Data-driven inference of unknown tilt and azimuth of distributed PV systems

Citation for published version (APA):
Meng, B., Loonen, R. C. G. M., & Hensen, J. L. M. (2020). Data-driven inference of unknown tilt and azimuth of distributed PV systems. Solar Energy, 211, 418-432. https://doi.org/10.1016/j.solener.2020.09.077

Document license:
CC BY

DOI:
10.1016/j.solener.2020.09.077

Document status and date:
Published: 15/11/2020

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 04. Jun. 2022
Data-driven inference of unknown tilt and azimuth of distributed PV systems

B. Meng*, R.C.G.M. Loonen, J.L.M. Hensen

Building Physics and Services, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, the Netherlands

ARTICLE INFO

Keywords:
PV orientation estimation
Data-driven method
Distributed photovoltaics
Solar irradiance

ABSTRACT

Information about the orientation (i.e. tilt and azimuth angles) of PV modules is a fundamental input for PV performance studies. However, this type of metadata is difficult to obtain for distributed in-use PV systems, which considerably impedes monitoring and diagnostics of PV systems and power grid management. Recently proposed parameterization methods to derive PV tilt and azimuth have limited practical applicability because they rely on data that is often not accessible. Hence, the aim of this research is to develop a novel method to infer tilt and azimuth angles of distributed PV systems, utilizing widely available data. The proposed method, which is based on a curve-matching procedure, is designed to be scalable because it only requires PV generation data and off-site irradiance data at a 1-hour time interval. The accuracy of this method has been tested using notional PV systems with a wide variety of orientations, as well as with data from real PV roofs distributed across the Netherlands. These tests show that the proposed method can parameterize azimuth and tilt angles of PV panels as far as 195 km away from the irradiance measurement site with mean absolute errors of 4.5° and 4.3° respectively. A demonstration case of the proposed approach written in Python is uploaded online for other researchers to use.

1. Introduction

Currently, nearly 40% of newly-installed PV systems worldwide is installed in the built environment, most notably on the rooftops of residential, commercial, and industrial buildings (PVPS IEA, 2018). According to the projections by the International Energy Agency (IEA), the growth of distributed generation capacity will further accelerate and account for almost half of PV capacity growth by 2023. Continuous price reductions, extended policy support, and emerging business models are key factors contributing to this transition (IEA, 2018).

The growth in distributed PV installations coincides with an increasing awareness of the role that PV output monitoring can play in practice. Energy yield information of a large number of PV systems is routinely being monitored, and this data is then uploaded to online repositories as part of regular system operation (Feng et al., 2019). As indicated below, this monitoring data contains a wealth of information for many stakeholders. With the intermittent and irregular nature of PV generation and intensive penetration of PV systems into the power grid, the difficulty of grid management has significantly increased. Noticeable effects on the grid can be seen in many aspects: voltage rise and fluctuations, power fluctuations and reverse flow, power factor changes, etc. (Karimi et al., 2016). In this respect, monitoring data of grid-connected PV systems in combination with forecasting techniques are used by power system operators to optimize power flow management, ensure the grid stability and achieve economic dispatch (Antonanzas et al., 2016; Rifonnoeur et al., 2011). In-depth insights into PV generation characteristics are also requested by owners and asset managers to evaluate the profitability of PV investments (Overholm, 2015; Vimpani and Junnila, 2017). Besides, PV monitoring methods are adopted by PV service providers for automated anomaly detection of PV systems (Drews et al., 2007; Ortega et al., 2019), in which monitored PV power is compared with simulated values. Failure analysis can then be conducted when significant differences are observed between the two data sets (Chine et al., 2014; Gokmen et al., 2012; Müller et al., 2016; Silvestre et al., 2013).

Despite the large potential and wide availability of monitored PV data, the use of this resource is restricted in practice. It is argued that one of the main obstacles is a lack of accurate information about module orientation (i.e. tilt and azimuth) of distributed PV systems. Tilt and azimuth angles of PV arrays are basic inputs for transposition models that are used to calculate plane of array irradiance ($E_{poa}$) from
The applicability of this method, however, is limited by the following factors. Firstly, as the essential information required in the approach, precise locations of PV modules are not provided in many databases for confidentiality reasons. Also, the accuracy of this method primarily depends on the resolution of DEM, while complete models are not publicly available for many places. Moreover, the method is time-consuming and impractical to parameterize for a large number of PV systems.

An alternative way to obtain PV metadata is to use PV monitoring datasets (Lorenz et al., 2011). Many databases attempt to collect PV metadata from monitoring companies and inverter manufacturers at the system design stage or allow PV owners or installers to register information themselves. However, as mentioned above, unlike electricity generation data which is intensively collected, the detailed metadata on tilt and azimuth tends to be incomplete or partially wrong. According to the inspection conducted by Killinger et al. (2018), in a number of datasets, a disproportionately large number of tilt values with 0° or 180° and azimuth values of 0° (due south in the northern hemisphere and due north in the southern hemisphere) are reported. The authors conducted sampling checks based on aerial images and found that these angles are incorrectly reported in most cases. It is argued that these orientation values close to 0° are default settings of datasets when the data is not reported. Besides, PV azimuth metadata is generally reported with a 45° interval in databases. Therefore, uncertainties of PV tilt and azimuth in datasets arise from both input error and default settings of the datasets.

Given the limited applicability of the methods mentioned above, a few studies have explored the parameterization of angular characteristics of PV systems utilizing PV power measurements and other recorded data.

Saint-Drenan et al. (2015) developed an algorithm to identify the orientation of PV modules based on PV yield and meteorological measurement data. In this method, PV performance model requires three parameters (module tilt, azimuth and angular loss coefficient). Simulated PV output and effective irradiance are calculated for all the time stamps in the year. PV parameters that maximize the likelihood that simulation outcomes match with power measurements are accepted as estimation results. According to a test on two actual PV plants, this method led to average estimation errors of tilt and azimuth of 0.75° and
3° respectively.

Ruelle et al. (2016) also propose a customized optimization method to estimate PV orientations by minimizing the error between simulated PV output and measurement data. To simulate PV performance, the single diode array performance model is used based on several subroutines from the pvlib python toolbox (F. Holmgren et al., 2018). However, as an essential input of the model, details about the equivalent circuit of the specific PV modules are required, but this information is often not available for large-scale studies. The authors have solved this issue by assuming the electrical parameters of a generic polycrystalline module as default for all systems, but the potential error of this assumption was not quantified.

Killinger et al. (2017) present a parameterized PV system metadata approach that requires one exogenous input (ambient temperature). Measurement data with high temporal resolution (5 mins) is utilized in their study to generate statistical clear sky curves for detecting clear sky periods. The statistical clear sky curve is defined as the 95th percentile measurement for each time step from the past 30 days. Then, PV output is simulated using a simplified quadratic model. The constant coefficients in the quadratic model are specific to different module technologies and inverter efficiencies. In the parameterization step, a nonlinear least square solver is used to derive system orientation. Applicability of this approach is limited by the relatively high quality requirements on PV output data with a short time interval and the need for detailed technical information about PV configurations. The method reaches an average PV orientation parameterization error of 4°.

Lingfors et al. (2018) compare two fundamentally different methods for identifying PV system orientation, a Light Detection and Ranging (LiDAR) based model, and their QCPV-Tuning model. The conclusion is drawn that manually measured PV orientation from high-resolution LiDAR data can be used as a benchmark, while the QCPV-Tuning model results in a mean accuracy of 10° in terms of module orientation (i.e., normal vector) taking 10-minute yield data and detailed electrical characteristics of PV systems as inputs.

Haghdadi et al. (2017) express PV generation as a function of latitude, tilt, and azimuth to estimate any unknown parameters by fitting the measured data to parametric curves that minimize the error. Prior to the curve fitting, multiple years’ worth of data is required to generate theoretical output and detect clear sky timestamps. Then, the output of PV systems is simulated by using NREL’s PVWatts model (Marion, 2008), which requires the characteristics of PV modules as input. However, there is a general lack of multi-year PV yield measurement, especially for newly installed PV systems. Simulation and curve fitting of long-term data is also time-consuming for large-scale studies in practice. It is reported that this method can estimate array tilt and azimuth with mean absolute errors of 2.75° and 5.85°, respectively, for a typical PV system.

Mason et al. (2020) develop a deep neural network (DNN) approach for estimating PV tilt and azimuth using only behind-the-meter data. The authors combine customer load profiles with simulated PV profiles to generate net load data. Then features that have a relationship with the tilt and azimuth are extracted from the net load data to develop the DNN model. The approach achieves a mean absolute error of 2.55° and 4.71°, respectively to estimate tilt and azimuth. However, PV systems that are utilized for training and validating the DNN model in this research are simulated from the irradiance data of the same location. On the one hand, the effectiveness of the developed DNN model to parameterize PV systems without irradiance data is unavailable is not tested. On the other hand, the uncertainty of actual PV performance in real working environments is not considered, which could cause a significant increase in parameterization errors.

Although the abovementioned parameterization methods have demonstrated the capability to estimate PV configuration parameters (tilt and azimuth) from measured PV output and meteorological data with reasonable accuracy, their applicability is limited by the high requirements in terms of time resolution and duration of input data. For those methods that require meteorological data as input, irradiance and ambient temperature need to be measured in parallel to the AC power production. However, pyranometers and thermometers are generally not installed on-site for many distributed PV systems. For other methods that pursue the independence of any exogenous data, multiple years of high-time-resolution PV data is required. In this case, however, the long-term data with a short time interval is not available for many PV systems due to limited running time or storage capacity. In addition, several of the existing methods depend on PV systems simulation as an important part of the process. However, a general lack of knowledge about the specific electrical characteristics of distributed PV systems, which is a necessary input for detailed PV simulation models, constitutes a main obstacle with respect to scalability potential.

1.2. Research objective and paper outline

Considering the practical limitations of existing methods described in the literature review, there is a clear need for approaches that enable the inference of a large number of missing PV orientation metadata without weather measurement on site. The goal of the presented research is to develop a novel method to parameterize the tilt and azimuth of PV systems utilizing widely available data. The proposed method only requires PV generation data and off-site irradiance data at a 1-hour time interval for no more than one year, and therefore is widely applicable. Beyond that, this proposed method aspires to make a significant contribution towards improving scalability. Different from existing methods, PV performance simulation is not conducted in this approach. Therefore, detailed information regarding the electrical characteristics of the PV systems is not required in the computation process. The errors caused by default settings that would be chosen for unknown PV electrical configurations in existing methods can therefore be avoided.

This paper is structured as follows. Section 2 presents the main working principles that enable the parameterization of PV orientations, followed by the procedure of the proposed method. Section 3 demonstrates the parameterization method using simulated PV prototypes with a wide variety of PV orientations. Section 4 presents the empirical validation based on 13 actual rooftop PV systems distributed across the Netherlands. A sensitivity study that assesses the performance of the parameter estimation algorithm with different assumptions for several influencing factors is then discussed in Section 5. The advantages and limitations of the approach are briefly summarized in Section 6, also including future prospects.

2. Proposed methodology

The basic idea of the proposed method is to identify the set of tilt and azimuth angles that maximizes the similarity between normalized Epoa and PV output during clear days in different seasons. The working principle of the method relies on the distinct shape of normalized Epoa profiles for various PV orientations as well as the strong correlation between Epoa and PV yield across different seasons, which is explained in Section 2.1.

2.1. Influence of PV tilt and azimuth on normalized Epoa

To illustrate the characteristic effect of PV tilt and azimuth angles on normalized Epoa profiles, solar irradiance data measured at the outdoor BIPV research and development facility ‘SolarBEAT’ (Valckenborg et al., 2015) in Eindhoven, the Netherlands, is transposed to Epoa for different combinations of PV tilts and azimuths. The pvlib python implementation of the Perez model is used as the transposition model for determining total in-plane irradiance and its beam, sky diffuse, and ground reflected components (Yang, 2016). The accuracy of the Perez model has been validated extensively in various climatic environments (Cameron et al., 2008; Harrison & Coombes, 1989; Loutzenhiser et al., 2007).
By transposing measured solar irradiance of a typical summer day to different planes, Fig. 1 illustrates the effect of tilt angles on the $E_{\text{poa}}$ profile incident on PV panels in four main directions. The results show that, for north-facing PV panels in the northern hemisphere, as the tilt angle increases, the amount of $E_{\text{poa}}$ decreases dramatically. On the clear summer day, as the tilt angle of north-facing PV panels increases to 15°, 30°, and 45°, the maximum $E_{\text{poa}}$ is approximately 82%, 62%, and 38% of the maximum irradiance available on the horizontal plane. Moreover, with the inclination angles higher than 60°, not only the maximum value of $E_{\text{poa}}$ decreases significantly, but there is also a shift in the time at which the peak value happens from noon to sunrise and sunset. These unique effects indicate that north-facing PV panels, which are uncommon in the northern hemisphere, can be easily identified through the low amount of irradiance on the plane.

For PV panels facing south, east, or west, normalized curves are obtained by dividing $E_{\text{poa}}$ by the corresponding maximum value of the whole day (Fig. 1). The normalization process eliminates the influence of absolute value differences and enables us to recognize shape changes of the $E_{\text{poa}}$ profile as a function of PV tilt. Fig. 1(b) shows that, for PV panels facing due south, as the tilt angle increases from 0° to 90°, the time when maximum $E_{\text{poa}}$ happens remains the same. At the same time, the time duration with direct solar irradiance, and thus relatively high normalized incident irradiance values gets slightly shorter. In comparison, for systems oriented towards east and west, the time at which maximum $E_{\text{poa}}$ appears is advanced and delayed, respectively, with increasing inclination degrees. On this specific date, comparing horizontal PV panels with vertical ones facing due west or east, the difference between the time of maximum $E_{\text{poa}}$ occurrence is 3 h. This time gap will be shortened in winter due to shorter daily sunshine hours. In addition, as the tilt angle increases, the time duration of relatively high $E_{\text{poa}}$ is shortened to a larger extent than for south-facing panels. Specifically, on this clear summer day, for PV panels facing east or west, vertical panels receive normalized $E_{\text{poa}}$ with a value above 0.3 for 6 h less per day.

To further illustrate the effect of PV azimuth on normalized $E_{\text{poa}}$ profiles, 30°-tilted PV panels with different azimuths are taken as examples. Fig. 2 shows the normalized transposition results on a clear winter day. It indicates that, even on a winter day with few sunlight hours, the time when maximum in-plane irradiance occurs is remarkably influenced by the panel azimuth.

In summary, it can be drawn from above transposition results that, as PV tilt and azimuth vary, significant changes can be observed in the normalized in-plane irradiance curves. Such changes are noticeable in two aspects: (a) the position of peak of the curve, (b) the time duration with direct irradiance when normalized $E_{\text{poa}}$ is relatively high.

Apart from the notable influence of panel orientation on the profile of normalized $E_{\text{poa}}$ curves, the strong relationship between $E_{\text{poa}}$ and PV yield also needs to be explained as the basis of curve matching in this parameterization method. Since PV systems absorb solar radiation as the energy source to generate electricity, there is a strong dependence between PV output and $E_{\text{poa}}$ (AlSkaif et al., 2020). However, in the process of energy conversion, PV systems are subjected to environmental losses, transfer losses through the wire and system losses within electrical components. Among these systematic loss effects, several prominent components such as angle of incidence effects, temperature effects, soiling cover and low light behavior show obvious seasonal features. These effects are decisively influenced by the seasonally dependent factors such as solar position, air temperature, wind speed, rainfalls and radiation intensity. As a result, there is a strong correlation between $E_{\text{poa}}$ and PV yield with seasonal features. According to the sensitivity analysis conducted by Hansen et al. (2013), variations in $E_{\text{poa}}$ are the dominant contributor to the PV yield fluctuations. The correlation coefficient between these two parameters remains high and fluctuates seasonally. Given the notable dependence between $E_{\text{poa}}$ and PV yield, the basic idea of the proposed method is to conduct curve matching between them in different months. To take the secondary and seasonal PV loss effects into consideration, an appropriate range of orientations that all lead to low curve mismatch results are taken as the monthly result. The monthly

Fig. 1. Effect of tilt angle on $E_{\text{poa}}$ profile of PV panels in four directions: (a) north, (b) south, (c) east, (d) west.

Fig. 2. Effect of azimuth angle on normalized $E_{\text{poa}}$ profile of PV panels with an inclination of 30°.

For PV panels facing south, east, or west, normalized curves are obtained by dividing $E_{\text{poa}}$ by the corresponding maximum value of the whole day (Fig. 1). The normalization process eliminates the influence of absolute value differences and enables us to recognize shape changes of the $E_{\text{poa}}$ profile as a function of PV tilt. Fig. 1(b) shows that, for PV panels facing due south, as the tilt angle increases from 0° to 90°, the time when maximum $E_{\text{poa}}$ happens remains the same. At the same time, the time duration with direct solar irradiance, and thus relatively high normalized incident irradiance values gets slightly shorter. In comparison, for systems oriented towards east and west, the time at which maximum $E_{\text{poa}}$ appears is advanced and delayed, respectively, with increasing inclination degrees. On this specific date, comparing horizontal PV panels with vertical ones facing due west or east, the difference between the time of maximum $E_{\text{poa}}$ occurrence is 3 h. This time gap will be shortened in winter due to shorter daily sunshine hours. In addition, as the tilt angle increases, the time duration of relatively high $E_{\text{poa}}$ is shortened to a larger extent than for south-facing panels. Specifically, on this clear summer day, for PV panels facing east or west, vertical panels receive normalized $E_{\text{poa}}$ with a value above 0.3 for 6 h less per day.

To further illustrate the effect of PV azimuth on normalized $E_{\text{poa}}$ profiles, 30°-tilted PV panels with different azimuths are taken as examples. Fig. 2 shows the normalized transposition results on a clear winter day. It indicates that, even on a winter day with few sunlight hours, the time when maximum in-plane irradiance occurs is remarkably influenced by the panel azimuth.

In summary, it can be drawn from above transposition results that, as PV tilt and azimuth vary, significant changes can be observed in the normalized in-plane irradiance curves. Such changes are noticeable in two aspects: (a) the position of peak of the curve, (b) the time duration with direct irradiance when normalized $E_{\text{poa}}$ is relatively high.

Apart from the notable influence of panel orientation on the profile of normalized $E_{\text{poa}}$ curves, the strong relationship between $E_{\text{poa}}$ and PV yield also needs to be explained as the basis of curve matching in this parameterization method. Since PV systems absorb solar radiation as the energy source to generate electricity, there is a strong dependence between PV output and $E_{\text{poa}}$ (AlSkaif et al., 2020). However, in the process of energy conversion, PV systems are subjected to environmental losses, transfer losses through the wire and system losses within electrical components. Among these systematic loss effects, several prominent components such as angle of incidence effects, temperature effects, soiling cover and low light behavior show obvious seasonal features. These effects are decisively influenced by the seasonally dependent factors such as solar position, air temperature, wind speed, rainfalls and radiation intensity. As a result, there is a strong correlation between $E_{\text{poa}}$ and PV yield with seasonal features. According to the sensitivity analysis conducted by Hansen et al. (2013), variations in $E_{\text{poa}}$ are the dominant contributor to the PV yield fluctuations. The correlation coefficient between these two parameters remains high and fluctuates seasonally. Given the notable dependence between $E_{\text{poa}}$ and PV yield, the basic idea of the proposed method is to conduct curve matching between them in different months. To take the secondary and seasonal PV loss effects into consideration, an appropriate range of orientations that all lead to low curve mismatch results are taken as the monthly result. The monthly
results across the year are eventually overlapped to eliminate the error caused by the PV loss effects in curve matching and to obtain an accurate estimation of PV orientation. Each step of the parameterization procedure will be explained in detail in Section 2.2.

2.2. Parameterization

Fig. 3 illustrates the main steps of the proposed method to infer the tilt and azimuth of PV panels from widely available irradiance data and local PV generation data. The core idea is that, through transposing global irradiance measurements to E_{poa} of any possible plane, the method evaluates the mismatch between normalized E_{poa} curves and the measured PV output curve for the clearest day of each month. Then, an appropriate range of PV orientations that minimizes curve mismatch is taken as the monthly result. The patterns of these monthly results can vary significantly over the year. In the end, to narrow down the range of parameterization results, the monthly results are overlapped, and finally, the average value of the subset with most overlapping PV orientations is calculated as the final inference result. Details of each step are provided in the following subsections.

2.2.1. Selecting the clearest day of each month

The first step of the proposed approach is to select the clearest day of each month based on hourly irradiance data. To evaluate the clearness of a day, daily diffuse fraction (k_d) is calculated as follows:

\[ k_d = \frac{DHI_{daily}}{GHI_{daily}} \]

where: \( DHI_{daily} \) = daily cumulative diffuse horizontal radiation (kWh/m^2); and \( GHI_{daily} \) = daily cumulative global horizontal radiation (kWh/m^2).

By accumulating and simply processing the hourly irradiance measurement data, \( k_d \) is calculated to quantify clear conditions of the sky. High values of \( k_d \) mean that diffuse irradiance accounts for a large proportion of global irradiance, while direct normal irradiance is low due to overcast skies. In the proposed method, input irradiance data is collected from a nearby weather station. Due to the separation in space, passing clouds during partly cloudy days could induce significant mismatch between the off-site irradiance measurement and actual irradiance on PV panels. To reduce the interference effects of passing clouds on the curve matching process, clear days which have the lowest \( k_d \) value per month are detected. During these clear days, the irradiance mismatch is reduced due to a smaller occurring probability of passing cloud, which enables the off-site irradiance measurements to be used for curve matching. Compared with other clear sky detection methods to identify clear-sky periods in the time series of global horizontal irradiance (GHI) (Reno et al., 2012), \( k_d \) was identified as a suitable approach for the proposed research because this indicator can evaluate overall clearness of a whole day.

As described above, \( k_d \) can be conveniently calculated with DHI and GHI measured in the weather station. However, for places where DHI measurement is not available, root mean square error (RMSE) between GHI and its clear-sky counterpart can be calculated as an alternative indicator to evaluate daily clearness. The RMSE is defined as:

\[ RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (GHI_i - GHI_{clear})^2} \]
RMSE = \sqrt{\frac{1}{T \sum_{i=1}^{T} (GHI(i) - GHI_{\text{norm}}(i))^2},}
\tag{2}

where $GHI_{\text{norm}}$ denotes the modeled clear-sky GHI at each timestamp $i$ during the day $T$. According to the comparison and analysis conducted by Yang (2020), McClear is taken as the clear-sky irradiance model in this work considering both accessibility and accuracy. Specifically, the modeled clear-sky irradiance data can be obtained from various web services (Schroeder-Homscheidt et al., 2016) or thecamsRad package in R (Lundstrom, 2016). Using this model, different components of clear-sky irradiance are available globally with the time coverage from 2004-01-01 to 2 days ago at a time resolution from 1 min to 1 month. Benefitting from the global coverage and convenient accessibility of McClear’s method, the RMSE calculated from Eq. (2) is taken as an alternative evaluation indicator of daily clearness when diffuse irradiance measurement is not available.

\subsection*{2.2.2. Transpose GHI to $E_{\text{poa}}$ of any possible plane}

After picking out the clearest day of each month, hourly measured irradiance is transposed to in-plane irradiance for any combination of PV tilt and azimuth at 1° intervals to investigate their combined effect on the resulting $E_{\text{poa}}$ profile. The Perez model (Perez et al., 1990) is selected as the transposition model to determine total in-plane irradiance and its beam, ground reflected, and sky diffuse components. To take all possible scenarios into account, the upper and lower limits of the parameters are set from 0° to 35° for azimuth, and from 0° to 90° for tilt, with an interval of 1°. In total, 32,760 time series of $E_{\text{poa}}$ are obtained for the clearest day of each month.

\subsection*{2.2.3. Normalization and data filtering of $E_{\text{poa}}$ and PV output}

Hourly irradiance measurements have been transposed to $E_{\text{poa}}$ for all possible scenarios in the previous step. Due to different dimensions of $E_{\text{poa}}$ and PV yield data, normalization will be conducted as a preprocessing step before evaluating curve mismatch. In the process of min–max normalization, data is rescaled to the range from 0 to 1. Since both irradiance data and measured AC power grow from a minimum of 0 during the day, normalized in-plane irradiance ($E_{\text{poa, norm}}$) and normalized AC power ($AC_{\text{norm}}$) are calculated from Eqs. (3a) and (3b),

\begin{equation}
E_{\text{poa, norm}}(t) = \frac{E_{\text{poa}}(t)}{E_{\text{poa, max}}},
\tag{3a}
\end{equation}

\begin{equation}
AC_{\text{norm}}(t) = \frac{AC(t)}{AC_{\text{max}}},
\tag{3b}
\end{equation}

Then, the normalized data is further filtered according to the position of the sun. Data is excluded from further processing when solar zenith angles are higher than 70° for three reasons. On the one hand, compared with solar irradiance sensors equipped in weather stations that are typically elevated or installed in an open environment, distributed PV systems on site are often shaded by buildings or vegetation in the neighborhood when the solar elevation is low. The shadows cast by surrounding obstacles can have a significant impact on PV performance and lead to inconsistencies between irradiance data and PV output. Specifically, for the typical rooftop PV systems in the residential area studied in this paper, significant power loss due to shading effects can be noticed at solar zenith angles larger than 70°. On the other hand, due to the difference in sensor thresholds, when the intensity of radiation is low at sunrise and sunset, irradiance measurement and PV yield records have respective starting and ending times during the day. This variance between time series can cause unexpected errors in evaluating curve similarity. Last but not least, the absolute values of both time series are very small during this period, which causes small deviations to have a significant relative impact. In the current research, data on solar zenith angle and irradiance is obtained from meteorological sensors installed in ‘SolarBEAT’. Future users of the proposed method could also work with calculated solar position data, for example following the algorithm described in (Reda and Andreas, 2008). 

\subsection*{2.2.4. Evaluating curve mismatch between normalized curves}

After normalization and data filtering, curve matching is conducted between normalized $E_{\text{poa}}$ and AC power. The root mean square error (RMSE) is taken as the cost function to evaluate curve mismatch. For the clearest day of each month, RMSE is calculated to compare all 32,760 possible time series of $E_{\text{poa, norm}}$ with the $AC_{\text{norm}}$ at hourly time step $i$ spanning the whole day $T$. For each combination of azimuth angle ($\alpha$) and tilt angle ($\beta$), the RMSE is calculated using Eq. (4). PV orientation scenarios with low RMSE are expected to minimize the disagreement between curves.

\begin{equation}
RMSE(\alpha, \beta) = \sqrt{\frac{1}{T \sum_{i=1}^{T} (E_{\text{poa, norm}}(\alpha, \beta, i) - AC_{\text{norm}}(\alpha, \beta, i))^2},}
\tag{4}
\end{equation}

\subsection*{2.2.5. Calculating monthly results}

PV output is not only influenced by $E_{\text{poa}}$ but is also affected by the angle of incidence, air temperature, wind speed, and other parameters with seasonal characteristics. Due to these seasonal effects, the PV orientation inferred from the best-matching $E_{\text{poa}}$ curve in the previous step may deviate from the actual value. In this research, a range of PV orientations with relatively low RMSE is accepted as monthly results. It should be noted that an acceptable range is essential for inference accuracy. With a range too small, the exact PV orientation may not be included in monthly inference results due to seasonal and secondary effects. On the other hand, the accuracy of inference results can reduce if the threshold for RMSE acceptance is too large. During the establishment process of the method, it has been tested that the top 1% of PV orientations with the lowest RMSE, namely the most fit 327 out of 32,760 possible tilt and azimuth combinations in total, are identified as the monthly inference results. This acceptable range is verified to moderately reflect the influence of secondary and seasonal effects. Further discussion on the selection of acceptable RMSE range is conducted in Section 5.1.

\subsection*{2.2.6. Overlapping monthly results}

In the previous step, to take secondary and seasonal effects into consideration, the best-matching 1% of PV orientations are taken as monthly parameterization results. In order to narrow down the range of monthly inference results and reduce the error caused by such effects in the curve matching, monthly results across the year are overlapped. For each of the 32,760 possible combinations of tilt and azimuth angle, its number of occurrences in monthly results is counted. The count of overlaps for each PV orientation angle is an integer from 0 to a possible maximum of 12. A higher count of overlaps means greater consistency with the monthly parameterization results. The overlapping process is illustrated in the polar contour plot in Fig. 3 (f). As represented by the peak area in the contour plot, the maximum count of overlaps generally corresponds to a few orientation angles. In this study, all orientation angles which exhibit the maximum count of overlaps are picked out, and their average tilt and azimuth are calculated as the final parameterization result (Eq. (5)):

\begin{equation}
X_{\text{norm}} = \text{average}(X_i(\text{ove = ove}_{\text{max}})),
\tag{5}
\end{equation}

where $X_{\text{norm}}$ is the set of final azimuth and tilt estimation result, $X_i$ are the sets of tilt and azimuth whose count of overlap (ove) is equal to the maximum count of all (ove$_{\text{max}}$).

The method is carried out under the following assumptions. Firstly, the objects to be identified are distributed PV systems with a single orientation angle per site. Then, the PV systems are free from local shading at solar zenith angles lower than 70°. This is because the
shading effect can have a significant impact on PV performance and bring considerable error in the curve matching process. In addition, the mismatch of solar radiation between the PV system and the nearest weather station is low during clear days. In the proposed parameterization method, there is no need to install pyranometers on site. During the clear days detected by the weather station nearby, the irradiance mismatch is reduced due to a smaller occurring probability of passing clouds.

3. Simulation-based demonstration with modeled PV systems

In this section, the performance of the proposed method is tested using simulated notional PV systems with representative orientations. Hourly irradiance, air temperature, and wind speed data are collected from SolarBEAT, the outdoor BIPV(T) research facility located in Eindhoven. The simulated PV yield, together with irradiance measurements, are used to infer the orientations of the notional PV systems.

3.1. Simulation of PV prototypes

In order to evaluate the method in different scenarios, PV systems with typical tilt and azimuth angles are simulated in this case study. The representative azimuths investigated include due east, south, west, and north. Tilt angles considered include 0°, 15°, 30°, 45°, 60°, and 90°. For any combination of tilt and azimuth, firstly, $E_{poa}$ is computed using the Perez transposition model, while PV module temperature ($T_m$), as well as cell temperature ($T_c$), are calculated from the empirically-based thermal model described by Eqs. (6) and (7) (King et al., 2004). The meteorological measurement data over an entire year collected from SolarBEAT is given as the boundary condition to the models.

$$T_m = E_{poa} \cdot \left\{ e^{a+b \cdot WS} \right\} + T_a,$$

$$T_c = T_a + \frac{E_{poa}}{E_0} \cdot \Delta T,$$

where: $a$, $b$ - empirically-determined coefficients regarding module operating temperature; $WS$ - wind speed measured at standard 10 m-high, m/s; $T_a$ - Ambient air temperature, °C; $E_0$ - reference solar irradiance, 1000 W/m²; $\Delta T$ - predetermined temperature difference between the back surface and the cell, °C. The coefficients ($a$, $b$, $\Delta T$) used in the model are determined by the module types and mounting configurations.

Then, NREL’s PVWatts Version 5 DC power model, loss model and inverter model are employed to estimate DC and AC output (Dobos, 2014; Marion, 2008). The validation and verification of the model chain in PVWatts have been conducted extensively (Saracoglu, 2018). According to the PVWatts module model presented in Eq. (8), DC power ($P_{ac}$) is calculated from $E_{poa}$ total system losses ($L_{total}$) and cell temperature. $L_{total}$ is calculated from Eq. (9) by multiplying the percentage energy reduction of each loss factor ($L_i$). It should be noted that although the setting of $L_{total}$ has a drastic effect on PV modeling accuracy, its effect on the orientation estimation is significantly attenuated. This is because $E_{poa}$ and AC power are normalized before curve matching in the method, and different settings of $L_{total}$ have only a slight effect on the characteristic shapes of both curves. In this case, the value of $L_i$ is taken from the default settings of PVWatts Version 5 (Dobos, 2014), as listed in Table 1. Besides, inverter loss is explicitly modeled using Eqs. (10) and (11), which indicates that the simulated AC output ($P_{ac}$) is clipped to the nameplate ($P_{ac, 0}$) value if it exceeds the nameplate rating.

$$P_{ac} = \frac{E_{poa}}{E_0} \cdot P_{dc, 0} \cdot \left[ 1 + \gamma \cdot (T_c - T_0) \right] \cdot (1 - L_{total}),$$

$$L_{total} = 1 - \prod_{i=1}^{n} \left( 1 - L_i \right),$$

Table 1: Loss factors for PV system simulation.

| Loss factor of PV system | Default value in PVWatts system loss model |
|--------------------------|-------------------------------------------|
| Soiling                  | 0.02                                      |
| Shading                  | 0.03                                      |
| Snow                     | 0                                         |
| Mismatch                 | 0.02                                      |
| Wiring                   | 0.02                                      |
| Connections              | 0.005                                     |
| Light-induced degradation | 0.015                                     |
| Nameplate rating         | 0.01                                      |
| Age                      | 0                                         |
| Availability             | 0.03                                      |
| Total losses             | 0.14                                      |

$$\eta = \frac{\eta_{nom}}{\eta_{ref}} \left[ -0.0162 \frac{P_{dc}}{P_{dc, 0}} - 0.0059 \frac{P_{ac}}{P_{ac, 0}} + 0.9858 \right],$$

$$P_{ac} = \min(\eta P_{dc}, P_{ac, 0}),$$

where: $P_{ac, 0}$ - DC power at Standard Reporting Conditions (SRC), which consists of irradiance of 1000 W/m² with a spectral distribution conforming to the air mass 1.5 spectrum and a PV cell temperature of 25 °C; $\gamma$ - temperature coefficient, °C⁻¹; $T_0$ - cell reference temperature, 25 °C; $\eta_{nom}$ - nominal inverter efficiency; $\eta_{ref}$ - reference inverter efficiency.

The settings of the technical parameters required by the PV simulation model are shown in Table 2. In this case study, the proposed method is demonstrated to infer PV orientation from hourly irradiance measurement data and simulated PV yield. Since tilt and azimuth angles of the simulated PV prototypes are customized, the usability of this approach in different scenarios can be tested.

3.2. Parameterization of simulated PV system prototypes

In this section, the parameterization results of 21 simulated PV system prototypes are reported. According to Eq. (1), $k_T$ is calculated from hourly irradiance data to select the clearest day of each month in 2017. As shown in Table 3, the minimum $k_T$ per month ranges from 0.12 to 0.49 within the year.

Then the performance simulation and parameterization of PV prototypes are conducted on the clearest day of each month. Overlapping is conducted on the monthly results. For the PV orientations with the highest number of overlaps, the average tilt and azimuth are calculated as the final inference results. The exact orientations of these typical PV systems are known since they are artificially set. Table 4 compares final parameterization results with the exact values. The azimuth of 180° corresponds to due south in this paper. It should be noted that errors are calculated not only for tilt and azimuth separately but also for orientation, which represents the angle between the normal vector of the PV module and that of inference results. The difference in orientation, which combines both tilt and azimuth, can better represent the accuracy of the parameterization method. For instance, for slightly tilted PV panels, large azimuth errors do not result in significant orientation differences, as well as significant differences in in-plane irradiance and power generation.

Table 2: Settings of technical indicators for PV system simulation.

| Parameters                      | Value                      |
|---------------------------------|----------------------------|
| Cell temperature model          | 'open_rack_cell_glassback' |
| Temperature coefficient (°C⁻¹)  | -0.0035                   |
| Cell reference temperature (°C) | 25                         |
| Nominal inverter efficiency (-) | 0.96                       |
| Reference inverter efficiency (-) | 0.9637                   |
| DC power at reference conditions (W) | 1000                     |
The derivation results in Table 4 show that, the mean absolute errors (MAE) of tilt, azimuth, and orientation for 21 simulated PV panels in this case study are 4.8\(^\circ\), 3.1\(^\circ\), and 5.0\(^\circ\). One thing to note here is that the accuracy of this method varies for PV systems with different azimuths. PV panels facing due south demonstrate greater parameterization errors than other scenarios. Specifically, the MAE of the tilt result is 7.9\(^\circ\) for south-facing PV panels, compared with 3.0\(^\circ\), 4.1\(^\circ\), and 5.1\(^\circ\) for north, east, and west facing systems. The same trend can also be found from the inference results of azimuth and orientation. This phenomenon can be caused by the fact that as PV tilt varies, the E\(_{\text{poa}}\) incident on south-facing PV panels remains relatively stable. As illustrated in Fig. 1 (b), the E\(_{\text{poa}}\) curves of south-oriented PV panels show stronger similarity than others, which could cause errors in the curve mismatch evaluation process.

To have a more in-depth insight into the proposed method, the east-facing 30\(^\circ\)-tilted PV panel, a common scenario in the Netherlands, is taken as an example. Fig. 4 depicts the monthly results and their overlapping. As shown in Fig. 4 (a), a seasonal characteristic can be noticed from monthly inference results due to the seasonally dependent effects. However, no result is obtained in January and December. This is because solar zenith does not satisfy the requirement of less than 70\(^\circ\) at any time during the clearest days in the two months, so that curve mismatch assessment is not conducted then. Through merging monthly results, the contour plot in Fig. 4 (b) counts the overlapping times of different PV orientations. The central area has a maximum number of 10 overlaps, centered at 92\(^\circ\) azimuth and 28\(^\circ\) inclination. It can therefore be concluded that overlapping helps narrow the range of monthly curve matching results and eliminate seasonally dependent errors to obtain more accurate inference results.

### Table 3
The clearest day of each month and corresponding daily diffuse fraction in SolarBEAT in 2017.

| Day     | Jan 19 | Feb 14 | Mar 27 | Apr 9 | May 26 | Jun 19 | Jul 21 | Aug 27 | Sep 23 | Oct 15 | Nov 6 | Dec 28 |
|---------|--------|--------|--------|-------|--------|--------|--------|--------|--------|--------|-------|--------|
| \(k_d\) | 0.18   | 0.30   | 0.15   | 0.23  | 0.12   | 0.13   | 0.18   | 0.23   | 0.31   | 0.23   | 0.33  | 0.49   |

The derivation results in Table 4 show that, the mean absolute errors (MAE) of tilt, azimuth, and orientation for 21 simulated PV panels in this case study are 4.8\(^\circ\), 3.1\(^\circ\), and 5.0\(^\circ\). One thing to note here is that the accuracy of this method varies for PV systems with different azimuths. PV panels facing due south demonstrate greater parameterization errors than other scenarios. Specifically, the MAE of the tilt result is 7.9\(^\circ\) for south-facing PV panels, compared with 3.0\(^\circ\), 4.1\(^\circ\), and 5.1\(^\circ\) for north, east, and west facing systems. The same trend can also be found from the inference results of azimuth and orientation. This phenomenon can be caused by the fact that as PV tilt varies, the E\(_{\text{poa}}\) incident on south-facing PV panels remains relatively stable. As illustrated in Fig. 1 (b), the E\(_{\text{poa}}\) curves of south-oriented PV panels show stronger similarity than others, which could cause errors in the curve mismatch evaluation process.

To have a more in-depth insight into the proposed method, the east-facing 30\(^\circ\)-tilted PV panel, a common scenario in the Netherlands, is taken as an example. Fig. 4 depicts the monthly results and their overlapping. As shown in Fig. 4 (a), a seasonal characteristic can be noticed from monthly inference results due to the seasonally dependent effects. However, no result is obtained in January and December. This is because solar zenith does not satisfy the requirement of less than 70\(^\circ\) at any time during the clearest days in the two months, so that curve mismatch assessment is not conducted then. Through merging monthly results, the contour plot in Fig. 4 (b) counts the overlapping times of different PV orientations. The central area has a maximum number of 10 overlaps, centered at 92\(^\circ\) azimuth and 28\(^\circ\) inclination. It can therefore be concluded that overlapping helps narrow the range of monthly curve matching results and eliminate seasonally dependent errors to obtain more accurate inference results.

### Table 4
Inference results of typical simulated PV systems.

| Actual metadata | Inference result | Error |
|----------------|-----------------|-------|
| Tilt (\(^\circ\)) | Azimuth (\(^\circ\)) | Tilt (\(^\circ\)) | Azimuth (\(^\circ\)) | Tilt (\(^\circ\)) | Azimuth (\(^\circ\)) | Orientation |
| 15 | 90 | 13.7 | 89.1 | 1.3 | 0.9 | 1.4 | |
| 30 | 90 | 28.0 | 92.0 | 2.0 | 2.0 | 2.2 | |
| 45 | 90 | 40.9 | 89.1 | 4.1 | 0.9 | 4.1 | |
| 60 | 90 | 54.8 | 89.0 | 5.2 | 1.0 | 5.3 | |
| 90 | 90 | 82.3 | 87.4 | 7.7 | 2.6 | 8.1 | |
| 15 | 180 | 6.6 | 170.2 | 8.4 | 9.8 | 8.6 | |
| 30 | 180 | 25.0 | 177.0 | 5.0 | 3.0 | 5.2 | |
| 45 | 180 | 41.3 | 178.3 | 3.7 | 1.7 | 3.9 | |
| 60 | 180 | 46.8 | 179.6 | 13.2 | 0.4 | 13.2 | |
| 90 | 180 | 81.0 | 179.9 | 9.0 | 0.1 | 9.0 | |
| 15 | 270 | 12.7 | 269.1 | 2.3 | 0.9 | 2.3 | |
| 30 | 270 | 26.4 | 270.3 | 3.6 | 0.3 | 3.6 | |
| 45 | 270 | 39.5 | 272.5 | 5.5 | 2.5 | 5.8 | |
| 60 | 270 | 52.0 | 273.0 | 8.0 | 3.0 | 8.4 | |
| 90 | 270 | 84.1 | 271.5 | 5.9 | 1.5 | 6.1 | |
| 15 | 0 | 13.2 | 2.0 | 1.8 | 2.0 | 1.8 | |
| 30 | 0 | 30.0 | 1.0 | 0.0 | 1.0 | 0.5 | |
| 45 | 0 | 37.0 | 1.0 | 8.0 | 1.0 | 8.0 | |
| 60 | 0 | 59.5 | 0.1 | 0.5 | 0.1 | 0.5 | |
| 90 | 0 | 85.2 | 2.0 | 4.8 | 2.0 | 5.2 | |
| 0 | N/A | 1.2 | 28.5 | 1.2 | N/A | 1.2 | |

The MAE is 4.8\(^\circ\), 3.1\(^\circ\), and 5.0\(^\circ\). One thing to note here is that the accuracy of this method varies for PV systems with different azimuths. PV panels facing due south demonstrate greater parameterization errors than other scenarios. Specifically, the MAE of the tilt result is 7.9\(^\circ\) for south-facing PV panels, compared with 3.0\(^\circ\), 4.1\(^\circ\), and 5.1\(^\circ\) for north, east, and west facing systems. The same trend can also be found from the inference results of azimuth and orientation. This phenomenon can be caused by the fact that as PV tilt varies, the E\(_{\text{poa}}\) incident on south-facing PV panels remains relatively stable. As illustrated in Fig. 1 (b), the E\(_{\text{poa}}\) curves of south-oriented PV panels show stronger similarity than others, which could cause errors in the curve mismatch evaluation process.

To have a more in-depth insight into the proposed method, the east-facing 30\(^\circ\)-tilted PV panel, a common scenario in the Netherlands, is taken as an example. Fig. 4 depicts the monthly results and their overlapping. As shown in Fig. 4 (a), a seasonal characteristic can be noticed from monthly inference results due to the seasonally dependent effects. However, no result is obtained in January and December. This is because solar zenith does not satisfy the requirement of less than 70\(^\circ\) at any time during the clearest days in the two months, so that curve mismatch assessment is not conducted then. Through merging monthly results, the contour plot in Fig. 4 (b) counts the overlapping times of different PV orientations. The central area has a maximum number of 10 overlaps, centered at 92\(^\circ\) azimuth and 28\(^\circ\) inclination. It can therefore be concluded that overlapping helps narrow the range of monthly curve matching results and eliminate seasonally dependent errors to obtain more accurate inference results.

### 4. Empirical validation based on actual PV systems

#### 4.1. Details of monitored PV sites

In this section, 13 real PV systems installed on the roofs of the standardized and pre-fabricated dwellings ‘Morgen Wonen’ distributed across the Netherlands are taken as a case study. For each dwelling, a smart metering system is installed on-site to monitor energy production and consumption. As shown from the geographical distribution in Fig. 5, these PV applications are located 9.5 km to 195 km from the weather station, SolarBEAT. The technical specifications of these identical PV systems are summarized in Table 5. Each system has a capacity of 6.63 kWp consisting of 24 modules rated at 265 Wp. These PV panels installed...
in different cities all have the same tilt of 42°, but with varying azimuths which are measured from satellite images. The metadata is presented in Table 6. As the input to this parameterization method, hourly PV yield in 2018 is downloaded from the online monitoring database of ‘Morgen Wonen’, and irradiance data is collected from SolarBEAT.

Compared with the simulated PV systems in the previous section, the PV systems studied in this section are installed in the real built environment, subjected to actual pre-photovoltaic losses, module losses and system losses. The empirical validation not only tests the usability of the proposed method considering the uncertainty of PV performance in real working environments, but also investigates the accuracy of parameterization results when PV sites are further away from the irradiance measurement site.

### Table 5
Technical specifications of distributed PV systems in empirical validation.

| Component          | Module                  | Number of modules | Modules in series | Number of parallel strings | Module capacity | PV module type          | Inverter type                  | Inverter type                      |
|--------------------|-------------------------|-------------------|-------------------|---------------------------|----------------|-------------------------|----------------------------------|-----------------------------------|
| Module type        | BLU 255–280 Wp          | 24                | 12                | 2                         | 265 Wp         | Multicrystalline        | Growatt 6000MTLP - US            | Growatt 6000MTLP - US              |

### Table 6
Characteristics of the distributed PV systems.

| PV system ID | Distance from the weather station (km) | Tilt (°) | Azimuth (°) |
|--------------|---------------------------------------|----------|-------------|
| 1            | 9.5                                   | 42       | 227.0       |
| 2            | 13                                    | 42       | 242.0       |
| 3            | 29.9                                  | 42       | 151.8       |
| 4            | 43.1                                  | 42       | 222.0       |
| 5            | 44.7                                  | 42       | 125.0       |
| 6            | 76.2                                  | 42       | 105.0       |
| 7            | 100                                   | 42       | 125.0       |
| 8            | 109.2                                 | 42       | 212.0       |
| 9            | 110.5                                 | 42       | 122.3       |
| 10           | 130                                   | 42       | 159.1       |
| 11           | 171                                   | 42       | 164.0       |
| 12           | 171                                   | 42       | 248.5       |
| 13           | 195                                   | 42       | 132.7       |

### Table 7
Daily solar radiation and diffuse fraction of the clearest day of each month in SolarBEAT in 2018.

| Timestamp  | \(DHI_{day}(kWh/m^2)\) | \(GHI_{day}(kWh/m^2)\) | \(k_d\) |
|------------|--------------------------|--------------------------|---------|
| Jan 14     | 0.50                     | 1.23                     | 0.41    |
| Feb 25     | 0.52                     | 3.40                     | 0.15    |
| Mar 20     | 0.84                     | 4.47                     | 0.19    |
| Apr 19     | 1.12                     | 6.21                     | 0.18    |
| May 6      | 0.82                     | 7.42                     | 0.11    |
| Jun 29     | 1.01                     | 8.18                     | 0.12    |
| Jul 2      | 0.80                     | 8.42                     | 0.10    |
| Aug 6      | 0.95                     | 6.96                     | 0.14    |
| Sep 27     | 0.50                     | 4.23                     | 0.12    |
| Oct 5      | 0.58                     | 3.70                     | 0.16    |
| Nov 17     | 0.36                     | 1.63                     | 0.22    |
| Dec 24     | 0.27                     | 1.08                     | 0.25    |

### 4.2. Parameterization results of the empirical cases

The empirical validation is based on the measurement data for 2018. The minimum \(k_d\) per month calculated from Eq. (1) is presented in Table 7. Parameterization results of the 13 distributed PV systems are shown in Table 8. The MAE of tilt, azimuth, and orientation results are 4.3°, 4.5°, and 5.4° respectively.

Fig. 6 compares the error distribution of the parameterization approach in the above two case studies. The box plot depicts the 25th and 75th quantiles at the bottom and top edges of boxes and the median as the band inside boxes. It shows that the proposed method has similar accuracy for simulated and actual PV panels. However, the errors of tilt and azimuth for simulated PV prototypes exhibit greater variability. One possible reason is that the real PV panels studied in the empirical validation part are mounted on a standardized roof with a fixed inclination of 42°. In comparison, the simulated PV prototypes contain a wide range of tilt and azimuth angles, including south-facing modules and horizontal modules, which exhibit relatively higher errors in parameterization.

### 5. Sensitivity analysis of influencing factors

The accuracy of the proposed parameterization method is assessed in the above case studies. However, in practice, uncertainties and other limiting factors may exist in many aspects, such as the time resolution and duration of input data, the distance between PV sites and the irradiance measurement site. To have a deeper insight into the usability of this method in actual situations, the impact of these uncertainties on parameterization results is investigated in this section.

#### 5.1. Influence of acceptable range of monthly results

In the proposed method, during the curve mismatch evaluation process, the best 1% of tilt and azimuth combinations with the smallest RMSE are accepted as monthly results. The acceptable range should reasonably reflect the extent of secondary and seasonal effects. These effects include the negative impact of temperature increases on the energy conversion efficiency of PV cells, due to the intrinsic property of silicon and the influence of angle of incidence on reflection, soiling and inverter losses (Dubey et al., 2013). Taking a PV panel with 42° tilt and 212° azimuth as an example, the overlapping results adopting different acceptable ranges are presented in Fig. 7. The actual PV orientation value is marked by a black dot in the contour figure. According to the method, the average of the most overlapping PV orientations is calculated as the parameterization result. As shown in Fig. 7, for a small acceptable range of 0.5%, only very few overlapping areas are obtained, and the actual PV orientation is not included in the most overlapping area because the impact of seasonal effects is underestimated. On the other hand, as the acceptable range increases from 1% to 5%, the error...
in PV orientation results increases from 3.7° to 5.8°. It illustrates that an appropriate acceptable range is essential for inference accuracy. It should be neither too narrow to exclude correct values, nor too large to bring errors and excessive uncertainty.

To evaluate the sensitivity of the method to monthly acceptability ranges, and to verify the feasibility of the currently adopted 1% threshold in practice, parameterization is conducted on the actual PV systems in the empirical study with the acceptable range varying from 0.5% to 10%. As shown in Fig. 8, the variance of parameterization error increases drastically when the acceptable range exceeds 1%. Further more, compared with 0.5%, the 1% acceptable range demonstrates higher accuracy in azimuth inference and lower variance in the inference accuracy. According to the results, a range of 1% is found to be most appropriate, considering both accuracy and stability aspects.

5.2. Influence of the distance between the irradiance measurement site and PV panels

Compared with existing parameterization approaches, one of the significant advantages of the proposed method is that it does not rely on the availability of on-site irradiance measurements. In Section 4, irradiance data collected in SolarBEAT is used to parameterize PV panels 9.5 km to 195 km apart with a mean orientation error of 5.4°. The error of inference results for PV panels at different distances is presented in Fig. 9. The bar chart shows that the smallest error of 1.6° is obtained from the 110.5 km away PV system, while the maximum of 10.1° is achieved from the 130 km away PV system. As distance increases, the accuracy of the parameterization results does not change significantly. A key feature that enables the proposed method to maintain accuracy for remote PV systems is that only the clearest day of each month is selected for parameterization. On these clear days, without passing clouds, the irradiance mismatch between weather station and PV sites is significantly reduced. The slight variation in the parameterization errors among PV systems can be caused by the different weather conditions on site. Other factors, such as noise in the measurement data, module soiling, and different maintenance efforts, also cause fluctuations in the accuracy of the results.

5.3. Influence of the time resolution of data

In theory, smart meters can record real-time electricity generation at 1-minute intervals or finer. However, these data are not available in most actual cases, due to bandwidth limitations and the cost of data storage. As a result, much of the publicly available PV yield data is at a
30-minute or 1-hour temporal resolution (Abdulla et al., 2017). In addition, most of the publicly available irradiance datasets also adopt an hourly time interval. However, a low temporal resolution could possibly pose a risk to parameterization accuracy. In the empirical study in Section 4, PV output is recorded every 15 min, while SolarBEAT collects irradiance data every minute. To investigate the sensitivity of parameterization accuracy to the temporal resolution of input, measurement data is aggregated over several time intervals: 15 min, 30 min, and 60 min. The distribution of parameterization errors at different temporal resolutions is shown in Fig. 10. In general, the orientation parameterization error decreases slightly as the temporal resolution decreases from 15 min to 1 h. With hourly input data, the variance of parameterization errors also reaches minimum values. It indicates that a finer time resolution of 15 min does not improve accuracy, and the widely-available hourly data is recommended for lower parameterization errors.

5.4. Influence of seasonal effects

In the proposed method, monthly results across the year are overlapped to capture various solar inclination angles, and to reduce the influence introduced by seasonal variation in the performance of PV systems. However, for newly installed PV systems, the data of power generation lasting for one year is not available. To explore the possibility of extending the method to newly installed PV panels with only several months of recorded yield data, the accuracy of monthly results is examined in this section.

As mentioned, the best-matching 1% of PV orientations are accepted as the monthly results in this method. To evaluate the parameterization accuracy of each month, the deviation is calculated between the average value of monthly inference results and the exact PV orientation. As
shown in Fig. 11, generally, parameterization accuracy reaches its highest level in May, June, and July, and lowest in March. The minimum orientation error achieved in July has a median of 5.4°, compared to the maximum of 33.2° achieved in March. It can be concluded that, for the cases examined, orientation results with errors less than 10° can be derived even though input data is only available in one month among May, June, and July. The high accuracy achieved in these three months is likely explained by the fact that the working conditions of the PV systems during this period are close to SRC. According to Eq. (6)-(11), PV yield and in-plane irradiance have a non-linear relationship with seasonal characteristics, which causes seasonally dependent errors in monthly curve matching results. As has occurred in May, June, and July, the seasonal error is attenuated when the PV working environment approximates SRC. According to Eq. (6)-(11), PV yield and in-plane irradiance have a non-linear relationship with seasonal characteristics, which causes seasonally dependent errors in monthly curve matching results. As has occurred in May, June, and July, the seasonal error is attenuated when the PV working environment approximates SRC. Also note in Fig. 11 that no result is obtained in November, December, and January. This is because the requirement for solar zenith of less than 70° is not satisfied at any time of the clearest days in these months. The monthly parameterization error of azimuth reveals sharp variations across the year, with the median monthly error ranging from 5.4° to 33.2°. Higher errors in specific months do not mean they are unfavorable for parameterization. On the contrary, the fluctuation of monthly results reflects the seasonally dependent relationship between $E_{poa}$ and PV yield, which forms the basis for the proposed method to eliminate seasonal effects through overlapping.

5.5. Influence of using satellite-derived irradiance

In Section 4, parameterization of PV orientation has been conducted on PV roofs up to 195 km away from the ground-based weather station with a mean error of 5.4°. However, for places where solar measurement stations are sparsely distributed, a greater distance between the target PV system and irradiance measurement site could pose a risk to the accuracy of parameterization. As an alternative source of solar irradiance data, satellite-derived irradiance with extensive spatial coverage is receiving increasing interest (Yang and Bright, 2020). Comparisons of ground-based vs. satellite-derived irradiance have been performed for many applications in solar energy research, such as solar resource assessment and forecasting (Yang and Perez, 2019). In this section, the feasibility of the proposed method to infer PV orientation from satellite irradiance data is investigated.

The Copernicus Atmosphere Monitoring Service (CAMS) radiation service which provides time series of the commonly used solar irradiance components is taken as the satellite-derived irradiance dataset in this section. CAMS data is publicly available from different platforms (Lundstrom, 2016; Schroedter-Homscheidt et al., 2016). From the dataset, hourly GHI, DHI and DNI in 2018 are obtained for the exact location of each PV system studied in Section 4. Taking the hourly satellite irradiance data and measured PV output as input, the proposed method is applied on the distributed PV systems in the Netherlands. Monthly clear days for each system are detected by entering CAMS data to Eq. (1). Fig. 12 compares the parameterization quality based on CAMS data with the estimation result in Section 4 which is calculated from ground irradiance measurement data.

Fig. 12 shows that the CAMS-based parameterization result exhibits greater error and uncertainty compared with the result generated from ground-based data. The reduction in the parameterization quality can be caused by the fact that satellite-derived irradiance has lower accuracy compared with ground-based measurement. Higher bias in the CAMS dataset deteriorates the accuracy of the proposed parameterization approach. It is found that although the ground-based solar measurement stations are scarce, they are still preferred to estimate panel orientations of PV systems as far as 195 km away from the measurement station. It should be noted that many methods have been proposed to correct the
5.6. Influence of the location of PV systems

In the empirical validation presented in Section 4, the proposed method has been tested on rooftop PV systems distributed across the Netherlands. To further explore its applicability in other regions, this section presents the results of an empirical test that was conducted on a PV system in Portland, USA. Hourly PV output and on-site solar irradiance measurements are obtained from a publicly available dataset, the University of Oregon Solar Radiation Monitoring Laboratory (UO SRML) solar monitoring network. According to the detailed description in the dataset, the selected PV system is free from local shading with the solar zenith angle lower than 70°. The exact tilt and azimuth angles are 33° and 200° respectively.

Using the proposed parameterization method, the monthly inference results and their overlapping are obtained as shown in Fig. 13. In the contour plot of Fig. 13 (b), the central area has a maximum number of 10 overlaps, centered at 30.4° tilt and 199.1° azimuth. The proposed method leads to an estimation error of panel orientation of 2.6° for this PV system in the United States. Compared with the Dutch scenario illustrated in Fig. 4, monthly inference results of January and December are not empty in Fig. 13 (a). This is because the Portland case has a lower latitude of 45.5° N as compared to 51.5° N of the Dutch case. As a result, solar zenith angle satisfies the requirement of less than 70° during part of the clearest days in the two months in Portland, and curve matching is conducted then. Through comparing the slightly different working mechanism of the proposed method in different locations, it can be drawn that validation and minor parameter adjustment in the algorithms is necessary to extend the method to other locations in the world.

The source code of the proposed parameterization algorithm written in Python can be accessed from the link https://gitlab.tue.nl/bp-tue/inference-of-unknown-tilt-and-azimuth, where an east-facing 45° tilted PV module is taken as a demonstration case. Through this example, other researchers can understand the procedures of the method more easily, and extend it to the target PV panels of their own with little modification.

6. Conclusion

This paper has presented a data-driven method to parameterize tilt and azimuth angles of distributed PV systems using widely available input data. Based on the seasonally dependent correlation between characteristic shapes of in-plane irradiance and PV yield, curve matching is conducted for the clearest day of each month. Monthly parameterization results are then overlapped to obtain the final inference result. This method is tested in both a simulation study using notional PV panels with a wide variety of orientations and an empirical validation using monitoring data from actual PV roofs distributed across the Netherlands. Taking hourly irradiance data collected by the meteorological sensors in SolarBEAT as input, the MAE in azimuth and tilt for empirical PV systems are 4.5° and 4.3° respectively, which are comparable to the accuracy of existing methods. However, the proposed method shows significant improvements in scalability potential. Compared with existing parameterization methods that require historical PV yield and on-site irradiance data with high temporal resolution, the proposed method only needs hourly PV output and off-site irradiance data for no more than one year. For a PV panel up to 200 km away from the weather station, a maximum PV orientation inference error of less than 10.1° can be obtained. On the other hand, as an essential part of the proposed method, which avoids the parameterization errors brought by the modeling process and the need to estimate unknown PV system properties. With these key features, the method has the advantage of easy implementation and broad applicability to significantly improve the lack of metadata in PV databases. A more in-depth insight into the method is obtained to investigate the implication of time resolution, time duration, use of satellite-derived irradiance and other potential influencing factors on parameterization accuracy. It is found that for PV panels within 195 km of the weather station, the error of inference results does not change significantly. The proposed parameterization algorithm reveals higher accuracy using ground-based measurement data than using satellite-derived irradiance obtained from CAMS. In addition, a finer time resolution of 15 min does not provide gains in parameterization accuracy. The use of hourly input data is therefore recommended in this method for its wide availability and low parameterization error.

The novel method presented in this paper features several main limitations. Firstly, only 42° tilted PV panels were considered in the empirical validation, which was restricted by the standardized roof structure of the prefabricated dwellings that were considered. The
seasonal effects on PV performance might, however, vary with different PV orientations in real working conditions, which could potentially affect the accuracy of parameterization results. Future tests with the proposed method would consider real PV panels with comprehensive angular characteristics. Another limitation is that, in this research, the feasibility to eliminate monthly parameterization errors caused by seasonal effects by means of curve matching and overlapping is only investigated for Dutch cases and a case in the United States. The applicability of current parameter settings in the method for different climate conditions needs to be further explored. In addition, in the empirical validation, no significant increase in the parameterization errors is noticed for the most remote case, 195 km away. The validity of this method over greater distances could be further tested. Moreover, PV panels studied in this research are free from partial shading by nearby buildings or other obstacles with the solar zenith angle less than 70°. However, in many practical cases, it is quite common that shading effects will result in a mismatch between PV yield and off-site irradiance measurements, potentially causing significant errors in evaluating curve mismatch and parameterization outcomes.

There are various opportunities to improve the proposed method in future work. One such direction could be improving the parameterization accuracy by interpolating ground-based irradiance measurements or using corrected satellite-derived irradiance data. In this research, hourly irradiance data measured by the weather station or obtained from the CAMS dataset is used to parameterize PV panels 9.5 km to 195 km away from the weather station. However, there are alternative methods to estimate solar radiation for locations without direct measurement through interpolating measurement data of several nearest weather stations or correcting the systematic bias in the satellite-derived irradiance (Polo et al., 2016). Hence, it is a topic of future development to improve the parameterization accuracy with other irradiance data sources. Another improvement direction in future work is to extend the proposed method to PV systems under local shading. Methods have, for example, been developed to identify locally shaded periods of PV installations, using measured AC power and regional irradiance data (Bognár et al., 2018; Branco et al., 2020). Thus, the potential opportunity to extend the proposed method is to detect shadow timestamps from measurement data and explore the parameterization of PV modules that experience daytime shadow.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material
Supplementary data to this article can be found online at https://doi.org/10.1016/j.solener.2020.09.077.

References
Abdulla, K., Steer, K., Wirth, A., Hoog, J. de, Halgamuge, S., 2017. The importance of PV orientations in real working conditions, which could potentially affect the accuracy of parameterization results. Future tests with the proposed method would consider real PV panels with comprehensive angular characteristics. Another limitation is that, in this research, the feasibility to eliminate monthly parameterization errors caused by seasonal effects by means of curve matching and overlapping is only investigated for Dutch cases and a case in the United States. The applicability of current parameter settings in the method for different climate conditions needs to be further explored. In addition, in the empirical validation, no significant increase in the parameterization errors is noticed for the most remote case, 195 km away. The validity of this method over greater distances could be further tested. Moreover, PV panels studied in this research are free from partial shading by nearby buildings or other obstacles with the solar zenith angle less than 70°. However, in many practical cases, it is quite common that shading effects will result in a mismatch between PV yield and off-site irradiance measurements, potentially causing significant errors in evaluating curve mismatch and parameterization outcomes.

There are various opportunities to improve the proposed method in future work. One such direction could be improving the parameterization accuracy by interpolating ground-based irradiance measurements or using corrected satellite-derived irradiance data. In this research, hourly irradiance data measured by the weather station or obtained from the CAMS dataset is used to parameterize PV panels 9.5 km to 195 km away from the weather station. However, there are alternative methods to estimate solar radiation for locations without direct measurement through interpolating measurement data of several nearest weather stations or correcting the systematic bias in the satellite-derived irradiance (Polo et al., 2016). Hence, it is a topic of future development to improve the parameterization accuracy with other irradiance data sources. Another improvement direction in future work is to extend the proposed method to PV systems under local shading. Methods have, for example, been developed to identify locally shaded periods of PV installations, using measured AC power and regional irradiance data (Bognár et al., 2018; Branco et al., 2020). Thus, the potential opportunity to extend the proposed method is to detect shadow timestamps from measurement data and explore the parameterization of PV modules that experience daytime shadow.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material
Supplementary data to this article can be found online at https://doi.org/10.1016/j.solener.2020.09.077.

References
Abdulla, K., Steer, K., Wirth, A., Hoog, J. de, Halgamuge, S., 2017. The importance of temporal resolution in evaluating residential energy storage. In: 2017 IEEE Power & Energy Society General Meeting. pp. 1–5. https://doi.org/10.1109/PESGM.2017.8274019.

Abdulla, K., Steer, K., Wirth, A., Hoog, J. de, Halgamuge, S., 2017. The importance of temporal resolution in evaluating residential energy storage. In: 2017 IEEE Power & Energy Society General Meeting. pp. 1–5. https://doi.org/10.1109/PESGM.2017.8274019.

Abdulla, K., Steer, K., Wirth, A., Hoog, J. de, Halgamuge, S., 2017. The importance of temporal resolution in evaluating residential energy storage. In: 2017 IEEE Power & Energy Society General Meeting. pp. 1–5. https://doi.org/10.1109/PESGM.2017.8274019.

Abdulla, K., Steer, K., Wirth, A., Hoog, J. de, Halgamuge, S., 2017. The importance of temporal resolution in evaluating residential energy storage. In: 2017 IEEE Power & Energy Society General Meeting. pp. 1–5. https://doi.org/10.1109/PESGM.2017.8274019.

Abdulla, K., Steer, K., Wirth, A., Hoog, J. de, Halgamuge, S., 2017. The importance of temporal resolution in evaluating residential energy storage. In: 2017 IEEE Power & Energy Society General Meeting. pp. 1–5. https://doi.org/10.1109/PESGM.2017.8274019.

Abdulla, K., Steer, K., Wirth, A., Hoog, J. de, Halgamuge, S., 2017. The importance of temporal resolution in evaluating residential energy storage. In: 2017 IEEE Power & Energy Society General Meeting. pp. 1–5. https://doi.org/10.1109/PESGM.2017.8274019.
Reda, I., Andreas, A., 2008. Solar Position Algorithm for Solar Radiation Applications (Revised). United States. https://doi.org/10.2172/15003974.
Reno, M.J., Hansen, C.W., Stein, J.S., 2012. Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis. https://doi.org/10.2172/1039404.
Riffonneau, Y., Bacha, S., Barruel, F., Ploix, S., 2011. Optimal power flow management for grid connected PV systems with batteries. IEEE Trans. Sustain. Energy 2, 309–320. https://doi.org/10.1109/TSTE.2011.2114901.
Ruelle, V.D., Jeppesen, M., Brear, M., 2016. Rooftop PV Model Technical Report. Melbourne.
Ruf, H., 2016. Computation of the Load Flow at the Transformer in Distribution Grids with a Significant Number of Photovoltaic Systems using Satellite-derived Solar Irradiance Data.
Ruf, H., Stetter, D., Lanig, S., Hoyer-Klick, C., Meier, F., Heilscher, G., 2012. Active grid planning based on solar power roof potential analysis. In: 27th European Photovoltaic Solar Energy Conference and Exhibition, pp. 3782–3787. https://doi.org/10.4229/27thEUPVSEC2012-SCD.7.2.
Saint-Drenan, Y.M., Bofinger, S., Fritz, R., Vogt, S., Good, G.H., Dobschinski, J., 2015. An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation. Sol. Energy 120, 479–493.
Saracoglu, B., 2018. Solar star projects SAM version 2017.9.5 PVwats version 5 model case study & validation. Int. J. Energy Appl. Technol. 5, 13–28. https://doi.org/10.31593/ijeat.403748.
Schroedter-Homscheidt, M., Wandji, W., Killius, N., Wald, L., Wey, E., 2016. The Copernicus Atmosphere Monitoring Service (CAMS) Radiation Service in a nutshell.
Silventre, S., Chouder, A., Karatepe, E., 2013. Automatic fault detection in grid connected PV systems. Sol. Energy 94, 119–127. https://doi.org/10.1016/j.solener.2013.05.001.
Valckenborg, R.M.E., Hensen, J.I.M., Folkerts, W., de Vries, A., 2015. Characterization of Bipv(T) applications in research facility ‘Solarbeat’. Proceedings of EU PVSEC.
Vimpari, J., Junnila, S., 2017. Evaluating decentralized energy investments: Spatial value of on-site PV electricity. Renew. Sustain. Energy Rev. 70, 1217–1222. https://doi.org/10.1016/j.rser.2016.12.023.
Yagil, G.M., Yang, D., Gandhi, O., Srinivasan, D., 2020. Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance?, 114122 Appl. Energy 259. https://doi.org/10.1016/j.apenergy.2019.114122.
Yang, D., 2020. Choice of clear-sky model in solar forecasting. J. Renew. Sustain. Energy 12. https://doi.org/10.1063/5.0003495.
Yang, D., 2016. Solar radiation on inclined surfaces: Corrections and benchmarks. Sol. Energy 136, 288–302. https://doi.org/10.1016/j.solener.2016.06.062.
Yang, D., Bright, J.M., 2020. Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years. Sol. Energy 1–17. https://doi.org/10.1016/j.solener.2020.04.016.
Yang, D., Perez, R., 2019. Can we gauge forecasts using satellite-derived solar irradiance? J. Renew. Sustain. Energy 11. https://doi.org/10.1063/1.5087588.