Review of the Operational Performance of Grid-Connected PV Systems

Utrecht University - Faculty of Geosciences
Sustainable Development Master's Program
Energy and Resources

Author

Moraitis Panagiotis
3717208

Supervisor

Dr. Wilfried van Sark

2014
UTRECHT

Abstract

The main objective of this paper is to review the operational performance of grid connected residential PV systems by comparing the Performance Ratio and Annual System Yield indicators. In order to collect a large number of high resolution data web scraping techniques were applied and as a result a total of 2693 systems form Netherlands, Germany, Belgium, Italy and France are analyzed. The installations are examined according to geographical and seasonal criteria and a further performance analysis is conducted in comparison with the type of module and the inverter that are used. The Performance Ratio in the Netherlands has an average value of 78% and the Annual Specific Yield was 874 kWh/kW_p for the year 2013. As a result of the local meteorological variations, the coastal part of the country produces 7% higher yields is 3% more efficient compared to the mainland. The higher average annual specific yield was observed in South Italy in 2011 (1352 kWh/kW_p) and the lower in the Netherlands in 2012 (865 kWh/kW_p).

Contents

1 Introduction

1.1 Motivation

The main objective of this paper is to review the operational performance of grid-connected photovoltaic (PV) systems in the Netherlands and to analyze the results in comparison with other European countries. The constantly growing PV market demands high energy yields throughout a constant increase in the performance. Over the last years the development of the solar technology, and the low installation and maintenance costs made solar panels a popular form of renewable energy production. Especially, small and medium size domestic users have started to embrace solar technology in order to reduce their utility bills.

In addition to other renewable energy forms, such as wind power or geothermal, that require both a high capital and expert's knowledge for their operation, PV demands only a few square meters of rooftop and it is affordable for the average house owner. This form of decentralized energy production, might appear like a prosperous perspective for a future energy scenario but the result might not be as flourishing as it was initially projected.

The main issue that rises from the current situation is that most of the owners are non experts, and consequently their lack of experience and technical skills might be a hindering factor for the optimum operation of their systems. As the majority of domestic production is coming from systems that are less than $5kW_p$ and due to economical reasons, these systems are lacking any kind of monitoring system, failures and energy losses remain undetected for a long time. [1]

Therefore it is essential to find a way to monitor and study the system performance of the scattered individual producers in order to reach higher yields and identify system failures and losses.

1.2 System Monitoring

System performance monitoring shows a steady progress over the years, for example the Performance Ratio indicator (PR) in the late 1980's had average values between 50% to 70% and nowadays it exceeds 80% in most cases [2]. However, systematic recording and evaluation of data occurred in very few places.

The first and most important attempt was the 1000-roofs program in Germany that provided operational performance data for grid connected systems from the beginning of 1990's till 2002. Different sites, applications, mounting types and installation capacities were evaluated for more than 10 years to reveal the trends in performance and annual yields [3].

Also in the Netherlands 500 decentralized building integrated systems of total $1MW_p$ were installed and monitored for 5 years in Amersfoort from 1999. Despite the fact that the program started as an attempt to familiarize architects and building inhabitants with PV technology, it was also a great opportunity for large scale performance monitoring. The process revealed system design errors and equipment failures that were remaining unnoticed but also seasonal variations were observed, methodological approaches were tested in practice and evaluated but also behavioral and social conclusions were extracted [4].

The above studies revealed the great progress in performance the last decades. The question that is now raised is how high the target could be and it is inspired by technological improvements in inverter efficiencies and system design. For that reason, system simulations to detect and fully understand loss mechanism among the top performing systems were performed to prove that further optimization is still possible. [5].

1.3 Research Questions

However, regardless of how important the further improvement of performance using high technology equipment and thorough system analysis is, the attention should remain on the decentralized nature of solar energy which includes a variety of sites and operational conditions among with owners that are usually lacking of proper training and knowledge.

The focus of monitoring attempts should remain on gathering systematic representative data from various locations to track the performance of individual producers in real life situations with all the uncertainty factors that are included. The rapid development of the domestic PV market along with the diffusion of the World Wide Web in our daily life increased the availability of data. Independent users can automatically or manually upload their energy production on the web and therefore a network of small and large scale installations is constructed.

The aim of this paper is the creation of an international data base that makes feasible to record the trends of the performance indicators of each site and analyze the results in order to determine the level of efficiency of the domestic PV energy production. The results will be a reference point for evaluation of different sites among countries, or between individual countries to reveal weak points that cause unsatisfactory performance. For the aforementioned purpose the following research questions will be investigated.

- How efficient is the energy production from PV systems in the Netherlands? What does the PV systems quality analysis reveal?
  - What is the monthly mean value of PR and annual system yield per site?
  - How do the performance indicators change for each region through the year?
  - What is the distribution of the results?
  - Are there seasonal variations through a yearly period?
  - How do the specific characteristics of each system affect the performance?
- Comparison of the results with other European countries and previous studies, how performance varies by region, was there any progress during the last years?

- Are there major differences between regions? How could this be explained? Is it possible to create a map showing regional variations of performance?
- What causes quality differences among PV systems? How does the climate conditions affect the performance?
- How much electricity do PV systems produce in terms of kWh per installed kWp?

1.4 Theoretical Background

The time and weather dependent nature of solar power makes it difficult to apply the conventional performance indicators that we are using for regular power plants. Capacity Factor, Efficiency and the Availability are the following dimensionless equations that represent the traditional way to measure the output of a utility power plant. However, when it comes to evaluating the performance of a PV system the above indicators may lead to wrong conclusions as they could leave the reader wondering whether the values are satisfactory or not.

$$Capacity\ Factor = \frac{Generation}{RatedPower\ Hours} \tag{1}$$

$$Efficiency = \frac{GenerationOut}{SolarFuelIn} \tag{2}$$

$$Availability = \frac{ServiceHours}{TotalHours} \tag{3}$$

As an example, the Capacity Factor of a PV system is in the range of 14%-24%[6], while a nuclear power plant can reach up to 90%[7].

Consequently, without deeper knowledge of the specific features of each individual PV system the small percentages could result in misinterpretation. Moreover, when it comes to solar energy, the above figures vary considerably for PV, even when the system is operating according to design. Any of the above values could denote excellent performance at one site and poor performance under different settings and climate conditions. Therefore there is a need for a dimensionless indicator that yields a 100% value for proper operation. [8].

1.4.1 Final System Yield

In order to be able to compare and evaluate different PV systems normalized performance indicators are necessary. For energy yield and efficiency, the values are normalized according to the nominal array power and to the PV area, respectively[9].

According to IEC Standard 61724 the final PV system yield Y_f is the net energy delivered for the specific time period, divided by the rated power output of the installed array and it has units of kWh/kWp[10].

$$Y_f = \frac{E}{P_0} \tag{4}$$

It is a convenient way to compare the energy produced by different PV systems as it normalizes the energy produced according to system size[11]. It has the advantage to be a straightforward indicator as the only measurement that it requires is the actual produced energy. However, it varies widely by climate, by the length of the calculation period and by how the two parameters are defined (e.g., array DC level or inverter AC output)[8].

For PV systems, the system yield is more widely used than the capacity factor despite the fact that the latter has the appealing property of being dimensionless. In fact the two terms are fundamentally very similar, as by dividing the annual system yield with the constant factor of 8,760 hours per year we produce a term that is equivalent in meaning with the annual capacity factor[8].

1.4.2 Performance Ratio

The performance ratio (PR) is a quality factor that was introduced as the degree of utilization of an entire PV system. It is stated as a percentage rate and describes the ratio between the actual and the theoretical expected energy yield. The actual energy yield is the utilizable AC electricity that it is measured at the feed in meter and it is divided by the amount of energy that could be generated if the PV system operated under Standard Test Conditions (STC) [2].

The main purpose of PR is to indicate the overall effect of losses on the overall performance of the PV system, and includes the effects of PV array temperature, incomplete utilization of irradiation, system component limited efficiencies, and failures[5]. The PR is a dimensionless quality and it is calculated by dividing the final system yield Y_f by the reference yield Y_r[10].

$$PR = \frac{Y_f}{Y_r} \tag{5}$$

The final yield is defined by the load efficiency of the array (n), which is the power output divided by the total input, multiplied with the energy recording interval (τ) and the ratio of the total measured power output (P_A) to the rated power output (P_0). The reference yield is defined by the total plane of array irradiance divided by the reference irradiation which is $1000W/m^2$.

$$Y_f = n_{Load}\tau\frac{\sum_{day} P_A}{P_0} \tag{6}$$

$$Y_r = \tau\frac{\sum_{day} G_{POA}}{G_{ref}} \tag{7}$$

Consequently, the PR has the advantage to be independent of the irradiation conditions and as a result to be also independent of the specific site and orientation of the module and finally, independent from the local weather conditions. Therefore it is a step further towards the evaluation of differently designed systems or equal systems installed in different locations. The difference between 1 and PR aggregates all the possible energy losses including inverter efficiency, wire losses, real power of the PV modules below nominal rating, mismatch, shades, dust, thermal, failures and in larger systems mid-voltage transformer losses that are also influencing the final value[12].

All the above factors are limiting the final value of PR in the range of $70\% - 80\%$ but the versatility of the indicator is mainly affected by the dependence on temperature. System performance will fluctuate depending on the climate conditions as it is expected to observe significantly lower PR in higher temperatures compared with colder ones and even within the same year seasonal variations may occur for the same reason.

To avoid such thermal fluctuations the PR_{STC} was introduced and includes the temperature under which the modules operate. It requires more complex calculations but it becomes practically independent from time and site and therefore more strictly appropriate for technical quality evaluation[12]. However, even under these conditions an excellent quality and properly maintained PV system will still have a PR lower than 1 as there will still be technical losses and inefficiencies.

2 Methodology

2.1 Web Data Collection Techniques

For the creation of an international database with solar performance indicators it is essential to have access to a large number of legitimate and systematic data recordings. The monitoring market consists of inverter manufacturers, project developers and independent monitoring vendors that integrate software and hardware in order to provide better customer service. Moreover, in that way they are able to manage their portfolio of systems through web-platforms and compare the actual system output with the estimated performance.

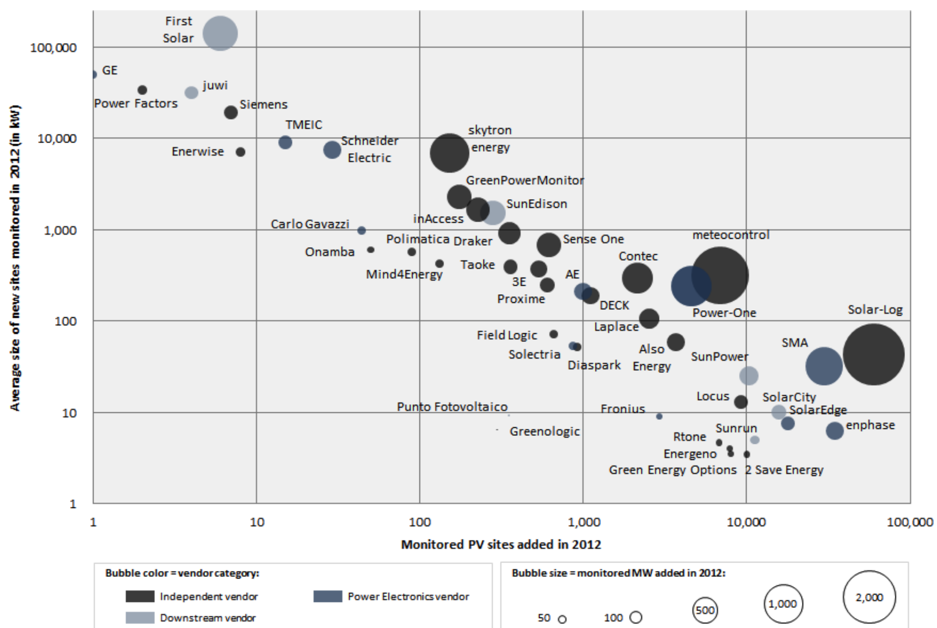

Figure 1: Market share of monitoring companies for 2012. Source: *Global PV Monitoring: Technologies, Markets and Leading Players, 2013-2017* [15]

Solar-Log is one of the major key players in monitoring [15] and one of the very few companies that offers free access to the online web-platform. More than 80,000 systems are using this service globally and approximately 800 of them are located in the Netherlands (by the end of 2013) (http://home.solarlog-web.nl/plants.html). Beyond the AC and DC energy yield that is uploaded daily for each system, the service also provides further detailed information such as the geographical location which consists of the post code, the capacity of the system in kW_p, the type of the module and the inverter, the orientation and the slope of the panel and the installation year. For the data collection process, a data logger that measures the power is used for every installation, the measurements then are sent to Solar-Log for processing and the user is able to watch the performance indicators of his plant through a computer or a smart phone.

Unfortunately, there is no automated way to download any of the above data as they are only presented individually in separate tables and the only option to store them is to manually copy and paste them from the web browser to a local file in the computer. This is a very tedious procedure that could take months or years to be completed. The most efficient technique that can be applied is web data collection, which is a computer software method of extracting information from websites. Through that, it is possible to simulate the human navigation through the web sources and also, automatically locate and save scattered information that is available to the user in to organized and structured data sheets. However, there is not a specific tool for that, as all the web sources have a different structure and therefore a unique software had to be developed.

In order to extract online data, first it is necessary to create a computer program that is able to read the HTML (Hyper Text Markup Language) code that is used for the creation of a web page. For that, Python 3.3.2 was used with the module BeautifulSoup 4.3.2. Python was chosen as it is one of the most popular object oriented languages and it is distinguished by its large and active scientific library support [16].

Figure 2: Data Collection scheme.

The software that was created (Appendix A), is able to locate which part of the HTML code refers to a PV system, identify the operational details of each installation, make a copy of them in to a txt file and then to link with the individual page of the specific site where the today's yield is uploaded. The DC and the AC value of the produced energy is copied below the site's information in the same txt file with the today's date. Finally there is a loop function that is repeating that process for the previous dates like a user would click the "Previous" button on the screen. This method, of getting the present day and then going back in the past was chosen in order to be able to easily reuse the software at any point in the future. For example, if the user wants all the values of the previous year he will set the function to be repeated 365 times.

Moreover, as there are some malfunctioning installations, where for unspecified reasons a large number of data entries is missing the software has a safety trigger and is automatically skipping the problematic installation if there are more than 100 zero values. This function is working as a primary quality control to ensure the legitimacy of the data base.

For optimum performance the data base was formatted in two tables using Python Pandas 0.12.0. The first one is a multindex structure where the yields are placed from the txt file according to recording date and the name of the installation site, and the second one is a look up table with the individual operational details of each one of them.

2.2 Irradiation Measurements

2.2.1 Irradiation on Inclined Surface Models

According to PR definition and the given formula (5) the total plane of array irradiation is necessary. For that reason the Royal Meteorological Institute of the Netherlands was used, as they are measuring hourly the global horizontal irradiation from 1951 till the present day.

The incident global horizontal irradiation can be divided in to three components, the beam component from the direct irradiation on the horizontal surface, the diffuse component and the component from ground reflections. The contribution of the diffuse component to the total value could be from 25% on a sunny day up to 80% on a cloudy day[17]. Furthermore, as it is dictated by best practice techniques, the majority of the solar panels are tilted toward the sun to maximize the amount of solar radiation on the cell surface.

Therefore, solar radiation incident on an inclined surface has to be calculated by converting the value measured on a horizontal surface to that incident on the tilted surface of interest. However, this is not possible by just applying the geometrical relationship between the two surfaces as the diffuse radiation comes from every point of the sky [17].

A number of models for determining the solar global irradiation on inclined surfaces derived from the global horizontal have been developed and according to studies, Olmo et al. model was found to have better match between the predicted and the experimental values [18]. Moreover, it has the advantage to depend only on the clearness index and avoids the separation of the solar beam in to direct and diffuse components. The global irradiance I_β on on an inclined surface derived from the corresponding global radiation I on a horizontal surface is given by the following equation [19]:

$$I_\beta = I\psi_0 F_c \tag{8}$$

Where ψ_0 is a function of the incident angle θ and the solar zenith angle θ_z, and F_c is the component of the anisotropic reflections from the ground (see Figure 3).

$$\psi_0 = exp[-k_t(\theta^2 - \theta_z^2)] \tag{9}$$

$$F_c = 1 + \rho sin^2(\theta/2) \tag{10}$$

Where, ρ is the albedo of the surface and in this research a constant value of 0.25 was used. The hourly clearness index k_t is the ratio of the global horizontal irradiance to the extraterrestrial horizontal irradiance which has an average value of $G_0 = 1367W/m^2$. The θ and the θ_z angles are given by the following formulas:

$$cos\theta = sin\delta sin(\phi - \beta) + cos\delta cos\phi cos\omega \tag{11}$$

$$cos\theta_z = sin\delta sin\phi + cos\delta cos\phi cos\omega \tag{12}$$

Where β is the tilt of the panel and ϕ is the latitude. The solar hour ω is the angle through which the Earth has rotated since solar noon [20]:

$$\omega = (15°h^{-1})(t_{zone} - 12h) + (\psi - \psi_{zone}) \tag{13}$$

Where t_{zone} is the local civil time, ψ is the longitude and ψ_{zone} is the longitude where the solar and the civil time coincide. Declination, δ is defined as the angle between the Sun's direction and the equatorial plane [20]:

$$\delta = \delta_0 sin\left(\frac{360°(284 + n)}{365}\right) \tag{14}$$

Where n is the day of the year and δ_0 is 23.45°. The definition of the angles that are used as coordinates are presented on the following figure.

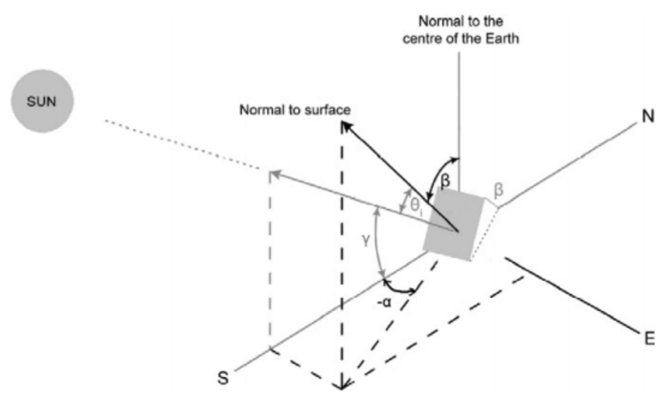

Figure 3: Sky dome showing the solar zenith geometry. [18]

2.2.2 Irradiation in the Netherlands

In the Netherlands there are 31 stations that are measuring the global horizontal irradiation and therefore the country was separated in 31 areas based on the proximity of each post code to the local meteorological station according to the first two digits of the post code. The result is the map below:

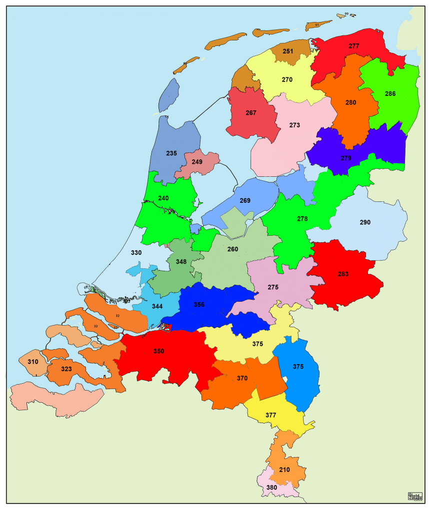

Figure 4: The Netherlands divided in areas according to the nearest meteorological station. The number refers to the meteorological station code.

The solar panel installations have geographical information based on the postcode of each location, therefore according to that and the map above each one of them was automatically linked to the nearest meteorological station to retrieve irradiation data. The plane of array irradiation was calculated on an hourly basis based on the above formulas and then it was summed daily. The geographical latitude ϕ of each installation was calculated using Python's pygeocoder 1.2.1.1 and it was derived from the post code of each location.

2.3 PR calculation, Quality Control and Error Propagation

The PR values for the Netherlands were calculated according to formula (5). The annual AC energy yield (kWh) divided by the total system capacity (kW_p) is giving the final system yield Y_f. Then the total plane of array irradiation for the given year (kWh) divided by the reference value ($1/kW_p$) is the reference yield Y_r. Unfortunately, from the list of countries that were selected for this research only the Royal Meteorological institute of the Netherlands (KNMI) is publishing solar irradiation measurements with a high resolution (hourly) and therefore the calculation of PR was not feasible for the rest. However, the final system Yield was calculated according to formula (4).

Every value that is measured has a level of uncertainty, for example the uncertainty of the energy yield measurement can be as low as $0.1 - 0.2\%$ for inspection equipment but for the regular domestic production the electricity meter that is installed in households has accuracy of $1 - 2\%$ [2]. The same level of uncertainty applies for the irradiation equipment, however for the Olmo's model the uncertainty level could reach 7%, especially in seasons with a lot of rainfall and less sunshine [18]. As every value is affected by a large number of errors, the Mean Square Error was selected to quantify the uncertainty around the estimate of the mean measurement.

The first quality control of the data set includes a primary test that sets a quantitative boundary to all the sites that were not operational for 95% of a year. Practically, it excludes any installation with less than 340 day-entries annually, and in that way locations that were offline for some period or that were subscribed at some point in the year are not taken in to consideration as they have not completed a full operational year.

Moreover, a second quality check was performed after the calculation of PR and Y_f in order to further detect possible mistakes during the data entry of every installation's operational details and exclude measurements beyond the samples readability boundaries. The reliability boundaries for each distribution is set to 3 times the standard deviation (σ) . Therefore, locations with higher or lower PR and Y_f values were checked manually and those that could not be interpreted neither statistically or physically were ignored.

3 Results

3.1 Analysis of the Dutch PV systems

The Dutch sample consists of 728 installations and the majority of them has been monitored since 2011. As it is shown in Figure 5 bellow that was constructed according to the geographical coordinates of each location, the sample is well distributed all over the country.

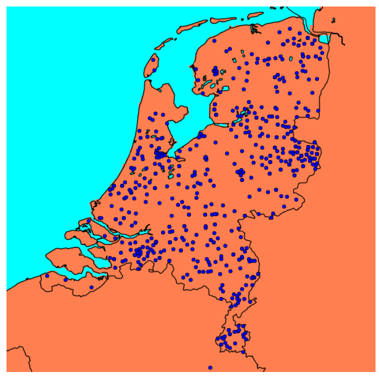

Figure 5: Sample locations in the Netherlands.

The total capacity installed of the monitored systems is 8072 kW_p which is only a small fraction of the total installed capacity in the Netherlands that was 370 MW_p by the end of 2012 (EPIA, Annual Report 2012). The average system size is $11kW_p$ with 40% of the installations being smaller than $5kW_p$ placing the sample in to a mid-range category of grid-connected PV systems.

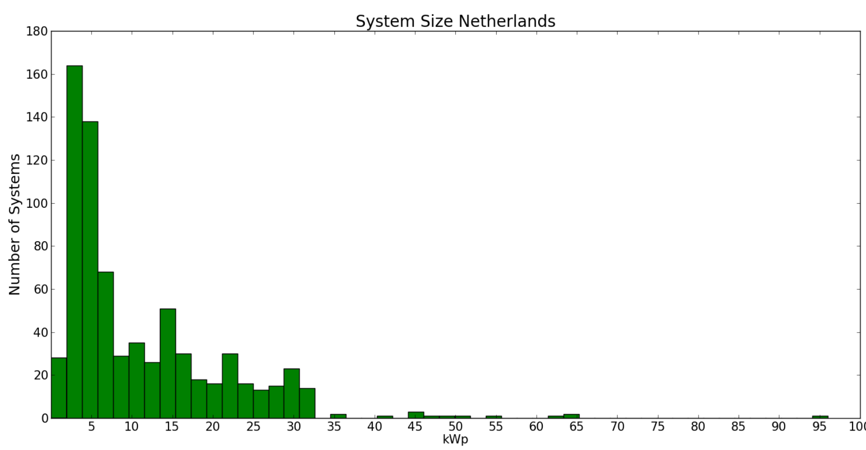

Figure 6: The system capacity distribution.

The majority of the modules is made out of polycrystalline silicon cells which is the most popular type as they can be manufactured rather easily by composing small silicon grains of various crystallographic orientation [9]. The result is a relatively cheap panel but less efficient compared to the competitive technologies. On the contrary mono crystalline modules are fabricated by a homogeneous form of silicon that is able to maintain the electronic properties throughout the material and consequently be more efficient and more expensive. Amorphous solar cells were known till recently for small scale applications like pocket calculators, but improvements in construction techniques made them more attractive for solar panels as their inherit low efficiency can be made up by their thinness and the low production cost. Hybrid solar cells is a new technology which combines the advantages of organic and inorganic semiconductors, it is not fully commercial yet and only 1% of the sample is using it.

A major part of the modules is made of polycrystalline silicon (66.9%) and 25.4% of mono crystalline which contributes to 63% and 21% respectively to the total installed capacity. The solar panels that are based on Amorphous silicon technology are 5.1% of the sample but they correspond to 10% of the installed capacity. Therefore its seems to be an appealing option for producers with bigger systems as the average system capacity of an installation with this type of panels is $22kW_p$. Only 1.6% of the owners did not specify what kind of panels have been installed.

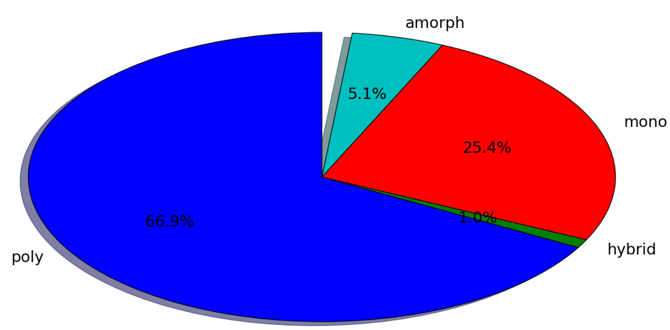

Figure 7: Panel type of the modules

The orientation and the slope of the panel is a critical factor for the performance of a PV system. The optimum orientation for the Northen hemisphere is directly South (180°) and for the Netherlands the optimum slope is 37°. As it shown in the graph below 43% of the systems are oriented directly South and 95% is in a range of ±15°. However, there is a large variation on the slope of the panels and this is mainly due to the slope of the roof that they are mounted on.

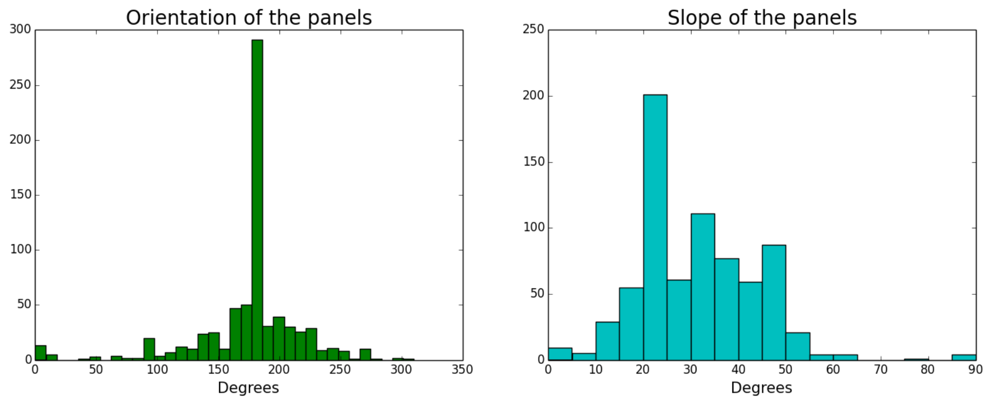

Figure 8: Orientation and slope of the panels

3.2 Performance Indicators for the Netherlands

The sample provided sufficient data to calculate the performance indicators for 2012 and 2013. The figure below presents the distribution of Y_f for 2012 for 160 sample installations with a mean value of $865kWh/kW_p$.

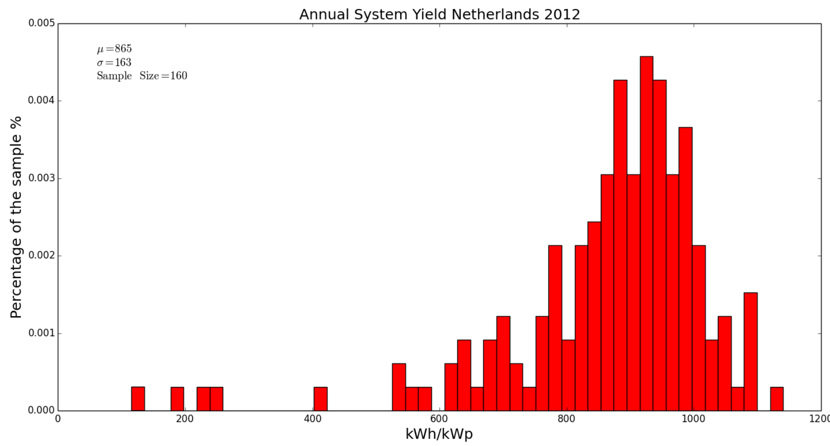

Figure 9: Distribution of Y_f for 2012

A relatively small but still considerable amount of PV systems that represents 5.6% of the sample appears to be malfunctioning and has annual yield less than $600kWh/kW_p$. On the other hand 12.5% of the installations perform exceptionally by producing more than $1000kWh/kW_p$ for the given year.

The above values were divided with the reference yield Y_r of each installation in order to acquire the PR. The average PV systems utilizes 78% of the incoming solar irradiation and the relatively low standard deviation reveals uniformity and a high concentration of all the values around the mean of the distribution. However, still 11% of the sample has PR values less than 55% and 15.6% was operating in the range $55\% - 70\%$. This percentage reveals weak points in the installation and operation of the aforementioned systems, and as only two of the systems were installed in the 70's, they are the only ones that could have been affected by the degradation of the panels. On the other hand 47% of the systems have PR in the range of $80\% - 95\%$.

According to the results, it is shown that despite the technological advances and the monitoring program, there was still a considerable amount of PV systems that was operating below the average regular standards.

The following year, the sample doubled in size as the result of the rapidly developing monitoring market in the Netherlands. The average annual yield is slightly bigger, reaching $876kWh/kW_p$ and only 3.2% appears serious malfunctions causing insufficient yield with less than $600kWh/kW_p$. In comparison with 2012 results, 7 out of 10 installations that were in this category have sig-

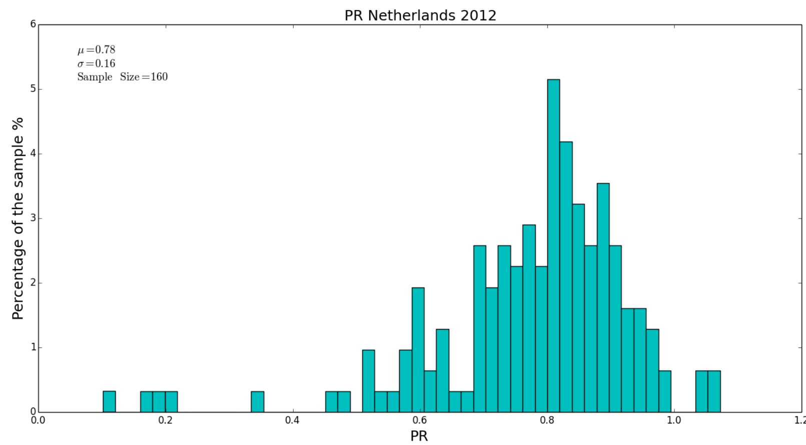

Figure 10: Distribution of PR for 2012

nificantly improved their performance in 2013, a sign that monitoring helps to detect a problem.

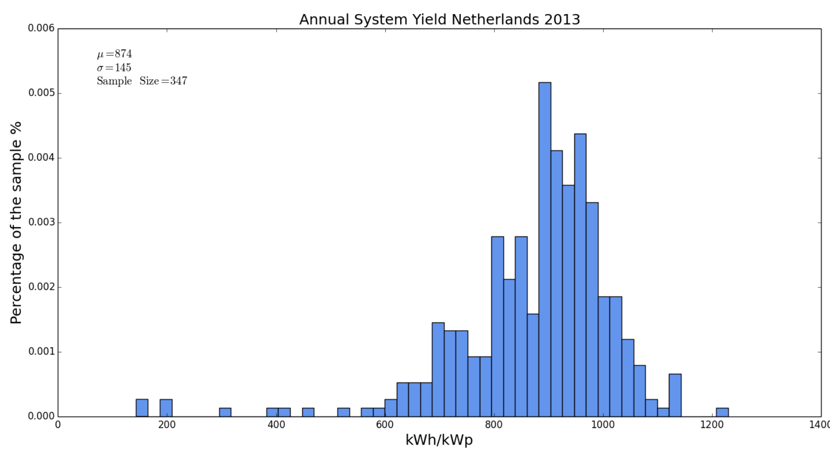

Figure 11: Distribution of Y_f for 2013

Despite the fact that the average PR remained unchanged, the monitoring program was beneficial for the majority of the installations that had inadequate performance the first year. The study reveals that systems that were operating with less than 70% in 2012, increased by 10% on average their performance indicators.

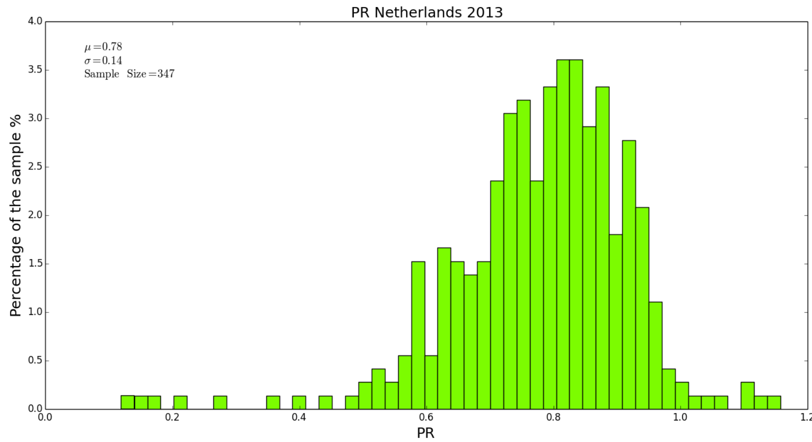

Figure 12: Distribution of PR for 2013

3.2.1 Geographical Variation

The solar irradiation in the Netherlands differs between the coastal part of the country and the mainland as it is shown on the PVGIS (Photovoltaic Geographical Information System) map below that presents the average irradiation for the years 2001-2008 [10]. On the left part of the figure is the geographical map, divided according to the postcode of each installation in order to adjust to the different irradiation zones that are highlighted on the right part of the figure. The postcodes of the coastal region with the higher solar activity, are represented with red colour and the postcodes from the mainland with green.

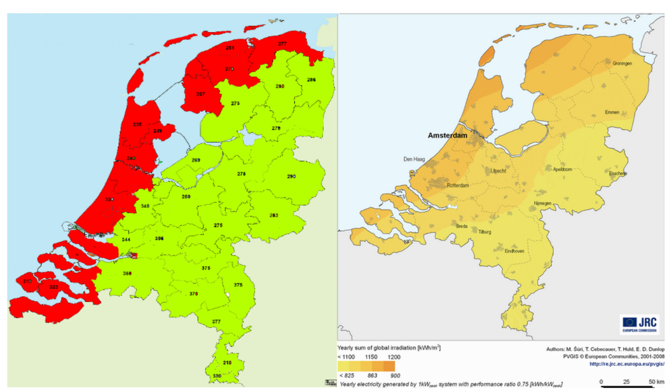

Figure 13: Geographical division of the Netherlands according to different solar irradiation zones

As listed in Table 2, modules that are installed on the coastal part of the Netherlands in 2012 produced 7% higher yields and were 3% more efficient in comparison with modules that are installed on the mainland.

Region	Performance Ratio	Annual Yield (kWh/kW_p)	Irradiation kWh/m^2
Coast	$79.3\% \pm 2\%$	895 ± 21	1165
Mainland	$77.1\% \pm 2\%$	833 ± 19	1021

Table 1: Geographical division according to solar irradiation

The following year, the total solar irradiation was approximately the same and therefore the difference in yield and PR was much smaller.

Region	Performance Ratio	Annual Yield (kWh/kW_p)	Irradiation kWh/m^2
Coast	$76.7\% \pm 2\%$	882 ± 21	1075
Mainland	$77.5\% \pm 2\%$	861 ± 19	1079

Table 2: Geographical division according to solar irradiation

3.2.2 Performance Analysis

The overall performance of a PV system depends on various factors and one of the most critical is the type of the module that is used on the system. The type of the panel is crucial not only for the performance of the system, but also for the total cost. As it was noted in section 3.1 the sample consists of installations that use four different type of panels.

Type of Panel	Number of Systems	Performance Ratio	Error
Polycristalline	236	80%	±0.1%
Monocrystalline	75	76.3%	±1%
Amorph	26	64.5%	±2%
Hybrid	4	69.4%	±3%

Table 3: PR values per type of panel module

The systems that use Polycrystalline and Moncrystalline modules have significantly higher performance (see Table 3) than the ones that use Amorph or Hybrid as it is shown form the table above and the ANOVA method that it was applied. However, statistically between the first two there is not significant difference and therefore in practice the difference in price is not depicted in the overall performance.

An other component that is crucial for the proper operation of a PV system is the inverter. Any malfunctions on this part may lead to inconsistency between to AC and DC current and therefore a considerably reduced performance. Three major manufacturers of inverters have the greatest market share in the Netherlands and the average PR of the systems that are using them is presented on the table below.

Type of Inverter	Number of Systems	Performance Ratio	Error
Mastervolt	83	73.6%	±1.1%
SMA	104	79%	±1%
Power One	83	83.5%	±1.2%

Table 4: PR values per inverter manufacturer

According to the ANOVA method that was applied, PV systems that are using SMA or Power One inverters have significantly better performance.

3.2.3 Seasonal Variation

The temperature dependence of the modules is affecting the overall performance of the PV system and as it shown in the graph below the average PR values are higher during the cold winter months than they are in the summer. As an example, during November the average PR Value was 82.1% but in July had dropped significantly to 73.1%.

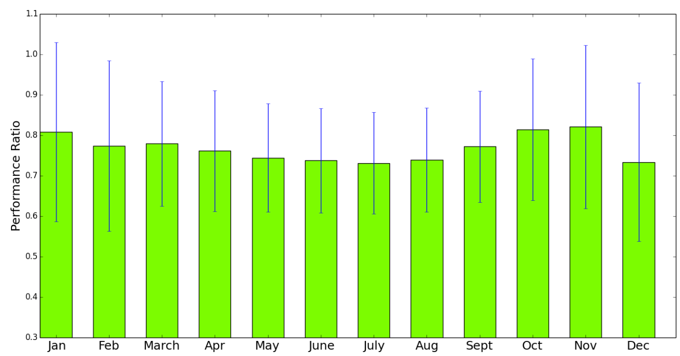

Figure 14: Monthly PR variation January-December 2012

The average daily temperature (°C) and the daily irradiation (kWh/m^2) are given in the graph below. By comparing the two graphs it is clear that the high solar activity of the summer months can not be sufficiently converted into electrical output due to the high temperatures.

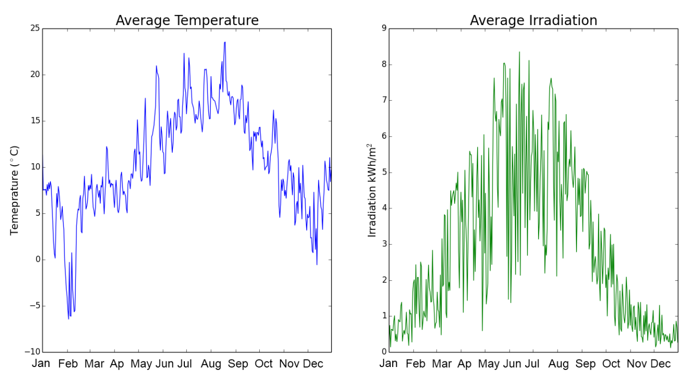

Figure 15: Annual temperature and solar irradiation 2012

The dependence of the performance on the temperature is also highlighted on the following scatter plot, where the performance of an individual system is plotted over the temperature of the surrounding environment. The PV system is $4.32kW_p$ and is located in Groningen, the average PR is 83% but it varies considerably as the temperature fluctuates. According to the regression line the maximum performance is reached when the temperature is below -5°C and it gradually falls to 65% when the temperature is more than 25°C. It is obvious that a rise of 30°C is able to reduce the overall performance up to 1/3.

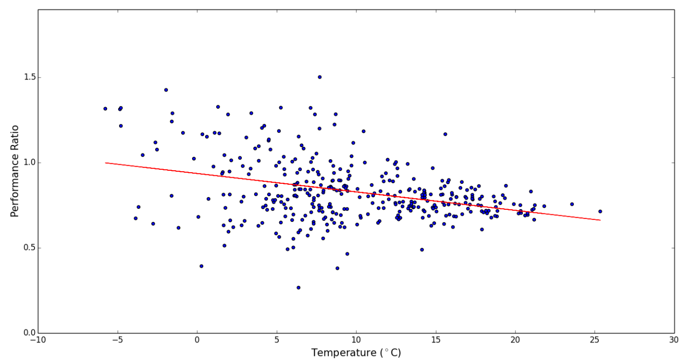

Figure 16: Performance Ratio as a function of ambient temperature

3.2.4 Inverter Efficiency and System Losses

PV systems in theory are not able to reach 100% PR as the final value aggregates all the possible energy losses and it is influenced by a number of factors such as inverter efficiency, wire losses, real power of the PV modules below nominal rating, mismatch, shades, dust, thermal and system failures [12]. The inverter efficiency of the sample was studied by comparing the AC and DC energy yield of each system and the results are presented on the graph below.

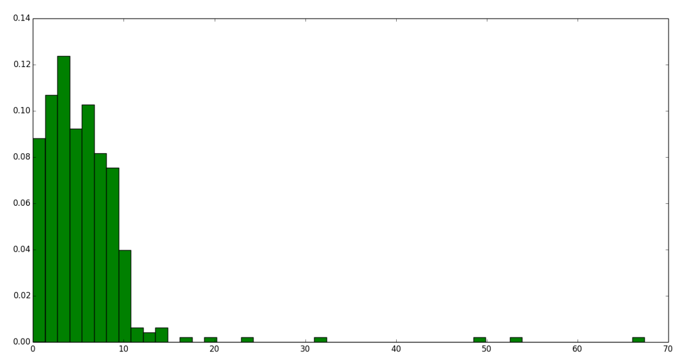

Figure 17: Average inverter losses of the sample

The average energy losses that occur during the transformation from DC to AC current are 5.7%. However, there is a number of systems that sustain losses from 10% up to 67%. That also explains the poor performance indicators that were observed for 15.6% and 18.1% of the installations for 2012 and 2013 respectively. During normal operation the AC and DC PR should have a constant difference that is explained from the inverter losses and is independent from any fluctuations of the system's performance. Even when the panel is exposed to various weather conditions and temperature fluctuations as the one that is depicted in Figure 18 the behavior of the two variables remains the same.

Figure 18: Systems AC and DC Performance Ratio for normal operation

On the other hand, wiring problems or inverter's malfunction may cause serious performance issues. In the following example for the period February to October 2013 the average energy losses were approximately 50% for the given PV system.

Figure 19: Systems AC and DC Performance Ratio for malfunctioning system

Any kind of similar system failure could easily be detected by monitoring the AC and DC energy yield. If the panel is functioning properly but there are great losses during the transmission or conversion of the energy there would be a large mismatch of the aforementioned values. However, technical malfunctions are not the only cause of low performance issues. Panel degradation, soiling and shading might be responsible as well. In the following example (Figure 20) despite the excellent conversion efficiency, the overall performance of the PV system is below 50% and since the conversion losses are below 8% one of the above reasons is likely to be responsible.

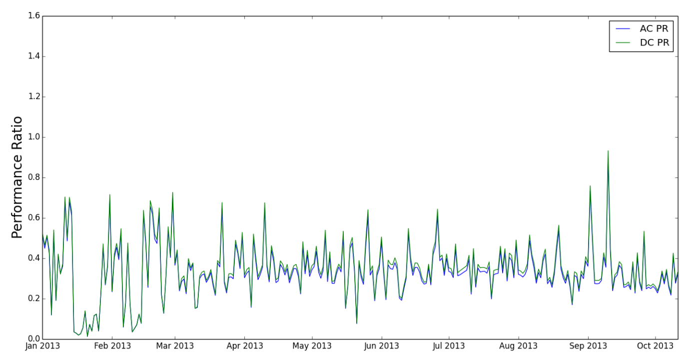

Figure 20: AC and DC Performance Ratio for low PR system

4 Comparison between European Countries

4.1 PV Systems Analysis

Despite the large number of data that was available for the Netherlands, the solar-log web platform was free to access for a number of other European countries as well. For this research data from Germany, Belgium, Italy and France was collected and analyzed. As it shown in the table below Germany has provided the largest number of installations and this is mainly due to the fact that the monitoring market is already well established there. All of the samples belong to mid-range PV systems with the one coming from Belgium being closest to the average domestic installation.

Country	Number of Samples	Average System Size (kW_p)
Netherlands	728	11.08
Germany	764	15.6
Italy	488	13.1
France	325	15.09
Belgium	388	6.52

Table 5: Sample size per country

The market share of each module type varies also per region, in Belgium and Germany the popularity of polycristalline and monocrystalline panels is balanced where in Netherlands and Italy the market is dominated by polycristalline modules. Only in the French sample the monocrystalline technology is more popular.

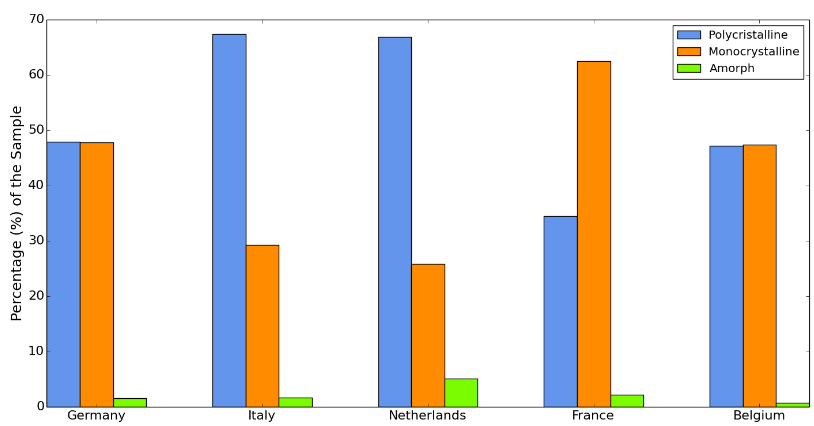

Figure 21: Silicon type market share percentage for each country

The Amorph silicon technology despite being the less favorable choice among PV owners, holds a great share among large PV installations. Especially in the Dutch and German Sample the average size of an Amorph system is more than $20kW_p$

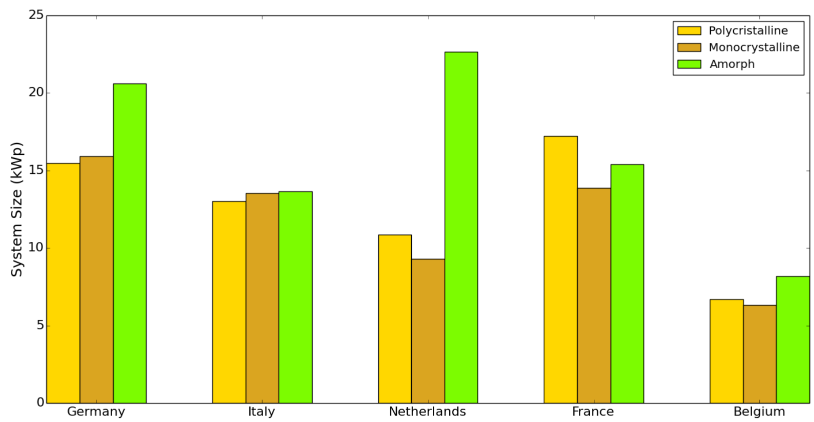

Figure 22: Average installation size per silicon type module

4.2 Performance Comparison

The following graph presents the average system yield of the given sample within a time frame of 3 years. As expected, Southern countries have achieved higher yields than the Northern ones. However, there is a descending tendency between the years 2011 and 2013 that ranges from 2% (the Dutch Sample) up to 11% (the German Sample). That could be explained either by the degradation that some of the panels have sustained or lower irradiation conditions.

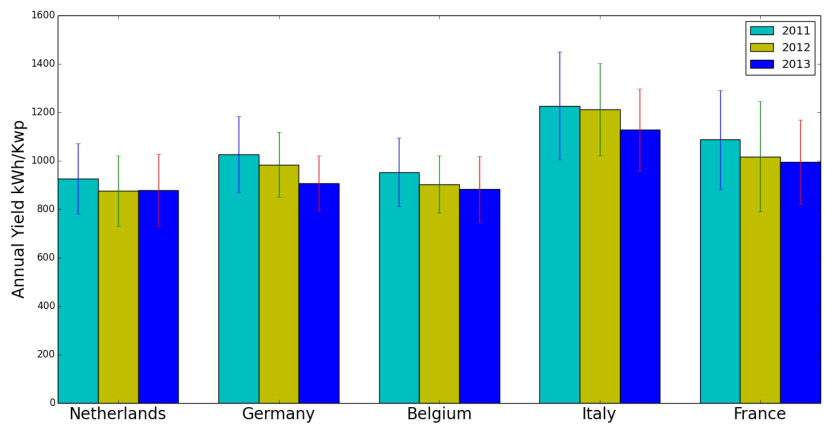

Figure 23: Annual Yield for the time period 201

The above figure does not represent the same number of systems over the years, the sample is enriched as there is a steady increase in the number of installations that are monitored over the years. By comparing Table 5 and Table 6, it is important to highlight that only a fraction of the total sample can provide reliable and continuous data for the whole period. The rest were either providing insufficient operational details or they were off line for some time and therefore they were not taken into consideration. A coherent conclusion could be that the monitoring industry it self has still further room for improvements.

Year	Country	Number of Samples	Annual Yield kWh/kW_p	Error $\pm$
2011	Netherlands	110	895	161
2012		147	865	145
2013		348	874	149
2011	Germany	335	1026	157
2012		335	983	134
2013		304	906	114
2011	Belgium	170	953	142
2012		191	903	118
2013		181	883	136
2011	France	23	1086	203
2012		84	1087	227
2013		91	995	174
2011	Italy	70	1226	223
2012		290	1211	170
2013		277	1128	170

Table 6: Mean Annual Yield per Country

For a more comprehensive analysis the above countries were separated in Northern and Southern parts according to the solar maps of each area. The separation was done according to irradiation measurements and not according to geographical criteria and therefore the areas are not equal in size.

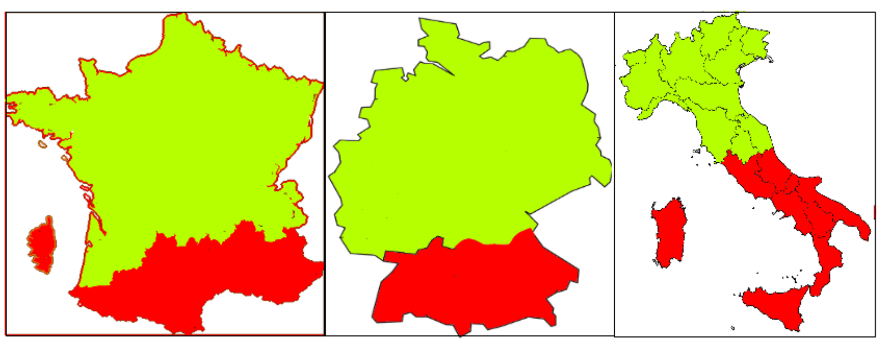

Figure 24: North-South separation for France, Germany and Italy

As it is depicted in the table below, the Southern part could reach 10-20% higher yields than the Northern one depending on the diversity of the solar activity and not by the size of the country. Moreover, countries with similar latitude boundaries like Germany and France achieve comparable level of energy production in comparison with Italy which is located Southern.

	Country	2011 kWh/kW_p	2012 kWh/kW_p	2013 kWh/kW_p
North	Germany	979 ± 153	937 ± 126	882 ± 109
South		1081 ± 154	1044 ± 121	922 ± 125
North	France	1030 ± 362	993 ± 201	959 ± 154
South		1099 ± 96	1092 ± 224	1103 ± 166
North	Italy	1219 ± 170	1177 ± 157	1094 ± 148
South		1352 ± 113	1337 ± 199	1288 ± 203

Table 7: Annual System Yield per Region

5 Conclusion

5.1 Conclusion

The sample that was chosen for this research was comparable in capacity with the average domestic installation in order to study the efficiency indicators for small scale systems. The results revealed a steady PR for Dutch systems with a value of 78% and the system yield being higher than 860 kWh/kW_p for the last 2 years. In this time period, the percentage of malfunctioning installations reduced significantly which is a sign that monitoring is beneficial for the proper system operation.

However, the efficiency and yield indicators appear to have a large seasonal and geographical variation. Even in a small country like the Netherlands, the coastal part receives approximately 14% higher solar irradiation than the mainland which is translated to 7% higher energy yields. Moreover, during the summer months the average PR drops from 82% to 75% due to higher temperature condition. Moreover, and by studying systems independently, it was revealed that they can reach the optimum operational efficiency when the ambient temperature is less than $-5\,^{\circ}\mathrm{C}$ and then it drops radically as the temperature rises.

The overall performance also depends on the individual components of each system, and as it is shown the type of the module and inverter that are used have significant impact on the final output.

On a bigger scale, the energy yield depends mainly on the geographical location of each country. Unfortunately comparison of the PR was not feasible for the current study as there were no high resolution irradiation data available. However, the analysis revealed a constant decrease of the total energy yield after 2011 that could either be explained by the degradation of the panels or by reduced solar activity. Furthermore, variations occur not only between different countries but also within different areas of the same country.

By comparing the results with previous studies there is a clear increase in PR values. The research that took place in Amersfoort in the time frame 2000-2005 concluded that the average PR was in the range of 65-70% [4], a value that is very close to a previous study (69.4%) that took place between 1992-2000 [3]. Moreover, the energy yield values for France and Belgium are approximately on the same level as in 2010 (1163 kWh/kW_p and 852 kWh/kW_p) according to a previous large scale study [12]

5.2 Recommendations for further research

For the conduction of this research a high resolution data base with solar yields was created. The quality and the quantity of the data gives the opportunity for further research in the field of performance assessment and error detection of PV installations.

Solar irradiation measurements of higher resolution can be used based on satellite images, in that way more precise calculations of key indicators will be achieved. Also, the spatial scope can be increased. A major hindering factor for this research was the availability of solar data outside of the Netherlands. By retrieving irradiation measurements for the rest of the European countries, PR could be calculated on an international scope.

Moreover, the growth of online monitoring applications can be used to en reach the existing data base. Similar programming scripts can be used to exploit the available data that are available to the public not only on SolarLog web page but also on other web platforms. Higher time resolution can also be achieved, in most cases 5 minute data are available. This kind of information can be used for further research on error and shading detection.

Finally, it is known that 'Big Data' can be used as input for artificial intelligence algorithms to create modeling and forecasting applications. PV yield measurements, meteorological data and solar irradiation could be combined to create a framework for solar energy forecasting and smart grid models.

5.3 Acknowledgements

The completion of this research was a difficult, demanding and challenging task. First of all I would like to thank my father E.Moraitis and my mother P.Styliari for their support to chase my dreams abroad and for their help during all these years. I would also like to thank my amazing friends Chris Fraggoulides, Stathis Grigoropoulos and Nick Tasios for their assistance and most of all their patience when they were dealing with my programming questions. They provided crucial help in critical parts of this research, and they helped me to develop high level programming skills. Of course, I would like to express the deepest appreciation to my supervisor Dr. Wilfried van Sark for his support but above all for the opportunity that he gave me to work in such a demanding project. Finally, I would like to thank my roommates and my friends in Utrecht for the great time we had.

Appendices

A Data Collection Script

```
import urllib.request

from bs4 import BeautifulSoup
import re
import time
import datetime
from datetime import timedelta


#adress without the last digit, th last digit is the indicator of each separate
#page with data

adress="http://home.solarlog-web.nl/sds/modul/SolarLogReference2/_sysreturn.php?
m=SolarLogReference2&s=8a0712a866935855021676cfa2b494ec&cid=5&action=result&mo=
result&cc=&pc=&mk=&of=&ot=&rf=&rt=&pf=&pt=&cy=&wp=0&wd=0&ws=0&cp="


#get the total amount of pages from the pointer on top of the page

Number_Pages=urllib.request.urlopen(str(adress)).read()
First_Soup=BeautifulSoup(Number_Pages)
Number_1=First_Soup.find('div', {'class':'footer'})
Number_List=re.findall(r'\d+', Number_1.text)

#identifies where each installation is located in the php code, first it saves the
#name and then seperatly each one of the operational details
for i in range(130,140):
    Names=[]
    Input_File=open("Data_Table%d.txt" %i, mode = 'w', encoding = 'utf8')
    print(i)
    currentadress=adress+str(i)
    webpage=urllib.request.urlopen(str(currentadress)).read()
    soup=BeautifulSoup(webpage)
    for Installations in soup.findAll('div', {'style':"display:inline;float:left;"}):
        for Names in Installations.find('td',{'style':"width:316px;background:url('/sds/modul/SolarLog
                            \Reference2/images/s_mitte_blauer_balken.png')
                            repeat-x;height:19px;overflow:hidden;padding:
                            1px 0px 0px 4px;color:#FFF;font-size:12px;"}):
            Input_File.write("\n")
            Installation_Name=Names.text
            Input_File.write('Name='+Installation_Name+'\n')
            for Location in Installations.findAll('div', {'title':re.compile('Locatie.')}):
                Installation_Location=Location.text
                Input_File.write('Location='+Installation_Location+'\n')
            for Size in Installations.findAll('div', {'title':re.compile('Installatievermogen:')}):
                Installation_Size=Size.text
                Input_File.write('System Size='+Installation_Size+'\n')
            for Inverter in Installations.findAll('div', {'title':re.compile('Omvo')}):
                Installation_Inverter=Inverter.text
                Input_File.write('Inverter='+Installation_Inverter+'\n')
            for Panel in Installations.findAll('div', {'title':re.compile('Modules')}):
                Installation_Panel=Panel.text
                Input_File.write('Panel='+Installation_Panel+'\n')
            for Orientation in Installations.findAll('div', {'title':re.compile('Uitri')}):
                Installation_Orientation=Orientation.text
                Input_File.write('Orientation='+Installation_Orientation+'\n')
            for Slope in Installations.findAll('div', {'title':re.compile('Dakhel')}):
                Installation_Slope=Slope.text
                Input_File.write('Slope='+Installation_Slope+'\n')
            for Date in Installations.findAll('div', {'title':re.compile('Bouwjaar')}):
                Installation_Date=Date.text
```

```
                Input_File.write('Installation Date='+Installation_Date+'\n')
#finally it locates the link for the individual page of each installation
            for link in Installations.findAll('a',{'style':"text-decoration:none;color:black;cursor:
                                            pointer;line-height:20px;vertical-align:top;"}):
                pvlink=(link['href'])
                print(pvlink,'\n')
                Data_PV_link=re.findall(r'\d+', pvlink)
#the full link of each installation includes further operational details tha are also located and
#stored
                Full_Link="http://home.solarlog-web.nl/"+str(int(Data_PV_link[-1])-1)+".html"
                Foul_Link_URL=urllib.request.urlopen(Full_Link)
                Soup_DataLink=BeautifulSoup(Foul_Link_URL)
                for Two_Headings in Soup_DataLink.findAll('td', {'width':'663'}, {'class':'headline'}):
                    Module_Type=Two_Headings.next_element
                    for i in range(19):
                        Module_Type=Module_Type.next_element
                    if Module_Type is not None:
                        try:
                            Input_File.write(Two_Headings.text+u"Count="+Module_Type.text+u'\n')
                            Input_File.write(Two_Headings.text+u"Type="+Module_Type.next_sibling.

                        except AttributeError:
                            print("No available Data")
#It navigates through different links and pages in order to detect the yield data page
#of each installation
                pv_linkslocation_2=urllib.request.urlopen(pvlink)
                soup=BeautifulSoup(pv_linkslocation_2)
                statistik=soup.findAll('script')
                next_adress=re.search(r'(var\ssvgsrc\s\W\s\W)(http|ftp|https):\/\/([\w\-_]+(?:(?:\.
                                                                [\w\-_]+)+))([\w\-\.,@?^=%&amp;:/~\+#]*
                                                                [\w\-\@?^=%&amp;/~\+#])?', str(statistik))
                help_me=next_adress.group()
                last_step=re.search(r'(http|ftp|https):\/\/([\w\-_]+(?:(?:\.[\w\-_]+)+))([\w\-\.,@?
                                                                ^=%&amp;:/~\+#]*[\w\-\@?^=%&amp;/~\+#])?'
                                                                , str(help_me))
                Move_On=last_step.group()+"&flag=33"
                Go_Back=re.search(r'(http|ftp|https):\/\/([\w\-_]+(?:(?:\.[\w\-_]+)+))([\w\-\.,@?^
                                                                =%&amp;:/~\+#]*[\w\-\@?^=%&amp;/~\+#])
                                                                ?offset=', str(Move_On))
                Go_Back_2=Go_Back.group()
                Go_Back_3=re.search(r'&r=(\w*)&flag=(\w*)', str(Move_On))
                Go_Back_4=Go_Back_3.group()
                Number_Of_Zeros=0
#It repeats the following loop for 1100 times (3 years) to get data, in case of invalid or no entry the
#Number_Of_Zeros is increased by one, up to 60(two months). At that point the prgram moves to the
#next installation
                for i in range (0,1100,-1):
                    All_the_links=Go_Back_2+str(i)+Go_Back_4
                    One_More_Step=urllib.request.urlopen(All_the_links)
                    One_More_Step_2=BeautifulSoup(One_More_Step)

                    for Find_AC in One_More_Step_2.findAll('g', {'transform':'translate(0 5)'}):
                        for Guess_what in Find_AC.findAll('text', {'text-anchor':'start'} ):
                            AC_Value=Guess_what.next_sibling
                            if AC_Value is not None:
                                try:
                                    DC_Value=AC_Value.next_sibling
                                    Input_File.write(str(datetime.date.today()+timedelta(days=i))+' '+
                                   ( Guess_what.text)+u' '+( AC_Value.text)+' '+
                                   (DC_Value.text )+u'\n')
                                except AttributeError:
                                    print(i)

                            else:
                               Number_Of_Zeros=+1
                               if Number_Of_Zeros==60:
                                   break
                Input_File.write('\n')
```

```
        Input_File.write('\n')
        Input_File.write('\n')
        print('\n')

Input_File.close()
```

B Data Base Script

```
#It reads all the .txt files with the data yields and creates two csv tables
#the first one is a look up table with all the opperational details of each installation
#it also links each installation with the closest meteorogical station
#the second one has the daily yields.

import re
from pandas import *
import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
from matplotlib.ticker import FuncFormatter
import datetime
import itertools
import collections
import random


Details_Table2=DataFrame()
Yield_Table2=DataFrame()
for i in range(120):
    print(i)
    Yield_Array=[]
    Yield_Table=[]
    PV_Name=[]
    Station_N=[]
    PV_NameList=[]
    PV_PostcodeList=[]
    Capacity_List=[]
    Dates_List=[]
    Inverter_Man=[]
    Panel_Man=[]
    Orientation_List=[]
    Slopes_List=[]
    Inverter_Numb=[]
    Inverter_Model=[]
    Module_Numb=[]
    Module_Type=[]
    PV_NameList2=[]
    Details_Table=[]
    Yield_Table=[]
    Silicon_Type=[]
    Yield_Array_DC=[]
    Installation_Date=[]

    Open_Yield_File=open(r'Data_Table%d.txt' %i,'r')
    text_line=Open_Yield_File.readline()

    while text_line!='':
            #the Post_Codes.txt is a file that includes the meteorological station locations
        if text_line[:4]=='Name':
            Name=text_line[5:10]
            print(Name)
            PV_Name.append(Name)
            for i in range(1):
                text_line=Open_Yield_File.readline()
                postcode=text_line[14:20]
                Open_PostCodes=open(r'Post_Codes.txt','r',encoding='utf-8')
                text_postcode=Open_PostCodes.readline()
                while text_postcode!='':
                    if postcode[:2] in text_postcode[5:150]:
                        station_number=text_postcode[2:6]
                        Station_N.append(station_number)
```

```
            text_postcode=Open_PostCodes.readline()
#step by step gets the installation info
    for i in range(1):
         text_line=Open_Yield_File.readline()
         Capacity=re.findall(r'\d+.\d+|\d+', text_line[14:22])
    for i in range(1):
         text_line=Open_Yield_File.readline()
         Inverter_M=re.search(r'.*', text_line[9:25])
    for i in range(1):
         text_line=Open_Yield_File.readline()
         Panel_M=re.search(r'.*',text_line[6:39])
    for i in range(1):
         text_line=Open_Yield_File.readline()
         orientation=re.findall(r'\d+', text_line[12:16])
         if orientation==[]:
             orientation=[0]

    for i in range(1):
         text_line=Open_Yield_File.readline()
         slopes=re.findall(r'\d+', text_line[6:15])
         if slopes==[]:
             slopes=[0]
    for i in range(1):
         text_line=Open_Yield_File.readline()
         installation_D=re.findall(r'\d+', text_line[18:22])
    for i in range(1):
         text_line=Open_Yield_File.readline()
         if text_line[:2]=='Om':
             inverternumb=re.findall(r'\d+', text_line[14:16])
         else:
             inverternumb=[0]
             break
    for i in range(1):
         text_line=Open_Yield_File.readline()
         if text_line[:2]=='Om':
             invertermodel=re.search(r'.*', text_line[13:40])
         else:
             invertermodel=re.search(r'.*','Unknown')
             break
    for i in range(1):
         text_line=Open_Yield_File.readline()
         if text_line[:2]=='Mo':
             modulenumber=text_line[13:15]
             if modulenumber!=int:
                 modulenumber=0
         else:
             modulenumber=0
             break
    for i in range(1):
             text_line=Open_Yield_File.readline()
             if text_line[:2]=='Mo':
                 moduletype=re.search(r'.*', text_line[12:45])
                 silicon=re.search('\((.*?)\)',str(moduletype.group())).group(1)
             else:
                 moduletype=re.search(r'.*','Unknown')
                 silicon=['Unknown']
                 break



    text_line=Open_Yield_File.readline()
    Orientation_List.append(int(orientation[-1]))
    Panel_Man.append(Panel_M.group())
    Inverter_Man.append((Inverter_M.group()))
    PV_NameList2.append(Name+postcode+(Capacity[-1])+Inverter_M.group())
    Capacity_List.append(float(Capacity[-1]))
    Slopes_List.append(int(slopes[-1]))
```

```
            Inverter_Numb.append(int(inverternumb[-1]))
            Inverter_Model.append(invertermodel.group())
            Module_Numb.append(int(modulenumber))
            Module_Type.append(moduletype.group())
            Silicon_Type.append(silicon)
            PV_PostcodeList.append(station_number)
            Installation_Date.append(int(installation_D[-1]))
                #the Details_Table includes all the opperational details
            Details_Table=DataFrame({'Installation_Date':Installation_Date,'Location':PV_PostcodeList,
                                    'Capacity_kWp':Capacity_List, 'Inverter_Man':Inverter_Man,
                                    'Panel_Man':Panel_Man, 'Orientation':Orientation_List,
                                    'Slope':Slopes_List, 'Inverter_Numb':Inverter_Numb, '
                                    Inverter_Model':Inverter_Model, 'Module_Numb':Module_Numb,
                                    'Module_Type':Module_Type,'Silicon_Type':Silicon_Type},
                                    index=PV_NameList2)

#getting AC and DC yield data. also sepparates Wh,kWh and MWh
            while text_line[:2]!='':
                if text_line[:2]=='20':
                    string_line=text_line.split()
                    Dates=string_line[0]
                    Test_Numb_2=float(string_line[2])
                    Indicator_AC=str(string_line[3])
                    if Indicator_AC=='Wh':
                        Yield_Array.append((Test_Numb_2*0.001))
                    if Indicator_AC=='MWh':
                        Yield_Array.append((Test_Numb_2*1000))
                    if Indicator_AC=='kWh':
                        Yield_Array.append(Test_Numb_2)
                    Test_Numb_2_DC=float(string_line[4])
                    Indicator_DC=str(string_line[5])
                    if Indicator_DC=='Wh':
                        Yield_Array_DC.append((Test_Numb_2_DC*0.001))
                    if Indicator_DC=='MWh':
                        Yield_Array_DC.append((Test_Numb_2_DC*1000))
                    if Indicator_DC=='kWh':
                        Yield_Array_DC.append(Test_Numb_2_DC)
                    Dates_List.append(datetime.datetime.strptime(Dates,'%Y-%m-%d'))
                    PV_NameList.append(Name+postcode+(Capacity[-1])+Inverter_M.group())
                else:
                    break
                text_line=Open_Yield_File.readline()
                Yield_Table=DataFrame({'Name':PV_NameList,'AC_Yield_kWh':Yield_Array,'DC_Yield_kWh':
                                                Yield_Array_DC}, index=Dates_List)



        text_line=Open_Yield_File.readline()

    Details_Table2=Details_Table2.append(Details_Table)
    Yield_Table2=Yield_Table2.append(Yield_Table)

#saves the csv files to the hard drive
Details_Table2.to_csv('Details_Table_NL',index_label='PV_NameList2')
Yield_Table2.to_csv('Yield_Table_NL',index_label='Dates_List')
```

C Acquiring Meteorological Information

```python
import re
from pandas import *
import numpy as np
import pandas as pd
import matplotlib
import datetime
import itertools
import math

#importing the file from KNMI
Input_File=open(r'KNMI_20140114_hourly.txt', 'r')

for i in range(58):
    line=Input_File.readline()

#skipping the first 58 lines that include additional info
#scraping the rest of the txt for irradiation,wind speed,precipitaion
#and temperature data devided by station.

Station=[]
Date=[]
Hour=[]
Irradiation=[]
Wind_Speed=[]
Temperature=[]
Precipitation=[]
Date_Year=[]
Hour_Daily=[]


line=Input_File.readline()

while line !='':
    station_num=line[2:5]
    time_date=line[6:14]
    time_hour=line[18:20]
    wind_speed=line[24:26]
    wind_speed=wind_speed.replace(' ','0')
    temperature=line[35:38]
    time_hour = time_hour.replace(' ','0')
    if time_hour=='24':
        time_hour='00'
    time_q=line[42:44]
    precipitation=line[48:50]
    Date.append(datetime.datetime.combine(datetime.datetime.strptime((time_date),'%Y%m%d').
             date(),datetime.datetime.strptime(time_hour, '%H').time()))
    Date_Year.append((datetime.datetime.strptime((time_date),'%Y%m%d')).strftime('%j'))
    Hour_Daily.append((datetime.datetime.strptime(time_hour, '%H')).strftime('%H'))
    Irradiation.append((time_q))
    Station.append(int(station_num))
    Wind_Speed.append(int(wind_speed))
    Temperature.append(int(temperature)*0.1)
    Precipitation.append(int(precipitation))
    line=Input_File.readline()


My_Table=DataFrame({'Station':Station, 'Irradiance':Irradiation,'Wind Speed':Wind_Speed, 'Temperature'
:Temperature,'Precipitation':Precipitation, 'Date':Date_Year, 'Hour':Hour_Daily }, index=Date)
My_Table2=My_Table.groupby('Station')
```

D Applying Olmo Model

```
import re
from pandas import *
import numpy as np
import pandas as pd
import matplotlib
import datetime
import itertools
import math
import pandas as pd


#calculating the declination for a whole year
Delta_Table=[]
for i in range(367):
    Delta_Value=23.45*math.sin(math.radians(0.9863*(284+i)))
    Delta_Table.append(Delta_Value)

#importing the irradiation and the yield csv
My_Table=pd.read_csv('Irradiation_2013', index_col='Date', parse_dates=True)
My_Table2=My_Table.groupby('Station')

Details_Table2=DataFrame()
Yield_Table2=DataFrame()





Yield_Table2=pd.read_csv('Yield_Table_NL', index_col='Dates_List', parse_dates=True)
Yield_Table3=Yield_Table2.groupby('Name')
Details_Table2=read_csv('Details_Table', index_col='PV_NameList2')


Names_List=[]


for i in Details_Table2.index:
    if i not in Names_List:
        Names_List.append(i)

Table2=DataFrame()
Final_Table=DataFrame()
#starts the procedure for each installation in the Names_List
for i in Names_List:
    try:
        #making correction for the orientation
        bita=Details_Table2.ix[i].get_value('Slope')
        angle=Details_Table2.ix[i].get_value('Orientation')
        if angle>180:
            new_angle=angle-180
        if angle<180:
            new_angle=-(180-angle)
        if angle==0:
            new_angle=0
        if angle==180:
            new_angle=0
#calculating the A,B,C,D,E factors for equation (11)
        Alfa=math.sin(math.radians(52))*math.cos(math.radians(bita))
        Bi=math.cos(math.radians(52))*math.sin(math.radians(bita))*math.cos(math.radians(gama))
        Ci=math.sin(math.radians(bita))*math.sin(math.radians(gama))
        Delta=math.cos(math.radians(52))*math.cos(math.radians(bita))
        Epsilon=math.sin(math.radians(52))*math.sin(math.radians(bita))*math.cos(math.radians(gama))
#finding the station according to the post code and then adds in Total table all the days of the year
        Code=int(Details_Table2.ix[i].get_value('Location'))
```

```
        Dates_Group=My_Table2.get_group(Code).groupby('Date_Year')
        Total=[]
        for j in My_Table2.get_group(Code).Date_Year:
            if j not in Total:
                Total.append(j)
        Table1=DataFrame()
#for every day of the year calculates the Irradiation on the inclined surface according to Olmo model
#calculates for every single hour according to formulas in 2.2.1
        for m in Total:
            Single_Days=Dates_Group.get_group(m)
            Irradiation_List=[]
            for k in Single_Days.index:
                Hour=int(Single_Days.ix[k]['Hour'])
                delta=Delta_Table[int(m)]
                solar_time=15*(Hour-12)+10
                Cos_thita=(Alfa-Bi)*math.sin(math.radians(delta))+(Ci*math.sin(math.radians(solar_time))
                                                                +(Delta+Epsilon)*math.cos(math.radians
                                                                (solar_time)))*math.cos(math.radians(delta))
                Cos_thita_Z=math.sin(math.radians(52))*math.sin(math.radians(delta))+math.cos(math.radians
                                                                (52))*math.cos(math.radians(solar_time))*
                                                                math.cos(math.radians(delta))
                thita=math.acos(Cos_thita)
                thita_z=math.acos(Cos_thita_Z)
                Irradiation=int(Single_Days.ix[k]['Irradiance'])
                Fc=1+0.25*(math.sin(thita/2))*(math.sin(thita/2))
                Clearness_Index=(Irradiation*2.77777)/1367
                Y_zero=math.exp((-Clearness_Index)*((thita**2)-(thita_z**2)))
                New_Irradiation=Y_zero*Irradiation*Fc
                Irradiation_List.append(New_Irradiation)
#for every hour imports the calculated values in to Irradiation_List and then appends it to the
#Irradiation_Table which is then combined with the original file with the global irradiation
#measurements.All the values are recalculated on a daily basis
            Irradiation_Table=DataFrame({'New_Irradiation':Irradiation_List} ,index=Single_Days.index)
            Table1=Table1.append(Irradiation_Table)
        Join_Table=My_Table2.get_group(Code).join(Table1, how='inner')
        Join_Table2=Join_Table.resample('D', how={'New_Irradiation':sum, 'Wind Speed':np.mean,
                                    'Temperature':np.mean})
#The table with the meteorological and irradiation data is then combined with the yield table for the
#calculation of the PR valu(daily PR)
        TestTable=Join_Table2.join(Yield_Table3.get_group(i), how='inner')
        TestTable['PR']=TestTable.AC_Yield_kWh/((TestTable.New_Irradiation*0.002777)*
                                    Details_Table2.ix[i]['Capacity_kWp'])
        TestTable['Location']=Details_Table2.ix[i]['Location']
        TestTable['Name']=i
        Final_Table=Final_Table.append(TestTable)
        print(Details_Table2.ix[i]['Location'],np.mean(TestTable['PR']))
        Table2=Table2.append(TestTable)
        Table2.to_csv('PR_Table_NL_2013',index_label='Date')
    except KeyError:
        print('Key Error in Number', i)
    except TypeError:
        print('Type Error in Number', i)
    except ValueError:
        print('Value Error',Cos_thita )
```

References

[1] Drews, A., De Keizer, A.C., Beyer, H.G., Lorenz, E., Betcke, J., Van Sark, W.G.J.H.M., Heydenreich, W., Wiemken, E., Stettler, S., Toggweiler, P., Bofinger, S., Schneider, M., Heilscher, G., Heinemann, D, 2007. *Monitoring and remote failure detection of grid-connected PV systems based on satellite observations*, Solar Energy81 (2007) 548-564.

[2] W.G.J.H.M. van Sark, 2012. *Review of Performance Ratio development.*Copernicus Institute.

[3] Jahn, U., Nasse, W., 2004. *Operational performance of Grid-Connected PV systems on Buildings in Germany*, Progress in Photovoltaics: Research and Applications 12 441-448.

[4] DeKeizer, A.C., Ter Horst, E., Molenbroek, E.C, Van Sark, W.G.J.H.M., 2007. *Evaluating 5-years performance monitoring of 1 MW building integrated PV project in Nieuwland, Amersfoort, the Netherlands*, Proceedings of the 22nd European Photovoltaic Solar Energy Conference, Milano, 2007, 2960-2965

[5] W.G.J.H.M. van Sark, 2011. *Performance ratio revisited: is $PR > 90\%$ realistic?.* PROGRESS IN PHOTOVOLTAICS: RESEARCH AND APPLICATIONS.

[6] Kristen Ardani and Robert Margolis, 2011. *Solar Technologies Market Report,*US Department of Energy

[7] US Department of Energy, 2011. *Electric Power Annual 2009*

[8] Behnke, Erdman, and Whitaker Engineering, Inc., Commonwealth Energy Biogas/PV Mini-Grid, Renewable Resources Program, 2005. *Interim Exposure and Operations Report for the Comprehensive Large PV System Comparison*, California Energy Commission Public Interest Energy Research Renewable Program

[9] Tom Markvart, 2003. *Practical Handbook of Photovoltaics: Fundamentals and Applications*, Elsevier Ltd.

[10] IEC 61274, 1998.*Photovoltaic system performance monitoring - Guidelines for measurement, data exchange and analysis.*

[11] B. Marion, 2005. *Performance Parameters for Grid-Connected PV Systems* . US Department of Energy.

[12] Jonathan Leloux, Luis Narvarte, David Trebosc, 2011. *Performance Analysis of 10,000 Residential PV Systems in France and Belgium.*

[13] Leloux, J., Narvarte, L., Trebosec, D., 2012. *Review of the performance of residential PV systems in France*, Renewable and Sustainable Energy Reviews16 1369-1376

[14] Colienne Demain, Michel Journe, Cadric Bertrand, 2013. *Evaluation of different models to estimate the global solar radiation on inclined surfaces.* Renewable Energy 50 (2013) 710-721

[15] Cedric Brehaut, 2013. *Global PV Monitoring: Technologies, Markets and Leading Players, 2013-2017.* GTM Research and SoliChamba Consulting.

[16] Wes McKinney, 2012. *Python for Data Analysis.* OâĂŹReilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

[17] Colienne Demain, Michel JournÃľe, CÃľdric Bertrand, 2013. *Evaluation of different models to estimate the global solar radiation on inclined surfaces.* Renewable Energy 50 (2013) 710-721.

[18] M. Muzathik, M.Z. Ibrahim, K.B. Samo, W.B. Wan Nik, 2011. Estimation of global solar irradiation on horizontal and inclined surfaces based on the horizontal measurements. Energy 36 (2011) 812-818.

[19] F.J. Olmo, J. Vida, I. Foyo, Y. Castro-Diez, L. Alados-Arboledas,1998. *Prediction of global irradiance on inclined surfaces.* Energy 24 (1999) 689-704.

[20] John Twidell, Tony Weir, 1986. *Renewable Energy Resources.* Taylor and Francis Group