TESE DE DOUTORAMENTO

BIG DATA ANALYSIS APPLICATION IN RENEWABLE ENERGY MARKET: WIND POWER

Hamid Bastani

ESCOLA DE DOUTORAMENTO INTERNACIONAL DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOCTORADO EN

ENERGIAS RENOVABLES Y SOSTENIBILIDAD ENERGETICA

SANTIAGO DE COMPOSTELA

2021

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

D. JUAN JOSÉ CASARES LONG, Catedrático de Universidad, y **D. JOSÉ ANTONIO SOUTO GONZÁLEZ**, Profesor Contratado Doctor, en calidad de DIRECTORES y TUTOR de la Tesis Doctoral titulada **Big Data Analysis Application in Renewable Energy Market: Wind Power**,

INFORMAN:

Que la presente Tesis se corresponde con el trabajo realizado por **D. HAMID BASTANI** bajo nuestra dirección/tutela en el Departamento de Ingeniería Química, Escuela Técnica Superior de Ingeniería (ETSE) de la USC; y autorizamos su presentación, considerando que reúne los requisitos exigidos en el **Regulamento de Estudos de Doutoramento de la USC**, y como directores/tutor de ésta no incurren en las causas de abstención establecidas en la Ley 40/2015.

En Santiago de Compostela, a 22 de Julio de 2021.

Prof. Dr. D. Juan José Casares Long Prof. Dr. D. José Antonio Souto González

DECLARACIÓN DO AUTOR/A DA TESE

D./Dna. **Hamid Bastani**

Título da tese: **Big Data Analysis Application in Renewable Energy Market: Wind Power**

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1) A tese abarca os resultados da elaboración do meu traballo.
2) De ser o caso, na tese faise referencia ás colaboracións que tivo este traballo.
3) Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.
4) A tese é a versión definitiva presentada para a súa defensa e coincide a versión impresa coa presentada en formato electrónico

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En **Santiago de Compostela**, **20 de Xullo de 2021**.

Sinatura electrónica

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

This thesis is dedicated to my loving parents to whom I owe everything.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Table of contents

Declaration

hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or partially for consideration of any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing that is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgments. This dissertation contains fewer than 82,000 words including appendices, bibliography, footnotes, tables and equations, with fewer than 110 figures.

Hamid Bastani
October 2021

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Acknowledgements

I would like to extend my sincere gratitude to my loving supervisors, Prof. Juan Casares Long and Prof. José Antonio Souto González, whose guidance, support, and confidence in me, brought me this chance to develop and terminate this thesis. A special thank from my heart to Juan for trusting in me and for his unconditional care and support with his amazing heart. A sincere appreciation to José Antonio for his tremendous support, scientific and technical supervision, and for his special kindness. It was a great experience working under your supervision.

I would like to thank my parents. Thank you for your unconditional love. Thank you for raising and nurturing me, and for dedicating all your time, energy, and life on me to be flourished and successful. This thesis is dedicated to you.

During the past few years, I had the opportunity to meet a lot of brilliant and smart people working in different fields of environmental modeling, renewable energies, and machine learning techniques in Spain, Iran and Germany. The list of names is endless. Thus, I wish to express my appreciation for their kind involvement, and their patience on resolving any doubts, and problems.

This thesis has been developed under European Union (EU) financial support through the MARHABA Lot 3 project of Erasmus-Mundus programme. The objective of MARHABA project directly corresponds to the development of further collaboration among EU and Asian Institutions for sending and hosting mobility of talented students and staff, and enhancement capacity. The program was open to the five top Iranian Universities including Sharif and Tehran University, for all engineering fields. After applying for the fellowship, I was nominated by the Tehran University and was so fortunate to be selected as one of the six granted doctorate candidates. I would like to thank the EU's programme to support education, training, youth and sport in Europe for the financial support to develop this work. I would also like to thank all the MARHABA project team members in the University of Santiago de

Compostela headquarter for their administrative help and support.

As a researcher, I would also like to extend my sincere gratitude to the following national institutes and companies, whose open-access data policy paved the path to develop this work: Meteogalicia, the regional meteorological agency for Galicia, Spain; Ministry of the environment, territory and housing of the xunta de Galicia, Spain; National renewable energy laboratory (NREL), United States; ENGIE-Green company located in France; and Sotavento wind company located at Galicia, Spain.

Last but not least, I would like to thank Spain and Galicia, Spanish and Galician. Thank you Spain for not only being my home during the previous several years, but also my launchpad for discovering the world, and for more professional opportunities. Thank you Galicia and Santiago de Compostela for your unparalleled beauty and lovely atmosphere.

RESUMO

As enerxías renovables son fontes de enerxía limpa, inesgotable e cada vez máis competitiva, desempeñando un papel clave no logro da Axenda 2030 para o Desenvolvemento Sustentable, o Acordo de París sobre o Cambio Climático e a Emisión Neta Cero para 2050. Recentemente, os diversos organismos sobre enerxía da ONU publicaron unha folla de ruta orientada ao logro do Obxectivo de Desenvolvemento Sustentable 7 (ODS 7) e as Emisións Netas Cero a través dunha iniciativa chamada Diálogo de Alto Nivel sobre Enerxía, que se celebrará en setembro de 2021 en Nova York. Esta folla de ruta contén cinco informes técnicos que abordan diferentes aspectos do sector enerxético, baseándose no desenvolvemento das enerxías renovables cara aos obxectivos antes mencionados. Entre os diferentes factores que derivan nun maior interese na transición ás enerxías renovables, os máis importantes son o impacto do sistema enerxético no cambio climático e a diminución dos recursos que ameazan a seguridade da subministración enerxética. As emisións de dióxido de carbono, como consecuencia de quéima de combustibles fósiles, desempeñan un papel considerable no cambio climático, ao crear efectos adversos nun clima cambiante, como o aumento da intensidade e a frecuencia dos desastres naturais. Como estratexia de mitigación do cambio climático, a transición completa a enerxía renovable para 2050 indícase como un obxectivo do Panel Intergobernamental de Expertos sobre o Cambio Climático (IPCC) para manter o quecemento global moi por baixo dos 2 graos centígrados, e co obxectivo de 1,5ºC. Doutra banda, a incerteza da subministración de combustibles fósiles, como recurso limitado, e o risco de alcanzar un pico do seu consumo cando o descenso da súa produción sexa a tendencia dominante, derivan na inseguridade da industria enerxética e do seu investimento en combustibles fósiles. Estas limitacións son fortes motivacións para reconsiderar o futuro do sector enerxético e avanzar cara a transición ás enerxías renovables. Además, o acceso global á enerxía, como meta 7.2 dos Obxectivos de Desenvolvemento Sustentable das Nacións Unidas (ONU), é outra motivación para investir en centrais eléctricas de enerxía renovable descentralizadas, para proporcionar acceso á electricidade a poboacións remotas a custos máis baixos.

Entre as enerxías renovables, a enerxía eólica é unha das tecnoloxías de rápido crecemento en todo o mundo coa caída dos custos de deseño, materiais, fabricación e instalación. A enerxía eólica foi a primeira fonte de enerxía renovable que se utilizou a gran escala na industria eléctrica e pioneira como as tecnoloxías máis estendidas no sector enerxético. O vento é un recurso enerxético flutuante, polo que a enerxía eólica non se pode despachar baixo demanda. De feito, a enerxía eólica xerada é moi variable (propiedade de alta volatilidade) nunha escala de tempo curta, con todo, é máis consistente dun ano a outro. As características intermitentes da velocidade do vento e a enerxía eólica fan que sexa bastante complexo equilibrar a oferta (mix de xeración) e a demanda en redes eléctricas intelixentes. Manter o equilibrio é aínda máis laborioso e custoso con fontes máis descentralizadas e volátiles conectadas á rede. Noutras palabras, unha proporción cada vez maior de fontes renovables volátiles no mix de xeración, incluída a enerxía eólica, daría lugar a fortes caídas de subministración non desexadas con maior frecuencia e a un aumento das fluctuaciones na subministración de enerxía. Por tanto, a predición da enerxía eólica requírese o máis precisa posible para garantir unha combinación de enerxía que proporcione unha subministración confiable, en combinación con outras fontes de enerxía. Isto pódese lograr mediante a aplicación de técnicas de xestión de enerxía, utilizando varios mecanismos como: fontes de enerxía de dispoñibilidade rápida (como centrais eléctricas de gas natural e hidroeléctricas), turbinas distribuídas xeograficamente, exportación e importación de enerxía cara a / desde áreas próximas, almacenamento na rede, exceso de capacidade, xestión e redución da demanda (cando a produción eólica é baixa). Canto maior sexa a proporción de capacidade de enerxía eólica instalada nunha rexión ou rede, máis instalacións alimentadas con fontes de enerxía convencionais necesítanse para apoiala. Noutras palabras, a medida que a penetración da xeración de enerxía eólica aumenta no mix enerxético as fluctuaciones na produción de enerxía serán máis visibles no sistema eléctrico e, en consecuencia, afectarán o prezo e ao mercado da electricidade. A predición e xestión da variabilidade da enerxía eólica utilizando modelos precisos de predición meteorolóxica, de velocidade do vento e de enerxía eólica permiten aos operadores das redes eléctricas (i.e., REE) equilibrar unha oferta e unha demanda máis fáciles nos sistemas de rede rexionais e nacionais. En países europeos como España, Alemaña e Dinamarca, onde a enerxía eólica ten unha alta participación no mix enerxético, os operadores e xestores das redes eléctricas necesitan coñecer a produción futura dos seus parques eólicos, que se utilizan para programar as operacións doutras centrais eléctricas e tamén con fins comerciais. A medida que creza o nivel de capacidade instalada de enerxía eólica, a previsión da produción de enerxía eólica crecerá en importancia. A industria eólica debe facer todo o posible para permitir que os operadores e xestores utilicen a enerxía eólica nas súas redes da maneira máis eficiente. Isto significa que as predicción de produción

agregada dos parques eólicos deben ser precisos.

Para facer fronte á incerteza da xeración de enerxía eólica desenvolvéronse numerosos modelos de predición de enerxía eólica e velocidade de vento baseados en métodos físicos, estatísticos e de aprendizaxe automática (machine learning, ML) para achegar o valor predito o máis posible ao valor observado. Os modelos físicos, tamén coñecidos como predición numérica do tempo (numerical weather prediction, NWP) son a base física dos programas de software que resolven as ecuacións físicas das condicións atmosféricas e os seus cambios ao longo do tempo. Estes modelos clasifícanse en dúas categorías principais: modelos globais que resolven as ecuacións primitivas para todo o sistema do planeta Terra, e modelos de área limitada (LAM) que cobren só un dominio limitado. O segundo grupo é diferente en varios termos, como simplificaciones de ecuacións, suposicións e formulación matemática. A aplicación de modelos físicos é máis útil para horizontes de tempo máis longos, principalmente predicións meteorolóxicas subestacionais a estacionales, que para predicións meteorolóxicas a moi curto, curto e medio prazo. Estes modelos poden abarcar todo o globo terráqueo, con todo, as súas saídas teñen baixas resolucións espaciais e non son directamente adecuados para a predición de velocidade do vento aplicable á predición da enerxía xerada en parques eólicos. Ademáis, os modelos NWP requiren un tempo de computación elevado.

Neste traballo, utilízanse os resultados de modelo de área limitada Weather Research and Forecasting (WRF) que cobren dominios e períodos de tempo seleccionados, como un modelo físico aplicado en todo o mundo, xa sexa como datos de entrada a os novos modelos de ML desenvolvidos como para comparar a exactitude dos novos modelos ML cos resultados WRF.

As solucións alternativas para reducir os inconvenientes da NWP na predición da velocidade do vento son os modelos estatísticos paramétricos e non paramétricos. Estes modelos utilizan NWP, medidas e datos históricos para realizar a predición da velocidade do vento mediante a estimación das relacións estatísticas entre as series temporais pasadas e actuais, e a aplicación dos patróns atopados á predición en períodos futuros. Nesta categoría de modelado estatístico aplícase unha ampla gama de técnicas de regresión e series temporais, como a autoregresión (AR), a media móbil (MA), ou a media móbil integrada autoregresiva (ARIMA). Aínda que un modelo específico pode proporcionar resultados precisos baseados nas series temporais (localización e período) que se analizaron, pasar a outras series temporais require unha nova análise estatística que ofrecerá outros resultados, e só esa técnica de

análise pódese exportar a diferentes localizacións / períodos, pero non os seus resultados.

Nos últimos anos, a aparición de técnicas baseadas en datos ou de aprendizaxe automática (ML) facilitou a capacidade de proporcionar predicións espaciais e temporais de alta resolución de velocidade de vento e enerxía eólica. Desenvolvéronse unha ampla gama de técnicas de aprendizaxe automática supervisadas e non supervisadas para a predición da xeración de enerxía renovable a curto e longo prazo, incluída a velocidade do vento e a enerxía eólica. Varios tipos de métodos de regresión e clasificación que forman modelos combinados, híbridos e conxuntos de predición de velocidade de vento e enerxía eólica, utilizando unha ampla gama de técnicas de ML, incluidas redes neuronais artificiais (ANN), máquinas de vectores de soporte (SSLM), sistemas difusos, Random Forest (RF), árbores de decisión e regresión de procesos gaussianos. Tendo en conta a aplicación exitosa e xeneralizada dos modelos ANN, tanto simples como complexos, na predición da velocidade de vento e a xeración de enerxía eólica, neste traballo desenvólvense tres modelos diferentes baseados na rede neuronal como núcleo, abordando tres problemas principais da predición de series temporais utilizando métodos de aprendizaxe automática: garantía de calidade dos datos empregados e imputación de datos non válidos, asignación de hiperparámetros dos modelos de predición e selección de características dos datos de entrada aos modelos de predición. Co fin de potenciar o núcleo dos modelos ANN utilízanse varias técnicas matemáticas de agrupación, optimización e procesamiento de sinais para crear novos modelos híbridos de predición de velocidade de vento e xeración eólica. Logo, estes modelos híbridos aplícanse a varios conxuntos de datos meteorolóxicos e de enerxía eólica en diferentes dominios, proporcionando predicións de velocidade do vento e enerxía eólica para moi curto, curto e medio prazo (desde minutos até horas por diante). En particular, a predición de enerxía eólica a curto prazo (poucas horas) xeralmente aplícase á integración operativa de plantas de enerxía eólica en redes eléctricas.

O rendemento dos novos modelos de aprendizaxe automática (ML, machine learning) desenvolvidos (fronte ás observacións) comparouse co doutras técnicas de predición, incluída a predición numérica rexional do tempo co modelo WRF, obténdose mellores estatísticos cos modelos ML. Como hoxe en día as plantas de enerxía eólica operativas xeralmente aplican métodos estatísticos específicos ou NWP para a predición da xeración de enerxía eólica, a aplicación de métodos ML pode aumentar a precisión das súas predicións, proporcionando unha operación da rede eléctrica máis estable; é dicir, unha predición menos precisa da enerxía eólica deriva nun uso máis convencional e flexible das fontes de enerxía, principalmente os combustibles fósiles, co fin de encher a brecha esperada entre a oferta e a demanda de electricidade. E, calquera erro na predición da enerxía eólica pode producir: (a) uso

excesivo de combustibles fósiles, con perdas simultáneas de xeración da enerxía eólica non conectada; (b) a falta de dispoñibilidade da subministración á rede eléctrica, o que supón un grave inconveniente.

Como consecuencia, este traballo representa un avance na análise de datos meteorolóxicos e de enerxía eólica, utilizando diferentes técnicas de aprendizaxe automática (redes reuronales artificiais, Artificial Neural Network, ANN) para comprender mellor as condicións climáticas en varios casos de estudo sobre diferentes dominios, cunha base sólida previa na revisión exhaustiva da aplicación de redes neuronais en modelos de predición de velocidade de vento e xeración eólica, desenvolvemento de novos modelos, axuste de modelos, e validación de modelos fronte a medidas operativas de velocidade de vento e de xeración de enerxía eólica.

En primeiro lugar, a revisión bibliográfica sistemática e exhaustiva realizada considérase unha base sólida para o desenvolvemento de tres modelos de ANN neste estudo. En particular, investígase e compárase o estado da arte da construción de redes neuronais (NN), incluídas as estruturas das redes, as funcións de activación e as combinacións de conxuntos ou híbridos de diversas técnicas matemáticas e modelos de predición. Recompiláronse ao redor de 300 artigos relevantes publicados en revistas de alto impacto entre 2010 e 2020 e filtráronse a través dun proceso sistemático para chegar a unha lista final de ao redor de 100 artigos que conteñen toda a información requirida relevante previamente identificada sobre a construción de NN, como características de datos de entrada, parámetros relacionados cos modelos e criterios de avaliación de resultados. Co fin de seleccionar as mellores opcións de construción para os novos modelos desenvolvidos, discútese o rendemento dos resultados de varios modelos tendo en conta as diferentes categorías dos compoñentes para a construción dunha ANN: arquitectura, función de activación, proceso de aprendizaxe e selección de características. Discútense e compáranse os rendementos dos modelos con varias arquitecturas de redes neuronais, considerando diferentes números de capas e neuronas por capa; os resultados mostraron que a arquitectura máis simple (1,2,1) proporciona resultados máis precisos con menos esforzo computacional. Sobre as funcións de activación, compárase un amplo rango (con capas ocultas e de saída), observando unha forte influencia na precisión da predición. Varios algoritmos de retropropagación e técnicas de optimización heurística próbanse no proceso de aprendizaxe da ANN, e identifícanse as súas vantaxes e inconvenientes. Finalmente, comparáronse varias técnicas de selección de características para a aprendizaxe supervisada aplicadas en diferentes estudos previos, incluídos os métodos baseados en envolturas e filtros; a partir desta comparación, os métodos híbridos, combinados e de conxunto demostraron

ser máis exitosos en comparación co modelo individual de ML, incluídos os modelos ANN individuais.

Esta revisión exhaustiva tamén cobre os criterios de avaliación dos modelos, co fin de permitir a comparación do rendemento dos novos modelos desenvolvidos co rendemento doutros modelos na literatura. Ademáis, proponse un novo criterio de avaliación, como a porcentaxe de mellora do rendemento do modelo (PI) fronte ao rendemento do modelo de persistencia. Como os resultados do modelo de persistencia sempre se obteñen directamente das series temporais de medidas dispoñibles, este criterio pódese aplicar en calquera dominio observado, polo que é posible comparar calquera rendemento do modelo aplicado en calquera outra parte onde se dispoña de conxuntos de medidas de velocidade do vento e / ou enerxía eólica.

Seguindo a metodoloxía ML, grandes conxuntos de datos de series temporais (Big Data) son necesarios para o desenvolvemento e adestramento dos modelos ML, e posteriormente, para a medición do rendemento dos modelos e a súa intercomparación. Por iso, neste traballo elíxense seis casos de estudo, que rexistran series temporais de datos meteorolóxicos en España, Francia e Estados Unidos, e datos de xeración de enerxía eólica en dous parques eólicos situados en España e Francia. Sobre os conxuntos de datos meteorolóxicos, selecciónanse tres estaciones de MeteoGalicia en Galicia, España, como representantes de diferentes características do terreo: áreas urbanas, rurais de montaña e costeiras. Ademais, selecciónase a estación meteorolóxica estadounidense de M2 Tower, como unha instalación específica para a monitorización da velocidade do vento a varias alturas. Respecto dos conxuntos de datos de enerxía eólica xerada, os parques eólicos de Sotavento e La Haute Bourne en Galicia e Francia, respectivamente, selecciónanse como dous casos de estudo diferentes.

Para comprender mellor as propiedades destas series temporais aplícanse varias técnicas de análises, como segue: Os parámetros meteorolóxicos de cada estación meteorolóxica visualízanse para períodos de tempo específicos, utilizando representacións gráficas uni e multiparamétricas, e compáranse series temporais anuais de velocidade do vento de varias estacións meteorolóxicas. Investíganse as autocorrelaciones de series temporais de velocidade do vento e as correlacións cruzadas entre varios parámetros, e tamén as observacións a diferentes alturas dispoñibles na estación meteorolóxica M2 Tower (EE.UU.). Analízanse as rosas dos ventos típicas, e tamén se obteñen as funcións de distribución da velocidade do vento de Weibull, Gamma e Rayleigh, e compáranse entre as diferentes estacións meteorolóxicas. Finalmente, realízanse diferentes análises espectrales de series temporais, utilizando os

métodos da Transformada Rápida de Fourier (Fast Fourier Transform, FFT) e Welch, para revelar a súa forte periodicidade ou estacionalidade. A análise destes conxuntos de datos dos distintos casos de estudo proporcionou unha visión xeral e unha mellor comprensión das súas características, o que é moi útil no desenvolvemento de modelos de predición de velocidade de vento e de xeración eólica.

Tendo en conta a revisión bibliográfica sobre os modelos ANN, a calidade do conxunto de datos de entrada é un paso importante no desenvolvemento e a aplicación do modelo, co fin de mellorar a súa precisión de predición. Por tanto, nos conxuntos de datos de observación dispoñibles, os datos faltantes ou non válidos deben completarse (algoritmo de imputación) antes de usalos como entrada ao modelo. Neste traballo, desenvólvese e aplícase un novo modelo de validación reconstrución de datos meteorolóxicos para completar as series temporais meteorolóxicas dispoñibles nos distintos casos de estudo. En primeiro lugar, a calidade dos datos dispoñibles mídese a través dun proceso de validación multinivel para marcar valores sospeitosos, faltantes e inválidos, é dicir, identificar períodos sen datos válidos. Aplicouse unha ampla gama de probas de validación, incluídas probas básicas e de rango, e probas de consistencia interna, temporal e espacial, aos seguintes parámetros meteorolóxicos: velocidade do vento, dirección do vento, temperatura, precipitación, radiación solar e humidade relativa. En segundo lugar, os períodos de datos xeradas por valores etiquetaxes como non válidos ou faltantes enchéronse en cada localización utilizando un modelo de rede neuronal de reconstrución cara adiante (Feed-forward Neural Network, FFNN) coa técnica de agrupamiento de k-medias multidimensionales, utilizando como datos de entrada: medidas dispoñibles na localización e resultados de predición operativa do modelo Weather Research and Forecasting (WRF) na cela da cuadrícula onde se atopa a localización considerada.

Este novo modelo de reconstrución foi validado fronte a varias series temporais medidas completas e validadas. Estas series temporais modificáronse para producir novos modelos sintéticos de datos con diferentes lonxitudes e patróns de dispersión de períodos de datos perdidos: 5%, 10%, 20% e 50%. Aplicouse a estratexia Missing Completely at Random (MCAR) para crear os períodos sintéticos de datos perdidos completamente ao azar. Ademáis, para simular datos consecutivos faltantes ou non válidos nunha serie temporal, utilizouse a estratexia Missing at Random (MAR), é dicir, producir tres patróns de datos perdidos consecutivos diferentes con lonxitudes pequenas (1-6h), medianas (6-24h) e grandes (24-72h) en series de tempo horarias. Como parte do desenvolvemento e validación do modelo de reconstrución, con estas series temporais sintéticas de medidas construíronse e adestráronse tres modelos FFNN diferentes utilizando como datos de entrada, tanto a saída WRF como as

series temporais sintéticas de medidas. O algoritmo de imputación resumido previamente mostrou mellores resultados que os outros dous modelos FFNN, e tamén mellor que a substitución directa de datos perdidos por resultados da predición WRF. Ademáis, observouse que agrupar os resultados de predición WRF como parte dos datos de entrada mellora o rendemento do modelo de reconstrución. Como resumo de resultados, obtéñense os seguintes rangos de estatísticos dos resultados do modelo de reconstrución, MSE e RMSE, utilizando as diferentes lonxitudes de períodos de datos perdidos descritos anteriormente: MSE (0,4, 0,6) e RMSE (0,6, 0,75) utilizando o novo modelo de reconstrución desenvolvido; MSE (0,83, 0,88) e RMSE (0,91, 0,94) usando o modelo ARIMA (1,1,1); e MSE (1,8, 2,3) e RMSE (1,3, 1,5) utilizando a saída directa do modelo WRF.

Despois de que se reconstruísen todas as series temporais de velocidade do vento correspondentes aos diversos casos de estudo, desenvolveuse un novo modelo ML de predición da velocidade do vento, denominado BBO-MLP, para horizontes temporais de predición moi curtos e curtos. Este modelo de predición híbrido está estruturado nun núcleo de predición de perceptrones multicapa, utilizando a optimización baseada nos sistemas biolóxicos (algoritmos xenéticos) para adestrar os seus parámetros. Desenvolveuse un novo algoritmo de procura semiexhaustivo para seleccionar o mellor conxunto de hiperparámetros en tres subespacios de solución diferentes: parámetros do modelo, algoritmo de optimización e datos de entrada. Defínense varios conxuntos de hiperparámetros, que compoñen o espazo de solución global, para cada unha do tres partes deste modelo híbrido: (a) para o núcleo do modelo de predición: número de nodos de entrada, número de nodos ocultos e función de activación; (b) para o algoritmo de optimización: tipo de método, rango de valores e número de pais e xeracións; e (c) para o conxunto de datos para adestramento: duración das series temporais, porcentaxe de proba (validación do adestramento) e características seleccionadas. Na maioría dos modelos previos de ML de predición de velocidade de vento e xeración eólica os hiperparámetros axústanse por separado para cada parte do modelo, o que descoida os posibles vínculos entre eles. Neste traballo, a optimización simultánea do tres conxuntos de hiperparámetros permite estudar os seus efectos mutuos, xa que tanto o conxunto de hiperparámetros óptimos como os conxuntos de datos para adestramento non son necesariamente independentes entre si, por tanto, recoméndase un proceso de optimización simultáneo para chegar conxuntamente ao mellor grupo de hiperparámetros que inclúe as diferentes partes do modelo. De feito, os resultados do modelo desenvolvido confirman esta dependencia entre o tres conxuntos diferentes. Tanto os criterios de información Bayesianos como os de Akaike xunto coa validación cruzada k-fold utilízanse para a selección da estrutura do modelo en cada fase iterativa do algoritmo de procura proposto. Aplícanse técnicas de media

móbil simples, ponderadas e exponenciais a diferentes combinacións de pasos de tempo anteriores para atopar o mellor conxunto de datos de entrada para o desenvolvemento do modelo de predición. Ademáis, seleccionáronse períodos altamente inestables nas series temporais anuais como conxuntos de datos de entrada, como un reto de maior dificultade para o modelo de predición.

Unha vez que o conxunto de hiperparámetros foi definido e o modelo adestrado, o modelo resultante aplicouse á predición de series temporais de velocidade de vento no catro estaciones meteorolóxicas seleccionadas, para a predición a curto e medio prazo da velocidade de vento. Estacións Santiago-EOAS (contorna urbana), Muralla (contorna rural de montaña), Camariñas (contorna costeira) en Galicia, España, e M2 Tower en Colorado, EE.UU. (coa singularidade de dispor de medidas de velocidade de vento a diferentes Alturas). Os resultados do novo modelo de predición desenvolvido BBO-MLP compáranse coas medidas de velocidade de vento dispoñibles no catro estaciones, usando os clásicos estatísticos MSE e RMSE. Comparados con outros modelos de predición, BBO-MLP mellora as predicións con 1-3-horas de adianto, cun paso adiante. Con todo, as predicións do modelo BBO-MLP a moi curto prazo (10-minutos) e con varios pasos adiante non son tan boas. Os resultados de MSE e RMSE do modelo BBO-MLP para as súas predicións con 1-hora de adianto melloran os do modelo de persistencia nun 28,75% e 12,73%, os do modelo ARIMA (1,1,1) nun 73,91% e 48,60%, os do modelo MLP BR nun 10,58% e 4,66%, respectivamente. No caso da predición con 3-horas de adianto, novamente para MSE e RMSE, os seus resultados melloran os do modelo de persistencia nun 47,50% e 31,86%, os do modelo ARIMA (1,1,1) nun 53,99% e 32,04%, e os do modelo MLP-BR nun 27,62% e 14,61%, respectivamente. No caso das estacións meteorolóxicas en Galicia, a predición con 1-hora de adianto do modelo BBO-MLP mellora sensiblemente a predición do modelo WRF, coas seguintes melloras en MSE e RMSE: En Camariñas (estación costeira), nun 88,27% e 66,56%, respectivamente; en Muralla (estación rural de montaña), nun 82,90% e 58,64%, respectivamente; e, en Santiago-EOAS (estación urbana), nun 83,27% e 59,09%, respectivamente.

Para a predición de enerxía eólica desenvolveuse un modelo de predición híbrido diferente, denominado EMD-QBPSO-ENN-GA, xa que este modelo baséase no uso de datos meteorolóxicos e de xeración eólica. O problema de selección das características nas redes neuronais identificouse como relevante para lograr bos resultados, polo que neste modelo desenvólvese e aplica un novo algoritmo de optimización de enxame de partículas binarias (QBPSO) baseado na teoría cuántica. Este novo algoritmo está composto por unha rede neuronal de Elman (ENN), un tipo de rede recursiva, que é optimizada mediante un algoritmo

xenético (GA), e tamén está equipada coa técnica de descomposición de modo empírico (EMD) para obter as características de inestabilidade dos datos de entrada aplicados ao desenvolvemento e adestramento do modelo. Este novo modelo híbrido EMD-QBPSO-ENN-GA aplicouse sobre dous parques eólicos diferentes: o parque eólico experimental Sotavento en Galicia, España, e o parque eólico La Haute Borne en Billancourt, Francia. Nestas aplicacións do modelo probáronse e aplicado tres conxuntos de datos de entrada diferentes para a lograr a predición da enerxía eólica: (1) medidas meteorolóxicas, (2) medidas de xeración eólica e; (3) no caso de Sotavento, tamén se teñen en conta os datos meteorolóxicos dispoñibles procedentes da predición do modelo operativo WRF (Meteogalicia).

Antes de aplicar este novo modelo probáronse varias configuracións de modelos EMD-QBPSO-ENN-GA utilizando diferentes conxuntos de datos de entrada, utilizando o erro cuadrático medio (RMSE), o erro absoluto medio (MAE) e R-cadrado, como estatísticos de medida do rendemento de cada modelo. En particular, empregáronse oito funciones de transferencia diferentes para o algoritmo QBPSO, de maneira que os resultados destas probas permitiron definir a combinación óptima de funcións de transferencia.

Unha vez definida a mellor estrutura do novo modelo, aplícase este modelo EMD-QBPSO-ENN-GA para predicir as dúas series temporais de enerxía eólica dos dous casos de estudo, tanto no parque eólico Sotavento con horizontes de predición curtos a medios (1 a 24-horas de antelación), como no parque eólico La Haute Borne, cun horizonte de predición a moi curto prazo, 10-minutos. No parque eólico Sotavento, seleccionando o mellor conxunto de datos de entrada os valores de RMSE de enerxía eólica obtidos con 1, 3, 6, 12 e 24 horas de adianto son 105,78, 163,98, 178,41, 215,99 e 414,74 (kWh), respectivamente; os valores de MAE de enerxía eólica cos mesmos horizontes temporais son 71,96, 115,39, 128,50 e 321,22 (kWh), respectivamente. Estes resultados mostran unha mellora significativa respecto ao modelo de persistencia, 61,85%, 50,01%, 41,62% e 47,27% para os mesmos horizontes temporais, respectivamente. En particular, tendo en conta o impacto das diferentes partes do modelo no seu rendemento, os resultados comparativos mostraron que o uso de GA e EMD melloraron os valores do RMSE para a predición con 1-hora de antelación nun 4% e un 21%, respectivamente. No parque eólico La Haute Borne, os valores de RMSE e MAE de enerxía eólica obtidos para a predición de 10-minutos de antelación son 89,46 e 56,42 (kWh), respectivamente.

Como resumo de resultados, o modelo de predición da velocidade do vento BBO-MLP mostra a súa capacidade para obter predicións precisas a curto prazo (1 3 horas por diante),

que son adecuadas para alimentar as ecuacións de deseño dos xeradores eólicos para estimar a súa xeración de enerxía eólica correspondente. Tamén o modelo de predición de enerxía eólica EMD-QBPSO-ENN-GA mostra a súa capacidade para obter predicións a curto prazo (1-6 horas con antelación) sobre o parque eólico ensaiado. Ademáis, ambos os modelos pódense executar sobre computadores persoais de elevada prestacións con tempos de resposta relativamente curtos. Por tanto, ambas as solucións son adecuadas para ser aplicadas á predición de enerxía eólica a curto prazo, como tecnoloxía necesaria para integrar a xeración dos parques eólicos nas redes eléctricas actuais, co fin de garantir a subministración eléctrica e minimizar o uso doutras fontes (combustibles fósiles).

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

RESUMEN

Las energías renovables son fuentes de energía limpia, inagotable y cada vez más competitiva, desempeñando un papel clave en el logro de la Agenda 2030 para el Desarrollo Sostenible, el Acuerdo de París sobre el Cambio Climático y la Emisión Neta Cero para 2050. Recientemente, los organismos sobre energía de la ONU han publicado una hoja de ruta de cara al logro del Objetivo de Desarrollo Sostenible 7 (ODS 7) y las Emisiones Netas Cero a través de una iniciativa llamada Diálogo de Alto Nivel sobre Energía, que se celebrará en septiembre de 2021 en Nueva York. Esta hoja de ruta contiene cinco informes técnicos que abordan diferentes aspectos del sector energético basándose en el desarrollo de las energías renovables hacia los objetivos antes mencionados. Entre los diferentes factores que derivan un mayor interés en la transición a las energías renovables, los más importantes son el impacto del sistema energético en el cambio climático y la disminución de los recursos que amenazan el aseguramiento del suministro energético. Las emisiones de dióxido de carbono, como consecuencia de la quema de combustibles fósiles, desempeñan un papel considerable en el cambio climático, al crear efectos adversos en un clima cambiante, como el aumento de la intensidad y la frecuencia de los desastres naturales. Como estrategia de mitigación del cambio climático, la transición completa a energía renovable para 2050 se indica como un objetivo del Panel Intergubernamental de Expertos sobre el Cambio Climático (IPCC) para mantener el calentamiento global muy por debajo de los 2 grados centígrados, y con el objetivo de 1,5ºC. Por otro lado, la incertidumbre del suministro de combustibles fósiles, como recurso limitado, y el riesgo de alcanzar un pico de su consumo cuando el decrecimiento de su producción sea la tendencia dominante, derivan en la inseguridad de la industria energética y de su inversión en combustibles fósiles. Estas limitaciones son fuertes motivaciones para reconsiderar el futuro del sector energético y avanzar hacia la transición a las energías renovables. Además, el acceso global a la energía, como meta 7.2 de los Objetivos de Desarrollo Sostenible de las Naciones Unidas (ONU), es otra motivación para invertir en centrales eléctricas de energía renovable descentralizadas, para proporcionar acceso a la electricidad a poblaciones remotas a costos más bajos.

Entre las energías renovables, la energía eólica es una de las tecnologías de rápido crecimiento en todo el mundo con la caída de los costes de diseño, materiales, fabricación e instalación. La energía eólica fue la primera fuente de energía renovable que se utilizó a gran escala en la industria eléctrica y pionera como las tecnologías más extendidas en el sector energético. El viento es un recurso energético fluctuante, por lo que la energía eólica no se puede despachar bajo demanda. De hecho, la energía eólica generada es muy variable (propiedad de alta volatilidad) en una escala de tiempo corta, sin embargo, es más consistente de un año a otro. Las características intermitentes de la velocidad del viento y la energía eólica hacen que sea bastante complejo equilibrar la oferta (mix de generación) y la demanda en redes eléctricas inteligentes. Mantener el equilibrio es aún más laborioso y costoso con fuentes más descentralizadas y volátiles conectadas a la red. En otras palabras, una proporción cada vez mayor de fuentes renovables volátiles en el mix de generación, incluida la energía eólica, daría lugar a fuertes caídas de suministro no deseadas con mayor frecuencia y a un aumento de las fluctuaciones en el suministro de energía. Por lo tanto, la predicción de la energía eólica se requiere lo más precisa posible para garantizar un mix de energía que proporcione un suministro confiable, en combinación con otras fuentes de energía. Esto se puede lograr mediante la aplicación de técnicas de gestión de energía, utilizando varios mecanismos como: fuentes de energía de disponibilidad rápida (como centrales eléctricas de gas natural e hidroeléctricas), turbinas distribuidas geográficamente, exportación e importación de energía hacia / desde áreas cercanas, almacenamiento en la red, exceso de capacidad, gestión y reducción de la demanda (cuando la producción eólica es baja). Cuanto mayor sea la proporción de capacidad de energía eólica instalada en una región o red, más fuentes de energía convencionales se necesitan para respaldarla. En otras palabras, a medida que la penetración de la generación de energía eólica aumenta en el mix energético las fluctuaciones en la producción de energía serán más visibles en el sistema eléctrico y, en consecuencia, afectarán al precio y al mercado de la electricidad. La predicción y gestión de la variabilidad de la energía eólica utilizando modelos precisos de predicción meteorológica, de velocidad del viento y de energía eólica permiten a los operadores de las redes eléctricas (i.e., REE) equilibrar una oferta y una demanda más fáciles en los sistemas de red regionales y nacionales. En países europeos como España, Alemania y Dinamarca, donde la energía eólica tiene una alta participación en la combinación energética, los operadores y gestores de las redes eléctricas necesitan conocer la producción futura de sus parques eólicos, que se utilizan para programar las operaciones de otras centrales eléctricas y también con fines comerciales. A medida que crezca el nivel de capacidad instalada de energía eólica, la previsión de la producción de energía eólica crecerá en importancia. La industria eólica debe hacer todo lo posible para permitir que los operadores y gestores utilicen la energía eólica

en sus redes de la manera más eficiente. Esto significa que los pronósticos agregados de producción de los parques eólicos deben ser precisos.

Para hacer frente a la incertidumbre de la generación de energía eólica se han desarrollado numerosos modelos de predicción de energía eólica y velocidad de viento basados en métodos físicos, estadísticos y de aprendizaje automático (machine learning, ML) para acercar el valor predicho lo más posible al valor observacional. Los modelos físicos, también conocidos como predicción numérica del tiempo (numerical weather prediction, NWP) son la base física de los programas de software que resuelven las ecuaciones físicas de las condiciones atmosféricas y sus cambios a lo largo del tiempo. Estos modelos se clasifican en dos categorías principales: modelos globales que resuelven las ecuaciones primitivas para todo el sistema del planeta Tierra, y modelos de área limitada (LAM) que cubren solo un dominio limitado. El segundo grupo es diferente en varios términos, como simplificaciones de ecuaciones, suposiciones y formulación matemática. La aplicación de modelos físicos es más útil para horizontes de tiempo más largos, principalmente predicciones meteorológicas subestacionales a estacionales, que para predicciones meteorológicas a muy corto, corto y medio plazo. Estos modelos pueden abarcar todo el globo terráqueo, sin embargo, sus salidas tienen bajas resoluciones espaciales y no son directamente adecuados para la predicción de velocidad del viento aplicable al pronóstico de la energía generada en parques eólicos. Además, los modelos NWP requieren un tiempo de computación elevado.

En este trabajo, se utilizan los resultados de modelo de área limitada Weather Research and Forecasting (WRF) que cubren dominios y períodos de tiempo seleccionados, como un modelo físico aplicado en todo el mundo, ya sea como datos de entrada los nuevos modelos de ML desarrollados como para comparar la exactitud de los nuevos modelos ML con los resultados WRF.

Las soluciones alternativas para reducir los inconvenientes del NWP en el pronóstico de la velocidad del viento son los modelos estadísticos paramétricos y no paramétricos. Estos modelos utilizan NWP, medidas y datos históricos para realizar la predicción de la velocidad del viento mediante la estimación de las relaciones estadísticas entre las series temporales pasadas y actuales, y la aplicación de los patrones encontrados al pronóstico en períodos futuros. En esta categoría de modelado estadístico se aplica una amplia gama de técnicas de regresión y series temporales, como la autoregresión (AR), la media móvil (MA), o la media móvil integrada autoregresiva (ARIMA). Si bien un modelo específico puede proporcionar resultados precisos basados en las series temporales (ubicación y período) que se analizaron,

pasar a otras series temporales requiere un nuevo análisis estadístico que ofrecerá otros resultados, y solo esa técnica de análisis se puede exportar a diferentes ubicaciones / períodos, pero no sus resultados.

En los últimos años, la aparición de técnicas basadas en datos o de aprendizaje automático (ML) ha facilitado la capacidad de proporcionar predicciones espaciales y temporales de alta resolución de velocidad de viento y energía eólica. Se han desarrollado una amplia gama de técnicas de aprendizaje automático supervisadas y no supervisadas para la predicción de la generación de energía renovable a corto y largo plazo, incluida la velocidad del viento y la energía eólica. Varios tipos de métodos de regresión y clasificación que forman modelos combinados, híbridos y conjuntos de predicción de velocidad de viento y energía eólica, utilizando una amplia gama de técnicas de ML, incluidas redes neuronales artificiales (ANN), máquinas de vectores de soporte (SSLM), sistemas difusos, Random Forest (RF), árboles de decisión y regresión de procesos gaussianos. Teniendo en cuenta la aplicación exitosa y generalizada de los modelos ANN, tanto simples como complejos, en la predicción de la velocidad de viento y la generación de energía eólica, en este trabajo se desarrollan tres modelos diferentes basados en la red neuronal como núcleo, abordando tres problemas principales de la predicción de series temporales utilizando métodos de aprendizaje automático: garantía de calidad de los datos empleados e imputación de datos no válidos, asignación de hiperparámetros de los modelos de predicción y selección de características de los datos de entrada a los modelos de predicción. Con el fin de potenciar el núcleo de los modelos ANN se utilizan varias técnicas matemáticas de agrupación, optimización y procesamiento de señales para crear nuevos modelos híbridos de predicción de velocidad de viento y generación eólica. Luego, estos modelos híbridos se aplican a varios conjuntos de datos meteorológicos y de energía eólica en diferentes dominios, proporcionando pronósticos de velocidad del viento y energía eólica para muy corto, corto y medio plazo (desde minutos hasta horas por delante). En particular, la predicción de energía eólica a corto plazo (pocas horas) generalmente se aplica a la integración operativa de plantas de energía eólica en redes eléctricas.

El rendimiento de los nuevos modelos de aprendizaje automático (ML, machine learning) desarrollados (frente a las observaciones) se comparó con el de otras técnicas de predicción, incluida la predicción numérica regional del tiempo con el modelo WRF, obteniéndose mejores estadísticos con los modelos ML. Como hoy en día las plantas de energía eólica operativas generalmente aplican métodos estadísticos específicos o NWP para la predicción de la generación de energía eólica, la aplicación de métodos ML puede aumentar la precisión de sus predicciones, proporcionando una operación de la red eléctrica más estable; es decir,

una predicción menos precisa de la energía eólica deriva en un uso más convencional y flexible de las fuentes de energía, principalmente los combustibles fósiles, con el fin de llenar la brecha esperada entre la oferta y la demanda de electricidad. Y, cualquier error en la predicción de la energía eólica puede producir: (a) uso excesivo de combustibles fósiles, con pérdidas simultáneas de generación de la energía eólica no conectada; b) la falta de disponibilidad del suministro a la red eléctrica, lo que supone un grave inconveniente.

Como consecuencia, este trabajo representa un avance en el análisis de datos meteorológicos y de energía eólica, utilizando diferentes técnicas de aprendizaje automático (redes reuronales artificiales, Artificial Neural Network, ANN) para comprender mejor las condiciones climáticas en varios casos de estudio sobre diferentes dominios, con una base sólida previa en la revisión exhaustiva de la aplicación de redes neuronales en modelos de predicción de velocidad de viento y generación eólica, desarrollo de nuevos modelos, ajuste de modelos, y validación de modelos frente a medidas operativas de velocidad de viento y de generación de energía eólica.

En primer lugar, la revisión bibliográfica sistemática y exhaustiva realizada se considera una base sólida para el desarrollo de los tres modelos de ANN en este estudio. En particular, se investiga y compara el estado del arte de la construcción de redes neuronales (NN), incluidas las estructuras de redes, las funciones de activación y las combinaciones de conjuntos o híbridos de diversas técnicas matemáticas y modelos de predicción. Se recopilaron alrededor de 300 artículos relevantes publicados en revistas de alto impacto entre 2010 y 2020 y se filtraron a través de un proceso sistemático para llegar a una lista final de alrededor de 100 artículos que contienen toda la información requerida relevante previamente identificada sobre la construcción de NN, como características de datos de entrada, parámetros relacionados con los modelos y criterios de evaluación de resultados. Con el fin de seleccionar las mejores opciones de construcción para los nuevos modelos desarrollados, se discute el rendimiento de los resultados de varios modelos teniendo en cuenta las diferentes categorías de los componentes para la construcción de una ANN: arquitectura, función de activación, proceso de aprendizaje y selección de características. Se discuten y comparan los rendimientos de los modelos con varias arquitecturas de redes neuronales, considerando diferentes números de capas y neuronas por capa; los resultados mostraron que la arquitectura más simple (1,2,1) proporciona resultados más precisos con menos esfuerzo computacional. Sobre las funciones de activación, se compara un amplio rango (con capas ocultas y de salida), observando una fuerte influencia en la precisión de la predicción. Varios algoritmos de retropropagación y técnicas de optimización heurística se prueban en el proceso de aprendizaje de la ANN, y

se identifican sus ventajas e inconvenientes.Finalmente, se compararon varias técnicas de selección de características para el aprendizaje supervisado aplicadas en diferentes estudios previos, incluidos los métodos basados en envolturas y filtros; a partir de esta comparación, los métodos híbridos, combinados y de conjunto han demostrado ser más exitosos en comparación con el modelo individual de ML, incluidos los modelos ANN individuales.

Esta revisión exhaustiva también cubre los criterios de evaluación de los modelos, con el fin de permitir la comparación del rendimiento de los nuevos modelos desarrollados con el rendimiento de otros modelos en la literatura. Además, se propone un nuevo criterio de evaluación, como el porcentaje de mejora del rendimiento del modelo (PI) frente al rendimiento del modelo de persistencia. Como los resultados del modelo de persistencia siempre se obtienen directamente de las series temporales de medidas disponibles, este criterio se puede aplicar en cualquier dominio observado, por lo que es posible comparar cualquier rendimiento del modelo aplicado en cualquier otra parte donde se disponga de conjuntos de medidas de velocidad del viento y / o energía eólica.

Siguiendo la metodología ML, grandes conjuntos de datos de series temporales (Big Data) son necesarios para el desarrollo y entrenamiento de los modelos ML, y posteriormente, para la medición del rendimiento de los modelos y su intercomparación. Por ello, en este trabajo se eligen seis casos de estudio, que registran series temporales de datos meteorológicos en España, Francia y Estados Unidos, y datos de generación de energía eólica en dos parques eólicos ubicados en España y Francia. Sobre los conjuntos de datos meteorológicos, se seleccionan tres estaciones de MeteoGalicia en Galicia, España, como representantes de diferentes características del terreno: áreas urbanas, rurales de montaña y costeras. Además, se selecciona la estación meteorológica estadounidense de M2 Tower, como una instalación específica para la monitorización de la velocidad del viento a varias alturas. Respecto a los conjuntos de datos de energía eólica generada, los parques eólicos de Sotavento y La Haute Bourne en Galicia y Francia, respectivamente, se seleccionan como dos casos de estudio diferentes.

Para comprender mejor las propiedades de estas series temporales se aplican varias técnicas de análisis, como sigue: Los parámetros meteorológicos de cada estación meteorológica se visualizan para períodos de tiempo específicos, utilizando representaciones gráficas uni y multiparamétricas, y se comparan series temporales anuales de velocidad del viento de varias estaciones meteorológicas. Se investigan las autocorrelaciones de series temporales de velocidad del viento y las correlaciones cruzadas entre varios parámetros, y también

las observaciones a diferentes alturas disponibles en la estación meteorológica M2 Tower (EE.UU.). Se analizan las rosas de los vientos típicas, y también se obtienen las funciones de distribución de la velocidad del viento de Weibull, Gamma y Rayleigh, y se comparan entre las diferentes estaciones meteorológicas. Finalmente, se realizan diferentes análisis espectrales de series temporales, utilizando los métodos de Transformada Rápida de Fourier (Fast Fourier Transform, FFT) y Welch, para revelar su fuerte periodicidad o estacionalidad. El análisis de estos conjuntos de datos de los distintos casos de estudio ha proporcionado una visión general y una mejor comprensión de sus características, lo que es muy útil en el desarrollo de modelos de predicción de velocidad de viento y de generación eólica.

Teniendo en cuenta la revisión bibliográfica sobre los modelos ANN, la calidad del conjunto de datos de entrada es un paso importante en el desarrollo y la aplicación del modelo, con el fin de mejorar su precisión de predicción. Por lo tanto, en los conjuntos de datos de observación disponibles, los datos faltantes o no válidos deben completarse (algoritmo de imputación) antes de usarlos como entrada al modelo. En este trabajo, se desarrolla y aplica un nuevo modelo de validación-reconstrucción de datos meteorológicos para completar las series temporales meteorológicas disponibles en diferentes sitios de prueba. En primer lugar, la calidad de los datos disponibles se mide a través de un proceso de validación multinivel para marcar valores sospechosos, faltantes e inválidos, es decir, identificar períodos sin datos válidos. Se aplicó una amplia gama de pruebas de validación, incluidas pruebas básicas y de rango, y pruebas de consistencia interna, temporal y espacial, a los siguientes parámetros meteorológicos: velocidad del viento, dirección del viento, temperatura, precipitación, radiación solar y humedad relativa. En segundo lugar, los períodos de datos generadas por valores etiquetados como no válidos o faltantes se rellenaron en cada localización utilizando un modelo de red neuronal de reconstrucción hacia delante (Feed-forward Neural Network, FFNN) con la técnica de agrupamiento de k-medias multidimensionales, utilizando como datos de entrada: medidas disponibles en la localización y resultados de predicción operativa del modelo Weather Research and Forecasting (WRF) en la celda de la cuadrícula donde se encuentra la localización considerada.

Este nuevo modelo de reconstrucción fue validado frente a varias series temporales medidas completas y validadas. Estas series temporales se modificaron para producir nuevos modelos sintéticos de datos con diferentes longitudes y patrones de dispersión de períodos de datos perdidos: 5%, 10%, 20% y 50%. Se aplicó la estrategia Missing Completely at Random (MCAR) para crear los períodos sintéticos de datos perdidos completamente al azar. Además, para simular datos consecutivos faltantes o no válidos en una serie tempo-

ral, se utilizó la estrategia Missing at Random (MAR), es decir, producir tres patrones de datos perdidos consecutivos diferentes con longitudes pequeñas (1-6h), medianas (6 24h) y grandes (24-72h) en series de tiempo horarias. Como parte del desarrollo y validación del modelo de reconstrucción, con estas series temporales sintéticas de medidas se construyeron y entrenaron tres modelos FFNN diferentes utilizando tanto la salida WRF como esas series temporales sintéticas de medidas como datos de entrada. El algoritmo de imputación resumido previamente mostró mejores resultados que los otros dos modelos FFNN, y también mejor que la sustitución directa de datos perdidos por resultados de la predicción WRF. Además, se observó que agrupar los resultados de predicción WRF como parte de los datos de entrada mejora el rendimiento del modelo de reconstrucción. Como resumen de resultados, se obtienen los siguientes rangos de estadísticos de los resultados del modelo de reconstrucción, MSE y RMSE, utilizando las diferentes longitudes de períodos de datos perdidos descritos anteriormente: MSE (0,4, 0,6) y RMSE (0,6, 0,75) utilizando el nuevo modelo de reconstrucción desarrollado; MSE (0,83, 0,88) y RMSE (0,91, 0,94) usando el modelo ARIMA (1,1,1); y MSE (1,8, 2,3) y RMSE (1,3, 1,5) utilizando la salida directa del modelo WRF.

Después de que se reconstruyesen todas las series temporales de velocidad del viento correspondientes a los diversos casos de estudio, se ha desarrollado un nuevo modelo ML de predicción de la velocidad del viento, denominado BBO-MLP, para horizontes temporales de predicción muy cortos y cortos. Este modelo de predicción híbrido está estructurado en un núcleo de predicción de perceptrones multicapa, utilizando la optimización basada en la los sistemas biológicos (algoritmos genéticos) para entrenar sus parámetros. Se ha desarrollado un nuevo algoritmo de búsqueda semiexhaustivo para seleccionar el mejor conjunto de hiperparámetros en tres subespacios de solución diferentes: parámetros del modelo, algoritmo de optimización y datos de entrada. Se definen varios conjuntos de hiperparámetros, que componen el espacio de solución global, para cada una de las tres partes de este modelo híbrido: (a) para el núcleo del modelo de predicción: número de nodos de entrada, número de nodos ocultos y función de activación; (b) para el algoritmo de optimización: tipo de método, rango de valores y número de padres y generaciones; y (c) para el conjunto de datos para entrenamiento: duración de las series temporales, porcentaje de prueba (validación del entrenamiento) y características seleccionadas. En la mayoría de los modelos previos de ML de predicción de velocidad de viento y generación eólica los hiperparámetros se ajustan por separado para cada parte del modelo, lo que descuida los posibles vínculos entre ellos. En este trabajo, la optimización simultánea de los tres conjuntos de hiperparámetros permite estudiar sus efectos mutuos, ya que tanto el conjunto de hiperparámetros óptimos como los

conjuntos de datos para entrenamiento no son necesariamente independientes entre sí, por lo tanto, se recomienda un proceso de optimización simultáneo para llegar conjuntamente al mejor grupo de hiperparámetros que incluye las diferentes partes del modelo. De hecho, los resultados del modelo desarrollado confirman esta dependencia entre los tres conjuntos diferentes. Tanto los criterios de información Bayesianos como los de Akaike junto con la validación cruzada k-fold se utilizan para la selección de la estructura del modelo en cada fase iterativa del algoritmo de búsqueda propuesto. Se aplican técnicas de media móvil simples, ponderadas y exponenciales a diferentes combinaciones de pasos de tiempo anteriores para encontrar el mejor conjunto de datos de entrada para el desarrollo del modelo de predicción. Además, se seleccionaron períodos altamente inestables en las series temporales anuales como conjuntos de datos de entrada, como un reto de mayor dificultad para el modelo de predicción.

Una vez que el conjunto de hiperparámetros fue definido y el modelo entrenado, el modelo resultante se aplicó a la predicción de series temporales de velocidad de viento en las cuatro estaciones meteorológicas seleccionadas, para la predicción a corto y medio plazo de la velocidad de viento. Estaciones Santiago-EOAS (entorno urbano), Muralla (entorno rural de montaña), Camariñas (entorno costero) en Galicia, España, y M2 Tower en Colorado, EE.UU. (con la singularidad de disponer de medidas de velocidad de viento a diferentes Alturas). Los resultados del nuevo modelo de predicción desarrollado BBO-MLP se comparan con las medidas de velocidad de viento disponibles en las cuatro estaciones, usando los clásicos estadísticos MSE y RMSE. Comparados con otros modelos de predicción, BBO-MLP mejora las predicciones con 1-3-horas de adelanto, con un paso adelante. Sin embargo, las predicciones del modelo BBO-MLP a muy corto plazo (10-minutos) y con varios pasos adelante no son tan buenas. BBO-MLP Los resultados de MSE y RMSE del modelo BBO-MLP para sus predicciones con 1-hora de adelanto mejoran los del modelo de persistencia en un 28,75% y 12,73%, los del modelo ARIMA (1,1,1) en un 73,91% y 48,60%, los del modelo MLP-BR en un 10,58% y 4,66%, respectivamente. En el caso de la predicción con 3-horas de adelanto, nuevamente para MSE y RMSE, sus resultados mejoran los del modelo de persistencia en un 47,50% y 31,86%, los del modelo ARIMA (1,1,1) en un 53,99% y 32,04%, y los del modelo MLP-BR en un 27,62% y 14,61%, respectivamente. En el caso de las estaciones meteorológicas en Galicia, la predicción con 1-hora de adelanto del modelo BBO-MLP mejora sensiblemente la predicción del modelo WRF, con las siguientes mejoras en MSE y RMSE: En Camariñas (estación costera), en un 88,27% y 66,56%, respectivamente; en Muralla (estación rural de montaña), en un 82,90% y 58,64%, respectivamente;

y, en Santiago-EOAS (estación urbana), en un 83,27% y 59,09%, respectivamente.

Para la predicción de energía eólica se ha desarrollado un modelo de predicción híbrido diferente, denominado EMD-QBPSO-ENN-GA, ya que este modelo se basa en el uso de datos meteorológicos y de generación eólica. El problema de selección de las características en las redes neuronales se ha identificado como relevante para lograr buenos resultados, por lo que en este modelo se desarrolla y aplica un nuevo algoritmo de optimización de enjambre de partículas binarias (QBPSO) basado en la teoría cuántica. Este nuevo algoritmo está compuesto por una red neuronal de Elman (ENN), un tipo de red recursiva, que es optimizada mediante un algoritmo genético (GA), y también está equipada con la técnica de descomposición de modo empírico (EMD) para obtener las características de inestabilidad de los datos de entrada aplicados al desarrollo y entrenamiento del modelo. Este nuevo modelo híbrido EMD-QBPSO-ENN-GA se ha aplicado sobre dos parques eólicos diferentes: el parque eólico experimental Sotavento en Galicia, España, y el parque eólico La Haute Borne en Billancourt, Francia. En estas aplicaciones del modelo se han probado y aplicado tres conjuntos de datos de entrada diferentes para la lograr la predicción de la energía eólica: (1) medidas meteorológicas, (2) medidas de generación eólica y; (3) en el caso de Sotavento, también se tienen en cuenta los datos meteorológicos disponibles procedentes de la predicción del modelo operativo WRF (MeteoGalicia).

Antes de aplicar este nuevo modelo se probaron varias configuraciones de modelos EMD-QBPSO-ENN-GA utilizando diferentes conjuntos de datos de entrada, utilizando el error cuadrático medio (RMSE), el error absoluto medio (MAE) y R-cuadrado, como estadísticos de medida del rendimiento de cada modelo. En particular, se emplearon ocho funciones de transferencia diferentes para el algoritmo QBPSO, de manera que los resultados de estas pruebas permitieron definir la combinación óptima de funciones de transferencia.

Una vez definida la mejor estructura del nuevo modelo, se aplica este modelo EMD-QBPSO-ENN-GA para predecir las dos series temporales de energía eólica de los dos casos de estudio, tanto en el parque eólico Sotavento con horizontes de predicción cortos a medios (1 a 24-horas de antelación), como en el parque eólico la Haute Borne, con un horizonte de predicción a muy corto plazo, 10-minutos. En el parque eólico Sotavento, seleccionando el mejor conjunto de datos de entrada los valores de RMSE de energía eólica obtenidos con 1, 3, 6, 12 y 24 horas de adelanto son 105,78, 163,98, 178,41, 215,99 y 414,74 (kWh), respectivamente; los valores de MAE de energía eólica con los mismos horizontes temporales son 71,96, 115,39, 128,50 y 321,22 (kWh), respectivamente. Estos resultados muestran una

mejora significativa respecto al modelo de persistencia, 61,85%, 50,01%, 41,62% y 47,27% para los mismos horizontes temporales, respectivamente. En particular, teniendo en cuenta el impacto de las diferentes partes del modelo en su rendimiento, los resultados comparativos mostraron que el uso de GA y EMD mejoraron los valores del RMSE para la predicción con 1-hora de antelación en un 4% y un 21%, respectivamente. En el parque eólico La Haute Borne, los valores de RMSE y MAE de energía eólica obtenidos para la predicción de 10-minutos de antelación son 89,46 y 56,42 (kWh), respectivamente.

Como resumen de resultados, el modelo de predicción de la velocidad del viento BBO-MLP muestra su capacidad para obtener predicciones precisas a corto plazo (1-3 horas por delante), que son adecuadas para alimentar las ecuaciones de diseño de los generadores eólicos para estimar su generación de energía eólica correspondiente. También el modelo de predicción de energía eólica EMD-QBPSO-ENN-GA muestra su capacidad para obtener predicciones a corto plazo (1-6 horas con antelación) sobre el parque eólico ensayado. Además, ambos modelos se pueden ejecutar sobre ordenadores personales de elevada prestaciones con tiempos de respuesta relativamente cortos. Por lo tanto, ambas soluciones son adecuadas para ser aplicadas a la predicción de energía eólica a corto plazo, como tecnología necesaria para integrar la generación de los parques eólicos en las redes eléctricas actuales, con el fin de garantizar el suministro eléctrico y minimizar el uso de otras fuentes (combustibles fósiles).

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

SUMMARY

Renewable energies are sources of clean, inexhaustible, and increasingly competitive energy, playing a key role in achievement of 2030 Agenda for sustainable development, the Paris Agreement on climate change, and the net-zero emission by 2050. Recently a roadmap towards achievement of both sustainable development goal 7 (SDG7) and net-zero emission has published by the UN-energy organizations through an international initiative named high-level dialogue on energy, that was hold on September 2021 in New York. The roadmap contains five technical reports addressing different aspects of the energy sector relying on renewable energy development towards the aforementioned goals. Among different factors deriving increased interest in renewable energy transition, the energy system impact on climate change and the diminishing resources that threaten energy security are the most important. Carbon dioxide emissions, as a consequence of burning fossil fuels, play a considerable role in climate change, by way of creating adverse effects on a changing climate such as increased intensity and frequency of natural disasters. As a climate change mitigation strategy, complete renewable energy transition by 2050 is indicated as a target by Intergovernmental Panel on Climate Change (IPCC) to keep global warming well below 2 Celsius degrees, and on 1.5 ºC trajectory target. From the other side, uncertainty with the supply of fossil fuels as a limited resource, and the risk of reaching a consumption peak when diminishing returns will become dominant, derive in the insecurity of the energy industry and its investment in fossil fuels. These driving limitations are strong motivations to reconsider future of the energy sector and moving towards renewable energy transition. Furthermore, global energy access, as target 7.2 of United Nations (UN) sustainable development goals, is another motivation to invest on decentralized renewable energy power plants to provide electricity access to remote populations at lower costs.

Among renewable energies, wind power is one of the fast-growing worldwide technologies with falling costs of design, material, manufacturing and establishment. Wind power was the first renewable energy source to be used on a large scale in the energy industry and pioneer as the most widespread technologies in the energy sector. Wind is a fluctu-

ated energy resource, so wind power cannot be dispatched on demand. Indeed, generated wind power is highly variable (high volatility property) over short time scale, however, it is more consistent from year to year. Intermittent characteristics of wind speed and wind power makes it quite complex to balance out supply (energy mix) and demand in smart electric grids. Maintaining the balance is even more laborious and expensive with more decentralized and volatile sources connected to the grid. In other words, a rising share of volatile renewable sources in energy mix, including wind energy, result in more often and undesirable heavy drops and increase (fluctuations) in energy supply. Therefore, wind energy prediction is required as accurate as possible in order to guarantee an energy mix that provides a reliable supply, in combination with other power sources. This one can be achieved by applying power-management techniques, using several mechanisms as: quick availability power sources (like natural gas-fired and hydroelectric power plants), geographically distributed turbines, power export and import to/from close areas, grid storage, excess capacity, demand management, and demand reduction (when wind production is low). The more the proportion of installed wind power capacity in a region or grid, the more conventional power sources needed to back it up. In other words, as the penetration of wind power generation increases in the energy mix, fluctuations in energy output will be more visible on the electric system and will consequently affect the electricity price and market. Predicting and managing wind energy variability using accurate weather, wind speed and wind power forecasting models allow transmission system operators (TSOs) to balance easier supply and demand on regional and national grid systems. In European countries, such as Spain, Germany and Denmark, where wind energy has a high share in energy mix, operators, managers and TSOs are routinely forecasting the output from their wind farms, which are used to schedule the operations of other power plants and also for trading purposes. As the level of installed wind power capacity grows, forecasting the wind energy production will grow in importance. The wind industry must do its very best to allow the TSOs to use wind energy to its best effect. This means aggregated wind farms output forecasts must be accurate.

To deal with wind power generation uncertainty, numerous wind energy and speed forecasting models based on physical, statistical, and machine learning (ML) methods are developed to bring the predicted value as closely as possible to the observational value. Physical models, also known as numerical weather prediction (NWP), are the physical basis of software programs that solve the physical equations of atmospheric conditions and their changes over time. These models are classified in two main categories: global models that solve the primitive equations for the whole Earth system, and limited area models (LAM) covering only a limited domain. The second group are different in several terms, such as

equations simplifications, assumptions, and mathematical formulation. Physical models application is more useful for longer time horizons, mainly sub-seasonal to seasonal weather predictions, than for very-short, short-, and medium-term weather forecasting. These models cover all around the world geographically. However, their outputs have low spatial resolutions, and they are not directly suitable to wind speed forecast to be applied for wind energy prediction at wind farms. In addition, NWP models require extensive computational time.

In this work, LAM outputs of Weather Research and Forecasting (WRF) covering selected domain and time periods, as a worldwide applied physical model, are used, either as inputs to the new developed ML models or to compare the accuracy between WRF and ML models.

Alternative solutions to reduce the NWP drawbacks in wind speed forecast are parametric and non-parametric statistical models. These use NWPs, observational and historical data to perform wind speed prediction by estimating statistical relationships between the past and current time series and applying discovered patterns for the future time step forecasting. A wide range of regression and time series techniques, such as auto-regression (AR), moving average (MA), autoregressive integrated moving average (ARIMA) are applied in this statistical modelling category. While a specific model can provide accuracy results based on the time series (location and period) were analysed, moving to other time series requires a new statistical analysis, and only the analysis technique can be exported to different locations/periods, but not its results.

In the last years, the emergence of data-driven or machine learning (ML) techniques has given the capability to provide high-resolution spatial and temporal predictions for wind speed and power. A wide range of supervised and un-supervised machine learning techniques are developed for short- to long-term renewable energy prediction, including wind speed and power. Various types of regression and classification methods that form combined, hybrid and ensemble wind speed and power prediction models can be found in literature, using a wide range of ML techniques, including artificial neural networks (ANN), support vector machines (SVMs), fuzzy systems, random forest (RF), decision trees, and Gaussian process regression. Considering the successful and widespread application of shallow and deep ANN models in wind speed and power prediction, in this work three different neural network core-based models are developed, addressing three major problems of time series prediction using machine learning methods: data quality assurance and invalid data imputation, prediction models hyperparameters assignment, and feature selection as input to the prediction models. In order to empower the ANN-core models, various mathematical techniques of

clustering, optimization, and signal processing are used to create new hybrid wind speed and power forecasting models. Then, the developed hybrid models are applied to several meteorological and wind energy datasets at different domains, providing wind speed and wind energy forecasts for very short-, short-, and medium-term (from minutes to hours ahead). Particularly, short-term wind energy prediction is usually applied to operational integration of wind power plants in electric grids.

New developed machine learning models performance (vs. observations) were compared to other prediction techniques, including WRF regional numerical weather prediction, with better ML statistics. As nowadays operational wind power plants usually apply either specific statistical methods or NWP for wind energy generation prediction, ML methods application can increase their predictions accuracy, providing a more stable electric grid operation; that is, less wind energy prediction derives in more conventional and flexible power sources use, mainly fossil fuels, in order to fill the expected gap between electricity supply and demand. And, any error in wind energy prediction can produce: (a) overuse of fossil fuels, with simultaneous losses of unconnected wind energy generation; (b) electric grid supply unavailability, which is a severe drawback.

As a consequence, this work represents an advance in meteorological and wind power data analysis, using different machine learning techniques (ANN) to better understand weather conditions during several case studies at different domains, with a strong basis in extensive review of neural networks application in wind speed and power prediction models, new models development, models adjustment, and models performance testing against both wind speed and wind energy operational observations.

First, a conducted comprehensive systematic review is considered as a solid basis for the three ANN models developed in this study. Particularly, the state-of-art of neural networks (NN) construction is investigated and compared, including networks structures, activation functions and ensemble or hybrid combinations of various mathematical techniques and prediction models. About 300 relevant articles published in high impact journals between 2010 and 2020 were gathered and filtered through a systematic process to reach a final list of around 100 articles which contain all the relevant pre-identified required information about NN construction, as input data characteristics, models-related parameters, and results evaluation criteria. In order to select the best construction options for the new developed models, several models results performance is discussed considering the different ANN construction components categories: architecture, activation function, learning process, and

feature selection. Models performances with various neural networks architectures, considering different number of layers and neurons per layer, are discussed and compared; results showed that the simplest architecture (1,2,1) provide more accurate results with less computational effort. About activation functions, a wide range (with hidden and output layers) is compared, observing strong influence in the prediction accuracy. Various backpropagation algorithms and heuristic optimization techniques are tested as ANN learning process, and their advantages and drawbacks are identified. Finally, various feature selection techniques for supervised learning applied in different previous studies were compared, including wrapper and filter-based methods; from this comparison, hybrid, combined, and ensemble methods have shown more successful compared to ML individual model, including individual ANN models.

This comprehensive review also covers models evaluation criteria, in order to allow the comparison of new developed models performance against other models performance in the literature. In addition, a new evaluation criterion, as the improved performance percentage (IP) against persistence model performance, is proposed. As persistence model results are always directly obtained from observational time series, this criterion can be applied at any observed domain, so it is possible to compare any model performance applied everywhere observed wind speed and/or wind energy datasets are available.

Following ML methodology, large time series datasets (Big Data) are necessary for ML models development and training, and afterwards, models performance measurement and intercomparison. Therefore, in this work six case studies are chosen, recording meteorological data time series in Spain, France and the United States, and wind power generation data in two wind farms located in Spain and France. About meteorological datasets, three MeteoGalicia stations in Galicia, Spain, are selected as representatives of different terrain characteristics: urban, highland rural, and coastal areas. In addition, M2-tower U.S. meteorological station is selected, as a specific site for wind speed monitoring at several heights. About wind energy datasets, Sotavento in Galicia and La Haute Bourne wind farms at Galicia and France, respectively, are selected as two different case studies.

In order to better understanding the properties of these time series, several analysis techniques are applied, as follows: Weather parameters from each meteorological station are visualized for specific time periods, using single and multi-parameter plots, and yearly wind speed time series from several meteorological stations are compared. Wind speed time series self-correlations and cross-correlations between various parameters, and also observations at different heights available at M2-tower meteorological station (U.S.), are

investigated. Typical wind roses are analysed, and also Weibull, Gamma, and Rayleigh wind speed distribution functions are obtained and compared between the different meteorological stations. Finally, different time series spectral analysis, using Fast Fourier Transform (FFT) and Welch methods, are conducted to reveal their strong periodicity or seasonality. These case studies datasets analysis has provided a general overview and better understanding of their characteristics, which is very useful in developing wind speed and power forecasting models.

Considering ANN models review, the quality of input dataset is an important step in model development and application, in order to improve its prediction accuracy. Therefore, in the available observation datasets, missing or invalid data must be completed (imputation algorithm) before using as model input. In this work, a new meteorological data validation-reconstruction model is developed and applied to complete available weather time series at different testing sites. First, available data quality is measured through a multi-level validation process to flag suspicious, missing and invalid values, that is, identifying data gaps. A wide range of validation tests, including basic and range tests, and internal, temporal and spatial consistency tests, were applied to the following weather parameters: wind speed, wind direction, temperature, precipitation, solar radiation, and relative humidity. Second, data gaps generated by invalid or missing labeled values were filled at each site using a reconstruction feed-forward neural network (FFNN) model with multi-dimensional k-means clustering technique, using as inputs: available observational data at the site, and Weather Research and Forecasting (WRF) model operational forecast output at the grid cell where the site is located.

This new reconstruction model was validated against several complete and validated observational time series. These time series were modified to produce new synthetic model inputs with different scattered removed values lengths: 5%, 10%, 20%, and 50%. Missing completely at random (MCAR) strategy was applied to create the completely-at-random synthetic gap patterns. Also, to simulate consecutive missing or invalid data in a time series, missing at random (MAR) strategy was used, that is, producing three different consecutive gap patterns of small (1-6h), medium (6-24h), and large (24-72h) data gaps in hourly time series. As part of the reconstruction model development and validation, with these synthetic observational time series three different FFNN models were build and trained using both WRF output and those synthetic time series as inputs. Imputation algorithm described above showed better results than the other two FFNN models, and also better than just gaps substitution by WRF output values. Also, it was observed that clustering WRF output as part of input data improves the reconstruction model performance. As a results summary, the

following final reconstruction model statistics MSE and RMSE ranges using the different gap ranges previously described are obtained: MSE (0.4, 0.6) and RMSE (0.6, 0.75) using the new developed reconstruction model, MSE (0.83, 0.88) and RMSE (0.91, 0.94) using ARIMA (1,1,1) model, and MSE (1.8, 2.3) and RMSE (1.3, 1.5) using direct WRF model output.

After all wind speed time series are reconstructed for the several case studies, a new wind speed prediction ML model, namely BBO-MLP, for very-short and short time horizon is developed. This hybrid prediction model is structured in a multi layer perceptron prediction core, using biological-based optimization to train the parameters. A new semi-exhaustive search algorithm is developed for selecting the best set of hyperparameters in three different solution spaces: parameters of model, optimization algorithm, and input data. Several hyperparameters sets, composing the global solution space, are defined for each of the three parts of this hybrid model: (a) for prediction core-model: input node numbers, hidden node number, and activation function; (b) for optimization: type of method, range of values, and number of parents and generation; and (c) for training dataset: length of time series, (train-validation)-test percentage, and selected features. In most of previous wind speed and power prediction ML models, hyperparameters are tuned separately for each part of the model, which neglects possible inter-linkages between them. In this work, simultaneous optimization of the three hyperparameter sets allows studying their mutual effects, as both optimal hyperparameter set and datasets are not necessarily independent of each other, therefore simultaneous optimization process is recommended to jointly reach out the best hyperparameter set including the different model parts. Actually, model results confirm this dependence between the three different sets. Both Bayesian and Akaike information criteria together with k-fold cross-validation are used for the model structure selection in each iterative phase of the proposed search algorithm. Simple, weighted, and exponential moving average techniques are applied to different combinations of previous time steps in order to find out the best input dataset to the prediction model development. In addition, highly unstable periods of yearly time series are selected as input datasets, as a more difficult challenge for the prediction model.

After the optimum model hyperparameters set is defined and the model is trained, the resultant prediction model is applied to the prediction of wind speed time series at four different meteorological stations, for short- and medium-term wind speed forecasting: Santiago-EOAS (urban environment), Muralla (highland rural environment), Camariñas (coastal environment) stations in Galicia, Spain, and M2 Tower in Colorado, USA (with its singularity that wind

data at different heights are available). The new developed model BBO-MLP prediction results are compared to the available wind speed observations at the four stations, using typical MSE and RMSE statistics. Compared to other prediction models, BBO-MLP outperforms for 1- to 3-hours ahead forecasting with one step ahead. However, very short-term (10 minutes) and multi-step ahead predictions obtained with BBO-MLP are not as successful as results from other prediction models. BBO-MLP MSE and RMSE results for 1 hour ahead prediction in M-2 tower meteorological station improves persistence model with 28.75% and 12.73%, ARIMA (1,1,1) with 73.91% and 48.60%, and MLP-BR model with 10.58% and 4.66%, respectively. In case of 3-hour ahead prediction, again for MSE and RMSE, persistence model results are improved by 47.50% and 31.86%, ARIMA (1,1,1) by 53.99% and 32.04%, and MLP-BR by 27.62% and 14.61%, respectively. In case of Galician meteorological stations, BBO MLP model 1-hour ahead prediction significantly outperforms WRF model outputs, with the following MSE and RMSE improvements: in Camariñas (coastal station), by 88.27% and 66.56%, respectively; in Muralla (high-land rural station), by 82.90% and 58.64%, respectively; and, in Santiago-EOAS (urban station), by 83.27% and 59.09%, respectively.

In the case of wind energy prediction, a different hybrid forecasting model is developed, namely EMD-QBPSO-ENN-GA, as this model is based on the use of both meteorological and wind generation data. Feature selection problem in neural networks is identified as a relevant issue to achieve good results, so a novel quantum-based binary particle swarm optimization (QBPSO) algorithm is developed and applied. This new algorithm is composed by an Elman neural network (ENN), a type of recursive network, that is optimized by a genetic algorithm (GA), and also equipped with empirical mode decomposition (EMD) technique to get the unsteady state features of the input data. This new hybrid EMD-QBPSO-ENN-GA model is applied over two different wind farms: Sotavento experimental wind farm in Galicia, Spain, and La Haute Borne wind farm in Billancourt, France. In these model applications three different input datasets are tested and used for wind energy prediction: (1) observed meteorological data, (2) wind generation data, and; (3) in the case of Sotavento, also meteorological data from the available WRF operational model output (MeteoGalicia) is considered. Previous to the new model application, several EMD-QBPSO-ENN-GA model setups using different input datasets were tested, using mean square error (RMSE), mean absolute error (MAE), and R-squared, as performance measurements. Particularly, eight different transfer functions for QBPSO algorithm, tests results allow to define the optimal transfer functions combination.

After the best new model structure was defined, EMD-QBPSO-ENN-GA model is applied to predict the two cases studies wind energy time series, both in the Sotavento wind farm at short to medium forecasting horizons (1 to 24 hours ahead), and in the La Haute Borne wind farm, at very short-term prediction horizon, 10 minutes. In Sotavento wind farm, selecting the best input dataset, wind energy RMSE obtained for 1, 3, 6, 12, and 24-hour ahead are 105.78, 163.98, 178.41, 215.99, and 414.74 (kWh), respectively; wind energy MAE for the same time horizons are 71.96, 115.39, 128.50, and 321.22 (kWh), respectively. These results show a significant improvement respect to persistence model, 61.85%, 50.01%, 41.62%, and 47.27% for the same time horizons, respectively. Particularly, considering the impact of different model parts in its performance, comparative results showed that using GA and EMD improved RMSE criteria for 1-hour ahead prediction with 4% and 21%, respectively. In La Haute Borne wind farm, wind energy RMSE and MAE obtained for 10-minute ahead prediction are 89.46 and 56.42 (kWh), respectively.

As a results summary, BBO-MLP wind speed prediction model shows its capability to obtain accurate short-term predictions (1-3 hours ahead), which are suitable to feed wind mills design equations to estimate their corresponding wind energy generation. Also EMD-QBPSO-ENN-GA wind energy prediction model shows its capability to obtain short-term (1-6 hours ahead) predictions over the tested wind farm. Additionally, both models can be run in short elapsed times over powerful PCs. Therefore, both solutions are suitable to be applied to short-term wind energy prediction, as a necessary technology to integrate wind farms generation in the current electric grids, in order to guarantee the electricity supply and to minimize other sources (fossil fuels) use.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

List of figures

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

List of tables

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Nomenclature

Acronyms / Abbreviations

ABL Atmospheric boundary layer

AI Artificial intelligence

ANFIS Adaptive neuro-fuzzy inference system

ANN Artificial neural network

ARIMA Autoregressive Integrated Moving Averag

COC Correlation of coefficient

DNN Deep neural network

EEMD Ensemble empirical mode decomposition

ELM Extreme Learning Machine

EMD Empirical mode decomposition

FEEMD Fast Ensemble empirical mode decomposition

FFNN Feed forward neural network

GA Genetic algorithm

HAWT Horizontal axis wind turbines

LR Linear regression

MAE Wavelet transform

MAPE Mean absolute percentage error

MC Markov Chain

MLP Multi-layer perceptron

MSE Mean square error

NARX Nonlinear autoregressive exogenous model

PCA Principal Component analysis

PSO Particle swarm optimization

RBF Radial based function

RMSE Root mean square error

SA Simulated annealing

SLP Single-layer perceptron

WD Wavelet Decomposition

WNN Wavelet Neural Network

WPD Wavelet Packet Decomposition

WPF Wind power forecasting

WS&PF Wind speed and power forecasting

WSF Wind speed forecasting

WT Wavelet transform

Chapter 1

Introduction

1.1 General overview

In 2015, the United Nations (UN) state members adopted the 2030 agenda for sustainable development, including 17 sustainable development goals (SDGs) considered an urgent call for action by countries. Among those, SDG7 is about energy aiming to ensure access to affordable, reliable, sustainable, and modern energy for all [1]. Also, based on the Paris Agreement signed in 2016 within the United Nation Framework Convention on Climate change, the countries are committed to moving towards a net-zero greenhouse gas emission by 2050 [2]. Together, these two targets are considered the main global goals in the energy sector, leading the world towards a more green and sustainable energy generation era.

Renewable energies have got much more attention in the energy sector during the past decade as several factors have pushed them. Some of these triggers are global warming due to the carbon dioxide (CO_2) emission from the combustion of fossil fuels, limiting greenhouse gas emissions due to the Kyoto protocol signed in 1997, energy security concern, aversion to the traditional fission nuclear power, and its unsuccessful progress in the application [3]. The growing population and the rising energy demand from the other side led to a higher renewable energy push. The ever-increasing rate of green power generation is faster than overall power demand. A new record was set in 2019 when for the first time, the increased rate of clean electricity generation overtook the increased rate of electricity demand while fossil-fuel electricity generation decreased [4]. This was the first time in decades that electricity demand increased, fossil-fuel electricity generation lessened.

Figure 1.1 shows a global map of the global primary energy share from renewable sources in 2019, including solar, wind, hydropower, geothermal, bioenergy, wave, and tidal. Inter-

preting from the figure, the renewable energies distribution around the world is not uniform. This non-uniform distribution has different origins, such as policies of the countries, access to technology, economic situation, access rate to different types of renewable energies, etc. The figure also shows the importance of global agreements, action plans, and global capacity development workshops in this sector to leave no one behind in just energy transition and global access.

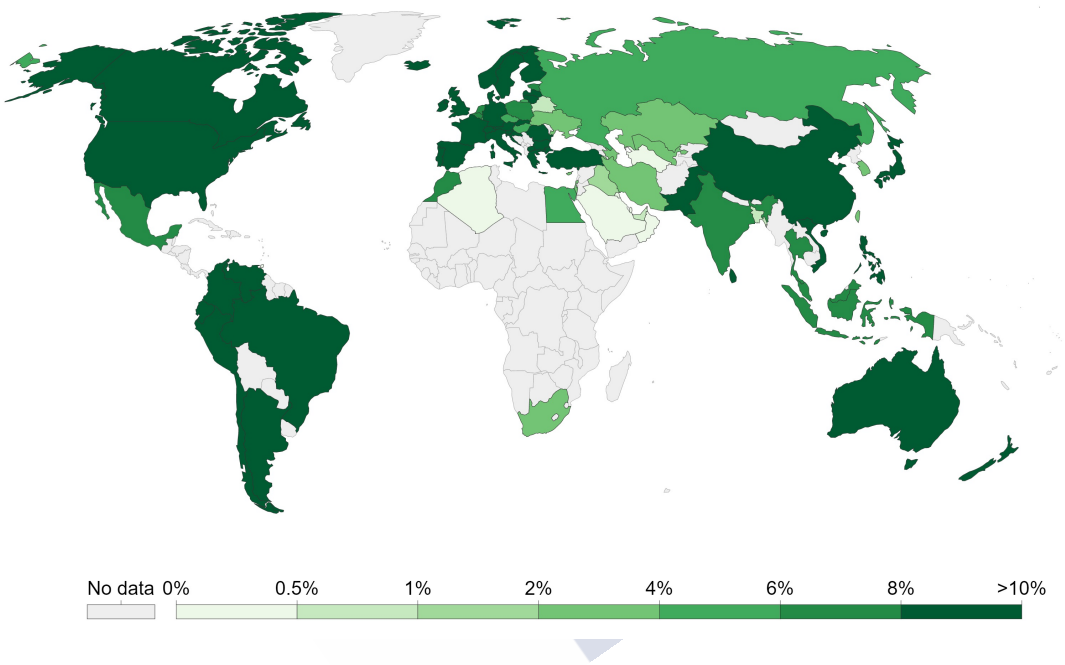


Fig. 1.1 Global share of primary energy from renewable sources, 2019 [5]

Among different renewable energy sources, wind and solar play a most active role during the next two decades. It refers to the global energy transformation roadmap 2050 published by the international renewable energy agency (IRENA) [6]. Figure 1.2 shows that the worldwide share of non-renewable sources (coal, oil, gas, and nuclear) in electricity generation will decrease significantly from 76% in 2015 to 15% in 2050. The figure also shows that wind energy will have a share of 36% of the renewable energies by 2050, making wind the first source of global energy. Another critical shift in traditional energy sources utilization is the significant decreased in coal's share of the global energy basket from 39% in 2015 to almost 0 in 2050. Oil will also lose its importance by shifting from 4% to 1%. However, natural gas keeps its share experience a limited loss of 8% from its share in 2015, changing from 16% in 2015 to 12% in 2050.

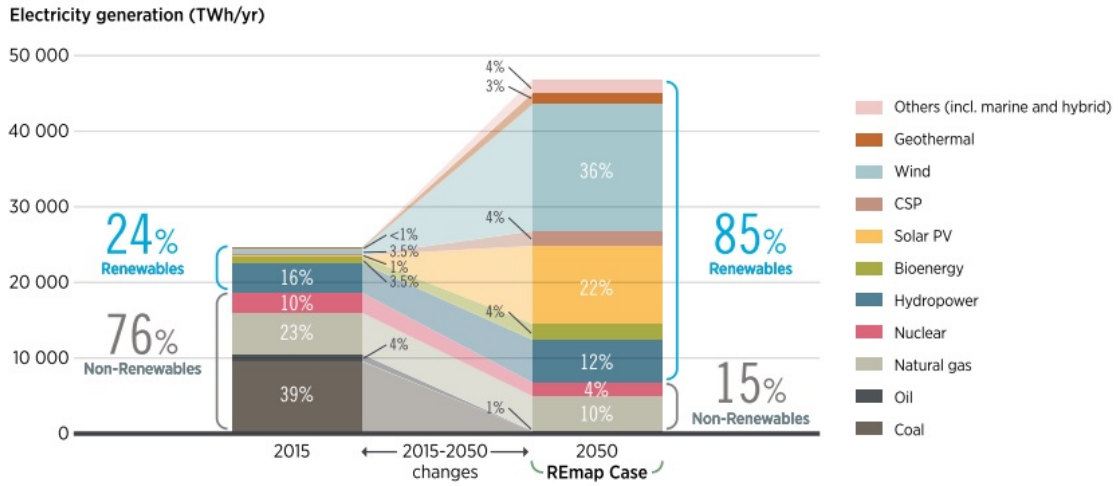


Fig. 1.2 Global renewable and non-renewable energy shares in electricity generation from 2015 to 2050 [6]

Investigating the ongoing global situation of wind energy, China is the world's most enormous producer, followed by Germany and United States, as shown in figure 1.3. As it is shown, ten first pioneer countries in wind energy installation and utilization are listed. Spain is the world's fifth-biggest and European Union's second-biggest producer with 25.808 MW installed capacity by the end of 2019. Also, Spain is the first country in Europe that has established the most onshore wind turbines throughout 2019 with an approximate 2.4 MW capacity [7].

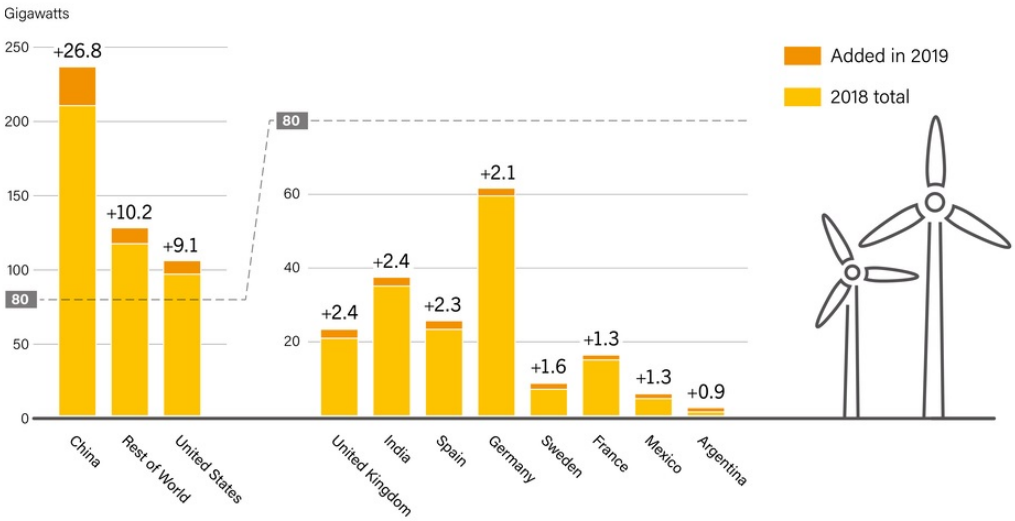


Fig. 1.3 Wind power capacity installations, Worldwide top 10 countries in 2019 [8]

1.2 Introduction and motivation

The scarcity of fossil fuels, the climate change environmental concerns, and the 2°C target of the Paris Agreement [2] require rapid decarbonization of the energy sector, contributing to the development and use of renewable energies in the world. [9]. Wind and solar power are the most promising and worldwide used sources because of their global availability cite. These types of renewable energies can constitute a keystone component for micro-grids towards intelligent infrastructures. Each MWh of generated wind energy prevents the emission of at least 500kg of greenhouse gasses [10]. Integration of renewable energies into the electricity grid comes with a significant challenge due to volatile and inconsistent wind speed in different weather conditions. To deal with these uncertainties, several technical measures such as virtual inertia, batteries, or intelligent grid applications can balance the grid in a time horizon of seconds to minutes. However, for a longer time horizon from minutes to hours, pumped hydro storage can provide the backup power [9]. Accurate wind speed and power prediction is another powerful tool to deal with this problem in advance.

In order to decrease this uncertainty of generated wind power estimation, numerous forecasting models can be found in the literature for different time horizons and time-steps ahead prediction. Dealing with a wide range of wind speed and power prediction problems, these models are comprised of physical methods, which application is mostly on global and regional long-term forecasting, traditional statistical models which capture seasonality and patterns in time series, and new promising techniques of artificial intelligence and machine learning (ML) that have been mushrooming overtaking the other two groups, especially for shorter time horizons [11]. In order to increase prediction models' accuracy, data-related and model-related factors are two main groups of factors that need to be considered. The quality of input data, or the chosen set of input data, known as feature selection problem, is as important as the model-related factors such as model's hyperparameters assignment, training method selection, and combined or hybrid structure definition. Important to mention that feature selection, known as the dimensionality reduction problem in a broader concept, is a transversal issue between data and models. Various types of ML methods have been applied successfully so far, addressing the above-mentioned issues with the final goal of prediction accuracy increasing, including ANNs shallow and deep models [11–13], SVR [14, 15], SVM [16, 17], fuzzy [18, 19], decision trees[20], random forest [20], Bayesian [21, 22], and Markov models [23, 24].

Despite the success of these models, a few essential facts need to be considered; First, as we are dealing with uncertainties in the complex atmospheric system, it is impossible to reach zero error in forecasting models. So there is always space and justification for developing new forecasting models. Second, most of the works in literature have focused on the mathematical models side instead of actively evolving meteorology science into the prediction models. In other words, wind or other weather parameters are treated simply as a non-stationary signal for an ML prediction model. However, no need to mention that including more weather and climate information into the models could result in more accurate results. Third, most of the proposed models in the literature, even the combined and hybrid ones, are focused on combining different ML models [25], and there is a lack of combination between numerical weather predictions (NWP) and ML models.

Focusing more on the modeling and data quality side of forecasting models, this work tries to develop some ANN models addressing data reconstruction, hyperparameters assignment, and feature selection problems. These down-scaled wind speed and power prediction models can be used in wind power plants, providing a more accurate estimating of the near-future wind power generation than widely used NWP models.

1.3 Objectives of this work

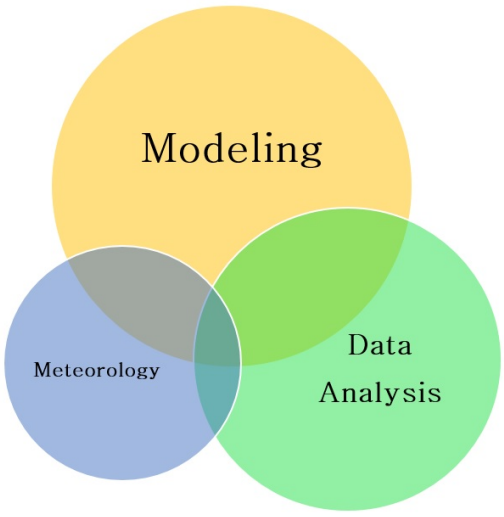


Fig. 1.4 Contribution of different fields in this study: modeling, data analysis, and meteorology

The primary purpose of this work is to develop a cohesive and technical process for short to medium-term wind speed and power prediction models, following a process in which observational and numerical weather prediction (NWP) data will go through validation, reconstruction, and artificial neural network (ANN) wind speed and power prediction procedures by addressing machine learning critical issues, including core model structure, hyperparameters optimization, and feature selection. In the direction of the main purpose, to develop this work a combination of modeling, data analysis, and meteorology information is used. This concept is shown in figure 1.4. The size of each field roughly represents its contribution to the current work.

To achieve the primary purposes of this work, we are looking for the answers to the following critical questions are developed:

1. **From Meteorological point of view**, How is the relationship of observational data with each other? Also, the relationship of the parameters between different height levels of the boundary layer? (Chapter 3)

2. How to design a weather stability condition indicator to classify and compare forecasting models based on their atmospheric condition? (Chapter 4)

3. **From the data point of view**, how important is improving the quality of input data before feeding them into a prediction model? (Chapter 5)

4. Can numerical weather prediction (NWP) data improve the accuracy of wind forecasting and data reconstruction models? (Chapter 5)

5. How to address the feature selection problem differently in a wind speed and power prediction model? (Chapter 7)

6. **From the modeling point of view**, Which models in the literature have been more successful and what are the main challenges with wind speed and power neural network models? (Chapter 4)

7. How to address differently hyperparameters assignment problem in a wind speed prediction models? (Chapter 6)

8. How effective is it to apply data clustering techniques for improving the accuracy of a wind speed prediction model? (Chapters 5, and 7)

9. How successful is a hybrid neural network wind power prediction model compare to a single model? (Chapter 7)

Regarding the meteorological point of view, the wind has some specific characteristics that distinguish it from other weather parameters. Temperature and pressure differences, air density, topography, and other factors involved in wind origin, nominates the wind speed as one of the most complex meteorological parameters to predict [18]. Numerous works in the literature treat wind speed and power as a regular time series prediction problem. Still, it is proved that including meteorological characteristics of wind into the prediction model makes notable accuracy improvements [26]. The relation between different heights of wind speed is reported to be meaningful [27]. It is tried to include the above-mentioned proved results into the current study to provide higher accuracy models.

Related to the data analysis sector, two types of data can be used in a wind speed and power statistical or ML prediction model. First, observational data acquired from meteorological stations and environmental sensors, and second, simulated data as the output of numerical weather prediction (NWP) models. Contrary to the simulated data, the observational type needs a quality check process. Observational data can be corrupted in different stages of acquiring, transmission and storage. No matter the origin of erroneous data, this part of the information should be cleaned and replaced before being used in the model. Different imputation techniques are investigated in the literature to complete invalid data [28, 29]. However, fewer have used machine learning techniques to address the problem. A validation-reconstruction algorithm is provided in this work to deal with this critical issue.

Regarding the modeling sector, choosing an optimal structure for the neural network is a tedious task. Addressing this issue, most of the available research has followed a test and trial procedure to determine the best structure, which does not necessarily guarantee the optimal solution. In the light of an extensive review of the applied neural networks for wind speed and power forecasting (WS&PF), a semi-exhaustive search algorithm was developed to optimize the model's hyperparameters in conjunction with other optimization-related and input-related sets of hyperparameters. Besides, different structures and activation functions of neural networks are used in a hybrid platform with other optimization and signal processing techniques. Two main challenging topics of neural network models, hyperparameters assignment and feature selection, are addressed by providing novel algorithms as part of the developed hybrid ANN models. These methods were usually reported as continuously successful in the literature [30–32].

1.4 Structure of this work

This manuscript is structured in eight chapters. In Chapter 2 a general overview of wind speed and power meteorology is presented, from wind speed origin and its characteristics in the boundary layer to local and global effects on wind speed, wind turbulence, and wind shear. The history of wind turbines and information about the modern wind turbines, structure, and power generation, are provided right after. Different wind speed and power prediction models, including physical, statistical, and data-driven techniques, are introduced, followed by explaining the various types of machine learning models. In Chapter 3, different case studies used in this work are introduced, followed by applying data analysis methods to better understanding the characteristics of the applied data better. In Chapter 4 a comprehensive review of all types of artificial neural networks applied in wind speed and power prediction models are proposed. In this chapter, the weather stability index is also introduced as an atmospheric stability class indicator for models under which they are developed. After that, Chapter 5 proposes a validation and reconstruction procedure for missing, wrong and invalid data for wind speed and other weather parameters using numerical weather prediction outputs, in particular WRF model. This step is considered as a need before inputting data to prediction models. Subsequently, Chapter 6 deals with hyperparameters assignment problem in a proposed hybrid wind speed forecasting model by providing a new problem formulation. A semi-exhaustive search algorithm is applied to achieve optimal solutions in three different hyperparameters sets of the model, optimization, and data simultaneously. In Chapter 7, the feature selection problem is addressed, developing a hybrid wind power prediction model. A novel quantum-based binary particle swarm optimization technique is combined with empirical mode decomposition signal analysis and recurrent Elman neural network. Finally, Chapter 8 provides a conclusion of the current study followed by future works recommendations.

Chapter 2

An overview on wind speed and power meteorology and prediction

2.1 Introduction

This chapter provides a comprehensive view of wind speed meteorology, wind power generation, and different categories of wind speed and power prediction models. The wind speed meteorology section explains how the wind is generated in origin under a global-scale view and how atmospheric and geographical local effects can change its characteristics. In the wind power section, a history of wind turbines is provided, followed by a discussion on modern wind turbine characteristics and the wind power output. Also, the environmental and human effects of wind turbines are discussed briefly. The last section talks about different wind speed and power prediction models, from physical to statistical, to machine learning models.

2.2 Meteorology of wind speed

2.2.1 The origin of wind

Sun is the leading cause of global wind generation by heating the earth unevenly and creating hot and cold spots. The thermal gradient of the air created by the sun heats forms a low and high pressures system in the atmosphere. Moving air from high- to low-pressure areas is generally known as wind. In other words, the wind is caused by differences in atmospheric pressure. The other major reason for global-wind patterns is the earth's rotation

which causes Coriolis force. In theory, the exact balance between pressure gradient force and Coriolis force, called geostrophic equilibrium, results in the geostrophic wind parallel to isobars, the lines of constant pressure at a specific height. However, in reality, the true wind is different from geostrophic wind because of surface friction forces. A general schematic overview of the prevailing global wind is shown in figure 2.1.

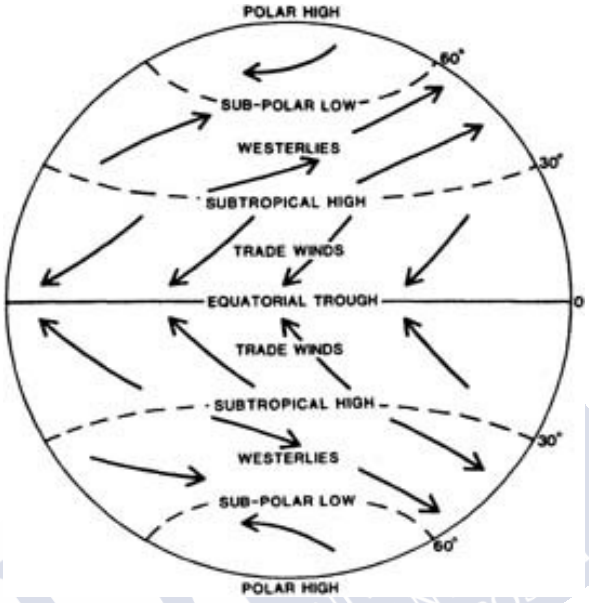


Fig. 2.1 Prevailing global winds [33]

Above the Atmosphere boundary layer (ABL) is the "free atmosphere," where the wind is approximately geostrophic, while within the ABL, the wind is affected by surface drag and turns across the isobars. In meteorology, the boundary layer is the closest air layer to the ground affected by diurnal heat, moisture, or momentum transfer to or from the surface. Most changes in boundary layer depth over oceans are caused by synoptic and mesoscale processes of vertical motion and advection of different air masses over the sea surface. ABL is where the meteorology of wind power occurs and has a depth of 100 meters at night with low wind and up to 2, or 3 km at daytime with a strong solar radiance. ABL can be principally divided into three layers: The lowest layer with only a few millimeters deep above the ground, called laminar. The second layer, known as the surface layer or flux layer, is located up to 100 meters above the ground, where turbulent viscosity forces dominate and wind speed increases with height. The third layer is called the Ekman layer, which is located on top of the other two layers and usually covers up to 90% of the boundary layer [26]. In general, the boundary layer is thicker in low pressure and thinner in high-pressure systems, as shown in the figure

2.2.

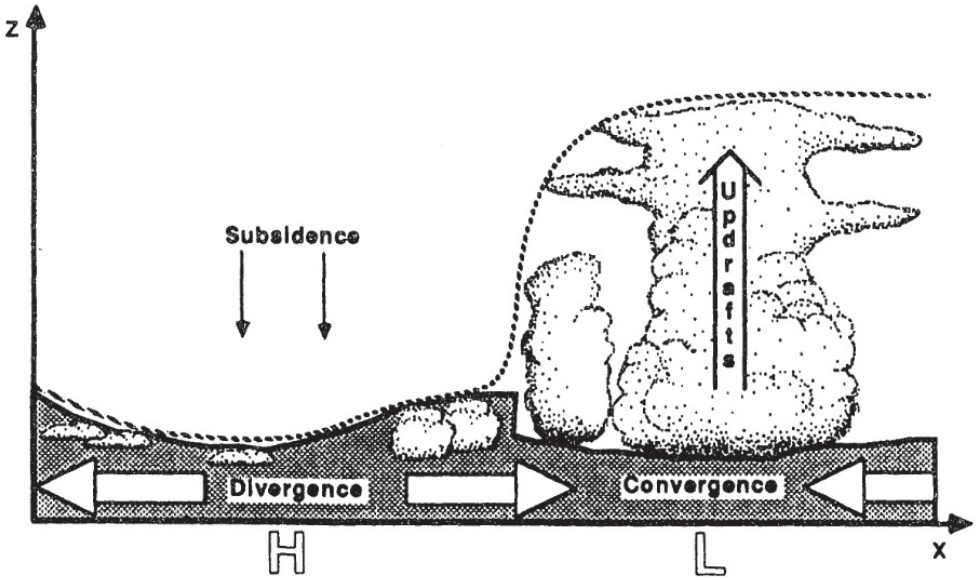


Fig. 2.2 Schematic illustration of boundary layer depth in high (H) and low (L) pressure. Dotted line is the boundary of maximum height reached by surface modified air during one-hour period. Solid line is most studied by meteorologists [34]

2.2.2 Wind components

Airflow can be divided into two general components: **mean wind**, and **turbulence**. In time series decomposition, however, a signal decomposes into three main components: trend, seasonal, and residuals which make it easier to detect trends, cycles, or treat with non-stationary residuals. The mean wind is responsible for rapid horizontal transport of quantities such as moisture, heat, and pollutants or advection in the boundary layer. Vertical wind means from the other side are much smaller, and wind turbulence takes the responsibility of vertical movement.

2.2.2.1 Turbulence

From a theoretical perspective, turbulence comprises many different sizes of irregular swirls of motion called eddies; a gustiness superimposed on the mean wind. The turbulence spectrum shows the relative strengths of these different scale eddies. As the primary origin, turbulence is generated by forces upcoming from the ground. Solar heating on sunny days

causes thermals of warmer air, which are large eddies, to rise. Frictional drag on the air flowing is another source of turbulence, which causes wind shears to develop. Obstacles like streets and buildings, which deflect the flow adjacent to and downward of the obstacle, are also the cause of forming turbulence.

Even though there is uncertainty in mean wind and turbulence, dealing with uncertainties in the latter is more complex and challenging than the first one. Turbulence has a duration of 10 seconds to 10 minutes. In order to characterize it, turbulence standard deviation and intensity are typically utilized, which latter one obtained by dividing the ratio of the standard wind velocity's fluctuations to the mean speed. The widely-used method to investigate turbulence is Reynolds-average [35]. It is found in many research that turbulence intensity is related to the mean wind speed, following relations summarized in table 2.1 in which σ is mean wind speed's standard deviation, v is mean wind speed, I_{15} is turbulence intensity at 15 m/s wind speed, and α, β, c are fitting parameters. More detailed information about the mean wind speed- turbulence intensity relationship can be found in [36].

Table 2.1 Mean wind speed and Turbulence intensity relationship

item	Mathematical equations
1	$I = \frac{\sigma}{v}$
2	$I = I_{15}(\alpha + 15 \times v^{-1})/(\alpha + 1)$
3	$I_{IEC} = I_{ref}(0.75 + 0.56 \times v^{-1})$
4	$I = \alpha \times v^{\beta} + c$

2.2.2.2 Wind shears and wind gusts

Wind shear or gradient is the difference of wind speed or direction in a relatively short vertical or horizontal distance in the atmosphere. Although wind shear is a local meteorological phenomenon, it can also occur in mesoscale or synoptic scales such as squall lines and cold fronts. As mathematically formulated later in this chapter, 2.2, wind shear coefficient is the same α parameter in the power low.

However, wind gust is a tiny increase in horizontal wind speed during a more extended sample period that usually lasts for less than 20 seconds. In general, gustiness is a dominant wind characteristic over lands, particularly near large buildings than over large water surfaces. Mathematically, it can be defined as the maximum of the moving average with a moving average window length equal to the gust duration t_g. In origin, the characteristics of old

anemometers with response times about 2 to 5 seconds have been the origin of wind gust duration [37].

2.2.3 Height effect on wind speed

An important phenomenon of wind speed concerning wind power is the increase of wind speed with height. Because of surface friction, wind speed value changes from zero at the ground level to tens of meters per second in the higher heights. Depending on the time of day and atmospheric condition, the range up to where the wind speed is undisturbed is between 600 and 2000 m above ground located inside the boundary layer. The conventional logarithmic height formula is proposed as equation 2.1, under a statistical mean of a steady-state wind distribution. It should be noticed that this assumption is logical for long-term statistical wind speed, usually applied in the calculation of the energy delivery by a wind turbine.

$$V_H = V_{ref} \cdot \frac{ln\frac{H}{z_0}}{ln\frac{H_{ref}}{z_0}} = \frac{u_*}{\kappa} ln(\frac{H - H_{ref}}{z_0}) \tag{2.1}$$

Where V_{ref} is wind speed at reference height (m/s), H stands for the desired height (m), z_0 is roughness length of the earth surface ,H_{ref} is reference height (m), ln stands for natural logarithm, u_* is friction velocity. The parameter κ, known as von Kármán constant, is a dimensionless value describing the distribution of longitudinal velocity in a wall-normal direction of a turbulent fluid flow near a no-slip condition boundary.

To calculate wind speed at the desired height, V_H, the roughness length of the surface is required. Roughness length, Z_0, does not represent any physical length but is considered as a length-scale representation of the surfaces' roughness. Indeed, it is a parameter of some vertical wind profile equations that model the horizontal mean wind speed near the ground. In the log wind profile, roughness length is equivalent to the height at which the wind speed theoretically becomes zero under natural conditions and in the absence of wind-slowing obstacles, as shown in figure 2.3. In this figure, wind speed is obtained from equation 2.1 using its simplest way. Table 2.2 shows the roughness lengths for different terrains [38].

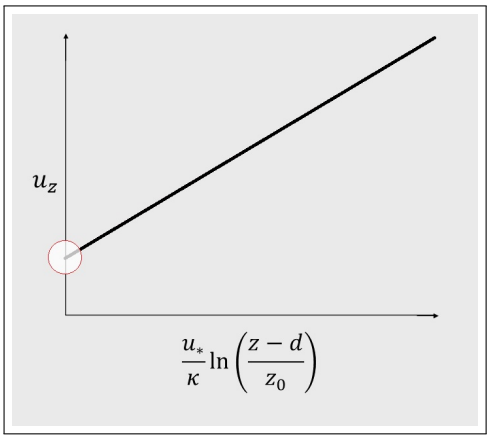


Fig. 2.3 Log wind profile

Table 2.2 Classes index and description of terrain roughness length (z_0)

Class index	**Terrain description**	$Z_0(m)$
1	Open sea, fetch at least 5 km	0.0002
2	Mud flats, snow, no obstacles and vegetation	0.005
3	Open flat terrain, grass, few isolated obstacles	0.03
4	Low crops, occasional large obstacles	0.10
5	High crops, scattered obstacles	0.25
6	Parkland, bushes, numerous obstacles	0.50
7	Regular large obstacle coverage (suburb, forest)	1.0
8	City center with various tall buildings	≤ 2

A standard simple description of wind speed correlation with altitude is the power law approximation proposed by Hellman, as formulated in equation 2.2 [39].

$$V_H = V_{ref} \cdot (\frac{H}{H_{ref}})^{\alpha} \tag{2.2}$$

Where V_H (m/s) is the mean wind speed at elevation H, V_{ref} (m/s) is the mean wind speed at the reference elevation, H (m) stands for the desired height, H_{ref} (m) is the reference elevation, and α stands for the Hellman's exponent or wind shear coefficient. It is essential

to mention that the calculated values are often underestimated for greater hub heights in both equations 2.1, and 2.2. This fact is even more critical on inland sites where the roughness lengths have larger values [40].

2.2.4 Air density and thermal stratification

The atmosphere's kinetic energy, equivalently the wind power generation, depends on the wind speed and air density, referring to the equations 2.9, and 2.10, which are presented later in this chapter. Near-surface air density itself is a function of air pressure P and temperature T as described below, in the equation 2.3. This formulation is equivalent to the Hydrostatic equation, 2.4, which is the primary explanation of the wind with the origin of horizontal heat gradient. The origin of the horizontal temperature gradient is the upward heat transportation from the surface originate by the sun heats the Earth's surface differently according to surface properties and latitude, equation 2.5. This equation represents a neutral atmosphere condition by an adiabatic vertical temperature [26]. Hence, the air density is different in such a way that a cooler air mass is denser and has a smaller vertical distance between two given pressure surfaces than a warmer air mass. The air pressure and density are closely related, shown in the following equation as an equivalent for the air density equation.

$$\rho = \frac{P}{RT} \tag{2.3}$$

$$\frac{\delta P}{\delta z} = -g\rho = -\frac{gP}{RT} \tag{2.4}$$

$$\frac{\delta T}{\delta z} = -\frac{g}{c_p} \tag{2.5}$$

Where ρ is the air density, z represents vertical coordinate, g is the earth's gravity, P stands for the air pressure, T is temperature, R stands for the universal gas content with a value of 287 $J\,kg^{-1}K^{-1}$, c_p is the specific heat of the air at constant pressure, which is equal to 1005 $J\,kg^{-1}K^{-1}$.

The vertical temperature gradient is the origin of different atmospheric stability conditions: neutral, stable, and unstable. An unstable stratified atmosphere usually forms when the cooler air flows over warmer surfaces connected to an upward turbulent heat flux to the

atmosphere. When a downward turbulent heat flux from the atmosphere towards the surface happens, we usually have a stably stratified atmosphere. Wind speed from the other side is totally in correlation to the atmospheric conditions in such a way that low wind speeds are mostly correlated to stable atmospheric conditions, whereas unstable atmospheric conditions favor higher wind speeds away from the surface. In a nutshell, for small wind speeds, the impact of thermal stratification is more significant, and for higher wind speeds gets smaller. However, stability becomes a more significant issue for deeper higher atmospheric layers, particularly for more giant and higher hub wind turbines.

2.2.5 Richardson number

The Richardson number is a non-dimensional measure, explaining the atmospheric stability, as shown in equation 2.6. In thermal convection problems, Richardson number represents the importance of natural convection relative to the forced convection. Typically, the forced convection is dominant when $Ri < 0.1$, natural convection is dominant when $Ri > 10$, and neither is negligible when $0.1 < Ri < 10$. It is important to mention that the forced convection is usually large relative to natural convection except in the case of extremely low forced flow velocities. [34]

$$Ri = \frac{g\frac{\delta\Theta_v}{\delta z}}{\Theta_v\left[\frac{\delta u}{\delta z}\right]^2} \tag{2.6}$$

Where Θ_v stands for virtual potential temperature, contains the effect of humidity into the potential temperature, u is the wind component in the mean wind direction, z represents vertical coordinate, and g is the Earth's gravity.

2.2.6 Local effects on wind speed

Contrary to roughness lengths that determine the local wind characteristics in a larger area, obstacles' effect is more limited, but sill important. Buildings in rural areas, trees in forests, mountains, hills, and valleys have the potential to create turbulence. Wind speed fluctuation can heavily impose adverse effects on the operation and lifetime of wind turbines. If obstacles such as buildings are the cause of flow wakes, as shown in figure 2.4, the turbulent airflow increases to twice the obstacle's height and extends to a distance to twenty times of

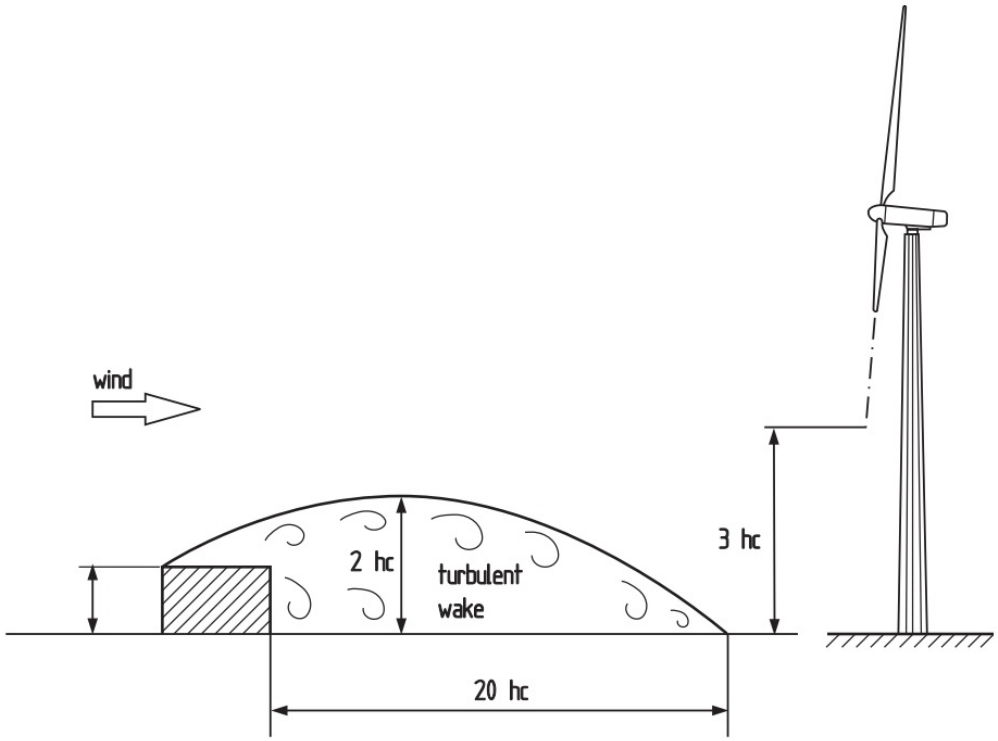


Fig. 2.4 Flow wake - Obstacle effect on wind turbulence [40]

the obstacles' height to the direction of the down-side. In order to prevent harmful effects on the wind turbine, the height of the rotor should be three times longer than the obstacle's height and sufficiently far away down-wind [40].

As mentioned before, the cause of local wind systems is the local or regional thermal properties of the Earth's surface. These local wind systems' origin is different from the large-scale pressure systems considered the primary origin of winds. Complex terrain like mountains is another reason for thermally driven secondary circulation systems where there is a weak large-scale pressure gradient which generally happens when there is no cloud in the sky. In this situation, three local and regional wind systems are induced, as shown in figure 2.5, which change the vertical structure of the mountainous atmosphere boundary layer. Slope winds (thin arrows in the figure) develop from a few meters up to 1 km on a slope spacial. Mountain and valley winds (big full arrows) emerge on a spatial scale from hundred meters up to a hundred kilometers in long valleys. Mountain-plan winds (open arrows) are the most larger scale winds from tens to one hundred kilometers. The latter two types have the potential for wind energy generation.

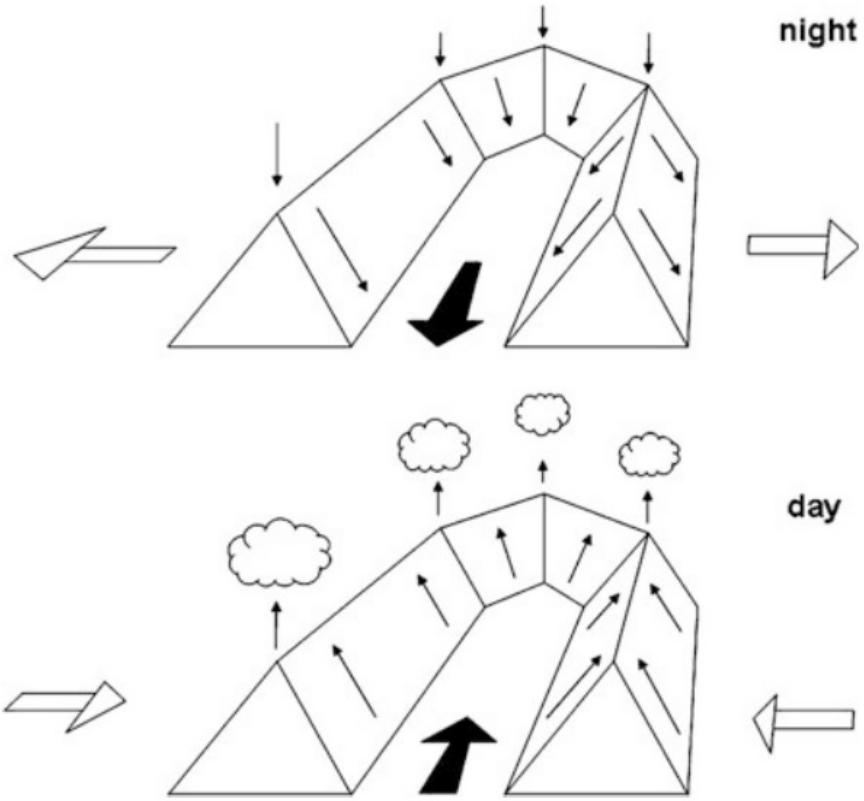


Fig. 2.5 Mountainous origin of local and regional wind systems at night and daytime. Description about three types of arrays is given in the paragraph above [26]

It is not easy to understand and model winds with accuracy over complex terrain as wind shows significant spatial and temporal variations. Non-linear numerical flow models provide better results to model flow separation than linear models [41] .

2.2.6.1 Wind wake and park effect

Another similar problem in wind farms caused by the first row turbines to the wind as obstacles are wind wake. The upstream wind turbines swirl the air flows and subsequently cause wake areas in the downstream regions. This phenomenon can cause extreme power losses and effects the total power potential of a wind farm. Hence, it is essential to either use computational fluid dynamics (CFD) or analytical models which consider wake to design turbine blades and calculate their installation distance from each other in the wind farm so as to prevent such an energy generation loss [42].

Considering equations 2.7, and 2.8, and figure 2.6, there is an overall mass-specific momentum consumption, m, of the turbines. This loss in the wind parks can only be compensated by a turbulent momentum flux s from above. Here, K_m is a coefficient of momentum or turbulence exchange that describes the atmosphere's ability to transfer momentum vertically by turbulent motion. u_0 is the unchanged wind speed above the wind park.

$$m = c_t {u_h}^2 \tag{2.7}$$

$$\frac{\tau}{\rho} = K_m \frac{u_0 - u_h}{\Delta Z} \tag{2.8}$$

Where c_t is turbine drag, u_h is the wind speed at hub height h, K_m is viscosity describes an atmosphere conductivity giving mass-specific momentum flux and has a value between 1 and 100 m^2/s. By rewriting the exchange coefficient as a function of other outer and inter conditions in wind park such as surface roughness, thermal stratification, turbulence generation of the turbines, and drag of turbines, the mean wind speed can be calculated [26].

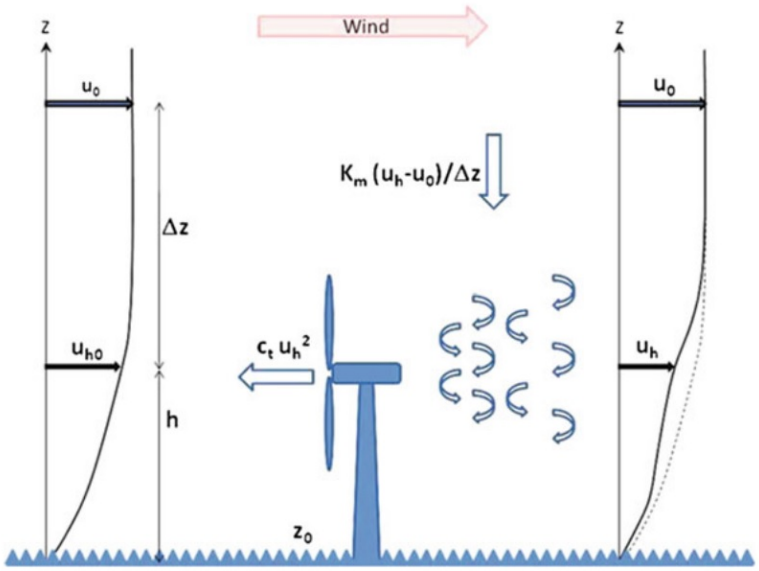


Fig. 2.6 Momentum loss in wind parks and replenishment [26]

2.3 Wind power generation

Wind energy or wind power refers to generating electricity by wind passing through the blades of a turbine. Onshore wind energy refers to wind turbines established on the lands, no matter as utility-scale that ranges in size of 100 kilowatts (kW) to several megawatts (MW), or single small wind turbines below 10 kW. The other main category of wind energy is offshore wind, which refers to the turbines erected in the open waters, especially near the coasts. Offshore winds are larger than onshore turbines, which in turn, generate more power. In the following, we mainly focus on onshore winds; however, most of the discussed topics and models are applicable to the offshore wind adjusting the forecasting tools under the wind characteristics of the offshore areas.

2.3.1 Historical development of wind turbines

The history of using wind energy goes back to $5,000$ BC to propel boats along the Nile River. The earliest recorded practical windmill design is of Persian origin and was invented around 700–900 AD. This design was the Panemone, consisting of a wall with slits surrounding a vertical axle containing four to eight fabric sails. This vertical-axis wind machine was first built to pump water and subsequently modified to grind grain [43]. These particular windmills, Asbad, are located in Nashtifan, Razavi Khorasan Province, Iran. Made from clay, straw, and wood, the windmills or Asbad, as it is pronounced in Persian, stand up to 20 meters in height with grinding stones in a room below for pressing grain into flour, as shown in figure 2.7 [44].

Later in the 12th century, the first horizontal-axis windmills were designed and used in northwestern Europe, most probably under the influence of vertical-axis design spread from the middle east and central Asia [43]. Later, post-, tower-, and stock-mills were designed and operated in different parts of Europe. Until the end of the 19th century, all different windmills were used, grinding grain, pumping water, and cutting wood at sawmills.

In the late 1800s and early 1900s, using wind power for electrical generation started to develop interest. The main reason was an enormous expense of electricity distribution and transmission from the central station. In 1887 and 1888, the two very first horizontal and vertical axis wind turbines were built, one in Scotland and the other in the United States. In 1887, James Blyth built the very first vertical-axis wind power generator, Blyth's windmill, in the back garden of his holiday home in Aberdeenshire, Scotland. He utilized the technologies

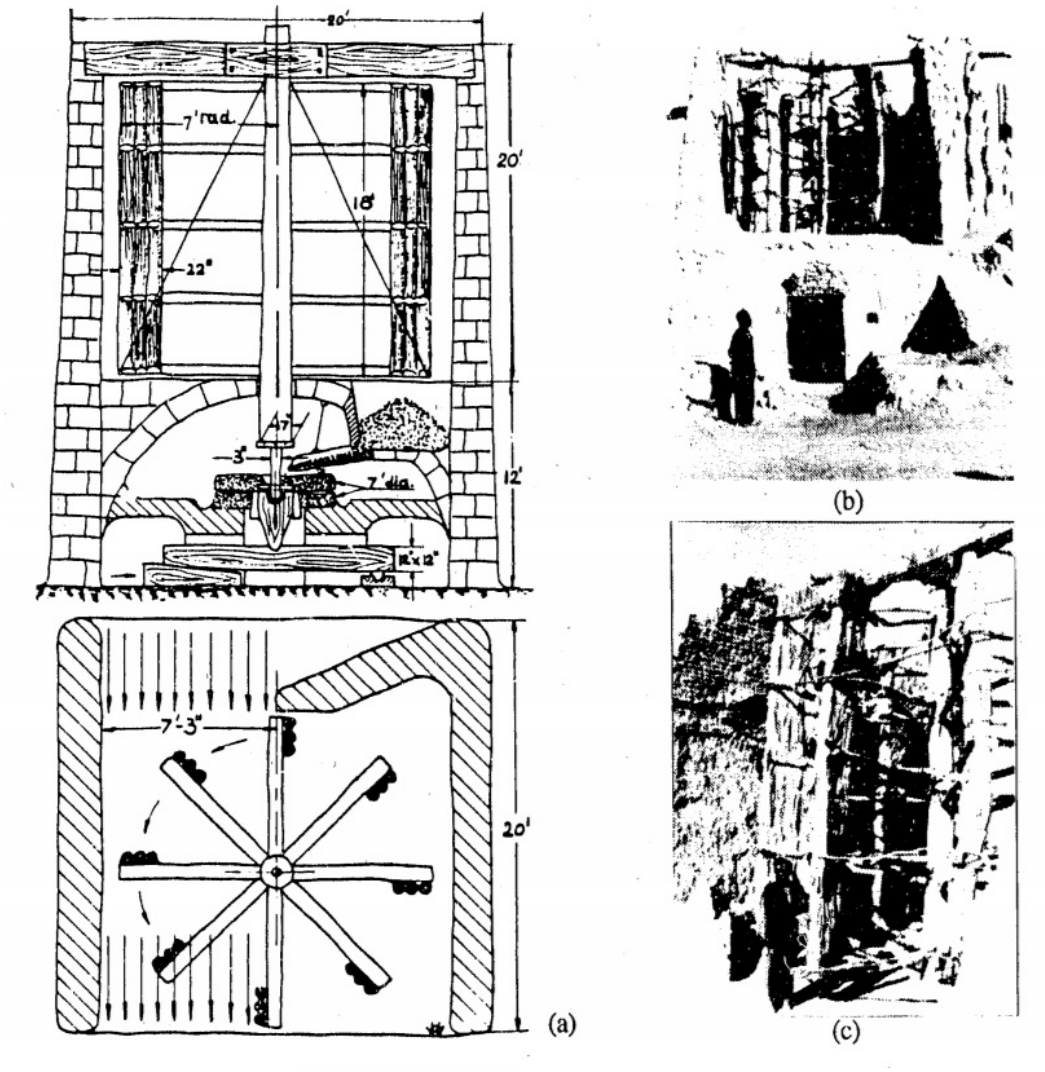
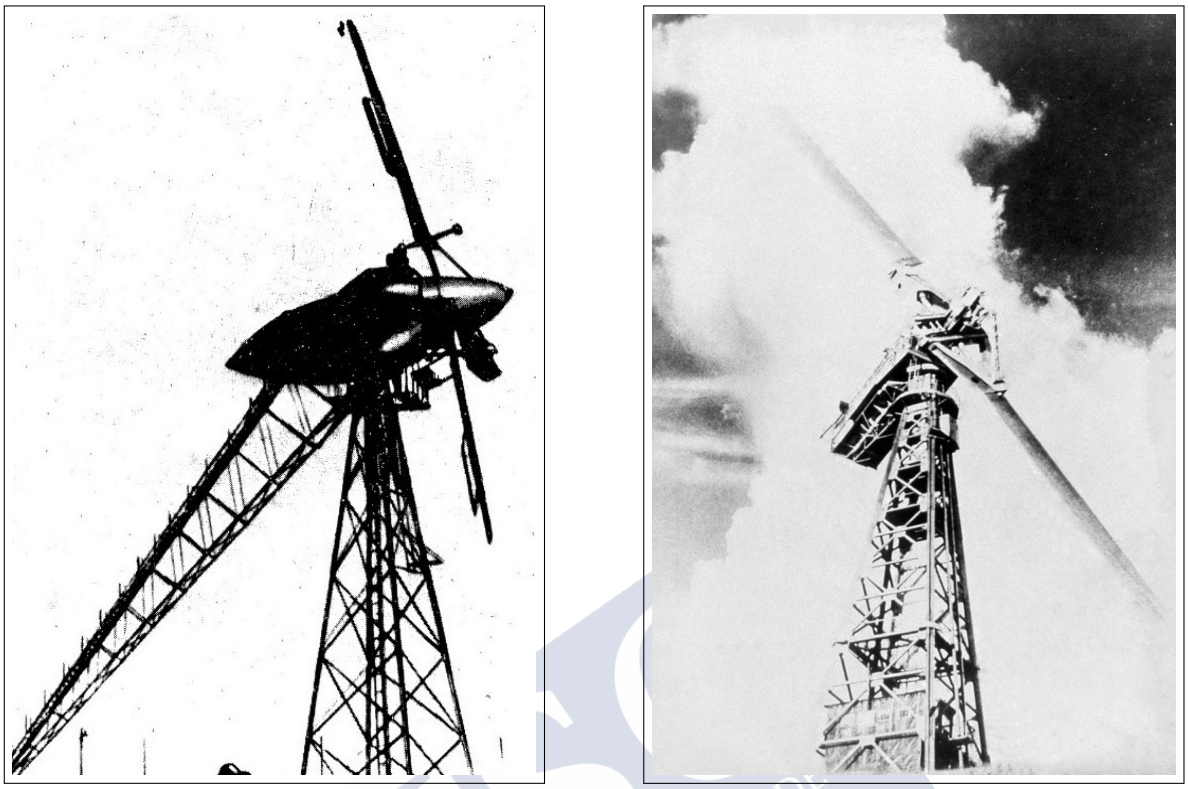


Fig. 2.7 A Persian vertical axis windmill, a) the millstones are below the rotor and the sails are bundles of reeds. b and c) a general view of the windmill [44]

of Robinson anemometer, the dynamo generator, and the Faure accumulator to build his 10 m high, cloth-sailed wind turbine to build his windmill in order to supply electricity for his home [45]. A year later, in 1888, Charles F. Brush invented the very first horizontal-axis small wind-electric generator in his backyard in Ohio, United States. The generated electricity was used as supply power for his mansion like Blyth. The configuration he used was a post mill with a tower of 18 m, a wheel with 144 blades of 18m in diameter with a capacity of 12 kW electricity generation [46].

Denmark was one of the pioneer countries in Europe implementing electricity-generating wind turbines in the early twenty. Seventy-two wind power turbines with a range from 5 to



(a) First global attempt towards large wind farms, Balaclava, Russia, 1931

(b) Smith-Putnum turbine, The world's first megawatt sized turbine, Vermont, USA, 1942

Fig. 2.8 History of large wind turbines in the world [44]

25 kW were used all around the country by 1908. The giant machines were on 24-meter towers with a four-bladed rotor. One of the first steps in developing large-scale wind power plants for electric utility applications was taken in Russia in 1931, with the construction of a 100 kW, a 30-meter diameter wind turbine at Balaclava, on the Black Sea. The turbine is shown in figure 2.8a. Later in 1940, around 40,000 wind power mills with a capacity of 50-100 kW were used in large farms in Russia. In 1941, the Smith-Putnam turbine, the first megawatt-sized wind machine with a capacity of 1.25 MW, was connected to the electricity grid for the first time in history in Vermont, United States. The turbine, shown in Figure 2.8b, had a two-bladed rotor swept an area of 53.3 meters in diameter [44].

Until the OPEC oil embargo triggered an energy crisis in 1973, the widespread use of individual wind generators competed against fossil fuel plants and centrally-generated electricity. From then, a new stage of wind turbine design and development started. This new stage is named the modern era of wind turbines.

2.3.1.1 Modern wind turbines

The 1970s can be considered a logical start point for the modern wind turbines, during which the shortage of oil changed the energy environment for the world. The oil shortage directed more investments in developing renewable energies such as wind turbines to generate electricity. Economically, windmills got more attention compared to solar cells for many people who began to desire a self-sufficient lifestyle.

From 1974 through the mid-1980s, NASA started a wind turbine program to produce utility-scale wind turbines, which is considered a breakpoint in designing and developing electricity-generating wind turbines. The giant wind turbines developed under this program have set several world records in terms of diameter and power output. Many of today's multi-megawatt turbine technologies, including aerodynamic, structural, and acoustic engineering design capabilities as well as steel tube towers, variable-speed generators, composite blade materials, partial-span pitch control, are the results of this program.

Soon after that, the emergence of modern wind turbine technology appeared in Europe, where Denmark, Sweden, and Germany were the pioneers. Denmark was the only country after the energy crisis in 1973, during which there was a particular tradition of successfully operating small wind turbines for power generation. Until 1985, 30% of the turbine purchase values were paid to the operators as a direct subsidy. Moreover, generated electricity from wind sources was tax-free. This economic background, introducing rural Danish settlements, increased the processing speed of granting building permission and the test certificate of technical maturity and safety of the units, encouraged widespread use of wind turbines in Denmark. Other facts like publishing the Danish wind atlas have also been effective. [40].

The development of turbines is characterized by their wind power production, which requires enlargement of their size, blade diameter, and height. Figure 2.9 shows a modern wind turbine in a mountainous wind farm in Galicia, Spain. The region is named Alto do Xiabre. The photo was taken by the author.

2.3.2 Wind turbine components

Wind turbines operate by transforming air kinetic energy into mechanical power uses for electricity generation by spinning a generator. Figure 2.10 shows different components of a modern wind turbine, explained in the following.

Fig. 2.9 A three-blade modern wind turbine located in a mountainous wind farm in Galicia, Spain

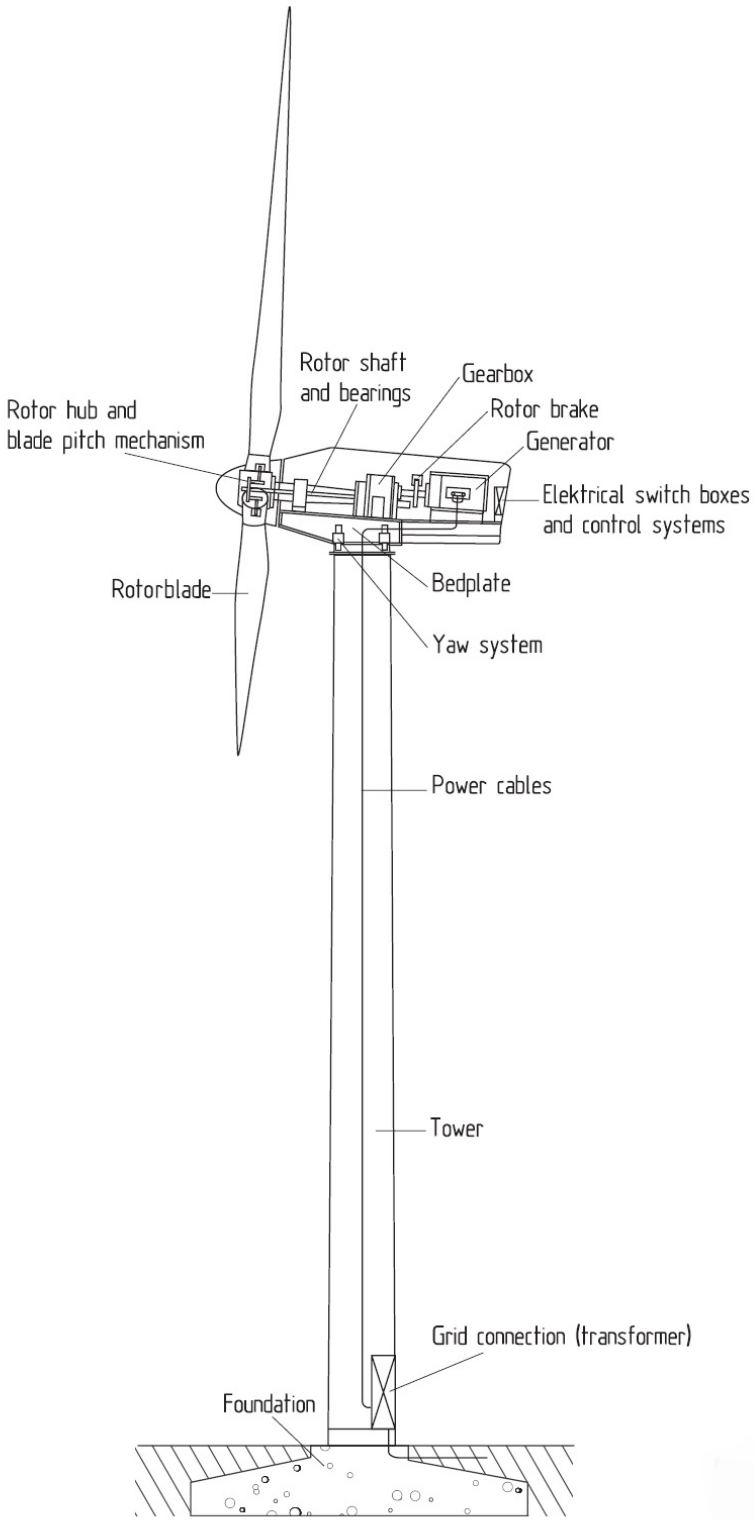


Fig. 2.10 Components of a modern horizontal axis wind turbine [40]

- **Rotorblade:** Lifts and rotates when the wind is blown over them, causing the rotor to spin. Most turbines have either two or three blades.
- **Rotor hub:** Holds the blades and connects them to the main shaft of the wind machine. It is a crucial component not only because it holds the blades in their proper position for maximum aerodynamic efficiency, but it also rotates to drive the generator.
- **Blade pitch:** Turns (or pitches) blades out of the wind to control the rotor speed and keep the rotor from turning in winds that are too high or too low to produce electricity.
- **Rotor shaft:** Drives the generator.
- **Gearbox:** Connects the low-speed shaft to the high-speed shaft and increases the rotational speeds. The gearbox is a costly part of the wind turbine. Engineers are exploring "direct-drive" generators that operate at lower rotational speeds and do not need gearboxes.
- **Rotor break:** Stops the rotor mechanically, electrically, or hydraulically in emergencies.
- **Generator:** Produces 60-cycle AC electricity. It is usually an off-the-shelf induction generator.
- **Control system:** Start up the machine at normal wind speed and shut it down at high wind speed. Turbines do not operate at high wind speeds because they may be damaged.
- **Bedplate:** holds the upper part of the system and transfers the loads to the tower.
- **Yaw system:** Orients upwind turbines to keep them facing the wind when the direction changes.
- **Tower:** Supports the turbine structure and is made of steel or concrete. Because wind speed increases with height, taller towers enable turbines to capture more energy and generate more electricity.

2.3.3 Locations to install onshore wind turbines

Sitting wind turbines on ridges higher than the surrounding areas or placing them on hills is a common way and location to install them. Mainly, installing the turbine wherever as broad a view as possible in the prevailing wind direction in the area is always an advantage.

Because the wind becomes compressed, passing around the hills, expanding again as it soars down into the low-pressure area, the wind speed is higher on hills than its surrounding. Those places are ideal for wind turbine installation [47]. Valleys are another possible good location that needs to be studied in the feasibility and design phase [48]. Besides the altitude, other factors such as topography, porosity, roughness, turbulence intensity, wind shear, and atmospheric stratification are responsible for the high wind speed magnitudes. Frost and Shieh [49], investigated wind characteristics over complex terrain related to wind energy conversion system and presented an outstanding description for a good selection of wind turbines based on theoretical and laboratory investigations, as shown in table 2.3.

Table 2.3 Wind flow regions classification for wind turbine installations

Wind speed strength	**Wind speed turbulence**	**Site selection**
High	Low	Good
Low	High	Poor
Periodic	Very high	Avoided

Approaching scientifically to wind turbines and farms optimal location, a wind assessment is needed to calculate the wind energy potential of a region besides doing risk analysis for extreme or maximum wind speeds [50]. The importance of wind resources assessment to define wind turbines or farms' installation location is that a minor change in wind speed causes a significant deviation in generated power. Various methods and software are using for mapping, modeling, and energy estimation for the potential locations, including WindPro, Computational Fluid Dynamics (CFD), Wind Atlas Analysis (WAA), and Geographical Information System (GIS). There are different wind assessment methods discussed in the literature [51]; however, using statistical analysis and applying Rayleigh and Weibull distributions are among the widely-used techniques.

2.3.4 Wind Power output

The available power from wind is the flow rate for kinetic energy and is described below, in the equations 2.9 and 2.10.

$$U = \frac{1}{2} m_a v^2 \quad (J) \tag{2.9}$$

$$P_{ideal} = \frac{dU}{dt} = \frac{1}{2}\rho A v^2 \frac{dx}{dt} = \frac{1}{2}\rho A v^3 \quad (W) \tag{2.10}$$

Where U is the kinetic energy, m_a is air mass, v stands for the average moving wind speed, P_{ideal} is the theoretical extracted wind power, ρ is the air density, A is the area of moving air packet, and J and W represents Joule and Watt, respectively.

There is a theoretical maximum power generation coefficient for any turbine, named Betz's Law. Based on the conservation of the mass principle, the amount of air entering and exiting wind turbines is equal. Bet's law limits the maximum achievable power by a wind turbine as $16\,/\,27$ or 0.593. In other words, an actual turbine can not extract more than 59.3% of power in a specified tube of air of the same area. In practice, the fraction of extracted power is always less because of mechanical imperfections. The power coefficient, C_p, reflects the ratio of actual energy production to the available energy in the area of wind turbine blades. However, modern turbine designs can approach this limit, reaching up to 70% - 80% of this theoretical limit. In equation 2.11, dependency of C_p on weather-related inputs makes the power coefficient a functional curve, often plotted against the tip speed ratio, shown in figure 2.11. Qualitative differences of power coefficient are shown for the rotors of different configurations. The advantage of modern rotors compare to traditional ones can be seen in the figure.

$$P_{practical} = P_{ideal} C_p = \frac{1}{2}\rho A v^3 C_p \quad (W), \quad max\, C_p = 0.593 \tag{2.11}$$

Where $P_{practical}$ stands for the practical extracted wind power from the turbine and C_p is the coefficient power, also called the performance coefficient.

Power coefficient is typically a function of the turbine parameters, including tip speed ratio, attack angle, air density, and atmosphere parameters. The tip-speed ratio, λ, is the ratio between the tangential speed of a blade's tip and the actual wind speed, v, as formulated in the equation 2.12.

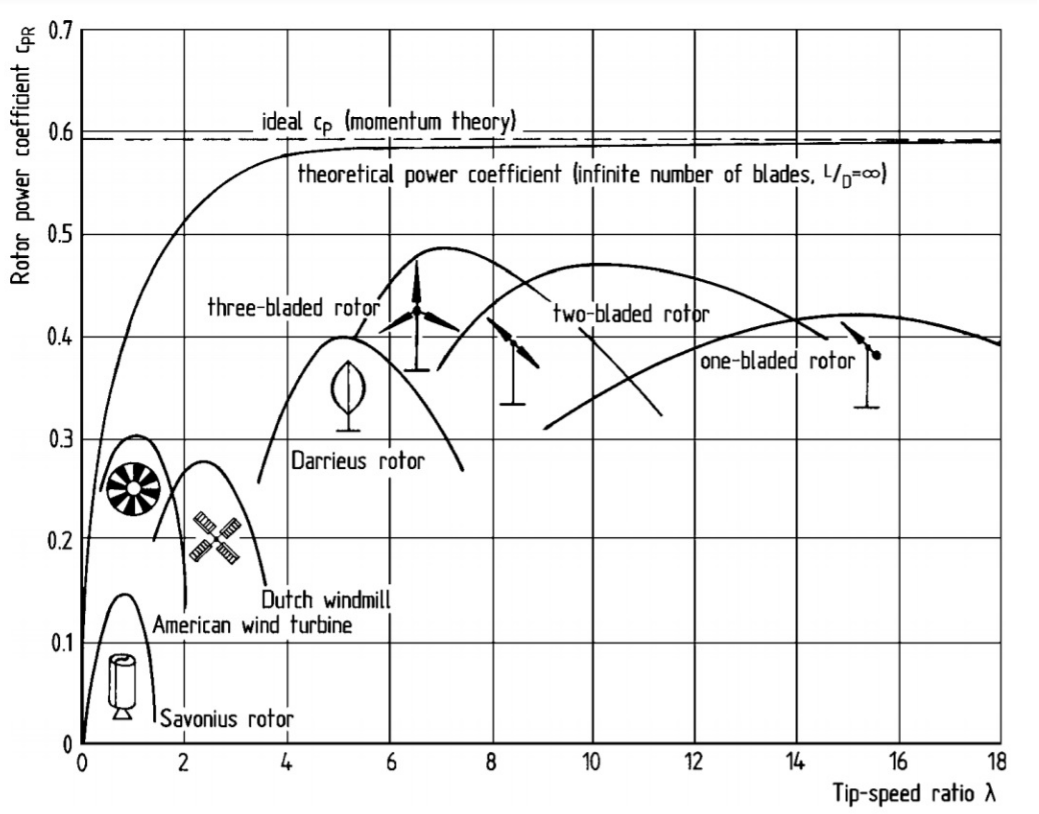


Fig. 2.11 Power coefficient curve for different rotor designs [40]

$$\lambda = \frac{Tip\ speed\ of\ blade}{wind\ speed} = \frac{\omega R}{v} \tag{2.12}$$

Where λ is tip speed ratio, ω is the rotational speed of the rotor in radians/second, R is rotor radius in meters, and v represents wind speed.

Another important design parameter for blades is attack angle, which is an aerodynamic parameter. Maintaining rotor speed at a predetermined range is vital in high wind speed situations or when the generator torque is suddenly lost during a grid outage. To fulfill this task, the rotor should have an aerodynamically effective means to limit its power and rotational speed. Changing the angle of attack, turning the rotor out of the wind to reduce the rotor's swept area (furling), and changing effective free-stream velocity at the rotor blades could control the aerodynamic forces. The latter approach is minimal, and the second one is applicable only on small turbines [40].

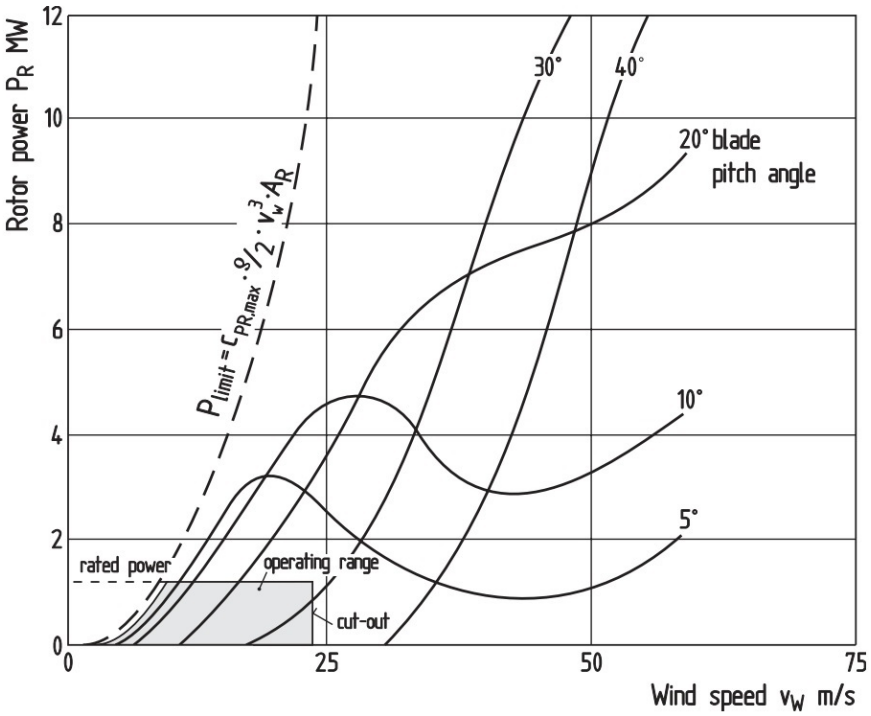


Fig. 2.12 Rotor's power curve vs. wind speed for different blade pitch angles and fixed rotor speed (Wind turbine model: WKA-60) [40]

Betz's law imposes a fundamental limitation to the extracted wind power. There is another technical limitation due to the design strength of the rotor structure, which controls the rotor power output by the maximum permissible power of the generator. A control system imposes this limitation. Figure 2.12 shows the extent to which the power input of the rotor increases when it is in the operation period, and the wind speed is less than the cut-out range. Changing pitch angle is the most effective way of controlling the aerodynamic forces by changing the angle of attack. The contents of the figure are provided for the WKA-60, which is a wind turbine model manufactured by WKA company, with a rotor diameter of 60 meters and a rated power of 1200 kW.

The practical power curve reflects the power response of a turbine to wind speed. Power curves are different for each specific turbine and provide by its manufacturer. As shown in figure 2.12, there is an operating range and a specific wind speed value for each line, above which the turbine output power is consistent. A rescheduled threshold by the turbine power control incurs in order to prevent damages to the turbine components. In figure 2.13 this concept is shown in more detail. When the wind speed is less than a threshold minimum,

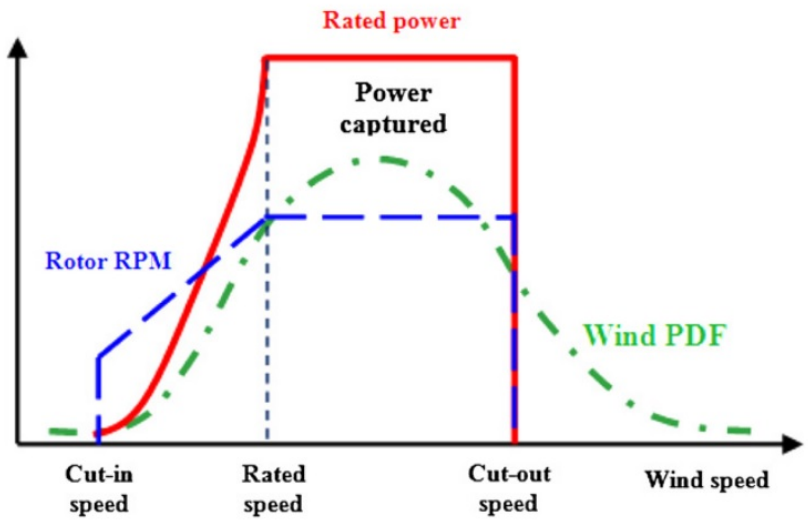


Fig. 2.13 Typical power output versus wind speed [52]

known as cut-in speed, the power output becomes zero. Between cut-in and rated speed, there is a rapid growth of power produced. A constant output (rated) produces until the cut-off speed attains, and beyond this speed, the turbine takes out of operation.

It is important to consider the variability and intermittency of wind speed in the wind power assessment since low wind speeds can lead to interruptions of electricity generation, and very high speeds exceeding 25 m/s need the cessation of power turbines. Hence, this variability is the main uncertainty associated with the mean power production of wind turbines [53].

2.3.5 Environmental and human-effects of wind turbines

Electricity generation from wind resources has the potential to reduce environmental pollution by reducing CO_2 emission and thermal pollution compared to the traditional sources of energy as coal, oil, and gas. However, on the other side, renewable energies leave some impacts on the environment, humans, ecosystems, and atmosphere, which need to be carefully considered. Below, the wind turbines effect on the environment and human is discussed.

Visual, noise, shadow flicker, and economic are four substantial human impacts of wind power plants. There is a delicate point in the aesthetic impacts of wind turbines to interpret it as either positive or negative. Wind turbines are found to be visually beautiful by many people, unlike some forms of construction such as cell towers. However, when a beautiful huge

object stands in one's current surroundings, it might not be desirable anymore. Generally speaking, although wind power generation is scientifically and socially supported, there is less support for projects close to our own home [54].

Both the noise and shadow flicker can be aesthetically troubling for the people who live nearby. The noise problem is usually serious within a half-mile of the project and amplifies as the wind speed increases. The nearby sound power level is usually about 90-105 decibel (dB) [40], while the comfortable sound level is between 40 and 60 dB. Shadow flicker is defined as a variation of brightness at a given location in the presence and absence of a shadow. It is a function of several factors such as turbine geographic latitude, location of nearby people due to the turbine, diurnal variation of sunlight, wind speed and direction, site's topography, and whether or not there is any obstruction. During the mornings and evenings, when shadows are long, the problem can get even more severe. This phenomenon should be modeled in the project's design phase to prevent problems for the nearby living people. In general, negative human impacts of wind-energy projects are mainly experienced by people living close to the sites. Regarding that fact, wind energy generation projects' benefits can increase while temper socioeconomic impacts by actively engaging stakeholders and undertaking local interactions in the very first steps of the feasibility studies of the project [54], and strictly adhering to the human-based impacts standards in the design phase of the project.

From an ecosystem point of view, a wind energy project development can either indirectly impact habitat structure and functioning or directly impact individual organisms. Birds and bats fatalities, soil disruption, and vegetation clearing are the main areas affected by wind farms. Referring to the numbers aggregated by the United States Fish and Wildlife Service, as the top threats to birds and their mortality rate in the country during 2017, wind turbines are not considered as a fundamental cause of birds mortality [55]. Average estimated numbers for different hazards threaten birds live are extracted from the same reference and provided in table 2.4. However, due to establishing more wind power plants, the role of wind turbines might be increasing. In order to prevent such ecosystem damage and reduce birds and bats' mortality rate, some solutions are provided. As an example, black paint on wind turbines is proved to be helpful. A recent study at Smøla wind-power plant in Sweden showed the success of this solution by over 70% reduction of birds mortality rate [56].

Table 2.4 Top threats to birds and their mortality rate in the U.S. during 2017

Hazard type	Average estimated number
Cats	2,400,000,000
Collision- building glass	559,000,000
Collision-vehicles	214,000,000
poison	72,000,000
Collision-electrical lines	25,500,000
Collision-communication towers	6,600,00
Oil puts	750,000
Collision- Onshore wind turbines	234,012

The construction of wind farms also impacts the ecosystem through soil disruption or erosion and vegetation clearing. As wind farms are usually constructed in remote areas, the risk of deforestation is severe in some cases. Considering all the construction and transportation needs, it might cause a loss of habitat for some species. Besides, changes in forest structure and create openings alter micro-climate and increase the forest edges. Different responses from plants and animals to this ecosystem change may occur, and subsequently, there might be essential interactions between habitat alteration and the risk of fatalities, such as bat foraging behavior near turbines [54].

More than human and ecosystem impacts, wind power plants, particularly wind turbines, cause atmospheric influence so that by passing the airflow through a wind turbine, the mean speed would slow down with more turbulence. More turbulence means better vertical exchange in the atmosphere and thus less vertical wind shear. In the larger wind farms, agricultural conditions would change because the near-surface cooling is reduced at night, and evaporation from the ground is enhanced [26]. Increased surface roughness and drag modifies surface-atmosphere exchanges and the transfer of energy, momentum, mass, and moisture within the atmosphere.

2.4 Wind data and prediction models

Both wind speed and power prediction models can be categorized into three groups. Wind speed forecasting methods are introduced as physical, statistical, and machine learning models. However, for wind powers, instead of physical models, the mathematical formula of

wind power, equation 2.2, is applied and the other two groups (statistical and ML techniques) are identical. It is essential to mention that different articles have provided different views about categories of wind speed and power prediction models. In some works, neural networks are classified as statistical models [57]; however, in others [22], they are considered nonlinear artificial intelligence or machine learning models. We follow the second classification in this work and session, considering linear time series models as statistical and nonlinear methods as machine learning models.

2.4.1 Wind data acquisition sources

Surface wind speed can be acquired through a range of technologies, some of which are widely used and some are still under development as new-born methods. Weather stations with observational sensors and different types of anemometers are the commonly used approaches to collect surface weather information, wind speed included. Anemometers are generally placed at the nacelle of a turbine or at a meteorological station or a met mast for which data can be used for controlling a wind turbine or, in general, for wind assessment. As wind speed rises considerably with height, particularly over rough terrain, the standard elevation of wind instruments installation is 10 meters above the ground, where measured wind speed represents the wind over an open area of a few kilometers. The wind direction shift over such a height interval is negligible [38].

Wind vane and cup or propeller anemometers are the two most commonly used instruments that are simple and low-cost compare to other types [58]. Cup or Robinson anemometer, invented in 1845 by Dr. John Thomas Romney Robinson, consists of a vertical pillar and three or four concaves that capture the horizontal movement of air particles. Propeller and vane anemometer, on the contrary, have their axis parallel to the direction of the wind and therefore is horizontal. Vane is used as the tail of a propeller anemometer to measure wind's direction. The wind direction is estimated to the nearest of the 16 points of the compass. As a minor technical problem, the response of both cup and propeller wind sensors is faster for acceleration than for deceleration. So there is an overestimation of the actual average wind speed. More detailed technical information about these instruments can be found in [38]. There are different commercial types of single and combined traditional anemometers, some of which are shown in figure 2.14. As another type of wind speed measurement instrument, pitot-tube anemometers measure the overpressure in a tube in the direction of the wind vector, kept aligned utilizing a vane. Without requiring electrical power, this instrument also provides proper gustiness records [59].

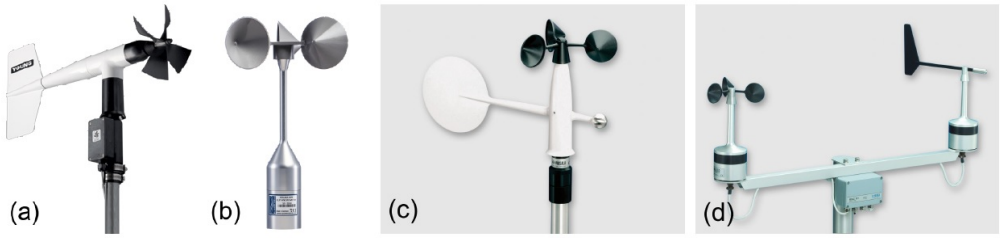


Fig. 2.14 Different types of traditional wind speed anemometers; a) Propeller anemometer b) Cup anemometer c) Combined cup-vane anemometer d) Cup and vane anemometer [37]

Sonic anemometers were first developed in the 1950s based on using ultrasonic sound waves to measure wind velocity. Sonic anemometers are appropriate for harsh environments, for example, on top of mountains or at the nacelle of wind turbines, as they do not have any moving parts and are maintenance-free instruments. The speed of acoustic signals, ultrasound waves traveling between two transducers, uses as the indicator for wind speed measurement. Wind speed may speed up the ultrasound's movement or slow it down. Sonic anemometers can capture turbulence as they can take measurements of 20 Hz or more, with very fine temporal resolution. They also have slight accuracy deterioration, and high durability [60]. Different types and shapes of (ultra) sonic anemometers are shown in figure 2.15.

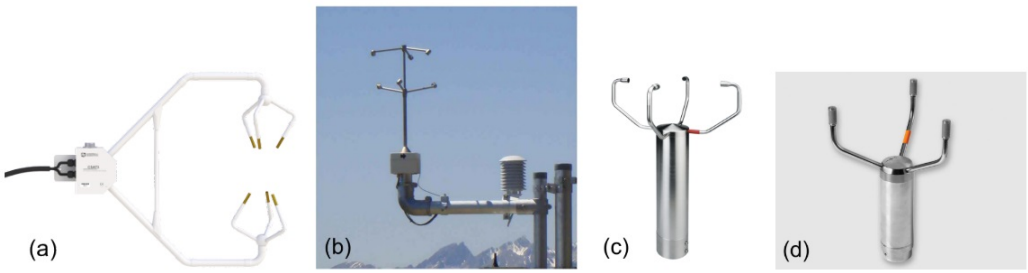


Fig. 2.15 Different types of (ultra) sonic anemometers; a) CSAT3 3D b) uSonic-3 Scientific c) Ultrasonic Anemometer 2D d) WINDCAP 2D Ultrasonic [37]

Hot-disc and hit-wire anemometers are two other wind speed measuring instruments sharing some similarities. The first one is a solid-state that measures temperature gradient across a chip arrangement. However, this type is steady and firm in uncertainties; the operational experience has been limited recently. The hit-wire instrument measures the cooling of thin heated wires. Because of their fragility and also rapid calibration needs in wet or unclean conditions, they are also operationally unreliable and are using less [38].

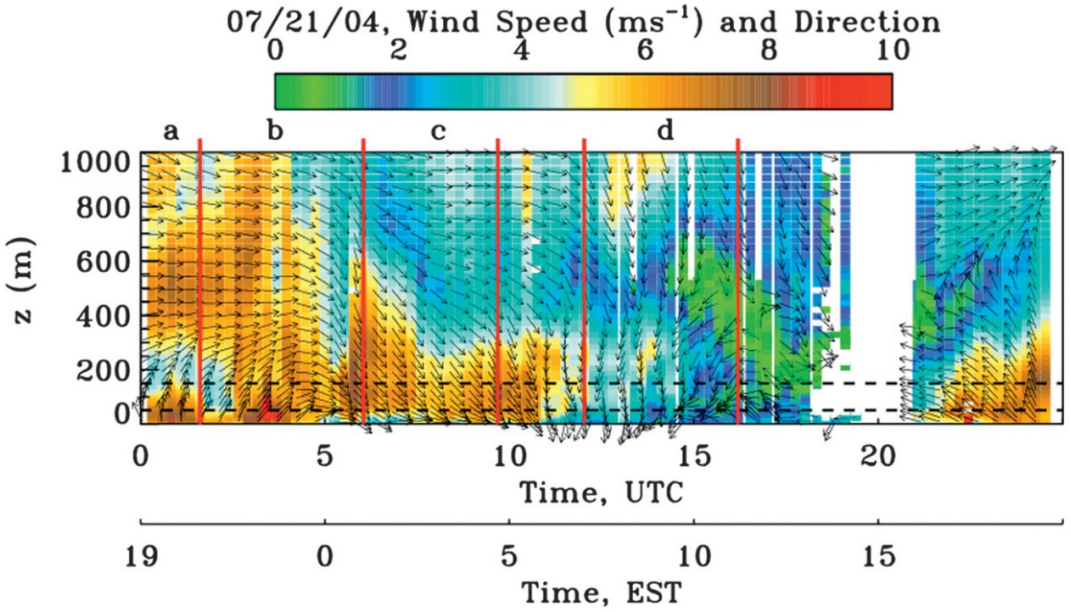


Fig. 2.16 Computed mean wind speed and direction by high-resolution Doppler Lidar-based conical scans, July 2004, Boston, U.S. [62]

As recently developed methods for wind speed measurement, remote sensing (RS) techniques are precise and accurate, which use sound waves or electromagnetic signals, named SODAR, and LIDAR, respectively. These techniques, however, are less used in routine meteorological networks as surface observational tools. However, these two techniques are suitable for low wind speeds from 1 to 70 m/s measurement and perform more accurately than cup anemometers, but they are low resolution in terms of space and time [61]. Wind masts can typically cover up to 100-meter height. However, RS goes beyond up to a few hundred meters in different directions, measuring wind speeds at the turbine's hub height. Both Sodar and Lidar techniques detect air movement in the atmospheric boundary layer (ABL) and calculate wind speed and direction by employing the Doppler effect. The difference is that in Lidar, electromagnetic wave, accurately saying laser beam, is reflected off particles, but in Sodar, a wave of sound reflects off the varying temperature structure in the atmosphere. Figure 2.16 shows an example of wind speed and direction measurement of the Lidar technology.

Flight data acquisition using manned and unmanned devices is another type of weather and meteorological data gathering with more recent attention due to their popularity for civilian, military, and research applications and the integration of an increasing amount of sensing technologies into these vehicles. For instance, in cooperation with dozens of airlines worldwide, the World Meteorological Organization (WMO) is running an aircraft-based

meteorological observation system named AMDAR program [63]. These collected and transmitted data to the ground via very high frequency (VHF) or satellite links using the aircraft communication system as a valuable source of observational data to improve forecast models. However, this program and other similar procedures may be considered as an upper-air data gathering method. Otherwise, there are research aircraft at a lower height lever to collect meteorological data [37]. Unmanned aerial vehicles (UAVs) are more flexible and economical than manned aircraft, especially for measuring properties around wind turbines to provide observational data for wind turbine design methods [64].

2.4.2 Physical models

Physical or numerical weather prediction (NWP) models are the physical basis of software programs that solve the physical equations of atmospheric conditions and changes over time. NWP models are in origin an initial value problem proposed by Vilhelm Bjerknes in 1904 [65], in a way that by knowing the initial condition of the atmosphere, changes can be simulated by applying physical force that acts over time. The atmospheric motion-related principal equations, also known as primitive equations, including momentum, mass, energy, in addition to the state and water conservation equations, are simplified models of the actual atmosphere's condition [66]. NWP models are computationally expensive due to their non-linear and the large-scale that they cover.

NWPs are classified as either global models which solve the primitive equations for the whole Earth glob or limited area models (LAM) covering only a limited domain. As mentioned before, running physical models is computationally expensive, and because of that, there are just a few centers around the world that can run global models. As some of these centers, we can name the European Center for Medium-range Weather Forecasts (ECMWF) with 25 km resolution, and National Center for Environmental Prediction (NCEP/NCAR), and the German Meteorological Center (GMC), both with 40 km resolution. Models developed and run by these centers are low resolution. Hence they are not capable of detecting small-scale phenomena and are mainly used for macroscale forecasting. However, many meteorological departments run LAMs for limited areas such as countries and regions with a higher resolution up to 2 or 3 km, used for mesoscale forecasting by only inputting the boundary condition from a global model. LAMs are different in several terms, such as equation simplification, assumptions, and numerical mathematical formulation. As the most widely-used ones, we can name ETA, HRM, MM5, ALADIN, COSMO, and Weather Research & Forecasting or WRF [67], between which WRF, ETA, and MM5 are freely available for download. As a common approach, most LAMs usually run twice a day at 00:00 and

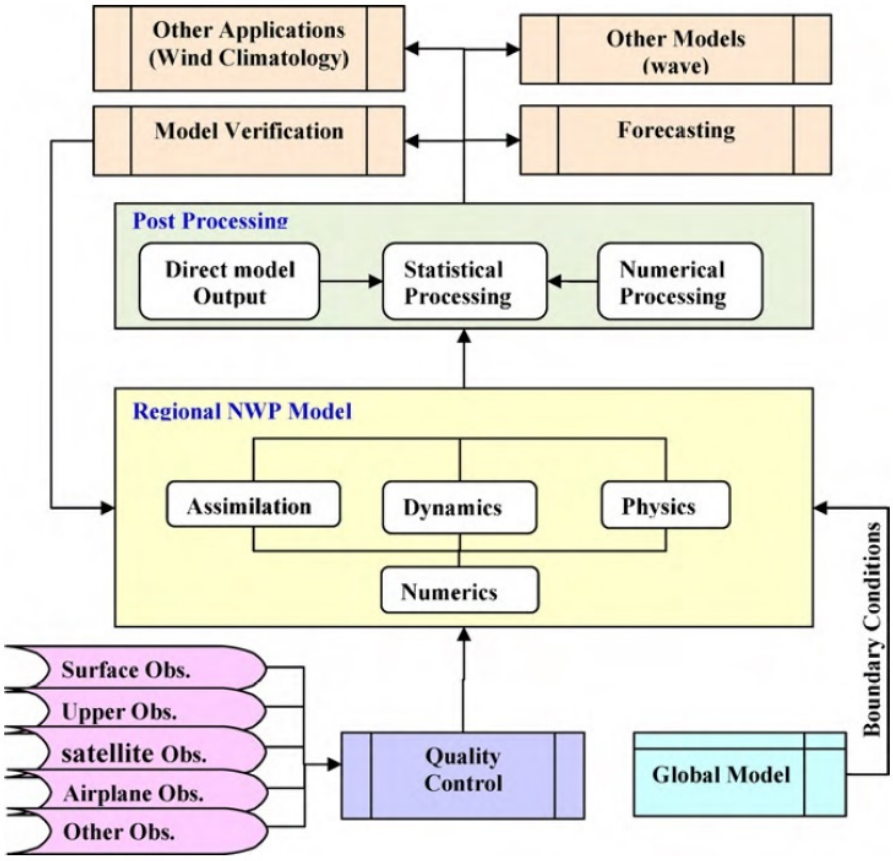


Fig. 2.17 High-resolution NWP models forecasting process [67]

12:00 and perform a three-days ahead (72h) prediction.

NWP models solve the equations for each atmosphere's vertical layer, and for this reason, wind speed data is also calculated at higher levels of the atmosphere. In figure 2.17, a typical high-resolution NWP process data flow is shown, which consists of dynamical, physical, and numerical processes. The dynamic process includes pressure gradient forces, adiabatic cooling and heating, and convection. The physical process contains a lower scale than the model resolution, such as precipitation and cloud micro-physics. The numerical process is about handling physical and dynamical processes by formulation and resolution handling. Assimilation creates the initial condition of the atmosphere using all sources of data (satellite, airplane upper, and surface observations) and forecasts to prepare the NWP model to start the forecast process.

Physical models need to be down-scaled before being applied to wind farms. This down-scaling can include a detailed physical description of the wind farms such as roughness,

obstacles, orthography as terrain characteristics, and wind farm layout and power curves as wind farm characteristics [57]. Defining these physical descriptions is one of the most significant drawbacks of the NWP models for their application in the wind energy sector which nominates statistical and machine learning models as more convenient and precise solutions. The wind turbine's power calculates subsequently through the power curve of the wind turbines when using physical models for wind prediction at hubs' height. Earlier in this chapter, a power curve of turbine WKA-60 is provided in figure 2.12.

2.4.3 Statistical models

Statistical modeling is one of the main substituted approaches for physical wind speed and power prediction models, which use NWPs, observational and historical data to perform prediction. These models are known to be easy to use and inexpensive methods. Statistical or time series models such as autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), Autoregressive conditional heteroscedastic (ARCH), etc., are the most successful for very short-term and short-term forecasting. However, nowadays are using mainly as reference models [68] as more advanced machine learning regression methods are developed [57]. A general mathematical formulation of the ARMA model is presented in equation 2.13.

$$X_t = k + \varepsilon_t + \sum_{i=1}^{p} \gamma_i X_{t-i} + \sum_{i=1}^{q} \phi_i \varepsilon_{t-i} \tag{2.13}$$

Where X_t stands for the predicted value at time t, γ_i represents an autoregressive parameter, ϕ_i is the moving average parameter, ε_i is white noise random variable, and k is a constant. These models are written as $ARMA(p,q)$, where p and q are the orders of autoregressive and moving average, respectively.

Different statistical models have been applied successfully [69, 70] for wind speed and power predictions. In recent works, statistical models are used in combination with machine learning techniques, capturing the linear part of signals successfully [71].

2.4.4 Machine learning or data-driven models

Machine learning (ML), data-driven, or artificial intelligence techniques are newer than the other two methods to the wind speed and power prediction sector. These models are

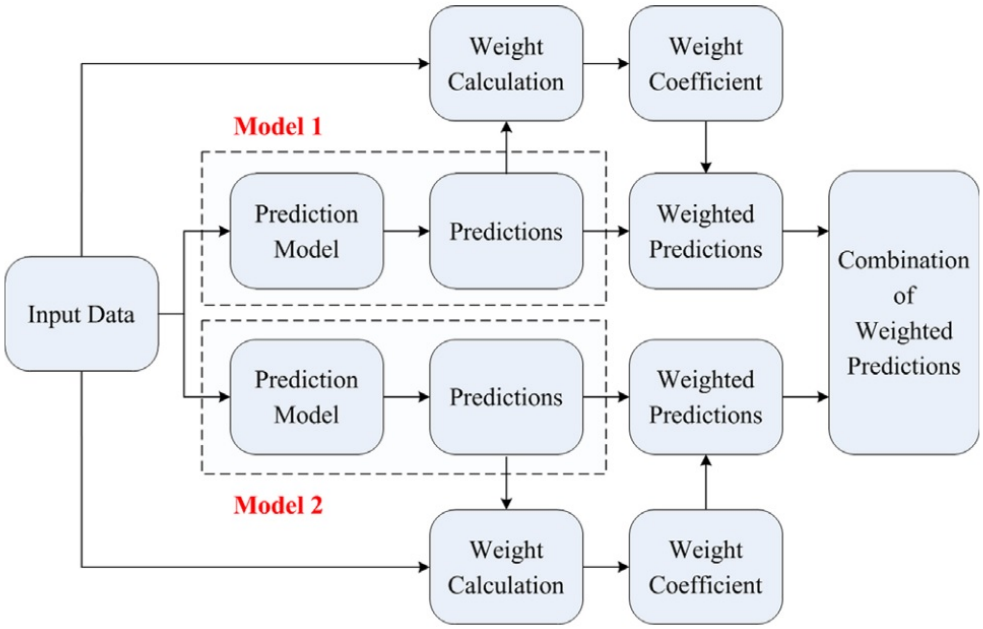


Fig. 2.18 A schematic view of combined models with weight-based approach [79]

mushrooming by inventing new algorithms and developing their training processes. Besides, as physical models produce coarse forecasting for shorter time horizon predictions, ML and statistical techniques are reported to be successful in overtaking NWP models. There is a wide range of ML techniques in the literature applied in WS&PF, including artificial neural networks, fuzzy logic methods, support vector regression and machines, random forests, Bayesian networks, and metastatic algorithms like genetic algorithms, particle swarm optimization, etc. However, hybrid, and combined ML models [72], are repeatedly reported to overcome single models.

Among different types of ML methods, neural networks are the most widely used models. ANN's performance depends on many factors such as their structures, learning methods, how neurons are connected, and data preprocessing. There are many types of neural networks both in shallow and deep learning techniques applied for WS&PF such as multi-layer perceptron (MLP) [73], wavelet neural networks (WNN) [74], convolutional neural network (CNN) [75], long-short-term memory (LSTM) [13], Elman neural network (ELMAN) [76], Back-propagation neural network (BPNN) [77], Radial basis function (RBF) [78] , etc. A comprehensive review of neural networks applied in wind speed and power prediction is provided in Chapter 4.

The exact definition of hybrid and combined models are not agreed upon. However, referring to the accomplished review in Chapter 4, these two terms can be defined as follows. Different types of core prediction models are incorporated in the combined algorithms producing various predictions with different performances. Then, through a combination approach such as weighting-based, the outputs of all individual models combines to provide final forecasting. Hybrid models, however, consist of a unique prediction core model or a linear model in conjunction with a nonlinear one [79]. A flowchart of combined models with a weight-based approach is illustrated in figure 2.18. Different types of optimization techniques, signal reprocessing, and error correction can be included in both hybrid and combined models in this provided categorization. Although there is no guarantee that a combined or hybrid model always performs better than single models, the success of these models is frequently reported [80].

Signal prepossessing techniques are considered an important part of the wind speed and power ML prediction models due to their capability to deal with the input signals' stationary and non-stationary characteristics. Various types of these techniques are used successfully in hybrid and combined models of wind speed and power prediction such as singular spectrum analysis (SSA) [81], variational mode decomposition (VMD) [82], wavelet transform (WT) [83], empirical mode decomposition (EMD) [84], ensemble empirical mode decomposition (EEMD) [85], etc. Combining a decomposition method into a prediction model will help to decrease the non-stationary feature of the original wind speed data [84].

2.5 Summary

In this chapter, a general overview of wind speed and power meteorology, wind power generation, and different groups of models for WS&PF was presented. By starting from wind origins at the global and local levels, more information is provided about the wind components and the atmospheric and environmental causes for wind fluctuations. In the next step, a history of wind power turbines was presented followed by more discussions on modern wind turbines, their components, and the power output dependencies. Finally, a summary of physical, statistical, and machine learning models provided, and the superiority of different ML techniques, especially hybrid and combined models for short-term forecasting over small-scale areas, were discussed.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 3

Case Studies

3.1 Introduction

In this chapter, six different case studies used in the developed models later in this work are introduced and analyzed using data analysis techniques to understand their characteristics. First, Three meteorological stations in Galicia representatives for urban, coastal, and rural highland areas, including one-year weather data of wind speed, wind direction, temperature, humidity, precipitation, pressure, and global radiation during 2017. Besides, Weather Research and Forecasting (WRF) model numerical weather prediction data for the same time and parameters are collected for all three stations. Second, Sotavento wind farm, located in Galicia, Spain, is the forth case study introduced in this chapter. The fifth case study corresponds to a meteorological station, M2 Tower, is located in Colorado, the USA, including one-year high-resolution data with 1-minute time-frequency and six different height levels. Finally, the sixth case study is Le Haute Borne wind farm located in France, including one year of observational weather parameters and the output wind energy from several wind turbines. These six case studies provide a wide range of characteristics used in different models with several purposes.

3.2 Case studies

3.2.1 Meteorological stations in Galicia

Wind energy has been widely developed in the last decades in the Galician territory that enjoys enormous wind energy resources. This process involved installing more than 4,000 wind turbines in many of the mountains areas of about 100 municipalities linked to more than

150 wind farms. Since 1995, the wind turbines and wind farms implementation and their associated infrastructures have started on many mountainous areas in Galicia. Until 2017, more than 3,300 MW capacity in operation is installed with the majority of turbines put into operation during the period from 1997 to 2008. The unit power of the wind turbines in the pipeline is increasing significantly, in a way that in many cases it is about 3 MW per wind turbine and even reaches 4.5 MW in some areas. Considering that, it is expected that the average power of the windmills of Galicia will experience significant growth. As a matter of fact the average power per wind turbine in operation in Galicia has experienced a continuous increase from 1998 to 2009, and a second increase starting in 2018 [86].

3.2.2 Meteorological stations in Galicia

Three meteorological stations located in Galicia were selected from MeteoGalicia network (Xunta de Galicia), as they represent different environmental characteristics: coastal, urban, and highland rural areas. Figure 3.1 shows these stations locations in Galicia, and pictures of them. Also, table 3.1 is a summary of the geographical information of these meteorological stations, from the MeteoGalicia website [87]. While these stations are different in many factors such as location, year of establishment, sensors type and quantity, etc., there are some general characteristics, which are identical for all of them and summarized in table 3.2. Besides representing different environments, these shared characteristics following listed made these stations good references to be selected for this study,

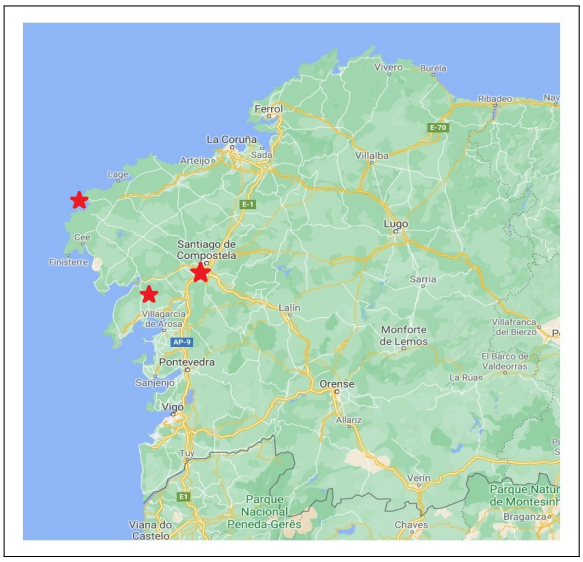
1. **Time series length**: Data for the whole year of 2017 is acquired from all of the selected stations, from 1 January to 31 December.

2. **Data acquisition frequency**: Data is gathered based on a 10-minutes time interval for all the stations.

3. **Altitude of WS and WD**: All stations provide wind speed and direction in standard height (10 above ground level meters).

4. **Other parameters**: Besides wind speed and direction, other weather parameters such as temperature, precipitation, pressure, humidity, and solar radiation are collected for the same period.

As wind speed and direction is the main focus in this work, the three stations install the same wind anemometer and vane type: 05106-5 MA model, classified as vane-propeller type. Its wind speed and direction uncertainties are ± 0.3 m/s and $\pm 3.0°$, respectively, and they are capable of measuring a maximum wind velocity of 100 m/s.



(a) Geographical location

(b) Camariñas, Coastal

(c) Muralla, Highland rural

(d) Santiago, Urban

Fig. 3.1 Geographical locations and images of the selected meteorological stations in Coastal, Highland rural and Urban areas in Galicia, Spain [87]

Table 3.1 Geographical information of the selected meteorological stations in Galicia

Station's Name	Latitude	Longitude	Altitude(m)	Terrain's type
Camariñas	43.124447	-9.178318	5	Coastal
Santiago-EOAS	42.87596	-8.559434	255	Urban
Muralla	42.745583	-8.776306	661	Highland rural

3.2.3 WRF numerical weather prediction data

Besides observational data acquired from meteorological stations, also numerical weather prediction data from operational WRF model (MeteoGalicia, [86]) covering Galicia at 4x4 km2 resolution were collected. Both observational and WRF data characteristics used in this study are summarized in table 3.2.

Table 3.2 Santiago-EOAS, Muralla, Camariñas data characteristics used in this study

	Items	Description
Obs.	Format	CSV, JASON, PDF
	Parameters	WS, WD, T, H, Rad, P, Prec
	Duration	1 Jan- 31 Dec 2017
	Time interval	10-minutes
WRF	Format	NetCDF
	Resolution	4 km
	parameters	WS, WD, WG, T, P, H, Rad, Prec, Cfl
	Duration	1 Jan- 31 Dec 2017
	Time interval	hourly
	Downscale modeling	IDW

Nomenclature: Obs.: Observational, WS: Wind speed, WD: Wind direction, WG: Wind gust, T: Temperature, P: Pressure, H: Humidity, Rad: Solar radiation, Prec: Precipitation, Cfl: Cloud cover at low levels, IDW: Inverse distance weighting.

3.2.4 Sotavento wind farm

Promoted by the Galician government, the Sotavento company was established in 1997. This partnership created the Sotavento Experimental Wind Farm, a unique facility that has made additional efforts in three areas of investigation, education-information, and training besides having common business objectives such as projects. The company is owned by a consortium of different public and private parties. The wind farm has an installed capacity of 17.56 MW, consisting of 24 wind turbines of 9 different models, belonging to the five manufacturers deployed at that time in Galicia (Gamesa, Made, Ecotecnia, NegMicon, and Bonus).

Three months of data for wind speed, wind direction, and wind power collected from a site turbine will be used to develop a wind power prediction model. Besides these observational data, also WRF data of 4 km resolution for the same period were considered. Figure

3.2 shows the Sotavento wind farm location in Galicia and an image of one site's turbine. In addition, geographical information and gathered data specifications are summarized in table 3.3.

(a) Geographical location (b) Sotavento, Galicia

Fig. 3.2 Geographical location and an image of the Sotavento wind power plant located in Galicia, Spain

Table 3.3 Sotavento site information and data characteristics used in this study

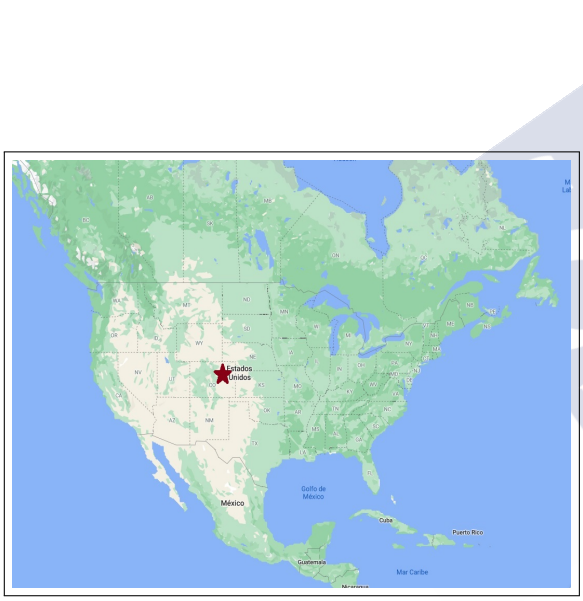
St. Name	Lat.	Lon.	Alt.(m)	Param.	Dur.	Frq.
Sotavento	43.3544	-7.8812	703	Obs.(WS,WD,WP) WRF (WD,WD,T,H,P)	Jan to Dec 2016	H

Nomenclature: St.: station, Lat.: latitude, Lon.: longitude, ALt.:Altitude, Param.: parameter, Dur.: Duration, Frq.: Frequency, Obs.: Observational data, WS: Wind speed, WD: Wind direction, WP: Wind power, T: Temperature, H: Humidity, P: Pressure.

3.2.5 M2 tower, Colorado U.S.

Another case study used in this works is the M2 tower located about 10 kilometers west of Broomfield, Colorado, United States. This site belongs to the America National Wind

Technology Center (NWTC), a world-class research and investigation center managed by the National Renewable Energy Laboratory (NREL) for the Department of energy. This site data are managed and put freely open to the public through the NREL measurement and instrument data center. Station instruments are mounted on a mast of 82 meters. Data are collected with a very high resolution of 2 seconds and is averaged over a minute before being stored. Wind speed and direction have collected at six different heights of 2, 5, 10, 20, 50, 80 meters above the ground level. Other weather data such as temperature, humidity, radiation, isometric pressure, and precipitation are also measuring at various limited heights [89]. Figure 3.3 shows the site's geographical location and a tower image. More information about this site and acquired data is summarized in table 3.4.



(a) Site's location

(b) Site's surrounding

Fig. 3.3 M2 tower location and surrounding - located in Colorado, U.S. [89]

3.2.6 La Haute Borne wind farm

ENGIE-Green company, the leader in wind energy in France, installed and manages La Haute Borne wind farm including four wind generation turbines from Senvion MM82 tech-

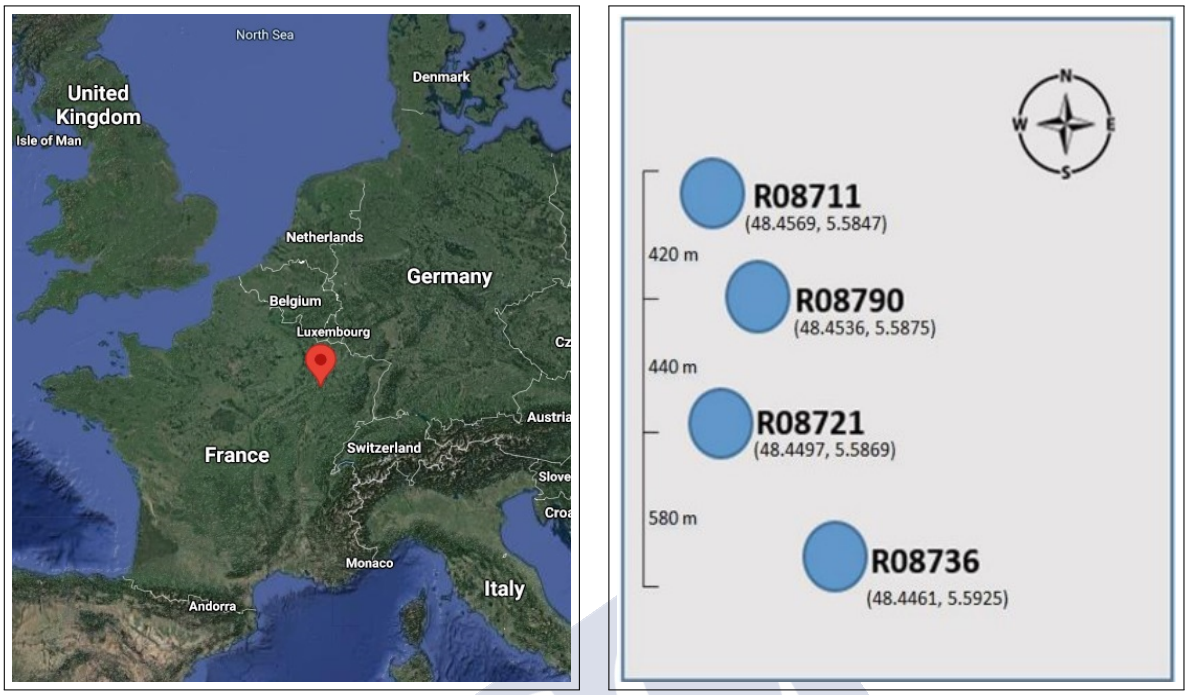
Table 3.4 M2-tower site and observational data characteristics used in this study

Items	Description
Name/Location	M2 tower/ Colorado, U.S.
Latitude	39.91
Longitude	-105.235
Altitude (m)	1855
Parameters	Obs. (WS, P-WS, S-WS, Turb, WD, T, H, P, Prec, Rad)
Duration	1 Jan. - 31 Dec. 2015 - 2018
Time interval	minutal

Nomenclature: WS: Wind speed, P-WS: Wind speed peak, S-WS: Wind shear, Turb: Wind turbulence, WD: Wind direction, T: Temperature, H: Humidity, P: Pressure, Prec: Precipitation, Rad: Solar radiation.

nology, as shown in figure 3.4b. The site is located in the Grand East region in north-eastern France. Usually, 31 different parameters are collected every ten minutes using a SCADA system. Some of these parameters are: active power, wind speed, absolute wind direction, vane position, nacelle angle, gearbox bearing temperature, hub temperature, generator speed, generator bearing temperature, pitch angle, and torque. The average, minimum, maximum, and standard deviation of each parameter are also computed and stored. In this study average values of wind speed, wind direction, and wind power from four turbines, in the period 2017-2020 are applied. The site geographical information and its applied data characteristics study are summarized in table 3.5.

Each of the four wind turbines (R80711, R80780, R80721, and R80736) has a nominal power of 2050 kW, with a diameter of 82m and a hub height of 80m. Cut-in wind speed, nominal wind speed and cut-out wind speed are 3.5 m/s, 14.5 m/s, and 25 m/s, respectively [90]. The geographical location of this wind farm in France and the position of the turbines, with their distances from each other in meters, are shown in figure 3.4.



(a) Geographical location (b) Wind farm layout [90]

Fig. 3.4 La Haute Borne wind farm, Grand East region, France

Table 3.5 La Haute Borne wind farm geographical and data information used in this study

Items	Description
Latitude	48.464644
Longitude	5.584008
Altitude (m)	438
Parameters	Wind speed, wind direction, wind power
Duration	1 Jan. - 31 Dec. 2017 - 2020
Time interval	10-minutes

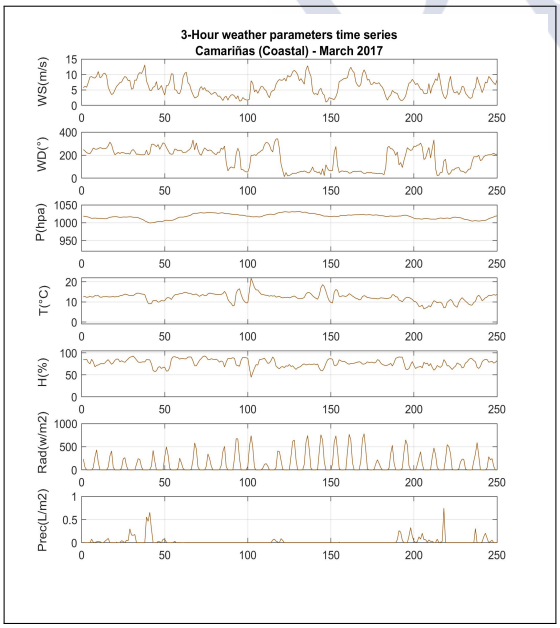
3.3 Case studies datasets analysis

In this section, case study parameters, focusing on wind speed and direction, are visualized and analyzed in different aspects using different techniques. This section also provides the concepts and mathematical formulas behind the analysis, and results with discussions are provided in each subsection. Wind speed and direction data analysis such as fitting the best

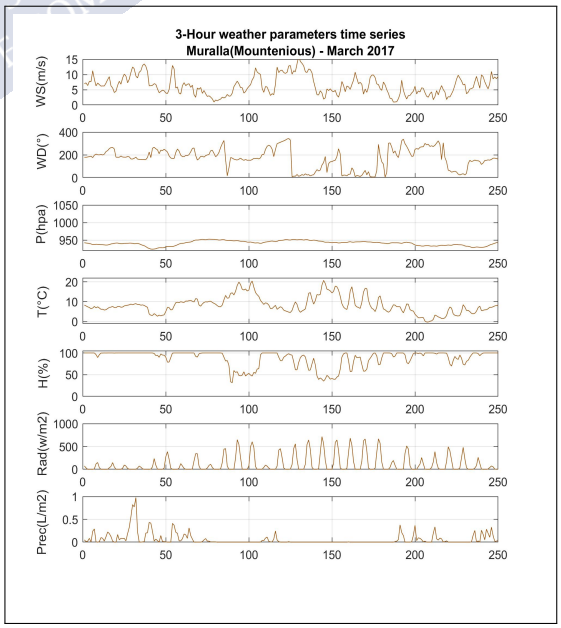
probability distribution to the data, determining cycles as seasonality of wind speeds or direction, and investigating the cross-correlations, are essential factors to assess the wind energy potential.

3.3.1 Time series visualization

Plotting time series helps to get a better insight into a signal characteristics and behavior, especially when plotted jointly to other related parameters. As an example, figure 3.6a shows various meteorological parameters of Santiago-EOAS station during March 2017. In order to capture detailed information simultaneously preventing excessive fluctuations in the signals, a 3-hour moving average window was applied from the original time series with a 10-minutes time interval. Some correlations are easy to detect between the parameters such as humidity with solar radiation: when the maximum radiation occurs at noon, the humidity reached its minimum range and vice versa. Also, during precipitation time, there is a significant drop in the pressure and temperature, but humidity increases to 100%. As another visually detectable information, there are two different wind direction patterns during the first 10-days and the second 10-days periods. About the wind speed, as it can be seen, it does not exceed 10 m/s and has a rough average of about 5 m/s. Figure 3.5 shows the same parameters visualization for Camariñas and Muralla stations.

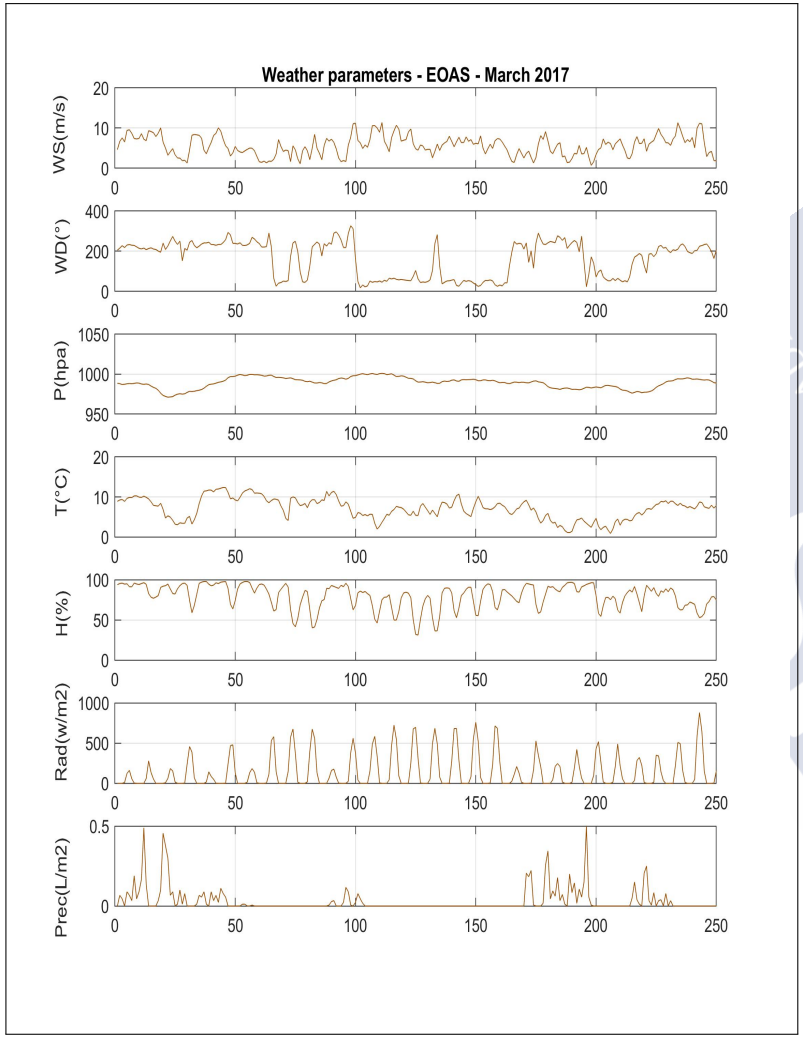


(a) Camariñas (Coastal)

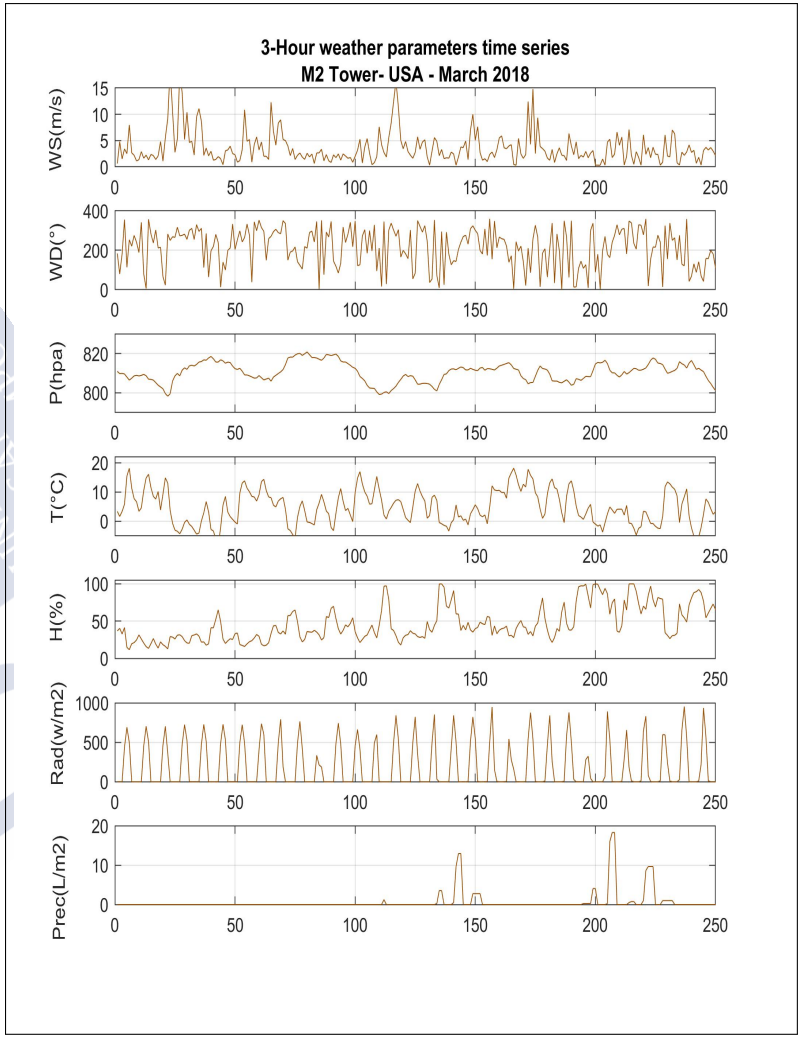


(b) Muralla (Highland rural)

Fig. 3.5 Camariñas and Muralla parameters visualization, March 2017



(a) Santiago-EOAS, March 2017



(b) M2-tower, March 2018

Fig. 3.6 Santiago-EOAS and M2 tower parameters visualization in March 2017 and 2018, respectively

Comparing the visualized parameters of Santiago-EOAS and M2 tower in figure 3.6 reveals the inherent differences between the two geographically different stations. However, considering figures 3.5a, 3.5b, and 3.6a, there is a general pattern for each parameter which is shared between the three stations located in Galicia.

Visual plots are also used to understand the relationship between the parameters better. A Scatter plot is a type of mathematical diagram using Cartesian coordination to compare values against each other. Figure 3.7 shows a scatter plot for seven variables (wind speed (WS), wind direction (WD), humidity (H), temperature (T), pressure (P), solar radiation (Radd), and sky cloud cover (Cfl)) at the Santiago-EOAS station during Spring 2017. The last parameter, Cfl, is not an observational value (obs), but a WRF parameter includes a range of colors. Considering WS-WD-Cfl, it can be inferred that two wind directions (with a difference of almost 180°) are dominant. It means that in the majority of time, there is north-east, south-west wind directions, which is consistent with the wind roses (figures 3.14b, 3.14c, 3.14d) discussed later in this Chapter. Interestingly, the full cloud cover at a low level (up to 2 km from the surface) occurs mostly from the southwest. Comparing precipitation with wind direction, the majority of the precipitation happens given the same wind direction. Considering WS-P-Cfl, there is a strong reverse relationship between the pressure and low-level cloud cover in a way that high pressure happens mostly in a clear sky and vice versa. Also, very high wind speed values happen in low-pressure systems with almost 100% cloud coverage.

Another 2-D scatter plot form of visual relationship comparison is presented in figure 3.8. Wind speed, wind direction, pressure, and temperature parameters are compared in figure 3.8a for the Camariñas station during January and February 2017. likewise, figure 3.8b investigates another combination of parameters, WS-H-P-T, in Muralla during the same time. Below, the inferred relationships are briefly listed:

- Higher wind speeds mainly occur in low-pressure systems in a limited range of temperature with a dominant south-west direction origin. (Camariñas)
- Low temperature mainly occurs in high-pressure systems followed by low wind speed and north-east dominant wind direction. (Camariñas)
- There is a correlation between higher temperature and pressure and lower humidity and wind speed. (Muralla)
- A meaningful relationship exists between higher wind speed and humidity and lower pressure. (Muralla)

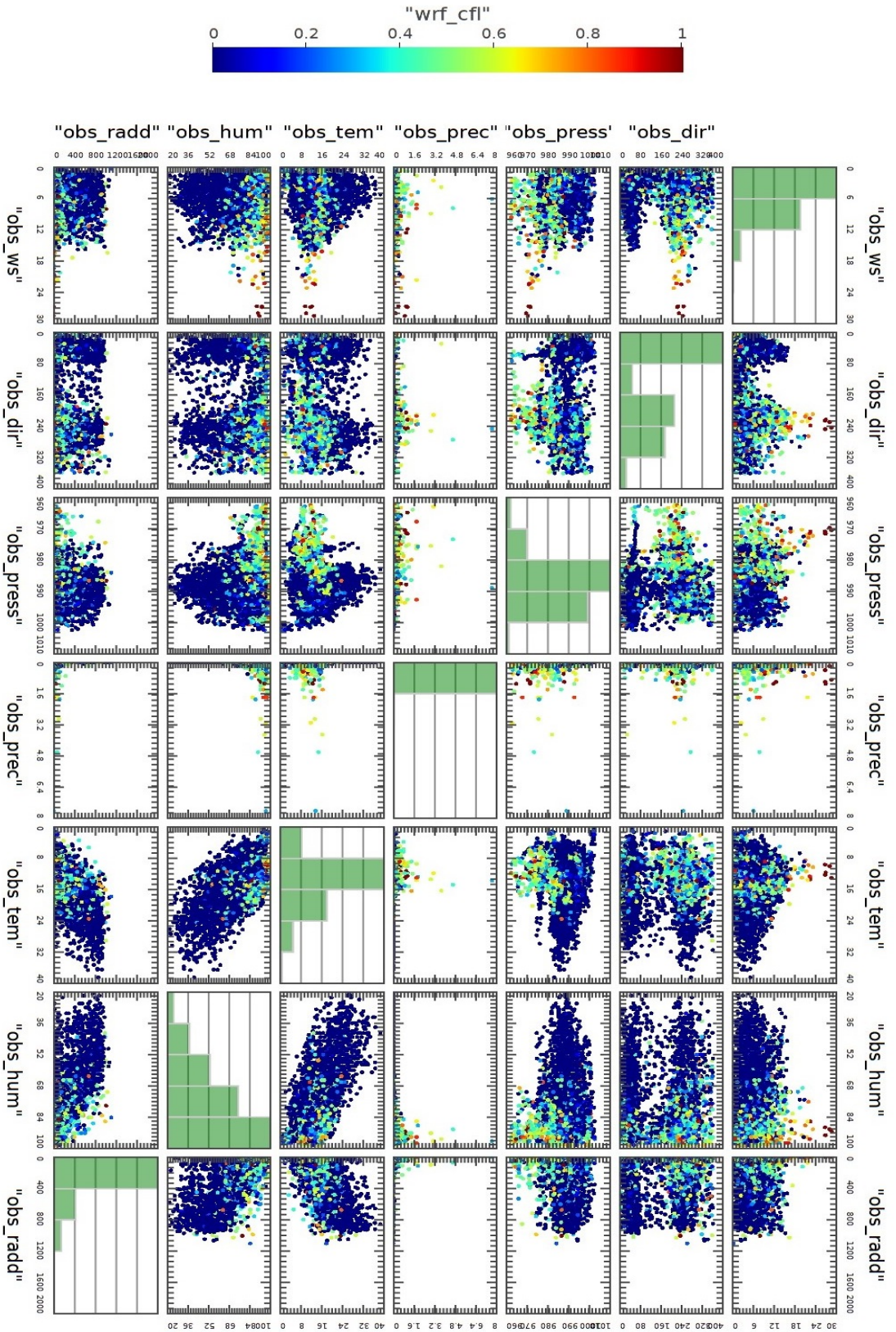
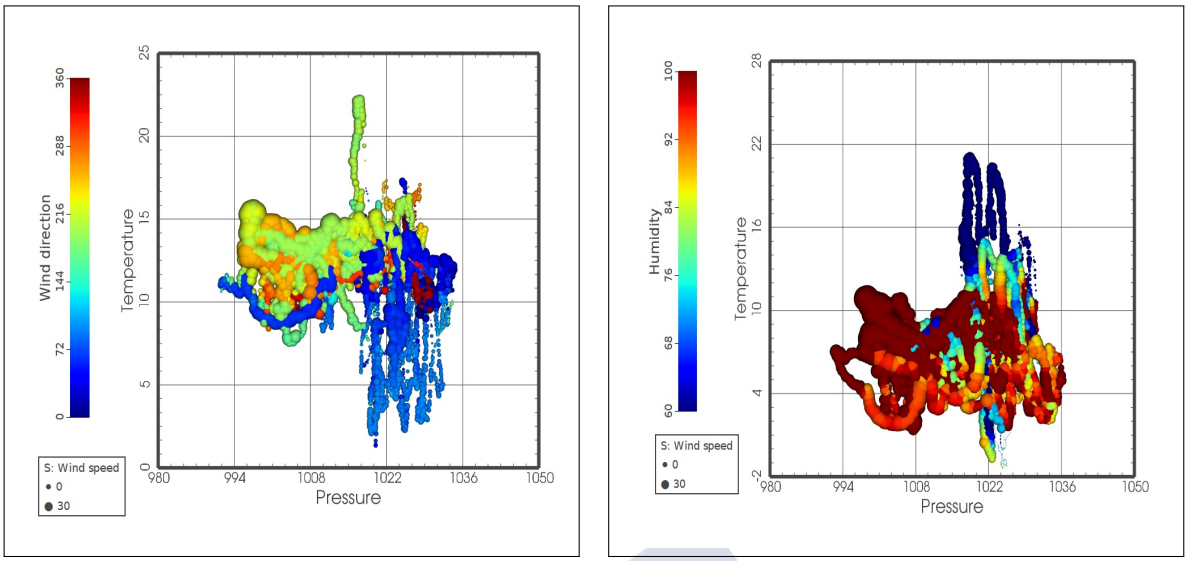


Fig. 3.7 Scatter-plot of weather parameters (WS-WD-H-T-P-Radd-Prec-Cfl) at Santiago-EOAS, Spring 2017



(a) Camariñas, WS-WD-P-T relationship (b) Muralla, WS-H-P-T relationship

Fig. 3.8 Camariñas and Muralla weather parameters relationship, Jan-Feb 2017

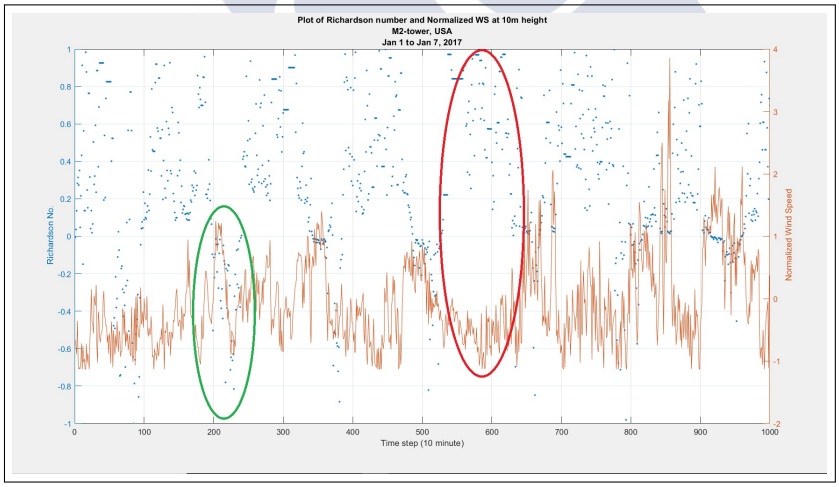


Fig. 3.9 Wind speed and Richardson number relationship, M2-tower, 1-7 January, 2018

As defined in subsection 2.2.5, Richardson number (Ri) is linked to the atmospheric stability. Generally speaking, moving from highly negative to positive Ri numbers, we move from extremely unstable to highly stable atmospheric conditions. As shown in the figure 3.9, there is a direct relationship between wind speed and atmospheric stability. Repeatable patterns of relatively lower wind speed (red ellipsoid) occur in more stable atmospheric conditions, and unstable weather is followed by relatively higher wind speed. It is important to mention that wind speed values are *locally* lower or higher than their next or previous time step neighbors.

Figure 3.10 provides a visual comparison between yearly wind speed at four different meteorological stations (Santiago-EOAS, Camariñas, Muralla, M2 tower).

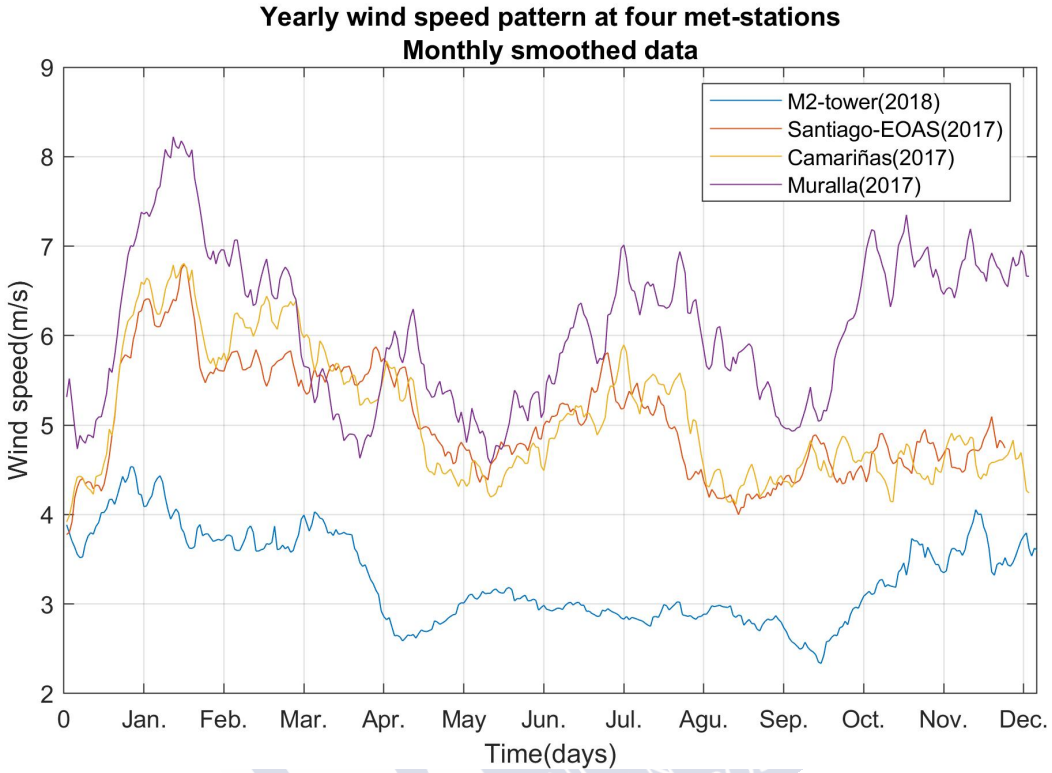


Fig. 3.10 Yearly wind speed comparison of four stations

3.3.2 Weather parameters correlation

As shown visually in the previous part, all the weather parameters are somehow connected as they are part of a whole in a dynamic atmospheric system. Investigating this correlation helps us better understanding how the atmospheric system behaves and affects the wind speed and direction. Besides, more accurate results are produced from the prediction modeling perspective, more correlated inputs feeding to a prediction model.

Correlation is a measure of the relationship between two mathematical variables or measured data values. There are different methods in the literature to measure the correlation. Here, we use normalized cross-correlation function (NCCF) or Pearson correlation, equation 3.1, and autocorrelation function (ACF) to show the relationship of the wind speed with other parameters and with itself. By replacing Y in the cross-correlation function with X, we obtain the autocorrelation function.

$$\rho_{XY}(\tau) = \frac{K_{XY}(\tau)}{\sigma_x \sigma_y} = \frac{E[(X_t - \mu_X)\overline{(Y_{t-\tau} - \mu_Y)}]}{\sigma_X \sigma_Y} \tag{3.1}$$

Where ρ_{XY} is the Pearson correlation coefficient, X, and Y are two parameters time series, σ_x and σ_y are their standard deviations, and μ_X and μ_Y are their mean values, respectively, E is expected value operator and τ and t are the lag and the current time step, respectively. The other form of Pearson correlation for the discrete data or samples is equation 3.2, where mean values of parameters are shown with a hat on top and n is the sample size.

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{3.2}$$

Figure 3.11 shows wind speed auto-correlation considering four different weather stations: Camariñas, Muralla, Santiago-EOAS, and M2 tower. By distancing from the current time, the auto-correlation coefficient decreases significantly. Diurnal patterns can be seen in the figure, which has a higher auto-correlation coefficient compared to their neighbors. Data with ten-minute time intervals during ten days of March 2017 and 2018, respectively, for the Galician (Santiago-EOAS, Camariñas, Muralla) and American (M2 tower) stations are used.

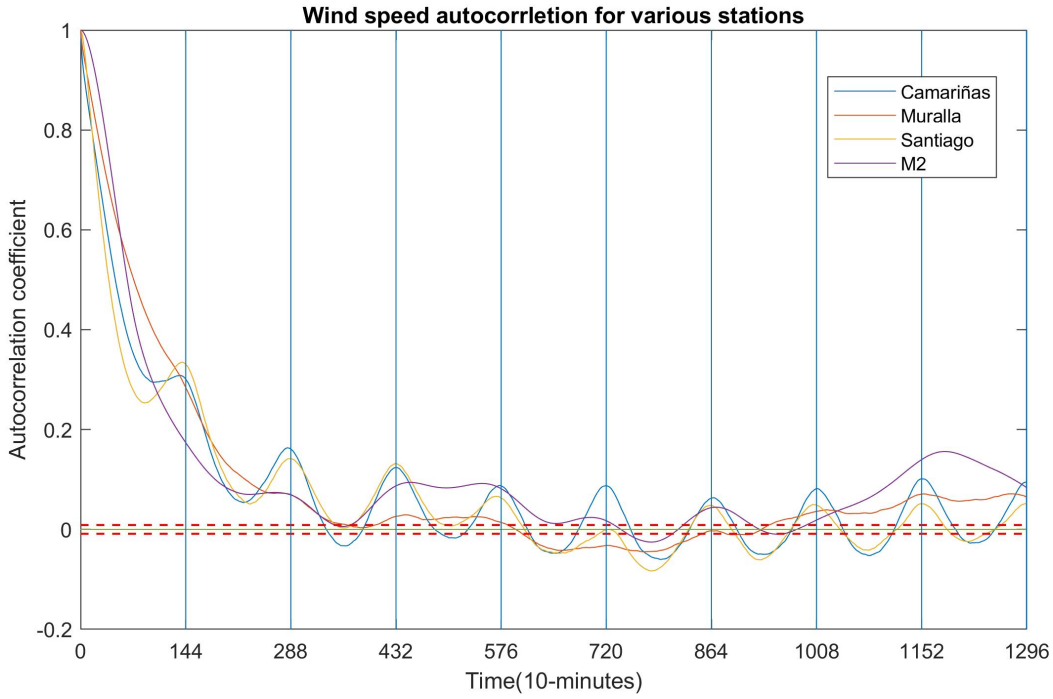


Fig. 3.11 Wind speed auto-correlation at four stations, during 10 days in March

Wind speed correlation for the M2 tower station at different altitudes is demonstrated in figure 3.12 for a 24-hour time horizon. The direct relationship between the wind speed's magnitude and the altitude is apparent. In addition, Lower consistent wind speed is experiencing at midnight and early morning and higher wind speed during the day. As the wind speed changes following a similar pattern at different altitudes, it can be inferred that there is a high correlation between different heights at the same time.

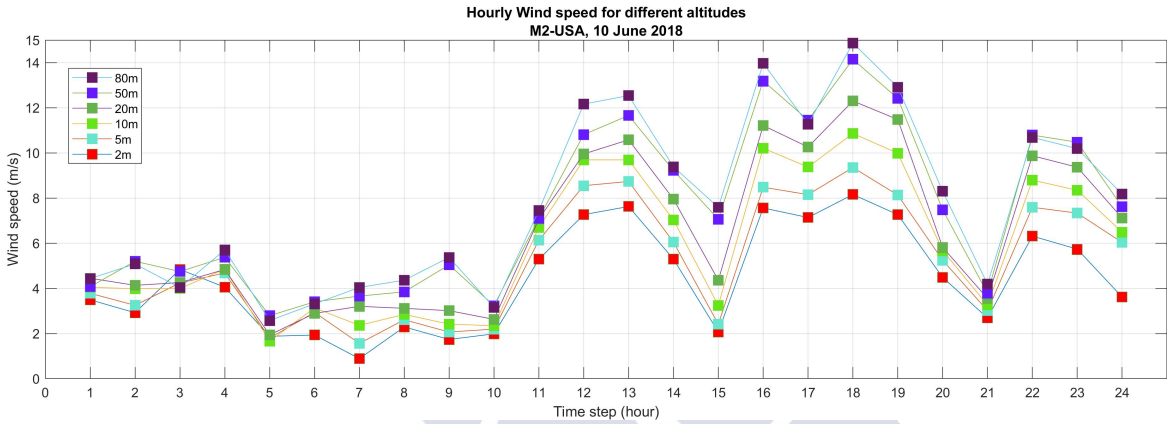


Fig. 3.12 M2-tower Wind speed at different heights

Figure 3.13 shows the reverse relationship between turbulence intensity and altitude for different seasons. Generally speaking, turbulence intensity decreases significantly by increasing the altitude with a bigger rate between 0-5 m and 20-50 meter height. Considering 10 meter height as standard altitude, the fall season has the highest turbulence intensity, followed by summer, spring, and winter.

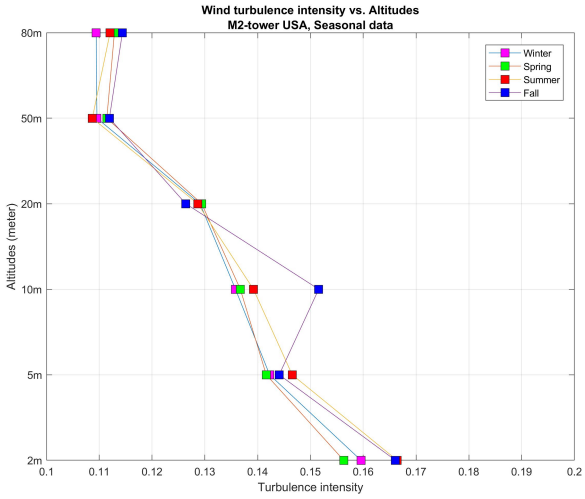


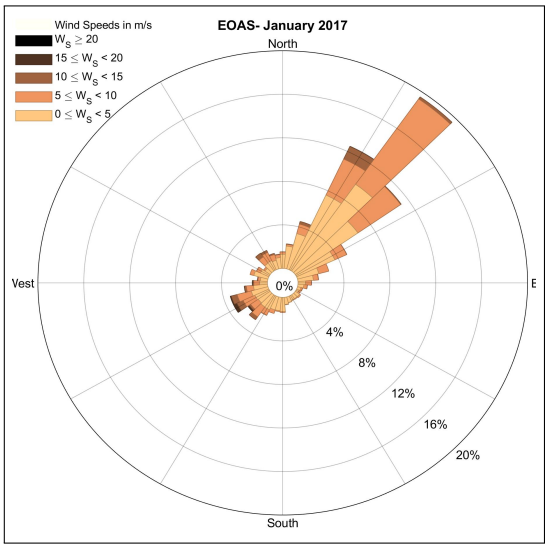
Fig. 3.13 M2-tower altitude vs. turbulence intensity diagram

3.3.3 Wind roses

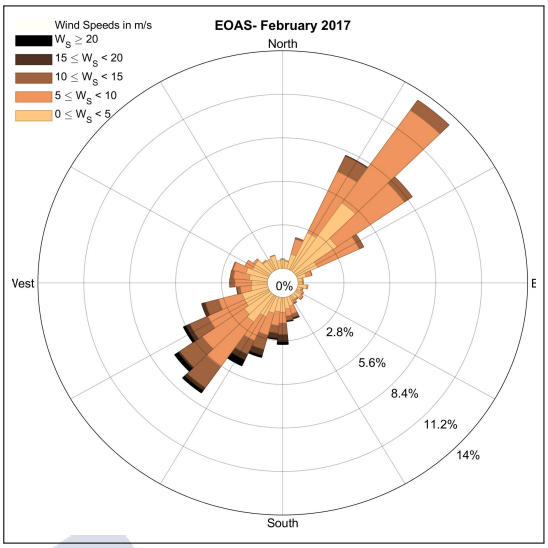
Wind roses are a helpful tool to show the distribution and frequency of the wind speed and direction. The percentage values shown by internal circles represent the probability of wind coming from that direction. It is always better to make an average of several years before depicting a rose diagram as the wind patterns may change from year to year. This may apply when a wind resources assessment is required. A shorter period for wind rose is also applicable. In the following, wind roses for selected stations and time horizons are depicted.

Monthly wind rose diagrams for the Santiago-EOAS station are shown in figures 3.14 and 3.15. As it can be seen, the two dominant wind directions are either from southwest or north-east with a slight dominance of the latter one. However, the strong winds above 15 m/s have blown from the southwest direction. February has recorded the highest wind speed among the other months. Wind speeds between 5 to 10 m/s are the most frequent range during all months. The periodicity between the southwest and north-east wind directions is consistent with the Ekman transport periodicity proposed by Alvarez and Descasto in [91], where seasonal fluctuations between 50 to 80 days is observed. These periodicities at the A Coruña harbor and the open ocean have been previously described by Blanton et al. [92] using wind stress.

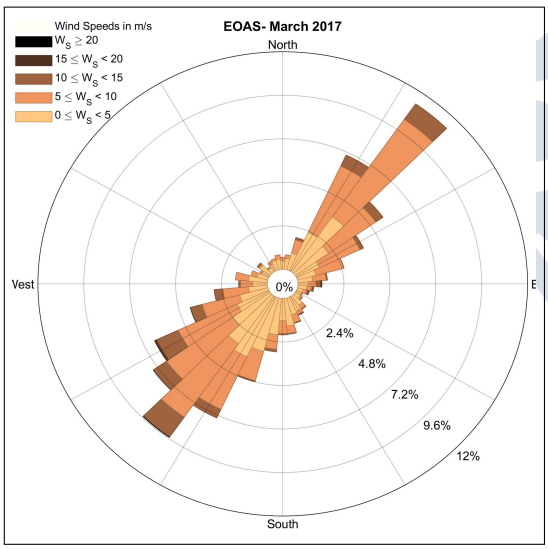
As another example, yearly wind roses at the M2-tower station for different heights are shown in figure 3.16. The wind shear effect can be well observed in this figure. The average and particularly the wind speed are slightly changing by increasing heights, as expected. As another fact well represented in the figure, the wind is more dispersed in the lower heights, which should be an effect of surface roughness and, also, possible obstacles and barriers. In other words, wind direction changes less in higher levels because of lower surface roughness effect and less obstacles flow crashing, contrary to the low-level winds. The dominant wind blows from south southeast (SSE) direction. However, comparing to the Santiago-EOAS, which wind pattern is strictly adhered to two specific directions, in the M2 tower site, the wind is frequently blowing from all directions. The longest spoke shows the wind blew from the SSE at speeds between 0-5 m/s about 4.5% of the time (light blue), 5-10 m/s about 2.5% of the time (dark blue), 10-15 m/s about 0.5% of the time (orange), between 15-20 m/s about less than 0.1% of the time (red) and for more than 20 m/s it is about 0% of the time (dark brown).



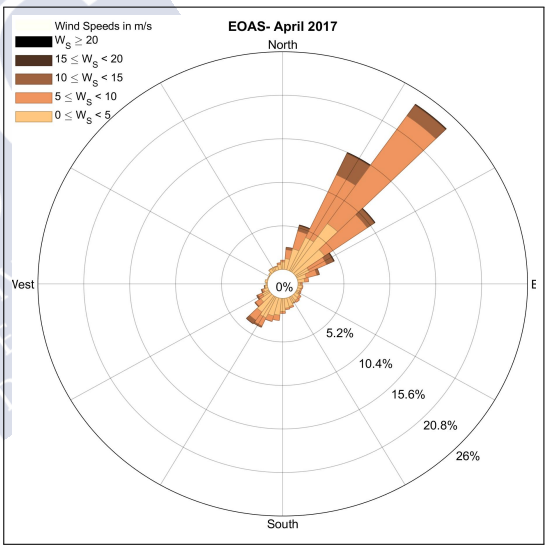
(a) January



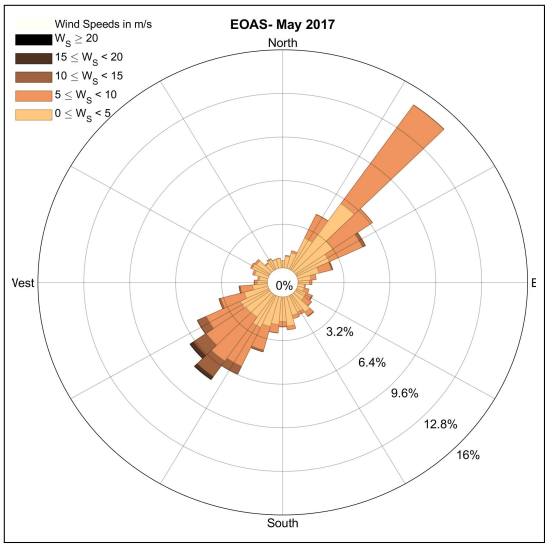
(b) February



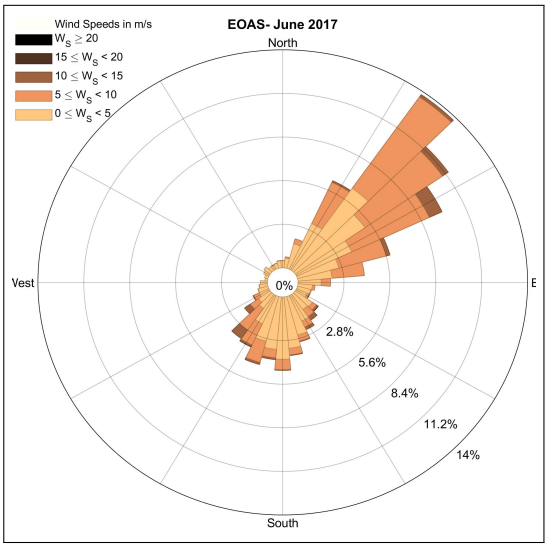
(c) March



(d) April

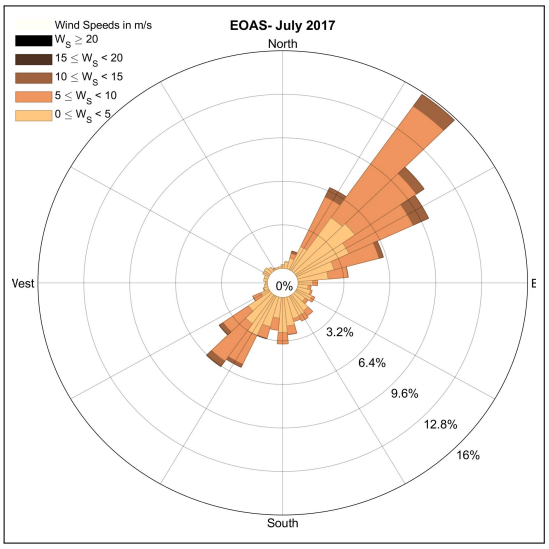


(e) May

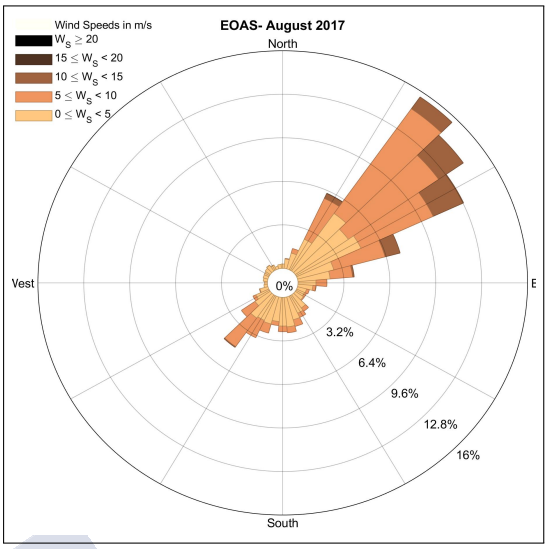


(f) June

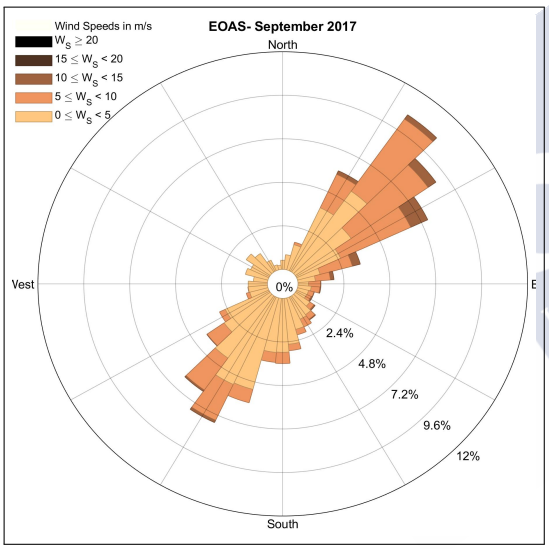
Fig. 3.14 Santiago-EOAS monthly wind roses from January to June 2017



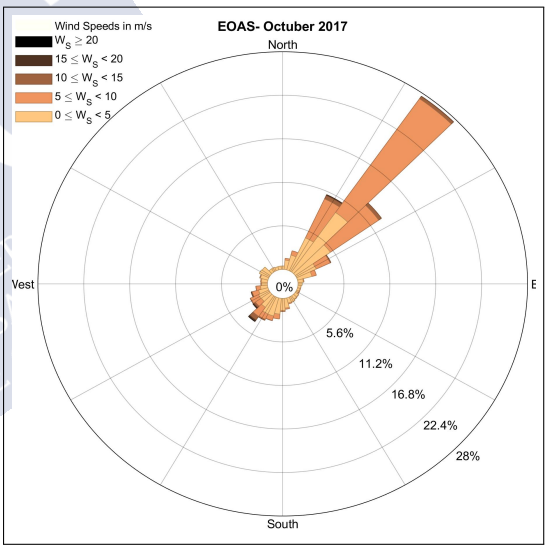
(a) July



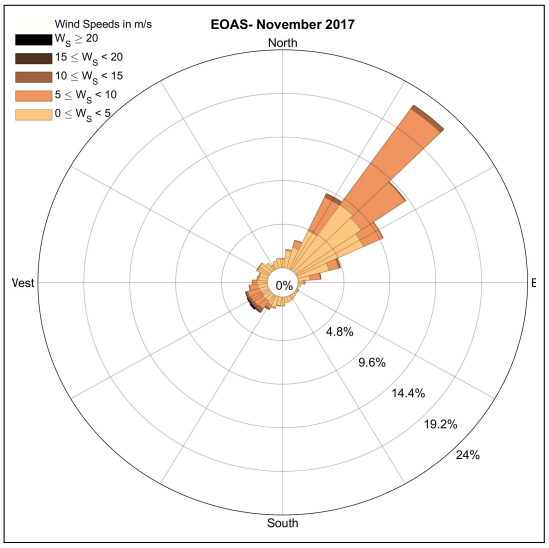
(b) August



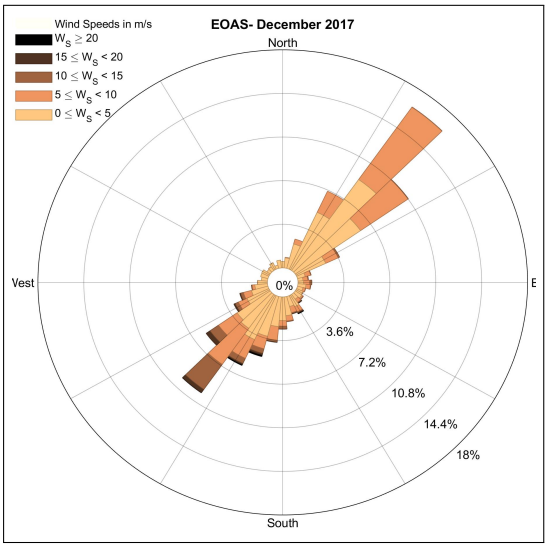
(c) September



(d) Octuber

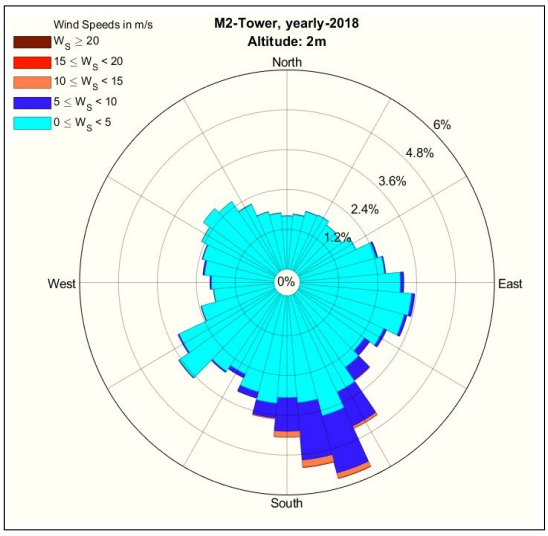


(e) November

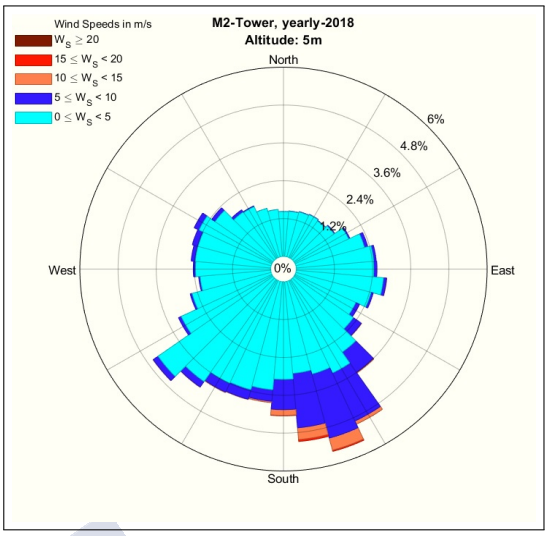


(f) December

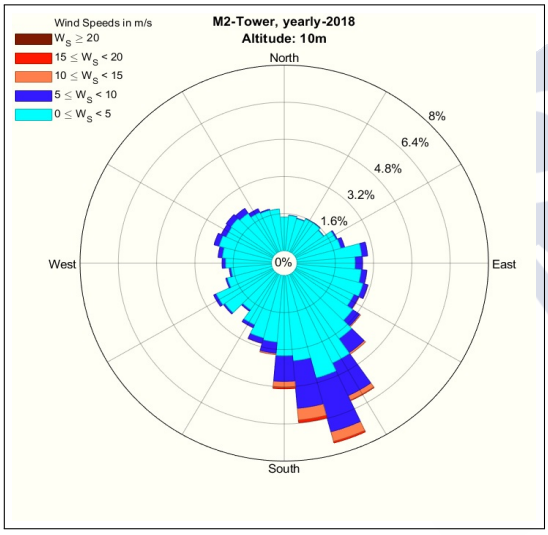
Fig. 3.15 Santiago-EOAS monthly wind roses from July to December 2017



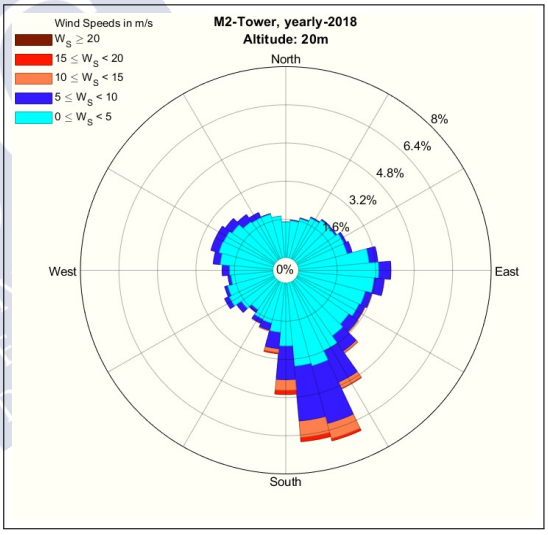
(a) 2m



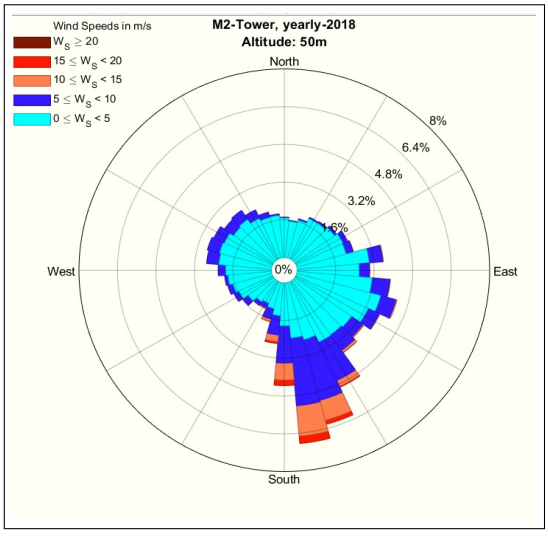
(b) 5m



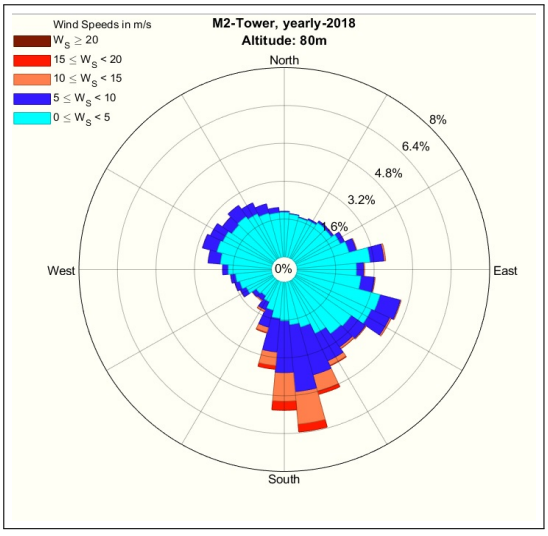
(c) 10m



(d) 20m



(e) 50m



(f) 80m

Fig. 3.16 M2-tower yearly wind roses at different height: a)2m, b)5m, c)10m, d)20m, e)50m, f)80m, 1 Jan-31 Dec 2018

3.3.4 Time series moving average analysis

Moving average is used to smooth time series and to put more focus on the trend or cycles patterns of the signal by eliminating fluctuations. A simple moving average (SMA) technique is used here, as described in equation 3.3. SMA is an unweighted mean of the previous k observations which based on definition could be placed at the beginning, end or in the middle of averaging period. In meteorological application, the mean value normally places at the end of the window. Considering a time series of n data-points as $p_1, p_2, ..., p_n$, the mean over the last k observations is defined as follows:

$$SMA_k = \frac{p_{n-k+1} + p_{n-k+2}, \cdots p_n}{k} = \frac{1}{k} \sum_{i=n-k+1}^{n} p_i \tag{3.3}$$

The visual effect of moving average or rollmeans technique is shown in figure 3.17. As it can be seen, by expanding the averaging window from 10 minutes to 1 and 3 hours, the short-term fluctuations are smoothed out, and longer-term trends or cycles are shown better. As it is evident from the trend, the mean wind speed is increasing from the first week of January, and a weekly pattern or cycle is evident.

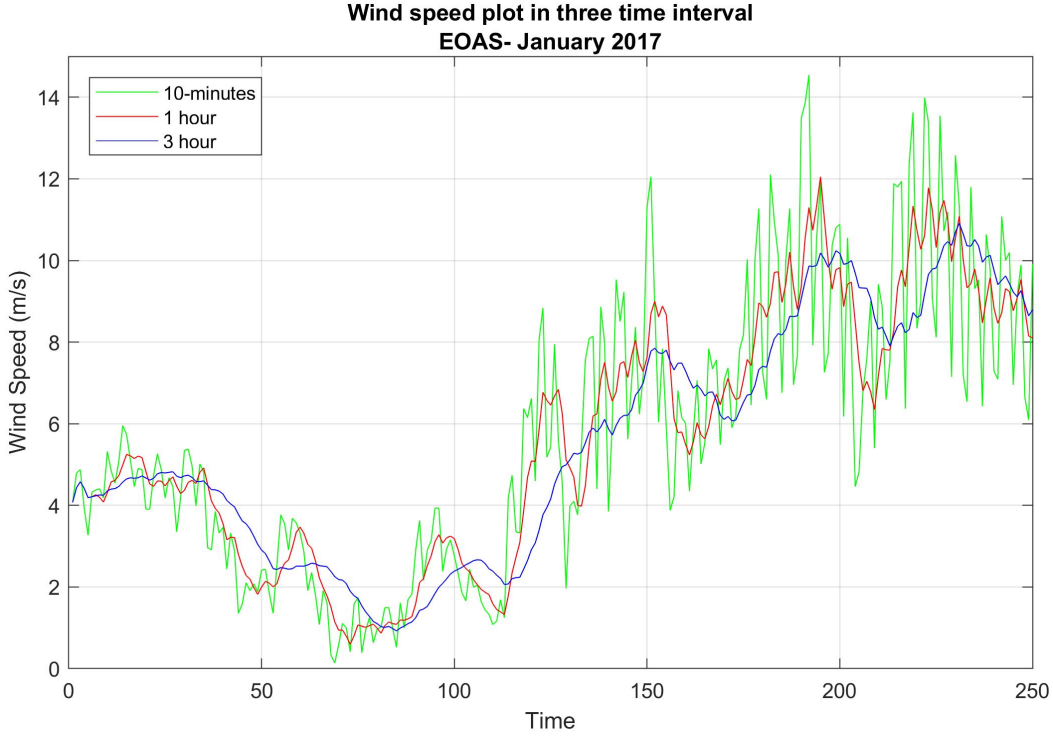


Fig. 3.17 EOAS-Santiago wind speed visualization for different data time step: 10-minutes, 1-hour, and 3-hours

3.3.5 Wind speed distribution

The wind speed distribution of a specific location during a period, usually annual, determines the wind energy systems. Fitting the best distribution function to data and drawing distribution functions is necessary for wind energy potential assessment of a region. A variety of probability density functions are used in the literature to describe wind speed frequency distributions. [93]. In a general view, there are two groups of probability density functions: parametric and non-parametric. Parametric distributions are divided into two main groups: one-component probability density function (pdf) and mixture pdf. Weibull [9], Gamma [94], and Rayleigh [95] distribution functions, as the widely used techniques in the wind power energy sector, are introduced and applied below.

In a primary comparative context, figure 3.18 shows the histogram of annual wind speed at four meteorological stations. It can be seen that wind energy production potential at Muralla, as a highland rural location, is higher than the other coastal and urban locations in Galicia, and the M2 tower site in Colorado.

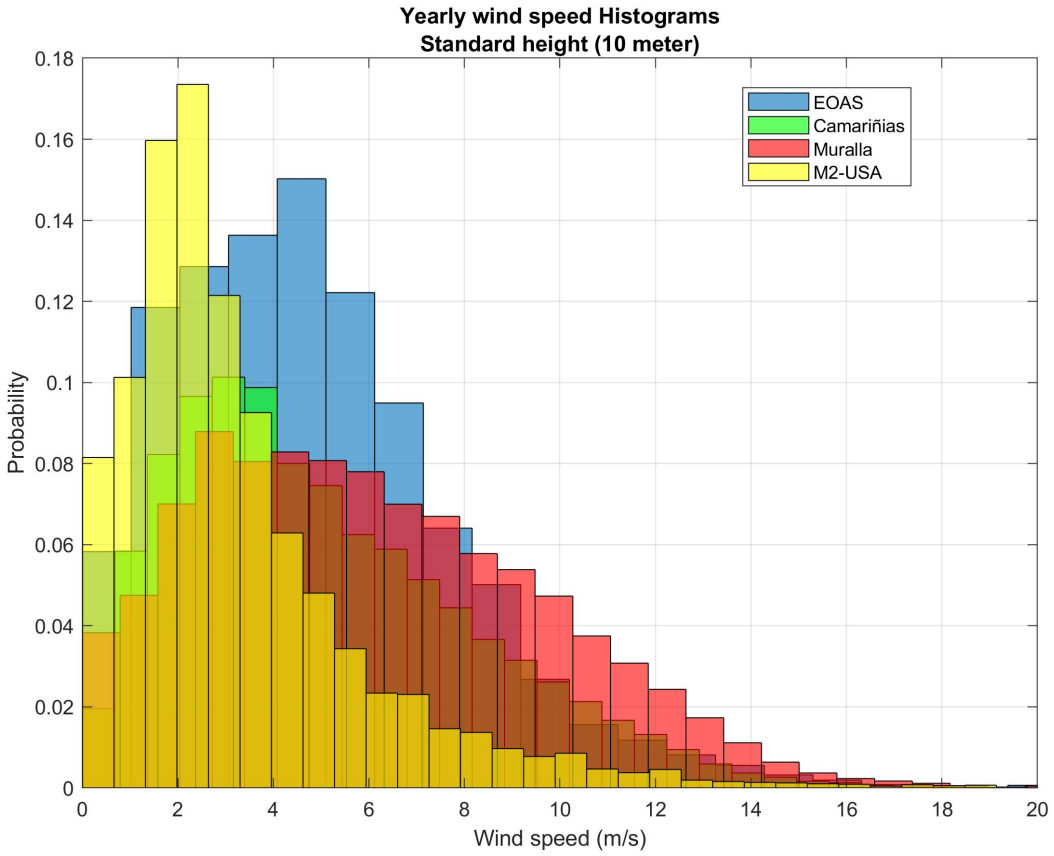


Fig. 3.18 Yearly wind speed probability distribution at four case studies

3.3.5.1 Weibull, Gamma and Rayleigh distribution functions

Weibull distribution is related to a number of other distributions which interpolates between exponential distribution ($\lambda = 1$) and Rayleigh distribution ($\lambda = 2, k = \sqrt{2\sigma}$), where $\lambda > 0$ is scale parameter and $k > 0$ is shape parameter of distribution.

$$f(x;k,\lambda) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} exp\left(-(\frac{x}{\lambda})^{k}\right) \tag{3.4}$$

By substituting ($\lambda = 2, k = \sqrt{2\sigma}$) in the Weibull distribution, we obtain Rayleigh distribution function as follows:

$$f(x;\sigma) = \frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}} \tag{3.5}$$

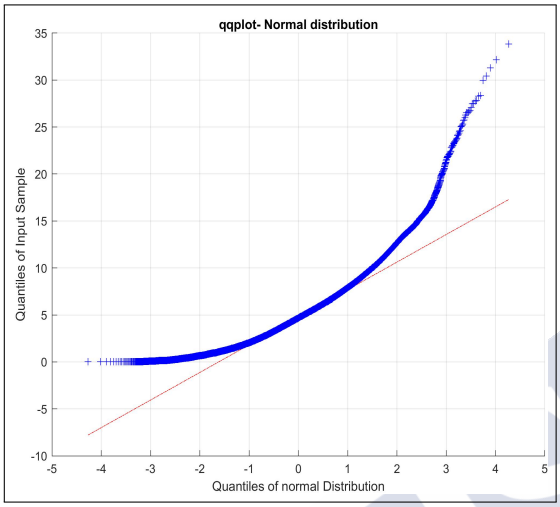
The following equation represents Gamma probability distribution, where β is the shape parameter, and ζ is the scale parameter.

$$f(x;\zeta,\beta) = \frac{x^{\zeta-1}}{\beta^{\zeta}\Gamma(\zeta)} exp(-\frac{x}{\beta}), \quad x,\zeta,\beta > 0 \tag{3.6}$$

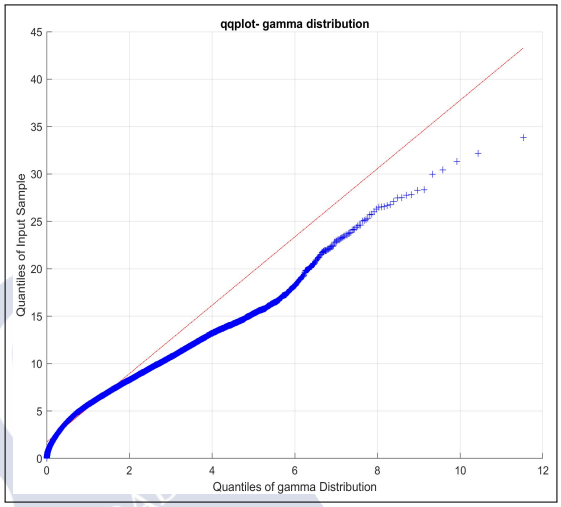
In order to have an idea about which model may fit better to wind speed data, quantile-quantile (QQ) plot [96] is a useful tool to be applied for different distribution functions. QQ-plot is a graphical tool to help us assess if a set of data plausibly came from some theoretical distribution such as a Normal or exponential. It is just a visual check, not a mathematical proof. However, it allows us to see rapidly if our assumption is appropriate and, in case of not, how the assumption is violated and what data values have caused this violation. As an example, figure 3.19 shows QQ-plots of four distribution functions applied on Santiago-EOAS yearly wind speed. As shown, Gamma and Weibull distributions appear to be fairly safe assumptions compare to Rayleigh and normal distributions. The slight downward distortion at the middle shows more concentration in the real data than the theoretical distribution would suppose. Considering the Weibull function, a light-tailed can be seen in the data. Figure 3.20 shows plotted distribution on the histogram of data.

In figure 3.21, the Weibull function is applied to the monthly-based data of Muralla station. As it can be seen, the function fits perfectly, following the linear data for some months like February, July, and October. For the rest of the months, the assumption is reasonably

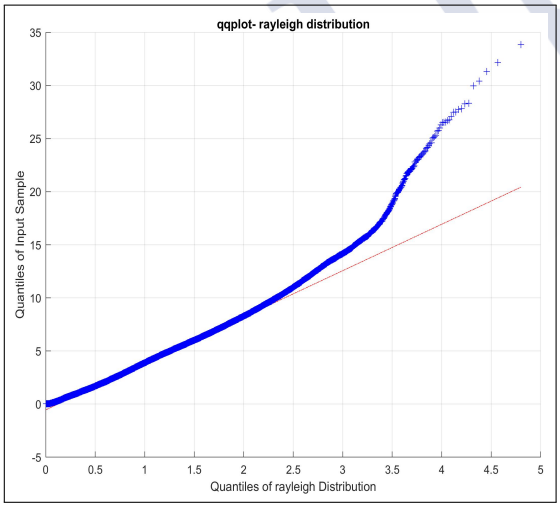
acceptable, and in some cases like January, maybe it is better to look for alternatives. The less matched first quantile shows that there is more concentration in data than Weibull is assuming. However, in order to apply a unique function for all months and due to the good performance of Weibull distribution on yearly data and most of the months, we just apply this function as the best-fitted function on our data.



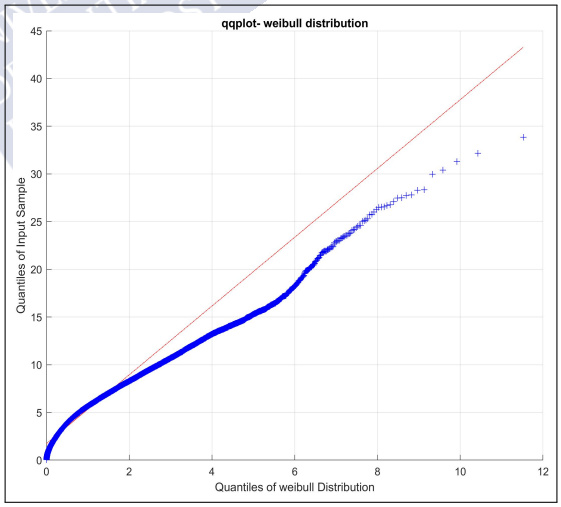
(a) Normal distribution



(b) Gama function



(c) Rayleigh distribution



(d) Weibull distribution

Fig. 3.19 QQ-plots of four commonly used distribution functions on Santiago-EOAS yearly data: Normal, Gamma, Rayleigh and Weibull

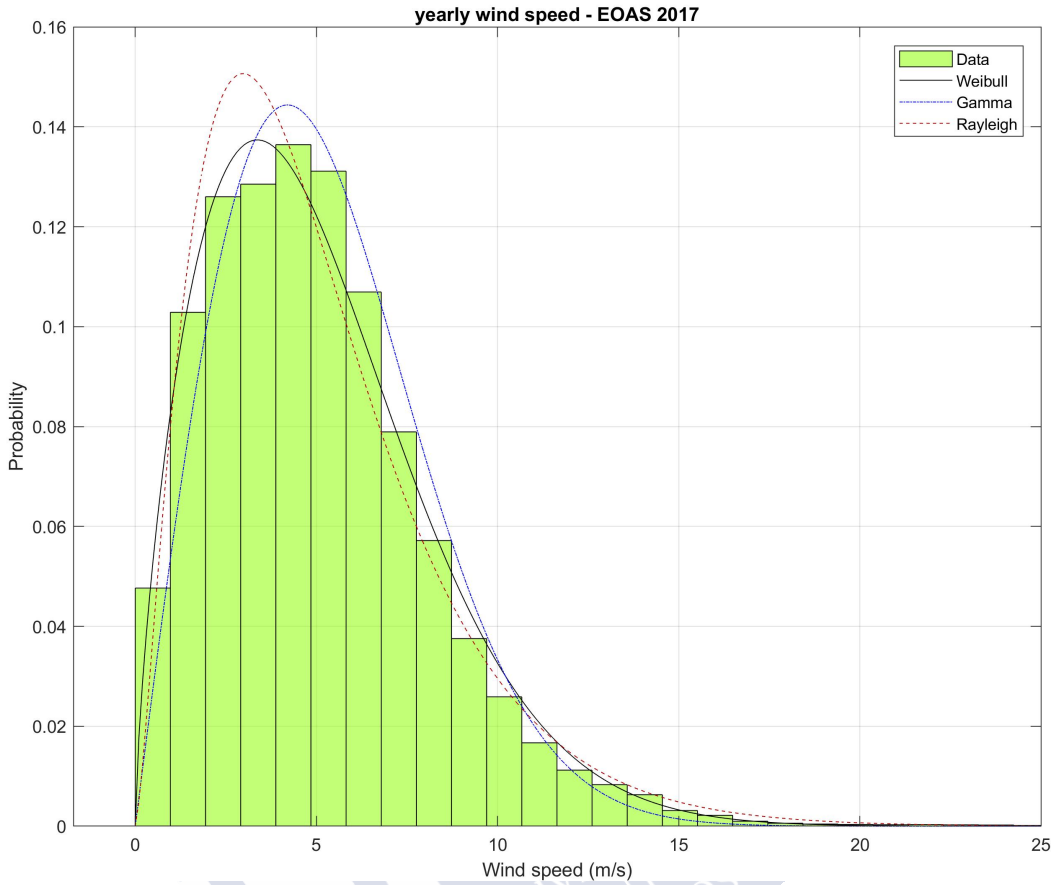
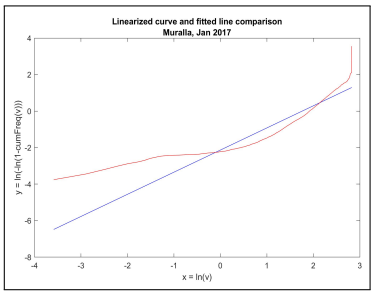
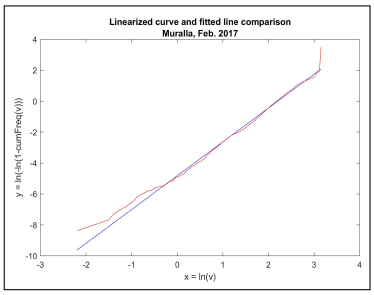


Fig. 3.20 EOAS-yearly wind speed histogram and imposed distributions: Weibull, Gamma, and Rayleigh

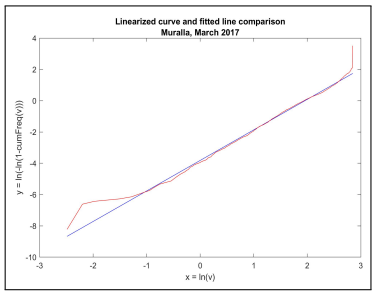
In table 3.6 are included the parameters of Weibull distribution, k and λ, shape and scale parameters respectively, for monthly-based wind speed data of four meteorological stations: Santiago-EOAS, Camariñas, Muralla, and M2 tower at 10 meters height. The yearly average amount of each parameter is also provided in the last column used for fitting distribution on annual-based data.



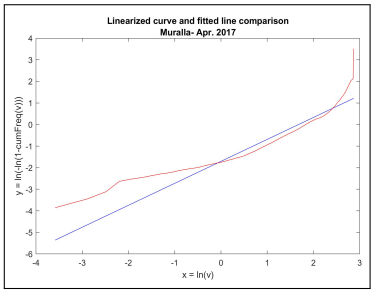
(a) Jan.



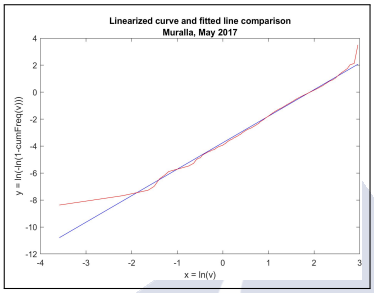
(b) Feb.



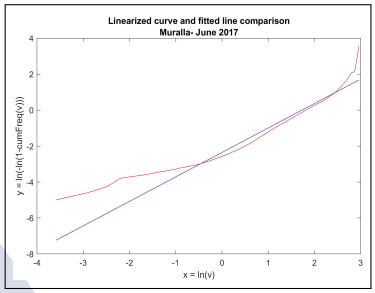
(c) Mar.



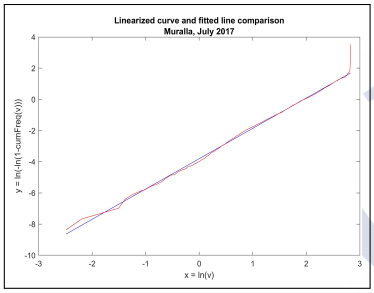
(d) Apr.



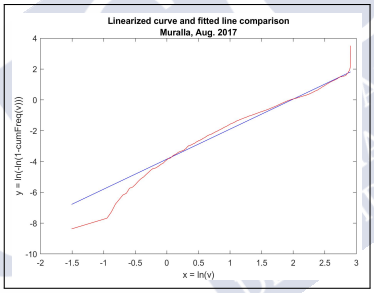
(e) May.



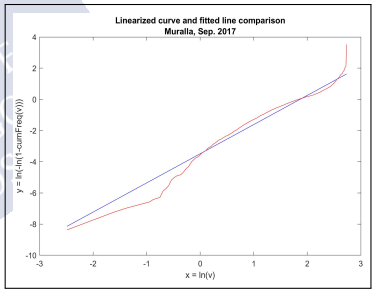
(f) Jun.



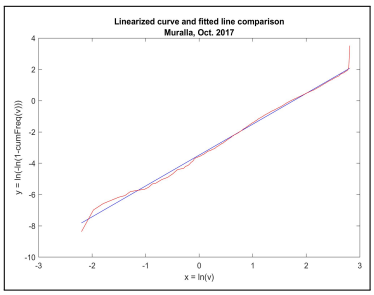
(g) Jul.



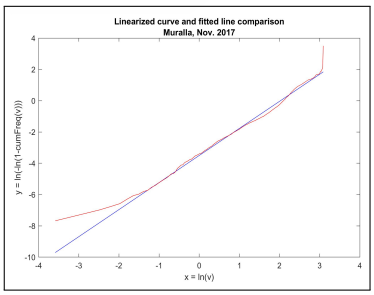
(h) Aug.



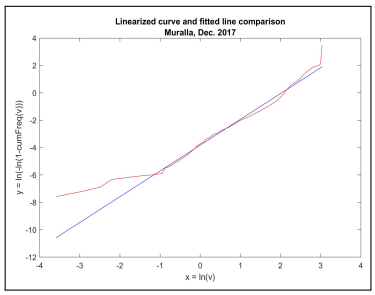
(i) Sep.



(j) Oct.



(k) Nov.



(l) Dec.

Fig. 3.21 QQ-plots of applied weibull distribution function applied on monthly-based Muralla data

Table 3.6 Weibull distributions' parameters for monthly wind speed at different stations

Stations & dist. parameters	**Jan.**	**Feb.**	**Mar.**	**Apr.**	**May.**	**Jun.**	**Jul.**	**Aug.**	**Sep.**	**Oct.**	**Nov.**	**Dec.**	**Annual average**
Santiago-EOAS													
$V(m/s)$	4.60	6.63	5.56	5.65	4.98	5.17	5.12	5.24	4.03	4.89	4.60	4.86	5.10
$k(-)$	1.600	1.689	1.852	1.749	1.635	1.673	1.850	1.671	1.590	1.876	1.632	1.476	1.691
$\lambda(m/s)$	5.178	7.598	6.166	6.336	5.507	5.712	5.675	5.722	4.466	5.546	5.158	5.489	5.713
Camariñas													
$V(m/s)$	4.85	6.66	5.87	5.74	5.18	5.11	5.19	5.48	4.17	4.62	4.59	4.60	5.17
$k(-)$	1.932	1.918	2.097	1.875	1.784	1.862	1.863	1.776	1.956	1.603	1.576	1.649	1.779
$\lambda(m/s)$	5.677	7.424	6.516	6.538	5.761	5.815	5.812	6.071	4.779	5.159	5.195	5.144	5.824
Muralla													
$V(m/s)$	5.54	7.97	6.29	5.14	5.99	5.35	6.31	6.38	5.71	5.10	6.92	6.94	6.13
$k(-)$	1.213	2.187	1.950	1.018	1.964	1.361	1.945	1.948	1.872	1.973	1.730	1.883	1.754
$\lambda(m/s)$	5.806	9.034	7.070	5.349	6.732	5.634	7.066	7.203	6.444	5.822	7.569	7.589	6.777
M2-Tower(10m)													
$V(m/s)$	4.13	3.64	3.77	3.82	2.64	3.18	3.02	2.86	2.85	2.41	3.22	4.00	3.29
$k(-)$	1.752	1.441	1.564	1.607	2.413	1.876	2.216	2.004	1.741	1.502	1.273	1.215	1.717
$\lambda(m/s)$	4.522	3.871	4.052	4.134	2.947	3.515	3.365	3.170	3.178	2.659	3.482	4.128	3.585

3.3.6 Spectral Analysis

Since analyzing wind speed in time-scale of several weeks there is no evidence on any strong periodicity or seasonality, we use spectral analysis [10] can be used to reveal such information by depicting power spectra for time series using Fast Fourier Transform (FFT), which mathematical formula is presenting below:

$$X(f,T) = \int_0^{T0} x(t)e^{(2\pi fti)dt} \tag{3.7}$$

Where x is a random variable in the time domain, f is frequency, X is complex Fourier components in the frequency domain, t is elapsed time, i is complex number unity, and T is time sequence.

Figure 3.22 is an example of power spectral density estimation using FFT for Santiago-EOAS yearly wind speed data. The results are depicted in a diagram with the normal axis. As it can be obviously seen, there is a high value for power density at one cycle/day, which shows diurnal periodicity in wind speed data.

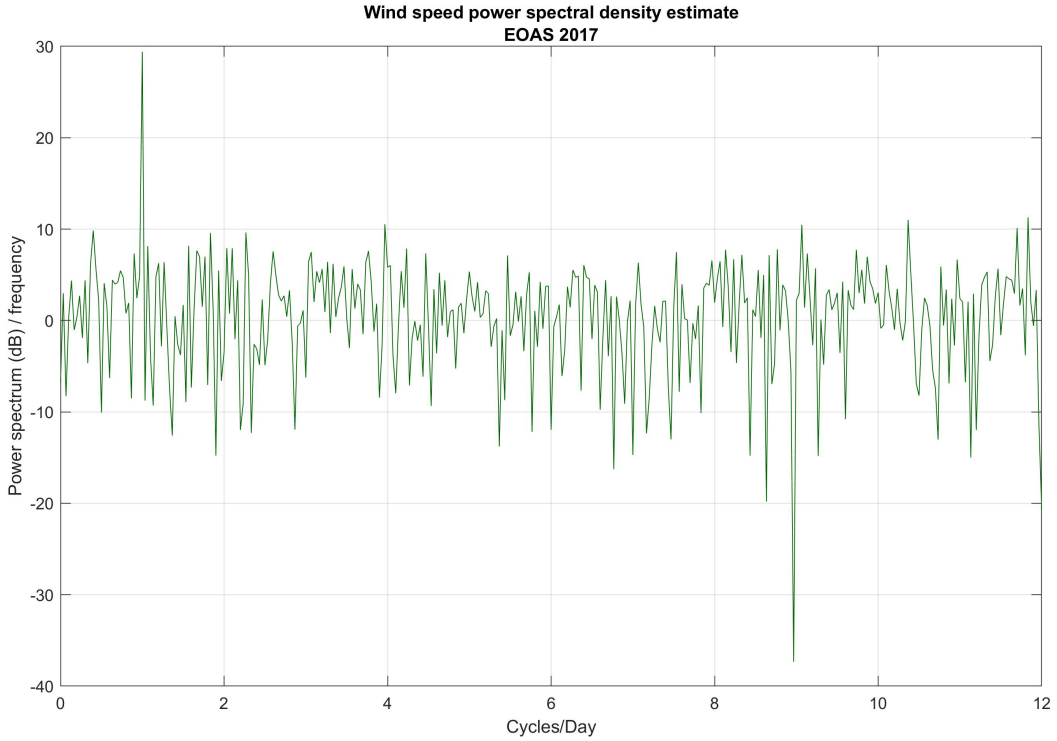
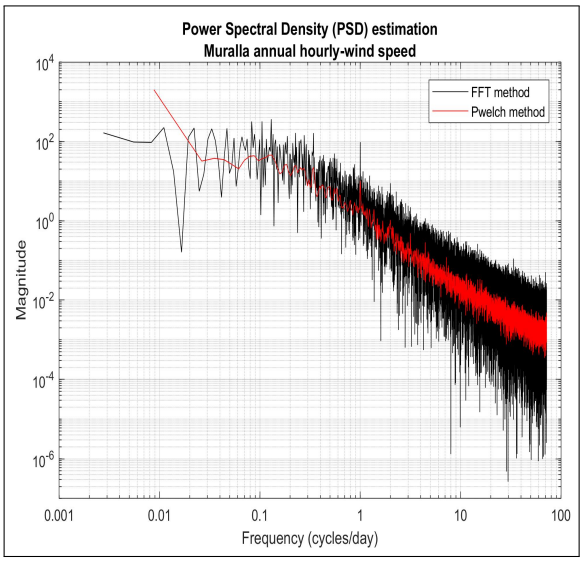
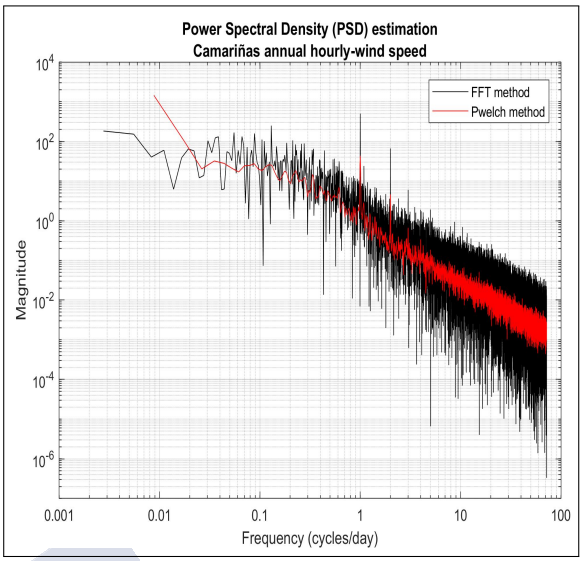


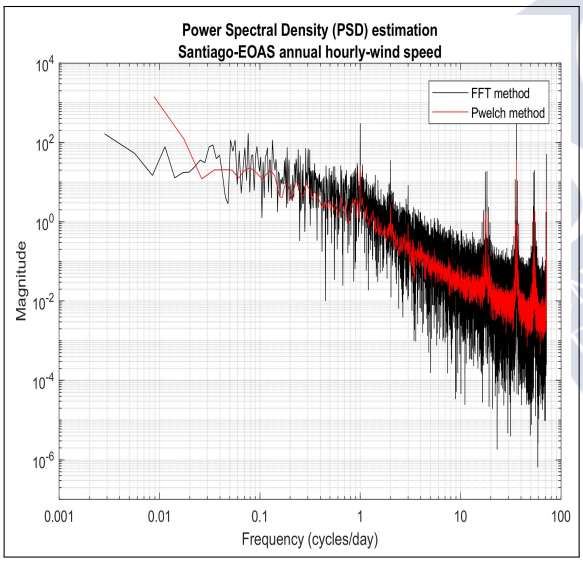
Fig. 3.22 Power spectral density estimation for Santiago-EOAS yearly wind speed



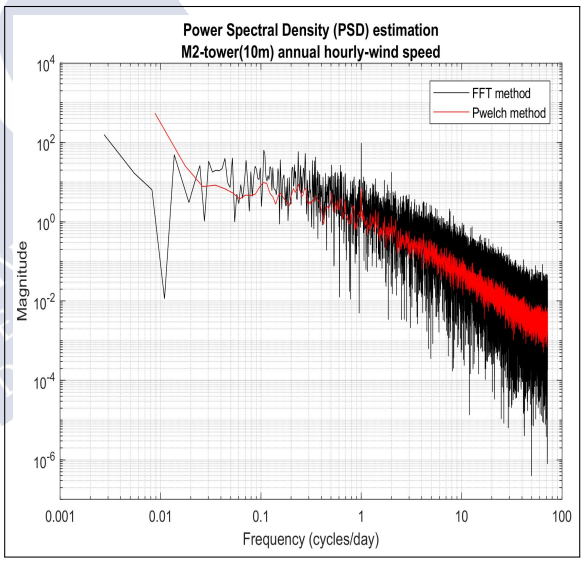
(a) Muralla



(b) Camariñas



(c) Santiago-EOAS



(d) M2-Tower

Fig. 3.23 Welch/FFT power spectral density estimation for wind speed at Santiago-EOAS, Camariñas, Muralla, and M2-tower(10m)

Using Fast Fourier Transform (FFT) on pure mathematical functions it is observed as deterministic signals performs satisfactorily. However, applying FFT on a real signal like the wind as a stochastic time series, with highly noise component, is not the best choice as this noise is not interesting in analysing the wind spectrum. In order to get rid out of the noise, an averaging algorithm like Welch, provided by Peter D. Welch, is helpful. The method slices the original signal into several pieces and averages their spectra. By shortening the signal to which FFT is applying and by averaging, the spectra are low-passed, and the peaks are not

so narrow as the price to pay, avoiding noise reflection in the power spectrum. Equations 3.8-3.10, formulate the Welch method and represents the power spectral density estimation by this method [97]:

$$\hat{S}_x^W(w_k) \triangleq \frac{1}{K}\sum_{m=0}^{K-1} P_{x_m,M}(w_k) \tag{3.8}$$

$$P_{x_m,M}(w_k) = \frac{1}{M}|FFT_{N,k}(x_m)|^2 \triangleq \frac{1}{M}\left|\sum_{n=0}^{N-1} x_m(n)e^{-j2\pi nk/N}\right|^2 \tag{3.9}$$

$$x_m(n) \triangleq w(n)x(n+mR), \qquad n = 1,2,...M-1, m = 1,2,...K-1 \tag{3.10}$$

Where $x_m(n)$ is zero-padded frame from the signal x, $P_{x_m,M}(w_k)$ is periodogram of the mth block, $\hat{S}_x^W(w_k)$ represents the Welch estimate of the power spectral density, R is the window hop size, K represents the number of available frames, and $w(n)$ is the rectangular window.

Figure 3.23 show a comparison between the FFT and Welch method for power spectral density estimation of annual hourly-averaged wind speed signals at four stations: Santiago-EOAS, Camariñas, Muralla, and M2-tower(10m). As it was shown, the Welch method has less noise. One cycle/day has the highest magnitude in all plots, which confirms diurnal periodicity in wind speed data. In Santiago-EOAS, 24 cycles/day is also among the high-density values, which show hourly periodicity in data.

3.4 Modeling environment

The two main software used for modeling and data analysis in this work is Matlab R2017b and R-3.6.6. Data analysis is mainly carried out in the R software. However, all the prediction models are coded and run in Matlab. The analytical diagrams are either in R, Matlab, or DiscoveryDV, an open-source python-based library for visualization [98]. Informative figures are designed using PowerPoint Microsoft Office. All codes and analyses have run using a personal computer. Software and hardware characteristics are summarized in table 3.7.

Table 3.7 Modeling and analysis environment platforms

Software platform	Hardware platform
Matlab-R2017b R-3.6.3 DiscoveryDV-1.2 PowerPoint-2016	Windows10, 64-bit, Core(TM) i7-6700HQ

3.5 Summary

In this chapter, description and data analysis of six case studies to be used in models development in this work are presented: Santiago-EOAS, Camariñas, and Muralla, as representatives of urban, coastal, and highland rural areas, three meteorological stations and Sotavento wind power plant located in Galicia, Spain. For all these Galician stations, hourly WRF model were collected for the same period and same weather parameters. Also, M2 tower meteorological station located in Colorado, U.S., providing 1-minutes weather data at six different heights. Finally, La Haute Borne wind farm was presented, providing several years of weather and power generation data of four wind turbines. Different data analysis methods were applied to better understanding the meteorological behavior of these case studies, in order to distinguish their characteristics.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 4

Artificial neural networks in wind speed and power prediction

In this chapter, an extensive systematic review [99] is carried out, focusing on artificial neural networks (ANN) applications to wind speed and wind power prediction. In addition, a comprehensive discussion is provided on the success of different ANN models, structures, activation functions, and feature selection methods for different time horizons, steps ahead, and probabilistic-deterministic approaches. In order to facilitate this systematic analysis, simplicity is considered a principle in all parts, such as diagrams, tables, and figures.

4.1 Introduction and background

To further reduce the cost of wind energy from both the wind energy performance and the increase of the lifetime of the turbine blades, accurate forecasting of wind power would take a significant role in better control of the energy supply. As a consequence, accurate forecasting of wind speed (including the wind direction) could lead to applications of advanced control and regulation technologies for better performance in wind energy production and supply, and for reduction of the unsteady loads on the turbine blades for prolonging the lifetime of wind turbines and for reducing the noise level in wind energy production. So far, forecasting wind speed and power remains one of the important issues for industry because of the unsteady and fluctuating nature of the wind (from small temporal scale in seconds to large temporal scales in days or even seasonal changes). Different prediction methods, i.e., physical, statistical, and machine learning-based, have been investigated, aiming to reduce the forecasting errors and increase the forecasting reliability.

Recently, by the revolution of machine learning (ML) and data mining techniques, most of the scientific fields, forecasting included, have taken advantage of their less complexity and higher accuracy in comparison to the conventional methods. Neural networks, as one of the most cited ML methods, is vastly used in different science areas such as medicine [100], astronomy [101], geophysics [102], nanotechnology [103], energy systems [104], solar radiation forecasting [105], and electricity load prediction [106].

Compared to other types of data-driven artificial intelligence methods, artificial neural networks (ANN) are the most applied techniques in wind speed and power forecasting due to their successes in both research stages and real-time applications. Previous reviews were mainly focused in specific aspects of wind speed and power forecasting (WS&PF) such as time horizon [107, 108], data-mining [109], combined methods [79], probabilistic models[110], multi-step ahead [111], and optimization techniques [112]. Few studies have also proposed a more general perspective covering most of the topics [113–115].

4.1.1 Review strategy

Four strategies were used to select studies that are potentially relevant to this work. First, as a pilot search, different keywords and their combinations were used to find as many recent papers as possible in the selected publication databases. Second, focusing on the energy scope, a manual search carried out on the authors' collective observations during the pilot research. Third, the articles cited in the two previous groups were investigated to populate the selection basket. Fourth, the references of all the selected articles in the previous steps were checked to mark any new article.

4.1.2 Publications selection

Key data were extracted from all collected studies using a standardized format, structured in three main groups: a) prediction model information, including time horizon, time interval, WS-WP approach, deterministic-probabilistic approach, and error indicators; b) neural network information, including architecture, activation function, learning and training methods, feature selection, sample size, and processing time; and, c) other wind energy related information, such as dataset and its characteristics, case study location, year and journal of publication.

In order to achieve current state of art in ANN wind energy applications, publications period from 2010 to 2020 was considered. The selecting process, shown in figure 4.1, was

applied in two rounds: First, a number of articles were excluded, as their topic was not totally in the line of neural networks application in wind speed and power prediction. Second, another set of articles were excluded, as they have not reported most of the information mentioned in the categories above.

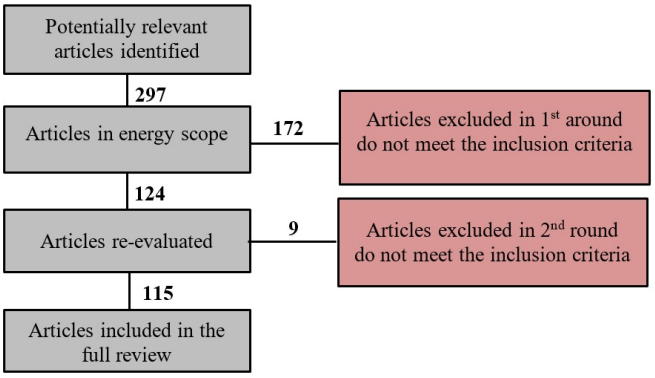


Fig. 4.1 Flow diagram of the systematic review carried out

4.2 Artificial neural networks

As a brief history, the concept of artificial neural network (ANN) was firstly proposed in 1943 by McCulloch and Pits [116], which was constructed for the problems using a network of neurons as a computational system. Inspired by biological neurons, the base structure of an ANN is the artificial neuron described by equation 4.1.

$$f(x) = f_{act}(w.x + b) \tag{4.1}$$

Where $f(x)$ is the predicted result, f_{act} is called activation function (here step function), w represents weights, b stands for bias, and x is the input. This unit is still using today for single-layer neural networks and as a key component in more complex networks. As it is shown in figure 4.2, the proposed ANN contains three main features: input, main cell, and output. Initially, a step function, namely, perception, has been applied. Whereas, nowadays a diverse group of functions can be used.

Three sets can generally characterize an ANN: the number of neurons and hidden layers (architecture), methods of determining weights of the connections (training/learning), and its activation functions [117]. In addition to these three items, feature selection is also considered

as important as the other layers. Then, ANN designers can build models from a wide range of possibilities. Table 4.1 briefs the neural networks' sets and provides a classification for their applications.

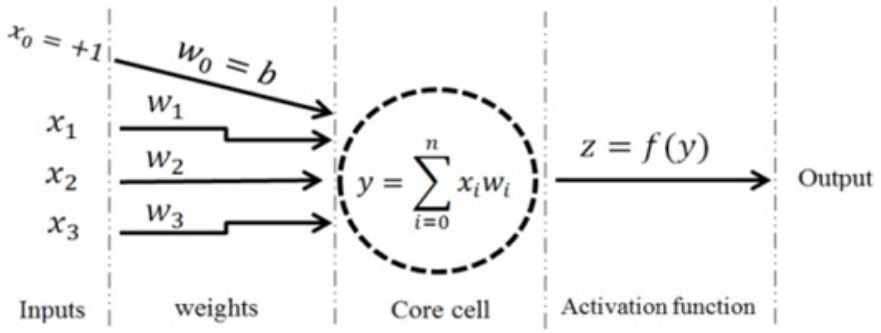


Fig. 4.2 Four components of a neural network

From the neural network architecture point of view, many different models have been proposed. Feed-forward neural network (FFNN) is the simplest one but most popular type of neural network with full connectivity between the adjacent layers. Data moves in just one direction, from the input to the output side. Another popular kind of NNs s is called recurrent neural networks (RNNs), which contain one or more loops in the network serving bi-directional data flow. Compared to FFNNs, an RNN propagates data from input towards output and sends direct data from later processing stages to earlier ones in feedback.

The learning methods are the most attractive parts of neural networks. Defining a specific task, for instance, a cost function, learning in the neural network means to find out the

Table 4.1 Neural networks components and diverse application

Architecture	FFNN *SLP, MLP*	RNN *SRN,ESN,ELMAN, NARX*
Activation function	Default *Sigmoid*	New *RBF,WNN, emotional...*
Feature selection	Wrapper Models *Exponential and sequential search*	Filter Models *Correlation, MI, IG,mRMR*
Learning	Deep learning DBM,RBM	Shallow learning ELM,BP,Optz
	Unsupervised learning	Supervised learning
	Online learning	Offline learning

solution by using a set of observations or iterations. As a classification, a neural network may be either a shallow NN (SNN, with only one hidden layer) or a deep NN (DNN, with more than one hidden layer). These extra layers in DNNs enable the composition of features from lower layers, providing more potentials in modeling complex data with fewer units than a similarly performing shallow network [118]. DNNs were mainly designed in the context of FFNNs, but also successful deep RNNs are reported in the literature [119]. Another learning method classification divides them into supervised, unsupervised, and reinforcement methods. In supervised learning, which is normally used in classification, forecasting, and modeling, an error function is used to guide the learning process, as an automatic teacher, following a backpropagation (BP) process. Unsupervised learning involves targeting values, and it is only based on the correlations among the input data, which are used to find any significant features or patterns without any "teacher" help. Generally, estimation problems, including clustering, statistical distribution analysis, compression, and filtering, are considered unsupervised training applications. However, reinforcement learning can be regarded as a special case of supervised learning where the exactly desired output is unknown, as the *teacher* supplies only feedback about success or failure of an answer [120, 121]. Following a different perspective in a learning context, online and offline (batch) learning are also important to be considered. Using online learning, data becomes available in sequential order and improves the prediction result for future data. In offline or batch learning, the model generates the prediction on the whole training dataset at once. Online learning is more related to the concept of unsupervised learning and is common for biologically inspired algorithms [122]. Generally, those ANNs that use the backpropagation (BP) algorithm for training are normally categorized as offline learning.

Feature selection (FS) is a critical issue in machine learning, as in neural networks. It can significantly increase the accuracy of the results and improve the model's performance by a good feature selection, such as eliminating redundant and irrelevant features. In an ANN, given a set of N features, there are 2^N possible choices. Finding such an optimal sub-set that performs the least error is usually unfeasible [123]. In a general categorization, FS can be divided into two main groups: Filter and Wrapper models. The feature is based on the selected data in a filter model, so it is independent of the classification performance; however, in wrapper models, the selection process is dynamic, improving the regression or classification performance. The search of the best feature subset can be performed utilizing any search algorithm, like hill-climbing, greedy or genetic algorithms [124]. Wrapper models are more powerful but with higher computational costs [125]. Correlation, Information Gain (IG), maximum-Relevance-minimal-Redundancy (mRMR), Relief Score (RS) are categorized

as filter models; and, exponential search and sequential search are considered as wrapper models, which are intensively reviewed in [126].

4.3 Construction of neural networks

An ANN can be customarily constructed using a combination of different network architectures, activation functions, learning methods, and input sets. In this section, we have discussed these network components highlighted in previous works are discussed.

4.3.1 Architecture

To define the parameters of neural networks, designers should be aware of the available options. Weights, connections, and neuron models are three aspects of a network that should be treated. In general, three methods can be used to define the parameters mentioned above: Discretionary, analytic, and data-driven. The latter is based on learning or training data and is the widely used parameters tuning. It derives from a particular set of problem examples using a parameter optimization algorithm namely learning algorithm. The discretionary method chooses the network and the neuron model based on individual experience, rules of thumb, or trial-and-error. In contrast, for the analytic approach, mathematical theories play the leading role [122]. The way that neurons connect in a neural network defines the structure of the model. Considering the WS&PF models published in the literature, different neural network structures are summarized below.

4.3.1.1 Feedforward neural network (FFNN)

As mentioned previously neural networks can be divided into two main groups in terms of structure: Feed-forward and Recurrent networks. considering FFNNs, there is no any loop in the network. This type of NNs is the main and the simplest type of neural networks that allows the information to move in only one direction, from the input nodes, through the hidden nodes (if any), and to the output nodes. Based on the network layers, FFNN itself can be categorized into two main groups: single-layer perceptron (SLP) and multi-layer perceptron (MLP).

4.3.1.1.1 Single layer perceptron

The simplest type of NNs is single layer perceptron in which inputs are fed directly to the output via a series of weights, as shown in figure 4.2. There is not a general rule for the

best structure of NNs. Hence as a reasonable procedure, it is better to start with the most straightforward configurations. In [127], the effect of different learning rates and iteration numbers of four NNs structure to predict next-hour wind speed (WSP) in Oaxaca, Mexico, as studied. Between four structures the most simple has shown superiority over the others, with MSE and MAE values of 0.0016 and 0.0399, respectively. In another study [32], the same authors modeled wind speed size applying ARIMA as a predictor and used the same four ANN structures to correct the prediction errors as the nonlinear tendency detector of the time series. Again the most straightforward model, (2,1), performed more accurately.

4.3.1.2 Multi-layer perceptron

Multi-layer perceptron (MLP) consists of one or more hidden layers, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. The utilized activation function for training is called backpropagation (BP). Hervás-Martínez et al. [128] proposed a hyperbolic tangent unit FFNN in which architecture, weights, and topology are defined using hybrid evolutionary programming in conjunction with a local optimization algorithm, Levenberg- Marquardt (L-M). In this work mesoscale model MM5 is nested to a global forecasting model to produce a limited area numerical weather prediction over a 15 km $\times$ 15 km resolution grid at different height levels. Wind speed and direction, temperature, and solar cycle are considered as input to predict the wind speed in the location of several wind turbines in a wind park located in the south of Spain. MSE of 3.44 was reported as the best result of the proposed model, which overcame MLP and RBF performances. In another study, four different FFNNs with 10, 20, 30, and 40 neurons in the hidden layer were used for a short-term wind speed forecast (WSF) in the Batman area of Turkey [129]. Eleven inputs are considered for all structures, including wind speed, humidity, air pressure, air temperature, and soil temperature. Maximum accuracy was achieved by a structure of (11,40,1). The reported MSE decreased by increasing the number of the hidden neurons, proportionally, from 2.03 to 0.59. In [130] a feed-forward neural network was organized with only one hidden layer including nine neurons, using hyperbolic tangent activation functions in an entropy training context of neural networks; this approach takes advantage of entropy criteria instead of least mean square error (LMSE), because of the non-Gaussian error distribution observed. A (31,63,31) MLP model proposed in [77] based on seasonal exponential adjustment (SEA) to forecast a daily WSF for the next one month. Kolmogorov–Smirnov (K–S) test method was applied to test the feasibility of using wind speed of a specific month during several years to predict wind speed in the corresponding month of the next year. The effect of the dataset on the outputs was also investigated. 21% of MAPE is reported as the best result of the model for December. Alexidas et

al. [131] proposed a unique correlation MLP model for WSF to predict up to 2-hours ahead of time steps using the data from 800 m to 40 km windward direction, with an average error of 20% to 40% better than the persistence model (PM). In [132] a hybrid Fast Block Least Mean Square (FBLMS)-MLP model is proposed, which applied on a 7-year dataset from 2007 to 2013 recorded from two sites in Turkey. The model performed better than three and five-layer conventional ANN structures, approximately between 2.1% and 3.3%.

Agrawal et al. [133] worked on an ANN structure based on yearly auto-regression (ANNYAR) to figure out the most influent parameters affecting wind prediction. They considered the corresponding range of an annual dataset to extend the time horizon from 6 to 96-hour ahead. Sixty-three different combination of the six parameters (time of the day, pressure, humidity, daily temperature, wind speed, and dew point temperature) were collected and analyzed to determine the optimal number of hidden layers (1 or 2), based on MSE, MAE and COC diagrams; single input hidden layer suited well. Dai et al. [134] proposed a model using meteorological numerical weather prediction (NWP) data from two main models, Weather Research and Forecasting (WRF) and Global Forecasting System (GFS), exporting their modified data by a Kalman filter (KF) to a neural network for wind power forecast. Five different MLP models were defined for five turbines, and a fair value of 16.47% of the normalized root mean square error (NRMSE) was reported. Kitajima and Yasuno [135] investigated a complex value feedforward neural network (CVNN) for 24-hours ahead wind speed forecasting using outcomes of the national Japanese numerical weather prediction (NWP) model to produce the inputs for the proposed model. Wind speed and wind direction were converted to a complex format and used as inputs into a three-layer feedforward NN. Prediction results demonstrated the usefulness of this proposed output prediction system, with a relevant improvement in prediction accuracy for the real-valued neural network (RVNN) results.

Considering combined MLP with other machine learning methods, Tasnim et al. [136] established an LR-MLP model with the linear regression part modified by using fast fourier transform (FFT) and principal components analysis (PCA). Their results for up to two days WPF showed 4.26%, 11.29%, and 4.24% improvement compared to LR (raw), LR (FFT), and LR (PCA), respectively. Lee and Shukur [137] proposed a hybrid KF-ANN model based on ARIMA to improve the accuracy of WSF. The outcome MAPE indicated that the hybrid KF-ANN model was the most effective for improving the accuracy of WSF compared to ARIMA, ANN, and hybrid ARIMA-KF models for two meteorologically different datasets located in Iraq and Malaysia for a period of four to five years of recorded data. In a very short-

term WSF study, which is essential for wind turbine control, Pourmosavi Kiani and Ardehali [138] proposed an ANN-MC model which could successfully struggle with over-training and extrapolation, as the common problems for ANN. As MC uses to capture long-term patterns of wind speed besides ANN captures short-term patterns. MAPE (%) for 2.5, 5, and 7.5 seconds ahead reported as 3.14, 8.03, and 11.33, respectively. In combination with supervised and unsupervised methods, Sheela and Deepa [117] proposed a self-organizing map (SOM)-MLP model to predict short-term wind speed. SOM is used for clustering input data into four categories for determining the relationship between input vectors and four MLPs for the prediction of each class of data. The reported RMSE of the proposed model was 0.0828, overcoming conventional MLP, BPN, and RBF neural network models. In a modified empirical mode decomposition (EMD)-FNN model [139] the high frequency of the decomposed wind speed time series is eliminated from the inputs. Both monthly and daily-based multi-step ahead wind forecasting considered for three models. The improvement in daily based was higher than in monthly based, considering Multi-step (M)-EMD-FNN and EMD-FNN model, proving the positive effect of eliminating high-frequency decomposition. Data is collected from Gansu province in China in mean daily and mean monthly format for three years, and MAPE (%) is reported 14.92 and 17.29, respectively.

In a comparative study, Ghorbani et al. [140] did a comparison study between MLP, genetic expression programming (GEP), multi-layer regression (MLR), and persistence model (PM) utilizing four criteria: COC, RMSE, Nash–Sutcliffe efficiency coefficient (E), and Akaike information criterion (AIC). Fourteen different model structures were tested to determine the best structure: (8,8,1). However, AIC provided the best result using $2n+1$ hidden neurons and n neurons in the input layer.

Considering more than one hidden layer study, Grassi and Vecchio [141] proposed a two-hidden layer neural network (TNN) to predict the output power energy from three experimental wind power plants in South of Italy. The number of neurons in each hidden layer and the type of activation functions defined by experiments is considering, comparing the MSE and MAE obtained by each model. As inputs, a monthly average of five variables namely wind speed, temperature, humidity, generation, and maintenance hours, were considered, showing that the latter variable can significantly improve model results. The lowest reported MSE and MAE values from a training test were 0.0047 and 0.0156, respectively. In another study [142], a wavelet transforms–two hidden layer Neural Network (WT-TNN) was applied to predict the hourly wind speed. Neuron numbers of the hidden layer for each decomposed series and for each season were obtained through experiments using MSE as the selection

criteria. Results from four case studies approved its success compared to PM, one hidden layer NN, and simple TNN. The minimum MAE reported as 0.23 belongs to the fall with a Wilcoxcon signed-rank of 3.03 and a p-value of 0.00. Another ensemble model built to predict up to 48 hours ahead wind power generation [143]. The model used 52 two hidden layers NNs sub-models to train wind information and five Gaussian Process (GP) sub-models to give the initial power level to NN sub-models. The Scaled Conjugate Gradient Backpropagation (SCG-BP) Method is used to train this NN, which is less accurate but faster than the LM Backpropagation Method.

4.3.1.3 General regression neural network (GRNN)

GRNNs have a similar architecture as MLPs, with the difference that traditional networks perform a classification, where the target variable is categorical. In contrast, GRNNs perform a regression, where the target variable is continuous. In a hybrid model of the k-means cluster algorithm and GRNN, Niu et al. [144] proposed a day-ahead WPF model. NWP data categorized depending on daily similarity between observed and predicted subsets were chosen as training samples to build a GRNN model. This ANN model building is based on mathematical statistics that can approach the mapping relationship according to the sample data. Even if there are only a few sample data, the network output can converge to the optimal regression surface. Moreover, only a parameter θ needs to be determined in GRNN, which can avoid the effect of subjective assumptions on forecasting results. 11.45% of improvement is achieved in comparison to the persistence model (PM). Kumar and Malik [145] proposed another GRNN model for WSF in 67 different sites in India using various NWP data as inputs. Accuracy of this model reported 99.9% and 97.97% for the training and testing phase, respectively, showing significant progress compared to the MLP model. Elattar [146] combined an Evolutionary Optimized Local General Regression Neural Network (EOLGRNN), using EA to optimize the smoothing parameter that produced a training time value of 400s; however, less than 40-seconds was required for persistence model, seasonal autoregressive integrated moving average (SARIMA), RBF and local GRNN.

4.3.1.4 Recurrent Neural network (RNN)

Contrary to FFNNs, recurrent neural networks form directed cycles in the connections between units, and they can use their internal memory to process arbitrary sequences of inputs. Theocharis and Barbounis [147] built a local RNN using global recursive prediction error (RPE) and three local versions as learning algorithms to predict wind power and wind speed in a long term context, 72-hours ahead on the Greek island of Grete. Wind speed

and wind direction of four different points around the main point, examined in terms of distribution correlation with the real data of the main point, to choose the most appropriate inputs coming out from the atmospheric modeling system, named SKIROM. The model was run for 3, 5, and 7 inputs from northern (N), eastern (E), and western (W) points which were locating on the sea contrary to the excluded southern (S) point, with different wind speed distribution. The study concluded that, in terms of convergence and accuracy, the local PRE schemes have a slightly lower performance than the global one, while its computational cost is considerably less.

In another research [148], three RNN model types and four learning approaches were tested through extensive experiments to find out the best state. In this study, global and decoupled recursive prediction error (GRPE and DRPE) were used as learning algorithm in another RNN based model proposed for wind speed and power prediction, with considerable computation time and storage capacity savings in comparison to real-time recurrent learning (RTRL), the most common algorithm for online training of RNN. Goh et al. [149] proposed a complex-valued pipelined recurrent neural network (CPRNN) architecture for wind speed prediction in RTRL, reporting more accurate results than univariate real-valued methods. 10, 20, and 30 steps ahead prediction performance with a one-hour step, reported by less used measures of multiple determination (r^2) and prediction gain (r_p) formula. For 1, 3 and 6 steps ahead, r^2 was reported as 0.70, 0.76 and 0.77, respectively. Enjoying the ability of ridgelets, an RNN model is proposed [150], using ridgelet as its activation function of the hidden layer nodes; a new statistical search engine, new differential evolution (NDE), is used to train and determine its free parameters, requiring a small set of training and validation samples, and low computational cost. The superiority of this technique was proved by comparing the accuracy of the prediction, trained by different optimization algorithms as SA, GA, and PSO. Two wind power plants located in Ireland and Spain were considered for testing the model with two and four months recorded data. The average MSE (MW) was reported 75.24 and 0.428 and the average NMAPE (%) 15.28 and 11.14, respectively. It is also proved that by expanding the training dataset from 30 to 70 days, with a proportion of 10 days in each experiment, the average MSE and NMAPE decrease to its minimum for 50 days training set and increase the number of training days again. Richard et al. [73] investigated a comparison between three different neural networks, namely MLP, Elman NN, and simultaneous RNN, all of them optimized by particle swarm optimization (PSO), for short-term wind speed forecasting. Results show that all of them improved the forecast accuracy. In general, RNN achieved higher accuracy, especially against data out of the training range, with the drawback of a higher training time.

4.3.1.4.1 Elman neural network

Elman neural networks consider as simple RNNs that composed of three layers of input, hidden, and output with a context layer that feeds back the hidden layer outputs in the previous time steps. Senjyu et al. [151] proposed an RNN model for wind speed and power prediction and compared its results to FFNN, showing the better performance of RNN in both short to long term horizons. The MAPE, was reported as 4.87%, 5.19%, 7.31%, 19.53%, 23.89% and 28.22% in 3-h, 6-h, 9-h, 1-day, 2-days and 3-days ahead. Wind power is then predicted using wind speed prediction and the kinetic energy equation. Wang et al. [152] constructed a model with an ENN and empirical mode dcomposition (EMD) by applying a novel secondary decomposition algorithm (SDA) using WPD and EEMD. Partial autocorrelation function is used to determine the potential relationship between the input vectors of hourly wind speed, recorded for 15 months from January 2011 in Gansu province of China. The number of hidden neurons was defined according to Kolmogorov's theorem: 2n + 1 hidden neurons are sufficient to map any function for n inputs. Thirty-six different ENN models with different sets of inputs were constructed, divided by four different seasons and eight decomposed wind speed signals. The prediction was more accurate for the fall season with an MAE, MSE, and MAPE of 0.37(m/s), 0.44(m/s), and 12%, respectively. Liu et al. [153] proposed an Elman Neural Network (ENN) model to predict the short-term wind speed. Results also show that the strategy of decomposing the wind speed time series using two successive signal processing stages improves effectively the forecasting performance of the Elman model. Qin et al. [154] constructed a medium-term WSP model using an Elman Recurrent Neural Network (ERNN). Support vector regression (SVR) is used to detect data outliers, and Kruskal-Wallis (K-W) test is conducted to explore the distribution of continuous random variables. At the last stage, Seasonal Index Adjustment (SIA) is applied to classify the seasonal and trend data components. Reported results in three sites of china with average values are 1.11/(m/s), 1.94(m/s), and 15.3% for MAE, MSE, and MAPE, respectively.

4.3.1.4.2 Eco state network

Echo state network (ESN) is one of the most plausible recurrent neural networks commonly applied to nonlinear time series prediction. The most attractive feature of ESN is its dynamical reservoir, a recurrent layer of nonlinear processing elements. In ESN, a recursive network with a large-scale stochastic connection is adopted instead of a hidden layer applied in traditional neural networks. In [155], Liu and Zhang used ESN as a part of their hybrid neural network to predict short-term wind power generation. A deterministic and probabilistic wind power prediction proposed by [16]. A hybrid model of echo state network (ESN), least-square

support vector machine (LSSVM), and regularized extreme learning machine (RELM) was proposed. A self-adaptive multi-strategy differential evaluation (SAMDI) algorithm is used to dynamically adjust the weight matrix to the models on the optimal virtual prediction scheme. Wind power data collected from Jiangsu province in China, with a 10-minutes interval. Variational mode decomposition (VMD) was used to decompose input data and showed more success than EEMD. The hybrid model performed better than every single model by normalized RMSE (%) of 5.68, 8.44, and 10.12 for one, two, and three steps ahead, respectively.

4.3.1.4.3 Nonlinear Autoregressive Exogenous model

Rivera et al. [156] proposed a nonlinear autoregressive exogenous model (NARX) model being applied for one step ahead short-term wind speed forecast in Oaxaca, Mexico. Different numbers of input and hidden layers were tested to determine the best model structure. Mahalanobis distance was applied as outlier detection. Later, the granger test was used to determine the variables included in the model: wind speed, radiation, temperature, pressure, and wind direction. The first two parameters were selected. NARX results showed a performance of 4% over NAR (Nonlinear Autoregressive model) results and 11% over the persistence model. In another related study, a NARX model was used to provide inputs for another neural network combined with wavelet transform proposed by Prema et al. [157]. Three different training methods, including recursive, conditional, and parallel training, were applied to find the best training strategies. Campos-Amexcua et al. [158] investigated the performances of linear and nonlinear models through ARIMA and NARX models. Although ARIME linear model provided reasonable step-ahead wind speed prediction, NARX performed more accurate results, with an average improvement of MAE and MSE of 5.5% and 10% for an hourly dataset, and 2.3% and 12.8% for ten minutes dataset, respectively.

4.3.1.5 Artificial Neuro-Fuzzy Inference System

A fuzzy neural network is an artificial neuro-fuzzy Inference System (ANFIS) with the structure of a neural network that creates a synergized model, enjoying the human-like reasoning style of fuzzy systems with the learning and connectionist structure of neural networks [19]. Abu-al-ez et al. [18] proposed a neuro-fuzzy system using modified fuzzy C-means (MFC) to cluster data implementing a hybrid optimization method to obtain an optimal number of fuzzy rules. Input data of temperature, wind speed, air density, and pressure were categorized into four seasonal groups. The reported average RMSE and ReErr were 0.003743 and 5.86, respectively. Catalão et al. [159] proposed a particle

swarm optimization (PSO)-ANFIS model to predict one hour ahead power generation of a case study in Portugal by using PSO to tune the membership functions of ANFIS. The average MAPE improvement over four seasons in comparison to the persistence model was 71.60%. In another study [159], the same authors improved their previous model by using a preprocessing step, namely Wavelet transform (WT)-PSO-ANFIS model, which improved MAPE by 73.85% respect to the persistence model. Wang et al. [85] worked on an experimental investigation of a fuzzy neural network optimized by cuckoo search (CS) and combined with ensemble empirical mode decomposition (EEMD) for the short term wind power prediction. Results showed that the EEMD decomposition part included in the optimized CS-EEMD-ANFIS model makes more contribution than the CS part. Compare to EEMD-FNN, CS-EEMD-FNN could improve the MAPE for one (10 minutes) to six steps (1h) ahead prediction by 4.31%, 5.28%, 5.9%, 6.77%, 7.63%, and 8.64%, respectively. Catalao et al. [160] proposed a hybrid evolutionary adaptive (HEA) mutual information (MI) methodology, namely MI-WT-EPSO-ANFIS, to predict short-term wind power generation. Mutual information (MI) was used as the feature selection, and the wavelet transform was applied to manage the unsteady character of the wind. Evolutionary PSO (EPSO) was used to assist in tuning the membership function parameters of the ANFIS model. The average MAPE value was only 3.75%, for an average error variance of 0.0013 and a NRMSE value of 2.66%. Haduq et al. [161] proposed a combination model of WT+ANFIS+SVM+GS. SVM was used to reduce the short-term wind power forecasting error coming out from ANFIS, and grid search (GS) was applied to optimize the SVM algorithm. This model was applied in Alberta, Canada. Results were compared to other 27 models built with different combinations of machine learning techniques, using MAPE, NRMSE, and NMAE seasonal statistics, showing the higher performance of the new proposed model. Faghihnia et al. [162] worked on a local neuro-fuzzy (LNF) approach trained by a polynomial model tree (POLYMOT) learning algorithm to predict the next 24 hours of power generation. The POLYMOT algorithm increased the complexity of the model by applying either higher-order polynomial or performing input space splitting until a satisfactory performance was achieved. The average NRMSE improvement with respect to persistence achieved at four different months was 67.3%. Another multi-step ahead WS prediction from 15 min to 3 hours ahead was developed by Theocharis and Barbounis [163], using a locally feedback dynamic fuzzy neural network (LF-DFNN). The network includes Takagi-Sugeno-Kang (TSK) fuzzy rules, where the consequent sub-models are implemented by recurrent neural networks of the infinite impulse response (IIR)-MLP. A 33.87% of MSE improvement over the persistence model was achieved for 15 min ahead prediction.

4.3.2 Activation function

Activation functions that are used in the aforementioned ANNs are summarized in table 4.2, taking [166] as a reference. More than that, several other models are proposed in the literature based on new successful activation functions, which are discussed below.

4.3.2.1 Radial Base Function (RBF)

Chen et al. [78] proposed an RBF-based model trained by an orthogonal least-square algorithm (OLS) for one hour ahead wind forecast. Six different time steps of wind speed data with 2-hours frequency used as inputs and the RMSE reported within 2 m/s . The advantage of using OLS was to choose the centers of hidden layer nodes automatically and more efficiently. Zhang et al. [164] proposed a hybrid model of the wavelet transform, seasonal adjustment method, and RBF (WT-SAM-RBF) was applied successfully for a short-term wind speed forecasting model. The low frequency decomposed signal obtained from WT is used to remove the seasonal component derived by SAM. The calculated trend component is input to an RBF, but the prediction is composed of both the RBF output and the seasonal indices previously obtained. In another similar study by the same authors [142], SAM-RBF with exponential smoothing methods (ESM), SAM-ESM-RBF, was developed. ESM captures the linear pattern of trend components before applying RBF for predicting the non-linear properties after seasonal adjustment of input data. Moreover, z-statistic values and p-values of Wilcoxon signed-rank tests are used to evaluate the predictive performances of the proposed model and other seven models considered. The numerical results indicated that the MAPE values of the new proposed approach were 12% and 16% for two different real wind speed datasets. Sideratos and Hatziargyriou [165] proposed a combination of RBF and fuzzy sets to predict power generation from 1 to 48-hours ahead.

Table 4.2 Widely-used activation functions in ANNs' hidden layers

Function	Definition	Range
Linear	x	$-\infty < x < +\infty$
Logistic(Logsig)	$\frac{1}{1-e^{-x}}$	$0 \leq x \leq +1$
Hyperbolic(Tansig)	$\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$	$-1 < x \leq +1$
Exponential	e^{-x}	$0 \leq x < +\infty$

A self-organized map classifies the NWP input data in three classes, depending on the magnitude of the wind speed, the prediction hour, and the hour after prediction. The output

quality of a primary RBF model was estimated using fuzzy rules, and the results are improved using an additional RBF structure. The proposed system enhances the persistence statistics up to 46% for both NMAE and NRMSE after 10-hours. This deterministic model was modified by Sideratos et al. [167] to build an RBFNN probabilistic model applied to power prediction. A self-organized map (SOM) classified the outputs uncertainties in multiple levels, and PSO trains input to related RBFNN models. More uncertainties are involved using data from NWP. The vertical profile of the wind speed is measured by the standard deviation of the wind speed values that correspond to six different horizontal levels of the NWP to the nearest point to the wind farm. The evaluation of the proposed model on two other wind farms shows that the model can perform satisfactorily and robustly in various weather conditions and different terrains. Another probabilistic model of short-term WSP was proposed by Zhang et al. [168]. The model uses a direct interval forecasting method as probabilistic wind speed forecasting for up to the next 5-days. First, the centers of the RBF are determined by a k-means clustering algorithm, and the hidden output weights of the RBF are pre-trained using the least-squares error (LSE) algorithm. Second, the hidden-output weights were further adjusted by NSGA-II to minimize the width and to maximize the coverage probability of PIs simultaneously. To choose the best input set, Box-Jenkins, ACF and PACF are used to make the first guess about the potential order, and then bayesian information criteria (BIC) is employed to select the best one. The model was applied on three sites in China, and the results were satisfactory in comparison to the performance of MLP in terms of PIs quality. Some comparison studies have been done with RBF and other methods, which, in general, have resulted in better results by using the radial base function model.

In a recent NN structural study, Vafaeipour et al. [169] analyzed different structures of MLP and RBF in terms of several neurons and training algorithms, using the sliding window technique, to find the most valuable/type network for wind speed forecasting. Results showed that the higher the number of hidden neurons for the RBF networks, the lower the mean square error, following a trend defined by the $y = 1/x$ relationship. Comparing the results of the best MLP (4-7-13-1) and RBF (4-25-1) models, applied on a one-year recorded wind speed data in Iran showed the superiority of MLP over RBF with MSE (m/s) of 1.19 and 1.36, respectively. Sheela and Deepa [117] compared RBFN and BPNN for the same time horizon wind speed forecasting, both with a (3,3,1) architecture. For RBFN, sigmoid was applied as an activation function, and the training method was a gaussian function. For the BPNN, LM and new radial base were applied to the activation function and training method. The experimental results on a real case showed that RBFN has a significant advantage over the BP model, with MSE equal 0.0013 from RBFN in comparison to 0.0397 from BP. Also,

RBFN is better in terms of stability, accuracy, and training time. The same conclusion was established by Wu et al. [170], for a short-term wind power prediction. MAPEs of the next hour wind power prediction for RBF, in contiguous 24-hours, 48-hours, 72-hours, and 96-hours are reported as 7.12, 7.30, 8.97, and 10.06, respectively. Li and Shi [171] investigated the accuracy of three different ANNs using RBF, BPNN, and ADAptive LINear Element network (ADALINE) for one hour ahead wind speed prediction. The number of hidden neurons in BP and ADALINE was defined as log(t), where t stands for training vectors' numbers. The effects of the learning rate, the number of inputs, convergence time, and different model structures were investigated in two sites, considering other evaluating criteria and several datasets. ACF and PACF were used to select inputs among the one-year hourly wind speed series recorded in Dakota, USA. Although BPNN showed better performance in some criteria for one site, RBF overcomes in another site. Among all the 160 BP models tested, using six previous observations as input and learning at a rate of 0.1 generated the smallest RMSE value (1.254), while using eight previous time steps and a 0.075 learning rate resulted in smaller MAE (0.945) and MAPE (0.206). In another study, Yan et al. [172] investigated prediction accuracy of three different machine learning models, namely RBF, genetic algorithm – backpropagation (GA-BP), and support vector machine (SVM), in a short-term power prediction context. The effects of sample size, number of training iteration, and accuracy of training samples (NWP) were analyzed in detail at three different wind farms. In general, RBF and SVM showed higher accuracy than GA-BP, with an RMSE range of 10 -20% for the three sites. The authors concluded that a limited number of samples should be used to train GA-BP (7000-8000), RBF (800-1000), and SVM (400-600). Minimum monthly average RMSE(%) between all three fields were reported as 10.92 for GA-BP, 10.67 for RBF, and 10.54 for SVM.

4.3.2.2 Wavelet neural network (WNN)

A wavelet activation function NN, using the multi-dimensional Morlet function, is proposed by Chitsaz et al. [173] to predict short-term wind power in Alberta, Canada. The model used the improved clonal selection algorithm (ICSA) to optimize the free parameters of WNN. The effectiveness of the proposed optimization method was proved in a set of experiments using other optimization techniques, such as PSO, simulated annealing (SA), and differential evolution (DE). More than 14 different models in two real-world datasets were tested, and the effect of three different training sample sizes (50, 40, and 30 days) was compared. As the error measure in the forecasting model, the maximum correntropy criterion (MCC) was used instead of MSE. The superiority of the Morlet function compare to the Mexican hat (MH) is shown, and the yearly average normalized (N)RMSE and

(N)NMAE of the proposed model reported as 12.10 and 8.41, respectively. Wang et al. [174] proposed a combined model of EEMD-GA-APSO-WNN in order to short-term WSF. The joint genetic algorithm-adaptive particle swarm optimization algorithm was used to optimize the parameters of the wavelet neural network (WNN) model and compensate for the disadvantages of each single optimization technique with a higher fitness value. MAE (m/s), MAPE (%), and MSE (m/s) of the proposed model for four sites in China were reported as 0.1347, 2.4907%, and 0.0271; 0.1281, 2.1562%, and 0.0302; 0.156, 2.074%, and 0.038; and 0.1344, 2.8895%, and 0.0262, respectively. Doucoure et al. [175] investigated the predictability analysis of the time series of wind power and its application in a forecasting method based on the adaptive wavelet neural network (AWNN). The Hurst predictability and multi-resolution analysis (MRA) of time series decomposition are applied to the wind power profile. As a result, computational time decreased up to 30%, keeping similar performance. The model applied on one-year wind speed data recorded from Trois-Rivieres weather station in Quebec, Canada showed almost similar performance compare to the persistence model, for a prediction of 1-5 hours ahead. For one-hour ahead prediction, symmetric (S) MAPE, MAE, MSE, and RMSE were reported as 16.90, 5.78, 0.01, and 7.06, respectively, all based on (km/h). In another study, Ricalde et al. [176] proposed a High Order Wavelet Neural Network (HOWNN) constructed by using high order terms in the neural network. Extended Kalman Filter (EKF) is used to update the weights of this model, showing high compatibility to the wind forecasting complexity.

4.3.2.3 Emotional neural network (ENN)

ENN is motivated by the neurophysiological knowledge of the human's emotional brain. Lotfi et al. [177] proposed an ENN that categorizes as brain emotional learning base (BEL) networks. A genetic algorithm (GA) was also applied for optimal tuning of crisp numerical parameters of ENN. GA was used to tune the parameters, and the error of different hidden layer neuron numbers was compared to find the best setup. A one-year hourly wind power dataset from a capital wind farm located in South Wales of Australia was used to test a short-term wind power prediction model. The proposed method performed better than conventional NN by improving RMSE about 1% and decreases standard deviation by more than 85%. In another research, focusing on simple days (SD) and wavelet transforming (WT) methods, Peri et al. [178] applied ENN as the core of such a combined model, reducing its errors up to 20% in comparison to ENN simple model. A WT-SD-ENN model was used for 10, 20, and 30 minutes ahead wind power forecasting of a wind farm located in Texas, USA, containing 150 wind turbines. One year dataset of wind speed and wind power with an interval of 10 minutes considered as input; wind power is passed through WT process and the output

besides wind speed time series and SD-based wind power forecasting, considered as input for ENN model. The minimum MAPE , RMSE (MW), and MAE (MW) were reported as 4.25, 7.98, and 6.37, respectively, for the winter season.

4.3.2.4 Other activation functions

A genetic algorithm trained a hysteric chaotic neural to predict multi-step wind speed [179]. The function of the output layer is hysteric, and the hidden layer neurons are defined as chaotic operators. As many different parameters need to be trained, like connection weights, chaos control, and hysteretic parameters, conventional training methods based on gradient algorithms do not fit properly. Zjavka [180] constructed a differential polynomial neural network (D-PNN). The model was trained with the output of two different NWP models, which significantly increased its accuracy. Amjadi et al. [150] proposed a neural network using ridgelet as its activation function of the hidden layer nodes, as it is described in subsubsection 4.3.1.4.

4.3.3 Learning process

4.3.3.1 Extreme learning machine (ELM)

Wan et al. [181] used ELM for probabilistic wind power forecasting. Three different bootstrap methods with several iterations were investigated to find the best combination of a bootstrap-based ELM approach (BELM) to forecast the prediction intervals (PI). RMSE and MAE were applied in a validation test to define the number of hidden layer neurons, resulting in ELMs achieving stable generalization performance once the hidden nodes exceed 63. Prediction interval coverage probability (PICP) and average coverage error (ACE) are used as a measure of the prediction interval reliability besides measuring the sharpness of the model (θ). PICP and ACE are reported as 91.51% and 1.51% for PI nominal confidence of 90% and 10-minutes time horizon. Liang et al. [17] proposed a hybrid SVM-ELM in which the former part is used as a predictor, and the latter is used as an error forecast model for 1-h forecast interval and 6-h forecast horizon. Their results showed that although SVM-ELM got the best result for multi-step ahead forecasting, SVM-SVM overcame different combinations in one-step forward. The reported improvements against the persistence model in terms of NMAE were 3.48%, 22.27%, 29.09%, 33.65%, and 37.36%, for a two to six timesteps ahead, respectively. Wang et al. [182] constructed an ELM-LBQ-SARIMA model to predict monthly and daily wind speed. Ljung-Box Q-test is applied to investigate the residual correlation after ELM was applied to the original time series. Then, seasonal auto-regressive integrated moving average (SARIMA) was used to enhance wind speed forecasting accuracy

as a diagnosis to the residuals that came out from the LBQ test. The model applied on three fields in China, and for the mean monthly, the best result is reported as 0.05 (m/s) for RMSE, 0.12 (m/s) for MAE, and 11.97 % for MAPE.

Coral reef optimization (CRO) was successfully applied by Salcedo-Sanz et al. [124] to train an ELM model for a short-term wind speed forecast. WRF input data used with an emphasis on the feature selection problem. The same authors later improved the model [183] adding Harmony Search (HS) to CRO. Recently, a Group Genetic Algorithm (GGA) was applied as a search engine by Salcedo-Sanz et al. [184] for several subsets of features jointly to ELM, in which Gaussian Processor (GPR) and ELM were used to solve the prediction problem. An atmospheric reanalysis dataset derived from ERA-interim reanalysis from the European Center for Medium-range Weather Forecasts (ECMWF) was developed and applied successfully. Wind speed and temperature data were collected from four points around three different wind farms in Spain for a 6-hours ahead WPF. GPR shows better results than ELM in the three cases, and in general, the performance of the ELM algorithm is reported 58.56% for r^2 and 4.9755MW for MAE. For the best case, r^2 was reported 71.13% and the measure of MAE as 3.9902MW.

In another study focused on characteristic components and using multi-input, multi-output (MIMO) strategy, Zhang et al. [185] proposed an EEMD-ELM model to WSF in a multi-step ahead context. The proposed method grouped the obtained IMFs to form the characteristic components with lower-order complexity. The combination of EEMD and entropy technique has shown an effective result in determining the characteristic component of wind speed time series. Liu et al. [84] studied four different signal decomposition algorithms jointly to ELM. Results proved that all four decomposition models improve results accuracy. However, fast ensemble empirical mode decomposition (FEEMD) in one-step-ahead prediction and wavelet pack decomposition (WPD) in two and three steps-ahead overcame others. As another outcome, this study showed that the proposed ELM provides better results than classical MLP and ARIMA methods due to its better generalization capacity. Considering ELM as a tuning-free method, Sun et al. [186] proposed a hybrid FEEMD-RELM model. A monthly wind speed dataset, measured at height 47 (m) with an interval of 20 minutes, was used for short and medium WSF in two wind farms in China. FEEMD compared to WT and EMD to shows its superiority and r^2, MAPE and MAE reported as 0.992, 11.142 % and 0.214 (m/s) for the proposed model of short term forecasting, respectively. Abdoos [82] proposed a combined model of variational mode decomposition (VMD) as pre-processor, Gram-Schmidt orthogonalization (GSO) as feature selector, and ELM as the central part of short-term WSF.

Cross-validations were applied to determine the optimal structure of the forecast engine by optimizing different numbers of feature selection and decomposition modes. Evaluation statistics such as NRMSE and NMAE, reported 5.51% and 3.58%, respectively.

4.3.3.2 Deep learning machine

Deep neural networks are multi-layer networks with many hidden layers whose weights are fully connected and often initialized or defined stacked Restricted Boltzmann Machine (RBM) or deep belief networks (DBN). Wang et al. [187] proposed a deterministic and probabilistic WSF model of WT, deep belief network (DBN), and spine quantile regression (QR). For each wind speed frequency obtained from WT, a deep independent framework based on DBN is driven by a layer-wise pre-training rule designed for structure mining and feature extraction. Probabilistic prediction is obtained by applying QR to deterministic forecasting. In another related study, the same authors [188] developed a hybrid model of WT-Convolutional NN constructed for probabilistic short-term WPF. An independent deep CNN network designed for each frequency came out from WT. As classical deep CNN could not be applied directly to WPF because wind power data is 1D data in a time domain and should be converted into a 2D image for feature extraction, then reconverted to a 1D vector again for prediction. Quantitatively, in a 15-min ahead prediction, the CRPS average performance of this proposed approach achieved improvements of 13.48%, 21.74%, and 9.86%, compared to the persistence method, BP, and SVM, respectively. In another study, Hu et al. [189] proposed a shared hidden-layer deep neural network (SHL-DNN). This method fused the WS information from multi-sources to build a DNN, and hidden layers shared across many wind farms, while the output layers are farm dependent, using the model as a transfer learning model that transfers knowledge from data-rich wind farms to newly-built wind farms. Tests with different training sizes of experimental data demonstrated that when the training dataset is not large, DNN models may be performed worse than other shallow models, such as SVR and ELM, in terms of accuracy and absolute results. However, when the training dataset is enough, this method effectively combines SDA unsupervised pre-training and supervised fine-tuning strategies. The model was applied to three different regions in China. Best results for MAE (m/s) are reported as 0.44679 (6-month training), 0.8017 (3-months training), 1.1364 (3-months training), and 1.16406 (6-months training), for 10-minutes, 30-minutes, 1-h, and 2-h ahead, respectively. Zhang et al. [190] proposed a sophisticated deep learning technique, namely deep Boltzmann machine (PDBM), and the corresponding learning algorithms, for short and long-range WSF. The PDBM models' structures in their experiments were (10, 50, 50, 10, 1) for short-time prediction and (6, 50, 50, 10, 1) for long-term prediction. Results showed that the PDBM method improves the

prediction accuracy by 10% above, not only in short-term prediction but also in the long-term, compared to AR, ANFIS, and SVR.

4.3.3.3 Backpropagation algorithms

A common method of training artificial neural networks is the so called backpropagation, usually associated to an optimization method, such as: gradient descent (GD) [77, 191, 131, 192], Levenberg–Marquardt (LM) algorithm [133, 132, 193, 134, 194, 128], conjugated gradient (CG) [195], Bayasian regularization (BR) [196, 197, 52], Broyden–Fletcher–Goldfarb–Shanno (BFGS), resilient backpropagation (RBP), and fast back propagation (FBP) [11, 198]. In fact, most of the recurrent and ordinary feed-forward networks used nowadays are trained with some form of back-propagation, typically associated to gradient descent. This combination is called steepest descent, as a good derivation technique described in the literature [183].

However, the LM algorithm trains a ANN between 10 to 100 times faster than the gradient descent backpropagation algorithm [141]. Catalão et al. [194] built a WT-ANN model for short-term WSF using the LM algorithm for training. The MAPE had an average value of 6.97%, outperforming persistence, ARIMA, and basic NN approaches, while the average computation time was less than 10-seconds. The dominance of the LM algorithm was proved over CGD, Scaled CG, and BR in a NARX model proposed by Rajanada et al. [196]. A similar conclusion was reported [197] for the use of an MLP structure trained by LM, BR, and SCG for short term WSF, with MAPE values of 23.78% and 24.26%, for LM and BR, respectively.

CGD, also called CG, is a medium method between GD and Newton Method. It demands a relatively small storage size and provides fast convergence and effective solutions of massive equation groups. Li et al. [195] proposed a new ANN model trained by the conjugated gradient (CGNN) for mid and long-term power generation. They tried to show the considerable low time consumption of the proposed model in comparison to ELM, RBF, and steepest-gradient NN, with 56.23%, 73.31%, 76.97% improvements, respectively, by setting the MSE value to 0.002.

A Bayesian neural network means adopting a Bayesian learning method for the implementation of a neural network model. In [52], an intensive review of Bayesian application in wind conversion systems, WSF included, is proposed, and the successful application of the Bayesian Regularization (BR) as a backpropagation method has reported in [196, 197].

Resilient backpropagation (RBP) is a first-order optimization algorithm of learning heuristic for supervised learning in FFNN which, similar to the LM algorithm, is one of the fastest weight update mechanisms. A compression based on quantum (QBP) is proposed in the BP learning rate concept, and it is primarily aimed at speeding up the error minimization because a high value of learning rate can cause a considerable change, causing the minimum to be missed while a very small learning rate slows the training unnecessarily. Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) is the most effective method for implementing a Quasi-Newton optimization technique. Liu et al. [11] established a hybrid model of EMD-ANN trained by four different algorithms of BP, RBP, QBP, and BFGS; and, the later method showed better performance. Based on this previous study and focused on preprocessing methods, three models of WT-BFGS, WT-ARIMA-BFGS, and WP-BFGS were proposed and compared in terms of accuracy in a multi-step ahead short term WSF [198]. Wavelet packet-ANN showed more accuracy comparing three evaluation criteria: MAPE, MSE, and MAE, with 3.55%, 0.5433, and 0.4396, respectively, for 30 minutes forecast horizon. However, WT-ARIMA-BFGS showed just a few lower accuracy than the WP-ANN model but recorded less computing time using ARIMA instead of ANN in subseries approximation. A hybrid Adaboost-MLP model trained by four different methods, including GD-ALR-BP, GDM-ALR-BP, CG-BP-FR, and BFGS algorithms proposed in [84]. The improved percentages of the MLP neural networks by the Adaboost algorithm increased step by step with the following sequence of training algorithms: GD-ALR-BP, GDMALR-BP, CG-BP-FR, and BFGS.

4.3.3.4 Optimization methods

Many different types of optimization methods are used in wind speed and power forecasting field for different components such as feature selection. However, they have been mostly applied for tuning the core model of the prediction. Genetic algorithm (GA) [196, 197, 52], particle swarm optimization (PSO) ([11, 187, 198]), simultaneous perturbation stochastic approximation (SPSA) [84], mind evolution algorithm (MEA) [199], artificial bee colony (ABC) [200], crisscross optimization (CSO) [201], chaotic shark smell optimization (CSSO) [202], and improved clonal selection algorithm (ICSA) [173] are applied as different optimization methods in WS&PF.

GA is used in a hybrid model of EEMD-GA-BPNN proposed by Wang et al. [203] to train a neural network model for short-term WSF. The proposed method showed satisfactory results with RMSE and MAPE values of 6.82% and 0.59 (m/s) for an ultra-short-term, and 8.08% and 0.71 (m/s) for a short-term, respectively. Sensitivity analysis on parameters

related to EEMD, namely white noise amplitude and ensemble number, can improve the forecasting results. Trashkar et al. [204] studied four different ANN methods of RBF, ANFIS, ANN-GA, and ANN-PSO to find the best model for short WSF in Tehran, Iran. ANN-GA overcame the others with 0.002 (m/s), 0.046 (m/s), and 0.95 values of MSE, RMSE, and coefficient of determination, respectively. In a comparative study, two different machine learning approaches with the main core of ANN were constructed and compared by Ak et al. [205]. Prediction Intervals (PIs) of short-term WSF were estimated by an MLP trained by a multi-objective genetic algorithm (MOGA) and combined to ELM with the nearest neighbor approach. Both algorithms showed good accuracy, high coverage, and relatively small interval size of PIs for winter and summer categorized wind speed datasets.

PSO was used to train a BPNN model developed by An et al. [68]. It is highlighted that the maximum iteration number of the PSO method should be chosen properly. Otherwise, the iteration could end prematurely, causing the neural network ill-suited for making the prediction. Jursa and Rohrig [192] investigated two different Machine learning models of near-neighbor search and a three-hidden layer neural network using two types of evolutionary optimization, PSO and DE, to predict short-term WPF in 30 wind farms. Results of 10 wind farms showed a significant improvement applying optimized ANN in comparison to the persistence model and manually selected input and several neurons of neural networks. However, for a few farms, any improvement was not observed. In another comparison study, Liu et al. [206] proposed two different models based on GA and PSO optimization methods attached with a Wavelet-MLP core. As a result, even though optimization algorithms improved the model's accuracy, they did not statistically improve MLP as much as wavelet decomposition.

SPSA was used to train an MLP model by Hong et al. [207], for short-term WPF. Three different multi-layers NN were constructed based on the two outputs: wind speed and wind power. The results nominated cascaded structure as the best model in terms of MAE, RMSE, and r^2 criteria. Besides, r^2, MAE, and RMSE, obtained by the BP-based MFNN with fuzzy inputs and time-interval averaging approach (TIAA), are comparable with those obtained by the SPSA-based MFNN with fuzzy inputs. However, the CPU times required for the BP-based MFNN with fuzzy inputs and TIAA are 7603 (s) and 214 (s), respectively, which are much longer than the times (25-35 s) for the SPSA-based MFNN with fuzzy inputs. Mind evolution algorithm (MEA) was used in a FEEMD-MEA-MLP model by Liu et al. [199] to select initial weights and thresholds for NN core trained by BFGS algorithm, outperform FEEMD-GA-MLP in a short term multi-step ahead WSF. ABC optimization method was used to train a BPNN network in a short-term WSF context, achieving more accurate results

compared to the use of GA [200]. CSO was applied by Meng et al. [201] in a WPD-CSO-BPNN hybrid model for short-term multi-step WSF. This comparative study showed the superiority of CSO compare to GA and PSO methods. A (6, 10, (1 or 3 or 5)) structure was applied as a BPNN model for one, three, or five steps ahead prediction using hourly wind speed data collected from a Netherland-based wind farm. MAPE (%) reported as 1.214, 2.9181 and 4.7443, RMSE (m/s) measured as 0.2216, 0.4525 and 0.7245 and MSE (m/s) reported as 0.1365, 0.2791 and 0.4895 for one step, three-step and five-step ahead, respectively. CSSO was developed and used in a BPNN engine WPF to find the optimum number of hidden layer neurons and enhance the NNs training performance's efficiency [202].

4.3.4 Feature selection

Most of the models applied for wind speed are applicable for the wind power and vice versa. ACF [17, 208] and PACF [185, 186, 162, 16] are widely used in literature, both individually and together [11, 140, 205]. Other methods are also applied, as mutual information [193, 150, 160, 162], information gain (IG) [202], ARIMA [153, 192], modified relief (MR) [193], kernel principal component analysis (KPCO) [146], Gram-Schmidt orthogonalization (GSO) [82], k-means [209]. A conditional mutual information (CMI) based feature selection was proposed by Wang et al. [12] to ensure that the selected features are individually informative and mutually weakly dependent. Besides the structure of the network and learning methods, feature selection is a crucial issue in supervised learning, which directly affects the accuracy of the results. A non- suitable FS can increase the computational cost of the prediction system and make poorer its generalization performance [125]: Individually informative and mutually weakly dependent. Using the CMI approach, MAPE of 1 hour ahead WPF was improved concerning other FS methods, CA, MI, RRrelifF, and mRmR by 15.5%, 21.0%, 19.7%, and 67.5%, respectively. Coral reef optimization was applied for a feature selection process by Salcedo et al.[124], and later improved using a harmony search (HS) by the same authors [183]. Ren et al. [68] investigated a novel approach of lateral and longitudinal data selection for an input parameter selection (IS)-BPNN-PSO to predict wind speed. The proposed parameter selection method comprised the following methods: The input dataset selection, input data dimension determination, and PSO parameters. Hecht-Nelson method $(2n+1)$ was used to determine the hidden layer neuron numbers. Mean daily and hourly wind speed data from two sites in China were considered as input, and the reported MAPE (%) was 15.51 (with 31 number of training) and 21.02 (with 124 number of training) for each site. Table 4.3 summarizes selected studies based on their completeness and superiority compared to other works.

Table 4.3 An error-comparison of short-term wind speed prediction models between 2012 and 2021. Terms with asterisks are described below the table

Model	No. Of Neurons	Activation Function (hidden layer)	Training Algorithm	Inputs	MAE (m/s)	MAPE (%)	RMSE (m/s)	Ref
WP-BFGS	6,13,1	-	BFGS	WS	0.0416	0.21	-	[198]
SOM-MLP	3,7,1	hyperbolic sigmoid	-	WS, WD, T	-	-	0.0828	[117]
EMD-BFGS	4,9,1	-	BFGS	WS	0.1293	0.69	-	[11]
WPD-CSO-BPNN	6, 10, (1/3/5)	sigmoid	BP-CSO	WS	0.1365	1.214	0.2216	[201]
EEMD-GA-APSO-WNN	-	WNN(with morelet wavelet function)	GA-APSO	WS	0.1862	2.42	-	[174]
FEEMD-(GA or MEA)-MLP	6,12,1 with(GA)	-	BFGS Quasi-Newton Algorithm	WS	0.2944	2.53	0.3884	[199]
EEMD-GA-BPNN	(6,10,1) (3,5,1)	-	BP	WS	-	8.08	0.71	[203]
NARX-ARIMA	5,10,1	sigmoid	LM	WS, WD, T, P, (R/H)	0.43	-	-	[158]
EMD-Elman NN	24,50,1	hyperbolic tan sigmoid transfer	BP	WS	0,6175	20	-	[152]
WTT-SAM-RBFNN	24, ?, 1	RBF	steepest GD	WS	0.65	23	0.88	[164]
EEMD-(T-S) fuzzy NN-CS	-	ANFIS	CS	WS	0.66	6.47	0.92	[85]
Adaboost–GD-ALR-BP–MLP	6,12,1	-	GD + Adaptive Learning Rate BP	WS	0.6797	8.11	-	[210]
NARX	(4,9,1)	sigmoid	GD	WS, SR	0.8614	11,448	-	[156]
SAM–ESM–RBFN	-	RBF + Gaussian function	steepest GD	WS	-	12	0.34	[142]
compariosn (MLP, GEP, Regressio	8,8,1	logarithmic sigmoid	BP	WS	-	-	0,899	[140]
comparison of MLP & RBF	4-7-13-1	Tansig–Logsig	LM	WS	-	-	1.19	[169]
MLP-SPSA Fuzzy input	3,2,1	sigmoid	SPSA*	WS	0.9472	-	1,3662	[207]
ANN-YAR	3,2,7 W* 2,3,12 S*	tansig	LMBA*	WS,T	1,4031	-	-	[133]
AWNN (adaptive)+ERA	7,?,1	Multi-dimensional wavelet	GD	WS	5,7822	-	7,0679	[175]

Nomenclature: AE: Applied Energy; N: Neurocomputing; RE: Renewable Energy; EC&M: Energy Conversion and Management; IET(RPG): IET Renewable Power Generation; ASC: Applied Soft Computing; NC&A: Neural Computing and Applications; IS: Information science; MAP: Meteorology and Atmospheric Physics, International Journal of Electronics and Electrical Engineering; E: Energy; W: Winter; S: Summer; SPSA: Simultaneous Perturbation Stochastic Approximation; LMBA: computing based on least square fitting technique and a combination of Gauss-Newton algorithm and gradient descent technique.

4.4 Hybrid and ensemble models

In this section, different hybrid and ensemble models of WS&PF are discussed. An ensemble model is constructed using a set of machine learning techniques to train a portion of a given problem and integrate these machine learning techniques to solve the whole problem. Hybrid models are defined as any possible combination of pre-processing, core machine learning models, and post-processing items, as it is shown in figure 4.3. An intensive literature review on hybrid methods in wind speed and power forecasting is done by Tascikaraoglu, and Uzunoglu [79]. Figure 4.4 shows a scheme of ANN combined or ensemble models.

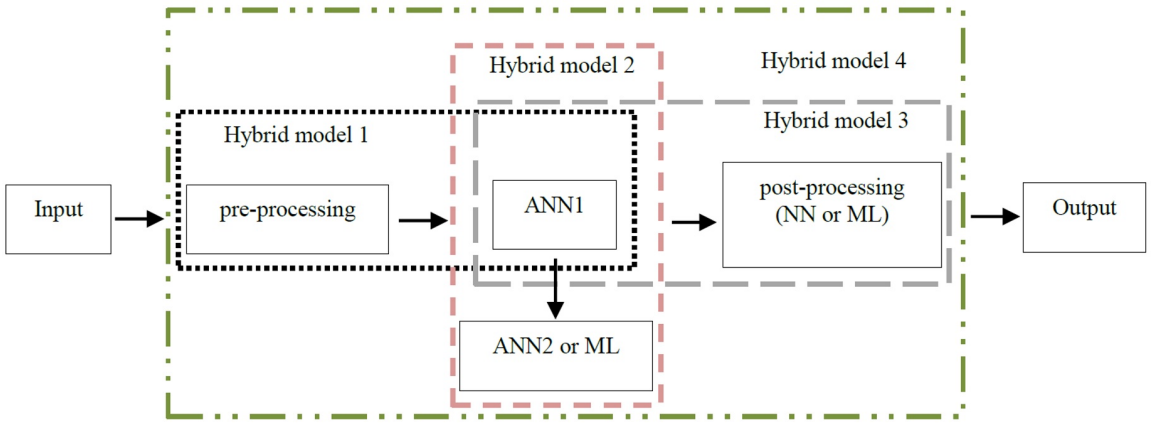


Fig. 4.3 A general scheme of ANN hybrid models

Buhan, and Çadırcı [211] proposed a multi-stage ANN-SVM hybrid model for WPF of 25 wind farms in Turkey, up to 48-h ahead. First, forecasted wind speed and direction from three different NWP sources were used as input dataset, and also to select the best numerical grid that gives the minimum MAE for each WPF. Second, combined SVM-ANN models improved the minimum MAE prediction; finally, the forecast errors were corrected

by applying model output statistics (MOS). For the 25 WPFs, an improvement range between 26% and 56% in NMAE over persistence was obtained. Zhang et al. [16] proposed a one-step-ahead turbine-based ensemble model to predict short-term wind power. EEMD was used as a pre-processor, and intrinsic mode function (IMF) components, except residue, were predicted by using four different methods: Add-weighted One-rank Local-region method (AOL), ESN, RELM, and SVM. The residue prediction was obtained from the nonlinear grey Bernoulli model based on PSO. A combined forecasting model based on optimal virtual prediction was developed in which a self-adaptive differential evolution algorithm (SADE) was used to optimize the model weight matrix. MAPE, NRMSE, and NMAE values reported for the second analyzed wind were 18.11%, 7.07%, and 4.91%, respectively.

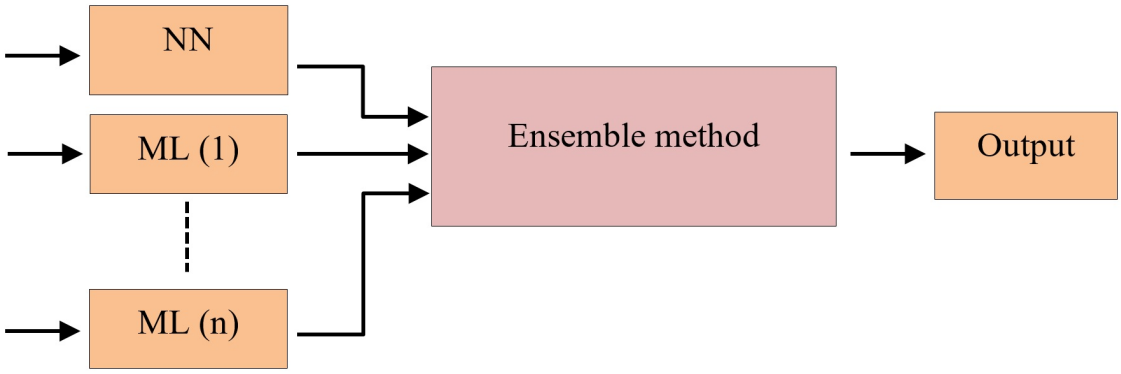


Fig. 4.4 A general scheme of ANN combined models

Liu et al. [16] extended this previous method to multi-step WPF with a core of LSSVM-ESN-RELM. Variation mode decomposition (VMD) was used as preprocessing method, and (SAMDE) algorithm was applied to dynamically adjust the weights matrix of each predictor in the optimal virtual prediction scheme. Besides, the quantile regression averaging (QRA) method was used for a probabilistic interval prediction based on deterministic forecasting. The results showed QRA-VMD superiority against QRA-EEMD. Their simulation results indicated that the NMAE of the proposed combined model from one-step to three-step forecasting were 4.34%, 6.49%, and 7.76%, respectively. A new traditional combination method based on No-Negative Constraint Theory (TCM–NNCT) was proposed by Wang et al. [212] for medium-range WSF. Different combinations of ARIMA, ARCH, BPNN, SVM, and KF, as commonly used machine learning methods, were tested in new proposed and traditional combination models. Derived results at three turbine locations at Chengdu province in China showed the new proposed model's superiority against traditional models. Besides, two NNCT-based combination models, CPSO–CM–NNCT and GA–CM–NNCT,

were compared to the new TCM-NNCT model, showing that the former model was better, and the latter showed high volatility.

Wang and Hu [71] worked on a hybrid model of probabilistic short-term WSF. Empirical wavelet transform (EWT) was used as pre-processor, and Gaussian process regression (GPR) was used to combine independent forecasts generated by ELM, SVM, LSSVM, and ARIMA. The deterministic model results for a 15-minutes ahead WSF reported as 7.75%, 0.46(m/s), and 0.46(m/s) for MAPE, MAE, and RMSE, respectively. The effect of using EWT was investigated, and at least for one step ahead, this method improved results accuracy. Bouzgou and Benoudjit [25] proposed a multiple architecture system (MAS) of MLR-MLP-RBF-SVM through three fusion strategies, simple, weighted, and non-linear, for WSF. The proposed method was tested in 7 different sites in Algeria for ten years, and in all the experiments, the three proposed fusion strategies presented better performance than single predictors. Applying a Bayesian combination forecast (BCF), Li et al. [52] developed a hybrid model of adaptive linear element network (ADALINE), two hidden layer BPNN, and RBF, for WSF, and tested it in two sites located in North Dakota, USA. In order to define the posterior probabilities of component models by estimating the corresponding model parameter vector, Box-Cox transformation was used to transform the non-Gaussian skewed distribution of wind speed data to a Gaussian distribution, and also to perform the data transformation required by the implementation of EM algorithm, which was applied to each observation-forecast combined time series. A two-stage wind power forecasting method was proposed by Bhaskar and Singhto [208]. For the WSF step, wavelet decomposition of NWP wind series was applied, and an adaptive wavelet neural network (AWNN) was used to regress upon each decomposed signal. As the second stage, this forecasted wind speed was used for WPF through a non-linear input-output mapping using FFNN.

Okumus and Dinler [108] proposed a hybrid model of ANFIS-FFNN-PER for short-term WPF inside a brief review of wind power forecasting studies. A linear equation of ANFIS, ANN, and persistence results with different weights was constructed, and the minimum least square error was applied. The proposed model outperformed the persistence prediction by 5% MAPE on average. Their results also showed that error rises to 15% MAPE for between 6 and 24-hours ahead predictions, although a value lower than 10% should be a good challenge.

The proposed model gave an average NMAE and NMSE of 7.08% and 10.22%, respectively, for 30 look-ahead hours over one year of evaluation. For each decomposed signal, a specific architecture of AWNN and FFNN was defined. Amjadi et al. [193] trained three

MLPs by LM, BFGS, and learning algorithms optimized by enhanced PSO for WPF at Alberta, Canada. A RBF model was applied as an auxiliary predictor. Besides, the main predictor of the model was able to capture both global and local behaviors of the target variable. Different combinations of NNs, optimization algorithms, and feature selection methods were compared in terms of average RMSE (MWh) and normalized mean absolute error NMAE, and the proposed model resulted in 4.06 and 2.21 values, respectively.

Ranaganayaki et al. [213] constructed an ensemble NN with MLP, BPNN, PNN, and Madaline (adaptive linear neuron network), focusing on the efficient number of hidden neuron numbers by satisfying 102 defined criteria through convergence theorem. To fix the number of neurons in the hidden layer, computed mean square error (MSE) of all the tested criteria compared and $(6n+7)/(n-3)$ showed a better operation in which n indicates the number of inputs. The ensemble model averages the output of each NN to produce the ensemble output. An adaptive boosting (AB), optimized BPNN model combined with FEEMD was proposed by Wang et al. [214]. Flower pollination algorithm and conjugate gradient (CG) (CGFPA) are used to optimize the BPNN model. AB strategy integrated with BP neural networks was adopted to overcome the uncertainty of the outcomes that can be set to the randomness of the initialization of the BP neural networks. After that, Li et al. [12] proposed an ensemble NNs model based on decomposed wavelets. Conditional mutual information (CMI) was used as feature selection of each sub-series of wavelets and the forecasted MLPs ensemble using a partial least square regression (PLSR) method for short to medium-term WPF.

In a hybrid model (k-means cluster algorithm and GRNN), Dong et al. [215] proposed a day-ahead WPF. NWP data were categorized based on daily similarity, and samples in the subset similar to the predicted day were chosen as training samples to build the GRNN model. This ANN model is made based on mathematical statistics that can approach the mapping relationship according to the sample data. Even though a few sample data are available, the network output can converge to the optimal regression surface. Moreover, only the parameter θ needs to be determined in GRNN, avoiding the effect of subjective assumptions on forecasting results. In comparison to the persistence model, the results' accuracy improved by 11.45% applying the proposed model.

4.5 Discussions

Forecasting of wind speed and wind power generation is multi-disciplinary research in which computer science, meteorology, statistics, and power systems engineering have to be

applied to achieve a reliable prediction. Recently, computer science techniques, i.e., machine learning methods, are playing an essential role in comparison to conventional methods, thanks to the fast and robust development of information technologies. Artificial Neural Networks (ANN) are the pioneers in artificial intelligence application in WS&PF, and, due to their potential, they would continue to be one of the most important methods in the future in this field. Also, Big Data analysis is one of the most successful and promising techniques which should get more attention. ANNs can be divided into two main categories, namely shallow and deep neural networks, with the latter one as the newest and promising approach because of its connection to the emergence of Big Data analysis. Following, in the present work, a number of significant aspects are identified in the literature as well as shortcomings with solutions are recommended.

4.5.1 Shallow neural network

Most of the published works on WS&PF are categorized as a shallow neural network, addressing manual structure mining or feature extraction. In the following, different items of shallow learning are discussed, and some solutions are proposed to improve their efficiency.

4.5.1.1 Inputs

As one of the hot spot issues in WS&PF, the input dataset plays a vital role in the accuracy of the results. Wind speed is the most used parameter in ANN models to predict both wind speed and wind energy. Other variables fed into the ANNs models are wind energy, wind direction – sin and cosine values or complex values –, day of the year, month, humidity, air temperature, soil temperature, minimum and maximum temperature, air density, pressure at different levels, and solar radiation. Each of the parameters above has been considered in different time steps of its time series. In some models, their input dataset comes from the output of other models, e.g., NWP [211]. However, in other models, raw collected observations are fed directly into the model. Some studies developed an input-dataset structure to find the optimum input series in fewer errors [11]. Of course, the defined prediction metrics for the ANN models change with the influence of their inputs.

Besides the variety of parameters used in previous studies, other variables still can be considered, such as wind shear or wind speed gradient, terrain slope index, latitude, longitude, average cloudiness, rain patterns, reference clearness index, and so on. Therefore, efforts to assess the influence of these parameters in the models performance are not yet considered. It is crucial to notice that selecting appropriate data between a large number of different

possible sources of input data has a direct effect on the accuracy of the model results. On the other hand, more input sources require more effort in selecting the most appropriate, both in terms of people involved and or computational cost.

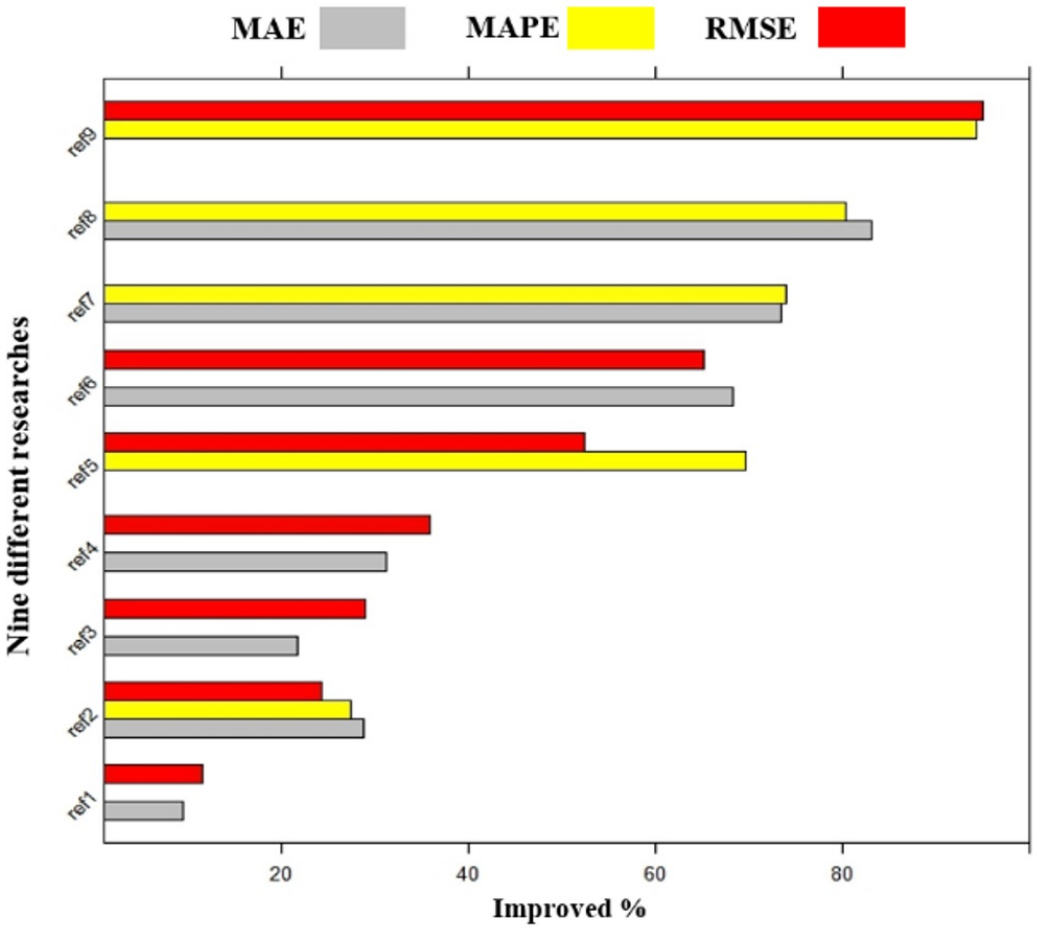


Fig. 4.5 Improvement Percentage (IP) of short–term WPF models against persistence model (PM) in terms of mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE) in nine different studies, 2010-2016. ref1: [173], ref2: [178], ref3: [207], ref4: [17], ref5: [202], ref6: [208], ref7: [85], ref8: [161], ref9: [216]

4.5.1.2 Preprocessing

As a preprocessing method, in WS&PF different signal preprocessing techniques was applied, such as WT [194], EMD [11], EEMD [194], WD, WPD, and FEEMD [153]. Reported results support their positive effect in increasing the accuracy of the results compared to the same model results without signal preprocessing techniques. In most studies, the only shortcoming is that multiprocessing techniques were limited to the input parameters, which

itself is specified to wind speed and wind power. However, due to their influential role in the results, it is worthwhile to apply them to other input variables.

4.5.1.3 Feature selection

The feature selection methods discussed in section 4.3.4, are categorized as manual feature selection in a way that some inputs consider, and through different methods, a defined number of inputs is selected to feed the model. However, the combination of these techniques with signal preprocessing improved the accuracy of the models. Still, due to the limited number of inputs or processes, these methods can not handle the necessity for unsupervised feature learning brings up [120]. Applied to this issue, deep neural networks (DNN), along with the big data concept, can provide more accurate predictions in a way that the more unlabeled data we use, the better the learned features will be.

4.5.1.4 Prediction metrics

A majority of the reviewed literature used mean absolute error (MAE), mean absolute percentage error (MAPE), mean square error (MSE), and root mean square error (RMSE) to evaluate the accuracy of their proposed models. However, some other individual error indexes were applied in some studies.

The persistence model (PM), as the simplest and reference prediction model, is not reported in a considerable number of researches. Still, the accuracy of the model is compared with other models in terms of different error metrics. It seems that calculating PM metrics is an essential step in any WS&PF model assessment. First, the constructed complex model should prove its predominance against PM, as if this condition fails, the model does not have any practical value. Second, by reporting PM errors, the different atmospheric conditions of the wind power plant location should help to compare the accuracy of the models based on the atmosphere stability index of the area on which the model is applied. Third, the Improvement Percentage (IP) of the proposed model against PM could be considered a reliable index of the superiority, in other words, the model's success. Since IP is not dependent on atmospheric stability, it is easier and more accurate to compare different proposed models - no matter where they are applied - in terms of accuracy. In figure 4.5, the IP index of other WPF models against PM is depicted for three different metrics, depending on the reported study. Not all of the studies could be taken into this comparison as either not saying PM or not considering at least two evaluation criteria of MAE, MAPE, or RMSE, which is recognized as the most commonly used metrics in the literature.

4.5.1.5 Limitations of shallow ANN models

Overtraining of network and extrapolation errors are mentioned as two main ANNs limitations in [72]. Over-training happens when the model does too many training iterations. The number of training cycles and training data should be optimized to avoid such a problem. Extrapolation error occurs because of the deficiency of ANN models in extrapolation beyond the data. Training data needs to be selected to represent the entire operating range of the modeled system to reduce this error source.

4.5.1.6 Network definition and parameter optimization

From the shallow neural network point of view, defining the structure of the network, activation function and learning approach is an important task in ANN modeling, besides optimal parameters selection of the network, such as the number of hidden layers, number of neurons in hidden layers, learning rate, and number of training data.

To define the structure of the network, the main applied models were discussed in section 4.3. Many studies compared their proposed structure with other possible model structures. Usually, the proposed models are primarily simple and do not enforce with different techniques like the preprocessing, optimization, or feature selection algorithms. However, combined models gather the advantage of different structures, and they can be considered a good option to be chosen. A robust model by a combination of different forecasting models is required to overcome any inconsistency issue due to model structure selection [171].

After selecting the network structure, to overcome other problems like the number of the neurons in hidden layers or learning rate, different optimization methods were discussed in subsubsection 4.3.3.4. GA and PSO are commonly identified as the best optimization methods for ANNs. Still, other creative algorithms have also applied, and their results even outperformed GA [199], PSO [173], or both of them [150, 201]. The fact is that different algorithms behave differently in different datasets. Therefore it is required to test other optimization techniques to find out the most successful method for each specific dataset, both for defining the number of hidden layers and the optimal input layer, for instance, [31].More than optimization techniques, some empirical equations, shown in table 4.4, have been proved helpful in the literature to define the number of hidden layer neurons.

Table 4.4 Empirical equations for hidden layer neuron numbers. L is the number of hidden neurons, n indicates the number of inputs, m shows the number of outputs, and T is the number of training vectors.

Equations	Ref.
$L = n$	[140]
$L = 2n$	[151, 158, 206]
$L = 2n+1$	[212, 137, 152, 77, 214, 11, 68]
$L = round(\sqrt{(m+n)} + rand(1 \sim 10))$	[82, 206]
$L = \frac{1}{2}(m+n+\sqrt{T})$	[217]

4.5.2 Deep neural networks

Big Data computing falls into two major categories, based on how data is analyzed concerning the time constraint. First, batch processing of large volume of on-disk data with no time constraints (e.g., MapReduce), and second, the streaming process of in-memory data in a real-time or short period. To make this process, there are several computing frameworks, e.g., Hadoop, ComMapReduce, and IBM parallel machine learning toolbox. Such systems have the ability to scale up machine learning [218]. The combination of deep learning and parallel training implementation techniques provides potential ways to process big data [219]. About deep neural networks (DNN), it is crucial to consider that feeding DNN with a small training dataset produces a prediction less accurate than using shallow learning methods [189]. Therefore, DNN should be used only when large input datasets are available; therefore, DNN must be only applied jointly to big data processing techniques. In addition, DNNs are more expensive in terms of computational cost, although parallel or distributed deep learning can overcome this problem. Considering all these issues between ANN models, the combination of big data processing and deep neural networks is a promised new technology to improve WS&PF.

4.6 Summary

Among the various data-driven methods, artificial neural networks (ANN) are mainly applied and often give more satisfactory results. However, the selection of their types and parameters could significantly affect the prediction's accuracy. So their successful applications due to available choices of structure and parameters are case-dependent. This chapter

presented an overview of the state-of-the-art ANN technologies in wind speed and power forecasting. In addition, Different types of architecture, learning algorithms, activation functions, feature selection methods, and hybrid and combined models using ANNs were discussed. In light of this extensive review, in the following chapters wind speed and power prediction models are developed. As an important conclusion, the accuracy of ANN models is not only dependent on the network layers, characteristics, and the models coefficient, but on the whole process of the prediction: including acquiring, validating, and reconstructing data, time horizon, single or multi-steps ahead forecast and the stability index of the local meteorological parameters, all have their substantial effects on the prediction accuracy.

It is not easy to compare the overall performance of different models due to the lack of a unique evaluation standard. However, a comparison is provided for several studies in section 4.5 considering the same criteria. It can be inferred that adding preprocessing and optimization techniques to the ANN core of the model can improve the forecasting accuracy. In this Chapter, the improvement percentage of the evaluation criteria between the proposed and persistence models is considered as a new criterion for models accuracy, which advantage is eliminating local weather conditions and characteristic effects on the models output. In other words, by classifying the prediction models based on weather and climate data characteristics, i.e., stability index (SI), comparing the models in each category makes more sense than making a global and general comparison. As a general fact, applied models on unstable atmosphere perform worse than those applied on stable and more stationary time series. So comparing the results models being applied on different atmospheric condition is a strategic error which most of the review articles have committed.

Similarly, it is not easy to single out any specific shallow ANN model as the best model. That is because the results are highly dependent on the dataset characteristics. However, unsupervised learning might get more attention considering the emergence of big data concepts besides combined and hybrid models.

Chapter 5

Weather data quality evaluation and reconstruction

In this chapter, the quality of weather data, particularly wind speed time series, is examined through a validation quality assurance process. The missing or low-quality data are replaced through a proposed reconstruction algorithm. The chapter is comprised of two main sections: the meteorological data validation, and the reconstruction algorithm and its application. Previously, the necessity of quality assurance in meteorological data is introduced. A validation process, inspired by other relative works in the literature, is introduced right after, followed by the proposed reconstruction algorithm that uses both observational and WRF data to impute gaps in weather time series. The derived reconstruction model is successfully evaluated compared with other benchmark models. Finally, this reconstruction model is applied to the meteorological parameters required in the different case studies, to impute the missing or invalid data points.

5.1 Introduction and background

Wind speed and power prediction models need high-quality observational data to provide reliable predictions. The weather sensors, that are the main sources of collecting observational data, should be reliable enough to capture and provide high-quality data under different meteorological conditions such as high-temperature difference, solar radiation, strong winds, storms, etc. A sensor observes a physical parameter along the time as a time series, then converts it into an electric signal and finally into a digital value called raw data. Meteorological stations observe, collect, and transfer data to a datalogger, while wind turbines also have a system to handle this process through supervisory control and data acquisition systems

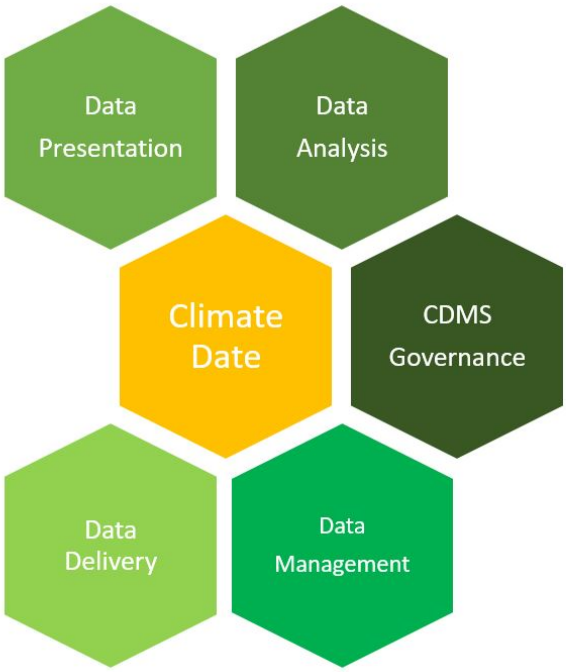


Fig. 5.1 A component-based graphical depiction of the climate data management system (CDMS)

(SCADA). Data capturing and transition fulfill in a specific time interval depends on the instrument specification and the potential of the transition system (from seconds to hours).

5.1.1 Data validation

Referring to the WMO guideline [38] for climatological practice, any selected process for meteorological data acquisition, processing, transmission, and monitoring should addresses the following concerns: reliability, accuracy, design simplicity, and suitability for operating in the environment. Reliability is a factor of long-term and continuous functioning and being robust to be working in weather extreme to provide qualified data and prevent data gaps, biases, and other inhomogeneities in data series. Instruments should also have a high level of accuracy and maintain this factor during the time as instrument drift can cause severe deterioration in a climate record. Simpler instruments are easier to use and have less operational and maintenance costs. On the other side, complex systems can easily cause data inhomogeneities and data losses. Hence, to cover the whole process of climate data collection, transformation, storage, tailoring, and visualization in a secure and quality assured way, a climate data management system (CDMS), shown in figure 5.1, is required [38].

To ensure the quality of acquired meteorological data, some fundamental standards need to be carried out as a part of the CDMS process: a) Proper station sitting to avoid obstacles, artificial heat sources, site slope, and influence of irrigation. b) Adequate and regular site maintenance to physically keep the station clean from dust, vegetation encroaches, etc. c) Routine calibration of sensors to avoid any possible drift in data acquisition. d) Storaging of the original data and flagging its quality instead of changing the suspicious observations. e) Use standard time (e.g., UTC) and observation units (e.g., m/s) to ease the intercomparison process between the close stations. f) last but not least, use of redundant sensors if possible to compare between two or more sensors of the same type [220].

A rich library of weather data validation procedures can be found in the literature. Ehrlinger et al. [221], developed a quality assessment for integrated information environments, namely QuaIIe, explicitly designed for imperfect data, using general quality assurance (QA) tests for quality control processes such as accuracy, completeness, and pertinence in data. Fiebrich et al. [220], presented the mainly used automated QA tests that can be used for every weather parameter such as sensor and climate-based range tests, temporal checks including step and persistence tests, particular checks, consistency checks, and adjustment check. Some variable-specific tests are also provided as modifications of or additions to general tests. Gandin [222], provided complex quality control (CQC) procedure for meteorological data as an integral part of a data assimilation system. In the proposed method, the so-called rough errors caused by malfunctioning instruments or by mistake during data processing, transmission, or reception are classified according to their origin. The provided method is pointed exclusively against these large error values. Data errors can arise primarily in each of the following processes: instrumental observer, data transmission, key entry, and data validation processes, as well as changing data formats and data summarizing.

The main objective of data quality control is to verify if an acquired data value is representative of what it is supposed to be, i.e., whether it is contaminated or not. In a CMDS system, the quality control steps should be taken before recording or transmitting an observation. However, this process should also be applied in the storage phase to prevent any transmission error when data is transferred from the observational station to the center. The validation process is mainly automatized. However, it is highly recommended to include a human confirmation step before setting an invalid label to a value. Providing a set of quality tests and potential errors to a human analyst is required to judge the cause of errors and determine any correction that should be applied. All observations should be appropriately flagged. Table

5.1 shows an example of data quality codes provided by WMO [38].

Table 5.1 Data quality labeling codes for meteorological data based on WMO standard

Data type code	Interpretation
0	Original data
1	Corrected data
2	Reconstructed data
3	Calculated data

Yang et al. [223] proposed an ANFIS-based interpolation model for missing wind data using wind speed and wind direction data from other altitudes. Two indicators, integrity and reasonableness of data, were defined to check the integrity rate of the observational data and check the range of value, correlation, and trend of data in a validation process. Estévez et al. [224], developed a validation procedure guideline for data collected from agroclimatic information networks in Andalusia, Spain. Different tests were used in the validation process, such as range test, limits check, time and internal consistency tests, persistence check, and spatial consistency tests. The provided guideline is broad enough to be applied to all climate data for different applications in agriculture, hydrology modeling, etc.

Another QA procedure for mesoscale meteorological data was proposed by Feibrich et al. [220]. A complete set of tests, including general automated QA tests and variable-specific QA tests were provided, justifying that not all the general tests are applicable for all variables. Hence, some modifications of or additions to the general tests for particular variables are required. Range, temporal, spatial, and internal consistency tests were discussed for temperature, air pressure, relative humidity, soil moisture and temperature, rainfall, solar radiation, net radiation, and winds. Jiménez et al. [225], provided another QA procedure for surface wind observations. Four stages of validation tests were introduced to test manipulation errors, limits consistency, temporal consistency, and biases in the wind speed and direction time series, with a detailed discussion on abnormal low and high variations tests, including temporal consistency. The method was applied successfully on 41 automated stations from 1992 to 2005 in a complex terrain region located in the northern part of Spain.

5.1.2 Data reconstruction

Besides data validation and quality control procedures, some works in the literature have focused on the reconstruction part of missing, wrong, or low-quality data. Before

reviewing the successful investigations in this field, it needs to be clarified that reconstruction, assimilation, and imputation terms, frequently used in the literature, share the same meaning. From a general perspective, we can categorize missing data reconstruction methods into seven groups, which are listed below [226]:

- *Method of ignoring instances with unknown feature values:* Ignoring the instances, which have at least one unknown feature value.
- *Most common feature value:* The value of the feature that occurs most often is selected to be the value for all the unknown values of the feature.
- *Concept most common feature value:* the value of the feature, which occurs the most common within the same class, is selected to be the value for all the unknown values of the feature.
- *Mean substitution:* Substitute a feature's mean value computed from the available values to fill the gaps.
- *Regression or classification methods:* Develop a regression or classification model based on complete case data for a given feature, treating it as the outcome and using all other relevant features as predictors.
- *Hot deck imputation:* Identify the most similar cases in terms of missing and not-missing values to substitute the missing value.
- *Method of treating missing feature values as special values:* Treating "unknown" itself as a new value for the features that contain missing values.

Machine learning methods can be added at the end of the above list, which is considered as a research gap in the weather data reconstruction literature. Indeed, most of the developed ML techniques have been applied in the prediction problem. A few works are specifically developed to deal with weather missing data and imputation problem. Below, the most relevant recent results are discussed.

Liguori et al. [227], developed an autoencoder neural network for indoor environment time series missing data reconstruction. Indoor air temperature, relative humidity, and CO_2 were used during a four-year monitoring campaign from 84 different rooms. The RMSEs for the reconstructed parameters were reported as $0.42^{\circ}C$, 1.3 %, and 78.41 ppm, respectively. Amiri and Jensen [228] proposed a general fuzzy-rough model for missing data imputation,

which can be applied to different environmental dataset.

Dealing with weather and climate missing data, some methods were developed for specific weather parameters such as solar radiation or rainfall. Demirhan and Renwick [229] compared 36 imputation methods for solar radiation time series of minute-based, hourly, daily, and weekly time resolution collected under 16 different experimental conditions in Australia. Monte Carlo simulation was used to generate missing values randomly. The weighted moving average method overcame other methods for long-term imputations: daily and weekly. However, linear and Stineman interpolation and Kalman filtering with structural model and smoothing provided better results for minutely and hourly imputations. In another study, a k-nearest neighbor (KNN) method was used for weather data imputation (wind speed, wind direction, solar irradiation, temperature, precipitation, humidity and cloud cover, both monthly and hourly) before feeding them to a photovoltaic (PV) forecasting SVR model. Its prediction performance yields results close to those for the original data with no missing gap [230]. Zhang et al. [231] proposed a generative adversarial network (GAN) for multivariate solar data imputation with at least 23.9% reduction of MSE for up to 40% data gaps. A hybrid technique of k-nearest neighbor, weighting KNN methods, and other regression methods used for processing missing rainfall data in Algeria [232]. In a comparative approach, Miro et al. [233] compared 10 different linear, non-linear, and hybrid imputation models for daily and monthly rainfall data, where particle component analysis methods outperformed the self-organizing map (SOM) significantly.

Dealing with all types of weather data missing gaps, particularly wind speeds, some traditional and data-driven imputation techniques have addressed the problem. A low-rank matrix approximation model for imputing wind speed missing data overcame other traditional imputation techniques such as mean, multiple imputations, and KNN [234]. Alsaber et al. [235], used multiple imputation (MI) approach to deal with different levels of missing air quality data from 5% to 40%. NO_2, CO, PM_{10}, SO_2 and O_3 parameters where used for reconstruction and climatological data as control variables for better estimation. A combined statistical method using Kalman smoothing, ARIMA, and multiple linear regressions was used to complete hourly wind speed, temperature, and humidity of four locations during 12 months at the coast of Western Australia [236]. In a comparative study, Flores et al. [237] developed six recurrent neural network models, such as LSTM and gated recurrent unit (GRU) with different sequential layers, to impute hourly wind speed, wind direction, and temperature time series. A competitive performance was reported between the developed method and the moving average technique. However, between LSTM and GRU- based mod-

els, the first one has shown greater heterogeneity in results. In another study, deep neuron networks, combining convolutional neural networks (CNNs) and bi-directional convolutional long-short term memory networks (BiConvLSTMs), developed by Gao et al. [238], was used to reconstruct missing data of weather radar image sequences. The model proved robustness, uninfluenced by less qualified input data, dealing with arbitrary missing patterns.

Addressing the corrupted SCADA data problem in wind farms, Yang et al. [239], developed a data reconstruction model using another type of data-driven technique, support vector regression (SVR). Temperature signals relevant to wind power faults were used to estimate generator drive. In another study, a de-noising auto-encoder (DAE) model developed by Ahmed et al. [240], was applied for the reconstruction of original sensor measurements from a non-reliable SCADA system under a cyber-physical attack. Khan et al. [241], established a generalized model to reconstruct missing or corrupted SCADA data of wind power using machine learning techniques based on linear regression, random forest, and CGP evolved neural networks. Although random forest results were better for a particular wind turbine, but GCPANN showed a higher generality and performed better when trained on one turbine and tested on others. Focusing on different missing data scenarios, Liu et al. [242], developed a wind power prediction model using a probabilistic nonparametric mixture of Gaussian function to define a distribution of plausible values for each sample. The expectation-maximization algorithm was then used for parameter estimation. The proposed multiple imputation (MI) outperformed the deletion method, mean substitution (MS), and k-nearest neighbor (KNN).

5.2 Developed validation-reconstruction methodology and methods

Figure 5.2 shows a general schematic view of the new proposed methodology to detect and reconstruct unreliable and missing data in this work, with the aim of providing high-quality input data for wind speed and power prediction models. This procedure is comprised of three main steps for inconsistent data: First, time series are checked using validation techniques to find missing and unreliable data. In the second step, all the suspicious data are labeled as missing data. In the third step, missing data are reconstructed using a rebuilding model. A feed-forward neural network in conjunction with k-means clustering for the input data is developed for the last reconstruction step, as it is explained with more details in subsection 5.2.2.

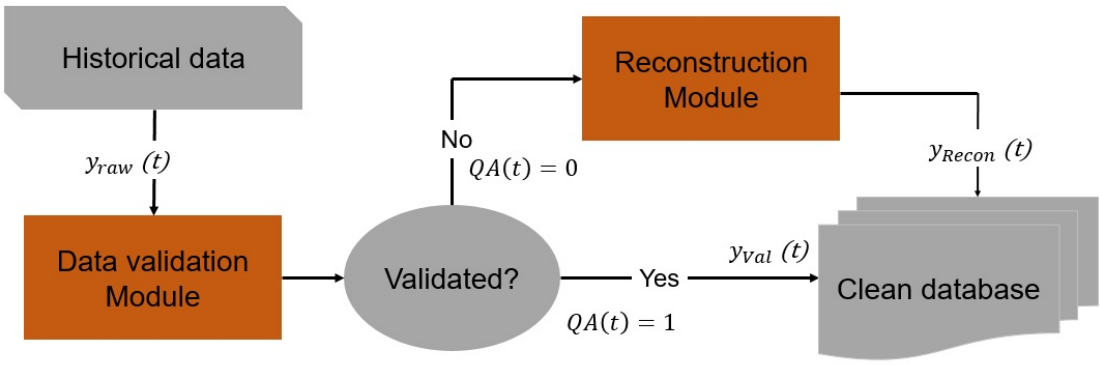


Fig. 5.2 Schematic depiction of the proposed data validation-reconstruction methodology

5.2.1 Data validation module

In general, observational data errors are categorized into two groups: random and systematic errors. Random errors are inherent to the measurement, and they are influenced by many factors. The instrument might be the origin of random errors as all measurements are approximate by their very nature and are always recorded with some errors. Random errors are more or less symmetrically distributed with zero mean $\mu_{RE} = 0$, assumed to be described by Gaussian distribution. It is impossible to exclude random errors from data, so the only possibility is to consider noise level when dealing with the data. Contrary to random errors, systematic errors or biases are not symmetrically distributed with respect to zero and usually their mean values differ significantly from zero. A scale shift of the instrument which requires calibration is considered as the main systematic errors origin. Systematic errors are usually persist in time, allowing to determine them by some quality control test using time-averaged data. In meteorological data, the situation is even more complex as they vary in both time and space. In addition, many observational instruments are remote from the points where they have received for processing. Micrometeorological errors are deviations caused by small-scale perturbations and manifest themselves as white noise. So random errors in meteorological data are composed of observational and micrometeorological errors. The noise level, σ_e, is therefore more than that of random observational errors [222].

Inspired by the previous studies, a two-stage validation module was developed, including general and variable-specific parts extracted from the literature. The general or basic test is applied to all-weather parameters. However, some value-specific tests, mainly for wind speed and wind power, were applied as the second part of the validation procedure. The

different validation tests used in this work are introduced below. Derived validation codes are defined in table 5.2.

5.2.1.1 Basic test

A basic, general check is necessary to determine any missing value or duplication in the parameters' date. This kind of problem may occur at the time of storage. This test is applied to all the parameters and is not value-specific. More information is provided in table 5.3.

5.2.1.2 Range test

The range test is based on both the sensors' specifications and the extreme values of meteorological parameters in a location. So both sensor-based and climate-based limits should be considered in this test. The climate-range test distinguishes into two types of fixed and dynamic limits. Considering the fixed type, global established extreme values of a parameter take specific limits, and any observation that occurs outside of an allowable range is flagged as invalid. However, dynamic ranges are site and time-specific. In this work, fixed limits for each parameter were adopted, following previous works listed in table 5.3.

5.2.1.3 Internal consistency test

Considering several weather parameters at the same station simultaneously or different parameters at various geographical locations, consistency test checks if they are reasonably integrated. Any violations of logical or physical relationships between two or more parameters, P, such as $P_{min} > P_{max}$, will fail the test, and the value label changes to invalid. For example, maximum temperature on one day can not be less than the minimum observed temperature for the immediately following data [243].

5.2.1.4 Temporal consistency test

Temporal tests will check the validity of fluctuations in a weather parameter time series at a specific station [244]. Both step and persistence tests are categorized as temporal consistency tests. In step tests, changes in magnitude in sequential data points are compared due to different time intervals. The threshold values are more subjective than range tests, and usually they are climate-based dependent, varying from one climate ragtime to another. For example, a decreased air temperature of 9°C in five minutes was reported in Oklahoma, the U.S., in a low front system [244]. In the persistence test, the minimum fluctuation of the observations alongside time is taken as criteria. For example, ice accumulation on a wind

Table 5.2 Data validation codes applied in this work

Validation code	Interpretation
1	Valid data
2	Missing value
3	Invalid data (Failed to pass one or more QA tests)

sensor can cause a persistence error problem in the acquired data. Again like the step test, there is a threshold defined for the persistence duration, above which data will be flagged incorrectly. For example, the persistent threshold for solar radiation may be set to about 12 hours, depending on the geographical location.

5.2.1.5 Spatial consistency test

Spatial consistency tests check the required correlation between observations at different heights at the same station, or between several surrounding stations. Observations in which difference is more than a predefined threshold are flagged as wrong data. Thresholds definition is usually based on location and terrain characteristics. In general, spatial comparisons between different stations are most reliable during the well-mixed portion of the day during periods of weak horizontal gradients [245]. Spatial consistency tests can detect gross errors in individual observations. In another work by Shafer et al. [246], a distance-weight-based method is proposed for the nearby observations to estimate the values and compare with the real observations at the intended station. The difference is compared to the standard variation, and observations with a difference above a predefined threshold are flagged as suspicious.

In this work, a weather data quality assurance procedure is developed using several combinations of different tests previously published. For some parameters, only one or two test types are available; for other parameters, more than two test can be considered. In this procedure, if any test is failed, data is set as invalid. As this procedure output, each data is labelled as valid, missing, or invalid, following table 5.2 labels. All missing and invalid data are treated as missing data to be reconstructed in the next step.

Table 5.3 Applied quality assurance tests on weather parameters

Validation tests	Parameters	Reference condition	Ref.
Basic test	All parameters	Missing and duplicated data	
Range test	Wind speed(m/s)	$0 < WS_{(hourly\,average)} < 40$	[223]
	Wind direction	$0 < WD_{(hourly\,average)} < 360°$	[223]
	Pressure(Pa)	-	
	Relative Humidity(%)	$0.8 < RH < 103$	[246]
	Temperature(°C)	$-30 < T < 50$	[224]
	Precipitation(mm)	$0 \leq Prec \leq 120$	[243]
		$0 \leq Prec < 508$	[246]
Step test	Wind speed(m/s)	Slope of $WS_{(hourly\,average)} < 6m/s$	[223]
	Pressure(Pa)	Slop of $P_{(hourly\,average)} < 1kPa$	[223]
	Relative Humidity()	Slope of $H_{(Semihourly\,average)} < 45$	[247]
	Temperature(°C)	Slop of $T_{(hourly\,average)} < 5°C$	[223]
	Radiation(wm^{-2})	$0 \leq Rad_{(hourly\,average)} \leq 555$	[248]
Persistence test	Wind speed	$WS_{daily\,mean}(d) \neq WS_{daily\,mean}(d-1)$ $WS_{daily\,max}(d) \neq WS_{daily\,max}(d-1)$	[248]
	Relative Humidity(%)	$\sigma_{hourly} H_{hourly} > 1$	[247]
	Temperature(°C)	$T_{daily\,mean}(d) \neq T_{daily\,mean}(d-1)$	[248]
Internal const.	Wind speed	$WS_{daily\,mean} < WS_{daily\,max}$	[224]
		$WS = 0 \,\&\, WD = 0 \,\|\|\, WS \neq 0 \,\&\, WD \neq 0$	[249]
	Temperature(°C)	$daily\,average:\ T_{min} < T_{mean} < T_{max}$ $T_{mean}(d-1) < T_{max}(d)$ $T_{min}(d) \leq T_{max}(d-1)$	[224]
	Precipitation(mm)	$daily\,average: Prec_{(0-3h)} \leq Prec_{(0-6h)}$ $Prec_{(0-12h)} \leq Prec_{(0-24h)}$	[250]
Spatial const.	Wind speed(m/s)	$0 < WS_{60m-hour.ave.} - WS_{10m-hour.ave.} < 0.5$	[223]
	Wind direction	$WD_{60m-hour.ave.} - WD_{10m-hour.ave.} < 22.5°$	[223]

5.2.2 Data reconstruction module

The reconstruction proposed model is composed by two steps: a classification phase and a prediction step. In the first step, a multidimensional k-means clustering algorithm will be applied to the WRF numerical weather prediction data to form several groups with similar characteristics. Then FFNN model will be trained on each cluster and reconstruct

the observational gap whose group is identical to its corresponding WRF value. Figure 5.3 shows a schematic view of the developed reconstruction model. The model applies to two groups of missing patterns which will be presented in the following.

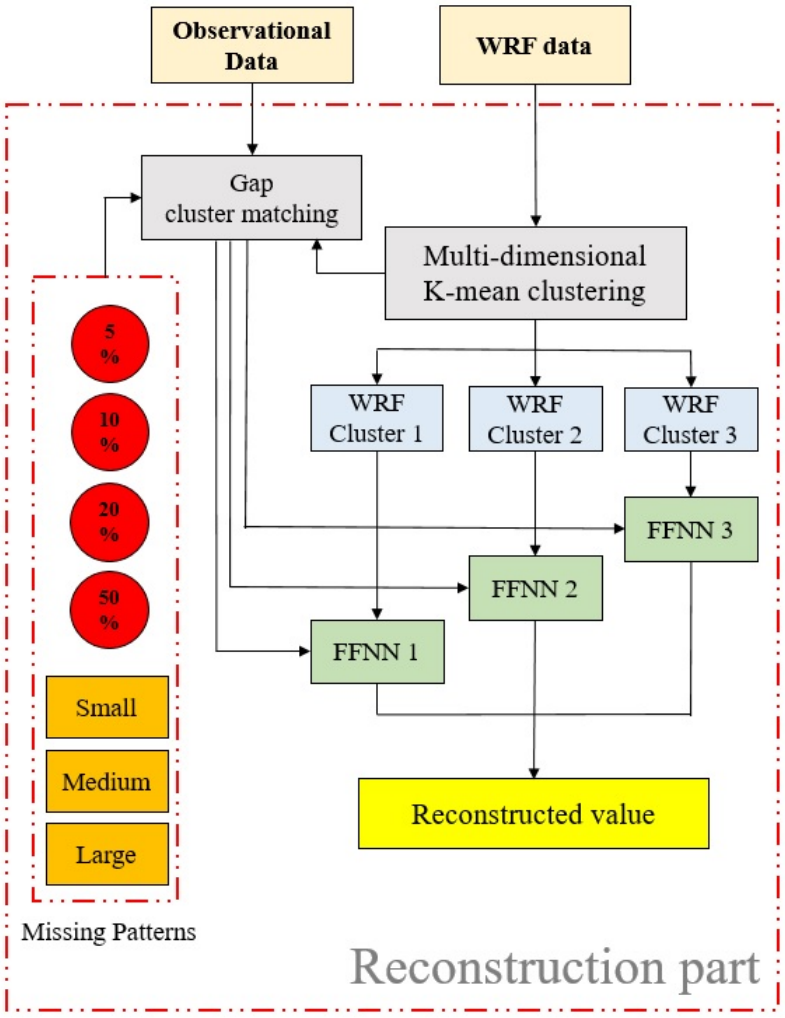


Fig. 5.3 Reconstruction module flowchart

5.2.2.1 Inverse distance weighting of WRF data

Simulated data from weather research and forecasting (WRF) model is collected from the open access repository of the Galician regional meteorological agency (MeteoGalicia) [87]. These WRF data are based on a mesh of 4km horizontal resolution. In this work special interpolation is chose as we need to have simulated data down-scaled at the location of the meteorological station.

Inverse distance weighting (IDW) is a type of deterministic multivariate interpolation with a known scattered set of points. With a weighted average technique of the values available at the known points, the value of the unknown points is assigned. Indeed, the weights are assigned to the inverse of the distance to each known point. This method was developed by Shepared, in 1965 as a subsidiary project in Harvard laboratory for computer graphics and spatial analysis [251]. Equation 5.1 represents IDW mathematically and equation 5.2 shows how to find an interpolated value ϕ given a set of spatial points.

Considering S as a study region, the expected result is a discrete assignment of the unknown function Φ in S:

$$\Phi(x) : x \to R, \quad x \in \mathbf{S} \subset R^n \tag{5.1}$$

The expected results can be demostrated as a set of n-tuple:

$$[(x_1, \phi_1), (x_2, \phi_2), (x_3, \phi_3), ...(x_n, \phi_n)].$$

A general mathematical form of finding an interpolated value ϕ at a given point x based on samples $v_i = v(x_i)$ for $i = 1, 2, ..., N$:

$$\phi(x) = \begin{cases} \frac{\sum_{i=1}^{N} \omega_i(x) v_i}{\sum_{i=1}^{N} \omega_i(x)} & \text{if } \forall i : d(x, x_i) \\ v_i, & \exists\, i : d(x, x_i) \end{cases} \tag{5.2}$$

Where $\omega_i(x)$ is the weight assigned to the (x, x_i) pare and is equal to the inverse distance between x and x_i:

$$\omega_i(x) = \frac{1}{d(x, x_i)^p}, \tag{5.3}$$

5.2.2.2 Missing data patterns

For the classification phase, different missing patterns should be as different groups in a clean, complete dataset. Gap patterns should be representative of the real missing patterns in data. In this work, only random errors in weather data and missing gaps in the data time series were considered; with the lattest also considered as random errors.

In literature, there are several mechanisms for missing data: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). For the MCAR, the chance of missing data values is identical across all the time series for any time

step. It can be interpreted as the causes not being related to the data collected. In addition, this type of gap appears more diverse and pointwise in the time series with fewer consecutive time steps [235]. MAR gives an equal chance to all of the categories instead of all instances. Selecting samples from a population based on certain characteristics is an example of MAR. An example is a sensor that occasionally fails during the data acquisition process due to a power outage where the missingness depends only on the observed input data. In NMAR, the results of missing data are because of some unknown reasons. If a sensor cannot acquire information outside a certain range, its data are missing due to NMAR factors. Then the data are said to be censored. If missing data are NMAR, valuable information is lost from the data; and there is no general method of handling missing data properly [252].

5.2.2.3 K-means multidimensional clustering

Clustering is the task of dividing a set of data into several groups with more similarities between the members of each group compared to other groups. k-means is an iterative and the most popular clustering algorithm which stores p centroids as the identifier of the clusters. MacQueen first used this algorithm in 1967 [253]. A data point is then assigned to a cluster that has a closer distance to that centroid. The centroids get updated in each iteration to find the best by alternating between giving data points to the clusters based on the current centroids and choosing centroids based on the current assigned value to clusters. Equation 5.4 describe the mathematical form of the k-means algorithm.

Given a set of data points $(X_1, X_2, ..., X_k)$, which each data point is a d-dimensional real vector, k-means clustering partitions the k observations in to p sets $S = S_1, S_2, ..., S_p$ in a way to minimize the within-cluster sum of squares as follow:

$$\arg\min_{\mathrm{S}} \sum_{i=1}^{p} \sum_{X \in S_i} ||x - \mu_i||^2 = \arg\min_{\mathrm{S}} \sum_{i=1}^{p} |S_i| \, Var \, S_i \tag{5.4}$$

Where, μ_i stands for mean of data points in S_i. The above equation can be calculated as pairwise squared deviations of points in the same cluster:

$$\arg\min_{\mathrm{S}} \sum_{i=1}^{p} \frac{1}{2\,|S_i|} \sum_{X,Y \in S_i} ||X - Y||^2 \tag{5.5}$$

5.2.2.4 Feed forward neural network (FFNN) - MLP

As explained in subsubsection 4.3.1.1, multi-layer perceptron belongs to the FFNN family.However, FFNN can be used to refer MLP in this context, since the other group of FFNN, single-layer perceptron, is rarely used for wind speed and power prediction models. The two most highlighted characteristics of FFNN are: a) the neurons of each layer are fully connected to the previous or next layers; b) the inputs move only in one direction from the input to the output layer. MLP is vastly applied in the literature for the WS&PF problem with successful results [77, 128, 129]. In a comparative study by Ghorbani et al. [140], a new MLP model (8,8,1) outperformed genetic expression programming, multi-layer regression, and persistence models. A schematic view of the MLP in the model developed in this work as a reconstruction model is presented in figure 5.4.

(I) Hidden-Layer Process.

Equation 5.6, show the output result of all neurons in the hidden layer.

$$O_{1,j} = \phi_{H1}\left(\sum_{i=0}^{n} W_{i,j}x_{i,j}\right) \qquad j = 1,2,3,...,L_1 \tag{5.6}$$

Where $X = (x_1, x_2, x_3, ...x_n)$ is the input value dataset feed into input layer, $O_{1,j}$ is the output values of the jth node in in the hidden layer. $W_{i,j}$ is the weight values between the input and the hidden-layer, and ϕ_{H1} is the activation function which is logarithmic sigmoid transfer function as represented below.

$$\phi_{H1}(x) = \frac{1}{1+exp(-\alpha x)} \qquad \alpha > 0 \tag{5.7}$$

Where x is any real value, and $alpha$ is normally equal to one.

(II) Output Layer Process.

Considering n, m as the numbers of input nodes and hidden layer nodes, respectively, the single output of the FFNN MLP is calculated as follows:

$$Y = \theta\left(\sum_{j=0}^{m} c_k O_{1,j}\right) \tag{5.8}$$

where Y is the predicted output, c_k is the weight value between the hidden-layer and the output layer and θ is the activation function which normally considered linear function, $\theta(x) = x$.

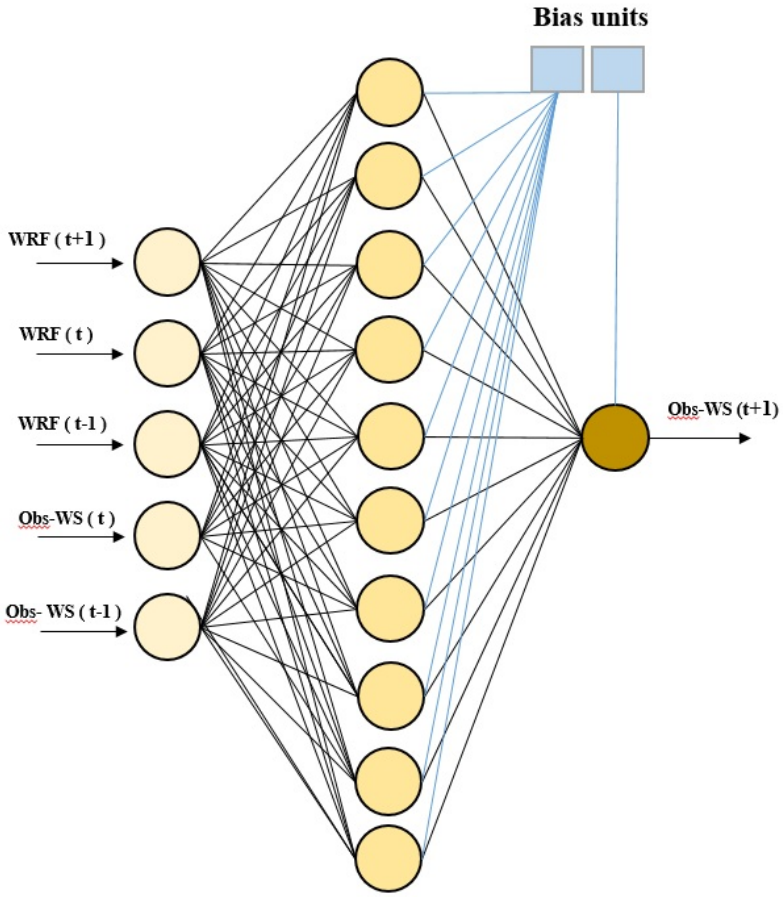


Fig. 5.4 A schematic view of FFNN used in the reconstruction phase

5.2.2.5 Bayesian regularization learning algorithm

The proposed learning algorithm here is Bayesian regularization [254], which is a network training function that updates the weight and bias values according to Levenberg-Marquardt optimization [197]. It minimizes a combination of squared errors and weights, and then determines the correct combination so as to produce a network that generalizes well. The process is called Bayesian regularization. Typically, training aims to reduce the sum of squared errors $F = E_D$. However, regularization adds an additional term; the objective

function becomes:

$$F = \beta E_D + \alpha E_W, \tag{5.9}$$

where E_W is the sum of squares of the network weights and α, and β are objective function parameters. The relative size of the objective function parameters dictates the emphasis for training. If $\alpha << \beta$, then the training algorithm will drive the errors smaller. If $\alpha << \beta$, training will emphasize weight size reduction as the expense of network errors, thus producing a smoother network response. In the Bayesian framework, the weights of the network are considered random variables. After the data is taken, the density function for the weights can be updated according to Bayes' rule, as follows,

$$P(\mathbf{w}|D,\alpha,\beta,M) = \frac{P(D|\mathbf{w},\beta,M)P(\mathbf{w}|\alpha,M)}{P(D|\alpha,\beta,M)} \tag{5.10}$$

Where D represents the dataset, M is the particular neural network model used, and w is the vector of network weights. $P(\mathbf{w}|\alpha,M)$ is the prior density, which represent our knowledge of the weights before any data is collected. $P(D|\mathbf{w},\beta,M)$ is the likelihood function, which is the probability of the data occurring, given the weights w. $P(D|\alpha,\beta,M)$ is a normalization factor, which assure that the total probability is 1. In Bayesian framework, the optimal weights should maximize the posterior probability $P(\mathbf{w}|D,\alpha,\beta,M)$, which is equivalent to minimizing the regularized objective function, equation 5.9. Applying the Bayes' rule to optimize alpha and beta in equation, it is obtained,

$$P(\alpha,\beta|D,M) = \frac{P(D|\alpha,\beta,M)P(\alpha,\beta|M)}{P(D|M)} \tag{5.11}$$

By assuming a uniform prior density $P(\alpha,\beta|M)$ for α and β, then maximizing the likelihood function, $P(D|\alpha,\beta,M)$, the posterior probability $P(\alpha,\beta|D,M)$ is maximized. Assuming that both the noise in the training set data and the prior distribution for the weights follow Gaussian distributions, the probability densities in Bayes' rule can be expressed as follows,

$$P(\mathbf{w}|D,\alpha,\beta,M) = \frac{1}{Z_D(\beta)e^{(-\beta E_D)}},\ and$$
$$P(\mathbf{w}|\alpha,M) = \frac{1}{Z_W(\alpha)e^{(-\alpha E_W)}} \tag{5.12}$$

Where $Z_D(\beta) = (\pi/\beta)^{n/2}$ and $Z_W(\alpha) = (\pi/\beta)^{N/2}$. By substituting above mentioned probabilities into Bayes' rule equation, it is obtained,

$$P(\mathbf{w}|\alpha,M) = \frac{1}{Z_F(\alpha,\beta)e^{(-F(\mathbf{w}))}} \tag{5.13}$$

By substituting the above probabilities into the likelihood function of equation 5.11, it is obtained,

$$P(D|\alpha,\beta,M) = \frac{P(D|\mathbf{w},\beta,M)P(\mathbf{w}|\alpha,M)}{P(\mathbf{w}|D,\alpha,\beta,M)} =$$
$$\frac{\left[\frac{1}{Z_D(\beta)}e^{(-\beta E_D)}\right]\left[\frac{1}{Z_W(\alpha)}e^{(-\alpha E_W)}\right]}{\frac{1}{Z_F(\alpha,\beta)}e^{(-F(\mathbf{w}))}} = \tag{5.14}$$
$$\frac{Z_(\alpha,\beta)}{Z_D(\beta)Z_W(\alpha)}.\frac{e^{(-\beta E_D-\alpha E_W)}}{e^{(-F(\mathbf{w}))}}.\frac{Z_F(\alpha,\beta)}{Z_D(\beta)Z_W(\alpha)}$$

The only unknown part in the above equation is $Z_F(\alpha,\beta)$, which can be estimated by Taylor series expansion. As the objective function's shape is quadratic in a small limited area around the minimum point, $F(\mathbf{w})$ can be expanded around the minimum point of the posterior density $\mathbf{w}_{MP}$, where we have zero gradients. Solving for the normalizing constant we have:

$$Z_F \approx (2\pi)^{(N/2)}(det((\mathbf{H}_{MP}))^{-1})^{1/2}\, e^{(-F(\mathbf{w}_{MP}))} \tag{5.15}$$

Where $\mathbf{H} = \beta\nabla^2 E_D + \alpha\nabla^2 E_W$ is the Hessian matrix of the objective function, which by being substituted in the equation 5.14 the optimum values of α and β at the minimum point can be obtained. In order to get the minimum, logarithmic of equation 5.14 is derived and set to zero; then, alpha and beta values for the minimum are obtained,

$$\alpha_{MP} = \frac{\eta}{2E_W(\mathbf{w}_{MP})} \quad and, \quad \beta_{MP} = \frac{n-\eta}{2E_D(\mathbf{w}_{MP})} \tag{5.16}$$

Where $\eta = N - 2\alpha_{MP}\, tr(\mathbf{H}_{MP})^{-1}$ is called the effective number of parameters. N is the total number of network's parameters. η is a measure of how many parameters in the neural network are effectively used in reducing the error function and have a range of $[0\, N]$.

5.2.3 New developed validation-reconstruction model

Figure 5.5 shows an scheme and algorithm 5.1 shows the pseudocode of the new developed validation-reconstruction model, applied to the WRF weather dataset. After that, several model conditions and parameters are detailed.

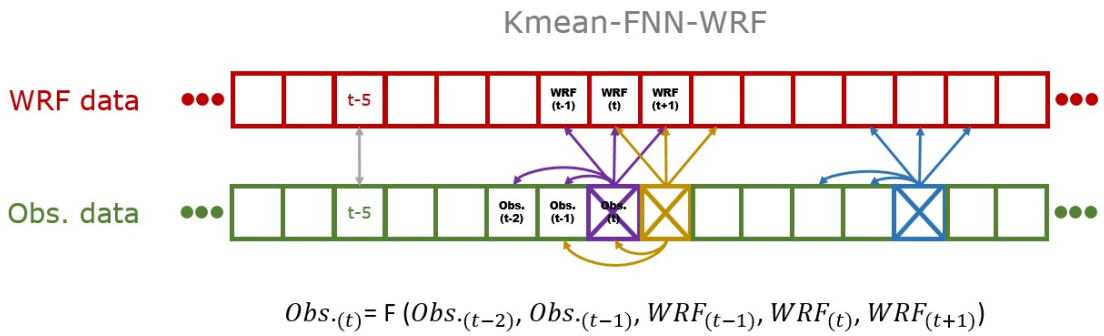


Fig. 5.5 Schematic depiction of input data feeding into the reconstruction model

Missing pattern: This model is focused in MCAR and MAR selecting 5%, 10%, 20%, and 50% missing values for the time series. In MCAR, the missing values are chosen based on a random selection. To model the MAR, two randomly selected values will be generated. The first one represents the missed value location and the second one (limited to 10) considers as consecutive steps after the missing value. Usually when there is a large gap, models can not deal with it properly. To test the model capability to deal with different length of consecutive gaps, three gaps categories were defined: small gaps containing 1 to 6 consecutive missing

values (up to one hour), medium gaps containing 6 to 24 consecutive missing values (up to one day), and long gaps containing 24 to 72 consecutive missing values (up to 3 days). Results are based on averaging 100 iterations for each category and group of missing patterns.

FFNN model: The developed model structure is MLP (5,10,1). The training algorithm is Bayesian regularization, and the activation functions are sigmoid and linear for the hidden and output layer. Different sets of structures and learning method were tested over the weather dataset, to select the above mentioned network parameters, as its best set.

5.2.4 Evaluation criteria

To evaluate the reconstruction model and compare it with other benchmark methods, three criteria were selected: MSE, RMSE, MAE, and MAPE, which are developed in equations 5.17 to 5.20.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - P_i)^2 \tag{5.17}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - P_i)^2} \tag{5.18}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|e_i| \tag{5.19}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{P_i - \hat{P}_i}{P_i}\right| \tag{5.20}$$

Where N is the total number of samples, $\hat{P}_i$ and P_i are predicted and real values at the time step i, respectively, and e_i stands for the residual error, obtained from predicted and real values reduction at each time step.

Algorithm 1 Validation-Reconstruction procedure

validation Module

for s as Stations = 1 to S **do**
  Consider P parameters of a station as $P_i = \{p_1, p_2, \ldots, p_p\}$
  Consider L validation levels as $L_i = \{l_1, l_2, \ldots, l_k\}$
  for p = 1 to P **do**
    for l = 1 to L **do**
      if p is not valid **then**
        Mark p as invalid time series and store it: $\{R_{s,p,i}\} \leftarrow P_{s,p}$
      end if
    end for
  end for
  $\{R_s\} \leftarrow \{R_{s,p}\}$
end for
Return $\{R\}$

Reconstruction Module

while R is not empty **do**
  for s' ($\in R$) as station = 1 to S' **do**
    for p' ($\in S' \in R$) as parameter = 1 to P' **do**
      if p' is wind speed time series **then**
        Apply k-means clustering on p', and same period of wind direction and humidity WRF dataset to form 3 clusters SC_j:
      else if For any other p' parameter except WS **then**
        Apply k-means clustering on single p' WRF dataset to form 3 sub-clusters SC_j:
      end if
      Create M missing data points based on missing patterns of MCAR or MAR: $R_{s',p'}$ as $M_{s',p'} = \{m_1, m_2, \ldots, m_M\}$
      for m = 1 to M **do**
        Find the sub-cluster SC_j which M_m belong to; train the model on that; Save trained network parameters W, B in $K_{s',p'}$; Apply trained model to find reconstructed value of M_m
      end for
    end for
  end for
end while

Application Module

while R is not empty **do**
  Apply $K_{s',p'}$ to reconstruct invalid or missing data in $R_{s',p'}$
end while

5.3 Results and discussion

In this section, the application of the new validation-reconstruction model to the different case studies described on Chapter 3 is presented. Including numerical results and their discussion.

5.3.1 Observations validation module

Unfortunately, there is no any metadata available for the case studies to find out the reason of the data gaps or invalid data. Figure 5.6 shows an example of Basic test failure in Muralla station during January 2017 for wind speed data, in which some data are missing, and the rest inside the red rectangular are not valid. Besides using tests to identify such invalid data, visual inspection is useful to detect any gap in time series. This type of error may be related to sensor failure as the period is around a few days, and there is a sudden disruption in data. On the other hand, figure 5.7 shows an example of the range test failure for wind direction at Camariñas station, on 4 October 2017. There are three consecutive wrong data out of the normal range of wind direction, $0-360°$. This type of error, as it is relatively short, can not be related to sensor miscalibration or substantial failure; possible weather extreme effect is withdrawn, as other simultaneous weather data are valid at the same station. As the numbers are recorded -999 in the database, this should be a flag value for a problem of transmission, storage, or a short temporary problem of the sensor.

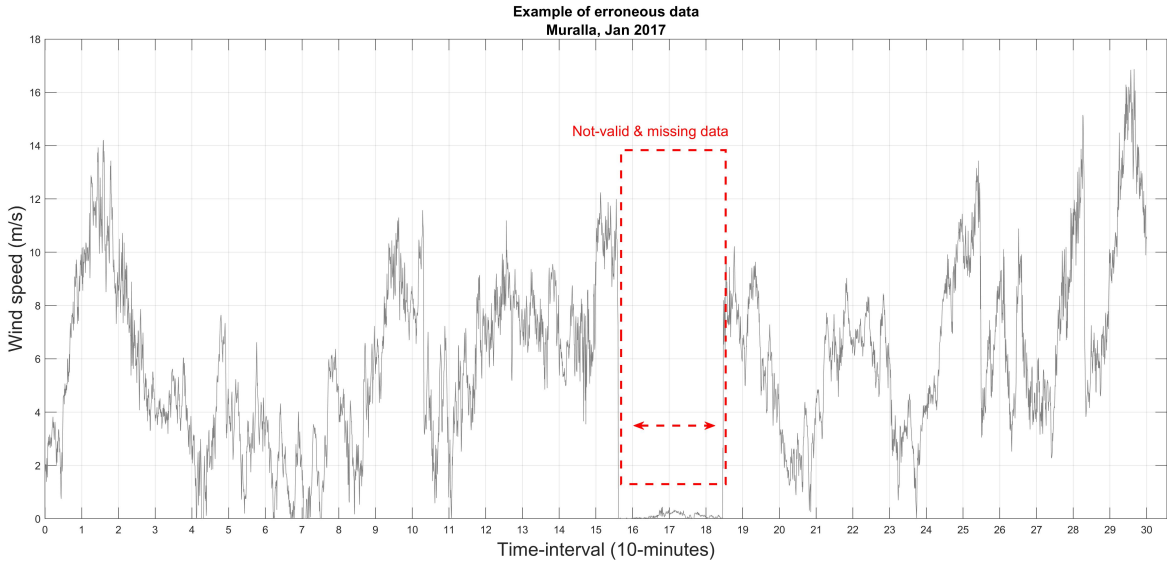


Fig. 5.6 Wind speed basic test failure, Muralla, Jan 2017

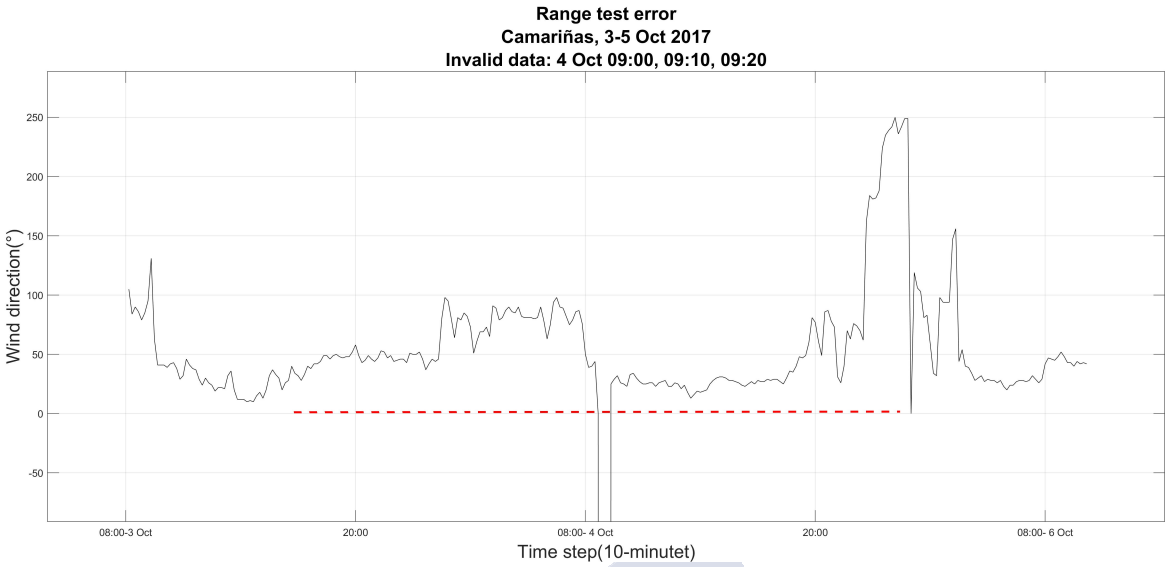


Fig. 5.7 Wind direction range test failure, Camariñas, 4 Oct 2017

Another test failure example, internal consistency, is shown in figure 5.8. Refer to table 5.3, and following [248], there should be an internal consistency between wind speed (WS) and wind direction (WD) in a way that whenever one parameter becomes zero, the other one should be zero as well and vice versa. As shown in this figure, for the M2-tower at 10-meter height, on several occasions (inside of the black line squares), wind speed and direction violate this rule. These erroneous data could be either because of a wind speed or wind direction. As data with a 1-minute frequency is so volatile, it has a low probability to persist at zero value for a continuous duration of tens of minutes, as in this case. So, the problem is highly related to wind speed. Again, since the wrong data period is not long and it is repeated during a few hours, there might be a connection or transmission problem.

An example of spatial test failure according to the criteria proposed by Yang et al. [223] is shown in figure 5.9 to the M2-tower time series. However, in that study, the difference between wind directions at 10- and 60-meter height is is supported by the wind shear increment with height [255]. The two horizontal red lines at ± 22.5 are the above and below limits the wind direction is not considered valid. Significant violation of validation criteria is observed in October, which during a continuous period there are large wind direction variations. According to these criteria, there are 2124 invalid data along one year. Because of that, the origin of the problem could be instrumental: It is possible that sensors at different heights might be either highly sensitive or easily be affected by environmental and weather conditions.

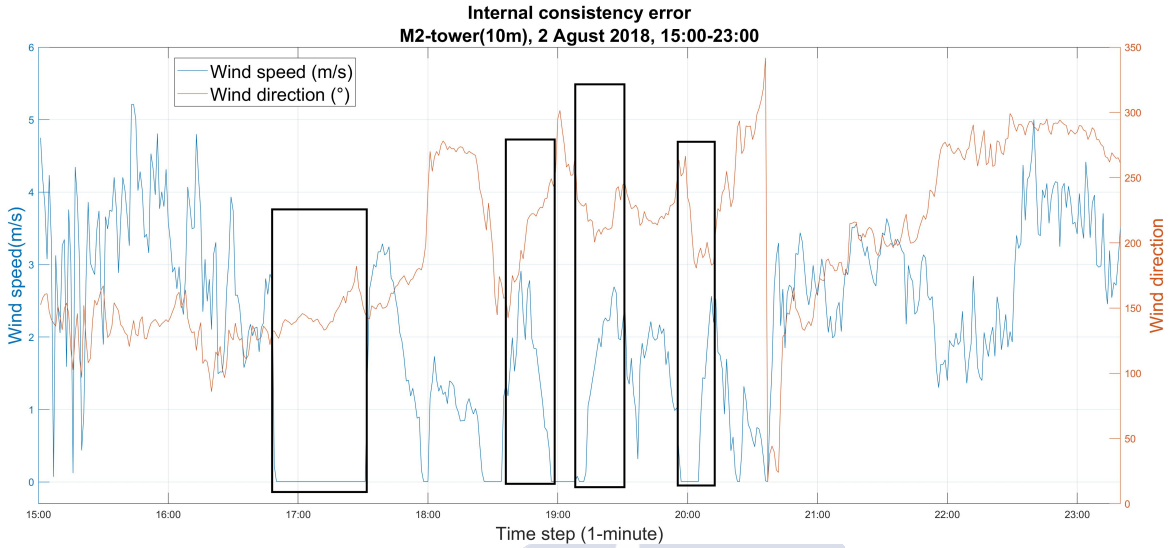


Fig. 5.8 Wind speed and direction internal consistency test failure, M2 tower, 2 Aug. 2018

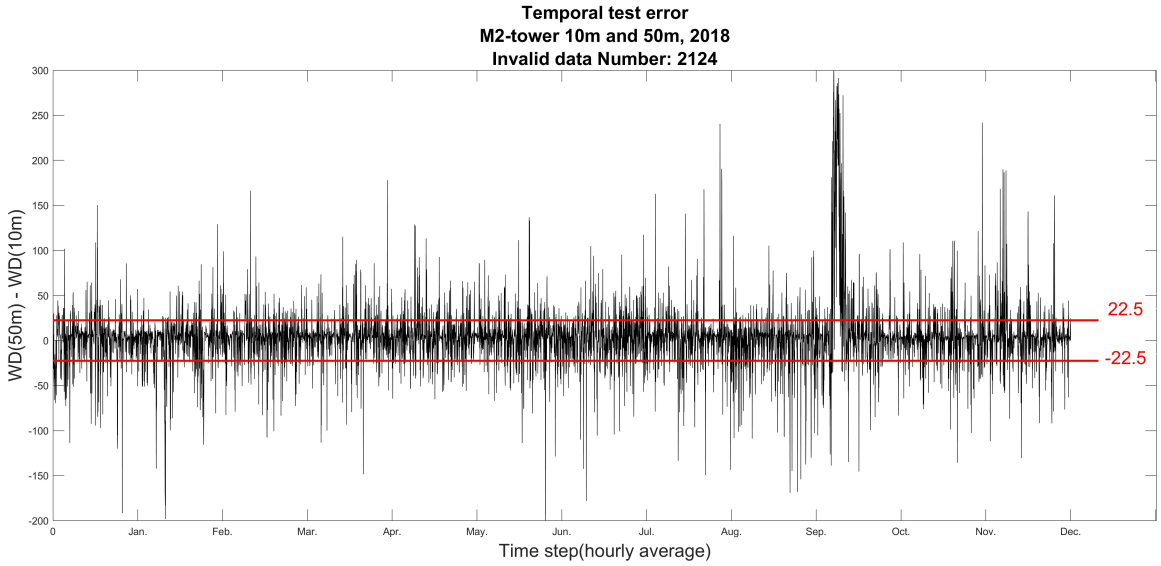


Fig. 5.9 Wind direction spatial consistency test failure, M2 tower, 2018

Considering all the meteorological stations data and applying validation tests summarized in table 5.3, the numbers and percentages of the invalid data for all parameters is provided in table 5.4. Notice that overall time steps at M2 tower (1-minute time step) and Galician stations (10-minutes time step) are different, so their results cannot be compared.

5.3.2 Invalid data reconstruction module

The data reconstruction module includes various steps through which different scenarios and conditions are compared. While results over different stations are obtained, Santiago-EOAS station results are mainly shown.

5.3.2.1 MLP structure and training selection

In the first step of building the reconstruction model, three different NN structures were used, following previous works. Referring to the table 4.4, the number of hidden layer neurons, L, can be defined based on input neuron numbers N, such as $(L = N)$ [140], $(L = 2N)$ [151, 158, 206], and $(L = 2N + 1)$ [152, 77, 11, 68]. For each structure, the proposed Bayesian regularization (BR) learning method was compared to other four training algorithms widely and successfully used in the literature for wind speed and power prediction. As it can be seen from table 5.5, the proposed BR overcame other algorithms in terms of MASE, MAE, and MAPE. This one confirms the results in [197] in which BR is reported to perform better than Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG) in a short term wind speed prediction model. Also, new model results support the outcomes of [256], where BR has provided more accurate prediction values than other five learning algorithms, including gradient descent and BFGS in a multi-step ahead wind speed forecasting model using a multi-layer perceptron feed-forward neural network. The performance of MLP structure (5,10,1) and (5,11,1) for BR learning method is competitive. However, the first structure performs slightly better than the other one for most of the cases, which is aligned with findings in some previous works [214, 212, 137]. It is important to mention that evaluation criteria were averaged over 100 iterations for each model under MCAR missing data pattern. Globally, 1,500 different models were run in these tests with the weather case studies datasets.

Table 5.4 Missing, duplicated and invalid weather data share and percentage at four meteorological stations

Stations QA	Wind speed	Wind direction	Humidity	Pressure	Temperature	Precipitation	Radiation
Santiago-EOAS							
Missing/Dubp.	589(0.01%)	1209(0.03%)	553(0.01%)	1824(0.04%)	509(0.00%)	1788(0.04%)	1850(0.04%)
Invalid	12(0.00%)	2894(0.06%)	653(0.01%)	1350(0.03%)	369(0.00%)	6305(0.12%)	1580(0.03%)
Camariñas							
Missing/Dup.	1252 (0.24%)	2506(0.00%)	8069(0.00%)	206(0.00%)	25(0.00%)	609(0.00%)	1589(0.00%)
Invalid	180 (0.00%)	3848(0.00%)	5924(0.00%)	1208(0.00%)	609(0.00%)	1897(0.00%)	5690(0.00%)
Muralla							
Missing/Dup.	890(0.02%)	1809(0.00%)	3560(0.00%)	12890(0.00%)	1023(0.00%)	2569(0.00%)	890(0.00%)
Invalid	126(0.00%)	5478(0.00%)	1561(0.00%)	1590(0.00%)	39(0.00%)	484(0.00%)	209(0.00%)
M2-Tower(10m)							
Missing/Dup.	18607(0.04%)	15050(0.00%)	4458(0.00%)	17584 (0.00%)	1673(0.00%)	25550(0.00%)	32449(0.00%)
Invalid	8589(0.00%)	1689(0.00%)	15845(0.00%)	5415(0.00%)	81079(0.00%)	14259(0.00%)	41685(0.00%)
M2-Tower(50m)							
Missing/Dup.	12855(0.03%)	18052(0.00%)	4896(0.00%)	45894(0.00%)	49875(0.00%)	45723(0.00%)	6359(0.00%)
Invalid	1892(0.00%)	36980(0.00%)	47890(0.00%)	1507(0.00%)	1468(0.00%)	14895(0.00%)	578(0.00%)
M2-Tower(80m)							
Missing/Dup.	10023(0.02%)	5002(0.00%)	1480(0.00%)	45789(0.00%)	12489(0.00%)	4125(0.00%)	4896(0.00%)
Invalid	14890(0.00%)	8790(0.00%)	7488(0.00%)	48752(0.00%)	35890(0.00%)	18998(0.00%)	589(0.00%)

Table 5.5 Performance comparison of the developed BR training algorithm with other learning methods

Gap Percentage / Model	MSE	RMSE	MAE(m/s)	MAPE(%)
MLP Structure (5,5,1)				
Levenberg-Marquardt (LM)	0.9359	0.967	0.755	30.853
Bayesian regularization (BR)	0.872	0.933	0.748	27.379
BFGS Quasi-Newton (BFGS)	0.904	0.951	0.752	25.763
Gradient descent (GD)	12.108	3.479	3.047	66.407
Scaled Conjugate Gradient (SCG)	0.874	0.935	0.765	31.834
MLP Structure (5,10,1)				
Levenberg-Marquardt	0.942	0.971	0.766	29.987
Bayesian regularization	**0.860**	**0.927**	**0.744**	27.876
BFGS Quasi-Newton	0.913	0.955	0.756	27.876
Gradient descent	42.998	6.557	6.243	64.606
Scaled Conjugate Gradient	0.940	0.969	0.758	26.715
MLP Structure (5,11,1)				
Levenberg-Marquardt	0.885	0.941	0.771	27.410
Bayesian regularization	0.861	0.928	0.748	**27.355**
BFGS Quasi-Newton	0.960	0.979	0.774	29.912
Gradient descent	125.67	11.21	10.174	69.751
Scaled Conjugate Gradient	1.281	1.131	0.883	31.671

5.3.2.2 Parameters selection for input data clustering

Clustered input data for neural networks is proved to increase the accuracy of the results using different clustering methods such as Fuzzy C-means [18], and k-means [168, 191]. In this study, multivariate k-means were selected due to their success in the other works. Apart from cluster number, K is another parameter needs to be defined. In this study, $K = 3$ was selected due to the time series length considered, which is 720 time steps. More clusters number will end in a short training set, which can result in an underfitting model. Otherwise, the best combination of weather parameters for wind speed prediction was obtained by setting the best clusters number. One hundred models for each possibility were run, and it turned out that wind speed, wind direction, and humidity combination outperforms other combinations,

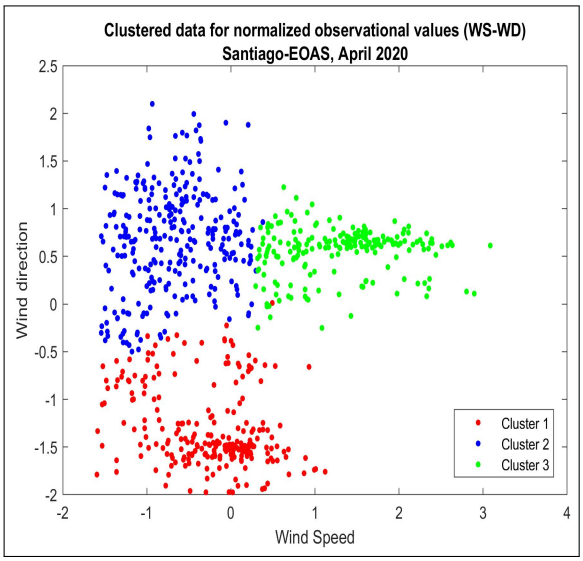
in terms of evaluation criteria. Table 5.6 shows the results for all possible combinations of 1,500 models.

Table 5.6 k-means clustering (3-clusters) applied on different combination of weather parameters for wind speed reconstruction using MLP

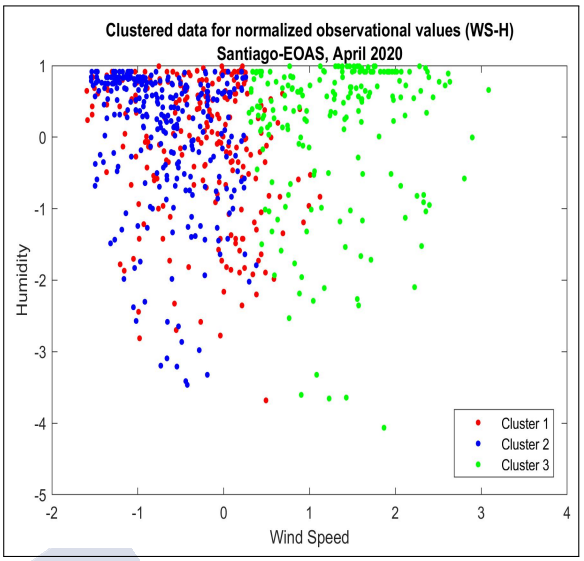
combinations	**MSE**	**RMSE**	**MAE(m/s)**	**MAPE(%)**
WS	0.499	0.684	0.600	20.202
WD	0.829	0.875	0.739	25.530
P	0.582	0.731	0.573	18.94
H	0.402	0.613	0.521	19.533
WS-WD	0.778	0.845	0.708	28.935
WS-P	0.706	0.822	0.677	21.080
WS-H	0.534	0.729	0.595	17.562
WD-P	0.598	0.768	0.606	22.827
WD-H	1.086	1.027	0.857	28.384
P-H	0.697	0.823	0.671	24.600
WS-WD-P	0.903	0.926	0.770	33.023
WS-WD-H	**0.396**	**0.604**	**0.497**	**17.321**
WS-P-H	0.799	0.853	0.652	19.808
WD-P-H	0.963	0.948	0.727	28.180
WS-WD-P-H	0.972	0.971	0.832	39.721

Figure 5.10 shows a two-by-two combination of wind speed, wind direction, and humidity. The three-dimensional figure 5.10d shows a perfect separation of data due to the three parameters.

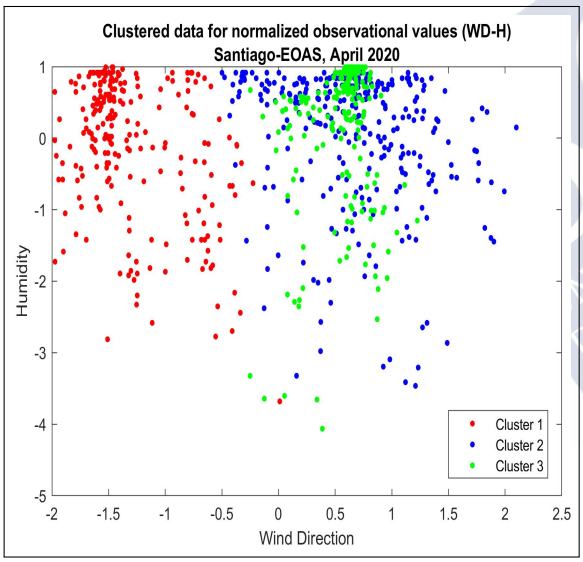
It is essential to mention that the same comparison between different combinations of clustered weather parameters is required for accurate results. However, the k-means clustering method is applied on each single parameter except WS, above discussed, because of two reasons: First, the applied k-means on a single parameter provides highly accurate results compared to other combinations, as it is observed in table 5.6. Second, to obtain a general simplified procedure that can be logically applied for different parameters, it is more straightforward to apply the clustering method only to the same parameter. In figure 5.11, wind direction, pressure, and humidity are clustered in three groups and plotted versus wind speed to be visualized in a 2D projection. As it can be seen, the wind direction, pressure and humidity are clustered simply by dividing their values range. Plotting those variables against wind speed shows clearly parameters' data separation.



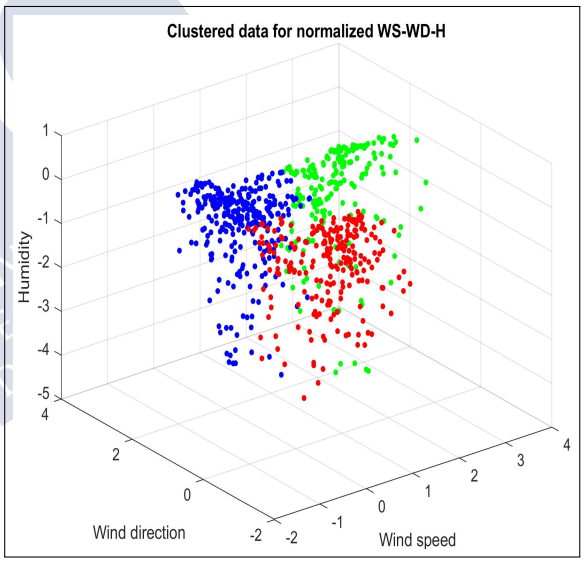
(a) WS-WD



(b) WS-H



(c) WD-H

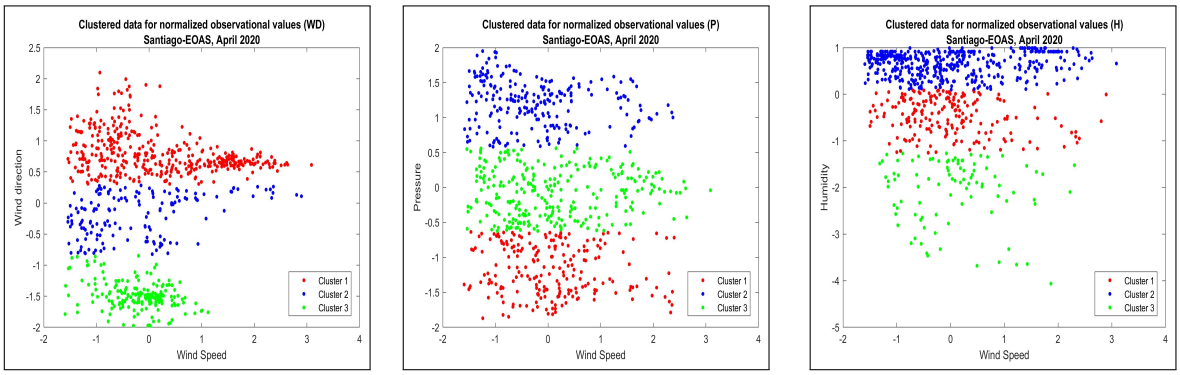


(d) WS-WD-H

Fig. 5.10 K-means clustering applied to WS using WS-WD-H combination

5.3.2.3 FNN model performance

The developed FNN model uses the Bayesian regularization training algorithm. The model has three layers: In the input layer, there are five neurons as the model inputs, including three WRF data points at the current, next, and previous time steps of the same parameter (figure 5.5), and two observational data from the last two steps of the same parameters.



(a) Wind direction (b) Pressure (c) Humidity

Fig. 5.11 Single parameter K-means clustering visualization for weather parameters except WS

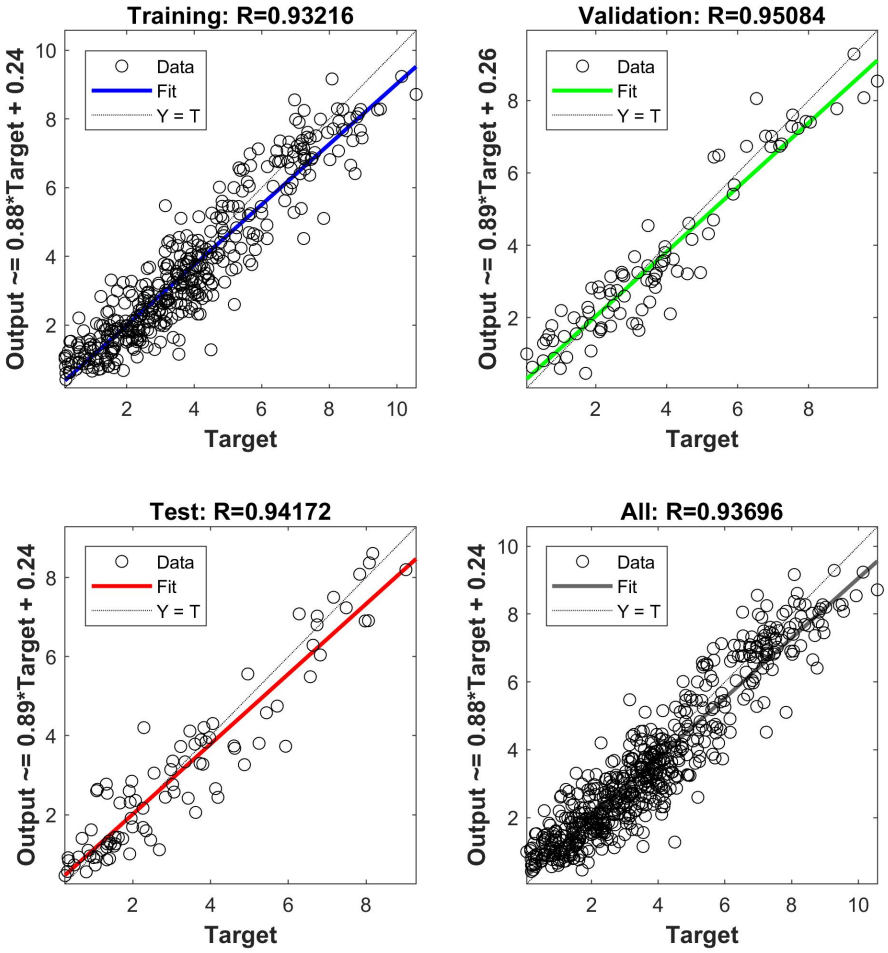
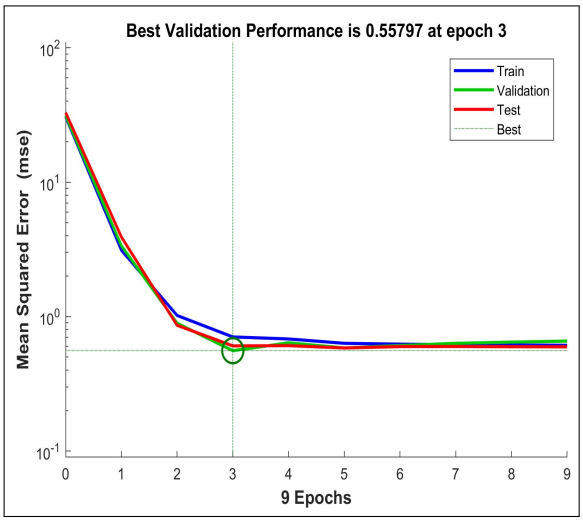
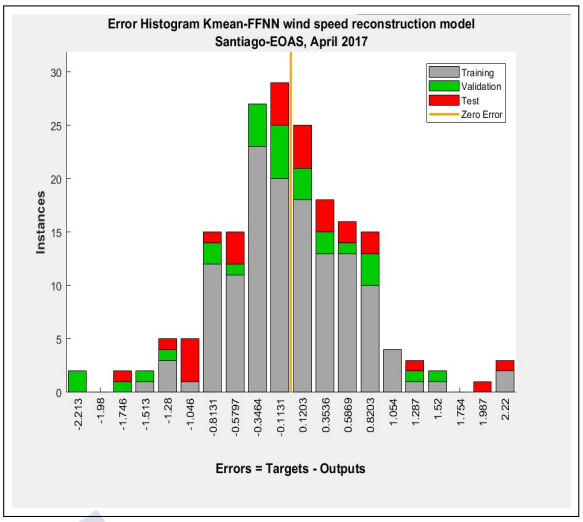


Fig. 5.12 Regression plot of the (k-means)-FNN-WRF model for Training, Validation and Test datasets. And combined regression plot with All results.



(a) Best validation performance of the proposed model vs. Epoch



(b) Training, Validation and test errors' histogram

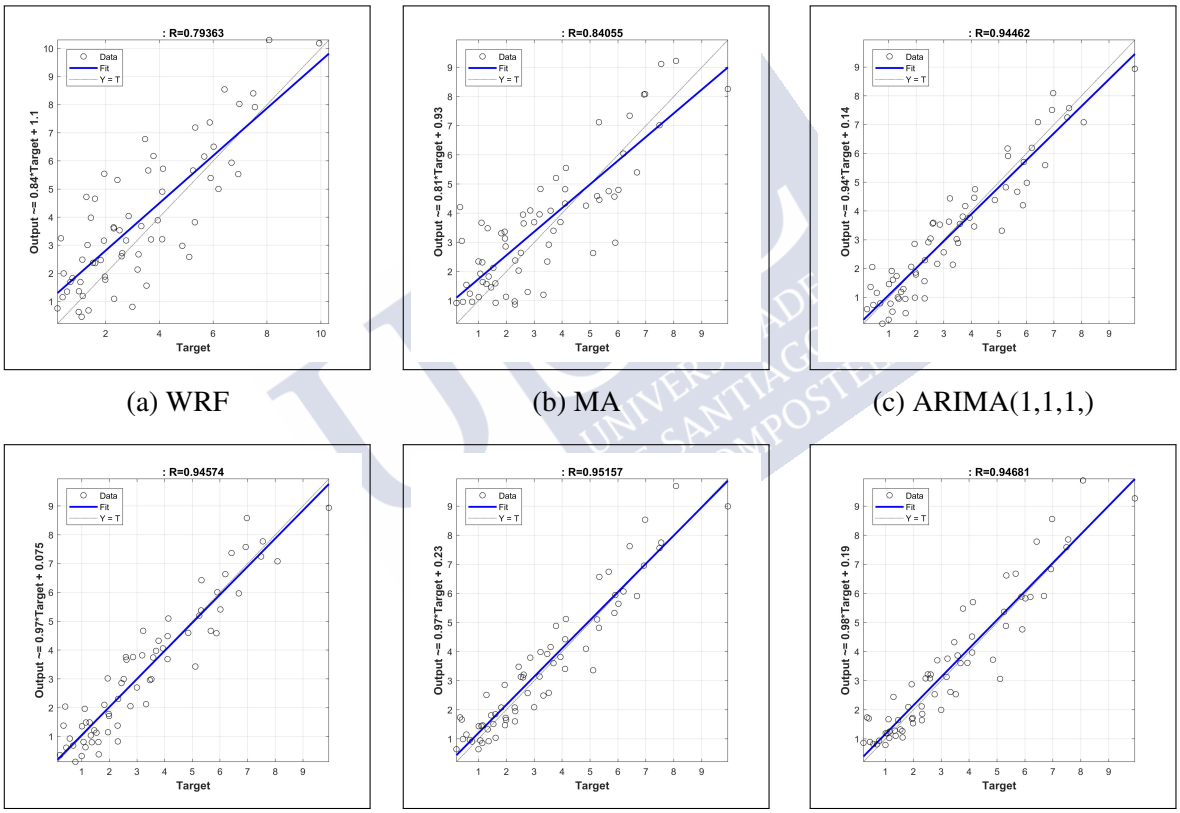
Fig. 5.13 Errors histogram of the proposed model on training, validation and test datasets

All neurons from one layer are fully connected to the next layer. Linear regression analysis is commonly used to test the performance of a model in approximation applications. As it can be seen in figure 5.12, total dataset is divided into three main groups of training (30%), validation (15%), and test (15%). Also, the regression coefficients for each group and the regression line between the target and the calculated output variables are shown. For a perfect fit (outputs exactly equal to targets), the slope should be 1, and the y-intercept would be 0. If the correlation coefficient is equal to 1, then there is a perfect correlation between the outputs from neural network and the targets in the testing subset. Therefore, it is observed a strong correlation between the model predictions and its actual results for all training, validation, and test dataset. As the model is performing well on not only the training data set but also the validation and test datasets, neither overfitting nor underfitting problema are observed [257]. In figure 5.13a MSEs values for the different datasets are shown at different Epoch number. As it is observed, at Epoch = 3 validation is ok. Also, figure 5.13b shows the error histograms for training, validation and test datasets. Errors follow a normal distribution with their mean at zero.

5.3.2.4 Comparison with other benchmark models

Other six models are used to compare their results with the developed model. WRF is a well common weather forecast model, including wind speed forecast, so it is widely applied in wind energy assessment [258, 259]. Moving average is reported to give highly precise

imputations for daily and weekly solar radiations [229]. It is reported to outperform recurrent neural networks for wind speed, wind direction, and temperature time series imputation [237]. Two ARIMA models (1,1,1) and (1,1,0) were proposed in [158] for wind speed prediction model and reported to be successfully applied to missing high-resolution weather data [236]. The following two models are also used to compare their results against the proposed kmeans-FNN-WRF model: First, the single FFNN model is the basis of the new developed model. Second, kmeans-FNN-Obs, which inputs are only observational data. In the developed model, a combination of WRF and observational data is used as input for the MLP model trained on WRF clustered data, as it was described in section 5.3.2.2.



(a) WRF (b) MA (c) ARIMA(1,1,1,)

(d) ARIMA(1,1,0) (e) FNN (f) Kmean-FNN-WRF

Fig. 5.14 Regression plot of different models: a) WRF, b) MA, c) ARIMA(1,1,1), d) ARIMA(1,1,1), e) FNN, f) Developed model

Each model, except WRF and MA, was run for 100 times and the mean values of MSE, RMSE, MAE, and MAPE for all of the above mentioned models, and for different MCAR values are shown in table 5.7. Besides, figure 5.14 shows an example of regressions plots and R correlation coefficient for different models in one random iteration. In a time series of

720 hourly data points for one month, 5%, 10%, 20%, and 50% is equal to 36, 72,144, and 360 missing values. By increasing the number of missing values, the error should increase as predicted previous time step values are using to predict other missing data. The error's increase when the gap percentage increases, can be seen obviously in the table. This result is perfectly consistent with [252] and [238], however contradicts with [235]. The outcome of the proposed method for 5% gap is 0.434, 0.648, 0.544 and 17.552 for MSE, RMSE, MAE and MAPE, respectively, for 10% gap is 16.033 for MAPE, for 20% gap is 0.391, 0.624, 0.493 and 17.492 for MAE, RMSE, MAE and MAPE respectively, and for 50%gap is 0.591, 0.755, 0.586, and 20.376 for MSE, RMSE, MAE and MAPE, respectively which outperforms reported results by [235, 238], considerably.

Figure 5.16 summarize table 5.7 for the MSE criteria using bar diagram and makes it easier to interpret the results. As a general fact, WRF and MA(36-hours) model produce larger errors for all of the gap percentages than the other methods. The positive effect of using k-means clustering is obvious for all of the gap percentages scenarios applied to the proposed model using observational and WRF data, and the model that uses five previous time steps of observational data (k-means -FNN-Obs). Using WRF data improves the results compared to using only observational data. However, in one case for the 10% missing data, the MAE criteria is 0.45 for observational inputs and 0.509 for mixed WRF and observational inputs. According to the table 5.7 regarding this case, MSE and RMSE are better for the observational data, but MAPE criteria is less for the proposed model, 16.03 against 21.90.

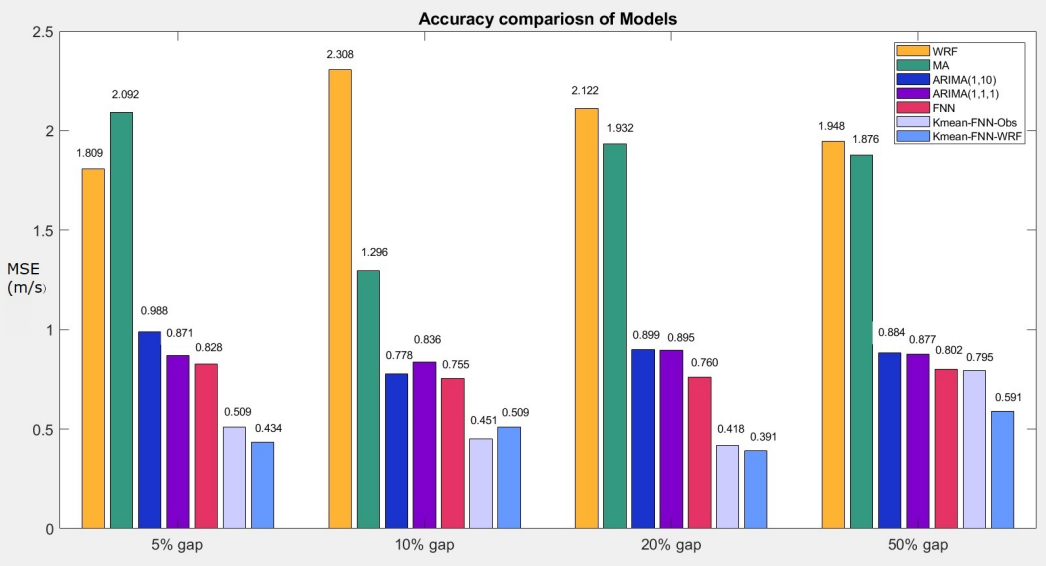


Fig. 5.15 Mean square error comparison at different data gap percentages (MCAR), for developed model (k-means-FNN-WRF) and benchmark models

Table 5.7 Models performance comparison between developed model (Kmean-FNN-WRF) and benchmark models based on MSE, RMSE, MAE, and MAPE, for several MCAR values

Gap Percentage / Model	MSE	RMSE	MAE	MAPE
5% **Gap**				
WRF	1.809	1.345	1.083	33.417
$MovingAverage$	2.092	1.445	1.123	38.677
$ARIMA(1,1,0)$	0.988	0.994	0.780	25.757
$ARIMA(1,1,1)$	0.871	0.933	0.747	25.720
FNN	0.828	0.910	0.729	23.248
$Kmean-FNN-Obs$	0.509	0.692	0.574	18.242
$Kmean-FNN-WRF$	**0.434**	**0.648**	**0.544**	**17.552**
10% **Gap**				
WRF	2.308	1.519	1.192	38.195
$MovingAverage$	1.296	1.138	0.873	28.115
$ARIMA(1,1,0)$	0.778	0.882	0.701	34.633
$ARIMA(1,1,1)$	0.836	0.915	0.691	28.929
FNN	0.755	0.869	0.651	21.886
$Kmean-FNN-Obs$	**0.451**	**0.640**	**0.496**	21.922
$Kmean-FNN-WRF$	0.509	0.684	0.509	**16.033**
20% **Gap**				
WRF	2.122	1.457	1.171	44.059
$MovingAverage$	1.932	1.389	1.099	39.879
$ARIMA(1,1,0)$	0.899	0.948	0.779	41.030
$ARIMA(1,1,1)$	0.895	0.946	0.779	39.212
FNN	0.760	0.872	0.709	26.322
$Kmean-FNN-Obs$	0.418	0.646	0.516	25.756
$Kmean-FNN-WRF$	**0.391**	**0.624**	**0.493**	**17.492**
50% **Gap**				
WRF	1.948	1.396	1.083	43.438
$MovingAverage$	1.876	1.370	1.064	45.576
$ARIMA(1,1,0)$	0.884	0.940	0.744	44.806
$ARIMA(1,1,1)$	0.877	0.937	0.740	43.113
FNN	0.802	0.895	0.696	27.869
$Kmean-FNN-Obs$	0.795	0.891	0.678	25.40
$Kmean-FNN-WRF$	**0.591**	**0.755**	**0.586**	**20.376**

5.3.3 Different gap length in datasets

MCAR doesn't necessarily represent the reality of the missing patterns in datasets. Using MCAR, we mainly deal with pointwise missing values, which will end up in high accuracy reconstructed results in a range of 0.434 to 0.591 from 5% to 50% missing percentage gap. To apply the model on more real data missing cases, a new missing pattern was considered as consecutive data points are dominant in the time series. Performance results in table 5.8 show the developed method outperforms all the other models for three different gap lengths. The developed model outcomes for small gaps between 1 to 6-hours of consecutive missing values, mostly representative of environmental and temporal sensor errors, are 1.058, 0.777, 0.654, and 28.634 for MSE, RMSE, MAE, and MAPE, respectively. The same statistics for a medium gap from 6 hours up to one-day of consecutive missing data points, mostly related to communication and storage errors, are 1.184, 0.859, 0.736, and 33.000, respectively. For large data gaps from one day up to three days of consecutive gaps, mostly related communication or calibration errors, their values are 1.250, 0.897, 0.774 and 31.138, respectively.

Table 5.8 Accuracy comparison of the developed model against other benchmark models for three groups of consecutive gaps: Small(1-6h), Medium(6-24h), and Large(24-72h).

Gap Percentage / Model	MSE	RMSE	MAE	MAPE
Small Gap (1-6 h)				
WRF	2.003	1.415	1.096	39.850
MovingAverage	2.254	1.501	1.200	42.227
ARIMA$(1,1,0)$	2.509	1.506	1.170	73.788
ARIMA$(1,1,1)$	1.613	1.269	1.037	110.591
Developed model	**1.058**	**0.777**	**0.654**	**28.634**
Medium Gap (6-24h)				
WRF	2.012	1.412	1.099	40.201
MovingAverage	4.015	1.976	1.542	48.952
ARIMA$(1,1,0)$	4.285	2.070	1.573	313.26
ARIMA$(1,1,1)$	4.125	2.002	1.549	234.41
Developed model	**1.184**	**0.736**	**0.859**	**33.000**
Large Gap(24-72 h)				
WRF	1.929	1.383	1.082	35.536
MovingAverage	5.124	2.164	1.728	59.321
ARIMA$(1,1,0)$	7.210	2.586	2.072	106.63
ARIMA$(1,1,1)$	8.640	3.105	2.446	162.11
Developed model	**1.250**	**0.897**	**0.774**	**31.138**

Particularly, figure 5.16 shows a graphical representation of MSE values, at three different gap groups. It is clear that by increasing the size of the gap, the average error increases. Not surprisingly, the WRF model accuracy doesn't have any relationship with the data gap and has a uniform range of $[1.929, 2.012]$. MA and two ARIMA models, however, are strongly dependent on the gap size. For the new developed method, errors increase by amplifying gap length, but these changes are not as significant as in MA and ARIMA models. The MSE values for the proposed model is 1.058, 1.181, and 1.250 for small gap(1-6h), medium gap(6-24h), and large gap(24-72h), respectively. The outcome result is consistent with [237] and [238] previous results.

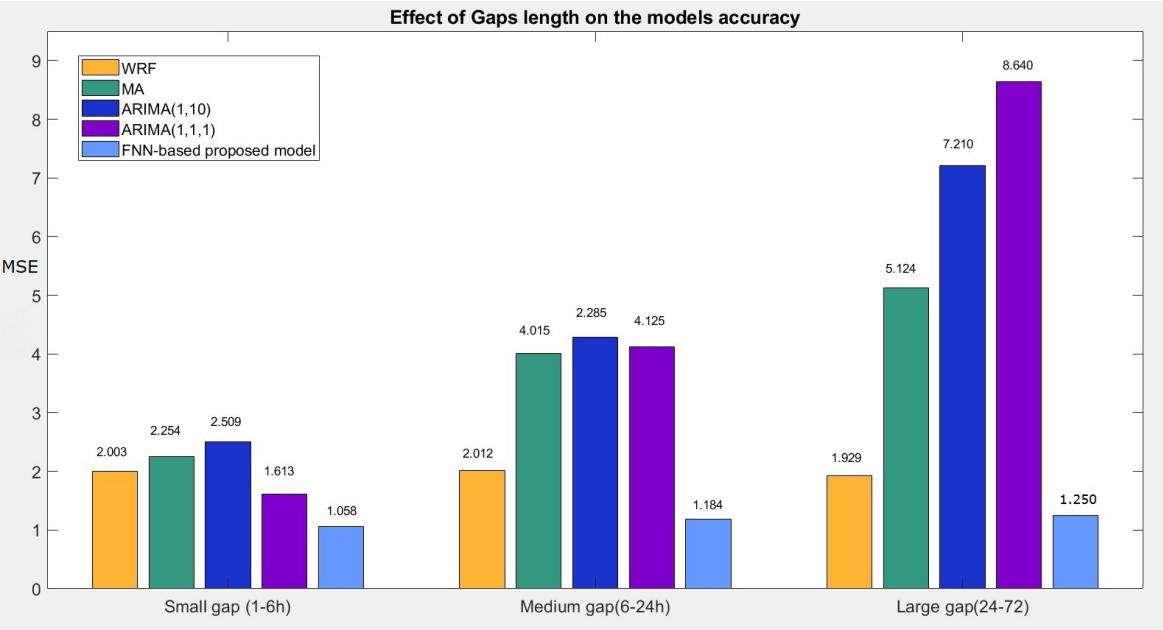


Fig. 5.16 Mean square error comparison for different gaps length, using FFN-based developed model and benchmark models

In figure 5.17 and 5.18 two reconstruction examples of the different models on small and large gaps are presented. ARIMA(s) and MA models are highly dependent on the previous time steps. As predicted values are considered input for the next time steps, the prediction's errors generated in each time step will accumulate and result in poor performance for longer consecutive gaps. As previously mentioned, p, d, q parameters in these ARIMA models are set based on a successfully applied model in [158] build for wind speed prediction of hourly averaged data.

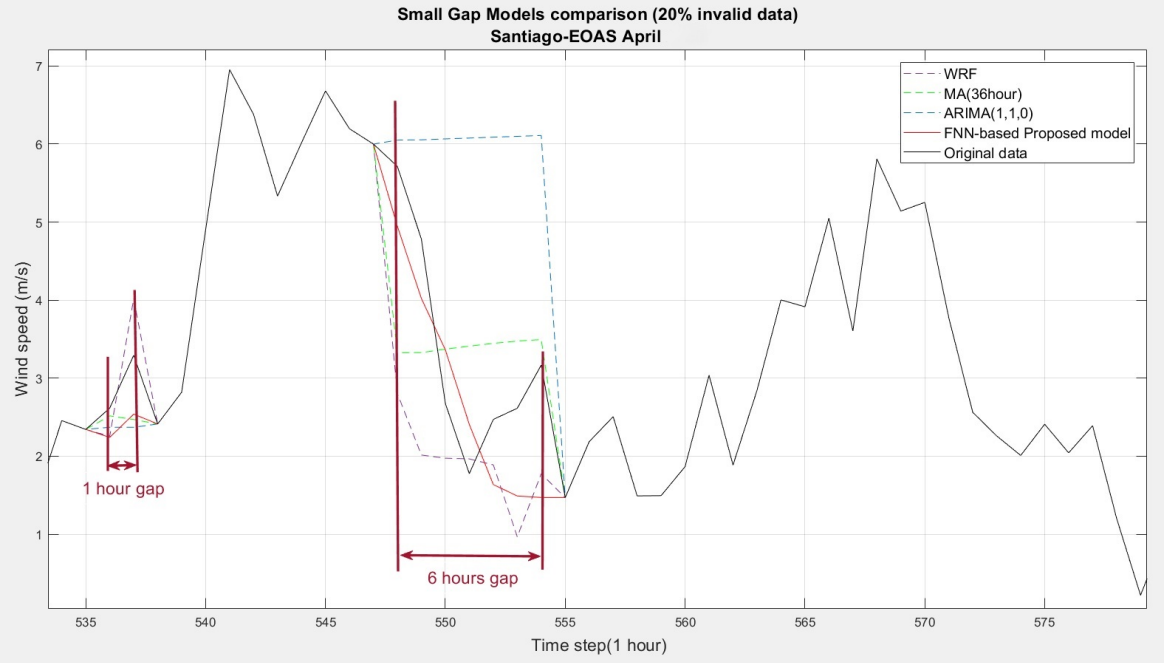


Fig. 5.17 An example of reconstruction models' results on small data gap

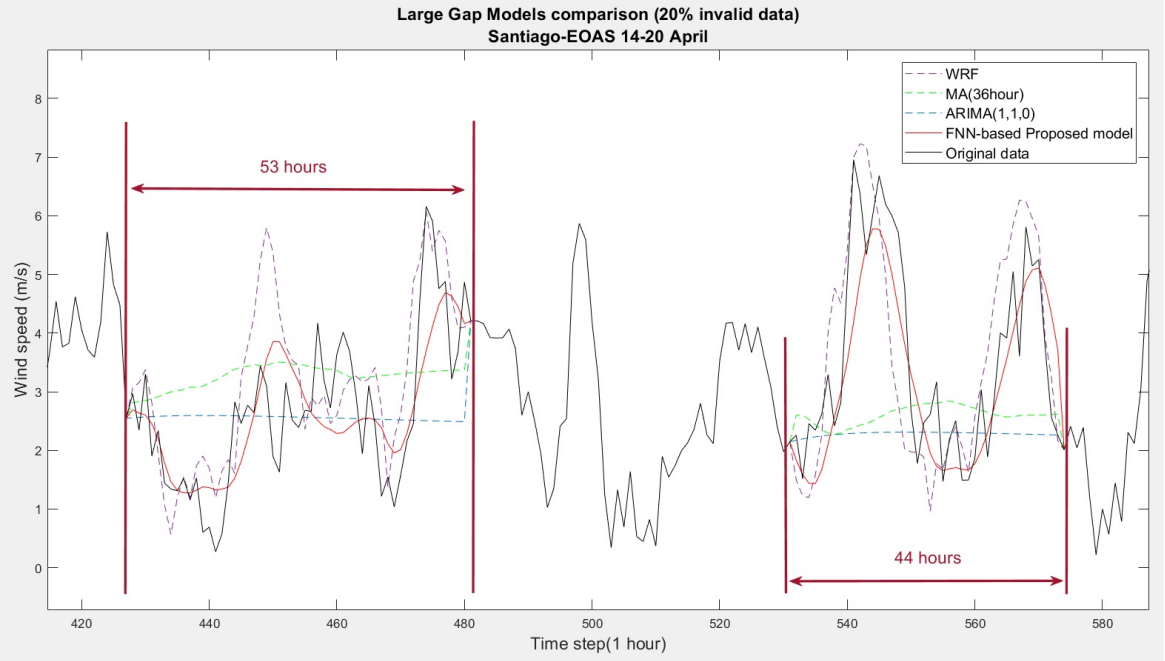


Fig. 5.18 An example of reconstruction models' results on large data gap

5.4 Summary

Quality of weather data, particularly wind speed and direction, is vital as they are inputs of wind speed and power prediction models. Considering this fact, in the current chapter, a new developed validation-reconstruction model was described and applied to guarantee the quality of weather data to be used as the WS&PF models' inputs. In this model development, several datasets from the different case studies (weather stations) were applied to a set of quality control tests such as Basic, Manipulation, and Range tests. In addition to Internal, Temporal and Spatial consistency tests were used to classify the dataset values as valid, missing, and invalid data. Then, the last two coded data groups were transformed to a new developed reconstruction module composed by a clustering part and a regression part. Using WRF data as input improved the accuracy of data being reconstructed by the feed-forward neural network empowered with k-means clustering. This new model was applied on different percentages of the missing patterns, and its results always outperformed other benchmark models, particularly for short consecutive data gaps.

Chapter 6

Short-term wind speed forecasting model: Hyperparameters optimization

Machine learning prediction models are comprised of various parameters, known as hyperparameters, which need to be determined before applying the model. Hyperparameters have a direct effect on the models' performance and their results accuracy. In a hybrid model composed of various techniques, each part has its own set of hyperparameters that needs to be tunned. In most literature studies, hyperparameters are determined separately for each technique, mostly running the model for a limited set of possible solutions. However, this simplified tunning approach ignores the inter-linkages between hyperparameters of different techniques. In chapter 4 hybrid models were introduced, including the different techniques applied, and the new MLP model developed in this work. In this chapter 6, the best set of hyperparameters for the three techniques included in this MLP model structure is obtained. A new formulation of hyperparameter selecting problem as a global optimization applied using a semi-exhaustive search algorithm. Then, this tunned new BBO-MLP model is applied to predict wind speed at four meteorological stations, Santiago-EOAS, Camariñas, Muralla, and M2-tower, for a range from very-short (10-minute) to short-term (1-3 hour) time horizon.

6.1 Introduction and background

Hyperparameter selection and learning algorithm are two important topics in machine learning prediction models. Almost any neural network-based model performs well with the correct hyperparameters values from its training on a particular dataset. However, selecting and tunning an algorithm for model training require significant efforts and experiments, in order to obtain the best hyperparameters values. In the following sections, some neural network

learning methods used in the literature to forecast short-term wind speed are summarized in conjunction with the hyperparameter selection problem in neural networks.

6.1.1 Short-term ANN wind speed forecasting models

Wind speed prediction has a wide range of applications in the meteorology and wind energy industry. In the energy sector, all the power system operations such as regulation, load and energy forecasting, unit commitment and scheduling, etc., are carried out within the following time horizons: Ultra short-term (seconds to minutes), short-term (up to a few hours), and medium-term (up to a few days) [113]. However, longer predictions such as sub-seasonal (up to one month) and seasonal (up to several months) are mainly used for wind energy assessment to deal with the energy resources uncertainties and guarantee low-risk investment in wind energy projects. Focusing on short-term wind speed forecasting, several machine learning methods are available in the literature. The process of training an ML model involves providing a learning algorithm with training data to learn from. In a neural network, as first step random values are set to its weights, which obviously will not provide very good results. In the process of training, the best set of weights that result in high prediction accuracy will be obtained. Supervised, unsupervised, and reinforced learning are three main groups of training methods in machine learning, with different applications. Considering the goal of training a wind speed prediction model, supervised learning is normally used for classification and prediction models based on an error function that acts as a teacher to guide the learning process [120]. Backpropagation (BP) methods (section 4.3.3.3) and optimization techniques (section 4.3.3.4) are two main groups of supervised learning used in neural networks.

Backpropagation is a widely used class of training algorithms in feed-forward neural networks. Several BP methods were introduced in section 4.3.3.3 and tested in section 5.3.2.1. BP algorithm mechanism is learning by the errors from the output layer to decrease the errors in the hidden layer, and are highly successful in problems without direct relationship between the input and output [260]. Training algorithms, BP included, have different characteristics and performance in terms of memory requirements and processing speed. For example, a Levenberg–Marquardt (ML) algorithm trains a neural network between 10 to 100 times faster than a GD [141]. Also, GD methods have several shortcomings,

- GD backpropagation does not guarantee a global minimum but a local minimum. Also, there is a problem with cross plateaus in the error function landscape caused by

non-convexity in the error functions of neural networks. However, Yann et al. [261] argued that in many systems this drawback is not a problem.

- As it was mentioned above, GD backpropagation algorithm is generally very slow because it requires small learning rates for stable learning.

Meta-heuristics algorithms are good alternatives for BP methods covering most traditional training methods drawbacks such as their low speed and the problem of being trapped in a local minimum. Meta-heuristic algorithms have been successfully implemented on shallow neural networks to speed up the training process by substituting the GD strategy with iterative, evolutionary strategy or swarm intelligence strategy [262, 197, 187, 204, 205].

Zhang et al. [263] proposed a short wind speed prediction model using GA as a training method instead of BP in a VMD-GA-MLP model. It was reported that prediction errors of GA-MLP are smaller than BP neural network. Wang et al. [203] used GA to optimize an FNN model for each intrinsic mode functions (IMFs) signal decomposed by empirical mode decomposition. Using the meta-heuristic method was to decrease the computation time. Liu et al. [206] developed two wavelet decomposition (WD)-MLP model using GA and PSO to determine the weights sets of the NN. Using the optimization techniques before BFGS training improved the results slightly. However, the effect of using WD techniques was more substantial. In order to prevent the GA/PSO components from being over-optimized, it is recommended to select the best combinations considering both the accuracy requirements and the characteristics of the original wind speed signals. Yan et al. [172] used the same optimization methods to improve the BP-MLP model. Tian et al. [264] improved the PSO algorithm by adapting inertia weight and acceleration factors of the algorithm and used it as a preliminary step training for an ML-MLP model. After applying this model on 300 time steps of hourly wind speed data (m/s), short term wind speed forecast achieved 0.1774, 0.1622, and 0.2354 for RMSE, MAE, and MAPE, respectively.

Liu et al. [199] used a mind evolution algorithm (MEA) to optimize weights of an EMD-MLP model for short-term wind speed forecasting, which outperformed the GA-MLP model. The proposed model improved results compared to a single MLP by 67.71%, 68.69%, and 67.52% for MAE, MAPE, and RMSE, respectively. Jia et al. [200] used an artificial bee colony (ABC) meta-heuristic algorithm to train a BP-MLP model for 10, and 30 minutes time intervals. The results outperformed both MLP and GA-MLP. MAE and RMSE for 20-min ahead were reported as 0.3232 and 0.3648, respectively (wind speed in m/s). Meng et al. [201] applied a crisscross optimization (CCO) algorithm in a hybrid model

(WPD-CCO-MLP) to train neural network weights. The proposed crisscross optimization algorithm showed an advantage over the backpropagation algorithm and PSO, and regardless of one-, three-, or five-step ahead prediction, the model outperformed other BP methods. MAE, MAPE, and RMSE, for one-hour ahead prediction, were 0.704, 10.532, and 0.938, respectively.

As wind speed forecasting models, there are many optimization techniques to train neural network models for wind power prediction. Abednia and Amjadi [202] proposed a short-term wind power prediction based on a neural network optimized by a novel optimization algorithm: chaotic shark smell optimization (CSSO). The algorithm is based on distinct shark smell abilities for localizing the prey. The method is used for one-hour-ahead wind power prediction. Averaged NMAPE, NMAE, NRMSE for different times of year were reported as 10.22, 4.95, and 6.11, respectively. In addition, a wavelet neural network was trained by an improved clonal selection algorithm (ICSA) by Chitsaz and his colleagues. The proposed model is used for the 6-hours time-interval wind power time series. Averaged NRMSE and NMAE for the proposed model were 13.24, and 9.70, respectively; while for the persistence model, 14.95 and 10.71, respectively.

6.1.2 Hyperparameters selection

Hyperparameter optimization in machine learning refers to choosing a set of parameters values for a learning algorithm or model. In other words, all the constraints, weights, learning rates, etc., which are defined in a model, are considered hyperparameters that need to be tuned to achieve acceptable results from the model. Various techniques there exist in the literature for solving hyperparameter optimization as: grid search, random search, Bayesian and gradient-based optimization, early-stopping based, etc.

In nonlinear problems, feature and hyperparameter subsets can evolve discontinuously in case of using ranking methods, e.g., mutual information in which subset is built by adding features in the order defined by ranking. This discontinuous behavior means that the best subset of size d is not necessarily repeated in the best subset of size $d+1$ [265]. Considering this fact, the space of possible subsets should be searched in an optimized way to find a set that results in global maximum model performance. Hyperparameter optimization finds a set of hyperparameters that yields an optimal model which minimizes a predefined loss function on given test data. The objective function takes a tuple of hyperparameters and returns the associated loss [266]. In machine learning, a hyperparameter is a parameter whose value is used to control the learning process. By contrast, the values of other parameters (typically

node weights) are derived via training. Therefore, hyperparameters can be defined as model hyperparameters that cannot be inferred while fitting the machine to a training set. That is because they refer to the model selection task, or algorithm hyperparameters, that initially have no influence on the model's performance but affect the speed and quality of the learning process. An example of a model hyperparameter is the topology and size of a neural network. Examples of algorithm hyperparameters are learning rate and mini-batch size. Generally speaking, the selection of optimization methods parameters used to train ANN models is still dependent on experience and experimental work [264].

The number of hyperparameters is normally below five. However, it could be easily increased to hundreds for complex learning algorithms, or when processing steps are also subjected to optimization [266]. It has been empirically demonstrated that in many cases, only a limited number of hyperparameters have a significant effect on the model performance, although it is not an easy task to identify them in advance [267]. Some hyperparameters are conditional upon others. For example, in optimizing the neural network structure [268], the number of hidden layers is one hyperparameter, and the size of each layer is another hyperparameter conditional upon the first one. However, a fully automatic self-configurable machine is the ultimate goal in prediction models, but the current state-of-the-art shows that there is still a long way to walk before reaching this goal. There are many techniques used in the literature to address this problem, mainly in other areas than WS&PF, using meta-heuristic algorithms such as GA, PSO, ACO, etc. [266]. In general, algorithms closer to grid search provide more accurate results, but more computational time is required. For this reason, in this work a semi-exhaustive search algorithm is developed as a compromise between the computational time and a high level of certainty in obtaining the global maximum. Taking into account that all the listed hyperparameters in this work are integer numbers and not a continuous function, increasing the computation complexity.

In most of the ANN models used for WS&PF, hyperparameters are either extracted from the literature or set by test, and trial [269–272]. Gan et al. [273] developed a temporal convolution neural network (TCN) for wind speed forecasting, defining the number of nodes per layer in shallow and deep networks and the proposed TCN model empirically, and determined through trial and error in order to avoid dimensional disasters when using PSO. Li et al. [274] extracted hyperparameters from the literature and defined some of them by experiment with a hybrid deep interval prediction model for wind speed forecasting. Ye et al. [22] set multiple starting points to tune the hyperparameters of a Bayesian emulator as a nonparametric, probabilistic model for wind speed forecasting. The reason was to solve the

defined optimization problem avoiding suffering from local optimum. Tian et al. [275] used an improved whale optimization algorithm (IWOA) to optimize the (hyper)parameters of an eco state neural network (ESN) model, including weight spectrum radius (SR), reservoir size (N), input scale (IS) and sparsity degree (SD). MAPE, MAE, and RMSE for one-hour-ahead wind speed prediction (m/s) of the proposed model composed of variation model decomposition and IWOA-ESN techniques were reported as 10.2424, 0.7425, and 0.8544, respectively. Different structures such as number of sub-layers, number of main components of a sub-layer, and number of inputs are investigated through sensitivity analysis to find out the best structure in a singular spectrum analysis (SSA) in a deep learning neural network model proposed by Mi and Zhao [276] for wind speed prediction. By comparing RMSE of 1-step ahead with 3-step ahead, it revealed that changing structure has less effect than changing the method from SSA-BP to the proposed model.

6.2 New wind speed forecasting model and methods

In this section, the new model jointly to the methods applied are introduced. Considering the previous works above discussed, the new forecasting model and its associated methods are based on the following main points,

1. Provide a new formulation to address hyperparameters optimization problem in shallow neural networks, including training algorithm
2. Apply a new and promising optimization technique (BBO) to train an MLP model for short-term wind speed prediction (BBO-MLP)
3. Conduct a comparative study on different MLP structures, activation functions, inputs features, input data lengths, optimization methods with different generation numbers, parents, and range values.
4. Introduce successful new features and activation function for the forecasting model

6.2.1 Methods

6.2.1.1 Search algorithm

Grid search, is the basic method of optimization problem solving that considers all possible solutions of the problem to obtain the best result. This method is also known as exhaustive search or brute-force search. The positive point about grid search is that no

possible solution is left behind. However, the main drawback is that the larger the number of hyperparameters are used to be optimized, the longer that this method takes to run [277]. On the other hand, random search, refers to parameters value update based on the results of the previous trial. Exhaustive search is not viable in the majority of cases due to the high number of possible solutions. Particularly, prediction model training using grid search optimization requires a huge time amount, around several months in a personal computer. Although random search can get a solution very fast, it does not guarantee the optimum and, also, the achieved solution could not be representative for all the parameters range [278].

The critical step in a feasible hyperparameter optimization is to choose the set of trials $\lambda(1)...\lambda(S)$. The most widely used strategy is a combination of grid search and manual search as well as the use of machine learning software packages such as libsvm [279]. Considering K configuration variables, then grid search requires to choose a set of K values for each variable {I(1)...I(K)}. In grid search the set of trials is formed by assembling every possible combination of those values, so the number of trials in a grid search is $S = \prod_{i=1}^{I} |I^{(K)}|$ elements. This product over L sets makes grid search to suffer the dimensionality curse because the number of joint values grows exponentially with the number of hyperparameters [280].

6.2.1.2 Multi-layer perception neural network

The new prediction model uses a multi-layer perceptron neural network, as in the reconstruction model described in Chapter 5, which neurons in each layer are fully connected to the adjust layers. Values move in only one direction from the input layer to the output layer. A weight is assigned to each connection between the neurons and a bias value is assigned to each neuron. This type of neural network is commonly-used in ANNs. The output of an MLP model y_i could be obtained through a the non-linear formulation as follows:

$$y_t = \zeta \left(\sum_{h=1}^{H} k_h \left(\eta \sum_{i=1}^{I} w_{i,h} x_i + b_i \right) + bb_h \right) \tag{6.1}$$

Where ζ and η are output and hidden activation functions, respectively. h is a hidden layer neurons from 1 to H, and i is an input layer neurons, from 1 to I. w and k are input-hidden and hidden-output connection weights, respectively. b and bb are biases for the hidden

and output layer neurons, receptively. Finally, y_t is the predicted value at the time step t.

Figure 6.1 shows a multi-layer perceptron with three different layers and unlimited neurons in input and hidden layers. The output layer normally has one neuron, although it could have more. Bias is added to the hidden and output layers neurons.

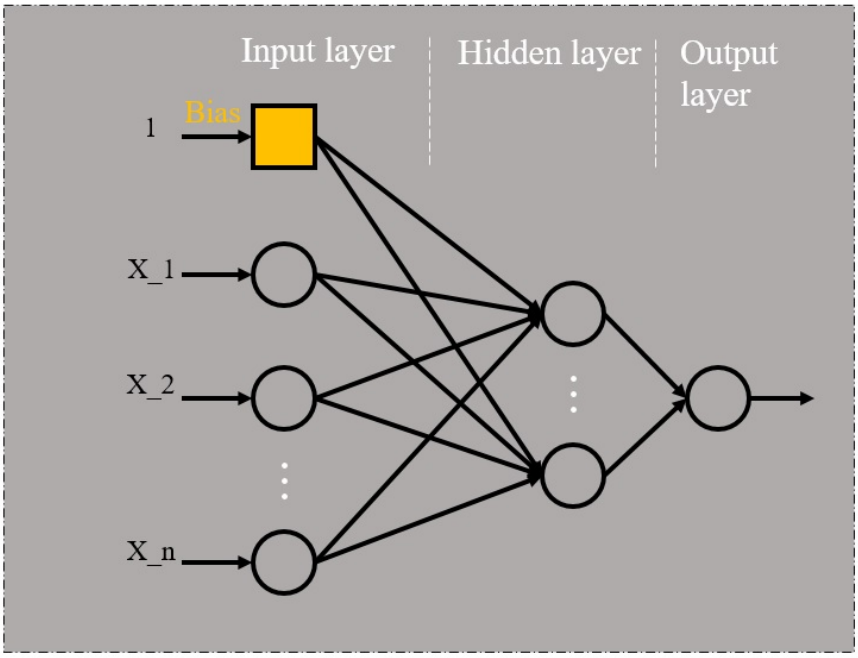


Fig. 6.1 Schematic view of a Multi-layer perceptron network

6.2.1.3 Biological-based optimization

In order to train the weights of this new MLP model, a novel meta-heuristic optimization BBO algorithm is presented, as it is not used for wind speed prediction before. Inspiring by bio-geography, BBO algorithm was proposed by Simon [281] in 2008. Simon proved that BBO outperformed some commonly used algorithms such as GA, PSO, ACO, ES and PBIL on 14 benchmark functions and a real problem. The core idea in BBO is based on biological evolution over time and space. It might include different ecosystems like habitats or territories to find out the relationships between different species and the effects of mutation, immigration or emigration on their evolution. As other evolutionary algorithms (EA) like GA [199], in BBO there are some agents called habitats with a similar role as chromosomes in GAs. As a representative of the problem variables, a vector of habitants is set to each habitat; in addition, an index defines the overall fitness of an habitat, and called Habitat Suitability Index (HSI). Habitats evolve over time due to three main rules [282]:

- The possibility of immigration between habitants living in habitats with high HSI is higher than in habitats with low HSI;
- Habitats with high HSI are less prone to attract new immigrant habitants from those with low HSI and vice versa;
- Random changes might happen in habitants of all habitats regardless of their HSI values.

In natural environment, these rules cause balance between different ecosystems. using these rules, the BBO improves the HSI of all habitats which makes evolving from the initial random solution for the problem.

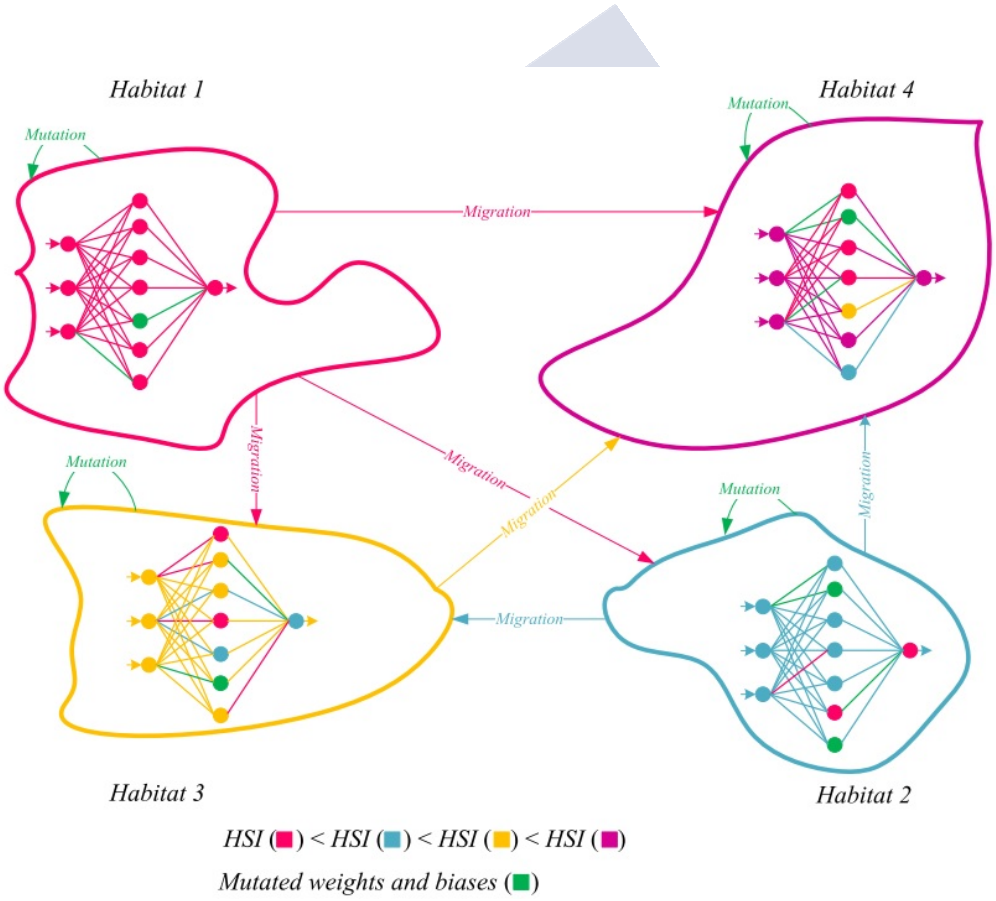


Fig. 6.2 BBO: Conceptual model of migration between the habitats for training MLP [262]

An initial random set of habitats with n different habitants inside of each of them is defined as the first step. Different immigration β_t, emigration ν_t, and mutation τ_t rates are defined for each habitat to mimics the characteristics of various nature geographically separated locations. Considering n as the number of current habitants, N as the maximum

number of habitants, ME as the maximum number of emigration rate, MI as maximum immigration rate, m as the initial mutation's value, p_n as the mutation probability of the nth habitat, and $p_{max} = argmax(p_n)$ where $n = 1,2,3,...,N$, the following equations are given:

$$\beta_t = \frac{EM \times n}{N} \tag{6.2}$$

$$\nu_t = IM \times \frac{1-n}{N} \tag{6.3}$$

$$\tau_n = m \times \left[1 - \frac{p_n}{p_{max}}\right] \tag{6.4}$$

In this neural network optimization algorithm, the fitness function is the mean square error (MSE) for all training samples, defined as follows,

$$E = \sum_{t_1}^{b} \frac{\sum i = 1^j (Out_i^t - d_i^t)^2}{b} \tag{6.5}$$

Where b is the number of training samples, j is the number of outputs, d_i^t and Out_i^t, are the desired and actual output of the ith input when the tth training sample is used.

6.2.1.4 Time series trend and moving averaging methods

An important issue in MLP is the trend of the time series that describe the system behavior. This trend is usually determined by several averaging methods [283]. Give an time series of n time steps $x = \{x_1, x_2, ...x_t, ...x_n\}$, simple moving average (SMA) and weighted moving average (WMA) for the last k parameters are defined in equations 6.6 and 6.7, respectively. Exponential moving average (EMA) is shown in equation 6.8, as a recursive method. WMA_k is the value of EMA at any time period k. Two types of EMA can be used, depending on the parameters definition: Using last k observations, in the first one $\alpha = 2/k+1$ [283], while in the second one $\alpha = 1/k$ [284].

$$SMA_k = \frac{1}{k} \sum_{i=n-k+1}^{n} x_k = \frac{x_{n-k+1} + x_{n-k+2} + ... + X_n}{k} \tag{6.6}$$

$$WMA_k = \sum_{i=n-k+1}^{n} x_k = \frac{(1 \times x_{n-k+1}) + (2 \times x_{n-k+2}) + ... + (n \times X_n)}{\sum_{j=1}^{n} j} \tag{6.7}$$

$$WMA_k = \begin{cases} x_1 & t = 1 \\ \alpha x_1 + (1-\alpha) x_{t-1} & t > 1 \end{cases} \tag{6.8}$$

Given a dataset $S = \{X_1, X_2, X_3, ...x_n\}$ containing n multivariate time series $X_1 = \{x_1, .., x_t, ..\}$, the arithmetic mean (AM) and weighted arithmetic mean (WAM) for time step t are defined in equations 6.9 and 6.10, respectively. As an example of AM and WAM application, some features of the new wind speed forecast are defined using those averages, as it will be detailed in the corresponding section.

$$AM_t = \frac{1}{n} \sum_{i=1}^{n} X_{i,t} = \frac{X_{1,t} + X_{2,t} + ... + X_{n,t}}{n} \tag{6.9}$$

$$WAM_t = \sum_{i=1}^{n} \omega_i X_{i,t} = \frac{\omega_1 X_{1,t} + \omega_2 X_{2,t} + ... + \omega_n X_{n,t}}{\sum_{j=1}^{n} \omega_i}, \quad \sum_{j=1}^{n} \omega_i = 1 \tag{6.10}$$

6.2.1.5 Model selection criteria

Model selection is an important step of developing a prediction model which deals with two main problems of overfitting (good performance on the training data, but poor generalization to other datasets) and underfitting (poor performance on the training data, and poor generalization to other datasets). Indeed, a trade-off between the goodness of fit of the model and the simplicity of the model should be considered.

Considering different models defined by their corresponding set of parameters θ applied on a historical dataset $X = \{x_1, x_2, \ldots, x_n\}$, model error can be defined by $\mathscr{T}(X, \theta)$ which is named **in-sample** loss or experimental risk. The simplest way to chose a model is looking for such a model that minimize this experimental risk. However, for a prediction model this approach is not completed, as prediction model performance requires to minimize model errors over future data, not only historical data. Then, minimizing **out-of-sample** loss $\mathscr{T}(\mathscr{X}, \theta)$ is required. Hence, expected loss on new data, $\mathbb{E}[\mathscr{T}(\mathscr{X}, \theta)]$, is the error

of interest, called **risk** or **generalization error**. In conclusion, the in-sample loss can be expressed as,

$$\mathscr{T}(\mathscr{X},\theta) = \mathbb{E}[\mathscr{T}(\mathscr{X},\theta)] + \eta_n(\theta) \tag{6.11}$$

Where $\eta_n(\theta)$ is random noise with mean zero. Hence, the selection model problem can be formalized as equation 6.12 assuming $n \to \infty$, $\eta_n(\theta) \to 0$ which rests on the law of large numbers.

$$\hat{\theta} = \underset{\theta \in \Theta}{argmin}\, \mathbb{E}[\mathscr{T}(\mathscr{X},\theta)] + \eta_n(\theta) \tag{6.12}$$

In order to solve the above optimization problem, some criteria were defined using the out-of-sample error estimation. Two of them are widely applied in previous research works, so they are described in the next two sections.

6.2.1.5.1 Akaike information criteria (AIC)

The underlying idea of information criteria is to find an optimal trade-off between an unbiased approximation of the underlying model and the loss of accuracy caused by estimating an increasing number of parameters. To achieve this, information criteria combine some measure of fitness with a penalty term to account for model complexity. AIC is found based on Information Theory by a Japanese statistician Hirotugu Akaike. This criteria estimates the quality of various models by estimating the out-of-sample error [285]. In other words, AIC provides a means for model selection through which the model with minimum AIC is considered to be the preferred model. Indeed AIC makes a trade-off between the simplicity of the model and the goodness of model fitness (likelihood function).

Suppose that θ is the number of estimated parameters in the model and $\hat{\mathscr{T}}(\mathscr{X},\theta)$ as the maximum value of likelihood function for the model. Then, AIC is defined as follows,

$$AIC = 2\theta - 2ln(\hat{\mathscr{T}}) \tag{6.13}$$

Given a set of candidate models $\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$, the model with the minimum AIC value is the selected model. As it is shown in the equation 6.13, AIC puts value on the goodness of model fitness and impose negative mark for increasing the numbers of estimated parameters, which indirectly prevent the model to be overfitted.

6.2.1.5.2 Bayesian information criteria (BCI)

BIC is another model selection criterion, developed by Schwartz in 1978 [286]. Similar to AIC, there is a penalty term in BIC for increasing estimating parameters which result in overfitting. However, the weight of this penalty is bigger in BIC compared to AIC. Also, as in ACI, between a series of candidate models the one with lower BIC value is the preferred model.

Given θ as the number of estimated parameters in the model, $\hat{\mathscr{T}}(\mathscr{X},\theta)$ as the maximum value of likelihood function, and n as the sample size, BIC is presented as follows:

$$BIC = \theta ln(n) - 2ln(\hat{\mathscr{T}}) \tag{6.14}$$

BIC has two main disadvantages: First, sample size, n,s should be much bigger than the number of estimating parameters, θ and second, BIC criteria can not handle a complex collection of models, like high-dimension features prediction models.

In the new developed prediction model, after each step in solutions space that provides different candidate models in a specific subspace, both AIC and BIC based on out-of-sample loss are applied to sort the models, and the first 5 best models are chosen to the net phase.

6.2.2 Problem formulation

At this point, mathematical formulation of the model fitness is required, divided in two parts: First, formulating the hyperparameters optimization problem, using a semi-exhaustive search algorithm to find the best set of hyperparameters. Second, formulating a training-hyperparameter optimization problem to find the best set of hyperparameters values, conditioned on optimized weights of the neural network. The general formulation of the problem as a conditional stochastic optimization is represented below:

$$\underset{x \in X}{\text{Argmin}}\, F(x) = \mathbb{E}_{\zeta} f_{\zeta}(\, \mathbb{E}_{\eta|\zeta}\, g_{\eta}(x,\zeta)\,) \tag{6.15}$$

Where ζ and η are both random vectors, $X \subset R^d$, $g_{\eta}(x,\zeta) : R^d \to R^k$ is a vector-valued function dependent on ζ and η , $f_{\zeta} : R^k \to R$ is a function dependent on ζ. The innermost expectation is taken respect to the conditional distribution of $\eta|\zeta$. The assumption here is to have access to samples from the distribution $P(\zeta)$ and the conditional distribution $P(\zeta|\eta)$

and not to the cumulative distribution function of these distributions.

Adapting this general formulation to neural networks application, the goal of hyperparameter search optimization in conditioned on network's weights optimization is to find a set of hyperparameters ζ^* and network weights $\eta^*|\zeta^*$ that yield an optimal model $\mathscr{M}$ minimizing the loss function ($\mathscr{L}(X^{(valid)};\mathscr{M})$) on the validation dataset. This formulation is provided below:

$$\{\zeta^*,\eta^*|\zeta^*\} = \arg\min_{\zeta} \mathscr{L}(X^{(valid)}; g_{\eta^*}(Y^{(valid)},\zeta)) = \arg\min_{\zeta} F(\zeta, g_{\eta^*}, X^{(valid)}, Y^{(valid)}, \mathscr{L}) \quad (6.16)$$

$$\eta^* = \arg\min_{\eta} F(\zeta, g_{\eta}, X^{(valid)}, Y^{(valid)}, \mathscr{L}) \quad (6.17)$$

Where loss function $\mathscr{L}$ is defined based on mean square error, and $X^{(valid)}$ and $Y^{(valid)}$ are test and train datasets, respectively.

6.2.3 Search algorithm

Following previous formulation, a flowchart of the semi exhaustive search algorithm developed in this work is shown in figure 6.3. Also, the corresponding pseudocode (Algorithm 2) is shown.

As shown in figure 6.4 by dividing the general solution space into subspaces of the model, optimization, and data hyperparameters, a grid search algorithm will be applied to each subspace accordingly. The best results from each cycle will be chosen through a model selection process and used for the next cycle. Outcome solutions of each subspace will go through a model selection process with two stages. In the first stage, models and hyperparameters are ranked based on the loss function's outcome. Top 10% of the best models (about 250 models) are qualified for the next step in which a 10-fold cross-validation will be applied to the models. 10-fold cross-validation would perform the fitting procedure a total of ten times, with each fit being performed on a training set consisting of 90% random selected of the total training set selected at random, with the remaining 10% used as a hold out set for validation. The best five models based on BIC and AIC criteria are selected and

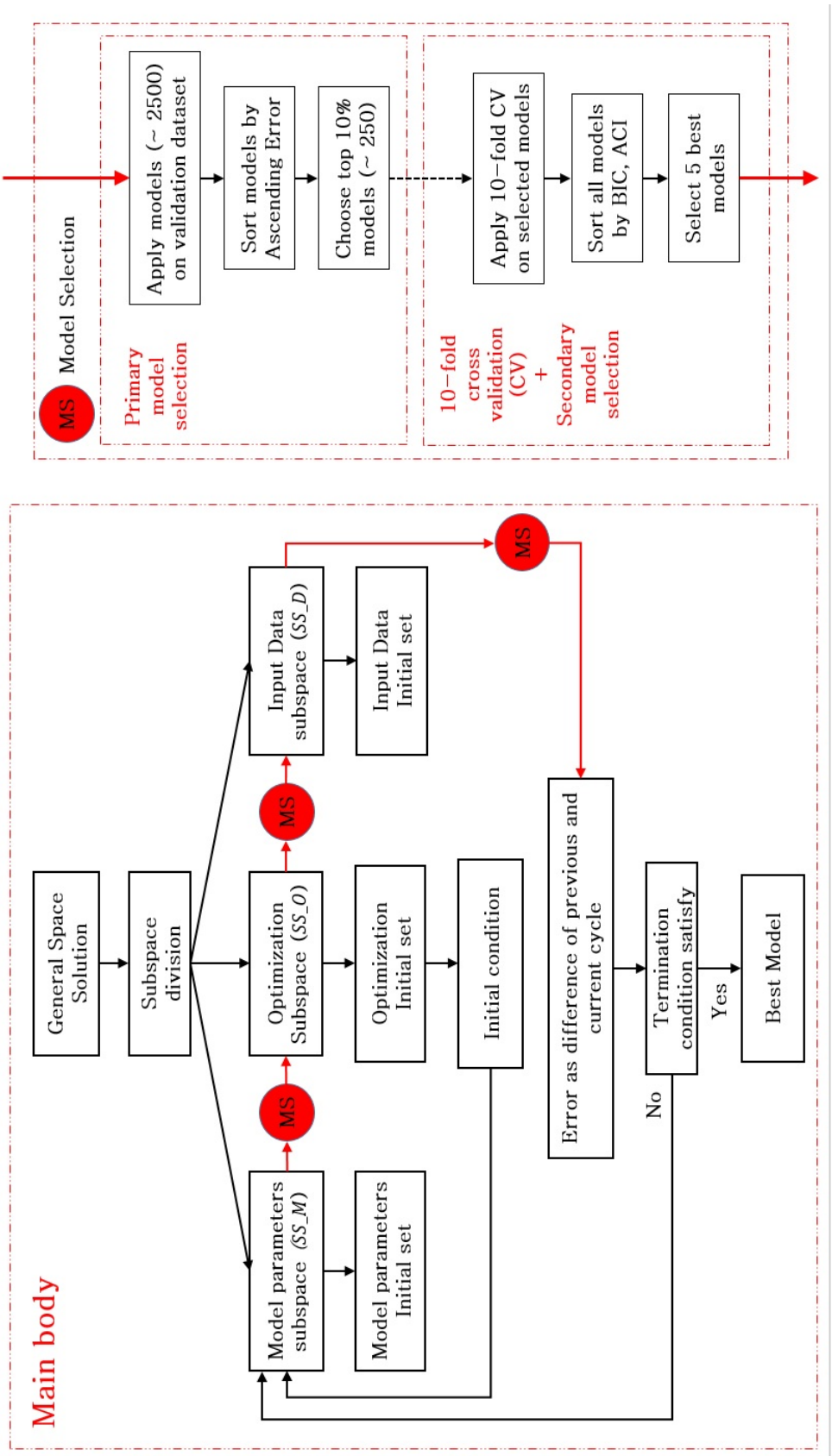


Fig. 6.3 Flowchart of the proposed methodology based on semi-exhaustive search

Algorithm 2 Semi-exhaustive search algorithm

Define the best solution, *Global_Best*, as *GB*
Define solution space composed of three sub-spaces (model, optimization, data):
$S_i = \{S_1, S_2, \ldots, S_n\}$ = $\{M_i = \{m_1, m_2, m_3\}, O_i = \{o_1, o_2, o_3\}, D_i = \{d_1, d_2, d_3\}\}$
Each Subsubspace, $\{m_i\}$, $\{o_i\}$, $\{d_i\}$ is a set of hyperparameters with different dimensions
Assign initial conditions to sub-spaces and put them in $GB = \{M_{init}, O_{init}, D_{init}\}$
while Stop condition is not true **do**
  for each subspace $\{SS_M\}$, $\{SS_O\}$, $\{SS_D\}$ in order **do**
    Brute-force search and update *GB* with the outcomes
  end for
end while

fed to the next subspace, where the hyperparameters of the other two subspaces are fixed, based on the best outcome of the previous step. In each iteration around 5,000 models need to be computed. By comparing the loss function results from the current subspace grid-search with the previous subspace, the termination criteria are achieved if the MSE difference is below 0.01. Otherwise, subspace searching will continue.

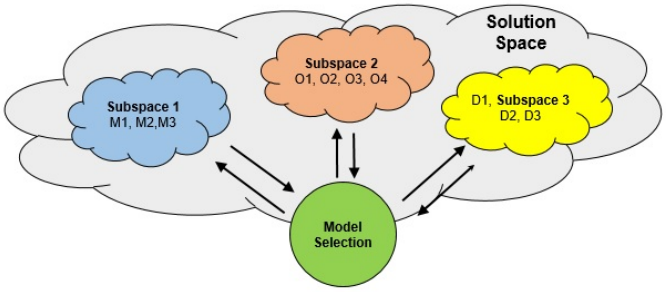


Fig. 6.4 schematic view of solution space and subspaces

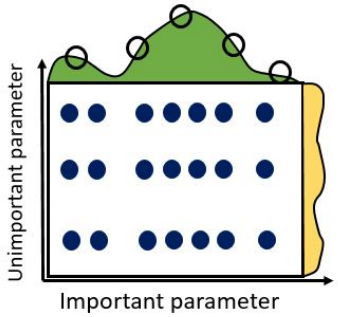


Fig. 6.5 Schematic depiction of grid search layout

Figure 6.5 shows the search process inside each subspace [267]. Some hyperparameters play a more important role than others. The X-axis shows an example of an important parameter which function is depicted on top of the box. The Y-axis shows a less important hyperparameter that is shown on the left side of the box. The dots represent a hyperparameter values, as clearly more values means more probability to achieve the best result. As mentioned in section 6.2.1.1, either adding new values to a hyperparameter or adding new hyperparameters to a subspace, the number of required calculations for grid search grows exponentially.

6.2.3.1 Final prediction model components

In table 6.1 all hyperparameters of three different model spaces are listed. As shown, subspace of models hyperparameters SS_M contains: M_1: six different input node numbers, M_2: four approaches to define hidden nodes numbers, and M_3: five different combinations of activation functions. Of course, the learning method belongs to the models, but in order to better representing in the model structure, and taking into account the complexity of the optimization techniques adopted, a separated group of optimization hyperparameters is defined as a new subspace $SS_O\}$, which is composed by: O_1: A novel optimization technique (BBO) in conjunction with three more commonly-used methods: GA, PSO, and ACO, O_2: Five different parents number for the optimization methods, O_3: Twelve different numbers of

Table 6.1 Subspaces of hyperparameters: model, optimization, and data

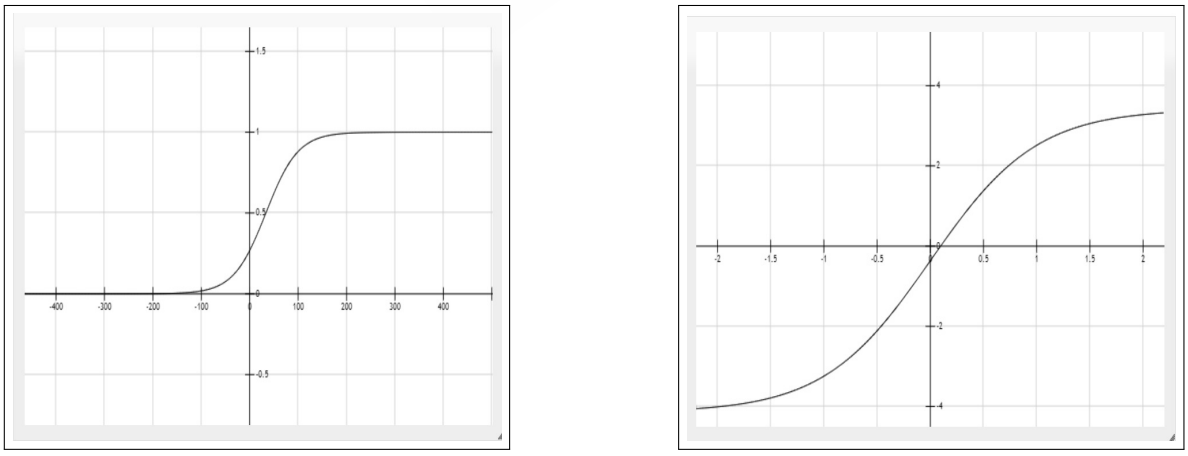
H.P.	Description	Values
M_1	input nodes number	$\{1,3,5,10,20,50\}$
M_2	hidden nodes number	$\{(n,n,1),(n,2n,1),(n,2n+1,1),(n,random,1)\}$
M_3	activation function	{Sigmoid-proposed, sigmoid-tanh, sigmoid-ReLU, tanh-ReLU}
O_1	optimization method	{GA, ACO, PSO, BBO}
O_2	parents number	$\{25,50,100,150,200\}$
O_3	Generation number	$\{25,50,75,100,125,150,175,200,225,250,275,300\}$
O_4	weights value range	$\{5,10,20\}$
D_1	Length of data (time step)	{60, 180, 360, 720, 1440, 4320, 10080, 20160, 40320}
D_2	(train-validaton) / test%	{50/50 , 60/40, 70/30, 80/20, 90/10}
D_3	pre-selected features	combination of different weather parameters (Just for single input, twelve different inputs are created using SMA, WMA, and EMA)

generation length for the optimization methods, and O_4: three different ranges for the network weights. The effect of different data hyperparameters is investigated through data subspace SS_D with the following three items: D_1: Nine different possibilities of time series length, i.e., its number of time steps, D_2: five different possibilities of the (training-validation)-testing percentage of the dataset, and D_3: ten different feature selection approaches.

Table 6.2 shows different activation functions and their formulas used as M_3 model hyperparameters. Considering various activation functions for hidden and output layer, more than 100 combinations were tested in a preliminary stage, and the best five nominated combinations were selected as the values of M_3 shown in table 6.2, including a new function as modified hyperbolic tangent. Also, in figure 6.6 both modified sigmoid and modified hyperbolic tangent (new function) activation functions are plotted.

Table 6.2 Applied activation functions description

Name	Function	Range
Rectified linear unit (ReLU)	$\begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}$	$[0, \infty)$
Sigmoid	$\frac{1}{1+e^{-x}}$	$[0,1]$
Modified sigmoid	$\frac{1}{1+e^{-0.03x-1}}$	$[0,1]$
Hyperbolic tangent (tanh)	$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$[-1,1]$
Modified tanh	$3.78 \times tanh(x) - 0.364$	$[-4,4]$



(a) Modified sigmoid function

(b) Modified hyperbolic tangent

Fig. 6.6 Modified sigmoid and hyperbolic tangent activation functions

Table 6.3 shows single inputs as created features under data hyperparameter D_3. As explained before in section 6.2.1.4, simple moving average (SMA), weighted moving average (WMA), exponential moving average (EMA) [283], Arithmetic mean (AM), and weighted arithmetic mean (WAM) are used to generate different combinations of wind speed (WS) and wind peak (WP) as input data.

It was observed that wind time series (TS) is dependent on the atmospheric stability, particularly TS are more fluctuating in unstable conditions, difficulting the wind prediction. Table 6.4 [287] shows the relationship between Richardson number values and Pasquill stability classes. As evident, the most unstable conditions correspond to $Ri < 0.86$, so a time window for such conditions was selected at any time, as the most wind fluctuating conditions.

Table 6.3 Simple and hybrid features using SMA, EMA, WMA, AM, and WAM

Item	Input features
1	$WS_{10m,(t-1)}$
2	$SMA_{10m}(WS_{(t-3)}, WS_{(t-2)}, WS_{(t-1)})$
3	$SMA_{10m}(WS_{(t-5)}, WS_{(t-4)}, ..., WS_{(t-1)})$
4	$WMA_{10m}(WS_{(t-5)}, WS_{(t-4)}, ..., WS_{(t-1)})$
5	$EMA1_{10m}(WS_{(t-5)}, WS_{(t-4)}, ..., WS_{(t-1)})$
6	$EMA2_{10m}(WS_{(t-5)}, WS_{(t-4)}, ..., WS_{(t-1)})$
7	$MWM_{10m}(WS_{(t-3)}, WP_{(t-3)}, WS_{(t-2)}, WP_{(t-2)}, WP_{(t-1)})$
8	$MSM(WS_{20m,(t-1)} - WS_{5m,(t-1)})$
9	$MSM(WS_{10m,(t-1)}, WS_{20m,(t-1)}, WS_{5m,(t-1)})$
10	$MWM(WS_{10m,(t-1)}, WS_{20m,(t-1)}, WS_{5m,(t-1)})$
11	$MWM(WS_{10m,(t-1)}, WP_{10m,(t-1)}, WS_{5m,(t-1)}, WS_{20m,(t-1)})$
12	$MWM(WS_{10m,(t-1)}, WP_{10m,(t-1)}, WS_{5m,(t-1)}, WP_{5m,(t-1)}, WS_{20m,(t-1)}, WP_{20m,(t-1)})$

Table 6.4 Atmospheric stability classes based on Richardson number

Stability class	Atmospheric description	Richardson number
A	Highly unstable	$R_i < -0.86$
B	Unstable	$-0.86 \leq R_i < -0.37$
C	Slightly unstable	$-0.37 \leq R_i < -0.1$
D	Natural	$-0.16 \leq R_i < 0.053$
E	Slightly stable	$0.053 \leq R_i < 0.134$
F	Stable	$0.134 \leq R_i$

6.2.3.2 Initial conditions

The proposed search algorithm will start by setting initial conditions for two subspaces of optimization and data, and applying a grid search on the model subspace. It is important that initial hyperparameters values will be optimal, in order to assure the search algorithm achieves the optimum in all subspaces. Best initial hyperparameters values were collected from previous works, as they are shown in table 6.5.

Table 6.5 Initial condition for optimization and data subspace

H.P.	Description	Initial value	Ref.
O_1	Optimization method	GA	[199]
O_2	parents number	200	
O_3	Generation number	50	
O_4	weights value range	5	
D_1	Data length	One week	[263]
D_2	(train-validaton) / test%	80/20	[288]
D_3	pre-selected features	WS_{t-1}	[289]

6.2.4 Evaluation criteria

To evaluate the results accuracy and compare the performance of the new model against other models, the commonly-used evaluation criteria in the literature are selected: Mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), as defined in equations 6.18 to 6.21, respectively.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - P_i)^2 \tag{6.18}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - P_i)^2} \tag{6.19}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{P}_i - P_i| \tag{6.20}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{P_i - \hat{P}_i}{P_i}\right| \tag{6.21}$$

Where P_i stands for the real value at ith time step, $\hat{P}_i$ is the predicted value, and $\bar{P}$ is the average value of N data points [194].

6.3 Results and discussion

This section has two main parts, that correspond to model training and test prediction model phases. In the first phase, M2-tower case study with 1-minute time interval data (high resolution), is used to find out the best set of hyperparameters based on a Pareto optimal approach of AIC, BIC, and MSE. In the second phase, four different meteorological stations are used to test the model for very short (10-minute) and short-term (1-3 hours) wind speed prediction. Pareto optimal applied in a situation where no individual or preference criterion can be better off without making at least one individual or preference criterion worse off or without any loss thereof [290].

6.3.1 Hyperparameter optimization

6.3.1.1 Hyperparameters definition

Considering all the possible combinations in the new prediction model, the space solution has $3,110,400$ possible solutions, which needs a total number of $9,331,200,000$ model runs (including heuristic optimization generations of 300 and 10-fold cross-validation) to find the best result using a grid algorithm. Running CPU for a long time may cause an overheating problem which results in slowing down the microprocessor speed [291]. As

a rough estimation ignoring the effect of time and heat effect on CPU functionality and by considering an underestimated average of 0.01 seconds for each model run (based on the results), it will last about 12 years to run a grid search for all the possible solutions of the problem, using a typical personal computer with characteristics listed in table 3.7. Although applying the described semi-exhaustive search , it will last 2 weeks in getting the global optimum. Table 6.6 shows a comparison between grid search and the developed semi-exhaustive search algorithm in terms of the required models to be run and the computation time as discussed above.

Table 6.6 Required computation time for grid search vs. proposed search algorithm

Search algorithm	**Space Solutions**	**No. models**	**Aprox. run time**
	$\|S\| = \|M\| \times \|O\| \times \|D\|$	$G \times (n \times \|S\|) + CV$	0.01 sec/model
Brute-force	$3,110,400$	$9,331,200,000$	12 years
Semi-exhaustive	$3,110,400$	$417,600$	14 days

Note: S stands for solutions space, M, O and D are abbreviations for Model, Optimization, and Data subspace solutions, respectively, G stands for optimization generation number = 300, and CV represents 10-fold cross validation = 10.

6.3.1.2 Model hyperparameters

Graphical comparison between different members of each model parameter is shown in figure 6.8. It includes all three parameters of input number of neurons and hidden number of neurons (which is the outcome of different structures).

In figure 6.7 it is clear that the new sigmoid function for hidden and output activation functions (Sigmoid-pro) is performing better (with the highest correlation index) than the other functions. Also, in figure 6.7 inputs number is represented by a colors scale. It is clearly shown that both time and (train/validation) error (mean square error) is much less for the lower number of inputs. Obviously, the higher the number of inputs and outputs are, the more time is required to (train/validation) the model. These results is aligned with [127] and [32], in which the simplest model outperforms other structures. It is important to mention that the validation error is higher than the (train/validation) error, which is taken as shown in the figure 6.7. However, the selected training/validation methodology (previously described) is not based on the (train/validation) error but a combination of model selection criteria (AIC,

BIC) and a 10-cross validation strategy.

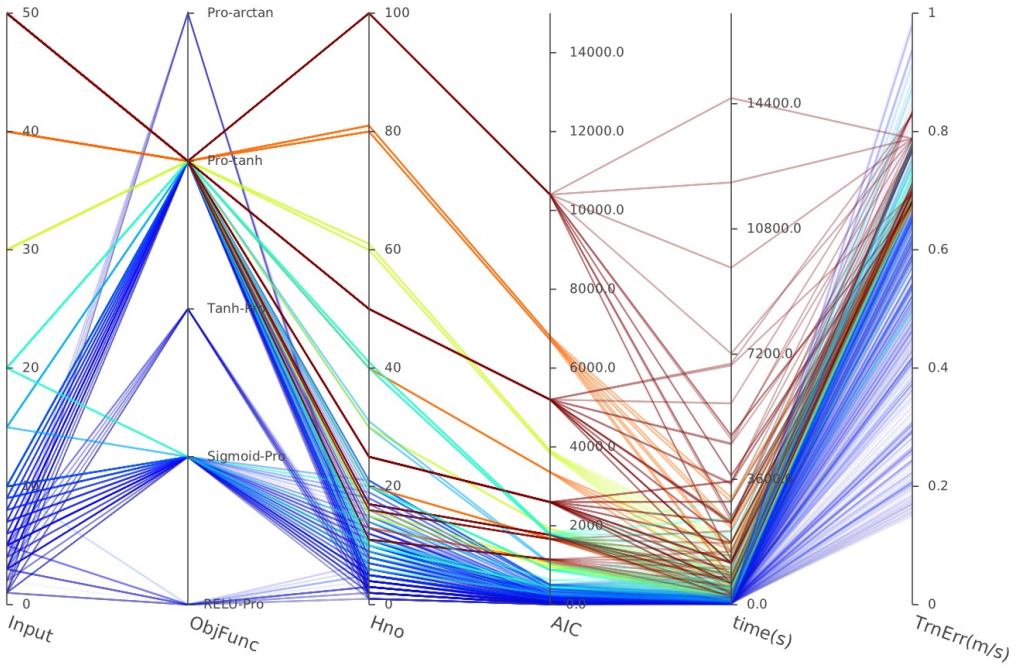


Fig. 6.7 Parallel coordinate plot of model hyperparameters

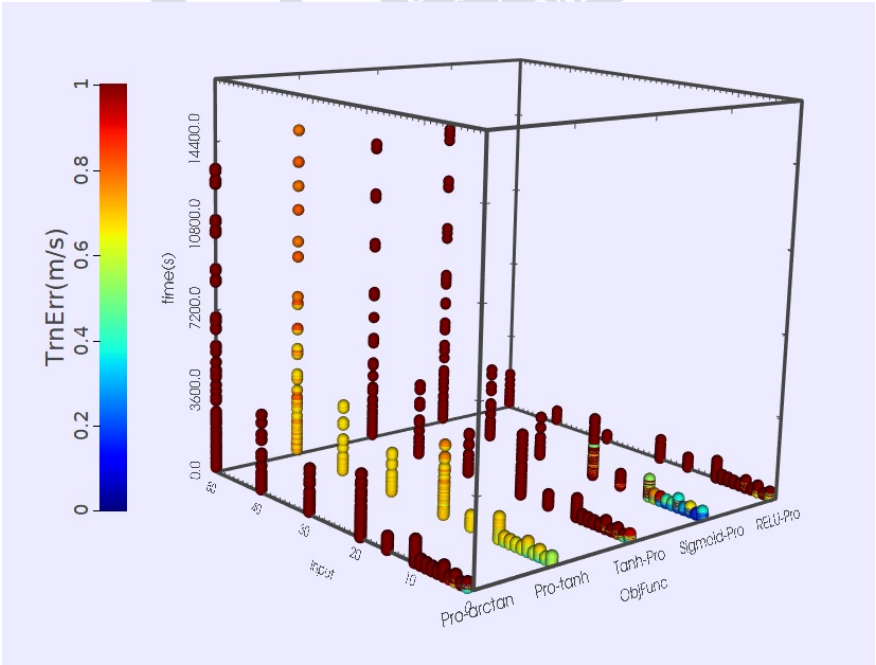


Fig. 6.8 Correlation of different parameters in model, for different activation functions

6.3.1.3 Optimization hyperparameters

To determine the best set of the optimization hyperparameters, model and data hyperparameters are set as follows: Input and hidden neurons are set to 1 and 2, respectively; Activation function is the new modified sigmoid and the modified hyperbolic tangent functions for both hidden and output layers; and data hyperparameters are set based on the initial condition. Following the four different optimization methods included on table 6.5 (O1 to O4), four different hyperparameters solutions are obtained. These results are based on the first phase of grid-search for the optimization hyperparameters solution space.

Four model optimized hyperparameters sets are shown in figure 6.9. Each of the four different optimization methods are colored differently. It can be observed that, in general, BBO and GA are performing better than PSO, and PSO is performing better than ACO. The results are aligned with the success of BBO in other comparative studies of regression and classification problems in other sectors [281, 262]. In figure 6.10 BBO and GA convergence plots (based on MSE) for a different number of parents and generation vs. computing time are shown. As a general fact, fewer parents need more generations and more computational time to reach an acceptable MSE threshold. The number of generations has much more effect on the computing time than the number of parents. Hence, for the BBO the best combination is 200 for parents and 75 for generations. For GA, 200 and 50, respectively. However, as the computing time is not too long, in order to assure minimum MSE higher number of generations will be applied.

In figure 6.11 three different range of values for MLP weights are compared considering a different number of parents in BBO method. Indeed, the weights allocated to the MLP are selected within the range of this parameter. For example, for $value = 5$, the weights can be chosen as any number in [-5, 5]. As it can be seen, $value = 20$ is performing worse than the other two values regardless of parents' number. In general $value = 5$ outperforms other higher values for most of the parents numbers, except for parents number = 100 (figure 6.11), in which the difference in MSE is negligible.

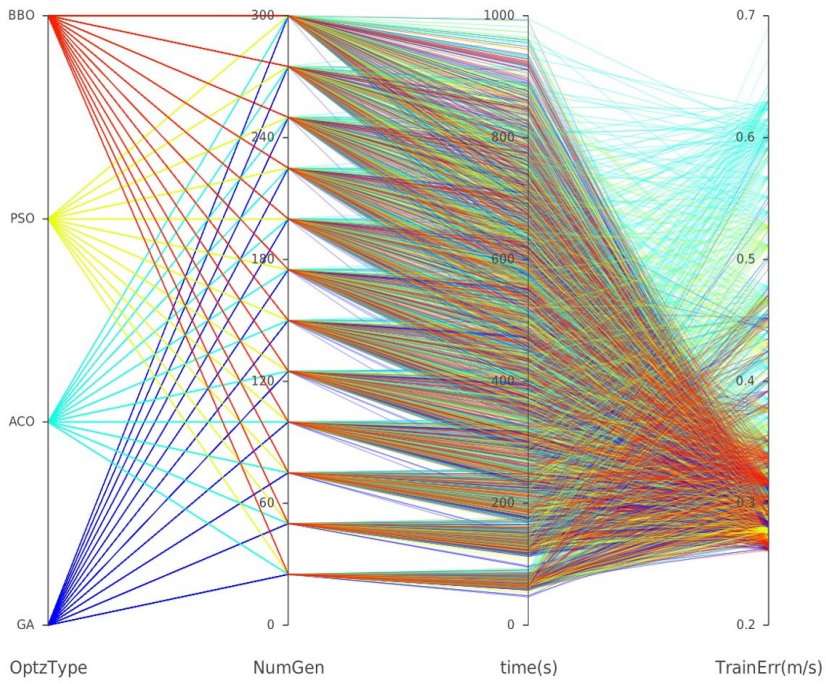
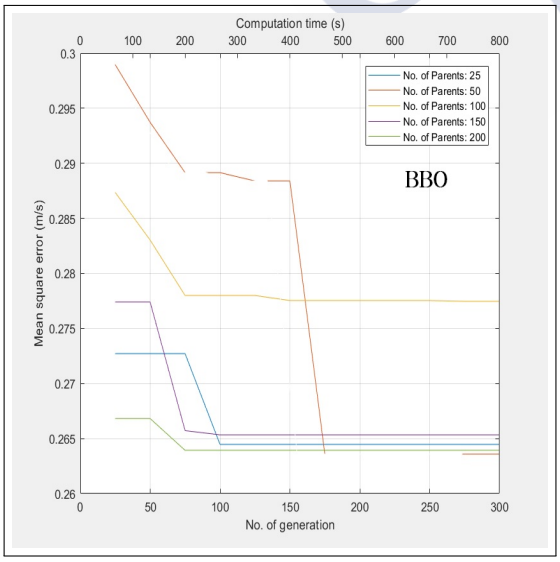
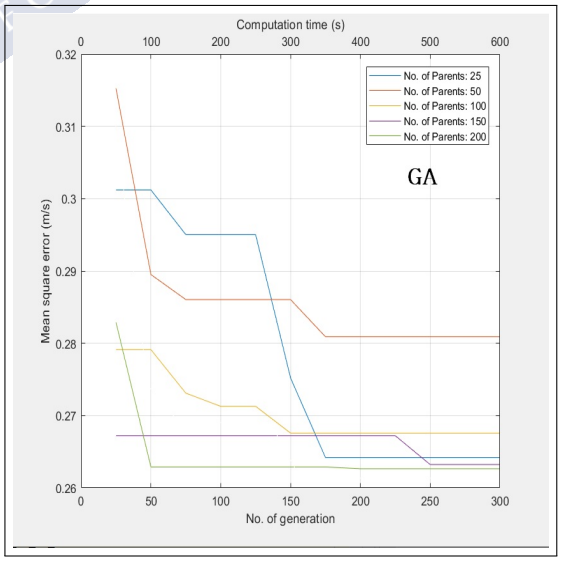


Fig. 6.9 Parallel coordinate plot of optimization hyperparameters using the four proposed methods: GA, ACO, PSO, and BBO

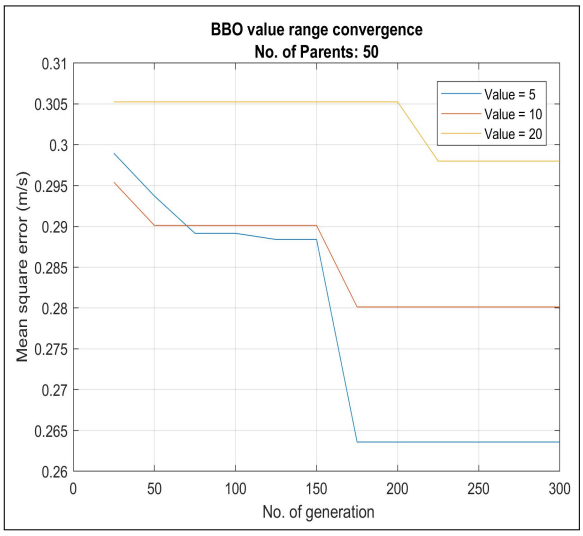


(a) BBO convergence

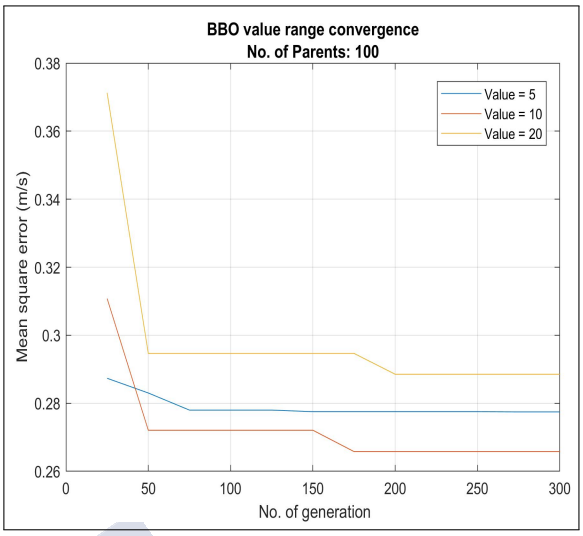


(b) GA convergence

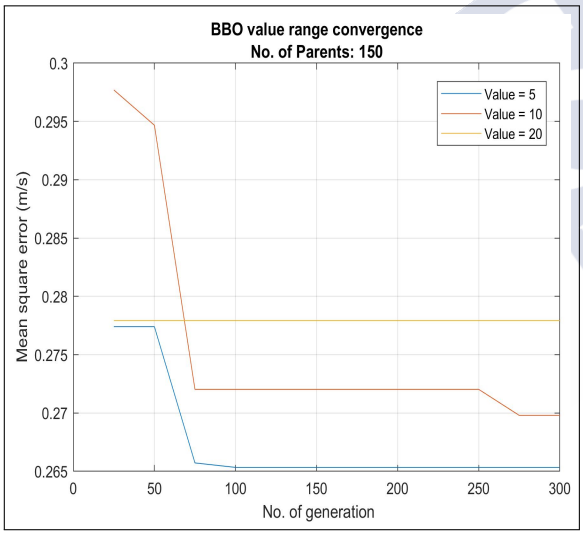
Fig. 6.10 Number of parents vs. number of generations for BBO and GA for convergence rate



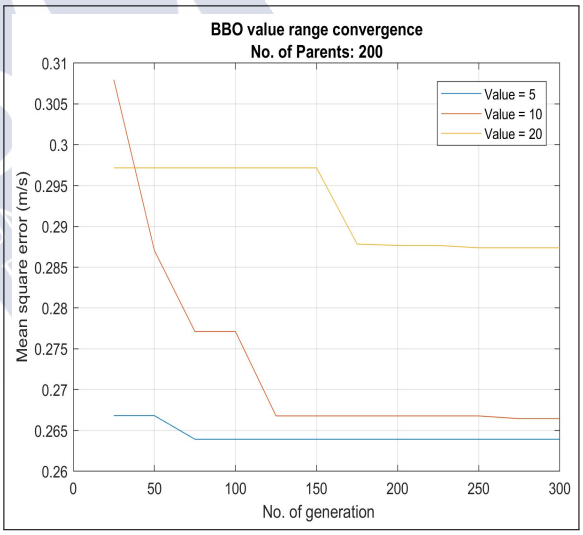
(a) BBO convergence- 50 parents



(b) BBO convergence- 100 parents



(c) BBO convergence 150 parents



(d) BBO convergence 200 parents

Fig. 6.11 Comparing convergence rate of BBO method for different range of values, parents' numbers, and generations' numbers

6.3.1.4 Data hyperparameters

As previously described, data hyperparameter selection is based on the outcomes s of the model and optimization hyperparameter phases. Twelve different feature codes, listed in table 6.3, are considered as different inputs to the model. In figure 6.12 three sets of data hyperparameters are compared, considering validation error and computing time. It can be observed that a few of short length time series are resulting in accurate (low error) results, with the lowest validation errors. However, this also might be related to the over-fitting problem. On the other hand, under-fitting problem is also clearly evident for short length time series, as most of them achieve the highest validation errors. From the different tested lengths, $Length = 1440$ usually performs well, so it is the best election. Also, computing time for this length is reasonable, below 1 minute.

Regarding the input feature code, it is clearly shown that the code 1 in table 6.3, which is a simple previous time step of wind speed, performs much better than the other combinations. However, code 9, a simple moving average of three different heights, (5, 15, and 20 meters), perform outstandingly. Also, features codes 10, 11, and 12 are performing very well. For the (train/validation)-test percentage, almost all the portions perform in a wide range of low to high accuracy depending on the other hyperparameters. It is also highly dependent on the length of time series.

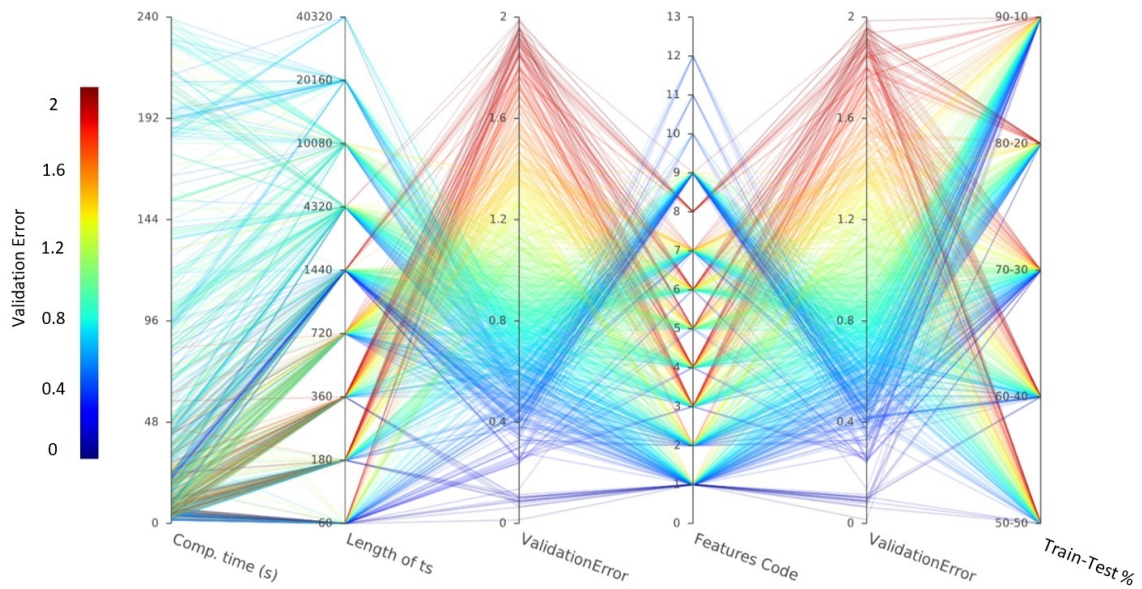
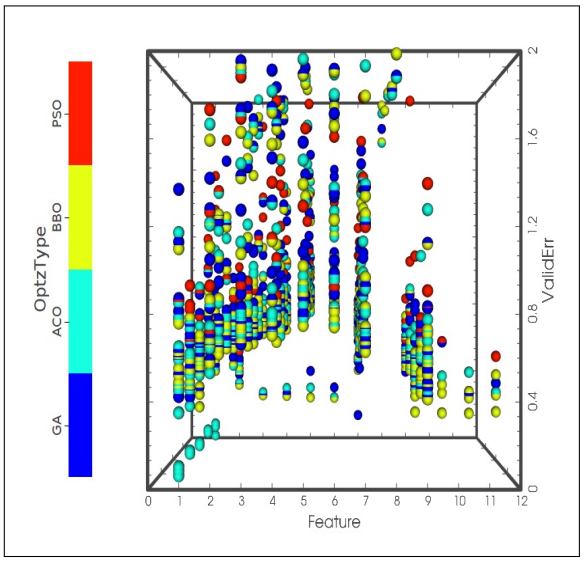
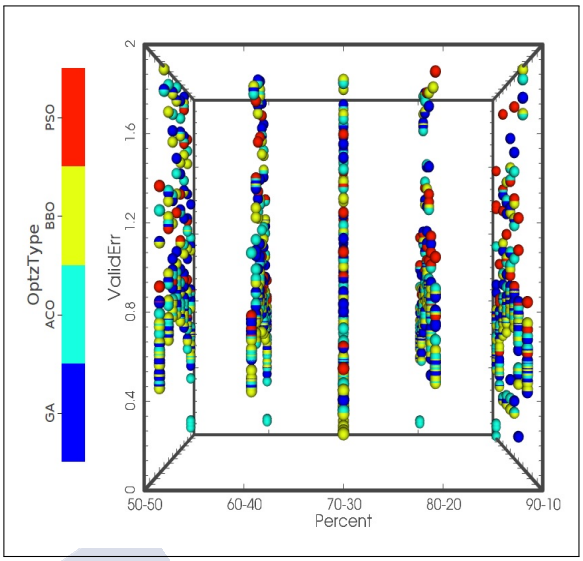


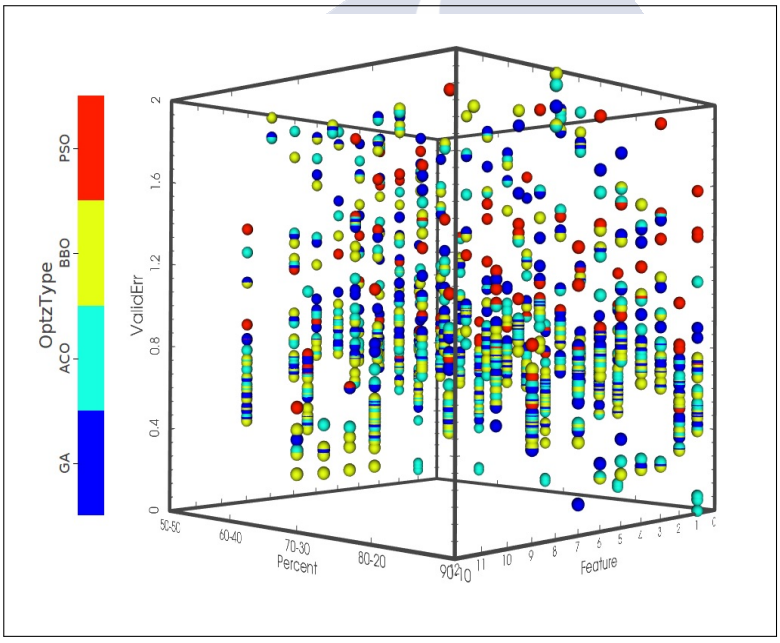
Fig. 6.12 Parallel coordinate plot of data hyperparameters, considering first data step, space search



(a) Feature, optz. method, validation error



(b) Percentage, optz. method, validation error



(c) 4-D visualization of three hyperparameters vs. validation error

Fig. 6.13 Four dimensional visualization of validation error using different data hyperparameters (feature, percentage of train/validation test) vs. optimization method: GA, ACO, BBO, and PSO

Figure 6.13 shows three different views of 4-D visualization of data and optimization changing input feature, percentage of (train/validation)-test, validation error, and optimization method. It can be seen in figure 6.13a that the code 1 (table 6.3), the previous time step of wind speed, provides the minimum validation error for ACO, although considering

other factors as computing time and selection model, ACO might not be the best selected optimization method. For a validation error around 0.4, BBO is dominant, followed by GA and ACO. In general, feature codes 1, 2, and 9 provide the most accurate results for any optimization method. In figure 6.13b, it is clear that 70-30 percentage for (train/validation) and testing performs the minimum validation error for all types of optimization methods, with BBO overcoming the others. Figure 6.13c shows that PSO is performing much worse than the other methods for different values of feature and percentage. And BBO is performing slightly better than the other two methods.

Figure 6.14 shows , using several data lenghts, results of validation error and required computing time. For lengths more than 1440, the computing time increases considerably. This time is an average for all percentages and feature codes. The rest of the hyperparameters are fixed based on the best results of the previous search spaces. Except for the very short lengths of 60, 180, and 360, which might cause an under-fitting problem as only a few of runs produce low validation errors, data length = 1440 is confirmed to perform better than the lower and higher number of time series length.

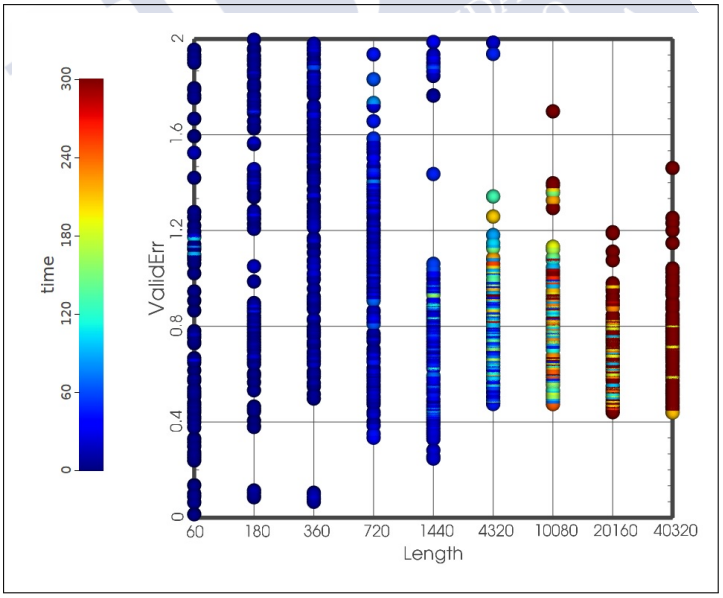


Fig. 6.14 Three dimensional visualization of time, data length and validation error

6.3.2 Selected model

After each subspace search is completed, the model selection phase starts as a two-stage process. In the first phase, the best 10% models in MSE will be chosen for the second phase, in which a 10-fold cross-validation method is applied to each single selected model. The models based on their validation error go through AIC and BIC calculation, and the hyperparameters of the best five sorted models based on the Pareto of AIC, BIC, and validation error, go for the next sub-space grid-search. The process should continue for at least one round of each sub-space to be searched until the validation error (MSE) difference will be equal to or less than 0.01.

The following results explain the outcome after the third round and final round of subspace searches, which are shown in figures 6.15 and 6.16. The ranked results of the final 10-fold cross-validation for the first forty selected models are presented in table 6.7.

Considering both figures 6.15 and 6.16, AIC and BIC are not supporting each other, and in some cases, one provides the best result while the worst result for another one. This possibility was previously reported in [285]. The data points close to the axes tops in figure 6.16a correspond to that condition Considering equations 6.13 and 6.14 and the fact that all of the nominated models in subspaces of hyperparameter solution space have the same model parameters, the only parameter which makes a difference is the models error. Taking into account Pareto validation method is applied, it is observed in figure 6.15 that BBO models

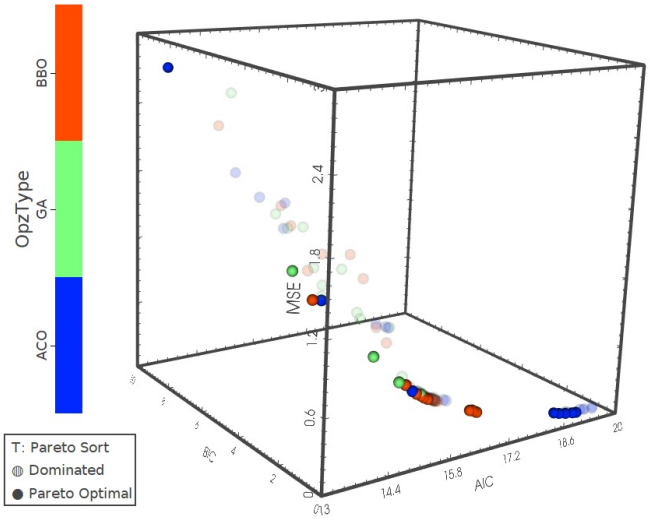
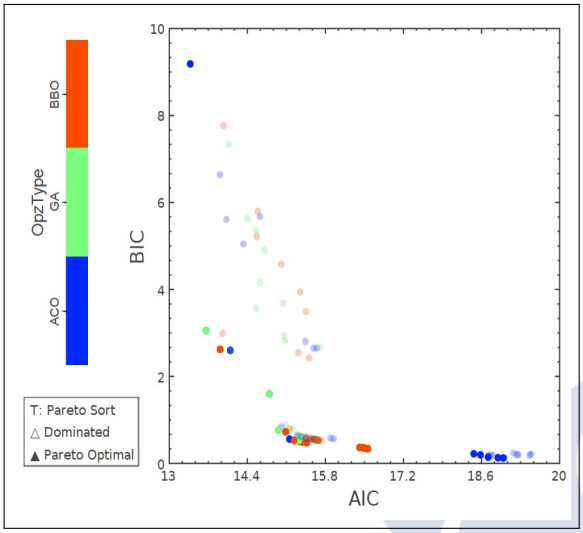
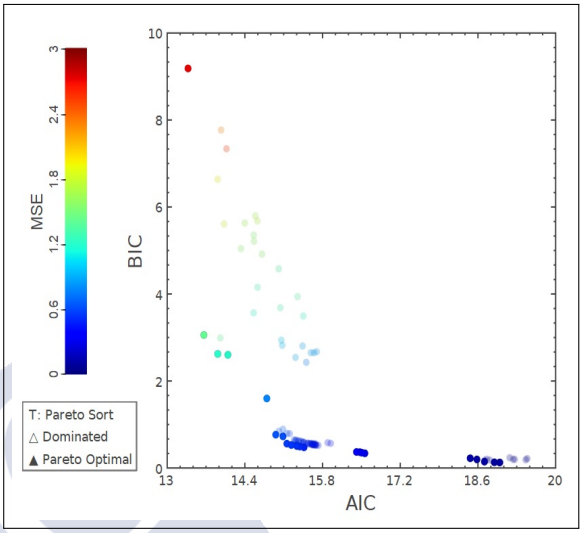


Fig. 6.15 Four-dimensional Pareto validation results for model selection criteria

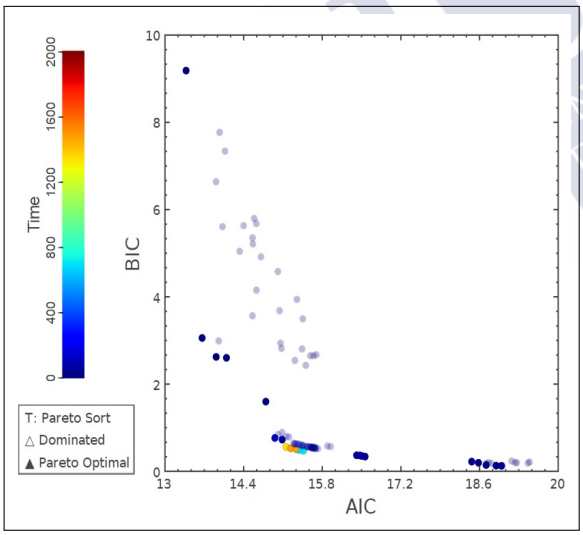
(orange color) perform better, jointly considering AIC, BIC, and MSE validation error. This jointly approach, instead of single MSE results, guarantees to satisfy all criteria and prevent those results that perform worse in one criterion while best in the another one.



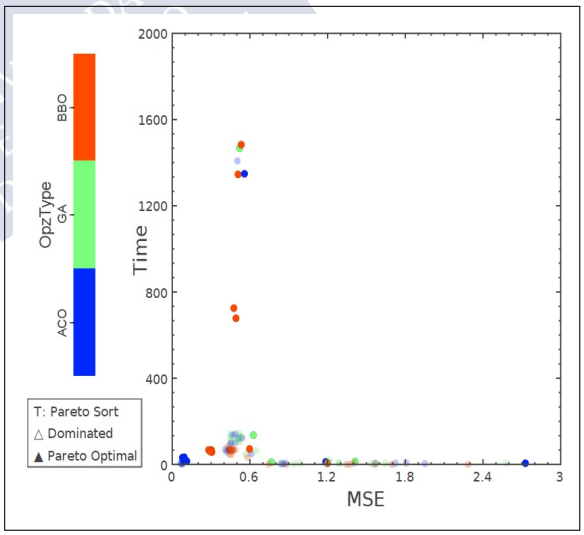
(a) AIC,BIC,Optz. method



(b) AIC,BIC,MSE



(c) AIC,BIC,Time



(d) MSE,Time,Optz. method

Fig. 6.16 Pareto validation results of BIC, AIC, and MSE for model selection

Considering the results in table 6.7, for the model hyperparameters the most straightforward configuration is the best set among all different combinations of hidden and input neurons. It means that for input number equal to 1 and hidden number equal to 2 and the new sigmoid modified activation functions the best results are gained regardless of other

hyperparameters configurations. This result is in agreement with the findings of [127] and [32]. Related to the activation functions, the five nominated functions are chosen from a combination of 144 functions in a pre-selection phase, applying them in a benchmark neural network. Between the functions, it is important to notice that proposed (new sigmoid-modified function table 6.2) performs better than any other combination of commonly used activation functions.

Related to the optimization hyperparameters, all types of optimization methods exist in the nominated solutions, except PSO. ACO performs the minimum MSE, but following Pareto validation criteria also AIC and BIC must be jointly considered, in order to prevent contradictory results. Following this validation criteria, the *BBO* method is dominant in all the first five ranked results. The number of generations, parents, and weights values are 100, 75, and 10, respectively, for the best five nominated solutions. The number of generation and parents for the other types of optimization methods is limited to 150 and 10.

Related to data hyperparameters, it is important to highlight that the time interval of data samples is one minute. In such a data sample frequency, 1440 is dominant for the best results and many of the other listed sets in the table. Also, for the rows with the lowest data length = 60 (table 6.7), the MSE is much beyond the MSE results for length=1440 and even for length=720. This shows that the lower data length produces either under- or over-fitting on the test dataset and not performs well on the validation dataset. About the feature code (table 6.3), proposed simple and weighted averaging of higher and lower altitudes alongside with ten standard meter height performs as good as simple previous wind speed time step (code 1). Regarding (train/validation)-test percentage, for all length=1440, the 70-30 percentage provides the best result.

It must be notice the relevance of Pareto validation method in model selection phase. In some cases, like the last two rows in table 6.7, MSE is very low. However, AIC achieves its maximum value. Also, notice the Richardson number average always correspond to very unstable conditions (table 6.4), as the most fluctuating wind speed conditions, that is, with more difficulties to get better model results.

From the results obtained in the training/validation of the model using the M2 tower case study as sample input, the following conclusions about wind speed prediction model can be established,

Table 6.7 Results for model hyperparameters selection based on Pareto validation criteria (AIC, BIC, and MSE). The best five model settings are shown in the first five results rows.

	Model Hyperparameters			Optimization Hyperpatameters				Data Hyperparameter			10-fold Cross Validation - Model Seletion				
Pareto Rank	Input No.	Hidden No.	Activation Funct	Type	Gen. No.	Parents No.	Value range	Feature code	Length (1-minute)	train-test percentage	Av. Stability index	Comp. Time(s)	MSE	AIC	BIC
1	1	2	Sigmoid-Pro	BBO	100	75	10	10	1440	70-30	-0.55	60.65	0.31	16.42	0.39
2	1	2	Sigmoid-Pro	BBO	100	75	10	11	1440	70-30	-0.55	63.19	0.3	16.52	0.37
3	1	2	Sigmoid-Pro	BBO	100	75	10	9	1440	70-30	-0.07	86.83	0.44	15.74	0.54
4	1	2	Sigmoid-Pro	BBO	100	75	10	1	1440	70-30	-0.07	66.27	0.47	15.59	0.58
5	1	2	Sigmoid-Pro	BBO	100	75	10	9	720	70-30	0.04	36.43	0.59	15.17	0.81
	1	2	Sigmoid-Pro	GA	150	100	20	9	180	90-10	-0.67	14.53	0.77	14.81	1.62
	1	2	Sigmoid-Pro	GA	150	100	20	1	1440	60-40	-0.07	146.64	0.52	15.32	0.64
	1	2	Sigmoid-Pro	GA	150	100	20	2	1440	90-10	-0.07	139.38	0.63	14.97	0.78
	1	2	Sigmoid-Pro	GA	150	100	20	1	720	70-30	0.04	67.17	0.65	15.1	0.91
	1	2	Sigmoid-Pro	GA	150	100	20	7	60	90-10	-0.75	8.57	1.58	14.72	4.93
	1	2	Sigmoid-Pro	BBO	100	75	10	6	60	90-10	-0.75	3.65	2.29	13.98	7.79
	1	2	Sigmoid-Pro	BBO	100	75	10	5	60	90-10	-0.75	3.66	1.57	14.58	5.23
	1	2	Sigmoid-Pro	ACO	150	100	10	6	60	90-10	-0.75	7.79	1.96	13.92	6.65
	1	2	Sigmoid-Pro	BBO	100	75	10	4	60	90-10	-0.75	3.57	1.38	15.36	3.96
	1	2	Sigmoid-Pro	BBO	100	75	10	3	60	90-10	-0.75	3.68	1.7	14.6	5.81
	1	2	Sigmoid-Pro	BBO	100	75	10	9	1440	70-30	-0.55	67.91	0.31	16.49	0.38
	1	2	Sigmoid-Pro	BBO	100	75	10	12	1440	70-30	-0.55	70.42	0.29	16.57	0.35
	1	2	Sigmoid-Pro	GA	150	100	20	3	60	90-10	-0.75	9.44	1.65	14.41	5.65
	1	2	Sigmoid-Pro	GA	150	100	20	4	60	90-10	-0.75	9.43	1.29	15.05	3.7
	1	2	Sigmoid-Pro	ACO	150	100	10	3	60	90-10	-0.75	10.41	1.81	14.04	5.63
	1	2	Sigmoid-Pro	BBO	100	75	10	1	1440	70-30	-0.55	70.36	0.31	16.48	0.38
	1	2	Sigmoid-Pro	ACO	150	100	10	4	60	90-10	-0.75	9.19	1.73	14.64	5.7
	1	2	Sigmoid-Pro	ACO	150	100	10	5	60	90-10	-0.75	9.15	2.73	13.39	9.2
	1	2	Sigmoid-Pro	GA	150	100	20	6	60	90-10	-0.75	9.83	1.29	14.64	4.17
	1	2	Sigmoid-Pro	BBO	100	75	10	2	60	90-10	-0.75	4.44	1.35	15.02	4.6
	1	2	Sigmoid-Pro	GA	150	100	20	2	60	90-10	-0.75	9.54	1.2	14.57	3.58
	1	2	Sigmoid-Pro	ACO	150	100	10	2	60	90-10	-0.75	9.29	1.57	14.34	5.06
	1	2	Sigmoid-Pro	ACO	150	100	10	1	720	60-40	-0.61	74.05	0.42	15.96	0.58
	1	2	Sigmoid-Pro	BBO	100	75	10	9	1440	70-30	-0.44	71.55	0.47	15.54	0.58
	1	2	Sigmoid-Pro	BBO	100	75	10	12	1440	70-30	-0.44	71.36	0.45	15.63	0.56
	1	2	Sigmoid-Pro	BBO	100	75	10	1	1440	70-30	-0.44	71.74	0.49	15.47	0.61
	1	2	Sigmoid-Pro	GA	150	100	20	11	1440	70-30	-0.07	128.33	0.45	15.65	0.56
	1	2	Sigmoid-Pro	ACO	150	100	10	1	720	50-50	-0.61	62.38	0.42	15.9	0.6
	1	2	Sigmoid-Pro	BBO	100	75	10	10	1440	70-30	-0.44	64.24	0.46	15.63	0.57
	1	2	Sigmoid-Pro	ACO	150	100	10	1	60	90-10	-0.59	6.37	0.08	19.25	0.22
	1	2	Sigmoid-Pro	ACO	150	100	10	1	60	50-50	-0.59	7.5	0.08	19.18	0.25

- Simplest configuration of MLP (1,2,1) performs better than any other complex one.
- New activation function (table 6.2), which are modified versions of basic activation functions (tanh and sigmoid), outperform both basic and small modified functions.
- BBO optimization model performs better than GA, ACO, and PSO for MLP weights.
- The optimum numbers of parents and generations in the optimization method can be relatively small, with good results and lower computing time.
- TDataset length for both training and validation plays a vital role in the accuracy of results, with probably dependence on its sampling frequency. For minute-based (M2-tower case study), 1440 samples (24 hours, a complete daily cycle), a complete day, provides the best results. However, the best dataset length could be different using other sampling frequencies, but this parameter was discussed in very few studies. The training period of data is discussed in very few studies. As an example, Hu et al. [292] tested different dataset lengths for training, from two weeks to six months, for a deep neural network. The longer lengths provided more accurate results than shorter ones. However, the new developed model uses a shallow NN and, as it was shown in table 6.7, longer training datasets are not the best option.
- Pareto validation method for model selection rather than individual criteria methods (i.e., MSE only) plays an essential role in preventing contradictory results.

Table 6.8 summarizes the new developed short term wind speed prediction model.

Table 6.8 New developed wind speed prediction model: Selected hyperparameters from semi exhaustive search optimization

H.parameter	Description	Values
M_1	No. of input neurons	1
M_2	No. of hidden neurons	2
M_3	activation function	Sigmoid-proposed function
O_1	optimization method	*BBO*
O_2	No. of parents	75
O_3	No. of generation	100
O_4	weights value range	[-10 10]
D_1	No. of time steps	1440 (minute-based)
D_2	(train/validation)/test% percentage	70%-30%
D_3	Input features	Code:1 (WS_{t-1})

6.3.3 Wind speed prediction model testing: Application to cases studies

In this section, the best hyperparameters set described in table 6.8 are used for training and testing the new developed short-term wind speed forecasting model. Four meteorological stations are considered as case studies: Santiago-EOAS, Camariñas, and Muralla in Galicia, Spain; and M2-tower in Colorado, U.S.. As previous model selection process was based on short term ahead periods, three different time ahead prediction periods of 10-min, 1-hour, and 3-hours are considered for the M2-tower seasonal based data, and 1-hour, 2-hours, and 3-hours ahead predictions are used for the Galician case studies; taking into account that M2-tower provides 1-min. average data, while at Galician stations 10-min. average data are available.

For M2-tower, one year of 1-minute data during 2018 at three different heights of 5, 10, and 20 meters were fed into the model. At Galician stations only 10-m standard height data are available; Spring 2017 time series were applied. Also, at these Galician stations WRF numerical weather prediction data were applied both as a benchmark model results (for comparison) and as input data to the new developed model. That is, results from the new developed model will be compared to the direct WRF NWP, in order to evaluate any achieved improvements.

In addition, in the last section a comparison between the four optimization methods previously tested (BBO, GA, PSO, and ACO) is also developed using Galician stations datasets and three different time steps, considering several criteria.

6.3.3.1 M2 Tower, USA

A one-year data comprised of various parameters at six different height levels are gathered during 2018, as described in section 3.2.5. Model testing statistics for both one year and for seasonal periods were computed, in order to better understanding the model accuracy.

Tables 6.9 to 6.11 show the results from the new developed model and other models: persistence (PM), as a benchmark method [288], ARIMA (1,1,1) [293], and MLP BR, as the best model described in section 5.3.2.1. Prediction ahead time periods are: 10 minutes, 1 hour, and 3 hours. Statistics considered are: MSE, RMSE, MAE, and MAPE. The best results are highlighted in bold text.

Table 6.9 Very short-term prediction: 10-minutes ahead wind speed predictions statistics for M2-tower at 10m height

Season	**Models**	**Criteria**			
		MSE$(m/s)^2$	**RMSE(m/s)**	**MAE (m/s)**	**MAPE%**
Spring	BBO-MLP	0.4917	0.7012	0.5208	26.86
	PM	0.4942	0.7030	0.5053	23.43
	ARIMA	4.7228	2.1732	1.7243	63.96
	MLP-BR	0.4788	0.6919	0.5023	25.07
Summer	BBO-MLP	0.6588	0.8116	0.5935	26.18
	PM	0.6360	0.7976	0.5703	24.45
	ARIMA	6.6144	2.5718	2.0655	89.14
	MLP-BR	0.5973	0.7729	0.5503	24.89
Fall	BBO-MLP	0.8665	0.9308	0.6142	28.88
	PM	0.7020	0.8379	0.5589	27.06
	ARIMA	7.3700	2.7147	3.4457	42.51
	MLP-BR	0.6519	0.8073	0.5534	28.99
Winter	BBO-MLP	1.4775	1.2155	0.7722	25.69
	PM	1.4270	1.1946	0.7002	17.03
	ARIMA	7.6630	2.7682	5.1372	45.64
	MLP-BR	1.4071	1.1862	0.7358	23.43
Average	BBO-MLP	0.8736	0.9148	0.6252	26.90
	PM	0.8148	0.8833	0.5867	**22.99**
	ARIMA	6.5918	2.56751	3.0932	60.31
	MLP-BR	**0.7837**	**0.8645**	**0.5855**	25.59

As it is shown above in table 6.9, considering annual average the new developed model (BBO-MLP) is not performing as good as backpropagation MLP-BR method described in section 5.3.2.1. Also, the new model even has lower annual average accuracy than persistence model [95]. However, this result is in agreement with Wegley et al.[294], where the persistence model performs better than other methods as generalized equivalent Markov model and autoregressive model . In addition, new model accuracy is much lower in winter,

which shows the highly fluctuating characteristics of wind speed during that season at M2-Tower.

Table 6.10 Short-term prediction: 1-hour ahead wind speed predictions statistics for M2-tower at 10m height

Season´	**Models**	**Criteria**			
		MSE$(m/s)^2$	**RMSE(m/s)**	**MAE (m/s)**	**MAPE%**
Spring	BBO-MLP	0.7020	0.8378	0.6618	24.84
	PM	0.9899	0.9949	0.7205	26.04
	ARIMA	2.4379	1.5614	1.3067	53.09
	MLP-BR	0.7381	0.8594	0.6570	27.05
Summer	BBO-MLP	0.8357	0.9141	0.6260	24.41
	PM	0.9825	0.9912	0.7045	27.49
	ARIMA	4.5295	2.1283	1.7578	81.38
	MLP-BR	0.8469	0.9203	0.6486	26.49
Fall	BBO-MLP	0.8089	0.8994	0.7125	25.64
	PM	0.9431	0.9711	0.7522	28.72
	ARIMA	3.9858	1.9964	2.4822	73.18
	MLP-BR	0.8549	0.9837	0.7462	26.15
Winter	BBO-MLP	1.7566	1.3254	0.9353	34.23
	PM	2.5574	1.5991	1.0325	38.23
	ARIMA	4.2050	2.0506	1.5208	77.88
	MLP-BR	1.9815	1.4077	1.0541	37.32
Average	**BBO-MLP**	**0.9884**	**0.9941**	**0.7339**	**27.28**
	PM	1.3682	1.1391	0.8024	30.12
	ARIMA	3.7895	1.9341	1.7669	71.38
	MLP-BR	1.1053	1.0427	0.7765	29.25

Table 6.10 compares the results of the new developed model against PM, ARIMA(1,1,1), and MLP-BR model. On the contrary to very short term prediciton results, the new model is

quite successful compared to the other models, for all seasons and in all evaluation criteria. MSE, RMSE, MAE, and MAPE of the proposed model for annual average are 0.9884, 0.9941, 0.7339, and 27.28, respectively, for the next step in hourly-averaged wind speed (m/s). New model statistics are even much better (lower) during Spring, Summer, and Fall, while higher in Winter. As expected, persistence model rapidly losses accuracy when ahead prediction interval increases, proving the necessity of other modeling approaches. As a literature reference, although sample datasets for testing are different, the new model outperform the results reported by Li and Shi [295] that developed and applied three ANN models (adaptive linear element, backpropagation, and radial basis function) on two sites in Chine for 1-hour ahead wind speed prediction. For example, the best results of MAE and RMSE of Kumn site (China) were reported as 1.112, 1.444, respectively.

Table 6.11 shows the prediction test statistics for one-step-ahead prediction in 3-hours time interval. Again, the new model outperforms other models, even with higher accuracy than its own 1-hour ahead prediction. The MSE, RMSE, MAE, and MAPE are 1.562, 1.0749, 0.8950, and 34.01, respectively. PM model statistics are much higher than 1-hour due to the time interval increment, as expected. Also, it can be notice that increasing the interval time also decreases the time series fluctuations, as time interval is not only applied to the prediction time ahead, but also to the average input data. As a consequence, using 3-hours as time interval also increases Winter prediction accuracy, which now is similar to the new model accuracy during other seasons.

In order to better comparing the new model results and other models results, table 6.12 shows the improvement percentage obtained with the new model compared to the other models in the different statistics criteria, for 1-hour and 3-hours ahead time intervals. As shown, the proposed model improves the PM results of MSE, RMSE, MAE, and MAPE, by 28.75%, 12.73%, 8.50%, 9.43%, for 1-hour time interval; and by 47.50%, 31.86%, 18.90%, and 25.55% for 3-hour time interval, respectively.

In order to show the wind speed variability at T2-tower, figure 6.17 shows an example of next step wind speed along Spring 0218 for a 10-minutes time interval. Asterisks on the red line are the predicted values for the first week of Spring 2018. Figure 6.18 shows the improved percentage of MSE, RMSE, MAE, and MAPE for the proposed method against the PM. By increasing the time interval, the improved percentage increase. The best result is for MSE with 47.5% of improvement for a 3-hour time interval which is significant.

Table 6.11 3-hours ahead wind speed predictions statistics for M2-tower at 10m height

Season	**Models**	**Criteria**			
		MSE$(m/s)^2$	**RMSE(m/s)**	**MAE (m/s)**	**MAPE%**
Spring	BBO-MLP	1.1025	1.0500	0.8151	31.11
	PM	1.6460	1.2830	0.9596	37.86
	ARIMA	2.0734	1.4366	1.1956	44.11
	MLP-BR	1.3463	1.1603	0.9087	41.16
Summer	BBO-MLP	1.2557	1.1206	1.1058	38.98
	PM	2.6598	1.6309	1.2657	42.88
	ARIMA	2.9133	1.7068	1.3548	64.51
	MLP-BR	2.0691	1.4384	1.2166	54.74
Fall	BBO-MLP	1.1892	1.0905	0.8940	37.94
	PM	1.5574	1.2479	0.9609	50.45
	ARIMA	2.4492	1.5650	1.2816	75.57
	MLP-BR	1.3474	1.1608	0.9464	54.02
Winter	BBO-MLP	1.0783	1.0384	0.7649	28.02
	PM	3.3985	1.8435	1.2286	51.52
	ARIMA	2.6192	1.6184	1.3824	52.33
	MLP-BR	1.6273	1.2757	0.8969	36.34
Average	BBO-MLP	**1.1564**	**1.0749**	**0.8950**	**34.01**
	PM	2.2029	1.5776	1.1037	45.68
	ARIMA	2.5138	1.5817	1.3036	59.38
	MLP-BR	1.5975	1.2588	0.9919	46.57

Table 6.12 Percentage improvement of MSE, RMSE, MAE, and MAPE obtained with the new developed model, respect to the three benchmark models, for one and three hour ahead, M-2 tower, 2018

Step	**Criteria**	**Model**		
		Persistence model	**ARIMA (1,1,1)**	**MLP-BR**
1-hour	MSE	28.75 %	73.91 %	10.58 %
	RMSE	12.73 %	48.60 %	4.66 %
	MAE	8.5 0%	58.46 %	5.49 %
	MAPE	9.43 %	61.78 %	6.74 %
3-hour	MSE	47.50 %	53.99 %	27.62 %
	RMSE	31.86 %	32.04 %	14.61 %
	MAE	18.90 %	31.34 %	10.77 %
	MAPE	25.55 %	42.72 %	26.97 %

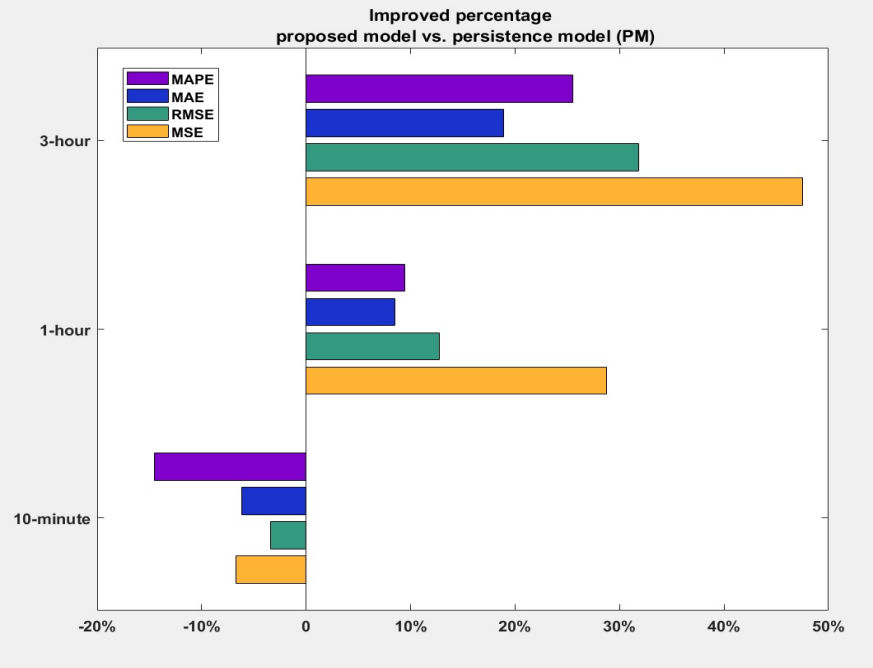


Fig. 6.18 Improvement percentage of the new developed model against persistence model for different statistic criteria and the three tested different time intervals,M2-tower

6.3.3.2 Galician meteorological stations, Spain

The developed model with the best set of hyperparameters is applied on Santiago-EOAS, Camariñas, and Muralla weather stations to predict 1-hour, 2-hours, and 3-hours time interval wind speed. The results are compared to the following models results: models, including the WRF model. The following five models are used to compare their results in terms of MSE, RMSE, MAE, and MAPE with the developed model: persistence model (PM), ARIMA(1,1,1), MLP-BR, WRF operational regional numerical weather forecast [REF], and also the developed model with WRF forecast as additional input (BBO-MLP-WRF). Some models use previous time steps of observational (Obs.) data as common input, and others use WRF data, both previous and future time steps, following table 6.8. As models comparison criteria, MSE, RMSE, MAE, and MAPE are applied.

In tables 6.13, 6.15, and 6.17 the different prediction models statistics results at the three stations are presented. Also, the corresponding improvement percentages in statistics obtained with the new developed model respect to the other models are shown in tables 6.14, 6.16, and 6.18. Figure 6.19 shows the same improvement percentages for one, two, and three hour ahead prediction at Santiago-EOAS station, considering all evaluation criteria. Finally, in figure 6.20 MSE improvements at Muralla and Camariñas stations are shown.

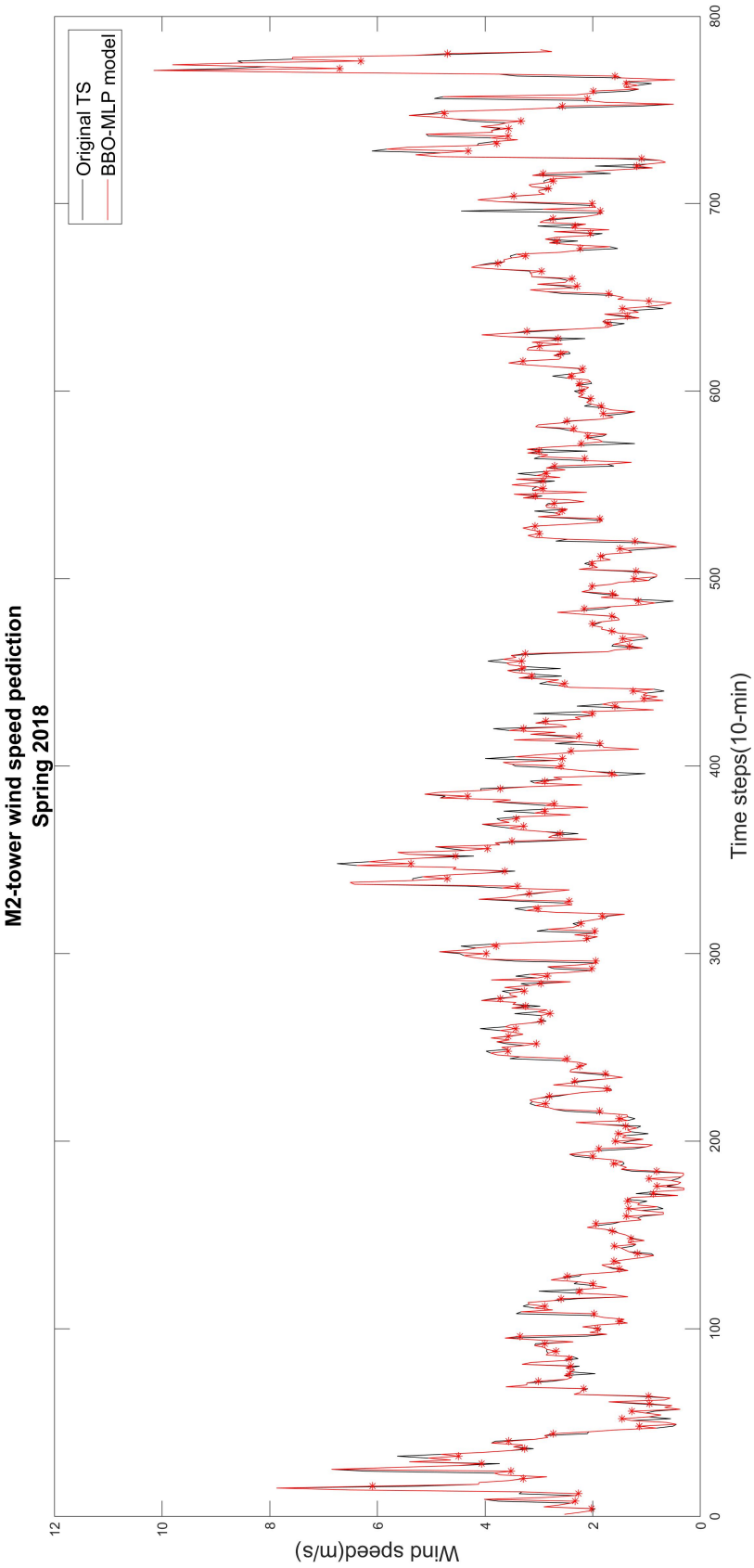


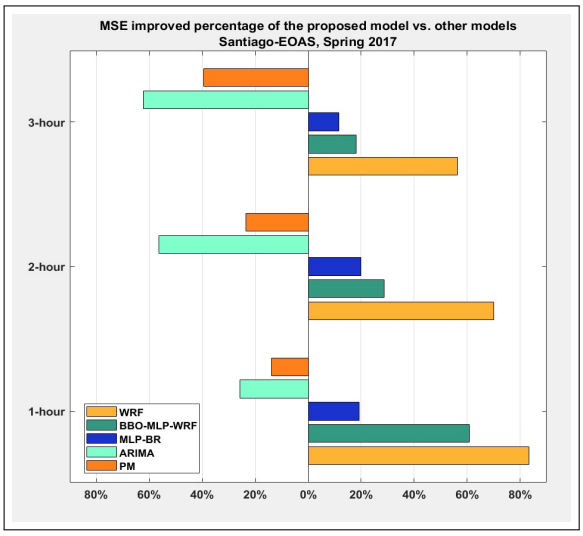
Fig. 6.17 M2-tower 10-minutes ahead wind speed prediction, Spring 2018

Table 6.13 One to three-hour ahead wind speed forecasting statistics, Santiago-EOAS, Spring 2017. Best results in bold text

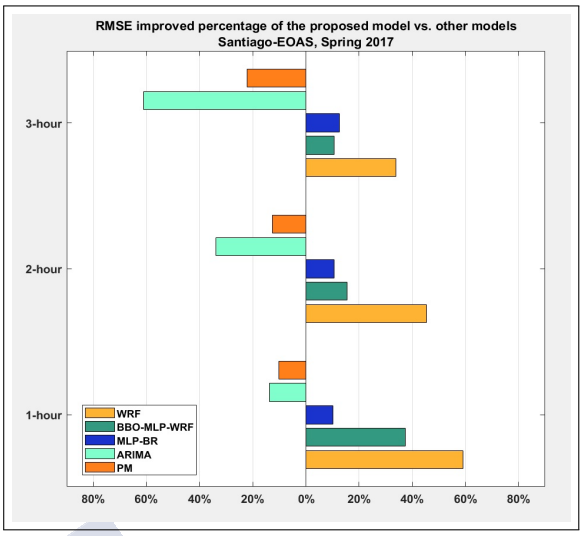
Step	Model	MSE$(m/s)^2$	RMSE(m/s)	MAE(m/s)	MAPE(%)
1-hour ahead	BBO-MLP (Obs)	**0.7706**	**0.8778**	**0.6681**	**15.32**
	BBO-MLP (WRF)	1.9713	1.4040	1.1504	21.17
	PM	0.8956	0.9463	0.7891	18.28
	WRF	4.6052	2.1460	1.7554	27.12
	ARIMA	1.0388	1.0192	0.7968	18.27
	MLP-BR	0.9544	0.9769	0.7899	17.89
2-hour ahead	BBO-MLP (Obs)	**1.1333**	**1.0646**	**0.8106**	**17.75**
	BBO-MLP (WRF)	1.5894	1.2607	1.0186	20.14
	PM	1.4818	1.2173	1.0014	21.58
	WRF	3.7934	1.9477	1.7859	28.15
	ARIMA	2.5970	1.6115	1.1879	25.26
	MLP-BR	1.4160	1.1899	0.9969	23.45
3-hour ahead	BBO-MLP (Obs)	**1.5037**	**1.2262**	**1.0352**	**22.14**
	BBO-MLP (WRF)	1.8383	1.3558	1.2015	28.72
	PM	1.8752	1.3693	1.3914	26.53
	WRF	3.4404	1.8548	1.5446	27.45
	ARIMA	3.8574	1.9640	1.6420	28.77
	MLP-BR	1.6998	1.4038	1.2559	25.75

Table 6.14 Improvement percentage (%) of different criteria, comparing the developed model with to other models, for one, two, and three hours ahead, Santiago-EOAS, Spring 2017

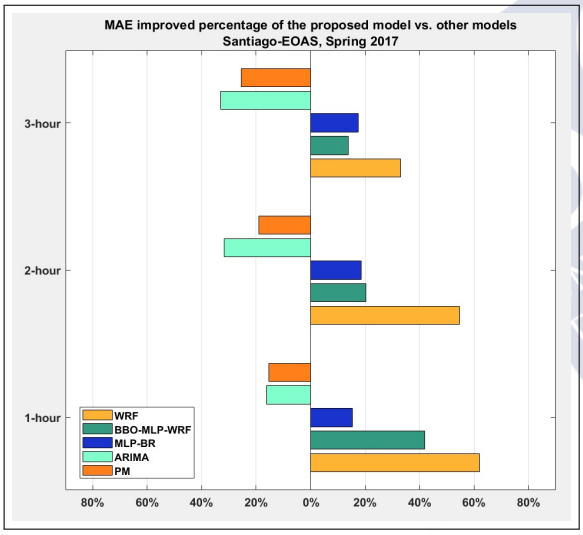
Step	Criteria	BBO-MLP(WRF)	Persistent	WRF model	ARIMA	MLP-BR
1-hour	MSE	60.90 %	13.95 %	83.27 %	25.82 %	19.25 %
	RMSE	37.48 %	07.24 %	59.09 %	13.78 %	10.14 %
	MAE	41.92 %	15.33 %	61.94 %	16.15 %	15.41 %
	MAEP	27.63 %	16.19 %	43.50 %	16.14 %	14.36%
2-hour	MSE	28.69 %	23.52 %	70.12 %	56.36 %	20.02 %
	RMSE	15.56 %	12.54 %	45.34 %	33.94 %	10.53 %
	MAE	20.42 %	19.05 %	54.61 %	31.76 %	18.68 %
	MAPE	11.86 %	17.75 %	36.94 %	29.73 %	24.31 %
3-hour	MSE	18.20 %	39.56 %	56.29 %	62.16 %	11.53 %
	RMSE	10.56 %	22.25 %	33.89 %	61.02 %	12.65 %
	MAE	13.84 %	25.60 %	32.98 %	36.95 %	17.57 %
	MAPE	13.92 %	20.54 %	23.04 %	33.93 %	14.02 %



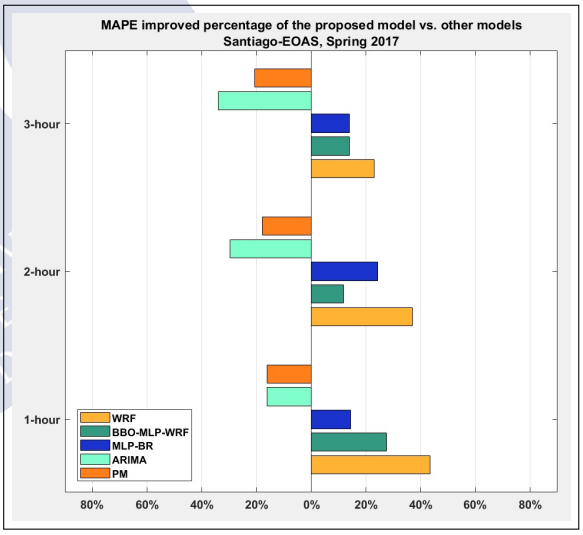
(a) IP(%) MSE



(b) IP(%) RMSE



(c) IP(%) MAE



(d) IP(%) MAPE

Fig. 6.19 Improvement percentage of different criteria obtained with the developed model against other models at Santiago-EOAS for Spring 2017

As it can be seen in figure 6.19, all evaluation criteria follow the same pattern, although the overall improvement percentage (IP) is different. PM and ARIMA models are grouped to the left since by increasing the time interval, also IP increase. While WRF and BBO-MLP-WRF, and MLP-BR are grouped to the right side. IP of the first two models decreases by moving from 1 to 3 hour time interval. However, there is not a unique pattern for the last MLP-BR. Finally, it is important to highlight that, for 3-hours time interval, the new developed model improves MSE respect to WRF NWP and PM in 56.29% and 40%, respectively.

Table 6.15 One to three hours ahead wind speed forecasting statistics, Muralla, Spring 2017. Best results in bold text

Step	Model	MSE$(m/s)^2$	RMSE(m/s)	MAE(m/s)	MAPE(%)
1-hour ahead	BBO-MLP (Obs)	**0.7499**	**0.8659**	**0.5870**	**10.33**
	BBO-MLP (WRF)	4.0030	2.0008	1.5373	34.29
	PM	0.7642	0.8742	0.6342	13.89
	WRF	4.3856	2.0942	1.6656	48.58
	ARIMA	0.7844	0.8857	0.6256	19.57
	MLP-BR	0.7514	0.8668	0.6083	13.12
2-hour ahead	BBO-MLP (Obs)	**1.0956**	**1.0467**	**0.8295**	**19.01**
	BBO-MLP (WRF)	4.6642	2.1597	1.8045	35.56
	PM	1.2834	1.1329	0.9349	22.07
	WRF	5.2038	2.2812	1.9580	36.45
	ARIMA	2.1506	1.4665	1.0842	31.43
	MLP-BR	1.3391	1.1572	1.0071	29.52
3-hour ahead	BBO-MLP (Obs)	**1.3137**	**1.1462**	**0.8609**	**23.69**
	BBO-MLP (WRF)	4.6328	2.1523	2.0748	37.79
	PM	1.6487	1.2840	1.0865	29.92
	WRF	5.2589	2.2932	2.0389	39.26
	ARIMA	3.7994	1.9492	1.6766	36.84
	MLP-BR	1.5708	1.2533	1.0355	28.21

Table 6.16 Improvement percentage (%) of different criteria, comparing the developed model to the other models, for one, two, and three hours ahead, Muralla, Spring 2017

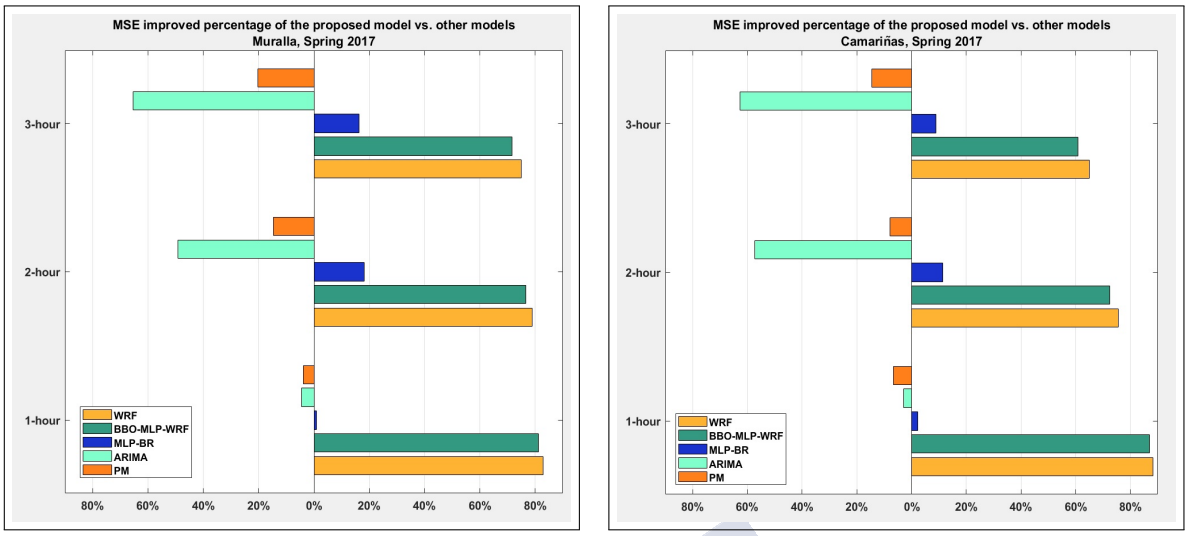
Step	Criteria	BBO-MLP(WRF)	Persistent	WRF model	ARIMA	MLP-BR
1-hour	MSE	81.26 %	03.80 %	82.90 %	04.40 %	01.00 %
	RMSE	56.71 %	01.00 %	58.64 %	02.23 %	01.00 %
	MAE	61.81 %	07.43 %	64.75 %	06.17 %	03.50 %
	MAEP	69.78 %	25.69 %	78.73 %	47.25 %	02.13 %
2-hour	MSE	76.51 %	14.63 %	78.94 %	49.06 %	18.18 %
	RMSE	51.53 %	07.06 %	54.12 %	28.63 %	09.82 %
	MAE	54.03 %	11.27 %	57.63 %	23.49 %	17.63 %
	MAPE	46.45 %	15.58 %	47.84 %	39.51 %	35.60 %
3-hour	MSE	71.64 %	20.32 %	75.02 %	65.42 %	16.37 %
	RMSE	46.75 %	10.73 %	50.05 %	41.19 %	8.55 %
	MAE	63.77 %	20.76 %	57.77 %	48.64 %	16.85 %
	MAPE	37.31 %	20.82 %	39.66 %	40.63 %	16.02 %

Table 6.17 One to three-hours ahead wind speed forecasting statistics, Camariñas, Spring 2017. Best results in bold text

Step	Model	MSE$(m/s)^2$	RMSE(m/s)	MAE(m/s)	MAPE(%)
1-hour ahead	BBO-MLP (Obs)	**0.5561**	**0.7280**	**0.5912**	**14.25**
	BBO-MLP (WRF)	4.2824	2.0693	1.9708	32.44
	PM	0.5964	0.7723	0.6263	15.45
	WRF	4.7411	2.1774	1.8965	28.44
	ARIMA	0.7945	0.8935	0.6983	15.58
	MLP-BR	0.5689	0.7542	0.6022	15.43
2-hour ahead	BBO-MLP (Obs)	**1.3607**	**1.1665**	**0.9377**	**18.32**
	BBO-MLP (WRF)	4.9409	2.2228	2.0057	33.51
	PM	1.4762	1.2150	1.1120	19.62
	WRF	5.5589	2.3577	2.2598	39.22
	ARIMA	3.1940	1.7872	1.4252	21.92
	MLP-BR	1.5357	1.2392	0.9662	20.42
3-hour ahead	BBO-MLP (Obs)	**1.8410**	**1.3568**	**1.1791**	**25.09**
	BBO-MLP (WRF)	4.6956	2.1669	1.9317	38.95
	PM	2.1520	1.4669	1.4166	28.85
	WRF	5.1946	2.2792	2.0433	43.54
	ARIMA	4.9403	2.2227	1.7875	31.98
	MLP-BR	2.2022	1.4840	1.2367	27.74

Table 6.18 Improvement percentage (%) of different criteria, comparing the developed model with the other models, for one, two, and three hours ahead, Camariñas, Spring 2017

Step	Criteria	BBO-MLP(WRF)	Persistent	WRF model	ARIMA	MLP-BR
1-hour	MSE	87.01 %	06.76 %	88.27 %	03.00 %	02.22 %
	RMSE	64.82 %	05.74 %	66.56 %	18.52 %	03.47 %
	MAE	69.98 %	05.62 %	68.82 %	15.32 %	01.83 %
	MAEP	56.07 %	07.77 %	49.89 %	08.53 %	07.65 %
2-hour	MSE	72.46 %	07.82 %	75.52 %	57.40 %	11.39 %
	RMSE	47.52 %	03.99 %	50.52 %	34.73 %	08.79 %
	MAE	53.24 %	15.67 %	58.51 %	34.20 %	02.94 %
	MAPE	45.32 %	06.63 %	53.29 %	16.42 %	10.28 %
3-hour	MSE	60.79 %	14.45 %	65.06 %	62.74 %	08.95 %
	RMSE	37.39 %	07.48 %	40.47 %	38.95 %	08.57 %
	MAE	38.96 %	16.76 %	42.29 %	34.04 %	04.65 %
	MAPE	35.58 %	13.04 %	42.37 %	21.54 %	08.89 %



(a) Muralla IP(%) MSE

(b) Camariñas- IP(%) MSE

Fig. 6.20 MSE improvement percentage of the developed model against other models at Muralla and Camariñas stations, Spring 2017

Comparing the figures 6.19a, 6.20a, and 6.20b comparing MSE improvement respect to PM (standard reference benchmark) the new developed model performs better at Santiago-EOAS urban station, then at Muralla rural station and, as the lowest improvement, at Camariñas coastal station. But, if WRF NWP is used as reference, the new developed model performs similarly at the three stations. However, considering 1-hour time interval predictions as PM as reference, the new developed model performs better at the coastal station than at rural and urban stations. Therefore, improvement percentage is not a general criterion to evaluate the model performance between different locations. It is a more reliable criterion to compare different models, more than to compare the same model performances at different locations.

As it was shown in previous tables, MSE, RMSE, MAE, and MAPE of the developed model for Santiago-EOAS one-hour ahead prediction are 0.7706, 0.8778, 0.6681, and 15.32, respectively. The same statistics at Muralla are 0.7499, 0.8659, 0.5870, and 10.33, and at Camariñas are 0.5561, 0.7280, 0.5912, and 14.25, respectively. Conducting a direct comparison, outcome MAE and RMSE outperform results from the WPD-CCO-MLP model developed by Meng et al. [201] with 0.704 and 0.938 values, respectively. MAE and MAPE values from the new developed model for 1-hour ahead prediction at all tested Galician stations are compared to the best results in the literature from another model using transfer learning with a deep neural network (Hu et al. [292]). In [292], MAE and MAPE are 0.8057, 15.92 at Ningxa station, 0.7098, 21.90 at Jilin station, and 1.0738, 22.26 at Mongolia

station, respectively . In another study, Zhao et al. [296] improved WRF NWP using a Kalman filter. MAE and RMSE of yearly averaged data for NWP are reported as 1.1913 and 2.491, which decreased to 1.452 and 1.926 using Kalman filter, showing 17.95% and 22.65% of improvement for hourly data. The same comparison of MAE and RMSE in the new developed model gets improvements of 65.71% and 61.43% at the three testing stations. In [297], using Bayesian adaptive combination of short-term wind speed forecasts from a neural network, the best MAE, RMSE, and MAPE for 1-hour ahead prediction reported were 1.117, 1.508, 18.1 at site 1, and 1.050, 1.381, 14.1 at site 2, respectively. Compared to the new developed results for the same statistics at the three testing stations, MAE and RMSE are much better and, also, MAPE is lower in some cases.

6.3.3.3 BBO compare to other optimization methods

After finishing the new developed model testing using different case studies, also it was considered the effect of optimization methods in the model performance. As optimization is a key step in model development and training. MSE, RMSE, MAE, and MAPE different results were obtained, using BBO, GA, PSO, and ACO optimization methods for three different time horizons and at the three Galician meteorological stations with Spring 2017. As shown in table 6.19, BBO overtakes the other three optimization in most cases. These results are similar to others obtained by Mirjalili et al. [262] optimizing an MLP for classification problem. However, it must be noticed that sometimes evaluation criteria results are contradictory between them, so in these cases it is necessary to include more criteria to get a significant comparison between models and methods.

The results in table 6.19 are similar to the values in table 6.7 for the same optimization methods applied to the 1-minute M2-tower time series, although in table 6.19 results correspond to 3-hours ahead time interval at Galician meteorological stations. For one step ahead, BBO outperforms other optimization methods at Santigo-EOAS, considering MSE, RMSE, MAE, and MAPE, with values of 0.72, 0.85, 0.65, 12.86, respectively. At Muralla, MSE, RMSE, and MAE are 0.82, 0.90, and 0.65, and MAPE is 51.65, which is less than 38.84 provided by the ACO method. At Camariñas, MSE, RMSE, and MAE using BBO achieves is the best results, with 0.56, 0.75, and 0.59 values, respectively. However, MAPE using GA and BBO are 14.38 and 15.25, respectively, which are very similar. Also, 2-step and 3-step ahead results are shown in table 6.19, which are very similar to 1-step ahead results already discussed.

As an example, figure 6.21 shows the prediction model results for 2-hours time interval, using the four different optimization methods during the first week of March 2017. As it can be observed, BBO optimization provides the prediction results closer to the original time series, also with significant differences respect to other optimization methods at the highest and lowest wind speed values. Apart from the comprehensive statistical comparison, it confirms the election of BBO method as the best one between the four methods tested.

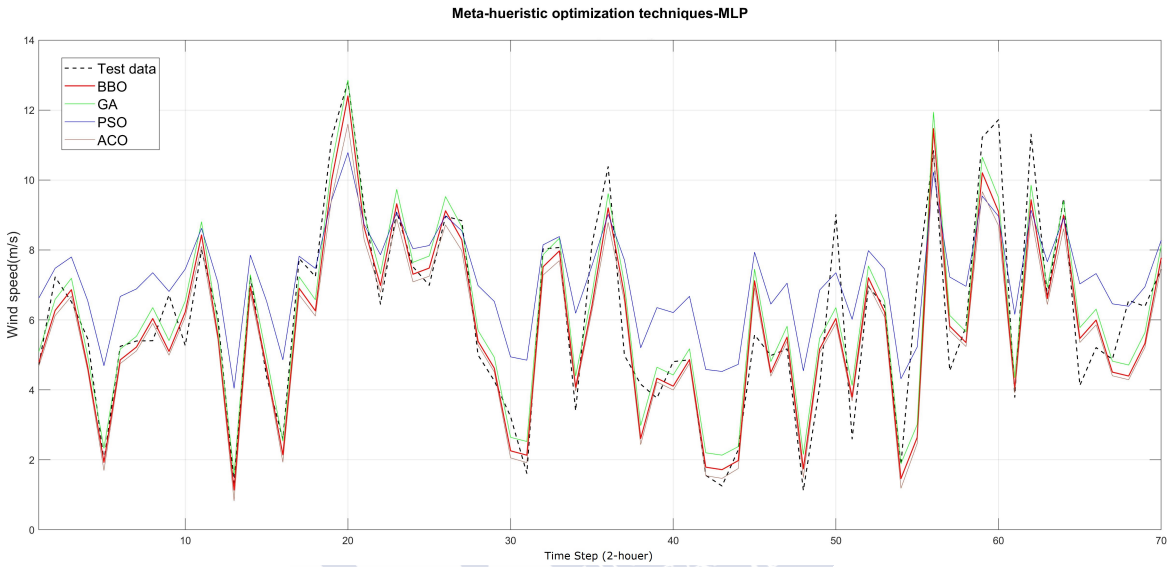


Fig. 6.21 Prediction example of optimized MLP using BBO, GA, ACO, and PSO, Muralla, Spring 2017

The following suggestions can be considered for future works to improved the accuracy of model and hyperparameter optimization:

- Expanding the range of hyperparameters possible values and using high-performance computing or parallel computing to tackle with the computation time;
- Adding components to the developed BBO-MLP model to increase the accuracy of results such as signal pre-processing techniques, error correction component, etc;
- Developing ensemble models with BBO-MLP or extended BBO-MLP with other signal processing techniques, as one of its prediction models;
- Clustering input data before feeding into the model.

Table 6.19 1-step, 2-steps, and 3 steps ahead wind speed forecasting of hourly data at Galician meteorological stations. MSE, RMSE, MAE and MAPE are used to report **MLP** performance optimized with different methods: BBO, GA, PSO, and ACO

Step	Site	Optz	MSE$(m/s)^2$	RMSE(m/s)	MAE(m/s)	MAPE(%)
1-Step ahead	**EOAS**	BBO	**0.72**	**0.85**	**0.65**	**12.86**
		GA	0.74	0.86	0.66	13.64
		PSO	1.11	1.06	87.15	18.78
		ACO	0.86	0.93	0.72	14.92
	Muralla	BBO	**0.82**	**0.90**	**0.65**	51.65
		GA	0.83	0.92	0.69	48.05
		PSO	1.44	1.19	0.91	80.97
		ACO	0.88	0.94	0.70	**38.84**
	Camariñas	BBO	**0.56**	**0.75**	**0.59**	15.25
		GA	0.59	0.77	0.60	**14.38**
		PSO	0.89	0.95	0.78	20.72
		ACO	0.72	0.85	0.71	17.23
2-Step ahead	**EOAS**	BBO	1.18	**1.08**	**0.79**	**15.00**
		GA	**1.12**	1.12	0.81	16.70
		PSO	3.36	1.83	1.56	45.10
		ACO	1.32	1.17	0.86	15.40
	Muralla	BBO	**1.39**	**1.18**	**0.86**	**83.07**
		GA	1.40	1.19	0.90	83.58
		PSO	1.42	1.20	0.89	89.64
		ACO	1.46	1.21	0.93	90.77
	Camariñas	BBO	**1.31**	**1.14**	**0.92**	**18.15**
		GA	**1.31**	**1.14**	0.93	18.36
		PSO	2.07	1.44	1.23	25.30
		ACO	1.46	1.21	0.95	18.76
3-Step ahead	**EOAS**	BBO	**1.49**	**1.22**	**1.02**	24.29
		GA	1.54	1.24	1.05	**23.45**
		PSO	4.02	2.00	1.74	35.42
		ACO	1.50	1.23	1.02	23.94
	Muralla	BBO	1.38	1.17	0.92	14.55
		GA	1.46	1.21	0.95	14.88
		PSO	3.84	1.96	1.51	24.81
		ACO	**1.34**	**1.16**	**0.84**	**13.35**
	Camariñas	BBO	**2.40**	**1.51´**	1.13	26.98
		GA	2.46	1.57	**1.06**	26.46
		PSO	3.37	1.84	1.29	26.34
		ACO	2.51	1.58	1.15	**26.40**

6.4 Summary

In this chapter, a semi-brutal force algorithm is described and developed to find the best hyperparameters set of an optimized MLP model. Three groups of hyperparameters are defined considering MLP model structure, optimization techniques, and data-related hyperparameters, and a new formulation of a global optimization problem. Different numbers of the input and hidden neurons with various activation functions were tested to find out the best set for the model hyperparameters. Also, four optimization methods, namely GA, ACO, PSO, and BBO, with different parents' numbers, generations' numbers, and values ranges were tested, in order to defined the optimization sets of hyperparameters. Additionally, several different input data features, time series lengths for training, and (train/validation)-test percentages compose the data-related hyperparameters. 10-fold cross-validation, AIC, and BIC were used for the model selection process, each of them including a specific set of hyperparameters.

This new developed BBO-MLP model was optimized and trained using 1-year wind speed time series from M2-tower at Colorado, U.S., with special interest in unstable conditions periods. Then, trained model was applied to predict short-term wind speed at the same U.S. station, and at three different Galician meteorological stations in urban, rural, coastal areas. The new model performance was compared to other models using MSE, RMSE, MAE, and MAPE as statistical criteria. Models for comparison include persistence model, ARIMA, MLP-BR and, at Galician stations, WRF operational numerical weather prediction from MeteoGalicia. In all cases the new model BBO-MLP performed better than the other models, in 1-hour to 3-hours ahead predictions. Also, BBO-MLP model statistics are similar and even better than other results currently published from other ML models and methods at different locations around the world.

Chapter 7

EMD-QBPSO-ELMAN-GA wind power prediction model: Feature selection

In this Chapter, a hybrid wind power prediction model is developed in light of the outcomes from the previous Chapters and the literature. The developed model combines different techniques to provide short- to medium-term wind power prediction. Both observational data and Numerical Weather Prediction (NWP) data compose the model input space. A novel Quantum-based Binary Particle Swarm Optimization (QBPSO) technique is developed as a feature selection technique to find out the best input dataset for the Elman neural network (ENN) model structure, which is optimized by using a Genetic Algorithm (GA). In addition, An Empirical Mode Decomposition (EMD) technique is used to deal with unsteady weather time series.

Weather and generated wind power generation datasets from two wind farms, Sotavento at Galicia, Spain, and La Haute Borne at France, are used as for both model tuning and testing. Finally, the results of this new developed hybrid wind power forecasting model are compared with other models results, in order to test its quality and comparative model performance.

7.1 Introduction and background

Feature selection, a process of selecting relevant features or variables in a machine learning (ML) model, is an essential and well-discussed topic in ML models, particularly neural networks (NN) [123]. In wind speed and power prediction models, the input set of

variables, i.e., the model features, plays a vital role in the model performance and directly affect the accuracy of the results. Larger size of features requires larger NN model in terms of nodes in the input and hidden layer(s), increasing model complexity and the computational efforts in both tuning and application. Therefore, there is a trade-off between the number of selected features and the generalization error: selecting a few of features, model produces inaccurate results; while using all features, the model may be overfitted, with less generalization [298]. Hence, to develop an accurate wind speed and power forecasting, addressing feature selection is as important as hyperparameter optimization, thoroughly discussed in Chapter 5. In addition, several examples of hybrid and combined models which proved to be more successful than individual techniques [79, 136, 146, 144]. In fact, most of the recently developed models for wind speed and power prediction are composed of different techniques of neural network optimization [299], feature selection [216] and/or signal decomposition [20]. In addition, hybrid models built with several prediction core models are also proved to be more successful than individual models [274, 80]. Therefore, according to the state-of-art, hybrid prediction model seems to be a necessary election.

Regarding neural network structure and its supervised learning process, there are two main groups: FeedForward neural networks (FNN) and recurrent neural networks (RNN). The main difference between the two groups is that the latter can have signals traveling in both directions by introducing loops in the network (as a "memory"), while FFNN information can only move in the forward direction. Welch et al. [73] proved the superiority of Elman neural network (ENN) as a popular type of RNN compared to all types of NNs, with the disadvantage of increasing running time because of its feedback loops. Also, Xu and Mao [76] proposed a short-term wind power forecasting based on Elman NN optimized with particle swarm optimization (PSO). Its maximum error was reported as 936.88 (kWd). Zhang [300] used GA-Elman and support vector machine- generalized regression neural network (SVM-GRNN) hybrid model for short-term wind power prediction. In this case, a small RMSE equal to 0.16 (power in W) was reported. As a consequence, Elman neural network (ENN) is a good election for NN structure in wind speed and wind power prediction models.

When a model prediction input dataset includes weather time series, the unsteady and non-linear nature of weather parameters can produce chaotic behavior in the model results; which can be a problem in model applications. This is why signal decomposition techniques can help to deal with non-linear weather input datsets before using them in a prediction model, both wind speed and wind power models. For instance, Liu et al. [153] proposed a double signal decomposed technique (DT) combining Elman NN with wavelet package

decomposition (WPD) and fast ensemble empirical mode decomposition (FEEMD). The latter one was used to go one step further on decomposing the signals obtained by the first method. The derived wind speed prediction method was reported to be more accurate than other individual models, such as MLP, ELMAN, ARIMA, and Persistence Random Walk Model (PRWM). In another study, Liu et al. [301] compared two decomposition methods in a hybrid wind speed forecasting model using Back-Propagation NN and Radial-Based Function (RBF). An average of 50% improvement was reported using hybrid decomposition models compared to individual ones for 1-hour ahead wind speed prediction. Wang et al. [152] used empirical model decomposition (EMD) to decompose input data before feeding into an Elman network to predict hourly wind speed data on seasonal datasets. Both combined and hybrid models comparisons against individual models, i.e., ENN against backpropagation NN, and EMD-ENN against ENN, show that always individual models produced worse results in all seasons.

As it was highlighted above, feature selection problem is a key step in these prediction models. Regarding the feature selection problem, Salcedo-Sanz et al. [124] developed and applied a novel coral reefs optimization algorithm as feature selection (FS) technique in a wind speed forecasting model with an extreme learning machine as the prediction method. These developed FS techniques were used to identify the best model input dataset from the outputs of a meteorological mesoscale model. In another work, Cornejo-Bueno et al. [184] proposed a grouping genetic algorithm-extreme learning machine hybrid model to predict wind power time series. In addition, a Gaussian Processes (GPR) method was developed to compare model results using different input datasets. ERA-Interim reanalysis from the European Center for Medium-Range Weather Forecasts was used as the input data sources. Results showed that GPR model with all the available 48 input variables outperforms any input data subset. Model testing at three sites in Galicia, Spain, produced an average R-square of 71.13% and MAE of 3.99 MW, for the best case. Li et al. [216] used a conditional mutual information-based technique to select the best set of inputs in a wavelet-based neural network ensemble wind power prediction model. A partial least-square regression method was also used to aggregate the prediction results of different prediction parts. This model signficantly improved persistence model (PM) results: MAPE and NRMSE improvements were around 90% for 1-hour ahead prediction, and from 70% to 85% for up to two days ahead. In another recent work, Zhang et al. [302] developed and applied a Seq2Seq deep neural network on the Sotavento wind farm, at Galicia, Spain, using a combination of observational and NWP data. K-means clustering was applied to create three main clusters of NWP data as input variables to the model. The best results for the RMSE and MAE statistic metrics and

the 1-day forecast horizon with 10-min data resolution were reported as 129.3 (kWh) and 81.1 (kWh), respectively. According to these examples, feature selection technique can significantly change the model performance.

Considering these previous works, Elman recurrent neural network was chosen as the core of the new developed short- to medium-term wind power prediction model. In addition, a newly developed quantum-based binary particle swarm optimization was developed as a feature selection method. This method follows the classical wrapper method, where a global search algorithm looks for the best set of features that minimize a likelihood function. About model optimization, genetic algorithm (GA) was used to train the Elman NN. Finally, EMD was added to decompose the input signals before feeding them into the prediction core of the model.

As findings described in previous chapters about the new developed wind speed prediction model proved to achieve very good model performance, most of the defined wind power model hyperparameters are based on them.

7.2 New developed EMD-QBPSO-Elman-GA model

Following, the different parts of the new developed wind power prediction model are described in detail.

7.2.1 Empirical mode decomposition (EMD)

Empirical mode decomposition is a signal analysis to breaks down unsteady time series into a finite number of sub-series or components sharing more statistical properties. These components are called intrinsic mode functions (IMFs), with the last component named residual. Given a set of data points in a original time series $\mathscr{X}(t) = \{x_1, x_2, \ldots, x_k\}$, and considering C_i as an IMF at the ith level of decomposition, and R_n as the residual after n number of decomposition phases, the original time series can be expressed by equation 7.1 [303],

$$\mathscr{X}(t) = \sum_{i=1}^{n} C_i + R_n \tag{7.1}$$

IMFs are defined as functions and should satisfy two requirements: 1) the number of extrema and the number of zero-crossing should not exceed one in the whole dataset and, 2)

the mean value of the enveloped defined by local maxima and the envelope value defined by local minima should be zero at any point.

To obtain the IMFs components, the following iterative process must be applied ([11] [144]),

1. All the local extrema (local minima and local maxima) of the time series $\mathscr{X}(t)$ should be identified.
2. Connect all the local minima and local maxima together using a Cubic Spline Line to generate lower envelop $\{\mathscr{X}_{low}(t)\}$ and upper envelop $\{\mathscr{X}_{up}(t)\}$, respectively.
3. Averaging the high and low envelops to compute the mean envelope: $M(t) = \frac{\mathscr{X}_{low}(t)+\mathscr{X}_{up}(t)}{2}$
4. Extract the residual as $Z(t) = \mathscr{X}(t) - M(t)$
5. Apply equation 7.2 as convergence criterion:

$$\sum_{t=1}^{m} \frac{[Z_{j-1}(t) - Z_j(t)]^2}{[Z_{j-1}(t)]^2} \leq \theta(t = 1,2,...;t = 1,2,...,m) \qquad (7.2)$$

   This criterion is applied as follows: Check whether $Z(t)$ is an IMF or not. If positive, set $C(t) = Z(t)$, and replace $\mathscr{X}(t)$ with the residual, $\mathscr{X}(t) - M(t)$. If it is not the case, replace $\mathscr{X}(t)$ with $Z(t)$ and repeat steps 2 to 4 until the convergence criterion is satisfied. Where m is the signal length, and θ is the termination parameter, with a value between 0.2 and 0.3.
6. Repeat steps 1 to 5 until all the IMFs are found.

7.2.2 Elamn Neural network

Elman neural network (ENN) is one of the various types of recurrent neural networks (RNN), as it was introduced by J.Elman [152] in 1990. As shown in figure 7.1, in this type of NN, hidden nodes are connected to a new layer (context layer) through which feedback of the previous time step inputs is carried, as some kind of memory. Therefore, the previous stage of the hidden neurons is saved, which makes possible to consider the effect of the previous time step in the current one. Input and hidden layers neurons are fully connected exactly, similar to MLP networks. Equation 7.3 describes the mathematical formulation of an Elman neural network,

$$\mathscr{T}(t+1) = \omega'_{1,k} + \sum_{j=1}^{J} \phi \left(\omega_{1,j} \gamma \left(\sum_{c=1}^{C} \sum_{j=1}^{J} \omega_{c,j} S_c(t-1) + \sum_{i=1}^{I} \sum_{j=1}^{J} \omega_{i,j} V_i(t) \right) + \omega'_{1,j} \right) \tag{7.3}$$

Where $\mathscr{T}(t)$ is the model output, i, j, k and c refer to input, hidden, output, and context layer neurons, respectively. ω stands for the all connections weights except biases, ω' represents the bias connection weights, S_k and V_i represent the outputs of context and input layer, respectively, and ϕ and γ stand for output and hidden activation functions, respectively. The output of a specific node in the context layer is obtained by equation 7.4,

$$S_{c'}(t) = \omega_{c'} \left(\gamma \left(\sum_{i=1}^{I} \omega_{i,j'} V_i(t) + \sum_{c=1}^{C} \omega_{c,j'} S_c(t-1) \right) + \omega'_{1,j'} \right) \tag{7.4}$$

Where c' and j' are the two corresponding nodes at the context and hidden layers, respectively.

As shown in figure 7.1, the feedback loops are from the hidden layer to the context layer. As another important point, there is no activation function in the context layer (or it can be considered equal to 1), so the neuron numbers in the context layer are equal to corresponding hidden layer neurons. In addition, Bios units, which value is equal to 1, are fully connected to the hidden and output layer. Of course, input nodes can be more than one, depending on the input variables number.

Given a training set of time series with T time steps, $X = \{x_1, x_2, ..., x_T\}$, the best set of the weights for all connections in the network are achieved during the learning process by minimizing the mean square error (MSE) cost function, described in equation 7.5 as follows,

$$\mathscr{E}(w) = \sum_{t=1}^{T} (y_p(t) - y_R(t))^2 \tag{7.5}$$

Where $\mathscr{E}(w)$ is the cost function for all w weights, $y_p(t)$ is the outcome of the Elman network at time step t, and $y_R(t)$ is the actual value at the same time step.

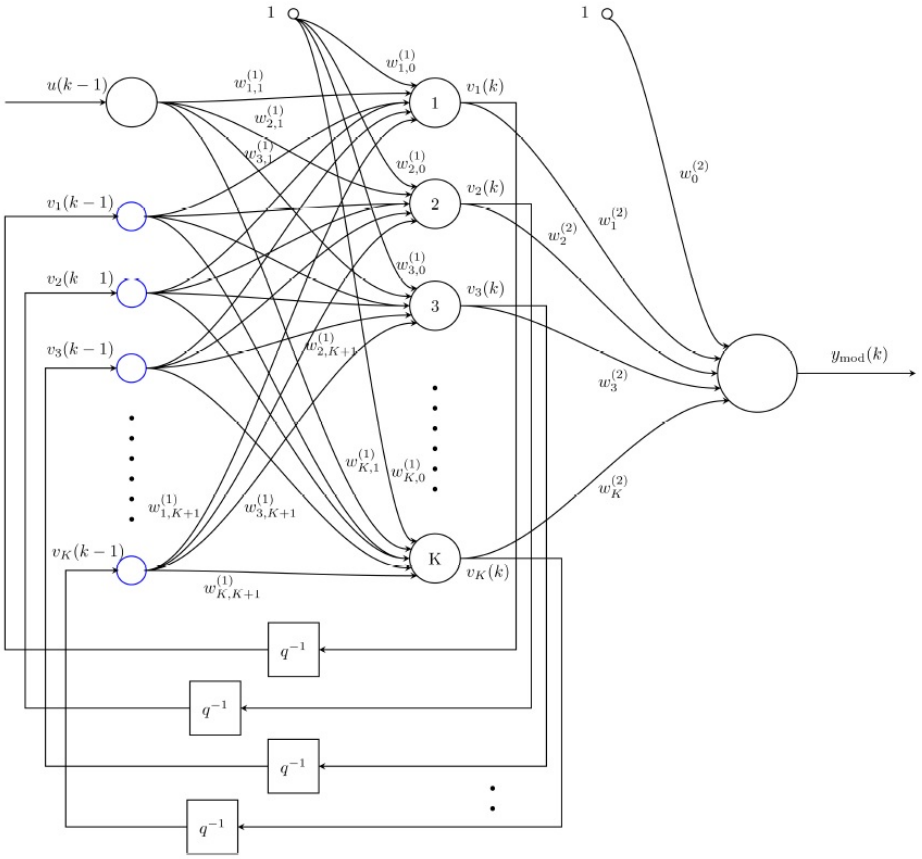


Fig. 7.1 Schematic view of an Elman neural network- Redistributed figure from [304]

7.2.3 Elman-GA

Genetic algorithm, originally proposed by J.Holland in 1975, is an heuristic optimization method is inspired by the genetic mechanism of *survival of the fitness* in the biological world. In order to formulate an optimization problem in the form of GA each possible solution will be considered as a parent, which produces new generations through several operations, such as elite selection, crossover, mutation, etc., between the parents through their chromosomes, with different variable values for each specific parent or solution. The best results are those with the highest fitness values, and the process will continue until a the following termination criterion is achieved: Either the number of generations will reach to the end (maximum number) or the fitness value of the two consecutive generations are lower than a pre-defined threshold [263].

Successful application of GA in training ANN models are reported in the literature for wind speed, and wind power prediction models [206, 305, 199, 203, 13]. As an example, GA was also reported to outperform PSO in training MLP in a hybrid wind speed forecasting model [206].

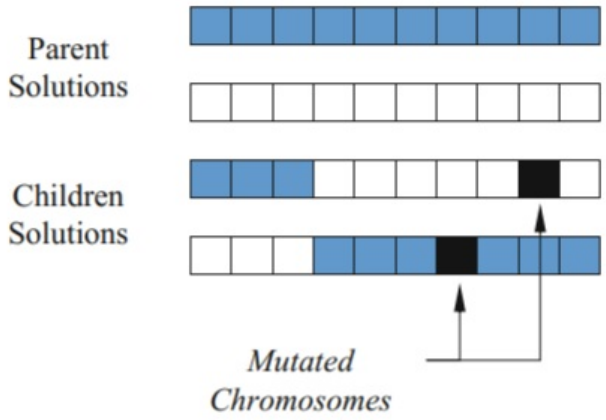


Fig. 7.2 Schematic representation of GA mutation operation [306]

In the new developed wind power model, first Elman neural network (ENN) was trained using BP method. This best outcome was set as the initial condition of the Elman-GA training. Problem formulation and its algorithm for training ENN by weights optimization is described as follows,

1. Define all the n number of weights (ω and ω') of the ENN as a possible solution (Parent): $P = \{c_1, c_2, ...c_n\}$, where c_i represents the i_{th} chromosome of a parent.
2. Set the best outcome of the Elman-BP training process, as the initial condition for the GA optimization $P_{initial,GA} = P_{Best,BP}$, and produce the first generation of solutions with m different parents: $G_1P = \{P_1, P_2, ...P_m\}$
3. Apply elite selection (EO), crossover (CO), and mutation (MO) operations on the parents to create a new generation, $G_iP = \{EO, CO, MO\}(G_{i-1}P)$, as follows [306],

    - EO: α percent of the previous generation best solutions will be selected and directed to the new generation.
    - CO: A single-point crossover is applied. Figure 7.3 shows how the single-point and double-point crossover operation works.

- MO: As shown in figure 7.2, one or multiple genes are altered after creating children's solutions in mutation operation. The mutation rate is set to low in GA because high mutation rates convert GA to a primitive random search.

4. Check if there is any duplication in the results to be taken out.
5. Check if all the chromosome values in all parents,G_iP_j, are in the pre-defined range: $U_{min} \leq c_k \leq U_{max}$, where $j = 1 : m$, $k = 1 : n$, and i represents the i_{th} generation of parents.
6. Run the ENN model with all the solutions (parents) of generation i, G_iP and form the fitness value vector by calculating this value for each of the m parents: $\mathscr{E} = \{e_1, e_2, ...e_m\}$ using equation 7.5.
7. Sort the parents based on fitness values and check if the termination criteria are achieved.
8. Repeat steps from 3 to 7.

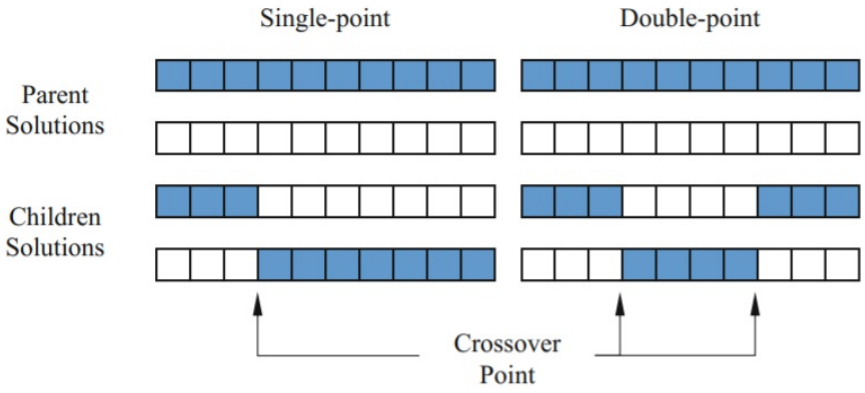


Fig. 7.3 Schematic representation of GA single-point and two-point crossover operation [306]

7.2.3.1 Quantum binary particle swarm optimization (QBPSO)

Particle swarm optimization (PSO) was first proposed by Kennedy and Eberhart [307] in 1995 as an evolutionary optimization algorithm that is based on the sociological and biological behavior of animals searching for food. The main idea comes from the social behavior of schools of fish and flocks of birds. The method has been applied with modifications in many wind speed and power prediction models with successful results [159, 206, 308].

In traditional PSO, each particle is considered as a possible solution of the global optimization problem. Particle and velocity as set and continuously updated through consecutive iterations. Given $X_i^t = \{X_{i,1}^t, X_{i,2}^t, ..., X_{i,d}^t\}$ as position of a particle i at iteration t in a $d-dimensional$ space, each particle can remember its best position and velocity along each dimension as $V_i^t = \{V_{i,1}^t, V_{i,2}^t, ..., V_{i,d}^t\}$. The best position of all particles in the swarm is called global best (gbest) and expressed as $G^t = \{G_1^t, G_2^t, ..., G_d^t\}$; and the best position for particle i is called personal best (pbest) and expressed as $P_i^t = \{P_{i,1}^t, P_{i,2}^t, ..., P_{i,d}^t\}$. Updated position and velocity at iteration $t+1$ are defined by the following equations,

$$V_{i,j}^{t+1} = \omega V_{i,j}^t + c_1.r_1(P_{i,j}^t - X_{i,j}^t) + c_2.r_2(G_j^t - X_{i,1}^t) \tag{7.6}$$

$$X_{i,j}^{t+1} = X_{i,j}^t + V_{i,j}^{t+1} \tag{7.7}$$

$$\omega = \omega_{max} - t\frac{\omega_{max} - \omega_{min}}{T} \tag{7.8}$$

Where i and j are particle and dimension index, respectively, c_1 and c_1 are two positive numbers known as accelerating coefficients, r_1 and r_2 are random values in range of $(0,1)$, ω is called inertia weight, used for the balance between gbest and pbest, ω_{max} and ω_{max} are predefined values normally set to 0.9 and 0.4, respectively [307], and t and T are the numbers of current and total iterations, respectively.

As a new particle-based optimization method, Quantum behaved PSO was first introduced by Sun et al. in 2004 [309]. As shown in figure 7.4, the behavior of particles in QBPSO is different from PSO. In QBPSO, a wave function $\psi(x,t)$ is assigned to each particle, therefore iterations are changing this wave function. More details on QPSO can be found in [310]. In the new wind power prediction model developed in this work a new QBBPSO optimization algorithm was done. It is inspired in the work by Jeong et al. [311], and combines the binary version of traditional PSO with the concept and principles of quantum computing, such as a quantum bit, Q-bit, and superposition, $\{o,1\}$, of states. However, any property state keeps either 0 or 1 value, as a binary system. But particle velocity is now recalculated following quantum computing. As a consequence, both inertia weight and acceleration coefficients are removed, and only a rotation angle, $\Delta\theta$, is required. Following [312], the position of an agent/particle needs to change with the probability of its velocity. In order to do this, a transfer function is necessary to map velocity values to probability values, for updating the positions, which equations are provided in table 7.2. Considering Q-bit as the smallest unit of information for each particle, it is defined with a pair of numbers, (α, β), in a way that

$|\alpha|^2+|\beta|^2=1$. The position vector of the ith particle, $X_i=\{X_{i,1},X_{i,2},...,X_{i,d}\}$ is updated by probability of $|\beta|^2$ stored in the correspondent Q-bit, q_i. The jth element of the ith particle will be updated by equation 7.9 [313],

$$X_{i,j}=\begin{cases}1, & if\ rn_{i,j}\leq|\beta_{i,j}^2|\\ 0, & Otherwise\end{cases} \tag{7.9}$$

Where $rn_{i,j}$ is the uniformly distributed random number between [0,1]. The rotation angle, $\Delta\theta$, is used to update Q-bits α and β values, as described in equation 7.10,

$$\Delta\theta_{i,j}=\theta\times\{\psi_{1,i}\times(X_{i,j}(P)-X_{i,j})+\psi_{2,i}\times(X_{i,j}(G)-X_{i,j})\} \tag{7.10}$$

Where θ is the magnitude of rotation angel, $X_{i,j}(P)$ and $X_{i,j}(G)$ are the personal best and global best, and $\psi_{1,i}$ and $\psi_{1,i}$ obtained by comparing the fitness of current position of particle i with those of $Pbest_i$ and $Gbest$, respectively, as described in equations 7.12 and 7.12,

$$\psi_{1,i}=\begin{cases}1, & if\ Fitness(X_{i,j})<Fitness(Pbest_i)\\ 0, & Otherwise\end{cases} \tag{7.11}$$

$$\psi_{1,i}=\begin{cases}1, & if\ Fitness(X_{i,j})<Fitness(Gbest)\\ 0, & Otherwise\end{cases} \tag{7.12}$$

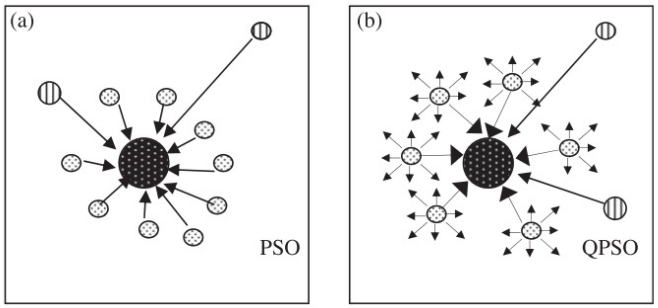


Fig. 7.4 movement of particles in a) PSO and b) QPSO [310]

7.2.4 EMD-QBPSO-ENN-GA framework

The general framework of the new developed model is illustrated in figure 7.5. The input dataset, composed by observational and numerical weather prediction WRF model data, will form the feature space, as shown in figure 7.6. After finding the best set of using QBPSO optimization, it will pass through the EMD technique, and the resulting IMFs sets from all the input variables will be fed into an Elman-GA method to be tunned with that specific signal properties. Finally, the prediction results for different IMFs will be aggregated to obtain the predicted wind power. Previous 48-time steps of the observational data 48 previous and 48 future time steps of the WRF data will form the feature space (Figure 7.6), with a size of $3 \times 2^{48} + 5 \times 2^{96}$.

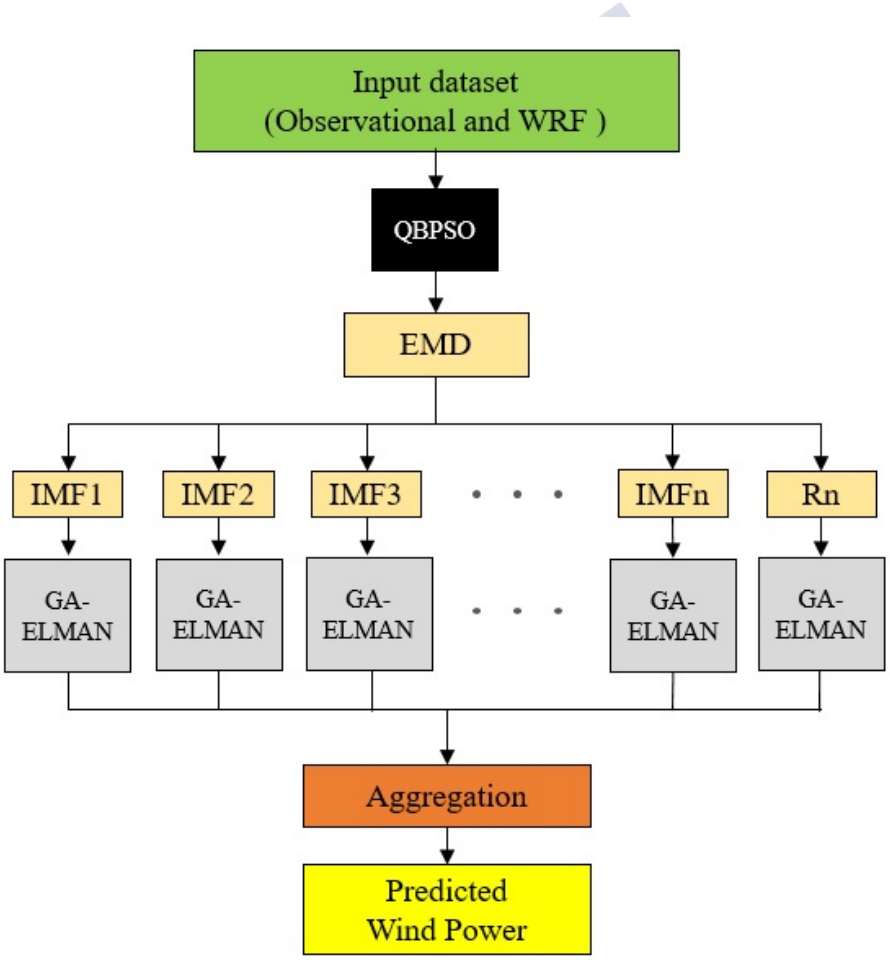


Fig. 7.5 Schematic presentation of the new proposed EMD-QBPSO-ELMAN-GA wind power prediction model

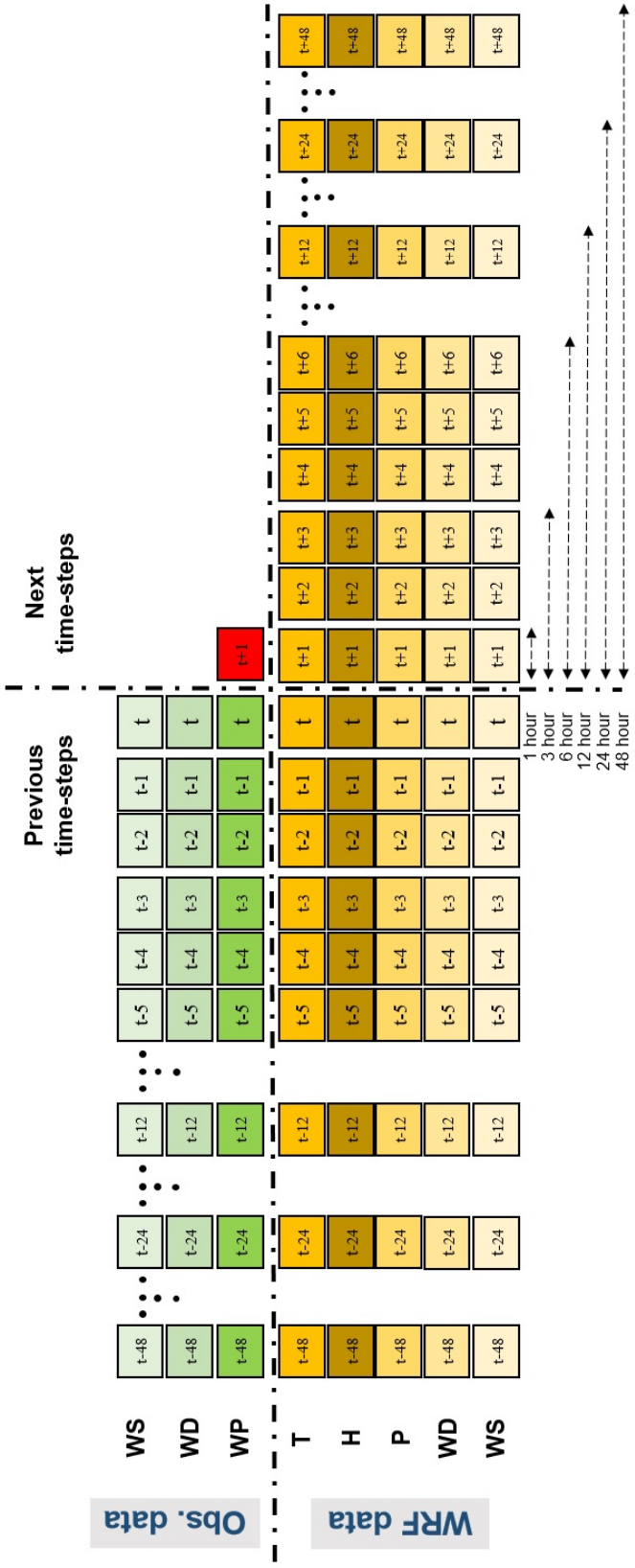


Fig. 7.6 Schematic representation of the feature space in the new developed model

7.2.5 Wind power models criteria

To evaluate the different wind power prediction models to be tested, including the new developed EMD-QBPSO-Elman-GA model, mean absolute error (MAE), root mean square error (RMSE) [16], and correlation coefficient or R-squared [140] the different wind power prediction models to be tested, including the new developed EMD-QBPSO-Elman-GA model, in equations 7.13 to 7.15, respectively.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - P_i)^2} \tag{7.13}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{P}_i - P_i| \tag{7.14}$$

$$R^2 = 1 - \frac{RSS}{TSS} \tag{7.15}$$

Where P_i represents the actual wind power value at ith time step, $\hat{P}_i$ is the corresponding predicted value of N data points, RSS and TSS mean Sum of Squares of Residuals, and Total Sum of Squares, respectively.

7.3 Results and discussion

The new developed wind power prediction model was tuning and tested on two wind farms, Sotavento in Spain and La Haute Borne in France, and their results were compared to other models. In Sotavento, predictions from 1-hour to 24-hours were obtained, while in La Haute Borne, short-term wind power forecast (10-minutes time interval) was tested.

7.3.1 Initial conditions

Regarding the ENN hyperparameters, general structure follows $(n, f(n), 1)$ structure, where $f(n)$ is described in equation 7.16 [82, 206]. In this new developed model, sigmoid and identity activation functions are used in the Elman hidden and output layers, respectively. For the GA-Elman method, modified sigmoid and hyperbolic tangent (tanh) functions already shown in Chapter 6 and inspired by [314] are used, as described in equations 7.17, and 7.18.

$$f(n) = round(\sqrt{(m+n)} + rand(1 \sim 10)) \tag{7.16}$$

$$\phi(x) = \frac{1}{1+e^{-0.03x-1}}, \quad [-1,1] \tag{7.17}$$

$$\gamma(x) = 3.78 \times tanh(x) - 0.364, \quad [-4,4] \tag{7.18}$$

Observational and numerical weather prediction (NWP) input data are divided into two main categories: training-validation dataset, and test dataset, with a portion of 70% and 30%, respectively. For the training-validation process, 10-fold cross-validation method is used [315]. Besides, input data normalization is required to achieve a mean value close to zero; as normalization speeds up learning and leads to faster convergence [316].

The optimization techniques hyperparameters are summarized in table 7.1. These results were obtained by combination of several tests (QBPSO) and other published models values (GA).

Table 7.1 QBPSO and QA hyperparameters definition

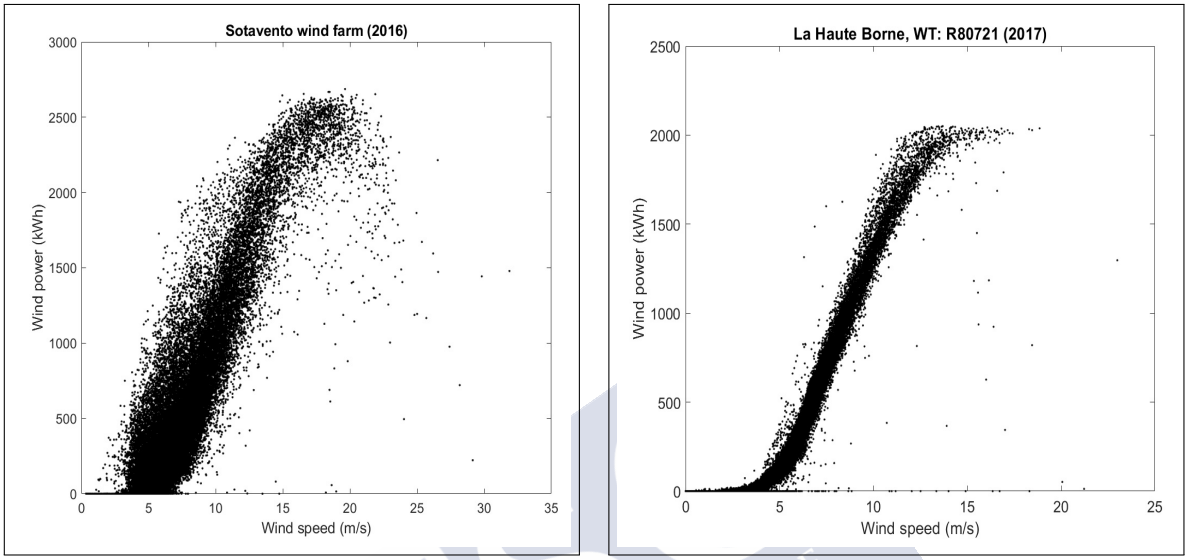
QBPSO	GA [262]
Generation number: TBD	Generation number: 200
Number of particles: TBD	Number of parents: 200
Number of parents: TBD	Minimum and maximum value: 5
	Elite selection percentage: 1%
	Crossover: Single-point(probability=1)
	Mutation: Uniform(Probability=0.01)

TBD: to be determined.

7.3.2 Wind power curves

Wind power depends not only on the wind speed and other weather conditions, but also on turbine characteristics, described by its wind power curve. Wind power depends not only on the wind speed and other weather conditions, but also on turbine characteristics, described by its wind power curve. In figure 7.7, the power curve for two wind-specific turbines in the two wind farms selected as case studies are shown. The cut-in speed for the Sotavento wind turbine seems to be lower than the R80711 wind turbine in La Haute Borne wind farm. Also, maximum wind speed and the turbine's cut-out speed in Sotavento are higher than in La Haute Borne turbine. Considering the produced wind power for a specific wind speed, the turbine installed in Sotavento should be more powerful than the one in La Haute borne.

However, wind power generation between 5 and 10 m/s in Sotavento turbine is wiser than in La Haute Borne turbine, which could have some reasons for controlling the turbine to produce a specific amount of energy demand change.



(a) Sotavento, Spain (2016) (b) La haute borne, France (2017)

Fig. 7.7 Power curves from Sotavento and Haute Borne wind turbines with 10-minutes time-interval data

7.3.3 New model QBPSO optimization in Sotavento case study

Eight different functions were used to test the QBPSO-Elman network with Sotavento dataset, in order to identify the best function. Three functions performed well, while the results of the other five functions were not acceptable. RMSE and MAE results for 1-hour ahead wind power prediction using Sotavento wind farm datasets are listed in table 7.2 for both S-shaped and V-shaped functions. Three of them performed well (S3, V2, and V4), while results from the other functions are not acceptable. As shown, the V2 function provided the best results with 181.33 (kWh) and 126.63 (kWh) for RMSE and MAE, respectively. Figure 7.8 shows wind power prediction results using prediction results for the three best functions. V2 function provides the best results, with 181.33 (kWh) and 126.63 (kWh) for RMSE and MAE, respectively..

Figure 7.9 shows the convergence range for different numbers of particles (P) for the best QBPSO transfer function, V2. Regarding the number of generations has a linear relationship

with the computation time, and considering the error evolution, 50 is selected as the optimal number of generations, and P is set to 250.

Table 7.2 Evaluation criteria for different QBPSO functions in wind power QBPSO-ELMAN model: 1-hour ahead. Best results in bold text

Func.	Mathematical Formula	Evaluation Criteria	
		RMSE	MAE
S-shaped	$S1 = \frac{1}{1+exp(-V)}$	640.3639	585.1063
	$S2 = \frac{1}{1+exp(-2\times V)}$	329.2336	254.7054
	$S3 = \frac{1}{1+exp(\frac{-V}{2})}$	252.0137	157.4214
	$S4 = \frac{1}{1+exp(\frac{-V}{3})}$	583.9917	513.6948
V-shaped	$V1 = \left\|erf(\sqrt{\frac{\pi}{2}\times V})\right\|$	340.4882	302.4369
	$V2 = \|tan(V)\|$	**181.3218**	**126.6317**
	$V3 = \left\|\frac{V}{\sqrt{1+V^2}}\right\|$	418.3774	329.6119
	$V4 = \left\|\frac{2}{\pi}tan^{-1}(\frac{\pi}{2}\times V\right\|$	253.0029	165.2870

Nomenclature: erf: error function: $erfz = \frac{2}{\pi}\int_0^z e^{-t^2}dt$

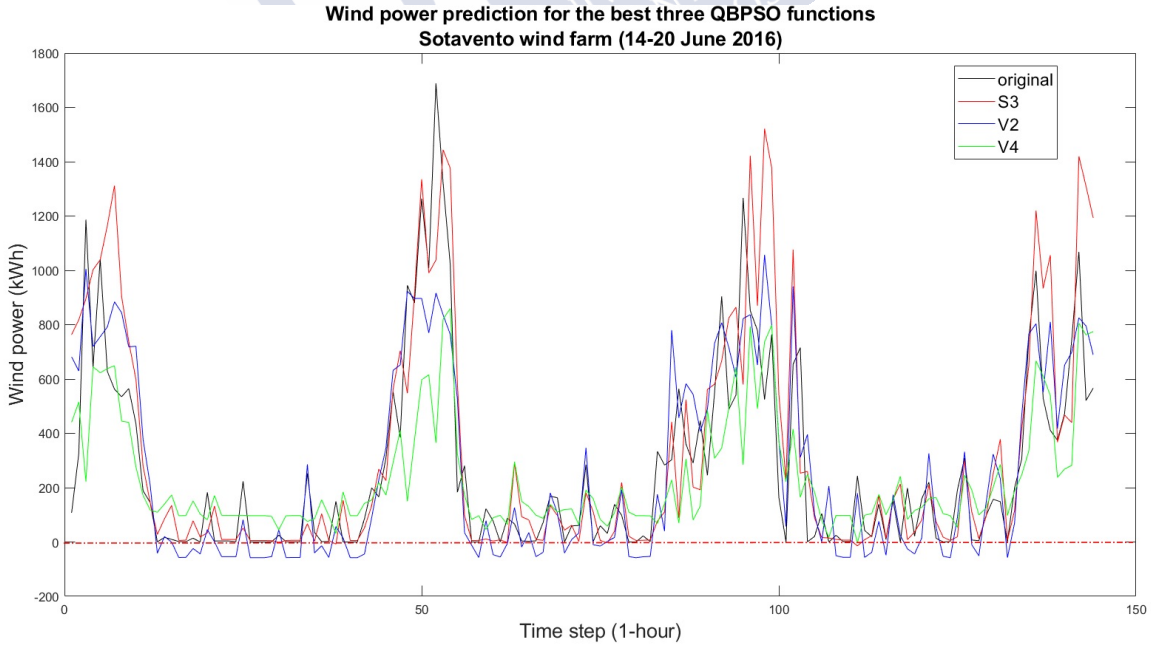


Fig. 7.8 Wind power forecasting visual comparison using the three best QBPSO functions (S3, V2, and V4) and original 1-hour wind power time series from Sotavento wind farm, 14-20 June 2016

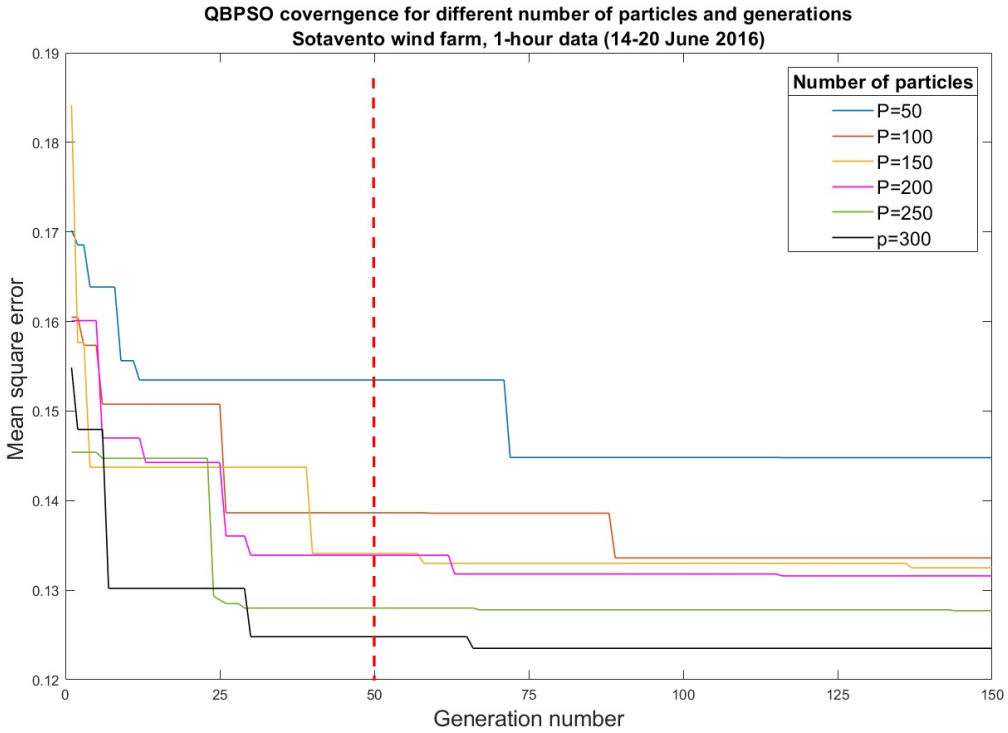
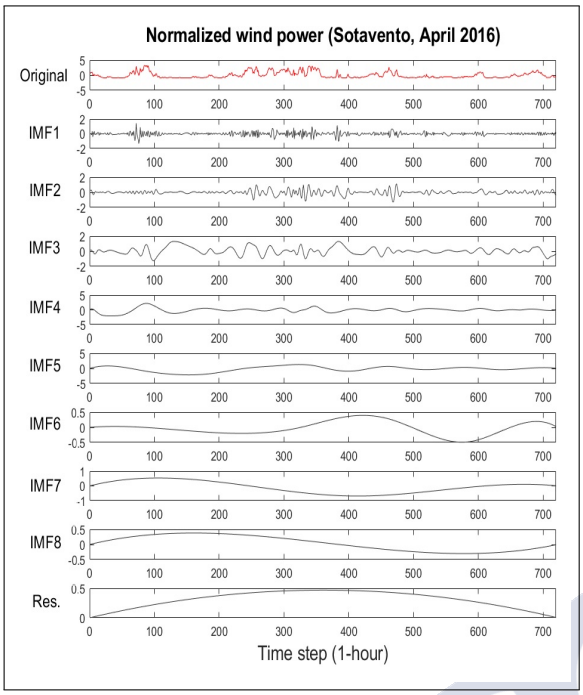


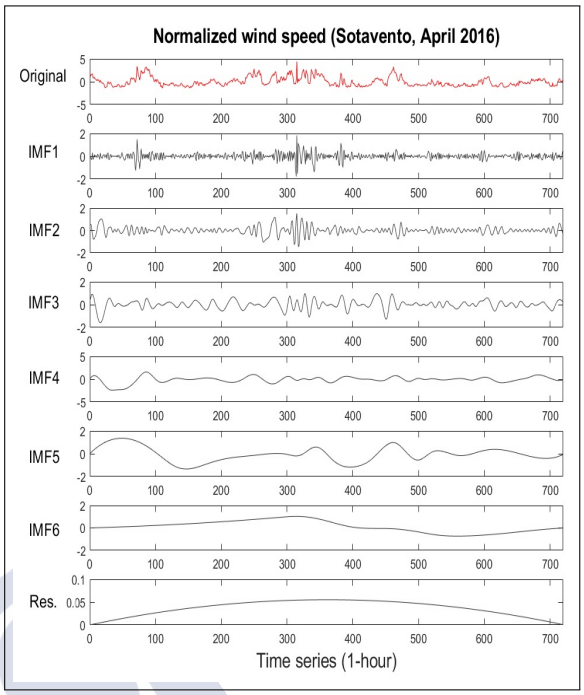
Fig. 7.9 QBPSO-V2 convergence rate using different number of particles, P

7.3.4 Empirical mode decomposition in Sotavento case study

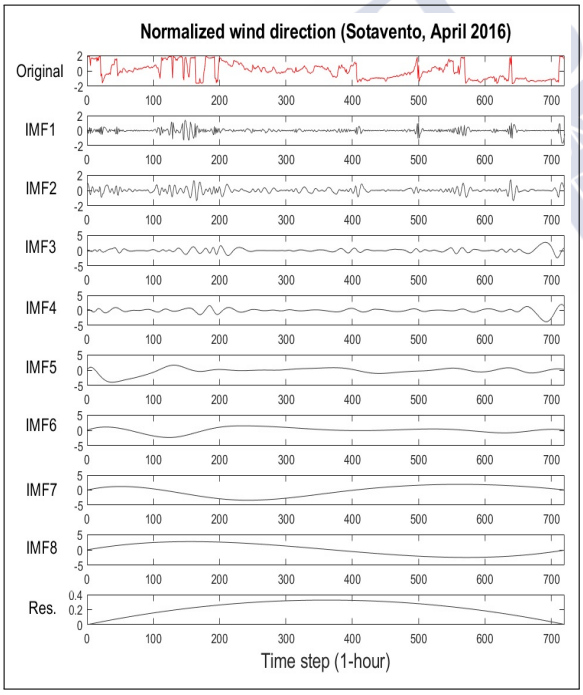
Figure 7.10 shows empirical mode decomposition (EMD) using both Sotavento observations and WRF numerical weather prediction normalized datasets as model input. Wind speed, wind direction, wind power, temperature, humidity, and pressure hourly time series in April 32016, are decomposed in intrinsic mode functions (IMFs) and a residual component. It is important to highlight that different time series have different numbers of IMFs depending on the unsteady level of the signal. From the other side, each group of IMFs from various parameters will be input to a separate QBPSO-Elman-GA model. Hence, a non-identical number of IMFs for different variables could make a problem. In order to solve this, the maximum number of IMF considering all variables will be set as a reference, and for those variables with less IMF than the reference time step, more IMFs defined as $y = 0$ (as empty IMFs) will ll be added. As an example, in figure 7.10, wind power has 8 IMFs, while wind speed has 6. So, Residual will be considered as IMF7, and two additional components with $y = 0$ definition will be added as IMF8 and new Residual component.



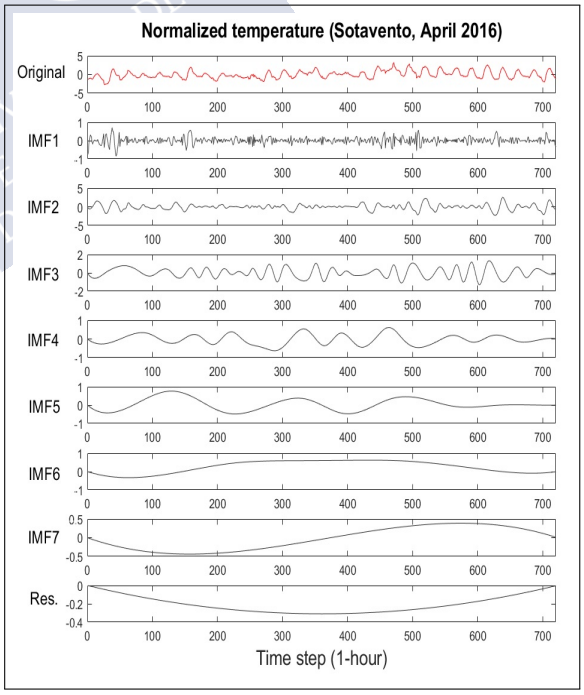
(a) Decomposed wind power (Obs.)



(b) Decomposed wind speed (Obs.)

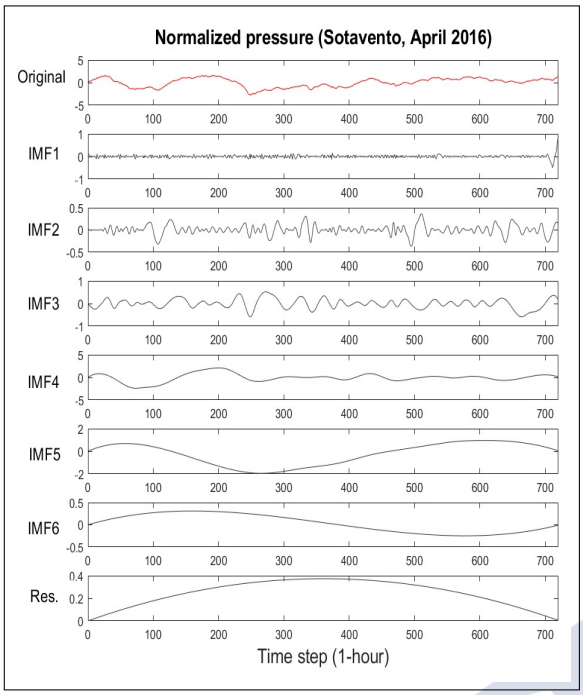


(c) Decomposed wind direction (Obs.)

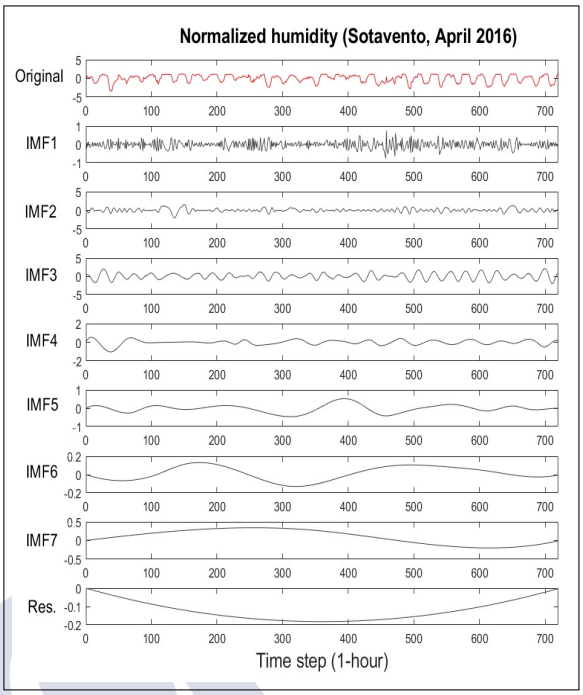


(d) Decomposed temperature (WRF)

Fig. 7.10 Decomposed observational time series (WP, WS, and WD) and WRF parameters (T, H, and P) for Sotavento wind farm on April 2016 using EMD technique



(e) Decomposed pressure (WRF)



(f) Decomposed humidity (WRF)

Fig. 7.10 (Continued) Decomposed observational time series (WP, WS, and WD) and WRF parameters (T, H, and P) for Sotavento wind farm on April 2016 using EMD technique

7.3.5 New wind power prediction model testing at Sotavento wind farm

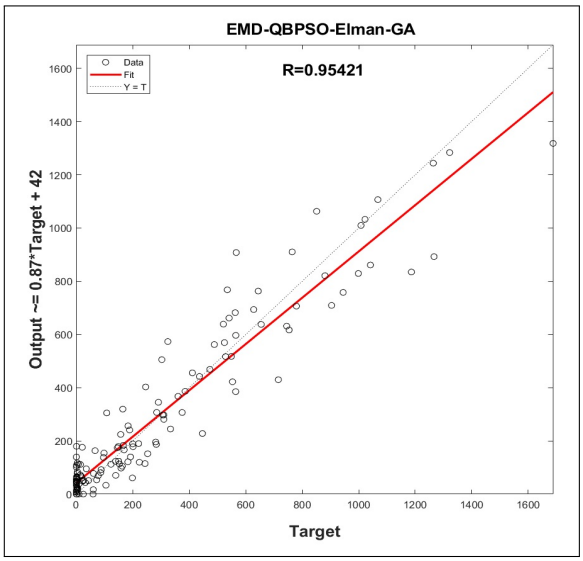
The new developed model, EMD-QBPSO-Elman-GA, was applied in Sotavento wind farm using both observational and WRF data as input dataset. Evaluation results are presented in table 7.3. RMSE and MAE statistical metrics are used to compare the results for 1-, 3-, 6-, 12-, and 24-hour ahead wind power prediction of the developed model against persistence model (PM), MLP-BR (described in Chapter 5, for wind speed prediction), QBPSO-Elman and QBPSO-Elman-GA models. As it can be seen, the developed model outperforms all the other models with RMSE between 105.78, and 414.47 (kWh) from 1 to 24 hours ahead, MAE results for the same time horizons are from 71.96 to 321.22 (kWh). Significant improvements were obtained compared to the PM model, with 61.85%, 50.01%, 41.62%, 47.27%, and 57.56% for RMSE for the corresponding 1- to 24-hours ahead time intervals, respectively. Ans adding GA has improved the QBPSO-Elman RMSE results between 4% and 9% for up to 24-hour ahead. Using the EMD technique to the QBPSO-Elman-GA model has improved the RMSE by 21% for 1-hour ahead prediction.

Unfortunately, these results cannot be compared to other results in the literature, because of different reasons: 1) in some works [194, 216, 317, 160, 215], the evaluation criteria are different; 2) in other works [143, 318], normalized data are used to produce RMSE and MAE, which does not a give real sense of the model accuracy; 3) none of those previous works include any comparison with PM model, which could be considered as a good criteria for models inter-comparison, based on their improvement percentage compared to PM.

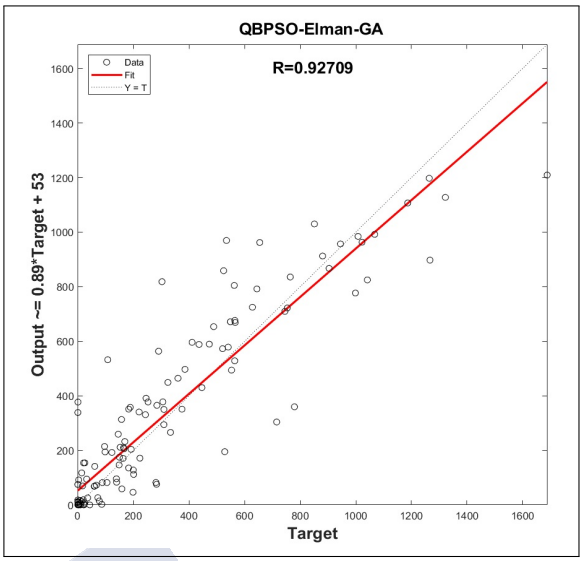
Table 7.3 Comparing WP forecasting accuracy of different tested models using RMSE and MAE evaluation criteria (EC) for 1, 3, 6, 12, and 24 hours ahead, with Sotavento dataset

	EC	Models	Time Horizon				
			1-hour	3-hour	6-hour	12-hour	24-hour
Sotavento, Spain	RMSE	PM	277.27	328.18	305.58	409.63	977.43
		MLP-BR	137.29	188.20	216.17	292.19	459.58
		QBPSO-ELMAN	140.10	201.71	218.27	339.31	502.18
		QBPSO-ELMAN-GA	135.28	181.47	212.93	257.47	458.73
		EMD-QBPSO-ELMAN-GA	**105.78**	**163.98**	**187.41**	**215.99**	**414.74**
	MAE	PM	130.27	328.18	305.58	390.52	977.44
		MLP-BR	85.52	127.98	149.11	203.47	341.68
		QBPSO-ELMAN	79.36	134.48	150.80	241.15	338.50
		QBPSO-ELMAN-GA	76.61	125.96	146.78	179.66	347.59
		EMD-QBPSO-ELMAN-GA	**71.96**	**115.39**	**128.50**	**161.70**	**321.22**

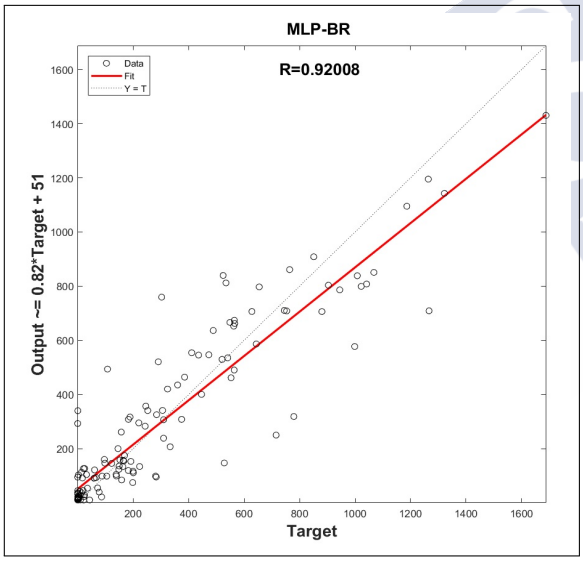
In figure 7.11 R-squared plots of the different new developed models (including the complete EMD-QBPSO-ELMAN-GA model), MLP-BR, and PM are shown. The complete model has the best correlation coefficient, with the value of 0.95. This coefficient values for the QBPSO-Elman-GA, MLP-BR, and PM models are 0.93, 0.92, and 0.75, respectively. Figure 7.12 shows error histograms of the new developed complete model and MLP-BR model for 1-hour ahead wind power prediction. Also, figure 7.13 to 7.15 shows wind power time series, both predicted with all the different tested models, and observed, at Sotavento wind farm, using 1-hour, 3-hours, and 6-hours of prediction ahead time intervals. The complete new developed models follows the original observed wind power using 1-hour ahead time interval, while other models show some differences in local extremes, along the time series. However, differences with original (observed) time series increase using 3-hours and 6-hours ahead, and also complete model errrors at some local extremes (specially, peaks) are observed.



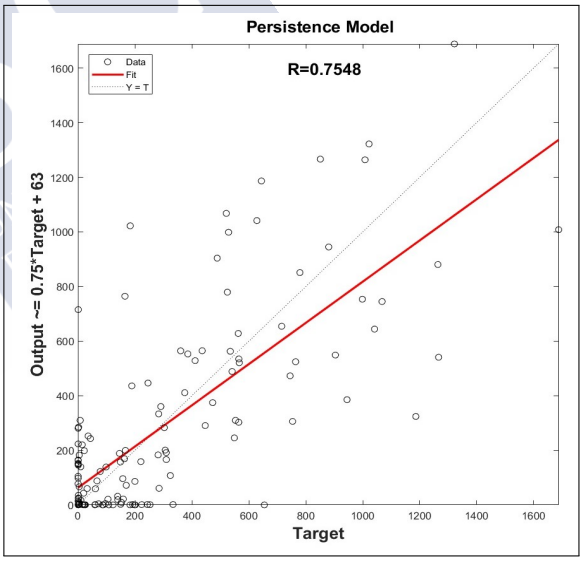
(a) EMD-QBPSO-Elman-GA Model



(b) QBPSO-Elman-GA Model

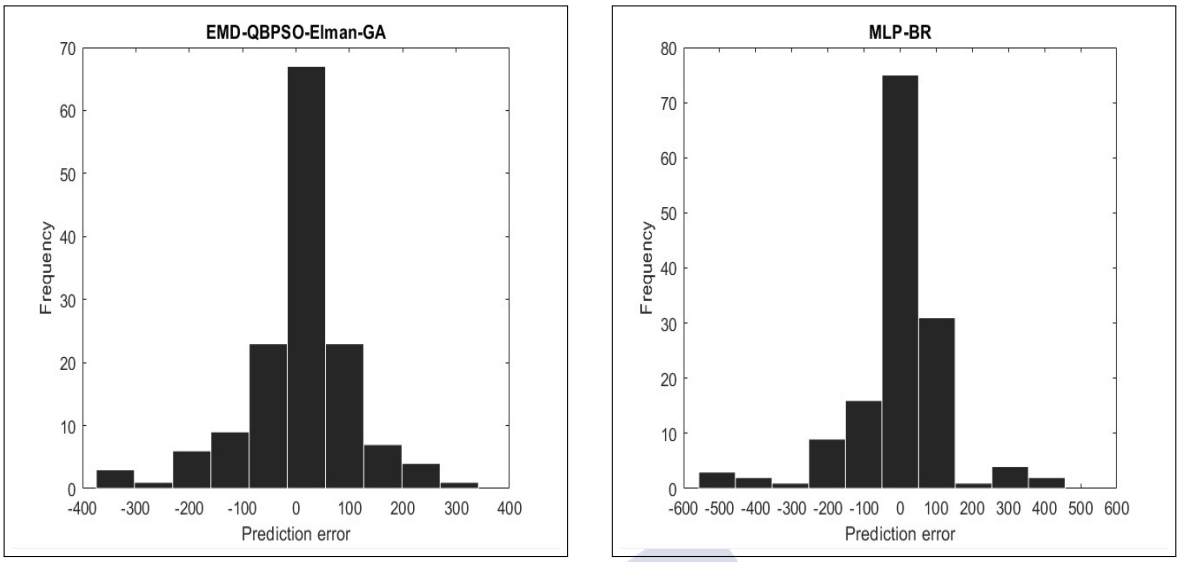


(c) MLP-BR Model



(d) Persistence Model

Fig. 7.11 R-squared values for the developed and models applied on hourly-based observational and WRF data from Sotavento wind farm, 2016



(a) EMD-QBPSO-Elman-GA Model (b) MLP-BR Model

Fig. 7.12 Residual error histograms of the developed and MLP-BR models tested with Sotavento dataset, 2016

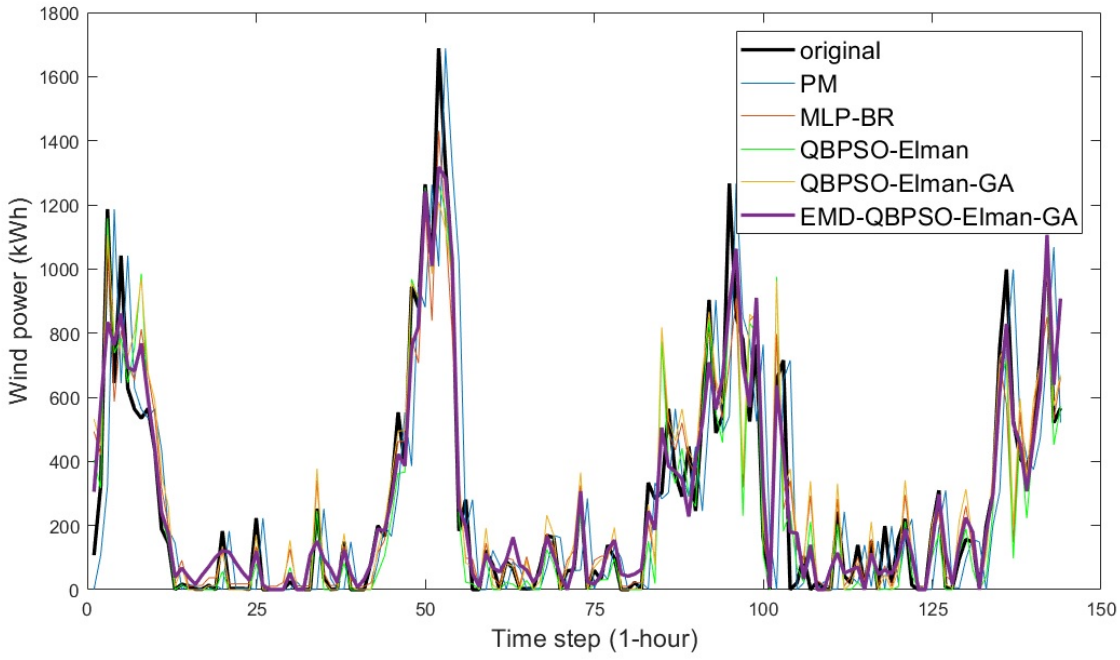


Fig. 7.13 1-hour ahead wind power prediction from developed and other models, compared to original observed, using Sotavento dataset, 2016

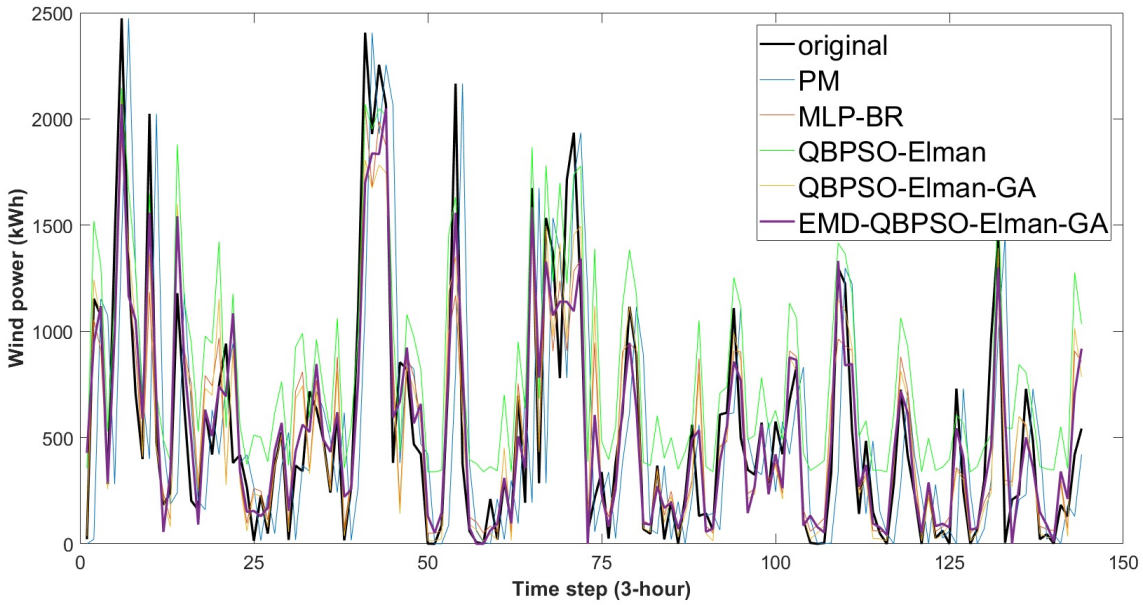


Fig. 7.14 3-hour ahead wind power prediction from developed and other models, compared to original observed, using Sotavento dataset, 2016

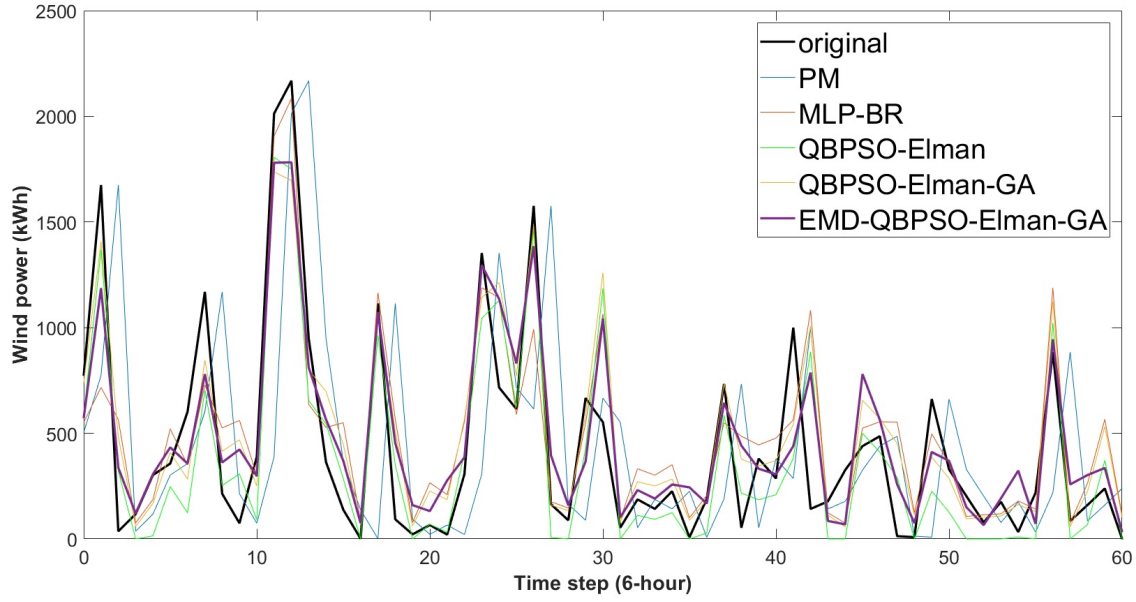


Fig. 7.15 6-hour ahead wind power prediction from developed and other models, compared to original observed, using Sotavento dataset, 2016

7.3.5.1 New wind power prediction model testing at La Haute Borne wind farm

In La Haute Borne wind farm only observational data are available, such as wind power, wind speed, wind direction, and temperature. So, the feature space is formed using these four parameters with up to 48 previous time steps. Again, the best combination was the simplest one with only the previous step of wind power. Figure 7.16 shows the prediction results from different models.

Table 7.4 Comparing accuracy of new developed models and other models for 10-minute ahead wind power prediction in two wind turbines at La Haute Borne wind farm. Best resutls in bold text

Site	Turbine	Model	Evaluation Criteria RMSE	MAE
La Haute Borne	R80711	PM	268.43	268.43
		MLP-BR	103.30	59.23
		QBPSO-Elman	146.74	107.06
		QBPSO-Elman-GA	94.04	55.67
		EMD-QBPSO-GA-Elman	**89.46**	**56.42**
	R80790	PM	207.42	207.42
		MLP-BR	114.41	60.18
		QBPSO-Elman	134.40	92.62
		QBPSO-Elman-GA	112.15	61.42
		EMD-QBPSO-GA-Elman	**104.81**	**53.54**

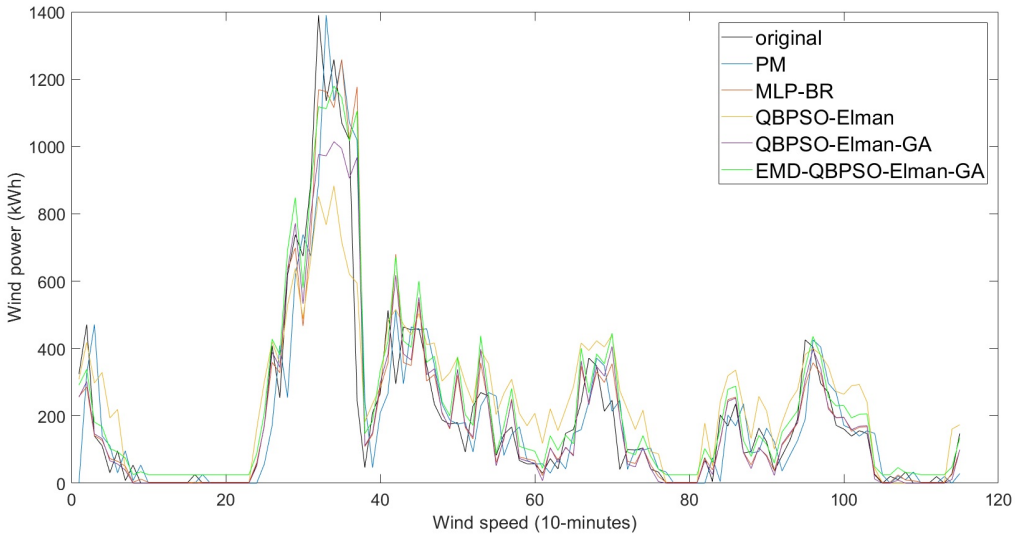


Fig. 7.16 10-minutes ahead wind power prediction from developed and other models, compared to original observed, using La Haute Borne dataset, March 2017

7.4 Summary

Regarding the use of ML techniques in wind power prediction, besides the structure of the network and learning methods, feature selection is a crucial issue in supervised learning, which directly affects the accuracy of the results [216]. In addition, hybrid and combined models are proved to be more successful in wind speed, and power prediction [299]. Considering these two facts, a new hybrid model was developed and successfully tested on two wind farms. The components of this new developed model are: Empirical Mode Decomposition (EMD), Quantum-based Binary Particle Swarm Optimization (QBPSO), Elman Neural Network (ENN), and Genetic Algorithm (GA). EMD was used to tackle with non-linearity of the weather time series and improved the results up to 10% in terms of RMSE and MAE. New developed QBPSO feature selection technique was used to find the best set of the input variables (among $3 \times 2^{48} + 5 \times 2^{96}$ possible input sets composed composed by both observations and numerical weather predictions). The simplest set of the previous time step of wind power variable outperformed the other sets; in agreement with similar wind speed prediction results described on Chapter 6. In addition, GA was used to train the ENN weights, improving model results compared to the original Elman-BP technique.

Chapter 8

Concluding remarks

8.1 Conclusions

As the main goal in this study, a coherent path to develop and evolve wind speed and power prediction models was done by addressing data validation, and reconstruction, hyper-parameters assignment, and feature selection problems in neural networks, as three critical issues need to be considered in a ML-based prediction model. In addition, prediction models developing followed a continuous improvement line, testing more specific and complex techniques, from a most widely used type of neural networks, FFNN, until an optimized MLP model, to a hybrid Elman neural network model.

Previous to any time series application to developing a ML wind speed prediction model, a comprehensive data analysis was required, in order to detect the main time series features and their drawnbacks. Also, from this analysis the necessity of time series reconstruction was established.

Therefore, a new data reconstruction model was developed to impute missing and invalid data. This model is based on feed-forward neural network trained with the Bayesian Regularization method, and equipped with a k-means clustering algorithm. Compared to other benchmark models, like WRF numerical weather prediction, MA, and ARIMA, the new reconstruction model achieved better results.

Then, a new wind speed BBO-MLP prediction model was developed and trainned with the reconstructed time series from four case studies: Three Galician standard meteorological stations, and one US station (M2-tower) with high frequency observations at several heights. In this model a semi-exhaustive search algorithm is applied simultaneously on three

different sets of model-related, optimization-related, and input data-related hyperparameters. BBO-MLP results at the four case studies were compared to other benchmark models using Pareto evaluation methodology, showing that the new models achieved better statistics that any other model. Also, the new model statistics values are in either similar or better than other recent models results in the literature.

The development of a new wind speed prediction model was a very good basis for developing a new wind power prediction model. While it could be possible to derive wind power from wind speed, taking into account specific wind farm characteristics, in this work a direct approach was adopted. That is, wind power time series, jointly to weather data series, was applied to develop and train a new wind power prediction model. Taking the advantage of the previous experience in developing a wind speed prediction model based on ML-techniques. The new developed wind power prediction model, namely EMD-QBPSO-ENN-GA, is a hybrid recurrent neural network model, addressing the feature selection problem. Optimized Elman neural network, using a genetic algorithm as a fast optimization technique, is combined with a newly developed quantum-based binary particle swarm optimization (QBPO) algorithm and equipped with an empirical mode decomposition (EMD) technique as a signal processing method. As input dataset both meteorological and wind power data are combined. Particularly, EMD helps to deal with the unsteady behaviour of wind speed, using cluster input signal to produce several intrinsic mode functions with similar characteristics. Then, different GA-ENN models were trained for each cluster. And, as an important step for achieving good results in feasible computing time, a new hygrid optimization technique, namely QBPSO, was developed to select the minimum required input dataset.

This new developed proposed hybrid model, EMD-QBPO-ENN-GA, was applied and tested for wind power prediction at two different case studies: (a) Sotavento wind farm at Galicia, Spain, using as input dataset both meteorological and wind power observations, and WRF numerical weather predictions; and, (b) La Haute Borne wind farm at France, using meteoroological and wind power observations. New model results were compared to other models, also with reduced versions of the new model, showing that the new model improved wind power prediction at both wind farms; and, also, proving the relevance of each of the new model components in improving the prediction.

8.2 Future work

Considering the deeply review of the state-of-art in ML techniques applied to wind speed and wind power prediction, and the successful results obtained with the new developed models, some recommendations for future work are provided in a top to down approach in a way that broader ideas are provided first, followed by the more detailed areas to work and suggestions.

All the developed and applied models in this study are categorized as shallow neural networks, which perform accurately using a short to medium length time series of either observational or simulated data from meteorological weather stations and physical models. More than these two data sources, there are other types of weather data, including Radar, Lidar, satellite images, regular street camera images. These diverse weather data sources provide a vast potential if they can be applied in a complementary context using deep learning and big data technologies. More information inclusion means providing more accurate wind speed and power predictions if this enormous amount of data can be handled and interpreted in a machine-understanding way. Hence, as the first valuable future work, using different weather and climate data sources in addition to big data technologies, such as Hadoop and Spark [319, 236] is highly recommended.

Developing a neural network for either wind speed or power prediction modeling requires running many nominated models that are different in structure and parameters to determine the best model. It might impose a high cost in terms of computing time which derived in unfeasible solutions to be run on a personal computer. As up to date solution, cloud computing has a high potential to address this problem. Microsoft Azure, Amazon Web Services (AWS), and Oracle Cloud are among the most popular cloud platforms providing helpful tools such as parallel computing, high-performance computing (HPC), big data technologies, etc. These tools, remarkably parallel computing, allows running a heavy comparative model, as BBO-MLP, simultaneously on several personal computers. Hence the computation time could decrease significantly. As an exciting area of work concerning computing time, cloud computing is highly recommended.

Additionally, Quantum Computing (QC), still under development, is an exciting area considered in any big data problem. While shallow neural networks do not take direct advantage on quantum computing, if large input datasets from different sources are used, QC can help to manage them. Although current deep and shallow binary coding prediction models

are achieved satisfactory accuracy levels, the grid networks are getting more complex and volatile renewable energy sources are increasingly also uncertainties in the electric energy availability arise, so more sophisticated power prediction models will be required.

In addition, physical weather forecast models, as WRF, can take advantage on Quantum Computing, increasing their grids resolutions and physical schemes complexity with feasible computing times.

Finally, wind speed and power prediction models as based in two different knowledge areas: Computer Science, including mathematical, statistical, and programming issues; and Meteorology. However, most of the efforts evolved until now are mainly focused on computer science. However, involving weather and climate knowledge into the model (not only numerical, but also human experience) and developing physical-machine learning combined models may result in more accurate prediction results, as it was briefly tested in chapter chapter 5. New models should focus on the meteorological side required to include wind share, wind turbulence, boundary layer conditions, terrain characteristics, etc., into the wind speed and power prediction models. Also, there is a research gap in developing NWP-ML combined and hybrid models, particularly for long-term wind speed and power forecasting.

Both models and methodologies applied in this work are in the line of the state-of-art. However, the combination of many other methods and methodologies is also possible. Hence, other novel optimization algorithms, signal prepossessing techniques, and machine learning methods, particularly in a combined and hybrid context, are worthwhile to be investigated in a comprehensive comparative procedure with other methods to illustrate whether or not a new developed technique is successful.

References

[1] R. B. Johnston, "Arsenic and the 2030 Agenda for sustainable development," *Arsenic Research and Global Sustainability - Proceedings of the 6th International Congress on Arsenic in the Environment, AS 2016*, pp. 12–14, 2016.

[2] *The 21st Conference of the Parties to the United Nations Framework, Convention on Climate Change (UNFCCC). The Paris Agreement*. 2015.

[3] M. D. Esteban, J. J. Diez, J. S. López, and V. Negro, "Why offshore wind energy?," *Renewable Energy*, vol. 36, no. 2, pp. 444–450, 2011.

[4] "International energy agancy." http://www.iea.org.

[5] H. Ritchie, "Renewable energy," *Our World in Data*, 2017. https://ourworldindata.org/renewable-energy.

[6] Martha Ekkert, *Global energy transformation: a roadmap to 2050*. 2018.

[7] E. Sesto and N. H. Lipman, "Wind energy in Europe," 2020.

[8] Reve, "Renewables 2020 global status report." https://www.ren21.net, July 2020.

[9] J. Weber, M. Reyers, C. Beck, M. Timme, J. G. Pinto, D. Witthaut, and B. Schäfer, "Wind Power Persistence Characterized by Superstatistics," *Scientific Reports*, vol. 9, no. 1, pp. 1–15, 2019.

[10] A. J. Zapata-Sierra, A. Cama-Pinto, F. G. Montoya, A. Alcayde, and F. Manzano-Agugliaro, "Wind missing data arrangement using wavelet based techniques for getting maximum likelihood," *Energy Conversion and Management*, vol. 185, no. January, pp. 552–561, 2019.

[11] H. Liu, C. Chen, H. Q. Tian, and Y. F. Li, "A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks," *Renewable Energy*, vol. 48, pp. 545–556, 2012.

[12] S. Li, P. Wang, and L. Goel, "Wind Power Forecasting Using Neural Network Ensembles with Feature Selection," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1447–1456, 2015.

[13] Y. Chen, Z. Dong, Y. Wang, J. Su, Z. Han, D. Zhou, K. Zhang, Y. Zhao, and Y. Bao, "Short-term wind speed predicting framework based on EEMD-GA-LSTM method under large scaled wind history," *Energy Conversion and Management*, vol. 227, no. August 2020, p. 113559, 2021.

[14] H. Cai, X. Jia, J. Feng, W. Li, Y. M. Hsu, and J. Lee, "Gaussian Process Regression for numerical wind speed prediction enhancement," *Renewable Energy*, vol. 146, pp. 2112–2123, 2020.

[15] J. Hu and J. Wang, "Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression," *Energy*, vol. 93, pp. 1456–1466, 2015.

[16] Y. Zhang, K. Liu, L. Qin, and X. An, "Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods," *Energy Conversion and Management*, vol. 112, pp. 208–219, 2016.

[17] Z. Liang, J. Liang, C. Wang, X. Dong, and X. Miao, "Short-term wind power combined forecasting based on error forecast correction," *Energy Conversion and Management*, vol. 119, pp. 215–226, 2016.

[18] A. E. Saleh, M. S. Moustafa, K. M. Abo-Al-Ez, and A. A. Abdullah, "A hybrid neuro-fuzzy power prediction system for wind energy generation," *International Journal of Electrical Power and Energy Systems*, vol. 74, pp. 384–395, 2016.

[19] A. Abraham, "Adaptation of Fuzzy Inference System Using Neural Learning," vol. 83, pp. 53–83, 2005.

[20] M. Yu, "Short-term wind speed forecasting based on random forest model combining ensemble empirical mode decomposition and improved harmony search algorithm," *International Journal of Green Energy*, vol. 17, no. 5, pp. 332–348, 2020.

[21] S. S. Eide, J. B. Bremnes, and I. Steinsland, "Bayesian model averaging for wind speed ensemble forecasts using wind speed and direction," *Weather and Forecasting*, vol. 32, no. 6, pp. 2217–2227, 2017.

[22] X. W. Ye, Y. Ding, and H. P. Wan, "Probabilistic forecast of wind speed based on Bayesian emulator using monitoring data," *Structural Control and Health Monitoring*, vol. 28, no. 1, pp. 1–14, 2021.

[23] W. Li, X. Jia, X. Li, Y. Wang, and J. Lee, "A Markov model for short term wind speed prediction by integrating the wind acceleration information," *Renewable Energy*, vol. 164, pp. 242–253, 2021.

[24] Z. Lu, S. Lu, M. Xu, and B. Cui, "A robust stochastic stability analysis approach for power system considering wind speed prediction error based on Markov model," *Computer Standards and Interfaces*, vol. 75, p. 103503, 2021.

[25] H. Bouzgou and N. Benoudjit, "Multiple architecture system for wind speed prediction," *Applied Energy*, vol. 88, no. 7, pp. 2463–2471, 2011.

[26] S. Emeis, *Green Energy and Technology Wind Energy Meteorology Atmospheric Physics for Wind Power Generation Second Edition*. 2018.

[27] S. S. Kutty, M. G. Khan, and M. R. Ahmed, "Estimation of different wind characteristics parameters and accurate wind resource assessment for Kadavu, Fiji," *AIMS Energy*, vol. 7, no. 6, pp. 760–791, 2019.

[28] R. Wesonga, "On multivariate imputation and forecasting of decadal wind speed missing data," *SpringerPlus*, vol. 4, no. 1, 2015.

[29] X. Li, L. Comerford, and A. Plater, "Utilizing sparse representations of wind data for reconstruction of historical records with gaps," no. June, p. 2016, 2016.

[30] Z. Ma, H. Chen, J. Wang, X. Yang, R. Yan, J. Jia, and W. Xu, "Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction," *Energy Conversion and Management*, vol. 205, no. September 2019, p. 112345, 2020.

[31] J. Z. Wang, Y. Wang, and P. Jiang, "The study and application of a novel hybrid forecasting model - A case study of wind speed forecasting in China," *Applied Energy*, vol. 143, no. 2015, pp. 472–488, 2015.

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

[33] J. G. Lockwood, *ATMOSPHERIC CIRCULATION, GLOBALAtmospheric circulation, global*, pp. 131–140. Boston, MA: Springer US, 1987.

[34] R. B. Stull, "An introduction to boundary layer meteorology," *An introduction to boundary layer meteorology*, 1988.

[35] M. T. Pierre Sagaut, Sebastien Deck, *Multiscale and Multiresolution Approaches in Turbulence - Les, Des and Hybrid Rans/Les Methods: Applications and Guidelines (2nd Edition)*. IMPERIAL COLLEGE PRESS; 2nd Revised ed. edition, 2013.

[36] F. Li, G. Ren, and J. Lee, "Multi-step wind speed prediction based on turbulence intensity and hybrid deep neural networks," *Energy Conversion and Management*, vol. 186, no. January, pp. 306–322, 2019.

[37] I. Suomi, "Wind Gust Measurement Techniques — From Traditional Anemometry to New Possibilities," no. April, pp. 1–27, 2018.

[38] *Guide to Climatological Practices*. No. 100, World meteorological organization, 2018.

[39] S. Emeis, "Wind Energy Meteorology - Second Edition," no. April, 2018.

[40] E. Hua, "Wind turbines," 2066.

[41] O. ZEMAN and N. O. Jensen, "Modification of turbulence characteristics in flow over hills," *J. Roy. Meteorol. Soc. 113, 55–80*, pp. 55–80, 1987.

[42] Z. Ti, X. Wei, and M. Zhang, "Arti fi cial Neural Networks based wake model for power prediction of wind farm," *Renewable Energy*, vol. 172, pp. 618–631, 2021.

[43] F. Eldridge, *Wind machines*. New York: Litton Educational Publishing, Inc. p. 15, 1980.

[44] D. G. Shepherd, *Historical development of the windmill*. 1990.

[45] T. J. Price, "James Blyth – Britain ' s first modern wind power pioneer," pp. 191–200, 1984.

[46] R. Righter, *Wind Energy in America: A History*. University of Oklahoma Press, 1996.

[47] G. o. I. . Centre of Wind Energy Technology, "Wind : Origin and Local Effects," pp. 7–16, 2007.

[48] F. Blaabjerg and K. Ma, "Wind Energy Systems," *Proceedings of the IEEE*, vol. 105, no. 11, pp. 2116–2131, 2017.

[49] W. Frost and C. F. Shieh, "Wind Characteristics Over Complex Terrain Relative to WECS Siting," vol. 5, no. 5, 1981.

[50] H. Cetinay, F. A. Kuipers, and A. N. Guven, "Optimal siting and sizing of wind farms," *Renewable Energy*, vol. 101, pp. 51–58, 2017.

[51] K. S. R. Murthy and O. P. Rahi, "A comprehensive review of wind resource assessment," *Renewable and Sustainable Energy Reviews*, no. July, pp. 0–1, 2016.

[52] G. Li and J. Shi, "Applications of Bayesian methods in wind energy conversion systems," *Renewable Energy*, vol. 43, pp. 1–8, 2012.

[53] T. Bakri and P. Jackson, "Statistical and synoptic analyses of offshore wind variations," *International Journal of Climatology*, vol. 39, no. 7, pp. 3201–3217, 2019.

[54] P. Gipe, "9 - design as if people matter: Aesthetic guidelines for a wind power future," in *Wind Power in View* (M. J. Pasqualetti, P. Gipe, and R. W. Righter, eds.), Sustainable World, pp. 173–212, San Diego: Academic Press, 2002.

[55] *Top Threats to Birds (U.S. only. Ordered by Median Estimate of Bird Mortality Annually. As of 2017.)*. United States Fish and Wildlife Service, 2017.

[56] R. May, T. Nygård, U. Falkdalen, J. Åström, Ø. Hamre, and B. G. Stokke, "Paint it black : Efficacy of increased wind turbine rotor blade visibility to reduce avian fatalities," no. February, pp. 8927–8935, 2020.

[57] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 762–777, 2014.

[58] S. Pindado, "Deviation of Cup and Propeller Anemometer Calibration Results with Air Density," pp. 683–701, 2012.

[59] M. Variables, *Guide to Instruments and Methods of Observation*, vol. I. World meteorological organization, volume i – ed., 2018.

[60] D. Porter and E. Producers, *Comprehensive renewable energy*, vol. 1. Association of American, 2012.

[61] J. Schneemann, D. Trabucchi, J. J. Trujillo, and M. Kuhn, "Comparing measurements of the horizontal wind speed of a 2D Multi-Lidar and a cup anemometer Comparing measurements of the horizontal wind speed of a 2D Multi-Lidar and a cup anemometer," *Journal of Physics: Conference Series 555*, 2014.

[62] R. M. BANTA, W. BREWER, S. P. SANDBERG, and R. M. HARDESTY, "Doppler Lidar – Based Wind-Profile Measurement System for Offshore Wind-Energy and Other Marine Boundary Layer Applications," pp. 327–349, 2012.

[63] "Aircraft Meteorological Data Relay (AMDAR) Reference Manual," *Secretariat of the World Meteorological Organization — Geneva —Switzerland*, no. 958, 2003.

[64] A. Nagy and I. Jahn, "Advanced Data Acquisition System for Wind Energy Applications," pp. 1–7, 2017.

[65] *UCAR. NWP model fundamentals. Meteorological Education. University Corporation for Atmospheric Research. https://www.meted.ucar.edu/ [online].*

[66] G. J. Haltiner, *Numerical weather prediction*. Wiley: New York, 1971.

[67] S. Al-yahyai, Y. Charabi, and A. Gastli, "Review of the use of Numerical Weather Prediction (NWP) Models for wind energy assessment," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 9, pp. 3192–3198, 2010.

[68] C. Ren, N. An, J. Wang, L. Li, B. Hu, and D. Shang, "Optimal parameters selection for BP neural network based on particle swarm optimization: A case study of wind speed forecasting," *Knowledge-Based Systems*, vol. 56, pp. 226–239, 2014.

[69] E. Erdem and J. Shi, "ARMA based approaches for forecasting the tuple of wind speed and direction," *Applied Energy*, vol. 88, no. 4, pp. 1405–1414, 2011.

[70] M.-d. Wang, Q.-r. Qiu, and B.-w. Cui, "SHORT-TERM WIND SPEED FORECASTING COMBINED TIME SERIES," pp. 15–17, 2012.

[71] J. Wang and J. 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 usi," *Energy*, vol. 93, pp. 41–56, 2015.

[72] R. Ata, "Artificial neural networks applications in wind energy systems: a review," *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 534–562, 2015.

[73] R. L. Welch, S. M. Ruffing, and G. K. Venayagamoorthy, "Comparison of feedforward and feedback neural network architectures for short term wind speed prediction',," pp. 3335–3340, 2009.

[74] P. P. Pradhan and B. Subudhi, "Wind speed forecasting based on wavelet transformation and recurrent neural network," no. July, pp. 1–11, 2019.

[75] K. Trebing and S. Mehrkanoon, "Wind speed prediction using multidimensional convolutional neural networks,"

[76] L. Xu and J. Mao, "Short-Term Wind Power Forecasting Based on Elman Neural Network with Particle Swarm Optimization," no. 2, pp. 2678–2681, 2016.

[77] Z. H. Guo, J. Wu, H. Y. Lu, and J. Z. Wang, "A case study on a hybrid wind speed forecasting method using BP neural network," *Knowledge-Based Systems*, vol. 24, no. 7, pp. 1048–1056, 2011.

[78] B. Chen, L. Zhao, X. Wang, J. H. Lu, G. Y. Liu, R. F. Cao, and J. B. Liu, "Wind speed prediction using OLS algorithm based on RBF neural network," *Asia-Pacific Power and Energy Engineering Conference, APPEEC*, pp. 9–12, 2009.

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

[80] S. Hanifi, X. Liu, Z. Lin, and S. Lotfian, "A Critical Review of Wind Power Forecasting," pp. 1–24, 2020.

[81] P. R. Kamath and K. Senapati, "Short-term wind speed forecasting using S-transform with compactly supported kernel," no. April 2020, pp. 260–274, 2021.

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

[83] A. Bagheri, H. Mohammadi Peyhani, and M. Akbari, "Financial forecasting using ANFIS networks with Quantum-behaved Particle Swarm Optimization," *Expert Systems with Applications*, vol. 41, no. 14, pp. 6235–6250, 2014.

[84] H. Liu, H. Q. Tian, and Y. F. Li, "Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms," *Energy Conversion and Management*, vol. 100, pp. 16–22, 2015.

[85] J. Wang, H. Jiang, B. Han, and Q. Zhou, "An experimental investigation of FNN model for wind speed forecasting using EEMD and CS," *Mathematical Problems in Engineering*, vol. 2015, 2015.

[86] "Observatorio eólico de galicia." http://observatorio.eolico.webs.uvigo.es/?lang=es.

[87] "Meteogalicia." https://www.meteogalicia.gal/web/inicio.action.

[88] "Sotavento company." http://www.sotaventogalicia.com.

[89] "National renewable energy labratory (nrel)." https://www.nrel.gov/wind.

[90] N. M. Khan, G. M. Khan, and P. Matthews, *AI Based Real-Time Signal Reconstruction for Wind Farm with SCADA Sensor Failure*, vol. 1. Springer International Publishing, 2020.

[91] M. Gomez-Gesteira, C. Moreira, I. Alvarez, and M. de Castro, "Ekman transport along the Galician coast (northwest Spain) calculated from forecasted winds," *Journal of Geophysical Research: Oceans*, vol. 111, no. 10, 2006.

[92] J. O. Blanton, K. R. Tenore, F. Castillejo, L. P. Atkinson, F. B. Schwing, and A. Lavin, "Relationship of Upwelling To Mussel Production in the Rias on the Western Coast of Spain.," *Journal of Marine Research*, vol. 45, no. 2, pp. 497–511, 1987.

[93] C. Jung and D. Schindler, "Global comparison of the goodness-of-fit of wind speed distributions," *Energy Conversion and Management*, vol. 133, pp. 216–234, 2017.

[94] P. P. Revheim, *Improving and enhancing NWP based wind power forecasts under Norwegian conditions*. 2015.

[95] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. Prem Kumar, "A comprehensive review on wind turbine power curve modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 30, no. October 2010, pp. 452–460, 2014.

[96] I. R. Young, E. Sanina, and A. V. Babanin, "Calibration and cross validation of a global wind and wave database of altimeter, radiometer, and scatterometer measurements," *Journal of Atmospheric and Oceanic Technology*, vol. 34, no. 6, pp. 1285–1306, 2017.

[97] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra," *Digital Signal Processing*, no. 2, pp. 532–574, 1975.

[98] "Discoverydv 1.2." https://pypi.org/project/DiscoveryDV/, September 2019.

[99] R. Malhotra, "A systematic review of machine learning techniques for software fault prediction," *Applied Soft Computing Journal*, vol. 27, pp. 504–518, 2015.

[100] D. Lin, A. V. Vasilakos, Y. Tang, and Y. Yao, "Neural networks for computer-aided diagnosis in medicine: A review," *Neurocomputing*, vol. 216, pp. 700–708, 2016.

[101] R. Tagliaferri, G. Longo, L. Milano, F. Acernese, F. Barone, A. Ciaramella, R. De Rosa, C. Donalek, A. Eleuteri, G. Raiconi, S. Sessa, A. Staiano, and A. Volpicelli, "Neural networks in astronomy," *Neural Networks*, vol. 16, no. 3-4, pp. 297–319, 2003.

[102] M. J. Misinterpreting, M. Journal, S. S. Dhaliwal, and M. J. Campbell, "This is a repository copy of Misinterpreting p-values in research . White Rose Research Online URL for this paper : Article : Misinterpreting P-Values In Research," 2010.

[103] A. Amani and D. Mohammadyani, "Artificial Neural Networks: Applications in Nanotechnology," *Artificial Neural Networks - Application*, no. March 2014, 2011.

[104] S. A. Kalogirou, "Applications of artificial neural-networks for energy systems," *Applied Energy*, vol. 67, no. 1-2, pp. 17–35, 2000.

[105] A. K. Yadav and S. S. Chandel, "Solar radiation prediction using Artificial Neural Network techniques: A review," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 772–781, 2014.

[106] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44–55, 2001.

[107] M. Negnevitsky, P. Mandal, and A. K. Srivastava, "Machine learning applications for load, price and wind power prediction in power systems," *2009 15th International Conference on Intelligent System Applications to Power Systems, ISAP '09*, 2009.

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

[109] I. Colak, S. Sagiroglu, and M. Yesilbudak, "Data mining and wind power prediction: A literature review," *Renewable Energy*, vol. 46, pp. 241–247, 2012.

[110] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, 2014.

[111] J. Wang, Y. Song, F. Liu, and R. Hou, "Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models," *Renewable and Sustainable Energy Reviews*, vol. 60, pp. 960–981, 2016.

[112] W. L. Theo, J. S. Lim, W. S. Ho, H. Hashim, and C. T. Lee, "Review of distributed generation (DG) system planning and optimisation techniques: Comparison of numerical and mathematical modelling methods," *Renewable and Sustainable Energy Reviews*, vol. 67, pp. 531–573, 2017.

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

[114] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 4, pp. 915–920, 2009.

[115] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 762–777, 2014.

[116] W. S. Mcculloch and W. Pitts, "A logical calculus nervous activity," *Bulletin of Mathematical Biology*, vol. 52, no. l, pp. 99–115, 1990.

[117] K. Gnana Sheela and S. N. Deepa, "Neural network based hybrid computing model for wind speed prediction," *Neurocomputing*, vol. 122, pp. 425–429, 2013.

[118] Y. Bengio, *Learning deep architectures for AI*, vol. 2. 2009.

[119] S. Hochreiter, "Long Short-Term Memory," vol. 1780, pp. 1735–1780, 1997.

[120] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.

[121] M. Martinez-Luengo, A. Kolios, and L. Wang, "Structural health monitoring of offshore wind turbines: A review through the Statistical Pattern Recognition Paradigm," *Renewable and Sustainable Energy Reviews*, vol. 64, pp. 91–105, 2016.

[122] A. D. Almási, S. Woźniak, V. Cristea, Y. Leblebici, and T. Engbersen, "Review of advances in neural networks: Neural design technology stack," *Neurocomputing*, vol. 174, pp. 31–41, 2016.

[123] G. H. J. Ron Kohavi, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 7920 LNCS, no. 97, pp. 654–678, 1997.

[124] S. Salcedo-Sanz, A. Pastor-Sánchez, L. Prieto, A. Blanco-Aguilera, and R. García-Herrera, "Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization - Extreme learning machine approach," *Energy Conversion and Management*, vol. 87, pp. 10–18, 2014.

[125] A. L. Blum and P. Langley, "Artificial Intelligence Selection of relevant features and examples in machine," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 245–271, 1997.

[126] H. Liu and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," *Knowledge Creation Diffusion Utilization*, vol. 17, no. 4, pp. 491–502, 2005.

[127] E. Cadenas and W. Rivera, "Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks," *Renewable Energy*, vol. 34, no. 1, pp. 274–278, 2009.

[128] C. Hervás-Martínez, S. Salcedo-Sanz, P. A. Gutiérrez, E. G. Ortiz-García, and L. Prieto, "Evolutionary product unit neural networks for short-term wind speed forecasting in wind farms," *Neural Computing and Applications*, vol. 21, no. 5, pp. 993–1005, 2012.

[129] T. C. Akinci, "Short term wind speed forecasting with ANN in Batman, Turkey," *Elektronika ir Elektrotechnika*, no. 1, pp. 41–45, 2011.

[130] R. J. Bessa, V. Miranda, and J. Gama, "Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting," *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1657–1666, 2009.

[131] M. C. Alexiadis, P. S. Dokopoulos, and H. S. Sahsamanoglou, "Wind speed and power forecasting based on spatial correlation models," *IEEE Transactions on Energy Conversion*, vol. 14, no. 3, pp. 836–842, 1999.

[132] U. B. Filik, "A New Hybrid Approach for Wind Speed Prediction Using Fast Block Least Mean Square Algorithm and Artificial Neural Network," *Mathematical Problems in Engineering*, vol. 2016, 2016.

[133] A. Agrawal and K. S. Sandhu, "Most influential parametrical and data needs for realistic wind speed prediction," *Renewable Energy*, vol. 94, pp. 452–465, 2016.

[134] P. Zhao, J. Wang, J. Xia, Y. Dai, Y. Sheng, and J. Yue, "Performance evaluation and accuracy enhancement of a day-ahead wind power forecasting system in China," *Renewable Energy*, vol. 43, pp. 234–241, 2012.

[135] T. Kitajima and T. Yasuno, "Output prediction of wind power generation system using complex-valued neural network," *Proceedings of the SICE Annual Conference*, pp. 3610–3613, 2010.

[136] S. Tasnim, A. Rahman, G. M. Shafiullah, A. M. T. Oo, and A. Stojcevski, "A time series ensemble method to predict wind power," *IEEE Symposium on Computational Intelligence Applications in Smart Grid, CIASG*, vol. 2015-Janua, no. January, 2015.

[137] O. B. Shukur and M. H. Lee, "Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA," *Renewable Energy*, vol. 76, pp. 637–647, 2015.

[138] S. A. Pourmousavi Kani and M. M. Ardehali, "Very short-term wind speed prediction: A new artificial neural network-Markov chain model," *Energy Conversion and Management*, vol. 52, no. 1, pp. 738–745, 2011.

[139] Z. Guo, W. Zhao, H. Lu, and J. Wang, "Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model," *Renewable Energy*, vol. 37, no. 1, pp. 241–249, 2012.

[140] M. A. Ghorbani, R. Khatibi, M. H. FazeliFard, L. Naghipour, and O. Makarynskyy, "Short-term wind speed predictions with machine learning techniques," *Meteorology and Atmospheric Physics*, vol. 128, no. 1, pp. 57–72, 2016.

[141] G. Grassi and P. Vecchio, "Wind energy prediction using a two-hidden layer neural network," *Communications in Nonlinear Science and Numerical Simulation*, vol. 15, no. 9, pp. 2262–2266, 2010.

[142] J. Wang, W. Zhang, J. Wang, T. Han, and L. Kong, "A novel hybrid approach for wind speed prediction," *Information Sciences*, vol. 273, pp. 304–318, 2014.

[143] D. Lee and R. Baldick, "Short-term wind power ensemble prediction based on gaussian processes and Neural networks," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 501–510, 2014.

[144] D. Niu, Y. Liang, and W. C. Hong, "Wind speed forecasting based on EMD and GRNN optimized by FOA," *Energies*, vol. 10, no. 12, 2017.

[145] G. Kumar and H. Malik, "Generalized Regression Neural Network Based Wind Speed Prediction Model for Western Region of India," *Procedia Computer Science*, vol. 93, no. September, pp. 26–32, 2016.

[146] E. E. Elattar, "Prediction of wind power based on evolutionary optimised local general regression neural network," *IET Generation, Transmission and Distribution*, vol. 8, no. 5, pp. 916–923, 2014.

[147] T. G. Barbounis and J. B. Theocharis, "Locally recurrent neural networks for long-term wind speed and power prediction," *Neurocomputing*, vol. 69, no. 4-6, pp. 466–496, 2006.

[148] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, and P. S. Dokopoulos, "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Transactions on Energy Conversion*, vol. 21, no. 1, pp. 273–284, 2006.

[149] S. L. Goh, M. Chen, D. H. Popović, K. Aihara, D. Obradovic, and D. P. Mandic, "Complex-valued forecasting of wind profile," *Renewable Energy*, vol. 31, no. 11, pp. 1733–1750, 2006.

[150] N. Amjady, F. Keynia, and H. Zareipour, "Short-term wind power forecasting using ridgelet neural network," *Electric Power Systems Research*, vol. 81, no. 12, pp. 2099–2107, 2011.

[151] A. Yona, T. Senjyu, N. Urasaki, and T. Funabshi, "Application of recurrent neural network to 3-hours-ahead generating power forecasting for wind power generators," *IEEJ Transactions on Power and Energy*, vol. 129, no. 5, pp. 1260–1265, 2009.

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

[153] H. Liu, H. Q. Tian, X. F. Liang, and Y. F. Li, "Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks," *Applied Energy*, vol. 157, pp. 183–194, 2015.

[154] J. Wang, S. Qin, Q. Zhou, and H. Jiang, "Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China," *Renewable Energy*, vol. 76, pp. 91–101, 2015.

[155] K. Liu, Y. Zhang, and L. Qin, "A novel combined forecasting model for short-term wind power based on ensemble empirical mode decomposition and optimal virtual prediction," *Journal of Renewable and Sustainable Energy*, vol. 8, no. 1, 2016.

[156] E. Cadenas, W. Rivera, R. Campos-Amezcua, and R. Cadenas, "Wind speed forecasting using the NARX model, case: La Mata, Oaxaca, México," *Neural Computing and Applications*, vol. 27, no. 8, pp. 2417–2428, 2016.

[157] V. Prema, B. S. Jnaneswar, C. A. Badarish, P. S. Ashok, S. Agarwal, and K. Uma Rao, "Novel training strategies for wavelet-neuro models for wind speed prediction," *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, vol. 2016-Janua, pp. 1–4, 2016.

[158] E. Cadenas, W. Rivera, R. Campos-Amezcua, and C. Heard, "Wind speed prediction using a univariate ARIMA model and a multivariate NARX model," *Energies*, vol. 9, no. 2, pp. 1–15, 2016.

[159] J. P. Catalão, H. M. Pousinho, and V. M. Mendes, "Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 1, pp. 50–59, 2011.

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

[161] A. U. Haque, P. Mandal, H. M. Nehrir, A. Bhuiya, and R. Baker, "A hybrid intelligent framework for wind power forecasting engine," *Proceedings - 2014 Electrical Power and Energy Conference, EPEC 2014*, pp. 184–189, 2014.

[162] E. Faghihnia, S. Salahshour, A. Ahmadian, and N. Senu, "Developing a Local Neuro-fuzzy Model for Short-Term Wind Power Forecasting," *Advances in Mathematical Physics*, vol. 2014, 2015.

[163] T. G. Barbounis and J. B. Theocharis, "A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation," *Neurocomputing*, vol. 70, no. 7-9, pp. 1525–1542, 2007.

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

[165] G. Sideratos and N. D. Hatziargyriou, "An advanced statistical method for wind power forecasting," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 258–265, 2007.

[166] A. Qazi, H. Fayaz, A. Wadi, R. G. Raj, N. A. Rahim, and W. A. Khan, "The artificial neural network for solar radiation prediction and designing solar systems: A systematic literature review," *Journal of Cleaner Production*, vol. 104, pp. 1–12, 2015.

[167] G. Sideratos and N. D. Hatziargyriou, "Probabilistic wind power forecasting using radial basis function neural networks," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1788–1796, 2012.

[168] C. Zhang, H. Wei, L. Xie, Y. Shen, and K. Zhang, "Direct interval forecasting of wind speed using radial basis function neural networks in a multi-objective optimization framework," *Neurocomputing*, vol. 205, pp. 53–63, 2016.

[169] M. Vafaeipour, O. Rahbari, M. A. Rosen, F. Fazelpour, and P. Ansarirad, "Application of sliding window technique for prediction of wind velocity time series," *International Journal of Energy and Environmental Engineering*, vol. 5, no. 2-3, pp. 1–7, 2014.

[170] X. Wu, B. Hong, X. Peng, F. Wen, and J. Huang, "Radial basis function neural network based short-term wind power forecasting with Grubbs test," *DRPT 2011 - 2011 4th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, pp. 1879–1882, 2011.

[171] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Applied Energy*, vol. 87, no. 7, pp. 2313–2320, 2010.

[172] J. Yan, X. Gao, Y. Liu, S. Han, L. Li, X. Ma, C. Gu, R. Bhakar, and F. Li, "Adaptabilities of three mainstream short-term wind power forecasting methods," *Journal of Renewable and Sustainable Energy*, vol. 7, no. 5, 2015.

[173] H. Chitsaz, N. Amjady, and H. Zareipour, "Wind power forecast using wavelet neural network trained by improved Clonal selection algorithm," *Energy Conversion and Management*, vol. 89, pp. 588–598, 2015.

[174] J. Wang, F. Zhang, F. Liu, and J. Ma, "Hybrid forecasting model-based data mining and genetic algorithm-Adaptive particle swarm optimisation: A case study of wind speed time series," *IET Renewable Power Generation*, vol. 10, no. 3, pp. 287–298, 2016.

[175] B. Doucoure, K. Agbossou, and A. Cardenas, "Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data," *Renewable Energy*, vol. 92, pp. 202–211, 2016.

[176] L. J. Ricalde, G. A. Catzin, A. Y. Alanis, and E. N. Sanchez, "Higher order wavelet neural networks with Kalman learning for wind speed forecasting," *IEEE SSCI 2011 - Symposium Series on Computational Intelligence - CIASG 2011: 2011 IEEE Symposium on Computational Intelligence Applications in Smart Grid*, pp. 55–60, 2011.

[177] E. Lotfi, A. Khosravi, M. R. Akbarzadeh, and S. Nahavandi, "Wind power forecasting using emotional neural networks," *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 2014-Janua, no. January, pp. 311–316, 2014.

[178] R. M. Peri, P. Mandal, A. U. Haque, and B. Tseng, "Very short-term prediction of wind farm power: An advanced hybrid intelligent approach," *IEEE Industry Application Society - 51st Annual Meeting, IAS 2015, Conference Record*, pp. 1–8, 2015.

[179] G. Xu, C. Xiu, and Z. Wan, "Hysteretic chaotic operator network and its application in wind speed series prediction," *Neurocomputing*, vol. 165, pp. 384–388, 2015.

[180] L. Zjavka, "Wind speed forecast correction models using polynomial neural networks," *Renewable Energy*, vol. 83, pp. 998–1006, 2015.

[181] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1033–1044, 2014.

[182] J. Wang, J. Hu, K. Ma, and Y. Zhang, "A self-adaptive hybrid approach for wind speed forecasting," *Renewable Energy*, vol. 78, pp. 374–385, 2015.

[183] S. Salcedo-Sanz, A. Pastor-Sánchez, J. Del Ser, L. Prieto, and Z. W. Geem, "A Coral Reefs Optimization algorithm with Harmony Search operators for accurate wind speed prediction," *Renewable Energy*, vol. 75, pp. 93–101, 2015.

[184] C. R. Alcantud, "Feature Selection with a Grouping Genetic Algorithm – Extreme Learning Machine Approach for Wind Power Prediction," pp. 479–488, 2016.

[185] G. Zhang, Y. Wu, and Y. Liu, "An advanced wind speed multi-step ahead forecasting approach with characteristic component analysis," *Journal of Renewable and Sustainable Energy*, vol. 6, no. 5, pp. 1–14, 2014.

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

[187] H. Z. Wang, G. B. Wang, G. Q. Li, J. C. Peng, and Y. T. Liu, "Deep belief network based deterministic and probabilistic wind speed forecasting approach," *Applied Energy*, vol. 182, pp. 80–93, 2016.

[188] H. zhi Wang, G. qiang Li, G. bing Wang, J. chun Peng, H. Jiang, and Y. tao Liu, "Deep learning based ensemble approach for probabilistic wind power forecasting," *Applied Energy*, vol. 188, pp. 56–70, 2017.

[189] Q. Hu, R. Zhang, and Y. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renewable Energy*, vol. 85, pp. 83–95, 2016.

[190] C. Y. Zhang, C. L. Chen, M. Gan, and L. Chen, "Predictive Deep Boltzmann Machine for Multiperiod Wind Speed Forecasting," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1416–1425, 2015.

[191] Q. Xu, D. He, N. Zhang, C. Kang, Q. Xia, J. Bai, and J. Huang, "A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1283–1291, 2015.

[192] R. Jursa and K. Rohrig, "Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models," *International Journal of Forecasting*, vol. 24, no. 4, pp. 694–709, 2008.

[193] N. Amjady, F. Keynia, and H. Zareipour, "A new hybrid iterative method for short - term wind speed forecasting A new hybrid iterative method for short-term wind speed forecasting," no. January, 2011.

[194] J. P. Catalão, H. M. Pousinho, and V. M. Mendes, "Short-term wind power forecasting in Portugal by neural networks and wavelet transform," *Renewable Energy*, vol. 36, no. 4, pp. 1245–1251, 2011.

[195] T. Li, Y. Li, M. Liao, W. Wang, and C. Zeng, "A New Wind Power Forecasting Approach Based on Conjugated Gradient Neural Network," *Mathematical Problems in Engineering*, vol. 2016, 2016.

[196] R. K. G, V. Prema, and K. U. Rao, "Multivariate Wind Power Forecast using Artificial Neural Network," pp. 159–163, 2014.

[197] R. Singh, K. B. Sahay, and S. A. Srivastava, "Short-term wind speed forecasting of Oak Park Weather Station by using different ANN algorithms," *Proceedings of the 2015 IEEE Innovative Smart Grid Technologies - Asia, ISGT ASIA 2015*, 2016.

[198] H. Liu, H. Q. Tian, D. F. Pan, and Y. F. Li, "Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Networks," *Applied Energy*, vol. 107, pp. 191–208, 2013.

[199] H. Liu, H. Tian, X. Liang, and Y. Li, "New wind speed forecasting approaches using fast ensemble empirical model decomposition, genetic algorithm, Mind Evolutionary Algorithm and Artificial Neural Networks," *Renewable Energy*, vol. 83, pp. 1066–1075, 2015.

[200] G. Jia, D. Li, L. Yao, and P. Zhao, "An improved artificial bee colony-BP neural network algorithm in the short-term wind speed prediction," *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, vol. 2016-Septe, no. 2, pp. 2252–2255, 2016.

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

[202] O. Abedinia and N. Amjady, "Short-term wind power prediction based on Hybrid Neural Network and chaotic shark smell optimization," *International Journal of Precision Engineering and Manufacturing - Green Technology*, vol. 2, no. 3, pp. 245–254, 2015.

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

[204] F. Fazelpour, N. Tarashkar, and M. A. Rosen, "Short-term wind speed forecasting using artificial neural networks for Tehran, Iran," *International Journal of Energy and Environmental Engineering*, vol. 7, no. 4, pp. 377–390, 2016.

[205] R. Ak, O. Fink, and E. Zio, "Two Machine Learning Approaches for Short-Term Wind Speed Time-Series Prediction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1734–1747, 2016.

[206] H. Liu, H. Q. Tian, C. Chen, and Y. F. Li, "An experimental investigation of two Wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization," *International Journal of Electrical Power and Energy Systems*, vol. 52, no. 1, pp. 161–173, 2013.

[207] Y. Y. Hong, H. L. Chang, and C. S. Chiu, "Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs," *Energy*, vol. 35, no. 9, pp. 3870–3876, 2010.

[208] K. Bhaskar and S. N. Singh, "AWNN-Assisted wind power forecasting using feed-forward neural network," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 2, pp. 306–315, 2012.

[209] C. Hervás-Martínez, P. A. Gutiérrez, J. C. Fernández, S. Salcedo-Sanz, A. Portilla-Figueras, A. Pérez-Bellido, and L. Prieto, "Hyperbolic tangent basis function neural networks training by hybrid evolutionary programming for accurate short-term wind speed prediction," *ISDA 2009 - 9th International Conference on Intelligent Systems Design and Applications*, pp. 193–198, 2009.

[210] H. Liu, H. Q. Tian, Y. F. Li, and L. Zhang, "Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions," *Energy Conversion and Management*, vol. 92, pp. 67–81, 2015.

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

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

[213] K. Gnana Sheela and S. N. Deepa, "An intelligent computing model for wind speed prediction in renewable energy systems," *Procedia Engineering*, vol. 30, pp. 380–385, 2012.

[214] J. Heng, C. Wang, X. Zhao, and L. Xiao, "Research and application based on adaptive boosting strategy and modified CGFPA algorithm: A case study for wind speed forecasting," *Sustainability (Switzerland)*, vol. 8, no. 3, 2016.

[215] L. Dong, L. Wang, S. F. Khahro, S. Gao, and X. Liao, "Wind power day-ahead prediction with cluster analysis of NWP," *Renewable and Sustainable Energy Reviews*, vol. 60, pp. 1206–1212, 2016.

[216] S. Li, P. Wang, and L. Goel, "Wind Power Forecasting Using Neural Network Ensembles with Feature Selection," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1447–1456, 2015.

[217] M. Mohanraj, S. Jayaraj, and C. Muraleedharan, "Applications of artificial neural networks for thermal analysis of heat exchangers - A review," *International Journal of Thermal Sciences*, vol. 90, pp. 150–172, 2015.

[218] O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis, and K. Taha, "Efficient Machine Learning for Big Data: A Review," *Big Data Research*, vol. 2, no. 3, pp. 87–93, 2015.

[219] C. L. Philip Chen and C. Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[220] C. A. Fiebrich, Y. R. Morgan, A. G. Mccombs, P. K. Hall, and R. A. Mcpherson, "Quality assurance procedures for mesoscale meteorological data," *Journal of Atmospheric and Oceanic Technology*, vol. 27, no. 10, pp. 1565–1582, 2010.

[221] L. Ehrlinger, B. Werth, and W. Wöß, "Automated Continuous Data Quality Measurement with QuaIIe," *International Journal on Advances in Software*, vol. 11, no. 3 & 4, pp. 400–417, 2018.

[222] L. S. Gandin, "Complex quality control of meteorological observations," 1988.

[223] Z. Yang, Y. Liu, and C. Li, "Interpolation of missing wind data based on ANFIS," *Renewable Energy*, vol. 36, no. 3, pp. 993–998, 2011.

[224] J. Estévez, P. Gavilán, and J. V. Giráldez, "Guidelines on validation procedures for meteorological data from automatic weather stations," vol. 402, pp. 144–154, 2011.

[225] P. A. Jiménez, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, "Quality assurance of surface wind observations from automated weather stations," *Journal of Atmospheric and Oceanic Technology*, vol. 27, no. 7, pp. 1101–1122, 2010.

[226] K. Lakshminarayan, S. A. Harp, and T. Samad, "Imputation of missing data in industrial databases," *Applied Intelligence*, vol. 11, no. 3, pp. 259–275, 1999.

[227] A. Liguori, R. Markovic, T. T. H. Dam, J. Frisch, C. van Treeck, and F. Causone, "Indoor environment data time-series reconstruction using autoencoder neural networks," *Building and Environment*, vol. 191, pp. 1–26, 2021.

[228] M. Amiri and R. Jensen, "Missing data imputation using fuzzy-rough methods," *Neurocomputing*, vol. 205, pp. 152–164, 2016.

[229] H. Demirhan and Z. Renwick, "Missing value imputation for short to mid-term horizontal solar irradiance data," *Applied Energy*, vol. 225, no. April, pp. 998–1012, 2018.

[230] T. Kim, W. Ko, and J. Kim, "Analysis and impact evaluation of missing data imputation in day-ahead PV generation forecasting," *Applied Sciences (Switzerland)*, vol. 9, no. 1, pp. 1–18, 2019.

[231] W. Zhang, Y. Luo, Y. Zhang, and D. Srinivasan, "SolarGAN: Multivariate solar data imputation using generative adversarial network," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 743–746, 2021.

[232] A. Aieb, K. Madani, M. Scarpa, B. Bonacorso, and K. Lefsih, "A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria," *Heliyon*, vol. 5, no. 2, p. e01247, 2019.

[233] J. J. Miró, V. Caselles, and M. J. Estrela, "Multiple imputation of rainfall missing data in the Iberian Mediterranean context," *Atmospheric Research*, vol. 197, no. 2016, pp. 313–330, 2017.

[234] Z. X. Xie and X. F. Sun, "Imputation of missing wind speed data based on low-rank matrix approximation," *2017 2nd International Conference on Power and Renewable Energy, ICPRE 2017*, pp. 397–401, 2018.

[235] A. R. Alsaber, J. Pan, and A. Al-Hurban, "Handling complex missing data using random forest approach for an air quality monitoring dataset: A case study of kuwait environmental data (2012 to 2018)," *International Journal of Environmental Research and Public Health*, vol. 18, no. 3, pp. 1–26, 2021.

[236] E. Afrifa-Yamoah, U. A. Mueller, S. M. Taylor, and A. J. Fisher, "Missing data imputation of high-resolution temporal climate time series data," *Meteorological Applications*, vol. 27, no. 1, pp. 1–18, 2020.

[237] A. Flores, H. Tito, and D. Centty, "Recurrent neural networks for meteorological time series imputation," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 482–487, 2020.

[238] L. Gao, Y. Zheng, Y. Wang, J. Xia, X. Chen, B. Li, M. Luo, and Y. Guo, "Reconstruction of missing data in weather radar image sequences using deep neuron networks," *Applied Sciences (Switzerland)*, vol. 11, no. 4, pp. 1–21, 2021.

[239] C. Yang, J. Liu, Y. Zeng, and G. Xie, "Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model," *Renewable Energy*, vol. 133, pp. 433–441, 2019.

[240] S. Ahmed, Y. D. Lee, S. H. Hyun, and I. Koo, "Mitigating the impacts of covert cyber attacks in smart grids via reconstruction of measurement data utilizing deep denoising autoencoders," *Energies*, vol. 12, no. 16, 2019.

[241] G. M. M. P. Khan, Nadia Masood Khan, "AI Based Real-Time Signal Reconstruction for Wind Farm with SCADA Sensor Failure," 2020.

[242] T. Liu, H. Wei, and K. Zhang, "Wind power prediction with missing data using Gaussian process regression and multiple imputation," *Applied Soft Computing Journal*, vol. 71, pp. 905–916, 2018.

[243] I. Durre, M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, "Comprehensive automated quality assurance of daily surface observations," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 8, pp. 1615–1633, 2010.

[244] N. Regional, "Complex Quality Assurance of Historical Hourly Surface Airways Meteorological Data," vol. 3505, no. Ncdc 2001, pp. 1156–1169, 2004.

[245] J. You, G. Hubbard, and S. Goddard, "Comparison of methods for spatially estimating station temperatures in a quality control system," vol. 787, no. August 2007, pp. 777–787, 2008.

[246] O. C. Survey, E. Verification, and O. C. Survey, "Quality Assurance Procedures in the Oklahoma Mesonetwork," pp. 474–494, 2000.

[247] I. Zahumenský, "Guidelines on Quality Control Procedures for Data from Automatic Weather Stations Guidelines on Quality Control Procedures for Data from Automatic Weather Stations," no. September, 2016.

[248] D. W. Meek and J. L. Hatfield, "Data quality checking for single station meteorological databases," *Agricultural and Forest Meteorology*, vol. 69, no. 1-2, pp. 85–109, 1994.

[249] N. Regional, "NOTES AND CORRESPONDENCE A Quality-Control Routine for Hourly Wind Observations," pp. 308–317, 1997.

[250] F. Vejen, C. Jacobsson, U. Fredriksson, M. Moei, L. Andresen, E. Hellsten, and O. Rissanen, "Quality Control of Meteorological Observations," 2002.

[251] SHEPARD D, "Two- dimensional interpolation function for irregularly- spaced data," *Proc 23rd Nat Conf*, pp. 517–524, 1968.

[252] P. J. García-Laencina, J. L. Sancho-Gómez, and A. R. Figueiras-Vidal, "Pattern classification with missing data: A review," *Neural Computing and Applications*, vol. 19, no. 2, pp. 263–282, 2010.

[253] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

[254] D. J. C. Mackay, "Bayesian Interpolation," *Computation and Neural Systems, California Institute of Technology 139-74, Pasadena,CA 91225 USA*, vol. 447, pp. 415–447, 1992.

[255] A. Sayigh.

[256] O. Baghirli, "Comparison of Lavenberg-Marquardt, Scaled Conjugate Gradient And Bayesian Regularization Backpropagation Algorithms for Multistep Ahead Wind Speed Forecasting Using Multilayer Perceptron Feedforward Neural Network," *Dissertation*, no. June, p. Uppsala University, 2015.

[257] N. Siebert, "Development of Methods for Regional Wind Power Forecasting," 2008.

[258] T. E. Kibona, "Application of WRF mesoscale model for prediction of wind energy resources in Tanzania," *Scientific African*, vol. 7, p. e00302, 2020.

[259] J. M. Cascón, "Local wind speed forecasting based on WRF-HDWind coupling," *Atmospheric Research*, vol. 248, no. July 2020, p. 105219, 2021.

[260] N. M. Nawi, F. Hamzah, N. A. Hamid, and M. Z. Rehman, "An Optimized Back Propagation Learning Algorithm with Adaptive Learning An Optimized Back Propagation Learning Algorithm with Adaptive Learning Rate," no. October, 2017.

[261] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," 2015.

[262] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Let a biogeography-based optimizer train your Multi-Layer Perceptron," *Information Sciences*, vol. 269, pp. 188–209, 2014.

[263] Y. Zhang, G. Pan, B. Chen, J. Han, and Y. Zhao, "Short-term wind speed prediction model based on GA-ANN improved by VMD," *Renewable Energy*, no. xxxx, 2019.

[264] Z. Tian, Y. Ren, and G. Wang, "Environmental Effects Short-term wind speed prediction based on improved PSO algorithm optimized EM-ELM," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 00, no. 00, pp. 1–21, 2018.

[265] G. Grégoire, "Elements of Statistics," *EAS Publications Series*, vol. 77, pp. 13–37, 2016.

[266] M. Claesen and B. De Moor, "Hyperparameter Search in Machine Learning," pp. 10–14, 2015.

[267] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," vol. 13, pp. 281–305, 2012.

[268] C. M. Bishop, "Neural Networks for Pattern Recognition," *Clarendon press. Oxford*, 1995.

[269] Y. Zhang, "Wind speed prediction research with EMD-BP based on Lorenz disturbance," vol. 70, pp. 198–207, 2019.

[270] Z. Zhang, L. Ye, H. Qin, Y. Liu, C. Wang, X. Yu, X. Yin, and J. Li, "Wind speed prediction method using Shared Weight Long Short-Term Memory Network and Gaussian Process Regression," *Applied Energy*, vol. 247, no. April, pp. 270–284, 2019.

[271] Y. Zhang, C. Zhang, Y. Zhao, and S. Gao, "Wind speed prediction with RBF neural network based on PCA and ICA Source signal S," vol. 69, pp. 148–155, 2018.

[272] L. Wang, X. Li, and Y. Bai, "Short-term wind speed prediction using an extreme learning machine model with error correction," *Energy Conversion and Management*, vol. 162, no. January, pp. 239–250, 2018.

[273] Z. Gan, C. Li, J. Zhou, and G. Tang, "Temporal convolutional networks interval prediction model for wind speed forecasting," *Electric Power Systems Research*, vol. 191, no. September 2020, p. 106865, 2021.

[274] Y. Li, X. Chen, C. Li, Z. Gan, and X. An, "A Hybrid Deep Interval Prediction Model for Wind Speed Forecasting," pp. 7323–7335, 2021.

[275] Z. Tian, H. Li, and F. Li, "A combination forecasting model of wind speed based on decomposition," *Energy Reports*, vol. 7, pp. 1217–1233, 2021.

[276] X. Mi and S. Zhao, "Wind speed prediction based on singular spectrum analysis and neural network structural learning," *Energy Conversion and Management*, vol. 216, no. April, p. 112956, 2020.

[277] K. B. Ensor and P. W. Glynn., "Stochastic optimization via grid search," 1997.

[278] S. Andradóttir, "A Review of Random Search Methods," pp. 277–292.

[279] H. G.E., "A practical guide to training restricted boltzmann machines," 2012.

[280] R. Bellman, *Adaptive Control Processes. A Guided Tour*. Princeton University Press., 1962.

[281] D. Simon, "Biogeography-based optimization," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 702–713, 2008.

[282] H. Ma, D. Simon, M. Fei, and Z. Xie, "Variations of biogeography-based optimization and Markov analysis," *Information Sciences*, vol. 220, pp. 492–506, 2013.

[283] S. Hansun, "A New Approach of Moving Average Method in Time Series Analysis,"

[284] J. W. Wilder, *New concepts in technical trading systems*. Trend research, 1978.

[285] J. Kuha, "AIC and BIC: Comparisons of assumptions and performance," *Sociological Methods and Research*, vol. 33, no. 2, pp. 188–229, 2004.

[286] H. Schwartz, "estimating the dimension of a model," *The annals of statistics*, 1987.

[287] A. Leelossy, F. Molnar, A. Havasi, and I. Lagzi, "Dispersion modeling of air pollutants in the atmosphere : a review Dispersion modeling of air pollutants in the atmosphere : a review," *Cent. Eur. J. Geosci.*, no. September 2015, pp. 257–278, 2014.

[288] J. Li and J. Wang, "Forcasting of energy futures market and synchronization based on stochastic gated recurrent unit model," *Energy*, vol. 213, p. 118787, 2020.

[289] H. Cai, X. Jia, J. Feng, Q. Yang, Y.-m. Hsu, Y. Chen, and J. Lee, "A combined fi ltering strategy for short term and long term wind speed prediction with improved accuracy," *Renewable Energy*, 2018.

[290] Y. Hao and C. Tian, "A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting," *Applied Energy*, vol. 238, no. January, pp. 368–383, 2019.

[291] S. Belhardj, S. Mimouni, A. Saidane, and M. Benzohra, "Using microchannels to cool microprocessors: A transmission-line-matrix study," *Microelectronics Journal*, vol. 34, no. 4, pp. 247–253, 2003.

[292] Q. Hu, R. Zhang, and Y. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renewable Energy*, vol. 85, pp. 83–95, 2016.

[293] Y. Liu, "Machine learning for wind power prediction," *MCS Thesis. University of New Brunswick . . .* , 2016.

[294] M. R. . F. W. J. Wegley, H. L.; Kosorok, "Subhourly wind forecasting techniques for wind turbine operations," 1984.

[295] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Applied Energy*, vol. 87, no. 7, pp. 2313–2320, 2010.

[296] P. Zhao, J. Wang, J. Xia, Y. Dai, Y. Sheng, and J. Yue, "Performance evaluation and accuracy enhancement of a day-ahead wind power forecasting system in China," *Renewable Energy*, vol. 43, pp. 234–241, 2012.

[297] G. Li, J. Shi, and J. Zhou, "Bayesian adaptive combination of short-term wind speed forecasts from neural network models," *Renewable Energy*, vol. 36, no. 1, pp. 352–359, 2011.

[298] F. Benoît, M. van Heeswijk, Y. Miche, M. Verleysen, and A. Lendasse, "Feature selection for nonlinear models with extreme learning machines," *Neurocomputing*, vol. 102, pp. 111–124, 2013.

[299] H. Liu, Y. Li, Z. Duan, and C. Chen, "A review on multi-objective optimization framework in wind energy forecasting techniques and applications," vol. 224, no. April, 2020.

[300] P. Zhang, "Short-term Wind Power Forecast Based on GA-Elman Neural Network and Nonlinear Combination Model," no. Ipemec, pp. 974–979, 2015.

[301] H. Liu, X. Mi, and Y. Li, "Comparison of two new intelligent wind speed forecasting approaches based on Wavelet Packet Decomposition , Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Arti fi cial Neural Networks," *Energy Conversion and Management*, vol. 155, no. October 2017, pp. 188–200, 2018.

[302] Y. Zhang, Y. Li, and G. Zhang, "Short-term wind power forecasting approach based on Seq2Seq model using NWP data," *Energy*, vol. 213, p. 118371, 2020.

[303] N. An, W. Zhao, J. Wang, D. Shang, and E. Zhao, "Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting," *Energy*, vol. 49, no. 1, pp. 279–288, 2013.

[304] A. Wysocki and M. Ławryńczuk, "Elman neural network for modeling and predictive control of delayed dynamic systems," *Archives of Control Sciences*, vol. 26, no. 1, pp. 117–142, 2016.

[305] K. Chen and J. Yu, "Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach," *Applied Energy*, vol. 113, pp. 690–705, 2014.

[306] S. Mirjalili, "Evolutionary Algorithms and Neural Networks," *Springer International Publishing AG, part of Springer Nature 2019*.

[307] J. Kennedy and R. Eberhart, "Particle Swarm Optimisation," *IEEE Int. Conf. Neural Networks*, vol. 4, pp. 1942–1948, 1995.

[308] Y. Zhang, G. Pan, Y. Zhao, Q. Li, and F. Wang, "Short-term wind speed interval prediction based on artificial intelligence methods and error probability distribution," *Energy Conversion and Management*, vol. 224, no. August, p. 113346, 2020.

[309] J. Sun, B. Feng, and W. Xu, "Particle swarm optimization with particles having quantum behavior," *Proceedings of the 2004 Congress on Evolutionary Computation, CEC2004*, vol. 1, pp. 325–331, 2004.

[310] J. Sun, W. Fang, V. Palade, X. Wu, and W. Xu, "Quantum-behaved particle swarm optimization with Gaussian distributed local attractor point," *Applied Mathematics and Computation*, vol. 218, no. 7, pp. 3763–3775, 2011.

[311] Y. W. Jeong, J. B. Park, S. H. Jang, and K. Y. Lee, "A new quantum-inspired binary PSO: Application to unit commitment problems for power systems," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1486–1495, 2010.

[312] S. Mirjalili and A. Lewis, "S-shaped versus V-shaped transfer functions for binary Particle Swarm Optimization," *Swarm and Evolutionary Computation*, vol. 9, pp. 1–14, 2013.

[313] K.-H. Han and J.-H. Kim, "Quantum-inspired evolutionary algorithm for a class of combinatorial optimization," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 6, pp. 580–593, 2002.

[314] S. Belciug and F. Gorunescu, "Error-correction learning for artificial neural networks using the Bayesian paradigm . Application to automated medical diagnosis," *JOURNAL OF BIOMEDICAL INFORMATICS*, 2014.

[315] Y. Ding, *Data cience for wind energy*. 2020.

[316] J. Sevilla, "Importance of input data normalization for the application of neural networks to complex industrial problems IMPORTANCE OF INPUT DATA NORMALIZATION FOR THE APPLICATION OF NEURAL NETWORKS TO COMPLEX," no. July 1997, 2012.

[317] S. Mi, "Short-Term Forecasting of Wind Speed and Wind Power Based on BP and Adaboost {_} BP," *Solar Energy*, no. August, 2014.

[318] F. Shahid, A. Zameer, A. Mehmood, M. Asif, and Z. Raja, "A novel wavenets long short term memory paradigm for wind power prediction," *Applied Energy*, vol. 269, no. April, p. 115098, 2020.

[319] I. Taleb, H. T. Kassabi, M. A. Serhani, R. Dssouli, and C. Bouhaddioui, "Big Data Quality: A Quality Dimensions Evaluation," *Proceedings - 13th IEEE International Conference on Ubiquitous Intelligence and Computing, 13th IEEE International Conference on Advanced and Trusted Computing, 16th IEEE International Conference on Scalable Computing and Communications, IEEE Internationa*, no. February 2018, pp. 759–765, 2017.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA