GIS-based Software Infrastructure to Model PV Generation in Fine-grained Spatio-temporal Domain

Original

Availability:
This version is available at: 11583/2677937 since: 2018-09-05T12:00:37Z

Publisher:
IEEE

Published
DOI:10.1109/JSYST.2017.2726350

Terms of use:
openAccess
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)
GIS-based Software Infrastructure to Model PV Generation in Fine-grained Spatio-temporal Domain

Lorenzo Bottaccioli, Edoardo Patti, Enrico Macii and Andrea Acquaviva.

Abstract—Nowadays, we are moving forward to more sustainable societies, where a crucial issue consists on reducing footprint and greenhouse emissions. This transition can be achieved by increasing the penetration of distributed renewable energy sources together with a smarter use of energy. To achieve it, new tools are needed to plan the deployment of such renewable systems by modelling variability and uncertainty of their generation profiles.

In this paper, we present a distributed software infrastructure for modelling and simulating energy production of Photovoltaic (PV) systems in urban context. Its core, performs simulations in a spatio-temporal domain exploiting Geographic Information Systems together with meteorological data to estimate Photovoltaic generation profiles in real operating conditions. This solution provides results in real-sky conditions with different time-intervals: i) yearly, ii) monthly and iii) sub-hourly. To evaluate the accuracy of our simulations, we tested the proposed software infrastructure in a real world case study. Finally, experimental results are presented and compared with real energy production data collected from PV systems deployed in the case study area.

Index Terms—Distributed Software Architecture, Geographic Information System, Solar Energy, Photovoltaic, Smart-city, Renewable energy sources.

I. INTRODUCTION

URING the international conference on climate changes (COP21) in 2015, the 196 parties attending the conference in Paris highlighted the need of reducing greenhouse gas emissions [1]. On this regard, in the last years, many countries are providing incentives to promote the deployment of low-carbon and sustainable technologies [2], such as Photovoltaic (PV) Systems. This implies both an increasing installation of Renewable Energy Sources (RES) and a smart use of energy in our cities [3]. Thus, specific tools to evaluate resource availability and uncertainty of RES are required: i) to understand their impact on power grids; ii) to perform load balancing; iii) to perform storage planning and management; iv) to perform demand-side management at different scales, from single user up to district or city; v) to provide generation profiles to electricity markets (e.g. day-ahead or intra-day market). By analysing the generation loads in fine-grained spatio-temporal domain (e.g. sub-hourly simulations of Photovoltaic systems at district scale), such tools should be able to overcome the current techniques estimating potential PV generation. The outcome of these tools is twofold. First, the impact on both power grids and electricity markets can be evaluated more accurately. Second, new policies for smart energy use can be developed. Information and Communication Technologies (ICT) have been identified as key player to foster this transition and to develop such innovative tools.

In particular, Geographic Information Systems (GIS) can become useful tools for planning renewable energy systems [4] in urban context and for evaluating their energy performances in spatio-temporal domain. In particular concerning PV systems, space and time domains must be considered to perform more accurate estimations of energy production [5]–[7]. Furthermore, the integration of these two domains helps to understand spatio-temporal dynamics in energy systems models [4], [8]. Literature solutions are limited in analysing solar potential in a spatial domain. Thus, they neglect temporal analysis (e.g. hourly or sub-hourly) to better model PV systems.

To overcome such limitations, we propose an innovative software infrastructure to estimate generation profiles of PV energy systems by performing fine-grained spatio-temporal simulations (e.g. 15-minutes simulations with a spatial resolution lower than 1 m). In its core, it computes real-sky conditions in urban context by simulating the incident radiation on tilted surface of buildings considering real meteorological data. Indeed, our solution integrates Volunteer Geographic Information (VGI) to exploit information available from personal weather stations [4].

Finally, we designed our infrastructure to ease the integration with third-party software that can exploit our results to develop tools for further analysis and evaluations (e.g. [9]–[12]).

Exploiting our software infrastructure, Single citizen can evaluate the economic and environmental savings achievable installing PV systems. Energy aggregators and Energy Communities can use simulation results to schedule consumption of their customers to maximize self-consumption. In particular, Energy Communities can exploit such infrastructure to perform feasibility studies and evaluate economic benefits [13]. Energy managers and PV system engineers can simulate the behaviour of their systems in real-sky conditions. This can help in sizing, validating and optimizing each system before and after the installation. Distribution System Operators (DSO) can take advantage of the proposed solution for network balancing and for planning retrofits and/or extensions of existing distribution grids. Finally, Energy and City planners can evaluate the impacts of large PV systems installations in city districts.

The rest of the paper is organized as follows. Section II reviews relevant background literature. Section III introduces the proposed software infrastructure to estimate PV energy production in urban contexts. Sections IV and V present respectively the case study and the experimental results. Finally, Section VI discusses concluding remarks and future works.
II. RELATED WORK

Geographic Information System (GIS) is considered a useful tool to plan the deployment of renewable energy sources, such as solar, wind, and biomass systems [4]. Particular emphasis is given to such technology for modelling solar potential in urban environments [8]. For example, the authors in [14] and [15] exploit GIS tools to estimate the yearly Photovoltaic (PV) potential starting from aerial and satellite images. Another approach consists in exploiting Digital Surface Models (DSMs) or 3D city models obtained from LiDAR data. DSM represents the earth’s surface and includes all objects and buildings on it. For example starting from DSM, Hofierka et al. [16] estimated monthly and yearly solar potential in urban areas using r.sun tool [17]. While exploiting LiDAR data, Brito et al. [18] estimated yearly PV potential in Lisbon using ESRI Solar Analyst tool. These solutions just perform a time-domain analysis to estimate yearly solar potential. However, to provide more precise PV estimations, both space and time domains must be taken into account [5]–[7]. In [8] and [4], authors highlight that the integration of these two domains, with higher time and space resolutions, is needed to better understand spatio-temporal dynamics in energy systems models. This is required to i) plan deployment activities; ii) evaluate business plans; iii) monitor existing plants and iv) promote smart energy use.

On this premise, in order to provide simulation tools to wide range of users, GIS solutions have been developed exploiting a Web Service approach. Li et al. [19] have developed a service-oriented environment for sharing geoscientific algorithms. They exploited both SOAP (Simple Object Access Protocol) technologies and OGC (Open Geospatial Consortium) standards to make available GRASS-GIS [20] features through Web Services. Gwass [21] is a distributed web-based GIS built on top of the GeoBrain Web Services. This platform exploits a service-oriented architecture to offer an alternative to commercial desktop solution.

Literature provides also web-based solutions [22]–[29] to give PV energy potential information and to foster assessments of environmental and economic benefits as pointed out by Freitas et al. [8]. PVWatts [24] is a web application developed by the National Renewable Energy Laboratory that estimates yearly (Y), monthly (M) and hourly (H) PV production using a Typical Meteorological Year (TMY) and a topographic model of 40 km². PVGIS [22], [23] is a solar web map that offers information on yearly and monthly PV production in Europe and Africa. It provides also sub-hourly radiation information in clear-sky conditions. To perform this computations, it exploits r.sun starting form a DSM with a resolution of 1 km. i-GUESS [26] is a web based system for urban energy planning in smart cities. It provides maps for yearly solar radiation on rooftops and for yearly PV potential. Mapwell Solar System [25] computes solar radiation and PV potential considering also a TMY.

It also provides information on rooftops and Region Of Interest (ROI). I-SCOPE [27] is an integrated platform to give 3-D smart city services. In particular, it offers a solar map with Yearly and Monthly PV potential. Finally, Brumen et al. [28] developed a web application for PV potential assessments by exploiting r.sun starting from DSM. This platform provides data about yearly and monthly PV Potential together with information on rooftops and ROI. Such information on rooftops and ROI are also provided by I-SCOPE and i-GUESS services. Table I report the main features of the relevant literature solutions.

SAM [30] is another literature solution for PV modelling. It is a non-GIS tool that has been a standard for several years. It integrates third-party weather data (given as off-line files) to perform sub-hourly simulations. However, SAM does not simulate possible scenarios for PV deployment at urban scale that take into account surrounding buildings.

The main limitation of presented solutions consists on overlooking a fine-grained spatio-temporal domain in simulating and modelling energy production and performance of PV systems. Indeed, they are mainly focused on spatial domain by performing yearly or monthly simulations. On the other hand, existing spatio-temporal solutions perform hourly simulations with a low-resolution in spatial domain (e.g. resolution > 1 m). To provide more accurate estimations, hourly and sub-hourly simulations with fine-grained resolution (e.g. DSM with resolution < 1 m) are instead needed. This high-resolution DSM allows to recognize and exclude encumbrance in rooftops, such as chimneys and dormers. Moreover, such simulations have to take into account real-sky conditions. To do so, they need real weather data (e.g. data from personal weather stations [4]) to compute incident radiation on tilted surface of rooftops and estimate PV performance and energy production.

As such, our contributions with respect to state-of-the-art, detailed in Table I, include the following: i) analysing together spatial and temporal domain in fine-grained resolution; ii) providing real-sky sub-hourly simulations, with 15-minutes time intervals; iii) integrating real meteorological data gathered from (personal) weather stations. Compared to current solutions that are monolithic software, we propose a distributed modular architecture based on a microservices approach [31], [32], where each module provides its Web Services to retrieve data. This approach eases the integration of our solution with third-party software and fosters the development of new services.

In our previous work [33], we provide a high-level description of the overall idea to develop a software infrastructure to model PV Generation in fine-grained spatio-temporal domain. With respect to [33], in this paper, we provide detailed description of the both methodologies i) to identify the suitable area where PV system can be installed and ii) to compute
the energy production of PV systems in real-sky conditions. In this paper, we also describe and compare the integrated decomposition models to estimate direct and diffuse solar radiation from the global solar radiation and how these have been applied to compute the incident radiation on pitched rooftops in real-sky conditions. Finally, we present the results about the extensive characterization we performed in a real-world case study carried out in Turin, Italy.

The results of the proposed solution can be conveniently applied to many contexts. For example, they can provide a better estimation of PV systems’ impact on the power grids [34] as well as on the electricity market [10], [35]. In addition, they can be exploited to implement novel control policies for a smart energy usage, such as demand response.

III. DISTRIBUTED SOFTWARE INFRASTRUCTURE TO ESTIMATE PV ENERGY PRODUCTION

In this work, we propose a distributed and modular software infrastructure that expose REST Web Services [36] to perform solar energy simulations (see Figure 1). Its modularity takes advantage of the microservices approach, which consists on developing software as a suite of small services, each running in its own process and communicating with lightweight mechanisms [31]. This increases flexibility and maintainability because services are small, highly decoupled and focus on doing a small task [32]. Our solution needs as main inputs the Digital Surface Models (DSMs) and the Cadastral maps. DSM is a raster image representing terrain elevation with buildings. A DSM with a high-resolution (in the order of sub-meters) permits i) to better recognize encumbrance in rooftops, such as chimneys and dormers, that would not allow deployment of PV panels and ii) to have a better simulation of shadows that will affect the PV energy production. Thus, higher is the DSM resolution, higher is the accuracy of the energy production estimation. Cadastral map is a vector image reporting the square footage of buildings in the area of interest. It also exploits meteorological data coming from third-party sources, such as personal weather stations deployed in cities. In particular, the needed inputs are solar radiation, ambient temperature and wind speed.

A. Data-source Integration Layer

The Data-source Integration Layer (the lowest layer in Figure 1) is in charge of integrating in the infrastructure the following heterogeneous information: i) Digital Surface Model (DSM); ii) Cadastral map; iii) Linke Turbidity coefficients [38] and iv) Weather data. The Linke Turbidity coefficients express the attenuation of solar radiation related to air pollution. This information can be automatically retrieved by third-party services, such as [23], [39], or can be specified by end-users before executing the simulation. Finally, Weather data are also retrieved by third party services, such as [40], and collected by personal weather stations deployed in cities. In particular, the needed inputs are solar radiation, ambient temperature and wind speed.

B. Services Layer for real-sky simulations

The Services Layer (the middle layer in Figure 1) is in charge of i) simulating solar radiation in real-sky condition; ii) identifying rooftops areas suitable for deploying PV systems and iii) evaluating their energy production. It consists of different software modules. It is worth highlighting that each of them exposes REST Web Services. Hence, each module can be invoked by third-party software to retrieve information and simulation results.

1) Clear-sky condition Service: In order to compute clear-sky solar radiation we exploit GRASS-GIS open-source software, which embeds in its core r.sun [17]. The r.sun tool provides an accurate simulation of solar radiation in urban contexts [8], [16], [41]. The resulting outputs of this module are set of direct and diffuse solar radiation maps in clear-sky condition with 15 minutes time interval. Such maps are stored in the Maps Data-store ready to be used by the Real-sky calculation service.

It performs such computation considering the monthly average days identified by [42] and reported in Table II with the related Day of the Year (DOY). DOY is a 1 to 365 non-dimensional sequential index starting from January 1st. For instance: January 17th is day 17, February 16th is day 47, December 31st is day 365.

The module requires as inputs: i) the DSM, ii) monthly Linke turbidity coefficients, iii) Slope and Aspect maps, that
are produced with GRASS-GIS r.slope.aspect tool. Monthly Linke turbidity coefficients are retrieved by using third party web services such as [23], [39] or can be specified by end-users as parameters in the execution request. The Slope and Aspect maps represent respectively the inclination and the orientation (expressed in degrees) of each pixel of the DSM. After their calculation, they are stored in the Maps Data-store.

2) Suitable surface identification service: Thanks to the high-resolution DSM given as input to the overall simulation process, this module identifies available surface for deploying solar systems on pitched rooftops excluding, for instance, dormers and chimneys. The Suitable surface identification service uses as inputs both Slope and Aspect maps for identifying the suitable area. These maps are retrieved from the Maps Data-store. By default, it identifies areas representing tilted rooftops oriented between South-Est and South-west. Commonly, a tilted rooftop has a slope ($\theta$) in-between $10^\circ \leq \theta \leq 45^\circ$. This range is also suitable to install a PV system. Furthermore, facades with an orientation ($\gamma$) in-between $220^\circ \leq \gamma \leq 320^\circ$ (considering South facing roofs having $\gamma = 270^\circ$) are more exposed to solar radiation. For these reasons, we chose these ranges for $\theta$ and $\gamma$ as the default values for the Suitable surface identification service. However, the end-user can give as input new ranges for $\theta$ and $\gamma$ to select the desired suitable surfaces.

The output of this selection is a binary map where pixels with 1 as value represent the available areas. Such map is then filtered with GRASS-GIS r.neighbors tool to smooth noise and to remove small areas that are too small for installing a PV system (e.g. areas where deployable PV system are smaller than 1 kW). This resulting map is vectorized and clipped with the cadastral map stored in Maps Data-store. The information on the size of area of the resulting polygons is a 2D projection of the real roof surface. To estimate the real surface of polygons, the Formula (1) is applied to correct the value of the area with inclination angle of rooftops.

$$S = \frac{S_{2D}}{\cos(\theta)}$$

where $S_{2D}$ is the 2D area of the polygon and $\theta$ is the roof inclination obtained from the slope map, again retrieved from the Maps Data-store.

The output of this module is a GeoJSON reporting a number of polygons that represent deployable areas.

3) Solar radiation decomposition service: Nowadays in our cities, weather stations are pervasively deployed and their information are also provided by third-party services, such as [40]. Normally such stations are equipped with global horizontal solar radiation sensors and do not provide information on direct normal and diffuse horizontal radiation. However, to simulate real-sky solar radiation on a pitched surface, information on direct and diffuse radiation is needed [17]. To overcome this limitation, the Solar radiation decomposition services integrates six different solar radiation decomposition techniques in literature [43]–[48] to compute both direct normal and diffuse horizontal radiation starting from measurements of global horizontal radiation.

These models can be categorized by the number of predictors. In the following we briefly introduce them. Erbs et al. [43], Reindl et al. [47] and Karatasou et al. [46] use only the clearness index $k_t$ as predictor. $k_t$ is the ratio between global radiation and extraterrestrial radiation both on a horizontal plane. Skartveit et al. [48] use $k_t$, solar altitude and solar zenith angles as predictors. Ruiz-Arias et al. [45] use $k_t$ and the air mass as predictors. In their work, Engerer et al. [44] provide three different models. Our solution implements the second model with the following predictors: $k_t$, the zenith angle, the time of the day, the clear-sky global radiation on horizontal plane and a variability index, which represents the deviation of the observed $k_t$ value from the clear sky value of the clearness index. The end-user can specify the model suitable for the area of interest. Indeed, as reported by [49]–[51], the accuracy of decomposition models is strongly affected by different latitude, longitude and environmental conditions. Hence, the integration of these decomposition models and third-party meteorological services makes our infrastructure flexible in performing simulation in different geographic locations. The input of this module is a time series of the global horizontal radiation provided by third-party meteorological services in Data-source Integration Layer. The output of this process is a JSON with the values of Direct Normal Incident radiation (DNI) and Diffuse Horizontal Incident radiation (DHI) for the requested time interval.

4) Real-sky calculation service: This module produces maps of incident global radiation on pitched rooftops in real-sky conditions. The inputs of this service are values of DNI, DHI, direct and diffuse solar radiation maps retrieved from the Maps Data-store. Through the Data-source Integration layer, the Real-sky calculation service retrieves information on solar radiation from third-party services. Both values of DNI and DHI radiation are required to calculate solar radiation on tilted surface [17]. If meteorological services provide only global horizontal radiation and not its direct and diffuse components, the Real-sky calculation service invokes the Solar radiation decomposition services to compute such information. Thus, Real-sky calculation service uses the values of DNI and DHI for each time interval to calculate the clear-sky indexes $k_{b}^{c}$ (2) and $k_{d}^{c}$ (3).

$$k_{b}^{c} = \frac{DNI_{\text{overcast}}}{DNI_{\text{clear-sky}}} \quad (2)$$

$$k_{d}^{c} = \frac{DHI_{\text{overcast}}}{DHI_{\text{clear-sky}}} \quad (3)$$

$k_{b}^{c}$ (2) represents the ratio between DNI in overcast conditions and DNI in clear-sky conditions. While, $k_{d}^{c}$ (3) is the ratio between DHI in overcast conditions and DHI in clear-sky conditions. Finally for each time interval, diffuse and direct radiation maps, produced by the Clear-sky condition Service, are multiplied by clear-sky indexes $k_{b}^{c}$ and $k_{d}^{c}$. Then, both maps are summed together to obtain a global incident radiation map in real-sky condition. The resulting output of this process is a set of GeoTIFF images representing the maps with incident global solar radiation in real-sky condition.

5) Photovoltaic energy estimation service: This module is in charge of estimating PV production for each area identified by Suitable surface identification service. The required inputs are i) the maps provided by the Real-sky calculation service and ii) the GeoJSON given by the Suitable surface identification service where polygons representing deployable areas are
α

η

U

G

T

wind speed is not provided by the nearest weather station. It

Data source integration layer

values (i.e.

layer, third-party solutions can retrieve simulation results and

Web-Map displayers

proposed infrastructure (see Figure 1). It is dedicated to end-

case study. Turin is a city located in Piedmont, north-west of Italy. The

study district. The weather station collects global horizontal radiation by a first class

pyranometer that samples every minute. Then, these samples are averaged and provided every 15 minutes. As

proposed by [55], we exclude samples of solar radiation with:

i) the altitude lower than 7° and ii) the clearness index lower than 0 and higher than 1. Furthermore, we excluded the measured samples of global horizontal radiation with higher values than in clear-sky condition, again as suggested by [55].

To validate our solution, we considered three different PV systems and a tilted solar radiometer located in the case study district. Thus, we compared results on PV energy

IV. CASE STUDY

In order to test and validate the simulation of the proposed software infrastructure, we selected a district in Turin as case study. Turin is a city located in Piedmont, north-west of Italy. The district under analysis is La Crocetta, where there is our University campus. La Crocetta is located in the city centre with an area of about 3.7 km² and around 2200 residential buildings. It has been selected because of its buildings, which are heterogeneous for construction type and period. DSM and Cadastral maps for this area have been provided by the city council. DSM has a resolution of 0.25 m², which gives the possibility to describe with high accuracy rooftops, highlighting encumbrance like chimneys and dormers. Throug third-party web services, meteorological data (i.e. solar radiation and air temperature) are retrieved by the weather station in our University campus that is located in the middle of the case study district. The weather station collects global horizontal radiation by a first class pyranometer that samples every minute. Then, these samples are averaged and provided every 15 minutes. As proposed by [55], we exclude samples of solar radiation with: i) the altitude lower than 7° and ii) the clearness index lower than 0 and higher than 1. Furthermore, we excluded the measured samples of global horizontal radiation with higher values than in clear-sky condition, again as suggested by [55].

To validate our solution, we considered three different PV systems and a tilted solar radiometer located in the case study district. Thus, we compared results on PV energy

estimations on PV energy production to perform further analysis and evaluations.

C. Application Layer

The Application Layer represents the highest layer of the proposed infrastructure (see Figure 1). It is dedicated to end-user applications, such as Web-Map displayers and Dashboards, that can provide information about performed simulation across the city with different level of details. In addition at this layer, third-party solutions can retrieve simulation results and
In this section, we present the experimental results performed in the case study described in Section IV. First, we present the results achieved by the Real-sky calculation service compared with the real data collected by the digital pyranometer in our University campus. This test has been performed for all the solar radiation decomposition methods integrated in the proposed infrastructure in order to select the best method suitable for the case study area. Then, the resulting generation profiles, obtained by the Photovoltaic energy estimation service, are compared with measured data of Campus, Sommellier and GalFer PV systems. Finally to highlight the advantages of our solution, we performed simulations with one day time-interval for Sommellier and GalFer and we present the comparison of our simulation results with the simulation results obtained in PERSIL [56]. To evaluate the performance of our simulations, we exploited the indicators reported by Güeymard [57]. In particular, we considered the following four indicators of dispersion. i) The Mean Bias Difference (MBD) measures the average squares of errors between predicted and measured values. ii) The Root Mean Square Difference (RMSD) represents the standard deviation of differences between predicted and observed values. iii) The Mean Absolute Difference (MAD) is defined as the average of the absolute difference between the variance and the predicted variable. These indicators of dispersion, with exception of $r^2$, are expressed in percentage of mean measured values rather than in absolute units as suggested in [57]. Furthermore, we considered the following indicators for the overall performance. i) The Willmott’s Index of Agreement (WIA) is the standardized measure of the degree of model prediction error. It varies between 0 and 1. ii) The Legates’s Coefficient of Efficiency (LCE) is the ratio between the mean square error and the variance in the observed data, subtracted from unity. $LCE$ can vary between $-\infty$ and 1, where 1 represents the perfect model.

### A. Selection of best decomposition model for case study area

In order to select the best decomposition model for the case study area, all the models implemented in the infrastructure have been tested because their accuracy is strongly affected by the geographic location (see Section III-B3). The best model has been selected by comparing solar radiation simulations with measurements sampled by the solar radiometer in our University campus. The tests consisted on simulating solar radiation from the 1st of January 2014 to 31st of December 2015. Table IV reports the performance indicators of simulations for each decomposition model under analysis with time-intervals of 1 hour and 15 minutes respectively.

### V. Experimental Results

In this section, we present the experimental results performed in the case study described in Section IV. First, we present the results achieved by the Real-sky calculation service compared with the real data collected by the digital pyranometer in our University campus. This test has been performed for all the solar
The analysis of performance indicators for hourly simulations highlights that the best model for the case study area (Turin, Italy) is Karatasou [46]. Indeed, among its performance indicators, this model has the best values with exception of \(MBD\), which is \(-1.88\%). Indeed for \(MBD\) only, Karatasou is worst than the Ruiz-Aris model [45] with the \(MBD = -0.68\%\). In 15 minutes simulation, Karatasou again has the best performance indicators with exception of \(MBD\) and \(MAD\). Indeed, Ruiz-Aris achieves the best \(MBD\), which is \(-0.5\%\); while Reindl [47] has the best \(MAD\) equal to 16.21%.

Table IV highlights that all these decomposition models have better performance in simulation with 1 hour time-interval. This is also confirmed by Gueymard et al. [49], where authors prove that solar decomposition models developed to have the best performance with slow variations of \(k_t\) (e.g. hourly resolution) cannot provide same performance with faster variations of \(k_t\) (e.g. 15 minutes resolution).

Finally, Table V details the performance indicators for season and weather conditions (Sunny, Cloudy and Rainy) applying the Karatasou model [46]. Table V highlights that best results are achieved in sunny days of summer and autumn. Indeed for all seasons, \(MAD\), \(MBD\) and \(RMSD\) have lower values in sunny days. The \(r^2\) is higher or equal to 0.9 for all seasons except for of rainy days in summer. On the other hand, simulations in winter season provide the worst results with respect to the other seasons, in particular during rainy days. More in general, rainy days for the fous seasons do not have the same good performance than in sunny days, with \(RMSD\) higher than 30%. This is due to the accuracy of the integrated decomposition models in evaluating direct and diffuse components of solar radiation in rainy days. However, the results of 15-minutes simulations in rainy day (worst case) are still acceptable. We strongly believe that the exploiting sensors to sample direct and diffuse radiation will improve the performance of the simulations.

### B. Evaluation of PV system simulation against measured data

In order to validate the estimation of generated instant power, we perform simulations for the three PV systems introduced in Section IV: i) Campus, ii) Sommellier and iii) GalFer. Then we compared our results with the real instant power sampled every 15 minutes. In our simulations for each PV system, we considered an yearly degradation factor of 1% for the efficiency coefficient \(\eta\). For calculating the temperature \(T_{\text{sol-air}}\), we set the loss factor \(k\) to 0.05 as reported by [5].

Figure 3 shows the comparisons among our results and measured generation load profiles for the Campus PV system in a time period from the 8th of May 2014 to the 31st of December 2015. In particular, Figure 3 reports the plots of the instant power for three generic days in autumn: i) sunny, ii) cloudy and iii) rainy. It is worth noting that the trends of our results follow the real behaviour of the PV system with a good accuracy.

This is also highlighted by Table VI that reports the results in terms of performance indicators of instant power for the Campus PV system. The best results are achieved in sunny days of autumn. Indeed, \(r^2\) is higher than 0.9; \(RMSD\) and \(MAD\) are respectively lower than 13% and 10%.

On the other hand, simulations performed in summer provide the worst results with respect to the other seasons, because the temperature strongly affects the efficiency coefficient \(\eta\) of the PV module.

Considering the whole time period (18 months), the accuracy of simulation performance for daily energy production increases with respect of 15 minutes time resolution. Daily simulations are computed as the integral of 15-minute simulations; thus, errors tend to be attenuated.

The other PV systems under analysis, both Sommellier and GalFer, have been monitored from the 1st of March 2010 to the 22nd of February 2011 and the data sampling has been done by the PERSIL project consortium [56]. Table VII and VIII report the performance indicators of our simulation compared with...
measured data for Sommellier and GalFer respectively. The simulations have been performed with different time-frames: i) 15 minutes, ii) 1 day and iii) 1 year. In both Sommellier and GalFer, the accuracy of performance indicators for simulations with daily and yearly time-frames increase with respect to 15 minutes simulations. LCE, WIA and $r^2$ are not calculated for yearly values because they can be applied to series and not to a single value. With respect to 15 minutes simulations, the accuracy of performance indicators for Sommellier simulations is in line with the one of Campus. Apparently, GalFer presents the worst accuracy. However, if the analysis is restricted to the period between 1st of March 2010 and 15th of May 2010 the results are in line with the other PV systems. This is due to a malfunction of the PV system. Further details on this particular behaviour are discussed in next Section V-D.

### C. Comparison with PERSIL simulation methodology

During our studies and tests, we also compared our solution with PERSIL simulation methodology [56], with whom we partially share the same case study (i.e. Sommellier and GalFer PV systems). Figure 4 and Figure 5 show the real measured daily energy production compared with results of our solution and with results of PERSIL methodology for July 2010. Both Figures point out that PERSIL overestimates daily energy production with respect to our solution. In particular comparing PERSIL with our solution, RMSD and MAD for Sommellier are 22.92% and 21.77% respectively. Regarding GalFer, it is worth noting that both solutions overestimate daily energy production due to the malfunction in the system occurred after the 15th of May 2010 (see in Section V-D). However also in this case, our solution improves the estimations as show in Figure 5. Indeed comparing both methodologies, RMSD and MAD for GalFer are 23.31% and 22.15% respectively. This is also highlighted and quantified by comparing the accuracy of performance indicators obtained by PERSIL in Table IX with ours in Table VII and VIII. This analysis underlines how our software infrastructure better describes the energy production with respect to PERSIL. Regarding the estimation of daily energy production for Sommellier, our solution reduces RMSD of 32%, MBD of 28.88% and MAD of 24.34%. Considering the yearly energy production for the same PV system, our solution decreases RMSD of 35.66%, both MAD and MBD of 35.79%. About the estimation of daily energy production for GalFer before the malfunction, our solution reduces RMSD of 34.30%, MAD of 29.44% and MBD of 25.72%. Considering the yearly energy production for GalFer, our solution increases the accuracy by reducing the RMSD of 42.79%, both MAD and MBD of 42.76%.

In our case, the accuracy increases because our simulation process takes as inputs an high resolution DSM that is needed to simulate possible shadow behaviours. In addition, our solution consider $T_{sol-air}$ to better estimate the temperature for the PV cells. Finally, we computes the daily energy production as a sum of 15 minutes energy simulations. Thus, the effects of temperature on the PV efficiency is evaluated with real and actual values; while PERSIL exploits daily mean value.

### D. PV system operation assessment

As mentioned in Section I, the proposed distributed software infrastructure can help *Energy managers and PV system engineers* in monitoring the performance of already deployed PV systems. This can help in planning maintenance activities...
by identifying systems that are not working properly. During our tests on GalFer and after a deep analysis of its generation loads, we identify a malfunction of the system that affects its performance from 17Th of May 2010. Indeed, this anomaly decreases the efficiency of the modules reducing the maximum peak power. This is pointed out by analysing the peak power in a sunny day before and after 17Th of May 2010. Indeed till 17Th of May 2010, the simulation trends follow the real behaviour of the PV system. In particular, the peak power in 17Th of May 2010 is 6354.77 [W], with a global horizontal radiation of 844.10 [W/m²] in sunny day. Instead, in 20Th of April 2010, still a sunny day, the peak power is 7545.36 [W] with a global horizontal radiation of 810.5 [W/m²]. This anomaly is also highlighted in Figure 6, where simulation results are compared with real measured values in sunny, cloudy and rainy days before and after 17Th of May 2010. It is worth highlighting that this anomaly is not verified for Sommellier, which has almost the same system characteristics and geographic location of GalFer (see Section IV). Indeed, in the same days under analysis, the peak power is 10027.05 [W] and 9937.60 [W] respectively. In conclusion, we believe that results of the proposed solution can be used also to feed algorithms for detecting faults and/or malfunctions in PV systems.

VI. CONCLUSION AND FUTURE WORKS

In this work, we presented our distributed software infrastructure for modelling PV energy production in urban contexts. It overcomes the limitations of literature solutions by providing fine grained real-sky simulations considering also meteorological data from weather stations. Thus, addressing the challenges highlighted in [4]. Furthermore, the proposed solution performs simulations on spatial and temporal domains, providing energy profiles of PV systems with a good accuracy, as reported by the experimental results. Finally, our infrastructure can be use as a tool to help Energy managers and PV system engineers in monitoring the performances of existing PV systems as depicted in the analysed GalFer system, where a malfunction has been identified.

As future works, we will enhance the Photovoltaic energy estimation service by introducing detailed hardware models for the most relevant PV systems in the market. These hardware models are also needed to consider the effect of partial shading on PV systems.