Forecasting of wind power production in the Netherlands

THESIS

by
Yme Joustra
Master of Science in Computer Science
University of Twente, the Netherlands

August 26, 2014

Supervisors from the University of Twente:
Dr.ir. Maurice van Keulen
Dr. Mannes Poel
Dr.ir. Dolf Trieschnigg
Supervisors from Raedthuys:
Drs. Martijn Tielkes
Ir. Erik Veer

Preface

After working for several years at Raedthuys Pure Energie, it became clear how interesting the renewable energy industry is. Searching for interesting areas in the energy world is not hard (since there are plenty), but to create a decent scientific research of it is. Combining my interest in the area of machine learning with the possibility of performing a research at Raedthuys narrowed the amount of areas down to a few certain areas in the field of forecasting wind power. At the company I mainly focus on the analysis of producing wind power. During this analysis the company noticed that the forecasting methods they used did not perform as well as they expected. The challenge finding a suitable forecasting technique for day-ahead wind power predictions motivated me to conduct research in this area.

I would like to thank Maurice van Keulen and Mannes Poel, my supervisors from the University of Twente for guiding me through the process of this research and for providing me feedback. I would also thank Dolf Trieschnigg for his guidance and feedback during the literature study part of this research.

Furthermore, I would like to thank Martijn Tielkes and Erik Veer, my supervisors from Raedthuys for guiding me through the process of wind power forecasting and for helping me to understand the energy market in detail. I have learned many things during my research.

Finally, I would like to thank my fellow students, my best friends and especially my girlfriend for supporting me during the past few months, for their feedback and useful discussions on the subject.

Yme Joustra

Abstract

Wind power has become an important source of power for some countries because wind is renewable, wind power is clean and no pollutants are produced compared to fossil fuels which are mainly used for the generation of energy today. Because of these reasons also in the Netherlands attention towards the use of wind power has grown. In the past decade, a lot of research has been performed on the forecasting of wind power production over a period of minutes, days, months and years. This paper focuses on day-ahead forecasting and starts with a theoretical and economical overview of the electrical grid and energy market. The main reasons to focus day-ahead forecasting is to ensure the balance between the demand and supply of electricity and because the energy needs to be sold against a day-ahead spot price. Based on a literature study in the field of forecasting wind power it has been found that factors such as geographical location, data sources and grid sizes show influence on the accuracy of the data and therefore influence the prediction of wind power. Furthermore, based on the literature input parameters such as wind speed, wind direction, weather stability, availability, relative humidity and seasonal data have been found useful as input data for forecasting methods to forecast wind power day-ahead. From a large set of forecasting methods it has been found that the most used techniques to predict wind power day-ahead are physical methods, and statistical or hybrid methods such as neural networks.

This research has obtained forecasting results from a Random forest, Feed forward neural network and a hybrid model consisting of a combination of unsupervised k-nearest neighbour clustering and a neural network. These results have been compared with the forecasting results obtained from an external organization. Based on the comparison of monthly and average monthly MAPD and RMSPD we have found that the Feed forward neural network and the hybrid model are able to obtain a performance equally or even better compared to the external forecasting for a single turbine. The input parameters that made the difference were the u-vector, v-vector, the use of SCADA data and the wind speed time lag 1.

Furthermore, the three forecasting models did perform less good compared to the external forecasting on forecasting wind power generated by a wind farm. Main reasons are because we did not take shadowing effects from other turbines into account and also the lack of fuzzy rules overfitted the neural networks at higher wind speed values. The random forest however was more robust and performed best of the three models.

Glossary

Electrical grid: The grid through which electricity is being transmitted.

TSO: Transmission System Operator who is keeping the balance between supply and demand of electricity on country level.

DSO: Distributed System Operator who is keeping the balance between supply and demand of electricity on region level.

Energy trader: Is responsible for optimising the production and supply forecast by buying and selling electricity on the wholesale trading market.

PR: Program Responsible is responsible for informing the TSO based on the most actual forecast and trading position in order to support the balance between demand and supply.

Short: There is being less energy produced than forecasted. In other words an underproduction.

Long: There is being more energy produced than forecasted. In other words an overproduction.

APX: A day ahead (or spot) market price based on submitted orders of demand and supply of electricity on a hourly basis.

Helper: A producer whose portfolio is short when the TSO is long or whose portfolio is long when the TSO is short.

Causer: A producer whose portfolio is short when the TSO is short or whose portfolio is long when the TSO is long.

Imbalance: The difference between the forecasted amount and the allocated amount of energy produced.

NWP: Numerical weather predictions (NWP) are based on the current weather conditions of the atmosphere and are calculated using models. Numerical means that each data value is represented as a number (a series of numbers).

Wind power: Energy produced by wind turbines, also called Wind energy.

Contents

1 Introduction **1**

2 Background **3**
2.1 Infrastructure of the electrical grid 3
2.2 Parties and their roles in the energy market 4
2.2.1 Physical flow 5
2.2.2 Information flow 6
2.2.3 Cash flow 6
2.2.4 Horizontal dashed line 7

3 The importance of forecasting **7**
3.1 Balancing demand and supply 7
3.2 Economic perspective 8
3.2.0.1 Scenario 1 - $P_t > 0$ and $P_t < P_a$: 9
3.2.0.2 Scenario 2 - $P_t > 0$ and $P_t > P_a$: 10
3.2.0.3 Scenario 3 - $P_t < 0$: 10
3.3 Discussion 11

4 Introduction to forecasting models **12**
4.1 Random forest model 12
4.1.1 Process of the Random Forest 12
4.2 Feed forward Neural Network 13
4.2.1 Process of the feed forward neural network 13
4.2.2 Learning the algorithm 14
4.2.3 Hidden neurons and layers 14

5 Related work **14**
5.1 Important factors for forecasting wind power 15
5.1.1 Data sources 15
5.1.2 Grid area 16
5.1.3 Geographical location 16
5.2 Input parameters for forecasting models 17
5.2.1 Wind speed 17
5.2.2 Wind direction 18
5.2.3 Weather stability 19
5.2.4 Availability 19
5.2.5 Relative Humidity 20
5.2.6 Seasonal 20
5.2.7 Temperature and pressure 21
5.3 Forecasting models 21
5.3.1 Statistical models 22
5.3.1.1 Regression trees 22
5.3.1.2 Time series models 22
5.3.1.3 Artificial Neural networks (ANN) 23
5.3.1.4 Support vector regression/machine 24
5.3.1.5 Discussion 25
5.4 Feature selection methods 26
5.5 Evaluation metrics 27

6 Methodology **27**
6.1 Data sources . . . 27
6.1.1 Data description . . . 28
6.1.2 Meteorological data . . . 28
6.1.3 SCADA data . . . 29
6.1.4 Production data . . . 29
6.1.5 Data tools . . . 31
6.2 Methods of analysing data . . . 31
6.3 Training, validation and testing forecasting models . . . 32
6.4 Performance measurements . . . 33

7 A hybrid model **34**
7.1 Step 1: Unsupervised k-nearest neighbor clustering . . . 34
7.1.1 Process of the algorithm . . . 34
7.1.2 Applying unsupervised k-nearest neighbor clustering algorithm . . . 35
7.2 Step 2: Applying feed forward neural network . . . 36

8 Data analysis **37**
8.1 Grid point reduction . . . 37
8.2 Selection of input features . . . 38
8.3 Correlation studies . . . 39
8.3.1 Correlation coefficient . . . 39
8.3.2 Discussion . . . 40
8.3.3 Autocorrelation coefficient . . . 45
8.4 Cook's distance measure . . . 46
8.5 Scada data and wind speed distribution analysis . . . 46
8.6 Conclusion . . . 48

9 Results **49**
9.1 Single turbine . . . 51
9.2 Wind farm of 27 Turbines . . . 53
9.3 34 Turbines . . . 54
9.4 Financial impact of one turbine . . . 55

10 Discussion **56**
10.1 Single turbine . . . 56
10.2 Wind farm of 27 turbines . . . 58
10.3 34 Turbines . . . 60
10.4 Input parameters . . . 60

11 Conclusion **62**

12 Future work **64**

1 Introduction

Consumers have become accustomed to a stable electricity supply. This electrical supply is produced using different sources such as burning fossil fuels, using solar panels or wind turbines. This research will focus on energy used for the electricity supply which is produced by wind turbines. This form of energy is also called wind power. A major difference between wind power and fossil fuel energy is the *predictability* of producing energy. A predictable source can be used to balance the supply to the demands. Balancing the demand and supply of electricity is important to ensure continuous electricity supply. For instance one expects to start one's computer when one puts the plug into the socket. Fossil fuels such as gas differ from wind power in terms of predictability, as is indicated below:

Predictability - Fossil fuels (gas): Energy can be produced by a fossil fuel such as gas. Fossil fuels have a limited capacity and are in stock. Therefore the production of energy is predictable.

Predictability - wind: Energy can also be produced by wind turbines. However wind is unpredictable and therefore one cannot ensure the availability of wind power when needed. In other words: one cannot rely on producing energy using wind if one cannot predict this source.

The predictability of wind power production is therefore a major drawback, because one cannot ensure the availability of wind power when needed. However, compared to fossil fuels, wind has its advantages; there is plenty of wind available, wind is renewable, wind power is clean and no pollutants are produced.

Using only energy produced by wind turbines is not yet possible since no country is able to provide enough wind power to ensure a continuously electricity supply. The Netherlands for instance, produced only 4.9% wind power by wind turbines of its total electricity consumption in 2012 [6].

Therefore nowadays the supply of electricity is based on the production of both forms of energy, fossil and renewable. However an issue arises using this combination, giving a scenario.

Scenario - combination: Energy will be produced by fossil fuels and wind turbines. Since wind is a variable source one does not know how much energy is produced by wind turbines. This makes it difficult to keep the balance between demand and supply of energy, because one does not know how much fossil fuels to use to keep this balance.

Clearly, it is difficult to keep the balance between the demand and supply of energy when at least one energy source is uncontrollable, which is addressed by Soman et al. [43]. A solution to deal with the uncontrollability of renewable energy is the use of accurate forecasting techniques to predict the production of energy by these sources. Forecasting techniques provide forecasts about the amount of wind power produced by wind turbines.

This research focuses on the subject 'forecasting the production of wind power'. Forecasting wind power can be performed for different time scales, from thirty minutes to a week, month or a year into the future. In this research we focus on a time scale called day-ahead (24 to 48 hours) forecasting. The reason focusing on this time scale is because regulators of the net; like the Transmission System Operator (TSO) and the Distributed System Operators (DSO), need to know how much wind power will be produced day-ahead so they can ensure the balance between the demand and supply of electricity. Using these forecasts regulators can respond easier on balancing the demand and supply of electricity, because now the amount of wind power generated by turbines does not come as a surprise.

Another reason focusing this time scale day-ahead is because the energy needs to be sold against a day-ahead spot price. Both reasons are discussed in more detail in section 3.

A lot of research in the field of forecasting wind power has been performed. Literature overviews (e.g. [27],[49]) have identified different prediction models for different time scales in different countries. Also research has been conducted to find the right input parameters that influence the outcome of the prediction model. The amount of different prediction models and the use of factors in literature studies lead to the following research questions:

RQ1 Which factors and input parameters to predict wind power have been described in literature? And which of those have been found successful?

RQ2 Which forecasting models have been found the most relevant by previous literature to predict wind power generated by wind turbines?

These first two research questions are answered performing a literature study which is given in section 5. Determining which factors and input parameters have been found successful is based on what literature recommends to use to predict wind power. If parameters increase the accuracy of the prediction model they will be found successful and the other way around.

Besides the literature study, this research will be conducted at the company Raedthuys Energie BV. Since more attention is given to renewable energy in the Netherlands, more research in the field of forecasting wind power generated by turbines can be conducted, specifically in the Netherlands.

Raedthuys Energie BV, located in Enschede, is a renewable energy producer in the Netherlands. About 50 employees are working at Raedthuys. Their mission is to stimulate the use of renewable energy and the goal is the delivery of sustainable energy from wind and sun to its customers using wind turbines and solar panels. Raedthuys earns his money by realizing a large set of activities which are developing, investing, building, managing and ensuring renewable energy projects and the delivery of renewable energy [38].

Forecasting wind power is important for Raedthuys because they sell their forecasted energy day-ahead. It is a 'risk' to sell the production of energy real-time, which will be explained in section 3.2. It is therefore important to have a model that can forecast wind power generated by wind turbines as accurate as possible. Currently Raedthuys is using forecasts of wind power provided by external organizations. However they want to have their own forecasting model so they do not depend on other organizations. Therefore we have to build a forecasting model that performs at least equal to the current forecasts, which gives us the overall research questions of this research:

RQ3 How do the recommended forecasting methods identified in literature studies perform to the forecasts provided by external organizations?

RQ4 Which input parameters and optimizations have to be applied on the recommended forecasting models to achieve an as accurate prediction model compared to the forecasting model from the external organizations?

Identifying the input parameters mentioned in research question three is done using several correlation studies see also section 8. The input parameters are used to find the forecasting model that performs best.

The structure of this document is as follows. Section 2, explains the background of the subject. Section 3 explains in detail the importance of forecasting wind power generated by wind turbines. Section 4 describes which forecasting techniques are used to predict wind power. Section 5 provides a literature overview about the subject. Section 6 describes the methodology of this research, what data and methods have been used and how they have been used. Section 7 describes our proposed hybrid model. Section 8, explains stepwise how data has been selected and extracted. Section 9 provides the results obtained by the different forecasting techniques with the right parameter estimation. Section 10 discusses the obtained results. Section 11 presents the conclusion of this study. Finally, some future work is given in section 12.

2 Background

To understand the research questions and our related work 5 we explain the infrastructure of the electrical grid (section 2.1) and the energy market (section 2.2).

2.1 Infrastructure of the electrical grid

Figure 1 shows the infrastructure of the electrical grid in the Netherlands. As one can see, the figure presents different parties who are part of the grid. In this figure only renewable energy producers wind and solar have been taken into account. Other renewable energy sources such as hydro energy or biomass energy are not included in the picture because these are beyond the scope of this research.

Furthermore, the figure also presents different levels of electrical power transmission (the load transmission of electrical energy). Each of the load transmission levels of electrical energy will be explained below:

220 - 380 kV: This is the top level of the electrical power transmission. The power plants produce a high voltage of electricity using fossil fuels. Using high voltages is because they can be transferred over large distances with less losses. The electricity is transferred via transmission lines. The load of electricity in the transmission lines is regulated by the Transmission System Operator (TSO), TenneT in the case of the Netherlands [45].

50 - 150 kV: The TSO converts high voltage to a low voltage [45]. This is the second level of the electrical power transmission. The electricity is merged with the electricity produced from wind turbines (and other renewable sources, which are not included into the picture); the reason why turbines are connected to this level of voltage is because the turbines do not produce as much as power plants and therefore it is not necessary to

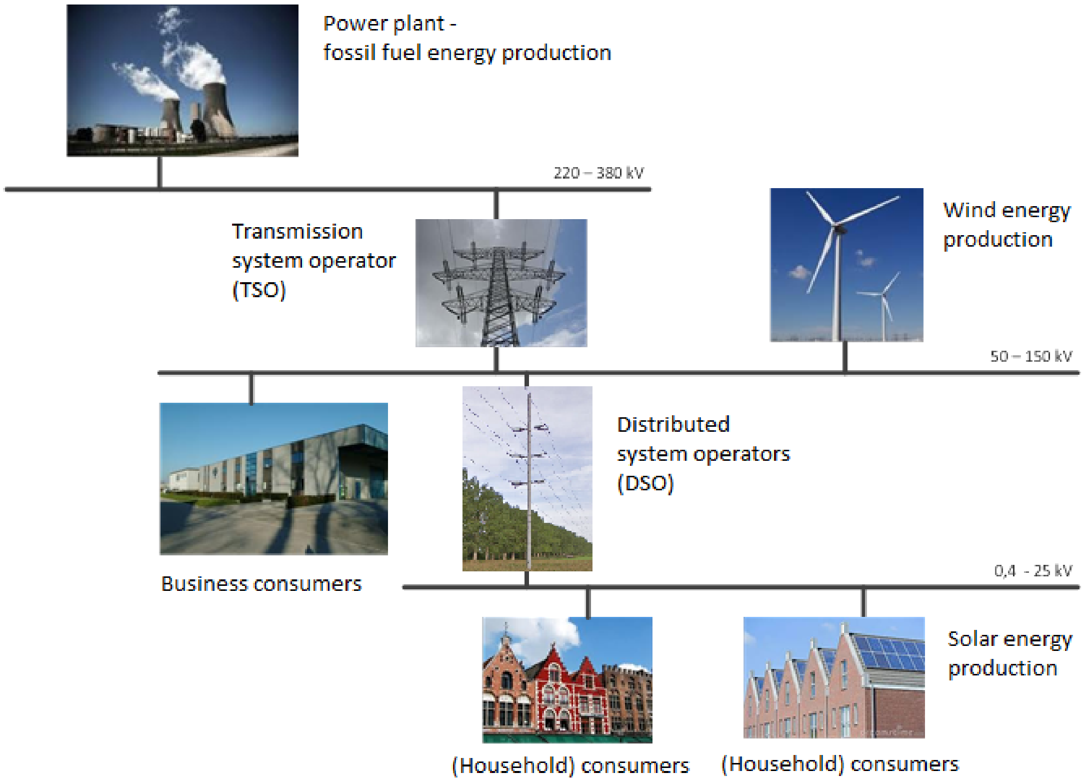


Figure 1: Infrastructure of the electrical grid (provided by Raedthuys Energie BV.)

connect turbines to the highest level of voltage. The electricity is transmitted to large consumers of electricity (e.g. industrial consumers) and Distributed System Operators (DSO). DSOs regulate the grid on a smaller (regional) area than the TSO does.

0.4 - 25 kV: The electricity transmission from the second level is converted by the DSO to a much lower voltage. This load of electricity is distributed to end users (e.g. household consumers) that belong to the area of the DSO. End users with solar panels produce electricity which flows into the grid or is consumed by the end user itself.

The TSO also collaborates with the other European TSOs to compensate electricity shortages and surpluses [45]. A discussion of the European market is beyond the scope of this research.

2.2 Parties and their roles in the energy market

Figure 2 shows the participating parties in the energy market. In this figure you see a graph containing vertices, edges and one horizontal dashed line. The vertices are the parties participating in the electricity market. One vertex contains a horizontal line. This party is divided into two different roles, the Energy

Trader and the Program Responsible (PR). The job of the energy trader is to buy energy from the producers, sell energy to the suppliers and trade energy with other traders on the Energy trading market. The PR is responsible for informing the TSO based on the most actual forecast and trading position in order to support the balance between demand and supply. [9][46].

The edges show the flow between the parties. There are three types of flows visible in the figure, namely the physical flow on the grid (black lines (MWh)), the Information flow (Green lines (Info)) and the Cash flow (Red lines (€)). To understand the graph each flow will be described by explaining the edges.

The horizontal dashed line in the middle divides the picture into an energy market (upperside) and the electricity grid (underside). The upper side and underside always have to be balanced. More information about this line will be given in section 2.2.4.

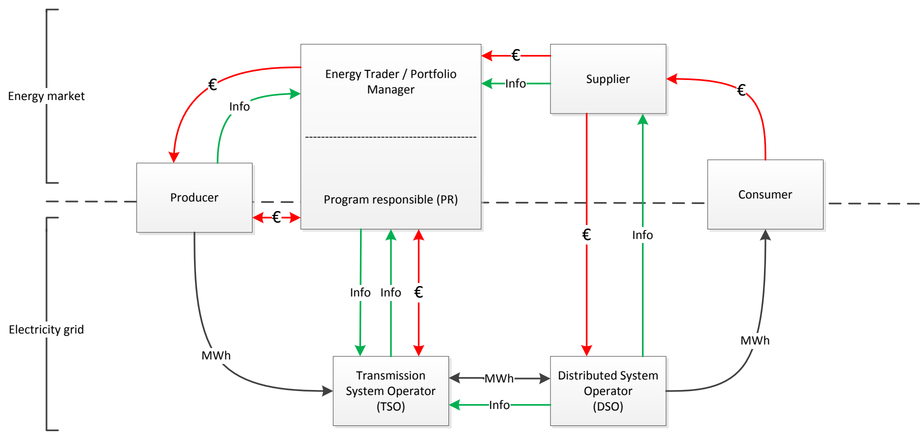


Figure 2: The structure of parties in the energy market

2.2.1 Physical flow

The electricity produced by the wind power producer is being delivered at the consumer via the electrical grid. Before the electricity gets to the consumer it passes the Transmission System Operator (TSO) and the Distributed System Operator (DSO). The electricity can flow from the TSO to the DSO and the other way around to ensure decentralized production of energy. The TSO and DSO both are independent companies. Their main job is to manage the balance between the demand and supply of electricity. The electrical grid is being balanced by regulating the amount of electricity through the electrical grid. To ensure the quality and continuous supply of electricity the grid needs to be maintained [34][45]. Maintaining the grid costs the TSO and DSO money. The difference between the TSO and DSO is that the TSO is managing the transport of electricity on the electrical grid on country level, whereas the DSO is managing the electrical grid on a specific region.

2.2.2 Information flow

The flow of information between parties consists of information about the forecasted values to be produced and consumed and the actual values produced and consumed. For each of the parties the following information is being shared.

Producer → Energy trader: The producer informs the Energy trader about the forecasted amount of electricity produced for the next day.

Supplier → Energy trader: The supplier informs the Energy trader about the forecasted amount of electricity consumed by the consumers.

PR → TSO: The PR informs the TSO about its purchase and sale transactions of electricity. The PR tries to balance the demand and supply of electricity using the information from the supplier and the producer. To keep the balance the PR is responsible for buying and selling electricity on the market.

DSO → TSO: The DSO regulates the electricity grid real-time in its own region and informs the TSO about the actual amount of electricity consumed by the consumers.

DSO → Supplier: The DSO informs the supplier about the actual amount of electricity consumed by the consumer.

TSO → PR: The TSO regulates the high voltage electricity load through the transmission lines real time and informs the PR about the actual amount of electricity produced and consumed.

2.2.3 Cash flow

In the cash flow a party buys or sells energy, settles imbalances or maintains the electrical grid. How the cash between parties flows is explained here:

Energy trader → producer: The Energy Trader pays the producer for selling his amount of electricity.

Supplier → Energy trader: The supplier is responsible for buying electricity from the energy trader.

Consumer → Supplier: The consumer buys the electricity from the supplier and pays for the services of the DSO, which is indirectly done via the supplier.

Supplier $\rightarrow$ DSO: The supplier pays the DSO for its services. These costs have already been paid by the consumer, since the consumer has paid the supplier for the DSO its services.

Producer $\longleftrightarrow$ PR: the difference between the forecasted amount and the allocated amount of electricity produced is called the imbalance and is financially settled. One has to pay the other depending on the so called imbalance price, which can be positive or negative.

Producer $\longleftrightarrow$ TSO: the difference between the forecasted amount and the allocated amount of electricity produced and consumed is called the imbalance and is financially settled. One has to pay the other depending on the so called imbalance price, which can be positive or negative.

2.2.4 Horizontal dashed line

The horizontal dashed line divides figure 2 into the energy market/cash flow (top) and the physical flow on the grid; the electricity grid (bottom). The cash flow represents the forecasted amount of energy bought and sold. This means in an ideal situation where the forecasts are 100% accurate, the cash flow of energy bought and sold represents the energy transmitted through the electrical grid. However since forecasts are never 100% accurate, there exists a difference between the forecasted and allocated amount of energy, which unbalances the demand and supply. This difference is also called the imbalance. To balance the demand and supply two extra cash flows have been added. One between the producer and PR and one between the PR and the TSO (underside of the figure). These cash flows are the settlement of the imbalance by the TSO and are required to ensure the balance between the demand and supply of energy in the overall cash flow as in the electricity flow.

The next section explains the importance of forecasting wind power, by discussing several scenarios to understand the energy market in practice.

3 The importance of forecasting

As mentioned in the previous section the demand and supply of energy needs to be balanced. This is the main reason using forecast energy. However there are also costs bound to these forecasts. These are the two cash flows between the producer and the PR and the PR and the TSO, see also figure 2. In this section we will explain the importance of forecasting wind power by discussing several scenarios. The first subsection discusses balancing the demand and supply of energy and the second subsection discusses the economic perspective of forecasting wind power.

3.1 Balancing demand and supply

As mentioned in the previous section the PR tries to balance the purchases and sales of the demand and supply of electricity using the information about the predicted consumption and production from the supplier and the producer respectively. This information is being passed from the PR to the TSO. The

TSO uses this information to regulate the electrical grid [45]. Without knowing the predicted wind power issues will arise, because the TSO will use only fossil fuels for energy production. Issues such as regulating the grid to keep balance between the demand and supply of electricity can occur. Therefore the importance of knowing the predicted wind power gives the TSO the possibility to estimate the amount of fossil fuels needed to ensure the balance between demand and supply of energy. In other words to regulate the grid easier. When the actual wind power production is known the electrical grid is *short* or *long*. A grid being short means there is less wind power produced by the turbines than forecasted. In other words an under production. A grid being long is the other way around, an over production. The reason of this under or over production is because predictions are almost never 100% accurate. Actions taken when the electrical grid is *short* or *long* are explained in the following scenarios.

Scenario 1: There is a *short* on the electrical grid. The TSO needs to ramp up the energy by producing energy using the fossil fuel power plants to ensure the balance between demand and supply of energy.

Scenario 2: If there is a *long* on the electrical grid, then the TSO needs to ramp down the energy to ensure the balance between demand and supply of energy. For example by transmitting the energy to other TSOs.

Both scenarios solve a problem. The first scenario solves the problem of ensuring continuously electricity supply, the second scenario prevents the problem of overloading the capacity of the transmission system.

3.2 Economic perspective

From an economic perspective view, parties have interest in an accurate forecasting model. Producers sell their forecasted amount of energy (V_p) one day ahead against a hourly varying market spot price called the APX (P_a). When a producer knows its allocated amount of energy (V_a) produced there exists a difference ΔV between the forecasted and allocated amount of energy, see also equation 1. This difference is also called the imbalance.

$$\Delta V = V_a - V_p \tag{1}$$

A negative ΔV means a producer or TSO being *short* ($V_a < V_p$) and a positive ΔV means a producer or TSO being *long* ($V_a > V_p$)

Furthermore, a producer can be a ‘causer’ or a ‘helper’. When the electrical grid is long then all the producers being long are the ‘causers’ and all the producers being short are the ‘helpers’. When the electrical grid is short then the producers being short are the ‘causers’ and the producers being long are the ‘helpers’ [13]. Each fifteen minutes the TSO determines a price (P_t) called the imbalance price which can be positive or negative, and differs from the spot price (APX). This price is based on the production, consumption and the regulation of electrical grid. The ΔV (imbalance) of a producer will be sold against this price (P_t).

The profit and loss of a producer depends on the ΔV, P_t and P_a. Therefore three scenarios can be sketched. The first scenario discusses $P_t > 0$ and $P_t < P_a$, the second scenario where $P_t > 0$ and $P_t > P_a$, and the third scenario where

$P_t < 0$. For each scenario an example will be given using the characteristics of ΔV, P_t and P_a and the following terms: causer, helper, short and long.

Furthermore, assume for each scenario that the spot price is 40 €/MWh and assume the following predicted production and consumption values:

	V_p	**Sold against P_a (40€/MWh)**
Producer A	300 MWh	€12000
Producer B and C	700 MWh	€28000
Consumption	1000 MWh	-

Table 1: Predicted production and consumption values

The forecasted amounts of wind power produced have been sold against the spot price of 40€/MWh.

An imbalance price determined by the TSO can be positive or negative. When the price is positive the TSO will pay the producers and when the price is negative the producers will pay the TSO.

Finally, the prices of the consumption have not been included in the scenarios since this is beyond of the scope of this research.

3.2.0.1 Scenario 1 - $P_t > 0$ and $P_t < P_a$:

This scenario outlines a positive imbalance price ($P_t > 0$) and is smaller than the spot price ($P_t < P_a$). The following results have been obtained after knowing the actual production and consumption values:

	V_p	V_a	ΔV	**Long / Short**	**Helper / Causer**
Producer A	300 MWh	400 MWh	+100 MWh	Long	Helper
Producer B and C	700 MWh	500 MWh	-200 MWh	Short	Causer
Consumption	1000 MWh	1000 MWh	0 MWh	-	-
TSO (totals)	0 MWh	-100 MWh	-100 MWh	Short	-

Table 2: Results for scenario 1

From table 2 one can see that there is an under production. Producer A is a helper and cannot ensure the balance between demand and supply. This means that the missing production needs to be produced by for example power plants using fossil fuels. The TSO has to pay money to regulate the grid balancing the production and the consumption. The price P_t will be €10 per MWh. Based on this price the cost can be calculated for the producers which are shown in table 3.

As one can see in table 3 the TSO has paid producer A for his over production. Therefore producer A has made a profit of €1000. However if producer A had a more accurate forecast he would have sold his production against the spot price, which would have resulted in a profit of €4000. This scenario shows that a more accurate forecast would have been fortunate.

	ΔV	**Sold against** P_t	**When sold against** P_a
Producer A	+100 MWh	€1000	€4000
Producer B and C	-200 MWh	€-2000	€8000

Table 3: Costs calculated using table 2

The producers B and C have to pay the TSO €2000. However producers B and C still have made a profit of €6000 since they have sold their predicted energy for €8000. In this case a wrong forecast was not unfortunate. Even though the costs from the TSO are passed to the PR which passes the costs to the producers. From the perspective of the TSO a wrong forecast is unfortunate since they have to regulate the grid.

3.2.0.2 Scenario 2 - $P_t > 0$ and $P_t > P_a$:

This scenario outlines a positive imbalance price ($P_t > 0$) and is larger than the spot price ($P_t > P_a$). For this scenario the results from table 2 are being used. The missing production needs to be produced by for example power plants using fossil fuels. The TSO has to pay money to regulate the grid balancing the production and the consumption. Instead of having a price of €10 per MWh the price is now €50 per MWh, resulting into the following costs shown in table 4.

	ΔV	**Sold against** P_t	**When sold against** P_a
Producer A	+100 MWh	€5000	€4000
Producer B and C	-200 MWh	€-10.000	€8000

Table 4: Cost calculated based using table 2

As one can see in table 4 the TSO has paid producer A for his over production. For this over production producer A has received €5000. In this case the wind power forecast of producer A was fortunate, because if producer A had a more accurate forecast he would have sold his production against the spot price, which would result in a profit of €4000.

The producers B and C have to pay the TSO €10.000 for their under production. Since they have sold their production against the spot price for €8000 they have made a loss of €2000. In this case their forecast was unfortunate.

3.2.0.3 Scenario 3 - $P_t < 0$:

This scenario outlines a negative price ($P_t < 0$) determined by the TSO. The following results have been obtained after knowing the actual production and consumption values:

From table 5 one can see that there is an over production. Producers B and C are helpers but cannot ensure the balance between demand and supply. This means that the over production needs to be removed by for example transmitting energy to other TSOs. The TSO has to regulate the grid to balance the production and the consumption which costs money. Therefore the TSO price P_t will be -€10 per MWh.

Based on this price the following cost can be calculated:

As one can see in table 6 producer A has paid the TSO €3000 for his over production. In this case the wind power forecast of producer A was unfortunate,

	V_p	V_a	ΔV	**Long / Short**	**Helper / Causer**
Producer A	300 MWh	600 MWh	+300 MWh	Long	Causer
Producer B and C	700 MWh	600 MWh	-100 MWh	Short	Helper
Consumption	1000 MWh	1000 MWh	0 MWh	-	-
TSO (totals)	0 MWh	+200 MWh	+200 MWh	Long	-

Table 5: Results for scenario 3

	ΔV	**Sold against P_t**	**When sold against P_a**
Producer A	+300 MWh	-€3000	€12000
Producer B and C	-100 MWh	€1000	€4000

Table 6: Cost calculated based using table 5

because if producer A had a more accurate forecast he would have reduced his loss.

Producers B and C are being paid by the TSO for their under production. Furthermore, since they have sold their predicted production also against the spot price for €4000 they have made a total profit of €5000.

3.3 Discussion

For the TSO it is satisfying if the forecast of wind power is as accurate as possible, because this way the regulation of the net is reduced and therefore the costs are reduced. Since the TSO is a non-commercial company they do not profit from regulation of the net. The costs are passed to the PR which passes it to the producers. The TSO is therefore an independent company.

From the perspective of the producer we want to make clear that a helper always receives money and a causer has to pay money. However, at the moment of forecasting wind power, it is unknown for a producer if he is a helper or a causer. The reason is because a producer does not know what the imbalance price will be; positive or negative, and how this price is compared to the spot price. A producer does not know the consumption of energy, and what other producers predict to produce. Therefore it is a risk one takes when selling or buying energy against the imbalance price, because the imbalance price is dependent on real time units of production and consumption. To avoid this risk, an accurate forecast of energy production is needed.

There is also the fact of gambling. Basically using forecast wind power generated by turbines is gambling with a limited certainty. This certainty has been obtained by the production of wind power based on specific circumstances such as the weather. Of course a producer can adjust its forecast before selling it against the spot price. But the uncertainty of being a helper or a causer still remains unless you are a lucky gambler. Since a company is dealing with large amounts of money (tons to millions) it is not recommended to gamble with forecasts, but rather use the forecasts which gives some certainty.

4 Introduction to forecasting models

In our research we apply two forecasting models to forecast wind power generated by wind turbines. Since the main goal is to build a forecasting model which outperforms the current forecast Raedthuys is using, we will use forecasting models which have been recommended by literature the most. In this section we discuss two forecasting models. The first model is a random forest and the second one a feed forward neural network, two recommended techniques to predict wind power. The following subsection explains the process of the forecasting models in more detail.

4.1 Random forest model

According to Breiman [3] a Random Forest is a collection of tree-structured classifiers. The trees are random vector sampled independently and are identically distributed. They cast a vote for the most popular class at input x. To understand the idea behind Random Forest we will explain the process of the algorithm.

4.1.1 Process of the Random Forest

The process of the Random Forest works as follows [29]. Assume we have a dataset D containing n samples. Each sample has a vector X of input variables $x_1, x_2, \ldots, x_n$ and one output variable y_1.

Step 1: First T number of trees has to be defined.

Step 2: Then draw T bootstrap samples of size n from the original training dataset. We mean by bootstrap samples the following: Each time randomly a sample is taken from the dataset. The sample is not removed but remains in the dataset. After selecting n samples it might occur that there are duplicates in the dataset or that there are samples missing which do exist in the original dataset.

Step 3: For each of the bootstrap samples a regression tree is build. For each node select randomly m variables from the X variables, this is also called Bagging [12]. Pick the best split among all the predictors in m. This is done recursively for each node.

Step 4: After creating all the trees new data can be predicted. The prediction of the new data is performed by the aggregation of the predictions of the T trees. In the case of regression the average is taken from all the predictions [12][29], see equation 2. Here $\hat{Y}_t(x)$ is the predicted outcome of tree t for observation x.

$$\hat{Y} = \frac{1}{T} \sum_{t=1}^{T} \hat{Y}_t(x) \tag{2}$$

To obtain the lowest error rate we have to know the number of trees occurring this error. To determine this number of trees T we increase the amount of trees each time by ten up to 500. Each time a random forest has been generated we estimate the error rate based on the validation set. The validation set has not been included in the bootstrap sample, this is according to Breiman also called the "out-of-bag"(OOB), data. To obtain the error rate percentage we apply the evaluation metrics RMSD and the MAE since these have been applied most by previous research.

After knowing the amount of trees needed to obtain the minimum error on the validation set we can predict our testing samples. According to Liaw and Wiener [29] have found that random forest performs very well compared to other forecasting techniques such as, neural networks or support vector machines. Furthermore, according to Fugon et al. [12] and Liaw and Wiener [29] random forest is robust against overfitting.

4.2 Feed forward Neural Network

Different types of neural networks can be used to forecast wind power. In our research we design a feed forward neural network (FNN).

4.2.1 Process of the feed forward neural network

The basic structure of a neural network is that it is an ensemble of neurons connected to levels called layers. This structure is based on the human brain [12]. In this section we will explain the process of the FNN used in this research.

In figure 3 is the structure given of the feed forward neural network. The neural network contains three different layers, called the input layer, hidden layer(s) (optional) and the output layer. The neural network is completely connected. Every node in a layer is connected with every node in the next layer, but the nodes are not connected among each other in the same layer. Each connection between two nodes contains a weight w_{ij} (i is the node, j is the layer) [26].

The input layer corresponds with the input variables x_i, in our research these are the weather variables. Each neuron in this layer represents a variable. The neurons from the input layer are connected with the hidden layer and are each affected by a weight w_{ij}. The input of a hidden layer is a weighted linear combination of the output of the neurons from the previous layer[12]. This linear combination is a summation of the inputs, see equation 3[26]. The output of hidden or output layer is a transformation of the weighted linear combination based on a specific transfer/activation function. The most used transfer function is the sigmoid function, see equation 4. In this function y is the weighted linear combination, see equation 3. The output of the hidden layer is affected by a weight and passes to the input of the next hidden layer or the output layer. The feed forward neural network uses in the output layer a linear regression function as transfer function to create its final output. More information about neural networks can be found in [26].

$$y = \sum_{i=1}^{n} x_i w_{ij} \tag{3}$$

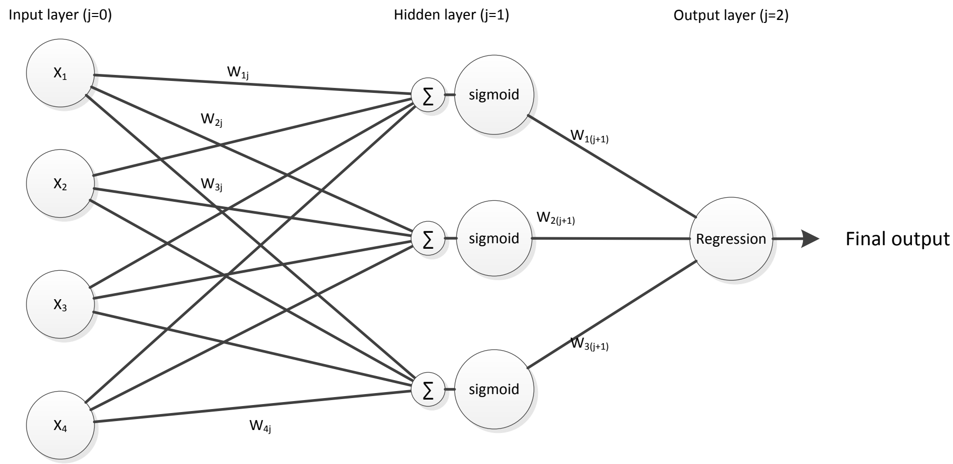


Figure 3: Feed forward neural network

$$S(y) = \frac{1}{1 + \exp^{-y}} \tag{4}$$

4.2.2 Learning the algorithm

The neural network learns based on a back-propagation algorithm. The basic idea is to adjust the different neuron weights by back-propagating the error between the predicted output and the actual output. In this research adjusting the weights is conducted applying a training function called the Levenberg-Marquardt back-propagation algorithm [31]. This training function minimizes the error to a local minimum.

4.2.3 Hidden neurons and layers

According to Fugon et al. [12] the choice of the number of hidden neurons and layers is important, because a high number of neurons creates complex relations in the model between inputs and outputs and this can lead to overfitting of the data. Trial and error will be applied in this research to obtain the optimal amount of hidden neurons. According to [26] one hidden layer is sufficient for most purposes. Therefore in this research we will use one hidden layer.

5 Related work

This section provides a literature overview on the subject of this research 'Forecasting of wind power production generated by turbines'. The goal of this research is to find the answers on the first research questions stated in the introduction (section 1). We identify which factors and input parameters have been used in previous literature to forecast wind power. Furthermore, we are identifying which forecasting models have been found the most relevant.

The first subsection 5.1 describes which factors have been used by previous literature. The second subsection 5.2 describes which input parameters have been found useful to predict wind power. The final subsection 5.3 describes which forecasting techniques have been used by previous literature to predict wind power.

5.1 Important factors for forecasting wind power

To forecast wind power generated by turbines we have to find out which factors have influence on the forecast of wind power. Based on a literature study we have found three factors which have been found useful to predict wind power. The first factor is the use of different data sources (section 5.1.1), the second factor is the prediction of wind power on different grid areas (section 5.1.2) and the third factor is taking into account the geographical location (section 5.1.3). For each of the factors we describe its importance and relevance for forecasting wind power generated by turbines.

5.1.1 Data sources

Literature has shown that different data sources can be used to predict wind power generated by turbines. Each of these data sources have shown to be useful to predict wind power and therefore we will describe for each data source its use.

Firstly, a large amount of previous research has been done on meteorological data such as HiRLAM and ECMWF [10][36][41][42][44]. Meteorological data is important for day-ahead forecasts since they are covering a horizon of 48 to 72 hours ahead [42]. The data are numerical weather predictions (NWP), measured at 10 meter height which describe the condition of the atmosphere, including important information like wind speed, wind direction, temperature etc. According to Pinson and Kariniotakis [36] and Sideratos and Hatziargyriou [42] NWPs are indispensable for an acceptable performance on short term and long term forecast and their accuracy contributes to the accuracy of wind power predictions.

Secondly, weather stations surrounding the wind turbine or farm [10] have been used as data source to obtain weather data observations. The advantage of using these weather stations is that they provide weather data in a local area near the wind turbine.

Thirdly, the online supervisory control and data acquisition (SCADA) system has been used to obtain data. The SCADA system can provide measurements of wind power, wind speed, wind direction and other variables on a real-time basis every minute [42]. The data provided by the SCADA system is measured at the location of the wind turbine or farm and provides the actual operational status. This makes SCADA data valuable since it describes the actual performance of the wind turbine [36]. SCADA data can therefore be used to map the meteorological or weather station data to the state of the turbine and can be used as training data for the prediction model.

Finally, to obtain weather observations at a specific location a Laser Imaging Detection and Ranging (LIDAR) can be used Wagner et al. [47]. A LIDAR measures the weather conditions at different heights using a laser on real-time basis. It can be used to decide placing a turbine in a certain area by measuring the wind profile of that area. An advantage of a LIDAR compared to the SCADA system is that a LIDAR can measure weather conditions on various heights, up to 200 meters, while a SCADA system measures only on the hub height, the height of the turbine rotor.

The combination of different data sources (weather stations and meteorological data) can be mapped to the SCADA data or LIDAR data to create specialized local models for wind power production in specific turbine locations, which might help to improve the prediction of wind power.

5.1.2 Grid area

The prediction of wind power has been applied on different sizes of grid areas. A large amount of research has been focusing on forecasting wind farm production (e.g. [30], [36], [42]). A wind farm is a group of turbines located in the same area producing wind power. However H. Holttinen and Sillanpaa [14] showed that the aggregation of areas lowers relative share of prediction errors. Their prediction model lowered the prediction error of wind power up to 60%. This result has been obtained when comparing the mean average error (MAE) of 52%-56% from a single turbine with the aggregation of three areas of about 20%. Also Brand et al. [2] and Focken et al. [11] have found that aggregation of wind power improves the quality of the forecast. According to Focken et al. [11] integrating over an extended area, weakly correlated errors underlying prediction and measurement cancel out partly due to statistical effects. This results into a reduced prediction error for an area compared to a single turbine.

The size of a aggregation area proposed by H. Holttinen and Sillanpaa [14] is roughly the size of the entire Netherlands. Therefore one aggregation grid area of the Netherlands could decrease the prediction error of wind power. Another grid area which might be applicable is the aggregation on province level.

5.1.3 Geographical location

In many different countries research has been conducted on the prediction of wind power. Because models have been proposed in different countries makes it difficult to evaluate the performance of models [27]. However according to Wang et al. [49], research has been conducted comparing 11 models (which models is unclear), running the same forecasting case. The models were evaluated based on six test cases in Spain, Germany, Denmark and Ireland using the same numerical weather predictions (NWP) as input. Numerical means that each data value is represented as a number.

The results have shown that no forecasting model can perform perfect in any condition, no model was the best in all the cases. Furthermore, the results show that the forecasting accuracy gets worse in complex terrain.

To find out which forecasting model performs the best in the Netherlands, models proposed in countries with similar topological characteristics should be considered. Therefore models proposed in other surrounding European countries, like United Kingdom, France, Germany or Denmark might be useful for further research.

5.2 Input parameters for forecasting models

Forecasting wind power is performed by applying forecasting models. These forecasting models need input data to predict wind power generated by turbines. The input parameters which have been found a successful predictor of wind power by literature are taken into account. The input parameters have been selected based on correlation studies reported in literature. This section discusses the input parameters by explaining its importance and relevance.

5.2.1 Wind speed

Wind speed is the most used input parameter to predict wind power generated by turbines. Literature has used average values of wind speed, such as hourly average wind speed [41][10] or monthly average wind speed [30]. To predict wind power for a certain moment in time previous values of wind power and wind speed have been used as input parameters. For example Senjyu et al. [41] have used wind speed predictions on several-hour-ahead, such as data of every six hours interval wind speed has been used for the prediction of six hours ahead and data of one day interval wind speed has been used for the prediction of one day ahead. The number of lagged hours or days required to predict wind power accurately has been determined by performing an autocorrelation and cross-correlation analysis between different variables [10].

A well known formula transforming wind speed into wind power is given in equation 5 [30][41]. The A (m^2) is the sweep area of the blades. The ρ is air density (kg/m^3) and the V is wind speed (m/s). The air density can be calculated as a function of the temperature and pressure.

$$P = \frac{1}{2} A \rho V^3 \tag{5}$$

In equation 5 one can see that wind power output is proportional to the cube of the wind speed [41]. Therefore a method is required to predict wind speed as accurate as possible, because the error between the predicted and actual wind power value is also proportional to the cube of the error of the predicted and actual wind speed.

Why literature rather use forecasting models to predict wind power rather than wind speed is because of several possible reasons.

Firstly, equation 5 can be applied to forecast wind power for one specific turbine. However the equation has problems to deal with the total wind power output generated from a wind farm. A wind farm is a group of turbines located in the same area producing wind power. It is possible to calculate the wind power generated by one turbine using the predicted wind speed and multiply it by the number of turbines in the wind farm, but this would result in a larger forecasting error because important details have not been taken into account, such as shadowing effects or wake effects caused by other turbines [23]. Wakes are invisible ripples and waves in the atmosphere that can damage turbines and decrease efficiency [22].

Secondly, equation 5 uses a wind speed value measured at one height and therefore does not take into account the wind speed profile, which is the relation between wind speed values on different heights. Wagner et al. [47] state that it is

[1] `http://www.wind-power-program.com/turbine_characteristics.htm`

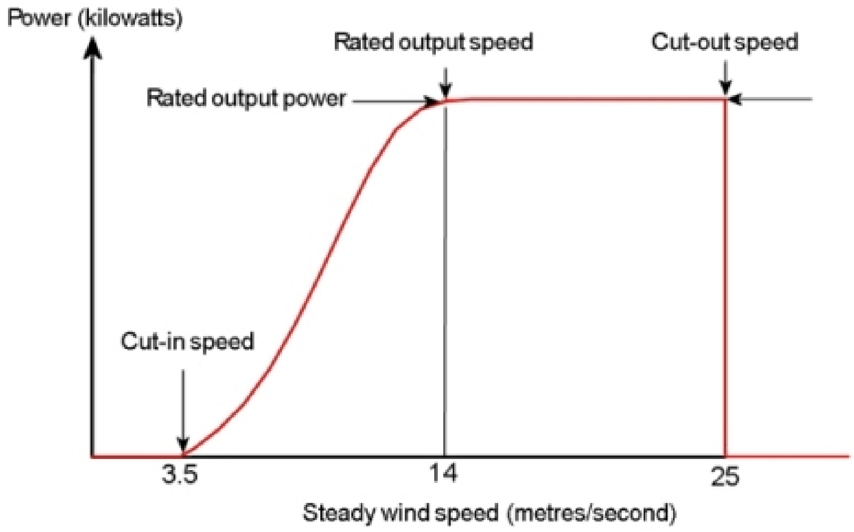


Figure 4: Example of a power curve: turbine power output with steady wind speed[1]

common to assume that the wind speed profile is continuous based on the wind speed measurements at hub height, which is the height of the turbine rotor. Therefore to predict wind power generated by turbines the wind speed values at hub height are used.

A problem with using the wind speed at hub height is that it ignores the wind speed shear, which is the change of wind speed between two heights. Wagner et al. [47] have found ignoring the wind speed shear could lead to a misinterpretation of the power performance of the turbine. They have found this result based on measurements using a Laser Imaging Detection and Ranging (LIDAR). A LIDAR can measure the weather conditions at certain heights using a laser. Based on the measurements obtained by the LIDAR they have identified a wind speed profile taking into account the wind speed shear which is different from the wind speed profile ignoring the wind speed shear (only using the wind speed at hub height). This resulted into two different power curves. A power curve shows the relation between the wind speed and the wind power output, see also figure 4 [17][30][44].

The wind speed profiles have been used for deriving an equivalent wind speed, which resulted into a reduction of the scatter in the power curve and therefore into in an improve of the power performance measurement.

5.2.2 Wind direction

Kulkarni et al. [19] have used wind direction; recorded by a cup anemometer at a level of 100 meters, as an input parameter to predict wind speed. Barbounis and Theocharis [1] have obtained wind speed and wind direction from four nodes surrounding the wind farm. The nodes are located north, south, east and west from the wind farm. Based on their correlation studies between the wind direction and the real wind power generated by the turbines they have found that the nodes did not all have an adequate level of correlation. Using the irrelevant nodes could result in a poor performance of wind power forecast. Sideratos and Hatziargyriou [42] have found that when the forecast wind direction is between defined limits, NPWs can be considered as relatively accurate. Their reason is because strong winds are difficult to predict but come from known directions due to the topography of the area where the wind farm is located.

5.2.3 Weather stability

Weather stability influences the accuracy of the forecast of among others numerical weather prediction (NWP) models. Numerical means that each data value is represented as a number. According to Pinson and Kariniotakis [36] an unstable atmospheric, such as unstable pressure, temperature and/or relative humidity can lead to poor numerical weather predictions and the other way around. To evaluate the global atmospheric situation they have defined a unique representative index for the following N_h hours, called the Meteo-Risk index (MR-index). The MR-index measures the spread of the weather forecasts at a given time. The most recent forecast is used as a reference and reflects the variability of the older forecast [37]. Low MRI-index indicate there is a stable atmosphere and high MRI-index indicate there is an unstable atmosphere. Pinson and Kariniotakis [37] have calculated the MRI index on a horizon of 24 hours. Since they have used HiRLAM data which provides data every six hours, they have used four sets of wind speed predictions. Plotting the distribution of MRI values against the prediction error they have found that the prediction error increases linearly with the MRI values. Based on this linear relationship they have made the following empirical relation:

$$e = e_0 + s\text{MRI} \tag{6}$$

The first part of the right side of equation 6 is the basic part of the error, e_0, this is the point where the error line crosses the y-axis, the second part is a direct consequence of the prediction model sensibility to the weather stability. The sensibility s is the slope of the linear fitting model. Using this equation a scale factor can be defined for the confidence interval depending on the value of the MRI. The scale factor can be used to enlarge or narrow the interval width for a number of hours N_h. Based on their findings Pinson and Kariniotakis [37] have defined rules concerning the expected prediction error depending on the MRI values. They have binned the data by MRI values and calculated the cumulative distribution function of the prediction errors for each bin. The results given by this function give the probability with which an error larger than a defined threshold occurs. Based on the defined rules and the results from the distribution function permits one to derive signals that large prediction errors might occur.

5.2.4 Availability

According to Mabel and Fernandez [30] the amount of wind power generated by a turbine is dependent on the amount of generation hours of the turbine. They have found a correlation of 0.7 between the monthly generation hours and the monthly wind power output. The availability of a turbine depends on factors such as mechanical break down or scheduled maintenance. Therefore it is important to plan these factors at moments when wind power prediction is low. This means that the availability of the turbine is very important for the production of wind power, which is also addressed by Mohandes et al. [33]. Mabel and Fernandez [30] calculate generation hours as follows: generation hour = (total numbers of hours in a month) − (low wind hours + wind turbine maintenance hours + turbine breakdown hours + grid maintenance hours + grid breakdown hours).

For hourly wind power prediction hourly values of availability are required. In case of a single turbine one can correct the predicted wind power output with the availability of the turbine afterwards. But when predicting the total wind power output generated by all the turbines it is important which turbine is active and which one is not. Some turbines produce more wind power than other turbines, because they are larger and have a larger capacity to produce wind power or because wind turbines are located in a wind farm. This means larger turbines have more influence and turbines located in wind farms have less influence on the total produced amount of wind power, because of shadowing effects. Therefore it is important to take each turbine its availability into account.

5.2.5 Relative Humidity

Relative humidity of air is dependent on the amount of water vapor in the air, which affects the air density[30]. Mabel and Fernandez [30] and Park et al. [35] have found that relative humidity improves forecasting models to predict wind power. Mabel and Fernandez [30] have found that relative humidity has a dependence on wind power output with a correlation value of 0.4. Furthermore, the monthly variation of relative humidity through the year lays between the 60% and the 90%. Based on these results they have found it important to include relative humidity as one of the input parameters for the prediction of wind power.

5.2.6 Seasonal

Kwon [21] has investigated if there is a seasonal effect in the data. Therefore they have segmented the data in three month sets. The four seasons reveal significant variations of the average error percent. They have found that during the winter the annual production of wind power is relatively high among the four seasons . Furthermore, they have found that the summer exhibits a low wind regime. Also Taylor et al. [44] has found seasonality in their dataset. They have plotted the time against the wind speed and based on this plot one can see that the wind speed is low in the summer months compared to the winter months. Finally, Mohandes et al. [33] have identified a seasonal effect between wind speed and wind power, which they have used for their time series model.

Including seasonal effects as input parameters helps the forecasting model to understand in what kind of weather circumstances the turbines are. This could decrease the prediction error per season. Another way to include seasonal effects is to design different models for different seasons.

5.2.7 Temperature and pressure

The temperature and pressure have influence on the wind power, since both have influence on the air density. However Kulkarni et al. [19] did not use the feature temperature because the complicated influence of the temperature on wind power would make the selection of a function for a regression model and its fitting difficult. Furthermore, Mabel and Fernandez [30] and Fan et al. [10] have found a low correlation between the temperature and the wind power output and between the pressure and the wind power output. Adding these features did not improve the performance of the models and slowed down the learning process. Therefore these two features have been left out of the models.

5.3 Forecasting models

Forecasting wind power can be performed for different time horizons (e.g. minutes, hours, days, months). A lot of literature studies have been found in the field of forecasting wind power or wind speed. Literature studies discussing forecasting models for day-ahead (24 to 48 hours) or longer forecast have been analysed. The reason discussing day-ahead forecasting has been explained in the introduction 1. Different types of forecasting models have been identified by literature:

Persistence model: This model is also called Naive predictor model. The wind speed at time $t + \Delta t$ will be the same as it was at time t. [43]

Physical models: These models are using a detailed description of the atmosphere, topological information and characteristics of the wind turbines. The description of the atmosphere are numerical weather predictions (NWP) given by a weather service (HiRLAM or ECMWF etc.) and contain information, such as hourly average wind speed, pressure, temperature and relative humidity. The topological information contains data about the surroundings of the turbine such as obstacles, roughness and orography. Characteristics of the wind turbine are for instance the height of the rotor of the turbine (hub height) or its location in a wind turbine park. In this research the focus is not on the physical models and therefore we will not discuss these. However more detailed information can be found in references [24][25][23],

Statistical models: These models such as artificial neural network or regression trees, are based on using a training dataset containing historical measurement data. For example to predict wind power, historical measurements of weather data is needed such as wind speed, wind direction, etc.

In this research we are going to focus on the prediction of wind power using machine learning techniques. Therefore we are discussing only statistical models. The discussion about physical models and persistence models are beyond the scope of this research.

For each of the statistical models we discuss what kind of information has been applied to predict wind power, what the advantages and drawbacks of the model are, in what country the model was used and what prediction performance compared to the actual data was.

5.3.1 Statistical models

Many different statistical forecasting models to forecast wind power have been proposed in the literature. To answer the research questions 'Which forecasting models have been proposed in literature?' and 'What are the advantages and drawbacks of the proposed models?' we discuss these proposed statistical forecasting models in detail. Each statistical forecasting model is discussed in a separate subsection and in tables 7, 8 or 9. The subsections and the tables provide the following detailed information: advantages and disadvantages of the model, the used dataset, the input parameters, the output parameter(s) and the country in which the model has been performed.

The statistical forecasting models are discussed in the following order: regression trees, time series models, recurrent neural networks (RNN), feed forward neural networks, fuzzy-neural networks (FNN) and support vector regression/machine (SVR/SVM).

5.3.1.1 Regression trees

Clifton et al. [7] have performed their research in the USA using the machine learning technique 'regression trees, see also table 7. They used this technique because it is a technique that can capture non-linear changes. Regression trees are models using a branching structure based on attribute-value pairs to predict an outcome, they are quick to run and easy to update.

Clifton et al. [7] have used the mean value predicted by an ensemble of 100 regression trees to increase the accuracy of the power predictions. Subsets of the training data are used to generate different trees and the predicted power output is calculated by each tree using the same input data. The input data is generated by a turbulence simulator, TurbSIM which creates wind fields. 1796 10 minute wind fields have been created. A wind field consist of a random combination of hub height wind speed between the 3 and 25 m/s (height of the turbine rotor), the hub height turbulence intensity (from 5% to 45%) and the wind shear exponent (α) (from -0.5 to 0.5). The wind shear exponent is a value which transforms the wind speed (v_0) at the height (h_0) to a wind speed (v) at height (h), see also equation 7. This equation is also called the power law [17]. Besides these three parameters also the operating region has been chosen as input parameter. The output is the prediction of 10 minute mean power values.

$$\frac{v}{v_0} = (\frac{h}{h_0})^{\alpha} \tag{7}$$

The results show that the ensemble regression tree method predicts two to three times more accurate than the traditional power curve. A power curve is the relation between the wind speed and the output power of a turbine, which is also shown in section 5.2.1.

5.3.1.2 Time series models

Time series models such as Autoregressive (AR) or Autoregressive Integrated Moving Average (ARIMA) are intended to investigate trends, seasonal patterns [4] etc. Comparing Artificial Neural Network (ANN) and AR models on computational complexity it has been found that ANN have a higher complexity compared to AR models [33]. Kavasseri and Seetharaman [17] have proposed the use of a fractional-ARIMA model. A unique ability of this model is to capture time series measurements in the presence of correlations [17]. They have forecast hourly average wind speed up to 48 hours ahead. Table 7 provides information about the dataset, input- and output parameters. Using the power curve they have found the corresponding wind power value. They have found an improving accuracy of 42% compared to the persistence model. More information about the model can be found in reference [17].

Ref.	Forecasting model	Dataset	Input parameters	Output parameters	Country
[7]	Regression tree	1796 10 Minute wind field values	Hub height wind speed Turbulence intensity Wind shear Operating region	wind power 10 minutes ahead	United States
[17]	fractional-ARIMA	Four weeks of wind speed data obtained from four wind generation locations	Average wind speed (granularity unkown)	Hourly average wind speed values day-ahead (24 hours) and two day-ahead (48 hours)	United States

Table 7: Overview of forecasting models: regression trees and fractional-ARIMA

5.3.1.3 Artificial Neural networks (ANN)

Different types of artificial neural networks (ANN) have been discussed in the literature. Table 8 gives a detailed overview of literature studies which have used ANN to forecast wind speed or wind power. ANN have the ability to learn from experience [8][33], to handle noisy, incomplete or corrupted data[5] and are able to deal with non-linear problems [30]. Furthermore, based on historical data an input-output mapping is constructed [33]. In this section we will discuss the following neural networks: recurrent neural network (RNN), feed forward neural network (FFNN) and fuzzy neural network (FNN).

5.3.1.3.1 Recurrent neural networks (RNN)

Senjyu et al. [41] compared different on-line learning algorithms for training recurrent neural networks in Japan (table 8). They have compared a feed forward neural network with a recurrent neural network (RNN). The difference between those two neural networks is that feed forward networks travel one way from the input to the output, whereas recurrent neural networks can travel both directions, meaning they can use their output value as an input value for predicting the next output value. The virtue of recurrent neural networks compared to feed forward networks is that they can establish a dynamic mapping relating input to output sequences [41].

Senjyu et al. [41] have predicted wind speed on several-hour-ahead, using a 3-hour interval dataset of one year containing average wind speed values. For the prediction of wind speed for six hours ahead they have used the data of every six hours. For the prediction of wind speed for nine hours ahead they have used the data of every nine hours etc. For each x-hour ahead forecasting they have found that the RNN performs better than the feed forward neural network.

Barbounis and Theocharis [1] have also forecasted wind power and wind speed values using a recurrent neural network. For more details see table 8. They have found that RNN provide better forecasts compared to the persistent method, the atmospheric and time-series models.

5.3.1.3.2 Feed forward neural network (ANN)

Mohandes et al. [33] have compared a feed forward neural network versus a time series autoregressive model (AR) in Saudi Arabia. The results show that the feed forward neural network outperforms the AR model. Other research conducted by Kulkarni et al. [19] have compared four different statistical techniques: curve fitting, Auto Regressive Integrated Moving Average Model (ARIMA), extrapolation with periodic function and feed forward artificial neural network. They have found that wind speed can be predicted based on previous knowledge of wind speed and the local time of the day. Kulkarni et al. [19] have found a periodic behaviour between both parameters by plotting the wind speed of five successive days against the hours of the day. The extrapolation with periodic function and the feed forward neural network performed the best compared to the other two models.

5.3.1.3.3 Fuzzy-neural network (FNN)

Pinson and Kariniotakis [36] have proposed a fuzzy neural network to forecast wind power. This neural network has the ability to adapt and to fine-tune its parameters during on-line operation. Advantages of a FNN compared to the physical models are that it avoids temporal correlations between past and future data, conversion of wind speed from the height of measurement to the hub height of wind turbines, spatial projection of the meteorological wind speed forecast from the NWP to the level of the wind farm and correction of the wind park output for factors affecting the total production [36].

5.3.1.4 Support vector regression/machine

Fan et al. [10] have proposed a hybrid forecasting model. The model is a combination of a Bayesian clustering and support vector regression. They mention that the model is well suited for capturing the dynamics of wind generation / wind speed time series, has strong robustness and can easily by modified for different wind farms. The Bayesian clustering has the ability to partition the input training dataset into subsets and the SVR is a technique for data regression based on recent advances. SVRs have been found very resistant to the overfitting problem [10]. Other research conducted by Mohandes et al. [32] and SangitaB and Deshmukh [40] have used a Support Vector Machine (SVM), with a gaussian and RBF kernel respectively. The idea behind a SVM is to map the data into a high dimensional feature space through a non-linear mapping. After the mapping a linear regression is done in this feature space [40]. A disadvantage of SVM addressed by Cao and Tay [5] is that the training time is somewhere between quadratic and cubic compared to the number of training samples. Therefore SVM have a large amount of computation time for solving large-size problems.

Ref.	Forecasting model	Dataset	Input parameters	Output parameters	Country
[19]	Extrapolation with periodic function and Artificial Neural Networks (ANN)	Ten years of data (1992-2002) wind speed of previous hours recorded by a cup anemometer at 100 meter height	Average hourly wind speed	Hourly wind speed up to 48 hours ahead	India
[1]	Recurrent multi-layer network	Wind predictions provided from four nodes nearby the wind park	Hourly average wind speed and wind direction	Hourly wind speed from 1 to 72 hours ahead	Greece
[41]	Recurrent neural network	Ground observation data from Japan Meteorological Business Support Center from one year	Average wind speed values	Forecasting wind speed 3,6 and 9 hours ahead, and 1, 2 and 3 day-ahead	Japan
[33]	Artificial neural network	Mean monthly wind speed from 1970 to 1982 and daily wind speed from 1970	Mean monthly and daily wind speed	Mean monthly wind speed prediction and mean daily wind speed prediction	Saudi Arabia
[30]	Feed forward neural network	Field data collected from seven wind farms over a period of 3 years (2002-2005)	Monthly average wind speed, relative humidity and generation hours	Monthly average wind power	India
[8]	Artificial neural network with fuzzy set-based classification	Hourly historical weather data (size unknown)	Day of week, hour of day, load shape, temperature, humidity, wind speed	wind power prediction from 1 to 120 hours ahead	United States
[42]	Artificial neural network with fuzzy logic	Dataset power measurements and meteorological forecasts	Wind speed, Wind direction	wind power forecasting from 1 to 48 hours ahead	Greece
[4]	Artificial neural network and Autoregressive Integrated Moving average (ARIMA)	7 Years of wind speed measurements	Monthly average wind speed values	Monthly wind speed and wind power	Mexico
[36]	Fuzzy neural network (FNN)	Online SCADA measurements and NWP data	Unclear which input parameters	wind power prediction 48 hours ahead	Ireland

Table 8: Overview of forecasting models with its input and output features

5.3.1.5 Discussion

As one can see literature have applied different statistical models on the

Ref.	Forecasting model	Dataset	Input parameters	Output parameters	Country
[10]	Bayesian clustering with Support vector regression	10 minutes data from the wind farm, observations from a meteorological tower, hourly observations from surrounding weather stations and hourly meteorological forecast at the wind farm its location	Wind speed, wind direction, humidity	Hourly wind power generation from 1 to 48 hour ahead	United States
[32]	Support vector machine vs. Multilayer perceptron	Daily wind speed data of 12 years between 1970 and 1982	Mean daily wind speed data	Mean daily wind speed data	Saudi Arabia

Table 9: Overview of forecasting models with its input and output features

forecast of wind power generated by turbines. It is hard to compare forecasting techniques and determine which one works best, because the previous research have analysed different techniques on different datasets in different countries.

Therefore in our research we compare two forecasting models, feed forward neural network and a random forest which have been explained in section 4. Based on our analysis on the forecasting models we have found that different types of neural networks are good predictors on long term forecasting and according to Mohandes et al. [33] and Barbounis and Theocharis [1] neural networks perform better than time series. Furthermore, Barbounis and Theocharis [1] have found that Recurrent Neural Networks performs better than the Feed forward neural network.

The main advantages using neural networks are because neural networks have the ability to learn from experience, to handle noisy and incomplete data. Furthermore, maybe one of the most important advantages is neural networks can handle non-linear problems, because the prediction of wind power is a non-linear problem. A drawback of neural networks is that they have a high computational complexity.

Other forecasting models such as SVM or Recurrent neural network have not been taken into account to use as a forecasting models in this research. The main reason is because lack of time.

5.4 Feature selection methods

To identify the right input parameters different selection methods have been applied in the literature. Fan et al. [10] have selected their features using the correlation function shown in equation 8. In this equation the $Cov(g, s)$ is the covariance of generation g and wind speed s and σ_g and σ_s are the standard deviations for g and s. In time series models proposed by Kavasseri and Seetharaman [17], Mohandes et al. [33] and Taylor et al. [44] the autocorrelation function is used (equation 9).

$$\rho_{g,s} = \frac{Cov(g,s)}{\sigma_g \sigma_s} \tag{8}$$

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \overline{x})(x_{k+t} - \overline{x})}{\sum_{t=1}^{n} (x_t - \overline{x})^2} \tag{9}$$

The autocorrelation function is a tool that characterizes patterns of wind persistence. The function gives an indication about the order of the model to best fit the data [33]. The r_k is the sample autocorrelation coefficient, k is the time lag, $\overline{x}$ is the mean wind speed, and n is the number of data samples.

5.5 Evaluation metrics

Previous literature have used several evaluation metrics to measure the performance of a forecasting model. According to Clifton et al. [7] the metrics Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) are commonly used to judge forecast accuracy. Therefore in this research we are going to apply these two metrics to judge the accuracy of our forecasting models.

According to H. Holttinen and Sillanpaa [14] the RMSE gives more weight to larger errors and the MAE gives information about the accuracy because it ensures not to cancel out the positive and negative errors.

Also other research conducted by Taylor et al. [44] and Sideratos and Hatziargyriou [42] have used the RMSE and MAE metrics.

6 Methodology

In our literature study we have obtained an answer for the first two research questions. Identifying the factors, input parameters, correlation studies, performance metrics and forecasting models provide insight about what is important to forecast wind power generated by turbines. Using this information we can continue our research to answer our research questions three and four. Which steps have to be taken to obtain a forecast as accurate as the forecast from the external organization Raedthuys is currently using.

This section will explain stepwise our approach to obtain the answers for these research questions. Firstly, we will discuss our data sources with a data description of each data source. Secondly, we introduce our methods we will use for our data analysis. Thirdly we discuss our forecasting method and the division of our data into training, validation and test data. Finally, we discuss which performance metrics we use to compare the performances of the different forecasting models.

6.1 Data sources

To forecast the production of wind power using machine learning techniques historical weather data is needed to find relations between the weather forecasts and the produced wind power generated by wind turbines. During our literature study (section 5) we have found data sources such as meteorological data and SCADA data are useful for the prediction of wind power. These data sources contain weather variables such as wind speed, temperature, pressure etc. In this section we describe the data sources and its data variables used for this research.

For our research we have access to the following data sources: meteorological, SCADA and production data. More detailed information about these data sources is given in section 6.1.1. In section 6.1.5 we introduce the tools used to import and export the data from the database.

6.1.1 Data description

The data used for this research consists out meteorological data, SCADA data and production data. The data sources provide data for different locations and each of them contains a set of weather or turbine variables. In the following sections each of the data sources will be described.

6.1.2 Meteorological data

The meteorological data model HiRLAM has been used for the prediction of wind power and has been provided by the Royal Netherlands Meteorological Institute (KNMI). The KNMI has used the meteorological model to calculate the weather forecast for different geographical latitude and longitude coordinates. The coordinates form a grid covering the Netherlands as can be seen in figure 5. The dots in the figure are the locations of a subset of the turbines from Raedthuys. This research focuses on the prediction of wind power generated by these turbines.

The KNMI has stored a lot of historical data generated by the HiRLAM model and they were willing to provide some of this historical data to conduct this research. How the HiRLAM model has calculated the data is beyond the scope of this research.

We have requested one dataset, which we call 'HiRLAM dataset' from now on. The HiRLAM dataset contains grid points having step size of 0.25. This grid is also shown in figure 5. The dataset contains data covering a period from January 2009 until April 2014. In our research we will use only the data from January 2009 until the end of December 2013.

The HiRLAM dataset provides weather data four times a day (hour 0, 6, 12, 18). Each of these times contain a forecast of 48 hours ahead having step size of six hours (0, 6, 12, 18, 24, 30, 36, 42, 48 hours ahead), which we call FH from now on. This means only forecast weather data is included and no actual data. The weather forecasts from three years ago are the forecasts that were made back then. The reason including no actual data is because we want to forecast wind power based on forecast weather data.

Per geographical location, each forecasting sample contains a dataset of numerical weather predictions (NWP), measured at 2 meter or 10 meter height describing the condition of the atmosphere at the forecast moment in time. Table 10 gives an overview about the weather variables used in this research.

Wind speed and wind direction have been calculated using two vectors (u and v) provided by the HiRLAM dataset, see also equation 10 and 11. We will not further explain those formulas since this is beyond the scope of this research.

$$windspeed = \sqrt{u^2 + v^2} \tag{10}$$

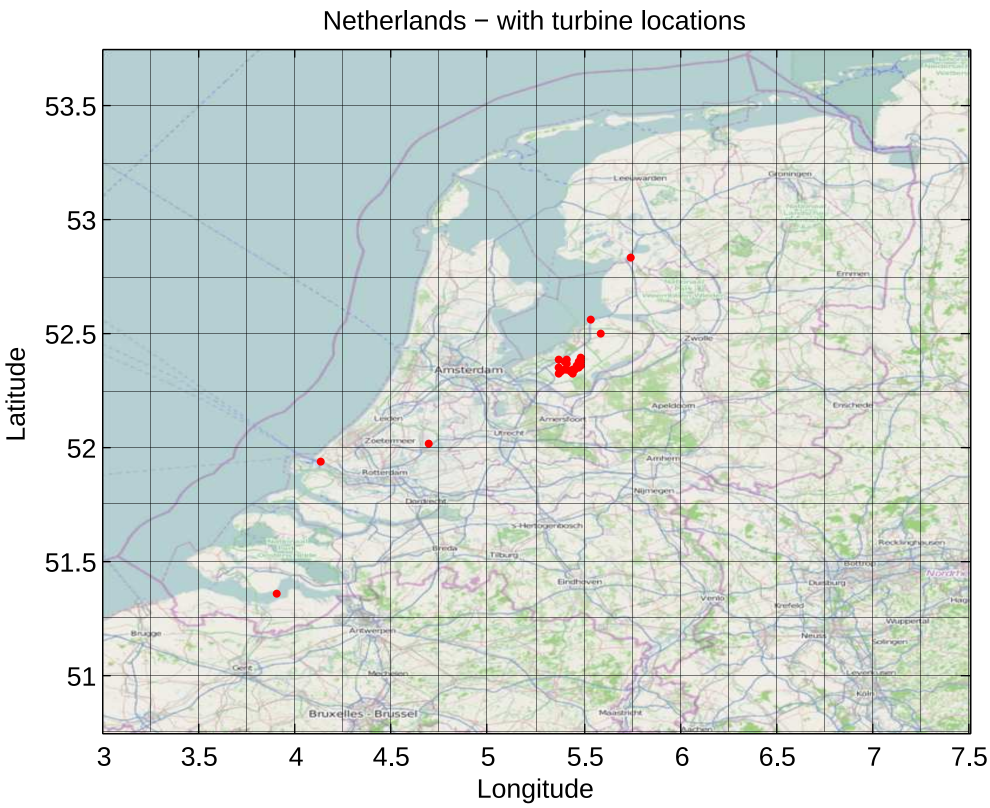


Figure 5: Grid of latitude and longitude points over the Netherlands. The dots are wind turbines from Raedthuys

$$winddirection = arctan2(v, u) \tag{11}$$

6.1.3 SCADA data

The online supervisory control and data acquisition (SCADA) system has been used to obtain data provided by the wind turbines. The data provided by the SCADA system is measured at the location of the wind turbine or farm and provides the actual operational status. Unfortunately not all wind turbines provide SCADA data. In table 11 the available turbine variables are listed. Each data sample contains average information of 10 minutes.

6.1.4 Production data

The production data is information about the amount of wind power generated by turbines. This variable is the output of our prediction models. Each 15 minutes the production measured in kWh is stored per turbine or wind farm in the database. This means for each turbine or wind farm 96 values are stored in the database per day. The production data used in this research covers the

Variablename	Description	Unit measured
Date forecast (DF)	The moment when the forecast takes place	Coordinated Universal time (UTC)
Forecast hours (FH)	The forecast of x-hours ahead from the moment of date forecast	x is integer
Latitude	Specifiying the north-south position of a point on the earth's surface	float
Longitude	Specifiying the east-west position of a point on the earth's surface	float
u vector	The zonal component and represents the horizontal flow of wind in the west-east direction	float
v vector	The meridional component and represents the vertical flow of wind in the north-south direction	float
Wind speed	Wind speed measured 10 meter above the surface	(m/s)
Wind direction	Wind direction measured 10 meter above the surface	Degrees (0 - 360)
Pressure	Pressure measured 10 meter above the surface	Pascal (Pa)
Temperature	Temperature measured 2 meter above the surface	Kelvin (K)
Dewpoint	Dewpoint temperature measured 2 meter above the surface	Kelvin (K)

Table 10: Available weather variables

Variablename	Description	Unit measured
Measured time	The moment of measuring the data	Coordinated Universal time (UTC)
AvgWindSpeed	The average wind speed over the past 10 minutes	(m/s)
AvgRotorSpeed	The average speed of the turbine rotor over the past 10 minutes	(m/s)
MeanPowerkW	The average kW the turbine has produced over the past 10 minutes	(m/s)

Table 11: SCADA variables

period of the HiRLAM dataset. As mentioned earlier HiRLAM provides data on specific forecast hours (FH). This research is focusing on day-ahead forecast. For Raedthuys this means the forecast of wind power needs to be sold against the APX spot price day-ahead. To sell this forecast they have to know the forecast day-ahead before ten o'clock in the morning. This means for example if we want to predict the wind power on January 3th, 2013. We use the forecast weather data predicted on day January 2nd, 2013 on time zero o'clock in the mornings (0:00) and we use the FH 30, 36, 42, 48 (meaning we will forecast the wind power on January 3th 6:00, 12:00, 18:00 and 24:00). Forecasting at 0:00 means we have ten hours to forecast the wind power for the next day. Note, that one can also use the forecast weather data predicted on Januari 2nd, 2013 on time six o'clock (6:00) and use the FH 24, 30, 36, 42. In this case only four hours are left to predict the forecast for Januari 3th, 2013. In this research we will use the most actual forecast moment of six o'clock to forecast wind power for the next day.

Since we are in this research only forecasting four timestamps a day it means only four 15 minute values of the production data that correspond to those forecast times will be used. Since this amount is quite low compared to the 96 values, we will also use the aggregated production over the corresponding hour. This way we are able to cover 16 15-minute values.

Furthermore from all the turbines and wind farms we select a subset of turbines. This subset has been active in the period from January 2009 until December 2013. The reason selecting a subset of turbines is because, in case of predicting the total production of wind power generated by the aggregation of the turbines, new turbines which have been added during the period can influence the total production. Because these turbines have been added later in time they have influence on the performance of the model. For example if the model has been trained on the data from the subset without the new turbine and the model has to predict the total production including the production from the new turbine, then we expect the model lacks performance in predicting the right production amount since it has not been trained on the complete turbine dataset. Therefore if one wants to predict the production of the new turbine then a specific forecasting model for this turbine is necessary to predict the wind power from this turbine.

Another solution is to build for each turbine individually a forecasting model and then aggregate the forecast wind power of each model.

6.1.5 Data tools

Microsoft tools such as SQL Server Integration Services (SSIS) and Microsoft SQL Server (MSSQL) are used to import and export data in databases. Before importing the meteorological data we first had to extract the data from .grib files. We have used the nctoolbox for Matlab to extract the data.

6.2 Methods of analysing data

Before applying the data for the forecasting models as input data we perform a data analysis on the data. First the data will be normalized, because the dataset contains variables represented by different units.

Each data object in the dataset D has to be normalized by its own variable dataset v_i, which is done using equation 12. In this equation $X_{j(v_i)}$ is the object j of variable v_i which will be normalized, μ_{v_i} is the mean over the objects in the set of variable v_i and σ_{v_i} is the standard deviation of the set of variable v_i. $z_{j(v_i)}$ is the normalized object $X_{j(v_i)}$ of variable v_i.

$$z_{j(v_i)} = \frac{X_{j(v_i)} - \mu_{v_i}}{\sigma_{v_i}} \quad \boldsymbol{D} = (v_1, v_2, \ldots, v_n) \tag{12}$$

After normalization we apply two correlation functions on the data to identify the relation among variables. The correlation functions are the correlation coefficient and the auto correlation coefficient. More information about the correlation studies can be found in section 8.3.

The correlation studies are being conducted on data variables of the aggregation of turbines. In our literature study, see also section 5.1.2 we have found that the aggregation of wind power generated by turbines improves the quality of the forecast [2][11][14]. Therefore in this study we perform a data analysis on the prediction of wind power generated by a single turbine and on the prediction of wind power generated by a set of aggregated wind turbines.

We also will apply the Cook's distance measure which is used to detect outliers in the dataset. Using these analysis we can find outliers. More about Cook's distance measure will be explained in section 8.4.

6.3 Training, validation and testing forecasting models

This research will investigate three different forecasting models, namely: Random forest, Neural network and a hybrid model consisting of an unsupervised k-nearest neigbor algorithm and a neural network. A description of the first two models is given in section 4. The hybrid model is described in section 7. The reason to investigate these three models is because neural networks have shown to forecast the best on long term forecasting and because random forest is an easy and fast forecasting method.

Each of the models will be trained, validated and tested. The main difference between the validation of the model and testing the model is that the validation of the forecasting model provides an estimation on how good the model has been trained based on specific parameters/properties. Obtaining the best forecasting model is therefore based on selecting the model providing the lowest validation performance error. Having obtained the best forecasting model we can test this final model and determine its performance when new data is being used as input data, testing the model gives an indication about its performance when using the model life.

Validation of forecasting models using times series datasets cannot be performed by randomly selecting a training, validation and test set. The reason is that the algorithm can be trained by data samples later in time compared to data of the test data samples. In other words the algorithm uses future samples to predict the past which is unacceptable in this time series dataset.

To avoid this problem we will apply a dataset in chronological order and split the dataset into three subsets, the training, validation and test set. The training dataset is the first part of the chronological dataset, the validation set is the second part and the test set is the final part. The sizes of the dataset will be initialized as follows. The training dataset is 60% of the dataset, the validation is 20% and the test set is also 20%. The reason for these percentages is because we have five years of data. The first three are used as training data, one year as validation data and one year as test data.

While moving through time, the forecasting model is trained and validated using the training and validation set. Then the samples of the first day from the test set are predicted. After prediction the three datasets will be resized as follows:

Training set: the size of the training dataset increases by shifting the validation samples of the oldest day from the validation set into the training set.

Validation set: the size of the validation set will remain the same by removing the oldest day samples and adding these to the training set and receiving the oldest test samples (the predicted samples) from the test set.

Test set: the size of the test set decreases by shifting its predicted samples to the validation set.

The next test samples are predicted after retraining the model using the updated training set and validated using the validation set. This process is an iterative process until all the test samples are predicted. In figure 6 this prediction process is visualized.

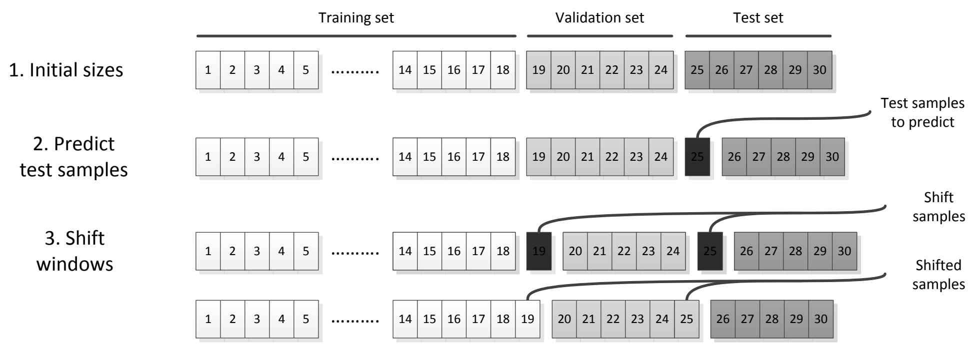


Figure 6: The window shifting prediction process of test samples. The numbers represent day-numbers for the ease of understanding.

6.4 Performance measurements

The forecast of the test set is obtained from the different forecasting models using different input parameters. After having obtained all the forecast results from the forecasting models we apply the performance metrics the root mean square deviation (RMSD) or root mean squared error (RMSE) and the mean absolute error (MAE), see equation 13 and 14 respectively.

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2} \tag{13}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\|\hat{Y}_i - Y_i\| \tag{14}$$

We also calculate the RMSD and MAE of the forecast Raedthuys is currently using. We express the MAE and RMSD in percentages by dividing the errors by its actual production, see equation 15 and 16. We will discuss our metrics in percentages and we will use the terminology mean absolute percentage error (MAPD) and root mean square percentage error (RMSPD).

The MAPD is calculated dividing the total error MAE by the total allocated production Y_i, see also equation 15. The RMSPD is calculated dividing the total error RMSD by the root mean square of the total allocated production Y_j, see also equation 16.

$$MAPD = \frac{MAE}{\sum_{i=1}^{n} Y_i} \tag{15}$$

$$RMSPD = \frac{RMSD}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}(Y_j)^2}} \tag{16}$$

Based on all the performance values we can determine which model is the most accurate (lowest performance error). Using this information gives us the answer to our research questions three and four. Research question three can be answered by comparing the accuracy of the forecasting models and the forecast from the external organizations. Research question four is answered by analysing the different parameters that have been used to obtain the lowest performance error.

7 A hybrid model

As mentioned in section 4 two forecasting model are being used to forecast wind power generated by turbines. Besides these two models a third model is being used. This model is a hybrid forecasting model, a two stage model using a combination of an unsupervised k-nearest neighbour clustering with a neural network.

The next sections explain the model in two steps.

7.1 Step 1: Unsupervised k-nearest neighbor clustering

The first step of this model is to apply the unsupervised k-nearest neighbor clustering algorithm. The main goal of this step is to split the dataset in k clusters in a unsupervised way. Unsupervised means that the algorithm identifies k clusters based on the patterns/similarities among the input features (variables) instead of the output value (target variable) [48].

7.1.1 Process of the algorithm

The basic idea behind the unsupervised k-nearest neighbor clustering algorithm is to divide the dataset in k clusters where each cluster is a subset of the dataset. The process of unsupervised k-nearest neighbor clustering can be explained in several steps [26]:

Step 1: First k centroids are being initialized. These centroids represent the centres of each cluster.

Step 2: For each data point the distance between the data point and each centroid is calculated using a distance function. The most common distance function is the Euclidean distance shown in equation 17. Here x_i is the data point and c_i is the centroid.

Step 3: Each data point is being assigned to the cluster with its centroid having the smallest distance.

Step 4: After having assigned each data point to a cluster we can re-position the centroids. For each centroid the mean of all the distances between the data points in cluster C_i and centroid c_i is calculated. The calculated mean is the new value being assigned to the centroid c_i.

Step 5: Knowing the new positions of the centroids each data point is re-assigned to a cluster by starting again from step 2.

After a number of iterations the positions of the centroids remain stable and the clustering process stops. At this point we have divided the dataset into k subsets.

$$d_{Euclidean}(x, c) = \sqrt{\sum_i (x_i - c_i)^2} \tag{17}$$

7.1.2 Applying unsupervised k-nearest neighbor clustering algorithm

The unsupervised k-nearest neighbor algorithm is the first step in our forecasting model. Before performing the k clustering algorithm we firstly normalize our dataset using equation 12 explained in section 6.2. This is necessary since we are dealing with different variables being presented in different units. Large values such as wind direction have a greater influence than smaller values such as wind speed when applying the euclidean function. We assume our dataset is normal distributed.

After normalization we apply the unsupervised clustering algorithm. The training and validation data are distributed among k clusters. When the positions of the centroids are known the test data is distributed among the k clusters.

In figure 7 one can see an example of the distribution of the data among four clusters.

In this figure only the wind speed variable has been used to distribute the data against the wind power. As one can see the the clusters have on their boundaries a small overlap. It might therefore be the case that some of the test data points are being assigned to the wrong cluster. To over come this problem we calculate the probability of a data point belonging to cluster i using equation 18. C_i stands for cluster i and x is the data point. $P(x|C_i)$ stand for the likelihood (equation 19[28] in case of a univariate distribution and equation 20 [28] in a multivariate distribution) of x in cluster i and $P(C_i)$ stand for the number of observations in cluster i.

$$p_i = P(x|C_i)P(C_i) \tag{18}$$

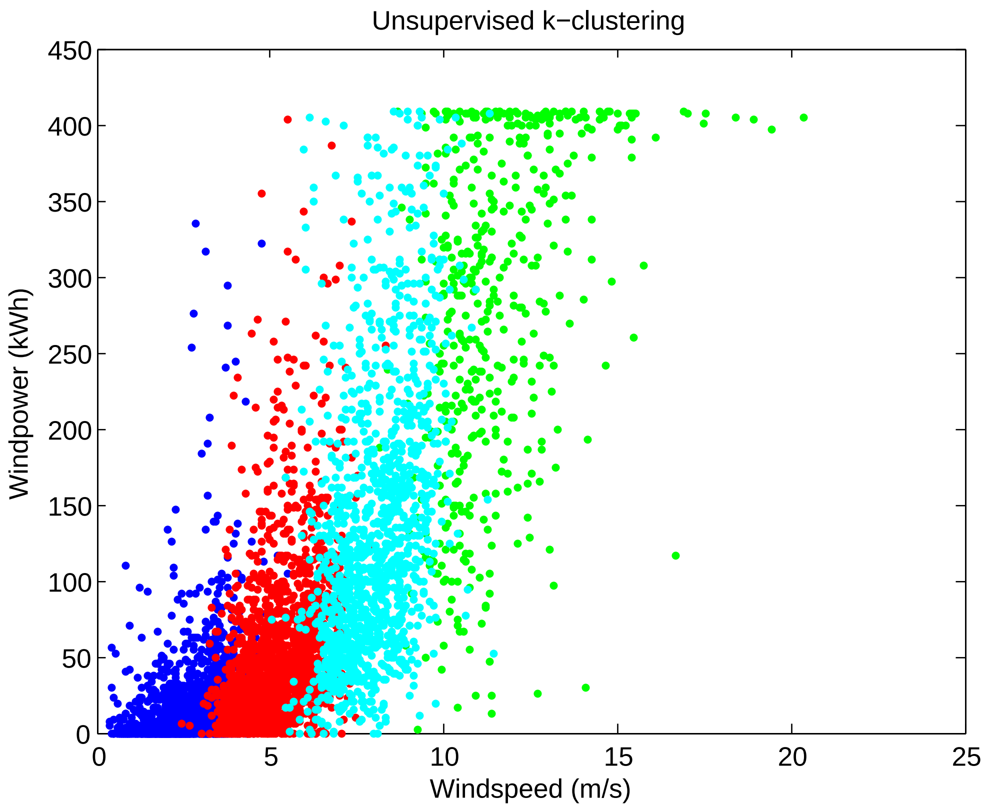


Figure 7: Wind speed versus wind power applying unsupervised k-clustering ($k = 4$)

$$P(x|C_i) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \tag{19}$$

$$P(x|C_i) = \frac{1}{(2\pi)^{d/2}\left|\sum_i\right|^{1/2}} e^{-\frac{1}{2}(x-\mu_i)^T \sum_i^{-1}(x-\mu_i)} \tag{20}$$

After having obtained the probability of the testdata points in each cluster, we can perform our second step of this forecasting model, which is applying a neural network. The description of the feed forward neural network is given in section 4.2.

7.2 Step 2: Applying feed forward neural network

The second step of this model is applying the feed forward neural network (FNN). For each cluster a neural network is trained and validated based on the training and validation data assigned to that cluster.

When testing a data point we obtain from each neural network its predicted output. This means we have k number of predicted outputs. The final output is calculated using equation 21. The sum of all the probabilities (p_j) of a test data point belonging to a cluster j times the outcome of its FNN ($\hat{Y}_j$) divided by the sum of all the probabilities (p_i) of the test data point belonging to a cluster i.

$$\hat{Y} = \frac{1}{\sum_{i=1}^{k} p_i} \sum_{j=1}^{k} p_j \hat{Y}_j \tag{21}$$

8 Data analysis

To forecast the production of wind power using machine learning techniques historical weather data is needed to find relations between the weather forecasts and the produced wind power generated by wind turbines. In section 6.1.1 we have discussed the available data sources and its content. In this section we analyse the data from these the data sources. We will perform correlations studies on the data to identify which variables and which dimensions have great influence on the prediction of wind power generated by wind turbines. In this section we will analyse the data from the aggregation of a subset of the wind turbines from Raedthuys.

The analysis of the data was carried out as follows. First the number of grid points has been reduced. Only those grid points which are related to the total wind production are being selected. Secondly, we select the variables which are recommended by previous literature. Thirdly, two correlation studies are conducted, which are the correlation and the autocorrelation coefficient. Finding the correlation values between the weather variables of the grid points and the aggregated production of wind power generated by the turbines provides us information about which variables have any dependency with the prediction of the production of wind power generated by the aggregation of turbines. Finally we describe the detection of outliers using the Cook's distance measure.

8.1 Grid point reduction

The HiRLAM dataset contains 228 grid points. Each point contains five weather variables (wind speed, wind direction, pressure, temperature and dewpoint), which will result for each forecasting time in 1140 dimensions. Many grid points are not even located near the turbines as can be seen in figure 5. Kusiak et al. [20] has found that the nearest four surrounding grid points are being selected by among other the boosting tree algorithm. Also Barbounis and Theocharis [1] have used the nearest four grid points to predict wind power generated by wind turbines. Therefore we remove all the grid points except the nearest grid points surrounding the locations of the turbines as can be seen in figure 8.

Applying the reduction of grid points results into 21 grid points from the HD1 dataset representing 105 dimensions.

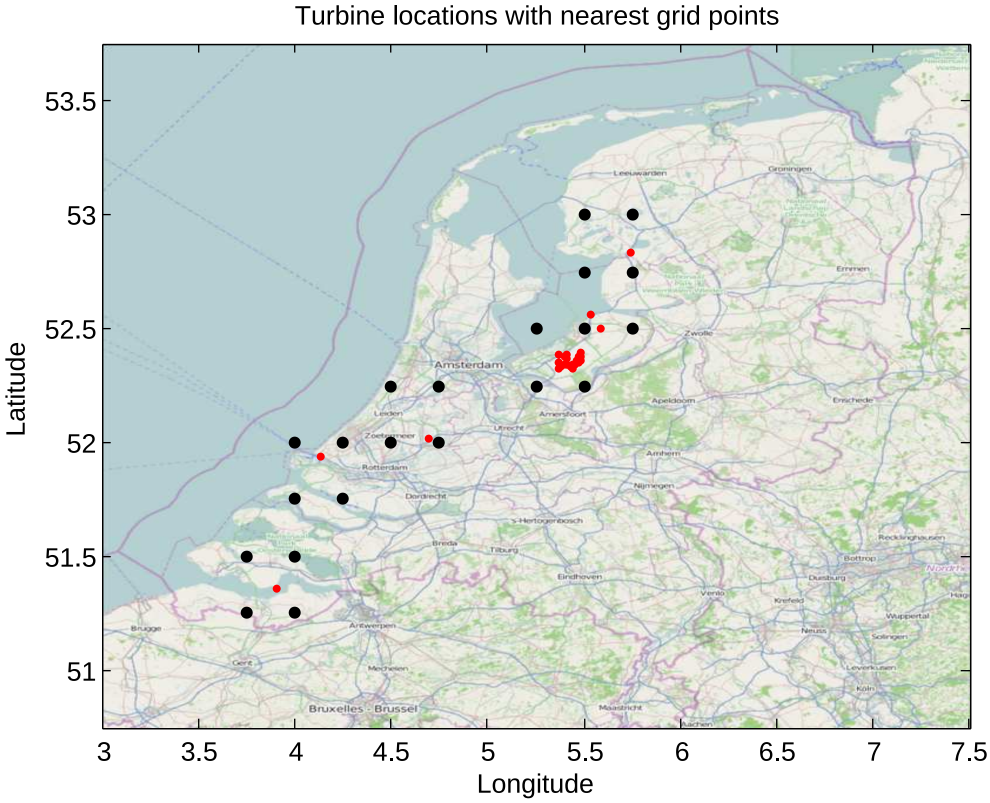


Figure 8: Black dots are the grid points nearest the turbines

8.2 Selection of input features

Based on our literature study we have identified a set of weather features which are: wind speed, wind direction, temperature, pressure and relative humidity. Unfortunately, according to literature the variables temperature and pressure have not found to be of use to predict the production of wind power generated by wind turbines. Therefore we will not include these two variables into our selection of variables.

Furthermore we will add three new variables which are the air density, the maximum amount of energy that can be produced and the relative humidity. Air density is part of equation 5 (see section 5.2.1). The air density can be calculated using equation 22. Here ρ is the air density in kg/m^3, p is the absolute pressure, $R_{specific}$ is the specific gas constant of dry air which is 287.058 J/kgK and T is the absolute temperature. Even tough temperature and pressure do not have any correlation with the production of wind power separately. They might show some dependence when using those variables in a combination of a function.

$$\rho = \frac{p}{R_{specific}T} \tag{22}$$

Input variables	Output variables
Wind speed Wind direction Air density Relative humidity Maximum amount of power available	Wind power (15-minute value) Aggregated wind power (Past hour)

Table 12: Input and output variables

The maximum amount of energy is used to determine if a turbine is active or not at a certain moment in time. The aggregation of turbines consist out turbines of different heights or the same height depending on the selection of the turbines. Larger turbines have a higher capacity of producing energy. This means they have a higher influence on the total maximum amount of energy.

Furthermore, the relative humidity is being calculated using the dew point and temperature variables from the HiRLAM dataset, see also equation 23 [39].

$$RH = 100 \left(\frac{112 - 0.1T + T_D}{112 + 0.9T} \right)^8 \tag{23}$$

The above mentioned data variables also listed in table 12 will be analysed performing two correlation studies, which are being explained in the next subsection.

8.3 Correlation studies

Our dataset contains variables expressed in different units as can be seen in table 10. Therefore to ensure all the variables are measured on the same scale we first normalize all the data. Applying the normalization equation 12 (section 6) on each variable in the dataset ensures that all the data is being normalized where μ is 0 and σ is 1. Having normalized all the data we have performed two correlation studies the correlation coefficient and the auto correlation coefficient.

8.3.1 Correlation coefficient

The correlation coefficient can be calculated applying equation 24. This equation is also called the Pearson‘s Correlation Coefficient. The correlation coefficient is a univariate method which explains the dependency between two variables. It identifies the dependency of one variable X with a variable Y. The correlation coefficient is within -1 and 1. -1 Means that the value is negatively correlated and 1 means that the values are positive correlated.

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{24}$$

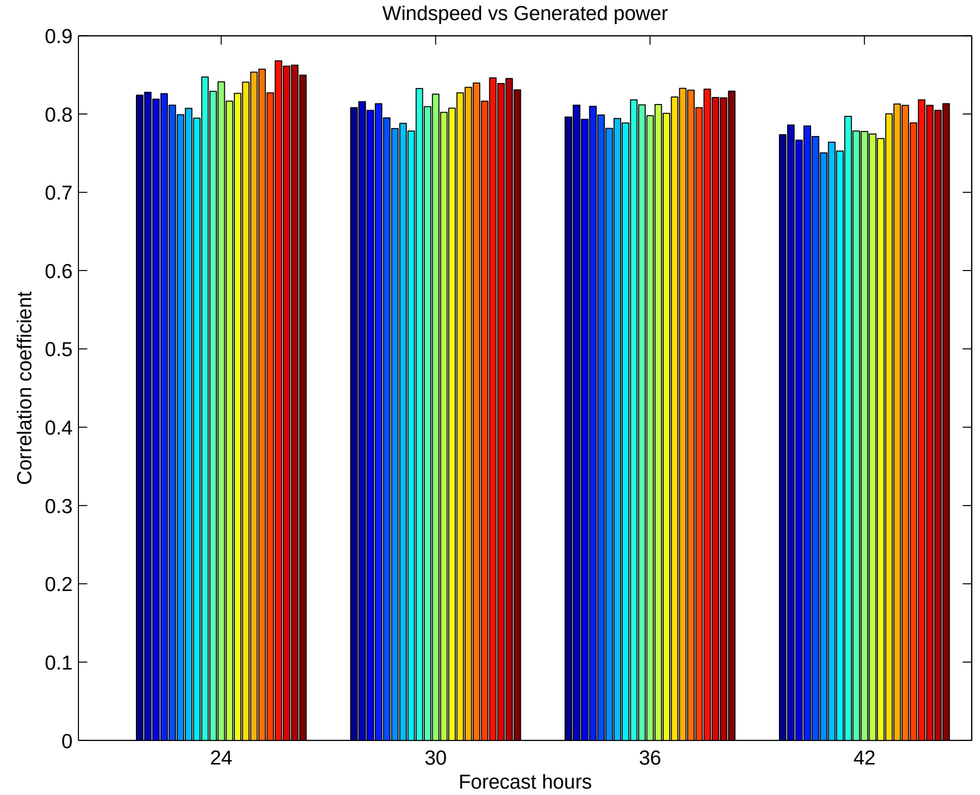
For each input variable shown in table 12 we have calculated the correlation coefficient with the output variable wind power 15-minute basis. In figures 9 and 10 the correlation coefficient is shown for each location and for each forecasting hour day-ahead (24, 30, 36, 42). The correlation coefficient has also been calculated for the generated power on hourly-basis. This has resulted in similar results compared to the coefficients of the 15-minute production. Therefore we do not show these results in this research.

8.3.2 Discussion

Based on the correlation study we have found that wind power generated by turbines has the highest correlation value of 0.8 (see figure 9. This however does not come as a surprise, since also previous research has identified high correlation values between wind speed and the production of wind power generated by wind turbines.

Also the variable wind direction does show a positive correlation coefficient for all the locations. Looking closely to the wind direction it can be seen that not all locations have an equal correlation coefficient as is shown with the wind speed. The twelve grid points on the left (see figure 8) near the shore show to have the highest correlation values compared to the grid points located in the centre of the Netherlands. The reason for this higher correlation value is because we assume that the wind turbines located near the North Sea are less affected by geographical obstacles or other barriers and therefore show possibly a higher correlation value compared to the wind turbines located in the centre of the Netherlands. Furthermore, based on figure 11 one can see that the wind direction mostly is coming from South-West and West (200 - 270 degrees).

Furthermore in figure 10 we see the other variables air density and relative humidity showing to have a negative correlation value and fluctuate around the zero correlation. Possible reason is that these two variables are really location specific and therefore perform worse on the aggregation of turbines. To confirm this reason we have also calculated the correlation values for air density and relative humidity with the generated power for one turbine from the set of turbines as can be seen in figure 12. Based on this figure we can conclude with some certainty that the air density is location specific since it shows a positive correlations values for a single wind turbines, but still the correlation value is near zero, meaning there is no linear coherency between the variables. The relative humidity however performs worse compared to the correlation values of the aggregated set of turbines.



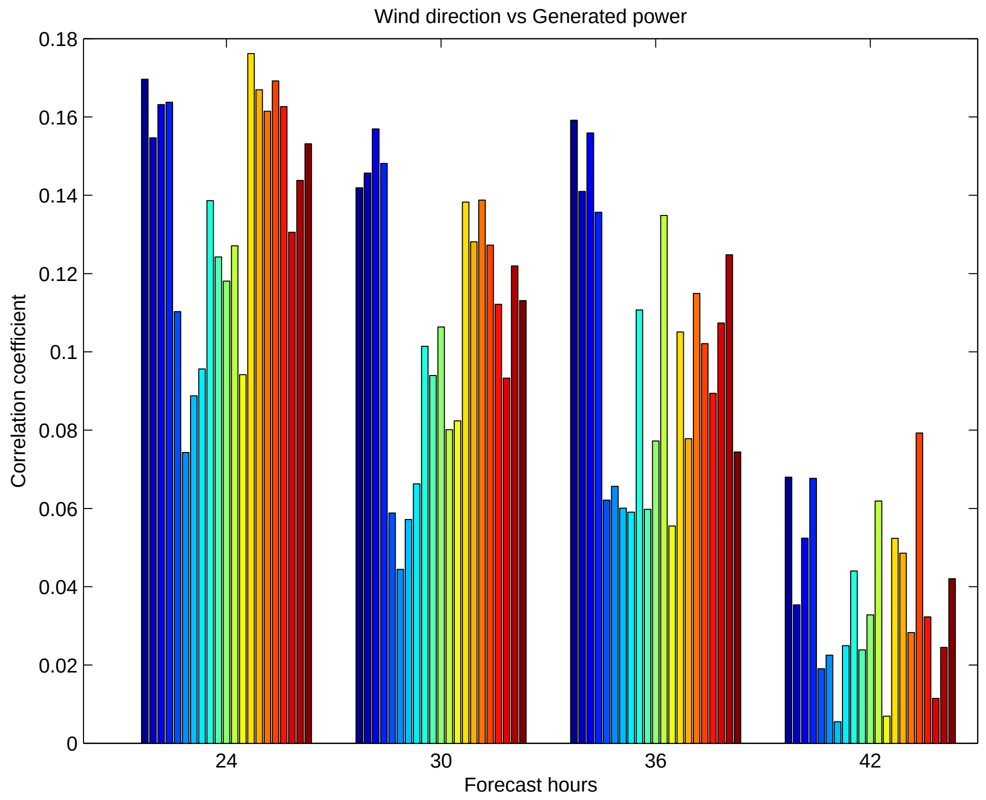
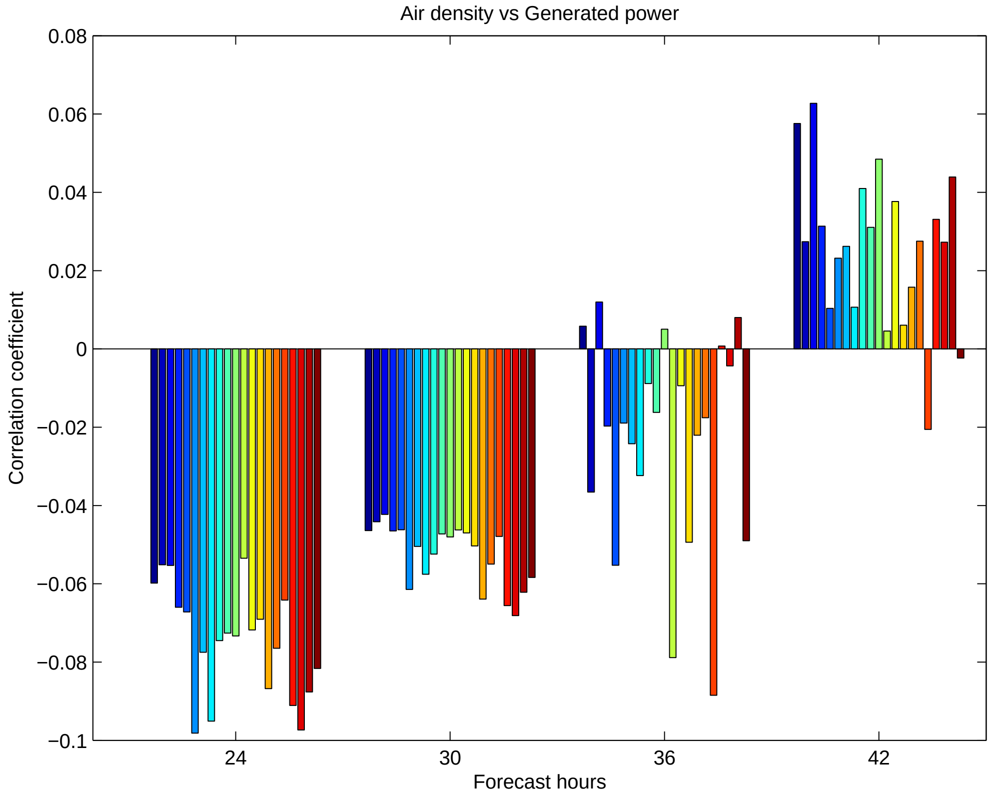


Figure 9: Correlation coefficient between wind speed and generated power (15-minute) and wind direction and generated power (15-minute) - each color stands for one geographical data point from figure 8



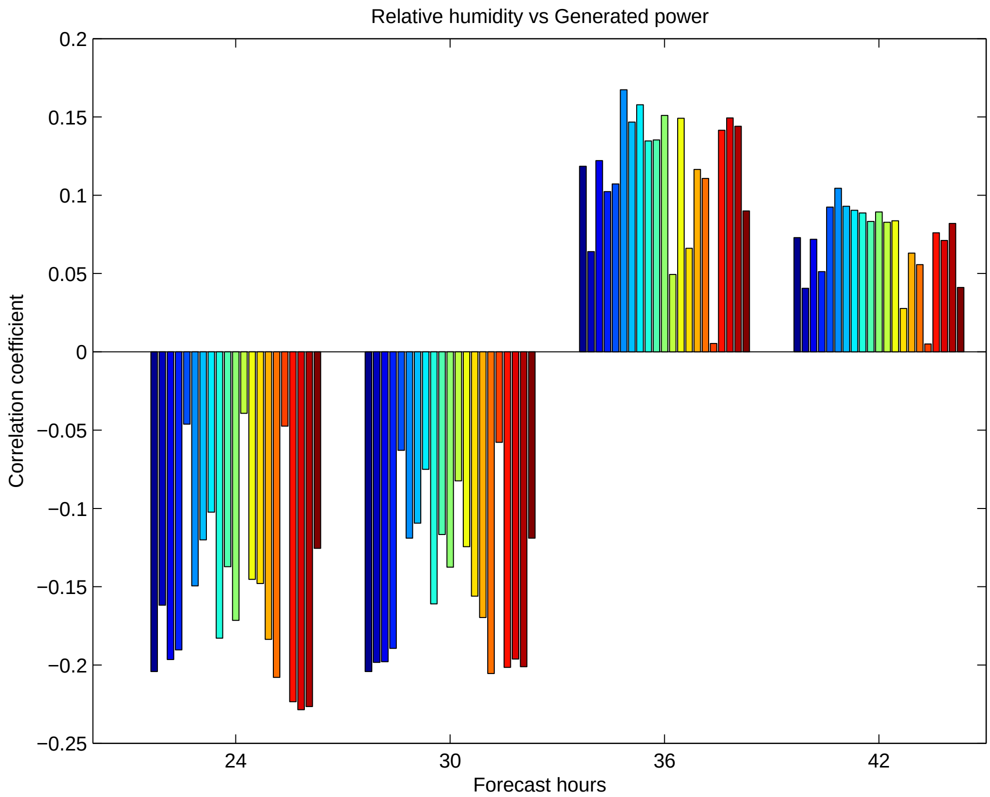


Figure 10: Correlation coefficient between air density and generated power (15-minute) and between relative humidity and generated power (15-minute) - each color stands for one geographical data point from figure 8

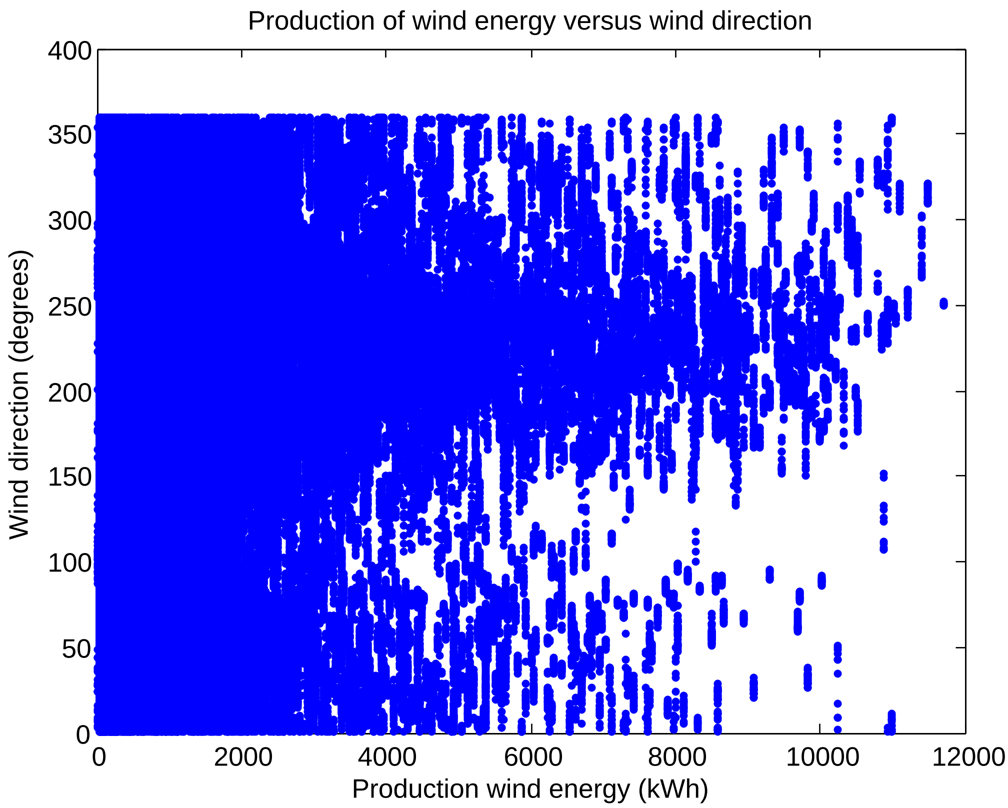


Figure 11: Total wind power generated by wind turbines versus the wind direction

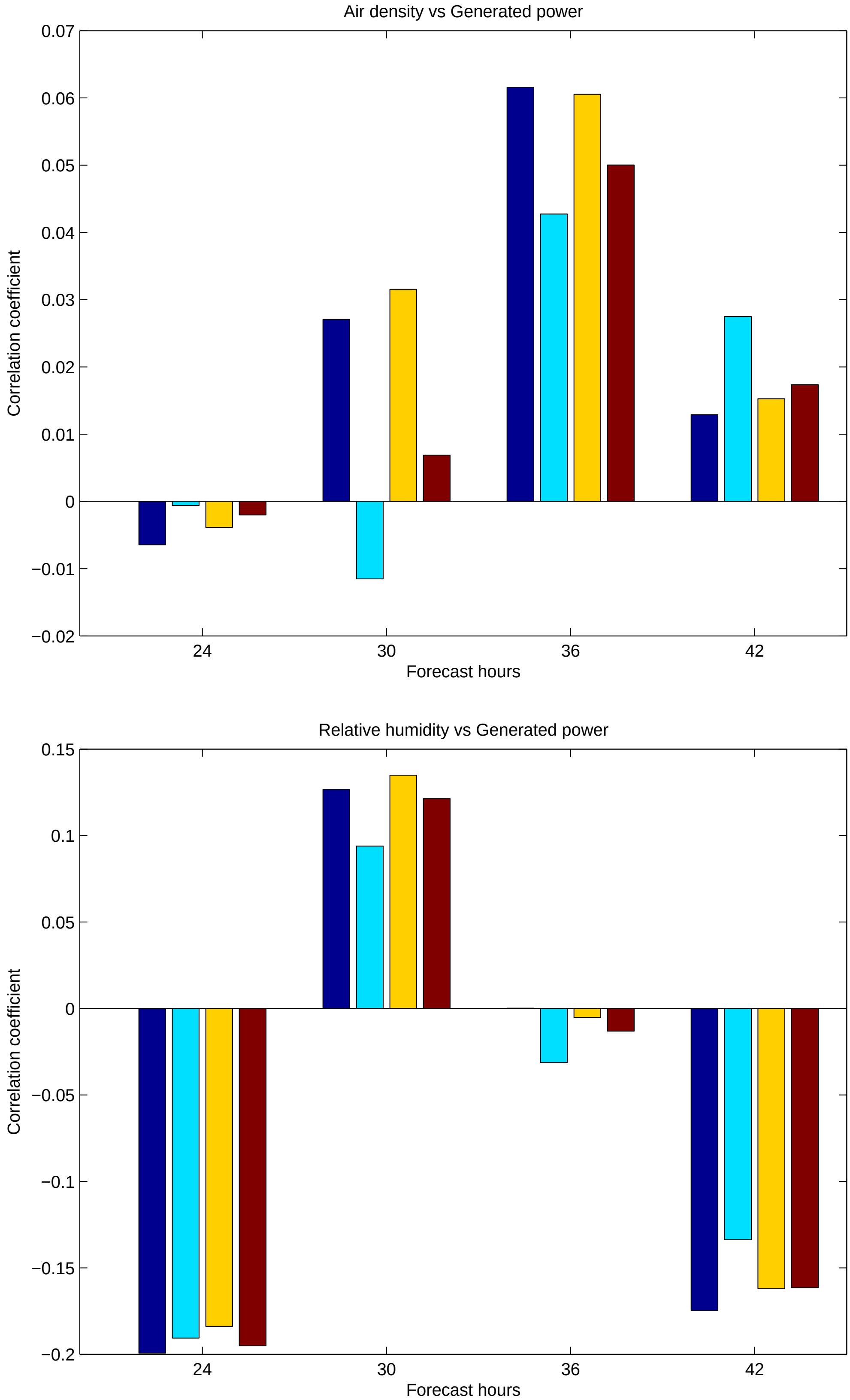


Figure 12: Correlation coefficient between air density and generated power (15-minute) and between relative humidity and generated power (15-minute) of one turbine - each color stands for one geographical data point

8.3.3 Autocorrelation coefficient

The second correlation study we performed is the autocorrelation coefficient. From our literature study (section 5.4) we have found that this is a common correlation study to perform on time series datasets. Literature has performed the correlation study mainly on the wind speed variable to identify if there is a relation between two adjacent time stamps. Since we only have data every six hours, we have performed our autocorrelation study on six hourly data, meaning that one time lag is equal to a time shift of six hours. In figure 13 our autocorrelation results from the complete set of locations is viewed. As can be seen in the figure lag 0 has a correlation value of 1 which is correct since this is the correlation with itself. Furthermore lag one shows to have a correlation value of about 0.75, therefore we will take this feature in consideration when forecasting wind power generated by turbines, using our forecasting models.

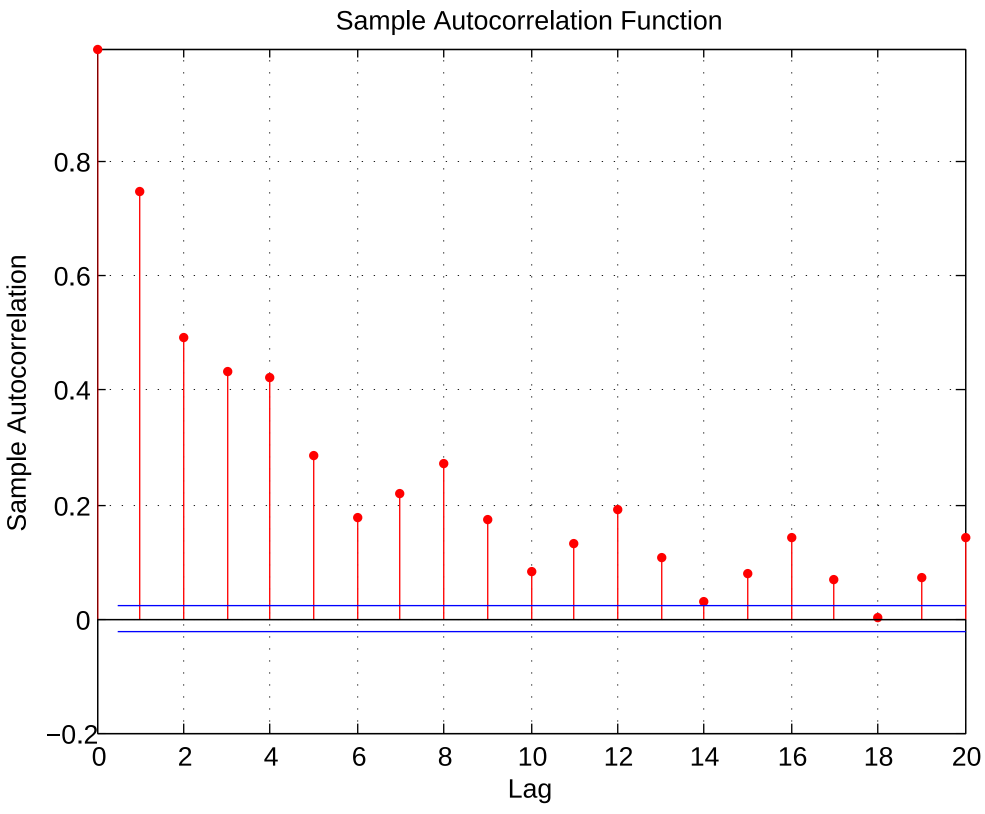


Figure 13: Autocorrelation between wind speed values (1 time lag is a time difference of 6 hours)

8.4 Cook's distance measure

When the dataset contains wrong data, forecasting algorithms being trained with this wrong data are being biased. Before we can remove this bias we first have to identify this wrong data (hereafter called outliers) and remove these outliers before training the forecasting model.

In this research we will apply the Cook's distance measure, see equation 25 to identify outliers. An outlier is according to Jonsson [16] a data point which is unrepresentative for the data set as a whole.

The Cook's distance measure detects the influence of observations in linear regression and is therefore useful to detect outliers. It measures the effect of deleting a given observation [18]. In equation 25 the $\hat{Y}_j$ is the predicted output from the full regression model for observation j. $\hat{Y}_{j(i)}$ is the prediction for observation j from a refitted regression model where observation i has been deleted. MSE is the mean square error of the regression model and p is the number of fitted parameters in the model [18].

The higher the Cook's distance of a data point the more likely the omitted data point i is an outlier and need to be examined. Heiberger and Holland [15] state that if the Cook's distance is greater than one for a data point, it is assumed to be examined.

$$D_i = \frac{\sum_{j=1}^{n} (\hat{Y}_j - \hat{Y}_{j(i)})^2}{pMSE} \tag{25}$$

In our research we will apply the Cook's distance in the middle of our forecasting process as can be seen in figure 14. After training and validating the forecasting model we can obtain the predicted outputs of the training and validation samples. The predictions of our samples will be plotted against the actual outcome and a linear regression function is created. On this linear regression function we will apply the Cook's distance measure. We assume predictions showing to have a high error value are probably outliers. Data samples which provide a high Cook's distance value will be deleted and the forecasting model will be completely retrained without those deleted samples. This way we train our forecasting models without creating the bias for wrong data and hopefully this will increase the accuracy of our prediction models. To obtain the accuracy of the retrained forecasting model all the samples have to predicted again including the outliers, since the outliers are still timestamps which need a forecast.

Furthermore determining which predictions are outliers is based on a threshold. This threshold can be determined by trial and error. As can be seen in the figure the threshold is based on i times the mean of all the Cook's distances. If a prediction sample has a higher distance than the threshold we will remove this sample from the dataset. Figure 15 gives an example of how the datasamples can be distributed. The black dotted line at 0,003 is the threshold, meaning that the samples above this threshold are classified as outliers.

8.5 Scada data and wind speed distribution analysis

In this section we analyse the SCADA data. From our selection of wind turbines we have selected one wind turbine of which we have scada data available. In our research we have used the variable AvgWindSpeed as is mentioned in table 11.

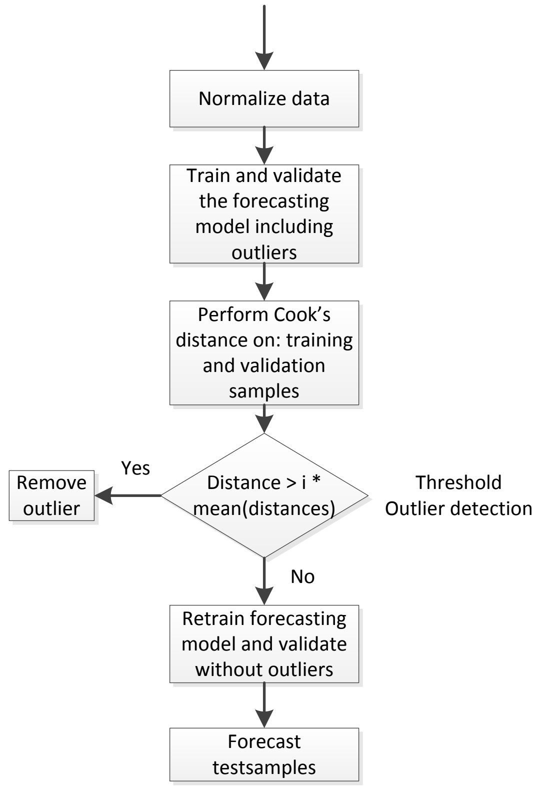


Figure 14: The Cook's distance measure included in the forecasting process

The purpose using this variable is to predict the wind speed at turbine location, to obtain the relation between the wind speed from the surrounding grid points and the SCADA wind speed. In figure 16 the distribution of wind speed against the wind power from the grid points for different forecast hours is compared with the wind speed from the SCADA data. As one can see from this figure the distributions of the forecasted wind speed from the grid points show to be similar for different hours. Based on this information will not take the forecasting hours as input parameters for our forecasting models.

Furthermore, as one can see the AvgWindSpeed scada data does show the ideal relation between wind speed and wind power. The figure gives a clear indication about the difference between the wind speed distributions from the forecast hours and the wind speed distribution obtained from the SCADA data. We will build a neural network to predict scada data using the predicted u-vector and v-vector from the grid points as input parameters and training parameters, and we use the actual SCADA data as output parameters. This way the network can be build to predict SCADA data for the test set. The predicted SCADA data will then be used to predict wind power generated by the turbine. The reason first predicting SCADA data is because the distribution of wind speed where the maximum is around 20 m/s is lesser than the prediction of wind power which is around 400 kWh. This might therefore decrease the prediction error.

8.6 Conclusion

Based on our correlation studies we have found that wind speed, wind direction and wind speed time lag one show to have positive correlation values with the wind power output. Therefore we will include these weather variables into our forecasting model. Since the u and v vector are the input variables for calculating wind speed and wind direction we will also take into account these vectors and use these as replacements for the current wind speed and wind direction. The main reason to perform this action is because of the wind direction. This weather variable is presented in degrees where 0 degrees is equal to 360 degrees (like a circle). It is therefore questionable if the forecasting models can identify the difference and similarities between different wind directions.

Another variable we will take into account is the maximum amount of power that can be delivered by the aggregation of wind turbines. Finally, identifying outliers we will perform Cook's distance measure.

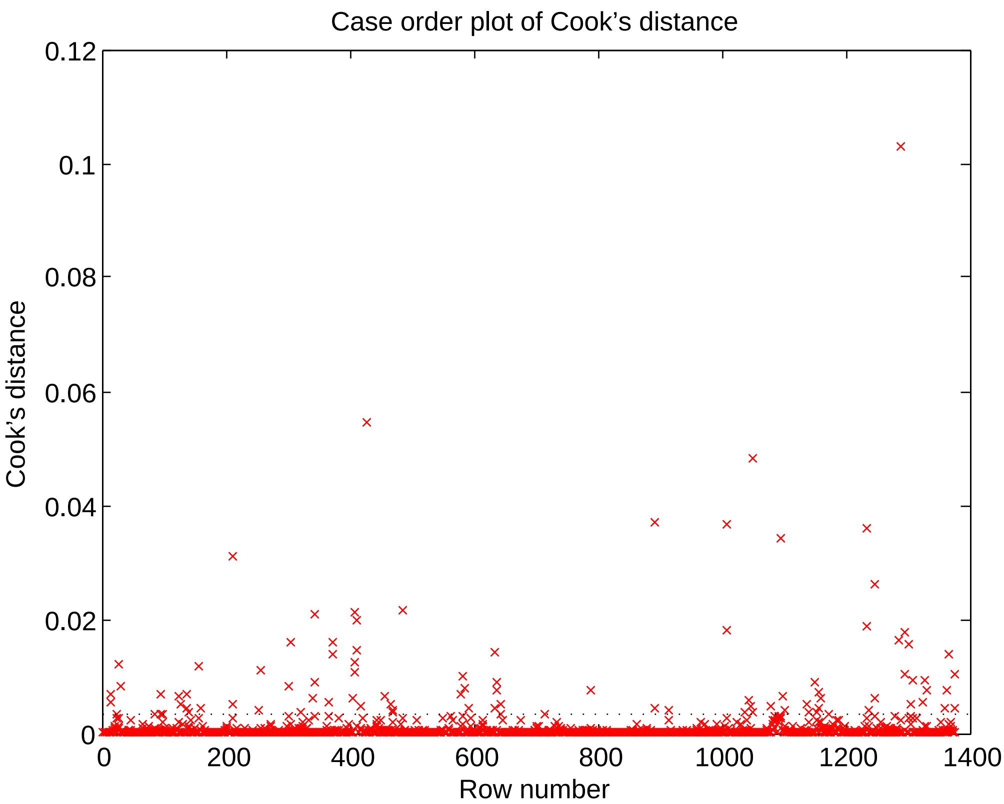


Figure 15: An example of how distances from the samples are plotted (red cross). The black dotted line is the threshold.

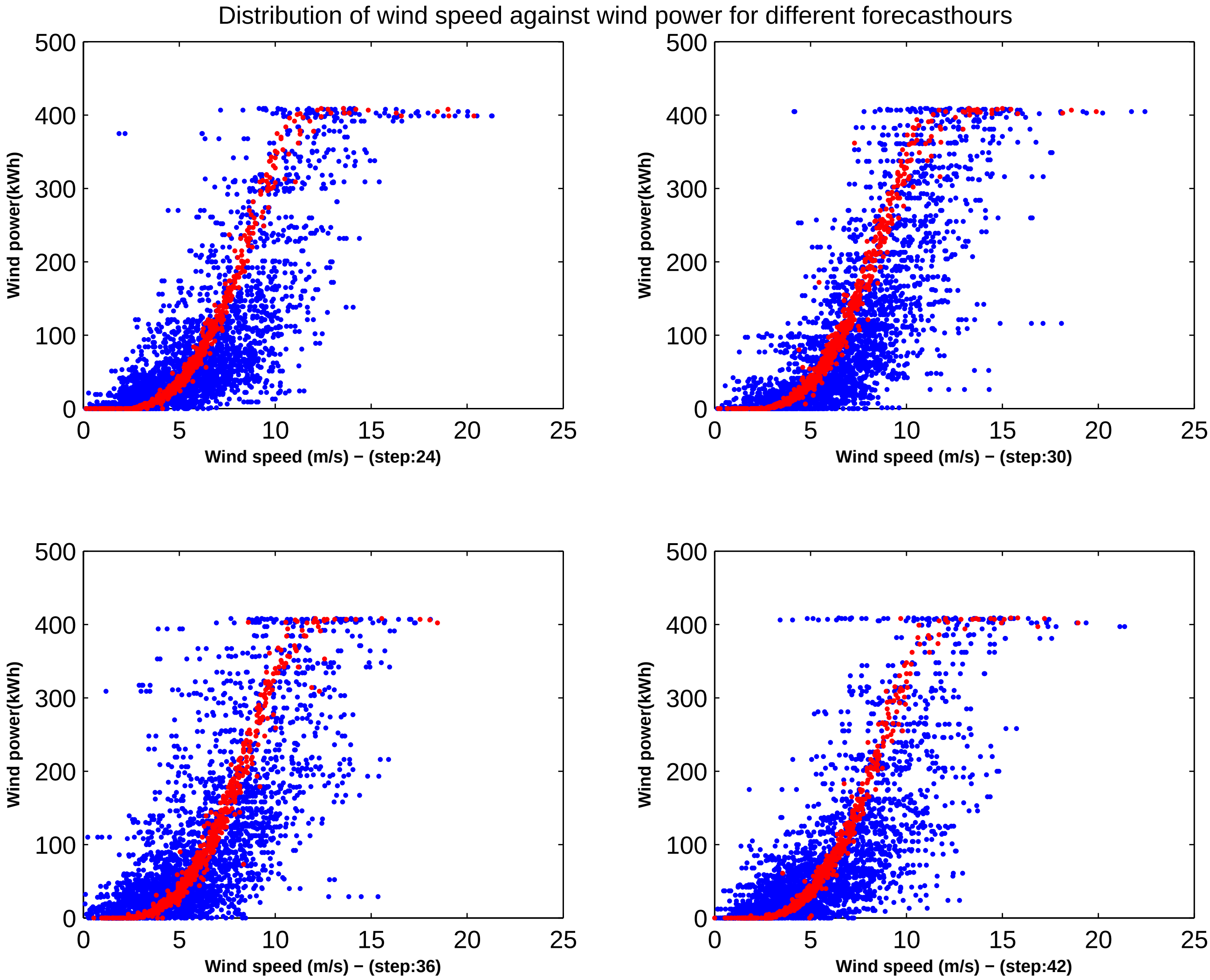


Figure 16: The blue dots represent the wind speed distribution of the grid points, the red dots is the AvgWindSpeed scada data.

9 Results

We have applied the three forecasting models (Random Forest, Feed Forward Neural Network and the Hybrid model) on three different aggregation areas of which more detail is given in table 13. Our window shifting method (as explained in section 6.3) has been applied on each forecasting model. Different combinations of input parameters such as the wind speed, wind direction, u-vector, v-vector and wind speed time lag 1 have been used to predict the wind power generated by these different areas. In table 14 a summary is given of the parameters we have used in our research with their corresponding ID's. The ID's in table 14 will be used in our results to represent the parameters.

Obtaining our results five years of data have been divided into three separate datasets. The years 2009 until 2011 have been used as training data, year 2012

Number of turbines	Capacity (kWh)	Scada data available
1	1 × 1600	Yes
27	27 × 850	No
34	1 × 1600 1 × 12000 1 × 660 1 × 9000 30 × 850	No

Table 13: Three predicted areas

ID	Input parameter
2	Forecast hours
10	Wind direction
13	Wind speed at hub height
17	Scada data
18	Wind speed of time lag 1
20	u-vector
21	v-vector

Table 14: Input parameters and their ID's

as validation data and year 2013 as test data (60%, 20%, 20%).

Furthermore, to obtain the results we first have identified the best Random forest, Feed forward neural network and Hybrid model for the specific set of input parameters. We have obtained these best models as follows.

We have used the training and validation sets as data to obtain the best neural networks and forests. In case of the Random forest we have iteratively increased the number of trees by 10 each time up to 500 trees. The Random forest showing to have the lowest validation performance (lowest OOB error) was selected as the 'best' Random Forest to use for the specific set of input parameters.

In case of finding the best neural network a similar technique has been applied. Iteratively increasing the number of hidden neurons by 1 up to 10 hidden neurons. For each number of hidden neurons Cook's distance has been applied by iteratively increasing the threshold by 1 up to 10 also. This gives us 100 neural networks. The neural network providing the lowest performance error obtained using the validation set has been selected as the 'best' neural network for that specific set of input parameters.

Furthermore, we have set the number of clusters for the Hybrid model on three. The reason selecting this number is because of the structure of the power curve, see also figure 4 in section 5. The first cluster is based on the wake of the turbine (Cut-in speed), the second cluster is based on the linear part and the third cluster is based on the Rated output power (the maximum a turbine can produce).

Our results show the monthly average MAPD and RMSPD for four 15-minute values and for 16 15-minute values (the corresponding hours) of the test data.

The first three subsections present the results of each forecast area. Each area is forecast by the three forecasting models introduced in this research. A fourth model is the external forecasting Raedthuys is currently using (hereafter called Forecasting E.). The results of all the four models are compared per forecast area.

The final section contains results on the financial impact by comparing the results of one single turbine. These results have been included because of curiosity but are not further analysed in this research, since it is beyond the scope of this research.

9.1 Single turbine

This section discusses the results obtained from a single turbine connection where SCADA data is available. For each forecasting model different combinations of imput parameters have been used to predict wind power and to obtain the average month MAPD and RMSPD are calculated for four 15-minute values and their corresponding hours (16 15-minute values), as is shown in table 15. The MAPD and the RMSPD represent the performance test error obtained using the test data.

Model	**Parameters ID**	**Clusters**	**4 15-minute values**		**16 15-minute values**	
			MAPD	**RMSPD**	**MAPD**	**RMSPD**
External org.	-	-	40.7%	42.8%	39.4%	41.3%
Random forest	17	-	41.8%	43.4%	41.3%	42.3%
	17,18	-	42.1%	43.2%	41.2%	41.8%
	20,21	-	44.7%	43.9%	43.3%	42.4%
	18, 20, 21	-	44.8%	43.6%	43.0%	41.8%
	10,13	-	44.7%	43.8%	43.0%	42.1%
Feed forward neural network	17	-	41.6%	43.3%	40.5%	41.8%
	17,18	-	42.0%	43.3%	40.0%	41.4%
	20,21	-	44.8%	44.2%	43.2%	43.1%
	18, 20, 21	-	45.0%	44.4%	42.1%	41.8%
	10,13	-	46.0%	45.2%	43.4%	43.1%
Hybrid model	17	3	41.7%	43.3%	40.7%	41.9%
	17,18	3	41.7%	43.3%	40.6%	41.8%
	20,21	3	44.0%	43.8%	43.0%	41.8%
	18, 20, 21	3	43.1%	42.8%	40.6%	40.8%
	10,13	3	44.5%	44.5%	43.2%	42.9%

Table 15: The monthly average MAPD and RMSPD are shown for the Forecasting E. and the three forecasting models for the four 15-minute values and their corresponding hours (16 15-minute values).

As can be seen in table 15 we have calculated the results for five different combinations of parameters. The combinations containing ID 17 are results obtained using the average wind speed from the SCADA data. Since SCADA data are real-time measurements we first have predicted the SCADA data for the test set based on using the forecasted u-vector and the v-vector. A neural network has been created to predict the SCADA data. The training and validation set have used the real time wind speed to train the forecasting model.

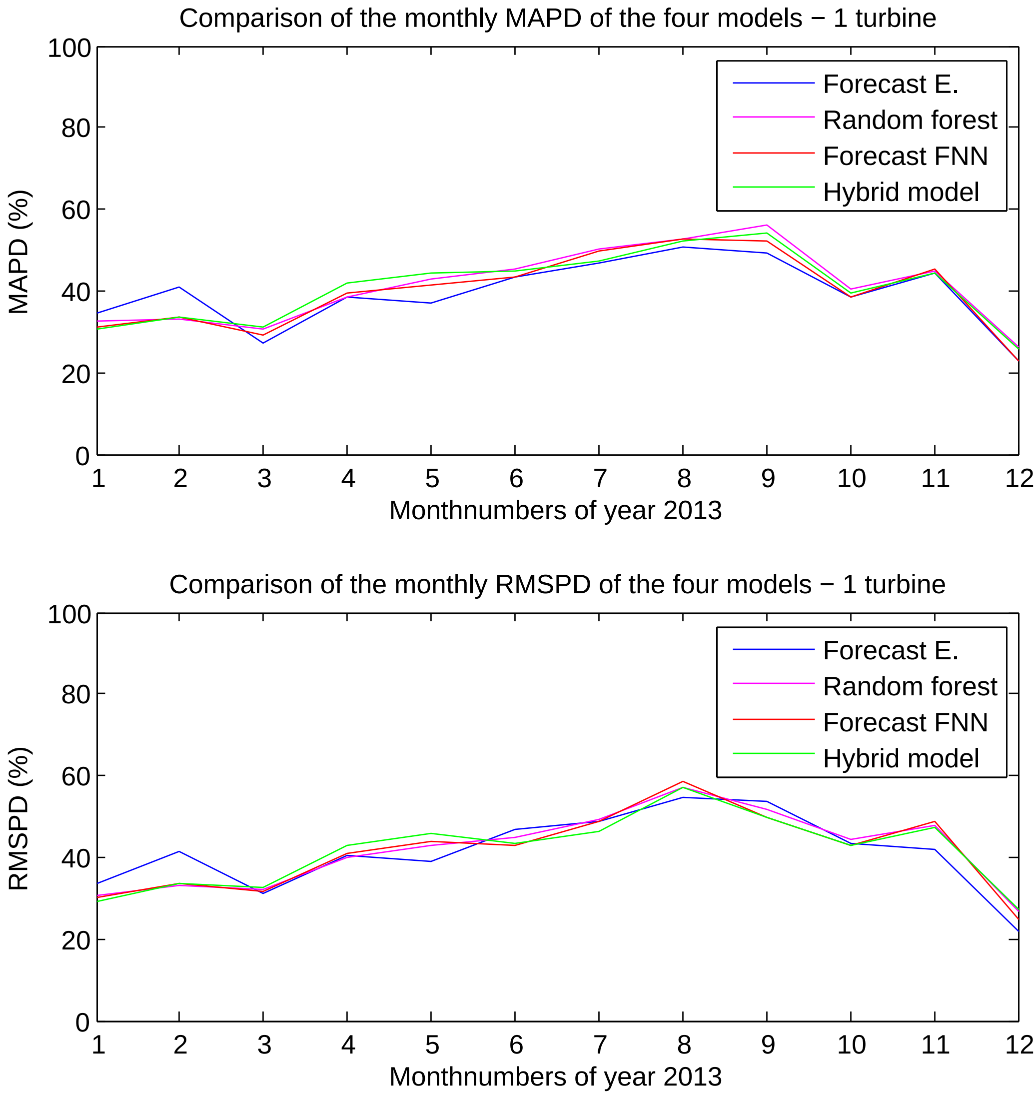


Figure 17: Best obtained monthly MAPD and RMSPD performances for the year 2013

As can be seen in table 15 the results from the three forecasting models show a similar pattern in case of the monthly MAPD. For all the three forecasting models, prediction of the SCADA data first improves the performance with about 2% compared to other parameter combinations. The pattern is visible for the four 15-minute values as well for the 16 15-minute values. The performances of the feed forward neural network, with a MAPD of 41.6% for four 15-minute values using parameter 17 and 40.0% for 16 15-minute values using parameters 17 and 18 comes closest to the performance of Forecasting E.

While looking at the RMSPD the performances of the three forecasting models remain stable for different combinations of input parameters. However the hybrid model using three clusters and the input parameters 18,20,21 performs equal compared to the Forecasting E on the four 15-minute values and performs even better on the 16 15-minute values with 40.8%. Since RMSPD gives more weight to larger errors we assume that the hybrid model can forecast large amounts of wind power slightly better than Forecasting E..

Besides only comparing the monthly average of our performance measurements we also have compared the monthly performances individually as is shown in figure 17.

While looking at the monthly MAPD results we see that our three forecasting models show equal to better results in the seasons winter and autumn. In the seasons spring and summer the performance errors are slightly higher compared to the Forecasting E. Looking at the monthly RMSPD results no pattern can be identified since the forecasting models and Forecasting E. fluctuate every month.

9.2 Wind farm of 27 Turbines

This section discusses the results obtained from a wind farm of 27 turbines located in the province Flevoland in the Netherlands. All these turbines are of the same size and for these turbines no SCADA data is available. The average monthly MAPD and RMSPD performance results are shown in table 16.

Model	**Parameters ID**	**Clusters**	**4 15-minute values**		**16 15-minute values**	
			MAPD	**RMSPD**	**MAPD**	**RMSPD**
External org.	-	-	37.0%	38.0%	35.0%	36.3%
Random forest	20,21	-	43.5%	42.8%	41.0%	41.1%
	18, 20, 21	-	44.0%	43.0%	41.0%	40.9%
Feed forward neural network	20,21	-	44.0%	43.0%	42.0%	40.5%
	18, 20, 21	-	44.0%	42.9%	41.0%	42.0%
Hybrid model	20,21	3	44.0%	44.0%	42.0%	42.3%
	18, 20, 21	3	43.0%	42.5%	41.0%	41.2%

Table 16: 27 Turbines: The monthly average MAPD and RMSPD are shown for the Forecasting E. and the three forecasting models for the four 15-minute values and their corresponding hours (16 15-minute values).

From table 16 we can see that no forecasting model shows similar performance compared to the Forecasting E. The three models however perform equally among each other. Performances which come the closest near the Forecasting E. are from the hybrid model and the Random forest.

To get a more detailed overview of the performances among the models we have figure 18 which shows the monthly MAPD and RMSPD of the 16-minute values in 2013. As one can see in the figure the MAPD and the RMSPD of the

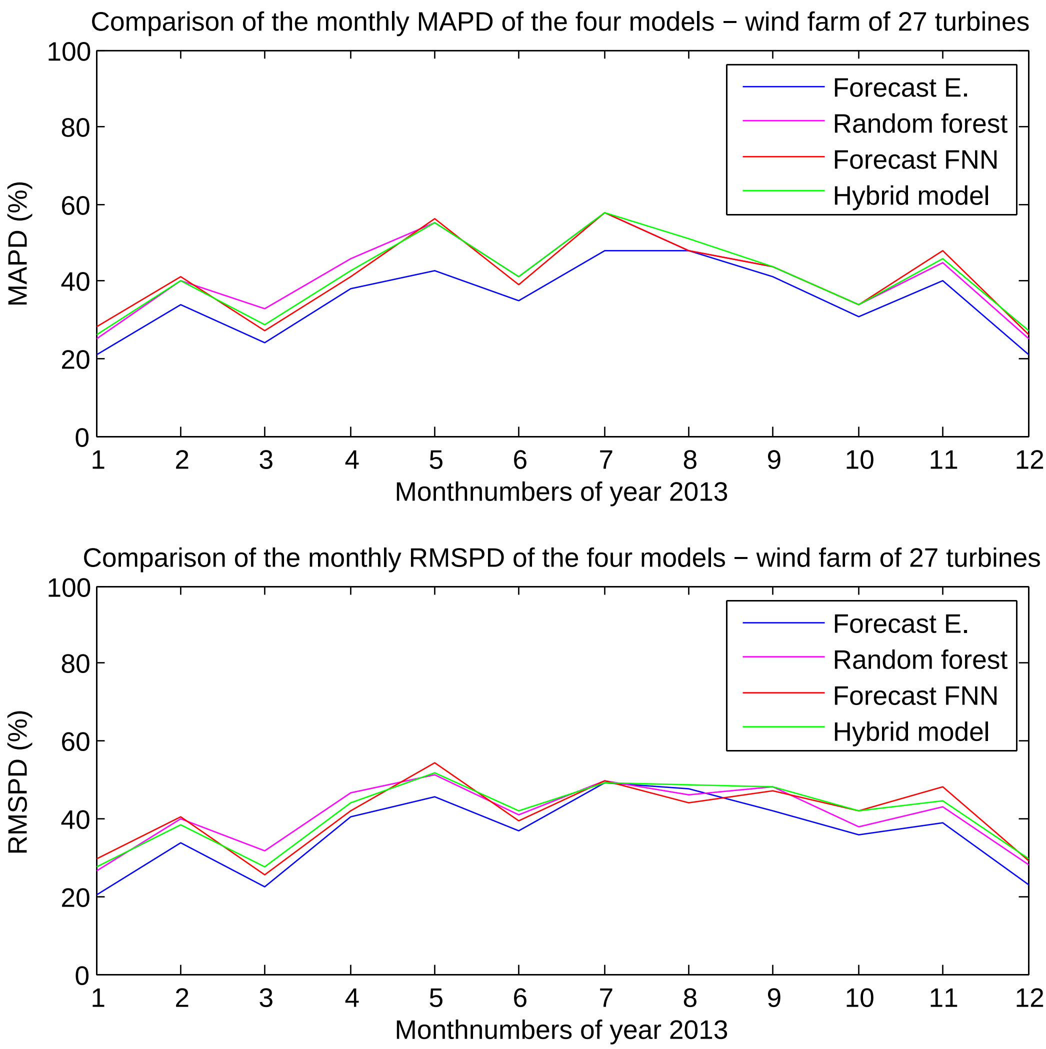


Figure 18: The monthly MAPD and RMSPD of the 27 turbines for 2013.

three forecasting models perform structurally worse compared to the forecasting E. except in the month August. The main reason is probably because the production of the turbines is the lowest in this time of the year as can be seen in figure 19 and is therefore easier to forecast. Another reason can be that the forecast delivered by the KNMI for August is more accurate compared to other months in 2013.

9.3 34 Turbines

This section discusses the results obtained from all the turbines. The locations of all the turbines are shown in figure 5 (section 6.1.1). The average monthly MAPD and RMSPD performance results are shown in table 17. As one can see none of the three forecasting models achieves the performance obtained by the

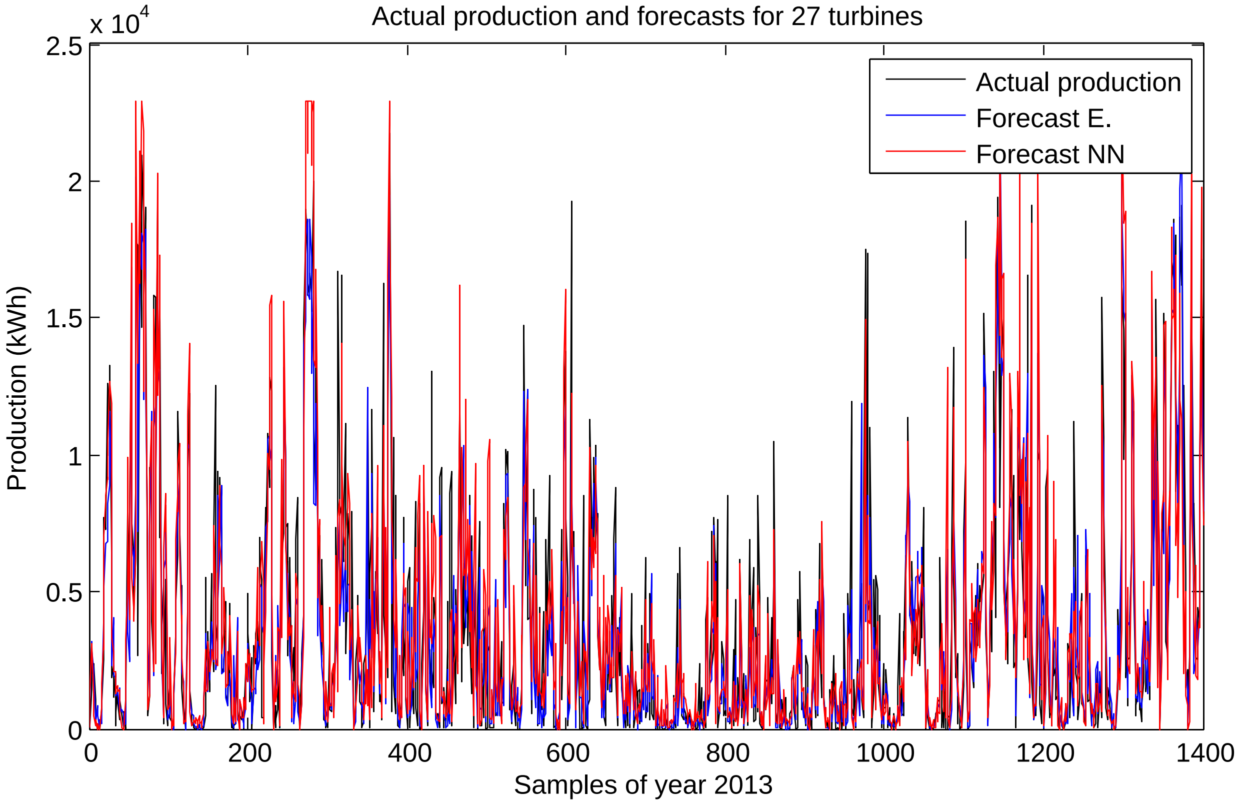


Figure 19: Production and forecasts of 2013. Samples 850 to 950 show the production and its forecasts for August 2013.

Forecasting E.

Figure 20 provides a more detailed overview of the months of the year 2013. As is shown in this figure also here the forecasting models perform structurally worse compared to forecasting E. except for the months October, November and for the random forest the month July.

9.4 Financial impact of one turbine

The actual production, the Forecasting E. and the Feed forward neural network using parameters 17 and 18 have been transformed into financial components. In figure 21 one can see the monthly profits using the forecast of one single turbine. The actual production is the profit line showing an ideal forecast having an error of 0%. The profits of the forecast wind power are calculated by multiplying the forecast against the spot price and add imbalance costs, which are calculated by taking the difference between the actual production and the forecast times the imbalance price.

As one can see having an ideal forecast of 0% gives almost structural (except for March 2013) more profit than having a forecast error in our case. Furthermore, comparing the profits of the forecasts we can find no structural pattern

Model	Parameters ID	Clusters	4 15-minute values		16 15-minute values	
			MAPD	RMSPD	MAPD	RMSPD
External org.	-	-	29.0%	30.7%	28.0%	29.4%
Random forest	20,21	-	33.0%	34.0%	32.0%	33.5%
	18, 20, 21	-	33.5%	34.1%	32.5%	33.4%
Feed forward neural network	20,21	-	34.0%	34.7%	34.0%	33.7%
	18, 20, 21	-	34.0%	34.0%	33.0%	34.0%
Hybrid model	20,21	3	36.0%	36.3%	35.5%	36.2%
	18, 20, 21	3	37.0%	37.8%	34.5%	36.0%

Table 17: 34 Turbines: The monthly average MAPD and RMSPD are shown for the Forecasting E. and the three forecasting models for the four 15-minute values and their corresponding hours (16 15-minute values).

in the figure. The difference between the total profit of both forecasts is about €100.

10 Discussion

In this section we will analyse our results from the previous section more briefly. As in the results section we divide this section into three subsections each providing a discussion about its size of turbines. We will take a close look why our forecasting models perform equally with the Forecasting E. for a single turbine and why the forecasting models perform worse for larger areas.

10.1 Single turbine

Based on our results we have seen that the two forecasting models feed forward neural network and the hybrid model perform the best compared to the Forecasting E..

Lets first take a look at the results of the 4 15-minute values. The best results have been obtained by the use of the SCADA data wind speed. Using the actual wind speed parameter from the SCADA data for the training and validation of our neural networks and predicted SCADA data for our test set has increased the performance with about 2% to 3% compared to the combination of other variables such as using the u-vector and v-vector.

The prediction of SCADA data for the test set has been conducted by creating a neural network being trained and validated using the u-vector and v-vector as input parameters and the actual SCADA wind speed as output parameter. Since the wind speed varies between 0 and 25 m/s the range of prediction wind speed compared to the production of wind power is much smaller and therefore we assume that the probability of large errors decreases. Furthermore, the prediction of SCADA data is valuable since it provides a description of the actual performance of the turbine as is also stated by Pinson and Kariniotakis

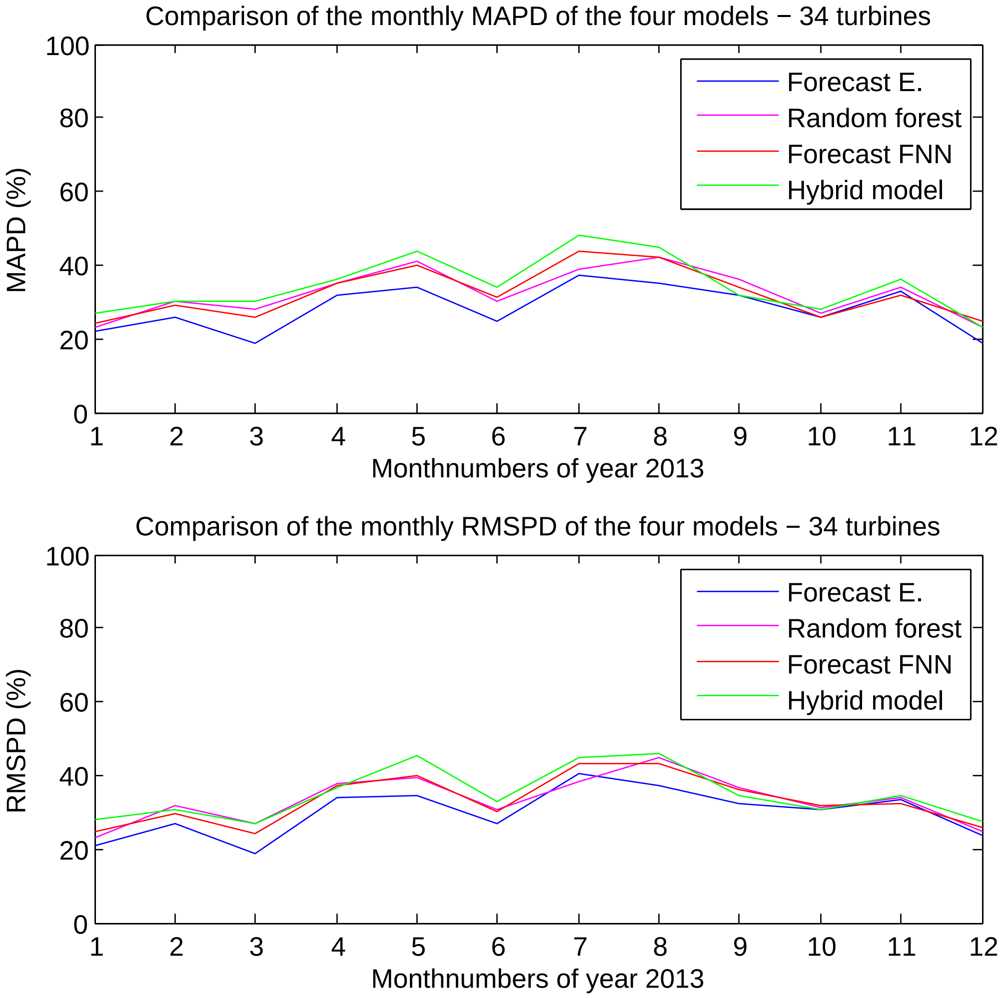


Figure 20: The monthly MAPD and RMSPD of all turbines for 2013.

[36]. This performance description might reduce the error and 'help' the neural network to understand which wind power output is related to the wind speed.

While looking at the 16 15-minute values we find our results obtained by the forecasting models more promising compared to the Forecasting E. Best results have been obtained by the feed forward neural network and the hybrid model using SCADA data and the input parameters u-vector, v-vector and wind speed time lag 1.

Through the year the MAPD and the RMSPD performance errors of the forecasting models converge to the performance error of the Forecasting E.. The hybrid model does show a RMSPD performance that is even better than the RMSPD performance error of the Forecasting E.. A possible reason why the RMSPD of the hybrid model performs better than the Forecasting E. is because our hybrid model is able to forecast wind power generated at high wind speeds.

Furthermore, our forecasting models converge on 16 15-minute values. This

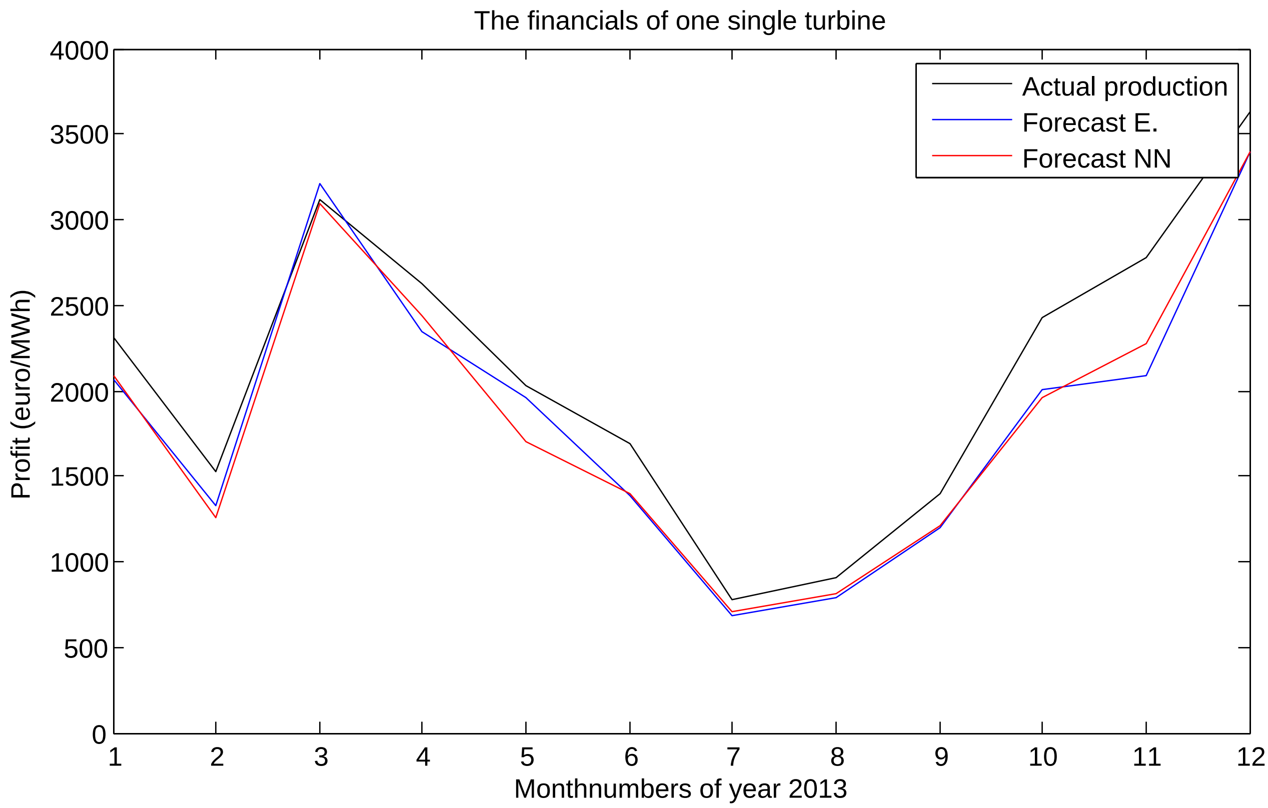


Figure 21: Profit per month in 2013. Actual production is the profit having a forecast error of 0%.

is probably because the Forecasting E. is providing forecasts in periods of 15-minutes, whereas our forecasting models provide total forecast over the past hours.

This means that the Forecasting E. is taking wind speeds and other effects into account at 15-minute basis, which can be misleading at high wind speeds. The reason behind this is that we assume that a turbine rotating at high wind speed does not easily slow down if the wind speed is decreasing for a few minutes. It might therefore be interesting to use hourly wind speed to forecast hourly production and divide this production by four if one requires 15-minute production.

10.2 Wind farm of 27 turbines

Unfortunately none of the forecasting models have performed well when forecasting the wind power generated by a wind farm of 27 turbines of the same size. The MAPD and RMSPD have been found higher for the four 15-minute values and the 16 15-minute values. There are some reasons to clarify these unfortunate results.

First of all, as explained in section 5.2.1 each wind turbine generates a certain amount of wind power based on the wind speed. This can be visualised into a power curve. The aggregation of turbines that use the same power curve can be transformed into a new aggregated power curve, where for each wind speed value the corresponding aggregation of the wind power at that specific wind speed can be calculated. As you can image at lower wind speeds the production remains low, but at higher wind speeds the generation of wind power is linearly increased by the multiplication of the turbines until their maximum power capacity, as can be seen in figure 22. The aggregation of turbines makes the difference between the generation of wind power at low wind speed and high wind speed x times higher compared to the the generation of wind power of one turbine. In this case we assume that all the turbines are equal and produce each of them exact the same amount of power for a specific wind speed. However this is not the case. Therefore it is more difficult to forecast wind power accurately.

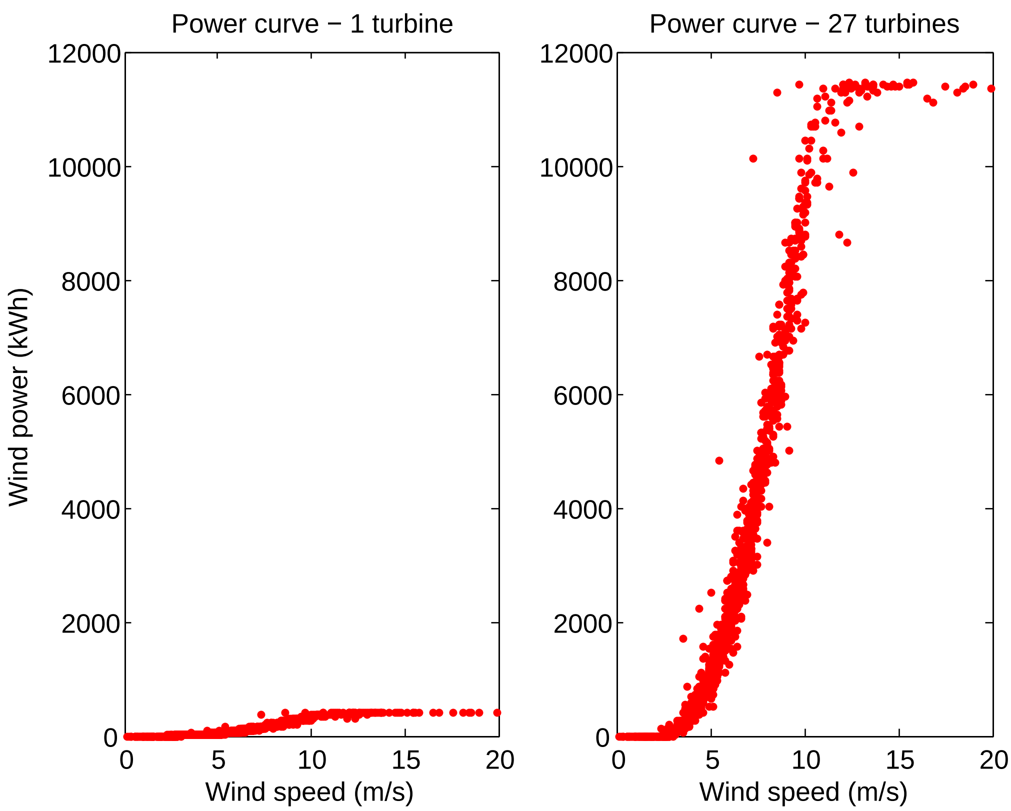


Figure 22: Power curve of one turbine (left) and the power curve of the aggregation of 27 turbines all of the same size (right)

A second reason is the forecast of wind power at high wind speed. The wind turbines have a maximum amount of power they can deliver in a certain amount of time (mostly expressed in maximum kWh per hour). The neural network however does not contain this kind of information and therefore will ignore this feature. This means at high wind speeds the neural network will forecast an amount of wind power which cannot even be produced by the wind turbines.

The third reason is that we found in our prediction results that at high wind speed values the maximum amount of wind power was not even reached by the turbines. The only reason we can think of in this case is that the performance of certain turbines are being affected by other turbines, such as shadowing effects (wind turbulence). This has also been mentioned by literature [23] [22]. We assumed that the forecasting models were able to identify the shadowing effects by using training and validation data. Unfortunately this was not the case.

Finally, as mentioned earlier the Forecasting E. was based on the MAPD and RMSPD able to predict better than our forecasting models. The reason we can find to substantiate this result is that the Forecasting E. has forecast each turbine individually, while we have forecast the aggregation of the turbines. We assume therefore that Forecasting E. does take more detailed information of the specific turbines, such as shadowing effects and their geographical location into account.

10.3 34 Turbines

The MAPD and RMSD results obtained for all the 34 turbines have also shown no fortunate results. All the forecasting models perform about 4% to 8% worse compared to the Forecasting E.. From the three models the Random Forest shows to perform the best among the three forecasting models, whereas the hybrid model performs worst. The reason why Random forest performs best is because it is a robust forecasting model and prevents overfitting.

Furthermore, the reasons mentioned in the previous section apply also on this set of turbines. However some additions have to be made here. The first reason, multiplication of the power curve from the previous section does apply on the a set of turbines having the same height. In case of using all the turbines, different power curves have to be taken into account since we are dealing with different turbine sizes.

To clarify the second reason from the previous section we have figure 23 that shows the effect of a neural network which is at a few time samples forecasting higher wind power than actually is allowed by the turbines at such wind speed. For example at time sample 100 as one can see the forecast of the neural network using the hybrid model is rising above the maximum capacity of all the turbines in one hour (which is 48760 kWh). To prevent these surprising outliers we have adjusted our prediction set in such way that when the prediction exceeds its maximum capacity we modify the value to its maximum. One can see this as a form of a fuzzy rule to optimize the performance of the forecasting models.

10.4 Input parameters

Forecasting the wind power of one turbine has been conducted by using five different combinations of input parameters. The input parameters mostly recommended in literature to forecast wind power are wind speed and wind direction. We have taken these parameters into account, but we have also used the

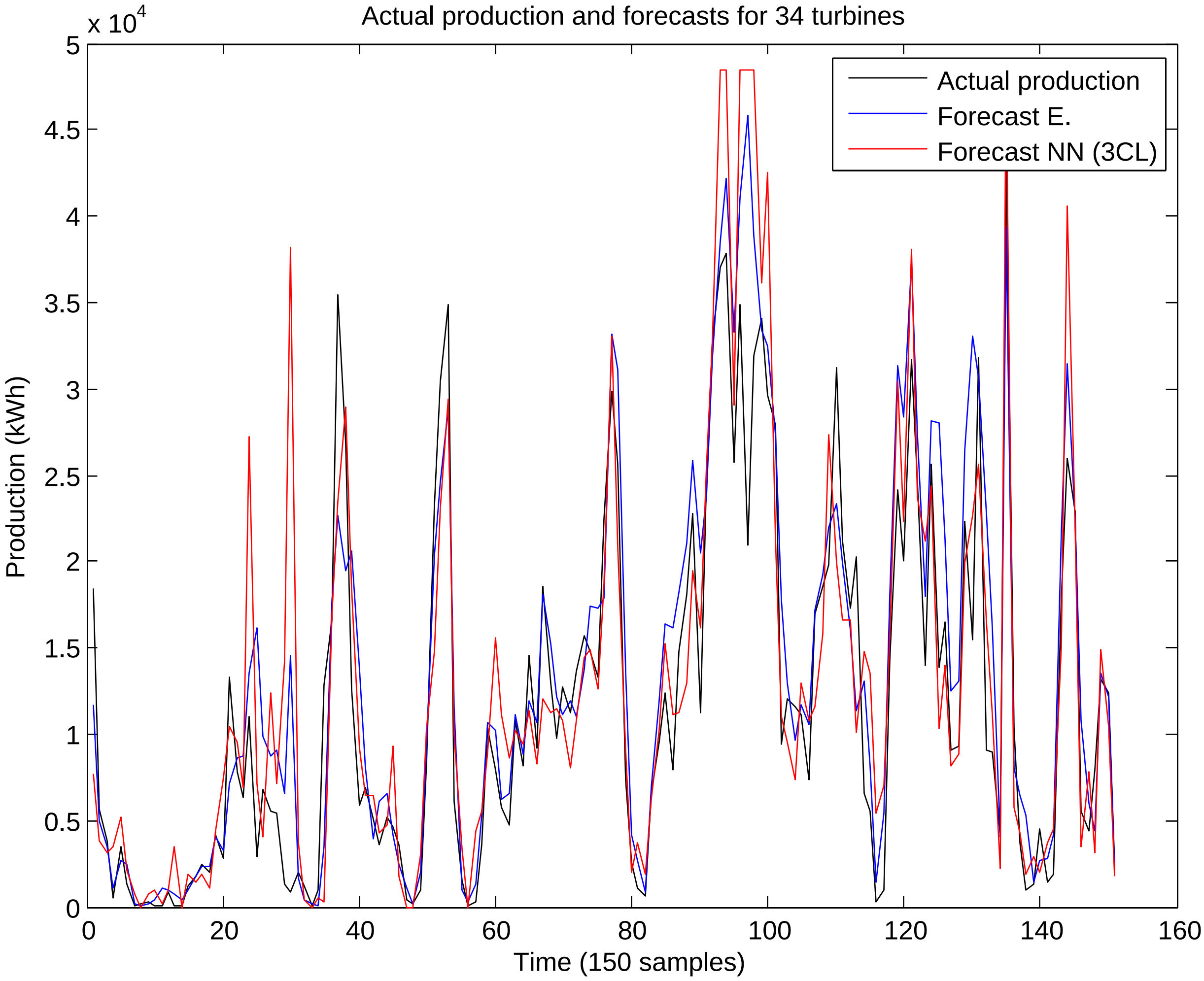


Figure 23: Production and forecasts of a part of 2013 (16 15-minute values have been used here).

u-vector and the v-vector as input parameters. These two vectors are needed to calculate the wind speed and wind direction as explained in section 6.1.1.

The reason why we have used the u-vector and v-vector as input parameters is because calculating wind speed and wind direction first could result in that certain information is being lost. This is similar to for example rounding a double in the middle of an equation.

Based on the results from a single turbine (table 15) we find that using the u-vector and v-vector as input parameters perform slightly better than the input parameters wind- speed and direction. The random forest however performs almost equal for the four 15-minute values and slightly worse for the 16 15-minute values. Based on these results of the neural networks we assume that there might get information lost when calculating the wind speed and direction. Therefore we have not used the input parameters wind speed and wind direction to forecast wind power generated by a wind farm.

At last we discuss the input parameter 'wind speed time lag 1'. The reason using the time lag parameter is to identify trends in the wind speed data and therefore to forecast better than using only the u-vector and the v-vector.

The results are fortunate for a single turbine. The time lag parameter improves the performance with about 1% using a feed forward neural network for four 15-minute values and 3% of the hybrid model for 16 15-minute values. Based on these results we assume that the neural networks can identify wind speed trends and are capable to forecast the wind power better.

The results are unfortunate for the wind farm and for the aggregation of all the turbines. Including the time lag parameter slightly decreases the performance of the forecasting models. A possible reason is because the wind speed time lag 1 has a different effect on each of the turbines and therefore no relation can be found. Therefore it is for a neural network hard to determine how many turbines are being influenced by this time lag.

11 Conclusion

In our introduction we have stated four research questions. For each research question we explain the results and conclude with the answer on the research question.

RQ1 Which factors and input parameters to predict wind power have been described in literature? And which of those have been found successful?

A literature study has been conducted to identify factors and input parameters to predict wind power. The literature has identified three factors of importance which are, the use of different data sources, such as meteorological, SCADA or LIDAR data sources, the aggregation of turbines (different grid sizes) and the geographical location of the turbines. Furthermore, literature has identified a large set of input parameters to predict wind power. These input parameters are, wind speed, wind direction, weather stability, availability, relative humidity, seasonal patterns, temperature and pressure. Of all those input parameters literature has found no improvement of the performance of the forecasting model when using the temperature and/or pressure.

In our research we have used the meteorological data source HiRLAM and SCADA data source. Furthermore, we have forecast wind power for one single turbine, a wind farm consisting of 27 turbines and the complete set of 34 turbines. The geographical location has been taken into account, by selecting the four grid points surrounding the turbines. Geographical obstacles have not been take into account.

Based on the correlation coefficient we have selected the input parameters wind speed with a correlation value of approximately 0.85 and wind direction with a correlation value between 0.08 and 0.16 for different locations. Since wind speed and wind direction use the u-vector and the v-vector as input parameters we also have used these two vectors as input parameters. Furthermore based on the results obtained using the autocorrelation coefficient we have selected the input parameter wind speed with a time lag 1. The correlation value of the time lag 1 was about 0.75.

The availability has not been taken into account when forecasting wind power, because the Forecasting E. Raedthuys is using does also not take the availability of the turbines into account.

Finally, we have not selected the parameters air density and relative humidity since these two have shown a negative correlation value with the wind power.

RQ2 Which forecasting models have been found the most relevant by previous literature to predict wind power generated by wind turbines?

Many different forecasting models have been researched. The forecasting models which have been researched most are time series models, neural networks, regression trees or SVR. In our research we have used the feed forward neural network, a random forest model and a hybrid model to forecast wind power generated by turbines. The results obtained by these three forecasting models have been compared with the forecast from the external organization.

RQ3 How do the recommended forecasting methods identified in literature studies perform to the forecasts provided by external organizations?

In our research we have tested our forecasting models on three aggregation areas of turbines. The results obtained for a single turbine are promising. The best MAPD performance was achieved using the Feed Forward neural network with a performance of 40.0% which is 0.6% worse compared to the Forecasting E. (being 39.4%) for the 16 15-minute values. In case of the RMSPD, the best performance was achieved using the hybrid model with a 40.8% which is 0.5% better compared to the RMSPD results of the Forecasting E. (being 41.3%). Furthermore, based on monthly performances the MAPD results have shown that our forecasting models show equal to better results in the seasons winter and autumn.

Unfortunate results have been found when forecasting wind power generated by a wind farm and forecasting wind power for all the turbines. In both cases the performance errors MAPD and RMSPD results performed about 4% to 8% worse compared to the Forecasting E. The forecasting models did perform equally among each other.

To clarify our results we have stated the fourth research question.

RQ4 Which input parameters and optimizations have to be applied on the recommended forecasting models to achieve an as accurate prediction model compared to the forecasting model from the external organizations?

Based on our discussion, results and the answer on the previous research question we come to the conclusion that we are able to forecast wind power generated by a single turbine. The input parameters u-vector and v-vector are the most important parameters to forecast wind power, because these parameters are used to forecast SCADA data, wind speed and wind direction. Therefore, the u-vector and the v-vector have shown their contribution to obtain the best RMSPD performance of 40.8% and have indirectly by forecasting SCADA data obtained the best MAPD performance of 40.0%.

Based on the poor performance (answer on research question three) obtained by forecasting wind power generated by a wind farm we can conclude that forecasting the aggregation of wind power of a wind farm does require further research. The input parameters u-vector, v-vector and wind speed time lag 1 were not able to perform at least equally to the Forecasting E.. Also applying clustering from our hybrid model or the use of a robust random forest did not show any improvement.

Several reasons have been mentioned in the discussion that can clarify the performance of forecasting wind power generated by the aggregation of turbines. The power curve of the aggregated set of turbines is becoming more steeper when the wind speed increases. In this case we assume all the turbines of the wind farm perform equally, while this is not the case. Furthermore, the neural networks are not aware of the fact that wind turbines can produce a maximum amount of energy for a specific wind speed. An optimization for both reasons is to apply fuzzy rules. The only two fuzzy rules applied in this research are setting the minimum amount of energy forecasting on zero and the maximum amount on the maximum of the aggregation of turbines. Finally, an optimization which is required is taking into account turbines that are possible being affected by other turbines (shadowing effect). The forecasting models were not able to identify this shadowing effect.

Altogether this research has shown that the the three forecasting models each have shown their purposes. The feed forward neural network and the hybrid model show similar performances with the Forecasting E. when forecasting wind power generated by a single turbine. Whereas the random forest shows its robustness on the forecast of wind power generated by the aggregation of turbines. Based on this research we recommend in the first place to develop forecasting models for turbines on individual level, as long no improvement has been found on the aggregation of wind turbines.

12 Future work

A lot of research has been conducted on the forecast of wind power. This research has shown its contribution on day-ahead forecasting wind power generated by turbines in the Netherlands. However further research is required to improve forecasting wind power generated by turbines.

First of all, feed forward neural networks using input parameters u-vector and v-vector have shown satisfying results. However other forecasting models should be considered, such as Recurrent neural networks or support vector machine, while using the same input parameters.

Secondly, in our research we were hoping to capture seasonal patterns by using the clustering step of the hybrid model. Other techniques to include seasonal patterns can be applied, such as: clustering based on season or including the season as input parameter.

Thirdly, when applying the aggregation of turbines one should take into account the shadowing effects of turbines and possible geographical obstacles. This can be done by first identifying which turbines are being affected, based on for example comparing the production from the turbines located on the edge of the wind farm with the production from the turbines located in the middle of the wind farm.

Fourthly, using fuzzy rules can be useful when forecasting wind power generated by the aggregation of turbines. Over fitting the model can therefore be prevented. Another solution is using the power curve to get an indication about the production of wind power and use this value as input parameter.

Fifthly, in our research we have only used the meteorological data source HiRLAM. There are however other meteorological data sources such as ECMWF.

Using a combination of those data sources could give more insight on the accuracy of both data sources.

Sixthly, in our research we have not used LIDAR data as data sources because it was not yet available. It is recommended to conduct future research using also the LIDAR data as data source.

Finally, in this research we have taken only one time lag as input parameter into account. It might therefore be interesting to include two or more time lag and compare results.

References

[1] T.G. Barbounis and J.B. Theocharis. Locally recurrent neural networks for long-term wind speed and power prediction. *Neurocomputing*, 69(46): 466 – 496, 2006. ISSN 0925-2312. URL http://www.sciencedirect.com/science/article/pii/S0925231205000408.

[2] AJ Brand, M Gibescu, and WW de Boer. Variability and predictability of large-scale wind energy in the netherlands. In *Wind Power*. Croatia: InTech, 2010.

[3] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[4] Erasmo Cadenas and Wilfrido Rivera. Wind speed forecasting in the south coast of oaxaca, mxico. *Renewable Energy*, 32(12):2116 – 2128, 2007. ISSN 0960-1481. URL http://www.sciencedirect.com/science/article/pii/S0960148106002801.

[5] L. J. Cao and F.E.H. Tay. Support vector machine with adaptive parameters in financial time series forecasting. *Neural Networks, IEEE Transactions on*, 14(6):1506–1518, Nov 2003. ISSN 1045-9227.

[6] Centraal Bureau voor de Statistiek. Elektriciteit; productie per energiebron, 2014. URL http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=80030NED&D1=0,3&D2=0&D3=a&D4=8-14&HD=140304-1229&HDR=T,G1,G3&STB=G2. Visited on 2014-03-07.

[7] A Clifton, L Kilcher, J K Lundquist, and P Fleming. Using machine learning to predict wind turbine power output. *Environmental Research Letters*, 8(2):024009, 2013. URL http://stacks.iop.org/1748-9326/8/i=2/a=024009.

[8] M. Daneshdoost, M. Lotfalian, G. Bumroonggit, and J. P. Ngoy. Neural network with fuzzy set-based classification for short-term load forecasting. *Power Systems, IEEE Transactions on*, 13(4):1386–1391, Nov 1998. ISSN 0885-8950.

[9] DeEnergieGids.nl. Marktpartijen, 2012. URL http://www.deenergiegids.nl/marktpartijen/Marktpartijen-in-de-Energiemarkt.aspx. Last visited on: 2014-01-31.

[10] Shu Fan, James R Liao, Ryuichi Yokoyama, Luonan Chen, and Wei-Jen Lee. Forecasting the wind generation using a two-stage network based on meteorological information. *Energy Conversion, IEEE Transactions on*, 24 (2):474–482, 2009.

[11] Ulrich Focken, Matthias Lange, Kai Mnnich, Hans-Peter Waldl, Hans Georg Beyer, and Armin Luig. Short-term prediction of the aggregated power output of wind farmsa statistical analysis of the reduction of the prediction error by spatial smoothing effects. *Journal of Wind Engineering and Industrial Aerodynamics*, 90(3):231 – 246, 2002. ISSN 0167-6105. URL http://www.sciencedirect.com/science/article/pii/S0167610501002227.

[12] Lionel Fugon, Jérémie Juban, Georges Kariniotakis, et al. Data mining for wind power forecasting. In *Proceedings European Wind Energy Conference & Exhibition EWEC 2008*, 2008.

[13] Gasunie transport services B.V. About pos and sbs, 2014. URL http://www.gasunietransportservices.nl/en/transportinformation/about-balancing/about-pos-and-sbs. Last visited on: 2014-01-31.

[14] J. Miettinen H. Holttinen and S. Sillanpaa. Wind power forecasting accuracy and uncertainty in finland, 2013.

[15] Richard M Heiberger and Burt Holland. *Statistical analysis and data display: an intermediate course with examples in S-plus, R, and SAS*. Springer, 2004.

[16] Christoffer Jonsson. Statistical analysis of winddata regarding long-term correction. 2010.

[17] Rajesh G. Kavasseri and Krithika Seetharaman. Day-ahead wind speed forecasting using f-arima models. *Renewable Energy*, 34(5):1388 – 1393, 2009. ISSN 0960-1481. URL http://www.sciencedirect.com/science/article/pii/S0960148108003327.

[18] S Upadhye PG Khot. Identification of outliers by cook's distance in agriculture datasets. *International Journal Of Engineering And Computer Science ISSN:2319 - 7242*, pages 2045–2049, 2013.

[19] MakarandA. Kulkarni, Sunil Patil, G.V. Rama, and P.N. Sen. Wind speed prediction using statistical regression and neural network. *Journal of Earth System Science*, 117(4):457–463, 2008. ISSN 0253-4126.

[20] Andrew Kusiak, Haiyang Zheng, and Zhe Song. Wind farm power prediction: a data-mining approach. *Wind Energy*, 12(3):275–293, 2009. ISSN 1099-1824. doi: 10.1002/we.295. URL http://dx.doi.org/10.1002/we.295.

[21] Soon-Duck Kwon. Uncertainty analysis of wind energy potential assessment. *Applied Energy*, 87(3):856 – 865, 2010. ISSN 0306-2619. URL http://www.sciencedirect.com/science/article/pii/S0306261909003717.

[22] DOE/Lawrence Livermore National Laboratory. Wind turbines: In the wake of the wind, 2014. URL www.sciencedaily.com/releases/2011/04/110426151040.htm. Visited on: 2014-03-13.

[23] L. Landberg. Short-term prediction of the power production from wind farms. *Journal of Wind Engineering and Industrial Aerodynamics*, 80(12): 207 – 220, 1999. ISSN 0167-6105. URL http://www.sciencedirect.com/science/article/pii/S0167610598001925.

[24] Lars Landberg. A mathematical look at a physical power prediction model. *Wind Energy*, 1(1):23–28, 1998. ISSN 1099-1824.

[25] Lars Landberg and SimonJ. Watson. Short-term prediction of local wind conditions. *Boundary-Layer Meteorology*, 70(1-2):171–195, 1994. ISSN 0006-8314.

[26] Daniel T Larose. *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2005.

[27] Ma Lei, Luan Shiyan, Jiang Chuanwen, Liu Hongling, and Zhang Yan. A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, 13(4):915 – 920, 2009. ISSN 1364-0321. URL `http://www.sciencedirect.com/science/article/pii/S1364032108000282`.

[28] Roger Levy. Probabilistic models in the study of language. *Ms, University of California, San Diego*, 2010.

[29] Andy Liaw and Matthew Wiener. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[30] M. Carolin Mabel and E. Fernandez. Analysis of wind power generation and prediction using ann: A case study. *Renewable Energy*, 33(5):986 – 992, 2008. ISSN 0960-1481. URL `http://www.sciencedirect.com/science/article/pii/S0960148107002261`.

[31] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441, 1963.

[32] M.A. Mohandes, T.O. Halawani, S. Rehman, and Ahmed A. Hussain. Support vector machines for wind speed prediction. *Renewable Energy*, 29(6): 939 – 947, 2004. ISSN 0960-1481. URL `http://www.sciencedirect.com/science/article/pii/S0960148103003860`.

[33] Mohamed A Mohandes, Shafiqur Rehman, and Talal O Halawani. A neural networks approach for wind speed prediction. *Renewable Energy*, 13(3): 345–354, 1998.

[34] NUON. Netbeheerder, 2014. URL `http://www.nuon.nl/energie/energieprijzen/netbeheerder.js`. Last visited on: 2014-01-31.

[35] D.C. Park, M.A. El-Sharkawi, II Marks, R.J., L.E. Atlas, and M.J. Damborg. Electric load forecasting using an artificial neural network. *Power Systems, IEEE Transactions on*, 6(2):442–449, May 1991. ISSN 0885-8950.

[36] P. Pinson and G.N. Kariniotakis. Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment. In *Power Tech Conference Proceedings, 2003 IEEE Bologna*, volume 2, pages 8 pp. Vol.2–, June 2003.

[37] Pierre Pinson and Georges Kariniotakis. On-line assessment of prediction risk for wind power production forecasts. *Wind Energy*, 7(2):119–132, 2004.

[38] Raedthuys BV. Business units, 2013. URL `http://www.raedthuys.nl/organisatie/business-units.html`. Visited on 2014-02-20.

[39] Samaneh Sabetghadam, Farhang Ahmadi-Givi, and Yahya Golestani. Visibility trends in tehran during 19582008. *Atmospheric Environment*, 62 (0):512 – 520, 2012. ISSN 1352-2310. doi: http://dx.doi.org/10.1016/j.atmosenv.2012.09.008. URL `http://www.sciencedirect.com/science/article/pii/S1352231012008679`.

[40] P. SangitaB and S.R. Deshmukh. Use of support vector machine for wind speed prediction. In *Power and Energy Systems (ICPS), 2011 International Conference on*, pages 1–8, Dec 2011.

[41] T. Senjyu, A. Yona, N. Urasaki, and T. Funabashi. Application of recurrent neural network to long-term-ahead generating power forecasting for wind power generator. In *Power Systems Conference and Exposition, 2006. PSCE '06. 2006 IEEE PES*, pages 1260–1265, Oct 2006.

[42] G. Sideratos and N.D. Hatziargyriou. An advanced statistical method for wind power forecasting. *Power Systems, IEEE Transactions on*, 22(1): 258–265, Feb 2007. ISSN 0885-8950.

[43] S.S. Soman, H. Zareipour, O. Malik, and P. Mandal. A review of wind power and wind speed forecasting methods with different time horizons, 2010.

[44] J.W. Taylor, P.E. McSharry, and R. Buizza. Wind power density forecasting using ensemble predictions and time series models. *Energy Conversion, IEEE Transactions on*, 24(3):775–782, Sept 2009. ISSN 0885-8969.

[45] TenneT Holding B.V. About the electricity grid, 2014. URL `http://www.tennet.eu/nl/grid-projects/gridmap.html`. Last visited on: 2014-01-31.

[46] TenneT Holding B.V. Programmaverantwoordelijkheid, 2014. URL `http://www.tennet.eu/nl/nl/klanten/diensten/systeemdiensten/programmaverantwoordelijkheid.html`. Last visited on: 2014-01-31.

[47] R. Wagner, M. Courtney, J. Gottschall, and P. Lindelw-Marsden. Accounting for the speed shear in wind turbine power performance measurement. *Wind Energy*, 14(8):993–1004, 2011. ISSN 1099-1824.

[48] Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schrödl, et al. Constrained k-means clustering with background knowledge. In *ICML*, volume 1, pages 577–584, 2001.

[49] Xiaochen Wang, Peng Guo, and Xiaobin Huang. A review of wind power forecasting models. *Energy Procedia*, 12(0):770 – 778, 2011. ISSN 1876-6102. URL `http://www.sciencedirect.com/science/article/pii/S1876610211019291`. The Proceedings of International Conference on Smart Grid and Clean Energy Technologies (ICSGCE 2011).