MSc Program
Renewable Energy in Central and Eastern Europe

Compilation of the best regional 3D clustering model for forecasting of PV electricity generation respecting the geographical and climatological specifics of the location of PV installations (on the example of Styria/Austria)

A Master's Thesis submitted for the degree of
"Master of Science"

supervised by
FH-Prof. DI Hubert Fechner, MAS, MSc

Michaela Leonhardt, Ph.D.

1227681

20th September 2014, Vienna

AFFIDAVIT

I, **Michaela Leonhardt** hereby declare

1. that I am the sole author of the present Master Thesis, "Compilation of the best 3D clustering model for forecasting of PV electricity generation respecting the geographical and climatological specifics of the location of PV installations (on the example of Styria/Austria)", 54 pages, bound, and that I have not used any source or tool other than those referenced or any other illicit aid or tool, and
2. that I have not prior to this date submitted this Master Thesis as an examination paper in any form in Austria or abroad.

Vienna, _______________ ____________________________

Date Signature

ABSTRACT

Assuming that the number of PV installations in Austria will further increase, a well performed area forecast of PV electricity generation will gain in importance especially for the grid operating companies. To create the forecast as precise as possible contributes to a successful integration of higher amount of PV into the electricity system.

The main focus of this master thesis is to compile the best 3D regional clustering model for forecasting of PV electricity generation based on the geographical position of the PV installations and reflecting the micro-climatological regional specifics, and to evaluate the potential for increasing the accuracy of PV electricity generation forecast in the future. This is applied on the example of the federal state of Styria. Due to its mountainous landscape, defining the clusters by the height above mean sea level, is assumed to be the key-criteria. Particularly, this is connected to various micro-climatic zones and reflects the important input factor characteristics, which influence the PV electricity output, like the intensity of global horizontal irradiance, ambient temperature, frequency of cloud cover, fog and inversion.

To be able to evaluate the influencing factors and their interactions, various prediction models are defined and proved. The forecast is based on the regressions analysis and the statistical MARS-model (multivariate adaptive regression splines). The criteria of a normalised Root Mean Square Error (nRMSE) referring to installed capacity (kWp) is used for verification of the accuracy of the prediction models.

Finally, the best results are achieved if the region is divided in two (poss. three) clusters according to the position of the PV systems above sea level defined by 500m (poss. by 1.000m in addition) and overall 6 representing weather measuring stations (providing the forecasted values of global irradiance and ambient temperature) are taken into account.

The results confirmed that the idea of finer clustering, respecting the geographical and climatological specifics of the 3D location of PV installations, opens the door to an improvement of the forecast accuracy in the future. With an increasing number of clusters, necessarily the number of representative weather measuring stations has to be increased as well. Due to the additional meteorological information, the local variations are reflected more precisely, which leads to more exact forecasts results.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Motivation

Increasing amount of renewable energy sources in the electricity system is both the European and the Austrian trend in the last years. In Austria, next to well established hydro power plants, also wind parks and photovoltaic gain in importance.

In case of electricity production by photovoltaic, especially the fluctuating energy output at all timescales (from seconds to months over the whole year) and uncertainty in dependence on current weather conditions bring new challenges for grid operation by keeping the balance between electricity demand and supply at any time.

The yearly sum of global irradiance (GI) in Austria varies significantly especially in correlation with the specific geographical conditions (e.g. mountainous vs. flat areas, cloudy or foggy areas). Depending on the region, the average value of global horizontal irradiance in Austria is between 1.050 and 1.250 kWh/m^2/year, which has a direct impact on the energy output of a PV system. As explained later on in this master thesis, this variability is clearly noticeable even within one relatively small area of a federal state.

Since we can expect further increasing amount of PV installations in the next years in Austria, it will be more and more important for the grid operators to deal with this specific issues. To enable the integration of higher amount of photovoltaic into the electricity system, a well performed area forecast of PV electricity generation is essential.

Reflecting the experience from everyday grid operation in last years – working by the Austrian transmission system operator (TSO) Austrian Power Grid AG – the ever-increasing requirements to the accuracy of the forecast, especially in case of RES generation, are more than evident.

1.2 State of the art

Based on the amendment to the Green Electricity Act in Austria in 2006, "*Abwicklungsstelle für Ökostrom AG*" (OeMAG) was established as a company which has by the law the obligation to take over the green electricity produced from installations which are registered by the "*Öko-Bilanzgruppe*" (Öko-BG) [Eco Balance Group].[1] To fulfil this requirement, the electricity production of these PV installations has to be forecasted and scheduled regularly. Based on the service agreement between OeMAG and Austrian Power Grid AG (APG), APG takes over the forecasting and scheduling process for the Öko-BG.

In case of photovoltaic, taking the current overall installed capacity and the distribution of PV installations in Austria into consideration, the forecasting process of the electricity generation for the Öko-BG is done on the federal state basis. This forecast is based on meteorological data, statistical methods and PV system data. In the first step, regressions model and MARS model (multivariate adaptive regression splines) are applied taking into consideration the standardised meteorological values. In the second step, the overall installed capacity is taken into account, which gives the absolute values. Finally, the best estimation gives the day-ahead (D+1 and D+2) PV electricity generation forecast (kW).[2] The whole process is described in the graphic below:

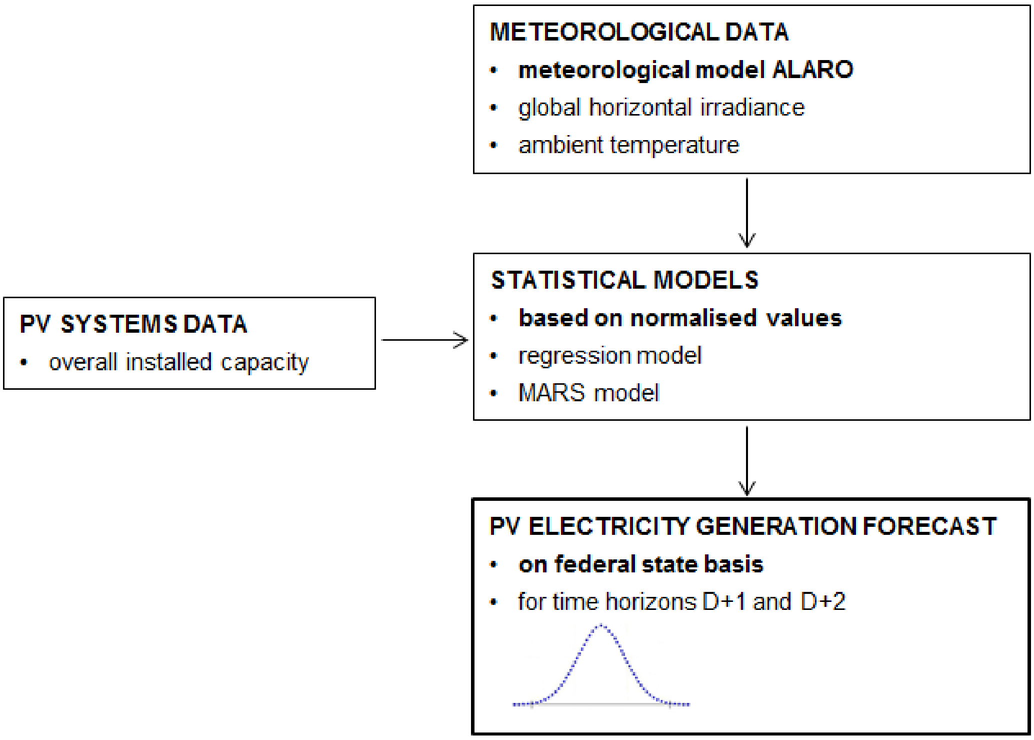


Figure 1.2: Main steps of the PV electricity generation forecasting process. (Own figure)

[1] For a period of time and fix specified price which are both determined by the law. Free according to http://www.oem-ag.at/de/home/ (18.05.2014). See this page also for further information on current Austrian Green Electricity Act and feed-in tariff regulation for PV installations (http://www.oem-ag.at/de/gesetze-regelwerk/).

1.3 Core question and main objectives

From today's perspective, regarding to the current overall installed capacity and the distribution of PV installations in Austria (as mentioned above), the classification based on a federal state basis suits very well and provides satisfactory results. With further increasing number of PV installations, the question on finer clustering in connection with the level of accuracy of PV electricity generation forecast will arise in the future. Then, especially the location of PV system(s) should be considered more in detail. Due to the Austrian mountainous landscape which is characterised by different micro-climates, the height above mean sea level – e.g. in connection with the intensity of global horizontal irradiance and ambient temperature as well as with frequency of cloud cover, fog and inversion – seems to be a significant factor for the accuracy of PV electricity generation forecast. In this context, the aim of this master thesis is to take the geographical position of the system (3D) and the micro-climatic conditions into account and to compile a finer sub-regional clustering model based on these additionally information. This is applied on the example of PV systems which are located in federal state of Styria, registered by the Öko-BG and provide the real measured data of their generation by a "feed-in profile meter" at least over a period of 1 year.[3] Both the size of the geographic area and the number of PV installations considered, contribute statistically to the reduction of forecast inaccuracy and base one of the main challenges by composing appropriate clusters. Naturally, based on the best optimal cluster estimation, meteorological data from several measuring stations for representatively selected sub-regions have to be implemented into the modelling and forecasting process.

In this context, the core questions of this master thesis are defined following:

How to compile the best regional 3D clustering model for forecasting of PV electricity generation respecting the geographical and climatological specifics of the location of PV installations, on the example of PV systems which are located in the

[3] The electricity generation data of these PV systems are provided on monthly basis in 15-minutes raster. Free according to http://www.oem-ag.at/de/oemag/service/faqs/ (18.05.2014). The relevant date of this 1 year period is defined by the date of registration by Öko-BG by 30th April 2013 at the latest. For the selection of an appropriate federal state, according to the basic conditions, the number of PV systems and the total installed capacity (kWp) was evaluated for each federal state in Austria. The best proportion of these parameters was identified in case of federal state of Styria (the highest total installed capacity and the second highest number of "feed-in profile meters"). Additionally, the region of Styria provides geographically and climatologically variable conditions, which matches the focus of this study.

federal state of Styria, registered by the "Öko-Bilanzgruppe" [Eco Balance Group] and provide the real measured data of their generation by a "feed-in profile meter" at least over a period of 1 year? How is the forecast accuracy of different prediction models in dependence on number of defined (sub-)clusters and number of weather measuring stations taken into account?

Additionally, following points are in close relation to this topic and also considered:

- Evaluation of the density of PV systems in different geographical sub-regions.
- Reflection of some specific issues regarding the forecast accuracy, e.g.:
  - What factors influence a forecast quality?
  - Is there a link between meteorological forecast, seasonality and PV power output forecast?
  - What are the possibilities and challenges by up-scaling of PV electricity generation forecast.

2 PV ELECTRICITY GENERATION FORECAST

2.1 Definition and application of forecast

Related to the grid operation, a generation forecast announces the expected amount of electricity fed into the grid at a specific timeframe (typically 15-minutes, hourly and daily) and provides an important information for the system operators, who are responsible for keeping the balance between electricity demand and supply at any time. In case of renewable energy sources (RES), regarding to their dependence on current weather situation and unscheduled occurrence of generation, the forecasting process is the most challenging task in comparison to conventional energy sources.

Following graphics emphasise the great seriousness of this topic and the necessity to create the PV electricity generation forecast as precise as possible, which contributes to a successful grid integration of PV systems.

First, a symbolic 24-hours profile of electricity consumption and PV electricity production (see the graphic below) illustrates the challenges for grid operation by keeping the electricity system in balance. The green line is for PV electricity production, the red line for electricity consumption. The grey areas show periods when the electricity consumption is higher than the PV production, which means the system operator has to supply the required electricity amount by substitution through other energy sources. The yellow area on the top illustrates a period when the PV electricity production is higher than the electricity consumption, which means the grid operator has to take care about the surplus electricity fed into the grid.

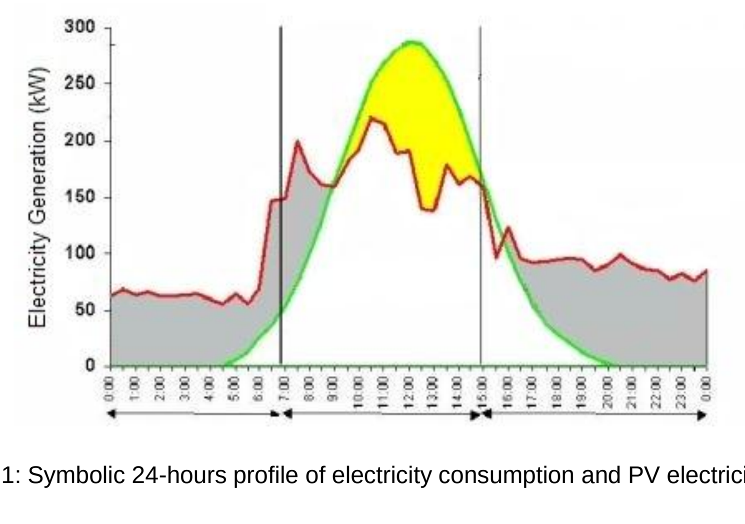


Figure 2.1: Symbolic 24-hours profile of electricity consumption and PV electricity production.[4]

[4] http://www.pvaustria.at/wp-content/uploads/2013/07/286_Energieprofil.jpg (07.06.2014), own adaptation.

Next, the relatively low correspondence of the average yearly models of electricity consumption and PV electricity production – especially day vs. night, summer vs. winter intensity – is more than evident, if the graphics below are considered.

The electricity consumption increases from the dark green colour to the dark red colour with the peaks in the mid-day, but also in the morning and evening hours, especially in the winter months.

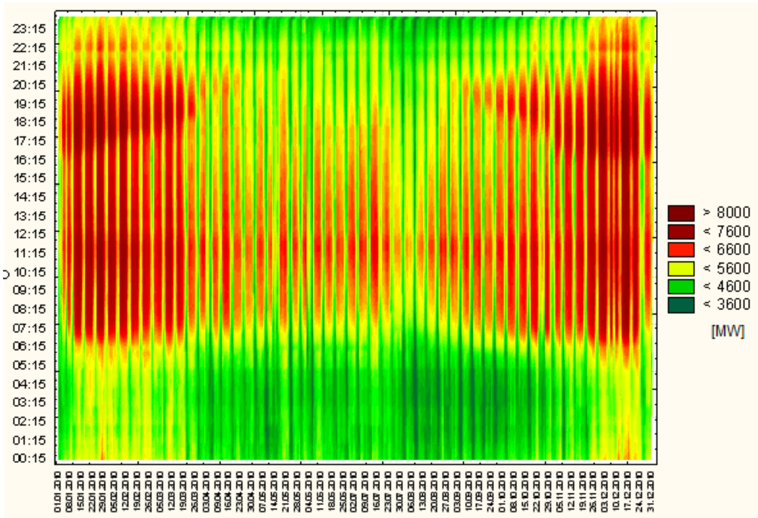


Figure 2.2: Electricity consumption model (yearly average).[5]

The PV electricity production increases from the dark green colour to the dark red colour with the peaks at the mid-day hours and in the highest intensity between March and September.

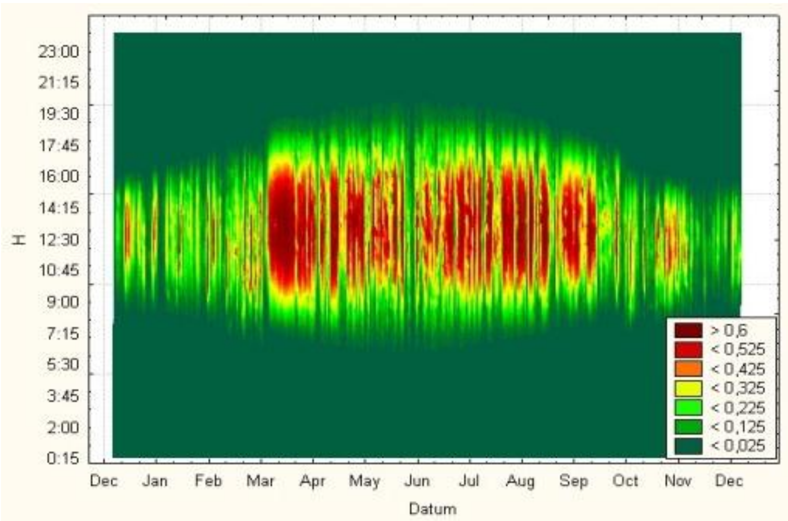


Figure 2.3: PV electricity production model (yearly average).[6]

[5] APG, internal company presentation, unknown author.
[6] APG, internal company presentation, unknown author.

2.2 Important factors

Lots of factors of technical and natural character affect the energy yield of PV systems with different importance. In this part, the influenced factors are listed and the decision which of them can and can't be applied reasonably in terms of this study is done. Since the number and quality of input parameters are very relevant for the accuracy of the forecast, this decision has a crucial role.

Primarily, the energy yield of a PV system depends significantly on two factors – global irradiance and PV-module temperature.

2.2.1 Global irradiance (GI)

The global irradiance is the amount of solar power falling on a horizontal surface in a specific area (consists of direct irradiance and diffuse irradiance and is measured in W/m^2) and gives an idea of "intensity" of the sunshine. The exact values fluctuate quite high in dependence of the period of a day and year (day vs. night, during the day, summer vs. winter) and weather conditions (e.g. cloud cover, wind, snow, fog and inversion). The sum of this solar power over a period of time gives the energy yield (kWh/m^2). In Austria, the average sum of global horizontal irradiance depending on the region is between 1.050 and 1.250 kWh/m^2/year.[7] The fluctuation of the energy yield of a PV system between years is relatively low.[8]

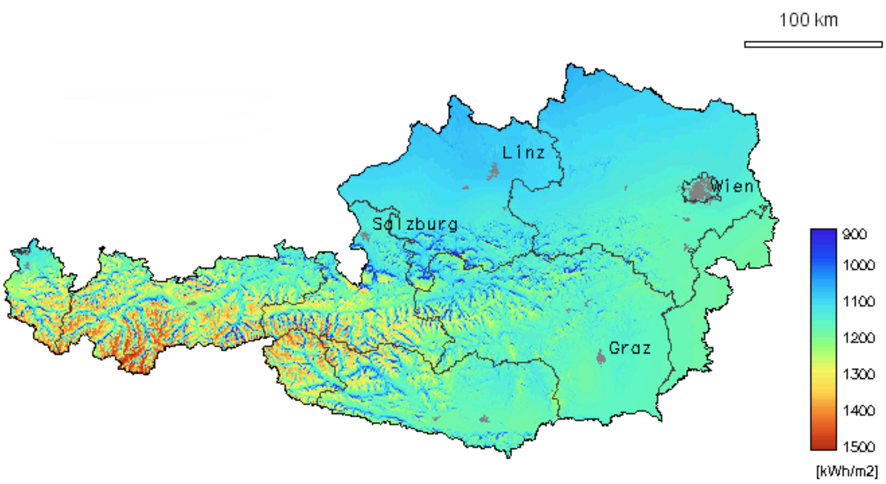


Figure 2.4: Austria, yearly sum of global horizontal irradiance.[9]

[7] Free according to Häberlin (2010), p. 11, p. 52 to 59.
[8] According to http://www.pvaustria.at/daten-fakten/technologie/pv-auslegung/ (07.06.2014).
[9] http://re.jrc.ec.europa.eu/pvgis/countries/europe/g13y_at.png (14.07.2014).

The strong dependence between the intensity of global irradiance and the PV-module energy output illustrates the graphic below. This dependence is nearly linear; if the GI (W/m^2) doubles, also the energy output is twice as high.[10]

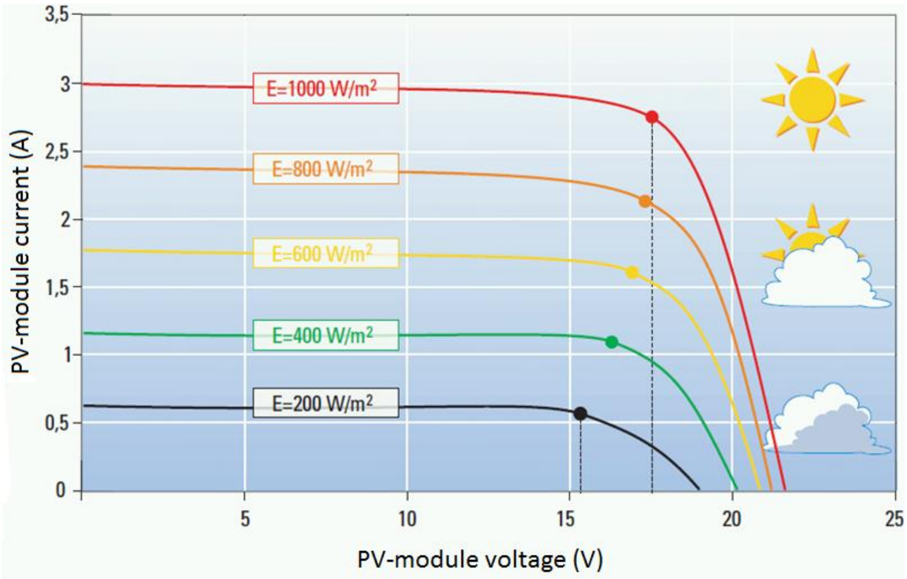


Figure 2.5: Influence of global irradiance on PV-module energy output.[11]

The global irradiance is one of the most significant factors for the PV electricity generation forecast and is taken as very relevant parameter in case of this study.

2.2.2 Temperature

In general, the energy output of a PV system decreases with increasing PV-module temperature (see the graphic below). In this context, the ambient temperature and ventilation of the PV system (see later under "mounting") influence the efficiency. Taking into consideration the ambient temperature (T_a), the temperature coefficient (c) and the global irradiance at the module surface (GI_{mod}), the PV-module temperature can be calculated as: $T_{mod} = T_a + c \cdot GI_{mod}$.[12] The temperature coefficient describes the behaviour of a PV-module with changing temperature. E.g. in case of crystalline PV-modules the temperature coefficient is round minus 0,4% per $^{\circ}$C, which means, with each 10°C increasing PV-module temperature the

[10] Free according to http://www.solaik.ch/photovoltaik/grundlagen/index.php# (14.07.2014).
[11] http://www.solaik.ch/photovoltaik/grundlagen/index.php# (14.07.2014), own adaptation.
[12] According to Drews (2007), p. 554. The temperature coefficient can vary e.g. in relation with material of solar cells, mounting of the PV system (ventilation).

power output decreases by round 4%. In very hot summer months, the module power output can be reduced by up to 30% or higher, depending on the mounting.[13]

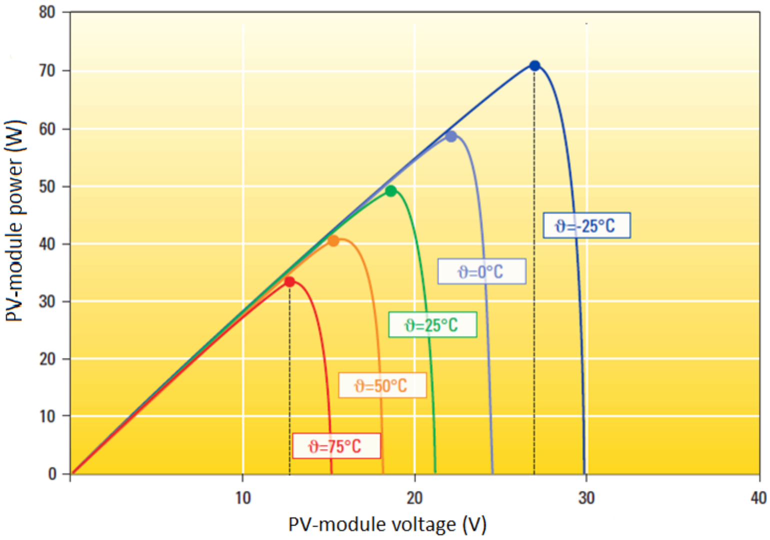


Figure 2.6: Influence of PV-module temperature on PV-module power output.[14]

In practice, it is not possible to monitor the temperature of each PV-module for the purpose of the wide area PV electricity generation forecast. Instead of that, the ambient temperature (T_a) of representative selected weather measuring stations (WMSs) is taken as a relevant parameter in case of this study.

2.2.3 Geographical position of the PV system

Especially in relation to the global irradiance and next to the energy yield, the geographical position of the PV system is a very important factor. In general, the yearly GI value increases towards the south. Also the height above sea level – basically the thickness of air mass, which the sun rays have to go through – and local climatic conditions influence this value crucially.[15]

The geographical location of the PV systems – specified by latitude and longitude coordinates (°) and meters above sea level (m) – is considered in special in this master thesis, if its focus is to compile the best regional 3D clustering model for forecasting of PV electricity generation respecting the geographical and climatological specifics of the PV installations.

---

[13] Free according to http://www.solaik.ch/photovoltaik/grundlagen/index.php# (14.07.2014). An example: by the 70˚C PV-module temperature and the PV-module temperature coefficient minus 0,4% per ˚C, the electricity power output will be reduce by approx. 28% (70 x -0,4).
[14] http://www.solaik.ch/photovoltaik/grundlagen/index.php# (14.07.2014), own adaptation.
[15] According to Podesser (2010), p. 4, and http://de.wikipedia.org/wiki/Globalstrahlung (07.06.2014).

2.2.4 Mounting

The energy yield of a PV system depends on its orientation and inclination. The maximal energy yield – in mid-Europe – is achievable at 30° inclination and south orientation.

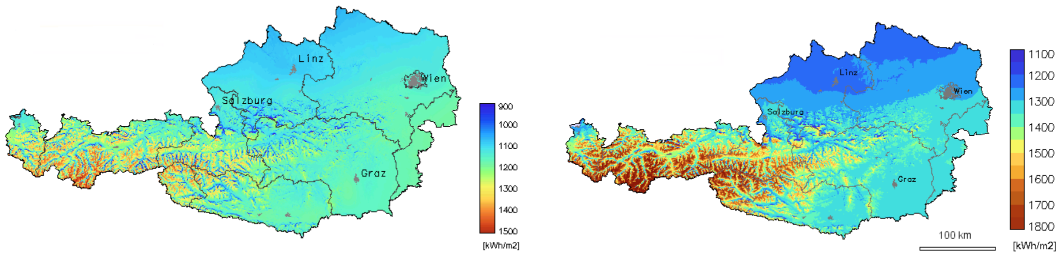


Figure 2.7: Austria, comparison of yearly sum of global horizontal irradiance (left) and global irradiance received by optimally inclined PV modules, at 30° inclination and south orientation (right).[16]

As shown in the graphic below, the energy yield (in its yearly average) is still very high by inclination in the range of 0° to 50°, the reduction varies from 5% to 25%. Next, the ± 45° deviation from the south decrease the energy yield by only 5% to 10%.[17]

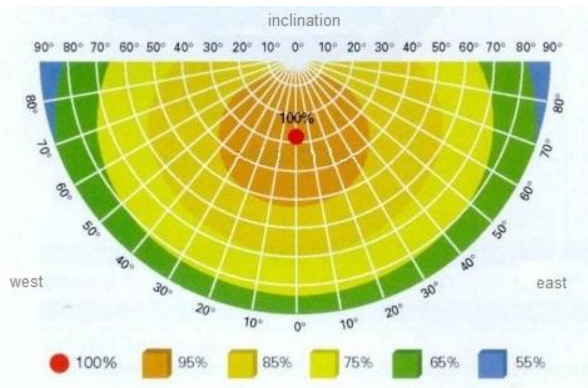


Figure 2.8: The energy yield of a PV system in dependence of its orientation and inclination (referring to the yearly average).[18]

Further, the type of mounting – ground based mounting, roof and facade mounting or integration, influence the energy yield especially in relation to ventilation. As described above, the module temperature increase can cause the energy yield reduction (for examples see the table below).

[16] http://re.jrc.ec.europa.eu/pvgis/countries/europe/g13y_at.png (14.07.2014), http://re.jrc.ec.europa.eu/pvgis/countries/europe/ffg13.optyc.at.png (14.07.2014).
[17] Free according to http://www.pvaustria.at/daten-fakten/technologie/pv-auslegung/ (07.06.2014).
[18] http://www.pvaustria.at/wp-content/uploads/2013/07/160_Energieertrag-Ausrichtung-Dachneigung.jpg (07.06.2014), own adaptation.

Mounting	Temperatur-increase K	Reduction in energy yield %
Facadeintegration (without ventilation)	55	8,9
Roofintegration (without ventilation)	43	5,4
Facade (insufficient ventilation)	39	4,8
Facade (good ventilaton)	35	3,9
Roof (insufficient ventilation)	32	2,6
Roof (good ventilation)	29	2,1
On roof (no integration)	28	1,8
Free mounting	22	0,0

Table 2.1: Temperature increase (K) and reduction in annual yields (%) at different mountings.[19]

From the investors perspective, to ensure the best optimal mounting is an important factor for the economy of each single PV system. As the experience shows, the energy yield varies mostly by not more than ± 5% to 10% (in a yearly average). In this study, the mounting of considered PV systems is an unknown parameter. Regarding to their installation for a "business" purposes, the (nearly) optimal mounting and therefore only small deviations in energy yield can be expected, which have rather negligible impact referring to statistical distribution by wide area forecast.

2.2.5 Operation mode (fixed or tracked)

The position of sun varies depending on the time of a day and season. A solar tracker always adjusts the PV panel to the sun. Depending on meteorological conditions, a tracked PV system – compared to an optimal mounted system – can increase the annual energy output approx. up to 20% to 25% (for Central Europe).[20] The necessary additional effort – especially acquisition and maintenance costs – is currently still very high, which has a prolonging effect to the amortisation time of the PV installation. Due to this fact, the percentage of tracked installations in Austria is still very low; therefore this factor is neglected in case of this study (it is also an unknown parameter in case of considered PV systems).

2.2.6 Shadowing effects

There are different reasons for shadowing effects – just to remind the typical ones following examples should be mentioned – shading caused by neighbouring buildings, chimneys, trees, masts, antennas, other PV systems, hills and mountains

[19] According to http://www.sev-bayern.de/content/Objektforum.pdf (14.06.2014), Becker G., *Gebäude-integrierte Photovoltaik (GIPV/BIPV)*, presentation, p. 19, p. 23.

[20] According to Häberlin (2010), p. 43, p. 172 to 173, and http://www.pvaustria.at/daten-fakten/technologie/pv-auslegung/ (14.06.2014).

etc. In practice, shadings are the major reason for deficient in energy output of PV systems and should be avoided already in the planning phase. In case of wide area forecast, it is not possible to monitor the shadowing effects. Basically, these effects appear as a part of forecast deviation.

2.2.7 Performance ratio (PR)

Next to various kinds of shadowing effects, which can cause very high percentage of losses, and temperature losses, there are other factors causing additional loses for the PV system, which affects that the real energy yield of a PV system is lower than a reference energy yield. To evaluate the quality of a PV system, the performance ratio (PR) is an important value, which respects all kind of losses and is defined as a convergence of real and ideal case. In this context following factors should be mentioned – material (laboratory vs. real conditions), components and conversion losses (invertor), losses due to mismatching (serial connected PV-cells and PV-modules), cabling (DC/AC), reflexions at the module surface, dirt and ageing process etc. For a well-designed PV system the PR coefficient should reach the value higher than 85%. In general, the exact value is not predictable and applicable in practice, especially in terms of wide area forecast and appears as a part of forecast deviation.[21]

2.3 Meteorological models and data

Appropriately chosen, precise and complete numerical weather prediction model (NWPM) is the crucial point for the quality of a PV electricity generation forecast. In case of this study, the weather prediction model ALARO, which is the newest version of model ALADIN (Aire Limitée Adaptation dynamique Développement InterNational), is used for modelling of the day-ahead PV forecast. It gives a NWP for Central and East Europe in resolution of 5 x 5 km considering the regional effects. This model is available for next 72 hours in hourly raster. An update of these data occurs 4 times a day (every 6 hours).[22] It provides information about global irradiance, ambient temperature, cloud cover, probability of precipitation, relative humidity, wind speed and direction. As discussed in chapter 2.2, global irradiance and ambient temperature are the weather variables, which are defined as relevant for a PV electricity generation forecast in accordance with objectives of this study.

[21] Free according to Häberlin (2010), p. 485 to 487.
[22] The creation time is 00:00 UTC (model ALA00) and provides the meteorological data for the next 72 hours. Free according to http://www.zamg.ac.at/cms/de/forschung/wetter/alaro (14.05.2014).

Further, the global model ECMWF (European Centre for Medium-Range Weather Forecasts) offers the weather prediction up to 15 days ahead in hourly raster (updated twice a day) with resolution 25 x 25 km. This model could be foreseen for the medium-range PV forecast in the future. In addition, the relatively new developed INCA (Integrated Nowcasting through Comprehensive Analysis) model provides an hourly updated regional prediction for Austria in 1 x 1 km raster.[23] In the future, this model could be foreseen for the short-term (now-casting, hour-0 to hour-6) PV forecast.

The Austrian Institute for Meteorology "*Zentralanstalt für Meteorologie und Geodynamik*" (ZAMG) provides both the weather prediction models mentioned above and real measured values by ZAMG measuring stations.[24] The real measured values are available in 10-minutes raster.

Following ZAMG weather measuring stations (WMSs) are used for the purpose of this study (see the map below):

- Bad Radkersburg (SL 216m), Fürstenfeld (SL 267m) and Graz Universität (SL 379m) for cluster 1,
- Leoben (SL 545m) and Seckau (SL 872m) for cluster 2, sub-cluster 2A,
- Stolzalpe (SL 1.289m) for cluster 2, sub-cluster 2B.

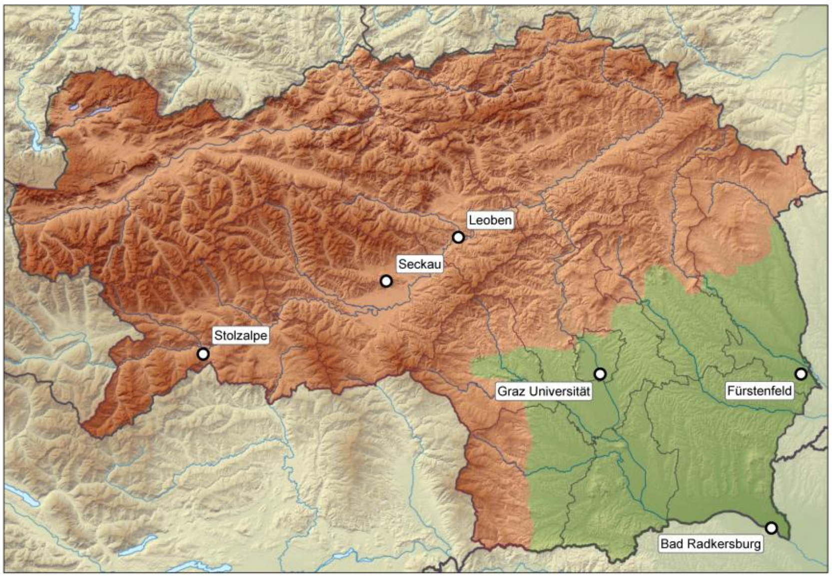


Figure 2.9: The regional 3D clustering model of Styria. (Own figure)

[23] Free according to http://www.zamg.ac.at/cms/de/forschung/wetter/inca (14.05.2014).

[24] The necessary meteorological data are provided by ZAMG (resp. in case of real measured values in cooperation with TU Graz) via APG.

2.4 Forecast horizons and procedures

In case of PV electricity generation, the day-ahead time frame is the forecast horizon applied in practice. In context of this master thesis, basically the horizons from hour-24 to hour-48, which provide the forecast for the next day (D+1) and the day after the next day (D+2), are defined as a relevant day-ahead timeframes and taken into consideration.[25] The forecast provides predicted values on 15-minutes basis (in kW) and is updated every 6 hours (respecting the updates of the weather prediction models).

The forecast is based on the regressions analysis and the statistical MARS-model (multivariate adaptive regression splines), which takes into account one dependent variable (in this case electricity production) and several influencing factors (in this case global irradiance and ambient temperature) plus historical generation data.

Mainly Microsoft Excel and STATISTICA (StatSoft) are used for collecting, statistical processing and visualisation of the data. The software tool mP Energy (Meta Predictor Energie, Metalogic) provides the prediction models based on multivariate regression (MARS-model) and enables to represent the non-linear influence factors like in case of various meteorological data. The software tool TSM (Time Series Manager, Hakom) is used for time-sets processing (e.g. transformation of dissimilar time-intervals)[26] and for establishing the access between databases with energy and meteorological data and mP Energy modeling tool.

2.5 System and energy data of considered PV systems

Since this master thesis is written in close cooperation with Austrian Power Grid AG (APG) and OeMAG, both of these companies give a support by providing the necessary data. The system and energy data of the considered PV systems – geographical position (address of the connection point), installed capacity (kWp), real measured data on electricity production (kWh) in 15-minutes steps on the monthly basis – are provided by OeMAG via APG. Since these are information of strict confidential nature, they are treated **anonymised** in this master thesis.

[25] The definition of a day-ahead timeframe depends on the day type (e.g. weekends, bank holidays can be treated differently).
[26] E.g. in opposite to the 15-minutes raster of the PV electricity forecast output, the weather prediction models are available in hourly raster.

3 REGIONAL 3D CLUSTERING MODEL BASED ON GEOGRAPHICAL AND CLIMATOLOGICAL SPECIFICS

3.1 Description and evaluation of the chosen sample

As already mentioned above, this study considers the PV systems which are located in the federal state of Styria, registered by the Öko-BG and provides the real measured data of their electricity generation by a "feed-in profile meter" at least over a period of 1 year. The relevant date of this 1 year period is defined by the date of registration by Öko-BG by 30th April 2013 at the latest.[27] The total installed capacity of the PV systems which match the given requirements is approx. 25MWp. Closer evaluation of these data shows that approx. 90% of the number of all considered PV systems are in a dimension less than 500kWp each, and the rest approx. 10% more (or equal) 500kWp each.[28] In comparison to this extremely high difference, the total installed capacity of each of these two groups is nearly the same.

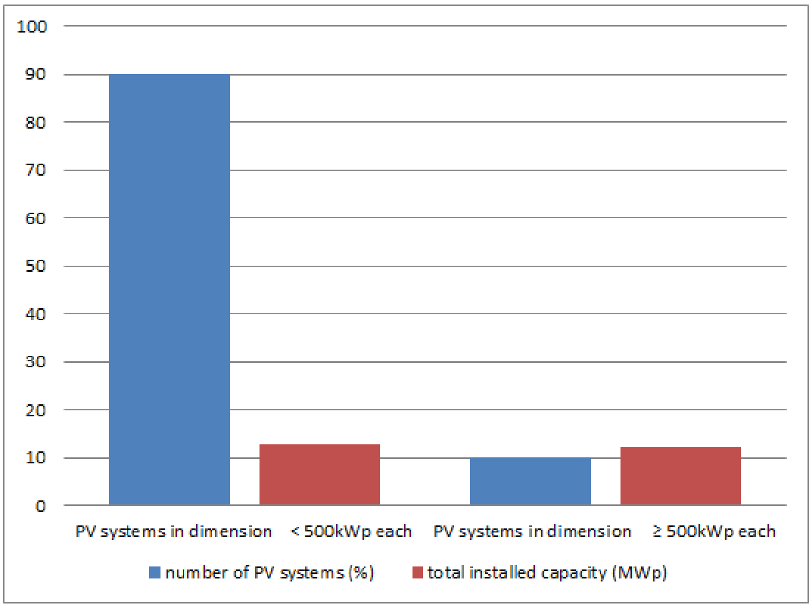


Figure 3.1: Evaluation of the chosen sample: PV systems in dimension < 500kWp each and ≥ 500kWp each – number of systems vs. total installed capacity. (Own figure)

3.2 Analysis of the geographical allocation of considered PV systems

In the next step the geographical position of the PV systems and their installed capacity (kWp) were evaluated. Since these are information of strict confidential nature, they are treated **anonymised** in this master thesis. The position of each PV

[27] The observation period for modelling of the forecast is defined from 1st May 2013 to 30th April 2014.
[28] The minimum and maximum values are approx. in the range of 4kWp and 2.900kWp.

system is defined by the address of the connection point.[29] Based on this information the geographical location was subsequently specified by latitude and longitude coordinates (°) and meters above sea level (m) with the help of the Google Earth program.[30] Visualisation of the geographical values in the map below gives the first idea about the density and size of PV systems in different geographical regions of Styria.

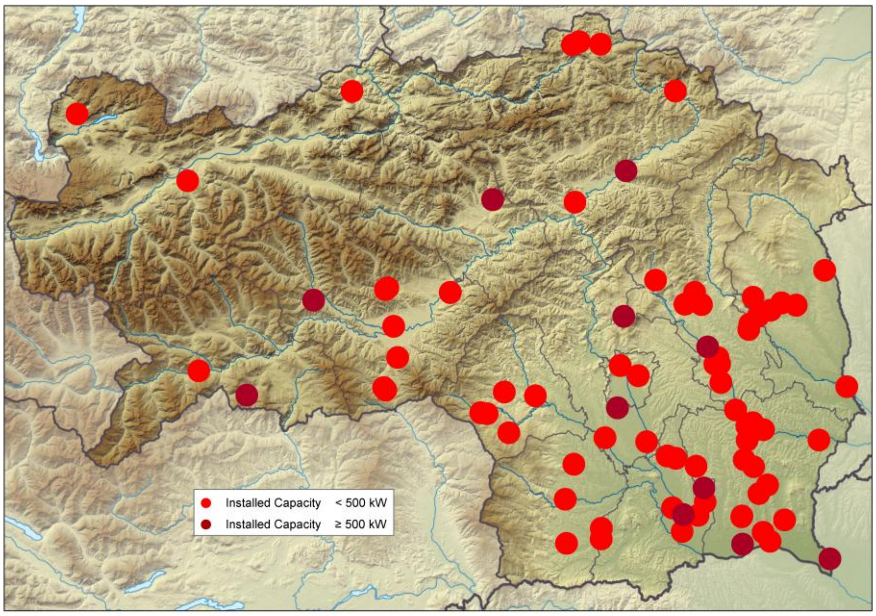


Figure 3.2: Evaluation of the chosen sample: density and installed capacity of considered PV systems in different geographical regions of Styria. (Own figure)

The variable landscape of Styria (area of 16.400 km^2) is characterised by three main zones – valleys and flat area in the south, alpine area (up to max. 3.000m) in the north, and low mountainous area in between. The evaluation of the geographical allocation of considered PV systems gives following outcomes: The lowest/highest position of a PV system is about 200m/1.900m above sea level. Respecting the micro-climatic specific of the region (see further on) and results of the correlation analysis based on real measured values of GI and PV electricity output (see later), the heights of 500m and of 1.000m above sea level are the approximate reasonable boarders for the geographically based clusters. Obviously, the highest density of PV installations is in the flat, southern part of the land. Nearly 70% of the considered PV systems are located in this area.

[29] Small deviations can occur due to the variations between the address of the connection point and exact position of the PV system; nevertheless these are mostly of minor relevance.

[30] Especially in case of the large scale PV systems (typically > 500kWp), some more precise information (e.g. sea level) are also published by the operating company on the internet. If it was the case, this information was taken into account.

3.3 Analysis of the micro-climatic conditions

As discussed in chapter 2.2, global irradiance (GI) and ambient temperature (T_a) are the weather variables, which are defined as relevant for the PV electricity generation forecast in accordance with objectives of this master thesis. In particular, the global irradiance is taken as the most significant factor for the establishing of the regional 3D clustering model and therefore analysed more in detail, since it reflects also the related aspects like cloud cover (cumulus) or mist and fog.

Basically, the irradiance intensity depends on the geographical position and the seasonal sun elevation angle as well as weather situation and varies pretty much during the year. In this context, there are some variations in GI values according to the regional micro-climatic conditions if distinguishing the flat areas and mountainous or alpine areas of Styria (see the graphic below). These regional specifics are reflected in the clustering model (compare the graphics 3.3 and 3.10; later on the cluster 1 is defined by the flat areas and the cluster 2 contains the mountainous and alpine areas).

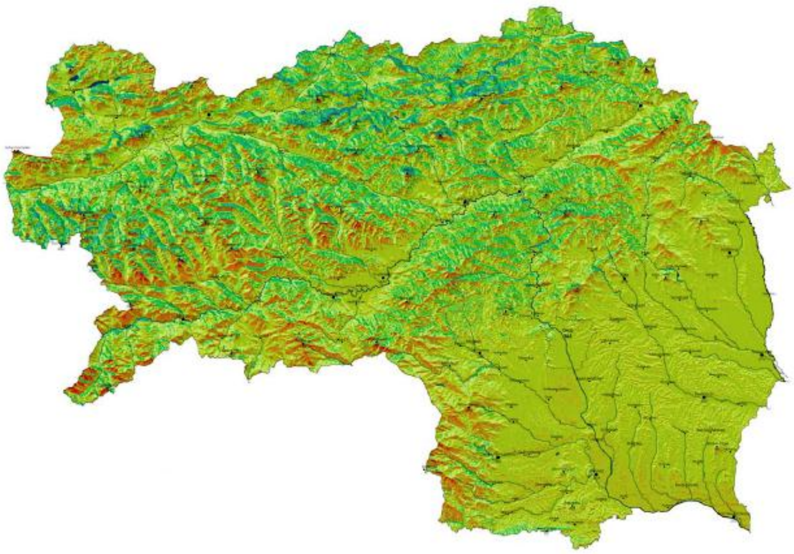


Figure 3.3: Yearly average on the global horizontal irradiance in Styria.[31]

The process of the determination of regional clusters consists of extensive data analyses (see later, correlation analysis, determination, optimisation and evaluation of prediction models etc.). In practice, a subset on real measured meteorological values (GI) and electricity generation values for a limited period of time is taken into consideration. In optimal case, one to three years; in the case of this master thesis

[31] Podesser (2010), p. 44.

the data for the defined one year observation period from 1st May 2013 to 30th April 2014 are available, evaluated and visualised following.

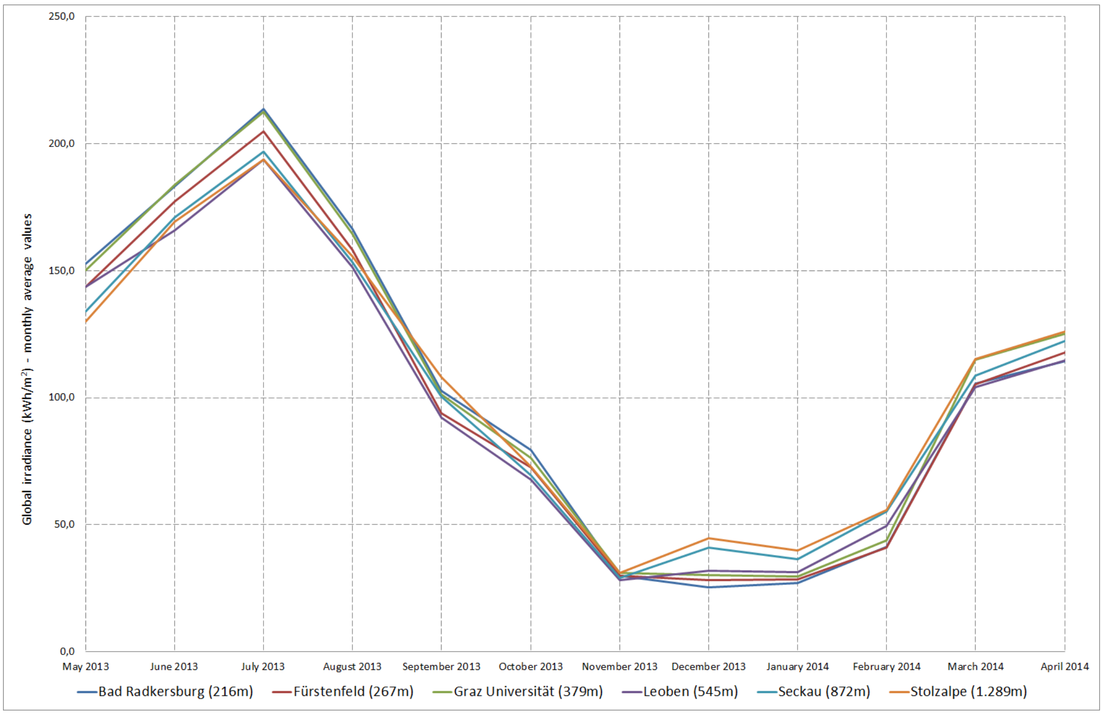


Figure 3.4: Graphical representation of real measured values of global horizontal irradiance in selected WMSs in Styria, on monthly basis from 1st May 2013 to 30th April 2014. (Own figure)

Year	Month	Bad Radkersburg (216m)	Fürstenfeld (267m)	Graz Universität (379m)	Leoben (545m)	Seckau (872m)	Stolzalpe (1.289m)
2013	May	152,8	143,6	150,1	143,5	134,0	130,0
	June	183,2	177,1	183,8	165,9	171,0	169,2
	July	213,6	204,8	212,6	193,8	196,8	193,8
	August	166,5	158,1	164,4	151,2	153,2	155,6
	September	102,8	94,0	101,2	92,2	100,5	108,1
	October	79,5	72,5	76,2	67,6	69,5	72,7
	November	29,8	29,8	31,1	28,4	29,1	31,2
	December	25,4	28,2	30,3	32,1	41,1	44,7
2014	January	27,2	28,6	29,6	31,3	36,5	39,9
	February	41,5	41,0	44,0	49,5	55,2	55,8
	March	105,5	105,2	115,0	104,0	108,7	115,1
	April	114,5	117,7	125,2	114,7	122,4	126,0

Table 3.1: Real measured values of global irradiance (kWh/m²) in selected WMSs in Styria, on monthly basis from 1st May 2013 to 30th April 2014. (Own analysis, own table)

Following analysis describes the differences in GI intensity in connection with regional micro-climatic variances if comparing the flat and mountainous or alpine areas of Styria, with the focus on the defined boarders for the 3D regional clusters – by the height above sea level of 500m and of 1.000m. The statements are based on the results of the long-term observation performed by ZAMG, which is presented in Podesser (2010), as well as on the evaluation of real measured values within the 1 year period defined for the purpose of this master thesis as described above.

In context with increasing sea level, the higher values of global irradiance (GI) are typical for mountainous and alpine areas during the winter month, especially from December to February. In March and April, the frequency of mist and fog in flat areas regresses and the GI values increase rapidly. In alpine areas, although the cloud cover increases slowly, it has not a huge influence on irradiation losses in case of sufficient snow cover. On the contrary, beginning with May, the high cloud cover in alpine areas corresponding to the sun peak is the main important factor for the GI values reduction. The highest position of the sun in summer month (up to 66° at midsummer) ensures further increasing amount of irradiation. Comparing the GI values in defined zones, the alpine areas remain a little bit behind. The GI maxima are reached in July. Increasing frequency of mist and fog as well as more often high cloud cover, which are typical especially for flat areas in autumn, caused pretty high irradiance losses in this period.

Similar to the regional intensity of GI values, the values of the average ambient temperature over the year in Styria reflect the determination of 3D clusters. If the graphic below is compared with the clustering model (figure 3.8), the determined micro-climatic boarders at the seal level 500m and 1.000m are clearly evident.

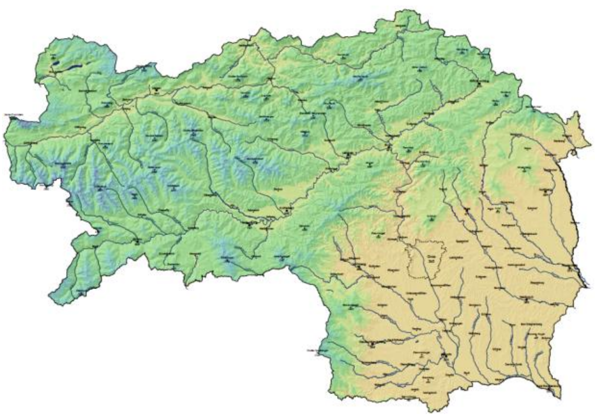


Figure 3.5: Yearly average on the ambient temperature in Styria.[32]

[32] Wakonigg (2010), p. 22.

Following graphics illustrate the described micro-climatic effects in Styria – like cloud cover and fog. The intensity of cloud cover (increasing with the dark blue colour) and the high frequency of fog are more obvious in the flat area of Styria (later on the cluster 1).

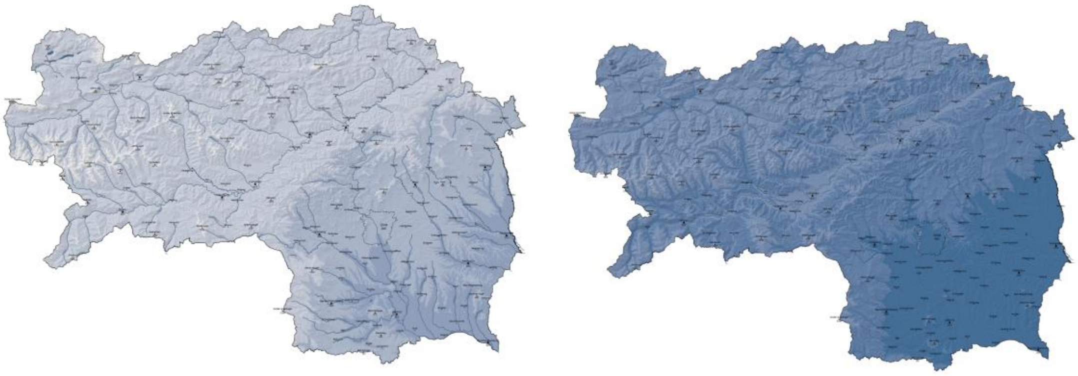


Figure 3.6: Grade of cloud cover in Styria from early-spring to early-summer (left) and from mid-summer to late-summer (right).[33]

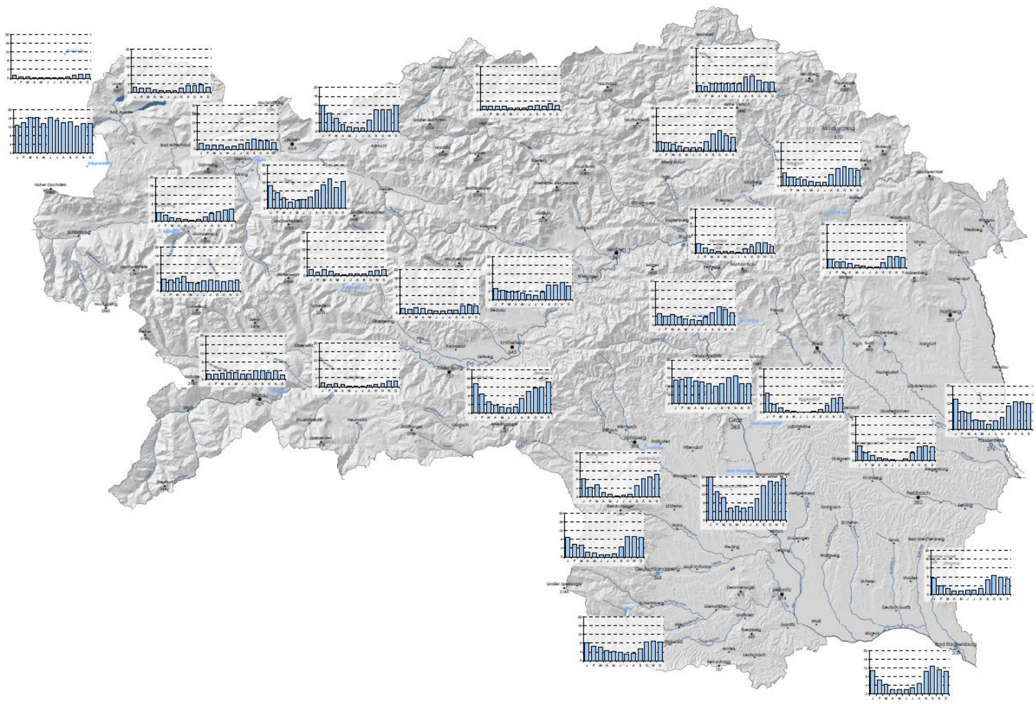


Figure 3.7: Frequency of fog in Styria in its yearly cycle.[34]

[33] Podesser, Wölfelmaier (2010), p. 67, p. 69.
[34] Podesser, Wölfelmaier (2010), p. 89.

3.4 Proposal of the best regional 3D clustering model

Since the observation period of this study is defined from 1st May 2013 to 30th April 2014, all (pre-)analyses and conclusions and afterwards the modelling process of the PV electricity generation forecast are performed for this timeframe.

Both the size of the geographical area and the number of considered PV installations contribute statistically to the reduction of a forecast inaccuracy and present one of the main challenges by composing appropriate clusters (see also up-scaling in chapter 5.2). There are 95 PV systems, which match the requirements given in this master thesis and are considered as a basis for determination of the regional clusters (see 3.1 and 3.2 for details on the chosen sample).

In cooperation with the Austrian Institute for Meteorology "*Zentralanstalt für Meteorologie und Geodynamik*" (ZAMG), the best choice of 16 representative weather measuring stations (WMSs) was made in first step.[35] This selection respects results of analysis of geographical allocation of the considered PV systems, particularly their position above sea level, further their distribution and density in geographical regions and the micro-climatic specifics, which are presented above. In the second step, the real measured values of global irradiance in defined WMSs (mean values in hourly raster, provided by ZAMG) and the measured values of electricity production of all considered PV systems (mean values in hourly raster, provided by OeMAG via APG) were taken into account and evaluated by means of correlation analysis (for results overview see the table below).[36]

It means, the values of electricity production of each PV system and the values of real measured GI of each proposed representative WMSs were correlated together. In the figure below, on the left the PV systems and at the top the WMSs are

[35] The representative WMSs are following: Bad Radkersburg (216m), Fürstenfeld (267m), Hartberg (337m), Deutschlandsberg (355m), Graz Universität (379m), Kapfenberg (506m), Leoben (545m), Hall/Admont (637m), Aigen/Ennstal (652m), Murau (814m), Oberwölz (834m), Seckau (872m), Fischbach (1.037m), Präbichl (1.214m), Stolzalpe (1.289m), Katschberg (1.637m).

[36] The data sets were checked on their plausibility and improved if necessary (e.g. measurement errors due to the transmission defects). Mostly, there are two approaches by the correlation analysis performance: either to take only day-hours into account (the best way is to pre-define a fixed timeframe in such case, e.g. from 5 a.m. to 9 p.m.; nevertheless there is still some variability over the year) or to take into account all 24-hours (day and night hours), which gives a bit higher correlation coefficients (due to the high correlation at night-hours; both the GI and the electricity generation are zero at that time), but the absolute correlation rate remains the same. In case of this study all 24-hours are taken into account for the correlation analysis.

arranged in ascending order according to their position above the sea level. The assumption is that PV systems and WMSs situated at a similar sea level correlate in the best way. The highest correlation coefficients, means the best "overlapping", are highlight by the red colour. The darker the red colour the higher the correlation of the electricity generation of each considered PV system with the measuring station in a specific region.[37] Based on the correlation analysis there is clear evidence that the height at approx. 500m above sea level builds a very significant boarder for the regional 3D clustering model. Also, the height at approx. 1.000m seems to be relevant for the determination of regional clusters (see later on).

ZAMG-station(SL)/PV-system			Bad Radkersburg (216m)	Fürstenfeld (267m)	Hartberg (337m)	Deutschlandsberg (355m)	Graz Universität (379m)	Kapfenberg (506m)	Leoben (545m)	Hall/Admont (637m)	Aigen/Ennstal (652m)	Murau (814m)	Oberwölz (834m)	Seckau (872m)	Fischbach (1037m)	Präbichl (1214m)	Stolzalpe (1289m)	Katschberg (1637m)
PV_01	203	lower 500m	97,5%	92,6%	91,9%	91,3%	92,3%	89,4%	89,6%	86,8%	86,7%	87,6%	86,8%	89,4%	87,9%	85,7%	86,9%	86,2%
PV_02	227		97,3%	94,1%	93,3%	93,3%	93,9%	91,1%	91,1%	87,9%	87,8%	88,8%	88,3%	90,6%	89,6%	87,5%	87,7%	87,4%
PV_03	235		96,5%	93,7%	92,4%	92,3%	93,1%	90,5%	90,2%	87,6%	87,6%	87,4%	86,7%	89,3%	88,9%	86,2%	86,8%	86,4%
PV_04	235		96,9%	94,3%	93,2%	93,9%	94,0%	91,3%	91,5%	88,5%	88,5%	89,1%	88,6%	91,1%	89,9%	87,7%	87,7%	87,5%
PV_05	235		95,7%	92,8%	91,6%	91,8%	92,6%	89,5%	89,9%	86,8%	86,7%	87,8%	87,3%	89,5%	88,2%	85,9%	86,9%	86,3%
PV_06	242		97,6%	94,7%	93,5%	93,4%	94,2%	91,7%	91,5%	88,6%	88,6%	89,1%	88,5%	90,7%	89,9%	87,9%	88,3%	88,2%
PV_07	242		97,9%	95,0%	93,7%	93,9%	94,6%	92,0%	91,6%	88,9%	89,0%	89,2%	89,0%	91,1%	90,0%	88,6%	88,4%	88,3%
PV_08	250		94,1%	93,7%	93,7%	91,6%	92,8%	90,1%	89,5%	86,6%	86,9%	87,5%	87,1%	89,8%	89,1%	86,7%	86,5%	87,0%
PV_09	257		96,1%	94,2%	92,7%	92,6%	93,8%	90,8%	90,7%	87,7%	87,5%	88,4%	88,1%	90,2%	89,6%	87,1%	87,9%	87,4%
PV_10	260		97,0%	95,3%	93,9%	93,5%	94,7%	91,6%	91,3%	88,5%	88,6%	89,0%	88,7%	90,7%	90,4%	87,8%	88,4%	88,5%
PV_11	262		95,7%	94,6%	93,0%	94,8%	94,6%	91,2%	91,4%	87,7%	87,9%	89,4%	89,2%	91,1%	90,2%	87,8%	88,5%	88,2%
PV_12	264		93,5%	96,2%	93,1%	93,0%	97,1%	91,2%	91,1%	87,2%	87,5%	88,9%	89,1%	90,8%	90,1%	86,8%	88,6%	87,7%
PV_13	270		95,6%	93,9%	93,2%	91,6%	93,0%	90,1%	89,7%	87,3%	87,1%	87,5%	86,8%	89,1%	88,6%	86,3%	87,1%	86,8%
PV_14	271		95,5%	93,8%	93,2%	91,6%	92,9%	90,0%	89,6%	87,1%	86,9%	87,4%	86,7%	89,1%	88,6%	86,2%	86,8%	86,6%
PV_15	275		95,9%	95,5%	93,0%	94,9%	95,5%	92,2%	91,8%	88,6%	88,8%	89,6%	89,7%	91,2%	90,4%	88,1%	89,2%	89,0%
PV_16	278		95,1%	95,4%	93,5%	94,7%	95,2%	91,5%	91,9%	87,9%	88,2%	89,9%	89,9%	91,9%	90,8%	87,7%	88,9%	88,5%
PV_17	285		94,6%	94,2%	92,0%	93,1%	94,1%	90,1%	90,4%	86,6%	86,7%	88,6%	88,6%	90,0%	89,6%	86,4%	88,3%	87,5%
PV_18	287		95,7%	95,2%	93,1%	93,7%	94,9%	91,2%	91,2%	87,9%	87,8%	89,0%	88,7%	90,6%	90,2%	87,4%	88,8%	88,4%
PV_19	288		94,8%	92,8%	90,8%	90,1%	92,0%	89,3%	88,4%	86,5%	86,5%	85,9%	85,6%	87,0%	87,0%	85,0%	86,3%	86,3%
PV_20	290		95,6%	95,6%	94,3%	93,3%	94,9%	91,2%	91,2%	87,6%	87,8%	88,9%	88,8%	91,0%	90,5%	87,4%	88,2%	88,1%
PV_21	291		94,3%	93,8%	92,6%	93,8%	93,7%	89,9%	90,6%	85,9%	86,2%	88,6%	88,3%	90,8%	89,8%	86,7%	87,0%	86,9%
PV_22	298		94,9%	95,7%	94,5%	93,8%	95,0%	91,4%	91,3%	87,2%	87,4%	88,9%	88,9%	91,5%	90,6%	87,7%	87,6%	87,5%
PV_23	303		94,9%	94,6%	93,0%	93,9%	94,5%	90,7%	91,2%	87,1%	87,1%	89,0%	88,9%	91,0%	90,1%	87,3%	88,1%	87,8%
PV_24	304		95,9%	96,1%	94,2%	93,5%	95,2%	91,9%	91,6%	88,3%	88,4%	89,1%	89,0%	91,0%	90,2%	88,0%	88,7%	88,5%
PV_25	304		82,6%	82,8%	81,9%	80,9%	82,6%	77,7%	77,9%	75,2%	75,1%	77,3%	77,6%	79,0%	79,0%	76,6%	77,7%	77,3%
PV_26	310		93,0%	92,7%	91,9%	90,2%	92,0%	87,1%	87,2%	83,9%	84,2%	85,0%	84,6%	87,3%	88,2%	83,9%	84,4%	84,5%
PV_27	312		91,1%	89,8%	87,2%	90,5%	90,1%	87,4%	86,9%	85,8%	85,6%	85,2%	84,4%	85,4%	84,2%	83,0%	85,9%	84,7%
PV_28	313		93,3%	94,6%	92,5%	94,0%	95,3%	90,2%	90,9%	87,0%	87,0%	89,0%	88,8%	90,8%	89,4%	86,6%	88,3%	87,4%
PV_29	315		95,6%	96,3%	94,8%	94,3%	95,6%	92,2%	92,0%	88,0%	88,3%	89,5%	89,5%	92,0%	91,2%	88,5%	88,5%	88,4%
PV_30	315		95,4%	96,1%	94,5%	94,1%	95,4%	91,9%	91,8%	87,7%	87,9%	89,1%	89,1%	91,6%	90,8%	88,2%	87,9%	88,0%
PV_31	317		94,0%	95,8%	94,0%	92,5%	94,9%	90,8%	90,8%	87,2%	87,1%	88,5%	88,6%	90,6%	90,0%	87,2%	88,0%	87,5%
PV_32	323		96,2%	95,3%	93,5%	92,8%	94,6%	91,4%	90,9%	88,1%	88,2%	88,6%	88,3%	90,2%	90,0%	87,2%	88,4%	88,0%
PV_33	327		94,1%	93,9%	92,1%	96,2%	94,6%	90,9%	91,4%	87,6%	88,2%	89,5%	89,2%	91,1%	89,5%	87,2%	88,6%	88,0%
PV_34	332		84,7%	86,6%	87,1%	83,5%	85,6%	84,6%	82,9%	80,3%	80,7%	79,1%	79,1%	81,5%	80,6%	80,0%	78,0%	78,0%
PV_35	332		94,9%	95,7%	93,6%	94,1%	95,5%	91,2%	91,3%	87,7%	87,8%	89,2%	89,1%	91,0%	90,3%	87,3%	88,6%	88,3%
PV_36	333		96,2%	95,4%	93,4%	92,7%	94,6%	91,5%	90,8%	88,1%	88,3%	88,5%	88,2%	90,0%	89,8%	87,5%	88,4%	88,1%
PV_37	336		93,3%	94,1%	92,1%	96,5%	95,3%	91,3%	91,7%	88,0%	88,4%	90,0%	89,4%	91,3%	89,9%	87,7%	88,9%	87,9%
PV_38	345		93,3%	93,7%	91,9%	98,5%	94,4%	91,0%	91,3%	88,1%	88,1%	89,3%	88,7%	90,4%	89,2%	87,3%	88,5%	87,5%
PV_39	347		92,5%	94,9%	95,7%	91,7%	93,9%	91,3%	90,4%	87,0%	87,6%	87,8%	87,4%	89,8%	90,2%	86,9%	86,8%	86,6%
PV_40	349		93,5%	97,6%	94,2%	93,3%	95,7%	91,7%	91,0%	88,1%	88,5%	88,8%	88,8%	91,1%	90,1%	88,3%	88,4%	87,5%
PV_41	349		93,9%	98,3%	94,2%	92,6%	96,1%	92,0%	91,3%	88,2%	88,5%	88,5%	88,5%	90,6%	90,3%	87,3%	88,5%	87,9%
PV_42	350		93,5%	96,6%	96,6%	92,4%	95,1%	91,9%	91,3%	87,8%	87,9%	88,3%	88,3%	90,5%	91,0%	87,4%	87,8%	87,0%
PV_43	353		70,0%	73,6%	69,2%	67,7%	71,0%	68,7%	67,7%	65,4%	64,7%	63,4%	62,9%	65,7%	66,1%	63,7%	63,0%	63,2%
PV_44	353		88,8%	93,4%	89,5%	88,3%	91,1%	87,1%	87,2%	83,5%	83,8%	84,7%	84,7%	86,7%	85,2%	82,6%	83,7%	82,6%
PV_45	360		92,5%	95,5%	93,0%	93,3%	97,8%	91,9%	92,0%	86,8%	87,8%	89,6%	89,9%	91,7%	91,4%	87,2%	88,9%	88,7%
PV_46	362		92,7%	95,2%	95,8%	91,6%	94,3%	91,3%	90,4%	87,2%	87,2%	87,6%	87,5%	89,7%	89,9%	86,8%	87,1%	86,5%
PV_47	367		92,0%	97,7%	93,5%	91,7%	95,1%	90,7%	90,4%	86,5%	86,7%	87,9%	87,7%	90,2%	89,8%	86,2%	87,4%	86,7%
PV_48	370		91,9%	97,0%	94,2%	93,0%	95,6%	91,0%	91,1%	85,9%	86,6%	88,3%	88,7%	91,3%	91,1%	86,7%	87,0%	86,5%
PV_49	370		92,3%	97,3%	94,5%	93,3%	95,9%	91,4%	91,5%	86,5%	87,0%	88,7%	89,1%	91,6%	91,3%	87,1%	87,5%	86,8%
PV_50	374		93,3%	95,4%	92,9%	94,2%	96,9%	91,3%	91,9%	88,0%	88,1%	89,9%	89,6%	91,3%	90,3%	87,6%	89,2%	88,3%
PV_51	374		93,3%	95,4%	93,0%	94,3%	96,9%	91,3%	91,9%	88,0%	88,1%	89,8%	89,5%	91,3%	90,4%	87,6%	89,0%	88,2%
PV_52	374		88,3%	91,7%	92,3%	89,7%	91,1%	87,3%	87,8%	82,4%	83,0%	85,0%	84,8%	88,2%	88,2%	83,8%	82,6%	82,3%
PV_53	377		91,7%	93,5%	94,5%	89,0%	92,3%	89,6%	88,6%	85,9%	86,0%	85,7%	85,5%	87,4%	88,0%	84,6%	86,1%	84,9%
PV_54	381		90,9%	95,0%	93,5%	90,0%	93,6%	89,2%	89,2%	85,4%	85,5%	86,7%	86,6%	88,9%	88,9%	85,0%	86,6%	85,2%
PV_55	382		90,6%	93,7%	90,7%	90,6%	96,3%	89,4%	89,6%	85,5%	85,4%	87,3%	87,9%	89,3%	87,9%	85,3%	87,7%	86,2%
PV_56	386		93,5%	94,2%	91,1%	91,8%	94,0%	90,1%	89,5%	86,8%	86,9%	87,1%	87,0%	88,4%	88,1%	85,3%	87,5%	87,0%
PV_57	392		92,9%	94,3%	91,6%	94,0%	95,4%	92,4%	92,7%	89,0%	89,5%	90,8%	90,7%	91,8%	90,0%	88,7%	90,8%	89,6%
PV_58	401		90,8%	91,1%	89,8%	94,7%	92,0%	87,8%	89,0%	85,8%	85,3%	88,2%	87,7%	89,4%	87,5%	85,3%	87,5%	86,0%
PV_59	405		89,8%	92,6%	95,9%	89,7%	91,6%	88,8%	88,6%	84,0%	84,1%	86,0%	86,2%	88,8%	89,9%	84,9%	84,6%	83,7%
PV_60	406		94,1%	96,9%	94,4%	93,2%	95,7%	91,3%	90,9%	87,5%	87,7%	88,5%	88,6%	90,8%	90,7%	87,3%	87,9%	87,5%
PV_61	426		88,7%	91,8%	93,0%	89,5%	91,1%	87,0%	86,7%	81,9%	82,7%	84,5%	84,2%	87,1%	88,6%	82,5%	83,7%	83,0%
PV_62	432		91,2%	95,2%	94,1%	92,2%	94,9%	91,3%	91,3%	86,1%	86,7%	88,2%	88,3%	90,8%	91,6%	86,7%	87,1%	86,2%
PV_63	448		65,3%	70,3%	71,2%	69,8%	70,8%	64,5%	67,5%	60,6%	61,1%	66,6%	66,5%	69,8%	70,2%	64,8%	63,9%	62,9%
PV_64	475		88,9%	90,9%	90,0%	92,4%	92,1%	89,3%	90,7%	85,7%	86,0%	89,6%	89,5%	91,5%	89,5%	87,3%	87,9%	86,9%
PV_65	484		91,5%	95,2%	93,6%	91,8%	95,4%	91,8%	91,5%	86,6%	86,9%	88,7%	88,7%	90,9%	92,2%	87,1%	88,3%	86,9%
PV_66	484		91,5%	95,2%	93,5%	91,7%	95,3%	92,1%	91,4%	87,1%	87,3%	88,5%	88,5%	90,7%	91,7%	87,4%	88,4%	86,8%
PV_67	512	from 500m to 1.000m	91,6%	93,0%	91,9%	90,8%	92,6%	97,9%	94,4%	90,6%	90,7%	89,8%	90,1%	92,3%	91,5%	90,4%	89,4%	88,2%
PV_68	526		88,3%	88,8%	88,3%	88,9%	88,3%	90,1%	90,8%	94,4%	93,1%	88,3%	87,3%	89,6%	85,8%	90,1%	86,4%	85,2%
PV_69	560		86,9%	89,1%	88,3%	88,2%	89,2%	92,2%	91,8%	86,4%	86,6%	88,5%	89,1%	90,8%	91,4%	89,1%	88,1%	85,9%
PV_70	560		86,4%	88,4%	87,7%	87,7%	88,7%	91,6%	91,4%	86,0%	86,2%	88,3%	88,9%	90,5%	90,9%	88,6%	87,9%	85,6%
PV_71	612		91,5%	92,0%	90,9%	91,6%	92,0%	93,5%	95,0%	90,6%	91,2%	91,4%	91,1%	94,7%	89,8%	90,0%	90,0%	89,2%
PV_72	649		90,3%	93,2%	91,4%	90,3%	93,3%	92,2%	91,4%	86,8%	86,8%	88,4%	88,6%	90,2%	91,9%	86,9%	88,0%	87,1%
PV_73	683		91,9%	92,3%	91,1%	91,4%	92,6%	92,3%	93,1%	89,7%	90,0%	92,1%	92,3%	94,0%	89,9%	89,2%	91,0%	89,9%
PV_74	683		90,8%	91,4%	90,2%	90,7%	91,7%	91,9%	93,1%	89,7%	90,0%	92,1%	92,0%	94,0%	89,4%	89,1%	91,2%	90,3%
PV_75	692		90,3%	91,5%	90,1%	90,8%	91,8%	92,0%	93,2%	89,5%	90,0%	92,5%	92,7%	95,3%	89,4%	89,3%	91,6%	90,2%
PV_76	698		85,8%	86,5%	85,8%	86,1%	86,2%	88,8%	90,1%	93,0%	95,4%	89,3%	88,1%	89,7%	84,2%	88,7%	87,8%	86,2%
PV_77	755		65,7%	66,7%	65,9%	65,4%	66,5%	68,3%	70,6%	74,0%	74,4%	71,6%	70,2%	70,7%	67,9%	69,9%	72,1%	69,6%
PV_78	802		81,6%	82,9%	81,9%	80,9%	82,0%	84,4%	84,2%	85,5%	84,1%	81,9%	81,0%	82,7%	82,2%	84,6%	82,1%	78,9%
PV_79	814		87,4%	88,8%	87,7%	88,0%	89,2%	92,1%	94,1%	88,5%	88,0%	90,0%	90,4%	92,3%	88,8%	90,8%	89,6%	87,7%
PV_80	832		84,0%	84,9%	84,1%	85,5%	85,7%	85,9%	88,2%	85,2%	85,8%	95,0%	91,8%	89,3%	85,3%	85,4%	93,3%	89,0%
PV_81	845		84,8%	87,4%	84,7%	83,7%	88,3%	85,8%	85,5%	80,9%	81,1%	83,1%	84,2%	84,7%	86,2%	81,0%	84,4%	82,9%
PV_82	858		83,5%	85,6%	84,8%	85,5%	86,2%	87,5%	89,5%	84,8%	85,0%	89,2%	90,2%	93,4%	86,6%	86,7%	88,6%	86,2%
PV_83	860		87,8%	89,9%	88,9%	90,6%	91,2%	88,4%	89,9%	85,1%	85,4%	89,2%	89,3%	91,5%	88,8%	86,6%	88,4%	87,2%
PV_84	861		83,6%	83,9%	82,9%	81,5%	83,1%	85,4%	84,4%	86,1%	85,0%	81,4%	80,5%	82,4%	82,3%	83,6%	81,9%	79,5%
PV_85	864		86,4%	87,8%	87,2%	85,6%	87,4%	89,9%	89,4%	87,1%	86,7%	86,8%	86,6%	87,7%	88,8%	88,4%	86,4%	84,0%
PV_86	867		89,0%	90,3%	88,9%	90,1%	91,2%	90,7%	91,8%	87,5%	88,0%	91,4%	92,0%	93,2%	88,7%	87,5%	91,1%	89,8%
PV_87	873		75,5%	75,8%	74,5%	73,0%	75,0%	77,6%	76,6%	78,8%	77,6%	74,4%	73,2%	74,5%	75,2%	75,8%	75,7%	73,2%
PV_88	890		87,0%	88,4%	87,7%	89,1%	89,4%	88,7%	90,4%	85,1%	85,7%	89,9%	90,2%	91,9%	87,8%	86,0%	88,4%	87,5%
PV_89	892		83,4%	84,7%	82,7%	82,9%	85,2%	86,5%	86,9%	84,4%	84,6%	86,0%	87,3%	89,3%	83,7%	83,2%	87,1%	84,7%
PV_90	942		86,3%	88,5%	87,6%	89,3%	89,9%	87,0%	88,8%	83,9%	84,0%	88,3%	88,5%	90,7%	87,8%	85,6%	87,6%	86,2%
PV_91	972		74,6%	76,3%	75,3%	74,7%	75,8%	77,3%	77,3%	77,9%	76,7%	74,9%	74,4%	76,2%	76,0%	78,5%	74,6%	71,5%
PV_92	1067	higher 1.000m	83,8%	85,8%	84,7%	86,9%	87,5%	84,1%	85,9%	80,4%	81,0%	85,7%	85,8%	87,3%	85,5%	82,6%	85,4%	83,9%
PV_93	1207		87,2%	88,2%	86,7%	87,7%	88,8%	88,3%	89,5%	85,5%	86,6%	92,7%	91,9%	90,3%	87,4%	86,1%	92,6%	89,8%
PV_94	1207		76,1%	77,5%	76,2%	77,2%	77,8%	77,7%	78,9%	75,3%	76,2%	81,9%	81,2%	79,9%	76,7%	76,2%	81,6%	79,0%
PV_95	1900		76,9%	79,1%	78,4%	78,8%	79,9%	79,8%	81,7%	78,5%	78,7%	84,4%	86,5%	83,9%	81,9%	81,5%	84,8%	82,9%
cluster 1			92,7%	93,6%	92,1%	91,7%	93,3%											
sub-cluster 2A								87,9%	88,5%	86,1%	86,1%	87,0%	86,8%	88,5%				
sub-cluster 2B															82,9%	81,6%	86,1%	83,9%
cluster 2								87,2%	87,9%	85,2%	85,3%	86,8%	86,7%	88,0%	85,6%	85,2%	86,4%	84,5%
ALL			90,3%	91,4%	90,0%	89,7%	91,2%	88,8%	89,0%	85,7%	85,9%	87,1%	87,0%	88,8%	87,7%	85,6%	86,5%	85,6%

Table 3.2: Evaluation of the chosen sample: Correlation analysis (real measured values) between electricity generation of considered PV systems and global irradiance by representative weather measuring stations (ZAMG). (Calculation made by STATISTICA, own table)

[37] The grey fields marked particular situations, where the highest calculated correlation is not directly in the due to the sea level defined region. In such cases, the deviation to the highest correlation in the defined region is mostly minimal.

Respecting the results of the correlation analysis and also the micro-climatic specifics of the region and the density / distribution of the considered PV system installations, the proposal of the best regional 3D clustering model for Styria is based on two main clusters, which are defined following (see the map below):

- Cluster 1: PV systems situated on the sea level below 500m.
- Cluster 2: PV systems situated on the sea level above 500m.

  Later on, the cluster 2 is splitted into two sub-clusters for deeper analysis:

  - Sub-cluster 2A: PV systems situated on the sea level from 500m to 1.000m.
  - Sub-cluster 2B: PV systems situated on the sea level above 1.000m.

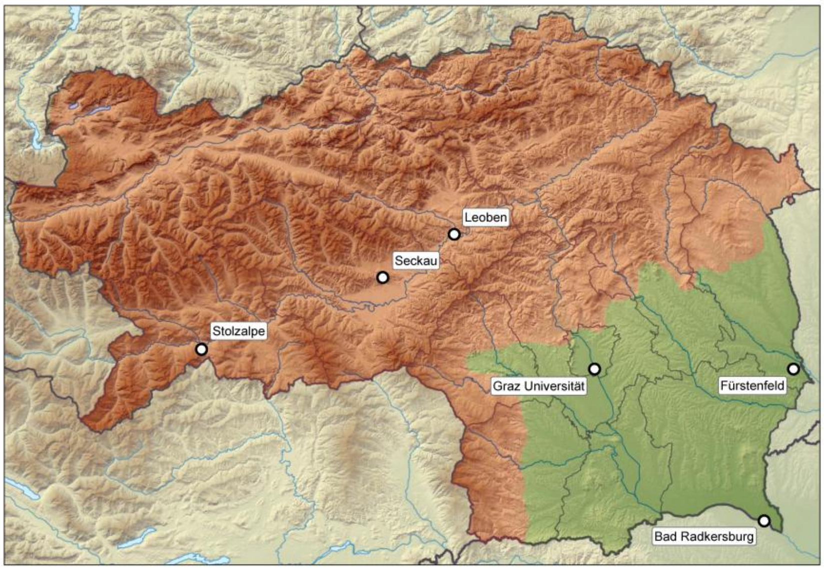


Figure 3.8: Final proposal of the regional 3D clustering model of Styria and defined representing WMSs. (Own figure)

The cluster 1 is represented by the southern part of Styria; valleys, flat areas and the sea level below 500m are typical for this cluster. The density of PV installations in this cluster is much higher than in the cluster 2; nearly 70% of the considered PV systems are located in this area. Due to its characteristics – like large amount of PV installations, relatively small area and difference in height above sea level in range of only 300m (from 200m to 500m) – the cluster 1 is determined with a homogenous sample. Regarding to that, the basic idea was to cover this area with appropriate number of WMSs – finally, with 3 representing WMSs – to be able to monitor the local deviations and to minimise its influence on the forecast.

From the climatological point of view, mist, fog and inversion are the typical weather conditions from (late) autumn to spring, usually from mid-October or early-November to April. Especially, in October and November these phenomena are mostly pretty stable and long-lasting. Further, the local cloud cover occurs often in higher intensity in warm, summer months in general. Both of these effects vary considerable within one season as well as from one year to the next. This is the main challenge for this region in context of PV electricity generation forecasting process. The forecast mismatches go in both directions and affect the GI forecast and PV electricity generation forecast directly. Basically, the high rate of predicted fog and/or cloud cover, which unexpected doesn't occur (which can be also linked with wind force again), means higher values of GI and suddenly much more PV electricity generation than expected (and vice-versa).

The cluster 2 is characterised by (low) mountainous and alpine areas and the sea level above 500m. This includes large and very variable area with difference in the height above sea level in range of 1.400m (from 500m to 1.900m). The amount of PV installations in this cluster is noticeably lower in comparison to the cluster 1. In context with this large distribution and variability, the most challenging issue, in the first step, is to find or to determine the representing weather measuring stations respecting the similar height above sea level, especially in the alpine region. The volatile frequency and intensity of cloud cover in late summer and autumn as well as the snow fall and snow covered PV-modules in winter are the most challenging issues for the forecasting process.

According to the correlation analysis described above, the best representative weather measuring stations for cluster 1 are Bad Radkersburg, Fürstenfeld and Graz Universität, for sub-cluster 2A Leoben and Seckau and finally for sub-cluster 2B Stolzalpe. Following table presents the correlation values (in %):

WMS	**WMS's SL (m)**	**Correlation (%) with PVs in (sub-)cluster** (incl. night hours)				
		1	2	2A	2B	all
Bad Radkersburg	216	92,7	.	.	.	90,3
Fürstenfeld	267	93,6	.	.	.	91,4
Graz Universität	379	93,3	.	.	.	91,2
Leoben	545	.	87,9	88,5	.	89,0
Seckau	872	.	88,0	88,5	.	88,8
Stolzalpe	1.289	.	86,4	.	86,1	86,5

Table 3.3: Evaluation of the chosen sample: Results of the correlation analysis – the best representative weather measuring stations. (Own results, own table)

4 FORECASTING PROCESS AND EVALUATION

4.1 Prediction models

The prediction models are based on the statistical MARS-model (multivariate adaptive regression splines), which takes into account one dependent variable (in this case electricity production) and several influencing factors (in this case global irradiance and ambient temperature) plus historical generation data.

The forecasting models are created and run with the software tool mP Energy (Meta Predictor Energie, Hakom), which provides prediction models based on multivariate regression and enables to represent the non-linear influence factors like in case of various meteorological data. Basically, the modeling process with mP Energy consists of a sequence of 4 steps:

- Definition of the prediction model (input and output factors), poss. sub-model(s)
- Optimisation of the model
- Calibration and validation of the model (“training” of a model)
- Forecasting process with the model

Following values are defined as the input factors: global irradiance forecast values (GL) per weather measuring station (WMS) and calculated regional average (all WMSs in defined region), ambient temperature forecast values (TL) per WMS and calculated regional average (all WMSs in defined region). The output factor is electricity generation.

To be able to evaluate the influencing factors, their interactions, weight and poss. patterns in the model very precisely, various prediction models, sub-models and their combinations were defined. Using the mP Energy Optimiser tool, each separate model was optimised (i.e. the best model settings were defined using the cross-validation technique and the minimal RMSE) and trained, before the forecasting process was started.

4.1.1 Cluster 1

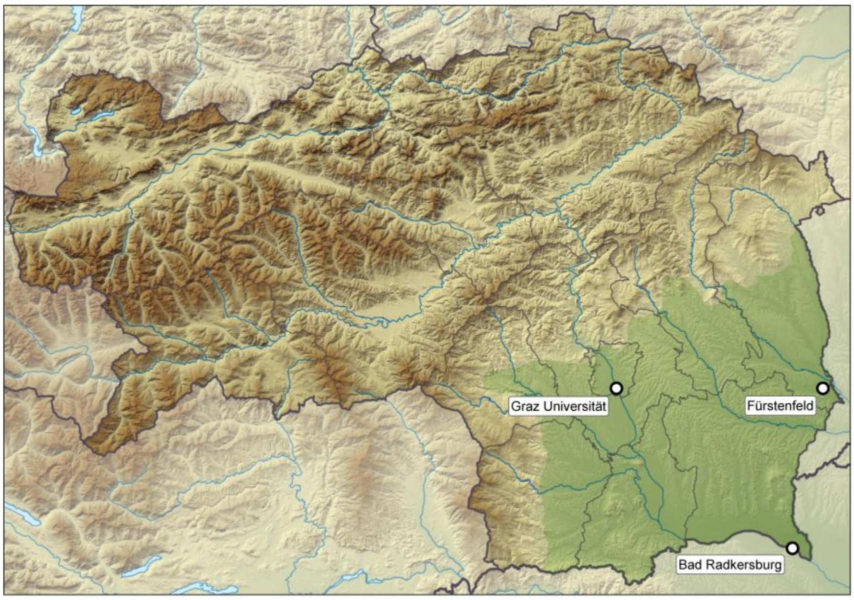


Figure 4.1: Visualisation of the cluster 1 and defined representing WMSs. (Own figure)

As described above, there are 3 weather measuring stations (Bad Radkersburg, Fürstenfeld, Graz Universität) defined as relevant for the forecasting process for cluster 1. Global irradiance forecast values (GL) and ambient temperature forecast values (TL) of these stations are taken as the input factors. The electricity generation of PV installations in this region is taken as the output factor. The basic question for the forecasting procedure in this region is, whether the calculated average values for GL and TL should be taken into consideration in addition. To evaluate this issue, there are 3 models defined and analysed for cluster 1:

1. The model "Cl1_WMSs" includes solely GL and TL for the weather measuring stations (but no average values).
2. The model "Cl1_average" includes only GL and TL calculated average values (but no GL and TL for each particular WMS).
3. The model "Cl1_all" includes GL and TL for all weather measuring stations (WMSs) as well as the GL and TL calculated average values.

4.1.2 Cluster 2

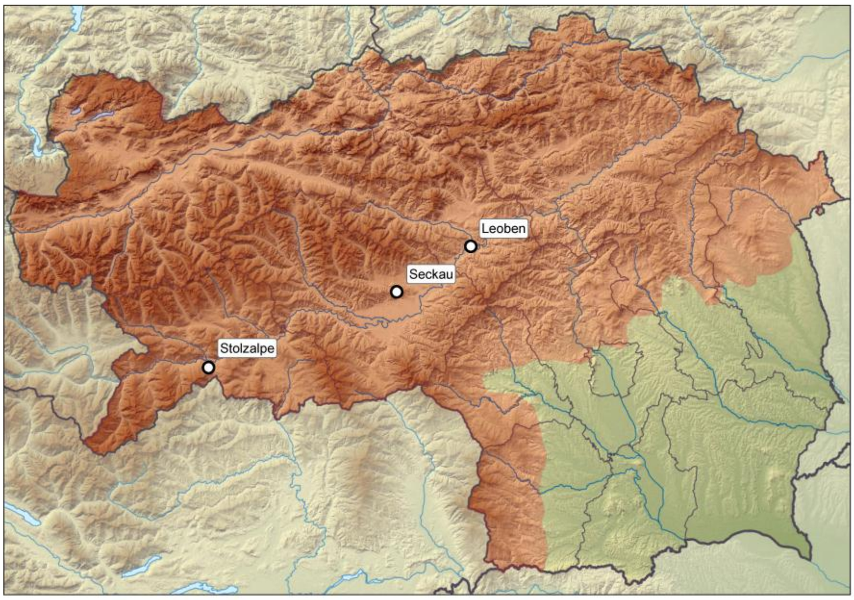


Figure 4.2: Visualisation of the cluster 2 (poss. 2A, 2B) and defined representing WMSs. (Own figure)

As described above, there are several reasons, why the cluster 2 is more complicated in relation to the forecasting process in comparison to the cluster 1. Basically, the cluster 2 is divided into two sub-clusters: sub-cluster 2A for PV installations situated at the sea level from 500m to 1.000m represented by two WMSs (Leoben, Seckau), and sub-cluster 2B for PV installations above 1.000m sea level represented by one WMS (Stolzalpe). Global irradiance forecast values (GL) and ambient temperature forecast values (TL) of these stations (and if reasonable the average values) are taken as the input factors. The electricity generation of PV installations in this region is taken as the output factor. The core question for the forecasting process in cluster 2 is, which approach will give better accuracy:

- to split the cluster 2 into two independent areas (sub-clusters 2A and 2B), make a forecast for these two models before getting the common forecast for cluster 2, or
- especially regarding to the very small number of installations in sub-cluster 2B, to merge the sub-clusters 2A and 2B and make only one common forecast for cluster 2 (taking into account only the common criteria – installations above 500m), and if so, which WMSs should be taken into account.

To evaluate this issue, there are 3 main models defined and analysed for cluster 2:

1. The model "Cl2A & Cl2B" represents two separate models for sub-cluster 2A (WMSs Leoben, Seckau) and 2B (WMS Stolzalpe). In addition, the sub-models "Cl2A_WMSs", "Cl2A_avarage" and "Cl2A_all" are defined in the same way as in cluster 1 (see above).
2. The model "Cl2_WMSsCl2A" represents one common model for cluster 2 (includes both sub-clusters 2A and 2B) and includes only the WMSs of sub-cluster 2A (in order to that, the relevance of WMS of sub-cluster 2B in the common model is verified). Also, the options "_WMSs", "_average" and "_all" are taken into consideration.
3. The model "Cl2_WMSsCl2ACl2B" represents one common model for cluster 2 (includes both sub-clusters 2A and 2B) and includes all WMSs of sub-clusters 2A and 2B. Also in this case, the options "_WMSs", "_average" and "_all" are considered.

4.2 Interpretation of results

The accuracy of a PV electricity production forecast depends on numerous factors, such as meteorological and climatological conditions (number of input factors), forecast horizons, size of the geographical area and the number of installations considered. The criteria of a normalised Root Mean Square Error (nRMSE) and a normalised Mean Absolute Error (nMAE) both referring to installed capacity (kWp) are used for verification of the accuracy of the prediction models. These criteria are frequently used for comparing the forecast accuracy of different models (the lower the nRMSE or nMAE the better the forecast results) while taking into consideration estimated values ($P_{estimated}$, predicted by a model) and real measured values ($P_{measured}$) for the electricity generation output.

- $$nRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{(P_{estimated,i}-P_{measured,i})^2}{P_{installed,i}{}^2}}$$

- $$nMAE = \frac{1}{N}\sum_{i=1}^{N}\frac{|P_{estimated,i}-P_{measured,i}|}{P_{installed,i}}$$

The nMAE gives an average value of forecast errors, while nRMSE gives more weight on the largest error(s) (see Pelland (2013)). The values of nRMSE and nMAE are presented in the form of standardised data; both as an outcome without consideration of night hours. To simplify this approach, the period from 9 p.m. to 5 a.m. is defined as the night hours constantly over the whole year. Naturally, the model without consideration of night hours – comparing to the model, which takes also night hours into account – gives by a few percent worse, nevertheless more realistic and meaningful values.

4.2.1 Cluster 1

In case of cluster 1, the model "Cl1_average" provides the best results, i.e. the lowest nRMSE (7,6%) and nMAE (4,91%) are achieved if solely GL and TL average values of all considered WMSs are taken into consideration.

Cluster 1	nRMSE (%)	nMAE (%)
Cl1_WMSs	7,87	5,47
Cl1_average	**7,60**	**4,91**
Cl1_all	8,08	5,38

Table 4.1: nRMSE and nMAE evaluation of the cluster 1. (Own results, own table)

4.2.2 Cluster 2

In case of cluster 2, the model "Cl2_WMSsCl2ACl2B_WMSs" provides the best results, i.e. the lowest nRMSE (9,19%) and nMAE (5,85%) are achieved if the forecast is done for a common model for cluster 2 (without differing between sub-clusters 2A and 2B) and GL and TL (not the average values) of all WMSs of sub-clusters 2A and 2B are taken into consideration.

Cluster 2	nRMSE (%)	nMAE (%)
Cl2A_WMSs	9,30	6,12
Cl2A_average	9,53	6,42
Cl2A_all	9,48	6,32
Cl2B	15,38	10,10
Cl2A & Cl2B	9,87	6,35
Cl2_WMSsCl2A_WMSs	10,38	6,44
Cl2_WMSsCl2A_average	11,03	6,58
Cl2_WMSsCl2A_all	10,40	6,81
Cl2_WMSsCl2ACl2B_WMSs	**9,19**	**5,85**
Cl2_WMSsCl2ACl2B_average	9,62	6,00
Cl2_WMSsCl2ACl2B_all	9,38	5,98

Table 4.2: nRMSE and nMAE evaluation of the cluster 2. (Own results, own table)

4.2.3 Comparison of results for cluster 1 and cluster 2

As mentioned, there are lots of factors, which influence the accuracy of a PV electricity production forecast. Comparing the results of cluster 1 and cluster 2 following facts are worth to be mentioned:

- Finally, 3 representing WMSs were used as the input factors for each cluster. In case of cluster 1, the best results were achieved if solely calculated average values for GL and TL were taken into consideration, by contrast to cluster 2, the best results were achieved if the forecasting process was done with GL and TL of each single WMS (not the average values). The vice versa situation would mean a deterioration of 0,27% in case of cluster 1 and a deterioration of 0,43% in case of cluster 2. The conclusion is, that there is not really a statistical overall main principle for the determination of clusters and input factors and it is necessary to evaluate and optimised these for each model individually.

Region of Styria	**nRMSE (%) Cluster 1**	**nRMSE (%) Cluster 2**
_WMSs	7,87	**9,19**
_average	**7,60**	9,62

Table 4.3: Comparison of nRMSE results of cluster 1 and cluster 2. (Own results, own table)

- The high number of PV installations and overall installed capacity in cluster 1 (both nearly 70%) are essential for the forecasts matching. In the case of cluster 2, the higher distribution of PV installations within the region – especially larger area and higher sea level differences – are of a significant statistical value.
- Comparing the full load hours (MWh/MWpeak) per (sub-)cluster, the normalised values are very similar – 1.099 FLH for the cluster 1, 1.074 FLH for the cluster 2A and – as expected, due to the favourable micro-climatic conditions, the highest value – 1.159 FLH for the cluster 2B.

4.2.4 Region of Styria (chosen sample)

To come back to the core questions – *How to compile the best regional 3D clustering model for forecasting of PV electricity generation respecting the geographical and climatological specifics of the location of PV installations, on the example of PV systems which are located in federal state of Styria, registered by the “Öko-Bilanzgruppe” [Eco Balance Group] and provide the real measured data of* *their generation by a "feed-in profile meter" at least over period of 1 year? How is the forecast accuracy of different prediction models in dependence on number of defined (sub-)clusters and number of weather measuring stations taken into account?* – the PV electricity generation forecast based on different prediction models is implemented and evaluated for the chosen sample of the region of Styria.

In the first step, the forecast is based on global irradiance forecast values (GL) and ambient temperature forecast values (TL) of one WMS for the whole region of Styria (chosen sample). As described in 3.4, the WMSs with the overall best correlation are Fürstenfeld (91,4%) and Graz Universität (91,2%), which gives a basis for following prediction models:

- ST_Fürstenfeld
- ST_Graz Universität

In the second step, the idea of finer clustering and additional representing WMSs is implemented. Based on the best results of above analysed prediction models for cluster 1 and cluster 2, the forecasting models for the region of Styria (chosen sample) consist of following:

- ST like Cl1 & Cl2-WMSsCl2A
- ST like Cl1 & Cl2-WMSsCl2ACl2B
- ST like Cl1 & Cl2A & Cl2B

The models "ST like Cl1 & Cl2A & Cl2B" and "ST like Cl1 & Cl2_WMSsCl2ACl2B" have nearly the same basis and provide the best results, i.e. the lowest nRMSE (7,07% or 7,08%) and nMAE (4,49% or 4,51%) are achieved if the region is divided in clusters according to the position of the PV systems above sea level and several representing WMSs – in this case overall 6 WMSs – are taken into account. In comparison to the forecast process based on one measuring station, this approach gives an nRMSE improvement of at least 0,61% (compared to the model "ST_Fürstenfeld").

Number of WMSs / Model	**nRMSE (%)**	**nMAE (%)**
1 / ST_Fürstenfeld	**7,68**	**5,06**
1 / ST_Graz Universität	7,72	5,23
3 / ST like Cl1 & Cl2-WMSsCl2A	7,28	4,64
6 / ST like Cl1 & Cl2-WMSsCl2ACl2B	7,08	4,51
6 / ST like Cl1 &Cl2A & Cl2B	**7,07**	**4,49**

Table 4.4: Results of nRMSE and nMAE evaluation of region of Styria. (Own results, own table)

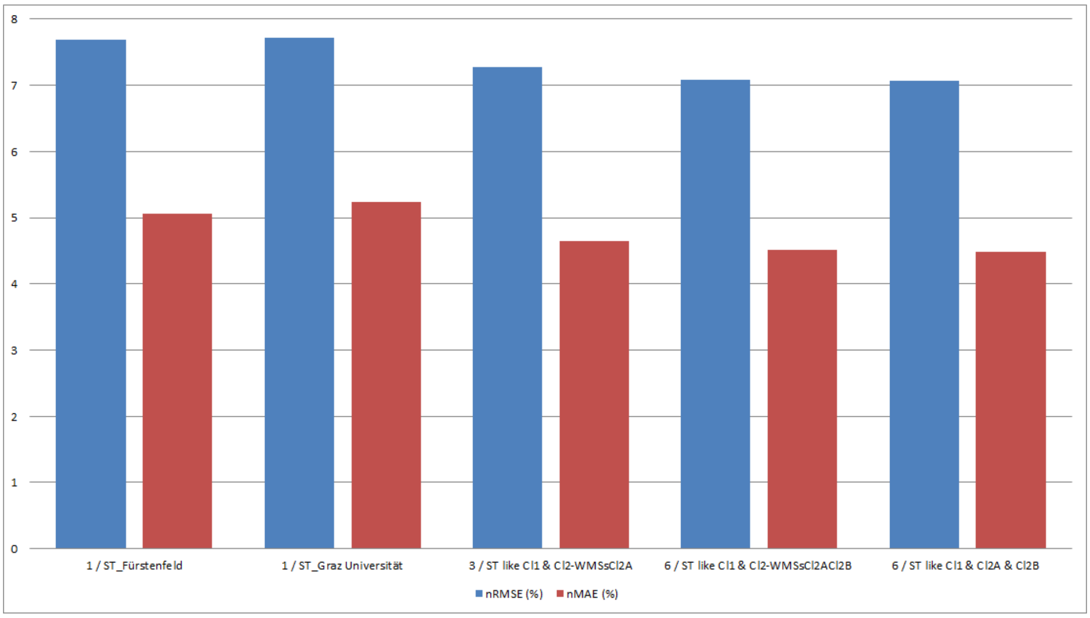


Figure 4.3: Graphical representation of nRMSE and nMAE evaluation of region of Styria. (Own results, own figure)

As explained before, in case of determination of regional clusters as well as of prediction models, there is no overall method, which could be applied as a prototype in all cases. It is necessary to evaluate and optimised all the input parameters, influencing factors and conditions very individually. Based on the results of this case study, following observations should be pointed out:

- With an increasing number of weather measuring stations (WMSs) taken into account, the forecast accuracy improves, due to the additional meteorological information, which reflects the local variations more precisely.
- Not always, the additional information of any nature has an added value for the forecast accuracy. As the results of this study showed, taking the forecasted values of global irradiance (GL) and ambient temperature (TL) for defined measuring stations only or the calculated average values only, provide the best results (the lowest nRMSE, nMAE). Taking into account both of these in one prediction model – to have some additional input factors – provide not necessarily better results (see the graphic below). Again, there is no general rule – once, this additional information gives the worst nRMSE results (see the example of cluster 1), most probably due to the overloading of the prediction model with redundant data in the very homogenous region; next, the results are close to the best nRMSE results (see the example of cluster 2 and WMSs for Cl2A), most likely due to the higher distribution of PV

installations as well as of defined representative WMSs within the large area (for details see 4.1.1, 4.1.2 and 4.2.1, 4.2.2).

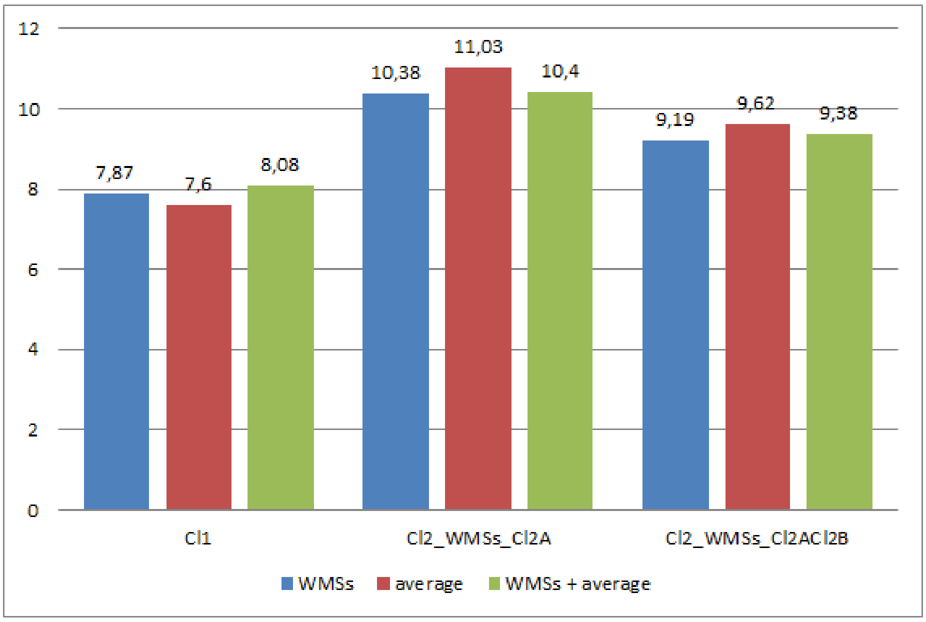


Figure 4.4: Representation of nRMSE results (%) for the cluster 1 and cluster 2 of region of Styria. (Own results, own figure)

4.3 Evaluation of the best prediction model

In the last step, the best prediction model for the PV electricity generation in region of Styria (chosen sample) – "ST like Cl1 & Cl2A & Cl2B" with the nRMSE 7,07% – is evaluated from the perspective of daily and 15-minutes raster volatility. The best prediction model is reached if there are 3 clusters established, respecting the geographical location of PV installations in the view of the sea level (by defined boarders at 500m and 1.000m), and the data from overall 6 weather measuring stations are taken in account (for details see above). Following evaluation and statements are based on the defined time period of this master thesis, from 1st May 2013 to 30th April 2014.

The calculation of nRMSE on a daily basis (daily average) shows that 61,38% of all events – which is the case of approx. 224 days in year – occurs with nRMSE lower than the yearly average value. From the nRMSE limit of 10% the number of occurrence falls down rapidly. The number of outliers with high nRMSE is relatively low. The instances with nRMSE higher than 14%, which is twice high as a yearly average, arise approx. at 5 days in year.

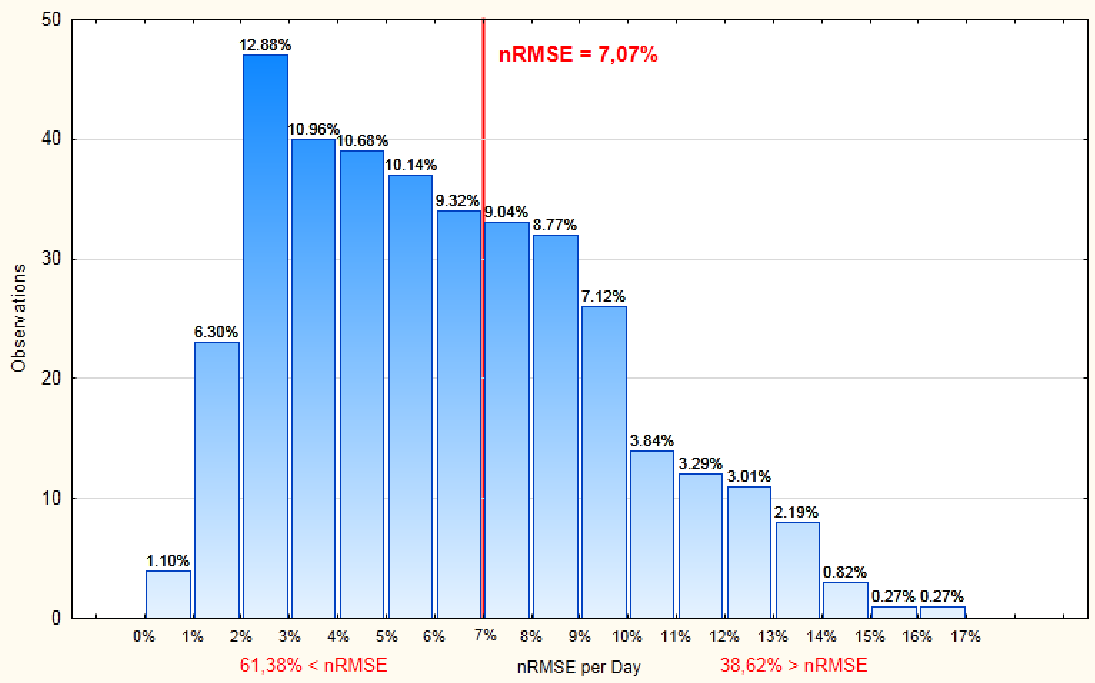


Figure 4.5: nRMSE evaluation of the best prediction model for region of Styria (chosen sample) on a daily basis. (Own results, own figure)

The evaluation of nRMSE on a 15-minutes basis (15-minutes average) illustrates the volatility during a day. This shows that 78,1% of all events occurs with nRMSE lower that the yearly average value and also the percentage of the events with higher nRMSE is quite low.

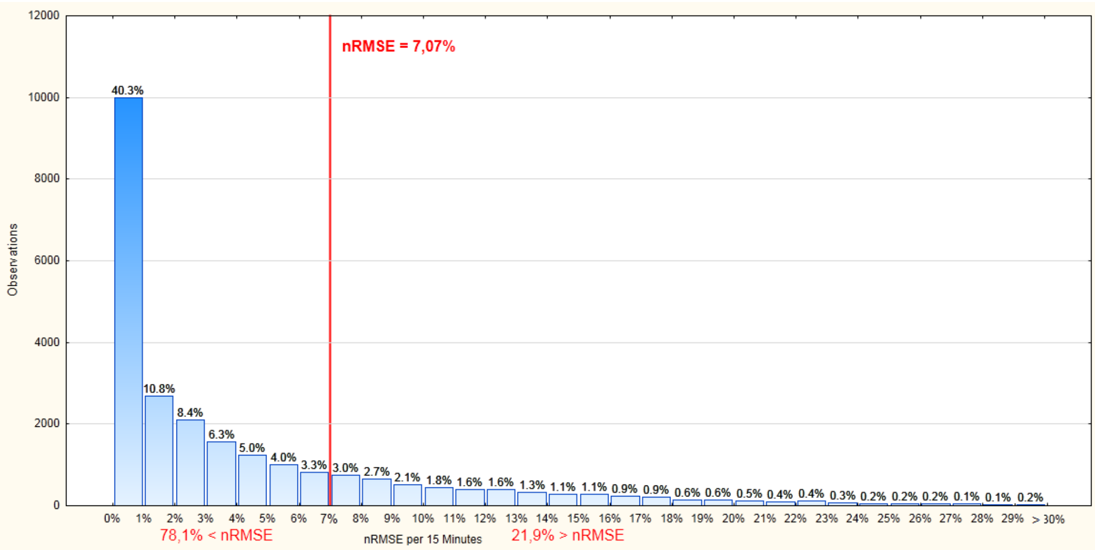
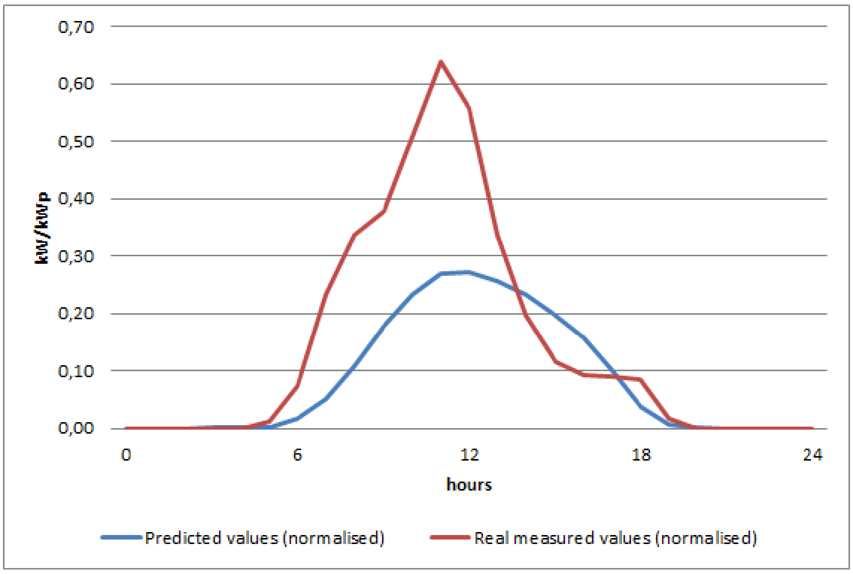


Figure 4.6: nRMSE evaluation of the best prediction model for region of Styria (chosen sample) on a 15-minutes basis. (Own results, own figure)

The nRMSE range up to 30% (and more, in very rare particular cases) shows that there are still some periods of time when the predicted and real measured values of electricity production vary extremely. The deviations can go both directions, which illustrate the graphics below. Naturally, these values reflect all the short term meteorological deviations like heavy cloud cover and fog, or snow cover of PV-modules. One example: there is a thick, long-lasting cloud cover in reality instead of a sunny day with clear sky, which was expected (or vice versa). The next example: the PV modules are covered with snow after some days of snowfall, which cause that the electricity generation is totally nullified or much lower than expected.



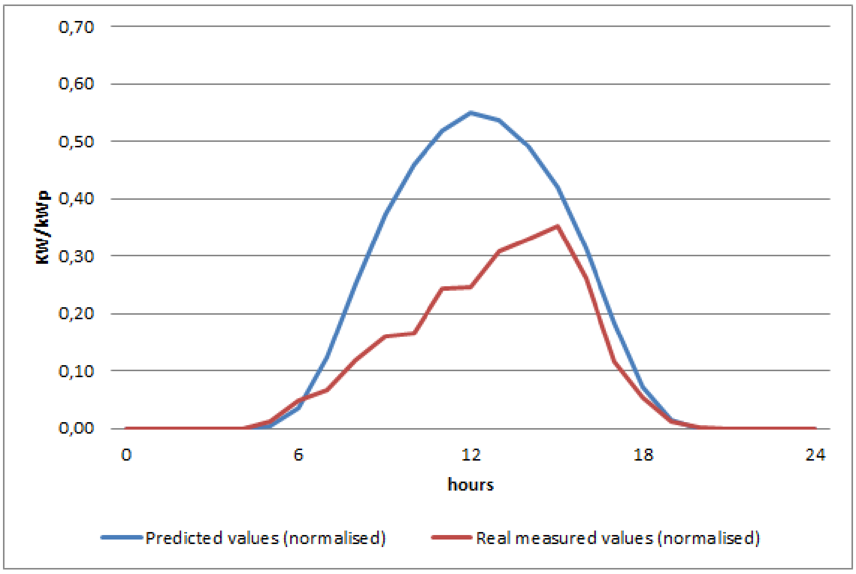


Figure 4.7: An example of very high deviations between real measured and forecasted values of PV electricity generation (normalised values). (Own results, own figures)

Closer analysis of the data based on the daily average and 15-minutes values shows that the differences between predicted and real measured values of electricity production occur more often and in stronger form in autumn and winter – as mentioned before, especially in context with heavy cloud cover, mist, fog and inversion. These deviations are in both directions – the real measured values are higher than predicted values and vice versa – and occur mostly cumulated in a period of several days running. In summer months the deviations are in general not so intensive and if so than they arise rather as isolated events.

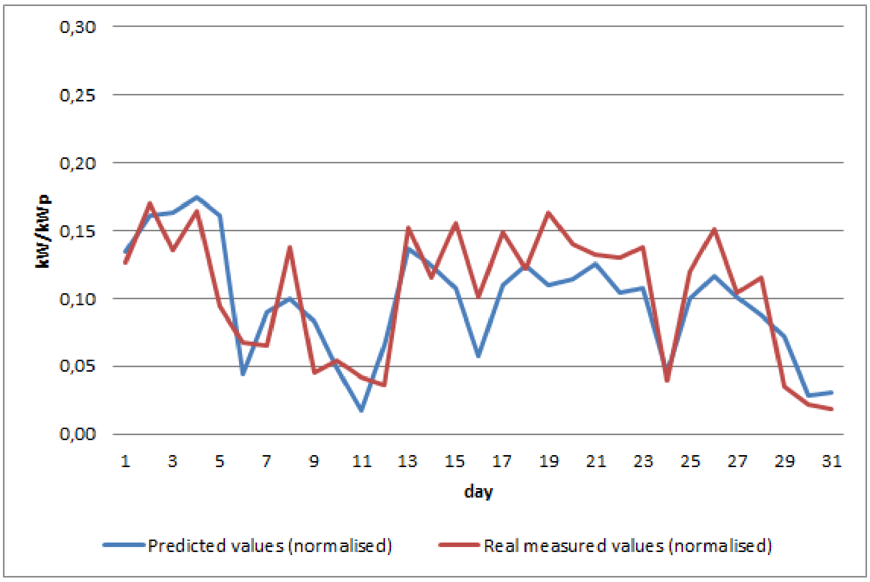


Figure 4.8: Daily average deviations between real measured and forecasted values of PV electricity generation (normalised values) in October 2013. (Own results, own figure)

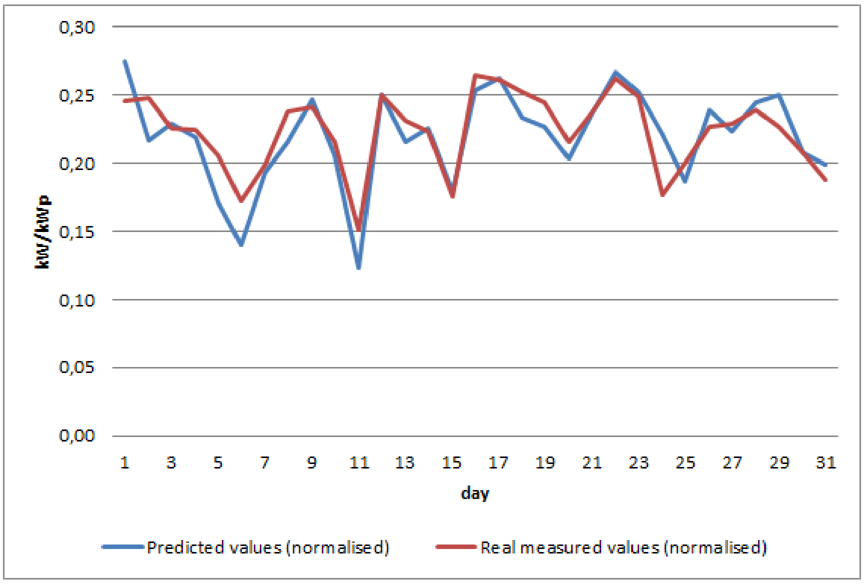


Figure 4.9: Daily average deviations between real measured and forecasted values of PV electricity generation (normalised values) in July 2013. (Own results, own figure)

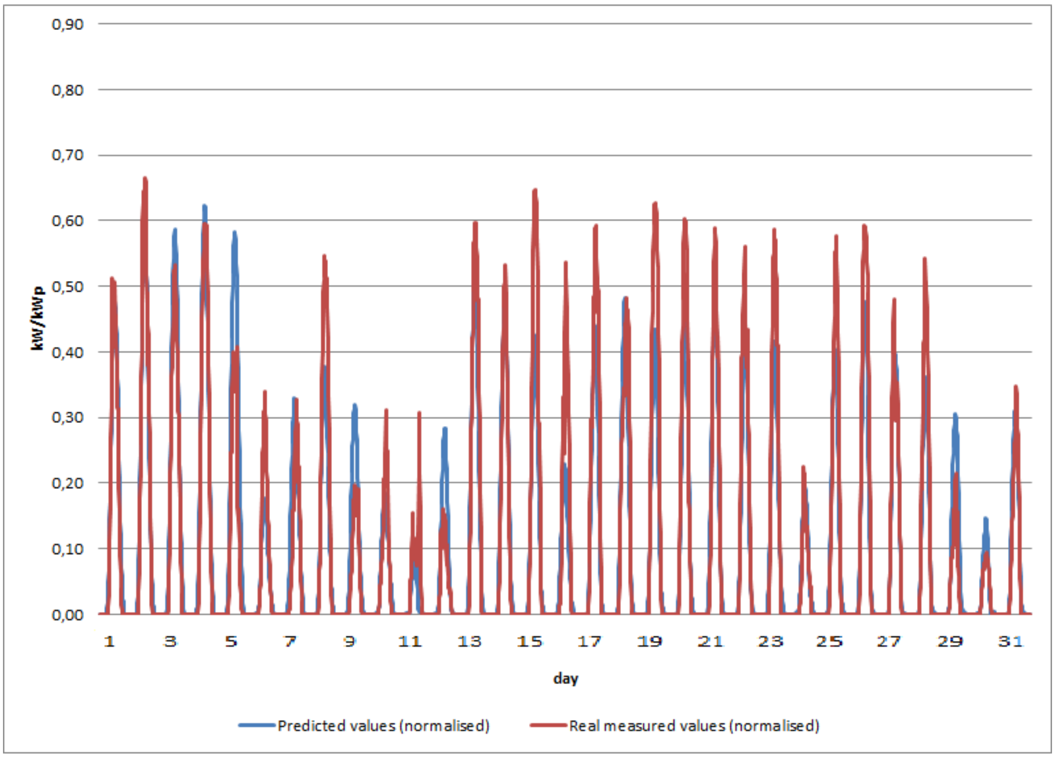


Figure 4.10: 15-minutes deviations between real measured and forecasted values of PV electricity generation (normalised values) in October 2013. (Own results, own figure)

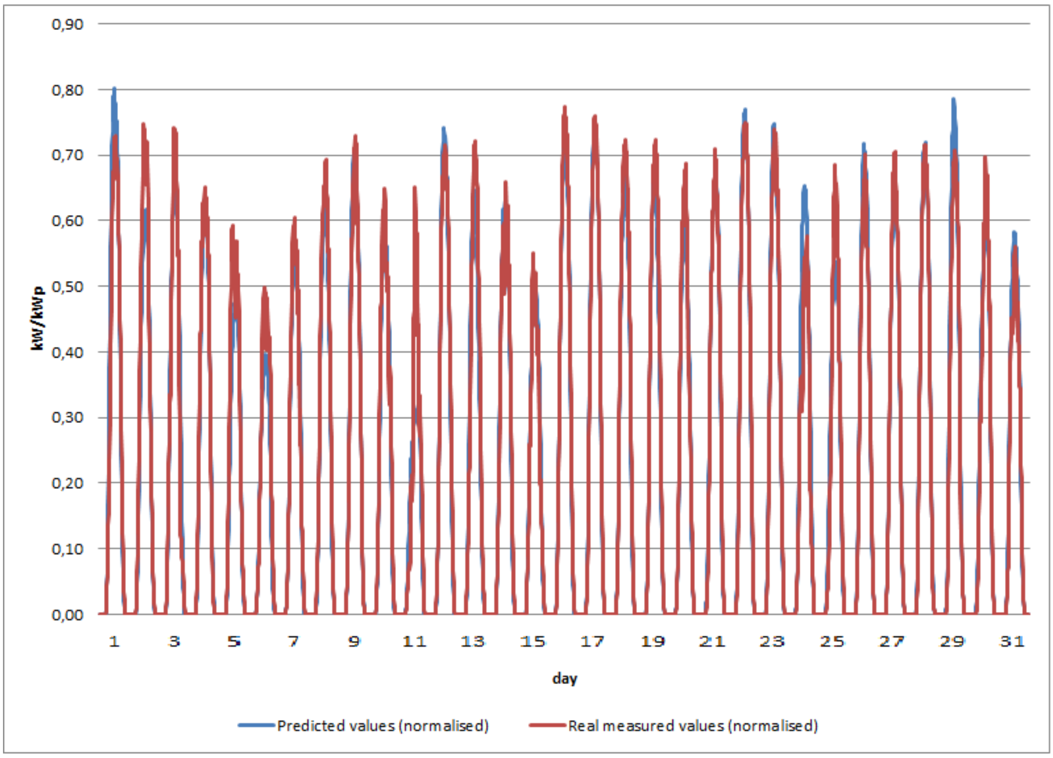


Figure 4.11: 15-minutes deviations between real measured and forecasted values of PV electricity generation (normalised values) in July 2013. (Own results, own figure)

5 SPECIFIC ISSUES OF FORECAST ACCURACY

5.1 Seasonal peaks in PV electricity production, influence of global irradiance and ambient temperature

Based on the experience from the grid operation, the seasonal influence of global irradiance and of ambient temperature on the PV electricity production is an important issue, which traces back to the geographical location of PV installations and to the clustering process. If monitoring the PV electricity output over the year (in this case based on the defined time period of this master thesis, from 1st May 2013 to 30th April 2014) under the consideration of the location of PV systems (defined by the sea level), it is obvious that the micro-climatic conditions are more favourable from approx. September to March in the alpine areas (sub-cluster 2B) and from approx. April to September in the flat areas (cluster 1); (see chapter 3 for detailed explanation). If considering the overall PV electricity output (the red line in the graphic below), it is evident, that this is determined mainly by the PV electricity generation in the flat areas (cluster 1, the green line in the graphic below), which traces back to the considerably higher density of PV installations in this area.

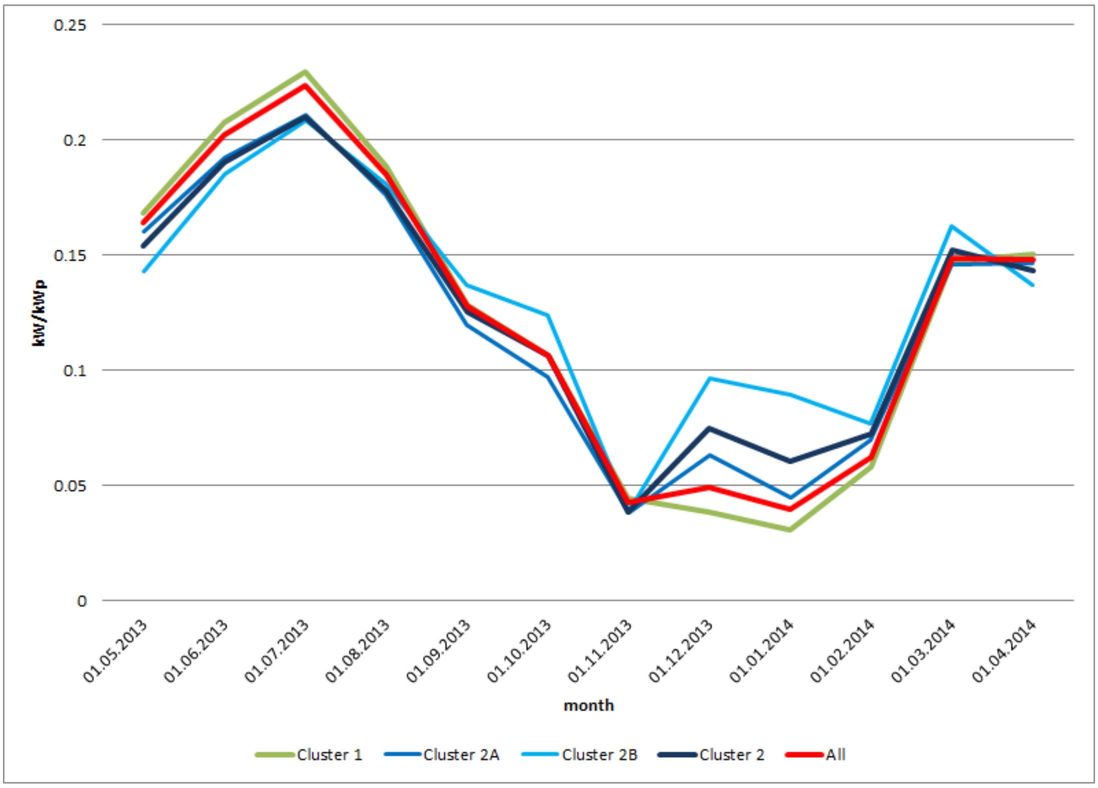


Figure 5.1: Comparison of the monthly sum of PV electricity production (normalised values) in dependence on the PV system location (defined by sea level and regional cluster).
(Own results, own figure)

To illustrate the influence of seasonal peaks of GI and Ta on the PV electricity output more in detail, the electricity generation values in months July and March are compared further on. As expected, there are some daily peaks, where the PV electricity production in the winter months is nearly so high or even higher than in summer months (see the green areas in the graphics below).

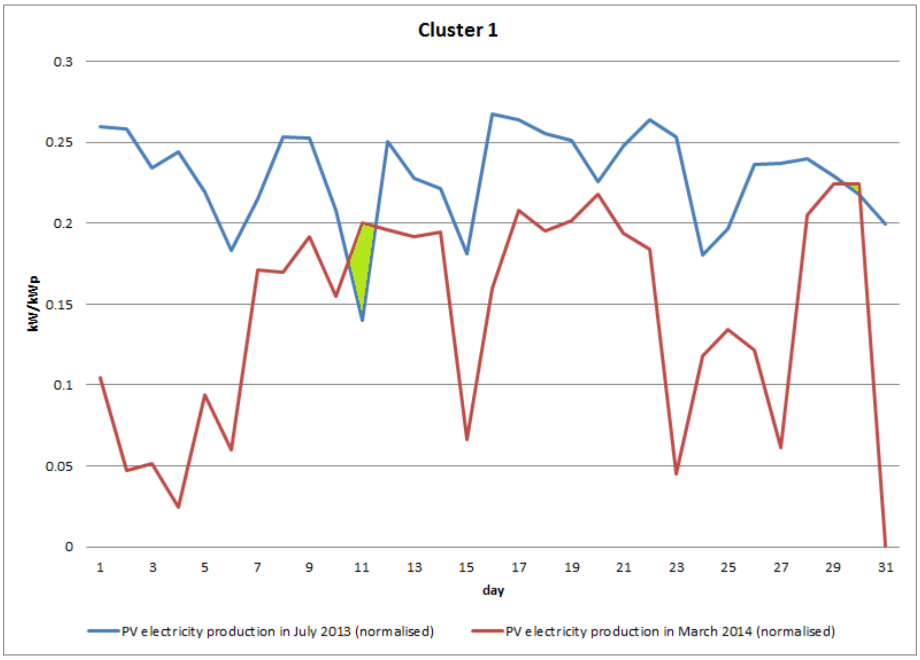


Figure 5.2: Comparison of the daily sum of PV electricity production in cluster 1 (sea level below 500m) in July 2013 and March 2014. (Own results, own figure)

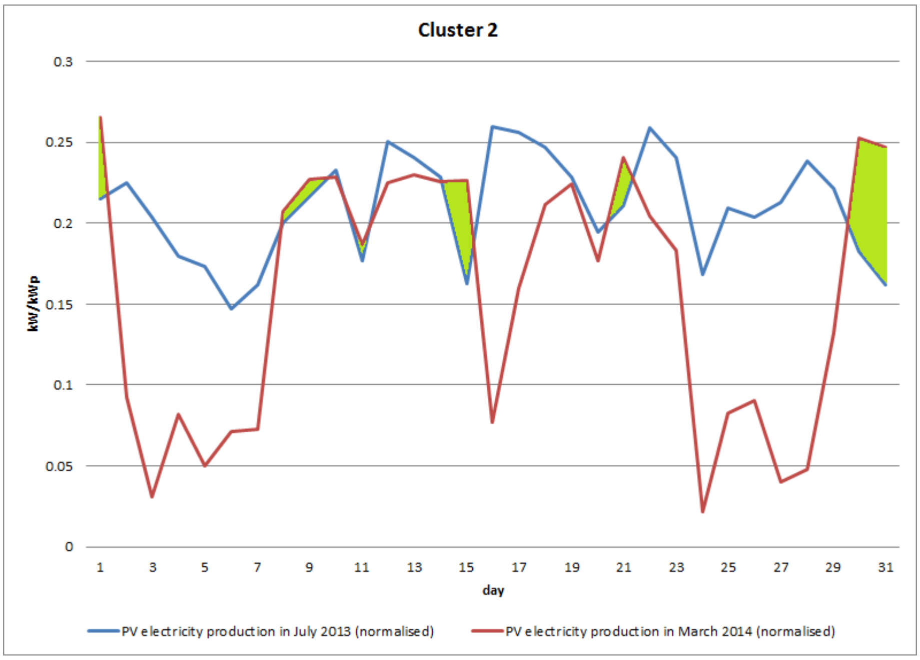


Figure 5.3: Comparison of the daily sum of PV electricity production in cluster 2 (sea level above 500m) in July 2013 and March 2014. (Own results, own figure)

As described in 3.3., the highest irradiance values are reached in July. At this time, also the ambient temperature is mostly very high, which causes losses in PV electricity production, as explained in 2.2.2. In March, there are only few days with "summer-high" irradiance values, but if so, the low temperature influences the PV electricity production in a positive way. Regarding to the advantageous micro-climatic conditions – higher GI values and lower cloud cover – this phenomenon is much more visible in mountainous and alpine areas (see the green areas in the graphic for cluster 2). The higher the PV installed capacity in this region, the higher the relevance of this issue for the grid operation. These seasonal peaks – as well as their volatility – are definitely a challenge for a forecasting process and demonstrates again the priority of carefully defined clusters and appropriate number of representative weather measuring stations.

5.2 Up-scaling

Respecting the 31th December 2013 as the date of the last available statistical data published, the overall PV installed capacity in Austria was approx. 626MWp, therefrom approx. 321MWp was registered in Öko-BG. The installed capacity in Styria was approx. 67MWp, therefrom approx. 47MWp was registered in Öko-BG.[38]

This master thesis considers PV systems which are located in federal state of Styria, registered by the Öko-BG and provides the real measured data of their electricity generation by a "feed-in profile meter" at least over a period of 1 year. The relevant date of this 1 year period is defined by the date of registration by Öko-BG by 30th April 2013 at the latest. The total installed capacity of the PV systems which match the given requirements is approx. 25MWp.

Naturally, the question on electricity production forecast of all PV systems in Styria (or even in whole Austria) appears in this context. As explained at the beginning, this is a very important issue for grid operation in charge of keeping balance between electricity production and consumption. In the case of wide area forecast it is nearly impossible to obtain detailed information of each PV system in a given area.[39] In this case, up-scaling techniques are applied in practice. In up-scaling process, a representative subset of all systems is used to generate a forecast covering all systems for a limited number of representative systems. This method

[38] Statistical data on the overall installed capacity are based on Biermayr (2013), the data on the PV systems registered in Öko-BG are provided by OeMAG via APG.
[39] Additionally, "the errors of area forecasts are typically significantly lower than the corresponding errors of forecast of single systems". Pelland (2013), p. 19.

decreases the need for extended and detailed characteristics of all PV systems and also simplifies the calculations.

Principally, the up-scaling techniques are based on two key issues. First, it is necessary to define an appropriate large subset in a way that these randomly selected systems represent well the characteristics (like location or specifications) and behaviour of the complete set of PV systems. Second, to scale the predicted electricity output of the subset to get the forecast power output of the complete set; which is described by following formula:[40]

$$P_{all,predicted} = \frac{P_{all,nominal}}{P_{representative,nominal}} \cdot P_{representative,predicted}$$

If the representative subset reflects the basic properties of the complete data set correctly, the forecast accuracy of the complete set and the subset is almost the same.

This study, the applied methods, results and statements are a good example for a detailed process of determination and evaluation of a representative subset (in case of Styria), which is the first step in the up-scaling process. In practical terms, the major challenge for up-scaling appears at the very basic level of the second step and relates to data availability and their updating.

In Austria, the registration of each PV system by a distribution system operator (DSO) is prescribed by the law. Currently, these data of overall PV installations are collected in the process of national statistics and international reporting, updated on a yearly(!) basis only and because of the incompleteness of the data partially based on extrapolations of real data. Obviously, these issues are directly linked to the forecast accuracy and following also to the costs for the balancing of the electricity system. To be able to manage these challenges, especially in context with constantly increasing amount of PV installations, a closer cooperation of all involved parties and an adaptation of legal provisions will be necessary in near future.

[40] Free according to Pelland (2013).

6 CONCLUSION

The main focus of this master thesis was to compile the best 3D regional clustering model for forecasting of PV electricity generation based on the geographical position of the PV installations and reflecting the climatological regional specifics, and to evaluate the potential for increasing the accuracy of PV electricity generation forecast in the future. Assuming that the number of PV installations in Austria will further increase, a well performed area forecast of PV electricity generation will gain in importance especially for the grid operating companies, which are responsible for keeping the balance between electricity demand and supply at any time. In this context, to create the PV electricity generation forecast as precise as possible contributes to a successful integration of higher amount of photovoltaic into the electricity system.

Due to the Austrian mountainous landscape, defining the clusters for the PV electricity production forecast on the basis of the PV system location, defined by the height above mean sea level, is assumed to be the key-criteria. Particularly, this is connected to various micro-climatic zones and reflects the important input factor characteristics, which influence the PV electricity output, like the intensity of global horizontal irradiance, ambient temperature, frequency of cloud cover, fog and inversion.

The results of this master thesis proved those assumptions and confirmed that the idea of finer clustering, respecting the geographical and climatological specifics of the 3D location of PV installations, opens the door to an improvement of the forecast accuracy in the future. With an increasing number of clusters, necessarily the number of representative weather measuring stations (provided the forecasted values of global irradiance and ambient temperature) has to be increased as well. Due to the additional meteorological information, the local variations are reflected more precisely, which leads to more exact forecasts results.

In context with the PV systems distribution / density as well as with the variability of the micro-climatic conditions over the year in different regions, one of the most challenging issues is to find or to determine the representative weather measuring stations in an appropriate number and quality. In case of this study, especially, to find representative WMSs in the alpine region (due to the height above sea level) and in the mountainous region (due to the large distribution of the PV systems) was a challenge, which required deep analysis of the regional conditions as well as mutual influences and afterwards very precise definition and optimisation of prediction models.

It was shown repetitively, that there is no overall main principle, which could be applied as a general method leading to the best regional clustering and prediction models. Since the well-determined clusters and prediction models have a crucial role in the context of the PV electricity forecast accuracy, it is necessary to evaluate and optimise all the input parameters, influencing factors and conditions for each model very individually, which is essential to bear in mind for further projects.

In the case of wide area forecast, it is nearly impossible to obtain detailed information of each PV system in a given area. Additionally, “the errors of area forecasts are typically significantly lower than the corresponding errors of forecast of single systems”.[41] In practice, the up-scaling techniques are applied to generate a wide area forecast covering all PV systems from a representative subset. This method decreases the need for extended and detailed characteristics of all PV systems and also simplifies the calculations. The results and statements of this master thesis are a good example for a detailed process of determination and evaluation of a representative subset (in case of Styria), which is the first step in the up-scaling process.

As mentioned, also a closer cooperation of PV systems operators and grid operators as well as an adaptation of legal provisions both in terms of the data availability – e.g. PV systems data like precise geographical position and overall installed capacity – is required for further development.

Since the focus of this master thesis was the day-ahead PV electricity forecast, as a consequence of the further increasing amount of PV installations in Austria and their increasing influence on the grid operation, the progress will be extended more and more to the real-time. The development of the medium- and short-term forecast procedures are the expectations for further research in this area.

[41] Pelland (2013), p. 19.

ABBREVIATIONS

ALADIN	Aire Limitée Adaptation dynamique Développement InterNational
ALARO	the newest model of ALADIN
APG	Austrian Power Grid AG
c	temperature coefficient (%)
Cl	cluster
DSO	distribution system operator
ECMWF	European Centre for Medium-Range Weather Forecasts
FLH	full load hours
GI	global (horizontal) irradiance (kW/m^2 or kWh/m^2)
GI_{mod}	global irradiance at the module surface (W/m^2)
GL	global irradiance forecast value(s)
INCA	Integrated Now-casting through Comprehensive Analysis
MARS	multivariate adaptive regression splines
mP Energy	Meta Predictor Energie (software tool)
nMAE	normalised Mean Absolute Error
nRMSE	normalised Root Mean Square Error
NWP	numerical weather prediction
NWPM	numerical weather prediction model
OeMAG	Abwicklungsstelle für Ökostrom AG
Öko-BG	Öko-Bilanzgruppe [Eco Balance Group]
P	power (kW, MW)
PR	performance ratio
PV	photovoltaic
PVGIS	Photovoltaic Geographical Information System
RES	renewable energy source(s)
SL	sea level
ST	Styria
T_a	ambient temperature (°C)
T_{mod}	module temperature (°C)

TL	ambient temperature forecast value(s)
TSM	Time Series Manager (software tool)
TSO	transmission system operator
WMS(s)	weather measuring station(s)
ZAMG	Zentralanstalt für Meteorologie und Geodynamik [the Austrian Central Institute for Meteorology and Geodynamics], short: The Austrian Institute for Meteorology

REFERENCES

List of Literature

Biermayr (2013): Biermayr P., et al., *Innovative Energietechnologien in Österreich, Marketentwicklung 2013*, Bundesministerium für Verkehr, Innovation und Technologie, Berichte aus Energie- und Umweltforschung, 26/2014.

Drews (2007): Drews A., et al., *Monitoring and remote failure detection of grid-connected PV systems based on satellite observations*, Solar Energy 81(4), 2007.

Häberlin (2010): Häberlin H., *Photovolatik – Strom aus Sonnenlicht für Verbundnetz und Inselanlagen*, VDE Verlag, 2010.

Pelland (2013): Pelland S., et al., *Photovoltaic and Solar Forecasting: State of the Art*, Report International Energy Agency (IEA) Photovoltaics power systems programme (PVPS), Task 14, Subtask 3.1, 2013.

Podesser (2010): Podesser A., et al., *Klimaatlas Steiermark – Kapitel 1: Strahlung*, Zentralanstalt für Meteorologie und Geodynamik (ZAMG), 2010.

Podesser, Wölfelmaier (2010): Podesser A., Wölfelmaier F., et al., *Klimaatlas Steiermark – Kapitel 3: Luftfeuchtigkeit, Bewölkung, Nebel*, Zentralanstalt für Meteorologie und Geodynamik (ZAMG), 2010.

Wakonigg (2010): Wakonigg H., et al., *Klimaatlas Steiermark – Kapitel 2: Temperatur*, Zentralanstalt für Meteorologie und Geodynamik (ZAMG), 2010.

Internet Sources

OeMAG:

http://www.oem-ag.at/de/home/ (18.05.2014)

http://www.oem-ag.at/de/gesetze-regelwerk/ (18.05.2014)

http://www.oem-ag.at/de/oemag/service/faqs/ (18.05.2014)

PV Austria:

http://www.pvaustria.at/daten-fakten/technologie/pv-auslegung/ (07.06.2014)

http://www.pvaustria.at/wp-content/uploads/2013/07/286_Energieprofil.jpg (07.06.2014)

http://www.pvaustria.at/wp-content/uploads/2013/07/160_Energieertrag-Ausrichtung-Dachneigung.jpg (07.06.2014)

PVGIS:

http://re.jrc.ec.europa.eu/pvgis/countries/europe/g13y_at.png (14.07.2014)

http://re.jrc.ec.europa.eu/pvgis/countries/europe/ffg13.optyc.at.png (14.07.2014)

Solaik:

http://www.solaik.ch/photovoltaik/grundlagen/index.php# (14.07.2014)

Solarenergieförderverein Bayern:

http://www.sev-bayern.de/content/Objektforum.pdf (14.06.2014),
Becker G., *Gebäudeintegrierte Photovoltaik (GIPV/BIPV)*, presentation, p. 19, p. 23.

ZAMG:

http://www.zamg.ac.at/cms/de/forschung/wetter/alaro (14.05.2014)

http://www.zamg.ac.at/cms/de/forschung/wetter/inca (14.05.2014)