Universidad de Alcalá
COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

ACTA DE EVALUACIÓN DE LA TESIS DOCTORAL

Año académico 2016/17

DOCTORANDO: **SAAVEDRA MORENO,BEATRIZ**
D.N.I./PASAPORTE: ******1119K**

PROGRAMA DE DOCTORADO: **D347-TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES**
DEPARTAMENTO DE: **TEORÍA DE LA SEÑAL Y COMUNICACIONES**
TITULACIÓN DE DOCTOR EN: **DOCTOR/A POR LA UNIVERSIDAD DE ALCALÁ**

En el día de hoy 14/10/2016, reunido el tribunal de evaluación nombrado por la Comisión de Estudios Oficiales de Posgrado y Doctorado de la Universidad y constituido por los miembros que suscriben la presente Acta, el aspirante defendió su Tesis Doctoral, elaborada bajo la dirección de **SANCHO SALCEDO SANZ // JOSE ANTONIO PORTILLA FIGUERAS.**

Sobre el siguiente tema: *NEW SOFT-COMPUTING TECHNIQUES IN WIND ENERGY: CONTRIBUTIONS TO WIND SPEED RECONSTRUCTION, PREDICTION AND WIND FARM DESIGN*

Finalizada la defensa y discusión de la tesis, el tribunal acordó otorgar la CALIFICACIÓN GLOBAL[4] de (**no apto, aprobado, notable y sobresaliente**): SOBRESALIENTE

Alcalá de Henares, 14 de OCTUBRE de 2016

EL PRESIDENTE

Fdo.: LUCAS CUADRA

EL SECRETARIO

Fdo.: SILVIA JIMENEZ

EL VOCAL

Fdo.: ANTONIO CAAMAÑO

EL VOCAL

Fdo.: SERGIO DAMAS

EL VOCAL

Fdo.: LUIS PRIETO

FIRMA DEL ALUMNO,

Fdo.: BEATRIZ SAAVEDRA

Con fecha 27 de octubre de 2016 la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado, a la vista de los votos emitidos de manera anónima por el tribunal que ha juzgado la tesis, resuelve:

☒ Conceder la Mención de "Cum Laude"
☐ No conceder la Mención de "Cum Laude"

La Secretaria de la Comisión Delegada

[4] La calificación podrá ser "**no apto**" "**aprobado**" "**notable**" y "**sobresaliente**". El tribunal podrá otorgar la mención de "cum laude" si la calificación global es de sobresaliente y se emite en tal sentido el voto secreto positivo por unanimidad.

INCIDENCIAS / OBSERVACIONES:

Universidad de Alcalá

COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

En aplicación del art. 14.7 del RD. 99/2011 y el art. 14 del Reglamento de Elaboración, Autorización y Defensa de la Tesis Doctoral, la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado y Doctorado, en sesión pública de fecha 27 de octubre, procedió al escrutinio de los votos emitidos por los miembros del tribunal de la tesis defendida por *SAAVEDRA MORENO,BEATRIZ*, el día 14 de octubre de 2016, titulada *NEW SOFT-COMPUTING TECHNIQUES IN WIND ENERGY: CONTRIBUTIONS TO WIND SPEED RECONSTRUCTION, PREDICTION AND WIND FARM DESIGN*, para determinar, si a la misma, se le concede la mención "cum laude", arrojando como resultado el voto favorable de todos los miembros del tribunal.

Por lo tanto, la Comisión de Estudios Oficiales de Posgrado **resuelve otorgar** a dicha tesis la

MENCIÓN "CUM LAUDE"

Alcalá de Henares, 31 de octubre de 2016
EL PRESIDENTE DE LA COMISIÓN DE ESTUDIOS
OFICIALES DE POSGRADO Y DOCTORADO

UNIVERSIDAD DE ALCALÁ - COMISIÓN DE ESTUDIOS OFICIALES DE POSGRADO Y DOCTORADO

Juan Ramón Velasco Pérez

Copia por e-mail a:
Doctorando: SAAVEDRA MORENO,BEATRIZ
Secretario del Tribunal: SILVIA JIMÉNEZ FERNÁNDEZ
Directores de Tesis: SANCHO SALCEDO SANZ // JOSE ANTONIO PORTILLA FIGUERAS

Universidad de Alcalá

Campus Universitario
Dpto. de Teoría de la Señal y Comunicaciones
Ctra. Madrid-Barcelona, Km. 36,6
28805 Alcalá de Henares (Madrid) Telf: +34 91 885 88 99
Fax: +34 91 885 66 99

Dr. D. Sancho Salcedo Sanz, Profesor Titular de Universidad del Área de Conocimiento de Teoría de la Señal y Comunicaciones de la Universidad de Alcalá, y Dr. D. José Antonio Portilla Figueras, Profesor Titular de Universidad del Área de Conocimiento de Teoría de la Señal y Comunicaciones de la Universidad de Alcalá,

CERTIFICAN

Que la tesis "New Soft-Computing techniques in wind energy: contributions to wind speed reconstruction, prediction and wind farm design", presentada por Beatriz Saavedra Moreno y realizada en el Departamento de Teoría de la Señal y Comunicaciones bajo nuestra dirección, reúne méritos suficientes para optar al grado de Doctor, por lo que puede procederse a su depósito y lectura.

Alcalá de Henares, Junio 2016.

Fdo.: Dr. D. Sancho Salcedo Sanz

Fdo.: Dr. D. José Antonio Portilla Figueras

Universidad de Alcalá

Campus Universitario
Dpto. de Teoría de la Señal y Comunicaciones
Ctra. Madrid-Barcelona, Km. 36,6
28805 Alcalá de Henares (Madrid) Telf: +34 91 885 88 99
Fax: +34 91 885 66 99

Beatriz Saavedra Moreno ha realizado en el Departamento de Teoría de la Señal y Comunicaciones y bajo la dirección del Dr. D. Sancho Salcedo Sanz y del Dr. D. José Antonio Portilla Figueras, la tesis doctoral titulada "New Soft-Computing techniques in wind energy: contributions to wind speed reconstruction, prediction and wind farm design", cumpliéndose todos los requisitos para la tramitación que conduce a su posterior lectura.

Alcalá de Henares, Junio 2016.

EL DIRECTOR DEL DEPARTAMENTO

Fdo: Dr. D. Javier Acevedo Rodríguez

Universidad de Alcalá

Escuela Politécnica Superior

Departamento de Teoría de la Señal y Comunicaciones

Tesis Doctoral

New Soft-Computing Techniques in Wind Energy: contributions to wind speed reconstruction, prediction and wind farm design

Autor:

Beatriz Saavedra Moreno

Directores:

Dr. D. Sancho Salcedo Sanz
Dr. D. José Antonio Portilla Figueras

Junio, 2016

Abstract

This Ph. D. Thesis deals with different problems that arise when planning, designing and managing a wind farm. This work proposes several solutions to these problems using soft-computing techniques. First of all, the reconstruction of wind speed values based on synoptic pressure values is tackled. We propose the application of a support vector machine for regression, in order to be able to establish a relationship between the measured wind speed values and the synoptic pressure values, available on a grid over the area under study. A Weather Regimes Classification Technique (WRCT) is applied for comparison. Both algorithms are evaluated in several real problems of wind speed reconstruction at three sites, obtaining excellent results in terms of wind speed reconstruction.

Then, wind speed reconstruction of wind series using data from in-situ measuring towers is carried out. In this work, we present the application of two state-of-the-art neural networks which have shown a very fast training time with an excellent performance in terms of accuracy. Specifically, we show the application of Group Method of Data Handling and Extreme Learning Machines in the reconstruction of wind speed series, in a real wind farm in Spain. A comparison in terms of computation time and accuracy with alternative algorithms in the literature is also carried out.

Surface wind speed distribution over a wind farm area is a key parameter related to several different processes in wind farm prospecting, design and micrositing. This information is often obtained from mesoscale models simulations, that include variables from global models. We present several techniques to correct surface wind speed simulations from mesoscale models, using data from measuring stations in wind farms. Specifically, we propose different heuristic corrections of the mesoscale models output by means of a surface fitting between the wind speed series Weibull parameters (from the mesoscale model) and those from the measuring stations (real wind speed) in the wind farm. The good performance of our proposal is shown in the correction of the surface wind speed from mesoscale models in two wind farms facilities in Spain where several measuring towers are installed.

Finally, a novel evolutionary algorithm for optimal positioning of wind turbines in wind farms is proposed. A realistic model for the wind farm is considered in the optimization process, which includes orography, shape of the wind farm, simulation of the wind speed and direction, and costs of installation, connection and road construction among wind turbines. Regarding the solution of the problem, this work introduces a greedy heuristic algorithm which is able to obtain a reasonable initial solution for the problem. This heuristic is then used to seed the initial population of the evolutionary algorithm, improving its performance. It is shown that the proposed seeded evolutionary approach is able to obtain very good solutions to this problem, which maximize the economical benefit which can be obtained from the wind farm.

Resumen

En esta tesis se presentan diferentes problemas que surgen durante la planificación, el diseño y la operación de un parque eólico. Este trabajo propone abordar estos problemas utilizando diferentes técnicas de soft-computing. En primer lugar se aborda la reconstrucción de valores de velocidad de viento basada en valores de presiones sinópticas. Se propone la utilización de máquinas de vectores de soporte, para establecer una relación entre los valores de viento y los valores de presión, disponibles en un grid sobre el área de estudio. Además, se aplica una técnica de Weather Regimes, para comparar. Ambos algoritmos son evaluados en varias situaciones reales en tres localizaciones, obteniendo resultados excelentes en términos de reconstrucción de viento.

Se propone también la reconstrucción de series de viento utilizando datos de torres reales de medida. En este trabajo se presenta la aplicación de dos redes neuronales, con un entrenamiento extremadamente rápido y con un rendimiento excelente en términos de precisión. En concreto, se aplican Extreme Learning Machines y Group Method of Data Handling para la reconstrucción de series de viento en un parque eólico real en España. Se realiza una comparación con otros algoritmos alternativos de la literatura en términos de tiempo de computación y precisión.

Por otro lado, la distribución espacial del viento sobre un parque eólico proporciona información muy útil. Esta información se obtiene normalmente de simulaciones de modelos de mesoescala. En este trabajo se presentan diferentes técnicas para corregir estas simulaciones de modelos de mesoescala, utilizando datos de torres de medida situadas en el propio parque. En concreto, se proponen diferentes métodos heurísticos para ajustar mediante superficies los parámetros Weibull de la serie de viento del modelo de mesoescala con los de la serie de viento real medida por las torres. Se muestra el buen rendimiento obtenido por los algoritmos propuestos en dos parques eólicos en España.

Por último se propone un novedoso algoritmo para el posicionamiento óptimo de las turbinas en un parque eólico. Se considera un modelo de parque eólico más realista, incluyendo orografía del terreno, forma del parque eólico, simulación de velocidad y dirección de viento y costes de instalación y conexión entre turbinas. Este trabajo propone la aplicación de un algoritmo heurístico que obtiene una solución inicial razonable para el problema. Esta solución se utilizará para inicializar la población inicial de un algoritmo evolutivo. Se muestra que el algoritmo evolutivo inicializado con la solución inicial del heurístico, obtiene muy buenos resultados, maximizando el beneficio económico.

A person who never made a mistake never tried anything new

- Albert Einstein

Acknowledgements

This work has been partially supported by the project TIN2014-54583-C2-2-R of the Spanish Ministerial Commission of Science and Technology (MICYT), and by Comunidad Autónoma de Madrid, under project number S2013ICE-2933_02. I also acknowledge the funding from Universidad de Alcalá in the period 2013-2016, through a FPI grant.

Agradecimientos

Parece mentira que esté aquí escribiendo los agradecimientos de esta tesis. Que después de estos años, tanto esfuerzo, quebraderos de cabeza y satisfacciones, por fin la línea de meta esté tan cerca.

Sería injusto por mi parte no dedicar unas palabras a todos aquellos que de algún modo u otro me han ayudado a llegar hasta aquí. También es la parte más difícil de la tesis, sintetizar en unas pocas líneas toda la gratitud que siento.

Me gustaría empezar con Dr. Sancho Salcedo, codirector de mi tesis. Han sido muchos años juntos y no puedo expresar aquí toda mi gratitud, por tanta dedicación y entrega, por su disposición absoluta, por contagiarme sus ganas de profundizar en el mundo del *Machine Learning* y animarme cada día a completar mi tesis. No puedo dejar de nombrar a mi otro codirector, Dr. José Antonio Portilla, que siempre me ha contagiado sus ganas de trabajar y de perfeccionamiento. Además, aunque no haya participado directamente, Dra. Silvia Jiménez siempre ha estado ahí, dándome consejos. También me gustaría agradecer a todos los miembros del grupo GHEODE, y a aquellos que en algún momento pasaron por aquí, y que no nombraré por miedo a dejarme a alguno, porque en mayor o menor medida, todos me han aportado algo y me han ayudado en este camino.

A mis padres y mi hermana, que siempre me han apoyado y me han animado a seguir adelante. Y por supuesto, a Arturo, que me da fuerza y me apoya en mis momentos de dudas.

Bea

Contents

List of Figures

List of Tables

Part I

Motivation, objectives and state-of-art

Chapter 1

Introduction

Existing reserves and undiscovered sources of fossil fuels and other combustible or radioactive minerals will eventually be exhausted or too expensive to extract in the mid-term future. Therefore, new energy sources are needed, and that is here, where renewable energies play a very important role. The use of renewable energies is also linked with environmental issues. As the Intergovernmental Panel on Climate Change (IPCC) and the International Energy Agency (IEA) point out, CO_2 and other greenhouse gas (GHG) emissions must peak and begin to decline in less than a decade if we plan to have any chance of meeting the target of staying below 2° C of global mean temperature rise - the target to which the 192 member governments of the UNFCCC (United Nations Framework Convention on Climate Change) have committed themselves to.

1.1 Some notes on wind energy economy

Due to climate change and in order to meet the established targets, a transformation on the global energy system is needed. This energy transformation is becoming already a fact: renewable energy provided an estimated 19.1% of global energy consumption in 2013 (Figure 1.1) and continued to grow in 2014, despite the dramatic decline in oil prices. Out of this total share in 2013, modern renewables accounted for approximately 10.1%, with the remainder coming from traditional biomass [142].

Since 2009 installed renewable capacity grew at rapid rates. Over this period, solar photovoltaics (PV) experienced the fastest capacity growth rate of any power generation technology, while wind experienced the most power capacity added of any renewable technology. Besides, there was a number of significant and positive developments: wind power was deeply developed in Africa and Latin America; solar thermal power has shifted its focus to the Middle East and North Africa region and to South America and solar PV has continued to grow around the world. It is evident then, that renewables are no longer in hands of a small bunch of countries. Major renewable energy companies focused their efforts in Africa, Asia and Latin America, where strong new markets are emerging. Renewable energy is considered crucial for meeting current and future needs. In developing countries, it can provide and expand access to modern energy services. On the other hand, a growing number of cities around the world aim to switch to a 100% Renewable energy system (See Table 1.1)

According to the Medium-Term Market Report (MTRMR) 2014, published by the International Energy Agency (IEA) [121], it is expected that the deployment of renewable technologies

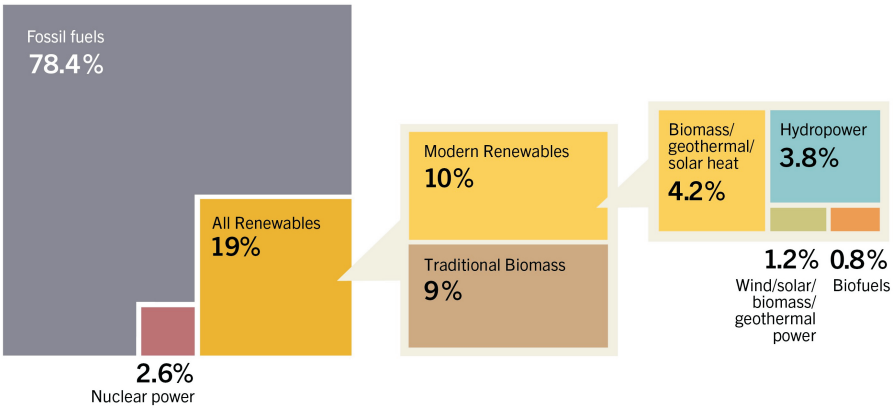


Figure 1.1: Estimated Renewable Energy Share of Global Final Energy Consumption. *Source: REN21.*

Table 1.1: Cities and Countries moving towards 100% RE in electricity systems.

Country	*Target*
Iceland (OECD)	Produces 100% of its electricity from hydropower and geothermal energy.
Cook Islands (Small Island State)	50% by 2015 and 100% by 2020 renewable energy goal.
Costa Rica (Central America)	95% renewables goal for 2014, mostly from indigenous hydro resources.
Denmark (OECD)	More than 50% of its electricity supply with renewables by 2020, 100% of electricity and heat by 2035, and 100 per cent in transport by 2050.
Maldives (Small Island State)	Completely carbon neutral by 2020. This entailed embracing an almost 100% renewables based energy system.
Scotland (OECD)	100% renewable power supply by 2020.
Tokelau (Small Island State)	100% of its electricity from renewable energy. Now Tokelau produces over 10% of its electricity from solar energy.
Tuvalu (Small Island State)	100% of its electricity from renewable energy by 2020.
Tasmania (Australian Territory)	100% renewables by 2020 and a 35% reduction in emissions.
City	*Target*
Sydney (Australia)	100% of the City's electricity, heating and cooling from RE sources by 2030
Malmo (Sweden)	100% RE by 2030
Greensburg, Kansas (US)	Today the town sources a 100% of electricity from a 12.5 MW wind farm.

over the medium term rises. Global renewable energy generation is projected to grow from 5070 TWh in 2013 to 7315 TWh in 2020 (Figure 1.2). China should be responsible for the biggest

part of this growth within most categories.

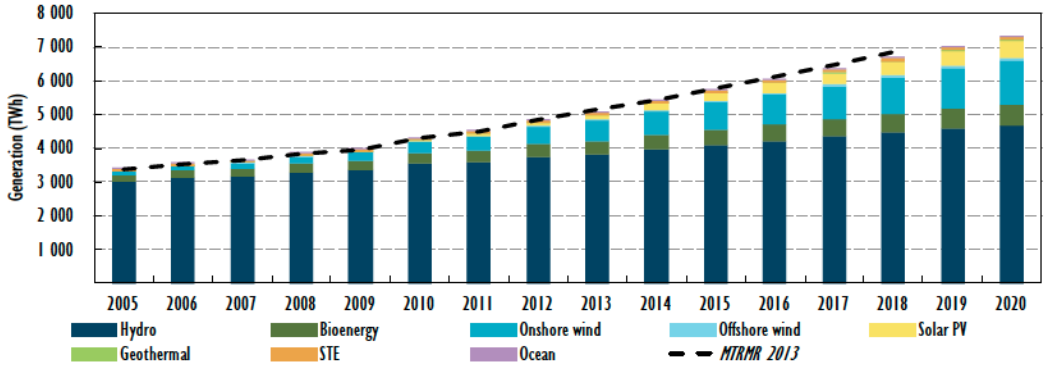


Figure 1.2: Global renewable electricity generation. *Source: MTRMR 2014.*

In 2020, the number of countries with cumulative renewable electricity capacities above 100 MW is expected to notably increase. Onshore wind is expected to be deployed in nearly 75 countries by 2020. The deployment of solar PV is expected to reach 75 countries by 2020, and bioenergy should reach 60 countries by 2020, up from 50 in 2013. The spread of offshore wind, CSP, geothermal and ocean deployment should remain relatively lower (Figure 1.3).

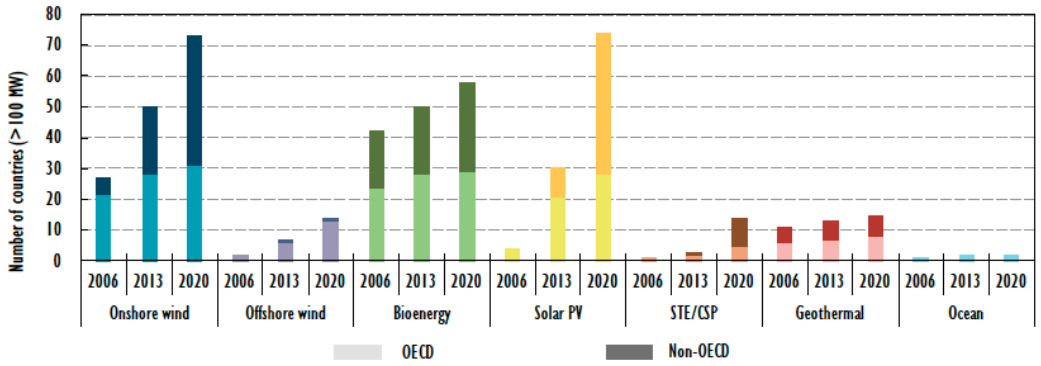


Figure 1.3: Number of countries with non-hydro renewable capacity above 100 MW. *Source: MTRMR 2014.*

1.1.1 The Global Status of Wind Power

According to the Global Wind 2014 Report [66], 2014 was a record year for the wind industry as annual installations crossed the 50 GW mark for the first time. More than 51 GW of new wind power capacity were installed, a sharp rise in comparison to 2013, when global installations were just over 35.6 GW (Figure 1.4).

The new global total capacity at the end of 2014 was 369.6 GW, representing cumulative market growth of more than 16 percent (Figure 1.5). China, the largest wind market since 2009, had a good year and retained the top spot in new installed capacity in 2014. The majority of

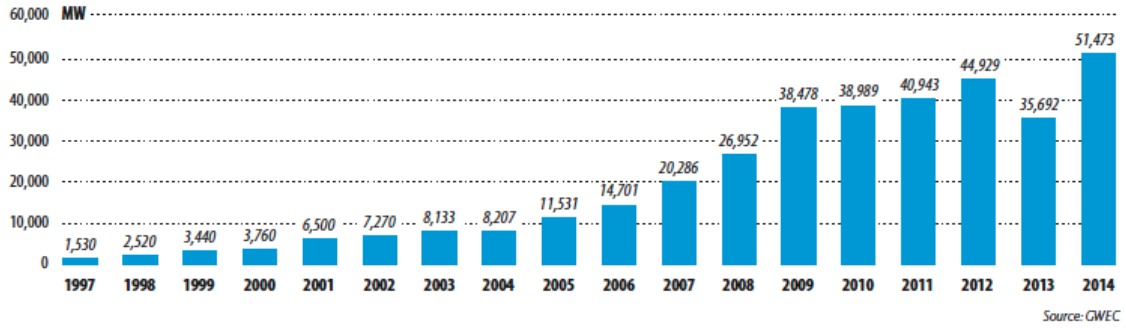


Figure 1.4: Global annual installed wind capacity 1997-2014 *Source: GWEC 2014*

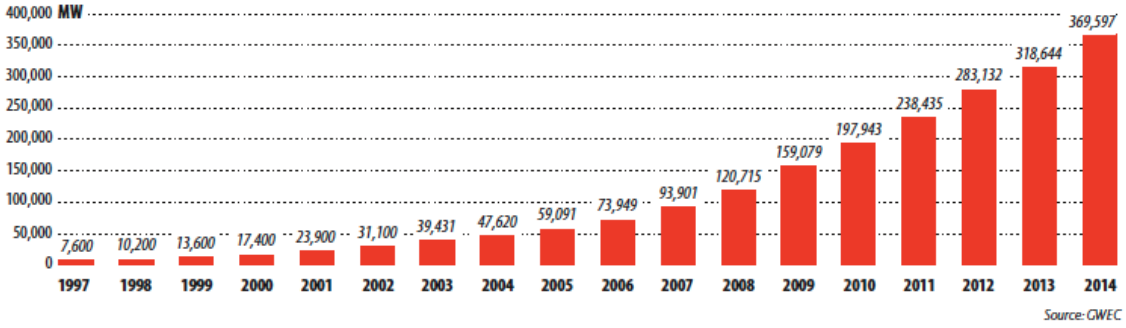


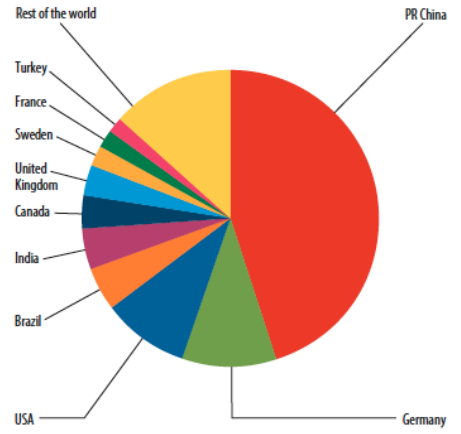
Figure 1.5: Global annual installed wind capacity 1997-2014 *Source: GWEC 2014*

wind farm facilities were installed outside the OECD[1].

The big five markets - China, USA, Germany, Spain and India, still represent a 72% share of global wind capacity. The Chinese market crossed the 100 GW mark, adding another milestone to its already exceptional history of renewable energy development since 2005 [66].

China controls 31% of the global installed capacity, adding 23 GW in 2014. Germany added 5.3 GW anticipating the changes in the renewable energy legislation, which may lead to a slow-down of the German market in the coming years. For the first time, Brazil has entered the top group by becoming the third largest wind market for new capacities, with 2.5 GW, representing 5% of all new wind capacities. India kept the second Asian position, with 2.3 GW of new wind capacity. The Spanish market, however, has not contributed to the overall growth in 2014, with only 0.1 MW of new wind capacity added [185].

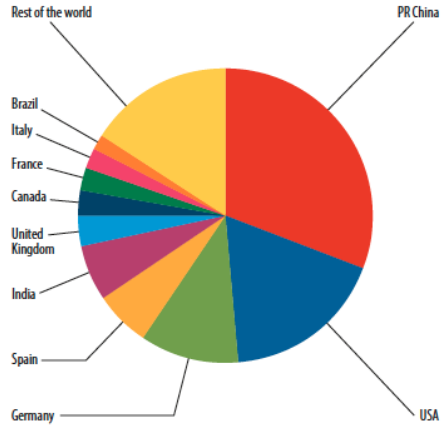
[1]Organisation for Economic Co-operation and Development. Members: EU-28, except Bulgaria, Croatia, Cyprus, Latvia, Lithuania, Romania and Malta; Australia, Canada, Chile, Israel, Japan, South Korea, New Zealand, Mexico, Turkey, Switzerland, Norway, Iceland, United States of America.



Country	MW	% SHARE
PR China	23,196	45.1
Germany	5,279	10.2
USA	4,854	9.4
Brazil*	2,472	4.8
India	2,315	4.5
Canada	1,871	3.6
United Kingdom	1,736	3.4
Sweden	1,050	2.0
France	1,042	2.0
Turkey	804	1.6
Rest of the world	6,852	13.3
Total TOP 10	**44,620**	**87**
World Total	**51,473**	**100**

** Projects fully commissioned, grid connection pending in some cases* *Source: GWEC*

(a)



Country	MW	% SHARE
PR China	114,609	31.0
USA	65,879	17.8
Germany	39,165	10.6
Spain	22,987	6.2
India	22,465	6.1
United Kingdom	12,440	3.4
Canada	9,694	2.6
France	9,285	2.5
Italy	8,663	2.3
Brazil*	5,939	1.6
Rest of the world	58,473	15.8
Total TOP 10	**311,124**	**84.2**
World Total	**369,597**	**100**

** Projects fully commissioned, grid connection pending in some cases* *Source: GWEC*

(b)

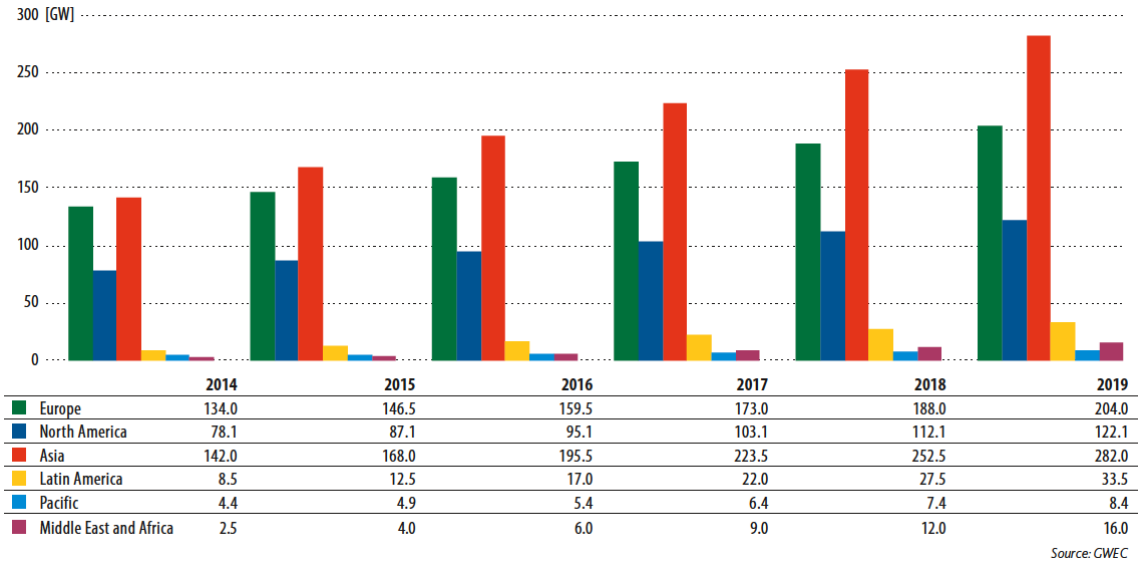
Figure 1.6: (a) Top 10 new installed capacity Jan-Dec 2014; (b) Top 10 cumulative capacity Dec 2014. *Source: GWEC 2014.*

1.1.2 Wind market forecast for 2015-2019[1]

All industry experts saw wind energy as a very strong market for decades [141]. One of the the most optimistic projections was 1000 GW globally by 2020. Another expert saw 50% of global electricity from wind by 2050. Some utilities are equally optimistic. For example, German RWE said that "wind power is well on its way to becoming competitive even in a non-regulated market". And Spanish Gas Natural Fenosa said that "wind energy is one of the most mature renewable technologies and is the most widespread and has the greatest growth horizon world-wide". According to the Global Wind Energy Comission [66], Asia will continue to dominate and having surpassed Europe in terms of cumulative installed capacity at the end of 2014, it will continue to lead markets with 40-45% of the annual global total. Cumulative growth will be near 15% in 2015, but it is expected to be in average 11-13% from 2016 to 2019. Total installation should almost double today's capacity, going from 370 GW to just about 670 GW by the end of 2019 (Figure 1.7).

On the other hand, the potential of off-shore wind energy remains high, but technical, financial and grid connection issues pose challenges to its deployment. The reliability of offshore turbines needs to be improved. To do so, more robust turbines specifically designed to operate in offshore conditions and floating turbines to operate in deep waters are needed. Despite the lack of nowadays reliability, wind offshore capacity is expected to overcome these difficulties and

[1]Note that at the time of writing this Ph.D. Thesis, most recent reports on Wind Energy were from 2014.



	2014	2015	2016	2017	2018	2019
Europe	134.0	146.5	159.5	173.0	188.0	204.0
North America	78.1	87.1	95.1	103.1	112.1	122.1
Asia	142.0	168.0	195.5	223.5	252.5	282.0
Latin America	8.5	12.5	17.0	22.0	27.5	33.5
Pacific	4.4	4.9	5.4	6.4	7.4	8.4
Middle East and Africa	2.5	4.0	6.0	9.0	12.0	16.0

Figure 1.7: Cumulative market forecast by region 2014-2019 *Source: GWEC 2014.*

expand globally. According to EWEA, there are twelve projects currently under construction, that will add 2.9 GW, bringing cumulative capacity in Europe to 10.9 GW in 2015-16. Beyond 2016, EWEA has identified future offshore wind farms, resulting in more than 98 GW.

1.1.3 Wind energy in Spain

Spain is the fourth country in the world in terms of wind energy capacity installed. In fact, it has a history of continuous growth during the past few years (Figure 1.8).

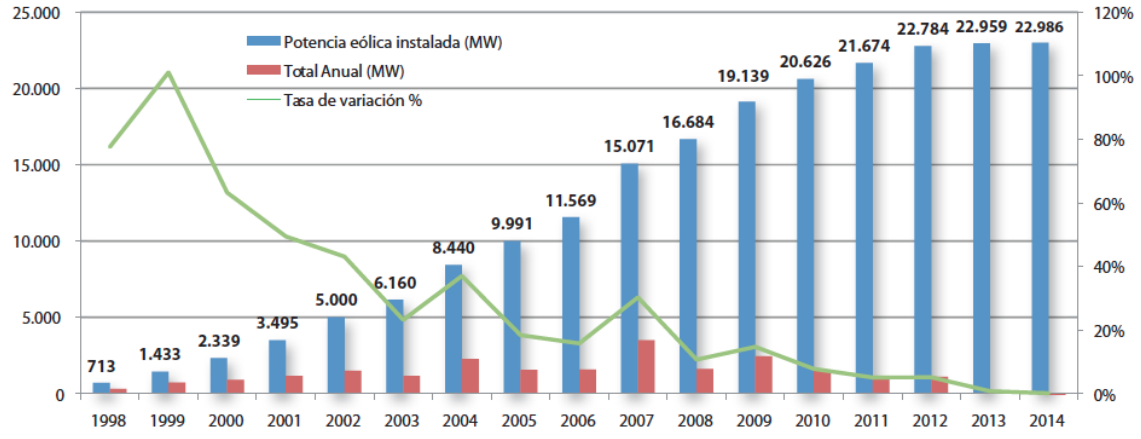


Figure 1.8: New and cumulative annual capacity and variation rate in Spain 1998-2014. *Source: AEE.*

However, the year 2014 will be remembered as the worst year for the wind sector. The energy reform distanced new possible investors from Spain. The short term therefore offers little hope for the wind power sector given the lack of objectives for 2020, the lack of incentives and the

newly imposed taxes that do not apply to other technologies. New installed capacity reached the minimum of the latest 20 years, with only 27 MW installed, 14 MW out of them corresponding to a new wind farm in Galicia. On the other hand, according to REE, 42% of the electricity demand in 2013 was covered by renewable energies. In fact, wind energy covered 21.1% of the Spanish electricity demand in 2013, reaching a historical record. For the first time, wind energy was the technology with the highest share in electricity demand (Figure 1.9) and Spain became the first country in the world, in which wind energy was the main energy power. In 2014 however, wind energy covered 20.3% of the electricity demand. According to the preliminary report of 2015, coal-fire generation, with a contribution of 20.3% to demand coverage, placed second in the generation mix, displacing wind energy to third place with 19.1% of the total share.

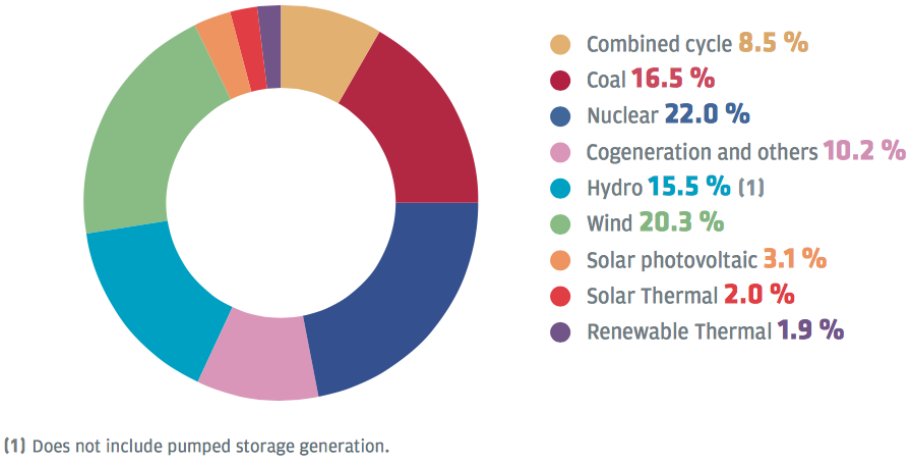


Figure 1.9: Spanish electricity demand coverage 2014. *Source: REE.*

1.2 Wind farms

The design of a wind farm and its later operation is an extraordinarily complex task, in which many different parts participate: electrical and electronic engineering, mechanic and aeronautical engineering, topography, meteorology, environmental sciences, as well as economics and law. Simplifying, there are three key factors to consider in the construction of a wind energy facility [26].

- *Availability of wind (Wind resource).* A site must have a high annual average wind speed to guarantee a certain amount of generated energy, and a low level of turbulences.
- *Availability of existing transmission lines.* Whenever possible, access to existing lines should be considered to avoid the installation of new high voltages lines, that can cost thousands of euros per kilometer.
- *Access to the site.* The construction of a wind farm needs the use of heavy industrial equipment and turbine components have to be transported to the site. Therefore, a good road infrastructure is needed.

While every project is unique, the life cycle of a wind farm generally has the following steps, in which the key factors explained above should be taken into account [4, 32].

1. Development.

   (a) Site assessment. As a first step, a study of a potential site is undertaken to know its suitability, looking at wind speed (using databases, or data from near airports, for example), landscape features (orography), environmental issues and proximity to grid lines. It is particulary useful here the use of Geographical Information Systems (GIS), given the amount of information and considerations involved. There are some articles in the literature, that deal with GIS systems to assess and evaluate the wind potential of an area [72, 83, 102]. This is a *pre-development* phase, which ends with an agreement with the landowner.

   (b) Feasibility studies. One of the most important steps, once the site is chosen, is the study of the wind resource to confirm the initial wind speed assessment. To do so, typically, measuring towers are installed on the site and data is gathered for 12-24 months. Sometimes these towers have technical failures, so the gathered data may have gaps. Other times the developer cannot afford such long measuring campaigns and only three or six months of data are available. Therefore, a good analysis of wind speed data, that deals with lacks of data or short data series, is needed to study the trend of the wind resource. Besides, there are other feasibility considerations to take into account: if there are endangered or protected species in the area, if noise and aesthetics will be an issue for locals, if the site's geology is appropriate for industrial development, if the turbines will obstruct the flight path of air traffic, etc.

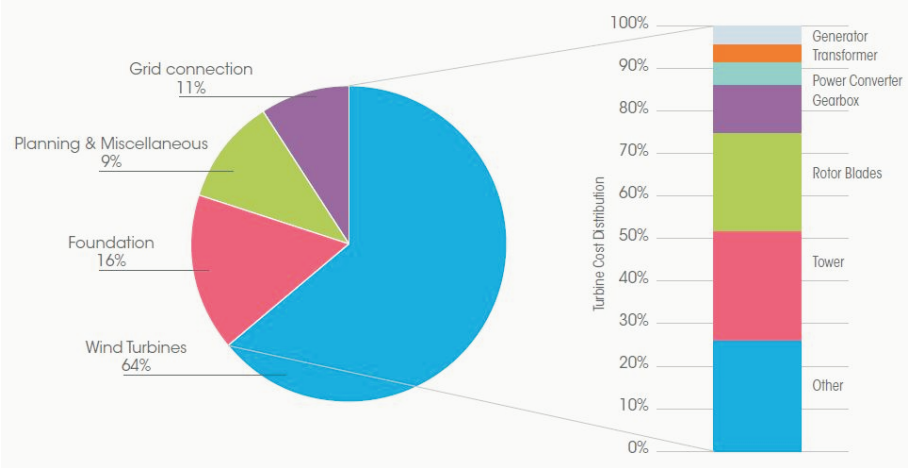


Figure 1.10: Capital cost breakdown for a typical onshore wind farm. *Source: IRENA*.

On the other hand, financing is another key point in terms of feasibility. First of all, developers need to make a cost analysis. Wind energy is capital intensive, but has no fuel costs. The key parameters governing wind power economics are:

- Investment costs (including project financing).
- Operation and maintenance costs (fixed and variable).
- Capacity factor (based on wind speeds and turbine availability factor).
- Economic lifetime.
- Cost of capital.

The installed cost of a wind power project is dominated by the upfront capital cost for the wind turbines (Figure 1.10). These high upfront costs can be a barrier to their development, despite the fact that there is no fuel price once the wind farm is built. The capital costs can be divided into turbine cost, civil works (including construction for site preparation and foundation for the towers), grid connection costs and other costs, that include the construction of buildings, control systems, etc.

Altough wind energy is capital intensive, it is one of the most cost-effective renewable technologies in terms of the cost per kW of electriciy generated. The levelised cost of energy (LCOE) is the main metric for describing and comparing the economics in energy projects. In wind energy, the LCOE represents the sum of all costs of a fully operational wind power system over the lifetime of the project with financial flows discounted to a common year. The principal components of the LCOE of wind power systems include capital costs, operation and maintenance costs and expected annual energy production (Figure 1.11). If the reader is interested in a deeper analysis, the International Renewable Energy Agency (IRENA) provides a series of cost analysis of different renewable energy technologies, with a specific study on wind power cost [80]. Further details are not included because cost analysis is beyond the scope of this PhD. Thesis.

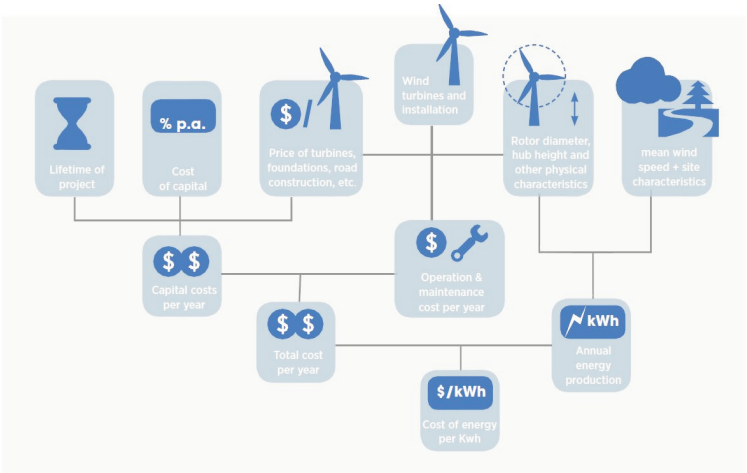


Figure 1.11: The economics of wind systems. *Source: IRENA.*

(c) Construction. Once the permits and licenses have been obtained, it is needed to project the location of the turbines within the wind farm. It is usually made in a heuristic way, with trial and error. With this preliminary layout, developers use a commercial software for micro-sitting [176, 181, 182] or novel computational intelligence techniques to evaluate the potential power obtained with the layout, considering the orography and wind speed distribution of the wind farm. Then, slight modifications are made by trial and error to adjust and maximize the power output. With the layout established, construction works start: access roads, turbines foundations, transportation of equipment and turbines, etc. In parallel with this, an agreement with the local electricity distribution company is made to connect the wind farm to the grid.

2. Operation and management. Modern wind farms are fully automatic and are managed using remote telemetry. Maintenance requirements are minimal and are carried out by specialists. However, it could be interesting the development of a failure predicting system, to anticipate and manage more efficiently failures and breakdowns. Besides, in some countries, governments or electricity companies require wind power prediction for the next few hours, in order to balance the electricity system based on offer and demand.

1.3 Objectives of this Ph. D. Thesis

This Ph. D. Thesis tackles several problems related to different aspects of the wind farm design previously described. Specifically, the following problems are the core of this work:

- Wind speed reconstruction, based on synoptic pressure values, rather than past wind speed values. It is useful to study the wind resource in the area, when only measurements for a short period of time have been taken. A relationship between wind speed and past synoptic pressure values is achieved by applying different soft-computing techniques.

- Wind speed reconstruction based on wind speed values from neighbour towers. Measuring towers usually suffer technical problems and gathered data have gaps. In this part of the work, we propose several algorithms for missing values reconstruction using the available information from neighbour measuring towers. Different techniques such as fast training neural networks or support vector machines will be analyzed.

- Modification of values from a mesoscale model using information from measuring towers. Mesoscale models are used in wind farm prospection to study the spatial distribution of wind speed. However, they tend to have a bias and data from mesoscale models differ from real data measured by towers. To solve this problem, an heuristic correction of the mesoscale model values is proposed using on-site measuring towers.

- Wind farm design. A novel approach to design the layout of the turbines is proposed, using computational intelligence techniques and including parameters like wind farm shape, orography and costs, to improve the profit of the installation.

1.4 Structure of this Ph. D. Thesis

This Ph. D. Thesis is structured as follows:

- In Chapter 2 the State of the Art of each problem tackled in this work is reviewed, giving references and examples of previous works in the literature.

- Chapter 3 includes the description of the different soft-computing techniques applied to solve the problems.

- Chapters 4, 5, 6 and 7 contain the work related to each of the proposed problems. Each chapter is self-contained and includes a general description of the problem, the chosen theoretical approach, the results obtained and some final conclusions. When necessary, appendices to tables are included.

- Finally, some general remarks and possible lines of work are included. Besides, references from the scientific publications produced as a result of this research work are provided.

Chapter 2

State-of-the-art

This chapter follows the structure of this Ph. D. Thesis. The objective is to follow the line of the whole process of installing a new wind farm facility. This process implies a huge variety of tasks in order to have a detailed study of the wind at the given location, as it has been shown in Chapter 1. Thus, in the next sections we review some previous works on indirect wind speed reconstruction from synoptic pressure patterns, direct wind speed reconstruction from measured data, spatial reconstruction of wind speed with mesoscale models and intelligent design of wind farms.

2.1 Indirect models based on synoptic pressure for wind reconstruction

Wind speed series reconstruction is an important problem currently faced by companies exploiting wind farms. Basically this problem is usually faced by obtaining a model for characterizing the wind speed based on previous real wind measures and then, applying it to obtain values in the past in order to reconstruct wind speed series. Different techniques have been used to obtain these wind speed models, such as statistical methods [50, 90, 105, 117, 166], neural networks [52, 103, 104, 110, 114], support vector machines [113] or hybridization of some of these algorithms [17].

The majority of the existing techniques to reconstruct wind speed series (and also for long-term wind speed prediction problems) are based on past wind speed data [21], and some of them include other atmospheric variables such as local temperature, radiation or pressure at the measuring point. The main problem with this approach is that these prediction variables are not always available for all the places, so it is sometimes difficult to extrapolate the current techniques or studies to new locations. This problem with local measures is common all over the world, so the idea of considering synoptical information combined with local information has been of interest in the last few years. In fact, this idea has been successfully applied in a huge variety of prediction problems: pollution [28, 129], ozone levels [29, 189], precipitation [133, 147, 167], temperature [132], forest fires [135] , dust storms [45], car accidents [96]. These works tackle the relationship between synoptic pressure patterns and the variable under study in lots of different ways: support vector machines, weather types approaches or simply statistical analysis, among others.

Since synoptic information (mainly atmospheric pressure) is available all over the world and there are reliable records of synoptic pressure fields back to more than one century ago, it is

a great source of information for wind speed, because it has also been proved that the wind at a given point is a direct function of the pressure gradient over time. Therefore, different works tackle the problem of wind speed prediction using synoptic pressure data. In [161], the authors propose a method to automatically classify objective synoptic processes, not to predict wind speed, but to study the mesoscale atmospheric circulatory patterns. Surface pressure and geopotential height data at 500 hPa are studied, considering three day groups. The method is based on the minimization of the distance among the groups, obtaining an average pressure and geopotencial maps for each of the groups. Finally, a high-resolution mesoscale model is run, with the characteristic maps (average pressure and geopotential maps calculated) for each situation (group), so circulatory patterns are obtained for a certain synoptic situation.

In another different work, [71], an approach to model wind speed data using atmospheric pressure by means of hidden Markov models (HMMs) is developed. The authors state that the correlation between atmospheric pressure and wind speed is not sufficient to build a nonlinear predictor with past and current wind speed and pressure data, as most of the works try to establish. Therefore, a HMM is used, to exploit the dependencies between wind speed and pressure to be able to build an accurate wind speed model. It is worth noting that a single value of atmospheric pressure in the location under study is used as input data in the HMM. Experimental results show the performance of the proposed model: consistent predicted values, in accordance with the real measured data, and parameters to model accurately the wind speed distribution.

In [22, 23] an evolutionary algorithm to carry out the synoptic pressure clustering for wind speed reconstruction is proposed. The algorithm operates in a search space formed by grids of pressure measurements, and classifies the different pressure situation into classes. Each class has a mean wind speed and wind direction assigned, so the wind speed reconstruction is possible for a new grid of synoptic pressures, by classifying it. A matrix of synoptic pressure data around the location is used, specifically, a matrix of 182 values (14×13) surrounding the Iberian Peninsula is considered. The authors propose a way to reduce the number of points in the grid and encode the information for the evolutive algorithm to be used: differences of pressure between points of the grid, specifically, 4 pressure differences (8 points) are considered. The evolutionary algorithm tries to find the best possible encoding of the synoptic situation in order to classify them and obtain the most concentrated clusters, to predict wind speed in the most accurate way. Both works operate in a similar way, but [22] reconstructs wind speed module while [23] reconstructs wind roses. In the reconstruction of wind speed module, a comparison with a weather types approach is considered. In the reconstruction of wind roses, two different objective functions are proposed.

Another recent and interesting work is [64], that tackles the problem of wind speed estimation in a very different way from the previous works. In this work, wind speed is considered as a discrete variable. Thus, the wind speed is discretized into different levels of wind and the problem is now a classification problem. According to the authors, in some cases it is not necessary to know the exact wind speed value, it is enough with a general idea of the level of speed for the manager to set functional operations in the wind farm. Daily wind speed is discretized in four different classes and a synoptic pressure grid is considered. The objective of that work is to test several methods to compare nominal and ordinal classification. Also a comparison with the hidden Markov models in [71] is given. In order to reduce the curse of dimensionality (synoptic pressure grid has a high number of variables), a principal component analysis is applied, so as to reduce the inputs for the different algorithms. The results obtained in the work show that the

best method is support vector machines, while ordinal and nominal classifiers do not show big differences.

2.2 Direct models for wind reconstruction from real measured data

The analysis of wind speed in the area under study can also be carried out by installing one or several wind measuring towers in the area of interest to take measures of wind speed and direction, over a time period, long enough to obtain revealing results. In [38] and more recently in [100] these measures are then used to evaluate the best sites to place a wind farm. In other works these measures are applied to the wind speed reconstruction using two-sites correlation models [154], or to estimate mean wind speed from short-term data [101]. Note that these measuring stations often have different operational or technical problems, so in many cases the measured wind speed series are incomplete in some periods. Therefore, the estimation of missing values in wind speed series at a particular site, i.e. the reconstruction of a wind speed time series, is an important problem in wind farms site prospection [35].

On the one hand, very short-term wind speed prediction is an important problem in wind farms since in countries like USA hourly energy prediction is required. On the other hand, some previous works have tackled long-term prediction in wind farms using wind speed data from measuring towers. In [11] the wind speed prediction from reference stations is carried out using neural networks, and applied to wind energy evaluation in Turkey. The authors use data from 8 stations and perform the prediction of each measuring stations using data from some other reference stations. It is concluded in this work that the correlation factor between the target station and the reference stations used to obtain the prediction is of extremely importance. Thus, the prediction is made using data from the stations with the highest correlation factor. In [25] a probabilistic method using measures from a reference site is used to estimate long-term wind speed in the Canary Islands, Spain. The same authors also carried out different studies on long-term wind speed prediction from reference sites using neural networks [170, 171]. In [170] mean hourly wind speeds and directions from twenty two measuring stations are used. The authors conclude that the estimation errors tend to decrease when the number of reference measuring towers used increases and the correlation factor between these reference towers and the objective tower does not matter. It is also concluded that, when both the wind speed and the direction are used as input signals, the number of measuring towers needed to obtain a certain estimation error is lower than when only wind speed is used. On the other hand, [171] focuses on the estimation of the cost per kWh produced so wind speed, electrical energy and energy cost are analysed. As it can be seen, wind speed prediction and reconstruction of wind series from in-situ measuring towers is currently a hot topic in wind energy research, covered under the generic name of Measure-Correlate-Predict methods (MCP).

Many different techniques have been applied to wind speed prediction or reconstruction from MCP: traditional ones, such as linear MCP techniques have been applied in different studies of wind farm site assessment [118, 162] or wind speed reconstruction [146], in which a variety of sites (offshore, coastal, complex terrain) and different Measure-Correlate-Predict methods are compared. Non-linear approaches have also been considered to this end, including different computational intelligent methods like neural networks. In [87] a review of neural networks applied to renewable energy system is provided: solar water heating systems, photovoltaic systems,

solar radiation and wind speed prediction are tackled. Focusing on artificial neural networks for wind speed prediction, a lot of works can be found in the literature in which, specifically, multi-layer preceptrons have been applied to estimate wind speed in different countries such as Turkey [11, 130], Spain [106] or Nigeria [52], among others. It should be pointed out that in [106] the authors showed the importance of using the wind direction as an input in complex terrain locations. In the work [60], a two hidden layer neural network is proposed to carefully predict the wind energy output. To do so, not only the authors use monthly mean wind speed data from measuring towers, but also humidity, temperature, generation hours and maintenance hours as inputs of the neural network. The novelties proposed by the authors are the use of a two hidden layer neural network with two different nonlinearities (i.e., the hyperbolic tangent function and the logarithmic sigmoid function) and the consideration of five input signals that enable a more accurate energy prediction. Some other works propose the usage of Support Vector Machines [113], Abductive networks [1], Bayessian networks [24], or generalized mapping regression [8]. Usually the performance of each method has been evaluated on different real data from wind farms. Besides there are also several works comparing different MCP techniques [148].

In spite of this huge work in MCP processes and algorithms, the majority of articles do not cover real MCP operations (RMCPO), but refined versions on the problem, in which data have been treated to consider complete sets, where all the data from reference towers are available. A RMCPO problem consists of tackling the raw reconstruction or prediction problem from measuring stations data, i.e., training a large number of reconstruction or prediction models from the existing data (with missing values). Since huge amounts of data are received in a control center managing several wind farms, RMCPO are dynamic processes which often require the continuous re-training of the algorithms to consider new data. Thus, very fast training approaches must be used in RMCPO problems, and that is why linear regression approaches are usually applied in industry for RMCPO problems.

2.3 Spacial reconstruction of wind speed with Mesoscale models

In order to obtain the wind speed trend over the years, regression methods are usually applied. The objective is to extend the series of the measured wind speed back to previous years. This process can also be used to obtain long-term wind speed estimations, which will help support the appropriateness of the area when it is presented to investors. There are many works in the literature presenting regression techniques for this issue (Section 2.2). Estimation of wind speed can also be tackled using indirect measurements from data acquired from Reanalysis or proxies such as synoptic pressure patterns (Section 2.1). Besides, regarding the specific geographical distribution of surface wind speeds over the area of study, Computational Fluid Dynamics (CFD) simulations are usually carried out, but in large complex areas simulations by mesoscale models provide better results [5] and are quite often employed. The methodology to use mesoscale models consists of considering a grid of points, defined over the area under study, and the wind speed distribution, calculated by means of the mesoscale model. This model is often initialized from global-scale models outputs and fed with local information and/or parameterizations. The main problem with this approach is that mesoscale models can produce wind speed data with some drift, due to the characteristics of the model such as its resolution and discretization process. Therefore, it is important to include different corrections to the models in order to obtain surface wind fields as similar to real wind as possible.

Recently, two works have proposed the statistical correction of the surface wind speed from a mesoscale model using measuring stations [123, 124]. In those papers, a statistical correction of the geostrophic monthly wind speed is carried out through linear transformation of the mesoscale model data in a grid in Iceland. The rescaling factor and offset in the correction are determined at measuring stations by comparison with measurements. The values of the rescaling factor and offset are then interpolated in the model grid through distance-weighted horizontal averaging. These works have not been proposed to be applied in wind farm prospecting, but to improve wind speed fields analysis for Iceland, in the frame of meteorological analysis of the area.

Another recent work tackling mesoscale models correction is [93]. The main difference with the previous commented work, is that this one does not use data from measuring stations to correct mesoscale models, only to validate their proposed method. Instead, satellite-derived variables are used: land surface temperature (LST) derived from satellite thermal infrared signals and normalized difference water index (NDWI). The authors state that these two variables can represent the land surface characteristics corresponding to temperature and moisture conditions. Simulated values of the mesoscale models matching the measuring station location are corrected applying a multiple linear regression in order to include the LST and NDWI parameters. Data from 74 meteorological stations operated by the Korean Meteorological Administration are used to check the performance of the proposed algorithm. An analysis between simulated and measured data is done, before doing any correction. Then, the correction is applied and compared again with measured data. Overall, corrected wind speed values show stronger correlation with the observed values. The authors suggest that their method can improve the accuracy of mesoscale models in the Republic of Korea. The main advantage is that measuring towers are unnecessary to carry out the correction of mesoscale models, they are needed just to validate and check the performance of the method, while the technique proposed in [124] needs measuring towers. Therefore, [124] can be used in wind farms studies, where measuring towers will be available and the method proposed in [93] is a more general method, to be applied in a huge area where no real data is available.

Although correction of mesoscale models is a very interesting field and can be a very resourceful tool as shown in the cited works [93, 124], it is not an exploited issue in the literature.

2.4 Wind farm design

When a completed study of wind resource in the location under study, and once the economical viability of the wind farm facility has been checked, it is time for the engineers to design the wind farm. Automated wind farm design is a topic gaining popularity in the last few years. Therefore, there is an increasing number of articles tackling this problem successfully applying computational intelligence techniques. Evolutionary algorithms are mainly applied, although other approaches have also been used.

The seminal paper in the use of evolutionary computation techniques for wind turbines layout in wind farms is the work by Mosetti et al., [119]. This paper proposed a genetic algorithm to tackle the problem of the optimal positioning of turbines in a wind farm. The model proposed in [119] consists of modeling the wind farm as a square divided into cells in which turbines can be situated. A new wake model was proposed and several experiments considering different average wind speed and direction were presented. This initial work has been the base of different recent

approaches which have improved the initial model. For example, in [61], the authors showed that better results can be obtained in the problem by improving the genetic algorithm used, using the same model as in [119]. Another improvement with the same model has been recently proposed in [48]. This paper proposes a modification of the objective function of the problem, to take into account deployment cost and efficiency of the turbines. The authors show that this modification leads to better design results than previous approaches using a standard genetic algorithm.

Another interesting and recent work, including a different optimization model is the work described in [144]. In this work a variable-length genetic algorithm with novel procedures of crossover is applied to solve a problem of optimal positioning of wind turbines considering monetary cost as the objective optimization function. The authors show that their variable-length evolutionary approach is able to obtain good results in terms of the objective function considering different types of wind turbines. A similar approach using a hybrid evolutionary algorithm was previously presented in [109]. This approach has been further studied in [155, 156]. The work in [116] also proposed a genetic algorithm with variable size chromosomes, in which each solution uses a different number of wind turbines.

It is also significant the work proposed in [175], in which the authors proposed new improved models for wind and turbines. Specifically, a more elaborated wind speed simulation is considered based on a Weibull distribution. The authors have shown that this new model is able to produce better results than previous approaches in the literature. Also following this trend, [98] proposes to incorporate different sophisticated models in different parts of the problem. Specifically, a complete study of the problem including different costs of turbines and their maintenance, a wind turbine wake model similar to the one described in [119], a Weibull distribution for modelling the wind speed and direction in each point of study are considered. The authors propose then an evolutionary programming approach to solve the continuous optimization problem of optimal positioning of wind turbines using these novel models.

In the last few years other approaches to wind farm design based on evolutionary algorithms and related techniques have been published, such as the work in [173], based on real-coded genetic algorithms. The use of real encoding is interesting in wind turbines positioning, since it allows to improve the accuracy of the optimal turbine sitting, i.e. the location of each wind turbine in a given cell can be modified to maximize the energy produced by that turbine. The paper [159] is based on a multi-objective evolutionary algorithm which maximizes the power production capacity while constrains the budget of installed turbines. The algorithm evolves individuals of the Pareto optimal solution, which are evaluated in terms of the different criteria considered in the optimization. This is one of the first works dealing with multi-objective evolutionary techniques applied to wind farm design problems. Another interesting paper is the work in [79] based on hybrid genetic algorithms, in which the genetic algorithm is combined with a steepest ascent hill-climbing local search technique and with an heuristic method to reduce the computation time in finding the local optimum. There are also other bio-inspired approaches (alternative to evolutionary algorithms) that have been successfully applied to the wind farm design problem, for example Particle Swarm Optimization (PSO). In [174] the authors proposed a PSO with inertia weight, maximum velocity and constriction factor. The main novelty of this work is the fact that a continuous space is considered, while previous works considered discrete positions for the wind turbines (the center of the cells). In [140] there are a lot of similarities with the work in [119] and [61] because a comparison between those methods is made. Another work that also tackle wind farm layout problems is the one in [30, 31]. In their proposed PSO

a combination of different rotors diameters is included. A study of the influence of the size of the wind farm and the number of turbines is also carried out. A most recent work that tackles the optimal placement of wind turbines with a PSO is the one proposed in [137]. A binary PSO (BPSO) with time-varying acceleration coefficient (TVAC) is applied to different scenarios, in which both uniform and non-uniform wind characteristics are set. A comparison with different methods, such as genetic algorithms, BPSO-TVIW (time-varying inertia weight factor), BPSO-RANDIW (random inertia weight factor) and BPSO-RTVIWAC (random time-varying inertia weight and acceleration coefficients) is carried out. It is also worth mentioning the novel and recent work proposed in [63], that introduces a new method to take into account irregular-shape wind farms. Even though the optimization algorithm is a traditional Gaussian PSO, the main proposed novelty is the introduction of an edge detection algorithm to extract wind farm contour data from digital maps. Another bio-inspired technique, Ant Colony Optimization, is used in the work described in [51]. Although Simulated Annealing (SA) is a well-known optimization technique applied to numerous optimization problems, only two works in the literature use SA to tackle the problem of onshore wind farm layouts, [10] and [70]. A different approach is proposed in [172], where a covariance matrix adaptation evolution strategy (CMA-ES) is proposed as the optimization algorithm. An excellent review of the most significant papers focused on onshore wind farm design has been recently published [91]. Another interesting review has been published in [157]. It offers a review of the latest advances and the main aspects that need to be taken into account when dealing wih the wind farm design problem.

Although in this Ph.D. Thesis the problem that has been tackled is onshore wind farm design, a quick review on offshore design is made. Multiple studies have been carried out to account for the growth of offshore wind power. In the study [12] a historical review of the evolution of wind power in onshore and offshore facilities is carried out. Also, the current status of the wind power installed in Europe (onshore and offshore) is presented and a study of investment cost is done. A most recent work, [163], explores the current development of offshore wind power around the world and the economic, technical and environmental issues that this technology involves. Different computational optimization methods has been also applied to the design of offshore wind farms in different works [46, 47, 99, 145, 190]. In the works [46, 47, 99], a novel model for the design of off-shore wind farms is presented and several approaches were compared in this work. A greedy algorithm, a genetic algorithm, a pattern search approach and a simulated annealing technique were tested in this article. In [190], the authors presented a different approach to the design of an offshore wind farm, focused on minimizing the connections between wind turbines, considering a layout where turbines are previously settled. The authors tested their approach in a real design of an offshore facility in Liverpool Bay. Three different genetic algorithms were tested with diverse selection and initialization mechanisms, such as rank-based selection or the niching method. Finally, [145] used a simulated annealing algorithm to solve a problem of optimal turbine sitting in wind farms. In a quite recent paper [136], the authors have proposed an specific approach to a problem of offshore wind farm design, based on mathematical programming techniques, specifically a combination of heuristic and gradient-based algorithms, that provides a good solution to the design of a real wind farm in northern Europe. A novel bio-inspired approach based on the simulation of coral reefs is applied in [153] to offshore wind farm design. A complete description of this novel algorithm is provided as well as a comparison with different algorithms, such as Evolutionary Algorithm, Differential Evolution and Harmony Search. The proposed Coral Reef Optimization algorithm obtains better results in the discussed problem.

Finally, to finish this section, a brief note on useful software for wind farm design is given. Currently, the most used free software for wind farm turbines layout is *OpenWind* [127]. This is a wind farm design software for engineers and scientists. This software is based on the following heuristic algorithm:

1. The heuristic searches for a valid (which fulfils all the problem's constraints) layout. Once a starting valid layout is available, the heuristic does a full test of the layout to get the starting energy. It then tests the layout again to get its first optimizing benchmark. It then begins to optimize the layout to evolve to better situations after a number of iterations.

   Each iteration of the optimizer consists of the following steps:

   - The heuristic attempts to find a new legal position for each turbine. If the turbine made a good move in the last iteration, it will attempt to move in the same direction. Otherwise, it finds a new random movement adding a Gaussian-distributed random noise to the turbineś x and y coordinates. If the new position is not a legal position or it obstructs another turbine, then a new random movement is made and so on until all the turbines have new valid positions.
   - The heuristic then runs an energy capture and if the total energy is greater than the benchmark energy, it accepts the entire new layout and the heuristic returns to step 2.
   - If the new layout was not accepted as a whole, the heuristic analyzes each turbine in a separate way: if the turbine has less energy than in its benchmark position, the movement for this turbine is rejected, and it is returned to its last position and benchmark energy. After analyzing all the turbines, the heuristic sums the total energy from all the turbines and if it is equal or greater than the benchmark energy, it runs another energy capture to see if it really constitutes an improvement. If no,t then all the perturbations are discarded and return to step 2. If so, then all these new positions and energies are accepted as the new benchmark energies and the heuristic returns to step 2.

The Openwind heuristic is fast and can be used in combination with evolutionary algorithms to improve the quality of the latter. A hybrid evolutionary algorithm with the Openwind heuristic as initial point could be a very attractive algorithm to face difficult wind farm design problems.

Chapter 3

Techniques and methods

Lots of different real-world problems can be modelled as optimization or regression models. Usually, these models have huge search-spaces, constraints and are difficult to model and to find an exact solution in a reasonable time. Thus, applying conventional algorithms is very arduous, if not impossible. In these cases, Soft Computing (SC) techniques have shown a big potential to overcome these limitations and tackle these problems. As L. Zadeh defined in [187]: "*Soft computing is a collection of methodologies that aim to exploit the tolerance for imprecision and uncertainty to achieve tractability, robustness and low solution cost*". So SC appears to be a good way to tackle real-world problems.

The term SC does not refer to an homogenous group of concepts and techniques. On the contrary, the classification of SC methods is flexible and even a technique can belong to different groups. Mainly, SC is divided in three different branches: Neural Computation, Evolutionary Computation and Fuzzy Techniques (Figure 3.1).

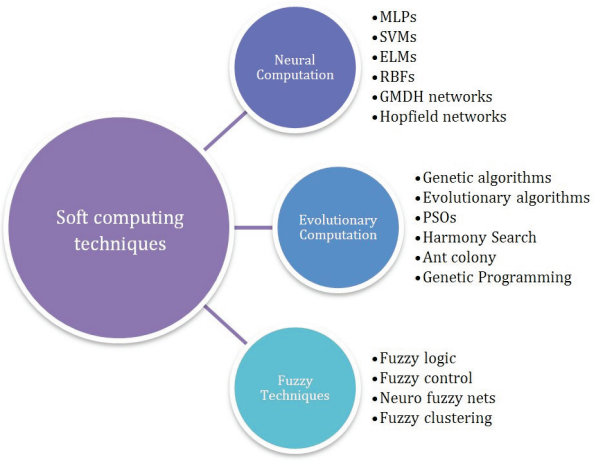


Figure 3.1: Classification of Soft Computing techniques.

For the aim of this Ph.D. Thesis, Neural Computation and Evolutionary approaches are deeply studied.

3.1 Neural Computation-based Approaches

Neural Computation is the first branch of Soft Computing techniques. It includes algorithms inspired in the human brain metaphor and tries to mimic how the biological neurons work. So these techniques will, given a set of data (inputs and outputs), analyze it and recognise patterns, learning the relationship between them. Artificial Neural Networks (such as Multi-layer perceptrons and Extreme Learning Machines) and another approaches devoted to classification and regression problems, such as Support Vector Machines, are included in this branch.

3.1.1 Multi-layer perceptron

A Multi-Layer Perceptron (MLP) is a particular kind of Neural Network (NN), a massively parallel and distributed information processing system, successfully applied in modeling a large variety of nonlinear problems [13, 69]. The MLP is probably the class of neural network most widely used in classification and regression problems. An MLP is a parallel information processing network consisting of an *input layer*, a number of *hidden layers*, and an *output layer*. The leftmost one represents the input layer, and, as shown in Figure 3.2, receives a number of N data inputs, which usually are arranged forming an *input vector*. On the other side, the rightmost layer, named the output layer, produces an output signal. In Figure 3.2, for the sake of clarity, we have represented a simple situation in which the MLP generates only one output signal although, in a more general case, it could produce a number of output signals. Those layers in between the input and output layers are the so-called hidden layers. In turn, the layers forming an MLP are basically composed of a number of especial processing units, called *neurons*, whose internal behavior will be described below. Prior to this, and as shown in Figure 3.2, as important as the processing units themselves is the connectivity among them: note that the neurons within a given layer are connected to those of other layers by means of weighted links. These weights are just the parameters that determine to what extent a neuron is connected to other. In this respect, the value of each weight is related to one of the most important properties that an MPL can exhibit: the ability to learn and generalize from a sufficiently long number of examples.

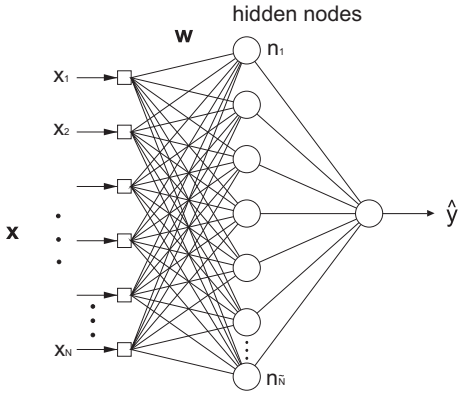


Figure 3.2: Example of the structure of a MLP/ELM network.

Such a learning process demands a proper database containing a variety of input examples or patterns and their corresponding known outputs. The adequate weight values are just those that minimize the error between the output generated by the MLP (when fed with input patterns in

the database) and the corresponding expected known one in the database. Or in other words, the weights of the links are adjusted to learn the function relating the input samples to the corresponding known output in the database. Regarding this, it is well known that MLPs (like most of the neural networks) are universal approximators of a wide range of functions, which gives them a great versatility. For instance, in many regression problems and in time series predictions, MLPs with a single hidden layer, as illustrated in Figure 3.2, are profusely used. The number of neurons in the hidden layer is a parameter to be optimized when using this type of neural networks [13, 69].

As represented in Figure 3.2, the input data consist in N samples, which usually are arranged forming an *input vector*, $\{x_1, \ldots, x_N\}$. As mentioned before, once an MLP has been properly trained, validated and tested, when fed with an input vector different from those contained in the database, it is able to generate a proper output y. The relationship between the output and the input signals of a neuron is

$$y = \varphi\left(\sum_{j=1}^{n} w_j x_j - \theta\right), \tag{3.1}$$

where y is the output signal, x_j, for $j = 1, \ldots, n$ are the input signals, w_j is the weight associated with the j-th input and θ is a threshold [13, 69]. The transfer function φ is usually considered as the logistic function:

$$\varphi(x) = \frac{1}{1 + e^{-x}}. \tag{3.2}$$

Usually, the well-known Levenberg-Marquardt algorithm is used to train the MLP [68]. The Levenberg-Marquardt algorithm was designed to approach second-order training speed, without having to compute the Hessian matrix. This matrix is estimated using the Jacobian matrix instead, which can be computed through a standard back-propagation technique, much less complex than computing the Hessian matrix [68]. The Levenberg-Marquardt algorithm works by using the following Newton-like update:

$$x_{k+1} = x_k - \left(J^T J + \mu I\right)^{-1} J^T e, \tag{3.3}$$

where J is the Jacobian matrix, e is a vector of network errors and μ is a parameter which controls the process: note that when $\mu = 0$, we have the Newton's method, while if μ is large, it becomes a gradient descent method with small step size.

MLPs have been successfully applied to many different classification and regression problems in science and engineering applications. The main drawback related to this algorithm is the lack of a general rule to come up with an optimal network structure to solve a given problem. The optimal number of hidden layers and the number and type of neurons in this layer is an open problem, which has been tackled massively in the literature. The training algorithm is another open question, though the existing approaches provide good results, research in this field is still open, and different training algorithms, many of them based on meta-heuristic search, have been proposed. The Extreme Learning Machine, described in the following section, is one of the most successful training approaches suggested in the last few years.

3.1.2 The Extreme Learning Machine

The Extreme Learning Machine (ELM) is a novel and fast learning method based on the structure of multi-layer perceptrons, recently proposed in [73] and applied thereafter to a large number of classification and regression problems [75, 76, 77]. The ELM approach is a novel way of training feedforward neural networks, with perceptron structure similar to the network given in Figure 3.2. The most significant characteristic of the ELM training is that it is carried out just by randomly setting the network weights, and then obtaining the inverse of the hidden-layer output matrix. The advantages of this technique are its simplicity, which makes the training algorithm extremely fast, and also its outstanding performance when compared to a priori learning methods, usually better than other established approaches such as classical multi-layer perceptrons or support vector machines. Moreover, the universal approximation capability of the ELM network, as well as its classification capability, have been already proven [74, 78].

The ELM algorithm can be described as follows: given a training set $\aleph = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbb{R}^n, \mathbf{t}_i \in \mathbb{R}^m, i = 1, \cdots, N\}$, an activation function $g(x)$ and number of hidden nodes ($\tilde{N}$),

1. Randomly assign input weights $\mathbf{w}_i$ and biases b_i, $i = 1, \cdots, \tilde{N}$.

2. Calculate the hidden layer output matrix $\mathbf{H}$, defined as

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x_1} + b_1) & \cdots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x_1} + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x_N} + b_1) & \cdots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x_N} + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \tag{3.4}$$

3. Calculate the output weight vector β as

$$\beta = \mathbf{H}^{\dagger}\mathbf{T}, \tag{3.5}$$

   where $\mathbf{H}^{\dagger}$ stands for the Moore-Penrose inverse of matrix $\mathbf{H}$ [73], and $\mathbf{T}$ is the training output vector, $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_N]^T$.

Note that the number of hidden nodes ($\tilde{N}$) is a free parameter of the ELM training, and must be estimated for obtaining good results. The solution to this problem is usually to evaluate a different number of values for $\tilde{N}$. It is well known that the ELM is an algorithm with a low computational complexity because it just involves the calculation of the output weights by means of the Moore-Pennrose matrix and other minor calculations.

Because of their excellent performance along with their extreme fast training time, ELMs are perfect for hybrid algorithms that requires fast classifiers or regressors, such as in the case of feature selection. The research in ELMs has also been focussed on improving their performance, by including small modifications to their training algorithm. There are some parameters that must be optimized in ELM, i.e. the number of neurons in the hidden layer, and the random initialization of their input weights. Different approaches have focused on improving ELM performance by tuning these points using optimization algorithms [77]-[78].

3.1.3 The Group Method of Data Handling

The Group Method of Data Handling (GMDH) is a self-organized heuristic technique developed by A. G. Ivakhnenko in 1968 [81], which have been successfully applied to different prediction problems, including applications in wind energy [1] and also solar energy [88]. In self-organized algorithms such as GMDH, the model is generated in an adaptive way by using the data, getting more complex and fitting it to a specific problem, this process continues until the model reaches the optimal complexity degree.

Generally speaking, the relationship between input-output variables can be approximated by Volterra functional series, the discrete analogue of which is Kolmogorov-Gabor polynomial [3].

$$p = a_0 + \sum_{i=1}^{m} a_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} a_{ijk} x_i x_j x_k + \dots \tag{3.6}$$

where, $x = (x_1, x_2, \dots, x_m)$ are the inputs and $A = (a_0, a_1, a_2, \dots, a_m)$ are the coefficients (weights). The Kolmogorov-Gabor polynomial is a universal format for non-linear function modeling as they can approximate any continuous function on a compact data set to an arbitrary precision, in an average squared residual sense (ASR), if there are enough terms [55]:

$$ASR = \frac{1}{N}\sum_{i=1}^{N}(y_i - p(x_i))^2 \tag{3.7}$$

The Kolmogorov-Gabor polynomial has an important drawback, it is necessary to have a large numbers of samples in order to calculate all the coefficients a_i [81]. In order to overcome this drawback, Ivakhnenko proposed a new algorithm which approximates the Kolmogorov-Gabor polynomial by using low order polynomials in an iterative method very similar to the Multilayer Perceptron. This method does not need so many samples and the time consuming is much lower. In fact, Ivakhnenko proved that using second order polynomials can reconstruct the complete Kolmogorov-Gabor Polynomial.

There are several GMDH types [3], in this work the Multilayer GMDH algorithm [81] is considered. This algorithm constructs a hierarchical tree graph of bivariate second order polynomials in the nodes and variables in the leaves. The steps followed to create a GMDH network (Figure 3.3) are:

1. Obtain the data set of the problem $D = (x_i, y_i)_{i=1}^{N}, x_i = (x_{i1}, x_{i2}, \dots, x_{iK})$, where N is the number of samples and K the input dimension.

2. Split the data set in two subsets, one of them is used to calculate the polynomial coefficients of the nodes. The second subset checks the goodness of every polynomial created before.

3. Make all combination of variables in pairs (x_i, x_j) in order to generate all the possible bivariate polynomials $L = \frac{K(K-1)}{2}$.

4. Calculate the coefficients of every polynomial with Ordinary Least Square method.

$$A = (X^T X)^{-1} X^T Y \tag{3.8}$$

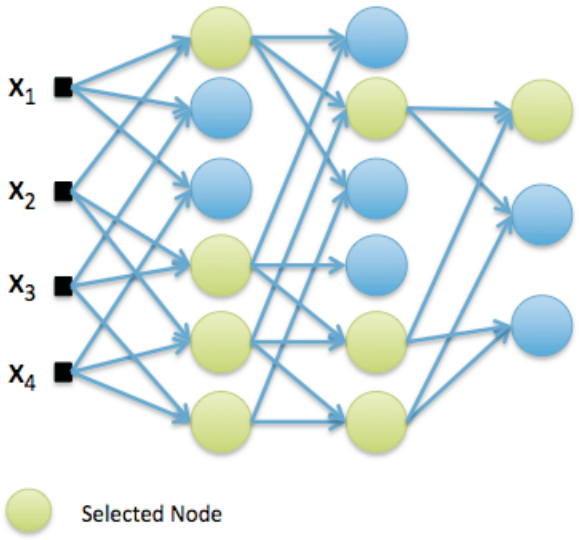


Figure 3.3: Basic structure of a GMDH neural network.

5. Apply an external criterion to choose the best nodes in the current layer. This election must be based on new available information (samples not previously used in the process of coefficients calculation) in order to avoid over-fitting. There are several possible criteria to carry out this node selection, the most popular in GMDH is called *regularity criterion*, which consists of splitting the training set in two subsets A and B. One is used to calculate the coefficients and the other to apply the ASR by using the second order polynomial (Equation (3.7)).

6. The outputs of the selected nodes are the inputs for the new layer.

This process is repeated until a stopping criterium is fulfilled. The stopping criterium is usually based on the ASR: when, at some point of the network construction, the ASR does not decrease from one layer to the next one, the algorithm is stopped, because the structure generated is considered as enough complex in order to make good estimations.

3.1.4 Support vector regression algorithms

One of the most important forecasting statistic models are the support vector machines for regression (SVMr) [160]. The SVMr are used in a large amount of regression problems, including wind speed prediction [113]. The SVMr uses kernel theory to increase the quality of regression models and, in most cases can be solved as a convex optimization problem. The SVMr does not only take into consideration the prediction error of the data, but also the generalization of the model, i. e., its capability to improve the prediction of the model when new data are evaluated by it (avoiding over-fitting). Also there are several versions of SVMr, in this case the classical model presented in [160] is described.

The ϵ-SVM method for the SVMr [168] consists of, given a set of training vectors $S = \{(\mathbf{x_i}, y_i), i = 1, \ldots, l\}$, obtaining a model of the form $y(\mathbf{x}) = f(\mathbf{x}) + b = \mathbf{w}^T \phi(\mathbf{x}) + b$, to minimize

a general risk function of the form

$$R[f] = \frac{1}{2}\left\|\mathbf{w}\right\|^2 + \frac{1}{2}C\sum_{i=1}^{l} L\left(y_i, f(\mathbf{x})\right) \tag{3.9}$$

where $\mathbf{w}$ controls the smoothness of the model, $\phi(\mathbf{x})$ is a function of projection of the input space to the feature space, b is a parameter of bias, $\mathbf{x_i}$ is a feature vector of the input space with dimension N, y_i is the output value to be estimated and $L\left(y_i, f(\mathbf{x})\right)$ is the selected loss function. In this work, the L1-SVR (L1 support vector regression) is used, characterized by an ϵ-insensitive loss function [168]:

$$L\left(y_i, f(\mathbf{x})\right) = |y_i - f(\mathbf{x_i})|_\epsilon \tag{3.10}$$

In order to train this model, it is necessary to solve the following optimization problem [168]:

$$\min\left(\frac{1}{2}\left\|\mathbf{w}\right\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right) \tag{3.11}$$

subject to

$$y_i - \mathbf{w}^T\phi(\mathbf{x_i}) - b \leq \epsilon + \xi_i, \quad i = 1, \ldots, l \tag{3.12}$$

$$-y_i + \mathbf{w}^T\phi(\mathbf{x_i}) + b \leq \epsilon + \xi_i^*, \quad i = 1, \ldots, l \tag{3.13}$$

$$\xi_i, \xi_i^* \geq 0, \quad i = 1, \ldots, l \tag{3.14}$$

The dual form of this optimization problem is usually obtained through the minimization of the Lagrange function, constructed from the objective function and the problem constraints. In this case, the dual form of the optimization problem is the following:

$$\max\left(-\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(\mathbf{x_i}, \mathbf{x_j}) - \epsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*)\right) \tag{3.15}$$

subject to

$$\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \tag{3.16}$$

$$\alpha_i, \alpha_i^* \in [0, C] \tag{3.17}$$

In addition to these constraints, the Karush-Kuhn-Tucker conditions must be fulfilled, and also the bias variable, b, must be obtained. This process is not detailed here for simplicity, the interested reader can consult [168] for reference. In the dual formulation of the problem the function $K(\mathbf{x_i}, \mathbf{x_j})$ is the kernel matrix, which is formed by the evaluation of a kernel function, equivalent to the dot product $\langle\phi(\mathbf{x_i}), \phi(\mathbf{x_j})\rangle$. An usual election for this kernel function is a Gaussian function, as follows:

$$K(\mathbf{x_i}, \mathbf{x_j}) = exp(-\gamma \cdot \left\|\mathbf{x_i} - \mathbf{x_j}\right\|^2). \tag{3.18}$$

The final form of function $f(\mathbf{x})$ depends on the Lagrange multipliers α_i, α_i^*, as follows:

$$f(\mathbf{x}) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(\mathbf{x_i}, \mathbf{x}) \tag{3.19}$$

Note also, that, in addition to the obtention of the final model given by Equation (3.19), the hyper-parameters C, γ and ϵ must be estimated before a complete application of SVMr, usually using grid search with k-cross validation, so the computational cost of SVMr is eventually quite high.

3.2 Evolutionary computation-based algorithms

Evolutionary Computation-based approaches are inspired by the principles of Genetics and Natural Selection. The most representative strategy in this subset is perhaps the concept of Genetic Algorithm (GA) [44], although there are also other paradigms that have been recently introduced such as the social computation by swarms, like in the Particle Swarm Optimization approach (PSO) [41], or the simulation of an orchestra composition, like in the Harmony Search algorithm [57]. There are many algorithms belonging to the family of evolutionary computation. Here genetic and evolutionary approaches are described

3.2.1 Genetic and evolutionary algorithms

Evolutionary computation algorithms have been widely used for solving combinatorial optimization problems, which primarily work in discrete search spaces. These algorithms are based on an encoding of the problem by strings of numbers. All genetic and evolutionary algorithms are based on the evolution of a population of candidate solutions by applying a series of evolutionary operators.

In general, a genetic algorithm is based on the evolution of a population of individuals (called chromosomes, indistinctly), each of them representing a candidate solutions of the problem. Now, different operations inspired by the natural evolving process are iteratively applied: mutations, recombinations among different individuals, selection of the most suited to the environment ones (the so-called *survival of the fittest*). Figure 3.4 shows this process, in a simply way.

- **Encoding candidate solutions**

  Encoding candidate solutions is the very first step when applying a genetic algorithm. We have to choose a feasible codification that allows the encoding of the different solutions of the problem.

  In nature, all of the genetic information which encodes and causes the external characteristics of a living organism (or individual) is called genotype. Any particular characteristic produced by a piece of this genetic information is encoded by a gene. Usually, a chromosome will be a vector of numbers, being each number a gene. Each gene is located at a particular position on the chromosome and can have different values, called allele. Note

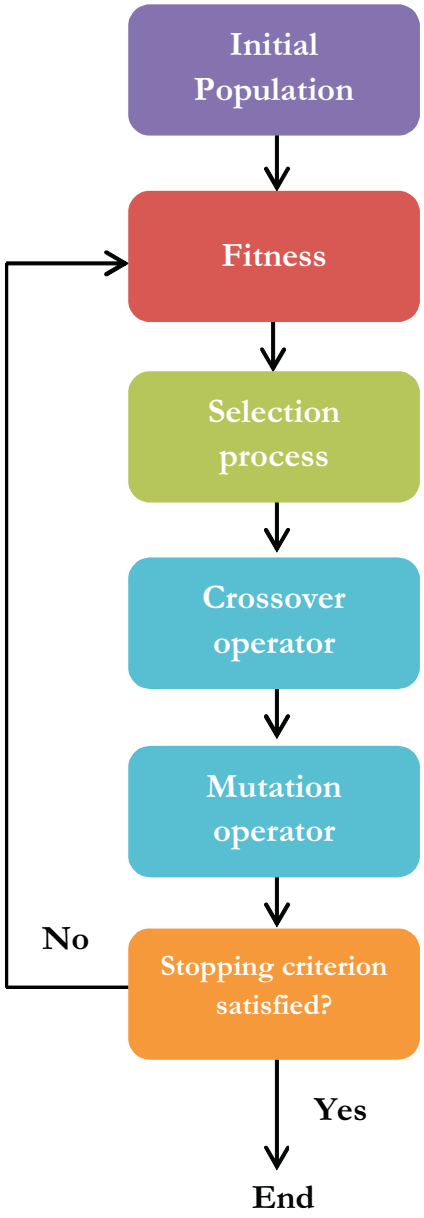


Figure 3.4: Example of the process of a genetic algorithm.

that this strategy can be considered as transforming the real search space into another in which working is much easier.

Normally, the values of the different genes of a chromosome are independent. The value a gene adopts does not influence the set of values that another gene can take. However, sometimes, because of the chosen encoding system, the values of the different genes depend on each other, being that the case of permutations. In that case, a certain value in a gene cannot appear in another gene of the chromosome.

The existence of this kind of constraints is a determining factor for the design of some operators involved in the evolving process of a genetic algorithm, such as the recombination or mutation operators. This is due to the fact that the dependence on the values of different genes requires operators designed to fulfill the required constraints. The dependence between the genes depends on the encoding system. Therefore, the operators must be designed for each specific type of encoding. The following operators are applicable to those problems without encoding constraints, and should be modified in order to fulfill the specific requirements of each problem.

- **Fitness function**

  The fitness function is the one used to measure how much an individual is suited to its environment. An infinity of possible fitness functions can lead to the solution of a problem. The reach of good solutions depends on the chosen function and the convergence of it.

  We have to be very careful when defining a fitness function because of the constraints of each problem. This problem can be tackled with two different approaches:

  – In this first approach, the solutions that do not fulfill the requirements are unfeasible. Therefore, the solution is repaired until a feasible solution is obtained or the fitness function value is set to zero.
  – This second approach is based on the penalization of the fitness function. The value of the fitness function is divided by a certain value (the penalty), related with the unfulfillment of the constraints.

- **Selection**
  The selection operator generates a new population randomly choosing chromosomes of the original population, in such a way that the better an individual is suited to an environment, the higher its probability of being selected (survive) is. Lots of methods have been proposed to carry out the selection process, but two of them can be highlighted:

  – Roulette-wheel: in this case, each individual has a probability of selection related to its fitness value. A portion of the wheel is assigned to each chromosome, being proportional to the value of the fitness. This is usually carried out dividing the fitness value of each individual by the total fitness of all the individuals. Then, a random selection is made to choose the individuals of the new population. With this method, weak individuals are less likely to survive, but there is a chance they may be chosen. That includes genetic material, which causes genetic diversity and by means of the recombination process can produce new and better individuals.

  – Tournament: a subset of chromosomes among the initial population is chosen at random to "fight in a tournament". The fittest individual of this subset will survive the next generation. This process is carried out until the whole new population is completed. The larger the tournament size is, the smaller probability weak individuals have to be selected. Figure 3.5 shows the process.

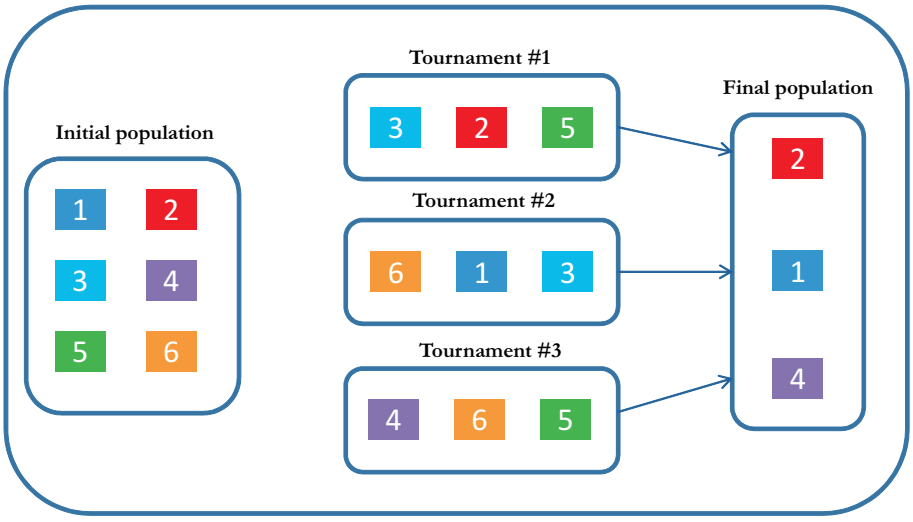


Figure 3.5: Example of the tournament selection process. Numbers on each individual represent the fitness value, being 1 the fittest individual.

Even though the evolution in a genetic algorithm tries to find the fittest individual, it is sometimes not found in an exact way. Due to the intrinsic randomness of the selection operator, it is possible that, although the optimum individual appears in a certain generation of the evolution, it does not survive, and later it may be impossible to reach that solution again.

To avoid the loss of optimum individuals, the elitism operator is used. The fittest individual (the elitist) in each generation is placed on a privileged position that assures its survival. The elitist cannot be replaced or modified by any operator. In that way, the best individual on each generation is able to survive.

- **Crossover and recombination**
  In the procreation process, the parent chromosomes are combined to provide a novel chromosome. Each individual is chosen with a probability of recombination to be a part of the crossover process. Once the "parents" chromosomes are chosen, they are grouped in random pairs, so as to each pair generates new descendants. The objective of this crossover process is to join the potentially advantageous features of different chromosomes in a new fitter individual. There are a lot of recombination schemes, but commonly, the ones most used are:

  – One-point crossover: A random point is selected. Data of each parent chromosome beyond that point is swapped. The resulting individuals are the children.

  – Two-point crossover: In this case, the genetic material swapped, is the one between two genes, randomly selected.

  – N-point crossover: It is a generalization of the two previous cases. Each parent individual is split along N random points. The $N+1$ segments in which each chromosome

is sectioned into, are swapped between parents.

- Uniform crossover: Each gene of the chromosome is swapped between parent with a certain probability.

- **Mutation**
  The mutation operator introduces genetic variability, slightly changing some genes of a certain individual. This operator allows the algorithm to evolve, without getting stuck in local minimums, inducing a genetic diversity and thus, a higher search space. However, the mutation probability must be set low, because if not, the search will turn into a random search. There are different kinds of mutation operators:

  - Gene changing: The value of the genes that are going to be mutated, are replaced with another feasible value, according to the encoding system. In the particular case of binary encoding, the mutation implies the inversion of the gene ('1' turns '0' and viceversa).

  - Swap: The mutation is carried out over two different genes of an individual. These genes, randomly chosen, are swapped.

  - Probabilistic distribution: In real-value encoding, the mutation is carried out adding random noise, defined by a probabilistic distribution, such as Gaussian or Cauchy.

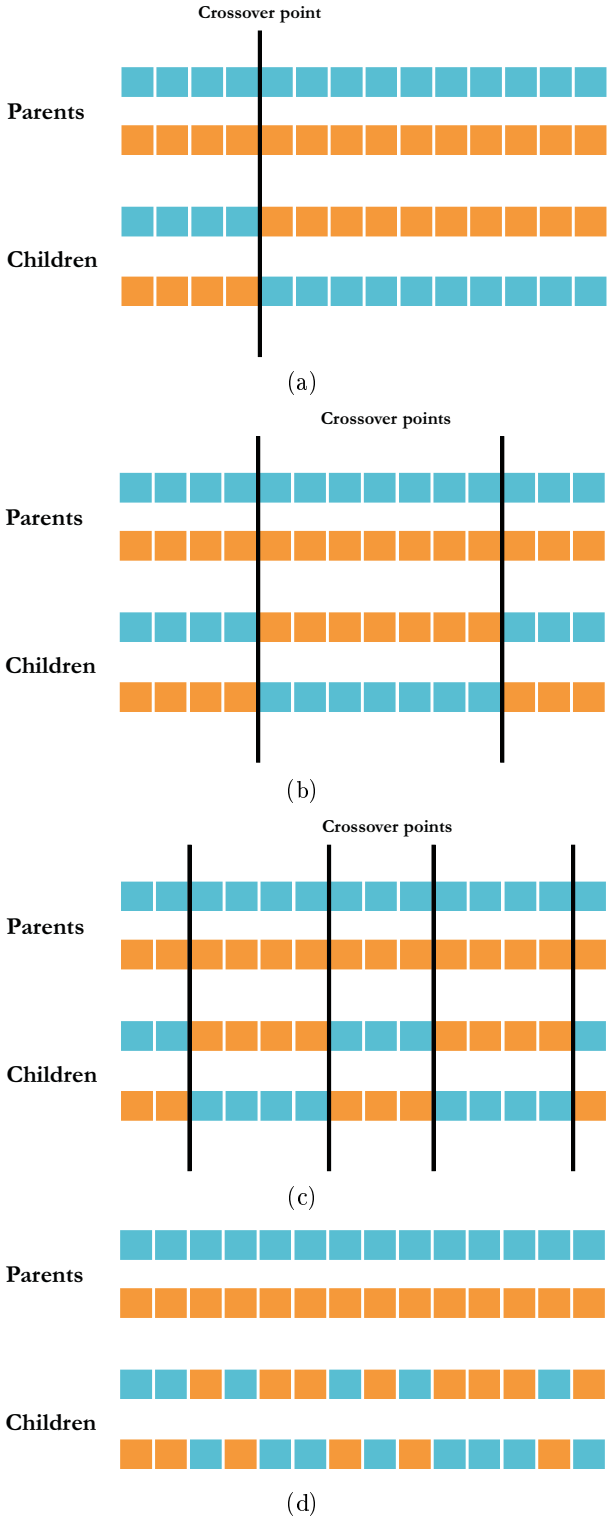


Figure 3.6: Examples of crossover process; (a) One-point crossover; (b) Two-point crossover; (c) N-point crossover; (d) Uniform crossover.

Part II

Proposed contributions with numerical results

Chapter 4

Wind speed regression based on Synoptic-scale pressure fields

4.1 Introduction

Wind speed reconstruction is an important problem in wind energy, related to key procedures in wind farms, such as production analysis [56, 97], planning [126], wind resource [139] or even micro-siting of new wind turbines and wind farms. Existing approaches to tackle wind speed reconstruction are mainly based on historical data and wind speed measurements, usually taken in-situ, using a measuring tower within the wind farm. From these wind speed measurements, statistical models are then constructed in order to carry out the desired wind speed reconstruction [34]. The main disadvantage with this approach is that, in some cases, wind measurements are not fully available at a given location, due to failures in the observation systems, short measuring campaigns or even meteorological mast dismantling in wind farm areas. In these problematic cases, we could use indirect measurements (not direct measured wind speed) in order to construct the statistical models (this way we need less wind speed measurements to construct these models). The synoptic mean sea-level pressure (SLP), or patterns thereof, is the most commonly used indirect variable for wind speed reconstruction [19, 22, 150], and it has also been used as prospective variable in different applications such as pollution [28] or rainfall analysis [147].

Different methods have used SLP as a predictive variable for meteorological prediction. We can find in the literature works dealing with physical approaches [84], machine learning techniques [23, 64] or hybrid approaches involving both physical and machine learning approaches [94]. Physical approaches refer to those methodologies that solve atmospheric equations to construct the prediction/reconstruction system. On the other hand, machine learning approaches do not consider the physical properties of the problem to construct the prediction/reconstruction system, but instead it is just trained with a set of predictive and goal variables.

We propose a comparison of the performance of a machine learning approach versus a weather oriented model, in a problem of wind speed reconstruction from Synoptic Pressure variables. Specifically, the following problem of wind speed reconstruction in a given point (wind farm) from a related synoptic sea-level pressure values in a grid is tackled: Let $\mathbf{d}_t,\ t = 1\ldots,T$, be a series of wind speed, measured in a given point (a wind farm in this case), for a given period of time T. Let $P_t,\ t = 1\ldots,T$, be a series of synoptic-scale pressure measurements in a given grid. In our case, each component of P_t is a matrix of $M \times N$ surface pressure values, measured in a grid surrounding the wind farm location. Both, the time interval of reconstruction and the range

(dimensions and resolution) of the pressure grid are parameters to be fixed in the problem. The problem consists of obtaining the relationship between the wind speed and the synoptic-scale pressure situation using a given regressor, in such a way that a measure of error between the wind speed reconstruction and a real value is minimized.

The algorithms here compared are the following: in the case of the machine learning approach (a hybrid Support Vector Regression - Genetic Algorithm, SVMr-GA), the problem involves six-hour time pressure patterns in a synoptic grid and a wind speed module measurement. In this case, the SVMr is used to carry out the wind speed reconstruction from the mean sea-level pressure values, that act as input (predictive) variables. A GA is used to carry out a feature selection (selecting what points of the pressure grid are useful to obtain a good wind speed prediction), in order to improve the SVMr results. The physical-based approach considered in this work is a Weather Regimes Classification Technique (WRCT), based on fields of the 850 hPa geopotential height in order to classify the synoptic pattern in a reduced number of Weather Regimes (WR), using a k-means algorithm. A previous step of Principal Component Analysis reduces the dimensions of the problem. Once these WRs are obtained, a wind speed predictor (module of the SLP gradient in this case), is used to obtain the wind speed reconstruction within each WR.

We test both SVMr-GA and WRCT systems in several real problems of wind speed reconstruction at three different sites: Cabauw (The Netherlands), Capel (Wales, UK) and Kaegnes (Denmark). The performance of both approaches is comparatively analyzed, in terms of performance in wind speed, Weibull parameters of a wind speed model and wind power reconstruction, and also in terms of computational cost, obtaining results that identify the advantages and disadvantages of each approach in this problem.

The structure of the rest of the Chapter is the following: Section 4.2 presents the detailed description of the algorithms discussed here. The experimental part is then presented in Section 4.3, where we compare the results obtained by both approaches in a real wind speed reconstruction problem. Finally, Section 4.4 closes the Chapter giving some concluding remarks.

4.2 Material and Methods

4.2.1 A hybrid SVMr-GA algorithm over SLP variables for wind speed reconstruction

4.2.1.1 Feature Selection in the SLP grid

One of the main issues when tackling a regression problem is to obtain a set of input (predictive) variables that are significant. This problem is usually known as Feature Selection Problem (FSP), and it is also important in other problems such as supervised and unsupervised classification. Obtaining a good set of input features is key in regression, since irrelevant features can cause different distortions in the regressor performance, such as lack of generalization or increase of the computation time [16].

A general formulation of the FSP is the following: Let $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)$ be a set of data points and their corresponding objective values, where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$. The FSP consists of choosing a subset of m features ($m < l$), that produces the lowest error in the prediction of y_i.

There are two basic approaches to the FSP: the first one tries to identify an appropriate set of features, independently of its regression performance, i.e. independently of the regressor used, which preserve as much information from the original data as possible. This approach is known as the *filter method* for feature selection [16]. The second approach uses the regressor to obtain a subset of m features out of the total l, trying to improve the performance of the regressor (or at least not degrade it). This method is known as the *wrapper* method, and it has been proven to be more powerful than filter methods, though it is usually computationally more demanding [95, 177].

In this work we consider a wrapper procedure for FSP, that uses a SVMr as a regressor and a genetic algorithm (GA) (described in Chapter 3) as a search algorithm to obtain the optimal set of features (SLP points in the grid).

4.2.2 WRCT: Weather Regimes and SLP gradient for wind speed reconstruction

4.2.2.1 Principal component analysis

Empirical Orthogonal Function (EOF) analysis is a methodology used to calculate the principal directions of variability followed by a particular variable [134]. This methodology, also named as Principal Component Analysis (PCA), is applied in this case to the anomalous fields of covariance matrix of a synoptic scale meteorological field (850 hPa geopotential height in this study), by calculating its eigenstructure. The diagonalization of this matrix produces a diagonal matrix with the fractions of variance given by each of the components, and a set of eigenventors or EOFs which show the regions in which the covariance is organized. A number of eigenvectors are chosen, which make up a new orthogonal base in which the data is oriented explaining a percentage of variance given by the sum of the fraction of variances of the selected components. The projection of each of the EOFs onto the original data gives a time series (Principal Components (PCs)), in which each of the scores indicates the weight of each EOF in each of the days of analysis.

The time domain in the available data is then divided into two sub-domains: a period used to train the model, and a validation period to test the results of the wind speed reconstruction. The PCA is performed with all data available from a Reanalysis model (the one by the European Center for Medium range Weather Forecasting, ECMWF), except for the validation periods, considering groups of three months each: January, February, March (JFM); April, May, June (AMJ); July, August, September (JAS) and October, November, December (OND). A spatial domain has been chosen for the calculation of the PCA, centered at the meteorological mast position and with an area of $10.5° \times 10.5°$ in longitude and latitude. For each group of months, the fraction of the total variance explained by the chosen leading EOFs was established to be greater than 98%. Thanks to the PCA and to the possibility of expressing the geopotential height anomalies at 850hPa fields with a few PCs in the EOFs basis, the computational time used to perform the following cluster analysis is considerably reduced.

4.2.2.2 k-means cluster analysis

The clustering method used is based on the well known k-means clustering [39, 67]. Hence, a k-means cluster analysis using Euclidean distance was applied for each group of months over the

selected PCs, whose result leads to a distribution in k clusters of the geopotential height anomaly fields. The final partition can be sensitive to the algorithm initialization, which requires a first guess for the cluster centroids. Thus, the algorithm was repeated 50 times [143], each one with different initial seeds, and the best partition was retained. The clusters represent commonly occurring patterns of the geopotential height anomalies at 850-hPa field and are so-called weather types or WR [112, 115]. These WRs are responsible for the behaviour of the synoptic-scale motions during several days or weeks [111], and therefore, affect the local weather. This study is based in the assumption that the orography of the geopotential field associated with each WR is the main driving force of surface wind at local scale.

4.2.2.3 Regression of local surface wind against modulus of SLP gradient

Once the training period is selected, the 6 hourly averaged surface wind data for each WR are grouped conforming 4 time-based populations per WR, i.e., data at 0, 6, 12 and 18 UTC. Each of these wind data groups are then linearly regressed with their corresponding modulus of the SLP gradient, which is calculated from 4 points at a distance of 0.75° from the meteorological mast position, two of which are located in the same parallel and the other two in the same meridian (Figure 4.1). The interpolated field at point i (W, E, N, or S point in Figure 4.1) is equal to a weighted mean of the value among the $j = 1, \ldots, 4$ closest points of the grid:

$$SLP_{i=W,E,N,S} = \frac{\sum_{j=1}^{4} \frac{SLP_j}{d_{ij}^2}}{\sum_{j=1}^{4} \frac{1}{d_{ij}^2}} \tag{4.1}$$

where d_{ij} is the linear distance between points i and j.

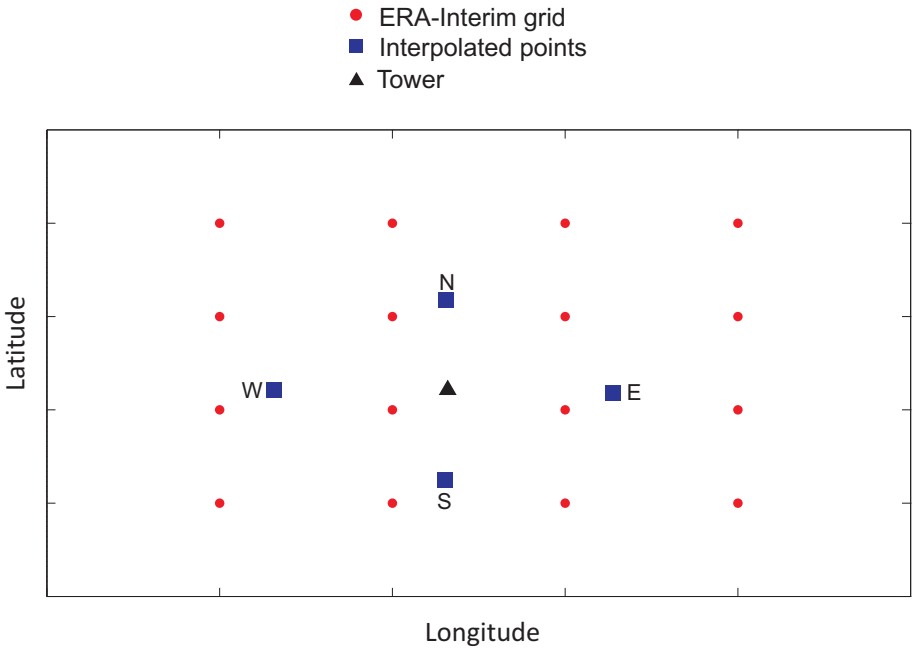


Figure 4.1: Schematic representation of the grid with reanalysis data (ERA-Interim from ECMWF), tower's location (triangle), and the interpolated points used to calculate the modulus of SLP gradient series (W, E, N and S squares).

Once the SLP values in the W, E, N and S points are known, each gradient component was

calculated at the meteorological mast location as a centered finite difference approximation given by Equation (4.2). Then, the components of the vector are composed in order to obtain the wind speed.

$$\left(\frac{\partial SLP}{\partial x}, \frac{\partial SLP}{\partial y}\right) = \left(\frac{SLP_E - SLP_W}{R_T 2\Delta\lambda cos\phi}, \frac{SLP_N - SLP_S}{R_T 2\Delta\lambda}\right) \tag{4.2}$$

where R_T is the Earth radius, $\Delta\lambda$ is the angular distance in radians between each point and the mast and ϕ is the mast's latitude.

4.2.2.4 Wind speed reconstruction

Once the model is trained, the surface wind speed series is reconstructed for every instant of the test period (6-hour time step). The first step is to determine to which WR belongs the 850-hPa geopotential height anomalies map at every moment. Depending on the group of months considered, a projection of the map onto the appropriate EOFs basis was made, calculating its representation in PCs. Then, the Euclidean distance, l, between these PCs and the PCs of all centroids of the seasonal k-means cluster analysis is computed:

$$l_j = \sqrt{\sum_{i=1}^{n_m} [PC_i - PC_{CTR_{j,i}}]^2} \tag{4.3}$$

where the subscript j marks one particular centroid CTR_j and n_m is the number of components retained in the PCA. After this, the module of the SLP gradient is calculated (Equations (4.1) and (4.2)), and then, the corresponding WR regression equation is applied in order to obtained a reconstructed surface wind speed.

4.2.2.5 Choosing the optimal number of clusters

Probably one of the most important issues in cluster analysis is to choose an appropriate number of clusters in the classification. To deal with this challenge, there are a large amount of methods in the literature, but it does not seem to exist an absolutely objective criterion to determine the optimal number of clusters. Thus, a different number can be used depending on the specific application of cluster analysis. The final purpose of this work is the wind speed reconstruction at local scale, so we have developed a new criterion in order to guarantee a better estimation of this variable.

The procedure basically consists of reconstructing the wind speed during a certain period, in our case, each training period for each scenario, and assessing the quality of estimations depending on the number of clusters employed. Therefore, the first step is to choose a time period that comprises a number n of complete years (at least two) with available wind data. With the data of $n-1$ years the WRCT model is trained, using different numbers of clusters (k) in the k-means algorithm. Then, for each choice of k, a wind speed reconstruction is done during the remaining year, calculating the associated RMSE and MAE. The procedure is repeated for all possible combinations of n-1 years to train and 1 year to test, conforming a cross validation of n iterations (n-fold cross-validation). A set of values of RMSE(i,k) and MAE(i,k) are obtained

(see Section 4.3.1), one for each iteration i and for each number of clusters k. Thus, we compute the following index I for each k:

$$I(k) = \frac{1}{2}\left[\frac{\overline{RMSE}(k) - \overline{RMSE}_{min}}{\overline{RMSE}_{max} - \overline{RMSE}_{min}} + \frac{\overline{MAE}(k) - \overline{MAE}_{min}}{\overline{MAE}_{max} - \overline{MAE}_{min}}\right] \tag{4.4}$$

where the overline indicates an arithmetic mean for each number of clusters, made over the n iterations:

$$\overline{RMSE} = \frac{\sum_{i=1}^{n} RMSE(i,k)}{n} \tag{4.5}$$

$$\overline{MAE} = \frac{\sum_{i=1}^{n} MAE(i,k)}{n} \tag{4.6}$$

and the max and min subscripts refer to the maximum and minimum values of the RMSE and MAE means over k. Finally, we identify the optimal number of clusters as the value of k which minimizes the I index.

4.3 Experiments and results

This section presents the experimental part of this study. Observed wind data at three sites in the Netherlands (Cabauw), Denmark (Kaegness) and UK (Capel) are used, together with meteorological variables (geopotential height at 850 hPa and mean Sea Level Pressure) at grids centers in these places (see Figure 4.2). First, we present the data used for this study and then, we show the comparative performance of the two algorithms proposed here.

4.3.1 Data used for the study and methodology

In order to analyze the atmospheric variability during recent-past conditions, high resolution data of the third ECMWF (ERA-Interim) reanalysis generation has been used [158, 169]. The 6-hourly fields (at 0, 6, 12 and 18 UTC) of the geopotential height at 850hPa (WRCT method), and Mean Sea Level Pressure (SLP) for both SVMr-GA and WRCT methods, have been used as predictor fields. All these datasets were used on $0.75° \times 0.75°$ grids centered at each site considered. In all the sites considered, 6-hour averages of the original hourly wind time series were made, centered at 0, 6, 12 and 18 UTC. Note, however, that only those averages calculated with at least 50% of the data in each 6-hours interval were taken into account for the subsequent analysis (there are missing data in the wind series considered at all sites). Specific characteristics of each site are as follows:

Cabauw: The observed wind data at Cabauw (The Netherlands, $51°58'14''N$, $4°55'35''E$) come from a 213m meteorological mast. Wind speed and wind direction were measured with propellor anemometers at several heights. However, only derived hourly measurements of wind speed and direction at 10m were used. The dataset at this site has been downloaded from [180]. Specifically, wind speed data from Cabauw is available from 1st December 1988 until 30th November 2000. We have divided the available data at this site into four different training and test sets within this period (C1 to C4), in order to have some diversity for results discussion:

C1. Training: 01/12/1988-30/11/1991; Test: 01/12/1991-30/11/2000.

C2. Training: 01/12/1991-30/11/1994; Test: 01/12/1988-30/11/1991+01/12/1995-30/11/2000.

C3. Training: 01/12/1994-30/11/1997; Test: 01/12/1988-30/11/1994+01/12/1997-30/11/2000.

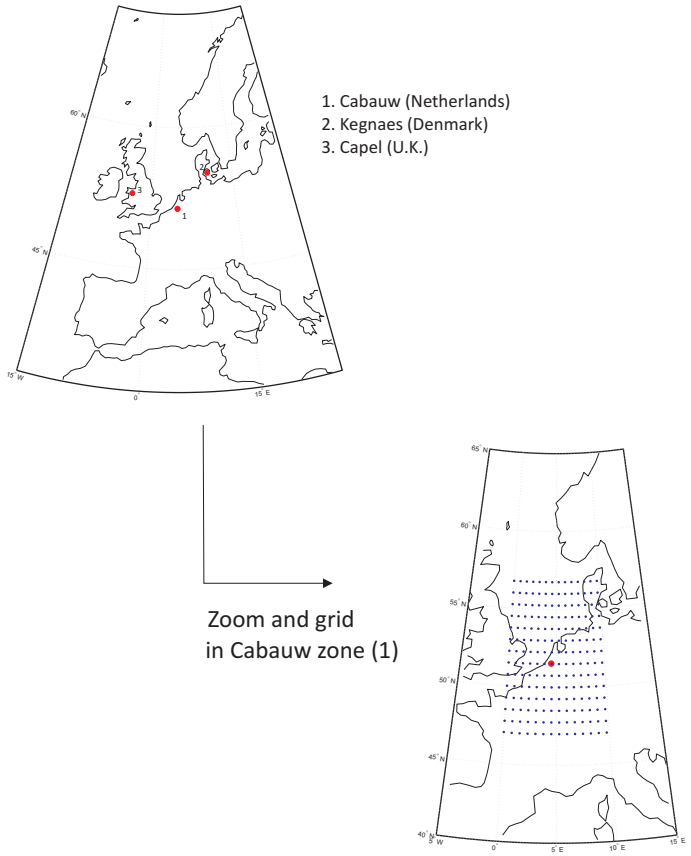


Figure 4.2: Locations of the sites considered (Cabauw, Kaegnes and Capel), and grid (0.75°) considered over Cabauw site.

C4. Training: 01/12/1997-30/11/2000; Test: 01/12/1988-30/11/1997.

Capel: This site is located in Wales (UK), ($52°8'20''N$, $4°21'15''W$). In this case the period of available data is quite reduced, from 01/01/1990 to 31/12/1992. Thus, we only consider one train and test period (Cp1):

Cp1. Training: 01/01/1990-31/12/1991; Test: 01/01/1992-31/12/1992.

Kaegnes: This site is close to the shore of Kaegnes island, Denmark ($54°51'20''N$, $9°56'10''E$). In this case we have available data from 01/03/1991 to 29/02/2000. We have considered three training and test sets at this site:

K1. Training: 01/03/1991-28/02/1994; Test: 01/03/1994-29/02/2000.

K2. Training: 01/03/1994-28/02/1997; Test: 01/03/1991-28/02/1994+01/03/1997-29/02/2000.

K3. Training: 01/03/1997-29/02/2000; Test: 01/03/1991-28/02/1994.

Note that the results obtained at all sites will be referred to these training and test sets. For these sites, the average number of weather types (clusters) obtained by applying WRCT methodology are displayed in Table 4.1. Note that depending on the training period considered, the number of clusters varies, so average values are displayed in this table.

On the other hand, we will use several statistical measures to compare the results obtained by the WRCT and SVMr-GA approaches proposed. Specifically, we use the root mean squared error, the mean absolute error, the bias and R^2 measure, defined as follows:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(v_{p_i} - v_{r_i})^2}{N}} \tag{4.7}$$

$$MAE = \frac{\sum_{i=1}^{N} |v_{p_i} - v_{r_i}|}{N} \tag{4.8}$$

$$Bias = \frac{\sum_{i=1}^{N} v_{p_i} - v_{r_i}}{N} \tag{4.9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (v_{p_i} - v_{r_i})^2}{\sum_{i=1}^{N} (v_{r_i} - \bar{v_r})^2} \tag{4.10}$$

where v_{p_i} stands for the predicted wind speed value, v_{r_i} stands for the observed wind speed value, $\bar{v_r}$ is the mean observed wind speed value and N is the number of samples of wind speed considered in the test set.

Regarding SVMr tuning, usually the SVMr parameters (C, ϵ and γ) selection is done by standard grid search in the space of parameters [128], though there is also the possibility of using a global search algorithm in order to obtain these parameters [152]. In this case, we finally decided to apply a grid search approach for SVMr parameters tuning, to keep the GA as simple as possible (binary encoding with the standard operators), and only focused on the feature selection of synoptic pressure points. The methodology applied to tune the SVMr parameters is to carry out a first grid search to tune the SVMr parameters before running the GA for feature selection. Finally, once the best set of features have been selected we also carry out a final more accurate grid search to obtain the SVMr parameters before the final application of the SVMr in the test set.

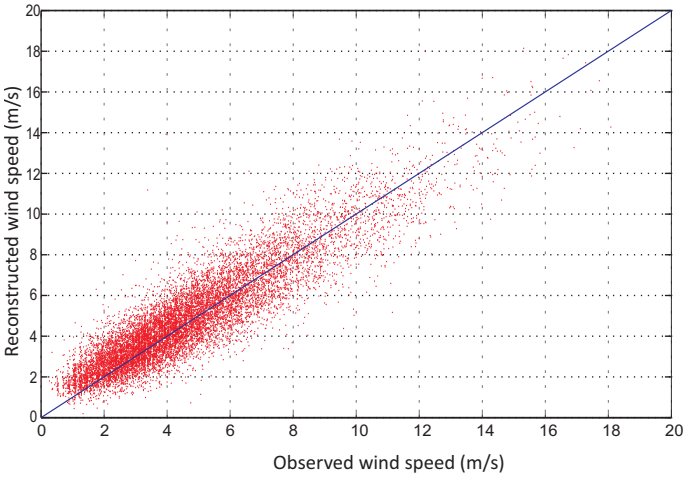
GA parameters are also important to make the results of this work reproducible. We have used a population of 30 individuals, during 50 generations. Multi-point crossover and flip-based mutation are used to form an offspring population. The parents of next generation are obtained from the joined population of parents and offspring by using a tournament selection. We have followed the algorithm's dynamics outlined in [186] in the proposed GA.

4.3.2 Results and discussion

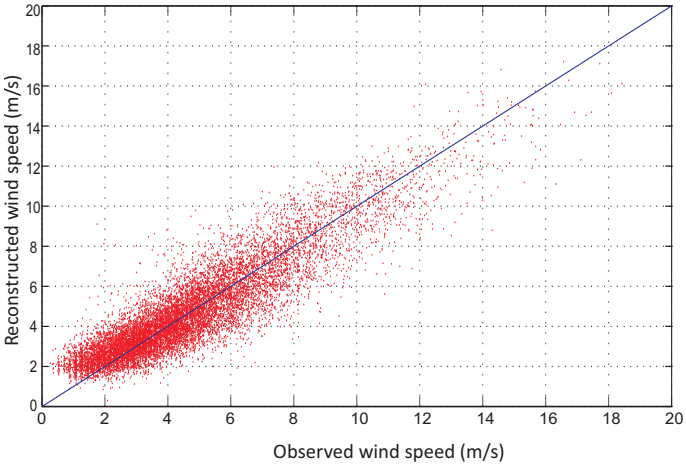
The structure of this subsection is based on the results obtained in each site, Cabauw, Capel and Kaegnes.

4.3.2.1 Cabauw

The results shown at this site are based on the 4 different training and test periods (C1 to C4) defined above for Cabauw. Table 4.2 shows the obtained results with the WRCT and SVMr-GA methods compared, respectively at this site. Note that the WRCT approach obtains slightly better values of the statistical indices considered in all the periods. However, there are no major differences between the compared approaches in any of the periods considered. Figure 4.3 (a) and (b) show an example of the dispersion figure (real wind speed vs. reconstructed wind speed)



(a)



(b)

Figure 4.3: Wind speed reconstruction in C1 test period at Cabauw; (a) WRCT approach; (b) SVMr-GA approach.

with the two compared algorithms, in the test period C1. It seems that the WRCT performs better in for low wind speeds, whereas the dispersion is smaller with the SVMr-GA for high wind speeds. Note that the wind speed reconstruction is made in 6-hours time averages.

It can be seen that both methodologies tend to miss maximum and minimum values of the wind speed in the reconstruction. However, note that the aim of both approaches is to obtain a good reconstruction of the wind speed in general, and none of them are specifically focused on maximum and minimum values of the wind. In this sense, the proposed methodologies are not applicable to locate extreme values in the reconstruction, but to obtain a good average reconstruction of the wind speed, which is enough in many cases. Figures 4.4 and 4.5 show the wind speed histogram reconstruction, and the observed one at Cabauw, for both approaches and for all the test periods considered. As it can be seen, the histogram reconstruction is quite accurate. We also can model the reconstructed histograms of wind speed to a Weibull distribution for each test period, and obtain the parameters of the distribution. Table 4.3 (top of the table)

shows these parameters for each test period at Cabauw and the two considered approaches. Note that this allows a complete and accurate reconstruction of the wind speed at Cabauw, and also it brings out the possibility to extend the wind speed analysis back to the past, and in years where there are not wind speed measures at this site.

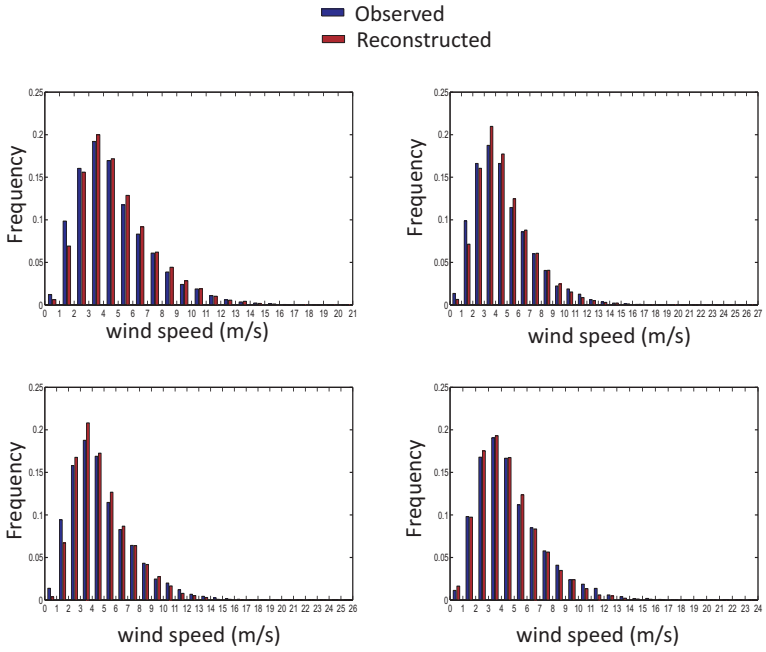


Figure 4.4: Histograms of the wind speed (reconstructed and observed ones), for all the considered periods at Cabauw (WRCT approach).

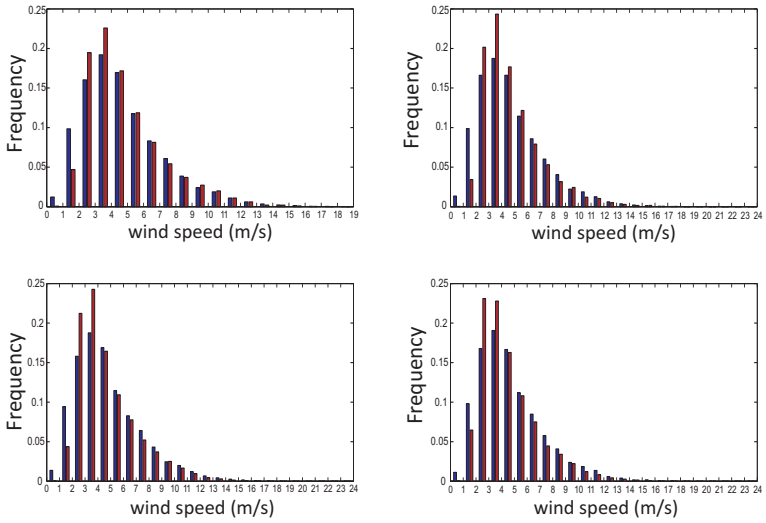
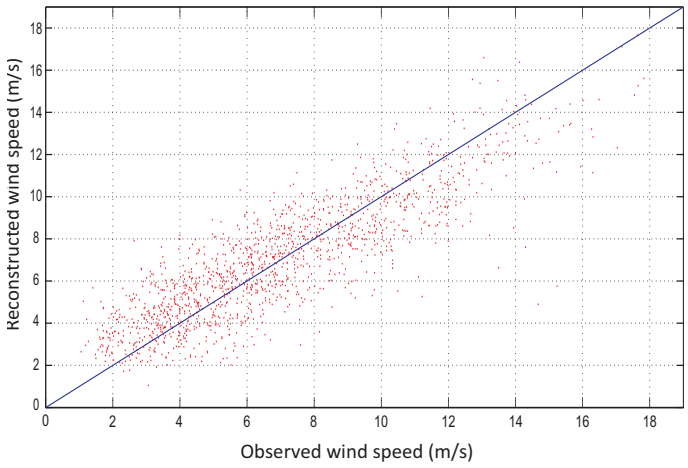


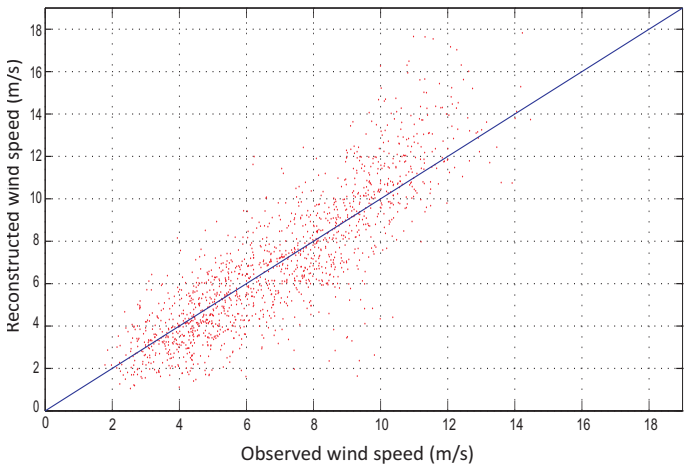
Figure 4.5: Histograms of the wind speed (reconstructed and observed ones), for all the considered periods at Cabauw (SVMr-GA approach).

The algorithms complexity and computational cost of the compared approaches are also

important elements to be discussed. The SVMr-GA approach has 3 parameters (C, ϵ and γ) that must be estimated previously to its application (see above). On the other hand the WRCT approach needs to estimate the optimum number of clusters k in the k-means approach, which is also done in previous tests. Regarding the computational time required by each approach, both require approximately the same computation time, that includes all the preprocessing of files and algorithm's parameters, about 20 minutes.



(a)



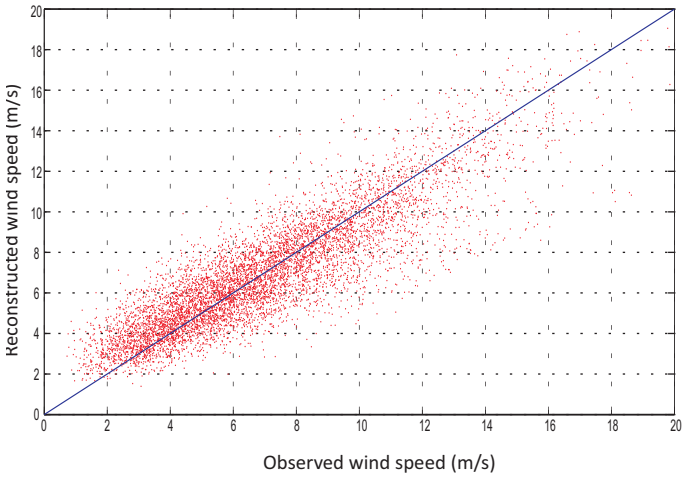
(b)

Figure 4.6: Wind speed reconstruction in Cp1 test period at Capel; (a) WRCT approach; (b) SVMr-GA approach.

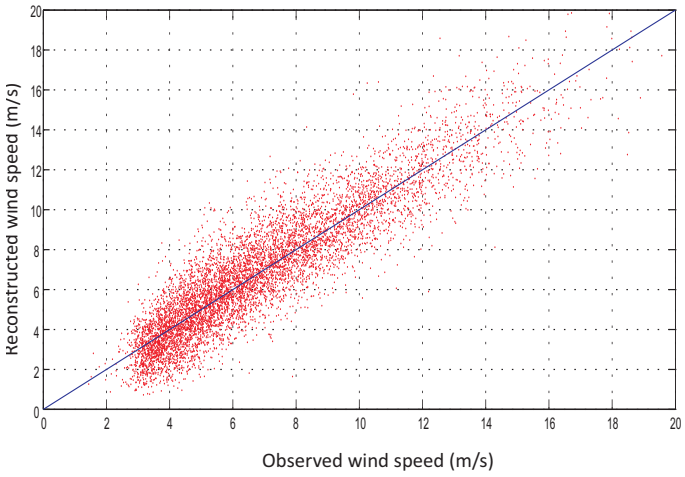
4.3.2.2 Capel

Results in Capel site are reduced to just one test period (Cp1). Table 4.2 (middle of the table) shows the obtained results with the WRCT and SVMr-GA methods compared, respectively at this site. In this case the performance of both approaches is quite similar, and no real differences can be detected in the statistical measures used. This point is certified by Figure 4.6, that shows the dispersion graph for Capel, where it is not possible to observe strong differences between the performance of either algorithms in this case. Finally, Table 4.3 (middle of the table) shows the

reconstruction of Weibull distribution parameters (A and k) for the wind speed at Capel, with the WRCT and SVMr-GA. In this case, it seems that the wind speed reconstruction is slightly more accurate with the WRCT, since the SVMr-GA slightly overestimates these parameters.



(a)



(b)

Figure 4.7: Wind speed reconstruction in K1 test period at Kaegnes; (a) WRCT approach; (b) SVMr-GA approach.

4.3.2.3 Kaegnes

The final tests of this comparative are carried out at Kaegnes, where three test periods are considered. Table 4.2 (at the bottom of the table) shows the results obtained with the two approaches considered. In this case, the performance of the algorithms varies with the test period considered. In the period K1 the SVMr-GA outperforms the WRCT, whereas in periods K2 and K3 it is the WRCT that obtains the best results. The differences in all cases are small. Table 4.3 (at the bottom of the table) shows the reconstruction of the Weibull distribution parameters (A and k) of the wind speed at Kaegnes. In this case the best reconstruction of these parameters are given by the SVMr-GA, and it is the WRCT that overestimates these parameters

at this site. Figure 4.7 shows the dispersion graph for Kaegnes (period K1) obtained with the WRCT and SVMr-GA algorithms.

4.4 Conclusions

We have presented a comparison between two different methods for wind speed reconstruction based on synoptic variables. Specifically, the comparison between a machine learning approach (SVMr-GA) and a physical-based algorithm (WRCT) is carried out. The SVMr-GA approach is a hybrid algorithm composed by a Support Vector Regression and a Genetic Algorithm. The WRCT works by carrying out a first clustering step over a synoptic field. Then a regression technique between wind speed and any meteorological variable is applied. We have fully described both approaches and compared them within the frame of a real wind speed reconstruction problem at three different sites: Cabauw (The Netherlands), Capel (Wales, UK) and Kaegnes (Denmark). The results obtained have shown a good performance of both approaches, with slightly prevalence of one or another depending on the site or reconstruction objective considered. The proposed approaches have shown to be excellent tools for wind speed analysis in sites where there are long past periods without wind speed measurements.

These two tools open the possibility of improving long term estimation of the wind resource of a particular area given the fact that this is a very challenging problem when developing a wind farm. Specifically, this problem is aggravated when there are few years of measurements and there are no nearby meteorological masts to apply classical MCP (Measure-Correlate-Predict) methodologies. This situation is very usual, especially in countries with recent development of wind energy and scarce data from meteorological stations. Moreover, these two methodologies can help on the financing stage as well as to reduce the uncertainty of private investment in wind energy projects.

Appendix for this chapter

4.A Tables

Table 4.1: Average number of weather types (clusters) determined for each group of three months (JFM, AMJ, JAS, OND) at each site considered with the WRCT methodology.

site	JFM	AMJ	JAS	OND
Cabauw	3	4	3	5
Capel	3	2	3	2
Kaegnes	3	5	3	6

Table 4.2: Wind speed reconstruction results for various statistics obtained with the WRCT and the SVMr-GA approach at all the sites considered.

	Set	MAE	RMSE	Bias	R^2
Cabauw (WRCT)					
	C1	0.8368	1.1017	0.1796	0.8205
	C2	0.8137	1.0805	0.0667	0.8241
	C3	0.8227	1.0908	-0.0089	0.8262
	C4	0.8078	1.0909	-0.2018	0.8279
Cabauw (SVMr-GA)					
	C1	0.8536	1.1093	0.0659	0.8086
	C2	0.8902	1.1593	-0.0136	0.7949
	C3	0.8775	1.1534	-0.1441	0.8057
	C4	0.8657	1.1392	-0.2564	0.8057
Capel (WRCT)					
	Cp1	1.32	1.68	0.05	0.74
Capel (SVMr-GA)					
	Cp1	1.28	1.71	0.02	0.73
Kaegnes (WRCT)					
	K1	1.17	1.56	-0.01	0.77
	K2	1.16	1.55	0.12	0.78
	K3	1.22	1.65	-0.05	0.75
Kaegnes (SVMr-GA)					
	K1	1.04	1.35	-0.03	0.83
	K2	1.15	1.66	-0.03	0.74
	K3	1.21	1.86	0.04	0.68

Table 4.3: Comparison of the Weibull distribution parameters of the wind (A and k) reconstructed versus observed ones at all the sites considered, obtained with the SVMr-GA and WRCT approaches.

	Set	A Observed	A Reconstructed	k Observed	k Reconstructed
Cabauw (WRCT)					
	C1	5.32	5.40	1.96	2.12
	C2	5.32	5.32	1.94	2.17
	C3	5.41	5.27	1.93	2.13
	C4	5.33	5.07	1.93	2.08
Cabauw (SVMr-GA)					
	C1	5.32	5.52	1.96	2.09
	C2	5.32	5.40	1.94	2.07
	C3	5.41	5.42	1.93	2.12
	C4	5.33	5.12	1.93	2.00
Capel (WRCT)					
	Cp1	7.86	7.87	2.25	2.62
Capel (SVMr-GA)					
	Cp1	7.86	7.82	2.25	2.88
Kaegnes (WRCT)					
	K1	7.86	7.82	2.28	2.53
	K2	7.94	8.06	2.28	2.51
	K3	7.95	7.88	2.28	2.43
Kaegnes (SVMr-GA)					
	K1	7.86	7.81	2.28	2.46
	K2	7.94	7.90	2.28	2.46
	K3	7.95	7.99	2.28	2.35

Chapter 5

Wind speed reconstruction and prediction from neighbour towers data

5.1 Introduction

Wind farm management and prospection usually tackle with two important problems: wind speed prediction and wind series reconstruction. Wind speed prediction and reconstruction of wind series are usually carried out in wind farms using data from in-situ measuring towers, using the so-called Measure-Correlate-Predict methods (MCP). MCP processes consist, therefore, in the wind speed prediction or reconstruction from neighbor stations, using different methods as it has been explained in Section 2.2. Traditional MCP processes deal with pre-processed data from measuring towers in order to have a full set of data from all the towers without gaps. But in real operations in a wind farm, real time applications are needed, so the processing of data to tackle the problem with a traditional MCP method is not feasible. We deal with Real MCP Operations (RMCPO) in a wind farm, which are dynamic processes which often require the continuous re-training of the algorithms to consider new data. Our method will consider direct data from the measuring stations and will apply the corresponding model to reconstruct or predict the wind speed. The algorithms used to reconstruct or predict the wind series must be extremely fast in order to reuse new input data. Thus, we propose the application of two state-of-the-art neural networks which have shown a very fast training and excellent performance in terms of accuracy. Specifically, Group Method of Data Handling and Extreme Learning Machines are applied to the MCP reconstruction and prediction of wind speed series, in a real wind farm in Spain.

This chapter is structured as follows: next section summarizes the RMCPO problem definition. Section 5.4 presents the experimental part, with a comparison with alternative multilinear regression and different computational intelligence-based regression algorithms, such as multilayer perceptrons or support vector machines. This section also discusses the performance of the proposed approach in reconstruction and prediction problems. In this section we also present the real software developed for RMCPO in wind farms. Section 5.5 closes the chapter giving some final conclusions.

5.2 Problem definition

Let $\mathcal{D}$ be a $N \times M$ matrix, where N stands for the number of measuring stations in the wind farm, and M stands for the total number of wind speed measures available. The RMCPO problem consists of, given $\mathcal{D}$, training a set of $n(\mathcal{D})$ regression models $\mathcal{R}$, for covering all the possible

combinations of available wind speed data and missing values to be reconstructed (or to predict a wind speed value in a future time at a given station), in such a way that each regression model $\mathcal{R}_i,\ i = 1, \ldots, n(\mathcal{D})$ must contain, at least, two input features from $\mathcal{D}$.

Just as an example, consider a wind farm in which 4 measuring towers have been installed. Let us consider 10 values of wind in each tower to illustrate the neural models generation in order to reconstruct missing values.

$$\mathcal{D} = \begin{bmatrix} 5.1 & \mathbf{X} & 3.8 & 2.9 \\ 6.0 & 7.2 & \mathbf{X} & X \\ X & X & 1.3 & 1.2 \\ 9.4 & 6.2 & 4.8 & 7.9 \\ X & 9.6 & X & 8.5 \\ 6.2 & 2.9 & 4.0 & X \\ 4.1 & 7.2 & 4.3 & 6.8 \\ 4.5 & 5.2 & 6.1 & X \\ 2.1 & 3.6 & 5.1 & 3.8 \\ 8.6 & 9.2 & 10.0 & X \end{bmatrix}. \tag{5.1}$$

For example, let us suppose that we want to reconstruct the missing value (marked in boldface) in the first row of $\mathcal{D}$, corresponding to the wind speed measured by tower #2. Note that in this case, we need to train a model in which all the values in the four towers are available (measures of towers #1, #3 and #4 as features, and measure of tower #2 as objective variable). Thus, we must use all rows in $\mathcal{D}$ which have all the values complete, i.e, rows (samples) 4, 7 and 9 in this case. Imagine now that we need to reconstruct the variable in the second row of $\mathcal{D}$ marked in boldface. In this case we need all the samples in $\mathcal{D}$ which contains features #1, #2 and #3 (#1 and #2 as input features and #3 as objective variable) to train the network. We can use rows (samples) 4, 6, 7, 8, 9 and 10 to train this model.

Note that the total number of models needed ($n(\mathcal{D})$) to ensure the complete reconstruction of a matrix $\mathcal{D}$ (having at least two input variables), completely depends on size N of $\mathcal{D}$ (number of wind speed measuring stations). The expression for $n(\mathcal{D})$ is the following:

$$n(\mathcal{D}) = \sum_{i=0}^{N-3} \frac{N!}{i!(N-i-1)!} \tag{5.2}$$

Note that for $N = 5$, the total number of neural models that we have to train to cover all possible combinations of missing and existing values is $n(\mathcal{D}) = 55$. This number, however, increases dramatically with N, so for $N = 8$ we have to train 960 models to cover all the possible combinations in $\mathcal{D}$.

5.3 Fast training neural networks for RMCPO problems

In order to tackle RMCPO instances when the value of N increases, very fast training neural methods must be applied to obtain results in a reasonable time. This necessity is even harder when we tackle MCP short-term prediction problems, where often we have to incorporate new data and refresh the models to carry out the predictions. We propose two models of neural

networks that share the property of having an extremely fast training process. The first one is a polynomial self-organized neural network called Group Method of Data Handling, whereas the second proposal is the Extreme Learning Machine. Both approaches have recently received the attention of the research community due to their good performance in accuracy and excellent computation time required and have been explained in Chapter 3.

5.4 Experimental part

In this section a real case of study in a Spanish wind farm is tackled. The methodology used is explained and alternative MCP algorithms are proposed for comparison tasks.

5.4.1 Alternative MCP algorithms for comparison

Different modern MCP techniques based on computational intelligence algorithms are proposed to compare the fast training approaches. Specifically, a multi-layer perceptron and a support vector machine are used (both described in Sections 3.1.1 and 3.1.4). We briefly describe a classical linear MCP technique, the multiple linear regression that can be used as a reference algorithm.

5.4.1.1 Multiple Linear Regression

Multiple linear regression (MLR) is an extension of the simple linear regression model, in which the relationship between a dependent variable and several independent variables can be calculated. The general model for k variables is:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_k x_{i,k} + e_i, \quad i = 1, \ldots, n \tag{5.3}$$

as shown in [18]. In multiple linear regression, the least-squares method is used also to determine the regression coefficients, β_i, in Equation (5.3).

The error term in Equation (5.3), e_i, is unknown because the true model is unknown. Once the model is predicted, the residual errors are defined as:

$$e_i = y_i - \hat{y}_i \tag{5.4}$$

where y_i stands for the real value and $\hat{y}_i$ stands for the predicted value. These residual errors give an idea of how accurate the prediction obtained is. In this work the Matlab implementation of the MLR algorithm described in [18] is used. An example in the *carsmall* dataset available in Matlab is shown in Figure 5.1. Given the horsepower and the weight of different vehicles, MPG (miles per gallon) is regressed (from the given MPG data). A surface of regression is obtained in order to predict MPG of new vehicles given its horsepower and weight. In the RMCPO problem considered, the MLR is the basic reference algorithm to be improved, since many companies devoted to managing wind farms use this approach in MCP problems.

5.4.2 Experimental evaluation in a real wind farm: methodology

In order to analyze the performance of the two proposed algorithms, several experiments of wind series reconstruction and prediction have been tackled, over real data from a wind farm in Spain.

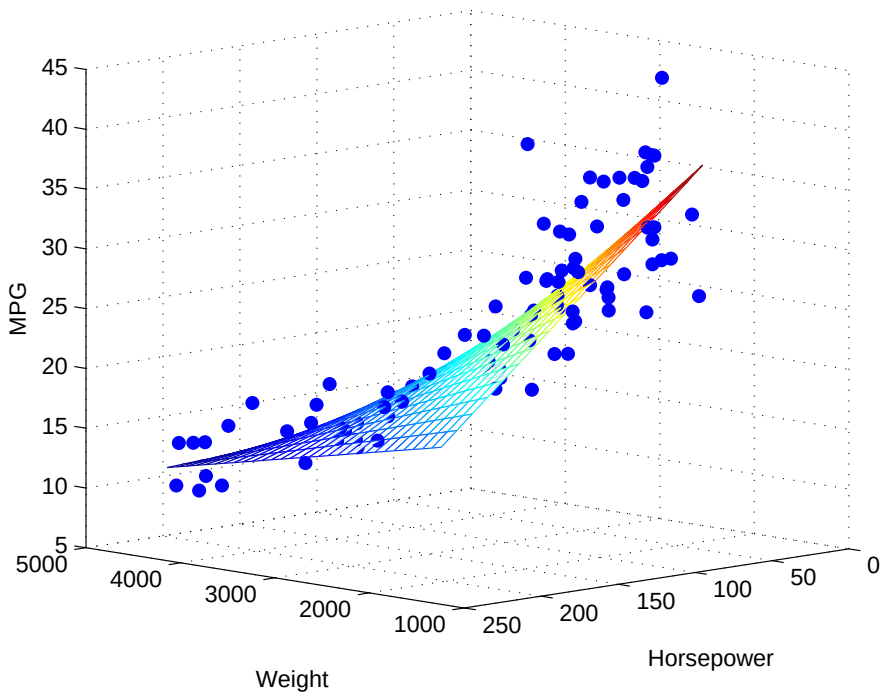


Figure 5.1: Example of Multiple Linear Regression, with *carsmall* MatLab dataset.

Figure 5.2 shows the situation of the eight measuring towers considered, within the same wind farm, situated in northern Guadalajara, Spain.

Measured hourly wind data from November 2008 to November 2010 are available. We have carried out the following experiments: first, in order to focus on the specific reconstruction and prediction performance of the algorithms proposed, we have tackled a round of experiments consisting in MCP reconstruction and prediction using only complete samples (samples without missing values). There is available a total of 239 complete temporal series, of different length, with a total of 4391 hours of average wind speed data for the eight towers. The distribution length of the considered wind series is shown in Figure 5.3.

In the same way, the problem of predicting the wind speed in each of the towers from the measures of the other towers is then tackled. In order to face these first two problems with complete samples, the initial data have been split in train, validation and test sets. The validation set is used in the MLP training, in order to avoid over-fitting. To obtain independent sets with a composition balanced over the seasons, the 239 time series have been split as shown in Figure 5.4. These sets will be used to train and test the different methods considered. The distribution in seasons of the samples in the train, validation and test sets considered is shown in Figure 5.5. In a second round of experiments, we analyze the performance of the ELM and GMDH in a RMCPO, consisting in the reconstruction of the missing values in the set of eight considered towers without processing data.

5.4.3 Evaluation of the GMDH and ELM in MCP wind speed reconstruction

First, the problem of wind speed series reconstruction using the complete data from neighbor towers is tackled. In this problem, the wind values of different neighbor towers are used as input

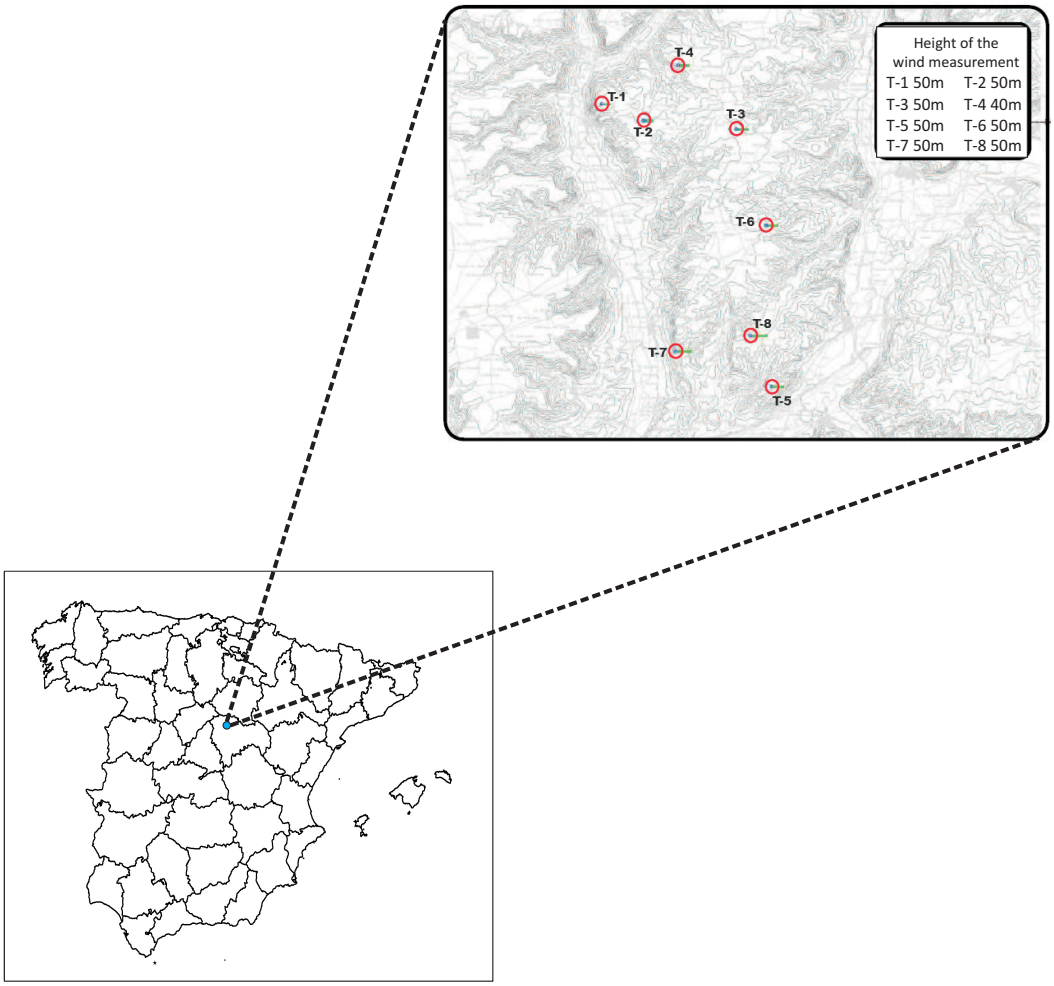


Figure 5.2: Situation of the wind measuring towers in Spain and within the wind farm.

in the regression techniques, in order to estimate the wind at the same time in an objective tower. Since we have available all the data in each tower, we can evaluate the accuracy of each considered method. This evaluation is carried out in terms of different well-known statistical evaluation indices in each tower: such as root mean square error (RMSE), which provides information about the mean error in the wind speed reconstruction, the Mean Bias Error (MBE), which checks whether the model overestimates or underestimates the wind speed, the Coefficient of Determination (R^2), which provides information about the percentage of the variance that the model is able to explain, and the Index of Agreement (IoA), which gives information about how close the predicted wind speed values are to the observed ones. These performance indices have been used profusely before in different studies, including in environmental applications [40, 120, 179]. We also compare the algorithms in terms of the final computation time for training each method, i.e., the time taken by each method to generate the model to reconstruct each tower (one model). We have carried out wind speed reconstruction and prediction for all the towers by using a different number of neighbor towers: the 3 nearest towers to the target one (called 3T case), the 5 nearest (called 5T case) and all the available towers (called 7T case). Table 5.1 shows a summary of the towers used to reconstruct/predict each target tower. This table can be seen together with Figure 5.2, where the actual disposition of the towers is shown. The table details which

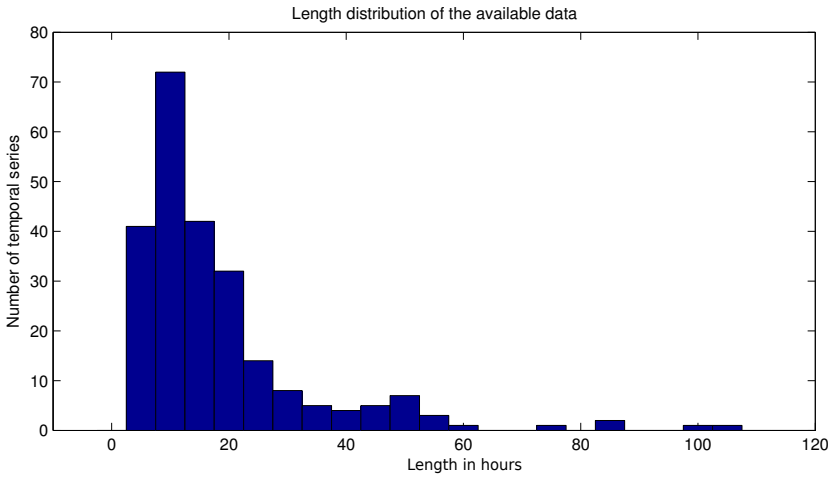


Figure 5.3: Distribution of the wind speed time series (complete samples) in terms of their duration.

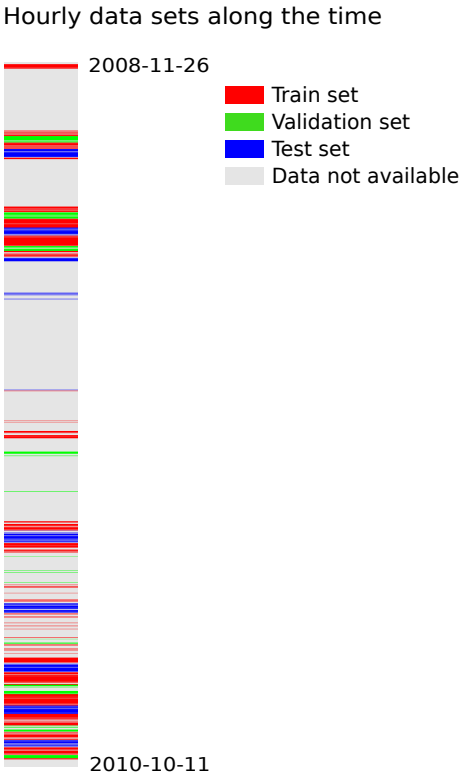


Figure 5.4: Distribution of the sets for training, validation and test.

towers are used in the 3T case (using the 3 nearest towers to do the reconstruction/prediction) and in the 5T case (using the 5 nearest towers to do the reconstruction/prediction). Note that in the 7T case all the towers but the target one are used in the reconstruction/prediction process.

Tables 5.2 to 5.6 show the performance of the GMDH and ELM, and the alternative method for comparison (MLR, MLP and SVMr), in terms of the different evaluation indices considered.

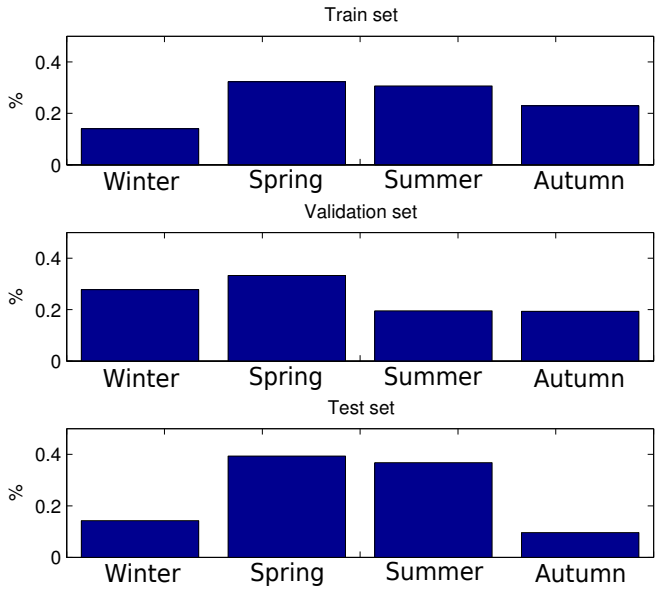


Figure 5.5: Balanced distribution over the time of the sets.

Table 5.1: Towers used to reconstruct/predict the target tower, by using a different number of neighbor towers (the 3 nearest (3T case) or the 5 nearest (5T case)). Note that in the 7T case all the towers but the target one are used in the reconstruction/prediction process.

Target tower	Nearest towers (3T case)	Nearest towers (5T case)
1	2, 3, 4	2, 3, 4, 6, 7
2	1, 3, 4	1, 3, 4, 6, 7
3	1, 2, 4	1, 2, 4, 6, 8
4	1, 2, 3	1, 2, 3, 6, 8
5	6, 7, 8	2, 3, 6, 7, 8
6	2, 3, 8	2, 3, 5, 7, 8
7	5, 6, 8	2, 3, 5, 6, 8
8	5, 6, 7	2, 3, 5, 6, 7

Note that the performance of both the ELM and GMDH is quite acceptable, with evaluation indices comparable to the obtained by the compared approaches. The SVMr seems to be the most accurate method among all tested and the MLR (reference method) is the one which provides the poorer results. Regarding the differences between the wind speed reconstruction using 3, 5 or 7 towers, it is interesting to see how the reconstruction using 3 reference towers (3T case) provides poorer results than the reconstruction using 5 (5T case) or 7 (7T case). However, the reconstruction using 5 towers (5T case) is, in many cases, better than the one using 7 towers. This indicates that the information provided by 5 towers is enough to obtain a good quality reconstruction of the wind speed, whereas the information of the 3 nearest towers is not enough to provide the best possible reconstruction.

Table 5.7 shows the training time in the reconstruction problem, in the case of data from 7

towers (7T case). This table shows one of the main advantages of using GMDH or ELM networks in wind speed reconstruction. Note that the ELM approach is extremely fast, with training time comparable to the MLR approach (about 30 ms per training). The GMDH takes about 1 second per training model, but it is still less than the MLP, and of course less than the SVMr approach, which is the algorithm that employs more time in a single model training.

5.4.4 Evaluation of the GMDH and ELM in MCP wind speed prediction

The prediction problem is quite similar to the reconstruction one, but in this case the regressors include wind measures in past time (last hour) at the neighbor towers in order to predict the current wind value at the current tower. Tables 5.8 to 5.12 show the results obtained by the different techniques considered in the prediction problem, with different number of towers to train the predictors (3T, 5T and 7T cases). Note that the results obtained in the prediction problem are slightly worse in terms of accuracy (in all evaluation indices) for the compared algorithms, as expected. However, again the results obtained by the ELM and GMDH are quite competitive to the other tested algorithms, and improve the results of the reference method (MLR). We can also observe in this problem the same effect regarding the number of reference towers used to do the prediction: the results with 3 towers are worse than with 5 and 7 towers in the prediction, but the results using 5 towers are, in average, slightly better than the results obtained using 7 towers in the prediction. Table 5.13 shows the training time obtained by the compared algorithms in each tower, for the case of predictions using data from 7 towers. As in the previous case, the training time of the ELM and GMDH is less than the other neural approaches compared.

5.4.5 A software for RMCPO problems: real wind series reconstruction and prediction

This section presents a software tool implementing the ELM and GMDH, for RMCPO problems in wind farms. The tool has been developed in Matlab, and it is currently under operation in different wind farms in Spain and USA. Figure 5.6 shows the frontend of the tool, where we can set different parameters and options (ELM or GMDH selection, MCP reconstruction or prediction, generation of the models, selection of the file with the wind speed values (matrix $\mathcal{D}$, etc.). A first step in any RMCPO problem is the generation of the regression models for the chosen algorithm (ELM or GMDH), that is showed in the tool with a progress bar, as shown in Figure 5.7. At this stage, the tool is able to reconstruct or predict the wind speed values of the matrix $\mathcal{D}$. As an example, if we consider the GMDH network, the tool is able to provide the network's structure obtained (Figure 5.8), and finally, a complete reconstruction of matrix $\mathcal{D}$ if required, or a prediction of the wind speed in each tower from the neighbor ones. For example, Figures 5.9 shows the wind speed prediction in Tower 6 obtained by our software (with the GMDH network) and the comparison with the real wind speed measured in that tower. On the other hand, the ELM network is probabilistic, so we can obtain different predictions (depending on network initialization). Figure 5.10 shows the best and worst prediction obtained for Tower 6.

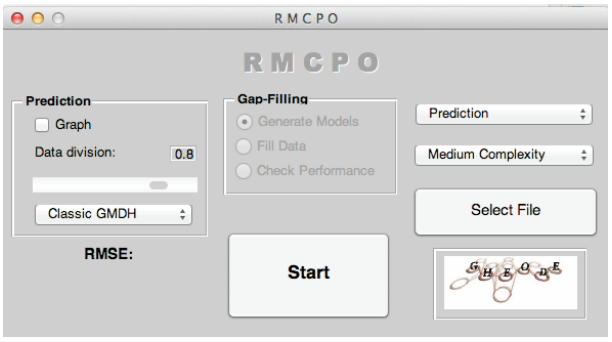


Figure 5.6: Frontend of the proposed software for wind speed reconstruction and prediction.

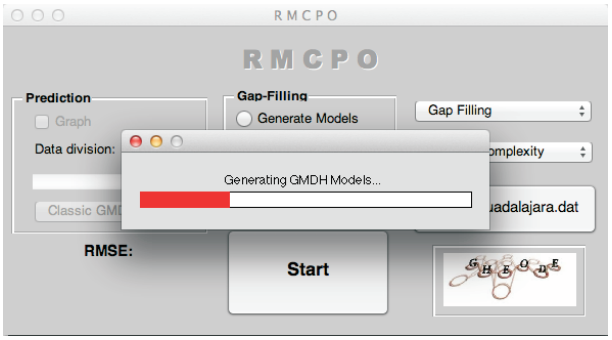


Figure 5.7: Progress bar in the proposed software indicating that GMDH or ELM models for wind speed reconstruction or prediction are being computed.

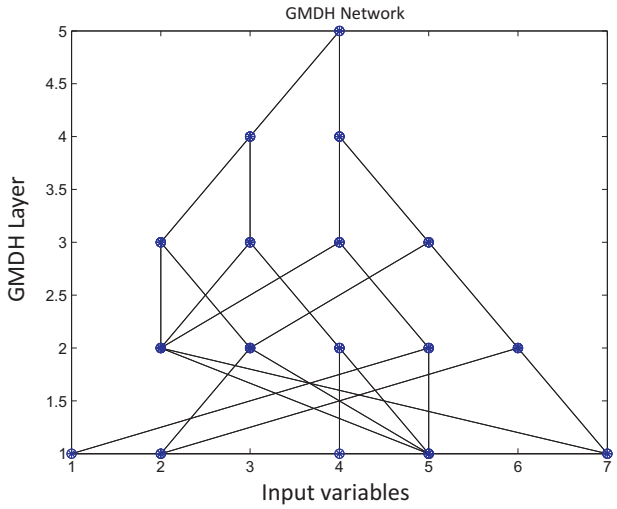


Figure 5.8: GMDH structure obtained by the proposed software tool in a wind speed prediction problem in the considered wind farm.

5.5 Conclusions

We have applied two very fast training neural network to two real MCP operation problems (wind series reconstruction and prediction), i.e. wind series management from neighbor stations

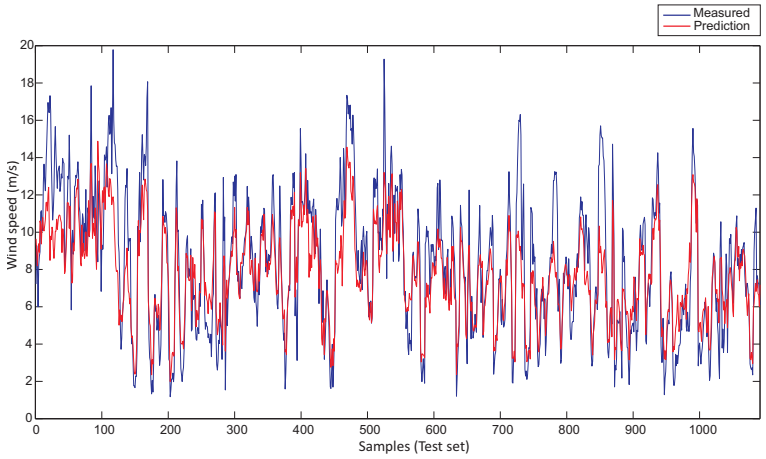
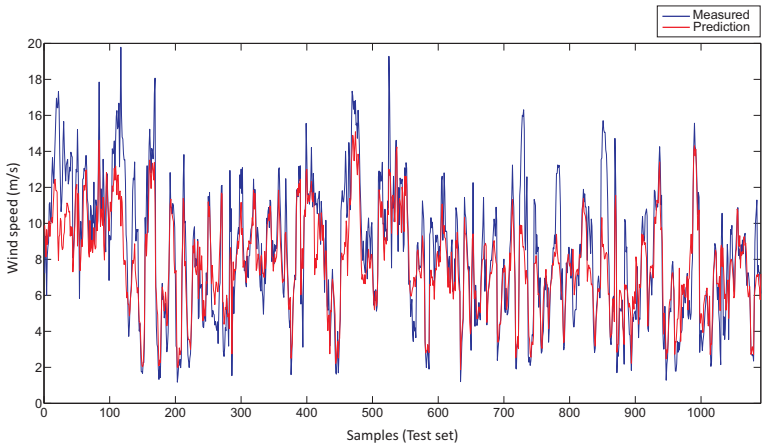
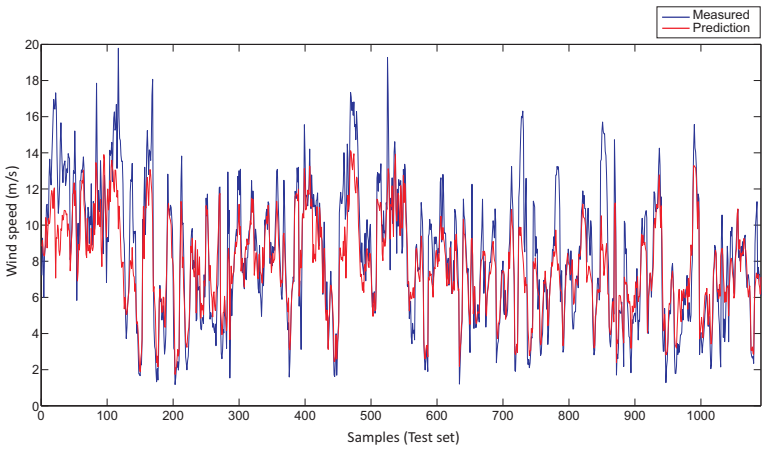


Figure 5.9: Wind speed prediction in tower 6 of the considered wind farm obtained by the GMDH network (prediction using data from 7 towers).

in a wind farm. Specifically, we have applied a Group Method of Data Handling network, and an Extreme Learning Machine, which are able to obtain good results in terms of accuracy, within a extreme fast computation time. A complete set of experiments have evaluated the performance of these two approaches in a real application of wind series reconstruction in a real wind farm at Guadalajara, Spain, with excellent results. A software that implements fast versions of the GMDH and ELM has also been described and tested. This software can be used to carry out fast wind speed reconstruction and short term prediction of wind speed from reference measuring towers in wind farms. This software tool is therefore an useful tool for wind farm managers to improve the processing of wind speed data in wind farms.



(a)



(b)

Figure 5.10: Wind speed prediction in tower 6 of the considered wind farm obtained by the ELM network (prediction using data from 7 towers); (a) Best prediction obtained; (b) Worst prediction obtained.

Appendix for this chapter

5.A Tables

Table 5.2: Wind speed reconstruction results obtained by the GMDH network.

GMDH reconstruction 3T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.955	0.019	0.977	0.950
2	0.952	0.001	0.987	0.676
3	0.946	0.002	0.9823	0.738
4	0.938	0.058	0.983	0.761
5	0.965	-0.001	0.985	0.818
6	0.781	-0.009	0.976	0.847
7	0.951	0.001	0.969	1.263
8	0.972	-0.006	0.986	0.789
GMDH reconstruction 5T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.928	0.010	0.978	0.928
2	0.950	-0.003	0.990	0.599
3	0.974	0.003	0.985	0.690
4	0.941	0.047	0.984	0.730
5	0.969	-0.009	0.987	0.773
6	0.787	-0.018	0.975	0.868
7	0.987	0.013	0.970	1.256
8	0.985	-0.010	0.987	0.770
GMDH reconstruction 7T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.928	0.010	0.978	0.928
2	0.949	-0.003	0.990	0.585
3	0.973	0.003	0.985	0.690
4	0.941	0.047	0.984	0.730
5	0.969	-0.009	0.987	0.773
6	0.796	-0.030	0.975	0.871
7	0.926	-0.041	0.971	1.248
8	0.985	0.005	0.987	0.760

Table 5.3: Wind speed reconstruction results (average values of 30 runs) obtained by the ELM network.

ELM reconstruction 3T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.955	-0.090	0.978	0.943
2	0.970	0.086	0.989	0.616
3	0.941	0.019	0.985	0.676
4	0.922	-0.064	0.983	0.748
5	0.950	0.009	0.987	0.768
6	0.849	-0.049	0.978	0.835
7	0.839	-0.186	0.966	1.289
8	0.983	0.059	0.988	0.737
ELM reconstruction 5T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.978	-0.062	0.978	0.949
2	0.999	0.094	0.989	0.621
3	0.942	0.017	0.985	0.675
4	0.937	-0.044	0.984	0.740
5	0.962	0.014	0.986	0.785
6	0.858	-0.055	0.977	0.849
7	0.831	-0.199	0.968	1.249
8	0.978	0.046	0.987	0.744
ELM reconstruction 7T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.974	-0.034	0.975	0.994
2	0.985	0.054	0.989	0.618
3	0.955	0.017	0.983	0.717
4	0.965	-0.044	0.982	0.771
5	0.967	0.023	0.986	0.800
6	0.865	-0.048	0.975	0.880
7	0.838	-0.157	0.967	1.264
8	0.988	0.039	0.986	0.791

Table 5.4: Wind speed reconstruction results obtained by the MLR method (Reference).

MLR reconstruction 3T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.937	-0.084	0.976	0.966
2	0.961	0.080	0.989	0.611
3	0.944	0.038	0.983	0.712
4	0.959	-0.059	0.984	0.747
5	0.968	0.021	0.987	0.770
6	0.830	-0.053	0.977	0.835
7	0.822	-0.190	0.964	1.312
8	0.975	0.071	0.986	0.780
MLR reconstruction 5T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.962	-0.059	0.978	0.936
2	0.981	0.087	0.990	0.599
3	0.953	0.032	0.9841	0.702
4	0.949	-0.045	0.984	0.727
5	0.963	0.016	0.986	0.766
6	0.824	-0.059	0.978	0.830
7	0.806	-0.202	0.966	1.280
8	0.959	0.057	0.987	0.764
MLR reconstruction 7T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.963	-0.052	0.978	0.943
2	0.983	0.057	0.990	0.590
3	0.954	0.031	0.984	0.700
4	0.948	-0.048	0.984	0.727
5	0.963	0.019	0.987	0.767
6	0.826	-0.058	0.977	0.837
7	0.812	-0.169	0.966	1.271
8	0.962	0.041	0.987	0.755

Table 5.5: Wind speed reconstruction results (average values of 30 runs) obtained by the MLP.

MLP reconstruction 3T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.958	-0.080	0.977	0.950
2	0.968	0.084	0.989	0.616
3	0.938	-0.010	0.985	0.677
4	0.911	-0.097	0.981	0.783
5	0.942	0.014	0.986	0.762
6	0.836	-0.066	0.977	0.846
7	0.834	-0.169	0.966	1.292
8	0.992	0.069	0.987	0.755
MLP reconstruction 5T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.965	-0.056	0.978	0.930
2	0.982	0.086	0.989	0.605
3	0.937	-0.003	0.985	0.669
4	0.933	-0.039	0.984	0.724
5	0.946	0.006	0.987	0.760
6	0.840	-0.052	0.976	0.852
7	0.819	-0.238	0.967	1.269
8	0.970	0.043	0.987	0.751
MLP reconstruction 7T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.968	-0.059	0.978	0.935
2	0.981	0.056	0.990	0.590
3	0.952	0.012	0.986	0.668
4	0.931	-0.046	0.983	0.748
5	0.945	0.018	0.987	0.766
6	0.853	-0.073	0.975	0.878
7	0.844	-0.130	0.969	1.232
8	0.982	0.041	0.988	0.729

Table 5.6: Wind speed reconstruction results obtained by the SVMr.

SVMr reconstruction 3T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.952	-0.141	0.977	0.961
2	0.972	0.088	0.989	0.616
3	0.953	0.038	0.985	0.672
4	0.931	-0.011	0.983	0.747
5	0.965	0.081	0.987	0.772
6	0.833	-0.015	0.977	0.844
7	0.832	-0.129	0.967	1.275
8	0.984	-0.036	0.988	0.734
SVMr reconstruction 5T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.979	-0.106	0.978	0.930
2	0.978	0.069	0.990	0.599
3	0.955	0.039	0.986	0.665
4	0.944	-0.019	0.984	0.726
5	0.965	0.077	0.987	0.768
6	0.807	-0.043	0.977	0.844
7	0.789	-0.211	0.967	1.252
8	0.981	-0.036	0.988	0.738
SVMr reconstruction 7T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.984	-0.099	0.978	0.941
2	0.976	0.058	0.990	0.590
3	0.958	0.039	0.986	0.661
4	0.940	-0.029	0.984	0.727
5	0.965	0.078	0.987	0.767
6	0.817	-0.048	0.977	0.845
7	0.831	-0.121	0.970	1.214
8	0.979	-0.036	0.989	0.704

Table 5.7: Computation time of wind speed reconstruction (using data from 7 towers) in each considered tower, in the example with complete wind speed samples (in seconds).

Tower	ELM	GMDH	MLP	SVMr	MLR (Reference)
T1	0.045	1.078	20.001	584.5	0.028
T2	0.033	1.149	21.315	617.1	0.023
T3	0.034	0.788	20.975	586.5	0.029
T4	0.033	0.891	22.850	601.3	0.030
T5	0.033	0.982	21.153	596.7	0.027
T6	0.036	1.034	22.179	589.9	0.027
T7	0.035	1.101	20.336	598.1	0.031
T8	0.042	0.972	22.891	619.2	0.024

Table 5.8: Wind speed prediction results obtained by the GMDH network.

GMDH prediction 3T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.836	0.004	0.945	1.409
2	0.875	-0.023	0.957	1.182
3	0.842	-0.012	0.939	1.305
4	0.833	0.022	0.946	1.282
5	0.884	-0.032	0.947	1.470
6	0.702	-0.054	0.934	1.346
7	0.843	-0.030	0.930	1.823
8	0.867	-0.004	0.951	1.406
GMDH prediction 5T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.775	0.016	0.943	1.400
2	0.846	-0.022	0.956	1.181
3	0.867	-0.003	0.943	1.272
4	0.833	0.023	0.946	1.285
5	0.897	-0.043	0.952	1.410
6	0.717	-0.031	0.939	1.305
7	0.899	-0.015	0.933	1.804
8	0.920	-0.015	0.955	1.361
GMDH prediction 7T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.777	0.017	0.944	1.395
2	0.868	-0.029	0.957	1.185
3	0.867	-0.003	0.943	1.272
4	0.841	-0.036	0.944	1.310
5	0.926	-0.075	0.949	1.459
6	0.717	-0.031	0.939	1.305
7	0.934	-0.074	0.936	1.785
8	0.948	-0.044	0.954	1.387

Table 5.9: Wind speed prediction results (average values of 30 runs) obtained by the ELM network.

ELM prediction 3T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.835	-0.094	0.946	1.394
2	0.870	0.070	0.957	1.187
3	0.813	-0.004	0.943	1.261
4	0.822	-0.023	0.944	1.299
5	0.849	0.049	0.950	1.416
6	0.760	-0.048	0.941	1.297
7	0.748	-0.157	0.927	1.813
8	0.888	0.095	0.955	1.358
ELM prediction 5T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.861	-0.060	0.948	1.374
2	0.889	0.063	0.957	1.187
3	0.826	0.033	0.941	1.281
4	0.850	-0.014	0.946	1.285
5	0.860	0.038	0.950	1.424
6	0.762	-0.061	0.941	1.300
7	0.754	-0.172	0.931	1.760
8	0.891	0.073	0.955	1.358
ELM prediction 7T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.895	-0.033	0.946	1.412
2	0.898	0.057	0.956	1.205
3	0.842	0.037	0.940	1.298
4	0.867	-0.001	0.944	1.317
5	0.866	0.071	0.949	1.435
6	0.775	-0.033	0.940	1.309
7	0.755	-0.126	0.930	1.768
8	0.891	0.079	0.955	1.365

Table 5.10: Wind speed prediction results obtained by the MLR method (Reference).

MLR prediction 3T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.853	-0.081	0.945	1.408
2	0.876	0.063	0.958	1.175
3	0.832	0.037	0.939	1.305
4	0.869	-0.035	0.947	1.282
5	0.861	0.057	0.949	1.436
6	0.746	-0.053	0.941	1.297
7	0.733	-0.154	0.924	1.834
8	0.877	0.106	0.952	1.394
MLR prediction 5T				
Tower	R^2	MBE (m/s	IoA	RMSE (m/s)
1	0.885	-0.051	0.949	1.377
2	0.890	0.068	0.958	1.170
3	0.840	0.048	0.941	1.291
4	0.863	-0.021	0.948	1.273
5	0.849	0.040	0.950	1.413
6	0.743	-0.060	0.941	1.295
7	0.714	-0.174	0.927	1.789
8	0.856	0.081	0.953	1.370
MLR prediction 7T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.885	-0.047	0.948	1.381
2	0.891	0.045	0.959	1.171
3	0.838	0.040	0.941	1.284
4	0.860	-0.026	0.948	1.270
5	0.850	0.048	0.949	1.422
6	0.740	-0.052	0.942	1.281
7	0.726	-0.132	0.927	1.795
8	0.857	0.077	0.954	1.362

Table 5.11: Wind speed prediction results (average values of 30 runs) obtained by the MLP.

MLP prediction 3T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.833	-0.081	0.945	1.401
2	0.877	0.072	0.957	1.188
3	0.802	-0.003	0.942	1.265
4	0.821	-0.046	0.944	1.295
5	0.834	0.060	0.949	1.416
6	0.755	-0.028	0.939	1.314
7	0.750	-0.147	0.926	1.819
8	0.886	0.100	0.954	1.371
MLP prediction 5T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.850	-0.075	0.948	1.374
2	0.881	0.055	0.958	1.180
3	0.817	0.058	0.944	1.253
4	0.842	-0.018	0.946	1.278
5	0.851	0.004	0.951	1.402
6	0.763	-0.057	0.941	1.302
7	0.738	-0.182	0.930	1.768
8	0.873	0.089	0.955	1.349
MLP prediction 7T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.845	-0.060	0.947	1.379
2	0.881	0.048	0.958	1.179
3	0.832	0.065	0.945	1.246
4	0.844	-0.003	0.947	1.272
5	0.835	0.079	0.949	1.425
6	0.758	-0.069	0.940	1.312
7	0.752	-0.074	0.931	1.760
8	0.893	0.080	0.956	1.350

Table 5.12: Wind speed prediction results obtained by the SVMr.

SVM prediction 3T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.865	-0.052	0.948	1.378
2	0.886	0.024	0.959	1.168
3	0.799	0.041	0.942	1.263
4	0.804	-0.002	0.946	1.272
5	0.884	0.031	0.951	1.411
6	0.769	0.005	0.943	1.279
7	0.785	-0.120	0.932	1.769
8	0.898	0.105	0.956	1.349
SVM prediction 5T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.850	-0.054	0.948	1.377
2	0.862	0.057	0.958	1.173
3	0.792	0.051	0.941	1.270
4	0.867	-0.067	0.947	1.279
5	0.887	0.018	0.951	1.417
6	0.764	0.009	0.942	1.295
7	0.764	-0.247	0.931	1.771
8	0.885	0.098	0.955	1.353
SVM prediction 7T				
Tower	R^2	MBE (m/s)	IoA	RMSE (m/s)
1	0.852	-0.065	0.945	1.407
2	0.881	0.053	0.958	1.175
3	0.854	-0.021	0.943	1.272
4	0.858	-0.091	0.946	1.289
5	0.884	0.017	0.950	1.437
6	0.775	0.006	0.942	1.296
7	0.730	-0.162	0.926	1.809
8	0.904	0.122	0.953	1.390

Table 5.13: Computation time of wind speed prediction (using data from 7 towers) in each considered tower, in the example with complete wind speed samples (in seconds).

Tower	ELM	GMDH	MLP	SVMr	MLR (Reference)
T1	0.030	0.992	17.254	501.4	0.024
T2	0.033	1.091	17.973	507.9	0.025
T3	0.030	0.699	18.597	505.1	0.023
T4	0.038	0.783	17.366	499.3	0.027
T5	0.029	0.647	17.426	517.9	0.025
T6	0.033	0.546	18.007	499.6	0.024
T7	0.029	0.981	17.884	534.4	0.027
T8	0.030	1.127	17.677	525.9	0.028

Chapter 6

Heuristic Correction of Wind Speed Mesoscale Models Simulations

6.1 Introduction

Wind farms prospection is a complex process that involves different previous studies to be accomplished. These prospection studies include analysis and estimation of wind resource in the area, evaluation of possible problems and cost in wind turbines installation and wind farm exploitation or several environmental and impact studies [85]-[165]. The important economical investments involved in this process make extremely important, that previous analyses that ensure the optimal location of the wind farm [2, 92] are carried out. These analyses usually start with a wind speed modeling in the study area, from different points of view: first, wind speed trends are to be assessed, to check if they are maintained (or even increased) over the years. Second, it is important to know the geographical wind speed distribution in the study area. Finally, it is needed a complete study of surface wind speeds in the study area in order to carry out an effective micrositing of the wind turbines, allowing tto obtain the maximum profit from the wind farm. Usually this prospective wind farm process starts with the installation of a set of measuring towers in the zone, in order to collect wind speed data during a period of time, long enough to perform the different studies involved in the process.

In this Chapter we tackle the geographical distribution of surface wind speed, using mesoscale models. Previous works have been proposed to improve wind speed fields analysis, in the frame of meteorological analysis of the area (Chapter 2, Section 2.3). The idea, however, can be extrapolated to wind farms prospection in a direct way, and that is the objective of our work. Thus, we propose the statistical correction of mesoscale models to estimate the geographical distribution of surface wind speeds for wind farms prospection. We introduce several novelties in comparison to the previous works in the literature: first, instead of managing complete wind speed series in each point of the grid, we deal with probability distributions: we consider the parameters of a Weibull distribution in each point of the grid (representing the wind speed at that point), obtained from the mesoscale model. We also consider that the measured wind speed follows a Weibull distribution, so the Weibull parameters of the measuring stations are used to modify the mesoscale models. An heuristic approach is proposed, that carries out a rescaling of the Weibull parameters at each grid point, depending on the distance to the different measuring towers in the wind farm. We also show the performance of an Evolutionary Strategy to select the best parameters for the heuristic search. We discuss the performance of the proposed approach by means of different experiments in two wind farms prospection sites in

Spain, where several measuring towers are available to show the appropriateness of our approach.

The rest of the Chapter is structured in the following way: next section presents the problem definition, including the notation used and the objective functions considered for each Weibull parameter. Section 6.3 discusses the different heuristics proposed in this work for mesoscale models correction. Section 6.4 shows the performance of the heuristics in grid data obtained from measuring towers located in two different wind farms in Spain, one with an important number of measuring towers installed, and another with a reduced number of towers available. Comparative results are discussed in both cases. Section 6.5 closes the chapter by giving some final conclusions and remarks.

6.2 Problem definition

Let us consider a grid $M \times N$ in a prospective area under study to install a wind farm. In each node of the grid we consider a wind speed series, obtained from a given mesoscale physical model Ξ. Since the treatment of large wind speed series in each grid node is complicated, we make the assumption that each wind speed series follows a Weibull probability distribution:

$$f(v; A, k) = \frac{k}{A}\left(\frac{v}{A}\right)^{k-1} e^{-(\frac{v}{A})^k} \tag{6.1}$$

where v is the wind speed value (variable), A is the scale parameter of the distribution and k is its shape parameter. So each point in the grid keeps a value of A and k to model the complete wind speed series.

Let $\mathcal{G}$ be an $M \times N$ grid with the wind speed measures from the mesoscale model. The wind speed series in the grid can be represented by using two matrices: $\mathcal{G}_A$ with values of the Weibull distribution of each grid node (parameter A), and another matrix $\mathcal{G}_k$ with values of parameter k. In addition, let $\mathcal{T} = \{t_i\}, i = 1, \ldots, K$, be the set of K measuring towers installed on the studied area, which gives a set of real wind speed measures. We also represent the wind speed in each measuring tower t using the Weibull distribution, i.e., a value t_{A_i} for the parameter A and a value t_{k_i} for the parameter k.

The aim of the problem is to obtain two different $M \times N$ matrices, $\mathcal{S}_A$ and $\mathcal{S}_k$ that contain modified values of the mesoscale model. The modification is carried out by using information of the real wind speed values of the measuring towers, in such a way that minimizes a given error function, (e_A or e_k, depending on the Weibull parameter to be modified), defined by the following equations:

$$e_A^{\mathcal{S}} = \frac{1}{K^*}\sum_{i=1}^{K^*} |A(t_i) - \mathcal{S}_A(t_i)| \tag{6.2}$$

$$e_k^{\mathcal{S}} = \frac{1}{K^*}\sum_{i=1}^{K^*} |k(t_i) - \mathcal{S}_k(t_i)| \tag{6.3}$$

where $A(t_i)$ and $k(t_i)$ stand for the real value of the Weibull parameters that represent wind speed series measured at tower t_i (value of A or k, respectively), K stands for the total number of towers in the wind farm and K^* stands for the number of towers selected to train or test the results. $\mathcal{S}_A(t_i)$ and $\mathcal{S}_k(t_i)$ stand for the value of the modified mesoscale model wind speed series Weibull parameter at the point where tower i is installed. Note that if the modification process

of the mesoscale model is done correctly, the values of $e_A^{\mathcal{S}}$ and $e_k^{\mathcal{S}}$ should be better than the values of the original mesoscale model $e_A^{\mathcal{G}}$ and $e_k^{\mathcal{G}}$ (non-corrected values), defined as:

$$e_A^{\mathcal{G}} = \frac{1}{K^*} \sum_{i=1}^{K^*} |A(t_i) - \mathcal{G}_A(t_i)| \tag{6.4}$$

$$e_k^{\mathcal{G}} = \frac{1}{K^*} \sum_{i=1}^{K^*} |k(t_i) - \mathcal{G}_k(t_i)| \tag{6.5}$$

Figure 6.1 shows an example of a distribution of measures. Grid points with data form the mesoscale model are represented by red crosses. The location of measuring towers are represented by blue circles. Note that the measuring towers do not coincide with any point in the mesoscale model grid, so at this point we cannot calculate Equations (6.4-6.5). A transformation from discrete to continuous values is therefore needed, in order to extend the values of the mesoscale model to the points where the measuring towers are located. There are different possibilities to do this transformation from discrete to continuous values, and the study in this work is general, so any procedure which does the transformation can be applied. In this work we use a surface fitting procedure able to construct accurate 2D surface models from scattered (discrete) data, by means of the software tool available for MatLab, *gridfit* [62].

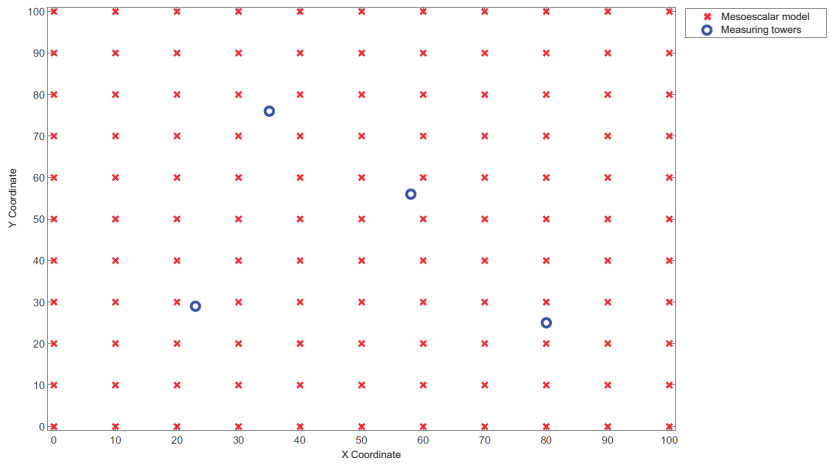


Figure 6.1: Example of mesoscale grid and measuring towers situation in a wind farm.

6.3 Proposed heuristics for mesoscale models correction

The correction of the mesoscale model with measured data can be done in different ways. First, we propose a constructive heuristics to carry out this modification. Second, we introduce an Evolutionary Strategy in order to refine the search for the best modification of A and k surfaces in terms of the values of the measuring towers.

6.3.1 Constructive heuristic

Intuitively, we can think of modifying the output values of the mesoscale model by applying some function that determines the influence of the real value measured by the towers in each point on the mesoscale grid. Note that this function depends on the distance to the tower, so the closer a point of the mesoscale model is to a measuring tower, the more similar both values should be. This procedure can be iteratively done, by modifying each value of the mesoscale model to take into account the influence of all the towers installed in the wind farm. Using the definitions above, the steps of the modification procedure are defined in Algorithm 1. In this algorithm w_h is a parameter that controls the correction level of each measuring tower h. This parameter is introduced for the sake of completeness: if all the stations are considered as equally important, then $w_h = 1, \quad \forall h$, as in our case. Note that the heuristic construction is easy: in the first step of the proposed heuristic, each point of the mesoescalar model output is modified by a factor, which depends on distance to tower #1 ($w_h\,(t_h - s_{ij}) \cdot f(d)$). This value is modified again by a second factor, which depends on the distance to tower #2, and so on. The process is repeated in a loop fashion until tower $\#K^*$ is reached. Note also that all the points in the grid are modified in this way. The procedure is carried out for modifying both $\mathcal{G}_A$ and $\mathcal{G}_k$ grid values, that characterized the wind speed series from the mesoscale model. In addition, note that there are Weibull distribution data ($\mathcal{G}_A$ and $\mathcal{G}_k$) for different sectors of a wind rose in a given wind farm. Thus, this procedure is carried out for each sector of the wind rose.

Algorithm 1 Constructive heuristic

Require: An initial grid of A or k values from a mesoscale model ($\mathcal{G}$), and set of measuring towers ($\mathcal{T}$).
Ensure: A modified grid of A or k measures of the mesoscale model, ($\mathcal{S}$).

```
for i = 1 to N do
  for j = 1 to M do
    for h = 1 to K* do
      if h = 1 then
        s_ij = g_ij
      end if
      Calculate the distance d between the point g_ij and the tower t_h.
      s_ij = s_ij + w_h (t_h − s_ij) · f(d).
    end for
  end for
end for
```

We can think of a large variety of functions that depends on distance in the desired way, i.e., the larger the distance the less the function values. A negative exponential function has been chosen in this work, following the suggestion in [123], where it was shown to be effective in mesoscale model corrections. Therefore, the correction function used is:

$$f(d) = \exp(-\alpha \cdot d) \tag{6.6}$$

The constant value α has been chosen in such a way that, when the distance of a given point in the grid to a given tower is larger than the maximum distance between the set of measuring towers, d_{max}, the value of $f(-\alpha \cdot d)$ is less than 0.01, i.e. α is obtained by solving the equation $f(-\alpha \cdot d_{max}) = 0.01$. In this case, the $f(-\alpha \cdot d)$ value is considered insignificant and therefore,

$f(-\alpha \cdot d) = 0$ if $d > d_{min}$, so the influence of towers situated larger than a distance d_{max} is not considered at any point in the grid (Figure 6.2).

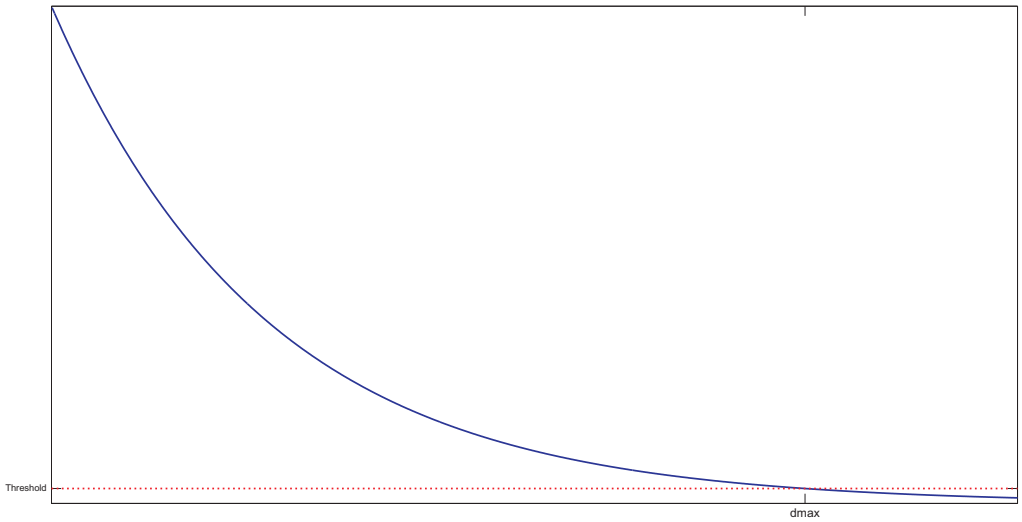


Figure 6.2: Graphic explanation of α choice.

6.3.2 An Evolutionary Strategy to set α

We propose an alternative method to set α in Equation (6.6), by means of an evolutionary strategy (ES) [9, 54]. ESs are robust meta-heuristic search algorithms based on Darwinian principles of natural evolution and survival of the fittest. Several types of ESs have been proposed [9], and the fields of application are huge, in almost all areas of science and engineering optimization. We apply this algorithm to obtain the optimal α parameter in the correction function of our approach for adapting mesoscale wind speed values to measuring stations.

The encoding of each individual in the population of the strategy is a real value $\alpha > 0$. Also, each individual has associated a fitness value, given by Equation (6.2) if we consider a modification of the Weibull parameter A, or Equation (6.3) if we consider parameter k. The modified mesoscale model is obtained by applying the constructive heuristic presented in the previous section, with the α value of each individual. Only Gaussian mutations are considered in the evolutionary strategy carried out, and the value for α is bounded between 0 and two times the maximum distance among towers ($2 \cdot d_{max}$). The evolutionary strategy procedure is the following:

1. Generate an initial population of μ individuals (solutions). Each individual is formed by a parameter α (to be found) and a matrix $\mathcal{S}$ containing the values obtained by applying the constructive heuristic with the exponential function (Equation (6.6)). Let b be a counter for the number of generations, set it to $b = 1$.

2. Evaluate the fitness value for each individual by using the problem's error equation (Equations (6.2) or (6.3), for A and k, respectively, with K^* equal to the number of towers used to train the algorithm). Before this, the surface that best fits to the $\mathcal{S}$ values needs to be calculated using the Gridfit software.

3. Generate an offspring population, of length μ, by applying a mutation operator to the individuals (Gaussian mutation is considered).

4. Evaluate the fitness of the offspring population using Equations (6.2) or (6.3).

5. Selection is based on the procedure described in [186]: Conduct pairwise comparison over the union of parents and offspring remaining: for each individual, p opponents are chosen uniformly at random from all the parents and offspring. The best individual in these p is selected to survive for the next generation. This process is repeated until a new parent generation of μ individuals is obtained.

6. Stop if the stopping criterion is satisfied. Otherwise set $b = b + 1$ and go to Step 3. In this case, the stopping criterion established is that the best solution found by the algorithm is not improved during $\mathcal{X}$ generations, or, alternatively, the algorithm reaches a maximum number of generations.

6.4 Experimental part

In order to show the performance of the proposed algorithms a large amount of experiments have been analyzed. Experiments in two locations where there are currently installed two wind farms in Spain are carried out. This ensures the presence of measuring towers which provide real wind speed data in the area under study. Data from the Weather Research and Forecasting mesoscale model [184] are considered in the two locations, obtaining a Weibull modeling of the wind in a grid defined in both wind farms.

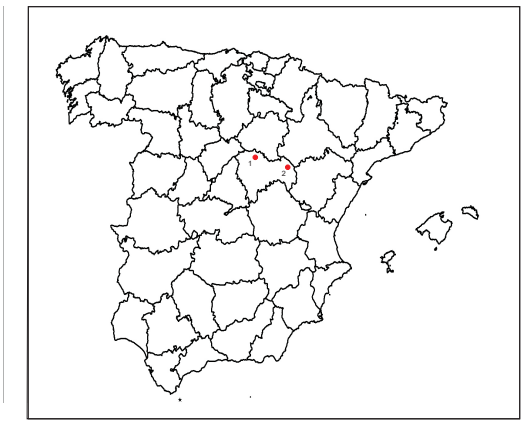
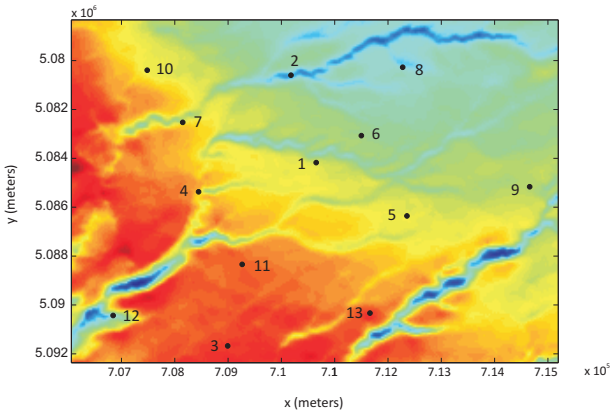


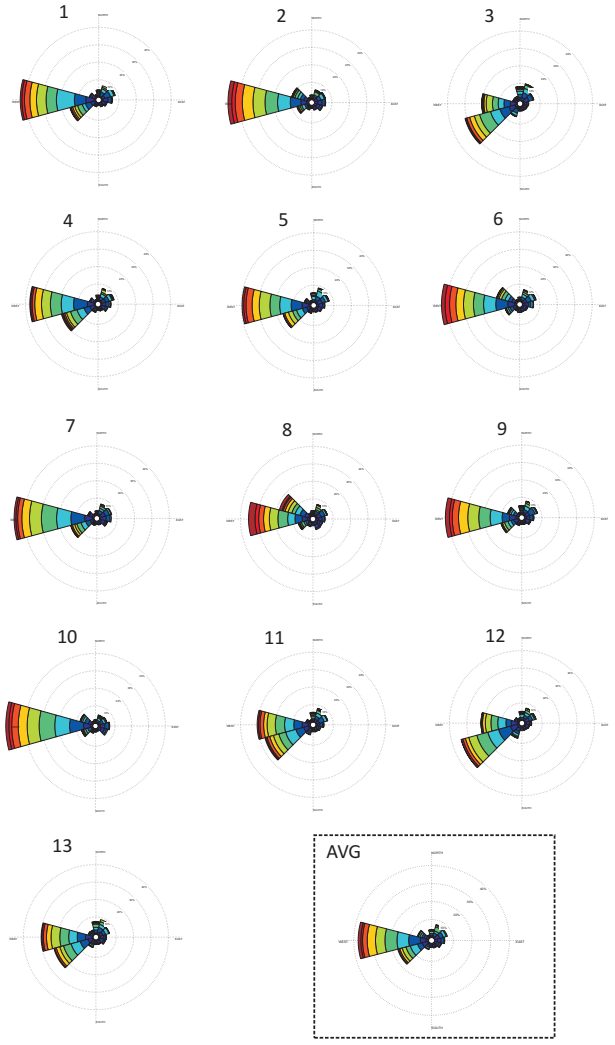
Figure 6.3: Location of the two wind farms under study in the center of Spain.

6.4.1 Results in wind farm # 1

The first wind farm area considered is located in the northern part of Guadalajara province, in Spain. There are 13 measuring towers available in this wind farm ($K = 13$) and a wind rose of 12 sectors is considered. Figure 6.4 shows the location of all the measuring towers in this case, and the wind rose obtained from data collected by the measuring tower #8.



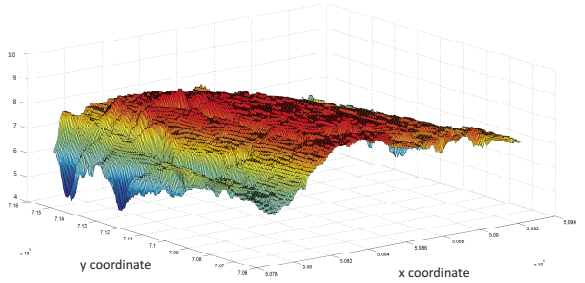
(a)



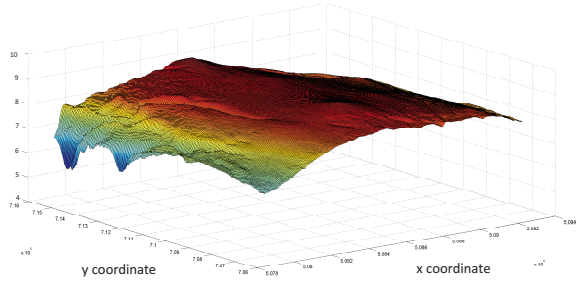
(b)

Figure 6.4: Location of measuring towers and average wind rose for wind farm #1; (a) Location of measuring towers. Colors stand for orography of the area under study, red colors indicate higher zones than blue colors; (b) Average wind rose.

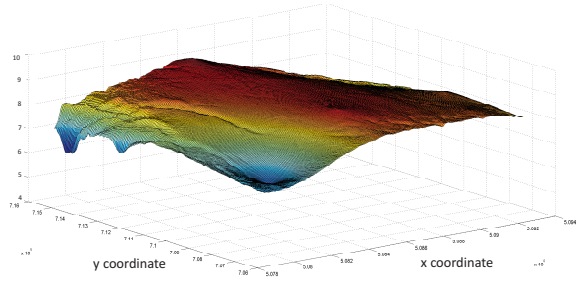
In order to evaluate the performance of the proposed algorithms, a cross-validation procedure has been carried out. The idea is to use a number of wind towers to obtain the mesoscale model output correction, and then a reduced test set in order to compare the algorithms' performance. In this case, a cross-validation procedure that uses 10 towers for training and 3 towers for test is carried out ($K^* = 10$ in train and $K^* = 3$ in test). This process has been carried out a number of times (10 permutations have been considered), in which we have used different towers in the training and test sets).



(a)



(b)

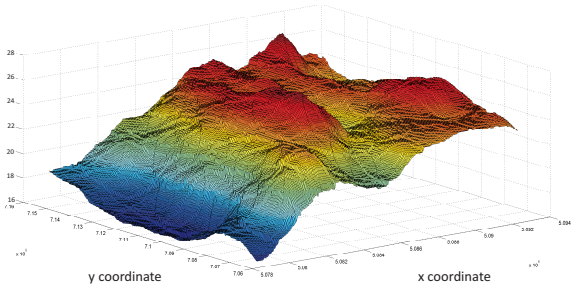


(c)

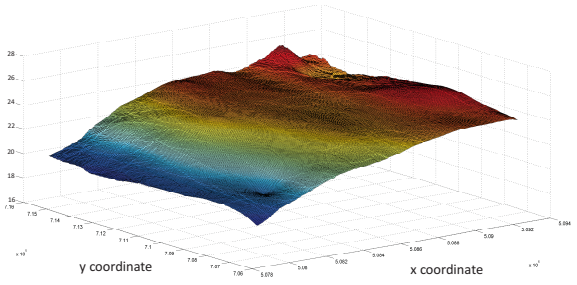
Figure 6.5: Surface of Weibull parameter A obtained from the uncorrected data of the mesoscale model output(a), the heuristic approach (b) and the ES considered (c). x and y coordinates are expressed in UTM coordinates (meters).

Tables 6.1-6.6 show the results obtained comparing the proposed heuristic and ES. These tables show the fitness values calculated following Equation (6.2) or (6.3), that can be interpreted as a average mean error in the test set. The values of $e_A^{\mathcal{G}}$ and $e_k^{\mathcal{G}}$ (average absolute error between the mesocale model without correction and the measurements at the K^* towers) are also shown

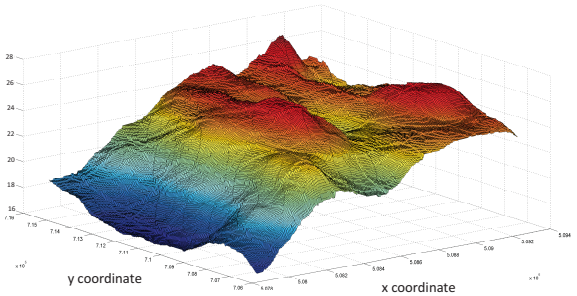
in these tables, for reference. It is possible to see how the statistical correction carried out with the heuristic algorithms proposed improves the mesoscale model value (in terms of error in the test measuring towers considered). Depending on the parameter considered, sector and tower, the results obtained by the heuristic are better than those by the ES, or vice-versa: there is not a clear pattern of outperforming in this case. Note that the ES chooses the best possible α in the training phase, however, this does not mean that the correction made using that α value is the best when new data (test) are considered. The average results in the 10 permutations taken into account (lower part of Tables 6.1-6.6) confirm that the heuristic and ES corrections of the mesoscale model improves its performance in terms of error with the available in-situ measurements.



(a)



(b)



(c)

Figure 6.6: Surface of Weibull parameter k obtained from the uncorrected data of the mesoscale model output(a), the heuristic approach (b) and the ES considered (c). x and y coordinates are expressed in UTM coordinates (meters).

The adjusted surfaces obtained after the application of the algorithms proposed in this work can be also represented, and compared to the surface of the mesoscale model data (without correction). This gives an idea of how the heuristic and ES correction affect the values of parameters A and k. We have chosen a specific case (parameter A, sector 1, permutation 1). The uncorrected data from the mesoscale model for this case are shown in Figure 6.5 (a). On the other hand, Figures 6.5 (b) and 6.5 (c) show the resulting surfaces obtained after applying the heuristic and ES methods, respectively. Note the differences obtained when the correction procedures are applied, and how the surface is smoother after the application of the heuristic corrections. Figures 6.6 (a), (b) and (c) show the uncorrected mesoscale model values, heuristic and ES corrections, respectively, for parameter k. In these figures it can also be seen how the heuristic correction is quite important, while the ES correction is not so deep.

6.4.2 Results in wind farm # 2

The second wind farm area studied in this work is located in the Eastern part of the Guadalajara province, Spain. In this wind farm, there are only 6 measuring towers available (represented in Figure 6.7 (a), $K = 6$). The wind rose in this case has been measured with 16 sectors (Figure 6.7 (b) shows the wind rose obtained from data of tower #1). In this case, 4 towers have been used to train the algorithms, and 2 for test purposes ($K^* = 4$ in train and $K^* = 2$ in test). Again 10 different permutations of the training and test towers are considered. Results are presented in Tables 6.7-6.14 for Weibull parameters A and k. It is easy to see that in this scenario, the ES consistently obtains better results than the heuristic approach tested. Again, both proposed approaches improve the performance of the mesoscale model without correction, in terms of error in test. These results show that, in the case of scarce information available to carry out the correction (low number of measuring towers), the optimal selection of the α parameter through the ES procedure provides better results that the heuristic choice of this parameter.

Figures 6.8 (a), (b) and (c) show the parameter A surfaces for the mesoscale model without correction, heuristic correction and ES, respectively. Figures 6.9 (a), (b) and (c) show their counterpart for parameter k. These figures show the effect of the proposed corrections, and how in this case the corrections by the heuristic and the ES are small.

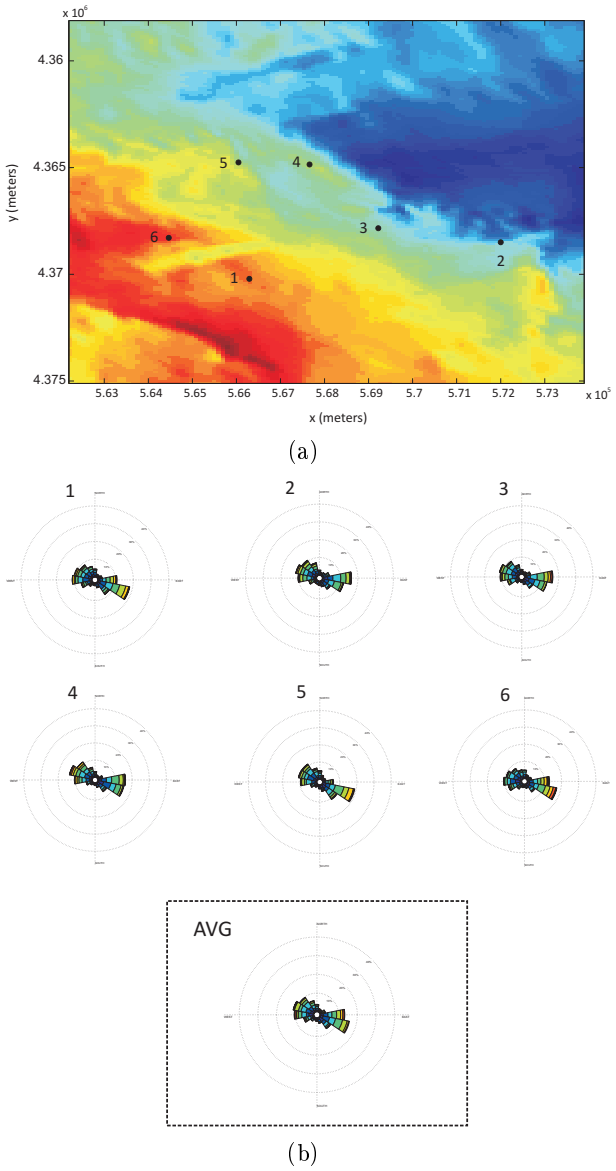


Figure 6.7: Location of measuring towers and average wind rose for wind farm #2; (a) Location of measuring towers. Colors stand for orography of the area under study, red colors indicate higher zones than blue colors; (b) Average wind rose.

6.5 Conclusions

In this Chapter we have discussed two different heuristic methods for the statistical correction of wind speed outputs from mesoscale models. The methods developed here have direct application in prospecting wind farms and in the process of wind farm design and turbines micrositing. In these processes, the exact knowledge of wind speed field in the area under study is extremely important, so correction of the mesoscale models output is a must. In this study we have tackled the correction of the model with probability distributions, by considering the parameters of a Weibull distribution (obtained from the mesoscale model) in each point of a grid. The statistical

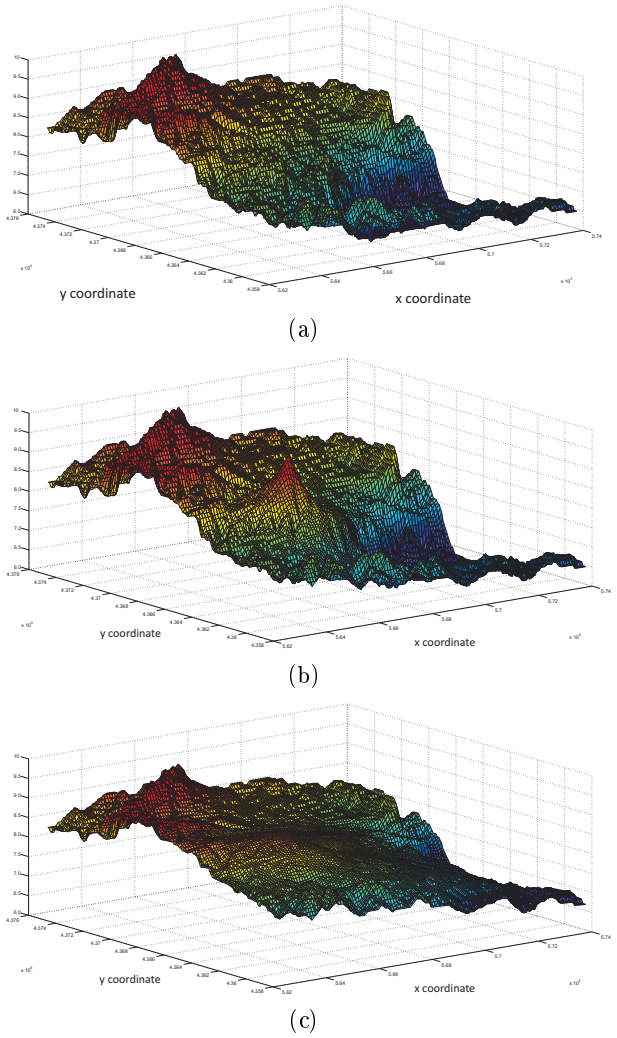


Figure 6.8: Surface of Weibull parameter A obtained from the uncorrected data of the mesoscale model output(a), the heuristic approach (b) and the ES considered (c). x and y coordinates are expressed in UTM coordinates (meters).

correction has been carried out in terms of similarity of these Weibull parameters to some measuring stations. This technique allows using a surface fitting algorithm to obtain the goodness of mesoscale model parameters' correction. The effectiveness of the proposed heuristic approaches in the correction of wind speeds from mesoscale models has been successfully tested in two real wind farm areas in the center of Spain, where measuring towers are installed, and real wind speed measurements are available.

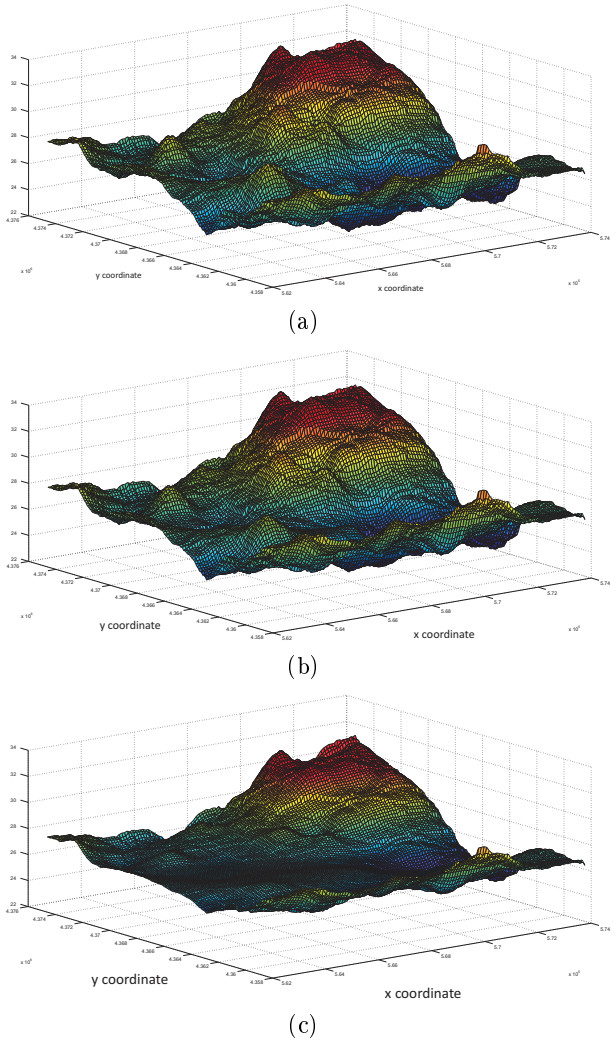


Figure 6.9: Surface of Weibull parameter k obtained from the uncorrected data of the mesoscale model output(a), the heuristic approach (b) and the ES considered (c). x and y coordinates are expressed in UTM coordinates (meters).

Appendix for this chapter

6.A Tables

Note that for every table $e_A^{\mathcal{G}}$ stands for the value of the error in test for the uncorrected mesoscale model. Avg stands for the average error of the 10 permutations considered.

Table 6.1: Wind farm #1. Correction of Weibull parameter A (Sectors 1 to 4) in terms of average absolute error in the test set ($e_A^{\mathcal{S}}$, $K^* = 3$ towers), in m/s for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 1			Sector 2			Sector 3			Sector 4		
	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES
1	0.83	0.52	0.54	2.02	0.63	0.67	0.65	0.34	0.35	0.33	0.11	0.09
2	1.10	0.69	0.64	2.12	0.64	0.51	1.35	0.71	0.74	0.53	0.28	0.32
3	0.64	0.43	0.43	1.43	0.24	0.24	0.83	0.24	0.28	0.42	0.49	0.48
4	0.87	0.54	0.56	1.64	0.35	0.45	0.42	0.40	0.47	0.39	0.17	0.19
5	0.90	0.41	0.37	1.59	0.36	0.43	0.63	0.37	0.36	0.44	0.34	0.41
6	0.46	0.43	0.45	1.27	0.46	0.52	0.69	0.32	0.34	0.26	0.37	0.31
7	1.22	0.87	0.86	2.07	0.34	0.25	0.73	0.25	0.24	0.47	0.22	0.31
8	0.81	0.35	0.35	1.94	0.47	0.38	0.89	0.46	0.42	0.17	0.17	0.10
9	0.41	0.38	0.38	1.57	0.47	0.43	0.35	0.38	0.45	0.19	0.43	0.37
10	0.69	0.63	0.63	1.79	0.48	0.30	0.56	0.12	0.08	0.35	0.31	0.20
Avg	0.79	0.52	0.52	1.74	0.44	0.42	0.71	0.36	0.37	0.35	0.29	0.28

Table 6.2: Wind farm #1. Correction of Weibull parameter A (Sectors 5 to 8) in terms of average absolute error in the test set ($e_A^{\mathcal{S}}$, $K^* = 3$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 5			Sector 6			Sector 7			Sector 8		
	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES
1	0.47	0.31	0.33	0.30	0.18	0.21	0.45	0.44	0.44	0.67	0.35	0.32
2	0.28	0.13	0.15	0.32	0.25	0.21	0.45	0.46	0.44	1.65	0.90	0.93
3	0.23	0.28	0.25	0.27	0.06	0.07	0.25	0.13	0.21	1.11	0.18	0.07
4	0.33	0.04	0.11	0.37	0.16	0.22	0.27	0.23	0.30	0.64	0.11	0.10
5	0.35	0.17	0.24	0.16	0.08	0.08	0.32	0.13	0.19	0.54	0.38	0.47
6	0.33	0.33	0.32	0.20	0.09	0.08	0.32	0.30	0.33	1.02	0.23	0.22
7	0.22	0.09	0.10	0.24	0.17	0.17	0.26	0.24	0.22	1.07	0.24	0.24
8	0.26	0.15	0.13	0.33	0.20	0.20	0.40	0.34	0.42	1.41	0.55	0.67
9	0.23	0.15	0.14	0.29	0.10	0.10	0.17	0.09	0.12	0.80	0.17	0.15
10	0.31	0.30	0.30	0.27	0.11	0.18	0.37	0.33	0.37	1.04	0.35	0.34
Avg	0.30	0.19	0.21	0.28	0.14	0.15	0.33	0.27	0.30	0.99	0.35	0.35

Table 6.3: Wind farm #1. Correction of Weibull parameter A (Sectors 9 to 12) in terms of average absolute error in the test set ($e_A^{\mathcal{S}}$, $K^* = 3$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 9			Sector 10			Sector 11			Sector 12		
	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES
1	0.35	0.45	0.46	0.68	0.74	0.90	0.93	1.59	1.38	1.33	0.44	0.38
2	0.85	0.54	0.58	1.14	0.50	0.52	1.28	0.74	0.58	1.46	0.56	0.48
3	0.40	0.17	0.23	1.85	0.96	1.02	3.02	1.90	2.18	1.36	0.69	0.69
4	0.37	0.55	0.57	1.29	0.56	0.64	0.63	0.34	0.42	1.30	0.80	0.65
5	0.43	0.42	0.43	1.36	0.69	0.56	0.46	0.58	0.65	0.78	0.58	0.56
6	0.33	0.21	0.18	1.65	0.71	0.77	1.81	1.17	1.21	1.18	1.11	1.09
7	0.69	0.39	0.40	1.45	0.47	0.59	1.25	0.29	0.16	1.57	0.54	0.46
8	0.41	0.07	0.09	1.41	0.57	0.70	1.31	1.36	1.15	1.35	0.63	0.63
9	0.70	0.21	0.36	1.73	0.80	1.03	2.25	1.33	1.47	1.30	0.39	0.40
10	0.36	0.16	0.16	1.55	1.01	1.08	1.83	1.62	1.51	0.89	0.45	0.45
Avg	0.49	0.32	0.35	1.41	0.70	0.78	1.48	1.09	1.07	1.25	0.62	0.58

Table 6.4: Wind farm #1. Correction of Weibull parameter k (Sectors 1 to 4) in terms of average absolute error in the test set ($e_k^{\mathcal{S}}$, $K^* = 3$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 1			Sector 2			Sector 3			Sector 4		
	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES
1	1.78	1.30	1.71	2.22	1.68	2.03	1.05	1.25	1.06	1.03	0.69	0.95
2	1.53	1.68	1.70	2.69	2.24	2.68	1.26	1.55	1.27	2.39	1.83	2.28
3	0.71	1.92	1.17	1.53	3.05	3.73	1.93	2.69	2.24	2.31	1.84	2.10
4	2.94	2.15	2.64	2.36	1.44	2.18	0.78	1.51	0.79	1.22	1.14	1.00
5	2.88	1.94	2.71	3.19	2.60	3.21	1.04	0.78	1.06	3.10	2.35	2.87
6	0.98	1.61	1.00	2.68	2.94	2.65	1.77	2.12	1.76	1.39	1.71	1.58
7	1.10	0.91	1.11	1.54	1.27	1.48	0.89	0.63	0.91	2.20	1.12	1.94
8	0.75	0.67	0.71	1.43	2.12	1.51	1.18	1.78	1.22	1.93	1.77	1.95
9	0.63	1.79	1.08	1.65	3.53	4.16	1.58	2.28	1.99	0.96	1.16	0.82
10	0.94	0.86	0.94	0.74	2.39	2.48	1.67	2.40	1.98	2.36	1.87	2.05
Avg	1.42	1.48	1.48	2.00	2.33	2.61	1.31	1.70	1.43	1.89	1.55	1.75

Table 6.5: Wind farm #1. Correction of Weibull parameter k (Sectors 5 to 8) in terms of average absolute error in the test set ($e_k^{\mathcal{S}}$, $K^* = 3$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 5			Sector 6			Sector 7			Sector 8		
	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES
1	0.90	1.50	1.07	2.24	2.09	2.27	1.23	1.40	1.21	0.42	0.53	0.45
2	2.27	1.42	2.05	0.88	0.80	0.88	0.76	1.34	0.88	1.15	1.55	1.19
3	2.03	1.61	2.01	0.50	0.35	0.36	1.61	1.57	1.66	1.00	1.32	1.05
4	0.69	0.90	0.33	2.45	1.81	1.90	1.36	0.82	1.28	0.70	1.18	0.70
5	2.38	1.39	2.06	1.61	0.68	1.55	1.30	1.05	1.31	1.71	1.77	1.71
6	0.32	1.32	0.38	0.99	0.36	0.75	1.97	1.66	2.00	1.62	1.91	1.66
7	0.69	0.65	0.56	1.34	1.84	1.76	0.24	1.15	0.47	1.12	1.06	1.12
8	1.63	2.02	1.47	1.14	0.51	1.02	1.13	0.32	1.13	1.53	1.76	1.58
9	0.76	1.19	0.71	0.82	0.82	0.70	0.86	0.60	0.86	1.22	1.89	1.41
10	1.76	1.62	1.69	1.23	0.57	1.10	1.92	2.19	2.02	1.11	1.51	1.12
Avg	1.34	1.36	1.23	1.32	0.98	1.23	1.24	1.21	1.28	1.16	1.45	1.20

Table 6.6: Wind farm #1. Correction of Weibull parameter k (Sectors 9 to 12) in terms of average absolute error in the test set ($e_k^{\mathcal{S}}$, $K^* = 3$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 9			Sector 10			Sector 11			Sector 12		
	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES
1	1.22	1.20	1.29	1.18	0.76	1.16	1.50	1.06	1.34	1.59	1.39	1.61
2	0.71	0.92	0.68	0.75	0.64	0.75	0.84	1.20	0.83	1.55	1.46	1.54
3	1.42	1.15	1.42	0.87	1.10	0.93	1.66	1.13	1.53	1.46	1.68	1.49
4	1.82	0.84	1.73	1.70	1.07	1.53	0.45	0.30	0.45	2.53	2.40	2.61
5	1.09	0.63	0.86	1.92	1.09	1.75	0.74	0.54	0.63	3.15	2.86	3.14
6	1.35	0.94	1.30	0.53	1.04	0.58	1.55	1.13	1.52	1.55	1.66	1.58
7	0.19	0.48	0.24	0.57	0.91	0.49	0.72	0.38	0.66	0.99	0.96	1.03
8	1.22	0.88	1.05	0.34	0.75	0.34	0.72	1.20	0.75	1.84	1.65	1.77
9	0.99	0.57	0.93	0.83	0.72	0.72	0.94	0.71	0.72	0.46	0.58	0.31
10	1.38	0.92	1.34	1.17	0.87	1.15	0.99	1.09	1.05	1.39	1.03	1.25
Avg	1.14	0.85	1.08	0.99	0.90	0.94	1.01	0.87	0.95	1.65	1.57	1.63

Table 6.7: Wind farm #2. Correction of Weibull parameter A (Sectors 1 to 4) in terms of average absolute error in the test set ($e_A^{\mathcal{S}}$, $K^* = 2$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 1			Sector 2			Sector 3			Sector 4		
	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES
1	0.53	0.56	0.46	0.90	0.86	0.72	0.65	0.55	0.51	0.98	0.82	0.54
2	0.70	0.66	0.55	0.86	0.85	0.64	0.14	0.21	0.33	0.48	0.29	0.30
3	0.41	0.46	0.52	0.50	0.49	0.40	0.58	0.56	0.53	1.12	1.05	0.80
4	0.48	0.50	0.49	1.28	1.25	1.10	0.39	0.31	0.31	1.22	1.15	0.93
5	0.83	0.82	0.65	0.82	0.77	0.50	0.16	0.12	0.29	0.54	0.42	0.41
6	0.41	0.46	0.53	0.50	0.49	0.42	0.58	0.57	0.55	1.12	1.05	0.77
7	0.26	0.28	0.31	0.76	0.71	0.58	0.79	0.69	0.60	1.18	1.06	0.86
8	0.71	0.72	0.63	0.74	0.78	0.63	0.67	0.63	0.55	1.18	1.07	0.81
9	0.53	0.56	0.46	0.90	0.86	0.72	0.65	0.55	0.51	0.98	0.82	0.54
10	0.18	0.19	0.27	1.16	1.14	0.95	0.63	0.67	0.64	1.55	1.51	1.25
Avg	0.50	0.52	0.49	0.84	0.82	0.66	0.52	0.49	0.48	1.04	0.93	0.72

Table 6.8: Wind farm #2. Correction of Weibull parameter A (Sectors 5 to 8) in terms of average absolute error in the test set ($e_A^{\mathcal{S}}$, $K^* = 2$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 5			Sector 6			Sector 7			Sector 8		
	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES
1	1.13	1.09	0.77	1.11	0.97	0.27	0.24	0.22	0.16	0.20	0.18	0.30
2	1.33	1.21	0.85	0.93	0.72	0.49	0.53	0.57	0.54	0.89	0.93	0.98
3	1.40	1.42	1.24	1.22	0.94	0.57	0.09	0.16	0.29	0.61	0.65	0.67
4	0.55	0.59	0.80	1.82	1.76	1.45	0.68	0.68	0.59	0.57	0.56	0.56
5	1.46	1.45	0.92	0.99	0.86	0.43	0.47	0.47	0.53	0.52	0.49	0.38
6	1.40	1.42	1.21	1.22	0.94	0.65	0.09	0.17	0.24	0.61	0.65	0.69
7	1.34	1.28	1.06	1.52	1.32	0.97	0.65	0.66	0.69	0.67	0.66	0.77
8	1.04	1.03	0.77	1.41	1.31	0.95	0.16	0.14	0.12	0.38	0.39	0.27
9	1.13	1.09	0.77	1.11	0.97	0.27	0.24	0.22	0.16	0.20	0.18	0.30
10	0.72	0.69	0.61	1.28	1.33	1.16	0.93	0.92	0.92	0.78	0.73	0.64
Avg	1.15	1.13	0.90	1.26	1.11	0.72	0.41	0.42	0.43	0.54	0.54	0.55

Table 6.9: Wind farm #2. Correction of Weibull parameter A (Sectors 9 to 12) in terms of average absolute error in the test set ($e_A^{\mathcal{S}}$, $K^* = 2$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 9			Sector 10			Sector 11			Sector 12		
	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES
1	0.37	0.34	0.19	0.92	0.84	0.49	0.91	0.89	0.57	0.56	0.45	0.24
2	0.60	0.57	0.52	1.13	1.03	0.76	0.81	0.80	0.70	0.30	0.19	0.18
3	0.73	0.68	0.54	1.36	1.27	1.11	1.15	1.09	0.85	0.44	0.38	0.29
4	0.57	0.55	0.46	0.93	0.88	0.65	0.70	0.70	0.56	0.65	0.57	0.65
5	0.40	0.42	0.47	1.00	0.95	0.70	1.18	1.16	0.80	0.13	0.18	0.52
6	0.73	0.68	0.55	1.36	1.27	1.12	1.15	1.09	0.86	0.44	0.39	0.32
7	0.72	0.68	0.59	1.21	1.06	0.78	1.06	1.00	0.77	0.64	0.45	0.37
8	0.57	0.55	0.40	1.11	1.07	0.84	0.98	0.97	0.70	0.41	0.34	0.17
9	0.37	0.34	0.19	0.92	0.84	0.49	0.91	0.89	0.57	0.56	0.45	0.24
10	0.93	0.93	0.84	1.00	0.99	0.68	0.66	0.63	0.45	0.80	0.75	0.57
Avg	0.60	0.57	0.48	1.09	1.02	0.76	0.95	0.92	0.68	0.49	0.41	0.35

Table 6.10: Wind farm #2. Correction of Weibull parameter A (Sectors 13 to 16) in terms of average absolute error in the test set ($e_A^{\mathcal{S}}$, $K^* = 2$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 13			Sector 14			Sector 15			Sector 16		
	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES	$e_A^{\mathcal{G}}$	H	ES
1	1.62	1.54	0.99	0.66	0.68	0.60	0.83	0.82	0.71	0.50	0.56	0.60
2	1.51	1.35	0.99	0.65	0.63	0.56	0.98	0.94	0.75	0.56	0.54	0.55
3	0.99	0.74	0.59	0.18	0.17	0.16	0.37	0.36	0.23	0.30	0.30	0.41
4	1.38	1.30	1.04	0.38	0.37	0.33	0.61	0.59	0.41	0.39	0.40	0.39
5	1.61	1.52	0.82	0.61	0.60	0.35	0.83	0.80	0.37	0.51	0.50	0.43
6	0.99	0.74	0.58	0.18	0.17	0.15	0.37	0.36	0.22	0.30	0.30	0.40
7	1.42	1.23	0.90	0.21	0.16	0.07	0.38	0.29	0.12	0.14	0.16	0.16
8	1.02	1.00	0.89	0.43	0.44	0.42	0.60	0.62	0.58	0.60	0.60	0.60
9	1.62	1.54	0.99	0.66	0.68	0.60	0.83	0.82	0.71	0.50	0.56	0.60
10	1.66	1.62	1.23	0.36	0.35	0.23	0.46	0.44	0.30	0.12	0.14	0.16
Avg	1.38	1.26	0.90	0.43	0.43	0.35	0.63	0.61	0.44	0.39	0.41	0.43

Table 6.11: Wind farm #2. Correction of Weibull parameter k (Sectors 1 to 4) in terms of average absolute error in the test set ($e_k^{\mathcal{S}}$, $K^* = 2$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 1			Sector 2			Sector 3			Sector 4		
	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES
1	3.76	4.30	3.25	0.65	0.60	0.42	0.58	1.92	0.70	1.23	6.56	1.38
2	2.75	3.72	2.96	0.87	0.97	0.49	1.24	1.86	1.24	0.90	5.08	0.90
3	2.01	3.09	1.15	0.69	0.45	0.52	1.43	1.63	1.23	1.76	7.68	3.07
4	2.77	2.70	2.07	0.45	0.94	0.42	2.22	1.00	2.07	1.46	4.98	2.14
5	2.26	2.99	1.68	1.01	0.77	0.58	0.75	1.72	0.79	1.03	6.60	0.89
6	2.01	3.08	1.19	0.69	0.44	0.46	1.43	1.63	1.25	1.76	7.67	2.92
7	2.36	3.60	1.67	0.98	0.90	0.82	0.56	1.78	1.13	0.92	5.29	0.87
8	2.59	2.62	1.76	0.95	0.27	0.75	1.60	1.68	1.40	2.35	7.27	3.83
9	3.76	4.30	3.25	0.65	0.60	0.42	0.58	1.92	0.70	1.23	6.56	1.38
10	2.39	2.40	1.82	0.67	0.62	0.52	1.79	1.81	1.63	2.65	2.75	2.73
Avg	2.67	3.28	2.08	0.76	0.66	0.54	1.22	1.69	1.21	1.53	6.04	2.01

Table 6.12: Wind farm #2. Correction of Weibull parameter k (Sectors 5 to 8) in terms of average absolute error in the test set ($e_k^{\mathcal{S}}$, $K^* = 2$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 5			Sector 6			Sector 7			Sector 8		
	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES
1	0.91	1.85	0.83	0.50	2.02	0.55	0.53	2.48	0.50	2.44	3.28	2.48
2	1.19	0.79	0.87	0.22	0.38	0.26	1.25	2.45	1.34	1.65	1.84	1.39
3	1.69	1.36	1.35	1.75	1.01	1.12	1.83	2.06	1.75	1.77	2.79	1.73
4	1.47	1.51	1.26	0.56	0.73	0.65	1.35	1.39	1.39	1.07	2.10	1.00
5	1.69	1.17	1.03	1.97	1.31	1.84	1.28	2.05	1.25	2.90	2.89	3.19
6	1.69	1.36	1.36	1.75	1.01	1.18	1.83	2.07	1.82	1.77	2.80	1.79
7	0.89	0.58	0.46	0.42	0.52	0.24	0.76	1.07	0.86	1.78	2.22	1.49
8	1.82	1.81	1.65	1.62	0.98	1.27	1.50	2.23	1.56	2.02	2.26	2.14
9	0.91	1.85	0.83	0.50	2.02	0.55	0.53	2.48	0.50	2.44	3.28	2.48
10	1.15	1.12	0.92	0.68	0.75	1.00	1.20	1.19	1.04	0.89	0.84	0.61
Avg	1.34	1.34	1.06	1.00	1.07	0.87	1.21	1.95	1.20	1.87	2.43	1.83

Table 6.13: Wind farm #2. Correction of Weibull parameter k (Sectors 9 to 12) in terms of average absolute error in the test set ($e_k^{\mathcal{S}}$, $K^* = 2$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 9			Sector 10			Sector 11			Sector 12		
	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES
1	1.92	3.21	0.98	3.54	4.91	2.42	1.10	5.20	0.78	1.56	2.94	1.29
2	1.91	2.89	1.55	2.99	3.53	1.80	1.63	4.77	1.68	1.94	2.12	1.71
3	2.46	4.66	1.80	3.53	3.63	2.77	2.62	4.66	2.52	1.82	2.02	1.70
4	2.20	2.93	1.89	4.63	4.62	4.03	1.32	3.14	1.95	0.72	1.36	0.89
5	2.03	2.80	0.91	4.96	4.32	4.21	2.27	4.35	1.64	2.17	2.93	2.04
6	2.46	4.65	1.75	3.53	3.66	3.21	2.62	4.65	2.41	1.82	2.02	1.72
7	1.96	3.44	0.99	3.14	2.97	2.15	1.86	2.72	1.83	1.61	1.57	1.37
8	2.44	4.37	1.71	4.43	4.41	4.26	1.90	4.96	1.61	1.05	1.91	0.84
9	1.92	3.21	0.99	3.54	4.91	2.42	1.10	5.20	0.78	1.56	2.94	1.29
10	2.33	2.26	1.89	3.66	3.71	3.49	0.62	0.60	0.70	1.23	1.18	0.88
Avg	2.16	3.44	1.45	3.79	4.07	3.08	1.70	4.02	1.59	1.55	2.10	1.37

Table 6.14: Wind farm #2. Correction of Weibull parameter k (Sectors 12 to 16) in terms of average absolute error in the test set ($e_k^{\mathcal{S}}$, $K^* = 2$ towers), in m/s, for the proposed heuristic (H) and the Evolutionary Strategy (ES).

# P	Sector 13			Sector 14			Sector 15			Sector 16		
	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES	$e_k^{\mathcal{G}}$	H	ES
1	0.35	0.81	0.15	1.02	1.66	1.01	0.30	1.28	0.50	3.59	4.20	1.79
2	0.97	1.54	0.68	1.84	2.10	1.61	1.69	1.99	1.69	5.16	5.16	3.64
3	0.81	2.04	0.82	0.99	1.74	0.49	1.48	1.98	1.10	3.85	3.82	2.33
4	0.53	1.55	0.64	1.43	1.42	1.19	0.86	0.89	0.89	4.53	4.44	3.24
5	0.87	0.73	0.45	1.45	1.47	0.91	1.45	0.62	0.42	3.98	3.17	1.92
6	0.81	2.05	0.75	0.99	1.74	0.55	1.48	1.99	1.02	3.85	3.81	2.21
7	0.75	1.34	0.71	0.75	1.21	0.60	1.12	1.19	0.54	4.14	3.81	2.87
8	0.69	1.14	1.01	1.12	1.56	0.61	0.97	0.97	0.51	3.13	3.03	1.50
9	0.35	0.81	0.15	1.02	1.66	1.01	0.30	1.28	0.50	3.59	4.20	1.79
10	0.72	0.69	0.65	1.07	1.07	0.93	0.91	0.87	0.67	3.58	3.56	2.79
Avg	0.69	1.27	0.60	1.17	1.56	0.89	1.06	1.31	0.78	3.94	3.92	2.41

Chapter 7

On-shore wind farm design with evolutionary algorithms

7.1 Introduction

A problem directly related to wind energy production and the improvement of wind farms efficiency is the optimal design of wind farms taking into account wind speed data in the area, costs and expected profits. This wind farm design problem is crucial for companies in the energy sector, since it is expected the construction of hundreds of wind farm facilities in Europe, United States and Middle East in the next few years. Therefore, tackling wind farm design is extremely interesting from an economic point of view. That is why automatic wind farm design is a topic gaining popularity among wind farm designers and engineers in the last years. Wind farm design has opened a huge new line of research. There is an increasing number of articles tackling this problem, successfully applying computational intelligence techniques, mainly evolutionary algorithms, though other approaches have also been used, as it has been reported in Chapter 2.

Several novelties are proposed in this Thesis in order to make the problem closer to reality: a wind farm shape model, an orography model and the inclusion of benefit/cost terms in the objective function are the main new points included in this work. In addition, a novel evolutionary algorithm is presented, initially seeded with the solution from a greedy heuristic for the optimal solution of the problem.

The rest of the Chapter is structured as follows: next section summarizes the main previous approaches on wake and cost models used in the literature. Section 7.3 presents the main novelties included in the optimization model. Section 7.4 presents the greedy-constructive heuristic proposed, which will be used to seed the evolutionary algorithm presented in Section 7.5. The experimental part of the study is shown in Section 7.6, and to sum up, some conclusions are presented at the end of the chapter.

7.2 Background: turbines' wake and most used cost models in the literature

The first wake model (and maybe the most common one) was proposed by Mosetti et al. [119]. It was applied later in many following works such as [48, 61, 98, 108]. Though it is simple (and therefore somehow far away from a realistic wake) it has been profusely used, since it can be complicated with extra constraints to make it closer to reality, and it can be used to compare

different algorithms in the same conditions. Figure 7.1 shows and schematic of the wake model considered. This model has been simplified by applying the continuity equation in the control volume in Figure 7.1:

$$\rho u_0 A_0 = \rho u_1 A_1 = \rho u_i A_i \tag{7.1}$$

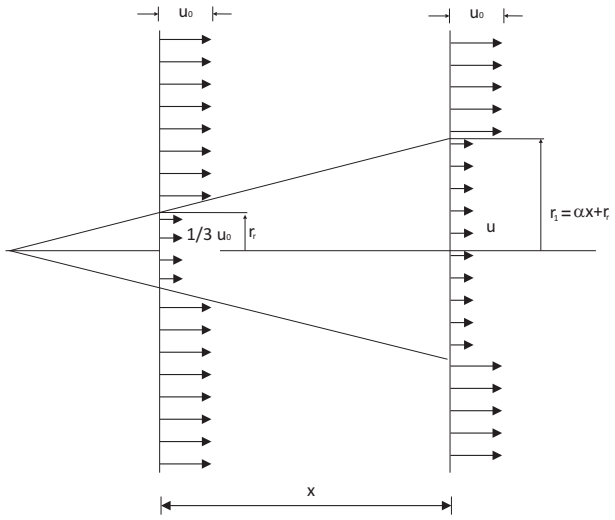


Figure 7.1: Schematic of Mosetti's wake model.

So assuming that wind speed will reduce in a units its speed after passing through a turbine:

$$\rho(au_0)A_r + \rho u(A_1 - A_r) = \rho u A_r \tag{7.2}$$

where $A_1 = \pi r_1^2$, $A_r = \pi r_r^2$ and $r_1 = \alpha x + r_r$. Substituting in Equation (7.2):

$$u = u_0 \left[1 - \left((2a)/\left[1 + \alpha(x/r_1)\right]^2\right)\right] \tag{7.3}$$

where u_0 is the mean wind speed, a is the axial induction factor, x is the distance downstream from the turbine, r_r is the downstream rotor radius and α is the entrainment constant. In addition, r_1 and and the turbine coefficient C_T can be calculated from r_r and a, through the so-called $Betz$ equations:

$$r_1 = r_r \sqrt{((1-a)/(1-2a))}, \tag{7.4}$$

$$C_T = 4a(1-a) \tag{7.5}$$

Finally, the entrainment constant α can be empirically calculated as:

$$\alpha = (0.5)/(ln(z/z_0)), \tag{7.6}$$

where z is the hub height of the wind turbine, and z_0 is the surface roughness.

Using these equations, and assuming that the kinetic energy deficit of a mixed wake is equal to the sum of the energy deficits, the resulting wind speed downstream of N turbines can be calculated as follows:

$$\left(1-\frac{\overline{u}}{u_0}\right)^2=\sum_{i=1}^{N}\left(1-\frac{u_i}{u_0}\right)^2 \tag{7.7}$$

The power equation given in [48, 61, 119] is the following:

$$P_{total}=\sum_{i=1}^{N}0.3u_i \tag{7.8}$$

Regarding the cost modeling, in [61, 119] is assumed that the non-dimensionalized cost/year of a single turbine is 1, and a reduction in the cost of each turbine when a large number are installed, the total cost/year for the entire wind farm is:

$$cost=N\left(\frac{2}{3}+\frac{1}{3}e^{-0.00174N^2}\right) \tag{7.9}$$

So the genetic algorithms presented in [61, 119] use as objective function the following:

$$g=\frac{cost}{P_{total}}. \tag{7.10}$$

In [48], using the same wake and cost modeling, a different objective function is considered:

$$g=w_1 cost_m+w_2\frac{1}{P_{total}} \tag{7.11}$$

$$w_1+w_2=1 \tag{7.12}$$

where $cost_m$ is the per unit value of cost/year of the whole wind farm. Equation 7.11 not only optimizes the placement of wind turbines, but also has control on cost.

In [144] a novel cost model based on profitability of investments in the wind farm was presented. Basically, this model is based on the following objective function to be maximized:

$$NPV(x,i,t)=\frac{N_1(x)}{i+1}+\ldots+\frac{N_t(x)}{(i+1)^t}+IC(x), \tag{7.13}$$

where IC is the initial capital investment, N_k stands for the net cash flow of the kth year, i is the discount rate (capital cost), t is the number of years spanned by the investment and finally x is the solution vector containing the location and height of the wind turbines.

7.3 Optimization model

Several points have been included in this model to make the problem closer to a real wind farm design than previous approaches. The main novelties in this model are the inclusion of the wind farm shape, an orography model and a cost model based on benefit/investment terms.

7.3.1 Wind farm shape model

Previous approaches in the literature have not taken into account the problem of the wind farm shape. Basically the majority of previous approaches consider squares wind farms, divided into cells, where turbines could be positioned [61, 119]. This square-based approach is interesting, since it introduces a nice way of managing the different possible points where a turbine can be installed, but the problem is that it cannot model the design of a real wind farm in a realistic way. An easy manner to consider different shapes for the wind farm is proposed, while keeping square cells to model a possible point to locate a wind turbine. The idea is really simple: over a square of length $K \times K$ cells which serves as a background, a binary template $\mathcal{T}$ $(K \times K)$ which describes the *zones allowed* to install turbines is defined. The elements of $\mathcal{T}$ are defined in such a way that $\mathcal{T}_{ij} = 1$ stands for a point included in the wind farm area, and $\mathcal{T}_{ij} = 0$ stands for a point outside of the wind farm. Note that with this simple idea almost any shape for the wind farm can be considered. As an example, Figure 7.2 shows the square background in black, and the allowed zone (described by binary matrix $\mathcal{T}$) in white.

7.3.2 Wake, orography model and wind speed simulation

In this work the wake model previously described in Section 7.2 is considered. Though it is a simple model, it works really well to simulate a real turbine's wake, obtaining a good balance between model's complexity and final performance. Moreover, several new concepts are considered in the problem formulation tackled in this work: first of all, note that none of the previous approaches to the problem takes into account the wind farm's orography or variations on wind speed. The existence of hills within the wind farm makes that the wind speed is different at the top of the hill or at the bottom of the corresponding valley. In order to take this important point into account, the concept of *wind speed multipliers* is included in the problem definition, in such a way that a higher point will be characterized by having a larger wind multiplier. Thus, when the wind speed associated to a given point in the wind farm is modified by means of the wind multiplier, the orography of the wind farm is being taken into account. Figure 7.3 shows and example of the wind multipliers in the previous wind farm example. Red areas stand for the largest wind multipliers, whereas blue areas stand for the smallest wind multipliers. This can be obtain by means of different software like CFD etc.

The wind speed modeling in a given point of the wind farm has been calculated in the following way: for each direction of the wind rose in the wind farm considered, a set of Monte Carlo simulations of t years wind are carried out, using a Weibull probability density function for the wind speed module. The result of the Monte Carlo simulation is weighted by the corresponding probability extracted from the wind rose and by the wind speed multipliers (in order to include the orography of the wind farm). The power curve shown in Figure 7.4 is used to obtain the power production associated to the wind speed in a given wind turbine.

7.3.3 Cost model

The cost model used in this work is based on a simplified model of investment/benefit considerations, similar to the one proposed in [144]. Specifically, the considered cost model includes wind turbines installation cost (C_i) and connection between turbines and road construction costs (C_{ij}^C), modeled as the Euclidean distance between turbine i and j. Also,

Figure 7.2: Example of the template to generate the wind farm shape.

the considered model includes the net benefit obtained from the energy produced in t years (B_t). All these parameters are measured in Euros. The objective function to be maximized is:

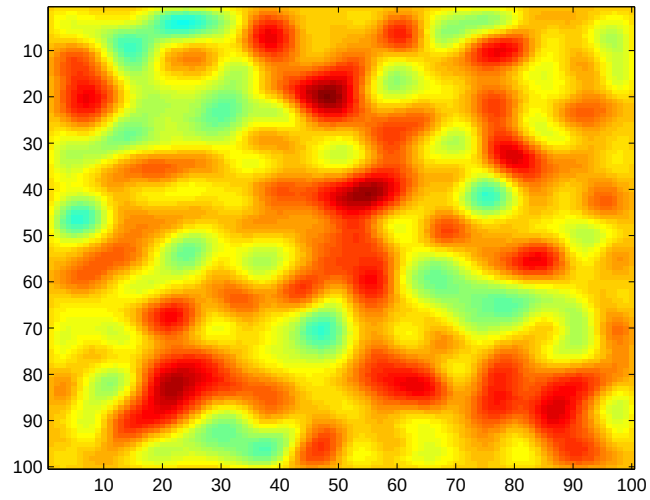
$$\varphi(\Xi) = B_t - N \cdot C_i - \sum_{i=1}^{N} \sum_{j<i} C_{ij}^{C} \tag{7.14}$$

where N stands for the number of wind turbines installed in the wind farm. Note that no alternative costs such as the operational costs (OPEX) are considered in this objective function. However, it is good enough to show the performance of the different compared algorithms.

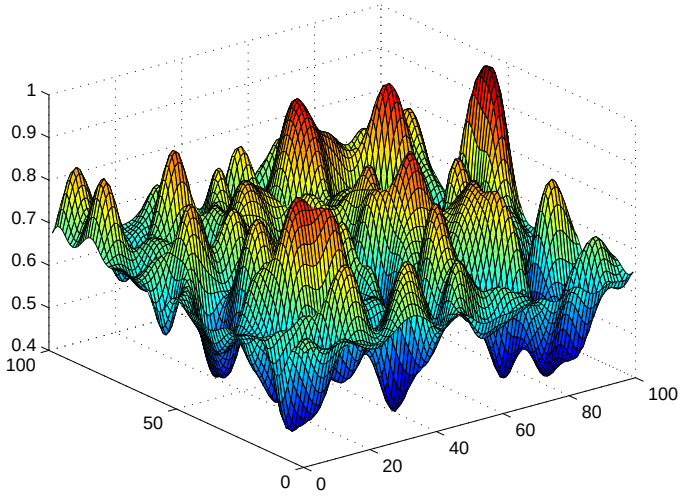
7.4 GHWTP: a greedy-constructive heuristic for wind turbines positioning

An heuristic approach to solve the wind turbine location problem with the proposed optimization model(see previous section) is first proposed: the Greedy Heuristic for Wind Turbines Positioning (GHWTP). There are previous works in the literature dealing with heuristics for optimal positioning of wind turbines in wind farms [131]; here, an ad-hoc heuristic approach is presented, which takes into account all the peculiarities of the considered optimization model.

The proposed heuristic can be considered as a greedy-constructive approach, based on exploiting the best locating points in terms of the objective function φ. The heuristic starts with the location of the point in the wind farm with the maximum wind speed (simulated in the way described in Section 7.3.2). A wind turbine is located at this point. The wind speed values in other points of the wind farm are then modified applying the wake model (Figure 7.5), taking into account the wind rose, and then another point to locate the second wind turbine is selected by considering the point which maximizes the objective function φ, i.e. point with maximum wind



(a)



(b)

Figure 7.3: Example of wind speed multipliers and orography model induced; (a) wind speed multipliers; (b) Orography model induced by the wind multipliers.

speed and with the minimum possible value of the connection term among turbines. This procedure is carried out until the desired number of wind turbines (N) are positioned in the wind farm.

1. Locate the point in the wind farm with maximum wind and positioning there the first wind turbine.

2. Correct the wind speed in neighbor points to turbines taking into account the wake model and wind rose considered.

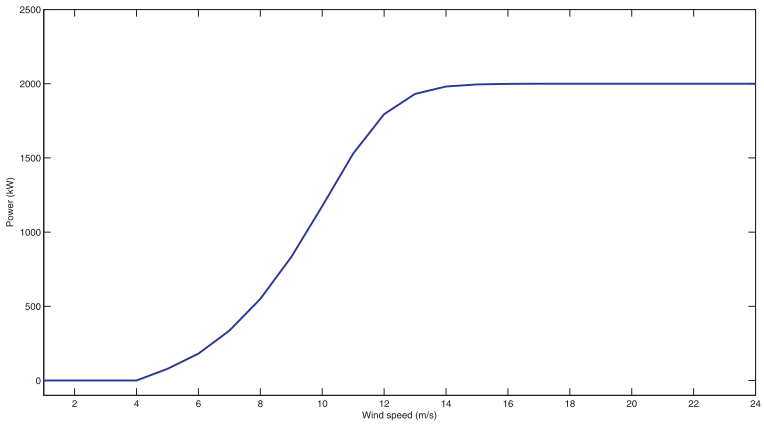


Figure 7.4: Power curve used in the simulations.

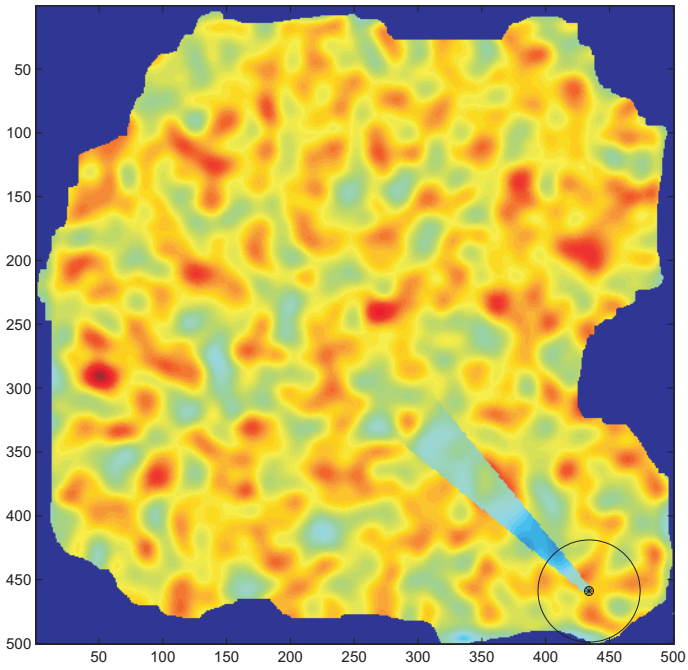


Figure 7.5: Modification of wind speed values when applying the wake model.

3. Locate the point in the wind farm which maximizes Equation (7.14), and set there the following wind turbine.

4. If N wind turbines have been installed, then stop. Otherwise go to step 2.

The distance between turbines is calculated in each iteration of the algorithm and as a part of the fitness calculation. To help us understand how the algorithm works, Figures 7.6 and 7.7 show different steps of the algorithm, in which a new turbine is added.

Note that the GHWTP takes into account the best points in terms of wind speed, but also in terms of the distance among wind turbines in Equation (7.14). The main characteristics of this heuristic is that it is really fast, and provides a reasonable solution in terms of the objective function given by Equation (7.14). However, it is not an optimal approach, since the positioning

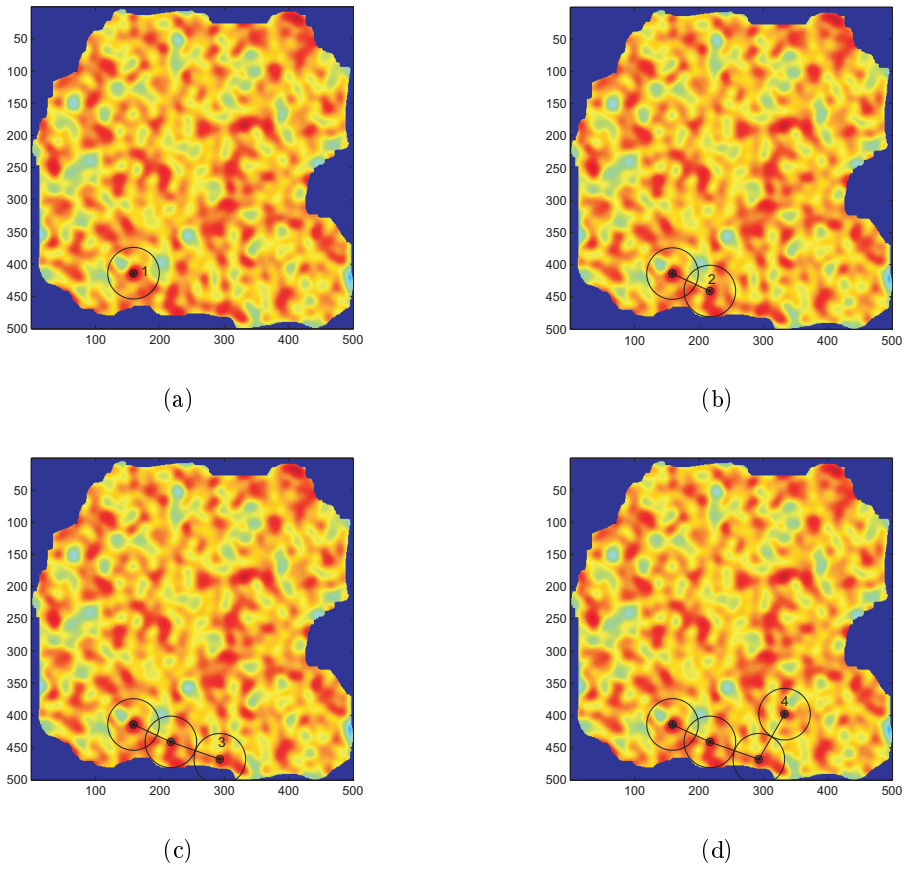


Figure 7.6: Example of turbines connection for a given wind farm. (a) Location of the first turbine in the maximum wind point; (b) Location of the turbine 2; (c) Location of the turbine 3; (d) Location of the turbine 4.

of the first wind turbines affect the positioning of the last ones, and suboptimal solutions may appear.

7.4.1 Case of study. How GHWTP works

To check the performance of the constructive heuristic and see the dependence between the objective function and the distance between turbines, the differential benefit obtain in each step of the algorithm is represented in Figure 7.9. This differential benefit is the benefit obtained when having i turbines, the $i+1$ turbine is installed, for a wind farm given in Figure 7.8. Note that locating a new turbine always produces a benefit, it is nonsense that a new turbine generates losses. Moreover, with this example, it is clarified and exemplified the dependence of the objective function (Equation 7.14) with the distance. In Figure 7.9, when installing the turbine # 9 or # 13, the benefit is significantly lower. This is due to the fact that the distance between turbines 8 and 9 and between turbines 12 and 13 is bigger than in other cases, and thus the benefits are lower because of the penalization of the distance (huge road construction costs).

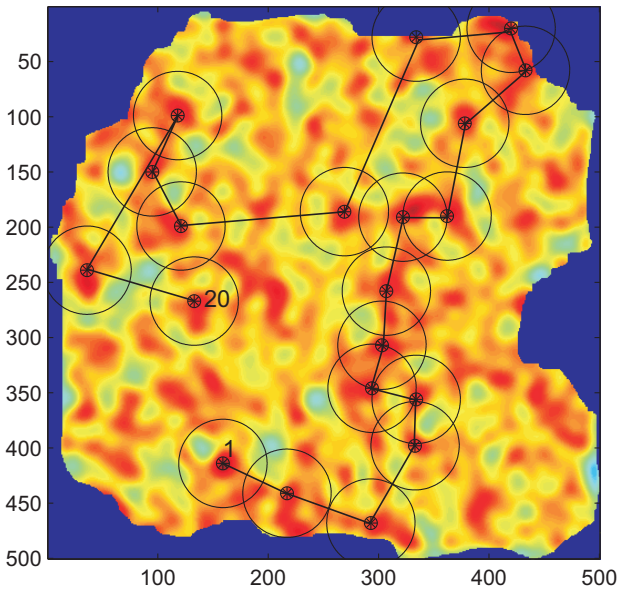


Figure 7.7: Connection after the location of 20 turbines.

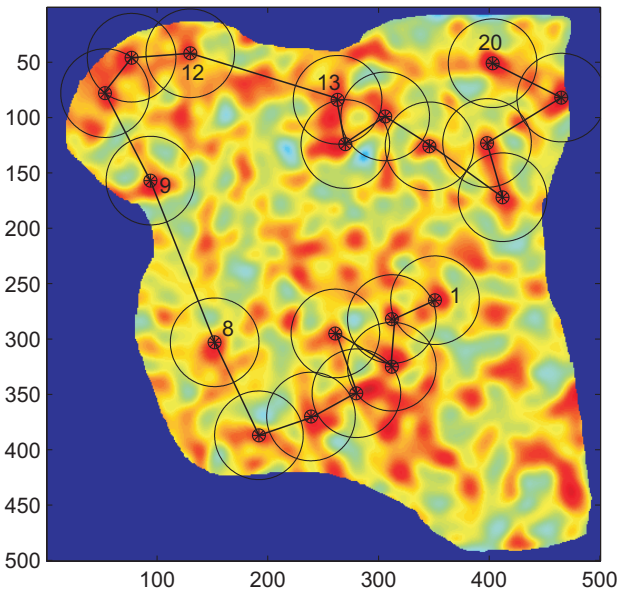


Figure 7.8: Wind farm under study.

7.5 The Evolutionary algorithm proposed

In Chapter 3, evolutionary algorithms (EAs) have already been described. However, in order to adapt the EA to the characteristics and constraints of the tackled problem, the evolutionary operators must be modified. Next the main characteristics of the evolutionary algorithm proposed are described, including the algorithm's initialization and selection, crossover and mutation operators proposed.

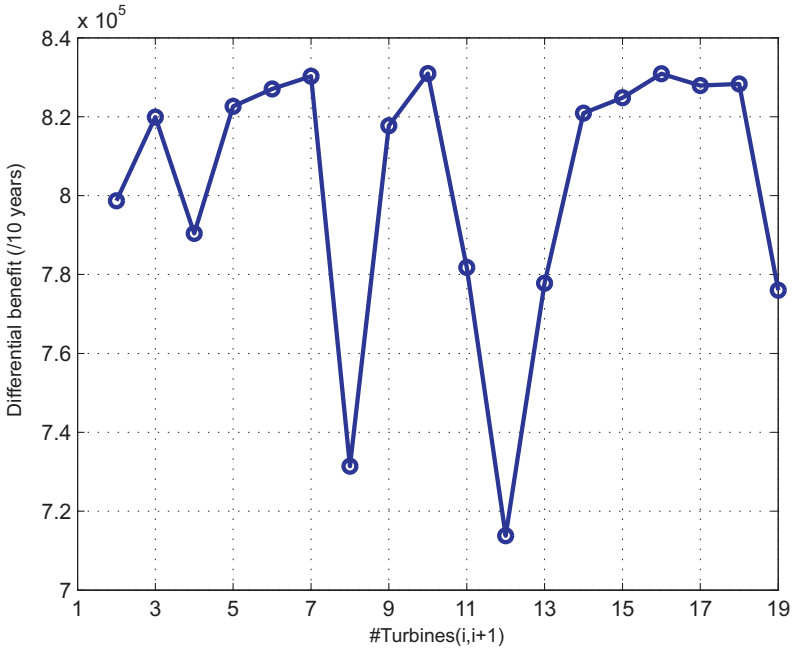


Figure 7.9: Differential benefit for 10 years of simulation, using the proposed GHWTP.

1. Generate an initial population of μ individuals (solutions). Let t be a counter for the number of generations, set it to $t = 1$. Each individual is taken as a matrix of integer vectors $\Xi = (x_i, y_i)$, $i = 1, \ldots, N$, where each x_i stands for the x-coordinate of turbine i in the background square considered, and each y_i stands for the y-coordinates ($x_i = 1, \ldots, K$, $y_i = 1, \ldots, K$). Note that every location point of a given solution Ξ, lets say (x, y), must fulfil a number of requisites to be considered as feasible: first, all the location points should be within the wind farm surface, i.e, the associate value in matrix $\mathcal{T}$ must be 1 ($\mathcal{T}_{xy} = 1, \forall\ x, y$). Second, a given turbine situated at a point (x, y) must be at least at a distance $\mathcal{D}$ of any other turbine. The initial individuals of the population are generated in such a way that these constraints are fulfilled.

2. Evaluate the fitness value for each individual Ξ of the population using the problem's objective function φ.

3. Generate an offspring population, of length μ, applying one-point crossover operator [44] and mutation operator. The crossover operator is applied in the traditional way (Figure 7.10 shows an example). On the other hand, mutation operator is carried out by randomly changing couples of specific points (x_i, y_i) to (x_i', y_i'), (see Figure 7.11 as an example).

4. Correct the offspring population in such a way that all their individuals are feasible (fulfil the problem's constraints). Note that the crossover operator may produce solutions within a distance $\mathcal{D}$ of another turbine, and mutation operator, in addition, may produce solutions which are outside the wind farm points defined by matrix $\mathcal{T}$. In order to correct these unfeasible solutions, a modification of unfeasible points is applied after each round of crossover and mutation, by using two random numbers $r, s \in [-2\mathcal{D}, 2\mathcal{D}]$: a given unfeasible point (x, y) is modified to $(x + r, y + s)$ until it is feasible.

5. Selection: Pass the best individual found so far in the evolution to the next generation. Conduct then pairwise comparison over the union of parents and offspring remaining: for each individual, p opponents are chosen uniformly at random from all the parents and offspring. The best individual in these p is selected to survive for the next generation. This process is repeated until a new parent generation of μ individuals is obtained.

6. Evaluate the fitness value for each individual Ξ of the new parent population using the problem's objective function φ.

7. Stop if the stopping criterion is satisfied, and if not, set $t = t + 1$ and go to Step 3. In this case, the stopping criterion established is that the best solution found by the algorithm is not improved during $\mathcal{K}$ generations, or, alternatively, the algorithm reaches to a maximum number of generations max_ite.

When working with the GHWTP, the fitness value of the solution is calculated iteratively, and in each step of the algorithm the objective function is calculated (and its associated value of distance between turbine $i-1$ and turbine i). However, in the evolutionary algorithm, all the N turbines are located at once, so a new mechanism to connect the turbines has to be implemented.

The fitness will depend on the way the connection is made. To put it simple, the first installed turbine is chosen (the one appearing first in the coordinates vector). Now, among the rest of the turbines, the one that maximices the fitness function is chosen. That way, those two turbines are the endpoints of the path. Then, among the rest of the turbines, we have to select the turbine that maximices the fitness function, taking as the distance value the distance between the turbine and each of the endpoints of our path. The chosen turbine will now be the new endpoint of the path, and the former endpoint is now in the path. Iteratively, the connection is made. Note that turbines can be added to the path in both of the sides of the path, so the path is not built in a single direction as it is with the GHWTP.

7.5.1 Seeding the EA with the GHWTP

This work also deals with the idea of using the solution obtained by the GHWTP as starting point for the EA (seed the EA). Many researchers have proposed to seed EAs with good initial solutions whenever it is possible, obtaining important improvements in the algorithm's convergence and quality of the solutions obtained. In this case it is proposed to seed part of the initial EA population with the GHWTP solution and some variations of this solution (obtained by means of mutation), and also to include randomly generated individuals to complete the initial population of the EA.

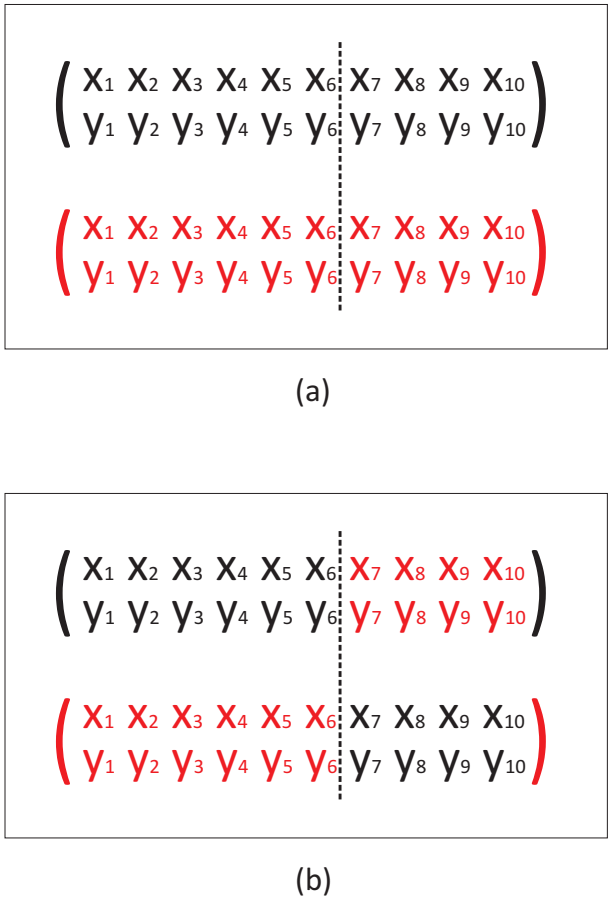


Figure 7.10: Example of the one-point crossover implemented in the proposed EA, in an example with $N = 10$ wind turbines; (a) inial couples of individuals and random-picked crossover point; (b) Final crossed individuals.

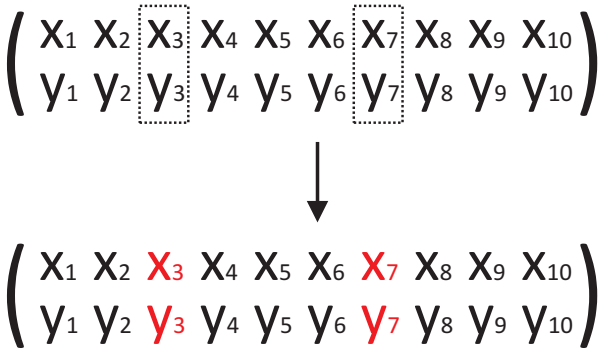


Figure 7.11: Example of the mutation operator implemented: first a number of points are randomly selected to be mutated. Second, new values are randomly selected in $K \times K$ and substitute previous values.

This way of initializing the evolutionary algorithm should improve the final solution obtained by the EA, since due to the elitist selection implemented, the best individual in a generation is maintained in the next generation. Thus, at least the EA seeded by the GHWTP will obtained

this solution as the best one. Moreover, it has been found that the seeded EA improves significantly the GHWTP solution, and the convergence of the algorithm is much better than the EA without this intelligent initialization, as will be shown in the experimental results.

7.6 Experiments and results

The experiments carried out and the results obtained, when the optimization model and the proposed algorithms are applied to different wind farms (different wind farm shape, orography and wind speed), are presented here. Two different wind farms are obtained, generating shape matrices $\mathcal{T}$, with 15 different orographies each.

7.6.1 General features

In order to characterise a wind farm, a grid of $K \times K, K = 500$ is set. A basic cell size of $10m \times 10m$ is considered, so the wind farms used in the experiments are framed by a $5km \times 5km$ surface. It should be pointed out that the number of cells and the resolution of them can be adjusted, being able to deal with larger wind farms or with higher precision. Another important parameter to be taken into account when locating the turbines is the minimum deployment distance $\mathcal{D}$. Because of the rotor diameter, at least, a distance $\mathcal{D}$ of 90 meters has to exist between turbines. Besides, a fixed number of turbines $N = 20$ are being located.

The wind speed is simulated, as described in Section 7.3.2, with a Weibull probabilistic distribution. The Weibull parameters considered are: $\lambda = 10$ and $\beta = 1,6$. The parameter λ is the mean value of the Weibull function, the mean wind speed value in this case; β is the shape parameter. The value of $\beta = 1,6$ is the one that best fit the real wind performance. Figure 7.12 shows the wind rose considered, extracted from a real wind farm in southern Spain, except from the wind speed module, constant in our case.

To characterise the objective function (Equation 7.14) the values of the parameters are set. The cost for the connection of wind towers and road construction (C_{ij}^c) has been set to 10^5 Euros/Km. The cost of each tower has been estimated to be $C_i = 10^6$ Euros. Regarding the calculation of the estimated benefit $B_t(\Xi)$ in the wind farm: 75 Euros/Mwatt-h, 3000 effective-hours/year and a period of $t = 10$ years of simulation are considered in the simulations.

The parameters of the evolutive algorithm are:

- The population size is $\mu = 50$.
- During the recombination process, μ new individuals are generated. The mutation probability is $p = 0,01$.
- In the selection process, a probabilistic tournament, 20 individuals randomly chosen fight. The tournament is carried out until a population of μ individuals is reached.

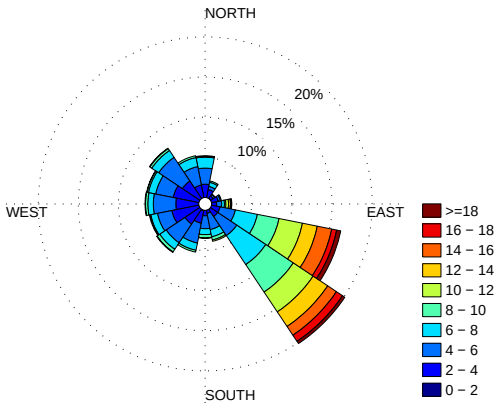


Figure 7.12: Wind rose considered in the simulations.

7.6.2 Results

Table 7.1 and 7.2 show the comparison between the GHWTP, EA and SEA algorithms in the two wind farm shapes considered, respectively. Note that the EA is able to obtain better solutions than the GHWTP in all the simulations. The SEA, however, outperforms the GHWTP and the EA in all the experiments carried out. Note the differences in performance obtained between the SEA and EA. It indicates that a good initialization procedure leads to a much better performance of the algorithm. On the other hand, in these tables it is possible to see the differences introduced by considering the orography of the wind farm, since the only differences between the different instances in both cases is the orography considered (random values of wind speed multipliers).

Figures 7.13 (a), (b) and (c) show an example of the solution obtained by the GHWTP, EA and SEA, respectively, in one instance of the wind farm #1. It is possible to see that the solutions obtained by the three algorithms considered fulfil the constraint given by the minimum deployment distance $\mathcal{D}$ (marked in the figures by the corresponding radius around a positioned wind turbine). Figures 7.14 (a), (b) and (c) shows an example for wind farm #2.

7.7 Conclusions

A novel evolutionary algorithm has been considered, initially seeded with the solution of a greedy approach, in a problem of optimal location of wind turbines in wind farms. A novel optimization model has also been proposed, which includes some new aspects such as wind farm shape, orography and different costs in the objective function. The wind farm shape models implies that any shape of any wind farm can be considered, while previous approaches only considered squared farms, limiting the design of a real wind farm. The orography model allows us to simulate the effect of the topography in the wind speed. Moreover, the wind does not blow always with the same intensity or direction, thus a good simulation of wind speed is needed. Finally, the costs

Table 7.1: Objective function values (in Euros/10^7), in the 15 different simulations performed, obtained by the GHWTP heuristic, Evolutionary Algorithm and Seeded Evolutionary Algorithm proposed for the wind farm #1.

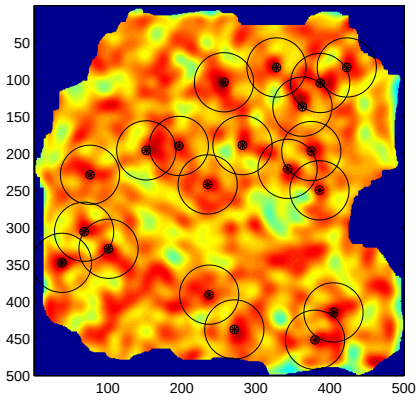
# instance	GHWTP	EA	SEA
1	2.063	2.748	4.992
2	1.844	4.569	5.852
3	2.454	4.671	5.936
4	1.752	4.035	4.728
5	3.264	4.565	4.873
6	1.945	3.768	5.576
7	1.899	2.913	3.813
8	1.841	4.520	4.992
9	1.854	4.368	4.731
10	1.670	3.953	4.467
11	1.755	3.242	4.881
12	0.883	2.399	3.374
13	2.247	3.660	4.853
14	1.684	3.206	4.040
15	2.851	3.648	4.685

Table 7.2: Objective function values (in Euros/10^7), in the 15 different simulations performed, obtained by the GHWTP heuristic, Evolutionary Algorithm and Seeded Evolutionary Algorithm proposed for the wind farm #2.

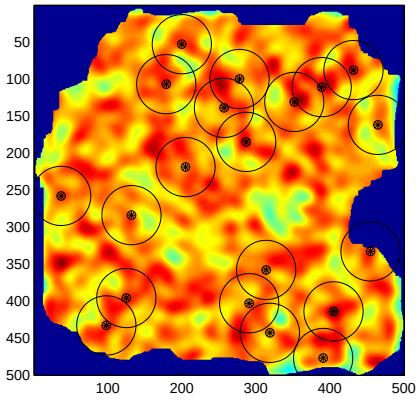
# instance	GHWTP	EA	SEA
1	2.283	3.613	3.483
2	3.305	5.029	5.262
3	1.002	3.741	3.642
4	1.697	4.364	5.662
5	2.171	3.818	3.990
6	1.849	3.964	4.403
7	1.683	3.334	4.350
8	1.776	4.530	3.915
9	1.201	4.046	5.642
10	2.813	3.934	4.517
11	1.676	3.590	4.039
12	1.778	3.647	4.371
13	1.767	3.170	4.959
14	2.837	4.534	4.834
15	1.575	4.072	4.832

model includes several novelties, and is based on the benefit obtained from the energy production and the costs of wind turbines, installation and road construction.

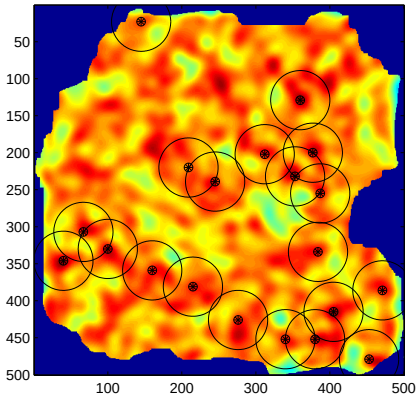
Several experiments have been carried out, where the performance of the proposed algorithms is shown. A greedy heuristic, an evolutionary algorithm and an evolutionary algorithm seeded with the solution found by the heuristic are tested. It is shown the good performance of the seeded evolutionary in the design of wind farms.



(a)

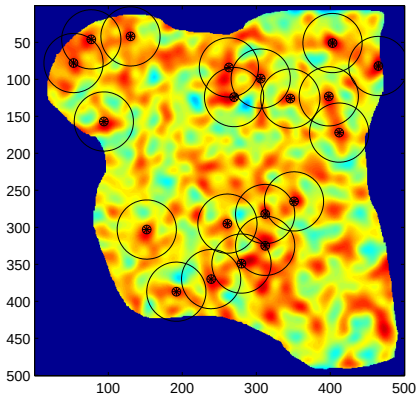


(b)

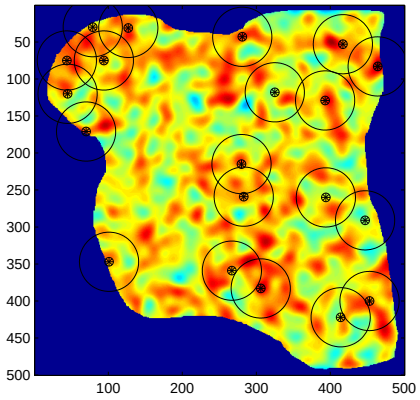


(c)

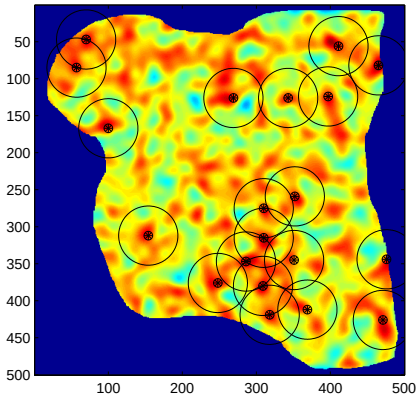
Figure 7.13: Final wind turbines disposition, security radius and wind multipliers; (a) GHWTP; (b) EA; (c) SEA.



(a)



(b)



(c)

Figure 7.14: Final wind turbines disposition, security radius and wind multipliers; (a) GHWTP; (b) EA; (c) SEA.

Part III

Final remarks and future lines of work

Chapter 8

Final remarks and future lines of work

In this thesis we have tackled a number of problems that arise in the process of prospection, planning and management of on-shore wind farms. Specifically, we have solved a problem of wind speed reconstruction from pressure patterns, useful in wind farms prospection to avoid issues related to scarce of wind speed data. A novel problem of spatial reconstruction of wind speed statistics (through the reconstruction of the Weibull parameters at different points of the wind farm) has been also tackled. This problem is useful both in prospection and also in micrositing (final layout) of the wind farms. A novel algorithm for turbines layout in on-shore wind farms has been proposed, and finally a problem of gap reconstruction in wind speed series and short-term wind speed prediction from neighbor towers is tackled using fast-training neural computation techniques.

In wind speed reconstruction, we have developed a hybrid genetic-Support Vector Regression algorithm, which is able to locate the best points at a grid of pressure, and use these data to obtain a robust wind speed reconstruction from this pressure pattern. The algorithm has been tested in three different sites and an algorithm based on Weather Regimes is used for comparison purposes. Both techniques show a good performance,obtaining an excellent average reconstruction of wind speed values, but missing maximum and minimum values in most cases. However, the aim of these algorithms is to obtain a general good reconstruction, that is completely obtained by the developed approach. The problem of spatial reconstruction of wind speed statistics is also useful in the prospection process of the wind farm. It consists of modifying wind speed numerical models outputs using the data from some measuring towers installed in the prospection zone. The final objective is to obtain a surface of wind speed values with tunable resolution, by means of computational techniques. In this case we have developed a number of novel heuristic approaches to obtain this surface. An additional novelty of the proposal is to work with matrices of Weibull parameters (A and k) in the prospection zone, which can be later converted into wind speed values. We have developed two different heuristics, one of them based on evolutionary computation to take into account the correction from the different measuring towers in this process. Results in two different prospection sites with different number of measuring towers have been obtained. First, in a case with a large number of measuring towers (13), the results of the surface obtained is excellent in terms of a measure of error using a cross validation process. In the second case with a lower number of towers (6), the results are poorer as expected, and also we have shown that the surface cannot be generated with information from less than 4 measuring stations. In the problem of on-shore wind farm micrositing we have developed a new approach based on evolutionary computation. Specifically, the problem tackled included a model of orography, a model of wind farm shape, and we have also included costs from civil

works needed to connect the wind turbines. An evolutionary approach seeded with and heuristic algorithm at the initialization has been developed. We have shown the excellent performance of the proposal in several simulations, outperforming the behaviour of a greedy algorithm and also the performance of the evolutionary algorithm without smart initialization. Finally, we have shown the performance of a number of fast-training neural algorithms in wind speed series reconstruction from neighbor measuring towers. A software based on Extreme Learning Machines, GMDHs (fast-training approaches), and also on SVR and MLPs networks have been developed and tested by comparison with the multi-linear regression, which is the technique currently used by the majority of wind speed managers. Clear advantages of using the fast-training and alternative machine learning algorithms instead of the multi-linear regression have been shown.

In terms of results, the works developed in this thesis has produced 4 articles in top international journals (see Appendix A), an invited review for the initial number of another international journal, 4 articles in national and international conferences and a software license (RMCPO, see Chapter 5), currently under exploitation by Iberdrola in the management of wind farms.

Regarding future lines of research, this thesis, due to its wide range of applications, opens a broad amount of possible research lines in the near future.

1. A line we have already followed has been to extend the meteorological variables for wind speed reconstruction to other different from pressure patterns. This implies to reduce the size of the grid considered (because the number of variables grow exponentially), and improve the feature selection process in order to locate the best set of variables to carry out the wind speed reconstruction. The idea of extend this problem to medium and long-term wind speed prediction has been also considered as a valuable future research line of this thesis.

2. A second line of future work we like to highlight is related to the problem of spatial reconstruction of wind speed numerical models in wind farms. In this case, it is possible to change the heuristics proposed in a number of ways, for example by changing the influence of the neighbor towers or even including exogenous variables that may be available in the area.

3. Maybe the most promising future work is related to the extension of the micrositing approach (turbines layout) in this work. Off-shore wind farms can be also design using the proposed techniques, but new constraints must be included, such as different wakes behaviour, influence of neighbor wind farms, dead zones design (for example to plan heliports or ports for wind farm maintenance, etc.). Also, in on-shore design, the improvement of the proposed approach is possible by including alternative constraints and costs (financing of the wind farm to be included in the final revenue from the facility). This problem can also be tackled as a multi-objective problem, exploiting the large number of approaches existing for this type of problems. The exploration of alternative algorithms different from evolutionary computation is also a must: since the publication of the article describing our proposal for on-shore layout, more than 50 works have cited our paper (Google Scholar data), and a large number of novel approaches have been proposed. The idea is to extend our proposal by including some of the new ideas given in these works for specific problems or sites, etc. The exploration of variable-length algorithms can also be a promising future line of research, which would deal with problems in which the number of turbines is open, instead of fixed, as our current approach.

4. A future line of research out of the specific problems tackled in this thesis is to evaluate the effect of climate change in the production of wind energy, and the other way around, the effect of renewable energy in climate change mitigation. We are aware of different works related to this interesting area, such as [53, 138]. The idea is to exploit our know-how in Machine Learning to obtain new results in this topic.

Part IV

Appendix

Appendix A

List of publications

This section is a compilation of the scientific publications produced as a result of this research work, apart from those other studies conducted during the training process.

A.1 Papers related to the research work performed in this PhD.

Papers in international journals

1. B. Saavedra-Moreno, A. Iglesia, J. Magdalena-Saiz, L. Carro-Calvo, L. Durán, S. Salcedo-Sanz, "Surface wind speed reconstruction from synoptic pressure fields: machine learning versus weather regimes classification techniques", *Wind Energy*, 2014. (JCR: 3.069)

2. B. Saavedra-Moreno, S. Salcedo-Sanz, C. Casanova-Mateo, J. A. Portilla-Figueras and L. Prieto, "Heuristic correction of wind speed mesoscale models simulations for wind farms prospecting and micrositing", *Journal of Wind Engineering and Industrial Aerodynamics*, vol.130, pp. 1-15, 2014. (JCR: 1.414)

3. B. Saavedra-Moreno, S. Salcedo-Sanz, L. Carro-Calvo, J. Gascón-Moreno, S. Jiménez-Fernández, L. Prieto, "Very fast training neural-computation techniques for real measure-correlate-predict wind operations in wind farms", *Journal of Wind Engineering and Industrial Aerodynamics*, vol 116, pp.49-60, 2013, (JCR: 1.414).

4. S. Salcedo-Sanz, B. Saavedra-Moreno, A. Paniagua-Tineo, L. Prieto, A. Portilla-Figueras, "A review of recent evolutionary computation-based techniques in wind turbines layout optimization problems", *Central European Journal of Computer Science*, 2011.

5. B. Saavedra-Moreno, S. Salcedo-Sanz, A. Paniagua-Tineo, L. Prieto, A. Portilla-Figueras, "Seeding evolutionary algorithms with heuristics for optimal wind turbines positioning in wind farms", *Renewable Energy*, vol. 36, pp. 2838-2844, 2011, (JCR: 3.476).

International conferences

1. B. Saavedra-Moreno, S. Salcedo-Sanz, L. Carro-Calvo, J. A. Portilla-Figueras and J. Magdalena-Saiz, "Reconstruction of wind speed based on synoptic pressure values and Support Vector regression", *The 14th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'2013)*, Heifei, China, October, 2013.

2. B. Saavedra-Moreno, S. Salcedo-Sanz, A. Paniagua-Tineo, J. Gascón-Moreno, A. Portilla-Figueras, "Optimal Evolutionary Wind Turbine Placement in Wind Farms Considering New

Models of Shape, Orography and Wind Speed Simulation", *International Work Conference on Artificial Neural Networks*, IWANN 2011, Part I, LNCS 6691, pp. 25-32, 2011.

National conferences

1. B. Saavedra-Moreno, S. Salcedo-Sanz, A. Paniagua-Tineo, A. Portilla-Figueras, "Diseño inteligente de parques eólicos mediante nuevos modelos de computación heurística", *I Encuentro de Investigadores en Infraestructuras Inteligentes*, Guadalajara, España, 2011.

2. B. Saavedra-Moreno, S. Salcedo-Sanz,"Introducción a los algoritmos de tipo evolutivo y sus aplicaciones", *Workshop on Mathematical and Soft-Computing Techniques in Risk Management. Applications to Actuarial Science*, Universidad Complutense, 2010.

A.2 Other publications archieved during the training process

Papers in international journals

1. S. Salcedo-Sanz , R. C. Deo, L. Carro-Calvo, B. Saavedra-Moreno, "Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms", *Theoretical and Applied Climatology*, pp.1-13, 2015, (JCR: 2.015)

2. J. Gascón-Moreno, S. Salcedo-Sanz, B. Saavedra-Moreno, L. Carro-Calvo, A. Portilla-Figueras, "An Evolutionary-based Hyper-Heuristic Approach for Optimal Construction of Group Method of Data Handling Networks", *Information Sciences*, vol.247, pp.94-108, 2013, (JCR: 4.038).

3. J. Gascón-Moreno, E.G. Ortiz-García, S. Salcedo-Sanz, L. Carro-Calvo, B. Saavedra-Moreno, A. Portilla-Figueras, "Evolutionary optimization of multi-parametric kernel ϵ - SVMr for forecasting problems", *Soft Computing*, vol. 17, pp.213-221, 2013, (JCR: 1.271).

4. A. M. Ahmadzadeh, J. E. Sánchez-García, B. Saavedra-Moreno, A. Portilla-Figueras, S. Salcedo-Sanz, "Capacity estimation algorithm for simultaneous support of multi-class traffic services in Mobile WiMAX", *Computer Communications*, vol.35, pp.109-119, 2012, (JCR: 1.695).

International conferences

1. J. Gascón-Moreno, S. Salcedo-Sanz, E.G. Ortiz-García, L. Carro-Calvo, B. Saavedra-Moreno, A. Portilla-Figueras, "A binary-encoded tabu-list genetic algorithm for fast Support Vector Regression hyper-parameters tuning", *International Conference on Intelligent Systems Design and Applications (ISDA)*, 2011.

2. J. Gascón-Moreno, S. Salcedo-Sanz, E.G. Ortiz-García, A. Paniagua-Tineo, B. Saavedra-Moreno, A. Portilla-Figueras, "Multi-parametric Gaussian Kernel Function Optimization for ϵ-SVMr Using a Genetic Algorithm", *International Work Conference on Artificial Neural Networks*, IWANN 2011, Part II, LNCS 6693, pp. 113-120, 2011.

3. A. Paniagua-Tineo, S. Salcedo-Sanz, E.G. Ortiz-García, J. Gascón-Moreno, B. Saavedra-Moreno, A. Portilla-Figueras, "On the performance of μ -GA Extreme Learning Machines in Regression Problems", *International Work Conference on Artificial Neural Networks*, IWANN 2011, Part II, LNCS 6693, pp. 153-160, 2011.

4. A. Paniagua-Tineo, S. Salcedo-Sanz, E.G. Ortiz-García, A. Portilla-Figueras, B. Saavedra-Moreno, G. López-Díaz, "Greenhouse Indoor Temperature Prediction based on Extreme Learning Machines for Resource-constrained Control Devices Implementation", *Highlights in Practical Applications of Agents and Multiagent Systems Advances in Intelligent and Soft Computing*, vol. 89, pp. 203-211, 2011.

5. J. E. Sánchez-García, A. M Ahmadzadeh, B. Saavedra-Moreno, S. Salcedo-Sanz and A. Portilla-Figueras, "Strategic Methods for Radio Access Design in 2G/3G Networks", *3rd International ICST Conference on Mobile Lightweight Wireless Systems (Mobilight 2011)*, Bilbao, Spain, 2011.

Part V

Bibliography

Bibliography

[1] R. E. Abdel-Aal, M. A. Elhadidy and S. M. Shaahid "Modeling and forecasting the mean hourly wind speed time series using GMDH-based abductive networks," *Renewable Energy*, vol. 34, no. 7, pp. 1686-1699, 2009. 16, 25

[2] S. Al-Yahyai, Y. Charabi, A. Gastli and A. Al-Badi, "Wind farm land suitability indexing using multi-criteria analysis," Renewable Energy, vol. 44, pp. 80-87, 2012. 75

[3] L. Anastasakis and N. Mort, "The development of self-organization techniques in modelling: A review of the group method of data handling (gmdh)". Research Report, Department of Automatic Control and Systems Engineering, The University of Sheffield, (813), 2001. 25

[4] American Wind Energy Association,"!0 Steps in building a wind farm", Wind Energy Fact Sheet. 9

[5] J. Badger, A. Hahmann, X. Guo-Larsen, A. Peña-Díaz, E. Batchvarova, S. E. Gryning, et al. "Comprehensive utilization of mesoscale modelling for wind energy applications," In *Proc. of the EWEA annual event*, 2011. 16

[6] 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.

[7] R.J. Barthelmie, F. Murray and S.C. Pryor, "The economic benefit of short-term forecasting for wind energy in the UK electricity market," *Energy Policy*, vol. 36, no. 5, pp. 1687-1696, 2008.

[8] M. Beccali, G. Cirrincione, A. Marvuglia and C. Serporta, "Estimation of wind velocity over a complex terrain using the generalized mapping regressor," *Applied Energy*, vol. 87, no. 3, pp. 884-893, 2010. 16

[9] H. G. Beyer and H. P. Schwefel, "Evolution Strategies: a comprehensive introduction," *Natural Computing*, vol. 1, no. 1, pp. 3-52, 2002. 79

[10] M. Bilbao and E. Alba, "CHC and SA applied to wind energy optimization using real data", *In. Proceedings of the IEEE COnference on Evolutionary Computation*, pp. 1-8, 2010. 19

[11] M. Bilgili, B. Sahin and A. Yasar, "Application of neural networks for the wind speed prediction of target station using reference stations data", *Renewable Energy*, vol.32, no. 14, pp. 2350-2360, 2007. 15, 16

[12] M. Bilgili, A. Yasar and E. Simsek, "Offshore wind power development in Europe and its comparison with onshore counterpart", *Renewable and Sustainable Energy reviews*, vol.15, pp. 905-915, 2011. 19

[13] C. M. Bishop, *Neural networks for pattern recognition*, Oxford University Press, 1995. 22, 23

[14] I. D. Bishop and C. Stock, "Using collaborative virtual environments to plan wind energy installations," *Renewable Energy*, vol. 35, no. 10, pp. 2348-2355, 2010.

[15] R. Fletcher, *Practical Methods of Optimization.* John Wiley & Sons, Inc. 2nd Edition, 1987.

[16] A. Blum and P. Langley, "Selection of relevant features and examples in Machine Learning," *Artificial Intelligence*, vol. 97, pp. 245-271, 1997. 38, 39

[17] H. Bouzgou and N. Benoudjit, "Multiple architecture system for wind speed prediction," *Applied Energy*, vol. 88, no. 7, pp. 2463-2471, 2011. 13

[18] S. H. Brown, "Multiple linear regression analysis: a matrix approach with MATLAB", *Alabama Journal of Mathematics*, Spring/Fall, 2009. 55

[19] M. Burlando, M. Antonelli and C. F. Ratto, "Mesoscale wind climate analysis: identification of anemological regions and wind regimes," *International Journal of Climatology*, vol. 28, pp. 629-641, 2008. 37

[20] E. Cadenas and W. Rivera, "Wind speed forecasting in the south coast of Oaxaca, Mexico," *Renewable Energy*, vol. 32, pp. 2116-2128, 2007.

[21] O. Carranza, E. Figueres, G. Garcera and L. G. Gonzalez, "Comparative study of speed estimators with highly noisy measurement signals for Wind Energy Generation Systems," *Applied Energy*, vol. 88, no. 3, pp. 805-813, 2011. 13

[22] L. Carro-Calvo, S. Salcedo-Sanz, N. Kirchner-Bossi, A. Portilla-Figueras, L. Prieto, R. Garcia-Herrera and E. Hernández-Martín, "Extraction of synoptic pressure patterns for long-term wind speed estimation in wind farms using evolutionary computing," *Energy*, vol. 36, no. 3, pp. 1571-1581, 2011. 14, 37

[23] L. Carro-Calvo, S. Salcedo-Sanz, L. Prieto, N. Kirchner-Bossi, A. Portilla-Figueras and S. Jiménez-Fernández, "Wind speed reconstruction from synoptic pressure patterns using an evolutionary algorithm," *Applied Energy*, vol. 89, no. 1, pp. 347-354, 2012. 14, 37

[24] J. A. Carta, S. Velázquez and J.M. Matías, "Use of Bayesian networks classifiers for long-term mean wind turbine energy output estimation at a potential wind energy conversion site," *Energy Conversion and Management*, vol. 52, no. 2, pp. 1137-1149, 2011. 16

[25] J. A. Carta and S. Velázquez, "A new probabilistic method to estimate the long-term wind speed characteristics at a potential wind energy conversion site," *Energy*, vol. 36, no. 5, pp. 2671-2685, 2011. 15

[26] J. Castro-Mora, "Optimización global de parques eólicos mediante algoritmos evolutivos", Tesis Doctoral, Universidad de Sevilla, 2008. 9

[27] C.-C. Chang and C.-J. Lin, "LIBSVM : a library for support vector machines", 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm

[28] Z.H. Chen, S. Y. Cheng, J.B. Li, X.R. Guo, W.H. Wang and D. S. Chen, "Relationship between atmospheric pollution processes and synoptic pressure patterns in northern China," *Atmospheric Environment*, vol. 42, no. 24, pp. 6078-6087, 2008. 13, 37

[29] W. L. Cheng "Synoptic weather patterns and their relationship to high ozone concentrations in the Taichung Basin," *Atmospheric Environment*, vol. 35, no. 29, pp. 4971-4994, 2001. 13

[30] S. Chowdhury and J. Zhang, "Exploring key factors influencing optimal farm design using mixed-discrete particle swarm optimization," *In Proceedings of the 13th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, pp. 1-16, 2010. 18

[31] S. Chowdhury, J. Zhang, A. Messac and L. Castillo, "Unrestricted wind farm layout optimization UWFLO: investigating key factors influencing maximum power generation," *Journal of Renewable Energy*, vol. 38, pp. 16-30, 2012. 18

[32] Coriolis Energy, "The wind farm cycle". 9

[33] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.

[34] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable and Sustainable Energy Reviews*, vol. 12, pp. 1725-1744, 2008. 37

[35] A. Coville, A. Siddiqui and K. O. Vogstad, "The effect of missing data on wind resource estimation," *Energy*, vol. 36, no. 7, pp. 4505-4517, 2011. 15

[36] L. D. Davis, "Handbook of Genetic Algorithms,"Van Nostrand Reinhold Company, 1991.

[37] "Principio n.3 de la Declaración de Río sobre el Medio Ambiente y el Desarrollo, "Conferencia de las Naciones Unidas sobre el Medio Ambiente y el Desarrollo, Río de Janeiro, Junio, 1992.

[38] A. Derrick, "Development of the measure-correlate-predict strategy for site assessment", *Proceedings of the BWEA*, 1992. 15

[39] E. Diday and J.C. Simon, "Clustering analysis. In: K.S. Fu (ed.): Digital Pattern Classification", *Springer Verlag*, pp. 47-94, 1976. 39

[40] A. L. Dutot, J. Rynkiewicz, F. E. Steiner and J. Rude, "A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions," *Environmental Modelling and Software*, vol. 22, pp. 1261-1269, 2007. 57

[41] R. Eberhart and Y. Shi, Particle swarm optimization: developments, applications and resources, In *Proc. IEEE Congress on Evolutionary Computation*, 2001. 28

[42] http://data-portal.ecmwf.int/data/d/interim_full_daily

[43] "2011 Wind Technologies Market Report," EERE, U.S. Department of Energy, http://www1.eere.energy.gov/wind/pdfs/2011_wind_technologies_market_report.pdf

[44] A. E. Eiben and J. E. Smith, Introduction to evolutionary computing, Springer-Verlag, 2003. 28, 104

[45] M. Ekstrom, G.H. McTainsh and A. Chappell, "Australian dust storms: temporal trends and relationships with synoptic pressure distributions (1960-99)," *International Journal of Climatology*, vol. 24, pp. 1581-1599, 2004. 13

[46] C.N. Elkinton, J.F. Manwell and J. G. McGowan, "Optimization algorithms for offshore wind farm micrositing," In *Proc. of the WINDPOWER Conference and Exhibition*, Los Angeles, CA, USA, 2007. 19

[47] C. N. Elkinton, J. F. Manwell and J. G. McGowan, "Algorithms for offshore wind farm layout optimization ," *Wind Engineering*, vol. 32, no. 1, pp. 67-84, 2008. 19

[48] A. Emami and P. Noghreh, "New approach on optimization in placement of wind turbines within wind farm by genetic algorithms," *Renewable Energy*, vol. 35, no. 7, pp. 1559-1564, 2010. 18, 95, 97

[49] D. P. Dee, S. M. Uppal, A. J. Simmons et. al., "The ERA-Interim reanalysis: configuration and performance of the data assimilation system", *Quarterly Journal of the Royal Meteorological Society*, vol. 137, no. 656, pp. 553-597, 2011.

[50] 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. 13

[51] Y. Eroglu and S. U. Seckiner, "Design of wind farm layout using ant colony algorithm," *Renewable Energy*, vol.44, pp. 53-62, 2012. 19

[52] D. A. Fadare, "The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria", *Applied Energy* vol. 87, no. 3, pp. 934-942, 2010. 13, 16

[53] C. Fant, C. A. Schlosser, K. Strzepek, "The impact of climate change on wind and solar resources in southern Africa", *Applied Energy* vol. 161, no. 1, pp. 556-564, 2016. 117

[54] D. B. Fogel, "An introduction to simulated evolution," *IEEE Transactions Neural Networks*, vol. 5, pp. 3-14, 1994. 79

[55] D. Gabor, W. Wildes and R. Woodcock, "A universal non-linear filter predictor and simulator which optimices itself by a learning process", *Proc IEEE*, 108B, pp. 422-438, 1961. 25

[56] E. García-Bustamante, J. F. González-Rouco, P. A. Jiménez, J. Navarro and J. P. Montávez, "A comparison of methodologies for monthly wind energy estimation," *Wind Energy*, vol. 12, no. 7, pp. 640-659, 2009. 37

[57] Z. W. Geem, J. Hoon Kim and G. V. Loganathan, A New Heuristic Optimization Algorithm: Harmony Search, *Simulation*, 76(2) (2001) 60-68. 28

[58] S. L. Goh, M. Chen, D.H. Popovic, K. Aihara, D. Obradovic and D.P. Mandic, "Complex-valued forecasting of wind profile," *Renewable Energy*, vol. 31, pp. 1733-1750, 2006.

[59] D. E. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Reading, MA: Addison-Wesley, 1989.

[60] G. Grassi and P. Vecchio, "Wind energy prediction using a two-hidden layer neural network", *Commun. Nonlinear Sci. Number. Simulat.*, vol. 15, no. 9, pp. 2262-2266, 2010. 16

[61] S. A. Grady, M. Y. Hussaini and M. M. Abdullah, "Placement of wind turbines using genetic algorithms," *Renewable Energy*, vol. 30, no. 2, pp. 259-270, 2005. 18, 95, 97, 98

[62] J. R. D'Errico, "Understanding GRIDFIT", December, 2006. 77

[63] H. Gu, J. Wang, "Irregular-shape wind farm micro-siting optimization", *Energy*, vol. 57, pp. 535-544, 2013. 19

[64] P. A. Gutierrez, S. Salcedo-Sanz, C. Hervï£¡s-Martï£¡nez, L. Carro-Calvo, J. A. Sï£¡nchez-Monedero and L. Prieto, "Ordinal and nominal classification of wind speed from synoptic pressure patterns," *Engineering Applications of Artificial Intelligence*, vol. 26, pp. 1008-1015, 2013. 14, 37

[65] "GWEC Global Wind Statistics 2011," Global Wind Energy Commission, http://gwec.net/wp-content/uploads/2012/06/GWEC_-_Global_Wind_Statistics_2011.pdf

[66] "Global Wind Report: Annual Market Update 2014", Global Wind Energy Comission. 5, 6, 7

[67] J.A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm", *J. Roy. Stat. Soc. Series C (Applied Statistics)*, vol. 28, no. 1 , pp. 100-108. 39

[68] M. T. Hagan and M. B. Menhaj, "Training feed forward network with the Marquardt algorithm", *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989-993, 1994. 23

[69] S. Haykin, *Neural networks: a comprenhensive foundation*, Prentice Hall, 1998. 22, 23

[70] J. Herbert-Acero, J. Franco-Acevedo, M. Valenzuela-Rendon and O. Probst-Olezewski, "Linear wind farm layout optimization through computational intelligence", *In. Proceedings of the Mexican International Conference on Artificial Intelligence*, pp. 692-703, 2009. 19

[71] F. O. Hocaoglu, Ö. N. Gerek, M. Kurban, "A novel wind speed modeling approach using atmospheric pressure observations and hidden Markov models," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 98, no. 8-9, pp. 472-481, 2010. 14

[72] J. Hossain, V. Sinha and V.V.N. Kishore, "A GIS based assessment potential for windfarms in India", *Renewable Energy*, vol. 36, no. 12, pp. 3257-3267, 2011 10

[73] G. B. Huang, Q. Y. Zhu and C. K. Siew, "Extreme Learning Machine: Theory and Applications," *Neurocomputing*, Vol. 70, N. 1–3, pp. 489–501, 2006. 24

[74] G. B. Huang, L. Chen and C. K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879-892, 2006. 24

[75] G. B. Huang and L. Chen, Convex incremental extreme learning machine, *Neurocomputing*, 70 (2007) 3056-3062. 24

[76] G. B. Huang and L. Chen, Enhanced random search based incremental extreme learning machine, *Neurocomputing*, 71 (2008) 3460-3468. 24

[77] G. B. Huang, D. H. Wang and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107-122, 2011. 24

[78] G. B. Huang, H. Zhou, X. Ding and R. Zhang, "Extreme learning machine for regression and multi-class classification," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 42, no. 2, pp. 513-529, 2012. 24

[79] H. Huang, "Efficient hybrid distributed genetic algorithms for wind turbine positioning in large wind farms," In *Proc. of the 2009 IEEE International Symposium on Industrial Electronics*, pp. 2196-2201, 2009. 18

[80] "Renewable Energy Technologies: Cost Analysis series", Volume 1: Power Sector, Issue 5/5 Wind power, June 2012, *available online*. 11

[81] A.G. Ivakhnenko, "The group of method of data handling, a rival of the method of stochastic approximation", Soviet Automatic Control c/c of Automatica, vol. 1, no. 3, pp. 43-55, 1968. 25

[82] S. H. Jangamshetti and V. G. Rau, "Optimum siting of wind turbine generators," *IEEE Transactions on Energy Conversion*, vol. 16, no. 1, pp. 8-13, 2001.

[83] S. Janjai, I. Masiri, W. Promsen, S. Pattarapanitchai, P. Pankaew, J. Laksanaboonsong, I. Bischoff-Gauss, N. Kalthoff, "Evaluation of wind energy potential over Thailand by using an atmospheric mesoscale model and a GIS approach", *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 129, pp. 1-10, 2014. 10

[84] P. A. Jiménez, J. F. Gonzalez-Rouco, J. P. Montavez, E. García-Bustamante and J. Navarro, "Climatology of wind patterns in the northeast of the Iberian Peninsula," *International Journal of Climatology*, vol. 29, pp. 501-525, 2009. 37

[85] M. J. Kaiser and B. F. Snyder "Modeling offshore wind installation costs on the U.S. outer continental shelf," *Renewable Energy*, vol. 50, February 2013, pp. 676-691, 2013. 75

[86] J. K. Kaldellis and D. Zafirakis "The wind energy (r)evolution: A short review of a long history," *Renewable Energy*, vol. 36, no. 7, pp. 1887-1901, 2011.

[87] S.A Kalogirou, "Artificial neural networks in renewable energy systems applications: a review", *Renewable and Sustainable Energy Reviews*, vol. 5, no. 4, pp. 373-401, 2001. 15

[88] S. Kalogirou, "Optimization of solar systems using artificial neural-networks and genetic algorithms," *Applied Energy*, vol. 77, pp. 383-405, 2004. 25

[89] G. S. Karioniotakis and E. F. Nogaret, "Wind power forecasting using advanced neural networks models," *IEEE Transactions on Energy Conversion*, vol. 11, no. 4, pp. 762-767, 1996.

[90] M. Khashei, M. Bijari and G. Raissi-Ardali, "Improvement of Auto-Regressive Integrated Moving Average Models Using Fuzzy Logic and Artificial Neural Networks (ANNs)," *Neurocomputing*, vol. 72, no. 4-6, pp. 956-967, 2009. 13

[91] S. A. Khan and S. Rehman, "Iterative non-deterministic algorithms in on-shore wind farm design: a review," *Renewable and Sustainable Reviews*, vol. 19, pp. 370-384, 2013. 19

[92] J. Y. Kim, K. Y. Oh, K. S. Kang and J. S. Lee, "Site selection of offshore wind farms around the Korean Peninsula through economic evaluation," *Renewable Energy*, vol. 54, pp. 189-195, 2013. 75

[93] D. Y. Kim, J. Y. Kim and J. J. Kim, "A regression-based statistical correction of mesoscale simulations for near-surface wind speed using remotely sensed surface observations", *Asia-Pacific Journal of Atmospheric Sciences*, vol. 48, no. 4, pp. 449-456, 2012. 17

[94] N. Kirchner-Bossi, L. Prieto, R. García-Herrera, L. Carro-Calvo and S. Salcedo-Sanz, "Multi-decadal variability in a centennial reconstruction of daily wind," *Applied Energy*, vol. 105, pp. 30-46, 2013. 37

[95] R. Kohavi, and G. H. John, "Wrappers for Features Subset Selection," *International Journal of Digital Libraries*, vol. 1, pp. 108-121, 1997. 39

[96] M. Kuchcik, "Deaths and car accidents under different synoptic situations," *In Proc. of the 15th Conf on Biometeorology and Aerobiology and the 16th International Congress of Biometeorology*, Kansas City, United States, 2002. 13

[97] A. Kusiak, H. Zheng and Z. Song, "Wind farm power prediction: a data-mining approach," *Wind Energy*, vol. 12, no. 3, pp. 275-293, 2009. 37

[98] A. Kusiak and Z. Song, "Design of wind farm layout for maximum wind energy capture," *Renewable Energy*, vol. 35, pp. 685-694, 2010. 18, 95

[99] M. A. Lackner and C. N. Elkinton, "An analytical framework for offshore wind farm layout optimization," *Wind Engineering*, vol. 31, no. 1, pp. 17-31, 2007. 19

[100] M. A. Lackner, A. L. Rogers and J. F. Manwell, "The round robin site assessment method: A new approach to wind energy site assessment," *Renewable Energy*, vol. 33, no. 9, pp. 2019-2026, 2008. 15

[101] M. A. Lackner, A. L. Rogers, J. F. Manwell and Jon G. McGowan "A new method for improved hub height mean wind speed estimates using short-term hub height data," *Renewable Energy*, vol. 35, no. 10, pp. 2340-2347, 2010. 15

[102] D. Latinopoulos, K. Kechagia, " A GIS-based multi-criteria evaluation for wind farm site selection. A regional scale application in Greece," *Renewable Energy*, vol. 78, pp. 550-560, 2015. 10

[103] G. Li and J. Shi, "Application of Bayesian model averaging in modeling long-term wind speed distributions," *Renewable Energy*, vol. 35, no. 6, pp. 1192-1202, 2010. 13

[104] 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. 13

[105] H. Liu, E. Erdem and J. Shi, "Comprehensive evaluation of ARMA-GARCH(-M) approaches for modeling the mean and volatility of wind speed," *Applied Energy*, vol. 88, no. 3, pp. 724-732, 2011. 13

[106] P. López, R. Velo and F. Maseda, "Effect of direction on wind speed estimation in complex terrain using neural networks", *Renewable Energy*, vol. 33, no. 10, pp. 2266-2272, 2008. 16

[107] M. C. Mabel and E. Fernández, "Analysis of wind power generation and prediction using ANN: A case study," *Renewable Energy*, vol. 33, no. 5, pp. 986-992, 2008.

[108] G. Marmidis, S. Lazarou and E. Pyrgioti, "Optimal placement of wind turbines in a wind park using Monte Carlo simulation," *Renewable Energy*, vol. 33, no. 7, pp. 1455-1460, 2008. 95

[109] J.L. Martínez-Ramos, J. Castro, J. Riquelme-Santos, M. Burgos-Payán, "A hybrid evolutive algorithm for wind farm optimum network design," *Artificial Intelligence in Energy Systems and Power*, pp. 1-5, Madeira, Portugal, 2006. 18

[110] A. Mellit, S. A. Kalogirou, L. Hontoria and S. Shaari, "Artificial intelligence techniques for photovoltaic applications: A review," *Progress in Energy and Combustion Science*, vol. 34, no. 5, pp. 406-419, 2009. 13

[111] P. Michelangeli, R. Vautard and B. Legras, "Weather regimes: recurrence and quasi-stationarity", *Journal of the Atmospheric Sciences*, vol. 52, pp. 1237-1256, 1995. 40

[112] K. Mo and M. Ghil, "Cluster analysis of multiple planetary flow regimes", *Journal of Geophysical Research: Atmospheres (1984-2012)*, vol. 93, no. D9, pp. 10927-10952, 1988. 40

[113] M. A. Mohandes, T. O. Halawani, S. Rehman and A. A. Hussain, "Support vector machines for wind speed prediction", *Renewable Energy*, vol. 29, no. 6, pp. 939-947, 2004. 13, 16, 26

[114] M. Mohandes, S. Rehman and S. M. Rahman, "Estimation of wind speed profile using adaptive neuro-fuzzy inference system (ANFIS)," *Applied Energy*, vol. 38, no. 11, pp. 4024-4032, 2011. 13

[115] F. Molteni, S. Tibaldi and T.N. Palmer, " Regimes in the wintertime circulation over northern extratropics. I: Observational evidence", *Quarterly Journal of the Royal Meteorological Society*, vol. 116, no. 491, pp. 31-67, 1990. 40

[116] J. Mora, J. Baron, J. Santos and M. Payan, "An evolutive algorithm for wind farm optimal design," *Neurocomputing*, vol. 710, pp. 2651-2658, 2007. 18

[117] J. M. Morales, R. Minguez and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Applied Energy*, vol. 87, no. 3, pp. 843-855, 2010. 13

[118] A. A. Mortimer, "A new correlation/prediction method for potential wind farm sites", *Proceedings of the BWEA*, 1992. 15

[119] G. Mosetti, C. Poloni and B. Diviacco, "Optimization of wind turbine positioning in large wind farms by means of a genetic algorithm," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 51, no. 1, pp. 105-116, 1994. 17, 18, 95, 97, 98

[120] K. P. Moustris, I. C. Ziomas and A. G. Paliatsos, "3-days-ahead forecasting of regional pollution index for the pollutants NO_2, CO, SO_2 and O_3 using artificial neural networks in Athens," *Soil Pollution Journal*, vol. 209, pp. 29-43, 2010. 57

[121] Medium-Term Renewable Energy Market Report 2014, International Energy Agency, *online available* 3

[122] I. Mustakerov and D. Borissova, "Wind turbines type and number choice using combinatorial optimization," *Renewable Energy*, vol. 35, pp. 1887-1894, 2010.

[123] N. Nawri,G. N. Petersen, H. Björnsson and K. Jonasson, "Statistical correction of WRF mesoscale model simulations of surface wind over Iceland based on station data," Icelandic Meteorological Office Report no. 2012-11, 2012. 17, 78

[124] N. Nawri,G. N. Petersen, H. Björnsson and K. Jonasson, "Evaluation of WRF mesoscale model simulations of surface wind over Iceland," Icelandic Meteorological Office Report no. 2012-11, 2012. 17

[125] http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml

[126] P. Nolan, P. Lynch and C. Sweeney, "Simulating the future wind energy resource of Ireland using the COSMO-CLM model," *Wind Energy*, in press, 2013. 37

[127] http://www.awsopenwind.org/ 20

[128] E. G. Ortiz-García, S. Salcedo-Sanz, A. M. Pérez-Bellido and J. A. Portilla-Figueras, "Improving the training time of support vector regression algorithms through novel hyper-parameters search space reductions," *Neurocomputing*, vol. 72, pp. 3683-3691, 2009. 44

[129] S. Osowski, and K. Garanty, "Forecasting of the daily meteorological pollution using wavelets and support vector machine," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 6, pp. 745-755, 2006. 13

[130] A. Öztopal, "Artificial neural network approach to spatial estimation of wind velocity data", *Energy Conversion and Management*, vol. 47, no. 4, pp. 395-406, 2006. 16

[131] U. A. Ozturk and B. A. Norman, "Heuristic methods for wind energy conversion system positioning," *Electric Power Systems Research*, vol. 70, no. 3, pp. 179-185, 2004. 99

[132] A. Paniagua-Tineo, S. Salcedo-Sanz, C. Casanova-Mateo, E.G. Ortiz-García, M.A. Cony, E. Hernández-Martín, "Prediction of daily maximum temperature using a support vector regression algorithm," *Renewable Energy*, vol. 36, no. 11, pp. 3054-3060, 2011. 13

[133] D. Paredes, R. M. Trigo, R. García-Herrera and I. F. Trigo, "Understanding precipitation changes in Iberia in early spring: weather typing and storm tracking approaches," *Journal of Hydrometeorology*, vol. 7, pp. 101-113, 2006. 13

[134] J.P. Peixoto and A. H. Oort, "Physics of climate", Springer-Verlag, 1992. 39

[135] M.G. Pereira, R.M. Trigo, C.C. da Camara, J.M.C. Pereira and S.M. Leite, "Synoptic patterns associated with large summer forest fires in Portugal," *Agricultural and Forest Meteorology*, vol. 129, pp. 11-25, 2005. 13

[136] B. Pérez, R. Mínguez and R. Guanche, "Offshore wind farm layout optimization using mathematical programming techniques," *Renewable Energy*, vol. 53, pp. 389-399, 2013. 19

[137] S. Pookpunt and W. Ongsakul, "Optimal placement of wind turbines within wind farm using binary particle swarm optimization with time-varying accelaration coefficients", *Renewable Energy*, vol. 55, pp. 266-276, 2013. 19

[138] S. C. Pryor and R. J. Barthelmie, "Climate change impacts on wind energy: A review," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 1, pp. 430-437, 2010. 117

[139] S. C. Pryor and R. J. Barthelmie, "Assessing climate change impacts on the near-term stability of the wind energy resource over the United States," *Proceedings of the National Academy of Sciences*, vol. 108, 8167-8171, 2011. 37

[140] R. Rahmani, A. Khairuddin, S. Cherati and H. Pesaran, "A novel method for optimal placing wind turbines in a wind farm using particle swarm optimizaction (PSO)," *In. Proceedings of the IEEE International Conference on Power Engineering*, pp. 134-139, 2010. pp. 2651-2658, 2007. 18

[141] "Renewables 2014. Global Futures Report", Renewable Energy Policy Network for the 21st Century (REN21). 7

[142] "Renewables 2014. Global Status Report", Renewable Energy Policy Network for the 21st Century (REN21). 3

[143] E.E. Riddle, M.B. Stoner, C.J. Johnson, M.L. L'Heureux, D.C. Collins and S. B. Feldstein, "The impact of the MJO on clusters of wintertime circulation anomalies over the North American region", *Clim. Dyn.*, doi: 10.1007/s00382-012-1493-y 40

[144] J. Riquelme-Santos, M. Burgos-Payan, J. M. Calero and J. Castro Mora, "An evolutive algorithm for wind farm optimal design," *Neurocomputing*, vol. 70, no. 16-18, pp. 2651-2658, 2007. 18, 97, 98

[145] R. A. Rivas, J. Clausen, K. S. Hansen, and L. E. Jensen, "Solving the turbine positioning problem for large offshore wind farms by simulated annealing," *Wind Engineering*, vol. 33, no. 3, pp. 287-297, 2009. 19

[146] A. L. Rogers, J. W. Rogers, J. F. Manwell, "Comparison of the performance of four measure-correlate-predict algorithms", *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 93, no. 3, pp. 243-264, 2005. 15

[147] R. Romero, G. Summer, C. Ramis and A. Genoves, "A classification of the atmospheric circulation patterns producing significant daily rainfall in the Spanish Mediterranean area," *International Journal of Climatology*, vol. 19, 765-785, 1999. 13, 37

[148] A. Romo-Perea, J Amezcua and O. Probst, "Validation of three new measure-correlate-predict models for the long-term prospection of the wind resource," *Journal of Renewable and Sustainable Energy*, vol. 3, 023105, 2011. 16

[149] F.Rothlauf, "Representations for genetic and evolutionary algorithms," Springer, 2006.

[150] B. Saavedra-Moreno, S. Salcedo-Sanz, L. Carro-Calvo, A. Portilla-Figueras and J. Magdalena-Saiz, "Reconstruction of Wind Speed Based on Synoptic Pressure Values and Support Vector Regression," *Lecture Notes in Computer Science*, vol. 8206, pp. 310-317, 2013. 37

[151] R. Saidur, M.R. Islam, N.A. Rahim, K.H. Solangi, "A review on global wind energy policy," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 7, pp. 1744-1762, 2010.

[152] S. Salcedo-Sanz, E. G. Ortiz-García, A. M. Pérez-Bellido, A. Portilla-Figueras and L. Prieto, "Short term wind speed prediction based on evolutionary support vector regression algorithms," *Expert Systems with Applications*, vol. 38, no. 4, pp. 4052-4057, 2011. 44

[153] S. Salcedo-Sanz, D. Gallo-Marazuela, A. Pastor-Sánchez, L. Carro-Calvo, A. Portilla-Figueras and L. Prieto, "Offshore wind farm design with the Coral Reefs Optimization algorithm", *Renewable Energy*, vol. 63, pp. 109-115, 2014. 19

[154] J. R. Salmon and J. L. Walmsley, "A two-site correlation model for wind speed, direction and energy estimates," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 79, no 3, pp. 233-268, 1999. 15

[155] J. Serrano-González, A. G. González-Rodríguez, J. Castro-Mora, J. Riquelme-Santos and M. Burgos-Payán, "Optimization of wind farm turbines layout using an evolutive algorithm," *Renewable Energy* vol. 35, no. 8, pp. 1-11, 2010. 18

[156] J. Serrano-González, J. Riquelme-Santos and M. Burgos-Payán, "Wind farm optimal design including risk," In *Proc. of the Modern Electric Power Systems*, Wroclaw, Poland, 2010. 18

[157] J. Serrano-González, M. Burgos-Payán, J. Riquelme-Santos and F. González-Longatt, "A review and recent developments in the optimal wind-turbine micro-siting problem", *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 133-144, 2014. 19

[158] A. Simmons,C. Uppala, D. Dee and S. Kobayashi, "ERA-Interim: New ECMWF reanalysis products from 1989 onwards", *ECMWF Newsletter*, no. 110, pp. 25-35, 2007. 42

[159] S. Sisbot, Ö. Turgut, M. Tunï£¡, and Ü. ï£¡amdali, "Optimal positioning of wind turbines on Gökceada using multi-objective genetic algorithm," *Wind Energy*, vol. 13, no. 4, pp. 297-306, 2010. 18

[160] A. J. Smola, N. Murata, B. Schölkopf and K. Muller, "Asymptotically optimal choice of ϵ-loss for support vector machines", *In proc. of the 8th International Conference on Artificial Neural Networks, Perspectives in Neural Computing*, 1998. 26

[161] C. Soriano, A. Fernández and J. Martin-Vide, "Objective Synoptic classification combined with high resolution meteorological models for wind mesoscale studies," *Meteorology and Atmospheric Physics*, vol. 91, pp. 165-181, 2006. 14

[162] E. Sreevalsan, S. S. Das, R. Sasikumar and M. P. Ramesh, "Wind farm site assessment using Measure-Correlate-Predict (MCP) analysis", *Wind Engineering*, vol. 31, no. 2, pp. 111-116, 2007. 15

[163] X. Sun, D. Huang and G. Wu, "The current state of offshore wind energy technology development," *Energy*, vol. 41, pp. 298-312, 2012. 19

[164] R. Swisher, C. Real de Azua and J. Clendenin, "Strong winds on the horizon: wind power comes of age," *Proceedings of the IEEE*, vol. 89, no. 12, pp. 1757-1764, 2001.

[165] L. I. Tegou, H. Polatidis and D. A. Haralambopoulos, "Environmental management framework for wind farm siting: Methodology and case study," *Journal of Environmental Management*, vol. 91, no. 11, pp. 2134-2147, 2010. 75

[166] J.L. Torres, A. García, M. De Blas and A. De Francisco, "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)," *Solar Energy*, vol. 79, pp. 65-77, 2005. 13

[167] R. M. Trigo and C. C. DaCamara, "Circulation weather types and their influence on the precipitation regime in Portugal," *International Journal of Climatology*, vol. 20, pp. 1559-1581, 2000. 13

[168] A. J. Smola and B. Schölkopf,"A tutorial on support vector regression", *Statistics and Computing*, 1998. 26, 27

[169] S. Uppala, D. Dee, S. Kobayashi, P. Berrisford and A. Simmons, "Towards a climate data assimilation system: status update of ERA-Interim", *ECMWF Newsletter*, no. 115, pp. 12-18, 2008. 42

[170] S. Velázquez, J. A. Carta and J.M. Matías, "Influence of the input layer signals of ANNs on wind power estimation for a target site: a case study", *Renewable and Sustainable Energy Reviews*, vol. 15, no. 3, pp. 1556-1566, 2011. 15

[171] S. Velázquez, J. A. carta, J. M. Matías, "Comparison between ANNs and linear MCP algorithms in the lorg-term estimation of the cost per kWh produced by a wind turbine at a candidate site: a case study in the Canary Islands", *Applied Energy*, vol. 88, no. 11, pp. 3869-3881, 2011. 15

[172] M. Wagner, K. Veeramachaneni, F. Neumann, U. OâĂšReilly, "Optimizing the layout of 1000 wind turbines", *European Wind Energy Association Annual Event*, 2011. 19

[173] C. Wan, J. Wang, G. Yang and X. Zhang, "Optimal siting of wind turbines using real-coded genetic algorithms," in *Proc. of the EWEC*, Marseille, France, 2009. 18

[174] C. Wan, J. Wang, G. Yang, and X. Zhang, "Optimal micro-siting of wind farms by Particle Swarm Optimization," in *Advances in Swarm Intelligence*, Berlin, Heidelberg, 2010, pp. 198-205. 18

[175] C. Wan, G. Yang, X. Li and X. Zhang, "Optimal micro-siting of wind turbines by genetic algorithms based on improved wind and turbine models," In *Proc. of the 48th IEEE Conference on Decision and Control*, Shanghai, China, pp. 5092-5096, 2009. 18

[176] WAsP, http://www.wasp.dk 11

[177] H. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio and V. Vapnik,"Feature Selection for SVMs," *Advances in NIPS 12*, MIT Press, pp. 526-532, 2000. 39

[178] D.S. Wilks, "Statistical Methods in Atmospheric Sciences", Elsevier, 630 pp, 2006.

[179] C. J. Wilmott, S. G. Ackleson and R. E. Davis,"Statistics for the evaluation and comparison of models," *Journal of Geophysics Research*, vol. 90, pp. 8995-9005, 1985. 57

[180] http://www.winddata.com 42

[181] WindFarm, http://www.resoft.co.uk/Spanish/index.HTM 11

[182] WindPro, http://www.emd.dk 11

[183] "2011 Wind Power European Statistics," European Wind Energy Association, pp. 1-11, Feb. 2012,
http://www.ewea.org/fileadmin/files/library/publications/statistics/-Wind_in_power_2011_European_statistics.pdf

[184] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, et al., "A description of the Advanced Research WRF Version 3," NCAR Technical Note NCAR/TN-475+STR, *National Center for Atmospheric Research*, Boulder, Colorado, USA, 2008. 80

[185] "Half-year Report 2014", World Wind Energy Association. 6

[186] X. Yao, Y. Liu and G. Lin, "Evolutionary Programming made faster," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 2, pp. 82-102, 1999. 44, 80

[187] L. A. Zadeh, "Soft Computing and Fuzzy Logic," *Software*, IEEE, 1994, vol. 11, no 6, p. 48-56. 21

[188] J. Zhang, S. Chowdhury, A. Messac and L. Castillo, "A response surface-based cost model for wind farm design," *Energy Policy*, vol. 42, pp. 538-550, 2012.

[189] Y.Zhang, H. Mao, A. Ding, D. Zhou and Congbin Fu, "Impact of synoptic weather patterns on spatio-temporal variation in surface O3 levels in Hong Kong during 1999ï£¡2011," *Atmospheric Environment*, vol. 73, pp. 41-50, July 2013. 13

[190] M. Zhao, Z. Chen and J. Hjerrild, "Analysis of the behaviour of genetic algorithm applied in optimization of electrical system design for offshore wind farms," In *Proc. of the 32nd IEEE Conference on Industrial Electronics*, pp. 2335-2340, 2006. 19