Spatial Aggregation of Small-scale Photovoltaic Generation Using Voronoi Decomposition


Published in:
IEEE Transactions on Sustainable Energy

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2020 IEEE. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Spatial Aggregation of Small-scale Photovoltaic Generation Using Voronoi Decomposition

Javier López Lorente, Student Member, IEEE, Xueqin (Amy) Liu, Member, IEEE, and D. John Morrow, Member, IEEE

Abstract—In this paper, a methodology based in Voronoi decomposition is proposed to spatially aggregate small-scale solar generation. The locations of relevant electrical infrastructure are used to manage the uncertainty on locating solar photovoltaic installations. The known coordinates of step-down high- to medium-voltage electrical substations (i.e. bulk supply points) are used to divide the territory and find multiple representative locations where solar resource can be assessed. Modelling solar photovoltaic generation from global solar radiation observations permits the estimation of power output and degradation factor due to age for the entire small-scale photovoltaic fleet. The results are validated against multiple solar installations across the region of study (Northern Ireland, UK) and show a relatively low root mean square error with monthly values ranging from 0.036 to 0.123 kW/kWp. The proposed method is scalable to larger and smaller geographical areas and transferable to other categories of solar photovoltaic installations. This methodology can serve as a basis for multiple applications, such as solar generation forecasting. System operators could utilise this method to improve knowledge of when, where and in what amount additional resources would be required to manage solar penetration in favour of a robust, low-carbon and efficient power network.

Index Terms—small scale solar photovoltaic, distributed generation, upscaling method, Voronoi decomposition.

I. INTRODUCTION

Solar photovoltaic (PV) technology plays an essential role in decarbonising our societies and economies. In 2017, it was the single renewable energy technology with the highest annual growth [1]. Solar PV systems have brought along new challenges to the management of power networks at different levels, from low-voltage feeders to grid-scale. The number of small-scale solar PV installations have largely expanded due to incentive schemes and levelised cost of electricity reductions in the recent years [2]. In light of these challenges and growth, solar generation forecasting models can be used to quantify solar production of both individual and fleets of PV systems. Solar generation forecasting help system operators and aggregators to efficiently manage and plan the electricity network that they operate [3].

As solar radiation varies in both a spatial and temporal dimension, the higher the resolution of these dimensions, will a better understanding be obtained on how generation of PV systems varies. Ideally, production from PV installations should be considered individually, where particular characteristics of each site would be studied. This would permit the assessment of the independent effect that each PV system produces and its influence in a larger PV systems’ fleet. Aggregation or scale-up methods are utilised to estimate the whole generation of PV fleets to assess their impact on the electricity networks. These methods consider the spatial distribution of PV fleets and are mainly based on up-scaling using clustering, subsetting or random allocation techniques. An up-scaling method using k-means clustering of PV sites based on their known location was proposed by Pierro et al. [4] to estimate PV generation at regional level. This method made use of weather forecasts at the geometric centre of the cluster of the installations to estimate solar production. Clustering and sub-clustering techniques had been also used by Wolff et al. [5] for PV power forecasting. Subsetting combined with interpolation was evaluated by Saint-Drenan et al. [6], where the performance was analysed based on the average distance between reference PV plants. Clustering and subsetting techniques require the coordinates of the PV systems to be assessed. Random spatial allocation of PV fleets usually considers the total installed capacity in a region, as well as a typical installation’s size and assigns PV locations based on certain criteria. These criteria may include: (i) type of land use; (ii) type of area: urban or rural; (iii) number of customers or load density; and (iv) distance to transmission and/or distribution networks. Random allocation of residential PV systems considering urban and rural areas and PV concentration based on load density was performed in [7]. The authors assessed the impact of PV on medium-voltage distribution networks for power losses and voltage level analysis. The main shortcoming of these methods is the requirement of detailed and extensive amount of data, as they are data-driven approaches.

Characteristics of each plant, e.g. hardware configuration, orientation, tilt angle, among others, are not readily available. Due to general data protection legislations, the exact location or address of individual small-scale solar PV installations are generally withheld information. Distribution system operators or aggregator companies would not exchange, when available, such data from their customers due to this type of legislation. There are exceptions when individuals voluntarily decide to share their data and participate in crowdsourcing platforms. For example, technical characteristics, location information and sometimes historic generation data from small-scale PV installations in the residential sector are available in platforms like PVOuput.org [8]. The number of installations in such platforms are often a minority for a particular region or country. In large geographical areas, e.g. a town or a city, the available data of a relatively small spatial dimension regarding...
small-scale generators and abiding to the requirements of personal data protection legislations, may be sites referred to by postcode as studied in [9] or at the bulk supply points. Bulk supply points are the step-down high- to medium-voltage electric substations, e.g. 110/33 kV, supplying electricity from the transmission to the distribution network. The installed capacity of small-scale solar PV connected to these substations would represent the starting point from where PV generation and its impact on the power network could be analysed for a particular area. The industrial need to efficiently estimate the expected output generation is now more than evident, since the thousands of installed PV systems scattered across large geographical areas make aggregated forecasting challenging.

In this paper, the installed capacity of solar PV connected to bulk supply points and their known coordinates are used to obtain representative locations using Thiessen or Voronoi decomposition. In those local representative locations, solar generation is then computed using global solar irradiance data following a physical approach. Solar production can be then estimated for each local area and as a whole. The results are validated against available power output data from solar PV installations in the region of study and the proposed approach is compared to existing scale-up methods from the literature.

The novelty of this paper is the use of Voronoi decomposition to find representative locations to scale-up solar generation from a large fleet of small-scale PV generators with unknown location. The contributions are manifold. This proposed solution manages the uncertainty in the location of PV sites, sometimes protected by general data protection legislations. This method could serve as a primary approach to spatial distribution of renewable sources due to its flexibility in terms of: (i) scalability to larger or smaller areas; (ii) applicability to PV fleets of different sizes, small- and large-scale; and (iii) the reduced number of data types required. Besides this scaling-up model, the paper also contributes to the understanding of spatial variability of solar resource. A new use to an existing PV performance metric, namely the variability index [10], is proposed to identify days when solar generation can pose a challenge to the power networks. The spatial distribution of small-scale PV systems are also analysed based on the installed capacity in each bulk supply point and the relationship with the human settlements, e.g. cities and towns, is illustrated. This work also serves as a groundwork in relation to the operation of power systems, such as regional solar generation forecasting and electrical load management, which are core tasks of the electricity system operators. The paper is organised as follows: Section II describes the methodology, case study and data sources used. The results are presented, validated and discussed in Section III. Finally, conclusions are drawn in Section IV.

II. METHODOLOGY, CASE STUDY AND DATA SOURCES

A. Methodology

The methodology and its methods can be grouped in four differentiated stages: (1) spatial division and estimation of representative locations; (2) estimation of global plane-of-array solar irradiance; (3) estimation of aggregated electricity generation; and (4) selection of days of interest.

1) Spatial division and estimation of representative locations: The proposed approach for the spatial distribution of PV systems is based on Voronoi decomposition, a stochastic model for space division. In a Voronoi or Thiessen diagram, the space is divided into polyhedral tessellations or polygons where each of them consists of all points closer to the known point, also called nucleus or generator point, than any other point. The polygon created by those points defines an area of influence around its nucleus [11]. A generic Voronoi diagram and its elements is illustrated in Fig. 1, where each of the vertices in the resulting Voronoi polygon corresponds to the point of intersection of the perpendicular bisectors formed by the triangulation of the generator point \( p_i \) and other two points.

Mathematically, let \( \phi = \{x_1, x_2, \ldots, x_i, \ldots, x_n\} \) be a locally finite system of points in the d-dimensional Euclidean space \( \mathbb{R}^d \). Each location in \( \mathbb{R}^d \) is associated to its nearest point(s) belonging to \( \phi \). The neighbourhood or Voronoi cell \( C(x_i, \phi) \) of a point \( x_i \), the nucleus or generator, of \( \phi \) is defined by:

\[
C(x_i, \phi) = \{ y \in \mathbb{R}^d : \|y - x_i\| \leq \|y - x_j\| \forall j \neq i \}. \tag{1}
\]

The \( C(x_i, \phi) \) are all convex polygons with the \( x_i \) as inner points. The Voronoi diagram represents the union of all Voronoi polygons developed from the generating points \( \phi \) [12].

Voronoi diagrams are an interdisciplinary concept that have been broadly used in many different fields including, but not limited to, archaeology, astrology, cartography, computational geometry, geography, geology, marketing, meteorology, physics, and urban and regional planning [13]. In the field of power systems, Voronoi decomposition has been used for the design and planning of large distribution networks. Optimal locations and sizes for power distribution substations and transformers using Voronoi polygons were used in [14]–[16]. Spatial division with Voronoi diagrams was also used as an approach for planning infrastructure of electric vehicles in a trade-off among economic cost, traffic congestion and power

![Fig. 1: Generic Voronoi diagram for a given point \( p_i \) and its components.](image-url)
grid constraints [17]. In those cases, Voronoi decomposition helped to find an optimised spatial division based on the considered planning zones.

In this paper, Voronoi decomposition is applied to the known locations of bulk supply points, i.e. 110/33 kV electric substations. Bulk supply points (BSPs) are considered for two main reasons: (i) BSPs may be the smallest available granularity of installed PV data as per anonymity and personal data protection; and (ii) distribution codes often establish that distributed generation issues (e.g. reverse power flows) in the distribution network system shall not reach the transmission system, that is, go upstream of BSPs. In this manner, electrical infrastructure related to renewable distributed generation is considered as part of the scale-up stage to estimate regional PV production. This establishes an approach that links the cause and effect to evaluate the impact of small-scale generation on power networks. Developing the understanding of PV installations beyond their spatial distribution.

Using as input points the geographical location of the BSPs, Voronoi polygons were obtained using an open-source tool for geographical information systems (QGIS [18]). The produced Voronoi tessellations were further analysed to obtain their centroid as a representative local point in that area, where solar irradiation from satellite observations could then be analysed and PV generation computed using a photovoltaic physical performance model. Finally, the regional PV generation was obtained based on the installed capacity connected to each BSP and corrected by a factor reflecting the age degradation of the ATM system, that is, go upstream of BSPs. In this manner, electrical infrastructure related to renewable distributed generation is considered as part of the scale-up stage to estimate regional PV production. This establishes an approach that links the cause and effect to evaluate the impact of small-scale generation on power networks. Developing the understanding of PV installations beyond their spatial distribution.

The geometric centre or centroid of each resulting Voronoi tessellations is a representative local point of the area of study. The use of the centroids as representative study points has been performed. The results will be illustrated in Section III.

Obtaining the centroids of the resulted Voronoi polygons.

The centroid as a representative local point in that area, where solar irradiation from satellite observations could then be analysed and PV generation computed using a photovoltaic physical performance model. Finally, the regional PV generation was obtained based on the installed capacity connected to each BSP and corrected by a factor reflecting the age degradation of the ATM system, that is, go upstream of BSPs. In this manner, electrical infrastructure related to renewable distributed generation is considered as part of the scale-up stage to estimate regional PV production. This establishes an approach that links the cause and effect to evaluate the impact of small-scale generation on power networks. Developing the understanding of PV installations beyond their spatial distribution.

Using as input points the geographical location of the BSPs, Voronoi polygons were obtained using an open-source tool for geographical information systems (QGIS [18]). The produced Voronoi tessellations were further analysed to obtain their centroid as a representative local point in that area, where solar irradiation from satellite observations could then be analysed and PV generation computed using a photovoltaic physical performance model. Finally, the regional PV generation was obtained based on the installed capacity connected to each BSP and corrected by a factor reflecting the age degradation of the ATM system, that is, go upstream of BSPs. In this manner, electrical infrastructure related to renewable distributed generation is considered as part of the scale-up stage to estimate regional PV production. This establishes an approach that links the cause and effect to evaluate the impact of small-scale generation on power networks. Developing the understanding of PV installations beyond their spatial distribution.

3) Estimation of Plane-Of-Array Solar Irradiance: This stage includes the steps used to model the total or global solar plane-of-array irradiance \( (I_{poa}) \). First, the solar global horizontal irradiance \( (GHI) \) data were obtained at the coordinates of the centroids. Second, the diffuse solar radiation fraction was decomposed from GHI using the Erbs [19] model. The Erbs decomposition model is used to estimate direct normal irradiance \( (DNI) \) from measured GHI. This model applies an empirical relationship based on a piecewise segmented function between the clearness index \( (k_l, \text{Eq. 2}) \) and the diffuse fraction \( (k_d, \text{Eq. 3}) \). The clearness index is the fraction of the solar extra-terrestrial radiation that is transmitted through the atmosphere to strike the surface of the Earth. It is a dimensionless number between zero and one.

\[
k_l = \frac{GHI}{E_a \cdot \cos(\theta_Z)} \quad (2)
\]

\[
k_d = \frac{DHI}{GHI} \quad (3)
\]

Where, \( E_a \) is the extra-terrestrial irradiance at the top of the Earth’s atmosphere and \( \theta_Z \) is the solar zenith angle. DHI is the diffuse horizontal irradiance and DNI can then be calculated using Eq. 4.

\[
GHI = DHI + DNI \cdot \cos(\theta_Z) \quad (4)
\]

Third, plane-of-array (POA) components were calculated with the Hay and Davies’ transposition model [20]. This model determines the diffuse irradiance from the sky on a tilted surface using DHI, DNI, extra-terrestrial irradiance, solar azimuth angle, solar zenith angle, surface tilt angle, and surface azimuth angle. These models, Erbs and Hay and Davies, were selected as their combination has been shown to achieve higher accuracy and less bias than other combinations of models. Comparison details can be found in [21], [22]. Finally, the total POA irradiance was computed using the following equation:

\[
I_{poa} = I_{beam} + I_{d,sky} + I_{d,ground} \quad (5)
\]

where, \( I_{beam} \) is the direct or beam irradiance, \( I_{d,sky} \) is the diffuse irradiance produced by the direct sunlight scattered through molecules and particles in the atmosphere and \( I_{d,ground} \) is the ground diffuse irradiance, all three modelled at a designated plane-of-array. The ground diffuse irradiance is modelled using Eq. 6 [23], which depends on the ground albedo \( \rho \), surface tilt angle from horizon \( \beta \), and GHI.

\[
I_{d,ground} = GHI \cdot \rho \frac{1 - \cos \beta}{2} \quad (6)
\]

The characteristics of the tilted surface in terms of typical azimuth and tilt angles used in the transposition model were selected considering average values in the UK. These values were selected as an approximation to the characteristics of the PV systems in the region of the study and had been presented in [24]. An azimuth angle of 178.93° and a tilt angle of 31.8° were used. The outputs of this stage are multiple time-series of total POA irradiance at the centroids of the resulted Voronoi polygons.

3) Estimation of Aggregated Electricity Generation: The estimation of the electricity generation of the entire regional PV fleet consisted of three steps: (a) solar power generation is modelled from total POA irradiance, (b) estimation of the PV fleet effective capacity and (c) estimation of the aggregated PV generation.

a) Estimation of power generation from solar radiation: This stage involves the estimation of the electricity generation of the entire regional PV fleet. The conversion from sunlight to electricity must consider affecting factors to the performance of the PV system, namely: (i) radiation, (ii) module temperature, (iii) local shadows produced by nearby trees, buildings or structures, (iv) the aging of the modules, (v) losses in electrical wiring, and (vi) solar inverter’s efficiency.
Estimating the electricity generation of the PV module from solar radiation requires mathematical algorithms that model its physical behaviour. This type of algorithm are known as photovoltaic performance models. The model presented by Huld et al. [25] was used in this study, which has been widely validated and applied in the solar energy industry. In this model, the power output is assumed to depend on irradiance \( G \) and the module temperature \( T_m \), as described by Eq. 7 and Eq. 8.

\[
P = \frac{G}{G_{STC}} \cdot P_{STC} \cdot \eta_{rel}(G', T'_m) \tag{7}
\]

\[
\eta_{rel}(G', T'_m) = 1 + k_1 \ln(G') + k_2 \ln(G')^2 + k_3 T'_m + k_4 G'_m \ln(G') + k_5 G'_m \ln(G')^2 + k_6 T'_m^2 \tag{8}
\]

where, \( G_{STC} = 1.000 \text{ Wm}^{-2} \), \( P_{STC} \) is the rated power of the module, which was set to 1 kW/kWp to obtain normalised values, and \( \eta_{rel} \) is the relative efficiency of the module as function of corrected values for irradiance and module temperature according to standard test conditions (STC). \( G' \) and \( T'_m \) are defined as: \( G' = \frac{G}{G_{STC}} \) and \( T'_m = T_m - 25 \).

The coefficients \( k_1 \) to \( k_6 \) depend on the PV module technology. It is assumed that the distribution of the PV fleet in the region corresponds to the market shares of the different PV modules’ technologies. Wafer-based crystalline silicon (c-Si) are about 92% of the commercial modules, while thin film represent the remaining 8% of the market. Amorphous silicon (a-Si) modules account for less than 1% of the thin film market, cadmium telluride (CdTe) for 5%, and copper indium gallium selenide (CIGS) for 2% [26]. The module temperature, \( T_m \), depends on air temperature and irradiance, and changes in wind speed can alter the cooling effect on the module. The module temperature can be modelled according to Faiman’s model [27]:

\[
T_m = T_a + G \cdot (U_0 + U_1 \cdot W)^{-1} \tag{9}
\]

where, \( T_a \) is the air temperature in degrees Celsius and \( W \) is the wind speed expressed as \( \text{m}\cdot\text{s}^{-1} \). The heat loss coefficients \( U_0 \) and \( U_1 \) equal 25.0 \( \text{W}\cdot\text{m}^{-2}\cdot\text{K}^{-1} \) and 6.84 \( \text{W}\cdot\text{m}^{-3}\cdot\text{S}\cdot\text{K}^{-1} \), respectively [27].

Concerning the losses, these were calculated using typical values in PV systems and were computed as for Eq. 10 [28]. The values used were: soiling losses due to dirt in the cells (2%), shading (3%), mismatch due to manufacturing imperfections that slightly modify the cells current-voltage characteristics (2%), wiring losses (2%), resistive losses in electrical connections (0.5%), light-induced degradation of the PV cells (1.5%), nameplate rating accuracy (1%), and availability losses due to system’s scheduled and unscheduled shutdowns for maintenance or other operational reasons (3%) [28]. The normalised total power output \( P_{out} \), that considers all the losses was calculated according to Eq. 11.

\[
L_{total}(\%) = 100 \left(1 - \prod_{i} \left(1 - \frac{L_i}{100}\right)\right) \tag{10}
\]

\[
P_{out} = P \cdot \left(1 - \frac{L_{total}}{100}\right) \cdot \eta_{inv} \cdot \eta_{age} \tag{11}
\]

Where \( L_{total} \) is the overall percentage of losses, \( L_i \) are the different losses types expressed in percentage, \( P \) is the electric power output of the module as for Eq. 7, 8 and 9; and \( \eta_{inv} \) is the efficiency of the inverter, where a typical value of 0.96 was selected. \( \eta_{age} \) corresponds to the age degradation factor for the whole installed PV fleet and is described below.

b) Estimation of the PV fleet effective capacity: Deterioration over time for a single PV site could be modelled considering its installation date. However, for an entire PV fleet of thousands, this information would not be readily available. Therefore, a theoretical effective installed photovoltaic capacity \( P_{eff} \), defined as per Eq. 12, was modelled assuming linear performance based on typical values of minimum guaranteed performance over time from manufacturers of PV modules. Manufacturers would guarantee that the maximum output power is not less than a percentage of minimum values specified in the data-sheet of the PV module. Generally, these percentages are, at least 97% after the first year, 90% by year 10 and 80% by year 25. Given those three reference points, a first order polynomial of the type \( y = mx + b \) could be fitted, where \( y \) is the performance of the PV module in percentage and \( x \) is the time of the installation expressed in months. The points used were: \( Year(x, y) = 1(12, 97), 10(120, 90), 25(300, 80)\); being the resulted function: \( y = -0.0587x + 97.449 \). There is a different \( \eta_{age} \) for each time step considered, as the existing and newly installed PV sites modify this factor every time the total PV fleet capacity changes.

\[
P_{eff} = \sum_{i=0}^{k} \frac{mx + n}{100} \cdot P_i \tag{12}
\]

where \( P_{eff} \) is the effective installed capacity in the region corrected by age degradation, \( k \) is the number of the month for which the degradation is estimated, \( m \) and \( n \) are the above mentioned coefficients of the fitted first order polynomial, \( P_i \) is the new installed capacity for each \( k \) step, and \( x = k - i \). After calculating the effective installed capacity, the age degradation factor is estimated as \( \eta_{age} = P_{eff}/P_{total} \), where \( P_{total} \) is the accumulated installed capacity in the region of study up to the time of \( k \).

c) Estimation of the aggregated PV generation: After estimating the normalised PV generation \( P_{out} \) in the centroid of each polygon according to the available solar resource and the age degradation factor \( \eta_{age} \), the aggregated solar PV generation in each Voronoi polygon can be calculated multiplying those values by the installed capacity in each BSP \( P_{installed} \). The all-region solar PV generation \( P_{region} \) would thus be equal to the sum of the generation of each area:

\[
P_{region} = \sum_{i=1}^{n} \left( P_{out} \cdot \eta_{age} \cdot P_{installed} \right).
\]

4) Selection of days of interest: The methods in the previous stages of the methodology are sufficient to estimate the overall solar PV generation of a large fleet of installations scattered across large territories. These methods provide a time-series that can be extended as long as the timeline of the data available. Certain criterion was then defined to select days of interest for further analysis. The mean daily variability index (VI) was used for this purpose. This index developed by Stein et al. [10] was proposed as a metric to measure variability in irradiance and power output of single PV sites.
considering both global horizontal and clear sky irradiance.

\[
VI = \frac{\sum_{k=2}^{n} \sqrt{(GHI_k - GHI_{k-1})^2 + \Delta t^2}}{\sum_{k=2}^{n} \sqrt{(CSI_k - CSI_{k-1})^2 + \Delta t^2}}
\]  (13)

where, \(GHI_k\) and \(GHI_{k-1}\) are vectors of length \(n\) of global horizontal irradiance values averaged at a stated time interval in minutes, \(\Delta t\), and \(CSI_k\) and \(CSI_{k-1}\) are vectors calculated from clear sky horizontal irradiance values for the same stated times as the previous vectors.

The daily VI score was applied to the totality of resulting areas, i.e., 33 areas, using 1-minute resolution time-series of GHI. After calculating the mean daily VI score for each day and location, the standard deviation across all the values was calculated for each day as an approach to define an inter-area metric to select days of interest considering solar resource and PV output variability. Using the standard deviation of daily VI scores helps understand how solar resource varies over time and space across large-geographical areas. At the same time, it provides insights of local weather factors affecting solar resource and consequently, highlighting differences in solar PV generation across subareas.

B. Case study and Data sources

Data from Northern Ireland (UK) for all 2018 were used. Small-scale solar generators accounted for 110.8 MW, connected to the distribution network and spread across 16,107 sites, in January 2019. The majority of small-scale solar generators are rooftop installations in domestic or small commercial premises with an installed capacity below 11 kW. Industrial, agricultural, commercial premises and small solar plants above that capacity and up to 5 MW also fall in the category of small-scale generators. The data used were:

- **Geospatial data.** Geospatial information from Northern Ireland was used to work out the spatial Voronoi decomposition in the boundaries of the region. Shapefiles that provided location, shape and attributes of geographical information were obtained from the open data portal OpenDataNI [29].

- **Solar radiation data.** The Copernicus Atmosphere Monitoring Service (CAMS) all-sky radiation service provided solar radiation data at the desired coordinates. Time series data of GHI and clear-sky irradiance with temporal resolution of 1 minute were obtained from the Copernicus portal [30].

- **Weather data.** Temperature and wind speed to be used in the PV performance model for each particular location were taken from their closest official weather station in the closest official weather station belonging to the UK land surface observing network [31].

- **Solar photovoltaic fleet data.** The distribution system operator of Northern Ireland, NIE networks, provided data on small-scale photovoltaic systems connected at the bulk supply points in Northern Ireland. The solar PV capacity and the number of sites connected to each bulk supply point are the data with the smallest spatial dimension made available adhering to general data protection laws. Data regarding the monthly deployment of solar photovoltaic capacity used to compute the age degradation factor of the whole PV fleet, are made available by the Department for Business, Energy & Industrial Strategy of the UK [32]. Electricity generation from small-scale PV systems in Northern Ireland were used to validate the methodology. These datasets were obtained from the open platform PVOutput.org [8] and from a test PV array available to the authors of this study.

III. RESULTS AND DISCUSSION

The results of the Voronoi decomposition to divide the country (Northern Ireland in this case study) in local areas for the assessment of solar resources and photovoltaic generation are presented in Fig. 2. The centroids of the Voronoi tessellations or polygons, the solar installed capacity connected to the electric bulk supply points in each resulting area and the human settlements over 500 people are illustrated. The resulting area associated to each bulk supply point (i.e. Voronoi polygons) was numbered to ease identification, in ascending order West to East according to the longitude of the bulk supply point. It can be observed that the distribution of the main human settlements and solar PV installed capacity across the country have a negative relationship. For example, the capital city of Belfast, areas 19 and 23-26 in the North East, and the second-largest main city, Londonderry, areas 3, 4 and 6 in the North West of the country, present the lowest numbers of installed solar PV capacity. This indicates that small-scale solar installations may be largely present in rural areas rather than in urban areas. This finding is consistent with that of Westacott and Candelise [33], who found a predominance of domestic grid-connected PV installations in rural areas. Additionally, our study extends this trend to all type of small-scale solar PV installations, instead of only those in the domestic sector.

A. Inter-area analysis of solar resource

The standard deviation of the daily VI scores across the 33 resulting Voronoi polygons for the year 2018 is presented...
in Fig. 3. The day with the highest deviation across all the areas was the 14th June, followed by the 1st March with 0.543 and 0.513, respectively. Contrarily, the days with the lowest standard deviation of their daily variability indexes were 29th June and 17th November with 0.005 and 0.019, respectively. It can be noted that similar dispersion of daily VI across areas can be found during very sunny days or periods (late-June) and completely overcast days (early January or mid-November). Furthermore, this shows that inter-area standard deviation does not relate to meteorological seasons as the maximum and the minimum standard deviation of mean daily VI scores were recorded over two weeks’ time in summer. The selection of the clear sky radiation model to compute clear sky irradiance in the VI score (Eq. 13) will intrinsically lead to variations thereof, as the atmosphere and its composition are modelled differently. The data source of GHI will also affect the outcome of the VI scores, in this case, using satellite observations for both clear and global horizontal radiation provided smoother, less dispersed VI scores in comparison to land observations. Nonetheless, the use of the standard deviation of daily VI scores across a number of areas can facilitate the assessment of spatial solar resource variability in large-geographical areas.

Intra-hourly spatial and temporal analysis was performed to the days of maximum (14 June) and minimum (29 June) standard deviation of mean daily VI score across all resulted areas. Fortuitously, both days are near and within the same month. This favours their comparison in terms of incoming solar radiation. Figure 4 presents heat maps illustrating the spatial and temporal evolution of the 15-minute mean global horizontal irradiance at the centroid of each area associated to a bulk supply point across the country. It can be observed that for the day of minimum standard deviation of mean daily variability index (Fig. 4-A) the distribution of global solar irradiance across all the areas is homogenous with the exception of some clouds in areas 9 and 10 in the early morning.

There are slightly lower GHI values during the core hours of the day in some areas, probably due to the presence of few high thin clouds, for instance, cirrus clouds. In comparison, on the day of maximum standard deviation of mean daily VI (Fig. 4-B) there is a strong contrast in the observed radiation values across the areas, which is noticeable from adjacent to distant areas. For example, some areas between area 1 to 20 experience differences over 300 W m$^{-2}$ during the core hours of the day.

B. Estimation of the overall small-scale solar generation

For the designated days of interest, normalised solar PV power per area was computed and it is shown in Fig. 5. It can be observed that PV systems in all the areas generated were subject similar power output per installed unit in 29 June. Contrary to the day of maximum standard deviation of daily VI score, 14 June, when the maximum daily solar energy recorded was 6.13 kWh per kilowatt peak installed in area 27, representing a difference 42.5% compared to the minimum value recorded in area 15 of 3.52 kWh/kWp. As shown in Fig. 4-B, areas 8, 9 and 21 to 33 experienced longer cloudless periods during that day generated more energy.

Considering the installed capacity connected to each BSP, the aggregated power output in all the region for the designated days of interest was estimated. The maximum power output ($P_{region}$) were 76.7 MW and 76.9 MW, in the days 14 and 29 June 2018, respectively. While the maximum electricity power output in both days are similar in range, the contribution of each area to the total drastically changed throughout the 14 June. The assessment of the individual power output of each area reveals where solar PV is subject to raise issues to the power network. According to the location and nature of small-scale generators, it is evident that large part of the electricity generated is self-consumed behind the meter. In addition, when surplus of solar production is fed back to the electricity network, this is probably consumed locally. Nevertheless, the aggregated solar generation of large PV fleets remains of high interest as it affects the instantaneous electricity demand. First, locally in low-voltage networks, and extending its impact along the power system up to the transmission system. At transmission system level, the total electricity demand of a region or country can be masked by small-scale solar generation. Consequences of this demand mask can be disturbances in the operation of the power system, requirements for load enhancement or load relief and alterations in the unit commitment and scheduling of large generators.

C. Validation of the results

1) Validation of the model: Many small-scale solar generators are generally not monitored nor controlled. This complicates the access to data from multiple PV sites to validate models. In this case, this data constraint was solved by accessing power generation from PV sites available in the open platform POutPut.org. Fourteen sites with available intra-day power output during 2018 were retrieved. Additionally, data from a rooftop PV array existing in the urban campus
Fig. 4: Daily evolution of 15-minute mean global horizontal irradiance (GHI) across all the areas for the day of minimum, 29 June 2018 (A), and maximum standard deviation of daily VI score, 14 June 2018 (B).

Fig. 5: Normalised PV power outputs (kW/kW\textsubscript{p}) per bulk supply point area for the days of interest.

of Queen’s University Belfast was included. Altogether, fifteen solar PV installations with an accumulated capacity of 71.08 kW, across twelve of the considered areas, and with resolution ranging from 5 to 15 minutes were used to validate the estimated power outputs of the model.

The mean 15-minute power output of the fleet formed by the available sites across the country was compared against the equivalent modelled output at the areas’ centroids that a PV fleet of similar size would obtain. The metric used was the root mean square error (RMSE), which is the standard deviation of the residuals or prediction errors and it is common and widely extended in the literature. The results of the overall data set, in Fig. 6, present a RMSE 6.834 kW, what represents a normalised RMSE of 0.096 kW/kW\textsubscript{p}. Monthly RMSE varies from 0.036 to 0.123 kW/kW\textsubscript{p} in December and May, respectively. The results are in range with those in [6] for a distance between reference PV plants of 20 km, where real data was utilised, RMSE: 0.0750 to 0.1287 kW/kW\textsubscript{p}. This distance is similar to the average separation between centroids.

The model is prone to estimating higher generation. These differences may be subject to the assumptions done in terms of losses of the system, for example, the real non-constant performance of solar inverters, which is considered constant in the model, or additional shadowing. Real differences between the azimuth and tilted angles of each installation compared to the typical values used may have also led to positive bias. In the same figure, it can be observed that the bias is higher closer to the winter period (more density of dark green to blue colours above the fitted line), what may be result of how GHI is calculated under all-sky conditions from satellite measurements.

2) Benchmarking — comparison to other models: There are several approaches to scale-up geographically distributed datasets, these can either use single- (e.g. spatial median) or multiple-representative locations (e.g. random allocation or clustering). In this subsection, two approaches available in the literature are compared to the proposed Voronoi method in terms of performance. These selected methods are the most robust of each category among the single- and multiple-representative locations’ techniques. In this case and for the purpose of validation, the differences in the application of these methods only relate to the selection of location/s utilised to model PV output from solar irradiance data. The same physical performance models and parameters were used to obtain the aggregated power output with each method. The classic techniques to assess geographically distributed datasets evaluated are: (i) spatial median or median centre [34] and (ii) clustering [35], with the latter having been used by others for the study of solar PV [4], [5].

The median centre or spatial median method identifies the location that minimises the accumulated travel distance among the features in a dataset. The element selected presents a central tendency that is robust to outliers, it is located in the
median of the dataset and corresponds to an existing location in the dataset [34]. Applied to this study, the accumulated distance from each BSP to the rest were obtained and the median distance among all would correspond to the location of the spatial median. The single location corresponding to the median centre resulted the BSP area 12 as numbered in Fig. 2. The clustering method was implemented using the coordinates of the validation PV sites, which were estimated from the location displayed in the online platform used for the validation data, i.e. PVOutput.org. Then, the analysis of solar resources were computed in each of the clusters’ centres. The clustering method requires a distance radius or a grid division to group the installations. As 30 km corresponds to the maximum horizontal distance for the validity of cloud cover in a local area [36], this distance and half this distance were selected to assess this method. The clustering method was a two-step process where: (i) the distance radius was applied to each location, where the areas that overlap conformed a cluster from which the centroid was computed; and (ii) solar power output was then modelled as per the methodology described at each cluster’s centroid through obtaining the generation output as the normalised output in kW/kWp times a proportional weight based on the installed capacity in each sub-cluster. The total solar power generation was modelled for a PV fleet with the same size than the total capacity of the validation sites (i.e. 71.08 kW).

The metrics for each of the two methods in comparison to the Voronoi method are presented in Table I. The clustering method with 15 km resolution is the technique with the highest accuracy among the methods reproduced from the literature. Nevertheless, the proposed methodology based on Voronoi decomposition outperforms both. While the coefficient of determination (R-square) of the proposed model and clustering with 15 km distance radius are almost equal, the RMSE is lower in the Voronoi decomposition.

![Fig. 6: Estimated vs observed 15-minute mean solar power output generated by the validation PV fleet.](image)

<table>
<thead>
<tr>
<th>Scale-up Method</th>
<th>RMSE (kW)</th>
<th>nRMSE (kW/kWp)</th>
<th>R-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Centre</td>
<td>10.796</td>
<td>0.1519</td>
<td>0.6700</td>
</tr>
<tr>
<td>Clustering (30 km)</td>
<td>9.629</td>
<td>0.1355</td>
<td>0.7323</td>
</tr>
<tr>
<td>Clustering (15 km)</td>
<td>8.940</td>
<td>0.1258</td>
<td>0.7499</td>
</tr>
<tr>
<td>Voronoi dec. (proposed)</td>
<td>6.834</td>
<td>0.0962</td>
<td>0.7492</td>
</tr>
</tbody>
</table>

**IV. CONCLUSIONS AND FUTURE WORK**

The objective of this paper was to propose a new method that could be employed by system operators and utility companies, to enable the calculation of aggregated PV fleets power output and energy production in the context of their operational and planning activities. This study explored the spatial variability of solar resource and the spatial distribution of small-scale PV sites according to the density of human settlements in large geographical areas.

A scaling-up methodology was proposed to distribute solar resources considering the location of electrical substations, in this case, bulk supply points. The methodology based on Voronoi decomposition permitted the spatially aggregation of electricity generation from small-scale solar PV in large territories considering electrical infrastructure. The novelty in the proposed methodology is the use of Voronoi decomposition in the electrical infrastructure where solar generators are connected and obtain local representative locations where assess solar resources. The proposed method is scalable to smaller and larger geographical areas and transferable across categories of solar PV installations, from small-scale (e.g. residential installations), as in this study, to large-scale (e.g. solar farms). This approach permits the utilisation of the spatial distribution of grid infrastructure to analyse grid-connected solar systems without knowing any exact location of the PV sites assessed. Considering multiple representative areas also facilitates dealing with the inherent uncertainty in defining the representativeness of a single location for the analysis of the solar generation of PV fleets in large territories. The study also contributes to the understanding of solar variability resource in large geographical areas. It has shown how inter-area spatial and temporal variability can be evaluated considering multiple representative areas to obtain insights on available solar resource and photovoltaic generation. This paper reaffirms previous work on the spatial distribution of domestic solar installations and expands such statements to all types of grid-connected small-scale solar, domestic and non-domestic, installations. Small-scale PV generators in the residential sector tend to concentrate in rural areas. The validation results were based on the aggregated 15-minute mean generation of a group of solar generators against the modelled production of a PV fleet of similar size and distributed in the same sub-areas. The Voronoi method produced relatively low error metrics RMSE and high R-square, and outperformed other scaling-up techniques reported in the literature.

Further research could elucidate the value of aggregated small-scale solar generators and their effect in the system’s
demand at a regional or country level. The proposed methodology can serve as a foundation for multiple power systems applications. For instance, it can be used for solar generation forecasting when using solar forecasted radiation values as inputs. Additionally, it could be used to effectively identify areas of the power network where solar PV might raise issues, allowing resources to be assigned (e.g. demand side response battery storage units) to maintain a robust and efficient power system.

**ACKNOWLEDGMENT**

The authors would like to thank Northern Ireland Electricity Networks for providing data on the PV deployment of small-scale installations. The authors would also like to acknowledge the financial support from SPIRE 2 (Storage Platform for the Integration of Renewable Energy) project. The SPIRE 2 project is supported by the European Union’s INTERREG VA Programme (Grant No. INT-VA/049), managed by the Special EU Programmes Body (SEUPB). The views and opinions expressed in this paper do not necessarily reflect those of the European Commission or the Special EU Programmes Body.

**REFERENCES**


