SiliconPV 2024, 14th International Conference on Crystalline Silicon Photovoltaics
Energy Yield, Reliability & Cost
https://doi.org/10.52825/siliconpv.v2i.1296

Published: 09 Jan. 2025

Improving the Accuracy of PV Yield Calculation by Exploitation of Real Weather Data

Andreas Schneider [1], Matthias Buschmann [1], Jorge Rabanal Arabach [2,3], Julia Chochollek [1], and Thomas Nierhoff [1]

[1] University of Applied Sciences Gelsenkirchen, Germany

[2] Universidad de Antofagasta, Chile

[3] ATAMOSTEC, Chile

*Correspondence: Andreas Schneider, andreas.schneider@w-hs.de

Abstract. The precision of yield calculation of modern design and simulation software for photovoltaic systems strongly rely, beside the accuracy of the specified module and inverter data, on the quality of the weather data. Since data from weather stations is not available for most locations world-wide this data is calculated by using modern interpolation methods. Beside this, simulation software typically uses historical weather data. In this work the mismatch of yield simulation results based on interpolated or also called synthetical data, and data coming from a weather station in proximity to the installation is evaluated. The simulated data sets are compared to measurement data as obtained by the inverter output and hence give a profound understanding how interpolated data may influence the simulation results. The outcome shows that the averaged differences between simulation and measurement are decreased by a factor of up to four if on-site weather data is used as input for the simulation instead of synthetic data, reaffirming the hypothesis of this work. The largest source of deviation is irradiation, which varies up to 10% if synthetical and measured irradiation on-site is compared. The second largest sources for simulation mismatches are power calculation and module temperature correction.

Keywords: Photovoltaic Modules, Temperature Coefficients, Field Measurement

1. Introduction

Popular design and simulation software for photovoltaic systems like PV*SOL from VALENTIN SOFTWARE offer various ways on how to import weather data. A popular source for weather data is Meteonorm which provides mean values for GHI and DNI for the period of 1996 to 2015 and more current data for temperature, wind, precipitation, etc. for the years of 2000-2019. Recently, since Meteonorm version 8 was published, it is also possible to use climatic data for distinct and more recent years. The data originates from thousands of weather stations world-wide and is supported by satellite measurement and further aerosol measurement data. For many major cities globally, this data directly represents the climatic situation at the origin. Nonetheless, for most locations the local weather data needs to be calculated by means of advanced interpolation models and is therefore referred to synthetical data. Any interpolation method, even advanced ones, will inevitably lead to uncertainties which impact the accuracy of the outcome of the yield simulation [1], [2]. Beside this, even for major locations with available climatic data the use of long-term time series automatically leads to deviations in the simulation results, specifically for recent years in which the effect of global warming on local weather situations become more pronounced.

This paper presents a detailed study on the differences in yield simulation results of PV plants. Third-party climatic data based on the interpolation method is compared to results for which actual measurement data from a weather station in close proximity to the PV installations is used. The main idea is to compare the simulation results with the measurement data of the individual solar installations, this way taking the measurement data as a reference. The simulation is carried out with the software PV*SOL. Results are presented for seven PV installations which are hooked up to the grid since 2011 at the location of Gelsenkirchen, Germany and one PV installation for the location of Antofagasta, Chile, which is installed at lat. −25.9343295, long. −71.93716 and operates since 2019. The term climatic data is in the following replaced by the term weather data since on-site and hence local environmental conditions are compared.

2. Methodology

A weather station, in close proximity to the solar panels, measures humidity, ambient temperature, wind speed, wind direction and horizontal as well as in-plane irradiation every five minutes at the location of Gelsenkirchen and every minute at the location of Antofagasta. All data is stored in long-term data bases. The validity of the measurement data is verified by regular internal round robin tests. Long term measurement data for seven selected installations out of 16 available for the year 2021 is used for the location of Gelsenkirchen. In Antofagasta four smaller PV plants are operating since 2019. For this study power data from a mono-crystalline PERC module plant of the year 2023 is used. Current, voltage and power data are provided by the inverter with a time resolution of five minutes. Table 1 shows the details on the selected installations for both locations.

Table 1. *Details on the selected installations*

Module type	Location	Cell technology	Number Modules	Peak Power (kW)
Sanyo 240 Wp	Ge, Germany	HIT	19	4.6
GET 85 Wp	Ge, Germany	Amorphous Silicon	80	6.8
Auria 115 Wp	Ge, Germany	Amorphous Silicon	60	6.9
PeakON 220 Wp	Ge, Germany	Poly crystalline	10	2.2
Scheuten 245 Wp	Ge, Germany	Poly crystalline	30	7.4
SolonBlue 230 Wp	Ge, Germany	Mono crystalline	30	6.9
SolonBlack 280 Wp	Ge, Germany	Mono crystalline	24	6.7
JA 345 Wp	An, Chile	Mono crystalline PERC	4	1.3

A 3D-model for the existing eight installations was recreated inside the simulation software PV*SOL supported by satellite imaging data, allowing for a very realistic model of the existing installation, such as building height, inclination, and orientation angle. Furthermore, were the specific electrical module and inverter specification inside the simulation tool selected, allowing for a calculation based on the original manufacturer data. To consider module degradation an in depth-study of the degradation was initially performed using long-term measurement data (>10 years), verified by measurements with an in-house solar simulator.

The fact that the inverters at the location of Gelsenkirchen are operating since 2011 requires a verification of the measurement accuracy of the internal power measurement of the SMA webbox which is connected to the SMA inverter. For this a calibrated Fluke 43 Power Quality Analyzer was installed over the course of one day to measure the power data of the

Sanyo HIT PV plant. In parallel the webbox acquired the electrical plant data. Initially the internal timers of both measurement devices were synchronized. Figure 1 shows the measurement data for both devices. Over the course of the day a total deviation in the calculated energy of 1.6% can be stated, which is, taking the operating time of the SMA inverter and webbox into account a very good result.

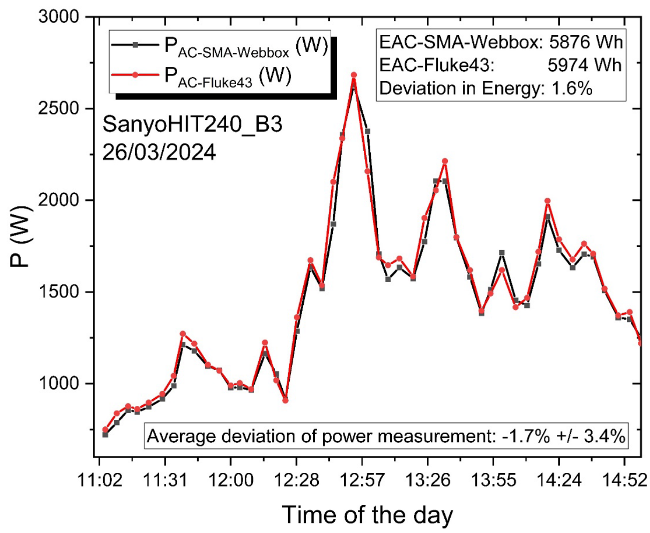

Figure 1. *Measured power data over the course of a partly casted day at the Sanyo HIT installation. The red line shows the Fluke 43 and the black one the webbox data.*

3. Results

The relative difference of yield calculations and measurement data for all installations as shown in Table 1 is calculated and compared. Figure 2 shows exemplarily the monthly relative differences of yield calculation and measurement data by a) importing weather data from the weather station in proximity to the installation (in darkseagreen) and b) taking the synthetical weather data (in burlywood). At Figure 2 (left) the 4.6 kWp Sanyo installation in Germany for the year of 2021 is displayed. The difference between the monthly simulation results in comparison to the measurement data is of a factor up to four higher for the simulation based on synthetical weather data. The annual difference improves from −13.4 to −9.3% for the simulation based on on-site weather data.

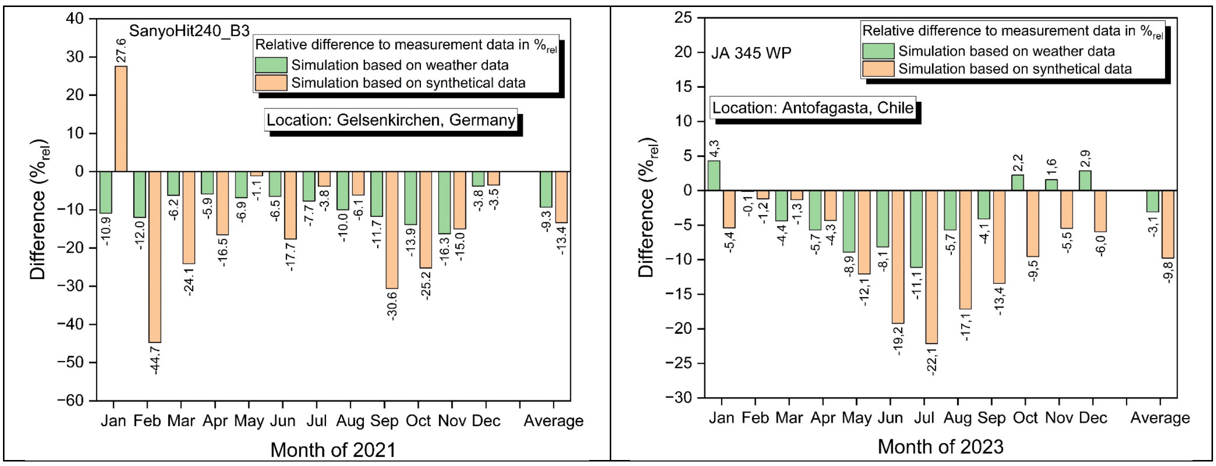

Figure 2. *Left: relative difference of yield calculation and measurement data for Gelsenkirchen (year of 2021) and right: for Antofagasta for the year of 2023.*

Figure 2 (right) displays the relative difference of yield calculations and measurement data for the 1.3 kWp JA installation in Chile for the year of 2023, showing that the annual difference improves from −9.8 to −3.1% for the simulation based on on-site weather data. Both examples show how yield simulation results are improved by importing more accurate and matching weather data.

Depending on the months of the year the deviation can be positive or negative. Interestingly to mention that for the month of February in Figure 2 (left) and even for four months in Figure 2 (right) the deviation for all installations as documented in Table 1 is highest for the synthetical weather data set. Since the monthly average can be positive or negative the average of the differences (Eq. 1) and the average of the absolute differences (Eq. 2) are calculated. On such equations, D_S and D_W are the average of the relative difference calculated over N datapoints; the former is between the simulated yield based on synthetic ($Y_{AC,S}$) and measured PV system yield (Y_{AC}), and the latter between weather data ($Y_{AC,W}$) and Y_{AC}. The subscript "abs" indicates that the average is based on the absolute value of the relative difference.

$$D_{\mathrm{S}} = \frac{1}{N}\sum_n^N \left(\frac{Y_{\mathrm{AC,S}_n}}{Y_{\mathrm{AC}_n}} - 1\right); \quad D_{\mathrm{W}} = \frac{1}{N}\sum_n^N \left(\frac{Y_{\mathrm{AC,W}_n}}{Y_{\mathrm{AC}_n}} - 1\right) \tag{1}$$

$$D_{\mathrm{S,abs}} = \frac{1}{N}\sum_n^N \left(\left|\frac{Y_{\mathrm{AC,S}_n}}{Y_{\mathrm{AC}_n}} - 1\right|\right); \quad D_{\mathrm{W,abs}} = \frac{1}{N}\sum_n^N \left(\left|\frac{Y_{\mathrm{AC,W}_n}}{Y_{\mathrm{AC}_n}} - 1\right|\right) \tag{2}$$

Table 2 shows the averaged differences between simulation and measurement for each of the eight installations over all months for the specific year. Results clearly demonstrate that the use of synthetically generated weather data leads to far less precise yield calculation results.

Table 2. *Average of the monthly differences and of the absolute monthly differences in yield calculations for the specific year*

Module type	**Location**	D_S	D_W	$D_{S,abs}$	$D_{W,abs}$
Sanyo 240 Wp	Germany	−13.4%	−9.3%	18.0%	9.3%
GET 85 Wp	Germany	−4.6%	−0.6%	14.6%	3.9%
Auria 115 Wp	Germany	10.6%	15.3%	20.2%	15.3%
PeakON 220 Wp	Germany	−12.9%	−8.7%	17.8%	8.7%
Scheuten 245 Wp	Germany	−8.3%	−4.1%	18.3%	7.9%
SolonBlue 230 Wp	Germany	−0.8%	3.9%	16.3%	4.1%
SolonBlack 280 Wp	Germany	−22.6%	−17.6%	23.7%	17.6%
JA 345 Wp	Chile	−9.8%	−3.1%	9.8%	4.9%

4. Statistical evaluation

4.1 Evaluation of potential deviations

Table 3 lists the main factors influencing the accuracy of the yield calculation by the simulation software, differentiated into weather data, mathematical model, and specific plant information. Important to note that the list is by far not complete, beside this, many factors are influencing each other. For weather data solar irradiance and wind speed have a significant effect on the power calculation. Furthermore, the wind speed is calculated based on the logarithmic wind profile formula, which primarily uses wind data measured at a standard height of 10 m. Together with the so-called roughness length which can only be a guess for the individual sites the wind speed is calculated for the specific settings. Critical within the mathematical model of the simulation software is the underlying algorithm for the determination of the module temperature. Various approaches exist for which many rely on the exact knowledge of physical parameters including the module and the installation [3], [4]. Module power and degradation was precisely determined by a lab measurement, module inclination and installation direction are well known. Therefore, the main contributing factor is the temperature coefficient of the modules which are used as originally given by the manufacturer and comes with a larger uncertainty.

Table 3. *Main factors influencing the accuracy of the yield calculation (↑ high impact; → medium impact; ↓ small impact)*

Weather data		Mathematical Model		Specific plant details	
Solar irradiance	↑	Module temperature	↑	Inclination and direction	↓
Solar spectra	→	Degradation	→	Module power at STC	↓
Ambient temperature	→	Module Power	↓	Temperature coefficients	↑
Wind speed	↑			Inverter efficiency	→

To account for the above listed sources impacting the accuracy of the simulated yield a deeper study on the high impact risks is performed. The only risk which cannot be assessed is wind speed because this data is not available from the simulation software. For the in-depth study of chapter 4.3 (second part) and chapter 4.4 the Sanyo HIT plant was used exemplarily.

4.2 Temperature coefficients and power calculation

To verify the validity of the underlying temperature coefficients and at the same time the accuracy of the mathematical model for calculating the power, the 4.6 kWp Sanyo HIT plant in Germany and the JA plant in Chile are exemplarily used. The simulated output power based on both weather data sets for the whole year (measured and synthetical) was filtered inside a narrow irradiance intensity band of (1000±50) W/m^2 at a simulated module temperature of (50±1) °C. The resulting power data set is averaged, and temperature corrected to 25 °C by using the temperature coefficient as originally given by the manufacturer. Additionally, the well-known power degradation for the Sanyo HIT modules, backed upon indoor measurements, is taking into account. Based on the lab measurement and hence after deduction of degradation the current power of the Sanyo HIT plant is 4.3 kWp. Power at 50 °C and corrected power is displayed in Table 4.

Table 4. *Power calculated with synthetical weather data and with registered data from weather stations for the Sanyo HIT plant in Germany and the JA plant in Chile. The calculations are performed for 1000 W/m^2, 50 °C, whilst "corrected" stands for 25 °C module temperature*

Module type	Power (kW) Synthetical	Power (kW) Weather station	Power (kW) Synthetical corrected	Power (kW) Weather station corrected
Sanyo 240 Wp	3.82±0.04	3.76±0.03	4.48	4.42
JA 345 Wp	1.28±0.02	1.21±0.03	1.39	1.32

Based on the results it can be stated that both simulations lead to comparable corrected results. The JA plant in Chile comes with a rated and verified total power of 1.3 kWp. Simulation results show a deviation of 1.3% if local weather data is used and 6.9% for the simulation based on synthetical weather data as input. The Sanyo HIT plant shows a deviation of +2.8% if local weather data is used and +4.2% for the simulation based on synthetical weather data as input. Wind speed data is not available, causing an unknow error in above's calculation.

4.3 Solar irradiation

As mentioned in section 4.1 solar irradiation is subject to be the main cause for the larger differences in yield calculation for both simulations. In terms of the total annual irradiation a difference of ~10% is seen for the location of Gelsenkirchen (Figure 3, left) and ~15% for the location of Antofagasta (Figure 3, right). For both locations the irradiation as determined by the weather stations was significantly higher if compared to the synthetically calculated irradiation.

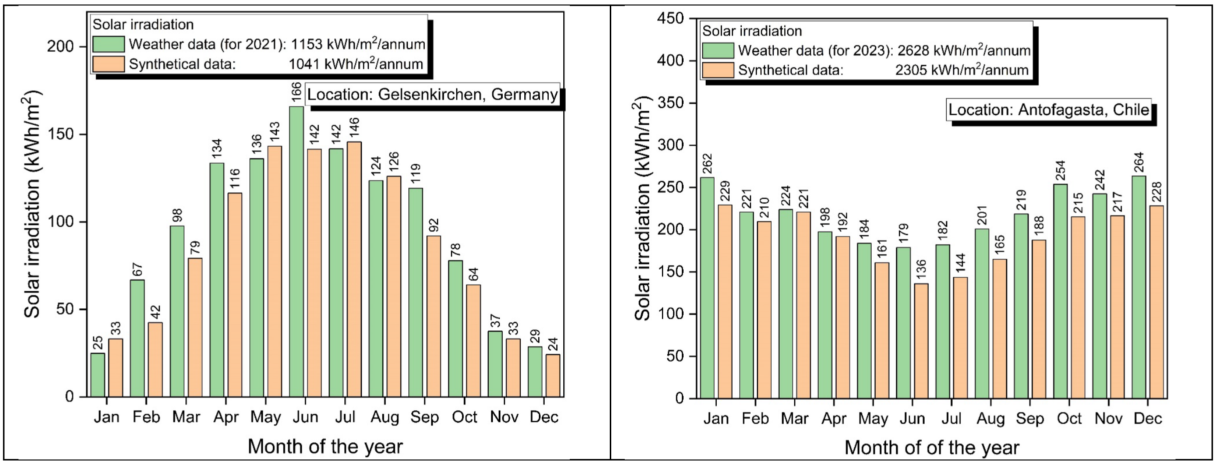

Figure 3. *Monthly and annual solar irradiation for both locations (left: Gelsenkirchen, Germany, and right: Antofagasta, Chile)*

If the irradiance is the key driver for the larger deviation seen in the simulation results based on synthetical weather data a correlation between irradiation difference and calculated yield must exist. The best procedure is to use the calculated power and measured power on at least hourly base to calculate the differences. However, these two sets of power data don't correlate in real-time. This lack of immediate correlation means that differences can only be compared over larger time scales. Therefore, for the purpose of this comparative investigation, energy yield, which is calculated on daily basis, is used. The data displayed in Figure 4 (left) shows the correlation between the absolute difference of measured and simulated yield to the difference between measured and synthetical irradiation, exemplarily for the Sanyo HIT plant in Gelsenkirchen. As can be seen a correlation exists for the simulation results based on synthetical weather data. No correlation can be stated for the simulation results based on on-site weather data (not shown), measured by the weather station. Figure 4 (right) shows that for months with large irradiation differences the difference of simulated to measured yield is high.

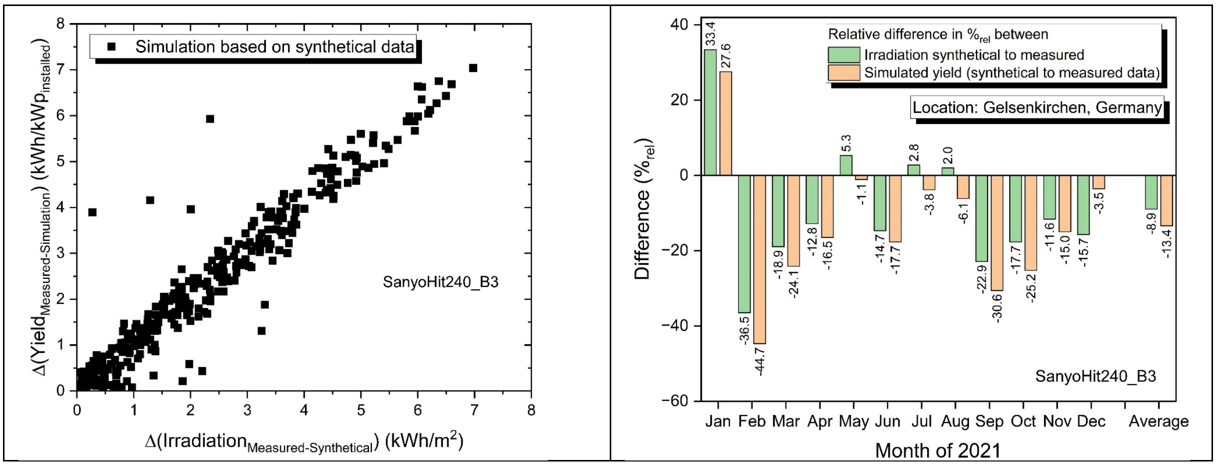

Figure 4. *Left: difference of measured and simulated yield in dependence to the difference between measured and synthetical irradiation, right: relative monthly irradiation and yield differences.*

4.4 Module temperature

The average annual module temperature is (30.23±9.85) °C for the simulation based on on-site weather data and (33.28±7.72) °C for the simulation based on synthetical weather data. This annual average difference in module temperature alone leads to a difference of ~1% in calculated power. The distribution of the temperature for both simulation scenarios is displayed in Figure 5 (left). Largest differences exist for temperatures ≤ 25 °C. The large deviation at ~25 °C has least impact since it is close to STC temperature and hence has no larger impact by the temperature coefficient correction.

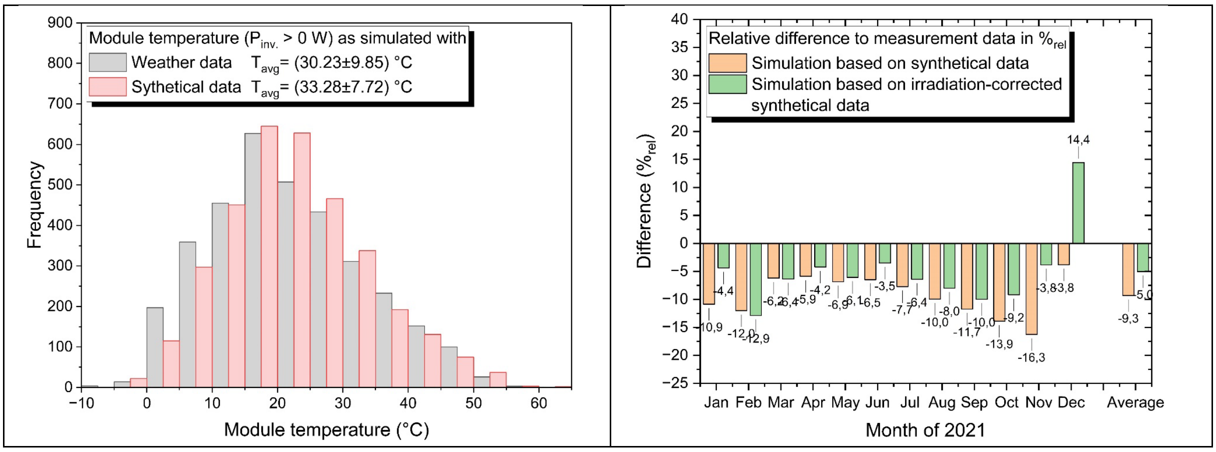

Figure 5. *Left: simulated module temperature distribution and right: relative difference of yield calculation for the original and irradiation-corrected data*

4.5 Discussion

Results of chapter 4 prove irradiation to be the main root cause for the large deviation found in simulations. This result also aligns with previously published results by Müller et al. where a deviation of at least 5% for yield simulation due to irradiation is reported [5]. The measurement uncertainty may contribute to about ~1.6% whereas power calculation and temperature correction may even have an effect of up to 4.6%. The calculation of module temperature may impact the power simulation results by about 1%. Based on the difference between synthetical and measured irradiance (Figure 2, left) the difference in yield between the simulation for the location of Gelsenkirchen was irradiation corrected by applying the data as displayed in Figure 3, left. The results are shown in Figure 5, right. By applying on-site data, the annual yield difference of −9.3% is reduced to −5.0%. For each individual month, with the exception of December, a strong improvement in yield simulation is achieved. Important to mention that this correction is only an irradiation correction based on monthly differences without taking any other factors into account. The reason why the month of December exceeds the original result is the extremely low irradiation level at this time of the year, at this geographical location, where on certain days the irradiance level even stayed well below 100 W/m^2. For days with a such low irradiance level yield simulation is not capable of delivering reliable results. For the sake of simplicity, the focus of this evaluation was set on the Sanyo HIT plant in Gelsenkirchen and the JA plant in Chile.

5. Conclusion

Yield calculations based on modern design and simulation software for photovoltaic systems strongly depend on the quality of weather data. For most locations world-wide on-site weather data is not available and synthetically generated weather data is used which leads to larger simulation deviations compared to measurement data as provided by the inverter.

Eight PV plants were evaluated, seven located in Gelsenkirchen, Germany and one in Antofagasta, Chile. The combined uncertainty of the yield simulation is off up to a factor of four for the simulation based on synthetical weather data if compared to the simulation based on on-site weather data. The use of synthetically generated weather data may even lead to an absolute deviation as large as 23.7% (SolonBlack 280 Wp plant).

An in-depth study for the Sanyo HIT plant revealed that the most significantly contributing factor is irradiation, followed by power calculation and temperature correction, contributing by up to 4.2%. Module temperature calculation contributes by ~1% and measurement uncertainty by ~1-2%. The import of on-site weather data strongly improves simulation accuracy. If on-site irradiation is used the deviation improves from 18% to 9.3%.

The outcome shows that the averaged differences between simulation and measurement are decreased by a factor of up to four if on-site weather data is used as input for the simulation, even for sites with wide clear sky conditions as it is for the Antofagasta Region, in northern Chile.

Data availability statement

The data that support the findings of this study are available from the corresponding author, [A.S.], upon reasonable request.

Author contributions

This study was designed, directed and coordinated by A.S. as the principal investigator. A.S and M.B. designed the experiments, planned and performed the data analysis. J.R.A. supported with measurement data from Chile and commented on the design of experiments. J.C. programmed evaluation software for the study. T.N. performed the outdoor measurements and maintained the measurement tools. The manuscript was written by A.S. and commented on by all authors. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

[1] D. Milosavljević, T. S. Kevkić, and S. J. Jovanović, “Review and validation of photovoltaic solar simulation tools/software based on case study“, Open Physics 2022; 20: 431–451.

[2] G. Bender, S. Pelland, and L. L., Vaisala, “Comparing photovoltaic yield prediction from synthetic weather years to those from hourly time series“, Vaisala Energy, 2017.

[3] A. Migan-Dubois, J. Badosa, F. C. Obaldía, O. Atlan, V. Bourdin, M. Pavlov, D. Y. Kim, Y. Bonnassieux, “Step-by-step evaluation of photovoltaic module performance related to outdoor parameters: evaluation of the uncertainty“, 44th IEEE Photovoltaic Specialists Conference (IEEE-PVSC), Jun 2017, Washington, United States. pp.626-631, 10.1109/PVSC.2017.8366615.

[4] M. Zouine, M. Akhsassi, N. Erraissi, N. Aarich, A. Bennouna, M. Raoufi, A. Outzourhit, “Mathematical models calculating PV module temperature using weather data: Experimental study“, Lecture Notes in Electrical Engineering 519:630-639, DOI:10.1007/978-981-13-1405-6_72, April 2018

[5] B. Müller, L. Hardt, A. Armbruster, K. Kiefer, C. Reise, “Yield predictions for photovoltaic power plants: empirical validation, recent advances and remaining uncertainties“, Progress in Photovoltaics Research and Applications 24(4):570–583, 2016, DOI:10.1002/pip.2616