HAL
open science

Solar irradiance estimations for modeling the variability of photovoltaic generation and assessing violations of grid constraints: A comparison between satellite and pyranometers measurements with load flow simulations

Fabrizio Sossan, Enrica Scolari, Rahul Gupta, Mario Paolone

To cite this version:

Fabrizio Sossan, Enrica Scolari, Rahul Gupta, Mario Paolone. Solar irradiance estimations for modeling the variability of photovoltaic generation and assessing violations of grid constraints: A comparison between satellite and pyranometers measurements with load flow simulations. Journal of Renewable and Sustainable Energy, 2019, 11 (5), pp.056103. 10.1063/1.5109076 . hal-02431344

HAL Id: hal-02431344

https://hal.science/hal-02431344v1

Submitted on 7 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Solar Irradiance Estimations for Modeling the Variability of Photovoltaic Generation and Assessing Violations of Grid Constraints: a Comparison between Satellite and Pyranometers Measurements with Load Flow Simulations

Fabrizio Sossan,[1, a)] Enrica Scolari,[2] Rahul Gupta,[2] and Mario Paolone[2]

[1)] *Centre Procédés, Energies Renouvelables et Systèmes Energétiques (PERSEE), Mines ParisTech, France*

[2)] *Distributed Electrical Systems Laboratory (DESL), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland*

(Dated: 7 January 2020)

Global horizontal irradiance (GHI) is typically used to model the potential of distributed photovoltaic (PV) generation. On the one hand, satellite estimations are non-pervasive and already available from commercial providers, but they have a limited spatiotemporal resolution. On the other hand, local estimations, e.g., from pyranometers, sky-cameras and monitored PV plants, capture local irradiance patterns and dynamics, but they require in-situ monitoring infrastructure and upgrading the asset of electrical operators. Considering that in most power systems, PV generation is typically the aggregated contribution of many distributed plants, are local GHI estimations necessary to characterize the variability of the power flow at the grid connection point (GCP) and detect violations of the limits of voltages and line currents accurately? To reply, we consider GHI measurements from a dense network of pyranometers (used to model the ground truth GHI potential), satellite estimations for the same area, and information about a medium and low voltage distribution system. We perform load flows at different levels of installed PV capacity and compare the nodal voltages, line currents, and the power at the GCP when the irradiance is from pyranometers and when from satellite estimations, deriving conclusions on the necessity, or not, of highly spatiotemporally resolved irradiance estimations.

I. INTRODUCTION

Current operational practices of electrical distribution utilities related to PV generation typically involve the use of solar irradiance estimations to evaluate its production potential. The kind of estimations usually depends on the level of aggregation. Distribution system operators (DSOs) typically use model-based estimates of PV generation for feasibility studies when connecting new PV power plants to verify that the grid can withstand the power injections without determining violations of statutory voltage limits and cables' ampacities (see e.g.[1]). Irradiance forecasts might be relevant for DSOs to schedule the operation of utility-scale storage facilities to mitigate the impact of PV generation on their grids, or if specific grid codes (see, e.g., the notion of balance group in the Swiss grid code[2]) to improve the forecast of their aggregated demand when they are penalized for incurring in unbalances.

At the level of the transmission system operators (TSOs) and load balance responsible, the need for forecasts of stochastic generation and demand stems from the requirements of dispatching conventional power plants and allocating power reserves in day-ahead and intra-day markets, typically at 1-hour or sub-hour resolution; due to the low time resolution of the dispatch processes, irradiance estimations from numerical weather predictions (NWPs) fed by satellite information may be used (e.g.,[3]).

However, future increased generation capacity levels from renewables and reduced values of system inertia will determine more significant reserve requirements and the compression of the time scale at which ancillary services need to be delivered. Future grid codes will promote the development of markets for enhanced regulation and congestion management (e.g.[4,5]), fostering the participation of distributed energy resources, including renewable generation, in the provision of grid ancillary services. Since PV generation is the fastest-growing among stochastic renewables in distribution grids, it is relevant to wonder whether satellite estimations will be able to capture PV production patterns accurately or if local estimations should complement them.

To reply to this question, we consider a low voltage (LV) and a medium voltage (MV) distribution feeder, and GHI estimations from both a dense network of pyranometers and satellites overlaid on a common geographical meshgrid. With load flow simulations, we determine nodal voltages, line currents, and power flows in the grids, first, when GHI is from pyranometers measurements (i.e., ground-truth case) and, second, with satellite estimations and compare the results. The comparison is performed from two standpoints: an analysis on the nodal voltages and lines currents to verify whether they respect the prescribed limits (i.e., a critical operational concern for distribution system operators), and evaluation of the volatility levels of the power flow at the grids' connection points (i.e., relevant in the context of scheduling power reserves).

The use of local and satellite estimations of GHI was

[a)] Electronic mail: fabrizio.sossan@mines-paristech.fr.; Formerly at DESL, EPFL.

considered in the existing literature in application to power system studies, however little effort was devoted to comparing their performance and to understand whether higher or lower spatiotemporal resolution plays a role. For instance, Ruf et al.[6] compare the total power production of a cluster of PV plants spread on a 0.6×0.7 km area against HelioClim-3 v.5 (HC3v5) estimates and ground observations from a single weather station at 11 km distance; they conclude that satellite estimations lead to a better performance than distant local observations. Leloux et al.[7] used satellite GHI estimates to characterize the performance of PV plants, with no focus however on grid voltage and current constraints and without considering local irradiance measurements. Conversely, Nguyen et al.[8] discusses applying local GHI measurements to determine the PV hosting capacity of MV grids and the position of on-load tap changer transformers for voltage regulation, with no performance comparison with satellite estimations. Rikos et al.[9] investigates the use of satellite estimations to characterize the variability of PV generation and anticipate shortcoming of generation, concluding that preemptive control based on early warnings from satellites are possible if below critical levels of installed PV capacity and ramping rates of PV generation, with no performance comparison against local GHI estimations. Lave et al.[10] propose a model to characterize the geographical smoothing of distributed irradiance time series based on measurements from 6 pyranometers within 3 km distance. They observed that the variance of the spatially averaged series is six times smaller than single series for time scales up to 256 sec; above that, the spatial smoothing effect decreases due to series becoming more correlated.

The main contribution of this paper is a quantitative comparison of the performance of satellite versus local estimations of GHI for detecting grid-constraint violations and estimating the variability at the GCP. As specific numerical findings depend on the topology and physical characteristics of the lines, it is generally not possible to derive analytic principles to determine apriori whether the use of readily available satellite estimations is reliable. However, the proposed results provide useful indications to determine whether local estimations should be considered.

The remainder of this paper is organized as follows. Section II describes the modelling principles and GHI data sources, Section III details the metric adopted for the comparison, sections IV and V presents and discusses the results, and Section VI outlines the conclusions.

II. SIMULATION MODELS AND DATA

A. A common geographical reference for measurements and networks

To the best of our knowledge, an experimental setup with geolocated information on both the electrical grid infrastructure and local irradiance is not publicly available at the time of this study. Thus, we collect information from different sources and overlay them on a common geographical mesh grid, as shown in Fig. 1. This allows to establish the link between GHI and grid topologies (as explained in Section II E), and to calculate the PV generation at each node. There are 3 layers of information: electrical networks, GHI pyranometer measurements, and GHI satellite estimations. The geographical area of interest extends for approximately 48 km^2. A mesh grid with 0.040 km resolution (this approximates the smallest distance in the model, i.e. the 0.035 km long segments of the LV grid) is overlaid on the area and used to compute a uniform map of approximated GHI with nearest-neighbor (NN) interpolation. The time resolution of the time series is 30 s, chosen as a trade-off between capturing fast GHI fluctuations, and tractable computational times.

The specifications of the benchmark electrical systems (described in Section II B) do not report the position of the nodes but only the lines length; these are used to establish the relative positions of the nodes with respect to each other; once they are determined, the location of the grids as a whole is chosen so that each falls in the perimeter delimited by the outermost pyranometers so to guarantee reliable GHI interpolations.

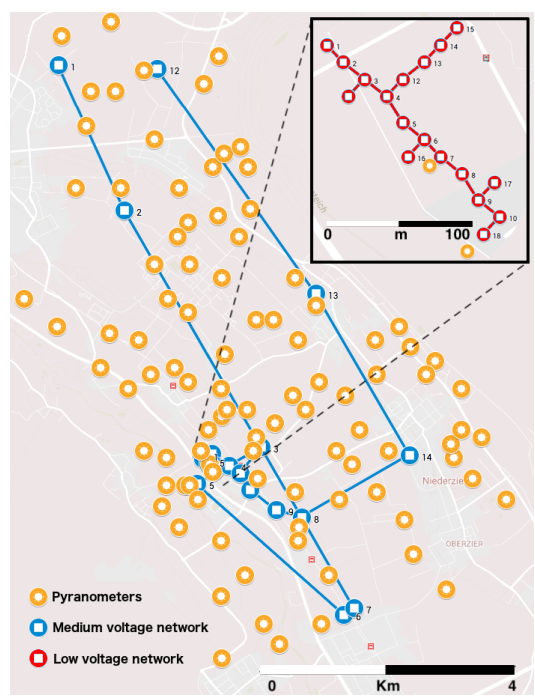

FIG. 1. Location of pyranometers and electrical networks. The zoomed region shows the LV grid. Information are from different sources and overlaid on a common geographical mesh grid.

B. Electrical networks

a. LV and MV systems To reflect the two mainstream trends for the connection to the grid of PV plants (i.e., utility-scale and rooftop installations), we consider a medium voltage (MV) and a low voltage (LV) distribution system, modeled according to the CIGRE specifica-

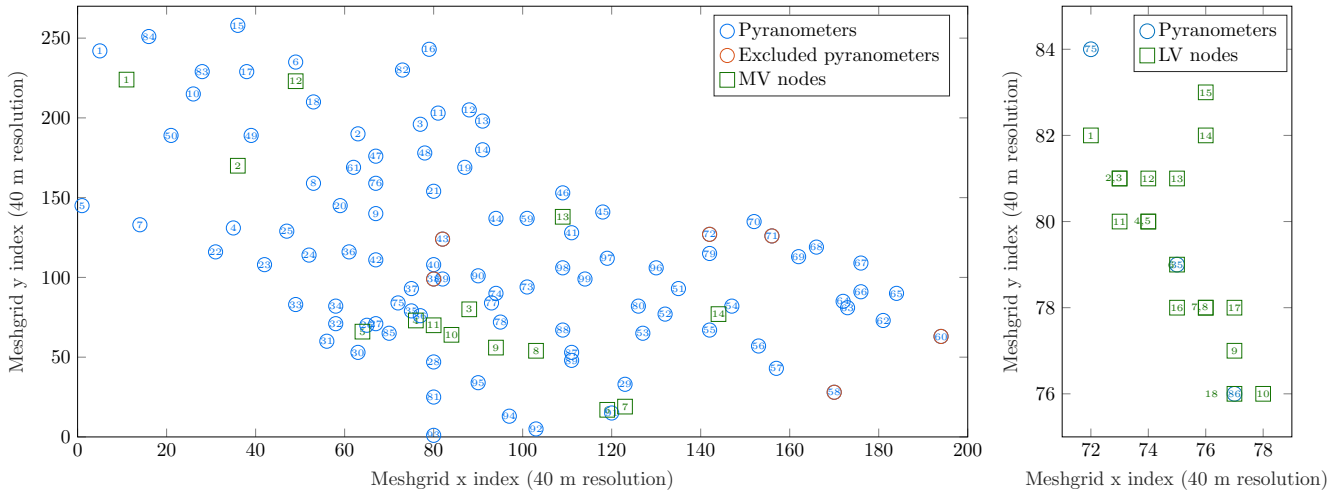

FIG. 2. Locations and labels of the pyranometers, MV nodes (in the left plot), and LV nodes (right). Excluded pyranometers refer to units that occasionally underestimated irradiance and were disregarded, as explained in the text. The distance between nodes {2, 3}, {7, 8}, and {4, 5} is less than the meshgrid resolution and share the same location on the meshgrid.

tions for the European grid benchmarks for three-phase MV and LV distribution systems[11] (figures 4 and 5). The MV grid is a 14-node 20 kV/25 MVA meshed system; the LV grid is an 18-node 0.4 kV/400 kVA radial system. The power flow at the GCP, nodal voltages and line currents of each grid are computed with single-phase load flows as a function of the grid admittance matrix and nodal injections. We thereby assume balanced three-phase systems: the inclusion of unbalances in the system is not of particular interest as most of the PV installations are interfaced to the grid with three-phase power converters. The nodal injection at each node is the algebraic sum of the electrical demand and PV generation, determined as discussed in the next two paragraphs.

We assume that excess PV generation flows in the upper-level grid. PV self-consumption schemes are not considered because not relevant in the context of comparing local versus satellite estimations.

b. Demand profiles In-line with the CIGRE specifications[11], the active power demand at the nodes is modeled by scaling the time series shown in Fig. 3 for the nominal demand of each node (reported in tables I and II) and according to the type of load (i.e., domestic or commercial). The same re-scaled series is applied to the demand at all nodes by considering perfect correlation among them. This choice is because our focus is comparing the effects of using satellite estimations against using local irradiance, and modeling spatial correlation would introduce a new element of variability in the analysis. Reactive power injections are modeled using a constant power factor, reported in Table I.

c. PV generation CIGRE specifications[11] include a moderate quantity of connected PV generation capacity. To reproduce a future scenario with large installed capacity levels of PV generation, we apply a procedure similar to work[12] to determine the maximum PV installed power than the grid can host without violating grid constraints (i.e., so-called PV hosting capacity of the grid). We consider fixed (non-tracking) PV plants. Nominal

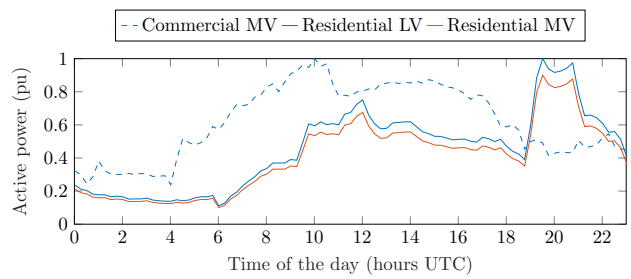

FIG. 3. Active power profiles of the demand in per unit (pu) of the nominal value of the demand.

capacity, and tilt and azimuth per node are specified in tables I and II . Plants' tilt and azimuth are, for the MV grid, the same for all the plants and optimized to achieve the highest yearly capacity factor, as typical for large-scale installations. For the LV grid, they are randomly distributed around the optimal configuration to account that in LV feeders PV generation is typically from non-uniform rooftop installations. The output of each PV power plant is determined as a function of its installed capacity, tilt and azimuth, and local GHI as detailed in Section II D. PV plants are assumed operated at unitary power factor.

In addition to the case where the grid is at the limit of its PV hosting capacity (that we denote as **Scenario 33**), we consider a more extreme second case where the PV installed capacity at each node is scaled by factor 4/3 to analyze the sensitivity of the results with respect to the amount of PV generation (**Scenario 43**).

C. Data sources for the global horizontal irradiance

We describe here the irradiance estimations and measurements processing. We introduce the following notation: $t = 1, 2, \ldots, T$ denotes the index of the T time inter-

TABLE I. MV grid: nominal loads, power factor (pf), PV installed capacity at STC, and ID and distance of the nearest pyranometer per node.

Node	Demand				PV Generation	Nearest	
	Residential		Commercial		Power	pyranometer	
	MW	pf	MW	pf	MWp	ID	dst (m)
1	15.3	0.98	5.1	0.95	5.71	10	700
2	0	–	0		5.02	49	769
3	0.28	0.97	0.26	0.85	4.39	77	256
4	0.44	0.97	0	–	2.23	86	126
5	0.75	0.97	0	–	1.28	26	165
6	0.56	0.97	0	–	0.09	86	268
7	0	–	0.09	0.85	0	78	544
8	0.6	0.97	0	–	0	78	641
9	0	–	0.67	0.85	1.13	89	400
10	0.49	0.97	0.08	0.85	1.62	91	200
11	0.34	0.97	0	–	5.71	91	89
12	15.3	0.98	5.28	0.95	5.71	55	408
13	0	–	0.04	0.85	0	59	322
14	0.22	0.97	0.39	0.85	5.02	18	544
Total	34.28		11.91		37.91		

TABLE II. LV grid: nominal load, power factor (pf), installed capacity at STC and configuration (i.e., azimuth and tilt) of PV generation, and ID and distance of the nearest pyranometer per node.

Node	Demand		PV Generation			Nearest	
	Residential		Power	Azimuth	Tilt	pyranometer	
	kW	pf	kWp	°	°	ID	dst (m)
1	200	0.95	227	167	34	75	80
11	15	0.95	114	161	43	75	165
15	52	0.95	107	185	38	75	165
16	55	0.95	112	182	42	86	113
17	35	0.95	26	190	41	86	80
Total	404		586				

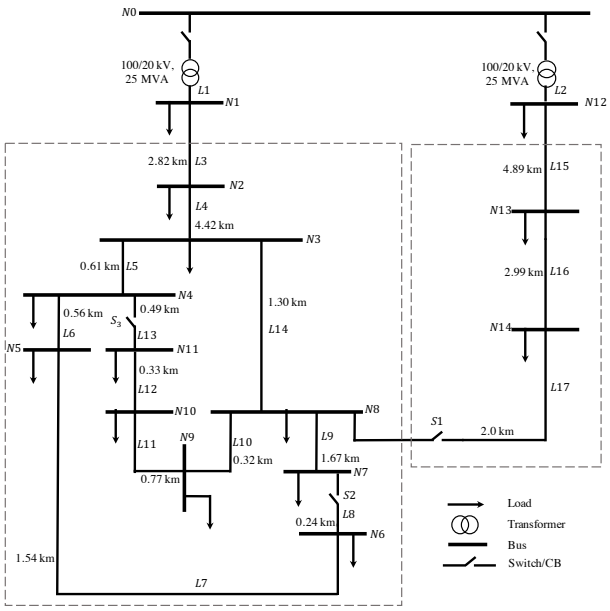

FIG. 4. The CIGRE benchmark grids for MV systems[11]; breaker S1 is operated closed.

vals, $d = 0, 1, \ldots$ the index of the day, and $p = 1, \ldots, 99$ the index of the pyranometers location; the time index t is split into daily sequences D_d for $d = 0, 1, \ldots$ that groups the indexes of all time intervals of day d. Pyranometer and satellite estimations for time t and location p are respectively denoted by x_{tp} and $\hat{x}_{tp}$.

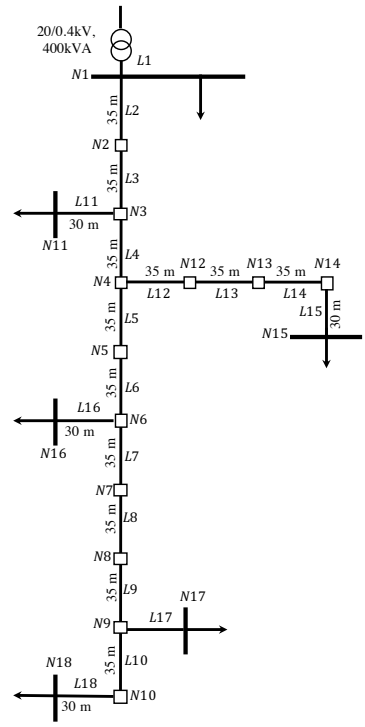

FIG. 5. The CIGRE benchmark grids for LV systems[11].

1. Pyranometers measurements and processing

a. Data selection Pyranometers data are from the HOPE measurements campaign[13] and refer to the period from 2 April to 24 July 2013. The setup included 99 open-field installations with silicon photodiode pyranometers distributed over an area of approximately 48 km^2. Measurements were labeled with quality flags according to the cleanliness of the sensor and horizontal alignment. We consider measurements marked as *good* and low incident angles with solar zenith angle smaller than 70° to avoid inaccurate readings. To filter out potential other bad data, we remove GHI values within a ±30 s range from a missing or non-valid data point. Additionally, we verify the presence of pyranometers with outlier measurements with respect to others and exclude them as explained next. Let $x_p^{(d)}$ be the reading of pyranometer p averaged over day d, formally:

$$x_p^{(d)} = \frac{1}{|D_d|} \sum_{t \in D_d} x_{tp}, \tag{1}$$

where $|D_d|$ denotes the number of elements in D_d, and $\bar{x}_d$ the average over space:

$$\bar{x}_d = \frac{1}{99} \sum_{p=1}^{99} x_p^{(d)}. \tag{2}$$

Daily average series $x_p^{(d)},\ d = 1, 2, \ldots$ are normalized by applying $\bar{x}_p^{(d)} = x_p^{(d)}/\bar{x}_d$, that essentially tell us how a measurements of a pyranometer compare with respect to average. The cumulative distribution function (CDF) of $\bar{x}_p^{(d)}, \forall p, d$ is shown in Fig. 6. It was observed that pyranometers number 38, 43, 58, 60, 71, 72 (shown in Fig. 2) occasionally underestimated the irradiance with respect to others; setting up an empirical symmetrical acceptance range 0.75-1.25 (shown with the dashed lines in Fig. 6 and that contains 99.5% of the measurements) was enough to exclude them. Variations around 1 in Fig. 6 can be explained by local weather patterns and non-uniform cloud passing over the area. To avoid periods with spatially sparse measurements, we consider time intervals where at least 70% of pyranometers are available.

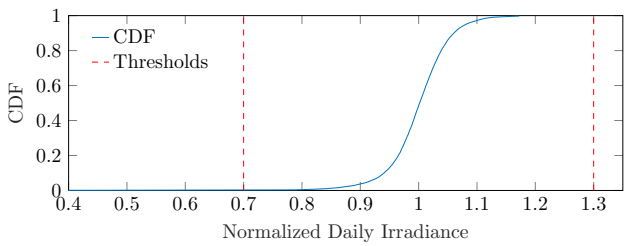

FIG. 6. CDF of normalized daily irradiance recorded by the pyranometers after filtering outliers.

b. Data processing GHI, originally at 1 Hz, is down-sampled to 30 s by decimation (samples average would smooth quick variations) to reduce the number of load flows to perform, and NN interpolation is applied to compute one data-point per pixel of the mesh grid. It is worth noting that, on the one hand, temporal down-sampling reduces temporal variations, thus biasing the analysis in favor of satellite estimations. On the other-hand, spatial NN interpolation exacerbates spatial correlation, thus biasing the results in favor of pyranometer estimations.

c. Accuracy of measurements Madhavan et al.[13] characterize the accuracy of pyranometers readings and report uncertainty levels of ±4.4 and 35.5 Wm^{-2} at 50 and 1000 Wm^{-2}, respectively, and zenith angle smaller than 70°, i.e., ±8% and 3.5%. We anticipate that, as it will be shown in Section IV C, the impact of measurements uncertainties on the performance assessment is negligible.

2. Satellite estimations

a. Data selection and processing We use commercial HC3v5 GHI estimations from Soda Service. They are obtained by estimating the clear-sky index using 15-min resolution satellite images from second generation Meteosat, applied to McClear's clear-sky model including information on water vapor, aerosols, and ozone levels. The commercial provider achieves a higher time resolution with an interpolation procedure[14] that considers the position of the sun in the sky, but not cloud movements. We query data at a 1-min resolution for each pyranometer location. We achieve the target 30-sec resolution with linear interpolation.

b. De-biasing Vernay et al.[15] correct the bias with a non-linear regression model fitted on a year of ground-based GHI measurements. This procedure is not applicable here because pyranometer data are for 4 months. Alternatively, we correct the bias by subtracting, for each pyranometer location, the difference between the satellite estimation for that location and the ground-based measurements, averaged on a daily basis, as explain next. The error of satellite estimation is:

$$e_{tp} = \hat{x}_{tp} - x_{tp} \tag{3}$$

The daily bias is the average of the error over one day:

$$e_p^{(d)} = \frac{1}{|D_d|} \sum_{t \in D_d} e_{tp}. \tag{4}$$

Satellite estimations are de-biased by applying:

$$\widehat{x}'_{pt} = \widehat{x}_{pt} - e_p^{(y(t))}, \tag{5}$$

where $y(t)$ denotes the index d of the day to which time interval t belong ($y(t) = d \mid t \in D_d$). Fig. 7 shows the empirical probability distribution function of the error of debiased satellite estimations: variations around zero reflect inaccuracies due to, e.g., reduced spatiotemporal resolution and parallax, as for instance reported in work[16]. Fig. 8 shows the standard deviation of the error as a function of the pyranometer location: the variability is within the range 125-145 Wm^{-2}. It indicates that the variations of the satellite estimation error is nearly uniform over space. This can be explained by the movements of clouds that determine similar shading patterns on the area over time and comparable levels of variability across locations.

The de-biasing process (3)-(5) might be inaccurate for late morning and evening hours, when irradiance values are lower than for the rest of the day. However, irradiance values with zenith angle larger than 70 degrees are removed. Moreover, in the context of the proposed analyses on variability and violations of grid constraints, low irradiance values are not expected to play a relevant role.

It is also worth noting that the process (3)-(5) is not viable in real-time applications because it is *non-causal* (it entails calculating the bias before knowing the realizations) and requires a dense network of measurements, that is generally unavailable. However, as the comparison focuses on quantifying the effect of spatial resolution and not of bias, we consider ideally de-biased satellite estimations.

c. Quality of satellite estimations Qu et al.[17] compare HC3v5 estimations against ground-based measurements from 23 sites and report relative bias and rRMSE in the ranges -4/4% and 14/38%. The same metrics (defined in work[17]) calculated for the current case study score in the ranges -7/3% and 28/41% at 15 min resolution, in-line with results reported in the literature.

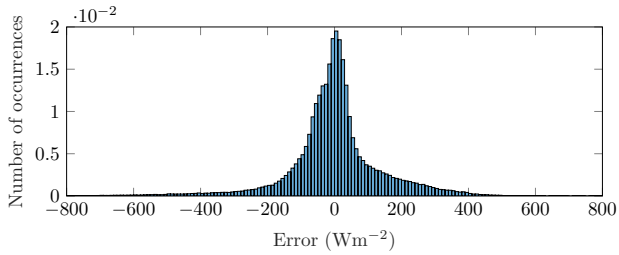

FIG. 7. Histogram of satellite estimation errors.

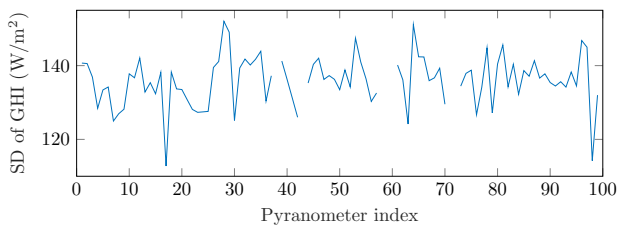

FIG. 8. Standard deviation (SD) of the satellite estimations error over time as a function of the location. Missing points refer to pyranometers that were excluded.

D. Computation of the nodal PV injections

We model PV generation with the empirical model[18–20] that accounts for the temperature effect by scaling the value of the POA irradiance. Let b denote the index of the electrical nodes, t the time interval, $I_t^{(\diagdown b)}$ in Wm^{-2} the POA irradiance. The PV generation and the cell temperature of the PV plant at node b and time t are modeled by (assuming constant unitary efficiency with respect to the irradiance[21]):

$$PV_t^b = \frac{I^{(\diagdown b)}}{1000}\left[1+\gamma\left(T_{\text{cell},t}^{(b)} - 25^\circ\right)\right] P_{\text{nom}}^{(b)} \tag{6}$$

$$T_{\text{cell},t}^{(b)} = T_{\text{air},t}^{(b)} + \beta \mathcal{I}_t^{(\diagdown b)}, \tag{7}$$

where $P_{\text{nom}}^{(b)}$ is the nominal capacity at the standard test condition (STC), $T_{\text{air},t}$ is the air temperature from measurements, and $\beta = 0.038$ and $\gamma = -0.0043$ are plant-specific parameters which were obtained by averaging the values for the close roof mount and open rack configurations[22] and values for polycrystalline modules[23], respectively.

The POA irradiance $I_t^{(\diagdown b)}$ is computed by applying a transposition model as a function of the GHI, plant tilt/azimuth and time of the day as described in II E 0 b. In turn, the GHI at the location b of the plant is computed as described next.

E. Computation of plane-of-array irradiance

This step allows us to establish the link between plant position and georeferenced measurements of the GHI and to determine the POA irradiance of each PV plant. For each plant, we determine first the geographical coordinates in the mesh grid that encompass its full extension as a function of the PV capacity. These coordinates are used to query GHI with the nearest neighbor interpolation scheme. Finally, GHI is transposed to get the POA irradiance. This whole process is detailed in the following.

a. Computation of GHI We first retrieve the coordinates of the pixels encompassing the full extension of the PV plant as a function of its nominal power capacity. For convenience, we assume square PV plants and centred on the pixel of the electrical node to which they are connected. The pixels which include the plants' extension are identified starting from the plant location by performing consecutive iteration of morphological dilatation until the pixels span an equal or larger area than the plant footprint area (see Fig. 9). Assuming horizontal panels, the footprint area A_b (m^2) of a plant with capacity $P_{\text{nom}}^{(b)}$ is approximated by:

$$A_b = P_{\text{nom}}^{(b)}/d, \tag{8}$$

where $d = 0.15$ is the PV conversion efficiency. The plant area in number of pixels approximated to the nearest larger integer is:

$$n_b = \text{ceil}\left(A_b/r^2\right), \tag{9}$$

where r is the resolution of the meshgrid in meters. Let (i_b, j_b) be the coordinates of bus b, k a non-negative integer denoting the algorithm iteration, $\mathcal{S}_k^{(b)}$ the set of coordinates at k, $\oplus$ the Minkowski sum, $\mathbf{1}_{3\times3}$ the all-one 3×3 matrix , and $|\mathcal{S}_k^{(b)}|$ the cardinality, or number of elements, of $\mathcal{S}_k^{(b)}$. The set of pixel coordinates that contains the PV plant at node b (i.e., gray pixels in Fig. 9) is found by:

$$\mathcal{S}_0^{(b)} = \{(i_b, j_b)\} \tag{10}$$

Repeat for k=1,2,...

$$\mathcal{S}_k^{(b)} = \mathcal{S}_{k-1}^{(b)} \oplus \mathbf{1}_{3\times3} \tag{11}$$

Until

$$|\mathcal{S}_k^{(b)}| \geq n_b. \tag{12}$$

Once the set is found, it is used to query the GHI values at each coordinate following the nearest neighbor interpolation scheme. GHI values are finally averaged to get equivalent value for the whole plant. Partial shading effect will be considered in future works where more dense measurements will be available.

b. Computation of POA irradiance For transposing irradiance starting from the GHI, we adopt the modeling tool-chain implemented in PVlib[24] to allow reproducibility of the results. For the computation of the direct normal irradiance from GHI, we use the DISC model[25]. For transposition as a function of the panels tilt and azimuth of each plant, we use the anisotropic Hay-Davies model[26], which is known to have better performance than isotropic models[27]. Results were also tested for the Perez model[28]; the change of the transposition model (that is

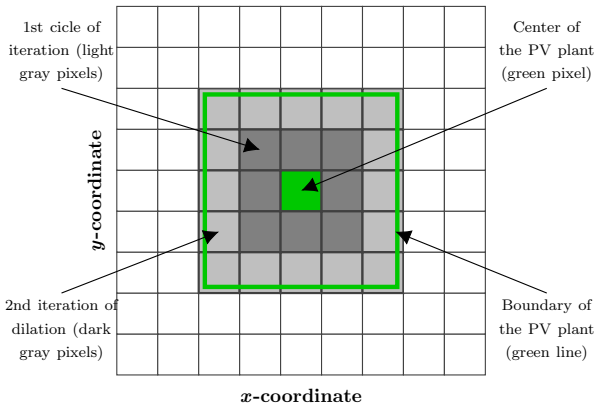

FIG. 9. A geographical mesh-grid with the PV plant location (green pixel), plant boundary (green line) and set of pixels encompassing the plant extension (gray and light gray pixels), obtained, in this case, with 2 iterations of morphological dilation.

known to be a critical component, see e.g. work[29]) was however found to impact results marginally, with slight changes on false negative detections and unaltered levels of variability-underestimation factors.

III. METRICS FOR THE COMPARISON

The objective of the comparative analysis is to evaluate whether the use of satellite estimations leads to underestimate reserve capacity requirements and if they can detect violations of grid constraints reliably.

For the former objective, we rely on the notion of variability of the power export, and we verify whether satellite estimations can record, or not, the same levels of variability as pyranometers. For the latter, we verify whether satellite estimations can correctly detect violations of grid constraints, and we analyze durations and inter-arrival times of false detections. The metrics are described in the following.

a. Variability-underestimation factor of satellite estimations Let $\boldsymbol{x}$ be a time series (for instance, irradiance) with elements $x_1, x_2, \ldots$ and $\bar{\boldsymbol{x}}$ its moving average on a 15-minute interval. We define the variability of $\boldsymbol{x}$ as the square root of the variance (Var) of the difference of the two series

$$\text{variability}(\boldsymbol{x}) = \sqrt{\text{Var}\,(\boldsymbol{x} - \bar{\boldsymbol{x}})}, \tag{13}$$

where $-$ is element-wise difference among vectors. In words, we say that (13) measures the intensity of the variations of x around a trend. Let $\boldsymbol{x}_{\text{pyr}}$ and $\boldsymbol{x}_{\text{sat}}$ refer to estimates from pyranometers and satellite, respectively. The variability-underestimation factor is the ratio between them:

$$\begin{aligned}\text{variability-underestimation factor}(\boldsymbol{x}) &= \\ &= \frac{\text{variability}(\boldsymbol{x}_{\text{pyr}})}{\text{variability}(\boldsymbol{x}_{\text{sat}})}.\end{aligned} \tag{14}$$

To further exemplify, we could assume that the variability is the metric that a TSO/load balance responsible uses to allocate power reserves. For instance, the higher the variability is, the larger the reserves should be. Assuming that the TSO uses satellite-derived information to allocate reserves capacity, the variability underestimation factor (14) can be interpreted as a scale coefficient that denotes the "inadequacy" of the allocated reserves with respect to real requirements. The variability-underestimation factor is calculated for the active power flow at the grid substation transformer (denoted by `Power at the GCP`), i.e., the quantity that a TSO would use to allocate reserve capacity. In addition, for the sake of following analysis, it is computed for the following time series:

- `Spatially averaged GHI`: GHI at the PV plants' locations averaged over all locations. It allows to compare the quality of satellite estimations compared to those of pyranometers;
- `Spatially averaged POA`: POA irradiance on PV plants' panels averaged over all locations. It allows to verify how the transposition model affects variability;
- `Total PV production`: the sum over all the plants of the PV power output. It allows to verify how local irradiance patterns are transformed in PV generation accounting for the modeled spatial distribution of PV panels;

b. False positive and negative detections of grid constraints violations A voltage violation occurs when the voltage magnitude at any node of the network is outside the range 0.9-1.1 pu. As we consider single-phase equivalent load flows and PV generation (i.e., voltage magnitudes monotonically increase with active power injections), violations occur in the form of trespassing the upper-bound 1.1 pu; we refer to this as over-voltage. A violation of a cable's ampacity limit occurs when its current magnitude exceeds 1 pu; we refer to this as overcurrent. For assessing the quality of satellite estimations in detecting over-currents and over-voltages, we introduce the notion of false negative and false positive. A false negative is when satellite estimations do not lead to a constraint violation and the pyranometers estimations do. A false positive is when satellite estimations lead to a constraint violation and the pyranometers estimations do not. False negatives are normalized by dividing for the total number of true positives, and false positives for true negatives. From a grid operator's perspective, false negatives are more of a concern than false positives because over-currents eventually lead to destroying cables.

c. Duration and inter-arrivals of false negative events It is of interest to characterize the *duration* of groups of false negatives and the *inter-arrival* between them to understand whether false detections tend to be sporadic or instead persistent over specific time intervals. The notion of duration and inter-arrival is graphically introduced in Fig. 10: the duration $d_1, d_2, \ldots$ is the series with the number of elements in each group of false negatives, whereas the inter-arrival $i_1, i_2, \ldots$ is the distance between each false negative and the next.

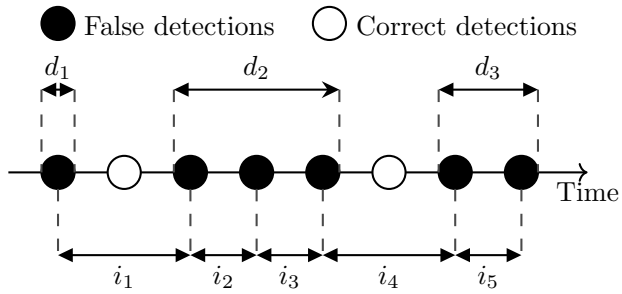

FIG. 10. Duration (d_1, d_2, d_3) and inter-arrival (i_1, i_2, i_3) of false detections.

IV. RESULTS ON A 3-DAY DATA SET

We analyze three days of measurements to identify and illustrate some mainstream results, that will be then further verified in Section V on a longer data set. Measurements were randomly selected among the available daily time series to reflect characteristic irradiance patterns because, as evidenced by the analysis, this was found to impact on results. In particular, they refer to a nearly clear-sky day (Day A, April 22), a partly cloudy day with frequent GHI variations (Day B, May 12), and a second nearly clear-sky day (Day C, July 2), with mild GHI variations in the afternoon. GHI series averaged over all the locations are shown in Fig. 11, where the thick lines and the envelopes show respectively averages and min/max values of the estimations over space.

A. Medium voltage voltage network

a. Variability-underestimation factor Table III reports the variability-underestimation factor for days A, B and C in scenario 33. It can be observed that:

1. largest underestimations happen in nearly clear-sky days A and C, denoting that satellite estimations cannot spot highly variable irradiance components caused by, e.g., small clouds, because of reduced spatiotemporal resolution and parallax effects;

2. underestimation levels of `spatially averaged POA` irradiance are different than of GHI. This is to be expected because irradiance transposition is nonlinear;

3. underestimation levels of `Total PV production` are different than of `Spatially averaged POA` due to temperature and non-uniformly distributed PV generation capacity that amplifies differently local irradiance patterns; to understand the role played by these two factors, `Total PV production (no T)` reports the underestimation factor when the temperature effect on PV generation is not modeled. In this case, the PV production is directly proportional to the local POA irradiance. It can be seen that modeling temperature leads to reducing the underestimation factor;

4. `power at the GCP` reflects the same trends as former quantities, however with different values than Total PV production because, *i*), the time-varying demand profile impacts on the variability and, *ii*), of path-dependant transmission losses. In Scenario 33, grid losses are approximately 5% of the total power flow in average.

Scenario 43 denoted increased PV power production levels. Results are not reported because they do not bring new relevant insights on variability.

TABLE III. MV grid: variability-underestimation factors. Metric Total PV production (**no T**) refers to PV production modeled without considering temperature effects.

Variability-underestimation factor of	Day A	Day B	Day C
Spatially averaged GHI	9.8	2.3	8.3
Spatially averaged POA	10.9	2.2	8.3
Total PV production	12.5	2.6	9.5
Power at the GCP	5.6	2.6	8.2
Total PV production (**no T**)	13.2	2.7	10.0

b. False-positive and false-negative detections Table IV summarizes the number of over-currents and over-voltages in the ground-truth load flows with pyranometers estimations, and of false detections with satellite estimations. Scenario 33 features no over-voltage and, in day B, a few over-currents, which are all undetected by satellite estimations, as denoted by 100% over-current false negatives; false detections in Day B are because satellite estimations do not record the large irradiance peaks that lead to over-currents. In Scenario 33, satellite estimations do not score any false-positive detection.

Scenario 43 features higher PV installed capacity, and few over-voltages and more over-currents than Scenario 33. Days B and C are still the most critical for false negatives due to highly volatile irradiance patterns. Scenario 43 reports less over-current false negatives than Scenario 33 because the larger PV capacity determines higher average line currents; thus smaller irradiance peaks are enough to trigger a valid detection of a grid constraint violation.

Scenario 43 reports some false positives, denoting that satellite estimations overestimates the PV production potential in some instances. This is the case, e.g., in the central part of Day A. as visible in Fig. 11, due to residual positive bias.

TABLE IV. MV grid: occurrences (%) of over-voltages and over-currents with pyranometers, and false positive (FPs) and false negative (FNs) detections of satellite estimations.

PV penetration scenario	Scenario 33			Scenario 43		
Day / Metric	A	B	C	A	B	C
Over-voltages (PYR)	0	0	0	0	0.32	0
Over-currents (PYR)	0	0.03	0	6	1.57	3.54
Over-voltage FNs (SAT)	–	–	–	–	100	–
Over-current FNs (SAT)	–	100	–	16.6	72.4	58.4
Over-voltage FPs (SAT)	0	0	0	0	0	0
Over-current FPs (SAT)	0	0	0	0.4	0.2	1.7

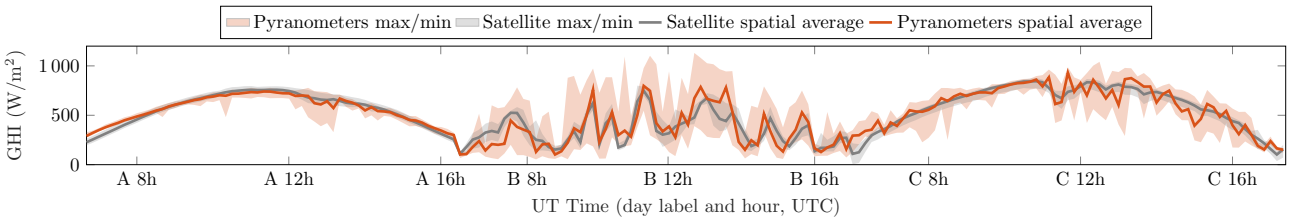

FIG. 11. Satellite (SAT) and pyranometer (PYR) estimations of GHI averaged over all locations for 3 days: A (22 Apr 2013), B (12 May 2013), and C (2 Jul 2013). Values with solar zenith angle larger than 70° are omitted. Solid lines denote spatial averages and color-filled areas denote the envelope between maximum and minimum values over space.

c. Irradiance and over-current patterns during false negatives We perform this analysis based on Scenario 43 and partly-cloudy Day B because it reports the largest number of over-current false negatives.

The upper-panel plot in Fig. 12 shows the CDFs of the GHI estimation error (3) of satellite estimations during correct, false negative and false positive detections. It shows that the GHI error for false negatives and false positives is respectively smaller and larger than for true detections. This in-line with the previous insight, according to which irradiance underestimations lead to false negatives and overestimations to false positives.

The bottom-panel plot in Fig. 12 shows the CDFs of the estimation errors of line currents for satellite estimations (i.e., the difference between line currents estimated from satellite and pyranometers). Current errors are in per unit of the ampacity limit of the respective cable, e.g. 0.5 pu corresponds to half of the cable rating. The plot denotes that negative errors (current underestimations) are conducive to false negatives, and positive errors to false positives, as to be expected. It also denotes that the error magnitude is generally larger for false negatives than for false positives. Nearly 20% of false-negative episodes have current errors larger than 0.6 pu.

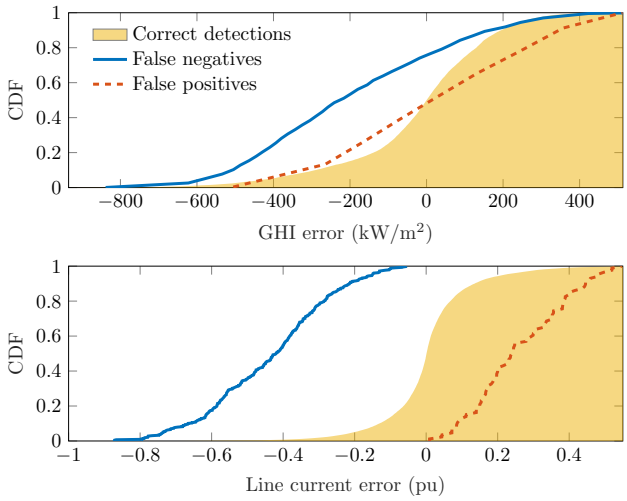

FIG. 12. MV grid: CDFs of GHI errors of satellite estimations (upper panel) and of line current errors of satellite estimations (lower panel) in Day B and scenario 43.

d. Duration and inter-arrival times of false negatives Also this analysis is performed on Scenario 43 and partly-cloudy Day B because it reports the largest number of over-current false negatives.

Fig. 13 shows the histograms of the duration (in the first row) and of the inter-arrival times (second row) of over-current false negatives. We first discuss the duration by referring to the histograms of the first row. The duration of events with persistent false negatives is generally less than 10 samples (i.e., 5 minutes), except for a single event in nearly clear-sky day C, which is 20-sample long. Duration patterns are similar in the 3 days; thus, they do not seem to exhibit specific variations according to the sky conditions.

The histograms in the second row of Fig. 13 show inter-arrival times mostly smaller than 10 samples, denoting that false positives tend to be in the proximity of each other. Day C features longer inter-arrivals, up to 120 samples, denoting that false negatives occur in 2 groups that are 1 hour apart.

B. Low voltage network

The results for the LV network were found similar to the MV case, except for some differences ascribed to the reduced spatial scale. In this section, we briefly summarize the findings and focus on details only in the case of major differences.

a. Variability-underestimation factor Table V reports the variability-underestimation factor for days A, B and C in Scenario 33. Compared to the MV case, the LV grid scores larger variability-underestimation due to its smaller extension and reduced spatial smoothing. Largest underestimation factors happen in the nearly clear-sky day. Average transmission losses (not reported in the table) are approx. 8% of the total active power flow at the GCP.

TABLE V. LV grid: variability-underestimation factors. Metric Total PV production (**no T**) refers to PV production modeled without considering temperature effects.

Variability-underestimation factor of	Day A	Day B	Day C
Spatially averaged GHI	12.6	3.8	13.5
Spatially averaged POA	13.5	3.9	13.4
Total PV production	9.0	3.9	14.3
Power at the GCP	7.4	3.9	13.7
Total PV production (**no T**)	9.0	4.1	14.8

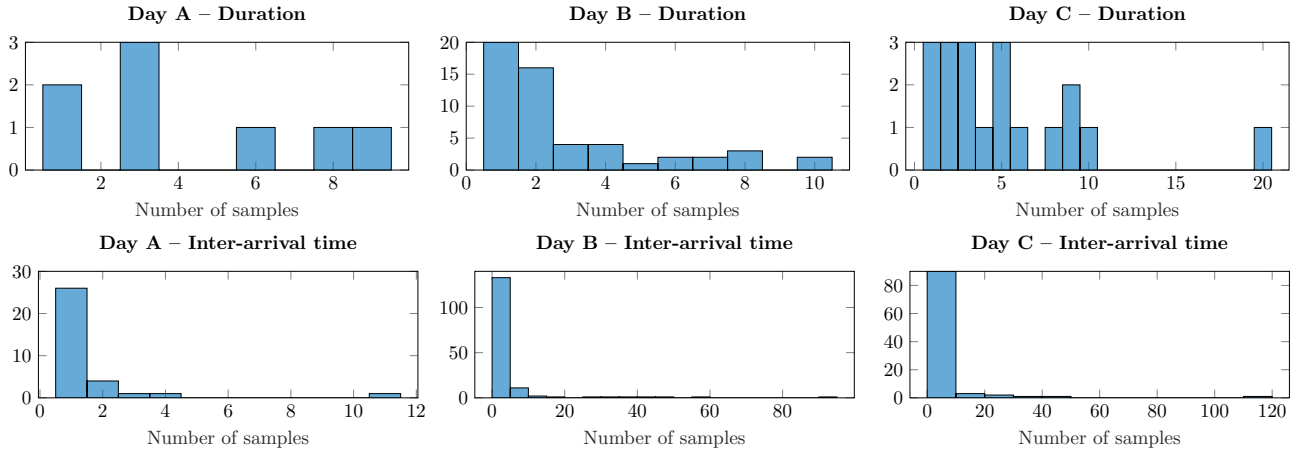

FIG. 13. Duration (first row) and inter-arrival (second row) of false negative events for the MV grid case.

b. False positive and false negatives Table VI reports the number of over-currents and over-voltages in the load flows with pyranometers, and of false detections with satellite estimations. Similarly to the MV case, days B and C are the most critical for false-negative detections. False positives occur only in the scenario with the largest PV installed capacity. The CDFs of the satellite estimations error and line currents error confirm identical findings as before and they are omitted.

c. Duration and inter-arrival times Statistics of duration and inter-arrival of over-current false negatives were found comparable with the previous case, denoting that differences in the geographical dimension at this scale does not seem to play a role. Similar results are also confirmed by the analysis of the longer data set in the next section.

TABLE VI. LV grid: occurrences (%) of over-voltages and over-currents with pyranometers, and false positive (FPs) and false negative (FNs) detections of satellite estimations.

PV penetration scenario	Scenario 33			Scenario 43		
Day / Metric	A	B	C	A	B	C
Over-voltages (PYR)	0	0	0	0	0.24	0.03
Over-currents (PYR)	0	0.01	0	1.59	1.44	1.05
Over-voltage FNs (SAT)	–	–	–	–	100	100
Over-current FNs (SAT)	–	100	–	10.1	89.9	53.1
Over-voltage FPs (SAT)	0	0	0	0	0	0
Over-current FPs (SAT)	0	0	0	0	0	0.2

C. Impact of pyranometers' inaccuracies on results

To assess if the accuracy of pyranometers' measurements could impact on results, we have repeated the simulations of the MV grid by altering all irradiance value by ±5%(i.e., given by averaging the accuracy figures reported in Section II C). It was verified that results were minimally impacted, and conclusions unchanged.

V. RESULTS ON 25 DAYS OF DATA

We consider 25 days of measurements and verify if considerations developed in the former section apply. Data are shown in Fig. 14, along with their classification according to prevailing sky conditions (clear-sky, nearly clear-sky, partly cloudy, cloudy), performed by visual inspection.

a. Variability Fig. 15 shows the variability-underestimation factor of `Total PV production` against the variability of `spatially averaged GHI` of pyranometers (defined in Section III). They denote that, *i)*, variability is generally higher in partly cloudy and nearly clear-sky days than in cloudy and clear-sky days, and *ii)*, largest to lowest underestimations are in nearly clear-sky, partly cloudy, clear-sky and cloudy days. Results for the MV and LV grids are similar, however the former features smaller axis values than the latter due to its larger extension and spatial smoothing. In this regard, it is interesting to observe that despite the MV grid uses information from 11 pyranometers on a surface of 15 km^2 (whereas the LV grid uses 2, which are less than 400 meters far apart, and extends for 0.035 km^2), the variability of its `spatially averaged GHI` is only 30% smaller than for the LV grid, denoting an overall small spatial smoothing effect.

The variability-underestimation factor was defined in (14) as the ratio of the variability recorded by pyranometers; hence, it expresses the proportion among the two. To complement that information, we define the variability difference

$$\text{variability difference}(x) = \tag{15}$$
$$\text{variability}(\boldsymbol{x}_{\text{pyr}}) - \text{variability}(\boldsymbol{x}_{\text{sat}}), \tag{16}$$

which can be interpreted as the magnitude of the uncaptured variations. Fig.16 shows the variability difference of `Total PV production` against the variability of `spatially averaged GHI` of pyranometers for the MV network. They stand in a nearly linear relationship, denoting that, the higher the variability of the pyranometers is, the more satellite estimations underestimates the

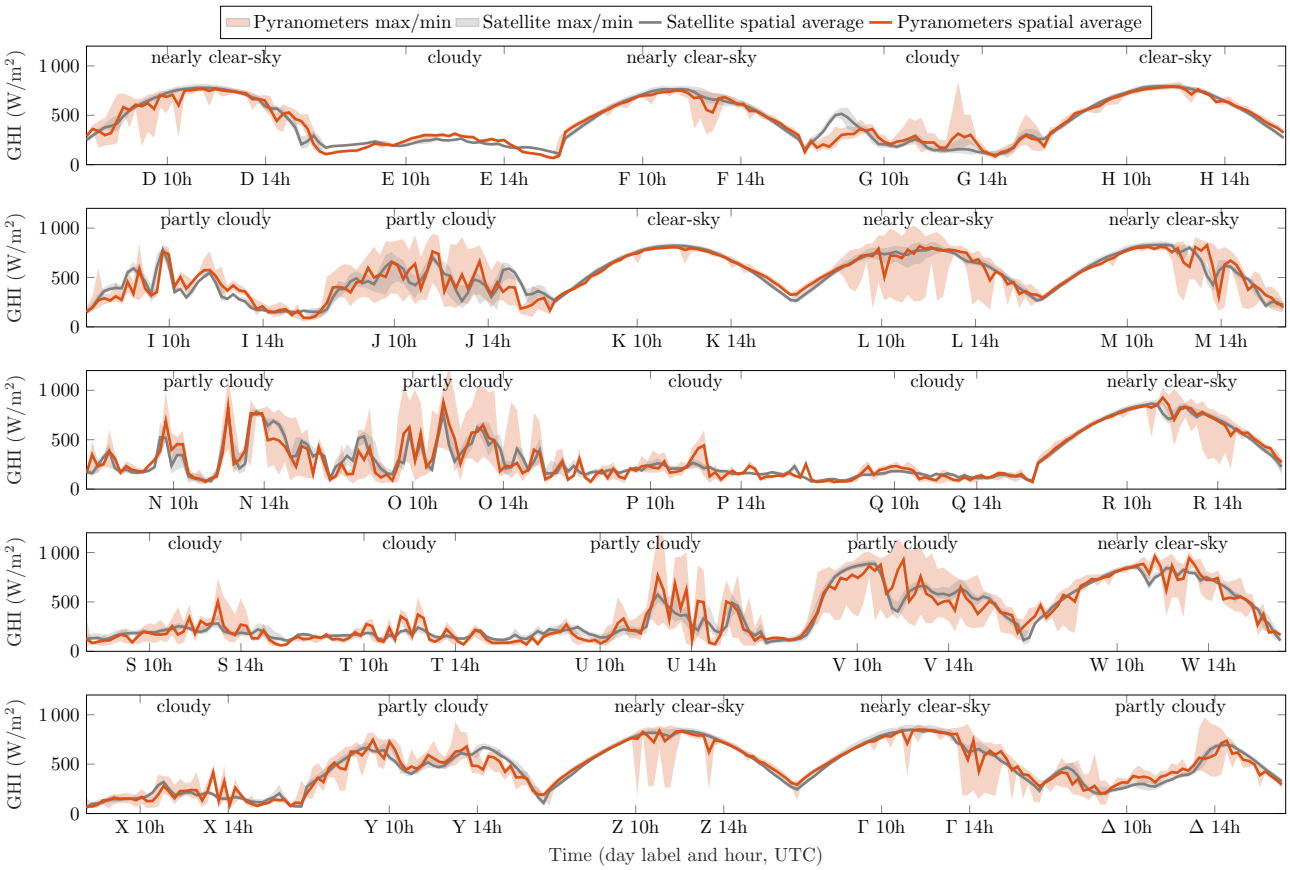

FIG. 14. Fifteen non-consecutive daily scenarios of GHI: D, E, F, G, H (20, 23, 24, 25, and 26 April 2013), I, J, K, L, M, N, P (2, 4, 5, 6, 12, 15, and 16 May 2013), Q, and R (4 June and 3 July). Values with solar zenith angle larger than 70° are omitted. Solid lines denote spatial averages and color-filled areas denote the envelope between maximum and minimum values over space.

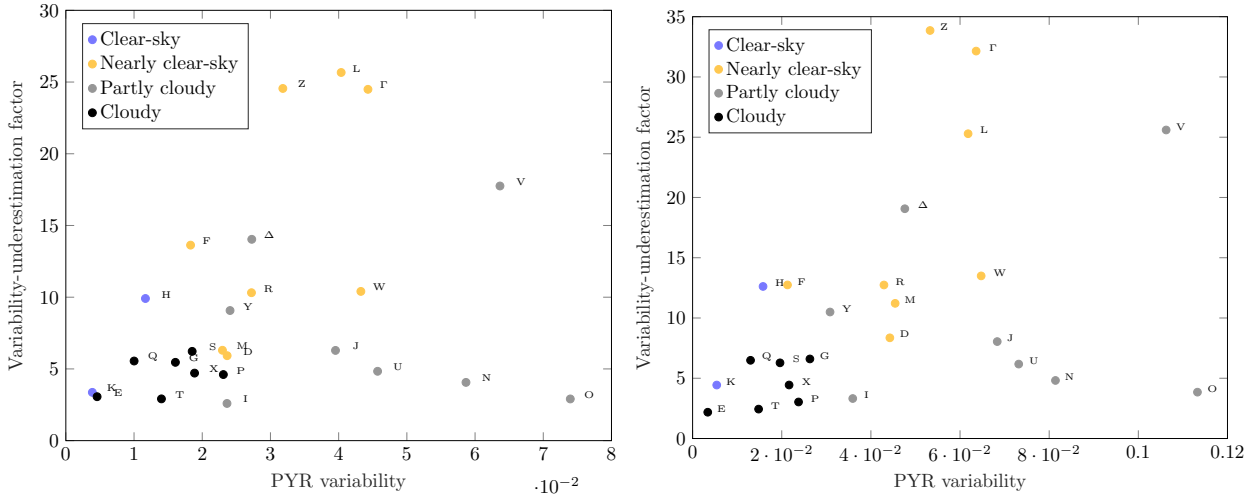

FIG. 15. Variability underestimation factor of `total PV generation` of satellite estimations vs variability of `Spatially averaged GHI` of pyranometers for the MV (left) and LV grid (right). Alphabetic labels refer to the 25 days of data.

variability of the aggregated PV production. The results for the LV network are similar and are omitted.

b. False detections Fig. 17 compares the shares of false positives versus false negatives for over-currents in the MV and LV grids. In the MV grid (left panel), false negatives are predominant in partly cloudy days (when irradiance variations are generally large), whereas false positives are predominant in clear-sky and nearly clearly-sky days due to the residual positive bias of satellite estimations. Cloudy Day U also scores 100% false nega-

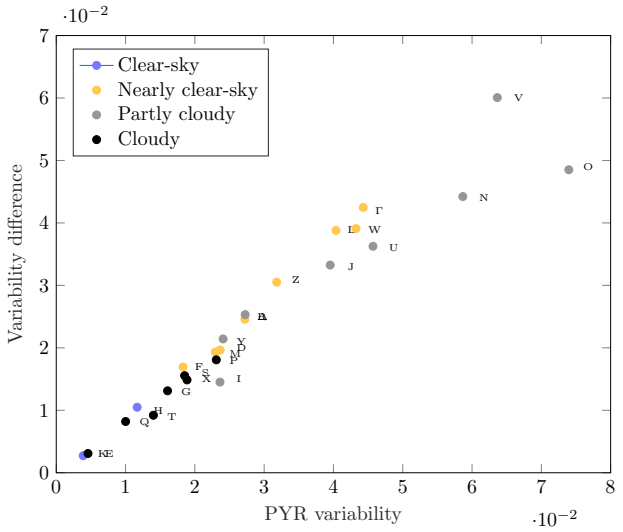

FIG. 16. Variability difference of `total PV generation` of satellite estimations vs variability of `Spatially averaged GHI` of pyranometers for the MV (left) and LV grid (right). Alphabetic labels refer to the 25 days of data.

tives due to large spatial variations of irradiance measurements, as the orange envelope in Fig. 14 shows, that are not captured by satellite estimations. Other cloudy days are not shown in the plots because either score zero false positive and negative, or their irradiance patterns do not determine grid violations. Similar considerations apply to the LV grid (right panel), with the differences that the proportion of FPs is lower and more cloudy days scores false negatives, denoting higher susceptibility to local irradiance conditions.

c. Duration and inter-arrival times The analysis on the duration and inter-arrival on the 25-day data set did not show specific patterns with respect to sky-conditions and are not reported here. Also, differences between the MV and LV grids did not emerge in the comparison.

d. Congested lines The plot in the upper left panel of Fig. 18 shows the occurrences of over-currents in the MV grid according to sky conditions. Over-currents mostly happen in nearly clear sky days (where irradiance feature very large values due to cloud enhancement) followed by clear-sky days and partly cloudy days. Congested lines are the number 3 and 15-17, which, from Fig. 4, are the closest to the GCP and the most critical because they are subject to the aggregated power flows. The plot in the bottom left panel of Fig. 18 shows the occurrences of over-currents false negatives according to sky conditions. They occur mostly in nearly clear-sky and partly cloudy days, and less in clear-sky days, when satellite estimations are able to detect irradiance dynamics better.

The plots in the right panels of Fig. 18 show the same results for the LV network. Congested lines are 1-3 and 11-16, which are those close the GCP and in the lateral feeders that interface a large amount of generation, respectively. Distribution of both over-currents and false negatives show the same features as discussed for the MV case.

VI. CONCLUSIONS AND PERSPECTIVES

a. Summary of the work We investigated the performance of HelioClim-3 v.5 satellite estimates of GHI compared to ground-based measurements from pyranometers (HOPE campaign) in applications to modeling the variability of distributed PV generation and detection of grid constraints violations (nodal voltages and line currents). The analysis was performed with load flow simulations considering the grid and demand specifications of the CIGRE benchmarks for LV and MV systems. Distributed PV systems were sited and sized with a procedure from the literature to saturate the hosting capacity of the grids. PV generation at each node was modeled as a function of installed capacity and GHI with a transposition model. Satellite estimations were de-biased on a daily basis to retain the focus the analyses on variability only. Pyranometers measurements were post-processed to remove outliers and potential other bad data close to missing measurements. Analysis were conducted at a 30 sec resolution.

b. Key results The analyses were carried out for 28 daily profiles of the irradiance with several typical sky conditions: clear-sky, nearly clear-sky, partly cloudy, and cloudy. It was found that:

- at the edge of the grids' PV hosting capacity, satellite estimations could not detect mild over-current episodes in both LV and MV grids, leading to false negative estimations;
- false negatives were more likely in partly cloudy (when local cloud enhancement phenomena affect PV production patterns) and cloudy days. False positives were more likely in nearly and clear-sky days due to a residual local positive bias of satellite estimations;
- false negative detections were persistent for long time intervals (e.g., tens of minutes) due to the residual negative bias of satellite estimations, as well as sporadic (e.g., single false negatives with inter-arrival times of minutes) due to fast fluctuations of the irradiance that were not captured by satellite estimations;
- satellite estimations underestimated the variability of the total PV production and the power flow at the GCP in both LV and MV grids. Underestimations were less severe in the MV grid case due to the larger geographical scale. Non-uniformly installed PV generation capacity and temperature were found affecting variability patterns, suggesting that these phenomena should be taken into considerations when, for instance, plan reserve requirements.

c. Implications and perspectives There is no disputing on the practical value of satellite estimations. However, results suggest that, when using irradiance estimations with similar attributes as those considered (e.g., spatiotemporal resolution and parallax), proper counter

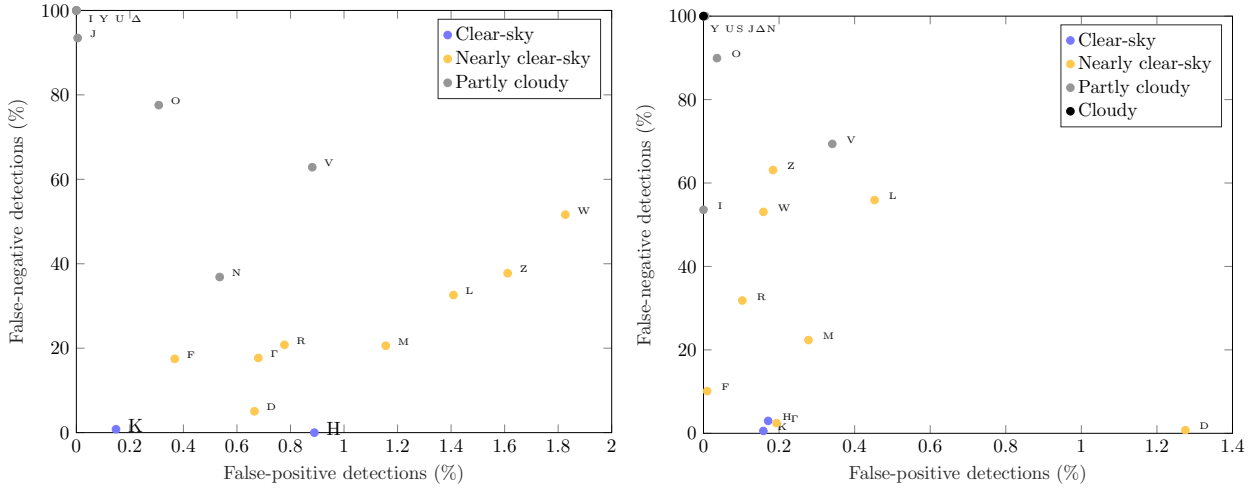

FIG. 17. False-negative versus false-positive detections of over-currents of satellite estimations for the MV (left) and LV grid (right). Alphabetic labels refer to the 25 days of data.

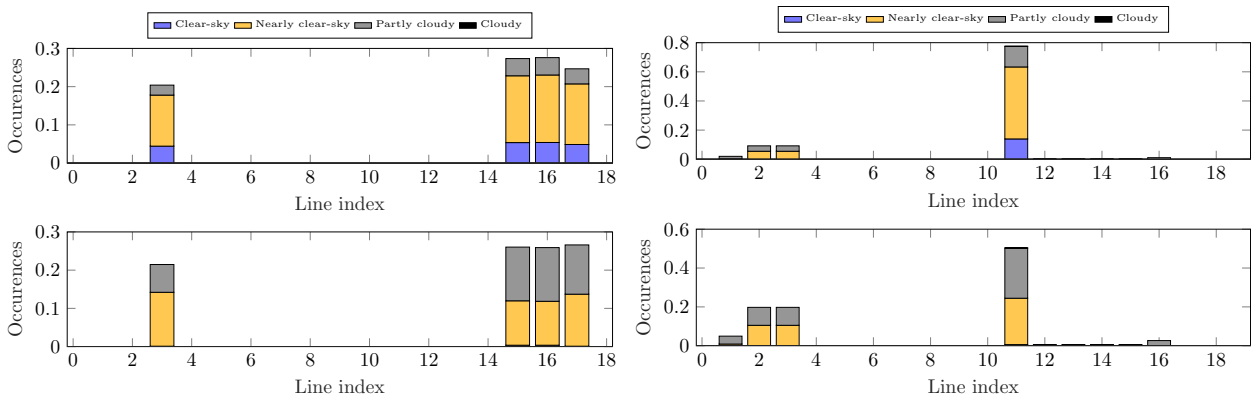

FIG. 18. Top panel: congested LV lines according to sky conditions in the MV (left) and LV grid (right). Bottom panel: LV lines with false negative detections according to sky conditions in the MV (left) and LV grid (right).

measurements or data reconciliation strategies with local estimates should be adopted to limit the shortcomings described above. These considerations also extend to forecasts, that are typically elaborated starting from satellite images and are delivered on relatively coarse geographical mesh grids.

Implications are case specific. For TSOs and balance responsible parties, coarse irradiance estimates in confluence with high levels of installed PV generation capacity can lead to underestimating the variability of PV generation, thus underestimating reserve requirements, which should be therefore scheduled adopting conservative margins (this can also apply to DSOs if grid codes require them to compensate deviations from pre-defined dispatch plans). DSOs, when planning the connection of new PV facilities, should also implement conservative margins from grid constraints or adopt realistic models of spatiotemporal variations, especially when near the PV hosting capacity of the grid, to avoid false negative detections. In the context of real-time and predictive control of grids for congestion management, dispatch, voltage control, and peak shaving, results suggest that spatiotemporal coarse estimations could be complemented with real-time local information at a higher resolution to perform better decisions.

Current results were elaborated from load flow simulations. The uncertainty related to pyranometers measurements and two different anisotropic transposition models (Hay-Davies and Perez) were found to impact on numerical results marginally without altering conclusions. If field data from an experimental setup with many grid-connected distributed PV plants will become available, a comparison with the proposed results would be of definite interest to confirm findings as they would inherently include all modeling uncertainties (i.e., direct/diffuse separation, transposition, panels' spectral response and efficiency, and distribution grid's parameters). Also, the effects of uncorrected bias should be further investigated; even if this is not expected to play a significant role on variability, it could certainly have a dominant impact on false positive and false negative detections of grid constraints violations.

ACKNOWLEDGMENTS

The research was partly funded under the FURIES and JASM programs of the Swiss Competence Center

on Energy Research. Authors warmly thank Dr. Deneke (Leibniz Institute for Tropospheric Research, Germany) for providing data of the HOPE campaign, Dr. Schmidt (EHF Department, University of Oldenburg, Germany) for having pointed out the HOPE campaign, and Prof. Jan Kleissl (University of California San Diego, USA) and the anonymous reviewers for providing us with relevant comments.

REFERENCES

[1] T. R. Ricciardi, K. Petrou, J. F. Franco, and L. F. Ochoa, "Defining customer export limits in pv-rich low voltage networks," *IEEE Transactions on Power Systems*, vol. 34, pp. 87–97, Jan 2019.

[2] J. Abrell, "The swiss wholesale electricity market,"

[3] A.-G. (TENNET), "Res forecasting from a tso perspective," As seen on July 2019. `http://www.ewea.org/events/workshops/wp-content/uploads/2014/01/EWEA-Forecasting-Workshop-2013-4-1-Alejandro-Gesino-Tennet-TSO.pdf`.

[4] AEMO Australian Energy market Operator, "Fast frequency response in the NEM," tech. rep., 2017.

[5] Q. Hong, M. Nedd, S. Norris, I. Abdulhadi, M. Karimi, V. Terzija, B. Marshall, K. Bell, and C. Booth, "Fast frequency response for effective frequency control in power systems with low inertia," *The Journal of Engineering*, vol. 2019, no. 16, pp. 1696–1702, 2019.

[6] H. Ruf, M. Schroedter-Homscheidt, G. Heilscher, and H. G. Beyer, "Quantifying residential pv feed-in power in low voltage grids based on satellite-derived irradiance data with application to power flow calculations," *Solar Energy*, vol. 135, pp. 692 – 702, 2016.

[7] J. Leloux, L. Narvarte, and D. Trebosc, "Review of the performance of residential pv systems in france," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 2, pp. 1369 – 1376, 2012.

[8] A. Nguyen, M. Velay, J. Schoene, V. Zheglov, B. Kurtz, K. Murray, B. Torre, and J. Kleissl, "High pv penetration impacts on five local distribution networks using high resolution solar resource assessment with sky imager and quasi-steady state distribution system simulations," *Solar Energy*, vol. 132, pp. 221 – 235, 2016.

[9] E. Rikos, S. Tselepis, C. Hoyer-Klick, and M. Schroedter-Homscheidt, "Stability and power quality issues in microgrids under weather disturbances," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 1, no. 3, pp. 170–179, 2008.

[10] M. Lave, J. Kleissl, and E. Arias-Castro, "High-frequency irradiance fluctuations and geographic smoothing," *Solar Energy*, vol. 86, no. 8, pp. 2190 – 2199, 2012.

[11] C. T. F. C6.04.02, "Benchmark systems for network integration of renewable and distributed energy resources," tech. rep., Cigre' International Council on large electric systems, July 2009.

[12] F. Sossan, J. Darulova, M. Paolone, A. Kahl, S. J. Bartlett, and M. Lehning, "Large scale deployment of pv units in existing distribution networks: Optimization of the installation layout," in *2016 Power Systems Computation Conference (PSCC)*, June 2016.

[13] B. Madhavan, J. Kalisch, and A. Macke, "Shortwave surface radiation budget network for observing small-scale cloud inhomogeneity fields," *Atmospheric Measurement Techniques Discussions*, vol. 8, no. 3, pp. 2555–2589, 2015.

[14] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. S. Homscheidt, C. Hoyer-Klick, A. Arola, *et al.*, "Mcclear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions," *Atmospheric Measurement Techniques*, vol. 6, pp. 2403–2418, 2013.

[15] C. Vernay, P. Blanc, and S. Pitaval, "Characterizing measurements campaigns for an innovative calibration approach of the global horizontal irradiation estimated by helioclim-3," *Renewable Energy*, vol. 57, pp. 339 – 347, 2013.

[16] I. Marie-Joseph, L. Linguet, M.-L. Gobinddass, and L. Wald, "On the applicability of the heliosat-2 method to assess surface solar irradiance in the intertropical convergence zone, french guiana," *International journal of remote sensing*, vol. 34, no. 8, pp. 3012–3027, 2013.

[17] Z. Qu, B. Gschwind, M. Lefevre, and L. Wald, "Improving helioclim-3 estimates of surface solar irradiance using the mcclear clear-sky model and recent advances in atmosphere composition," *Atmospheric Measurement Techniques*, vol. 7, no. 11, pp. 3927–3933, 2014.

[18] E. Skoplaki and J. Palyvos, "On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations," *Solar Energy*, vol. 83, no. 5, pp. 614 – 624, 2009.

[19] F. Sossan, L. Nespoli, V. Medici, and M. Paolone, "Unsupervised disaggregation of photo-voltaic production from aggregated power flow measurements of heterogeneous prosumers," *IEEE Transactions on Industrial Informatics, Accepted for publication*, 2017.

[20] L. Nespoli and V. Medici, "An unsupervised method for estimating the global horizontal irradiance from photovoltaic power measurements," *Solar Energy*, vol. 158, pp. 701–710, 2017.

[21] G. Friesen *et al.*, "Photovoltaic module energy yield measurements: Existing approaches and best practice," *Report IEA-PVPS T13-11*, 2018.

[22] J. A. Kratochvil, W. E. Boyson, and D. L. King, "Photovoltaic array performance model.," tech. rep., Sandia National Laboratories, 2004.

[23] Photon, "Photon databases of PV modules," 2017. https://www.photon.info/en/photon-databases.

[24] J. S. Stein, W. F. Holmgren, J. Forbess, and C. W. Hansen, "Pvlib: Open source photovoltaic performance modeling functions for matlab and python," in *Photovoltaic Specialists Conference (PVSC), 2016 IEEE 43rd*, pp. 3425–3430, IEEE, 2016.

[25] E. L. Maxwell, "A quasi-physical model for converting hourly global horizontal to direct normal insolation," *Unknown*, 1987.

[26] J. E. Hay and J. A. Davies, "Calculation of the solar radiation incident on an inclined surface," in *Proc. of First Canadian Solar Radiation Data Workshop (Eds: JE Hay and TK Won), Ministry of Supply and Services Canada*, vol. 59, 1980.

[27] J. E. Hay, "Calculating solar radiation for inclined surfaces: Practical approaches," *Renewable Energy*, vol. 3, no. 4, pp. 373 – 380, 1993. Solar radiation, environment and climate change.

[28] R. Perez, P. Ineichen, R. Seals, J. Michalsky, and R. Stewart, "Modeling daylight availability and irradiance components from direct and global irradiance," *Solar energy*, vol. 44, no. 5, pp. 271–289, 1990.

[29] M. Lave, W. Hayes, A. Pohl, and C. W. Hansen, "Evaluation of global horizontal irradiance to plane-of-array irradiance models at locations across the united states," *IEEE Journal of Photovoltaics*, vol. 5, pp. 597–606, March 2015.