Time Series Based Forecasting of Renewable Power Infeed for Operation of Microgrids

Master thesis

from

Elin Klages

born on

08. June 1989 in Hamburg, Germany

04. October 2016

Supervisor:	Christian A. Hans
Examiner:	Prof. Dr.-Ing. Jörg Raisch Control Systems Group Department of Energy and Automation Technology Faculty IV - Electrical Engineering and Computer Science Technische Universität Berlin
Second Examiner:	Prof.-Dr.-Ing. Clemens Gühmann Chair of Electronic Measurement and Diagnostic Technology Department of Energy and Automation Technology Faculty IV - Electrical Engineering and Computer Science Technische Universität Berlin

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeflührten Quellen und Hilfsmittel angefertigt habe.

Berlin, den 01. August 2016

Abstract

The proportion of renewable energy, in relation to conventional energy, is increasing. Especially the fluctuation of energy from renewable sources causes difficulties for the operation of electrical grids. In the present thesis, methods for time series based forecasts of renewable power infeed are investigated. Therefore, Autoregressive Integrated Moving Average (ARIMA) Models, Kernel, k Nearest Neighbour (kNN), and Support Vector (SV) Regression are applied to predict future values of wind speed and solar irradiance. Forecasts up to 2 h in advance with a resolution of 10 min are considered as relevant for the operation of a microgrid. The performance of the trained models is evaluated based on a comparison to a naive forecast. Results show that prediction of future values of solar irradiance is most accurate if the the observed values of solar irradiance measured at ground level are normalized using the estimated extraterrestrial solar irradiance. Support Vector Regression yields the most accurate forecasts for solar irradiance of the above mentioned techniques. For forecasting windspeed, the results do not indicate a clear tendency which regression technique to use. While Support Vector Regression works best for forecast from 10 min up to 2 h, ARIMA provides better results for the first predictions steps and Kernel Regression for the last prediction steps. For both techniques, improvement over the naive forecast is highest for prediction horizons close to 2 h (up to 20.9% for solar irradiance), while prediction of the next 10 min cannot be improved significantly compared with a naive forecast (<3%).

Kurzfassung

Der Anteil von regenerativen Energien, im Vergleich zu konventionellen Energien, nimmt stetig zu. Insbesondere die Fluktuation von Energien aus erneuerbaren Quellen, wirkt sich negativ auf die Regelung von elektrischen Netzen aus. In der vorliegenden Arbeit werden Methoden zur Vorhersage der Einspeisung von erneuerbaren Energien untersucht. Es werden "Autoregressive Integrated Moving Average (ARIMA)" Modelle, Kernel, k Nächste Nachbarn (kNN), und Stützvektor (SV) Regression verwendet um zukünftige Werte für Sonneneinstrahlung und Windgeschwindigkeit vorherzusagen. Vorhersagen bis zu 2 h in die Zukunft bei einer Auflösung von 10 min werden als relevant für die Regelung eines Microgrids angenommen. Die Bewertung der tranierten Modelle erfolgt im Vergleich zu einer naiven Vorhersage. Ergebnisse für die Vorhersage von Sonneneinstrahlung zeigen, dass die beste Vorhersagengenauigkeit erreicht wird, wenn die Daten zunächst mit der extraterrestrialen Sonneneinstrahlung normiert werden. Stützvektor Regression erreicht, von den oben genannten Techniken, die höchste Vorhersagegenauigkeit für die Vorhersage von Sonneneinstrahlung. Die Ergebnisse für die Windgeschwindigkeitvorhersagen sind weniger deutlich. Stützvektor Regression erreicht die höchste Genauigkeit für Vorhersagen von 10 min bis 2 h, während ARIMA bessere Ergebnisse für Vorhersagen bis zu 30 min erzielt und Kernel Regression für die Vorhersagen nah an 2 h am genausten ist. In beiden Fällen zeigt sich, dass die Vorhersagegenauigkeit im Vergleich zu einer naiven Vorhersage am meisten für Vorhersagen von 2 h in die Zukunft zunimmt (bis zu 20,9% für Sonneneinstrahlung), während bei Vorhersagen für 10 min in die Zukunft die Vorhersagegenauigkeit nur geringfügig verbessert werden kann (<3%).

Contents

List of Figures

List of Tables

Glossary

B Backwards shift operator. vii, viii, 12, 13

$\Phi(B^s)$ Seasonal autoregressive operator of a seasonal ARIMA model. 13

$\Theta(B^s)$ Seasonal moving average operator of a seasonal ARIMA model. 13

α Significance level of the hypothesis test. 6, 15, 16

γ Solar altitude. v, 6, 7

$\hat{y}$ Predicted time series. 11, 12, 16

μ Mean value, also called expected value. 3, 4, 5, 6, 20

$\phi(B)$ Autoregressive operator of an ARIMA model. 12, 13

ρ Correlation. 5

σ^2 Variance. 4, 5

σ Standard Deviation. 3, 4, 5

τ Transmissivity. Relation between solar irradiance that reaches the earth surface and that enters the earths atmospehere. 9

Cov Covariance. 4, 5

$\mathbf{D_N}$ Observed data. N pairs of measured independent variable x_i and dependent variable y_i, $D_N = \left\{(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)\right\}$. 11, 12, 16, 17, 19

$\mathbf{G_0}$ Extraterriestral solar irradiance. 6, 7, 9

$\mathbf{G_{dh}}$ Diffuse horizontal irradiance. 8, 9

$\mathbf{G_{dn}}$ Direct normal irradiance. 9

$\mathbf{G_{sc}}$ Solar constant. 7

G Global horizontal irradiance. 8, 9

$\mathbf{H_0}$ Null hypothesis. 5, 6

$\mathbf{N_{min}}$ Number of historical data points needed to predict future values.. 29, 30, 32

$\mathbf{N_p}$ Number of parameters of a model. 15, 16

$\mathbf{N}$ Number of samples. 11, 15, 16, 28, 29

$\mathbf{T}$ Test statistic of hypothesis test. 5, 6

$\mathbf{c}$ Critical value of hypothesis test. 5, 6

$\theta(B)$ Moving average operator of an ARIMA model. 12, 13

θ Parameter vector of a parametric model. 12

a Julian day.. 7

d Degree of differencing of an ARIMA model. 12

e Residuals. Prediction minus measured values. 12, 26, 28, 29, 30, 31, 32

f Regression function or index to indicate forecast. Clear from context.. 12

j Index to indicate prediction horizon. 5, 30, 31, 32

k Index to indicate time instant or number of neighbors used for k Nearest Neighbour Regression. Clear from context.. 5, 11, 12, 13, 15, 16, 20, 28, 32

$p(i)$ Probability associated with the ith value of the set of possible values of the random variable. 3, 4

p Prediction Horizon. 29, 32

r Autocorrelation. 5, 15

s Length of one season of an SARIMA model. 13

x_i ith value of the set of possible values of the random variable $\boldsymbol{X}$. 3, 4, 16

x Time series respectively independent variable.. 11, 12, 16, 20

y Time series respectively response/dependent variable.. 5, 6, 11, 12, 13, 16, 20, 32

z Zenith angle. v, 6, 9

$\boldsymbol{X}$ Random Variable. viii, 3, 4, 5

RMSE Root mean square error. 28, 30, 31, 32, 33, 35, 36, 40, 41

sr Index to indicate sunrise. 30, 31

ss Index to indicate sunset. 30, 31

Chapter 1

Introduction

Time series based forecasting, i.e., predicting future values based on previously observed values, has been used in various areas. In the context of the operation of a microgrid, it is of interest to estimate beforehand future power infeed and load. In contrast to infeed of conventional energy, infeed of renwable energy[1] cannot be set by the operators. While load is relatively predictable[2], the fluctuation of renewable power infeed make it difficult to manage in the context of the operation of electrical grids. Here, accurate forecasts of renewable power infeed could improve the operation of the grid and yield to decreased use of energy from conventional sources. Solutions, based merely on already provided information due to the operation of a microgrid (like current time, previously measured power infeed and basic knowledge about the location of the grid) keep additional costs and effort small. Therefore, the present work aims to present how short term prediction (here up to 2 h) of renewable power infeed (solar and wind power) can be realized accurately and efficiently.

Various studies on forecasting wind speed and solar irradiance have been published. However, most focus on either physical models or long term predictions or both. Especially for forecasting of solar irradiance, hardly any work can be found for forecasting intervals less than 1 hour. Still, previous research indicates that time series based forecasts perform better for short term predictions, while physical models work best for long term predictions, e.g., two days ahead, [Qiao and Zeng, 2012, Section 1.], [Heinemann et al., 2006, Section 1.], [Bacher et al., 2009, Section 7.]. This supports the approach of the present work, to explore the information provided by historical data instead of using complex numerical simulations to obtain estimates of future values of renewable power infeed.

In both fields, forecasting solar irradiance and wind speed, as for time series forecasts

[1] In 2015, already 32.6% of electric current in Germany came from renewable sources, [Statistik, 2016]. Until 2025, a percentage of 40-45% is planned, [Werdermann, 2016].

[2] For example, lights are switched on when it gets dark.

in general, classically used are Autoregressive Integrated Moving Average (ARIMA) models. Recently, research has been done on using other techniques, as e.g., Artificial Neural Networks (ANN), [Sfetos, 2001], [Cadenas and Rivera, 2006], [Qiao and Zeng, 2012], [Mellit and Pavan, 2010], Fuzzy Logic Models (especially for wind speed), [Monfared et al., 2009] and Support Vector Regression (SVR), [Qiao and Zeng, 2012], [Zhou et al., 2011]. There are various studies that compare two or more techniques, although comparison is done mainly for ARIMA models and one Artificial Intelligence related technique. However, no clear conclusion can be drawn from these. Firstly, results differ. For example, a study on forecasting wind speed, concludes that ANN achieves significantly better results than ARIMA models, [Sfetos, 2001]. Another study that compares ANN and ARIMA for forecasting wind speed, comes to the opposite conclusion, saying that Seasonal ARIMA models perform better than ANN, [Cadenas and Rivera, 2006]. Secondly, the use of different data sets measured at different locations, and the use of different error measurements, make it hard to compare results and to use the information provided by previous research. Sfetos suggests, that the improvement over the naive forecast could be used to compare the performance forecast techniques, evaluated on different data, [Sfetos, 2001, Section 1.]. However, the results of a naive forecast are often not provided in the articles.

The present work focuses on forecasting wind speed and solar radiation using exclusively data that is already provided by the operation of a microgrid. Different techniques for time series based forecasts with predictions steps of 10 min with prediction horizons up to 2 h are explored, to find the most adequate technique for the operation of a microgrid, based on high solution data. The chosen techniques are ARIMA models, k Nearest Neighbour Regression (kNN), Kernel Regression and Support Vector Regression. ARIMA models are selected to be able to compare the performance of other techniques to ARIMA models, since they are considered as the classical way to approximate time series. Kernel and kNN Regression are simple, intuitive techniques that can be implemented with little effort. Also, no training of a model is needed. For these reasons, these techniques were chosen to examine if with simple regression methods good forecast quality can be achieved. As a more complex regression technique, Support Vector Regression was chosen. Studies using SVR indicate that it could provide better results than ANN, e.g., [Qiao and Zeng, 2012]. As mentioned above, the trained models are evaluated based on the improvement relative to a naive forecast.

The reminder of this paper is organized as follows. Basic information on statistics and solar radiation are given in chapter 2. In chapter 3, the applied regression techniques are explained in detail. How the techniques were implemented is clarified in chapter 4. In chapter 5, achieved forecast accuracy is presented and discussed. As an example, results of a performed simulation of the operation of a microgrid, using Support Vector Regression to predict solar irradiance, is presented in chapter 6.

Chapter 2

Background

Before introducing the applied regression techniques to forecast future values of wind and solar power, information on statistics and solar radiation is provided by this chapter. This chapter aims to provide the needed background information for Chapter 3 and 4.

2.1 Statistics

In the following, some basics and some content of more advanced topics of statistic are given. This section primarly intends providing sufficient information to understand this work. It is also thought to simplifify understanding the cited literature.

Probability and Random Variables

A **random variable** $\boldsymbol{X}$ represents a set of possible different values x_i, where the occurrence of each value, for example as output of an experiment, is associated with a **probability** $p(i)$. A **probability distribution** describes the random variable completely, i.e., its values and their associated probabilities (see , e.g., figure 2.1, example is explained below), [Wasserman, 2005, Sec. 1.1.-1.4, 2.1]. Random variables can be continuous or discrete. However, in this work only discrete variables (time series) are considered.

One example of a random variable is the occurence of rain fall during one day in Berlin[1]. The two possible outcomes, the so called **sample space**, are *day with rainfall* or *day without rainfall*, represented by 1 and 0. Their probability distribution is[2]

$$p(i) = \begin{cases} 0.73, & \text{if } i = 0 \\ 0.27, & \text{if } i = 1. \end{cases} \tag{2.1}$$

[1] Another example of a random variable is the side of a coin, that occurs when a coin is flipped, where the sample space includes the two sides and the probability of each side is 0.5. An example with a sample space that includes more than two possible outcomes would be , e.g., the number of children in family.

[2] This is meant to serve as an example, for this reason the probabilities should not be taken too seriously.

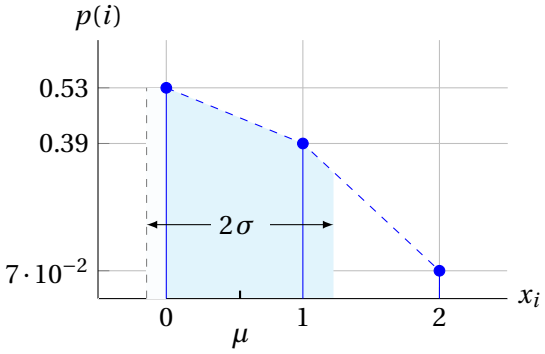


Figure 2.1: The discrete probability distribution, with mean value $\mu = 0.54$ and standard deviation $\sigma = 0.68$, of the example in (2.2). The colored part indicates the part of the probability distribution included by the standard deviation. Nonnegative values are not colored, since it is impossible to get a negative number of days with rainfall.

If two days instead of one are considered, the example changes. The probability distribution would be

$$p(i) = \begin{cases} 0.73 \cdot 0.73 = 0.53, & \text{if } i = 0, \quad \text{it does not rain on both days,} \\ 0.73 \cdot 0.27 + 0.27 \cdot 0.73 = 0.39, & \text{if } i = 1, \quad \text{it rains on one day,} \\ 0.27 \cdot 0.27 = 0.07, & \text{if } i = 2, \quad \text{it rains on both days.} \end{cases} \tag{2.2}$$

The probability is visualized in figure 2.1.

In the following it is assumed that the underlying random process is **stationary**, i.e. the probability distribution is time independent.

The **mean value** μ of a random variable $\boldsymbol{X}$, also called expected value, is defined as

$$\mu(\boldsymbol{X}) = \sum_{i=1}^{n} p(i)\, x_i. \tag{2.3}$$

As a measure of the spread of a random value, its **variance** σ^2 ,

$$\sigma^2\,(\boldsymbol{X}) = \mu\left(\left(\boldsymbol{X} - \mu_{\boldsymbol{X}}\right)^2\right), \tag{2.4}$$

respectively its **standard deviation** σ can be used. The variable $\mu_{\boldsymbol{X}}$ refers to $\mu(\boldsymbol{X})$. A high dispersion around the mean value is indicated by a high variance / standard deviation, while a small variance / standard deviation indicates that the results tend to lie close to the mean, [Wasserman, 2005, Def. 3.14], (see figure 2.1 for an example).

If, instead of one variable $\boldsymbol{X}$, two random variables $\boldsymbol{X}$, $\boldsymbol{X}^*$ are considered (e.g. if it rains in Berlin and if it rains in Potsdam), the variance (see (2.4)), turns into the **covariance**,

$$\mathrm{Cov}\,(\boldsymbol{X}, \boldsymbol{X}^*) = \mu\left(\left(\boldsymbol{X} - \mu_{\boldsymbol{X}}\right)\left(\boldsymbol{X}^* - \mu_{\boldsymbol{X}^*}\right)\right), \tag{2.5}$$

where $\mu_{\boldsymbol{X}}$ is the mean value of $\boldsymbol{X}$ and $\mu_{\boldsymbol{X}^*}$ is the mean value of $\boldsymbol{X}^*$. The **correlation** between these two variables is defined by

$$\rho(\boldsymbol{X},\boldsymbol{X}^*) = \frac{\mathrm{Cov}(\boldsymbol{X},\boldsymbol{X}^*)}{\sigma_{\boldsymbol{X}}\sigma_{\boldsymbol{X}^*}}, \tag{2.6}$$

where $\sigma_{\boldsymbol{X}}$ is the standard deviation of $\boldsymbol{X}$ and $\sigma_{\boldsymbol{X}^*}$ is the standard deviation of $\boldsymbol{X}^*$. The covariance and the correlation are both measurements of the linear relationship between the two variables $\boldsymbol{X}$ and $\boldsymbol{X}^*$, [Wasserman, 2005, Def. 3.18].

If the correlation (see (2.5)) is applied for the same random variable (e.g. if it rains during the day) but at different points of time, $\boldsymbol{X}_k, \boldsymbol{X}_{k-j}$ (e.g. if rains on Monday and if it rains on Tuesday) the **autocorrelation** is obtained, [Hyndman and Athanasopoulos, 2014, p.35],

$$r(\boldsymbol{X}_k,\boldsymbol{X}_{k-j}) = \frac{\mu\left((\boldsymbol{X}_k-\mu_{\boldsymbol{X}})(\boldsymbol{X}_{k-j}-\mu_{\boldsymbol{X}})\right)}{\sigma^2}. \tag{2.7}$$

As an alternative to autocorrelation as a measure between correlation of elements of a time series, the **partial autocorrelation** can be used. The partial autocorrelation of $\boldsymbol{X}_k$, $\boldsymbol{X}_{k-j}$ is the correlation after removing all correlation with $\boldsymbol{X}_{k-1},\ldots,\boldsymbol{X}_{k+1-j}$, [Hyndman and Athanasopoulos, 2014].

Statistical Inference

Statistical inference deals with infering information about the data generating process from the observed data, [Wasserman, 2005, preface]. This section does not serve as a general overview about statistical inference; the aim is merely to present some techniques, later on used in the context of time series forecasting using ARIMA models (see Section 3.1). For a general overview the reader is referred to the cited books in this section.

Hypothesis Testing

Hypothesis testing refers to a procedure used to measure how much evidence provides the observed data against a theory, typically named **null hypothesis** H_0, about the population the data is coming from, [Wasserman, 2005, p.94].

The **hypothesis test** itself is a rule that specifies when to reject or retain the null hypothesis. Usually a **test statistic** $\mathrm{T}(y)$, a function of a sample of the population y, in combination with a **critical value** c that functions as threshold for the test statistic, is used as test rule, [Casella and Berger, 2001, Chapter 8, p.374-375].

An example for a null hypothesis could be the theory that the mean height s of all habitants of a certain region is 170 cm, written as

$$\mathrm{H}_0 : \mu(s) = 170\,cm. \tag{2.8}$$

To test H_0, a group of 30 habitants is taken and the height s of each one is measured. As test statistic might be used the distance of mean height of these 30 habitants to the theoretical mean height,

$$T(s) = \left|170\,cm - \mu(s_i)\right|,\ i = 1,\ldots,30. \tag{2.9}$$

If the mean hight results to be , e.g., 172.34 cm ($T(s) = 2.34\,cm$), the observed data (the 30 habitants) would not provide much evidence to reject the null hypothesis, but neither proves it. But if 30 other people were chosen and the mean hight was , e.g., 145,12 cm ($T(s) = 24.88\,cm$), the null hypothesis would seem to be more likely false. But, as before, the test statistic doesn't prove that the null hypothesis is wrong or right.

To decide to reject or retain the null hypothesis, the test statistic is compared to a predefined critical value,

$$H_0 \text{ is } \begin{cases} \text{rejected}, & \text{if } T(y) \geq c, \\ \text{retained}, & \text{if } T(y) < c, \end{cases} \tag{2.10}$$

[Kalbfleisch, 1979, p.136].

The choice of the critical value is linked to the so called **significance level** α of the test. The significance level is defined by the probability that the test statistic $T(y)$ is smaller or equal to the critical value c, if the null hypothesis H_0 is true,

$$\alpha = p\left\{c \geq T(y)|H_0\right\}. \tag{2.11}$$

Hence, the significance level α does not refer to the probability that the null hypothesis is true, [Kalbfleisch, 1979, p.136-137].

A statement as

> *The hypothesis test xy was passed at the 5% significance level.*

means that the critical value c was chosen to obtain an α (see (2.11)) of 0.05 and that the test statistic $T(y)$ for the given data is as least as large as c. This means the test is also passed for all significance level less than 5%. This statement does not provide the information if the test would be also passed at the 10% or even 90% level. Therefore it might be more informative to pass on the lowest significance level the data passes. This lowest passed significance level is often referred to as p-value[3].

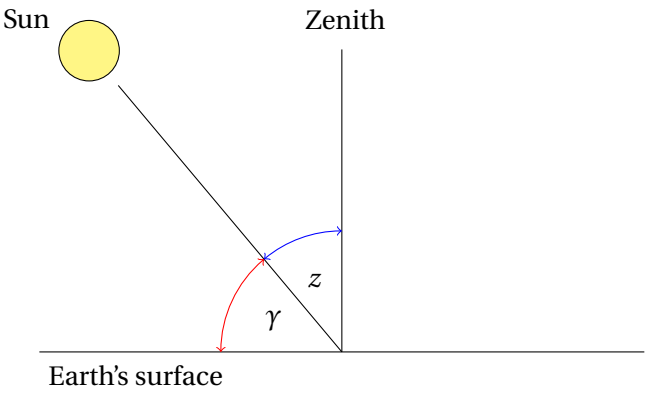


Figure 2.2: Zenith angle z and solar altitude γ.

2.2 Solar Radiation

This section intends to provide all information needed, to understand how the solar radiation data is used in this work and in work cited work by others. The solar energy that arrives at the limit of the earth's atmosphere is called extraterriestral solar radiation, [IS/ISO-9488, 1999]. It can be predicted by the amount of energy emitted by the sun and the position of the sun relatively to the earth. The extraterrestrial solar irradiance, irradiation per area, G_0 is given by

$$G_0 = \varepsilon G_{sc} \sin(\gamma) \tag{2.12}$$

where $G_{sc} = 1360.8\,W/m^2$ is the solar constant[4], γ is the solar altitude in degrees (see figure 2.2) , ε a correction factor for the distance of the sun to the earth[5]

$$\varepsilon = 1 + 0.0334 \cos(j' - 2.80^\circ), \tag{2.13}$$

j' the day angle,

$$j' = \frac{a}{365.25} 360^\circ, \tag{2.14}$$

[3]For example in this section cited book *All about Statistics*, [Wasserman, 2005], or by MATLAB. In the other here cited book *Probability and Statistical Inference*, [Kalbfleisch, 1979], p-value and significance level are used synonymously.

[4]Different references define slightly different values for the solar constant. Although in the *The European Solar Radiation Atlas*, where equation (2.12) is taken from, the solar constant is defined as $1367\,W/m^2$, [Scharmer and Greif, 2000, Chapter 3, p.28], the value used in this work, $1360.8\,W/m^2$, is chosen after Kopp and Lean, Kopp and Lean [2011].

[5]The distance between sun and earth varies over the year, because the earth's orbit isn't a circle.

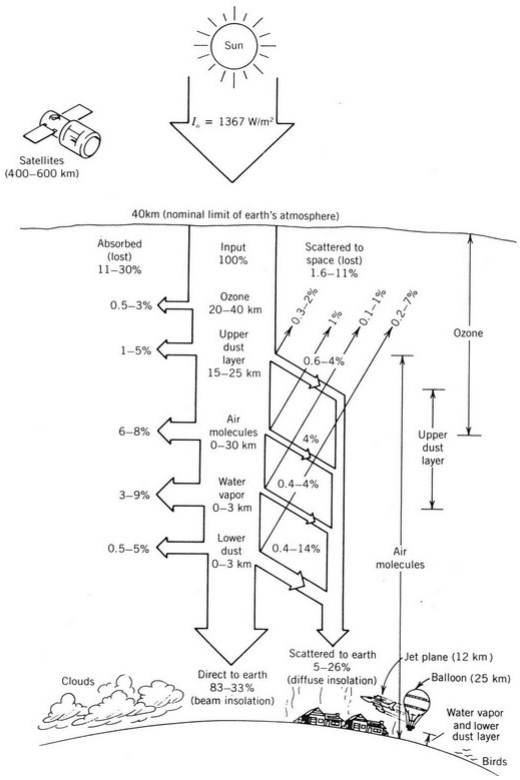


Figure 2.3: Nominal range of clear sky absorption and scattering of incident solar energy, [Stine and Harrigan, 1986, Figure 2.9]. Note: Insolation is a deprecated term for solar radiation, [IS/ISO-9488, 1999, 3.13].

and a is the *Julian Day* [6] the number of the day in the year (1-366) [Scharmer and Greif, 2000, Chapter 3, p.28].

Ground level solar radiation is more difficult to predict than the radiation that reaches the outer edge of the atmosphere, as only a part of solar radiation is transmitted directly while other parts are absorbed, reflected, and scattered by the atmosphere and clouds (see figure 2.3), [Stine and Harrigan, 1986, Section 2.2].

The solar radiation received on a horizontal plane on earth is called global solar radiation, [IS/ISO-9488, 1999, 3.2], or global horizontal radiation[7]. The global horizontal irra-

[6] A note to avoid confusion. In other references, the term *Julian Day* might be used for numbering the days when counting them continuously since some historical date (see `https://en.wikipedia.org/w/index.php?title=Julian_day&oldid=715401672`), while in the context of counting them from January 1st, the term *day number* might be used, as for example in the book *Solar Energy Systems Design*, [Stine and Harrigan, 1986, Section 2.1.3, equation 2.1].

[7] e.g. National Renewable Energy Laboratory (NREL)

diance G includes the direct irradiance G_{dh} (often called beam irradiance, [Quaschning, 2004, Section 2.1]), which comes directly from the sun, and the diffuse irradiance G_{dh}, which includes scattered and ground reflected[8] irradiance, [IS/ISO-9488, 1999, 3.20, 3.21, 3.25]. Direct irradiance is usually measured at normal incidence (see figure 2.2), [IS/ISO-9488, 1999, 3.17], and therefore in the context of this work called direct normal irradiance to avoid confusion. Hence, the relation between global horizontal irradiance and it's both components is,

$$G = G_{dn} + G_{dh} \cos(z). \tag{2.15}$$

Direct normal irradiance G_{dn} can be measured by a pyrheliometer, that is always aimed at the sun, [IS/ISO-9488, 1999, 4.7] with help of a solar tracking algorithm. A pyranometer, which measures solar irradiance on a plan surface, [IS/ISO-9488, 1999, 4.4] can be used to measure the global horizontal irradiance G or the diffuse horizontal irradiance G_{dh}. The latter can be obtained either by taking the difference between the global horizontal irradiance or by measuring the solar irradiance while covering the current normal to the sun.

Another possible representation of the solar radiation reaching the earth, instead of using it's absolute amount, is using the transmissivity for the comparison between solar radiation entering the earth's atmosphere and the radiation reaching earth's surface,

$$\tau = \frac{G}{G_0} \tag{2.16}$$

as for example used in [Qiao and Zeng, 2012]. By using also the extraterriestral irradiance, which can be easily predicted, as explained above (see equation (2.12)), this representation already includes meteorological information.

2.3 Summary

Hypotheis tests, as used for the evaluation of ARIMA models, are described. For this reason and also as background information for forecasting in general, the basics of probabilites are discussed shortly. To give an explanation of the particularity of solar radiation data in relation to wind speed data, solar radiation is discussed. In the same context, the transmissivity is introduced. An explanation of the used regression techniques, as for example ARIMA models, follows in the next chapter.

[8] The part reflected by the ground is usually an insignificant amount `http://rredc.nrel.gov/solar/glossary/gloss_g.html`, see *Global Horizontal Radiation*

Chapter 3

Time Series Based Forecast

This chapter aims to give an overview about the regression techniques used in this work. Regression in general refers to infer information about the data generating process from historical data. In other words, regression is about finding a regression function $f(x)$ that approximates y,

$$\hat{y} = f(x), \tag{3.1}$$

based on the observed data

$$\mathrm{D_N} = \left\{ (x_1, y_1), (x_2, y_2) \ldots (x_\mathrm{N}, y_\mathrm{N}) \right\}, \tag{3.2}$$

where x is the independent variable and y is the dependent variable, [Gyoerfi et al., 2002, p.1-3].

Autoregression is regression using only historical data of the response variable y. Therefore, the independent variable x represents one or more historical values of y,

$$x_k = [y_{k-1}, y_{k-2}, \ldots, y_{k-n}], \tag{3.3}$$

where n is the number of elements in x. In this chapter as in this work in general, the focus lies on autoregression. In contrast, regression in general can rely on any historical data. However, all in the following presented techniques can be also used, or at least extended, to regression in general.

In the following, it is differentiated between parametric and nonparametric regression. In parametric regression, the model, used to approximate the data, depends on the parameters derived from the data, while in non parametric regression, the model depends directly on the data.

On purpose, it is not focused on separating between Statistical Inference and Machine Learning. The main reason is that there is no clear difference between these two fields.

In contrast, they are often described as the same, e.g., by Larry Wasserman in *All about statistics*[1] or nearly the same, e.g., by Jerome H. Friedmann in *The Role of Statistics in the Data Revolution?*[2]. The main difference may lie in the way their techniques are presented.

Apart from the techniques used for the forecasts, this chapter also explains how multiple step forecasts are done and how a probability distribution for the forecasts is here estimated.

3.1 Parametric Regression

In parametric regression, the regression function f depends on the independent variable x and the parameters θ,

$$\hat{y} = f(x, \theta). \tag{3.4}$$

The parameters θ are estimated based on the observed data $\mathrm{D_N}$. In the following the only parametric models used in this work, ARIMA models, are presented.

The ARIMA Model

Often used linear models for time series forecasting are autoregressive integrated moving average (ARIMA) models. Making use of the backwards shift operator B and the backward difference operator ∇, which are defined by

$$B^m y_k = y_{k-m}, \tag{3.5}$$

$$\nabla^d y_k = \left(y_k - y_{k-1}\right)^d, \tag{3.6}$$

[Box et al., 1994, p.8], an ARIMA model can be defined as follows,

$$\phi(B)\,\nabla^d y_k = \theta_0 + \theta(B)\,e_k, \tag{3.7}$$

where y_k is a sample of the time series at time instant k, e_k the error of the forecast compared to the measured value, $\phi(B)$ the autoregressive operator, ∇^d the integrative part, and $\theta(B)$ the moving average operator. The autoregressive and the moving average operator are polynomials with degree p and q respectively,

$$\phi(B) = 1 - \phi_1 B^{-1} - \phi_2 B^{-2} - \cdots - \phi_p B^{-p}, \tag{3.8}$$

$$\theta(B) = 1 - \theta_1 B^{-1} - \theta_2 B^{-2} - \cdots - \theta_q B^{-q}, \tag{3.9}$$

[1] *The basic problem of statistical inference is about the inverse of probability, infer the data generating process from the observed data. Prediction, classification, clustering and estimation are all cases of statistical inference. Data analysis, data mining and machine learning are just different names given to the practise of statistical inference,* [Wasserman, 2005, Preface, page ix].

[2] *Had we incorporated computing methodology from its inception as a fundamental statistical tool (...) many of the other data related fields [e.g., machine learning] would not have needed to exist. They would have been part of our field [statistics],* [Friedmann, 2001]

[Box et al., 1994, p.96].

Due to its different parts, this class of models also includes autoregressive (AR) models, moving average models (MA), and the mixed autoregressive, moving average (ARMA) models. Due to differencing (the integrative part of the ARIMA model), ARIMA models can be also fitted to nonstationary data, if stationary data can be obtained by differencing of the original data, [Box et al., 1994, p.89].

To fit a model to data that exhibits seasonality, ARIMA models can be extended by seasonality of a certain period s. These models can be described by,

$$\phi(B)\,\Phi(B^s)\,\nabla^d\,\nabla_s^D\,y_k = \theta_0 + \theta(B)\,\Theta(B^s)\,e_k, \tag{3.10}$$

where $\Phi(B^s)$ and $\Theta(B^s)$ are the seasonal autoregressive and seasonal moving average operator, [Box et al., 1994, Chapter 9, p.332], defined as

$$\Phi(B^s) = 1 - \Phi_1 B^{-s} - \Phi_2 B^{-2s} - \cdots - \Phi_p B^{-Ps}, \tag{3.11}$$

$$\Theta(B^s) = 1 - \Theta_1 B^{-s} - \Theta_2 B^{-2s} - \cdots - \Theta_q B^{-Qs}, \tag{3.12}$$

and ∇_s^D represents seasonal differencing

$$\nabla_s^D\,y_k = \left(y_k - y_{k-s}\right)^D. \tag{3.13}$$

Model Identification and Estimation

To fit a model to a given time series, at first, the order of the autoregressive and the moving average operator of the model and the degree of differencing needs to be defined. This step is called Identification. The second step, Estimation refers to the process of obtaining the values of the parameters. Identification and estimation usually overlap, since identification refers to obtaining a rough idea of the kind of model needed, [Box et al., 1994, Chapter 6, p.183-184]. See also figure 3.1.

The first step of identification is to difference the data as many times as needed to get stationary data. As second step follows to identify the resulting ARMA model. Here, the autocorrelation and the partial autocorrelation can be used. For an AR model, the last significant lag of the partial autocorrelation of the sample represents roughly the order of the autoregressive operator, [Box et al., 1994, Chapter 6, p.184-185]. Likewise the autocorrelation can be used to identify the order of a moving average operator in a MA model. For mixed models such approaches are usually not very effective in pratice, [Liu and Hudak, 1992, Chapter 5, p.5.7-5.9].

For identified models, estimation of the parameters can be done for example by maximum likelihood estimation[3].

[3] As used for example by MATLAB, see documentation of function *estimate*.

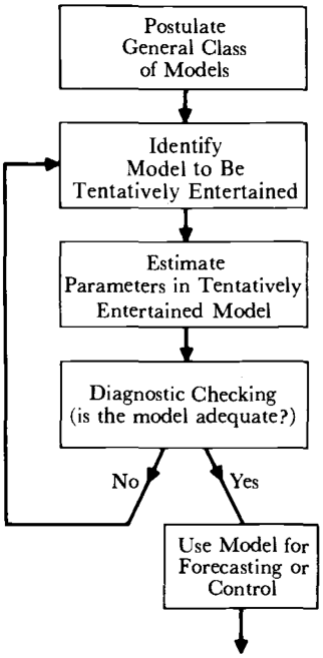


Figure 3.1: Stages in the iterative approach to model building as proposed by Box, Jenkins and Reinsel, [Box et al., 1994, Fig. 1.7]

Diagnostic Checks on the Model

The residuals from a well fitted model are expected to be independent and normally distributed. If this assumptions cannot be proved, the model needs modification, [Liu and Hudak, 1992, Chapter 5, p.5.15]. The independence of the residuals can be checked by the Ljung-Box Test, which tests whether the residuals exhibit autocorrelation or not. To access information about the distribution of the residuals, the Kolmogrov-Smirnov Test, which compares the distribution of two samples, can be used.

It is also of interest to check against overfitting, that means if some estimated parameters are not statistically different[4] from zero, [Liu and Hudak, 1992, Chapter 5, p.5.16]. To find an adequate number of parameters for the model, the Akaike information criterion, described in 3.1, can be also helpful.

The above mentioned hypothesis tests[5] are described in the following sections.

[4]The term *statistically different* refers to an assumption made based on a statistic. This does not prove that the parameter actually is different from zero, nor that the data generating process can be approximated by the chosen model.

[5]For general information about hypothesis tests see section 2.1

Ljung-Box Test

To test the null hypothesis that the residuals that result from the estimation of an ARIMA model, exhibit no autocorrelation, the Ljung-Box Test can be applied.

It is therefore used the test statistic

$$Q(m) = \mathrm{N}(\mathrm{N}+2) \sum_{k=1}^{m} \frac{\hat{r}_k^2}{T-k} \tag{3.14}$$

where $\hat{r}_k$ is the autocorrelation of the residuals and N the number of observed samples. Under the null hypothesis, $Q(m)$ follows a chi squared distribution χ_h^2, where the degree of freedom h equals the difference of the maximum lag of autocorrelation m and the number of parameters of the ARIMA model, [Enders, 1994, p.87]. A chi squared distribution is the distribution of a sum of the squares of h independent standard normal random variables [Woolridge, 2015, Section B-5d].

The test rule is to reject the null hypothesis if

$$Q(m) > \chi^2_{1-\alpha,\, m-\mathrm{N_p}},$$

[Enders, 1994, p.88], where α is the significance level, $\mathrm{N_p}$ the number of parameters of a model and $\chi^2_{1-\alpha,h}$ refers to the $1-\alpha$ quantile of the chi square distribution.

Kolmogrov-Smirnov Test

Generally, the Kolmogorov-Smirnov Test compares the distribution of two samples. However, in this case it is applied to compare the empirical cumulative distribution of the residuals that result from the estimation of an ARIMA model to a normal distribution.

The test statistic

$$d(\mathrm{N}) = \max\left|F_0(x) - S_\mathrm{N}(x)\right|, \tag{3.15}$$

is the maximum distance between the empirical cumulative distribution $S_N(x)$ and a hypothetical cumulative distribution (null distribution) $F_0(x)$. Depending on the chosen confidence interval, the test statistic $d(\mathrm{N})$ is then compared to the corresponding critical value $d(\mathrm{N})_\alpha$ (see e.g. table 1 in [Massey, 1951]). If the test statistic $d(\mathrm{N})$ exceeds the critical value, the null hypothesis is rejected, [Massey, 1951, Section 2].

Statistical Significance of Model Parameters

To access information about the significance of an estimated parameter, the null hypothesis, that the parameter equals zero, is considered. The used test statistic is defined as[6]

$$t_{\hat{\theta}} = \frac{\hat{\theta}}{\hat{\mathrm{se}}(\hat{\theta})} \tag{3.16}$$

[6] See documentation of MATLAB function *estimate*.

where $\hat{\theta}$ is the estimated parameter and $\hat{se}$ the estimated standard error of the parameter estimation. Under the null hypothesis the distribution of $t_{\hat{\theta}}$ follows the Student's t distribution [7].

The null hypothesis is rejected for

$$p(t_{\hat{\theta}}) \leq \alpha, \tag{3.17}$$

where α is the chosen significance level.

Selection of Model (Akaike Information Criterion)

Adding additional lags to a model reduces the training error, but, on the other hand, leads to more complicated models and doesn't necessarily improve the forecast quality. There exist various criteria that trade off a reduction of the trainings error for a simpler model. One of the most used is the Akaike Information Criteria, [Enders, 1994, Chapter 7, p.88]. The criterion is given by

$$\mathrm{AIC} = \frac{-2\ln(\text{maximum likelihood})}{\mathrm{N}} + \frac{2\,\mathrm{N_p}}{\mathrm{N}}, \tag{3.18}$$

where the maximum likelihood of the model comes from the estimation by maximum likelihood methods. Following the criteria, the best model is the one with the smallest AIC. Therefore the second term in equation (3.18) can be seen as a penalty factor for the inclusion of additional parameters, [Box et al., 1994, Chapter 6, p. 201].

3.2 Nonparametric Regression.

In nonparametric regression, the regression function f depends on the independent variable x and the observed data[8] $\mathrm{D_N}$,

$$\hat{y} = f(x, \mathrm{D_N}). \tag{3.19}$$

In the following, the nonparametric regression techniques, used in this work, are presented.

Kernel and K Nearest Neighbor Regression

K nearest neighbor and kernel regression are two similar algorithms. Both provide a regression function f by using the nearest neighbors of x in the training sample. The regression function can be written as,

$$\hat{y}(x) = f(x, \mathrm{D_N}) = \frac{1}{k} \sum_{x_i \in K(x)} y(x_i), \tag{3.20}$$

[7] `https://en.wikipedia.org/w/index.php?title=T-statistic&oldid=737116178`

[8] Recall that in parametric regression the regression function depends on x and the parameter(s) θ.

where $K(x_i)$ is the neighborhood of x in the trainings sample in (3.2) and k refers to the number of neighbors, [Altman, 1992, Equation 1]. For the k nearest neighbors algorithm, the number of neighbors k is fixed. While, the neighborhood for kernel regression is determined by a fixed maximum distance to x, [Altman, 1992, Section 2.].

An example is shown in figure 3.2. Let's assume one is interested in estimating the wage per hour based on the years of education and of experience of the considered worker. A data set, including various examples, is given. Based on this data set, an estimation of the wage per hour for two other employees, Andrew and Lilly, is requested.

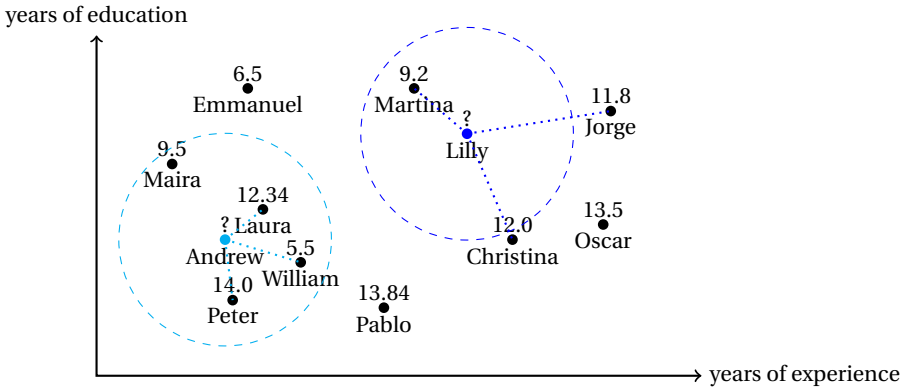


Figure 3.2: Example for local regression.

In figure 3.2 is shown which neighbors are selected, depending on the chosen regression technique. For k nearest neighbors, the same number of given examples are used to estimate the wage of Andrew and Lilly. k is here set to three. For kernel regression, the neighborhood $K(x)$ is described by the circle around x. For Andrew, four examples are included, for Lilly just one.

As an alternative to (3.20), a weighted average can be used, [Altman, 1992, Section 4.],

$$f(x,D) = \frac{\sum_{x_i \in K(x)} w_i \, y(x_i)}{\sum_{x_i \in K(x)} w_i} \tag{3.21}$$

In this work, the distance to the neighbor is used as weight.

Support Vector Regression

The Support Vector algorithm has its initials in Russia in the sixties. The today known form for classification was mainly developed in the nineties, focusing on optical character recognition. In the late nineties, good results were also obtained for regression and time series prediction, [Smola and Schölkopf, 2004, Section 1.1].

The Basic Idea.

The basic idea of Support Vector regression is to fit all given data D_N into a tube with width 2ε around an estimated regression function $f(x)$ (see figure 3.3). For the linear case, the model can be written as

$$\hat{y} = f(x) = w^T x + b. \tag{3.22}$$

The non linear case is considered in the next section.

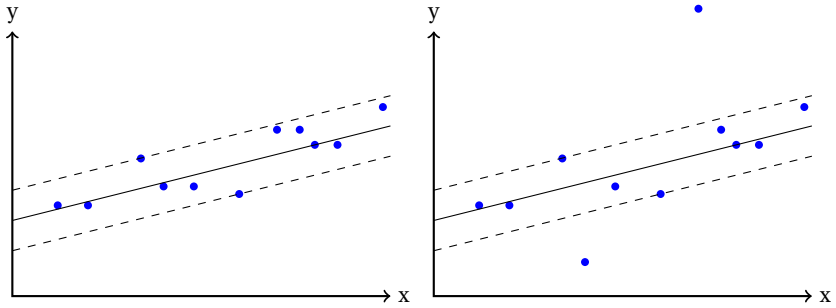


Figure 3.3: Linear one dimensional example for Support Vector regression.

The aim is to find a tube as flat as possible. This can be formulated as

$$\text{minimize } \frac{1}{2}\|w\|^2, \tag{3.23}$$

$$\text{subject to } \begin{cases} y_i - w^T x - b < \varepsilon, \\ -y_i + w^T x + b < \varepsilon, \end{cases} \tag{3.24}$$

where (3.23) is the objective function and (3.24) represents the corresponding constraints. Minimizing (3.23) is one option to search for a flat regression function. The square and the factor $\frac{1}{2}$ are not necessary but usually used, because this causes a simple deviation with respect to w.

For the optimization problem in equation (3.23)-(3.24), it is assumed that at least one feasible solution exists. Since this may not be the case using real data, an extension, the so called soft margin, is presented in the following. Let's consider the example in figure 3.3. Some data points are relatively far from the others. This data points can be included without choosing a large ε, by introducing slack variables ξ and ξ_i^*,

$$\text{minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*), \tag{3.25}$$

$$\text{subject to } \begin{cases} y_i - w^T x_i - b < \varepsilon + \xi_i, \\ -y_i + w^T x_i + b < \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* > 0. \end{cases} \tag{3.26}$$

The added slack variables ξ_i, ξ_i^* on the right hand side in (3.26) make it possible to have some data points with a larger distance than ε to the regression function $f(x)$. The additional term in the objective function penalizes the use of slack variables. This means that errors smaller than ε are not penalized, but that errors larger than ε are penalized proportional to its distance to ε, [Smola and Schölkopf, 2004, Section 1.2],

$$|\xi|_\varepsilon = \begin{cases} 0, & \text{if } |y_i - f(x_i)| < \varepsilon, \\ C\left(\xi_i - \varepsilon\right), & \text{if } |y_i - f(x_i)| > \varepsilon. \end{cases} \tag{3.27}$$

The Non Linear Case.

In (3.22), a linear SVR is stated. To represent non linear behaviour, the data $\mathrm{D_N}$ can be transferred to another space $\Phi(\mathrm{D_N})$, where it is possible to approximate it well by a linear function, [Smola and Schölkopf, 2004, Section 2.1].

Practically this isn't done by preprocessing the data, but by using so called Kernel[9] functions as part of the regression function in (3.22). First, it is necessary to have a look on how to resolve the optimization problem, before Kernel functions can be explained.

Including the constraints in the objective function using the Lagrange multipliers α_i and α_i^* yields

$$\begin{aligned} L = &\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\left(\xi_i - \xi_i^*\right) - \sum_{i=1}^{N}\eta_i\xi_i + \eta_i^*\xi_i^* \\ &- \sum_{i=1}^{N}\alpha_i\left(y_i - w^T x_i - b + \varepsilon + \xi_i\right) - \sum_{i=1}^{N}\alpha_i^*\left(-y_i + w^T x_i + b + \varepsilon + \xi_i^*\right). \end{aligned} \tag{3.28}$$

In an extrema of the objective function $L(w, \eta, \eta^*, \alpha, \alpha^*, \xi, \xi^*, b)$, the deviation with respect to all variables should be zero. From the derivative with respect to w

$$L_w = w - \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)\, x_i = 0, \tag{3.29}$$

one can conclude that the vector w can be represented by the Lagrange multipliers and the vectors of the trainings sample x_i,

$$w = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)\, x_i. \tag{3.30}$$

Therefore (3.22) can be rewritten as

$$\hat{y} = f(x) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)\,\Phi(x_i)\,\Phi(x_j) + b. \tag{3.31}$$

[9] Kernel functions should not be confused with Kernel Regression. Although names are really similar, they are not related.

for the general non linear case. The same can be done for the objective function (3.23), [Smola and Schölkopf, 2004, Section 1.3],

$$\text{minimize } \frac{1}{2}\sum_{i=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\Phi(x_i)\Phi(x_j). \tag{3.32}$$

This shows that neither the vector w nor the explicit transformation of the vectors x_i are necessary. Needed is only the scalar product of two transformed vectors x_i. These are Kernel functions, [Smola and Schölkopf, 2004, Section 2.2],

$$k(x_i, x_j) = \Phi(x_i)\Phi(x_j). \tag{3.33}$$

3.3 Multiple step forecasts.

To create multiple step forecasts, recursive strategy is applied. The trained regression model $f(x)$ is used for one step forecasts,

$$\begin{aligned}
\hat{y}_{k+1} &= f(x_k) = f(y_k, y_{k-1}, \ldots, y_{k-N_{\min}}),\\
\hat{y}_{k+2} &= f(\hat{x}_{k+1}) = f(\hat{y}_{k+1}, y_k, \ldots, y_{k-N_{\min}+1}),\\
\hat{y}_{k+3} &= f(\hat{x}_{k+2}) = f(\hat{y}_{k+2}, \hat{y}_{k+1}, y_k, \ldots, y_{k-N_{\min}+2}),
\end{aligned} \tag{3.34}$$

where the elements of x are adapted in each step, [Bontempi et al., 2013, Section 4.].

3.4 Estimating the Probability of a Forecast

Forecasts are often expressed as single values, so called point estimations, which include no information about their uncertainty. However, it is more informative to pass additionally a probability of the estimation value or a range of most probable future values, called prediction interval, or even a probability distribution. Since the real future value is unknown, probabilities can only be estimated, [Chatfield, 2001]. Different approaches are presented, e.g., by [Chatfield, 2001] and [Kumar and Srivastava, 2012].

However, here the simple approach is used to assume normally distributed errors of the prediction, independent of the observed behaviour of the residuals resulting from forecasting the training and test data. The mean value μ of the residuals is assumed to be zero; the variance of the distribution is estimated based on the data used for the training of the model used for prediction. Based on this assumption a discrete probability distribution of the forecasts is estimated. Randomly created fictive errors e_i are added to the estimated values,

$$\tilde{y}_{k,i} = \hat{y}_k + e_i,\ i = 1, \ldots, \mathrm{N_s}, \tag{3.35}$$

where N_s defines the number of estimated values for the probability distribution. In case of multiple step forecasts, the predictions with added noise are used for the prediction of the next time step.

3.5 Summary

This chapter presents the regression techniques that are used in this work for time series based forecasts of solar irradiance and wind speed. It is also presented how multiple step forecasts can be done. How these techniques are implemented is explained in the next chapter.

Additionally it was shortly discussed how to forecast the probability of a prediction. The estimated probability is used for the case study, as explained in Chapter 6.

Chapter 4

Implementation

Building on the previous chapter, the present chapter explains how forecasting using the different regression techniques is implemented. Additionally it is presented how the performance of the forecasts is evaluated.

4.1 Implementation of Regression Techniques

External libraries are used for the implementation of all regression techniques. In the following, for each technique is explained where the original functions come from and how they are used.

ARIMA Models

The Econometrics Toolbox provided by MATLAB is used to fit ARIMA Models to given training data and for forecasts using ARIMA Models.

Kernel and k Nearest Neighbors Regression

The public Statistical Learning Toolbox for MATLAB, provided by Dahua Lin includes a function to find the nearest neighbors, for k Nearest Neighbors Regressiond and Kernel Regression.

Using Kernel Regression, caused the problem that, depending on the chosen threshold for the distance to neighbors, no neighbor could be found. For this reason, Kernel Regression was implemented using k Nearest Neighbor as backup, in case no neighbors within the chosen distance could be found. The parameter k is therefore also used as an input of the, for the present work, implemented version of Kernel Regression.

Support Vector Regression

For Support Vector Regression, the libary LibSVM is used, [Chang and Lin, 2011]. LibSVM provides a MATLAB interface, but does not depend on MATLAB. Functions for Support Vector Classification and Regression are provided.

4.2 Data Used for Training and Testing.

To train and test the models, a data set from the Azores Island including wind speed and solar irradiance data, provided by the ARM Climate Research Facility, is used. The data provides data points of every minute from the 01. June 2009 until the 31. December 2010. In the present work, mean values for every ten minutes for

- downwelling shortwave hemispheric irradiance,
- artihmetic average wind speed

are used.

Due to quality issues, like values out of range or missing data, only part of the data with little quality issues is used. Three months are used for training and one month for testing,

- Data Training 01. April - 30. June 2010,
- Data Testing 01. July - 31. July 2010.

Using the Solar Position Algorithm (SPA) from the National Renewable Energy Laboratory, [Reda and Andreas, 2008], and a corresponding MATLAB Interface `mspa`[1], provided by Anders Lennartsson, the zenith angel is estimated. From the zenith angle, the extraterrestrial irradiance (see Section 2.2) and therefore the transmissivity can be obtained. Data made available by the National Renewable Energy Laboratory[2], recorded at a station in Hefei, China, that includes the zenith angle were used to verify the results.

To the end, data of the three variables

- solar irradiance,
- transmissivity,
- wind speed

is available. Additionally, copies of the data for solar irradiance and the transmissivity, where night data is removed, are created.

[1] `http://www.mathworks.com/matlabcentral/fileexchange/32507-mspa`
[2] http://www.nrel.gov/

4.3 Search for Models with Highest Forecast Accuracy

For each regression technique and variable a parameter space, a parameter space to search for the best model, is defined. An exhaustive search is conducted to obtain a trained model respectively a training sample and the corresponding root mean square error when forecasting the testing data.

ARIMA Model

For the best ARIMA model is searched in the parameter space presented in table 4.1. Training with only 100 respectively 1000 data points were included, based on the results of grid searches applied for Kernel, kNN and SV Regression. Originally the degree of differencing 2 was also included. But due to over all worse results, than for degree of differencing 0 and 1, it was excluded from final searches.

Autoregressive lags	0-20,
Moving average lags	0-20
Degree of differencing	0, 1
Data set training	night data kept or removed
	100/1000 data points for solar irradiance/wind speed or whole data set
Threshold t-statistic	1.96, 10^{-2}

Table 4.1: Parameter space to search for ARIMA models.

It is also searched for the seasonal ARIMA model that works best on the test data. The parameter space in table 4.1 is extended to the parameter space shown in 4.2.

Autoregressive lags	0-5,
Seasonal autoregressive lags	0-2,
Moving average lags	0-5,
Seasonal moving average lags	0-2,
Degree of differencing	0, 1,
Degree of seasonal differencing	0, 1
Threshold t-statistic	1.96, 10^{-2}

Table 4.2: Parameter space to search for seasonal ARIMA models with highest forecast accuracy. The length of one season is 144.

A flow chart of the performed grid search can be seen in figure 4.1. Firstly, during the training phase, all predefined ARIMA models are fitted to the data and evaluated (auto-correlation of residuals, normally distributed residuals, AIC). Secondly, the models that adapted best to the training data are passed. Eventually insignificant lags are removed and the models are trained again. However, it resulted that the preselection of just a few models that could present well the training data, favoured overfitted models, which did

not perform well for forecasting the test data. For this reason, preselection were skipped. During the following testing phase, forecasts of the test data are performed, and the results are saved to use them for model selection.

Kernel and k Nearest Neighbors Regression

For the best combination of parameters for Kernel and k Nearest Neighbors Regression is searched in the parameter space shown in table 4.3 respectively 4.4. In figure 4.2 a

k	1-20,
Autoregressive lags	1-20,
Number of feature vectors	50, 100, 500, 1000, 2000, whole data set,
Weights	uniform, distance.

Table 4.3: Parameter space for k Nearest Neighbors Regression.

ε	0.01, 0.05, 0.1, 0.5
k	1-3,
Autoregressive lags	1-20,
Number of feature vectors	50, 100, 500, 1000, 2000, whole data set,
Weights	uniform, distance.

Table 4.4: Parameter space for Kernel Regression. The threshold for the distance ε is multiplicated with 13.4 (maximal value in training data) for wind speed.

flowchart of the performed grid searches is shown. The training phase can be omitted, since no model needs to be trained.

Support Vector Regression

The parameter space used for the search for the Support Vector Regression model that achieves the best forecast, can be seen in table 4.5. The chosen values for ε depend on the chosen variable, since ε functions on a threshold when to accept errors of the model (see section 3.2) and should, therefore, depend on the range of values of the considered time series. For the grid search for the best combination of SVR, some combinations are omitted to decrease running time.

The flowchart in figure 4.3 visualizes the performed grid search. In the training phase, all models are trained. In the test phase, the models are used to predict the test data.

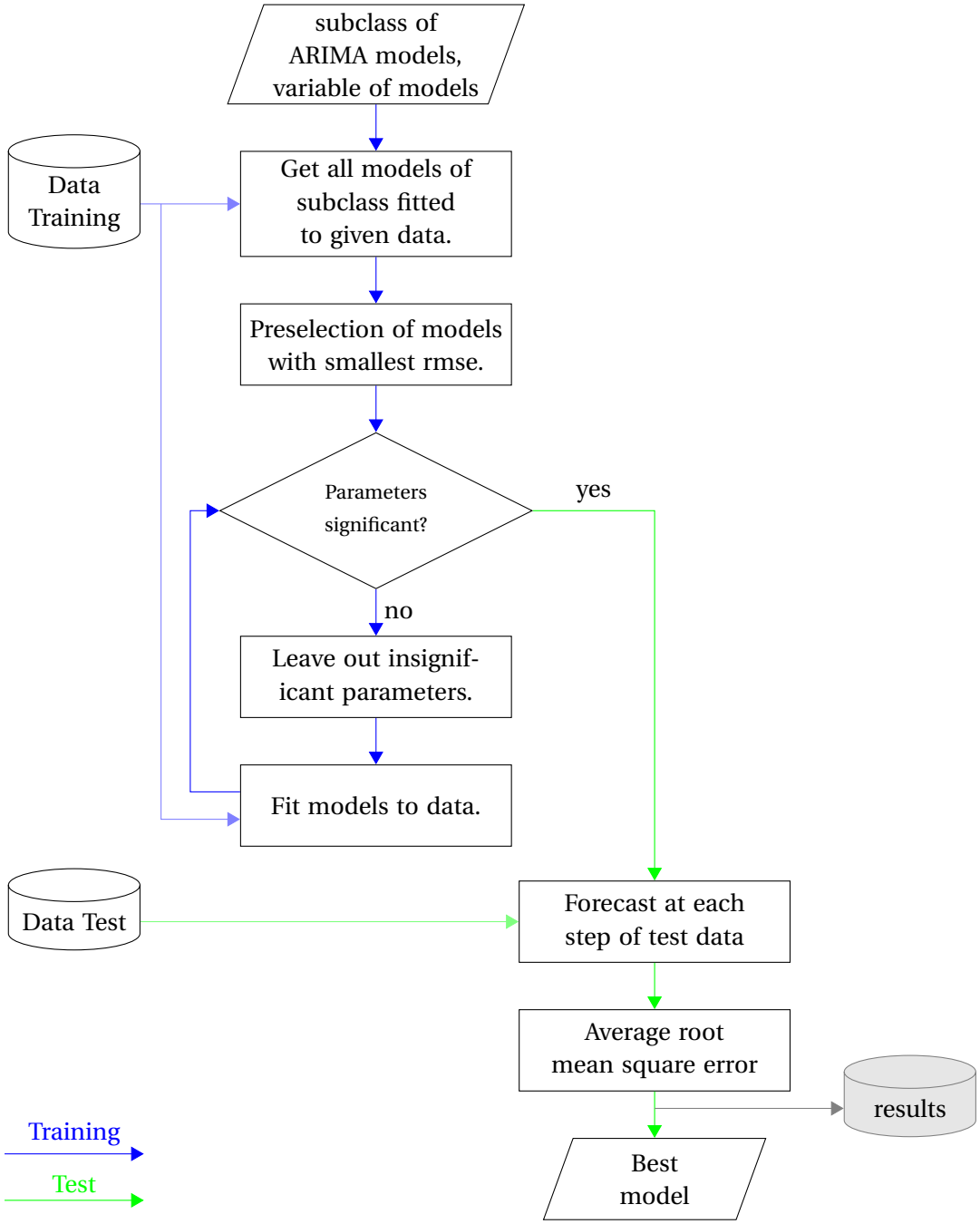


Figure 4.1: Flow chart of grid search for ARIMA models.

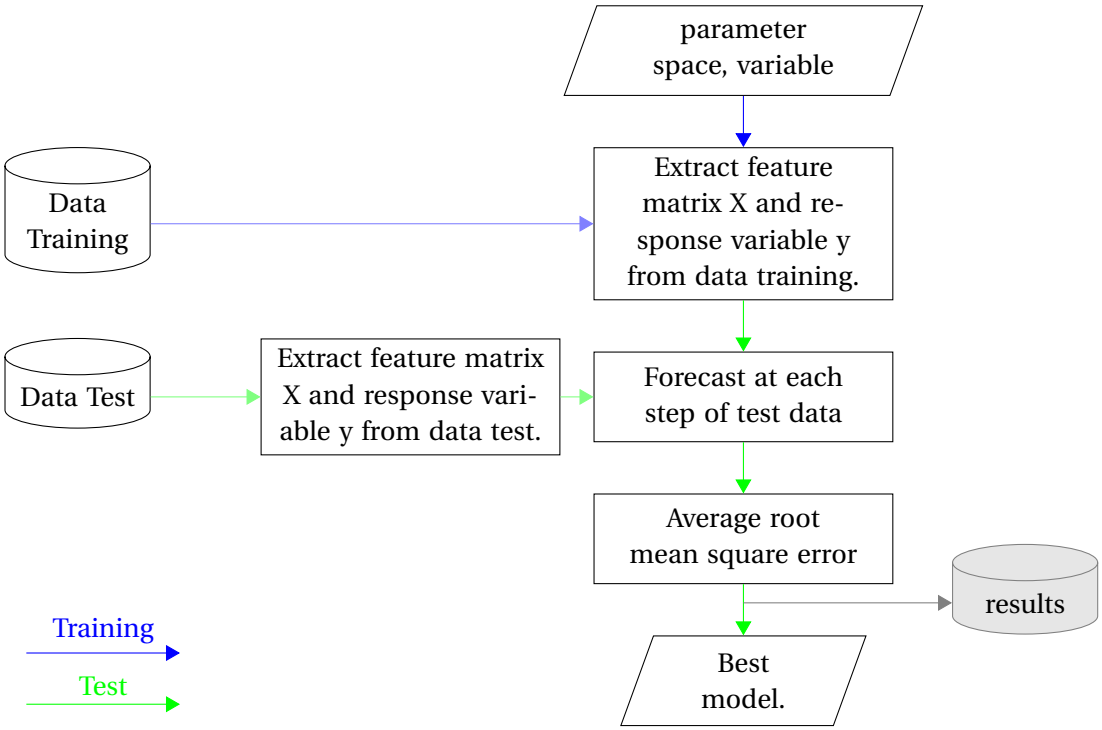


Figure 4.2: Flow chart of grid search for k Nearest Neighbors and Kernel Regression.

Autoregressive lags	1, 3, 5, 7, 9, 11, 13, 15, 17, 19
Number of feature vectors	100, 1000, whole data set,
ε	0.001, 0.01, 0.1, 0.3
C	1, 5, 10
Kernel function	linear, radial basis, sigmoid.

Table 4.5: Parameter space for grid search for SVR using the transmittivity (maximum = 1). For solar irradiance, the values for ε are multiplicated with 1300 (approximates the solar constant). For wind speed, the values for ε are multiplicated with 13.4 (the maximum wind speed in the training data).

4.4 Performance Evaluation

Root Mean Square Error for Multiple Steps Forecasts

To evaluate the performance of a chosen model, the residuals of a forecast performed by the model are used. The residuals e are defined as the difference between the predicted values $\hat{y}$ and the measured values y,

$$e_k = \hat{y}_k - y_k, \tag{4.1}$$

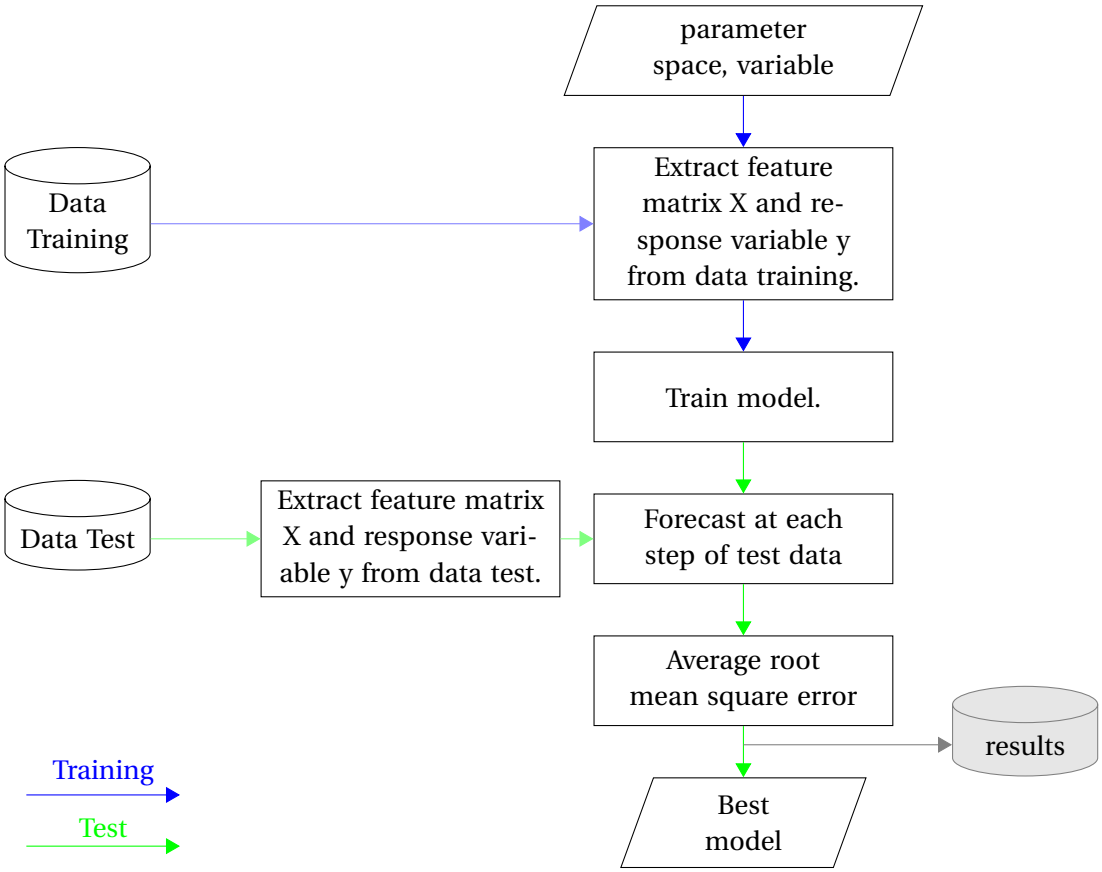


Figure 4.3: Flow chart of grid search for Support Vector Regression.

where k refers to the time instant. An often used measure for the distance between the measured and predicted time series is the root mean square error,

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}} \sum_{k=1}^{\mathrm{N}} \left(\hat{y}_k - y_k\right)}. \tag{4.2}$$

The symbol N refers in general to the number of samples of a time series. Here, N refers to the number of residuals.

Typically a time series with $\mathrm{N} \gg \mathrm{N}_{\mathrm{min}}$, where $\mathrm{N}_{\mathrm{min}}$ is the number of historical data points needed to predict future values, is used to evaluate performance. This means, that more than one multiple step forecast is done (see, e.g., figure 4.4) and each data point y_k is predicted more than once. Therefore, various residuals $e_{k|k-1}, \ldots, e_{k|k-p}$ are obtained, where the first index indicates the time instant of the predictions, the second index is used to define the last known historic value and p refers to the prediction horizon. As a

consequence, it isn't straight forward to apply the root mean square error (see equation (4.2)) to the result of multiple step forecasts. Three different adapted formulas are shown in the following.

Let n_f be the number of forecast that can be done for the given time series, f the index of the forecast, and j the steps ahead the last known value. The first approach is to calculate the RMSE of each forecast separately and to take the mean of them,

$$\text{RMSE1} = \frac{1}{N_f} \sum_{f=1}^{N_f} \underbrace{\sqrt{\frac{1}{p} \sum_{j=1}^{p} e^2_{j+\text{N}_{\min}+f|\text{N}_{\min}+f}}}_{\text{RMSE of forecast } f.} . \tag{4.3}$$

With the second formula, the mean of the RMSE for all predictions at the same prediction horizon is calculated,

$$\text{RMSE2} = \frac{1}{p} \sum_{j=1}^{p} \underbrace{\sqrt{\frac{1}{N_f} \sum_{f=1}^{N_f} e^2_{j+\text{N}_{\min}+f|\text{N}_{\min}+f}}}_{\text{RMSE at prediction horizon } j} . \tag{4.4}$$

To avoid taking the mean before calculating the square root, the RMSE over all the predictions can be used,

$$\text{RMSE3} = \frac{1}{p \cdot N_f} \sqrt{\sum_{j=1}^{p} \sum_{f=1}^{N_f} e^2_{j+\text{N}_{\min}+f|\text{N}_{\min}+f}} . \tag{4.5}$$

Root Mean Square Error for Predictions of Solar Irradiance

In case of predicting the solar irradiance, the matter gets slightly more complicated. In contrast to wind speed, it is already known that the solar irradiance during night is zero. Therefore, it is of interest to predict only part of the time series. Hence, it would not make sense to evaluate the model based partly on predictions the model is not used for.

For one step predictions, the procedure to omit data is straight forward. For multiple step forecasts, on the other hand, it is to decide how to deal with forecasts including daylight and night data. One option is to exclude each multiple step forecast that is not purely about daylight data. See, e.g., figure 4.4, forecast (b)-(d) would be included in the evaluation of the model. Another option is to include each multiple step forecast that includes at least one prediction during daylight (compare forecast (a) and (d) in figure 4.4). In this work the second option is used to avoid putting more emphasis on midday, than on the morning and evening of the day, for the model selection. Therefore

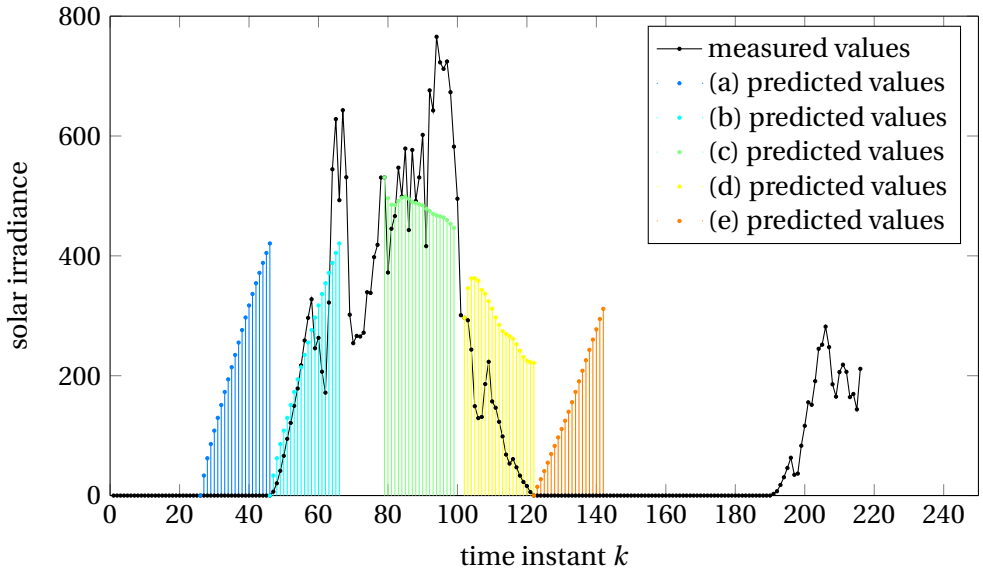


Figure 4.4: Five forecasts (a)-(e), with prediction horizon 20, of solar irradiance, with sample time = 10 min, beginning at different time steps. It is used an ARIMA(20,0,1) model, fitted to solar irradiance data without night data. Forecast (a) includes the calculated Sunrise (considered daylight), likewise forecast (e) includes the calculated Sunset (considered daylight). Forecast (b),(c) and (d) include only predictions of daylight data.

all predictions for solar irradiance during night are automatically set to zero,

$$\hat{y}_k \begin{cases} \text{output of model} & \text{if } \text{ss} \le k \le \text{sr}, \\ 0 & \text{if } k \le \text{sr},\ k \ge \text{ss}. \end{cases} \tag{4.6}$$

The Sunset is indicated by the time instant ss, likewise the Sunrise by sr. It is always referred to the Sunset and Sunrise at the same day as the time instant k.

Applying the first proposed RMSE to the forecasts of solar irradiance, as explained above in 4.3, results in

$$\begin{aligned} \text{RMSE1} = \sqrt{\frac{1}{p}} \Bigg(& e_{\text{ss}|\text{ss}-p} + \sqrt{e^2_{\text{ss}+1|\text{ss}-p+1} + e^2_{\text{ss}|\text{ss}-p+1}} + \cdots + \sqrt{\sum_{j=1}^{p} e^2_{\text{ss}-1+j|\text{ss}-1}} \\ & + \frac{1}{n_{f,day}} \sum_{f=1}^{n_{f,day}} \sqrt{\sum_{j=1}^{p} e^2_{\text{ss}+f+j\,|\text{ss}+f}} + \sqrt{\sum_{j=1}^{p-1} e^2_{\text{sr}-j\,|\text{sr}-p+1}} + \cdots \\ & + \sqrt{e^2_{\text{sr}-1|\text{sr}-2} + e^2_{\text{sr}|\text{sr}-2}} + e_{\text{sr}|\text{sr}-1} \Bigg). \end{aligned}$$

As one can see, the residuals have different influences on the error measurement, as of, e.g., the first predicted daylight data point $\hat{y}_{\mathrm{ss|ss}-p}$. The other two presented variants of the RMSE (see equation (4.7), (4.5)) for multiple step forecasts are not effected by omitting night data. Since it is also of interest to get information about the forecast accuracy depending on the prediction step, RMSE2 is used respectively the particular errors for each prediction step,

$$\mathrm{RMSE}(j) = \sqrt{\frac{1}{N_f}\sum_{f=1}^{N_f} e^2_{j+\mathrm{N_{min}}+f|\mathrm{N_{min}}+f}}, j = 1,..,p \tag{4.7}$$

Comparison of Performance of Different Models

The quality of the forecast obtained by different models is evaluated by comparing the RMSE2 to a naive forecast. Using a naive model means to assume, that the next value equals the current one,

$$\hat{y}_{k+1} = y_k. \tag{4.8}$$

For comparison the distance between the naive forecast and the forecast performed by the chosen model,

$$d = \sum_{j=1}^{p} \mathrm{RMSE}(j)_{\mathrm{naive}} - \mathrm{RMSE}(j), \tag{4.9}$$

is used. In the following chapter, the model with the highest distance d is considered the best model.

4.5 Summary

All regression techniques are implemented using external libraries. Due to practical issues with Kernel Regression, the implemented technique is slightly modified, using k Nearest Neighbour Regression as back up.

The data used here, provides wind speed and solar irradiance data. The transmissivity is calculated from the given data of solar irradiance and the estimated data for the extraterrestrial irradiance, using a solar position algorithm.

A grid search for each technique is performed to chose the model that achieves the most accurate forecasts. Performance of the forecasts provided by each of the models is evaluated comparing the resulting root mean square error to a naive forecast.

In the next chapter, the chosen models are presented. The performance of the chosen models is evaluated and analysed.

Chapter 5

Results and Analysis

In this chapter, results of the grid search (see Section 4.3) are presented. Multiple step forecasts are performed with chosen models and evaluated as explained in Section 4.4.

5.1 Results for Forecasting Solar Irradiance

For each technique and variable, the model that predicts most accurate the test data, in comparison to a naive forecast of transmittivity, is selected (see (4.9)). The naive forecast of transmittivity outperforms the naive forecast using the irradiance directly (see figure 5.1), therefore the naive forecast of transmittivity is chosen as reference.

Forecasting Solar Irradiance Using Solar Irradiance as Input

For all the in this work considered techniques, forecasting the solar irradiance, using the solar irradiance directly as input, yields worse results than a naive forecast using the transmittivity. Results are shown in figure 5.2. Therefore, in the following only the results for forecasting solar irradiance using the transmittivity are discussed in detail.

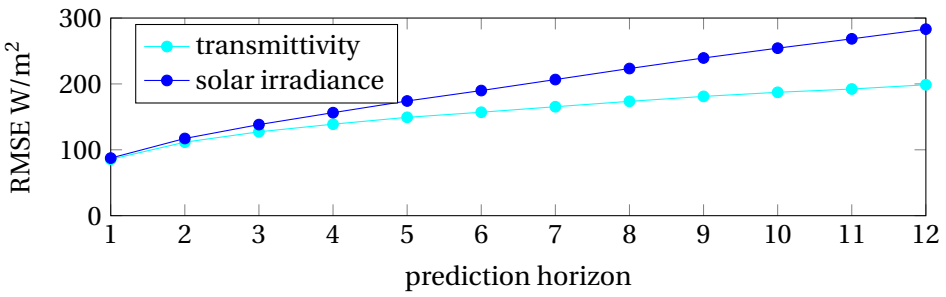


Figure 5.1: Performance of naive forecast predicting solar irradiance using the solar irradiance directly and using the transmittivity.

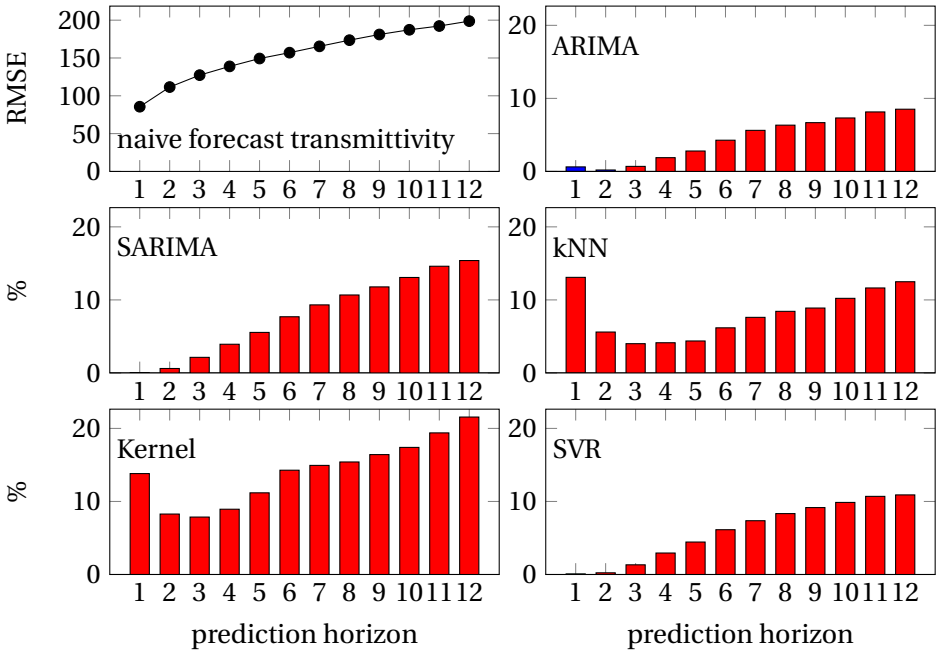


Figure 5.2: Performance of models forecasting the solar irradiance. For each technique the model with lowest RMSE (in comparison to a naive forecast using the transmittivity for predictions up to 2 h/12 prediction steps) is chosen. The red filled bars indicate that the forecast for the corresponding model and prediction step is worse than the naive forecast. Here, only an ARIMA model achieves just for the first prediction step a slightly higher accuracy (0.6%) than the naive forecast.

Forecasting Solar Irradiance Using Transmittivity as Input

In the following, it is distinguished between accuracy achieved for the following prediction steps

- 10 - 120 min, 12 prediction steps,
- 10 - 30 min, 3 prediction steps,
- 100 - 120 min, 3 prediction steps.

For each prediction range, the model is selected that predicts most accurate the considered prediction steps. The selected models are described in table 5.1. The performance of the particular models is visualized in figure 5.3, 5.4(a) and 5.4(b). Most importantly, the results show that Support Vector Regression, independently of the considered prediction range, achieves better results when forecasting the test data than the other considered techniques. For prediction step 12 (120 min ahead), the improvement is 20,90% in comparison to a naive forecast. However, for predictions 10 min ahead, the improvement is only 2,35%.

technique	data training	model specifications			distance to naive forecast
prediction steps 1 - 12			**RMSE 1 step**	**RMSE 12 steps**	
Naive			85.56	198.72	
ARIMA	WND, three months	AR = 0, MA = 10, d = 1	84.09	164.16	211.75
SARIMA	three months	AR = 3, SAR = 2, MA = 0, SMA = 0, d = 0, D = 0	83.55	193.84	59.41
kNN	WND, 100	AR = 2, k = 11, weight = distance	93.97	163.47	192.98
Kernel	WND, 100	AR = 2, ϵ = 0.1, weight = distance	94.36	165.28	185.32
SVR	100	AR = 7, ε = 0.1, radial basis kernel, C = 10	83.93	157.18	249.12
prediction steps 1 - 3			**RMSE 1 step**	**RMSE 3 steps**	
Naive			85.56	127.31	
ARIMA	WND, three months	AR = 4, MA = 4, d = 0, t-statistic $\geq$ 1.96	82.88	119.78	15.73
SARIMA	three months	AR = 3, SAR = 2, MA = 0, SMA = 0, d = 0, D = 0	83.55	122.28	11.13
kNN	2000	AR = 3, k = 20, weight = uni	84.29	119.05	14.74
Kernel	three months	AR = 3, ϵ = 0.1, weight = uni	84.32	119.94	13.98
SVR	100	AR = 7, ε = 0.1, radial basis kernel, C = 5	83.55	118.53	16.89
prediction steps 10 - 12			**RMSE 10 steps**	**RMSE 12 steps**	
Naive			187.28	198.72	
ARIMA	WND, 100	AR = 17, MA = 0, d = 1	157.09	162.11	97.98
SARIMA	three months	AR = 2, SAR = 2, MA = 4, SMA = 2, d = 1, D = 0	181.14	193.03	17.30
kNN	WND, 100	AR = 2, k = 11, weight = distance	157.86	163.47	96.88
Kernel	WND, 100	AR = 2, ϵ = 0.1, weight = distance	159.93	165.28	91.53
SVR	100	AR = 7, ε = 0.1, radial basis kernel, C = 10	152.75	157.18	113.4

Table 5.1: Models that predict most accurate the solar irradiance test data using the transmittivity. Models are chosen based on the distance between the model's error and the naive model's error forecasting the test data. The abbreviation WND stands for training without night data.

Independently of the technique, one can see, that there is only small improvement for predictions up to 30 min ahead in comparison to a naive forecast. Significantly better results are achieved for higher prediction horizons.

Another interesting results is, that often the models that predict most accurate the data, were trained with only 100 data points (see table 5.1). 100 data points include less than 17 h of data. Other models were trained with data covering three months. Since the 100 data points used for training, are the most recent data points, this result indicates that using less, but recent data, may yield to better results than using a large data set for training.

While the Kolmogorv-Smirnov test for all selected ARIMA models is not passed (not even at a 1% significance level), the ARIMA model chosen for forecasts from 10-120 min achieves a p-value for the Ljung-Box Test of 19.05%. However, the chosen ARIMA models for the first and last prediction steps do not achieve p-Values for the Ljung-Box Test of less than 5%.

5.2 Results for Forecasting Wind Speed

In contrast to solar irradiance forecasting, the improvement of wind speed forecasting using ARIMA Models, k Nearest Neighbour, Kernel or Support Vector Regression, in comparison to a naive forecast, is small. Results are visualized in figure 5.5, 5.6(a) and 5.6(b). Selected models are shown in table 5.2. As can be recognized considering all results, the different used regression techniques yield to similar forecasting accuracy. No technique obviously achieves best results for forecasting the wind speed test data. While a simple ARIMA model achieves best results for predictions up to 30 min, best results for predictions from 100 min to 120 min are obtained by Kernel Regression using the 5 last time steps. For predictions from 10 min up to 120 min, linear Support Vector Regression forecasts the data most accurate.

Improvement of accuracy relative to a naive forecast might be low in contrast to forecasting solar irradiance, because no additional information than historical data of wind speed itself is provided. Relatively good results for forecasting solar irradiance are provided if the transmittivity is used[1]. However, results for predictions 2 h ahead can be improved by 7.3%.

In contrast to the results of forecasting solar irradiance, here in each case, models trained using a t-statistic threshold of 1.96 perform best. Anyway, neither the p-value for the Ljung-Box Testnor for the Kolomogrov-Smirnov Test is significantly different from zero.

[1] The transmittivity already includes location and time specific information about the solar irradiance.

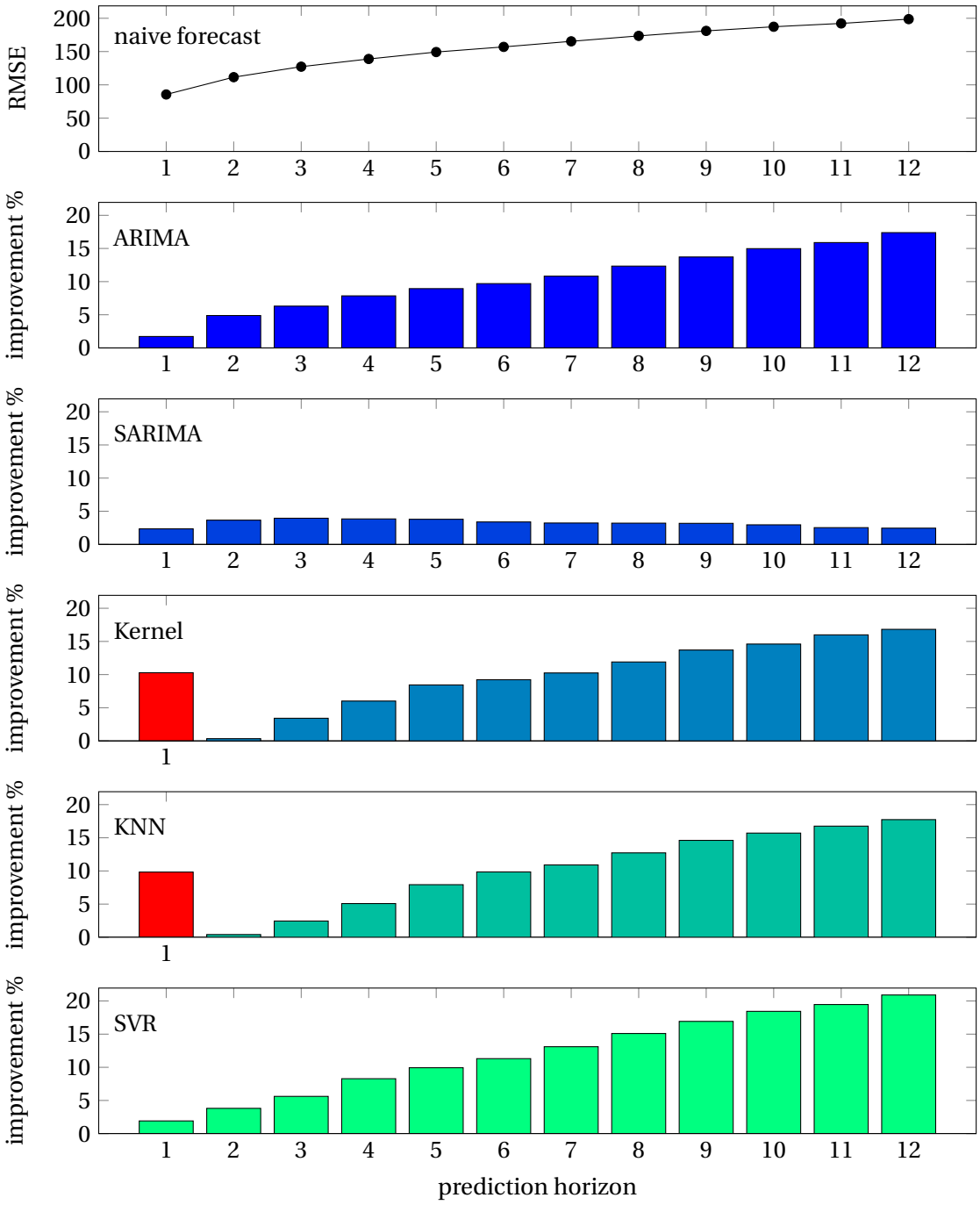
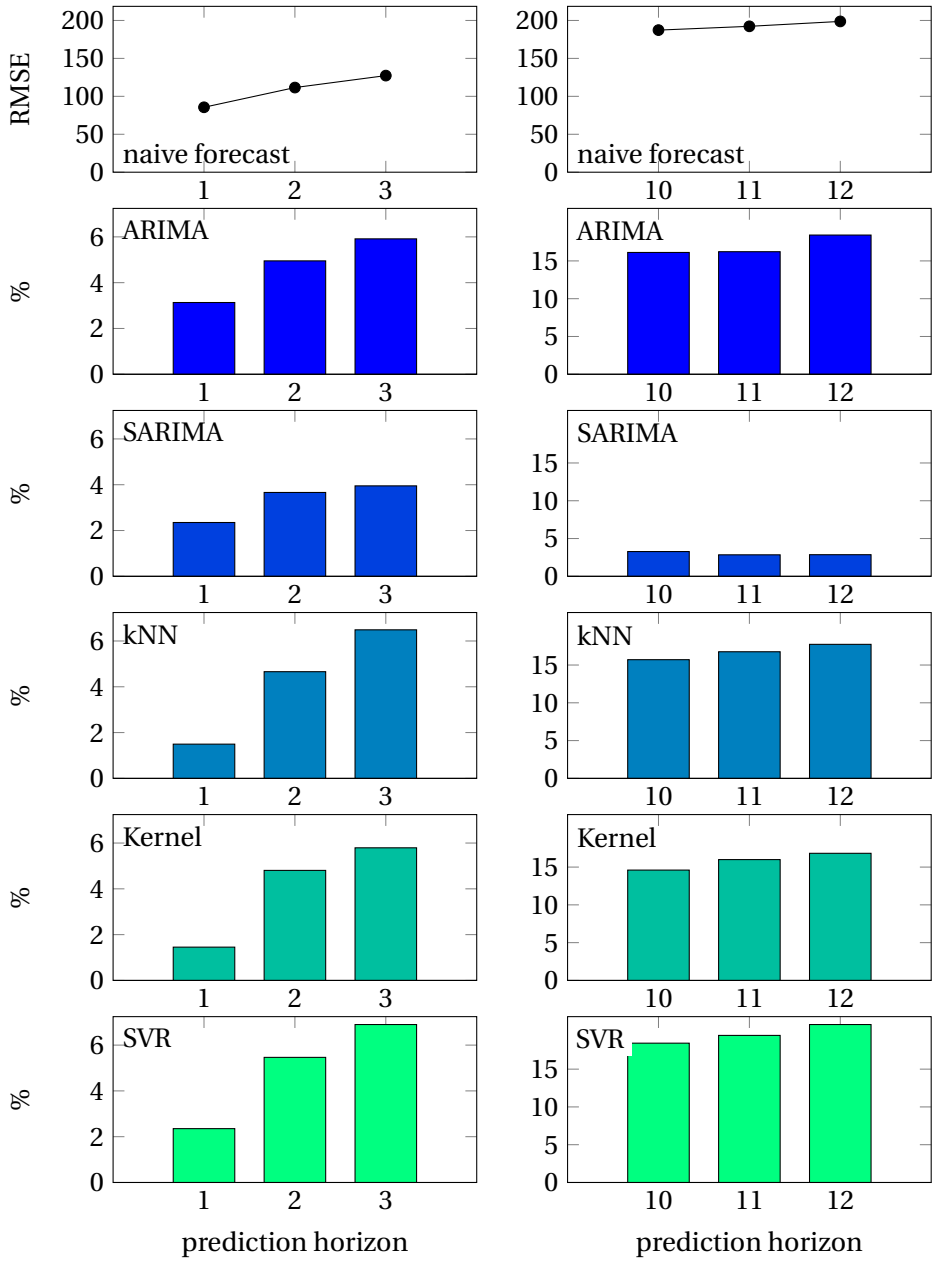


Figure 5.3: Performance of models forecasting the solar irradiance test data. For each technique the model with lowest RMSE (in comparison to a naive forecast using transmittivity for predictions up to 2 h/12 prediction steps) is chosen. The red filled bars indicate that the forecast for the corresponding model and prediction step is worse than the naive forecast. Here, SVR performs best, followed by ARIMA and Kernel.



(a) Prediction step 1-3.

(b) Prediction step 10-12.

Figure 5.4: Performance of models forecasting the solar irradiance test data using the transmittivity. For each technique the model with lowest RMSE (in comparison to a naive forecast) when forecasting the first free prediction steps / first 30 min (a) respectively the last three prediction steps / 100 - 120 min (b) is chosen. The red filled bars indicate that the forecast for the corresponding model and prediction step is worse than the naive forecast. Here, ARIMA performs best for predictions of the first three steps, followed by SVR. For forecasts of prediction step 10, 11, 12, Kernel Regression performs best, followed by SVR.

5.3 Summary

The regression techniques, yielding relatively accurate forecasts, are ARIMA models, Kernel Regression and Support Vector Regression. Overall, Support Vector Regression outperforms the other techniques. In case of forecasting wind speed, the improvement over the naive forecast is between 0-8%. For forecasts of solar irradiance, improvement is roughly in the range from 0 to 20%.

technique	number of data points training	model specifications			distance to naive forecast
prediction steps 1 - 12			**RMSE 1 step**	**RMSE 12 steps**	
Naive			0.46	1.10	
ARIMA	1000	AR = 6, MA = 6, d = 0, t-statistic $\geq$ 1.96	0.45	1.04	0.49
SARIMA	three months	AR = 1, SAR = 0, MA = 1, SMA = 0, d = 1, D = 0	0.46	1.08	0.19
KNN	three months	AR = 7, k = 18, weight = uni	0.48	1.05	0.39
Kernel	100	AR = 2, ϵ = 1.34, weight = distance	0.50	1.02	0.50
SVR	1000	AR = 7, ε = 0.134, linear kernel, C = 1	0.46	1.03	0.53
prediction steps 1 - 3			**RMSE 1 step**	**RMSE 3 steps**	
Naive			0.46	0.72	
ARIMA	1000	AR = 6, MA = 6, d = 0, t-statistic $\geq$ 1.96	0.45	0.69	0.070
SARIMA	three months	AR = 1, SAR = 0, MA = 1, SMA = 0, d = 1, D = 0	0.46	0.71	0.025
KNN	three months	AR = 6, k = 20, weight = uni	0.48	0.69	0.027
Kernel	three months	AR = 6, ϵ = 0.67, weight = uni	0.47	0.70	0.024
SVR	1000	AR = 7, ε = 0.134, linear kernel, C = 1	0.45	0.69	0.068
prediction steps 10 - 12			**RMSE 10 steps**	**RMSE 12 steps**	
Naive			1.03	1.10	
ARIMA	1000	AR = 6, MA = 6, d = 0, t-statistic $\geq$ 1.96	0.98	1.04	0.16
SARIMA	three months	AR = 1, SAR = 0, MA = 1, SMA = 0, d = 1, D = 0	1.01	1.08	0.05
KNN	three months	AR = 7, k = 17, weight = uni	0.98	1.05	0.14
Kernel	three months	AR = 6, ϵ = 1.34, weight = uni	0.96	1.02	0.21
SVR	three months	AR = 3, ε = 0.134, radial basis kernel, C = 10	0.96	1.03	0.20

Table 5.2: Models that predict most accurate the wind speed test data. Models are chosen based on the distance between the model's error and the naive model's error forecasting the test data.

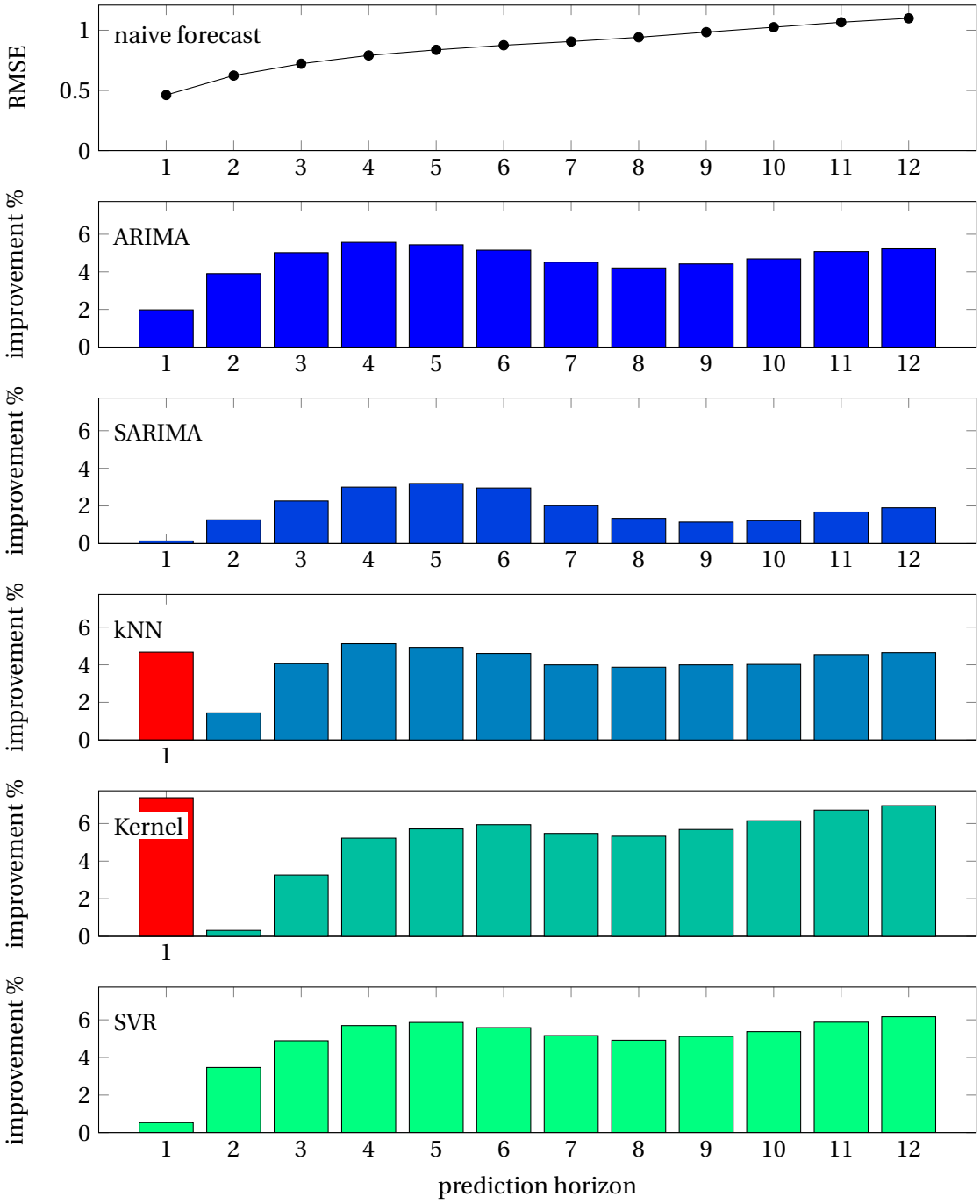
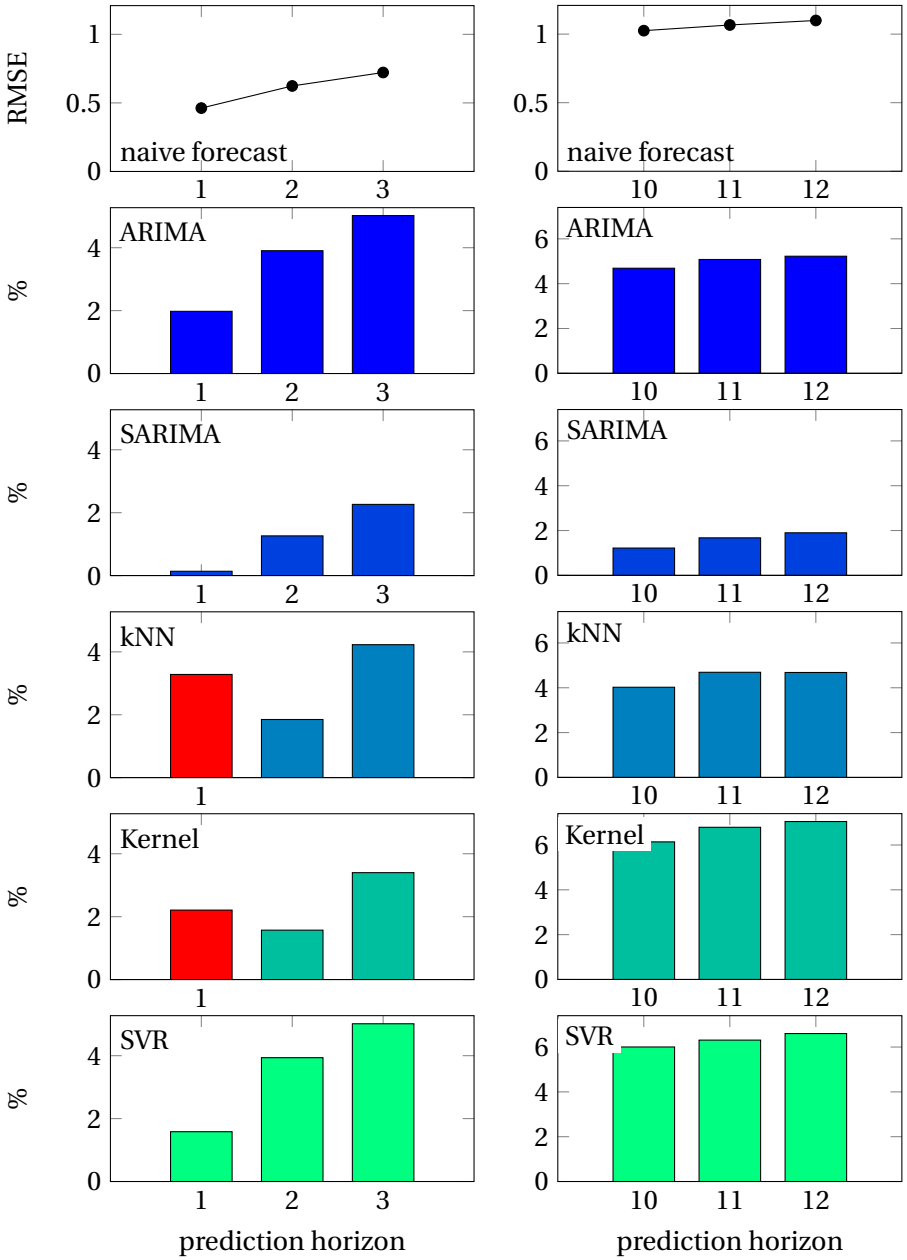


Figure 5.5: Performance of models forecasting the wind speed test data. For each technique the model with lowest RMSE (in comparison to a naive forecast for predictions up to 2 h/12 prediction steps) is chosen. The red filled bars indicate that the forecast for the corresponding model and prediction step is worse than the naive forecast. Here, SVR performs best, followed by Kernel and ARIMA.



(a) Prediction step 1-3.

(b) Prediction step 10-12.

Figure 5.6: Performance of models forecasting the wind speed test data. For each technique the model with lowest RMSE (in comparison to a naive forecast) when forecasting the first free prediction steps / first 30 min (a) respectively the last three prediction steps / 100 - 120 min (b) is chosen. The red filled bars indicate that the forecast for the corresponding model and prediction step is worse than the naive forecast. Here, ARIMA performs best for predictions of the first three steps, followed by SVR. For forecasts of prediction step 10, 11, 12, Kernel Regression performs best, followed by SVR.

Chapter 6

Case Study

Building on a previous performed case study for the operation of a microgrid, using wind power as renewable infeed and ARIMA models to perform the forecasts, [Hans et al., 2015], for the present work, an additional case study, using solar power as renewable infeed and Support Vector Regression to perform the forecasts for solar irradiance, is accomplished. Seven days in July are chosen, conform to the data used in the former sections to evaluate the performance of the regression techniques. For Support Vector Regression, the parameters that yield the most accurate forecasts on test data for 10 min to 2 h are used (see table 5.1). For the load forecast, an ARIMA model, provided by [Hans et al., 2015], is applied. The sampling time is set, as before, to 10 min.

The modelled microgrid includes a conventional generator, solar panels to provide renewable power infeed and a storage device. Its basic structure is shown in figure 6.1.

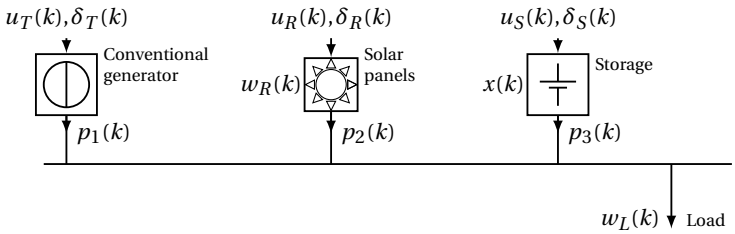


Figure 6.1: Exemplary microgrid, [based on a version by C. Hans.]. The set points are indicated by the symbol u, δ includes the information if the corresponding element of the grid is connected or not (0 or 1), p stands for the power, w for disturbances (load and solar power are considered disturbances) and x is the stored energy.

The general structure of the operation of the microgrid is shown in figure 6.2. Stochastic Model Predictive Control is applied, [Hans et al., 2015]. Therefore, a scenario fan, i.e., several different forecasts for the same time steps, based on the forecasts provided by Support Vector Regression, is created (see section 3.4). An example of a scenario fan,

created with Support Vector Regression, is shown in figure 6.3. It is aimed to provide the demanded power, while minimizing the generator costs (thermal energy and thermal switching), [Hans et al., 2015].

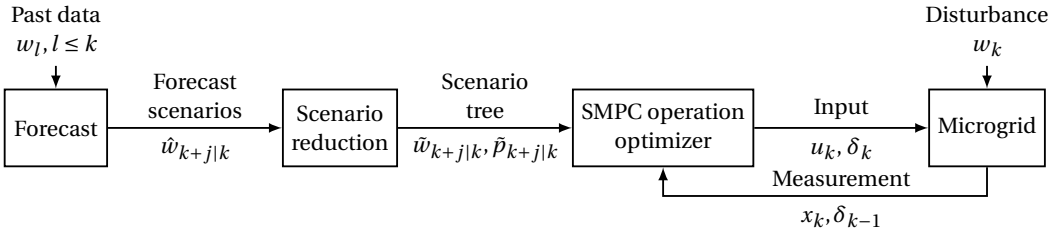


Figure 6.2: Block diagram for a stochastic model predictive control approach, [Hans et al., 2015].

To evaluate the performance, the results are compared with a case study performed under the ideal, not realistic conditions that a perfect forecast is performed. The results of the case study are presented in table 6.1. The simulated values are plotted in figure 6.4. The results show that the gap to the reference with perfect forecasts is small. Even less thermal switching is done. However, in case of perfect knowledge about future values of solar irradiance and load, the energy infeed of renewable energy is increased, while the thermal energy decreases.

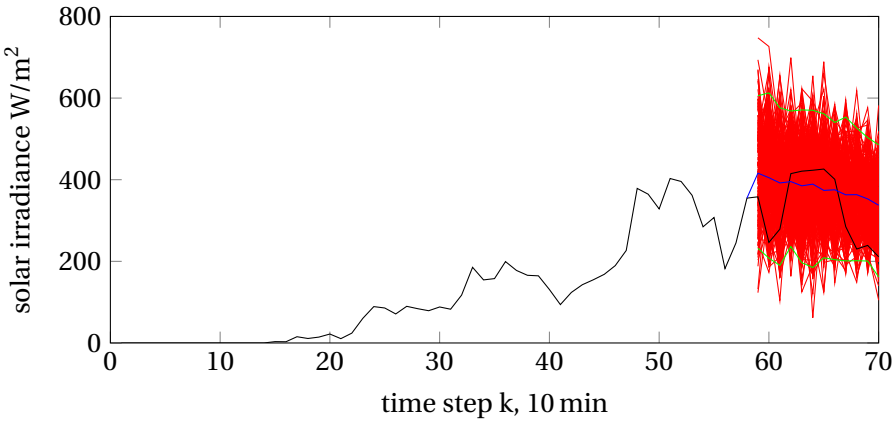


Figure 6.3: Scenario fan (—) for a forecast of solar irradiance on 01 July 2010. Support Vector Regression is used to perform the forecast (—). The 95% prediction interval (—) is estimated based on the scenario fan.

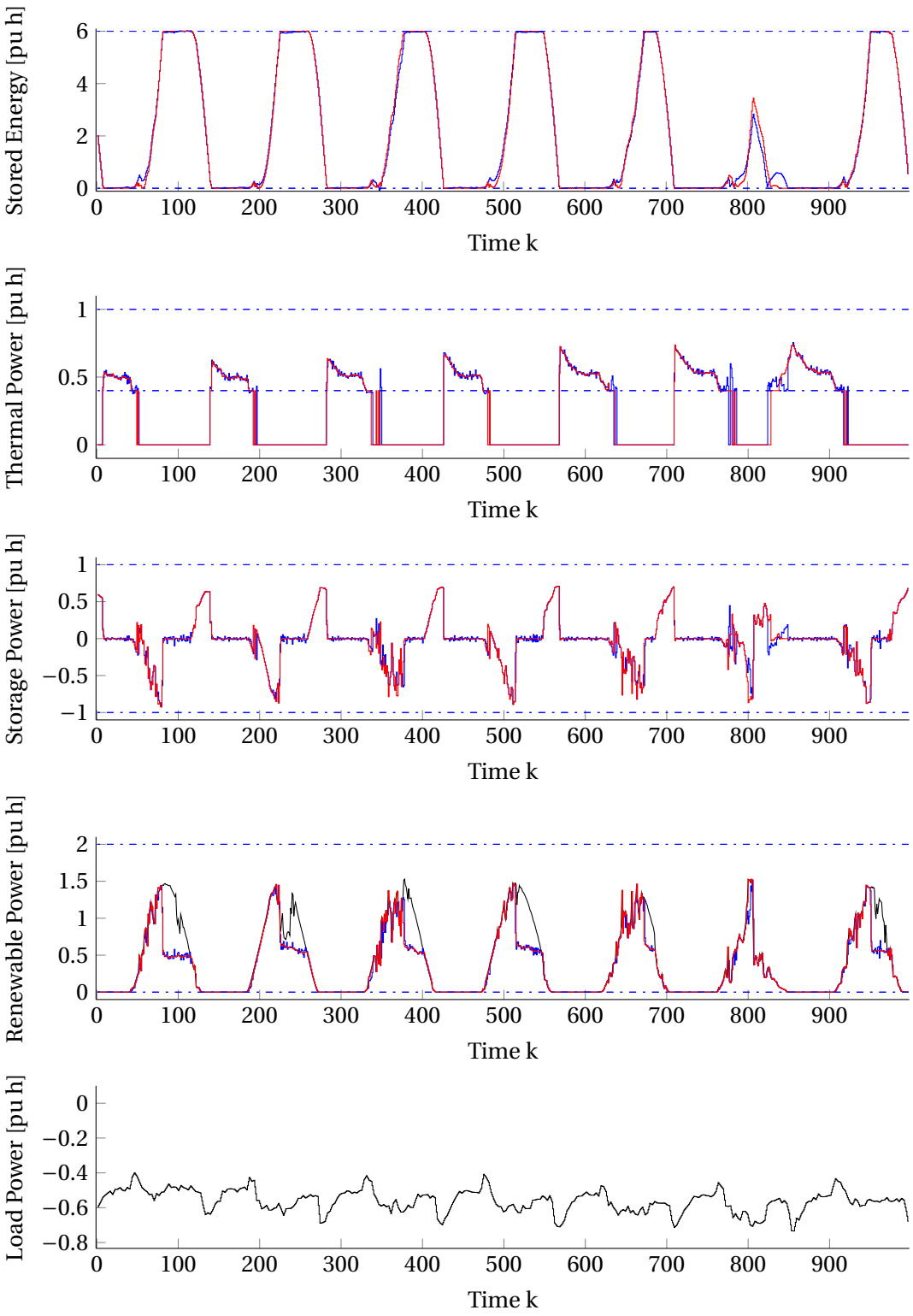


Figure 6.4: Thermal, storage, renewable and stored energy and load from the case study (—) in comparison to an ideal case study (—) with perfect forecasts.

	SMPC	**Perfect Forecast**
Solar energy infeed in puh	56.61	57.81
Thermal energy infeed in puh	39.09	37.81
Thermal switching	26	28

Table 6.1: Results of the simulation of the operation of a microgrid using Support Vector Regression in comparison to the results of a case study where the future values are known.

Chapter 7

Conclusion

Based on the results presented in chapter 5, Support Vector Regression yields, of all considered regression techniques, to most accurate forecasts. For solar irradiance forecasting, SVR outporformes for all considered prediction ranges (10-120 min, 10-30 min, 100-120 min) the other here considered techniques. Although, this is not the case for the wind speed forecasting, SVR here also achieves most accurate results considering predictions from 10 min up to 2 h.

Another interesting conclusion that can be drawn from the results, is that using recent data yields better forecast quality than using a large data set including the same recent data, but additionally less recent data. This indicates, that it might be a mistaken approach to train one model for forecasts of solar irradiance respectively wind speed in general. Preferably, during the operation of the microgrid, recently observed data can be used to search for and train a new model frequently.

For very short term prediction (up to 10 min), a naive forecast already performs well. Little improvement is achieved by applying other techniques. For this reason, for applications using only very short term predictions, it might be sufficient to use a naive forecast, instead of training a more complex model. For predictions with larger horizons, however, applying more complex techniques leads to significant improvement.

The poor performance of SARIMA models, in comparison to all other considered techniques including ARIMA models, leads to the conclusion that using seasonality cannot improve forecasts for solar irradiance nor wind speed. Anyway, here it might be helpful to consider first a larger parameter space to search for most accurate SARIMA models, to come to a definite conclusion. In the present work the last fifty minutes of the current day and the two days before are included.

However, since the result may largely depend on the chosen data and season, results should be validated first using different data sets, measured at different locations and

tested on data throughout the whole year.

To extend the considered regression techniques, it might be interesting to include Artificial Neural Networks. It could be also of interest to consider hybrid models. A study performed on wind speed forecasting achieves better results for a Hybrid model of an ARIMA model and an Artificial Neural Network, than for both of these techniques used alone, [Cadenas and Rivera, 2010].

In the case of forecasting solar irradiance using the transmissivity, instead of the solar irradiance directly, significantly improves the accuracy. This raises the question if the use of other additional variables than solar irradiance respectively wind speed forecasting could improve the forecast accuracy in a similar way. For the forecast of solar irradiance, it could be considered including cloud cover, since other studies are able to improve results by using it, e.g., [Qiao and Zeng, 2012].

Additionally, estimation of a probability distribution of the forecast could be probably improved. Here, it is to analyse how the quality of the probability estimation affects the control of the microgrid.

Bibliography

Altman. An intoduction to kernel and nearest-neighbor nonparametric regression. *The American Statistican*, 1992.

Peder Bacher, Henrik Madsen, and Henrik Aalborg Nielsen. Online short-term solar power forecasting. *Elsevier*, 2009.

Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Ael Le Borgne. *Lecture Notes in Business Information Processing*, chapter Machine Learning Strategies for Time Series Forecasting. Springer, 2013.

George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis - Forecasting and Control*. Prentice- Hall, Inc., 1994.

Erasmo Cadenas and Wilfrido Rivera. Wind speed forecasting in the south coast of oaxaca, mexico. *Renewable Energy, Elsevier*, 2006.

Erasmo Cadenas and Wilfrido Rivera. Wind speed forecasting in three different regions of mexico, using a hybrid arima-ann model. *Renewable Energy, Elsevier*, 2010.

George Casella and Roger L. Berger. *Statistical Inference*. Duxbury, 2001.

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Chris Chatfield. *Principles of Forecasting*, chapter Prediction Intervals for Time-Series Forecasting, pages 475–499. Springer, 2001.

Walter Enders. *Applied Econometric Time Series*. John Wiley & Sons, 1994.

J. H. Friedmann. The role of statistics in the data revolution? *International Statistical Review/Revue Internationale de Statistique, 5-10*, 2001.

Lazlo Gyoerfi, Michael Kohler, Adam Krzyzak, and Harro Walk. *A Distribution Free Theory of Nonparametric Regression*. Springer, 2002.

C. A. Hans, P. Sopasakis, A. Bemporad, J. Raisch, and C. Reincke-Collon. Scenario-based model predictive operation control of islanded microgrid. *2015 54th IEEE Conference on Decision and Control (CDC)*, 2015.

Detlev Heinemann, Elke Lorenz, and Marco Grirodo. Solar irradiance forecasting for the management of solar energy systems. *University Oldenburg*, 2006.

Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2014.

IS/ISO-9488. Solar energy - vocabulary. Technical report, Bureau of Indian Standads - Non-conventional Energy Sources Sectional Committee, 1999.

J.G. Kalbfleisch. *Probability and Statistical Inference*. Springer Verlag, 1979.

Greg Kopp and Judith L. Lean. A new, lower value of total solar irradiance: Evidence and climate significance. *Geophysical Research Letters*, 2011.

Sricharan Kumar and Ashok Srivastava. Bootstrap prediction intervals in non-parametric regression with applications to anomaly detection. *NASA*, 2012.

Lon-Mu Liu and Gregory B. Hudak. *Forecasting and Time Forecast Analysis using the SCA Statistic*. Scientific Computing Associates, 1992.

Frank J. Massey. The kolmogrov-smirnov test for goodness of fit. *American Statistical Association Journal*, 1951.

Adel Mellit and Alessandro Massi Pavan. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected pv plant at trieste, italy. *Solar Energy*, 84(5):807–821, 2010.

Mohammad Monfared, Hasan Rastegar a, and Hossein Madadi Kojabadi. A new strategy for wind speed forecasting using artificial intelligent methods. *Renewable Energy, Elsevier*, 2009.

Wei Qiao and Jianwu Zeng. Short-term solar power prediction using a support vector machine. *EElsevier- Renewable Energy*, 2012.

Volker Quaschning. Technical and economical system comparison of photovoltaic and concentrating solar thermal power systems depending on annual global irradiation. *Elsevier*, 2004.

Ibrahim Reda and Afshin Andreas. Solar position algorithm for solar radiation applications. Technical report, National Renewable Energy Laboratory, 2008.

K. Scharmer and J. Greif. *The European Solar Radiation Atlas*. Le Presses de l'Ecole de Mines, 2000.

A. Sfetos. A novel approach for the forecasting of mean hourly wind speed time series. *Renewable Energy, Elsevier*, 2001.

Alex J. Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 2004.

Arbeitsgruppe Erneuerbare Energien Statistik. *Zeitreihen zur Entwicklung der erneuerbaren Energien in Deutschland*. Bundesministerium für Wirtschaft und Energie, 2016.

W. B. Stine and R. W. Harrigan. *Solar Energy Systems Design*. John Wiley and Sons, Inc., 1986.

Larry Wasserman. *All of Statistics - A Concise Course in Statistical Inference*. Springer, 2005.

Felix Werdermann. Wind aus den segeln genommen. *Der Freitag*, May 2016.

Jeffrey M. Woolridge. *Introductory Econometric*, chapter The Normal and Related Distributions, page 665ff. Cengage Learning, 2015.

Junyi Zhou, Jing Shi, and Gong Li. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Conversion and Managemen, Elsevier*, 2011.