Forecasting Wind Power for the Day-Ahead Market using Numerical Weather Prediction Models and Computational Intelligence Techniques

GIOVANNA MARTÍNEZ ARELLANO

A thesis submitted in partial fulfilment of the requirements of
Nottingham Trent University for the degree of

Doctor of Philosophy

April 2015

To my husband and son; the apple of my eye...

Acknowledgements

ing as a professional and as a human being. Their understanding and love has always helped me to overcome the difficult moments over the last four years.

<div align="right">

Giovanna Martínez Arellano

April 2015

</div>

Abstract

Wind power forecasting is essential for the integration of large amounts of wind power into the electric grid, especially during large rapid changes of wind generation. These changes, known as *ramp events*, may cause instability in the power grid. Therefore, detailed information of future ramp events could potentially improve the backup allocation process during the Day Ahead (DA) market (12 to 36 hours before the actual operation), allowing the reduction of resources needed, costs and environmental impact.

It is well established in the literature that meteorological models are necessary when forecasting more than six hours into the future. Most state-of-the-art forecasting tools use a combination of Numerical Weather Prediction (NWP) forecasts and observations to estimate the power output of a single wind turbine or a whole wind farm. Although NWP systems can model meteorological processes that are related to large changes in wind power, these might be *misplaced* i.e. in the wrong physical position. A standard way to quantify such errors is by the use of NWP ensembles. However, these are computationally expensive. Here, an alternative is to use *spatial fields*, which are used to explore different numerical grid points to quantify variability. This strategy can achieve comparable results to typical numerical ensembles, which makes it a potential candidate for ramp characterisation.

A major disadvantage of most ramp events studies is that they are based on a binary classification, which specifies a percentage of change in power within a defined time window. This may produce artifacts, as ramp detection tools might miss potential changes due to errors in the forecasts. Moreover, a change just below the threshold could be

equally damaging as a change that meets the definition. The novel contribution of this project is the application of computational intelligence techniques for wind power forecasting and ramp event characterisation. To achieve this, two stages are required. In the first stage, Genetic Programming (GP) is used to generate an ensemble of wind power forecasts based on the idea of spatial fields. This in its own is an important contribution as the approach will allow the development of computationally cheap wind speed-to-power conversion models, without making any assumptions of their shape or properties. In the second stage, wind power forecasts are converted into a set of *filtered signals* in order to study ramp events at different time scales. These signals, when applied to a set of Fuzzy Logic rules, indicate the probabilities of a ramp event happening, avoiding the binary classification, which is another important contribution of this work.

The observation data used for this investigation was obtained from a real wind park in Galicia, Spain and some observation points in Illinois, USA. The numerical data was obtained by running locally a Mesoscale model. Experiments showed that the accuracy of wind power forecasts obtained using GP as a downscaling/conversion method are comparable to traditional forecasting tools as it is able to achieve an 87% of accuracy. At the same time the computational effort was significantly reduced. The novel ramp detection approach that is introduced here, is able to outperform a basic binary-based detection algorithm. In addition, the fuzzy rules can provide a probability of other events happening; events that might not meet the crisp definition. Using colour maps, which are easier to interpret by human non-experts, it is possible to show how an event is developing in different time windows. Finally, it is shown how neighbouring points can help modelling events that might not be detected using only the closest point of the grid. Having a detailed characterisation of future ramp events can help grid operators to make more informed decisions on the scheduling of back-up units needed and hence to potentially reduce costs and the environmental impact.

Publications

The following publications have been published as a direct result of this thesis:

Refereed Journal Papers

G. Martínez-Arellano, L. Nolle, R. Cant, A. Lotfi and C. Windmill. Characterisation of Large Changes in Wind Power for the Day-Ahead Market Using a Fuzzy Logic Approach. KI - Künstliche Intelligenz, August 2014. DOI: 10.1007/s13218-014-0322-3

Refereed Conference Papers

G. Martínez-Arellano, R. Cant and L. Nolle. Prediction of Jet Engine Parameters for Control Design Using Genetic Programming, UKSim 14th International Conference on Modelling and Simulation, Cambridge, UK, 2014 pp. 45-50, DOI: 10.1109/UKSim.2014.64.

G. Martínez-Arellano and L. Nolle. Short-term Wind Power Forecasting with WRF-ARW Model and Genetic Programming. In Radomil Matousek editor, 19th International Conference on Soft Computing, MENDEL 2013, Brno, Czech Republic, 2013, ISBN: 978-80-214-4755-4.

G. Martínez-Arellano and L. Nolle. Genetic Programming for Wind Power Forecasting and Ramp Detection. Proceedings of the Thrity-third SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence (AI 2013), pp. 403-417, Cambridge, UK, 2013, DOI: 10.1007/978-3-319-02621-

3_30.

G. Martínez-Arellano, L. Nolle and J. A. Bland. Improving WRF-ARW Wind Speed Predictions using Genetic Programming. Proceedings of the Thirty-second SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence (AI 2012), pp. 347-360, Cambridge, UK, 2012, DOI: 10.1007/978-1-4471-4739-8_27.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General Context

Ever since the discovery of fire, we have learnt to improve our quality of life making use of our natural resources as the energy supply. Before the Industrial Revolution in the 18th Century, human and animal force, wind and water were our energy sources, while firewood was our source of heat and light. With the invention of the steam engine, the development of electricity and the internal combustion engine, a historical transition was made in the way we use energy. Nowadays, our manufacturing, transportation, home commodities and lighting rely on the constant supply of energy and, as the population continues to increase, so does the demand for energy (Schobert, 2002).

In the last decades, renewable energy sources have attracted special attention due to a global concern over issues such as global warming and fossil fuel depletion. Wind power has had the strongest growth in the electricity markets over the recent years (World Wind Energy Association, 2012). With international agreements, such as the Kyoto Protocol (United Nations, 1997), industrialized countries and the European Union have committed to supply a certain percentage of their energy demand with renewable sources. In a first stage of the agreement, 37 countries committed to reduce green house gas (GHG) emissions to an average of five percent against 1990 levels. In a second commitment period, a higher number of countries joined the agreement and the reduction was set to 18 percent below

1990 levels from 2013 to 2020. With this target, there has been an important increase in the amount of installed wind energy capacity in Europe. Countries such as Germany, Denmark and Spain have successfully integrated large amounts of wind power generation. In 2013, the total installed capacity in Europe reached 117 GW, which in a normal wind year would produce 257 TWh of electricity. This is enough to cover 8% of the EU's electricity consumption.

To meet the 20% target of renewable generation by 2020, Europe is increasing its offshore installations. More than 90% of the world's offshore wind power is currently installed off northern Europe, in the North, Baltic and Irish Seas, and the English Channel. The potential of offshore wind is enormous and is generally greater than onshore wind, generating more energy with fewer turbines. Although offshore wind is often the most talked about part of the wind sector, today it represents only about 2% of global installed capacity (Global Wind Energy Council, 2013).

Regardless of the type of installation, offshore or onshore, the power generation of wind farms is intermittent, which is directly correlated to the natural variability of wind. The capacity factor (actual energy output) of a wind farm is between 25 and 40% of its total installed capacity, making this source of energy unreliable in the eye of the utilities. With the increasing growth in wind energy capacity, there is a major concern about the imbalances that high levels of wind power could create in the grid. For this reason, different solutions have been studied and are still in development to mitigate these.

One alternative for smoothing intermittent wind power is to distribute wind farms in a large area with different wind flows (Cochran et al., 2012). Another alternative is the use of energy storage devices (Díaz-González et al., 2012) which, until now, appear not to be feasible for large scale wind generation. A recent alternative is the demand response which can potentially help to manage the unexpected deviations of wind generation (Kowli and Meyn, 2011). Another alternative is the use of operating reserves, which are generation units that are flexible enough to be adapted to the variations of load and supply. The use of reserves is currently the most common way to mitigate the problem. This, however, implies additional costs and emissions.

The scheduling of operating reserves is a current practice of grid operators

to balance the electric grid whenever there is a disruption in the supply. The main challenge when integrating larger amounts of wind power into the grid (more than 10%) is during the *commitment* of these reserve units. If the power output of a wind farm could be predicted accurately, then it could be treated as a conventional power plant where the operating reserve could be estimated and units could be committed according to that prediction. However, the lack of accurate forecasts adds further uncertainty to the commitment decision and could result in the commitment of unnecessary reserves.

Meanwhile, several studies have been performed to understand the technical and financial difficulties that a high penetration of wind energy will bring to the electricity market (Australian Energy Market Operator, 2011; Cochran et al., 2012; Energinet, 2010). As a result, it has been suggested that the electricity market must change its practices in order to increase power systems flexibility. Practices such as moving to sub-hourly dispatch and reducing gate closure times in order to bid when more accurate forecasts are available have been recommended as they have been successfully applied in some markets. An example of this is the Danish market. This power market achieved more flexibility by joining the Nord Pool market, as well as by the use of a fast market design, combined heat and power (CHP) plants and negative prices. This has allowed the Danish market to develop equitable rules for curtailment during periods of excess renewable generation (Cochran et al., 2012).

The adjustment of market rules is a difficult process as stakeholders may have different economic interests. In this context, there is an important interest in improving wind power forecasts and quantifying their uncertainty as part of an important measure to integrate wind power into the grid, avoiding imbalances and reducing curtailment. Nowadays, there is an increasing demand in the development of forecasting tools that can meet the needs of end users such as transmission system operators (TSO), energy traders, power producers and utilities.

One of the major issues that the wind power generation industry is dealing with is ramp events, which are sudden and large changes in wind power that are particularly difficult to forecast. An unforeseen ramp event can leave a grid operator scrambling to balance supply and demand. The demand from grid operators is clear: "there is a need to produce forecasts designed to predict significant swings

in wind power generation and give operators better tools to aid in balancing these swings in the real-time market or scheduling reserves in the day-ahead market" (Francis, 2008).

The aim of this research is to investigate potential approaches to improving the characterisation of wind power ramp events. This is done by the integration of numerical weather prediction (NWP) models with computational intelligence techniques. This thesis explores various weaknesses of NWP models and studies how these can be addressed to improve ramp characterisation. The rest of the chapter is organised as follows: Section 1.2 presents an overview of the thesis; in section 1.3, the aims of this research and the proposed objectives are presented; the main contributions of this work are presented in Section 1.4 and, finally, Section 1.5 outlines the remaining chapters.

1.2   Overview of the Research

In our modern times, predictions are essential for making decisions in our every day life. However, predictions are not useful if they are not specially tailored to the application. Having said this, in order to provide a wind power forecast of good value, we need to consider the specific needs of grid operators and power producers. Short-term forecasts (from 6 to 48 hours into the future) have been of great interest in the last decades as the process of deciding the sources that will supply the demand takes place one day before the actual grid operation. There has already been great progress achieved in the development of forecasting tools. For the short-term, the majority of them would require two major steps: one meteorological stage where wind speed is predicted at a specific location and a second step where the wind speed prediction is converted to a power prediction. State-of-the-art forecasting tools are able to provide not only a *point forecast*, which is an exact numerical value of the power output at a point in time, but an estimation of the uncertainty of that forecast. This is specially of great use as probabilistic forecasts are being proved to be more useful than point forecasts. Most state-of-the-art forecasting tools are based on combinations of NWP models and statistical methods, as it has been demonstrated that, for horizons larger than 6 hours, the modeling of the atmospheric flow is essential. These NWP models

run at either very high spatial resolutions to provide forecasts very close to specific locations or run in "ensemble" mode at a much lower resolution. In both cases, this process is highly demanding of computational resources. In addition, NWP models may have "misplacement" errors, which are errors in the physical location of meteorological events. This leads to the question of whether a neighbourhood approach (considering a set of locations surrounding the point of study) could improve power forecasts and the characterisation of large changes.

In addition, wind power ramp prediction is a research area which is still in its infancy. One of the major problems for ramp detection and characterisation is that there is no strict definition of what a ramp event is. It is usually referred to as a large change in wind power in a short period of time. The amount of change and period of time would need to be defined by the end user (grid operator) according to the characteristics of the power system. For this reason, most studies that have been published work on an arbitrary fixed percentage and time window. This can potentially lead to the omission of ramp events that did not meet the criteria, but that were close enough to be considered potential events, and that could be equally damaging as those that do meet the definition. There is an evident necessity for new approaches to the definition of ramp events and their characterisation.

The main research question that is to be answered in this thesis is how could numerical models be used and their weaknesses addressed to improve the characterisation of ramp events. In particular, this study is trying to answer the following questions:

- Is it feasible to use a bio-inspired heuristic to develop models to downscale numerical wind speed predictions to exact locations?

- Can the same approach be used to model the relationship between numerical predictions and actual power output of a wind farm?

- Could misplacement errors of numerical models be addressed with a neighbourhood approach?

- Are there other ways to improve the detection of ramp events using the binary definition (ramp or not a ramp based on a percentage)?

- How could the crisp definition of ramp event be avoided in order to identify potential events that do not meet the binary definition but that could be equally dangerous to the stability of the grid?

To answer the above questions, the aims and objectives are outlined below.

1.3   Aims and Objectives

The aim of the work in this thesis is to investigate potential ways to improve the characterisation of wind power ramp events by means of computational intelligence techniques. The main idea is to be able to provide not only detail of duration, rate and timing of events, but to be able to identify those events that might be omitted by common ramp detection tools. This information could allow the grid operator to take informed decisions of what type of backup units are needed according to the characteristics of the ramp. This could bring several benefits, such as the adequate allocation of backup units to avoid blackouts and cost reductions by avoiding the allocation of unnecessary reserves.

In order to accomplish the aim of this research, the following objectives are identified:

- Use mesoscale NWP models as the main source of information in order to predict wind speed at specific locations.

- Study the feasibility of applying Genetic Programming (GP) as a means to model the relationship between wind speed predictions and wind power predictions of a wind park. The main idea of using GP as a regression tool is that this technique does not make any assumptions about how the model should be, making it site independent in that sense. Also, it provides a mathematical representation of the model, providing information that could improve the understanding of what is happening at the location of the study.

- Investigate the possibility of applying the GP algorithm at different neighbour locations and create "ensembles" in these set of forecasts. The idea

behind this is that, if there is a misplacement error in the numerical forecast, by looking at different neighbour points, one could estimate the possible power outcomes as if the location of the park was in each of these neighbour points.

- Study a probabilistic approach to define ramp events by means of how they develop in short windows of time.

1.4   Major Contributions of the Thesis

The main contributions of this thesis are:

- The development of a novel approach for wind power forecasting based on Genetic Programming and NWP models.

- The creation of ensembles using GP in a neighbourhood approach to help determine the possible outcomes of wind power output of the farm and also to estimate the timing of ramp events.

- Investigate and determine wind power prediction intervals by using Quantile Regression Forests and numerical predictions as explanatory variables.

- Determine if the consideration of the distribution of the error of wind power predictions could improve ramp detection when using a binary definition.

- Provide a probabilistic prediction of possible ramp events and present these results in a way that could be easily interpreted.

1.5   Thesis Outline

This thesis consists of eight chapters that are summarized as follows:

Chapter 2: Literature Review
This chapter provides an overview of the most relevant existing techniques used for wind power forecasting and ramp characterisation and their relevance in the

integration of high levels of wind power into the electric grid. The chapter begins by presenting the two main approaches to wind power forecasting. Then, it introduces different techniques used specifically for short term wind power forecasting and discusses their benefits and limitations. A review of ramp forecasting methods and uncertainty estimation is given and, finally, a discussion on the open problems arising due to the integration of wind power as an intermittent resource into the electric grid is presented.

Chapter 3: Data Collection
This chapter introduces the mesoscale numerical weather prediction model used for forecasting atmospheric variables at different locations and heights. The model has several steps and settings in order to produce a weather forecast. These settings are presented in detail. Once the model is configured, forecasts can be produced at any location using input data from a global model. With a weather forecast, variables such as wind speed, wind direction and temperature are extracted, processed and used as the main input for wind power forecasting and ramp characterisation.

Chapter 4: Genetic Programming: an Approach for Symbolic Regression
This chapter introduces genetic programming and how it can be applied for symbolic regression. Details about the algorithm and how genetic operators are implemented are given. Then, in order to validate the implementation, the approach is tested using a simple regression problem. The results obtained are presented and discussed.

Chapter 5: A GP Approach for Wind Speed Downscaling
In this chapter, the GP approach is used as a means to obtaining a wind speed forecast at a specific location from numerical wind speed predictions. The algorithm performs a "downscaling" process to refine the resolution of the mesoscale model grid to the exact location of the observation point. Different sets of experiments were designed to find the best settings of the algorithm to achieve a wind speed forecast with the smallest error possible.

Chapter 6: Wind Power Forecasting with Genetic Programming
This chapter presents two approaches to wind power forecasting. The first one consists in taking the raw wind speed forecasts from the mesoscale model grid directly to predict the power output of the farm in one step. The second one is a two step approach which consists in first forecasting wind speed at a specific location (downscaling) and then feeding the downscaled wind speed forecasts into a wind farm power curve model, which is also found by GP. Finally, both strategies are compared and the results are discussed.

Chapter 7: Identifying Large Variations in Wind Power
In this chapter, an approach to the characterisation of wind power ramp events is proposed. First, a benchmark algorithm is presented using the typical binary definition. This basic detection algorithm is improved by taking into account the forecasting errors. Furthermore, using a filtering process and fuzzy rules, changes at different time scales are analysed and characterised, providing a probabilistic forecast of the potential ramps. The results are presented and discussed.

Chapter 8: Conclusions and Future Work
This chapter provides the conclusions that arise from this thesis and formulates some future research for the integration of wind power forecasts into the decision making problems in the electricity market.

Chapter 2

Literature Review

2.1 Introduction

Renewable energy (RE) has become an attractive topic for politicians and researchers as they look for options to reduce CO_2 emissions which are contributing to issues such as global warming. Emissions produced by conventional power generation (e.g. coal-fired power plants) account for about one quarter of the total CO_2 emissions that are released into the atmosphere (Cassedy and Grossman, 1998; of Global Development, 2007). Wind power has had the strongest growth over the recent years (World Wind Energy Association, 2012) compared to other renewable sources. This might be motivated by various factors. This energy resource has no fuel cost as it depends basically on the wind. In addition, the main mechanism to create electricity with wind is a rotor turning a generator, which is a relatively mature and well understood technology. The amount of installed capacity is increasing and researchers, private companies and electricity market policy makers have made a great effort to understand what are the impacts in terms of security and reliability of introducing large quantities of this RE into the grid and how this integration can be carried out.

As with any other power generation process, wind power generation has its disadvantages. The inherent intermittency and variability of wind generates uncertainty about the real production of a wind farm. This chapter will present an overview of the different aspects that have been studied and technologies that

have been developed as an attempt to integrate wind power into the electricity market. The first section presents the concept of wind power in more detail and its challenges. Then an overview of the different aspects that need to be addressed in order to achieve integration is presented, followed by the background to the first solutions and current solutions under development. Throughout this overview, different areas that need further study are highlighted to introduce the reader to the main points that will be addressed in this thesis.

2.2 Wind Power and Its Challenges

In order to understand what the challenges of wind power are, it is important to establish first how wind power is obtained. The main energy source of wind power is the wind energy flux. Wind is derived from the sun, which heats the Earth unevenly, creating pressure differences in the atmosphere. These pressure differences cause the air to move in an attempt to restore equilibrium in the pressure, moving from places with high pressure to places with low pressure. At the same time, the air experiences forces from other sources such as the *coriolis* force (Metoffice, 2014).

The movement of the air can be on a small or large scale. Small scale winds, like sea breezes, happen due to the difference in air temperature on land and sea. As the sun heats both land and sea, the sea is able to absorb some of the heat, keeping the air at a lower temperature and higher density, in contrast with the land, which warms up, heating the air, and hence decreasing the pressure. Large scale winds occur due to the difference in temperature between the poles and the equator. The heated air rises at the equator while the cold air in the poles sinks. This difference in pressure sets up a global wind circulation as the cold air from the poles tend to move to the equator. These winds are also affected by the rotation of the earth, pulling them to left or right depending on the side of the hemisphere they are heading to.

Wind exists everywhere on the Earth and has been widely used in the past for mechanical power and transportation, so the potential of this source of energy has been well known for years. Some examples of this are the use of sails in vessels which goes back to the the ancient Egypt and Mesopotamia, and windmills,

which were used mainly for milling grain and pumping water for consumption. These uses started to decline with the invention of the internal combustion engine. However, technological development also brought a new conception of the windmill, which is the wind turbine. Some early electricity-generating wind turbines had some features of windmills incorporated into them. The wind turbine can extract power from the wind in the same way as a traditional windmill but uses this to produce electricity. As in traditional windmills, wind turbines can only produce energy in response to the resource that is immediately available. Wind cannot be stored to be used at a later time. The output of a wind turbine is thus inherently fluctuating and non-dispatchable, making this source of energy unreliable for the electricity market (Manwell et al., 2010).

It was not until several factors were put together that wind power started to emerge as a potential source of renewable energy. The awareness of the possible depletion of fossil fuels, the availability of technology, and finally, the emerging political support, which started in the United States, Denmark, Germany and exists now in much of the rest of the world (Manwell et al., 2010), led to the current development of this industry, which is now a reality. Wind turbines are nowadays installed mainly in large conglomerates or *wind farms*, as larger wind turbines are more cost effective, providing bulk power to the electrical grid.

While conventional power plants, such as thermal and nuclear power stations, have a reliable production capacity per hour (except when exceptional breakdowns happen), a wind farm will only work at its maximum capacity if certain wind conditions are met. Wind turbines need a minimum wind speed to start power production. This minimum wind speed could be different for different turbine models. As wind speed increases, so does the power output of the turbine. In most wind turbines, there is a range where the power output can remain constant. Outside that range, a higher speed could damage the turbine, leading to having to turn it off and halting the power generation process. This speed is referred to as the cut-off speed, and is usually at 25 m/s for most modern turbine models. This relationship between wind speed and power output can be represented with a *power curve* (Figure 2.1). Different turbine models have different power curves, based on an idealised installation which may not be reproduced in practice. It has been found that turbines might not work entirely at their capac-

ities due to a correlation with the roughness and orography of the terrain where the turbine is located (Kariniotakis et al., 2004). This means that if one needs to quantify the generation capacity of a turbine/wind farm, a site-dependent model would be preferable. This requires a considerable amount of effort for tuning the models, and here the experience of who performs the installation always makes a difference (Giebel et al., 2011).
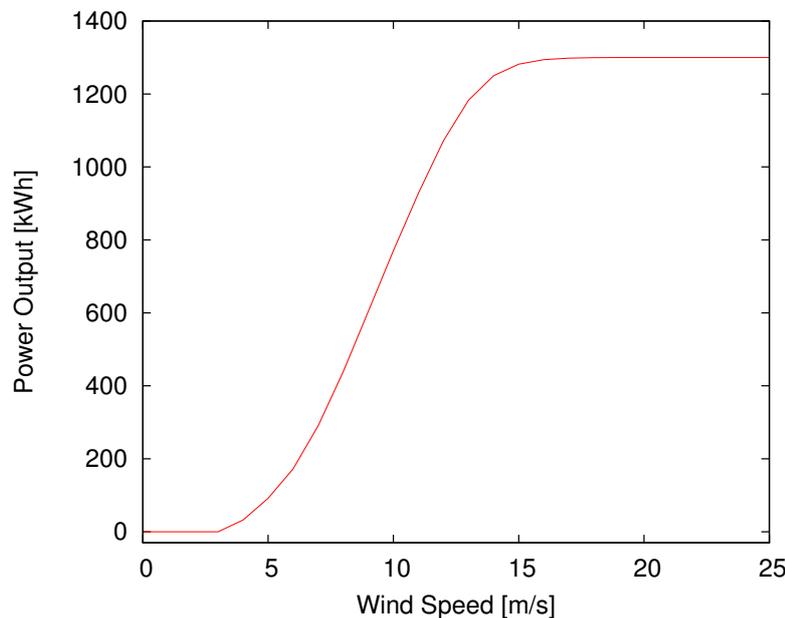
Figure 2.1: Example of a power curve from turbine model Izar-Bonus 1.3 MW.

Utility grids operate at a constant frequency to allow stability and interconnection with other utilities. This frequency is achieved if the total amount of power generated is equal to the amount of power consumed. The sudden halt of a wind turbine and the continuous fluctuations of the wind speed would translate into power output fluctuations, which can impact the frequency of the power system the turbine is connected to, as well as the interconnection tie-line schedules. As wind turbines are usually installed in conglomerates, a sudden change in wind speed can affect a large number of turbines, hence affecting considerably the power output. Frequency that is too high or too low from its nominal level can cause varying issues and in the most drastic instance lead to load shedding (intentional power shutdowns) or instability in the grid (Ela and Kemper, 2009).

In current electricity markets, fluctuations are addressed already as random changes in demand may occur. Fluctuations are managed by using backup generating units which can be started quickly and synchronised to the grid to balance the fluctuations. Different back up units respond to different speeds and the type of backup to use will depend on the type of fluctuation. Adding the wind power fluctuations raises the level of complexity of the balancing task. If an increase in demand happens at the same time as an increase in wind power, then the change in power output could be easily fed-in to compensate the demand. However, when these two events do not complement each other, either wind power is curtailed or, in a shortage of wind power, backup resources need to be quickly started up.

With the inherent variability and intermittency of the wind, it is not possible for an independent system operator/transmission operator (ISO/TSO) to treat this source of energy as a conventional power plant. If this source is to be integrated into the electric grid as any other source, it would need to be considered during the operational procedures of the electricity market. These procedures require taking decisions based on future demand and production. In order to provide information about the future production of a wind farm, a *forecast* is needed. Being able to predict how much power can be produced at a wind farm has become more important as the level of wind power penetration on the grid increases. With higher levels of wind power in the grid, there is a higher risk of intermittency and instability due to large forecasting errors. These large forecasting errors are likely to occur during the so-called *ramp events*, which are large changes of power output (increase or decrease), if these were not well forecasted.The more accurate and detailed the forecast and characterisation of future ramp events, the more prepared the ISO/TSO will be as the real time operation approaches.

2.3 Electricity Market Operations

Most electricity markets are designed to perform a set of procedures as part of the short-term operation of the grid. These procedures, also called Day-Ahead operations (DA), will allow the TSO to prepare for the real-time operation (RT) and to successfully supply the power demand (Monteiro et al., 2009). The fol-

lowing sections present the current state of the electricity market practices and the current developments in the integration of wind power into these practices.

2.3.1 Reserve Requirements

In order to run a power system in a reliable and secure manner, it is necessary to maintain a certain amount of operating reserves that can be used when there is an eventual outage of a generating unit (Monteiro et al., 2009). Part of the DA procedures include the allocation of these reserves, also known as ancillary services, and the amount will depend on the regulations of each market. Operating reserves are typically categorised depending on how quickly they respond. Regulating reserves, for example, respond immediately to generation adjustment needs. Contingency reserves need to respond within 10 minutes and are used to respond to contingencies such as forced outages of generators or transmission lines. These contingency reserves can be either spinning reserves, which are obtained by increasing the power output of generators that are already connected to the power system, and non-spinning reserves, which is extra generating capacity that is not currently connected. Figure 2.2 shows the time frame of use of reserves by type. For example, regulation reserves can be started within 1 minute but are also used for a short period of time. If the reserve needs to respond for a longer period, then spinning or non-spinning reserves would be used. It will depend on the rules of the market and the operator to decide when these reserves would need to be replaced by supplemental reserves or by ordinary power suppliers.

The type of reserves to allocate will not only depend on the time of response but on the availability of these types of units in the system. The more flexible the system is, the more variable generation (VG) like wind and solar power can be integrated into the grid. Pumped storage (PS) power plants are the most responsive technology and can generate electricity almost instantaneously (Eurelectric, 2011). Their load gradient (nominal output change rate in a given timeframe) is the fastest, as they can ramp up and down by more than 40% of the nominal output per minute. Nuclear power plants (NPP) have the second fastest load gradient, which make them a good option to perform load-following operations (units that can adapt to the power fluctuations) if they are already in operation.
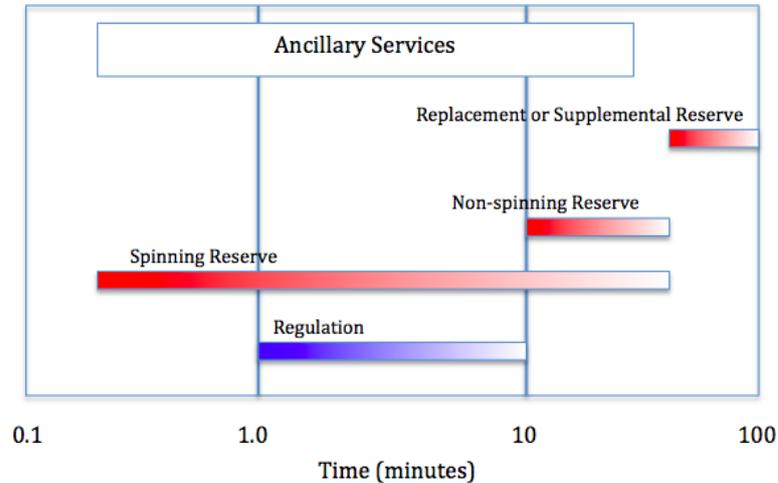
Figure 2.2: Ancillary services (reserves) by time frame of use.

However, they cannot be brought online from a cold or warm start as quickly as PS. Combined cycle-gas power plants (CCG) are also suitable for load-following operations as they have fast load gradients (4%/min) and can be brought online fairly quickly (less than 1.5 hours from warm conditions). Coal-fired power plants are less responsive than the previously mentioned technologies, although newer plants are more flexible than older units. Hard coal-fired and lignite-fired plants have similar load gradients, but the former are faster to respond to load changes from cold and warm start-up conditions. Hydroelectric power plants can operate as base load, load following or peaking power plants (plants used only during peak times). These plants can be started within minutes, and in some cases in seconds. The way the power plant is operated will depend on the amount of water that is available. Gas turbine power plants are very flexible units as their power level can be adjusted very quickly; however, they are the most expensive to operate. This is one of the reasons this kind of plant is used as a peaking unit. Figures 2.3 and 2.4 show the flexibility of these technologies in terms of start-up times and load gradient. Start-up times are presented for both cold and warm starts. Figure 2.4 shows the load gradient once the unit is already in operation. Operating units that provide load-following would not be at a 0% load as shown in the Figure, but the purpose of the figure is to show how quickly they can potentially get to a 100% load.
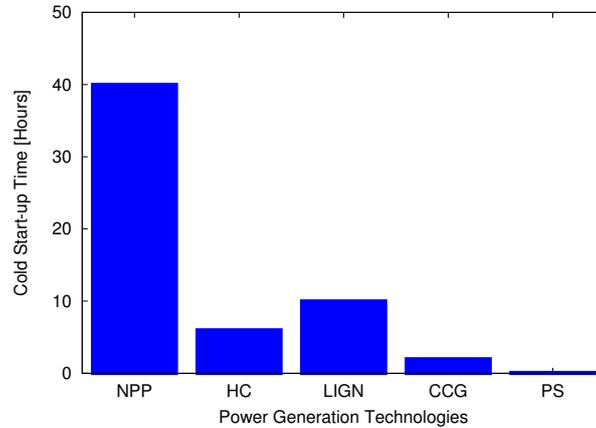
(a) Start-up time on "cold" conditions
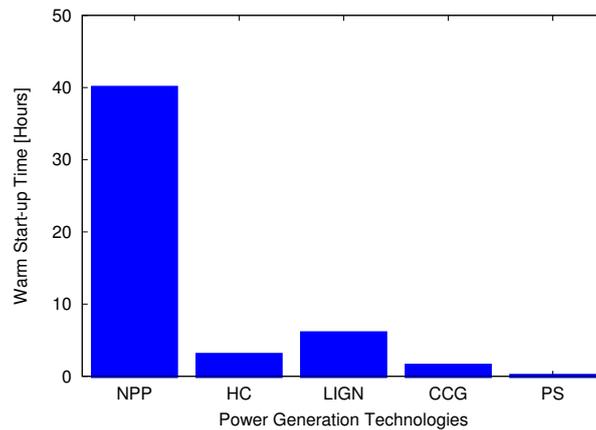
(b) Start-up time on "warm" conditions

Figure 2.3: Start-up times on cold and warm initial conditions for different technologies: nuclear power plants (NPP), hard coal-fired power plats (HC), lignite-fired power plants (LIGN), combined cycle gas-fired power plants (CCG) and pumped storage power plants (PS).

These types of ancillary services are currently used to cope with the demand-driven fluctuations. The problem with generation-driven fluctuations (such as those from wind energy) is that, during times of ramp events, the velocity of the fluctuations is so high that only storage facilities such as pumped storage and hydro storage schemes with peak generation can cope.

Over the recent years, the integration of intermittent renewable energy has prompted additional demand for reserve and response operations. Power markets
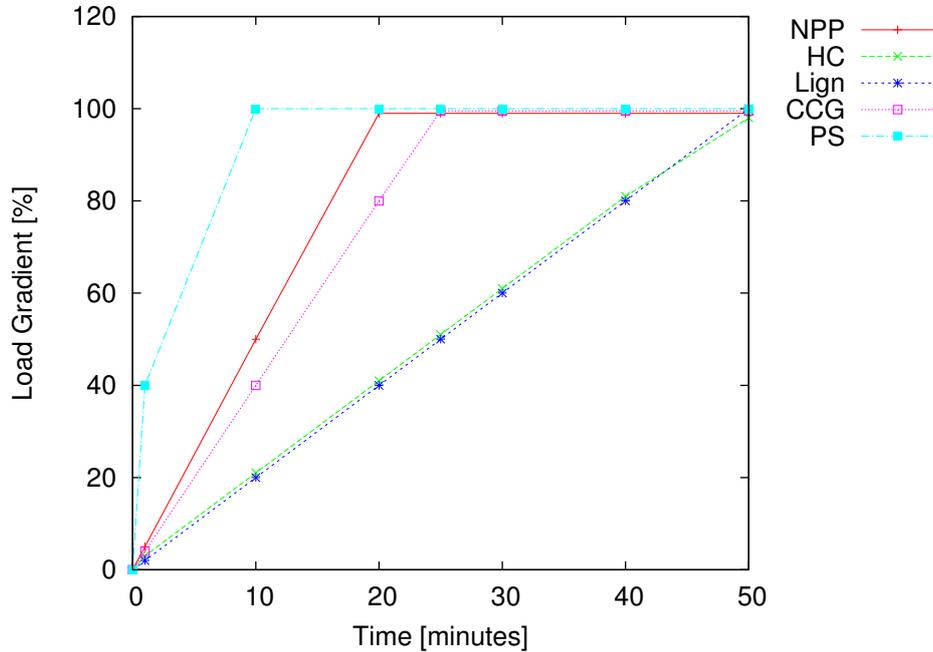
Figure 2.4: Load gradient for both upward and downward directions of different technologies once they are in operation. Figure based on the data reported by EURELECTRIC on the EU generation mix (Eurelectric, 2011).

that are unable to adapt to power inttermitency during the day need large volumes of real-time balancing reserves. This results in additional costs due to increased start-up and part-load costs to provide balancing power (Borggrefe and Neuhoff, 2011; Energinet, 2010). The better the changes in wind power are characterised, the better the selection of the appropriate back up generation will be. That is why it is not enough to only predict the changes, but to characterise how these events will develop, especially if the changes are rapid and large.

2.3.2 Unit Commitment and Economic Dispatch

As part of the DA procedures, market participants, from both the demand and supply side, submit their bids to the ISO/TSO by a certain deadline. This deadline varies according to the market. The bids will provide information about how much energy and operating reserves they can provide for the following day, as well as their constraints (ramping rates, start-up cost/times, minimum down-

time, etc.) (Monteiro et al., 2009). The clearing of the DA market will then consist of two procedures. The first step is a *unit commitment* (UC), which will commit resources of different providers, trying to minimise the operating costs, while meeting the total bidded demand and constraints. The second step is to run an *economic dispatch* (ED) algorithm. This procedure consists in determining the amount of power each of the scheduled generating units will produce at each hour, taking into consideration constraints of the transmission lines. After the generating units are dispatched and as the operating day approaches, the ISO/TSO will perform a re-commitment procedure in case of any forced outage and will take into consideration the forecasted load.

2.3.3 Wind Power Forecasting in the Market Operations

In most markets, intermittent resources are not bidding in the DA market, they are bidding only in the RT market, leading to large curtailments of wind power (Bessa et al., 2014). Wind power forecasting can provide important information to several of the grid operation procedures. As mentioned previously, the allocation of reserves has increased as a result of the high uncertainty of wind energy. A wind power forecast and uncertainty estimation could potentially be used to improve the process of determining the operating reserve requirements. The Electric Reliability Council of Texas (ERCOT) was one of the first markets in the USA to consider wind power uncertainty forecasts in its reserves determination process. ERCOT defines the non-spinning reserve as the 95th percentile of the observed hourly net load error from the previous 30 days, plus the size of the largest generation unit. The Spanish system operator Red Eléctrica de España (REE) defines the balancing reserve requirements as the sum of the generation shortage/surplus due to load and wind generation historical forecast errors and unplanned outages (Bessa et al., 2014).

Doherty and O'Malley (2005) presented an analytical approach to quantify reserve demand. In order to determine the required amount of system reserves, the authors take into consideration the generator outages and load and wind forecasting errors. This study finds that the amount of reserve must be increased if the wind power capacity increases in the system. Ortega-Vazquez and Kirschen

described a method for minimising the sum of the expected cost of energy not served and the operating cost (Ortega-Vazquez and Kirschen, 2009). Wind power and load forecast uncertainty is assumed to follow a normal distribution, which is highly questionable as several studies have demonstrated the high skewness and kurtosis of wind power forecast errors (Pinson, 2006). Matos and Bessa (2011) developed a reserve management tool to support the TSO in the online definition of operating reserve needs. Their probabilistic method does not assume normal distribution of the errors and is able to calculate risk/reserve and risk/cost curves that allow the TSO to make decisions on the amount of reserves. However, the authors rely on the quality of the probabilistic forecasts (obtained with prediction intervals), which, as is stressed in their work, is not evaluated. As later addressed in detail, large sudden changes in power output might not be placed on the correct physical position. The uncertainty estimation could be potentially improved by taking into consideration numerical weather prediction spatial fields (Cutler et al., 2009).

There has been some work on integrating wind power forecasting and uncertainty into the UC as well. Barth et al. (2006) proposed a stochastic model of the unit commitment problem, where the uncertainty of wind power is taken into account, to study the impact of large amounts of wind penetration on electricity prices in the area covering the following countries: Denmark, Finland, Germany, Norway and Sweden. They concluded that the integration of wind power has impacts on the resulting prices in the electricity markets. Generally, the electricity prices decline during hours with high wind power feed-in. During the hours where the wind power is able to cover the total electricity demand and the transmission capacities are working at full load, the prices become equal to zero.

Wang et al. (2008) propose a Security Constrained Unit Commitment model where the intermittence of individual wind farms is considered in order to ensure that prevailing constraints are satisfied. The authors found that iterating between the master unit commitment problem and wind power scenarios, used to model uncertainty, could identify a robust unit commitment and dispatch solution for accommodating this volatility.

Tuohy et al. (2009) did a more exhaustive study of the advantages of using a stochastic model over a deterministic model. They concluded that more frequent

scheduling of the system means wind and load forecasts are being updated more often and more of the uncertainty of the wind is captured in the model. This means that more of the costs caused by uncertainty will be minimised, leading to more optimal results and better performing schedules.

None of these previous stochastic optimisation studies present details on the wind power forecasting model and uncertainty used to support their conclusions. Constantinescu et al. (2011) developed a computational framework to integrate the state-of-the-art NWP model Weather Research and Forecasting (WRF) to incorporate wind uncertainty in stochastic unit commitment and economic dispatch procedures. The use of physical models is desirable because consistent and accurate uncertainty information can be obtained (Monteiro et al., 2009). On the other hand, one of the major limiting factors of using NWP is their computational complexity. The studies done in Constantinescu et al. (2011) indicated that using WRF forecasts and uncertainty information is critical to achieve high adoption levels with minimum reserves.

In general, most methods model uncertainty by using a set of scenarios. According to Bouffard and Galiana (2008), the reserves in the UC are defined internally and there is no need to specify *a priori* a reserve requirement. However, there are other studies which include constraints related to reserve requirements to model uncertainty explicitly in a deterministic UC problem (Restrepo and Galiana, 2011; Ruiz et al., 2009). Wang et al. (2011) compared different deterministic and stochastic UC strategies that deal with wind power uncertainty. Results showed that when wind power forecasts were not considered in a deterministic UC, the algorithm tends to overcommit conventional units. Results also showed that when point forecasts were included, the commitment led to the highest cost and load curtailment. It was also found that stochastic UC had relatively low cost, but deterministic UC with a reserve rule obtained a similar cost. Although the deterministic UC yielded to similar results when incorporating reserve rules, the use of scenarios on the stochastic UC could capture with more detail the ramps between hours and consequently compute a suitable reserve for dealing with temporal variability. The problem with scenarios is that solving the UC with these is expensive so the number of scenarios need to be reduced. In general, both constrained UC and stochastic UC have their advantages; however, there

is still a gap between what has been proposed in research and what is actually implemented in the electricity markets.

In order to incorporate probabilistic approaches for reserve requirements and UC, the electricity market must be redesigned. The smart grid project, which is looking to improve the reliability, security and efficiency of the grid, could incorporate these approaches in order to improve the integration of wind power, as well as other intermittent resources.

The need for wind power forecasts is a reality now. As has been noted, an integration on the DA market needs a forecast to be provided 36 to 48 hours before the RT operation. This forecast horizon is commonly referred as the *short-term* horizon. There has been a significant development in short-term forecasting over the last decade. The following section presents this in detail as well as the challenging open areas which are still under development.

2.4   Short-term Wind Power Forecasting

A wind power forecast made at time t for the look-ahead time $t + k$ is the average power $p_{t+k|t}$ the wind farm is expected to generate during a period of time (e.g. 1 hour) if it would operate under constant wind (Monteiro et al., 2009). The time resolution of a forecast is denoted by the time step k. Usually for horizons from 24 to 72 hours into the future, the time step is hourly. An exact amount of power at time $t + k$ is referred to as a *point* forecast and is denoted as $\hat{p}_{t+k|t}$.

A wind power forecast is characterised by the time horizon, which is the future time period for which the wind generation will be predicted. This horizon can be separated into three categories: *very short-term*, which ranges between a couple of hours up to 4/9 hours (depending on the market); *short-term*, which ranges from the very-short term up to 48/72 hours into the future; and *medium-term*, which ranges from the short-term up to 7 days (Monteiro et al., 2009).The short-term horizon is mainly interesting for trading in the day-ahead market, and is the horizon of interest in this research.

2.4.1   Single Time-Series Forecasts

Time-series forecasts are predictions containing a sequence of point forecasts that describe one possible future scenario. This scenario is treated as the most likely scenario or the one that minimises certain error metrics. The time-series forecast is the most common way to present forecast information and the most frequently used in the literature (Ernst et al., 2007). There are also other ways of representing forecasts, which are mostly complementary, such as the time-series forecast of a risk-index related to the likelihood of large instantaneous errors (Pinson and Kariniotakis, 2004), prediction intervals, ensembles (Pinson, 2006) and scenarios (Wang et al., 2011). These will be addressed in detail in the following sections. A typical time-series power forecast is presented in Figure 2.5.
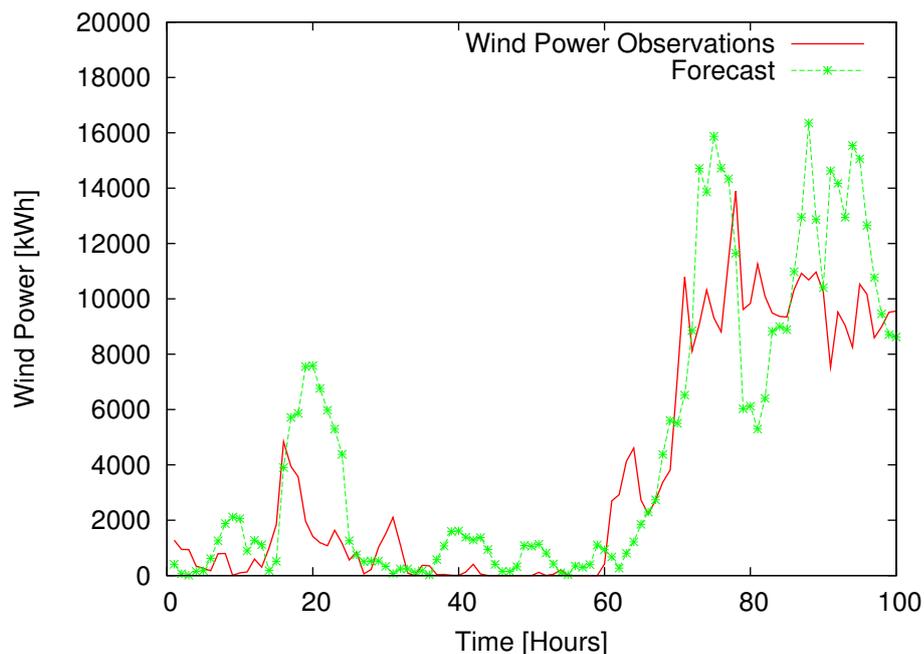
Figure 2.5: Example of single time-series forecast.

A single time-series forecast can be evaluated for its ability to minimize a certain metric if the evaluation is made over a period of time that is long enough to ensure that the most likely scenario actually occurs sufficiently often. None of them actually evaluate the forecast for its ability to predict the more likely scenario, which is a very difficult task as the most likely scenario is not known.

In general, a prediction error for look ahead $t + k$ is defined as:

$$e(t + k|t) = P(t + k) - \hat{P}(t + k|t) \tag{2.1}$$

where $\hat{P}$ is the predicted power at time t and P is the actual power output. This error is often normalised when compared among different wind farms.

The model bias (BIAS), one of the common error measures used in the literature, corresponds to the systematic error and it is estimated as an average error over the whole evaluation period. Its formula is presented in Equation 2.2.

$$BIAS(k) = \frac{1}{N} \sum_{t=1}^{N} (e(t + k|t)) \tag{2.2}$$

The Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) are the two basic criteria for illustrating the performance of a prediction method (Equations 2.3 and 2.4). Both systematic and random errors contribute to these criteria.

$$MAE(k) = \frac{1}{N} \sum_{t=1}^{N} (|e(t + k|t)|) \tag{2.3}$$

$$RMSE(k) = \left(\frac{1}{N} \sum_{t=1}^{N} (e^2(t + k|t)) \right)^{\frac{1}{2}} \tag{2.4}$$

Another error measure is the Standard Deviation of Errors (SDE), which deals with the random part of the error. The SDE is expressed as follows:

$$SDE(k) = \left(\frac{1}{N - (p + 1)} \sum_{t=1}^{N} ((e(t + k|t) - \bar{e}_k)^2) \right)^{\frac{1}{2}} \tag{2.5}$$

where p is the number of estimated parameters using the considered data and $\bar{e}_k$ is the bias or systematic error. For the test data, $p = 0$.

Madsen et al. (2005b) studied the use of different error measures to analyse the performance of predictions. They found that the accuracy of a forecasting method was influenced by the characteristics of the site and the time covered by the test set. They suggested that, in order to measure performance, a minimum

of three measures should be used (normalised BIAS, MAE and RMSE) and that a base performance evaluation should be done on the test set only.

There has been a substantial amount of work to improve time-series forecasts, which will be addressed later in this chapter. According to Giebel et al., there are two main types of errors in a time-series forecast, the amplitude errors and the timing (phase) errors (Giebel et al., 2011). A study presented in AWS Truewind (2008) shows how timing errors can occur in a time-series forecast for several hours. A current strategy to overcome errors is to estimate the uncertainty of a forecast. Most of the work published for handling the timing and amplitude errors use probabilistic forecasts, such as scenarios and prediction intervals, generated either by the error distribution of the point forecast or by the use of ensembles.

2.4.2 Reference Models

Due to the high influence of the location of the wind farm on its actual power production (flat, complex terrain, offshore), it is not possible to compare prediction systems based on available results. Some reasons for this is that a standard form of measure is still not adopted and it is very important that the data is exactly the same (Costa et al., 2008). Persistence is one of the models that has been widely used as a benchmark or reference. Its key features are:

- It is simple. It assumes that the wind (speed and direction) or power at a certain future time will be the same as it is when the forecast was made.

- It performs very well in the first 6 hours.

- It is difficult to beat in the very short term.

Persistence can be expressed as $\hat{p}_{t+k|t} = p_t$. Any method that outperforms Persistence is worth implementing (Monteiro et al., 2009).

Another typical model used for comparison is Climatology, which has been designed to capture the average hourly diurnal cycle for present weather regimes. This method has the following features:

- It is a simple approach, similar to Persistence.

- It is used for day-ahead prediction.

- The method involves using the average of weather statistics accumulated over past years.

The Climatology method only works when the weather pattern is similar to that expected for

2.4.3   Probabilistic Forecasts

Probabilistic forecasts have been developed as a means to address the estimation of uncertainty around point forecasts. They provide a level of confidence in a given forecast scenario and may provide multiple possible scenarios with associated probabilities. A probabilistic forecast can be either calculated from a single time-series forecast by understanding the error of the forecast model in past data, or can be obtained using ensemble forecasts, which will be addressed in more detail later in this chapter.

There are three main representations of the uncertainty of wind power forecasts; probabilistic forecasting, risk index and scenario forecasting (Zhang et al., 2014). A probabilistic forecast is the most commonly used uncertainty representation. It can be expressed by probability measures such as probability density functions (PDF) (Juban et al., 2007), quantiles and intervals (Khosravi and Nahavandi, 2013; Wan et al., 2013), discrete probabilities and moments of probability distribution (e.g. mean, variance and skewness). The probabilistic forecast is often visually presented as time-series prediction intervals around a single-time series forecast (Bremnes and Villanger, 2002; Lange et al., 2006; Pinson, 2006; WEPROG, 2007). An example of a time-series probabilistic forecast in shown in Figure 2.6.

In order to assess the quality of a probabilistic forecasts, there are three basic measures used in the literature. The first one is *reliability*, which measures the agreement between the probability of a forecast and the mean observed frequency of an event. The second one is *sharpness*, which is defined by Pinson (2006) as how narrow the probabilistic distribution is. The narrower the distribution, the more valuable the prediction in a decision-making context. The third one is *resolution*,
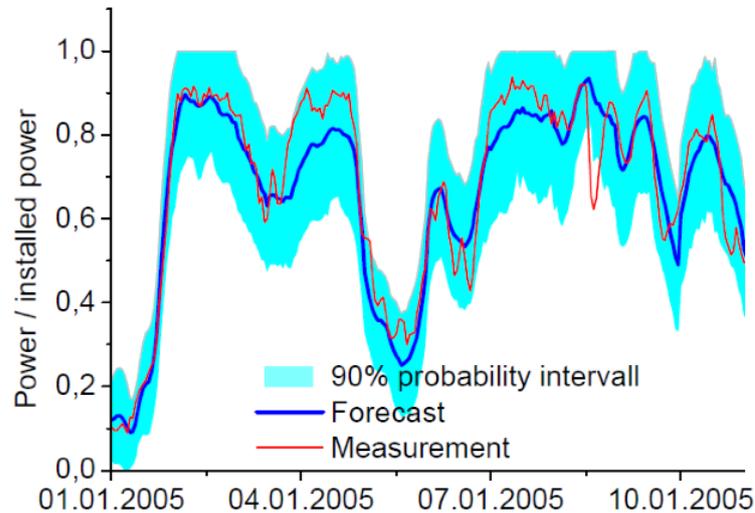
Figure 2.6: Example of a probabilistic forecast from the work of Lange et al. (Lange et al., 2006). Permission to reproduce this image has been granted by the author.

which is defined as the ability of providing situation-dependent assessment of the uncertainty. Pinson suggests the use of a unique measure that could give all the information on a given method performance. This is called the unique skill score, which rewards tight intervals and gives a penalty if an observation does not lie inside the estimated interval.

An example of a reliability diagram is presented in Figure 2.7. Deviations from the "perfect reliability", which is the case when the empirical coverage equals the nominal one, are given as a function of the quantile nominal proportions. The deviations shown in the figure are less than 2%, which is considered a good calibration.

2.4.4 Approaches to Short-term Forecasting

In order to design a wind power forecasting system, one needs to have a clear idea of what the end application will be, as this defines the type of output that is required. Each end user may have a different set of detailed requirements and these may be defined by the features of the power system, e.g. type of generating units and ramp rates, and market rules. It is also important to have knowledge
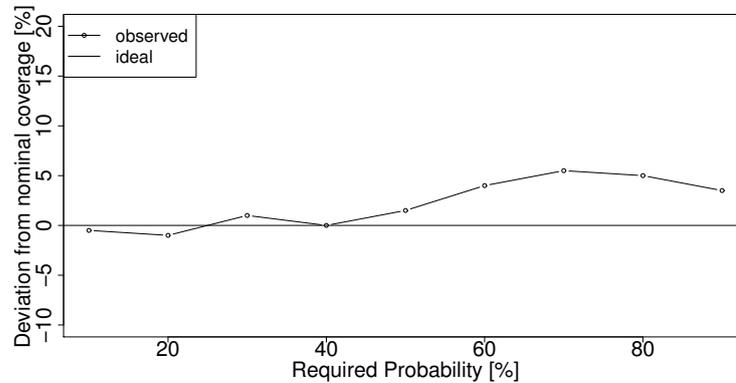
Figure 2.7: Example of a reliability diagram.

of the limitations, e.g. data availability, as these have a direct impact on the accuracy quality of the system.

There are two main approaches for wind power forecasting: those that involve a Numerical Weather Prediction model and those that do not. NWP models represent the atmospheric flow by a set of physical equations, which model the dynamics and thermodynamics of the atmosphere. Variables such as temperature, pressure, wind speed, wind direction, among others, are modeled. These models can be classified into three types: global, mesoscale and local models. Global models, such as the Global Forecasting System (GFS) from the National Oceanic and Atmospheric Administration (NOAA) in the USA (Kanamitsu et al., 1991), produce low space resolution predictions of the complete globe. Mesoscale models, such as the Weather Research and Forecasting (WRF) (Skamarock et al., 2001), improve the time and space resolution of a global model within a specific area down to 1km x 1km. Finally, local models produce the highest space resolution in a limited area. Whether their inclusion is worthwhile depends on the horizon to be predicted. Typically, prediction models using NWP forecasts outperform time series approaches after 3-6 hours look-ahead time (Giebel et al., 2011). Most utilities need a short-term forecast, meaning that they need a method that is based on NWP forecasts.

The are two main schools of thought in terms of short-term prediction: the physical and the statistical approaches. The physical approach focuses on the description of the wind flow around and inside the wind farm. This approach

commonly uses the manufacturer's power curve to estimate the power output. The statistical approach emulates the relation between meteorological predictions, historical measurements and generation output (obtained by monitoring systems i.e. SCADA) through statistical models whose parameters have to be estimated from data without taking into account any physical phenomena.

Whatever the approach, the main steps for short-term power prediction are three: downscaling, conversion to power and upscaling. Variables such as wind speed and direction are obtained from NWP models for the geographical point of the wind farm or for a grid of surrounding points. This involves finding the best NWP level, which might be the wind speed at 10 m or one of the lowest model pressure levels. Once the variables are obtained from the model, they are scaled to the hub height and interpolated to the exact location of the turbine. This downscaling step is different in physical and statistical methods. The physical approach uses a meso- or microscale model, which is able to resolve scales down to tens of meters. After the downscaling process, wind is converted to power with a power curve. The easiest way is to use the manufacturer's power curve; however, it has been shown in some studies that it is advantageous to estimate the power curve from the forecasted wind speed and the measured power. Finally, if only one farm is to be predicted, the prediction process ends here. However, since most utilities need a prediction for a total area, the upscaling process is the last step. This step could involve simple summation or the use of representative farms which are then upscaled using a factor. Usually the error of distributed farms is reduced compared to the error of a single farm (Costa et al., 2008).

2.4.4.1 The Physical Approach

The physical approach, shown in Figure 2.8, consists of the use of several submodels which all together are able to provide a wind speed profile at the location of the farm. It usually starts with a global model providing a forecast at several grid points covering an area. These forecasts are then extrapolated to higher resolution grids by the use of submodels that contain the mathematical description of the physical processes relevant to the translation. The refinement of NWPs is done by considering the physical aspects of the terrain such as roughness, orog-

raphy and obstacles. A common way to do this is by using Computational Fluid Dynamics (CDF), which enables the accurate computation of the wind field at the farm location considering the terrain (Magnusson and Wern, 2001).
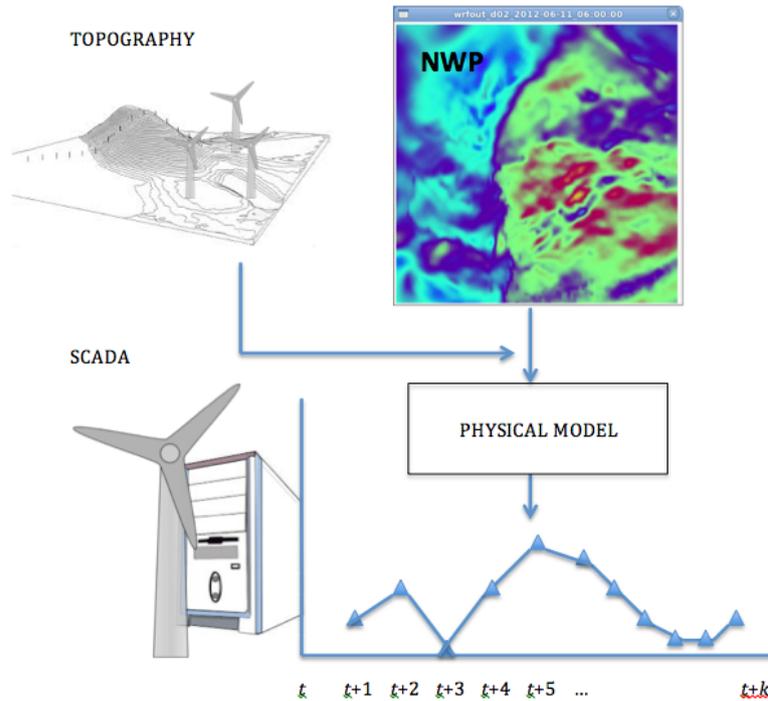
Figure 2.8: Structure of the Physical Approach.

Several NWP models have been applied for short-term forecasting. One of the first commercially available forecasting systems based on NWP was *eWind*, from the American Society TrueWind (Bailey et al., 1999). This system uses a numerical weather model called *ForeWind*, which produces accurate near-surface wind forecasts on a fine grid, using boundary conditions from a regional weather prediction model. The system uses adaptive statistics to correct the systematic errors of the wind forecasts. *Prediktor* is another NWP based method, developed at the Risø Laboratory in Denmark (Landberg, 1999, 2001). The predictions of this system are post-processed using MOS techniques to reduce the related error. Jørgensen et al. (2002) studied the coupling of a NWP model to wind power forecasts and examined especially badly forecasted days. He found that in all cases the error came from the NWP model and that higher resolutions using

HIRLAM did not improve. Enomoto et al. used the Local Circulation Assessment and Prediction System (LOCALS) to forecast the power production of a wind farm in Japan (Enemoto et al., 2001). Despite the model being used at a 500-m grid resolution, the RMSE was still at 15% of the installed capacity. The authors pointed out that the turbulence intensity between turbines was not modelled correctly. Other models have also been applied for wind speed forecasting, such as the ETA model, which has been applied for wind speed forecasting in Sweden (Lazić et al., 2010). Yamaguchi et al. (2007) compared 1km resolution Regional Atmospheric Modeling System (RAMS) model wind speeds to a lower resolution of the same model coupled with a simple transfer coefficient method. The authors were able to get the same accuracy in their results as that obtained with the high resolution RAMS model, decreasing the computational effort. Dierer et al. (2005) also found in a study using the MM5 model (Grell et al., 1994) that an increase in the horizontal resolution from 10km to 1km did not bring any large improvements. It is clear that in order for numerical models to improve their accuracy at very high resolutions, they need to be provided with the appropriate data, such as high resolution topographic data; otherwise, the physical downscaling process could be potentially replaced by some type of statistical method.

There is a chaotic nature in the atmosphere, and modeling it without any uncertainty is not possible as there is a strong sensitivity to small perturbations in the initial conditions. The way to overcome this issue is by the use of probabilistic weather forecasts. It has been suggested in the literature that the computational effort used for running high resolution models could be instead used to produce *ensemble forecasts* at lower resolution to address the uncertainty of point forecasts. Ensemble predictions correspond to multiple runs of a NWP model under slightly different initial conditions (Bourke, 2004; Toth and Kalnay, 1993), different physical parameterisations (Arribas et al., 2005) or from different time origins in the past (Giebel et al., 2003). The initial differences between the ensemble members are small, and consistent with the uncertainties in the observations. However, for several days ahead, ensemble forecasts can be quite different. If the ensemble members vary significantly, then this means there is a high uncertainty about what the weather will actually do. If the members agree significantly, then there is more confidence in modeling an event. Ensembles can be used to improve single

time-series forecasts in several ways. The simplest way is to use the mean of the NWP ensemble. It has been shown that this strategy reduces the average error compared to using a single time-series forecast (Möhrlen, 2004). More sophisticated methods use weighting to combine ensemble members. The Multi-Scheme Ensemble Prediction Model (MSEPS) is a NWP ensemble based on different physical parameterisations. MSEPS, developed by the Danish company Weprog, uses a probabilistic filter which combines long-term statistics with classical clustering methods to give a dynamic weighted combination of the ensemble members. The weightings change according to the weather conditions. Previento is another wind energy forecasting system based on ensembles and weightings developed by the German company Energy and Meteo Systems (Ernst et al., 2007). The system uses input from NWP models run by different institutions. The weighting is adjusted according to the models, which perform differently under different weather situations. NWP ensembles can also be converted to scenarios of wind power and these can be used for UC decisions (Nielsen et al., 2005). Each scenario can be treated as a sample following a predictive distribution. To convert NWP ensembles to wind power probabilistic forecasts, there are two crucial issues: first, how to estimate a wind to power curve, and second, how to estimate the PDF from the wind power ensembles. Taylor et al. (2009) applied a single and deterministic wind power curve. Nielsen proposed the logistical-shaped wind power curve based on non-linear regression (Nielsen et al., 2006). There has also been a significant effort in studying the use of Kernel Density Estimation (KDE)(Pinson and Madsen, 2009) and Bayesian Model Averaging (BMA) (Bao et al., 2010; Sloughter et al., 2010) for wind power curve estimation. Wind turbine/farm power curves are site dependent and, as for downscaling methods, power conversion methods could make use of non-parametric approaches so no assumptions of the relationship between wind speed predictions, wind power and terrain conditions need to be done, unless there is a complete understanding and mathematical representation of how the terrain and meteorological processes affect the power production of the farm.

2.4.4.2   The Statistical Approach

With the use of different linear and non-linear models, the statistical approach is able to provide a wind power forecast in a single step, also referred to as statistical block (Monteiro et al., 2009). This block, shown in Figure 2.9, combines NWP data such as wind speed, direction, temperature, pressure, and others from different model levels together with online measurements at the farm.
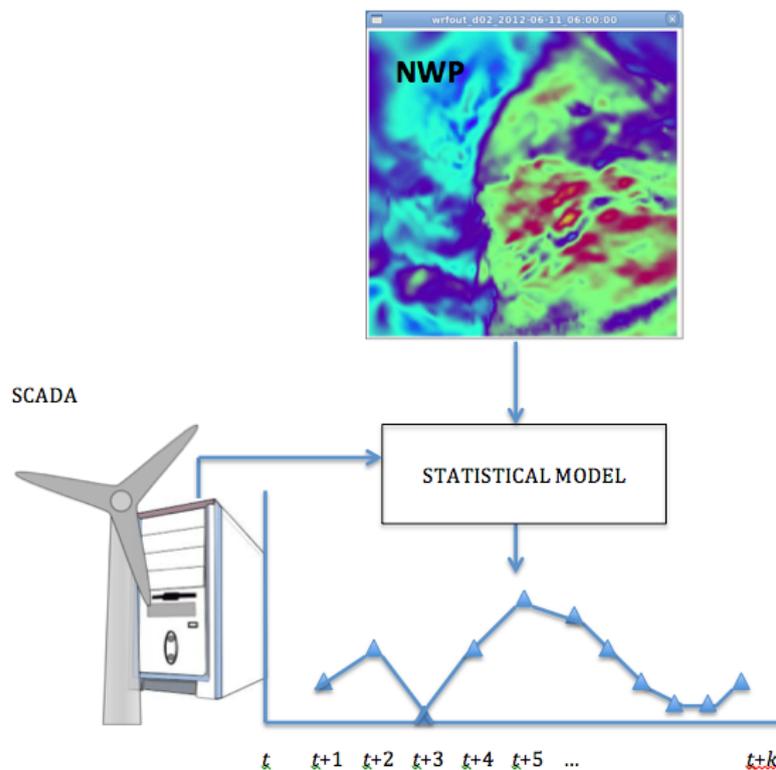
Figure 2.9: Structure of the Statistical Approach.

Not all short-term prediction models involve all steps or all types of inputs performed by physical approaches. In the early 70s, when the NWP were not widely available, the first approaches were done with time series analysis techniques. These used recently observed values of wind and other values to predict future wind speed. One of the first works published which applied time series forecasting for wind power applications was Brown et al. in 1984 (Brown et al., 1984). The authors used an autoregressive (AR) process for wind speed prediction and then

used a measured power curve to convert the predictions to power. Since then, different studies for very-short term prediction using time-series models have been developed. The most common approaches use Kalman Filters (Bossanyi, 1985), Autoregressive Moving Average (ARMA) models (Balouktsis et al., 1986; Kamal and Jafri, 1997; Kavasseri and Seetharaman, 2009; Schwartz and Milligan, 2002; Tantareanu, 1992; Torres et al., 2005), adaptive fuzzy logic and wavelet models (Kariniotakis et al., 1997, 1999), among others. Most of the methods developed were able to outperform Persistence. Nielsen et al. (1998) tried to introduce a new reference model based on a time-series model. Basically, it predicts the power $p(t)$ using the power $p(t-n)$ (being n time steps back) and the mean μ of the time series. The model is expressed as:

$$p(t) = a(n) * p(t-n) + (1 - a(n)) * \mu \qquad (2.6)$$

where $a(n)$ is the autocorrelation of the time series n steps back. This simple model can improve by 10% the error of Persistence.

Artificial intelligence (AI) based models have also been used for time series modeling. These include methods such as artificial neural networks (ANN) (Bilgili et al., 2007; Ripley, 1996; Sideratos and Hatziargyriou, 2007), fuzzy systems (Haque et al., 2012) and support vector machines (SVM) (Mohandes et al., 2004; Zeng and Qiao, 2012). Beyer et al. used a neural network for next-step forecasting of 1minute or 10 minute averages (Beyer et al., 1994). They were able to improve Persistence by 10% with a very simple network topology. Sfetsos (2001) compared different methods to forecast hourly mean wind speeds, such as the Box-Jenkins model, feed forward NN, radial basis function Networks, Adaptive Network based Fuzzy Inference Systems (ANFIS), among others. All non-linear models showed a comparable error which was better than any of the linear methods. Although ANN have been widely used by research groups as a time-series forecasting tool, the level of improvement over Persistence is not enough of a tradeoff for the effort of training such networks. Cadenas and Rivera (2007) used a NN approach to forecast monthly wind speeds at Oaxaca, México. The results were compared to a seasonal ARIMA model, the latter achieving better results. The authors suggested that an increase in the number of training vectors could improve the

results of the ANN approach. Flores et al. (2005) used a GP based strategy to perform a symbolic regression in the time series prediction problem, where the genetic program represents a combination of past wind speed data to calculate the wind speed in the near future $(t + 1)$. The authors presented different experiments varying the function set and found that the best results were obtained using basic operators, logarithmic, sine and cosine functions. The approach was able to improve the results obtained with an ARIMA model. Vladislavleva et al. (2013) proposed a GP based technique to predict the energy output of a wind farm from weather measurements taken at a close point. The forecast horizon was half an hour after the last measurements. The proposed approach was able to obtain 85.5% accuracy in unseen data using measurements of wind gust and dew point. Although the results look promising, these are only on the very short-term.

In general, a purely statistical approach has proven to outperform the Persistence model for up to 6-10 hours into the future. However, for larger time horizons, the accumulated prediction error increases. Therefore, meteorological models have to be taken into account. The atmospheric dynamics are a key piece of information when it comes to 48 hour horizon forecasts (Giebel et al., 2011).

Some of the most widely used prediction tools are based on a combination of physical and statistical methods. A classical model used in many studies is the Wind Power Prediction Tool (WPPT) (Madsen et al., 2005a). Based on advanced statistical models, this forecasting system is quite flexible as it can provide both single-farm and regional forecasts. WPPT uses self-adaptive models: this means that it is capable of adapting to changes such as the surroundings of the park, NWP model changes, even the number of turbines at the farm, disregarding old information as new information becomes available. The main inputs to the system are NWP predictions for a region and reference wind farms, power generation measurements, as well as measurements of climate variables at the farm. The meteorological forecasts used as input are not updated very frequently and the NWP models do not run at very high resolutions. Instead, the system uses interpolation methods to avoid high computational costs. In order to predict the power output of a region, the system divides the region into subareas and applies two different approaches to calculate the power output of those subareas. One approach is to select representative wind farms of each subarea and apply a model

that uses inputs from the wind farm power output as well as numerical weather predictions at the farm. The power output of the representative wind farms is calculated and then upscaled by multiplying the summarized power predictions for the wind farms in the subarea by a upscaling function which depends on wind speed and wind direction. The second approach is to use a model that links the total power production of the subarea to the numerical weather predictions. The advantage of this approach is that it uses the smoothing properties of the total production, as numerical weather models perform well in predicting the weather patterns but less well in predicting local weather at a particular farm. Apart from delivering point forecasts, WPPT provides uncertainty estimation via prediction intervals and scenarios. According to Madsen et al. (2005a), when this tool started to be used commercially in 2005, it was one of the few that provided an uncertainty forecast. Prediction intervals are calculated using ensembles provided by the European Centre of Medium-Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP), part of NOAA in the US. The WPPT started by operating in Denmark in Eltra/Energinet.dk, a SO in the west of Denmark, and in some CHP and wind farm systems. Nowadays, EFOR, the company that currently owns the tool, claims this is one of the most used forecasting tools worldwide. WPPT and Prediktor systems were unified under the Zephyr project in order to combine the advantages of both systems; however, WPPT has been the most successful one.

Another important tool, developed in Spain by the University Carlos III of Madrid for the Red Eléctrica de España (the Spanish TSO), is Sipreólico (Sánchez et al., 2002). This forecasting system is based on NWPs from the HIRLAM, NCEP and Metra (Meteorological service from New Zealand) models and using on-line data from 80% of all the Spanish wind turbines. This forecasting tool was designed to be used in the DA market as it provides hourly wind power forecasts for the following 48 hours. HIRLAM forecasts are used at a 20 km resolution for up to 24 hours ahead and at 40 km resolution for 24 to 48 hour horizons. One main feature of this system is that it can adapt easily to changes in the wind farm or in the NWP model. Sipreólico uses an adaptive combination of a set of statistical models and it updates the estimation of all models using the most recent information. The forecasting system is composed of 9 different models of

two types: dynamic linear models, which convert wind to power using polynomials of different degrees, and non-parametric models. Some models use NWP data, and some do not, and each has better results at different horizons. The models are obtained by using the Recursive Least Squares (RLS) algorithm and a Kalman Filter. Another feature is that there is no pre-calibration needed, so it can be used for different wind farms. The system uses different inputs such as NWPs, online data and manufacturers' power curves. Depending on the circumstances, it can deal with missing data. The system has an additional diagnosis tool to deal with uncertainty. This tool evaluates the accuracy of the predictions through the calculation of a variety of statistical parameters and error measurements (Sánchez et al., 2002). It was reported by these authors that one of future enhancements was to consider NWPs from different sources and apply a weighting strategy to add them as an attempt to improve the quality of their forecasts.

LocalPred, developed by the Centro Nacional de Energías Renovables (CENER), combines physical and statistical models to provide short-term forecasts for wind farms (Pérez, 2002). The system uses meteorological forecasts from HIRLAM and uses MM5 for high resolution physical modeling. Then, using MOS techniques, it removes systematic errors and applies a power curve model using ridge regression and fuzzy logic methods. The system quantifies uncertainty by the use of ensemble predictions which are obtained combining numerical predictions from various models.

Garrad Hassan, now DNV GL, developed a forecasting model called GH Forecaster based on the UK MetOffice NWPs. This system provides hourly forecasts of future production of UK wind farms and it implements linear regression techniques to convert NWP to local wind speeds and power models to finally convert to hourly power outputs. According to Parkes and Tindal (2004), the GH Forecaster is able to improve the Persistence model by 60% for periods of 12 hours and this accuracy is maintained even beyond 2 days in advance. The authors present different error measures to evaluate the accuracy of the system, such as MAE and RMSE; however, they do not provide any results regarding uncertainty estimations.

A four year research and development project, called the ANEMOS project, started as an important effort in the development of forecasting systems and

integration to market operations (Kariniotakis et al., 2006). Twenty two partners from seven countries, including institutes, universities, industrial companies, utilities, TSOs and agencies, participated. Kariniotakis et al. reported from this project that the software developed was installed to evaluate the online operation at some on- and offshore wind farms. As the major follow-up, two projects are running now, ANEMOS.plus and SafeWind. ANEMOS.plus is working on the integration of the forecasts in the electricity markets while SafeWind is a project oriented to the study of extreme events (i.e. ramp events). In this project, Kariniotakis emphasizes "the need of developing dedicated approaches to reduce large prediction errors or predict extremes at local scale as state-of-the-art wind power forecasting systems are focused so far on "usual" conditions rather on the extreme ones" (Kariniotakis, 2008).

Most commercial wind power forecasting systems have been published via fast track routes, such as conference papers/posters or technical reports. There is a great amount of work published in journals which does not aim to build complete forecasting tools but instead to introduce new approaches to improving the specific steps of the forecasting process. Some of these are focused on the wind speed downscaling step, some on the power conversion. Sweeney et al. proposed different post-processing methods to improve wind speed COSMO model forecasts (Sweeney et al., 2011) using different resolutions. These post-processing methods included a Kalman filter approach (Sweeney and Lynch, 2010), ANN approach (Salcedo-Sanz et al., 2009), bias correction methods and a combination of these. Sweeney et al. were able to obtain similar results at 3 and 7 km resolution, meaning that computational costs can be reduced by using post-processing methods instead of high resolution NWPs. Zhao et al. proposed the implementation of WRF model together with a Kalman filter method for wind speed and wind power forecasting for a wind farm in China (Zhao et al., 2012). Kalman filter approaches have also been applied by Monache et al. (2011) and by Cassola and Burlando (2012) as post-processing tools for correcting the bias of WRF wind speed predictions, reducing significantly the size of the training set, compared to ANN based methods.

As mentioned previously, an increase in the horizontal resolution of the mesoscale model is not always a guarantee for improvement in the quality of the forecasts

(Louka et al., 2008; Möhrlen, 2004). The same amount of computational resources used for increasing the horizontal resolution could be used for ensemble forecasting. Taking this into account, several downscaling methods have been proposed to avoid high resolution runs. Salcedo-Sanz et al. (2009) have proposed the integration of GFS and the mesoscale physical model MM5 (Grell et al., 1994) together with neural networks for short term forecasting. The neural network is used to perform the final downscaling from the mesoscale model to the observation sites, avoiding the execution of the numerical model at high resolutions. However, the neural network approach behaves as a black box, which does not provide information about the model that was found and needs a significant amount of training data to ensure generalisation. The same forecasting model was implemented replacing the ANN with an SVM approach (Salcedo-Sanz et al., 2011).

2.5 Forecasting and Characterising Large Changes in Wind Power

One of the current issues in wind power generation is dealing with ramp events. Ramp events, as mentioned before, are characterised by sudden increases or decreases in wind power. With high levels of wind penetration, ramp events can be quite large, becoming a great challenge to balancing the load and the generation. This has been well documented. For example, a rapid and large ramp-down event occurred in the ERCOT operations area in February 2008 that forced them to declare a system emergency (a high cost system condition) (Francis, 2008). Although wind power forecasts were available to the operator, they were not integrated into the ERCOT system operations. The use of forecasts was an important lesson learnt from this event, and, most of all, that the way the forecast is to be used should be just as important as the forecast itself (Ela and Kirby, 2008). System security is one of the main aspects that TSOs need to consider when managing the grid. The uncertainty of ramp events might prevent TSOs from incorporating wind power sources into the system, taking into account the potential risk these ramps represent. To address these events, TSOs, utilities

and wind power operators need to develop mechanisms that can allow them to satisfy the electricity demand while the economic and environmental benefits are maximized. If ramp events are identified and characterised early enough, the procedures will be more effective.

One of the problems when dealing with the characterisation of these events is that there is no standard formal definition. The literature reports different definitions depending, for instance, on the location and size of the farms or the characteristics of the power system the wind farms are connected to. In general, most authors agree that a ramp occurs when there is a change in power output that has a large enough amplitude over a relatively short period of time. Figure 2.10 illustrates the definition presented by Greaves et al. (2009), which according to the authors, is a change of 50% or more of the wind farm capacity in 4 hours or less.
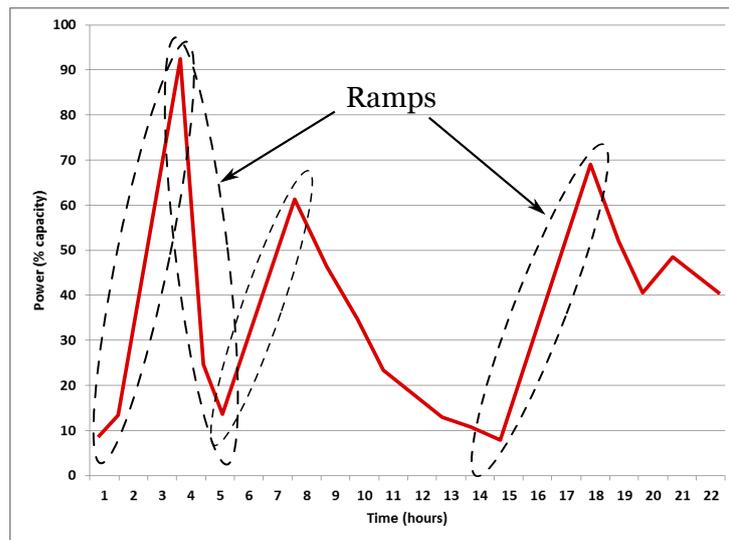
Figure 2.10: Ramp event definition: A change in more than 50% of the capacity in a maximum time window of 4 hours. Figure inspired by the work of Ferreira et al. (Ferreira et al., 2010).

In general, all these definitions seem a somewhat arbitrary. It would be good to have a characterisation that is related to the needs of the power grid in some way. As stated in the literature, a ramp event can be identified according to the power signal $P(t)$ and two user-defined parameters. The first parameter,

Δ_t, is related to the duration of a ramp and defines the time interval considered to be identified. This duration could range from minutes to hours, depending on the configuration of the system. The second parameter, P_{val}, is related to the type and magnitude of the ramp and provides a cut-off level on the power changes. This parameter is set usually as a percentage of the wind power nominal capacity. Kamath (2010) defines a ramp event as a change occurring at the start of an interval if the magnitude of the increase or decrease in the power signal at time Δ_t ahead of the interval is greater than a predefined threshold value:

$$|P(t + \Delta_t) - P(t)| > P_{val} \qquad (2.7)$$

This definition does not consider any ramp events that occur in the middle of the interval. To consider them, Kamath presents a second definition formulated as:

$$max(P[t, t + \Delta_t]) - min(P[t, t + \Delta_t]) > P_{val} \qquad (2.8)$$

This second definition looks for the maximum and minimum value of the power output in a time interval. However, it is not possible to determine how quickly the change occurred. This kind of information is provided by the slope of the curve or power ramp rate. To address this, Zheng and Kusiak propose characterising power ramps based on the power ramp rate PRR_{val} (Zheng and Kusiak, 2009). In this characterisation approach, a ramp event is flagged when the ratio between the absolute difference of the power measured at two time points (the initial and the final points of the interval Δ_t) and the size of the interval Δ_t is greater than a predefined reference value:

$$\frac{|P(t + \Delta_t) - P(t)|}{\Delta_t} > PRR_{val} \qquad (2.9)$$

The larger the absolute value of PRR, the faster the power surge (or drop). From Equations 2.7 and 2.9, the type or direction of the ramp can be identified: if $P(t) > P(t + \Delta_t)$ then the change represents a ramp-down event; otherwise, it is a ramp-up event. In Equation 2.8, on the other hand, it is difficult to be able to identify if the change represents a ramp-up or down event as the minimum

and maximum values are not identified in time. In order to identify the type, one would need to identify the relative position of the extreme time points within the interval.

Instead of working directly on the power signal, some ramp detection approaches pre-process the signal before identifying the ramps. A common transformation consists of considering k-order differences in the power amplitude. Bossavy et al. presented a study where filtered versions of a power signal are used to identify possible ramps (Bossavy et al., 2013). The filtered signal calculation can be expressed as:

$$p_t^f = mean(p_{t+h} - p_{t+h-n_{am}}; h = 1, ..., n_{nam}) \tag{2.10}$$

where n_{am} is the number of averaged power differences to consider. According to this filtered signal, the ramp event occurs if the absolute value of the filtered signal p_t^f exceeds a given threshold value P_{val}:

$$|p_t^f| > P_{val} \tag{2.11}$$

The time of the ramp is defined by the value that reaches the maximum value in the interval.

Regardless of the definition used, most ramp event studies look for a specific percentage of change in power output, which is a binary classification, since a decision needs to be made if the ramp exists or not. The disadvantage of this definition is that slightly lower changes might not be identified and these may be equally important to the grid operator. The following sections provide details of some of the most relevant studies carried out to understand the meteorological causes related to ramp events and the approaches developed to characterise them.

2.5.1 Understanding the Meteorological Causes of Ramp Events

A report from AWS Truewind presents an analysis of the frequency of large ramps in different wind farms in California, USA (Zack, 2007). The authors define large ramp events as changes in more than 50% of the capacity of the

farm in 2 hours or less. They found that different wind farms presented different levels of variation and that the ramp events could be associated with different atmospheric processes. Some of the large ramp events were associated with large-scale weather systems, such as low and high pressure areas and cold and warm fronts. These types of meteorological events usually have long life cycles, so they are easy to model by NWP models. A second type of atmospheric process identified to produce large ramp events were the local or mesoscale circulations. These circulations include sea-land breezes, mountain-valley winds, drainage flows and gap flows. These have a shorter life cycle and smaller spatial scales; hence, they are more difficult to diagnose. The authors suggest that, for the day ahead, NWP models are the best option to forecast ramp events associated with these kinds of meteorological processes. The third type of atmospheric process that was identified was the vertical turbulent mixing momentum. Large ramp events can occur in two ways in this situation. One occurs when there is a layer of high wind speeds above a low speeds layer at the height of the turbine rotors. The turbulent mixing of wind speed between the two layers can bring the high wind speeds down to the turbine rotor level, creating explosive ramp-up events. A second type can occur when the layer of high wind speeds from the surface up to the height well-above the top of the turbine is mixed with a low thermodynamic stability. A sudden cooling of the near-surface layer can drastically increase stability, which decreases the turbulent mixing. In this case, the wind speed can experience a sudden decrease below and at the rotor height, decreasing the amount of power output of the turbine for a short period of time. In a similar way to turbulent mixing, a thunderstorm can cause a ramp event, but involving water as well, which results in a process with a longer life time but still at a small spatial scale. This small spatial scale could mean that the thunderstorm could affect at a wind farm level, not at a region level, as large-scale processes do. Finally, ramp events were also identified to occur when the wind speed exceeds the turbine cut-off speed (usually 25 m/s). In this type of event, a very small variation in wind speed can result in a large ramp event as the wind turbine is shut down for safety reasons. This event is usually difficult to predict as small errors in the prediction can result in large errors of real power output. The AWS Truewind report emphasizes the need for developing dedicated ramp event forecasting systems as ramp events are

usually outliers in data samples used to train most forecasting models.

Another study by AWS Truewind performed in the ERCOT domain reported the meteorological causes of the ramp events observed in the region (AWS Truewind, 2008). To characterise the events, the authors defined ramp events as a change in 20% of the rated capacity of the wind farms in the region over a 30 minute period. The meteorological processes involved in the observed ramp events are similar to the ones reported in Zack (2007) such as thunderstorms, cold fronts, frontal passages, dry lines and weakening pressure gradients. Convective events were the primary cause of all ramp events, followed by frontal passages and weakening pressure gradients. In the case of the ramp-down events, the most frequent cause found was the weakening pressure gradient. Although all ramp events were mapped to a meteorological one, it was found that every meteorological event was temporally and spatially unique. This peculiarity does not allow us to make any generalisation of the duration of the ramp event based on its meteorological cause. This finding is interesting as it means that, even if a ramp forecasting system is developed based on the training over historical data of ramp events, it would not be possible to use this past information to characterise correctly the timings and duration.

Ela and Kemper presented a study on the characterisation of large ramp events in different areas of the U.S. With the characterisation of the observed ramp events, the authors studied the correlation of these between wind farms, and how that could provide information to an utility for predicting those ramps (Ela and Kemper, 2009). The study is focused on the Xcel Energy service areas. Xcel owns one of the largest amounts of wind capacity in the United States. The authors note the importance of wind power forecast accuracy during times when the wind power production is changing rapidly. During a ramp-down event, an operator needs to compensate for the loss of generation by either ramping up or turning on reserve units. During a ramp up event, the operator compensates for the excess of load by ramping down units, shutting them off, or, in the worst case, curtailing the high-producing wind. The different strategies of the operator to handle these situations will depend on the speed (i.e. ramp rate), duration, magnitude, and timing of the ramp. It will also depend on how far in advance the ramp is predicted and the different generating units available in the system. Units

with longer start times are usually cheaper than units with shorter start times. Also, units that can remain on for longer are cheaper than those with shorter minimum on times. The prediction of ramp events in advance could potentially reduce costs. The magnitude and duration of a ramp can help to determine the type of energy requirements to compensate for the ramp.

Ela and Kemper found that, from the total number of ramps observed in the area of Colorado (PSCO), there are more occurrences of ramp up events compared to ramp-down events. The authors concluded that the more strict the ramp definition (i.e. the larger the threshold), the higher the percentage of ramp up events compared to ramp-downs. It was also observed by the authors that there is a correlation between ramp direction and time of day in the area of Texas and New Mexico (SPS) that was not seen at PSCO. The majority of the ramps during the evening and night period were ramp-ups, while most ramp-downs were early in the day. These results demonstrate how different and site-dependent ramp events are, so, in order to develop tools, one first needs to understand what factors are involved in the development of the such ramp events. The authors also studied the correlation between ramp events in one farm and another, depending on the dispersion and terrain features. They concluded that geographic diversity lessened the impact on ramps.

2.5.2 Evaluation Metrics

In order to assess the accuracy of a ramp event forecast, there are several metrics used in the literature depending on whether it is a deterministic or probabilistic forecast. Two statistics that are widely used to evaluate deterministic forecasts are precision and recall. Precision is defined as the ratio between the events that were well forecasted and the total events forecasted. Recall is defined as the ratio between the number of well forecasted events and the number of observed ones (Ferreira et al., 2010). Precision is formally expressed as:

$$Precision = \frac{TP}{TP + FP} \tag{2.12}$$

where TP corresponds to true positives (well forecasted) and FP to false positives

(forecasts that did not occur). Recall is expressed as:

$$Recall = \frac{TP}{TP + FN} \qquad (2.13)$$

where FN corresponds to false negatives. Precision, also referred to as forecast accuracy in Greaves et al. (2009), determines the fraction of predicted events that really occurred. Recall, also referred to as ramp capture, determines the fraction of observed ramps that were correctly forecasted. To assess the performance of a classification system, precision and recall can be combined as follows:

$$F_{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (2.14)$$

An F_{score} of 1 would mean that all events were detected, while a score of 0 indicates that none were detected. Another metric widely used is the Hanssen & Kuipers skill score (KSS), also known as Peirce's skill score or the true skill score. This score measures the ability of the classification system to separate the correctly forecasted events from the false ones. It can be calculated by means of the hit rate (H) and false alarm rate (F) in the following way:

$$KSS = H - F = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \qquad (2.15)$$

where TN corresponds to the true negatives, which are the cases where no ramp was observed nor forecasted. The KSS values range between -1 and 1, where -1 indicates no skill and 1 is the perfect score.

As the binary classification of events is a definition that might lead to artificial results, the use of probabilistic forecasts is a potential way to introduce a degree of freedom. One technique that can be used to choose the optimal threshold is the receiver operating characteristic (ROC) curve. This technique shows graphically the sensitivity of the classifier system when the threshold is varied. The ROC is obtained by plotting the fraction of true positives versus the fraction of false positives as the criterion threshold changes (Hand, 2009).

For those ramp detection tools that assign a probability to each event, other types of metrics could be used. The Brier score (BS) (Brier, 1950) is a score function that measures the average squared deviation between predicted probabilities

for a set of events and their outcomes. It is expressed as:

$$BS = \frac{1}{n}\sum_{t=1}^{N}(F_t - O_t)^2 \qquad (2.16)$$

where F is the probability that was forecasted and O is the actual outcome of the event at instance t (0 if it does not happen and 1 if it does), and N is the number of forecasting instances. A low BS score will indicate a high accuracy of the forecasting method.

Bossavy et al. (2013) used the Brier skill score (BSS), which is based on the BS score, to compare the performance of their probability forecasts with a reference methodology. The BSS is expressed as:

$$BSS = 1 - \frac{BS}{BS_{ref}} \qquad (2.17)$$

where BS_{ref} and BS are the Brier score of the reference and proposed models. The BSS is a particular case of the ranked probability score (RPS), which is used to assess the performance of multicategory probabilistic forecasting systems.

2.5.3 Ramp Forecasting Approaches

The first approaches for wind power ramp characterisation are fairly recent. One of the first studies was presented by Zheng and Kusiak (2009), which reported a system for the prediction of power ramp rates using a time series approach. The authors apply different data-mining algorithms to build the prediction models for 10, 20, 30, 40, 50 and 60 minutes into the future ramp predictions. Some of the algorithms applied are the multi-layer perceptron (MLP), SVM, random forest, classification and regression(C&R) tree and the pace regression algorithm. The study used one month of wind speed and power measurements from a wind farm with 100 turbines. The data was collected every 2 seconds, but averaged and stored at 10 minute intervals. The predictors used to build the time-series models were the mean, standard deviation, maximum and minimum wind speed of a turbine, the wind farm power output and the power ramp rate. According to the results obtained, for horizons from 10 to 40 minutes into the future, the time series models can predict accurately the ramp rates. However, for a larger

horizon, this accuracy degrades, demonstrating the importance of the inclusion of weather forecasting data.

Potter et al. (2009) presented a study on the benefits of using a dedicated ramp forecasting tool during the reserves allocation process. The authors remarked that there are several challenges to be addressed when developing a ramp forecasting system. These challenges are:

- *Timing error* (phase error): when a ramp event is predicted accurately in magnitude but it occurs at a different (unexpected) time.

- *Intensity error*: when the ramp event is forecasted at the "right" time but with the wrong magnitude.

- *Location error*: when the weather model forecasts the event at another physical position or the event follows a different path than the one forecasted, resulting in timing and intensity errors.

These types of error are highly related to the numerical model, although they might be also related to the wind to power conversion process (Pinson, 2006). Potter et al. also suggested that the use of a strict "ramp or no ramp" binary definition can result in artificially low accuracy. To address this, Potter et al. proposed the use of probabilistic ramp forecasts, characterising each event with a potential risk which would need to be assessed by the grid operator. The authors compared the costs of using the probabilistic forecasts with the costs of having a constant backup and the costs of not having no backup protection at all. According to their results, there is a potential economic benefit in incorporating these kinds of systems into the grid management. Unfortunately, the authors did not provide any detail on how these probabilistic forecasts are to be obtained.

Greaves et al. (2009) presented a study where NWP models are used for characterising the timings of ramp events. The study was performed using data from wind farms in the US and the UK. The observed ramp events were treated separately by wind farm and as a portfolio, or group, of farms. The results showed that, in general, forecasts for a group of farms were significantly more accurate than for individual farms. However, despite the high accuracy, the method used to forecast in individual sites was not optimal for portfolios, implying that different

definitions of a forecast ramp event should be used in each case. The authors also suggested that, although using NWP data from different sources can reduce the overall error of the forecasts, the methods used for combining them are not optimal as some ramp event information may be lost.

Bossavy et al. (2013) proposed the use of NWP ensembles for wind power ramp predictions. The ensembles are used to improve the time prediction of the ramps by calculating the probabilities of the ramp happening at a certain time interval depending on the number of ensembles that model the change at that interval of time. The authors used the Brier skill score to assess the quality of the predictions. Their predictions are compared to the Brier core of climatology, which is calculated based on the number of observations that fell in each time interval. The results showed that the proposed approach has more skill than the climatology.

According to Cutler et al. (2009), the timing or *phase error* of numerical models can be addressed taking into account a wider area of the NWP grid, not only the closest point to the observation site, due to the NWP being unable to place the meteorological process in the correct physical position, causing a "misplacement" error. This wide area approach is what the authors denominated as *spatial fields*. Cutler et al. (2009) developed a visual tool that shows the possible power output of the farm if it was located at different points of the mesoscale grid. In order to calculate the potential available power in the surrounding points, each grid point is standardised to be equivalent to the terrain at the closest point assuming a homogeneous relationship between the wind speeds at the closest point and at the neighbor points. Then the wind speeds at each grid point are converted to wind power using a farm power curve obtained using wind speed and power observations at the farm. The purpose of this is to provide a graphic forecast of potential power outputs of the farms that could indicate changes that are not modeled at the closest point. The authors compared the visual results to specific case studies where large ramp events were detected. The tool is described by the authors as a complementary tool apart from the single time-series forecast that still provides useful information on smaller spatial scale events happening at a close point to the farm. One of the disadvantages of this visual tool is that it needs the interpretation of an expert to determine the potential risk that changes

represent in the neighbourhood to the farm.

Tastu et al. (2011) performed a spatio-temporal analysis of short-term wind power forecast errors in western Denmark. The authors suggested that a forecast error made at a given point in space and time might be related to forecast errors at other points in space in later periods of time. According to their study, the higher the wind speed, the stronger the dependency on more remote places, while in the case of lower wind speeds, the influence came from a local origin.

Gallego et al. (2013) presented a ramp forecasting approach using wavelets, which avoids using a fixed change percentage, analysing the power forecast at different magnitudes of change and different time windows. Despite their promising results, the application to the day-ahead market is not addressed in depth, nor how this could allow end users to interpret the different ramp intensities.

NWP models are necessary for the characterisation of ramp events one day in advance. However, it has been shown that a forecast based on the closest point of the NWP model to the location of the wind farm is not the best approach as the forecasts might be potentially misplaced. For this reason, there is potential in the use of neighbour points to see how these can be used to improve the characterisation of ramp events.

2.6 Discussion

There are two main aspects, or lessons that have been learnt as wind power has evolved as an important source of renewable energy. On the one hand, the first lesson learnt is the need for accurate forecasts in order to incorporate wind power into the electricity market. Wind power is an intermittent and uncertain source that, although quite appealing due to its vast availability and low cost, in order to treat it as a conventional power source, the incorporation of a prediction is crucial. For this reason, there has been great effort in the last decades to try to improve the accuracy of wind power forecasts. Most state-of-the-art forecasting systems achieve a 10-15% RMSE of the total installed capacity for a 36 hour horizon, although one of the latest reports from Germany mentions the reduction to less than 5% of the RMSE (Giebel et al., 2011). The literature shows there is still ongoing research to improve forecasts and that researchers in the field are

continuously looking for new approaches.

On the other hand, the second lesson learnt is that it has been shown that the "how" of using a forecast is as important as the forecast itself. The majority of TSOs are using forecasts in deterministic UC and ED decisions and the use of probability forecasts is still in its infancy in terms of the electricity market operations. Although the literature reports a vast amount of research on uncertainty forecasts, there is still a gap between the research and the actual application. As this situation is currently true for most markets, it has become relevant to introduce the use of dedicated ramp forecasting systems in order to improve the allocation of backup generation. This is a very critical aspect of the wind power integration. Security is one of the main concerns for operators as everyone relies on the constant supply of energy for their daily activities. In recent years, the increase of allocated backup units has been inevitable as wind power is bringing more uncertainty to TSOs. However, it is not acceptable for power markets to increase the amount of dedicated backup generation to ensure continuous supply if the resulting costs are too high.

Ramp forecasting is becoming a potential area to exploit as the recent research has shown the benefit of using these types of forecasts in the reduction of costs and, hence, for having a positive environmental impact, as more wind power is used and less backup is allocated. It is for this reason that it is interesting to investigate how to adequately use NWP predictions to get as much information as possible to characterise ramp events.

It has been well established in the literature that NWP models are essential for the day-ahead market. According to several studies, running NWP at very high resolutions might not improve significantly the quality of the meteorological forecasts, which has led to an increasing use of statistical methods to do the the final downscaling step to the location of interest. Forecasting systems such as WPPT use statistical methods for this. Some authors argue that it is better to use the computational effort to run ensembles rather than to run high resolution NWP forecasts. Ensembles have caught the attention of researchers as a way to characterise ramp events. Most ensemble forecasts reported in the literature are based either on perturbations of the initial conditions, different parameterisations or the combination of different forecast sources. Spatial fields is an area where

very little attention has been paid. There are some studies now emerging on the spatio-temporal errors which show the importance of taking into account information from wider areas to improve the characterisation of errors.

This research aims to explore two aspects of wind forecasting which, when combined, could potentially improve the characterisation of ramp events. One is the one-step conversion of wind speed to power by the use of a non-parametric technique and the other is to apply that procedure to a wide area in order to compute different possible power outputs of the farm using different grid points. Each neighbour point will provide a possible generating scenario as part of an ensemble of scenarios.

Chapter 3

Data Collection

3.1   Introduction

As the major interest in this research was to study and characterise wind power forecasts for the DA market, it was essential to incorporate numerical models as the main source of data into the characterisation process. For the time horizon of study, it was required to have availability of hourly forecasts for up to 48 hours into the future, wind speed being the most relevant variable to obtain. A first option was to use Global forecasts which are freely available for any location in the world and for a wide range of years. However, their spatial and time resolution was not enough for this application. The following option was to use mesoscale model predictions, which can achieve a high time and space resolution. It was then decided to run locally the Advanced Research WRF mesoscale model (WRF-ARW), which can be initialised with global inputs from the GFS system. The main advantage of running locally a mesoscale model like WRF-ARW is the availability of different meteorological variables at different pressure levels and locations. This was essential for the prediction of wind power and the characterisation of ramp events. In addition, this mesoscale model has good support from the University Corporation for Atmospheric Research (UCAR), which provides user assistance and maintains the code. Apart from the meteorological forecasts, wind speed and power observations were obtained from the locations of study.

This chapter presents in detail how both meteorological and observation data

was obtained and processed. Section 3.2 describes the sites of study and their characteristics. Section 3.3 introduces the WRF-ARW model, which was used for all meteorological forecasts. The parameterisation of the model was tested using several case studies provided by the NOAA, like for example the Hurricane Katrina. In Section 3.4, the acquisition of wind speed forecasts for the state of Illinois, USA, is explained, and the quality of the results are analysed. Then, in Section 3.5, a different setup of the model is introduced for wind speed forecasting at Galicia, Spain. Finally, a summary is provided in Section 3.6.

3.2 Sites of Study and Available Observation Data

The availability of free meteorological and observation data is very restricted, and it is specially difficult to find both for the same location. The best option in terms of meteorological data was to run the mesoscale model in a location where wind speed/power observations were available.

To run the mesoscale model locally, an adequate parameterisation was needed. To make sure the model was well set-up, the quality of the wind speed forecasts needed to be assessed, before any further experimentation into wind power forecasting was carried out. This required the availability of wind speed observations, if possible at multiple sites, to observe if the quality was consistent. At this stage it was not required to have access to wind power observations. For this reason, the data available from Illinois Wind seemed suitable. Illinois Wind is part of a program from the Illinois Institute for Rural Affairs that assists rural residents and communities of Illinois to build knowledge of wind power (Illinois Institute for Rural Affairs, 2014). Together with the Western Illinois University, they provide organisations and landowners the opportunity to assess the wind at a site under consideration for wind power. A limited number of sites is selected each year to undergo monitoring of wind velocity and direction for short periods. The sites depend on the applicants interest in developing a small scale farm. The project provides free access to historical observations at various locations. This was a good starting point as there were three locations in Illinois where observations

were available for the same period of time (9 months) and it only required one run of the model to cover the three locations, instead of having three independent runs, one per location. The three observation sites used in this study, Cuba, SIUE and Wilmington, are monitoring points where wind speed, wind direction and temperature were captured from November 2010 to July 2011 every 10 minutes at 10, 40 and 58 meters height . The latter is at the most appropriate height for wind power prediction, so this height was selected for the study. The latitude and longitude location of the three sites is shown in Table 3.1 and Figure 3.1 shows their geographic location.

Table 3.1: Location of the three observation sites in Illinois, USA

Site	Latitude	Longitude
Cuba	N 40 28.906	W 90 11.187
SIUE	N 38 48.098	W 90 00.473
Wilmington	N 41 17.999	W 88 7.818

Figure 3.1: Observation sites in Illinois. The blue place mark corresponds to SIUE, the red to Cuba and the green to Wilmington.

As the data was in a 10 minute frequency, averages were calculated to be compared with the hourly averages obtained by the WRF-ARW model for that same period. At this point, this data was not selected for its frequent incidents of ramp events. The comparison of the numerical prediction against these ob-

servations were used to evaluate the accuracy and ensure the correct settings of the numerical model. This 9 month period was selected as it presented more diversity in observed locations and could give more information of the accuracy of the numerical model in different terrain characteristics.

After experimenting with the setup of the numerical model, the following phase in terms of data collection/generation was to look for a wind farm location where wind power and wind speed/direction observations were available. Sotavento Experimental Wind Farm (Sotavento, 2014) is a wind energy project established in 1997 and promoted by the Government of Galicia, Spain. The main purpose of this wind farm is the promotion and implementation of projects, studies and research related to renewable energy. The wind farm is situated in north-western Spain at approximately 40 km from the Atlantic Ocean in a moderately complex terrain. The wind farm is composed by 24 wind turbines of 9 different models. The characteristics of the models are shown in Table 3.2 .

Table 3.2: Wind turbine general features

Turbine Type	Unit Power (kW)	Rotor Diameter (m)	Tower Height (m)	Blade Pitch
Neg Micon NM-48 750	750	48	45	Fixed
Gamesa G-47	660	47	45	Variable
Made AE-46	660	46	45	Fixed
Izar-Bonus MK-IV	600	44	40	Fixed
Ecotecnia 44/640	640 (2 x 320)	44	46	Fixed
Neg Micon NM-52 900	900	48	45	Fixed
Made AE-52	800	52	50	Variable
Made AE-61	1320	61	60	Fixed
Izar-Bonus 1.3 MW	1300	62	49	Variable

The rated capacity of Sotavento wind farm is 17.56 MW. The wind farm has two measuring stations located at the middle and south of the farm. However, the observations available are only from one of the anemometers. Wind speed and direction observations are available at a 10-min and hourly frequency and at a height of 45m. Hourly observations were again selected to match the frequency of the hourly forecasts of the numerical model. The reason for this is that the

day-ahead market works at an hourly frequency, which is the term of interest for this research. Sotavento has historical observations available for multiple years. As it will be explained in detail later on, generating numerical data is time and resource consuming. For this reason, the amount of data selected was restricted to six months, from January to June 2012. This period presents similar wind patterns and the presence of both ramp up and ramp down events.

3.3 Setting the Basic Configuration of WRF-ARW With a Known Case: Hurricane Katrina

The WRF modeling system is a multiagency effort intended to provide a next-generation mesoscale forecast model and data assimilation system for both understanding mesoscale weather and accelerate the transfer of research into operations (Skamarock et al., 2001). This model was developed as a collaborative effort among a number of agencies like the the NCAR Mesoscale and Microscale Meteorology (MMM) Division, NOAA, NCEP, among others, and a number of university scientists.

WRF has a modular single-source code that can be configured for both research and operations. Its principal components are shown in Figure 3.3. The WRF Software Framework (WSF) provides infrastructure that support multiple dynamics solvers (e.g. NMM and ARW), physic packages that plug into the solvers through a standard physics interface, programs for initialization, and a data assimilation (WRF-Var) system which can be used to produce optimal estimates of the true atmospheric state at analysis time.

In this research the Advanced Research WRF (ARW) solver was used. The reason for this is that this solver is very well supported with documentation, online tutorials and a user forum. WRF-ARW is a non-hydrostatic limited area model (with an hydrostatic option) based on an Eulerian mass dynamical core (Skamarock et al., 2005). It solves a system of differential equations that represent the dynamics of the atmospheric flow. The solver uses a third order Runge Kutta time integration scheme with a split-explicit 2nd-order small time inte-
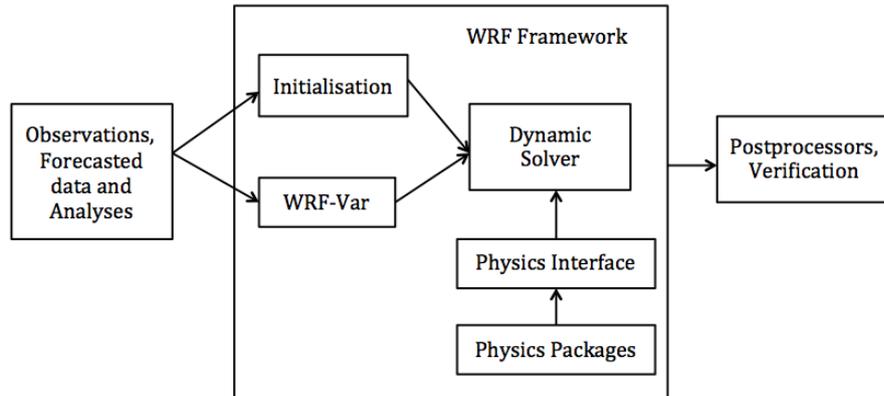
Figure 3.2: Principal components of the WRF Framework.

gration scheme for the acoustic and gravity-wave modes. Some of the variables predicted by the integration of the equations include velocity components u and v in Cartesian coordinate, vertical velocity w, potential temperature perturbations, among others. Figure 3.4 shows the program flow for a typical run.
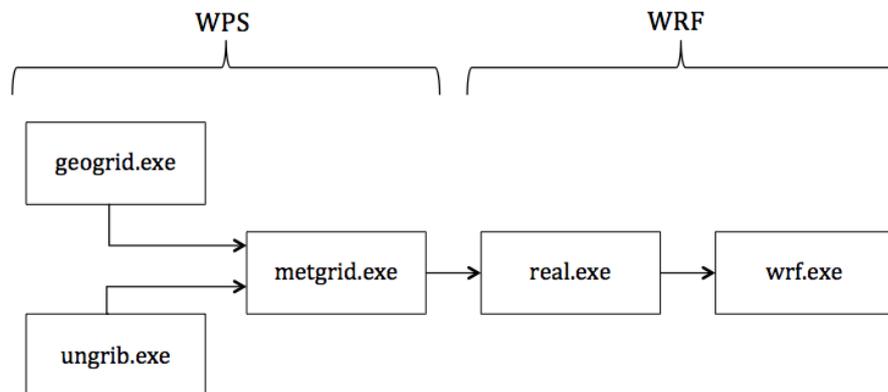
Figure 3.3: Program flow for a typical run of the WRF-ARW system.

The left side of the figure shows the steps needed to preprocess the data before running the model. This is done with the WRF Preprocessing System (WPS) that consists of three main programs. The first step is running geogrid.exe, which is used to define the simulation domain(s) and interpolate terrestrial data to the model grids. The domain refers to the area of interest, specified by latitude and

longitude locations within a minimum coverage area. Terrestrial data interpolated to the model grid include soil and land use categories, terrain height, among others. A list and definition of these is provided on Appendix A.

Figure 3.5 shows a domain setup centered in the Gulf of México for a hurricane simulation. Hurricane Katrina, the test case that will be used for the rest of this section, has been the most destructive hurricane to strike the U.S. This meteorological development happened on August 28, 2005. It strengthened to a Category 5 storm, based on the Saffir-Simpson hurricane scale, with winds estimated at 175 mph. Figure 3.6 shows the result of the interpolation of the albedo data into the grid points of the specified domain.
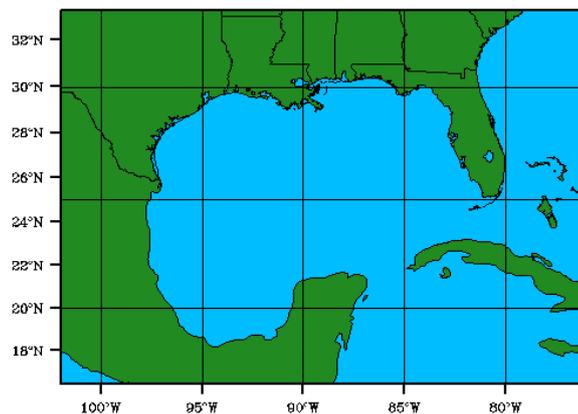
Figure 3.4: Domain setup centered on The Gulf of México, area of interest.

The ungrib module unpacks GRIB meteorological data (from the GFS), which for this case consists of 1 degree data at 26 pressure levels and a frequency output of 6 hours. After unpacking it, ungrib packs it into an intermediate file format that is used by the final program metgrid.exe which horizontally interpolates the meteorological data onto the model domain.

The ARW solver supports *horizontal nesting*, that allows to increase the resolution of the grid on smaller regions of the area of interest. This is done by adding an additional grid (or grids) into the simulation. The additional grids (nested grids) are rectangular and aligned with the parent (coarser) grid within which they are nested. In ARW, each nested region is entirely contained within a single parent grid. The finer nested grids are referred to as *child* grids. A parent
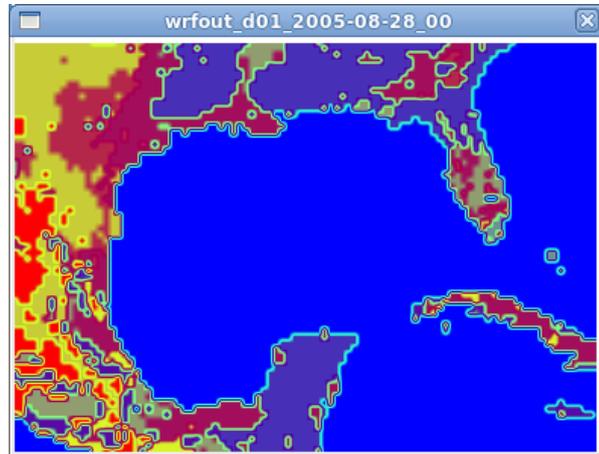
Figure 3.5: Albedo for the 28th of August 2005 at 00T (GMT).
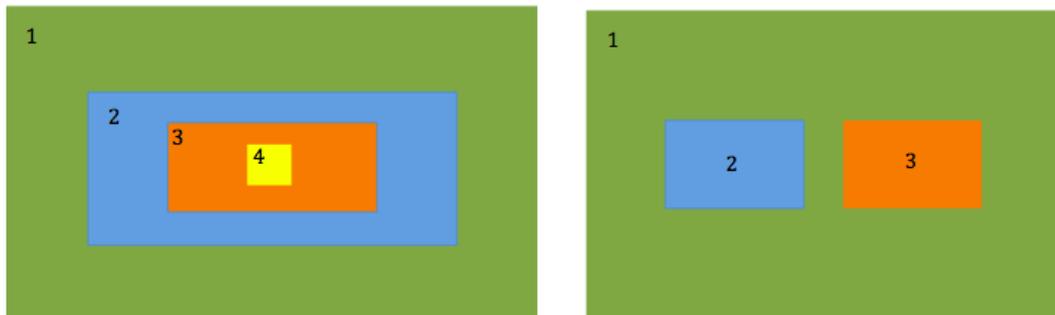
(a) Telescoping nests.

(b) Nests at the same level with respect to a parent grid

Figure 3.6: Possible WRF-ARW domain configurations.

grid may contain several child grids at the same level of nesting. These child grids cannot be overlapped. Figure 3.7 shows the possible configurations of the domains.

ARW uses *staggering*, which defines the way that the child grid is situated on top of the parent grid. For all odd ratios there is a coincident point for each variable: a location that has the coarse grid and the fine grid at the same physical point. The location of this point depends on the variable. Figure 3.8 shows the u and v components of horizontal velocity which are normal to the respective faces of the grid cell. Scalar variables are located at the center of the cell.

The nested grid simulations can be produced using either 1-way or 2-way grid
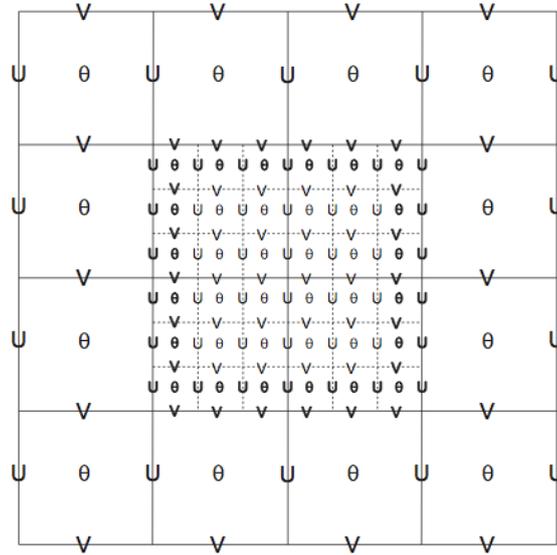
Figure 3.7: Example of two domain setting and the staggered variables u and v.

nesting. These options refer to how a coarse grid and the fine grid interact. In both cases, the fine grid boundary conditions are interpolated from the coarse grid forecast (from coarse grid to fine grid). In a 1-way nesting, boundary information is the only information that is exchanged between the grids. In the 2-way nest integration, the fine grid solution replaces the coarse grid solution for the coarse grid points that lie inside the fine grid. Figure 3.9 shows a two nested setting for the Hurricane Katrina case.
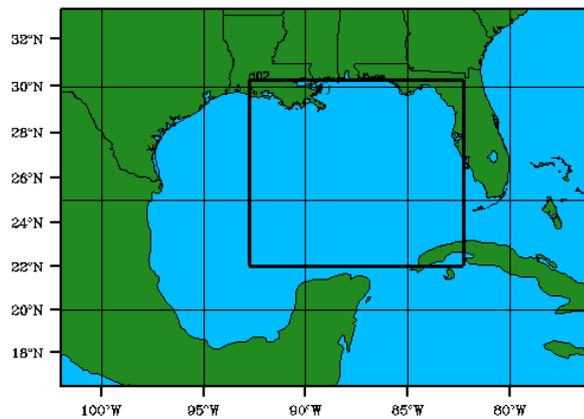
Figure 3.8: Two domain setting for the Hurricane Katrina simulation.

Once the input data is processed, the WRF-ARW module, which is the second part of the system depicted in Figure 3.4, can be executed. This module consists of two main programs, real.exe which interpolates vertically the data prepared by metgrid and creates the boundary and initial condition files, and wrf.exe, which generates the model forecast.

WRF-ARW has multiple physics options classified in different categories, which are (1) microphysics, (2) cumulus parameterisation, (3) land surface model, (4) planetary boundary layer and (5) radiation. The physics section, which is an independent module from the rest of the dynamics solver, involves the filling of arrays with physics-required variables that include the temperature, pressure, heights, layer thickness, and other state variables.

As the WRF-ARW model was installed in a supercomputer with a graphic processor (GPU), a CUDA based version of the WRF Single Moment 5 Cloud (WSM5) microphysics kernel was used as an attempt to decrease execution times as much as possible (Michalakes and Vachharajani, 2008). This was the only physics option that was available in a CUDA version, so the rest of the physics options were installed on the CPU with the rest of the WRF system. Table 3.3 shows the physics options used for Hurricane Katrina simulation.

Table 3.3: Physics options used in the WRF model

Domains	2 nested domains
Dynamics	nonhydrostatic Euler equations
Longwave Radiation	rapid radiative transfer model (RRTM)
Shortwave Radiation	simple downward integration
Surface Layer	MM5 similarity
Boundary Layer	YSU scheme
Cumulus	Kain-Fritsch (new Eta) scheme

The WRF-ARW real and wrf programs were executed with the Hurricane Katrina test case using GFS meteorological forecasts as input. Figure 3.10 shows the water vapor that was forecasted by WRF every 6 hours on August 28 2005.

In order to assess the benefits of running the microphysics module on the GPU, two experiments were designed. The first one consisted in running the system to

(a) Hurricane Katrina at 00Z.

(b) Hurricane Katrina at 06Z.

(c) Hurricane Katrina at 12Z.

(d) Hurricane Katrina at 18Z.
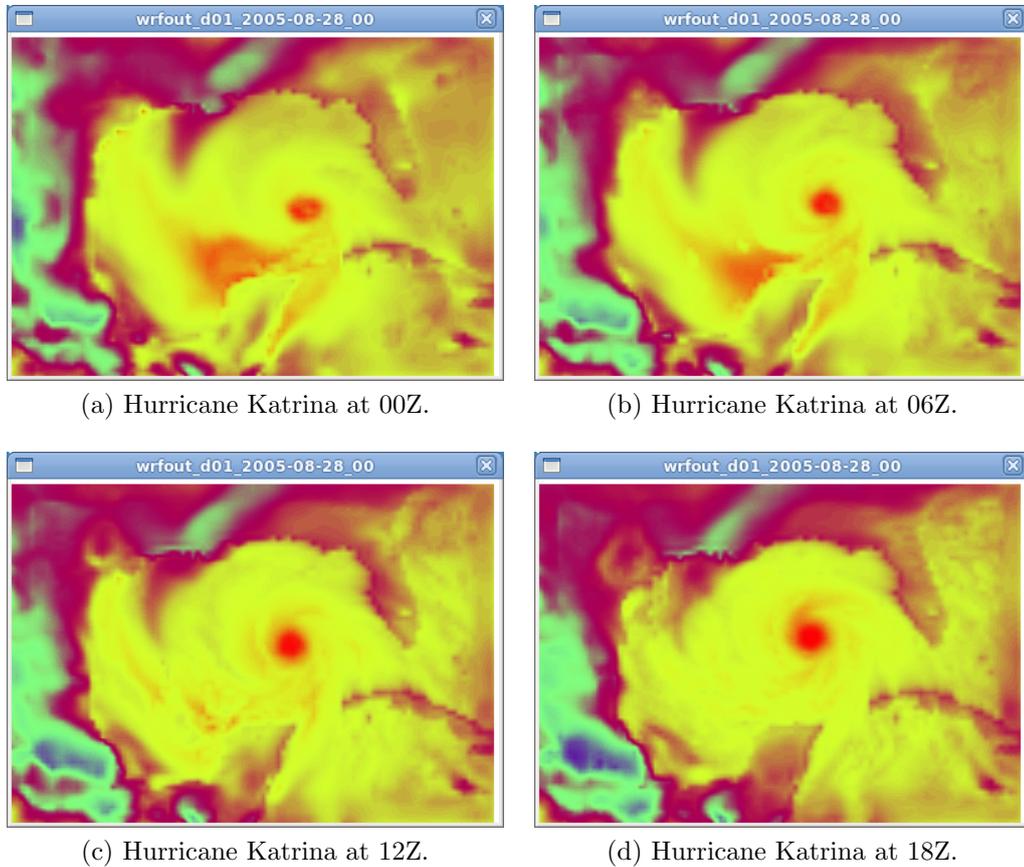
Figure 3.9: Hurricane Katrina forecasted with WRF on a single domain setup.

forecast one day (from 28/08/2005 00Z to 29/08/2005 00Z (Zulu or UTC time)) using a single domain with the microphysics module on the CPU and another run with the module on the GPU. The run with the module on the GPU took 283 seconds while the run with no GPU took 499 seconds. A second experiment was executed using a 2-way nested run with two domains for the same period of time. The execution using the GPU took 1952 seconds (approx. 32 minutes) while with no GPU, the execution took 3482 seconds (approx. 58 minutes). With these results, it was decided to continue with the GPU configuration.

3.4 Forecasting Wind Speed at Illinois, USA

With the WRF-ARW up and running, the model was ready to be configured
to run for any location. Before setting the model for wind speed forecasting in
Illinois, an analysis of the GFS input forecasts was performed. Global forecasts
were used previously for the Hurricane Katrina test case, however this case was
used only for configuration purposes. As the wind speed forecasts in Illinois were
the first formal evaluation of the numerical model, it was important to validate
the quality of the input data and use that as a reference when evaluating the
produced forecasts.

3.4.1 Initialisation Data for the WRF-ARW Model

Global models can be used as a first step towards the prediction of wind speed
at specific sites. Despite their low resolution, these forecasts can be used as the
start point and boundary conditions for a mesoscale model that could improve
the space and time resolution of these forecasts. Specifically, the GFS model was
used in this research (Kanamitsu et al., 1991). GFS runs four times a day, at
00Z, 06Z, 12Z and 18Z. At each run, it produces low resolution forecasts, this
means the entire globe is divided into a grid, of usually $1° \times 1°$ or $0.5° \times 0.5°$,
producing forecasts at each of the intersection points of the grid. Each GFS
model execution predicts up to 16 days into the future with a three-hour time
step in the first 8 days, and a 12-hour time step in the following days. GFS
data is freely available from the NCEP Products Inventory (National Centers
for Environmental Prediction, 2013). Their database usually maintains 1 year of
data online, but previous years can be requested.

Before further progress could be made, it was necessary to verify that a re-
lationship between GFS wind speed forecasts and real wind speed observations
exists. To establish this, a correlation analysis was carried out. Taking into ac-
count only the first 48 hours of each GFS forecast (predictions for the day-ahead
market), 50 contiguous runs of the GFS model were used, considering only those
values forecasted 48 hours into the future. The reason for this is that values fore-
casted 48 hours into the future are more likely to have a larger error than the rest
of the forecasted hours within the 48-hour horizon. The 50 runs corresponded

to GFS forecasts from the 6th to 18th of February, 2011. The data was selected in this month to match the observations, which were only available for limited periods.

Wind speeds at 58 meters height were extracted from the GFS model forecasts at the closest points in the grid to the three observation sites (Illinois Institute for Rural Affairs, 2014) in Illinois. This was achieved using the NCAR Command Language (NCL) post-processing tool, which is an interpreted language designed specifically for scientific data analysis and visualisation. It provides a set of libraries to process output files from models like GFS and WRF and extract the desired variables. The script developed using NCL consisted in the following steps: first, the height was converted to an approximate air pressure taking into account the terrain of the location. Subsequently, wind vectors (u,v) were interpolated vertically, using the pressure level, to obtain the wind speed prediction at 58m height.

Figures 3.11 and 3.12 show the correlation of GFS with observations at Cuba site. The GFS model has a correlation of 0.74668 and a RMSE of 1.90462 for values forecasted 48 hours into the future. The global model is able to capture the trend of local observations despite the low space resolution of the model. The same behavior occurred at the SIUE and Wilmington sites. This evident relationship between the GFS model and the observations proves the potential of the model for short-term prediction.

3.4.2   WRF-ARW Model Settings

To run the WRF-ARW system, the GFS data together with global terrestrial data at a 30″ resolution were used to build the domain, initial state and boundary conditions. The domain configuration is shown in Figure 3.13.

As shown in the figure, the model was set to run in two domains. The first domain D_1, which covers a major part of the United States, has a resolution of 30 km x 30 km and results from the first integration of the WRF model from the GFS grid (111km x 78km). The second domain D_2 (grey), is limited to the area of interest (state of Illinois) . It has a resolution of 10 km x 10 km and is obtained by a second model integration using the first domain as boundary conditions. A
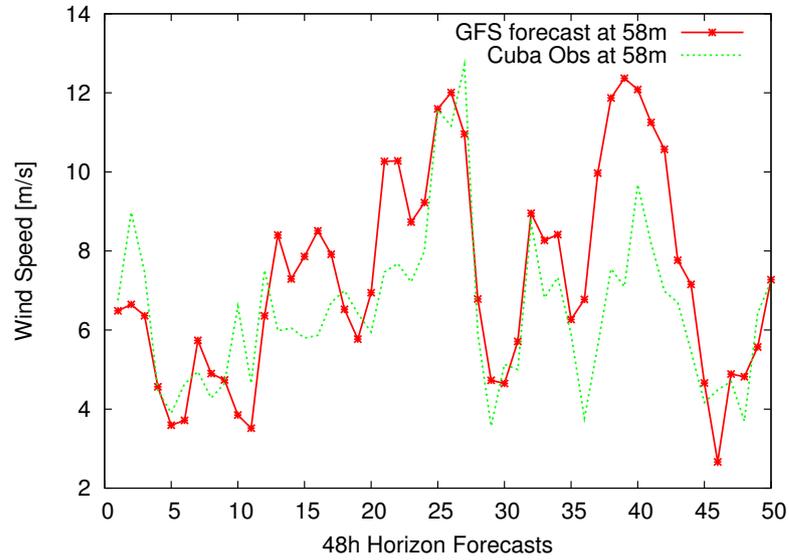
Figure 3.10: GFS 48h into the future wind speed forecasts at 58 meters and Cuba observations.
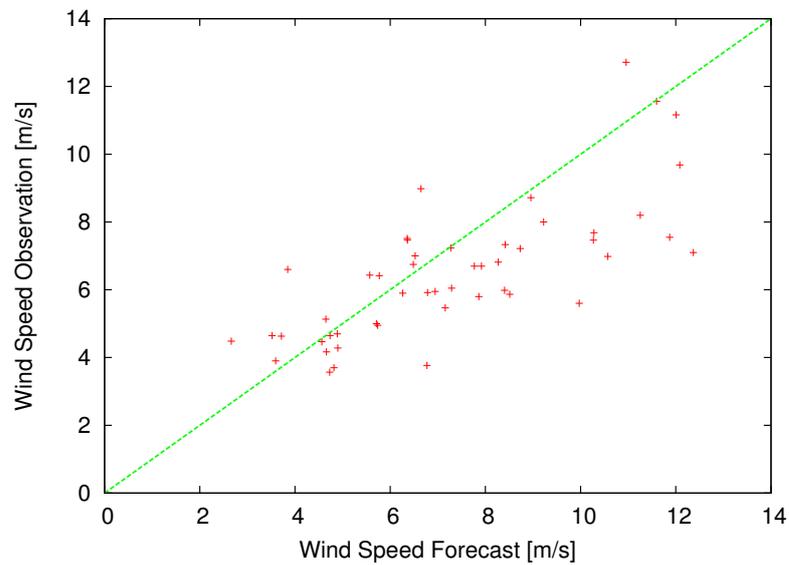
Figure 3.11: Correlation between GFS 48h into the future wind speed forecasts at 58 meters and Cuba observations.

third domain was considered, but the computational cost of running at 3 km resolution was too high. In a first attempt of running three domains, the system took approximately 12 hours to forecast a 12 hour horizon. It was decided then to
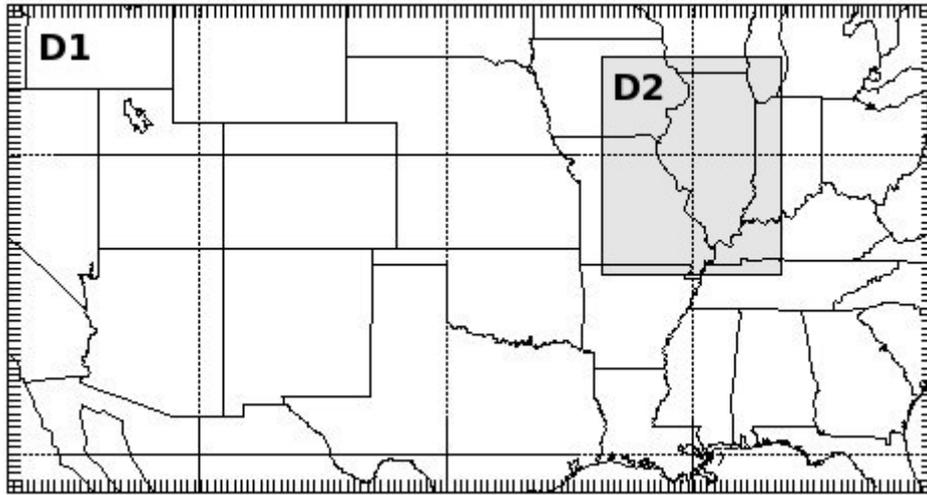
Figure 3.12: WRF-ARW two domain setting.

run in two domains and use an interpolation method for the final downscaling step to the location of interest. This step was performed with different interpolation techniques which will be presented later in this section and discussed in detail on Chapter 5. A similar domain setup is described in (Constantinescu et al., 2011).

From the data available at the three observations sites, the model was used to produce meteorological forecasts from February to July 2012, starting a new run every 6 hours (frequency of GFS forecasts). Each run produced complete 48-hour horizon forecasts with wind speed predictions every hour at a 10 km space resolution (second domain of the WRF grid). Using the NCL post-processing tool, wind speed predictions were extracted from the WRF-ARW forecasts and interpolated to 58 meters height. The most relevant steps of the NCL script to extract wind speed are the following (the complete script can be found on Appendix B):

- Obtain the x, y positions of the mesoscale grid that correspond to the closest point to the latitude and longitude of the observation site.

- Retrieve u and v wind variables from all levels at all grid points at a specific hour.

- Taking into account the terrain, interpolate u and v at 58m height. This

results in a plane of u and v components at 58 meters that covers all the domain.

- Extract from the plane the u and v components at the x,y position.

- Calculate wind speed and direction based on the u and v values obtained.

- Repeat the steps for the rest of the forecasting hours.

The same 50 contiguous GFS model runs that were used in the previous section, were the ones used for the WRF-ARW model to assess the results. Figures 3.14 and 3.15 show the WRF forecasts obtained for 48 hour into the future wind speeds. The correlation obtained was 0.73258, which is slightly less than the correlation of the global model. This could be related to the higher overestimation of the mesoscale model for the higher peaks and underestimation of the lowest. Overall, it can be observed that more detail is obtained with the mesoscale model comparing with the GFS forecasts.
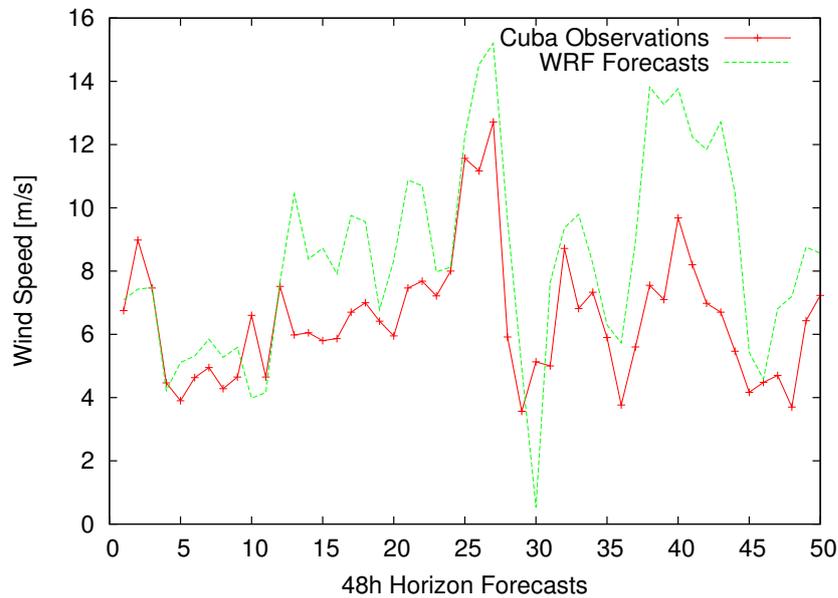
Figure 3.13: WRF-ARW wind speed forecasts and Cuba observations.

As the space resolution was still not high enough to obtain a forecast at the exact observation site directly from the grid, the forecasts obtained by the mesoscale
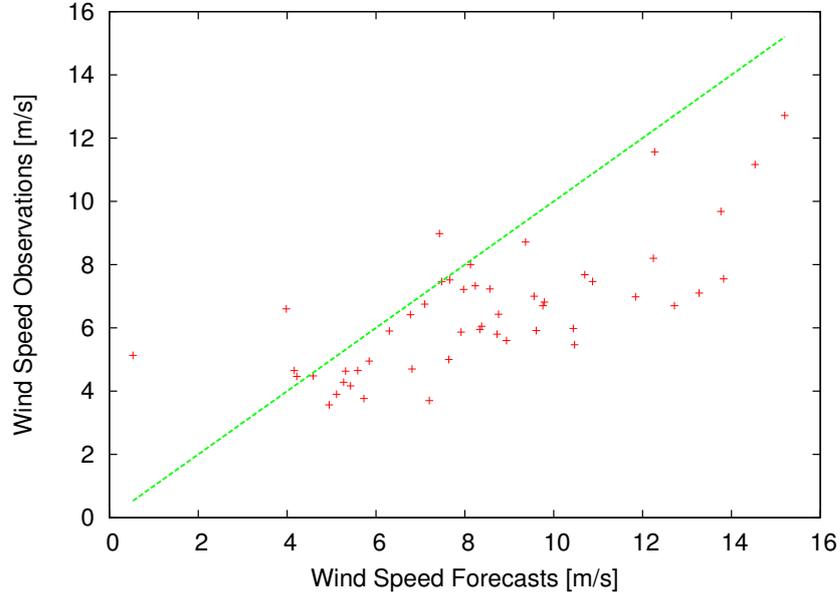
Figure 3.14: Correlation between WRF-ARW wind speed forecasts and Cuba observations.

model were interpolated using different techniques. The first one, Inverse Distance Weighting (IDW) (W. Luo, 2008), consists of calculating the wind speed at the observed location by adding the four wind speeds from the grid that surround the observation point and multiplying each windspeed by a weight depending on its distance to the observation point. IDW method is represented in Equation 3.1.

$$v_0 = w_1 v_1 + w_2 v_2 + w_3 v_3 + w_4 v_4 \tag{3.1}$$

where

$$w_i = \frac{\frac{1}{d_i^2}}{\sum\limits_{j=1}^{4} \frac{1}{d_j^2}} \tag{3.2}$$

The variables v_1, v_2, v_3 and v_4 represent the four forecasted wind speeds extracted from the mesoscale grid. v_0 is the new calculated wind speed at the observation site and d_i is the distance between v_i and the observation site. The

69
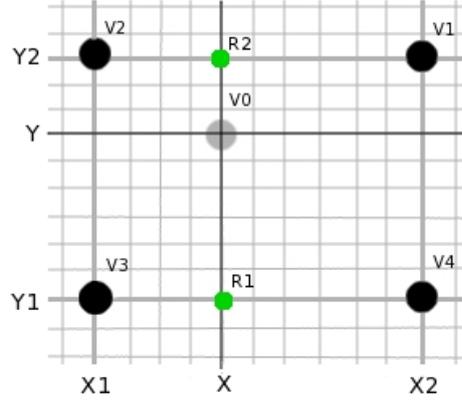
Figure 3.15: v_0 is calculated by multiplying each velocity in the grid by the area at the opposite to it.

second method, Bilinear Interpolation, consist of calculating the wind speed at the observed location by first performing linear interpolation in one axis, and then on the second axis, using the four surrounding points of the grid (see Figure 3.16). The interpolation in the x-direction is done by calculating:

$$f(R_1) = \frac{x_2 - x}{x_2 - x1}v_3 + \frac{x - x_1}{x_2 - x_1}v_4 \tag{3.3}$$

$$f(R_2) = \frac{x_2 - x}{x_2 - x1}v_2 + \frac{x - x_1}{x_2 - x_1}v_1 \tag{3.4}$$

After the linear interpolation is done in the x-direction, the values obtained are used to interpolate in the y direction as follows:

$$v_0 = \frac{y_2 - y}{y_2 - y_1}f(R_1) + \frac{y - y_1}{y_2 - y_1}f(R_2) \tag{3.5}$$

These two steps are simplified in Equation 3.6:

$$v_0 = \frac{1}{(x_2 - x_1)(y_2 - y_1)}(v_1(x - x_1)(y - y_1)+$$
$$v_2(x_2 - x)(y - y_1)+$$
$$v_3(x_2 - x)(y_2 - y)+ \qquad (3.6)$$
$$v_4(x - x_1)(y_2 - y))$$

The third method (Best Weights) consists of combining the four surrounding wind speeds giving a specific weight to each of them. The mathematical representation is as follows:

$$v_0 = w_1 v_1 + w_2 v_2 + w_3 v_3 + w_4 v_4 \qquad (3.7)$$

where, as with the IDW method, the sum of the weights is one. In order to find the values of the weights, past data is used such that the values of the unknowns will be those that minimise the error between the forecasts and the observations in a given sample of past data. Once the unknowns are found by solving a system of four linear equations, they could be used to calculate new forecasts.

The fourth mathematical method (Best Coefficients) is similar to the previous one, with the only difference that the sum of the weights, which in this case are called coefficients c_i, do not need to sum up to one.

$$v_0 = c_1 v_1 + c_2 v_2 + c_3 v_3 + c_4 v_4 \qquad (3.8)$$

These methods are evaluated on Chapter 5 and compared to a GP approach introduced in that same chapter.
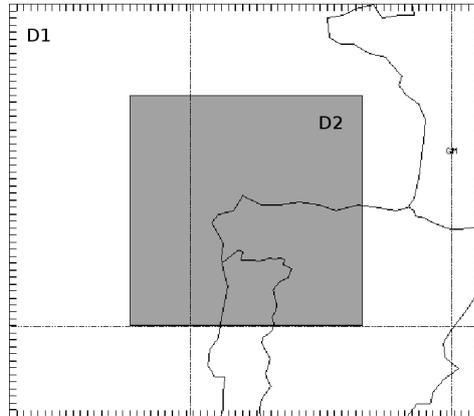
Figure 3.16: WRF-ARW two domain setting. The second domain (grey area) is centered in the point of interest, which is Galicia, Spain.

3.5   Wind Speed Forecasts at Sotavento Experimental Park

To generate the meteorological forecasts at Sotavento, global forecasts from the GFS model as well as terrestrial data were used. Both observations at Sotavento and GFS forecasts were available for the year 2012, so the first 6 months of the year, from January to June 2012 were used. Although there was more data available, the study was restricted to the amount of mesoscale data that could be produced locally, as this takes time. Sotavento provides 10 minute and hourly wind speed, direction and power observations so there was no need to average data as it was done for the Illinois case. Figure 3.17 shows the domain settings that were used. As shown in the figure, the model was set to run in two domains. The first domain, which covers a major part of Spain, has a resolution of 30km x 30km and results from the first integration of the WRF model from the GFS grid (111km x 78km). The second domain, which is centered on Galicia, the area of interest, has a resolution of 10km x 10km and is obtained by a second model integration that uses the first domain as boundary conditions. The parameterisations used for Sotavento were the same as the ones used for Illinois. Each run took approximately 3 hours. The six month period was not entirely forecasted as there was some data missing from the GFS model.

72
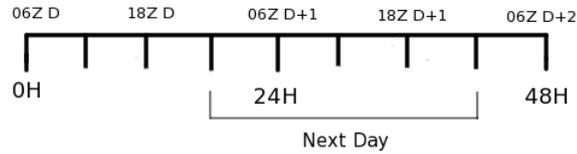
Figure 3.17: WRF-ARW 48 hour horizon in a 06Z run started at day D. The next day forecast corresponds to those values 19 to 42 hours. into the future.

For the Sotavento wind farm, only 06T runs of the GFS model were used. The reason for this was to ensure there was a consistent error for each day forecasted as it was seen in the previous GFS analysis that a 48 into the future forecast had a different level of error when using different runs. For each 06T GFS run, a WRF-ARW run was executed, producing a higher resolution forecast in time and space. Those values forecasted for the next day (19 to 42 hours into the future), as shown in Figure 3.18, are the values of interest for the DA power forecasting.

From the model output, wind speed and wind direction forecasts were to most relevant to consider for power prediction. To extract this data, another NCL script similar to the one used for the Illinois case was used. The data extracted was from the closest grid point and neighbour points to the location of the wind farm. This time forecasts were extracted at 45 meters height, which is the height of the anemometer at the wind farm.

3.6 Summary

This chapter has presented in detail how meteorological and observation data was obtained and processed. The amount of data that was produced with WRF-ARW was restricted mainly by two factors. One was the availability of observation data for the Illinois case study. The other factor was the time it takes for the model to run, restricting our time period to six months for the Sotavento case study. The wind speeds from Illinois will be used on Chapter 5 in order to assess the quality of the numerical model and the downscaling procedures at the three sites. The observations from Sotavento and numerical predictions of the area of Galicia will be used in Chapters 6 and 7 for wind power forecasting and ramp characterisation.

Chapter 4

Genetic Programming: an Approach for Symbolic Regression

4.1 Introduction

This chapter introduces Genetic Programming (GP) (Koza, 1992) as an approach for regression. This bio-inspired technique evolves a population of possible models that represent the relationship between an input or set of input variables and an output. The advantage of this approach over other regression techniques is that no assumptions on the model structure need to be made. GP can provide an alternative for the modelling of wind speed and wind power using numerical weather prediction model information. The layout of this chapter is organized as follows: Section 4.2 introduces the approach and gives details on the model representation and genetic operators necessary for evolution. Section 4.3 presents in detail the implementation of the algorithm. Finally, Section 4.4 presents a summary and the conclusions.

4.2 The Genetic Programming Approach

Getting a computer to solve a problem without explicitly being programmed would require a process through which the computer is able to build computer programs. These computer programs would perform operations, computations and iterations on different variables to solve a given problem. The construction and search process of computer programs, which is referred to as Genetic Programming, looks at the space of all possible computer programs and selects the desired program which best suits the problem. Looking at the complete search space would be a very exhaustive process, thus, the search needs to be done in an intelligent way (Koza, 1992).

Genetic Programming is a biologically inspired computation technique based on the evolution of individuals over time, through events such as crossover and mutation, which progressively refines them into better individuals. In GP, instead of evolving a population of binary chromosomes, as used in genetic algorithms (Goldberg, 1989a), a population of programs (in a binary tree layout) is evolved, each program representing a set of instructions to solve a specific problem. GP, like nature, is a random process, which cannot guarantee results but is that randomness which can lead it to escape traps, which deterministic methods may be captured by (Poli et al., 2008).

The range of real-world problems where GP has been applied is quite diverse. In general, GP has been used for problems where the following characteristics are present (Poli et al., 2008): the relationship between relevant variables is unknown; situations where there is a way to test solutions but it is difficult to derive them from the underlying theory; where the problem does not have an analytical solution and an approximation is enough; where the size and form of the solution is not known. Specifically, some of the areas were GP has been used are medicine, bioinformatics, computational chemistry, industrial process control, financial trading, image and signal processing, data modelling, symbolic regression, among others (Oakley, 1994; Poli et al., 2008; Tsakonas, 2006).

The prediction of wind power at a wind farm involves the development of mathematical models to estimate the power output of the wind farm. As many different factors such as the landscape, wind direction, temperature, pressure may
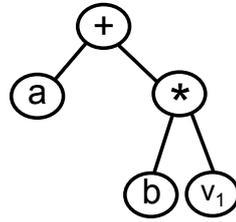
Figure 4.1: Example of a tree expression of the program $a + b * v_1$.

affect the power output, finding a mathematical model for a specific wind farm is a complex task. Genetic programming provides a means to find these models without knowing the shape or size of the problem, as it grows dynamically, compared to other approaches like genetic algorithms where the size of the solution is fixed.

4.2.1 Representation

Symbolic regression via GP is a non-parametric non-linear regression technique that, through evolution, looks for an appropriate model structure and model parameters that best fits a given sample of data (Kotanchek et al., 2010). This is different from conventional regression techniques that assume a certain model structure and estimate the optimal parameters. Symbolic regression involves finding a mathematical expression that fits the given sample data. In GP, the mathematical expression can be viewed as a computer program that takes as input the values of the independent variables and produces the values of the dependent variables. The structure of a program is presented in Figure 4.1.

As shown in the figure, a GP tree is formed by a set of terminals and functions. The functions may be basic arithmetic operators $(+, -, *, /)$, standard mathematical functions (sine, cosine, logarithmic, exponential), logical functions or domain-specific functions. The terminals may be a constant or any problem-related variable (b and v_1 in Figure 4.1). An initial population of randomly generated computer programs composed by these terminal and function sets will be created as a first step in the GP process. In this research the set of operators explored are the basic operators$\{+, -, *, /\}$, and the *sine*, *cos*, *exp* and *log* functions.
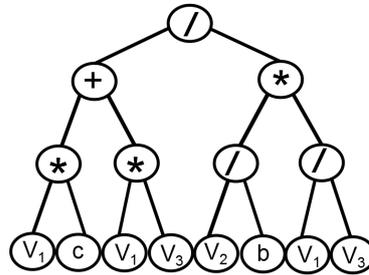
Figure 4.2: Example of a tree created with the full method. Variables v_i represent wind speed at different locations; variables b and c are constants.

4.2.2 Tree Initialisation

The creation of the initial population is achieved by generating each individual (tree) randomly like in other evolutionary algorithms. There are different methods for creating trees randomly. Two of the earliest methods, proposed by Koza (Koza, 1992), are the *full* and *grow* methods. Both methods are restricted by one parameter which is the depth. The *depth* of a tree is calculated by counting the number of nodes that need to be traversed from the root in order to reach the deepest leaf. This parameter needs to be set experimentally. The *full* method consists then in adding nodes randomly generated from the function set to the tree until the given depth is reached. The last level of the tree is constructed with nodes generated from the terminal set.

The *grow* method, unlike the full method, can build the tree by adding nodes generated from both function and terminal sets. The only restriction is that the root has to be always chosen from the function set. This adds variability to the shape of the trees. Nodes containing terminals would stop the growth of that branch even if the maximum depth is not reached. If the maximum depth is reached then only nodes from the terminal set are added. Figures 4.2 and 4.3 present an example of trees generated by the full and grow methods respectively.

Koza proposes a combination of the full and grow methods called *ramped half-and-half* in order to provide a wider range of sizes and shapes to the population. In this method, half of the population is created by the full method and the other half by the grow method. The ramped half-and-half has become one of the most common ways to create initial GP populations, however other proposed
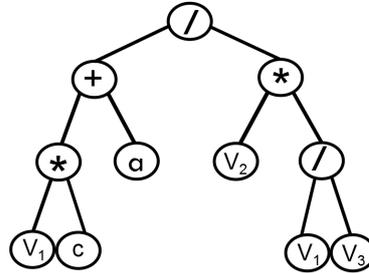
Figure 4.3: Example of a tree created with the grow method. Variables v_i represent wind speed at different locations; variables a and c are constants.

approaches for this can be found for example in (Bohm and Geyer-Schulz, 1996; Langdon, 2000).

4.2.3 Fitness

Following the theory of evolution, individuals with a high fitness have a high possibility to survive. In the GP approach, each program or individual of a population will be ranked (fitness measure) depending on how well it performs in solving the problem. This fitness can represent different things depending on the situation to be evaluated. For example, the fitness could express how accurate a program is for recognising or classifying objects. It could also represent the compliance of a structure or the time required to bring a system to a desired state. For symbolic regression, a common approach is to use fitness as a form of error between an input and an output. In this type of system, the best fitness corresponds to the lowest numerical value and, hence, the function is a *cost* rather than a fitness one.The error can be calculated using different criteria like the BIAS, the MAE, the RMSE, among others. In this research, the RMSE is used as a criteria to evaluate the cost of a model. Although taking the squared root would have no effect if this were the complete cost function, it has a substantial impact when a penalty is added to the overall cost. Additionally, the MAE is calculated from the best selected models to have a second measure to assess the accuracy.

To calculate the cost based on the RMSE, a collection of s data points that will be fitted or approximated is needed. The squared error between the model

output and these points is calculated, then averaged over the total number of points in the collection and finally the square root of this average is calculated. Finding an appropriate collection of points or *training set* in order to produce good generalisation models and reduce overfitting may be not be straight forward when the computational effort in training the models is an important factor. Although it is important to use a training set large enough to cover the search space to avoid overfitted solutions, very large training sets may increase computation time which might be critical in certain applications. The appropriate size of the set would need to be determined experimentally as it may vary according to the problem.

4.2.4 Genetic Operators

Creating an initial population by a random process would not be enough to produce the best program to solve a specific problem, unless the problem is simple and small. However, some of the individuals in the initial population may have characteristics that make them fitter than others and that could be used to produce better individuals. GP follows the Darwinian principle of reproduction and survival of the fittest, using genetic operations such as reproduction, crossover and mutation. The reproduction operator consists on selecting fit individuals, according to a specific criterion, and copy the selected individual into a new population (the next generation). The crossover operation creates variation in the population by producing new offspring from the combination of parts of fit individuals (parents). Each parent is selected according to a criterion, and could be of different size and depth. A typical criterion for selecting individuals in GP is the *tournament selection*, which consists in selecting a number of individuals at random from the population and choosing the best of them to be a parent. This procedure is done twice to obtain two parents which will be recombined to generate offspring. In each parent, a random crossover point is selected. The random point in each parent determines the root of the subtree that will be replaced with the subtree of the other parent. An example of crossover is shown in Figure 4.4.

The mutation operation can introduce diversity to the population when the algorithm is converging prematurely. This operator introduces random changes
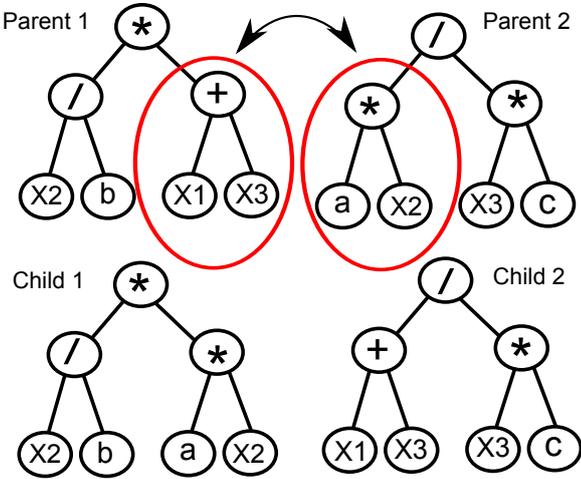
Figure 4.4: Crossover operation.
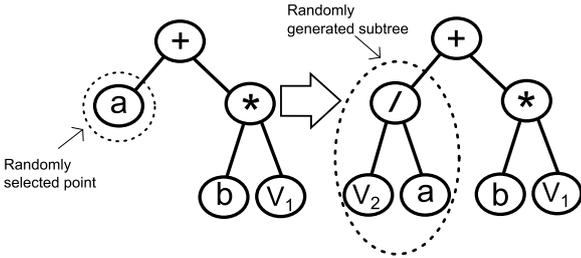
Figure 4.5: Mutation operation.

in individuals according to a probability. It starts by selecting a random point within the tree. This mutation point can be either a function or a terminal point. The mutation operation removes everything from the selected point and inserts a randomly generated subtree in this point. The size of the subtree inserted is controlled by a parameter that specifies the maximum size. Figure 4.5 presents an example of applying the mutation operator.

4.2.5 Overfitting

As any other Machine Learning (ML) technique, achieving good generalisation is one of the most important goals of the GP approach (Naik and Dabhi, 2013-12-01T00:00:00). Failure to generalize, or overfitting, happens when a program performs well in training cases but has a poor performance in unseen cases. It

has been noted that poor generalisation of programs is commonly related to their growth in complexity (increase in the number of nodes) as the number of iterations increases (Langdon and Poli, 1997a). Programs can grow large pieces of code that does not make any significant improvement in the overall fitness of the program. This is what is called the bloat problem (Poli et al., 2008).

There are three classic theories that explain the bloat problem:

1. The replication accuracy theory (McPhee and Miller) states that the success of a GP depends on its ability to replicate accurately, having offspring that are functionally the same. If the function set for a problem contains functions easily combined to form large, semantically irrelevant trees, the presence of these trees often inhibits the discovery of a correct individual.

2. The removal bias theory (Soule and Foester, 1998) states that code growth can occur whenever operations which remove and replace a variable sized section of code, like crossover or subtree mutation, are used. Inactive code (bloat code that if replaced, does not change the fitness) usually forms subtrees in the low part of the tree. These subtrees might be small in size at the early stage of the evolution. However, when these inactive pieces of code are replaced during crossover or mutation, there is nothing to prevent them from becoming arbitrarily complex because the fitness is not affected.

3. The nature of program search spaces theory (Langdon and Poli, 1997b) states that after a certain size, the distribution of fitness does not vary with size. Because there is a greater amount of larger programs with a given fitness than short programs of the same fitness in the solution space, the probability that GP selects larger programs is greater over time.

Due to the lack of a clear explanation to the bloat problem, several empirical methods to solve it have been proposed. One approach is to limit the size or depth of the trees during crossover. If the offspring violates the size or depth limit rule, then the parent can be returned instead of the new offspring. However this can lead to a population full of programs that nearly infringe the rule, which is not desired. One way to avoid this is to, instead of returning the parents, return the offspring with a fitness of 0, preventing it from being selected for crossover in

future generations. Another way to avoid this situation is by selecting a different crossover point after the failed first attempt. If the new crossover also infringes the rule, then a new crossover point can be selected until a certain number of trials. If after all the attempts, no valid offspring was successfully created, then a new pair of parents will be selected. Defining the limit of the size or depth of programs is not trivial. Setting a small limit could prevent programs to express any solution to the given problem.

Another empirical approach to the bloat problem is having anti-bloat operators, which constrain the choices made during the execution of the genetic operations. Size fair crossover (Crawford-Marks and Spector, 2002) and size fair mutation (Langdon, 2000) are two approaches based on this idea. Size fair crossover restricts the crossover point of the second parent according to the size of the subtree selected from the first parent. Size fair mutation generates random replacement subtrees according to a distribution, in such a way that, when replacement is done, the size of the tree is not altered.

The parsimony pressure (Koza, 1992; Zhang and Mühlenbein, 1995), another technique to avoid bloat, penalises the fitness of a program according to its complexity reducing the probability for it to be selected in future generations for crossover. The penalisation is done by adding to the regular computation of the fitness function a coefficient called parsimony coefficient, which is multiplied by the size of the program and then added (or substracted according to the problem) to the calculated fitness. In (Amil et al., 2009), a theoretical study about the bloat problem is presented and a parsimony pressure based method is proposed. This method consists of computing the complexity factor of the tree by approximating the $Vapnik - Chervonenkis$ dimension (Amil et al., 2009), which is the measure of the capacity of a classifying algorithm. For symbolic regression, this is done by calculating the upper bound on the generalisation error. The complexity factor is used as a penalisation of the fitness of each tree. The mathematical representation of this penalty function is shown on Equation 4.1.

$$f = \frac{1}{s} \sum_{i=0}^{s} e(i) + k \left(\frac{(t^2 log_2 (t))}{s} \right)^{\frac{1}{2}}$$ (4.1)

The first term of the fitness equation 4.1 is the sum of the errors between

the model output and the desired output in the test set (test set size = s). The second term is the complexity factor, where t is the number of nodes of the program tested and k is a trade-off weight that allows to control the level of pressure of the complexity factor. A small value of k (e.g. k=0.0001) would be translated into low complexity pressure, and higher values of k (e.g. k=0.1) will result in a strong pressure to the penalisation.

In some implementations of the GP algorithm, a second set, the *validation set*, is used as a way to ensure generalisation. The evolution process will depend on the training set, while the validation set would help to verify if the best solutions are also good solutions using *unseen data* (validation set). If during the evolution process the best solutions begin to learn the training set too well that their performance on the validation decreases, then the last solution that was best on both sets will be the output solution of the algorithm.

All these approaches to avoid bloat, as mentioned before, prevent the GP algorithm from producing programs that are complex and overfitted. Other approaches based on Random Sampling Technique (RST) have been proposed to manage overfitting (Gonçalves and da Silva, 2011; Liu and Khoshgoftaar, 2004). RST basically consists of using a random subset of the training set instead of the complete set, and calculate the fitness of the programs based only on this subset. The subset can be re-selected every iteration or at every t iterations. The programs that survive through the generations are those performing reasonably well on different subsets. These programs will be capturing the underlying relationship of the data instead of overfitting it.

The GP based approach developed in this research explores a combination of random sampling to achieve generalisation with a parsimony pressure to avoid the evolution of large programs. The reason for using two overfitting strategies is that, if only the RST strategy is considered, even getting good generalisation models, the problem of bloat or code growth can occur (Liu and Khoshgoftaar, 2004). This is due to the concept of introns, which are useless pieces of code that do not affect the fitness of the program but contributes to the development of large solutions.

4.3 The Implementation

The GP algorithm implemented in this research works as follows: First, an initial population p of n trees is created, selecting random combinations of variables and operators with the *ramped half-and-half* method. The fitness of each tree is calculated in two steps. First, the mathematical expression is evaluated for every point of the training set in order to calculate the RMSE between the model outputs and the real observations. Then, a penalisation is added according to the complexity of the expression (Equation 4.1). Once the tree with the best fitness is identified, the algorithm iterates x number of generations. At each iteration, a new population is created by copying the best tree of the previous generation and the rest of the individuals are created by means of selection, crossover and mutation operators. The new population replaces the previous population and the best tree of the new population is identified. This process is repeated until the specified number of generations has been reached and the best tree on the training set and validation sets are obtained. A flow chart with this algorithm is shown in Figure 4.6.

The algorithm was implemented in C++ using the GPC++ class library (Fraser and Weinbrenner, 1997) by Adam P. Fraser and Thomas Weinbrenner as a starting point. With these, different algorithms can be developed according to the application problem; image processing, process control, regression, time series prediction, etc. To implement the symbolic regression algorithm, the steps described previously were used. There were several aspects to address in order to ensure the evolution of valid models. One was the evaluation of each individual. As each tree represents a mathematical expression, the tree needs to be evaluated according to the function and terminal set that is defined during experimentation. Special situations need to be handled like the division by zero and the natural logarithm of negative values to avoid invalid expressions. Another important aspect of the implementation to consider is that the number of input variables may vary according to the application problem or type of experiment. The algorithm was implemented in such a way that few changes need to be done in order to use the algorithm with any number of input variables. Another aspect to consider is the strategy that will be used to prevent overfitting and bloat. It is
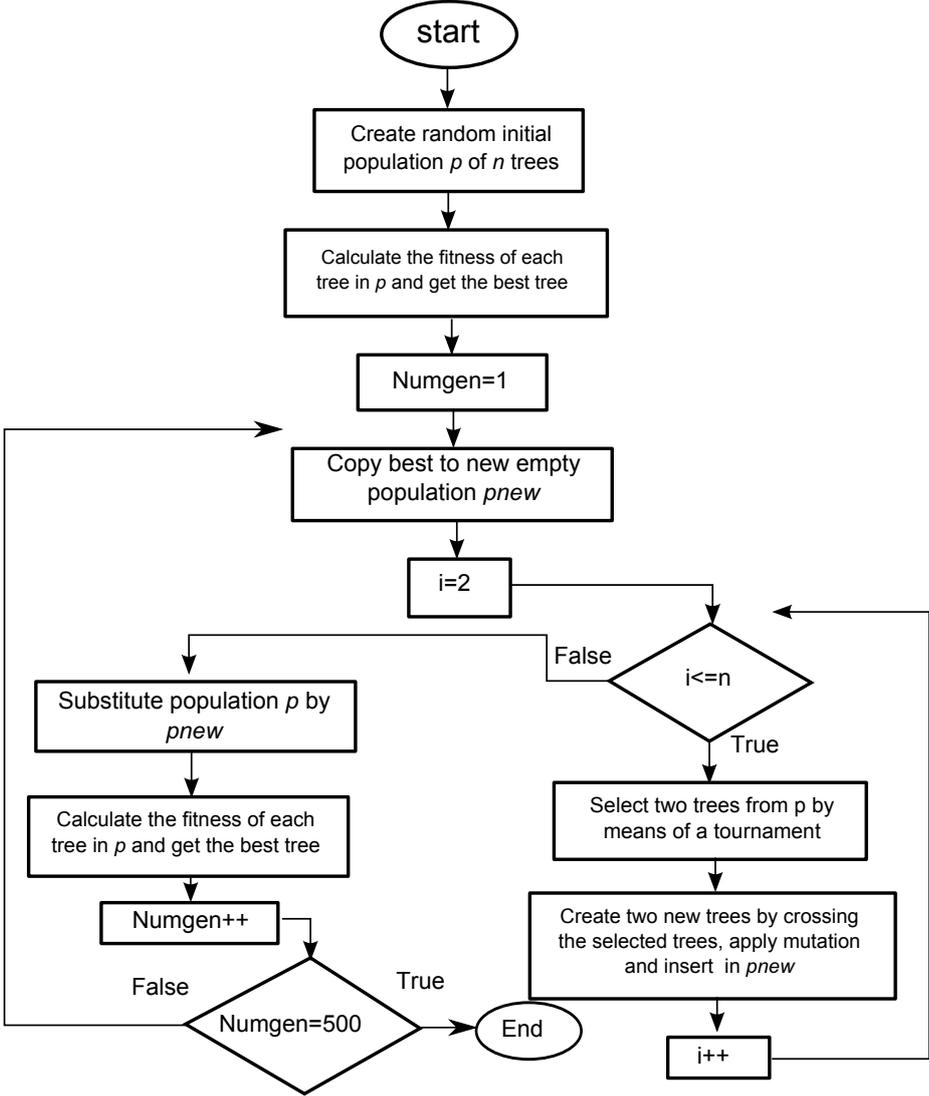
Figure 4.6: Genetic Programming Algorithm.

important to consider this situation in symbolic regression to avoid the evolution of models which contain terms that are zero and therefore do not contribute to the understanding of the variable being modelled. This problem was handled by applying a parsimony pressure that was determined experimentally. In addition to these three aspects, there are several parameters that need to be tuned in order to apply the algorithm. Parameters, such as the number of generations, size of population, etc., need to be determined experimentally. Further details on the tuning of these are presented in Chapter 5.

In order to validate the implementation of the algorithm, some experiments were designed to apply GP as modeller of different jet engine parameters. The problem and results are presented in Appendix C.

4.4   Summary

This section presents genetic programming as an approach to symbolic regression. In the one hand, the amount of decisions that need to be made during the implementation, like the fitness function, the overfitting strategy, the number of inputs, the size of training sets, among others, might be a complex process. As any other evolutionary algorithm, it needs tuning, it is not a plug and play strategy. However, as it was found with the jet engine application, this approach is able to provide some understanding of the relationship between variables which other machine-learning techniques might not. Compared to neural networks, GP provides a mathematical representation of the solution which can be used to understand the processes being modelled. It also has an advantage over genetic algorithms, in the sense that the solution is not restricted in size and form. As there are several aspects to tune, there is always different settings that could be explored to improve results. The following chapter introduces the use of GP as a final downscaling step from the numerical model grid to a specific location avoiding the execution of high resolution NWP model runs. This is the first step towards wind power prediction.

Chapter 5

A Genetic Programming Approach for Wind Speed Downscaling

5.1 Introduction

As it has been mentioned previously, numerical weather prediction models are essential for wind speed/power forecasting at horizons larger than 6 hours, where the Persistance model is no longer able to provide robust forecasts (Monteiro et al., 2009). There are two main aspects of numerical models that allow improvement. One aspect is the high resolution runs. In order to provide a wind speed forecast at an exact observation point or as close as possible, the resolution of the numerical model grid has to be very high. The high resolution runs are computationally demanding and therefore they may take too long to complete. A 0.5km resolution run (e.g. 5 domains) may take up to 48 hours to complete when resources are limited (Louka et al., 2008). In addition, it has been found that the benefits gained beyond a 6km resolution are not always worth the computational expenses as coarser domains (12km and 6km) capture mesoscale features of the airflow satisfactorily. GP as a downscaling technique could provide a way to find the relationship between the existing low resolution information from the numerical model and observations at a specific location as it is a non-parametric

technique. In addition, its model transparency can provide a quick understanding of the phenomena being modelled. This chapter presents the application of GP for wind speed downscaling as a first step towards wind power forecasting. The results presented were published as a conference paper in (Martínez-Arellano et al., 2012).

The second aspect to study are the so-called "misplacement errors". NWP models are good at modeling the spatial phenomena that cause spatial variations in wind energy flux, yet it may misplace these phenomena relative to the physical world (Cutler et al., 2009). The spatial variability of wind speeds in a NWP system could give a useful characterisation of uncertainty in future wind power. Most of the current wind power forecasting tools focus on the closest grid point to the observation site, while *neighbour* points could provide information of changes in wind speed which were not modelled on the closest point of the grid. This misplacement error will be addressed further in this thesis on Chapters 6 and 7.

The layout of this chapter is organized as follows: in Section 5.2, a basic GP approach for statistical downscaling is introduced. Section 5.3 presents an improvement of the basic GP downscaling approach by varying the function set. Sections 5.4, 5.5 and 5.6 extend the experimentation by exploring different aspects of the approach such as the stop condition, the size of the training set and the use of a sliding window for re-training. Section 5.7 presents a discussion and conclusions on the results obtained.

5.2 Basic GP Approach for Statistical Downscaling at Different Locations

Genetic Programming, as other machine learning techniques used for statistical downscaling like neural networks (Salcedo-Sanz et al., 2009), works as black box, as a-priori knowledge of the system is not required. However, in contrast to these techniques, it is able to provide a direct model of the system. This is important as the mathematical representation might provide some insight about the process being modelled. GP does not restrict the size and form of the solutions, which is an advantage over regression methods where the shape of the model must be

known in advance. For short-term forecasting, as the forecast horizon becomes larger, it is important to take into account numerical weather prediction models. Symbolic regression via genetic programming will be used as the final downscaling step from the mesoscale grid to the location of the wind farm to avoid the execution of high resolution model runs which may not add any additional improvement to the lower resolution forecasts. The algorithm will be trained to "learn" from the error of past forecasts to be able to compensate it in future forecasts. In addition, genetic programming is able to provide an ensemble of potential solutions after training. This is useful and can potentially be used as a means to quantify uncertainty of the modelled variable.

An exhaustive experimentation phase was carried out to analyse the potential of genetic programming for statistical downscaling and to compare it with the mathematical methods introduced previously. The objective at this stage was to improve the wind speed predictions obtained with the numerical model to convert these in a second step to wind power predictions by applying a power curve.

Many variables could be used as inputs for wind speed downscaling. In addition to wind speed, atmospheric variables such as wind direction, temperature, solar cycle, among others, could be used, as many approaches currently do (Monteiro et al., 2009). However, with the results obtained while testing the simulation of the jet engine parameters, it has been found that a selection of input variables for the training process is not a simple and straight forward task. Using a large set of input variables and letting the algorithm find out which one is the most relevant is not the best approach and would not guarantee any success in the first attempt. In this situation, the application of a feature selection algorithm may enable the regression process to operate faster and more effectively. There are three general schemes for feature selection: the embedded, the filter and the wrapper scheme. A survey on feature selection techniques can be found in (Chandrashekar and Sahin, 2014). In this research a naive forward sequential wrapper scheme has been adopted. This approach wraps the feature selection around the induction algorithm to be used, using cross-validation to predict the benefits of adding or removing a feature from the subset used. Forward selection greedily adds attributes that, when added to the feature set, yields to structures that generalise better. For this problem in particular, the first step was to work only

with wind speed and then to add other variables progressively. These other variables included wind direction, temperature and solar cycle, as these are the most common variables used by state-of-the-art methods.

The observations available from the state of Illinois were used for the experimentation. Three different sites could be tested for wind speed with the same numerical model run. This was very useful as each run takes approximately 3 hours to complete and producing enough training data for different observation sites at totally different locations in the globe would have been a long process.

There were two main stages during the experimentation. The first stage used observations at 10 meters height as the numerical model provides a specific variable to model wind speed at this height. In a second stage, observations at 58m were used, as this height is closer to a typical hub height of 80-120 meters. Although it would be more useful to work with observations at hub height, these are still not commonly available as most anemometer measurements are at 50m or lower.

The size of the training set plays an important role in machine learning algorithms so this was an important aspect to explore. Large training sets might be the best option to ensure generalisation. However, very large training sets may require a greater computational effort, so it is critical to find what is the minimum training size that could ensure that generalisation. The starting point was to take a small training set of 50 points to assess the capabilities of the approach; the same 50 points used to assess the GFS forecasts in Chapter 3. Later on, larger training sets were tested as well as the use of validation sets to avoid overfitting. A similar approach was presented by (Sweeney et al., 2011), where experiments started from a 2-day window size to up to 30 days.

The first set of experiments was conducted as follows. From the 50 forecasts generated using WRF-ARW (from the 6th to the 18th of February 2011), the four 10 meter wind speed predictions of the grid that surrounded the observation sites were extracted, considering only those values forecasted 48 hours into the future. At this point, only wind speed was selected as feature for regression. Hence the training set consisted of 50 sets (v_1, v_2, v_3, v_4) each one with its corresponding observation at the target site. The reason for considering only values forecasted 48 hours into the future is that these are more likely to have a larger error than the

rest of the forecasted hours within the 48-hour horizon. So at this stage, the interest was to observe how the GP approach would handle specifically these values at $t + 48$, which are probably the worst case scenario. A test set, which corresponds to the five following days after the period considered for training (from the 18th to the 22nd of February, 2011), was used to assess the generalisation achieved by the models when applied to new data. The concept of generalisation would only apply to the immediate available forecasts after the training period. This does not mean that the models obtained could be used for any time of the year. It is assumed that the immediate days after the training period would have similar characteristics and conditions. At this point there is no notion about what the best training size is and if it is site dependent.

The crossover probability was set according to a small experiment consisting in varying this parameter and observe the overall quality of the programs. Figure 5.1 shows how small crossover probabilities tend to slightly decrease the overall quality of the solutions. High probabilities, despite the higher variability in the first quartile, can achieve better solutions and may produce less outlier values. With these results, it was decided to set the crossover control parameter to 100%. The mutation probability was set to 3% so that the new material could be occasionally incorporated into the solutions.

For such a small training set, it could be assumed that all of the points would be needed during training to obtain the best generalisation model possible. However, to avoid any assumption, different subset sizes were used. From the training set, the subset size values that were used were 50, 40, 30, 20, and 10. For each of these subset sizes, three different parsimony pressures k were used, $k = 0.1$, $k = 0.01$ and $k = 0.001$ as it was identified that those values have a visible effect in the size of the models. Each combination of this two parameters was executed 50 times, keeping track of the best solution at each run and taking an overall average of the best solution. In order to assess the impact of different parsimony pressures and random sub sets, the rest of the control parameters on the GP run were fixed. The same control parameters used for the jet engine problem were used here. A substantial consistency in the choice of the same control parameters in all experiments would help to eliminate any concerns that the success of the algorithm depends on random or coincidental choices in a
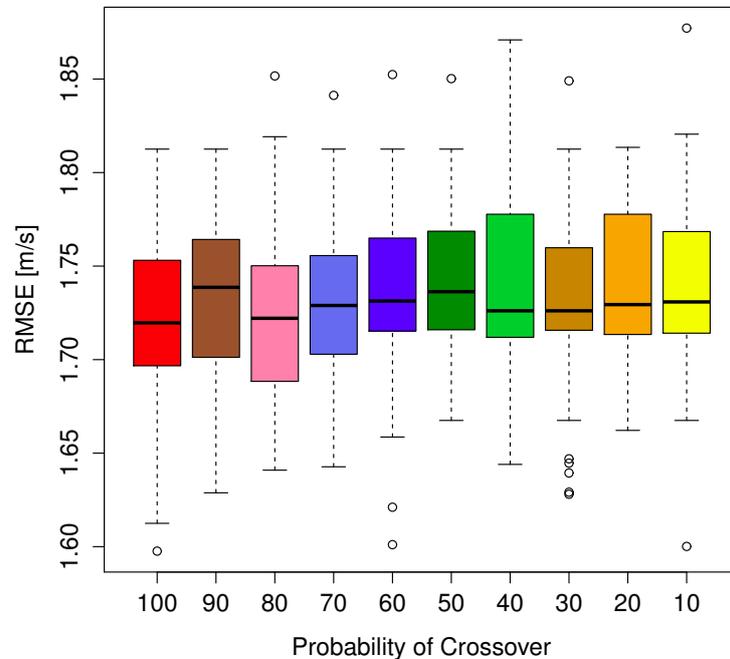
Figure 5.1: Quality of the best solutions on 50 runs of the GP varying the probability of crossover.

particular problem. These control parameters are shown in Table 5.1.

The first results using a pressure of $k = 0.1$ and different training subset sizes showed a clear site dependency. As it was expected, the best models on the training were achieved by using the complete training set. However, to achieve generalisation the best size was yet to be defined by the complexity of the site. At SIUE, the GP algorithm obtained a diverse quality of solutions on the test (unseen) data compared to the training data, obtaining the best average using a random training set of 40 points. At Cuba and Wilmington, a smaller variation in the quality of the solutions was observed. At Wilmington, there was an apparent improvement on the average fitness on test data when changing the training set to larger sizes. This opened the question if a better generalisation/quality could be obtained by increasing the training set for a size larger than the current maximum size.

Table 5.1: Fixed GP parameters used for the experiments.

Runs	50
Population	1000
Generations	100 and 500
Crossover operator	Standard subtree crossover, probability 100%
Mutation operator	Standard subtree mutation, probability 3%, maximum depth of new tree 17
Tree initialisation	Ramped Half-and-Half, maximum depth 6
Function set	+, -, *, /
Terminal set	v_1, v_2, v_3, v_4 and random constants
Selection	Tournament of size 20
Elitism	Best individual always survives

The next step was to decrease the parsimony pressure ($k = 0.01$) to see if a smaller pressure and more complex models could improve results without getting into an overfitting situation. The first aspect that was observed in these experiments was that the standard deviation of the best GP in the test data was in general smaller than previous experiments, meaning that the best programs were performing better on test data in all runs. This behavior was expected as using a smaller parsimony pressure can lead to more complex mathematical formulas, which can represent more accurate models. However, there was still the question if these could be further improved with the current training set. A final parsimony pressure of $k = 0.001$ was experimented. The results showed an overtraining in the models as the programs were getting very good at the training but were not performing well on the test data. Results using the complete training set are presented in Figure 5.2 for Cuba. The figure shows large variations on the quality of the results when using a small pressure. It also shows how a pressure of $k = 0.01$ leads to the smaller variability in the quality of the models. The tendency for most training sizes was that the models perform worse in test than in training. All figures obtained from SIUE and Wilmington data can be found in Appendix D.

With an adequate pressure parameter that could ensure good generalisation
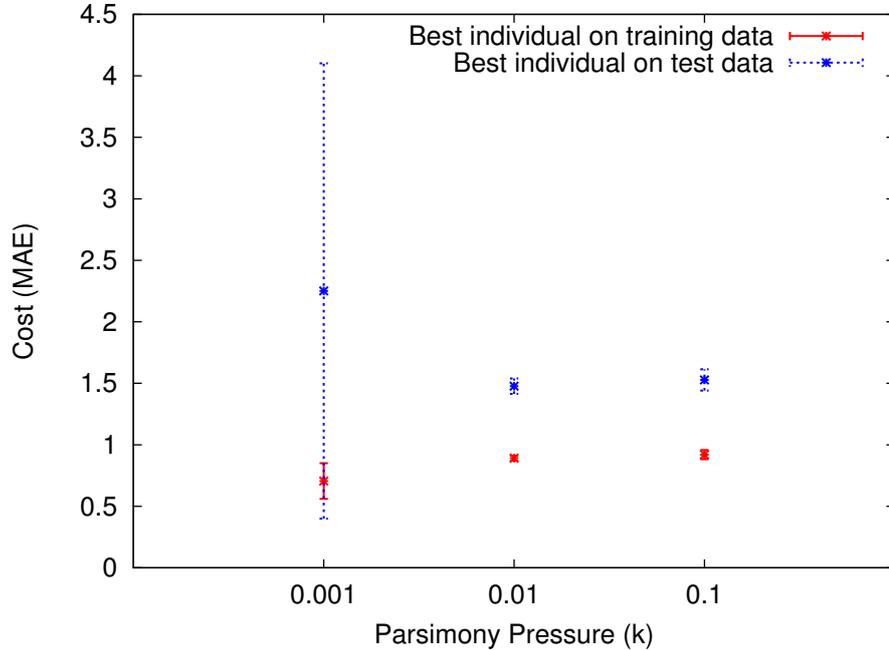
Figure 5.2: Average cost and standard deviation of the best GP at Cuba with the complete training set.

models, the approach, in its current basic implementation, could be compared to benchmark mathematical methods to be able to evaluate the performance of GP. Tables 5.2 and 5.3 summarise the results obtained with the basic GP implementation at the three observation sites over the complete training data set compared to mathematical methods presented in Chapter 3. The results of the GP approach were taken from the best of 50 runs with a parsimony pressure of $k = 0.01$, as this one achieved the best results. As the tables show, GP obtains the best results for the three sites over training data. In the case of test data, Table 5.4 shows that, in general, the results obtained with GP improve the previous methods due to its ability to generalise. It is interesting to note that the best result over new data is not obtained using the complete training set.

After finding the parameters that could allow a successful application of the algorithm to the downscaling problem, a second stage in the experimentation was focused on working with wind speeds closer to the wind turbine hub height, as the objective is to be able to convert wind speed forecasts to wind power forecasts. A

Table 5.2: MAE in meters per second of all downscaling methods over the training data.

Site	IDW	Bilinear Interp.	Best Weights	Best Coefficients	GP
Cuba	2.84723	2.84794	1.118538	0.900721	0.8891711(0.01, 49)
SIUE	1.8812	1.87679	1.630793	0.799778	0.7254781(0.01, 50)
Wilmington	2.78542	2.77693	2.530139	0.839859	0.76496(0.01, 50)

Table 5.3: RMSE in meters per second of all downscaling methods over the training data.

Site	IDW	Bilinear Interp.	Best Weights	Best Coefficients	GP
Cuba	3.27872	3.2765	1.386675	1.167496	1.153821(0.01, 49)
SIUE	2.35062	2.34336	2.121927	1.094980	0.974957(0.01, 50)
Wilmington	3.23827	3.23091	3.078297	1.059626	0.994099(0.01, 50)

Table 5.4: MAE and RMSE in meters per second of the best two methods over test data.

Site	Best Coefficients		GP(k,rss)	
	MAE	$RMSE$	MAE	$RMSE$
Cuba	1.4668495	1.761794	1.47404(0.01, 20)	1.712134
SIUE	1.042472	1.321985	1.01643(0.02, 20)	1.229757
Wilmington	1.3233674	1.569485	1.16398(0.01, 49)	1.38117

different performance of the GP could be expected as the airflow is highly affected by topographic effects and the quality of the numerical model predictions could be different (Ray et al., 2006). The wind speed predictions were obtained at 58m height as explained in Chapter 3. Similar training and testing sets were created as for previous experiments and different pressure parameters were used. The results in Tables 5.5 and 5.6 show that the GP approach was able to improve again the results of all the other techniques. It is interesting to observe that the best training sub set size to achieve generalisation in most cases is basically the complete set. This could mean that the error at 58 meters height of the numerical model could be larger than the one at 10m and the GP algorithm needs more data to get the best generalisation possible with that training set. This again leads to the question of possible improvement by increasing the amount of training data.

Table 5.5: RMSE in meters per second of all downscaling methods over the training data.

Site	IDW	Bilinear Interp.	Best Weights	Best Coefficients	GP
Cuba	2.83422	2.82729	2.156678	1.51196	1.282657
SIUE	2.54912	2.52729	2.120475	1.81081	1.326856
Wilmington	3.4226	3.41178	3.169136	1.576498	1.217808

Table 5.6: MAE and RMSE in meters per second of the best two methods over test data.

Site	Best Coefficients		GP(k,rss)	
	MAE	$RMSE$	MAE	$RMSE$
Cuba	1.04409	1.36384	0.86469(0.01, 50)	1.126627
SIUE	1.30877	1.50457	1.2604356(0.01, 40)	1.5825726
Wilmington	1.74079	2.11086	1.388853(0.01, 40)	1.826828

Figure 5.3 shows the relationship between the model output and observations at Cuba on training and test data using the best model found by the GP. The

correlation coefficient on training data is 0.60855 and on new data is 0.62860. This indicates the model is achieving a similar quality of the results when new data is presented and means that the parsimony pressure applied is maintaining the desired level of generalisation for 58 meter wind speeds.
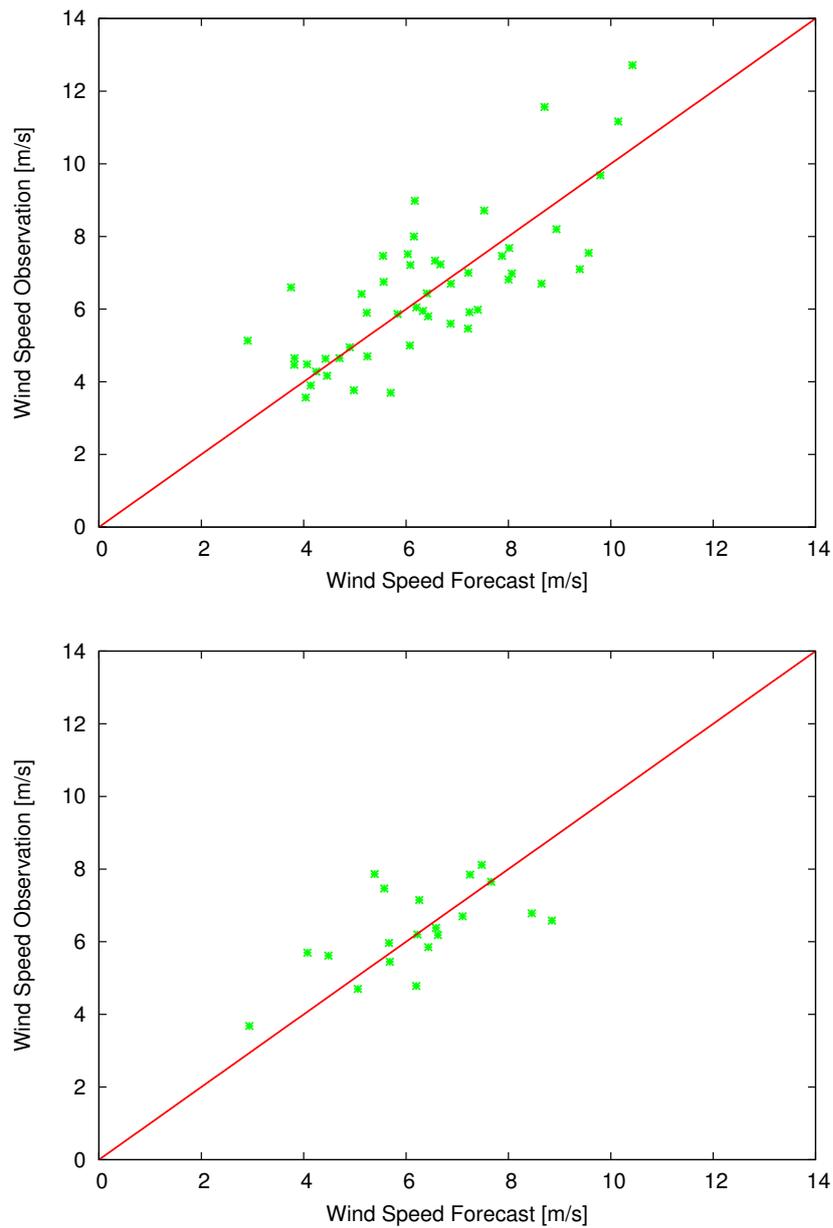
Figure 5.3: Correlation between the best GP model and wind speed observations at Cuba on training (left) and testing (right) sets.

As one of the main features of GP is that it outputs a mathematical representation of the model, it is interesting to see how the models for wind speed downscaling look. The results showed that models obtained with less training data (training size = 10) were more stable as most runs produced similar mathematical formulas. The quality of these solutions, however, was poor compared to models found using the complete training set. The models found at each site with the best set of parameters are shown in Table 5.7.

Table 5.7: Models found with the Basic GP implementation.

Site	Mathematical Representation of the Model
Cuba	$(v_4/v_1 * 0.570914 + v_3)/v_1 + v_3/(v_1 * 0.570914 + v_1) * v_4$
SIUE	$(6.58506/v_2 + 2 * v_3 - 2 * v_1)/1.38663$
Wilmington	$(3.5761 + 1.89108)/v_2 + v_2/1.89108 + 1$

The model for Cuba shows a combination of most of the input variables, compared to SIUE and Wilmington that tend to select v_2. This suggests that using only the closest point for a downscaling procedure might not be enough as the terrain conditions in each site are different. As there is a possibility that the numerical forecast contains misplacement errors, a forecasting approach based on several grid points might bring additional information.

5.3 Varying the Function and Terminal Sets

The GP implementation presented previously is able to construct models by means of the basic operators (+, -, *, /) and four wind speeds obtained from the WRF-ARW model at both 10 and 58 meters height. This is a basic implementation that could be potentially improved if other variables or operators were considered. As more functions or variables become available to the GP algorithm, it could be capable of constructing more complex programs that could represent better the downscaling procedure. This however, it is not a trivial task to perform. In order to investigate which variables and operators could improve the results based on the current data available, a set of experiments was designed.

The first aspect to study was the addition of other operators to the function set. For this, the logarithmic, exponential, sine and cosine functions were added to the set and experiments were executed in the same way as it was carried out with the basic GP implementation. Different parsimony pressures k and training set sizes rss were used, the same ones as in previous experiments, and each experiment was executed fifty times. The combination of logarithmic and exponential functions provided the best results. These are presented in Table 5.8 and 5.9.

Table 5.8: MAE and RMSE in meters per second of all downscaling methods over the training data at 58m.

Site	Best Coefficients		Basic GP		GP + log +exp	
	MAE	$RMSE$	MAE	$RMSE$	MAE	$RMSE$
Cuba	1.188412	1.51196	1.011829	1.282657	0.966050	1.222136
SIUE	1.276052	1.81081	1.034538	1.326856	1.054124	1.358985
Wilmington	1.152524	1.576498	0.907641	1.217808	0.920869	1.223930

Table 5.9: MAE and RMSE in meters per second of all downscaling methods over the new data at 58m.

$Site$	Best Coefficients		Basic GP		GP + log +exp	
	MAE	$RMSE$	MAE	$RMSE$	MAE	$RMSE$
Cuba	1.044095	1.36384	0.86469	1.126627	0.866885	1.122023
SIUE	1.308773	1.504599	1.260435	1.582573	1.232535	1.535281
Wilmington	1.740791	2.110863	1.388853	1.82683	1.3064	1.72599

The results obtained with the logarithmic and exponential functions improved the ones obtained with the basic GP implementation in the test set. These results are comparable with previously published results. In Salcedo-Sanz et al. (2009), the MAE obtained lies between 1.45 to 2.2 m/s, and the error obtained in the presented experiments lies between 0.86 and 1.30 m/s in average. Results obtained in Sweeney et al. (2011) in terms of RMSE lies between 0.96 and 1.89 m/s, which are comparable to the RMSE obtained by the GP which lies between

1.12 and 1.72 m/s. Figures of the average cost, standard deviation and correlation between the best GP obtained using logarithmic and exponential functions and observations in the tree sites can be found in Appendix C.
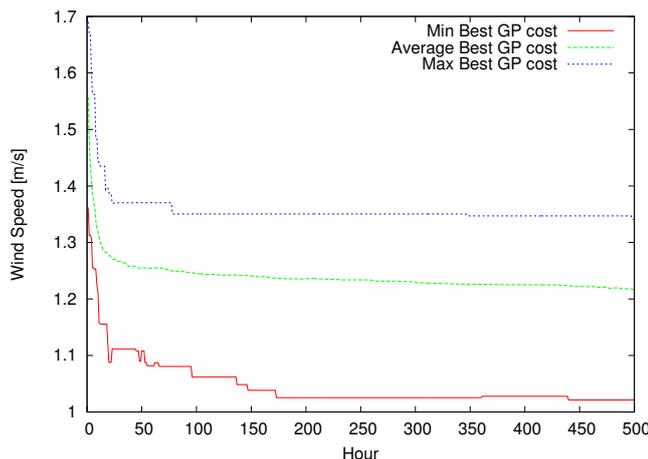
Figure 5.4: Minimum, average and maximum cost in m/s of the best GP on 50 runs at Cuba.

Figure 5.4 shows the minimum, average and maximum cost of the best GPs on 50 runs at Cuba. It can be seen that the worst GP found is closer to the average. In general, any solution that falls between the average and the worst cost, which are the majority, will provide a good model because the program is not overfitted. Over trained solutions are less frequent.

In terms of computation time, it could be observed that a lower parsimony pressure would increase the execution time of the algorithm as more complex models would evolve, requiring more time to evaluate them. Computation time was also influenced by the number of training points and the number of individuals in the population. In general, when using the complete training set, the cases with high parsimony pressure would run in 2-3 minutes while low pressures cases would run for up to 40-45 minutes.

After experimenting with the function set, new atmospheric variables were added to the terminal set to investigate if these could improve the model with the given training data. Seven types of experiments were conducted, using different combinations of variables at the four closest points additionally to wind speed: temperature, wind direction and solar cycle. In all experiments, logarithmic

and exponential functions were included in the function set. For each type of experiment, the same procedure as in past experiments was followed. Different values of k and rss were used and each experiment was executed 50 times.

Table 5.10: MAE in m/s on new data at 58m (WS = wind speed, T= temperature, WD = wind direction, SC = solar cycle).

Site	WS+T	WS+WD	WS+SC	WS+T+WD
Cuba	0.888935	1.08766	0.877463	1.019572
SIUE	1.295135	1.29692	1.265810	1.330596
Wilmington	1.315414	1.443895	1.313376	1.46964

Table 5.11: MAE in m/s on new data at 58m (WS = wind speed, T= temperature, WD = wind direction, SC = solar cycle).

Site	WS+T+SC	WS+WD+SC	WS+T+WD+SC
Cuba	0.8891106	1.01097	1.682021
SIUE	1.282913	1.313471	2.04603
Wilmington	1.331096	1.464797	2.39844

Table 5.12: RMSE in m/s on new data at 58m (WS = wind speed, T= temperature, WD = wind direction, SC = solar cycle).

Site	WS+T	WS+WD	WS+SC	WS+T+WD
Cuba	1.138327	1.397556	1.137904	1.268113
SIUE	1.624708	1.592776	1.562777	1.675193
Wilmington	1.72772	1.899298	1.743662	1.910552

As it can be seen in Tables 5.10 to 5.13, as more variables are added to the terminal set, there is a decrease in the quality of the models when applied to the test data. All these experiments were executed using a population of 2000 individuals, which, according to the results in Table 5.10 and Table 5.11 may be insufficient as the number of variables involved in the model increase. The

Table 5.13: RMSE in m/s on new data at 58m (WS = wind speed, T= temperature, WD = wind direction, SC = solar cycle).

Site	WS+T+SC	WS+WD+SC	WS+T+D+SC
Cuba	1.1309058	1.283128	2.315644
SIUE	1.6020632	1.637047	2.5275
Wilmington	1.7543704	1.919438	3.24581

hypothesis about adding these atmospheric variables was that, with more information known, better models could be constructed. The results show otherwise. In addition to the insufficient population size, it is important to have in mind that the training set contains only 50 data points, which seem a small number to be able to represent all the combinations of temperature, wind direction and solar cycle that could be presented to the model. To find out the appropriate number of training points, an exhaustive experimentation would be needed for each combination of variables. Another important aspect to take into account is the execution time. Population and training sets could be increased in size, however these would have a high execution cost.

Table 5.14 presents the mathematical representation of some of the models obtained at the three sites. The models present some repetition of functions as there were no constraints in how the program could grow. These models were used to forecast a complete 48-hour horizon (data starting immediately after the last training point). If the error of the numerical model was more or less consistent within the 48-hour horizon, then it would be interesting to observe how well could the same model be applied on the rest of the hours of the 48-hour horizon. Results in Figures 5.5, 5.6 and 5.7 show that the forecast is able to capture the trend of the observations, having an average MAE on the complete horizon of 0.783031 at Cuba, 0.995733 at SIUE and 0.809297 at Wilmington.

5.4   Analysing the Stop Condition

In all experiments, the algorithm has evolved for 500 generations. However the parameter was set arbitrarily. The approach has shown to work successfully, but it

Table 5.14: Models found with GP+Log+Exp.

Site	Mathematical Representation of the Model
Cuba	$v_3/v_1 * (exp(exp(exp(v_3 - v_1))/17.8952)+$ $exp(exp(v_2/17.8952)))$
SIUE	$exp(log(log(v_2 - 39.0888 - exp(exp(log(v_3) - v_1/39.0888))))$ $+log(v_2/v_4))$
Wilmington	$v_4/log(v_2 * v_3)+$ $log(exp(v_4)/(v_3 * v_2)/v_1 - v_1 + v_3 * v_2)$
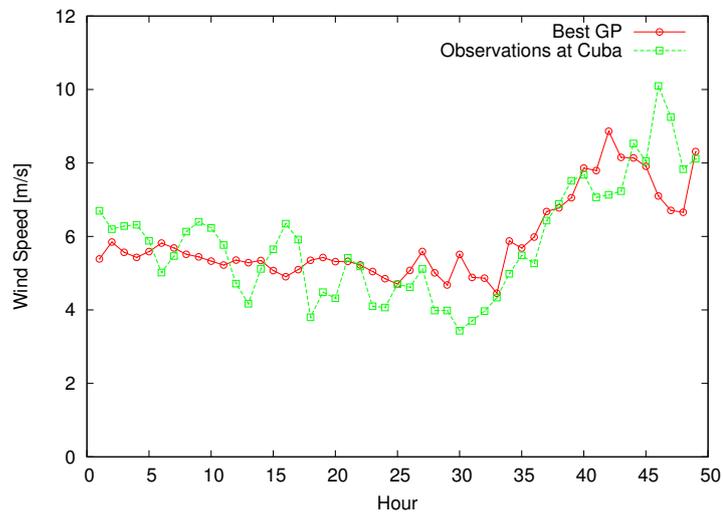
Figure 5.5: Forecast of a complete 48-hour horizon with the best GP found at the three different observation sites.
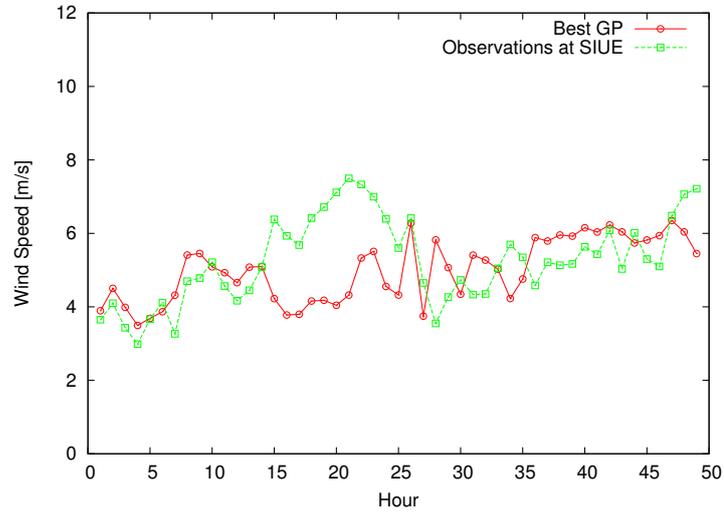
Figure 5.6: Forecast of a complete 48-hour horizon with the best GP found at the three different observation sites.
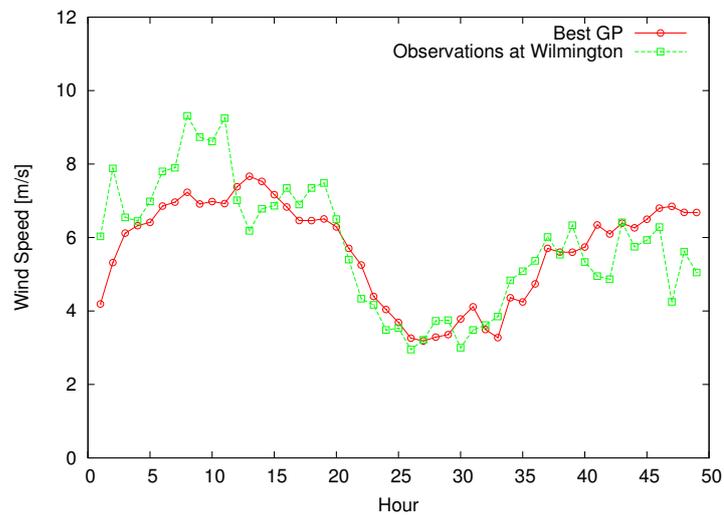
Figure 5.7: Forecast of a complete 48-hour horizon with the best GP found at the three different observation sites.

is important to make sure that the stop condition is allowing models enough time to evolve. In order to determine if this number of generations was large enough, the behavior of the population and the best solution was studied. Several runs of the GP were plotted to compare the population average cost versus the best program through time. Figure 5.8 shows the relationship between the evolution of the best solution and the population. As the best solution improves over the training and test sets, the average population cost has a decreasing tendency. After a certain number of generations, usually between 100 and 200, the best solution, which continues to improve over training set, decreases its quality over the test set, showing also an abrupt change in the tendency of the population average cost.
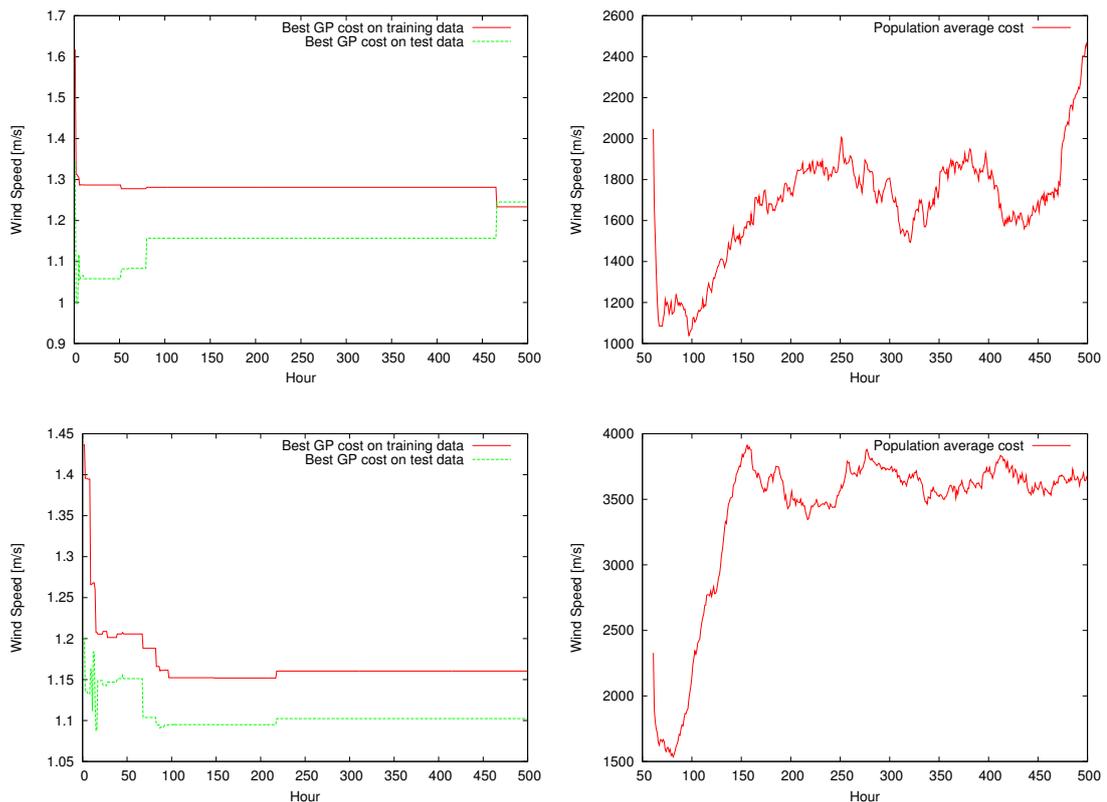
Figure 5.8: Best GP cost and population average cost on three independent runs at Cuba. Figures on the left hand side show the best GP on the training (solid line) and test (dotted line) sets on a typical run. Figures on the right show the moving average of the population on the corresponding run to the left.
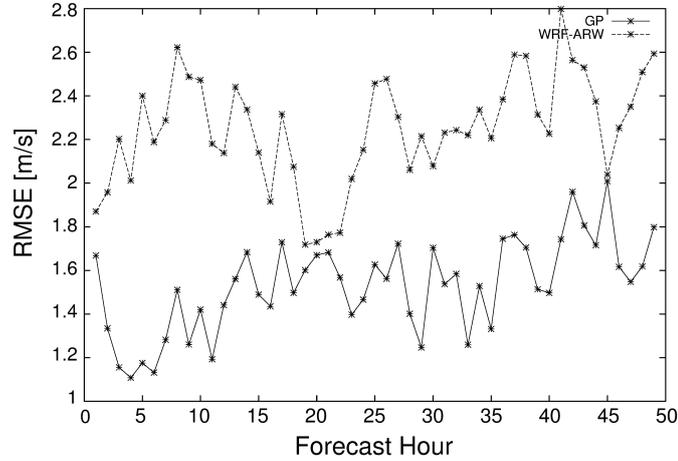
Figure 5.9: RMSE at each forecasting hour of the 48-hour horizon at Cuba.

It can also be observed in Figure 5.8 that after 100 generations, the models are not improving on the test set, and as the programs continue evolving, there is a higher probability of overfitting them. This suggests that a stopping condition of 100 generations might be just enough to let good generalisation models evolve and at the same time reduce the overfitting cases.

5.5 Re-training the Algorithm

With good generalisation models obtained, an important question that is raised is for how long can they be used, specially if they are obtained from a small training set. It was not expected that the model could be used for any time of the year. However, if the model was constructed in a type of sliding window process, where a new model is obtained as new data becomes available, the applicability of the model for the following day could been ensured. The main idea behind the sliding window approach is to re-train the model as soon as new information becomes available. This way, the models can continuously be capturing the changes in the trend of the climate and be applied in a different time.

To evaluate the quality of the downscaling models in a sliding window approach, the following experiment was designed. A training set was build using a window of 50 consecutive forecasts (corresponding to 13 consecutive days), start-
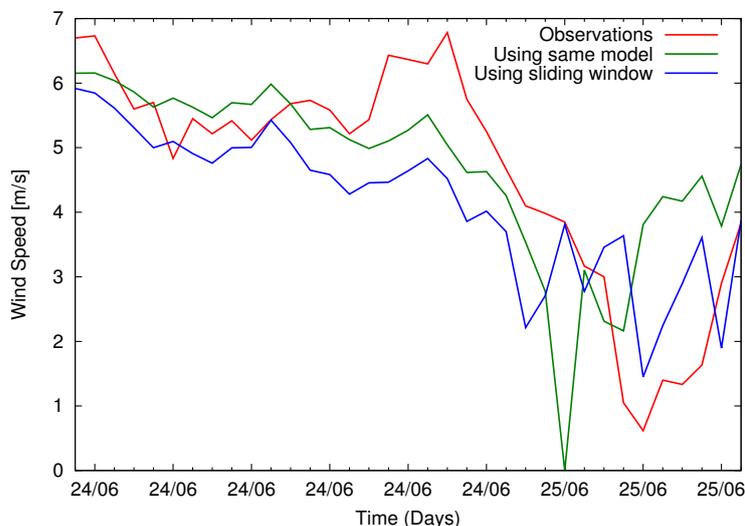
Figure 5.10: Using a static and retrained models before and during a wind speed ramp down event on the 24 and 25th of June 2011 at Cuba.

ing on May 2nd with the run at 12Z until May 14th last run, at 18Z. After running the GP on the training set, the best model found was used to forecast a complete 48 hour horizon, starting just after the last data point used in the training set (00Z May 15th). After testing, a new training set was created by sliding the window one day ahead (12Z run on May 3rd) and then tested on the following 48-hour horizon, starting after the last data point in the training set. This procedure was repeated until a period of 50 days was covered, producing a total of 50 complete 48-h horizon forecasts. The RMSE was calculated at each hour of the horizon and averaged over the 50 horizon forecasts. Figure 5.9 presents the average error per hour that was obtained for Cuba.

Figures 5.10, 5.11 and 5.12 show forecasts obtained during different wind speed ramp events at Cuba using the same model for every event (green) and daily retrained models (blue). It can be observed from Figure 5.10 that when the wind is more of less stable, both models give similar outputs. During the drop, however, the retrained model is slightly more accurate. The same situation is depicted in Figure 5.11. Figure 5.12 shows an increase in wind speed that, although both forecasts in general terms can predict, the sudden ramp ups are not well captured expect for the last day of June.
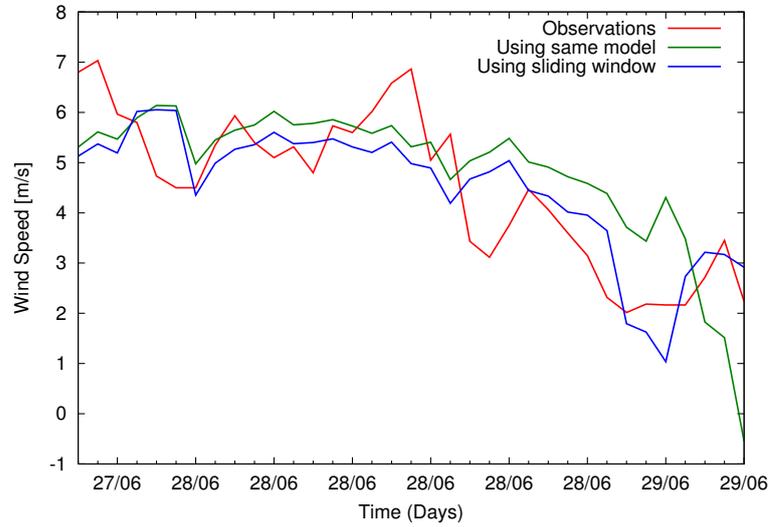
Figure 5.11: Using a static and retrained models before and during a wind speed ramp down event on the 27 and 27th of June 2011 at Cuba.
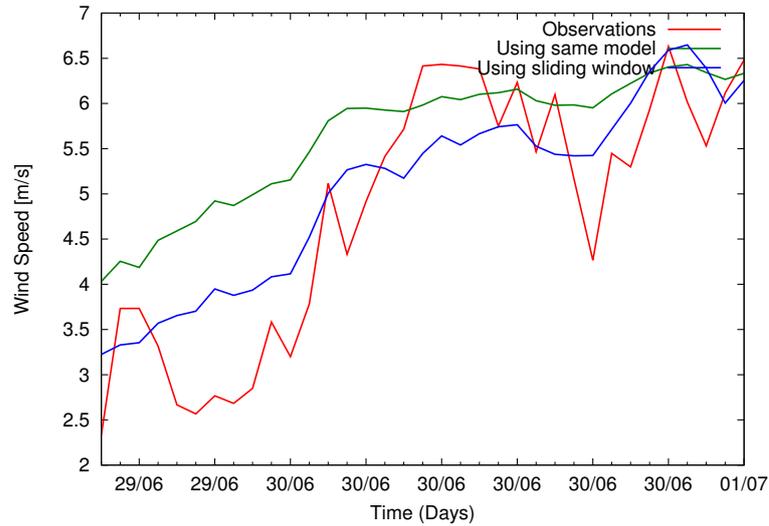
Figure 5.12: Using a static and retrained models before and during a wind speed ramp up event on the 29 and 30th of June 2011 at Cuba.

In terms of the error, using the same model in all events has a MAE of 1.05 and RMSE of 1.38 while using a retrained model each time obtains an overall MAE of 0.9364 and a RMSE of 1.20.

As it has been found in previous experiments, the results showed that the downscaling technique based on GP will have a different level of improvement depending on the observation site. Cuba and Wilmington sites showed better improvements than those obtained for SIUE. Nevertheless, in general, the technique was able to reduce the error of the WRF-ARW model, avoiding the execution of very high resolution forecasts, which decreases the time and resources cost.

5.6   Varying the Size of the Training Set

A final set of experiments was designed to evaluate the quality of the results using GP in another time of the year and varying the size of the training set. Three training sets were used: one with 2 months of data (May and June 2011), a second one with one month (June) and a third one with 15 days (last two weeks of June) of data. In the three cases, only the daily 06T run was used, taking the values forecasted from hour 19 to hour 42 (which correspond to the next day forecast). All next day forecasted hours were considered this time, as an attempt to increase the number of training points. The testing set consisted of the first 10 days of July. As in the training set, each day of the test set corresponds to a forecast done the previous day at 06Z. According to the RMSE obtained for the three sites, a training set of 15 days was enough to forecast wind speed at Cuba. Forecasts at Wilmington were best achieved with one month of data and the best results at SIUE were achieved with the two-month training set. This indicates that each place has a different level of complexity to forecast wind speed and, therefore, a different amount of training data is needed. Larger training sets were not used due to limited data availability.

Figure 5.13 shows the first 100 hours of the test set at Cuba. Additionally to the control forecast (best solution), this figure shows the best 50 models obtained on the best run. GP evolves a population of solutions, therefore, not only the best one but a subset of the population can be used.

As it can be observed in the top image of Figure 5.13, most of the models
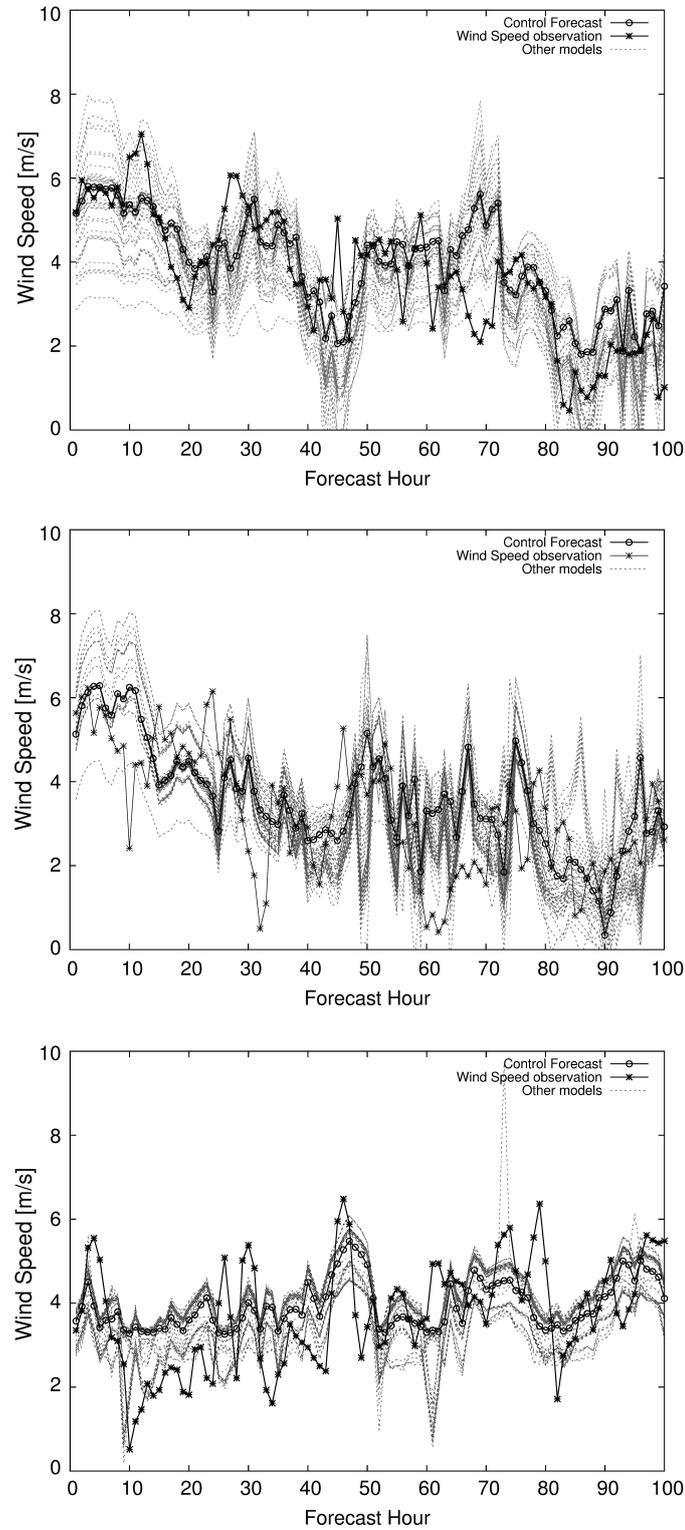
Figure 5.13: Wind Speed forecasts at the three observations sites with a different training set sizes: a 15-day training set for Cuba, a 1-month training set for Wilmington and a 2-month training set for SIUE.

agree with the control forecast. However, periods like hours 80 to 90 or 95 to 100, show how a visible amount of models can differ from the control forecast. This can be related to the use of different wind speed variables in the models, which is expected. The use of four wind speeds to build the downscaling model is based on the hypothesis that the closest point from the mesoscale grid is not necessarily the best point of reference for a downscaling model. Additionally, considering several grid points could help quantify the uncertainty of the wind in the observation site. The results in the three figures show that the same models found with one training set, could be used for a consecutive period of 10 days. However, this was only tested on the month of July, where atmospheric conditions might have been similar. If the models are to be tested on a different period of time, the re-training of the algorithm is preferable.

In terms of the speed of the algorithm, the average execution time of the GP approach was 1,348.394 seconds for a two months training set. This, added to the total execution time of the numerical model (approx. 3 hours), is still considerable lower than the high resolution runs which could take 48 hours for 5 domains (Louka et al., 2008).

5.7 Discussion

This chapter presents the successful application of GP as a downscaling method to forecast wind speed at specific observation sites. The implemented algorithm is able to construct programs that model the wind speed behavior at the observation sites from the wind predictions obtained by the mesoscale model WRF-ARW. The results obtained with GP improve the ones obtained with the mathematical methods explored before. Also, different function and terminal sets were used to investigate if these could improve the results of the basic GP implementation. It was found that adding logarithmic and exponential functions to the function turned to better results. However, including more variables to the terminal did not lead to better results. The reason for this could be that the training set is not large enough to be representative of all the wind speed and temperature combinations in the site, so new data presented would not necessarily be modelled by the program. Nevertheless, the results obtained by adding only logarithmic

and exponential functions are comparable to the ones published in (Salcedo-Sanz et al., 2009) and (Sweeney et al., 2011).

An important finding of the experimentation presented in this chapter was that the level of improvement using GP was different at each site and required a different training set size. The same NWP model can have a different level of accuracy, which was expected as each location has its own level of topographic complexity and the same finding applies for the downscaling technique. In general it can be concluded that the quality of the results are site dependent. However, the algorithm is not designed for a specific site. A different equation will evolve according to the complexity of the site, making it a general downscaling tool.

Results obtained at 58m height showed that more training data was needed in order to obtain good generalisation models. Finding the adequate size of the training set is an important aspect of the algorithm and if many inputs are to be considered, one has to ensure that the training set is able to be large enough to cover the search space. The more variables involved, the more complex the problem and larger the training set needed.

It was also found that for wind speeds at 10m height, models trained with a subset of the training set achieved better generalisation. This finding suggests the use of a validation set for further experimentation, dividing the original training set into two sets, training and validation, using the second one as a way to prevent overfitting. While models will evolve based on their fitness on the training set, the best individual will be the one that has the best fitness on both sets.

With a wind speed forecast at a specific location, a power curve could be applied to convert those forecasts to wind power forecasts of a given wind turbine/park. The following chapter introduces this idea and explores a second approach which consists in forecasting directly wind power from raw NWP wind speed predictions.

Chapter 6

Wind Power Forecasting with Genetic Programming

6.1 Introduction

In the previous chapter, GP was successfully applied as a downscaling step to produce a wind speed prediction at a specific location using low resolution NWP forecasts. With the execution of high resolution NWP runs being avoided, the second aspect to study was the misplacement errors of the model. Using a wider area from the mesoscale grid for wind speed/wind power prediction could provide information about rapid changes in wind speed that might not necessarily be modeled at the closest points. This chapter is focused on studying how wind speed forecasts can be converted into power forecasts and how a neighbourhood approach could be used to improve those forecasts. The results presented were published in (Martínez-Arellano and Nolle, 2013a,b).

The chapter is organized as follows: in Section 6.2, a look into some of the possible variables for modelling wind power with GP is provided. Then, section 6.3 introduces two main approaches to wind power forecasting. The first one is a one-step approach, which models the conversion of raw NWP forecasts into wind power directly. The second approach is a two-step one, which first applies GP to downscale wind speed and then applies a power curve to obtain the total wind power. The estimation of uncertainty is studied in section 6.4 using *Quantile*

Regression to estimate prediction intervals. Section 6.5 extends the application of GP to neighbour points in order to generate an ensemble of forecasts as a second approach to uncertainty estimation. Finally, section 6.6 discusses the results obtained.

6.2 Considering Wind Direction for Wind Power Modelling

One of the critical aspects of symbolic regression is to determine the variables that will be used for modelling. Wind is usually characterised not only by its speed but also its direction. Wind direction could potentially have an effect on the performance of wind turbines as these need to readjust when the direction changes. In a preliminary analysis of the wind speed, wind direction and power output at Sotavento, it has been identified that for the year 2012, from January to March, when the wind blows in the East Northeast direction (56.25 to 78.75 degrees) or in the West Southwest direction (236.25 to 258.75 degrees), the maximum wind power output was achieved for high wind speeds. On the other hand, when the wind blows in the North direction (348.75 to 11.25 degrees) the power output is lower even for the higher wind speeds. This can be seen in Figure 6.1. This suggests that the incorporation of the wind direction variable in the regression models could improve the accuracy of the predictions.

April had a predominant wind from the west southwest as the first three months of 2012. This tendency, however, does not apply to all the year. Figure 6.2 shows the wind roses from May and June, 2012. It can be observed that the direction that was predominant in previous months did not prevail during May. With a model dependent on the wind direction, it cannot be expected that it will apply for any other period, as wind patterns can be different, as well as the power output.

Given the characteristics of the wind during this period, it was decided to use data from January to April, 2012, as an initial experimental set, using April for the testing as the prevailing winds are similar. Before creating the training, validation and test sets, the January to March period was preprocessed to remove
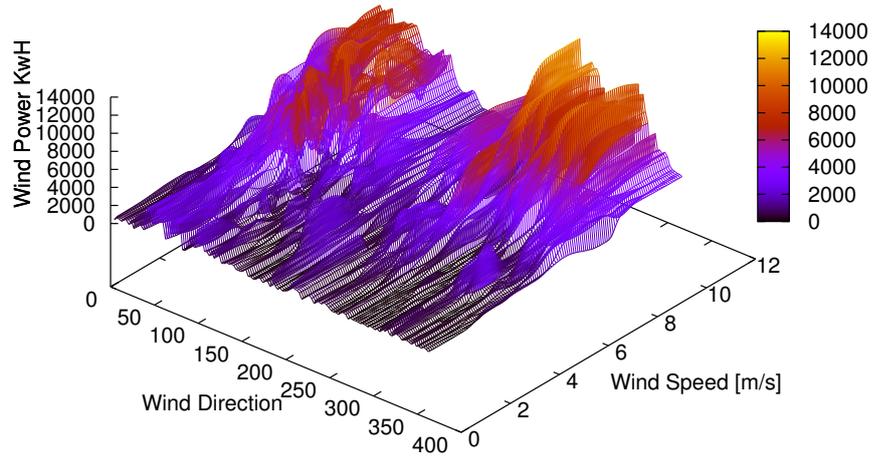
Figure 6.1: Relationship between wind speed, wind direction and the power output at Sotavento Wind Park for three months, January to March, 2012.

(a) May

(b) June
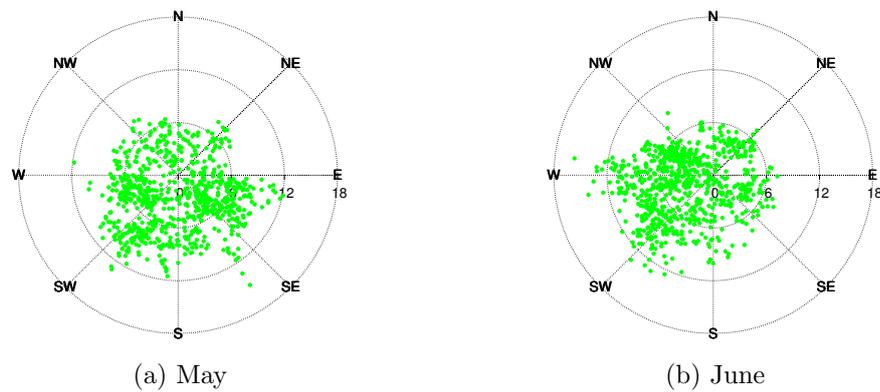
Figure 6.2: Wind speed and direction during May and June 2012 at Sotavento.

Table 6.1: Wind direction categories.

Cardinal Direction	Degree Direction	Cardinal Direction	Degree Direction
N	348.75 - 11.25	S	168.75 - 191.25
NNE	11.25 - 33.75	SSW	191.25 - 213.75
NE	33.75 - 56.25	SW	213.75 - 236.25
ENE	56.25 - 78.75	WSW	236.25 - 258.75
E	78.75 - 101.25	W	258.75 - 281.25
ESE	101.25 - 123.75	WNW	281.25 - 303.75
SE	123.75 - 146.25	NW	303.75 - 326.25
SSE	146.25 - 168.75	NNW	326.25 - 348.75

the outliers. Two types of outlier values that could affect the training process were identified. The first one relates to 'errors' on the power output measurements. As the power output observations used in this study are total production values of the wind farm, an outlier could represent the situation where some turbines have been turned off. The second type of outlier is related to unusually large errors in the wind speed predictions. These can be easily identified by plotting wind speed predictions versus power observations.

To remove these values, the relationship between wind speed predictions (from WRF-ARW) and power observations was approximated by a curve using Matlab and the points that did not fall at a 90% range from the curve were removed. As different direction patterns may have a different effect in the power output, it was decided to divide the training period by the different wind direction categories shown in Table 6.1 and, for each category, to remove the wind speed measurements that were considered outliers.

Figure 6.3 shows the training set before and after removing the outlier points. The total number of input records decreased almost by half from 1632 records to 975. From the remaining points, 80% was used for training and 20% for validation.
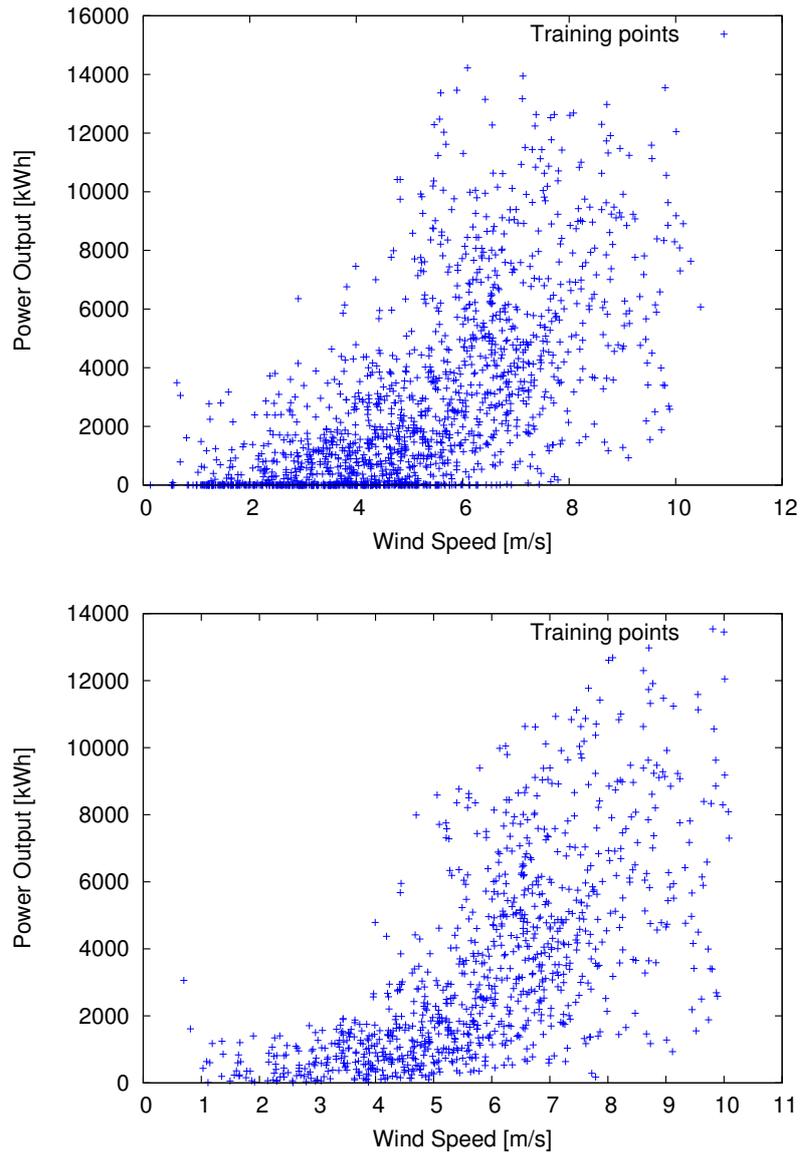
Figure 6.3: Training set points before (top) and after (bottom) removing outliers.

6.3 One-Step and Two-Step Approaches for Wind Power Forecasting

As it has been previously mentioned, NWP models need to be taken into account when forecasting wind power for horizons larger than 6 hours. As the main interest of this research is to study ramp events for the day-ahead market, an important aspect to investigate is how to incorporate these NWP predictions into the wind power forecasting process. There are two basic ways to convert the numerical model predictions to wind power forecasts. The first one consists in taking the raw NWP output together with observations of the local conditions to estimate directly the output of a wind turbine/farm in one step. The second approach consists in forecasting first the wind speed at a target location, either at the wind turbine location or somewhere in the wind farm, and then the wind speed forecast would be converted to a wind power forecast using a wind turbine/farm power curve. These two approaches are presented in more detail in the following sections.

6.3.1 The One-Step Approach

As the name states, this approach consists in finding a model that can convert in one step raw NWP forecasts into power predictions. The "downscaling" step is performed at the same time the wind speed prediction is converted to a wind power one. The initial benefit of doing the whole process in one step is that the algorithm will be executed once. This, however, involves the study, in detail, of the number and type of variables that will be used as input. As this is performed in only one step, it could imply the use of a large number of input variables, which, from experience, might sometimes not be the best approach.

The starting point to get a basic one-step approach was to simply use wind speed in the modelling process. Once this approach was implemented, it was further extended by adding wind direction. The inclusion of this variable could potentially improve the accuracy of the one-step approach and be fairly compared to the two-step approach which naturally incorporates the wind direction into the process. In order to assess the inclusion of wind direction, an experiment

was performed. This experiment was based only in the predictions at the closest point from the grid(denoted as v_1 and d_1). The reason for this is that if the four surrounding points are taken, as in the downscaling procedure, the number of inputs would grow to eight when direction comes into play. The GP was set to run with the basic operators and logarithmic and exponential functions, as this configuration has worked well in previous experiments. The experiment was executed with different pressure parameters and, as shown in Figure 6.4, a parsimony pressure of $k = 2.0$ was enough to obtain a similar performance on the validation and test sets.
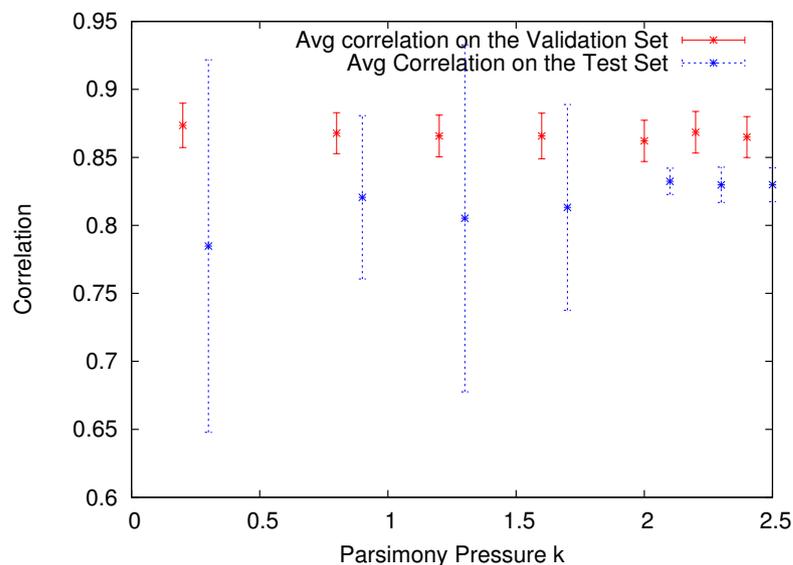
Figure 6.4: Average correlation of the best model to the validation and test sets and standard deviation in 50 runs applying different pressure parameter values.

In terms of model complexity, the lowest RMSE on the test set was achieved with a complexity between 50 and 100 (see Figure 6.5). This means that models do not need to grow very large to achieve a reasonable accuracy.

The results show that models using the wind direction variable provide an improvement over models that use only the wind speed. Table 6.2 shows the MAE and RMSE on the test set for both types of models. It can be observed that there is a larger improvement on the RMSE and that correlation parameters are also improved. The MAE using both variables represents the 11.7% of the
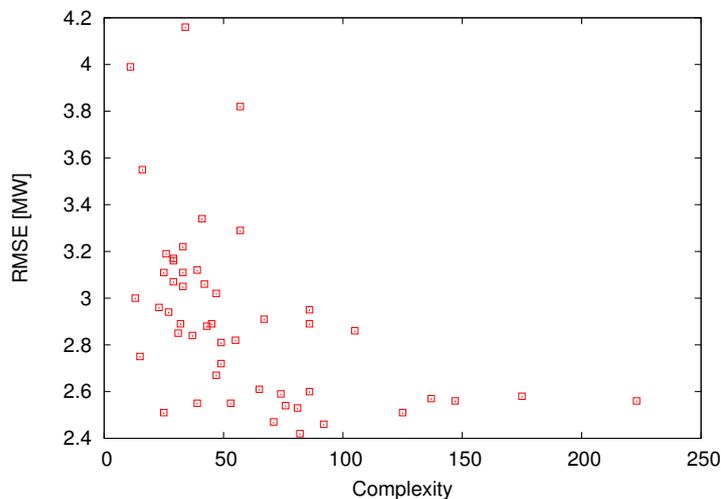
Figure 6.5: Relationship between complexity and RMSE in 50 runs using the pressure parameter $k=2.0$.

total capacity of the park. This percentage is very good compared to state-of-the-art tools that have been applied at Sotavento (Giebel et al., 2011).

Table 6.2: Average MAE and RMSE in kWh obtained in 50 runs on the test set using only wind speed and using wind speed and direction.

Training Set	MAE	RMSE	Correlation Coefficient	Determination Coefficient
Only wind speed	2066.3842	2810.64	0.7930	0.5869
Wind speed and direction	2055.4866	2679.6936	0.8521	0.6871

Another way to see this difference is by looking at the power curves obtained with the two models which are shown in Figures 6.6 and 6.7. These show that the use of wind direction can reproduce the power output variability of the real world within the model whilst the wind speed alone cannot.

With a wind speed-to-power model developed, wind power forecasts can be obtained as soon as new NWP predictions are available.
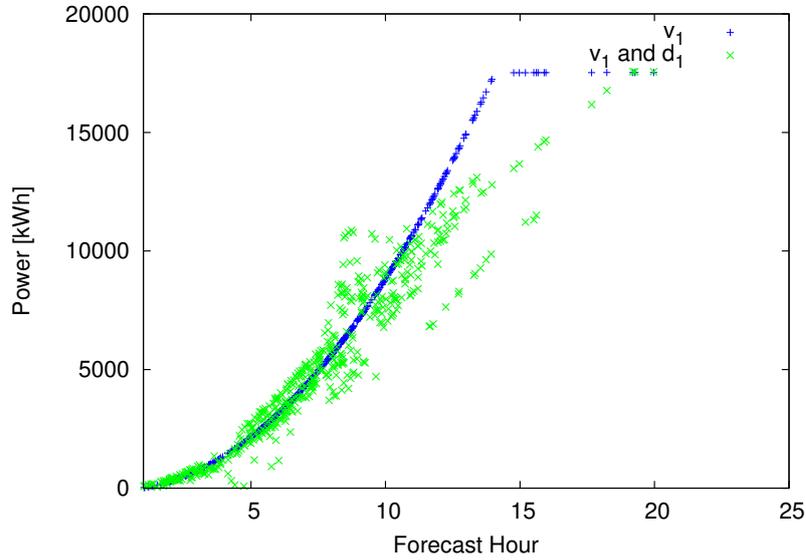
Figure 6.6: Power plots obtained on April 2012 using only wind speed (blue) and using both wind speed and wind direction (green).
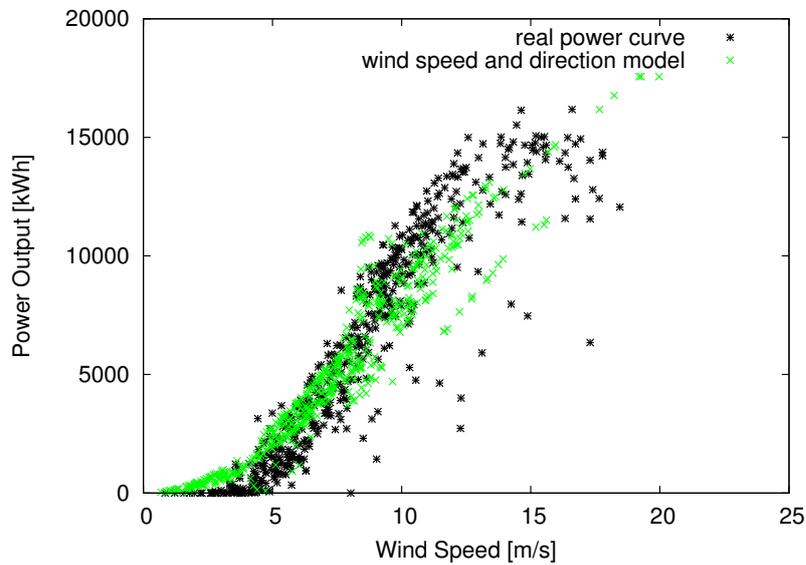
Figure 6.7: Observed and predicted power plots on April 2012.

6.3.2 The Two-Step Approach

An alternative to the direct conversion of raw numerical forecasts to wind power prediction, is to do the process in two separate steps. The first step will consist

in downscaling wind speed to a location of the wind park. This process will be done as introduced in the previous chapter, taking the four surrounding points to the park. Once a new wind speed forecast is obtained, it is possible to estimate the power output of the park at the time of the prediction by using a wind park power curve. Wind park power curves are an alternative to the use of wind turbine power curves, which would require the estimation of the power output of each turbine and then sum them all together. A park level wind-to-power model in this case is convenient as there is no available information of the power production of each turbine. To obtain the power curve model of the farm, the GP is trained with wind speed, wind direction and power observations from the park.

Figure 6.8 shows the Sotavento power model, found by GP, applied to wind speed observations at the farm on the test period (April 2012).

The plot at the bottom of the figure shows a well modeled power output. The top plot shows some discrepancies not in shape but in intensity, specially from the 14 to the 16 of April. An interesting thing to point out is that, even though the model is applied to wind speed observations, one can observe a small time misplacement during the 14th of April. The data that is provided from Sotavento does not specify any changes in the configuration of the turbines or the cases of turbines turned off or on. However, these discrepancies suggest two possible things. One is that the wind turbines might have taken longer than expected to adjust to the wind variability. In this case, further study into the modelling of the aerodynamics would be necessary. The other possibility is that the due to certain grid conditions, some of the wind turbines were powered off. Despite these discrepancies, the figure shows an overall agreement with the power observations. Some model examples are shown in Table 6.3

Once a wind park power curve was obtained, downscaled wind speed forecasts could be fed into the model to obtain wind power forecasts. These downscaled wind speed were obtained as it was explained in the previous chapter.
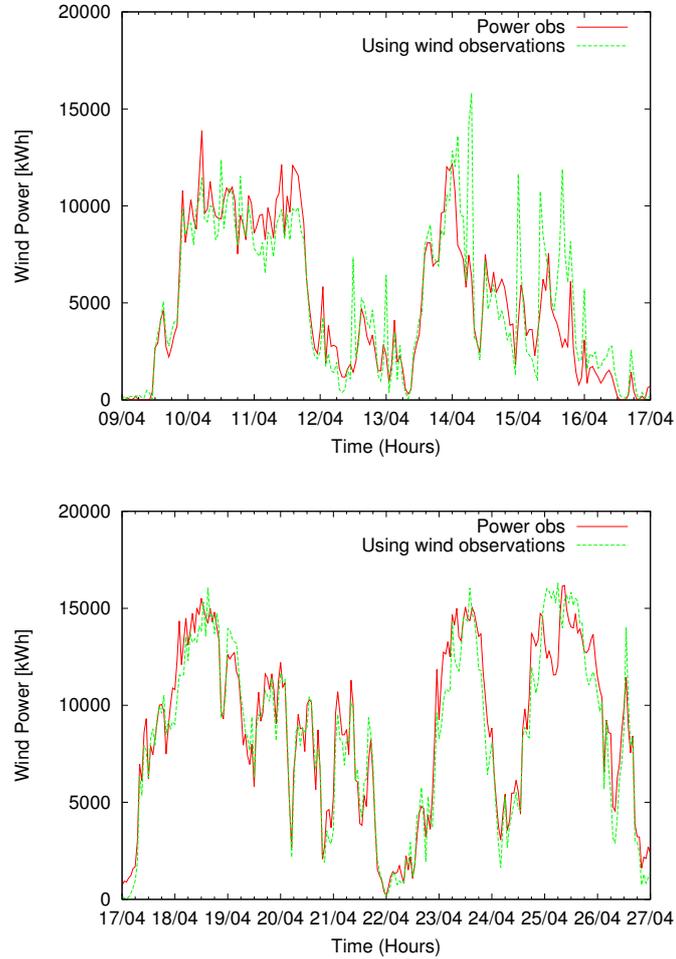
Figure 6.8: Sotavento power model, found by GP, applied to wind speed observations on April 2012.

Table 6.3: Examples of Sotavento power model found using GP and wind speed/direction observations at the farm (v_0, d_0).

Size	Mathematical Representation of the Model
25	$90.092v_0^2 + 2v_0 - 2d_0 - 170.183$
40	$v_0^2 + +2ln(v_0) + v_0 - d_0+$
	$(166.7914v_0ln(v_0) - ln(v_0) - d_0 - v_0)ln(v_0)$

6.3.3 Comparison

The power models obtained with both approaches were used to obtain the wind power forecasts for the testing period. This is shown in Figure 6.9 .
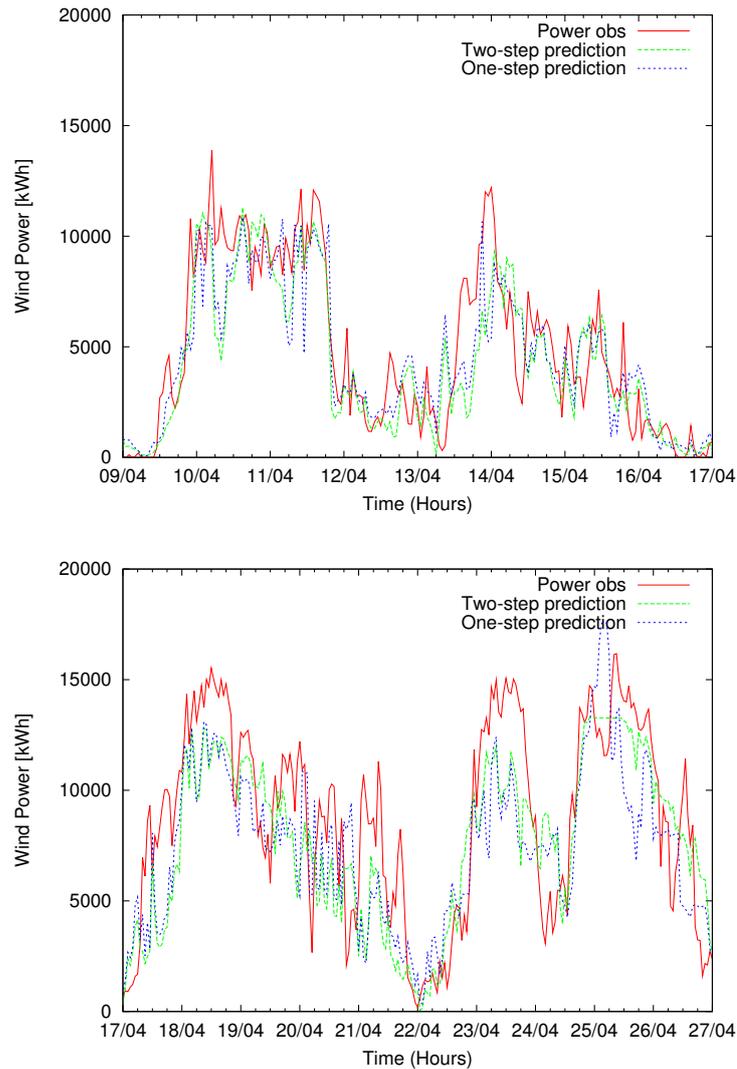
Figure 6.9: Sotavento power model, found by GP, applied to downscaled wind speed predictions on April 2012.

The results suggest that there is a similar performance, although some small differences can be observed. The computational cost of the two-step approach is higher than the one-step approach as the GP needs to run twice. In both cases

the algorithm would need to be "retrained" as the weather patterns change with the seasons. So, the two-step approach might not need to run twice if applied in the near future, but will necessarily need to obtain a new wind park curve if applied to a different time of the year. Despite the need of re-training, the training time of the GP is in average 2,856 seconds, which may not be part of the critical latency of the numerical prediction except when the system is deployed in one specific way. Even in those cases, the total execution time of the numerical model (approx. 3 hours) is still low compared to high resolution runs which for up to 5 domains could run for 48 hours. Regardless of the approach, the accuracy obtained for the Sotavento location is around 88% in terms of the MAE. This is comparable to current forecasting tools, as established in (Giebel et al., 2011). This, however, would require further experimentation as the summer and autumn seasons were not considered.

With a similar performance achieved by both strategies, it was decided to use the one-step approach to forecast the months of May and June. For May experiments, the training period was from February to April. For the month of June, the training period was from March to May. In both cases, the preprocessing of the data was performed to remove outliers. Table 6.4 shows the MAE and RMSE obtained for May and June.

Table 6.4: MAE and RMSE in kWh obtained for different testing periods.

Month	MAE	RMSE	Correlation Coefficient	Determination Coefficient
April	1934.69	2580.39	0.844564	0.697236
May	1459.78	2015.27	0.699818	0.338019
June	1450.3	2027.96	0.841638	0.696648

The table shows that the errors obtained for the new test periods were slightly lower than for April. However, results for May show a lower correlation. It is important to note that the wind pattern in this month showed more variability (as shown in Figure 6.2), with frequent winds from the east. As the January to April period has predominant winds from the west, it might be that for this period in particular, the training data was not containing a significant amount of

cases with winds from the east that may affect the farm differently. The month of June, however, shows similar results as April in terms of correlation.

6.4 Uncertainty Estimation of Wind Power Forecasts

With point forecasts like the ones obtained previously, the grid operator can take decisions for the daily unit commitment. However, these would have a very high uncertainty if no other information is provided. It has been shown that the point forecast is more useful if it is supplied with an uncertainty estimation (Pinson, 2006). This estimation can give the operator a better idea about the amount of backup needed.

The total error of the wind power forecast can be the product of two sources. The first source of error is the error introduced by the numerical model. This error tends to increase as the forecasting horizon increases. The second source of error is the conversion process from wind speed to wind power. This process can depend on local conditions such as the roughness and orography of the wind farm location (Kariniotakis et al., 2004). This type of error could be potentially decreased by the use of empirical power curves rather than the power curve provided by the manufacturer.

There are several approaches in the literature to quantify the uncertainty of wind power point forecast errors. In general, the error distribution is found by observing the behavior of the error on past forecasts and by using explanatory variables which are additional information such as wind speed, wind direction, temperature, that can improve the understanding and thus the modelling of the error.

In order to anaylse the behaviour of the forecast error obtained with the GP model, histograms were used. Figure 6.10 shows the empirical distribution of the error in the January-March training set for the first and last hour of the next-day horizon. Each hour of the horizon has been treated separately as a first attempt to study the use of the error estimation on the ramp characterisation. One can observe that the distribution of the error is different for the horizon $t + 19$ and
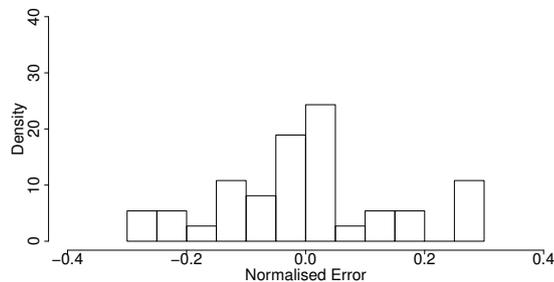
(a) Error at t+19
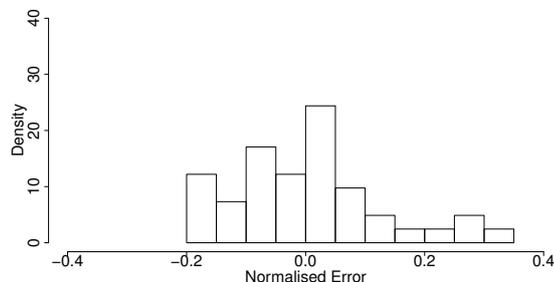
(b) Error at t+42

Figure 6.10: Empirical distributions of the prediction error obtained with v_1 for two look-ahead days. Prediction errors are normalised on a scale [-1,1].

$t+42$. The shorter horizon shows a more distributed error while the larger horizon shows more overestimation errors, as the frequency of negative errors is higher. It can also be observed that the upper bound of the empirical distribution at $t + 42$ is slightly higher. This means that at this horizon errors of 35% were observed, while at t+19 errors stayed lower than 30% of the nominal capacity of the farm.

Studies suggest that wind power forecast errors do not follow a normal distribution as wind speed forecast errors do. In fact, wind power error distributions have been found to have high kurtosis and skewness (Pinson, 2006). For this reason, an approach that makes no assumption of the distribution when estimating the uncertainty could be more appropriate. *Quantile Regression Forests* (Meinshausen, 2006) is a non-parametric technique to estimate conditional quantiles for high dimensional predictor variables of a response variable. In typical linear regression, the equations found are designed to estimate the mean of the response variable conditional to a predictor one. Quantile regression finds these equations

127

by examining how the relation of the predictor and response changes depending on the score of the response variable.

Quantile Regression Forests perform quantile regression based on the Random Forest algorithm (Breiman, 2001). Random forest is an ensemble approach, also thought as a form of nearest neighbor predictor, where the response of the ensemble members is averaged to obtain the value of the prediction variable when an input is given. Each ensemble member is a decision tree which is constructed based on a subset of the predictor variables (X which can be high-dimensional) to predict a real-valued response variable (Y). For each tree and node, random forests employs randomness when selecting a variable to split on. The split is done according to a training set which is created by sampling with replacement from a data set. Each leaf of the tree should, at the end of the construction phase, correspond to a specific subspace of the predictors search space. For regression, the prediction of a single tree is the weighted average of the original observations that fall in the same leaf. The weight is given by:

$$w_i(x, \theta) = \frac{1}{\#\{j : X_j \in R_{j(x,\theta)}\}} \tag{6.1}$$

where $R_{j(x,\theta)}$ is the node were X falls so the weight is 1 divided by the number of observed values that fell in the same leaf as X. The prediction of a single tree given $X = x$ is calculated as follows:

$$\hat{\mu}(x) = \sum_{i=1}^{n} w_i(x, \theta) Y_i \tag{6.2}$$

The prediction of a single tree is then the weighted average of the original observations $Y_i, i = 1, ..., n$. For the nodes were X do not fall, the weigh is zero. The prediction of a random forest is the weighted average of all trees. Compared to typical regression, Random forests approximates the conditional mean by a weighted mean over the observations of the response variable.

Quantile regression forests adds an additional step to find the conditional distributions of the variable to predict. While traditional random forests keeps only the mean of the observations that fall into a node, the quantile regression forests keeps the value of all observations that fall in the node, not just the mean,

(a) 90% percentil on out-of-the-bag observations at $t + 19$

(b) 90% percentil on out-of-the-bag observations at $t + 42$
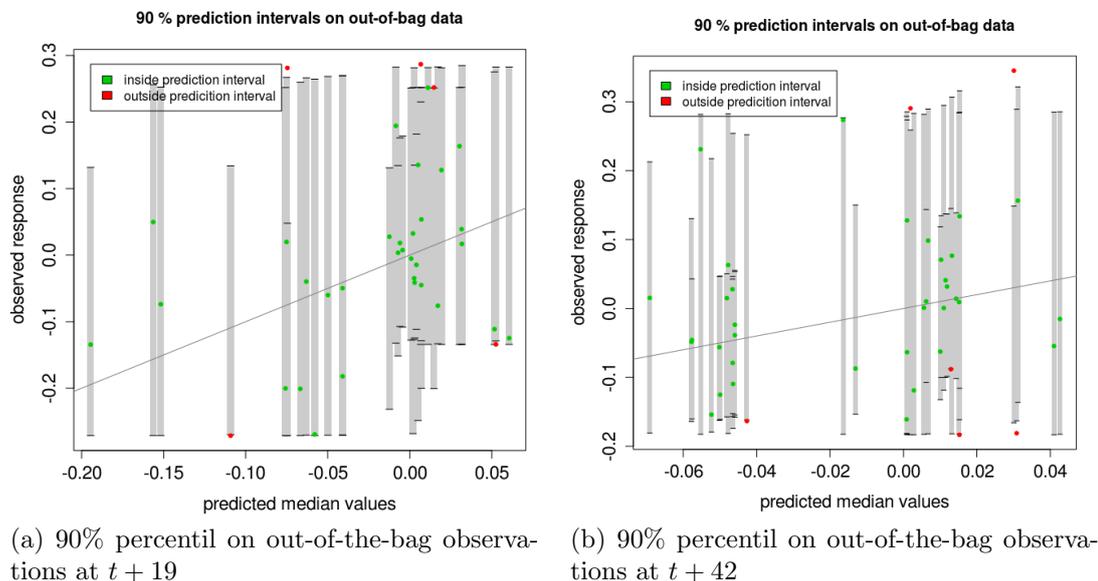
Figure 6.11: Intervals for out-of-bag observations. The green points correspond to observations that fell inside the intervals, while red points are those that fell outside. Prediction errors are normalised on a scale [-1,1].

and assesses the conditional distribution based on this information.

To investigate the training errors of the GP model, error quantiles were obtained by applying the quantile regression forest method for each hour of the next-day horizon. The distribution of the error was found using wind power, wind speed and wind direction forecasts as explanatory variables. Figure 6.11 shows the out-of-bag observations (observations not used for finding the intervals) and how they fall into the 90% interval. It can be observed that for $t + 19$ and $t + 42$ (the beginning and end of the next day horizon) the intervals are covering most of the points.

To see how these intervals found perform on unseen observations, these were applied to the April test set. Figure 6.12 presents the point forecasts for the first four days of the April test set as well as the prediction intervals (10%, 20%, ..., 90%) obtained. The power predictions are presented as a percentage P_n of the nominal power of the farm (maximum capacity).

The distributions found were applied to the test set which corresponds to one month after the last training point. It is important to consider that the
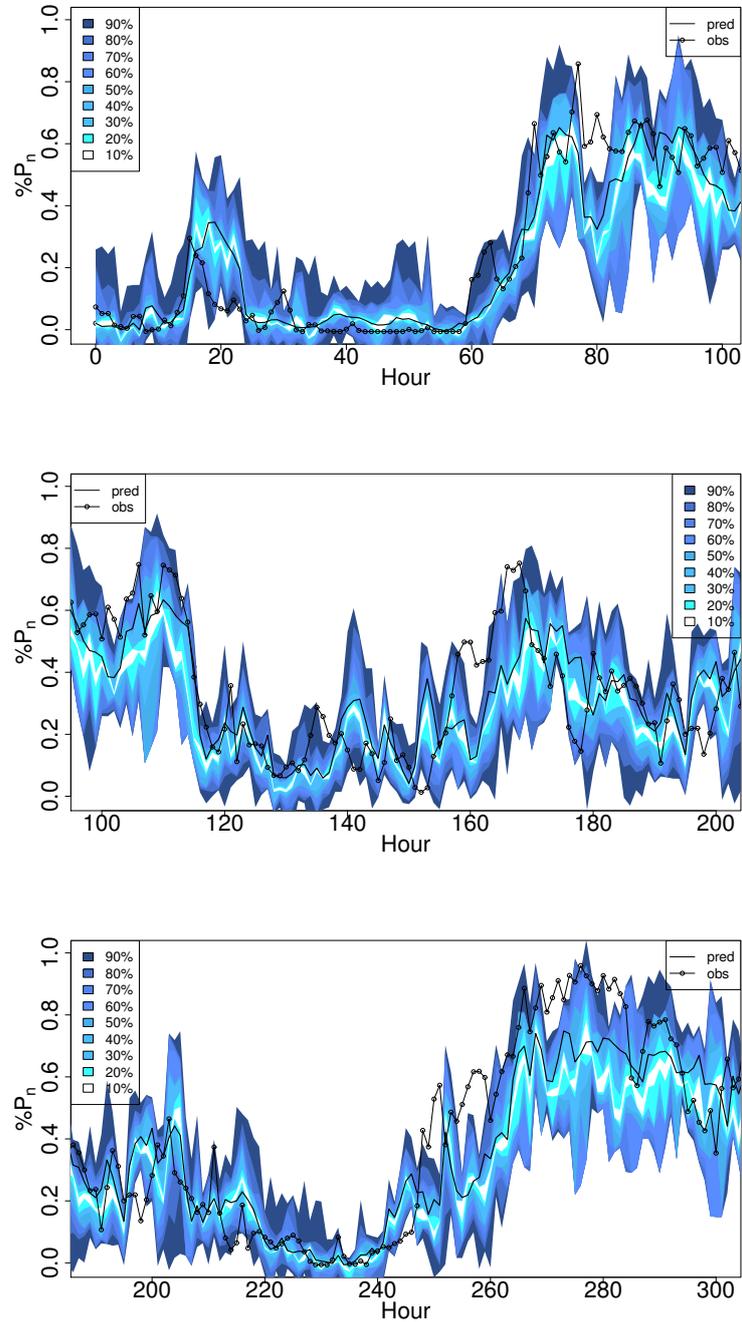
Figure 6.12: Wind power point predictions as percentage of the nominal power of the farm (P_n) and the associated interval forecasts using the closest point from the grid (v_1).
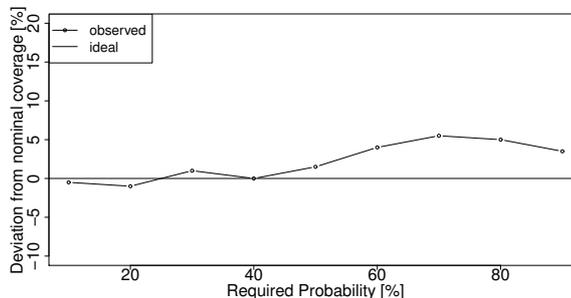
(a) Reliability on the first 200 hours of the test set
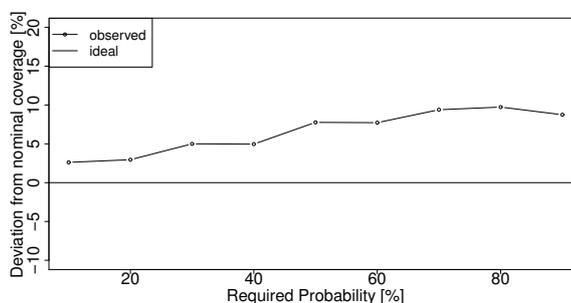
(b) Reliability over all the test set

Figure 6.13: Reliability diagrams of 5, ..., 95 percentiles estimations made with the *Quantile Regression Forest* procedure.

distributions found are considering a small set of history observations, which could probably indicate that these are only accurate for a short period after the last training point. In order to assess the influence of the sample size on the quality of the estimated intervals, reliability diagrams were used. These reliability diagrams provide information about the deviation of the actual coverage ($\hat{a}^{\alpha}$) of the forecasted intervals from the nominal proportions (α). This deviation is defined as

$$\alpha - \hat{a}^{(\alpha)} = \alpha - \frac{z^{(\alpha)}}{N} \tag{6.3}$$

where N is the total number of observations and $z^{(\alpha)}$ is the number of observations that fell in the interval with proportion α (Pinson, 2006).

Figure 6.13 depicts the reliability evaluation results for the predictive distributions obtained in the following 200 hours after the last training point and for

the complete test set. The diagrams are for the complete next-day horizon (hours 19 to 42). This means that all probabilistic forecasts for all look-ahead times were used with equal weight when calculating the deviations.

There are a couple of things to note from this figure. The narrower intervals (10% and 20%) tend to be overestimated on the first 200 hours, having a slightly higher coverage, while the larger intervals tend to be underestimated. In the same figure, diagram b shows how the reliability decreases as the horizon increases. The results confirm the behavior that was expected. The intervals are valid only for a short period of time after the training period. In order to maintain the reliability, the GP model would need to be retrained as new information becomes available and the error distribution needs to be recalculated. Similar results for the months of May and June can be found in Appendix E.
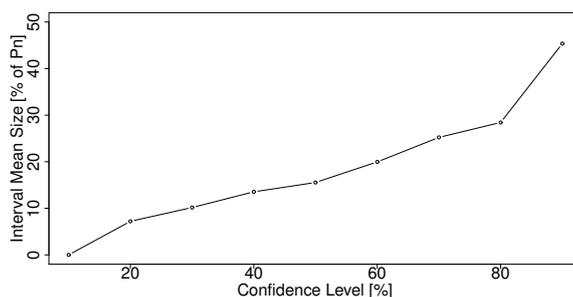
(a) Average size of prediction intervals at T+19
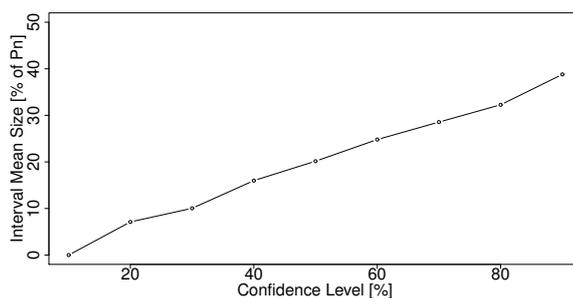
(b) Average size of prediction intervals at T+48

Figure 6.14: Average size of prediction intervals at T+19 and T+48 horizons.

Other aspects of the predicted intervals such as the sharpness and resolution were analysed. The sharpness, Figure 6.14, was calculated as the average of the
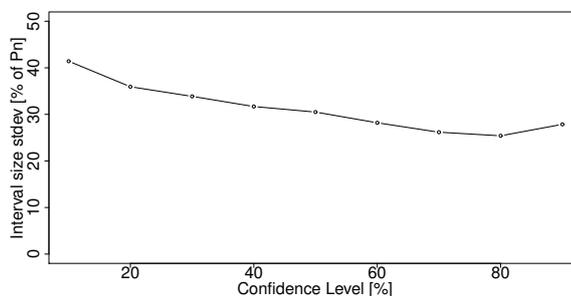
(a) Average size of prediction intervals at T+19
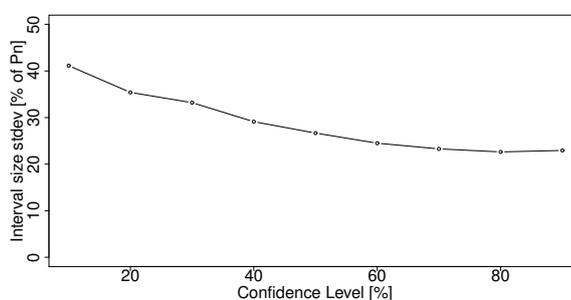
(b) Average size of prediction intervals at T+48

Figure 6.15: Average size of prediction intervals at T+19 and T+48 horizons.

interval size for a particular horizon. As it is shown, the average size of the intervals increases with the confidence intervals as it was expected.

The resolution metric shows how the intervals adjust in different situations. This was obtained by calculating the standard deviation of the size of the intervals. Figure 6.15 shows that the intervals vary between 30 to 40% at T+19 and between 20 to 40% at T+48.

With an accurate estimation of the intervals, the grid operator can determine what is the risk of sudden increases or decreases in power output that were not modeled by the point forecast. Although this information is already useful for decision making, it hasn't been used for the characterisation of ramp events. If the distribution of the error is known, this could improve the detection of possible ramp events that were not modeled by the point forecast but that can potentially develop into one. This idea will be developed in detail in the following chapter.

6.5 Using Neighbour Points to Study Misplacement Errors

With a wind-to-power modelling approach working and achieving a comparable accuracy to current forecasting tools, it was possible to move forward into the neighbourhood approach for ensemble prediction. The GP approach would allow the modelling process to be applied at different grid points, as if they were the actual location of the park. To investigate into this idea, two strategies were used. The first one consists in using a set of surrounding points to generate one model. It is assumed that the algorithm will be able to find the combination of points which better describes the power output of the wind farm. It has to be taken into account that with this approach, it is assumed that the best combination of variables found would be the best for the complete testing period. The second approach consists in generating one model per neighbour point, which will create a set of models to potentially generate an ensemble of forecasts. The assumption in this approach is that the misplacement error could be randomised by considering all neighbour points. This strategy would also deal with the time and season variability. While some points may be good representation of misplacement errors during certain times, other points could be better for other times.
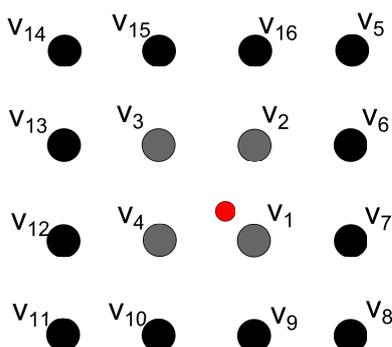
Figure 6.16: Neighbour points surrounding the wind farm location, where v_1 is the closest point to the farm (red point).

For the first approach, three different combinations of neighbour points were used. The training sets consisted of 4, 9 (v_1 at the centre of the grid) and 16 wind speed and direction points (shown in Figure 6.16) covering the January-

March period. As more variables are incorporated to the regression process, it can be expected that the resulting models will need to be more complex. For this reason, different parsimony pressures were experimented to find the ideal pressure for each case.

Table 6.5 presents the average MAE and RMSE obtained for each combination of input variables. It can be observed that the incorporation of more variables is not decreasing the error substantially.

Table 6.5: Average MAE and RMSE in kWh obtained in 50 runs on the April test set using different set of input variables.

No. of Neighbours Training Set	MAE	RMSE	Correlation Coefficient	Determination Coefficient
4	1985.5294	2597.5432	0.844361	0.671764
9	1988.14	2584.7988	0.837649	0.647170
16	2009.677	2627.9488	0.841045	0.647295

Figure 6.17 shows the set of models obtained in 50 runs for each of the configurations of the input variables.

Most models in the three training set configurations are very similar, which explains why the MAE and RMSE do not vary much. Looking at the different models, it has been observed that neighbour point v_9 is the most frequently picked up by the algorithm. This means that this point is the one that minimises the RMSE the most. However, this does not mean that this point improves the big misplacement errors like the one that occurs around hour 160. To verify this, neighbour point $v9$ was taken individually with its direction and used to find a model, as it was done with the closest point. Figure 6.18 shows the results obtained on the test set for this point.

In general, it can be observed that both points v_1 and v_9 provide a similar time placement on their predictions, however there are some areas where v_9 models some changes and magnitudes that were not seen with v_1. For example, the sudden increase around time $t + 500$, the observation shows a sudden peak which was not seen with v_1. Also, between times $t + 260$ to $t + 300$ it can be observed that v_9 can model the higher variations better. There is still an evident omission of some power peaks (increases and decreases). However, this indicates that if

135
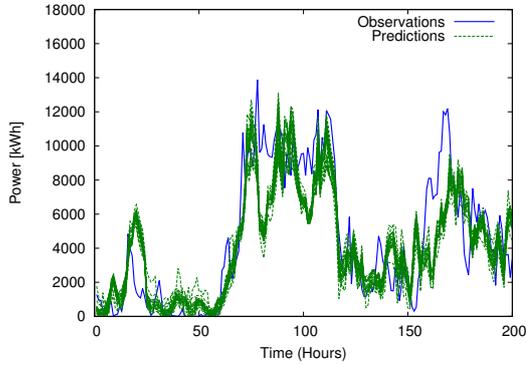
(a) Models on the first 200 hours of the test set with 4 neighbour points
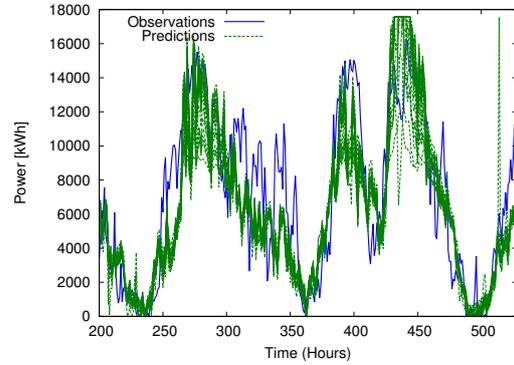
(b) Models on the following 328 hours of the test set with 4 neighbour points
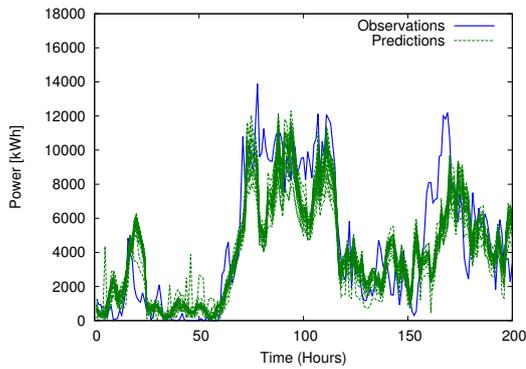
(c) Models on the first 200 hours of the test set with 9 neighbour points
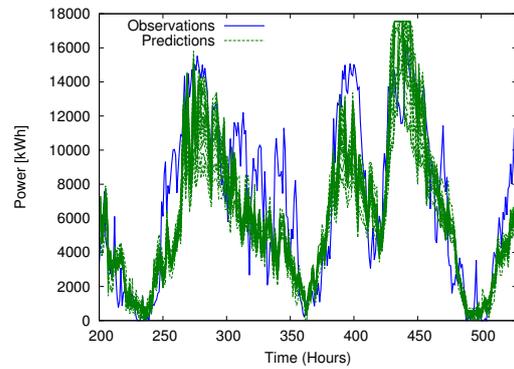
(d) Models on the following 328 hours of the test set with 9 neighbour points
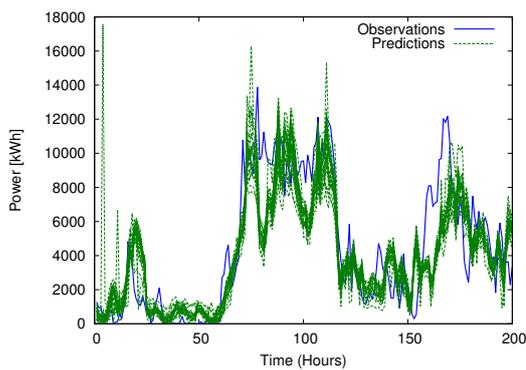
(e) Models on the first 200 hours of the test set with 16 neighbour points
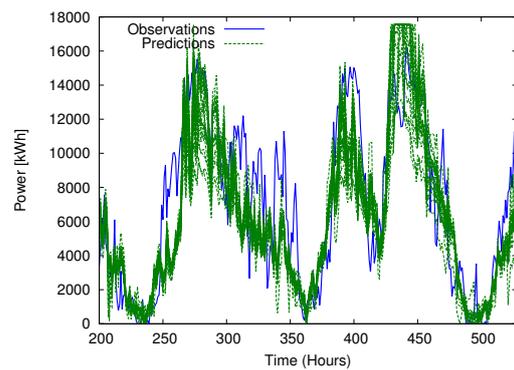
(f) Models on the following 328 hours of the test set with 16 neighbour points

Figure 6.17: Models obtained in 25 independent runs using 4 (a and b) , 9 (c and d) and 16 (e and f) neighbour points.
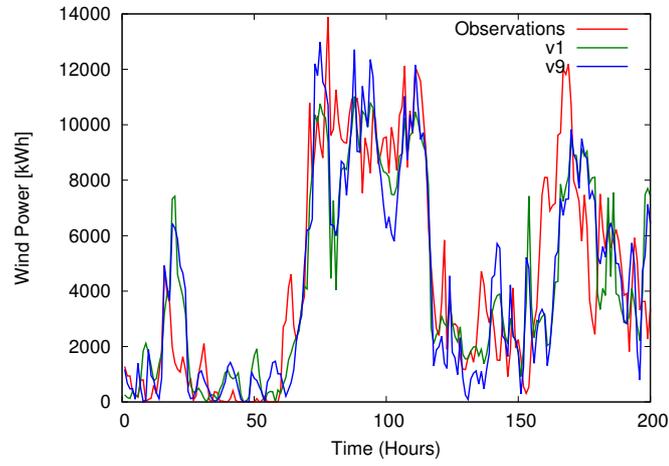
(a) t to $t + 200$
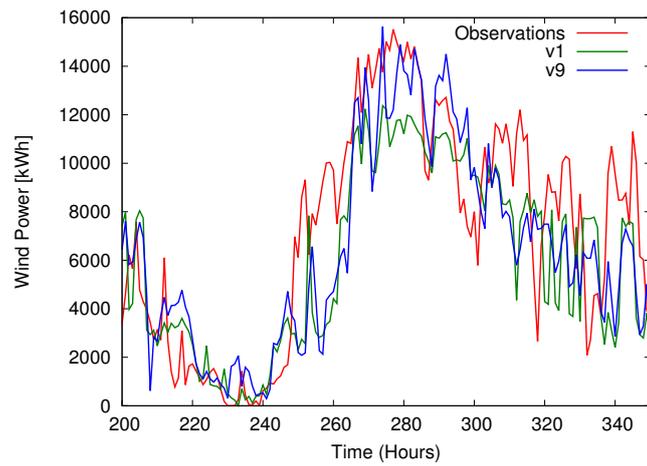
(b) $t + 201$ to $t + 350$
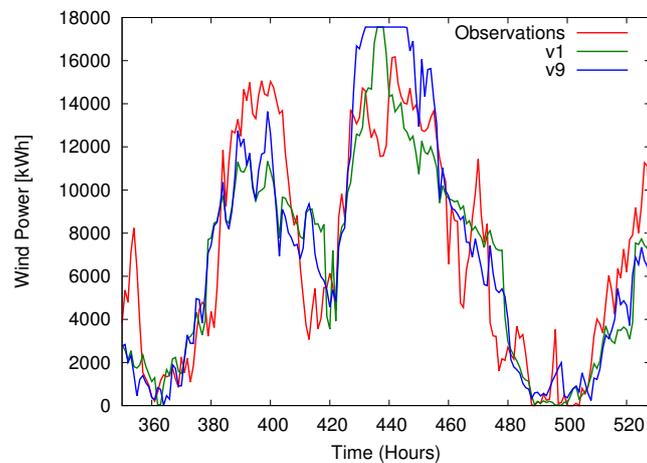
(c) $t + 350$ to $t + 528$

Figure 6.18: Wind power forecasts obtained on the test set using v_1 and v_9.

further neighbour points are explored, a better modelling could be obtained by using all points as a ensemble, not just one.
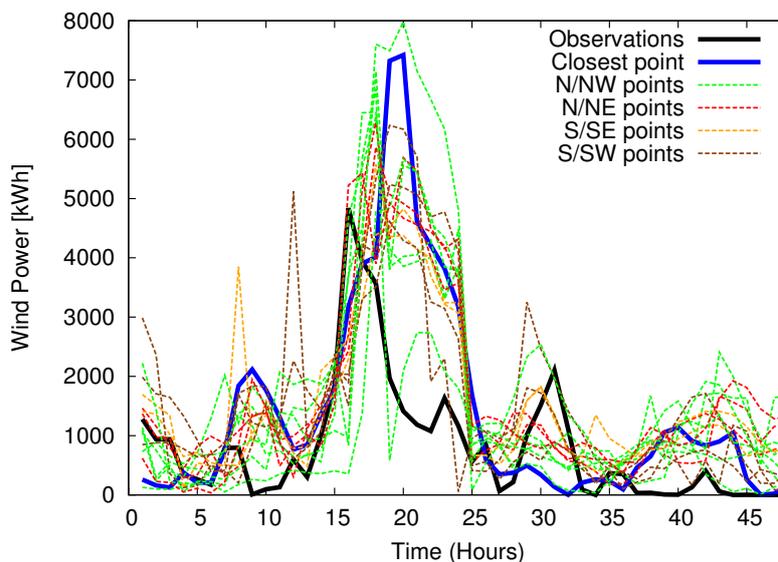
Figure 6.19: Forecasts obtained using neighbour points on the test set from t to $t + 48$. The N/NE points correspond to neighbours located at the north and northeast of the farm location. The N/NW correspond to those at the northwest, the S/SW points correspond to those located at the southwest of the farm and the S/SE corresponds to those located at the southeast of the wind farm.

For this reason, the second approach was explored. The sixteen points taken before together in one model were taken individually using the same pressure parameter that was found optimal for v_1 and using both wind speed and direction at the neighbour point. The experiment was further extended to 20 more points to have finally 36 surrounding points. Figure 6.19 show the forecasts obtained using different neighbour points that surrounded the wind farm location during the first 48 hours of the April test set.

It can be observed in the figure that the neighbours located north to the location of the farm predict the power output with a time placement just before the times of the points at the south of the farm. Figure 6.20 shows the wind direction observations at the Sotavento wind farm during this forecasted period. As the figure indicates, the wind during this period was predominantly coming from the north, which explains why neighbour points to the north would show a

time placement previous to the ones from the south.
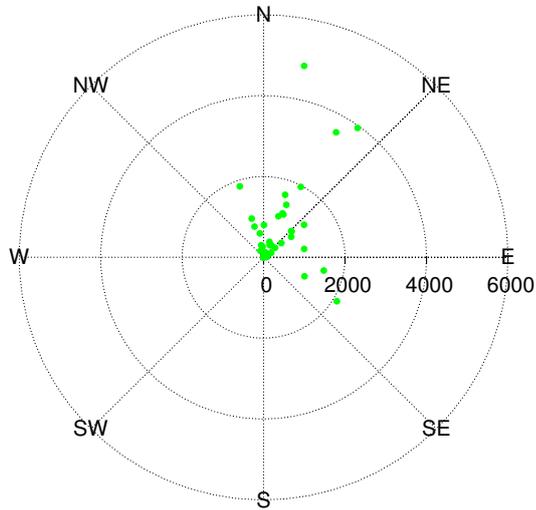
Figure 6.20: Wind power observations on the first 48 hours of the test set. Most points are located on the N/NE direction, meaning the predominant winds during this period were from the North.
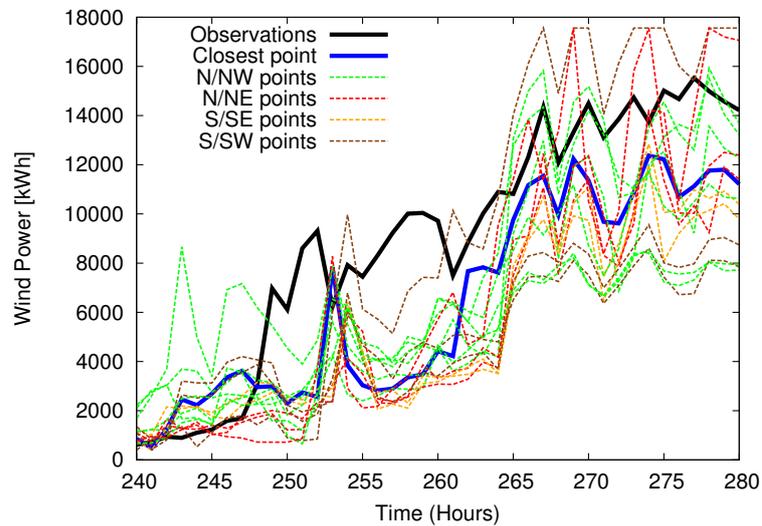
Figure 6.21: Forecasts obtained using neighbour points on the test set from $t+240$ to $t+280$.

Figure 6.21 shows another example of the differences in time placement using different neighbour points. This figure, which corresponds to the period $t+240$

to $t + 280$ (17th and 18th of April), shows that there is still a timing error in the forecasts. Between hours 240 and 255, there is a rapid increase which was modeled a few hours later. However, there are a couple of neighbors that modeled the increase with a slightly different time placement. The same situation happens at $t + 260$. Some points start the power increase before the rest. Using these forecasts as an ensemble, the spread could be used to calculate the probabilities of an increase/decrease happening in a certain period of time.

Figure 6.22 shows the wind direction corresponding to the wind predictions shown in Figure 6.21. It is interesting to see that during this period of consistent, although variable, power output increase, the wind was constantly predominantly from the west.
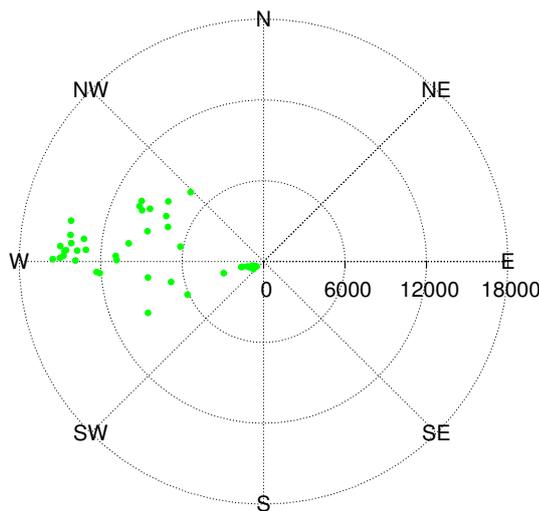
Figure 6.22: Wind power observations on the period corresponding to $t + 240$ to $t + 280$ hours of the test set. All points are located on the W direction, meaning the predominant winds during this period were from the West.

Figure 6.23 shows the wind speed and wind power observations in the April testing period. In general it can be observed that the winds come from the west and that those wind speeds observations with a east direction usually have a low power production. This wind direction tendency could mean that looking to the west points is like looking into the near "future" of what might happen at the farm. The use of a larger neighbourhood could potentially improve the

misplacement estimation by looking at the timing error of each forecast to build a distribution.

(a) Wind speed observations

(b) Wind power observations
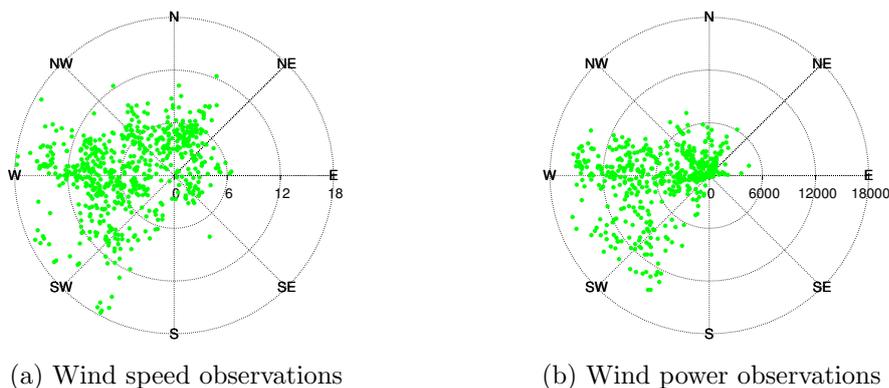
Figure 6.23: Wind speed and wind power observations on the April testing period.

Finally Figure 6.24 shows the forecasts of the most outer points of the neighbourhood during the testing period. It can be observed that there is an evident difference of each forecast that could be used to quantify the uncertainty of the power production. This will be further addressed in Chapter 7.

6.6 Discussion

This chapter introduces the use of GP for modelling the relationship between WRF-ARW forecasts and the total power production of Sotavento Wind Farm. One of the main findings of this chapter is the importance of the role that variables play in wind energy prediction, such as wind speed and direction. It has been shown that a change in direction has a direct impact on the power output. This could be specially observed on the May forecasts. The quality of the results for this period suggested that either the wind patterns in this month were not the same as for the training period, or the wind variability led to more frequent misplacement errors. Preliminary study using neighbour points suggested that some improvement could be achieved by considering other grid points. Although misplacement errors do not occur all the time, it has been found that the inclusion of other grid points can provide an idea of the possible variability of the future
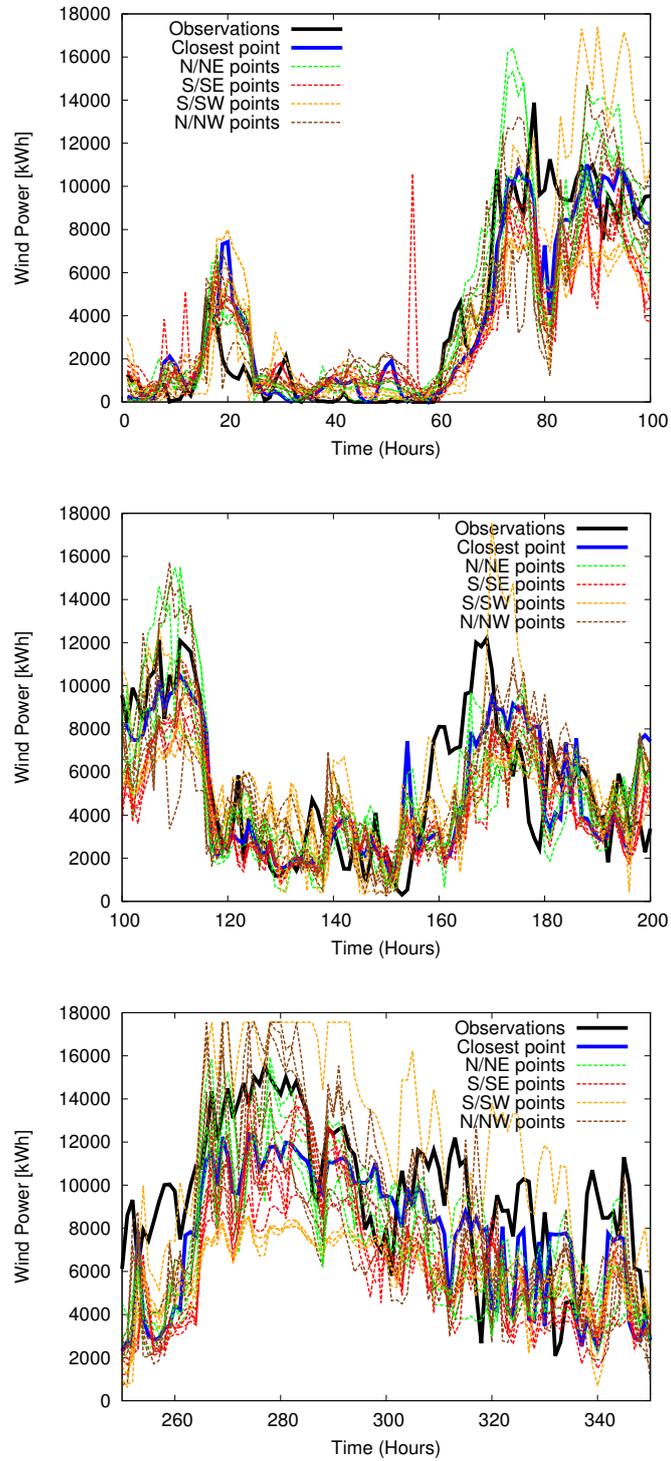
Figure 6.24: Wind power forecasts obtained using neighbour points on the April test set.

power production of the farm. The use of neighbour points was found to be better when taken individually than when taking them together as one input.

As it was previously found that a small complexity could produce models with better generalisation, the parsimony pressure in neighbour experiments was kept high. Although a low complexity was achieved, it could be noticed that for the cases of a high number of inputs, the search was biased to certain grid points; those that kept the error at the minimum level possible in such short model complexity. Neighbour points that were taken individually could show interesting power output events that were not modelled at the closest grid point. This aspect suggests that exploring a wider area could potentially be used to better characterise large changes in power output.

Both of the power prediction approaches presented in this chapter produced encouraging results in terms of point forecasts. This is an important achievement as it is a first step into the characterisation of ramp events. Another important result of this study was the quantification of the variability of the error, as this would be the second aspect to use for ramp characterisation. The results obtained show that the quantile regression forests technique is able to provide acceptable intervals, covering most of the observation points in the test set.

Chapter 7

Characterising Large Variations in Wind Power

7.1   Introduction

Wind power forecasting models are nowadays used by some grid operators for UC, however, existing tools need to be improved to be able to handle extreme situations related to wind power generation (Ferreira et al., 2010). These extreme situations, or ramp events, may be related to specific meteorological events, such as cold fronts or high pressure levels, which can produce drastic and unexpected increases or drops in the level of power production of a wind farm (Pinson, 2009). These sudden increases or drops may happen within a couple of minutes or a couple of hours. An early detection of the possibility of these events happening would let the grid operator prepare the most appropriate backup units according to the characteristics of the event. Different types of backup units respond at different speeds so an estimation of the intensity and the time scale of an event is important. The prediction of ramp events is commonly addressed using either point forecasts, obtained by running high resolution numerical models, or ensemble forecasts, which are obtained by running NWP models with different perturbations of the initial state (Bossavy et al., 2013). Despite high resolution numerical runs being avoided in ensemble forecasting, ensembles are still computationally expensive. NWP models may have misplacement errors and therefore

the closest point of the grid might not always be the best reference point from the numerical model. In addition, current ramp characterisation techniques are based on a crisp binary definition, which can leave out events that did not quite meet the amount of change expected but that could be equally important to consider.

This chapter puts together the wind power forecasting approach introduced in Chapter 6 and a ramp characterisation technique in order to address the weaknesses of current ramp characterisation approaches. The wind power forecasts and their corresponding error distribution are used during the characterisation process, which consists of a power signal filtering phase and the application of a fuzzy inference system to classify ramps. The fuzzy rules are based on the assumption that changes in power output that do not meet the strict definition of ramp event can be equally important and damaging and should not be ignored. To differentiate between highly probable events and events with lower probability, a score is given, which is the result of the application of fuzzy rules defined based on post experience. This work has been published as a journal paper in (Martínez-Arellano et al., 2014b).

The layout of this chapter is organized as follows: in Section 7.2, a basic ramp detection algorithm is presented and applied directly to the wind power forecasts obtained by using GP. Section 7.3 presents an improvement of the basic approach, taking into consideration the possible error of the forecasts. Section 7.4 presents a different detection/characterisation approach of power ramps based on fuzzy logic and is complemented by an uncertainty estimation of the timing of the events using neighbour points as ensemble forecasts. Finally, Section 7.6 presents a discussion and conclusion of the results obtained.

7.2 A Basic Ramp Detection Algorithm

One of the current challenges in wind power forecasting is the ability to handle extreme events, which can represent a different situation depending on the end user. In general, a ramp event is a rapid change in power output, either increase or decrease, within a small time window. Whether to use an increase of 50% or 30% of nominal power is, as mentioned before, up to the system operator. Wind power forecasts usually model very well the increases or decreases that happen

within large time windows (at a small change rate). However, changes in small time windows are more difficult to model correctly. Cutler et al. (2009) studied the weather phenomena that were causal of ramp events. They found that the majority of the events that were studied were associated with cold fronts, low pressure systems and troughs, which are well modelled by the numerical model but might be placed in the wrong physical position.

Another factor that might contribute to the inadequate characterisation of ramp events is the crisp definition of such events. In the literature, most studies are based on a binary definition, where the ramp is defined as a specific percentage of change in a specific time window. With this definition, events which might be slightly lower in change might not be identified but could be equally important in the eye of the operator. A good characterisation strategy should be more flexible and able to identify small but potential events.

In order to establish a benchmark for ramp detection precision, a basic ramp detection algorithm is presented, which follows the binary definition of ramp events. The approach consists in processing the power forecast several times using different time windows to look for changes of a certain percentage p until a maximum window size w is reached. To do this, a sliding window is used, which starts from hour 1 and is moved one hour ahead after verifying if the current time window has an increase or decrease of p percentage. A Pseudocode is presented in Algorithm 1.

The algorithm was applied on both forecasted and real power output of the Sotavento Experimental Park to identify the falsely and truely forecasted ramps. The time window was set to 5 hours, according to (Greaves et al., 2009). The percentage of change was set to 30% due to a very small number of real events of higher change. A total number of 21 ramps were observed in the available data from the month of April using the 30% of change determined previously. Of the 21 ramps observed, 8 of them were forecasted accurately in direction (ramp-up or down) and with a *phase error* less than ±12 hours. This time period of association is the maximum time difference between the timing of the forecast and the observed ramps, according to Greaves et al. (2009), that can maintain realistic connection between the forecast and the observed event. Events further apart might be representing totally different events. The total ramp precision and

---

Algorithm 1 Ramp detection

Require: a power signal, percentage of change and maximum window size

$p = maxPercentage$, $w = maxWindow$

$currentWindow = 1$

while currentWindow is less than w **do**

  set window pointer at the beginning of the power signal.

  while the end of the power signal is not reached **do**

    Calculate the amount of *change* from the starting point of the window to the end.

    if *change* is greater than or equal to p **then**

      identify the starting point of the window as the initial time of the ramp event

      The end time is detected in the following hours where the direction of the ramp event changes

    end if

    slide the start point of the window one hour forward

  end while

  $currentWindow = currentWindow + 1$

end while

In the final set of ramps identified, check overlaps and readjust the start and end times.

Output all ramp-up and ramp-down events with their start and end times

---

ramp recall percentage were calculated using the equations presented in Chapter 2.

Figure 7.1 shows an example of a true forecast of a ramp-up that occurred on the 9th of April. Although the observations indicate that the ramp started around 10:00 hrs, the forecast was able to model this event one hour later, from 11:00 hrs to 02:00 hrs the next day. One can also notice that the fluctuation at 10am was not modelled. It might have been caused by local conditions that were not captured at the mesoscale level, however the increase tendency was modelled effectively.

Table 7.1 shows the results of the approach described previously. It can be observed that the recall percentage obtained from the forecast is quite low. However, the result is not reflecting what could be appreciated in the signal at a first glance, as some of the ramps that were not detected by the algorithm were about 3% under the threshold. The ramps would have been detected using a smaller
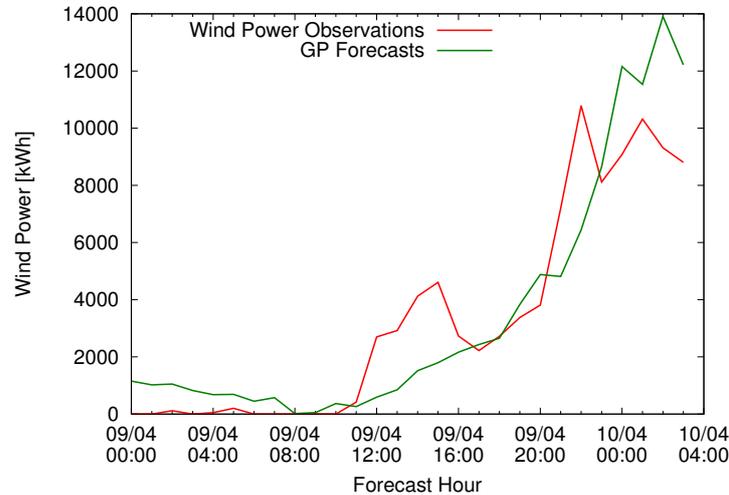
Figure 7.1: True ramp forecasted on April 9th, 2012.

Table 7.1: Forecast Precision and Recall on April, May and June 2012 at So-tavento using the model at the closest point. The observed ramps are obtained by applying the binary definition with a change of 30%.

	April	May	June
Number of true forecasts	8	3	9
Number of false forecasts	3	1	2
Number of missed ramps	14	14	15
Forecast Precision	72.72%	75%	81.81%
Recall	36.36%	17.64%	37.5%

percentage of change that could be equally important to the operator. However, taking into account the current definition of a ramp implemented in the algorithm, only those ramps that are of exactly 30% or more can be detected. Other ramps that were not captured had an error between the 5 and 10%, which can be attributed to an underestimation of the forecast. If the estimation of the error was considered in the process, it could potentially improve the ramp recall.

7.3 Incorporating the Error into the Ramp Detection Process

According to the previous analysis, the error of the model on the training set is most of the time negative (over estimation). However, the histograms show some large under estimation errors. This could be potentially related to the fact that the GP algorithm works by minimising the root mean squared error (RMSE), which during ramp events where there are timing errors might result into high penalisations. For this reason, it could be expected that some large changes in power output would be missed in subsequent forecasts. To investigate the potential of using the error distribution for the ramp characterisation, the ramp detection algorithm presented previously was adapted to incorporate the error during the calculation of the change percentage on the sliding window. The possible errors on the start and end of the window were taken into account only if by incorporating the error, the difference between start and end point increased. It is important to make clear that so far the error distribution is not time dependent, each hour is treated independently. From the quantiles obtained in Chapter 6, the median of the error (50th percentil) was used for the detection process. The reason for using the median instead of the mean, for example, is that with a high skewed distribution, the mean could be affected by few very large errors, so the use of the median helps to avoid the incorporation of very large errors that could potentially increase the number of false ramps.

The new version of the ramp detection algorithm was applied to April, May and June 2012. Results with this approach are shown in Tables 7.2 and 7.3. The ramp recall increases significantly for the months of May and June. There is not much increase in the number of false forecasts, which keeps the ramp precision at a sufficient level. For the case of May, which showed a lower precision, a neighbour point south to the location of the farm was tested. The precision shows an important improvement

These precision measurements are obtained using the binary definition on the real power output. This could mean that ramps that are apparently false forecasts could actually be "almost" a ramp on the real power output. This situation is shown on Figure 7.2. The figure shows a ramp-up and ramp-down which are

Table 7.2: Forecast Precision and Recall on April 2012 at Sotavento using the model at the closest point. The observed ramps are obtained by applying the binary definition using a change in 30%.

	Basic approach	Considering the error
Number of true forecasts	8	15
Number of false forecasts	3	5
Number of missed ramps	14	7
Forecast Precision	72.72%	75%
Recall	36.36%	68.18%

Table 7.3: Forecast Precision and Recall considering the error during May and June 2012 at Sotavento. The observed ramps are obtained by applying the binary definition using a change in 30%.

	May	May Neighbour Point	June
Number of true forecasts	6	11	14
Number of false forecasts	2	8	4
Number of missed ramps	11	6	10
Forecast Precision	75%	57.98%	77.77%
Recall	35.29%	64.70%	58.33%

not at the 30% change on the observations. The figure also shows four different forecasts using four different points from the grid. Most of the models show an increase and decrease which are identified as a ramp events. Although in the real power output this was not identified as a ramp due to the binary definition, it could still be of impact, if it is just slightly lower.

In order to avoid these "false" recalls, the binary definition of ramp events could be relaxed and changes could be categorised or scored according to certain criteria. To do this, a fuzzy rule based approach is proposed. This will allow to also consider those events, which are likely to become ramp events.
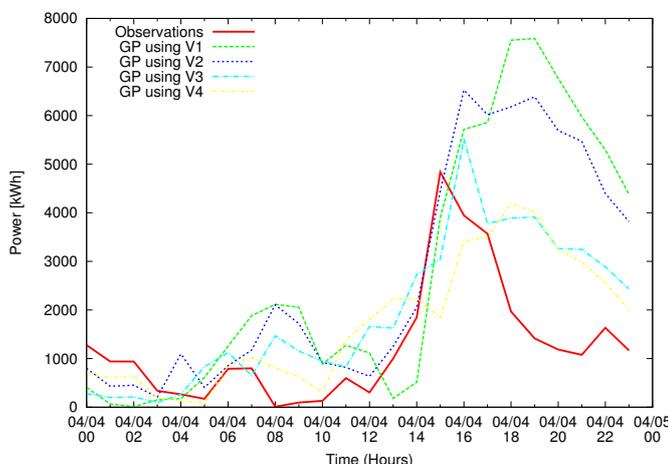
Figure 7.2: Ramp on the 4th of April using the four grid points surrounding the farm.

7.4 A Fuzzy Logic Approach for Ramp Characterisation

Fuzzy set theory, proposed by L.A. Zadeh, provides a methodology that allows to deal with the imprecision of practical systems (Zadeh, 1965). In a given system, where an output is produced according to certain inputs, those inputs or *elements* may have different states or values which represent ranges. These ranges which are not precisely defined, can be modelled using fuzzy sets. To decide whether the element belongs to one or another set, a membership function is used. The membership functions depict the degree of membership or one-to-one correspondence between an element in a domain and a truth value. Membership functions take the form

$$\mu_A(x) \leftarrow f(x \in A) \tag{7.1}$$

where μ_A is the membership function and x is an element of the set X, which may belong to a fuzzy set A. The membership functions may have different shapes according to the experience of the designer.

Once the input is mapped to a set, the process of deciding what the output should be is done by using a set of rules. Fuzzy rules describe in a high level

language how elements of the domain, which are inputs to the system, map to the outputs. This is done by applying a set of IF-THEN rules when an input has been mapped to its fuzzy set with the membership function. After the evaluation of these rules, a fuzzy set associated with each model solution variable is produced. Then, a process of *defuzzification* is used to find the value that best represents the information contained in the fuzzy set. This value is called the Fuzzy Inference System (FIS) score. Figure 7.3 represents the overall fuzzy inference system.

In terms of ramp characterisation, the decision to be made is whether a change in power is a potential ramp event. The binary classification can be avoided by characterising the event as a high, medium or low probable ramp event without discriminating it completely. The FIS score would be a measure of how probable is the ramp event. The following sections present the detail of how this is achieved.
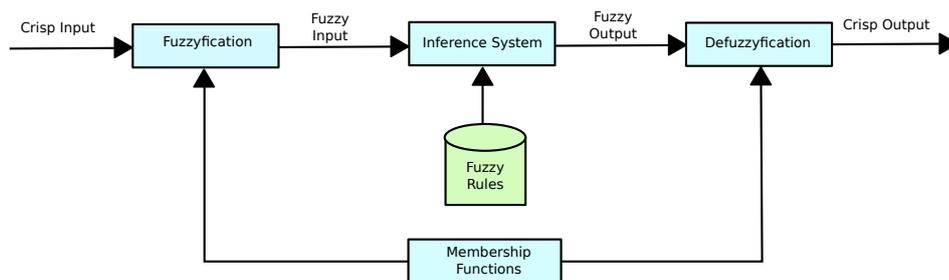
Figure 7.3: Fuzzy Logic System.

7.4.1 Inputs and the Filter Function

In order to use a fuzzy inference system for ramp characterisation, the first aspect to address is the type of inputs that will be used by the inference system to classify an event. The variables that are used in the decision making process, as well as the rules, are chosen by and depend on the knowledge and expertise of the designer. The experience of power grid operators would need to be taken into account in order to adjust the inputs, fuzzy sets and rules according to the situation of the grid.

As the purpose is to characterise a *change* in power (either increase or decrease), a straightforward option is to use the amount of change in a window of time, with the window size being a parameter that depends on the end user and

the conditions of the energy system that is being operated. This window size is important as the amount of change to be considered depends on the time frame one is looking at, which is related with how quickly the energy system can react. The window size is a clear concept drift problem as the conditions of the energy system are constantly changing, affecting on real time the size of the window. Concept drift refers to changes in the conditional distribution of the output or target variable, while the distribution of the input may stay the same (Gama et al., 2013). In the ramp characterisation problem, the classification ramp/no ramp will not always be the same for the same inputs (amount of change and window size). This research, however, is only addressing the problem as a static one, fixing the window size to 5 hours in order to develop a baseline for this method. The use of dynamic windowing is proposed as future work on Chapter 8.

The power forecast, which was obtained previously by GP, will be converted first to a percentage of total capacity, instead of using the actual power output in kWh. This is done by dividing the hourly forecasted power by the total capacity of the wind park. Once the signal is in terms of percentage, it is filtered to obtain a new signal which will indicate the amount of change at time t for a specific window of time.

For a given wind power time series P_t, the associated filtered signal P_t^f is calculated using the following equation:

$$P_t^f = |P_{t+w} - P_t|, w = 1, 2, 3...n \tag{7.2}$$

where w is the window size and n the maximum window of time to study. If the interest is to characterise changes in a time window of 5 hours, then 5 filtered signals would need to be obtained to study the changes in up to five hours. This is because if a percentage of change is only calculated in a window of size 5, then the changes that happened in between those 5 hours would be ignored. It might be the case that between hours 2 and 4, the change was produced but not noticed after the fifth hour if compared to the beginning to the time window. To avoid this situation, the procedure will be to generate multiple filtered signals using different window sizes until the maximum window size is reached. The first signal will then
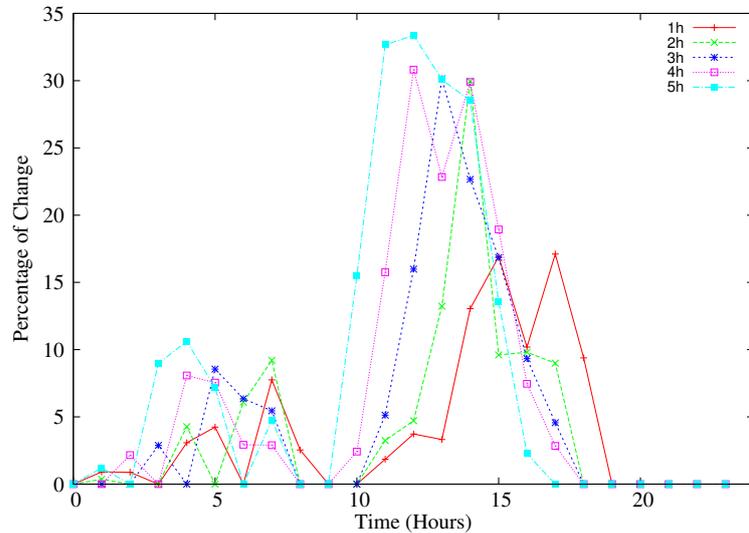
Figure 7.4: Filtered signals obtained for the first 24 hours of the test set. These show the possible ramp event on the 4th of April.

be obtained calculating the differences in power output taking a sliding window of one hour, the second signal will be obtained taking the differences on sliding windows of two hours and so on, until reaching the maximum window size. Each filtered signal will be slightly shorter than the original due to the window size. With the filtered signals ready, the next thing is to design the fuzzy inference system to process these one by one.

Figure 7.4 shows the five different filtered signals obtained for the first 24 hours of the April test set. It can be observed that a potential ramp-up event starting at $t+10$ is being predicted. All filtered signals show the ramp-up tendency at the same time and an end time can be identified at t_{18}, where the one hour window signal still shows a positive change.

Figure 7.5 shows a ramp-up event of long duration on the 9th of April. Taking a look at the signal in red (1-hour window), it can be observed that the quickest ramp rate happens at hour 71, where the signal is above the 15%. This signal also shows that the ramp-up that started at around hour 53, has a one-hour decrease (or no change), that was not visible on the other signals. The grid operator needs to make sure that the backup allocated would be quick enough to compensate for the sudden decrease and also that the rapid increase at hour 71 is not going to
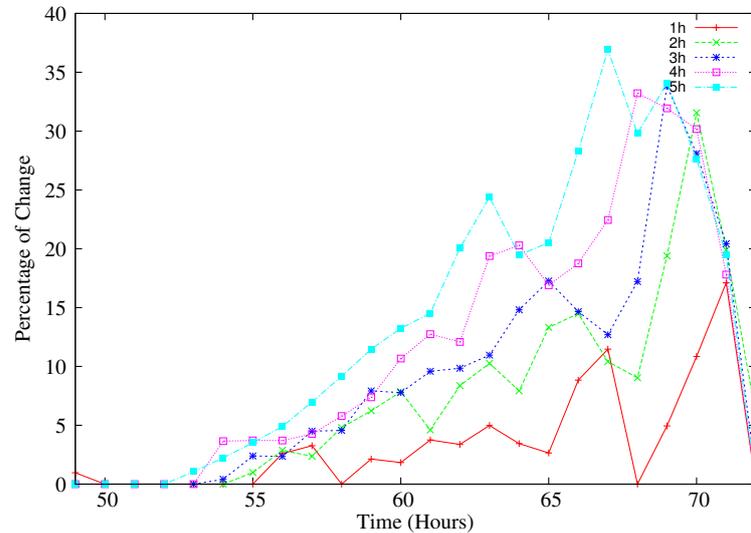
Figure 7.5: Filtered signals obtained on the 9th of April (test set).

cause instability in the system.

Another interesting aspect to point out is that in both figures, it can be observed that a change in a 1-hour window never reaches the 30%. However, from the two-hour window, one can start observing changes close to the expected percentage of change. This could help to determine the type of backup resource needed. If the rate of change during the first hour is small, the grid operator could decide to use non-spinning reserve. However, if it was the case of having a high probability of 30% changes within an hour, then a quick ramp-up backup resource would be more suitable.

7.4.2 The Fuzzy Inference System

As mentioned previously, two aspects are considered here, the amount of change in power and the time window. There is an interesting relationship between these two variables that can provide information about ramp events. When the time window is small and a large change is observed, there is a high probability that the tendency could continue in the following hours, marking the start of a ramp event. Moreover, when the time window increases to a medium size ($2 \sim 4$ hours), and if the change is high enough, there might still be a high possibility that a ramp event is happening. Finally, if the time window is at its maximum size and

the change was close to the percentage of change that defines a ramp event, then it is definitely a ramp. On the other hand, there are situations which are less likely of developing a ramp event. Situations where the time window is high and the change is low, either the increase/decrease event is occurring very slowly or the event happened in a smaller window, so in both cases the probability is low. This is opposite to a low change in a small window which can potentially still be a ramp in the coming hours.

This behaviour was translated into fuzzy sets and rules. There are different approaches for the derivation of fuzzy rules. As it is not always easy to derive these from human experts, several methods have been proposed for generating them automatically from numerical data. Most techniques involve the clustering or division of input and output data into subspaces (fuzzy regions, grid methods) and the inference of rules from these (Nozaki et al., 1996; Takagi and Sugeno, 1985; Wang and Mendel, 1992). AI approaches like Genetic Algorithms, Genetic Programming and Neural Networks have been proposed, which might require some notion of potential rules or might depend on the accuracy of clustering techniques (Akbarzadeh et al., 2008; Lee and Lee, 2005; Quek and Tung, 2001). Due to the limited size of the ramp event training set, it was decided to classify the existing training cases into few linguistic categories and derive the rules from there according to the training cases. Future work would address the automatic generation of these. Figure 7.6 shows the two input variables and member functions used. The time window size uses a triangular member function, as it is the standard. The three fuzzy sets for the time variable are slow, medium and fast, as depicted in the figure. A fast change would be a change that happened in the first three hours. A change in a medium time frame would be one taking between two and four hours, achieving its maximum score at hour three. A slow event would be one that takes the complete five hour window to happen.

The power change rate implements a gaussian membership function, as this function has a softer transition between sets. It also naturally introduces a "shoulder" near zero, suppressing the influence of small changes without introducing a third free parameter. As depicted in Figure 7.6, the fuzzy sets for the power change are divided in three: low, medium and high.

These depend in the amount of change as a percentage of the total power

Member functions for input number 1 named time

Member functions for input number 2 named change
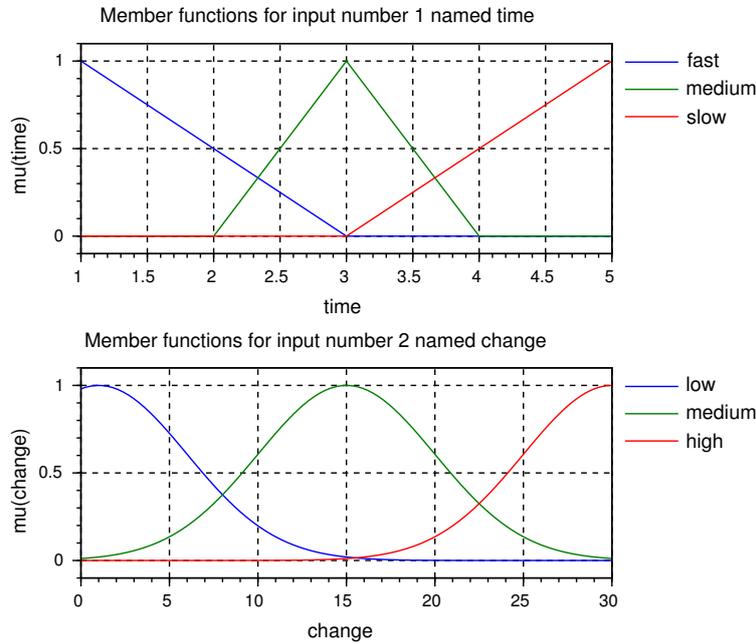
Figure 7.6: Member functions for variables time and rate.

production of the wind park. The rules are shown in Table 7.4 and using the Mamdani Inference Model, a surface plot of these was generated (Figure 7.7). Colours ranging from the magenta to orange, depict events that are ramp events or can potentially be ramp events, while yellow to blue depict changes which are unlikely to turn into ramp events.

With these rules and the fuzzy sets, a softer decision of what a ramp event is could be made, allowing the identification of smaller equally important events.

Table 7.4: Rules of the fuzzy inference system.

Rule 1.	If *change* is *high* then *rampSeverity* is *high*
Rule 2.	If *time* is *fast* and *change* is *medium* then *rampSeverity* is *midhigh*
Rule 3.	If *time* is *medium* and *change* is *medium* then *rampSeverity* is *midhigh*
Rule 4.	If *time* is *slow* and *change* is *medium* then *rampSeverity* is *medium*
Rule 5.	If *time* is *fast* and *change* is *low* then *rampSeverity* is *medium*
Rule 6.	If *time* is *medium* and *change* is *low* then *rampSeverity* is *medium*
Rule 7.	If *time* is *slow* and *change* is *low* then *rampSeverity* is *low*
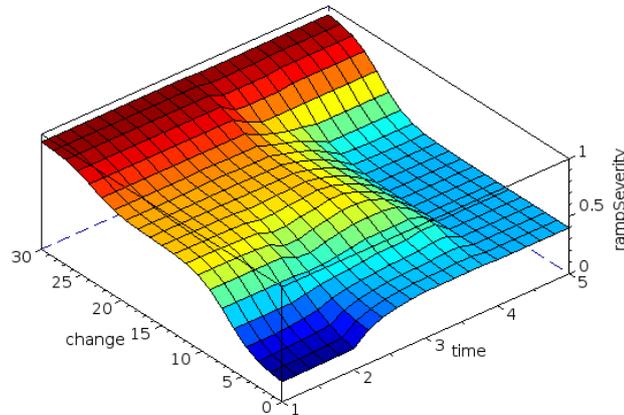
Figure 7.7: Surface plot of the solution domain according to fuzzy rules. Colours ranging from the magenta to orange depict potential ramp events, while yellow to blue depict changes which are unlikely to turn into ramp events.

7.4.3 Applying the FIS for Ramp Characterisation

To test the fuzzy inference system, the same characteristics of change were used as with the basic ramp detection approach studied previously, which considered ramp events as a change of 30% in 5 hours or less. For this kind of ramp events, 5 filtered signals were obtained and then the FIS was applied to each one individually. Each value of the filtered signal was taken as a new input, and the window size was chosen according to the filtered signal being tested. Some of the results are shown in Figure 7.8. Colour maps were used to facilitate the interpretation of the results. As the human eye is sensible to the difference in colour, differences can be captured more easily than when results are presented as quantities. In addition, the shapes provide some interesting facts about the characteristics of the event. A maroon colour corresponds to a FIS score of 1, which represents a high severity. A light maroon/crimson colour corresponds to a mid-high severity which is between 0.7 and 0.9 score. The yellow/green corresponds to low severity events, while the blue colour is definitely not a ramp event. The ramp-up and down events were separated into two graphs to avoid colour maps being too saturated. The top figures are the actual power signal, the figures in the middle show the ramp-up events and the last row shows the ramp-down events.

Focusing on the second line of the figure, the ramp-up events, it can be ob-
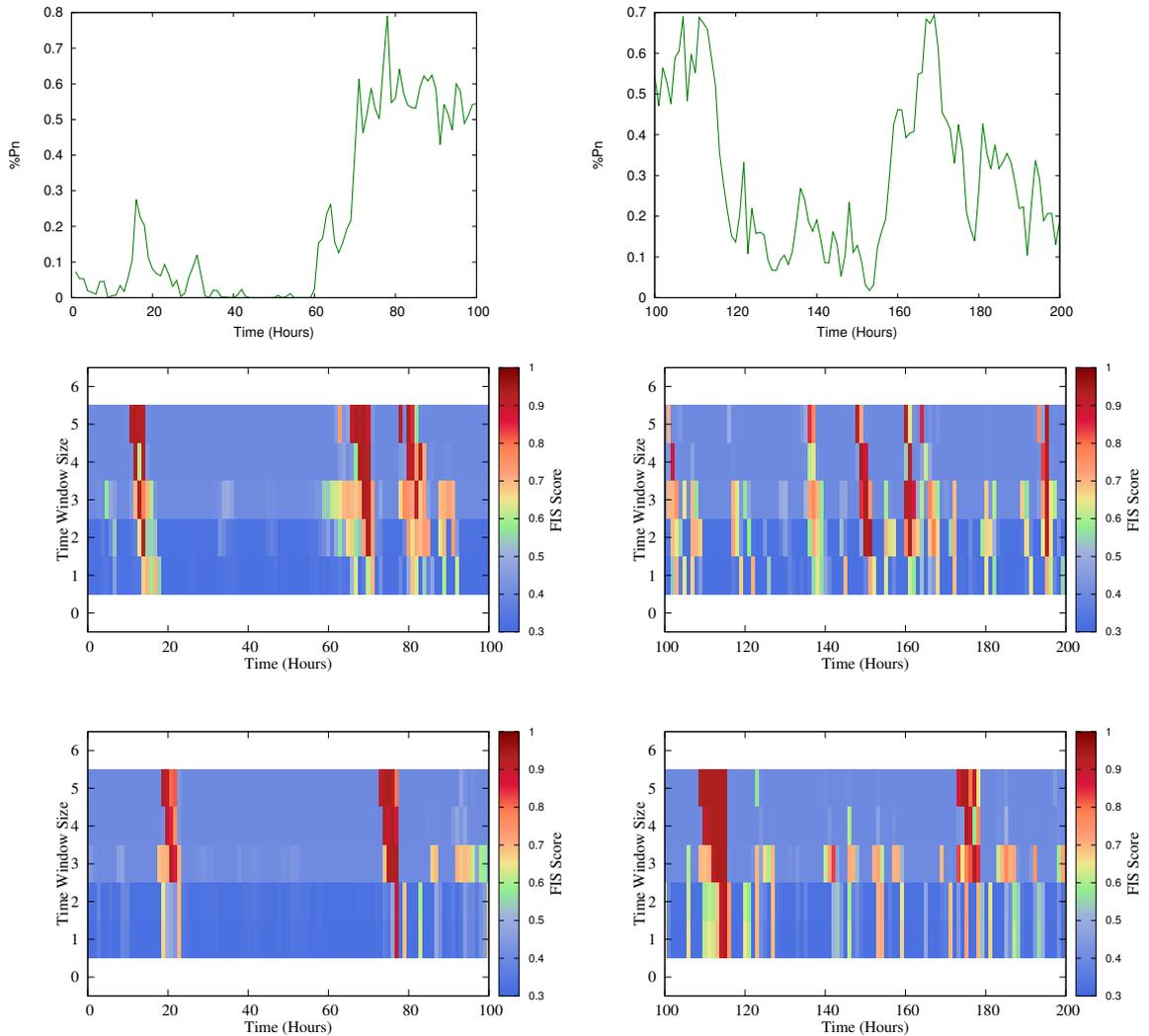
Figure 7.8: Real power output and fuzzy inference system scores corresponding *ramp-up and ramp-down* events on the same time period. The graphs on the left correspond to the first 100 hours after the last training point. The graphs to the right correspond to hours 100 to 200 after the last training point.

served that the largest change (between hour 60 and 80) is the strongest in FIS score along the 5 filtered signals. A dark colour in all filtered signals at the same time (a straight line from 1 to 5 in the Figure) would mean that the change started quickly with a high rate and lasted for some hours. These are the changes that might pose the worse threat as the grid operator might need backup units that

159

can be turned on or ramped up/down quickly. The step shape of some of the lines indicate the event started with a low change rate and progressively speeded up. A wide segment on the 5-hour signal would indicate that the change continued to increase even after the 5 hour window. If the event was shorter, strictly limited to 5 hours, the line would be thinner. The step shape is not reflected, for example, on the change detected after hour 160. A straight line is shown in this case, meaning that from the very start of the event it increased with a very prominent slope. The ramp-up observed at hour 80 is a false ramp due to large error on the forecast. The corresponding ramp-down of this false ramp-up can also be observed on the bottom of the figure. Another interesting thing to mention is that the "false" ramp-up before hour 20 is shown with a slightly lighter colour meaning that the total change in the 5 hour window did not reach the 30% change, nevertheless it is an important increase to take into account. Looking more into the bottom part of the figure, it can be observed that the largest ramp-down events are correctly identified, at hours 120 and 180. The original power signal has several small increases and decreases which can be seen in the filtered ones as noise.

All these characteristics that can be highlighted through the use of the colour maps can improve the identification of the ramps as more information about how the event will develop is given. The results show in general the capacity of the fuzzy system to identify the events that were previously identified using the ramp detection algorithm, and in addition to this, presents some additional events which are categorised as potential ones. These additional events are presented in lighter colours as they represent less risk and therefore less probability of actually happening. It might still be possible that the forecast signal even with the error taken into account could not represent the real intensity of the event due to misplacement errors, and that other points of the grid could highlight better these changes. It would be interesting to see if exploring other closer locations could provide even more important evidence of these changes.

The fuzzy inference system was applied to some neighbour points to see if other type of information could be revealed that the closest point was not showing. April power forecasts obtained with neighbour points v_3 and v_{13} (points to the west of the park) were processed, filtered and characterised with the FIS. Figure
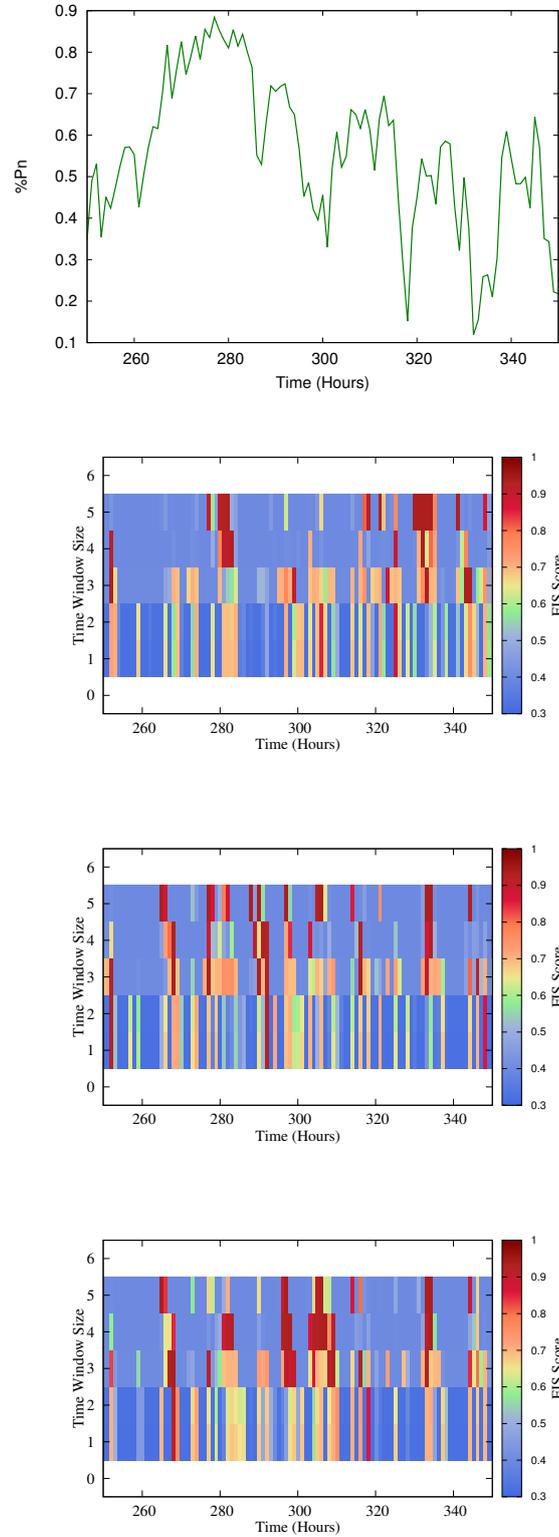
Figure 7.9: Real power output and fuzzy inference system scores corresponding to *ramp-down* events at three different points during the same time period (250 to 350 hours after the last training point).

7.9 shows some results applying the fuzzy approach to these points. In general, there is a match in the three points about the events shown, although they present different intensities. The interesting aspect to see in this figure is the detection of two events around the 300th hour by the second neighbour (bottom right). These frequent changes in power output were not well modelled by the closest point of the grid. However, a neighbour point to the west of the wind farm is starting to represent these events which are not well located in time but they can indicate a period of various changes that might be important to consider from the system operator's point of view. This is consistent with the wind direction observations which in general come from the southwest of the farm during this period.

A second subset of neighbour points was taken, this time further apart from the location of the farm. The points v_{17}, v_{22}, v_{27} and v_{32} which correspond to the far northeast, southeast, southwest and northwest of the 36 grid points explored before. Figure 7.10 shows the FIS scores obtained in the first one hundred hours of the test set. It can be observed that all four points agree on the possible development of a ramp-up event during the first 20 hours. The following three ramp-ups (at around hours 60, 70 and 80) are forecasted again by all neighbours but showing different probability. It can also be observed that the northeast point shows the higher probabilities. The same behaviour can be observed in Figure 7.11. The northeast neighbour highlights the events with higher probability and the larger ramps (around hours 150 and 160) are well identified. During the first 10 hours of this period, most neighbours indicate two possible ramp events, which in the real output are changes of around 20%. These were not detected as real ramp events with the binary definition, however, most neighbours highlight a moderate event.

Finally, it can be observed that during the test period on Figure 7.11, all increases are modelled in at least one neighbour, although different severity. This leads to think that if all 36 neighbours are considered, this could be used to compute an overall probability of the event happening at certain hour. Bossavy et al. (2013) propose a similar approach, however the authors use ensembles produced by the perturbation of the initial state of the numerical model, not ensembles by considering a set of points as proposed in this thesis. Also the authors base their approach on the binary definition of ramp event. The timing of
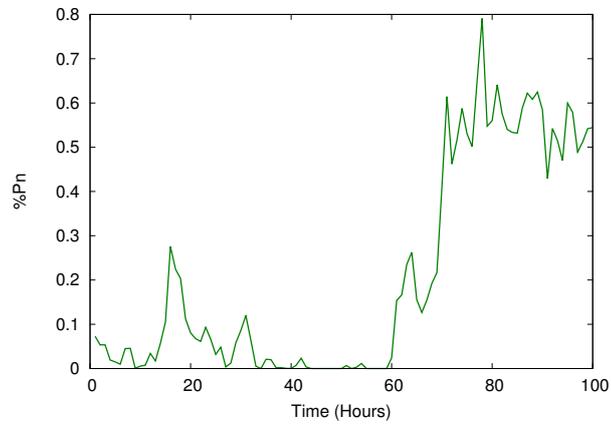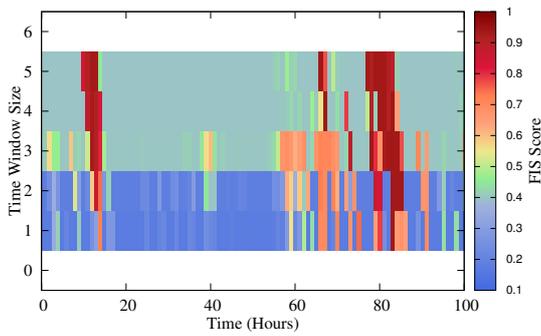
(a) Power Observations

(b) Northwest Neighbour
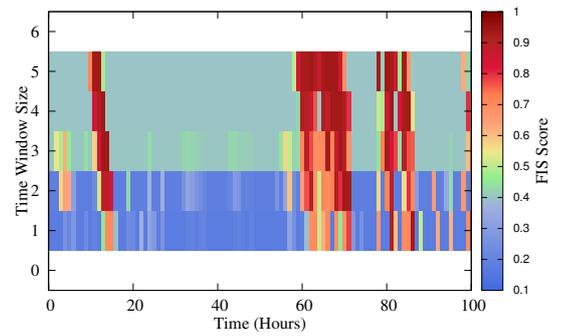
(c) Northeast Neighbour
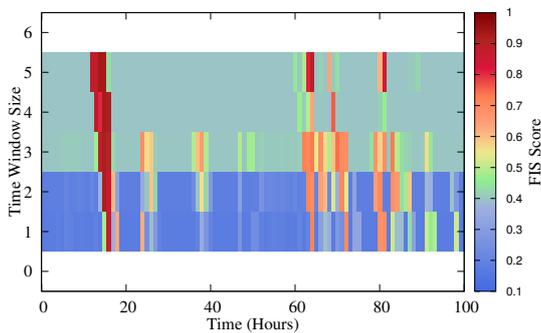
(d) Southwest Neighbour
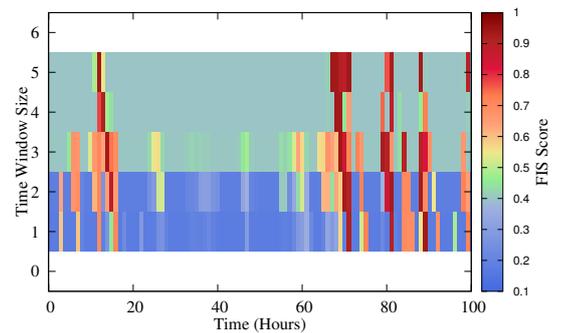
(e) Southeast Neighbour

Figure 7.10: Real power output and fuzzy inference system scores corresponding to *ramp-up* events at four neighbour points during the same time period (0 to 100 hours after the last training point).
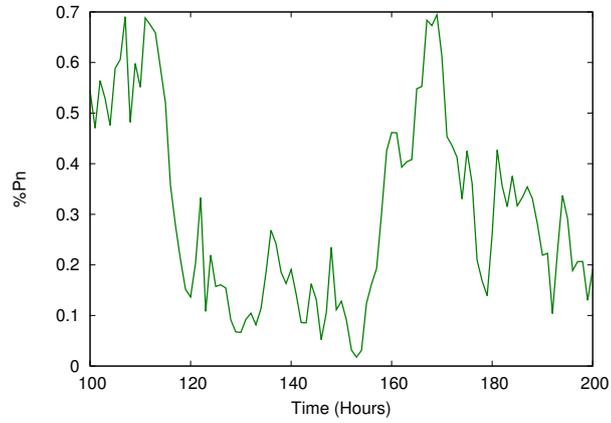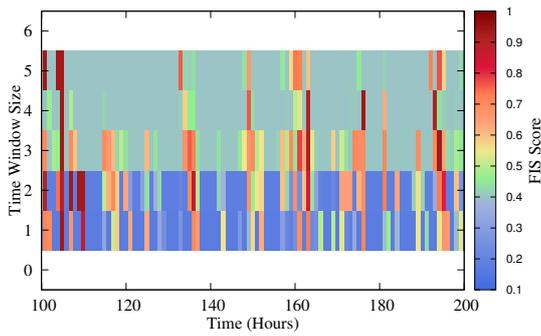
(a) Power Observations

(b) Northwest Neighbour
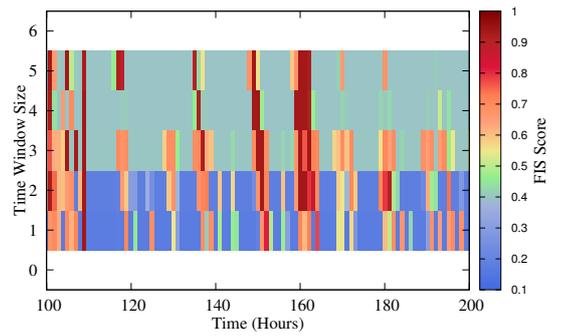
(c) Northeast Neighbour
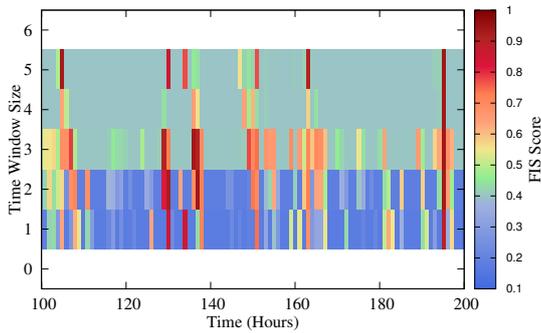
(d) Southwest Neighbour
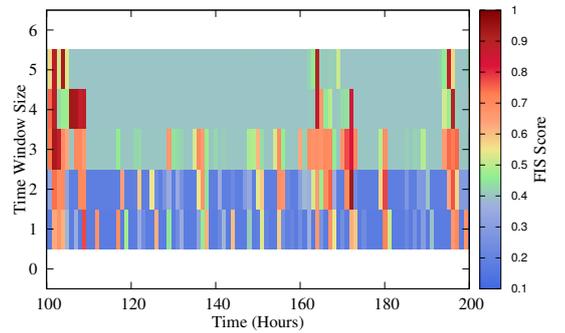
(e) Southeast Neighbour

Figure 7.11: Real power output and fuzzy inference system scores corresponding to *ramp-up* events at four neighbour points during the same time period (100 to 200 hours after the last training point).

the event is an important aspect to be characterised as the grid operator needs to define the available backup resources at each hour. The following section presents a neighbourhood approach for the estimation of the timing of the events.

7.5 Using A Neighbourhood Ensemble for Estimating the Ramp Timing Uncertainty

The prediction of the timing of a ramp event is a task even more challenging than predicting the duration. Phase errors have a high impact on the grid stability and their characterisation has become an important aspect to be addressed in future forecasting tools, as most state-of-the-art tools not have a dedicated module to characterise them. A neighborhood ensemble approach for determining the timing of an event could address the misplacement errors from numerical predictions by giving the grid operator more information of the possible time intervals where the event might happen.
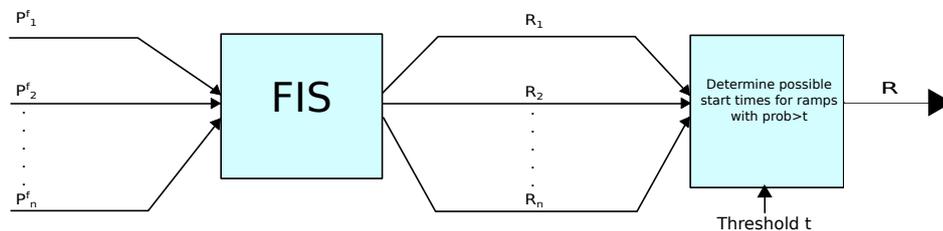
Figure 7.12: Ramp timing characterisation process.

The proposed approach for the timing characterisation, as depicted in Figure 7.12, consists in applying the fuzzy characterisation approach to each of the predictions obtained using each neighbour point individually and determining the different possible timings of the potential events (those above a certain FIS score) according to the ensemble. To determine the timings of ramp event for one grid point, the five filtered signals are used. First a FIS score threshold τ is defined. Then, for each forecasted hour t, the filtered signal that has the highest FIS score at that time and that is higher than the threshold, is marked as the beginning of the ramp event that will have a duration depending on how many hours the FIS score is not zero. This process is repeated for each neighbour point until for

Table 7.5: Forecasted timings for the ramp observed on the 9th of April at 5:00 pm.

Starting time	Number of Neighbour Points
11:00	2
12:00	1
14:00	1
15:00	3
16:00	1
17:00	1
18:00	4
19:00	7
Total	20

each ramp event a set of possible start times is obtained. The possible starting points for one event would determine an interval and the probability of the ramp event starting at a certain hour would be estimated according to the number of neighbour points that predicted that starting hour.

The process described above was applied to the 20 outer grid points to Sotavento for April 2012. Table 7.5 presents the different starting hours and number of neighbour points that forecasted the ramp-up observed on the 9th of April using a threshold of $\tau = 0.8$. This particular day, two contiguous ramp-up events were observed, although only one was detected under the binary definition. The general trend during the day was an increasing one, except for a slight decrease in the middle of the day. This can explain why some neighbours detect the increase from as early as 11 am. From the 20 ensemble members, 12 predict timings after 17:00 pm, which corresponds to the 60 % of them, 85% predict the starting time from 14:00 pm. With the FIS scores presented on the colour maps, the grid operator would still be able to see that there is a potential increase in the morning of the 9th of April, which is related to that 15% that is forecasting an early ramp-up. It would be up to the consideration of the grid operator to determine if this increase is representing a risk according to the scheduled back up.

Table 7.6 shows the starting hours and number of ensemble members that predict a ramp-up event on the 26th of April which was not forecasted by the closest point. From the 20 members, 12 predicted the event with high FIS scores,

Table 7.6: Forecasted timings for the ramp observed on the 26th of April at 08:00 am.

Starting time	Number of Neighbour Points
00:00	2
03:00	1
06:00	1
09:00	1
11:00	5
12:00	1
15:00	1
Total	12

indicating again the probability of an event. From the 12 members, 5 predict a start time at 11:00 am and the other seven are spread between 12:00 am and 15:00 pm, which is still in the 12 hour margin suggested by Greaves et al. (2009).

7.6 Discussion

This chapter presents a novel approach to wind power ramp characterisation. The approach introduces the use of the error distribution and fuzzy logic rules to improve the characterisation of ramp events which might not be identified by using a binary classification. According to the results shown in Table 7.2, the percentage of ramp recall improves as the error is taken into account. In addition, the introduction of fuzzy rules provides information about possible events which were not forecasted with a change of 30 % but could potentially be of interest to the grid operator especially during periods of constant fluctuations. The exploration of neighbour points could give information of events that are not predicted at the closest point, as it is the case of May predictions. The visual presentation of the results can help to interpret different aspects of the development of the event. Although the number of "false" ramps may increase with the fuzzy approach for ramp detection, it will be able to flag more potential events, which is equally important giving the operator the information to decide whether it should be considered a ramp event or not. The ramp rate of the events, which

can be determined by looking into the size of the time frame where the change happens, can help determine the type of backup needed. As it has been previously mentioned, the ramp rate is an important characteristic of ramp events that will determine how quick the backup resource should be able to start and reach its nominal load.

In addition to the characterisation of the duration and change rate of the event, it was possible to provide an estimation of the possible starting times of the events. Neighbour points were used as an ensemble, where the number of members were used to calculate a probability of an event happening at a certain hour.

It is important to take into account that these results are related specifically to the Sotavento Experimental Wind Farm. Each wind farm has different terrain characteristics that affect the power production of the farm in different ways. The advantage of GP is that no assumptions about the model are needed. This does not mean that the model could not be improved with local information, specially for the smaller fluctuations. However, for a day-ahead forecast, this approach provides a wider picture of the possible events at the farm.

In addition to the two variables used for the fuzzy inference system, there are other potential variables that could bring important information to the characterisation, such as temperature and pressure. It is well known that changes in these variables are related to the development of ramp events and these variables at different points may suggest changes that were not characterised at the closest point. Some detail of how these variables could be potentially used in the characterisation process will be developed in detail in the future work section. Finally, from the system operator's point of view, the fuzzy logic approach would allow a better understanding of how the events will develop over time in addition to the traditional point and probability day-ahead forecast that is provided by current state-of-the-art tools.

Chapter 8

Conclusions and Future Work

8.1 Summary

The work presented in this thesis is a novel approach to wind power forecasting and to the characterisation of wind power ramp events, which is currently an area of great interest because of its importance in the integration of wind energy into the electricity market. Based on the results of this research, it can be concluded that a good characterisation can be achieved by taking into account the information of the surroundings of the wind farm being studied. This, coupled with a softened definition of a ramp event, allowed the characterisation of potential events which the common binary definition used in the literature is not able to achieve.

The prediction of the duration and time of ramp events as well, as the characterisation of how these events will develop in small windows of time, is critical to the improvement of backup allocation and the stability of the electric grid. To achieve this characterisation, this research aimed to study the use of computational intelligence techniques to successfully integrate the information of numerical weather prediction models with observations from a wind park. This integration consisted of two main tasks: first, the use of genetic programming, which allowed the researcher to model the conversion of numerical predictions at different grid points to wind power forecasts. This on its own was a novel attempt for wind power prediction. Using GP as a downscaling technique or as

a wind-to-power conversion technique brought in its own interesting results on the importance of different meteorological variables in the conversion process. In addition, most state-of-the-art approaches, either point or ensemble forecasts, are based on the closest point. Few studies have been published regarding the use of neighbour points and especially of using these as an ensemble. The second important task conducted in this research was the introduction of a fuzzy logic based approach to characterise ramp events based on the way they evolve through a time window. The use of a fuzzy inference system provided the means to characterise ramp events which were not detected using a binary definition, but that might potentially become ramps as power forecasts usually present underestimations. The ramp events were classified from highly probable to unlikely, providing a percentage of the probability of the event happening. This, together with the use of colour maps, would allow the grid operator to decide either to take risks or allocate more backup units.

Another novelty introduced as part of the characterisation process was the use of the error distribution to calculate the possible amount of change in a time frame. This was an essential part of the signal filtering process that improved the ramp detection accuracy. Even the most precise forecasts provided by state-of-the art tools have a level of error. Most ramp detection techniques use a binary (crisp) definition directly on a wind power forecast, without taking into account that the forecast itself has an error that may cause false "unseen" events (Bossavy et al., 2013; Ferreira et al., 2010; Greaves et al., 2009). Estimating the distribution of the error is not straightforward, as the power forecasts do not follow typical distributions. This research explored the use of quantile regression forests to estimate the error distribution of the wind power forecasts obtained using GP. This distribution was then used to compute the possible large variations that the forecasts could present. Despite the slight increase of "false" ramps obtained after introducing the error in the filtering process, the total ramp capture increased, which is a fair tradeoff.

8.2   Conclusions

The following sections present in detail the most important points that were concluded during, and at the completion of, the project.

8.2.1   A Successful Application of Genetic Programming as a Downscaling and Wind-to-Power Conversion Technique

The major and most important part of the work undertaken in this research project was the study of Genetic Programming as a symbolic regression technique for downscaling and power conversion. A thorough experimentation was essential for the successful implementation of the technique. GP is not a "plug-and-play" technique as there are many aspects to address in order to apply this approach for regression. Aspects such as deciding on the type of input data and its pre-processing, if necessary, the selection of the function set and the type of strategy that will be applied to avoid overfitting are critical. The first experiments that were carried out were aimed at using GP only as a downscaling technique to see if numerical wind speed predictions could be improved and to avoid high-resolution numerical runs. The results were promising and showed the potential of GP, as it could improve methods that made assumptions of the mathematical model. The results obtained also showed that the size of the training set is critical especially when trying to use a large number of input variables. The use of temperature and the solar cycle did not decrease the overall RMSE; however, at that stage of the research, the experimental data available was not large enough. This means no conclusions could be made at this point regarding the incorporation of additional variables into the downscaling process.

On the other hand, in terms of the wind-to power conversion process, the experimental results showed the importance of taking into account the wind direction in the model construction. With the results obtained previously for wind speed downscaling, it was decided to carry out the first wind-to-power experiments using only wind speed and incorporate later on the wind direction to be able to conclude if this second variable was important or not. The results ob-

tained show how wind direction could add the variability that the power curve needed to better approximate the real behaviour of the power production of the wind park.

The fact that GP does not make any assumption of the model, facilitated the application of the approach to different grid points, which was one of the important aspects to study for ramp characterisation. The application of GP as an "ensemble generator" enabled treating each grid point as if it was the actual location of the wind park and to study potential displacements. The creation of the ensemble via GP allowed estimating different possible outcomes of the wind park, which were used to forecast potential ramp events and their timings.

The strategy for achieving good generalisation models was also a key aspect to study. It was shown that more complex models were better able to learn the training set but could perform badly on unseen data. Here, the training set played an important role, as the training data must represent as much of the search space as possible. This becomes more complex as variables are added to the terminal set. In terms of ramp events, it would be difficult to ensure that a training set is covering a complete search space. The reason for this is that most meteorological events are different from one another, so an increase in wind speed would not always be converted to a specific increase in power, because a slight change in wind direction may affect this conversion. This is one of the main factors that influenced the amount of training data that was used.

An interesting aspect of using GP that was not fully explored but could be of interesting further research is that it produces a mathematical representation of models that could help understand more about the process that is being represented. Some of the experiments that were done using multiple grid points in one run could show the grid points that were more likely to be picked up. This could provide more understanding of what was happening in terms of displacement errors.

8.2.2 Using Neighbour Points as One Input Set and Individually

In order to study the use of neighbour points introduced by Cutler et al. (2009), it was important to determine how these would be incorporated into the wind power and ramp prediction process. The first aspect that was studied was how to input these into the GP algorithm, all in a single run, or each one separately. The results confirmed what was expected. Using points together in a single run could give information about the most relevant points/variables in the neighbourhood. However, using only certain points to forecast the forthcoming month could not be the best strategy due to the high variability of the area of Galicia. Considering all points separately could give a better idea of the different possible outcomes of the wind park. In addition, according to the results, taking each point individually could be used as a way to estimate possible timings of large changes in power output.

8.2.3 The Ramp Detection Process

Most of the research published on wind power ramp detection and characterisation uses the binary definition as the main approach to the detection of such events. However, this crisp definition can lead to artifacts, missing events that might have been detected if a less rigorous definition was used. Even if a forecast was 100% accurate, one could think that a change for example between 28-29% could be as damaging or dangerous as a change of 30% (assuming the detection has been done using a 30% amount of change). For this reason, two major aspects were of interest to study. First, taking into account that forecasts always have a percentage of error, the research questions where how to calculate this error and how to use it to improve the ramp capture. Finding the error distribution of wind power forecasts is a research area on its own and a quick and effective approach was to use a state-of-the-art technique that could calculate this distribution without any assumption of its shape. Using wind speed and direction as explanatory variables, the error quantiles were obtained and used in the ramp detection algorithm. To quantify the level of ramp capture of this new approach, a basic detection algorithm based on the binary definition was used

as a benchmark. The experimental results showed that an improvement on the ramp capture could be obtained when considering the error of the forecast. It was expected to see an increase in "false ramps" that could potentially affect the performance of the approach; however, it was found that the increase in false ramps was a fair tradeoff compared to the number of ramps now being detected.

Another interesting aspect to look at was the use of neighbour points to identify potential ramps that were not detected at the closest point. Using these as an ensemble, a probability of a ramp happening at certain hour could be calculated.

8.2.4 A New Ramp Characterisation Approach Based on Filtered Signals and an FIS

As the FIS designed was using primarily the amount of change and time window as input variables, it was necessary to predict these changes with the best estimation possible. For this reason, it was decided to "break" the original power signal into different change signals (according to a time window) using the error distribution as done for the binary ramp detection. These signals, also referred to here as filtered signals, allowed the observation of the type of changes that were happening in the original signal at different time scales. The decomposition could tell us about the ramp rate and how quickly the change was happening. This information is already valuable on its own as this is critical for an adequate backup allocation.

The use of the FIS showed that it was possible to detect all ramp events that a common detection tool would detect. A major contribution into the ramp characterisation process was the use of the filtered signals together with the FIS to determine the probability of potential changes developing into real ramp events. This is an aspect that most state-of-the-art tools do not address. By providing a probability to the grid operator, combined with his expertise, a more informed decision regarding the allocation of backup resources could be made.

8.3 Suggestions for Future Work

Further investigations, in which future works could proceed, are listed below:

1. All numerical predictions were extracted at a specific height using the vertical interpolation tools from NCL. However, it is known that NWP models might have also vertical misplacement errors and that errors at different altitudes might be different due to topographic characteristics. Further investigation could be done using numerical predictions at different pressure levels to study the implications of using NCL interpolation methods compared to the use of different pressure levels into the regression process.

2. An important aspect to explore further is the incorporation of variables such as temperature and pressure into the fuzzy inference system. As different studies have shown, the cold fronts and changes in pressure can be major contributors to ramps. Figure 8.1 shows the wind power and temperature forecasts as well as temperature observations that were obtained from two meteorological stations, Marco da Curra and Coruña Dique (MeteoGalicia, 2014), close to Sotavento.

   The figure shows that, although both meteorological stations are located in the same region, the temperature varies significantly. It would be interesting to investigate if using the tendencies of change in temperature and pressure of the neighbourhood in the rules could improve the characterisation of those events that are not forecasted with a binary approach. This would require other tasks such as making sure the training and test data contains ramp events that developed during the meteorological events and also ones that were not caused by these.

3. There is potential in the use of the neighborhood forecasts to improve the time characterisation of ramp events. Further study could look into exploring wider areas and exploring other ways to merge all this information to generate one probabilistic forecast.

4. With no information available of a real market, it is difficult to measure the benefits of the probabilistic forecast in the backup allocation process.
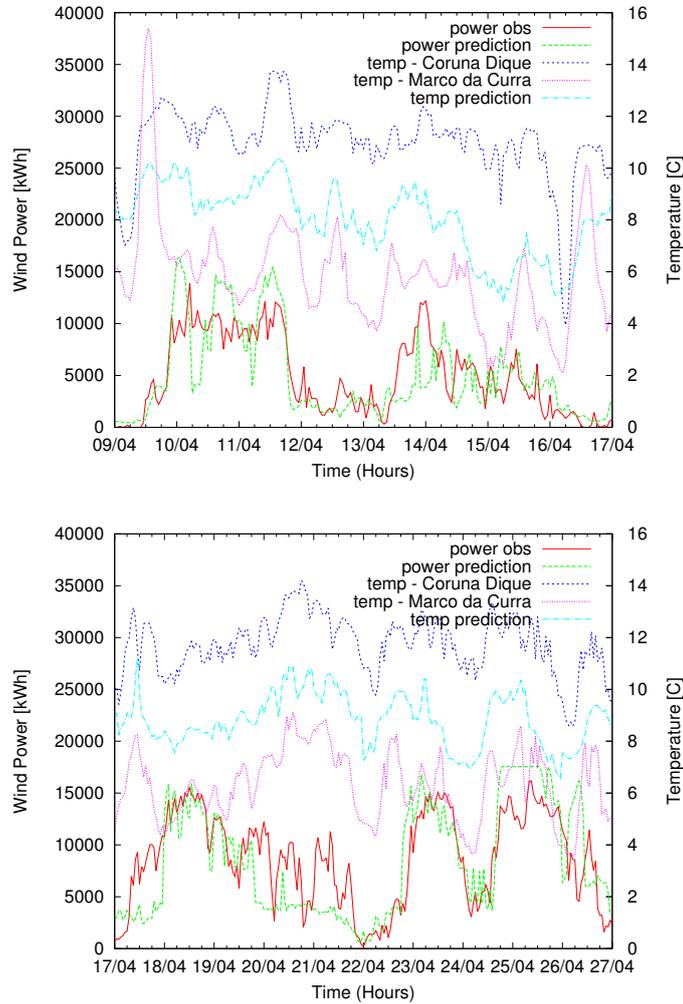
Figure 8.1: Wind power forecasts obtained on the test set using v_1 and v_9 .

There is the possibility of getting access to data from the Spanish market, which has a considerably good penetration of wind energy. With this data, case studies could be designed to quantify the level of resources that can be saved if probabilistic forecasts are provided.

5. The dynamic windowing for ramp characterisation is also an interesting aspect to explore as this mechanism will allow the ramp characterisation approach to adapt to the changes in the energy system. This could be applied not only to day-ahead predictions but also real-time ones. In addition to this, the use of large ramp events training sets could be used for the

automatic generation of fuzzy rules.

8.4   Final Remarks

In this thesis a novel approach to wind power forecasting and ramp character-isation has been presented. The experimental results showed how a Genetic Programming approach for symbolic regression is able to develop models that represent the relationship between the numerical weather predictions and wind power production of a wind park. The advantage of using this approach is that it can be applied in any location as it is not fixed to specific model shapes. Another important advantage of this approach is the fact that the output models can provide some knowledge of what is happening around the location of the park, especially when used in a neighbourhood. In addition to this novel wind power forecasting approach, there is an important contribution towards the character-isation of ramp events through the use of FIS and filtered signals which could provide information on potential ramp events that could be equally damaging as those that exceed a certain percentage of change, according to the binary definition.

References

V. Akbarzadeh, A. Sadeghian, and M. V. dos Santos. Derivation of relational fuzzy classification rules using evolutionary computation. *IEEE International Conference on Fuzzy Systems*, pages 1689–1693, 2008. 157

American Meteorology Society. Meteorology glossary, 2014. URL `glossary.ametsoc.org/wiki/`. 197

N. M. Amil, N. Bredeche, C. Gagn, S. Gelly, M. Schoenauer, and O. Teytaud. A statistical learning perspective of genetic programming. In Leonardo Vanneschi, Steven Gustafson, Alberto Moraglio, Ivanoe Falco, and Marc Ebner, editors, *Genetic Programming*, volume 5481 of *Lecture Notes in Computer Science*, pages 327–338. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-01180-1. doi: 10.1007/978-3-642-01181-8_28. URL `http://dx.doi.org/10.1007/978-3-642-01181-8_28`. 83

A. Arribas, K. B. Robertson, and K. R. Mylne. Test of a poor's man ensemble prediction system for short-range probability forecasting. *Monthly Weather Review*, 133:1825–1839, 2005. 31

Australian Energy Market Operator. Wind integration in electricity grids: International practice and experience. Technical report, 2011. 3

AWS Truewind. AWS Truewind's Final Report for the Alberta Forecasting Pilot Project. Technical report, 2008. 25, 44

B. Bailey, M. C. Brower, and J. C. Zack. Short-term wind forecasting - development and application of a mesoscale model. In *Proceedings of the 1999 European Wind Energy Conference EWEC'99*, pages 1062–1065, Nice, France, March 1999. 30

A. Balouktsis, D. Tsanakas, and G. Vachtsevanos. Stochastic simulation of hourly and daily average wind speed sequences. *Wind Engineering*, 10(1):1–11, 1986. 34

L. Bao, T. Gneiting, E. P. Grimit, P. Guttorp, and A. E. Raftery. Bias correction and bayesian model averaging for ensemble forecasts of surface wind direction. *Monthly Weather Reviews*, 138:1811–1821, 2010. 32

R. Barth, H. Brand, P. Meibom, and C. Weber. A stochastic unit-commitment model for the evaluation of the impacts of integration of large amounts of intermittent wind power. In *International Conference on Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006*, pages 1–8, June 2006. doi: 10.1109/PMAPS.2006.360195. 20

R. Bessa, C. Moreira, B. Silva, and M. Matos. Handling renewable energy variability and uncertainty in power systems operation. *Wiley Interdisciplinary Reviews: Energy and Environment*, 3(2):156–178, 2014. ISSN 2041-840X. doi: 10.1002/wene.76. URL http://dx.doi.org/10.1002/wene.76. 19

H. G. Beyer, T. Degner, J. Hausmann, M. Hoffmann, and P. Ruján. Short term prediction of wind speed and power output of a wind turbine with neural networks. In *Proceedings of the EWEC '94*, pages 349–352, Thessaloniki, 1994. 34

M. Bilgili, B. Sahin, and A. Yasar. Application of artificial neural networks for the wind speed prediction of target station using reference stations data. *Renewable Energy*, 32(14):2350–2360, 2007. 34

W. Bohm and A. Geyer-Schulz. Exact uniform initialization for genetic programming. In R. K. Belew and M. Vose, editors, *Foundations of Genetic Algorithms IV*, pages 379–407. Morgan Kaufmann, University of San Diego, CA, USA, 1996. ISBN 1-55860-460-X. 79

F. Borggrefe and K. Neuhoff. Balancing and intraday market design: Options for wind integration. Technical report, 2011. 18

E. A. Bossanyi. Short-term wind prediction using kalman filters. *Wind Engineering*, 9(1):1–8, 1985. 34

A. Bossavy, R. Girard, and G. Kariniotakis. Forecasting ramps of wind power production with numerical weather prediction ensembles. *Wind Energy*, 16(1): 51–63, 2013. 42, 47, 49, 145, 163, 171

F. Bouffard and F.D. Galiana. Stochastic security for operations planning with significant wind power generation. *Power Systems, IEEE Transactions on*, 23 (2):306–316, May 2008. ISSN 0885-8950. doi: 10.1109/TPWRS.2008.919318. 21

W. Bourke. Performance of the ecmwf and the bom ensemble systems in the southern hemisphere. *Monthly Weather Review*, 132:2338–2357, 2004. 31

L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001. 129

J. B. Bremnes and F. Villanger. Probabilistic forecasts for daily power production. In *Global Wind Power Conference*, Paris (FR), 2002. 26

G. W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950. 46

B. G. Brown, R. W. Katz, and A. H. Murphy. Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology*, 23(8):1184–1195, August 1984. 33

E. Cadenas and W. Rivera. Wind speed forecasting in the south coast of Oaxaca, México. *Renewable Energy*, 32(12):2116–2128, 2007. 34

Edward S. Cassedy and Peter Z. Grossman. *Introduction to Energy: Resources, Technology, and Society*. Cambridge University Press, 1998. 10

F. Cassola and M. Burlando. Wind speed and wind energy forecast through kalman filtering of numerical weather prediction model output. *Applied Energy*, 99(0):154–166, 2012. 38

G. Chandrashekar and F. Sahin. A survey on feature selection methods. *Computers and Electrical Engineering*, 40:16–28, 2014. 90

J. Cochran, L. Bird, J. Heeter, and D. J. Arent. Integrating variable renewable energy in electric power markets: Best practices from international experience. Technical report, 2012. 2, 3

E.M. Constantinescu, V.M. Zavala, M. Rocklin, Sangmin Lee, and M. Anitescu. A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation. *IEEE Transactions on Power Systems*, 26(1):431–441, Feb 2011. ISSN 0885-8950. 21, 67

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. *Reneable and Sustainable Energy Reviews*, 12:1725–1744, 2008. 25, 29

R. Crawford-Marks and L. Spector. Size control via size fair genetic operators in the pushgp genetic programming system. In W. B. Landgon et al., editor, *GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 733–739. Morgan Kaufmann Publishers, New York, July 2002. ISBN 1-55860-878-8. 83

N. J. Cutler, H. R. Outhred, I. F. MacGill, M. J. Kay, and J. D. Kepert. Characterizing future large, rapid changes in aggregated wind power using numerical weather prediction spatial fields. *Wind Energy*, 12(6):542–555, 2009. ISSN 1099-1824. doi: 10.1002/we.312. URL http://dx.doi.org/10.1002/we.312. 20, 49, 89, 147, 174

F. Díaz-González, A. Sumper, O. Gomis-Bellmunt, and R. Villaffila-Robles. A review of energy storage technologies for wind power applications. *Renewable and Sustainable Energy Reviews*, 16(4):2154 – 2171, 2012. ISSN 1364-0321. doi: http://dx.doi.org/10.1016/j.rser.2012.01.029. URL http://www.sciencedirect.com/science/article/pii/S1364032112000305. 2

S. Dierer, T. de Paus, F. Durante, E. Gregow, B. Lange, A. Lavagnini, M. Strack, and B. Tammelin. Predicting wind speed for wind energy; progress of the WINDENG Project. *Wind Engineering*, 29(5):393–408, 2005. 31

R. Doherty and M. O'Malley. A new approach to quantify reserve demand in systems with significant installed wind capacity. *Power Systems, IEEE Transactions on*, 20(2):587–595, May 2005. ISSN 0885-8950. doi: 10.1109/TPWRS.2005.846206. 19

E. Ela and J. Kemper. Wind plant ramping behavior. Technical report, NREL National Renewable Energy Laboratory, December 2009. 13, 44

E. Ela and B. Kirby. Ercot event on february 26, 2008: Lessons learned. Technical report, NREL National Renewable Energy Laboratory, July 2008. 39

S. Enemoto, N. Inomata, T. Yamada, H. Chiba, R. Tanikawa, T. Oota, and H. Fukuda. Prediction of power output from wind farm using local meteorological analysis. In *Proceedings of the European Wind Energy Conference*, pages 749–752, Copenhagen, Denmark, June 2001. ISBN 3-936338-09-4. 31

Energinet. EWIS: Towards a successful integration of large scale wind power into european electricity grids. Technical report, 2010. 3, 18

B. Ernst, B. Oakleaf, M. L. Ahlstrom, M. Lange, C. Möhrlen, B. Lange, U Focken, and K. Rohrig. Predicting the wind. *IEEE power and energy magazine*, 5:78–89, 2007. 23, 32

Eurelectric. Flexible Generation: Backing Up Renewables - part of the Renewables Action Plan (RESAP). Technical report, 2011. xii, 15, 18

C. Ferreira, J. Gama, L. Matias, A. Botterud, and J. Wang. A survey on wind power ramp forecasting. Technical report, ANL Argonne National Laboratory, December 2010. xii, 40, 45, 145, 171

J. J. Flores, M. Graff, and E. Cadenas. Wind prediction using genetic algorithms and gene expression programming. In *Proceedings of the International Conference on Modelling and Simulation in the Enterprises. AMSE 2005*, 2005. 35

N. Francis. Predicting sudden changes in wind power generation. *North American WindPower*, page 58, October 2008. 4, 39

A. P. Fraser and T. Weinbrenner. GPC++ - genetic programming C++ class library, 1997. 85

C. Gallego, A. Costa, A. Cuerva, L. Landberg, B. Greaves, and J. Collins. A wavelet-based approach for large wind power ramp characteristaion. *Wind Energy*, 16(2):257–278, 2013. 50

J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia. A survey on concept drift adaptation. *ACM Computing Surveys*, 1(1), January 2013. 154

G. Giebel, L. Landberg, J. Badger, K. Sattler, H. Feddersen, and T. S. Nielsen et al. Using ensemble forecasting for wind power. In *Proceedings of the European wind energy conference*, Madrid, Spain, 2003. 31

G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl. The state of the art in short-term prediction of wind power. Technical report, ANEMOS.plus and SafeWind, 2011. 13, 25, 28, 35, 50, 121, 126

Global Wind Energy Council. Global wind report 2013. Technical report, 2013. 2

D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison Wesley, Reading, MA, 1989a. 76

I. Gonçalves and S. G. Oliveira da Silva. Experiments on controlling overfitting in genetic programming. In *15th Portuguese Conference on Artificial Intelligence (EPIA 2011)*, pages 978–989, 2011. 84

B. Greaves, J. Collins, J. Farmes, and A. Tindal. Temporal forecast uncertainty for ramp events. *Wind Engineering*, 33(11):309–319, 2009. 40, 46, 48, 147, 168, 171

G. Grell, J. Dudhia, and D. Stauffer. A description of the fifth-generation Pennstate/NCAR mesoscale model (MM5). Technical report, NCAR, 1994. 31, 39

D. J. Hand. Measuring classifier performance: A coherent alternative to the area under the roc curve. *Machine Learning*, 77(1):103–123, 2009. 46

A. Haque, P. Mandal, J. Meng, M. Kaye, and L. Chang. A new strategy for wind speed forecasting using hybrid intelligent models. pages 1–4, 2012. 34

Illinois Institute for Rural Affairs. Wind monitoring program, June 2014. URL http://www.illinoiswind.org. 54, 65

J. Jørgensen, C. Mörlen, B. Ó Gallaghóir, K. Sattler, and E. McKeogh. HIRPOM: Description of an operational numerical wind power prediction model for large scale integration of on- and offshore wind power in denmark, 2002. 30

J. Juban, N. Siebert, and G.N. Kariniotakis. Probabilistic short-term wind power forecasting for the optimal management of wind generation. In *Power Tech, 2007 IEEE Lausanne*, pages 683–688, July 2007. doi: 10.1109/PCT.2007. 4538398. 26

L. Kamal and Y. Z. Jafri. Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. *Solar Energy*, 61(1):23–32, 1997. 34

C. Kamath. Understanding wind ramp events through analysis of historical data. In *Proceedings of the IEEE PES Transmission and Distribution Conference and Expo*, pages 1–6, New Orleans, La., April 2010. 41

M. Kanamitsu, J.C. Alpert, K. A. Campana, P.M. Caplan, D.G. Deaven, M. Iredell, B. Katz, H. L. Pan, J. Sela, and G. H. White. Recent changes implemented into the Global Forecast System at NMC. *Weather and Forecasting*, 6:425–435, 1991. 28, 64

G. Kariniotakis. Safewind: Multi-scale data assimilation, advanced wind modelling and forecasting with emphasis to extreme weather situations for a safe large-scale wind power integration, 2008. URL `http://www.armines.net/data/docs/cms/safewind_123382941118429800.pdf`. 38

G. Kariniotakis, E. Nogaret, and G. Stavrakis. Advanced short-term forecasting of wind power production. In *Proceedings of the European Wind Energy Conference*, pages 751–754, Dublin, Ireland, October 1997. 34

G. Kariniotakis, E. Nogaret, A. G. Dutton, J. A. Halliday, and A. Androutsos. Evaluation of advanced wind power and load forecasting methods for the optimal management of isolated power systems. In *Proceedings of the European Wind Energy Conference*, pages 1082–1085, Nice, France, 1999. ISBN 1-902916-00-X. 34

G. Kariniotakis, I. Martí, D. Casas, P. Pinson, T. S. Nielsen, H. Madsen, G. Giebel, J. Usaola, I. Sanchez, A. M. Palomares, R. Brownsword, J. Tambke, U. Focken, M. Lange, P. Louka, G. Kallos, C. Lac, G. Sideratos, and G. Descombes. What performance can be expected by short-term wind power prediction models depending on site characteristics? In *CD-Proc. of the 2004*

European Wind Energy Conference, EWEC04, London, UK, November 2004. 13, 127

G. Kariniotakis, G. Halliday, J. Brownsword, R. Martí, A. Palomares, and I. Cruz et al. Next generation short-term forecasting of wind power - overview of the ANEMOS project. In *Proceedings of European wind energy conference*, Athens, 2006. 38

R. G. Kavasseri and K. Seetharaman. Day-ahead wind speed forecasting using f-ARIMA models. *Renewable Energy*, 34:1388–1393, 2009. 34

A. Khosravi and S. Nahavandi. Combined nonparametric prediction intervals for wind power generation. *Sustainable Energy, IEEE Transactions on*, 4(4): 849–856, Oct 2013. ISSN 1949-3029. doi: 10.1109/TSTE.2013.2253140. 26

M. E. Kotanchek, E. Y. Vladislavleva, and G. F. Smits. Symbolic regression via gp as a discovery engine: insights on outliers and prototypes. In U. M. O'Reilly R. Riolo and T. McConaghy, editors, *Genetic Programming Theory and Practice VII*, chapter 4. Springer, 2010. 77

A.S. Kowli and S.P. Meyn. Supporting wind generation deployment with demand response. In *Power and Energy Society General Meeting, 2011 IEEE*, pages 1–8, July 2011. doi: 10.1109/PES.2011.6039780. 2

J. R. Koza. *Genetic Programming: on the programming of computers by means of natural selection*. MIT Press, Cambridge, MA, USA., 1992. 75, 76, 78, 83

L. Landberg. Short-term prediction of the power production from wind farms. *Journal of Wind Engineering and Industrial Aerodynamics*, 80:207–220, 1999. 30

L. Landberg. Short-term prediction of local wind conditions. *Journal of Wind Engineering and Industrial Aerodynamics*, 89:235–245, 2001. 30

W. B. Langdon. Size fair and homologous tree genetic programming crossovers. *Genetic Programming and Evolvable Machines*, 1/2:95–119, April 2000. ISSN 1389-2576. 79, 83

W. B. Langdon and R. Poli. Fitness causes bloat. In *Soft Computing in Engineering Design and Manufacturing*, pages 23–27. Springer-Verlag, 1997a. 82

W. B. Langdon and R. Poli. Fitness causes bloat. In P. K. Chawdhry et al., editor, *Soft Computing in Engineering Design and Manufacturing*, pages 13–22. Springer-Verlag, London, June 1997b. ISBN 3-540-76214-0. 82

B. Lange, K. Rohrig, B. Ernst, F. Schlögl, Ü. Cali, R. Jursa, and J. Moradi. Wind power prediction in germany - recent advances and future challenges. In *European Wind Energy Conference and Exhibition*, Athens (GR), 2006. xii, 26, 27

L. Lazić, G. Pejanović, and M. Živković. Wind forecasts for wind power generation using eta model. *Renewable Energy*, 35(6):1236–1243, 2010. 31

M. Lee and J. D. Lee. Optimization of fuzzy rules: Integrated approach for classification problems. *Journal of Security Engineering*, 2(4), 2005. 157

Y. Liu and T. Khoshgoftaar. Reducing overfitting in genetic programming models for software quality classification. In *Proceedings of the Eighth IEEE International Symposium on High Assurance Systems Engineering (HASE04)*, pages 56–65, 2004. 84

P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos. Improvements in wind speed forecasts for wind power prediction purposes using kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(12):2348 – 2362, 2008. ISSN 0167-6105. doi: http://dx. doi.org/10.1016/j.jweia.2008.03.013. URL `http://www.sciencedirect.com/ science/article/pii/S0167610508001074`. 39, 88, 112

H. Madsen, H. A. Nielsen, and T. S. Nielsen. A tool for predicting the wind power production of off-shore wind plants. In *Proceedings of the Copenhagen Offshore Wind Conference and Exhibition, Danish Wind Industry Association*, Copenhagen, Denmark, 2005a. 35, 36

H. Madsen, P. Pinson, G. Kariniotakis, H. A. Nielsen, and T.S. Nielsen. Standarizing the performance evaluation of short-term wind power prediction models. *Wind Engineering*, 29(6):475–489, 2005b. 24

M. Magnusson and L. Wern. Wind energy predictions using CDF and HIRLAM forecast. In *Proceedings of the European Wind Energy Conference EWEC'01*, pages 861–863, Copenhagen, Denmark, June 2001. 30

J. F. Manwell, J. G. McGowan, and A. L. Rogers. *Wind Energy Explained: Theory, Design and Application*. Wiley, 2nd edition edition, 2010. 12

G. Martínez-Arellano and L. Nolle. Short-term wind power forecasting with wrf-arw model and genetic programming. In *Proceedings of the 19th International Conference on Soft Computing MENDEL 2013*, 2013a. ISBN 978-80-214-4755-4. 114

G. Martínez-Arellano and L. Nolle. Genetic programming for wind power forecasting and ramp detection. In *Proceedings of the Thirty-third SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence AI 2013*, pages 403–417. Springer, 2013b. doi: 10.1007/978-3-319-02621-3_30. 114

G. Martínez-Arellano, L. Nolle, and J. A. Bland. Improving WRF-ARW wind speed predictions using genetic programming. In *Proceedings of the Thirty-second SGAI International Conference on Innovative Techniques and Applicatios of Artificial Intelligence AI 2012*, pages 347–360. Springer, 2012. doi: 10.1007/978-1-4471-4739-8_27. 89

G. Martínez-Arellano, R. Cant, and L. Nolle. Prediction of jet engine parameters for control design using genetic programming. IEEE, 2014a. 207

G. Martínez-Arellano, L. Nolle, R. Cant, A. Lotfi, and C. Windmill. Characterisation of large changes in wind power for the day-ahead market using a fuzzy logic approach. *Künstliche Intelligenz*, August 2014b. doi: 10.1007/s13218-014-0322-3. 146

M.A. Matos and R.J. Bessa. Setting the operating reserve using probabilistic wind power forecasts. *Power Systems, IEEE Transactions on*, 26(2):594–603, May 2011. ISSN 0885-8950. doi: 10.1109/TPWRS.2010.2065818. 20

N. F. McPhee and J. D. Miller. Accurate replication in genetic programming. In L. Eshelman, editor, *Genetic Algorithms: Proceedings of the Sixth International Conference.* 82

N. Meinshausen. Quantile regression forests. *Journal of Machine Learning Research*, 7:983–999, 2006. 128

MeteoGalicia. Meteogalicia - Xunta de Galicia Conselleria de Medio Ambiente, Territorio e Infraestructuras, 2014. URL http://www.meteogalicia.es. 176

Metoffice. What causes wind?, 2014. URL http://www.metoffice.gov.uk/learning/wind/what-causes-wind. 11

J. Michalakes and M. Vachharajani. Gpu acceleration of numerical weather prediction. *Parallel Processing Letters*, 18(4):531– 548, Dec 2008. 62

M. Mohandes, T. Halawani, S. Rehman, and A. A. Hussain. Support vector machines for wind speed prediction. *Renewable Energy*, 29(6):939–947, 2004. 34

C. Möhrlen. Uncertainty in wind energy forecasting. phd thesis., 2004. 32, 39

L. Delle Monache, T. Nipen, Y. Liu, G. Roux, and R. Stull. Kalman filter and analog schemes to postprocess numerical weather predictions. *Monthly Weather Review*, 139(11):3554–3570, 2011. 38

C Monteiro, R Bessa, V Miranda, A Botterud, J Wang, G Conzelmann, et al. Wind power forecasting: state-of-the-art 2009. Technical report, Argonne National Laboratory (ANL), 2009. 14, 15, 19, 21, 22, 25, 33, 88, 90

T. R. Naik and V. K. Dabhi. Improving generalization ability of genetic programming: Comparative study. *Journal of Bioinformatics and Intelligent Control*, 2(4):243–252, 2013-12-01T00:00:00. doi: doi:10.1166/jbic.2013.1063. URL http://www.ingentaconnect.com/content/asp/jbic/2013/00000002/00000004/art00001. 81

National Center for Atmospheric Research. Arw version 3 modeling system user's guide, 2014. URL www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/ARWUsersGuideV3.pdf. 197

National Centers for Environmental Prediction. Ncep products inventory, 2013. URL `http://www.nco.ncep.noaa.gov/pmb/products/gfs/`. 65

H. A. Nielsen, H. Madsen, T. S. Nielsen, J. Badger, G. Giebel, L. Landberg, K. Sattler, and H. Feddersen. Wind power ensemble forecasting using wind speed and direction ensembles from ecmwf or ncep. Technical report, Denmark Project PSO, 2005. URL `www.risoe.dtu.dk/rispubl/NEI/nei-dk-4552.pdf`. 32

H. A. Nielsen, T. S. Nielsen, H. Madsen, G. Giebel, J. Badger, and L. Landberg et al. From wind ensembles to probabilistic information about future wind power production - results from an actual application. In *Proceedings of the international conference on probabilistic methods applied to power systems*, Stockholm, Sweden, 2006. 32

T. S. Nielsen, A. Joensen, H. Madsen, L. Landberg, and G. Giebel. A new reference for predicting wind power. *Wind Energy*, 1:29–34, 1998. 34

K. Nozaki, H. Ishibuchi, and H. Tanaka. Adaptive fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, 4(3), 1996. 157

H. Oakley. Two scientific applications of genetic programming: Stack filters and non-linear equation fitting to chaotic data. In Kenneth E., editor, *Advances in Genetic Programming*, pages 369–389. MIT Press, 1994. 76

Center of Global Development. Carbon dioxide emissions from power plants rated worldwind, November 2007. URL `http://www.sciencedaily.com/releases/2007/11/071114163448.htm`. 10

M.A. Ortega-Vazquez and D.S. Kirschen. Estimating the spinning reserve requirements in systems with significant wind power generation penetration. *Power Systems, IEEE Transactions on*, 24(1):114–124, Feb 2009. ISSN 0885-8950. doi: 10.1109/TPWRS.2008.2004745. 20

J. Parkes and A. Tindal. Forecasting short term wind farm production in complex terrain. In *Proceedings of the European Wind Energy Conference and Exhibition EWEC'04*, London, UK, 2004. 37

I. Martí Pérez. Wind forecasting activities. In *Proceedings of the First IEA Joint Action Symposium on Wind Forecasting Techniques*, pages 11–20, Norrköping, Sweden, December 2002. FOI Swedish Defence Research Agency. 37

P. Pinson. Estimation of the uncertainty in wind power forecasting, PhD Thesis, 2006. xii, 20, 23, 26, 48, 127, 128, 132

P. Pinson. Catalogue of complex extreme situations. Technical report, EU Project SafeWind, 2009. 145

P. Pinson and G. Kariniotakis. On-line assessment of prediction risk for wind power production forecasts. *Wind Energy*, 7:119–132, 2004. 23

P. Pinson and H. Madsen. Ensemble-based probabilistic forecasting at horns rev. *Wind Energy*, 12:137–155, 2009. 32

R. Poli, W. B. Langdon, and N. F. McPhee. *A field guide to genetic programming, with contributions by J. R. Koza.* http://lulu.com, 2008. URL `http://www.gp-field-guide.org.uk`. 76, 82

C.W. Potter, E. Grimit, and B. Nijssen. Potential benefits of a dedicated probabilistic rapid ramp event forecast tool. In *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES*, pages 1–5, March 2009. doi: 10.1109/PSCE.2009.4840109. 48

C. Quek and W. L. Tung. A novel approach to the derivation of fuzzy membership functions using the Falcon-MART architecture. *Pattern Recognition Letters*, pages 941–958, 2001. 157

M.L. Ray, A.L. Rogers, and J.G. McGowan. Analysis of wind shear models and trends in different terrains. *University of Massachusetts, Department of Mechanical and Industrial Engineering, Renewable Energy Research Laboratory*, 2006. 97

J.F. Restrepo and F.D. Galiana. Assessing the yearly impact of wind power through a new hybrid deterministic/stochastic unit commitment. *Power Systems, IEEE Transactions on*, 26(1):401–410, Feb 2011. ISSN 0885-8950. doi: 10.1109/TPWRS.2010.2048345. 21

B.D. Ripley. *Pattern recognition and neural networks.* Cambridge University Press, 1996. 34

P.A. Ruiz, C.R. Philbrick, E. Zak, K.W. Cheung, and P.W. Sauer. Uncertainty management in the unit commitment problem. *Power Systems, IEEE Transactions on*, 24(2):642–651, May 2009. ISSN 0885-8950. doi: 10.1109/TPWRS.2008.2012180. 21

S. Salcedo-Sanz, E. G. Ortiz-García, A. Portilla-Figueras, L. Prieto, and D. Paredes. Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction. *Renewable Energy*, 34:1451–1457, 2009. 38, 39, 89, 100, 113

S. Salcedo-Sanz, E. G. Ortiz-Garca, ngel. M. Prez-Bellido, A. Portilla-Figueras, and L. Prieto. Short term wind speed prediction based on evolutionary support vector regression algorithms. *Expert Systems with Applications*, 38(4): 4052 – 4057, 2011. ISSN 0957-4174. doi: http://dx.doi.org/10.1016/j.eswa. 2010.09.067. URL `http://www.sciencedirect.com/science/article/pii/ S0957417410010249`. 39

I. Sánchez, J. Usaola, O. Ravelo, C. Velasco, J. Domínguez, M. G. Lobo, G. González, and F. Soto. SIPREÓLICO - a wind power prediction system based on flexible combination of dynamic models. application to the spanish power system. In *Poster on the World Wind Energy Conference*, Berlin, Germany, June 2002. 36, 37

V. Sanghi, B. K. Lakshmanan, and V. Sundararajan. Survey of advancements in jet-engine thermodynamic simulation. *Journal of Propulsion and Power*, 16: 797–807, 2000. 207

H. H. Schobert. *Energy and Society: an Introduction.* CRC Press, 2002. 1

K. Schreckling. *Gas Turbine Engines for Model Aircraft.* Traplet Publications, 1994. ISBN 0951058916. 206

M. Schwartz and M. Milligan. Statistical wind forecasting at the u.s. national renewable energy laboratory. In *Proceedings of the first IEA Joint Action Sym-*

posium on Wind Forecasting Techniques, pages 115–124, Norrköping, Sweden, December 2002. 34

A. Sfetsos. A novel approach for the forecasting of mean hourly wind speed time series. *Renewable Energy*, 27:163–174, 2001. 34

G. Sideratos and N. Hatziargyriou. An advanced statistical method for wind power forecasting. *IEEE Transactions on power systems*, 22(1):258–265, 2007. 34

W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers. A description of the Advanced Research WRF Version 2. Technical report, 2005. 58

W.C. Skamarock, J.B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers. A description of the advanced research wrf version 2. *AVAILABLE FROM NCAR*, 88:7–25, 2001. 28, 57

J. M. Sloughter, T. Gneiting, and A. E. Raftery. Probabilistic wind speed forecasting using ensembles and bayesian model averaging. *Journal of the American Statistical Association*, 105(489):25–35, 2010. 32

Sotavento. Sotavento Experimental Wind Farm, 2014. URL `http://www.sotaventogalicia.com`. xiii, 56

T. Soule and J. A. Foester. Removal bias: a new cause of code growth in tree based evolutionary programming. In *IEEE International Conference on Evolutionary Computation*, pages 781–786, Anchorage, Alaska, USA, May 1998. IEEE Press. 82

C. Sweeney and P. Lynch. Adaptive post-processing of short-term wind forecasts for energy applications. *Wind Energy*, 2010. doi: doi:10.1002/we.420. URL `http://dx.doi.org/10.1002/we.420`. 38

C. P. Sweeney, P. Lynch, and P. Nolan. Reducing errors of wind speed forecasts by an optimal combination of post-processing methods. *Meteorological Applications*, 2011. 38, 91, 100, 113

T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-15, 1985. 157

C. Tantareanu. Wind prediction in short term: A first step for a better wind turbine control. Technical report, Nordvestjysk Folkecenter for Vedvarende Energi, 1992. 34

J. Tastu, P. Pinson, E. Kotwa, H. Madsen, and H. Aa. Nielsen. Spatio-temporal analysis and modeling of short-term wind power forecast errors. *Wind Energy*, 14(1):43–60, 2011. ISSN 1099-1824. doi: 10.1002/we.401. URL `http://dx.doi.org/10.1002/we.401`. 50

J. W. Taylor, P. E. McSharry, and R. Buizza. Wind power density forecasting using ensemble predictions and time series models. *IEEE Trans Energy Convers*, 24:775–82, 2009. 32

J. L. Torres, A. Garcia, M. de Blas, and A. de Francisco. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Solar Energy*, 79 (1):65–77, 2005. 34

Z. Toth and E. Kalnay. Ensemble forecasting at nmc: The generation of perturbation. *Bulletin of the American Meteorological Society*, 74:2317–2330, 1993. 31

A. Tsakonas. A comparison of classification accuracy of four genetic programming-evolved intelligent structures. *Information Sciences*, 176:691–724, march 2006. 76

A. Tuohy, P. Meibom, E. Denny, and M. O'Malley. Unit commitment for systems with significant wind penetration. *Power Systems, IEEE Transactions on*, 24 (2):592–601, May 2009. ISSN 0885-8950. 20

United Nations. United nations framework convention on climate change - kyoto protocol, December 1997. URL `http://unfccc.int/kyoto_protocol/items/2830.php`. 1

E. Vladislavleva, T. Friedrich, F. Neumann, and M. Wagner. Predicting the energy output of wind farms based on weather data: Important variables and their correlation. *Renewable Energy*, 50:236–243, 2013. 35

S. R. Parker W. Luo, M. C. Taylor. A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. *International Journal of Climatology*, 28(7):947–959, 2008. 68

C. Wan, Z. Xu, and P. Pinson. Direct interval forecasting of wind power. *Power Systems, IEEE Transactions on*, 28(4):4877–4878, Nov 2013. ISSN 0885-8950. doi: 10.1109/TPWRS.2013.2258824. 26

J. Wang, M. Shahidehpour, and Zuyi Li. Security-constrained unit commitment with volatile wind power generation. *Power Systems, IEEE Transactions on*, 23(3):1319–1327, Aug 2008. 20

J. Wang, A. Botterud, R. Bessa, H. Keko, L. Carvalho, D. Issicaba, J. Sumaili, and V. Miranda. Wind power forecasting uncertainty and unit commitment. *Applied Energy*, 88(11):4014 – 4023, 2011. ISSN 0306-2619. doi: http://dx.doi. org/10.1016/j.apenergy.2011.04.011. URL `http://www.sciencedirect.com/ science/article/pii/S0306261911002339`. 21, 23

L. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*, 22(6), 1992. 157

WEPROG. Weprog: Aeso wind power forecasting pilot project pre-conference session, April 2007. 26

World Wind Energy Association. World wind energy half-year report 2012. Technical report, WWEA, 2012. 1, 10

A. Yamaguchi, T. Ishihara, K. Sakai, T. Ogawa, and Y. Fujino. A physical-statistical approach for the regional wind power forecasting. In *Proceedings of the European Wind Energy Conference EWEC*, Milan (IT), May 2007. 31

J. W. Zack. Optimization of wind power production forecast performance during critical periods for grid management. In *WINDPOWER 2007*, Los Angeles, CA, June 2007. 42, 44

L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965. 152

J. Zeng and W. Qiao. Short-term wind power prediction using a wavelet support vector machine. *IEEE Transactions on Sustainable Energy*, 3(2):255–264, 2012. 34

B. T. Zhang and H. Mühlenbein. Balancing accuracy and parsimony in genetic programming. *Evolutionary Computation*, pages 17–38, 1995. 83

Y. Zhang, J. Wang, and X. Wang. Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, 32(0): 255 – 270, 2014. ISSN 1364-0321. doi: http://dx.doi.org/10.1016/j.rser. 2014.01.033. URL `http://www.sciencedirect.com/science/article/pii/ S1364032114000446`. 26

P. Zhao, J. Wang, J. Xia, Y. Dai, Y. Sheng, and J. Yue. Performance evaluation and accuracy enhancement of a day-ahead wind power forecasting system in China. *Renewable Energy*, 43(0), 2012. 38

H. Zheng and A. Kusiak. Prediction of wind farm power ramp rates: A data-mining approach. *Solar Energy Engineering*, 131(3), 2009. doi: 031011-1-031011-8. 41, 47

Appendix A: Glossary of Meteorological Terms

Glossary of meteorological terms (American Meteorology Society, 2014; National Center for Atmospheric Research, 2014)

1. Albedo: fraction of solar energy reflected form the Earth back into space.

2. Annual mean deep soil temperature: the annual mean of the temperature of the soil that is at 1m depth from the surface.

3. Cartesian coordinates: a coordinate system in which the locations of points in space are expressed by reference to three planes, called coordinate planes, no two of which are parallel.

4. Eddy: any circulation drawing its energy from a flow of much larger scale, and brought about by pressure irregularities, as in the lee of a solid obstacle.

5. Geopotential: the potential energy of a unit mass relative to sea level, numerically equal to the work that would be done in lifting the unit mass from sea level tit he heigh at which the mass is located.

6. Land use categories: land use is characterised by the arrangements, activities and inputs people undertake in a certain land cover type to produce, change or maintain it. The geogrid.exe module interpolates land use categories from USGS 24-category data. Those categories are shown in Table 1.

7. Mixing ratio: the ratio of the mass of a variable atmospheric constituent to the mass of dry air. The term normally refers to water vapor.

Table 1: Land use categories

Land use category	Land use description
1	Urban and Built-up Land
2	Dryland Cropland and Pasture
3	Irrigated Cropland and Pasture
4	Mixed Dryland/Irrigated Cropland and Pasture
5	Cropland/Grassland Mosaic
6	Cropland/Woodland Mosaic
7	Grassland
8	Shrubland
9	Mixed Shrubland/Grassland
10	Savanna
11	Deciduous Broadleaf Forest
12	Deciduous Needleleaf Forest
13	Evergreen Broadleaf
14	Evergreen Needleleaf
15	Mixed Forest
16	Water Bodies
17	Herbaceous Wetland
18	Wooden Wetland
19	Barren or Sparsely Vegetated
20	Herbaceous Tundra
21	Wooded Tundra
22	Mixed Tundra
23	Bare Ground Tundra
24	Snow or Ice

8. Orography: branch of physical geography that studies the formation and features of mountains. The nature of a region with respect to its elevated terrain.

9. Roughness: the geometric characteristic of a surface with its efficiency as a momentum sink for turbulent flow, due to the generation of drag forces and increased vertical wind shear.

10. Soil categories: different types of soil can create differential heating of the earth surface, having different effects on the numerical model prediction.

To handle these, soil types are classified into 16 different types shown in Table 2.

Table 2: Soil categories

Soil Bategory	Soil Description
1	Sand
2	Loamy Sand
3	Sandy Loam
4	Silt Loam
5	Silt
6	Loam
7	Sandy Clay Loam
8	Silty Clay Loam
9	Clay Loam
10	Sandy Clay
11	Silty Clay
12	Clay
13	Organic Material
14	Water
15	Bedrock
16	Other (land-ice)

11. Snow albedo: fraction of solar energy reflected by ice or snow, which may vary according to the characteristics of the snow (freshly fallen, melting snow or dirty snow).

12. Terrain height: this is calculated by taking the difference between the surface pressure and the mean sea-level pressure and multiplying the result by 10m per hPa, which is the approximate relation of height to pressure and is a function of the air density.

13. Terrain slope: slope angle, or incline, is the measure of steepness or the degree of inclination of the terrain relative to the horizontal plane. It is one of the most significant terrain characteristics which affects weather and climate patterns.

14. Turbulent kinetic energy: the mean kinetic energy per unit mass associated with eddies in turbulent flow.

15. Vegetation fraction: the percentage of occupation of vegetation canopy in a given ground area in vertical projection. It is popularly treated as a comprehensive quantitative index to monitor respective land cover conditions.

16. Vertical velocity: in meteorology, the component of the velocity vector along the local vertical.

Appendix B: NCL Script to Extract Forecasts

NCL scripts

This section presents the NCL script used to extract data from the WRF-ARW output files. The extraction from GFS files was done in a similar way.

```
; ——————————— LOAD FUNCTIONS AND PROCEDURES ———————————

load "/usr/local/lib/ncarg/nclscripts/csm/gsn_code.ncl"
load "/usr/local/lib/ncarg/nclscripts/csm/gsn_csm.ncl"
load "/usr/local/lib/ncarg/nclscripts/csm/contributed.ncl"
load "/home/giovanna/Documents/WRF/WRFV3/WRFUserARW.ncl"


; ——————————— BEGINING OF NCL SCRIPT ———————————

begin
;********************************************************
; read in netCDF file and make a loop for all time steps
;********************************************************
    myFiles = systemfunc("ls /home/giovanna/Documents/WRF/DOMAINS/galicia/
wrfprd/wrfout_d02*")

    nfiles=dimsizes(myFiles)

  do numFiles=0,nfiles-1
   in      = addfile(myFiles(numFiles)+".nc","r")
   print(numFiles)

; get times in the file
```

```
times = wrf_user_list_times(in)
ntimes = dimsizes(times)           ; number of times in the file
wind_speed1 = new(ntimes, float)
wind_speed2 = new(ntimes, float)
wind_speed3 = new(ntimes, float)
wind_speed4 = new(ntimes, float)

temp45_1 = new(ntimes, float)
temp45_2 = new(ntimes, float)
temp45_3 = new(ntimes, float)
temp45_4 = new(ntimes, float)

wind_direction = new(ntimes, float)
wind_direction2 = new(ntimes, float)
wind_direction3 = new(ntimes, float)
wind_direction4 = new(ntimes, float)

lat = 43.351630
lon = -7.880116

do it = 0, ntimes-1              ;Loop for the time: it= starting time
time = it

;************************************************
; - Select lon & lat of the point of interest -
;************************************************

res = True
res@returnInt = True        ; False : return real values, True: return interger
point = wrf_user_ll_to_ij(in, lon, lat, res)
;printVarSummary(point)
x = point(0)
y = point(1)

; checking neighbors of x and y to get other grid points

x2 = x
y2 = y+1

x3 = x-1
```

201

```
y3 = y+1

x4 = x−1
y4 = y

;************************************************
;   − extract wind components −
;************************************************

  u = wrf_user_getvar(in, "ua", time)
  v = wrf_user_getvar(in, "va", time)
  height = wrf_user_getvar(in, "z", time)
  ter = wrf_user_getvar(in,"ter",time)
  tc = wrf_user_getvar(in, "tc", time)
  p = wrf_user_getvar(in "pressure",time)

  nheight = conform(height,ter,(/1,2/))
  height = height − nheight

  ;Interpolate U,V to 45 Meters
  u_plane = wrf_user_intrp3d(u,height,"h",45,0.,False)
  v_plane = wrf_user_intrp3d(v,height,"h",45,0., False)



  ; Calculate wind speed from Vectors
  spd = (u_plane*u_plane + v_plane*v_plane)^(0.5)

  wind_speed1(it) = spd(y − 1,x − 1)
  wind_speed2(it) = spd(y2 − 1,x2 − 1)
  wind_speed3(it) = spd(y3 − 1,x3 − 1)
  wind_speed4(it) = spd(y4 − 1,x4 − 1)

  ; Calculate wind direction from Vectors at first point
  r2d = 45.0/atan(1.0)   ; conversion factor (radians to degrees)
  dir = atan2(u_plane,v_plane) * r2d + 180

  wind_direction(it) = dir(y−1,x−1)
  wind_direction2(it) = dir(y2−1,x2−1)
  wind_direction3(it) = dir(y3−1,x3−1)
```

```
   wind_direction4(it) = dir(y4-1,x4-1)

  ; get Temperature

   tc_plane = wrf_user_intrp3d(tc,height,"h",45,0.,False)
   temp45_1(it) = tc_plane(y-1,x-1)
   temp45_2(it) = tc_plane(y2-1,x2-1)
   temp45_3(it) = tc_plane(y3-1,x3-1)
   temp45_4(it) = tc_plane(y4-1,x4-1)



end do                          ; end of time loop

;************************************************
;  - Write wind speed in ascii file -
;************************************************
  fName = "galiciaHorizon36Neigh" + numFiles + ".txt";
  data  = new( 24, "string")

  do it=0,ntimes-1

      if(it .gt. 17 .and. it .lt. 42) then
      data(it-18) = times(it)+ "␣"
      data(it-18) = data(it-18) + sprintf("%7.4f", wind_speed1(it)) + "␣"
      data(it-18) = data(it-18) + sprintf("%7.4f", wind_speed2(it)) + "␣"
      data(it-18) = data(it-18) + sprintf("%7.4f", wind_speed3(it)) + "␣"
      data(it-18) = data(it-18) + sprintf("%7.4f", wind_speed4(it)) + "␣"

      data(it-18) = data(it-18) + sprintf("%7.4f", wind_direction(it)) + "␣"
      data(it-18) = data(it-18) + sprintf("%7.4f", wind_direction2(it)) + "␣"
      data(it-18) = data(it-18) + sprintf("%7.4f", wind_direction3(it)) + "␣"
      data(it-18) = data(it-18) + sprintf("%7.4f", wind_direction4(it)) + "␣"


      data(it-18) = data(it-18) + sprintf("%f",temp45_1(it)) + "␣"
      data(it-18) = data(it-18) + sprintf("%f",temp45_2(it)) + "␣"
      data(it-18) = data(it-18) + sprintf("%f",temp45_3(it)) + "␣"
      data(it-18) = data(it-18) + sprintf("%f",temp45_4(it)) + "␣"
```

```
        end if

    end do

        asciiwrite (fName , data)

  end do

end
```

Appendix C: a Symbolic Regression Application

There is an important amount of research being carried out for the simulation of jet engines as an alternative to testing when developing control systems. Mathematical models needed for simulations are not always available as new engines become more complex. Considering that the GP approach provides some advantages over other machine-learning techniques, like the mathematical representation of solutions and unrestricted size solutions, this was an interesting way to validate the algorithm as a regression technique.

The Problem

In order to provide a smooth, stable and stall free operation of a jet engine, control systems were developed. These control systems should ensure a minimum required level of performance, which is defined as the thrust achieved at a given throttle setting. Engine constraints provide additional challenges for control design as the engine needs to operate safely; this means without rotor over-speed, compressor stall, combustor blowout or turbine over-temperature. Small scale jet engines, which operate on the same principles as the commercial jet engines, were initially developed by amateurs for use in model aircraft (Schreckling, 1994) but more recently commercially produced engines have been used for research and education purposes. Developing robust control systems for small jet engines is not practical without a simulation, especially if soft computing techniques are used, because of the need to explore the extreme of safety related parameters such as the Exhaust Gas Temperature (EGT).

As the level of sophistication of modern jet engines increased, the complexity of the mathematical models required for an accurate simulation grew. This led

Figure 2: Behotec j66 engine.

to the application of many knowledge-based systems such as Fuzzy Logic (FL), neural networks (NN), genetic algorithms (GA) and probabilistic reasoning (PR) for the generation of complex models. Further detail of control development and simulation techniques can be found on (Martínez-Arellano et al., 2014a; Sanghi et al., 2000).

For an amateur designer, developing a good controller without complete knowledge of the mathematical model of the engine behaviour would be a challenge. However, following a design/simulation/data acquisition iterative approach, could facilitate the process. The simulation in the iterative process could be achieved by applying the GP approach to find the mathematical models at each iteration with the data gathered in the previous iteration.

Using Genetic Programming for the Prediction of Jet Engine Parameters

As the thrust of an engine cannot be measured, in order to design a control system, the fuel flow or any other correlated variable could be used to control the shaft speed. For this reason, parameters such as the pump voltage, rpm, temperature, and pressure would be relevant during the simulation. The GP algorithm was used to model the relationship between three of these parameters using data available from a Behotec j66 small scale engine presented in Figure 2.

The approach used to test the algorithm is as follows: first, with a simple fuel flow controller, pump voltage, pressure, temperature among other data was

gathered from the engine. The GP approach was then used to generate a case pressure model using pump voltage and pressure observations. A second model was generated applying the GP approach to model the rotor speed from the case pressure, pump voltage and rotor speed observations. In a following stage, a third model was developed using the pump voltage, the pressure and the rotor speed to predict future values of EGT. The models found predict each parameter only one step ahead, so for simulation purposes, each value predicted by the model would have to be fed back into the model to obtain the next prediction.

Experimental Setup and Results

The starting point for the experiments was a set of data that had been logged during previous runs of the Behotec j66 engine. The information available consisted of pump voltage, case pressure, rpm and EGT logged at intervals of 0.2 seconds. The experimental setup was divided in three stages. The first stage involved the use of pump voltage observations in order to predict the case pressure. In a first attempt, four inputs, which corresponded to the last four observations of the pump voltage, were used. Preliminary results suggested that, as the variable presents very slight changes through time, a larger history would need to be considered. Instead of increasing the number of inputs to consider a larger history, the same four inputs were kept and the values of these inputs were averages of historical data. The first one was the actual observation at time t. The second input was the average of the four last observations of the pump voltage ($t = t, t-1, t-2, t-3$). The third input was the average of the last 8 observations ($t = t, t-1, ..., t-7$) and the fourth input was the average of the last 16 inputs. In this way, a general overview of the changes in the variable from the last 16 time steps could be considered without increasing the number of inputs of the algorithm to 16. The expected output would be the pressure at time t. Previous values of the pressure were not used to predict pressure at time t. The reason for this was that preliminary experiments showed that very small differences of the variable in different time steps could lead the search of the algorithm to pick up faster these previous values of the pressure rather than other variables because the algorithm is trying to decrease the RMSE between the forecast and the observation. As this was not the model output that was expected, the input was removed to avoid leading the search towards these values.

The second stage used the pressure and pump voltage in order to predict the rotor speed. For the second set of experiments, the number of inputs increased to 8 in order to consider the four values for both the pump voltage and the pressure. As it was done with the previous experiments, the rpm parameter was not fed into the algorithm to avoid picking up this variable over the other ones.

Finally, for the third set of experiments two settings were used. In a first attempt, pump voltage, pressure and rpm were used. For each variable, the four inputs as used in previous experiments were used, having a total of 12 inputs. In a second setting, history of the exhaust gas temperature was included. Due to the low correlation between the temperature and these three variables, it was also necessary to include historical data from the temperature. Three averages of the temperature were considered, an average on the previous 32 readings, an average on the last 16, and an average on the last 8. A longer history period was used for the temperature as this variable is changing at a lower pace compared to the other three variables. The total number of inputs to the algorithm grew to 15. Genetic programming is very sensitive to the input data. Feeding all variables to all experiments is not always the best approach. So it must be decided intuitively which variables are most important and the input must be limited to these ones only.

For the three sets of experiments the training and testing sets were designed as follows. Several runs of the engine were used. Those variable readings obtained by the sensors where the engine and/or pump were not running were excluded from the recorded information because they were unrepresentative. The remaining data was divided into training and testing sets. Both training and testing files contained data from startup and operation stages. From the total amount of records available 60% were used for training and 40% for testing. The training set was further randomly divided into 80% training and 20% validation.

Once the training, validation and test sets were obtained, the experiments were carried out as follows. First, for each type of experiment the best values of k were identified empirically. Each modelled variable has its own complexity so different values of k were applied. For the case pressure variable, a parsimony pressure of $k = 0.005$ was enough to avoid very complex trees. The raw fitness (fitness with no penalisation) of the models obtained for the case pressure is 0.045

Table 3: Fixed GP parameters used for the experiments.

Runs	50
Population	1000
Generations	100
Crossover operator	Standard subtree crossover, probability 1.0
Mutation operator	Standard subtree mutation, probability 0.03, maximum depth of new tree 17
Tree initialisation	Ramped Half-and-Half, maximum depth 6
Function set	+, -, *, / log, exp
Terminal set	pump_v, press, rpm and random constants
Selection	Tournament of size 20
Elitism	Best individual always survives

on average, so the parsimony pressure is such that does not provide an unwanted advantage over raw fitness. Larger models will be allowed to survive only if they provide a significant improvement over the raw fitness even when adding the penalisation. To model the rotor speed, the parsimony pressure used was around $k = 80$. For the EGT, values between $k = 0.2$ and $k = 0.3$ were used.

Once these parsimony pressures were obtained they were used to execute 50 runs of each type of experiment. At each run, the training and validation sets were selected randomly. The additional parameters that were set for the experiments are shown in Table 3 which are typical settings for the GP approach. The best models found for each variable are shown in Table 4. The subindexes of the variables are used to denote the type of input. A subindex of 1 means the current value of the variable at time t, 2 denotes the average of the last 4 observations, 3 corresponds to the last 8 observations and 4 to the last 16 observations. The models were applied to the test sets to evaluate the quality of the predictions.

Figure 3 presents the predictions of the case pressure obtained on the testing set with the best model found on the first set of experiments. The correlation between the observations and the predictions is very high, so the algorithm easily detected the relationship among them. Figure 4 presents the predictions of the rpm parameter using the best model found on the second set of experiments. It
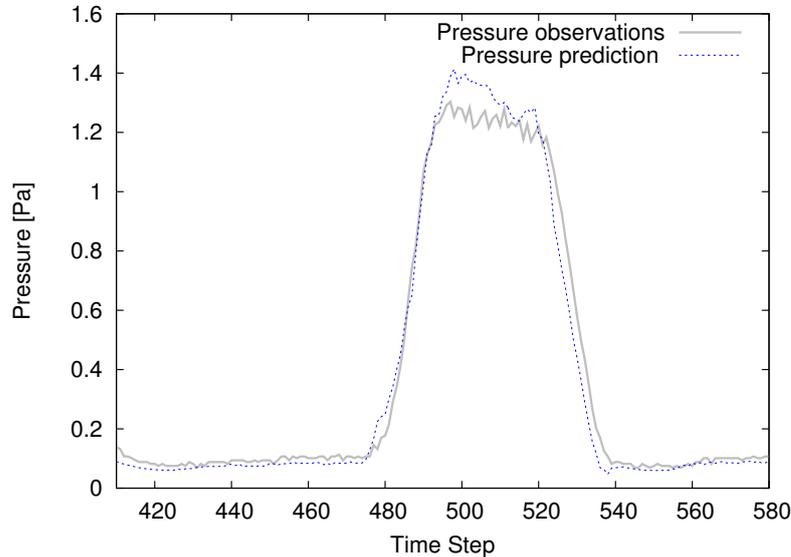
Figure 3: Pressure prediction one time step ahead with the best model found using pump voltage on test data.

can be seen that this relationship was also quite well caught by the algorithm producing an accurate model.

Figures 5 and 6 present the predictions of the exhaust gas temperature using the best model from the third set of experiments. The trend of the temperature is well caught in both settings of the experiment. However, the increment in temperature at around time step 480 was better achieved by the model that was trained with some information from previous data of the temperature included. This was done because it was observed that the temperature parameter is dependent on its own history in a way that could not easily be captured by working only from the pump voltage, pressure and RPM. The basic data for pump voltage and case pressure were somewhat noisy. This is of no consequence for the training process but the noise is inevitably transmitted through the model. Therefore noise of the prediction has been reduced in both cases by using the pressure readings that had been pre-filtered during the data-logging process and by using the demanded voltage from the pump controller rather than the actual measured pump voltage.

Overall the results are good for predicted pressure and RPM. The predicted EGT is less accurate but still models the actual engine behaviour in a manner
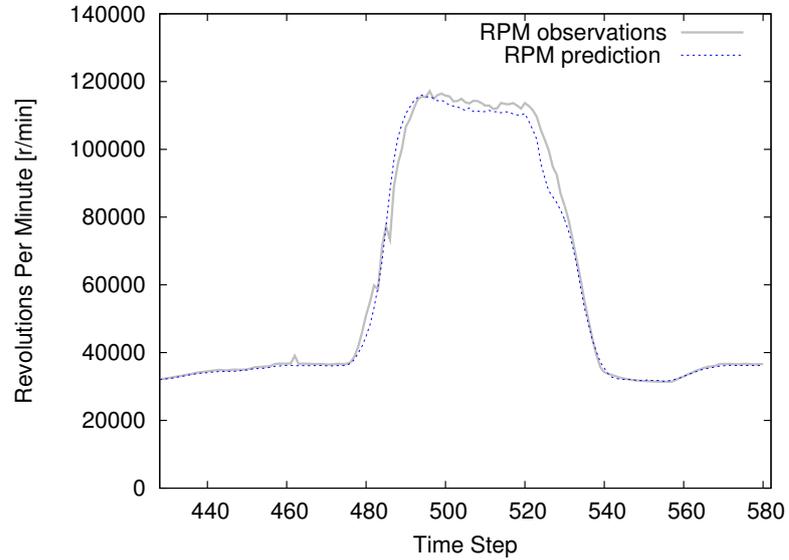
Figure 4: RPM prediction using pump voltage and pressure on test data.

which is qualitatively correct. When the EGT history was not included as an input as in Figure 5, the transient peak in EGT during acceleration was not strong enough. The model that included EGT history presented in Figure 6, shows a much stronger peak but with some time lag. This suggests that further experimentation with the way in which EGT history is presented as an input may be required. Another approach would be to increase the prominence of the transient events within the training data so that these inaccuracies are penalised more strongly during the selection process.
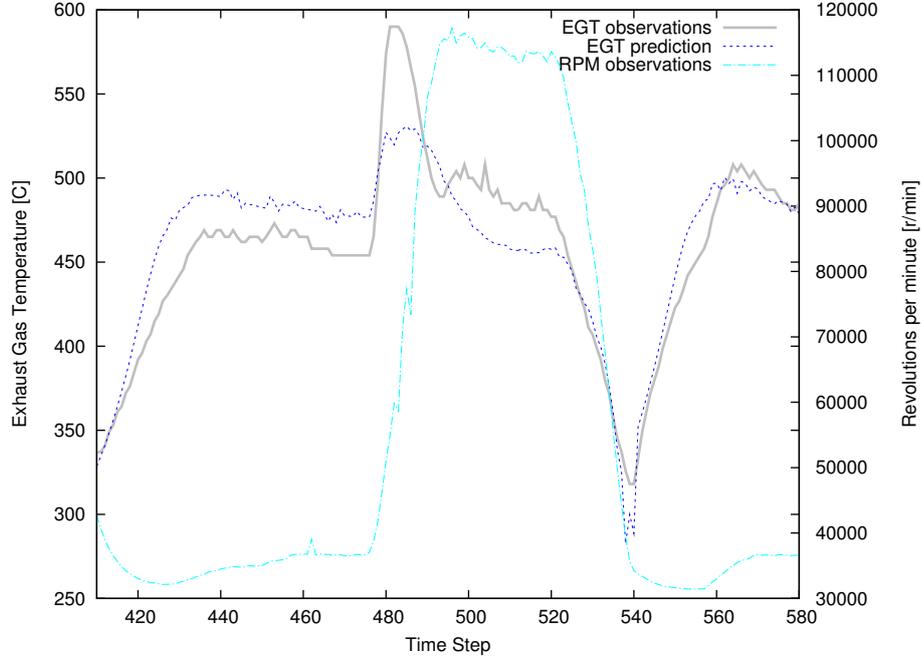
Figure 5: EGT prediction using pump voltage, pressure and rpm on test data. RPM observations in the same time period are also shown.

Table 4: Models found with the Basic GP implementation.

Predicted Variable	Mathematical Representation of the Model
Pressure	$pumpV_1/18.6264 + (pumpV_4 + ln(pumpV_2) + pumpV_2)$ $*pumpV_1/18.6264$
RPM	$exp(2press_1 + 11.61046 + press_3/exp(press_3^2(press_3*(press_1 + 5.80523) - press_3/(press_3/(press_1 + 5.80523) + press_3/ (press_3*exp(pumpV_2) - press_1))))) * 92.3692 + press_3/ exp(pumpV_2)) * 92.3692$
EGT	$6(pumpV_3/press_1) + 3(pumpV_3 * press_1)$ $+temp_8 - pumpV_4 - ln(temp_{16})/pumpV_3 - 6ln(temp_{16})$ $+pumpV_3 + pumpV_3/press_1 + pumpV_3 * press_1$
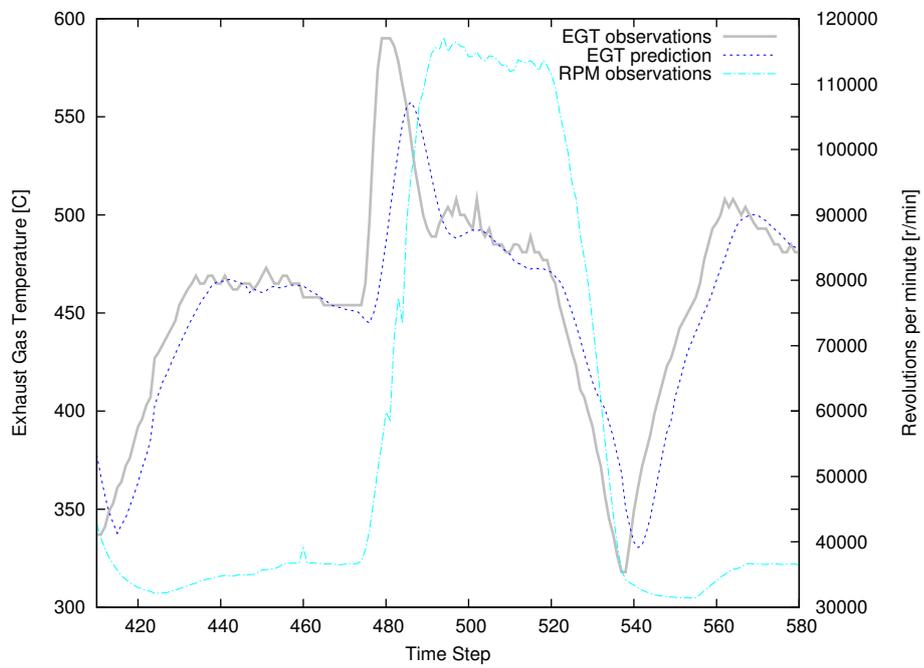
Figure 6: EGT prediction using pump voltage, pressure, rpm and temperature. RPM observations in the same time period are also shown.

Appendix D: Wind Speed Downscaling Experimental Results

Genetic Programming for wind speed prediction - Experimental Results

In the following are gathered all the results obtained using genetic programming for wind speed downscaling at Cuba, SIUE and Wilmington sites in Illinois, USA at 10 meters height. The experiments consisted in varying the parsimony pressure and the size of the training set as a way to avoid overfitting. The figures present the variation of the MAE when varying the size of the training set and the pressure parameter. The results of three different pressures are shown in Figures 7, 8 and 9.

Figures 10 to 13 present the results obtained using logarithmic and exponential functions in the function set.

Finally, Figure 14 shows the average improvement obtained per hour when retraining the algorithm.
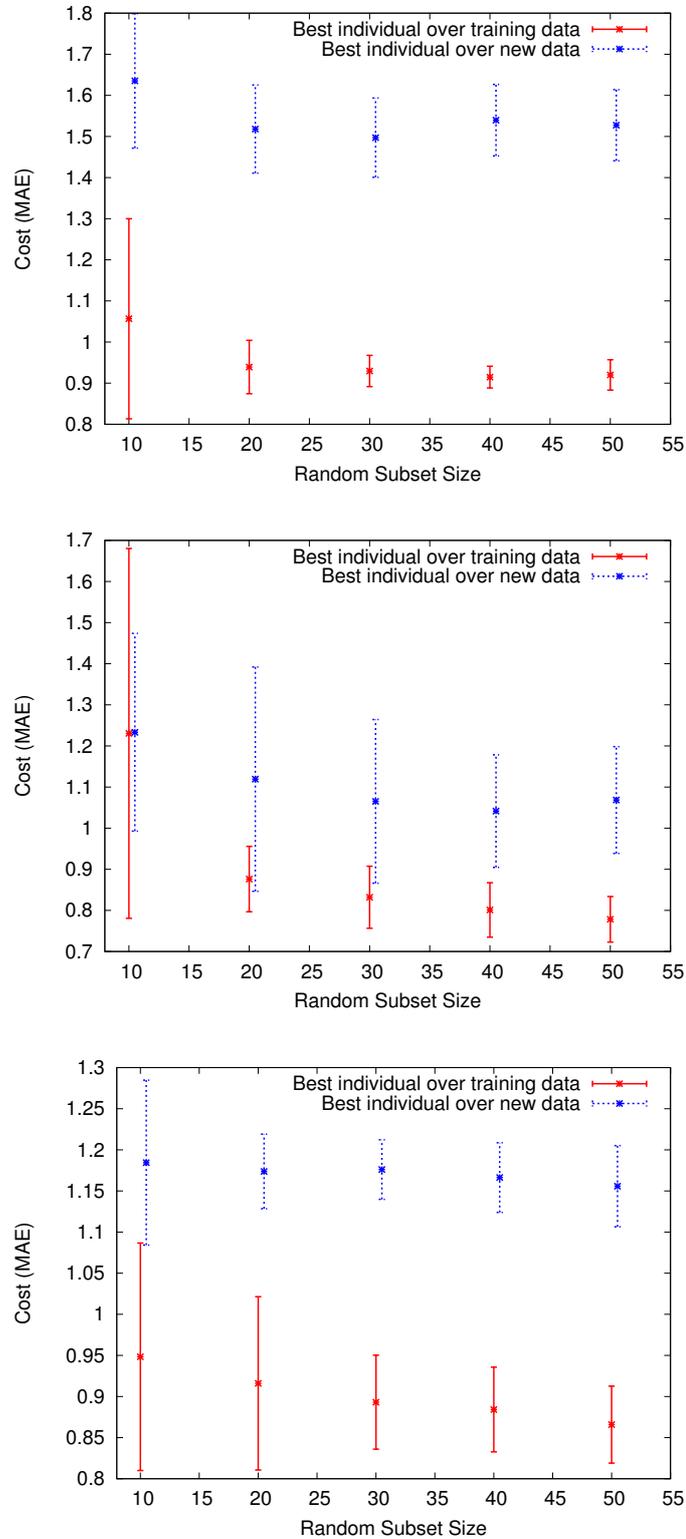
Figure 7: Wind Speed forecasts at the three observations sites using a parsimony pressure of $k = 0.1$. Top image corresponds to Cuba, middle image to SIUE and bottom image to Wilmington.
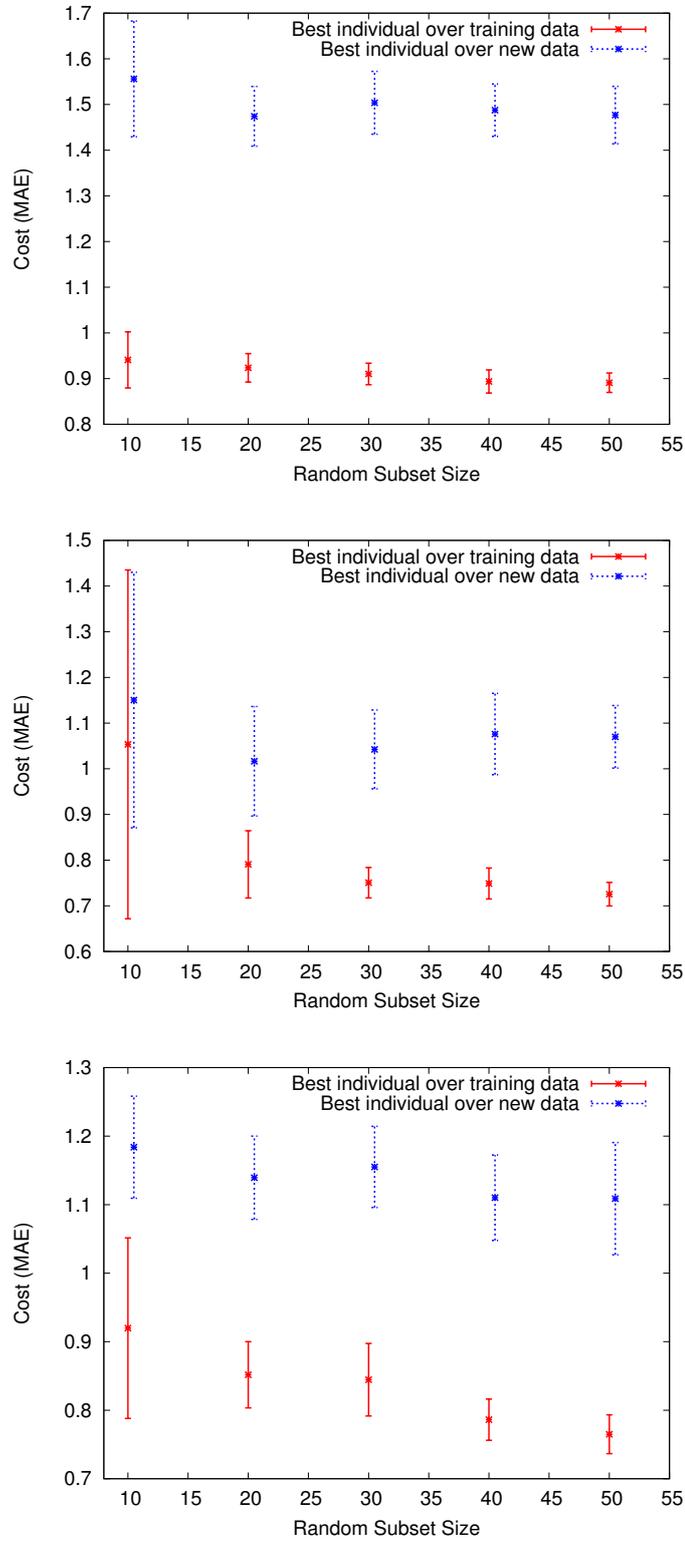
Figure 8: Wind Speed forecasts at the three observations sites using a parsimony pressure of $k = 0.01$. Top image corresponds to Cuba, middle image to SIUE and bottom image to Wilmington.
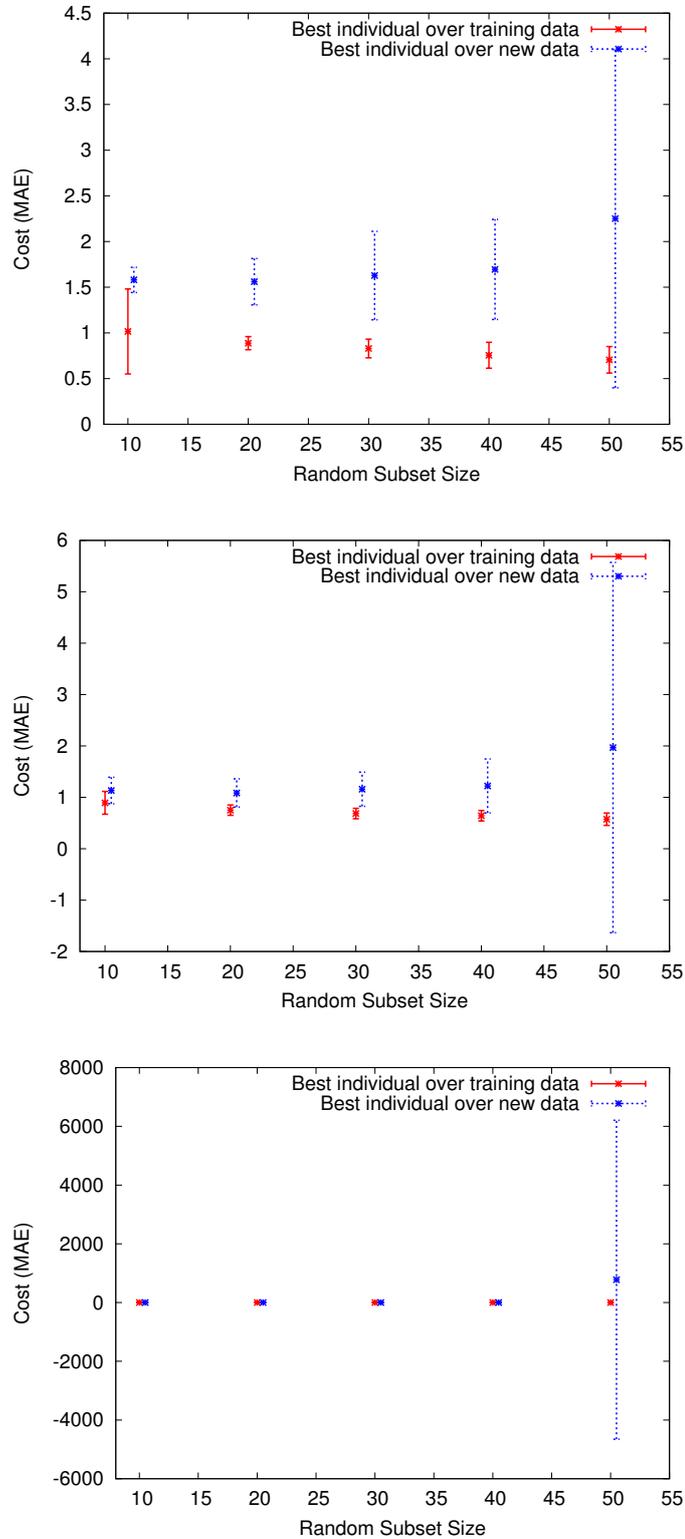
Figure 9: Wind Speed forecasts at the three observations sites using a parsimony pressure of $k = 0.001$. Top image corresponds to Cuba, middle image to SIUE and bottom image to Wilmington.
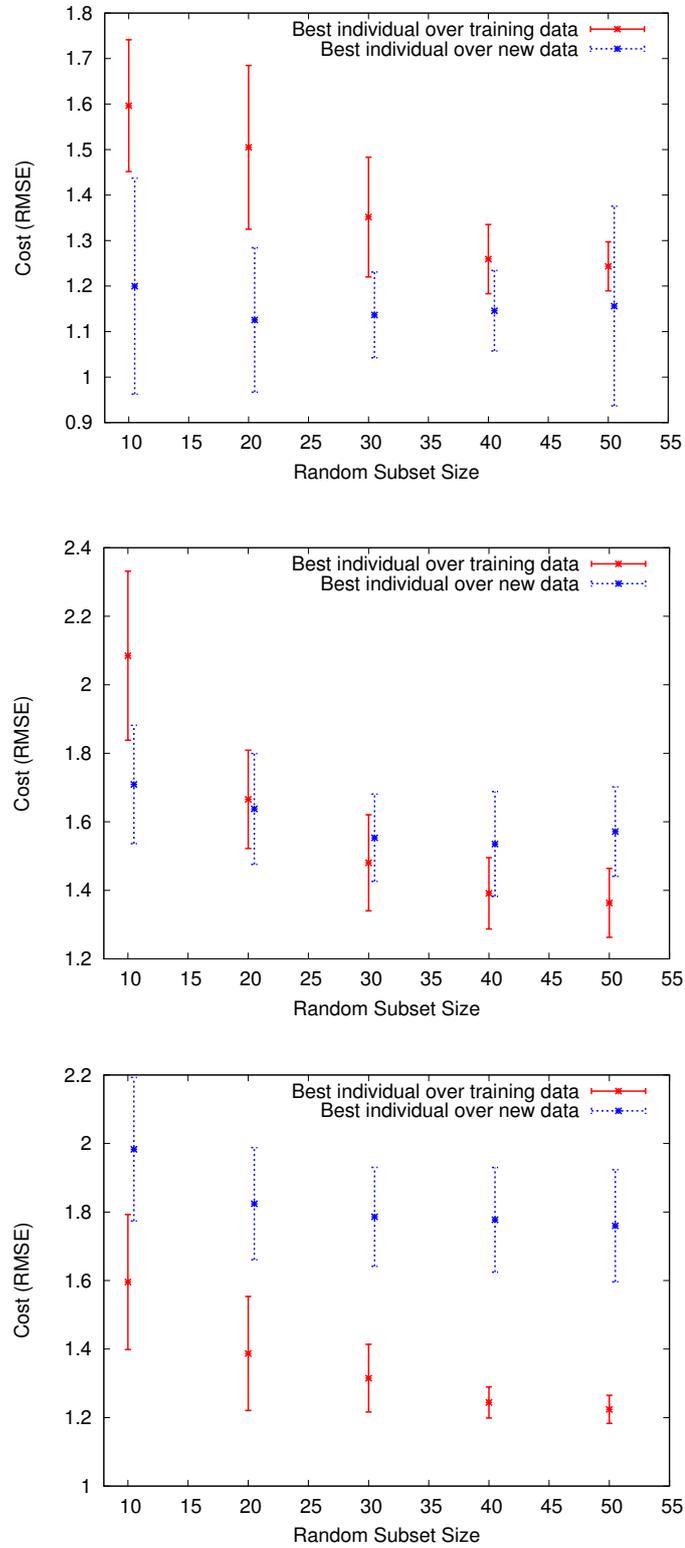
Figure 10: Wind Speed forecasts at the three observations sites using a parsimony pressure of $k = 0.01$.
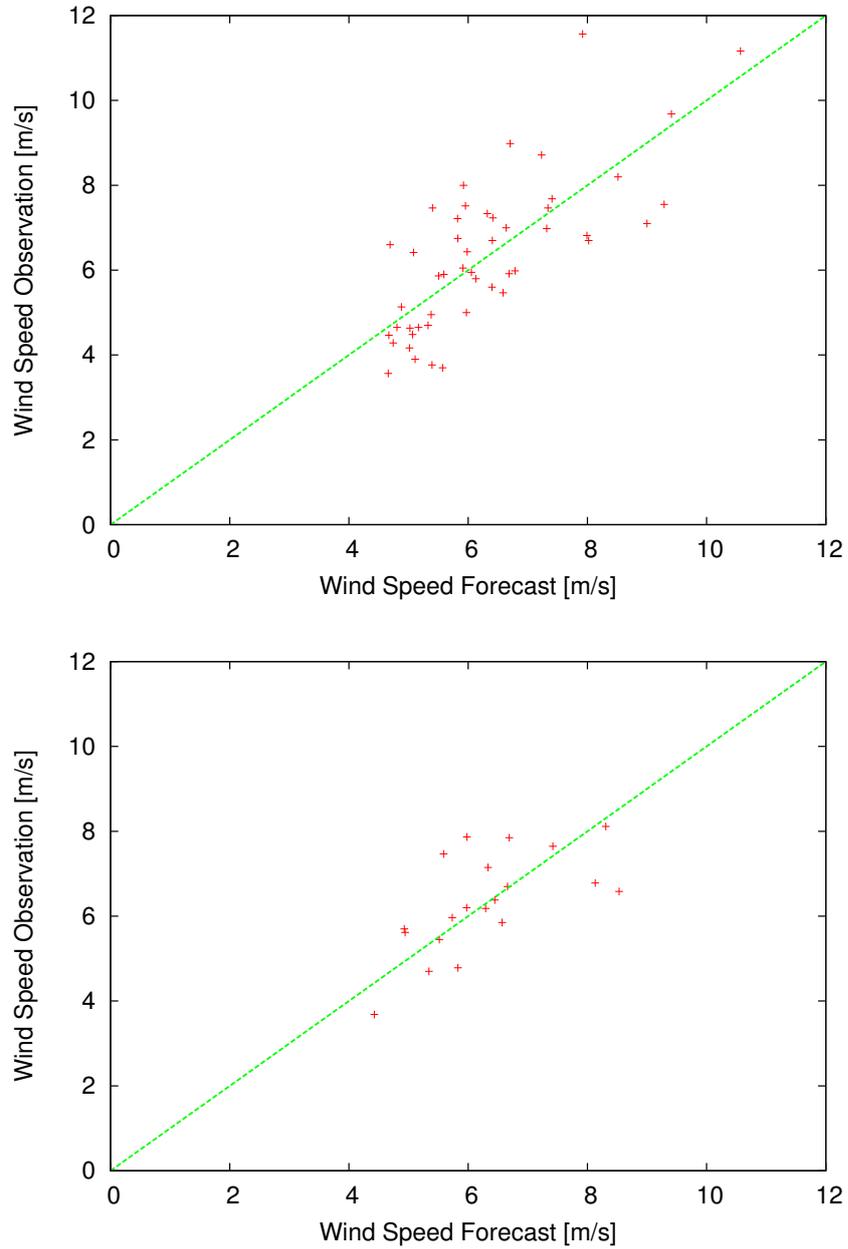
Figure 11: Correlation between the best GP model and wind speed observations at Cuba on training (left) and testing (right) sets.
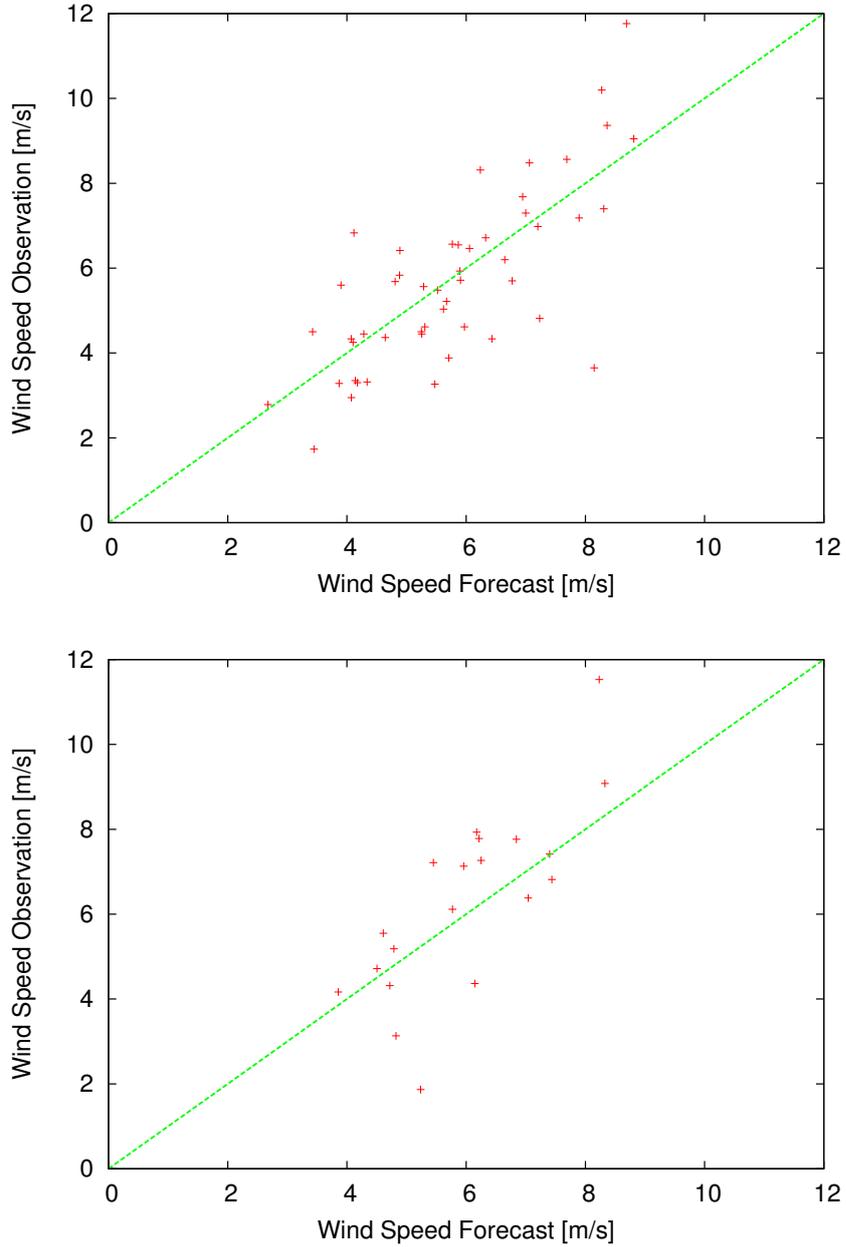
Figure 12: Correlation between the best GP model and wind speed observations at SIUE on training (left) and testing (right) sets.
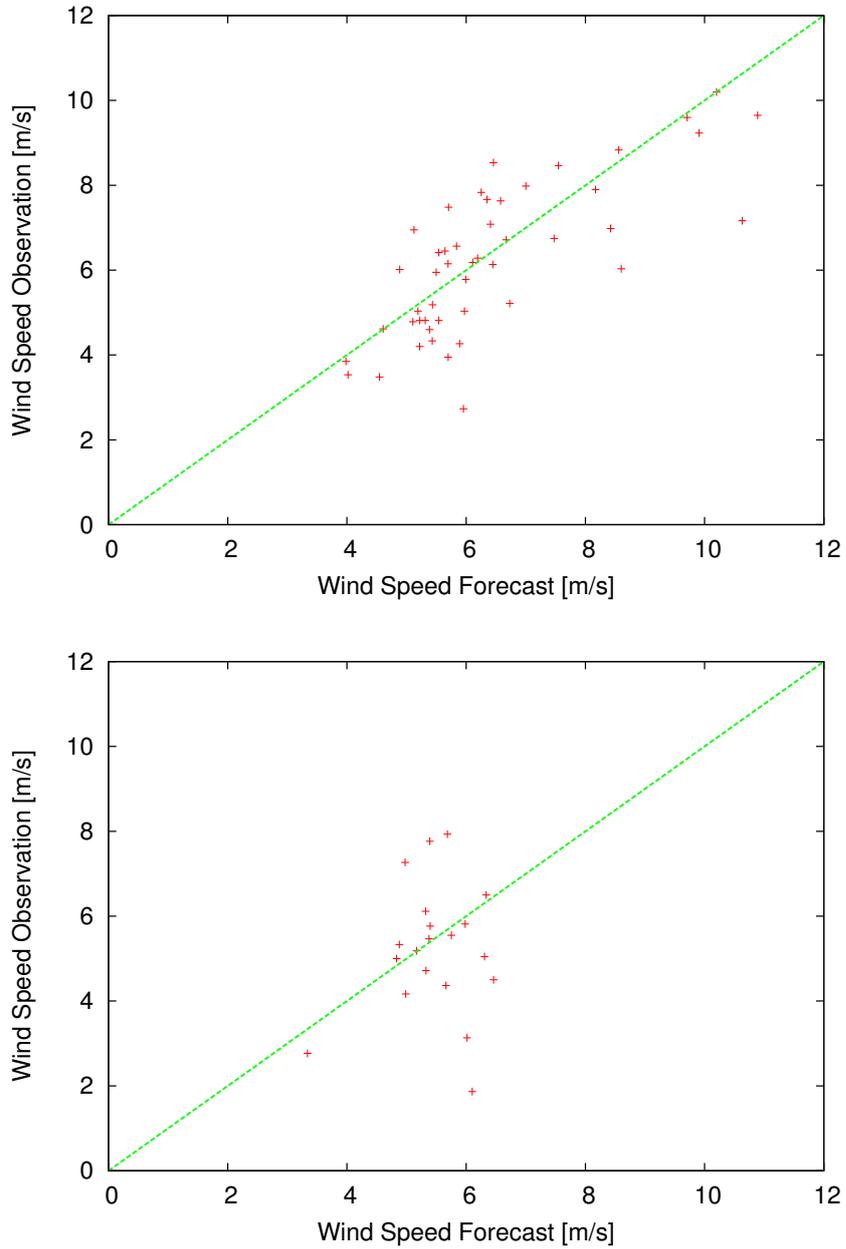
Figure 13: Correlation between the best GP model and wind speed observations at Wilmington on training (left) and testing (right) sets.
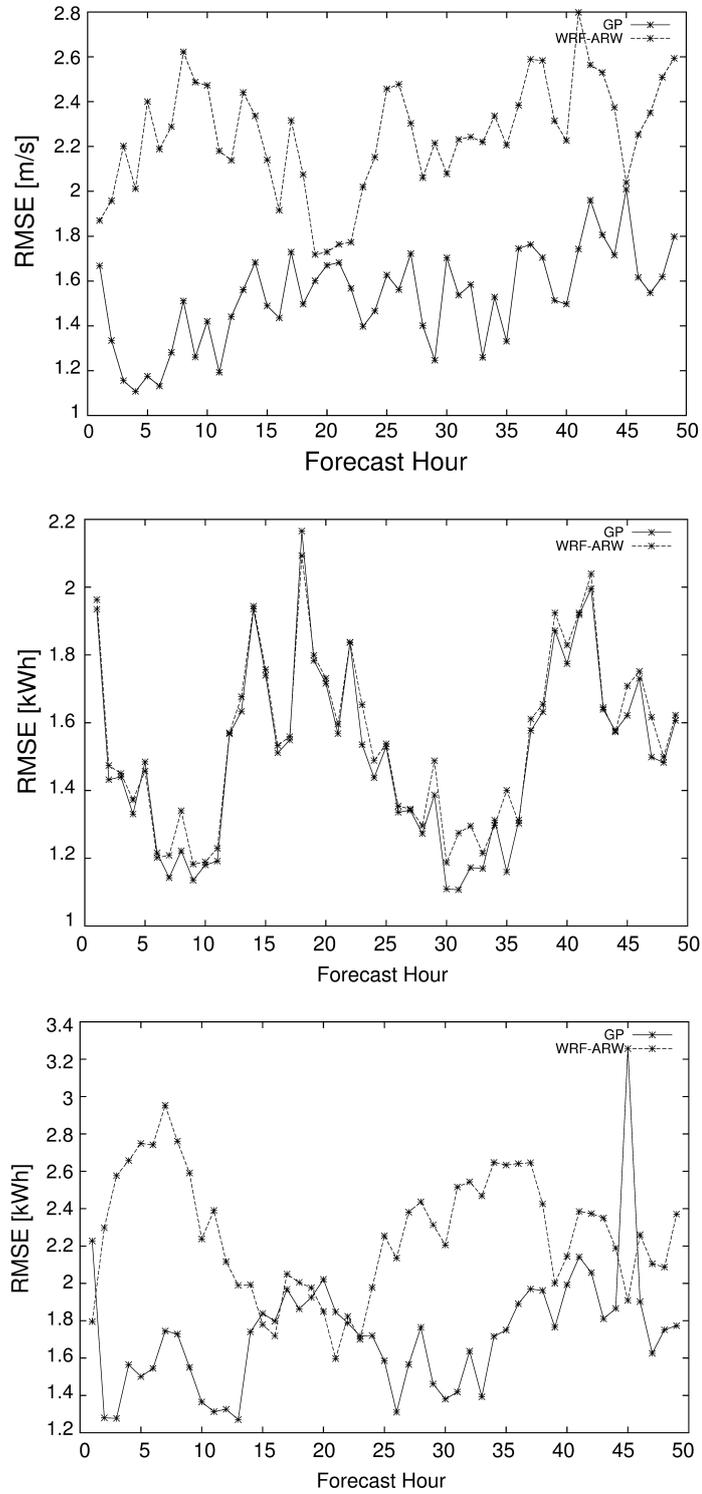
Figure 14: Average Error at each hour of the forecast horizon when retraining the algorithm.

Appendix E: Wind Power Forecasting Experimental Results

Genetic Programming for wind power prediction - Experimental Results

In the following are gathered all the results obtained using genetic programming to forecast the total power production of Sotavento Experimental Wind Farm for the Day-Ahead market. The approach was tested on the months of April, May and June. The following figures present the results.
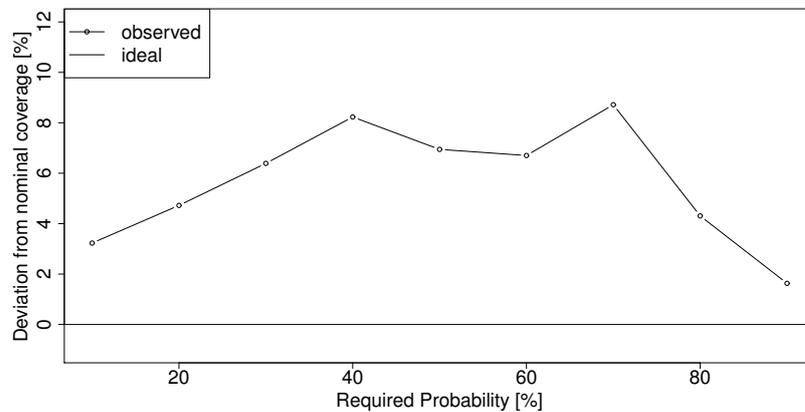
Figure 15: Reliability diagrams of percentile estimations made with the Quantile Regression Forest procedure for May forecasts using the closest point.
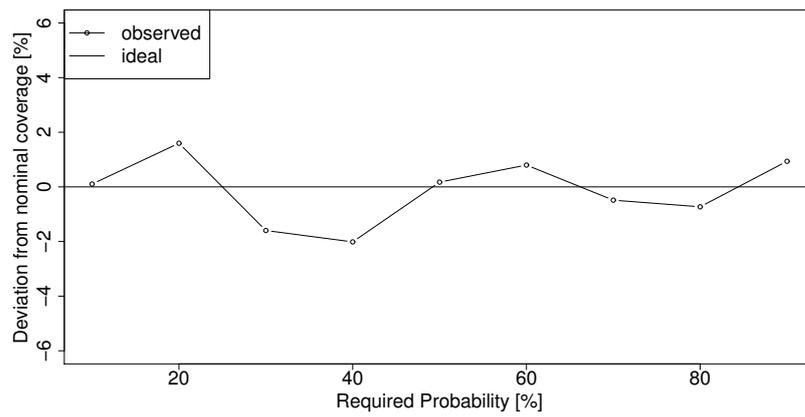
Figure 16: Reliability diagrams of percentile estimations made with the Quantile Regression Forest procedure for May forecasts using a neighbour point.
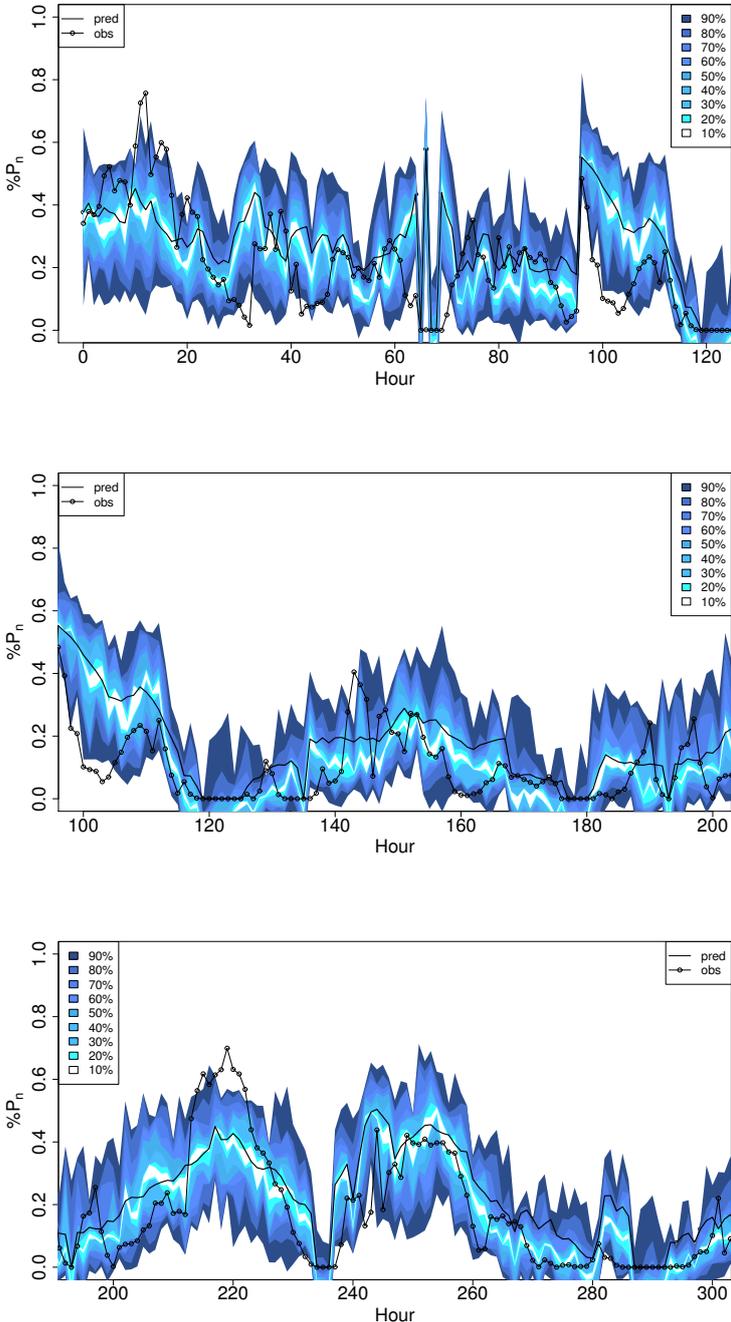
Figure 17: Wind power predictions and intervals for the month of May using the closest point.
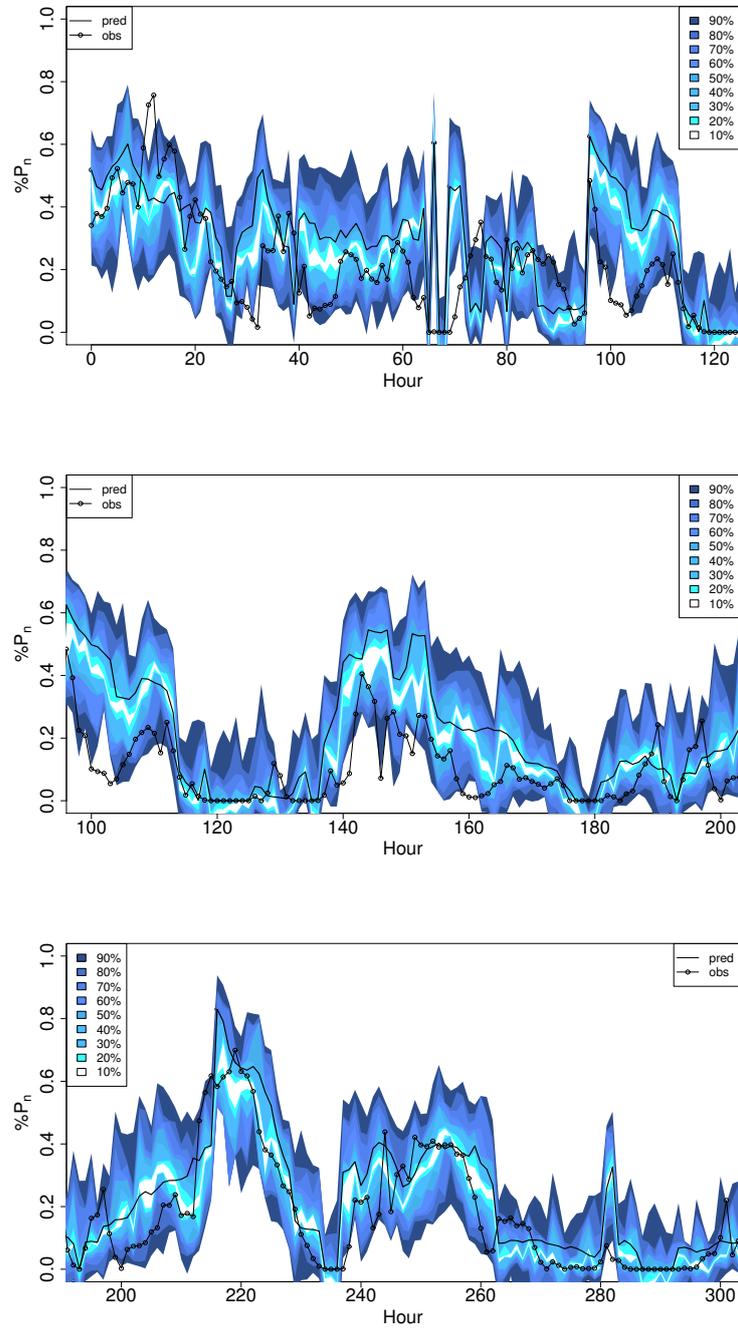
Figure 18: Wind power predictions and intervals for the month of May using a neighbour point.
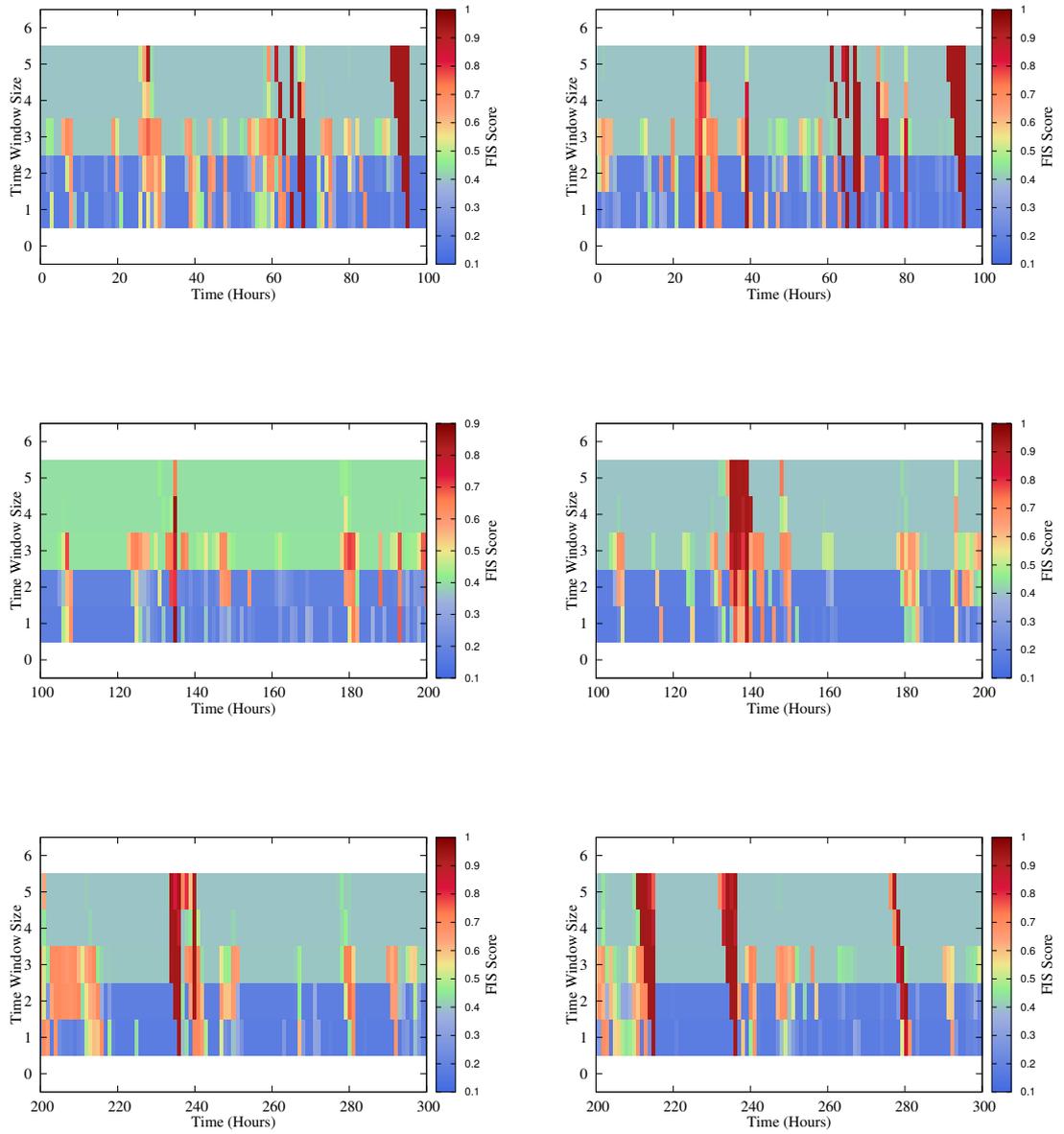
Figure 19: Fuzzy inference system scores corresponding to *ramp up* events at the closest point (left) and at a neighbour point (right) during the month of May.
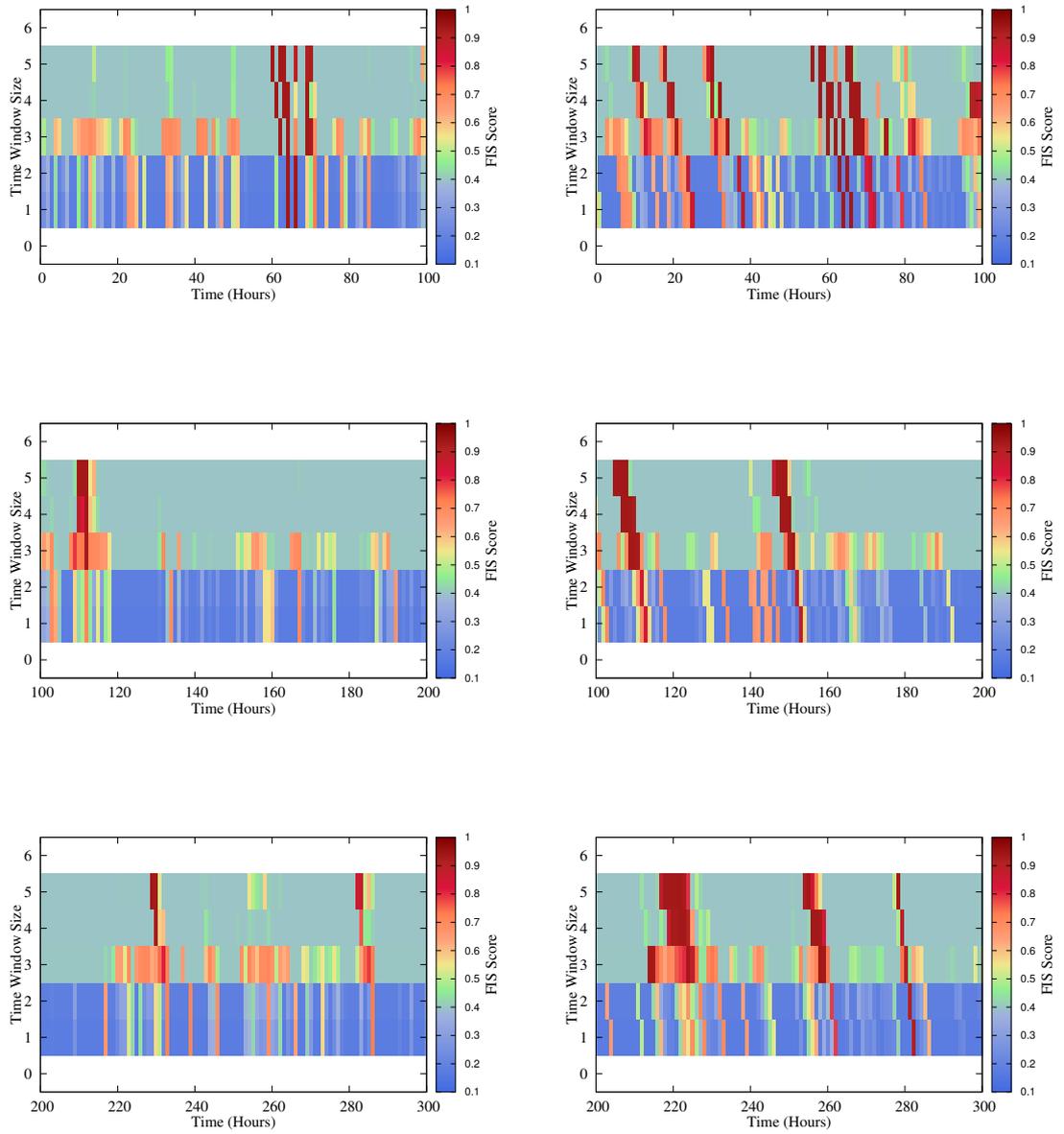
Figure 20: Fuzzy inference system scores corresponding to *ramp down* events at the closest point (left) and at a neighbour point (right) during the month of May.
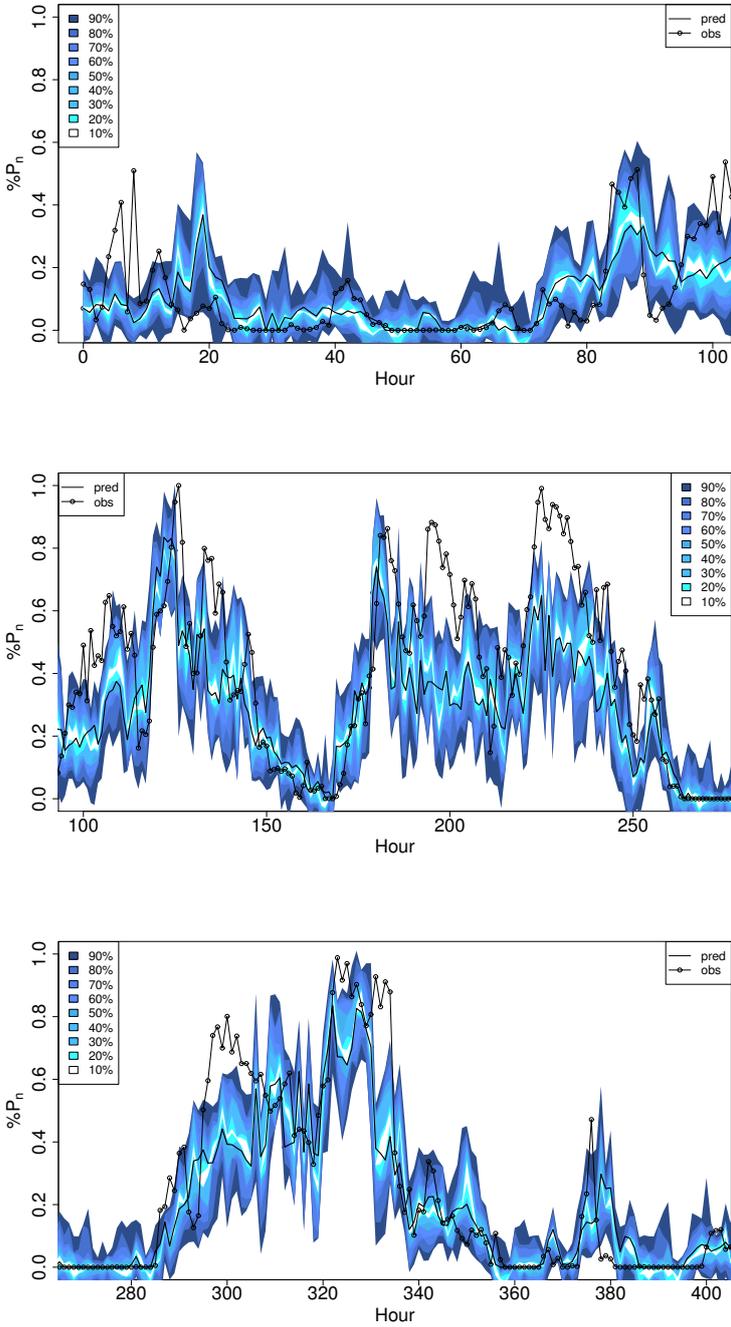
Figure 21: Wind power predictions and intervals for the month of June using the closest point.
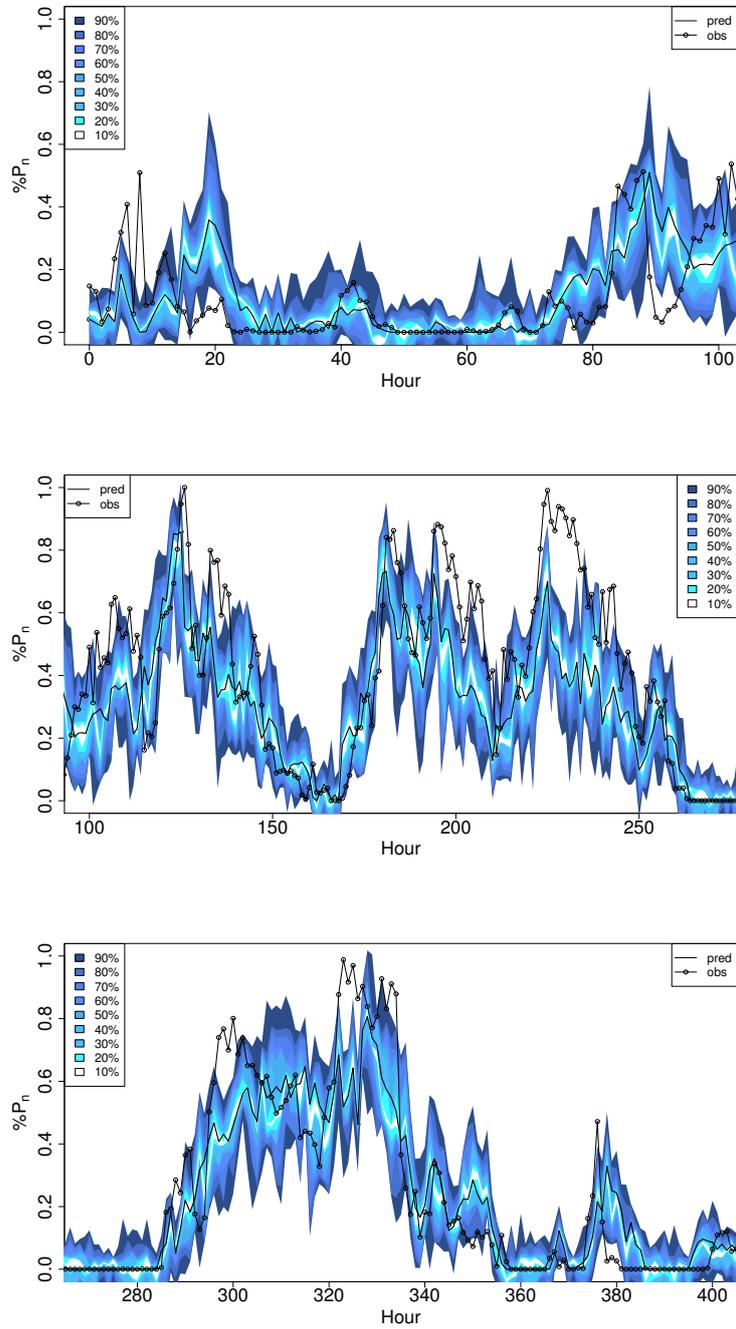
Figure 22: Wind power predictions and intervals for the month of June using a neighbour point.
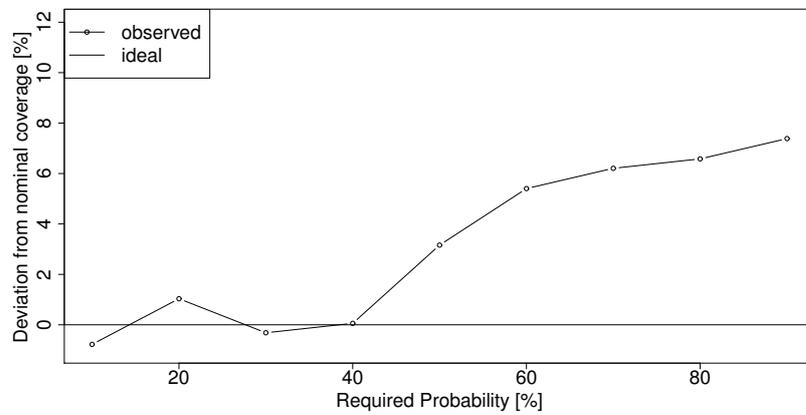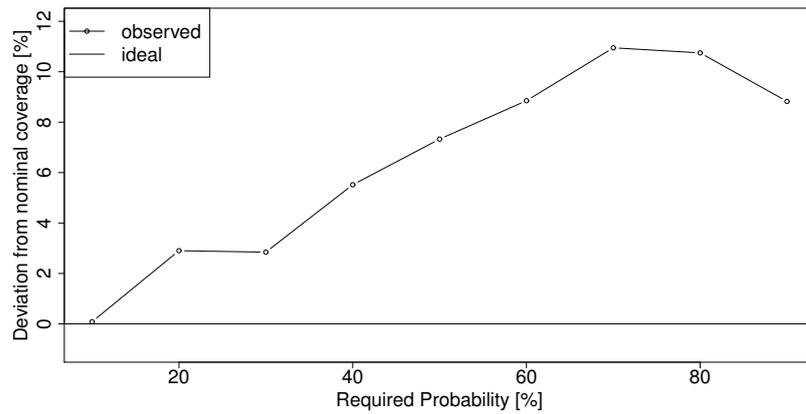
Figure 23: Reliability diagrams of percentile estimations made with the Quantile Regression Forest procedure for June forecasts using the closest point (top) and a neighbour point (bottom).
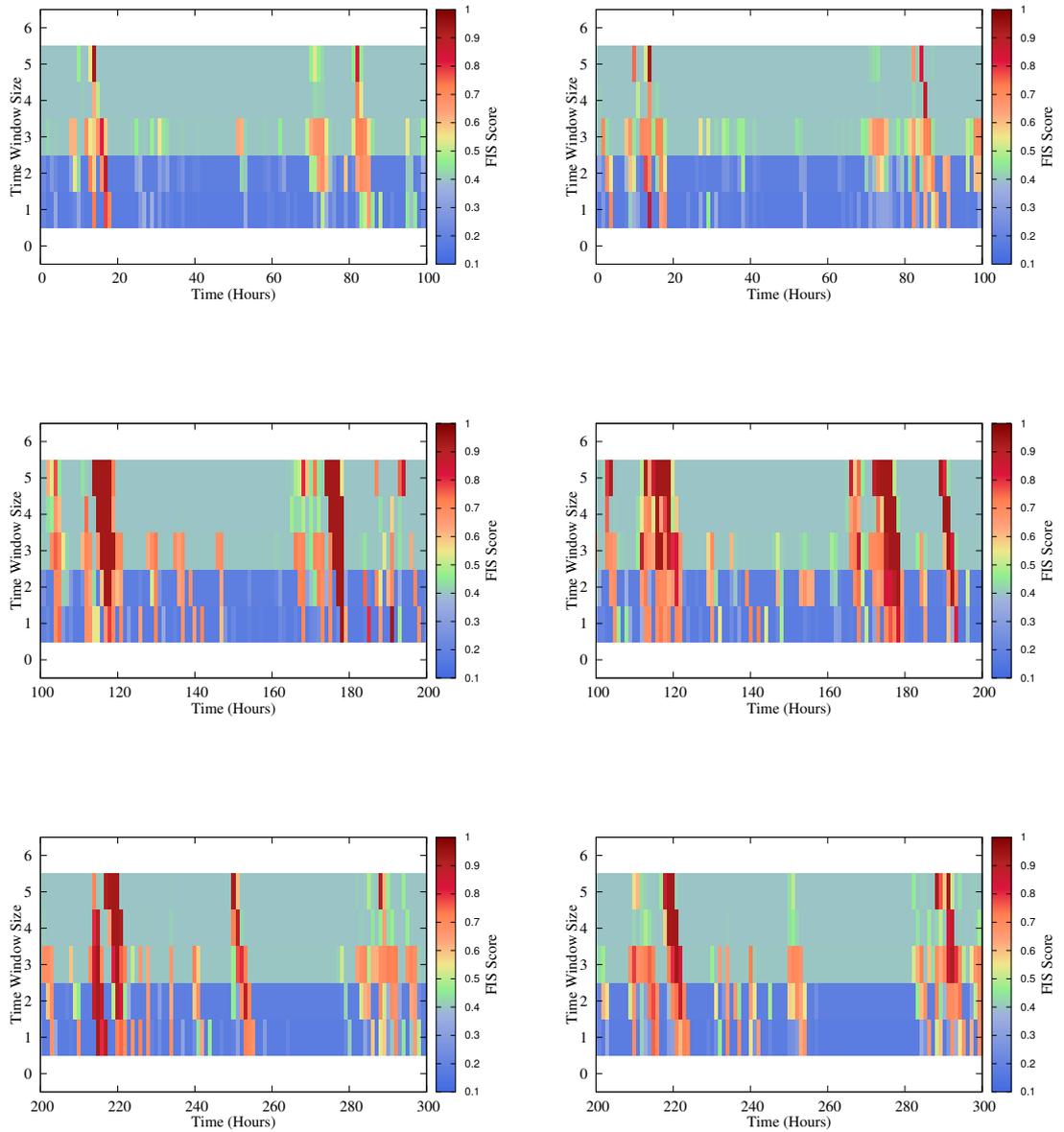
Figure 24: Fuzzy inference system scores corresponding to *ramp up* events at the closest point (left) and at a neighbour point (right) during the month of June.
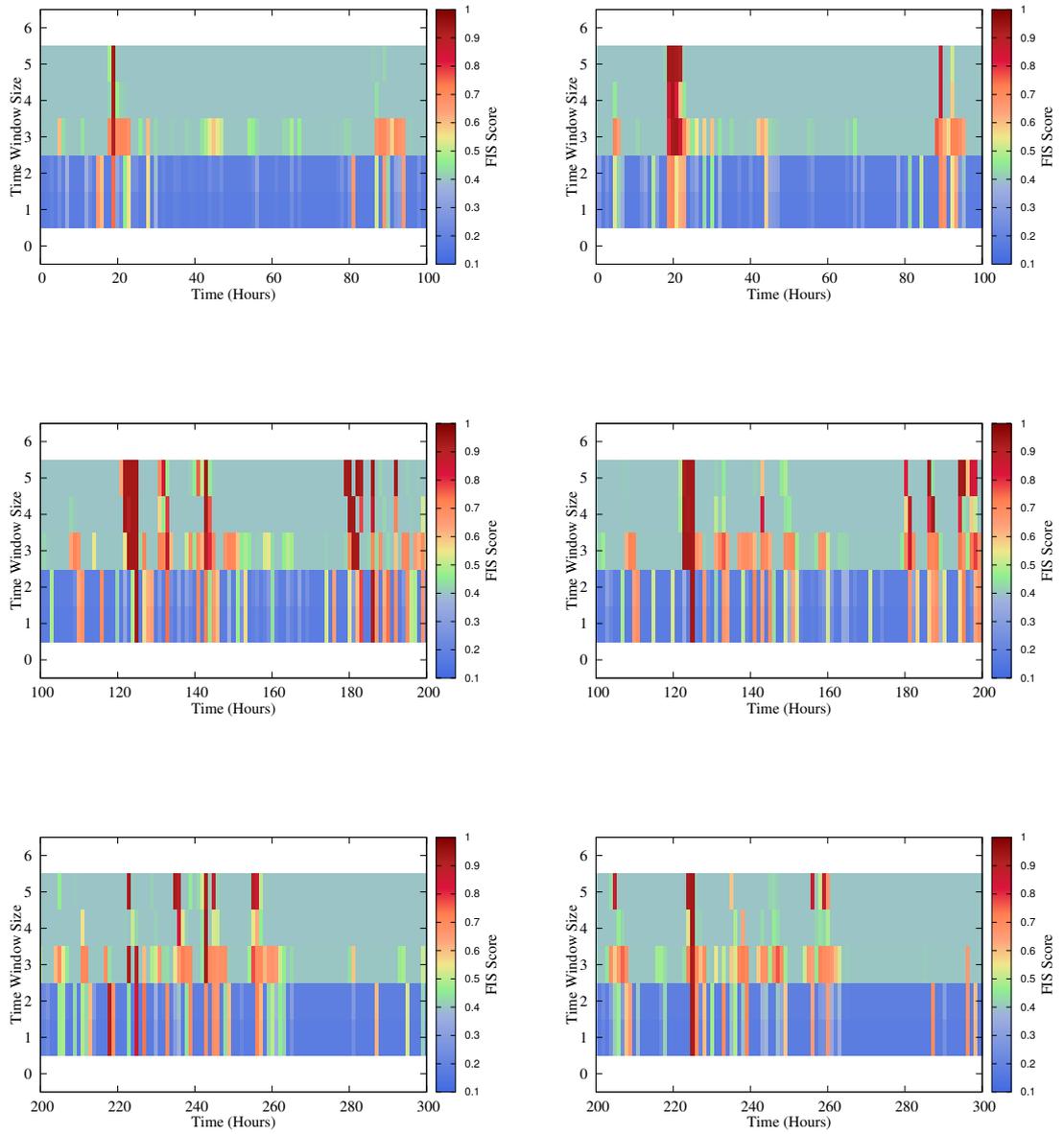
Figure 25: Fuzzy inference system scores corresponding to *ramp down* events at the closest point (left) and at a neighbour point (right) during the month of June.

Appendix F: The Wind Variability of Galicia

The Wind Variability of Galicia

   With the possibility of misplacement errors, it is interesting to take a look at the big picture to see what is happening in terms of wind direction and temperature in the surroundings of the park and in the whole Galicia region.

   Figure 26 shows a wind barb plot on April 11th in the Galicia region. The barbs show the speed and direction of the wind as well as the temperature (defined with a color) at the location where the barbs are placed. The two left plots in the figure show the wind forecasted one day before the actual observation (the 10th of April), and the right hand figures show the forecast produced on the same day. The wind barbs show a high variability of temperatures in the region, which suggest the possibility of micro-climates which are well known in the region. It can be observed in both forecasts that there is a change in the wind direction at the location of the farm.

   Figure 27 shows the forecasts for April 13th, 2012 on the day before and on the same day. This period of time corresponds to two power output increases that were not well forecasted in time (see Figure 8.1). This case is interesting as it is so far not identified if the error is related to a NWP misplacement error, to an inconsistency in the power output data or maybe both. So far, the temperature plot shows that the increase in temperature shown at Marco da Curra in this day is not of the same magnitude as the increase in the prediction which is also confirmed with the wind barbs. There is an evident underestimation which adds to the time misplacement characteristic of this day. The wind barb plot shows how at 17T there is a general increase in temperature (right plot) and a change
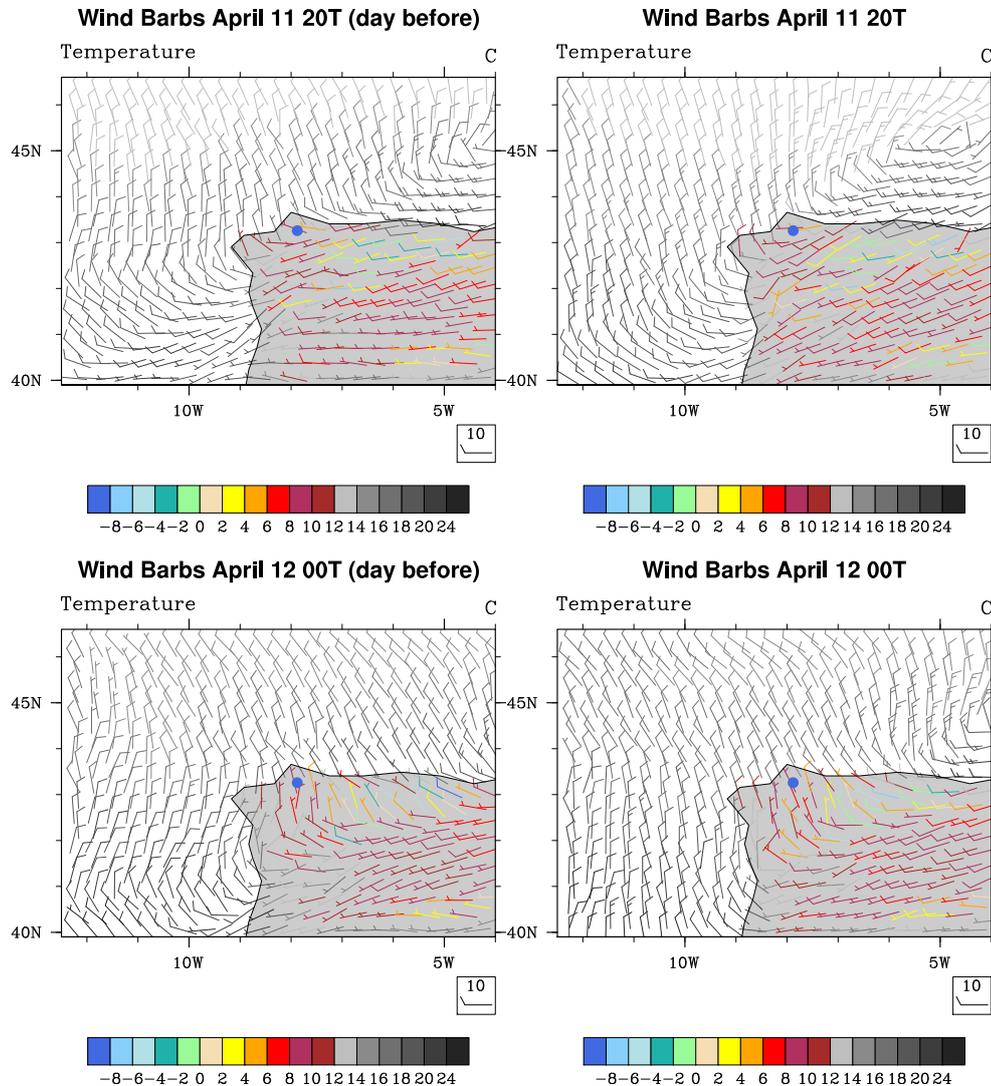
Figure 26: Wind barbs that show the wind speed, direction and temperature forecasted the day before (left) and on the actual day (right). The blue dot is to show the location of Sotavento Wind Farm.

in the direction of wind which changes from northwest to west and later on to southwest at the end of the day. A small timing misplacement of how wind direction changes at the location of the park could do a difference in the power output estimation.
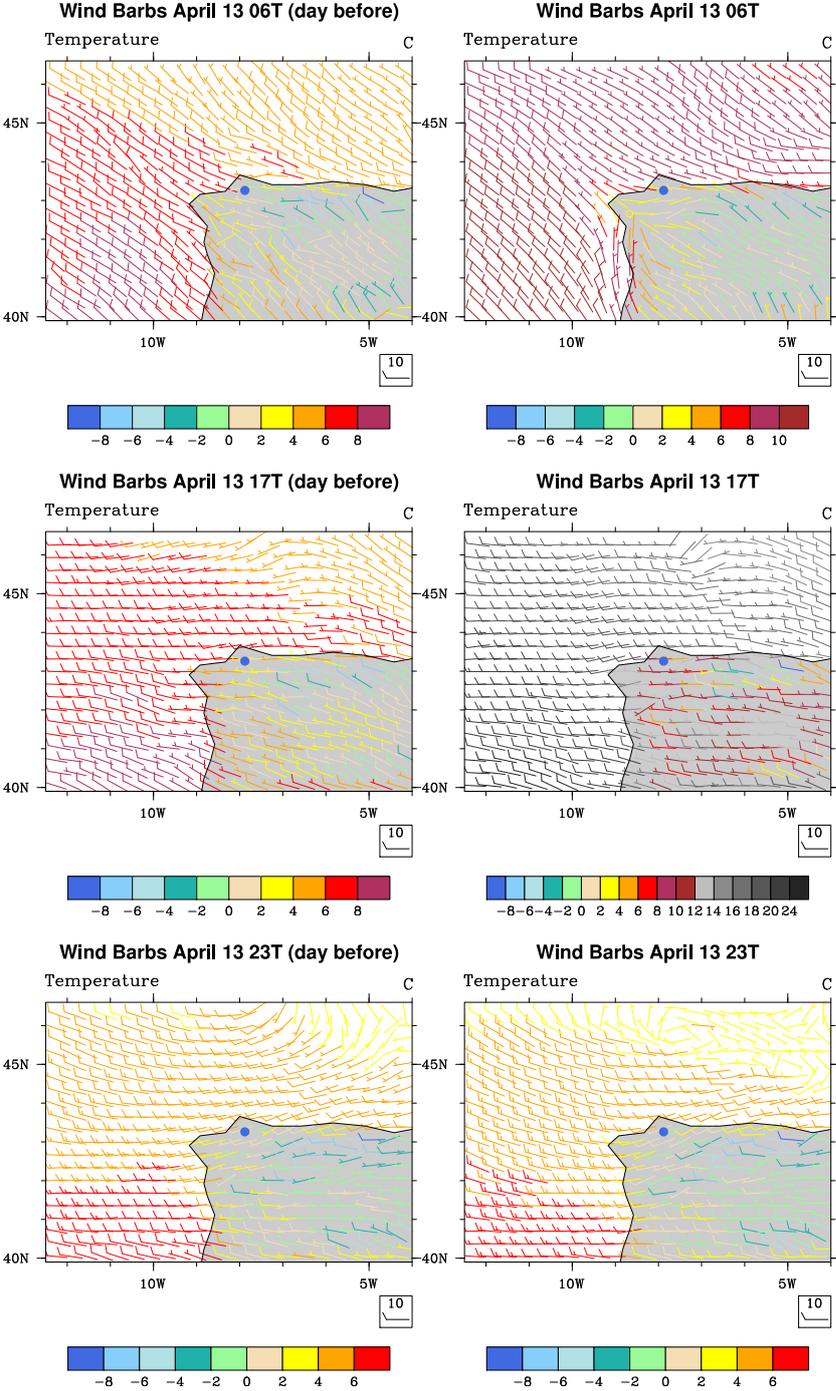
Figure 27: Wind barbs that show the wind speed, direction and temperature forecasted the day before and on the actual day. The blue dot is to show the location of Sotavento Wind Farm.