Master Thesis

Simulation of Synthetic Wind Power Time Series in the North-East of Brazil

Submitted by

Katharina Gruber

1026526

In partial fulfilment of the requirements for the degree of

“Diplom-Ingenieur”

Within the Master’s Programme “Umwelt und Bioressourcenmangement”

University of Natural Resources and Life Sciences, Vienna

Department of Economics and Social Sciences

Institute for Sustainable Economic Development

Supervised by: Schmid, Erwin, Univ.-Prof. Dipl.-Ing. Dr.

Co-Advisor: Schmidt, Johannes, Dipl.-Ing. Dr.

Vienna, August 2017

Statutory declaration

I declare that I have developed and written the enclosed Master Thesis completely by myself, and have not used sources or means without declaration in the text. Any thoughts from others or literal quotations are clearly marked. The Master Thesis was not used in the same or in a similar version to achieve an academic grading or is being published elsewhere.

Vienna, 21st of August 2017 Signature Katharina Gruber

Katharina Gruber

Abstract

Brazil is a country with lots of hydropower generation, which is at risk due to insufficient electricity generation capacities by droughts in recent years. Hence, other means of renewable energy generation need to be evaluated and promoted. Wind power has high potential, especially in the North-East region of Brazil. Therefore, the temporally explicit potential of wind power generation is examined in this thesis, to allow for an assessment of short- and long-term variability of this resource. A model is developed, which simulates wind power generation from reanalysis wind speed data provided by NASA. The wind speed data is corrected with measured wind speeds from the National Meteorological Institute (INMET) as well as with wind power generation data provided by the National Electrical System Operator (ONS). Results show that bias correction substantially improves the simulated time series. Although correlations are sometimes reduced, the deviation (RMSE) as well as absolute differences between simulated and measured wind power generation are reduced. Both measures are lowest when applying both bias correction methods. Results show that correlations between simulated and observed data are better in periods of low installed wind power capacity than in periods of higher installed capacities, i.e. at later time periods. The deviation at higher installed capacities is possibly due to different reasons, such as the use of new types of wind turbines (not considered in our simulation) or different locations of newer wind parks. The error could possibly be reduced by more detailed data on technologies and installation locations. In general, the model is suitable for estimating the wind power generation of a whole region, which can be interesting for simulation of different long-term energy scenarios in policy development.

Kurzfassung

Brasilien ist ein Land, das seit jeher von einem hohen Anteil an Wasserkraft geprägt ist, dessen Produktion jedoch starken Risiken durch Dürreperioden ausgesetzt ist. Daher ist es notwendig nach Alternativen in der Energieerzeugung zu suchen. Windkraft bietet dafür besonders im Nordosten Brasiliens hohes Potenzial, weshalb die mögliche Energieerzeugung aus Windkraft in dieser Arbeit untersucht wird. Zunächst wird auf Basis von durch die NASA zur Verfügung gestellten Reanalyse Winddaten die mögliche Windkraftproduktion modelliert. Diese Daten werden einer Fehlerkorrektur mit Windgeschwindigkeitsdaten vom Nationalen Meteorologischen Institut (INMET) sowie mit Energieerzeugungsdaten aus Windkraft vom Nationalen Netzbetreiber (ONS) unterzogen. Aus den Ergebnissen geht hervor, dass die Fehlerkorrektur die Simulation wesentlich verbessert. Auch wenn die Korrelationen zwischen gemessenen und simulierten Daten in manchen Fällen verringert werden, sinken sowohl die Abweichungen als auch absolute Differenzen zwischen simulierten und gemessenen Zeitreihen. Die Anwendung der Windgeschwindigkeits- und Windenergiefehlerkorrektur führen zu deutlich besseren Ergebnissen. In Zeiten niedrigerer installierter Kapazitäten korrelieren simulierte und gemessene Daten besser als bei höheren installierten Kapazitäten. Die Diskrepanz bei höheren Produktionswerten kann auf unterschiedliche Ursachen, wie den Ausbau unterschiedlicher Typen von Windkraftanlagen oder den genauen Standorten von neu installierten Windparks, zurückgeführt werden. Zusammenfassend lässt sich sagen, dass das Modell geeignet ist, die Windkraftproduktion auf regionaler Ebene in ausreichender Genauigkeit zu simulieren. Dies ist vor allem für Energiesystemmodelle, die der Politikberatung dienen, von Bedeutung.

Acknowledgement

First of all, I would like to thank my family, especially my parents, for supporting me not only during my thesis, but for the whole time of my studies and education. Thank you not only for your financial but also and even more for your moral support and for awakening my interest in science and learning since I was a little girl.

Furthermore I want to thank my advisor Erwin Schmid, who encouraged and motivated me.

I am also very grateful for the outstanding support of my co-advisor Johannes Schmidt, who provided me with many resources and gave me lots of useful inputs for this thesis. Thank you for helping me with all my concerns.

Moreover, many thanks to my friends and colleagues, who made my student days more pleasant.

And very special thanks to Peter Gsellmann for providing his knowledge and patience, whenever I needed him.

Table of Contents

List of Figures

List of Tables

1 Introduction

A global growth in demand for energy has been observed for many years. However, increasing energy consumption usually leads to environmental problems, as fossil fuels continue to be the major contributor to total energy supply. Therefore, several new policies and targets were recently set in the energy sector: In January 2016, the United Nations' 17 sustainable development goals (SDGs) came into force. The 7th goal refers to future energy provision: "Ensure access to affordable, reliable, sustainable, and modern energy for all" [1]. Furthermore, the World Energy Council formulated the key word "energy trilemma", indicating that energy should be provided in a secure, affordable and environmentally sustainable way [2]. The UN Climate Change Conference (COP 21) in Paris took place in 2015. The reduction of emissions in order to reach the two-degree-goal to enable a clean and stable development is paramount in the agreement met in this conference. In particular, developing countries shall be supported financially in helping with the achievement of this goal [3].

These agreements, among others, set greenhouse gas emission targets to help prevent pollution from energy consumption. In order to implement these goals, new policies have to be adopted and the energy system has to be completely restructured. To support political and management decisions in this fields, energy system models are frequently used to inform about possible future scenarios and trends [4].

In this thesis, we develop highly important input data for such models by simulating the wind power generation in the North-East of Brazil from reanalysis data. In this chapter, first of all insights into the current Brazilian energy landscape with a focus on wind power will be given. Furthermore, the use of energy system models and the suitability of reanalysis data as a source of weather data are discussed. Finally, the aim and structure of the thesis are presented.

1.1 Wind power and the Brazilian electricity system

Brazil is a country, where a large share of the electric power generation is provided by hydropower. In the period from 2004 to 2013, Brazilian electricity consumption was rising by about 4% per year and was estimated in 2013 to increase by 4.7% annually until 2022 [5] [6]. According to the official government expansion plan, around 2030 the load is expected to double, compared to the year 2013, with an annual growth of 4.2% [7]. The demand for energy in all of Brazil is in principle rising, however, a slight downturn was experienced in recent years due to profound economic recession [8]. Most of this demand has been and is still covered by hydropower: Between 2004 and 2013, the share of hydropower in Brazilian electricity generation ranged from 69% to 84% [5] [7] [9]. However, because of seasonal and annual variabilities, droughts and regulation failures, this source of energy is risky [7]

[9]. Between 2000 and 2002, there was an energy crisis due to several reasons including reduced rainfall. Later, it has become difficult to replenish the water reservoirs that were emptied during that time [10]. In 2013, the share of hydropower fell from 84.5% in the year before to 79.3%, due to unfavourable climatic conditions [8]. In 2014, another crisis in the electricity system occurred due to a long drought [7], which has caused lower electricity generation from hydropower than in previous years and therefore more energy had to be produced thermally from sources as coal, natural gas, or biomass [8].

The reasons for that lack of hydrological resources are sometimes anthropogenic, as the construction of large dams can cause a reduction of rainfalls due to less evapotranspiration as a consequence of forests that had to be deforested to provide space for the reservoir [9]. Especially the North-East of Brazil is a vulnerable region for relying mainly on hydropower, as hydrological resources show high inter-annual variability, which is attributed to the El Niño Southern Oscillation [7]. Apart from droughts, another problem of hydropower is that the construction of new reservoirs poses environmental as well as social problems, like deforestation or resettlement of local residents [7] [9]. An example, where problems like this cumulated during the construction of a hydropower plant, was the case of Belo Monte [11].

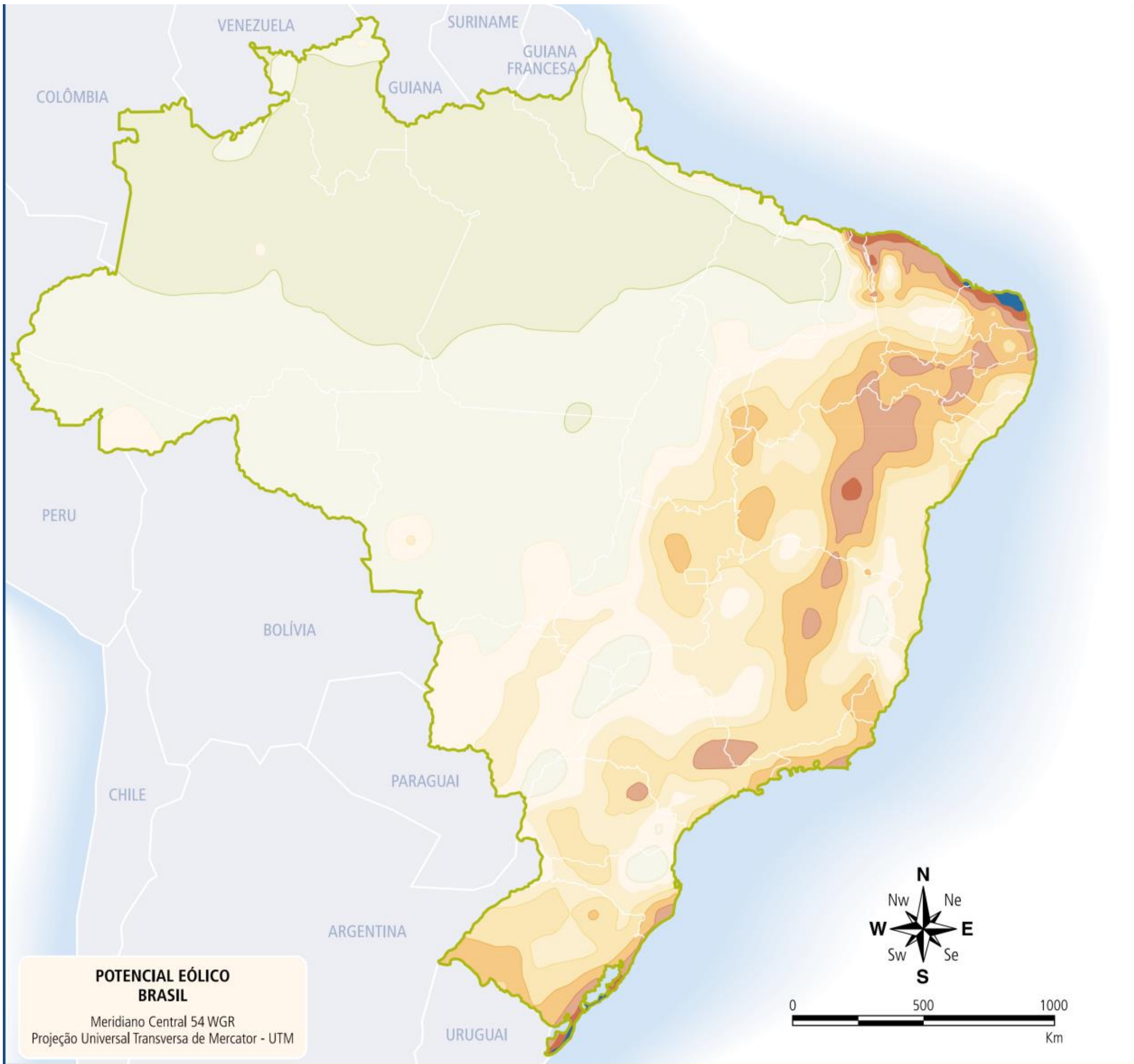
As electricity generation from fossil fuels should be reduced, or in the case of Brazil, at least not further expanded and as electricity generation from hydropower causes social and environmental problems, the necessary electricity has to be provided by other means. The supply of power from a portfolio of renewables is also more stable than a hydropower system only and there is less need for thermal power from fossil resources [7]. The optimal mix of PV and wind power with hydropower has been examined in [9] and results in 37% PV, 9% wind and 50% hydropower generation. In this case, only 2% of supply would have to be provided by thermal power and the overall risk of deficit would be decreased in comparison to a scenario with mainly hydro-power generation. In a hydro-power and thermal only scenario, the risk of deficit is enhanced tenfold, which shows the importance of incorporating enough wind power and PV into the Brazilian energy system.

According to the National Agency of Electrical Energy of Brazil (ANEEL) [12], the incorporation of wind power into the National Interconnection System (SIN) brings many advantages: reduction of emissions from thermal power plants, no necessity to build large reservoirs as well as a reduction of the seasonal risk of loss of load posed by hydropower. One disadvantage, however, are the costs of wind power, as they are about 30% higher, compared to those of hydropower [7].

Wind power has been used in Brazil since 2006, mainly in the North-East and South regions. Initially there was no notable generation, but expansion grew rapidly and in 2014 1.2% of national demand could be covered by that source [9]; in 2016 already 3.4% of demand could be covered by wind power, according to data of the National Electrical System Operator of Brazil [13] [14].

However, wind power shows strong seasonality in Brazil, where PV has a significantly lower seasonal variability [9]. It should also be considered, that wind power is a volatile source of electricity: if a system relies only or mostly on this source, some of the electricity has to be stored temporarily to be consumed later. This still poses a huge challenge in many electricity supply systems, as storing electricity is expensive [7]. However, in Brazil storage is less of a problem than in other regions, as big hydro-storage reservoirs are abundant with a total storage volume of 212 TWh [5] [7]. These reservoirs allow the balancing of volatile renewable electricity supply [9]. However, what may be a bottleneck to storing lots of intermittent energy in hydro-storage reservoirs, is the transmission grid, as electricity transmission all over the country between the locations of energy generation, storage, and consumption is limited [5]. The integration of wind power is especially a big advantage in a country like Brazil, with a high share of hydropower in the electricity generation, as wind power is temporally available in a complementary way to hydropower due to higher wind speeds in the dry than in the rainy period [15]. Schmidt et al. [5] [7] also come to the conclusion, that in the first half of the year wind power generation is lower than in the second half, contrary to hydropower generation.

The North-East of Brazil is an interesting region for wind power generation, as the geographical conditions often provide stable wind speeds. Figure 1 shows the distribution of average wind speeds in Brazil. It can clearly be seen, that the highest wind speeds occur in the North-East of Brazil. Furthermore, the expansion of wind turbines is not limited by extreme weather conditions, like in other countries where hurricanes, earthquakes or cold winters can cause disruptions to generation. This fact makes wind power generation in Brazil cheaper than in other areas of the world and has already led to lots of investments into wind energy in this region in the past years [8]. Modelling the supply of wind power is therefore a highly important measure to plan the future expansion of wind power in Brazil – and is the core aim of this thesis.



Velocidade média do vento (m/s)
50 m acima do nível da superfície

Classes de energia	Mata	Campo Aberto	Zona Costeira	Morro	Montanha
4	> 6,0	> 7,0	> 8,0	> 9,0	> 11,0
3	4,5 - 6,0	6,0 - 7,0	6,0 - 7,0	7,5 - 9,0	8,5 - 11,0
2	3,0 - 4,5	4,5 - 6,0	4,5 - 6,0	6,0 - 7,5	7,0 - 8,5
1	< 3,0	< 4,5	< 4,5	< 6,0	< 7,0

NOTAS:
Mata indica áreas de vegetação nativa, com arbustos e árvores altas.

Campo aberto refere-se a áreas planas de pastagens, plantações e/ ou vegetação baixa, sem muitas árvores altas.

Zonas costeiras são áreas de praia, normalmente com larga faixa de areia, onde o vento incide predominantemente no sentido mar-terra.

Morros são áreas de relevo levemente ondulado, relativamente complexo e de pouca vegetação ou pasto.

Montanhas representam áreas de relevo complexo com altas montanhas. O potencial eólico é dado para locais nos topos das montanhas em condições favoráveis para o fluxo de vento.

Figure 1: Mean yearly wind speed at 50 m height in Brazil (the legend for wind speeds is given in the Table in m/s, source: [12])

1.2 Use of synthetic time series in energy system models

To plan for climate change mitigation in the energy sector, it has to be assessed which shares of renewables, especially wind and solar power, should be optimally incorporated into the electricity generation system [16] [17] [18]. Energy system models are an important tool for studying the feasibility of different scenarios towards a low carbon society in order to help policy development and decision making [4] [17] [19]. They examine which impact long- as well as short-term changes in electricity generation from renewable resources have on the power grid and therefore can help in planning how to overcome the challenges they create [17] [19]. Because of these fluctuations in electricity generation from wind or solar energy, not only the average potential power generation is of interest, but also the time profile has to be determined in order to find out whether in periods of lower generation other forms of preferably renewable power generation can fill in those gaps [20]. The variability of renewable energy sources, particularly wind, influences the load in the electricity grid and thus has an effect on many factors: costs, investments in the transmission grid, emissions or electricity prices [20] [17]. Therefore, energy system models have to be fed with time series of power generation data. Observations of generation, however, are often not accessible, and of short duration, and therefore it can be a better option to simulate renewable electricity generation from weather data. As also explained in more detail in section 1.3, observed data are either not available in sufficient quality or very expensive, which is why it is a better choice to simulate electricity generation from wind or solar power based on freely available reanalysis data. For this reason, this thesis aims at creating a simulation of wind power generation to produce such time series, for the further use in simulation and optimization models of the Brazilian system.

1.3 Simulating renewable power generation

Models for renewable energy generation, which can be part of larger energy system models, play an important role for the investigation of future energy generation potentials. As observations of wind power generation are either not available or only for short time spans, such models have to be based on some kind of weather data, for which there are multiple possible sources. First of all, observed weather data, such as wind speed measurements or solar irradiation measurements, can be used. These measurements can be carried out either on or near the earth's surface or from space by satellites. One possibility is to use average values of weather data, which of course only provide average energy generation; another option is to use time series of weather data, which then also can depict changes of electricity generation throughout the day, year, or between years, which is in particular important to understand the integration of renewables into an energy system. Such measured datasets are provided for example by 3TIER, which offers remote sensing data as well as on-site measurements [21].

Another possible source of weather data are reanalysis data, which are based on measured data and interpolated to provide long-term consistent, validated time series of weather data, for example wind speeds at several heights above ground. An example for such a dataset are MERRA and MERRA-2 data;

the latter are used in this thesis. Other sources of reanalysis data are the 15 year reanalysis (ERA15) and the updated version ERA-Interim from the European Centre for Medium-Range Weather Forecasts (ECMWF), from the National Center of Atmospheric research (NCAR), from the National Centers for Environmental Protection (NCEP), from the National Center for Atmospheric Research (NCAR) as well as the NASA Goddard Earth Observing System 1 (GEOS1) reanalysis [9] [22] [23]. These different kinds of data sources have several advantages as well as disadvantages, which will be discussed in the following.

For observed data there are alternatives of high quality data, for example the before mentioned 3TIER weather data, which have a high spatiotemporal resolution and deliver very accurate results [24]. On-site measured data provide the advantage of considering local conditions, which can be useful in some cases, but not desirable in others. If only a very limited area is investigated, as for example in a study about the assessment of the wind energy potential in Maroua [25], it is possible and recommendable to use measured data, for the aforementioned reasons. If local conditions are not of interest, but rather weather data for larger areas shall be considered, it is better to use satellite data.

Noteworthy disadvantages of measured data are the limited availability of the data as well as the limited quality. For the case of wind speed measurements in Brazil, there are data available, however only for certain regions. Constraints in quality of data comprise low spatiotemporal resolution of data, often only three or six hourly mean values, incomplete data, with temporal or spatial gaps which create the time intensive need for selection of complete data, or erroneously measured data [26]. Also the change of measuring equipment can cause biases in measurement results [27]. For the case of wind speeds, measurements are often conducted only at one height, typically at 10 meters above ground and not at hub height, which can create a bias in the resulting data, as surroundings like buildings or vegetation influence near ground wind speeds. It would be more useful for the simulation of wind power generation to measure wind speeds at remote places or at heights near the hub height of wind turbines, as is also pointed out by Cannon et al. [27] as well as Pfenninger and Staffell [24] [26]. Another problem of observed weather data can be that they are measured at locations that are far from actual power plants and therefore do not represent the weather conditions there adequately [28]. A major downside of high quality measured data is that even if they may be easily obtainable, they are mostly costly, especially if larger amounts of data for longer time series or large areas are needed [24] [26].

For these reasons reanalysis data can be a better choice. One of the most important advantages is that they are easily, openly, and often freely obtainable, which makes them accessible for scientific verification and discourse. Moreover, reanalysis data are available without gaps in spatial and temporal coverage, contrary to observational data. The spatial or temporal resolution of underlying data do not influence the final modelled reanalysis data, as areas or times of missing data are filled with simulated data [23], which makes them very useful for a continuous model. Several more advantages of reanalysis data are described concisely by Cannon et al. [27], Pfenninger and Staffell [26] as well as Bengtsson et al. [23]: Reanalysis data often reproduce observed wind speeds better than measured data, because they are not biased by local conditions such as buildings or vegetation, which often influence wind speeds near the ground, or by change of equipment during the considered period of measurement. Due to the use of different kinds of data sources in reanalysis data, these kinds of biases are reduced compared to measured data. Also, though not relevant for this thesis, reanalysis data provide information for regions where measurements are difficult or impossible, like offshore. Additionally, the data are available for different heights, which is usually not the case for measured data.

An advantage of MERRA data in particular, which are used in this thesis, are their high temporal resolution of one hour and the availability of more than one height of wind speed data, which is very useful, as in other datasets wind speeds are usually only given at a height of 10 m above the ground, which is significantly lower than hub heights of commonly used wind turbines and makes it difficult to appropriately estimate wind power generation from these data [24]. Although MERRA data were found to not always be very accurate compared to measured weather data, especially for particular wind farms or wind turbines, they are more stable and also provide good results for the reproduction of electricity generation for a whole region as for example Sweden [24] [20].

What can be seen as a disadvantage of reanalysis data is that they are a global data source, where local conditions may not be perfectly reproduced. However, using additional local information can improve the quality of results [24]. In a study about wind speeds and wind power generation in the UK [27], where MERRA reanalysis data were used, it has been found that the mean value of wind speeds of measured and reanalysis data is very similar, but wind speeds for various locations sometimes are different. The authors conclude that reanalysis data often underestimate the variations in wind speed compared to observed wind speed data. Staffell and Pfenninger [24] list similar reasons why reanalysis data may be erroneous or not very accurate: as the data are just the output of a model, which does not reflect reality perfectly, there can be a systematic bias. The spatial resolution is not very high, so the local topography is disregarded. Furthermore it is questionable, whether the wind speeds are able to reproduce wind conditions for certain sites, as they are only based on ground measurements and satellite data. Furthermore they point out, that results from the use of reanalysis data have to be handled with care, as they may vary from one region to another and cannot be applied to all countries or areas around the world.

For several countries bias-corrected reanalysis data have already been used to simulate wind power generation, for example in Great Britain [27] or Ireland [28]. What is new for this thesis is that for the first time an hourly dataset of bias-corrected wind power generation in the North-East of Brazil will be made publicly available and furthermore not only wind power generation but also wind speed bias correction is conducted.

1.4 Aim and structure of the thesis

This thesis aims at simulating potential hourly wind electricity generation in selected areas in Brazil. These simulations will be compared to wind speed measurements and also observed wind power generation data in the subsystem “Nordeste” of Brazil, in order to determine bias correction factors for wind speed and power generation, in a similar way as it has been done in [24] for Europe. The North-East region of Brazil is chosen, because wind speeds, especially in coastal regions, are higher than in other regions of Brazil (see Figure 1) and therefore most of the wind power plants are located there [12]. Also, seasons of lots of wind coincide with seasons of little rainfall, which means that wind power could complement hydropower generation in this area [9] [10]. Wind power from the North-East region of Brazil is especially useful for this, more than from the South region, where also a lot of wind power capacity is already installed [5]. Hence, it will be possible to estimate the potential of wind power in case of a full expansion on a 37-year-basis. The analysis of the data is carried out using the statistical software R.

The following research questions will be analysed:

- Can a model based on MERRA reanalysis data adequately depict daily and monthly observed wind power generation?
- How well can the model be adapted to observed data by using bias correction methods? How does bias-correction with measured wind speeds compare to bias-correction with measured power generation?
- Which extreme events of very low or very high production can be identified in the final time series of wind power generation?

2 Data & Methods

This chapter describes the data and methods used in this thesis. In order to develop a model of wind power generation, the wind speed data from the MERRA reanalysis project will be bias-corrected using hourly measured wind speed data. From the modelled wind speeds, the potential power generation can be simulated and verified with observed monthly and daily generation data and a second set of bias correction factors is calculated. Finally, the calculated bias correction factors are used to model the power output over 37 years. In Figure 2, a scheme of the approach is depicted.

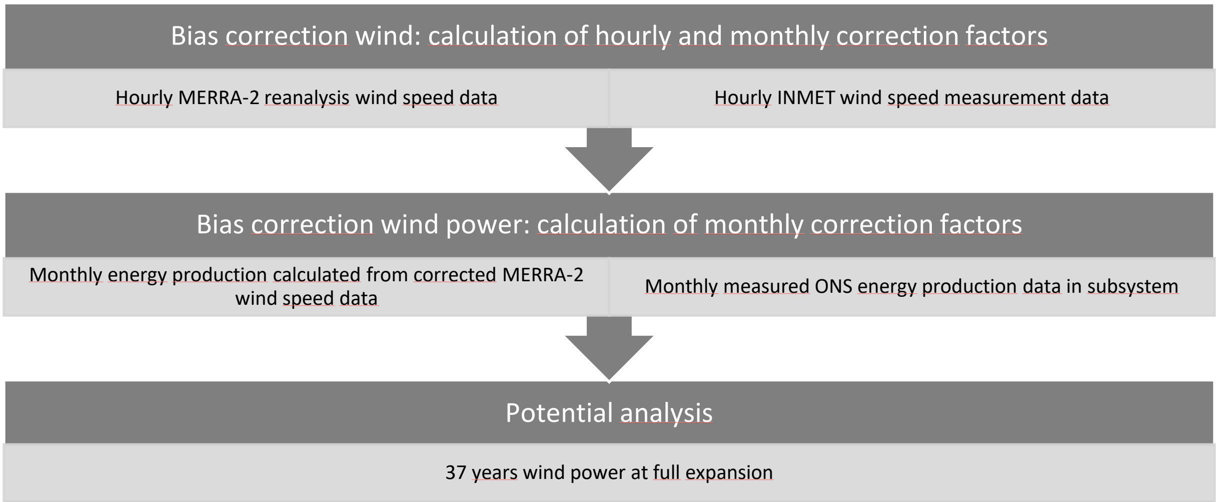


Figure 2: Approach for the analysis of the data (own representation)

2.1 Data: Access & Preparation

The data used for this thesis contain several wind measurement and reanalysis datasets as well as datasets on observed wind power generation. Altogether four different sources of data are utilised:

- MERRA-2 reanalysis data, which are modelled hourly wind speed data (hourly, 1980 - October 2016)
- INMET wind speed measurement data (hourly, 1999 – April 2016)
- Monthly generation data from the national electrical system operator of Brazil (monthly, 2002 – October 2016) and
- Daily generation data from the national electrical system operator of Brazil (daily, August 2015 – October 2016)

As the MERRA-2 data are no measured data and are produced globally, the other two datasets are used to adapt them to real local conditions. Each of these datasets will be described in more detail in the following.

2.1.1 MERRA-2 data

The MERRA-2 reanalysis data are provided for free online by NASA, the National Aeronautics and Space Administration. MERRA-2 stands for Modern-Era Retrospective analysis for Research and Applications, Version 2. Compared to the first version of the data set (MERRA), there have been improvements in the assimilation system. The data are available starting in since 1980 and is updated monthly [29].

In order to be able to download data, it is necessary to create an account at https://urs.earthdata.nasa.gov/. After registration, one has to log-in to the NASA GESDISC Data Archive before downloading the data. With the application wget64, which allows to download data from http and ftp servers, the data can be downloaded after creating the links to them with the help of the NASA MDISC application.

The used dataset MERRA-2 can be accessed via MDISC at http://disc.sci.gsfc.nasa.gov/daac-bin/FTPSubset2.pl%20, where the desired data product has to be chosen. The subset used is "tavg1_2d_slv_Nx", which stands for hourly, time averaged, single level, assimilation, single level diagnostics. It is two-dimensional and has a resolution of 1 hour. The data are averaged hourly, starting at 00:30 UTC (Universal Time Coordinated) every day [30]. For further calculations, we assume a shift of half an hour in generation (i.e. the 00:30 timestamp is assumed to be valid at 01:00).

As only the area of the North-East of Brazil is relevant for the calculations, an area as small as possible around this subsystem is selected. The chosen boundaries in geographical coordinates are North: 1°S, South: 19°S, East: 34°W and West: 49°W, as can be seen in Figure 3.

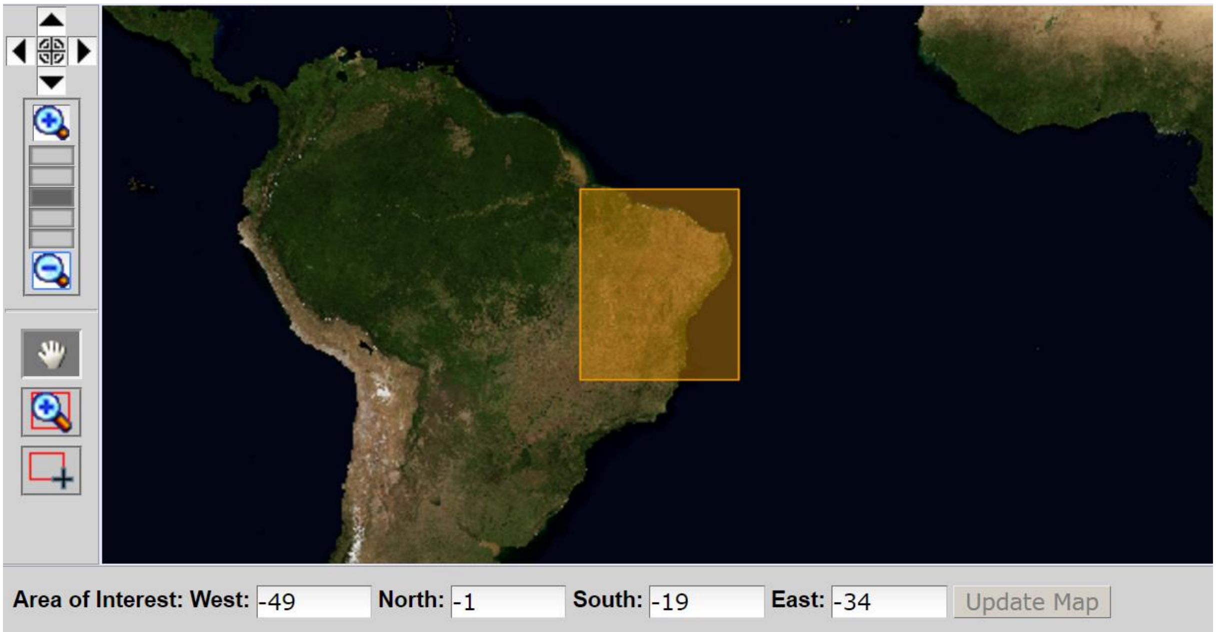


Figure 3: Selected area for MERRA-2 data in the North-East region of Brazil (download from [31])

As the spatial resolution is 0.625 degrees in longitude and 0.5 degrees in latitude, which is about 50 kilometres in distance [30], this amounts to 888 (24*37) points in the grid. Figure 4 shows the location of these points around the North-East region of Brazil.

Figure 4: Distribution of MERRA-2 grid points (own depiction)

The extracted parameters are the displacement height (DISPH) and the east- (U) and northward (V) wind at 2 and 10 meters above displacement height, as well as 50 meters above the surface. The time window used is 1st of January 1980 until 31st of October 2016, as this was the time span for which data was available at the time of the download. The format for the files to be downloaded is selected as ".nc-files". A list of filenames is created and downloaded.

To download the files, it is necessary to be logged in with the before created account, which is done via a log-in file. To create this log-in file the following commands are typed into the command line:

```
type nul .netrc

echo "machine urs.earthdata.nasa.gov login <uid> password <password>" >> .netrc
```

<uid> has to be replaced with the login name and <password> with the corresponding password. Also

```
type nul .urs_cookies
```

is used to save cookies. The download is started with the following command:

```
wget64 --load-cookies ~/.urs_cookies --save-cookies ~/.urs_cookies --keep-session-cookies --
content-disposition --user <uid> --password <password> -i myfile.dat
```

Again <uid> is replaced with the login name and <password> with the actual password.

The files downloaded amount to 13089 files, one for each day of the timespan regarded. Each of this files contains wind speeds at different heights and in two different wind directions, as well as the disposition height for each of the 888 points in the MERRA-2 grid in the selected area for all 24 hours of this day.

The format in which the files are saved is called netCDF-4, which stands for Network Common Data Form and is the fourth version of the netCDF [32] [33] file format. This is a special file format for saving and processing data used in geoscience education and research, developed by Unidata, which is part of the University Corporation for Atmospheric research. The advantage of using this file format is that it is “portable” and “self-describing”, which makes it independent from the machine it is used on, and it can be simply read with software libraries provided for a range of programming environments. NetCDF-4 is the latest version of netCDF and was sponsored by the Earth Science Technology Office of NASA. It is compatible with older versions of netCDF and offers some new functionalities like chunking [34].

2.1.2 Power Curve

For calculating power output from wind speeds, the power curve of a wind turbine is needed. The power curve indicates the energy output at a certain wind speed for a particular wind turbine. The actual wind turbines that are used in the single wind power plants are not known, therefore a standard turbine is chosen for the model. The wind turbine assumed for these calculation is the ENERCON E-82 with a rated power of 2000 kW, a rotor diameter of 82 m and a height of 108 m. This wind turbine was chosen, because it is a relatively cheap solution of generating electricity from wind, as mentioned in [25] and was also consulted in a study about the economic viability of the construction of a wind park in Parnaíba [35]. Figure 5 shows the power curve of the used wind turbine ENERCON E-82. A value of 0 kW power generation at 0 m/s wind speed has been added for simplification in the interpolation function.

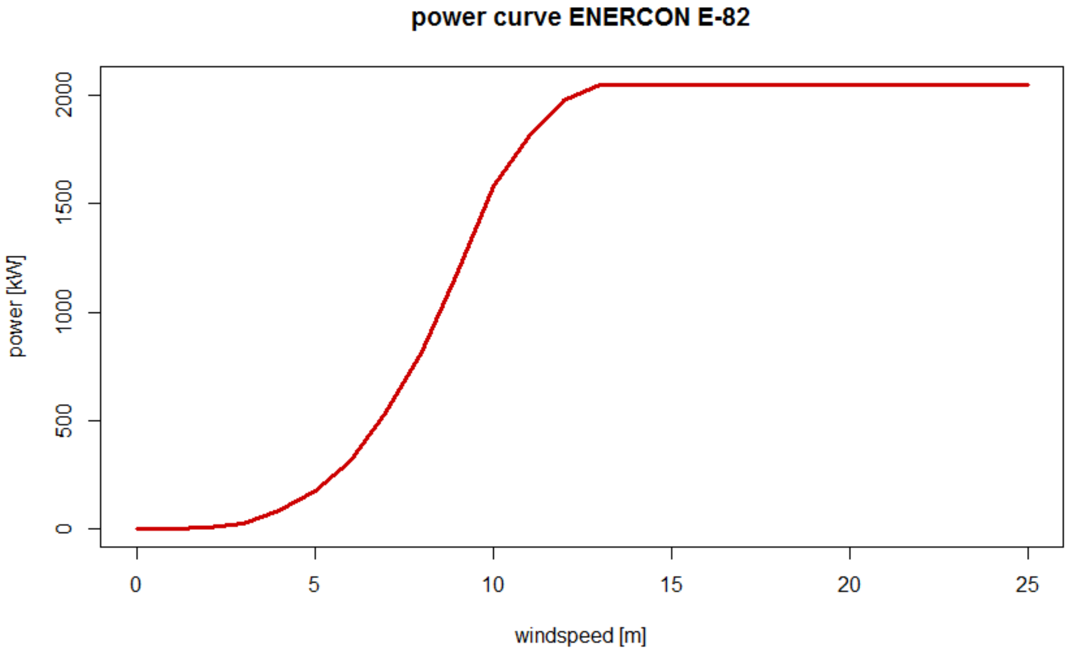


Figure 5: Power curve of ENERCON E-82 (own depiction with data from [36])

2.1.3 Data used for bias correction

As global reanalysis data alone are not necessarily representative of local conditions, measured data are used to validate and bias correct, respectively, the reanalysis data. In our approach, two types of data are utilised: measured wind speeds and observed wind power generation.

The measured wind speeds are provided by the meteorological institute of Brazil (Instituto Nacional de Meteorologia, INMET) and can be downloaded from its homepage[1]. The wind speed is measured at 10 m height above the earth surface at 841 stations in Brazil and in some cases outside of Brazil (as for example the Projeto Criosfera in the Antarctic). For each of these stations a separate file with time series of wind speed measurement data is downloaded. Only 128 of these stations are used (indicated by red dots in Figure 6), as the others are outside of the North-East region of Brazil and therefore irrelevant for the calculations in this model. The stations are chosen by state they are located in, indicated by two letters in the file name. The nine relevant states are coloured in blue in Figure 10 (see page 17): Alagoas (AL), Bahia (BA), Ceará (CE), Maranhão (MA), Paraíba (PB), Pernambuco (PE), Piauí (PI), Rio Grande do Norte (RN) and Sergipe (SE). For each INMET station, the geographical coordinates are given, as well as hourly wind speed measurements. Also other information, like temperature,

[1] http://www.inmet.gov.br/portal/

humidity or precipitation, is provided in the dataset. However, it is not used in this wind power simulation model.

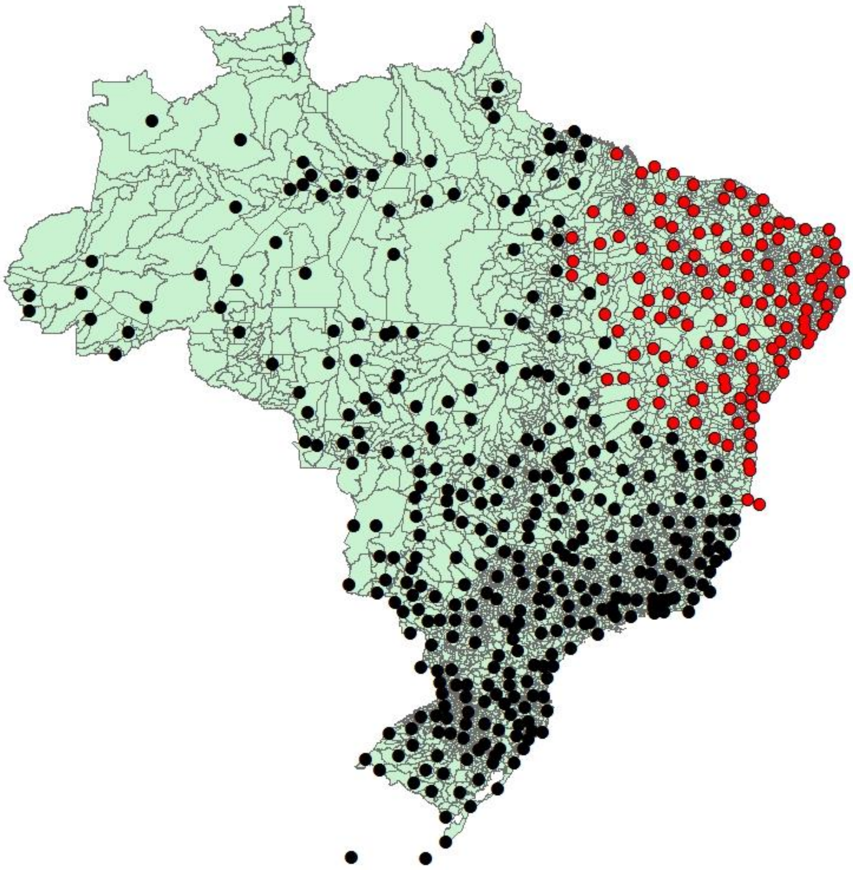


Figure 6: Location of INMET stations: the red dots indicate the stations considered (own depiction in ArcGIS with INMET meta data and a map of the municipalities in Brazil from the CodeGeo homepage [37])

The datasets are available for the period 1st of January 1999 00:00:00 - 20th of April 2016 00:00:00. However, continuous data are not necessarily available from the beginning of 1999. In most cases the measurements were started after the initial date and in addition, in many cases the series of values are interrupted in certain time periods. This means that the INMET data, unlike MERRA-2 data, are not complete, which makes it necessary to disregard the corresponding MERRA-2 data to prevent the distortion of results. Figure 7 shows an example of the hourly mean wind speed data of the year 2015 of the INMET station number 12 in Bom Jesus da Lapa, with hours starting with 1st of January 2015. The gaps in the data can be seen clearly.

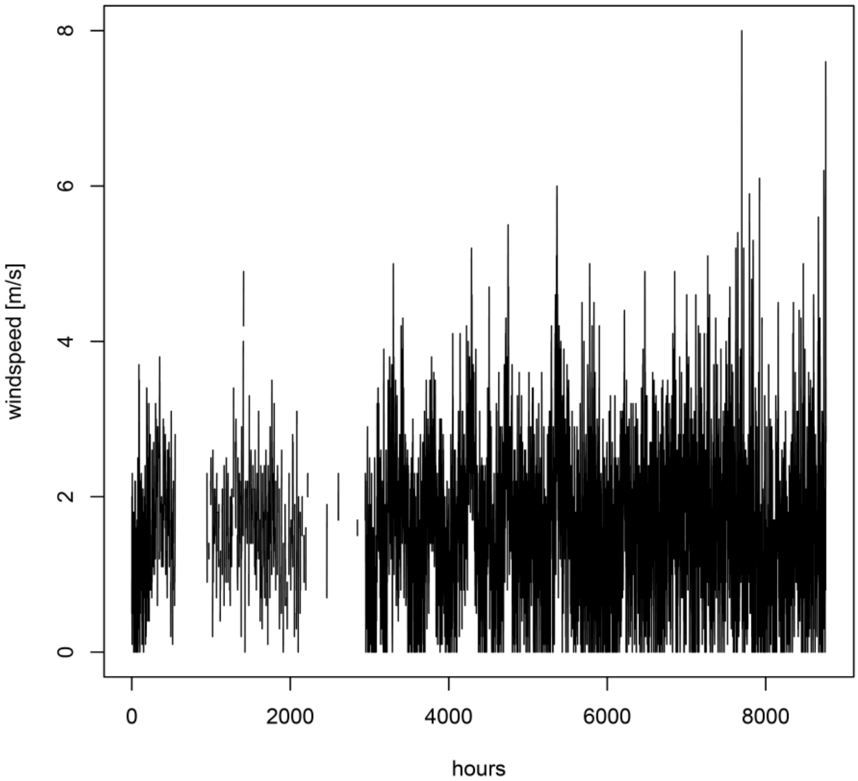


Figure 7: Wind speeds measured by INMET in 2015 in Bom Jesus da Lapa (own depiction of INMET data)

In contrast to that, the data at the same INMET station were complete in 2010, as can be seen in Figure 8.

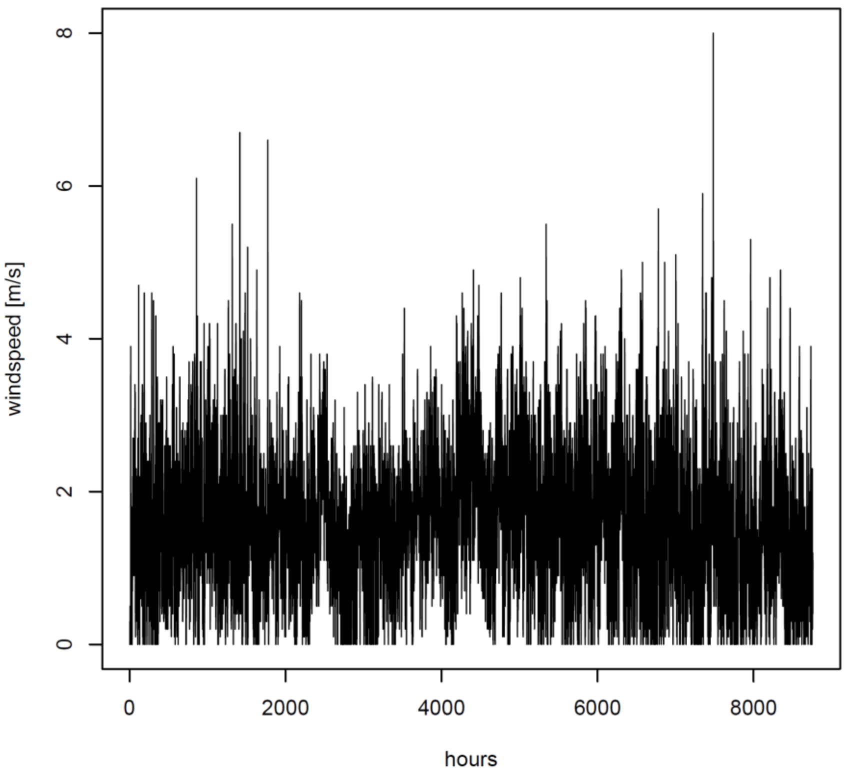


Figure 8: Wind speeds measured by INMET in 2010 in Bom Jesus da Lapa (own depiction of INMET data)

Some data are considered erroneous, as over longer periods of time a wind speed of 0 m/s is shown. Figure 9 shows an example for this in the year 2013 at station number 119 situated in Calcanhar. For a time span of about 4000 hours starting from 1st of January 2013 00:00:00 (about five to six months, i.e. until the middle of May) there are either data missing or a wind speed of 0 m/s is indicated, with only three peaks. These obvious errors in the data have to be considered in the further analysis.

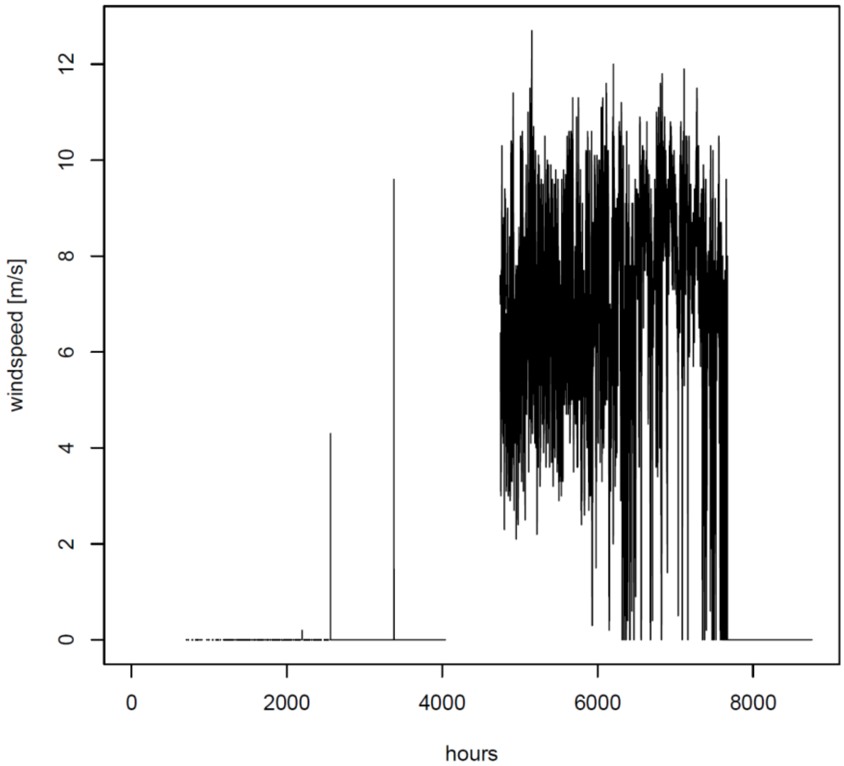


Figure 9: Wind speeds measured by INMET in 2013 in Calcanhar (own depiction of INMET data)

The second dataset used for bias correction is recorded wind power generation data. On the homepage of the national electrical system operator of Brazil (Operador Nacional do Sistema Elétrico, ONS) http://www.ons.org.br/, data on the monthly wind power generation per Brazilian subsystem between 2002 and 2016 can be found. The subsystem of interest here is North-East, whose location and federal states can be seen in Figure 10. For the North-East of Brazil only data since 2006 are available because significant wind power generation only started in that year.

Figure 10: Subsystems in Brazil (Subsistemas do Brasil, source: [38])

In order to be able to calculate the power generated at a certain wind speed, it is necessary to know the installed capacity. Installed capacities as well as the generation start dates of wind power plants in Brazil can be found on the homepage of the Brazilian Electricity Regulatory Agency (Agência Nacional de Energia Elétrica, ANEEL) [39], when selecting "EOL" in the first Table "Empreendimentos em Operação" (operating enterprises). Figure 11 shows a bubble chart of the installed capacities per municipality until November 2016. As can be seen in the image, the highest wind power plant density as well as most of the installed capacity are in the North of the North-East region of Brazil on the North(Eastern) coast of Brazil as well as in the centre of Bahia. Figure 12 shows the distribution of wind speeds in the North-East of Brazil and it is obvious, that wind power plants have been built where wind speeds are the highest.

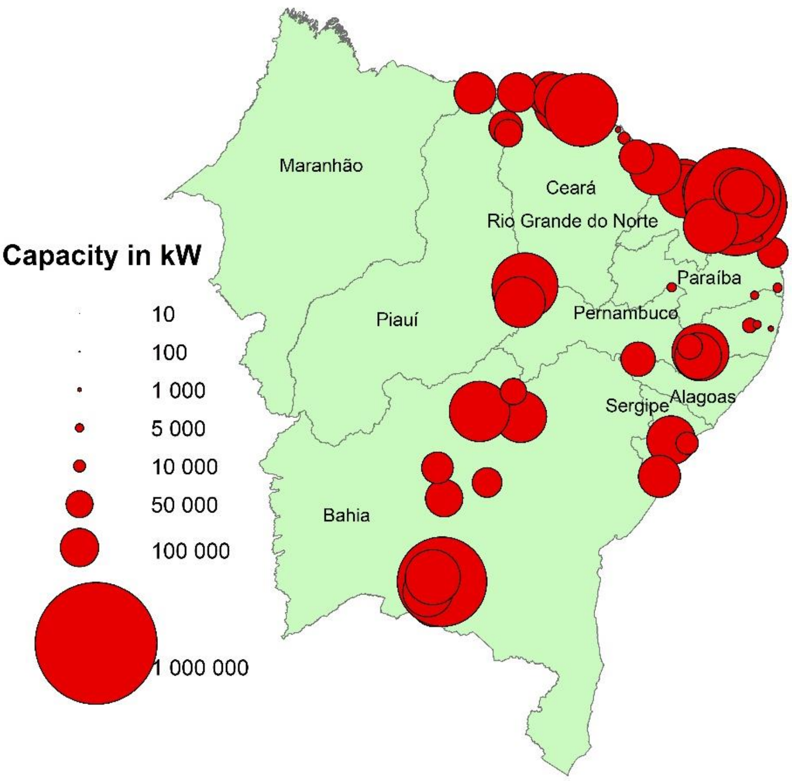


Figure 11: Installed capacities per municipality (own depiction with data from [39] in ArcMap, data until November 2016)

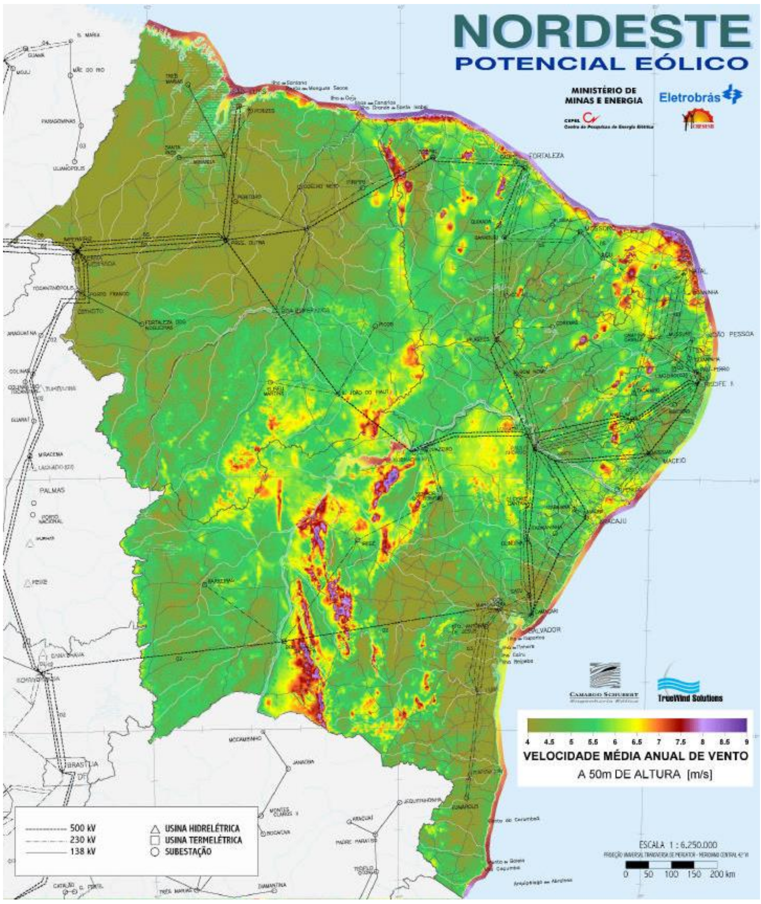


Figure 12: Mean annual wind speeds at 50 m height in the North-East of Brazil (source: [40])

The table of installed capacities is not directly downloadable. As a consequence, the homepage was saved as html and converted to a .csv-file online [41] with the settings ";" as a separator, "remove HTML tags in CSV" and "replace multiple spaces with 1 space in CSV", and downloaded. After the download, it is necessary to edit the data, to make them useful for the calculations: special characters are not converted correctly and therefore have to be replaced manually. Also, the "new lines" in the last column are removed by searching for them with the search function for "Strg+010". Furthermore, in the last column the "-" are replaced by ";" to separate municipality from the state code (two letter abbreviation), with the aim to select only the states in the North-East of Brazil. There are two entries (Ventos de Santo Augusto V, Ventos de São Clemente 7), where two municipalities are listed. In these cases, only the first municipality listed is adopted.

As the exact position of each wind power plant is not given in the data, the centre of the municipality in which it is located is assumed as geographical position. To obtain the geographical coordinates of the centres of the municipalities, the shapefile is loaded into ArcGIS and the x and y coordinates of the shapes' barycentres are determined and associated with the data of each wind power plant by the name of the municipality it is located in. This way, for each municipality the sum of installed capacities with according geographical coordinates is known. The data are adapted using Excel: The capacity has to be multiplied by 1000 to convert it from MW to kW, so that it is in the same dimension as the data from the power curve. The date is changed to another format of yyyy-mm-dd as well, to make it readable for RStudio. The file is saved as csv after the data have been ordered by municipality. A second file is created where the sums of the capacities installed per municipality are calculated by using Pivot tables in Excel.

Four of the start-up dates of the wind power plants in the North-East of Brazil were missing. Therefore, information about these wind power plants was researched on the Internet to find approximate start-up dates. The first of these power plants with a missing start-up date is located in Aquiraz. No exact start-up date could be found but on the website of the magazine Grandes Construções it is stated, that the wind park "Parque Eólico Prainha" in the municipality of Aquiraz with a capacity of 10 MW was inaugurated in 1999 [42]. Hence, a start-up date of 1st of January 1999 was assumed. The second missing start-up date was from a wind power plant in Brejinho. No reliable or exact data about this power plant could be found. It was assumed that the start-up date of the wind power plant "Ventos do Brejo A-6" was in 2012, as on the page of the power plant on ANEEL website the "Registro" has the number 2012 attached [43]. Thus, the 1st of January 2012 is assumed as start-up date. The next missing start-up date is from a wind power plant located near Fortaleza. The inauguration date of the "Usina Eólica do Mucuripe" was in 2002 [42], hence the start-up date is defined as the 1st of January 2002. The last power plant with a missing start-up date is situated in the municipality Macau. No official information about this wind power plant was found either, but in an article it was mentioned that the wind park was inaugurated in 2004 [44]. Therefore, the start-up date was determined as the 1st of January 2004.

2.1.4 Data used for validation

For the comparison of data after correction, also daily data of energy produced from wind provided by the ONS are used, which have been available only since recently. These data are available since 1st of August 2015 and are updated once per day. As MERRA-2 data were only downloaded until the 31st of October 2016, the data only until this date are used for comparison, which allows for a comparison of a period of one year and four months.

2.2 Methods

The following chapters describe the methods used in the development of the wind power simulation model. After collecting the data from the different sources as described in chapter 2.1, they are adapted to a format so that they can be used for further calculations. The procedure of simulation is structured into four main parts which will be explained in more detail:

- Extrapolation and correction of wind data
- Correction of wind power generation data
- Validation with daily wind power generation data and
- Simulation of 37 years of wind power generation and analysis of the results

Starting from MERRA-2 reanalysis data of wind speed at different points, an extrapolation to the desired height is conducted, followed by the first bias correction with wind speed data from INMET. After that correction, a second bias correction is performed with wind generation data from the national electrical system operator of Brazil. Finally, 37 years of wind power generation at full expansion can be simulated, using the correction factors calculated previously.

The calculations are performed with the statistical computing software RStudio version 1.0.136. As a prerequisite for RStudio, R version 3.3.2 is installed. Four packages are loaded that are needed in the development of the model and for the analysis of the data:

1) ncdf4: The package “ncdf4” is needed to read files of the format netCDF version 4, in which the MERRA-2 files are downloaded. The functions “nc_open” and “nc_close” are used to open

the netCDF files and by means of “ncvar_get” the desired variables are read from the particular files.

2) lubridate: The package “lubridate” contains, among others, the functions “year” and “month”, which are applied to dates of the class “POSIXct” to extract the year and month, respectively from the date. At several occasions, year or month need to be extracted to aggregate data per month or to correct data depending on the month they occur with the corresponding correction factor.
3) plotly: The package “plotly” is used for plotting heat maps, to analyse correction factors, and contour plots, which are useful for the comparison of observed and simulated power generation. By passing an argument of the required plot (in these cases either “heatmap” or histogram2dcontour” as type) to the function “plot_ly”, the desired graph is generated.
4) hydroGOF: The package “hydroGOF” (Goodness-of-fit functions for comparison of simulated and observed hydrological time series [45]) is used to calculate the RMSE (Root Mean Square Error) between simulated and observed wind power time series with the function “rmse”.

2.2.1 Extrapolation and correction of wind data

For working with MERRA-2 data it is necessary to load them into RStudio, using the “ncdf4”-package. Three reading functions are implemented to read data from the netCDF files: The first reading function “datum” reads the dates from each file and replicates them 24 times for each hour of the day. The hour is added and the date converted to the class “POSIXct”. A list of dates with 24 entries in each list element is created, later unlisted with the “unlist”-function of R and saved as a vector on the hard drive. The second reading function “getlonlatMRR” reads the 24 longitudes and 37 latitudes of the MERRA-2 grid points and expands a grid among those points, resulting in 888 geographical coordinates. The third reading function “readMRRvar” reads a variable, that has to be passed to the function, and matches it to the longitudes and latitudes, that were read by the second function. For each day a matrix with the 888 grid points in the rows and 24 hours in the columns is created. The whole dataset is saved as a list of matrices for each single day. This is done for each of the five variables (U50M, V50M, U10M, V10M and DISPH) separately because the memory was exceeded when trying to load all the data since 1980 at once.

In order to make the access to MERRA-2 data faster and easier, the data were rearranged in such a way that for each of the 888 grid points one file is saved with the time series of the five variables. This results in 888 files, each containing a data frame with six columns named “date”, “u50m”, “v50m”, “u10m”, “v10m” and “disph”. The files are named “ll***MRRdf.RData”, where *** is replaced by the position of the longitude and latitude in the list of longitudes and latitudes. This way data for one of the grid points can be easily accessed by opening one of those files, which will be needed often during the calculations. Although this claims a fair amount of time for calculation and saving, it is unavoidable, as the working storage is not large enough to deal with such a vast amount of data.

The first step of preparing the model intends to bias-correct the reanalysis data with observed wind speed measurements of INMET. To compare the data, the nearest neighbouring MERRA-2 grid point to each of the INMET stations is located. The method of the Nearest Neighbour is chosen in this case, as it is a simple method with relatively low computational cost. Other methods of interpolation were tested too, as also described and performed in [46]. The methods Bilinear Interpolation, Bicubic Interpolation and Inverse Distance Weighting were assessed with four INMET stations. The stations were selected randomly. Table 1 shows correlations of the different methods.

Table 1: Correlations of the four assessed interpolation methods for four INMET stations (own representation)

Method	**Station 1**	**Station 2**	**Station 42**	**Station 66**
Nearest Neighbour	0.70	0.64	0.34	0.22
Bilinear Interpolation	0.69	0.60	0.34	0.24
Bicubic Interpolation	0.61	0.47	0.33	-0.11
Inverse Distance Weighting	0.69	0.62	0.34	0.24

As can be seen in the table, more complex methods do not increase correlations significantly compared to the simple Nearest Neighbour method. Only in the case of station 66 slightly better correlations, except for the Bicubic Interpolation, were obtained. However, this may be due to a very low correlation of only about 22% with the Nearest Neighbour method. Therefore, the Nearest Neighbour method was selected as the final method, as its results are comparably good and also less computing effort is necessary compared to the other methods.

The method of the Nearest Neighbour calculates the distance of each INMET station to all the 888 MERRA-2 grid points according to the spherical law of cosines:

$$distance = r * acos(sin(lat1) * sin(lat2) + cos(lat1) * cos(lat2) * cos(lon2 - lon1)),$$

where r is the mean radius of the earth (6378.388 km), lon1 and lat1 are the longitude and latitude of the INMET station in radians and lon2 and lat2 are the longitudes and latitudes of the MERRA-2 grid points in radians. As RStudio trigonometric functions are implemented for calculating with radians, but the geographical coordinates are given in degrees of longitude and latitude, the values have to be multiplied by the factor $\frac{\pi}{180°}$ to convert them to radians. The spherical law of cosines thus is used for calculating the distance between two points on the surface of the earth and is also applied for example in GPS Tracking Systems [47].

The calculated distance is saved to a data frame along with the according longitudes and latitudes of the MERRA-2 grid point as well as a numbering from 1 to 888. Afterwards, the rows of the data frame are ordered by distance with the “order” function, which results in a data frame where the first entry is the nearest neighbour. The number then can be used to open the according file with the data of the nearest neighbour. For example, for INMET station number 63 with the longitude -44.79499° and the latitude -4.243056° in Bacabal in the state Maranhão, the nearest MERRA-2 grid point has the geographical coordinates -45° longitude and -4° latitude with a distance of about 35.4 km. Therefore, the index of the nearest MERRA-2 grid point is 727 and the file “ll727MRRdf.RData” is loaded to obtain the wind speed data for this point.

To compare the reanalysis data with the wind speed measurement data, it is also necessary to adjust them to the same level of height above ground. INMET data are measured 10 meters above the ground, MERRA-2 reanalysis data, however, are provided at 50 metres above the ground and 10 metres above displacement height. This makes it necessary to extrapolate the data, as the displacement height is not necessarily 0. For this purpose, the power law with the ground surface friction coefficient α is used:

$$v_2 = v_1 \left(\frac{h_2}{h_1}\right)^{\alpha},$$

with v_1 and v_2 being wind velocities at different heights h_1 and h_2. It would also be possible to use other methods of extrapolation, as described in [46], which would also consider more than two altitudes. However, as shown in [46], different vertical interpolation methods do not change results significantly.

The resulting wind speed (v_{res}) must be calculated from the east- (v_{east}) and northward (v_{north}) wind speeds for each height, by using the theorem of Pythagoras:

$$v_{res} = \sqrt{{v_{east}}^2 + {v_{north}}^2}$$

The friction coefficient α depends on the structure of the surface: the smoother the surface, the smaller the friction coefficient, the coarser the surface, the higher the friction coefficient [48]. Table 2 gives a few examples of typical friction coefficients depending on the surface.

Table 2: Friction coefficient α depending on surface of the earth (source: [48, p. 45])

Terrain Type	Friction Coefficient α
Lake, ocean, and smooth, hard ground	0.10
Foot-high grass on level ground	0.15
Tall crops, hedges, and shrubs	0.20
Wooded country with many trees	0.25
Small town with some trees and shrubs	0.30
City area with tall buildings	0.40

However, surface data for each point in the North-East of Brazil were not available. For this reason, the friction coefficient is calculated by using wind speeds in two heights (50 metres above the ground and 10 metres above displacement height). The power law equation is rearranged to the following equation:

$$\alpha = \frac{\ln(v_2) - \ln(v_1)}{\ln(h_2) - \ln(h_1)}.$$

v_2 is the resulting wind velocity at 50 metres above the ground, v_1 is the resulting wind velocity at disposition height plus 10 metres, h_2 is 50 metres and h_1 is disposition height plus 10 metres added. The calculated friction coefficient subsequently is inserted into the power law equation to calculate the wind velocity at 10 metres above the ground to compare it to the wind velocities measured by INMET. In the model this is done by the function "extrap".

Afterwards the INMET and MERRA-2 data are aligned so that they cover the same time span. As MERRA-2 data exist since 1980 and INMET data only since 1999, the first 19 years of MERRA are dropped. Also, MERRA-2 data were downloaded until the 31st of October 2016, but INMET data are only available until the end of April 2016, which makes it also necessary to cut the last six months of MERRA-2 data. Another operation that is needed, is to remove missing data, as INMET data are incomplete. For this purpose, INMET and MRERA-2 data are joined into a data frame and with the help of the function "na.omit" only full rows of data are retained.

The next step is the calculation of the hourly and monthly correction factors. For this purpose, the months of the dates are extracted. The hours were numbered before omitting missing values. Month and hour are joined in to the format "mmhh". This way, the monthly and hourly wind speeds can be added up by using the function "aggregate", resulting in 12*24 = 288 sums, one for each hour of every month. Then the two datasets are compared and 288 correction factors are calculated by dividing the hourly and monthly sums of the INMET data by the MERRA-2 data.

Afterwards the correction factors can be used to correct the wind speeds of the MERRA-2 dataset by multiplying each wind speed with the correction factor according to the corresponding hour of the day and month. Thus, the correlation after the correction compared to the correlation without correction can be enhanced. Furthermore, the sum of wind speeds after correction of MERRA-2 data is the same as for the INMET data.

Other methods of correction were tested as well: hourly and monthly correction separately, resulting in 24 respectively twelve correction factors. However, the results were worse than with combined hourly and monthly correction. Table 3 shows the correlations resulting from different correction methods (no correction, only hourly, only monthly and hourly and monthly correction combined) for four of the INMET stations. As can be seen, the best results (the highest correlations) are obtained with the hourly and monthly correction combined, which is why this method is selected for the further simulation. It can also be observed that the monthly correction alone mostly increases correlation only slightly, whereas the hourly correction does so to a higher extent. This may be due to higher differences in wind speeds depending on the hour of the day than depending on the season, despite noticeable differences in wind speeds in the different seasons [35].

Table 3: Correlations of time series of four INMET stations with time series of MERRA reanalysis data with and without different bias correction methods (own representation)

Correction	**Station 1**	**Station 2**	**Station 42**	**Station 66**
no	0.70	0.64	0.34	0.22
hourly	0.77	0.77	0.39	0.35
monthly	0.72	0.64	0.35	0.24
hourly and monthly	0.81	0.78	0.41	0.41

However, we do not use bias-correction for all locations, but apply a heuristic to decide whether to correct or not: The correction factors will only be applied if

1) there is a correlation of more than 50% after the correction and
2) the INMET station and the MERRA-2 grid point are no more than 80 km apart.

When checking the data, it was found that most correlations were above 50% and most distances below 80 km. As it is desired that most of the data are corrected, these limits were chosen, excluding only a few with very low correlations or far distances. Furthermore, it was assumed that the correlation should be at least 50%, because at lower correlations, data would rather be distorted than improved by applying the correction factors.

All in all, too low correlations applied to nine INMET stations, which were excluded (automatically by if-statement) for this reason. Causes of the low correlations were checked by assessing the location of the stations as well as the data by plotting them. As already described in chapter 2.1.2, some datasets were lacking lots of data or apparently erroneous data (lots of 0 m/s wind speed) were incorporated. Stations, where this was the case, are shown in Table 4. Remarks on the data quality are also shown.

Table 4: INMET stations with anomalies in the data (own representation)

Station Number	Location	Remarks
24	Ilheus	Lots of 0, 2008 only 0
78	Areia	2008 nearly only 0
90	Garanhuns	2013 fluctuation between 0 and higher values, beginning of 2014 only 0
98	Alvorada do Gurgeia	2009 outliers, 2013 and 2014 only 0
119	Calcanhar	Lots of 0, especially 2013

Moreover, Table 5 shows all stations excluded because of their low correlation and discusses possible anomalies in their location. They were determined by searching for the coordinates on Google Maps and examining the surroundings (height, vegetation, site density). High buildings, vegetation, mountains or valley location may influence the wind direction and intensity.

Table 5: INMET stations with anomalies in their location (own representation)

Station Number	Location	Remarks
12 36 66 78 98 103 119	Bom Jesus da Lapa Piatã Buriticupu Areia Alvorada do Gurgeia Esperantina Calcanhar	Ground objects such as buildings or trees are located close to the station and measurements may therefore be distorted in terms of representing wind speeds at hub heights
24 90	Ilheus Garanhus	On an elevation of about 100 m above the surroundings/ On high plateau of Borborema -> wind turbines are also located on elevation, therefore no obvious reason for low correlations was found

2.2.2 Correction of wind generation data

After determining the correction factors for the measured wind speeds, the next step is the bias-correction of the original reanalysis data and the reanalysis data corrected with measured wind speeds to observed generation data from the national electrical system operator of Brazil. As the data for a larger time span of several years exist only in monthly resolution for the whole North-East subsystem, the MERRA-2 data also have to be monthly aggregated for the whole North-East. The applied approach is depicted in Figure 13. In the following the single steps will be explained in more detail.

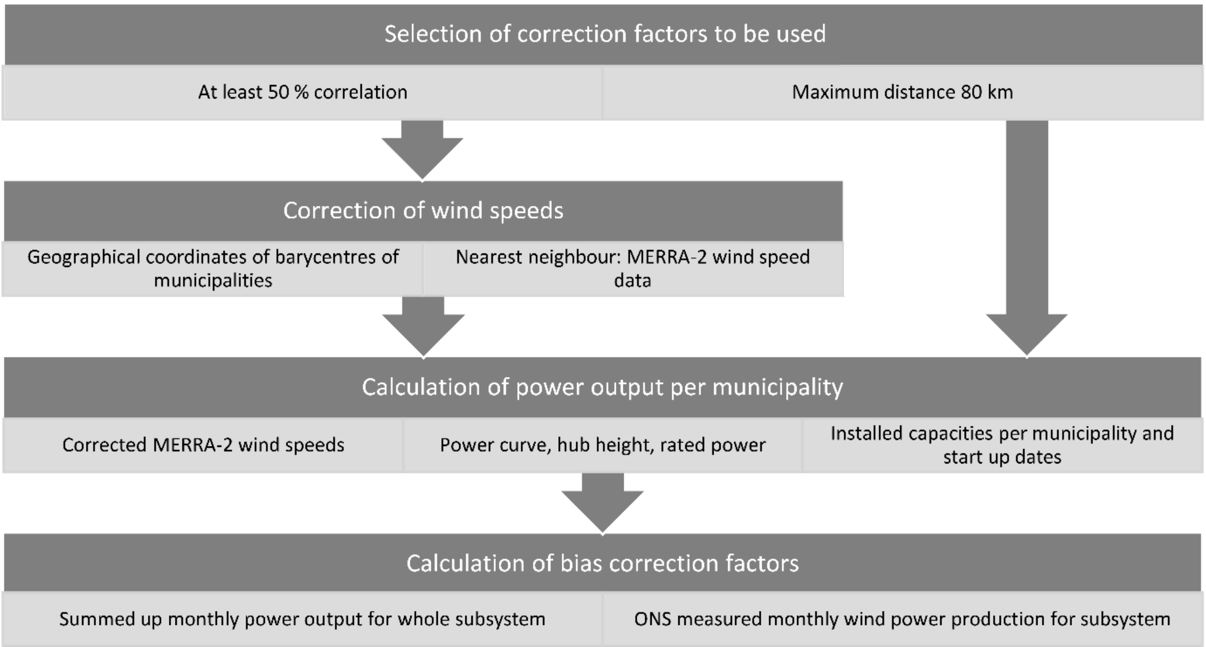


Figure 13: Approach of the wind power correction (own depiction)

To be able to bias-correct, the power resulting from the wind speed has to be calculated. This happens in the model by the function "calcmunpower". The rated power (2 MW), the hub height (108 m), the power curve (as two vectors of wind speed and the resulting power), as well as the correlation limit (50%) and the maximum distance of the INMET station (80 km) are passed to the function. The latter two are used, as explained before in chapter 2.2.1.

As mentioned before, the exact positions of the wind power plants are not known. For this reason, the centres of the municipalities they are located in are assumed as locations. The power is therefore calculated per municipality. It would be easy to just sum up the installed capacities per municipality, however, the wind power plants have different start-up dates. So, at first the data of the installed capacities are read into R, with the relevant data being the geographical coordinates of the centre of the municipality, the installed capacity and the start-up date of each of the wind power plants. As the data are sorted by municipality, the rows of the power plants in this municipality are extracted to a separate data frame. In that data frame, only the capacities and the start-up dates are saved. The longitude and latitude are saved in separate variables. The data frame is then sorted by start-up date and a list of the cumulative installed capacities per date is created. A list of capacities, in the length of the examined time span is created, where the entire capacity installed in the municipality at each time is stored.

Afterwards the nearest neighbour of the MERRA-2 grid points is searched by calculating the distances and applying the function "distanceorder", as described before. The data are loaded from the rearranged and saved files and are then corrected, if the limits of maximum distance between MERRA-2 grid point and INMET station, as well as the minimum correlation value are met.

Next, the power generation has to be calculated. This is done with the help of the power curve, shown in Figure 5, which indicates the power output from the selected turbine Enercon E-82 for a specific wind speed. This standard turbine is chosen, as no exact data of the installed turbines in the North-East of Brazil are known. Only certain points of the power curve are known. That is why the wind speeds between the given points have to be interpolated to receive the power output. For this purpose, the nearest point below and the one above the respective wind speed are determined. The interpolation is done for each point of time by a simple linear interpolation equation:

$$power = \frac{capacity}{ratedpower} * \left(\frac{power_1 - power_2}{wind_1 - wind_2} * (wind_m - wind_2) + power_2\right),$$

where "capacity" is the installed capacity in a municipality at a certain time, and "ratedpower" is the rated power of the wind turbine (2 MW). The number of turbines per wind power plant is calculated by dividing the installed capacity by the rated power and multiplying it with the power produced at that time. "$wind_1$" and "$power_1$" are the lower neighbouring data from the power curve, while "$wind_2$" and "$power_2$" are the upper neighbouring data from the power curve and "$wind_m$" is the (corrected) wind speed occurring at a certain time in the middle of the municipality. For wind speeds that are above the highest wind speed in the power curve, the power output is set to that of the highest wind speed (25 m/s) as the power curve flattens in the end and a maximum power output of 2 MW is reached. Finally, the energy produced can be determined from the power by multiplying with the time, which in this case is one hour, as the data are in hourly resolution.

When the electricity generated over time has been calculated for all the municipalities, it needs to be summed up and compared to the observed electricity generation. First, the calculated powerr output is dropped before the year 2006, as the data for comparison only exist since 2006. Then the calculated power output is aggregated by month for each municipality separately. Next the monthly sums over the municipalities are aggregated by month, resulting in 12 monthly sums for the whole North-East region for the period of 2006-2016. The sums are divided by 10^6 to convert them from kWh to GWh, the same unit as the observed data. The latter are also aggregated to 12 monthly sums. Then the measured power generation is divided by the calculated power generation and 12 correction factors for the whole North-East are obtained.

Afterwards, the whole data are corrected with the monthly correction factors for the North-East and the original correlation is compared to the correlation of the corrected data with the measured data.

A different approach was also examined: here, simulated as well as measured wind power generation are normalised with the installed capacity at that time before calculating the wind power generation correction factors. However, as can be seen in Figure 14, the simulated generation after wind and power correction (green line) underestimates the observed generation (black line), especially in recent years, more than the only wind corrected simulated generation (blue line). Therefore this approach does not result in better adapted time series, which is why it was discarded.

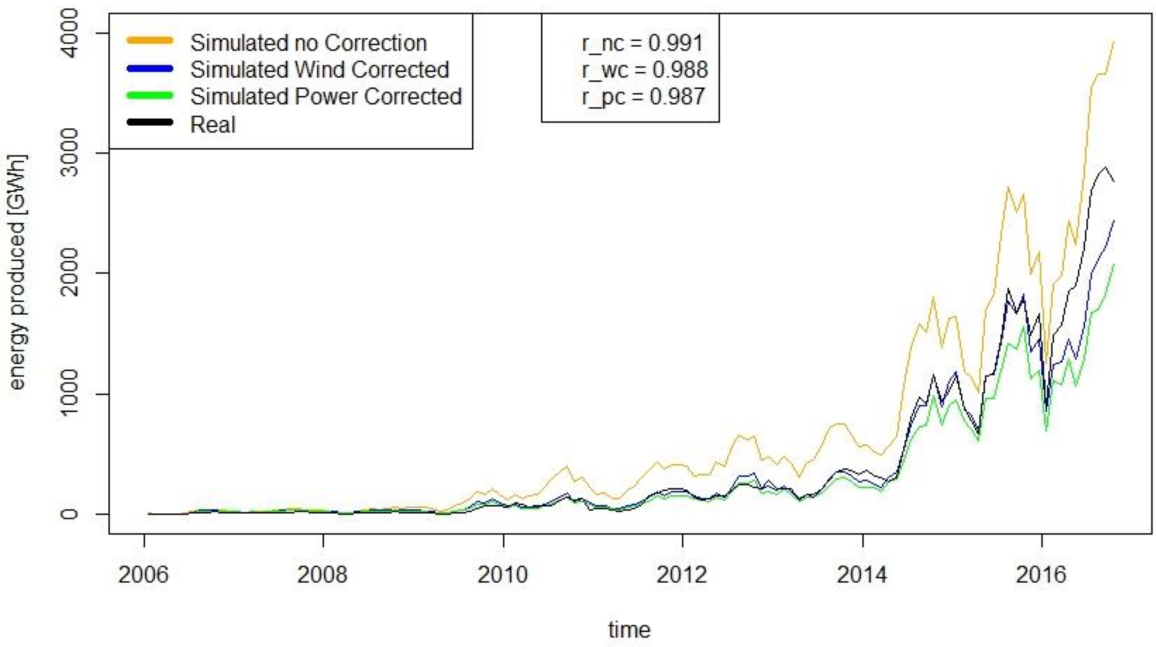


Figure 14: Illustration of the discarded approach of normalising generation with installed capacity (own depiction)

2.2.3 Validation with Daily Wind Power Generation Data

Similarly to the monthly wind power generation data, an analysis with daily generation data is conducted. The observed daily generation is compared to the simulated generation with and without application of different bias correction methods, in order to find out how well daily fluctuations in wind power generation can be simulated.

2.2.4 Simulation and Analysis

After all the correction factors have been calculated, finally a simulation of 37 years of wind power at full expansion can be simulated. For this purpose, for all the existing wind power plants, the sums of installed capacities are assumed as fixed and the electricity generated over the whole timespan of existing MERRA-2 data is calculated. The wind speeds at the centres of the municipalities are corrected with the hourly and monthly correction factors previously computed by comparing reanalysis data to INMET data. Again, the correlation limit of 50% and the distance limit of 80 km are assumed and data are only corrected if the respective values complied with the limits. Afterwards, the monthly correction factors for wind power generation for the North-East are applied as well. A sum over the municipalities as well as the time is calculated and the results indicate a possible wind power generation for 37 years if the current installed capacity stays the same.

Furthermore, a statistical analysis of the data is carried out. For the wind speed bias correction, scatterplots are created and correlations are calculated to determine interrelations of the correlations of INMET and MERRA-2 wind speeds with the distance to the nearest neighbours or the share of missing data. A possible impact on the change of correlation when applying the hourly and monthly correction factors is examined as well. The scatterplots are created using the "plot" function in RStudio and setting the type to "p", for points.

Moreover correlations before and after the hourly and monthly wind speed correction are compared in a boxplot. Also the changes in correlations are examined. Boxplots are generated with the "boxplot" function in R.

To find out how the wind speeds behave over time for certain stations, plots of hourly mean wind speeds for ten days are created, comparing the INMET wind speeds to the MERRA-2 uncorrected and corrected wind speeds. From these plots it can be visually determined how well the trajectory of the simulated and observed wind speeds match and also whether the simulation over- or underestimates the observed wind speeds and what changes are caused by the correction. As there are 128 INMET stations, not all of them could be examined, but certain ones were picked: The INMET stations with the lowest correlation with the MERRA-2 data before and after correction, the stations with the highest correlations with MERRA-2 wind speeds before and after bias correction, the stations with the lowest and highest change in correlation when applying the hourly and monthly correction factors as well as stations with medium correlation before and after the bias correction. The plots were generated with the "plot" function in RStudio, setting the type parameter to "l", for line.

For the same set of INMET stations, the hourly and monthly correction factors are examined. As it would be difficult to analyse 288 correction factors displayed as numbers, heat maps are drawn in R with the “plot_ly” function from the “plotly” package. The function parameter type is set to “heatmap” and a heat map is created where blue rectangles stand for low values and red rectangles for high values. However, the shade values differ between heatmaps. The heat maps help to determine in which times of the day or months of the year the wind speeds have to be lowered or heightened to bias-correct them and also when measured wind speeds are over- or underestimated (Examples can be seen in Figure 45 to Figure 53).

Furthermore, two boxplots are created for all correction factors, one with the correction factors segregated into twelve monthly boxplots and one segregated into 24 hourly boxplots. This way, it can be found out in which hours or months the wind speeds need to be lowered or increased. As some correction factors are significantly higher than the average, not the whole plot is displayed, but only the part where the boxes and therefore the main part of the data lie (see Figure 17 and Figure 16).

All these parameters and plots serve as illustration of the effect of the application of correction factors and furthermore show how observed and simulated data differ.

The next part of the analysis comprises the monthly wind power bias correction. The correction factors are presented in table, as there are only twelve of them for each method (with and without preceding wind speed correction). Furthermore, the daily, monthly and yearly aggregated electricity generations from wind power are compared with line plots and the correlations of the data before and after monthly bias correction are determined. The comparisons for the monthly and yearly generation data are in a time span between January 2006 and October 2016. The daily comparison is only exerted between August 2015 and October 2016, as no earlier data are available.

The corrected and uncorrected simulated data are compared to the measured data also by boxplots, to see how the measured data are over- or underestimated and also how the correction influences the data. Also the errors (differences between observed generation data and simulated or corrected simulated energy outputs, respectively) are compared as boxplots. There are no boxplots created for the aggregated yearly electricity generation from wind power in the North-East of Brazil, as there are too few data for these plots to be meaningful.

A histogram of power classes is created only for daily means of power, as only in the daily resolution sufficient data are available. The histogram is created using the “hist” function in RStudio.

To compare the observed and simulated power generation, contour plots are used. These plots are generated in R using the “plot_ly” function, by setting the parameter type to “histogram2dcontour”. Eight plots are generated, comparing the observed generation to the simulated generation before and after correction with and without previous wind speed correction, for monthly as well as for daily generation data. These plots help to determine whether the simulation over- or underestimates the observed generation and also in which amounts of generated energy.

Furthermore the statistical parameters correlation, mean, RMSE and minimum and maximum difference between measured and simulated wind power generation are determined and displayed in a table.

3 Results

This chapter presents the results from the analysis and correction of the MERRA-2, INMET and ONS data. Original reanalysis and corrected data are compared to observations and statistical parameters are calculated to show the effect of the bias correction. Also time series and correction factors are examined to find out whether the reanalysis data and the energy output calculated from them over- or underestimate the observed data.

3.1 Wind Speed Bias Correction

This chapter presents the results of the correction performed with hourly and monthly correction factors obtained by comparing MERRA-2 reanalysis data to observed wind speed data by INMET. Several factors are examined, which may have an influence on the fitness of the model. Furthermore the correction factors are examined to find out whether observed wind speed values are over- or underestimated.

The detailed analysis can be found in the appendix (section 6.1). The results are summarised in this section.

First of all, influences of different factors on the correlation between measured and simulated wind speed data were examined. It is shown, that neither the distance to the nearest neighbour, the share of missing values in the INMET data, nor the location (per state) have an influence on the correlation or change in correlation after bias-correction.

What could be determined, however, is that bias-correction in most cases has a very positive influence on the correlations. The correlations are always increased, only in a few cases, where correlation was very low initially, they remain low after correction. The improvement of correlations varies a lot, ranging from nearly 0 percentage points up to more than 50 percentage points (see Figure 15).

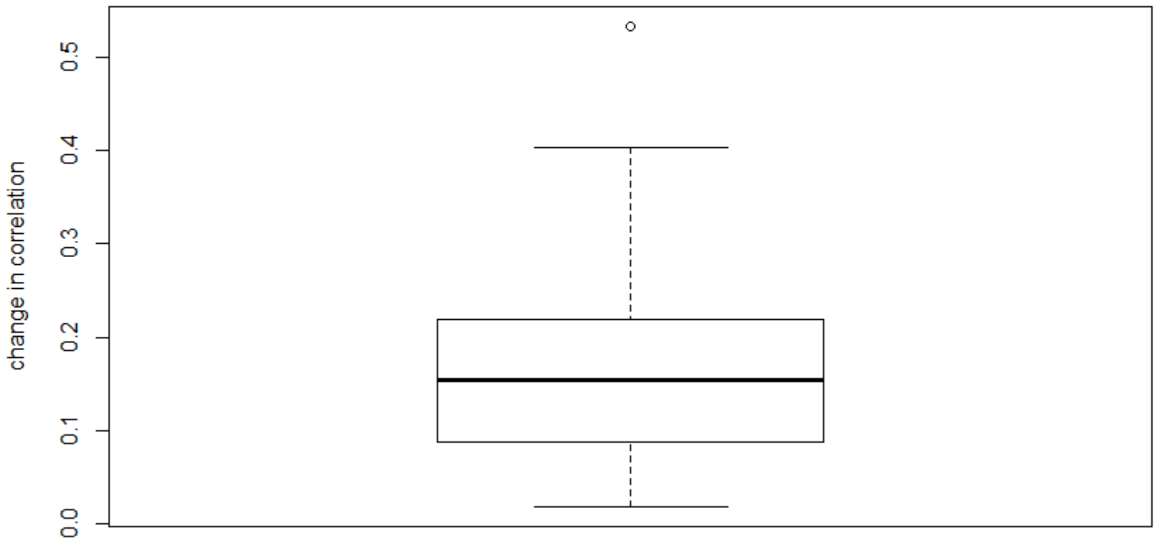


Figure 15: Boxplot of the changes in correlation with the monthly and hourly correction (own depiction)

When examining the correction factors it can be observed, that they show significant differences between locations, sometimes being lower than 1 and therefore reducing the simulated wind speeds, and other times being as high as 50, significantly increasing the simulated wind speeds. This shows, how important it is to calculate these factors for many different locations in order to adapt the simulated data well to measured data. Furthermore it could be determined that there is hardly any seasonal variation of correction factors, they are only slightly lower in summer months (see Figure 16).

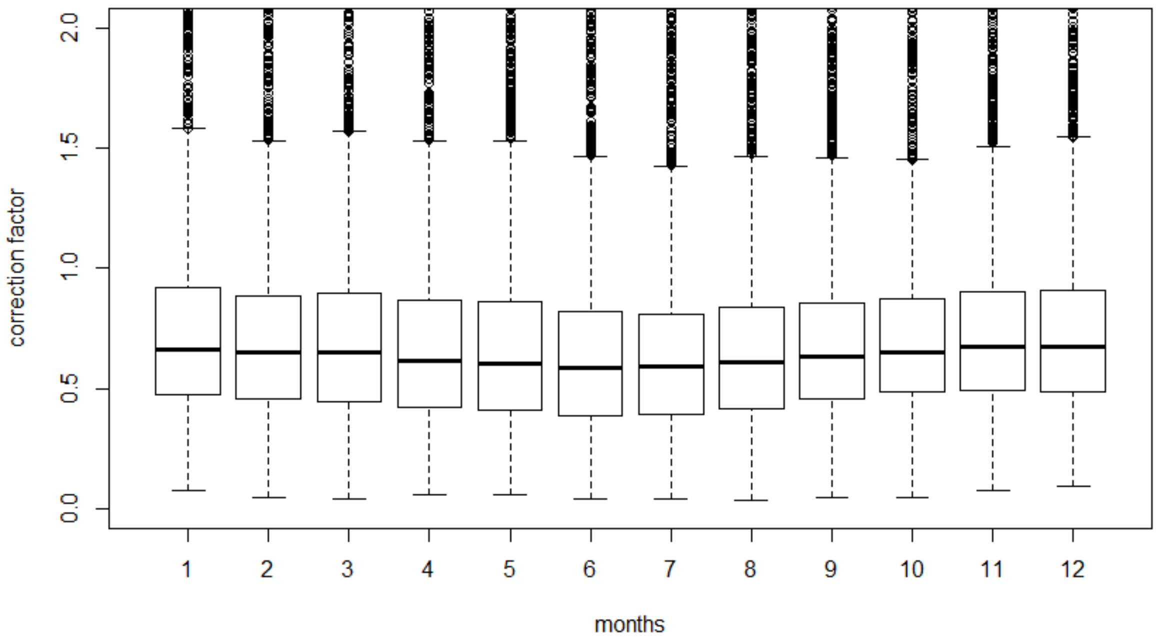


Figure 16: Boxplot for correction factors per month (own depiction)

However, there is a remarkable difference in correction factors during the day, as they are significantly lower in late morning hours and higher in afternoon and night hours (see Figure 17).

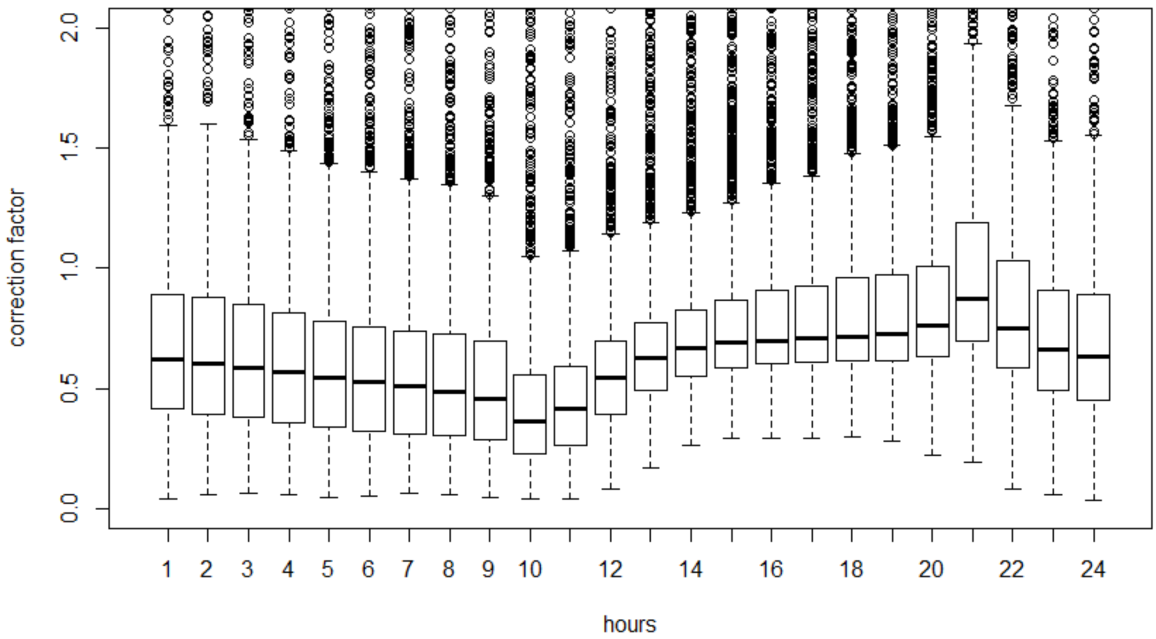


Figure 17: Boxplot for hourly and monthly correction factors per hours (own depiciton)

To sum up, applying the hourly and monthly correction for wind speeds, in most cases, has a very positive influence on the correlation of observed and simulated wind speeds, fitting the time series of reanalysis data better to time series of wind speed measurements. It can also be observed, that especially the hourly correction makes a big difference. The correction factors vary a lot, depending on the location. Sometimes there is a clear distinction between higher and lower correction factors in the evening or morning hours, respectively.

3.2 Wind Power Generation Bias Correction

This section of the results shows the results of bias-correction of wind power generation with monthly electricity generation data from wind power. The daily, monthly and yearly correlation and trajectories of simulated, corrected and observed wind power generation data are examined for the whole of the North-East region. In the implementation of the model, twelve monthly correction factors were determined twice, once with uncorrected wind speed data and once with wind speed corrected wind speed data, which are displayed in Table 6.

Table 6: Correction factors for the North-East of Brazil (own representation)

Month	Correction factors without Wind Correction	Correction factors with Wind Correction
January	0.63	1.02
February	0.66	1.08
March	0.66	1.11
April	0.63	1.11
May	0.65	1.15
June	0.61	1.15
July	0.60	1.13
August	0.63	1.12
September	0.64	1.08
October	0.63	1.02
November	0.63	1.05
December	0.63	1.03

All of the correction factors calculated from data without wind speed correction are between 0 and 1 and thus reduce the amount of energy produced when applying them, which indicates that the simulation overestimates the observed power generation. It can also be observed that in the months from February to May, apart from April, the correction factors are slightly higher (between 0.65 and 0.66) than in other months. The lowest correction factors are calculated for June and July where they lie at 0.61 and 0.60, respectively. This means that in the spring months the simulated generation is slightly better adapted to observed generation and in summer months the observed generation is overestimated the most by the simulation. When applying the wind speed correction beforehand, the correction factors are all slightly above 1 (between 1.02 and 1.15), which means that the simulated wind power generation underestimates the observed wind power generation. The highest correction factors occur in spring and summer months between March and August (between 1.11 and 1.15), whereas in the autumn and winter months between September and February they are all below 1.1 (between 1.02 and 1.08). This means that in spring and summer observed wind power generation is more underestimated by the simulated generation than in autumn and winter months. In general, it can be said that in summer months, especially June and July, simulated wind power generation needs to be corrected the most, compared to other months, as in these months the correction factors deviate the most from 1.

Figure 18 shows the comparison of daily wind power generation in GWh for the whole North-East region of Brazil. The orange line is wind power generation calculated from uncorrected MERRA data. The blue line is the wind power generation calculated from MERRA-2 data, where the wind speed has been corrected. The red line is the wind power generation calculated from uncorrected wind speed data but corrected with monthly wind power generation correction factors. The green line is the corrected electricity generation (corrected with hourly and monthly correction factors derived from INMET wind speed measurements as well as with monthly correction factors derived from ONS wind power generation data) and the black line shows the observed electricity generation from wind power provided by the national electrical system operator of Brazil. As daily comparison data are only available since 1st of August 2015, the comparison is done only for a timespan of one year and three months. It can be seen, that the trajectory of the not corrected line fits quite well to the observed electricity generation, it only is about twice as high, whereas the power generation calculated from wind speed corrected MERRA-2 data is lower than the observed power generation for the later months of 2016 but sometimes slightly lower or higher before that (in 2015 as well as in the earlier months of 2016).

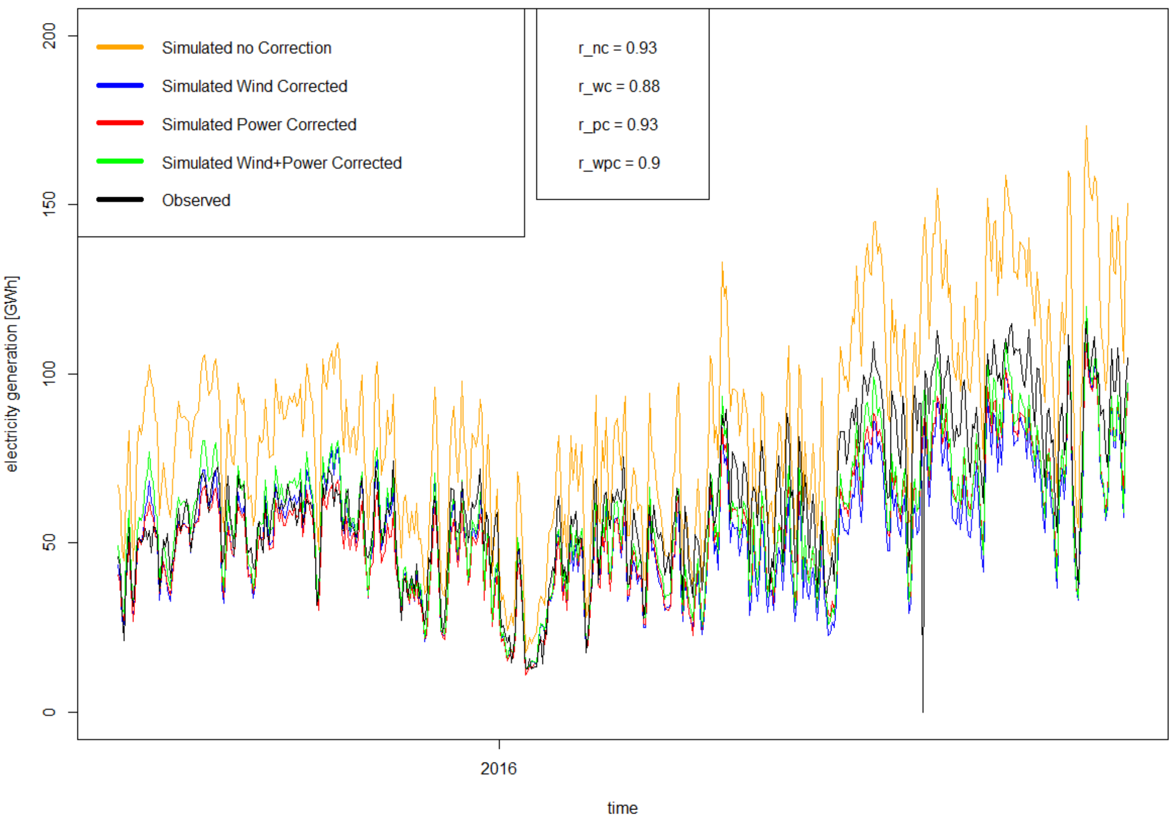


Figure 18: Daily observed and simulated wind power generation in the North-East of Brazil compared (own depiction)

When looking at the wind power generation data that have only been monthly power corrected, it can be seen that their trajectory fits observed generation much better than the uncorrected data. Also the trajectory of wind power generation calculated from wind speed corrected wind data seems to fit the observed wind power generation well. When comparing the two power generation corrected time series, it cannot be determined which one fits the measured daily wind power generation better: The trajectory of data with power correction only shows better fitting in the early months, where the simulation only slightly underestimates observed wind power generation, whereas the trajectory of the time series that have also been wind corrected overestimates the measured data. However, in the later months, the trajectory of the data on which correction factors, calculated from INMET data was well as the ones from ONS data, have been applied, seems to fit better to the trajectory of the observed wind power generation, as they underestimate the observed generation less than the only power corrected data. In terms of correlations, no correction (r_nc) or only power correction (r_pc) seem to bring the best results, as they correlate with 93%, whereas only wind corrected data (r_wc) correlate with only 88% and wind and power corrected data (r_wpc) with 90%. Without performing wind speed correction, the power generation correction does not improve correlation, whereas when wind speed correction is applied, the correlation deteriorates. However, one must bear in mind, that correlation alone is not an indicator for the goodness of fitting.

Figure 19 only partly supports what has been ascertained from the preceding figures: The positions of the boxplots show, that the simulated data without correction overestimate and the simulated data with correction underestimate the observed daily power generation data, at least for the higher and middle values. Furthermore, it can be seen, that with the correction the range of variation of the data is reduced, as the boxes as well as the whiskers are smaller. The variation of the uncorrected simulated data, which lie in a range between about 20 to nearly 180 GWh per day, is larger than the one of the measured data, which has values between about 20 (disregarding the outlier) and approximately 120 GWh per day. Also compared to the corrected simulated data, which range between about 20 and 100 to 110 GWh per day, the variation of uncorrected data is obviously higher. The simulation, which seems best adapted to the observed wind power generation data in terms of distribution, is the one with wind speed as well as wind power generation correction, as the median as well as the whiskers are closest to the measured data, compared to the other simulations.

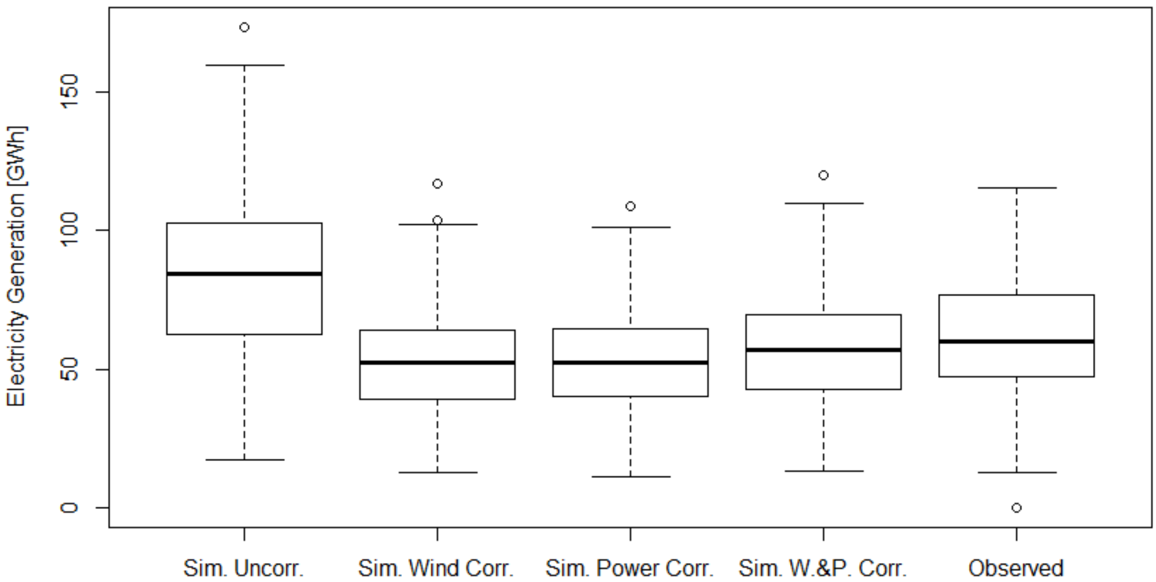


Figure 19: Boxplot for comparison of daily power generation in the North-East of Brazil (own depiction)

In Figure 20, from the left to the right, the absolute differences between observed daily wind electricity generation and simulated uncorrected, simulated wind corrected, simulated power corrected as well as simulated wind and power corrected wind power generation are displayed. When looking at the absolute differences in GWh represented as boxplots, a clear improvement in the differences can be observed when bias-correction is applied, as the median is lower and some differences are even at or very close to 0 GWh. The best results again seem to emerge from wind speed correction with subsequent wind power generation correction.

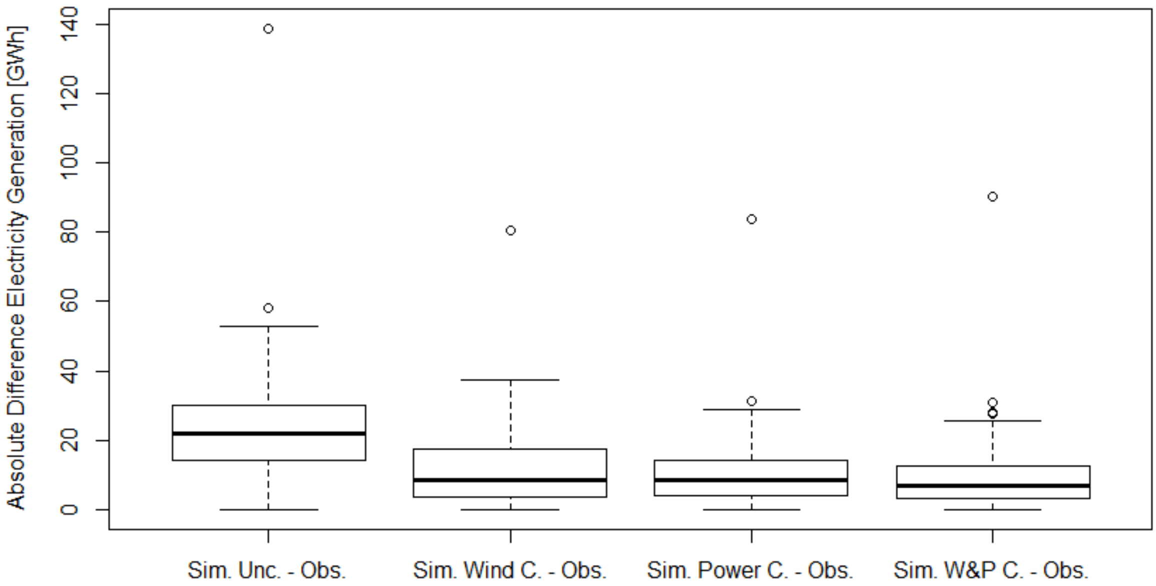


Figure 20: Boxplots for comparison of absolute differences between measured and simulated uncorrected or corrected daily wind power generation (own depiction)

In Figure 21, the real differences between measured and simulated (on the left) or wind and/or power corrected simulated (the three boxplots on the right) daily power generation are displayed in GWh. Here it can be seen, that before the bias correction the measured wind power generation data are overestimated and afterwards underestimated for the examined period, as the first lie above 0 GWh and those after correction mostly below 0 GWh (always nearly 75% of the daily generations). Furthermore, the differences are nearer to 0 GWh after the correction, which means that the simulated data better fit to the observed data in terms of absolute differences. The dataset with differences closest to 0 GWh is the one with wind speed and wind power generation corrected. Table 7 additionally displays the minimum and maximum difference of simulated and observed daily wind power generation.

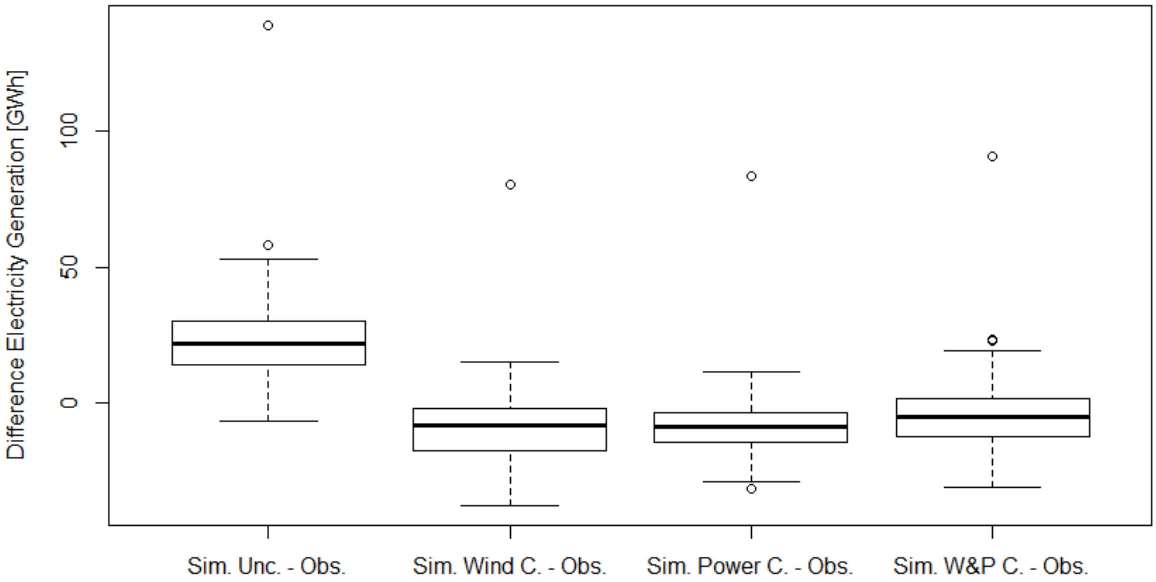


Figure 21: Boxplots for comparison of differences between measured and simulated uncorrected or corrected daily wind power generation (own depiction)

In Table 7, a comparison of the correlations between the resulting data from different corrections and observed data as well as the RMSE, the means, the minimum and the maximum difference to the observed data of each of the time series are displayed. The best correlations occur with uncorrected data as well as with only power corrected data. However, only looking at the correlations does not indicate how well the two time series fit, which is why also other parameters like mean and RMSE should be consulted. As can be clearly seen, the mean of the uncorrected data is significantly higher, at 85.67 GWh, than the mean of the measured wind power generation at 62.81 GWh. In contrast to the means of the monthly comparison in Table 8, the means of the daily comparison are not all the same despite bias-correction, because only a sub-period is considered. The mean closest to the mean of observed generation results from the wind speed and wind power generation correction, which again underpins the assumption that this simulation provides the best fitness to the observed wind power generation data. Also the lowest RMSE of only 11.32 GWh, which indicates the difference between measured and simulated time series, supports this.

Table 7: Comparison of correlations, means, RMSE, minimum and maximum of differences to observed data of different methods for daily wind power generation (own representation)

Method	**Correlation with Observed Generation**	**Mean [GWh]**	**RMSE [GWh]**	**Minimum Difference to Observed Data [GWh]**	**Maximum Difference to Observed Data [GWh]**
No Correction	0.93	85.57	26.12	-6.60	138.70
Wind Correction	0.88	53.02	14.65	-37.46	80.30
Power Correction	0.93	53.88	12.35	-31.22	83.60
Wind & Power Correction	0.90	57.61	11.32	-30.68	90.38
Observed	1.00	62.81	0.00	0.00	0.00

Some additional results from wind power production bias correction for the daily wind power generation in the form of contour plots can be found in section 6.3.

In Table 8, a comparison of the correlations between the resulting monthly wind power generation data from different corrections and measured data as well as the RMSE, the means, the minimum and the maximum difference to the observed data of each of the time series are displayed. For all cases, the correlations are very high at 99%. However, as mentioned before, only looking at the correlations is not sufficient to determine how well the simulated and observed time series fit, which is why also other parameters like mean and RMSE are consulted. As can be clearly seen, the mean of the uncorrected data is significantly higher at 645.12 GWh than the mean of the measured wind power generation at 406.92 GWh. The mean of wind corrected only simulation underestimates the observed wind power generation with only 374.83 GWh monthly mean. The means of the two wind power corrected simulations are the same as the measured wind power generation, due to bias-correction. The lowest RMSE of 108.31 GWh, which indicates the difference between measured and simulated time series, occurs for the wind speed and wind power generation corrected time series. This does not fully support the observations made before, as the only wind speed corrected data seemed to fit better. For more detailed information on the results of wind power generation bias correction for the monthly wind power generation see the Appendix (section 6.2).

Table 8: Comparison of correlations, means, RMSE, minimum and maximum difference to measured data of different methods for monthly wind power generation (own representation)

Method	**Correlation with Observed Generation**	**Mean [GWh]**	**RMSE [GWh]**	**Minimum Difference to Observed Data [GWh]**	**Maximum Difference to Observed Data [GWh]**
No Correction	0.99	645.12	340.67	1.71	1153.28
Wind Correction	0.99	374.83	147.50	-704.10	111.80
Power Correction	0.99	406.92	133.15	-547.62	184.86
Wind & Power Correction	0.99	406.92	108.13	-491.61	188.93
Observed	1.00	406.92	0.00	0.00	0.00

Figure 22 shows the yearly sums of wind power generation in TWh in the North-East of Brazil. The simulated generation without correction, displayed as an orange line, is the highest. When it is corrected with the hourly and monthly wind speed correction factors, illustrated by the blue line, the simulated wind power generation is reduced and therefore its trajectory fits the observed electricity generation better and sometimes over- and sometimes underestimates the observed wind power generation. When it is also corrected with the monthly correction factors (green line), the simulated power generation differs slightly more from the trajectory of the observed electricity generation, drawn in black, except for the last year and also only overestimates the observed wind power generation (again apart from the last year). If only the wind power generation correction is applied (red line) the differences to the observed power generation are obviously larger than when applying the wind speed correction (with or without wind power generation correction afterwards). Therefore it seems recommendable to apply the wind speed bias correction. According to the correlations, which are always at 99%, the corrections bring no improvements. However, as there are not many years to be compared, the evaluation of statistical parameters is not significant, which is why it is left out for the yearly comparison.

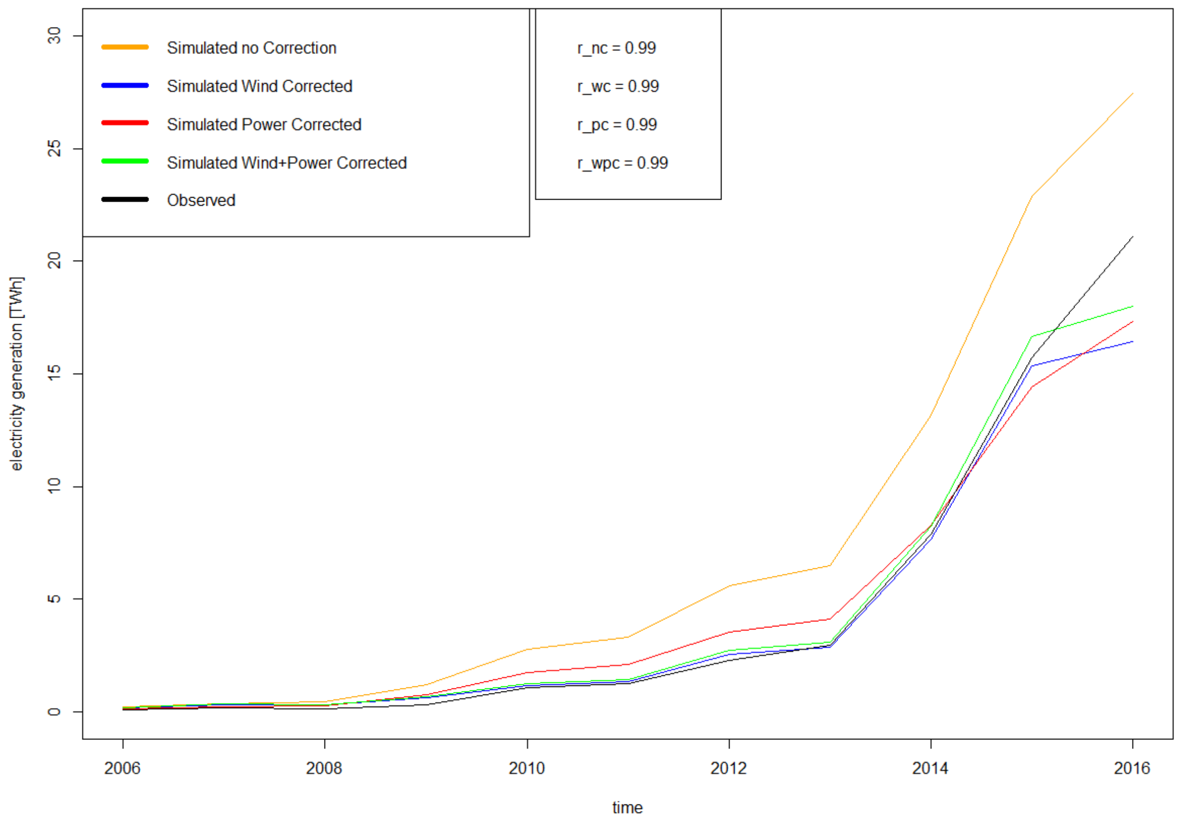


Figure 22: Yearly observed and simulated wind power generation in the North-East of Brazil compared (own depiction)

In the following, histograms, which compare the power classes of simulation to those of the observed data, will be displayed. They are only created for corrected data, as uncorrected data differ too far from observed data. The first histogram for the comparison of daily means of wind corrected power to observed power can be found in Figure 23. On the x-axis the power classes between 0 and 5 GW daily mean are shown, divided into ten classes each of 0.5 GW. The y-axis represents the frequency of each of these classes, indicating how often these power classes occur between the 1st of August 2015 and the 31st of October 2016. The highest frequency of about 120 to 130 occurrences for the simulated as well as the observed power is between 2 and 2.5 GW daily mean. The distribution of the power classes of the observed and simulated data is the same, as they lie both between 0 and 5 GW daily mean. The overlapping area is quite big, which means that the daily mean power classes of the observed electricity generation and of the wind corrected simulated electricity generation correspond well to each other, which underpins the observations made in the graph of daily power generation in Figure 18. However, what has to be emphasised, is, that the wind speed corrected wind power generation has higher frequencies in the lower power classes and the observed generation more occurrences in the higher power classes, from which one can conclude that the simulation tends to underestimate the observed power generation.

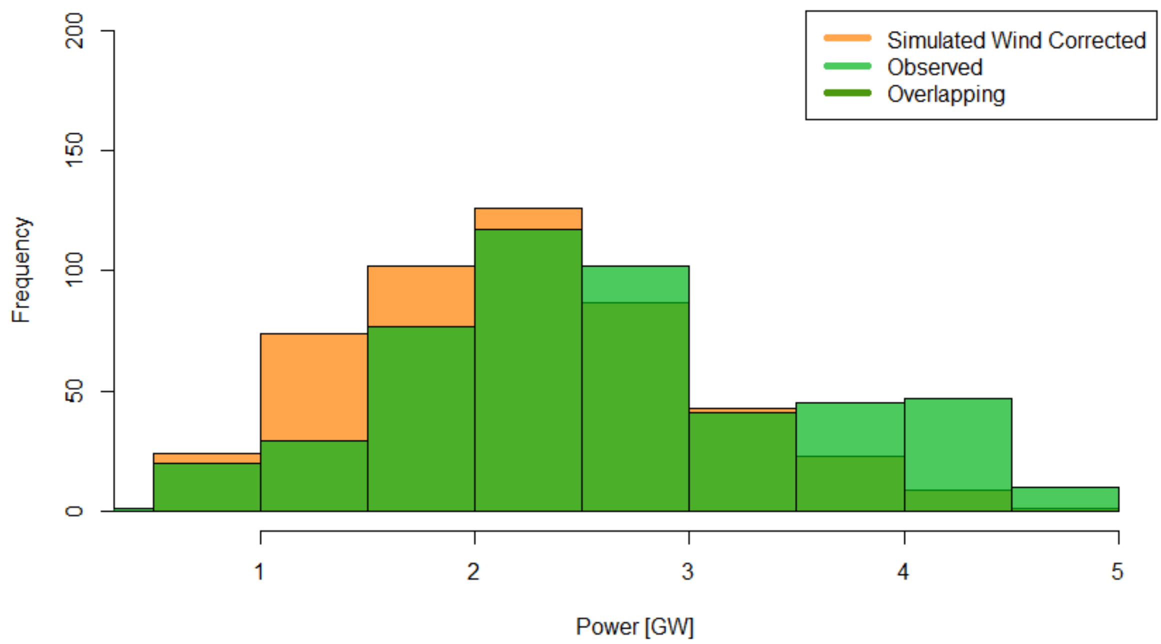


Figure 23: Histogram of mean daily simulated wind corrected and observed power (own depiction)

The following histogram (Figure 24) shows the comparison of mean daily power for the power corrected and the observed wind power generation. This graph looks very similar to the previous one, however, the simulated generation has slightly lower frequencies in the lower power classes and a few more occurrences in the higher power classes, which is hardly visible to the naked eye. Again the highest frequency is in the power class between 2 and 2.5 GW daily mean and also for lower power classes there are less occurrences for the observed data and in higher power classes there are more, compared to the simulated wind power. This again means that despite the power correction, the simulation tends to underestimate the observed wind power generation.

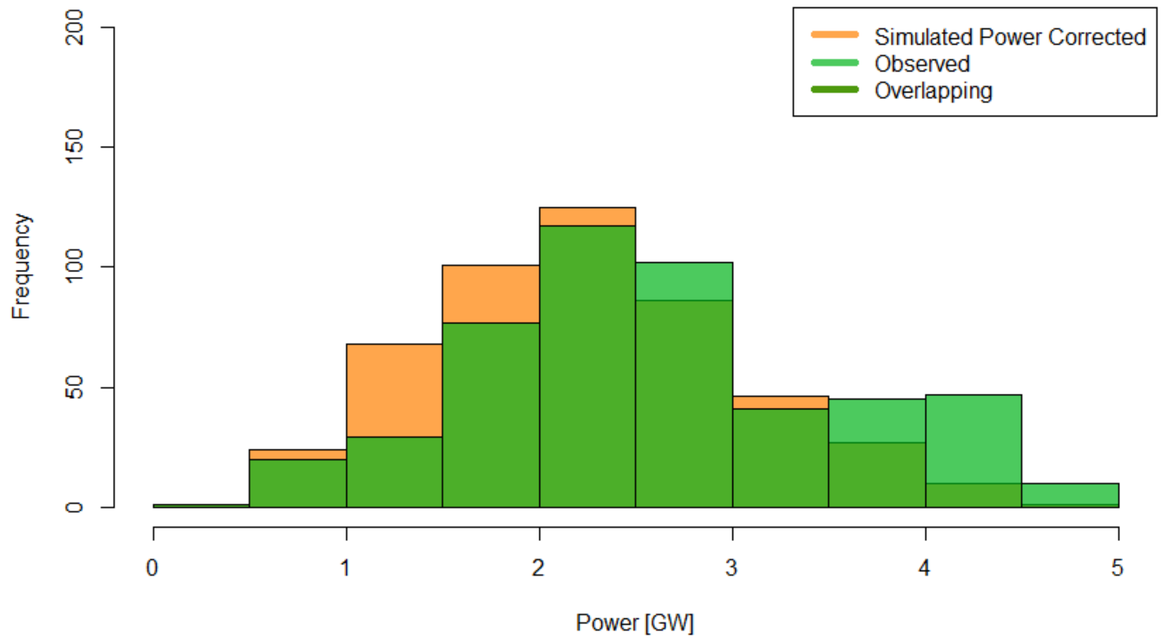


Figure 24: Histogram of mean daily simulated power corrected and observed power (own depiction)

Figure 25 compares the mean daily power of the observed to the wind speed and power generation corrected electricity generation. In this case, the highest frequency of the simulation is not between 2 and 2.5 GW (like the observed mean power) but between 2.5 and 3 GW. What can also be observed is that for lower power classes the frequency of the simulated daily mean power is increased and for higher power classes it is reduced, which means that application of wind speed and wind power generation bias-correction to the simulated power fits it better to the measured power.

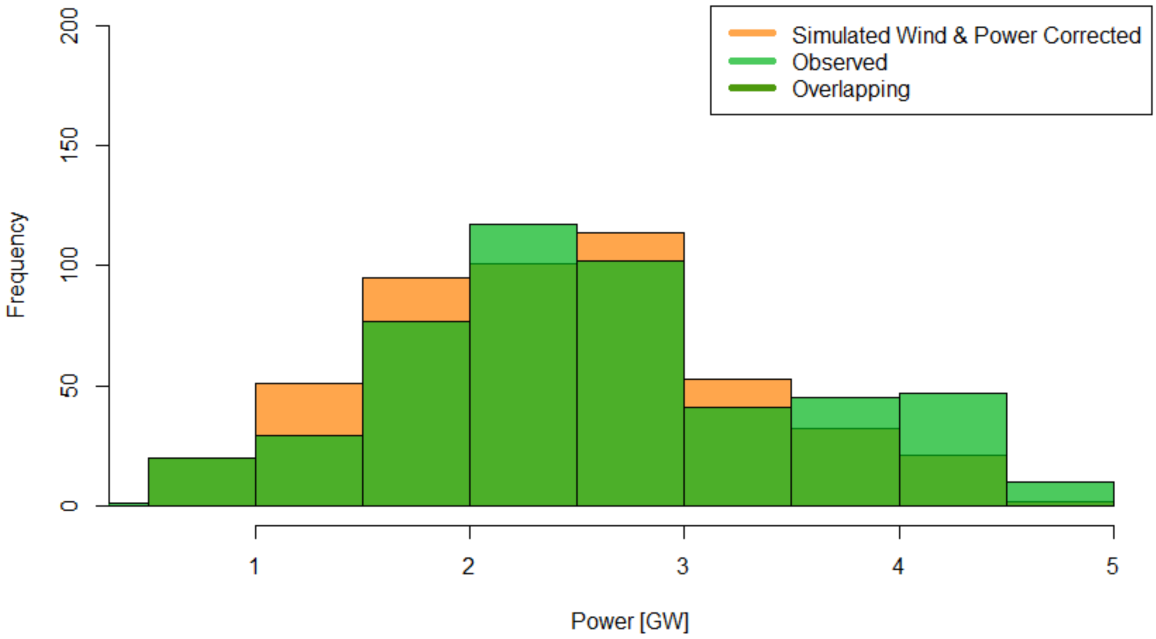


Figure 25: Histogram of mean daily simulated wind and power corrected and observed power (own depiction)

To sum up, one can say that in general the calculated wind power generation simulates the observed electricity generation quite well, but tends to underestimate the observed wind power generation for higher values of wind power generation. What can be determined from the monthly wind power generation correction factors, is, that if the wind speed correction is applied before the wind power generation correction, the simulation slightly underestimates the observed generation, which can be seen as the correction factors are all slightly above 1. The wind power generation correction factors without previous wind speed correction however, are all below 1, as before the correction the simulation overestimates the observed wind power generation. Table 9 provides a summary of the results from wind power generation bias correction. It can be clearly seen, that for the daily comparison, wind speed and wind power generation correction provides the best results, concluding from the different graphs as well as from the mean and RMSE. Only the correlation is worse for this type of correction. For the monthly comparison, from the plots the most recommendable methods seem to be wind speed correction and also the wind speed and wind power generation correction.

Correlation does not change by means of the correction methods and the mean of the simulations is the same as the one of observed data due to the method. The RMSE is lowest when applying wind speed as well as wind power generation correction. Therefore for daily as well as monthly comparisons both correction methods deliver good results.

Table 9: Summary of results from wind power generation bias correction (own representation)

	Correction	No	Wind	Power	Wind & Power
Daily Comparison	Correlation	0.93	0.88	0.93	0.90
	Mean (% Deviation)	36.67	-15.32	-13.94	-7.99
	RMSE	26.12	14.65	12.35	11.32
	Boxplot	--	+	+	++
	Boxplot Abs. Diff.	--	-	+	++
	Boxplot Difference	--	-	+	++
	Histogram	--	-	-	++
	Contour Plot	--	++	-	++
Monthly Comparison	Correlation	0.99	0.99	0.99	0.99
	Mean (% Deviation)	58.54	65.84	0.00	0.00
	RMSE	340.67	147.50	133.15	108.13
	Boxplot	--	++	-	++
	Boxplot Abs. Diff.	--	++	-	++
	Boxplot Difference	--	++	+	++
	Contour Plot	--	++	-	++

3.3 37 Years of Simulation

As the aim of the model is to be able to estimate future electricity generation potential from wind power, the past 37 years' (36 years and 10 months) reanalysis wind speed data are used to simulate wind, assuming that the following 37 years will be in similar weather conditions. The current (8.02 GW at the moment of download, in November 2016) capacities are assumed for the calculations, although they will probably increase. For the whole time of 36 years and 10 months, an overall energy output of between 783 TWh and 850 TWh was calculated, depending on the correction method used (see Figure 27). The lowest generation results from wind speed correction only, the highest from wind speed and wind power generation correction. These results yield a yearly mean of 21.26 TWh to 23.07 TWh of energy produced by wind power.

Figure 26 shows the monthly energy produced at full capacity for nearly 37 years in GWh. The maximum monthly power generation corrected as well as wind speed and wind power generation corrected energy produced is in August 2007, at 2904.61 GWh and 3076.44 GWh, respectively. The maximum power generation for the only wind speed corrected wind power prodcution occurs in September 2007, with 2747.85 GWh. The minima for wind speed, wind power generation and both corrections are at 596.02 GWh, 625.47 GWh and 661.69 GWh, respectively, all in April 2009. Clear seasonal fluctuations can be observed in the graph, with peaks in late summer or autumn and valleys in spring months.

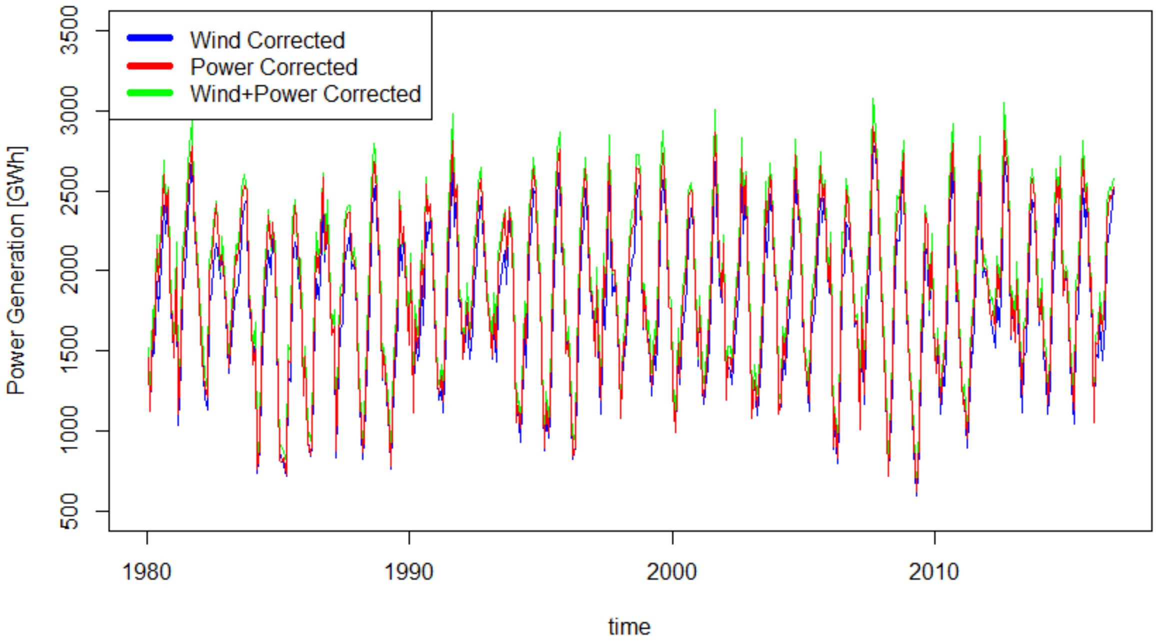


Figure 26: Simulation monthly power generation for 36 years and ten months (own depiction)

The following Figure 27 shows the yearly electricity generation of the same simulation in TWh. There are clear yearly fluctuations visible between the single years. The last year 2016 was removed from the graph, as two months are missing and the generation is very low due to this. Other minima occur in the years 1985, with yearly wind power generations between 17.34 TWh and 18.85 TWh, and 2009, with yearly wind power generations between 17.78 TWh and 19.23 TWh. The years with the highest wind speeds and therefore electricity generations are 1981 and 2012, where the yearly generation from wind power amounts to between 22.09 TWh and 24.02 TWh or between 23.61 TWh and 25.63 TWh, respectively.

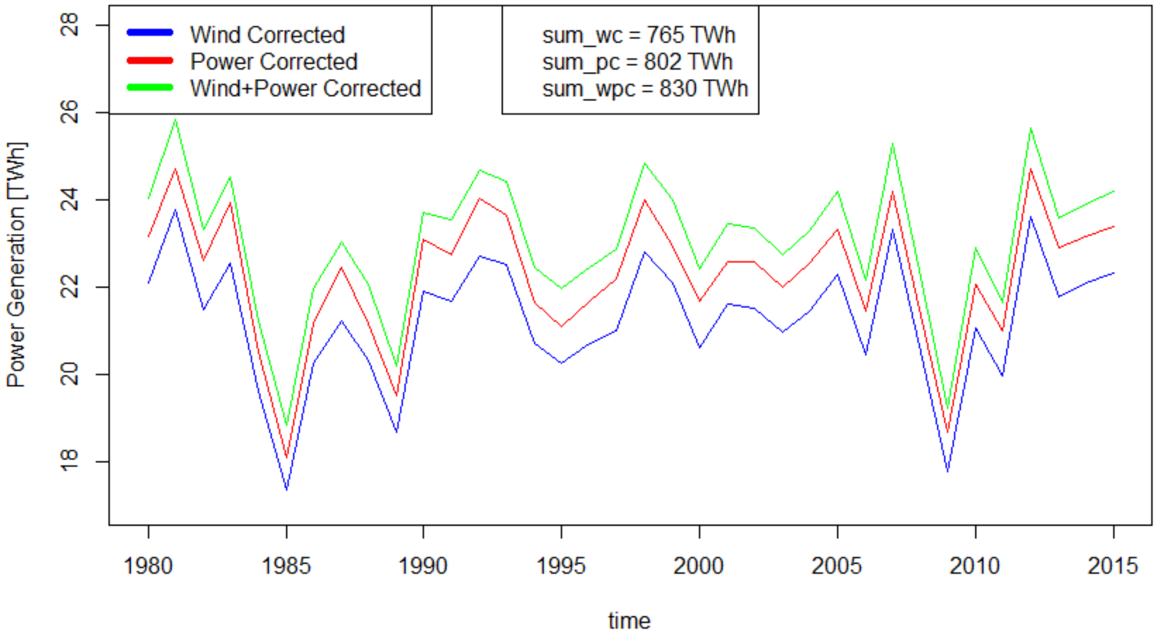


Figure 27: Simulation yearly power generation for 36 years (own depiction)

A simulation like this one can help to determine, how much of the electricity demand can be covered by wind energy, with current or future installed capacities. It can also help to decide, how much capacities need to be installed for estimated future demand, especially if other sources of electricity, like hydropower, cannot be used in the amount they have been used in the past.

4 Discussion & Conclusion

In this thesis, a model for simulating wind power generation in North-East Brazil based on publicly available wind speed and wind power generation data has been presented and validated with measured data. The model is optimised by applying hourly and monthly wind speed correction as well as monthly wind power generation correction. The results in the previous chapter show, that the model developed in this thesis is capable of simulating wind power generation in the North-East of Brazil, if at least one type of bias-correction is applied. If the data are not corrected at all, the observed wind power generation is overestimated, as can be seen in Figure 18, Figure 54 and Figure 22 as well as in the contour and boxplots, which underpins the necessity to apply at least one form of correction.

For the first correction, the wind speed bias correction, neither the distance to the nearest neighbour, the share of missing values in the INMET data nor the state, where the INMET stations are located in, had an influence on the correlation before or after correction or on the change in correlation. This means that improving any of these factors will probably not lead to better results. As could be seen, MERRA data alone do not simulate observed wind speeds very well and therefore should be corrected. The hourly and monthly correction delivers simulated wind speeds better adapted to observed wind speeds, as correlations rise with correction. Only hourly or only monthly correction cannot be recommended, as they do not result in that well adapted data, as can be seen in Table 3. In some cases, correlations are very low even after correction. This may not only be due to lack of accurateness of MERRA data, but maybe also to erroneous INMET data, as has been shown in section 2.2.1. However it should not be disregarded, that reanalysis data can contain a systematic bias, because they are also only output from a model, as emphasised by Staffell and Pfenninger [24].

In general, it has been observed, that simulated wind speeds often overestimate measured wind speeds, as the majority of the correction factors are between 0 and 1, but for some cases they are highly underestimated, with values of up to 50. In summer, wind speeds are mostly more overestimated than in winter months (see Figure 16) and around noon measured wind speeds also are more overestimated, resulting in lower correction factors (see Figure 17). This may be due to the assumption that reanalysis data often underestimate the observed variations of wind speeds, as stated by Cannon et al. [27].

It is not possible to determine exactly how well reanalysis data can simulate measured wind speeds (in this case from INMET), because, as mentioned before, either of them can be erroneous for various reasons.

Correlations of monthly and daily generation time series for the whole North-East region have been examined and show good results. Daily correlations vary between 88% and 93%, monthly correlations are all at 99% for any or no correction applied. What seems to be negative at the first glance is that the wind speed correction reduces the daily correlation. This can be due to the rather short time span regarded for daily comparison and furthermore because wind power generation correction factors were calculated for a longer timespan than the one examined in the daily comparison. However, correlation alone is not a sufficient measure to understand the quality of simulated time series, which is why also other statistical parameters like the RMSE are calculated and interpreted. For daily as well as for monthly data the RMSE is smallest when applying wind speed as well as wind power generation correction. From the boxplot in Figure 19 it can be determined, that the range of simulated daily generation is closest to the one of measured generation when applying both bias correction methods. Therefore this method is recommended for further use.

What makes the analysis of wind power generation difficult is the vast expansion of wind power capacity in recent years – and the technological and locational details of new turbines are unknown. Wind power generations at higher installed capacities are always underestimated by the simulations, even after correction. This is most likely due to applying the same correction factors as in times when lower capacities were installed to periods with higher capacities installed in the recent years. An approach for elimination of this problem was tried, by normalising with the installed capacities, but was discarded, as it led to worse results. The problem when applying the same correction factors for the whole time span is, that wind power generation structure can change, when new wind power plants are installed in different regions than the older ones, as there can be different weather conditions. To eliminate this problem, more precise data are necessary: Exact location of wind parks (not only the centre of the municipality they are located in), knowledge about the wind turbines (type, capacity, amount, rotor diameter, height), more accurate data for calculation of correction factors, for example data of measured wind power generation not only for a whole subsystem but maybe per state. The lack of more accurate (or higher resolved) data is also criticised by Pfenninger and Staffell [24] [26] as well as Schmidt et al. [9].

The importance of correcting wind speeds at many locations and not for a whole region is shown with the heat maps and boxplots of hourly and monthly correction factors in sections 3.1 and 6.1. Correction factors differ significantly at various stations and can depend highly on the location of the wind measurement station, as surroundings like buildings or trees influence the measurements at 10 metres height but not reanalysis wind speeds. Therefore these measurements do not represent the wind speeds correctly and in these cases it may be a better choice to use uncorrected reanalysis data, in order not to distort them, as also discussed by Bengtsson et al. [23].

Similar to two studies about the UK [27] and Sweden [20], it was also found that, in general, wind speeds can be reproduced quite well, but sometimes for single locations simulated wind speeds differ considerably from measured wind speeds, which may be due to the rather coarse resolution of the MERRA grid points at 50 km.

The incapability of representing observed data at single locations most likely also applies to electricity generation from wind power, but as data from single wind farms were not available, neither locations or the number and type of installed wind turbines, nor wind power generation time series, it cannot be determined whether it is possible to simulate wind power generation at turbine level. Mosshammer [49], Olauson and Bergkvist [20], as well as Pfenninger and Staffell [24] underpin the conclusion, that MERRA reanalysis data cannot represent local wind speeds and wind power generation adequately.

From the simulation of 36 years and 10 months of wind power generation, a total electricity generation of between 783 TWh and 850 TWh results, depending on the correction method used, which amounts to a mean yearly generation of 21.26 TWh to 23.07 TWh. For the electricity consumption of the North-East of Brazil in 2014 (67.1TWh [50]) this would make up for between 27.9% and 30.3%. For the 2015 electricity consumption of 79.9 TWh [8], only between 26.6% and 28.9% could be covered by wind power. In their 2024 Energy Plan the Brazilian Ministry of Mines and Energy estimates the 2024 electricity demand in the North-East of Brazil to rise to 111.0 TWh [50]. Therefore, if the capacity remains the same, only between 19.2% and 20.8% of demand can be covered, according to the calculations. Therefore, if the share of electricity from wind power in the (North-East) Brazilian electricity consumption shall remain at the same high level, capacity needs to be expanded.

It can be concluded, that it makes no relevant difference which bias correction method is applied, as results are similar. If computational cost shall be saved, only one method can be applied, applying both wind speed as well as wind power generation correction is however the recommended method.

For the future, several tasks remain: It is possible to automatize the acquisition of more recent data by using the R package "feeder", which automatically downloads new data, when running the code. Moreover, the model can be applied to other areas of Brazil (especially in the South) or in the world, if appropriate data are available and scenarios of expansion of wind power in Brazil can be developed with simulated data. As other methods of interpolation did not show better results, it is not considered worthwhile to try the model with these. However, what has not been attempted in this model is smoothing the power curve, which maybe could deliver better results. Also the parameters of maximum distance to the nearest neighbour as well as the minimum correlation could be varied in future research.

In order to facilitate research in this area, the data and the model are made public. It can be recommended for the future to search for methods for improving the data underlying this model, which means the quality of reanalysis as well as measured data, and especially the availability of site specific data, in order to increase the quality of this model.

5 References

[1] United Nations Economic and Social Council, “Progress towards the Sustainable Development Goals. Report of the Secretary-General,” United Nations, 2016.

[2] World Energy Council in partnership with Oliver Wyman, “World Energy Trilemma 2016. Defining Measures to Accelerate the Energy Transition,” World Energy Council, London, 2016.

[3] UN Climate Change Newsroom, “Historic Paris Agreement on Climate Change. 195 Nations Set Path to Keep Temperature Rise Well Below 2 Degrees Celsius,” UN Climate Change Newsroom, 12 December 2015. [Online]. Available: http://newsroom.unfccc.int/unfccc-newsroom/finale-cop21/. [Accessed 25 October 2016].

[4] L. M. H. Hall and A. R. Buckley, “A review of energy systems models in the UK: Prevalent usage and categorisation,” *Applied Energy,* pp. 607-628, May 2016.

[5] J. Schmidt, R. Cancella and A. O. J. Perieira, “The effect of windpower on long-term variability of combined hydro-wind resources: The case of Brazil,” *Renewable and Sustainable Energy Reviews,* pp. 131-141, March 2016.

[6] Ministério de Minas e Energia; Empresa de Pesquisa Energética;, “Plano Decenal de Expansão de Energia 2022,” Ministério de Minas e Energia; Empresa de Pesquisa Energética, Brasília, 2013.

[7] J. Schmidt, R. Cancella and A. O. J. Pereira, “The role of wind power and solar PV in reducing risks in the Brazilian hydro-thermal power system,” *Energy,* pp. 1748-1757, November 2016.

[8] Ministry of Economic Affairs, “Market Study: PV Energy in Brazil,” Rijksdienst voor Ondernemend Nederland, TRANSFER Latin Business Company, Den Haag, Netherlands, 2015.

[9] J. Schmidt, R. Cancella and A. O. J. Pereira, “An optimal mix of solar PV, wind and hydro power for a low-carbon electricity supply in Brazil,” *Energy,* pp. 137-147, January 2016.

[10] IRENA-GWEC, “30 Years of Policies for Wind Energy,” 2013. [Online]. Available: http://www.irena.org/DocumentDownloads/Publications/GWEC_Brazil.pdf. [Accessed 10 11 2016].

[11] I. Peters, “Der Belo Monte Staudamm: Paradebeispiel für eine erfolgreiche Zivilgesellschaft in Brasilien?,” GIGA German Institute of Global and Area Studies - Leibniz-Institut für Globale und Regionale Studien, Institut für Lateinamerika-Studien, Hamburg, 2011.

[12] ANEEL, Atlas de Energia Elétrica do Brasil. 2a Edição, Brasília: CRESESB, 2005.

[13] ONS, “Carga de Demanda,” 2017. [Online]. Available: http://www.ons.org.br/historico/carga_propria_de_demanda_out.aspx#. [Accessed 18 07 2017].

[14] ONS, “Geração de Energia,” 2017. [Online]. Available: http://www.ons.org.br/historico/geracao_energia_out.aspx?area=. [Accessed 18 07 2017].

[15] S. C. Guimarães, V. A. C. Ildeu, A. A. Paula and K. D. Rodrigues, “Estudo de Viabilidade Econômica da Construção de um Parque Eólico na Região do Triângulo Mineiro e Alto Parnaíba,” Universidade Federal de Uberlândia, Faculdade de Engenharia Elétrica, Uberlândia, 2014.

[16] S. Höltinger, J. Schmidt and E. Weterlund, “Using long term synthetic time series to assess the impact of meteorological extreme events on renewable energy systems: a case study of wind and hydro power in Sweden,” *Geophysical Research Abstracts,* 2017.

[17] R. C. Pietzcker, F. Ueckerdt, S. Carrara, H. S. de Boer, J. Després, S. Fujimori, N. Johnson, A. Kitous, Y. Scholz, P. Sullivan and G. Luderer, “System integration of wind and solar power in integrated assessment models: A cross-model evaluation of new approaches,” *Energy Economics,* pp. 583-599, 2017.

[18] T. Pesch, S. Schröders, H. J. Allelein and J. F. Hake, “A new Markov-chain-related statistical approach for modelling synthetic wind power time series,” *New Journal of Physics,* p. 055001, 06 05 2015.

[19] S. Weitemeyer, D. Kleinhans, T. Vogt and C. Agert, “Integration of Renewable Energy Sources in future power systems: The role of storage,” *Renewable Energy,* pp. 14-20, 03 2015.

[20] J. Olauson and M. Bergkvist, “Modelling the Swedish wind power production using MERRA reanalysis data,” *Renewable Energy,* pp. 717-725, April 2015.

[21] Vaisala, “Power and Energy,” Vaisala, 2017. [Online]. Available: http://www.vaisala.com/en/energy/Pages/default.aspx. [Accessed 24 07 2017].

[22] A. Boilley and L. Wald, “Comparison between meteorological re-analyses from ERA-Interim and MERRA and measurements of daily solar irradiation at surface,” *Renewable Energy,* pp. 135-143, March 2015.

[23] L. Bengtsson, S. Hagemann and K. I. Hodges, “Can climate trends be calculated from reanalysis data?,” *Journal of Geophysical Research,* 16 June 2004.

[24] S. Pfenninger and I. Staffell, “Using bias-corrected reanalysis to simulate current and future wind power output,” *Energy,* pp. 1224-1239, November 2016.

[25] M. Aillerie, B. Bogno, O. Hamandjoda, D. K. Kaoga, D. Raidandi, B. Tibi and S. D. Yamigno, “Assessment of Wind Energy Potential and Cost Estimation of Wind-generated Electricity at Hilltops Surrounding the City of Maroua in Cameroon,” American Institute of Physics, Maryland, USA, 2016.

[26] S. Pfenninger and I. Staffell, “Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data,” *Energy,* pp. 1251-1265, November 2016.

[27] D. J. Cannon, D. J. Baryshaw, J. Methven, P. J. Coker and D. Lenaghan, “Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in Great Britain,” *Renewable Energy,* pp. 767-778, March 2015.

[28] L. C. Cradden, F. McDermott, L. Zubiate, C. Sweeney and M. O'Malley, “A 34-year simulation of wind generation potential for Ireland and the impact of large-scale atmospheric pressure patterns,” *Renewable Energy,* pp. 165-176, June 2017.

[29] NASA, “Modern-Era Retrospective analysis for Research and Applications, Version 2,” 24 12 2015. [Online]. Available: https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/.

[30] M. Bosilovich and R. Lucchesi, “MERRA-2: File Specification,” Global Modeling and Assimilation Office, Greenbelt, Maryland, 2016.

[31] NASA, “MDISC Data Subset,” 23 11 2010. [Online]. Available: https://disc.sci.gsfc.nasa.gov/daac-bin/FTPSubset2.pl?LOOKUPID_List=M2T1NXSLV. [Accessed 06 03 2017].

[32] Unidata, “NetCDF 4.4.1.1,” Unidata, 21 11 2016. [Online]. Available: https://www.unidata.ucar.edu/software/netcdf/docs/index.html. [Accessed 17 03 2017].

[33] Unidata, “The NetCDF User's Guide,” Unidata, 21 11 2016. [Online]. Available: https://www.unidata.ucar.edu/software/netcdf/docs/user_guide.html. [Accessed 17 03 2017].

[34] C. Lee, M. Yang and R. Aydt, “NetCDF-4 Performance Report,” The HDF Group, 2008.

[35] S. C. Guimarães, V. A. C. Ildeu, A. A. Paula and K. D. Rodrigues, “Estudo de Viabilidade Econômica da Construção de um Parque Eólico na Região do Triângulo Mineiro e Alto Parnaíba,” Universidade Federal de Uberlândia, Faculdade de Engenharia Elétrica, Uberlândia, 2014.

[36] ENERCON, “ENERCON product overview,” 6 2015. [Online]. Available: http://www.enercon.de/fileadmin/Redakteur/Medien-Portal/broschueren/pdf/ENERCON_Produkt_de_6_2015.pdf. [Accessed 28 1 2017].

[37] G. Cruz, F. Estrela, B. Junior and M. Lima, “Shapefiles do Brasil para download,” CodeGeo, 16 4 2013. [Online]. Available: http://www.codegeo.com.br/2013/04/shapefiles-do-brasil-para-download.html. [Accessed 13 12 2016].

[38] Portal Brasil, “Administração de Empresas,” 4 2004. [Online]. Available: http://www.portalbrasil.net/2004/colunas/administracao/fevereiro.htm. [Accessed 26 10 2016].

[39] ANEEL, “Capacidade de Geração do Brasil,” 9 12 2016. [Online]. Available: http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm. [Accessed 9 12 2016].

[40] A. Leite de Sá, “Brasilien Spezial. Wind Energy ATLAS of Brazil. Windenergie Förderprogramm PROEÓLICO für 1050 MW in Brasilien. Wind / Hydro Complementary Seasonal Regimes in Brazil. Erfolgreiche DEWI Windenergie-Kurse in Brasilien,” *DEWI Magazin,* pp. 63-73, August 2001.

[41] Data Design Group, Inc., “HTML Table to CSV/Excel Converter,” [Online]. Available: http://www.convertcsv.com/html-table-to-csv.htm. [Accessed 9 12 2016].

[42] SOBRATEMA, “Ceará, potência eólica do Brasil,” 20 2 2013. [Online]. Available: http://www.grandesconstrucoes.com.br/br/index.php?option=com_conteudo&task=viewMateria&id=1101. [Accessed 30 1 2017].

[43] ANEEL, “Resumo do Empreendimento de nome "Ventos do Brejo A-6",” [Online]. Available: http://www2.aneel.gov.br/aplicacoes/Empreendimento/ResumoUsina.asp?lbxUsina=30553:Ventos%20do%20Brejo%20A-6. [Accessed 30 1 2017].

[44] J. Maria, “Energia Eólica Potiguar,” 26 10 2011. [Online]. Available: http://rn-energiaeolica.blogspot.co.at/. [Accessed 30 1 2017].

[45] M. Zambrano-Bigiarini, “Package ‘hydroGOF’,” 2 February 2017. [Online]. Available: https://cran.r-project.org/web/packages/hydroGOF/hydroGOF.pdf. [Accessed 1 May 2017].

[46] J. Baumgartner, *Modellierung der aggregierten Leistungsabgabe von Windparks im Vergleich zu gemessenen Leistungswerten anhand zweier Beispielstandorte in Österreich und Neuseeland,* Vienna: Universität für Bodenkultur Wien, Department für Wirtschafts- und Sozialwissenschaften, Institut für Nachhaltige Wirtschaftsentwicklung, 2016.

[47] S. H. Khondker, R. Mashiur, H. L. Abul, R. M Abdur, R. Tanzil and R. M Mahbubur, “Cost Effective GPS-GPRS Based Object Tracking System,” in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Hong Kong, 2009.

[48] M. R. Patel, Wind and Solar Power Systems: Design, Analysis, and Operation, Boca Raton: CRC Press Taylor & Francis Group, 2006.

[49] S. Mosshammer, “Assessing the validity of MERRA reanalysis data for simulation of wind power production,” Institute for Sustainable Economic Development, University of Natural Resources and Life Sciences, Vienna, 2016.

[50] Ministry of Mines and Energy, “Electricity in the 2024 Brazilian Energy Plan (PDE 2024),” 12 February 2015. [Online]. Available: http://www.mme.gov.br/documents/10584/3642013/02+-+Electricity+in+the+2024+Brazilian+Energy+Plan+(PDF)/96be552a-4a2c-4a32-839a-f51299c911fb;version=1.1. [Accessed 23 April 2017].

[51] D. Adler, C. Gläser, O. Nenadic, J. Oehlschlägel and W. Zucchini, “Package 'ff',” 19 2 2015. [Online]. Available: https://cran.r-project.org/web/packages/ff/ff.pdf. [Accessed 11 3 2017].

[52] GEA, Global Energy Assessment - Toward a Sustainable Future, Cambridge, UK and New York, NY, USA and Laxenburg, Austria: Cambridge University Press and International Institute for Applied Systems Analysis, 2012.

[53] International Energy Agency, World Energy Outlook 2014, Paris, France: International Energy Agency, 2014.

[54] S. Simoes, M. Zeyringer, D. Mayr, T. Huld, W. Nijs and J. Schmidt, "Impact of different levels of geographical disaggregation of wind and PV electricity generation in large energy system models: A case study for Austria," *Renewable Energy,* pp. 183-198, May 2017.

[55] I. Staffell and R. Green, "How does wind farm performance decline with age?," *Renewable Energy,* pp. 775-786, June 2014.

[56] G. D'Amico, F. Petroni and F. Prattico, "Wind speed forecasting at di erent time scales: a non parametric approach," *Renewable Energy,* pp. 59-66, 15 July 2014.

[57] D. Connolly, H. Lund, B. V. Mathiesen and M. Leahy, "A review of computer tools for analysing the integration of renewable energy into various energy systems," *Applied Energy,* pp. 1059-1082, 2010.

6 Appendix

6.1 Wind Speed Correction - Results

This section gives a more detailed overview of the results of the wind speed correction. First the possible influences on the correlations between measured and simulated data are examined, then the effect of correction on simulated wind speeds is demonstrated for several stations. Furthermore the hourly and monthly correction factors of particular stations are compared.

Figure 28 shows a scatterplot of the relation between the distance of each INMET station to its nearest neighbour and the correlation between the MERRA-2 data at the location of the nearest neighbour and the data of the respective INMET station, where data since 1999 were used. The correlations range from about 0.1, with one exception, up to nearly 0.8. The distances between the INMET stations and their nearest neighbours lie between slightly above 0 up to nearly 40 km. From the plot, no clear interrelation between the distance and the correlation can be determined, which is underpinned by the very low correlation of about 5.8% between these two factors. In general, a negative correlation could be expected, as the bigger the distance of the points of comparison the lower the correlation could be. However, this relation could not be found but even a slightly positive correlation. Therefore, it cannot be said that the correlation of INMET and MERRA-2 data depends on the distance between INMET station and nearest MERRA-2 grid point.

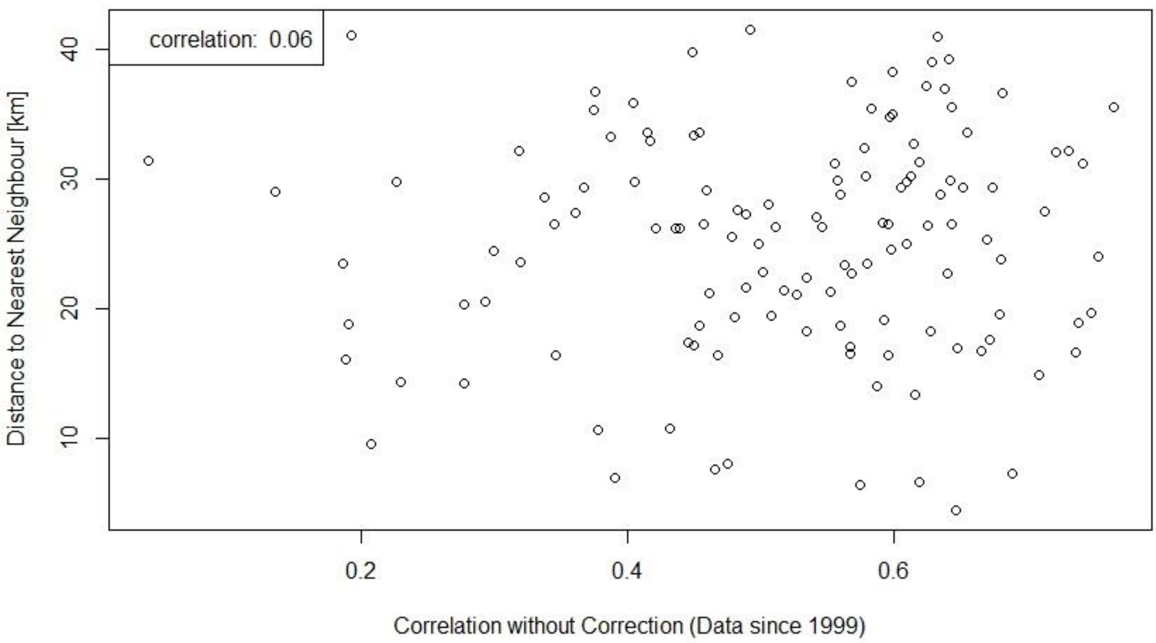


Figure 28: Relation between distance of INMET stations to their nearest neighbours and the correlation between MERRA-2 and INMET data without hourly and monthly correction (own depiction)

The next graph (Figure 29), similarly to the previous one, illustrates the relation of the distance between INMET stations to their nearest MERRA-2 neighbours and the correlation but with an hourly and monthly correction with the determined correction factors. The distances remain in the same range, but the correlation rises to a range of nearly 0.4 up to about 0.9. From the graph, still no relation of the distance between INMET stations and nearest MERRA-2 grid points and the correlation of INMET and MERRA-2 data after correction can be observed, which is supported by the low correlation of about 6.2%. It is slightly higher than before, but still too low to show a connection of the distance and the correlations after applying the hourly and monthly correction factors. Moreover, one could rather expect a negative than a positive correlation.

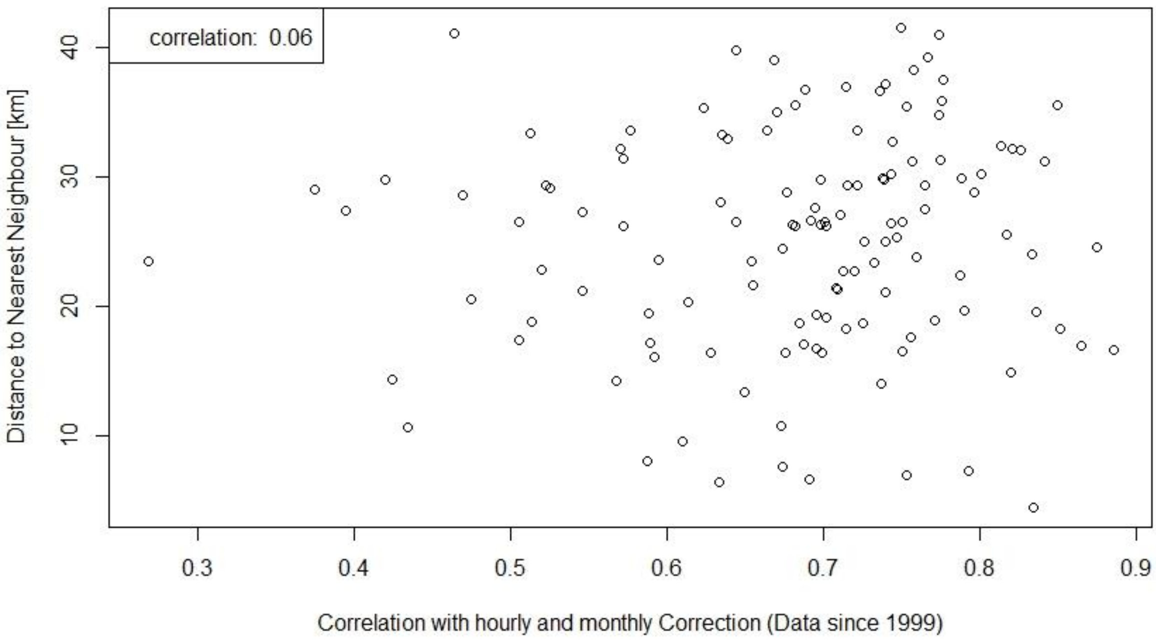


Figure 29: Relation between distance of INMET stations to their nearest neighbours and the correlation between MERRA-2 and INMET data with hourly and monthly correction (own depiction)

As another possible reason for low correlation of MERRA-2 wind speeds and INMET wind speeds the share of missing values (NAs) is assumed, also this factor and its possible influence on the correlations are examined. Figure 30 shows a scatterplot of the correlations between the INMET and the MERRA-2 data without hourly and monthly correction and of the share of NAs in the INMET datasets. The share of NAs lies between 30 and nearly 80% with an accumulation around 60%. However, a clear relation between these two factors cannot be determined, which is also underpinned by the slightly negative and low correlation of about -2.7% between the shares of NAs and the correlations between INMET and uncorrected MERRA-2 wind speeds. Therefore the hypothesis that the share of NAs has an impact on the correlation of INMET data and uncorrected MERRA-2 data is not supported by the data.

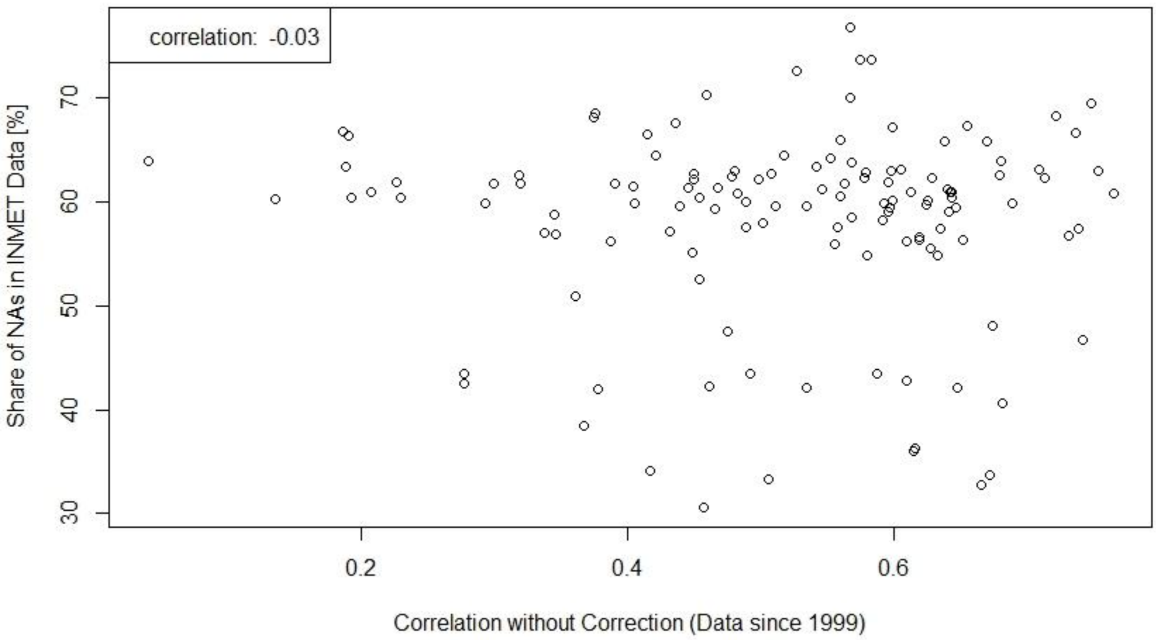


Figure 30: Relation between share of NAs in data of INMET stations and the correlation between MERRA-2 and INMET data without hourly and monthly correction (own depiction)

In Figure 31 the relation of the share of NAs in the INMET data and the correlation of INMET and MERRA-2 wind speeds with hourly and monthly correction is illustrated. An agglomeration around the point of 0.7 correlation and 60% share in NAs can be perceived, but no obvious relation between the share of NAs and the correlation between INMET and corrected MERRA-2 wind speeds. The correlation between these two factors is about 8.0%, thus higher than without correction and also positive. As this correlation still is very low, this supports the assumption, that there is no impact of the share of NAs on the correlation of INMET and MERRA-2 wind speeds after correction.

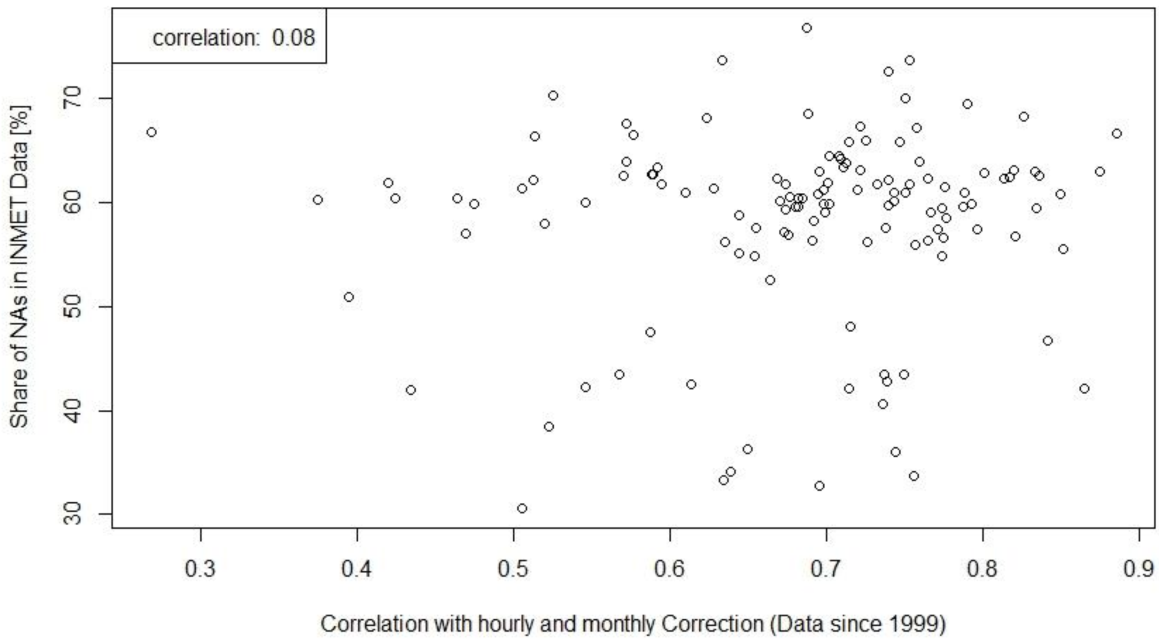


Figure 31: Relation between share of NAs in data of INMET stations and the correlation between MERRA-2 and INMET data with hourly and monthly correction (own depiction)

Furthermore, it is of interest, what might have an impact on the change in correlation of INMET and MERRA-2 when correction factors are applied. At first the relation between the distances of the INMET stations to their nearest neighbours in the MERRA-2 grid and the change in correlation is examined, which is displayed in Figure 32. The changes in correlation lie in a range of about 2 percentage points up to more than 50 percentage points. The points are widely spread and no relation between the distances between INMET stations and MERRA-2 grid points and the change in correlation can be determined. The slightly negative and low correlation of these two factors of about -1.7% underpins this observation.

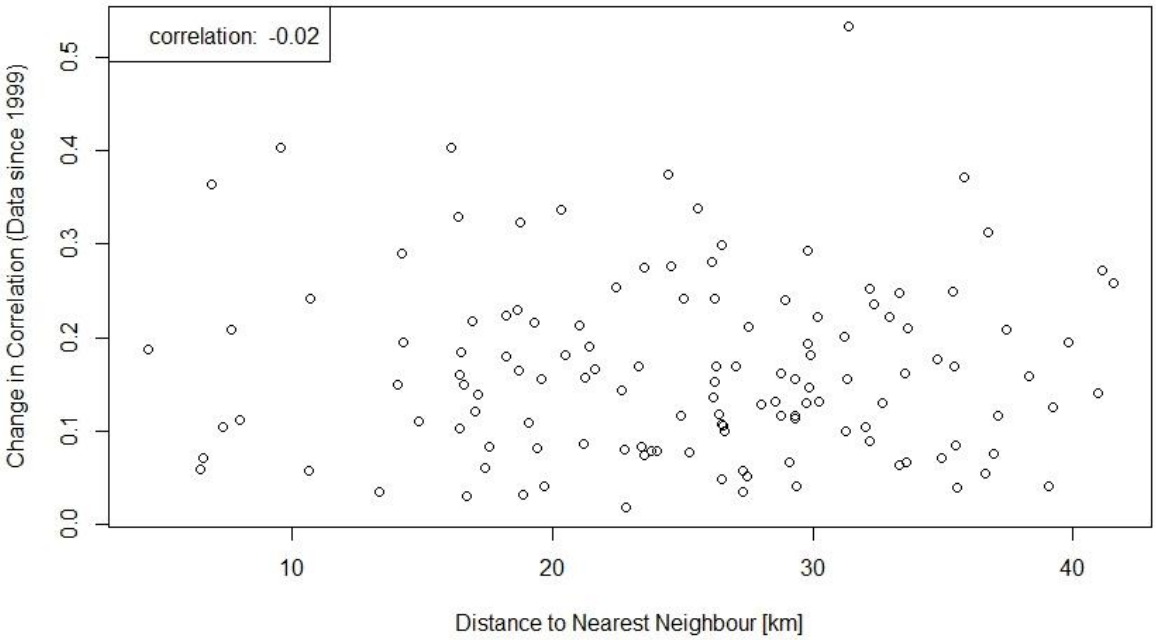


Figure 32: Relation between the distance to the nearest neighbour of INMET stations and the change in correlation between MERRA-2 and INMET data through hourly and monthly correction (own depiction)

Figure 33 shows the relation between the share of missing data in the INMET datasets and the changes in correlation when correcting the MERRA-2 data with the hourly and monthly correction factors. No relation between the two factors can be recognized. The correlation is higher than in the graphs presented before, but with about 13.9% still not very high. Tested with Steiger's test, the correlation results show no significance. So no real dependence of the distance to the nearest neighbour on the change on correlation after applying the hourly and monthly wind speed correction factors can be assumed.

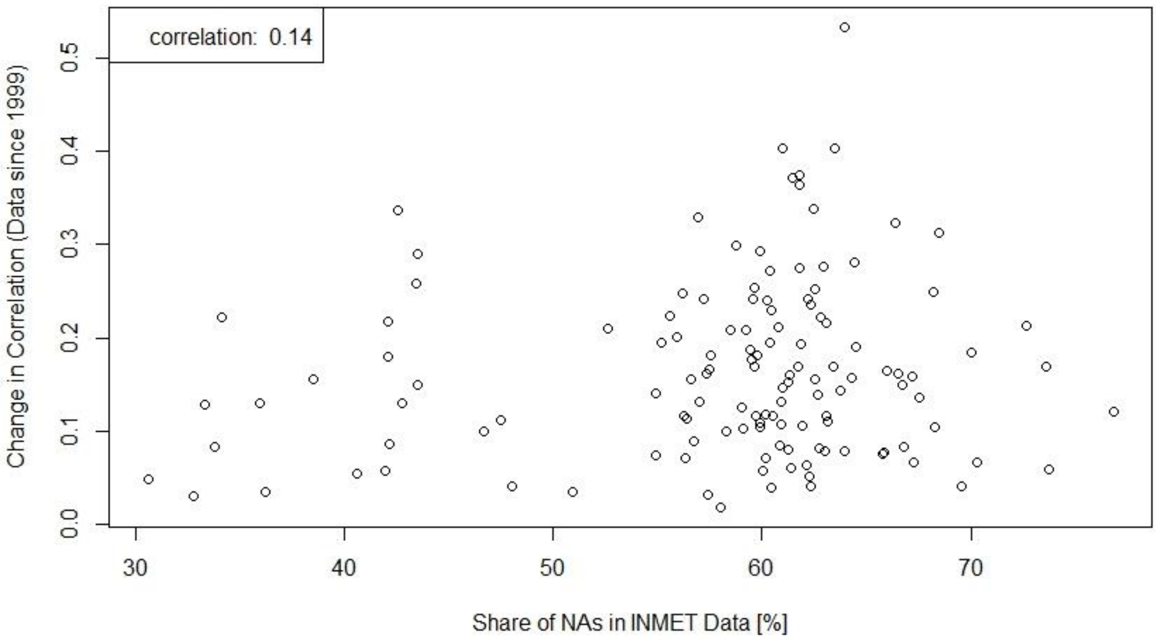


Figure 33: Relation between the share of NAs in the data of INMET Stations and the change in correlation between MERRA-2 and INMET data through hourly and monthly correction (own depiction)

Moreover, the location of the single states in the North-East of Brazil may have an influence on the correlation. Therefore the means of the calculated correlations between MERRA-2 wind speeds and INMET wind speeds per state are calculated, as displayed in Table 10. The lowest average correlation before application of the correction factors can be found in the state Maranhão, with about 0.43, the highest in Alagoas with about 0.67. After the correction the highest mean correlation remains in Alagoas with about 0.83, whereas the lowest average correlation after correction can be found not only in Maranhão but also in Rio Grande do Norte with about 0.65.

Table 10: Mean correlations before and after correction per state (own depiction)

State	Correlation before correction	Correlation after correction
Alagoas	0.67	0.83
Bahia	0.51	0.66
Ceará	0.58	0.74
Maranhão	0.43	0.65
Paraíba	0.55	0.69
Pernambuco	0.56	0.67
Piauí	0.49	0.68
Rio Grande do Norte	0.51	0.65
Sergipe	0.52	0.73

In order to find out how the application of hourly and monthly correction factors improves the correlation between the MERRA-2 and the INMET wind speed data, the boxplot in Figure 34 was created. It shows the correlation of MERRA-2 and INMET data before and after correction. It can be clearly seen, that the correlations are lower before the correction, between about 20% and less than 80% disregarding the outliers. After the correction, they lie in a range of around 50% up to 90% correlation, not considering the outliers. The box of the boxplot of corrected correlations is above the box of uncorrected correlations, which means that at least the highest 75% of the correlations between corrected MERRA-2 and INMET data are higher than the lowest 75% of the correlations between the uncorrected MERRA-2 and INMET data. This means that the correlations of reanalysis and measured wind speed data can be enhanced quite well by correcting them with hourly and monthly wind speed correction factors.

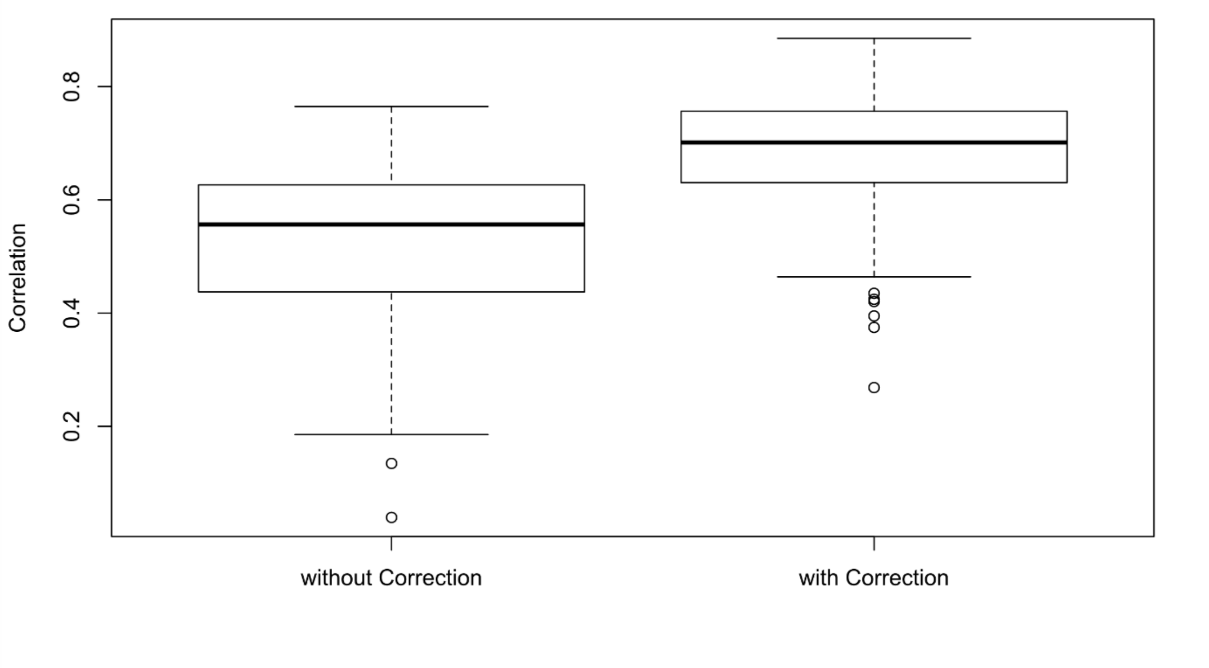


Figure 34: Boxplots of correlations before and after correction with hourly and monthly correction factors (own depiction)

Looking at the boxplot of changes in correlation (Figure 15) with hourly and monthly wind speed bias correction, it can be seen, that in half of the cases the correlation could be improved by more than 15 percentage points. The initially very low correlation of about 3.9% (INMET station 68) could even be increased to nearly 57.3%. Some data could hardly be improved (by only about 1.8 percentage points), whereas other show large improvements of up to over 40 percentage points, disregarding the outlying maximum. On average, the correlation could be improved by about 16 percentage points. However, correlation alone does not provide information whether this data actually fit to observed data, as INMET data are erroneous too. This is why if the correlation was very low, the correction was not performed as it was assumed, that the MERRA-2 reanalysis data fit better when left uncorrected.

The following figures show time series of simulated, corrected simulated and measured wind speed data for a random period of ten days to show how similar or different the shapes of the wind speed curves are. The correlation before correction (corn) and correlation between the data after the correction (corhm) is added to the figures. Only a few with extraordinary locations are chosen, which are shown in Figure 35.

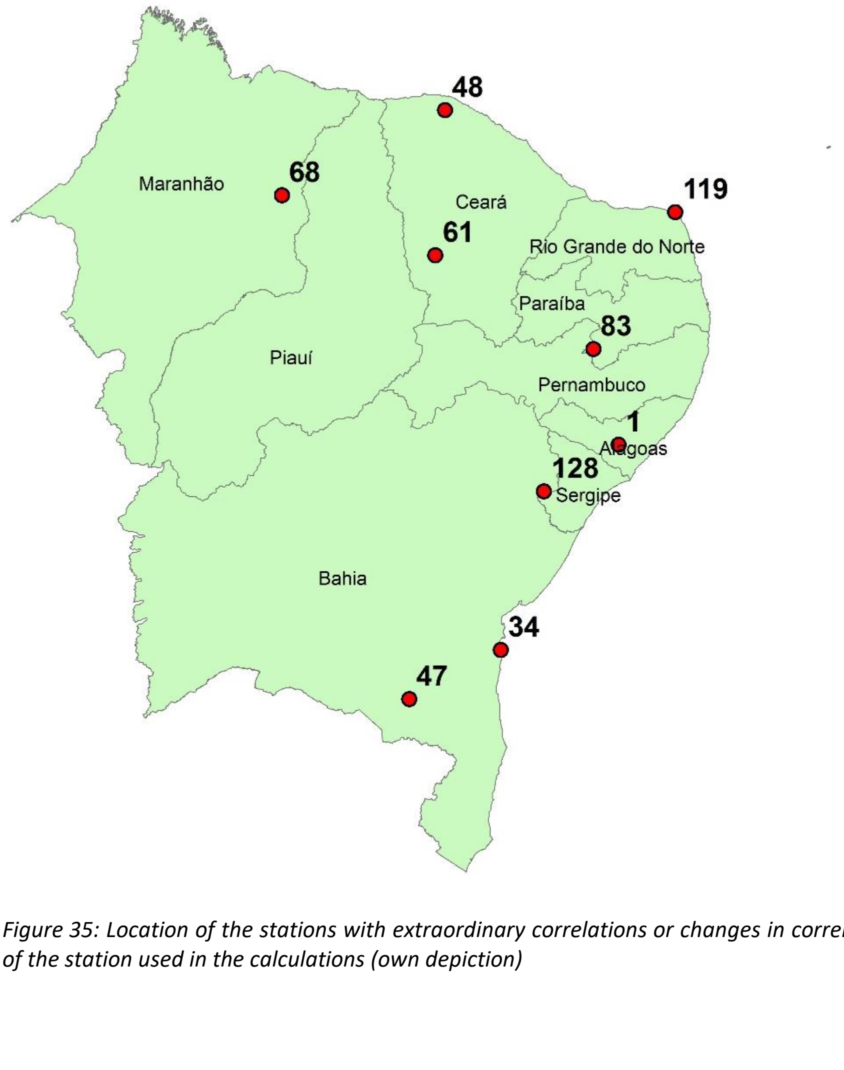


Figure 35: Location of the stations with extraordinary correlations or changes in correlations. Numbers indicate the number of the station used in the calculations (own depiction)

Figure 36 shows a comparison of time series of simulated, corrected simulated and observed wind speed data for a period of random ten days (5th of April 2013 08:00:00 until 15th of April 2013 08:00:00) for station 68, which is the INMET station with the lowest correlation with the data of the nearest neighbour in the MERRA-2 grid. Before correction, a correlation of only 0.04 (corn) was observed. However, the station experienced the highest change in correlation by 0.53, increasing the correlation to 0.57. The x-axis portrays the hours since the start date (5th of April 2013 08:00:00) and encompasses 240 hours, which corresponds to 10 days, the y-axis depicts the mean hourly wind speed in m/s. The INMET wind speed data are represented by a black line, the original MERRA-2 wind speed data of the nearest MERRA-2 grid points neighbour to the INMET station 68 by a blue line and the green line shows the corrected MERRA-2 wind speed data. The blue line of the uncorrected MERRA-2 wind speeds is mostly very flat and hardly above 0, whereas the INMET wind speeds have clear peaks and valleys that are not followed by the original MERRA-2 data, which explains the very low correlation between these data before the correction. When correcting the MERRA-2 wind speed data, the shape of the curve is significantly changed and better adapted to INMET data which is reflected in the improved correlation. Before the correction the measured wind speeds are vastly underestimated, after the hourly and monthly correction the INMET wind speeds are sometimes over- and sometimes underestimated, as they are corrected to the mean and the overall sum of wind speeds is the same for measured and simulated data. What can also be seen here, is that there are daily variations, as there are ten peaks, one for each day. This underpins the observation, that especially the hourly correction has a large effect on the correlations.

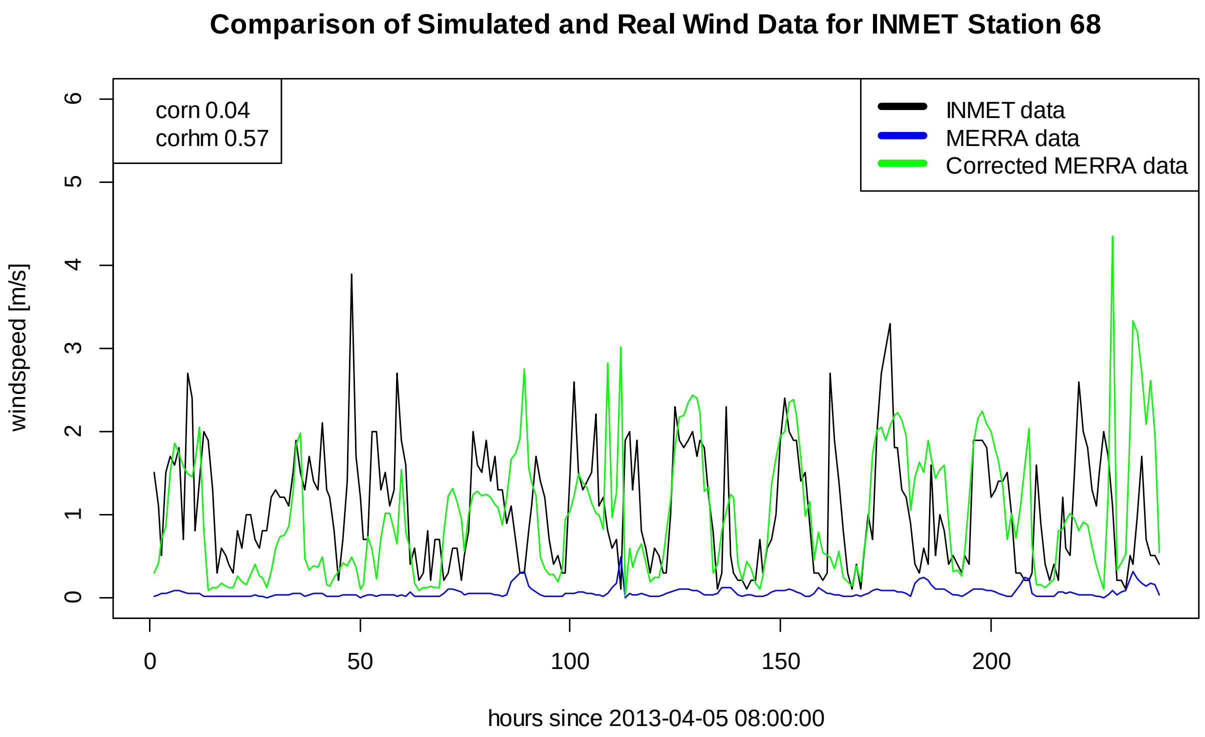


Figure 36: Comparison of a random period of ten days of INMET data with original and corrected MERRA-2 data for the station with with the lowest correlation before correction and the highest change in correlation (own depiction)

The following Figure 37 shows the mean hourly winds speeds of ten days in the same period of the INMET station 48, which is the station with the highest correlation after the hourly and monthly correction, as well as the original and corrected hourly MERRA-2 wind speeds. Here it can be observed that already before the correction the INMET and MERRA-2 data show a rather good correlation of about 73.6%, which can also be observed by similar shapes of the black and blue line. However, the original MERRA-2 wind speeds mostly overestimate the measured wind speeds. After the correction the measured and simulated wind speeds follow a similar course and the correlation is improved to nearly 90% (about 88.5%), which is a change in correlation of more than 15%, although the correlation was already quite high initially.

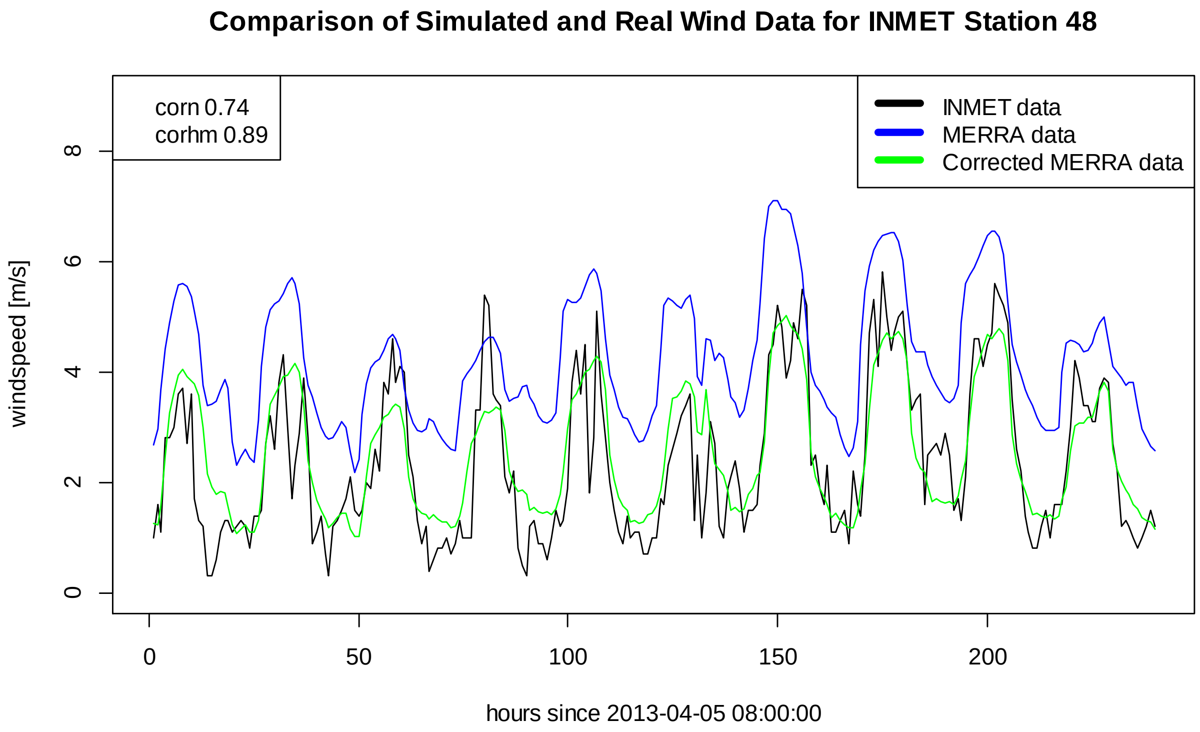


Figure 37: Comparison of a random period of ten days of INMET data with original and corrected MERRA-2 data for the station with the highest correlation after hourly and monthly correction (own depiction)

Figure 38 shows the shapes of wind speed curves for the same 240 hours as in the graphs before for the INMET station 1, which has the highest correlation of nearly 76.5% with MERRA-2 data before the hourly and monthly correction. When comparing the black line, representing the INMET wind speeds, to the blue line, representing the MERRA-2 wind speeds before correction, it can be seen, that the shapes are already very similar, even without applying the bias correction factors. As in the graph shown in Figure 37 the measured wind speeds are overestimated by the uncorrected MERRA-2 wind speed data. After the correction the wind speeds are slightly lowered, but the shape is hardly changed, which is not necessary, as the correlation is quite good before the correction and can therefore only be enhanced by around 8%.

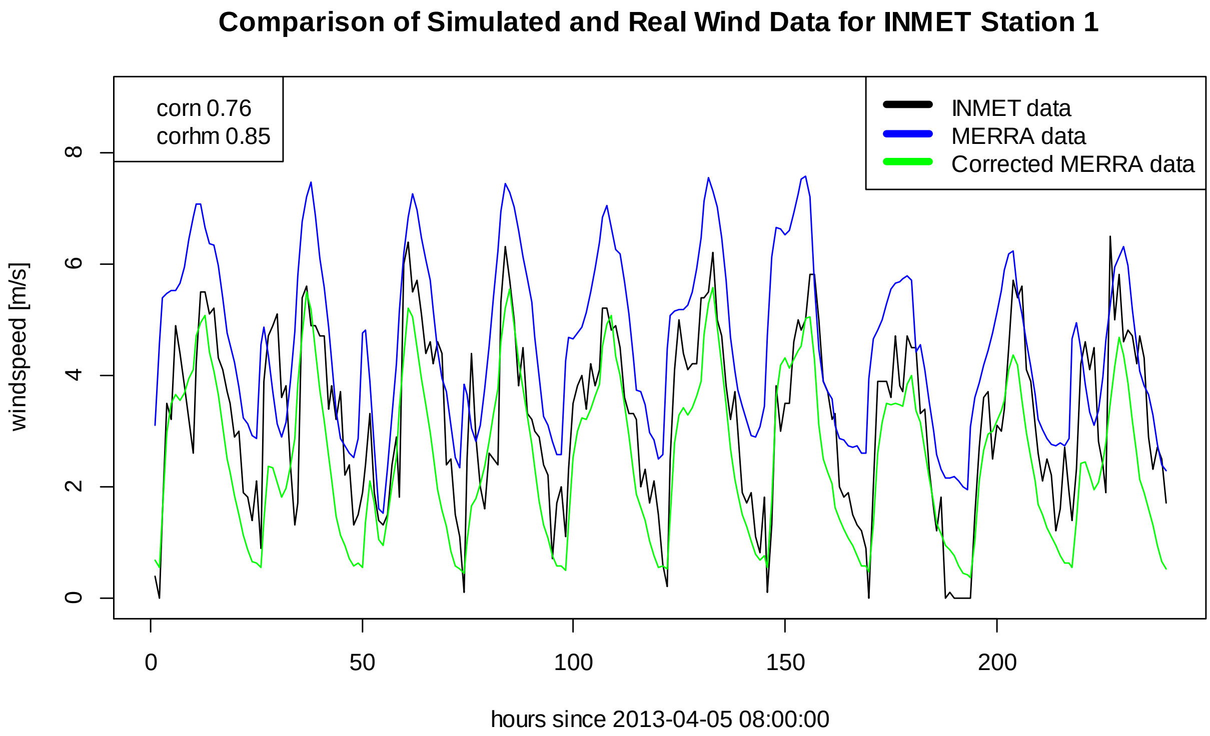


Figure 38: Comparison of a random period of ten days of INMET data with original and corrected MERRA-2 data for the station with the highest correlation before correction (own depiction)

Figure 39 shows ten days of hourly wind speeds of INMET and uncorrected and corrected MERRA-2 data for station 47, which is the station with the lowest change in correlation when applying the hourly and monthly correction factors. Before the correction, the MERRA-2 wind speed data underestimate the observed wind speed data, as the blue line is far below the black line. When correcting the data with the hourly and monthly bias correction factors, the green line, representing the corrected MERRA-2 wind speeds, seems to adapt a bit better to the black line, which represents the measured INMET wind speeds, at least in the range of 125 to 240 hours after the 5th of April 2013 08:00:00. The correlation initially was at about 50.2% and is increased to 52.0%. For this station the correction is not too effective, a reason for this might be that there are no obvious peaks in the wind speed data, as seen in the graphs before. If there were more obvious peaks and differences in wind speeds depending on the time of the day they probably could be corrected by the hourly correction, which usually is way more effective than the monthly correction.

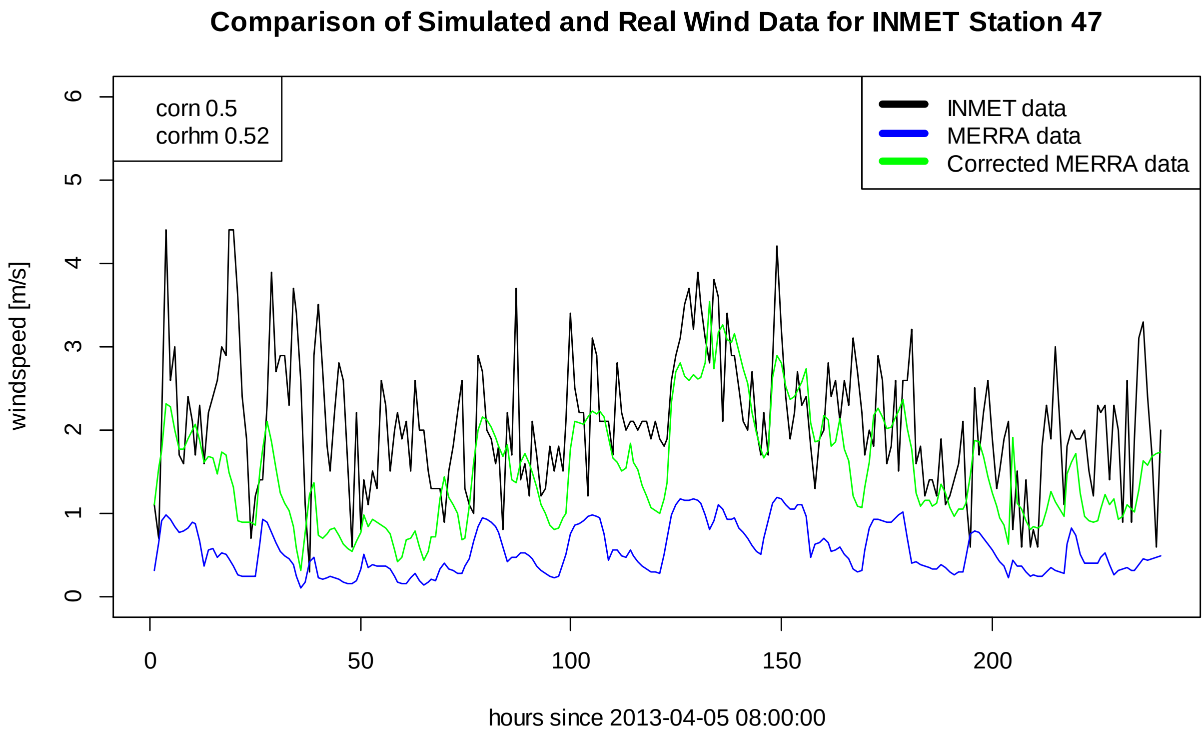


Figure 39: Comparison of a random period of ten days of INMET data with original and corrected MERRA-2 data for the station with the lowest change in correlation by hourly and monthly correction (own depiction)

Figure 40 shows an example of ten days of wind speeds of MERRA-2 data with and without correction and INMET data, with medium correlation of INMET and MERRA-2 wind speeds before applying hourly and monthly correction factors. The correlation before the correction is about 55.6% and after the correction rises to about 75.7%. Before the correction, the MERRA-2 wind speed data (blue line) already follow the measured INMET data (black line) quite well: Wherever there is a peak in the observed data, there is also a peak in the reanalysis data. Also the smaller peaks between the larger ones appear in the simulation. With correction of the simulation (green line) the height of the peaks becomes smaller, adapting them better to the observed data. Furthermore, it can be observed that the shape of the peaks changes: Before the correction the beginning of the peaks is higher than the end and after the correction the end of the peaks is higher than the beginning. This example of medium correlation before the correction, when comparing it to other graphs of data with low or high correlations, looks different, as the observed wind speeds are not significantly over- or underestimated as in the other examples.

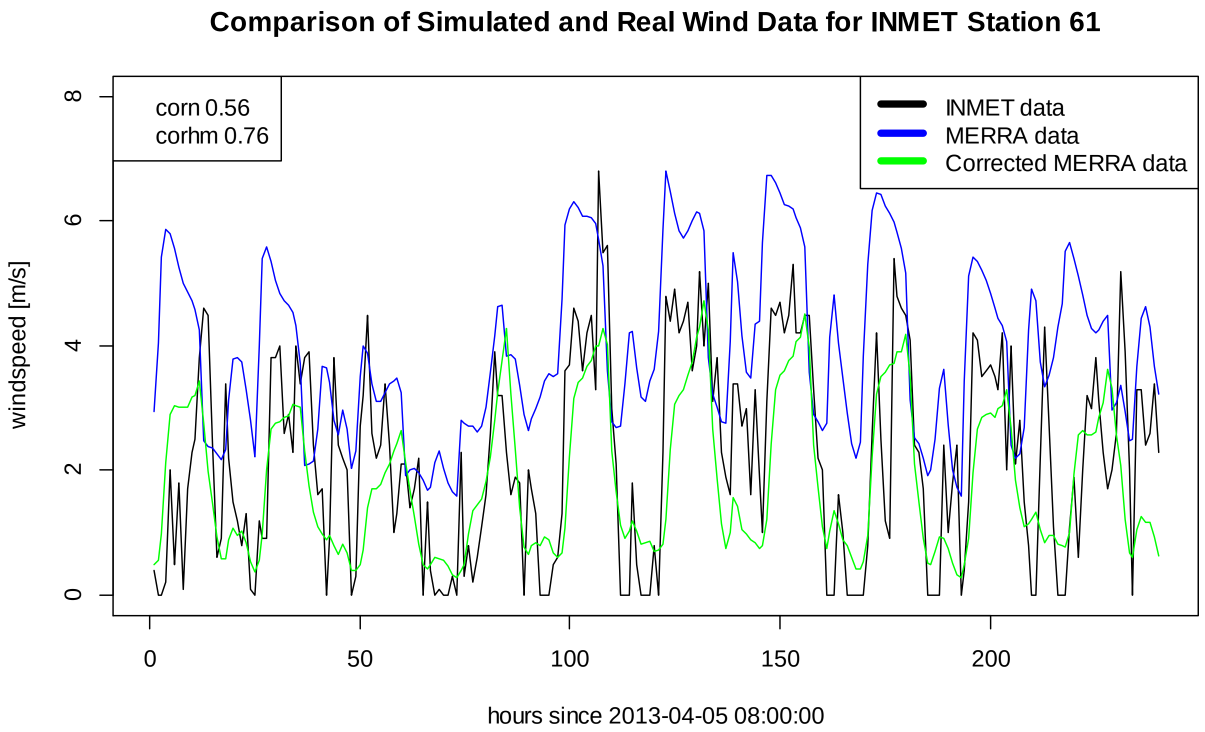


Figure 40: Comparison of a random period of ten days of INMET data with original and corrected MERRA-2 data for a station with medium correlation without hourly and monthly correction (own depiction)

Figure 41 shows a second example of a station with medium correlation (station 83) of INMET and MERRA-2 data without correction. Before the correction (blue line) the peaks that can be seen in the INMET data (black line) are not always clearly found in the uncorrected data. After the correction these peaks emerge more and at the same locations as in the INMET data. The correlation is improved from about 55.7% to about 73.8%. The correction does only have a slight influence on the height of the wind speeds, they are reduced a bit. However, the uncorrected as well as the corrected MERRA-2 wind speeds both are around the measured wind speed data and the correction mainly influences the shape of the peaks. As in Figure 40 also here unlike in the preceding graphs the observed wind speeds are not significantly under- or overestimated by the reanalysis data.

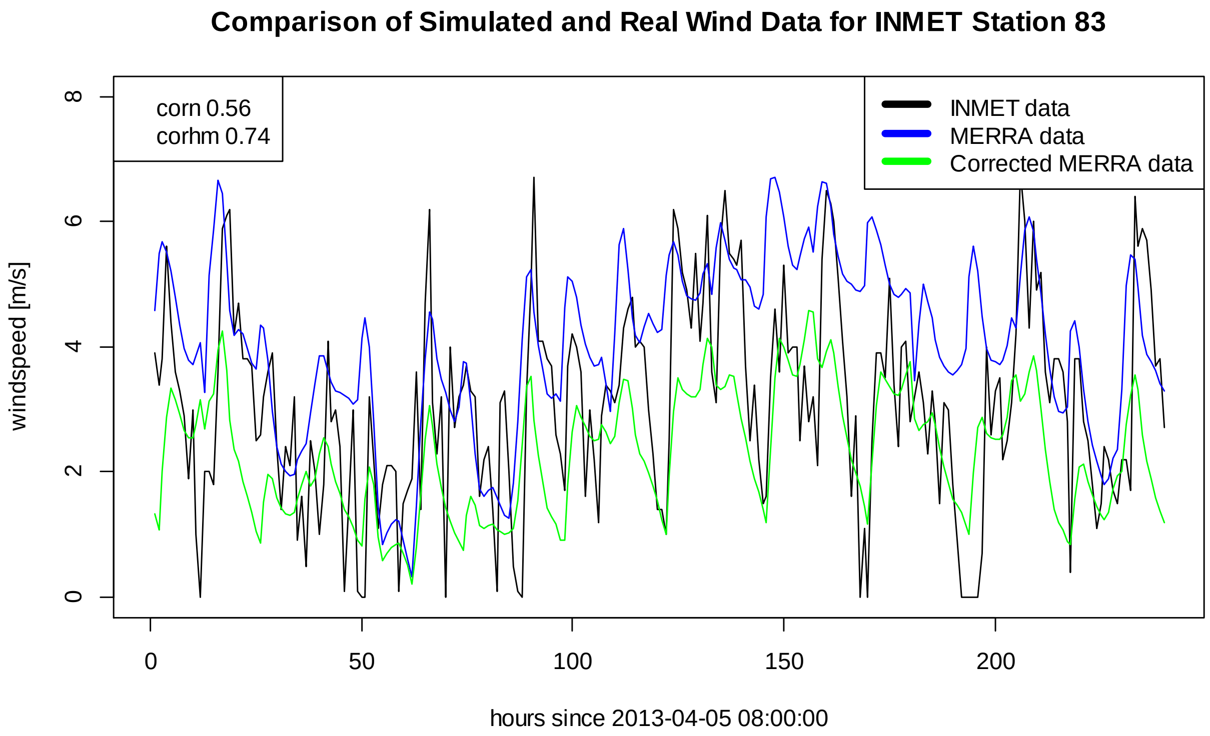


Figure 41: Comparison of a random period of ten days of INMET data with original and corrected MERRA-2 data for another station with medium correlation without hourly and monthly correction (own depiction)

The wind speeds of MERRA-2 data before and after hourly and monthly correction and measured INMET data with a medium correlation after the bias correction are compared in Figure 42. The uncorrected MERRA-2 wind speeds, represented as the blue line, do not seem to adapt to the measured wind speeds, represented as a black line, very well as sometimes the peaks of the INMET data meet the valleys of the MERRA-2 data (for example at 75, 150 or 175 hours after the 5th of April 2013 08:00:00). Therefore, the correlation of the uncorrected reanalysis wind speed data and the INMET data only show a correlation of about 42.1%. After applying the hourly and monthly correction factors the curve seems to fit better to the observed wind speed data, which is reflected in the higher correlation of about 70.2%. Moreover, the uncorrected data seem to overestimate the measured data.

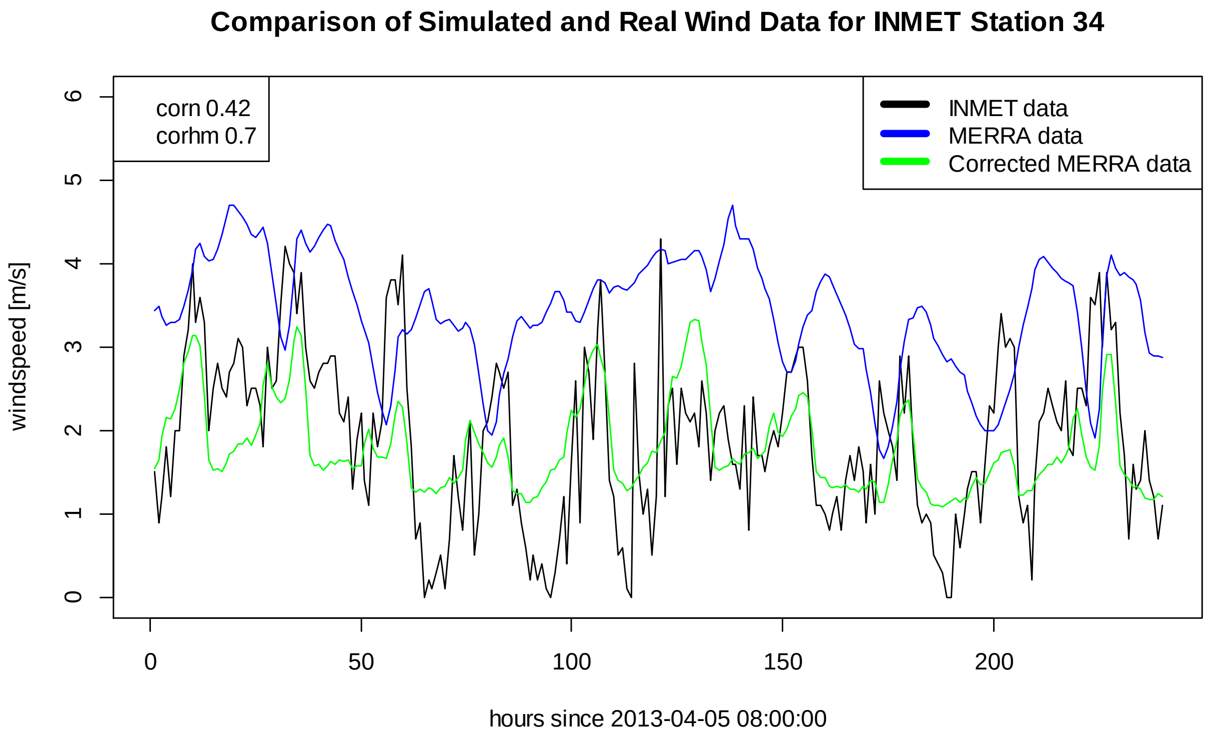


Figure 42: Comparison of a random period of ten days of INMET data with original and corrected MERRA-2 data for the station with medium correlation after hourly and monthly correction (own depiction)

Figure 43 shows another example of a ten day period of an INMET station with medium correlation between the MERRA-2 and INMET data after bias correction. When looking at the curves of wind speeds, they all seem to be in the same range, meaning that the measured wind speeds are hardly over- or underestimated. The peaks of the MERRA-2 data before the correction show a quite good correspondence with the observed wind speed data, which is reflected in the relatively high correlation of nearly 60%. However, it has to be pointed out, that sometimes there can be observed a slight shift to the left, meaning the wind speeds of the reanalysis data start to rise before they do in the INMET data. With the correction this deviation can be mostly eliminated, resulting in a higher correlation of 70.1%.

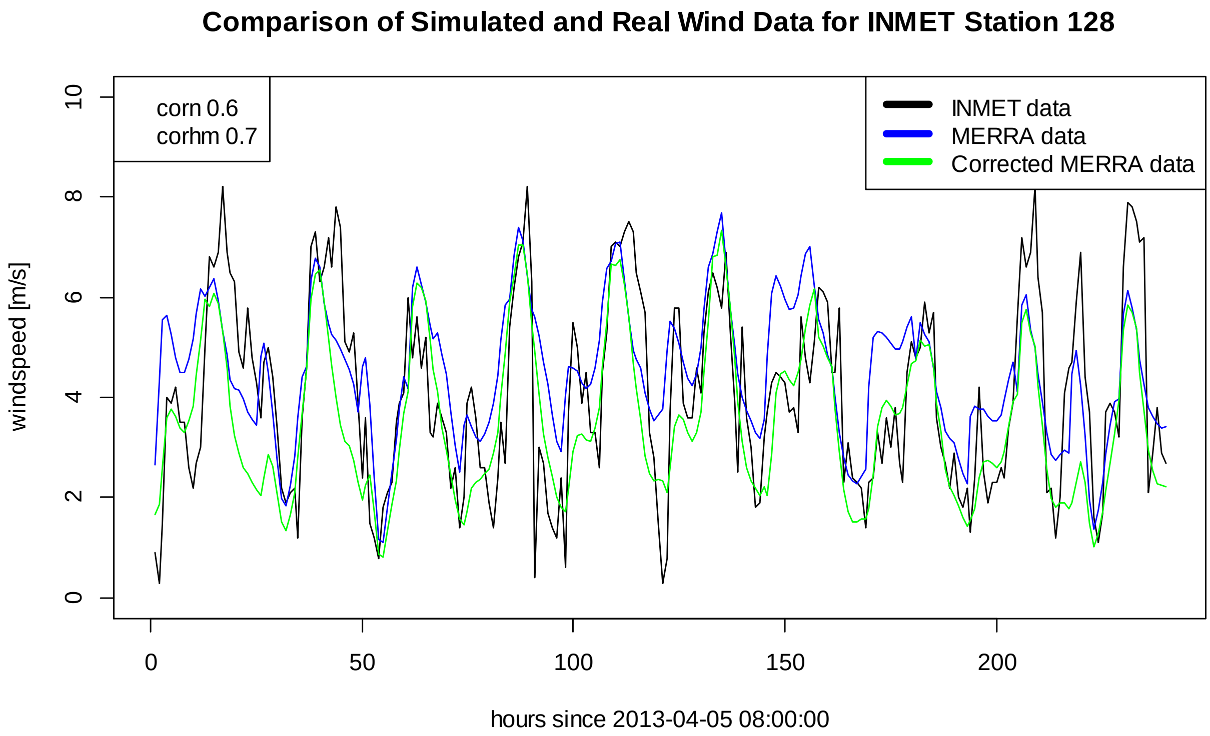


Figure 43: Comparison of a random period of ten days of INMET data with original and corrected MERRA-2 data for another station with medium correlation after hourly and monthly correction (own depiction)

An example for a very low correlation between MERRA-2 and INMET data even after the correction offers station 119. Figure 44 shows the comparison of wind speeds for uncorrected and corrected MERRA-2 and INMET data for a different period of ten days than in the other graphs, as for this station, the INMET data for this period were missing. The correlation before the correction is not very good, at about 18.6%, and can only be improved by nearly 8 percentage points to about 26.9% by means of the hourly and monthly correction. Although the peaks of reanalysis and measured data are matched relatively well, as they are always at the same positions, the shape of the curves is not so similar, as especially the corrected wind speed curve is much rounder than the curve of observed wind speeds. Moreover, what is extraordinary for this station, is that the wind speeds of the measured data fluctuate between 0 and 11 m/s for this period, meaning there are very high peaks. The original MERRA-2 data fluctuate between 6 and 8 m/s wind speed and the corrected between about 4 and 9 m/s, which may have an impact on the correlation. However, it cannot be clearly said that this is the only reason for the low correlation, as the INMET data of station 119 also have a rather large share of missing values of about 66.8%. Moreover, it is not sure whether INMET data are always correct, as in this case the fluctuations of wind speeds seem exceptionally high.

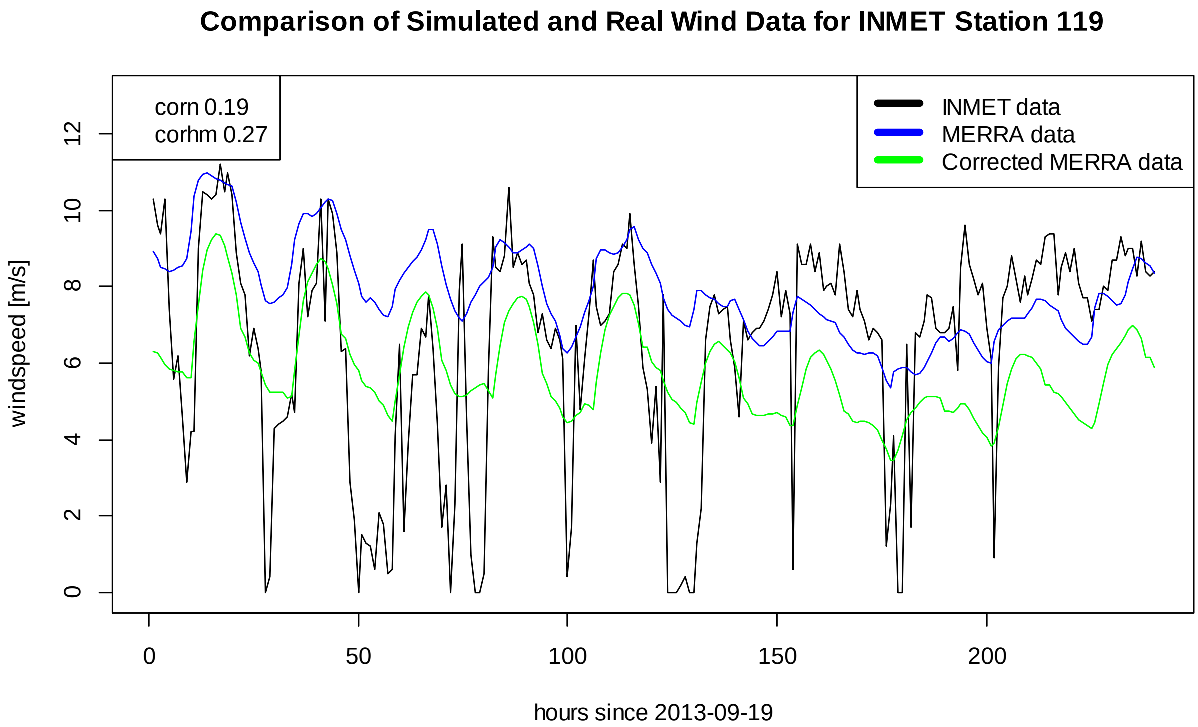


Figure 44: Comparison of a random period of ten days of INMET data with original and corrected MERRA-2 data for the station with lowest correlation after hourly and monthly correction (own depiction)

The following graphs show some examples of hourly and monthly correction factors displayed as heat maps. On the x-axis are the 24 hours and on the y-axis the 12 months, resulting in 288 correction factors, illustrated as coloured rectangles. The higher the correction factor, the more the rectangle is coloured in a red shade, the lower the correction factor, the more the rectangle is coloured in a blue shade. As the ranges of the correction factors vary considerably, one colour cannot be identified with a certain value of a correction factor, but has to be interpreted individually for each image.

The first heat map (Figure 45) shows the correction factors of INMET station 68, where the lowest correlation between INMET wind speeds and uncorrected MERRA-2 wind speeds as well as the highest change in correlation were determined. In this case, the correction factors are quite high, up to more than 50, although this only applies to one hour and month (June between 7 and 8 pm). Most of the other correction factors are below 30, but this is still high, considering the values of other correction factors which will be seen in the following. The high correction factors indicate that the observed wind speed data are highly underestimated by the reanalysis data, which could already be observed in Figure 36. As can be seen in the time series in the previous figure, the high peaks of the measured data are missing in the uncorrected MERRA-2 data, and are generated by the high correction factors especially in the evening hours. As the examined period was in April, the corresponding correction factors are rather high for this time of the day, being in darker shades of red. In the autumn and winter months (September to December) there are no high correction factors (only in December between 8 and 9 pm a slightly red colour is found). Apparently there is a high variation of wind speeds in the time of the day as it can be corrected well, when applying the correction factors, which increase the wind speeds in the evening hours. As the location of this station (latitude: -4.82, longitude: -43.34, in Caixas in the state of Maranhão) is not near the sea, these hourly variations cannot be associated with land and sea breeze.

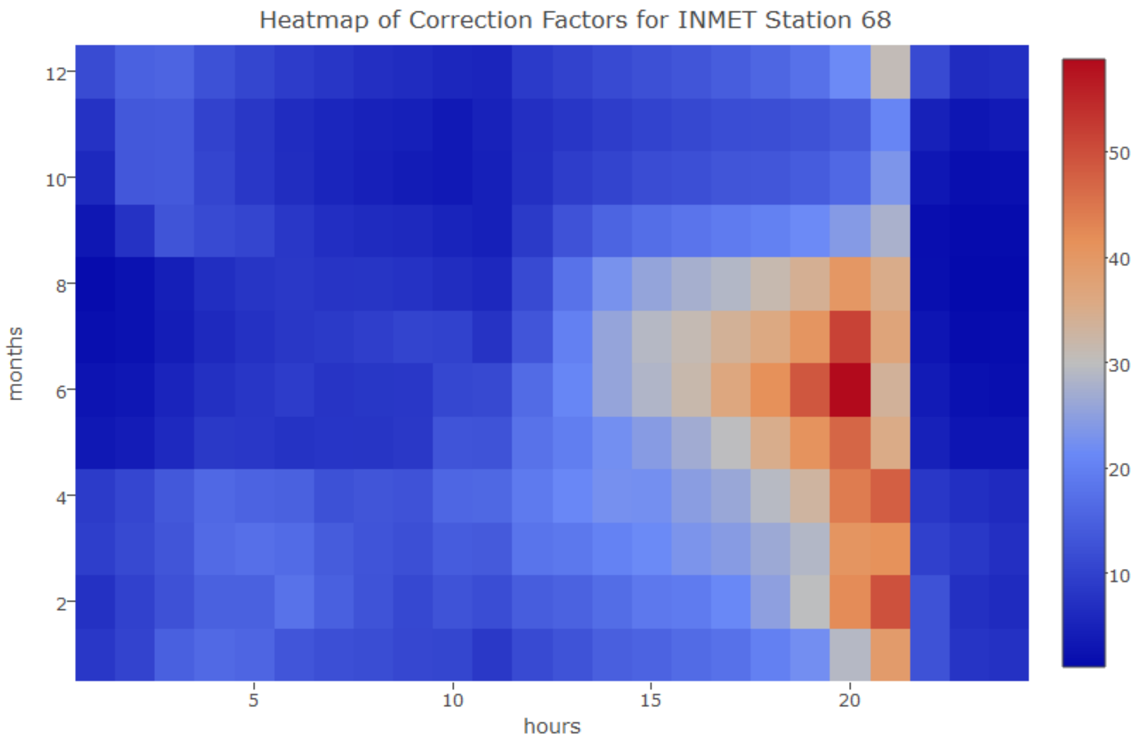


Figure 45: Heat map for for the hourly and monthly correction factors of the INMET station with with the lowest correlation before correction and the highest change in correlation (own depiction)

Figure 46 illustrates the correction factors of INMET station 48, which is the station with the highest correlation of nearly 90% after applying the hourly and monthly correction factors. In this example it can be seen very well, that in the afternoon and evening hours, the correction factors are higher than in the morning hours and slightly below 1, which means that at this time of the day the simulation better fits measured data than in morning hours. The heat map seems to be divided in a red part, with higher correction factors on the right, and a blue part, with lower correction factors on the left. Comparing this heat map to the previous one, the much lower correction factors have to be pointed out: In this case they are in a range between only 0.3 and 1. As could be seen in the graph before (Figure 37), there was only a slight overestimation of wind speeds.

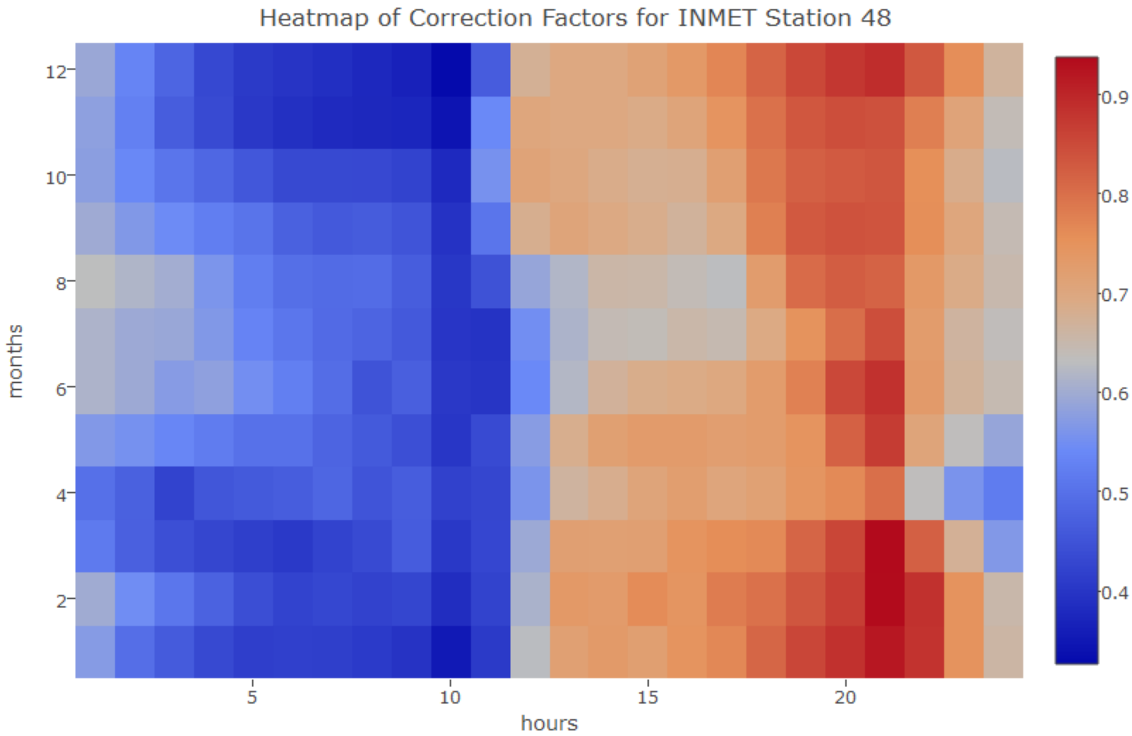


Figure 46: Heat map for hourly and monthly correction factors of the station with the highest correlation after hourly and monthly correction (own depiction)

Figure 47 shows a heat map of the correction factors of station 1, the station with the highest correlation before the application of hourly and monthly correction factors. The range of the correction factors is similar to the one in the previous figure, between 0.1 and 0.8, reducing the reanalysis wind speeds. Most of the correction factors are in the red range, meaning that the MERRA-2 wind speeds are reduced to less than half by the correction. Also here, similar to the plot in Figure 46, the lower correction factors (below 0.5, blue colours) can be found only in the morning hours (on the left side of the heat map). As the correlation was quite good before the correction (around 76.5%) the correction factors do not need to change a lot. Moreover, the observed wind speeds were only slightly overestimated (as seen in Figure 38), which results in correction factors only slightly below 1.

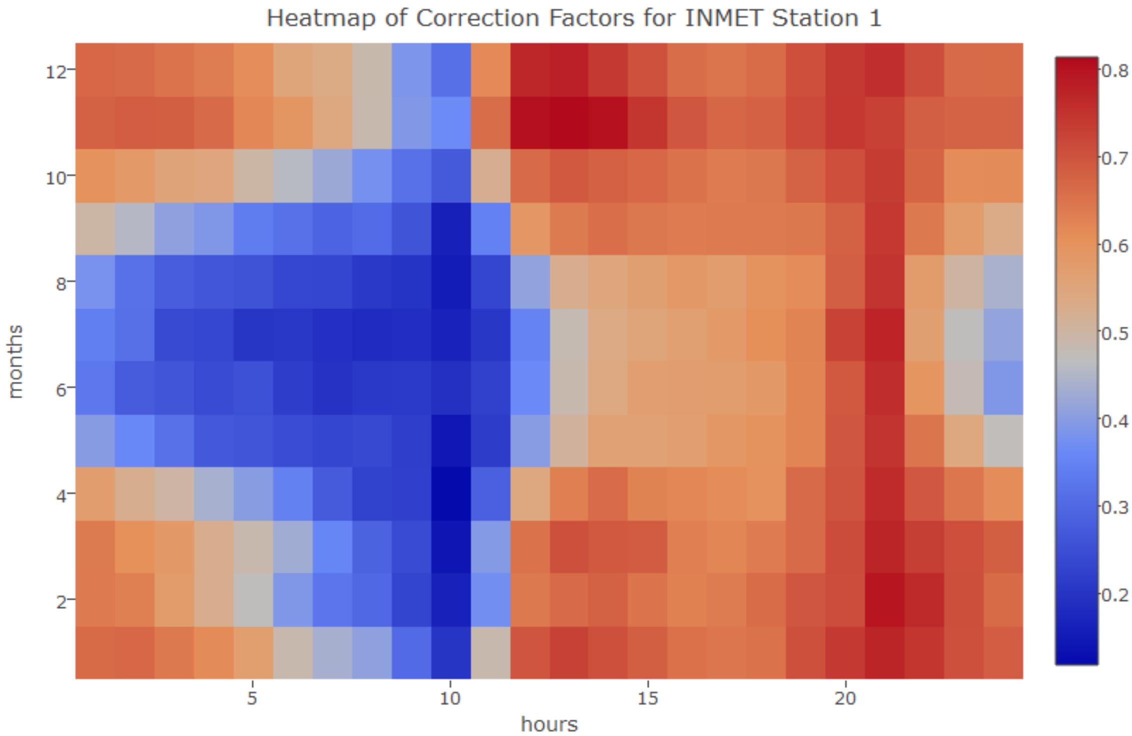


Figure 47: Heat map of the hourly and monthly correction factors of the station with the highest correlation before correction (own depiction)

Figure 48 shows a heat map for the correction factors of station 47, the station with the lowest change in correlation when applying the correction factors. No real pattern in the distribution of the correction factors can be determined, but the highest correction factors occur in December and the lowest (3 or less) between the hours of 9 am and 9 pm as well as in September and October. The correction factors are a bit higher than in the two heat maps before (Figure 46 and Figure 47), between 1.5 and 4.5. The distribution of the correction factors, which seems a bit random, may be a reason for the low improvement when applying these factors to correct the reanalysis data. Looking at the time series of this station (Figure 39), sometimes peaks are underestimated and sometimes they fit quite well. As there seems to be no real tendency in the over- or underestimation of peaks in wind speeds in the INMET data, these data can be difficult to correct properly and to improve the correlation of about 50% initially substantially.

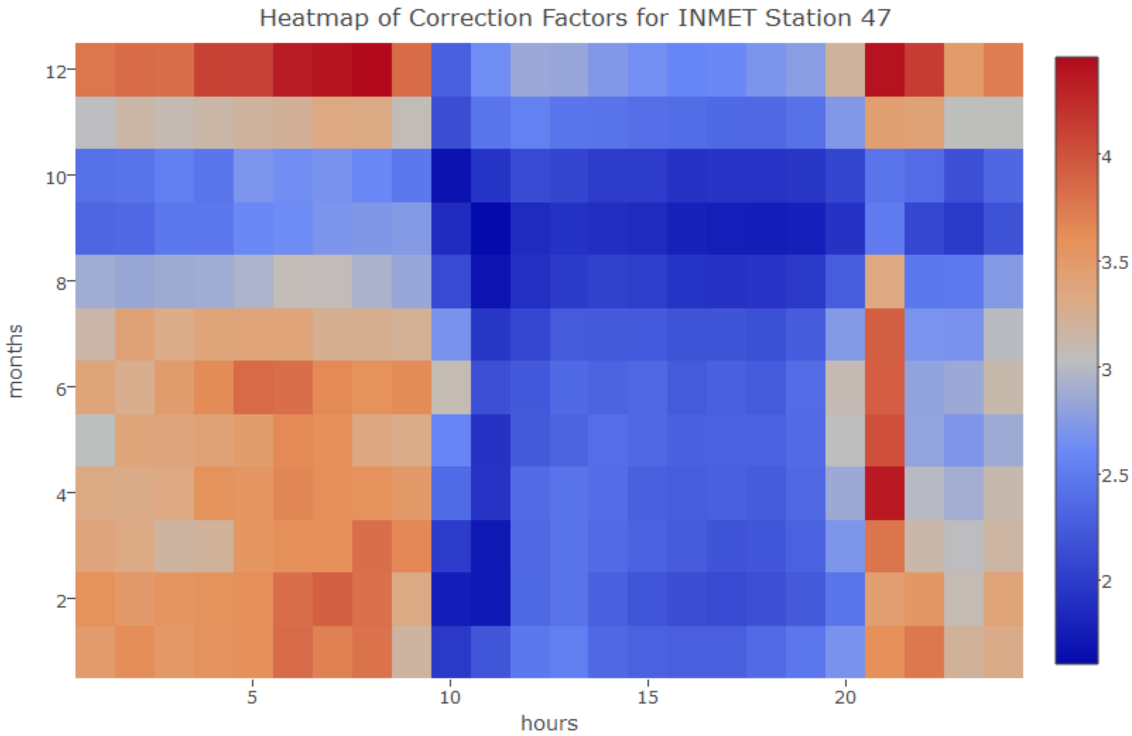


Figure 48: Heat map for the hourly and monthly correction factors of the station with the lowest change in correlation by hourly and monthly correction (own depiction)

The heat map of the correction factors for station 61, which is an example of an INMET station with a medium correlation before the hourly and monthly bias correction, can be seen in Figure 49. In this case the correction factors lie between 0.1 and 1.3. Most of the area in the heat map is coloured in blue to lilac colours, which stand for values below 1, meaning that observed wind speeds are mostly overestimated by MERRA-2 wind speeds. Only in the hours between 9 and 11 pm the correction factors reach values a bit above 1, meaning that they correct wind speeds that are slightly underestimating the measured data. This observation can also be made in Figure 40, where at the end of the peaks (which corresponds to the end of a day), the MERRA-2 wind speeds fall below the measured wind speeds and therefore need to be increased to adapt them to the observed data. As also observed in graphs before, the lowest correction factors are in the early hours of the day (on the left side), whereas the higher ones are in the afternoon and evening hours (on the right side).

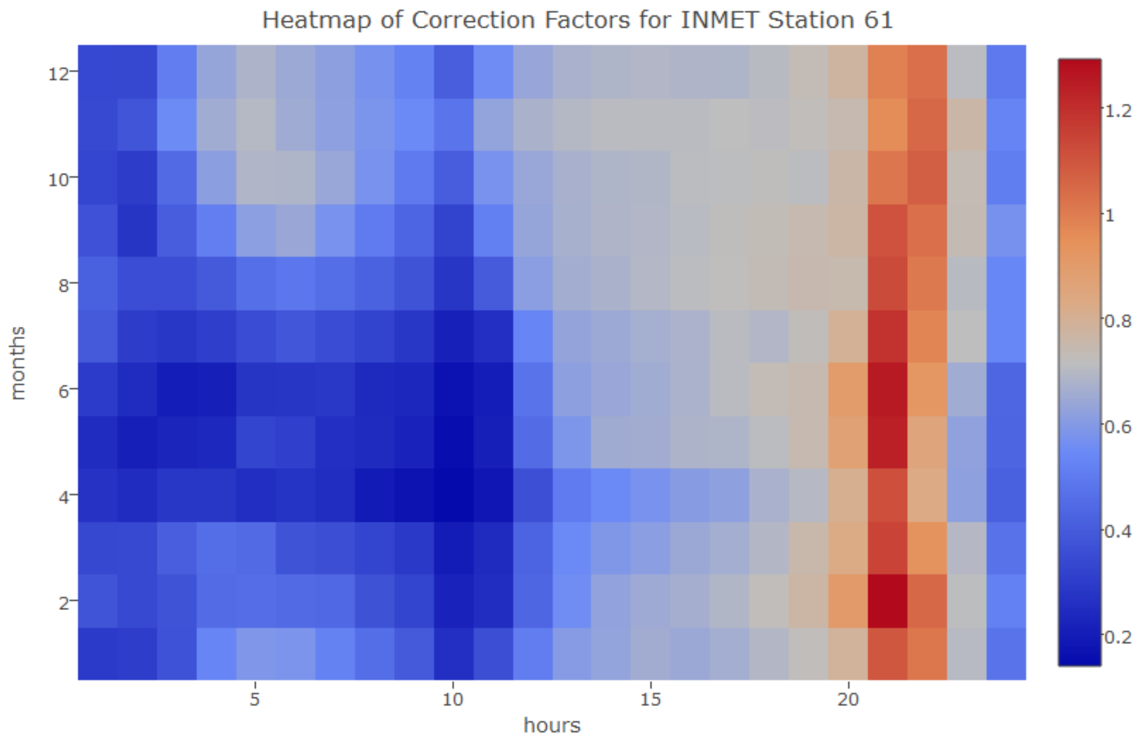


Figure 49: Heat map for the hourly and monthly correction factors of the station with medium correlation without hourly and monthly correction (own depiction)

The heat map of the correction factors for a second INMET station (station 83) with a medium correlation before the application of the correction factors can be found in Figure 50. For this station the correction factors are a bit lower than for the other station with medium correlation before the correction (Figure 49), in a range of about 0.1 to 0.9. As all the correction factors are below 1, the reanalysis wind speeds need to be reduced to adapt to the observed wind speeds, which means that without correction the MERRA-2 data overestimate the INMET data. Most of the correction factors are in a red or pink colour, meaning they are not far below 1. This means the wind speeds are only slightly overestimated, which meets with the observations made in Figure 41. This heat map looks similar to the one in Figure 47, the one of station 1 with a high correlation before the correction.

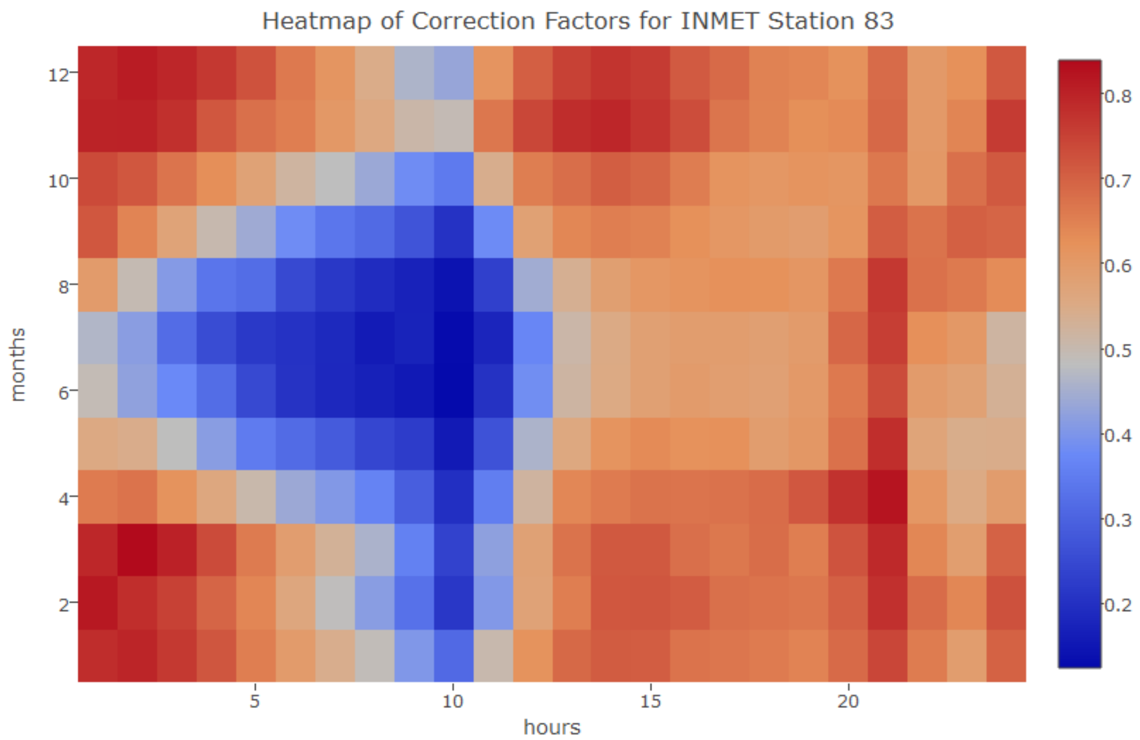


Figure 50: Heat map for the hourly and monthly correction factors of another station with medium correlation without hourly and monthly correction (own depiction)

Figure 51 shows an example of a station with a medium correlation of observed and simulated wind speeds after applying the correction factors: the heat map of station 34. The values of the correction factors range from 0.3 to more than 0.8, but are all below 1, which means that the reanalysis data overestimate the observed generation slightly. As also seen in the other examples, the highest correction factors occur in the afternoon and evening hours (between noon and 9 pm). At these times, the reanalysis data only overestimate the observed wind speeds slightly, in earlier hours they need to be corrected more, sometimes up to only a third of the original wind speed. Again, the right part of the heat map (the afternoon and evening hours), is mostly red, meaning there are higher correction factors than in the right part of the heat map (morning hours).

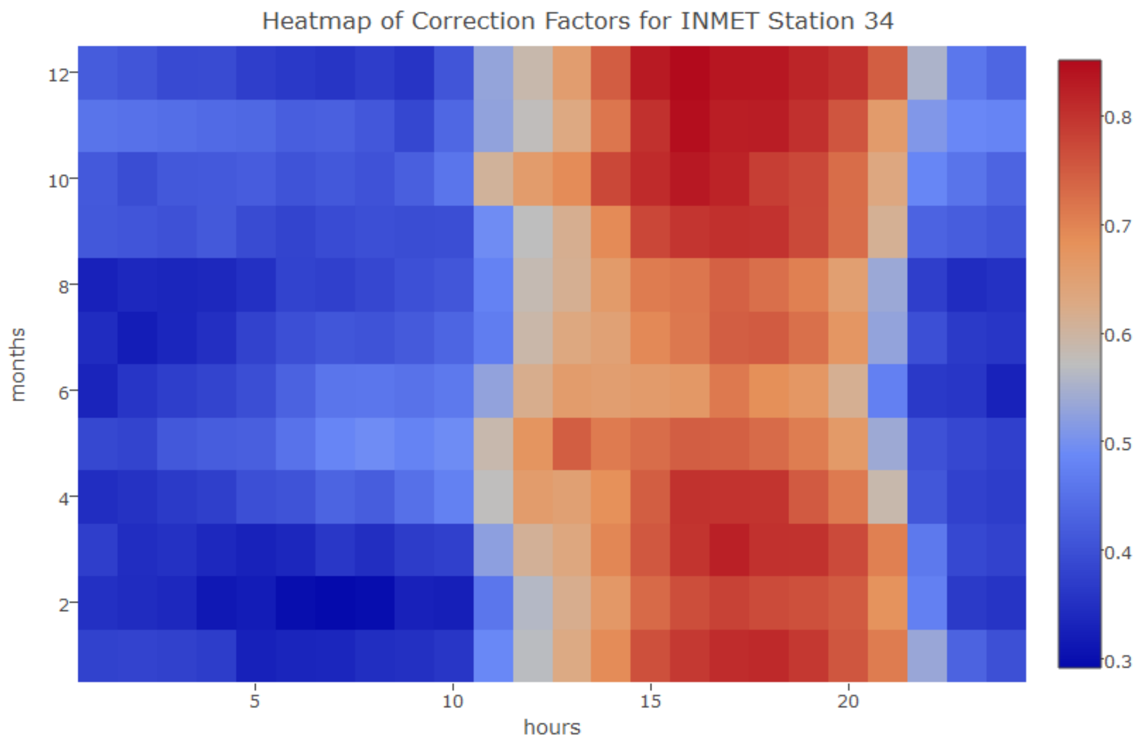


Figure 51: Heat map for the hourly and monthly correction factors of a station with medium correlation after hourly and monthly correction (own depiction)

Another example for the correction factors of a station with medium correlation after the bias correction offers Figure 52. The correction factors for this station lie between 0.3 and 1 or a little above. No clear separation between hours or months with lower or higher correction factors can be determined from the heat map. Very low correction factors occur between 5 and 10 am, the highest between 8 pm and 3 am. In general it can be said, that the observed wind speeds are overestimated, sometimes more and sometimes less, which fits the observations made in Figure 43.

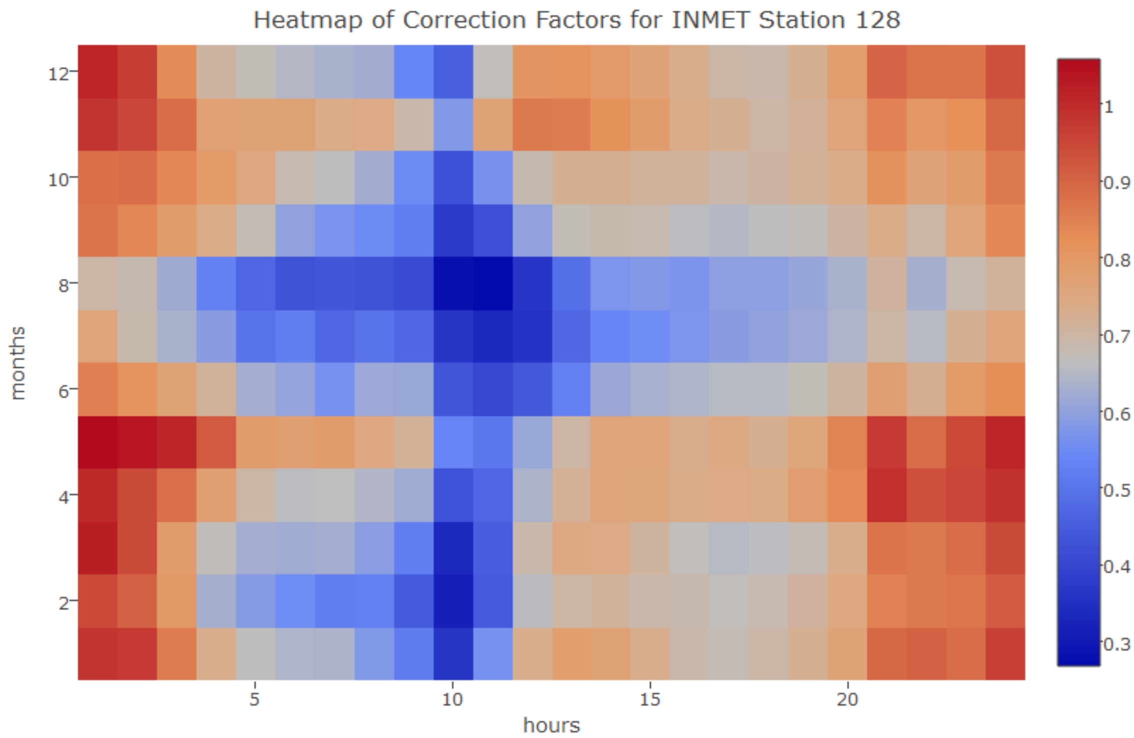


Figure 52: Heat map for the hourly and monthly correction factors of another station with medium correlation after hourly and monthly correction (own depiction)

The heat map of the hourly and monthly correction factors of station 119, the station with the lowest correlation of INMET data and corrected MERRA-2 data, can be seen in Figure 53. The correction factors all are below 1, in a rather narrow range between 0.5 and 0.95. Again, no pattern can be determined from the distribution of colours in the heat map. The highest correction factors occur in August and March. The arrangement of the correction factors with no clear pattern may be an indicator that the MERRA-2 and INMET data in general do not correlate very well and it is difficult to correct them.

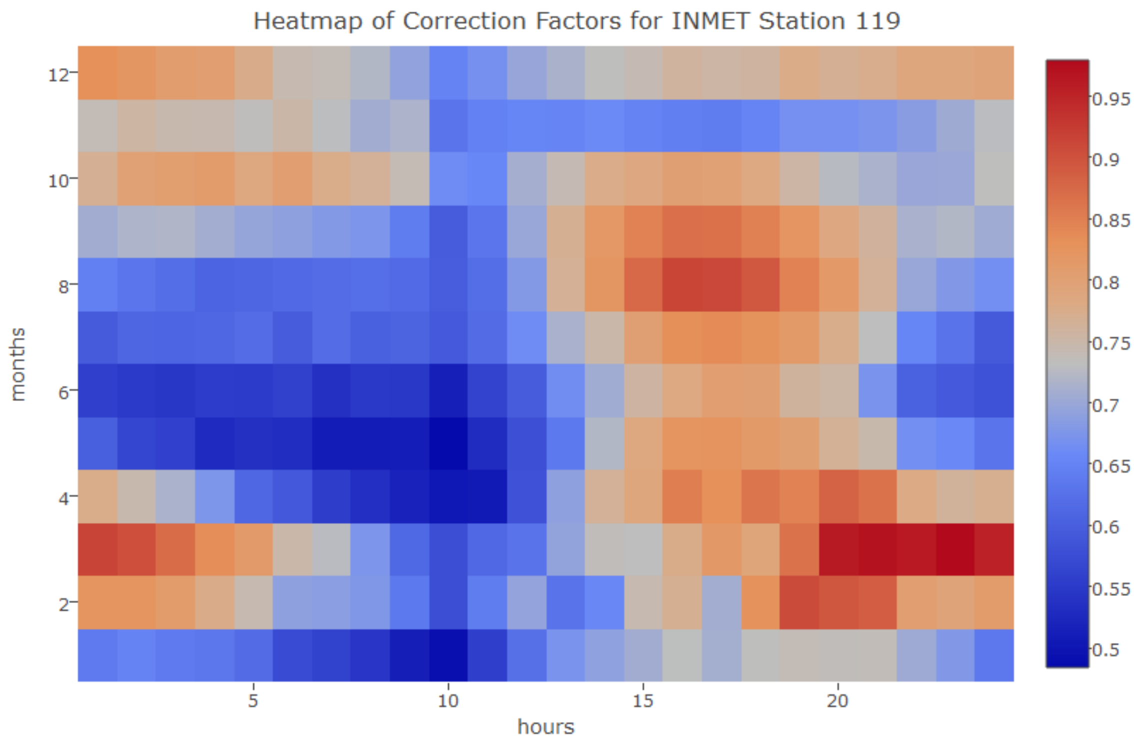


Figure 53: Heat map for the hourly and monthly correction factors of the station with lowest correlation after hourly and monthly correction (own depiction)

To be able to compare all the hourly and monthly correction factors for the different hours, a boxplot, as seen in Figure 17, is useful. As some correction factors were much higher than others, in ranges of more than 2 and up to 120, they are not all displayed in the graph, as the boxes would be too small to examine them. There are many outliers in the higher values, of which only some are included in the graph, for the previously mentioned reason. When looking at the medians and boxes, it can be seen clearly, that they are not all at the same height: The lowest correction factors can be found at 10 am, the highest at 9 pm. Between these hours the correction factors become continuously higher or lower, respectively. What also has to be pointed out, is that nearly for all hours, except for 9 pm, 75% or more of the correction factors are below 1, which means that in general it can be assumed that the reanalysis data overestimate the observed wind speeds and therefore need to be reduced to adapt to the measured values.

Figure 16 shows the distribution of correction factors per month. In this case, also the upper part of the diagram was cut, excluding several outliers, in order to be able to see the boxes better. In this boxplot the boxes and whiskers seem very similar: They are all at about the same height, only a very slight decrease of the correction factors can be seen in the summer, compared to the winter. This underpins the assumption, that the monthly correction does not have a very big influence on the correlation of the simulated and observed data, compared to the hourly correlation, which influences the course of the wind speed curve more intensely.

A summary of the presented results can be found in the main section in chapter 3.1.

6.2 Wind Power Generation Correction – Monthly Results

This section presents detailed results from the monthly analysis of wind power generation correction. First, simulated and observed monthly electricity generation from wind power are compared in a line diagram, to find out about the behaviour over time. Then the simulation and observations, as well as the differences between the simulated generation and the observed generation, are compared in boxplots. In the results (section 3.2) a short summary of the results in form of a table is presented.

In Figure 54, the comparison of monthly wind power generation in GWh for the whole North-East region of Brazil is shown. Monthly comparison data are available since 2006, however, in the first two to three years there were hardly any wind power plants installed in the North-East of Brazil, which is why there was nearly no electricity generation from wind. It seems that the trajectory of the uncorrected line adapts quite well to the observed electricity generation, only being considerably higher. The corrected electricity generation curves all are close to the one of the observed wind power generation curve. In the first years until the beginning of 2014, they all mostly overestimate observed generation. After that, the wind speed corrected and the wind speed and power corrected simulation overestimate the observed generation, whereas the only power corrected electricity generation underestimates it slightly. From the end of 2014, all corrected simulations underestimate the observed wind power generation, the one where both corrections are applied the least. The correlations of the uncorrected (r_nc), the wind corrected (r_wc), the power corrected (r_pc), and the wind and power corrected (r_wpc) simulations are all very high at 99%. In general, from this graph, it seems that the only wind speed corrected simulation adapts best to the observed wind power generation, as the blue line mostly is closest to the black line.

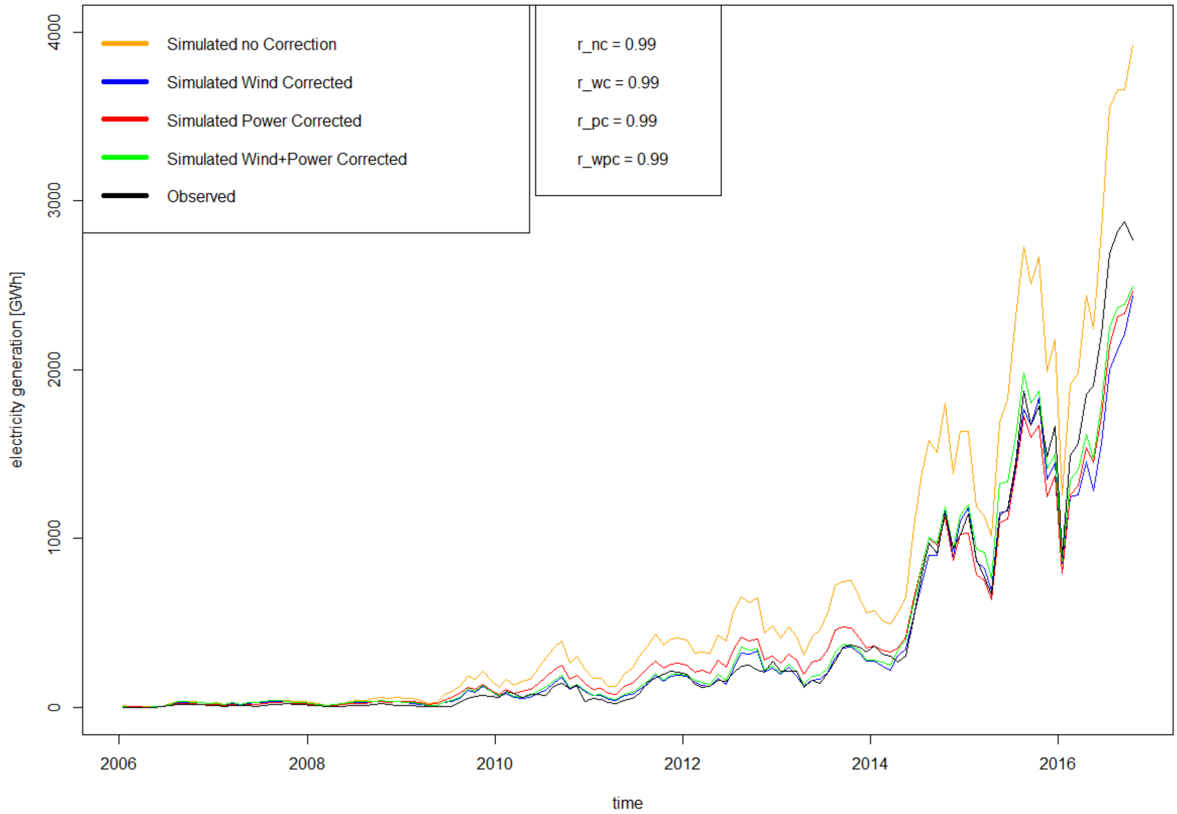


Figure 54: Monthly observed and simulated wind power generation in the North-East of Brazil compared (own depiction)

Figure 55 shows the boxplots for simulated uncorrected, simulated wind corrected, simulated power corrected, simulated wind and power corrected as well as observed monthly electricity generation in GWh from wind power in the North-East of Brazil. As there were many outliers, some of them are not displayed to make the location of the boxes better visible. It can be seen, that the uncorrected simulation, whose whiskers are between 0 and about 1500 GWh per month, overestimates the observed generation, which ranges between 0 and nearly 800 GWh per month, disregarding the outliers. The wind corrected simulated data, which lie in a range of 0 to about 750 GWh per month, however, seem to be near the observed data: When looking at the boxes, the simulation is slightly lower. The box of the only power corrected simulation seems to not fit too well to the box of the measured power generation, as the mean, the quantiles as well as the upper whisker are higher than those of the observed wind power generation. When applying the power generation correction to the wind speed corrected time series, however, the results look better than in the other cases: The boxplot of this simulation and the observed electricity generation from wind power look very similar and only differ slightly. Thus, all bias-corrections deliver better results compared to the simulation without correction, especially when wind speed correction is applied and the boxplots of simulated data fit observed data best when applying both corrections. All of the boxplots have many outliers in the higher range (the ones of the uncorrected simulation are outside the visible area of the graph), but as this is the case for all the plots, it should not be of great importance when interpreting the plots.

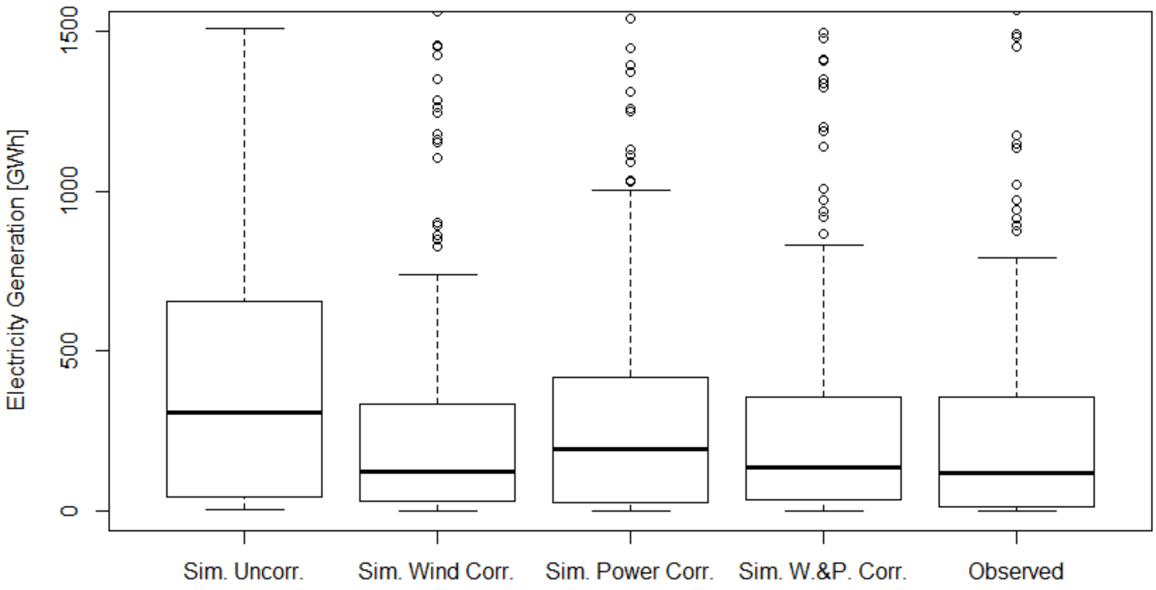


Figure 55: Boxplot for comparison of monthly power generation in the North-East of Brazil (own depiction)

Figure 56 shows the absolute differences in GWh between the uncorrected, wind and/or power corrected simulated monthly electricity generation from wind power and the observed wind power generation in the North-East of Brazil. For the uncorrected data, the differences range from 0 to nearly 900 GWh and are much higher, compared to the differences of the corrected to the observed data, which lie between 0 and about 100 GWh for the wind only or wind and power corrected electricity generation, disregarding the outliers, and between 0 and 300 GWh for the power generation corrected simulation, when not considering the outliers. Therefore, the distributions of the differences of corrected to simulated data are also less spread for the corrected data (indicated by significantly smaller whiskers and boxes) and also the medians are much lower, being below 100 GWh, whereas the median for the differences of uncorrected simulated data to the observed data is at about 200 GWh. What is moreover striking, is, that when looking at the differences of the simulated monthly generation data without correction to the observed monthly generation data, more than 50% of the differences are above 200 GWh, which indicates a considerable overestimation of the observed generation before correction. The smallest differences seem to occur when applying the wind speed correction only, however, there are more outliers than in the other cases. In general, as also observed in Figure 55, the best results are achieved when applying wind speed correction (and wind power generation correction optionally).

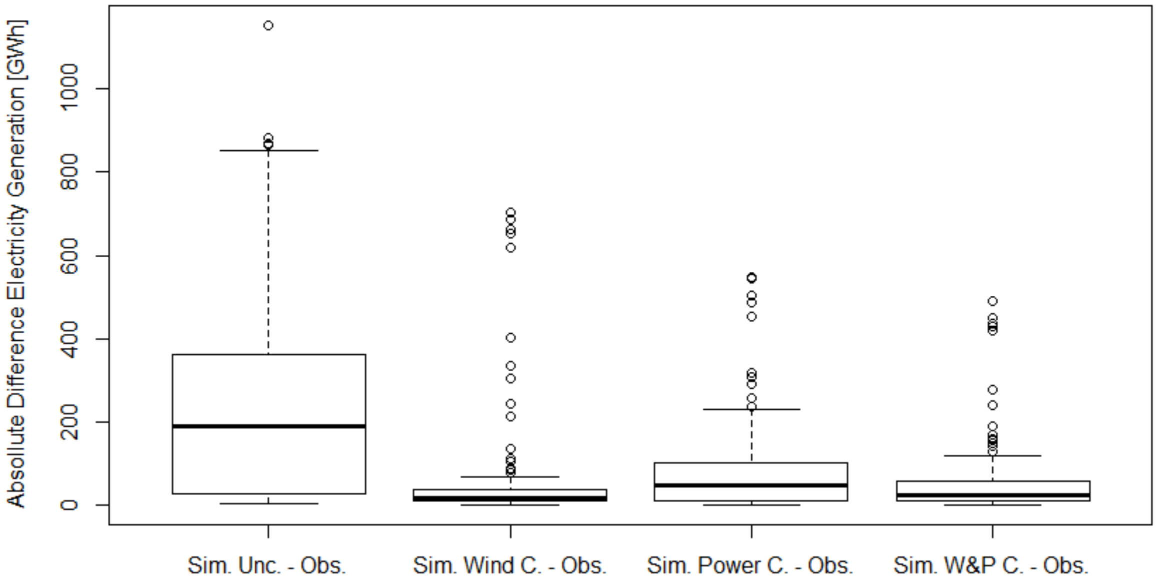


Figure 56: Boxplots for comparison of absolute differences between observed and simulated uncorrected or corrected monthly wind power generation (own depiction)

In Figure 57, the real differences between the uncorrected simulated or the corrected simulated and the observed monthly wind power generation are displayed. The uncorrected simulation remains in the range of 0 to about 900 GWh difference to the observed power generation because it only overestimates the observed wind power generation. In contrast to this, the differences of corrected monthly electricity generation from wind power to measured electricity generation shift to a range near 0 GWh, only with several outliers differing from that range in the negative area and therefore supply a better approximation to the observed power generation data. When correcting the data with monthly power generation bias correction factors, however, there are differences in the positive as well as in the negative range, meaning, that in some months the power generation is underestimated whereas in other months it is overestimated. Again it can be observed, that the correction significantly improves the results and correlation with observed wind power generation, especially when applying the wind speed correction. This supports the observations made before.

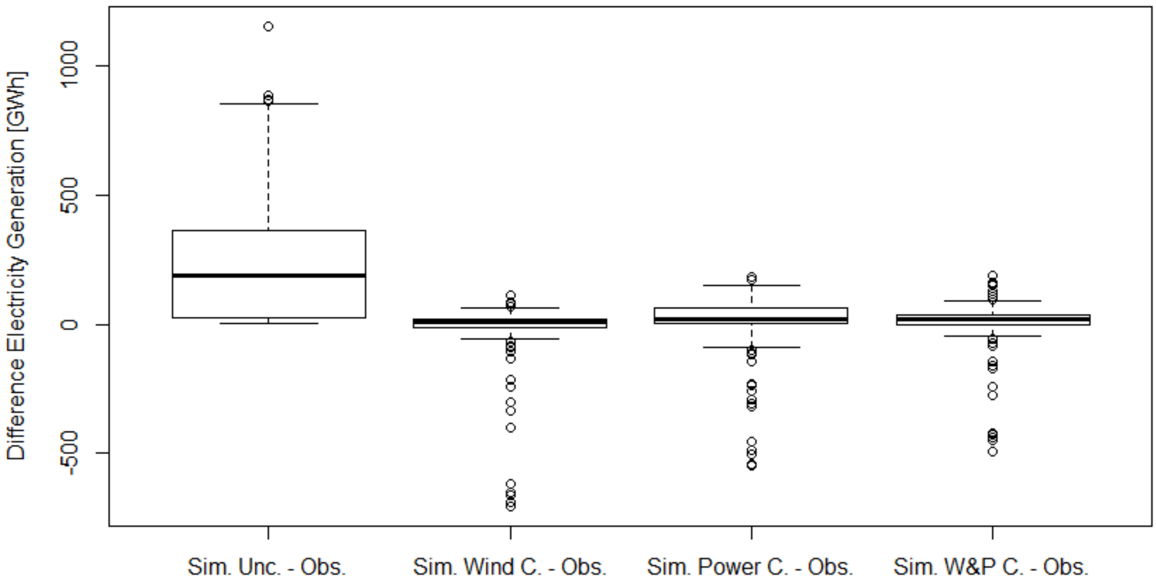


Figure 57: Boxplots for comparison of differences between measured and simulated uncorrected or corrected monthly wind power generation (own depiction)

The following contour plots compare observed and simulated monthly power generation. The first plot, Figure 58, shows the comparison of observed and simulated monthly electricity generation without applying any correction factors in a time span between January 2006 and October 2016. The contour plot shows only a slight dependence of the simulated wind power generation on the observed power generation. There is an accumulation at lower values, for the simulated generation between 0 and 1000 GWh monthly and for the measured wind power generation between 0 and 500 GWh. This and the range of the two time series (up to 3500 GWh for the ONS data and up to 5000 GWh for the calculated wind power generation) mean that the simulated production overestimates the observed power generation in the North-East of Brazil, which fits the observation made in Figure 54.

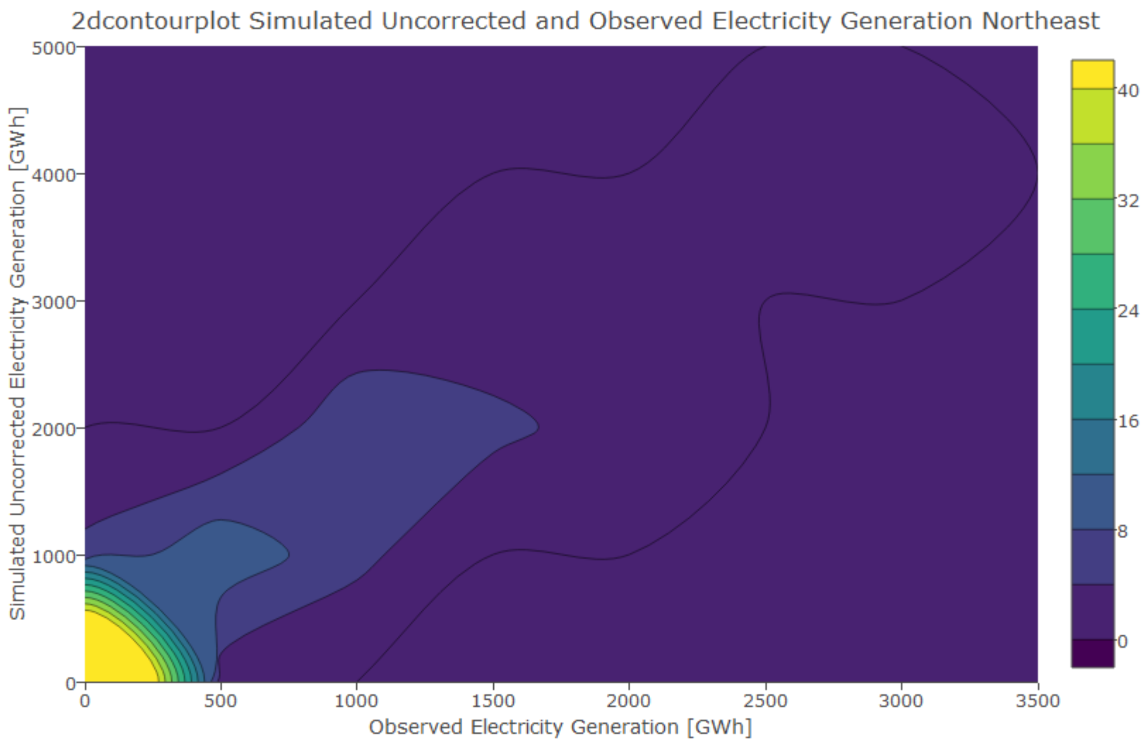


Figure 58: Contour plot for monthly observed and simulated power generation without any correction (own depiction)

Figure 59 illustrates the observed and wind corrected simulated monthly wind power generation. Compared to the previous contour plot, a clearer correlation can be determined. Still, there is an accumulation at low values, but this time for the measured as well as for the simulated data between 0 and 500 GWh monthly. A smaller agglomeration of generations can be found at 1000 GWh monthly generation. Only for the higher values of wind power generation, there cannot be found a clear dependence, which is due to a lower frequency of events for higher wind power generation. It can be assumed that the simulation slightly underestimates the observed wind power generation, as it ranges only up to 3000 GWh monthly (compared to 3500 GWh), at least for the higher values, which again fits the observations made in Figure 54. What can be observed too, is, that the wind speed correction improves the data, as after the correction the simulation better fits the range of measured data.

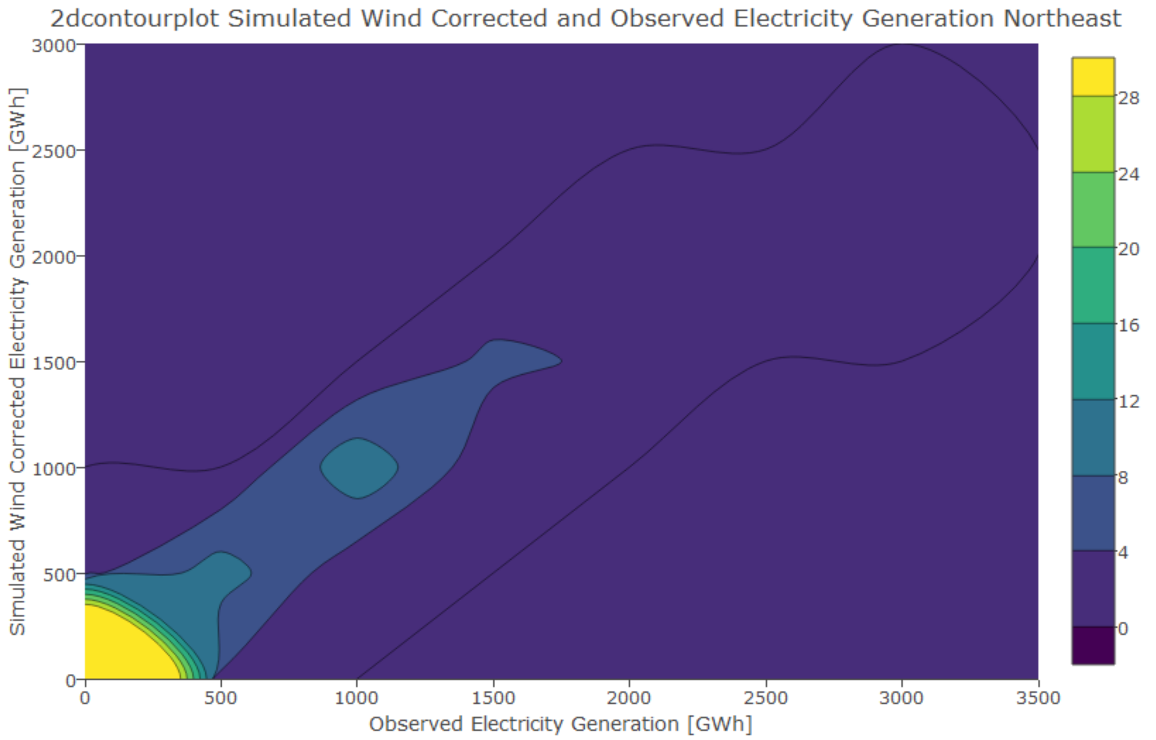


Figure 59: Contour plot for monthly observed and simulated power generation with wind speed correction (own depiction)

After bias correction with the wind power generation correction factors, the simulated power generation is reduced (Figure 60), similarly to the wind power generation after wind speed correction. However, in a few cases, the simulated and corrected electricity generation is lower, at about 1000 GWh monthly, than the observed electricity generation, which then is between 1500 and 2000 GWh. For lower values, though, the opposite is the case: Where the measured wind power generation is concentrated between 0 and 500 GWh monthly, the simulation provides monthly wind power generation of up to 1000 GWh. This fits the observation made in the graph in Figure 54, where the simulated generation after power correction until the middle of 2014 underestimates the observed generation, and later overestimates it.

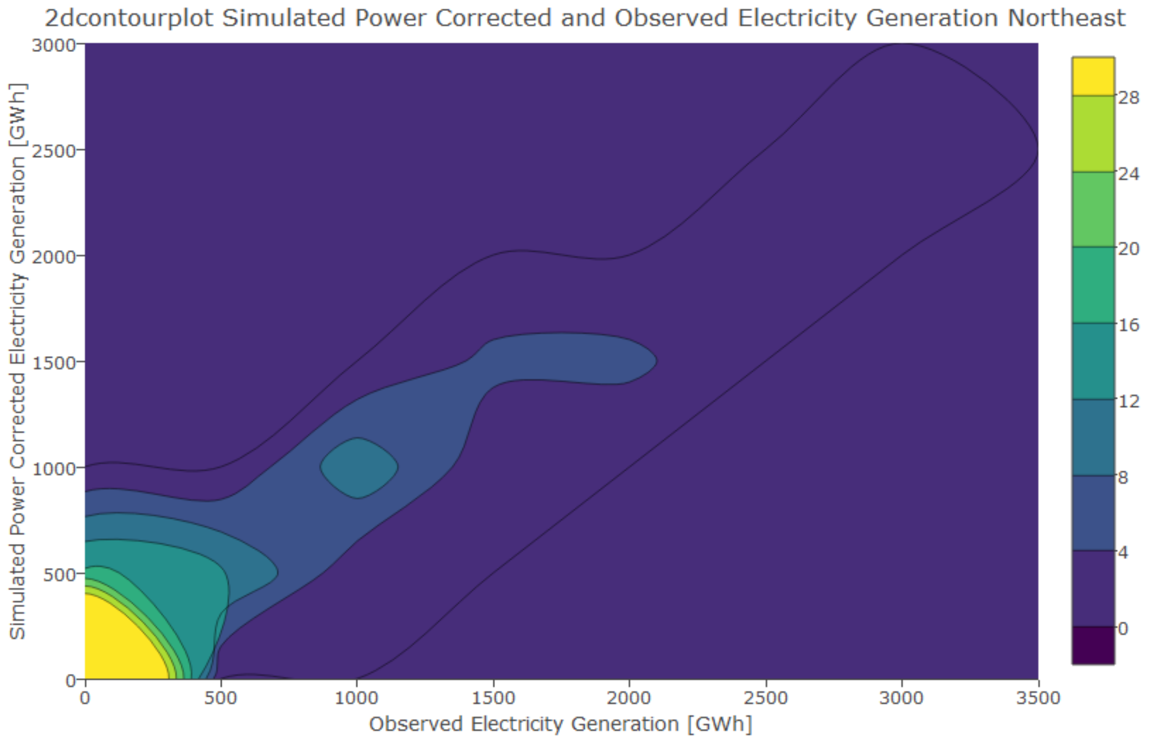


Figure 60: Contour plot for monthly observed and simulated power generation with wind power generation correction (own depiction)

When applying the wind power generation correction after the wind speed correction on the simulated power generation, which can be seen in Figure 61, there seems to be a better correlation than when only correcting with wind power generation correction factors. However, from the contour plot there can hardly be any improvement perceived, compared to wind speed correction only. As in the other simulations with only either wind speed or wind power generation correction, for higher wind power generation, the simulation underestimates the measured wind power generation, which also fits the observations made in Figure 54.

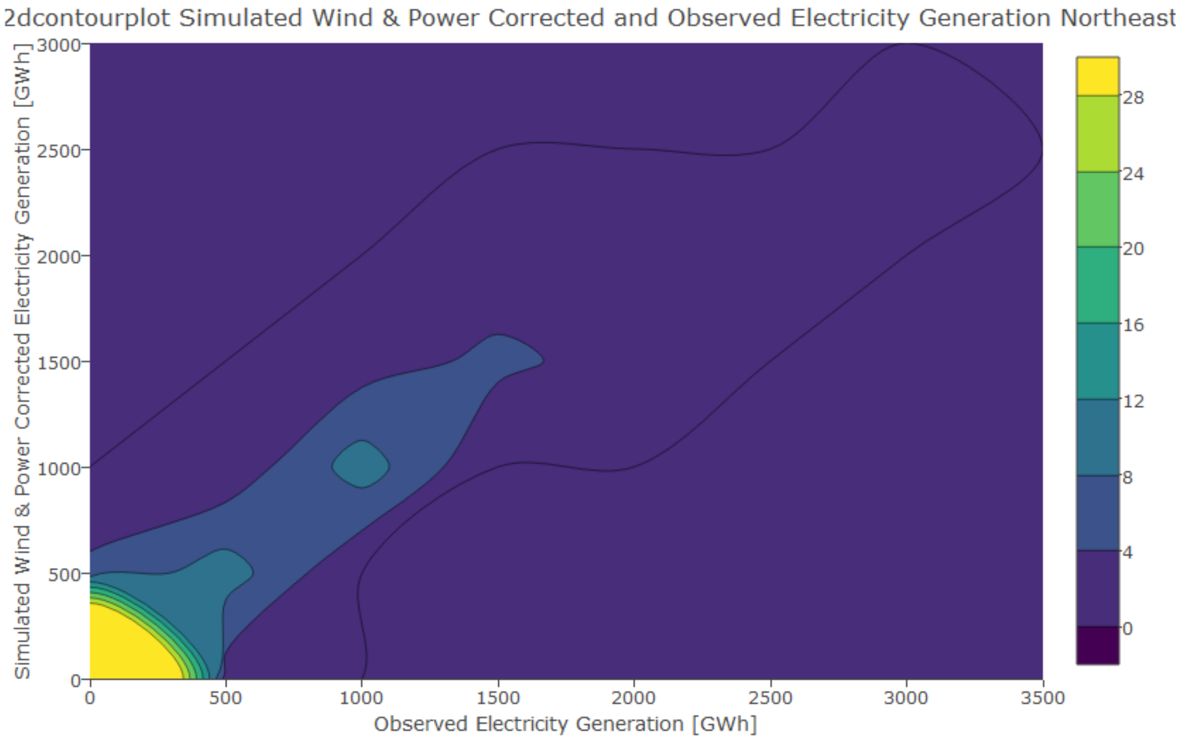


Figure 61: Contour plot for monthly observed and simulated power generation with wind speed and wind power generation correction (own depiction)

Considering these observations, it cannot be said that the corrected simulated electricity generation in general overestimates or underestimates the observed electricity generation. Nevertheless, it can be assumed, that, once more capacities are installed, generation generally deviates more.

6.3 Wind Power Generation Correction – Daily Contour Plots

This section presents an additional analysis of daily electricity generation for the results of wind power generation bias correction in the form of contour plots. Further results are found in section 3.2.

When looking at the daily generation, which is compared in Figure 62, for a time span between August 2015 and October 2016, the simulated and observed power generation seem to correspond well, as frequent occurrences form a nearly diagonal line. However, the simulated generation is in a range between about 20 and 180 GWh daily, whereas the observed generation lies between about 20 and 120 GWh daily, which means that, without any correction, the simulated electricity generation from wind power overestimates the observed electricity generation, which has already been observed in Figure 18. This is also underpinned by looking at the most frequent daily wind power generations: The simulated wind power generation mostly ranges between 50 and 100 GWh, whereas the observed wind power generation mostly lies between only 40 and 80 GWh per day.

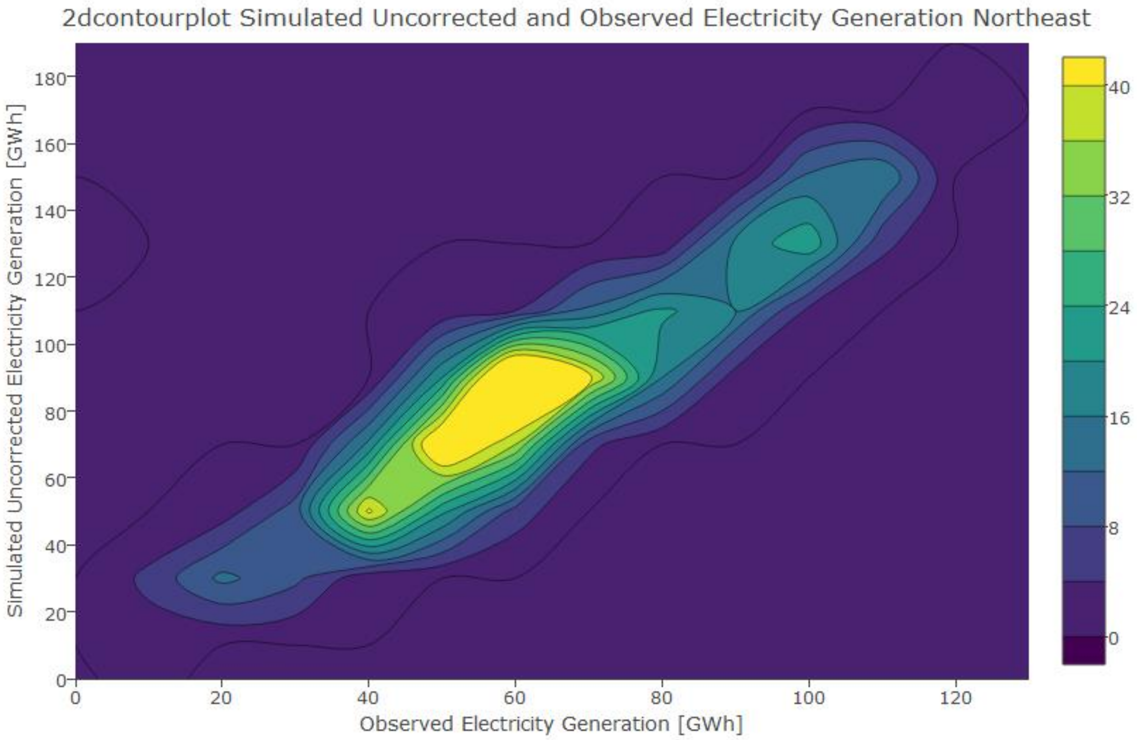


Figure 62: Contour plot for daily observed and simulated power generation without any correction (own depiction)

The following contour plot (Figure 63) shows the daily observed as well as wind speed corrected wind power generation. After the correction, the correlation between simulated and observed electricity generation is not that clearly linear any more, however, the range of simulated daily wind power generation is lowered to between 10 and 110 GWh and now fits the range of the measured electricity generation from wind power better. Now also for observed as well as for simulated wind speed corrected wind power generation, the most frequent (yellow and light green area) daily wind power generations lie between 40 and 80 GWh.

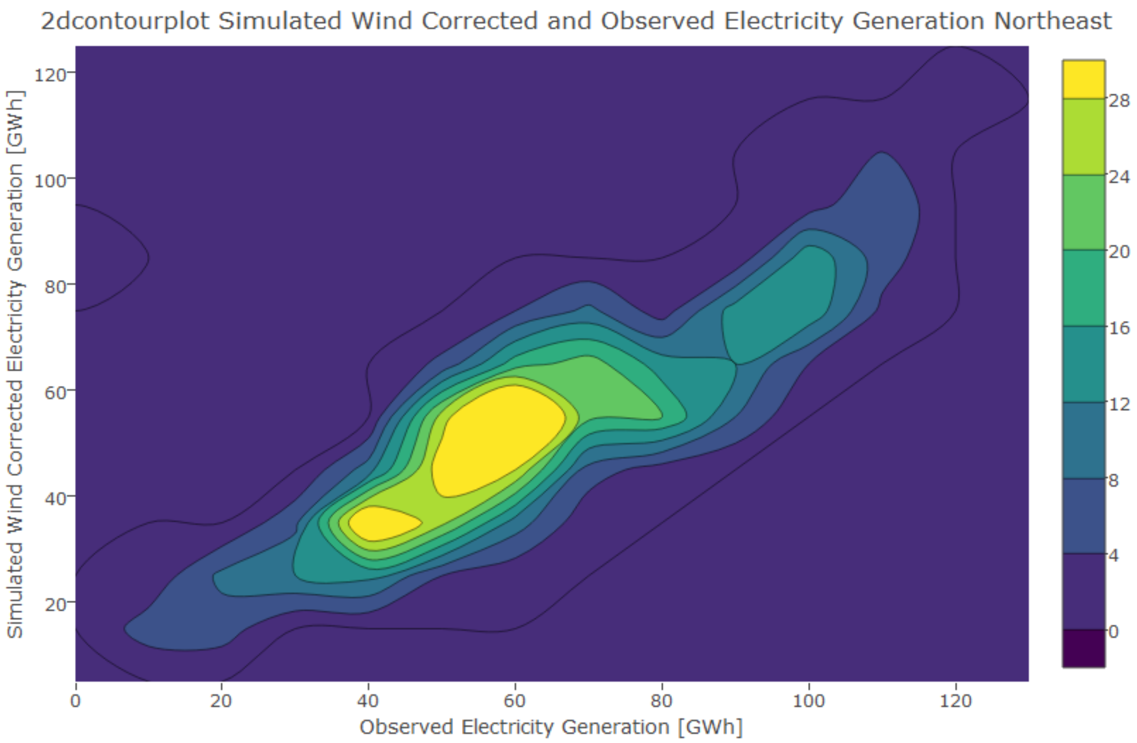


Figure 63: Contour plot for daily observed and simulated power generation with wind speed correction (own depiction)

When correcting the simulated wind power generation only with wind power generation correction factors and comparing it to the measured wind power generation, the correlation between these two seems to be more linear than when applying the wind speed correction. However, the observed electricity generation is underestimated at higher values between about 80 and 120 GWh daily. For lower values, the simulated and measured generation seem to correlate well, though. This is not supported by what was observed in Figure 18, as in the beginning of the considered timespan, where the generation is slightly lower than in the later days, the observed wind power generation is overestimated, and in the end of the considered timespan, where it is higher, it is underestimated by the wind speed corrected simulation. So only the underestimation for higher generation can be observed in both graphs.

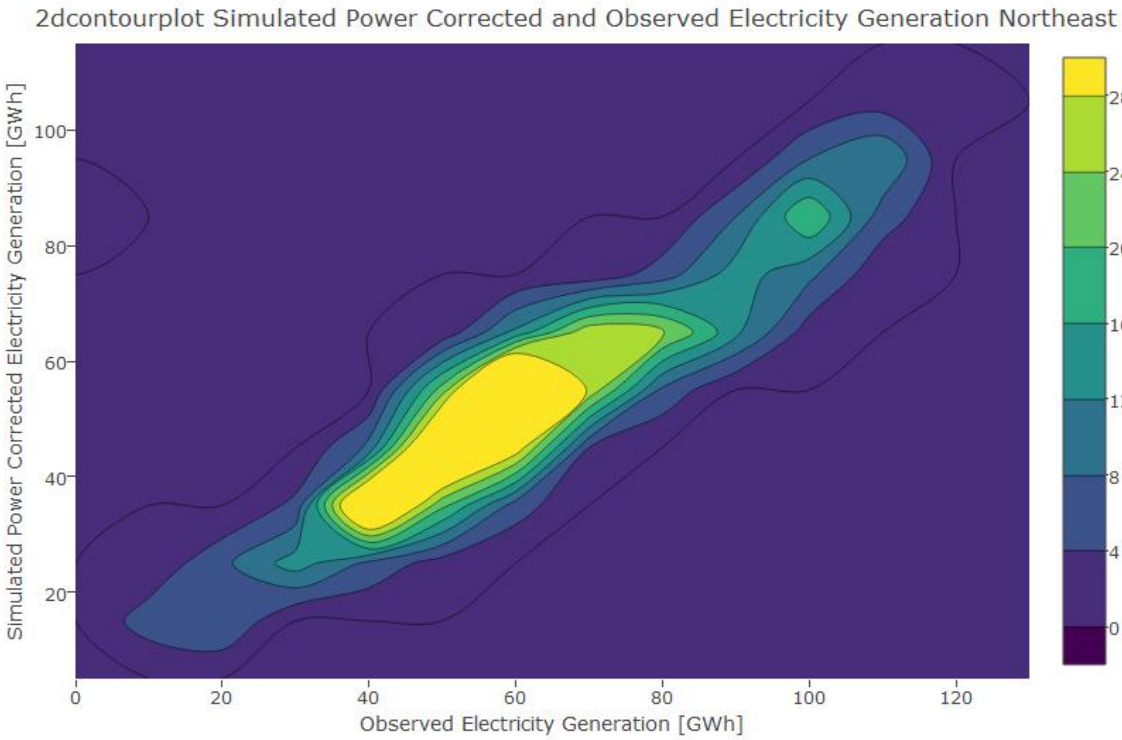


Figure 64: Contour plot for daily observed and simulated power generation with wind power generation correction (own depiction)

When the simulated wind speed corrected generation is corrected also with the monthly bias correction factors, which is illustrated in Figure 65, the relation between the simulated and the observed wind power generation in the North-East of Brazil seems not to be as direct as when only correcting with monthly wind power generation correction factors, as there is no clear diagonal line anymore. However, the simulated generation does not underestimate the observed electricity generation anymore for higher values, as both, the observed as well as the simulated daily energy outputs, lie between 20 and 120 GWh. What can be drawn as a conclusion from all the contour plots is, that the observed and simulated wind power generation mostly concur, especially when applying wind speed correction.

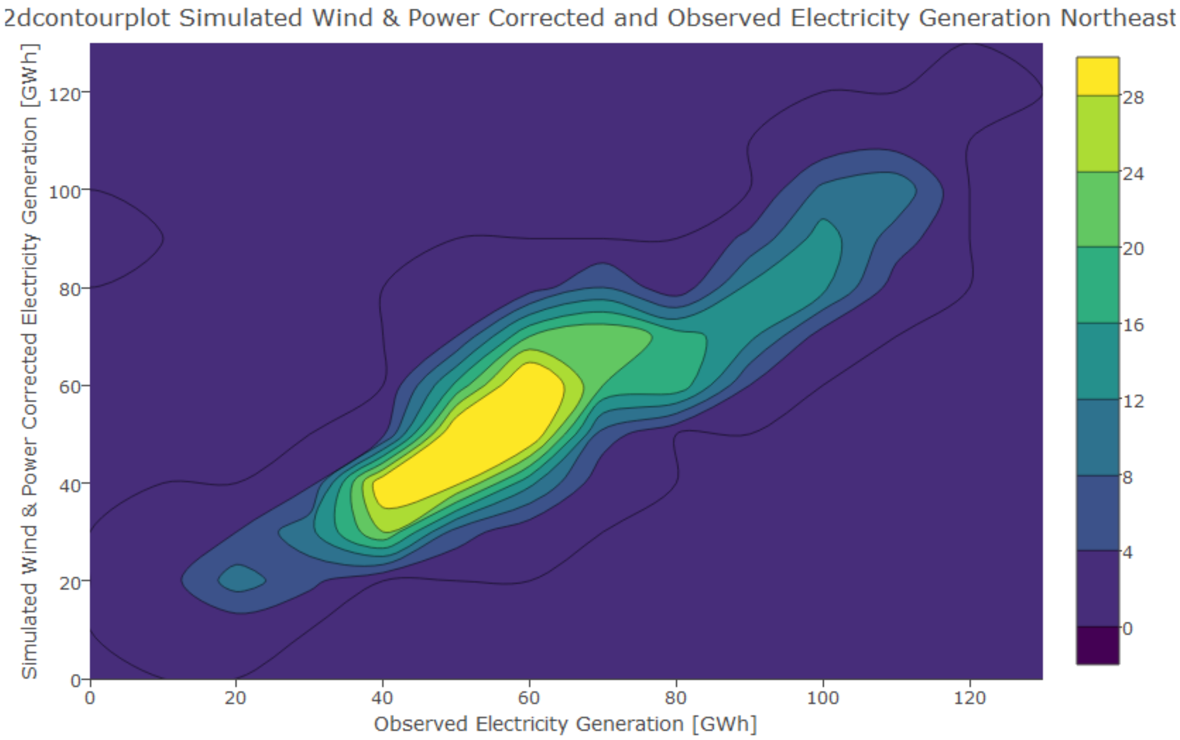


Figure 65: Contour plot for daily observed and simulated power generation with wind speed and wind power generation correction (own depiction)