Examining the Impact of Features in Short-Term Wind Power Prediction using Machine Learning

A thorough comparison of how different input patterns influence prediction accuracy for both wind turbine and wind park power prediction

Rune Skogstø Bryne

Thesis submitted for the degree of
Master in Network and system administration
60 credits

Department of Informatics
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

Spring 2019

Examining the Impact of Features in Short-Term Wind Power Prediction using Machine Learning

A thorough comparison of how different input patterns influence prediction accuracy for both wind turbine and wind park power prediction

Rune Skogstø Bryne

Examining the Impact of Features in Short-Term Wind Power Prediction using Machine Learning

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

Precise wind power output predictions is of great importance for a successful integration of wind power into the power grid and market operations. A popular approach to short-term wind power prediction is to use machine learning methods. Machine learning methods are data-drove, and the data it utilize as input is significant to prediction performance. There exits many proposed models in literature for short-term wind power prediction, but a variation in choice of input is apparent. Particularly a difference in the use of forecast weather data from numerical weather predictions models is observed.

In this thesis a wide variety of input types is examined by measuring their effect on prediction accuracy using a state of the art method for short-term wind power prediction. The inputs examined stems from an in-depth review of previous work on the subject. In addition, some input types not commonly utilized is considered. In order to give a general account on the importance of the specific feature types, experiments have been done for two wind turbines in separate wind parks located in Scandinavia using real world data measurements and correlating weather forecast data from a publicly available numerical weather prediction model. Further, in addressing the challenge of battling computation costs, this work compares two strategies for predicting the wind power output for the whole wind park. In the first, and expensive strategy, separate models are developed for each turbine, and the final wind park power prediction is obtained by summarizing the individual predictions from each turbine. The second strategy treats the wind park as a single entity by aggregating the input data from all turbines before model development.

Acknowledgments

First and foremost, I would like to thank my supervisor, Frank Eliassen, for his valuable guidance and patience throughout this process. I would also like to thank Arno Jacobsen his much appreciated input and advice along the way.

Statkraft have allowed me to use their data in my research, and for this I am very grateful.

Finally, I want to thank my partner Hanne for tirelessly supporting and motivating me throughout this challenging endeavour.

Contents

List of Figures

List of Tables

Part I

Introduction

Chapter 1

Introduction and Overview

Reliable short-term wind power predictions are important for a successful integration of wind power into the electricity grid. Popular approaches for this task are machine learning methods. The success of such methods depends greatly on the input they utilize. In this thesis we examine a variety of both commonly and rarely used input types in literature in order to measure their impact on prediction accuracy for both wind turbines and wind parks.

1.1 Motivation

There is an increased demand for clean and renewable energy across the globe. This was demonstrated with over 190 countries committing to the Paris agreement back in 2016. Dire reports from the United Nations climate panel tells us that renewable energy needs to make up somewhere between 70 to 85 percent of the worlds energy sources by 2050, in order to battle climate change [49]. It is expected that wind power will play a major role in this transition. Moving from dirty energy sources to wind is undoubtedly positive for the climate, but it also raises challenges which traditional energy sources, such as coal and oil, did not have to consider. Wind power is not a stored energy, and cannot be ramped up and down on demand. It exists fully at the mercy of the weather. If the wind does not blow, energy is not produced. This brings challenges to trade, which in many markets happens prior to production. In order to effectively trade energy, producers need reliable power predictions. An inaccurate power prediction might cause a producer to sell more that it can produce. Further, an increased integration of wind energy into the electricity grid introduces challenges to transmission system operators (TSO). At any time, the grid has to maintain a balance between energy generation and consumption. A large imbalance can in the worst case cause a blackout. The TSOs impose reporting requirements on expected energy generation from producers. Inaccurate reports are usually penalized in form of cost by the TSO. As more and more energy comes from non-dispatchable energy sources, such as wind, it is clear that balancing the grid will become harder, and more emphasis will be put on precise predictions.

In this context, better predictions will ease operations and boost revenue for the players in the energy market. Several methods exit for doing short term wind power prediction. A popular approach is the use of machine learning (ML) algorithms, which in literature has generally shown good performance when compared to the more traditional models. A vast number of ML-based models have been proposed through research in the quest for improving forecast accuracy. ML-algorithms are data-dependent and need to be trained. Consequently, the quality and selection of data used in developing ML-based models are highly important. In this scenario the question of which data should be used as input quickly pops up. In machine learning jargon, the specific inputs to a model are often referred to as a feature, and models for short-term wind power prediction proposed in literature utilize varying feature sets. Although present wind power is almost exclusively utilized as a feature, many models include additional data types such as wind speed, wind direction and other less obvious attributes. Other work considers lagged values as input, which are measurements of some attribute recorded prior to the time of prediction (e.g wind power measurement one and two hours ago), in addition to the most recent recordings. Finally, some models also consider data obtained from external numerical weather prediction (NWP) models as input, although this is mostly used when we want to predict several hours into the future. A tempting approach to chosen features is to just input all of them, but uncritically increasing the feature space can cause a model to become too complex and as a consequence, perform worse. It is therefore of interest to examine how different data types effect prediction accuracy, in order to discover which features are most useful for doing short-term wind power prediction. Further, a large feature set increases computation cost. In a real world application, new data becomes available continuously and predictions need to be done and reported on the fly. There is thus an incentive to have a time efficient model. Besides being conservative with the number of features, one way to lower computation cost is to treat the wind park as one entity, which means we only need to do a single prediction when predicting the power output for a wind park, as opposed to predicting each turbine individually. Both of these approaches can be found in literature.

1.2 Scope of the thesis

This work concerns short-term wind power prediction for two wind parks in the Nordic region. The choice of focusing on short-term predictions stems from considering the existing requirements in the Nordic energy domain. From talking to a commercial player in this market, short-term predictions was recognized as the most critical time frame. One hour ahead predictions was especially highlighted, since it addresses the need for balancing reporting to the responsible TSO. More specifically, this work looks at one to four hour ahead predictions. The Nordic market works on a 60 minute resolution, and as a consequence sub-hour prediction horizon is not addressed in this work.

The main objective of this thesis is to acquire knowledge about which features should be included for short-term wind power prediction by investigating a range of possible input configurations and their effect on prediction accuracy. This work considers hourly wind power predictions up to four hours into the future. It examines both frequently used and uncommonly used features in literature. From our perspective, a comprehensive comparison of feature importance is lacking in research. Knowledge on which features improve accuracy in prediction is useful several ways. Firstly, data acquisition and formatting can be time consuming. If, as an example, knowing that using data from NWP models as input on short-term predictions does not improve accuracy of any significance, then this is time saved from gathering the data and testing yourself. Secondly, it can expose the effectiveness of less obvious features which might not be considered by other developers of ML-models. Further, as briefly mentioned above, using a large set of features increases the complexity of the model, which can pose serious energy and time costs on optimization and training of the model. This work sets out to give a general account on which features should be considered when doing short-term wind power prediction.

A secondary objective focuses on the need for time-efficient models. In most cases we are only interested in the power output of the whole park, and not for a single turbine. One way to approach this is to aggregate the data from all wind turbines before deploying a single prediction model. In another, and more costly approach we would predict each turbine individually and add them up to get the final estimated power output. In this thesis we compare these two strategies for predicting the wind power output of a whole wind park. Finding out if, or how much, we lose in prediction accuracy by doing it the simple way is useful in cases where there are high constraints on computation time.

1.2.1 Research Questions

Specifically, this works seeks to answer the following questions:

RQ1: Which measured attributes should be included into the feature space for short-term wind power prediction?

RQ2: Should lag values be included into the feature space for short-term wind power prediction? In which case, how far back?

RQ3: Does the inclusion of data from neighbouring turbines improve prediction accuracy for a single turbine? In which case, how many neighbours should be included?

RQ4: Should NWP data be included into the feature space for short term wind power prediction? If so, which attributes?

RQ5: How does prediction of wind park power based on aggregated data compare, in regard to prediction accuracy, to predicting wind power of turbines individually and summing them up?

RQ1 regards the most recent measurement record at each turbine as input. In addition, wind direction measurements are examined, which are recorded at park level. The candidate features analysed in addressing this question are wind power, wind speed, temperature, wind direction and yaw. Besides yaw, these do all appear frequently in literature.

RQ2 concerns the inclusion of lag values into the feature and its effect on prediction performance. To set a limit, this work considers lag values for up to four hours back in time.

RQ3 concerns the use of measured data from neighbouring turbines in addition to the turbine in question as input. It further examines if increasing the number of neighbors considered has a favorable effect on prediction accuracy. We limit the scope to only considering the most significant features from the neighbor turbines, which is past wind power and wind speed.

RQ4 examines the effect of including NWP data from an external source into the feature set. The attributes examined are wind speed, wind speed of gust, temperature, pressure, humidity and wind direction.

The last question looks at how doing a single prediction on the aggregate of a wind park data performs when compared to predicting the power for each individual turbine and summing them.

1.2.2 Approach

The choice of features examined is in this works is inspired by an extensive survey done on proposed models in literature. In particular, the differences exposed in using NWP data versus excluding NWP data for short-term power predictions sparked curiosity, in addition to the inclusion, or exclusion of other attributes.

In order to evaluate the effect of the different features, a prediction approach utilizing a mutual information method to do an initial ranking of the candidate inputs was used to give an order on how candidate features are added to the different input patterns evaluated. This was in order to avoid having to test all possible combinations of feature candidates. The machine learning method used for prediction in this work is the support vector machine (SVM). Steps have been taken to optimize the prediction models developed by finding suitable hyperparameters to the SVM by using k-fold cross validation. Finally the different models developed, based on their input pattern, are evaluated with regards to prediction accuracy using the root mean squared error metric. In addition the models are compared against the persistence model, which is considered as a benchmark in the field of wind power prediction.

In applying this approach we show that NWP data are highly useful in short-term power predictions even for one hour ahead predictions. Other notable findings include the positive impact data from neighboring turbines has on prediction accuracy when predicting for a single turbine, however when this strategy is used for predicting all turbines, it is not nearly as effective.

1.3 Structure of the thesis

In this section a short presentation of how this thesis is structured is given.

Part II: Foundations

Chapter 2: Machine Leaning

In this chapter, machine learning along with some of its key concepts are introduced. Challenges concerning the choice of features and model development are also touched upon. In addition SVM is explained along with the importance of optimizing support SVM models by selecting suitable parameters. We also introduce K-fold cross validation and mutual information regression and how the two methods are used in this work. The former is used for tuning the parameters of the SVM, and the latter is used for ranking the candidate features. Finally, the root mean squared error metric and the persistence model is introduced, which are used for evaluating the different SVM models

Chapter 3: Wind Power Prediction

This chapter is a result of an extensive survey of the field of wind power prediction. The concept of forecast horizon is introduced, and a brief overview of existing approaches to wind power prediction is given. Further, a more in-depth analysis is given on the use of machine learning methods for short-term wind power prediction. Numerous models proposed in literature are reviewed, both based on the method they utilize and, more specifically related to this work, what features they utilize and at what scale predictions are done (turbine level or park level).

Part III: Short-Term Wind Power Prediction

Chapter 4: Data Sets and Experimental Environment

This work relies on data from two wind parks located in Scandinavia and forecast weather data from a publicly available NWP model. These data sets are introduced in this chapter. The pre-processing steps done to the data sets prior to experiments are also mentioned. Lastly, the experimental environment is explained by laying out the prediction pipeline.

Chapter 5: Effects of Measured Wind Park Data for Turbine Prediction

This chapter addresses RQ1, RQ2 and RQ3 from section 1.2.1, by doing experiments for two wind turbines located in separate wind parks. The first sections of the chapter examine the impact of the data recorded at turbine level (except for wind direction which is recorded at park level) has on prediction accuracy. The following sections look at how including lag values to the input effects prediction performance, followed by using data from neighboring turbines as part of the input. Finally, we combine the best input configurations from the use of lag values and neighbor data to see how this effects prediction performance

Chapter 6: Wind Park Prediction Using Measured Wind Park Data

In this chapter the two strategies examined in this work for doing power prediction of wind parks are presented. Further, and concerning RQ5, the two approaches are compared with regard to prediction accuracy. For both strategies, several input patterns are examined based on findings in earlier sections.

Chapter 7: Effects of External Weather Forecast Data

In this chapter we set out to answer RQ4 by introducing weather forecast data from a NWP model into the input patterns. Experiments are done for the same two turbines as in previous chapters. Altogether six different candidate features are considered, and the input configurations tested are partly based on an initial ranking of the candidates. We also revisit the two strategies for wind park prediction, in order to see if inclusion of NWP data has an impact on their performance.

Part IV: Summary

Chapter 8: Summary and Future Work

In this chapter the most important findings from this research are revisited. In addition, suggestions for future work are layed out.

Part II

Foundations

Chapter 2

Machine Learning

In this chapter we look more closely at the concept of machine learning and related concepts. In the first section the basic idea behind machine learning will be presented through a well known example. The following sections briefly introduce the concepts of supervised and unsupervised learning and the difference between classification and regression in order to establish which part of the machine learning universe this work resides in. Also touched upon is the role of data and how the choice of features can affect the effectiveness of a model, finally the concept of over- and underfitting is introduced and how it relates to the bias-variance-trade-off.

The following sections introduces methods utilized in this work, starting by first giving an introduction on support vector regression (SVR). Subsequently a short description on K-fold cross validation, and how it is used for optimizing the SVR models, is given. In relation to the feature selection method used in this work, mutual information regression is also briefly introduced. Finally, in relation to the evaluation of the SVR models, the root mean square error metric and the persistence model is introduced.

2.1 The Basics of Machine Learning

Machine learning stems from the field of pattern recognition and learning theory. The term was first coined by Arthur Samuel in 1959 in a paper in which he developed one of the first successful self-learning programs to solve the game of checkers [59]. As the names gives off, machine learning is at its core the ability to "learn" tasks by detecting patterns in often huge amounts of data not easily perceived by humans.

2.1.1 A Well Known Example

To explain the machine learning approach it is helpful to use an example. A well known problem in which machine learning has been used involves recognition of handwritten postal codes [74, Ch. 1]. Handwriting is indeed personal, and how one person writes a specific digit varies to another. The problem can be trivial to do manually, but not very time efficient. We therefore want to develop a model which solve this problem fast and

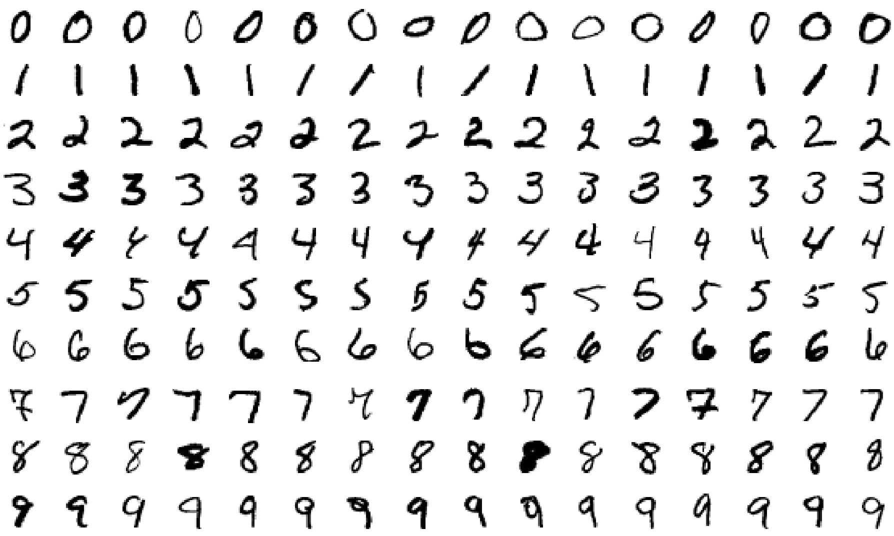


Figure 2.1: A few samples from the MNIST test dataset [67] .

automatically. To put it simply, we want a program that is fed a picture of a digit and then outputs what digit this picture illustrates in a digital format.

To build a prediction model in a machine learning setting we first need to collect appropriate data. In our example the MNIST data sets, which contains 70 000 handwritten digit samples, is suitable. Some samples of the data set can be seen in Figure 2.1. Next, the data are divided into two sets, a *train set* X_{train} and a *test set* X_{test}. Each sample is represented a 28x28 matrix of pixels, where each pixel contains a gray-scale value ranging from 0 to 255. Each sample can therefore be viewed as a vector consisting of $28x28 = 784$ *features*. Each sample is provided a corresponding label value, which is the digital solution to the handwritten sample. The label values are often referred to as *targets*. Let the corresponding targets be denoted by Y_{train} and Y_{test}. Next we choose a *learner* denoted by function A. The learner can be any sort of machine learning algorithm. A is given the training set S where $S = (x_{train1}, y_{train1}), ..., (x_{trainN}, y_{trainN})$ and returns a *prediction rule* $h : X \rightarrow Y$ [60, Ch. 2, Se. 1]. This function is also called a *predictor* or, depending on the problem, more precisely a *classifier* or a *regressor*. In plain language what happens in this stage, called the *training phase* is that our learner algorithm gets tweaked to best capture the correlating structures of the data set to give a good approximation of new data. The output from the *training phase* is called a *predictor*. We can now present the *predictor* with unseen data to in order to classify the digit. Finally, we score the developed *predictor* by comparing the Y_{test} to our predicted values Y_{pred} given by $h : X_{test} \rightarrow Y_{pred}$ to see how well our developed prediction model h generalize to the unseen data in X_{test}.

2.1.2 Supervised and Unsupervised Learning

At its most fundamental level machine learning is usually categorized into two types: Supervised and unsupervised learning. In the example described in section 2.1.1 the target values are known, which means that the machine learning algorithm can adjust the weights of the prediction model by looking at the target, or solution, for each specific train sample. This is called supervised learning and can be described as learning by example [74, Ch. 2, Se. 6.2]. We can picture us a teacher involved in the training process responding back to the algorithm telling it if it got to the correct answer, or in other cases how close it got to it. Based on the response from the teacher for each train sample, the algorithm "fits" the model in order to accurately predict the correct answer for future observations.[26, Ch. 2, Se. 1.4].

In contrast, in unsupervised learning the are no associated responses to the observations and can therefore be described as "learning without a teacher" [74, Ch. 14, Se. 1]. This makes the task of predicting more challenging since we do not have a clear way of measuring its success with the help of a target value. Unsupervised learning is mainly used for clustering and dimensionality reduction. In clustering the goal is to sort observations into distinct groups from which further information can be drawn. An example can be to group customers based on spending habits in a market study. If information about each customers spending is not available, a supervised analysis is not possible. What can be done however is to identity distinct groups on other available data, such as zip code and income, to identify interest groups [26, Ch. 2, Se. 1.4].

In this thesis we only consider supervised learning. Algorithms and challenges specifically related to unsupervised learning is therefore not further examined.

2.1.3 Regression and Classification Problems

A response variable can either be qualitative or quantitative in nature. A qualitative variables are discrete and can take form as yes or no answers, a persons gender or a medical diagnosis as examples. In the handwritten digit problem the response values are qualitative, since its responses are limited to discrete digits (0-9). Problems working with qualitative responses are referred to as classification problems [26, Ch. 2, Se. 1.5]. On the other hand, when our responses takes form of a continuous numerical values (quantitative) we are dealing with a regression problem. The type of problem being addressed has consequence on how we evaluate our model and also in selection of suitable machine learning methods.

In this work we want to predict the wind power output of wind farms, meaning our task is a regression problem.

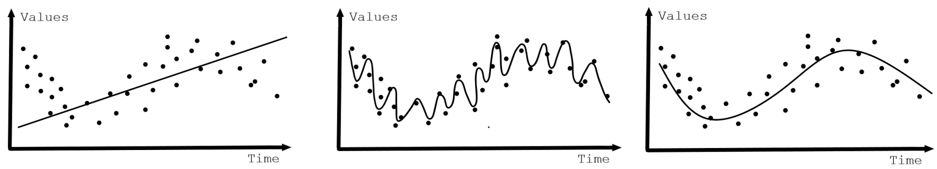
2.1.4 Features and Feature Selection

As mentioned, machine learning is purely data driven, which gives much emphasis on the amount and quality of data used as basis for training our predictive model. In machine learning the specific inputs to a model is often referred to as a feature, which is a numeric representation of raw data. The number of features is important in order to successfully adopt machine learning. If there are not enough informative features, then the model will not be able to perform well. If there are too many features, or if many features included are not relevant to the ultimate task, then the model will be more expensive and tricky to train [87, Ch. 2]. Putting some thought into which features to include when building a predictive model is therefore important. In one article the following reasons for doing feature selection were highlighted [5]:

- Reduces Overfitting: Less redundant data means less opportunity to make decisions based on noise
- Improves Accuracy: Less misleading data means modeling accuracy improves
- Reduces Training Time: fewer data points reduce algorithm complexity and algorithms train faster

But being able to locate which features has a negative or positive effect on accuracy is not necessarily easy. A variable that is presumably redundant can contribute to noise reduction and consequently better performance of our model. Further, a variable that is completely useless by itself can provide a significant improvement to performance when included with other variables. In addition, two useless variables can become useful when they are used together [30]. There exits many methods for feature selection, and in this work we have used a filter method named mutual information for doing a ranking of the available variables as a pre-processing step. This method will be covered in section 2.4.

2.1.5 Over- and underfitting



(a) Underfitted (b) Overfitted (c) Good fit

Figure 2.2: Simplified illustrations showing how over- and underfitting affect a learning curve. Adopted from [6].

When evaluation the performance of a model on the new unseen data we can sometimes encounter poor accuracy. A possible explanation to this is that our developed model is either under- or overfitting. This can be related to the data used in the training phase. Data sets often contains noisy data and outliers, and as a consequence of excessive training our model might pick up on this noise and give it explanatory value. This affects the models ability to generalize, leading the model to fit too much to the training data. Another possible source of overfitting is the inclusion of features which does not provide any predictive value. This is not easily avoidable, since which data might have an underlying relationship to the target is not easy to pinpoint. Machine learning models can also suffer from the underfitting if the model is too simple to capture the complexities of the data, but the former is by far the most common reason. Figure 2.2 shows some simplified illustration on how over- and underfitting can affect the learning curve.

The Bias-Variance Trade Off

A concept closely related to the over- and underfitting problem is the bias-variance-trade-off. Apart from irreducible error of a model, meaning error which can not be improved, we in machine learning terms often talk about bias error and variance error of a model. Bias refers to the error that occurs when the underlying machine learning method which tries to predict a complicated real-life problem is too simple [26, Ch. 2, Se. 2.2]. A simple linear regression method will undoubtedly not be able to successfully predict a complicated problem because of its lack of flexibility, leading the model to underfit. The solution in this case is to rather utilize a more flexible, model which is more suited to capture complex dependencies in the data. We can see from Figure 2.3 as model complexity rises, bias is reduced and the model error decreases, but after a point the error increases. At this point our model has become too complex, or in terms of this discussion the variance has become too high. By variance we mean the amount the prediction estimates will change if different training data was used in training the model [26, Ch. 2, Se. 2.2]. A model with high variance will put too much emphasis on the training data, leading to a model that does not generalize well to unseen data. Ideally the estimates should not vary too much between training sets. The goal for any successful deployment of a supervised machine learning is to balance the amount of bias and variance. This is referred to as a trade-off. It is easy to develop a model with low bias and high variance (for instance, by drawing a curve that passes through every single training observation) or a model with very low variance but high bias (by fitting a horizontal line to the data). The challenge lies in developing a model which balances good. Many machine learning algorithms provides means of balancing the bias and variance by configuring hyperparameters of the algorithm. This will be further elaborated on in the next section.

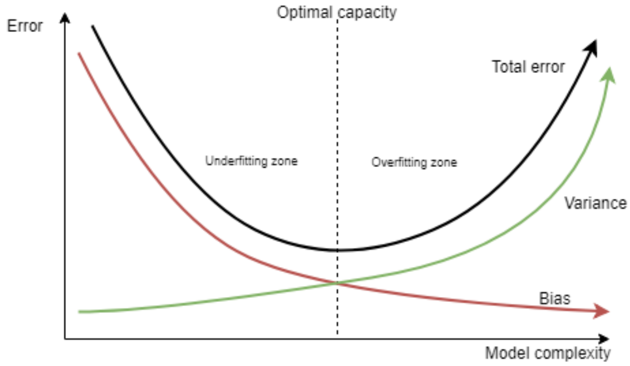


Figure 2.3: Bias and variance error in relation to model complexity. Adopted from [24]..

2.2 Support Vector Regression

In this thesis support vector machines (SVM) are used for prediction. More specifically, a regression variant of SVM are utilized, which has been shown to be successful in several fields of research, including time series forecasting [68].

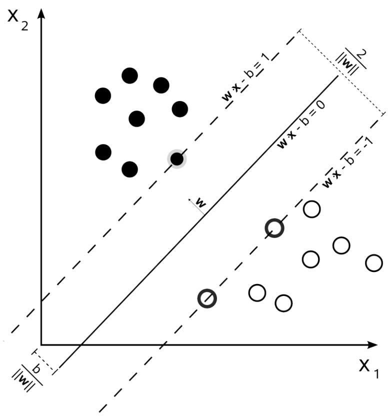


Figure 2.4: Graphic showing the maximum separating hyperplane and the margin [18] .

SVM are machine learning models using supervised learning. The SVM was first formulated back in the nineteen-sixties by Vapnik and Chervonenkis and was restricted to only doing linear classification [75]. It has since been extended to both handle non-linear modelling and regression. The basic idea of SVM is to represent the data as point in space and then find a function that correctly separates the data. The development of this function, or hyperplane, responds to the learning phase. The function is then used to predict new unseen data by mapped it onto the same space and classify the data points according to which category, or on which side of the hyperplane, it falls in. Figure 2.4 shows the hyperplane with the margin illustrated by the stippled lines. The same principle is adopted when using SVM for regression (SVR). Several

SVR algorithms have been proposed, and in this work the ϵ-sensitive SVR algorithm proposed by Cortes and Vapnik is utilized [17]. The ϵ parameter defines an area around the hyperplane in which samples labeled as correct even though they deviate from the actual target, allowing the model too loosen up demands on accuracy.

In SVR modelling the goal is to find weights w by solving the problem formulated in 2.2. Basically we are searching for two things: a hyperplane with the largest minimum margin which at the same time correctly describes the target data. These are two competing objectives, since when our margin gets larger, the chance of misrepresenting the data increases. This is where the C parameter in formula 2.2 comes into play. The C parameter controls the penalty imposed on observations that lie outside the ϵ margin. In simpler terms it controls how much we desire a correct representation. Figure 2.5 illustrates the impact of on the decision boundary with different values of C . Decreasing the C parameter allows for a large margin, but at the same time amount for a less accurate description of the data. Increasing our C parameter on the other hand gives a much better representation, but can have the consequence fitting too well to the training data and making the model worse at generalization. Preferably, we would like to generate models that represent the training data accurately, but not so much that it causes the model to overfit.

$$\text{minimize } \frac{1}{2}||w||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{2.1}$$

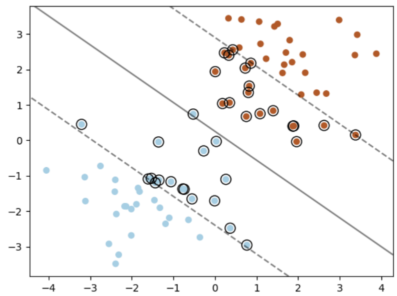
$$\text{subject to} \begin{cases} y_i - \langle w, x_i \rangle - b & \leq \epsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i & \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq 0 \end{cases}$$

2.2.1 Radial Basis Function Kernel

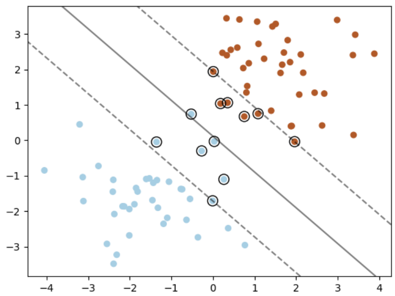
As mentioned, SVR has the ability to do non-linear modeling which is preferable when dealing with complex data sets. This is achieved by applying the *kernel trick*, which maps the data to a higher dimension space by the use of a kernel function. In this higher dimension a linear regression can be performed [74, Ch. 12, Se. 3] [68]. I will not go in depth into how this works mathematically as it is outside of scope of this work, however how the choice of kernel function can have an effect the model is worth pointing out. It exists several different choices for kernel function, where the polynomial and the radial basis function (RBF) are among the most popular.

$$K(x, x') = \exp\left(-\gamma||x - x'||^2\right) \tag{2.2}$$

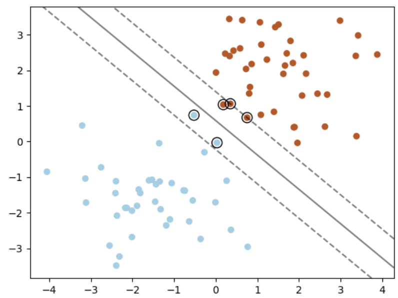
In this work the RBF kernel function, shown in 2.2, is used. This kernel introduces the gamma parameter γ. γ needs to be considered with care, as



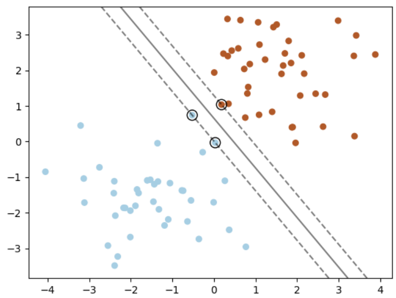
$C = 0.01$



$C = 0.1$



$C = 1.0$



$C = 10.0$

Figure 2.5: Effect of C on margin and decision boundary

it defines the area of influence of a single training example. A high γ value shrinks the distance of influence of a single training sample, leading to greater definition power of that particular sample over the fitting function in its close region, but at the same time limiting its influence over other training samples nearby [26, Ch. 9, Se. 3.2]. As a consequence the fitting function experiences high variances since few samples contribute to the curvature. Too low γ values will result in the smoother decision function by expanding the region of training samples who contribute to the curvature of our fitting function

2.2.2 SVR and the Bias-Variance-Trade-Off

Going back to the challenge of balancing the bias-variance described in section 2.1.5, we mention that many methods allow parameters to be set by the user to find a balance in which low bias and low variance can be achieved. In SVR, the most influential hyperparameter in this regard is the C parameter, which also is called the regularization parameter. In addition, we also mentioned how γ affects the curvature in the decision function. If γ is too high we can risk that no values C will prevent overfitting. We therefore have to consider both C and γ when optimizing the SVR model. In figure 2.6 we can see how different settings of two parameters effects the decision boundary. Since these hyperparameters controls how well

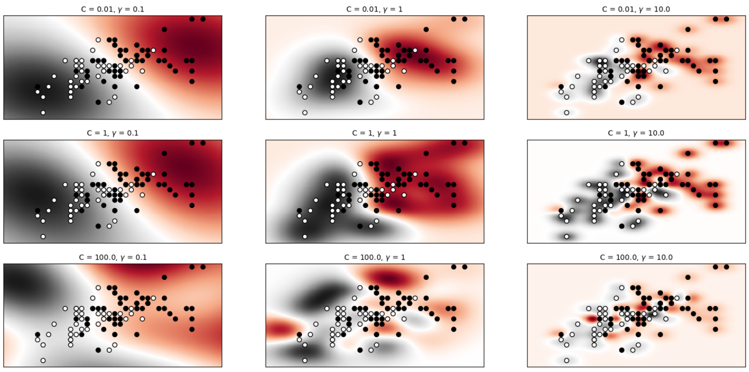


Figure 2.6: Example: Effect of C and γ on decision boundary

SVR generalize, finding suitable values is of utter importance. There exists many approaches in literature for finding suitable values of C and γ, and the strategy deployed in this thesis is discussed in the following section.

2.3 K-Fold Cross Validation

Cross-validation (CV) is probably the most used method for estimating prediction error of statistical learning methods [74, Ch. 7, Se. 10]. It is a technique for assessing how the results of a statistical analysis will generalize to an independent data set, and is used for both finding the suitable ML-methods for a problem, and for tuning the hyperparameters of a selected machine learning method. In the machine learning process it is common practice to leave the designated test set out of this process, since we want to treat this data as unseen samples and not have any influence on the choices in developing our model. In CV we therefore introduce a third set. The idea is simple: Put aside a part of the training samples in a hold-out set. The model is then trained on the remaining training set and validated by estimating its performance against the hold-out set.

Simple CV has two potential drawbacks. First, the error estimation can become highly variable depending on which part, and how much of the training samples are included in the hold-out set. Secondly, if our test data are limited in size, allocating parts of it to a independent hold out set might make our model to preform worse [26, Ch. 5, Se. 1]. To address these drawbacks the more refined k-fold CV is often preferred. In k-fold CV the training data is randomly divided into k subset of equal size. Then, in turn, each of the k subsets are treated as hold-out set and the rest rest as training set, leading to k error estimations with a unique training and hold out set. The idea is illustrated in figure 2.7. The final K-fold CV estimate is obtained by averaging all the k error estimations. The value of k can potentially have a major impact on the resulting estimates, and research has empirically shown that using k of size 5 or 10 shows test error rate estimates that suffer neither from excessively high bias nor from very high variance [26, Ch. 5, Se. 1].

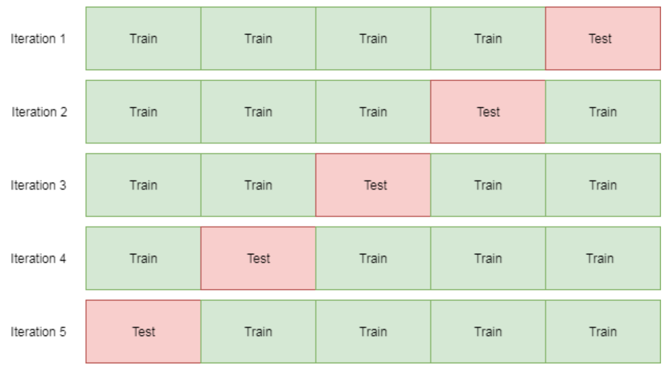


Figure 2.7: Illustration showing how the hold-out set (test set) is allocated for k = 5

For the purpose of optimizing a SVR model we can then define a range of possible hyperparameter configurations. By using k-fold cross validation we then validate the models using different C and γ settings in order to find the most suitable pair.

2.4 Mutual Information

Since the main focus of this thesis is the effect of different types of features, how I go about testing this must be addressed. The most solid strategy would be to just try out all the different combinations of the candidate features, but when the numbers of possible features becomes large, this strategy gets very time consuming. Instead a prior ranking of all the feature candidates is done. Then, based on their ranking, each feature is added on to the feature space to see how they affect the model. In work the ranking is done by finding the mutual information between a given candidate feature and the target.

Mutual information (MI) is a measure of the amount of information that one random variable contains about another random variable [69, Ch. 2, Se. 3]. The method has the advantage of being sensitive to dependencies which are do not manifest themselves in co-variance [37]. In other words it measures any relationship between variables, and not only linear relations. The MI between two variables is zero if and only if they are statistically independent of each other. A high MI score between a feature and the target would then mean that the feature is relevant. This method has been used in literature for selecting features for wind power prediction [37]. When using MI for feature selection directly, we typically set a predefined limit of of how many features we wish to use and then extract the highest ranked. However, in this work we only use MI as a guiding tool for how to expand the feature space in order to avoid testing all possible combinations. A potential drawback with this ranking procedure is that features that are relevant together but useless individually is not accurately spotted, so we have to avoid putting too much trust into their ranking.

2.5 Root Mean Squared Error

In order to evaluate the performance of a machine learning model, we need a way to measure how well its predictions actually match the observed data (test data). For regression tasks, the most common measure used is the mean squared error (MSE) [26, Ch. 2, Se. 2.1]. In this work a closely related measure is used, namely root mean squared error (RMSE). RMSE is easier to interpret since calculating the root brings back the measure to the actual unit. We can see from the equation 2.3 that the difference between the actual values and the predicted values are squared. This is in order to penalize larger differences more than smaller ones. When evaluation the performance of a model we seek a small RMSE, meaning the predicted value are very close to the true observations.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_{pred} - y_{true})^2}{N}} \tag{2.3}$$

2.6 The Persistence Model

In addition to comparing the RMSE of the developed SVR models, in this thesis we also evaluate our models in comparison to the persistence model. In wind power prediction the persistence method, also referred to as the naive approach, is often used as a benchmark. Although the model is extremely simple in nature it can be difficult to beat, especially on the short-term [86]. The model simply works by setting the future value to be the same as the current value, assuming that the future will hold the same as the present, hence persistence.

Chapter 3

Wind Power Prediction

The increased focus on wind energy in recent years has resulted in vast research on how to accurately predict energy output. In this chapter wind power prediction is introduced. Wind power prediction models can be classified by several factors, and in order to establish where this work resides in the landscape of wind power prediction, several of these factors will be briefly introduced. Later sections will dwell more into proposed models from literature in our specific scope of research, considering the methods utilized, the input they consider and on what scale they predict.

3.1 Forecast Horizons

The forecast horizon denotes how far into the future we wish to predict. It can typically vary from ten minutes up to several days depending on the case that it is used for. If the purpose of the prediction is to balance the grid of a power system we might need an estimation in a short time frame, while if the purpose of prediction is to better plan future maintenance it would make more sense to produce predictions reaching far longer into the future. Forecast horizons are usually classified into very-short-, short-, medium- and long-term predictions. Which prediction time-frame we are considering might influence the choice of methods and input. There exists no exact definition which also includes typical use-cases within each time-frame category. In this work we predict one to four hours ahead, which falls into the short-term category.

3.2 Prediction Approach

Roughly speaking there exists two main approaches to doing wind power prediction, namely the physical approach and the statistical approach. They differ in the sense of which input they take in and what kind of computation they do.

Table 3.1: Classification of Forecast Horizons

Class	Range	Applications
Very short-term	Few seconds to 30 minutes ahead	Electricity Market Clearing, Regulation Actions
Short-term	30 minutes to 6 hours ahead	Economic Load Dispatch Planning, Load Increment/Decrement Decisions
Medium-term	6 hours to 1 day ahead	Generator Online/Offline Decisions, Operational Security in Day-Ahead Electricity Market
Long-term	1 day to 1 week or more ahead	Unit Commitment Decisions, Reserve Requirement Decisions, Maintenance Scheduling to Obtain Optimal Operating Cost

Adopted from [58]

3.2.1 The Physical Approach

Physical approaches to wind power prediction is closely related to numerical weather prediction (NWP) models. NWP models are dependent on rich atmospheric data typically gathered from satellites, weather balloons and aircrafts. The gathered data is then used to model the initial atmospheric state on a three-dimensional grid from which the NWP model solves complex mathematical equations, based on physical laws, to produce the final forecast. As a consequence large NWP models are often drifted by governments and run only a few times a day. How accurate the forecast produced by NWP models is depends highly on the horizontal grid resolution it uses. Having a low grid resolution means that it will not be able to express local terrain variations, which might play a significant role for a single point forecast. As an example, the global intergovernmental NWP model ECMWF Integrated Forecasting System provides forecast with a 9 km horizontal resolution [52].
Physical models for wind power prediction are themselves NWP models, but they do forecasting on a much smaller scale, which covers only the area of interest. Typically this is done by using the outputs from global or regional NWP models to do a numerical down-scaling, which means producing a forecast on a higher grid resolution. This requires a detailed mapping of the geographical terrain of the area that has to be put into the model. The end result is a more reliable weather forecast. To finally predict the wind power, forecast wind speed is used as input to power curve functions. A model that uses this approach is the Prediktor[28] which utilizes the wind speed and direction forecasts from the NWP model HIRLAM(High Resolution Limited Area Model) and uses the WAsP model

to convert the wind to the local conditions. The converted wind values are finally converted to power output using the power curve . Another example is the Preventio[57] model which produces several wind power predictions in time intervals from 15 minutes up to 10 hours by integrating weather forecasts from several NWP models to the conditions in the wind farm's local environment. The final estimation is then optimized by combining the separate predictions.

3.2.2 The Statistical Approach

The basic idea of statistical models is to find a mapping between explanatory variables and the actual wind power [38]. This approach is typically less expensive, both considering computational costs and the building of the model. Mapping the local environment, which can be a complicated task, is not necessary in the statistical approach. However, statistical models depend on historical data, which is not the case for physical models [45]. Physical methods typically offer advantages in long-term prediction while statistical methods do well in short-term prediction [41]. Statistical models are often sub-classified into traditional methods based on time-series forecasting and learning methods based on machine learning [45] [38]. In addition, it is common to classify methods in to two additional classes, namely hybrid and ensemble methods. Hybrid methods try to combine the strengths of several different methods in order to improve prediction accuracy. The basic idea behind ensemble methods is to train multiple weak prediction units and then calculate the final wind power estimate based on all of the individual predictions.

Traditional Methods

The objective in traditional methods and machine learning methods are similar. They both aim to improve prediction accuracy by minimizing some loss function. Their difference lies in how the minimization is done. Typically traditional methods use linear processes while ML methods use non-linear algorithms [44]. Further, traditional statistical modeling was designed for a small number of input variables which demands more assumptions about the data and its distributions [8].

Several traditional models have been used in wind power prediction with success. Among the more popular choices is autoregressive moving average (ARMA), which has been successfully used to predict short-term wind power for wind farms in the US [50] and a single wind turbine [29]. The method has also been used to predict wind power in short time-frames [10]. The AR part involves regressing based on the recent values, while the MA part helps to filter out noise from short fluctuations by considering values further back in time. Another popular approach is autoregressive integrated moving average(ARIMA), in which data points are replaced with the difference between the actual value and the previous. [3][9][1][88][70]. Another similar method used for wind power prediction is the generalized autoregressive conditional heteroskedasticity (GARCH) model [88][89]. The GARCH model differs from ARMA and ARIMA in that it considers the variance error, and not the mean error, making use of the assumption that there are periods in the time series of stability and periods with high volatility. One paper combined ARIMA and GARCH for predicting power for a wind park, showing improved accuracy [70] over simply using ARIMA and GARCH. The mentioned models are all aimed at either very-short or short-term prediction, which is the case for most traditional models.

Machine Learning Methods

In contrast to the traditional methods, machine learning(ML) methods make use of general-purpose learning algorithms to find patterns in often rich and complex data, which makes them less sensitive to the quality of data [8]. ML-methods are widely used for wind power prediction, and can be utilized for predictions in all time frames. Popular methods for wind power prediction include artificial neural network (ANN), support vector machines (SVM), k-nearest neighbors (kNN) and gaussian process (GP). A wider account on the use of ML-methods will be given in section 3.3.

Hybrid and Ensemble Methods

Both hybrid and ensemble methods typically show good performance in research and are normally used for short-term and medium-term wind power predictions. As mentioned, hybrid methods combine several different methods to predict. The methods typically consist of a traditional statistical prediction model for the linear element of the time series, and a nonlinear prediction model for the nonlinear element.[14] Such an approach was used in [14] by combining ARMA and a radial basis function neural network (RBFNN) showing superiority over using ARMA and RBFNN individually. In [13] a hybrid model combining wavelet transform (WT), particle swarm optimization (PSO) and adaptive neuro-fuzzy inference system(ANFIS) was shown to outperform several other simpler approaches. WT is a filter technique used in several proposed models for transforming the data before prediction. PSO is a machine learning technique which tries to mimic the social behaviour of birds in flocks to do classification, and was used in this case to set the parameters of the ANFIS unit. In another proposed model, four different methods were hybridized, namely WT, fuzzy artmap network (FA), firefly (FF) and SVM [32].

Ensemble methods build on the assumption of strength in numbers by training several weak models for prediction and then using a strategy for getting the final prediction based on the individual models. One such method is boosting, in which the accuracy of an individual model effects the next model, in other words learning is done in sequence [31]. Another popular ensemble method is bootstrap aggregation (bagging). In this approach, models are trained independently and predictions from each independent model are aggregated in order to lower variance [31]. In [78] a bagging method consisting of ten individual SVR models was used. Other work has considered ensembles consisting of six individual NNs using the boosting strategy [31]. Another proposed model did not use boosting or bagging, but rather used the single predictions from six different NNs as input to a final GP unit [4].

3.3 Machine Learning Based Models for Short-Term Wind Power Prediction

We have seen that wind power prediction can be categorized by which forecast horizons they consider and the type of approach. This thesis focuses on the short-term wind power prediction and the use of machine learning methods, particularly support vector machines. This section goes into more detail on which methods are used in this category. Their use of features are also examined, looking at both historic measured attributes, lag features, attributes from neighboring turbines and external atmospheric data, as this thesis addresses the usefulness on all of these feature types. Finally, the thesis reviews on what scale they forecast, in other words if they predict turbines individually or for the whole wind park.

There exists a large number of machine learning based models proposed for short-term wind power prediction, and the majority of them build upon well established machine learning methods. The most common approach is using some kind of derivative from the concept of artificial neural networks (ANN). An ANN is a machine learning method which is based on the nervous system by having nodes that "fire off" information through the network as they get activated, or in other words if the input value to a node exceeds over a set threshold [60, Ch. 20].

The most "traditional" neural network is called the multilayer perceptron (MLP) which is a three-layered feed-forward ANN (FNN) which utilizes a sigmoid function in the activation step. Such a model was successfully developed to predict short term wind power output for a wind farm in Portugal by showing improvement over the persistence method [12]. The same approach has also been used to predict for wind farms in India [63]. Other works use the method of a radial basis function neural network [39][33]. RBFNN refers to a three layer FNN which uses a RBF as an activation function. The authors in [39] argued for the use of RBFNN on the fact that RBFF has shown advantages in learning speed and local minimum problem compared to the use of a sigmoid function. Other work has proposed using a wavelet function for the activation function for short term predictions [16][77]. In [16] such a neural network was used to predict up to 6 hours into the future. In this work they also compared the WNN against RBFNN and MLP, showing improvement over both methods.

Another popular class of neural networks are the so-called Recurrent neural networks. These differ from the three-layered feed-forward methods already mentioned in that their inner architecture contains loops. One such RNN is the Elman Network. In an Elman Network there is a feedback from the first-layer output to the first-layer input. This approach has been used in several works [35][83]. In [83] an Elman-based approach showed better performance compared to an RBFNN and WNN. Another recurrent architecture used in short term prediction is a Nonlinear Autoregressive Exogenous Neural Network (NARX NN), which uses feedback connections from several layers of the network. NARX has successfully been deployed to predict power output of wind farms located in the US [76]. A NARX NN was also used in [40] for doing short term predictions.

In recent years a deep learning approach based on RNNs, the Long-Term-Short-Memory network (LSTM) has become popular for wind power prediction [62][79][80][25]. LSTM replaces the nodes in the neural network with LSTM units, which have the ability to learn both long and short-term dependencies of the input data [79]. Several studies have shown its effectiveness for short-term wind power prediction prediction. In [79] the proposed LSTM model significantly improved accuracy when compared to ARIMA, a simple RNN and a NN. Similar improvements were shown in [79] against a SVM model and NN, and in [25] compared to a SVM and a ARIMA model.

Other approaches based on neural nets include the Adaptive neuro fuzzy inference system (ANFIS) architecture. ANFIS combines the concepts NN and fuzzy logics. Fuzzy logics consists of if-then rules based on human knowledge and membership functions. Combined with NN, the fuzzy rules can be learned automatically. Its use is seen in several works [55][56][20][41]. In [41] ANFIS was used to predict power output for a single turbine. The model was compared against a Feed forward neural network showing better accuracy for the ANFIS approach. In [20] ANFIS was used in a two step manner, first predicting wind speed based on NWP quantities, and then using this predicted wind speed into the final ANFIS for doing the wind power prediction. The proposed model showed effectiveness in predicting a single wind turbine in China.

The other main machine learning approach used for short term wind power prediction is support vector machines, which were introduced in section 2.2. Numerous studies have proposed models based on this method [84][46][61][48][19][42][34]. In [34] an SVM approach was compared against several other often used approaches, including a RNN, and FFNN, showing that the SVM approach yields better accuracy. In [19] the proposed SVM model showed better results when compared to an Elman network and the same was shown in [48] where their model gave better results than a BPNN (back propogation network). As discussed in section (LINK), an important step in developing a SVM model is the choice of kernel function. A clear majority of the SVR models utilize the Radial basis function as kernel. In [48] they justify this choice because of its effectiveness in power system load prediction. In [34] both RBF and a polynominal kernel fuction are used as they tend to perform better in real-world applications. Using a wavelet kernel has also been suggested in [42], but in this case it did not show improvement over the use of RBF.

Less popular machine learning approaches for short-term wind power prediction includes k-nearest neighbours (kNN). The kNN algorithm predicts by measuring the similarity of new unseen data to all available data. Based on the degree of similarity to other data points, a value is given. A kNN based model was used in [71] and [22]. Mainly, however, learning based short-term wind power prediction models are based on either the concepts of neural networks or support vector machines. Either as stand alone prediction units, as a hybridization with other methods, or as part of an ensemble approach.

3.3.1 Input Patterns for Short-Term WPP

As discussed above, a major contributing part to the effectiveness of a machine learning algorithm is its choice of input. By reviewing a vast number of proposed models we wanted to discover which attributes are commonly used for the task of short-term wind power prediction. In contrast to conventional time-series models, such as ARIMA and its extensions, most machine learning methods allow several features as input to the model.

For short term wind power prediction, the most common feature is historical measured wind power. Several models based on neural networks[16][2][56], SVM[85] [90] and kNN[71] utilized only this attribute as input. Another commonly used attribute is wind speed, and throughout literature both these quantities, both alone and in conjunction are widely used. In [91] and [81] they only predict the wind speed using past measured values, and then use a power curve to get the final wind power output. In [33] however, wind speed and wind power together were used as input. They found that using both attributes improved the accuracy of their model. Both wind speed and wind power were also used in [61], [82] and [7].

Other work has utilized additional atmospheric attributes. In [63] historic humidity measures, in addition to wind speed and wind power, was used as input. Because of low variations in temperature and air pressure, they excluded these attribute from the input. In [76] they included past wind direction and temperature measures along with wind power and wind speed. The NN model proposed in [40] was trained and tested using measured wind speed, temperature, pressure and air density. The LSTM proposed in [80] considered, in addition to historic wind power, wind speed, wind direction and temperature, a large set of operational attributes such as yaw angle. However, they found that after including ten separate attributes, the influence on prediction was constant.

Use of Lag Features

Another type of features used in several models reviewed are lag values. In other words, measured values not only at the time of prediction, but also close in history are used in predicting to the future. Both SVR models, and

NN models reviewed utilize lag features. However, this is not commonly used for RNN-based models, such as LSTM, since the information carried by previous inputs can be held inside the network when subsequent inputs are being processed [80]. One study compared using only the most recent wind power and wind speed measures to expanding the lag window up to 4, ending up with a feature vector of 8 features [39]. They found that increasing the feature vector boosted accuracy. In [82] they used lag features of both wind speed and wind power with a window width of five, ending up with ten inputs. An even wider lag window of size seven was utilized in [34], and in [61] the three last measures of wind speed and wind power, in addition to the most present, were considered as input. The majority of models, however, do not consider lag features.

Use of Neighbor Data

Not much research has been done on the impact of using data from neighboring turbines to predict a single turbine, but there exists two papers which examine this approach [72] [73]. In [73] they used power measurements of ten neighboring turbines as input in addition to the actual power measurement of the target turbine to predict the power output for a single turbine using support vector regression. The inclusion of neighbor turbines was shown to have a positive effect on accuracy.

Use of External NWP Data

All the work mentioned in this section up to this point only use past measured data as input to their prediction models. This seems to be the general approach for short-term wind power prediction. In one paper they write that learning methods based on historic data alone are used for short-term prediction of wind power within a six hour time frame [62]. Further, in a review of short-term models it was stated that prediction models using NWP forecasts outperform time series approaches after about three to six hours look-ahead time [27]. In [16] the authors deliberately did not choose to include NWP data into their model for short-term prediction, based on the assumption that this does not significantly improve accuracy. Despite this, some studies do consider NWP data for forecast horizons as short as one hour ahead. [62][25][20][19][21]. The obvious input candidate from NWP models is forecast wind speed, which all these models include, but other less explanatory attributes have also been considered. The LSTM model proposed in [79] for four hour ahead predictions included forecast pressure and temperature, in addition to forecast wind speed, as features with success. In [19] they expand the feature space even further by including forecast humidity. Other work has also included wind direction [21] as input in addition to wind speed, air pressure, temperature and humidity.

3.3.2 Difference in Scale of Predictions

The models reviewed differ regarding on what scale they do predictions. Several studies focus on single turbine prediction [39][33][35], [20][41][21][19][90]. However, in a real world application we are at most times interested in the predicted power output of a whole wind park. We therefore need a strategy to scale up to wind park level. In [23] they do this up-scaling by predicting each turbine individually and then sum them up to obtain the final wind park power output prediction. This approach was also examined in [73]. In addition, they compared the approach against just using one aggregate time series of the whole wind park. They found that when including lag values and corresponding differences between the lags, the latter strategy showed better results. Most models predict wind park power output by using aggregated data [7][12] [76][62][42][2][56][79]. How they go about calculating the aggregate and why they do not predict each turbine individually are in most cases not specified. A simple reason for this can be that the data sets used for testing were aggregated when obtained, or that measurements for individual turbines are not recorded. Another strategy for wind park prediction was deployed in [61]. In this case, they predicted wind power for a wind park consisting of 23 turbines by first finding the most correlating wind turbine to the aggregate wind power output. The most correlating turbine was then used to predict the power output for the whole wind park by using samples from turbine data, and aggregate data as targets.

3.4 Summary and Discussion

In this chapter, several approaches for wind power prediction have been introduced. Emphasis has been put on statistical methods, and particularly machine learning methods, for this task. We saw that combining several different methods into a hybrid model has shown to be effective.

The scope of this thesis, however, is machine learning methods for short-term wind power predictions, and in later sections models considering these constraints were further analysed. We saw that the vast majority of models in this scope are either based on a NN or a SVM method. It is difficult to single out one method that is better than others. The effectiveness of a model is highly dependent on the quality on the data used in training it, and the data sets used for testing the presented approaches are not the same. We saw that some of these reviewed works did compare their proposed model to other frequently used methods, but it is not always made clear how the models they are compared against are developed with regards to how and if they are optimized to be as effective as possible.

We also saw variation when analysing the input used for the reviewed models. Should only wind power measures be included, or only wind speed, or both? Should other attributes, such as wind direction, humidity and air pressure be included? Should lag values be included, or even data from neighbor turbines? We have seen that some studies compare different feature sets, but far from all. Our general perception is that most work does not reflect much around the choice of inputs. Most works only consider either wind power, wind speed or both as input. A simple explanation to this may be that additional attributes are not available.

Variation is also observed when examining the use of external NWP data as input to ML-based models. The vast majority of the work reviewed only considers measured historic data as input, and the general assumption seems to be that NWP data are not useful when predicting for the first couple of hours. However, we have seen that some studies include NWP data for forecast horizons as short as one hour with success. A reason for this "disagreement" might simply be that a reliable NWP model is not available in all cases. We briefly discussed how grid resolution affects the precision of NWP forecasts. It is not unlikely that if forecast weather data is extracted from a NWP model with a large grid resolution, the impact on prediction accuracy would be insignificant. Among the models utilizing NWP data we also saw a variation in terms of which attributes are used as input.

Finally, we compared the models reviewed on what scale they predict. Although a significant number of them considered wind power prediction for single turbines, most of the works regarded power output for the whole wind park. Different approaches to the latter scale of prediction were exposed. The majority of the works utilize aggregate data from the wind park, but we also reviewed a few papers using an approach of predicting each turbine individually and then summing them up to get the final wind power prediction for the whole wind park. Finally, in one paper data from a single turbine was used to predict power output for the whole park by using the aggregate wind power as target values. From analyzing the papers, our perception is that the choices of strategy are in general not argued for. One reason for this might be that data at wind turbine level is not available in the data sets utilized in the different works, in which case the choice of strategy is simple. Only one of the reviewed papers do a comparison of different strategies for wind park prediction.

This thesis is a results of the differences exposed particularly concerning the choice of input. Most interesting is the difference in use of NWP data. Our immediate intuition tells us that NWP would have a positive impact on prediction accuracy, even for predictions considering just a couple of hours ahead. Additionally, the use of data form neighbor turbines is interesting since this approach has not been much researched. Furthermore, since we have access to data recorded at turbine level, two of the strategies for doing predictions for a whole wind park are investigated.

Part III

Short-Term Wind Power Prediction

Chapter 4

Data Sets and Experimental Environment

In order to answer the questions raised in section 1.2.1, data has been acquired from two sources. The real measured wind park data from two wind parks has been provided by Statkraft and weather forecasts has been retrieved from the Norwegian Meteorological Institute's websites.
In the following sections a description of both these data sets will be provided.

4.1 Measured Wind Park Data

Measured data from two wind parks located in different parts of Scandinavia is utilized in this work. From the map in Figure 4.1 we can see that wind park I is located on an island close to the coast of Norway, and the second one is located inland in Sweden. The choice of these two wind parks is not coincidental. We assume that their differences in geographical location also reflects differences in weather conditions which gives us a better basis for giving a more general account on the effects of features. In addition to location, the two wind parks differ in size and capacity of turbines. Wind park 1 consists of 24 turbines with a full production capacity of 2.3 MW each, adding up to a total capacity 55 MW. The second wind park has a total capacity of 99 MW and is made up of 33 wind turbines with a capacity of 3.2 MW each.

Both data sets covers a time frame of two years and is recorded in ten minute intervals. It consists of the following variables:

- **Wind Power Production**
- **Wind Speed**
- **Temperature**
- **Yaw Drive**
- **Wind Direction**

Variables are recorded uniquely for each wind turbine with the exception of wind direction, which is only recorded at one spot in both wind parks. The variables listed are assumed self-explanatory, except for yaw drive which might need a further explanation. Yaw drive represents the position of the wind turbine head, which is typically rotated according to the wind direction in order to maximize energy production.

Figure 4.1: Location of wind parks. Case 1 refers to wind park I and case 2 refers to wind park II. The map is obtained from Google

4.1.1 Cleaning and Transformation of Measured Wind Park Data

Most data sets contains anomalies, and the data used in this thesis is of no exception. Therefore some simple preprocessing steps has been done before presenting it to the machine learning algorithm.

Cleaning

First, all instances which does not contain data are dropped. Also, obviously wrong records are removed from the data set (i.e a wind power measurement is larger than the total capacity of a turbine).

Another source of bad data is the fact that turbines can be shut off for a varying amount of time for several reasons (e.g maintenance). Steps have been done in attempting to recognize and remove these instances by comparing it with the measurements of neighboring turbines and considering the power curve of the turbine. From the scatter plot in Figure 4.2 we can see that if a record of a turbine shows zero output in power at the same time it records wind speed between 7 m/s and 18 m/s, then it is most likely shut off. From these analysis, some turbines have been considered unusable, and are removed from the data set entirely based on the frequency and length of shut down time. This concerns four turbines at wind park I and eleven turbines at wind park II. As a consequence, when wind power prediction for the whole park is performed, we are predicting in all for twenty turbines at wind park I and twenty-two at wind park II.

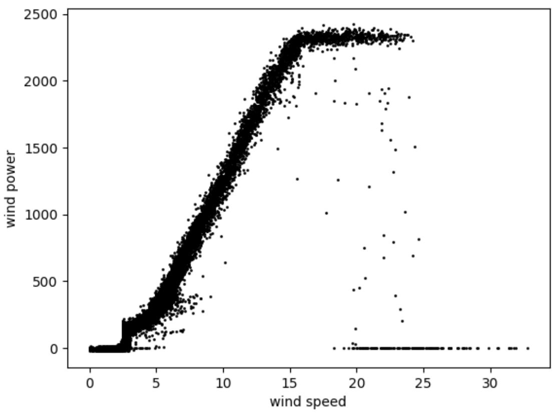
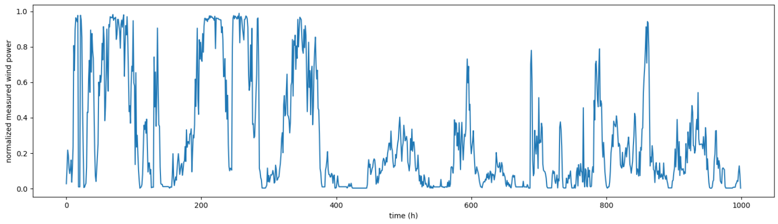


Figure 4.2: Scatter plot of wind speed and wind power

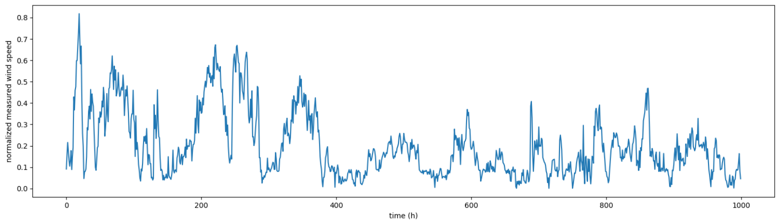
Transformation

When performing hourly predictions there is no need for all the ten minute measurements available in the acquired data sets. The data is therefore down sampled to the preferred frequency domain. There are two main approaches on how to do this. One way is to calculate an average from all the inter-hourly data-points for each hour hour. The other approach is to simply remove all the measurements recorded at times of no interest. In this work the latter option is used, which was based on some initial testing on which of the two approaches gave the most promising prediction results.

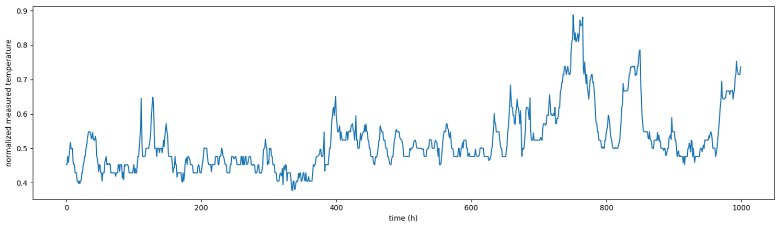
The final preprocessing step done is scaling of the input data. This is important since the different attributes are given in different units. The consequence of not doing this step is that some attributes gets emphasized more than others in the machine learning algorithm, This is especially true for the SVR-algorithm which is not scale invariant. The scaling approach used in this work is normalization, which simply transforms the data to the range between zero and one. Samples of the transformed data is shown in Figure **??**



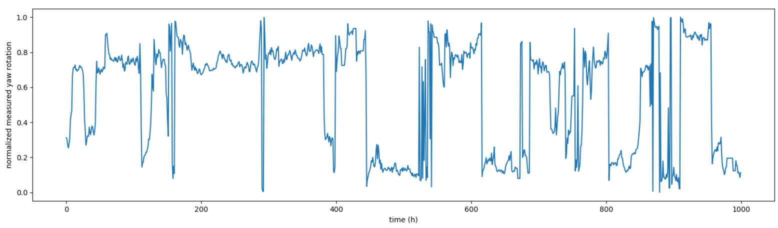
normalized wind power



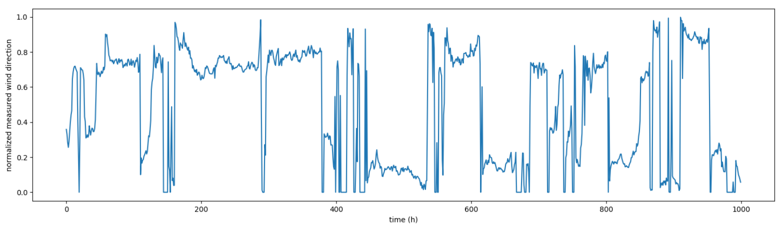
normalized wind speed



normalized temperature



normalized yaw drive



normalized wind direction

Samples of normalized measured data from wind park I

4.2 Weather Forecast Data

Weather forecast data has been gathered from https://thredds.met.no which is run by the Norwegian meteorological institutes. The forecast data utilized in this work is produced by the Arome MetCoOp (Meteorological Cooperation on Operational Numerical Weather Prediction) model, which is a bilateral cooperative effort between the Norwegian Meteorological Institute and the Swedish Meteorological and Hydrological Institute [51]. This is a mesoscale NWP model which covers both the concerning wind parks. The model is run four times a day and forecasts for every hour for up to 66 hours ahead.

As mentioned in section 3.2.1, the point accuracy of NWP models is related to their horizontal grid resolution. The smaller the grid points, a more accurate description of the geographical terrain is possible. Arome MetCoOp works on a horizontal resolution of 2.5 kilometers. Since both wind parks spans a bigger area than 2.5 km, weather data from several grid points has been collected by a simple algorithm developed for the purpose of finding the grid points closest to any given turbine. For wind park I, weather data from three grid points was gathered and for wind park II six points. The NWP model offers a large variety of attributes. In this work I have not considered all of the available data, but limited myself to the following:

- **Zonal 10 meter wind at 10 meters**
- **Meridional 10 meter wind at 10 meters**
- **Wind Speed of Gust at 10 meters**
- **Temperature at 2 meters**
- **Air Pressure at Surface**

This limitation is made on the basis of input types commonly by proposed models in literature for doing wind power prediction. Zonal wind denotes the wind blowing along the axis west-east axis. This value is positive if the wind blows from the west and vice versa [65]. Similarly meridional wind denotes the wind blowing from north to south [64].

4.2.1 Cleaning and Transformation of Weather Forecast Data

Cleaning

Similar with measured wind park data, simple steps has been done to remove outliers and anomalies from the NWP data sets. In addition, weather forecasts past six hours ahead from each run of the NWP model is removed such that there is an overlap to the next run. The forecast for the seventh hour is then acquired from a "fresher" model run, illustrated in figure 4.4. Its is worth noting that if wind power predictions with a forecast horizon past six hours in this scenario was considered it could be considered cheating, since the new model run most likely produces a more accurate weather forecast. However, since this work limits itself up to four hours it is in agreement with a real world scenario.

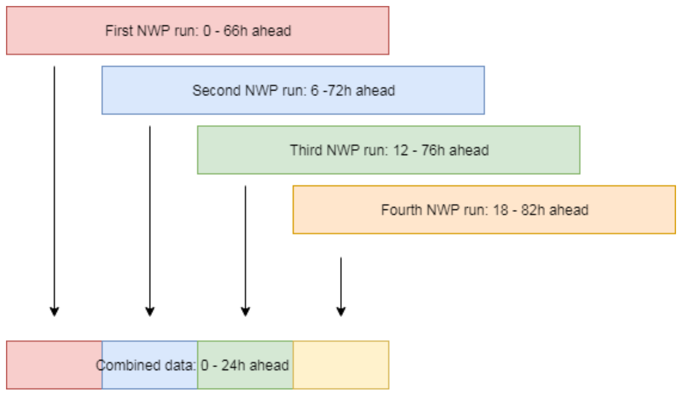


Figure 4.4: Only weather forecast up to six hours from each run of the NWP model is included in the final data set

Transformation

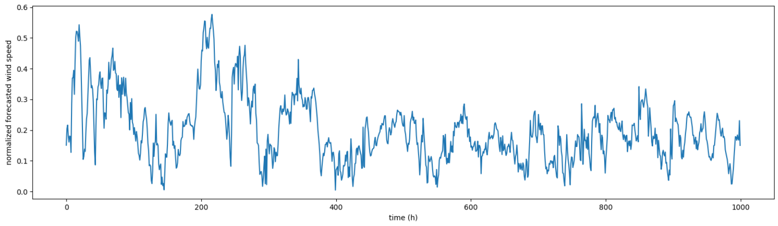
Zonal and meridional wind is not included in the final data set, rather wind speed and wind direction is calculated from these wind components using the following equations:

$$\text{wind speed} = \sqrt{u^2 + v^2} \tag{4.1}$$

$$\text{wind direction} = \text{atan2}(u, v) \tag{4.2}$$

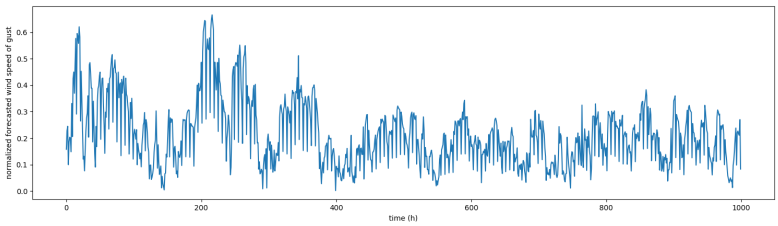
Where u is the zonal wind component and v is the meridional wind component

Finally, the data sets are normalized in range from zero to one. Samples from the data is shown in Figure 4.5.



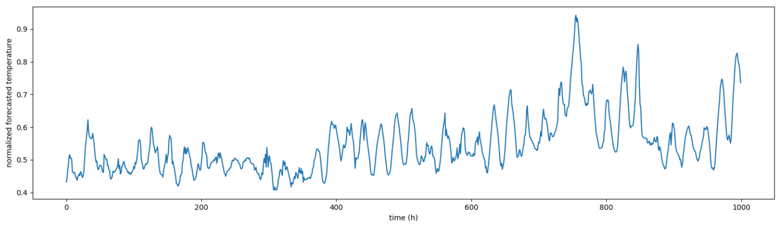
*

(a) normalized forecast wind speed



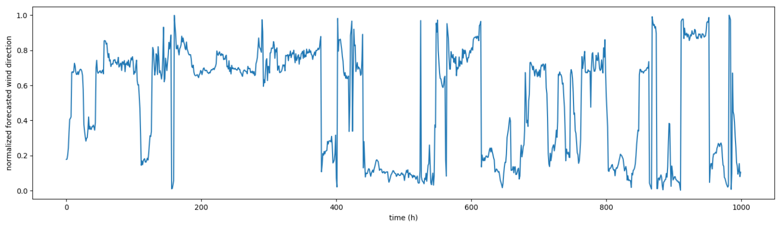
*

(b) normalized forecast wind speed of gust



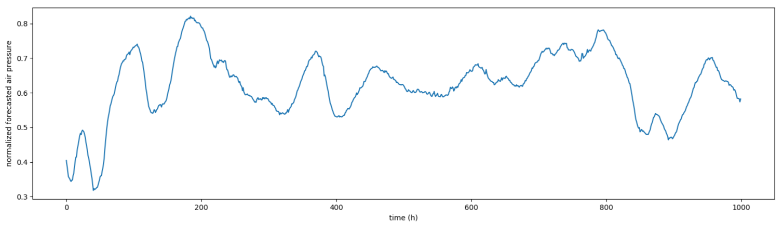
*

(c) normalized forecast temperature



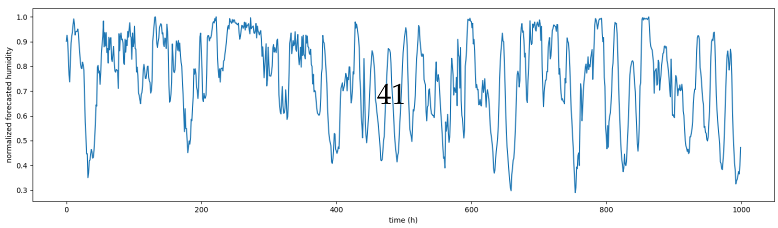
*

(d) normalized forecast wind direction



*

(e) normalized forecast air pressure



*

4.3 Experimental Environment

All the code written for this thesis is done so in Python (v. 2.7.12) primarily utilizing the following libraries:

- Pandas (v. 0.22.0)
- NumPy (v. 1.13.3)
- Scikit-learn (v. 0.19.1)
- Maplotlib (v. 3.0.3)

Pandas is used for loading and categorizing the data sets which are stored in csv files (comma separated values) [47]. NumPy is a library for handling and doing operations on matrices, which is much more memory effective compared to the native Python list [53]. Scikit-learn provides libraries for both the machine learning algorithms and grid search for optimizing their hyper parameters. It also provides tools scaling of data and ranking of potential features through mutual info regression [54]. Graphs presented in the following chapters is produced using matplotlib [36].

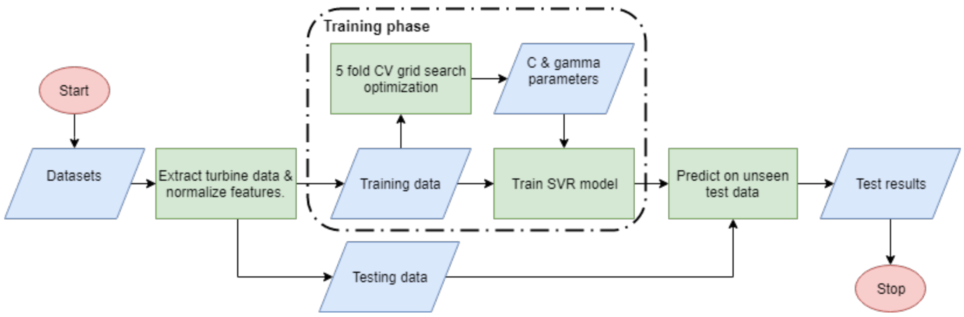


Figure 4.6: Simplified flowchart of the prediction process

Figure 4.6 shows a basic flowchart of the experimental environment. In the first step, measured wind park data is merged with the forecast weather data. Depending on the turbine and features chosen in the configuration, data is further extracted from the merged data set. Forecast weather data is only extracted from the grid point closest to the chosen turbine. Overall, eighteen months of data is used. The first twelve months is used for optimizing and training the models, and the remaining six months is used for testing. From the flowchart we can see that the test data is completely withheld from the training phase.

In the training phase the C and γ parameter of the SVR model are chosen from doing a 5-fold grid search on a predefined set of possible values. The computation cost to doing grid search can quickly become substantial based on the length of the training data and the range of possible hyper parameters. Since the inclusion of new features can effect the choice of hyper parameters, this has to been done for all the tests performed. To combat time consumption to some degree, grid search is only done on half of the training data. More specifically on every other record, such that data from all months is included. Initial experiments were done in order to find a suitable set of possible C and gamma values ending up with the following grid:

$$C = [0.01, 0.1, 1.0, 10.0, 100.0, 1000.0, 10000.0, 100000.0]$$

$$\gamma = [0.00001, 0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0]$$

The kernel function utilized in all the tests is Radial Basis Function (RBF). This choice is based on its popularity in literature.

After fitting hyper parameters have been chosen, the SVR model (configured with these parameters) are trained on the whole training set and then gets applied the unseen test data.

4.4 Initial Ranking of Features

Some shortcuts have been done when exploring the research questions which this thesis sets out to address. As an example, doing wind power forecasting regression using temperature alone as an explanatory feature would probably not yield satisfying results. Drawing from the intuition that excluding some features, such as past wind power output and past wind speed, from our feature vector will worsen our prediction accuracy, not all combinations are examined. This assumption is also reinforced by exploring the research done on the subject. We have not found a proposed ML-based model for doing short-term wind power prediction which does not either use past wind power output or past wind speed as feature. An initial mutual information (MI) regression analysis has been done to rank the available features after how much information they share with the target. Features are then in turn added to the feature space based on their ranking.

Chapter 5

Effects of Measured Wind Park Data for Turbine Prediction

Choosing appropriate features for the machine learning algorithm is essential for achieving good accuracy of predictions. When dealing with short term predictions, historic values are usually considered. As explained in chapter 3 there are differences in which features has been used as input for short-term wind power prediction throughout research. Intuitively, historic wind power is a critically important feature when predicting on short horizons with a statistical approach. Secondly, historic wind speed are in many cases included in the input. All work on this subject known to the author utilizes at least one of these two attributes, but from this point we have seen variation on whether other historic variables are used for prediction. In chapter 5 we saw that some work take into account wind direction, while others considers temperature as an input. Several work also includes several measurements further back in time of these attributes, called lag values and finally we saw a couple of proposed models which considered historic data from neighboring turbines as input. What motivates the choice, and for this matter, the exclusion of some attributes is not necessarily accounted for. It can be a question of availability of different types of attributes, intuition or testing.

In this chapter RQ1, RQ2 and RQ3 from section 1.2.1 are addressed by exploring different input configurations selected from the wind park data sets and their effect on prediction accuracy. For each wind park we examine one turbine chosen at random.

In section 5.1 we examine the effects that five different attributes have on prediction accuracy for two turbines in order to determine which attributes should be included as input when doing short term wind power prediction. The five attributes are wind power, wind speed, wind direction, temperature and yaw drive. Besides yaw drive, these are all attributes which have been used for this purpose.

In section 5.2 we explore the effects of using lag values of wind power and wind speed for the same two turbines. The maximum lag window width explored is limited to four, meaning wind power and wind speed measurements further back than four hours ago are not considered.

In section 5.3 we look at the effects of adding wind power and wind speed measurements from neighboring turbines to the input by varying the numbers of neighbors considered.

In the last section we combine the best configurations from previous sections, using both lag values and measures from neighboring turbines to see if further improvement in prediction accuracy is gained.

All results is compared against the persistence model, should we will refer to as the naive approach.

5.1 Present Measured Data

In this section we examine the effects of using all available attributes in the measured data sets on prediction accuracy for two wind turbines, one located at each park. Twelve months of hourly sampled data is used for training the models, and half of this data is used for finding suitable hyper parameters. The following six months of data is used for testing the performance of the models. In all, twenty models are trained and tested, ten for each turbine.

As mentioned in section 4.4, an initial MI regression analysis is done to determine which order the attributes is added to the input space. The result from this analysis for the turbines at wind park I and II is shown in Figure 5.1 and 5.2. Not surprisingly past wind speed and wind power gets a significantly higher score than the remaining attributes. For both turbines, wind power is ranked first, followed by wind speed and yaw direction. For the two remaining attributes we can see that wind direction is ranked before temperature for the turbine at wind park I, but not for the other turbine. Therefore, the order in which features are added to the input space is not the same for both turbines. Further we can see that temperature surpasses yaw direction in importance for the turbine at wind park II when the forecast horizon increases. For simplicity however, the feature ranking for forecast horizon 1 is followed for all test cases.

5.1.1 Turbine at Wind Park I

Table 5.1: Input Configurations - Turbine at Wind Park I

I	wp_0
II	wp_0, ws_0
III	wp_0, ws_0, yaw_0
IV	wp_0, ws_0, yaw_0, wd_0
V	wp_0, ws_0, yaw_0, wd_0, temp_0

The ranking from the MI regression analysis for the first turbine gives the input configurations shown in Table 5.1. The zero digit which follows

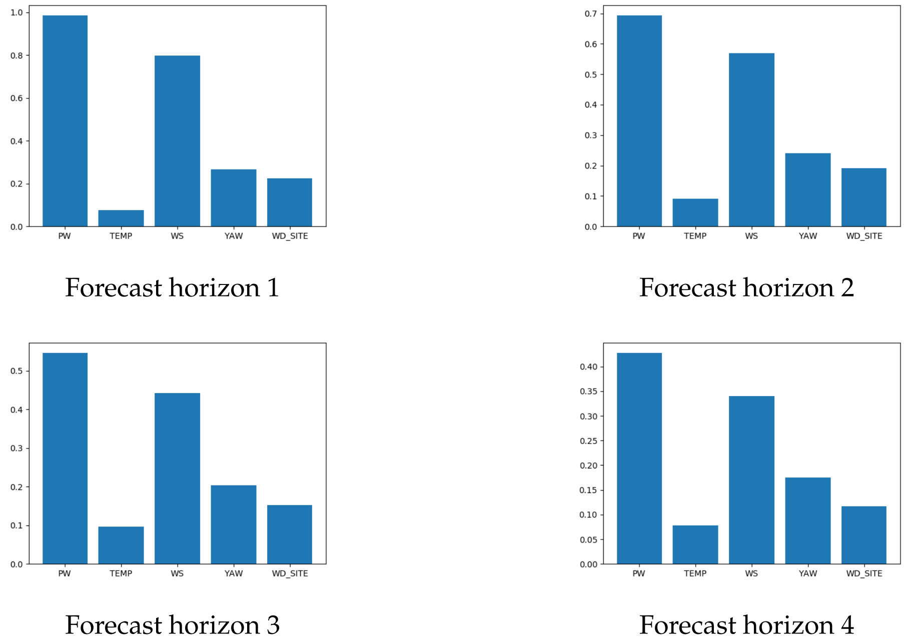


Figure 5.1: Mutual information scores of attributes for turbine at wind park I

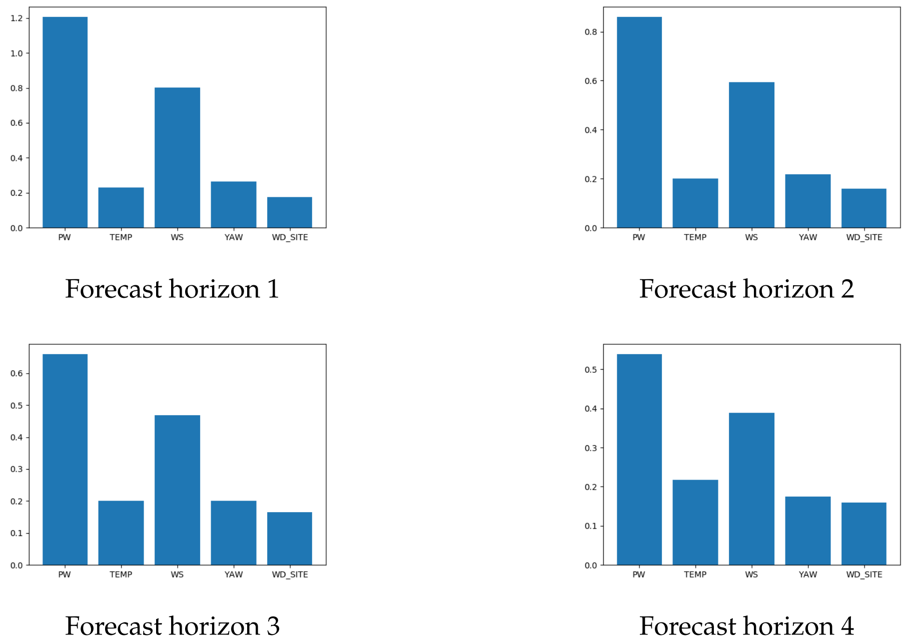


Figure 5.2: Mutual information scores of attributes for turbine at wind park II

the underscore on each feature tells us it is the most recent measurement available at the time. Prediction accuracy, using the different input configurations is scored by the RMSE metric. Scores are shown in Table 5.2 with the best results highlighted in bold.

Table 5.2: RMSE - Turbine at Wind Park I (KW)

Input	**1h ahead**	**2h ahead**	**3h ahead**	**4h ahead**
I	307.4	375.3	429.1	469.3
II	**300.2**	**365.3**	**423.1**	**461.8**
III	305.2	370.4	**423.1**	461.9
IV	302	370.6	423.5	461.9
V	305.1	370.4	423.6	462.6
naive	313.1	386.9	449.4	499.8

From the results we can see that including wind speed as a feature in addition to wind power has a positive effect on prediction accuracy for all considered forecast horizons. When compared to only using wind power as input, we see that improvement is most evident in one and two hour ahead predictions. A decrease in RMSE of 2.34% is obtained for one hour ahead prediction and 2.66% for two hour ahead prediction. For three and four hour ahead predictions the improvement is less evident, a decrease of 1.4% and 1.6%. Expanding the feature space further however shows no additional improvement. We can see that adding yaw to the feature space actually increases the error of the model in forecast horizon one, two and four. For input configuration IV, where, wind direction is included, there is a decrease in RMSE of 1.04% compared to the previous input configuration. However, it still performs worse compared to only using wind power and wind speed as input. This is also the case when predicting using all available attributes.

Further we can see that all feature configurations outperform the naive approach by a good margin. Not surprisingly this margin decrease when the horizon gets shorter, since it is in shortest time frames the persistence model can be difficult to beat. The best best performing input configuration II shows a decreases in RMSE of 4.12%, 5.58%, 5.85% and 7.6% in one, two, three and four hour ahead predictions when compared to the naive approach.

5.1.2 Turbine at Wind Park II

As mentioned in section 5.1, the attributes ranking from the MI analysis for the turbine at wind park II differs slightly from the turbine at wind park I. As a consequence we include wind direction last, and not temperature which was previously the case (shown in table 5.3). Besides from this, experiments are identical, using the same amount of training and testing

Table 5.3: Input Configurations - Turbine at Wind Park II

I	wp_0
II	wp_0, ws_0
III	wp_0, ws_0, yaw_0
IV	wp_0, ws_0, yaw_0, temp_0
V	wp_0, ws_0, yaw_0, temp_0, wd_0

data, and setting the hyperparameters of the SVR-models based on cross validation.

Table 5.4: RMSE - Turbine at Wind Park II (KW)

Input	**1h ahead**	**2h ahead**	**3h ahead**	**4h ahead**
I	430.2	528.1	602	645.9
II	425.6	525.8	595.5	641.5
III	425.5	525.5	595.4	640.7
IV	425.5	525.4	595.5	**640.5**
V	**425.4**	**525.2**	**595**	640.6
naive	437.8	549.5	633.3	690

For the turbine at wind park II we can see from table 5.4 that using wind speed in addition to wind power as input also in this case results in a more accurate model. For one hour ahead a decrease in RMSE of 1.07% is obtained. Two, three and four hours ahead predictions shows a 0.44%, 1.08% and 0.68% decrease. In contrast to the turbine at wind park I, we can see that by including yaw to the feature space, accuracy is improved even further in all forecast horizon. A slight improvement is also obtained when temperature is introduced. This configuration gave the best performing model for four hour ahead prediction. Finally, using all the available attributes as input resulted in the best performing model for the three shortest time frames. However, the amount of improvement gained beyond input configuration II is small. Between configuration II and V for one, two and three hours ahead there is a decrease in RMSE of 0.05%, 0.11% and 0.08%. The best model for four hours ahead, using configuration IV, decreased RMSE by 0.16% from the model utilizing only wind power and wind speed as input. Finally, we can also in this case see that all developed models performed better than the persistence model. The best performing models for each forecast horizon decrease RMSE by 2.83%, 4.42%, 6.04% and 7.17% for one to four hours ahead compared to the naive approach.

5.1.3 Summary and Discussion

For both tested turbines we can observe that including wind speed into the feature space has a positive effect on prediction accuracy in all forecast horizons. The improvement is most evident for the turbine at wind park I, and more specifically in the two first forecast horizons. The effects of yaw, temperature and wind direction as features is less clear. For the second turbine we observe a slight improvement as these attributes are added to the input, but for the first turbine they have a negative effect in all forecast horizons. From these results it seems reasonable to include measured wind speed as input, in addition to measured wind power. It is harder to give a general account on the effects of yaw, wind direction and temperature. For the turbine at wind park II one could argue they should be included. But even though these attributes had a positive effect on prediction accuracy in one of the tested turbines, the amount of improvement is small. In both cases we can see an improvement in accuracy over the persistence model with a comfortable margin, and as prediction horizon expands this gets even more evident.

5.2 Including Lag Values

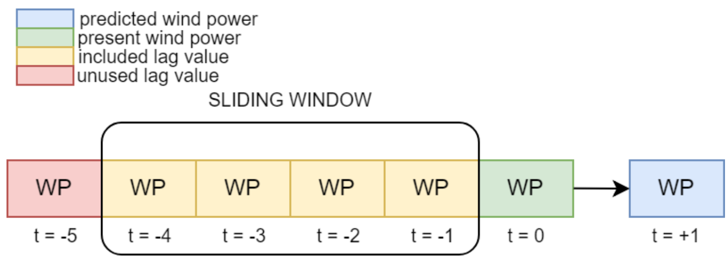


Figure 5.3: Included lag values with a sliding window width of four

Another frequently used feature when performing short term wind power prediction are lag values. The intuition behind the use of lag values is that by providing the machine learning algorithm with knowledge of what happened in recent history, and not just at the present moment, will give the model more context. The use of lag values is often referred to as the sliding window method. Figure 5.3 illustrates the basic idea.

In these experiments the maximum sliding window is limited to a width of four, meaning lag values further than four hours back in time is not examined. This decision is based on the assumption that a window of this size is sufficient enough to see if there is a trend in performance as the sliding window expands. and also the fact that we do not predict further than four hours into the future. Also, lags of which attributes to included in the window needs to determined. Based on the results from section 5.1, which showed little no improvement with the use of yaw, wind direction and temperature, only lags of wind power and wind speed is considered. Input configuration II is set as the initial features for both turbines, since this configuration gave successful results in both cases. All the tested lag configurations are shown in table 5.5.

Table 5.5: Input Configurations - Including Lag Values

i	wp_1
ii	wp_1, wp_2
iii	wp_1, wp_2, wp_3
iv	wp_1, wp_2, wp_3, wp_4
v	wp_1 ws_1
vi	wp_1, wp_2 ws_1, ws_2
vii	wp_1, wp_2, wp_3 ws_1, ws_2, ws_3
viii	wp_1, wp_2, wp_3, wp_4 ws_1, ws_2, ws_3, ws_4

5.2.1 Turbine at Wind Park I

The results for the turbine at wind park I is displayed in Table 5.6. For one hour ahead prediction we can observe that prediction accuracy was improved for two input configurations. In both these cases, the sliding window width is set to one. The best performing configuration, II + v includes lag values of both wind speed and wind power and gave av decrease in RMSE of 1.13% when compared to the initial II configuration. Using only lags of wind power shows a decrease of 0.77%. Using a wider sliding window increased RMSE for all cases. For two hours ahead, tests show no improvement in any of the lag configurations. The best performing models had an increase in RMSE of 0.82% compare to using only present values as input. This score is shared by three different configurations. In all the three cases lag values of wind power only was included. For the two remaining forecast horizons, we see that lag values gave a positive effect on accuracy. For both three and four hour ahead prediction, including lag values has a positive effect on accuracy. In both forecast horizons, the best configurations did not include lag values further back than one hour ago. For three hour ahead prediction the best performing model included lags of both wind speed and wind power and gave a decrease in RMSE of 0.26% over the initial configuration. For four hour ahead, the best configuration gave a decrease in RMSE of 0.84%, but used lag of wind power only.

Table 5.6: RMSE - Turbine at Wind Park I - Including Lag Values (KW)

Input	1h ahead	2h ahead	3h ahead	4h ahead
II + i	297.9	**368.3**	422.2	**457.9**
II + ii	300.4	**368.3**	422.3	461.3
II + iii	300.4	**368.3**	422.3	461.2
II + iv	300.4	368.4	422.4	460.9
II + v	**296.8**	368.5	**422**	460.9
II + vi	300.8	368.6	422.3	461
II + vii	300.8	368.7	422.3	460.9
II + viii	300.8	369	422.3	461
II	300.2	365.3	423.1	461.8
naive	313.1	386.9	449.4	499.8

5.2.2 Turbine at Wind Park II

For the turbine at wind park II, the first thing that can be noted from the results shown in Table 5.7 is that the inclusion of lag values improves accuracy in all forecast horizons. We can further observer that the best model in each time frame does only include lag values of wind power. For one hour ahead prediction, a decrease in RMSE of 0.33% is gained when using wind power lags up to three hours back in time. For the remaining forecast horizons we can observe that the best giving configuration uses a sliding window of width four. This suggests that using a sliding window wider than four, which I have set as a limit in these tests, could improve the accuracy even further for these time frames. For two hour ahead the best gave decrease in RMSE of 0.26% compared to the model based on configuration II. For three and four hours ahead, prediction accuracy is improvement by an even larger margin, showing a decrease of 0.50% (3h ahead) and 0.57% (4h ahead) in RMSE.

Table 5.7: RMSE - Turbine at Wind Park II - Including Lag Values (KW)

Input	1h ahead	2h ahead	3h ahead	4h ahead
II + i	424.7	525.7	595	639.9
II + ii	424.7	525.1	594.1	638.4
II + iii	**424.2**	525	592.8	637.9
II + iv	424.9	**524.4**	**592.5**	**637.8**
II + v	425	525.8	594.8	639.5
II + vi	425.9	525.6	594.3	638.4
II + vii	424.8	525.5	593.4	638.1
II + viii	425.7	525.3	593.1	637.9
II	425.6	525.8	595.5	641.5
naive	437.8	549.5	633.3	690

5.2.3 Summary and Discussion

When comparing the results from the two turbines it gets difficult to give a general account on the use of lag values. For the turbine at wind park I, including lag values from further back than one hour was fruitless. This is not the case for the second turbine, in which three out of the four best performing models has the sliding window width set to the limit of four. Another difference between the two turbines is in the use of only wind power or both wind power and wind speed values as lags. For the first turbine, the best result at forecast horizon 1 and 3, both attributes are utilized, while for two and four hour ahead predictions, only lags of wind power is used. For the turbine at wind park II however, it seems safe to say that only lags consisting of wind power should be included into the feature space. Overall, taking into account that in seven out of eight cases the inclusion of lag values improved accuracy over using only present measures, lags should be considered as an input in short term wind prediction. Which attributes should be lagged, and how wide the sliding window should be is difficult to say.

5.3 Including Measurements From Neighboring Turbines

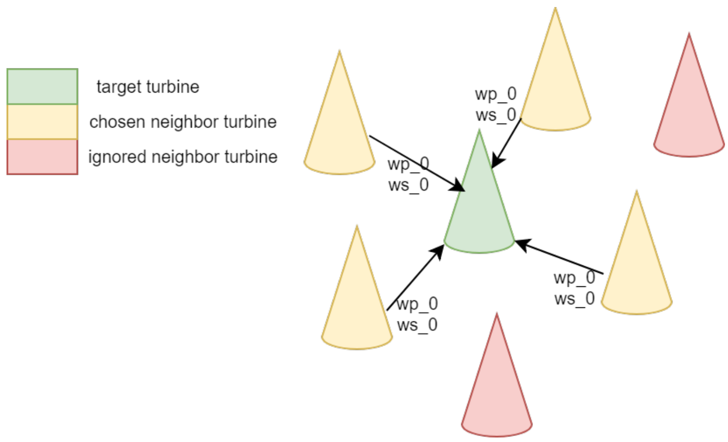


Figure 5.4: selection of the closest neighbors with a limit of four

In this section we examine the effect of incorporating measures from neighboring turbines to the input. The basic idea is illustrated in Figure 5.4. In this scenario we need to determine both what type of data we should include from the neighboring turbines, and from how many neighboring turbines. To address the former, we rely upon the findings from section 5.1 by only considering wind power and wind speed measurements. In order to determine the number of neighbors, six different sizes is examined by using measures from two to twelve neighbors, excluding odd numbers. The Selection of neighbors is based on their geographical positioning to the target turbine, such that the closest turbine is selected first. As in previous sections, present wind speed and wind power (input configuration II) is used as initial input on which the neighboring measures are added to.

5.3.1 Turbine at Wind Park I

The results for the turbine at wind park I is shown in 5.8. It is clear that using neighbor turbines is a successful strategy in this case. All configurations considered improves accuracy compared to only using measurements from the target turbine as input. We can further observe a steady decreases of RMSE as the neighborhood is expanded. In three of the four tested forecast horizons, the limit of twelve neighbors is considered in the input space, which suggest that considering an even bigger neighborhood could improve accuracy additionally. For one hour ahead the configuration including ten neighbors shows a significant improvement in RMSE with a decrease of 3.2% compared to configuration II. For two to four hours ahead, a decrease of 3.45%, 4.66% and 4.29% is obtained using wind power and wind speed measurements from twelve neighbors. The success of using data from neighbors is further emphasized when compared to the naive approach, which show a decrease in RMSE of

7.18%, 8.84%, 10.23% and 11.56% for forecast horizon one to four.

Table 5.8: RMSE - Turbine at Wind Park I - Including Neighbors (KW)

Input	1h ahead	2h ahead	3h ahead	4h ahead
II + 2 neighbors	298	362.6	415.2	453.8
II + 4 neighbors	294.8	355.9	407.5	449.9
II + 6 neighbors	292.4	354.7	405.4	443.7
II + 8 neighbors	292.3	355.8	407.6	446
II + 10 neighbors	**290.6**	353.8	404.7	442.9
II + 12 neighbors	293.2	**352.7**	**403.4**	**442**
II	300.2	365.3	423.1	461.8
naive	313.1	386.9	449.4	499.8

5.3.2 Turbine at Wind Park II

Including measurements from neighboring turbines also has a positive effect on the turbine located in wind park II. All except one configuration outperforms the use of only present wind power and wind speed as input. For one hour ahead we can see a steady decrease in RMSE up to using ten neighbors, with a slight increase when twelve neighbors are used. This steady decrease is not apparent for the remaining forecasts horizons, although the trend seems to be that accuracy is improved when the neighborhood is expanded. In the four hour ahead case, the model using ten neighbors actually gave the worse results, even worse than not using data from neighbors at all, while including twelve neighbors yield the best RMSE score. The effect of neighbor data is most apparent in the shorter forecast horizons. The best performing models shows a decrease in RMSE of 2.04%, 1.65%, 1.36% and 0.88 for one to four hour ahead predictions. This amounts a significant decrease in RMSE over the persistence model by 4.77%, 5.89%, 7.24% and 7.85%

Table 5.9: RMSE - Turbine at Wind Park II - Including Neighbors (KW)

Input	1h ahead	2h ahead	3h ahead	4h ahead
II + 2 neighbors	420.8	519.1	592.2	640.1
II + 4 neighbors	422	521.6	592.8	642.8
II + 6 neighbors	419.9	522.1	593.9	638.4
II + 8 neighbors	418.8	521.5	592.3	640.3
II + 10 neighbors	**416.9**	**517.1**	587.8	647.4
II + 12 neighbors	417.2	517.3	**587.4**	**635.8**
II	425.6	525.8	595.5	641.5
naive	437.8	549.5	633.3	690

5.3.3 Summary and Discussion

It is clear that including of neighbor features has a positive effect on prediction performance for both turbines. From all the potential feature types analyzed in this work up to this point, it has the most profound effect on performance. The improvement is especially apparent for the turbine at wind park I. Accuracy seems to be further improved with the increase of neighbors included into the input space. From all the eight cases, five of the best input configurations includes the limit of twelve neighbors, while the remaining three cases incorporates data from ten neighbors. It seems that forecast horizon influence the appropriate size of neighborhood, considering that in both cases using twelve neighbors did not improve performance in one hour ahead prediction. For the longer time frames the performance gained by using the limit of twelve neighbors suggest an even wider neighborhood than twelve could improve accuracy even further in these forecast horizons.

5.4 Combination of Lag Values and Neighboring Turbines

Up to this point we have treated the potential feature types partly separated. When the effect of neighboring measurements was examined, lag values of the target turbine was not included in the input and vice versa, even though they overall showed to have a positive effect on prediction accuracy for both turbines. In this section we use the results drawn from the previous three sections to see if accuracy can be improved even further by combining the different input types. Intuition may probably tell us that it will, but this is not necessarily the fact. We recall that for each test run a grid search is done to find suitable hyper parameters for the SVR model. The input utilized obviously has an effect on the hyper parameters chosen. Therefore, a suitable pair of hyper parameters for one type of input data might not be suitable for other input types, and could potentially lead to a worse performing model.

The input patterns tested in this section combines the best performing input configurations from section 5.2 and 5.3. For the turbine at wind park II, we saw that including yaw drive, wind direction and temperature gave a slight improvement in accuracy, but for the sake of comparison, these attributes is not utilized in this section. Secondly, for the turbine at wind park I we have included one lag value for two hour ahead predictions even though this did not yield improvement in accuracy compared to using only present measurements. The exact input configurations tested can be seen in Table 5.10 for the turbine at wind park I and in Table 5.12 for the turbine at wind park II.

5.4.1 Turbine at Wind Park I

Table 5.10: Input Configurations - Turbine at Wind Park I - Combination

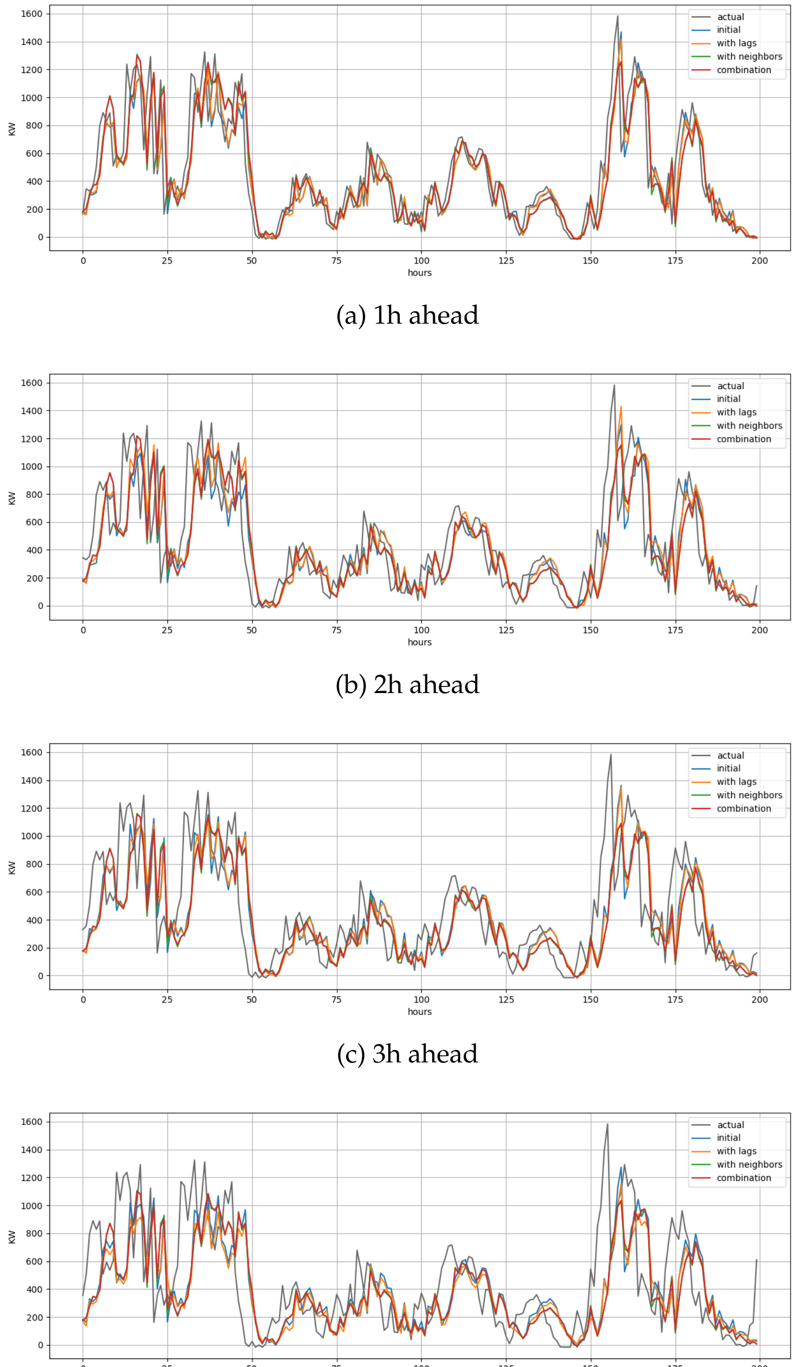
1h ahead	wp_0, ws_0, wp_1, ws_1, 10 neighbors(wp_0 & ws_0)
2h ahead	wp_0, ws_0, wp_1, 12 neighbors (wp_0 & ws_0)
3h ahead	wp_0, ws_0, wp_1, ws_1, 12 neighbors (wp_0 & ws_0)
4h ahead	wp_0, ws_0, wp_1, 12 neighbors (wp_0 & ws_0)

By combining data from neighbors and lag values we can see from table 5.11 that accuracy is improved even further for one hour and two hour ahead predictions. For one hour ahead we can observe a decrease of 0.62% in RMSE compared to the previous best performing model. For two hour ahead a slight decrease of 0.11% in RMSE is obtained, but for the remaining two forecast horizons no further improvement is gained. Recalling the results from section 5.2, which showed little improvement with the inclusion of lags for forecast horizon two to four, these results is not very surprising. A bit unexpected however is the fact that a slight increase in RMSE for four hour ahead prediction is observed, considering both input types showed improvement in isolation.

In figure 5.5 a sample from the prediction results in each forecast horizon is displayed along with the actual observed values. We can see in the results that the models including data from neighbor turbines generalizes better than the other models. Both the neighbor model and the combination model is less sensitive to the fluctuations in the power output, which can vary by a large amount from one hour to the next.

Table 5.11: RMSE - Turbine at Wind Park I - Combination (KW)

Input	1h ahead	2h ahead	3h ahead	4h ahead
initial (wp_0 & ws_0)	300.2	365.4	423.1	461.8
with lags	296.8	368.3	422	457.9
with neighbors	290.6	352.7	**403.4**	**442**
combination	**288.8**	**352.3**	**403.4**	442.3
naive	313.1	386.9	449.4	499.8



(a) 1h ahead

(b) 2h ahead

(c) 3h ahead

(d) 4h ahead

Figure 5.5: Sample data from prediction results for wind turbine at park I

5.4.2 Turbine at Wind Park II

Table 5.12: Input Configurations - Turbine at Wind Park II - Combination

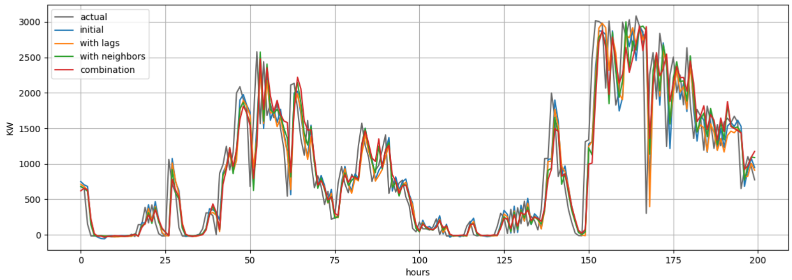
1h ahead	wp_0, ws_0, wp_1, wp_2, wp_3 10 neighbors(wp_0 & ws_0)
2h ahead	wp_0, ws_0, wp_1, wp_2, wp_3, wp_4 10 neighbors (wp_0 & ws_0)
3h ahead	wp_0, ws_0, wp_1, wp_2, wp_3, wp_4 12 neighbors (wp_0 & ws_0)
4h ahead	wp_0, ws_0, wp_1, wp_2, wp_3, wp_4 12 neighbors (wp_0 & ws_0)

From the results in table 5.13 we can see additional improvement in all forecast horizons by combining lag values and neighbor turbines for the turbine in wind park II. The improvement is most evident in four hour ahead prediction, showing a decrease in RMSE of 0.72%. Less significant is the improvement for one hour ahead, which decreased RMSE by 0.01%, although this is not very surprising considering the amount of improvement the inclusion of lag values had in isolation. For forecast horizon two and three, the decrease in RMSE amount to 0.25% and 0.18%.

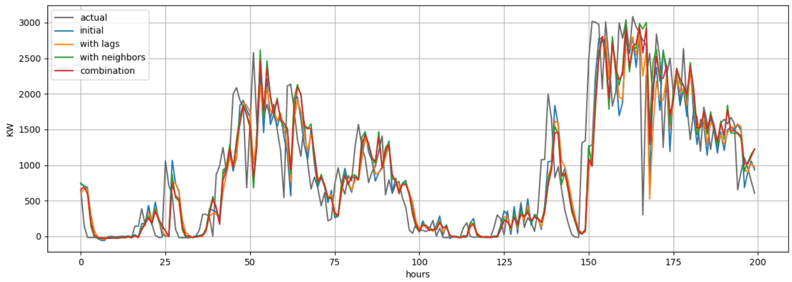
Samples from the prediction results for each model is shown in figure 5.6. As for the turbine at park I, we can also in this case see that the models utilizing neighbor data are less sensitive to the high variance in power output, resulting in better prediction results.

Table 5.13: RMSE - Turbine at Wind Park II - Combination (KW)

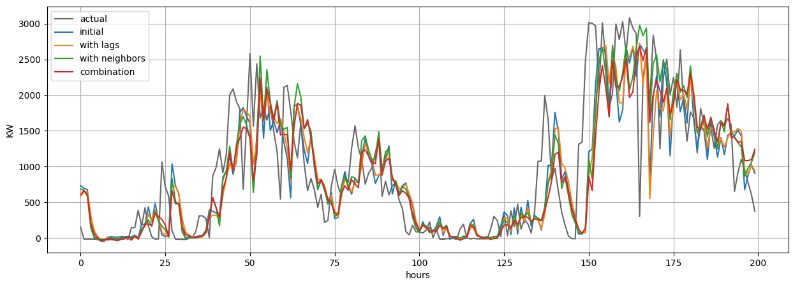
Input	1h ahead	2h ahead	3h ahead	4h ahead
initial (wp_0 & ws_0)	425.6	525.8	595.5	641.5
with lags	424.2	524.4	592.5	637.8
with neighbors	416.9	517.1	587.4	635.8
combination	**416.5**	**515.8**	**586.3**	**631.2**
naive	437.8	549.5	633.3	690



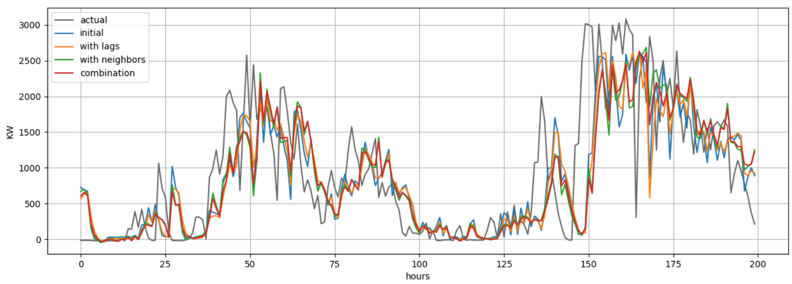
(a) 1h ahead



(b) 2h ahead



(c) 3h ahead



(d) 4h ahead

Figure 5.6: Sample data from prediction results for wind turbine at park II

5.4.3 Summary and Discussion

Combining data from neighbor turbines with lag values of the target turbine seems to be a reasonable strategy for both tested turbines. In six out of the eight cases (four horizons for each turbine), further improvement in accuracy is achieved. In the remaining two cases, one case shows a slight decrease in accuracy, while the other had no effect in either positive of negative direction. However, the amount of improvement is small. This is not unexpected considering the impact of lag values alone, addressed in section 5.2, had on performance. Nonetheless, these results do seem to tell us that data from neighbor turbines and lag values can safely be part of the same input pattern, even if the positive effect is even so slightly. Recalling the bias-variance trade off and its relation to model complexity, discussed in section 2.1.5, these results show that input size of well over twenty features does not make the model too complex. The combination model for the turbine at wind park II relies on as many as thirty inputs.

5.5 Conclusion

In this chapter we set out to answer RQ1, RQ2 and RQ from 1.2.1 by examining the effect a wide variety of feature candidates has on prediction accuracy using a machine learning method.

In the first section of this chapter we examined if including less obvious attributes recorded at a turbine to the input for our machine learning algorithm could help boost prediction accuracy. The attributes considered were wind power, wind speed, yaw drive, temperature and wind direction. For both turbines, including wind speed proved to be a positive addition to the input pattern along with present wind power. The two turbines had a differing effect on the rest of the attributes. While all the candidate features generally had a slight positive effect on the turbine at the second wind park, for the first turbine prediction accuracy got worse by their introduction. From this fact its not possible to give a general account on the use of these less explanatory attributes. Considering the small amount of improvement gained for the second turbine, it is tempting to conclude that their effect is too small to bother. What is clear however is that these attributes should not uncritically be added to the feature space without consideration. What seems conclusive from this section is that wind speed should be included in the input pattern if available.

In section 5.2 we examined the inclusion of lag values to the input pattern referencing RQ2 from section 1.2.1. Results showed that lag values had a positive effect in all tested cases expect for two hour ahead prediction for the turbine at wind park I. It is worth noting that the effect of lag values are in most cases not of big significance. In only one case, one hour prediction for the turbine at wind park I, the decrease in RMSE exceeded 1%. A part of RQ2 was also to see discover which data types should be should be lagged, and wide the lag window should be. These question are difficult to answer based on the present results. We saw that for the first turbine, only the most recent lag value had a positive effect on accuracy, while for the second turbine, a width of three and four seemed most promising. Differences in the two examined turbines were also apparent in the choice of lagged data types. For the first turbine, the best results for two forecast horizons, lag values of both wind speed and wind power were utilized. In the second turbine however, only lag values of wind power were included for the best yielding input configurations. From this it can be argued that from a general perspective at least lags of wind power should be considered as a part of the input pattern, but if lags of wind speed and how wide the lag window should be, needs to be examined for the specific turbine or park.

RQ3 concerned the use of data from neighboring turbines. This question was examined in section 5.3 for the two turbines using present wind speed and wind power measures for up to twelve neighbors in vicinity. For both turbines this had a positive effect on prediction accuracy, with improvements most apparent for the turbine at wind park I. For the second turbine improvement were most significant in the shortest time frame, with a decreased effect as the forecast horizon expanded. Results suggests that this strategy should be deployed when predicting the power output of a turbine. For both turbines, using neighbor data had a greater effect compared to using lag values. In addition, using neighbor turbines decreased by a greater margin than the inclusion of present wind speed. Further we discovered a trend which showed that predictions were further improved as data from a larger number of incorporated neighbors. By analyzing the results there seems to be a tipping point at ten neighbors for one hour ahead prediction for both turbines, while for three and four hour ahead the max limit of twelve neighbors yield the best result for both turbines. The last observations eludes to that for longer forecast horizons data from an even larger number of neighbors can improve prediction accuracy even further.

In the last section we examined the effect of using both lag values and neighbor turbines in the input pattern. Results showed that in all but one of the cases prediction accuracy was improved even further.

Chapter 6

Wind Park Prediction Using Measured Wind Park Data

So far in this work we have only focused on predicting wind power production for single turbines. For most use cases however we are interested in estimated wind power production for the whole wind park. In this chapter we look at how the feature configurations examined so far performs when scaled up to park level. In addition RQ5 is addressed by comparing two different strategies for predicting wind parks. In the first approach each turbine is predicted individually and then summed up to get the final wind power forecast. This strategy is illustrated in figure 6.1. The second, and simpler, approach we aggregate the data from all turbines and predict the power output using a single model. Figure 6.6 illustrates the strategy.

As mentioned in 4.1.1 some turbines have been left out of these experiments. This concerns four turbines at wind park I and eleven turbines and wind park II. The reason being that analysis of the data showed signs of some turbines being shut off for longer periods of time. The turbines in question would contribute to a misleading result, based on the assumption that in a real world scenario we would have prior knowledge to when turbines will experience down time. Overall twenty turbines is included for wind park I and twenty-two turbines for wind park II.

6.1 Strategy 1 - Predicting Individual Turbines

For predicting each turbine at a wind park, I have used the same approach as in chapter 5. To refresh, twelve months of hourly data is utilized for training the models. One model for each turbine. Half of this training data is used for finding suitable hyper parameters to the SVR model through grid search, and finally tested on 6 months of data. The input patterns chosen for each turbine is based on findings from chapter 5. The exact input configurations used for all turbines in the different wind parks are the same used in section 5.4 which can be found in Table 5.10 for wind park I and 5.12 for wind park II.

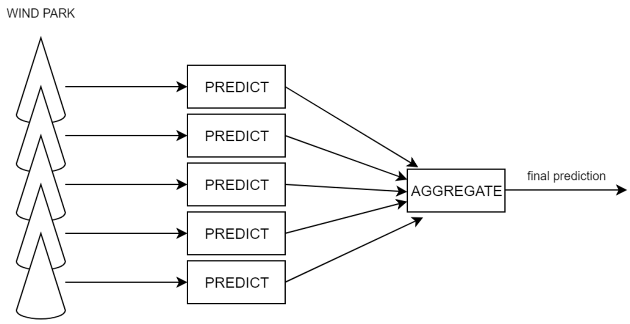


Figure 6.1: Strategy 1. Each turbine has its own predictive model. The prediction from every model gets summarized to produce the final wind park prediction

6.1.1 Wind Park I

Table 6.1: RMSE - Wind Park I - Strategy 1 (MW)

Input	1h ahead	2h ahead	3h ahead	4h ahead
initial (wp_0 & ws_0)	4.79	6.313	**7.473**	**8.346**
with lags	**4.771**	**6.31**	7.48	8.381
with neighbors	4.793	6.338	7.497	8.359
combination	4.779	6.328	7.484	8.372
naive	4.878	6.528	7.844	8.898

From the results in Table 6.1 we can see that the success of using data from neighboring turbines does not apply when scaled up to wind park level. In all forecast horizons the introduction of neighbor turbines actually decreases prediction performance. This is in direct contradiction to the results in section 5.3. In order to get more insight to these rather unexpected results, figure 6.2 shows the normalized RMSE (NRMSE) for each of the tested input patterns and how they compare to each other as the number of turbines included in the final one hour ahead prediction output is increased. We can see that using neighbor data in the input does provide a better performance when power output from fewer turbines are considered. Although, as number of turbines is increased, the models based on the simpler input patterns (lags or only present values) surpasses in accuracy. For the approach using lags, the accuracy becomes better than the neighbor approach when the final power forecast includes individual predictions from nine turbines and more. The combination approach is surpassed in accuracy after thirteen turbines.

The best prediction results was obtained by using lag values for the two shortest forecast horizons, but for three and four hour ahead predictions even this approach performs worse than using only present wind power and wind speed as input. This not in agreement with the results from turbine prediction in section 5.2. However, we can observe that all input

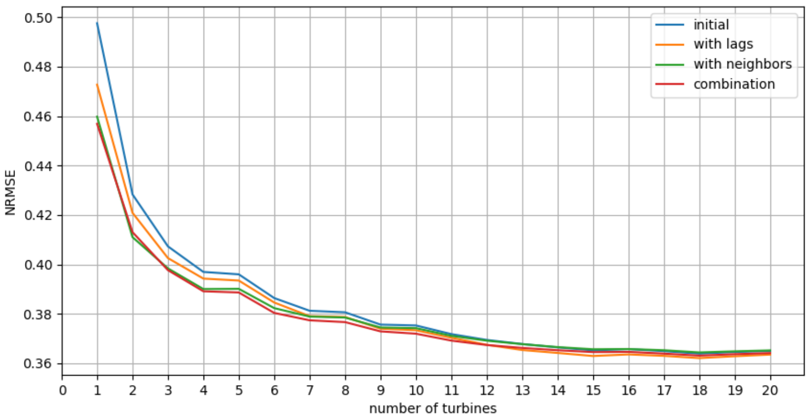
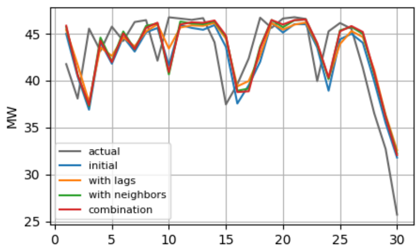
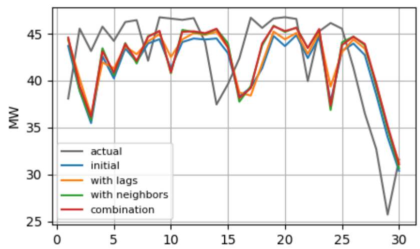


Figure 6.2: NRMSE

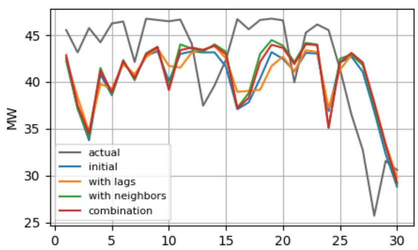
configurations performs better than the naive approach. For one hour ahead, there is an improvement of 2.19% with the use of lag values, and for two hour ahead a decrease of 3.34% is obtained. We can see that this margin increase the further we predict into the future. Using present wind power and wind speed decreased RMSE by 4.73% and 6.2% for three and four hours ahead predictions. Samples from predictions results of all input pattern is displayed in figure 6.3.



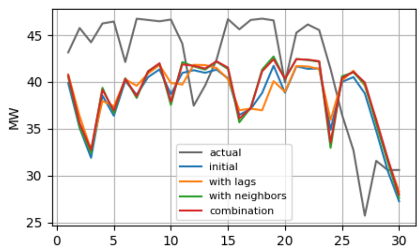
1h ahead



2h ahead



3h ahead



4h ahead

Figure 6.3: Prediction samples for wind park I

Table 6.2: RMSE - Wind Park II - Strategy 1 (MW)

Input	1h ahead	2h ahead	3h ahead	4h ahead
initial (wp_0 & ws_0)	**6.781**	**8.973**	**10.661**	**11.948**
with lags	6.803	9.078	10.761	12.092
with neighbors	6.885	9.121	10.79	12.122
combination	6.862	9.118	10.819	12.143
naive	7.019	9.387	11.304	12.826

6.1.2 Wind Park II

Prediction results from wind park II shows similar results. From table 6.2 we observe that also in this case using neighbor turbines in the input pattern has a negative effect on prediction accuracy compared to less complex input patterns. The NRMSE scores plotted in figure 6.4 shows, as was also the case for wind park I, that the approach using neighbor data rises above the simpler input patterns in NRMSE as more turbines are included in the overall prediction result. The fact that both wind park shows similar results, it seems that the use of neighboring turbines is most suitable for only predicting a single turbine, and not using this approach when predicting for the whole wind park.

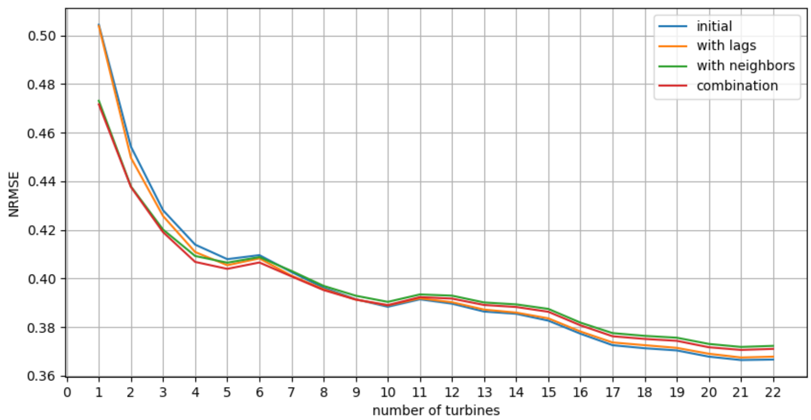
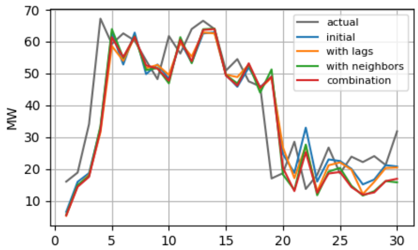
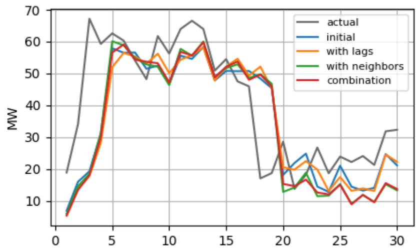


Figure 6.4: NRMSE

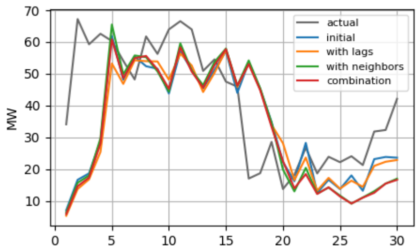
We can also see that the best performing input configuration in all forecast horizons only includes present values of wind power and wind speed, suggesting that lag values should not be included even for the shortest forecast horizons for this case.



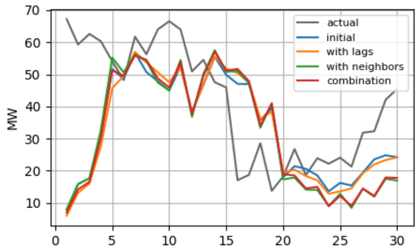
1h ahead



2h ahead



3h ahead



4h ahead

Figure 6.5: Prediction samples for wind park II

6.2 Strategy 2 - Predicting the Aggregate of All Turbines

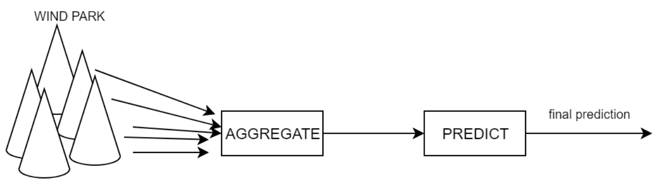


Figure 6.6: Strategy 2. The aggregate of all turbines are used to train a single prediction component for the whole wind park

A simple strategy for making our prediction process more effective time and computation-wise is to treat our wind park as one single entity. Instead of having to develop a model for each turbine, we instead train a single model which takes the aggregate from all turbines as input. When applying this strategy it makes no sense in using neighbor turbines, which we have considered as input for single turbine prediction. In addition yaw has not been included as a candidate feature, since this is an independent technical attribute of turbines. For the remaining candidate features, the average of all wind speed and temperature measure, and the sum of wind power measurements is calculated before normalization. For wind direction, which is only recorded at one spot in both wind parks, only normalization is necessary.

In order to set the best ground for comparison of the two strategies, I also in this section examine several input configurations. The patterns considered for both pars is displayed in table 6.3.

Table 6.3: Input Configurations - Strategy 2

I	wp_0
II	wp_0, avg_ws_0
III	wp_0, avg_ws_0, wd_0
IV	wp_0, avg_ws_0, avg_temp_0
V	wp_0, avg_ws_0, wd_0, avg_temp_0

6.2.1 Wind Park I

From the results in table 6.4 similar effects to performance by the less explanatory candidate features is observed. Using both wind speed and wind power in the prediction is further established as a good approach. A tiny improvement in accuracy is observed when temperature is also included to the feature space, but hardly of much significance. Wind speed and wind power in conjunction with lag values are also tested, showing no further improvement. Lastly, and most interestingly, we can see that strategy 1 outperforms strategy 2 in all forecast horizons. RMSE is decreased by 0.50% (1h ahead), 0.20%(2h ahead), 0.17%(3h ahead) and 0.22%(4h ahead) by adopting the method of predicting each turbine individually.

Table 6.4: RMSE - Wind Park I - Strategy 2 (MW)

Input	1h ahead	2h ahead	3h ahead	4h ahead
I	4.83	6.39	7.572	8.468
II	4.796	6.324	7.488	**8.365**
III	4.796	6.325	7.488	8.37
IV	**4.795**	**6.323**	**7.486**	**8.365**
V	4.796	6.324	7.488	8.367
II + lags (best)	4.833	6.365	7.507	8.394
strategy 1 (best)	4.771	6.31	7.473	8.346
naive	4.878	6.528	7.844	8.898

6.2.2 Wind Park II

Small amounts of improvements can be observed in some of the forecast horizons when including wind direction and temperature. In contrast to wind park I, we can see that using lag values improve the model further. Appropriate lags values where selected by examining the same input patterns as in section 5.2. The best performing model includes lags of wind power with a sliding window of width one. Also for this wind park, strategy 1 outperforms strategy 2 in all forecast horizons. In this case this with an even bigger margin, showing a decrease of 0.99% (1h ahead), 0.75% (2h ahead), 0.37% (3h ahead) and 0.87% (4h ahead) in RMSE compared to the best model using strategy 2.

Table 6.5: RMSE - Wind Park II - Strategy 2 (MW)

Input	1h ahead	2h ahead	3h ahead	4h ahead
I	6.949	9.197	10.903	12.206
II	6.915	9.095	10.775	12.072
III	6.916	9.097	10.775	12.072
IV	6.914	9.09	10.776	12.06
V	6.914	9.096	10.777	12.06
II + lags (best)	**6.849**	**9.041**	**10.701**	**12.053**
strategy 1 (best)	6.781	8.973	10.661	11.948
naive	7.019	9.387	11.304	12.826

6.3 Summary and Conclusion

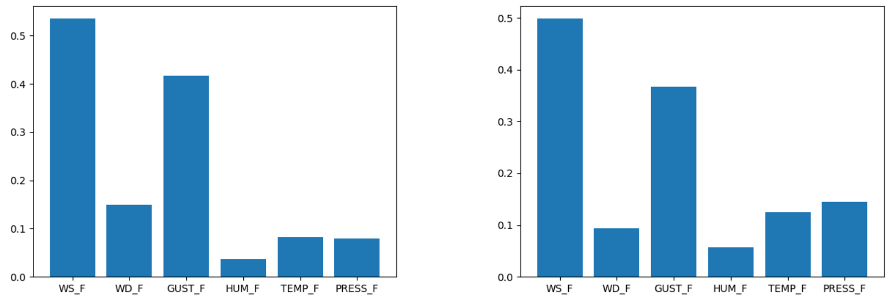
Recalling question five from 1.2.1, it seems clear that predicting each turbine individually as opposed to treating the wind park as one entity is a better strategy for wind power prediction. However, the difference in RMSE is not substantial. In a scenario where the number turbines in a wind park reaches far bigger numbers than the tested cases in this work, deploying strategy 1 in an effective manner can potentially demand huge computation powers and quickly become very costly. Considering the difference exposed in these experiments, using a simpler and more computationally efficient strategy might be more reasonable in such scenarios.

Chapter 7

Effects of External Weather Forecast Data

Surveying studies on wind power prediction, there seems to be an agreement that data from NWP models should be used as input in statistical models when the forecast horizons extends several hours. However, in short-term prediction, and particularly for one hour ahead prediction, there does not exist a consensus to the same extent on the usefulness of NWP data. In section 3.3 we saw that some studies do consider forecast weather data from external NWP models as input, although a majority of them do not. In this chapter we examine this diversity by studying the impact of external weather forecast data from a real world NWP model with respect to prediction accuracy. Several input configurations are examined in order to expose suitable features. The candidate features considered are forecast wind speed, gust, wind direction, temperature, air pressure and humidity. In the first two sections we examine the effect NWP data has on single turbine prediction. Predictions are done for the same two turbines considered in earlier chapters. In later sections, lag values and neighbor data are combined with NWP data too see if accuracy can be further approved. Finally, NWP data are used for predicting for the whole wind park, deploying both types of strategies examined in the previous chapter.

Experiments are done using twelve months of data for training, and six months of data for testing. Present wind power and wind speed measurements are used as initial features, and a ranking of the candidate attributes from the NWP model is done in advance to determine the order of inclusion. The result of the MI regression analysis is shown in figure 7.1. Not surprisingly, forecast wind speed is ranked in first place, followed by wind speed of gust for both turbines. From this point the order differs. At both turbines humidity is ranked last, but the order of temperature, air pressure and wind direction is not in alignment. Therefore, the input configurations used in this chapter ignore the ranking for these three attributes, consequently allowing for a larger set of tested input configurations. (shown in table 7.1).



(a) Turbine at wind park I

(b) Turbine at wind park II

Figure 7.1: Mutual information scores of NWP data for both turbines

Table 7.1: Input Configurations - NWP Data

ix	f_ws
x	f_ws, f_gust
xi	f_ws, f_gust, f_wd
xii	f_ws, f_gust, f_temp
xiii	f_ws, f_gust, f_press
xiv	f_ws, f_gust, f_wd, f_temp
xv	f_ws, f_gust, f_wd, f_press
xvi	f_ws, f_gust, f_temp, f_press
xvii	f_ws, f_gust, f_wd, f_temp, f_press
xviii	f_ws, f_gust, f_wd, f_temp f_press, f_hum

7.1 Turbine at Wind Park I

Results for the turbine located at wind park I are displayed in table 7.2. A significant improvement in accuracy is observed with the inclusion of NWP data. In one hour ahead prediction, the best performing model decreases RMSE by 9.26% when compared to using present wind speed and wind power only, and 13% compared to the persistence model. As expected, the significance of NWP data gets even more evident as forecast horizon expands. Showing a decrease in RMSE from input configuration II by 16.86% (two hours ahead), 23.09%(three hours ahead) and 27.50%(four hours ahead).

When looking at the effect of distinct candidate features, we observe that including gust improves accuracy in all horizons. Moreover, including wind direction (II + xi) and temperature (II + xii) further improves accuracy in all horizons. However, when using wind direction and temperature together (II + xiv), a slight increase in error is recorded for horizon one and three. Adding air pressure to the input has a negative impact on accuracy in all horizons, suggesting it is not suitable as a feature in this case. Finally we can observe that adding humidity improves accuracy (II + xviii) in all forecast horizons.

Table 7.2: RMSE - Turbine at Wind Park II - Including NWP (KW)

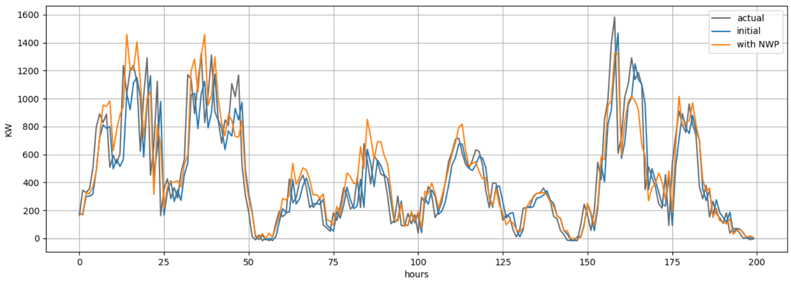
Input	1h ahead	2h ahead	3h ahead	4h ahead
II + ix	273.1	309.5	334	351.3
II + x	273.3	311.8	330.3	346.8
II + xi	273.2	309.7	327.8	341.7
II + xii	**272.4**	308.2	326.4	340
II + xiii	275	312.1	334.1	347.6
II + xiv	277	**303.7**	329.2	339.8
II + xv	276	309	328.5	341.6
II + xvi	274.6	309.7	327.4	340.7
II + xvii	280.3	307.1	328.7	341.6
II + xviii	277.1	304.5	**325.4**	**334.8**
II	300.2	365.3	423.1	461.8
naive	313.1	386.9	449.4	499.8

7.1.1 Including Lags and Neighbor Data

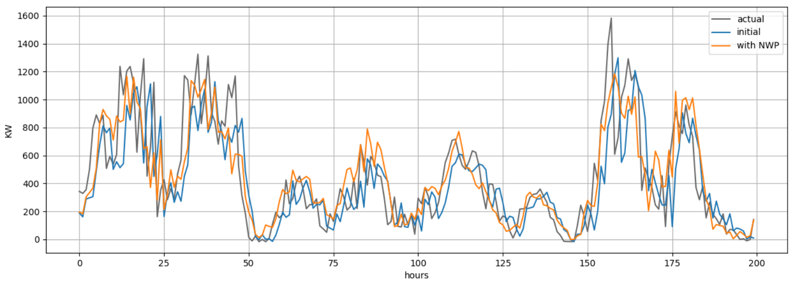
Experiments have been done using combinations of the best NWP, lags and neighbor configurations, to see if accuracy is further improved. From table 7.3 we see that this is not the case. The model utilizing only NWP data, in addition to present wind speed and wind power, provides the best results in all forecast horizons. This suggests that lag values and neighbor turbines should only be considered when not using NWP data.

Table 7.3: RMSE - Turbine at Wind Park I - All Input Types (KW)

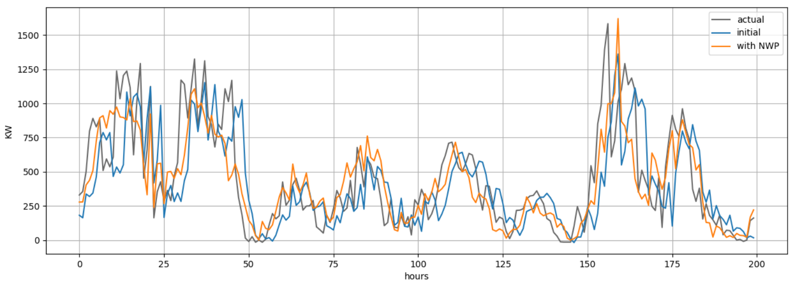
Input	1h ahead	2h ahead	3h ahead	4h ahead
Initial (wp_0 & ws_0)	300.2	365.3	423.1	461.8
with NWP	**272.4**	**303.7**	**325.4**	**334.8**
with NWP and lags	277.8	310.6	329.2	336.1
with NWP and neighbors	281.4	322.4	338.3	345.2
with all	285	320	340.3	346.8
naive	313.1	386.9	449.4	499.8



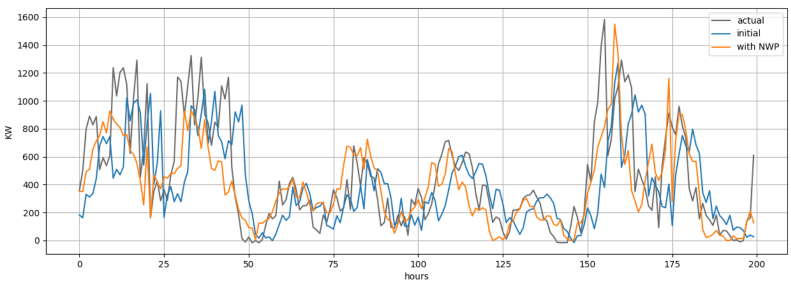
1h ahead



2h ahead



3h ahead



4h ahead

Figure 7.2: Sample data from prediction results using NWP data for wind turbine at park I

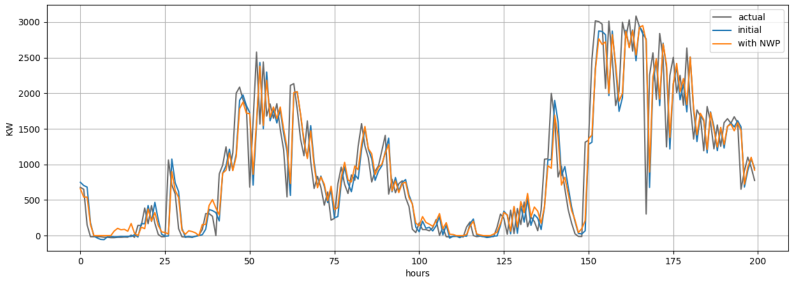
7.2 Turbine at Wind Park II

Similar to the turbine at wind park I, including NWP data shows noteworthy improvement in accuracy for the turbine at wind park II. For one hour ahead prediction the best NWP configuration (II + xvii) decreases RMSE of 7.51% from only using wind speed and wind power as input. For two hours ahead a decrease of 13.37% is registered, followed by 16.84% and 22.13% in for three and four hours ahead.

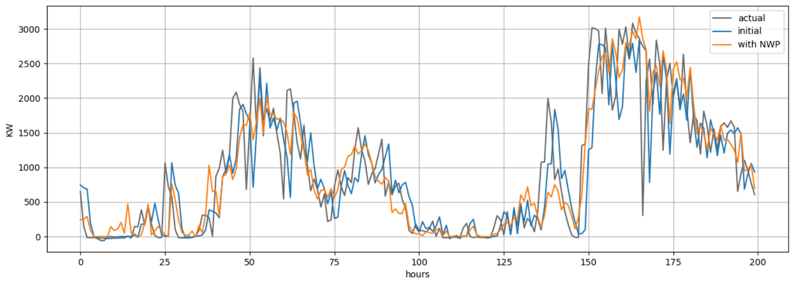
Including gust (II + x) further improves accuracy for all horizons also for the second turbine, establishing itself as a suitable feature for short-term wind power prediction. For the remaining candidate features, results do not exclusively point in one direction. When wind direction forecasts are first introduced (II + xi) accuracy is boosted for two- and four hour ahead predictions, but worsen for the one- and three hour ahead. Including forecast temperature (II + xii) has a positive impact on two, three and four hours ahead predictions, while pressure improves accuracy for two and three hours ahead (II + xiii). Finally, we see that including humidity (II + xviii) decreases error in two forecast horizons. However, combining the separate features seems have a positive effect on performance. In two out of four forecast horizons, all candidate features were included in the input. For the two remaining forecast horizons, humidity was left out of the feature space. From these results all features show signs of having some predictive value, although only gust was shown to exclusively give a positive impact.

Table 7.4: RMSE - Turbine at Wind Park II - Including NWP (MW) (KW)

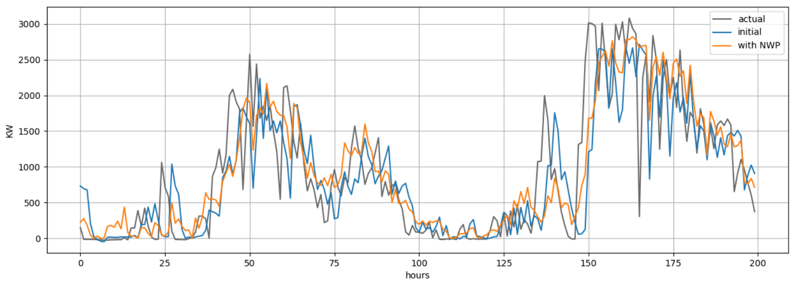
Input	1h ahead	2h ahead	3h ahead	4h ahead
II + ix	397.5	465	499	512.3
II + x	394.6	462.8	495.9	510.4
II + xi	395	460.9	497.2	503.8
II + xii	395.9	459	495.6	503
II + xiii	396.5	460.9	495.4	510.5
II + xiv	396.3	459	495.5	500.5
II + xv	394.1	459.9	497.1	504.9
II + xvi	403.1	458.4	496.5	505.2
II + xvii	**393.6**	455.8	**495.2**	500.5
II + xviii	394.1	**455.5**	495.7	**499.5**
II	425.6	525.8	595.5	641.5
naive	437.8	549.5	633.3	690



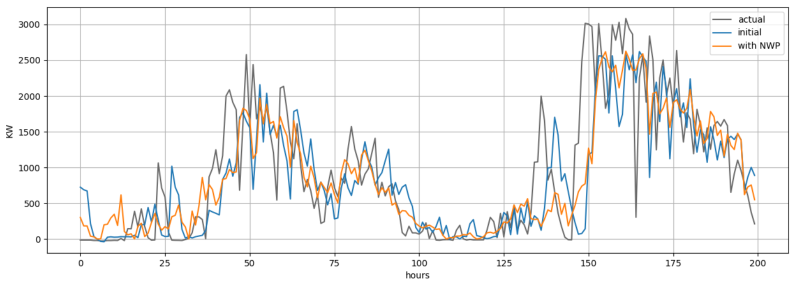
1h ahead



2h ahead



3h ahead



4h ahead

Figure 7.3: Sample data from prediction results using NWP data for wind turbine at park II

7.2.1 Including Lags and Neighbor Data

As also shown for the turbine at wind park I, combining NWP data with lags and neighbor data does not seems to be an effective strategy. From table 7.5 we can read that for all forecast horizons, using only NWP data, in addition to present wind power and wind speed, gave the best results.

Table 7.5: RMSE - Turbine at Wind Park II - All Input Types (KW)

Input	1h ahead	2h ahead	3h ahead	4h ahead
Initial (wp_0 & ws_0)	425.6	525.8	595.5	641.5
with NWP	**393.6**	**455.5**	**495.2**	**499.5**
with NWP and lags	394.1	459.5	513.1	521.4
with NWP and neighbors	397.7	466.1	510.1	523.5
with all	397.1	464.6	507.3	521.4
naive	437.8	549.5	633.3	690

7.3 Park Prediction with NWP Data

Finally, the impact of using NWP data when scaling prediction up to wind park level is examined. Both strategy 1 (predict each turbine individually) and strategy 2 (predict park from aggregate of turbines) from section 6.1 and 6.2 are followed and compared with respect to RMSE. For strategy 1, the best performing input configurations from section 7.1 and 7.2 are chosen for each turbine. In order to give the two strategies an even playing field, the same input patterns are examined for the strategy 2 approach as for the two turbines. These configurations are listed in table 7.1.

7.3.1 Wind Park I

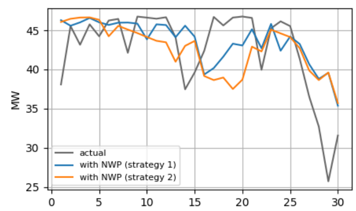
The impact of NWP data is also evident in power predictions for wind park, showing a decrease in RMSE of 9.28%(1h ahead), 17.14%(2h ahead), 23.26%(3h ahead) and 27.57%(4h ahead) comparing the best model from table 7.6 against input configuration II. We can see that strategy 1 gave the best result in three out of four cases, suggesting that this is the best approach, although the level of improvement gained by deploying strategy 1 over the simpler strategy 2 is minimal. For two and three hour ahead predictions a decrease in RMSE of 0.26% and 0.12% is obtained compared to the best results using strategy 2. The impact of strategy 1 is slightly more evident for four hour ahead prediction, showing a decrease in RMSE of 1.6%. For one hour ahead however, we can observe a slight increase in RMSE of 0.09%. Considering the small differences between the two approaches, it is difficult to see the advantage of following strategy 1 in this case, taking into account the computation cost following it. One could think the two approaches gave almost exact predictions, but by looking at figure 7.4 we observe that this is not the case.

From analyzing the effect of the specific candidate features when utilizing strategy 2, the most interesting fact is that including gust increases error in three of the four forecast horizons (II + x). This was not the case for single turbine prediction.

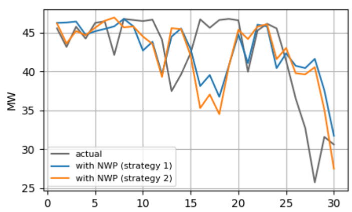
Table 7.6: RMSE - Wind park I - Strategy 2 with NWP (MW)

Input	**1h ahead**	**2h ahead**	**3h ahead**	**4h ahead**
Strat. 1 (best)	4.355	**5.24**	**5.746**	**6.058**
II + ix	4.376	5.298	5.809	6.184
II + x	4.395	5.328	5.805	6.274
II + xi	4.367	5.311	5.947	6.245
II + xii	4.364	5.246	5.753	6.148
II + xiii	4.384	5.337	5.986	6.188
II + xiv	4.393	5.254	5.879	6.157
II + xv	**4.351**	5.307	5.973	6.225
II + xvi	4.357	5.268	5.963	6.223
II + xvii	4.383	5.289	6.006	6.238
II + xviii	4.386	5.323	6.156	6.372
II	4.796	6.324	7.488	8.365
naive	4.878	6.528	7.844	8.898

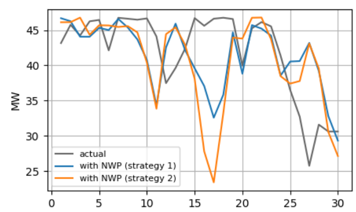
1h ahead



2h ahead



3h ahead



4h ahead

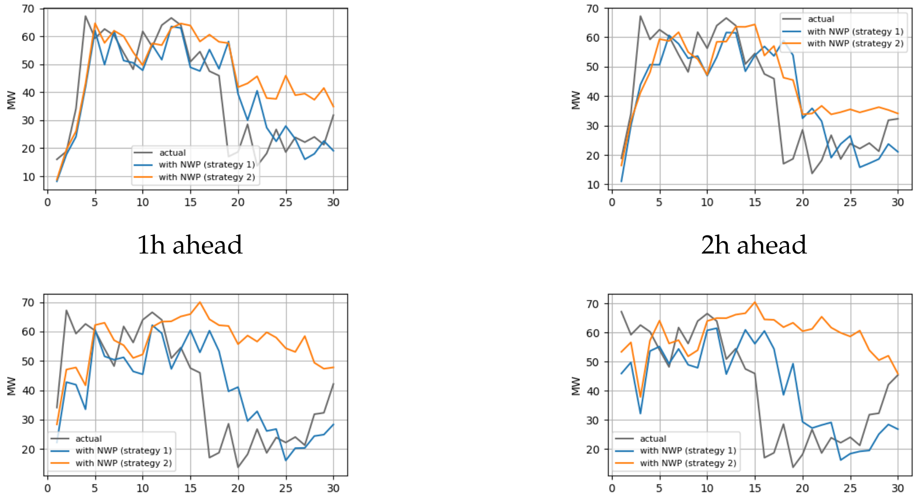
Figure 7.4: Prediction samples with NWP data for wind park I

7.3.2 Wind Park II

The prediction results for wind park II are displayed in table 7.7. RMSE are decreased by 8.02% (1h ahead) 13.21% (2h ahead) 18.45% (3h ahead) 23.10% (4h ahead) with the inclusion of NWP data compared to only using present data of wind speed and wind power measurement. Further, we can observe that the advantage of using strategy 1 over strategy 2 for wind park power prediction is much more evident in this case, showing a decrease in RMSE of 4.31% (1h ahead), 2.65% (2h ahead), 3.60% (3h ahead) and 2.90% (4h ahead) when compared to the best performing model applying strategy 2. Samples from the prediction results are displayed in figure 7.5. Regarding the impact of the specific features for strategy II, we can observe, as in the case of wind park I, that gust has a negative impact on prediction accuracy for all four forecast horizons(II + x). The best performing model for one and two hour ahead predictions using strategy 2 only included forecast wind speed from the NWP data. However as forecast horizon is expanded, including additional NWP data to the input showed to provide a better accuracy. Similar behavior was registered in section 7.1 for predicting the power output for the turbine at wind park I.

Table 7.7: RMSE - Wind park II - Strategy 2 with NWP (MW)

Input	**1h ahead**	**2h ahead**	**3h ahead**	**4h ahead**
Strat. 1 (best)	**6.36**	**7.893**	**8.786**	**9.283**
II + ix	6.647	8.108	9.383	9.926
II + x	6.766	8.294	9.448	10.031
II + xi	6.65	8.29	9.32	9.845
II + xii	6.66	8.349	9.344	9.88
II + xiii	6.762	8.379	9.538	10.063
II + xiv	6.636	8.366	9.115	9.575
II + xv	6.881	8.357	9.281	9.724
II + xvi	6.672	8.402	9.475	9.996
II + xvii	6.752	8.557	9.168	9.593
II + xviii	6.66	8.8	9.142	9.561
II	6.915	9.095	10.775	12.072
naive	7.019	9.388	11.304	12.826



1h ahead | 2h ahead

3h ahead | 4h ahead

Figure 7.5: Prediction samples with NWP data for wind park II

7.4 Summary and Conclusion

In this chapter we sought out to address RQ4 from section 1.2.1 by utilizing data from a NWP model in the prediction process. Results showed a significant drop in error for both turbines. Not surprisingly, the impact on accuracy increases as we predict further into the future. More interesting is the fact that it also provided noteworthy improvement even for one hour ahead predictions. Of all the feature types explored in this thesis, NWP data has the most evident effect on performance. We conclude from this that NWP data should definitely be included into the feature space, even for one hour ahead predictions. The second part of RQ4 involved exposing the specific attributes that should be included from the NWP data. Apart for the obvious choice of wind speed, the gust attribute had a positive impact in all cases for turbine power prediction. For the remaining candidate features it is tougher to draw any clear conclusions, although for the turbine at wind park I, all candidates were used as input for the best performing model in three and four hour ahead predictions, suggesting that wind direction, humidity, pressure and temperature are first useful when considering predictions further into the future. Another interesting observation was the fact that combining NWP data with lag values and neighbor data decreased performance in all cases, suggesting that these feature types should only be considered when NWP data is not available.

In this chapter we also revisited the task of predicting wind power for the whole wind park. Similar to when only measured data was considered, utilizing strategy 1 showed overall to be the best approach when applying NWP data. For wind park I the difference in RMSE between the two approaches was minimal, but for the second wind park, strategy 1 showed a noteworthy improvement over strategy two. Why this is the case is not clear, but it can be related to the fact that wind park II spans a larger geographical area. Calculating the average forecast weather data in this case for use in strategy 2 involved data from six different grid points from the NWP model, as opposed to only three points for wind park I. The loss of local precise forecast weather data is thereby more evident in the case of wind park II when deploying strategy 2.

Finally, we could observe that the features having a positive impact on turbine power prognosis do not necessarily do so when predicting wind park power using strategy 2. Overall, there is not much to gain in performance by including features beyond forecast wind speed. More specifically, gust was shown to have a negative effect in all cases, which contradicts with its influence on power prediction for the two turbines.

Part IV

Summary

Chapter 8

Summary and Future Work

8.1 On the Use of Measured Wind Park Data

In this work we studied the impact of measured data for predicting two turbines located in two different wind parks.

In addressing RQ1 the impact of wind power, wind speed, yaw drive, temperature and wind direction was examined for predicting two turbines located in different wind parks. We found that wind speed had a positive effect on prediction accuracy for turbines, and therefore should be considered as an input when doing short-term wind power prediction. Including the remaining candidate features showed to have a negative effect prediction accuracy for turbine located at wind park I and a slight improvement for the turbine at wind park II. However, it is worth pointing out that the increase in error experienced for the first turbine was more significant than the decrease in error gained for the second turbine. These findings nevertheless suggest that neither present measure of wind direction, temperature and yaw drive has a valuable impact on short-term wind power prediction.

In examining the use of lag values, referring to RQ2, we observed little or no impact on prediction accuracy. In a couple of the forecast horizons for the turbine at wind park I, using lag had a slight improvement on performance. For the second turbine lag values improved accuracy in all forecast horizons, although not by a significant factor. An inconsistency was observed in trying to determine how wide a suitable lag window should be. For the first turbine a window width of one gave the best result, while for the second turbine a wider lag window seemed to be preferred. Giving a general suggestion on the use of lag values, especially considering window width, is hard from these findings. However, taking into account that in most cases lag values did have a positive impact, although arguably small, they should not be excluded by default.

We also examined including data from neighbor turbines in predicting the wind power output of a single turbine. Experiments showed that this strategy had a positive effect on prediction accuracy, giving a noteworthy decrease in error, especially when considering data from ten or more neighbors. However we later found that this approach did not scale well to wind park prediction, showing an increase in error for both wind parks. This suggest that this approach should only be used when predicting a single turbine.

8.2 On the Use of NWP data

RQ4 regarded the impact that data from numerical weather prediction models have on short-term wind power prediction using a machine learning approach. Experiments showed that the inclusion of NWP data to the input had a significant effect on prediction accuracy. This fact was not surprising with concern to three and four hour ahead predictions, but results also showed a noteworthy impact on one and two hour ahead predictions. Based on these findings, it is not difficult to suggest the use of NWP data for this purpose. Which of the attributes to extract, however, is less clear. Apart from the obvious choice of forecast wind speed, wind speed of gust improved accuracy in all cases for turbine power prediction. However, when deploying strategy 2 for wind park power prediction, wind speed of gust had a negative impact for all tests. The remaining candidate features examined did all show signs of usefulness, but results showed that they tended to become more valuable when considering predictions further into the future.

8.3 On Strategies for Wind Park Prediction

In addressing RQ5 from section 1.2.1 we in this work compared two approaches for predicting the wind power output of a wind park. In the first strategy, we predicted the power output for each turbine individually and then summed up the predictions. This strategy is much more time consuming, since we have to develop a prediction model for each turbine. In the second strategy, we aggregated the data from each turbine before prediction, meaning that we only developed a single prediction model for the whole wind park. Experiments showed that the first strategy overall performs better than the second, when using inputs from measured data. The level of improvement, however, was shown to not be of a significant character.

Both strategies were significantly improved by including NWP data, and overall the first strategy showed better performance also in this case. However for one hour ahead prediction at wind park I, a slight improvement in accuracy was gained by deploying the second strategy. For wind park II however, the first strategy showed significantly better prediction results in all forecast horizons.

From these findings, predicting the wind power of each turbine individually and summing them up seems the most promising and should be considered, especially if the number of turbines in the wind park are reasonable low. However, as mentioned earlier in this thesis, if the number of turbines reaches hundreds, this approach might become too costly.

8.4 Future Work

In this section, some of the questions sparked from working on this thesis are presented and suggested for further research.

Improving Prediction by Focusing More on Method Selection

The main focus of this thesis was to look at the impact of input to prediction accuracy to a generic machine learning method. As a consequence, the prediction framework developed in this thesis can be regarded as rather simple. It is expected that more emphasis on the prediction unit and not only on the choice of input could further improve accuracy. It would therefore be of interest to research more on this part of the prediction process. Findings from other studies suggest that combining several methods into a hybrid model is a promising strategy. Also, in recent years deep-learning methods, particularly recurrent neural networks with long short-term memory units(LSTM), has risen in popularity. Expanding the knowledge on how well these types of methods perform at doing short-term wind power predictions would be of interest.

Correction of NWP data

Besides a couple of transformation steps explained in section 4.2.1 no pre-processing has been done on the NWP data utilized in this work. It is expected that more emphasis on down-scaling and correction of NWP data would further improve prediction accuracy, based on the assumption that most NWP models contain a certain degree of systematic and stochastic biases [15]. For correction purposes both gaussian process (GP) and kalman filter have been utilized with success in earlier work [15][11][43]. In [15] a GP was used to correct forecast wind speed with wind speed, wind direction, temperature, humidity and pressure as input before training the main prediction component. In [11] a kalman filter was used to correct 10 meter wind speed forecast with success. Exploring the impact correction methods for NWP data, such as the two mentioned, can have on the final prediction accuracy, would be a logical extension of this work.

Examining the impact of NWP data on very-short-term predictions

We have seen that much work in short-term predictions do not include NWP, but when considering even shorter forecast horizons NWP data is almost exclusively not used. Since the work on this thesis started, the transmission system operators in the Nordic countries have confirmed that a move from 60 to 15 minutes markets and settlement will be done by late 2020 [66]. This shift will demand reliable wind power predictions on even shorter time frames than examined in this thesis. It would therefore be of interest to see how NWP data impacts prediction accuracy for a statistical approach in very-short wind power prediction.

Examining Additional Strategies to Wind Park Prediction

This thesis was limited to examining two approaches for wind park prediction. At the time of experiments we were not aware of the strategy deployed in [61], in which case they trained the prediction model using samples from a single turbine at the wind park and the aggregate wind power for the whole park as target data. The specific turbine is chosen by finding the most correlating, among all the turbines, to the aggregate wind power. The most correlating turbine was then used to predict the power output for the whole wind park by using input samples from turbine measurements, and aggregated wind power data as targets. We were not aware of this strategy at the time experiments were conducted. In future research it would be interesting to compare this approach against the other two strategies examined.

Bibliography

[1] Almoataz Abdelaziz et al. 'Short Term Wind Power Forecasting Using Autoregressive Integrated Moving Average Modeling'. In: Dec. 2012.

[2] Rishabh Abhinav et al. 'Short-term wind power forecasting using wavelet-based neural network'. In: *Energy Procedia* 142 (2017). Proceedings of the 9th International Conference on Applied Energy, pp. 455–460. ISSN: 1876-6102. DOI: https://doi.org/10.1016/j.egypro.2017.12.071. URL: http://www.sciencedirect.com/science/article/pii/S1876610217357995.

[3] David Alencar et al. 'Different Models for Forecasting Wind Power Generation: Case Study'. In: *Energies* 10 (Nov. 2017), p. 1976. DOI: 10.3390/en10121976.

[4] J. Arshad, A. Zameer and A. Khan. 'Wind Power Prediction Using Genetic Programming Based Ensemble of Artificial Neural Networks (GPeANN)'. In: *2014 12th International Conference on Frontiers of Information Technology*. Dec. 2014, pp. 257–262. DOI: 10.1109/FIT.2014.55.

[5] Sudharsan Asaithambi. *Why, How and When to apply Feature Selection*. https://towardsdatascience.com/why-how-and-when-to-apply-feature-selection-e9c69adfabf2. Accessed: 2019-02-28.

[6] Anup Bhande. *What is underfitting and overfitting in machine learning and how to deal with it.What is underfitting and overfitting in machine learning and how to deal with it.* https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76. Accessed: 2018-10-02. 2018.

[7] K. Bhaskar and S. N. Singh. 'AWNN-Assisted Wind Power Forecasting Using Feed-Forward Neural Network'. In: *IEEE Transactions on Sustainable Energy* 3.2 (Apr. 2012), pp. 306–315. ISSN: 1949-3029. DOI: 10.1109/TSTE.2011.2182215.

[8] Danilo Bzdok, Naomi Altman and Martin Krzywinski. 'Statistics versus machine learning'. In: *Nature Methods* 15 (2018), pp. 233–234.

[9] Erasmo Cadenas et al. 'Wind speed prediction using a univariate ARIMA model and a Multivariate NARX model'. In: *Energies* 9 (Feb. 2016), p. 109. DOI: 10.3390/en9020109.

[10] Y. Cao et al. 'Wind power ultra-short-term forecasting method combined with pattern-matching and ARMA-model'. In: *2013 IEEE Grenoble Conference*. June 2013, pp. 1–4. DOI: 10.1109/PTC.2013.6652257.

[11] Federico Cassola and Massimiliano Burlando. 'Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output'. In: *Applied Energy* 99 (2012), pp. 154–166. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2012.03.054. URL: http://www.sciencedirect.com/science/article/pii/S0306261912002747.

[12] J. P. S. Catalao, H. M. I. Pousinho and V. M. F. Mendes. 'An Artificial Neural Network Approach for Short-Term Wind Power Forecasting in Portugal'. In: *2009 15th International Conference on Intelligent System Applications to Power Systems*. Nov. 2009, pp. 1–5. DOI: 10.1109/ISAP.2009.5352853.

[13] J. P. S. Catalao, H. M. I. Pousinho and V. M. F. Mendes. 'Hybrid Wavelet-PSO-ANFIS Approach for Short-Term Wind Power Forecasting in Portugal'. In: *IEEE Transactions on Sustainable Energy* 2.1 (Jan. 2011), pp. 50–59. ISSN: 1949-3029. DOI: 10.1109/TSTE.2010.2076359.

[14] G. W. Chang et al. 'A hybrid model for forecasting wind speed and wind power generation'. In: *2016 IEEE Power and Energy Society General Meeting (PESGM)*. July 2016, pp. 1–5. DOI: 10.1109/PESGM.2016.7742039.

[15] N. Chen et al. 'Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction'. In: *IEEE Transactions on Power Systems* 29.2 (Mar. 2014), pp. 656–665. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2013.2282366.

[16] Hamed Chitsaz, Nima Amjady and Hamidreza Zareipour. 'Wind power forecast using wavelet neural network trained by improved Clonal selection algorithm'. In: *Energy Conversion and Management* 89 (2015), pp. 588–598. ISSN: 0196-8904. DOI: https://doi.org/10.1016/j.enconman.2014.10.001. URL: http://www.sciencedirect.com/science/article/pii/S0196890414008814.

[17] Corinna Cortes and Vladimir Vapnik. 'Support-vector networks'. In: *Machine Learning* 20.3 (Sept. 1995), pp. 273–297. ISSN: 1573-0565. DOI: 10.1007/BF00994018. URL: https://doi.org/10.1007/BF00994018.

[18] Cyc. *Graphic showing the maximum separating hyperplane and the margin*. https://upload.wikimedia.org/wikipedia/commons/2/2a/Svm_max_sep_hyperplane_with_margin.png. Accessed: 2018-10-02. 2008.

[19] Maria Grazia De Giorgi et al. 'Comparison Between Wind Power Prediction Models Based on Wavelet Decomposition with Least-Squares Support Vector Machine (LS-SVM) and Artificial Neural Network (ANN)'. In: *Energies* 7.8 (2014), pp. 5251–5272. ISSN: 1996-

1073. DOI: 10.3390/en7085251. URL: http://www.mdpi.com/1996-1073/7/8/5251.

[20] A. T. Eseye et al. 'A double-stage hierarchical ANFIS model for short-term wind power prediction'. In: *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)(*. Mar. 2017, pp. 546–551. DOI: 10.1109/ICBDA.2017.8078694.

[21] A. T. Eseye et al. 'A double-stage hierarchical ANFIS model for short-term wind power prediction'. In: *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)(*. Mar. 2017, pp. 546–551. DOI: 10.1109/ICBDA.2017.8078694.

[22] R. S. F. Ferraz et al. 'Multi step forecasting of the wind generation by ARIMA and k-NN'. In: *2018 Simposio Brasileiro de Sistemas Eletricos (SBSE)*. May 2018, pp. 1–6. DOI: 10.1109/SBSE.2018.8395614.

[23] Aurélie Fischer et al. 'Statistical learning for wind power: A modeling and stability study towards forecasting'. In: *Wind Energy* 20 (Sept. 2017). DOI: 10.1002/we.2139.

[24] Scott Fortmann-Roe. *Understanding the Bias-Variance Tradeoff.* http://scott.fortmann-roe.com/docs/BiasVariance.html. Accessed: 2019-03-16. 2012.

[25] Y. Fu et al. 'Multi-step Ahead Wind Power Forecasting Based on Recurrent Neural Networks'. In: *2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*. Oct. 2018, pp. 217–222. DOI: 10.1109/APPEEC.2018.8566471.

[26] Trevor Hastie Gareth James Daniela Witten and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R*. Springer, 2013.

[27] Gregor Giebel et al. 'The State of the Art in Short-Term Prediction of Wind Power A Literature Overview, 2nd Edition'. In: (Jan. 2011). DOI: 10.13140/RG.2.1.2581.4485.

[28] Gregor Giebel et al. 'Wind Power Prediction using Ensembles'. In: 2005.

[29] Pedro Gomes and Rui Castro. 'Wind Speed and Wind Power Forecasting using Statistical Models: AutoRegressive Moving Average (ARMA) and Artificial Neural Networks (ANN)'. In: *International Journal of Sustainable Energy Development* 1 (Dec. 2012). DOI: 10.20533/ijsed.2046.3707.2012.0007.

[30] Isabelle Guyon and André Elisseeff. 'An Introduction to Variable and Feature Selection'. In: *J. Mach. Learn. Res.* 3 (Mar. 2003), pp. 1157–1182. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=944919.944968.

[31] S. Han, Y. Liu and J. Yan. 'Neural Network Ensemble Method Study for Wind Power Prediction'. In: *2011 Asia-Pacific Power and Energy Engineering Conference*. Mar. 2011, pp. 1–4. DOI: 10.1109/APPEEC.2011.5748787.

[32] A. U. Haque, M. H. Nehrir and P. Mandal. 'A Hybrid Intelligent Model for Deterministic and Quantile Regression Approach for Probabilistic Wind Power Forecasting'. In: *IEEE Transactions on Power Systems* 29.4 (July 2014), pp. 1663–1672. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2014.2299801.

[33] Y. Hong and C. Wu. 'Hour-ahead wind power and speed forecasting using market basket analysis and radial basis function network'. In: *2010 International Conference on Power System Technology*. Oct. 2010, pp. 1–6. DOI: 10.1109/POWERCON.2010.5666634.

[34] Q. Hu et al. 'Short-Term Wind Speed or Power Forecasting With Heteroscedastic Support Vector Regression'. In: *IEEE Transactions on Sustainable Energy* 7.1 (Jan. 2016), pp. 241–249. ISSN: 1949-3029. DOI: 10.1109/TSTE.2015.2480245.

[35] S. Hu and K. Li. 'Short-Term Wind Power Prediction Based on Principal Component Analysis and Elman Artificial Neural Networks'. In: *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*. July 2017, pp. 673–677. DOI: 10.1109/ICISCE.2017.146.

[36] J. D. Hunter. 'Matplotlib: A 2D Graphics Environment'. In: *Computing in Science Engineering* 9.3 (May 2007), pp. 90–95. ISSN: 1521-9615. DOI: 10.1109/MCSE.2007.55.

[37] 'Is mutual information adequate for feature selection in regression?' In: *Neural Networks* 48 (2013), pp. 1–7. ISSN: 0893-6080. DOI: https://doi.org/10.1016/j.neunet.2013.07.003. URL: http://www.sciencedirect.com/science/article/pii/S0893608013001883.

[38] Jaesung Jung and Robert P. Broadwater. 'Current status and future advances for wind speed and power forecasting'. In: *Renewable and Sustainable Energy Reviews* 31 (2014), pp. 762–777. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2013.12.054. URL: http://www.sciencedirect.com/science/article/pii/S1364032114000094.

[39] Wu Junli, Liu Xingjie and Qian Jian. 'Wind speed and power forecasting based on RBF neural network'. In: *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*. Vol. 5. Oct. 2010, pp. 298–301. DOI: 10.1109/ICCASM.2010.5619079.

[40] M. Nandana Jyothi et al. 'NARX Based Short Term Wind Power Forecasting Model'. In: 2015.

[41] Y. Kassa et al. 'Short term wind power prediction using ANFIS'. In: *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*. Oct. 2016, pp. 388–393. DOI: 10.1109/ICPRE.2016.7871238.

[42] Y. Liu et al. 'Short-Term Wind-Power Prediction Based on Wavelet Transform–Support Vector Machine and Statistic-Characteristics Analysis'. In: *IEEE Transactions on Industry Applications* 48.4 (July 2012), pp. 1136–1141. ISSN: 0093-9994. DOI: 10.1109/TIA.2012.2199449.

[43] P. Louka et al. 'Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering'. In: *Journal of Wind Engineering and Industrial Aerodynamics* 96.12 (2008), pp. 2348–2362. ISSN: 0167-6105. DOI: https://doi.org/10.1016/j.jweia.2008.03.013. URL: http://www.sciencedirect.com/science/article/pii/S0167610508001074.

[44] Spyros Makridakis, Evangelos Spiliotis and Vassilios Assimakopoulos. 'Statistical and Machine Learning forecasting methods: Concerns and ways forward'. In: *PLOS ONE* 13.3 (Mar. 2018), pp. 1–26. DOI: 10.1371/journal.pone.0194889. URL: https://doi.org/10.1371/journal.pone.0194889.

[45] Y. Mao and W. Shaoshuai. 'A review of wind power forecasting amp; prediction'. In: *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. Oct. 2016, pp. 1–7. DOI: 10.1109/PMAPS.2016.7764085.

[46] T. Mathaba, X. Xia and J. Zhang. 'Short-term wind power prediction using Least-Square Support Vector Machines'. In: *IEEE Power and Energy Society Conference and Exposition in Africa: Intelligent Grid Integration of Renewable Energy Resources (PowerAfrica)*. July 2012, pp. 1–6. DOI: 10.1109/PowerAfrica.2012.6498620.

[47] W. T. McKinney. 'Data Structures for Statistical Computing in Python'. In: 2010.

[48] Li Meng et al. 'SVM with improved grid search and its application to wind power prediction'. In: *2013 International Conference on Machine Learning and Cybernetics*. Vol. 02. July 2013, pp. 603–609. DOI: 10.1109/ICMLC.2013.6890363.

[49] Emma Foehringer Merchant. *IPCC: Renewables to Supply 70%-85% of Electricity by 2050 to Avoid Worst Impacts of Climate Change*. https://www.greentechmedia.com/articles/read/ipcc-renewables-85-electricity-worst-impacts-climate-change#gs.97sttq. Accessed: 2019-30-04.

[50] Michael Milligan, M N. Schwartz and Yih-Huei Wan. 'Statistical Wind Power Forecasting for U.S. Wind Farms: Preprint'. In: Nov. 2003.

[51] Malte Müller et al. 'AROME-MetCoOp: A Nordic Convective-Scale Operational Weather Prediction Model'. In: *Weather and Forecasting* 32.2 (2017), pp. 609–627. DOI: 10.1175/WAF-D-16-0099.1. eprint: https://doi.org/10.1175/WAF-D-16-0099.1. URL: https://doi.org/10.1175/WAF-D-16-0099.1.

[52] *New forecast model cycle brings highest-ever resolution*. https://www.ecmwf.int/en/about/media-centre/news/2016/new-forecast-model-cycle-brings-highest-ever-resolution. Accessed: 2019-03-30.

[53] T.E. Oliphant. *A Guide to NumPy*. Trelgol Publishing, 2006. URL: https://books.google.no/books?id=fKulSgAACAAJ.

[54] Fabian Pedregosa et al. 'Scikit-learn: Machine Learning in Python'. In: *J. Mach. Learn. Res.* 12 (Nov. 2011), pp. 2825–2830. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=1953048.2078195.

[55] C. W. Potter and M. Negnevitsky. 'Very short-term wind forecasting for Tasmanian power generation'. In: *IEEE Transactions on Power Systems* 21.2 (May 2006), pp. 965–972. ISSN: 0885-8950. DOI: 10.1109/TPWRS.2006.873421.

[56] H. M. I. Pousinho, J. P. S. Catalão and V. M. F. Mendes. 'Wind power short-term prediction by a hybrid PSO-ANFIS approach'. In: *Melecon 2010 - 2010 15th IEEE Mediterranean Electrotechnical Conference*. Apr. 2010, pp. 955–960. DOI: 10.1109/MELCON.2010.5475923.

[57] *Previento - The Reliable Wind Power Prediction*. https://www.energymeteo.com/products/power_predictions/wind_power_prediction.php. Accessed: 2019-03-30.

[58] Saurabh S. Soman et al. 'A review of wind power and wind speed forecasting methods with different time horizons'. In: Oct. 2010, pp. 1–8. DOI: 10.1109/NAPS.2010.5619586.

[59] A. L. Samuel. 'Some Studies in Machine Learning Using the Game of Checkers'. In: *IBM Journal of Research and Development* 3.3 (July 1959), pp. 210–229. ISSN: 0018-8646. DOI: 10.1147/rd.33.0210.

[60] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.

[61] Jie Shi et al. 'Genetic algorithm-piecewise support vector machine model for short term wind power prediction'. In: *2010 8th World Congress on Intelligent Control and Automation*. July 2010, pp. 2254–2258. DOI: 10.1109/WCICA.2010.5554305.

[62] 'Short-term prediction of wind power based on deep Long Short-Term Memory'. In: *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*. Oct. 2016, pp. 1148–1152.

[63] Vijendra Singh. 'Application of Artificial Neural Networks for Predicting Generated Wind Power'. In: *International Journal of Advanced Computer Science and Applications* 7 (2016). DOI: 10.14569/IJACSA.2016.070336.

[64] American Meteorological Society. *meridional wind*. http://glossary.ametsoc.org/wiki/Meridional_wind. Accessed: 2018-4-29.

[65] American Meteorological Society. *zonal wind*. http://glossary.ametsoc.org/wiki/Zonal_wind. Accessed: 2018-4-29.

[66] Statnett. *Nordic TSOs confirm 15 minutes resolution by end of 2020*. https://www.statnett.no/en/about-statnett/news-and-press-releases/News-archive-2018/nordic-tsos-confirm-15-minutes-resolution-by-end-of-2020/. Accessed: 2019-04-25.

[67] Josef Steppan. *A few samples from the MNIST test dataset*. https://upload.wikimedia.org/wikipedia/commons/2/27/MnistExamples.png. Accessed: 2018-10-01. 2017.

[68] U Thissen et al. 'Using support vector machines for time series prediction'. In: *Chemometrics and Intelligent Laboratory Systems* 69.1 (2003), pp. 35–49. ISSN: 0169-7439. DOI: https://doi.org/10.1016/S0169-7439(03)00111-4. URL: http://www.sciencedirect.com/science/article/pii/S0169743903001114.

[69] Joy A. Thomas Thomas M. Cover. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991, pp. 18–23.

[70] S. Tian et al. 'Wind Power Forecasting Based on ARIMA-LGARCH Model'. In: *2018 International Conference on Power System Technology (POWERCON)*. Nov. 2018, pp. 1285–1289.

[71] N. A. Treiber and O. Kramer. 'Evolutionary feature weighting for wind power prediction with nearest neighbor regression'. In: *2015 IEEE Congress on Evolutionary Computation (CEC)*. May 2015, pp. 332–337. DOI: 10.1109/CEC.2015.7256910.

[72] Nils André Treiber and Oliver Kramer. 'Wind Power Prediction with Cross-Correlation Weighted Nearest Neighbors'. In: *EnviroInfo*. 2014.

[73] Nils Treiber, Justin Heinermann and Oliver Kramer. 'Aggregation of Features for Wind Energy Prediction with Support Vector Regression and Nearest Neighbors'. In: Sept. 2013.

[74] Robert Tibshirani Trevor Hastie and J. H. Friedman. *The Elements of Statistical Learning : data mining, inference, and prediction*. 2nd ed. Springer, 2009.

[75] Vladimir Vapnik and Alexei Ya. Chervonenkis. 'A note on one class of perceptrons'. In: 1964.

[76] J. Varanasi and M. M. Tripathi. 'Artificial Neural Network based wind speed & power forecasting in US wind energy farms'. In: *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*. July 2016, pp. 1–6. DOI: 10.1109/ICPEICES.2016.7853622.

[77] Y. Wang. 'Short-term wind power forecasting by genetic algorithm of wavelet neural network'. In: *2014 International Conference on Information Science, Electronics and Electrical Engineering*. Vol. 3. Apr. 2014, pp. 1752–1755. DOI: 10.1109/InfoSEEE.2014.6946223.

[78] W. L. Woon and O. Kramer. 'Enhanced SVR ensembles for wind power prediction'. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. July 2016, pp. 2743–2748. DOI: 10.1109/IJCNN.2016.7727544.

[79] W. Wu et al. 'Probabilistic short-term wind power forecasting based on deep neural networks'. In: *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*. Oct. 2016, pp. 1–8.

[80] G. Xu and L. Xia. 'Short- Term Prediction of Wind Power Based on Adaptive LSTM'. In: *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*. Oct. 2018, pp. 1–5. DOI: 10.1109/EI2.2018.8582536.

[81] H. Xue et al. 'Short-Term Wind Power Prediction Based on Improved Chicken Algorithm and Support Vector Machine'. In: *2018 International Symposium on Computer, Consumer and Control (IS3C)*. Dec. 2018, pp. 137–140.

[82] L. Yang et al. 'Support-Vector-Machine-Enhanced Markov Model for Short-Term Wind Power Forecast'. In: *IEEE Transactions on Sustainable Energy* 6.3 (July 2015), pp. 791–799. ISSN: 1949-3029. DOI: 10.1109/TSTE.2015.2406814.

[83] Z. Yin et al. 'A new combination model for short-term wind power prediction'. In: *2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*. Nov. 2015, pp. 1869–1873. DOI: 10.1109/DRPT.2015.7432552.

[84] J. Zeng and W. Qiao. 'Support vector machine-based short-term wind power forecasting'. In: *2011 IEEE/PES Power Systems Conference and Exposition*. Mar. 2011, pp. 1–8. DOI: 10.1109/PSCE.2011.5772573.

[85] W. Zhang et al. 'A hybrid EMD-SVM based short-term wind power forecasting model'. In: *2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*. Nov. 2015, pp. 1–5. DOI: 10.1109/APPEEC.2015.7380872.

[86] Xin Zhao, Shuangxin Wang and Tao Li. 'Review of Evaluation Criteria and Main Methods of Wind Power Forecasting'. In: *Energy Procedia* 12 (2011). The Proceedings of International Conference on Smart Grid and Clean Energy Technologies (ICSGCE 2011, pp. 761–769. ISSN: 1876-6102. DOI: https://doi.org/10.1016/j.egypro.2011.10.102. URL: http://www.sciencedirect.com/science/article/pii/S187661021101928X.

[87] Alice Zheng and Amanda Casari. *Feature engineering for machine learning: principles and techniques for data scientists*. O'Rielly, 2018.

[88] H. Zhou and J. Fang. 'Wind power forecasting based on econometrics theory'. In: *2010 IEEE Electrical Power Energy Conference*. Aug. 2010, pp. 1–6. DOI: 10.1109/EPEC.2010.5697193.

[89] H. Zhou, J. Fang and M. Huang. 'Numerical analysis of application GARCH to short-term wind power forecasting'. In: *2010 International Conference on Power System Technology*. Oct. 2010, pp. 1–6.

[90] S. M. Zhu, M. Yang and X. S. Han. 'Short-term generation forecast of wind farm using SVM-GARCH approach'. In: *2012 IEEE International Conference on Power System Technology (POWERCON)*. Oct. 2012, pp. 1–6. DOI: 10.1109/PowerCon.2012.6401309.

[91] Ziming Zhu, Doudou Zhou and Zhong Fan. 'Short term forecast of wind power generation based on SVM with pattern matching'. In: *2016 IEEE International Energy Conference (ENERGYCON)*. Apr. 2016, pp. 1–6. DOI: 10.1109/ENERGYCON.2016.7514064.