Renewable Energy 125 (2018) 306–318

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Solar irradiation from the energy production of residential PV systems

Cédric Bertrand [a, *], Caroline Housmans [a], Jonathan Leloux [b], Michel Journée [a]

[a] Royal Meteorological Institute of Belgium, Brussels, Belgium
[b] Universidad Politécnica de Madrid, Madrid, Spain

ARTICLE INFO

Article history:
Received 24 October 2017
Received in revised form
11 January 2018
Accepted 6 February 2018
Available online 26 February 2018

Keywords:
Photovoltaic system
Solar radiation
Decomposition model
Transposition model
Remote sensing
Ground measurements
Crowdsourcing

ABSTRACT

Considering the dense network of residential photovoltaic (PV) systems implemented in Belgium, the paper evaluates the opportunity of deriving global horizontal solar irradiation data from the electrical energy production registered at PV systems. The study is based on one year (i.e. 2014) of hourly PV power output collected at a representative sample of roughly 1500 residential PV installations. Validation is based on ground-based measurements of solar radiation performed within the network of radiometric stations operated by the Royal Meteorological Institute of Belgium and the method's performance is compared to the satellite-based retrieval approach.

Our results indicate that the accuracy of the derived solar irradiation data depends on a number of factors including the efficiency of the PV system, the weather conditions, the density of PV systems that can be used for the tilt to horizontal conversion, other data sources that can be accessed to complement the PV data. In particular, the computed solar irradiation data degrade as the information about the orientation and tilt angles of the PV generator becomes more inaccurate. It is also found that there are certain sun positions (i.e. low solar elevations) for which the method fails to produce a valid estimation.

1. Introduction

Appropriate information on solar resources is very important for a variety of technological areas, such as: agriculture, meteorology, forestry engineering, water resources and in particular in the designing and sizing of solar energy systems. Traditionally, solar radiation is observed by means of networks of meteorological stations. However, costs for installation and maintenance of such networks are very high and national networks comprise only a few stations. Consequently the availability of solar radiation measurements has proven to be spatially and temporally inadequate for many applications.

Over the last decades, satellite-based retrieval of solar radiation at ground level has proven to be valuable for delivering a global coverage of the global solar irradiance distribution at the Earth's surface (e.g. Refs. [1–4]). The recent deployment of solar photovoltaic (PV) systems offers an opportunity to extend the in-situ characterization of the incoming solar radiation at the Earth's surface through the reverse engineering conversion of the energy production of the PV systems to global solar irradiation data.

During the last years, thanks to the newest developments in telecommunication technologies, an increasing amount of electrical (power, frequency, current, voltage etc …) data from the operation of PV systems has become available. These operational data are now typically available at time intervals of 10 min, opening the doors to a totally new set of data analysis procedures that were difficult to imagine just a few years ago. Smart energy meters are projected to become widespread in a couple of years, which will enable to monitor the energy production and other electrical outputs of millions of PV systems in the world. Therefore, in the near future, it will be possible to know the energy production of PV systems located at millions of points spread over the whole globe. This conveys the potential that solar irradiation data can also be retrieved at all these points.

However, using the energy production registered at PV systems as a solar irradiation sensor is not straightforward (e.g. Refs. [5,6]). It requires first to derive the solar irradiation from the energy production of the PV system (knowing that the power output of a PV system is not directly proportional to the solar irradiance that it receives). Second because modules are installed at a tilt angle close to local latitude to maximize array output (or at some minimum tilt to ensure self-cleaning by rain) this requires to convert the retrieved tilted global solar irradiance to horizontal. Towards this objective, operational data from a representative sample of Belgian

* Corresponding author.
E-mail address: Cedric.Bertrand@meteo.be (C. Bertrand).

https://doi.org/10.1016/j.renene.2018.02.036
0960-1481/

residential PV installations (note that tilt and azimuth of residential PV systems often depend on the particular rooftop on which they are installed, rather than being designed for optimal performance) have been considered to assess the performance of such an approach.

In Belgium, the total installed PV capacity has increased dramatically in recent years from 102.6 MW in 2008 (26.55 MW in 2007) to 3423 MW at the end of 2016, according to data collected by local renewable energy association APERe (http://www.apere.org/), which has combined the figures released by the country's three energy regulators Brugel (https://www.brugel.brussels/), VREG (http://www.vreg.be/) and CWaPE (http://www.cwape.be/). By contrast to the multi-megawatt PV plants installed in the South of Europe, residential and commercial rooftop PV systems with less than 10 kW capacity represent the majority of the installed capacity in Belgium. As an illustration, in 2016, the country installed about 170 MW across 25.000 PV systems.

2. Data

2.1. Residential PV systems data

This work is based on one year (i.e. 2014) of hourly PV power output collected at more than 2893 PV systems in Belgium installed from 2008 to 2013. PV generation data was collected via the Rtone company website (Rtone, http://www.rtone.fr). The PV energy production data provided by Rtone was monitored using the commercial Rbee Solar monitoring product, which measures the energy production with a smart energy meter at a 10-min time interval. The information concerning the PV systems (i.e. metadata) were supplied by their owners. Each PV system is localized by its latitude and longitude, completed with the corresponding altitude. The PV generator is characterized by its orientation and tilt angles, its total surface, and its total peak power. Additional information about the PV module manufacturer and model, inverter manufacturer and model, installer, year of installation, PV cell/module technology can also be provided.

Metadata has been subjected to several checks in order to isolate and remove as much erroneous information as possible. The standard set of filters employed before analyses are the following:

1. selection of single array systems since generation data cannot be decomposed into constituent arrays,
2. selection of systems located in Belgium,
3. selection of systems with and orientation between $-90°$ and $+90°$ from south and a tilt from horizontal smaller than $60°$.

In some cases, system details were investigated manually to verify to a good degree of confidence whether the systems orientation, tilt or peak power are correct. Indeed, metadata are prone to errors on part of the owner, by entering incorrect system parameters but also in some cases due to installer's conventions. As an example, tilt angle can be erroneously reported as $0°$ simply because 0 is used as default value in case of missing information. Another identified limitation is that many angles values are rounded as multiples of $5°$.

1470 systems have been selected and considered in this study after these preliminary checks (see Fig. 1 for the location of the selected PV systems). It is worth pointing out that due to availability reasons, most of the data comes from Wallonia and Brussels. As indicated in Table 1, the vast majority of the selected PV generators have a tilt angle between $20°$ and $50°$, which generally corresponds to the slope of the roofs on which they are installed ([7]).

2.2. Ground stations measurements

The Royal Meteorological Institute of Belgium (RMI) has a long term experience with ground-based measurements of solar radiation in Belgium (uninterrupted 30 min average measurements in Uccle since 1951, in Oostende since 1958, and in Saint-Hubert since 1959). Uccle is one of the 22 Regional Radiation Centres established within the WMO Regions. The incoming global horizontal solar irradiance is currently measured in 13 Automatic Weather Stations (AWS) in addition to the measurements performed in our main/reference station in Uccle (see Table 2).

At the reference station in Uccle, measurements of the global horizontal solar irradiance (GHI) are performed by Kipp & Zonen CM11-secondary standard pyranometers. The CM11 pyranometer has a thermopile detector and presents a reduced thermal offset (i.e. about 2 Wm^{-2} at 5 K/h temperature change). The pyranometer directional error (up to $80°$ with 1000 Wm^{-2} beam) is less than 10 Wm^{-2} and its spectral selectivity (300–1500 nm) is smaller than 2%. At the 13 RMI's AWS, GHI observations for the year 2014, were made with a Kipp & Zonen CNR1 net radiometer. It consists of a pyranometer (model type CM3 complying with the ISO Second Class Specification) and a pyrgeometer (model type CG3) pair that faces upward and a complementary pair that faces downward. The pyranometers and pyrgeometers measure shortwave and far infrared radiation, respectively. The CM3 pyranometer has a thermopile sensor and presents a limited thermal offset (i.e. ± 4 Wm^{-2} at 5 K/h temperature change). Its directional error (with 1000 Wm^{-2} beam) is ± 25 Wm^{-2} and its spectral selectivity (350–1500 nm) is $\pm 5\%$.

All ground measurements are made with a 5-s time step and then integrated to bring them to a 10-min time step. The 10-min data have undergone a series of automated quality control procedures ([8]) prior to be visually inspected and scrutinized in depth by a human operator for more subtle errors. Because our radiometric station in Saint-Hubert (AWS 6476) was known to operate deficiently in 2014, all GHI measurements from this station were discarded. The geographical location of the remaining 13 ground measurement sites is provided in Fig. 1 together with the selected residential PV systems. Note that because the data quality control revealed that the CNR1 net radiometer installed in the Buzenol station (i.e. AWS 6484) has only performed well intermittently during the year 2014, GHI measurements from this station were not used for validation purpose.

3. Methods

3.1. Conversion of PV system energy production to tilted global solar irradiation

The initial step of the approach consists in the derivation of global irradiance in plane of array, G_t, from the specific power output, P, of a PV system. Numerous models to calculate P from G_t exist in the literature (e.g. Refs. [9,10]). However, it is well known that the energy conversion efficiency of PV modules depends on a number of different influences. Losses in PV systems can be separated in capture losses and system losses (e.g. Refs. [11,12]). Capture losses are caused, e.g., by attenuation of the incoming light, temperature dependence, electrical mismatching, parasitic resistances in PV modules and imperfect maximum power point tracking. System losses are caused, e.g., by wiring, inverter, and transformer conversion losses. All these effects cause the module efficiency to deviate from the efficiency measured under Standard Test Condition (STC), which defines the rated or nominal power of a given module.

According to [13], the direct current (DC) power output of a PV

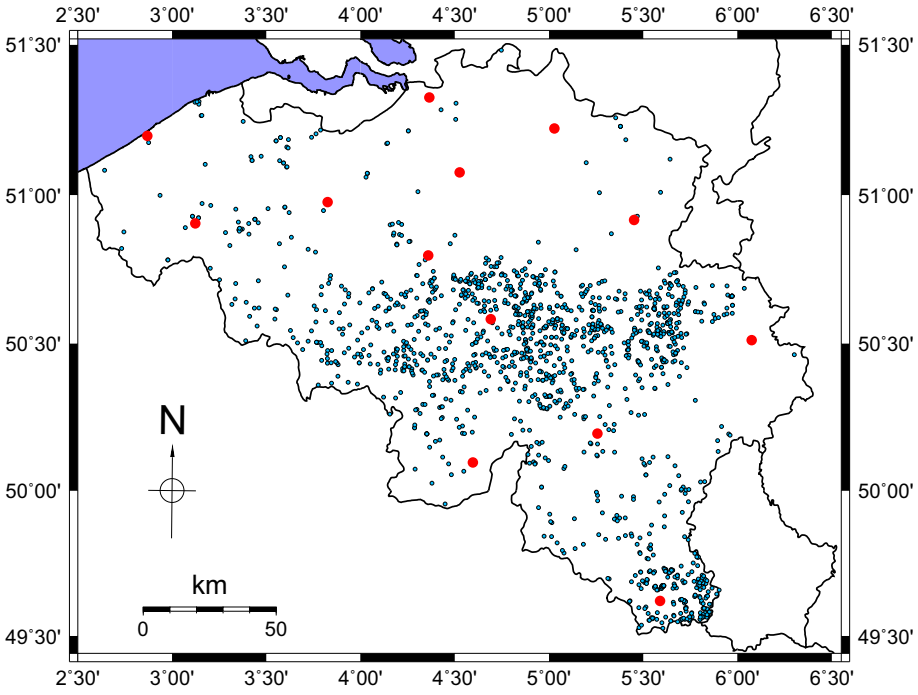

Fig. 1. Location of the selected 1470 residential PV systems within the Belgian territory. Red dots indicate the location of the RMI's ground radiometric stations considered in this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Distribution of the number of PV systems as a function of the orientation and tilt angle.

Tilt angle (°)	← East				Deviation from South (°)											West →			
	−90	−80	−70	−60	−50	−40	−30	−20	−10	0	10	20	30	40	50	60	70	80	90
0	0	0	0	0	0	1	0	0	0	3	0	0	0	0	0	0	0	0	0
10	4	0	0	0	0	6	0	2	2	7	0	2	0	0	2	0	0	0	0
20	4	2	1	2	0	19	10	12	10	20	8	5	9	0	18	2	4	3	0
30	13	4	14	4	0	50	24	24	18	71	9	21	14	0	47	7	6	6	0
40	67	8	21	13	0	188	38	36	30	223	16	32	23	0	147	24	12	15	0
50	4	2	2	2	0	12	3	6	4	16	2	3	1	0	17	5	6	1	0
60	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2
Location of the ground stations involved in the RMI solar radiation monitoring network in which the global horizontal radiation is recorded.

Station Code	Station Name	Lat. (°N)	Lon. (°E)	Alt. (m)
6407	MIDDELKERKE	51.198	2.869	3.0
6414	BEITEM	50.905	3.123	25.0
6434	MELLE	50.976	3.825	15.0
6438	STABROEK	51.326	4.365	5.0
6439	SINT-KATELIJNE-WAVER	51.076	4.526	10.0
6447	UCCLE	50.798	4.359	101.0
6455	DOURBES	50.096	4.596	233.0
6459	ERNAGE	50.583	4.691	157.0
6464	RETIE	51.222	5.028	21.0
6472	HUMAIN	50.194	5.257	296.0
6476	SAINT-HUBERT	50.040	5.405	557.0
6477	DIEPENBEEK	50.916	5.451	39.0
6484	BUZENOL	49.621	5.589	324.0
6494	MONT RIGI	50.512	6.075	673.0

generator can be properly described by:

$$P_{DC} = P^{\star}\frac{G_{eff}}{G^{\star}}\cdot\left(1+\kappa\left(T_c - T_c^{\star}\right)\right)\cdot\left(a + b\frac{G_{eff}}{G^{\star}} + c\ln\frac{G_{eff}}{G^{\star}}\right)\cdot f_{DC} \quad (1)$$

where the symbol * refers to STC [Irradiance: 1000 Wm^{-2}; Spectrum: AM 1.5; and cell temperature: 25 °C], P_{DC} is the DC power output of the PV generator (W), $P^{\star}$ is the nameplate DC power of the PV generator (i.e. power at maximum-power point, in W), G_{eff} is the effective global solar irradiance (Wm^{-2}) received by the PV generator (it takes into consideration the optical effects related to the solar incidence angle), $G^{\star}$ is the global solar irradiance under STC (Wm^{-2}), κ is the coefficient of power variation due to cell temperature (%/°C), T_c and $T_c^{\star}$ are respectively the cell temperatures under operating and STC conditions (°C), the three parameters a, b and c describe the efficiency dependence on irradiance, and f_{DC} is a coefficient that lumps together all the additional system losses in DC (e.g. technology-related issues, soiling and shading).

The first round brackets on the right-hand side of Eq. (1) takes into account the dependence of the efficiency with the temperature. The formulation goes a long way back ([14,15]) and considers that the PV module efficiency is affected by temperature, decreasing at a constant rate. Handling with this model formulation just requires standard information: $P^{\star}$ is the PV array rated power, which can be estimated as the product of the number of PV modules constituting the PV array multiplied by their nameplate STC power, and κ is routinely measured in the context of worldwide extended accreditation procedures: IEC Standard-61215 (2005) and IEC Standard-61646 (2008) for crystalline silicon and thin film devices, respectively. $P^{\star}$ and κ values are always included in PV manufacturer's data sheets or in more specific information as flash-reports.

The second round bracket on the right-hand side of Eq. (1) considers the efficiency dependence on irradiance. That was initially attempted by adding a base 10 logarithm ([14]) but it is better implemented by this empirical model proposed by Refs. [16,17] where a, b and c are empirical parameters. The efficiency increases with decreasing irradiance, due to series resistance effects, are represented by the term (b. $G_{eff}/G^{\star}$), providing $b \leq 0$, while the efficiency decreases with decreasing irradiance, due to parallel resistance effects, are represented by the term (c $\ln G_{eff}/G^{\star}$), providing $c \geq 0$.

The corresponding alternating current (AC) power output of the PV system from this DC power at the inverter entry is given by:

$$P_{AC} = P_{DC}\eta_{INV}f_{AC} \tag{2}$$

where P_{AC} is the AC power output of the PV generator, η_{INV} is the yield of the inverter, and f_{AC} is a coefficient that lumps together all the technology-related additional AC system losses.

The energy produced during a period of time T is finally given by:

$$E_{AC} = \int_{t=0}^{t=T} P_{AC}dt \tag{3}$$

To assess the technical quality of a particular PV system, energy performance indicators are obtained by comparing its actual production along a certain period of time with the production of a hypothetical reference system (of the same nominal power, installed at the same location, and oriented the same way). The **Performance Ration (PR)** which is the quotient of alternating current yield and the nominal yield of the generators direct current, is by far the most widely used performance indicator. It is defined mathematically as:

$$PR = \frac{\eta_{achieved}}{\eta_{spec}} = \frac{E_{AC}/G_t}{P_N^{\star}/G^{\star}} \tag{4}$$

where $P_N^{\star}$ is the nominal (or peak) DC power of the PV generator, understood as the product of the number of PV modules multiplied by the corresponding in-plane STC power. Because E_{AC}, $P_N^{\star}$ and G_t are given by the billing energy meter of the PV installation, the PV manufacturer and the integration of a solar irradiance signal, the PR value can be directly calculated. The difference between 1 and PR lumps together all imaginable energy losses (i.e. capture losses and system losses).

For a given PV system and site, the PR value tends to be constant along the years, as much as the climatic conditions tend to repeat. When sub-year periods are considered, the PR dependence on unavoidable and time-dependent losses requires corresponding correction in order to properly qualify the technical quality of a PV system. Based on Eqs. (1)–(3), we can reformulate Eq. (4) as:

$$\frac{1}{f_G \cdot f_T \cdot f_{AC} \cdot f_{PDC} \cdot f_{BOS}} \cdot PR = 1 \tag{5}$$

in which the losses have been lumped into five main categories:

1. f_G: PV module's yield in function of incident irradiance level,
2. f_T: PV module's yield in function of cell's temperature,
3. f_{AC}: yield of the conversion from DC to AC.
4. f_{PDC}: yield that represents the ratio of the real DC power and the rated DC power,
5. f_{BOS}: yield of the balance of system.

Three of these five losses parameters can be expressed analytically. Based on Eq. (1), the efficiency dependence on irradiance is:

$$f_G = a + b\frac{G_{eff}}{G^{\star}} + c\ln\frac{G_{eff}}{G^{\star}} \tag{6}$$

However, such a formulation of the f_G parameter is useless here since the effective irradiance, G_{eff}, is by definition unknown in our case. To overcome such a limitation, f_G is split into its two main contributing factors:

$$f_G = f_{irr} \cdot f_{inc} \tag{7}$$

where f_{irr} represents the variation in the PV module efficiency with the level of the solar irradiance and f_{inc} the variation in the PV module efficiency as a function of the incidence angle of the solar irradiance, respectively. Then, approximating the ratio $G_t/G^{\star}$ by the Capacity Utilization Factor (CUF) defined as:

$$CUF = \frac{E_{AC}}{T \cdot P_N^{\star}} \tag{8}$$

with E_{AC} the energy produced during a period of time T (see Eq. (3)) and $P_N^{\star}$ the nominal (or peak) DC power of the PV system, f_{irr} can be estimated by:

$$f_{irr} = a + b \cdot CUF + c\ln(CUF) \tag{9}$$

In this equation, the three parameters a, b and c vary according to the considered PV module technology. Values representative of crystalline silicon cells technology (i.e., $a = 1$, $b = -0.01$ and $c = 0.025$) have been assumed for all PV modules here. Finally, based on [18], f_{inc} can be expressed as follows:

$$f_{inc} = 1 - \frac{1 - \exp(-(\cos\theta_i)/\alpha_r)}{1 - \exp(-1/\alpha_r)} \tag{10}$$

where θ_i is the irradiance angle of incidence and α_r the angular loss coefficient, an empirical dimensionless parameter dependent on the PV module technology and the dirtiness level of the PV module. Typical α_r values range from 0.16 to 0.17 for commercial clean crystalline and amorphous silicon modules. In this work a value of 0.20 has been assumed for α_r which is a typical value for crystalline silicon PV modules presenting a moderate level of dirtiness.

The second factor, f_T, is defined as:

$$f_T = 1 + \kappa\left(T_c - T_c^{\star}\right) \tag{11}$$

where the operating temperature of the solar cell, T_c, is calculated from the ambient temperature, T_a, using the following equation based on the Nominal Operation Cell Temperature (NOCT) defined

as the temperature reached by the cells when the PV module is exposed to a solar irradiance of 800 Wm^{-2}, an ambient temperature of 20 °C, and a wind speed of 1 ms^{-1} (it is obtained from the manufacturer datasheets):

$$T_c = T_a + \frac{NOCT - 20}{800} \cdot G_t \tag{12}$$

Similarly to Eq. (6) the CUF approximation is used to estimate T_c reformulating Eq. (12) as follows:

$$T_c = T_a + ((NOCT - 20)/800) \cdot 1000 \cdot CUF \tag{13}$$

The third factor, f_{AC}, is computed from:

$$f_{AC} = \frac{P_{AC}}{P_N^{\star} \cdot \eta_{EUR}} \tag{14}$$

where, the so-called "European efficiency" of the inverter, η_{EUR}, is given by the formula:

$$\eta_{EUR} = 0.03\eta_5 + 0.06\eta_{10} + 0.13\eta_{20} + 0.1\eta_{30} + 0.48\eta_{50} + 0.2\eta_{100} \tag{15}$$

with $\eta_5, \eta_{10}, \eta_{20}, \eta_{30}, \eta_{50}, \eta_{100}$ the instantaneous power efficiency values at 5%, 10%, 20%, 30%, 50% and 100% load.

The fourth factor, f_{PDC}, as well as the f_{BOS} factor cannot not be directly estimated because the real energetic behavior of each PV system is a priori unknown. Lumping both factors together into a new losses factor, f_{PERF}, it follows from Eqs. (4) and (5) that

$$f_{PERF} = \frac{1}{f_G \cdot f_T \cdot f_{AC}} \cdot \frac{E_{AC}/G_t}{P_N^{\star}/G^{\star}} \tag{16}$$

where the f_{PERF} factor sums up all the performance losses that, on the first hand could be avoided and, on the other hand that cannot be modeled through a simple and general analytical expression. This factor can be estimated for each PV system from historical data of E_{AC} and G_t using the E_{AC} data directly from the energy meters and G_t data obtained from the combination of clear-sky radiative model simulations and cloud cover information. It is worth pointing out that to reduce the uncertainties in its estimation, the f_{PERF} factor was determined on a monthly basis from clear sky situations.

Similarly to [19], clear-sky situations were determined from the PV systems energy production time series using a modified version of the algorithm developed by Ref. [20]. For each PV system, f_{PERF}, was calculated as being the ratio between the electrical energy produced by the PV system corrected by the three other losses factors (i.e. f_G, f_T, and f_{AC}) together with the quotient $P_N^{\star}/G^{\star}$ and the calculated in-plane clear-sky irradiation received by the PV system during the considered month. Clear sky simulations were carried out by running the Ineichen and Perez clear-sky model [21] using monthly mean climatological Linke turbidity values from PVGIS/CMSAF (http://re.jrc.ec.europa.eu/pvgis/). Simulated clear sky global horizontal irradiances were then transposed to tilted clear sky global irradiance using the ERB decomposition model ([22]; see Appendices A.1) and the HAY transposition model ([23]; see Appendices B.1).

Finally, once all losses factors are estimated, the derivation of the in-plane hourly global solar irradiance from the hourly PV system energy production is given by:

$$G_t = \frac{1}{f_G \cdot f_T \cdot f_{AC} \cdot f_{PERF}} \cdot \frac{E_{AC}}{P_N^{\star}/G^{\star}} \tag{17}$$

where all losses factors except f_{PERF} are evaluated on an hourly basis using air temperature measurements performed within the RMI's AWS spatially interpolated at each of the PV system locations in the computation of the operating solar cells temperature, T_c (see Eq. (13)). f_{PERF} are determined monthly from the previous month E_{AC} data.

3.2. Tilt to horizontal global solar irradiance transposition

The next step consists in the conversion of the retrieved in-plane global solar irradiance values from the PV systems energy outputs to global horizontal solar irradiance at each of the PV systems location. If many transposition models have been proposed in the literature (see Ref. [24] for a review) to convert solar irradiance on the horizontal plane, G_h, to that on a tilted plane, G_t, the inverse process (i.e. converting from tilted to horizontal) is only poorly discussed in literature (e.g. Refs. [5,6,25–29]). The difficulty relies on the fact that the procedure is analytically not invertible.

Transposition models have the general form:

$$G_t = B_t + D_t + D_g \tag{18}$$

where the tilted global solar irradiance, G_t, is expressed as the sum of the in-plane direct irradiance, B_t, in-plane diffuse irradiance, D_t, and the irradiance due to the ground reflection, D_g. The direct component, B_t, is obtained from:

$$B_t = B_n \cos\theta_i = B_h \frac{\cos\theta_i}{\cos\theta_z} = B_h r_b \tag{19}$$

with B_n the direct normal irradiance and B_h the direct irradiance on a horizontal surface, respectively. θ_i is the incidence angle and θ_z the solar zenith angle, respectively. Parameter $r_b = \cos\theta_i/\cos\theta_z$ is a factor that accounts for the direction of the beam radiation. The diffuse component, D_t, and the irradiance due to the ground reflection, D_g, can be modeled as follows:

$$D_t = D_h R_d \tag{20}$$

$$D_g = \rho G_h R_r \tag{21}$$

where D_h is the diffuse horizontal irradiance, G_h the global horizontal irradiance (i.e., $G_h = D_h + B_h$), R_d the diffuse transposition factor and ρ the ground albedo. The transposition factor for ground reflection, R_r, can be modeled under the isotropic assumption (e.g. Ref. [30]) as follows:

$$R_r^{iso} = \frac{1 - \cos\beta}{2} \tag{22}$$

where, β, is the tilt angle of the inclined surface. See Fig. 2 for angles definition.

Considering the effective global horizontal transmittance, K_t, the direct normal transmittance, K_n, and the diffuse horizontal transmittance, K_d (i.e., $K_d + K_n = K_t$):

$$\begin{aligned} G_h &= K_t I_o \cos\theta_z \\ B_n &= K_n I_o \\ D_h &= K_d I_o \cos\theta_z \end{aligned} \tag{23}$$

where I_o is the extraterrestrial normal incident irradiance, Eq. (18)

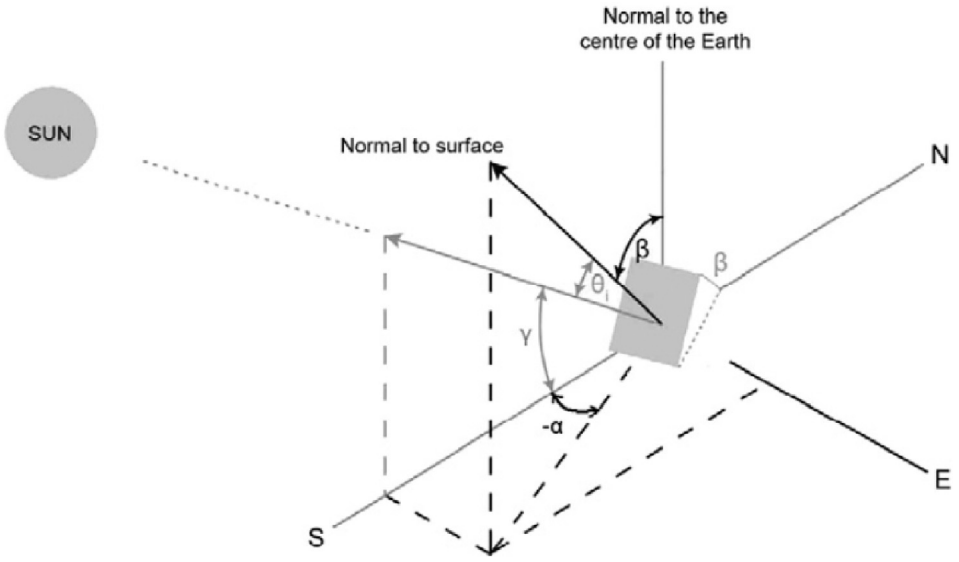

Fig. 2. Definition of angles used as coordinates for an element of sky radiation to an inclined plane of tilt β.

can be rewritten as:

$$G_t = K_t I_o \left[\cos\theta_i \left(1 - \frac{K_d}{K_t}\right) + \cos\theta_z \left(\frac{K_d}{K_t} R_d + \rho R_r\right)\right] \tag{24}$$

It comes from Eq. (24) that when only one tilted global solar irradiance measurement is considered, the conversion of G_t to G_h requires the use of a decomposition model (i.e. model that separate direct and diffuse solar components from the global one) to estimate K_d from K_t in addition to a transposition model to solve the inverse transposition problem. Eq. (24) is solved by an iteration procedure, varying the target quantity G_h (through K_t) until the resulting G_t' matches the input G_t (e.g. Refs. [5,26,28,29]). Note that an alternative method to Eq. (24) based on the Olmo model ([31]) that presents the particularity of being analytically invertible was proposed by Ref. [5]. But, if the overall performance of the inverted Olmo model was found comparable with the other approach, the results were slightly worse than those obtained by inverting the decomposition and transposition models in combination with an iterative solving process.

When two (or more) tilted irradiances values (with different tilt angles and/or orientations) are involved in the inverse transposition process, only a transposition model is required. The idea that simultaneous readings of a multi-pyranometers system can be used to disangle the various components of solar radiation on inclined surfaces was originally proposed by Ref. [25] to solve in remote locations the periodic adjustment required by normal incidence and shadow-band pyranometers to ensure that their readings remain accurate when long-term data acquisition is in progress.

Given n tilted pyranometers (with different inclinations and/or orientations), the inverse transposition problem can be represented in the matrix form (e.g. Ref. [27]):

$$\mathbf{x}^{\mathbf{T}}\, \mathbf{\Lambda}\, \mathbf{x} + \mathfrak{B} - \mathfrak{C} = 0 \tag{25}$$

where $\Lambda = \{\mathbf{A_i}\}$ is a $2 \times n \times 2$ third-order tensor, $\mathfrak{B} = \{B_i\}$ is a $n \times 2$ matrix, $\mathfrak{C}$ is a column vector with n given entries, and $\mathbf{x}$ a column vector with 2 variables:

$$\begin{aligned}
\mathbf{A_i} &= \begin{pmatrix} 0 & A_i \\ A_i & 0 \end{pmatrix} \in \Re^{2\times n\times 2} \\
\mathfrak{B} &= \begin{pmatrix} C_1 & B_1 \\ C_2 & B_2 \\ \vdots & \vdots \\ C_n & B_n \end{pmatrix} \in \Re^{n\times 2} \\
\mathfrak{C} &= \begin{pmatrix} G_{t1} \\ G_{t2} \\ \vdots \\ G_{tn} \end{pmatrix} \in \Re^{n} \\
\mathbf{x^T} &= \begin{pmatrix} D_h & B_h \end{pmatrix} \in \Re^{2}
\end{aligned} \tag{26}$$

where the coefficients A_i, B_i and C_i depend on the considered transposition model.

The least square (hereafter referred to as LS) solution to Eq. (26) is given by:

$$\min\left\{P(\mathbf{x}) = \frac{1}{2}\left|\left|\mathbf{x}^{\mathbf{T}}\, \mathbf{\Lambda}\, \mathbf{x} + \mathfrak{B} - \mathfrak{C}\right|\right|^2 : \mathbf{x} \in \Re^2\right\} \tag{27}$$

with $||.||$ referring to the Euclidean norm. However, the LS is hard to solve and a standard technique to resolve Eq. (27) is to use a Newton type iteration method (e.g. Ref. [32]). As an alternative, Eq. (26) can also be solved by minimizing the errors (this approach is hereafter denoted to as EM - Errors Minimization). In this case, the solution is to minimize

$$\min\left\{E(\mathbf{x}) = \sum_{i=1}^{n} \varepsilon_i^2(\mathbf{x}) : \mathbf{x} \in \Re^2\right\} \tag{28}$$

where, $\varepsilon_i(\mathbf{x}) = (A_i D_h B_h + B_i B_h + C_i D_h) - G_{t_i}$, with $i = 1, \ldots, n$ denoting the tilted pyranometer.

4. Results

The Mean Bias Error (MBE) and the Root Mean Square Error (RMSE) statistical error indexes have been used to evaluate the prediction of the global horizontal solar irradiance from the energy production of residential PV systems.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(e_i)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(e_i^2\right)}$$

where $e_i = (G_{i,e} - G_{i,o})$ is the residual value, $G_{i,e}$ are the estimated values and $G_{i,o}$ represent the observed measurements. A positive MBE (resp. a negative MBE) means that the model tends to overestimate (resp. underestimate) the observed measurements.

To obtain dimensionless statistical indicators we express MBE and RMSE as fractions of mean solar global irradiance during the respective time interval:

$$MBE[\%] = \frac{MBE}{\overline{M}}$$

$$RMSE[\%] = \frac{RMSE}{\overline{M}}$$

where $\overline{M} = \frac{1}{n}\sum_{i=1}^{n}(G_{i,o})$ is the measurements mean.

Statistical error indexes were computed between in situ hourly irradiance measurements and the estimations computed from the hourly energy productions of residential PV systems surrounding the measurement stations. An initial radius of 5 km centered on the station was considered to select the residential PV systems for the validation purpose. When less than 4 PV installations were found within the delimited area, the radius was extended to 10 km. Table 3 indicates for each of our measurement sites the number of neighboring PV installations used for validation. No PV system was found in the vicinity of the Sint-Katelijn-Waver, Retie and Mont-Rigi stations (i.e. AWS 6439, AWS 6464 and AWS 6494, respectively) and 3 others stations only have one surrounding residential PV system. At the opposite, the maximum number of installations surrounding a station is 37 for Ernage (i.e. AWS 6459).

Based on our former evaluation of the inverse transposition problem ([29]), two different approaches have been considered to compute the global horizontal solar irradiance from the PV systems energy production. In the first approach, the tilt to horizontal conversion is performed independently at each PV installations surrounding the validation site using Eq. (24) with the OLS decomposition model (see Appendices A.2) and the SKA transposition model (see Appendices B.2). The median value of the individual PV system estimates is considered in the validation against AWS observations. This approach is referred to as 1_PV-M hereafter. In the second approach all individual tilted global solar irradiance estimates are used simultaneously and the tilt to horizontal conversion is solved by EM (see Eq. (28)) using the Powell's quadratically convergent method ([33]) and the SKA transposition model (see Appendices B.2). Indeed the minimization carried out by using the Powell's method has been found to systematically outperform the LS solution ([29]). It is a generic minimization method that allows to minimize a quadratic function of several variables without calculating derivatives. The key advantage of not requiring explicit solution of derivatives is the very fast execution time of the Powell method. In order to avoid the problem of linear dependence in the Powell's algorithm, we adopted the modified Powell's method given in Ref. [34] and implemented in Ref. [35]. This second approach is hereafter denoted to as X_PV-EM.

Performance of the two approaches in the GHI estimation from PV systems AC power output was evaluated against in-situ observations and furthermore compared to the performance of GHI estimates retrieved from Meteosat Second Generation (MSG, [36]) satellite images as implemented on an operational basis at RMI. Description of the RMI's MAGIC/Heliosat-2 algorithm used to retrieve the solar surface irradiance at the SEVIRI imager spatial sampling distance above Belgium (e.g. about 6 km in the north–south direction and 3.3 km in the east–west direction) from MSG images is provided in Appendices C.

While the MSG based retrieval method always provides GHI estimates during day time, there are certain sun positions for which the PV systems power output method fails to produce a valid estimation. Unsuccessful tilt to horizontal conversions are systematically found at solar elevations lower or equal to 5° for both the 1_PV-M and the X_PV-EM approaches irrespective of the number of PV systems involved in the conversion process and their angular configurations (i.e. tilt and azimuth angles). Failure rates reported for the 1_PV-M and the X_PV-EM approaches at each validation sites are provided in Table 3 together with the total number of available hourly data points at each location for the year 2014. Unsurprisingly the largest failure rates (up to nearly 40% in the case of the Melle station -AWS 6434-) are found at validation sites where only one PV installation is available. With more PV systems, the number of unsuccessful conversions after sunrise and before sunset is decreased. Table 3 tends to indicate that 1_PV-M starts to produce valid results at lower solar elevation conditions than X_PV-EM (i.e. an overall failure rate of about 12.4% is reported for 1_PV-M and of 19.6% for X_PV-EM, respectively) but it is largely relying on the angular configurations of the PV installations found within the group of PV systems.

Table 3
Unsuccessful conversion (in %) reported for the PV systems 1_PV-M and X_PV-EM approaches, respectively. Also provided is the total number of hourly data points available at each validation sites and the number of PV installations found in the vicinity of the measurement stations. ⋆ indicates the PV installations located within a radius of 10 km surrounding the validation site.

AWS code	PV systems number	Hourly values	PV Method	
			1_PV-M	X_PV-EM
6407	1	4184	36.74%	/
6414	5	4529	12.89%	17.40%
6434	1	4477	39.74%	/
6438	1⋆	4666	22.29%	/
6439	0	/	/	/
6447	12⋆	4698	16.69%	18.09%
6455	4⋆	4346	13.71%	15.55%
6459	37⋆	4573	13.21%	21.21%
6464	0	/	/	/
6472	5	4583	17.72%	23.04%
6477	3⋆	4469	14.03%	22.04%
6484	10	/	/	/
6494	0	/	/	/

4.1. Hourly validation

Table 4 compares hourly GHI estimates derived from PV production data with the 1_PV-M and X_PV-EM approaches as well as from MSG images with the corresponding ground measurements. To ensure that the comparisons are made between comparable data, special attention was given to the coherence of the data, the precision of the time acquisition, and the synchronization of the different data sets with the ground measurements. Because of inaccuracies in the orientations and/or inclinations of the PV installations provided by the PV systems installers or owners, GHI computation from the energy production of only one installation can generate RMSE values as large as 189.34 Wm^{-2} or 57.8% (i.e. at the Middelkerke validation site, AWS 6407). Increasing the number of PV installations involved in the estimation process smoothes the

Table 4
Comparison between global horizontal solar irradiance produced by the PV systems power output method for both the 1_PV-M and X_PV-EM approaches and retrieved from the MSG satellite images with the corresponding ground measurements. The RMSE and MBE error statistics (in W m^{-2} and %) are calculated on an hourly basis over the full year 2014. ⋆ indicates the PV installations located within a radius of 10 km surrounding the validation site.

AWS code	PV system								MSG			
	Nbr	1_PV-M approach				X_PV-EM approach				RMSE		MBE	
		RMSE		MBE		RMSE		MBE					
6407	1	189.34	(57.76%)	−53.30	(-16.26%)	/	/	/	/	69.55	(21.22%)	0.54	(0.16%)
6414	5	116.83	(41.81%)	−24.06	(-8.61%)	113.81	(40.73%)	−9.00	(-3.22%)	56.68	(20.28%)	8.68	(3.11%)
6434	1	116.94	(39.11%)	−10.48	(-3.50%)	/	/	/	/	62.58	(20.93%)	11.61	(3.88%)
6438	1⋆	128.05	(46.26%)	−32.51	(-11.75%)	/	/	/	/	62.80	(22.69%)	13.86	(5.01%)
6447	12⋆	106.14	(39.23%)	1.85	(0.68%)	107.76	(39.83%)	−9.03	(-3.34%)	53.13	(19.64%)	7.67	(2.84%)
6455	4⋆	114.88	(43.17%)	20.86	(7.84%)	120.87	(45.42%)	14.63	(5.50%)	59.52	(22.37%)	15.63	(5.87%)
6459	37⋆	112.23	(39.66%)	−28.03	(-9.90%)	129.64	(45.81%)	−15.84	(-5.60%)	57.00	(20.14%)	1.01	(0.36%)
6472	5	120.55	(44.76%)	−9.08	(-3.37%)	129.87	(48.22%)	7.86	(2.92%)	52.95	(19.66%)	11.26	(4.18%)
6477	3⋆	110.13	(39.70%)	−17.95	(-6.47%)	129.15	(46.55%)	−3.56	(-1.28%)	55.05	(19.84%)	13.32	(4.80%)

GHI estimation to some extent. This is particularly apparent for 1_PV-M which globally presents lower RMSE values than found for X_PV-EM (i.e. an overall RMSE value of 113.5 Wm^{-2} or 41.4% is reported for 1_PV-M and of 121.9 Wm^{-2} or 44.4% for X_PV-EM, respectively). However, sensitivity experiments in which the number of PV installations involved in the GHI determination was varying revealed a larger variability in the resulting GHI estimations for 1_PV-M than found for X_PV-EM which produces a more stable solution.

[5] reported a somewhat similar mean RMSE error of about 40% for GHI estimates derived from 5 years (from 2010 through 2014) of 5-min resolution records of specific power of 45 PV systems in the region of Freiburg, Germany. In contrast, GHI retrieval from MSG images shows a better performance with an overall RMSE of 55.8 Wm^{-2} or 20.3%. Moreover, while the satellite retrieval tends to slightly overestimate the GHI values (i.e. MBE values ranging from 0.16 to 5.87%), there is no clear trend in the GHI computation from measured AC PV output power. Positive and negative biases are reported for both 1_PV-M and X_PV-EM approaches. Moreover it can appear that the sign of the bias even differs from one approach to the other (e.g. a negative MBE value of −9.08 Wm^{-2} or −3.37% is reported at the Humain validation site (AWS 6472) for 1_PV-M approach while an overestimation of 7.86 Wm^{-2} or 2.92% is found for X_PV-EM). In general, the magnitude of the bias is lower with X_PV-EM than with 1_PV-M (i.e. MBE values ranging from −5.6% to 5.5% and from −9.9% to 7.8%, respectively).

Table 5 compares the performance of the 1_PV-M and X_PV-EM approaches together with the MSG-based retrieval method in the hourly GHI computation as a function of the sky conditions. Sky-type classification is made upon the modified clearness index proposed by Ref. [37] and defined as

$$K'_t = \frac{K_t}{1.031\ exp\left(\frac{-1.4}{0.9+9.4/M}\right) + 0.1}$$

where $K_t = \frac{G_h}{E}$ is the usual clearness index (i.e. the horizontal surface solar irradiance, G_h, normalized by the corresponding extra-atmospheric irradiance, $E = I_o \cos\theta_z$). $M = (\sin\gamma + 0.15(\gamma + 3.885)^{-1.253})^{-1}$ is the optical air mass and γ is the solar elevation angle in degrees (i.e., $\gamma = 90 - \theta_z$). This zenith angle-independent expression of the clearness index, based on Kasten's ([38]) pyrheliometric formula, ensures that a given K'_t, unlike the classic index K_t, will represent meteorologically similar conditions regardless of the sun's position. Note that error statistics reported in Table 5 were calculated from validation sites accounting for at least three residential PV installations (i.e. the Middelkerke/AWS 6407, Melle/AWS 6434 and Stabroek/AWS 6438 measurement stations were excluded). Clearly, the relative accuracy of the GHI estimates varies noticeably as the sky conditions moves from overcast to clear sky situations irrespective of the calculation method. With a reported RMSE value of roughly 116%, the 1_PV-M and X_PV-EM approaches fail to produce reliable estimations in overcast conditions (i.e. $0.0 \leq K'_t < 0.2$) where the global radiation is mainly composed of diffuse radiation. With a reported RMSE in the order of 70% the satellite-based retrieval method also exhibits a poor performance in overcast conditions but shows a rapid performance improvement as the sky becomes clearer and presents a minimum RMSE value of 11.3% in partly clear conditions (i.e. $0.6 \leq K'_t < 0.8$). Similarly, the accuracy of GHI estimations from measured AC PV output power increases as the sky conditions becomes clear but the magnitude of the errors is still at least twice the one found for the satellite retrieval method irrespective of the

Table 5
Performance in term of RMSE of the PV systems power output method for both the 1_PV-M and X_PV-EM approaches and the MSG retrieval method in the global horizontal solar irradiance estimation as a function of the sky condition. Absolute (in Wm^{-2}) and relative (in %) RMSE are computed from 21716 hourly 2014 data points. Validation sites accounting for less than 3 residential PV installations were discarded for the errors indices computation (i.e. AWS 6407, AWS 6434 and AWS 6438).

Sky Condition	Method						Data Number	
	1_PV-M		X_PV-EM		MSG			
$0.0 \leq K'_t < 0.2$	73.83	(117.00%)	72.70	(115.21%)	46.26	(73.31%)	4164	(19.18%)
$0.2 \leq K'_t < 0.4$	106.96	(64.79%)	112.55	(68.17%)	54.82	(33.20%)	5494	(25.30%)
$0.4 \leq K'_t < 0.6$	117.59	(38.78%)	133.27	(43.95%)	63.48	(20.94%)	4981	(22.94%)
$0.6 \leq K'_t < 0.8$	133.60	(28.56%)	142.86	(30.54%)	53.04	(11.34%)	6214	(28.61%)
$0.8 \leq K'_t \leq 1.0$	128.89	(30.09%)	134.70	(31.45%)	72.45	(16.91%)	863	(3.97%)
All Sky	113.49	(41.37%)	121.87	(44.43%)	55.75	(20.32%)	21716	(100.0%)

computation approaches. As an example, RMSE values of roughly 30% are reported for the 1_PV-M and X_PV-EM approaches and of about 17% for the satellite-based method in the clear sky bin (i.e. $0.8 \leq K'_t \leq 1.0$) where the direct component largely domines. The poor performance of the PV systems power output method in overcast and cloudy conditions (i.e. $0.0 \leq K'_t < 0.4$) is mainly due to the reduced accuracy in the derivation of the in-plane hourly global solar irradiance from the hourly PV system energy production as computed from Eq. (17) when the global radiation is dominated by diffuse radiation (i.e. at low PV module's yield). This is supported by the recent evaluation of the tilt to horizontal global solar irradiance conversion by Ref. [29] indicating that the relative accuracy of the conversion does not vary noticeably as the sky condition moved from overcast to clear sky situations irrespective of the considered transposition model (e.g. a RMSE variation of about 3.5% was reported for the SKA transposition model we use in the current study).

Finally, both the PV systems power output method and the MSG based retrieval method present a positive bias when the global radiation is dominated by diffuse radiation (with MBE values of e.g. 43.1%, 37.3% and 45.4% reported in the overcast bin for the 1_PV-M approach, the X_PV-EM approach and the MSG based method, respectively) and a negative bias under clear-sky situation (with e.g. MBE values of respectively −18.6%, −13.% and −9.7% reported in the clear-sky bin). In overall the MSG based retrieval method present a positive bias in the order of 3.5% while the PV systems power output method shows a negative bias for both the 1_PV-M and X_PV-EM approaches (i.e. overall MBE of −3.4% and −1.%, respectively).

4.2. Daily validation

Computation of the statistical errors indexes on a daily basis is not as straightforward as for an hourly basis because as already mentioned both the 1_PV-M and X_PV-EM approaches fail to produce valid GHI estimates at low solar elevation conditions. In Table 6 RMSE and MBE indexes have been computed by assuming no incoming global horizontal solar irradiance in the computation of the daily global horizontal solar irradiation for data points where no valid hourly GHI estimates were obtained. Because unsuccessful GHI estimations from AC power output can be as large as 39.7% when only one PV installation is considered (see Table 3) the daily performance of the PV method can be very limited in some places in comparison to the satellite method. As an example, the RMSE of 1321.2 W h m^{-2} or 42.9% and the MBE of −984.8 W h m^{-2} or −31.9% reported at the Middelkerke validation site (AWS 6407) in Table 6 for the PV system method reduce to a RMSE of 296.4 W h m^{-2} or 9.6% and a MBE of 43 W h m^{-2} or 1.4% for the MSG-based retrieval method at this validation site.

Globally Table 6 indicates that the daily computation exhibits a better performance than hourly estimation irrespectively of the considered retrieval methods. An overall RMSE of 11% and a slight positive bias is reported for the satellite based method. 1_PV-M and X_PV-EM behave quite similarly in terms of RMSE (i.e. overall RMSE of 12%). Clearly the RMSE magnitude difference between the MSG-based and the PV systems power output methods is drastically reduced when considering daily global solar irradiation quantities rather than hourly GHI values. However the RMSE spatial variation (i.e. from one validation site to another) is larger in the PV systems based method and, while a systematic positive bias is reported for the satellite retrieval method, the sign of the bias can vary from one validation site to another in the PV systems based method and even between the 1_PV-M and X_PV-EM approaches on a given site.

4.3. Spatial validation

Finally, to assess the spatial distribution of the solar surface irradiation computed from hourly PV systems power outputs, the 1470 residential PV installations considered in our study (see Fig. 1) were spatially aggregated into 150 clusters (see black dots on panel C in Fig. 3 for the clusters spatial distribution) using the k-means algorithm ([39]). K-means clustering partitions a dataset into a small number of clusters by minimizing the distance between each data point and the center of the cluster it belongs to. A minimum of four PV installations by clusters was imposed except for two of them located in the vicinity of the Belgian coast because of the very low density of PV installations found in this area.

Fig. 3 presents an example of daily solar surface irradiation over the Belgian territory as computed from the interpolation of ground measurements using the ordinary kriging (OK) method (e.g. Ref. [40]), the MSG satellite derived estimation and the PV systems method using X_PV-EM. Clearly, due to the sparsity of the ground stations networks, interpolating ground data generates only a coarse distribution of the solar surface irradiation: large-scale variations of the solar irradiation (such as the south–east to north–west positive gradient) are identified but local fluctuations remain unseen (Fig. 3, panel a). Satellite-derived estimates, on the other hand, provide a global coverage and are therefore able to account for clouds induced small-scale variability in surface solar radiation (Fig. 3, panel b). Regarding the PV systems method, the daily solar irradiation estimated at the PV clusters level were then interpolated by OK method to cover the entire Belgian territory (Fig. 3, panel c). As we see, the south–east to north–west positive gradient is well apparent as well as some of the regional

Table 6
Comparison between daily cumulated surface solar irradiation produced by the PV systems power output method (for both the 1_PV-M and X_PV-EM approaches) and retrieved from the MSG satellite images with the corresponding daily ground measurements. The RMSE and MBE error statistics (in W h m^{-2} and %) are calculated on a daily basis over the full year 2014.★ indicates the PV installations located within a radius of 10 km surrounding the validation site.

AWS code	PV systems								MSG			
	Nbr	1_PV-M approach				X_PV-EM approach				RMSE		MBE	
		RMSE		MBE		RMSE		MBE					
6407	1	1321.23	(42.85%)	−984.80	(-31.94%)	/	/	/	/	296.38	(9.61%)	43.03	(1.40%)
6414	5	308.40	(14.25%)	−175.35	(-8.10%)	304.64	(14.08%)	−106.57	(-4.92%)	268.53	(12.41%)	116.29	(5.37%)
6434	1	534.81	(16.39%)	−257.06	(-7.88%)	/	/	/	/	337.75	(10.35%)	150.23	(4.60%)
6438	1★	640.95	(22.04%)	−434.83	(-14.95%)	/	/	/	/	322.61	(11.09%)	154.50	(5.31%)
6447	12★	250.26	(8.48%)	5.98	(0.20%)	242.89	(8.23%)	−115.75	(-3.92%)	232.35	(7.87%)	94.28	(3.19%)
6455	4★	348.20	(19.69%)	168.84	(9.55%)	306.30	(17.32%)	116.07	(6.56%)	260.73	(14.74%)	143.62	(8.12%)
6459	37★	225.21	(10.48%)	−137.49	(-6.40%)	180.90	(8.42%)	−64.94	(-3.02%)	231.97	(10.79%)	40.94	(1.90%)
6472	5	151.70	(9.48%)	−13.18	(-0.82%)	226.74	(14.18%)	43.76	(2.74%)	185.28	(11.58%)	112.68	(7.04%)
6477	3★	266.96	(10.56%)	−29.97	(-1.19%)	259.35	(10.26%)	100.60	(3.98%)	271.98	(10.76%)	161.13	(6.37%)

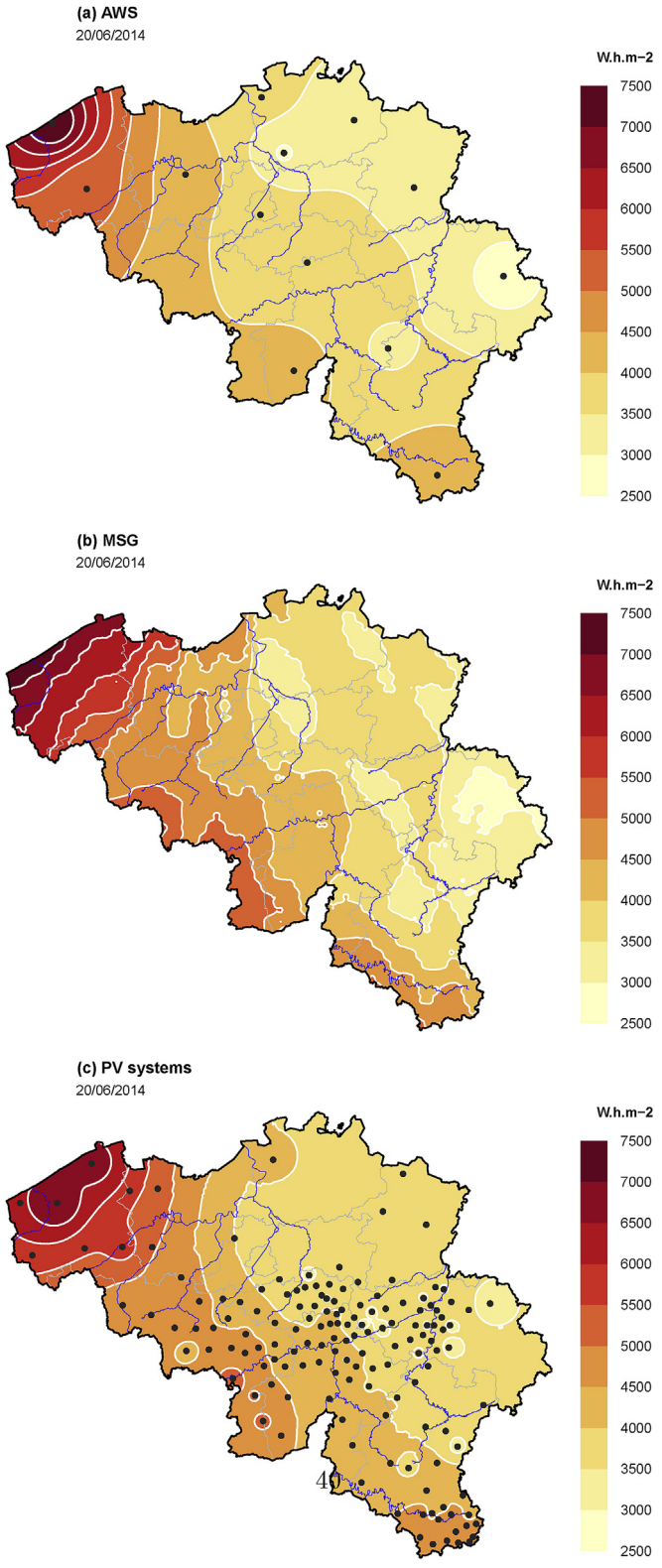

Fig. 3. Comparison between the daily spatial distribution of surface solar global irradiation (in $W\,h\,m^{-2}$) over Belgium as computed (a) by the interpolating ground measurements, (b) by the MSG satellite retrieval method and, (c) by the PV systems power output method. Illustrations are for the 20 June, 2014. Black dots in panel (a) indicate the location of the RMI's in-situ measurements sites. Black dots in panel (c) indicate the location of the 150 clusters of PV systems.

specificities. For instance, the Gaume region (area in the south-east of Belgium) located on the south side of the Ardenne (hilly mass) and that enjoys longer sunshine time appears clearly on the mapping. In general, the PV systems method provides small-scale patterns partly supported by the MSG derived mapping. Some others appear as the signature of an erroneous estimation at the cluster level.

5. Conclusions and perspective

The accuracy of solar irradiation data derived from the energy production of PV systems depends on a number of factors including the efficiency of the PV system, the weather conditions, the density of PV systems that can be used for the tilt to horizontal conversion, other data sources that can be accessed to complement the PV data. In particular, because the solar irradiation data obtained by reverse engineering of the PV energy output data degrade as the information about the orientation and tilt angles of the PV generator becomes more inaccurate, reliable angular characterization of the PV generator is of prime importance. Unfortunately, it was found that the provided angles can typically bear inaccuracies up to 5-10°. To overcome such a limitation, the development of a procedure able to assess the true azimuth and tilt angle of a PV generator from the sole knowledge of its hourly energy output data would be largely beneficial for the method. Towards this objective, it is worth mentioning that while the identification of a PV system's location and orientation from performance data has received limited attention in the past [41], have introduced a method for automatic detection of PV system configuration. And even more recently [42], have proposed a method to estimate the location and orientation of distributed photovoltaic systems from their generation output data.

Another limitation of the method is that there are certain sun positions for which the method fails to produce a valid estimation. As an example, unsuccessful tilt to horizontal conversions systematically occurs at low solar elevation (i.e., $\gamma \leq 5°$) irrespective of the number of PV systems involved in the conversion process. Increasing the number of PV installations involved in the computation process allows reducing the conversion failure rate and the smoothing the estimation to some extent. As an example, validation results computed on an hourly basis provide a mean RMSE value of about 40% when considering a group of neighboring PV installations in the estimation process while values as large as 60% are reported when using a single installation. By comparison, satellite-based global horizontal irradiance estimation exhibits a better performance with an associated overall RMSE of about 20%. By contrast the overall performance of both methods does not differ significantly on a daily basis. However, spatially the distribution of the solar surface irradiation computed from PV systems outputs presents small-scale pattern artifacts signature of erroneous estimations.

Beside a more accurate characterization of the orientation and tilt angles of the PV generator, the conversion of PV system energy production to tilted global solar irradiance could certainly be improved by taking into account the wind influence on the operating temperature of the solar cell as an example or by performing clear-sky radiative transfer computations with a more sophisticated model, etc. But because of the nature of the uncertainties that we have faced during our study, it could also be valuable to investigate on how recent data mining techniques (i.e. the extraction of hidden predictive information from large databases [43]) can be incorporated in the transformation of electrical variables into environmental variable such as solar irradiation. To this respect it is worth pointing out that artificial neural networks and machine learning techniques have already revealed to be powerful in the forecasting of solar irradiation data (e.g. Refs. [44–48]).

Acknowledgments

This study was supported by the Belgian Science Policy Office (BELSPO) through the Belgian Research Action through Interdisciplinary Networks (BRAIN-be) pioneer research project BR/314/PI/SPIDER Solar Irradiation from the energy production of residential PV systems.

Appendix A. Decomposition models

A.1. ERB model (Erbs et al., [22])

[22] developed a correlation between the hourly clearness index, K_t, and the corresponding diffuse fraction, K_d based on 5 stations data. In each station, hourly values of direct and global irradiances on a horizontal surface were registered. Diffuse irradiance was obtained as the difference of these quantities. The proposed correlation combines a linear regression for $0 < K_t \leq 0.22$, a fourth degree polynom for $0.22 < K_t \leq 0.8$ and a constant value for $K_t > 0.8$

$$K_d = \begin{cases} 1 - 0.09K_t & K_t \leq 0.22 \\ 0.9511 - 0.1604K_t + 4.388K_t^2 - 16.638K_t^3 + 12.336K_t^4 & 0.22 < K_t \leq 0.8 \\ 0.165 & K_t > 0.8 \end{cases} \quad (29)$$

A.2. OLS model (Skartveit and Olseth, [49])

Based on 15 years (i.e. 1965–1979) of hourly records of global and diffuse horizontal irradiances with mean solar elevation larger than 10° performed in Bergen (Norway: 60.4°N, 5.3°E) [49] proposed an analytical model expressing the hourly diffuse fraction of global irradiance in terms of hourly solar elevation and clearness index. In this model, the direct normal irradiance, B_n, is derived from the global horizontal irradiance, G_h, and the solar elevation angle, γ, using the following equation:

$$B_n = \frac{G_h(1 - \Psi)}{\sin\gamma} \quad (30)$$

where Ψ is a function of the clearness index, K_t. The model was validated with data collected in 12 stations worldwide. The function Ψ reads as:

$$\Psi = \begin{cases} 1 & \text{for } K_t < c_1 \\ 1 - (1 - d_1)\left[d_2 c_3^{1/2} + (1 - d_2)c_3^2\right] & \text{for } c_1 \leq K_t \leq 1.09c_2 \\ 1 - 1.09c_2\dfrac{1 - \Upsilon}{K_t} & \text{for } K_t > 1.09c_2 \end{cases} \quad (31)$$

where:

$$\begin{aligned} c_1 &= 0.2 \\ c_2 &= 0.87 - 0.56^{-0.06\gamma} \\ c_3 &= 0.5\left(1 + \sin\left(\pi\left(\tfrac{c_4}{d_3} - 0.5\right)\right)\right) \\ c_4 &= K_t - c_1 \\ d_1 &= 0.15 + 0.43\, e^{-0.06\,\gamma} \\ d_2 &= 0.27 \\ d_3 &= c_2 - c_1 \\ \Upsilon &= 1 - (1 - d_1)(d_2 cc_3^{1/2} + (1 - d_2)cc_3^2) \\ cc_3 &= 0.5\left(1 + \sin\left(\pi\left(\tfrac{cc_4}{d_3} - 0.5\right)\right)\right) \\ cc_4 &= 1.09c_2 - c_1 \end{aligned}$$

Appendix B. Transposition models

Each model develops the diffuse transposition factor (i.e., the ratio of diffuse radiation on a tilted surface to that of a horizontal), R_d, according to specific assumptions.

B.1. HAY model (Hay, [23])

In the HAY model ([23]), diffuse radiation from the sky is composed of an isotropic component and a circumsolar one. Horizon brightening is not taken into account. An anisotropy index, F_{Hay}, is used to quantify a portion of the diffuse radiation treated as circumsolar with the remaining portion of diffuse radiation assumed to be isotropic, i.e.,

$$R_d^{HAY} = F_{HAY} r_b + (1 - F_{HAY})\frac{1 + \cos\beta}{2} \quad (32)$$

where $F_{HAY} = B_h/(I_o\cos\theta_z)$ is the Hay's sky-clarity factor and $r_b = \cos\theta_i/\cos\theta_z$ the beam radiation conversion factor. θ_z is the solar zenith angle, θ_i is the incidence angle of the beam radiation on the tilted surface, B_h is the direct horizontal solar irradiance and I_o is the extraterrestrial normal incident irradiance.

B.2. SKA model (Skartveit and Olseth, [50])

Solar radiation measurements indicate that a significant part of sky diffuse radiation under overcast sky conditions comes from the sky region around the zenith. This effect vanishes when cloud cover disappears. [50] modified the HAY model (Eq. (32)) in order to account for this effect,

$$R_d^{SKA} = F_{HAY} r_b + Z\cos\beta + (1 - F_{HAY} - Z)\frac{1 + \cos\beta}{2} \quad (33)$$

where $Z = \max(0, 0.3 - 2F_{HAY})$ is the Skartveit-Olseth's correction factor. If $F_{HAY} \geq 0.15$, then $Z = 0$ and the model reduces to the HAY model.

Appendix C. Description of the RMIs MAGIC/Heliosat-2 algorithm

Information on GHI over Belgium is retrieved from MSG data by an adaptation of the well-known Heliosat-2 method ([51]) initially developed for Meteosat First generation satellites. The principle of the method is that a difference in global radiation perceived by the sensor aboard a satellite is only due to a change in the apparent albedo, which is itself due to an increase of the radiation emitted by the atmosphere towards the sensor (i.e. [52,53]). A key parameter is

the cloud index (also denoted as effective cloud albedo), n, determined by the magnitude of change between what is observed by the sensor and what should be observed under a very clear sky. To evaluate the all-sky GHI, a clear-sky model is coupled with the retrieved cloud index which acts as a proxy for cloud transmittance. Inputs to the Heliosat-2 method are not the visible satellite images in digital counts as in the original version of the method ([52]) but images of radiances/reflectances:

$$n^t(i,j) = \frac{\rho^t(i,j) - \rho^t_{cs}(i,j)}{\rho^t_{\max}(i,j) - \rho^t_{cs}(i,j)} \tag{34}$$

where $n^t(i,j)$ is the cloud index at time t for the satellite image pixel (i,j); $\rho^t(i,j)$ is the reflectance or apparent albedo observed by the sensor at time t; $\rho^t_{\max}(i,j)$ is the apparent albedo of the brightest cloud at time t; $\rho^t_{cs}(i,j)$ is the apparent ground albedo under clear-sky condition at time t.

The MAGIC/Heliosat-2 method implemented at RMI ([54]) is applied to the visible narrow-bands of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board of the MSG platform. $\rho^t_{cs}(i,j)$ is derived from the fifth percentile of the $\rho^t(i,j)$ distribution related to the last 60 days ([55]) while $\rho^t_{\max}(i,j)$ is estimated from theoretical radiative transfer model calculation (see Ref. [54]). GHI estimates are computed for each pixel and MSG time slot by the combination of the satellite cloud index, n, and the GHI in clear sky condition calculated by the Mesoscale Atmospheric Global Irradiance Code (MAGIC) clear-sky model ([56]). Daily global solar radiations are computed by trapezoidal integration over the diurnal cycle of the retrieved $GHI^t(i,j)$. The retrieval process runs over a spatial domain ranging from 48.0°N to 54.0°N and from 2.0°E to 7.5°E within the MSG field–of–view. In this domain, the SEVIRI spatial sampling distance degrades to about 6 km in the north-south direction and 3.3 km in the east-west direction.

References

[1] R. Perez, R. Seals, R. Stewart, A. Zelenka, V. Estrada-Cajigal, Using satellite-derived insolation data for the site/time specific simulation of solar energy systems, Sol. Energy 53 (1994) 491–495.
[2] R. Perez, R. Seals, A. Zelenka, Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data, Sol. Energy 60 (1997) 89–96.
[3] A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorentz, R. Mueller, et al., Solar energy assessment using remote sensing technologies, Rem. Sens. Environ. 86 (2003) 423–432.
[4] D. Renné, R. George, S. Wilcox, T. Stoffel, D. Myers, D. Heimiller, Solar Resource Assessment, Technical Report nrel/tp-581-42301, National Renewable Energy Laboratory, 2008, p. 1617. Cole Boulevard, Golden, Colorado:80401–3393.
[5] S. Killinger, F. Braam, B. Mueller, B. Wille-Haussmann, R. McKenna, Projection of power generation between differently-oriented pv systems, Sol. Energy 136 (2016) 153–165.
[6] B. Elsinga, W. van Sark, L. Ramakers, Inverse photovoltaic yield model for global horizontal irradiance reconstruction, Energy Sci. Eng. 5 (2017) 226–239.
[7] J. Leloux, L. Narvarte, D. Trebosc, Review of the performance of residential pv systems in Belgium, Renew. Sustain. Energy Rev. 16 (2012) 178–184.
[8] M. Journée, C. Bertrand, Quality control of solar radiation data within the rmib solar measurements network, Sol. Energy 85 (2011) 72–86.
[9] D. King, W. Boyson, J. Kratochvill, Photovoltaic Array Performance Model, Sandia National Laboratories (SANDIA REPORT SAND2004 3535), Albuquerque, New Mexico, 2004, p. 42.
[10] E. Lorenzo, Energy collected and delivered by PV modules, in: Antonio Luque, Steven Hegedus (Eds.), Handbook of Photovoltaic Science and Engineering, John Wiley and Sons, Ltd, 2011, pp. 984–1042.
[11] B. Decker, U. Jahn, Performance of 170 grid connected pv plants in northern Germany - analysis of yields and optimization potentials, Sol. Energy 59 (1997) 127–133.
[12] B. Marion, J. Adelstein, K. Boyle, H. Hayden, B. Hammond, T. Fletcher, et al., Performance parameters for grid-connected PV systems, in: Conference Record of the Thirty-first IEEE Photovoltaic Specialists Conference, 2005, pp. 1601–1606.
[13] F. Martínez-Moreno, E. Lorenzo, L. Narvarte, R. Moreton, N. Tyutyundzhiev, I. De La Parra, Technical specifications and quality control procedures for reducing the uncertainty in PV installations: results of the FP7 project PVCROPS, in: Proceedings of the 31st European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC), 2015, pp. 2217–2221.
[14] D. Evans, Simplified method for predicting photovoltaic array output, Sol. Energy 27 (1981) 555–560.
[15] C. Osterwald, Translation of device performance measurements to reference conditions, Sol. Cell. 18 (1986) 269–279.
[16] J. Randall, J. Jacot, Is am1.5 applicable in practice? modelling eight photovoltaic materials with respect to light intensity and two spectra, Renew. Energy 28 (2003) 1851–1864.
[17] S. Williams, T. Betts, T. Helf, R. Gottschalg, H. Beyer, D. Infield, Long-term module performance based on realistic reporting conditions with consideration to spectral effects, in: 3rd World Conference on Photovoltaic Energy Conversion: Osaka, Japan, 2003, pp. 1908–1911.
[18] N. Martìn, J. Ruiz, Annual angular reflection losses in pv modules, Prog. Photovoltaics Res. Appl. 13 (2005) 75–84.
[19] J. Taylor, J. Leloux, A. Everard, J. Briggs, A. Buckley, Monitoring thousands of distributed systems in the UK: energy production and performance, in: 11th Photovoltaic Science Application and Technology Conference: PVSAT-11, University of Leeds, Leeds, UK, 2015. Wednesday 15th-Friday 17th April 2015: Conference Proceedings C97.
[20] M. Reno, C. Hansen, J. Stein, Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis, Sandia National Laboratories (SANDIA REPORT SAN2012-2389), Albuquerque, New Mexico, 2012, p. 68.
[21] P. Ineichen, R. Perez, A new airmass independent formulation for the linke turbidity coefficient, Sol. Energy 73 (2002) 151–157.
[22] D. Erbs, S. Klein, J. Duffie, Estimation of the diffuse radiation fraction for hourly, daily and monthly-averaged global radiation, Sol. Energy 28 (1982) 293–302.
[23] J. Hay, Study of Shortwave Radiation on Non-horizontal Surfaces, Rep. No. 79-12, Atmospheric Environment Service, Downsview, Ontario, 1979.
[24] D. Yang, Solar radiation on inclined surfaces: corrections and benchmarks, Sol. Energy 136 (2016) 288–302.
[25] D. Faiman, A. Zemel, A. Zangvil, A method for monitoring insolation in remote regions, Sol. Energy 38 (1987) 327–333.
[26] D. Yang, Z. Dong, A. Nobre, Y. Khoo, P. Jirutitijaroen, W. Walsh, Evaluation of transposition and decomposition models for converting global solar irradiance form tilted surface to horizontal in tropical regions, Sol. Energy 97 (2013) 369–387.
[27] D. Yang, Z. Ye, A. Nobre, H. Du, W. Walsh, L. Lim, et al., Bidirectional irradiance transposition based on the perez model, Sol. Energy 110 (2014) 768–780.
[28] B. Marion, A model for deriving the direct normal and diffuse horizontal irradiance from the global tilted irradiance, Sol. Energy 122 (2015) 1037–1046.
[29] C. Housmans, A. Ipe, C. Bertrand, Tilt to horizontal global solar irradiance conversion: an evaluation at high tilt angles and different orientations, Renew. Energy 113 (2017) 1529–1538.
[30] C.A. Gueymard, Direct and indirect uncertainties in the prediction of tilted irradiance for solar engineering applications, Sol. Energy 83 (2009) 432–444.
[31] F. Olmo, J. Vida, I. Foyo, Y. Castro-Diez, L. Alados-Arboledas, Prediction of global irradiance on inclined surfaces from horizontal global irradiance, Energy 24 (1999) 689–704.
[32] C. Grosan, A. Abraham, A new approach for solving nonlinear equations systems, IEEE Trans. Syst. Man Cybern. Syst. Hum. 38 (2008) 698–714.
[33] M. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives, Comput. J. 7 (2) (1964) 155–162.
[34] F. Acton, Numerical Methods that Work, 1990, Corrected Edition, Mathematical Association of America, Washington, 1970, pp. 464–467.
[35] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes in C. The Art of Scientific Computing, second ed., Cambridge University Press, Cambrige, UK, 1992, p. 994 (Reprinted with corrections, 1997).
[36] J. Schmetz, P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, et al., An introduction to meteosat second generation (msg), Bull. Am. Meteorol. Soc. 83 (2002) 977–992.
[37] R. Perez, P. Ineichen, R. Seals, A. Zelenka, Making full use of the clearness index for parameterizing hourly insolation conditions, Sol. Energy 45 (1990) 111–114.
[38] F. Kasten, A simple parameterization of the pyrheliometric formula for determining the linke turbidity factor, Meteorol. Rundsch. 33 (1980) 124–127.
[39] J. Hair, W. Black, B. Babin, R. Anderson, Multivariate Data Analysis, seventh ed., Prentice Hall, 2009.
[40] H. Wackernagel, Multivariate Geostatistics: an Introduction with Applications, Springer-Verlag, Berlin, 1995.
[41] M. Williams, S. Kerrigan, A. Thornton, L. Energy, Automatic detection of PV system configuration, in: Presented at the World Renewable Energy Forum, 2012, p. 5.
[42] N. Haghdadi, J. Copper, A. Bruce, I. MacGill, A method to estimate the location and orientation of distributed photovoltaic systems from their generation output data, Renew. Energy 108 (2017) 390–400.
[43] J. Han, M. Kamber, Data Mining: Concepts and Techniques, second ed., The Morgan Kaufmann Publishers, 2006, ISBN 1-55860-901-6. March 2006.
[44] A. Sfetsos, A. Coonick, Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques, Sol. Energy 68 (2000)

169–178.
[45] M. Behrang, E. Assareh, A. Noghrehabadi, The potential of different artificial neural network (ann) techniques in daily global solar radiation modeling based on meteorological data, Sol. Energy 84 (2010) 1468–1480.
[46] A. Mellit, A. Massi Pavan, A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected pv plant at trieste, Italy, Sol. Energy 84 (2010) 807–821.
[47] M. Almeida, O. Perpiñán, L. Narvarte, Pv power forecast using a nonparametric pv model, Sol. Energy 115 (2015) 354–368.
[48] C. Voyant, G. Notton, S. Kalogirou, M.L. Nivet, C. Paoli, F. Motte, et al., Machine learning methods for solar radiation forecasting: a review, Renew. Energy 105 (2017) 569–582.
[49] A. Skartveit, J. Olseth, A model for the diffuse fraction of hourly global radiation, Sol. Energy 38 (1987) 271–274.
[50] A. Skartveit, J. Olseth, Modelling slope irradiance at high latitudes, Sol. Energy 36 (1986) 333–344.
[51] C. Rigollier, M. Lefèvre, L. Wald, The method heliosat-2 for deriving shortwave solar radiation from satellite images, Sol. Energy 77 (2004) 159–169.
[52] D. Cano, J. Monget, M. Albuisson, N. Guillar H Regas, L. Wald, A method for the determination of the global solar radiation from meteorological satellite data, Sol. Energy 56 (1986) 207–212.
[53] E. Raschke, A. Gratzki, M. Rieland, Estimates of global radiation at the ground from reduced data sets of international satellite cloud climatology project, J. Clim. 7 (1987) 205–213.
[54] C. Demain, M. Journée, C. Bertrand, Sensitivity of the rmi's magic/heliosat-2 method to relevant input data, Adv. Sci. Res. 10 (2013) 7–13.
[55] A. Ipe, N. Clerbaux, C. Bertrand, S. Dewitte, L. Gonzalez, Pixel-scale composite top-of-the-atmosphere clear-sky reflectances for meteosat-7 visible data, J. Geophys. Res. 108 (2003) 148–227.
[56] R. Müller, C. Matsoukas, A. Gratzki, H. Behr, R. Hollmann, The cm-saf operational scheme for the satellite based retrieval of solar surface irradiance-a lut based eigenvector hybrid approach, Rem. Sens. Environ. 113 (2009) 1012–1024.