Technische Universität München

Zentrum Mathematik

Time Series Modelling of Wind Speeds with View on Renewable Energies

Diplomarbeit

von

Sarah Maria Davoudabadi-Farahani

Themenstellerin:	Prof. Dr. Claudia Klüppelberg
Betreuer:	PD Dr. Gernot Müller
Abgabetermin:	01. Juni 2011

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Garching, den 01. Juni 2011

Acknowledgments

At this point I would like to thank all those people who supported me while writing the thesis.

First, I would like to thank Prof. Klüppelberg for the interesting topic and the opportunity to work on this thesis in her inspiring group.

Foremost I owe my deepest gratitude to my supervisor Gernot Müller, whose constant support and guidance enabled me to write this thesis.

Moreover, I would like to thank Dr. Florian Steinke for explaining the data set provided by Siemens AG.

I am grateful to the Institute for Energy Economy and Application Technology at Technische Universität München for their support.

Finally, I thank my parents for supporting me not only during writing this thesis, but in all situations throughout my life.

Abstract

Due to the recent events in Japan, the discussion about renewable energies and about the future energy supply in Germany and in Europe is, more than ever, in the focus of politics and science. To judge whether and to which extent wind power will be able to replace other traditional power plants, we need a deeper understanding of the dependencies of wind speeds in time and space. This is, in particular, important to guarantee a constant energy supply without any blackouts. In this thesis we develop a time series model for wind speed data. One major challenge is to find a model which is flexible enough to capture wind speed characteristics for different types of locations, as mountains, plains, and offshore regions. We apply our model to wind speed data from several monitoring stations in Germany, as, for instance, Munich and Helgoland.

Zusammenfassung

Erneuerbare Energien und die zukünftige Energieversorgung in Deutschland und Europa werden derzeit, ausgelöst durch die Ereignisse in Japan, in der Politik und Wissenschaft verstärkt diskutiert. Dabei stellt sich die Frage, in welchem Ausmaß man herkömmliche Kraftwerke durch erneuerbare Energieträger wie Windenergie ersetzen kann ohne die konstante Energieversorgung zu gefährden. Zur Beantwortung benötigt man unter anderem ein tieferes Verständnis von Windgeschwindigkeiten mit deren Abhängigkeiten in Zeit und Raum. In dieser Diplomarbeit entwickeln wir daher ein Zeitreihenmodell für Windgeschwindigkeitsdaten. Eine der großen Herausforderungen ist es, ein Modell zu finden, welches flexibel genug ist, Besonderheiten von Windgeschwindigkeiten an unterschiedlichen Orten wie zum Beispiel auf Bergen, im Flachland, oder in Offshore-Bereichen zu beinhalten. Wir wenden unser Modell auf verschiedene Wetterstationen in Deutschland an, unter anderem München und Helgoland.

Contents

Chapter 1

Introduction

"Wir können nicht so (weitermachen) wie es vorher war. Wir können nicht einfach zur Tagesordnung übergehen."
(We cannot proceed as done before. We cannot carry on as usual.)
Angela Merkel, Chancellor of the Federal Republic of Germany, 24.03.2011[1]

These words were spoken by Angela Merkel in the context of the nuclear catastrophe that began to unfold in the nuclear plants in Fukushima, Japan on March 11, 2011 after the severe Tohoku-earthquake. Reaching a highly sensitive public in Germany, it was clear that these events would lead to discussions with respect to power supply. And indeed, at the time of writing in May 2011, a panel which was established by the German Chancellor, urges the government to close all nuclear plants within the next ten years.

Even if the changes do not turn out as dramatic as suggested by the panel, it is clear that Germany will not only continue, but will also accelerate on its path from fossil fuels to renewable energy sources. Within the last ten years Germany has already seen a more than 100 percent increase in the number of wind energy facilities. *Europe 2020*, a 10-year strategy proposed by the European Commission on March 3, 2010, suggests a share of renewable energy of at least 20% until 2020.
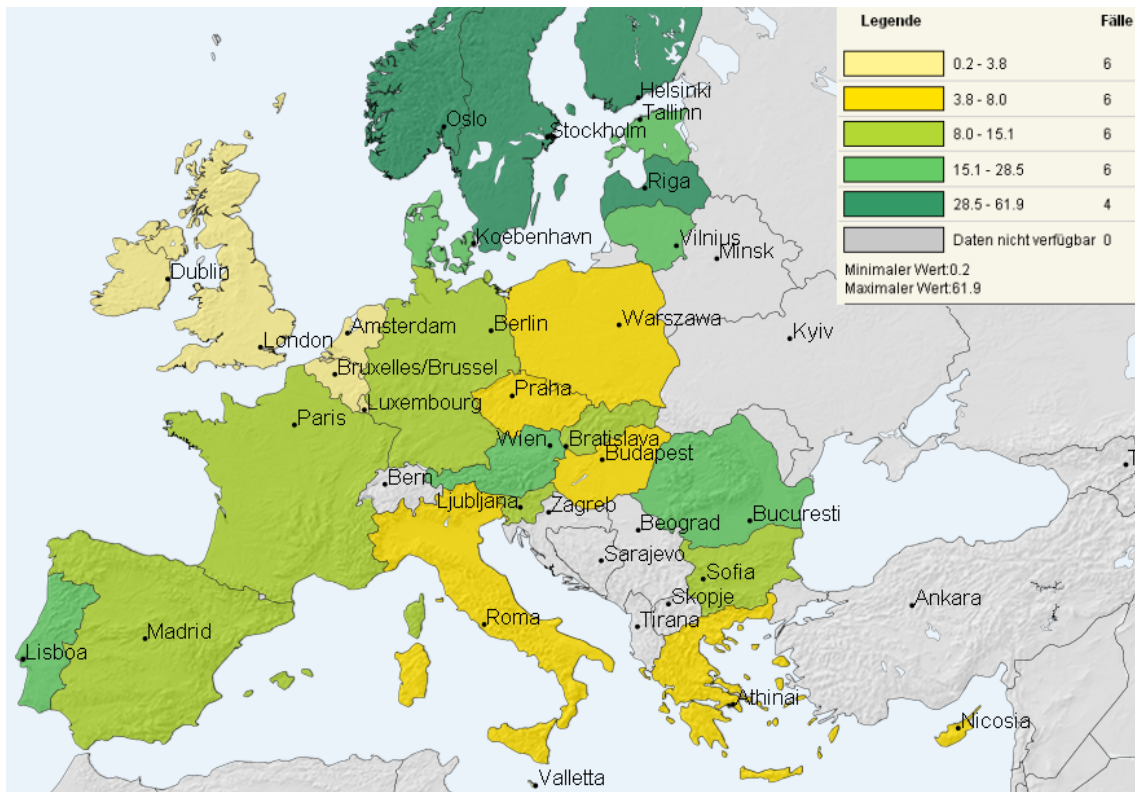
Figure 1.1[2] shows the amount of renewable energies as percentage of total energy produced in the member states of the European Union. Based on its existing renewable energy generation, its GDP and a flat-rate increase, each member state is given a national target to meet. However, it it will be extremely challenging for many member states of the EU to realize the national targets by 2020. The UK, for instance, has been set a target rate of 15%[3], while its current share of renewable energy has a maximum of 3.8% (cf. Figure 1.1). Consequently, renewable energies

---

[1] *http://www.bundeskanzlerin.de/nn_683698/Content/DE/Mitschrift/ Pressekonferenzen/2011/03/2011-03-24-statements-europaeischer-rat.html*
[2] *http://epp.eurostat.ec.europa.eu/tgm/mapToolClosed.do?tab=map&init1&plugin= 1&language=de&pcode=ten00081&toolbox=types*
[3] *http://www.publications.parliament.uk/pa/ld200708/ldselect/ldeucom/175/175.pdf*

Figure 1.1: Proportion of renewable energy in Europe, 2008

such as wind power, solar power, geothermal energy and hydropower will become even more important in science and politics. In order to guarantee a constant energy supply without risking any blackouts, several questions need to be answered.

Some work regarding the optimal mix of renewables, in particular wind and sun power, is presented in Heide et al. [2010]In this paper the requirements of storage for energy should be kept at a minimum level. Therefore following the seasonal load curve, the best point of counterbalancing the seasonal behavior of wind power and solar power, the seasonal optimal mix, is searched. As a result, the seasonal optimal mix for an EU, where 100% of energy production stems from renewable energies, consists of 55% of wind and 45% of solar power generation. For less than 100% renewable scenarios the percentage of wind power generation increases whereas percentage of solar power generation decreases. The required stored energy for all of Europe is reduced by a factor of two for the optimal mix and amounts to 1.5–1.8 times its monthly load.

Reducing requirements of storage for energy and improving the reliability of energy supply not only an optimal mix, but also a enhanced prediction of energy production will help. A regional forecasting for single wind farms based on principle component regression is discussed in von Bremen [2007]. In this paper wind power is considered approximately proportional to squared wind speed, which is additionally multiplied

with the normalized spatial distribution of wind power capacity. Using these wind power weighted squared wind speed maps as input to the principle component analysis, main patterns that explain most variance of wind power generation in Germany are identified. The variability of these weighted squared wind speed maps is expressed by principle components, which determine the weight of each eigenvector for a particular state of the maps. From the European Centre for Medium-Range Weather Forecast (ECMWF) some forecast calculation of the wind speed maps is used to compute the principle components. The total wind power feed-in for Germany is estimated by the principle components and regression coefficients gained from a multivariate linear regression between the first six principle components and the (historic) wind power feed-in, which are recalculated every 15 days. The first four eigenvectors explain about 84% of the observed variance. Another benefit of this wind power forecasting model is the moderate computational time.

As mentioned above renewable energies have huge potential for future development. However, traditional power plants could be replaced, if a constant energy supply can be guaranteed. Wind has become one of the fastest growing sources of renewable energy worldwide. As wind power is the conversion of wind speed into electric energy, wind speed can also be focused directly. In order to judge whether wind power will be able to replace other traditional power plants, it is imperative to understand the dependencies of wind in space and time.

One approach is the regime-switching space-time method used in Gneiting [2006].This paper focuses on modelling wind speed at a 2-hour horizon. The Regime-Switching Space-Time method (RST) identifies two forecast regimes, westerly and easterly ones, and fits a conditional predictive model for each regime. The examined time series of wind speed and wind direction are collected at Vansycle in northeastern Oregon (close to the Stateline wind energy center), Kennewick and Goodnoe Hills in southern Washington about 50 to 150 kms away, which are used to identify regimes as well. Different attempts to fit the model were made. The Regime-Switching Space-Time method without modelling a diurnal component (RST-N and RST-N-CH) uses homoscedastic and heteroscedastic (therefore the CH at the short cut) variance structures, respectively. The RST-D and RST-D-CH techniques fit a diurnal trend component (D), but do so only in the westerly regime. The variance structures are homoscedastic and conditionally heteroscedastic (CH), respectively. Generally no diurnal component is modelled in the easterly regime, because this did not improve the predictive performance. The technique of minimum continuous ranked probability score (CRPS) estimation for estimating predictive distributions is used. In minimum CRPS estimation, the continuous ranked probability score is expressed for the training data as a function of the model parameters, and this function is minimized with respect to the parameter values.

In this thesis we also model wind speeds directly. Since transforming the wind speeds with the adequate wind turbine curves is straightforward, this approach seems to have advantages for finding dependencies and influencing factors. One major chal-

lenge is to find a model which is flexible enough to capture wind speed characteristics for different types of locations as mountains, plains, and offshore regions. On the basis of different monitoring stations all over Germany local differences are observed. Diurnal and annual seasonalities exhibit different patterns for the different monitoring stations. Besides the structure caused by diurnal or annual seasonalities, further patterns are found to be inherent in wind speed. We will use an autoregressive process as well as a component for conditional heteroscedasticity to capture the whole structure of the data. For the error terms we assume an hyperbolic distribution. This distribution turns out to be flexible enough to model the different behavior of wind speed, depending on the location. As we will see from the parameter estimates for the hyperbolic distribution, we can identify different parameter patterns of the wind speeds at monitoring stations on mountains, plains and close to cities.

Chapter 2

Theoretical Basics: Hyperbolic Distribution Function

The hyperbolic distributions form a family of continuous distributions and can be considered as a subclass of the generalized hyperbolic distributions, which were introduced by Barndorff-Nielsen [1977]. Further information about hyperbolic distributions can be found in Eberlein [1995].One application for wind shear is demonstrated in Barndorff-Nielsen et al. [1989].

Definition 2.1 (Hyperbolic distribution)
A random variable Z has an hyperbolic distribution, $Z \sim hyperbol(\alpha, \beta, \delta, \mu)$, if its probability density function is

$$f_{HYP}(x) = \frac{\gamma}{2\alpha\delta K_1(\delta\gamma)} \exp\left(-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)\right) \qquad (2.1)$$

with $\gamma := \sqrt{\alpha^2 - \beta^2}, \qquad 0 < |\beta| < \alpha, \qquad \delta > 0 \quad and \quad x \in \mathbb{R}.$

The function $K_\lambda(x)$ appearing in the hyperbolic density is the modified Bessel function of the second kind. It is defined as

$$K_\lambda(t) = \frac{1}{2}\int_0^\infty x^{\lambda-1}\exp\left(-\frac{1}{2}t(x+x^{-1})\right) dx, \qquad t > 0$$

The function is one of the linear independent solutions to the ordinary differential equation $x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} - (x^2 + \lambda^2)y = 0$.

The parameter λ represents the order of the function. Graphs for modified Bessel function of the second kind with order 1, 2 and 3 are shown in Figure 2.1. Further information concerning Bessel functions can be referred to e.g. in Bayin [2006], Ch.6 and Cížek et al. [2011].
The parametrization by $(\alpha, \beta, \delta, \mu)$ in Equation (2.1) is the most common parameterization. Interpreting parameters, μ can be considered as location parameter and
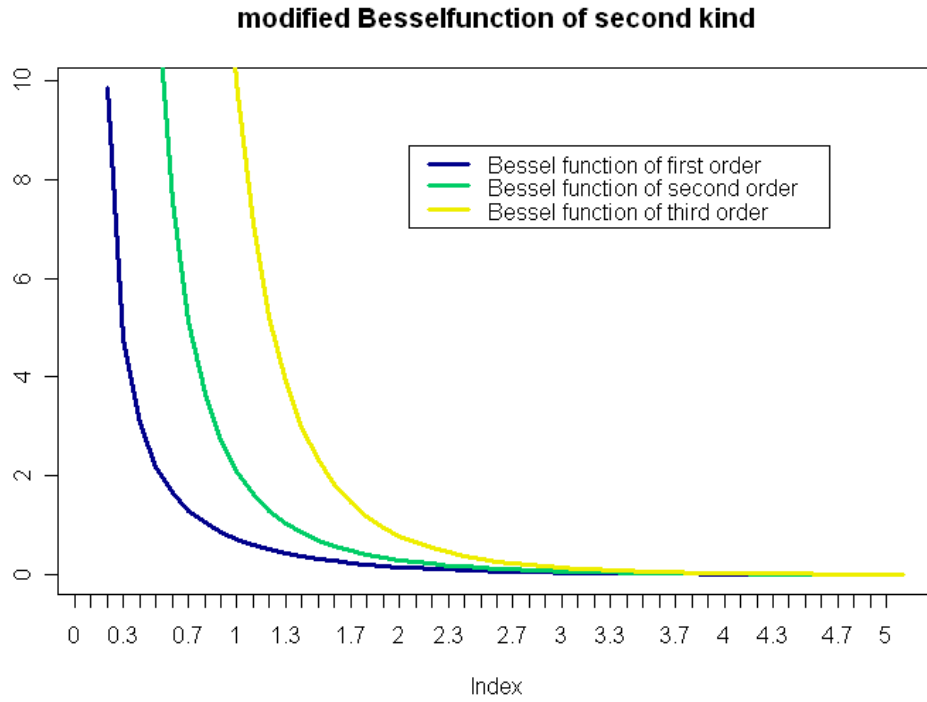
Figure 2.1: Modified Bessel function of the second kind of different order, $\lambda = 1$, 2 and 3
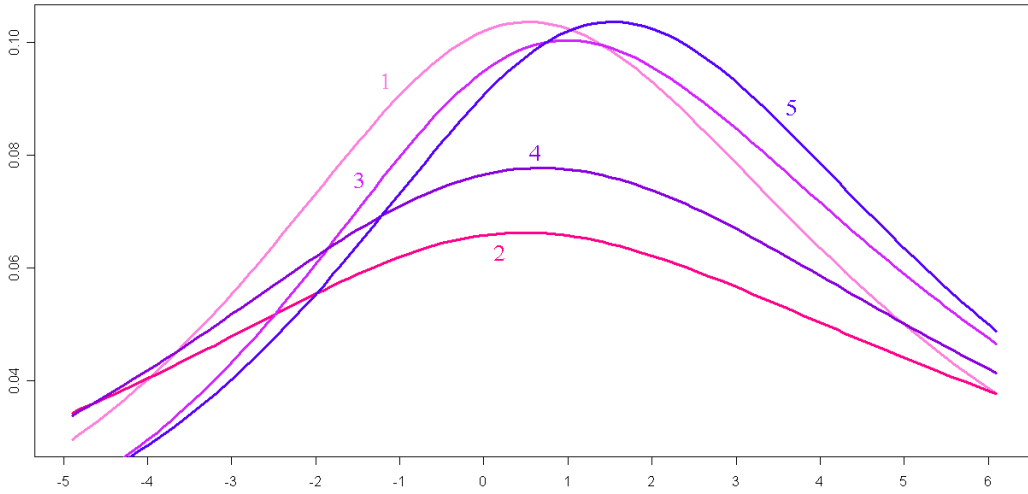
δ as scale parameter. The shape of the distribution is affected by α and β. Defining

$$\begin{aligned}
\zeta &:= \delta\gamma = \delta\sqrt{\alpha^2 - \beta^2} \\
\gamma &:= \sqrt{\alpha^2 - \beta^2} = \frac{\zeta}{\delta} \\
\xi &:= (1 + \zeta)^{-\frac{1}{2}} \\
\rho &:= \frac{\beta}{\alpha} = \frac{\xi}{\chi} \\
\chi &:= \xi\rho
\end{aligned}$$

we can parameterize the hyperbolic distribution using the parameter vector (ξ, χ, δ, μ). Deriving

$$\begin{aligned}
\alpha &= \frac{\xi\beta}{\chi} \qquad \text{and} \\
\beta &= \frac{1 - \xi^2}{\delta\xi^2\sqrt{\frac{\xi^2 - \chi^2}{\chi^2}}}.
\end{aligned}$$

the hyperbolic density in (ξ, χ, δ, μ) parametrization is

Figure 2.2: Density function of hyperbolic distribution, parameter values for parametrization $hyperbol(\xi, \chi, \delta, \mu)$ used are: (1) $(0.7, 0.1, 3, 0)$, (2) $(0.8, 0.1, 3, 0)$, (3) $(0.7, 0.2, 3, 0)$, (4) $(0.7, 0.1, 4, 0)$, (5) $(0.7, 0.1, 3, 1)$

$$
f_{HYP}(x) = \frac{1}{2\delta \frac{\xi}{\nu} K_1 \left(\frac{1-\xi^2}{\xi^2} \right)} \exp\left(-\frac{1-\xi^2}{\xi\nu} \sqrt{1 + (x-\mu)^2 \frac{1}{\delta^2}} + \chi \frac{1-\xi^2}{\xi^2\nu} \frac{(x-\mu)^2}{\delta^2} \right)
$$

(2.2)

with $\nu := \sqrt{\xi^2 - \chi^2}$, $\quad 0 < |\chi| < \xi < 1 \quad$ and $\quad x \in \mathbb{R}$.

The parameters μ and δ have the same meaning as before and can again be considered as a location and a scale parameter. The parameters ξ and χ affect the shape of the distribution. Examples can be seen in Figure 2.2. The influence of the parameter ξ can be seen by comparing curves 1 and 2, of the parameter χ by comparing curves 1 and 3, of the parameter δ by comparing curves 1 and 4 and of the parameter μ by comparing curves 1 and 5.

For data fitting, skewness and kurtosis of the distribution are of interest, too. However, due to the Bessel function appearing inside the density function of the hyperbolic distribution, a closed formulation of skewness is difficult. In Atkinson and Fienberg [1985], Ch.4 Barndorff gives an approximate relationship $\gamma_1 \approx 3\chi$ for skewness and $\gamma_2 \approx 3\xi^2$ for kurtosis at (ξ, χ, δ, μ) parametrization. Because of this interpretation of ξ and χ, the parametrization (ξ, χ, δ, μ) in Equation (2.2) is more convenient for our purpose.

Compared to the normal distribution, the hyperbolic distribution has a slower decay due to exponential descending. In Barndorff-Nielsen [1977] it is shown, that the hyperbolic distribution can be considered as a mixture of normal distributions.
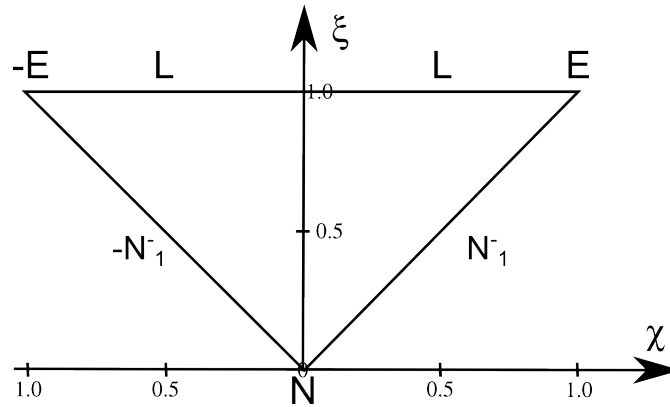
Figure 2.3: The triangle of shape, i.e. the domain of variation of invariant parameters ξ and χ of the hyperbolic distribution. Normal distribution (N), generalized inverse Gaussian distribution with $\lambda = 1$ (N_1^-), Laplace distribution (L) (symmetry of skewness) and exponential distribution (E) at the boundaries are limits of the hyperbolic distribution (cf. Atkinson and Fienberg [1985], Ch. 4).

The link between the normal distribution and the hyperbolic distribution can be expressed using generalized inverse Gaussian distributed random variables.

Proposition 2.2
If Z is an hyperbolically distributed random variable with parameter vector $(\alpha, \beta, \delta, \mu)$ and $Y \sim GIG(\lambda = 1, \xi, \psi)$ distributed a generalized inverse Gaussian (GIG) random variable with parameters $\lambda = 1, \xi$ and ψ, then

$$(Z \mid Y) \sim N(\mu + \beta Y, Y)$$

where $N(m, s^2)$ denotes the normal distribution with mean m and standard deviation s.

Further information can be found in Cížek et al. [2011]. A proof is shown in Taille [1981] by Barndorff-Nielsen and Blaesild.

Definition 2.3 (Generalized inverse Gaussian distribution)
A variable Y is said to be generalized inverse Gaussian distributed, $Y \sim GIG(\lambda = 1, \xi, \psi)$, if its probability density function is

$$f_{GIG}(x) = \frac{(\psi/\chi)^{\lambda/2}}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} \exp\left(-\frac{1}{2}(\frac{\chi}{x} + \psi x)\right) \tag{2.3}$$

defined on $(0, \infty)$ with $\chi, \psi \geq 0$.

Further information on GIG distribution can also be found in Cížek et al. [2011].

Chapter 3

Data

3.1 Siemens

The Siemens data set consists of three time series for all Europe, which are also available separated in regions. The time series are solar power, wind power and load data. The considered period of time is January 1, 2000 to December 31, 2007. During this period information is provided at hourly intervals. Sun and wind power are derived from a weather model by considering parameters like irradiance and wind speed. The required weather data are produced by Weather&Wind Energy Prognosis (WEPROG). WEPROG downscales medium-resolved analysis data from the National Center for Environmental Prediction (NCEP), the U.S. National Weather Service.

The third time series contains load data for the same space and period of time. Load profiles are retrieved from "Union for the Coordination of Transmission of Electricity" (UCTE), now "European Network of Transmission System Operators for Electricity" (ENTSO-E) or from national transmission providers. As original load profiles are not available for every region and year, missing load profiles are calculated with the help of electric power consumption (EPC). Figure 3.1 shows an average annual load in Europe for the observed period of time.

Hourly resolution can be found for the two years 2006 and 2007 at least. The prior six years are replicated using the
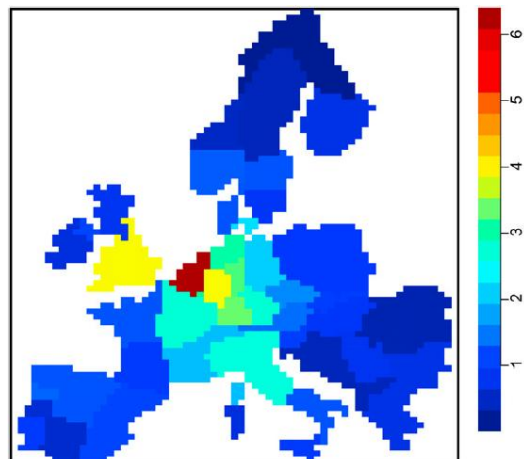
Figure 3.1: Average annual load in TWh for Europe during years 2000 to 2007. The total load across all 50 onshore regions sums up to 3130 TWh annual consumption (cf. Heide et al. [2010]).

wind turbine power curve
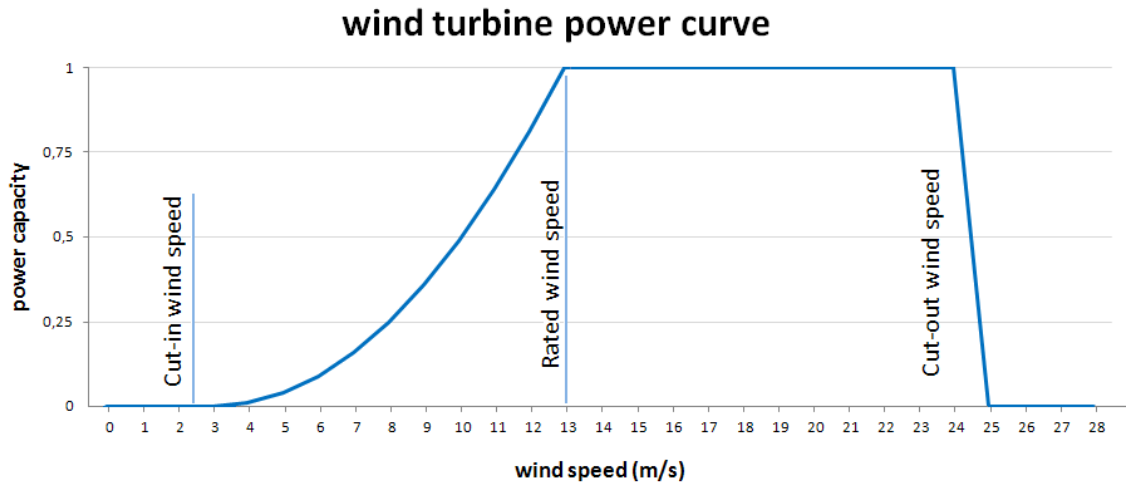
Figure 3.2: A typical wind turbine power curve. Two different curves are used to convert WEPROG wind speeds to wind power. One is used for onshore and one for offshore grid points.

relative annual electric power consumption, which is mostly known. For countries with large average load regional data profiles can be obtained from the country profiles. These are also used for smaller countries by calculating a factor from a linear regression. Linear regression analysis is performed between annual electric power consumption, population and the respective gross domestic product. Load time series are given in megawatt (MW).

As forementioned the time series starts on January 1, 2000 and every hour is taken into consideration until December 31, 2007. Time steps refer to Coordinated Universal Time (UTC). Europe is divided in sections by a grid of 0.45×0.45 degrees of latitude and longitude. This is equivalent to a spatial resolution of about 2500 km^2 per grid.

Using the weather model to derive wind generated power, several assumptions are made. Referring to the wind speed data of the WEPROG model, these wind speeds are converted into wind power by using typical wind power curves (see Figure 3.2) for each grid cell. A distinction in the curves is made between onshore and offshore grid points. Because of possible reductions in power production due to incalculable circumstances a factor of 0.93 is applied. The most important circumstances may be electrical losses and inoperable turbines. For offshore turbines power output is additionally reduced on account of wake effects. An estimation of local distribution of wind capacities in 2020 is shown in Figure 3.3.

Determining photovoltaic power, we have to assume a mixture of photovoltaic plant technologies per region. Characteristics of photovoltaic plants like tilt angle, orientation, fixed/with solar tracker, etc. are also assumed. Meteorological data like
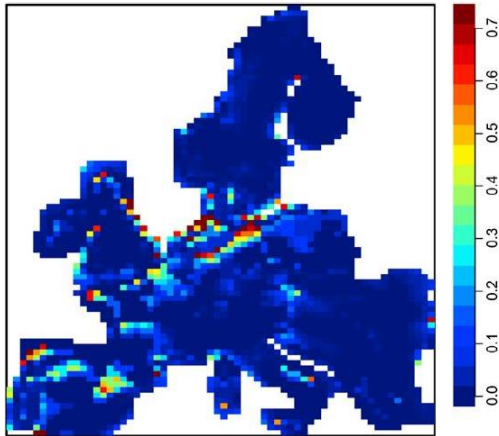
Figure 3.3: Targets of **wind power capacities** per spatial grid cell in 2020. Capacities larger than 0.73 *GW* are indicated in dark red. A total sum of 227 *GW* installed wind power capacity is expected across Europe in 2020, including 66 *GW* offshore installed capacity, cf. Heide et al. [2010].
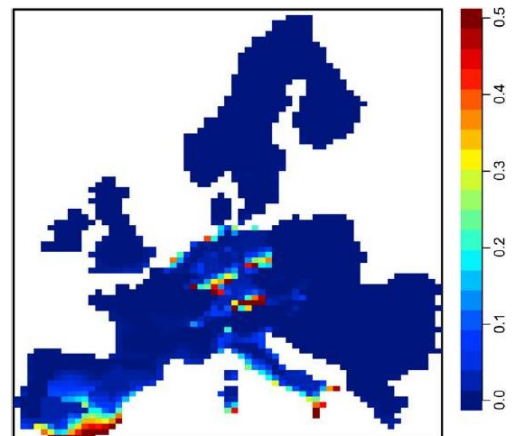
Figure 3.4: Targets of **solar photovoltaic power capacities** per spatial grid cell in 2020. Capacities larger than 0.50 *GW* are indicated in dark red. In total a sum of 68 *GW* installed sun power capacity is expected across Europe in 2020, cf. Heide et al. [2010].

global radiation, air temperature, total cloud cover, surface albedo, etc. are used to calculate photovoltaic plants affecting irradiance. An estimation of local distribution of wind capacities in 2020 is shown in Figure 3.4. Both, wind and solar power time series have been scaled by the installed capacity, resulting in a value range between 0 and 1.

3.2 Deutscher Wetterdienst

Germany's National Meteorological Service, the Deutscher Wetterdienst[1] (DWD), is a public institution with partial legal capacity under the Bundesministerium für Verkehr, Bau und Stadtentwicklung (Federal Ministry of Transport, Building and Urban Development). It is responsible for meeting meteorological requirements arising from all areas of the economy and the society in Germany. Therefore DWD is a major source of meteorological data.

Measurement points are meteorological stations of DWD, which are distributed all over Germany. The oldest data is from 1946, even though DWD was officialy founded in 1952. For the period of 2000 until 2007 a huge offer of local measurements is available. Data is available for different monitoring stations from very early on. Figure 3.5 shows locations of DWD. The locations marked with ● have monitoring stations. The symbol ▼ represents "Klimareferenzstationen" (reference stations for climate) which means that these monitoring stations are specially maintained and provide high quality data. Helgoland, Brocken and Fichtelberg belong to the category of reference stations. Munich is a regional center, therefore it is marked with ■. A research unit, that is also responsible for the administration, works there. The beginning and end of measurement activity of every single weather station is published by the DWD on its website. We choose the registration-only portal of DWD. Data can be retrieved from WebWerdis[2].

As observation points Munich airport, Helgoland, Brocken, Fichtelberg and Trier are selected. The exact coordinates can be found in Table 3.1.

Weather station	Coordinates	
Munich	48.36° N	11.81° E
Helgoland	54.18° N	7.89° E
Brocken	51.80° N	10.62° E
Fichtelberg	50.43° N	12.96° E
Trier	49.75° N	6.66° E

Table 3.1: Considered weather stations in Germany and their latitude and longitude.

Each time series starts on January 1, 2000 and ends on December 31, 2007. Hourly values exist for wind speed, temperature, sunshine duration, precipitation and air humidity (cf. Figure 4.10 for Munich and Figure 4.11 for Helgoland). Wind speed is provided in $\frac{m}{s}$, temperature in degree Celsius, precipitation in mm and air humidity in percent.

---

[1] $http://www.dwd.de/$

[2] $http://werdis.dwd.de/werdis\_de/WebWerdis\_start.do$

In almost every time series, there are some missing values. We replace the missing values by the respective measurements of neighborhood stations like Weihenstephan instead of Munich and St. Peter Ording instead of Helgoland and the average value for the still missing of the respective time series. This approach seems to be acceptable because of the small amount of missing values. Moreover, a final model of six hourly data will be used. The six hourly data originates from means of the prevailing six hours. This reduction in time resolution is caused by the lack of hourly resolved air pressure. Air pressure, which influences wind speed strongly, is only available on a daily basis at the DWD data base. In another source (cf. Section 3.3) air pressure is reported in six hourly intervals. Hence this additional data source is used, too.

Figure 3.5: Locations of DWD weather stations in Germany

3.3 European Centre for Medium-Range Weather Forecast

The European Centre for Medium-Range Weather Forecasts (ECMWF) is an inter-governmental organisation with headquarters in Reading in the United Kingdom. It is a union of national meteorological services in Europe supported by 34 States. The main focus of research is to develop numerical methods for medium-range weather forecasting. Another principal objective of the Centre is the collection and storage of appropriate meteorological data.

In the data archive services downloadable data sets can be found. For purpose of wind speed modelling we retrieved data from the ECMWF Interim Re-Analysis.[3] At the ECMWF Interim Re-Analysis data set data can be selected on a daily base. Further selection criteria are date range, time steps (maximal every six hours), and parameters like total precipitation, temperature, snowfall, surface solar radiation, among others. A list of parameters and their denomination in the files can be seen on the ECMWF homepage[4]. The surface of the earth is grasped as a grid. For every grid box the respective parameters are stored. Consequently, an enormous amount of data is available. The requested data is provided as a GRIB-file. GRIB stands for gridded binary, a data storage standard defined by the World Meteorological Organization (WMO) and known as GRIB FM 92-IX. This standard is commonly used in meteorology. The Manual on Codes (WMO-No. 306) (cf. WMO [1998]) explains the format in detail. Additional information on GRIB files can be found in WMO [2003]. On its web site, WMO offers further service manuals about the different WMO codes[5]. Instructions on how to retrieve data from grib files are given in Appendix A.

---

[3]$http://data-portal.ecmwf.int/data/d/interim\_daily/$
[4]$http://www.ecmwf.int/publications/manuals/d/gribapi/param/$
[5]$http://www.wmo.int/pages/prog/www/WMOCodes.html$

Chapter 4

Explorative Data Analysis

4.1   Siemens Data Set

As described in Section 3.1, the data set consists of wind power, sun power and load data. The three time series are plotted in Figure 4.1. As one can see the graphs have some cyclical behavior. For sun power higher values are observed in summer and lower values in winter, which is, of course, to expect on the northern hemisphere. It is to be noted an angle of incidence closer to the right angle, which causes more intensive sunshine.

An adverse behavior can be seen in wind power time series. All the peaks are realised during the winter month. Lower values can be observed during spring and reach minima in summer months. The same cyclical behavior can be seen at the load time series. During the winter months electricity consumption increases. Due to the lack of sunlight artificial light is required. Sometimes even heating is done by electricity. In summer month air-conditioning may increase electricity consumption. Due to the unpopularity of air-conditioning,the positive effects of additional daylight and due to the comfortable climate, electricity consumption is reduced during the summer. These factors may cause higher load levels in winter.

Considering the cyclical behavior during the year, one average year is investigated. Such an average year consists of daily averages over the eight years from 2000 to 2007. Every point in the Figure 4.2 is the mean of $24 \cdot 8 = 192$ values. In Figure 4.2 this average year is shown for each time series. Every curve has its specific pattern over the year. The yellow plot shows photovoltaic power all over Europe. The yearly cycle which is first noted in Figure 4.1 is also visible in Figure 4.2. Differences over the year's path are obvious. Naturally, higher photovoltaic power levels are reached in summer, as sunlight has higher intensity and duration at this time of the year.

The average wind power, represented by the aquamarine points in Figure 4.2, has higher values in winter and its trough in summer. Considering tight time intervals, the amplitude of values split is much bigger than for sun power. Due to a stronger variation of values the average wind power contains some local peaks as well.
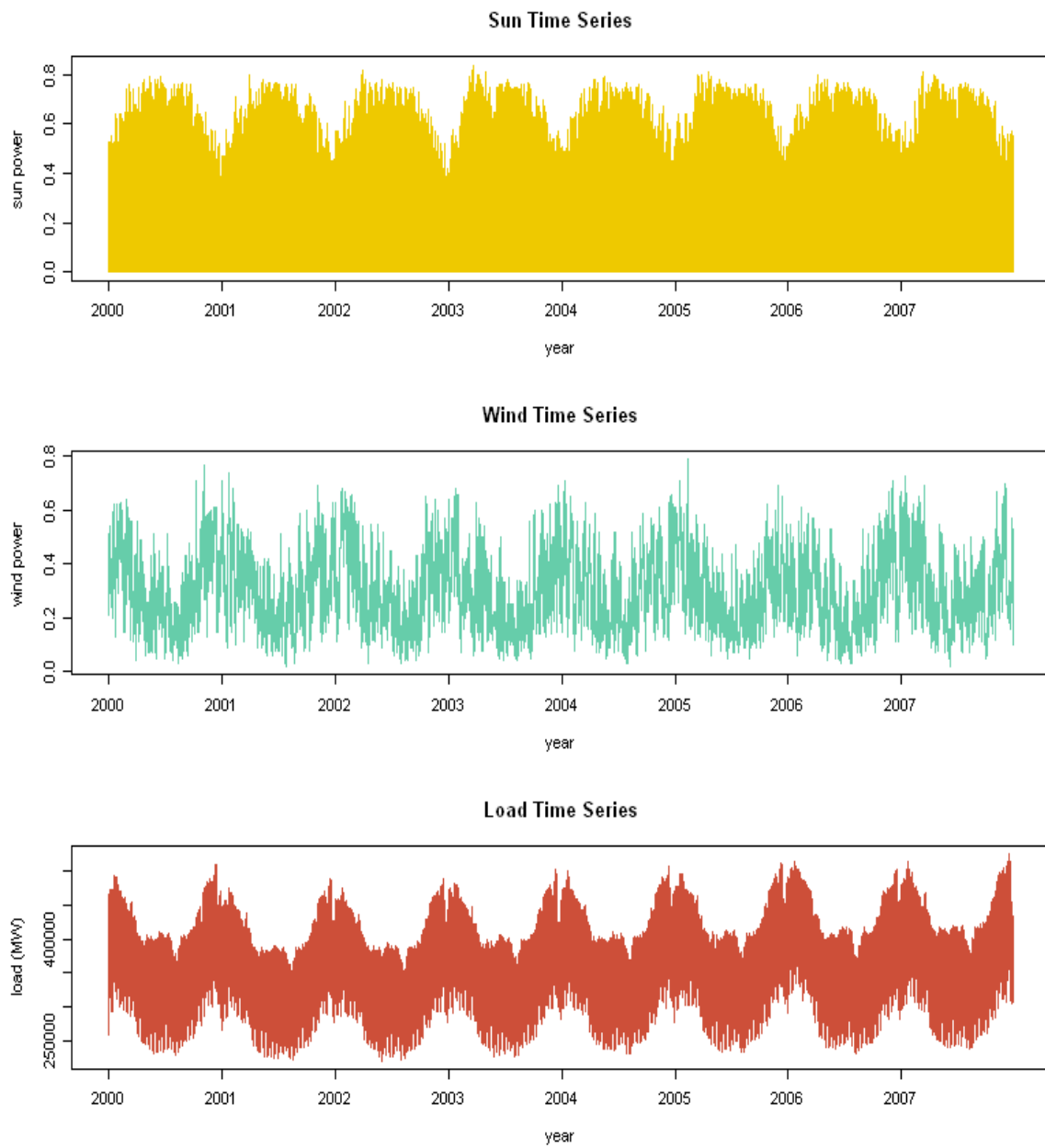
Figure 4.1: Wind power, sun power and load data time series all over Europe for the period 2000 to 2007. Sun and wind power are values scaled by capacity to be between 0 and 1. Unity of load is megawatt.
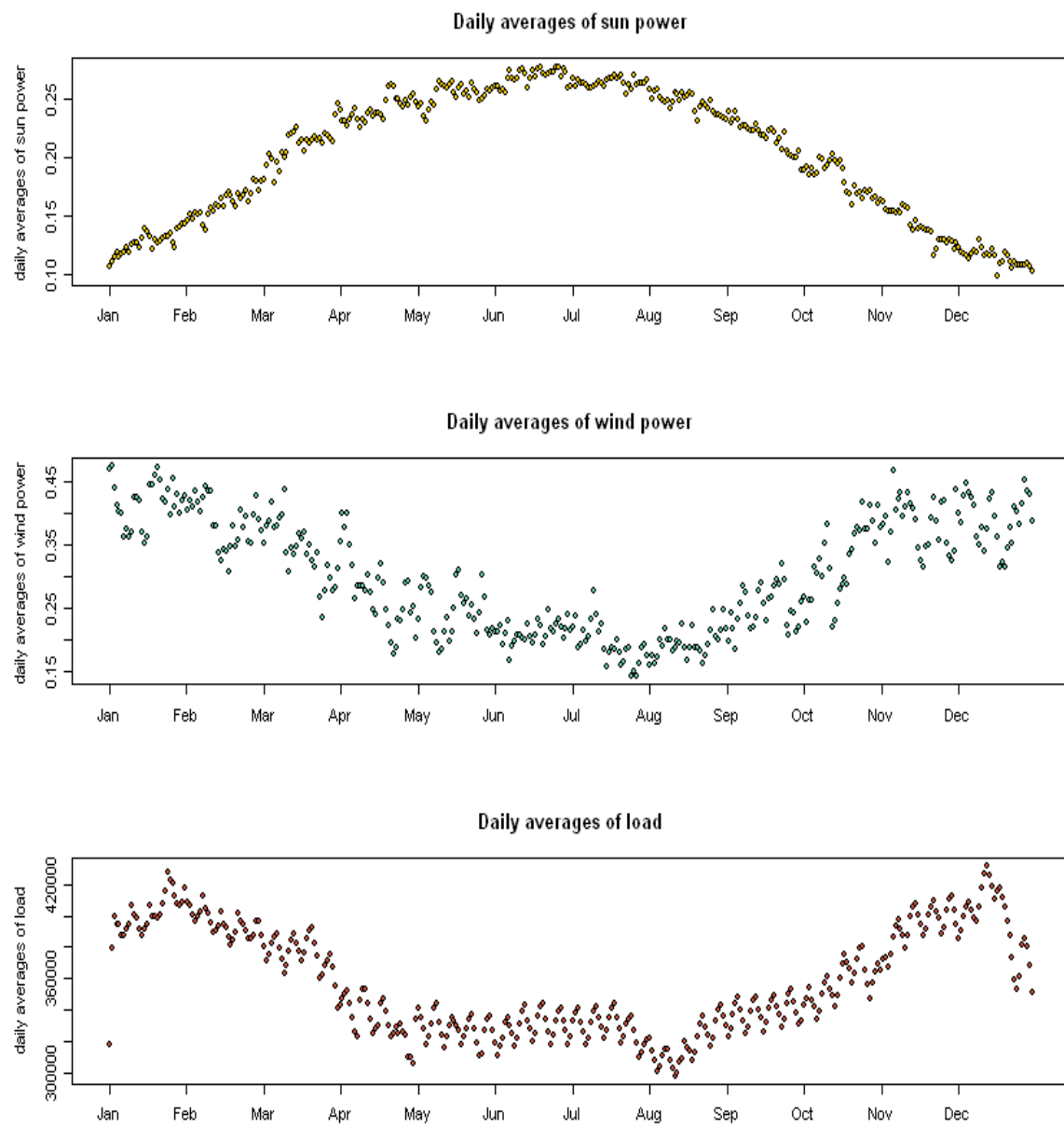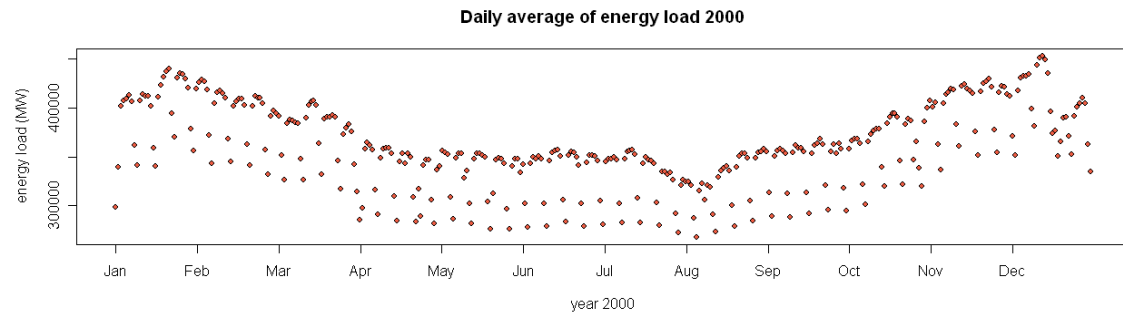
Figure 4.2: Average year of sun power, wind power and load data time series. The average year is calculated by taking the mean of 24 hours and again averaging these day means over the years 2000 to 2007. Sun and wind power are values scaled by capacity to be between 0 and 1. Unity of load is megawatt.
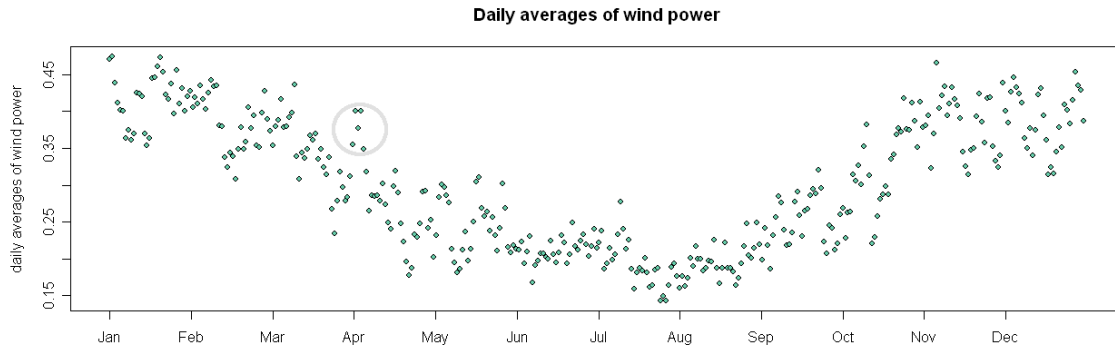
Figure 4.3: Average for the period of 2000 to 2007 for wind power. One point represents one average day. The grey circle marks the region which is further examined. In Figure 4.4 this period of time is shown.

For a more detailed analysis of this phenomenon a peak at the beginning of April is chosen. The peak is marked with a grey circle in Figure 4.3. Separately this period is shown in Figure 4.4. Original hourly values of wind energy are plotted from March, 28 to April, 10 for every single year. Therefore eight curves can be seen for the years 2000 till 2007. The peak seems to be caused by the values in the years 2000, 2002 and 2003. But in Figure 4.4, where one can see the curves for the years 2000 through 2007, a peak at the beginning of April seems not to be evident.
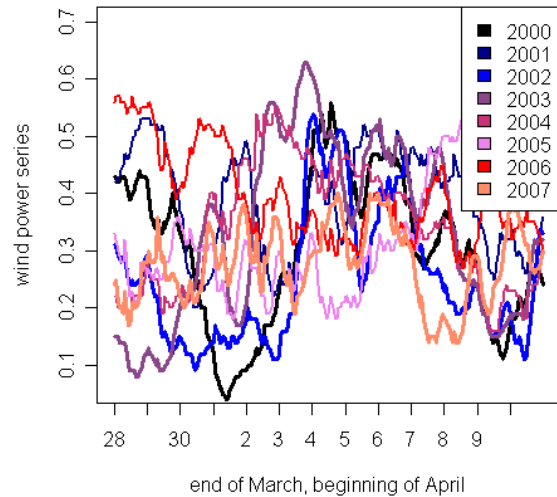
Figure 4.4: Wind power as end of March and beginning of April. Each year from 2000 to 2007 has a single curve.

Considering the lowest graph in Figure 4.2, energy load in Europe, the shift in the energy load pattern over the entire time series is attention caching. These shapes seem to originate from a weekly scheme. The date shift of the weekends over the reviewed year biases this effect at an average over all years. In Figure 4.5 averages over 24 hours of energy load in Europe are plotted for year 2000 only. Three different levels appear. The lowest level is present on Sundays and the second lowest on Saturdays. Higher values are measured on working days. As economies with industrial production have a huge energy load, the stop of the production process in many companies on the weekends is probably the principal factor for this phenomenon.

Figure 4.5: Average of energy load data (in MW) per day for year 2000.

In Figure 4.6 no average is built over the successive hours in contrast to Figure 4.2. Figure 4.6 shows the mean of every hour of a year for sun power, wind power and load. One mean point is calculated by eight values of one hour in the years 2000 till 2007. Each of the three plots has $365 \cdot 24 = 8760$ observation points showing average values over the eight examined years.

Having daily cycles like sunrise and sunset, the average hour has a certain value, which is similar to the value that can be measured at the same time on the following days. Thus a structure of different lines originating from hourly measurements is observed. The inherent daily seasonality of sun power data can be observed through clear separation of different day times. Different levels of value lead to the separated structure, that can be observed in the yellow graph of Figure 4.6. At most 24 lines of points could be drawn, but as all points before sunrise and after sunset coincide into zero, less levels are observed. Due to the increase and decrease in daylight during the seasons, the duration as well as the intensity change. This is reflected in the increase and the decrease of the amount of levels and the levels themselves. Most of the levels are observed in summer with the highest values, concomitantly.

Hourly wind power data for average years in all of Europe, which can be seen in the second plot of Figure 4.6, do not seem to have any daily seasonality, because no different levels are observed. However an annual seasonality can be observed again. This pattern has been seen already in Figure 4.2. The amount of wind power produced is obviously higher in winter than in summer. Due to meteorological conditions large winter storms, which are hardly predictable, occur in Northern Europe (cf. Quaschning[2007/2008]).

In Figure 4.6 the red points marking load profile in Europe increase during the winter times as well. Differences per hour seem to exist, but no well-defined levels can be found. As there is some free space at this broad data band, maybe a distinction into two groups could be done. Lower values may represent night and work free time where as higher values could originate from day and core working time. Having in mind, that main energy consumption comes from industry and business, this assumption seems to fit. Two cuttings which attract attention could be explained considering the same reason. One is around Christmas and New Year, the other one
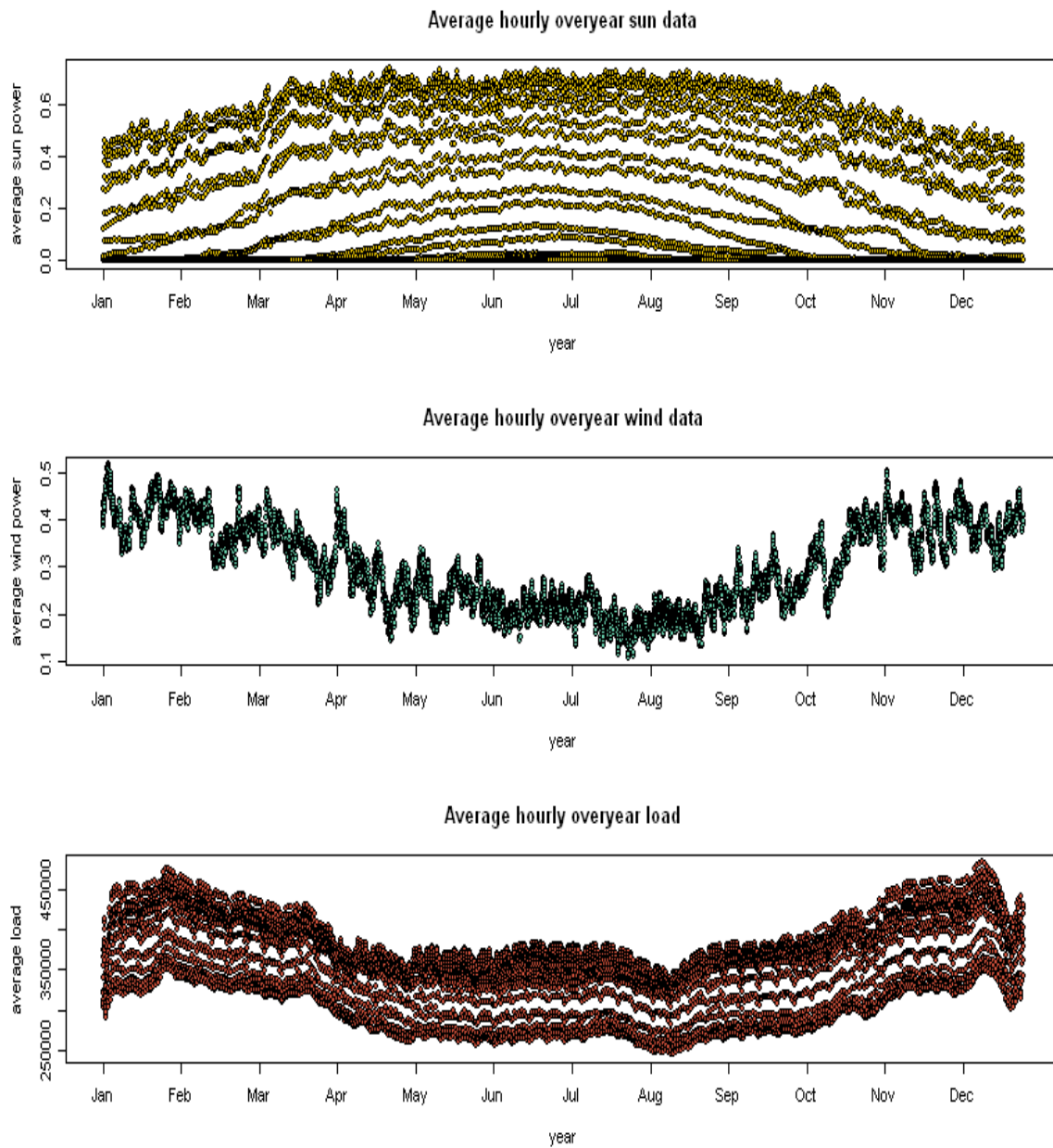
Figure 4.6: Average over the considered eight years per hour. Average is not built over successive hours. Instead mean is calculated by eight values of one hour in year 2000 to 2007. Each plot has $365 \cdot 24 = 8760$ observation points. Having different levels of values for different hours of the day at sun power and load data, a structure of at most 24 levels appears. Sun and wind power are values scaled by capacity to be between 0 and 1. Unity of load is megawatt.
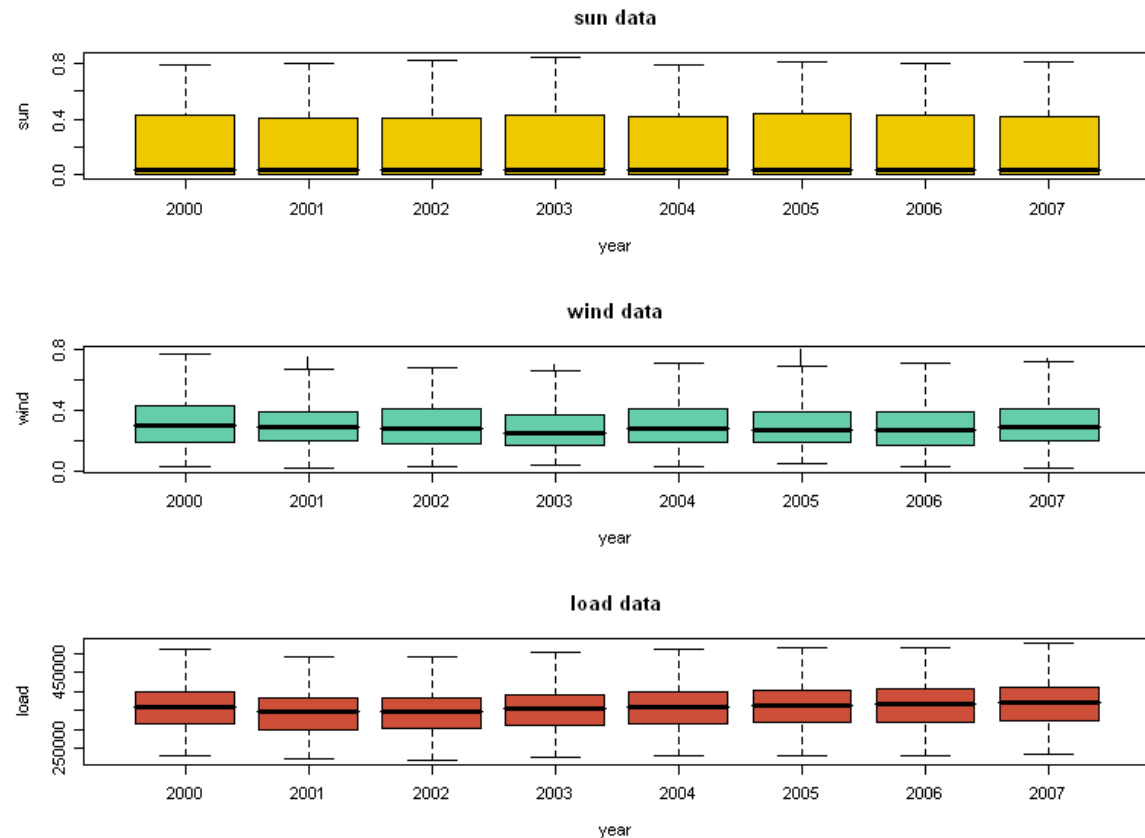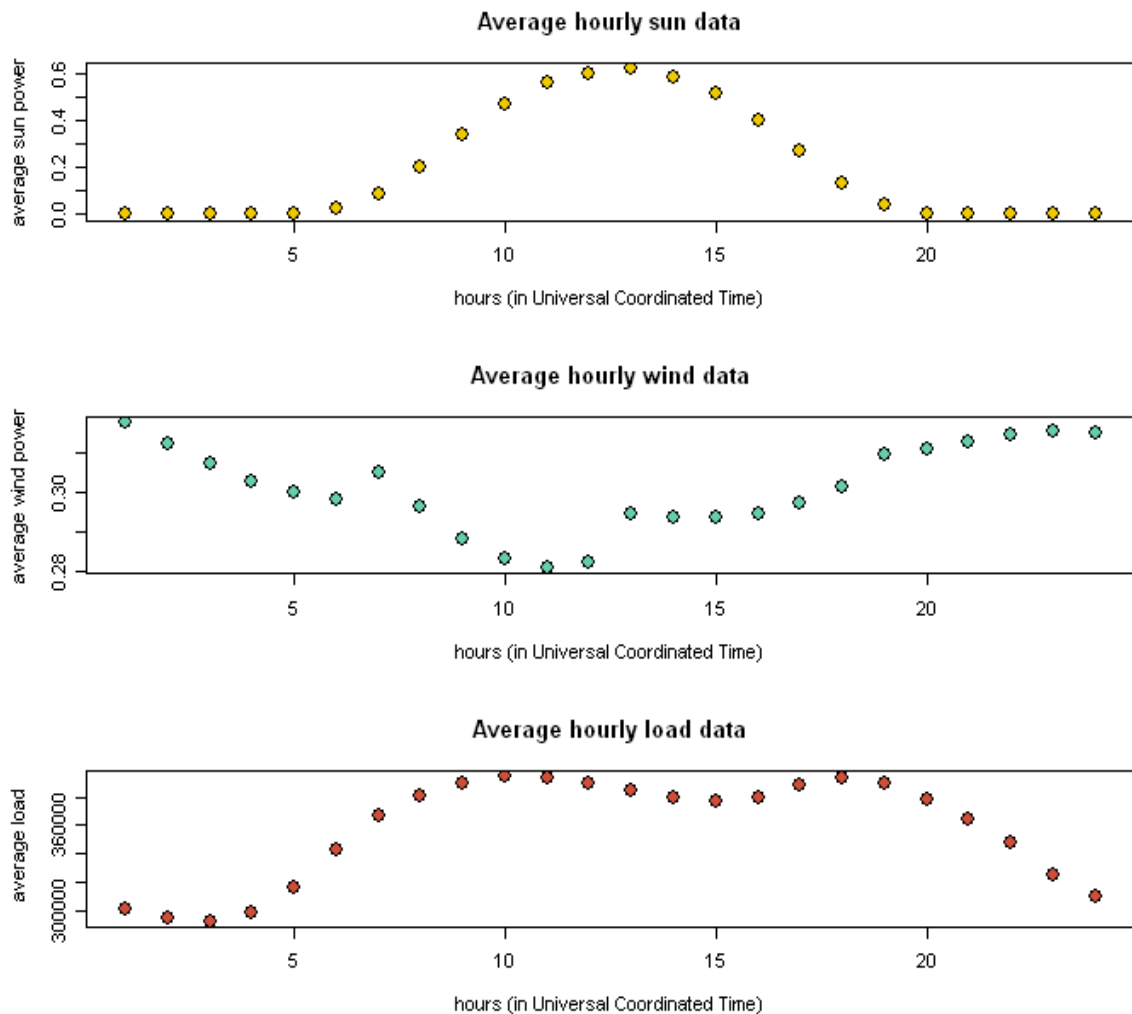
Figure 4.7: Boxplots of sun power, wind power and load data for the period 2000 to 2007. Sun and wind power are values scaled by capacity to be between 0 and 1. Unity of load is megawatt.

at the beginning of August. Both periods are classical holiday periods all over Europe and may contain temporary shutdowns of industrial production entities which could lead to the observed effect of load decrease.

The fact that sun energy production is restricted to daytime only is evident from the data. The median for the sun time series in the boxplots in Figure 4.7 is skewed strongly to zero. This behavior of the sun time series being clearly predictable to be zero every night, causes a special treatment of sun time series which is taken a closer look at the section of influences on photovoltaic power (cf. Section 6.1).

Besides seasonal trend per year in Figure 4.2 and per day in Figure 4.8 a slight trend in the load time series is implied by Figure 4.9. The red graph of load data shows a slightly increasing mean for all data of every single year and the belonging variance of this year.

Considering the mean of 24 hours over the whole time series, a concise and reasonable daily seasonality is visible for sun power, the first graph in Figure 4.8. There are variations over day, observable for wind power and load data, too. But the amplitude for the variations of both time series is much smaller than the amplitude for the sun power time series. Since sun power is, of course, only available when sun shines, we observe always zeros during night.

Figure 4.8: Averages over the whole time series of sun power, wind power and load data for the period 2000 to 2007 is calculated. The plotted points represent an average day with hours in Universal Coordinated Time. Sun and wind power are values scaled by capacity to be between 0 and 1. Unity of load is megawatt.
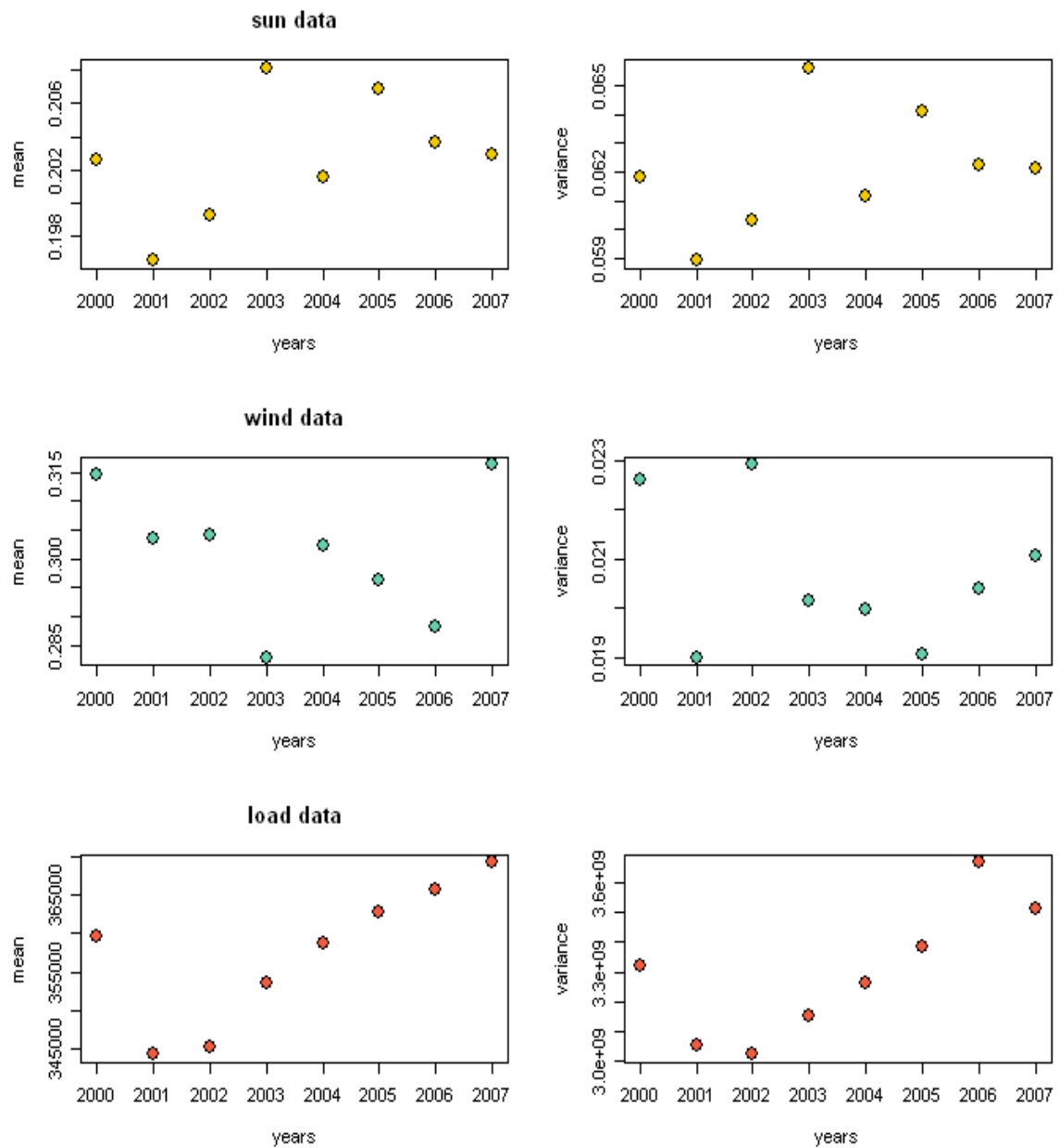
Figure 4.9: Mean and variance of sun power, wind power and load data for the period 2000 until 2007. Mean and variance are calculated over all values of prevailing year. Sun and wind power are values scaled by capacity to be between 0 and 1. Unity of load is megawatt.

4.2   DWD Data Set

Instead of wind power we now examine wind speeds, as wind power is the conversion of wind speed into electric energy wind speed can also be examined directly. Transforming the wind speeds with the adequate wind turbine curve may be more accurate than just estimating an approximate proportional relationship of wind power to squared wind speed. From now on we focus on wind speed, therefore DWD data are examined. The data set contains not only the wind speed, but also temperature, precipitation, air humidity and sunshine duration. As described in Section 3.2, we consider the period from 2000 to 2007. The different monitoring stations are reported in Table 3.1. Plots are shown exemplarily for weather stations in Munich (Figure 4.10) and Helgoland (Figure 4.11).

As can be especially seen in Figure 4.10, (monitoring station of Munich) air humidity corresponds with temperature and precipitation developments. Cross-correlations between air humidity and temperature show a strong dependency (cf. upper graph of Figure 4.12). The oscillating behavior is caused by daily fluctuations of temperature. The lower graph of Figure 4.12 has smaller cross-correlations for every lag. A wider time horizon of a year (cf. Figure 4.13) exposes annual seasonalities. Temperature and precipitation are higher during summer times. In winter mostly high air humidity is measured. Due to low temperatures low variation in air humidity occurs. In summer months the precipitation is higher and therefore air humidity is high too. But high temperatures cause more evaporation, which leads to more variation in air humidity with low and high values during summer.

We now consider wind speed in detail. Respective Figures can be found in Appendix B. As wind speed has a high fluctuation, we examine the differences in wind speed for each month. For each month, we calculate separately daily averages. E.g. for January 1, the average is based on all wind speeds measured on January 1, 2000, January 1, 2001, January 1, 2002, ... January 1, 2007. From the graphs for Munich (cf. Figure B.1) one can see, that the differences in the course of the year are smaller than for Helgoland. Generally, measured wind speeds are smaller in summer than in winter time. Any further fluctuations over the days of the month seem to be arbitrary.

In comparison to Munich, Helgoland (cf. Figure B.2) shows stronger differences between wind speeds in summer and winter. Generally, the measured wind speeds on Helgoland are higher, which is no surprise for an island located in the North Sea. Even in summer Helgoland has higher wind speeds than Munich in winter. In Helgoland values of about $12 \frac{m}{s}$ are standard during winter which is about twice as high as summer wind speeds.

In Figure B.3 for Brocken and Figure B.4 for Fichtelberg a behavior similar to that of Helgoland (cf. Figure B.2) can be observed. Wind speeds are quite high. They range between $6 \frac{m}{s}$ and $14 \frac{m}{s}$, whereas Trier (Figure B.5) has wind speeds between $2 \frac{m}{s}$ and $6 \frac{m}{s}$, which makes it more comparable to Munich (Figure B.1).
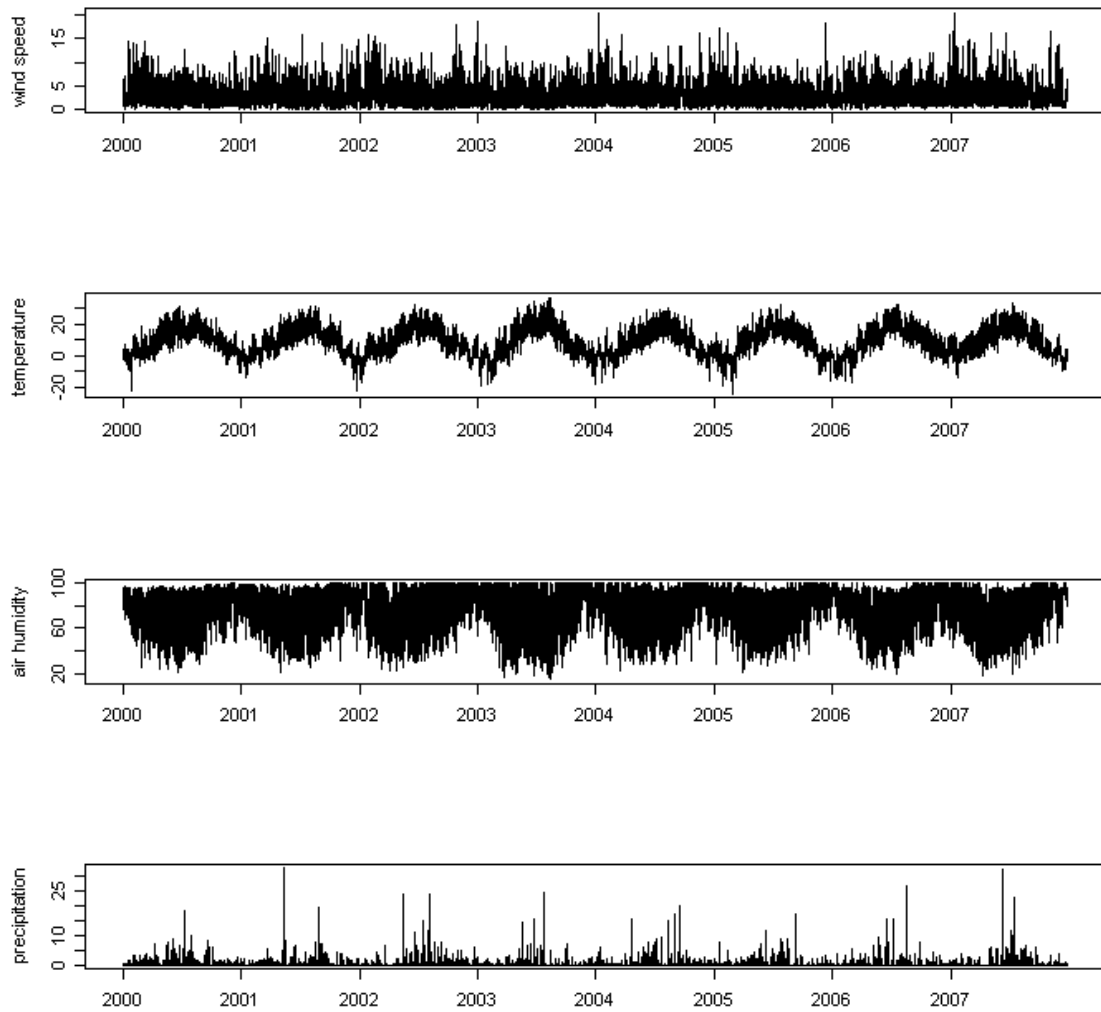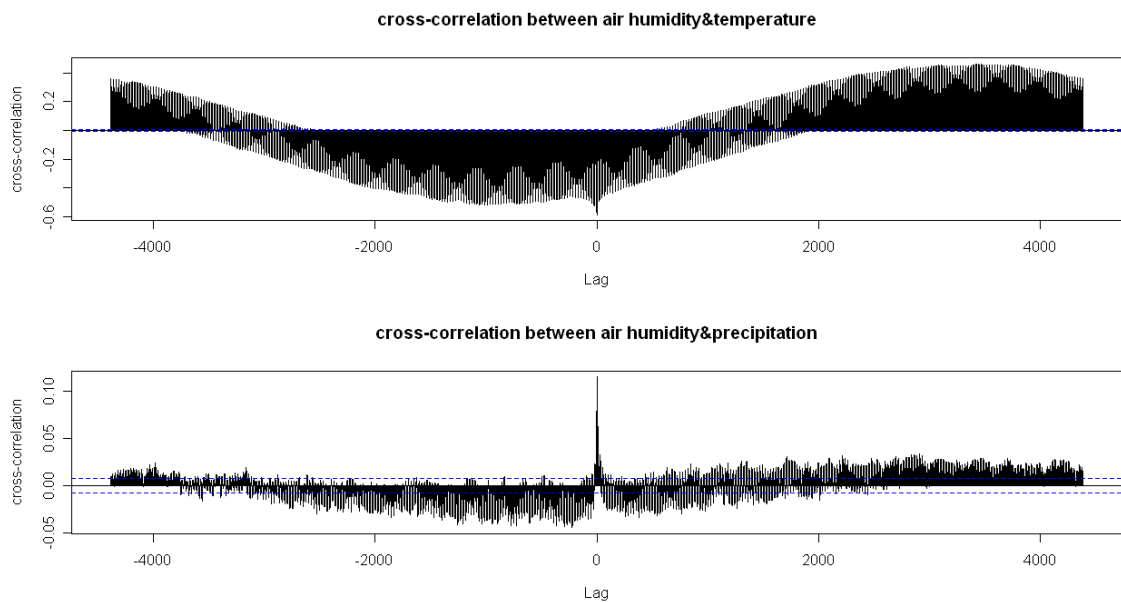
Figure 4.10: Wind speed ($\frac{\mathrm{m}}{\mathrm{s}}$), temperature (° C), air humidity (%) and precipitation (mm) reported at hourly intervals at the weather station of Munich airport
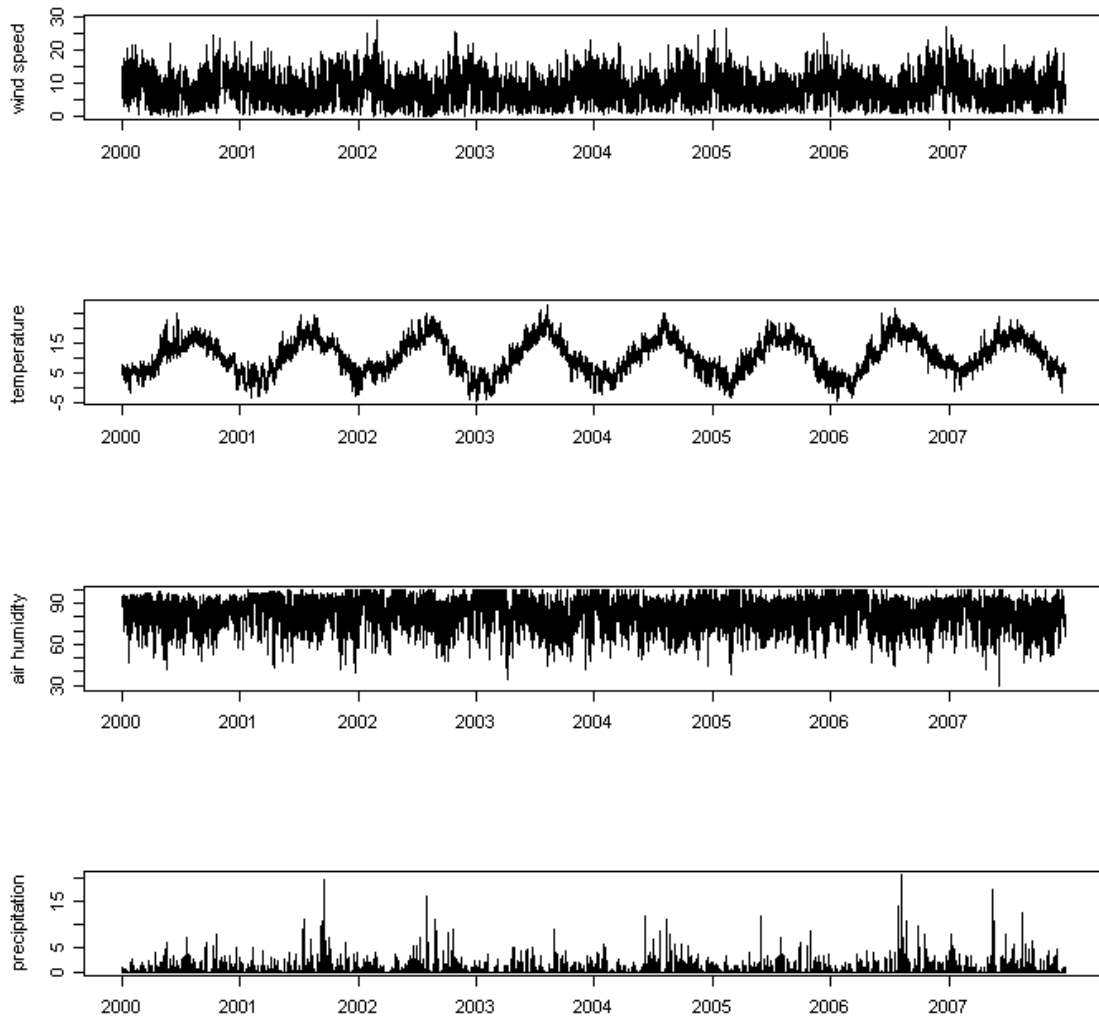
**Figure 4.11:** Wind speed ($\frac{\mathrm{m}}{\mathrm{s}}$), temperature ($^\circ$ C), air humidity (%) and precipitation (mm) reported at hourly intervals at the weather station of Helgoland
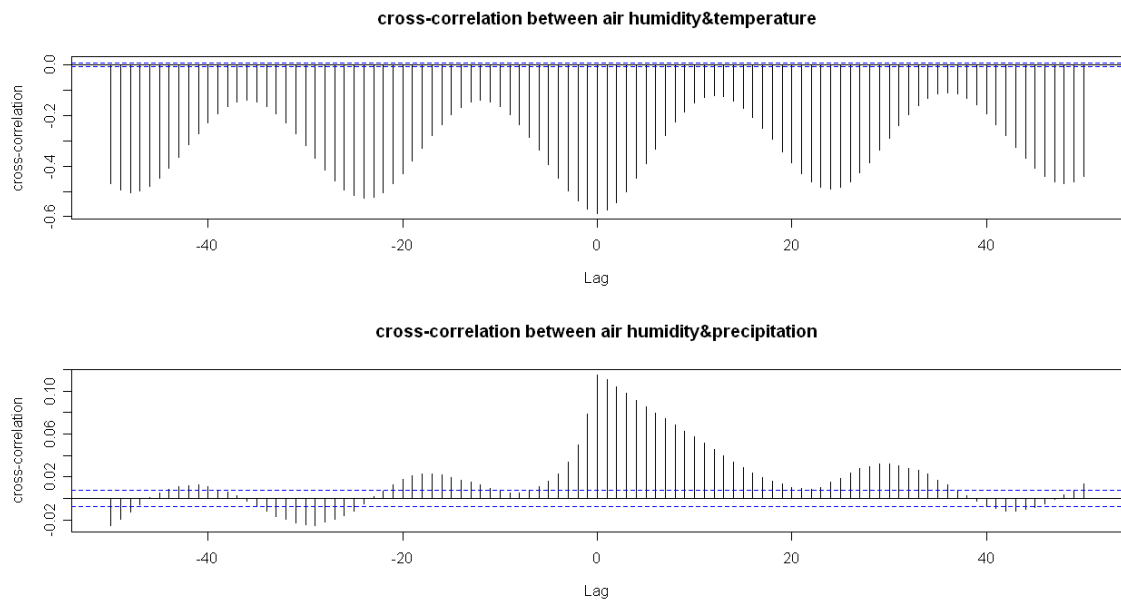
**Figure 4.12:** Cross-correlations between air humidity and temperature and between air humidity and precipitation in Munich. A time step has an hourly resolution. A period of four days is considered.

Figure 4.13: Cross-correlations between air humidity and temperature and between air humidity and precipitation in Munich. A time step has an hourly resolution. A period of one year is considered.
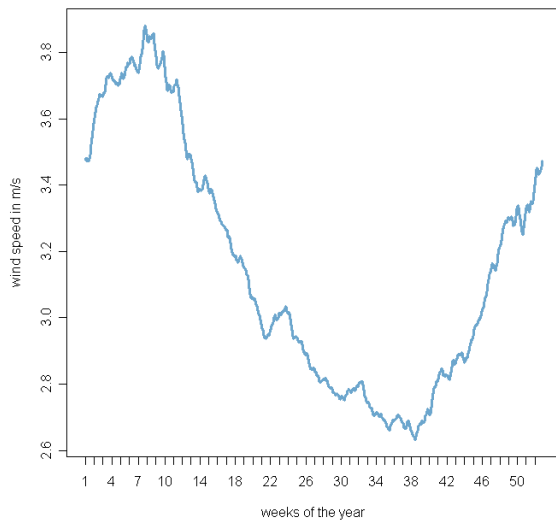
Figure 4.14: Moving average of wind speed in **Munich** for six weeks interval, a rough structure of wind speed over the year; Values vary between 2.6 $\frac{m}{s}$ and 3.8 $\frac{m}{s}$ over the year.
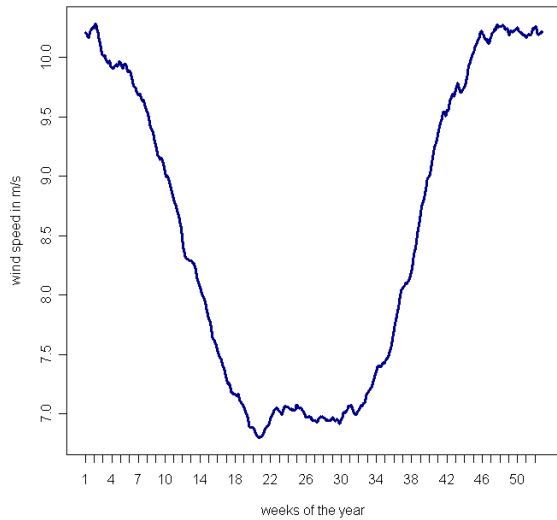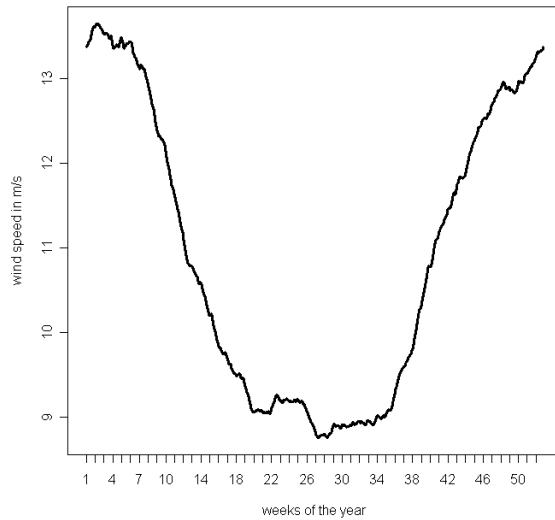
Figure 4.15: Moving average of wind speed in **Helgoland** for six weeks interval, a rough structure of wind speed over the year; Values vary between 6.8 $\frac{m}{s}$ and 10.2 $\frac{m}{s}$ over the year. Here an eminently higher amplitude than in Figure 4.14 is seen.

The patterns in Figure B.1, B.2, B.3, B.4 and B.5 suggest the presence of an annual seasonality component. Building a moving average over six weeks, this seasonality can clearly be seen (Figure 4.14, 4.15, 4.16, 4.17 and 4.18). Higher values are observed in winter. A decrease evolves in spring. During the summer months, wind speeds reach the bottom of their yearly averages. During autumn wind speed increases again.

Having considered the daily averages over all months, we consider the hourly averages. Recalling the annual variety of the graphs, we have a look at the daily behavior for each months. Figures B.6, B.7, B.8, B.9 and B.10 show average wind speeds for each hour of the day based on data from the month January to December, separately. For each month the hourly values are calculated as means of the respective hour over the period of eight years and then standardized.

As one can see the resulting values have some structure. For Helgoland (Figure B.7) and the two monitoring stations on mountains (Figure B.8 for Brocken and Figure B.9 for Fichtelberg) a trough can be observed in the afternoon. This trough is more refined in spring and autumn. As opposed to the monitoring stations on mountains, a peak is observed for Figure B.6 and B.10 in the afternoon, which are the measure-

Figure 4.16: Moving average of wind speed at **Brocken** for six weeks interval, a rough structure of wind speed over the year; Values vary between 8 $\frac{m}{s}$ and 14 $\frac{m}{s}$ over the year.
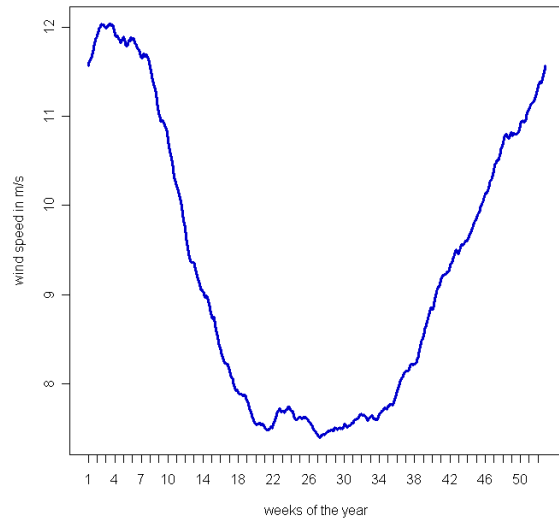

Figure 4.17: Moving average of wind speed at **Fichtelberg** for six weeks interval, a rough structure of wind speed over the year; Values vary between 7 $\frac{m}{s}$ and 12 $\frac{m}{s}$ over the year.
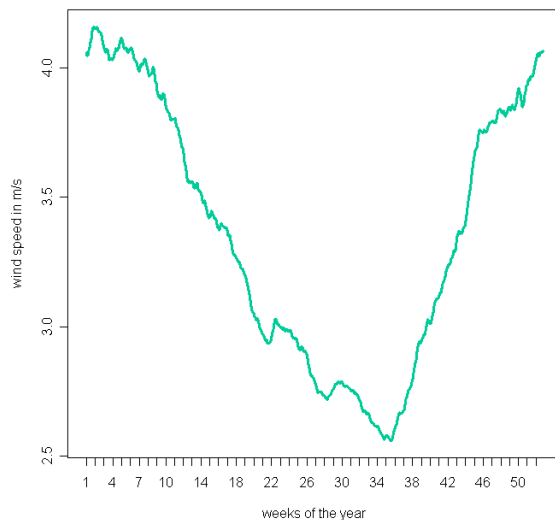
Figure 4.18: Moving average of wind speed in **Tier** for six weeks interval, a rough structure of wind speed over the year; Values vary between 2.5 $\frac{m}{s}$ and 4.2 $\frac{m}{s}$ over the year.
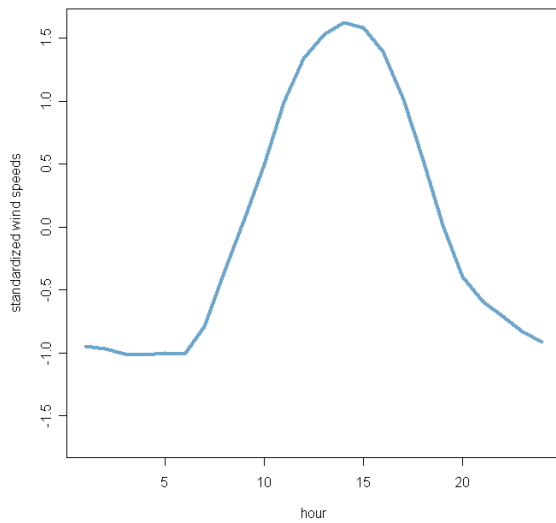
Figure 4.19: Centered average values of wind speed in **Munich** per hour, having an average day of underlying years 2000 to 2007 portrayed. The values vary quite strongly over one day. Maximum of the day is reached between 2:00 pm and 3:00 pm.
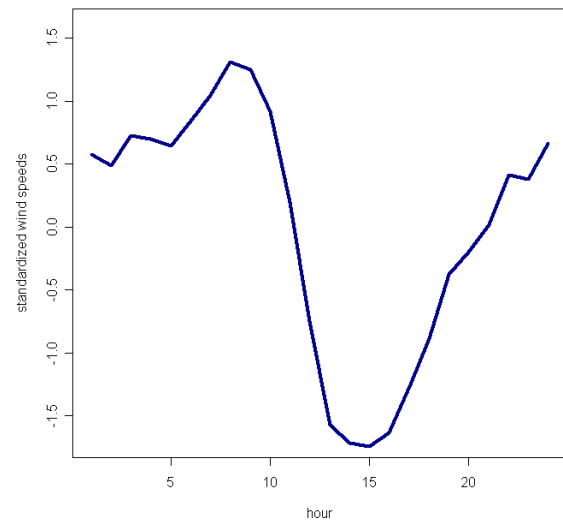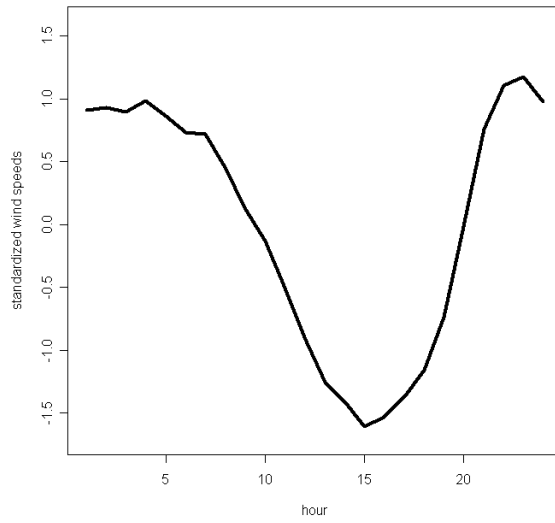
Figure 4.20: Centered average value of wind speed on **Helgoland** per hour, having an average day of underlying years 2000 to 2007 portrayed. The values develop adverse to the averages of Munich. Lowest point of the day is reached in the afternoon between 2:00 pm and 3:00 pm.

ment data of monitoring stations in cities. Having nearly the same diurnal structure the maximum in the afternoons is more revealing in spring, summer and autumn. In winter the same structure is evident, but the extent is smaller.

For Munich and Trier (Figure B.6 and B.10) the this structure does not alter a lot during the year. Monitoring stations in Helgoland, at Brocken and Fichtelberg (Figures B.7, B.8 and B.9) have more changes in structure. But calculating hourly averages, the diurnal seasonality is still evident. The graph for Helgoland is shown in Figure 4.20, for Brocken in Figure 4.21 and for Fichtelberg in Figure 4.22. For these monitoring stations at mountain and on the island in the North of Germany a different diurnal seasonality is observed in comparison to the cities. For the cities Munich and Trier the diurnal seasonalities resulting of hourly averages over all data are seen in Figure 4.19 and 4.23.
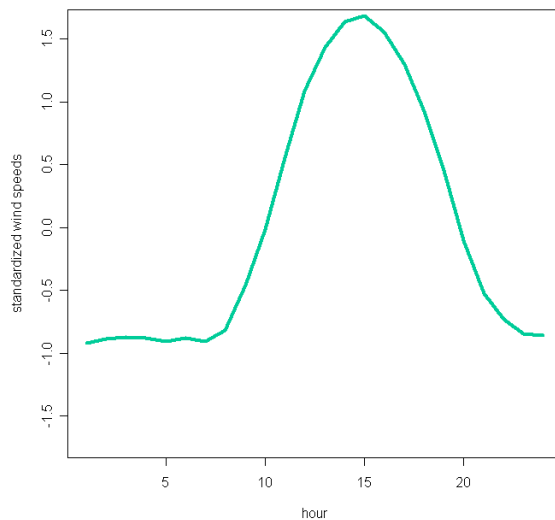
Figure 4.21: Centered average values of wind speed at **Brocken** per hour, having an average day of underlying years 2000 to 2007 portrayed. Lowest point is reached in the afternoon.
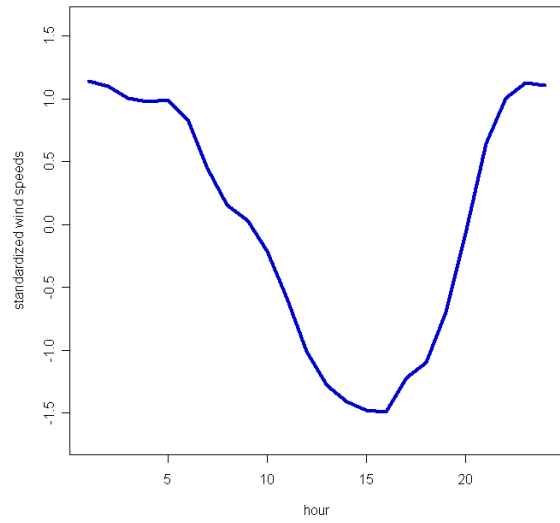
Figure 4.22: Centered average value of wind speed at **Fichtelberg** per hour, having an average day of underlying years 2000 to 2007 portrayed. Lowest point is reached in the afternoon.
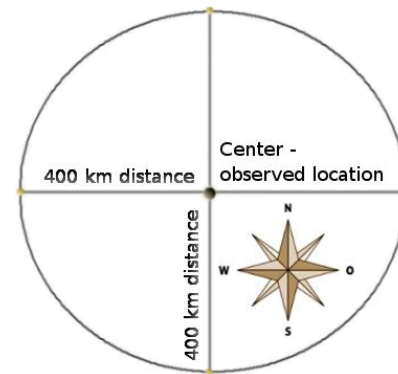

Figure 4.23: Centered average values of wind speed in **Trier** per hour, having an average day of underlying years 2000 to 2007 portrayed. The values vary quite strongly over one day. Maximum of the day is reached between 2:00 pm and 3:00 pm.
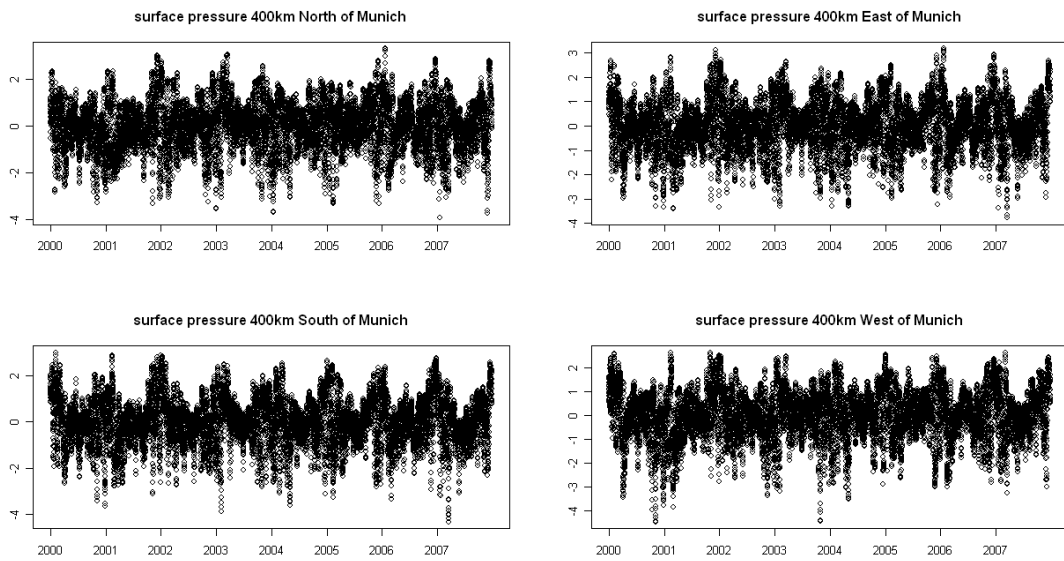
4.3 ECMWF Data Set

One factor effecting wind speed is pressure, as wind originates from movements of air masses to equiponderate pressure conditions at different locations. Thus wind speed is influenced by the extent of pressure differences. During high pressure differences a larger suction exists from area with high pressure to areas with low pressure. Since wind speed increases for higher suction, wind speed increases with higher differences of pressure in its surrounding.

Figure 4.24: Measurement points of surface pressure. A radius of 400 km was selected due to high correlations between wind speed and surface pressure differences. Munich, Helgoland, Brocken, Fichtelberg and Trier were set as prevailing center.

Therefore surface pressure was examined at the observation points, which are the five monitoring stations, and at four points on a circle around the observation points (see Figure 4.24) with a certain kilometer radius. Since retrieving data from a GRIB file requests coordinates, information in kilometers were transformed into coordinates by the approximation of $2\pi \cdot 6378 \cdot \frac{1}{360} \cos(\frac{\text{"current latitude"}}{180\pi})$ for one degree of longitude and 111 km per degree of latitude (cf. Longley et al. [2005]). Locating the data read out at the nearest grid point, this approximation is far enough. The four main points of the compass were selected as directions. Trying a distance of 200, 400, 600 and 800 kms, the 400 km radius was chosen because of the highest correlations for wind speed and surface pressure differences.

In Figures 4.25 one can see the standardized pressure differences which exist between Munich and 400 kms northwards, eastwards, southwards and westwards. In winter higher pressure differences can be observed than in summer. Same behavior is seen in Figure 4.25. This observation coincides with the observations for higher wind speeds in winter and lower wind speed in summer.

Figure 4.25: Standardized surface pressure differences at the celestial directions around Munich in a distance of 400 kms. The observed period of time is from beginning of 2000 to the end of 2007. Munich is considered as the center.
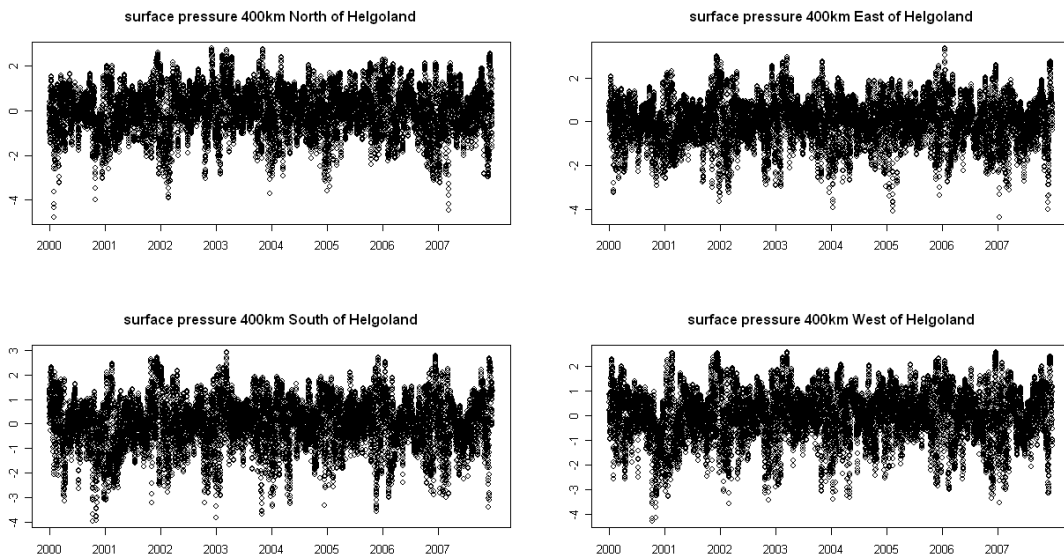
Figure 4.26: Standardized surface pressure differences at the celestial directions around Helgoland in a distance of 400 kms. The observed period of time is from beginning of 2000 to the end of 2007. Helgoland is considered as the center.

Chapter 5

Time Series Models for Wind Speed

In this chapter we focus on time series modelling for wind speed data. In Chapter 4 the data are described in detail. Some annual and daily seasonalities are noticed, which are reflected by differences in wind speeds over one day and in the course of the year. These persistent variations are to be detected by a model for wind speeds. The annual component is modelled by moving averages of two weeks over an average year in the period of 2000 to 2007. The diurnal component describes the fluctuations during one day, which are calculated by hourly averages on the whole time series. The daily course can be seen in Figures 4.19, 4.20, 4.21, 4.22 and 4.23 for each monitoring station. Likewise the annual course is shown in Figures 4.14, 4.15, 4.16, 4.17 and 4.18. Serving as daily and annual component to detect seasonalities, the values are standardized. Further information about methods for deseasonalizing and detrending data can be found in Brockwell and Davis [1991], Ch.1.4.

Using a daily and an annual component as covariates, modelling wind speed data starts with a linear regression of the form

$$W_i = \gamma_0 + \gamma_1 d_i + \gamma_2 a_i + Y_i$$

where

W_i are wind speeds in an hourly resolution.

d_i represents seasonality over one day. It is calculated as standardized hourly average on the whole time series. The shape for Munich can be seen in Figure 4.19, which is similar to the curve of Trier in Figure 4.23. At Helgoland the shape of the daily curve is different (Figure 4.20). Daily seasonalities of Brocken and Fichtelberg are plotted in Figures 4.21 and 4.22.

a_i is the annual component, which represents a standardized moving average of two weeks or 14 days of an average year originating from the observed years 2000 to 2007 (see Figures 4.14, 4.15, 4.16, 4.17 and 4.18).

Y_i are deseasonalized and centered wind speed data in an hourly resolution(cf. Figure 5.1).
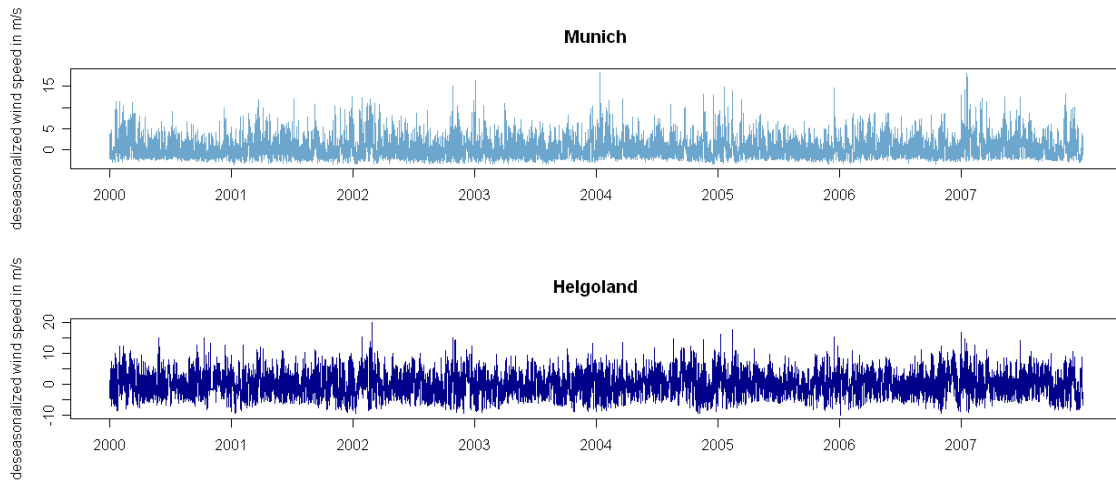
Figure 5.1: Deseasonalized and centered wind speed data for Munich and Helgoland in an hourly resolution.
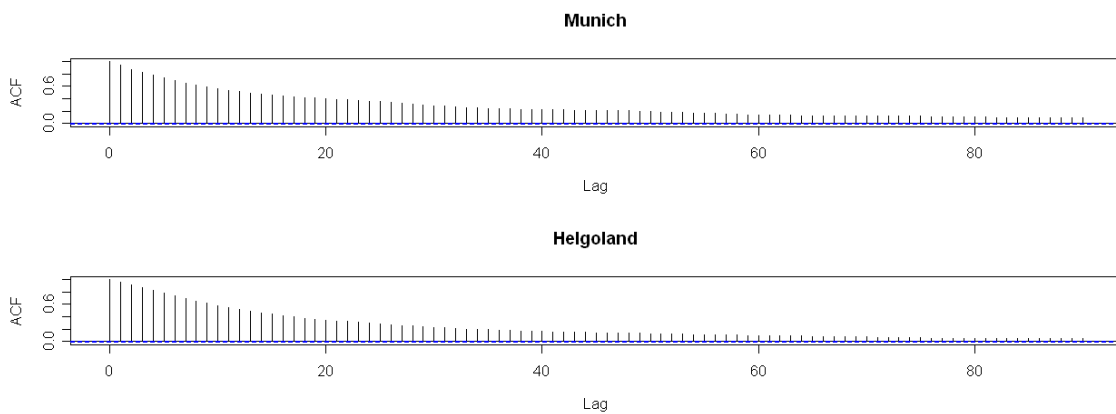

Figure 5.2: Autocorrelation function of the residuals Y_i for Munich and Helgoland. Y_i are the deseasonalized and centered wind speeds.

Daily and annual seasonalities are removed by these covariates. The resulting residuals of the regression should merely show seasonalities with daily or annual cycles (cf. Figure 5.1).

As wind speed data of DWD are available every hour, the parameter i measures time in units of 1 hour. For residuals Y_i of the regression the autocorrelation function is shown in Figure 5.2. The appearance gives notice of autocorrelations in the data. Obviously, there are strong dependencies left in the data after removing the daily and annual seasonalities.

We now try to explain the deseasonalized and centered wind speeds using covariates. As mentioned in Section 4.3 the main influence factors on wind speed are investigated. Keeping in mind that movements of air are due to pressure differences, the
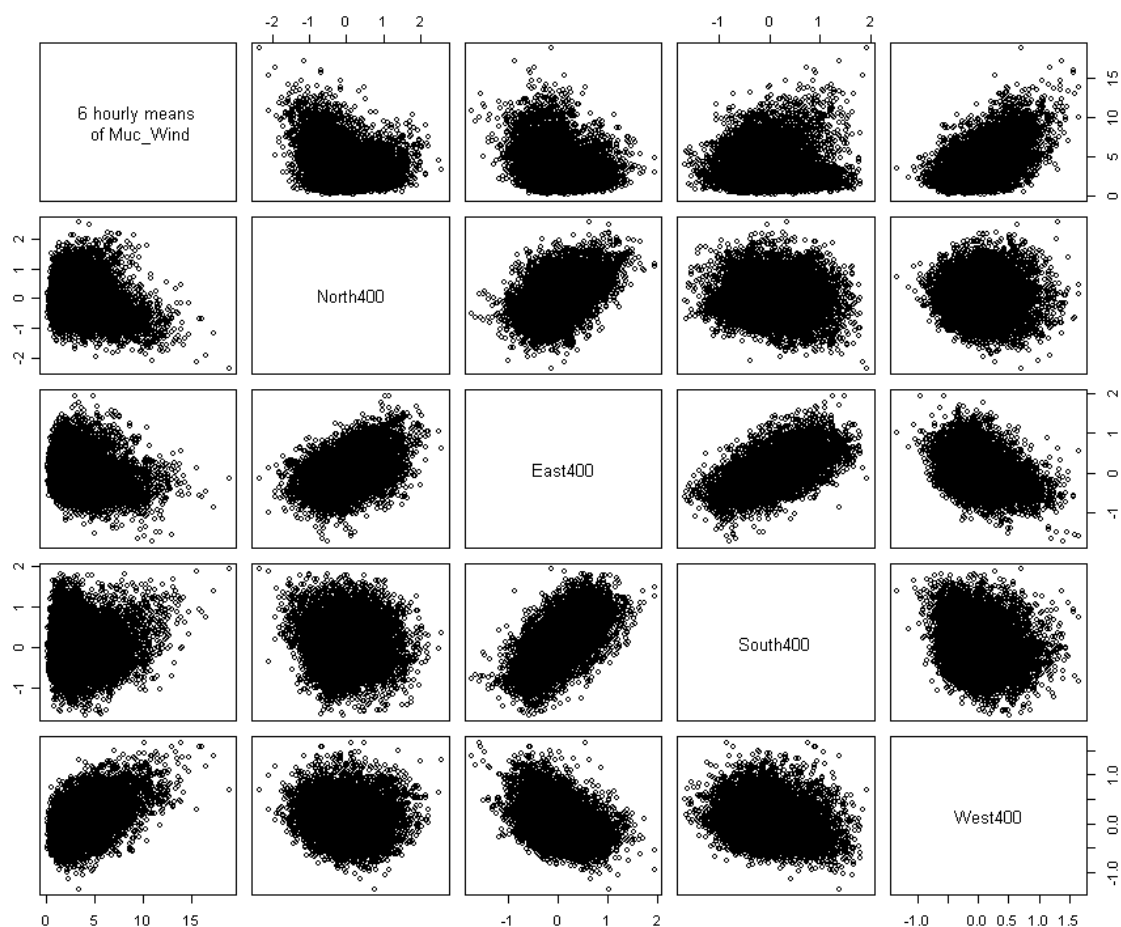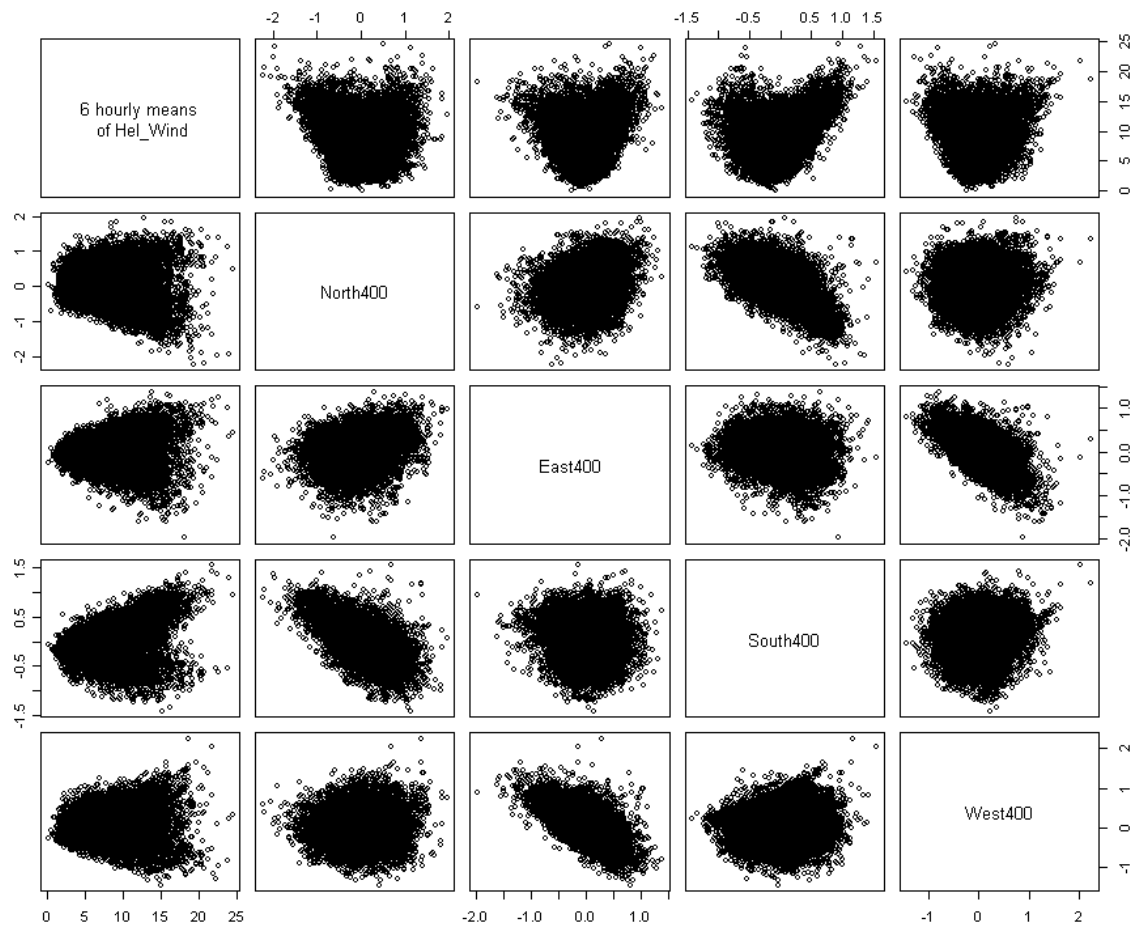
Figure 5.3: Pairs plot for wind speeds in **Munich** and standardized surface pressures 400 km away in every cardinal direction.

correlations between deseasonalized wind speeds and the standardized surface pressure differences were considered (Figures 5.3 and 5.4).

For Munich the highest absolute correlation is generated at westerly direction. The correlation has a value of 0.550. Northerly and easterly direction have both values of about -0.2, southerly direction just of 0.022.

The highest correlation being in westerly direction does not surprise, because in the system of global atmospheric circulation central Europe is in the cell of west winds, the Ferrel or mid-latitude cell with the prevailing westerlies. Major air circulations pump from the equator the heated air, which rises and causes high pressure, to the poles with cool air and lower pressure. As earth is a rotating reference frame, the coriolis effect applies. In the northern hemisphere the effect causes a deflection to the right of the motion of the object. This means for local areas of low and high

Figure 5.4: Pairs plot for wind speeds on **Helgoland** and standardized surface pressures 400 km away in every cardinal direction.
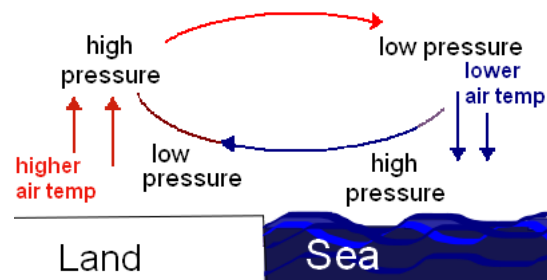
pressure, the pressure differences create equipondering wind movements which rotate counterclockwise around the low pressure point. The global wind currents are deflected westwards in the northern hemisphere to form the prevailing westerlies (cf. Etling [2008] and Quaschning [2010]).

Keeping the variable of surface pressure differences like a regime-switching factor, the direction with the highest correlation is selected as currents regime. So just the surface pressure differences of western direction are chosen, which also avoids multicollinearity.

Helgoland has the highest correlation of 0.333 between wind speed and surface pressure differences in southerly direction. Correlation between wind speed and the variable in westerly direction is lower at 0.026. In addition easterly direction has also a positive correlation of 0.145. In northerly direction the correlation is negative with the value -0.118, matching the positive correlation in southerly direction. Hence surface pressure differences in southerly direction are used for the model. Keeping in mind the location of Helgoland in the North sea, huge landmass places mainly south and a little bit in the east. Recalling the local wind currents phenomena of sea- or onshore-breeze, the result seems to match.

Sea breeze is given rise by the development of pressure fields due to unequal air temperature. As sea is able to absorb more heat than land, the land's surface warms up faster during day. The warm air has lower dense and starts rising, which causes lower pressure on surface and higher pressure above surface.

The compensating motion moves towards the sea with lower air pressure above the sea. The lower air pressure is caused by cooler air which has higher dense. This cooler air sinks, causing higher pressure on the sea surface. The compensating motion, the sea breeze, moves towards the land. At night land becomes cool faster than the sea and this cycle is reversed. Since the differences in temperature are impor-

Figure 5.5: Scheme of sea breeze during day.

tant, sea breeze does not necessarily depend on high surface temperature. This coincides with the diurnal pattern of Helgoland. Lowest wind speeds are described in the afternoon. At this time differences in temperature between land and sea have assimilated. During night differences in temperature do not change so much, therefore the reversed local wind currents stay at the same range, too. In Figure 5.5 the process is schematized.

Mountain breeze is a similar phenomenon caused by higher absorbtion of heat in

the valley. Cold air streams down the mountain slopes and heated air of the valley rises up to increase air pressure. In the afternoon differences of temperature have assimilated, which coincides with the diurnal pattern of Brocken and Fichtelberg. The smallest wind speeds are observed in the afternoon. As regime for Brocken, southern pressure differences are selected with a correlation of 0.45. For Fichtelberg northern differences are selected with a slightly higher absolute correlation than southern ones. By involving surface pressure differences we include phenomena like sea breeze, mountain breeze or other dominant local winds in our model naturally. Further information to local winds can be found in Simpson [1994] and Malberg [1997].

Since pressure differences are available only every six hours, a mean of six consecutive hours of the deseasonalized wind speeds Y_i is chosen as values of the new variable $\overline{Y}_t$.

$$\overline{Y}_t = \frac{1}{6} \sum_{i=6t-5}^{6t} Y_i$$

These variables $\overline{Y}_t$ are influenced by the standardized surface pressure difference Z_t, as we have seen above.

$$\overline{Y}_t = \beta_1 Z_t + u_t \tag{5.1}$$

Again the autocorrelation function of the resulting residuals u_t is inspected (cf. Figure 5.6). The shape with high autocorrelations for small lags and a rather quick decreasing for increasing lags suggests an autoregressive process of first or second order.
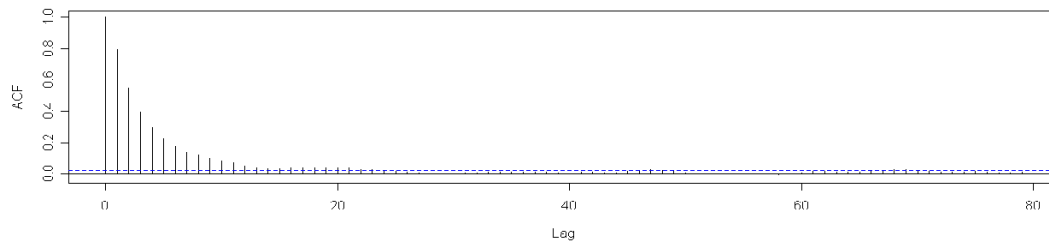
The autoregressive process of second order is built with the residuals u_t:

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + v_t$$

Comparing the autocorrelation functions of the residuals of an AR(1) process (Figure 5.7) and an AR(2) process (Figure 5.8), the values for AR(2) for small lags are even smaller. Hence an AR(2) process is considered and incorporated into the model for wind speed.

In Figure 5.8 are still some autocorrelations outside the confidence interval. Additionally, we have a look at the the autocorrelation function of the squared residuals v_t^2 in Figure 5.9. From the retained structure of the autocorrelation function, a GARCH(1,1) process is considered for the v_t:

$$v_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 v_{t-1}^2$$

Figure 5.6: Autocorrelation function of the variable u_t for Helgoland. The variable u_t origins from a regression of standardized western surface pressure differences to the variable of the deseasonalization regression in six hourly resolution as response. The autocorrelation function of the variable u_t for Munich has basically the same appearance. This shape seems to suggest an autoregressive process of first or second order.(e.g. Stralkowski.1970)
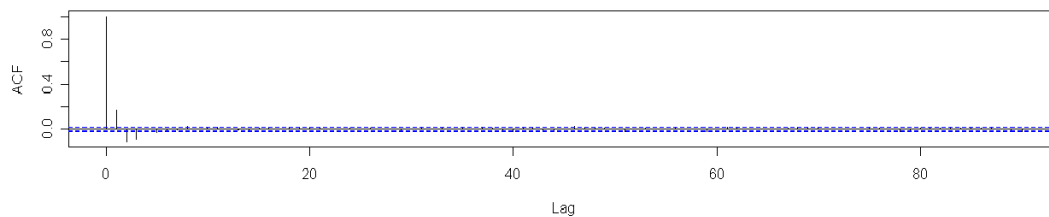
Figure 5.7: Autocorrelation function of the variable v_t of the AR(1) process of u_t for Helgoland. The autocorrelations for small lags outside the confidence interval indicate, that there is some more autocorrelation at the data.
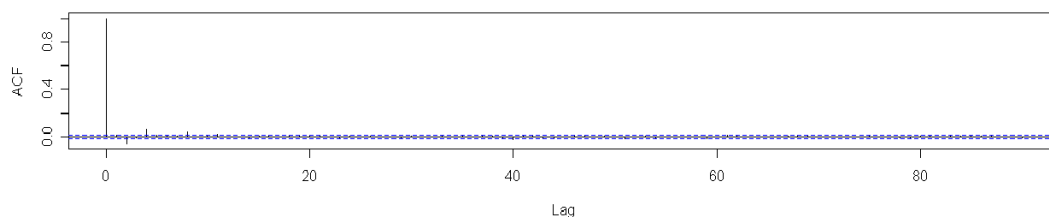
Figure 5.8: Autocorrelation function of the variable v_t of the AR(2) process of u_t for Helgoland. The autocorrelations for small lags are even smaller. That is the reason why we decided to incorporate an AR(2) process at our model.

Figure 5.9: Autocorrelation function of v_t^2.

Therefore we arrive at the following model

$$W_i \;\; = \;\; \gamma_0 + \gamma_1 d_i + \gamma_2 a_i + Y_i,$$

and, switching to a six hourly resolution by defining $\overline{Y}_t = \frac{1}{6} \sum_{i=6t-5}^{6t} Y_i,$

$$
\begin{aligned}
\overline{Y}_t &= \beta_1 Z_t + u_t \\
u_t &= \phi_1 u_{t-1} + \phi_2 u_{t-2} + v_t \\
v_t &= \sigma_t \varepsilon_t \\
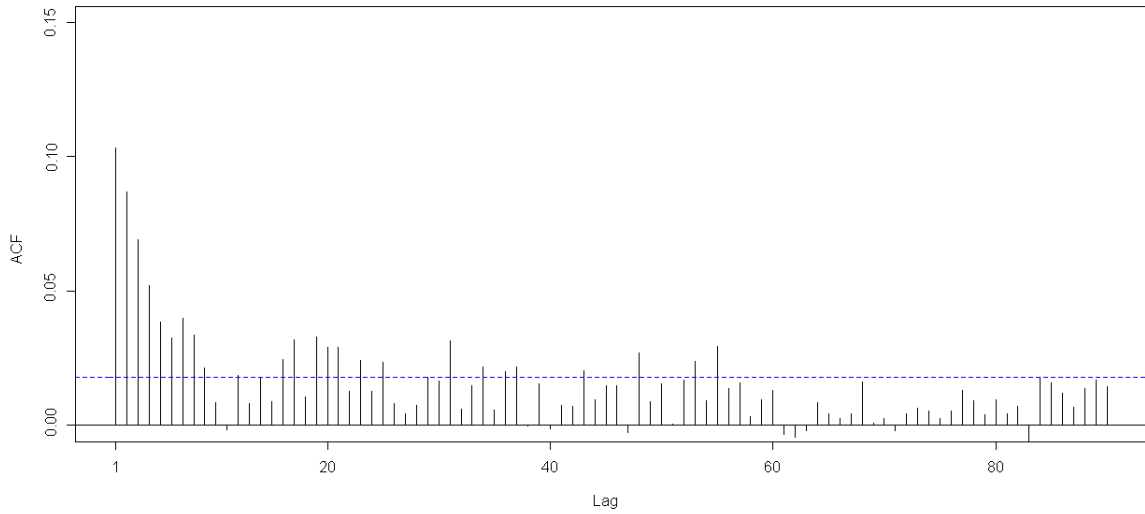\sigma_t^2 &= \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 v_{t-1}^2
\end{aligned}
$$

for $t = 1$, $u_{-1} := 0, u_0 := 0, v_0 := 0$ and $\sigma_0^2 := \frac{\alpha_0}{1-\alpha_1}$.

The variables of the model are

W_i:	wind speed (hourly resolution)
d_i:	diurnal seasonality component (hourly resolution) (cf. Figures 4.19, 4.20, 4.21, 4.22 and 4.21)
a_i:	annual seasonality component (houly resolution) (cf. Figures 4.14, 4.15, 4.16, 4.17 and 4.18)
Y_i:	deseasonalized wind speed (hourly resolution)
$\overline{Y}_t$:	deseasonalized wind speed (six hourly resolution)
Z_t:	standardized differences of surface pressure (six hourly resolution)
u_t:	AR(2) process
v_t:	GARCH(1,1) process

The transition from time scale i (one hour per step) to time scale t (six hours per step) is necessary as surface pressure differences are available only every six hours. Therefore, the model keeps an hourly base before using surface pressure differences. The transition to six hourly values takes place at the point where one uses surface pressure differences. At the point of using surface pressure differences the transition to six hourly values takes place. A maximum likelihood estimation was implemented in MATLAB for the process that was described above. For reasons of identifiability, constraints on ε_t have to be imposed. Moreover, the distribution of the variable ε_t must be fixed for the maximum likelihood estimation.

5.1   Model with Normal Distribution

We first assume that $(\varepsilon_t)_{t\in\mathbb{N}}$ is a sequence of normally distributed random variables with mean 0 and standard deviation 1, $\varepsilon_t \sim N(0,1)$. This implies that $v_t|\sigma_t$ is $N(0,\sigma_t^2)$ distributed.

Transforming Equation (5.1) of surface pressure differences to obtain the variable u_t, we get $u_t = \overline{Y}_t - \beta_1 Z_t$. To obtain the residuals v_t of the AR(2) process, we rewrite the process to $v_t = u_t - \phi_1 u_{t-1} - \phi_2 u_{t-2}$. These expressions can be used to rewrite the GARCH process v_t in the following manner:

$$v_t = [\overline{Y}_t - \beta_1 Z_t] - \phi_1[\overline{Y}_{t-1} - \beta_1 Z_{t-1}] - \phi_2[\overline{Y}_{t-2} - \beta_1 Z_{t-2}].$$

Let $\tilde{\mu}_t$ be an element of $\mathbb{R}$, $\tilde{\mu}_t := \beta_1 Z_t + \phi_1[\overline{Y}_{t-1} - \beta_1 Z_{t-1}] + \phi_2[\overline{Y}_{t-2} - \beta_1 Z_{t-2}]$, then

$$\overline{Y}_t = v_t + \tilde{\mu}_t. \tag{5.2}$$

Consequently, $\overline{Y}_t|\sigma_t$ is $N(\tilde{\mu}_t, \sigma_t^2)$ distributed.

For Munich we obtain the following model:

$$
\begin{aligned}
W_i &= 2.9302 + 0.5363 d_i - 0.0251 a_i + Y_i \\
\overline{Y}_t &= 3.6239 Z_t + u_t \\
u_t &= 0.7716 u_{t-1} - 0.0592 u_{t-2} + v_t \\
v_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= 0.2668 + 0.7027 \sigma_{t-1}^2 + 0.0588 v_{t-1}^2
\end{aligned}
$$

For Helgoland we get:

$$
\begin{aligned}
W_i &= 8.0214 + 0.0843 d_i + 1.2275 a_i + Y_i \\
\overline{Y}_t &= 2.9406 Z_t + u_t \\
u_t &= 0.9691 u_{t-1} - 0.2049 u_{t-2} + v_t \\
v_t &= \sigma_t \varepsilon_t \\
\sigma_t^2 &= 0.5477 + 0.6699 \sigma_{t-1}^2 + 0.0249 v_{t-1}^2
\end{aligned}
$$

In Table 5.1 the estimated parameters and corresponding standard errors are shown for the monitoring stations at Munich and on Helgoland.

Weather station	Munich	Helgoland
γ_0	2.9302	8.0214
(se)	(0.2605)	(1.1487)
γ_1	0.5363	0.0843
(se)	(0.0159)	(0.1242)
γ_2	-0.0251	1.2275
(se)	(0.0817)	(0.0526)
β_1	3.6239	2.9406
(se)	(0.0807)	(0.1380)
ϕ_1	0.7716	0.9691
(se)	(0.0103)	(0.0097)
ϕ_2	-0.0592	-0.2049
(se)	(0.0102)	(0.0096)
α_0	0.2668	0.5477
(se)	(0.0185)	(0.0580)
α_1	0.7027	0.6699
(se)	(0.0185)	(0.0326)
α_2	0.0588	0.0249
(se)	(0.0036)	(0.0020)

Table 5.1: Estimates and the corresponding standard errors for the model with normally distributed wind speeds Y_t.

A histogram of the residuals ε_t is shown in Figure 5.10 for Munich. The red dashed line is the density of the normal distribution with mean 0 and standard deviation 1. Obviously, the distribution of the ε_t does not coincide with the normal distribution. The plotted histogram gives the impression that a distribution with more shape parameters is needed. The distribution should be able to change skewness and kurtosis depending on shape parameters.

For Helgoland the corresponding histogram of ε_t with a red dashed line for the density of the normal distribution can be seen in Figure 5.11. The assumption of normally distributed ε_t's does not hold either. The deseasonalized wind speeds obviously require the incorporation of a different distribution.

As we checked, also the Weibull distribution is not suitable to capture the structure of the data. In the end, we selected an hyperbolic distribution which is described in detail in Section 2. The model with hyperbolic distribution is illustrated in the following Section 5.2.
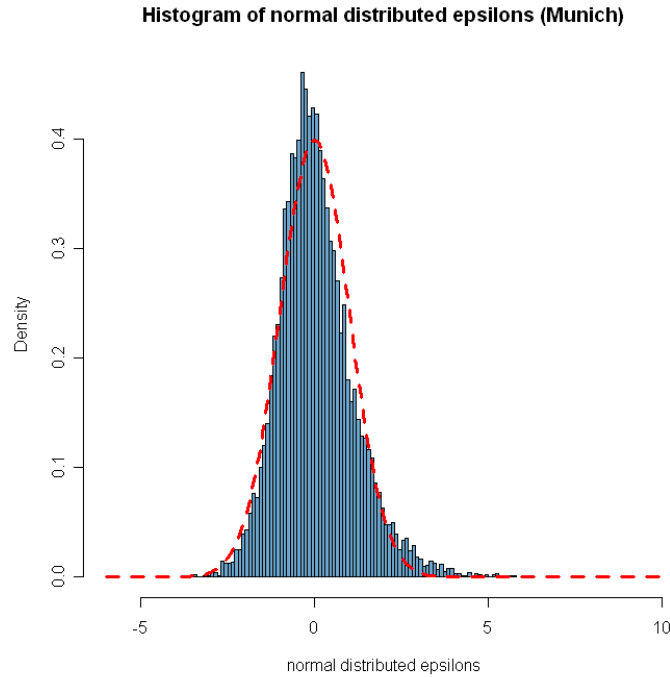
Figure 5.10: Histogram of resulting ε_ts for the wind speed modelling at Munich, assuming a normal distribution with mean 0 and standard deviation 1. The red dashed line is the distribution function of the Normal distribution for mean 0 and standard deviation 1. The graph does not fit the histogram exemplarily, some skewness seems to be observed.
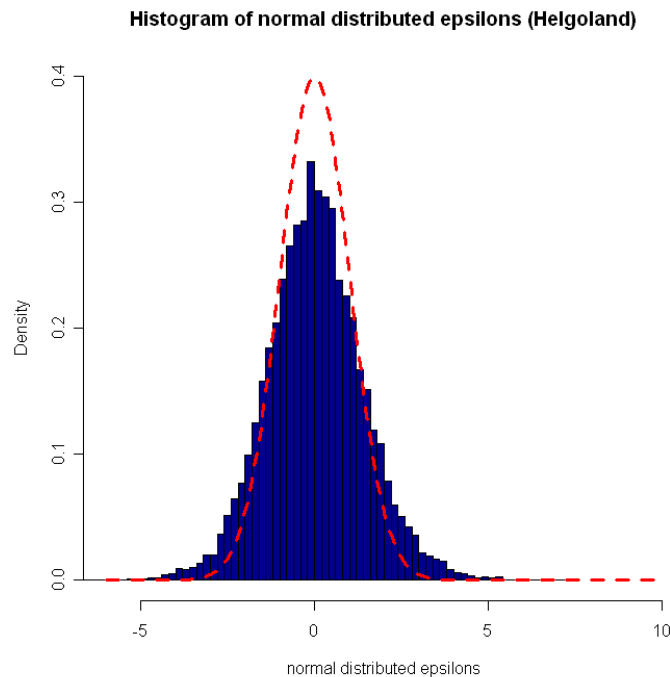
Figure 5.11: Histogram of resulting ε_ts for the wind speed modelling at Helgoland, assuming a normal distribution with mean 0 and standard deviation 1. The red dashed line is the distribution function of the Normal distribution for mean 0 and standard deviation 1. The graph does not fit the histogram exemplarily, the observed kurtosis seems to be different.

5.2   Model with Hyperbolic Distribution

Analogous to the model with the assumption of normal distributed wind speeds Y_t, we use the same model structure but assume a hyperbolic distribution for Y_t. Skewness and kurtosis of the hyperbolic distribution depend on the shape parameter ξ and χ (cf. Section 2). Having two more parameters than the normal distribution, the hyperbolic distribution is, of course, much more flexible. We recall the construction of the model

$$W_i = \gamma_0 + \gamma_1 d_i + \gamma_2 a_i + Y_i$$

and, switching to a six hourly resolution by defining $\overline{Y}_t = \frac{1}{6} \sum_{i=6t-5}^{6t} Y_i$,

$$\overline{Y}_t = \beta_1 Z_t + u_t$$
$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + v_t$$
$$v_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 v_{t-1}^2$$

for $t = 1$, $u_{-1} := 0, u_0 := 0, v_0 := 0$ and $\sigma_0^2 := \frac{\alpha_0}{1-\alpha_1}$.

The variables of the model are

W_i: wind speed (hourly resolution)
d_i: diurnal seasonality component (hourly resolution) (cf. Figures 4.19, 4.20, 4.21, 4.22 and 4.21)
a_i: annual seasonality component (houly resolution) (cf. Figures 4.14, 4.15, 4.16, 4.17 and 4.18)
Y_i: deseasonalized wind speed (hourly resolution)
$\overline{Y}_t$:   deseasonalized wind speed (six hourly resolution)
Z_t: standardized differences of surface pressure (six hourly resolution)
u_t: AR(2) process
v_t: GARCH(1,1) process

With Equation (5.2) $Y_t = v_t + \tilde{\mu}_t$ and $v_t = \sigma_t \varepsilon_t$,

$$\varepsilon_t = \frac{\overline{Y}_t - \tilde{\mu}_t}{\sigma_t}.$$

So is also $\overline{Y}_t$ hyperbolically distributed, characterized by the relation

$$f_{HYP(Y)}(Y_t) = f_{HYP(\varepsilon)}(\varepsilon(Y_t)) \frac{d\varepsilon(Y_t)}{dY_t} = f_{HYP(\varepsilon)}(\varepsilon(Y_t)) \frac{1}{\sigma_t}.$$

The function $f_{HYP(Y)}$ represents the hyperbolic density function for Y_t and $f_{HYP(\varepsilon)}$

is the density function for $\varepsilon_t \sim hyperbol(\xi, \chi, \delta, \mu)$. Additional constraints have to be defined to get an identifiable model. Therefore we assume $E[\varepsilon_t] \stackrel{!}{=} 0$. This assumption determines the location parameter μ, which can be calculated by

$$\mu = -\frac{\delta}{\sqrt{\xi^2 - \chi^2}} \frac{K_2(\zeta)}{K_1(\zeta)}.$$

Furthermore a value for the variance has to be assumed. By reason of numerical stability

$$Var(\varepsilon_t) = \frac{\delta^2 K_2(\zeta)}{\zeta K_1(\zeta)} + \frac{\delta^2}{\frac{\xi^2}{\chi^2} - 1} \Big(\frac{K_3(\zeta)}{K_1(\zeta)} - \frac{K_2(\zeta)^2}{K_1(\zeta)^2} \Big) \stackrel{!}{=} 4$$

is assumed. This condition initiates a separate optimization of δ, which uses least square optimization to find a solution.

Achieving δ by a single optimization depending on the model parameter, the parameters of the whole model are estimated by maximum likelihood optimization consistent with the δ. In Table 5.2 the achieved parameter estimates for the hyperbolically distributed wind speed model are shown.

 The intercept of the model, γ_0 could be interpreted as an average wind speed indicating the range of wind speed at the respective location. The parameter that scales the daily component is γ_1, which is quite similar for Brocken and Fichtelberg, as well as for Munich and Trier. For Helgoland the parameter of the daily component is relatively small compared to the other locations, which induces no huge effects of daily seasonality. For the parameter of the annual seasonality component, γ_2, a small value is estimated for Munich, inducing no huge effects of annual seasonality. The scale factor of surface pressure differences, β_1, depends on the covariate. As measurements and obtained surface differences differ for the observed locations, the values are non-uniform. The negative algebraic sign at Fichtelberg is caused by the selection of the direction of surface pressure differences. Recalling from the model description at the beginning of this chapter, the direction with the highest correlation is selected. For Fichtelberg the selected direction points at an area with lower surface pressure, which causes negative values for the pressure difference covariate Z.

The parameters of the AR(2) process, ϕ_1 and ϕ_2 show similar behavior on the one hand for the cities Munich and Trier and on the other hand for the mountains Brocken and Fichtelberg, as well the island Helgoland. The autoregressive parameter of first order, ϕ_1 is quite high for all locations. The autoregressive parameter of second order, ϕ_2, is rather small for the cities, but still significant. The mountain group has higher ϕ_2, containing the island Helgoland with the highest one of all observed locations. All locations have in common the negative influence of the second autoregressive parameter.

For the GARCH(1,1) parameter, α_0, α_1 and α_2, the locations again seem to be groupable to the cities fraction and the mountains fraction, which contains the

| Weather station	Munich	Helgo-land	Brocken	Fichtel-berg	Trier
γ_0	2.8061	8.2456	10.7005	9.0313	3.1376
(se)	(0.2017)	(1.1303)	(0.8911)	(0.7669)	(0.0377)
γ_1	0.4986	0.0712	0.2952	0.2993	0.4327
(se)	(0.0146)	(0.1213)	(0.0634)	(0.0562)	(0.0091)
γ_2	-0.0514	1.2730	1.9523	1.5831	0.4498
(se)	(0.0626)	(0.0531)	(0.0524)	(0.0619)	(0.0348)
β_1	2.7837	2.9656	4.0762	-1.4173	1.3022
(se)	(0.0793)	(0.1399)	(0.1207)	(0.1718)	(0.0709)
ϕ_1	0.7165	0.9844	0.9295	0.9825	0.7888
(se)	(0.0099)	(0.0097)	(0.0096)	(0.0095)	(0.0098)
ϕ_2	-0.0479	-0.2097	-0.1490	-0.1745	-0.0711
(se)	(0.0091)	(0.0095)	(0.0095)	(0.0094)	(0.0094)
α_0	0.0641	0.1458	0.0623	0.0400	0.0458
(se)	(0.0107)	(0.0378)	(0.0214)	(0.0128)	(0.0082)
α_1	0.6397	0.7620	0.9058	0.9253	0.7376
(se)	(0.0265)	(0.0397)	(0.0203)	(0.0142)	(0.0221)
α_2	0.0434	0.0215	0.0131	0.0109	0.0342
(se)	(0.0069)	(0.0045)	(0.0031)	(0.0025)	(0.0056)
ξ	0.5419	0.4543	0.3952	0.3974	0.4973
(se)	(0.0214)	(0.0230)	(0.0244)	(0.0239)	(0.0215)
χ	0.2365	0.0274	0.0534	0.0524	0.1730
(se)	(0.0094)	(0.0086)	(0.0085)	(0.0085)	(0.0091)
3χ	0.7095	0.0822	0.1602	0.1572	0.5190
$3\xi^2$	0.8810	0.6192	0.4685	0.4738	0.7419

Table 5.2: Parameter estimates for the different monitoring stations, assuming hyperbolically distributed deseasonalized wind speeds Y_t.

monitoring station on the island Helgoland. The estimates for the hyperbolic distribution parameters affirm this impression of two groups with different behavior fr wind speeds. As 3χ is an approximation for skewness, the skewness at the observation points Munich and Trier is within the same range, which is higher than for the mountain's group. Kurtosis, which is approximated by $3\xi^2$ is again higher for Munich and Trier. The mountains Brocken and Fichtelberg have identical values up to two decimals although being more than 300 kms apart.

Controlling for any remaining influence, autocorrelation functions of ε_t and of ε_t^2 are considered. In Figure 5.12 and 5.13 almost no structure can be observed. The fit of the model is reflected by the histograms of ε_t, too (cf. Figure 5.14 for Munich, 5.15 for Helgoland, 5.16 for Brocken, 5.17 for Fichtelberg and 5.18 for Trier). Also the shape of the QQ-plot (cf. Figure 5.20) supports the assumption of hyperbolic distribution.
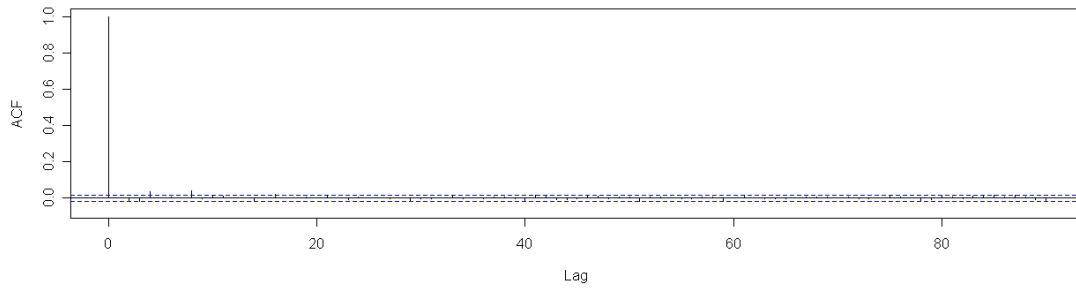
**Figure 5.12:** ACF of ε_t at monitoring station on Helgoland
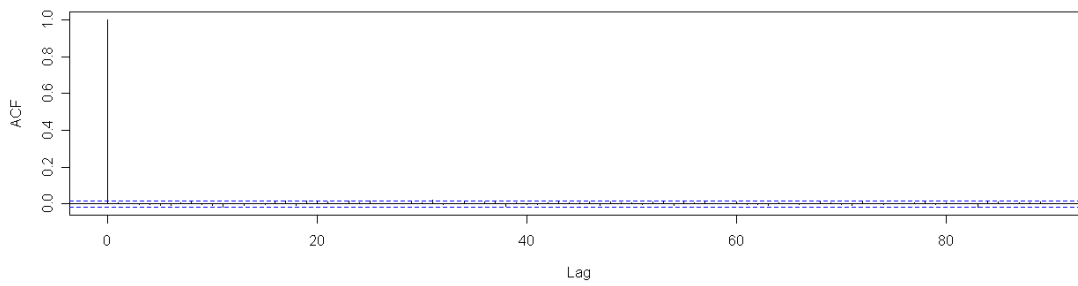
**Figure 5.13:** ACF of ε_t^2 at monitoring station on Helgoland
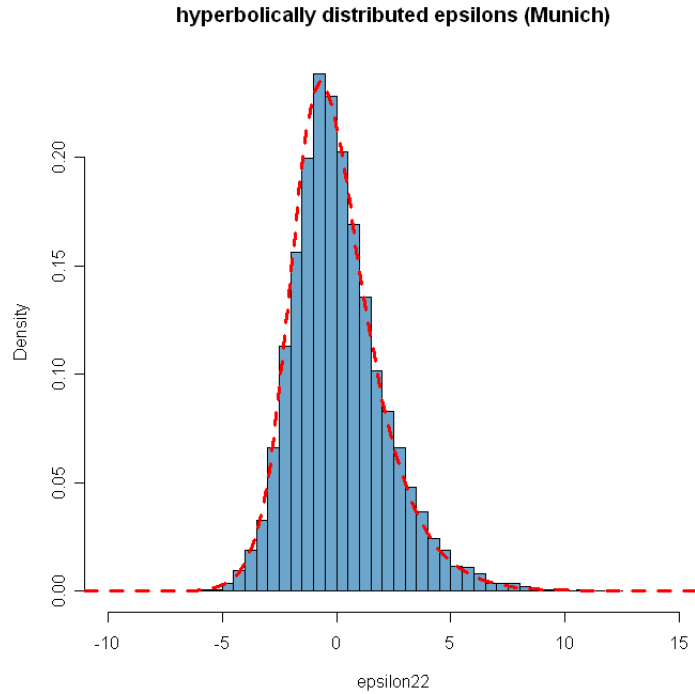
Figure 5.14: Histogram of resulting ε_ts for the wind speed measurements in **Munich**, assuming a hyperbolic distribution with $hyperbol(0.5419, 0.2365, 2.0774, -1.6828)$. The red dashed line is the distribution function of the hyperbolic distribution for $hyperbol(0.5419, 0.2365, 2.0774, -1.6828)$.
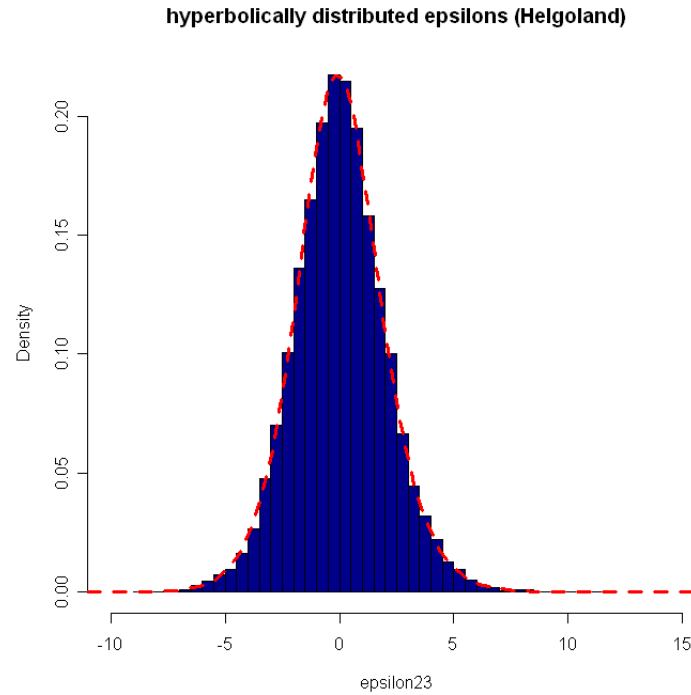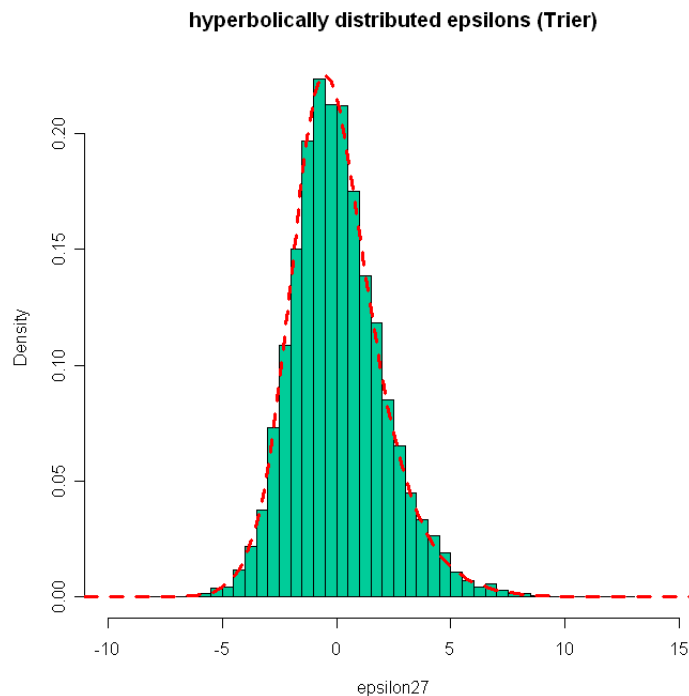
Figure 5.15: Histogram of resulting ε_ts for the wind speed measurements on **Helgoland**, assuming a hyperbolic distribution with $hyperbol(0.4543, 0.0274, 3.2944, -0.28038)$. The red dashed line is the distribution function of the hyperbolic distribution for $hyperbol(0.4543, 0.0274, 3.2944, -0.28038)$.
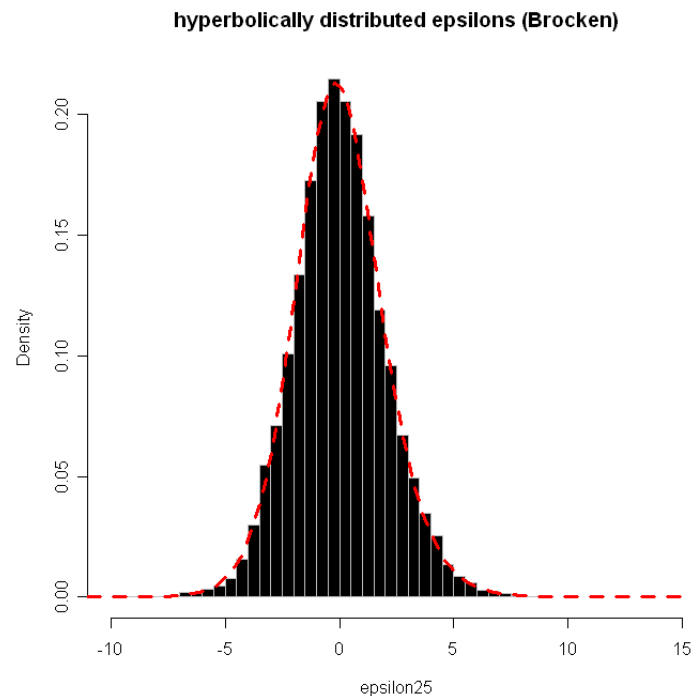

Figure 5.16: Histogram of resulting ε_ts for the wind speed measurements at **Brocken**, assuming a hyperbolic distribution with $hyperbol(0.3952, 0.0534, 4.0494, -0.71163)$. The red dashed line is the distribution function of the hyperbolic distribution for $hyperbol(0.3952, 0.0534, 4.0494, -0.71163)$.
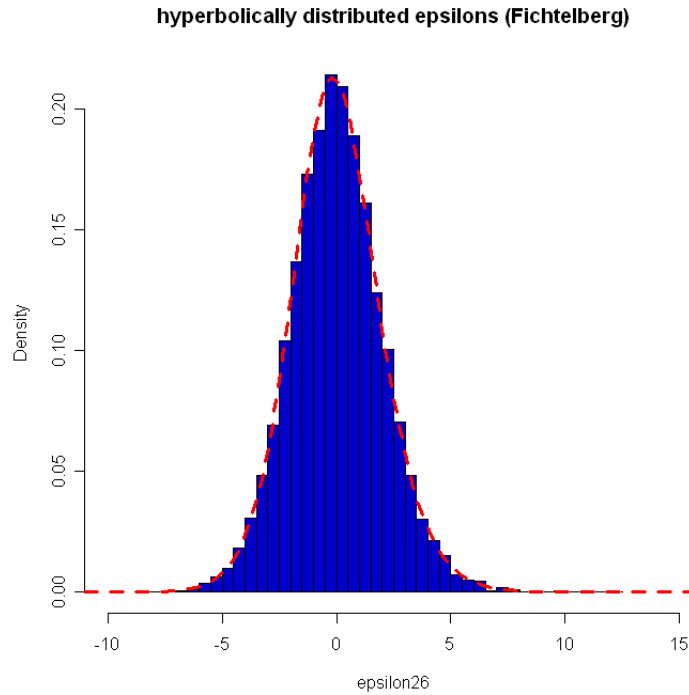
Figure 5.17: Histogram of resulting ε_ts for the wind speed measurements at **Fichtelberg**, assuming a hyperbolic distribution with $hyperbol(0.3974, 0.0524, 4.0185, -0.69071)$. The red dashed line is the distribution function of the hyperbolic distribution for $hyperbol(0.3974, 0.0524, 4.0185, -0.69071)$.


Figure 5.18: Histogram of resulting ε_ts for the wind speed measurements at **Trier**, assuming a hyperbolic distribution with $hyperbol(0.4973, 0.1730, 2.5942, -1.4668)$. The red dashed line is the distribution function of the hyperbolic distribution for $hyperbol(0.4973, 0.1730, 2.5942, -1.4668)$.
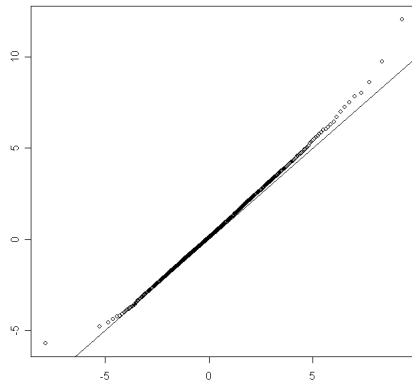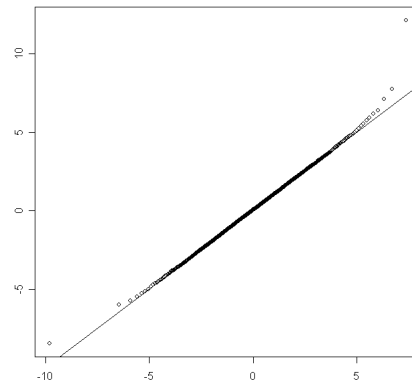
Figure 5.19: A QQ-plot comparing the quantiles of the set of ε_t at observation point **Munich** on the vertical axis to the hyperbolic theoretical quantiles on the horizontal axis.
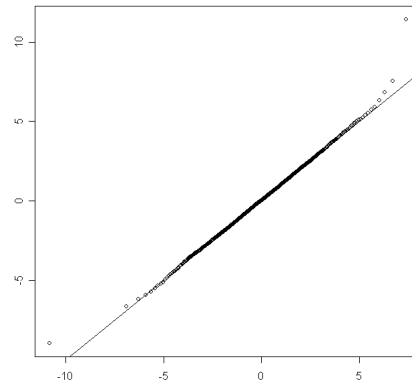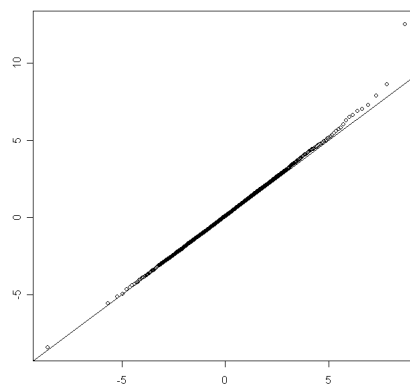

Figure 5.20: A QQ-plot comparing the quantiles of the set of ε_t at observation point **Helgoland** on the vertical axis to the hyperbolic theoretical quantiles on the horizontal axis.
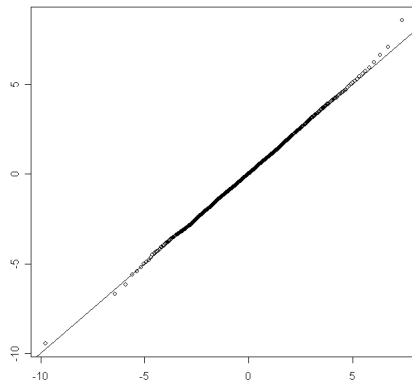
Figure 5.21: A QQ-plot comparing the quantiles of the set of ε_t at observation point **Brocken** on the vertical axis to the hyperbolic theoretical quantiles on the horizontal axis.

Figure 5.22: A QQ-plot comparing the quantiles of the set of ε_t at observation point **Fichtelberg** on the vertical axis to the hyperbolic theoretical quantiles on the horizontal axis.

Figure 5.23: A QQ-plot comparing the quantiles of the set of ε_t at observation point **Trier** on the vertical axis to the hyperbolic theoretical quantiles on the horizontal axis.

5.3 Forecasts

5.3.1 Forecasts Based on Seasonal Components Only

The simplest way of forecasting by means of our model is to use only the diurnal and annual seasonality components of the linear regression. No difference occurs between one-step and multi-step predictors, since forecasts are based only on the long-term average at the time under consideration, but not on specific wind speeds directly before.

$$\widehat{\overline{W}}_{t+1} = \widehat{\gamma_0} + \widehat{\gamma_1} \cdot \frac{1}{6} \sum_{i=6(t+1)-5}^{6(t+1)} d_i + \widehat{\gamma_2} \cdot \frac{1}{6} \sum_{i=6(t+1)-5}^{6(t+1)} a_i$$

The mean squared error (MSE) is

$$\text{MSE}\left(\widehat{\overline{W}}_n, \overline{W}_t\right) = \frac{1}{n} \sum_{t=1}^{n} \left(\widehat{\overline{W}}_t - \overline{W}_t\right)^2$$

where n indicates the number of observations taken into account. The true values are $\overline{W}_t = \frac{1}{6} \sum_{i=6t-5}^{6t} W_i$, i.e. the six hourly means of the hourly resolved wind speeds. For our data set $n = 11680$. The resulting mean squared error of Helgoland is $\text{MSE}\left(\widehat{\overline{W}}_{11680}, \overline{W}_t\right) = 12.10$.

5.3.2 Forecasts Based on Seasonal Components and Surface Pressure Differences

Using the diurnal and annual seasonality components of the linear regression and an additional covariate, the surface pressure differences, we get a better fit, but still nearly no difference exists between one-step and two-step predictors. Seasonality components are known constants and surface pressure differences are measurements which can easily be predicted at a time scale of six or twelve hours. For multiple-step predictors, estimation of surface pressure difference might be more difficult and vague. Therefore the MSE for higher multiple-step predictions gets worse.

$$\widehat{\widehat{\overline{W}}}_{t+1} = \widehat{\gamma_0} + \widehat{\gamma_1} \cdot \frac{1}{6} \sum_{i=6(t+1)-5}^{6(t+1)} d_i + \widehat{\gamma_2} \cdot \frac{1}{6} \sum_{i=6(t+1)-5}^{6(t+1)} a_i + \widehat{\widehat{\overline{Y}}}_{t+1}$$

$$\widehat{\widehat{\overline{Y}}}_{t+1} = \widehat{\beta_1} Z_{t+1}$$

For Helgoland we get $\text{MSE}\left(\widehat{\widehat{\overline{W}}}_{11680}, \overline{W}_t\right) = 11.08$

5.3.3 Forecasts Based on Seasonal Components, Surface Pressure Differences and AR(2) Process

Not only the seasonality components and the covariate are used, but also the autoregressive process is integrated in the prediction. As the variable u_t of the AR(2) process is unknown for future values, differences occur for each step-width, even for one- and two-step predictors. A one-step prediction can be calculated by

$$\widehat{\overline{\widehat{\widehat{W}}}}_{t+1}(6\text{h}) = \widehat{\gamma}_0 + \widehat{\gamma}_1 \cdot \frac{1}{6} \sum_{i=6(t+1)-5}^{6(t+1)} d_i + \widehat{\gamma}_2 \cdot \frac{1}{6} \sum_{i=6(t+1)-5}^{6(t+1)} a_i + \widehat{\overline{\widehat{\widehat{Y}}}}_{t+1}$$

$$\widehat{\overline{\widehat{\widehat{Y}}}}_{t+1} = \widehat{\beta}_1 Z_{t+1} + \widehat{\overline{\widehat{\widehat{u}}}}_{t+1}$$

$$\widehat{\overline{\widehat{\widehat{u}}}}_{t+1} = \widehat{\phi}_1 u_t + \widehat{\phi}_2 u_{t-1},$$

and a two-step prediction by

$$\widehat{\overline{\widehat{\widehat{W}}}}_{t+2}(12\text{h}) = \widehat{\gamma}_0 + \widehat{\gamma}_1 \cdot \frac{1}{6} \sum_{i=6(t+2)-5}^{6(t+2)} d_i + \widehat{\gamma}_2 \cdot \frac{1}{6} \sum_{i=6(t+2)-5}^{6(t+2)} a_i + \widehat{\overline{\widehat{\widehat{Y}}}}_{t+2}$$

$$\widehat{\overline{\widehat{\widehat{Y}}}}_{t+2} = \widehat{\beta}_1 Z_{t+2} + \widehat{\overline{\widehat{\widehat{u}}}}_{t+2}$$

$$\widehat{\overline{\widehat{\widehat{u}}}}_{t+2} = \widehat{\phi}_1 \widehat{\overline{\widehat{\widehat{u}}}}_{t+1} + \widehat{\phi}_2 u_t$$

$$\widehat{\overline{\widehat{\widehat{u}}}}_{t+1} = \widehat{\phi}_1 u_t + \widehat{\phi}_2 u_{t-1}.$$

Considering exemplarily Helgoland again, the MSE of a six hourly forecast is 3.84. For a two step prediction, a 12 hourly forecast, the MSE increases to 7.85. For these calculations we assumed a good prediction of future surface pressure differences, therefore we inserted the true values. Actually we need a prediction model for pressure differences to insert the predicted values. But even with the assumption of the last known period, which means $\widehat{\overline{\widehat{\widehat{Y}}}}_{t+1} = \widehat{\beta}_1 Z_t + \widehat{\overline{\widehat{\widehat{u}}}}_{t+1}$ for the one-step prediction, we get the reasonable results with a MSE $\left(\widehat{\overline{\widehat{\widehat{W}}}}_{11680}(6\text{h}, \text{Pr}_t^{t+1}), \overline{W}_t \right) = 4.03$. The symbol Pr_t^{t+1} indicates that for the prediction of period $t+1$ the known surface pressure differences of period t are assumed. For a two-step prediction $\widehat{\overline{\widehat{\widehat{Y}}}}_{t+2} = \widehat{\beta}_1 Z_t + \widehat{\overline{\widehat{\widehat{u}}}}_{t+2}$ when we use the last known pressure difference as future value. The MSE of the two-step prediction with pressure differences of period t, which is the value of the surface pressure differences two time steps backwards, is MSE $\left(\widehat{\overline{\widehat{\widehat{W}}}}_{11680}(12\text{h}, \text{Pr}_t^{t+2}), \overline{W}_t \right) = 8.11$ for Helgoland. The symbol Pr_t^{t+2} indicates that for the prediction of period $t+2$ the known surface pressure differences of period t are assumed.

The errors for each station and respective standard deviation can be found in Table 5.3. Therefore our predictions seem to be reasonable. The best forecast is always the one-step prediction based on seasonal components, surface pressure differences and an AR(2) process. For Munich, predicting only based on seasonal components and surface pressure differences leads to a better prediction as a two-step prediction based on seasonal components, surface pressure differences and an AR(2) process. Thus one can see, that wind speeds in Munich are highly dependent on surface pressure differences. Also for the wind speeds measured next to Trier (Figures 5.32, 5.33), the surface pressure differences seem to be more important than the AR(2) process. The two-step prediction based on the model with the AR(2) process is hardly better than the prediction based on seasonal components and surface pressure differences. This could be interpreted as a weaker dependency on local circulation or phenomenons. For Fichtelberg an adverse behavior is observed. The prediction with the AR(2) process is much better. This improvement of fit due to the AR(2) process indicates higher dependency of the local circulation. The surface pressure differences, which represent the wider surrounding, do not have such an huge impact. Considering the Figures 5.26 and 5.27 one can see the good impacts of surface pressure differences, whereas in Figures 5.30 and 5.31 the orange graph of the prediction with surface pressure differences hardly improves comparing to the forecast based on seasonal components only. For Brocken (Figures 5.28 and 5.29) the close predictions based on the model with AR(2) process can be seen. Generally, a similar behavior to Fichtelberg with local circulation dependencies is observed. However, the dependency on bigger circulation systems and therefore on surface pressure differences is a bit more distinctive for Brocken, which improves the fit.

| Weather station	Munich	Helgoland	Brocken	Fichtelberg	Trier
$\overline{W}_{11680}$	3.15	8.50	10.88	9.24	3.35
Std dev.	2.26	3.76	5.21	4.71	2.12
RMSE $\left(\widehat{\overline{W}}_{11680}\right)$	2.23	3.48	4.83	4.36	2.06
RMSE $\left(\widehat{\widehat{\overline{W}}}_{11680}\right)$	1.91	3.33	4.12	4.21	1.90
RMSE $\left(\widehat{\widehat{\overline{W}}}_{11680}(6\text{h})\right)$	1.53	2.17	2.44	2.27	1.23
RMSE $\left(\widehat{\widehat{\overline{W}}}_{11680}(12\text{h})\right)$	2.57	2.75	3.33	3.22	1.61

Table 5.3: Root mean squared errors for the different monitoring stations. For the forecasts in the last two rows a good prediction model for surface pressure differences is assumed. Therefore we calculated our wind speed forecast with the true surface pressure differences.
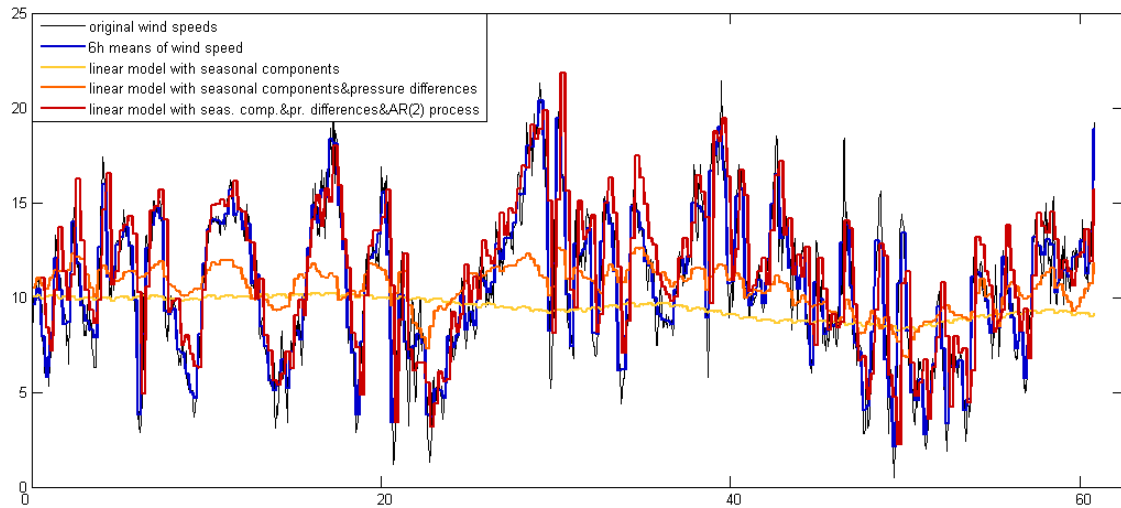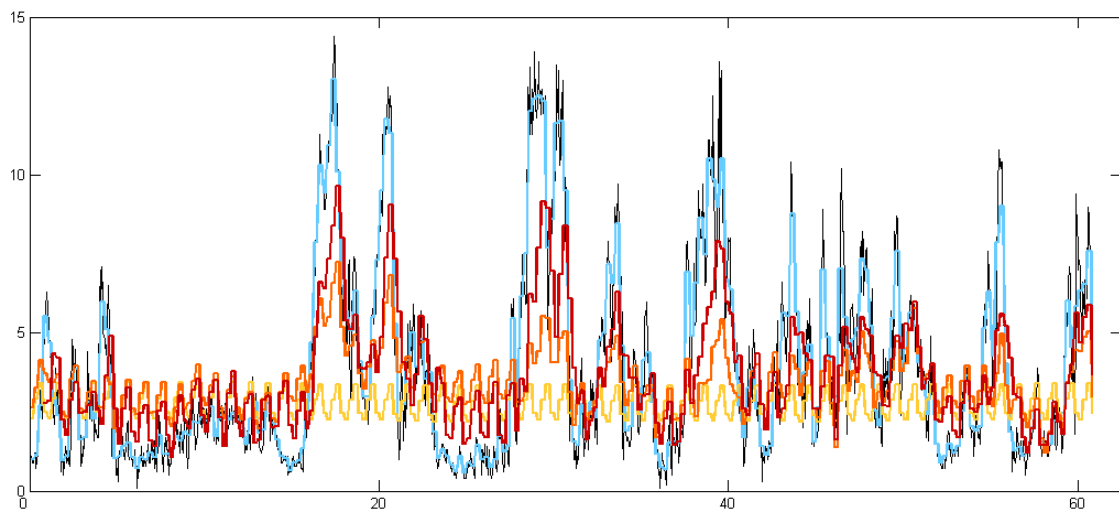
Figure 5.24: Forecast of six hourly wind speeds (one-step prediction) for **Helgoland**. The period of 60 days, starting January 1, 2000, is shown and x-axis is labeled per day, but plotted values are hourly values. The y-axis shows wind speeds in $\frac{m}{s}$. The thin black line are the originally measured wind speeds and the blue line are the six hourly means. The forecast of Subsection 5.3.1 is shown in yellow. The orange graph is the forecast described in Subsection 5.3.2. The red graph is the forecast described in Subsection 5.3.3.
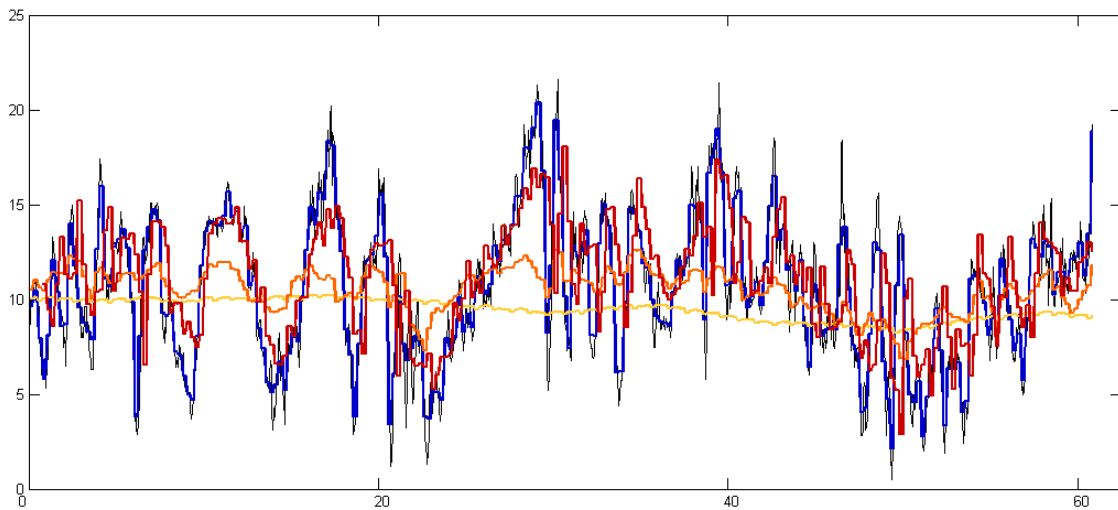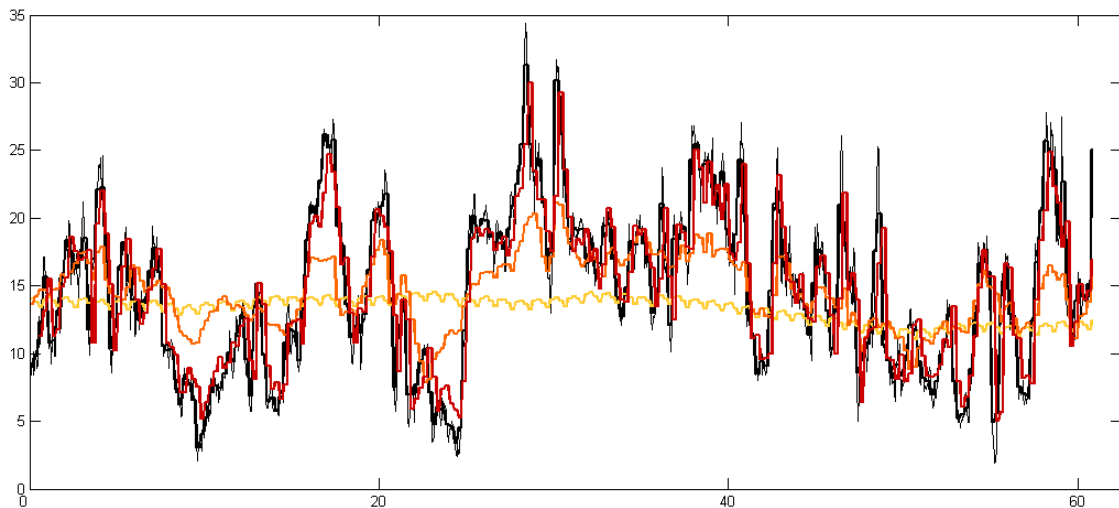

Figure 5.25: Forecast of 12 hourly wind speeds (two-step prediction) for **Helgoland**. The period of 60 days, starting January 1, 2000, is shown and x-axis is labeled per day, but plotted values are hourly values. The y-axis shows wind speeds in $\frac{m}{s}$. The thin black line are the originally measured wind speeds and the blue line are the six hourly means. The forecast of Subsection 5.3.1 is shown in yellow. The orange graph is the forecast described in Subsection 5.3.2. The red graph is the forecast described in Subsection 5.3.3.
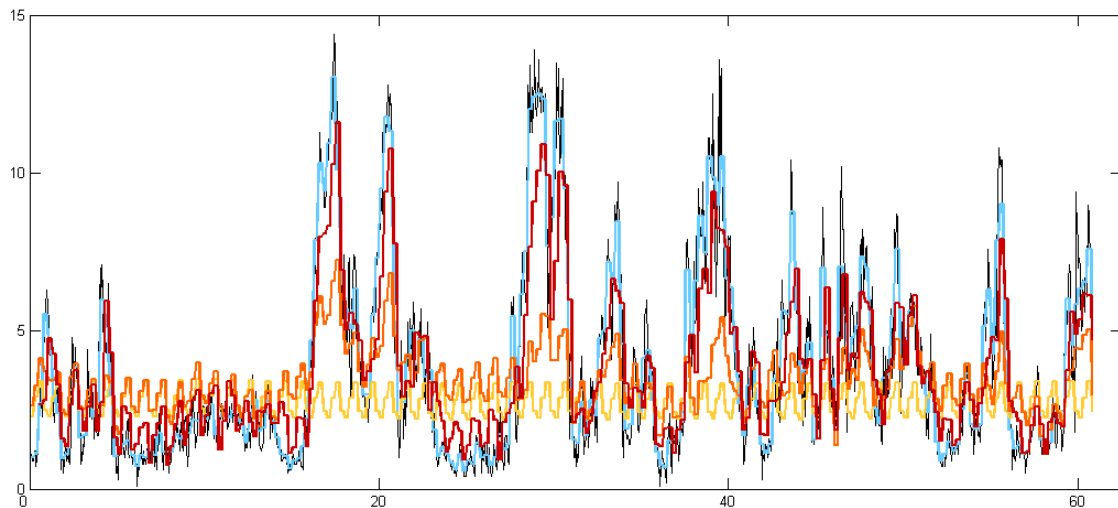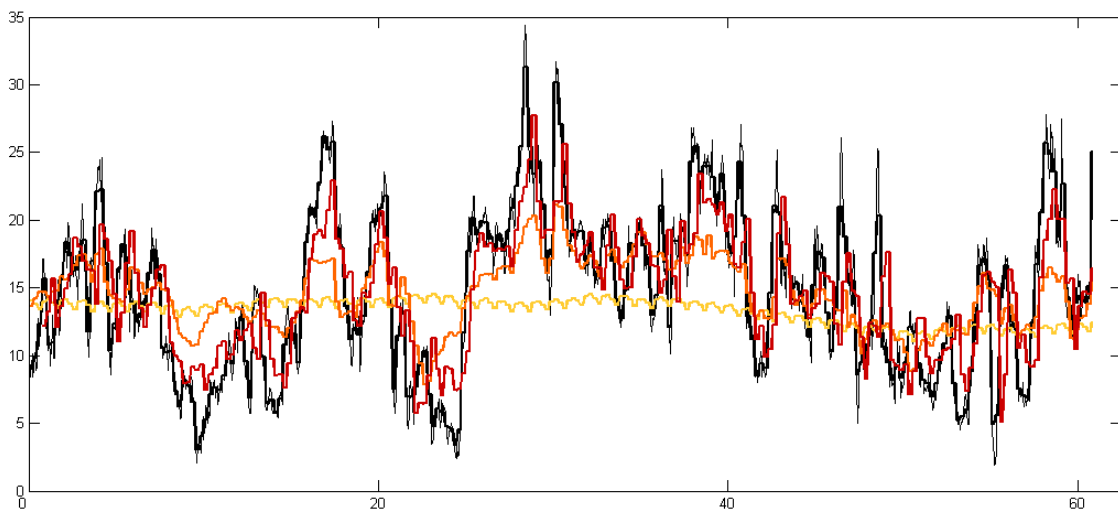
Figure 5.26: Forecast of six hourly wind speeds (one-step prediction) for **Munich**. The period of 60 days, starting January 1, 2000, is shown and x-axis is labeled per day, but plotted values are hourly values. The y-axis shows wind speeds in $\frac{\text{m}}{\text{s}}$. The thin black line are the originally measured wind speeds and the blue line are the six hourly means. The forecast of Subsection 5.3.1 is shown in yellow. The orange graph is the forecast described in Subsection 5.3.2. The red graph is the forecast described in Subsection 5.3.3.
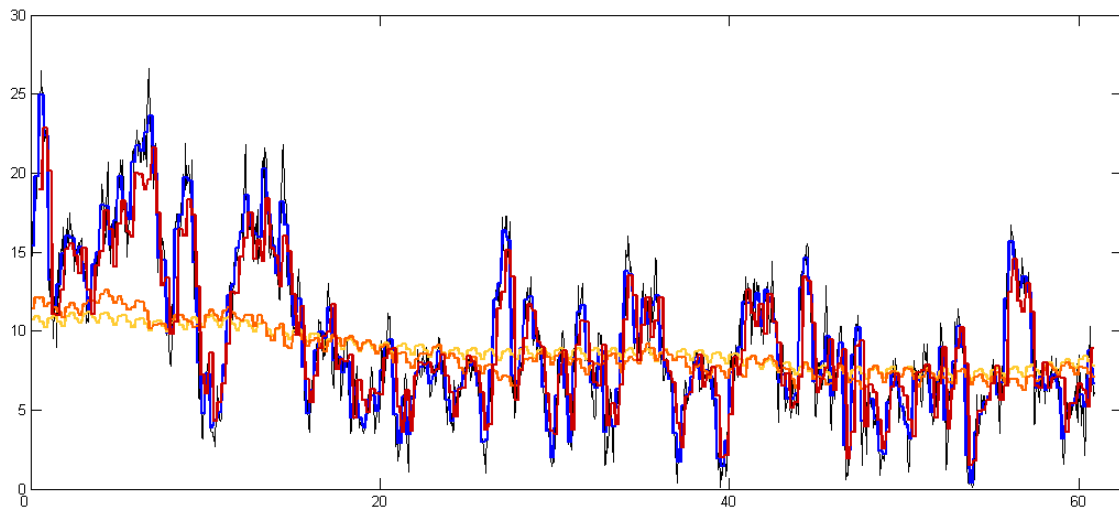

Figure 5.27: Forecast of 12 hourly wind speeds (two-step prediction) for **Munich**. The period of 60 days, starting January 1, 2000, is shown and x-axis is labeled per day, but plotted values are hourly values. The y-axis shows wind speeds in $\frac{\text{m}}{\text{s}}$. The thin black line are the originally measured wind speeds and the blue line are the six hourly means. The forecast of Subsection 5.3.1 is shown in yellow. The orange graph is the forecast described in Subsection 5.3.2. The red graph is the forecast described in Subsection 5.3.3.
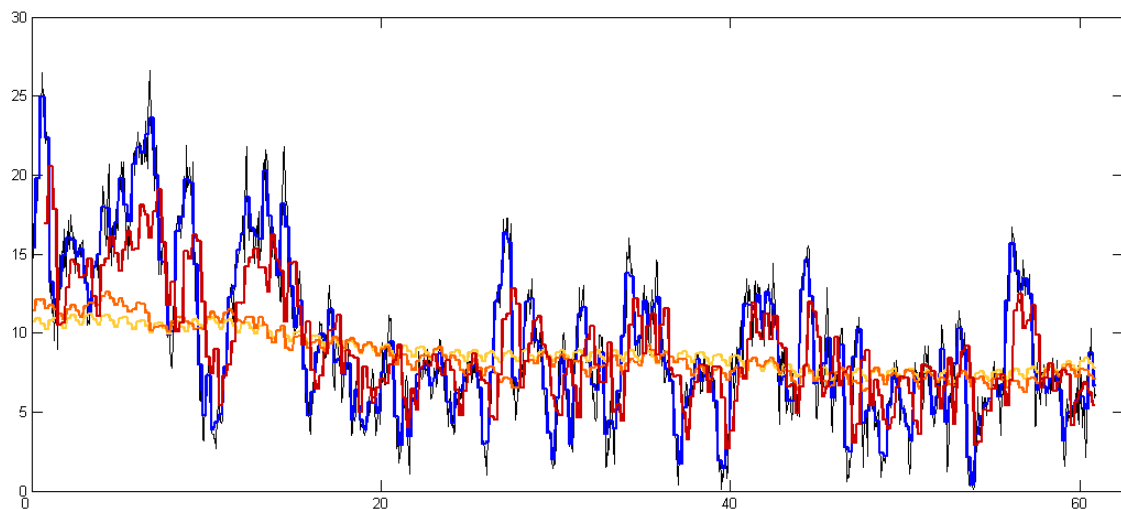
Figure 5.28: Forecast of six hourly wind speeds (one-step prediction) for **Brocken**. The period of 60 days, starting January 1, 2000, is shown and x-axis is labeled per day, but plotted values are hourly values. The y-axis shows wind speeds in $\frac{m}{s}$. The thin black line are the originally measured wind speeds and the blue line are the six hourly means. The forecast of Subsection 5.3.1 is shown in yellow. The orange graph is the forecast described in Subsection 5.3.2. The red graph is the forecast described in Subsection 5.3.3.
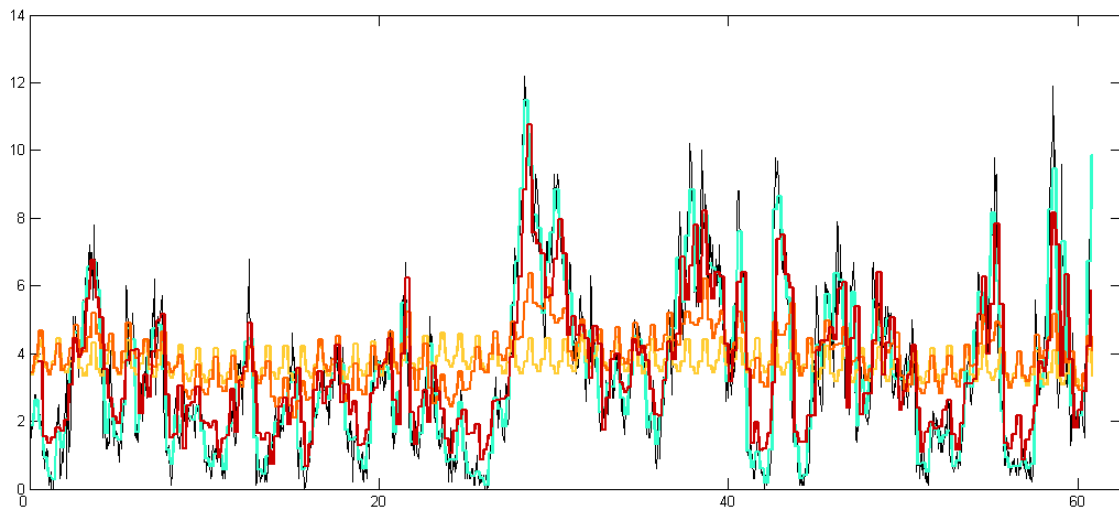

Figure 5.29: Forecast of 12 hourly wind speeds (two-step prediction) for **Brocken**. The period of 60 days, starting January 1, 2000, is shown and x-axis is labeled per day, but plotted values are hourly values. The y-axis shows wind speeds in $\frac{m}{s}$. The thin black line are the originally measured wind speeds and the blue line are the six hourly means. The forecast of Subsection 5.3.1 is shown in yellow. The orange graph is the forecast described in Subsection 5.3.2. The red graph is the forecast described in Subsection 5.3.3.

Figure 5.30: Forecast of six hourly wind speeds (one-step prediction) for **Fichtel-berg**. The period of 60 days, starting January 1, 2000, is shown and x-axis is labeled per day, but plotted values are hourly values. The y-axis shows wind speeds in $\frac{m}{s}$. The thin black line are the originally measured wind speeds and the blue line are the six hourly means. The forecast of Subsection 5.3.1 is shown in yellow. The orange graph is the forecast described in Subsection 5.3.2. The red graph is the forecast described in Subsection 5.3.3.
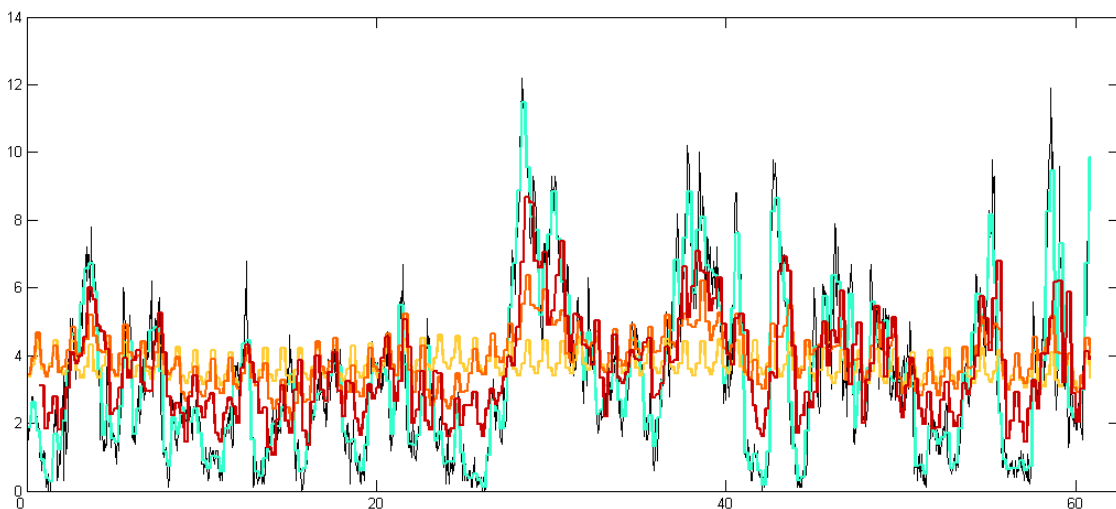

Figure 5.31: Forecast of 12 hourly wind speeds (two-step prediction) for **Fichtel-berg**. The period of 60 days, starting January 1, 2000, is shown and x-axis is labeled per day, but plotted values are hourly values. The y-axis shows wind speeds in $\frac{m}{s}$. The thin black line are the originally measured wind speeds and the blue line are the six hourly means. The forecast of Subsection 5.3.1 is shown in yellow. The orange graph is the forecast described in Subsection 5.3.2. The red graph is the forecast described in Subsection 5.3.3.

Figure 5.32: Forecast of six hourly wind speeds (one-step prediction) for **Trier**. The period of 60 days, starting January 1, 2000, is shown and x-axis is labeled per day, but plotted values are hourly values. The y-axis shows wind speeds in $\frac{m}{s}$. The thin black line are the originally measured wind speeds and the blue line are the six hourly means. The forecast of Subsection 5.3.1 is shown in yellow. The orange graph is the forecast described in Subsection 5.3.2. The red graph is the forecast described in Subsection 5.3.3.

Figure 5.33: Forecast of 12 hourly wind speeds (two-step prediction) for **Trier**. The period of 60 days, starting January 1, 2000, is shown and x-axis is labeled per day, but plotted values are hourly values. The y-axis shows wind speeds in $\frac{m}{s}$. The thin black line are the originally measured wind speeds and the blue line are the six hourly means. The forecast of Subsection 5.3.1 is shown in yellow. The orange graph is the forecast described in Subsection 5.3.2. The red graph is the forecast described in Subsection 5.3.3.

Chapter 6

Excursus - Photovoltaic Power Modelling

6.1 Influences on Photovoltaic Power

Irradiance, disturbed by cloud cover and other external factors, is converted into power. Blatantly sun exerts the dominating influence on photovoltaic power. Due to the earth's rotation we have strong diurnal fluctuations, e.g. after sunset and before sunrise no new photovoltaic power is produced.

6.1.1 Seasonal Behavior

The photovoltaic power has a clear seasonal behavior. During the day a maximum is achieved around noon (cf. Figure 6.1), during the year a maximum is achieved in summer (cf. Figure 6.2). The difficulty is how to model these patterns. The diurnal path of sun can be represented by altitude coordinate, a coordinate denomination in the horizontal coordinate system.

6.1.2 Horizontal Coordinate System

The horizontal coordinate system contains two coordinates: altitude (sometimes called elevation) and azimuth. So it is even referred to as altitude-azimuth (altaz) or azimuth-elevation (Az,El) system. It appoints the location of any celestial body from the observer´s point of view. So these coordinates are dependent on time and location.
The fundamental plane, which is designated as the horizontal plane, is the tangent plane of the earth at the observer´s location. Zenith is the perpendicular above the observer and nadir the antipodal one. Altitude measures the angle between the horizontal plane and the celestial body. At zenith per definition an altitude of 90° is observed. The range of altitude is $[-90°, +90°]$. Objects with negative altitude are above the horizontal plane and invisible to the observer. As a consequence no
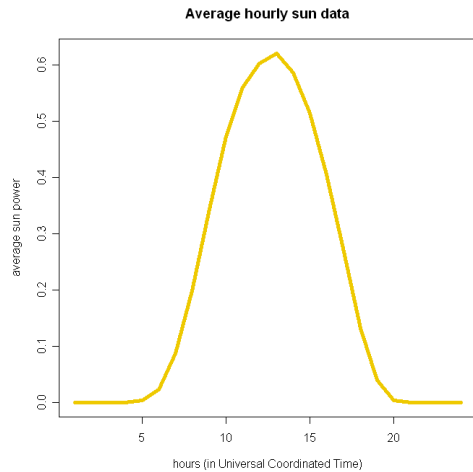
Figure 6.1: An average day of photovoltaic power production. Being standardized by capacity the values of photovoltaic power are between 0 and 1.
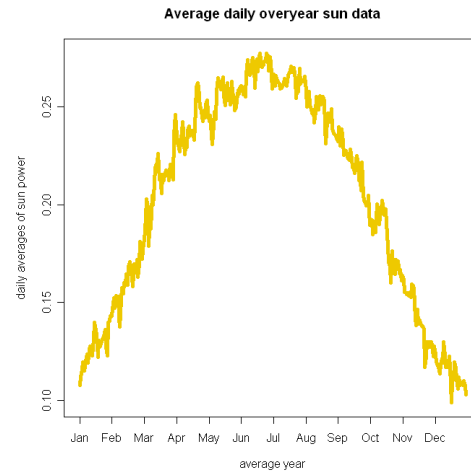
Figure 6.2: "An average year of photovoltaic power production. Being standardized by capacity the values of photovoltaic power are between 0 and 1.

sunlight can reach the observer´s location, if altitude of sun is smaller than 0° at this time. The zenith distance, which is more popular in astronomy, is defined as the angle between the celestial body and zenith. This angle is in a range of $[0°, 180°]$ and can be derived as $90° - altitude$.

Azimuth describes the angular distance between the defined point of origin and the observed celestial body in the horizontal plane. The range is $[0°, 360°]$, commonly starting from south or north as 0°. Measurement direction is mostly clockwise. As there are different conventions exact definition has to be ascertained on every occasion. Additional information and more elaborate explanation can be found in Karttunen [2000], Ch. 2.
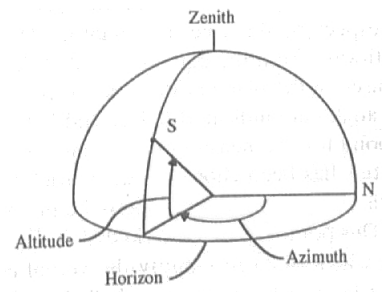
Figure 6.3: "Altitude and azimuth in the horizon system. S is the position of the star" Birney [1991], Ch.1.

6.1.3   Distance between Sun and Earth

The distance between sun and earth varies over the year. The annual perihelion happens in winter, which means in summer distance is wider than in winter. Thus the development of the distance is inversely proportional to the annual seasonality component of photovoltaic power (cf. Figure 6.4).
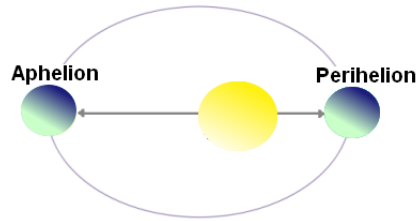
Figure 6.4: Distance between sun and earth changes in the course of the year, because the orbit of earth around the sun is elliptical. Considering the sun in the center, "perihelion", which names the shortest distance to earth occurs in summer and "aphelion", which names the longest distance happens in summer.

6.2 Modelling of Photovoltaic Power

Starting with the supposition of a connection between sun power and sun position we cut the sun power series by all values in which azimuth is smaller or equal to 0°. So the problem of zeros at night is solved. Clearly at night azimuth is smaller or equal to 0°. This new sun time series does not have equidistant time space, but we can have a look at the data without distortion towards zero.

We obtain the following model

$$S_i = \begin{cases} \beta_1 + \beta_2 alt_i^{\beta_3} + \beta_4(1 + \beta_5 \sin(\frac{(i+\beta_6)\pi}{4380}))Dist_i & \text{for} \quad alt_i > 0 \\ 0 & \text{else} \end{cases}$$

where

S_i:	solar power in Europe normalized by estimated capacity (hourly resolution, but not equidistant: no values at the dark)
alt_i:	altitude in Munich at time i
$Dist_i$:	distance between earth and sun

To get an at least approximate idea of the fit of the model, Adjusted-R^2 is calculated. Being slightly higher than 90% the approximation to the solar power time series over all Europe seems to be adequate with this model. One average year of measurement data and the estimates of the model are demonstrated in Figure 6.5.

equation data in gold
time series data in black
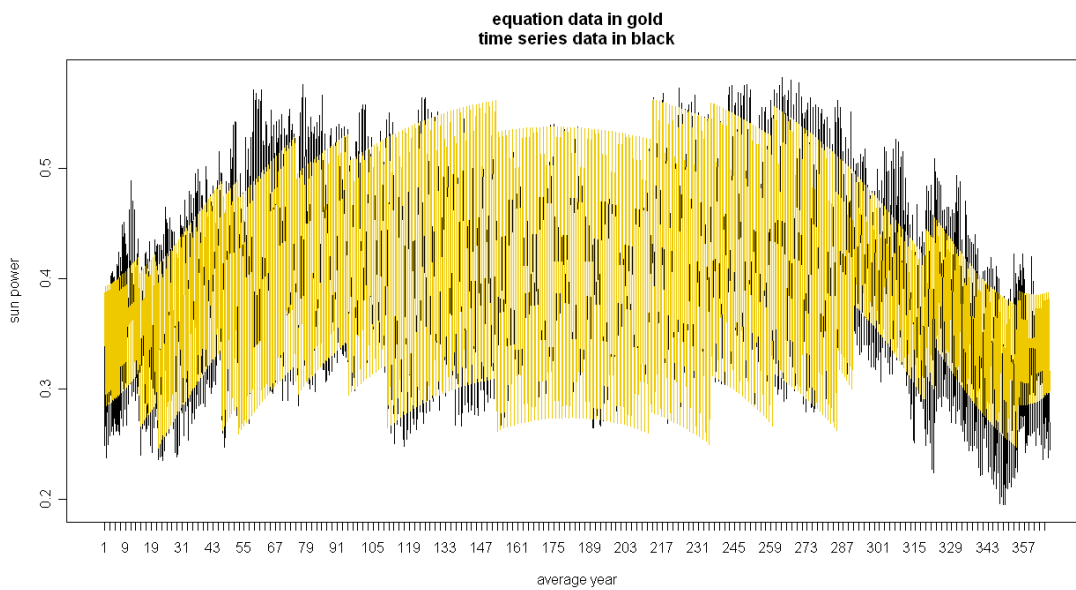
Figure 6.5: The horizontal axis shows one average year, the values of photovoltaic power are averaged over the period of 2000 to 2007. The vertical axis has values between 0 and 1, as the photovoltaic power is standardized by capacity. The yellow graph originates from the model, the black graph is drawn by the averaged measurement data.

Chapter 7

Conclusion

Renewable energies are essential for a sustainable future development. However, traditional power plants can only be replaced, if a constant energy supply is guaranteed. Therefore we need to improve our understanding of wind speed.

In this thesis we introduced a time series model for wind speeds. Seasonal behavior of wind speed is modelled by two seasonal components: one component models daily seasonalities and the other one models annual seasonalities. As expected, the seasonal components vary between the different monitoring stations in pattern and magnitude. In search for influencing factors we tried different covariates like temperature, air humidity, precipitation, sunshine duration and surface pressure, but only surface pressure differences could improve our model considerably. Moreover, an autoregressive process of order 2 is incorporated. Additionally, a GARCH process is used to capture the remaining dependencies. For the error terms, we assume an hyperbolic distribution. After fixing two of the four hyperbolic parameters for reasons of identifiability, we have two free parameters left, which are used to model the specific wind speed behavior at the different monitoring locations. Indeed, these two parameters turn out to be quite different for different types of locations. In particular, these parameters indicate a different wind speed pattern for locations close to cities (as Munich or Trier) than for mountainous and offshore regions (as Brocken, Fichtelberg, and Helgoland).

Close to cities as Munich and Trier we observe lower average wind speeds. The annual seasonality varies nearly in the same range as the daily seasonality and the parameter of annual seasonality implies only small effects. The average wind speed during one day reaches its peak in the afternoon for these two stations. As the second parameter of the AR(2) process is very small, wind speeds remain usually in the same range of values. The variance process has a higher impact for wind speeds at this locations and therefore the error terms get higher influence, too. These error terms are modelled with an hyperbolic distribution. For this group, the approximation of the skewness give a right-skewed distribution, with higher skew than for the other group. Few wind speeds are very high compared to the main part. Both groups have a smaller kurtosis than the normal distribution, which indicates more extreme

wind speeds. However, the kurtosis of this group is smaller than the kurtosis of the other group. In this case we consider the fact that only few high wind speeds occur, which influence the distribution to be right-skewed. In total, however, here we have more values closer to the mean that in the other group.

The second observed group is formed by the monitoring stations at the mountains, Brocken and Fichtelberg, which have nearly the same behavior. Similar behavior is also observed for Helgoland, which is an island in the North sea. Generally, these stations have much higher wind speeds. Seasonality over the year causes a higher change in wind speeds and is more important as the huge parameter shows. The average wind speeds during a day reach a low in the afternoons. These patterns represent the local circulation phenomenons at these locations. For the AR(2) process the first parameter mitigates the second parameter, which negatively influences the model. Observing values about three times bigger than than the parameter in the other group the impact on this group is higher. Hence, changes in wind speed occur more frequently than in the other group. However, wind speeds develop relatively smooth. The influence of the variance is bigger for this group. The variation in the variance is higher, but the variance process which is scaled by the error terms has a smaller influence. The estimation of the the hyperbolic distribution gives a right-skewed distribution, too. This shows the influence of few outlining high wind speeds. Observing a kurtosis smaller than the normal distribution, we can suppose the existence of more distinct values. The tails, which reflect very small and very high wind speeds, occur more often. This represents the location with some calm periods and strong storms especially in winter.

Our derived model is relatively flexible and includes characteristics arising from different locations. The estimates of our model give reasonable forecast values. In order to improve the fit, more covariates can be incorporated instead of only one direction of surface pressure. Having hourly pressure differences the time scale can be downsized to an hourly resolution. If wind directions are available, a regime-switching component can be included as well.

Appendix A

Retrieving Data from GRIB-files

In order to retrieve the required data from the GRIB-file an appropriate program is needed. GRIB API, the GRIB decoder offered by ECMWF is an application program interface to encode and decode WMO FM-92 GRIB edition 1 and edition 2 messages. A set of command line tools[1] are provided at ECMWF website to easily start using the data. To be able to use the main features of GRIB API, some examples can be found on the ECMWF homepage in Python, Fortran 90, C and Fortran 77.[2]

As GRIB API is a linux based program, one needs a linux environment to run the program. Operating with a windows system, a linux-like environment for a windows computer is needed. A program fulfilling these demands is e.g. Cygwin. During the installation process a minimum of Cygwin packages has to be downloaded and installed containing Archive, Base, Devel, Graphics, Interpreters, Jasper, Libs, Utils, Web and X11. As a minimum, these packages are needed to work with GRIB API. A list of all available packages, help for the installation process and documentation can be found on Cygwin´s web page.[3] Inside this interface GRIB API could be installed even on a windows operating system. Therein one can apply all the command line tools mentioned above and process with the data.

Since we needed a time series of surface pressure for the period from January 1, 2000 until December 31, 2007 at best in an hourly resolution, we requested a GRIB file from the ECMWF homepage containing the value of pressure at a restricted area. To achieve data from that file the grib_ls routine is used. Being at the storage location of grib_ls searched values can be retrieved by the command line order:

$$grib\_ls \; -l \; xx.xx, yy.yy, 1 \; -p \; paramID, \; dataDate, DataTime, name$$
$$/cygdrive/c/Documents/datafile.grib \; -> newfile.txt$$

Data is read out of $datafile.grib$ and saved in $newfile.txt$. The system path reflects the storage location of the GRIB file. By the assignment of $->$ the data read

---

[1] $http://www.ecmwf.int/publications/manuals/grib\_api/tools.html$
[2] $http://www.ecmwf.int/publications/manuals/grib\_api/grib\_examples.html$
[3] $http://cygwin.com/$

out will be saved at present system location in which grib_ls is also stored. Then *newfile.txt* contains the values of the variable at the grid point of the inserted coordinates, which is here symbolized as *xx.xx* and *yy.yy*. The placeholder *xx.xx* symbolizes degree of latitude, which is inserted positive for the northern latitude and negative for the southern latitude. The longitude *yy.yy* is inserted positive for the eastern longitude and negative for the western ones. In front of the coordinates option "−*l*" is inserted to obtain the nearest grid point belonging to these coordinates. After typing the coordinates, the numeral 1 is inserted. This numeral ensures one output value closest from the nearest grid point. The default modus would be 4, which means that the values at the four nearest grid points to the selected coordinates would be outputted. The "−p" option is used to set the keys which should be printed in the txt-file. The "Parameter ID" is a unique value which classifies every parameter. "DataDate" prints the date at which data is measured, e.g 20030201 for 01st of February 2003. "DataTime" shows the hour, 0000 for 00:00 am, 0600 for 06:00 am and so on. The key name prints out the names of the saved parameters.

Appendix B

Graphs of DWD Explorative Data Analysis

Here we report some graphs related to wind speed at the monitoring stations Munich, Helgoland, Brocken, Fichtelberg and Trier. The are used to find seasonal patterns in the data. Daily averages are considered per month to detect any seasonality over the year. As well one average day per month is considered to follow the changes in the seasonalities occurring during the year. In Section 4.2 these figures are explained in detail.
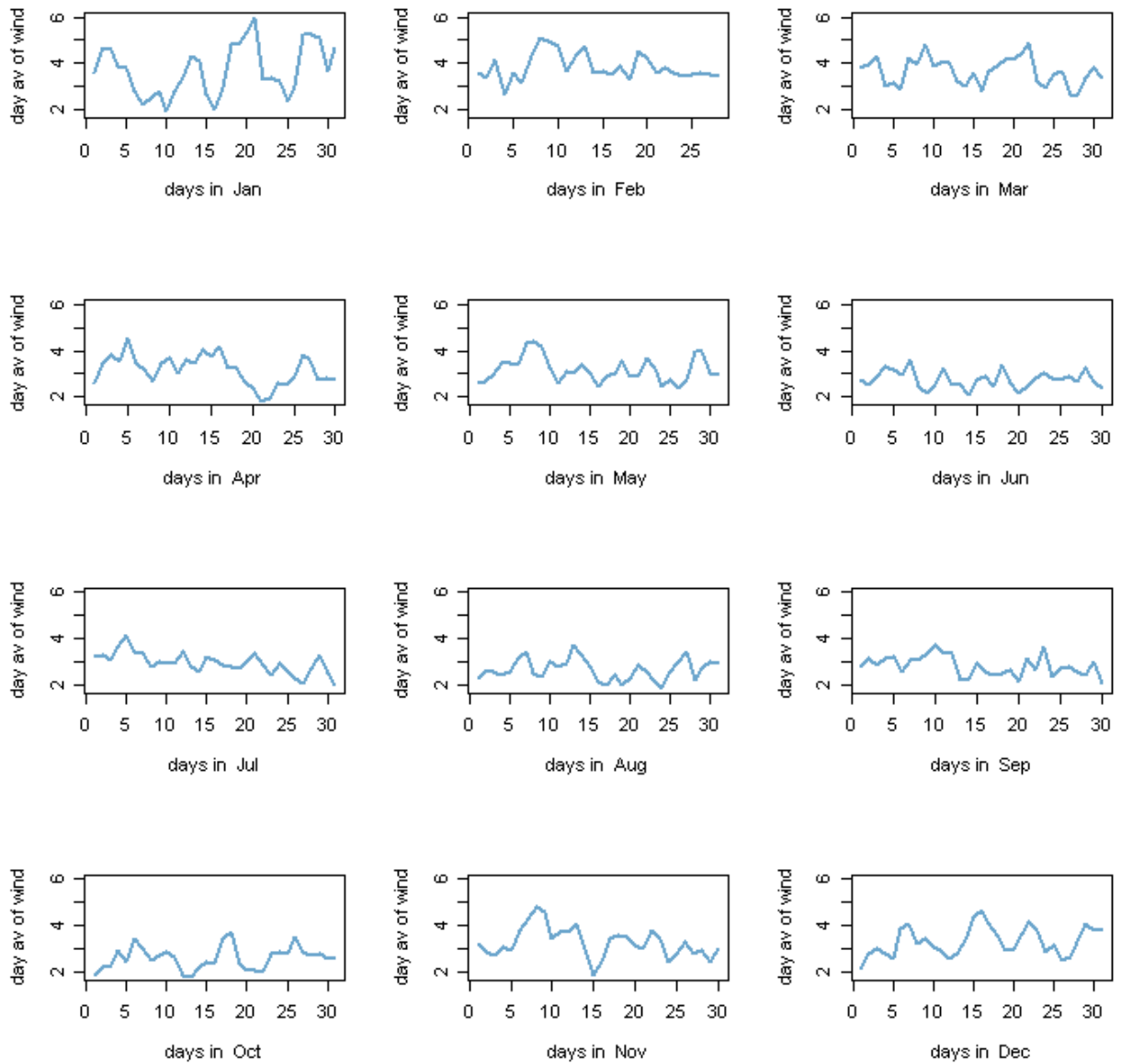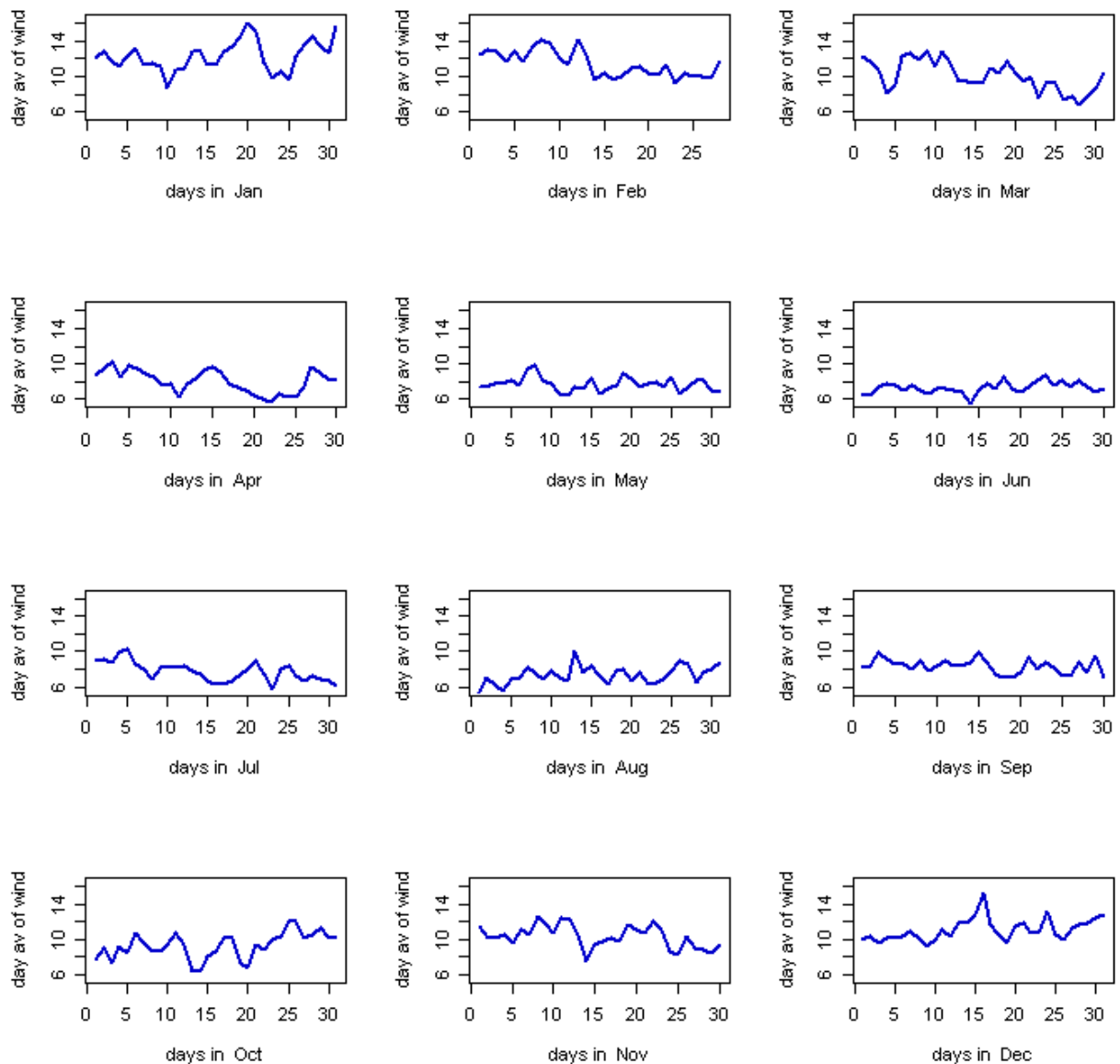
Figure B.1: Average value of wind speed in **Munich** per day in $\frac{m}{s}$, a 24 hours mean of every day is built for every single month. The considered months are as well average months, which come from one average year constructed by the values of the years 2000-2007. From these graphs one can see that the difference of levels in the course of the year can barely be noticed in comparison to Helgoland.
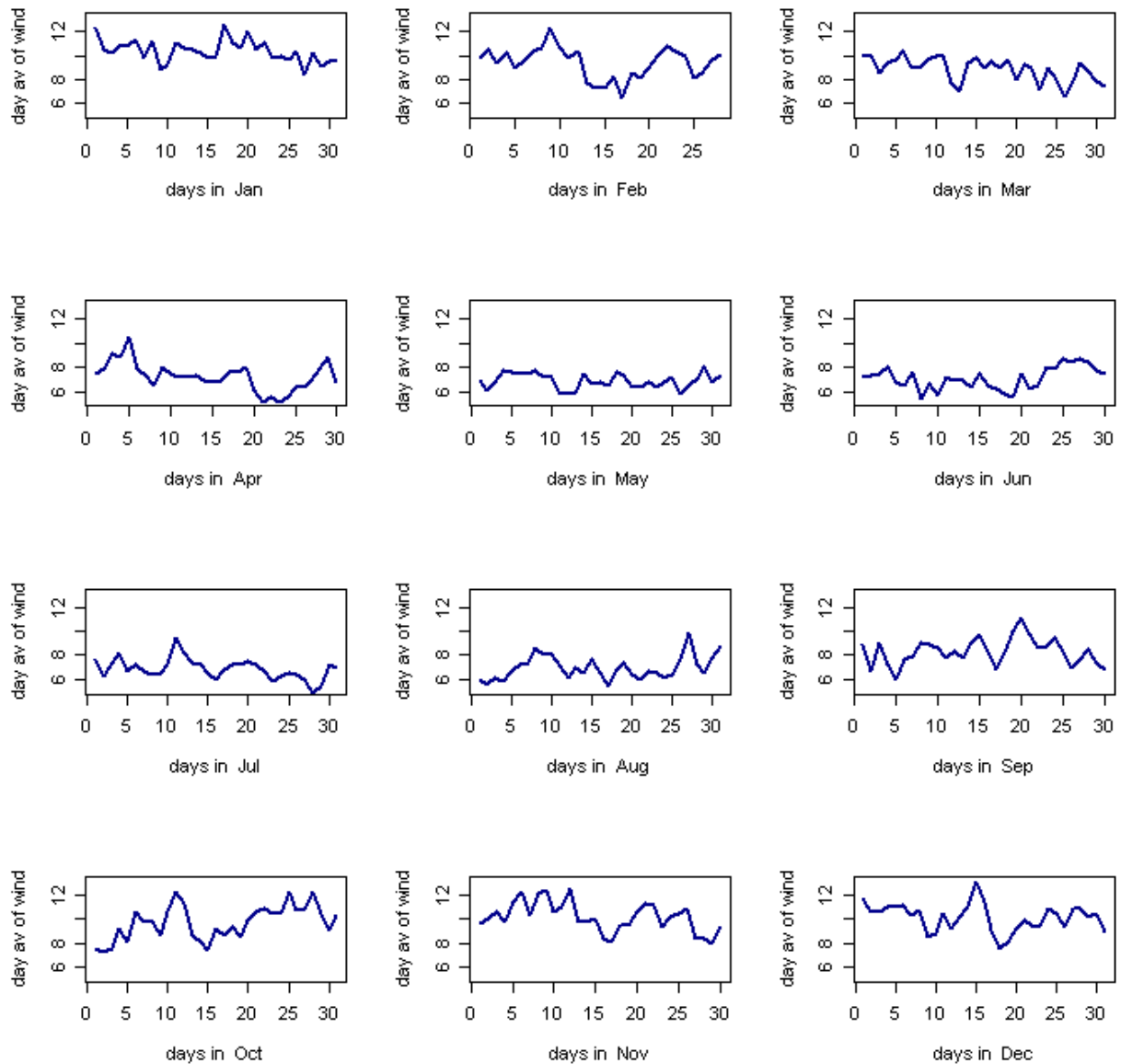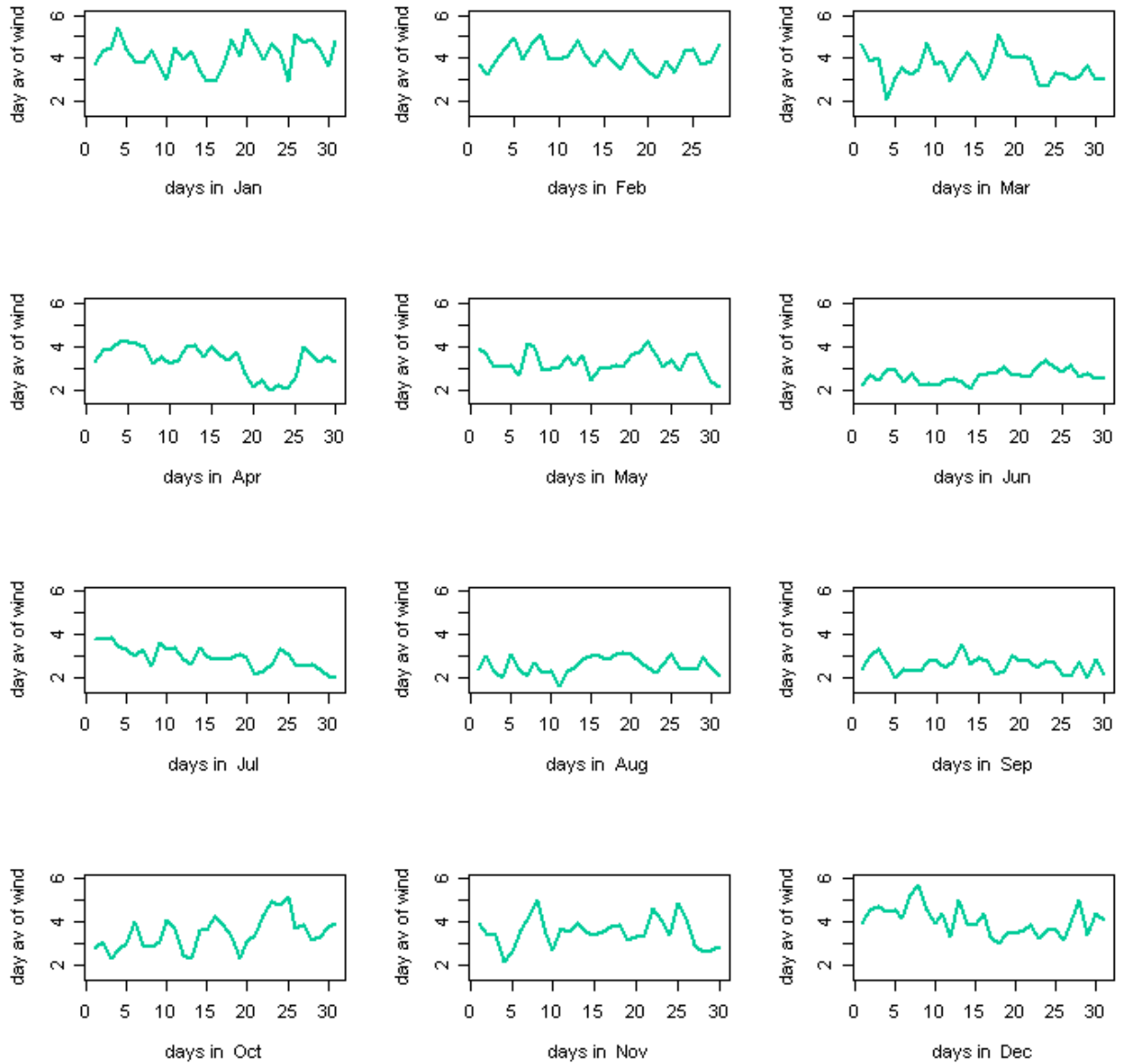
Figure B.2: Average value of wind speed on **Helgoland** per day in $\frac{m}{s}$, a 24 hours mean of every day is built for every single month. The considered months are as well average months, which come from one average year constructed by the values of the years 2000-2007. From these graphs one can see that the differences of levels in the course of the year are fairly high. In summer times values of about 6 $\frac{m}{s}$ seem usual, whereas in winter times values vary at about 12 $\frac{m}{s}$.
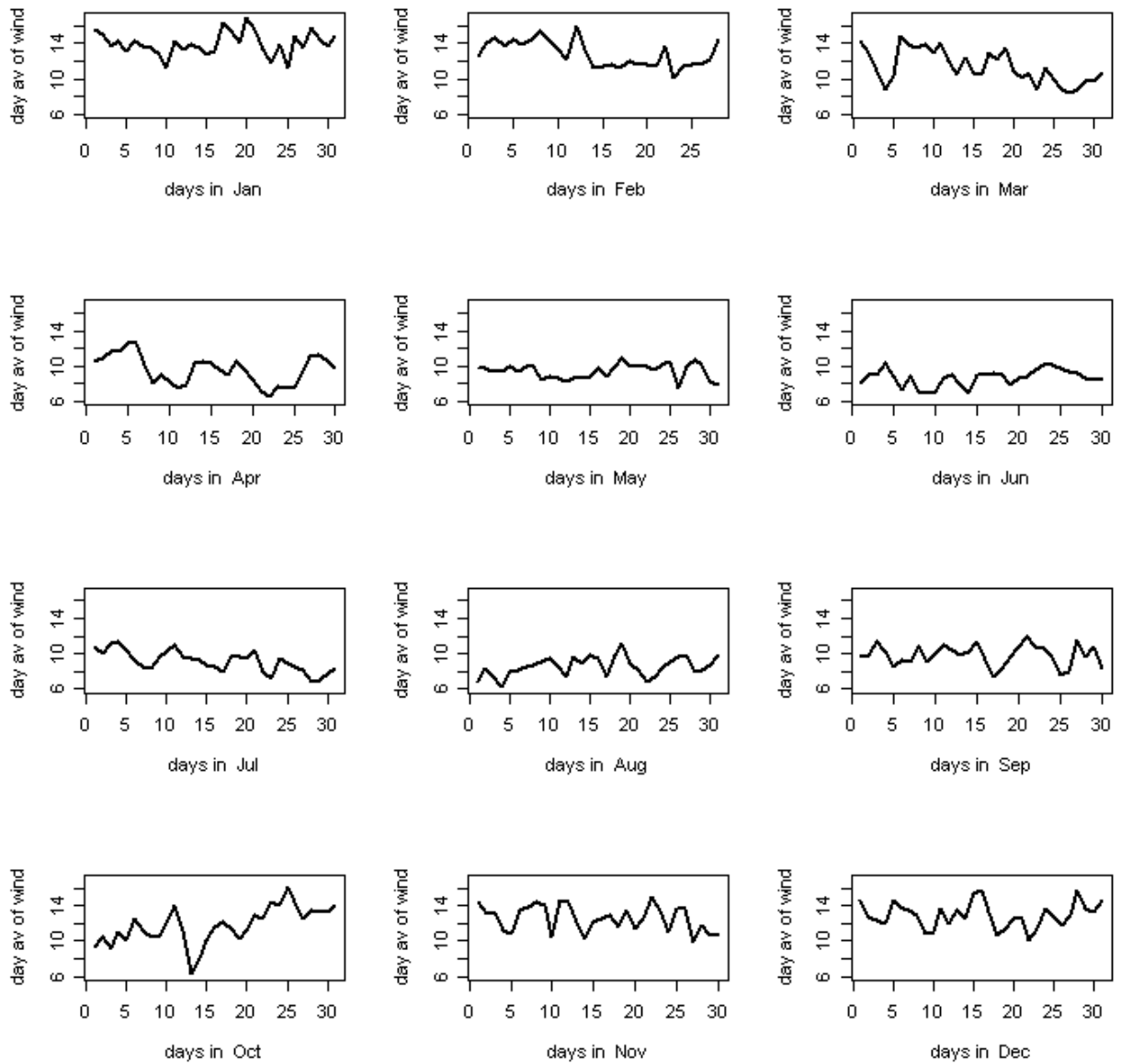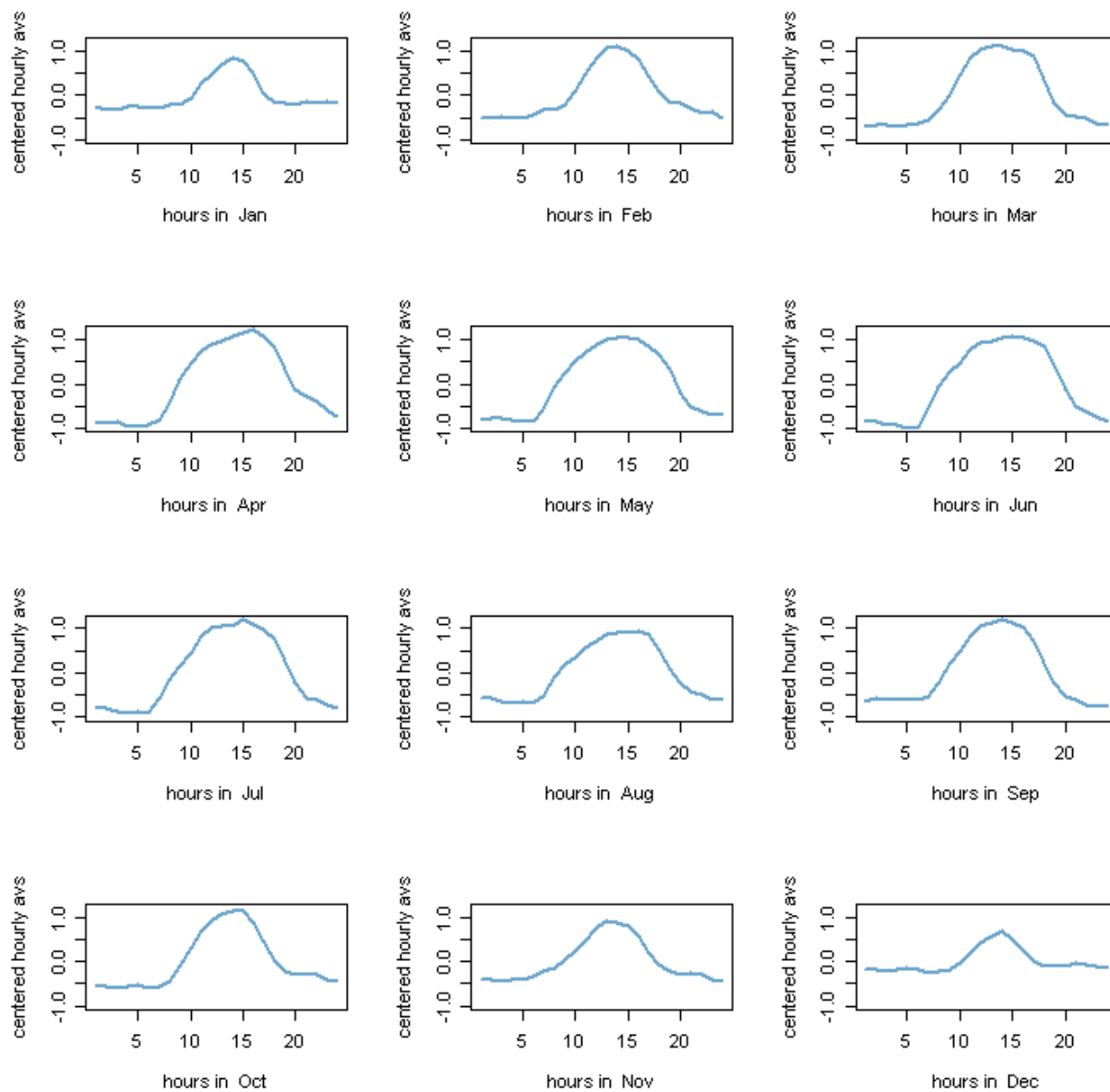
Figure B.3: Average value of wind speed at **Brocken** per day in $\frac{m}{s}$, a 24 hours mean of every day is built for every single month. The considered months are as well average months, which come from one average year constructed by the values of the years 2000-2007.
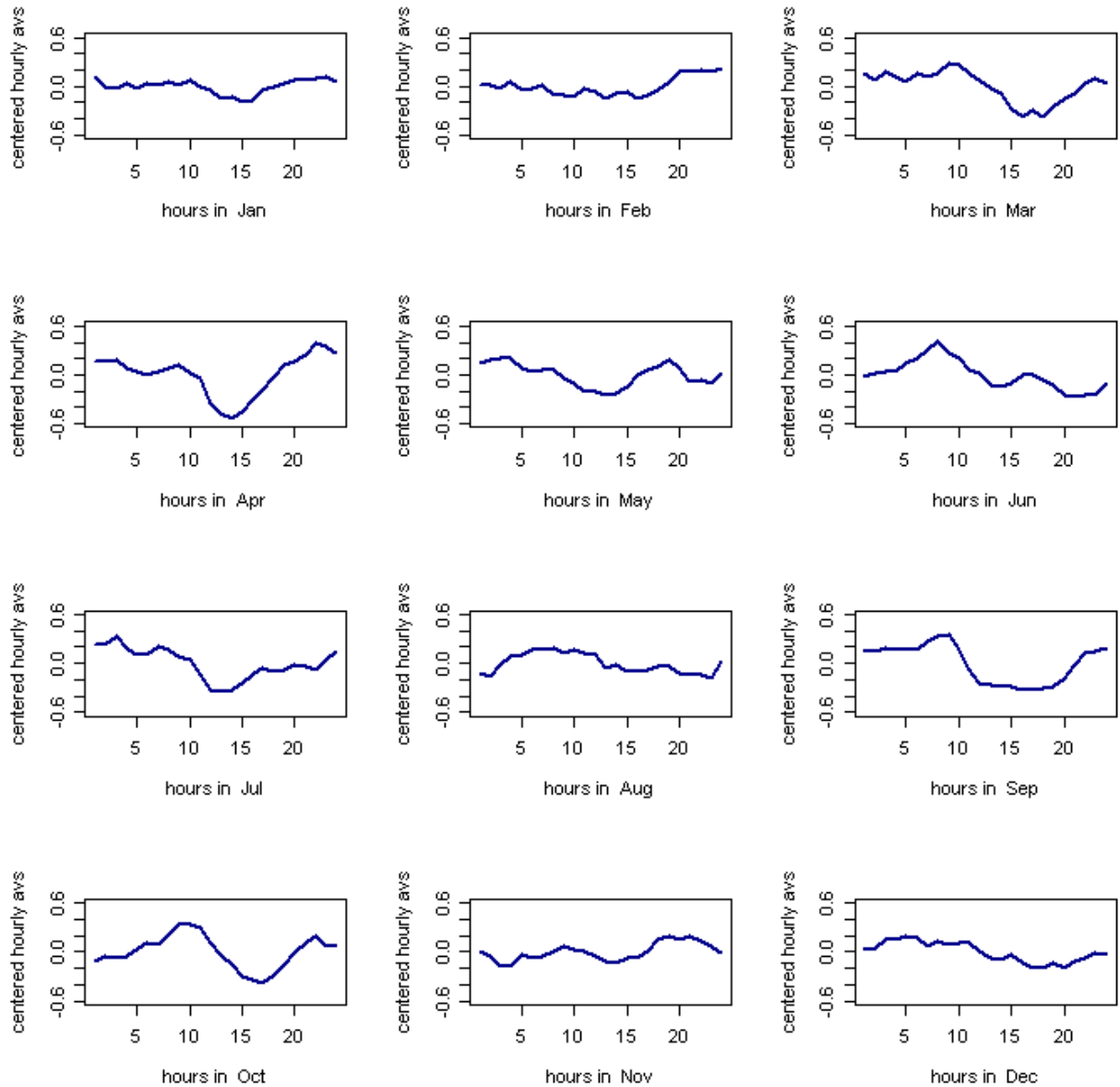
Figure B.4: Average value of wind speed at **Fichtelberg** per day in $\frac{m}{s}$, a 24 hours mean of every day is built for every single month. The considered months are as well average months, which come from one average year constructed by the values of the years 2000-2007.

Figure B.5: Average value of wind speed at **Trier** per day in $\frac{m}{s}$, a 24 hours mean of every day is built for every single month. The considered months are as well average months, which come from one average year constructed by the values of the years 2000-2007.
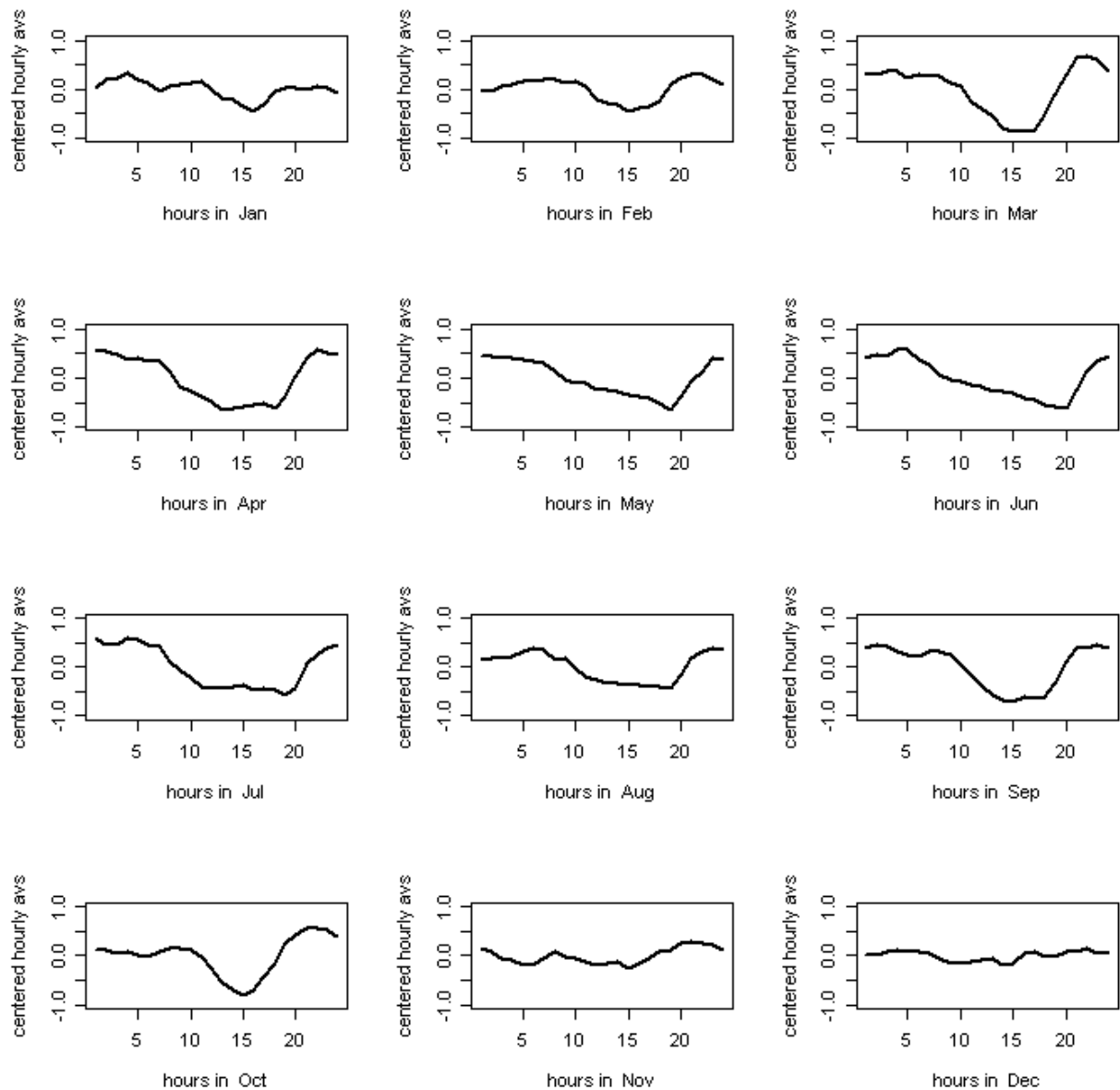
Figure B.6: One day average of wind speed for **Munich** for each month of the year. For every hour an average value separated per month of all the eight years is plotted. The hourly means are centered by an average wind speed value of the particular day in the prevailing month. The intra-daily oscillation partly reaches more than 2 $\frac{m}{s}$, which is quite big in comparison to the measured wind speeds of about 2-5 $\frac{m}{s}$.
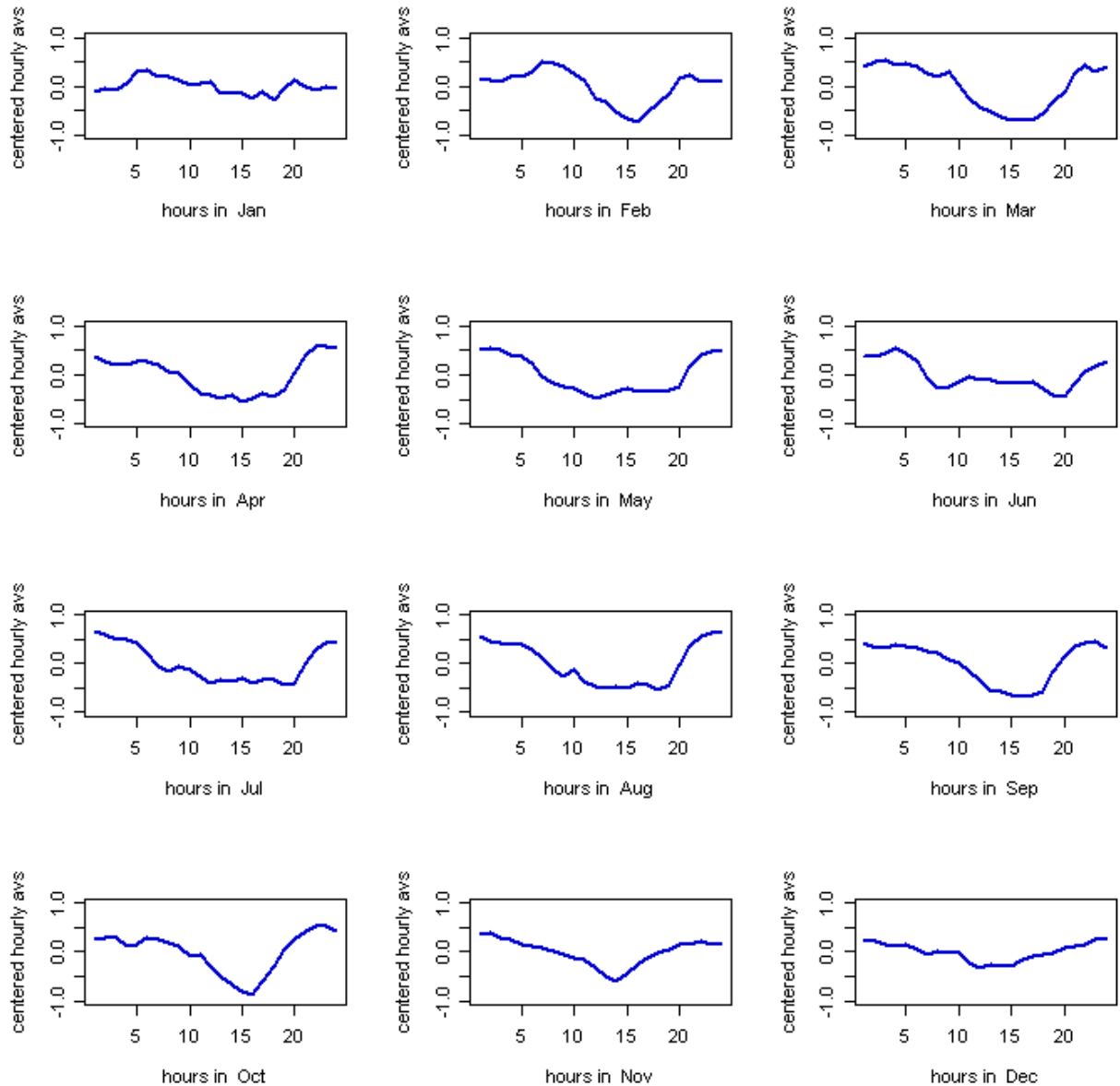
Figure B.7: One day average of wind speed for **Helgoland** for each month of the year. For every hour an average value separated per month of all the eight years is plotted. The hourly means are centered by an average wind speed value of the particular day in the prevailing month. The intra-daily oscillation hardly reaches more than 1 $\frac{m}{s}$, which is a slight change in comparison to the measured wind speeds of about 6-12 $\frac{m}{s}$.
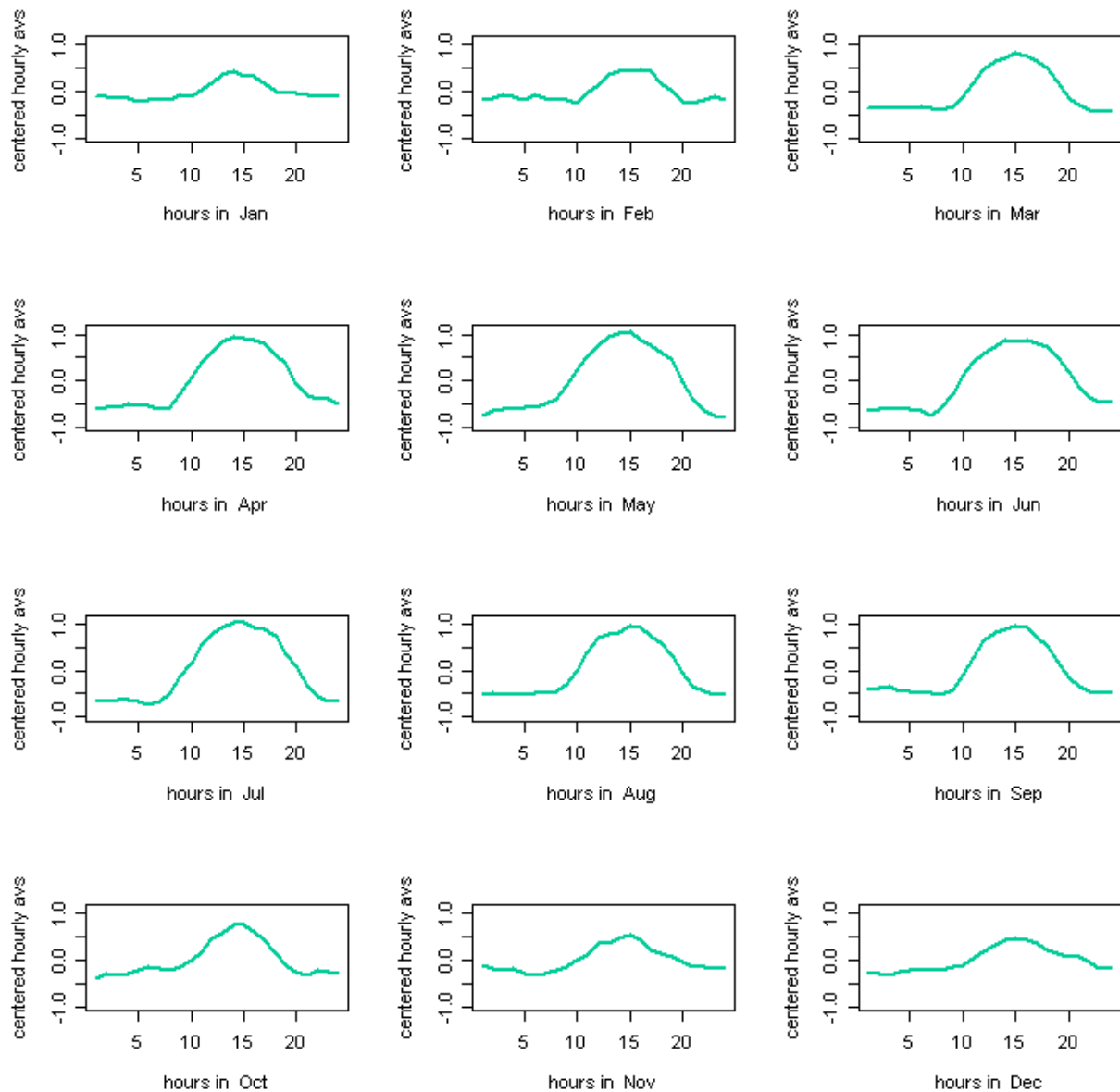
Figure B.8: One day average of wind speed for **Brocken** for each month of the year. For every hour an average value separated per month of all the eight years is plotted. The hourly means are centered by an average wind speed value of the particular day in the prevailing month.

Figure B.9: One day average of wind speed for **Fichtelberg** for each month of the year. For every hour an average value separated per month of all the eight years is plotted. The hourly means are centered by an average wind speed value of the particular day in the prevailing month.

Figure B.10: One day average of wind speed for **Trier** for each month of the year. For every hour an average value separated per month of all the eight years is plotted. The hourly means are centered by an average wind speed value of the particular day in the prevailing month.

Bibliography

Atkinson, A. C., Fienberg, S. E. (1985). *A celebration of statistics: The ISI centenary volume : a volume to celebrate the founding of the International Statistical Institute in 1885.* Springer, New York (USA).

Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **353** (1674) 401–409.

Barndorff-Nielsen, O. E., Jensen, J. L., Sørensen, M. (1989). Wind shear and hyperbolic distributions. *Boundary-Layer Meteorology* **49** (4) 417–431.

Bayin, S. S. (2006). *Mathematical methods in science and engineering.* Wiley, Hoboken NJ (USA).

Birney, D. S. (1991). *Observational astronomy.* Cambridge University Press, Cambridge (England).

Brockwell, P. J., Davis, R. A. (1991). *Time series: Theory and methods.* (2nd ed.) Springer, New York (USA).

Cížek, P., Härdle, W. K., Weron, R. (2011). *Statistical tools for Finance and insurance.* Springer, Berlin (Germany).

Eberlein, E., Keller, U. (1995). Hyperbolic distributions in Finance. *Bernoulli* **1** (3) 281–299.

Etling, D. (2008). *Theoretische Meteorologie: Eine Einführung.* (3., erweiterte u. aktualisierte Auflage). Springer, Berlin (Germany).

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., Aldrich, E. (2006). Calibrated probabilistic forecasting at the stateline wind energy center. *Journal of the American Statistical Association* **101** (475) 968–979.

Heide, D., von Bremen, L. , Greiner, M., Hoffmann, C., Speckmann, M., Bofinger, S. (2010). Seasonal optimal mix of wind and solar power in a future, highly renewable Europe. *Renewable Energy* **35** (11) 2483–2489.

Karttunen, H. (2000). *Fundamental astronomy.* (3., revised and enlarged, corrected 2. print ed.). Springer, Berlin (Germany).

Longley, P. A., Goodchild, M. F. C., Maguire, D. J. C., Rhind, D. W. C. (2005). *Geographical Information Systems and Science.* Wiley, West Sussex (England).

Malberg, H. (1997). *Meteorologie und Klimatologie: Eine Einführung.* (3., aktualisierte und erweiterte Auflage). Springer,Berlin (Germany).

Quaschning, V. (2007/2008). *Regenerative Energiesysteme: Technologie – Berechnung – Simulation.* (5., aktualisierte Auflage). Hanser, München (Germany).

Quaschning, V. (2010). *Renewable Energy and Climate Change.* Wiley, West Sussex (England).

Simpson, J. E. (1994). *Sea breeze and local winds.* Cambridge University Press, Cambridge (England).

Taille, J. (1981). *Statistical distributions in scientific work: 4. Models, structures, and characterizations.* D. Reidel Pub. Co., Dordrecht (Netherlands).

von Bremen, L., Saleck, N., Heinemann, D. (2007). Enhanced regional forecasting considering single wind farm distribution for upscaling. *Journal of Physics: Conference Series* **75** (2007) 012040.

WMO (1998). Manual on codes. *WMO PUBLICATION No. 306.* ftp://www.wmo.int/Documents/MediaPublic/ Publications/ CodesManual WMO no 306/WMO 306 VolII.pdf. WMO, Geneva (Switzerland).

WMO (2003). Guide to the WMO table driven code form used for the representation and exchange of regularly spaced data in binary form. http://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/ GRIB2_062006.pdf. WMO, Geneva (Switzerland).