Master's Thesis – master Energy Science

Investigating the variability of the Utrecht Science Park PV System

Lucas Bahnmüller
l.c.bahnmuller@students.uu.nl

Copernicus Institute of Sustainable Development, Utrecht University
Supervision: Prof. dr. W. van Sark
W.G.J.H.M.vanSark@uu.nl

Abstract

The increasing amount of installed PV capacity in the Netherlands is increasing the burden on the electrical grid. The intermittency, or output variability of PV systems is the reason why PV systems can create this large burden on the grid if the PV capacity keeps increasing. In this paper, the output variability of the different PV systems of the Utrecht Science Park was investigated, where exclusively the variability caused by the motion of the clouds was analyzed. Performance ratios were calculated for the different PV systems to see how these systems performed at different levels of output power. The theoretical output of the different PV systems was calculated with a transposition model to be able to determine their output variability. With this output variability, the correlation between the different pairs of PV systems was calculated, after which it could be concluded that the different systems strongly correlate which each other, with significant differences being present between the different pairs of PV systems. However, the correlations were not as strong as found in previous studies about the performance of the Utrecht Science Park PV systems. This is important to take into account for further studies, which could for example go into the forecasting of the output of the different PV systems.

Table of Contents

1 Introduction

The amount of installed PV (photovoltaic) capacity is increasing exponentially in the Netherlands, as prices are declining, and Dutch climate goals must be reached. CBS estimated that a total of 2352 MW PV capacity was added in 2019, increasing the total installed PV capacity in the Netherlands to 6874 MW [1].

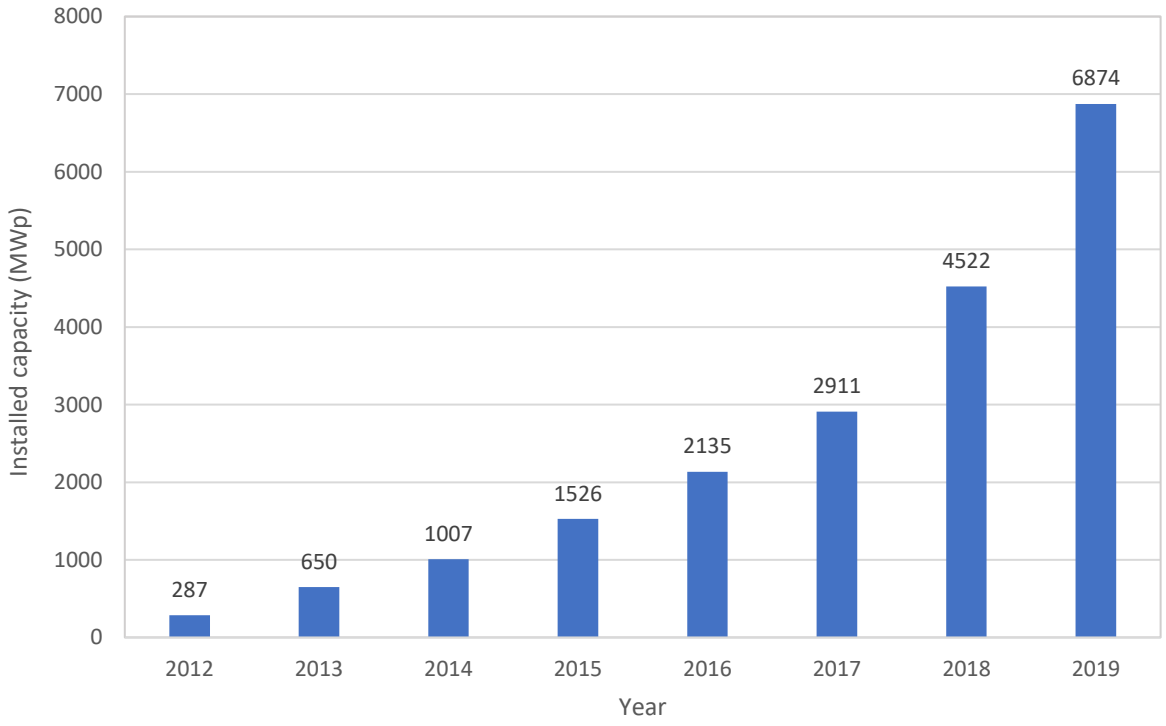

Figure 1: Installed PV capacity in the Netherlands since 2012, with 2019 being an estimation [1].

Most of this PV capacity, namely 79.1%, consists of PV systems with a peak capacity of less than 10 KWp, mostly being rooftop PV systems [2]. Primarily, the goal of these systems is to reduce the electricity bill of a household/company. These households/companies use the electricity that the PV system generates during the day, which is called self-consumption. The surplus energy that is generated throughout the day is delivered to the electrical grid for a certain price (presently identical to the retail price), generating an economic benefit. Self-consumption is ideal to decrease the burden on the electrical grid, as the fast increase in PV capacity is increasing the burden on the electrical grid [3].

The intermittency, or variability, of solar energy is the reason why PV systems may create a large burden on the grid if they are installed on every roof. This variability has two causes; the first cause is precisely predictable, as it depends on earth's distance from the sun and the apparent motion of the sun in the sky. The second cause is much less predictable, as it depends on the motion of clouds and weather systems. Being able to more accurately predict this second cause of variability could help to improve the stability of the grid, especially with the increasing amount of PV capacity [4].

1.1 Utrecht Science Park PV system

The Utrecht Science Park PV system is divided into eight systems. Utrecht University installed 4,600 PV modules on eight different buildings, as in Figure 2, in the Utrecht Science Park in 2016, working towards its goal to be CO_2 neutral by 2030. As the energy consumption of Utrecht University is responsible for half its carbon footprint, supplying its energy demand has a large potential in reducing this carbon footprint. Each module has a rated capacity of 270 Wp, resulting in a total system capacity of 1.2 MWp. This PV system is estimated to produce a total of 1 million kWh of electricity annually, which is around 2% of the annual electricity demand of Utrecht University [5].

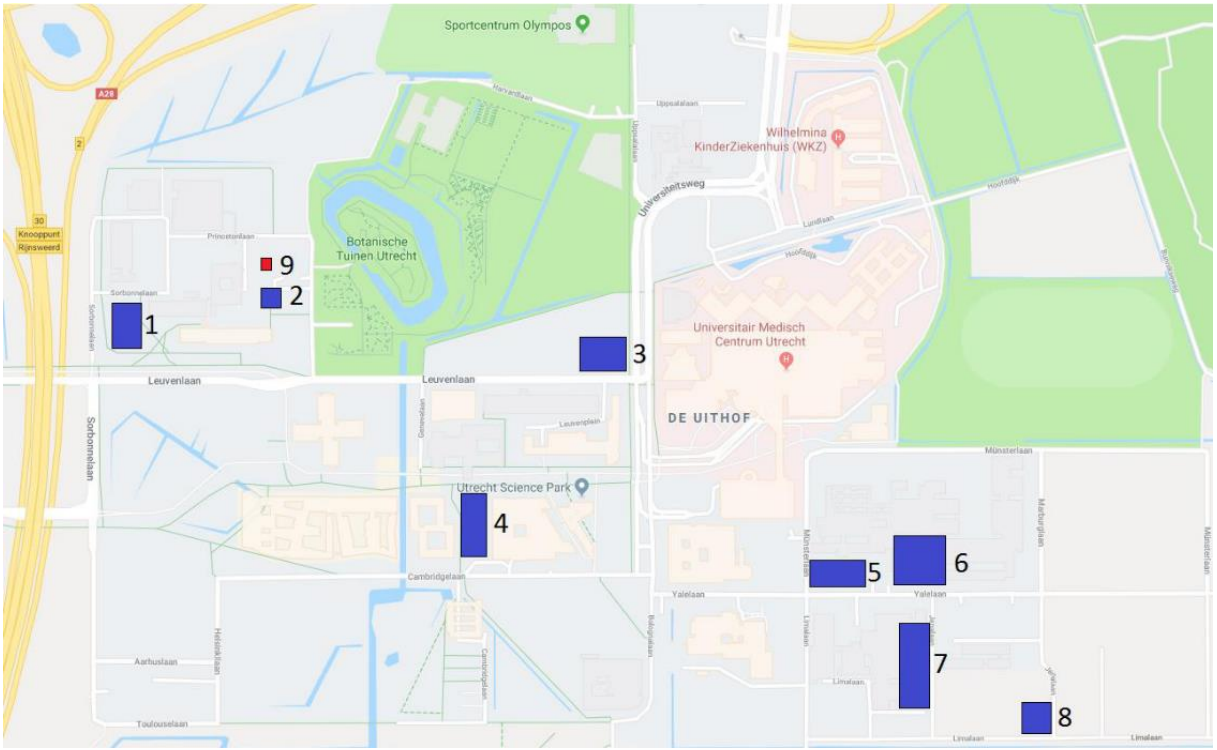

Figure 2: The location of each PV system on Utrecht Science Park (blue). 1: Caroline Bleeker building (CB), 2: Victor J. Koningsberger building (VJK), 3: David de Wied building (DdW), 4: University library (UB), 5: Jeanette Donker-Voet building (JDV), 6: Willem C. Schimmel building (WCS), 7: Martinus G. de Bruin building (MdB), 8: Waterberging diergeneeskunde/Tolakker (WGB), 9: Utrecht Photovoltaic Outdoor Test facility (UPOT) [5].

The parameters of the eight different systems are shown in Table 1.

Table 1: Parameters of the Utrecht Science Park PV system per building [5].

System	**Installed capacity (kWp)**	**Inverter capacity (kW)**	**Orientation (capacity per orientation(kWp))**	**Surface tilt (degrees)**
CB	132.30	107	W (72.37), E (59.93)	10
VJK	33.39	30	S (33.39)	10
UB	232.20	197	N (116.1), S (116.1)	10
DdW	123.12	99	W (49.68), S (23.76), E (49.68)	10
JDV	46.44	34	W (23.22), E (23.22)	10
WCS	234.36	200	W (117.18), E (117.18)	10
MdB	285.66	230	W (142.83), E (142.83)	10
WGB	131.22	107	S (131.22)	30

The influence of the different orientations can be considered small, as the tilt is only 10 degrees on all buildings, except for WGB [6].

1.2 Recent studies

Different studies have already been performed on the performance ratio of the Utrecht Science Park PV system. Van Sark et al. [6] described the performance of the PV systems, while also explaining the specific design and layouts of the system. The different buildings, as in Figure 2, were analyzed individually. Bruinewoud [5] also described the performance of the PV systems, but additionally investigated the rate of self-consumption of the systems and the correlation between each pair. These studies found different results for the performance ratio of the PV systems, caused by Bruinewoud [5] using production data with no gaps as opposed to the data that was used by van Sark et al. [6]. Bruinewoud (2018) found that all systems expect for one had a self-consumption rate of 100%, which was expected as the energy demand of the buildings were higher than the capacity of the PV systems. Finally, Bruinewoud [5] states that the different PV systems correlate strongly with each other.

There are no recent studies regarding the output variability of the Utrecht Science Park PV system, however, Hoff & Perez [7] have proposed a method to quantify the PV power output variability in general. This method revolves around relative output variability, which is the ratio between output variability of a PV system and the output variability of that PV system if it was concentrated in one single location. This relative output variability quantifies the noise reduction that is associated with the dispersion of the PV system.

Jamaly & Kleissl [8] have done research into the forecasting of irradiance data as they recognize the problems that unanticipated changes in PV output can cause for the electric grid. The method that they use is Kriging, which obtains irradiance at an arbitrary location and time by considering the correlations between observed data. Different functions have been used with this Kriging method, but in this research a new non-separable anisotropic covariance function is introduced. It is based on an advanced form of the Lagrangian covariance function developed by Schlather [9] and has the capability to consider complex cases with different weather conditions. Their proposed method leads to accurate results when the cloud motion is either steady or slowly varying in time and more accurate results compared to already existing Kriging methods when the cloud motions are unsteady. Results show that this anisotropic model has a 66% relative improvement over the persistence model.

1.3 Research aim

The aim of this research is to investigate the output variability of the Utrecht Science Park PV system, which consists of eight different PV systems, and see if there are differences between these different PV systems regarding output variability. This leads to the following research question:

What differences exist regarding output variability between the different PV systems within the Utrecht Science Park PV system?

Answering this research question provides insight in how the output variability varies between the different PV systems. How this research question is answered is explained in the methodology.

2 Methodology

The methods that are used to collect data and analyze the output variability of the Utrecht Science Park PV system are described in this section.

2.1 Programming

During this research, the programming language Python is used to analyze and restructure data within the Spyder editor. The choice for Python is made as it suitable for data analysis. Also, libraries that are useful for this research are available in python. The pandas library is used to manipulate and analyze the collected data [10]. The pvlib library is used to calculate the theoretical PV output of the different PV systems within the Utrecht Science Park as pvlib provides a set of functions and classes for simulating the performance of PV systems [11]. To clearly visualize the results, the matplotlib.pylot library is used, as it allows for plotting the data directly within the editor [12]. Lastly, the NumPy, SciPy, and scikit-learn libraries are used for mathematical functions, linear algebra routines, calculating correlations, and regression analysis [13] [14] [15].

2.2 Data collection and structuring

The eight different PV systems are equipped with inverters that log data, such as active power and inverter efficiency. This data is logged at a 5-minute interval and transmitted to a central SmartLogger through a wireless connection. Each system has its own SmartLogger, saving data for each individual inverter. This data can be accessed through the NetEco platform, where data can be shown in tables, visualized in graphs, and exported as CSV files. These CSV files can be used as input data for further analysis. After one month, the 5-minute data is replaced by 15-minute data by averaging the 5-minute data over 15 minutes. To collect this 15-minute time resolution data, the data has to be collected day by day, resulting in a high number of files. To concatenate this data, the pandas function 'concat' is used. Data from 2019 is used as this data is readily available in the NetEco platform. However, data for the MdB, VJK, and DdW buildings is missing due to a communication error within the SmartLogger, so the PV systems on these buildings aren't analyzed with this data.

Additionally, the PV production data is also collected by the energy department of the Utrecht University. This data has a 15-minute time resolution and logs the energy production instead of the active power each 15 minutes. This dataset only misses data for the DdW building, which means that the MdB and VJK buildings can be analyzed with this data in addition to the other five buildings. This data is converted to CSV files, and each file contains the data for one building for a whole year. These CSV files are then combined into one CSV file, which makes it possible to use it as input for Python. Both PV datasets are in local time, so they are converted to the UTC time zone with the 'dt.tz_convert' function to prevent errors. From now on, the first dataset is referred to as the NetEco dataset, and the second dataset is referred to as the UU dataset.

The Utrecht Photovoltaic Outdoor Test Facility (UPOT) pyranometer data is not available for 2019, so irradiance data is acquired through KNMI. Wind speed, temperature, and air pressure data is also acquired through KNMI [16]. The KNMI location that is used for this data is De Bilt. As this is relatively close (±2 km) to the Utrecht Science Park, it is assumed that the data from KNMI is an accurate representation of the actual situation in the Utrecht Science Park.

Before the KNMI data can be read correctly in Python, the date and time must be restructured. Instead of using the standard year/month/day/hour format, KNMI has the 'hour' value in a separate column. Also, this 'hour' value ranges from 1 to 24, instead of the usual 0-23, which Python (pandas) uses as

well. To fix this, a value of one is subtracted from all the 'hour' values in the KNMI data and the year/month/day column are combined with this 'hour' column. This new column can be read by Python correctly after using it as input for the 'to_datetime' function and setting this new column as the index of the dataframe. The KNMI data uses the UTC time zone, so it doesn't have to be converted anymore.

The data from KNMI has a time resolution of one hour, which makes it incompatible with the 15-minute time resolution PV output data. Therefore, the PV output data is resampled to match the KNMI data. As KNMI measures the total global horizontal irradiance per hour (J/cm^2) instead of the irradiance at a certain time (W/cm^2), the resampling method that is used for this is the 'mean' method for the NetEco dataset, and 'sum' for the UU dataset. These different methods are used, because the NetEco dataset logs active power and the UU dataset logs the energy production. This means that the average power during one hour in the UU dataset is the sum of the energy production after dividing by one hour. The old 15-minute time resolution is saved as well, as it can be used for calculations where KNMI data isn't used. For the UU dataset, all energy production values in the 15-minute time resolution dataframe are multiplied by four, as this results in the average power per 15 minutes.

The PV output with a one-hour time resolution represents the average power level per hour, so to compare this with the total plane of array irradiance data based on KNMI, this irradiance is calculated in the middle of every hour. Because of this, the index of the KNMI and PV output dataframes with a one-hour time resolution are offset by 30 minutes.

The wind speed, temperature, and air pressure data from KNMI data is also used in calculations where the time resolution is 15 minutes, so this data is resampled to a 15-minute time by using the 'ffill' method. This fills all the new empty cells with the most recent known value.

2.3 Definitions

2.3.1 PV System

PV system refers to one of the PV systems within the Utrecht Science Park, which consists of eight different systems. The different PV systems are distinguished by the names of the different buildings where the PV systems are installed. However, only seven of these PV systems are analyzed, namely: CB, UB, JDV, WCS, WGB, VJK, and MdB. Data for the DdW building is missing, as explained above.

2.3.2 Performance ratio

The performance ratios that are calculated here refer mostly to the efficiency of the inverter, but other potential losses are also considered, for example cable losses. The performance ratios of the different PV systems are calculated to see how the different systems compare with the efficiency curve of a Huawei inverter, as the different PV systems use the SUN2000 inverter from Huawei. This efficiency curve can be seen in Figure 3.

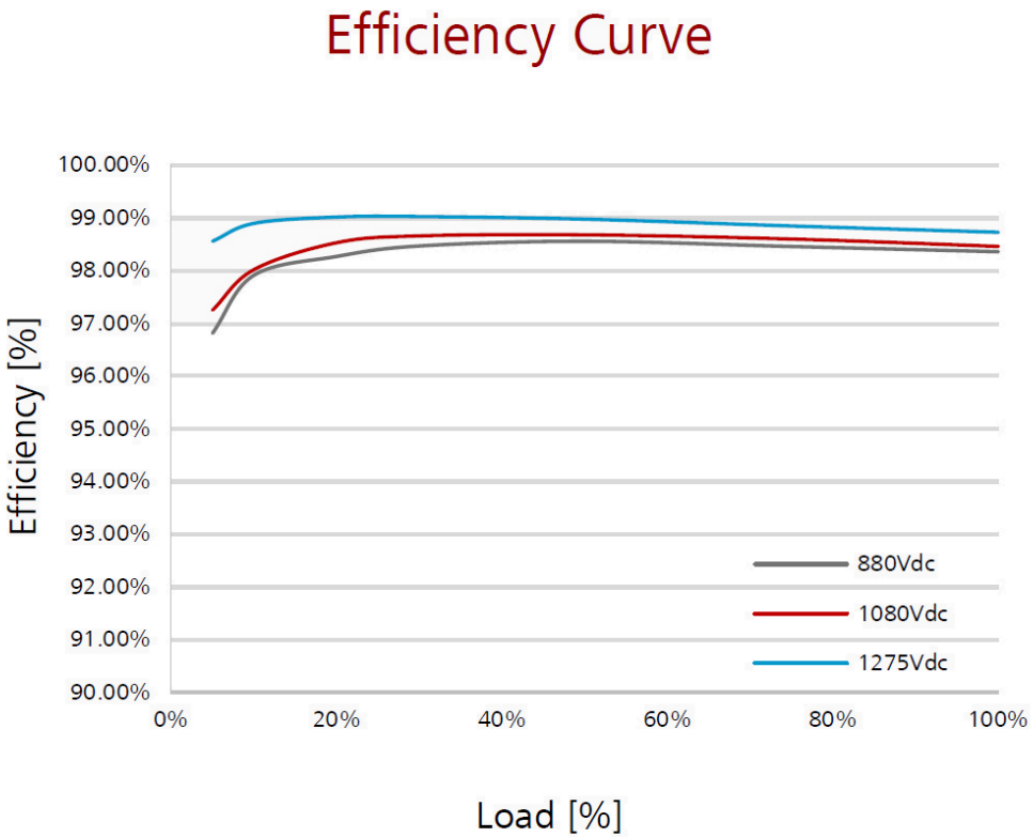

Figure 3: Efficiency curve of a Huawei inverter [17].

These calculations are done with KNMI data, which means that the time resolution is one hour. However, the irradiance data from KNMI consists of only the global horizontal irradiance (GHI). It measures the GHI during an hour in J/cm^2, so to convert this to W/m^2 it is divided by 3,600 seconds after being multiplied by 10^4.

To be able to compare this KNMI data to the actual output of the system, the GHI must be converted to the total plane of array irradiance. As each PV system has different characteristics, this is done for each PV system individually. To calculate the total plane of array irradiance, the direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI) must be calculated first. The DNI is calculated with the DIRINT model within pvlib, which uses the GHI, solar zenith angle, and air pressure [11]. The required solar zenith angle is calculated with the pvlib function 'solarposition.get_solarposition', which takes latitude, longitude, air pressure, and ambient temperature as input. The latitude and longitude are chosen approximately in the middle of the buildings that are analyzed, resulting in a latitude of 52.08 N and a longitude of 5.17 E.

The DHI can only be calculated after calculating the angle of incidence (AOI), however, this angle is the same as the solar zenith angle for a horizontal plane. This leads to Equation 1.

$$\mathrm{DHI} = \mathrm{GHI} - \cos(\theta_s) * \mathrm{DNI} \qquad \textit{Eq. 1}$$

With θ_s = solar zenith angle.

The total plane of array irradiance can then be calculated with the pvlib function 'irradiance.get_total_irradiance' for the different PV systems. The inputs for this function are the surface tilt and azimuth angle, the solar zenith and azimuth angles, and the GHI, DNI, and DHI.

Now, the system and reference yield can be calculated with Equation 2 and Equation 3 respectively.

$$Y_f = \frac{E_{AC}\,[\mathrm{Wh}]}{W_{peak}\,[\mathrm{W}]} \qquad \textit{Eq. 2}$$

With Y_f = system yield, E_{AC} = amount of energy delivered by the system, and W_{peak} = installed capacity of the system.

$$Y_r = \frac{H_{POA}\,\left[\frac{\mathrm{Wh}}{\mathrm{m}^2}\right]}{G_{STC}\,\left[1000\,\frac{\mathrm{W}}{\mathrm{m}^2}\right]} \qquad \textit{Eq. 3}$$

With Y_r = reference yield, H_{POA} = summed plane of array irradiance, and G_{STC} = Standard Test Conditions reference irradiance, which is 1000 W/m^2.

Finally, the performance ratio is calculated with Equation 4.

$$\text{Performance Ratio (PR)} = \frac{Y_f}{Y_r} \qquad \textit{Eq. 4}$$

However, as can be seen in Figure 3, the efficiency of inverters varies strongly, especially when the output power of the system relative to the inverter is between 0% and 20%. Therefore, using an average performance ratio of a year might not be useful for different kinds of calculations. Instead of calculating the average performance ratio, the performance ratios are calculated for different increments of percentages of output power. The range between 20% and 100% is handled as one increment, as the inverter efficiency only varies slightly within this range. The range between 0% and 20% is divided into multiple increments between 1% and 5%, depending on the amount of data that is available at these low percentages. The more data that is available within this 0% to 20% range, the smaller these increments can be to be able to make more precise calculations. The overall performance ratio is also calculated to show the difference compared to the increments.

Finally, the standard mean error of the mean is calculated for these performance ratios with the SciPy library [14].

2.3.3 Clear sky irradiance

The clear sky irradiance is calculated with the Ineichen and Perez clear sky model [11] to be able to calculate the theoretical output of the different PV systems with a 15-minute time resolution. First, the solar angles are calculated in the same way as in 2.3.2. This returns, among others, the following solar angles: apparent zenith, zenith, and azimuth. With the apparent zenith angle and the pressure data from KNMI, the relative and absolute air mass (AM) can be calculated with the pvlib functions 'atmosphere.get_relative_airmass' and 'atmosphere.get_absolute_airmass'. The Linke turbidity can be determined with the latitude and longitude with the pvlib function 'clearsky.lookup_linke_turbidity', which describes the amount of gaseous water vapour and aerosols in the atmosphere, which have an influence on solar irradiance [18]. Finally, the apparent zenith angle, absolute air mass and Linke turbidity are used to calculate the clear sky irradiance by using the pvlib function 'clearsky.ineichen'. This function returns the GHI, DNI, and DHI.

After the clear sky irradiance has been calculated, the total plane of array irradiance is calculated with the pvlib function 'irradiance.get_total_irradiance' for the different PV systems. The inputs for this function are the surface tilt and azimuth angle, the solar zenith and azimuth angles, and the GHI, DNI,

and DHI. The total plane of array irradiance is used to calculate the (simulated) theoretical output of the different PV systems.

2.3.4 Theoretical output

To calculate the theoretical output of the different PV systems, a transposition model is used. This model takes into account the following variables: temperature of the cell, angle of incidence, effective irradiance, and the type of module and inverter.

The cell temperature is determined by the pvlib function 'pvsystem.temperature.sapm_cell' with the incoming plane of array irradiance, ambient temperature, wind speed, and the module construction and its mounting. As the theoretical output is being calculated, the clear sky plane of array irradiance is used as the incoming irradiance. KNMI data is used for the ambient temperature and wind speed. All PV systems employ modules that are made of single glass and polymer and are mounted as open rack. This influences how and how fast heat is disposed of from the modules.

The pvlib function 'irradiance.aoi' determines the angle of incidence with the surface tilt and azimuth, and the solar zenith and azimuth angles.

The effective irradiance is determined by the pvlib function 'pvlib.pvsystem.sapm_effective_ irradiance' with the direct and diffuse irradiance, absolute air mass, angle of incidence, and the module type. Similar to above, the clear sky direct and diffuse irradiance is used. The absolute air mass and angle of incidence have been calculated before. As the specific type of module and inverter couldn't be retrieved, reference models are used in this calculation. 'Sanyo_HIP_225HDE1__2008__E__' is used for the module, and 'Enphase_Energy_Inc___ M210_84_240_Sxx__240V_' for the inverter.

With this information, the AC power can be calculated for the different PV systems with a 15-minute time resolution, which is considered the theoretical output of the PV systems.

2.3.5 Simulated output

To get insight in how much the different inverter to system capacity ratios and different orientations and tilts of the modules influence the output variability, the output of the PV systems is simulated in a way that all the modules are orientated towards the south with a 10-degree surface tilt.

First, the theoretical simulated output must be calculated. This is calculated in the same way as the theoretical output above, but now the surface azimuth and surface tilt are set to south and 10 degrees respectively for every PV system. This means that the clear sky irradiance is recalculated as well. When this theoretical simulated output is calculated, there is a certain ratio between this output and the theoretical output. However, this ratio differs significantly at different levels of output power, which can partly be explained by Figure 3, which illustrates how different levels of output power influence the efficiency. Because of this, ratios are calculated for different increments of system yield to make more precise calculations. As can be seen in Figure 3, the part between 0% and 20% requires the most increments as the efficiency varies strongly across these levels of output power. The exact increments that are chosen for this calculation are based on the results.

Finally, these ratios are used to convert the PV output of the different PV systems based on the input data to PV output that is simulated to be orientated towards the south with a 10-degree surface tilt.

2.3.6 Output variability

The output variability is a measure of the power output change of a PV system over a selected time interval, in this case 15 minutes, relative to the capacity of that PV system. Thus, output variability can be defined as a percentage of the PV system's capacity. However, as explained earlier, part of this variability is precisely predictable, which is not interesting for this research. Because of this, only the

much less predictable cause is considered, which depends on the motion of the clouds and weather systems. This is done by comparing the (simulated) theoretical output of the PV systems with the (simulated) actual output of the system. This leads to Equation 5.

$$P_{var}(g, t) = \frac{P(g, t)}{P_{theoretical}(g, t)} \quad \textit{Eq. 5}$$

With P_{var} = power output variability, P = power output, $P_{theoretical}$ = theoretical power output, t = time, and g = one of the PV systems that is analyzed.

Apart from calculating the output variability per 15 minutes, it is also done per day. For the daily variability, all 15-minute variability values are first subtracted from a value of 1, after which they are summed per day. This means that a higher value means more output variability.

2.3.7 Correlation

To determine the correlation between the different PV systems, the Pearson correlation coefficient, or Pearson's *r*, is calculated for all the different pairs of PV systems. This coefficient is a measure of the strength of the linear correlation between two variables. The Pearson correlation coefficient is calculated with Equation 6 [19].

$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}} \quad \textit{Eq. 6}$$

With m_x = mean of the vector x, and m_y = mean of the vector y

To calculate Pearson's *r*, the SciPy library is used, more specifically the 'stats' part. With SciPy, Pearson's *r* is calculated with the 'stats.pearonsr' function, which takes two arrays, X and Y, as input. These two arrays are two variables, which in this case are two different PV systems.

2.3.8 Regression

Linear and polynomial regression is used to visualize certain patterns in different graphs. The type of regression that is conducted is based on the graph. When there is no indication of a linear correlation, a polynomial regression is conducted instead of a linear regression. This regression analysis is done with the scikit-learn library, which has different built-in functions to carry out both linear and polynomial regression analyses [15].

2.4 Analysis

As mentioned earlier, only the second (less predictable) cause of output variability is analyzed. This means that days where the capacity of the inverter of one or more of the PV systems is reached aren't suitable for this analysis. These peaks happen because the inverters are 15-30% smaller than the peak capacity of the PV system that they are connected to. When these peaks happen the output variability as defined above changes if the incoming irradiance keeps increasing as a result of the first (predictable) cause of output variability, which is the reason for these days to not be analyzed. In the NetEco dataset, all PV systems reach the capacity of the inverter in some occasions, while this only happens for the JDV and WGB systems in the UU dataset. Days with missing data due to for example a communication error with the SmartLogger are also not analyzed, as incomplete data leads to problems within Python and the visualization of the data. The NetEco dataset is missing data for all PV systems for the 16th and 17th of February and from the 20th of July through the 3rd of August for the JDV

system, which means that these dates are not analyzed with the NetEco dataset. The UU dataset isn't missing any data, so all dates can be analyzed with the UU dataset.

3 Results

This section shows and explains the results of this research, visualizing them with different graphs.

3.1 Monthly yield

To visualize the difference between the PV systems, the monthly yield is plotted per PV system.

3.1.1 NetEco dataset

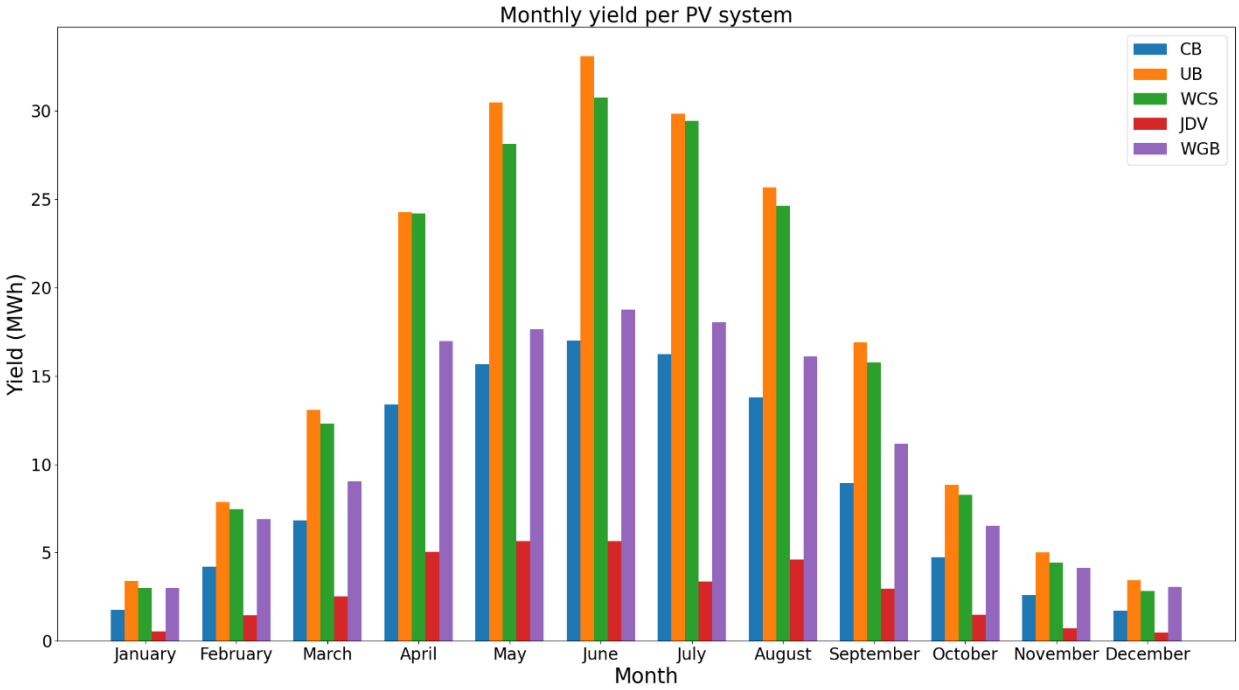

Figure 4: Monthly yield per PV system in 2019.

The missing data, mainly in July, explains the low yield of the JDV system in July. Apart from that, it can be noticed that the WGB system has a relatively high yield in November through February, considering its capacity compared to the UB and WCS systems. This is because of the 30-degree surface tilt and orientation towards the south of the WGB system. However, Figure 4 shows that the yield of the WGB system in the other months is significantly lower relative to the UB and WCS systems. The cause of this is analyzed and explained later in this report.

3.1.2 UU dataset

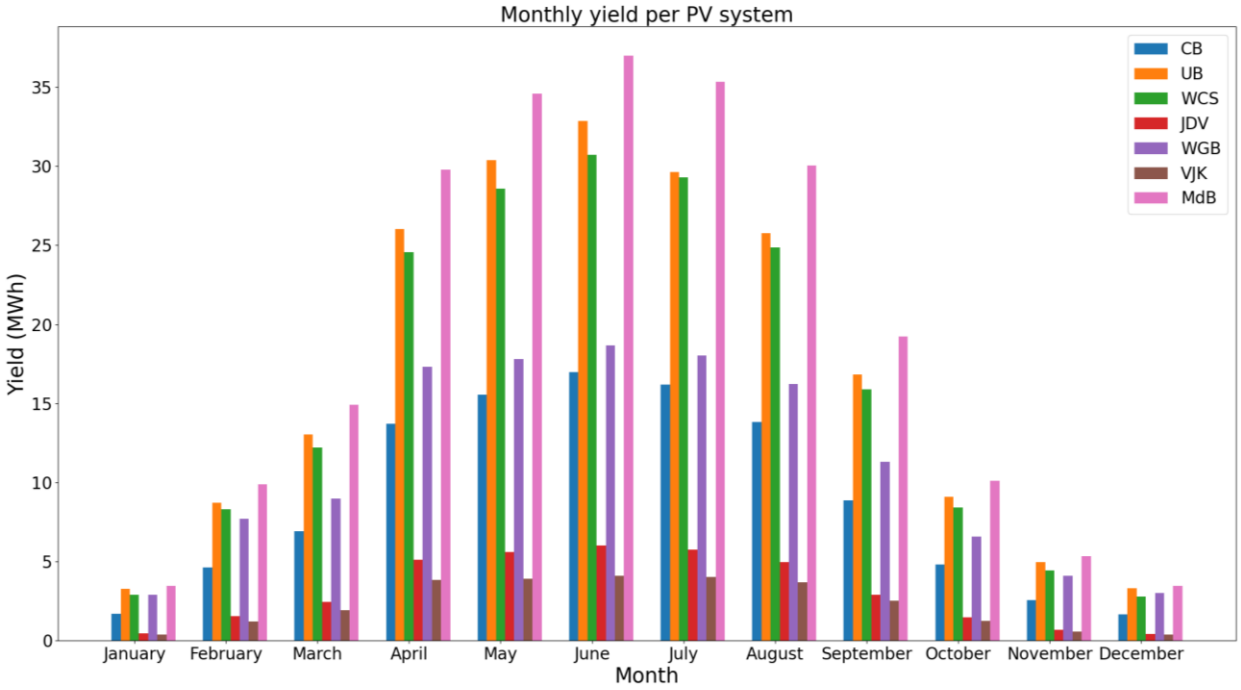

Figure 5: Monthly yield per PV system in 2019.

Figure 5 shows similar results to Figure 4, with the same pattern being visible for the WGB system.

3.2 Visualization of the input data

To visualize the input data, the system yield of the different PV systems is plotted against the reference yield based on the KNMI data. A linear regression is done and shown in the graphs, indicated by the red line, to show the relation between these two input variables.

3.2.1 NetEco dataset

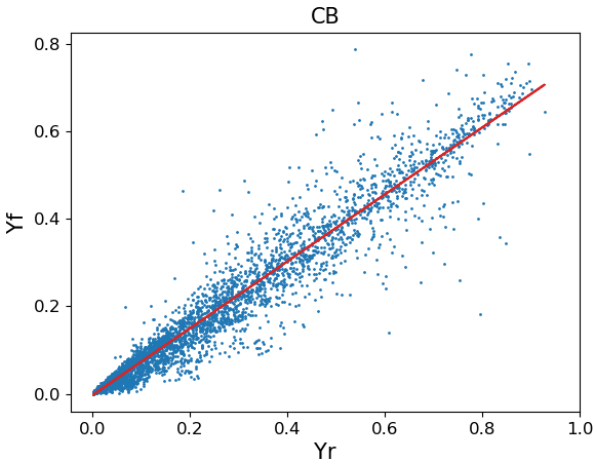

Figure 6: CB system yield and KNMI reference yield.

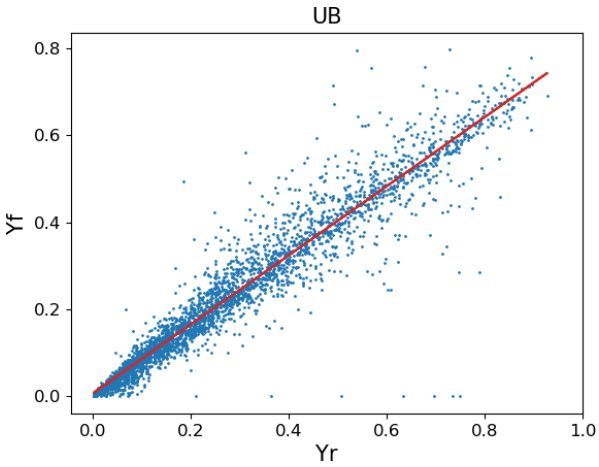

Figure 7: UB system yield and KNMI reference yield.

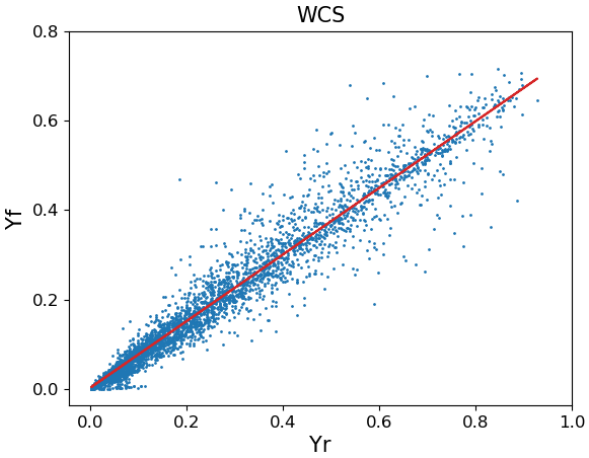

Figure 8: WCS system yield and KNMI reference yield.

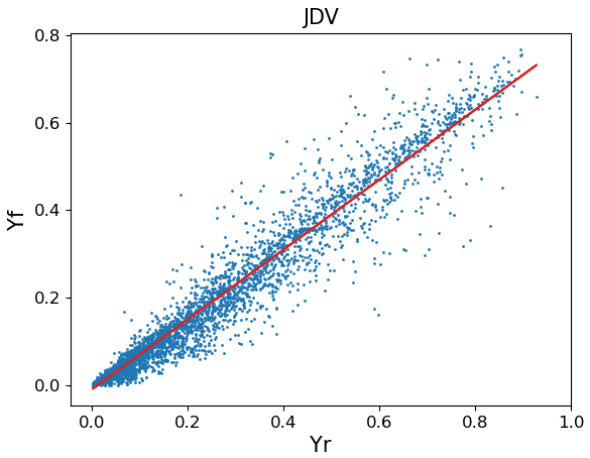

Figure 9: JDV system yield and KNMI reference yield.

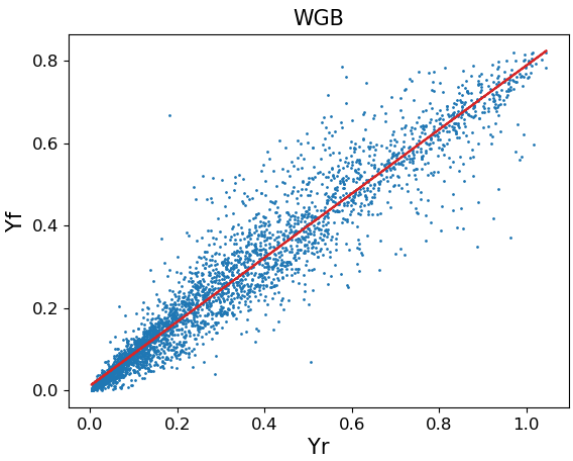

Figure 10: WGB system yield and KNMI reference yield.

Figures 6 through 10 look quite similar, except for Figure 10 showing a bit more spread. This can be explained by the WGB PV system being orientated towards the south with a 30-degree surface tilt. This leads to higher peaks both in the system yield and the reference yield, as the orientation and surface tilt are more favorable compared to the other PV systems. Therefore, the spread of the WGB values is a bit wider. Also, WGB is the only PV system where the reference yield reaches a value of 1.0 or higher, which can be explained by the same reasoning as above.

3.2.2 UU dataset

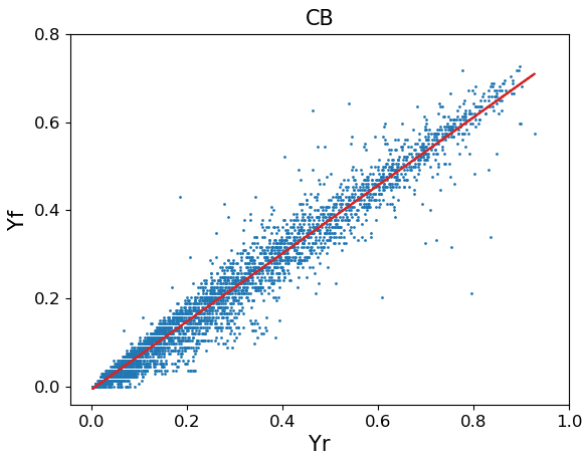

Figure 11: WCS system yield and KNMI reference yield.

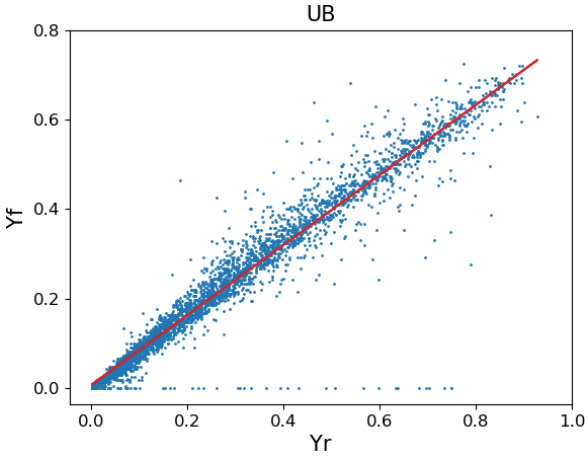

Figure 12: JDV system yield and KNMI reference yield.

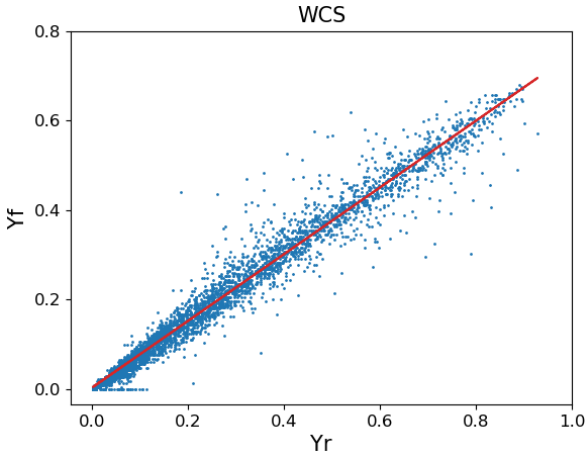

Figure 13: WCS system yield and KNMI reference yield.

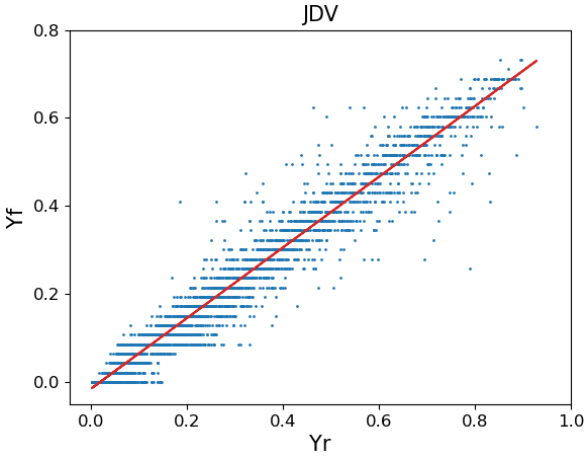

Figure 14: JDV system yield and KNMI reference yield.

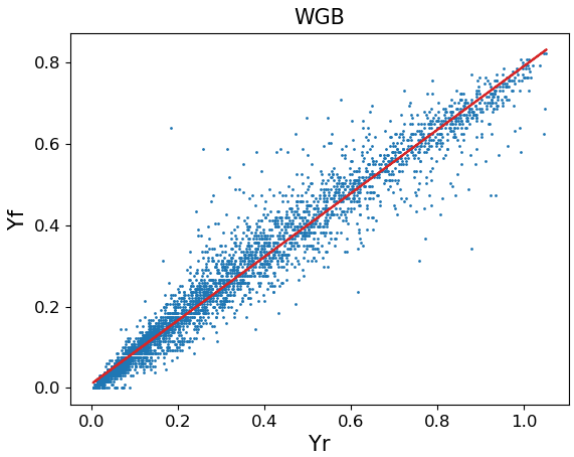

Figure 15: WGB system yield and KNMI reference yield.

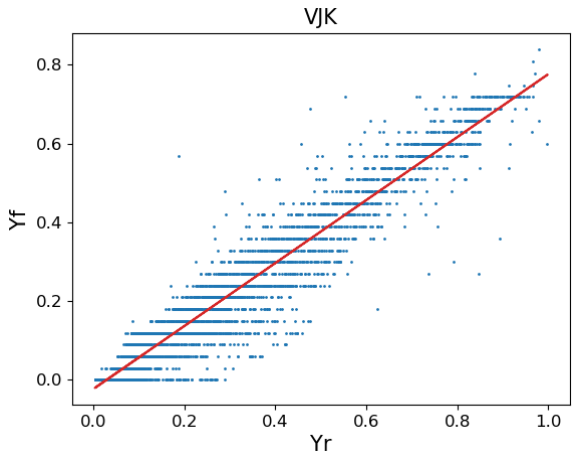

Figure 16: VJK system yield and KNMI reference yield.

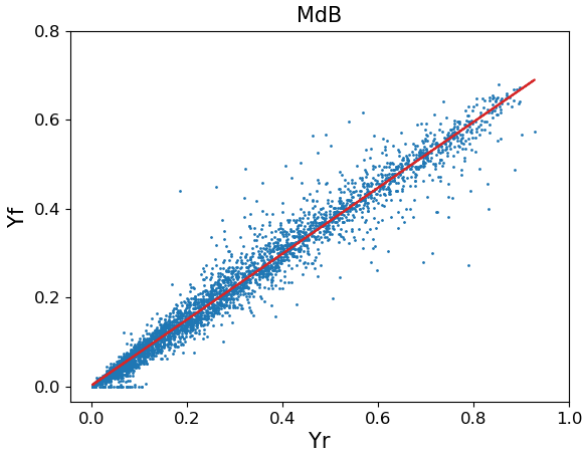

Figure 17: MdB system yield and KNMI reference yield.

Noticeable is that the graphs for the JDV and VJK PV systems look a bit different than the other graphs. This is caused by the fact that the UU dataset doesn't use decimal values and that these two PV systems are relatively small compared to the other PV systems. Measurement accuracy thus is lower compared to the other systems. This means that there is a low variety of different output numbers, which creates this different looking graph.

Apart from this, the different graphs look quite similar again. The WGB graph has a bit of a wider spread, but this is less noticeable than in Figure 10. Compared to the NetEco dataset, the spreads in these graphs look a bit narrower. This can be explained by the difference in logging active power instead of energy production. The logging of energy production can smooth out peaks or dips that might be present in the active power data, leading to the output numbers being closer together.

3.3 Performance ratio

Enough data is available to divide the part where the output power is between 0 and 20% into increments of 2.5%. System yields lower than 1% are ignored, as many unrealistic values show up in this segment, i.e. the performance ratio far exceeding a value of 1. Remaining performance ratio values higher than 1 are replaced by a value of 1 as higher values are unrealistic. These values probably show up as a result of the difference in locations of the different PV systems and De Bilt, where the KNMI data is measured.

As the graphs of the different PV systems for the performance ratios at different levels of system yield look quite similar again, like the visualization of the data, only two PV systems are shown per dataset. A polynomial regression is done to visualize the pattern of these graphs. Additionally, the increments and their corresponding performance ratios are shown in tables for every PV system, followed by a table with the standard error of the mean of these performance ratios.

3.3.1 NetEco dataset

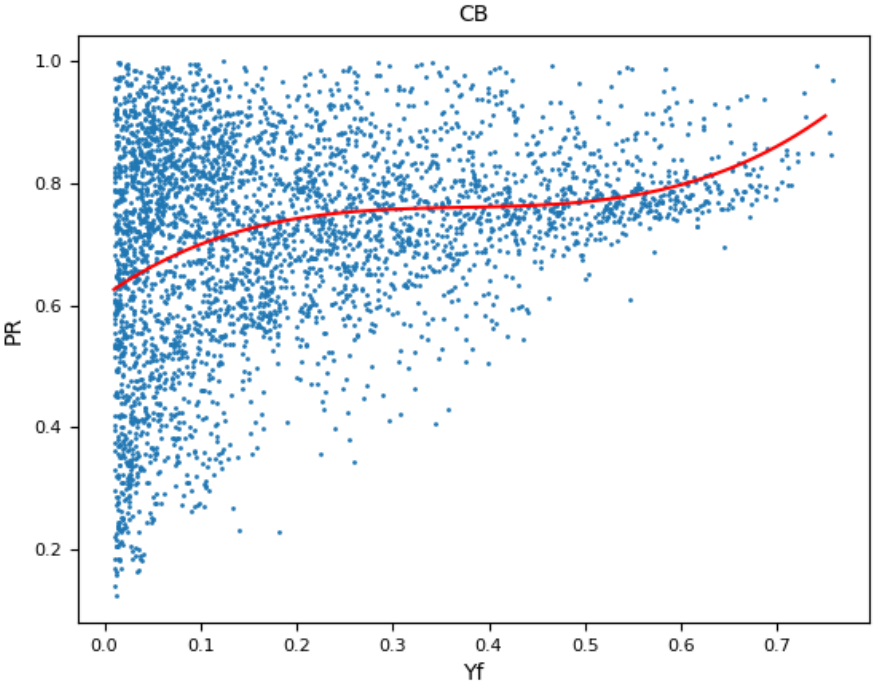

Figure 18: Performance ratio of the CB PV system at different levels of system yield, calculated per hour.

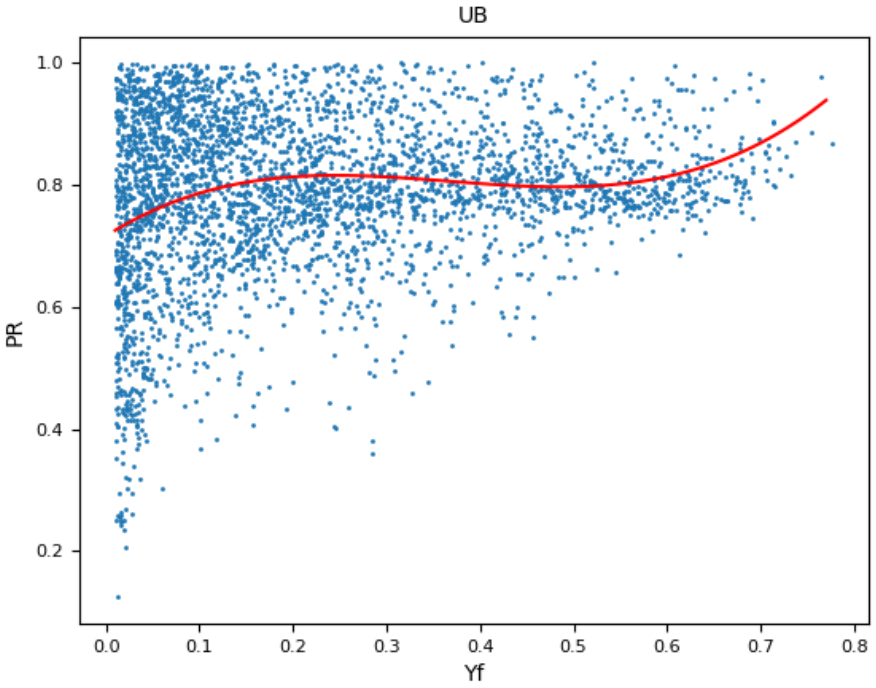

Figure 19: Performance ratio of the UB PV system at different levels of system yield, calculated per hour.

Figure 18 and 19 show a mostly similar pattern as in Figure 3; however, the performance ratio values are significantly lower. This is because Figure 16 and 17 don't just take into account the efficiency of the inverter, but also other factors, for example cable losses, rainfall, and unclean panels.

The increments and their corresponding performance ratios for every PV system are shown in Table 2, with their standard error of the mean being displayed in Table 3.

Table 2: Increments with their corresponding performance ratios.

Increment (Yf)	**CB**	**UB**	**WCS**	**JDV**	**WGB**
1% - 2.5%	0.65	0.76	0.71	0.60	0.76
2.5% - 5%	0.68	0.79	0.73	0.65	0.81
5% - 7.5%	0.73	0.85	0.78	0.68	0.82
7.5% - 10%	0.72	0.84	0.77	0.68	0.82
10% - 12.5%	0.72	0.82	0.76	0.71	0.85
12.5% - 15%	0.74	0.83	0.75	0.71	0.84
15% - 17.5%	0.72	0.82	0.76	0.72	0.83
17.5% - 20%	0.76	0.83	0.74	0.72	0.82
20% - 100%	0.78	0.83	0.77	0.80	0.83
1% - 100%	0.73	0.82	0.76	0.72	0.82

Table 3: Standard error of the mean of the values in Table 2.

Increment (Yf)	**CB**	**UB**	**WCS**	**JDV**	**WGB**
1% - 2.5%	0.012	0.011	0.011	0.011	0.010
2.5% - 5%	0.010	0.009	0.008	0.008	0.008
5% - 7.5%	0.010	0.007	0.008	0.009	0.010
7.5% - 10%	0.012	0.008	0.008	0.010	0.012
10% - 12.5%	0.011	0.009	0.009	0.010	0.012
12.5% - 15%	0.011	0.009	0.009	0.010	0.013
15% - 17.5%	0.010	0.009	0.010	0.012	0.012
17.5% - 20%	0.011	0.009	0.010	0.010	0.015
20% - 100%	0.003	0.003	0.003	0.003	0.003
1% - 100%	0.003	0.002	0.002	0.003	0.003

Figure 18 and 19 and Table 2 show that the absolute values of the performance ratios between the different PV systems differs, but the pattern exists for the different systems. This difference in absolute values has an impact on the output variability, which is further analyzed later in this paper.

Table 2 shows that the overall performance ratio can be significantly different from the performance ratios per increment. This can't be caused by the standard error of the mean, as the values in Table 3 are relatively low.

3.3.2 UU dataset

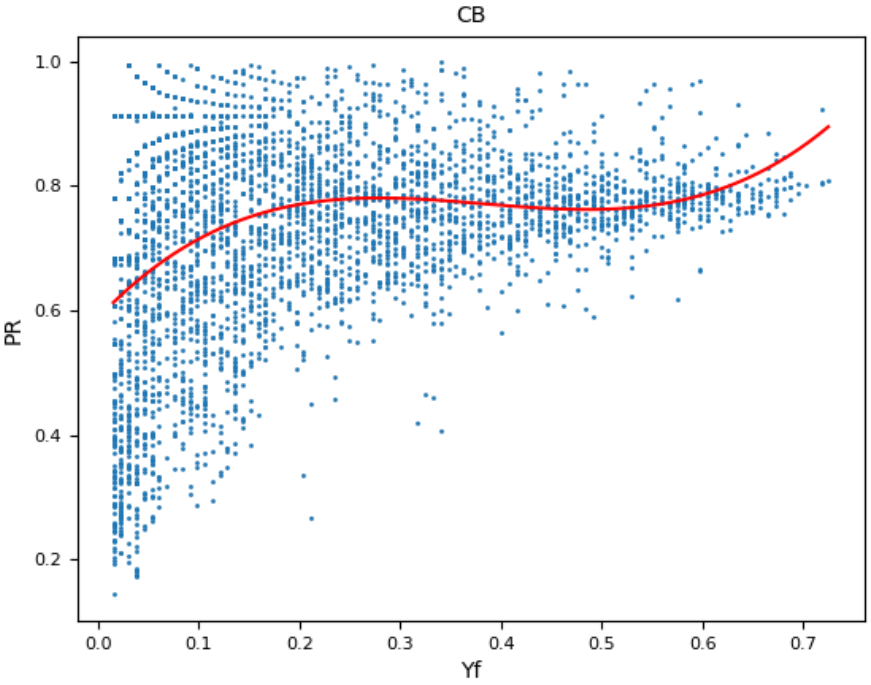

Figure 20: Performance ratio of the CB PV system at different levels of system yield, calculated per hour.

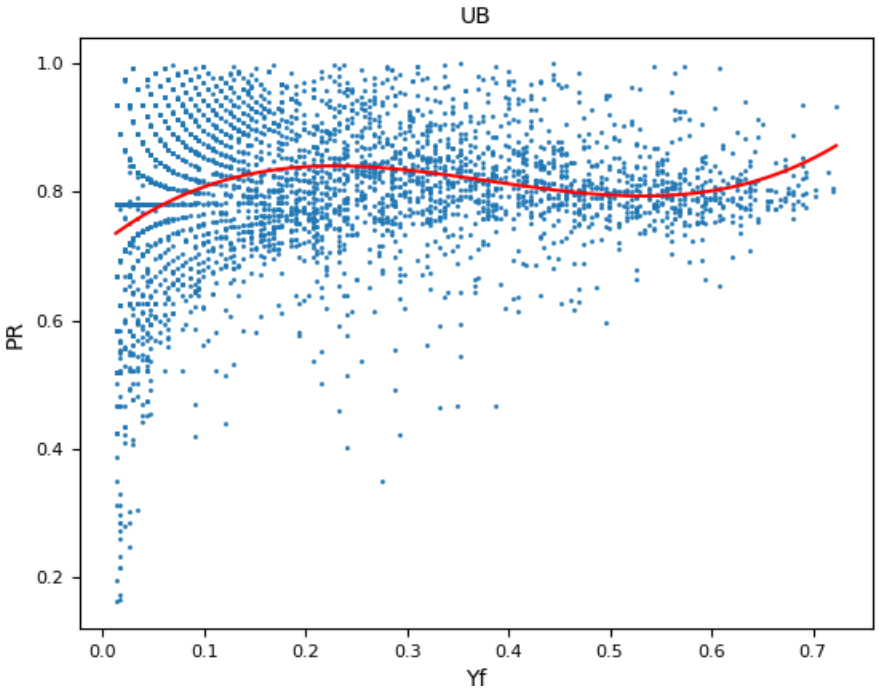

Figure 21: Performance ratio of the UB PV system at different levels of system yield, calculated per hour.

Similar to Figure 18 and 19, Figure 20 and 21 also show a mostly similar pattern as in Figure 3.

The increments and their corresponding performance ratios for every PV system are shown in Table 4, with their standard error of the mean being displayed in Table 3.

Table 4: Increments with their corresponding performance ratios.

Increment (Yf)	**CB**	**UB**	**WCS**	**JDV**	**WGB**	**VJK**	**MdB**
1% - 2.5%	0.56	0.72	0.69	0.40	0.75	-	0.67
2.5% - 5%	0.68	0.77	0.71	0.63	0.80	0.45	0.72
5% - 7.5%	0.73	0.83	0.77	0.75	0.84	0.59	0.76
7.5% - 10%	0.74	0.84	0.76	0.72	0.83	0.71	0.75
10% - 12.5%	0.73	0.85	0.77	0.69	0.84	0.74	0.77
12.5% - 15%	0.72	0.85	0.77	0.73	0.88	0.70	0.76
15% - 17.5%	0.76	0.83	0.77	0.73	0.86	-	0.76
17.5% - 20%	0.77	0.84	0.76	0.76	0.83	0.72	0.75
20% - 100%	0.78	0.83	0.77	0.79	0.83	0.79	0.77
1% - 100%	0.73	0.82	0.76	0.73	0.83	0.73	0.75

Table 5: Standard error of the mean of the values in Table 4.

Increment (Yf)	**CB**	**UB**	**WCS**	**JDV**	**WGB**	**VJK**	**MdB**
1% - 2.5%	0.012	0.011	0.010	0.008	0.010	-	0.008
2.5% - 5%	0.010	0.007	0.007	0.011	0.008	0.010	0.007
5% - 7.5%	0.010	0.006	0.007	0.013	0.009	0.014	0.007
7.5% - 10%	0.009	0.007	0.008	0.009	0.009	0.017	0.008
10% - 12.5%	0.011	0.007	0.007	0.011	0.011	0.010	0.008
12.5% - 15%	0.011	0.007	0.007	0.011	0.009	0.013	0.007
15% - 17.5%	0.009	0.007	0.007	0.007	0.010	-	0.008
17.5% - 20%	0.009	0.007	0.008	0.008	0.010	0.013	0.008
20% - 100%	0.002	0.002	0.002	0.002	0.002	0.003	0.002
1% - 100%	0.003	0.002	0.002	0.003	0.002	0.003	0.002

There are no output values in the 0% - 2.5% and 15% - 17.5% increments for VJK, which explains the absence of values in these cells.

Figure 20 and 21 and Table 4 also show the same pattern between the different PV systems, with differences in absolute values, which influences the output variability.

Again, Table 4 shows that the overall performance ratio can be significantly different from the performance ratios per increment. This can't be caused by the standard error of the mean either, as the values in Table 3 are relatively low.

3.4 Theoretical & simulated output

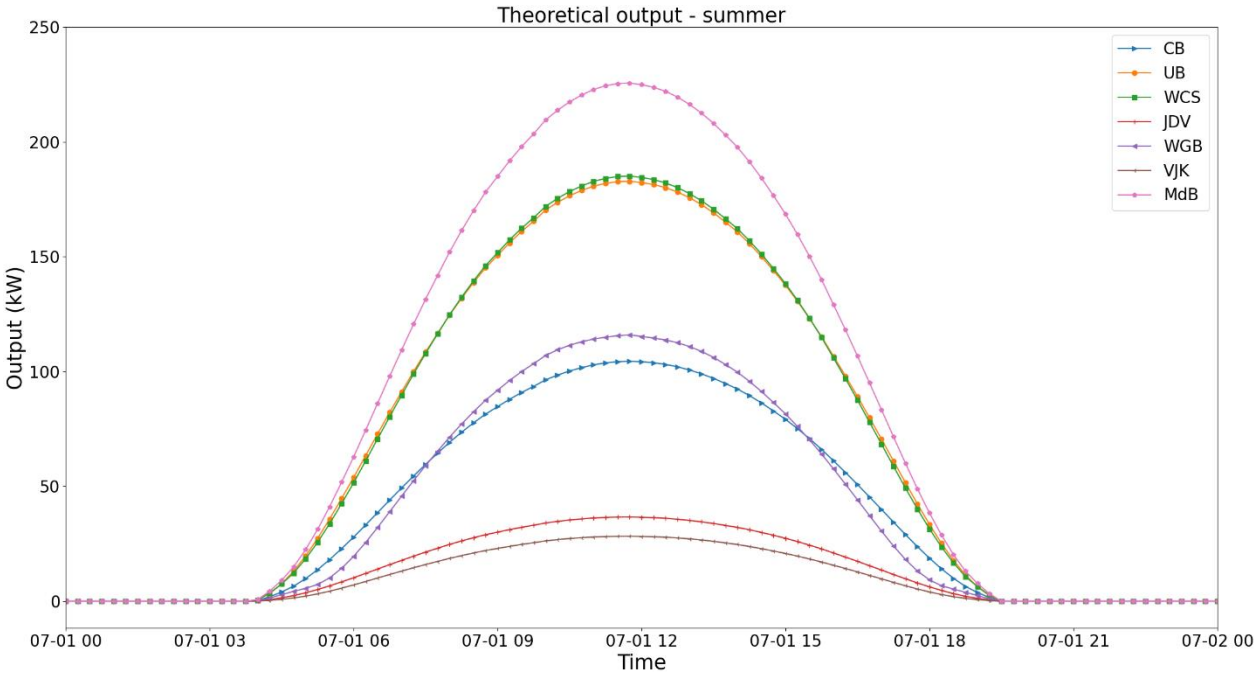

Figure 22: Theoretical output of the different PV systems on a summer day, 1st of July.

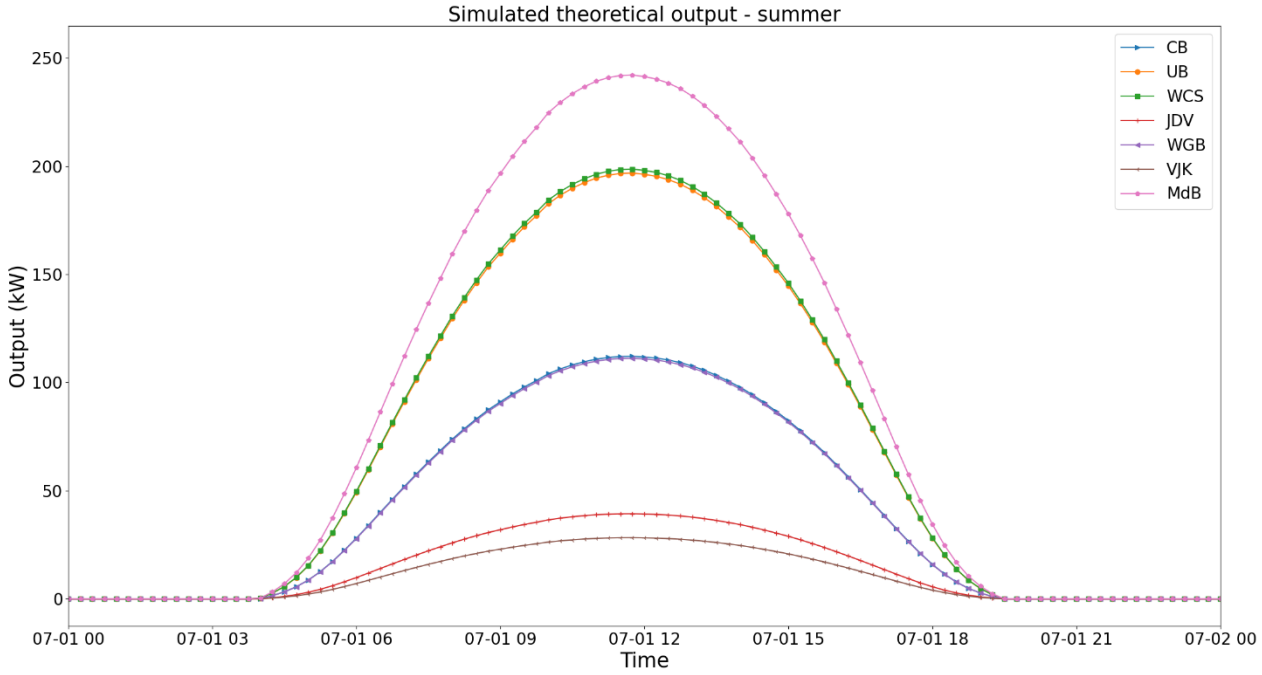

Figure 23: Simulated theoretical output (10-degree surface tilt, facing south) of the different PV systems on a summer day, 1st of July.

From these Figure 22 and 23 can be concluded that the difference in orientations of the different PV systems doesn't change the PV output significantly on a summer day in the theoretical model. This is due to the systems only having a 10-degree surface tilt, except for WGB which is already orientated

towards the south. The zenith angle in the summer is much smaller than in the winter, which makes a difference in orientation at a 10-degree surface tilt have a negligible impact on the PV output.

However, the change in surface tilt does have an impact, as the output curve for the WGB building is noticeably different when the surface tilt changes from 30 degrees to 10 degrees. Instead of having a higher peak and lower output values when the sun comes up and goes down, the WGB output curve is almost identical to the CB output curve at a 10-degree surface tilt. This makes sense, as the system capacities of the CB and WGB buildings differ less than 1%.

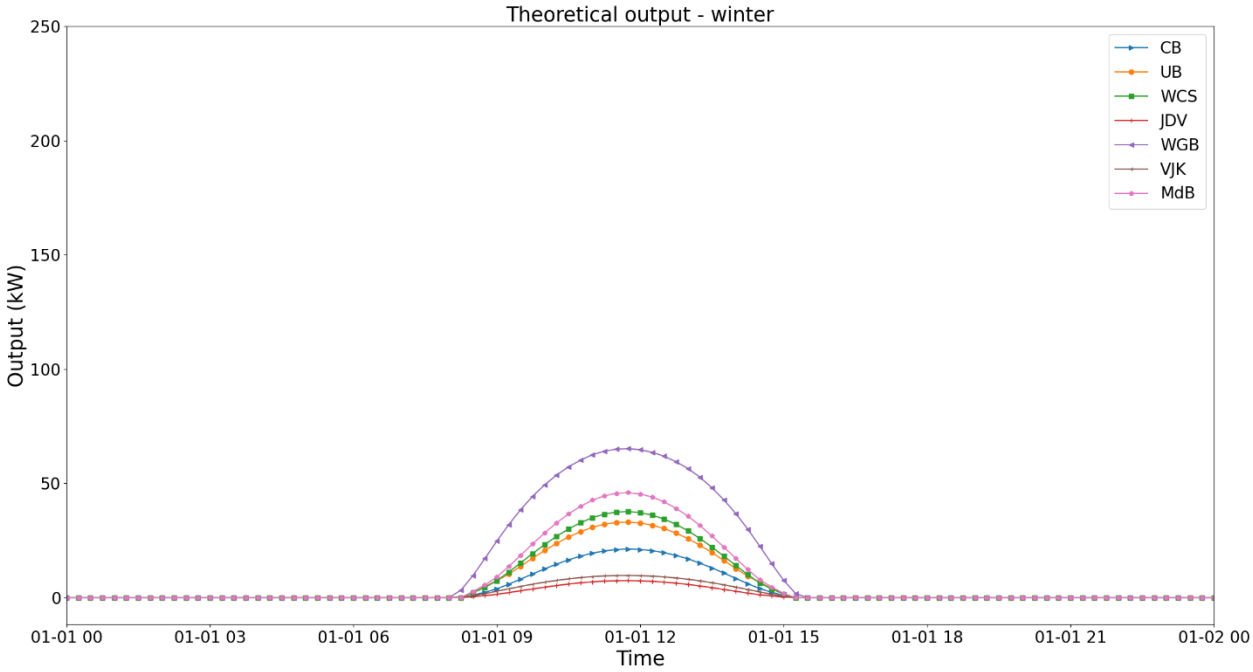

Figure 24: Theoretical output of the different PV systems on a winter day, 1st of January.

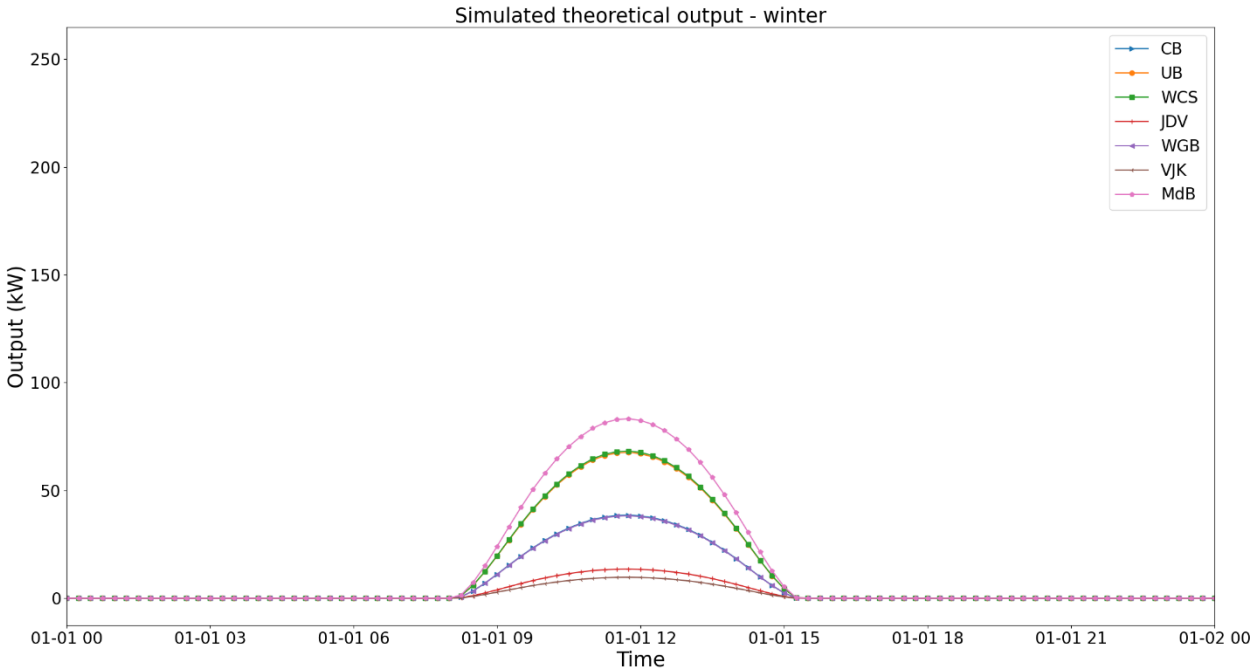

Figure 25: Simulated theoretical output (10-degree surface tilt, facing south) of the different PV systems on a winter day, 1st of January.

For the winter day, the difference in orientations does make a significant difference as can be seen in Figure 24 and 25. The 10-degree surface tilt is not negligible in the winter, as the zenith angle is much larger than in the summer. This results in a large part of the sun's irradiance not reaching the PV panels if they are not facing towards the sun. For example, the peak of the MdB output curve increases by approximately 40% as a result of the change in orientation from east and west to south.

The change in surface tilt also has more of an impact compared to the summer day. The peak of the WGB output curve decreases by approximately 41% after the surface tilt change from 30 degrees to 10 degrees. This can again be explained by the larger zenith angle in the winter.

As explained in the methodology, increments are used to determine the simulated output. These increments and their corresponding ratios between the simulated theoretical output and the theoretical output are presented in table 4.

Table 4: Increments with their corresponding ratios to calculate the simulated output.

Increment (Y_f)	**CB**	**UB**	**WCS**	**JDV**	**WGB**	**VJK**	**MdB**
0% - 2.5%	2.20	2.58	1.74	1.74	1.01	1.00	1.74
2.5% - 5%	1.50	1.59	1.49	1.49	1.03	1.00	1.49
5% - 7.5%	1.59	1.59	1.57	1.57	1.10	1.00	1.57
7.5% - 10%	1.60	1.58	1.60	1.60	1.04	1.00	1.60
10% - 12.5%	1.55	1.57	1.58	1.58	1.05	1.00	1.58
12.5% - 15%	1.54	1.55	1.54	1.54	0.97	1.00	1.54
15% - 17.5%	1.55	1.56	1.53	1.53	0.94	1.00	1.53
17.5% - 20%	1.54	1.63	1.55	1.55	0.92	1.00	1.55
20% - 30%	1.39	1.41	1.38	1.38	0.86	1.00	1.38
30% - 40%	1.24	1.25	1.24	1.24	0.81	1.00	1.24
40% - 50%	1.17	1.16	1.17	1.17	0.77	1.00	1.17
50% - 100%	1.10	1.10	1.10	1.10	0.87	1.00	1.10

3.5 Output variability

The days that are visualized are chosen based on how clear the differences between the PV systems can be shown. The graphs that show the summed daily output variability only have a timeline of one month, as showing multiple months decreases the clarity considerably. However, the output variability is summed per month for all the PV systems and shown in tables.

3.5.1 NetEco dataset

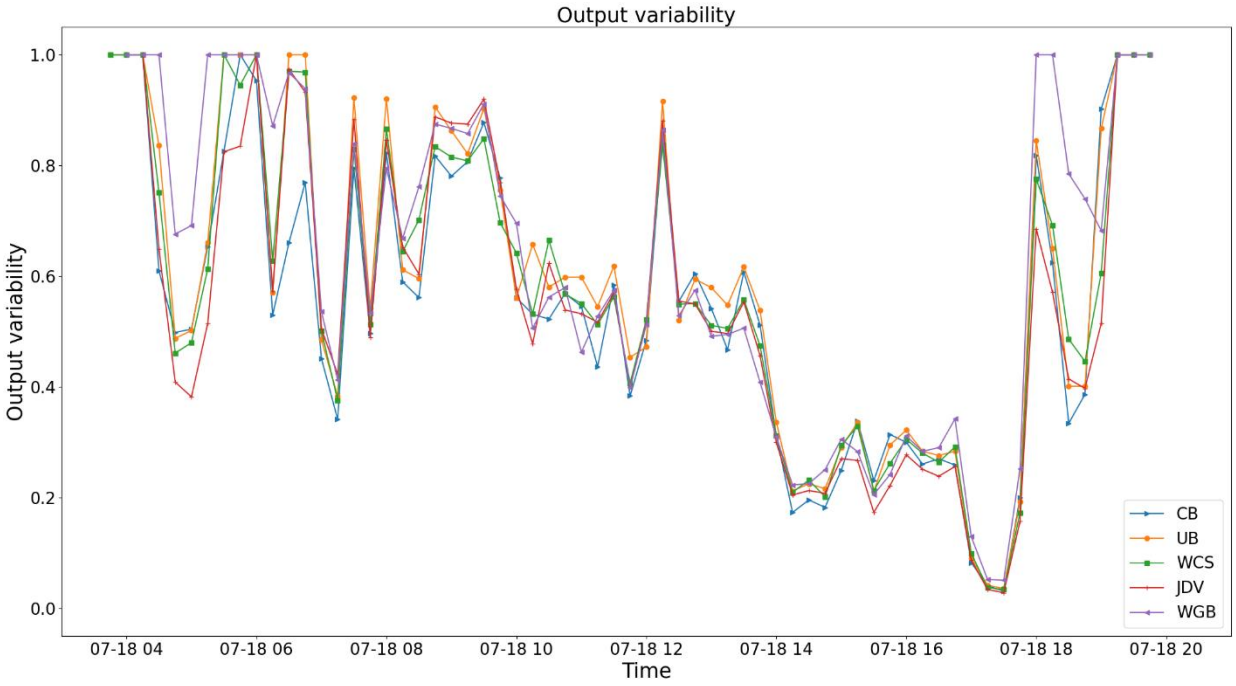

Figure 26: Output variability of the different PV systems on a summer day, 18th of July.

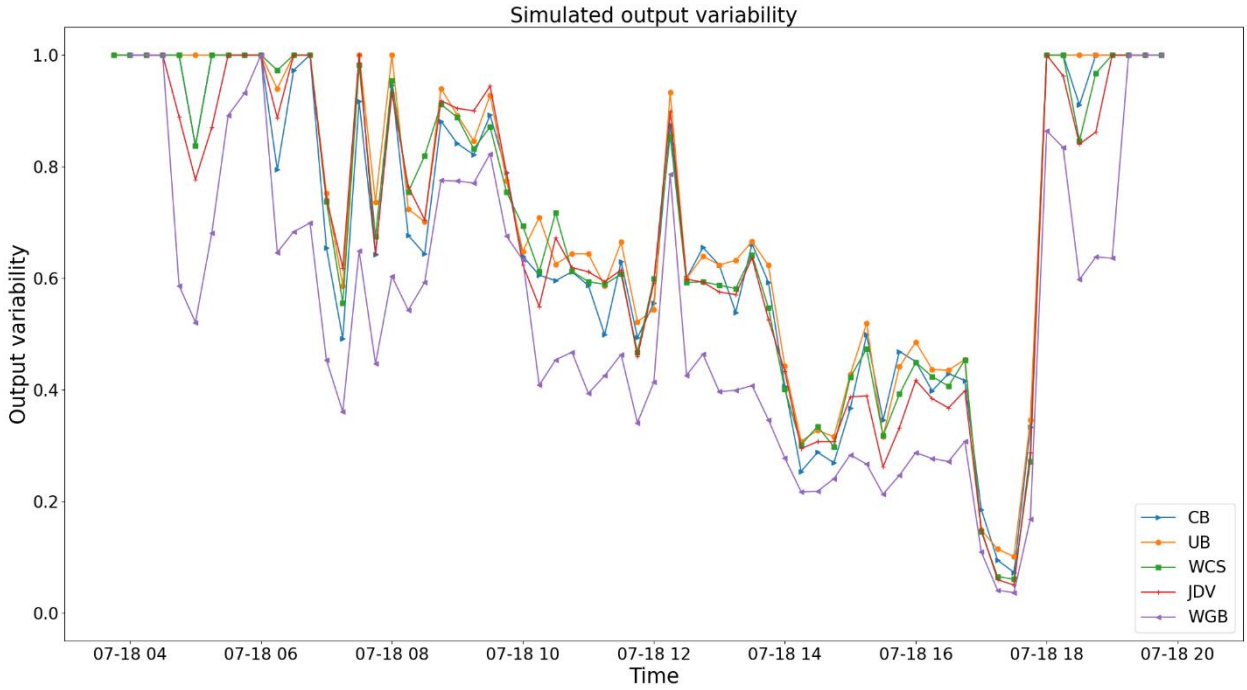

Figure 27: Simulated output variability of the different PV systems on a summer day, 18th of July.

The orientations of the PV systems with a 10-degree surface tilt have a negligible impact on the output in the summer according to Figure 22 and 23. The results that can be seen in Figure 26 and 27 are in line with this. The graphs for the different PV systems look quite similar, and the simulation doesn't change this by much, except for the WGB system. As the WGB system has a 30-degree surface tilt, the simulation effects it more than the other systems.

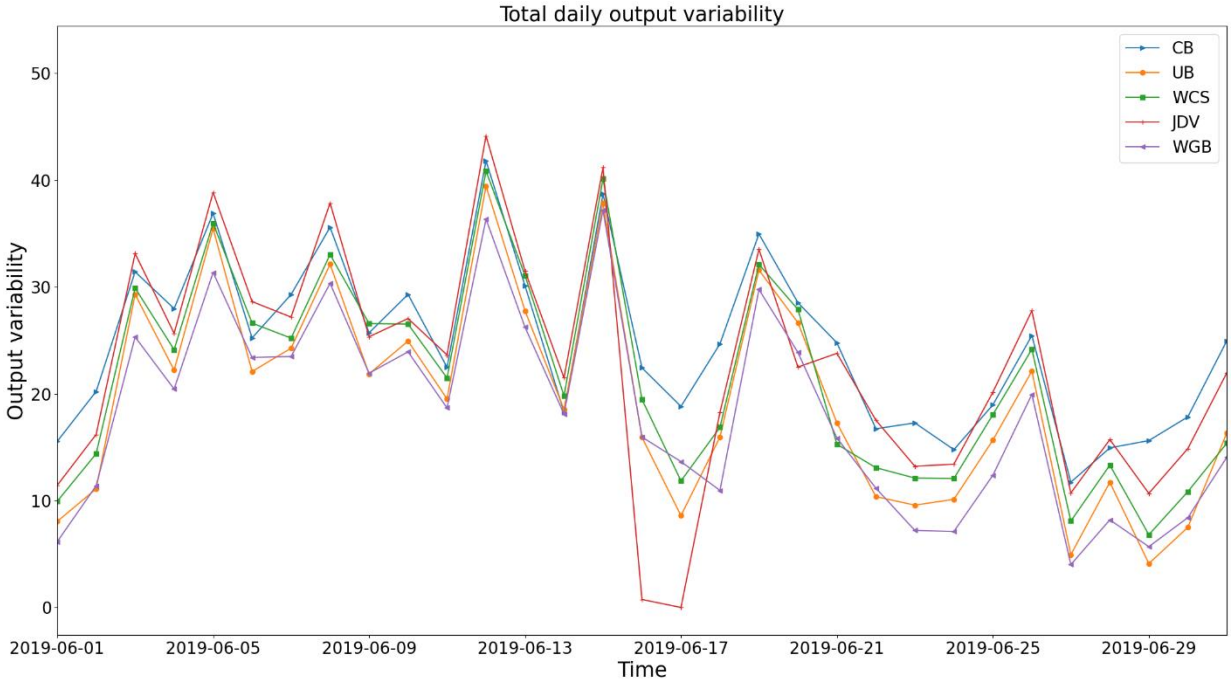

Figure 28: Total daily output variability of the different PV systems in the month June.

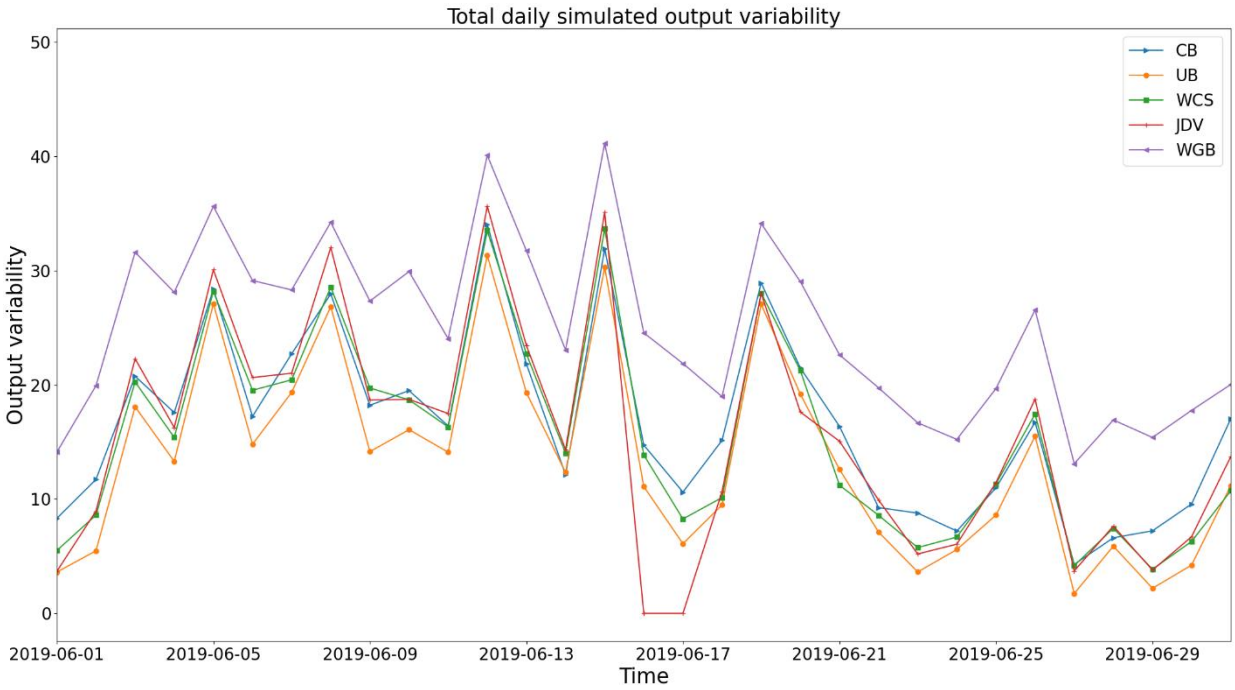

Figure 29: Total daily simulated output variability of the different PV systems in the month June.

When summing the output variability per day, the results look very similar. The simulation doesn't change the output variability by much, except for the WGB system. This was already noticeable in Figure 27, where the graph for the WGB system changed significantly compared to before the simulation.

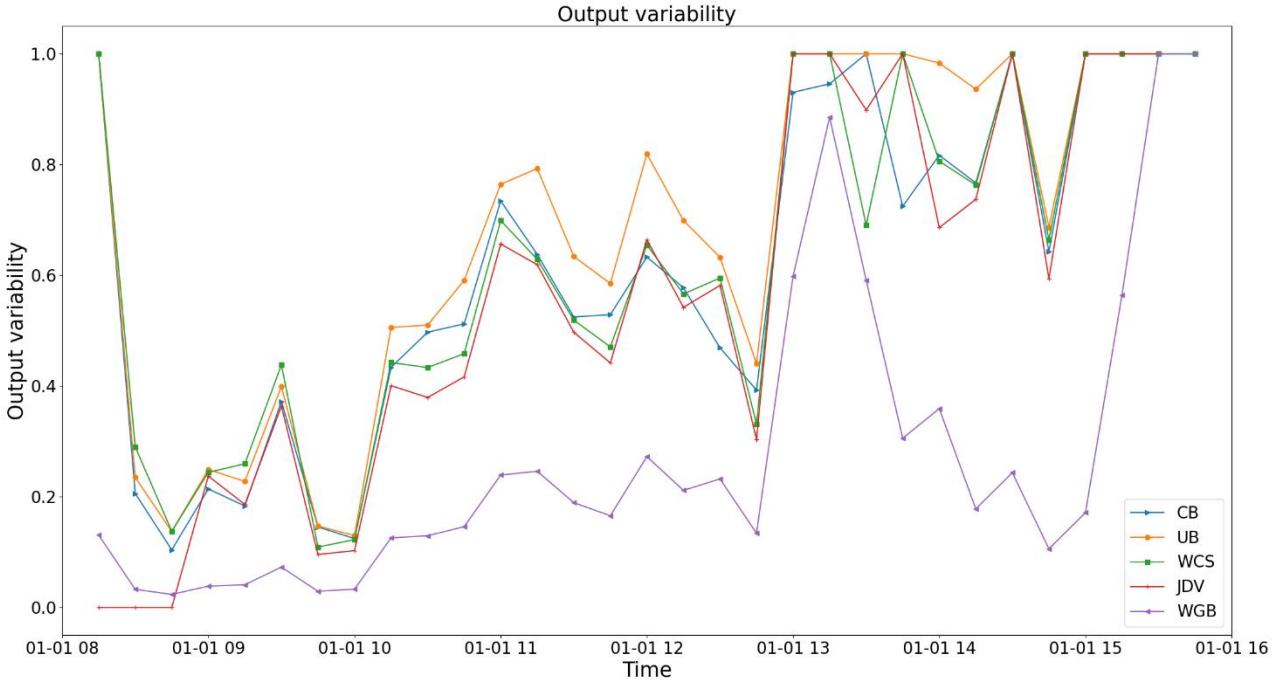

Figure 30: Output variability of the different PV systems on a winter day, 1st of January.

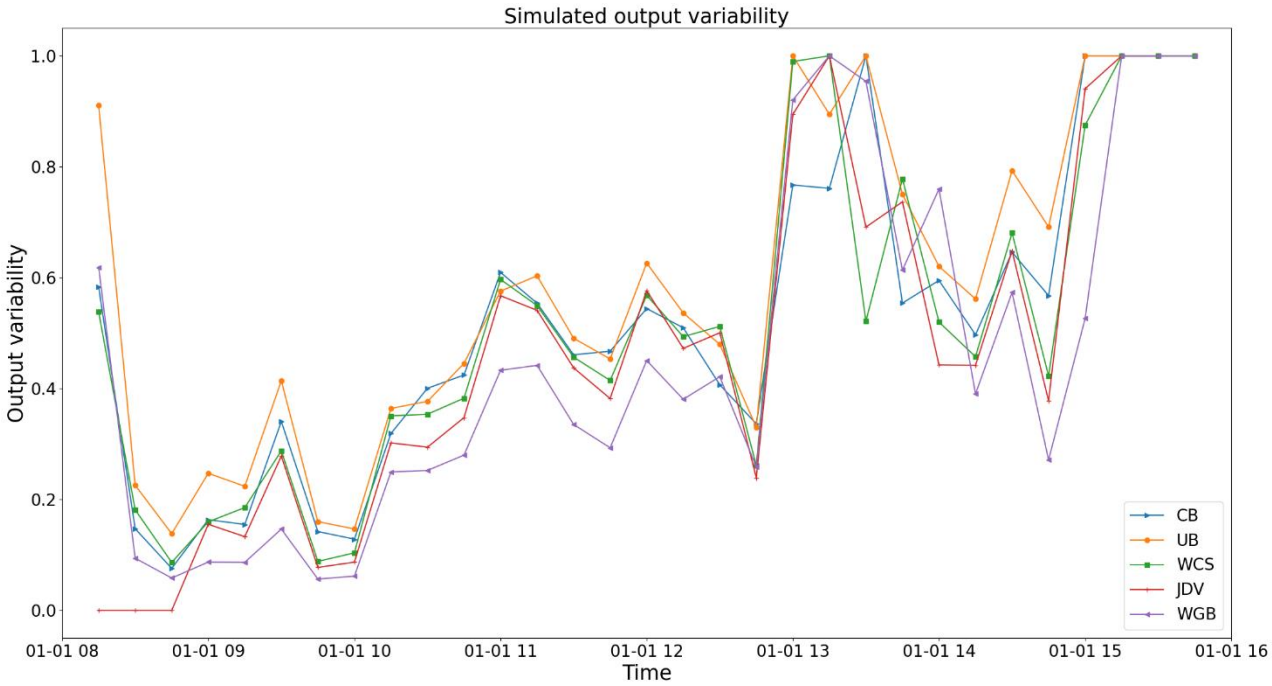

Figure 31: Simulated output variability of the different PV systems on a winter day, 1st of January.

According to Figure 24 and 25, the difference in orientations of the PV systems, even at a 10-degree surface tilt, does make a significant difference. Figure 30 and 31 are in line with this, as the graphs for the different PV systems are clearly closer together in Figure 31 than in Figure 30. Especially the CB, WCS, and JDV graphs look very similar, which can be explained by all these systems having both east and west orientated panels at a 10-degree surface tilt. The UB system has both north and south orientated panels, explaining the difference of its graph.

It is unknown why the output variability values for the WGB system are lower than the other systems, but a possible explanation of the relatively low output of the WGB system is that the WGB building is relatively low, which can result in the system being partly shaded regularly by for example surrounding trees.

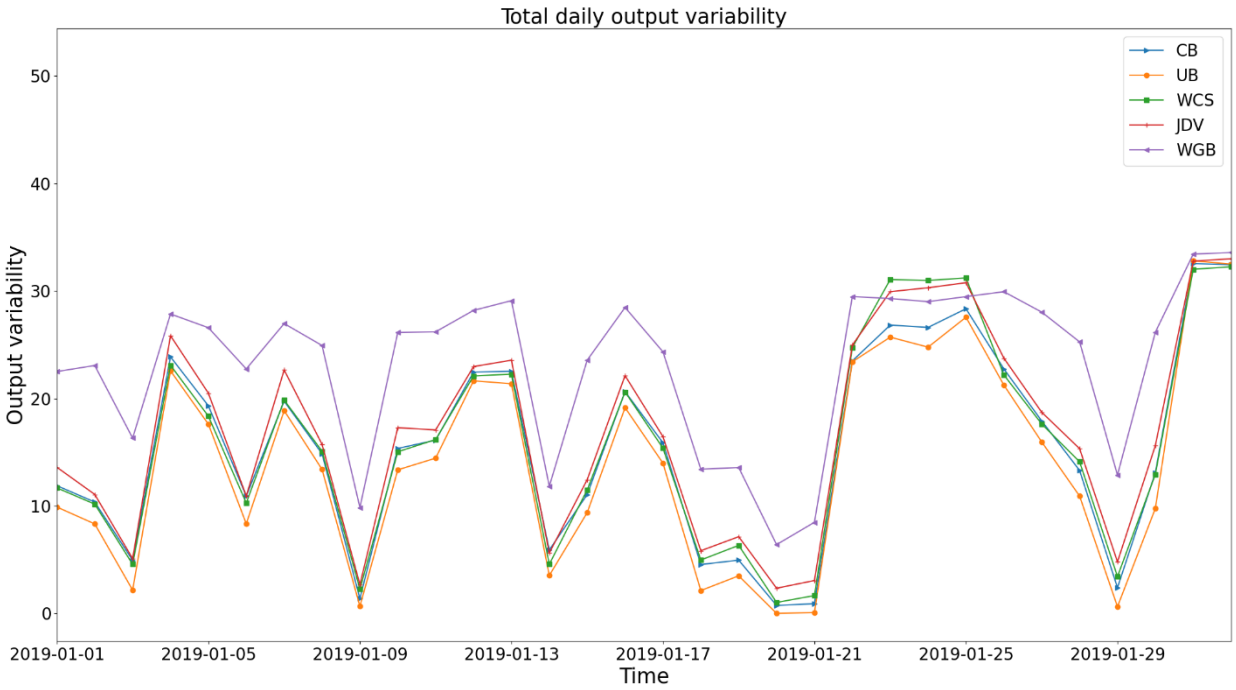

Figure 32: Total daily simulated output variability of the different PV systems in the month January.

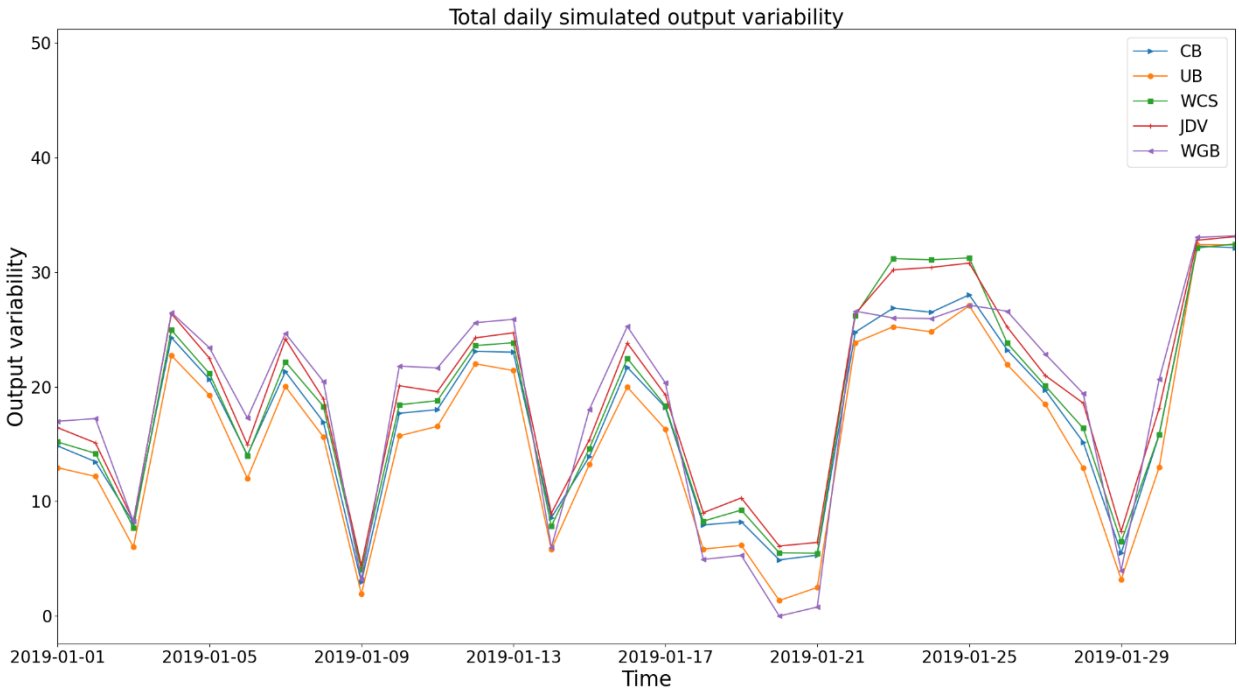

Figure 33: Total daily simulated output variability of the different PV systems in the month January.

Figure 32 and 33 further confirm what was observed in Figure 30 and 31, where the graph for the WGB system came significantly closer to the other graphs after the simulation. However, the graph for the

UB system coming closer can't be observed here, probably because the difference was already small in Figure 30 and 31, which only show one day instead of a month.

Finally, the values for all the months can be found in Table 5 and 6, clearly showing that the output variability values of the different PV systems are closer together after the simulation.

Table 5: Output variability of the different PV systems, summed per month.

Month	**CB**	**UB**	**WCS**	**JDV**	**WGB**
January	466	417	477	511	713
February	348	308	342	361	507
March	695	641	674	688	777
April	557	416	499	486	511
May	754	623	684	730	607
June	736	587	648	676	549
July	736	540	645	505	571
August	593	519	539	459	534
September	561	510	546	590	619
October	518	480	518	534	695
November	379	317	385	464	686
December	295	246	313	377	611

Table 6: Simulated output variability of the different PV systems, summed per month.

Month	**CB**	**UB**	**WCS**	**JDV**	**WGB**
January	525	472	552	580	566
February	371	329	368	387	391
March	602	548	601	619	716
April	415	313	398	388	557
May	530	439	511	527	759
June	497	406	470	463	751
July	490	366	460	351	750
August	409	383	406	346	615
September	456	414	464	502	593
October	476	434	487	503	578
November	453	388	471	539	499
December	391	327	429	483	419

3.5.2 UU dataset

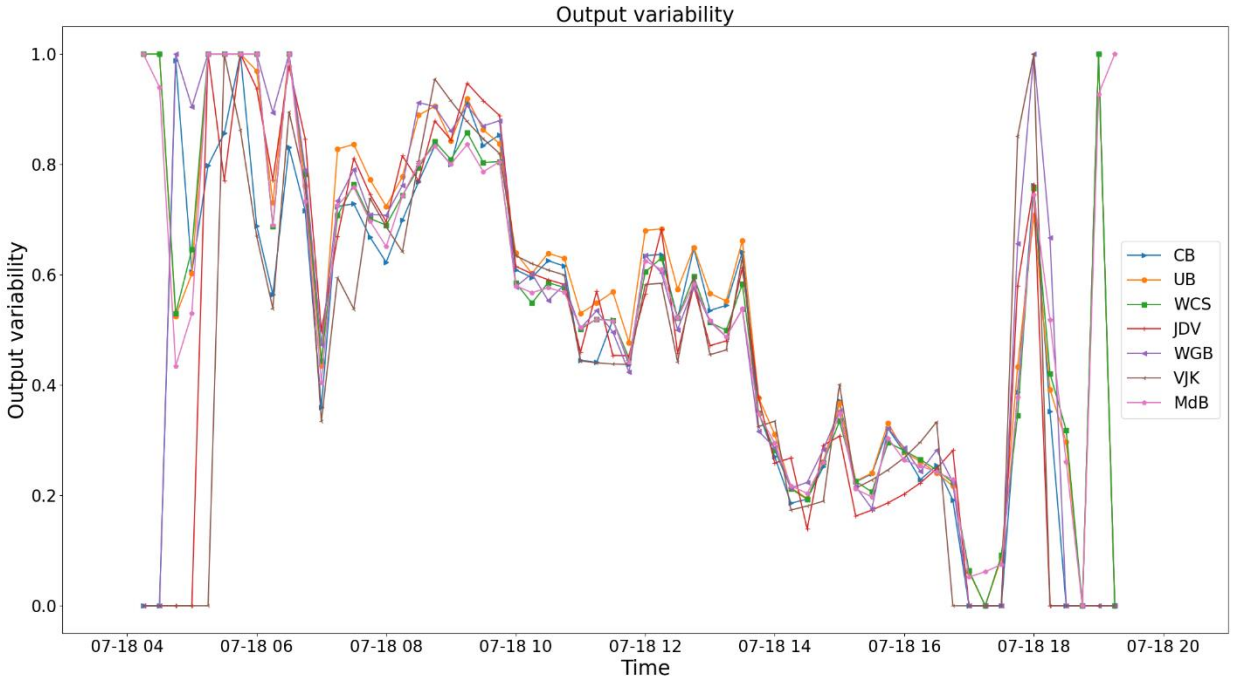

Figure 34: Output variability of the different PV systems on a summer day, 18th of July.

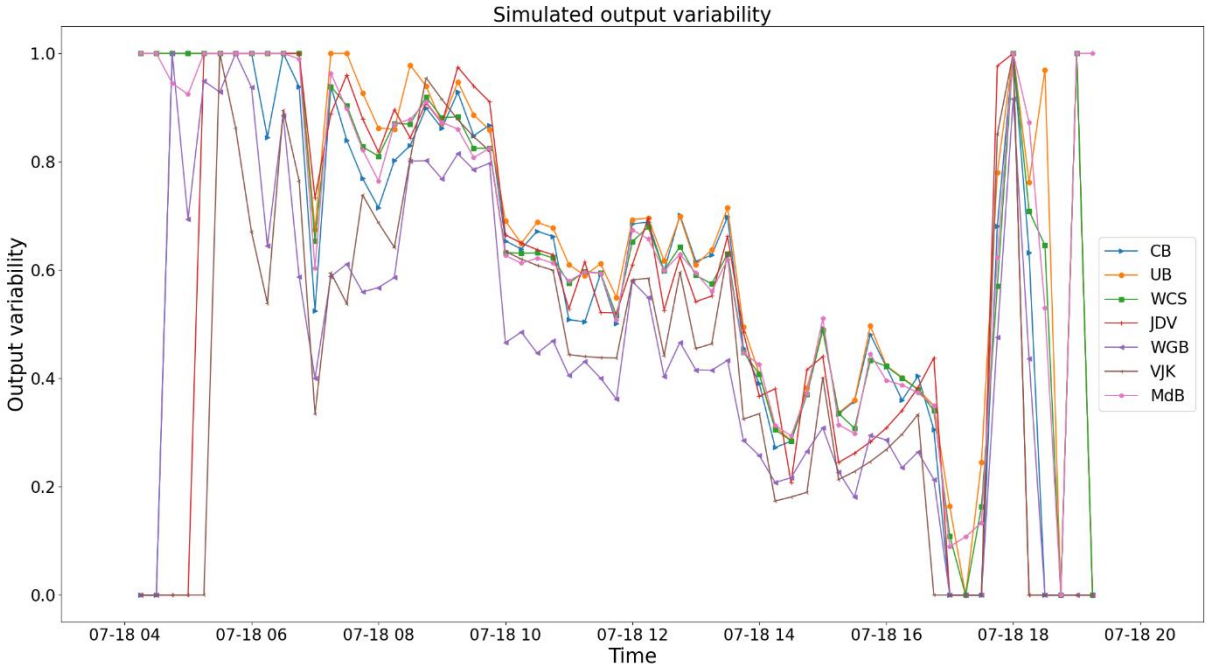

Figure 35: Simulated output variability of the different PV systems on a summer day, 18th of July.

Figure 34 and 35 show very similar results to Figure 26 and 27. The graphs for the different PV systems look similar again, with the WGB system sticking out after the simulation.

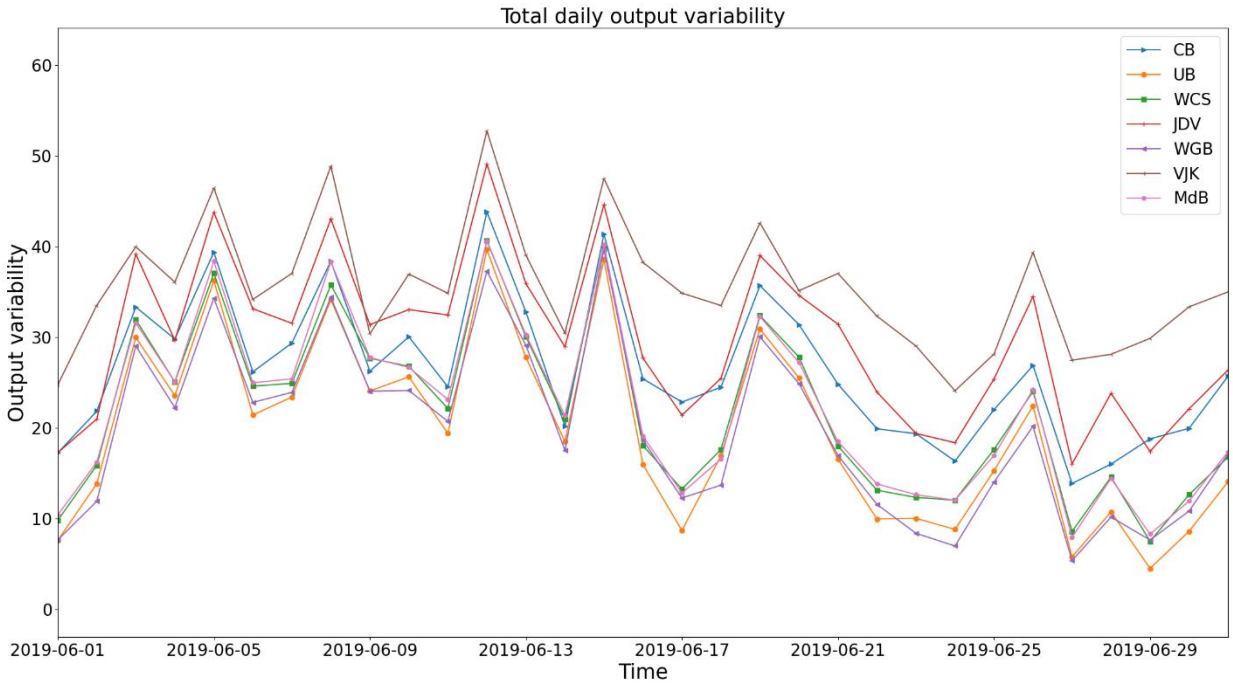

Figure 36: Total daily output variability of the different PV systems in the month June.

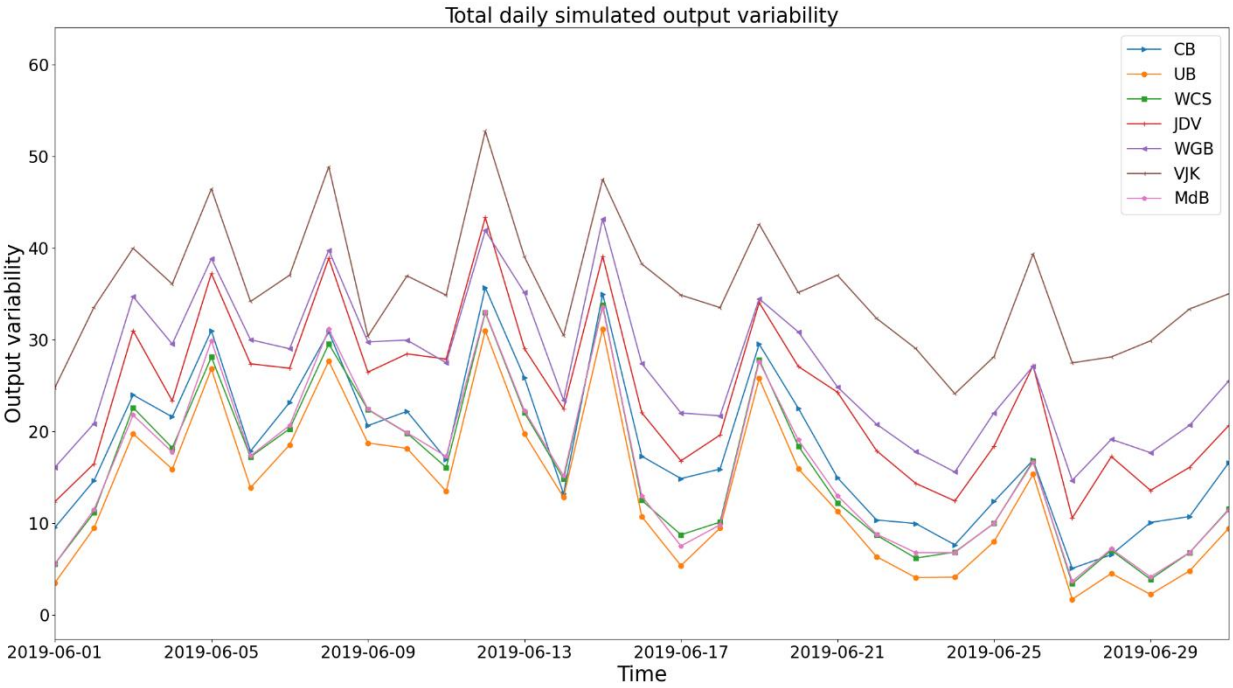

Figure 37: Total daily simulated output variability of the different PV systems in the month June.

Summing the output variability per day also results in similar looking graphs, where the graph for the WGB system clearly changed the most after the simulation.

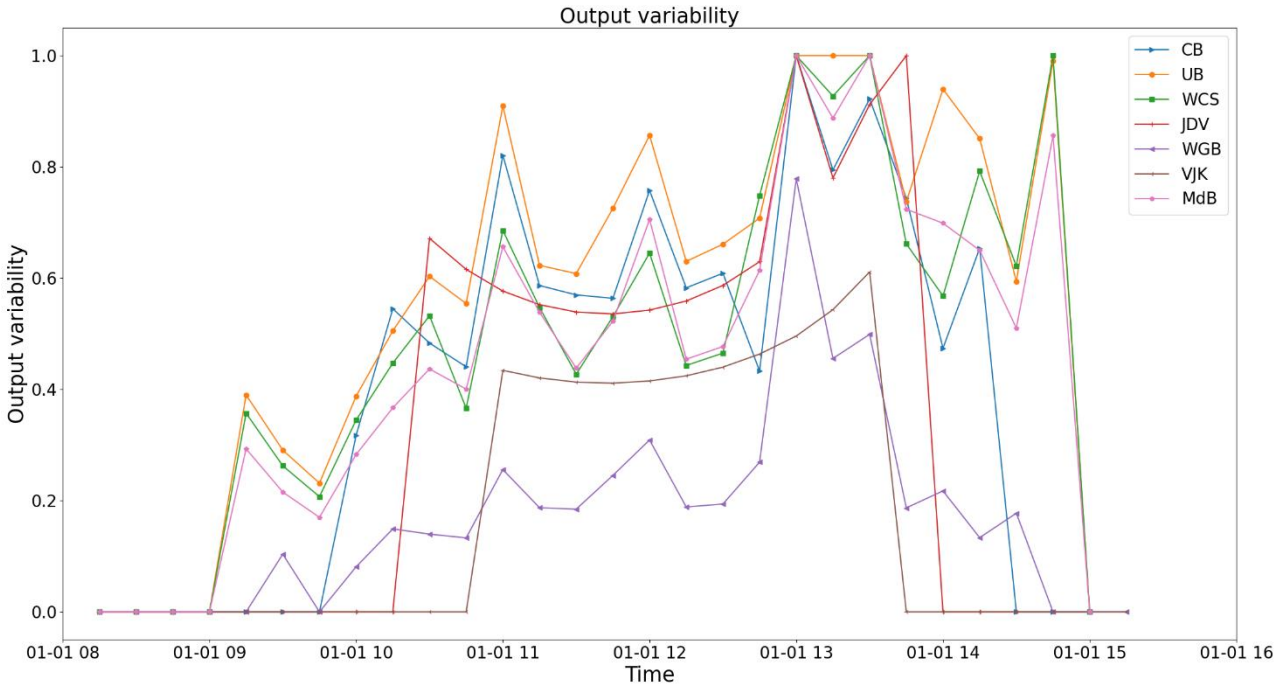

Figure 38: Output variability of the different PV systems on a winter day, 1st of January.

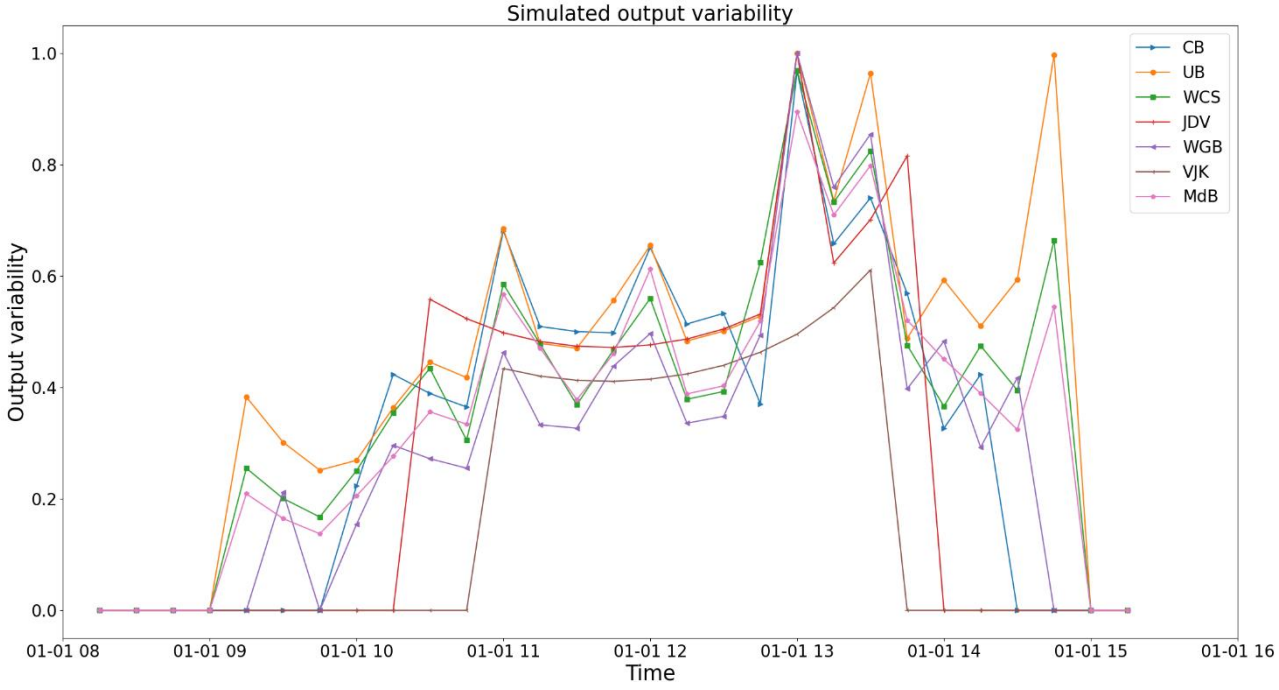

Figure 39: Simulated output variability of the different PV systems on a winter day, 1st of January.

The same applies here, as Figure 38 and 39 show similar results to Figure 30 and 31. All graphs are significantly closer to each other after the simulation, especially the graph for the WGB system.

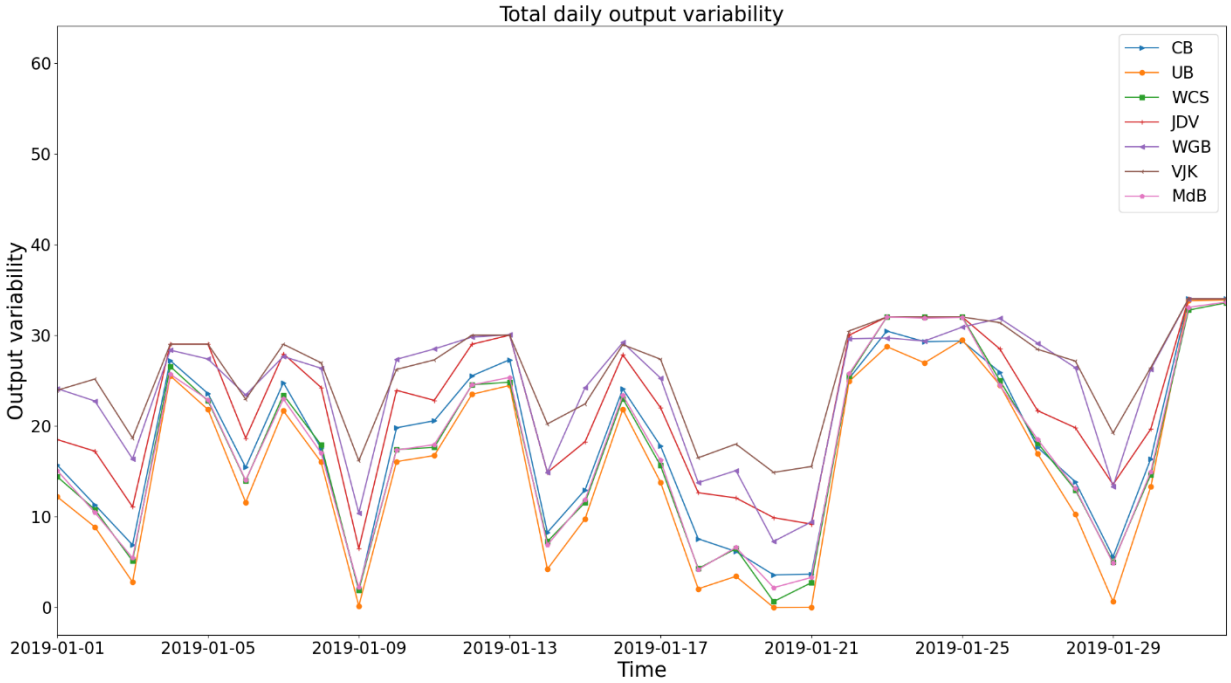

Figure 40: Total daily output variability of the different PV systems in the month January.

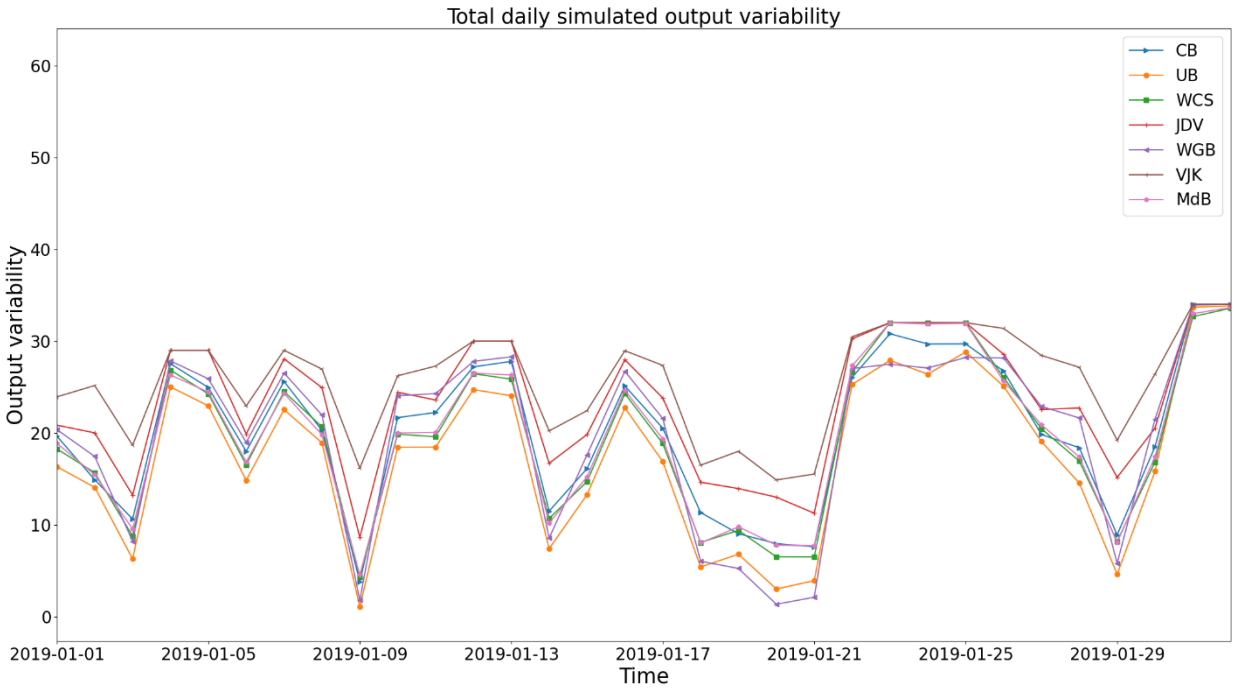

Figure 41: Total daily simulated output variability of the different PV systems in the month January.

In Figure 40 and 41, only the graph for the WGB system noticeably changes after the simulation, just like with the NetEco dataset, getting closer to the other graphs. Apart from that, Figure 40 and 41 look very similar, probably because of the small differences between Figure 38 and 39, except for the WGB system.

Finally, the values for all the months can be found in Table 7 and 8, clearly showing that the output variability values of the different PV systems are closer together after the simulation

Table 7: Output variability of the different PV systems, summed per month.

Month	**CB**	**UB**	**WCS**	**JDV**	**WGB**	**VJK**	**MdB**
January	549	466	523	678	742	791	526
February	417	325	372	504	550	663	375
March	729	646	695	799	806	871	686
April	579	451	507	608	515	744	510
May	827	650	700	883	655	1052	703
June	792	595	663	894	590	1066	669
July	785	665	664	872	612	1046	668
August	623	520	545	693	549	845	549
September	598	505	555	710	627	750	546
October	586	507	562	738	755	823	560
November	453	344	419	640	712	756	413
December	389	284	358	591	634	692	344

Table 8: Simulated output variability of the different PV systems, summed per month.

Month	**CB**	**UB**	**WCS**	**JDV**	**WGB**	**VJK**	**MdB**
January	616	528	594	713	606	791	601
February	446	353	400	512	420	663	399
March	655	561	619	742	744	871	611
April	445	344	398	525	567	744	398
May	612	468	527	735	824	1052	529
June	546	410	474	721	806	1066	480
July	546	489	471	704	810	1046	471
August	441	376	404	578	648	845	404
September	502	408	466	628	609	750	458
October	560	464	529	699	640	823	526
November	529	426	501	672	535	756	498
December	483	369	460	639	459	692	452

3.6 Correlation

Pearon's *r* is calculated for all the different pairs of PV systems, based on the output variability. Output variability values higher than one are ignored for this calculation, as they are assumed to be inaccurate, as explained earlier. These values aren't replaced by a value of one, as this influences the correlation coefficient significantly.

3.6.1 NetEco dataset

Table 9: Pearson correlation coefficients of the output variability between the different pairs of PV systems.

	CB	**UB**	**WCS**	**JDV**	**WGB**
CB	1	0.912	0.892	0.885	0.779
UB	0.912	1	0.929	0.919	0.789
WCS	0.892	0.929	1	0.963	0.846
JDV	0.885	0.919	0.963	1	0.853
WGB	0.779	0.789	0.846	0.853	1

Table 9 shows that the correlation between WCS and JDV has the highest correlation of all the system pairs. This can be explained by these systems having the same surface tilt, orientations, and ratio between different orientations. Additionally, these buildings are next to each other, increasing the correlation further.

WGB has the lowest correlation with the other systems, as its surface tilt is 30 degrees instead of 10 degrees, which has shown to make a significant difference in Figure 20 through 23.

Table 10: Pearson correlation coefficients of the simulated output variability between the different pairs of PV systems.

	CB	**UB**	**WCS**	**JDV**	**WGB**
CB	1	0.937	0.925	0.904	0.865
UB	0.937	1	0.931	0.902	0.862
WCS	0.925	0.931	1	0.963	0.903
JDV	0.904	0.902	0.963	1	0.885
WGB	0.865	0.862	0.903	0.885	1

After the simulation, all the correlation coefficients have increased as expected. WGB still has the lowest correlation coefficients, which could be explained by the WGB building being the furthest away from the CB and UB buildings.

3.6.2 UU dataset

Table 11: Pearson correlation coefficients of the output variability between the different pairs of PV systems.

	CB	**UB**	**WCS**	**JDV**	**WGB**	**VJK**	**MdB**
CB	1	0.896	0.898	0.848	0.848	0.842	0.899
UB	0.896	1	0.950	0.833	0.830	0.777	0.942
WCS	0.898	0.950	1	0.859	0.885	0.789	0.977
JDV	0.848	0.833	0.859	1	0.842	0.863	0.862
WGB	0.848	0.830	0.885	0.842	1	0.806	0.891
VJK	0.842	0.777	0.789	0.863	0.806	1	0.791
MdB	0.899	0.942	0.977	0.862	0.891	0.791	1

Table 11 shows that the correlation between the MdB and WCS systems is the highest, nearing a correlation coefficient of 1. Like the WCS and JDV pair, the MdB and WCS systems have the same surface tilt, orientations, and ratio between orientations. Also, the buildings are next to each other.

The correlation between the WCS and JDV systems is significantly lower than in the NetEco dataset, which is likely the result of the inaccuracy of the UU dataset, as it doesn't use decimal values. This theory is further confirmed by the correlation coefficients being the lowest for the JDV, WGB, and VJK systems. The values for the WGB system can again be explained by the surface tilt and location, but this is not the case for the JDV and VJK systems. As the JDV and VJK systems are much smaller than the other systems, the absence of decimal values has the biggest impact on the accuracy of the data for these systems.

Table 12: Pearson correlation coefficients of the simulated output variability between the different pairs of PV systems.

	CB	UB	WCS	JDV	WGB	VJK	MdB
CB	1	0.911	0.937	0.854	0.898	0.859	0.934
UB	0.911	1	0.944	0.806	0.877	0.791	0.934
WCS	0.937	0.944	1	0.876	0.914	0.843	0.981
JDV	0.854	0.806	0.876	1	0.827	0.876	0.880
WGB	0.898	0.877	0.914	0.827	1	0.833	0.913
VJK	0.859	0.791	0.843	0.876	0.833	1	0.845
MdB	0.934	0.934	0.981	0.880	0.913	0.845	1

Again, all correlation coefficients have increased as expected, except from some pairs, mainly involving the JDV system. The JDV, WGB, and VJK systems still have the lowest correlation coefficients, where the same reasoning applies as before.

4 Conclusion & discussion

This research has shown how the PV systems differ from each other regarding output variability. It's clear that even at a low surface tilt of 10 degrees that the output variability differs significantly in winter between the systems with a different layout regarding orientations, which was further confirmed by doing a simulation. The WGB system with its 30-degree surface tilt differs even more, as expected. In summer this is not the case, where the output variability graphs for the different PV systems look very similar.

The PV systems all strongly correlate with each other, although not as strongly as values found in a previous research by M. Bruinewoud [5]. He found correlation coefficients between 0.984 and 0.999 with 15-minute time resolution data and between 0.928 and 0.991 with 5-minute time resolution data for all the different pairs of PV systems, also using the NetEco dataset. The correlations he found were significantly higher for the 15-minute time resolution data, which could be caused by the different time resolution, or by the 5-minute data only stretching from December until March, which are months where there is more of a difference between the PV systems regarding output variability, as shown in this paper. The difference between the correlation values he found with the 15-minute time resolution data and the correlation values found in this research, can be explained by the fact that he calculated these values for the power series instead of the output variability. The differences in these values occur as a result of the different ratios between the inverter capacity and the capacity of the system of the different PV systems. As the performance ratios changes when the percentage of output power changes, this inverter to system ratio influences the output variability.

As UPOT data was not available for 2019, it couldn't be used for this research. The KNMI data from De Bilt was the second-best option after this, but because of the distance between De Bilt and the Utrecht Science Park, the data was not completely compatible with the PV output data. This occasionally led to unrealistic values when the performance ratios were calculated, mostly present at times when the sun came up or went down. Apart from this, the data was still usable as most values were logical.

The UU dataset was a useful addition to the research as it showed how a different type of logging can influence the results. However, due to the UU dataset not using decimal values, the accuracy of the dataset could have been better. This was especially noticeable for the relatively small PV systems JDV and VJK, which makes the results regarding these two systems less reliable.

The use of a reference module and inverter for the calculation of the theoretical output makes this theoretical output not fully accurate. However, as all the different PV systems are analyzed with this theoretical output, it is assumed to be sufficiently accurate for the comparison purposes that have been done in this research.

For further research, it would be interesting to see an analysis that is done with data that has a time resolution of 5 or less minutes to see how this impacts the results. Furthermore, research into the forecasting of the output of PV systems is needed, as this could ultimately be part of the solution of decreasing the burden on the electrical grid with the increasing amount of installed PV capacity.

5 Acknowledgements

I would like to thank my supervisor Wilfried van Sark for giving me the opportunity to work on this project and connecting me with the right people to help me throughout this research. Thanks to Odysseas Tsafarakis for providing me with regular supervision, mostly helping with the programming part of this research. Thanks to Lennard Visser for providing different ideas regarding methods that I could use during this research. Finally, thanks to Lex Schiebaan and Myria Ioannou for providing me with the NetEco and UU datasets respectively.

6 References

[1] CBS, “Hernieuwbare elektriciteit; productie en vermogen,” 2020. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82610NED/table?ts=1606480145574 (accessed Nov. 27, 2020).

[2] W. Georg *et al.*, “Existing and future PV prosumer concepts,” no. 764786, pp. 1–123, 2018, [Online]. Available: https://www.pvp4grid.eu/wp-content/uploads/2018/08/D2.1_Existing-future-prosumer-concepts_PVP4Grid_FV.pdf.

[3] R. Luthander, J. Widén, D. Nilsson, and J. Palm, “Photovoltaic self-consumption in buildings: A review,” *Appl. Energy*, vol. 142, pp. 80–94, 2015, doi: 10.1016/j.apenergy.2014.12.028.

[4] R. Perez *et al.*, “Spatial and Temporal Variability of Solar Energy,” *Found. Trends® Renew. Energy*, vol. 1, no. 1, pp. 1–44, 2016, doi: 10.1561/2700000006.

[5] M. Bruinewoud, “Assessing the performance of the Utrecht Science Park PV system,” 2018.

[6] W. van Sark *et al.*, “ENERGY PERFORMANCE OF A 1.2 MWP PHOTOVOLTAIC SYSTEM DISTRIBUTED OVER EIGHT BUILDINGS AT UTRECHT UNIVERSITY CAMPUS,” *Proc. 33rd Eur. Photovolt. Sol. Energy Conf.*, pp. 2284–2287, 2017.

[7] T. E. Hoff and R. Perez, “Quantifying PV power Output Variability,” *Sol. Energy*, vol. 84, no. 10, pp. 1782–1793, 2010, doi: 10.1016/j.solener.2010.07.003.

[8] M. Jamaly and J. Kleissl, “Spatiotemporal interpolation and forecast of irradiance data using Kriging,” *Sol. Energy*, vol. 158, no. February, pp. 407–423, 2017, doi: 10.1016/j.solener.2017.09.057.

[9] M. Schlather, “Some covariance models based on normal scale mixtures,” *Bernoulli*, vol. 16, no. 3, pp. 780–797, 2010, doi: 10.3150/09-BEJ226.

[10] W. McKinney, “Data Structures for Statistical Computing in Python,” *Proc. 9th Python Sci. Conf.*, vol. 1697900, no. Scipy, pp. 51–56, 2010, [Online]. Available: http://conference.scipy.org/proceedings/scipy2010/mckinney.html.

[11] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, “Pvlib Python: a Python Package for Modeling Solar Energy Systems,” *J. Open Source Softw.*, vol. 3, no. 29, p. 884, 2018, doi: 10.21105/joss.00884.

[12] J. D. Hunter, “Matplotlib: A 2D Graphics Environment,” *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007.

[13] T. E. Oliphant, “Python for Scientific Computing,” *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 10–20, 2007.

[14] P. Virtanen *et al.*, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nat. Methods*, no. 17(3), pp. 261–272, 2020.

[15] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, “Scikit-learn,” *GetMobile Mob. Comput. Commun.*, vol. 19, no. 1, pp. 29–33, 2015, doi: 10.1145/2786984.2786995.

[16] KNMI, “Uurgegevens van het weer in Nederland - Download,” 2020. http://projects.knmi.nl/klimatologie/uurgegevens/selectie.cgi (accessed Jun. 15, 2020).

[17] Sun&wind energy, “5-level topology – the next logical step for inverters,” 2017.

https://www.sunwindenergy.com/content/5-level-topology-next-logical-step-inverters (accessed Nov. 27, 2020).

[18] M. Chaâbane, M. Masmoudi, and K. Medhioub, “Determination of Linke turbidity factor from solar radiation measurement in northern Tunisia,” *Renew. Energy*, vol. 29, no. 13, pp. 2065–2076, 2004, doi: 10.1016/j.renene.2004.03.002.

[19] The SciPy community, “scipy.stats.pearsonr,” 2020. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html (accessed Oct. 23, 2020).