University of Natural Resources and Life Sciences, Vienna

H73 Department of Economics and Social Sciences

H731 Institute for Sustainable Economic Development

MASTER THESIS

The simulation of potential photovoltaic production with MERRA data in Germany

Author

Karin Egger, MSc

In partial fulfilment of the requirements for the degree of

Diplom-Ingenieur (Dipl.-Ing.)

Vienna, 2017

Code of study:	H 066 427
Field of study:	Environment and Bio-Resources Management
Adviser:	Univ.-Prof. Dipl.-Ing. Dr. Erwin Schmid
Co-adviser:	Ass.-Prof. Dipl.-Ing. Dr. Johannes Schmidt

Acknowledgement

To begin with I would like to thank all those who supported me during the development of this thesis.

First of all my adviser, Erwin Schmid, for supervising this thesis.

Furthermore I want to thank my co-adviser, Johannes Schmidt, for the ideas and time he put into this work, the great support in all issues and the patience when helping me with R.

A big thanks also to my parents, Andreas and Andrea, who always supported me in following my goals and fulfilling my dreams, not only by their financial but especially by their mental support, and my sister and friends, who are always there for me.

Special thanks to you, Christian, for the support, patience and encouragement all the time.

Affirmation

I hereby declare under oath that I have prepared the present work independently and without use of any aids except the given ones.

From other sources directly or indirectly taken thoughts are designated as such.

The work has not been submitted to any other examination board in the same or similar form and has not been published yet.

English abstract

The aim of this thesis is to assess the MERRA 2 reanalysis data for the simulation of photovoltaic production. Therefore, temperature and solar radiation data are used to simulate the photovoltaic production in Germany from 2010 to 2013. The simulations consider different settings regarding the inclination and orientation of the photovoltaic panels. The simulation results are then statistically compared to measured generation data of the same period in assessing the accuracy of the simulated data and investigating systematic bias. The results of the comparison show that the best setting of the panels for a correct simulation is an inclination of 15° and a South-West orientation. It further shows that the simulation with MERRA 2 data overestimates the measured production, whereby the overestimation is highest in winter with 43% while it is lower in the other seasons with 7% on average. A correction factor of 0.92 provides reasonable improvements in the simulation results. Extreme values and anomalies in photovoltaic production are assessed for the period of 1980 to 2015. The results indicate neither anomalies nor extreme values. Finally, the average hourly and daily capacity factor for all 35 years is calculated and compared to the results of Pfenninger & Staffell (2016a), they also used the MERRA 2 data for the simulation of photovoltaic production in Germany. A correlation coefficient of 0.95 shows a good consistency between the two simulations.

Kurzzusammenfassung

Das Ziel dieser Masterarbeit ist die Analyse von MERRA 2 Reanalyse Wetterdaten für ihre Eignung in Simulationen von Photovoltaikproduktion. Hierfür werden die Temperatur und Solarstrahlung aus dem Datensatz in einer Simulation der Photovoltaikproduktion für Deutschland im Zeitraum von 2010 bis 2013 verwendet. In der Simulation wurden verschiedene räumliche Ausrichtungen der Photovoltaikpanels untersucht, da diese nicht im Detail bekannt sind. Die Ergebnisse der Simulation wurden anschließend mit den gemessenen Produktionsdaten dieses Zeitraums statistisch verglichen, um die Genauigkeit der Simulation und eventuelle Über- oder Unterschätzungen zu bestimmen. Die Ergebnisse zeigen, dass die Ausrichtung mit einer Panelneigung von 15° und einer Orientierung nach Südwesten am besten die gemessenen Produktionsdaten wiedergeben können. Dennoch führt die Simulation zu einer klaren Überschätzung der gemessenen Produktion. Diese Überschätzung ist mit 43% im Winter am höchsten, während sie in den anderen Jahreszeiten durchschnittlich 7% beträgt. Ein Korrekturfaktor von 0,92 führt zur Verbesserung der Simulationsergebnisse. Die Simualtionsdaten für die Periode 1980-2015 werden auf eventuelle Anomalien und Extremwerte hin untersucht. Die Ergebnisse zeigen keine größeren Anomalien und Extremwerte. Darüber hinaus wurden die Simualtionsdaten mit den Ergebnissen von Pfenninger & Staffell (2016a) verglichen, die ebenfalls eine Simulation der Photovoltaikproduktion mit MERRA 2 Daten durchgeführt haben. Ein Korrelationskoeffizient von 0,95 zeigt eine gute Übereinstimmung zwischen den beiden Simulationen.

I Table of contents

II List of figures... XI

III List of tables... XIV

IV List of abbreviations... XV

1 Introduction...1

1.1 Comparative literature 3

1.2 Objectives of the thesis 6

2 Investigation area - Germany...9

2.1.1 Temperature, sunshine and solar radiation 9

2.1.2 Electricity market... 11

2.1.3 Photovoltaics 12

3 Methods and Data...14

3.1 Used data sets... 15

3.1.1 MERRA data 16

3.1.2 Installed capacity of photovoltaic systems....................................... 18

3.1.3 Measured generation data of photovoltaic production........................... 21

3.2 Methods of analysis... 23

3.2.1 Simulation of photovoltaic production with solaR............................... 23

3.2.2 Correction factor for simulation results 33

3.2.3 Comparison of simulated and measured generation data....................... 33

3.2.4 Time series analysis 34

4 Results 35

4.1 Comparison of simulated and measured generation data 35

4.1.1 Overview 39

4.1.2 Hourly comparison 40

4.1.3 Daily comparison 43

4.1.4 Monthly comparison 44

4.1.5 Seasonal comparison 45

4.2 Time series analysis 49

4.3 Comparison of simulated data to Pfenninger & Staffell (2016a) 52

5 Discussion & Conclusions 56

V Bibliography X

VI Appendix XIV

II List of figures

Figure 1: Simulated average hourly capacity factor for Germany and a comparison to measured data (Pfenninger & Staffell 2016a) 5

Figure 2: Average monthly temperature in Germany (Wetter.de 2016, database: Deutscher Wetterdienst) 9

Figure 3: Average monthly duration of sunshine in Germany (Wetter.de 2016, database: Deutscher Wetterdienst) 10

Figure 4: Average annual sum of Global Horizontal Irradiation in Germany (SolarGis 2016) 11

Figure 5: Spatial extent of the four German grid zones (Own representation) 12

Figure 6: Development of electricity production and installed power of photovoltaic systems in Germany 1990 – 2013 (BMWi 2017) 13

Figure 7: Working process of the simulation, comparison to measured data and time series analysis (Own representation) 14

Figure 8: Distribution of MERRA points over Germany (Own representation).............. 17

Figure 9: Installed photovoltaic capacity in Germany from 2010 - 2013 (Own representation, Data source: Bundesnetzagentur 2016)...................................... 20

Figure 10: Daily electricity production from photovoltaics in Germany 2010 – 2013 (Own representation, Data Source: EEX Transparency Platform).......................... 22

Figure 11: Monthly photovoltaic production for Germany and its four grid zones 2010 – 2013 (Own represenation, Data source: EEX Transparency Platform) 22

Figure 12: Process of simulation of photovoltaic production in the statistics software R (Own representation) 23

Figure 13: Comparison of monthly measured generation to production under different simulation settings (Own representation).. 37

Figure 14: Surplus of simulated to measured photovoltaic production for average monthly production in Germany from July 2010 to July 2013 (Own representation) 37

Figure 15: Surplus of simulated to measured photovoltaic production for average monthly production in Germany from July 2010 to July 2013 after application of correction factor (Own representation) 38

Figure 16: Comparison of measured generation data and simulation for average hourly photovoltaic production in Germand and the four grid zones from July 2010 to July 2013 (Own representation)..40

Figure 17: Comparison of measured generation data and simulation for average hourly photovoltaic production in the four grid zones from July 2010 to July 2013 (Own representation) 41

Figure 18: Scatter plot of hourly simulated and measured production per month in Germany July 2010 – July 2013 (Own representation)..42

Figure 19: Box plots of simulated and measured data for grid zones and Germany based on daily data (Own representation) 43

Figure 20: Measured (full line) and simulated (dotted line) monthly photovoltaic production for the four grid zones from July 2010 to July 2013 (Own representation) 44

Figure 21: Measured (full line) and simulated (dotted line) monthly photovoltaic production for Germany from July 2010 to July 2013 (Own representation) 45

Figure 22: Surplus of simulated to measured photovoltaic production for average seasonally production from July 2010 – July 2013 in Germany (Own representation) 47

Figure 23: Surplus of simulated to measured photovoltaic production for production in the winter periods from July 2010 – July 2013 in Germany (Own representation) 47

Figure 24: Comparison of measured generation data and simulation for average hourly photovoltaic production in the four seasons from July 2010 to July 2013 (Own representation) 48

Figure 25: Average daily capacity factor for photovoltaic systems in Germany from 1980 to 2015 (Own representation) 49

Figure 26: Hourly capacity factor for photovoltaic systems in Germany from 1980 to 2015 for the four seasons (Own representation) 50

Figure 27: Average annual electricity production of 1kWp installed photovoltaic capacity in Germany 2015 (Own calculation, Own representation) 51

Figure 28: Comparison of Pfenninger & Staffell (2016a) and own simulation for average capacity factor per month from 1985 – 2014 (Own representation) 52

Figure 29: Surplus of own simulation compared to Pfenninger & Staffell (2016a) for average capacity factor per month from 1985 – 2014 (Own representation)....... 53

Figure 30: Comparison of the daily capacity factor by the simulation of Pfenninger & Staffell (2016a) 1985 -2014, upper graph (red), to own simulation results 1980 – 2015, lower graph (orange). 54

Figure 31: Diurnal variability of hourly capacity factor in Germany 1985 - 2014 of Pfenninger & Staffell (2016b), upper graph (red), to own simulation results 1980 - 2014, lower graph (orange). 55

III List of tables

Table 1: Total installed photovoltaic capacity in the four grid zones in total Germany from 2008 – 2013 in GWp 18

Table 2: Electricity production of four grid zones and total Germany from 07/2010 to July 2013 in TWh (Own representation, Source: EEX Transperency Platform) 21

Table 3: Correlation coefficients of different settings for inclination and orientation in the simulation compared to measured generation data of the photovoltaic production (Own calculation) 36

Table 4: Correlation coefficients for the comparison between simulation and measured generation data from July 2010 to July 2013 with a significance level of 0.01 (Own calculations) 39

Table 5: Correlation coefficient (R) between simulation and measured generation data for the four seasons separated by year and based on hourly data (Own calculation) 46

IV List of abbreviations

GW	gigawatt
GWh	gigawatt hours
GWp	gigawatt peak
kW	kilowatt
kWh	kilowatt hours
kWh/m^2	kilowatt hours per square meter
kWp	kilowatt peak
MW	megawatt
MWh	megawatt hours
TW	terawatt
TWh	terawatt hours
W/m^2	watt per square meter
‰	per mill
°C	degree Celsius

1 Introduction

In December 2015 the 21st United Nations Climate Change Conference (COP 21) ended with the so called Paris Agreement that was signed by 158 of 195 member states of the United Nations so far (United Nations 2017). The main aim of the agreement, stated in article 2, is to hold the global temperature increase, that is caused by human climate change, below 2 °C of pre-industrial times, if possible even below 1.5 °C. Therefore the member states will submit their own national climate plans appropriate to their possibilities every five years and try to implement them (UNFCCC 2015).

The European Union set its pathway to reduce the climate change already in 2007, by implementing its 20-20-20 goals. According to these goals three main actions should be taken by the member states of the European Union: first a 20% reduction of greenhouse gas emissions compared to the level in 1990, second an increase of energy efficiency by 20%, and third the increased production of energy from renewables like biomass, wind, water, and solar power up to 20% of total energy consumption (European Commission 2017). Regarding renewables, especially solar power can play a major role in this transition. Per day about 1,367 W/m^2 of solar radiation reach the outer atmosphere of earth. This so called solar constant of 1,367 W/m^2 delivers about $1.5 * 10^9$ TWh of energy per year. Due to atmospheric effects like absorption, reflection or scattering through clouds and particles only 50% of this solar radiation reach the earth surface. Considering additionally a land surface of 35% that can be used for the collection of this solar energy, the solar radiation still delivers an amount of 262,500,000 TWh of energy per year (Zahoransky et al. 2012). In comparison to that, the primary energy consumption of the whole world was about 157,500 TWh in 2013 (OECD/IEA 2015), so only 0,6 ‰ (per mill) of the average energy delivered by the sun per year. So theoretically it is be possible to cover the global energy demand solely by solar power.

One problem concerning the production of energy by solar power is that it is not available continuously over a day and year. Fluctuations over a day are due to night times, when no sun is shining, or weather events like clouds, rain or snowfall, where most of the solar energy does not reach the earth surface. Furthermore the availability of solar energy over the year changes with seasons and the geographic position on the globe, that is the longitude and latitude of the location. While at locations close to the equator, the production of energy by solar power is very high, it is lower in regions closer to the poles, as solar radiation becomes lower due to the smaller angle of incidence. Therefore building solar heating systems and photovoltaic systems close to the equator would be an optimal solution to use solar energy. But a problem arises due to geopolitical conflicts between countries and debates about the wish of energy independency by single countries. Furthermore the infrastructure to transfer the produced energy (photovoltaic systems) or heat (solar heating systems) from one country to another is not always available or fully developed (Paschotta 2017). So for each country the search for suitable sites for the installation of solar systems, that bring a maximum yield, and the question of how to implement those systems into the national heat and electricity production arise (Pfenninger & Staffell 2016a). The present thesis focuses only on photovoltaic systems for electricity production and not on solar heating systems for heat production.

A simulation of their potential electricity production can be conducted to find appropriate places for the installation of photovoltaic systems. For such a simulation one needs to consider the factors that influence the production output of a photovoltaic system. First of all the photovoltaic module itself influences the output depending on the materials it is build of and the way it is positioned. While the material in most cases is given by the producers and cannot be influenced, the correct positioning of the system, that is the inclination and the orientation of the panel, is variable. The perfect orientation in Germany would be to South while the inclination is optimal when it is the latitude of the location minus 10° (Wirth 2017). Obviously, the output is mainly influenced by the solar radiation that reaches the photovoltaic system, as well as, to a minor extent, by the surrounding temperature and the solar radiation that reaches the photovoltaic system. Solar radiation hereby refers to the total solar radiation that reaches the surface of the earth and that is separated into direct radiation and indirect radiation. While direct radiation reaches the outer atmosphere and then goes straight to the surface of the earth, indirect radiation is scattered in the atmosphere by particles and afterwards reaches the earth surface (Zahoransky et al. 2012). So to simulate the production output at a certain location, it is necessary to gather information about solar radiation and temperature, and its change during the day and year at this location. Furthermore the correct orientation and positioning of the panel regarding its tilt angle needs to be taken into account. The gathering of the solar radiation and temperature data can be very time consuming if measured over a long period. But on the other hand, if the collection of data is done only over short periods extreme generation events that yield especially high or low generation are not included in the measurement. Therefore long term data available for several years would be a better measure to understand generation at a certain location. One source of such information is the MERRA dataset (Modern Era-Retrospective Analysis for Research and Applications) of the Global Modeling and Assimilation Office (GMAO). This reanalysis data in its second version (MERRA 2) delivers different meteorological factors as wind speed, solar radiation, or temperature on an hourly basis since the year 1980. The data is available on a point grid with a spatial resolution of 0.625° longitude and 0.5° latitude for the whole world (Bosilovich, Lucchesi, and Suarez 2016). As this data is available in such a high temporal and spatial resolution for the whole world, it is a perfect basis for the simulation of photovoltaic production output at different locations. But so far only a few studies have dealt with the topic of simulating potential photovoltaic production by using MERRA data.

1.1 Comparative literature

Only a few studies examined the topic of MERRA data as a basis for photovoltaic production simulation. Furthermore, as Pfenninger & Staffell (2016a) state in their study, *"… existing studies perform limited or no validation (in space and time) of their reanalysis based simulations against historical power output …"*. Three exemplary studies that do so are Juruš et al. (2013), Richardson & Andrews (2014) and Pfenninger & Staffell (2016a).

Juruš et al. (2013) used the first version of MERRA data to simulate the annually photovoltaic production (YPP) in parts of the Czech Republic for 33 years (1979 – 2011) and compared it to measured generation data. Furthermore they compared the variable of the surface incident shortwave flux from the MERRA data set to other more common datasets that are used for simulation and modeling of photovoltaic production. As a result of this comparison, they found that MERRA slightly overestimates the shortwave radiation flux, which leads to a higher potential production in a simulation. On the contrary they state that the advantage of MERRA is that it can be used for long term and global analysis due to its spatial extent and longtime records starting in 1979.

Richardson & Andrews (2014) first validated the accuracy of MERRA data by comparing the MERRA data on solar radiation with ground measurements of solar radiation. Their comparisons showed that MERRA data slightly overestimates the solar radiation, especially on an hourly basis, but is nevertheless suitable for the photovoltaic simulation. They also found that the error between MERRA and ground measurements is smaller in summer than in winter, which is due to less sunshine in the winter months that increases the relative error between MERRA data and measurements. Afterwards they conducted a simulation of possible photovoltaic production for different sites in the U.S. and validated the results with measured photovoltaic production data. Here they found similar results as in the first comparison of solar radiation. MERRA slightly overestimates production, especially on an hourly base but is suitable for use of daily simulations.

Pfenninger & Staffell (2016a) also did a simulation of potential photovoltaic production across Europe based on MERRA data and on satellite-derived SARAH data. For the MERRA data they first interpolated the values of irradiance and temperature of the MERRA point grid to the whole area. Then they used the BRL model to estimate the diffuse irradiance out of the irradiance MERRA data. Next, they calculated the irradiance on the photovoltaic panel and based on that simulated the power output of the panel for each MERRA point. For their simulation they first assumed an optimal position of the panel with a southwards-facing azimuth and a tilt angle depending on the latitude, but also did a simulation with non-optimal conditions, which means a random distribution of azimuth angle of 180° mean and a standard deviation of 40°, and a tilt angle of 25° mean and a standard deviation of 15°. Then they validated their simulation results once with gathered production data from 1,000 different photovoltaic systems across Europe and once with the national aggregated photovoltaic output provided by network operators. After the validation they used an annual correction factor in their simulation that adjusted the simulation results to the measured national photovoltaic output and then made a simulation of the European photovoltaic output over 30 years to investigate long-term patterns and anomalies of the production over the years. They found that results are closer to the measured generation under the non-optimal conditions mentioned above. Their results showed further that the satellite SARAH data delivers better accuracy, while the MERRA reanalysis data has a greater stability. They also found that MERRA overestimates the site output of the photovoltaic systems. They also did a seasonal comparison of the different results that can be seen in Figure 1. It shows that MERRA overestimates the production in all seasons, whereby the overestimation is highest in the winter months.

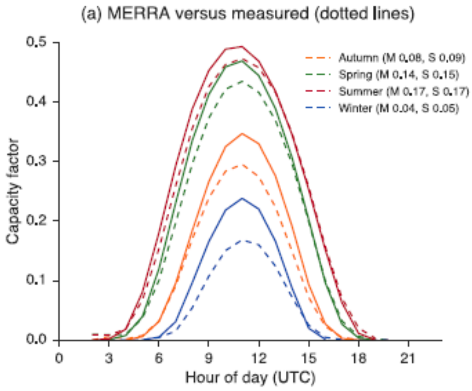


Figure 1: Simulated average hourly capacity factor for Germany and a comparison to measured data (Pfenninger & Staffell 2016a)

Another interesting aspect of the study is that it compares the old MERRA dataset with the new MERRA 2 dataset. According to the analysis the MERRA 2 data is not substantially better than the MERRA data, so they conclude that for photovoltaic output simulation there should be no big difference whether using the MERRA or the MERRA 2 data. The authors also provide their simulation data on their homepage (www.renewables.ninja). For each European country the capacity factor is available on an hourly basis from 1985 to 2014. Therefore a comparison of the simulation results in this thesis to that of Pfenninger & Staffell (2016a) will be done in chapter 4.3.

Further studies working with MERRA data do not deal directly with photovoltaic production but take a look at the accuracy of the MERRA data by comparing the solar radiation data of MERRA with in-situ measurements. Yi et al. (2011) investigated in their work the difference of certain MERRA variables to in-situ measurements and other satellite data. Amongst the investigated variables are temperature and the incident solar radiation. They found that MERRA overestimates solar radiation, especially in the middle latitudes of both hemispheres, which would apply to Germany as well. On the other side MERRA slightly underestimates maximum temperatures in the Northern hemisphere, while it overestimates it in the Southern hemisphere. Nevertheless they come to the conclusion that MERRA represents the variables appropriately. Especially the update from GEOS-4 to GEOS-5 model for the generation of the data delivers more accurate data for solar radiation and air temperature.

Boilley & Wald (2015) investigated the differences between daily solar radiation data of MERRA and qualified ground measurements from stations in Europe, Africa, and Atlantic Ocean. They found that MERRA overestimates solar irradiation due to an overestimation of the clearness index, which means that MERRA assumes a clear sky while actually it is cloudy. They suggest using MERRA data only in regions that are not cloudy, e.g. North Africa. In other regions satellite-based data should be preferred.

In conclusion the studies that have dealt so far with the topic of photovoltaic output based on MERRA data found that MERRA slightly overestimates solar radiation, among others due to an estimation of clear skies, when it is actually cloudy (Boilley & Wald 2015) and the non-consideration of local topography due to the low spatial resolution (Pfenninger & Staffell 2016a). This overestimation is higher in winter than in summer, which may be due to less sunshine in winter and connected to a higher relative error between MERRA data and measured data (Richardson & Andrews 2014).

1.2 Objectives of the thesis

The present paper aims to investigate if the MERRA 2 data is suitable for the simulation of the production output of photovoltaic systems. To answer this question a simulation of the production output is done using the example of Germany for the time span from July 2010 to July 2013. The simulated data is then statistically compared to measured generation data of the same period, to see the accuracy of the simulated data and to investigate over- or underestimations of production by the MERRA 2 data. If necessary a correction factor will be applied to the simulation results to adjust them to the measured generation, as Pfenninger & Staffell (2016a) did. The simulation is executed in the statistics software R with the software package SolaR and with the MERRA 2 dataset as input of meteorological factors. After this, the simulation will be done again for the period of 1980 to 2015 to investigate the annually development of the photovoltaic production for extreme values and anomalies. Furthermore the average hourly and daily capacity factor for all 35 years are calculated to obtain a value of possible production over a year for total Germany. The capacity factor is finally compared to the results of Pfenninger & Staffell (2016a) to derive a second validation of the simulation results.

The following research questions will be answered by the present master thesis:

- Is the MERRA 2 dataset a suitable basis for the simulation of the potential production of photovoltaic systems in Germany?
- How big are the differences between simulated and measured data?
- Are there bigger differences between simulated data and measured generation data over the year depending on the month or season?
- Which extreme values and anomalies occur over the 35 years on an average hourly and daily basis?
- How big are the differences of the simulation compared to the results of Pfenninger & Staffell (2016a)?

Chapter 2 gives an insight into the investigation area Germany, its climatic preconditions for photovoltaic production, and its current situation concerning the electricity production of photovoltaic systems. Chapter 3 describes the data that was used for the simulation. Besides the MERRA 2 data set also data on the installed capacity of photovoltaic systems in Germany and the measured generation data of electricity by photovoltaics has been used . Furthermore a detailed explanation of the processing of the data and of the simulation is given. Chapter 4 presents the results of the simulation and the comparison of these results to the measured generation data.

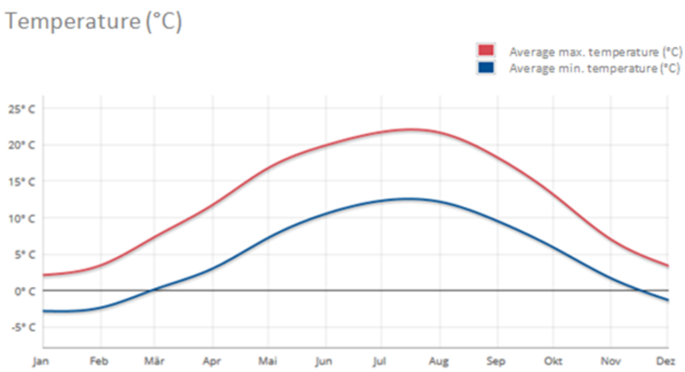
Furthermore chapter 4 shows the results for the 35 year simulation and the comparison of the average capacity factor of these 35 years to Pfenninger & Staffell (2016a). Chapter 5 and 6 close with a discussion of the results and a summary of the thesis.

2 Investigation area - Germany

The area of interest is Germany. It is located between the longitude of 6° east to 15° east and the latitude of 47° north to 55° north (Maps of World 2016). It is situated in the temperate climate zone, which is characteristic for its moderate temperatures and has four seasons, in comparison to tropical countries at the equator that have no seasons. The seasons will be relevant for the results of the production as in winter there is less solar radiation and a lower temperature and thus less production of photovoltaic systems.

2.1.1 Temperature, sunshine and solar radiation

Figure 2 shows the average maximum and minimum temperature per month for Germany. Here one can see the characteristics of the seasons. While in winter (December – February) temperature is very low and can be beyond 0 °C, temperature rises in spring (March – May) and reaches its top in summer (June – August) at 22 °C before it starts to fall in autumn (September – November).



Average max. temperature (°C)

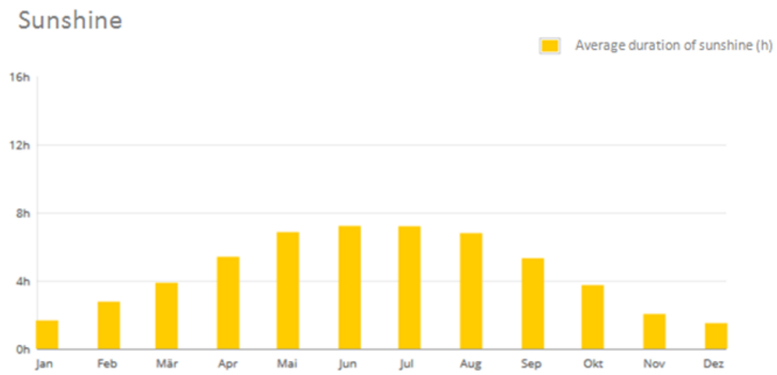
Jan	Feb	Mär	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez	Year
2,1	3,5	7,4	11,7	16,8	20	21,8	21,7	18,3	13,2	7	3,4	12,2

Average min. temperature (°C)

Jan	Feb	Mär	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez	Year
-2,8	-2,4	0,2	3	7,2	10,5	12,3	12,2	9,6	5,9	1,7	-1,3	4,7

Figure 2: Average monthly temperature in Germany (Wetter.de 2016, database: Deutscher Wetterdienst)

In the same way as temperature falls and rises over the year also the duration of sunshine develops in a similar pattern. Figure 3 shows the average hours of sunshine per day for the months of a year. It is obvious that there is most sunshine in summer with up to almost 8 hours, while in winter the duration of sunshine drops down to about 2 hours per day. This implies that the production of power by photovoltaic systems must be higher in summer than in winter.



Average duration of sunshine (h)

Jan	Feb	Mär	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez	Year
1,6	2,7	3,8	5,3	6,8	7,1	7,1	6,7	5,2	3,7	2	1,4	4,5

Figure 3: Average monthly duration of sunshine in Germany (Wetter.de 2016, database: Deutscher Wetterdienst)

Besides temperature and sunshine duration, the global radiation is an important factor determining the production potential of a photovoltaic system. As already mentioned at the beginning, the output of a photovoltaic system depends on its geographic position. As solar radiation is highest at the equator it slowly drops the closer it gets to the poles. The annual global radiation on a horizontal surface in South Germany is about 1,100 kWh/m^2 while in North Germany it is around 1,000 kWh/m^2. The seasonal global radiation varies in South Germany from 1 kWh/m^2 per day in January to 3.5 kWh/m^2 in March to 5 kWh/m^2 in June (Zahoransly et al. 2012). Figure 4 shows the average annual global horizontal irradiation for Germany. According to the map we can assume that production of photovoltaic power in North Germany will be a little less than in South Germany. This is on one side due to South Germany’s closer proximity to the equator and on the other side due to the higher altitude in the South because of the Alps. So due to its seasons and its spatial extent Germany presents an interesting investigation area for the research questions of this paper.

Figure 4: Average annual sum of Global Horizontal Irradiation in Germany (SolarGis 2016)

2.1.2 Electricity market

For the analysis in this thesis, data of the electricity production of photovoltaics was taken from official data provided by the German electricity market zone operators. This market consists of four grid zones named Amprion, Hertz50, Tennet, and Transnet BW. The spatial extent of the four grid zones can be seen in Figure 5. The reason why Germany was picked as investigation area is that there is measured generation data for the production of photovoltaic systems available for all four grid zones. The data is available from 19th July 2010 until 15th July 2013 for every quarter of an hour.

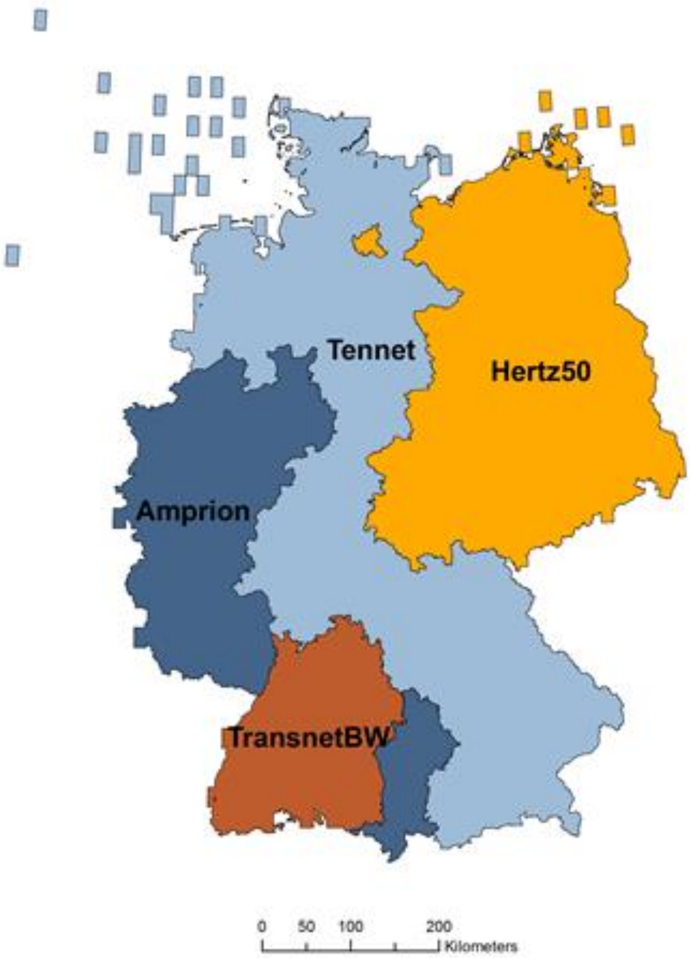


Figure 5: Spatial extent of the four German grid zones (Own representation)

2.1.3 Photovoltaics

Althouh Germany has not the best preconditions for solar energy production due to its location and climate, it is one of the biggest producers in Europe. In 2013 the highest installed capacity of photovoltaic systems in Europe was in Germany with about 36 GWp followed by Italy with 17.5 GWp, which is almost the half of the installed capacity in Germany. This means that in Germany there is an installed capacity of 447 Wp/inhabitant. Compared to that Austria reached a level of 82 Wp/inhabitant in 2013 (Buddensiek 2014).

Figure 6 shows the development of installed power and electricity production from photovoltaics in Germany from 1990 to 2013. One can see that the construction of photovoltaic systems became popular at the beginning of 2003 and highly increased over the years from about 1 GW installed power in 2003 to 39.8 GW installed power in 2015. Connected to this also the production of electricity by photovoltaic systems increased from about 1 TWh in 2003 to around 38.7 TWh in 2015 (BMWi 2017).

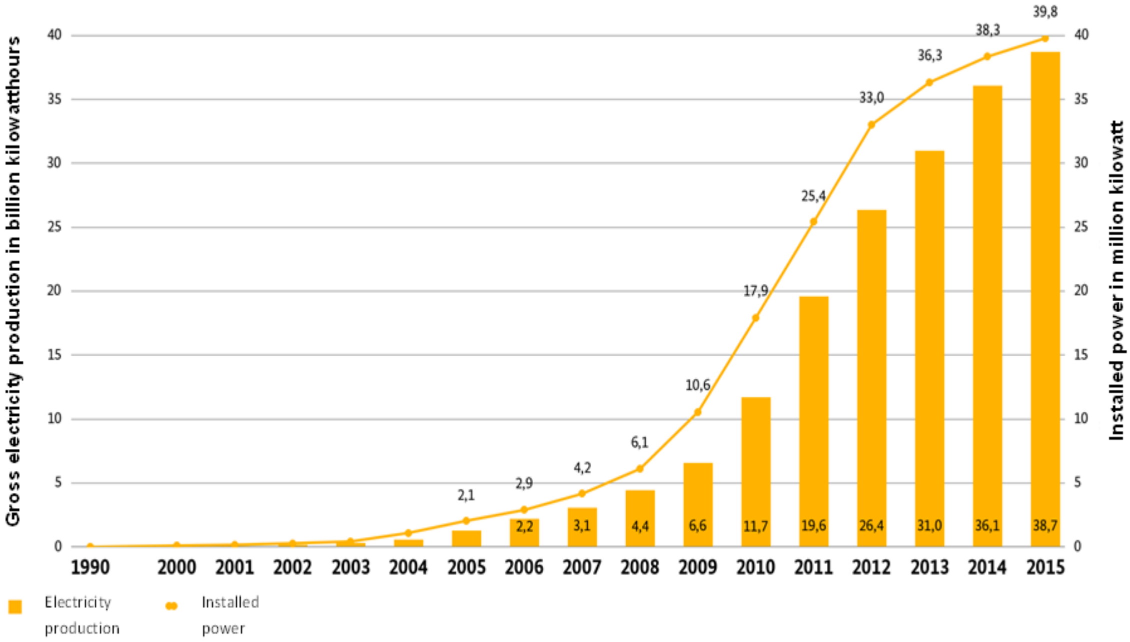


Figure 6: Development of electricity production and installed power of photovoltaic systems in Germany 1990 – 2013 (BMWi 2017)

3 Methods and Data

The following chapter describes the data and the methods used for the simulation of the photovoltaic production and the analysis and comparison of the results to measured generation data. An overview of the working process is given in Figure 7:

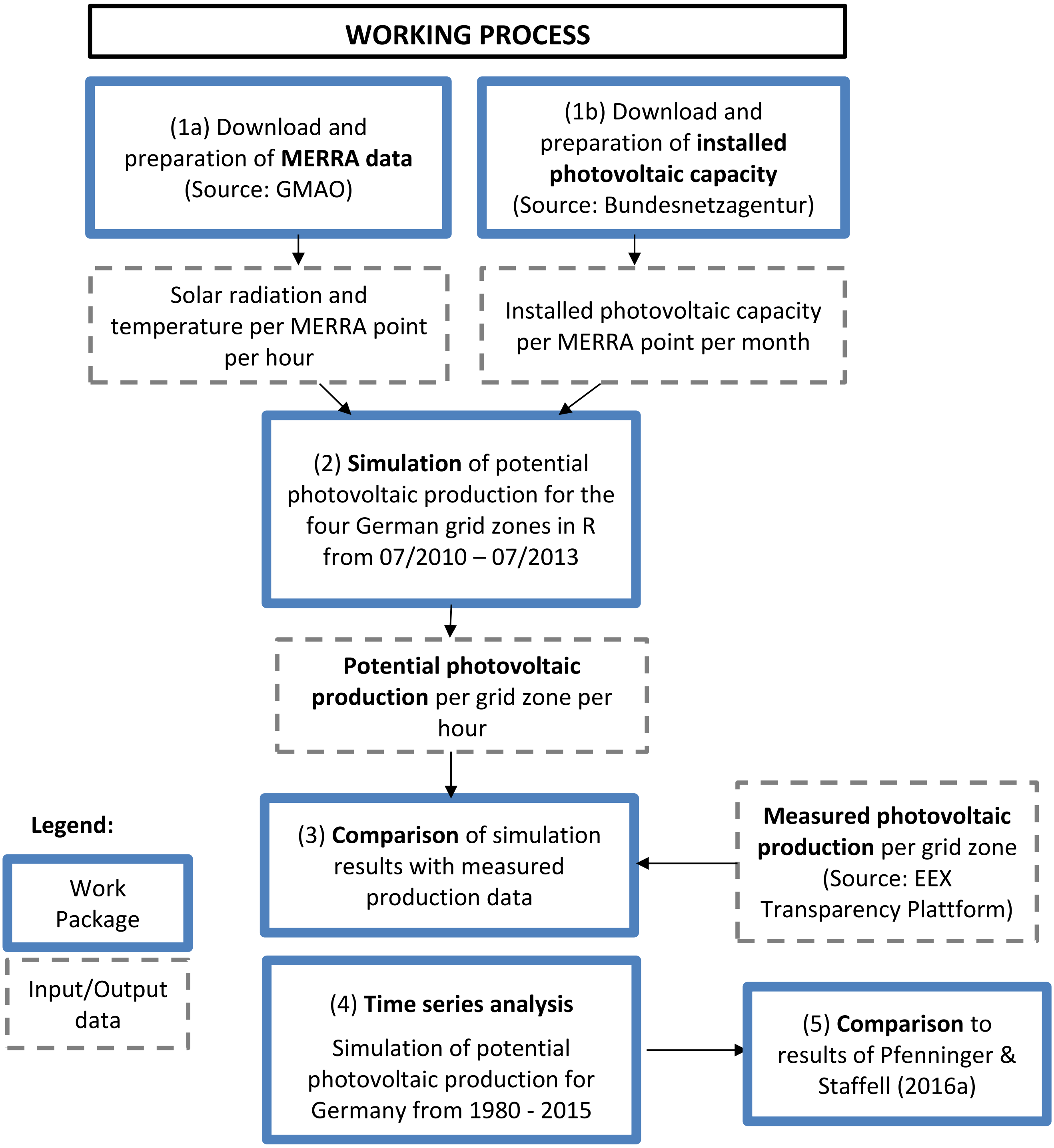


Figure 7: Working process of the simulation, comparison to measured data and time series analysis (Own representation)

The first objective of this thesis is to investigate if MERRA data is a potential data source for the simulation of photovoltaic production. Therefore in work package **(1a)** the MERRA 2 data that contains information on temperature and solar radiation, was downloaded from the website of the Goddard Earth Sciences Data and Information Services Center and prepared in R for further processing. Details on MERRA data are provided in chapter 3.1.1, the preparation is explained in chapter 3.2.1. At the same time in work package **(1b)** the installed photovoltaic capacity in Germany, that is provided by the Bundesnetzagentur in an Excel file, is prepared so that for each month the complete installed capacity per municipality is given. The municipalities and their capacity are then connected in ArcGIS to the next MERRA point. Details on the installed capacity and the connection to the MERRA points are described in chapter 3.1.2. Based on the MERRA data and the installed capacity of 1a and 1b, in work package **(2)** the simulation of the photovoltaic production for the four grid zones is done in R with the package solaR (see chapter 3.2.1 for a detailed explanation). As a result we get the potential photovoltaic production for each grid zone per hour per day. Work package **(3)** then deals with the comparison of the simulation results to measured generation data, as described in chapter 3.1.3. The method of comparison is described in subsection 3.2.2.

Afterwards the second objective of the thesis is covered by a time series analysis of the simulation results work package **(4)** to see if there are peak periods of very high or very low photovoltaic production. The method is described in 3.2.4. Furthermore the results are compared to Pfenninger & Staffell (2016a) in work package **(5)**.

3.1 Used data sets

The following sub chapters describe the three data sets that were used for the analysis in this thesis. First the MERRA 2 data, which provides, among many other variables, hourly time series of temperature and solar radiation for the whole world and is available from 1980 to 2015 (chapter 3.1.1), second the data on the installed capacity of photovoltaic systems in Germany that is available from 2009 to 2015 on a monthly basis (chapter 3.1.2) and third the measured generation data of photovoltaic production in Germany from July 2010 to July 2013 on a 15-minute basis (chapter 3.1.3). Due to the different length of the available timeseries, the simulation and comparison to measured generation data is done from July 2010 to July 2013.

3.1.1 MERRA data

The Modern Era-Retrospective analysis for Research and Applications (MERRA) is a dataset provided by NASA's Global Modeling and Assimilation Office (GMAO) and includes numerous climate variables like temperature, solar radiation, wind speed, precipitation, and so on. The reanalysis data is generated from historical data with the Goddard Earth Observing System Data Assimilation System Version 5 (GEOS-5). The historical data is often irregularly spatially distributed and therefore interpolated into a regular worldwide point grid. So it is transformed into a dataset that is temporally and spatially consistent (Pawson 2012). Since the beginning of 2016 a new dataset named "MERRA 2" is available. Compared to the old MERRA data, the projected grid and the method of assimilation are slightly different. The grid of the old MERRA data has a resolution of 0.667° longitude and 0.5 ° latitude (Lucchesi 2012) while the MERRA 2 grid was slightly changed to a resolution of 0.625° longitude and 0.5° latitude (Bosilovich, Lucchesi, and Suarez 2016). The MERRA data spans a time from 1979 until the beginning of 2016, while MERRA 2 spans a time from 1980 until today. The other changes from MERRA to MERRA 2 do not relate to the data sets for temperature and solar radiation (Bosilovich, Lucchesi, and Suarez 2016). As Pfenninger & Staffell (2016a) showed in their study, there is no big difference between the MERRA and the MERRA 2 data when using them for the simulation of photovoltaic production. So for the present thesis the MERRA 2 data was used. The MERRA 2 data can be downloaded under https://disc.sci.gsfc.nasa.gov/ datasets?page=2&keywords =merra-2 or https://disc.gsfc.nasa.gov/ daac-bin/FTPSubset2.pl?LOOKUPID List=M2T1NXRAD (GMAO 2015).

MERRA data is point data, where every point contains the information on the climate variables at this location. Figure 8 shows the distribution of the 399 selected MERRA points over Germany.

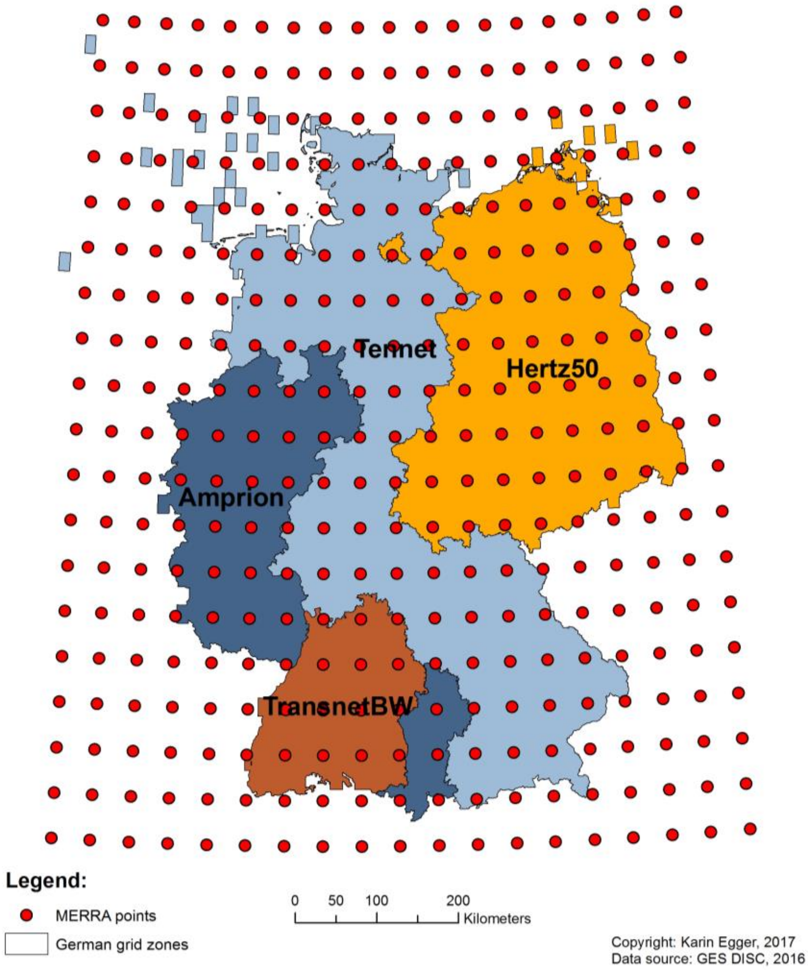


Figure 8: Distribution of MERRA points over Germany (Own representation)

The MERRA data is constructed by several sub-datasets that contain different climate variables. For the simulation of the potential photovoltaic production the two climate variables surface skin temperature (TS) and solar radiation (SWGDN) are used. These two variables have also been used by Richardson & Andrews (2014) for the simulation of photovoltaic production. The solar radiation (SWGDN) data represents the ground-level global irradiance (Pfenninger & Staffell 2016a). Both variables are part of the sub-dataset tavg1_2d_rad_Nx. So for MERRA 2 the data product "IAU 2d surface and TOA radiation fluxes (tavg1_2d_rad_Nx)" was chosen.

The spatial extent of the data was limited to the coordinates of Germany (West: 4°, North: 56°, South: 46°, East 16°). Due to the rectangle frame of the data selection the download also included MERRA points that are not within Germany. This must be considered in the analysis later on. The beginning of the dataset is set to 01. January

1980. The end date is set to 31. December 2015. The data was downloaded as NetCDF4 files. To speed up the download, a software named WGET was used for the automatic download of the files. The list of URLs was fed to WGET and the program downloaded the single files in a folder on the computer. The further processing of the data in the software R is described in chapter 3.2.1.

3.1.2 Installed capacity of photovoltaic systems

The installed capacity of photovoltaic systems in Germany is provided by the German Bundesnetzagentur and can be downloaded on their website (Bundesnetzagentur 2016). Since the beginning of 2009 the Bundesnetzagentur collects data of newly installed photovoltaic systems per postal code per month. The datasets of the newly installed capacity were downloaded for the period from January 2009 to December 2013 and then joined and cumulated in one file, so that for each postal code the total installed capacity per month was given. The capacity in the files is given in kilowatt peak (kWp). An installed capacity of 6.45 GWp, that was installed before 2009 and therefore is not included in the Bundesnetzagentur data, was distributed equally over all German postal codes as no concrete data of the distribution was available. Table 1 shows how much capacity was installed at the end of each year in the single grid zones and in total Germany according to the data of the Bundesnetzagentur. This data matches with Figure 6 of chapter 2.

Table 1: Total installed photovoltaic capacity in the four grid zones in total Germany from 2008 – 2013 in GWp (Own representation, Source: Bundesnetzagentur 2016)

	2008**	2009	2010	2011	2012	2013
Amprion	1.49	2.40	4.20	5.73	7.20	8.00
Hertz50	1.22	1.70	2.82	5.13	7.96	8.86
Tennet	2.80	4.66	8.12	10.93	13.57	14.81
TransnetBW	0.94	1.49	2.49	3.33	3.99	4.35
Germany	**6.45**	**10.25**	**17.63**	**25.12**	**32.72**	**36.02**

** For the years before 2009 only the total installed capacity of 6.45 GW is known, so that it was distributed equally to the postal codes.

Next the German postal codes, that are also available as shapefile, where connected to the next MERRA point via a NEAR analysis in the software ArcGIS. This is necessary as in the simulation the potential production will be calculated per MERRA point. So in the simulation we get for each MERRA point a potential photovoltaic production as if 1 kWp photovoltaic system was installed at this point. This production result must then be multiplied with the installed capacity next to this MERRA point. To connect the postal codes in ArcGIS to the MERRA points, the coordinates of the MERRA points are needed. Therefore one MERRA file has been opened in the software R and the longitude and latitude of the MERRA points were read out and saved into a text file. The R-code for this process is as follows:

```
> install.packages("ncdf4")
> library(ncdf4)
> setwd("Folder")
> ncfile <- nc_open("svc_MERRA2_400.tavg1_2d_rad_Nx.
20110101.nc4")

> lon <- ncvar_get(ncfile, "lon", verbose=FALSE)
> lat <- ncvar_get(ncfile, "lat", verbose=FALSE)
> lonlat <- expand.grid(lon,lat)
> colnames(lonlat) <- c("Lon","Lat")
> write.table(lonlat,"Folder/variables_lonlat_MERRA2.txt"
, sep="\t")
```

The text file with the coordinates was imported into ArcGIS. As the simulation is done on the basis of the four German grid zones and therefore the capacity per MERRA point per grid zone is needed, we first need to assign the postal codes to one of the four grid zones by a NEAR analysis. Afterwards all postal codes of one grid zone were assigned to one MERRA point again by a NEAR analysis. The resulting table, that contains the postal codes, their grid zone, and the nearest MERRA point was then exported and connected with the file containing the installed capacity per postal code per month in Excel. The installed capacity was summed up per MERRA point for each grid zone and separately stored. So in the end there were four files, each containing the installed capacity per MERRA point for one grid zone. The files were then imported into R to be used for the simulation of the photovoltaic production. The distribution of the installed capacity over Germany and its development from 2010 to 2013 can be seen in Figure 9. The capacities of the single postal codes where combined to regions in a grid of 300 km^2 cells.

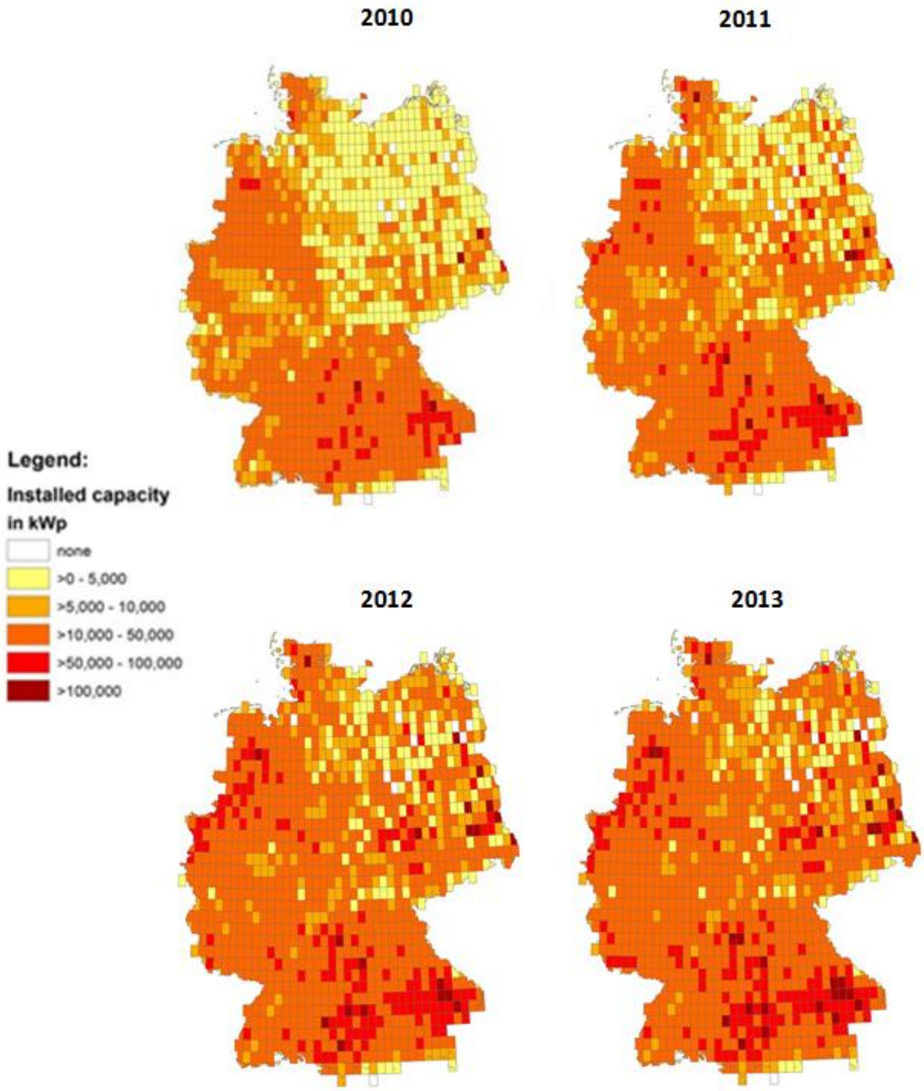


Figure 9: Installed photovoltaic capacity in Germany from 2010 - 2013 (Own representation, Data source: Bundesnetzagentur 2016)

The highest installed capacity per km^2 is in the South of Germany. In 2010 some regions in the South had an installed capacity of over 100,000 kWp while in the North East most of the regions had less than 5,000 kWp installed capacity. Up to 2013 the North East developed its installed capacity, so that most of the regions reach at least a level of 10,000 to 50,000 kWp of installed capacity, while in the South the installed capacity also grew, so that more regions reach a installed capacity over 100,000 kWp. As seen in Figure 4 of chapter 2.1.1 the South is the region with the highest solar radiation, so that it makes sense that the installed capacity is highest there.

3.1.3 Measured generation data of photovoltaic production

The measured generation data was provided by the EEX Transparency platform (https://www.eex-transparency.com). The given data shows the average actual performance of the photovoltaic systems in MW in 15 minute intervals from the 19. July 2010 until the 15. July 2013 as sum per grid zone. As the MERRA data for the simulation of the potential photovoltaic production is given in an hourly period, the measured generation data was aggregated to an hourly base. Furthermore, to use the data for the comparison with the simulation results, it must be transformed from power in MW to electricity production in MWh. Therefore the 15 minute intervals have been summed up and then divided by 4.

Table 2 shows the electricity production per grid zone and per year in TWh according to the data of the EEX Transparency platform. The year 2010 started at 19. July 2010, so the amount produced in this year (4,20 TWh) is not as high as in the following years as it only represents half of the year 2010. In 2011 there was a production of 18.56 TWh and in 2012 27.94 TWh. The 2013 data only lasts until the 15. July 2013, so just half of the annually production is given in the table with 16.83 TWh. This information matches almost with the data provided by the BMWi (2017), see chapter 2. There the electricity production in TWh for 2011 was 19.6 TWh and for 2012 26.4 TWh. As 2010 and 2013 in the table below are only half of the year they cannot be compared directly to the data of the BMWi (2017).

Table 2: Electricity production of four grid zones and total Germany from 07/2010 to July 2013 in TWh (Own representation, Source: EEX Transperency Platform)

	2010	2011	2012	2013
Amprion	1.06	4.67	6.75	4.00
Hertz50	0.42	2.38	5.13	3.54
Tennet	1.97	8.62	11.53	6.74
TransnetBW	0.76	2.89	4.54	2.56
Germany	**4.20**	**18.56**	**27.94**	**16.83**

Figure 10 illustrates the development of the daily electricity production from 19th July 2010 to 15th July 2013. The months with the highest production are May to August, while in the months September to February production goes down.

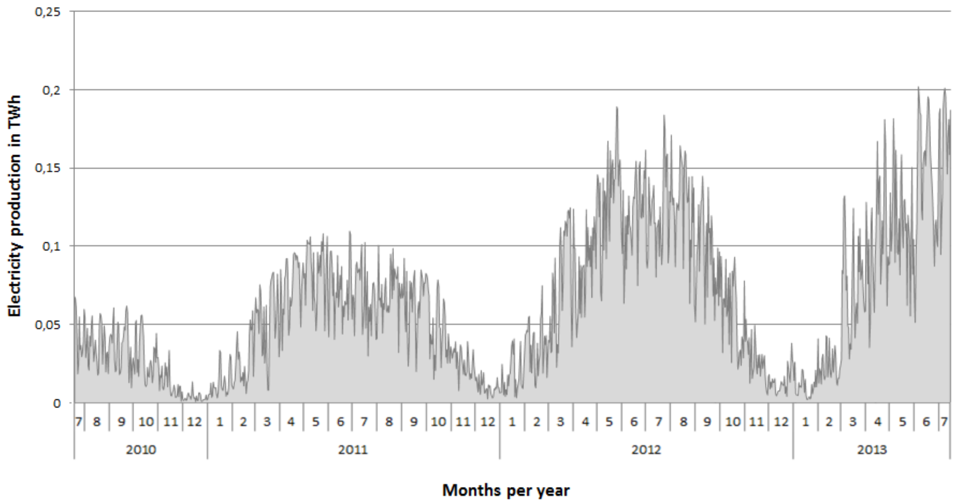


Figure 10: Daily electricity production from photovoltaics in Germany 2010 – 2013 (Own representation, Data Source: EEX Transparency Platform)

The next figure shows the photovoltaic production for total Germany and the 4 grid zones on a monthly base. The highest production of the 4 grid zones takes place in the zone Tennet, while the production in TransnetBW is lowest. Also the above mentioned monthly fluctuations can be seen in the table. While the production in December lies beyond 0.5 TWh for total Germany it goes up to above 4 TWh, e.g. in May 2012 and June 2013.

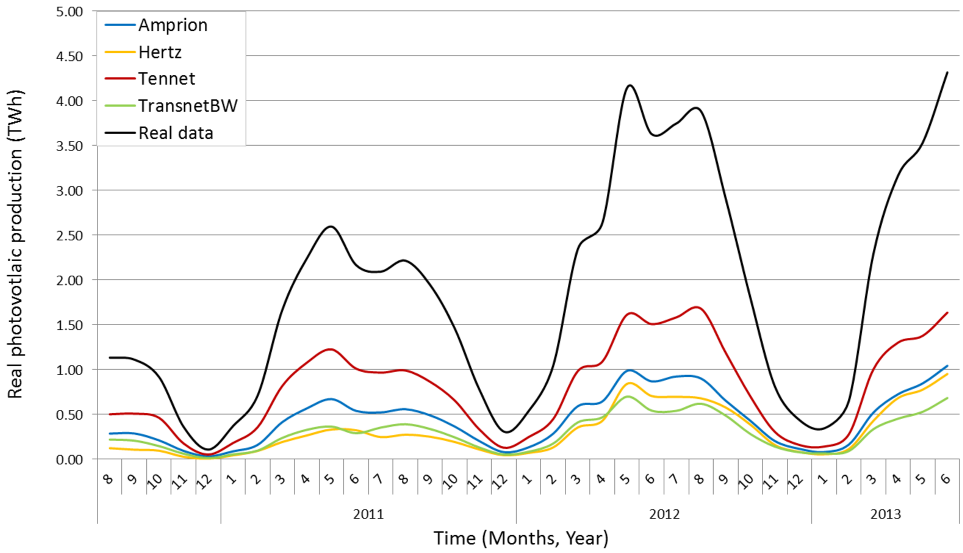


Figure 11: Monthly photovoltaic production for Germany and its four grid zones 2010 – 2013 (Own represenation, Data source: EEX Transparency Platform)

3.2 Methods of analysis

After the description of the used data in chapter 3.1, the next chapter gives an insight into the simulation and analysis of the photovoltaic production. The simulation in the statistics software R is described in chapter 3.2.1. Chapter 3.2.2 explains the correction factor that was applied to the simulation results, so that the results get more accurate in comparison to the measured production. Chapter 3.2.3 then describes how the measured generation data is compared to the simulation results in Excel while in chapter 3.2.4 the procedure of the time series analysis for the period from 1980 to 2015 is explained.

3.2.1 Simulation of photovoltaic production with solaR

Figure 12 shows the rough process of the simulation in the statistics software R. First the MERRA data must be imported and transformed to a format that is useable for the analysis. The import is done with the R-package NCDF4. Therefore two functions are created that read out the solar radiation and the temperature of a single MERRA file (step 1). One file contains the data of one day for all MERRA points.

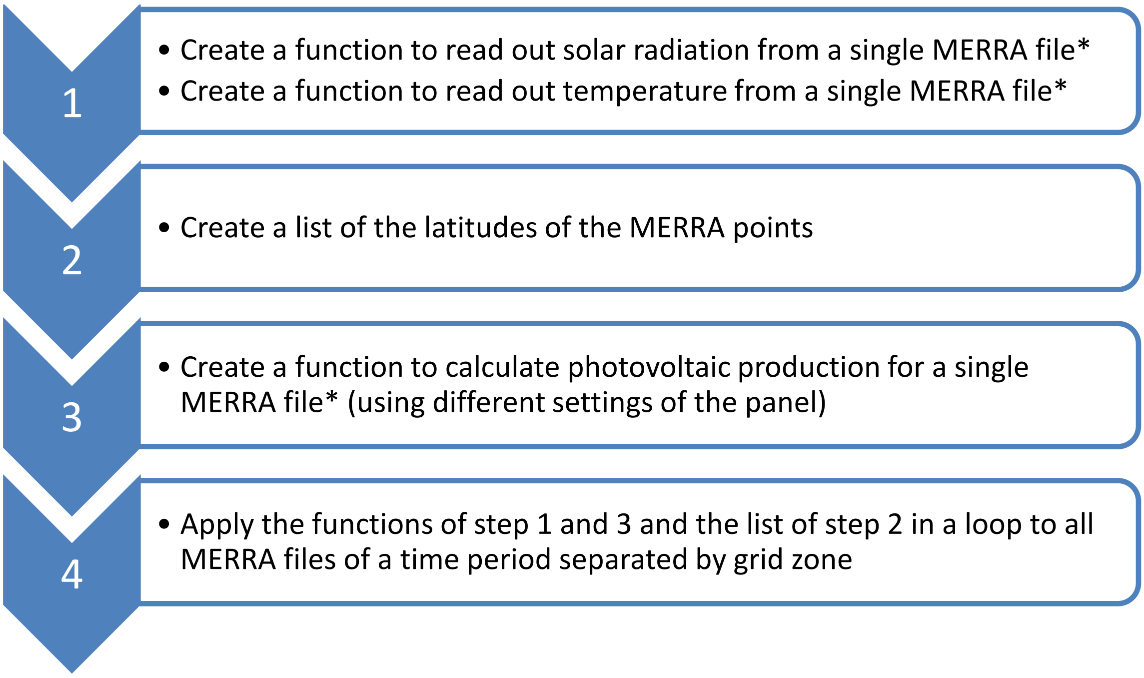


Figure 12: Process of simulation of photovoltaic production in the statistics software R (Own representation)
* A single MERRA file contains the data of one day for alle MERRA points

In step (2) a list of the latitutde of all MERRA points is made. The list doesn't need to be created over a function as it is the same for all MERRA files and therefore the same list can be applied to all files. In step (3) a function is created that calculates the photovoltaic production for a single MERRA file, thus for one day, under different settings. Later on in step (4) the above mentioned functions are used in a loop to apply them to all files over the whole time period and separated by the for grid zones. For step (3) and (4) the R-package solaR is used. The package was created by Perpiñán Lamigueiro (2012). It consists of different classes, methods and functions that help to calculate the geometry of the sun and the resulting direct and diffuse solar radiation at a certain location. Furthermore one can calculate the output of photovoltaic systems based on the geometry and calculated solar radiation of the package or by inserting already existing meteorological data of daily or intradaily solar radiation and temperature. For the calculation of the photovoltaic production the different settings of the photovoltaic panel, e.g. inclination and orientation of the panel, can be adjusted in the package. More information on that is given in the description of the code below. The present thesis uses the MERRA 2 data as input for meteorological data of solar radiation and temperature into solaR.

The described code in this chapter shows the simulation separated for the four grid zones on a monthly basis from July 2010 to July 2013. The results of the single zones are summed up afterwards to obtain the total production for Germany. If the code is needed for the simulation of just one country, the code must be adapted accordingly. The codes, which must be entered in R, are written in blue. Parts of the code that must be individually adapted are marked in red. The results that are given by R are grey. The complete code as entered in R can also be found in the appendix VI.1.

PREPARATION

Before starting to load the MERRA data into R and format it into the right form for the simulation, the necessary packages need to be installed. The first package (NCDF4) is necessary to read out the MERRA files, which are in a NCDF-format. The second one (solaR) is for the simulation of the photovoltaic production. Afterwards the packages must be started with the command library.

```
> install.packages("ncdf4")
> install.packages("solaR")
> library(ncdf4)
> library(solaR)
```

Next one MERRA file is opened to see the content of the file. Therefore the working directory is set to one folder that contains MERRA files. The NCDF4-package offers a code to open NCDF-files and read the data that is stored in these files. Therefore one of the files is selected and it's name is put into the code:

```
> ncfile <- nc_open("svc_MERRA2_400.tavg1_2d_rad_Nx.
20110101.nc4")
> print(ncfile)
```

As a result, the following information is shown:

```
File svc_MERRA2_400.tavg1_2d_rad_Nx.20110101.nc4
(NC_FORMAT_NETCDF4):

2 variables (excluding dimension variables):
     float SWGDN[lon,lat,time]    (Chunking: [19,21,1])
     (Compression: shuffle,level 1)
     standard_name: surface_incoming_shortwave_flux
            long_name: surface_incoming_shortwave_flux
            units: W m-2
            _FillValue: 999999986991104
            missing_value: 999999986991104
            fmissing_value: 999999986991104
            vmax: 999999986991104
            vmin: -999999986991104

     float TS[lon,lat,time]    (Chunking: [19,21,1])
     (Compression: shuffle,level 1)
     standard_name: surface_skin_temperature
            long_name: surface_skin_temperature
            units: K
            _FillValue: 999999986991104
            missing_value: 999999986991104
            fmissing_value: 999999986991104
            vmax: 999999986991104
            vmin: -999999986991104
```

```
3 dimensions:
        time   Size:24   *** is unlimited ***
            standard_name: time
            long_name: time
            units: minutes since 2011-01-01 00:30:00
            calendar: standard
        lat  Size:21
            standard_name: latitude
            long_name: latitude
            units: degrees_north
            axis: Y
        lon  Size:19
            standard_name: longitude
            long_name: longitude
            units: degrees_east
            axis: X
```

From this data it can bee seen that the file includes two variables, namely the *surface incoming shortwave flux*, which has the shortcut SWGDN and is given in W/m^2, and the *temperature*, which is named TS and given in Kelvin K. Furthermore it can be seen that the variables are projected on three dimensions that are time, longitude, and latitude. Each MERRA file includes one day of data. For this day the variables for all 24 hours are given. Depending on the spatial size of the file one or more MERRA points that are defined by their longitude and latitude are included. In the following case 19 values of longitude and 21 values of latitude are given. The multiplication of these values results in 399 MERRA points. So the file includes the solar radiation and temperature for each hour of 1. January 2011 for 399 MERRA points. For the distribution of these points see chapter 3.1.1.

STEP 1: READING FUNCTIONS

As now the data and the shortcuts of the variables and dimensions are, functions are created to read out the data of one file and save them as single variables. To begin with, a function is created for reading the temperature out of the file and transform it into a matrix with the values of temperature. The rows of the matrix are the 399 MERRA points and the columns are the time in hours, thus 24 columns.

```
> readts <- function(ncfile) {

    ncfile <- nc_open(ncfile)
    time <- ncvar_get(ncfile, "time", verbose=FALSE)
    ntime <- dim(time)
    lon <- ncvar_get(ncfile, "lon", verbose=FALSE)
    nlon <-dim(lon)
    lat <- ncvar_get(ncfile, "lat", verbose=FALSE)
    nlat <- dim(lat)
    ts.array <- ncvar_get(ncfile, varid="TS",
    verbose=FALSE)
    ts.vec.long <- as.vector(ts.array)
    ts.mat <-matrix(ts.vec.long, nrow=nlon *
    nlat,ncol=ntime)
    nc_close(ncfile)
  }
```

The same is done for the variable solar radiation:

```
> readswgdn <- function(ncfile) {
    ncfile <- nc_open(ncfile)
    time <- ncvar_get(ncfile, "time", verbose=FALSE)
    ntime <- dim(time)
    lon <- ncvar_get(ncfile, "lon", verbose=FALSE)
    nlon <-dim(lon)
    lat <- ncvar_get(ncfile, "lat", verbose=FALSE)
    nlat <- dim(lat)
    swgdn.array <- ncvar_get(ncfile, varid="SWGDN",
    verbose=FALSE)
    swgdn.vec.long <- as.vector(swgdn.array)
    swgdn.mat <-matrix(swgdn.vec.long, nrow=nlon *
    nlat,ncol=ntime)
    nc_close(ncfile)
  }
```

STEP 2: LIST OF LATITUDE

Next a list is created that contains the longitude and latitude of the single MERRA points. As all files contain the same MERRA points, this can be done with a single file and must not be applied to all files. The list will be used in the simulation later on.

```
> lon <- ncvar_get(ncfile, "lon", verbose=FALSE)
> lat <- ncvar_get(ncfile, "lat", verbose=FALSE)
> lonlat <- expand.grid(lon,lat)
> colnames(lonlat) <- c("Lon","Lat")
>head(lonlat)
```

STEP 3: SIMULATION OF PHOTOVOLTAIC PRODUCTION WITH PACKAGE SOLAR

Next a function is created to simulate the photovoltaic production with the data of a single MERRA file, thus with the data of one day for all MERRA points. Therefore commands of the R-package SolaR are used. This is a package that executes the whole calculation for a certain location from the solar angles, daily solar radiation and temperature up to the final productivity of a photovoltaic system for a certain location just by entering the latitude and time of this location. But the package can also be used with already existing data on solar radiation and temperature. Therefore it offers the command dfI2Meteo, which imports meteorological data. So by entering the variables solar radiation and temperature from the MERRA data, as well as the latitude of the points, the package calculates the potential photovoltaic production at this point. For this calculation the command prodGCPV is used (Perpiñán Lamigueiro 2016). Both commands, the dfI2Meteo and the prodGCPV, are combined in a new function named convertToPv. This function takes the MERRA data and calculates the photovoltaic production in kWh per MERRA point if a photovoltaic panel with 1 kWp would be installed at this place. In the loop afterwards the production of the single points is multiplied with the installed capacity around this point. To run the function convertToPV four main variables are necessary that will be inserted in the loop in step (4), namely the latitude of the MERRA point, the date and hour, the solar radiation and the temperature. Furthermore additional information as the average inclination of the panels, "beta" in solaR, or the average azimuth of the panels, "alfa" in SolaR, can be entered. For the following simulation these two variables are included as well. Another example for further variables would be the degree of dirtiness that is set to an average value in the default values. The list of latitude has been set out in Step (2), while the date and hours will be set out in a list in the loop in Step (4). The temperature and solar radiation have also been read out before in Step (1). The temperature must be subtracted with 273.15 to convert it from Kelvin (K) to Celsius degree (°C) to be used in solaR. The solar radiation is taken and transferred into direct and diffuse solar radiation with the function dfI2Meteo. The columns of the table are renamed to time ("Time"), solar radiation ("G0") and temperature in °C ("Ta"). The variable names are required by SolaR. The command prodGCPV then calculates different parameters regarding the productivity of photovoltaic panels, also including the potential photovoltaic production. Here the settings for the panel are included. This settings are adjusted in the loop in step (4). If no information is entered in the function, the package uses its default values. In the code developed in this thesis, the latitude (lat) can be entered into the calculation as well as the meteorological data through dfI2Meteo. But nor the average inclination angle of the panels neither the average azimuth angle is known. The default value for the inclination angle, named beta, is the latitude of the location minus 10, which is considered to maximize output of the system over the year. So if there is a latitude of 41, the system uses an inclination angle of 31. The default value for the azimuth angle, which is the direction the panel is facing, is set to South (alfa = 0). As Pfenninger & Staffell (2016a) found, the tilt angle and also the direction the panel is facing, may play a huge role in the presentation of the total annual power production and also on the development of the production over the day. As the default values may not represent the real average inclination and direction of the panels, as in practice the position and the tilt of the panel depend highly on the roof and the direction of the house, the simulation was also run with different settings. Besides the default values, also the production for 25° inclination and South (0° orientation), 25° and South West (45°), 20° and South (0°), 20° and South West (45°), 15° and South (0°), and 15° and South West (45°) were calculated. The comparison of the different settings will be given in chapter 4.1. To select the photovoltaic production of the produced table the command p@prodI[,8] is used. The result is given for a 25kWp panel in W. So at the end of the calculation the result must be divided by 25,000 to obtain the kWh production for a 1 kWp panel.

```
> convertToPv <-function(lat,datumCET,irradiation,
temperature,inclination,azimuth)
{ P1 <-cbind(datumCET,irradiation,temperature-273.15)
    colnames(P1) <- c("Time","G0", "Ta")
    irradiation <- dfI2Meteo(P1,time.col = "Time", lat =
   lat, source = "P1", format ='%Y/%m/%d %H:%M:%S')
   p <- prodGCPV(lat,modeRad = "bdI",dataRad =
irradiation, sample ="hour",beta=inclination,
alfa=azimuth)
   p@prodI[is.na(p@prodI[,8]),8] <- 0
    final_pv <- p@prodI[,8]/25000 #kwH

  return(final_pv) }
```

STEP 4: APPLYING FUNCTIONS IN A LOOP TO ALL MERRA FILES

Next the functions for a single MERRA file from step (1) – (3) are applied on all MERRA files and make a separate calculation for all grid zones. As the simulation only yields the production as if 1 kWp would be installed at the MERRA points we furthermore multiply this production by the installed capacity of the grid zones. The first task in the loop is to set the time period for which the calculation should be made, by entering the years and the months. Furthermore the grid zones that should be included in the calculation must be added with the shortcuts (Amprion = amp, Hertz50 = her, Tennet = ten, TransnetBW = tra). Then the working directory must be set to the folder of the MERRA files.

```
for(year in 2011:2012){
  for(month in c("01","02","03","04","05","06",″07″,″08″,
  ″09″,″10″,″11″,″12″)){
      for(zone in c("amp","her","ten","tra")){
setwd(paste("Folder",year,"/", month,sep=""))
```

The created reading functions are applied to all MERRA files that are in the folder. Therefore a list of all these files is created. Afterwards the functions of step (1) are applied to this list of files. So the variables temperature and solar radiation are read out from all files. The results are a lot of single matrices, where one matrix represents the variables for one day.

```
> allfiles <- list.files(pattern = "*.nc")
> list.of.ts <- lapply(allfiles,readts)
> list.of.swgdn <- lapply(allfiles, readswgdn)
> listlength<-length(allfiles)
```

After reading out all files, a data frame for the solar radiation and one for the temperature is created, so that all single matrices are stored in one file together. Furthermore a row is put at the beginning that holds the date and time.

```
> df.swgdn <- data.frame(matrix(unlist(list.of.swgdn),
nrow=listlength*24, byrow=T))
> colnames(df.swgdn)<-paste("MP",1:399,sep="")
> seq <-seq(as.POSIXct(paste(year,"-01-01
00:30",sep="")),as.POSIXct(paste(year,"-12-31
23:30",sep="")),"hours")
```

```
rownames(df.swgdn)<-seq[month(seq)==as.numeric(month)]

> df.ts <- data.frame(matrix(unlist(list.of.ts),
nrow=listlength*24, byrow=T))
> colnames(df.ts)<-paste("MP",1:399,sep="")
> rownames(df.ts)<-seq[month(seq)==as.numeric(month)]
```

The final two data frames (one for solar radiation, one for temperature) contain the values of the variables with the MERRA points in the columns and the days and hours in rows. For the analysis also the date is necessary. This date is created simply by a date sequence.

```
> seq<-seq(as.POSIXct(paste(year,"-01-01
00:30",sep="")),as.POSIXct(paste(year,"-12-31
23:30",sep="")),"hours")
datumCET <- seq[month(seq)==as.numeric(month)]
```

Furthermore the capacity of the single grid zones as presented in chapter 3.1.2. is included. Therefore the photovoltaic production for 1 kWp installed capacity per MERRA point must be multiplied with the capacity per MERRA point per grid zone. Therefore the text files of the capacity per grid zone are imported with the code:

```
> amp_cap_m <- read.delim("Folder/Amp_cap_m.txt",
stringsAsFactors=FALSE)
> cap_amp<-amp_cap_m[,as.numeric(month)-5+12*(year-2010)]
> cap_amp<-t(cap_amp)

> her_cap_m <- read.delim("Folder/Her_cap_m.txt",
stringsAsFactors=FALSE)
> cap_her<-her_cap_m[,as.numeric(month)-5+12*(year-2010)]
> cap_her<-t(cap_her)

> ten_cap_m <- read.delim("Folder/Ten_cap_m.txt",
stringsAsFactors=FALSE)
> cap_ten<-ten_cap_m[,as.numeric(month)-5+12*(year-2010)]
> cap_ten<-t(cap_ten)
> tra_cap_m <- read.delim("Folder/Tra_cap_m.txt",
stringsAsFactors=FALSE)
> cap_tra<-tra_cap_m[,as.numeric(month)-5+12*(year-2010)]
> cap_tra<-t(cap_tra)
```

Next the photovoltaic production per grid zone is calculated. Here the convertToPV function is applond on all MERRA points and at the end the result is multiplied with the capacity at the different MERRA points for the different zones. As already mentioned the convertToPv function needs four variables plus two extra variables for the calculation. While the latitude is taken from the list of latitude from Step (2), and the date, the solar radiation, and temperature are also already in a list by step (1), the inclination angle and the direction are entered via the code:

```
> pv_final<-list()
cap<-NULL
if(zone=="her"){
  cap<-cap_her }
if(zone=="amp"){
  cap<-cap_amp }
if(zone=="ten"){
  cap<-cap_ten }
if(zone=="tra"){
  cap<-cap_tra }

for(i in 1:ncol(df.swgdn)){
  print(i)
  lat <-lonlat[i,2]
  pv_final[[length(pv_final)+1]]<-
convertToPv(lonlat[i,2],datumCET,df.swgdn[,i],df.ts[,i],
15,45)*cap[i]}

View(pv_final)
```

The resulting table of pv_final gives the PV production per MERRA point per month per grid zone and is furthermore transformed into a data frame. At last the photovoltaic production of the single MERRA points is summed up, so that there is one value of production per hour and per day.

```
> df<-data.frame(matrix(unlist(pv_final), nrow=length
pv_final[[1]]),byrow=F))
> final_sum<-apply(df,1,sum)
```

This data is saved in a text file depending on the settings.

```
> write.table(final_sum,paste("Folder/15° Tilt 45°
West/",zone,year,month,".txt",sep=""), sep="\t")
```

In the end there is a text file for each month from July 2010 to July 2013 per grid zone that gives the total sum of production of the MERRA points per hour. The text files of the months and grid zones were joined so that for each grid zone there is a single column giving the production from 19. July 2010 00:00 to 15. July 2013 23:00. This format fits to the format of the measured data, see chapter 3.1.3.

For the time series analysis from 1980 to 2015 the simulation was done similar to the description above, but not separated by grid zones but for total Germany in one run. Therefore the capacities per MERRA point of the single grid zones were summed up, so that the capacity per MERRA point for total Germany is available. The capacities were taken from December 2013. To make the results comparable to the capacity factor of Pfenninger & Staffell (2016a) the final result of the simulation was divided by the total capacity of December 2013. The capacity factor represents the potential photovoltaic production if 1 kWp photovoltaic system was installed in total Germany.

3.2.2 Correction factor for simulation results

When comparing simulation results to measured generation data, a correction factor can be determined as Pfenninger & Staffell (2016a) did in their study. This correction factor should adjust the average simulation results to the average measured generation data. Therefore a ratio of the average total amount of photovoltaic production per year for the period from July 2010 to July 2013 is determined between the simulation results and the measured generation data. Depending on the total difference of the data sets an average annual difference in percent will be calculated and applied to the simulation results, so that the total produced amount of the simulation fits with the total produced amount of the measured generation data. The results are given in chapter 4.1.

3.2.3 Comparison of simulated and measured generation data

The comparison of simulated and measured generation data is done with a correlation analysis. The correlation is given on an hourly, daily, monthly, and seasonally basis separated by the four grid zones and for total Germany. The most used correlation coefficients are the Pearson and the Spearman correlation coefficient. For Pearson there should be a linear relation between the data and it should be normally distributed, thus not highly skewed. In comparison to that Spearman does not make any assumptions about the distribution of the data. Chok (2010) states, that for a moderately skewed distribution Pearson still is the best method. One further restriction for Pearson is the sample size. Bonett & Wright (2000) recommend using Pearson for a sample size bigger than 25. So in the present analysis the Pearson correlation is used for the hourly, daily and monthly analysis, while for the seasonally and annually analysis the Spearman correlation coefficient is taken. A further comparison between the data is given with a boxplot representing the minimum, 1st quartile, mean, 3rd quartile and maximum of the simulated and measured generation data on a daily and monthly basis.

3.2.4 Time series analysis

The time series analysis was done from 1980 to 2015 and aims to find long-term patterns. As explained at the end of chapter 3.2.1. the simulation was done for total Germany with the installed capacity from December 2013. The result was then divided by the total installed capacity to obtain the capacity factor. The results will then be analyzed regarding extreme values. For each day of the year the following statistical values over the 35 years are calculated: minimum, maximum, mean value and standard deviation. These values are then separated into quantiles of 0 – 10%, 11-25%, 26 – 50%, 51 – 75%, 76 – 90%, and 91 - 100%, as done in Pfenninger & Staffell (2016a). The same is done on an hourly basis separated by the seasons to see the average development over the day.

4 Results

The following chapter first presents the results of the different settings of the simulation and the correction factor that was applied to the simulation outcome in chapter 4.1. In the subchapters 4.1.1 to 4.1.5 the comparison of the simulation to the measured generation data is presented for the period from July 2010 to July 2013 on different time resolutions. Chapter 4.2 presents the results of the time series analysis for the period from 1980 to 2015. In chapter 4.3, the results of chapter 4.2 are compared to the simulation results of Pfenninger & Staffell (2016a).

4.1 Comparison of simulated and measured generation data

As mentioned in chapter 3.2.2 the simulated and the measured generation data were compared to each other by determining the correlation at different time resolutions (hourly, daily, monthly, seasonally, and annually). Hours with zero production in the simulation and the measured generation data at the same time were excluded from the correlation, as this is true for all night hours and the correlation coefficient thereby would be artifically increased. First an overview of the correlation results is given in chapter 4.1.1. Then the hourly, daily, monthly, and seasonally time resolutions are explained in more detail in chapters 4.1.2 to 4.1.5. The results are all based on a simulation with a setting of a 15° inclination of the photovoltaic panel and a position of the panel to the South-West (45°). The correlation of this setting was the best in three of the four grid zones compared to other settings, as can be seen in Table 3. Although the correlation coefficient for the setting with 20° tilt and South-West (45°) was better for one grid zone and total Germany, the differences were very small, so that 15° tilt and 45° South-West were chosen. From the table one can see that with a lower tilt angle the correlation results get better. Furthermore the correlation coefficients for each tilt position become better if faced to the South-West compared to the default value were it faces to the South.

Table 3: Correlation coefficients of different settings for inclination and orientation in the simulation compared to measured generation data of the photovoltaic production (Own calculation)

	Amprion	Hertz50	Tennet	TransnetBW	Germany
Default Value	0.865	0.879	0.882	0.842	0.886
25° Tilt – South	0.886	0.905	0900	0.866	0.906
25° Tilt – South West	0.923	0.925	0.929	0.905	0.939
20° Tilt - South	0.891	0.911	0.905	0.874	0.912
20° Tilt – South West	0.926	0.932	**0.933**	0.911	**0.943**
15° Tilt - South	0.895	0.916	0.908	0.880	0.916
15° Tilt – South West	**0.927**	**0.936**	0.924	**0.914**	0.942

Besides the better correlation results, the setting of 15° tilt and 45° South-West also yielded the best results when compared to the measured generation data in total amounts of photovoltaic production, see Figure 13. The figure shows that the results of the 15° tilt and 45° South-West setting are closest to the measured production. One can see that the setting with the default values, a tilt of the latitude minus 10° and an orientation to the South, highly underestimates the summer months, while it overestimates the winter months. Comparing the different settings one can see that with a lower degree of the tilt the production gets lower in winter and higher in summer. The same is true for the orientation. A orientation to the South-West yields lower production in winter, but higher production in summer compared to an orientation to the South. Further settings could have been tested, but as the results of the setting with 15° tilt and 45° South-West yielded results close to the measured production and the setting is realistic, as not all photovoltaic panels can be set to optimal conditions, the testing was stopped at this setting.

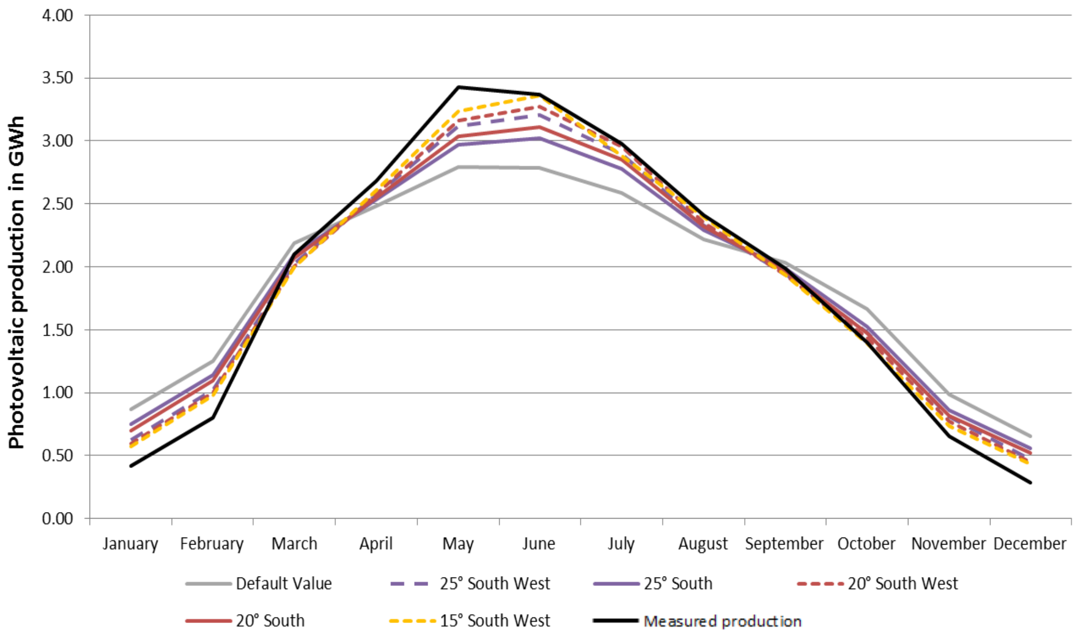


Figure 13: Comparison of monthly measured generation to production under different simulation settings (Own representation)

After choosing the right setting, the correction factor for the simulation was calculated, as explained in chapter 3.2.2. It is based on the average difference of total phtovoltaic production compared to measured generation data. Figure 14 shows the average surplus of the simulation results compared to the measured photovoltaic production in percentage before the correction factor for each month.

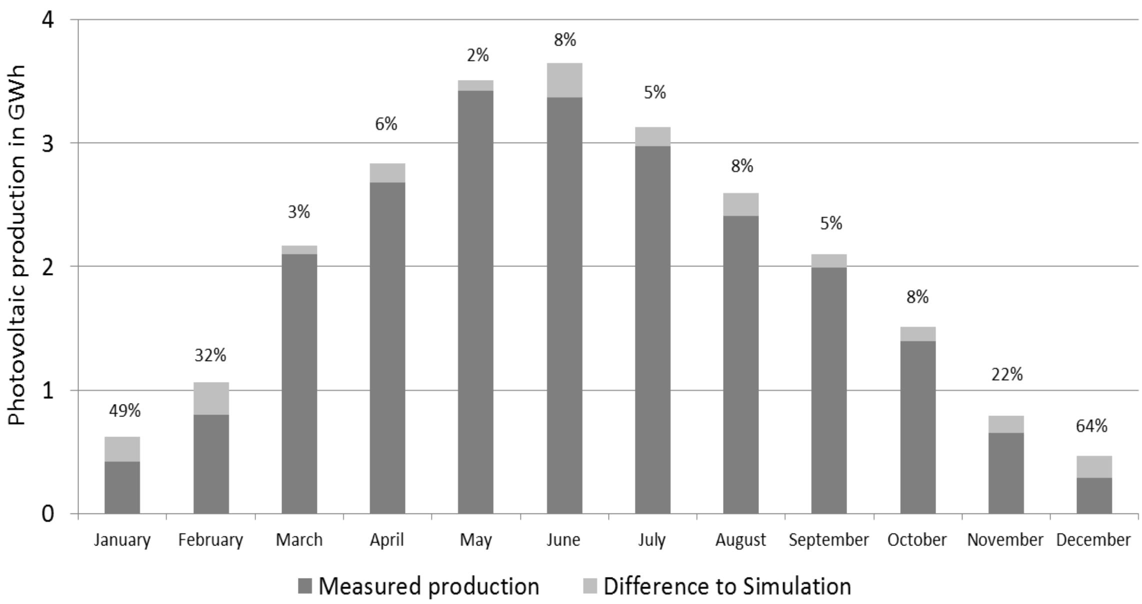


Figure 14: Surplus of simulated to measured photovoltaic production for average monthly production in Germany from July 2010 to July 2013 (Own representation)

It shows that there is a high overestimation of the measured production in the winter months, with 49% in January and 64% in December, while the summer production was simulated quite well, with a surplus of about 8%. Due to these results a correction factor over the year (average surplus) was calculated. The correction factor of 92% was then applied to the simulation by multiplying the results, so that in total the production of the simulation over the year is the same as the measured generation. The differences after the correction factor are shown in Figure 15. The simulation shows now a lower result in the summer months than the measured generation data ranging from -3% to 0%. In contrast to that the overestimation in the winter months could be reduced from 49% to 37% in January and from 64% to 51% in December. The correction factor of 92% is true only for the total German production. For the single grid zones different correction factors were calculated: Amprion 96%, Hertz50 79%, Tennet 92%, and TransnetBW 104%.

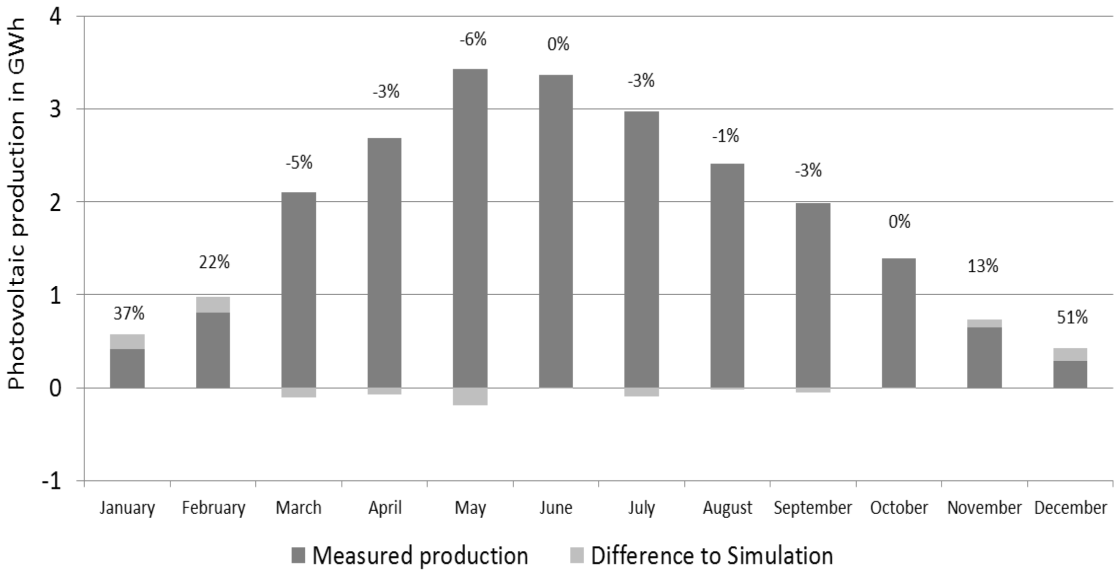


Figure 15: Surplus of simulated to measured photovoltaic production for average monthly production in Germany from July 2010 to July 2013 after application of correction factor (Own representation)

4.1.1 Overview

Table 4 shows the results of the correlation between simulation and measured generation data at the different time resolutions for the four grid zones and total Germany with the above mentioned setting of 15° inclination and an orientation to South-West and the correction factor of 92%. All results are significant at a 0.01 level. For the hourly, daily and monthly correlation the Pearson coefficient and for the seasonally and annually correlation the Spearman coefficient have been used, as there were less than 25 observations available. As already mentioned, hours with zero production in the simulation and the measured generation data at the same time, were not included in the results, to avoid an artificial increase of the correlation coefficient. While the coefficient in the hourly comparison for total Germany is 0.942, it rises to 0.948 at the daily resolution and to 0.986 at the monthly comparison. The four grid zones show the same pattern in their correlation coefficients as Germany. The seasonal comparison also delivers a high correlation with 0.974. In sum all correlation coefficients are very high (R > 0.9), which shows that the simulated results describe the measured generation data well in terms of their time profile. The highest values are achieved for the annual comparison, but due to the small number of compared observations (4 observations) these results should not be seen as meaningful. Therefore the annual data is not further investigated in the comparison of simulated and measured production.

Table 4: Correlation coefficients for the comparison between simulation and measured generation data from July 2010 to July 2013 with a significance level of 0.01 (Own calculations)

	Amprion	Hertz50	Tennet	TransnetBW	Germany
Hourly (26,232 h) Pearson	0.927	0.936	0.924	0.914	0.942
Daily (1,903 d) Pearson	0.935	0.954	0.918	0.923	0.948
Monthly (37 m) Pearson	0.988	0.988	0.972	0.977	0.986
Seasonally (14 s) Spearman	0.974	0.969	0.982	0.978	0.974
Annually (4 y) Spearman	1.000	1.000	1.000	1.000	1.000

4.1.2 Hourly comparison

Here, we analyse the hourly data of the measured production and the simulation. The average hourly production of the day was calculated over the period from July 2010 to July 2013. The results are presented in Figure 16 that shows that the simulation yields similar results to the measured generation data for the hourly average production. The simulation slightly underestimates the measured production in the morning hours from 4 a.m. to 7 a.m, while it overestimates it in the time from 8 a.m. to 14 p.m. In the afternoon it again underestimates it from 14 p.m. until the end of the sunshine at 19 p.m. Nevertheless the development of the production over the day is similar in the measured production and the simulation.

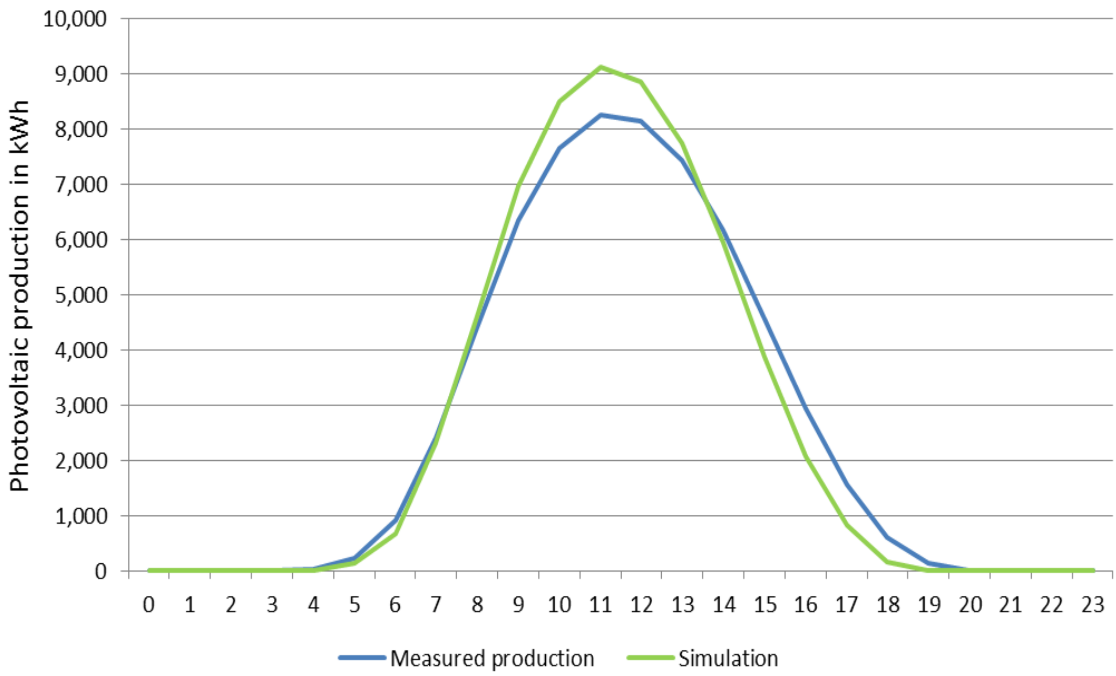


Figure 16: Comparison of measured generation data and simulation for average hourly photovoltaic production in Germand and the four grid zones from July 2010 to July 2013 (Own representation)

The same applies to the four grid zones, as seen in Figure 17. They all are overestimated in the simulation around midday, while they are underestimated in the afternoon. For the morning the simulation presents the measured production almost exactly in the zones Amprion, Tennet, and TransnetBW, while it slightly underestimates it in Hertz.

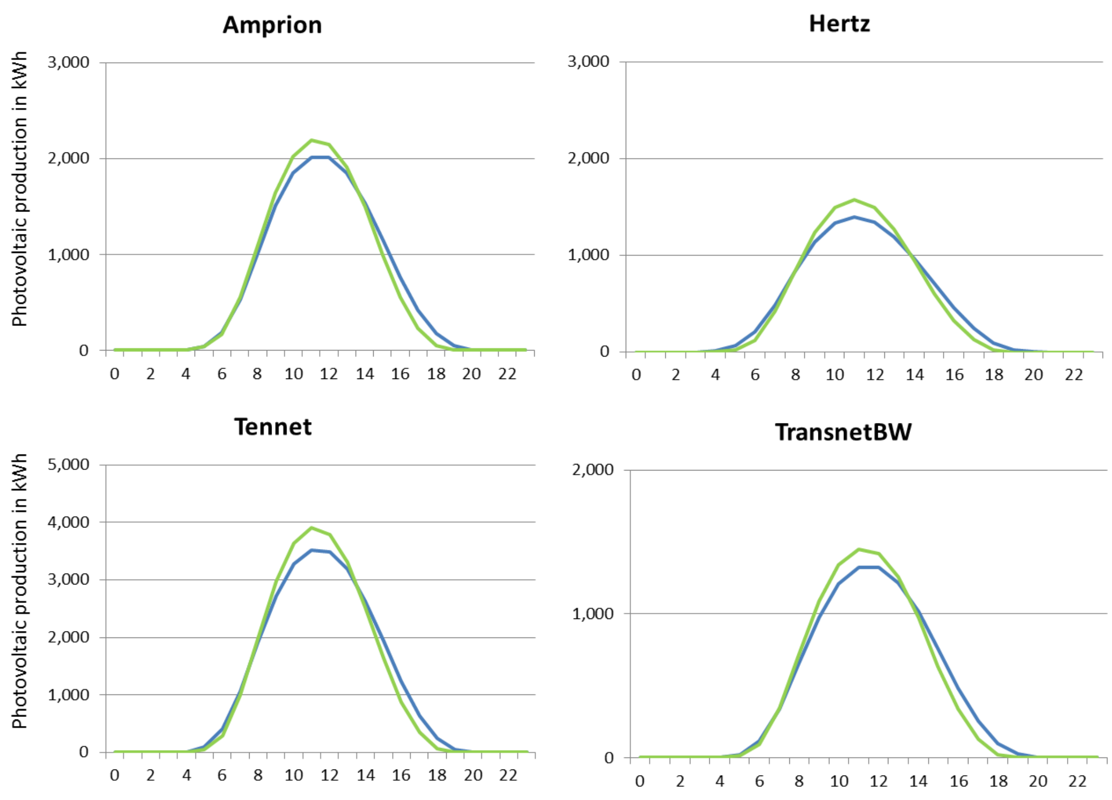


Figure 17: Comparison of measured generation data and simulation for average hourly photovoltaic production in the four grid zones from July 2010 to July 2013 (Own representation)

For the next step of the analysis, the simulation results and the measured produced amounts of all single hours were taken and plotted into a scatter diagram separated by months. The results are shown in Figure 18. The plots show that in January, February and December, i.e. the winter months, the simulated production is higher at the most hours compared to the measured production and that there are only a few hours at the right side of the 45°-line. Furthermore in January and December the hourly points are concentrated more around a production of 0 to 10 MWh. The correlation coefficient therefore is low with 0.82 in January and 0.84 in December. From March to October the correlation coefficient rises up to above 0.90 with the highest value in May with 0.97. The points also converge closer to the 45°-line which means that the simulated production is closer to the measured production. In the plots of January to March and October to December there are more hours at the left of the 45°-line than there are on the right side, which means that the simulated production is in almost all cases higher than the measured production, thus overestimates the measured values. In contrast to that in the months from April to September the points are equally distributed or tend more to the right side of the 45°-line, so production in this months is slightly underestimated.

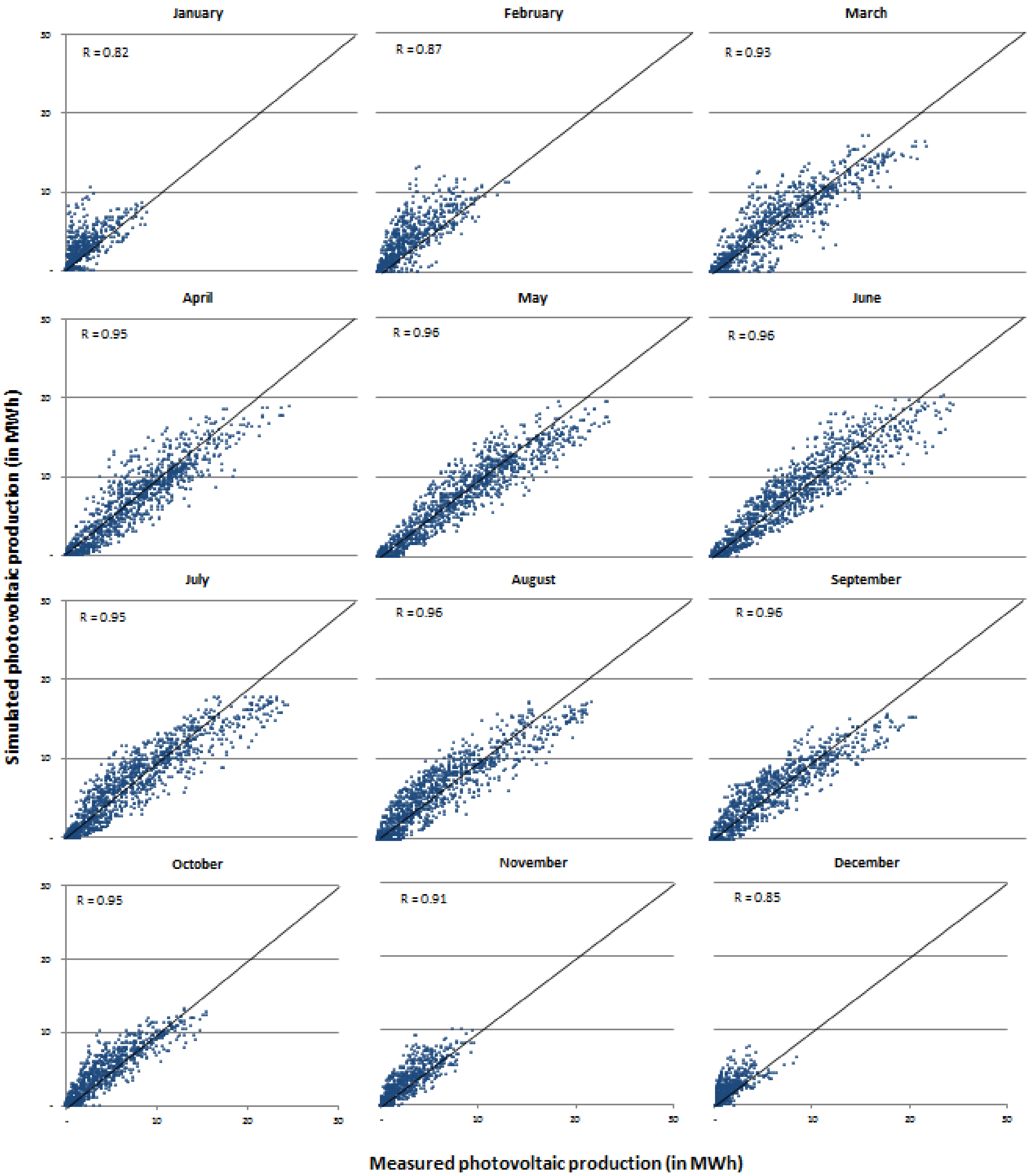


Figure 18: Scatter plot of hourly simulated and measured production per month in Germany July 2010 – July 2013 (Own representation)

4.1.3 Daily comparison

The analysis of the daily production was done based on the statistical characteristics (minimum, maximum, median, 25-% and 75%-quantiles) of the simulation results and the measured data. Figure 19 shows the box plots of the four grid zones and Germany with the measured generation data and with the simulation results. Most parameters of the distribution of the simulated data in all four grid zones and in total Germany are close to the values of the measured generation data. For all grid zones and Germany the median and the 25% quantiles of the simulation are higher than the measured generation values, although the difference is very small, while the maxima and the 75% quantiles of the simulation are lower than measured generation values. Furthermore the boxplots show that the simulation is not able to simulate the minima of daily production correctly. In all grid zones and Germany the simulated minimum is higher than the measured produced minimum. The box plot shows that the statistical values of the simulation are more compact than the measured production, but that the results are very similar to each other, stating that the simulation can quite well represent the measured generation.

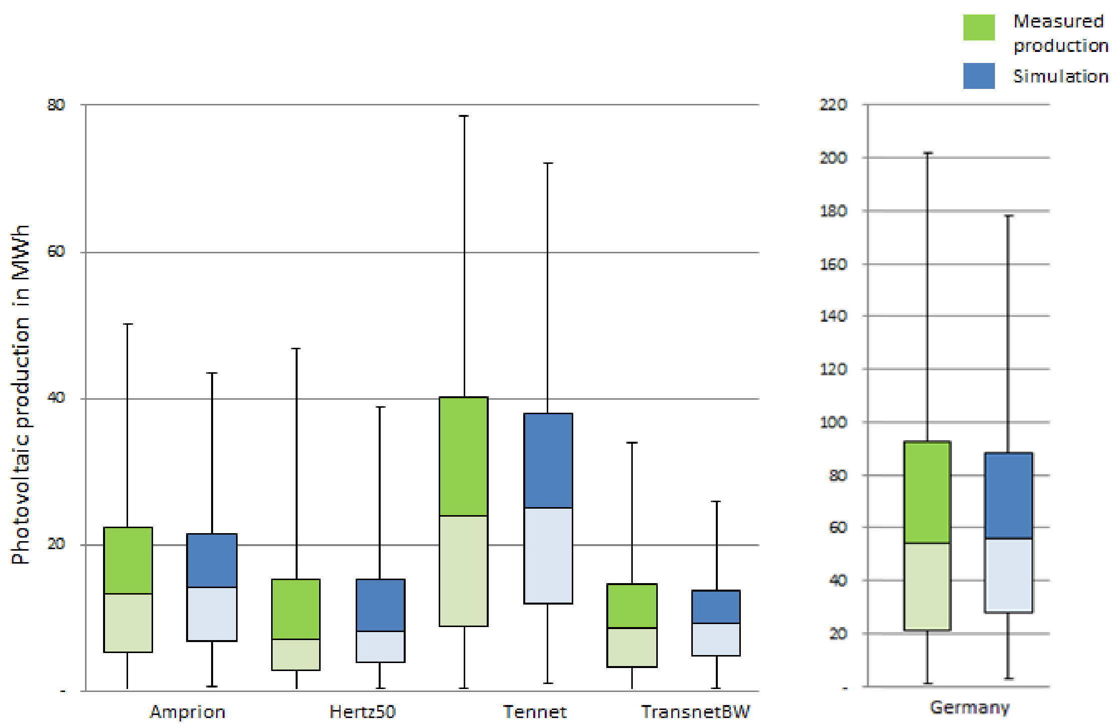


Figure 19: Box plots of simulated and measured data for grid zones and Germany based on daily data (Own representation)

4.1.4 Monthly comparison

In a next step, the data was aggregated from an hourly to a monthly basis. The results are shown in Figure 20 for the four grid zones and in Figure 21 aggregated for whole Germany. They show the comparison between measured production (full line) and simulated production (dotted line). It can be seen that the simulation overestimates the possible photovoltaic production in Germany and in each of the four grid zones in the months of January, February, November and December in all years, although the overestimation is smaller in December 2011 to February 2012 compared to the other years. The months March, April, September, and October are simulated quite well and therefore are close to the measured production in the years 2012 and 2013, while in 2011 March and April are underestimated by the simulation. The months from May to August differ over the years. While they are underestimated in one year, they are overestimated in the next year. On average this over- and underestimations add up, so that the average production is simulated close to the measured production, as has been seen in Figure 14 of chapter 4.1.

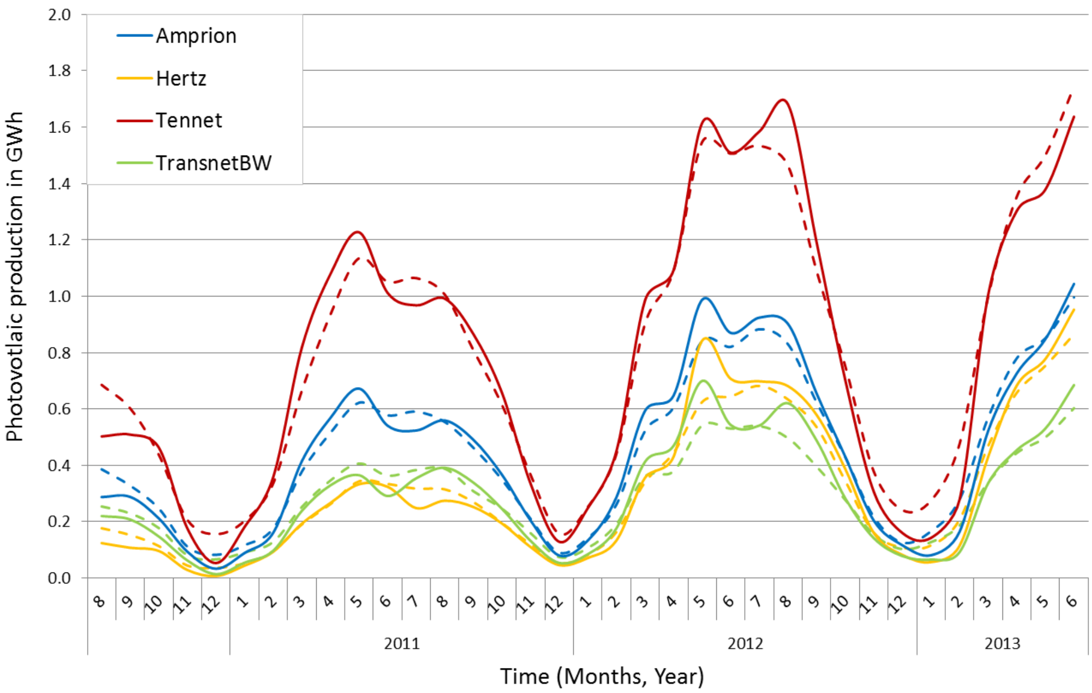


Figure 20: Measured (full line) and simulated (dotted line) monthly photovoltaic production for the four grid zones from July 2010 to July 2013 (Own representation)

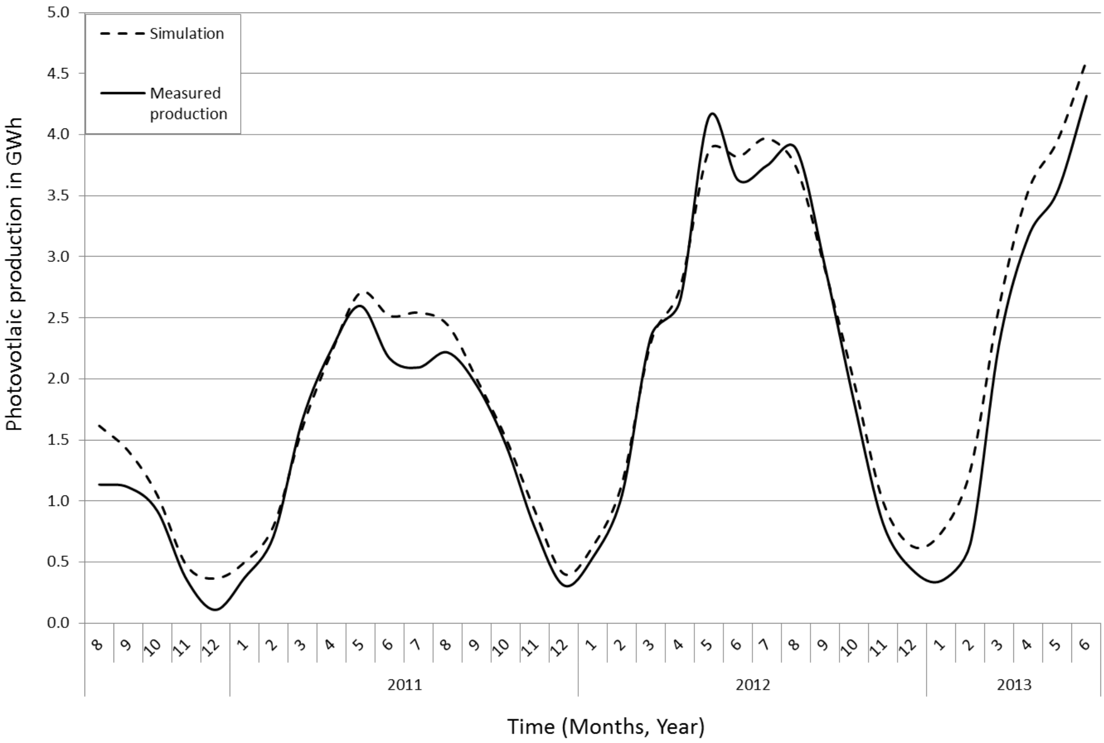


Figure 21: Measured (full line) and simulated (dotted line) monthly photovoltaic production for Germany from July 2010 to July 2013 (Own representation)

The difference in percentage for the average of the months was already shown in chapter 4.1 when the correction factor was explained. It showed that while the simulation reprents well the months from March to October with an overestimation of only 2% to 8%, the overestimation is much higher in the months from November to February with 22% to 64%.

4.1.5 Seasonal comparison

As already mentioned in the comparisons before there are bigger differences between the simulated production and the measured production in winter than in summer. Therefore a correlation of the hourly data was done separated by seasons. Table 5 shows that the correlation coefficient in summer, spring, and autumn is almost always very high (R > 0.9), while in winter the correlation between simulation and measured generation data is lower in most cases (R < 0.9). Especially winter 2010/11 and 2012/13 show a low correlation. While in 2010/11 at least one grid zone, Hertz 50, reaches a correlation value above 0.9, Amprion is lowest with 0.79 and in winter 2012/13 all correlation values are equal or below the 0.85 mark.

Table 5: Correlation coefficient (R) between simulation and measured generation data for the four seasons separated by year and based on hourly data (Own calculation)

	Amprion	Hertz50	Tennet	TransnetBW	Germany
Summer 2010	0.900	0.925	0.939	0.948	0.947
Autumn 2010	0.894	0.924	0.910	0.925	0.933
Winter 2010/11	0.786	0.910	0.843	0.813	0.855
Spring 2011	0.963	0.969	0.965	0.966	0.971
Summer 2011	0.953	0.950	0.948	0.952	0.962
Autumn 2011	0.952	0.951	0.953	0.946	0.962
Winter 2011/12	0.945	0.927	0.942	0.926	0.950
Spring 2012	0.953	0.942	0.956	0.957	0.964
Summer 2012	0.959	0.960	0.963	0.949	0.967
Autumn 2012	0.940	0.954	0.948	0.928	0.957
Winter 2012/13	0.815	0.804	0.833	0.795	0.846
Spring 2013	0.939	0.937	0.948	0.918	0.953
Summer 2013	0.950	0.960	0.892	0.939	0.954

Based on the results of the correlation the differences of the seasons were further investigated. Figure 22 shows the difference of the simulation and the measured generation data in percentage for the seasons. While in spring, summer, and autumn the overestimation of 4% to 9% is very small, the simulation is almost half as high as the measured generation data in winter with an overestimation of 43%. Regarding these results a closer look was taken at the single winter periods from 2010 to 2013. The result can be seen in Figure 23. While the winter period of 2011/12 (December 2011, January and February 2012) are simulated quite well with an overestimation of only 15%, the other two periods are more overestimated with the period of 2010/11 by 39% and 2012/13 by 83%. These results fit with the correlation table above, where the correlation of these two periods was worse compared to all other periods. Furthermore it shows that not every winter period is completely overestimated, but that some periods are highly overestimated, which leads to an higher overestimation compared to the other seasons.

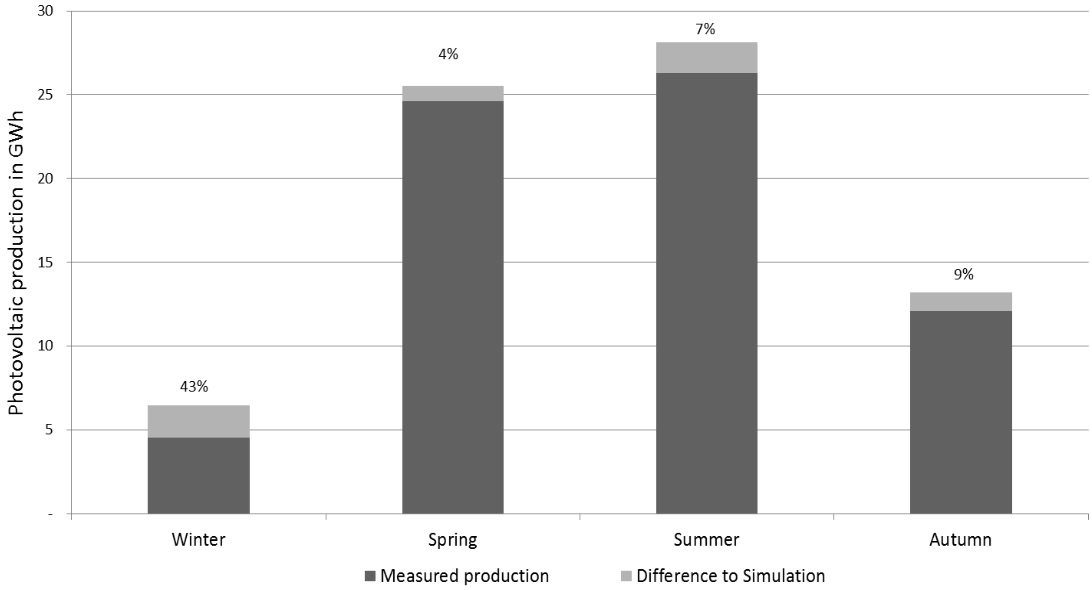


Figure 22: Surplus of simulated to measured photovoltaic production for average seasonally production from July 2010 – July 2013 in Germany (Own representation)

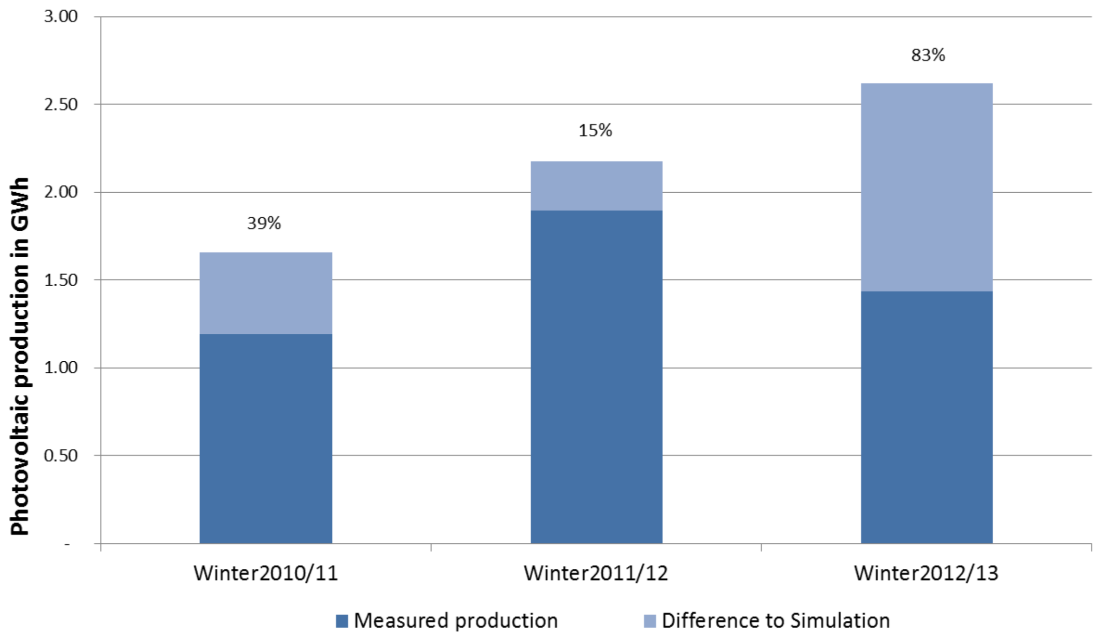


Figure 23: Surplus of simulated to measured photovoltaic production for production in the winter periods from July 2010 – July 2013 in Germany (Own representation)

The same results as above, can be found when comparing the simulation and measured generation data of the seasons on an hourly base. Figure 24 shows that while the average daily development and production is simulated almost perfectly, the simulation in winter highly overestimates the measured generation.

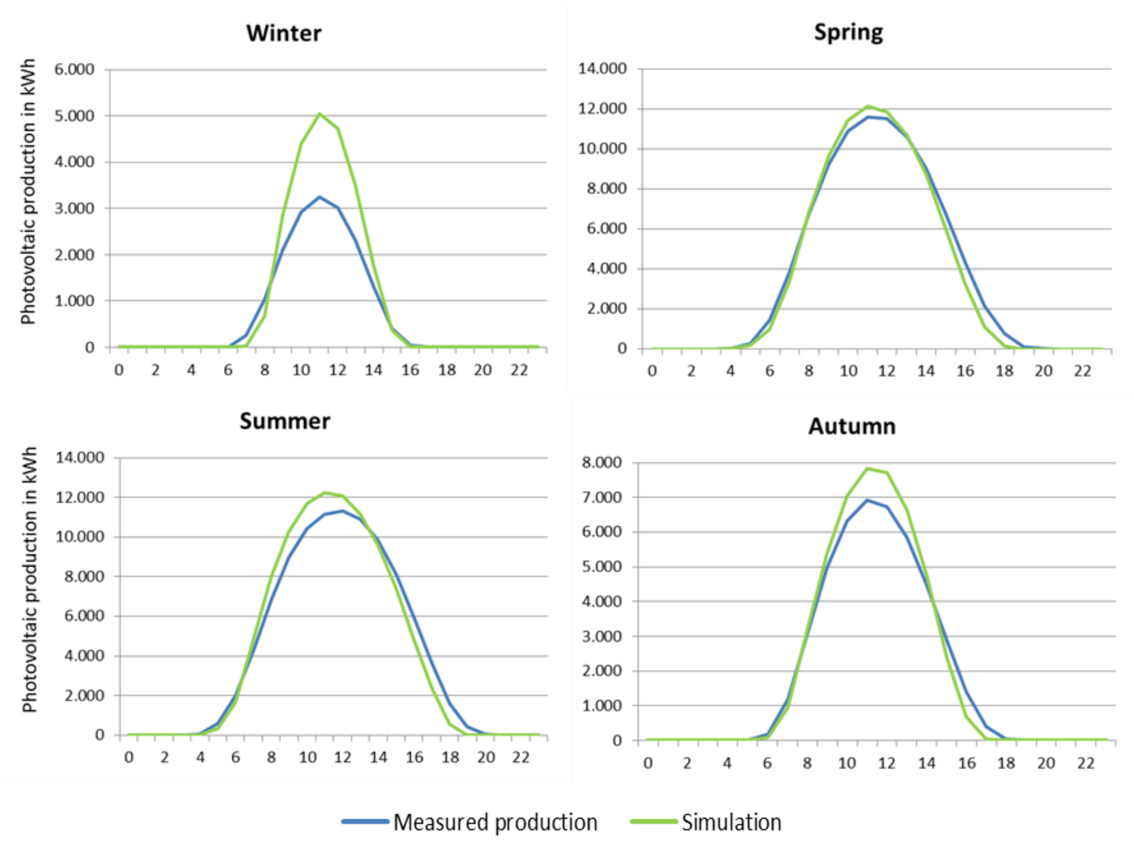


Figure 24: Comparison of measured generation data and simulation for average hourly photovoltaic production in the four seasons from July 2010 to July 2013 (Own representation)

All in all chapter 4.1 showed that the simulation was able to reproduce the development of the measured photovoltaic production quite well, shown by the high correlation coefficients and the low deviations between the total amounts of measured generation data and simulation. While the simulation highly overestimates production in the winter months, it simulates the production of the other months quite well, especially when using a correction factor, which leads to a balance of the winter months and therefore yields a good result between simulation and measured generation data for the total annual production.

4.2 Time series analysis

The time series analysis covers a period from the beginning of 1980 to the end of 2015. The simulation was done for these 35 years on an hourly basis and a correction factor of 92% was applied to the results based on the results of chapter 4.1. To make the time series comparable to the results of Pfenninger & Staffell (2016a), the capacity factor was calculated as explained at the end of chapter 3.2.1. Figure 25 shows the statistical values of the daily capacity factor for photovoltaic systems in Germany from 1980 – 2015.

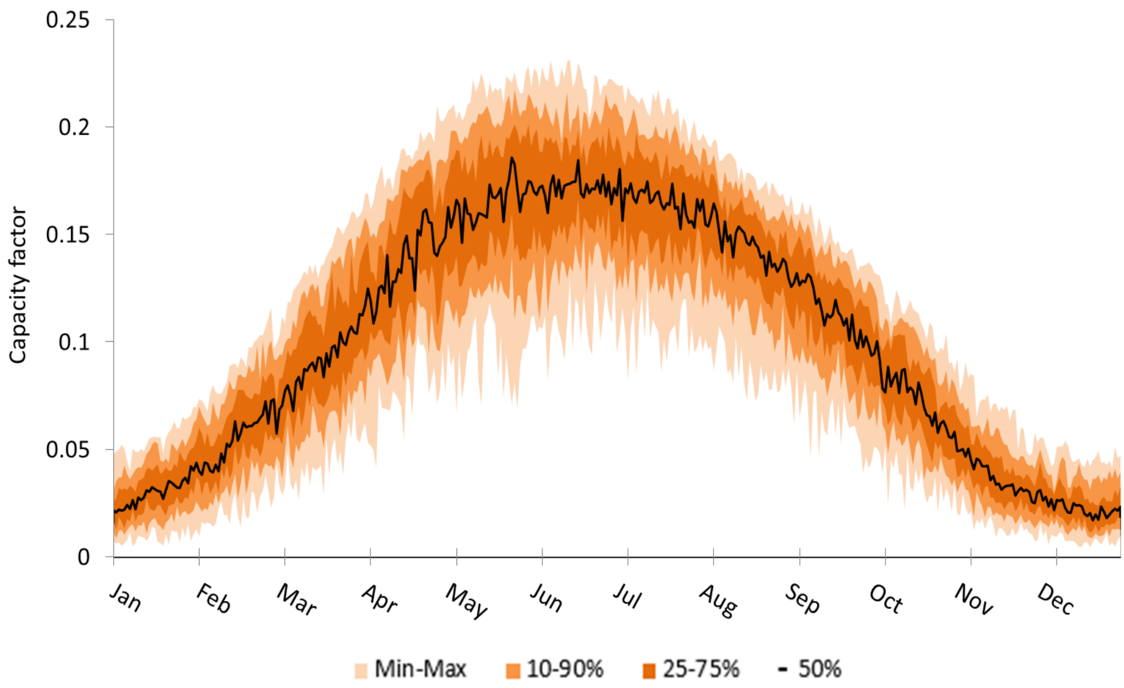


Figure 25: Average daily capacity factor for photovoltaic systems in Germany from 1980 to 2015 (Own representation)

The average daily capacity factor starts lowest in January at 0.02 and rises up to 0.18 in the months from May to August before it falls back to 0.02 in December. The other statistical values as minimum, maximum, and the quantiles (10%, 25%, 75%, 90%) show the same development. The minimum capacity factor is 0.00 in December, while the minimum is highest in June with 0.14. The lowest maximum value lies again in December with 0.04, while the highest maximum is reached in several days in June with 0.23. Regarding the variability of the capacity factor over the year there are no big anomalies visible in the development. At the beginning and the end of the year the production is low, while it is high in the mid of the year. Only in May there is a phase were the minimum capacity factor is quite low compared to the other warmer months with 0.07. Considerung the spread from minimum to maximum, it is smaller in January, February, November, and December with around 0.05 difference, compared to the months from March to October that show twice the difference with 0.10. So the variability between highest and lowest production on a certain day over the 35 years is higher in the months from March to October.

Figure 26 shows the statistical values of the hourly capacity factor over the day for the four seasons. All seasons show the expected development over the day. Production starts at zero at night and slowly rises in the morning up to the highest production at midday that then falls back to less procuction in the evening and zero production at night.

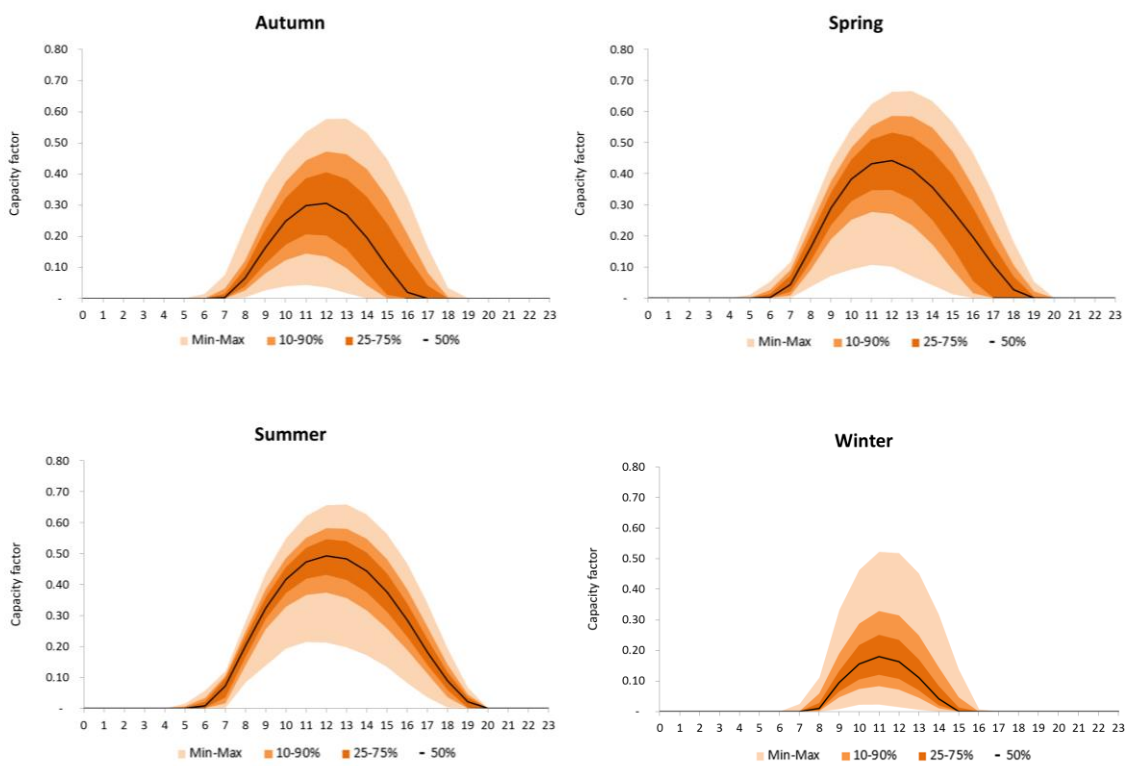


Figure 26: Hourly capacity factor for photovoltaic systems in Germany from 1980 to 2015 for the four seasons (Own representation)

The highest average capacity factors are reached at midday in summer with 0.50. Compared to that the highest average in spring is 0.44, in autumn 0.31 and in winter 0.19. These differences are normal related to the fact that in spring and summer there is more sunshine and thus more solar radiation for generation. The maximum capacity factor is reached in spring at 0.67, and summer at 0.66. In winter the maximum value is 0.53 and in autumn 0.58. All in all the development over the day shows the expected pattern in all four seasons. The variability of the capacity factor is lowest in summer with a difference of 0.44, where especially the minimum value lies above the minimum values of the other seasons with 0.22 compared to 0.11 in spring, 0.04 in autumn and 0.03 winter. The difference between maximum and minimum in spring is 0.56, in autumn 0.54, and in winter 0.50. Figure 25 and Figure 26 are also compared to the results of Pfenninger & Staffell (2016a) in the following chapter 4.3.

Figure 27 finally shows the simulated distribution of the electricity production in Germany for 2015 if 1 kWp photovoltaic capacity was installed. The results are based on a simple interpolation by inverse distance weighting. There is a higher production in the South West, while production gets lower to the North East. These results fit quite well with the representation of Figure 4 in chapter 2.1.1.

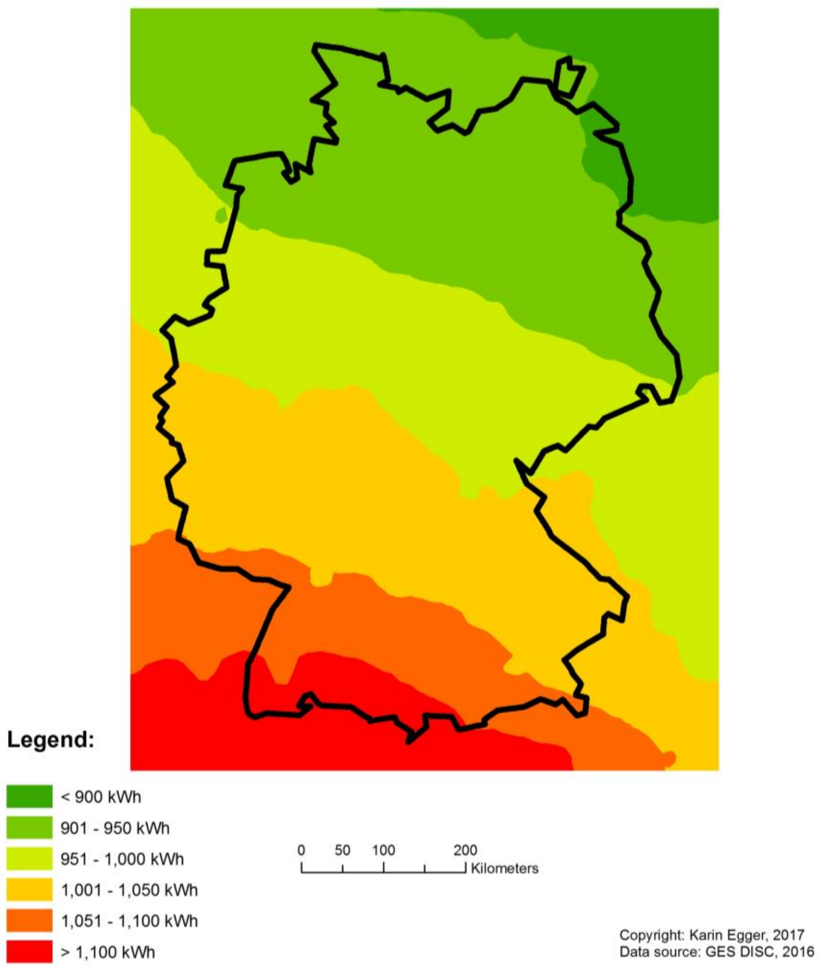


Figure 27: Average annual electricity production of 1kWp installed photovoltaic capacity in Germany 2015 (Own calculation, Own representation)

4.3 Comparison of simulated data to Pfenninger & Staffell (2016a)

The results of Pfenninger & Staffell (2016a) were compared to the results of the simulation of this thesis by a correlation analysis that yielded a correlation coefficient of 0.951. For the correlation hours with zero production in the data of Pfenninger & Staffell (2016a) and at the same time with zero production in the own simulation were not included to prevent creating an artificially higher correlation coefficient. Figure 28 shows a comparison of the monthly average capacity factor between the results of Pfenninger & Staffell (2016a) and the simulation of this thesis for the years 1985 – 2014. One can see that although the two lines have an equal development over the year, the results of the simulation of this thesis are below the results of Pfenninger & Staffell (2016a). While in the months from October to February the differences is 0.01 and 0.02, it is slightly higher in the months from March to September with 0.03. These differences may be due to a different way of simulation and calculation of the capacity factor or as different installed capacities were used.

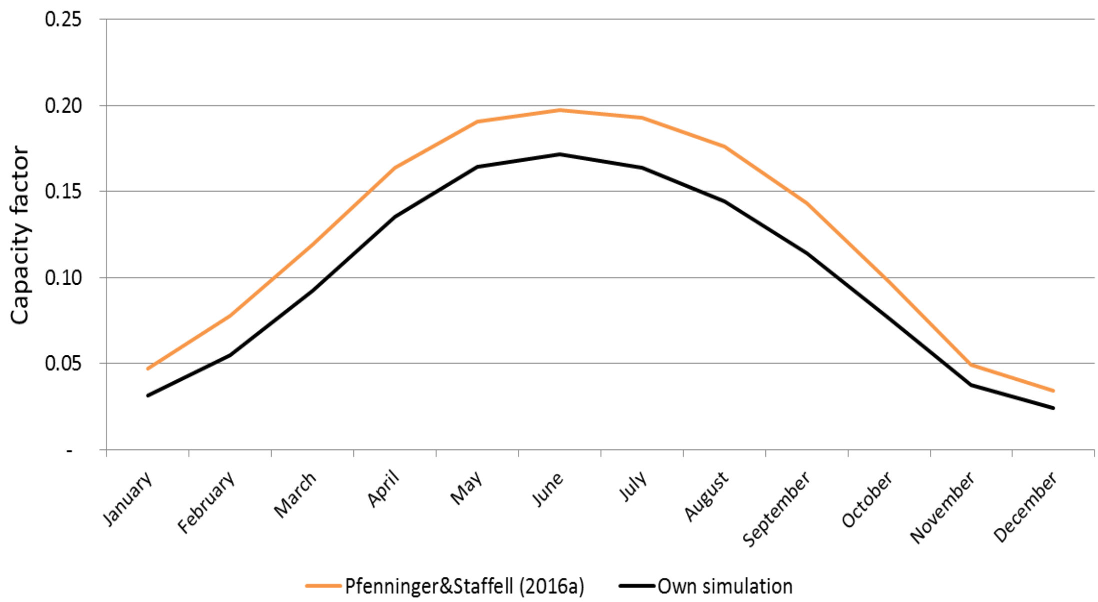


Figure 28: Comparison of Pfenninger & Staffell (2016a) and own simulation for average capacity factor per month from 1985 – 2014 (Own representation)

The surplus in the single months of the simulation compared to Pfenninger & Staffell (2016a) can be seen in Figure 29. It shows that the simulation is below the simulation results of Pfenninger & Staffell (2016a). While the difference in the months January and December is highest with -30% and -25%, it is lowest in May and June with -13%, while the other months vary around -15% to -27%. Comparing Figure 28 and Figure 29 it can be seen that although the difference between the months in total amounts is smaller in winter than in summer, due to the lower production in winter and the higher production in summer the difference in percentage is smaller in summer than in winter.

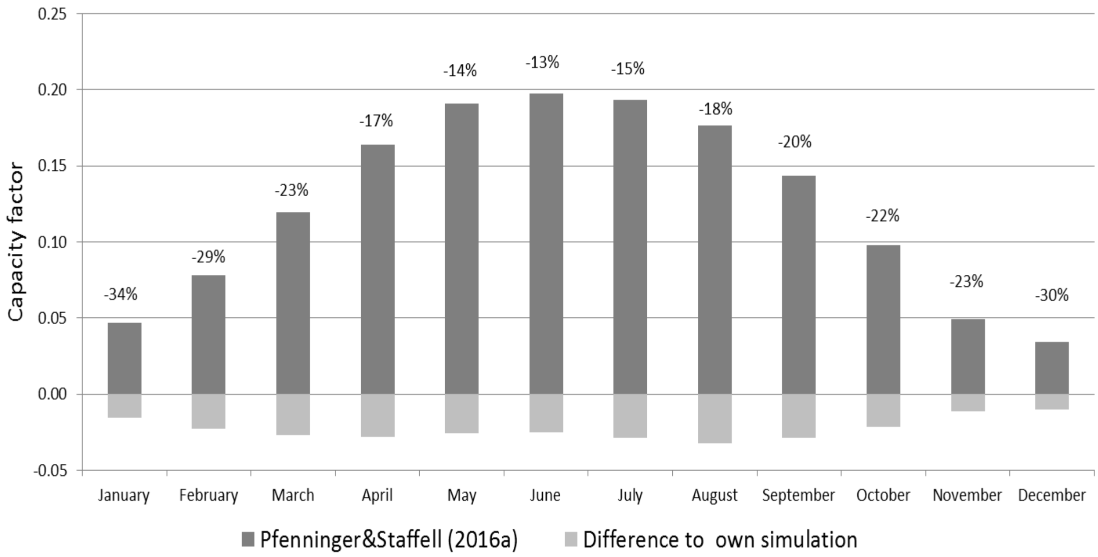


Figure 29: Surplus of own simulation compared to Pfenninger & Staffell (2016a) for average capacity factor per month from 1985 – 2014 (Own representation)

Another comparison was done for the daily capacity factor over the year. The data of Pfenninger & Staffell (2016a) covers the period from 1985 – 2014, while the simulation covers the period from 1980 – 2015. The results of the two simulations are seen in Figure 30. The lower graph of the simulation was already explained in chapter 4.2. The comparison shows that the development of the daily capacity factor over the year is similar in both simulations. The start of the average capacity factor is almost the same in the simulation of Pfenninger & Staffell (2016a) and the own simulation with 0.03 respectively 0.02. Both rise up to 0.20 respectively 0.17 in the summer months and fall back to 0.03 and 0.02 in November and December. As already seen in the above figures, the simulation yields a smaller average capacity factor in all months than Pfenninger & Staffell (2016a) do. Nevertheless there are no big differences or anomalies of the capacity factor over the year, neither in the average values nor the minima or maxima.

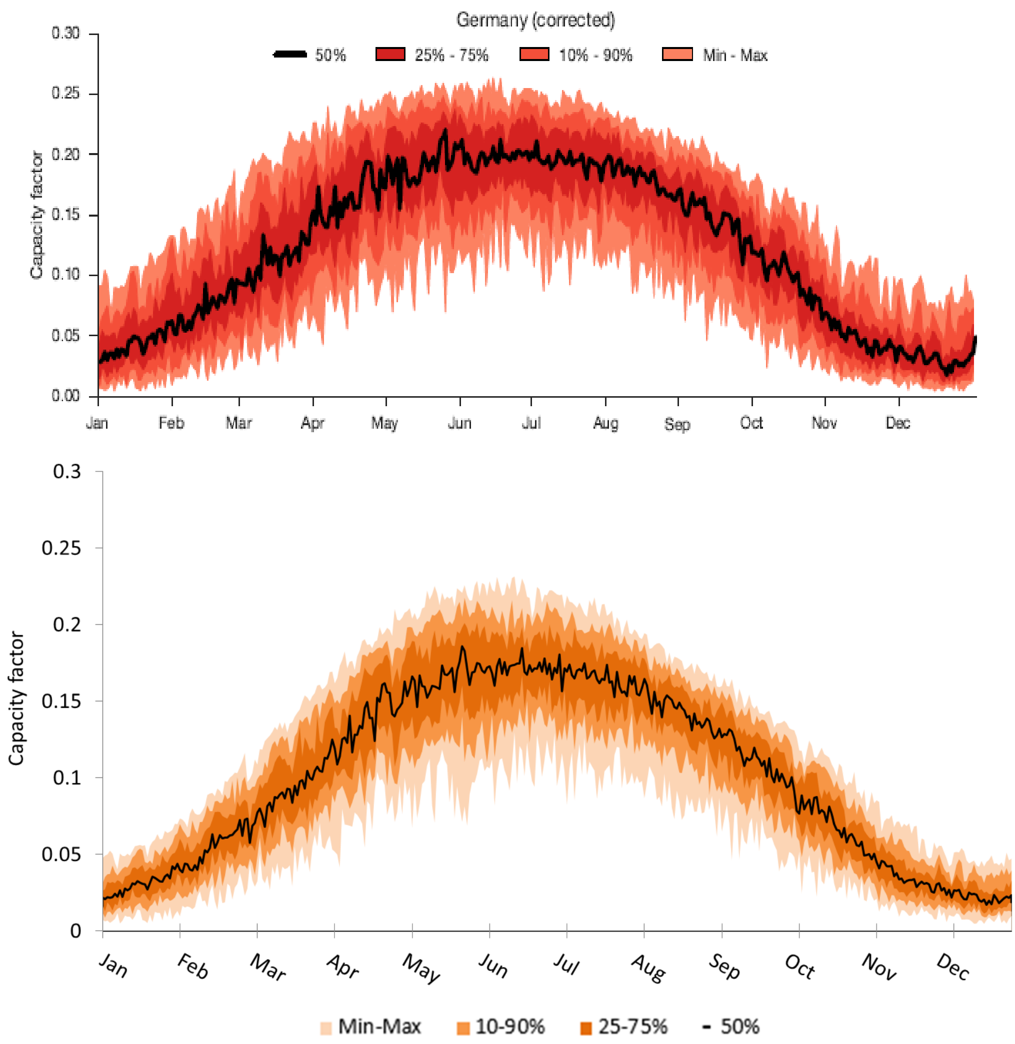


Figure 30: Comparison of the daily capacity factor by the simulation of Pfenninger & Staffell (2016a) 1985 -2014, upper graph (red), to own simulation results 1980 – 2015, lower graph (orange).

Pfenninger & Staffell (2016b) also investigated the interdaily variation of the production for summer and winter as seen in Figure 31. In summer the average daily production of Pfenninger & Staffell (2016b) starts at around 3 a.m. and rises up to around 0.53 until midday, before it drops and ends around 19 p.m. in the evening. The maximum capacity factor is reached at midday with around 0.73, while the minimum at this time is around 0.20. Compared to that the average factor of the simulation starts later at 5 a.m., rises to 0.50 at midday and drops until 20 p.m. The maximum factor for the simulation lies at 0.66, while the minimum is almost equal to Pfenninger & Staffell (2016b) with 0.22. In contrast to summer, the average winter capacity factor of Pfenninger & Staffell (2016b) starts later at around 6 a.m. and ends at around 15 a.m. It reaches its highest point around midday with a capacity factor of 0.22 which is far below the average midday capacity factor in summer. In contrast to that, the maximum capacity factor reaches almost the same level as the maximum capacity factor in summer with about 0.65, while the minimum factor is almost close to zero. The production in winter in the simulation starts at 7 a.m. while it ends at 15 a.m. The average factor reaches is highest value at 0.19. The maximum value of the simulation is highest with 0.53 while the minimum is 0.03. So once again the comparison of the figures showed that the simulation results of Pfenninger & Staffell (2016a) are higher than the simulation.

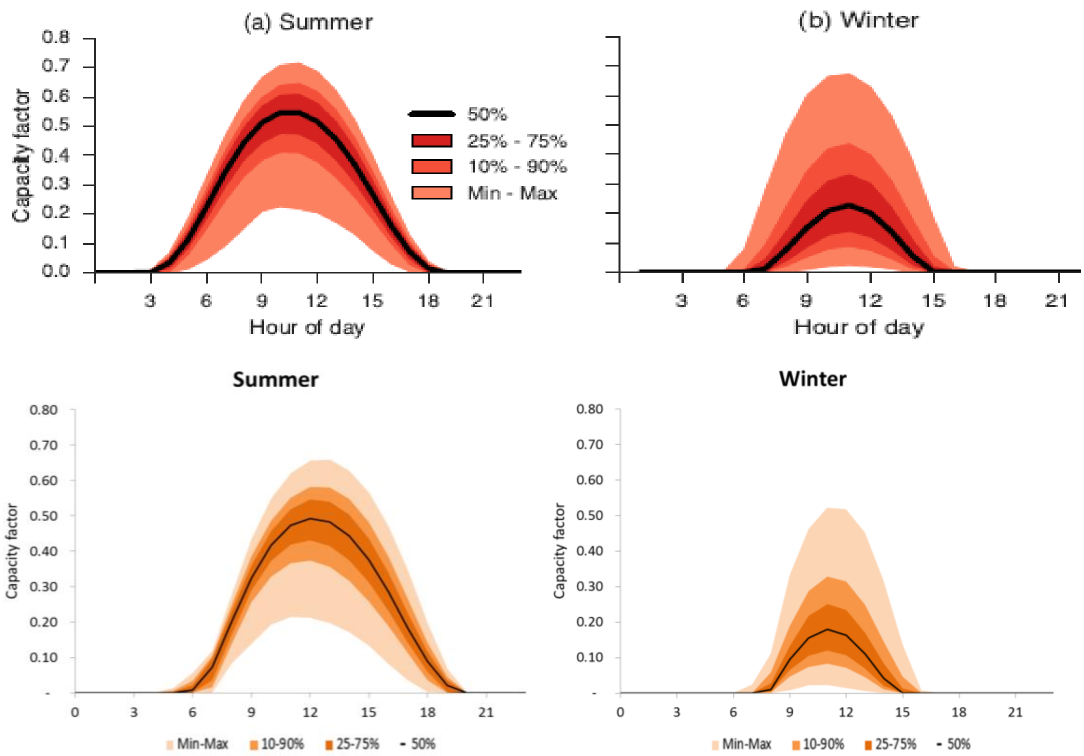


Figure 31: Diurnal variability of hourly capacity factor in Germany 1985 - 2014 of Pfenninger & Staffell (2016b), upper graph (red), to own simulation results 1980 - 2014, lower graph (orange).

Furthermore both simulations show that there is a higher variation between the days in winter than in summer. While the production in summer is on average almost three times as high as in winter, single days in winter can reach almost the same maximum production as average days in summer.

5 Discussion & Conclusions

The following chapter aims to answer the research questions of chapter 1.2 by discussing the results of the analysis in chapter 4.

The first two questions related to the difference between simulated and measured generation data and the development of these difference over the year. The comparison of the simulation to the measured generation data in chapters 4.1.1 to 4.1.5 showed a clear overestimation of the measured production by the simulation. This overestimation was small from March to October with 2% to 8%, while it was high from November to February ranging from 22% to 64%. Especially for January (49%) and December (64%) the simulation was a lot higher than the measured generation data. These findings correspond well to the findings of other studies, see chapter 1.1, that already found that MERRA data overestimates solar radiation and therefore will yield a higher production when simulating photovoltaic production with this data. Further the analysis showed clearly that the simulation is not able to correctly present the photovoltaic production in winter. While in spring, summer, and autumn the overestimation of 4% to 9% is very small, the simulation is almost half as high as the measured generation data in winter with an overestimation of 43%. These results were also found in other studies, mentioned in chapter 1.1, as for example Richardson & Andrews (2014) or Pfenninger & Staffell (2016a). Richardson & Andrews (2014) explained this overestimation with too little sunshine in the winter months, which increases the relative error between MERRA data and measured measurements. Pfenninger & Staffell (2016a) argued that weather events, that change the quantity of solar radiation arriving on the earth surface, are not modeled properly by MERRA. Furthermore they stated that topography is not included in the model and clouds are modeled inaccurate, which leads to an overestimation of clear-sky days. A comparison of the single winter periods from 2010 to 2013 showed that not all winters are overestimated that high. While the winter period of 2011/12 is simulated quite well with an overestimation of only 15%, the other two periods are clearly overestimated by 39% (period of 2010/11) and by 83% (period of 2012/13). A further analysis on an hourly and daily basis showed that the average hourly and daily production over a year is simulated quite well but again the hours and days in winter are overestimated.

Based on these results question three, regarding the suitability of MERRA 2 data for photovoltaic production simulation, can be answered with a yes. The MERRA 2 data, especially in combination with the solaR package of the statistics software R, is a suitable source for the simulation of photovoltaic production considering that the simulation results depend highly on the settings of inclination and orientation of the photovoltaic panel, which are often unknown. Furthermore a correction factor of 92% was applied to the simulation results to adjust the results to the measured generation. Such a correction factor was also used by Pfenninger & Staffell (2016a) as they also found that MERRA slightly overestimates the measured production. After applying the correction factor, the overestimation from March to October almost vanished or turned into an underestimation, ranging from -6% to 0%. Also the overestimation in the winter months was reduced to 13% to 51%.

The research question regarding extreme values and anomalies over a 35 year period can be answered with the results of the time series analysis in chapter 4.2. An investigation of the minimum, maximum, and average daily and hourly capacity factor showed that there are no big anomalies in the development over a day or year. The development over a year is as expected, as production is low in winter and rises in summer. The same is true for the daily development, where production is low in the morning and afternoon, while it is highest at midday. The comparison to Pfenninger & Staffell (2016a) in chapter 4.3 showed that these results are plausible. Furthermore the comparison of the results to the simulation results of Pfenninger & Staffell (2016a) showed that the authors overestimate the photovoltaic production slightly over the whole year, although this may be due to different calculation methods and different installed capacities that were used. The authors are aware of the overestimation and found that MERRA overestimates production due to an overestimation of clear sky days and a not proper modeling of weather events.

Nevertheless the results of Pfenninger & Staffell (2016a) and the results of this thesis show that the MERRA 2 data is able to simulate the development of the production on an hourly basis over the day, and a monthly basis over the year quite well. So the MERRA 2 data is suitable for the simulation of potential photovoltaic production, but the results need to be adjusted with a correction factor, to reduce the overestimation. Further investigation should therefore concentrate on the overestimation, especially in winter, and how to eliminate this overestimation.

To conclude, the MERRA 2 data can be seen as a good source of meteorological parameters for the photovoltaic production as it is able to simulate the development of the production over the day and year quite well considering that the setting of the system in the simulation (inclination and orientation of the panel) may have a huge influence on the results and that the simulation overestimates the measured production especially in winter. That is why the use of a seasonal correction factor, instead of an annual one, may be recommended in simulation models. Due to its global availability and a data range on an hourly basis from 1980 to 2015, covering 35 years, MERRA 2 data can be used for finding appropriate sites for the installation of photovoltaic production within countries or continents.

V Bibliography

BMWi (2017). Entwicklung der Stromerzeugung und der installierten Leistung von Photovoltaikanlagen. Available from: http://www.erneuerbare-energien.de/EE /Redaktion /DE/Textbausteine/Banner/banner_photovoltaik.html (Access on 12.07.2017)

Boilley, A., and Wald, L. (2015). Comparison between meteorological re-analyses from ERA-Interim and MERRA and measurements of daily solar irradiation at surface. In: *Renewable Energy*, 75, p. 135-143.

Bonett, D.G., and Wright, T.A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. In: *Psychometrika*, Bd. 65, Nr. 1, p. 23-28.

Bosilovich, M. G., Lucchesi, R., and Suarez, M. (2016). MERRA-2:File Specification. GMAO Office Note No. 9 (Version 1.1), 73 pp. Available from: http://gmao.gsfc.nasa.gov/pubs/office_notes or https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich785.pdf (Access on 12.07.2017)

Buddensiek (2014). 0.2 terawatt-hours of solar electricity generated in 2013 in the EU In: Sun & Wind Energy – The Plattform for Renewable Energies. Available from: http://www.sunwindenergy.com/photovoltaics/802-terawatt-hours-solar-electricity-generated-2013-eu (Access on 14.07.2017)

Bundesnetzagentur (2016). Photovoltaik – Datenmeldungen und EEG-Vergütungssätze. Available from: https://www.bundesnetzagentur.de/cln_1411/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/ErneuerbareEnergien/Photovoltaik/DatenMeldgn_EEG-VergSaetze/DatenMeldgn_EEG-VergSaetze_node.html;jsessionid=5E6CD848AD341EAB7A43CA4831744204 (Access on: 14.12.2016)

Chok, N. S. (2010). Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data. University of Pittsburgh.

European Commission (2017). 2020 climate & energy package. Available from: https://ec.europa.eu/clima/policies/strategies/2020_en (Access on: 26.07.2017)

GMAO (2015). MERRA-2 tavg1_2d_rad_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Radiation Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), (Access on: 13.07.2017), doi: 10.5067/Q9QMY5PBNV1T

Juruš, P., Eben, K., Resler, J., Krč, P., Kasanický, I., Pelíkan, E., Brabec, M., and Hošek, J. (2013). Estimating Climatological Variability of Solar Energy Production. In: *Solar Energy*, Volume 98, Part C, p. 255-264.

Lucchesi, R. (2012). "File Specification for MERRA Products. GMAO Office Note No.1 (Version 2.3), 82 pp. Available from: http://gmao.gsfc.nasa. gov/pubs/office_notes or https://gmao.gsfc.nasa.gov/products/documents/MERRA_File_Specification.pdf (Access on 12.07.2017)

Maps of World (2016). Germany Latitude and Longitude Map. Available from: https://www.mapsofworld.com/lat_long/germany-lat-long.html (Access on 14.12.2016)

OECD/IEA (2015). Key World Energy Statistics 2015. OECD/IEA. Available from: http://www.iea.org/publications/freepublications/publication/KeyWorld_Statistics_2015.pdf (Access on: 15.03.2017)

Paschotta R. (2017). Sonnenenergie. Available from: https://www.energie-lexikon.info/sonnenenergie.html (Access on: 25.08.2017)

Pawson, S. (2012). GMAO MERRA: Modern Era Retrospective-Analysis for Research and Applications. Available from: http://gmao.gsfc.nasa.gov/research/merra/intro.php (Access on: 14.12.2016)

Perpiñán Lamigueiro, O. (2012). solaR: Solar Radiation and Photovoltaic Systems with R. In: *Journal of Statistical Software*, 50(9), p. 1-32. Available from: http://www.jstatsoft.org/v50/i09/ (Access on: 26.07.2017)

Perpiñán Lamigueiro, O. (2016). Package 'solaR'. Available from: https://cran.r-project.org/web/packages/solaR/solaR.pdf (Access on: 13.07.2017)

Pfenninger, S., and Staffell, I. (2016a). Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. In: *Energy*, 114, p. 1251-1265.

Pfenninger, S., and Staffell, I. (2016b). Supplementary material for: Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. ETH Zürich and Imperial College London.

Richardson, D. B., and Andrews, R. W. (2014). Validation of the MERRA Dataset for Solar PV Applications. In. *Photovoltaic Specialist Conference* (PVSC), IEEE 40th.

SolarGis (2016). Global Horizontal Irradiation (GHI) Germany. Available from: http://solargis.com/products/maps-and-gis-data/free/download/germany (Access on: 14.12.2016)

UNFCCC (2015). Historic Paris Agreement on Climate Change: 195 Nations Set Path to Keep Temperature Rise Well Below 2 Degrees Celsius. Available from: http://newsroom.unfccc.int/unfccc-newsroom/finale-cop21/ (Access on: 05.08.2017)

United Nations (2017). Multilateral Treaties Deposited with the Secretary-General, Chapter XXVII Environment, 7.d. Paris Agreement. New York. Available from: https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7d&chapter=27&clang=_en (Access on: 05.08.2017)

Wetter.de (2016). Klima für Deutschland. Available from: http://www.wetter.de/klima/europa/deutschland-c49.html (Access on: 14.12.2016)

Wirth, H. (2017). Aktuelle Fakten zur Photovoltaik in Deutschland. Fraunhofer ISE. Available from: https://www.ise.fraunhofer.de/content/dam/ise/de/documents/publications/studies/aktuelle-fakten-zur-photovoltaik-in-deutschland.pdf (Access on: 26.07.2017)

Yi, Y., Kimball, J. S., and Jones, L. A. (2011). Evaluation of MERRA Land Surface Estimates in Preparation for the Soil Moisture Active Passive Mission. In: *Journal of Climate*, 24, p. 3797-3816.

Zahoransky, R., Allelein, H.-J., Bollin, E., Oehler, H., Schelling, U., and Schwarz, H. (2012). Energietechnik: Systeme zur Energieumwandlung. Kompaktwissen für Studium und Beruf. 6., Überarb. Erw. Aufl. 2013. Springer Vieweg.

VI Appendix

```
#SIMULATION OF PV-PRODUCTION WITH PACKAGE SOLAR AND MERRA-2
DATA

#This code is to import and read out the MERRA-2 data, that has
#been downloaded before and to transform it into the necessary
#format/data frame to do a simulation of the pv-production
#for a certain location/country with the package SolaR for a
#certain location.

#### PREPARATION ####

# Install necessary packages
install.packages("ncdf4")
install.packages("solaR")

# Open the necessary package for reading the NCDF File, if not
# installed, you need to install the package first
library(ncdf4)
library(solaR)

# Set Working Directory to folder where the MERRA data is stored
setwd("Folder/1991")

# Open one of the MERRA files to read the data and see which
# dimensions and variables are in the file.
# nc_open must contain the name of one file

ncfile<-nc_open("svc_MERRA2_100.tavg1_2d_rad_Nx.19910112.nc4")
ncfile

## 2 Variables -> SWGDN (surface_incominng_shortwave_flux, W/m2)
and TS (temperature, K) are important

#-----------------------------------------------------------
```

```
#### READING FUNCTIONS FOR ONE MERRA FILE ####
#later they are applied to all files

#Function for reading the temperature out of the ncfile
readts <- function(ncfile) {
  ncfile <- nc_open(ncfile)
#Read out dimensions time, longitude and latitude
  time <- ncvar_get(ncfile, "time", verbose=FALSE)
  ntime <- dim(time)
  lon <- ncvar_get(ncfile, "lon", verbose=FALSE)
  nlon <-dim(lon)
  lat <- ncvar_get(ncfile, "lat", verbose=FALSE)
  nlat <- dim(lat)
#Read out temperature data and transform into matrix
  ts.array <- ncvar_get(ncfile, varid="TS", verbose=FALSE)
  ts.vec.long <- as.vector(ts.array)
ts.mat <-matrix(ts.vec.long, nrow=nlon * nlat,ncol=ntime) #
Matrix of temperature with hours in columns and MERRA points
(lon+lat) in rows
return(ts.mat)
}

#Function for reading the solar radiation out of the ncfile
readswgdn <- function(ncfile) {
  ncfile <- nc_open(ncfile)
#Read out dimensions time, longitude and latitude
  time <- ncvar_get(ncfile, "time", verbose=FALSE)
  ntime <- dim(time)
  lon <- ncvar_get(ncfile, "lon", verbose=FALSE)
  nlon <-dim(lon)
  lat <- ncvar_get(ncfile, "lat", verbose=FALSE)
  nlat <- dim(lat)
#Read out solar radiation data and transform into matrix
  swgdn.array <- ncvar_get(ncfile, varid="SWGDN",
verbose=FALSE)
  swgdn.vec.long <- as.vector(swgdn.array)
```

```
  swgdn.mat <-matrix(swgdn.vec.long, nrow=nlon *
nlat,ncol=ntime) # Matrix of solar radiation with hours in
columns and MERRA points (lon+lat) in rowsMatrix der
Solarstrahlung mit Stunden in Spalten und Punkten (lon+lat) in
Zeilen
  return(swgdn.mat)
}
#-----------------------------------------------------------
#### SIMULATION OF PV PRODUCTION ####

#Part of the SolaR package, dfIMeteo, prodGCPV

convertToPv<-function(lat,datumCET,irradiation,temperature,
inclination,azimuth)
{
  P1 <-data.frame(datumCET,irradiation,temperature-273.15)
  colnames(P1) <- c("Time","G0", "Ta")
  irradiation <- dfI2Meteo(P1,time.col = "Time", lat = lat,
source = "P1", format ='%Y/%m/%d %H:%M:%S')
  p <- prodGCPV(lat,modeRad = "bdI",dataRad = irradiation,
sample ="hour",beta=inclination, alfa=azimuth)
  p@prodI[is.na(p@prodI[,8]),8] <- 0
  final_pv <- p@prodI[,8]/25000 #kwH

  return(final_pv)
}

#----------------------------------------------------------------

#### USING SIMULATION FUNCTION ON ALL MERRA-FILES ####

#Create a list of all the MERRA Files in the folder and apply
the functions to them

for(year in 2013:2013){
  for(month in c( "08","09","10","11","12")){
      for(zone in c("amp","her","ten","tra")){
```

```
setwd(paste("Folder",year,"/",month,sep=""))

allfiles <- list.files(pattern = "*.nc")
  list.of.ts <- lapply(allfiles,readts)
  list.of.swgdn <- lapply(allfiles, readswgdn) # Listen pro Tag
von allen Files, die in Spalten die Zeit und in Zeilen die
MERRA-Punkte enthalten
  listlength<-length(allfiles)



# Make a dataframe out of the list and save dataframe as txt
df.swgdn <- data.frame(matrix(unlist(list.of.swgdn),
nrow=listlength*24, byrow=T)) #nrow = Tage * 24 Stunden
colnames(df.swgdn) <- paste("MP",1:399,sep="")
seq<-seq(as.POSIXct(paste(year,"-01-01
00:30",sep="")),as.POSIXct(paste(year,"-12-31
23:30",sep="")),"hours")
rownames(df.swgdn)<-seq[month(seq)==as.numeric(month)]
print("df.swgdn")

df.ts <- data.frame(matrix(unlist(list.of.ts),
nrow=listlength*24, byrow=T))
colnames(df.ts)<-paste("MP",1:399,sep="")
rownames(df.ts)<-seq[month(seq)==as.numeric(month)]
print("df.ts")

#write.table(df.swgdn,"Folder/Solar.txt", sep="\t")
#write.table(df.ts,"Folder/Temperature.txt", sep="\t")

###DATE SEQUENCE###

#Make a list with date

seq<-seq(as.POSIXct(paste(year,"-01-01
00:30",sep="")),as.POSIXct(paste(year,"-12-31
23:30",sep="")),"hours")
```

```
datumCET <- seq[month(seq)==as.numeric(month)]
print("date")

#---------------------------------------------------------------

  #### SIMULATION OF PV-PRODUCTION IN SOLAR ####

#If you want to calculate pv-production simulation for one grid
zone, you need to import the capacity
# of the grid zone (amp, ten, her, tra)

#Data has been provided by calculation in ArcGIS

setwd("Folder")

#Monthly capacity
amp_cap_m <- read.delim("Folder/Amp_cap_m.txt",
stringsAsFactors=FALSE)
cap_amp<-amp_cap_m[,as.numeric(month)-5+12*(year-2010)]
cap_amp<-t(cap_amp)
#View(cap_amp)

ten_cap_m <- read.delim("Folder/Ten_cap_m.txt",
stringsAsFactors=FALSE)
cap_ten<-ten_cap_m[,as.numeric(month)-5+12*(year-2010)]
cap_ten<-t(cap_ten)
#View(cap_ten)

tra_cap_m <- read.delim("Folder/Tra_cap_m.txt",
stringsAsFactors=FALSE)
cap_tra<-tra_cap_m[,as.numeric(month)-5+12*(year-2010)]
cap_tra<-t(cap_tra)
#View(cap_tra)

her_cap_m <- read.delim("Folder/Her_cap_m.txt",
stringsAsFactors=FALSE)
cap_her<-her_cap_m[,as.numeric(month)-5+12*(year-2010)]
```

```
cap_her<-t(cap_her)
#View(cap_her)

# Calculate pv-production for one grid zone and all MERRA points
belonging to this grid zone

pv_final<-list()

cap<-NULL
if(zone=="her"){
  cap<-cap_her}

if(zone=="amp"){
  cap<-cap_amp}
if(zone=="ten"){
  cap<-cap_ten}

if(zone=="tra")
{
  cap<-cap_tra}

for(i in 1:ncol(df.swgdn)){
  print(i)
  lat <-lonlat[i,2]
  pv_final[[length(pv_final)+1]]<-
convertToPv(lonlat[i,2],datumCET,df.swgdn[,i],df.ts[,i],15,45)*c
ap[i]
  }



#convert list to dataframe
df<-
data.frame(matrix(unlist(pv_final),nrow=length(pv_final[[1]]),by
row=F))
final_sum<-apply(df,1,sum)
```

```
#Save calculated production per grid zone per year per month in
separate .txt

write.table(final_sum,paste("Folder/15° Tilt 45°
West/",zone,year,month,".txt",sep=""), sep="\t")
}
}
}
```