UNIVERSIDAD POLITÉCNICA DE MADRID

Instituto de energía solar

**ESCUELA TÉCNICA SUPERIOR
DE INGENIERÍA
Y SISTEMAS DE TELECOMUNICACIÓN**

**CONTRIBUTION TO THE ANALYSIS OF THE
PERFORMANCE OF DISTRIBUTED PV SYSTEMS**

PHD THESIS

**Jonathan Leloux
Mining and geology engineer
2019**

UNIVERSIDAD POLITÉCNICA DE MADRID

Instituto de energía solar

**ESCUELA TÉCNICA SUPERIOR
DE INGENIERÍA
Y SISTEMAS DE TELECOMUNICACIÓN**

**CONTRIBUTION TO THE ANALYSIS OF THE
PERFORMANCE OF DISTRIBUTED PV SYSTEMS**

PHD THESIS

Author: Jonathan Leloux
Mining and geology engineer
Advisor: Luis Narvarte Fernandez
Prof., Ph.D, Telecommunication engineer

2019

Tribunal nombrado por el Rector de la Universidad Politécnica de Madrid.

PRESIDENTE:

VOCALES:

SECRETARIO:

SUPLENTE:

Realizado el acto de defensa y lectura de la Tesis en Madrid,
el día de 2019.

Calificación:

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

To my mother,

Je t'aime maman

ACKNOWLEDGEMENTS

In March 2005, during my last year as an engineering student, I started an internship in the field of photovoltaic (PV) solar energy in a very small company located in Madrid called Fotosolar, and whose main activity was to promote and build large PV plants in Spain. When I graduated some months later, I was offered a job in Fotosolar, which I gladly accepted, thrilled to be finally out of school and to start working “in the real world out there”. These years at Fotosolar were a marvelous adventure, full of lessons that I will never forget and that marked the beginning of my career as a solar energy engineer. Two years later, with some help of the “Spanish solar bubble”, the company had enjoyed impressive growth, and it became part of the international group of EDF Energies Nouvelles. Meanwhile, at the great age of 25 years and owing to this long experience in the field of solar energy, my business card now stated “Technology and Engineering Manager”. In reality, during the first years of my career at Fotosolar, most of the time, I found myself in uncharted territory. My best friends would then often be Google, books, forums, and articles. I was not aware of it yet, but I would later look back at that period as the beginning of my PhD. Among these books and articles, I particularly liked the ones that were written by an author whose name was Eduardo Lorenzo, professor at Instituto de Energía Solar – Universidad Politécnica de Madrid (IES-UPM). I could never have enough of these articles and books. One day I happened to read his new article published in *Era Solar* and where he explained that IES-UPM had developed a novel method for assessing the productivity of PV plants in the context of their bankability. This is exactly what I was looking for at Fotosolar because we had recently completed the construction of several PV plants that we needed to sell. Therefore, I decided to contact Eduardo, and he kindly received all the members of the engineering department of Fotosolar at his office, and he offered us a visit to the research facilities of the institute. Some months later, the PV System Group of IES-UPM carried out the quality control measurements at two PV plants commissioned by Fotosolar in Andalusia. During that process and through Eduardo, I discovered a tiny part of the world of academic research, and I felt increasingly attracted to it. Eduardo became a mentor for the young engineer that I was and the idea of becoming a researcher acquired an increasing place in my mind until in September 2008 I took the bold decision to leave my job and start a PhD at IES-UPM. Some years later I even ended up forming part of that famous PV System Group, directed by that professor who had written these books and articles that had taught me so much, and with whom I was now going to co-author several articles. I have learned much from Eduardo, in the field of solar energy, as well as in so many other aspects of life. I would probably never have started a PhD if I had not met him.

I am thankful to my PhD advisor, prof. Luis Narvarte, and to all the members of the PV System Group. My memories are filled with wonderful moments shared with the team, for which I am thankful to my colleagues Rodrigo Moretón, Francisco Martínez, Aitor Makibar, José Manuel Carrillo, Luis Miguel Carrasco, Javier Muñoz, Oscar Perpiñán, and Isaac Carrêlo. I feel particularly grateful to Alberto Luna, who worked very closely with me when I had the good luck to become the director of his final course engineering project. Our collaboration saw the

birth of the very first version of the fault detection method that is presented in this thesis, and his help and friendship have meant a lot to me.

One of the wonders of the academic world is that it makes it possible to meet with interesting and passionate people around the world who share similar interests. At many occasions, I have deeply felt that famous quote from Newton “Standing on the shoulders of giants”. In particular, I want to express my more sincere gratitude towards Christian Gueymard, a world-class solar radiation scientist, for all the time that he spent over the last ten years guiding me in my journey to learning more about solar radiation. His knowledge and passion for research and solar energy are only matched by his kindness, and he is a wonderful source of inspiration for me.

This PhD was also made possible by people who do not even work in research or in the field of solar energy. In particular, this work would never have been possible without the thousands of altruistic people who generously and freely provided us with the data corresponding to the energy production of their PV system.

This PhD was also deeply influenced by many other persons that are completely foreign to the academic world. During most of my long journey as a PhD researcher, I have kept a parallel activity as a freelance consultant for the solar energy industry, and this double life had a profound influence on my research. My clients have made me aware of many interesting unsolved research problems and possible solutions, and these questions have always remained present in my mind while working on my research.

I also was very fortunate to be given the opportunity to lead WebPV, a spin-off company of IES-UPM, with a core business very close to my research topics, and in particular, related with the commercialization of automatic fault detection services for distributed PV systems. This experience has taught me many very important lessons that will stay with me during the rest of my life. WebPV also allowed me to associate with two business partners, David Trebosc and Jesús Robledo, who came from a world that is completely different from the university and from whom I have learned so much.

I recently realized that I will soon celebrate my fifteenth year working for solar energy. The time has passed so fast that I still cannot fully assimilate that reality. But I hope that this is still only the beginning of a long and passionating journey. The adventure now continues at LuciSun, a company that I recently co-founded and that will provide specialized consulting and simulation services to the solar energy sector. Since the day I started at Fotosolar, I have already witnessed such unbelievable evolutions in this field, and my dream is now to play an active part in the next chapters to be written.

“Be the change that you want to see in the world”.
Mahatma Gandhi.

ABSTRACT

This work aims to constitute a contribution to the field of the analysis of the performance of distributed photovoltaic (PV) systems and to the detection of faults of operation that can hinder their performance. With that objective in view, the work has been articulated around three main lines of research: the analysis of the performance of dozens of thousands of PV systems in Europe, the development of a novel fault detection method applied to PV system fleets, and the generation of solar irradiation data from the PV energy output.

We have based our analysis on the operational data monitored at more than 31,000 PV systems in Europe, with special focus over Belgium and France, and in a lesser extent to the UK and Spain.

For the analysis of the performance of PV systems, three main questions are posed: how much energy do they produce? What level of performance is associated with their production? Which are the key parameters have the most influence on their quality?

The mean Energy Yield of the PV systems located in each of the four reference countries is 1115 kWh/kW_p for France, 898 kWh/kW_p for the UK, 908 kWh/kW_p for Belgium, 1450 kWh/kW_p for the PV plants in Spain mounted on a static structure, and 2127 kWh/kW_p for those mounted on a solar tracker in Spain.

The quality of the PV systems is quantified using the Performance Ratio (PR), and the Performance Index (PI). After a mean exposure time of 2 years, the mean value of Performance Ratio is 78% in Belgium and 76% in France, and the mean Performance Index of the PV systems is 85% in both countries, which implies that the typical real PV system produces 15% less than a very high-quality PV system (or reference PV system) under the same conditions. On average, the real power of the PV modules falls some 5% below their corresponding nominal power announced on the manufacturer's datasheet. However, some modules show a real power more than 15% below the nominal power announced by their manufacturer. A brief analysis by PV module technology has led to relevant observations about two technologies. On the one hand, the PV systems equipped with Heterojunction with Intrinsic Thin layer (HIT) modules show performances higher than average. On the other hand, the systems equipped with Copper Indium Gallium Selenide (CIGS/CIS) modules show a real power that is 16 % lower than their nominal value.

We have shown that the distribution of the yearly integrated PR can be modeled well using a Weibull distribution for PR values ranging from 60% to 90%. This range of values represents the majority of the PV systems, and we suggest that they are representative of the state of the art for PV systems in Europe. The typical PR value for the PV systems installed in Europe before 2015 is 79% and its value for the PV systems installed in the last few years is 81%.

We have observed that the performance of the PV systems tends to increase when the peak power of the PV systems increases. We have found significant performance differences as a function of the inverter manufacturer, and the PV module manufacturer and technology. We have found an improvement of the state of the art, in the form of an increase in performance in the yearly integrated PR of around 3 to 4% over the last seven years, which represents an increase of about 0.5% per year.

This work presents a fault detection procedure applied to distributed PV system fleets, based on a novel performance indicator, designated as Performance to Peers (P2P), that can be constructed on the sole basis of the comparison of the energy production data of several neighboring PV systems. This work explains how to construct this performance indicator and how to use it to carry out automatic fault detections. This fault detection procedure has been developed in the context of the performance analysis carried out on approximately 6,000 PV installations located in Europe, and it is illustrated and discussed through real application cases. The P2P has been shown to be more stable than the Performance Ratio (PR), in particular in the presence of sub-par metadata on the PV systems, and it thus constitutes a more robust performance indicator for fault detection. The main limitations of this novel methodology are discussed, and several future lines of research are suggested.

Considering the dense network of residential photovoltaic (PV) systems implemented in Belgium, this work evaluates the opportunity of deriving solar irradiation data from the electrical energy production registered at PV systems. The study is based on one year (i.e., 2014) of hourly PV power output collected at a representative sample of roughly 1500 residential PV installations. Our results indicate that the accuracy of the derived solar irradiation data depends on a number of factors including the weather conditions, the density of PV systems, and other data sources that can be accessed to complement the PV data. In particular, the computed solar irradiation data degrade as the information about the orientation and tilt angles of the PV generator becomes more inaccurate.

RESUMEN

El objetivo de este trabajo es constituir una contribución al campo del análisis de la productividad energética de los sistemas fotovoltaicos (PV) distribuidos y a la detección de fallos de operación que pueden afectarla. Con ese objetivo en mente, el trabajo se ha articulado en torno a tres líneas principales de investigación: el análisis de la productividad de docenas de miles de sistemas fotovoltaicos en Europa, el desarrollo de un método novedoso de detección de fallos aplicado a las flotas de sistemas fotovoltaicos y la generación de datos de irradiación solar a partir de datos de producción de energía fotovoltaica.

Hemos basado nuestro análisis en los datos operativos monitoreados en más de 31,000 sistemas fotovoltaicos en Europa, con un enfoque especial sobre Bélgica y Francia, y en menor medida el Reino Unido y España.

Para el análisis de la productividad de los sistemas fotovoltaicos, se plantean tres preguntas principales: ¿cuánta energía producen? ¿Qué nivel de productividad está asociado con su producción energética? ¿Cuáles son los parámetros clave que más influyen en su calidad?

El productividad energética promedio de los sistemas fotovoltaicos ubicados en cada uno de los cuatro países de referencia es de 1115 kWh/kW_p para Francia, 898 kWh/kW_p para el Reino Unido, 908 kWh/kW_p para Bélgica, 1450 kWh/kW_p para las plantas fotovoltaicas en España montadas sobre una estructura estática, y 2127 kWh/kW_p para las plantas usando en seguidores solares en España.

La calidad de los sistemas fotovoltaicos se cuantifica utilizando el Performance Ratio (PR) y el Performance Index (PI). Después de un tiempo medio de exposición de 2 años, el PR promedio es del 78% en Bélgica y 76% en Francia, y el PI promedio de los sistemas fotovoltaicos es del 85% en ambos países, lo que implica que el sistema fotovoltaico real típico produce un 15% menos que un sistema fotovoltaico de muy alta calidad (o sistema fotovoltaico de referencia) en las mismas condiciones. En promedio, la potencia real de los módulos fotovoltaicos cae aproximadamente un 5% por debajo de su potencia nominal correspondiente anunciada en la hoja técnica del fabricante. Sin embargo, algunos módulos muestran una potencia real de más del 15% por debajo de la potencia nominal anunciada por su fabricante. Un breve análisis por la tecnología del módulo fotovoltaico ha llevado a observaciones relevantes sobre dos tecnologías. Por un lado, los sistemas fotovoltaicos equipados con módulos de tecnología Heterojunction with Intrinsic Thin layer (HIT) muestran rendimientos superiores a la media. Por otro lado, los sistemas equipados con los módulos de tecnología copper indium gallium selenide (CIGS/CIS) muestran una potencia real que es 16% más baja que el valor nominal.

Hemos mostrado que la distribución del PR anual se puede modelar de manera satisfactoria utilizando una distribución de Weibull para valores de PR entre el 60% y el 90%. Este rango de valores representa la mayoría de los sistemas fotovoltaicos, y sugerimos que es

representativo del estado del arte de los sistemas fotovoltaicos en Europa. El valor de PR típico para los sistemas fotovoltaicos instalados en Europa antes del 2015 es del 79% y su valor para los sistemas fotovoltaicos instalados en los últimos años es del 81%.

Hemos observado que la productividad de los sistemas fotovoltaicos tiende a aumentar cuando aumenta la potencia pico de estos sistemas. Hemos encontrado diferencias de productividad significativas en función del fabricante del inversor y de tanto el fabricante como la tecnología del módulo fotovoltaico. Hemos encontrado una tendencia hacia una mejora estado del arte con el tiempo, en la forma de un aumento en el rendimiento en el PR anual de alrededor de 3 a 4% en los últimos siete años, lo que representa un aumento de aproximadamente 0,5% por año.

Este trabajo presenta un procedimiento de detección de fallos aplicado a las flotas de sistemas fotovoltaicos distribuidos, basado en un indicador de productividad novedoso, designado como Performance to Peers (P2P), que puede construirse basándose únicamente en la comparación de los datos de producción de energía de varias instalaciones fotovoltaicas vecinas. Este trabajo explica cómo construir este indicador de productividad y cómo usarlo para realizar detecciones automáticas de fallos. Este procedimiento de detección de fallos se ha desarrollado en el contexto del análisis de la productividad energética aplicada a aproximadamente 6000 instalaciones fotovoltaicas ubicadas en Europa, y se ilustra y discute a través de casos de aplicaciones reales. Se ha demostrado que el P2P es más estable que el PR, en particular en presencia de imperfecciones en los metadatos sobre instalaciones fotovoltaicas, y por lo tanto constituye un indicador de productividad más robusto para la detección de fallos. Se discuten las principales limitaciones de esta nueva metodología y se sugieren varias líneas de investigación futuras.

Aprovechando la densa red de sistemas fotovoltaicos residenciales instalados en Bélgica, este trabajo estudia la oportunidad de obtener datos de irradiación solar a partir de la producción de energía eléctrica registrada en los sistemas fotovoltaicos. El estudio se basa en un año de datos (en concreto, 2014) de producción de energía fotovoltaica registrada con paso horario en una muestra representativa de aproximadamente 1500 instalaciones fotovoltaicas residenciales. Nuestros resultados indican que la precisión de los datos de irradiación solar derivados depende de una serie de factores que incluyen las condiciones climáticas, la densidad de los sistemas fotovoltaicos, y otras fuentes de datos a las que se puede acceder para complementar los datos de producción fotovoltaica. En particular, los datos calculados de irradiación solar se degradan a medida que la información sobre los ángulos de orientación e inclinación del generador fotovoltaico se vuelve más inexacta.

INDEX

1. INTRODUCTION	1
2. REVIEW OF THE PERFORMANCE OF PV SYSTEMS	7
2.1. Methodology	9
2.2. Belgium	15
2.2.1. Belgian residential PV market	15
2.2.2. Results	17
2.3. France	26
2.3.1. French residential PV market	26
2.3.2. Results	28
2.4. Europe	38
2.4.1. Results	38
3. FAULT DETECTION APPLIED TO PV SYSTEM FLEETS: PERFORMANCE TO PEERS (P2P)	53
3.1. Methodology overview	53
3.2. Looking for the best peers	55
3.3. Constructing the P2P	60
3.4. Applying the P2P to fault detection	62
3.5. Fault diagnosis	68
3.6. Application of P2P to PV system fleets	70
4. SOLAR IRRADIATION FROM PV ENERGY PRODUCTION	74
4.1. Methodology	74
4.2. Results	83
5. CONCLUSION	90
CONTRIBUTIONS OF THIS PHD THESIS	94
REFERENCES	98

1. INTRODUCTION

At the end of the year 2018, the total installed photovoltaic (PV) capacity in the world has surpassed the 500 GW (IEA PVPS, 2019), distributed over more than 50 million installations. Europe represented slightly more than 20% of the total installed capacity, with 114 GW installed, among which 8 GW were installed during the year in 2018. The leading country in Europe in terms of installed capacity is still, by far, Germany, with more than 45 GW, representing about 40% of the total European capacity. Most of the other countries also steadily keep adding PV power to their grid.

The performance of the PV installations has increased over time (Oozeki et al., 2010; Reich et al., 2012; Van Sark et al., 2012), owing to technological advances and system optimizations. Over the last decades, numerous monitoring campaigns have allowed to report and analyse the performance of thousands of PV systems worldwide (Anis and Nour, 1995; Baltus et al., 1997; Blaesser, 1997; Decker and Jahn, 1997; Erge et al., 2001; Pelland and Abboud, 2008; Ueda et al., 2009; Eltawil and Zhao, 2010; Hussin et al., 2012; Leloux et al., 2012a, 2012b; Wittkopf et al., 2012; Díez-Mediavilla et al., 2013; Woyte et al., 2013; Bonsignore et al., 2014; Colantuono et al., 2014; Leloux et al., 2015; Myong et al., 2015; Taylor et al., 2015a; Tsafarakis and Van Sark, 2015; Kausika et al., 2018; Seme et al., 2019). These studies have shown a wide disparity in performance between the PV systems, with Performance Ratio (*PR*) values typically comprised between 60% and 90%. Some of the causes explaining the performance losses have been identified and quantified, but more remains to be done.

Thanks to the development of Information and Telecommunication (IT) technologies, these last years have witnessed an increase in the quantity and in the quality of the operating data that have been monitored, offering new possibilities to detect performance problems. Automatic fault detection procedures have experienced major developments in the field of engineering (Isermann, 1997), and their application to PV systems has undergone intense development, and it is now a trending Research and Development (R&D) topic (Dালiento et al., 2017; Madeti and Singh, 2017).

A fault is understood as a decrease in a performance indicator, generally abrupt, during a specific time period, and due to abnormal functioning. According to this definition, in certain cases, the overall performance of a PV system can be relatively low without the presence of any fault. For example, this can be the case of an old PV system whose generator nameplate power has decreased over time because of the aging of the PV modules (Jordan et al., 2016). This can also happen when the real nameplate power of the PV modules has been overrated, and it is lower than the expected nameplate power, according to datasheets (Martinez-Moreno et al., 2010). These overall performance problems can be significant, but they are out of the scope of fault detection procedures strictly speaking.

The *PR* is the most commonly used performance indicator (Decker and Jahn, 1997; Jahn, 2004; Eltawil and Zhao, 2010; Reich, 2012; Van Sark, 2012; Chine, 2014; Jordan, 2014; Woyte et

al., 2014; Essah et al., 2015; Brecl and Topič, 2016; Dhimish and Holmes, 2016). It is based on the ratio between the energy output of a PV system and the solar irradiation that it receives. Several alternative fault detection procedures have been developed. The PR can be normalized in order to become more independent of the operating conditions (Herteleer et al., 2017). Another method relies on comparisons between measured and simulated energy yields (Perez et al., 2001; Stettler et al., 2006; Drews et al., 2008; Gaiddon et al., 2010; Huld et al., 2010; Phinikarides et al., 2016). Some procedures rely on the modeling of I-V curves (Charles et al., 2004; Forero et al., 2006; Chao et al., 2008; Gokmen et al., 2013; Hu et al., 2013; Spataru et al., 2015), and others make use of abnormal variations in power, voltage or current as a fault indicator (Sopitpan et al., 2001; Takashima et al., 2009; Chouder and Silvestre, 2010; Gokmen et al., 2012; Silvestre et al., 2013; Solórzano and Egido, 2013; Spertino and Corona, 2013; Silvestre et al., 2014; Tadj et al., 2014; Silvestre et al., 2016). Other methods make use of scatter plots, also called “stamp collections”, representing the outputs of the PV system (energy, AC or DC power, voltage, current,...) as a function of key input variables (solar irradiance, air temperature, cell temperature, solar elevation,...). These methods have shown promising capacities (Drews et al., 2007; Firth et al., 2010; Chouder et al., 2012; Chouder et al., 2013; Firman et al., 2014; Stein et al., 2014; Stein, 2015; Elkholy et al., 2016; Hachana et al., 2016; Hosseinzadeh and Salmasi, 2016; Killinger et al., 2016a). Another family of approaches relies on Artificial Neural Networks (ANN), fuzzy logic, or machine learning (Yagi et al., 2003; Mellit and Kalogirou, 2008; Mellit and Pavan, 2010; Almonacid et al., 2011; Riley and Johnson, 2012; Spiliotis et al., 2018).

In relation to their monitoring, PV systems can be classified into two main groups. The first group is constituted by large solar power plants, whose nameplate power stands between several hundreds of kW and several hundreds of MW. The second group is made of smaller installations, whose nameplate power typically lies between 1 kW and the hundreds of kW, generally mounted on the roof of a building, called distributed PV systems. The Operation and Maintenance (O&M) of PV plants benefits from a specific budget, and it is commonly assisted by Supervisory Control And Data Acquisition (SCADA) that registers sizeable quantities of data that are sometimes measured up to the PV module string level (Vargas et al., 2015; Garoudja et al., 2017), and a weather station that measures the operating conditions. Other tools and hardware specific to the O&M are also available, and they are used by trained professionals (Martínez-Moreno et al., 2012; Tyutyundzhiev et al., 2015).

Distributed PV installations consist of a large number of small PV generation units spread out over the territory. In Europe, these distributed PV installations represent about 40% of the total installed PV capacity and more than 90% of the total number of PV installations. They differ from one another in their size, components, orientation and tilt angles, topology, and quality. They operate under weather conditions that vary from one installation to another, both in space and in time. Each one of them usually belongs to a different owner. Most of the time its operator is the owner, who has no expertise in IT or in photovoltaics (Mukai et al., 2011). The metadata on the technical characteristics of the PV system, such as its peak power, orientation, PV module technology,... are often reported with limited accuracy, or even unknown (Kiefer et al., 2011). The budget allocated to the monitoring of the PV system is small, which makes cost-

prohibitive the installation of monitoring hardware, although low-cost solutions exist (Fuentes et al., 2014). Therefore, the energy output is often the only accurate information available. There is no on-site measurement of the operating conditions (Vergura, 2018a), which must be acquired through a remote data provider.

Unfortunately, most of the existing fault detection procedures rely on accurate input data on the PV system's characteristics and its operating conditions, and they have been developed and validated in large PV plants, or PV systems associated with research laboratories, whose technical characteristics and operating conditions were known with a high degree of accuracy. Therefore, they suffer from several limitations when the time comes to apply them to large groups of hundreds or thousands of PV installations, whose main technical characteristics and operating conditions are not known with accuracy. First, the inaccuracies in the metadata of the PV system, in particular, its peak power and its orientation, make it difficult to compare the energy production with the in-plane solar irradiation. Second, it is difficult to obtain reliable and low-cost solar irradiation data. The data measured by pyranometers are not available at a spatial resolution that is high enough, and the data retrieved from satellites convey inaccuracies that are due to their limited spatial and temporal resolution, cloudiness, snow cover, aerosols, and low solar elevation angles. Third, thermal losses affecting the PV systems depend on the weather conditions, notably including air temperature and wind speed. In fact, due to the uncertainties on solar irradiation at the hourly level, the *PR* can often go up to more than 100%, and it can show large variations. Under these conditions, it is difficult to build a reliable performance indicator that is stable enough to make it possible for accurate fault detections. As a result, many problems affecting distributed PV systems go undetected. As an example that is representative of the PV systems that were analyzed in this work, Figure 1 shows the daily-integrated *PR* of a residential PV system located in Belgium, and that has been affected by several important performance problems between July and October 2015 related to its inverters. The evolution of its *PR* shows important hints on some of the days when the problems occurred. Nevertheless, it is difficult to draw a clear distinction between correct operation and faults.

Figure 1: Daily-integrated PR of a residential PV system located in Belgium, and that has been affected by several important performance problems between July and October 2015.

Several independent research teams have undertaken the task of analyzing the monitoring data from several hundred or thousands of PV systems. Their findings show that they faced common problems and reached similar conclusions on the inherent limitations from the lack of reliable input data and the need to cope with that reality (Kiefer et al., 2011; Saint-Drenan et al., 2015; Koubli et al., 2016; Killinger et al., 2017, 2018; Bertrand et al., 2018). Several works have developed procedures that make it possible to correct for the inaccurate metadata, along with Quality Controls (QC) procedures and data inferences, making it possible to obtain more robust fault detections (Saint-Drenan et al., 2015; Koubli, 2016; Killinger, 2017). In parallel, other works have applied statistical techniques coming from other fields, such as data mining (Capozzoli et al., 2015) or kriging (Chiles et al., 2012), to construct a fault detector from the comparison of the performance of a group of PV systems, generally designed as PV system fleet (Golnas et al., 2011; Lonij et al., 2012; Elsinga and Van Sark, 2014; Engerer et al., 2014; Leloux et al., 2014; Popovic and Radovanovic, 2018). These studies have shown promising results, in particular in the presence of clear-sky conditions, and when the data come from several PV systems located in the same neighborhood and whose generators have similar orientations.

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

Over the last decades, satellite-based retrieval of solar radiation at ground level has proven to be valuable for delivering a global coverage of the global solar irradiance distribution at the Earth's surface (Perez et al., 1994, 1997; Hammer et al., 2003; Renné et al., 2008). The recent deployment of PV systems offers an opportunity to extend the in-situ characterization of the incoming solar radiation at the Earth's surface through the reverse engineering conversion of the energy production of the PV systems to global solar irradiation data.

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

The objectives of this work aim to constitute a further contribution in these fields of research. In particular, several specific topics are tackled:

- The assessment of the overall performance of the distributed PV systems in Europe;
- The development of a novel fault detection method applied to fleets of distributed PV systems;
- The development of a method makes it possible to deduce solar irradiation that is received on the plane of distributed PV systems from the measurement of their PV energy output.

Chapter 2 presents the assessment of the performance of the distributed PV systems which is carried out through the analysis of the monitoring data measured at more than 31,000 PV systems in Europe, mainly in France, Belgium and the UK. The results make it possible to extract conclusions about the expected energy production of PV residential systems, about the energy production losses due to their orientation and about state of the art and the quality of residential PV systems. The important quantity of PV systems analyzed makes it possible to extend the results to the European market and to PV installations worldwide, and hence they are of general interest. In fact, the conclusions are congruent with previous analyses of the operational performance of residential PV systems installed during the last two decades in Germany, Switzerland, Italy, Spain, Netherlands, Japan and USA (Wiles et al, 2002; Jahn and Nasse, 2004; Oozeki et al, 2010; Furkan, 2011) and can be useful for important works that are presently ongoing (IEA PVPS Task 13) and whose main purpose is the assessment of the performance and reliability of PV systems.

Chapter 3 presents a fault detection procedure applied to distributed PV system fleets, based on a novel performance indicator that can be constructed on the sole basis of the comparison of the energy production data of several neighboring PV systems. The resulting performance indicator has been designated as *Performance to Peers (P2P)* because it is based on comparisons between neighboring and similar, i.e., *peer* PV systems. This procedure does not require any specific monitoring hardware or the input of any operating conditions data, and it only requires the energy production data, which can be measured by the inverter or by the energy meter. The text explains how to construct this performance indicator and how to use it to carry out automatic fault detections. The procedure is applied to more than 6,000 PV installations and is illustrated and discussed through real application cases.

Chapter 4 presents a novel reverse engineering method that makes it possible to retrieve the solar irradiation received in the plane of a PV system from the PV energy output measured at this PV system by its energy meter. These developments were based on one year (i.e., 2014) of hourly PV power output data collected at 2893 PV systems in Belgium installed from 2008 to 2013. The solar irradiation data that were generated using this method were compared to site-measured data measured from several ground weather stations equipped with second class pyranometers in Belgium.

The conclusion section summarizes the major findings of this work and then discusses the main difficulties and limitations that have been encountered and suggests several possible paths for future research lines.

2. REVIEW OF THE PERFORMANCE OF PV SYSTEMS

This chapter presents a review of the state of the art of residential PV systems in Europe, with special focus on the cases of Belgium, France, and the UK, through the analysis of the operational data of a representative sample of 31,000 PV installations (see Figure 2).

Figure 2: Location of the 31,000+ PV systems in our database. Most of them are located in the UK, Belgium, and France.

The geographical distribution and type of these data are summarized in Table 1.

Table 1: Geographical distribution and type of PV systems in our database. Most of the PV systems in Spain are large PV plants. 300 PV systems are located in other countries from Europe.

Country	# systems	Pp [MWp]	Type
France	17672	65	Rooftop
Belgium	7648	50	Rooftop
UK	5835	23	Rooftop
Spain	29	116	PV plant
Others	307	3	Rooftop
TOTAL	31491	255	

The content of this chapter originated in several previous publications that dealt with these three countries separately and that have been gathered here into one single chapter (Leloux et

al., 2011, Leloux et al., 2012a, 2012b; Leloux et al., 2015; Taylor et al., 2014; Taylor et al., 2015). The reason for the choice of these countries was mainly driven by the availability of the data at the moment of the start of this research line. These studies were published some years ago, and the market has evolved since then, but the results are still representative of the state of the art of the PV installations that are currently in operation in these countries rather than the newest ones that were installed or to be installed in the near future. Nevertheless, the analyses carried out on a country basis do not include all the installations present in the database, because in particular in the case of Belgium and France all these data were not available at the time of the study. Therefore, the exact number of installations that were analyzed in each case is mentioned more specifically in the corresponding sections.

For Belgium, 993 PV installations were analyzed, totalizing a peak power of approximately 4MW, and installed between 2007 and 2010. Although Belgium is composed of three regions, the data come from Wallonia and Brussels and not from Flanders, due to availability reasons. Nevertheless, the data is still representative of the state of the art, since typologies are very similar. Flanders has developed towards both residential PV and solar plants. Since the end of 2007, Wallonia and Brussels established a supporting scheme (consisting of a mix of subventions and production-based support, called “green certificates”) to promote residential PV. The PV power connected to the grid consequently jumped from 200 kW in 2007 to 50 MW at the end of 2009 (see <https://www.apere.org/>). About half of the total power is installed in residential PV systems of less than 10 kW_p.

For France, the operational data of a representative sample of 6868 installations were analyzed, totalizing a peak power of approximately 20.5 MW, and installed between 2007 and 2010. Most of these installations are residential PV systems whose peak power is 3 kW_p or less.

For the UK, all the PV systems of the database were analyzed, totaling 5835 PV systems that represent about 23 MW_p.

About 300 PV systems are located in other countries from Europe. Among the other countries that were analyzed, Spain has few PV systems, but they are important ones because most of them are large PV plants, which sum up a larger total peak power than the rooftop installations from any other country. This study is therefore mostly based on the analysis of the data of 4 reference countries: France, UK, Belgium (mostly small rooftop PV systems) and Spain (mostly large PV plants). The diversity in climates and installation types among these 4 countries makes it possible to reach a representative picture of the state of the art in PV systems in Europe. The work focuses the analysis on the following three questions:

- a) How much electricity do PV systems produce in terms of kWh per installed kW_p?
- b) How good is this electrical production? The PV systems quality is analyzed using different performance indicators such as the Performance Ratio (PR), the Performance Ratio at STC conditions (PR_{STC}), and the Performance Index (PI)?
- c) Which are the key aspects that influence the quality of PV systems? Statistical tools are applied to find them out.

For the first question, we have used the monthly energy production data supplied by the PV systems' owners through several websites.

For the second question, the quality is evaluated using the aforementioned performance indicators, all of them consisting of comparing the actual production of each of the systems with the simulated electrical production of a corresponding hypothetical system used as a reference.

For the third question, we have investigated the main causes of the quality differences that are observed by applying an Analysis-of-Variance (ANOVA) to the *PI*. A general multidimensional ANOVA is realized by grouping the PV systems according to four characteristics: PV modules manufacturer, inverters manufacturer, installer, and PV generator power. The goal is to isolate the causes explaining the *PI* differences.

2.1. Methodology

Most of the data for the UK has been provided by Sheffield Solar and was acquired through the Microgen Database (MgDB) website (<http://www.microgen-database.org.uk/>). Most of the data for France and Belgium were provided by BDPV through its free public website for PV system owners (<https://www.bdpv.fr/>), and Rtone through its commercial monitoring service Rbee Solar (<http://www.rbeesolar.com>). The data from the PV plants in Spain have been provided by several PV plant operators through Instituto de Energía Solar - Universidad Politécnica de Madrid (IES-UPM). The data for the rest of Europe have been provided by Rtone and BDPV.

The data that were used as input for these analyses can be divided into three main categories: PV systems characteristics, PV energy production, and solar irradiation.

Some of the PV system characteristics were available for most of the PV systems: latitude, longitude, azimuth and tilt angles, and peak power. Some other characteristics were available, depending on the data provider: PV module manufacturer and model, inverter manufacturer and model, installer, year of installation, PV cell/module technology.

The PV energy production data provided by MgDB were collected via the MgDB website, with PV owners using the site as a portal to upload readings and in return receiving free monthly Performance Ratio (PR) analysis and peer-to-peer performance checking in the form of interactive maps and nearest neighbor comparisons. The majority of data is measured by the energy meters of the inverters and is collected from commercial data donors who own/monitor hundreds of systems using automated data transfers. PR calculations interact directly with the MgDB so as to provide regular updates to the live website.

The PV energy production data provided by Rtone have been monitored using the Rbee Solar monitoring product, which measures the energy production with a smart energy meter at a 10-min time interval. These installations have been installed from 2008 to 2014.

The PV energy production data provided by BDPV have been collected through a free, public website for PV system owners, where individuals can manually enter the monthly reading of their energy meter or inverter through an HTML web interface, or if their equipment allows for it, make use of an Application Program Interface (API) to automatically upload their data, at any time resolution.

The PV energy production data provided by the PV plant operators have been provided to IES-UPM in several different formats, mostly in the form of text/CSV files. The oldest of these PV plants was commissioned in 2005, but most of them date from the years 2006-2010.

Figure 3 represents the relative distribution of the azimuth and tilt angles of all the PV systems on a heatmap. The numbers are given in the format of a percentage. Most of the installations have an azimuth between South-East (-45°) and South-West (45°) and with a tilt between 20° and 50° , which correspond to the orientations for which the annual yearly yield is the highest.

		Azimuth deviation from South [°]																			
		(<--East)									(West-->)										
Tilt Angle [°]		-90	-80	-70	-60	-50	-40	-30	-20	-10	0	10	20	30	40	50	60	70	80	90	
	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	10	0.2	0.1	0.0	0.1	0.1	0.4	0.2	0.2	0.2	0.6	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	
	20	0.9	0.4	0.5	0.7	0.5	2.3	1.2	1.9	2.3	6.8	2.0	1.8	1.3	1.2	2.0	0.5	0.4	0.3	0.6	
	30	0.5	0.3	0.4	0.6	0.3	2.0	1.0	2.0	2.2	6.4	1.8	1.5	1.0	1.2	1.8	0.3	0.2	0.2	0.5	
	40	0.9	0.4	0.7	0.7	0.5	3.2	1.5	2.2	2.5	7.6	1.9	1.9	1.2	1.5	3.5	0.3	0.3	0.2	0.4	
	50	0.3	0.1	0.2	0.3	0.4	1.7	0.6	0.8	1.1	2.9	1.1	1.0	0.7	0.5	1.0	0.4	0.2	0.1	0.3	
	60	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	70	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	80	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	90	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

Figure 3: Relative distribution of the azimuth and tilt angles of all the PV systems on a heatmap. The figures are given as a percentage. Most of the installations have an azimuth between South-East (-45°) and South-West (45°) and with a tilt between 20° and 50° , which correspond to the orientations for which the annual yearly yield is the highest.

Figure 4 shows the frequency of the peak power of the PV systems in Europe. Most of them have a peak power between 1 and 6 kWp. A pronounced peak in frequency is visible around 3 kWp. This peak is mostly explained by more than 50% of installations in our database coming from France (see the section on France).

Figure 4: Frequency of the peak power of the PV systems in Europe. There is a pronounced peak around 3 kW_p. This is a direct result of the public incentive mechanism set up by the French government such that it is uneconomical to install a PV system with a peak power higher than 3 kW_p. As a consequence, most of the PV systems in France have a peak power of exactly 3kW_p or slightly lower.

The information reported by the data providers is not always very accurate. We have observed these inaccuracies in the information on some key parameters such as the azimuth angle, the tilt angle, or the peak power. We have also observed inaccuracies and inconsistencies in the energy production data and in the solar irradiation data. This leads to some important uncertainties in our results, although we have tried to minimize their impact, by applying filtering procedures, and by using analyses methods that are robust against erroneous data and outliers. As an example, Figure 5 shows the tilt angle information for all the PV systems in our database as reported by the data providers. The systematic peaks around values that are multiples of 5° are an indicator that most of the data have been rounded. Moreover, many tilt angles were reported as 0°, but they do not appear on the figure, because they corresponded to PV system orientations reported as 0° in tilt and 0° in azimuth, which for some data providers is a convention used as default when they do not know that information. As very few PV generators are installed completely horizontally, we have removed all of these data. It would, in any case, be bad practice to install the PV modules horizontally, because they would tend to accumulate soiling, and their energy yield would be greatly affected.

Tilt angle of PV generators as reported by data providers - Europe

Figure 5: Tilt angle of all the PV systems in Europe as reported by the data providers. The systematic peaks around values that are multiples of 5° are an indicator that most of the data have been rounded.

We have filtered out all the monthly integrated PR data that were higher than 1.1, because these values are very unlikely, and they are the consequence of uncertainties on some of the input data. On the other hand, we chose not to be too restrictive in the filtering operations in order to avoid skewing the results by eliminating some points that would look like outsiders but are due to other causes that are physically sound. We have filtered out all the yearly integrated PR data every time that at least one of the months of this year for a given installation had been filtered out.

We have analyzed the annual energy yield of the PV systems, and we have calculated and represented its mean value and dispersion for all the PV systems in Europe and for four of the reference countries.

The energy performance indicators that are used to assess the technical quality of a particular PV system are obtained by comparing its actual production along a certain period of time with the production of a hypothetical reference system (of the same nominal power, installed at the same location, and oriented the same way) somewhat free of certain kinds of losses. Table 2 presents three different performance indicators that are used to assess the quality of a PV system. All three compare the real energy production of the PV system during a certain period of time to the corresponding reference system. The variation between them comes from the different reference system that is chosen in each case.

Table 2: The quality of a PV system for producing energy can be described through three different performance indicators: PR , PR_{STC} , or PI .

Indicator	Definition	Reference system
performance ratio	$PR = \frac{E_{produced}}{P^{STC} \frac{\int G dt}{G^{STC}}}$ <p>where $E_{produced}$ is the net electrical energy produced by the PV system during a given period of time, P^{STC} is the rated power of the PV generator under STC, G^{STC} is the global solar irradiance under STC (i.e., 1000 W/m²), and G is the global solar irradiance received by the PV generator.</p>	Free of any kind of system losses. Its solar cells are always kept at 25° C.
performance ratio at STC	$PR_{STC} = \frac{E_{produced}}{P^{STC} \frac{\int G(1 - \Delta P_{STC}) dt}{G^{STC}}}$ <p>where ΔP_{STC} represents the thermal power losses in the PV cells due to their operational temperature which is different from STC.</p>	Free of any kind of system losses. Its solar cells operate at the same temperature that the ones of the system to be compared with.
performance index	$PI = \frac{E_{produced}}{P^{STC} \frac{\int G(1 - \Delta P_{STC})(1 - \Delta P_{DC/AC}) dt}{G^{STC}}}$ <p>where $\Delta P_{DC/AC}$ represents the conversion losses due to the inverter that equips the reference system.</p>	High-quality PV system. Almost free of system losses, except mainly the cell's temperature losses and the DC/AC losses corresponding to a very good inverter and considered somewhat unavoidable.

The PR is, by far, the most widely used performance indicator today, because the unitary energy production, which is of paramount importance for economic analyses, is simply given by the product of the irradiance, (or the number of “sun-hours”) by the PR. The difference between 1 and PR lumps together all imaginable energy losses (real power of the PV modules power below nominal rating, mismatch, wiring, shades, dust, thermal, DC/AC, failures, etc.). Because thermal losses are site-dependent (they depend on climate), the PR of a given, unchanged PV system fluctuates from one place to another, and along the course of a year or a day, which represents an obvious inconvenient for strictly qualifying its technical quality. The PR of a given PV system located in the North of France is higher than the PR of the same system installed in the South of the country. The PR_{STC} takes away such thermal losses, which requires to consider (measure or estimate) the temperature of operation of the solar cells. Because of

that, it is of more complex calculation than the PR, but it becomes practically independent from time and site, thus being more appropriate for strictly qualifying technical quality on a comparative basis. However, the PR_{STC} value corresponding to excellent quality and properly maintained PV system is lower than 1, mainly because real inverters always associate some energy losses to the DC/AC conversion. Hence, one further step can still be taken subtracting the DC/AC conversion losses corresponding to a top-class inverter, let us say, one whose European efficiency is 96%. That leads to the so-called PI (Behnke et al., 2005). It should be noted that a $PI = 1$ corresponds to a PV system composed by an inverter and a PV generator whose real power and characteristics coincide with their rated nominal value, free of shading, dust and wiring losses and also free of failures. Consequently, the difference between 1 and PI can be understood as a measure of the somewhat avoidable energy losses. The *PI* thus allows for a direct comparison of the quality of PV systems under different climatic and installation conditions. Because of that, this work pays particular attention to the analysis of *PI* values.

We have analyzed all the monthly integrated PR of all the PV systems, and for all the months for which the data were considered to be correct after filtering on a month-by-month basis. We have also analyzed the yearly integrated PR of all the PV systems, and for all the years for which there were 12 consistent monthly data for the PV system. We have applied yearly integrated PI analyses to the cases of the PV installations located in Belgium and in France.

The estimation of the energy production of the reference system is simulated with a tool developed at IES-UPM and based on widely-accepted models, whose details have been described elsewhere (Schmid and Schmidt, 1991; Hammond et al., 1997; Christian et al., 2002; Marion, 2002; Martin and Ruiz, 2006; Skoplaki and Palyvos, 2009; Simon and Meyer, 2011). The simulations require the input of the horizontal solar radiation and ambient temperature data, both on a monthly basis. For the analysis of the installations in Belgium and France, the data GHI data were obtained from SoDa (<http://www.soda-pro.com/>), and the air temperature data were obtained from PVGIS (https://re.jrc.ec.europa.eu/pvg_tools/en/tools.html). For the rest of the PV installations located in other countries, the data have been provided by the solar irradiation data provider Synermet and were acquired from CM SAF (<https://www.cmsaf.eu>). The Global Tilted Irradiation (GTI) received on the surface of each of the PV generators is estimated monthly using widely accepted decomposition and translation models (Liu and Jordan, 1960; Erbs et al., 1982; Hay, 2003).

To investigate more in-depth the main causes of the quality differences among the PV systems, they have been compared by grouping them by common properties. The statistical tool Analysis-of-Variance (ANOVA) has been used to study the causes of the dispersion of PI. ANOVA procedures rely on a distribution called the F-distribution. The key statistic is $F = MSTR/MSE$, where MSTR (Mean Square Treatment) represents the variation among the means of the different groups, and MSE (Mean Square Error) represents the variation within the groups. Large values of F indicate that the variation among the groups is large relative to the variation within the groups, and hence that the groups are significantly different. A general multidimensional ANOVA was first realized according to four criteria: PV modules manufacturer, inverters manufacturer, installer, and PV system peak-power. We have used

histograms and box plots to visualize the data, and we have used the ANOVA procedure to assess the statistical soundness of our results for the comparisons of the PV systems' performance as a function of several key parameters such as the inverter manufacturer, the PV module manufacturer, the PV module technology, the year of installation, or the country where it is located. This makes it possible to test whether the differences observed between different populations are significant enough relative to the overall dispersion among the populations that is due to uncertainties and/or the other parameters. We have used multi-ANOVA procedures in order to analyze the effects of one parameter on performance while normalizing as much as possible for the other relevant effects. For example, when analyzing whether there is a statistically significant relationship between a PV module manufacturer and the performance of the PV systems equipped with its PV modules, we have applied a multi-ANOVA to verify that the results obtained regarding the PV modules were not due to systematic associations between one PV module and one inverter, where one poorly performing inverter model or manufacturer could negatively influence the results for the PV modules associated with it.

2.2. Belgium

2.2.1. Belgian residential PV market

At the end of the year 2018, the total PV power installed in Belgium amounted to more than 4.2 GW_p, according to data collected by local renewable energy association APERe (<http://www.apere.org/>), which has combined the figures released by the country's three energy regulators Brugel (<https://www.brugel.brussels/>), VREG (<http://www.vreg.be/>) and CWaPE (<http://www.cwape.be/>). By contrast to the multi-megawatt PV plants installed in the South of Europe, residential and commercial rooftop PV systems with less than 10 kW capacity represent the majority of the installed capacity in Belgium.

In Wallonia, there is a fairly stable relation between installed PV power per province, the number of inhabitants and the net annual income per capita. However, the Region of Brussels-Capital, with a population nearly equivalent to one-third of the Walloon population, owns less than 5% of the PV power installed among the two regions. The urbanization style very likely explains this marked difference. In Wallonia, an important proportion of the population lives in a house that offers the surface necessary to install 1-5 kW_p of PV modules, while in Brussels-Capital people generally live in urban buildings with no individual roof where to install solar panels.

Figure 6 shows that residential PV systems of less than 10 kW account for 95% of the total installed PV power. The power of nearly two-thirds of the PV systems is lower than 5 kW. That range arises as a consequence of limiting the most interesting public financial support to systems of several kW, and from the surface typically available on the roofs. The “*green certificates*” are also limited in relation to the electricity consumption of the household, which in Belgium typically lies between 3000 and 4000 kWh/year. The feed-in tariff set-up by the government establishes that the owner of a PV system can sell as much PV-generated electricity as the amount of electricity that he/she consumes per year (i.e., this is an annual net-balance

mechanism). The distribution of the peak power shows that with most of the PV systems have a power between 1 kW_p and 10 kW_p, and it roughly follows a lognormal law, which arises as the product of several independent positive random variables, the most relevant being the annual energy consumption of the owner, the surface available on the roof to install PV modules, and the capital that the owner is willing/able to invest in a PV system. The feed-in tariff in Belgium was so attractive that numerous households have also found it economical to voluntarily increase their annual electricity consumption in order to be able to sell more PV electricity. They often did so by switching to electricity for water heating, space heating, or cooking.

Figure 6: Frequency of the peak power of the PV systems in Belgium. The distribution of the peak power shows a wider dispersion, with most of the PV systems between 1 kW_p and 10 kW_p, and it roughly follows a lognormal law.

PV modules based on classical crystalline Silicon (xSi) technology represent more than 90% of the total market shares. The rest of the market is distributed, by order of importance, among Copper Indium (di)Selenide (CIS) (3%), Heterojunction with Intrinsic Thin layer (HIT) (3%), and amorphous Silicon (aSi) (1%).

The database presents the number of PV market actors as follows: 85 PV modules manufacturers, 25 inverters manufacturers, and 210 PV systems installers. Figure 7 shows the relative market share of these inverter manufacturers. The relative market penetration within each of those three categories of actors is satisfactorily modeled by a power-law, indicating that the market is dominated by a reduced number of actors. The most extreme case of market

domination is the control by one inverter manufacturer alone of more than 50% of the market (the point corresponding to that inverter manufacturer is not represented because it is out of scale, but is taken into account in the power-law equation).

Figure 7: Relative portion of the PV industry outstripping a given market share. The industry is presented through the manufacturers of PV modules and inverters, and the installers. The relative market penetration within each indicator is satisfactorily modeled by a power-law. The R-squares of the fits yield 99.8% for the installers, 99.1% for the PV modules, and 96.9% for the inverters.

2.2.2. Results

A) Energy production

The energy production analysis is carried out for the year 2009 and for the 158 PV systems from which the monthly production was reported for the 12 months of the year. On average, the PV systems produced in 2009 a net annual energy of 902 kWh/kW_p.

In order to compare this production with other previous studies in the literature, we adjusted the production in 2009 by the ratio of the solar radiation received in 2009 and the mean solar radiation of the last decade according to SoDa database, resulting in a value of 836 kWh/kW_p. Figure 8 shows a histogram of the corrected productions. As a comparison, annual productions around 800 kWh/kW_p were reported for PV systems installed 5-10 years earlier in the North and East of Germany (Jahn and Nasse, 2004).

Figure 8: Histogram of the production of the PV systems in 2009, corrected to mean solar radiation in the last decade.

Figure 9 shows a linear regression between the annual energy production expected by the PV system owner (generally estimated by the installer, on the basis of the mean solar radiation corresponding the last decade) and the energy really produced during the year 2009 (adjusted to the mean solar radiation for the last decade as explained previously). Somewhat surprisingly, the linear regression shows no significant correlation between forecast and real production. The ratio $E_{\text{Forecast}}/E_{\text{Produced}}$ has a mean value of approximately 0.99, very close to 1, which indicates that this energy production is not overestimated taken as a whole. Nevertheless, the standard deviation of that ratio is close to 14%, which indicates that the installers generally do not estimate the energy production accurately.

Figure 9: Regression analysis between the energy produced in 2009 by the PV systems and the corresponding production expected by the PV system owner. No significant correlation is visible.

B) Tilt and orientation energy losses

The vast majority of PV generators have a tilt angle between 20 and 50 degrees, which generally corresponds to the configuration of the roofs on which they are mounted. At latitudes close to 50° North, a PV generator maximizes its annual energy produced when it faces South and benefits from a tilt angle around 40°. This would be the optimal orientation. When the orientation is different, which is usual in residential PV, the energy produced diminishes by an amount that is shown in Figure 10. The figure also shows the relative distribution, in percent, of the number of residential PV systems installed, as a function of the orientation and tilt. It is worth underlining that low tilt values favor dust accumulation (tilt angles of less than 10° have been reported to keep hold of important quantities of dust, as reported, e.g., in Hammond et al., 1997), but it is not frequent to find those low tilt values.

Figure 11 shows the relationship between the energy losses due to orientation and the proportion of PV systems installed. It is satisfactorily described by a power-law. Almost 70% of the PV systems lose less than 5% of their annual energy due to orientation, and less than one fifth lose more than 10%. As a whole, the orientation of residential PV causes energy productions to be 6% inferior to optimally oriented PV systems, which can be interpreted as the price to pay, in terms of energy losses, for installing PV systems on roofs instead of installing PV farms.

Figure 10: Relative distribution, in percent, of the number of PV systems installed, as a function of the orientation and tilt, together with the corresponding net annual energy produced by a PV system in Belgium with respect to the optimal inclination, in percent.

Figure 11: Proportion of PV systems (in %) oriented so that they lose less than a given percentage of net annual energy respect to the optimum orientation. Almost 70% of the PV systems lose less than 5% of annual energy due to their orientation, and less than one fifth lose more than 10%.

C) Technical quality

Figure 12 shows the evolution during the year 2009 of both PI and PR for a typical PV system of the sample, free of shading, not experiencing any lack of availability or other second-order problems, whose PI is 85%, whose PR_{STC} is 80.5% and whose PR is 78%. The PI is relatively constant throughout the year, while the PR varies of some 10% between winter and summer, mainly due to the evolution of cell's temperature. This lesser fluctuation of PI respect to PR suggests that PI is a better quality indicator of the intrinsic quality of a PV system than PR .

Figure 12: Evolution of PI , PR_{STC} , and PR for a PV system during the year 2009.

Figure 13 presents the histogram of PR and PI yearly values of 352 PV systems that provided at least 12 monthly produced energy data between January 2009 and August 2010. The mean value of PI is slightly under 85%, which indicates that, on average, the PV systems are producing about 15% less than the reference system on a yearly basis. Therefore, a PV system optimally oriented, located at Namur, and receiving the mean annual global horizontal solar radiation indicated by SoDa for this location, 975 kWh/m², would produce nearly 892 kWh/kW_p. This is the value that seems the most representative of the state of the art.

The distribution of PI is nearly normal between values from 70% to 100%. It is left-skewed, which physically arises from the existence of PV systems suffering from major issues and thus showing some PI values that are abnormally low, while even a very good PV system can hardly

have a *PI* much higher than 100%. The skewness can be modeled through a Weibull distribution. Further probability assessment plots of the *PI* have demonstrated that, at a confidence level of 95%, the distribution can be approximated by a Weibull (Anderson-Darling goodness of fit = 0.823) or normal (Anderson-Darling goodness of fit = 1.224) distribution. The Weibull fits better for extreme values, while the normal fits better for the central values. It was not possible to track *PI* values from previous works to compare them with the ones obtained in the present study. To make possible a direct comparison with the more widely spread concept of *PR*, Figure 13 also shows its corresponding histogram. The mean value of *PR* is 78%.

Figure 13: Histogram of the Performance Ratio (above) and Performance Index (below) of the PV systems analyzed. The distribution is nearly normal between a *PI* value of 70% and 100%. The distribution is left-skewed. The skewness is modeled through a Weibull distribution. The histogram of the Performance Ratio is very similar to the *PI* histogram because almost all the PV systems analyzed are subject to similar installation and climatic conditions.

As a comparison, values of *PR* between 48% and 93% have been reported in other works (Munro and Blaesser, 1994; Anis and Nour, 1995; Baltus et al., 1997; Blaesser, 1997; Decker and Jahn, 1997; Erge et al., 2001; Oozeki et al., 2010; Reich et al., 2012; Van Sark et al., 2012; Pelland and Abboud, 2008; Ueda et al., 2009; Eltawil and Zhao, 2010; Hussin et al., 2012; Wittkopf et al., 2012; Díez-Mediavilla et al., 2013; Woyte et al., 2013; Bonsignore et al., 2014; Colantuono et al., 2014; Myong et al., 2015; Tsafarakis and Van Sark, 2015; Kausika et al., 2018; Seme et al., 2019).

Both histograms show relative distributions that are very similar in the present case, because almost all the PV systems analyzed are subject to similar installation and climatic conditions.

In order to look for the causes that explain the *PI* differences among the different PV residential systems, we have applied an ANOVA to our database. It did not allow associating significant variations of *PI* to the nominal power of the installations, the inverters manufacturers, or installers. This failure to identify significant trends does not imply the absence of differences but simply suggests that the *PI* differences cannot be statistically attributed to any of these parameters.

The ANOVA did, however, allow to establish strong evidence that the PV modules explain the majority of the dispersion of *PI* ($F=9.94$ and $P\text{-Value} < 0.001$). The results of this ANOVA for PV modules present on at least 10 installations are detailed in Table 3. Manufacturers' names have been hidden under symbols for reasons of confidentiality: xSi stands for crystalline silicon; bcSi stands for back-contact silicon; HIT stands for Heterojunction with Intrinsic Thin layer; CIS stands for CuInSe₂ based solar cell (thin film). The systems equipped with the PV module tagged as "CIS" clearly show a *PI* pretty low respect to all the other groups.

Table 3: ANOVA on PV Modules present on at least 10 PV installations. *N* indicates the number of installations. The ANOVA analysis of PV modules shows significant differences between the mean power of several groups of PV modules.

PV Modules	N	Mean of PI (%)	StDev of PI (%)
bcSi	31	85.9	6.2
CIS	16	73.6	6.6
HIT	29	86.8	5.1
xSi1	26	87.8	6.6
xSi2	19	88.9	5.8
xSi3	11	88.9	5.6
xSi4	26	85.8	6.4
xSi5	58	83.8	6.2
xSi6	17	81.2	4.2
xSi7	10	84.6	5.7

Figure 14 shows a boxplot that allows visualizing the PI variations among and within the groups of PV modules.

Figure 14: Boxplot of Performance Index for PV modules present on at least 10 PV installations. The boxes show the first, second, and third quartiles represented respectively by the lower, medium and upper horizontal lines. The second quartile is also the median.

In order to estimate the real power of the PV modules, we assume that losses due to the Balance of System (BOS) are 10% higher than in the reference system. This assumption is supported by previous works that describe the losses typically present at a PV system. The soiling losses typically account for 3% (Hammond et al., 1997; Martin and Ruiz, 2006). The average inverter has a yield 2% lower than the high-quality inverter that equips the reference system (Muñoz et al., 2011). PV generator mismatch and wiring losses can typically be 2% higher than in the reference system (Martinez-Moreno et al., 2008). Shading can lead to important energy losses in some cases. The evaluation of shading losses is particular to each project and often implies complex models. The shading losses were not simulated for each PV system but were instead estimated to 2% on average, which seems a reasonable hypothesis for the typical residential PV systems in Belgium (Martinez-Moreno et al., 2011). Other losses, such as the ones due to the availability of the system, can account for 1% (Bletterie, 2007). Those losses can thus be estimated conservatively to account for 10% of annual energy losses. As the mean value of PI is 85%, there is a 5% left that is probably due to a power default in the PV modules.

Under those assumptions, the deviation of the real power of the PV modules, grouped by model, respect to the announced nominal power, is distributed as shown in Figure 15. In this figure, the analysis is extended to 46 different models of PV modules. It is worth mentioning that the PV modules analyzed here correspond to a mean exposure time of 2 years.

Figure 15: Histogram of the deviation of the real power of the PV modules respect to their nominal power. On average, the real power of the PV modules falls 4.7% below the corresponding nominal power. Some PV modules manufacturers provide very good quality modules. Some manufacturers deliver PV modules of poor quality, with a real power up to 16% below the nominal power.

On average, the real power of PV modules falls 4.7% below their corresponding nominal power. Other authors have reported the real nominal power of PV modules to be, on average 5% inferior to the nominal power stated by the manufacturer (Martinez-Moreno, 2011; Fraunhofer-ISE, 2007). The presence of PV modules showing a real power higher than their nominal power corresponds to PV modules delivered with positive power tolerances, or to a BOS better than the one considered in this analysis, or a combination of both factors. The majority of the PV modules have a real power between 2% and 8% lower than their nominal power. Some models of PV modules show poor quality, with a mean real power up to 16% below the nominal power. The gap between the poorest qualities and the rest corresponds to a change of technology between CIS-based modules and silicon-based modules. The multidimensional ANOVA allowed verifying that those conclusions about the real power of PV modules are not affected by other parameters of the installations, such as inverters or installers. Those differences between real power and nominal power suggest that it is profitable to implement quality control procedures to verify and improve the quality of PV systems (Caamaño et al., 1999; Coello et al., 2008).

2.3. France

2.3.1. French residential PV market

The data analyzed in the present study concerns Metropolitan France (i.e., excluding Overseas France). PV experienced there an important growth since the year 2004, with the establishment of a tax credit of 40% of the PV system cost. The growth was accelerated in 2005 with the rise of that tax credit up to 50%. But the decisive moment was the vote in 2006 of a new feed-in tariff specific to PV of at least 0.30 €/kWh, and up to 0.46 €/kWh for Building Integrated Photovoltaics (BIPV). As a direct consequence, from the year 2007, the number of residential PV systems started to take off, reaching 20 MW at the end of that year. At the end of March 2011, residential PV systems represented more than 550 MW. That power was distributed over more than 160,000 installations. At the end of the year 2018, the total installed PV power in France was 9.5 GW.

Figure 16 shows that 98% of the residential PV systems installed in France and contained in our database have a peak power of 3 kW_p or less, and more than half of the installations have a peak power very close to 3 kW_p. This situation arose as a direct consequence of a legal frame that strongly discourages installations of more than 3 kW_p, mainly for two reasons. First, the tax credits are denied for the PV systems of more than 3 kW_p. Second, a VAT of 5.5% is applied to systems of less than 3 kW_p, while it jumps to 19.6% for systems of more than 3 kW_p. The power distribution among residential PV systems in France is thus mainly explained by legal considerations, rather than technical ones.

Figure 16: Histogram of the peak power of PV systems. Nearly 98% of the residential PV systems installed in France have a peak power of 3 kW_p or less, and more than half of the installations have a peak power very close to 3 kW_p.

In our database, PV modules based on classical crystalline silicon (xSi) technology represent about 78% of the total market shares. The rest of the market is distributed, by order of importance, among heterojunction with intrinsic thin layer (HIT) (17%), amorphous silicon (aSi) (2%), copper indium (di)selenide (CIS) (2%), and cadmium telluride (CdTe) (1%).

The database contains 106 PV modules manufacturers, 21 inverters manufacturers, and 423 PV systems installers. Figure 17 shows that the relative market penetration within PV modules and inverters manufacturers is satisfactorily modeled by a power-law, indicating that the market is dominated by a reduced number of actors. The most extreme case of market domination is the control of more than 50% of the market by one single inverter manufacturer (the point corresponding to that inverter manufacturer is not represented in Figure 17 because it is out of scale, but is taken into account in the power-law equation). The leading PV modules manufacturer distributes a HIT technology and has a market share of 17%. The installers market does not follow a power-law mainly because many installers are small enterprises that only work at local scales, much smaller than the country.

Figure 17: Relative portion of the PV industry outstripping a given market share. The relative market penetration for PV modules and inverters manufacturers is satisfactorily modeled by a power-law. The R-squares of the fits yield 97.5% for the PV modules and 91.9% for the inverters.

2.3.2. Results

A) Energy production

For France, the data concerning the PV systems were supplied by their owners. Each PV system is localized by its latitude and longitude, completed with the corresponding altitude. The PV generator is characterized by its orientation and tilt angles, its total surface, and its total peak power. The data also provides information about the manufacturers of the PV modules and inverters that equip the systems and the installers. The net energy production is reported on a monthly basis and is read at the inverter (95% of the database), or at the meter (20%), or at both sources (15%). The PV owners also communicated the annual energy that they expected to produce, which was generally estimated by the installer. Not all the PV owners reported the energy production corresponding to each month, and only 35% of them reported it systematically and correctly.

Thanks to the PV owners that simultaneously provided the energy production data coming from both the inverter and the meter, it was possible to compare both sources of information. As shown in Figure 18, the ratio $E_{\text{inverter}}/E_{\text{meter}}$ shows a typical standard deviation of 3-4%, and a range of mean value of 1-1.07, depending on the inverter manufacturer. The ratios superior to 1 indicate that the inverters systematically overestimate the energy produced, up to 7% for

some inverter manufacturers. In the present study, the data provided by the inverter is adjusted by comparison with the data provided by the energy meter.

Figure 18: Histogram of the relation between the energy metering by inverter and meter.

The energy production analysis is carried out for the year 2010 and for the 1635 PV systems from which the monthly production was correctly reported for the 12 months of the year. Figure 19 shows a histogram of those energy productions. On average, the PV systems produced in 2010 a net annual energy of 1163 kWh/kW_p. The dispersion is mainly due to three factors: geography (and therefore solar radiation and temperature), orientation and performance. The solar radiation during the year 2010 in France was globally comparable to the mean radiation during the last decade. The energy productions reported are thus sufficiently representative to be compared with other previous studies in the literature. As a comparison, annual productions around 800 kWh/kW_p were reported for PV systems installed 5-10 years ago in the North and East of Germany (Jahn and Nasse, 2004). Two main causes explain the lower productions reported for the PV systems in Germany with respect to France. First, the solar radiation is globally higher in France. Second, the energy productions reported in Germany correspond to PV systems installed about 10 years ago, whose quality was probably lower, and whose power has decreased with time, mainly due to the light soaking.

Figure 19: Histogram of the production of the PV systems in 2010.

B) Energy losses related to tilt and orientation

The vast majority of PV generators have a tilt angle between 20 and 50 degrees, which generally corresponds to the configuration of the roofs on which they are mounted. At the latitudes of France, from 43° to 51° North, a PV generator maximizes its annual energy produced when it faces South and benefits from a tilt angle around 40°. When the orientation is different, which is usual in residential PV, the energy produced diminishes by an amount that is shown in Figure 20. That same figure also shows the relative distribution, in percent, of the number of residential PV systems installed, as a function of the orientation and tilt.

Figure 21 shows the relation between the energy losses due to orientation and the proportion of PV systems installed. It is satisfactorily described by a power-law ($R^2 = 97.5\%$). Almost 65% of the PV systems lose less than 5% of their annual energy due to orientation, and less than 10% lose more than 15%. As a whole, the orientation of residential PV causes energy productions to be 7% inferior to optimally oriented PV systems, which can be interpreted as the price to pay in France, in terms of energy losses, for installing PV systems on roofs instead of installing PV farms.

Figure 20: Distribution of the number of PV systems installed (out of a total number of 6868), as a function of the orientation and tilt, together with the corresponding net annual energy produced by a PV system in France with respect to the optimal inclination, in percent.

Figure 21: Proportion of PV systems (in %) oriented so that they lose less than a given percentage of net annual energy respect to the optimum orientation. Almost 65% of the PV systems lose less than 5% of annual energy due to their orientation, and less than 10% lose more than 15%.

C) Performance of PV systems

Figure 22 presents the histogram of PR and PI of 1635 PV systems that correctly provided the monthly produced energy for the year 2010. The mean value of PI is very close to 85%, which indicates that, on average, the PV systems are producing an amount of annual energy that is 15% inferior to the reference system. The PI observed in 2010 tends to be slightly higher for newer installations. The mean PI measured in 2010 for PV systems installed in 2007, 2008, and 2009 yielded respectively 82.9%, 83.5%, and 85.5%. Two main causes probably explain that trend. First, the power of PV modules is known to decrease with time due to the light soaking. Second, quality controls have been given growing importance during these last years. It was not possible to track PI values from previous works to compare them with the ones obtained in the present study. To make possible a direct comparison using the more widely spread concept of PR , Figure 22 shows its corresponding histogram. The mean value of PR is 76%.

In line with the observations made in Belgium, the distribution of PI is nearly normal between values from 70% to 100%. It is left-skewed, which physically arises from the existence of PV systems suffering from major issues and thus showing some PI values that are abnormally low, while even a very good PV system can hardly have a PI much higher than 100%. The skewness can be approximated through a Weibull distribution (at a confidence level of 95%, Anderson-Darling goodness of fit = 1.452). The distribution is more symmetrical for PR than for PI , mainly because the influence of the cell's temperature on the PR introduces wider variation among the PV systems installed in different climatic conditions.

Figure 22: Histogram of the performance ratio (above) and performance index (below) of the PV systems analyzed. The distribution is nearly normal between a PI value of 70% and 100%. The distribution is left-skewed. The skewness is modeled through a Weibull distribution. The distribution is more symmetrical for PR than PI, mainly because the influence of the cell's temperature on the PR introduces wider variation among the PV systems installed in different climatic conditions.

In order to look for the causes that explain the *PI* differences among the different PV residential systems, an ANOVA was applied to the whole database. It did not allow associating significant variations of *PI* to the nominal power of the installations, the inverter manufacturers or installers. This failure to identify significant trends does not imply the absence of differences. It simply means that the *PI* differences cannot be statistically attributed to any of these parameters with a sufficient confidence level.

The ANOVA did, however, allow to establish strong evidence that the PV modules explain the majority of the dispersion of *PI* ($F=23.21$ and $P\text{-Value} < 0.001$). The results of this ANOVA for PV modules that are present on at least 25 installations are detailed in Table 4. Manufacturers' names have been hidden under symbols for confidentiality reasons. xSi stands for crystalline silicon; bcSi stands for back-contact silicon; HIT stands for heterojunction with intrinsic thin layer; CIS stands for CuInSe₂ based solar cell (thin film). Among the results, it is possible to draw important observations about two PV modules technologies. On the one hand, the PV systems equipped with the module tagged as "HIT1" show *PI* values higher than average. This module is also the most represented on the PV systems of the database. On the other hand, the systems equipped with the PV module tagged as "CIS" clearly show a *PI* pretty low respect to all the other groups.

Table 4: ANOVA for PV Modules present on at least 25 PV installations. *N* indicates the number of installations. The ANOVA analysis of PV modules shows significant differences between the mean power of several groups of PV modules.

PV Module	N	Mean of PI (%)	StDev of PI (%)
bcSi1	47	83.7	6.2
CIS1	26	72.7	6.5
HIT1	283	88.7	6.1
xSi1	27	86.0	4.3
xSi2	32	83.6	8.8
xSi3	45	85.4	7.0
xSi4	70	79.3	8.1
xSi5	146	83.9	7.2
xSi6	137	87.2	6.5
xSi7	43	87.5	5.5
xSi8	127	87.9	6.8
xSi9	41	85.8	6.9

Figure 23 shows a boxplot that displays the PI variations among and within the groups of PV modules.

Figure 23: Boxplot of performance index for PV modules present on at least 25 PV installations. The boxes show the first, second and third quartiles, represented respectively by the lower, medium, and upper horizontal lines. The second quartile is also the median.

In order to estimate the real power of the PV modules, we assume that losses due to the Balance of System (BOS) are 10% higher than in the reference PV system, similarly to the analysis made for Belgium (see Section 2.2.2). As the mean value of PI is 85%, there is a 5% left that is probably due to a power default in the PV modules.

Under those assumptions, it is possible to group the PV modules by the manufacturer and to estimate the deviation of their real power respect to their corresponding nominal power announced by the manufacturer. Figure 24 shows the result of this exercise for 51 different manufacturers of PV modules. It is worth mentioning that the PV modules analyzed were subject to a mean exposure time of 2 years.

Figure 24: Histogram of the deviation of the real power of the PV modules with respect to their nominal power. On average, the real power of PV modules falls 4.9% below their corresponding nominal power.

The majority of the PV modules have a real power between 2% and 10% lower than their nominal power. The presence of PV modules showing a real power higher than their nominal power corresponds to PV modules delivered with positive power tolerances, or to a BOS better than the one considered in this analysis, or a combination of both factors. Two kinds of averages can be used to characterize the distribution as a whole. The first possibility is to give the same weight to the power deviations of each PV module manufacturer (unweighted average). The second possibility, more representative of the state of the art, is to weight the power deviations of each PV module manufacturer by the total power of its modules present in the sample (weighted average). The unweighted average yields 6.3%, and the weighted average yields 4.9%. The PV module that yields the best results is also the one that sold the best in 2010. It is a module based on HIT technology. Some models of PV modules show poor quality. A relevant observation concerns a PV module based on CIS technology, showing a mean real power up to 18% below the nominal power. To investigate the reasons for such low power, the PV systems equipped with this module have been grouped by year of installation. On average, a loss of power of 5% per year has been observed in these modules from 2007 to 2010. The low power is thus very probably due partly to an initial low power, and partly to a light soaking degradation higher than for the other technologies. The multidimensional ANOVA allowed verifying that those conclusions about the real power of PV modules are not affected by other parameters of the installations, such as inverters or installers. The comments received from the users of BDPV indicate that the geographical origin of the PV modules and inverters is often taken as an indicator of quality. In particular, they tend to consider as high quality the PV components manufactured in their country or in Europe, while they often turn down PV modules manufactured in China on the sole basis of their origin. Nevertheless, the analysis of the data of BDPV has demonstrated that no clear correlation exists between the performance of PV modules and the country where they were made.

D) Validation of the results against on-site measurements

The results of the present work have been compared to on-site measurements realized by the IES-UPM during the last years on more than 200 MW of PV modules equipping solar plants in Spain, Italy, and France (Martinez-Moreno, 2011). Among the modules manufacturers present on more than 25 installations of the database of BDPV, it was possible to identify 9 that were also measured by the IES-UPM. The relation between their real and nominal powers has been estimated by both methods and is shown in Table 5. The difference is generally lower than 2.3%. This is well within the uncertainties of those kinds of measurements. Additionally to the uncertainties on the measurements and on the estimations used in both methods, other sources of uncertainties are present from the fact that the PV modules that were rated here were not all installed at the same time, nor under the same climatic conditions, which possibly implies that some modules have already lost more power than others due to the light soaking process ongoing since they were exposed to outdoor conditions. Such agreements can thus be considered as very satisfactory. It is important to note that these conclusions are drawn only for PV modules present on at least 25 installations.

Table 5: Comparisons between real and nominal powers estimated in this work, and measured by IES-UPM on solar plants in Spain, Italy, and France.

PV Module (Hidden names)	Real vs. Nominal power (%) [This work]	Real vs. Nominal power (%) [on-site measurements]	Difference (%)
PV1	-6.4	-7.1	0.7
PV2	-4.6	-3.1	-1.5
PV3	-10.7	-12.3	1.6
PV4	-1.3	-2.1	0.8
PV5	-6.1	-4.7	-1.4
PV6	-2.9	-5.2	2.3
PV7	-2.1	-2.2	0.1
PV8	-4.2	-3.2	-1.0
PV9	-6.4	-6.6	0.2

2.4. Europe

The number of solar photovoltaic (PV) systems installed in Europe has drastically increased over the last few years, mostly thanks to the advantageous feed-in tariffs established by each country's government. A relatively small fraction of the energy production data of these PV systems has been analyzed, and as a consequence, there still remain wide gaps in the knowledge of the real-world performance of these PV systems. This feedback from the field is nevertheless important for the future development of the PV industry and for the establishment of new renewable energy development schemes by the respective governments.

In this work, we have tried to cast some light on some of the numerous remaining questions about the performance of PV systems in Europe. To do so, we have based our analysis on the operational data monitored at more than 31,000 PV systems in Europe. These installations comprise residential and commercial rooftop PV systems distributed over 9 different countries, and multi-megawatt PV plants installed in the South of Europe. These PV systems were installed between 2006 and 2014, and the data have been measured at time resolutions varying from minutely to monthly.

We have carried out an assessment of the energy production and the performance of all of these installations, from their commissioning date until the end of 2014.

We have looked for important trends on these installations, such as the decrease in performance with the time of operation, or the increase in performance due to improvements in the quality of PV system components and installation practices. We have also studied the distribution of the installed power per region, as a function of the type of installation and the external incentives.

We have characterized the state of the art of the PV installations, both in terms of PV system components, installation topology, orientation, and peak power, differentiating from the small residential PV market in northern and mid-latitude Europe and the PV plants in the southern and sunnier countries. We have applied statistical analyses on the whole database, which has allowed us to analyze the main parameters that explain the important differences that were observed among the installations in terms of performance, such as the PV module technology and manufacturer, the PV inverter, the region and year of installation.

2.4.1. Results

A) Annual energy yield

Figure 25 shows the distribution of the annual energy yield for Belgium, France, Spain, and the UK. The distributions in Belgium and in the UK are very similar, with a mean annual energy yield around 900 kWh/kW_p. This holds true in our database because a small proportion of the systems from the UK are located in the northern regions. The annual energy yield in France is

around 1,100 kWh/kW_p, with a wide distribution that largely reflects the difference in the solar resource between the north and the south of the country. The mean annual energy yield in Spain appears to be around 1,900 kWh/kW_p, but very few PV systems are encountered around that value, which corresponds to a valley between two peaks rather than to a peak. The explanation resides in the existence of two types of PV plants: static and trackers (See Figure 26). The mean annual energy yield for static PV plants is around 1,450 kWh/kW_p, and the corresponding value for PV plants with trackers is around 2,100 kWh/kW_p. There is a wide variation in the annual energy yield in Spain which, as with France, is largely due to the difference in solar resource between the north and the south of the country, but in the case of the PV plants with trackers, these differences also reflect the diversity in the tracking mechanisms.

Figure 25: Distribution of the annual energy yield for Belgium, France, Spain, and the UK. The differences observed are mainly due to the differences in the solar resource available. In the case of the PV plants in Spain, this also reflects the differences between static structures and tracking mechanisms.

Annual Energy Yield of PV plants in Spain vs static/tracker generator type

Figure 26: Distribution of the annual energy yield for PV plants in Spain. The differences are largely due to the difference in the solar resource between the north and the south of the country, but in the case of the PV plants with trackers, these differences also reflect the diversity in tracking mechanisms.

Table 6 summarizes the mean annual energy yield for the 4 reference countries.

Table 6: Mean annual energy yield data for the four reference countries.

Country	Mean yield (kWh/kW _p)
France	1115
UK	898
Belgium	908
Spain – PV plant – Static	1450
Spain – PV plant – Tracker	2127

B) Yearly integrated Performance Ratio

The yearly integrated PR has been calculated for all the PV systems, and for all the years for which data were available. Figure 27 shows the distribution of these yearly-integrated PR values. The distribution does not follow a normal (or gaussian) distribution, because an important fraction of the PV systems show an overall performance lower than average, and others are clearly subject to faults, which skews the distribution towards the low PR values. The distribution is better explained with a Weibull distribution, which often arises when the range of variation of a population is physically limited at one extremity, but not at the other. In the case of PV systems, it is very difficult to reach yearly integrated PR higher than 0.9, while it is much more likely that a PV system will have a PR much lower than average, because of performance problems.

Figure 27: The distribution does not follow a normal (or gaussian) distribution because it is skewed towards the low PR values. The distribution is better fit by a Weibull distribution.

We have tested the goodness of the Weibull distribution on our data. To do so, we have carried out a probability plot, whose result is shown in Figure 28. The probability plot shows that the Weibull explains very well the PR distribution from PR values ranging from 0.6 to 0.9. This range of values represents the majority of the PV systems. On the contrary, the Weibull distribution does not explain the data for PR values lower than 0.6 and higher than 0.9, which is clearly visible from the probability plot, and also from the P-value and from the result of the Anderson-Darling (AD) test. For PR values that are lower than 0.6, there is a significant

departure of the Weibull fit from the data, which shows that these data belong to a different population. These PV systems are subject to severe performance problems and faults. On the other hand, the Weibull law does not explain the PR values that are higher than 0.9, which also suggests that these PV systems belong to a different population. It is very likely that these PR values are simply not real, and that they are caused by uncertainties on solar irradiation data, as well as on the azimuth and tilt angles and the peak power of the PV systems. This probability plot allows us to suggest that the yearly integrated PR values ranging from 0.6 to 0.9 are representative of the state-of-the-art for PV systems in Europe. We suggest taking the scale (or peak or most frequent value) of the Weibull distribution as the typical value that represents the whole population. The shape of the Weibull fit to the data of all the PV systems from Europe is around 0.79, and we, therefore, suggest that this value is representative of the typical PV system in Europe installed over the last few years. Given the shape (the asymmetry) of this Weibull distribution, the mean value of the PR is some 3% lower than the most frequent value.

Probability plot of Yearly integrated Performance Ratio vs Weibull distribution - Europe

Figure 28: Probability plot of the yearly integrated PR against a fit on the data using a Weibull law. The probability plot shows that the Weibull law explains very well the PR distribution from PR values ranging from 0.6 to 0.9. This range of values represents the majority of the PV systems, and could represent the start of the art of the PV systems installed in Europe over the last few years.

Figure 29 represents the distribution of the yearly PR for the three reference countries whose installations are predominantly rooftop-mounted: France, UK, and Belgium. We take the typical (peak) values for yearly PR are 0.81 in the UK, 0.80 in Belgium, and 0.78 in France.

Figure 29: Distribution of the yearly PR for the three reference countries whose typology corresponds to rooftop installations: France, UK, and Belgium. The typical (peak) values for yearly PR are 0.81 in the UK, 0.80 in Belgium, and 0.78 in France.

Figure 30 shows the distribution of the yearly integrated PR for PV plants in Spain (both with static structures and trackers). The typical value of PR is 0.81. PV systems in Spain are more affected by thermal losses in the PV cells than the PV systems in the other countries of Europe. Notwithstanding, the PR obtained for the PV plants in Spain is still high when compared to the other countries in Europe. This high value for the PV plants in Spain in comparison with other countries reflects the higher efficiencies of larger PV installations, that take benefit from scale effects.

Figure 31 shows the comparison of the yearly integrated PR of the four reference countries on a boxplot. The middle line of the box represents the median or the 50th percentile of the distribution. The bottom line represents the 25th percentile, while the upper line represents the 75th percentile. The whiskers indicate the dispersion among the rest of the population, excluding outliers. The PV plants in Spain and the rooftop installations in the UK appear to have higher median values, followed by Belgium, and then France. The PR in France is some 3% lower than in the UK, and 2% lower than in Belgium.

Figure 30: Distribution of the yearly PR for PV plants in Spain. The typical value is 0.81.

Figure 31: Comparison of the yearly integrated PR of the 4 reference countries using boxplots. The PV plants in Spain and the rooftop installations in the UK appear to have the highest median value, followed by Belgium, and then France. The PR in France is some 3% lower than in the UK, and 2% lower than in Belgium.

Table 7 summarizes the result of the yearly PR calculations for the four reference countries.

Table 7: Yearly PR for the four reference countries.

Country		Mean PR [-]	Typical [-]
France	-	0.75	0.78
Rooftop			
UK - Rooftop		0.78	0.81
Belgium	-	0.77	0.80
Rooftop			
Spain	- PV plant	0.78	0.81

C) Influence of the peak power on performance

Figure 32 shows the relationship between the yearly integrated PR and the peak power of the PV system.

Figure 32: Relationship between the yearly integrated PR and the peak power of the PV system. We observe an increase in performance along with the increase in the peak power of the PV system. This is, in particular, true for PV systems whose peak power is lower than 4 kW_p.

We observe an increase in performance along with the increase in the peak power of the PV system. This is, in particular, true for PV systems whose peak power is lower than 4 kW_p. This is mostly explained by the yield of the inverter that increases as a function of their nominal

power, but in some cases, in particular, for large PV installations, the higher performance also arises from more demanding quality controls. These observations also advocate for more quality control procedures in small rooftop installations. This also suggests that when the policymaker decides to shape the PV landscape towards small systems, such as it was the case in France with systems limited to 3 kW_p, this is done at the price of lower PV system performance. An ANOVA allowed us to confirm that the trend observed is independent of external factors.

D) Influence of the inverter on performance

We have attempted to assess whether there was a significant difference in performance between PV systems equipped with inverters produced by different manufacturers. To do so, we have grouped the PV systems by inverter manufacturer, and we have calculated the yearly integrated PR for each one of these sub-groups. Figure 33 presents the result of this exercise, for the inverters that equip at least 100 PV systems in our database.

Figure 33: Boxplot representing the yearly PR of the inverters present on at least 100 PV systems in our database. We observe that there is a difference in the median value of PR of some 5% between the best performer and the worst performer.

We observe that there is a difference in the median value of PR of some 5% between the best performer and the worst performer. This seems to be partly in line with the datasheets of the inverter manufacturers, some of which deliver higher energy yields than others, but it also suggests that under real conditions, the real performance of some inverters departs more from their nominal performance than others. An ANOVA makes it possible to confirm that the differences observed between the best and the worst inverter manufacturers had enough statistical power.

E) Influence of the PV module on performance

We have also attempted to assess whether there was a significant difference in performance between PV systems equipped with PV modules produced by different manufacturers (except for the thin-film technologies, that are dealt with later on). Figure 34 presents the result of this exercise, for the PV modules that equip at least 100 PV systems in our database. The names of the manufacturers have been masked under a hidden code. The country of the manufacturers corresponds to the headquarters of the company, and not necessarily to the location of the manufacturing, which in some cases may be different. We observe that there is a difference in the median value of PR of some 6% between the best performer and the worst performer. This seems to be partly in line with the datasheets of the PV module manufacturers, most of which have a nominal power tolerance of $\pm 3\%$. Therefore, these differences could arise if some manufacturers sell their PV modules with a tolerance that is very close to $+3\%$, while some others sell with tolerances very close to -3% . An ANOVA made it possible to confirm that the differences observed between the best and the worst PV module manufacturers had enough statistical power. From the figure, we can also observe that there is no clear correlation between the geographical origin of a PV module manufacturer and the quality of the PV module.

Yearly integrated Performance Ratio vs PV module manufacturer - PV systems Europe

Figure 34: Boxplot representing the yearly PR for the PV modules that equip at least 100 PV systems in our database (excluding thin-film technologies). We observe that there is a difference in the median value of PR of some 6% between the best performer and the worst performer.

We have also attempted to study the impact of the PV cell/module technology on the performance of the PV systems. The results are shown in Figure 35.

Figure 35: Boxplot representing the yearly PR as a function of the PV cell/module technology. HIT = Intrinsic Thin layer; bcSi = back-contact silicon; xSi = classic crystalline silicon; xSi-Sat = Saturn silicon; CdTe = Cadmium Telluride. UMGSi = Upgraded Metallurgical Grade Silicon; aSi = amorphous silicon; CIS/CIGS = Copper Indium Selenide (CIS)/Copper Indium Gallium (di)selenide (CIGS).

The classic and most common crystalline silicon technology, marked as xSi, can serve as a reference to assess the performance of the other technologies relatively to it. We observe that the two technologies perform slightly better than the xSi technology. These technologies are the Heterojunction with Intrinsic Thin layer (HIT), and the back-contact silicon (bcSi). Other technologies show performances that are lower than the ones of the xSi. Immediately after the xSi technology comes the Saturn silicon (xSi-Sat). Then comes the Cadmium Telluride (CdTe). The dispersion of its performance appears to be higher, but this is mostly because the information that we possess on this technology comes from PV plants in Spain rather than in rooftops in Europe. The sample, therefore, appears as composed of fewer elements in the statistical analysis, although the total peak power represents around 30 MW. Furthermore, the dispersion mostly arises from one single PV plant that has been subject to severe performance problems for several years, that were due to the PV modules themselves (this PV plant appears at the lowest PR in Figure 30, with a value of 0.55). Nevertheless, as it can be observed, the median value of the entire technology is not much lower than for the xSi-Sat technology, and it is probably more representative of the newer modules made using this technology. Then comes the Upgraded Metallurgical Grade Silicon (UMGSi). The last two boxplots are for thin-film technologies and show a performance that is considerably lower than the rest of the technologies. The second poorest performer is the amorphous silicon (aSi). The worst

performer is the Copper Indium Selenide (CIS) and the Copper Indium Gallium (di)selenide (CIGS). Nevertheless, in this last case, the information that we had at our disposal only came from two PV module manufacturers, and both have now ceased their activities in PV, so our result should not be taken as representative of the latest CIS/CIGS technologies. An ANOVA made it possible to confirm that the differences observed as a function of the PV module technology had enough statistical power.

F) Evolution of the state of the art

We have investigated whether it was possible to observe any improvement in the state of the art of PV systems over time that would directly translate into an increase in overall performance. To do so, we have grouped the PV systems as a function of their year of installation. We have compared the yearly integrated PR of these groups during 2013 as a reference year. The result is shown in Figure 36, where we observe a remarkable increase in systems' performance over time, with an increase in yearly PR around 3 to 4% over the last seven years, which represents an increase of about 0.5% per year. For this result, we can also suggest that the state of the art of PV systems in Europe installed today is probably some 2% higher than the mean results that we have obtained with our data ranging from 2007 to 2014. Therefore, because the typical yearly PR for that period was assessed to be around 0.79, we can probably assert that the state of the art for PV systems installed in 2015 corresponds to a typical yearly PR around 0.81.

Yearly integrated Performance Ratio in 2013 vs year of installation of PV system - Europe

Figure 36: Yearly PR in 2013 for all the PV systems, grouped as a function of their year of installation. We observe an increase in systems' performance over time, with an increase in yearly PR around 3-4% over the last seven years, which represents an increase of about 0.5% per year.

G) Degradation of PV system performance

We have attempted to assess whether it is possible to observe any relevant degradation of the PV systems' performance over time. We have done so by analyzing the yearly integrated PR of all the PV systems installed in the year 2008, as a function of the year of production from 2009 to 2014. We chose the systems installed in 2008 because it is the first year for which we have at least 1000 PV systems with yearly energy production data that have passed the filterings. The result is shown in Figure 37, where we can observe that the degradation in PV systems' performance over the last six years ranges from unnoticeable to very low. The decrease in performance from 2009 to 2014 appears to be around 1.5%. Nevertheless, an ANOVA that was carried out on these results showed that it was not possible to conclude with a high statistical power that this observed decrease was not merely the effects of uncertainty, or reversely that the mean degradation is accurate. Therefore we have not been able to quantify degradation.

Yearly integrated Performance Ratio for PV systems installed in 2008 vs production year

Figure 37: Yearly integrated PR of all the PV systems installed in the year 2008 as a function of the year of production from 2009 to 2014. The degradation in PV systems' performance over the last six years ranges is very low, and we have not been able to quantify it.

H) Monthly data analysis

Apart from our analyses carried out on yearly integrated data, whose objective was to reach a global understanding of the performance of the PV systems in Europe, we have analyzed the distribution of the monthly integrated PR for all the PV systems in Europe to investigate whether other complementary observations could be obtained from them. The result is shown in Figure 38.

Figure 38: Dispersion in the monthly integrated PR. It is much higher than the one observed for the yearly integrated PR.

The dispersion in the monthly integrated PR is much higher than that observed for the yearly integrated PR for several reasons. First, the uncertainty in PR is higher during the winter months, which is made more explicit when monthly data are analyzed independently. Second, any performance problems that span one month or less are much more visible on the monthly data than on the yearly data. This dispersion is, therefore, a strong indicator of the presence of performance problems or faults on some of the PV systems. More evidence of short-lived faults appears when we move towards higher temporal resolutions, e.g., hourly data. The detection of this kind of performance problems and faults is the object of Chapter 3.

3. FAULT DETECTION APPLIED TO PV SYSTEM FLEETS: PERFORMANCE TO PEERS (P2P)

Effective fault detections require performance indicators that are as stable as possible in the absence of fault, and that vary significantly in the presence of a fault. Figure 39 illustrates an ideal (and fictitious) performance indicator, which is always equal to a constant value, e.g., 1 or 100%, when the PV system is operating correctly and decreases when a fault takes place. The evolution of this performance indicator can be analyzed as a function of time, at a specific time resolution, or also against other key parameters such as solar irradiance level, solar position, air temperature, etc.

Figure 39: Fault detector constructed on the basis on a hypothetical ideal performance indicator. This performance indicator is always equal to a constant value, e.g., 1 or 100%, when the PV system is operating correctly and decreases when a fault takes place.

Unfortunately, such a perfect performance indicator is not achievable in reality, due to the limitations of the models and the accuracy of the input data. This study has attempted to produce a performance indicator that makes it possible, as much as possible, to distinguish between normal operation and fault.

3.1. Methodology overview

Bearing in mind to produce a good performance indicator for fault detection, and for every PV system to be analyzed, called the *focus* PV system, the procedure is articulated around three main successive steps. First, the model determines which of the neighboring installations constitute the best peers, and it quantifies their relative goodness as peers. Second, the P2P is constructed from comparisons between the energy production of the focus installation and its

peers. Third, the P2P is applied to fault detection. The next sections explain these steps one by one.

This fault detection procedure has been developed in the context of the performance analysis carried out on approximately 6,000 PV (more precisely 6,169) systems located in Europe, mainly in Belgium, France, and the UK, as shown in Figure 40. The geographic distribution of the monitored PV installations made it possible to develop and validate the models for a fairly varied range of weather conditions. Most of these PV systems are residential rooftop installations whose generator peak power ranges from 1 to 10 kW. These installations are equipped with smart energy meters connected to General Packet Radio Service (GPRS). The energy output data represent approximately 7 years of operation (2011-2018) and were measured with a temporal resolution of 1 data every 10 minutes.

Figure 40: Location of the $\approx 6,000$ PV systems monitored in Europe.

Other data were also available for this study. In particular, key metadata on the PV systems was provided by the installers, such as their location, peak power, orientation, tilt, etc. These metadata were used for each PV system to obtain hourly solar irradiation data in the plane of the PV generators from the satellite-derived database HelioClim (Blanc et al., 2011) and

acquired through the web service SoDA. These data were used to associate a PR to the P2P and make it possible to compare both performance indicators.

Most of the specific developments have been based on PV installations located in the South of Belgium (Wallonia), where the database contains a higher density of installations, as illustrated in Figure 41. The year 2015 was chosen as a reference year for the analyses, and a fault detection validation campaign was carried out during that year. In fact, with an installed PV power of 374 W_p/inhabitant, Belgium is currently the ninth country in the world in terms of PV penetration, and its high population density also translates into a high density in terms of PV installations over the territory.

Figure 41: Closer view over one part of Wallonia presenting a high density of monitored PV installations contained in the database.

3.2. Looking for the best peers

For a given focus PV system, the best peer would ideally be its exact twin, installed just at its side, and whose performance is never subject to any problem. In practice, good peers are installations that are located relatively close to the focus, and whose most relevant technical characteristics are similar: tilt and azimuth angle, PV module technology, inverters, peak power, etc. Neighboring installations whose key characteristics are too different from the focus constitute bad peers. In between the two extremes, the quality of a neighboring system as a peer varies as a function of the similarity of its characteristics to the ones of the focus.

The challenge is to quantify the goodness of each potential peer and to determine which ones to use in the analysis of the performance of the focus. The parameters that can influence the goodness of a peer are numerous, and their combined effect is complex. Furthermore, these parameters vary over time, and so also does the relative goodness of the peers. Even if there existed a perfect physical model capable of explaining the effect of all of these parameters at every moment, this model could not be used in practical cases because all of these characteristics are often only known with limited accuracy, or they are even unknown. As a

consequence, this method has been developed so that it makes it possible to assess the goodness of the peers from nothing more than the comparison between the energy output data of neighboring installations.

The Capacity Utilization Factor (CUF) can be defined as:

$$CUF = \frac{E_{PV}}{P^*T}$$

The CUF makes it possible to normalize the energy output of a PV system (E_{PV}) by its peak power (P^*) and the time interval (T). It thus makes it possible to compare the energy output corresponding to PV systems of different peak powers, and during different time intervals. The peak power data is generally known for most of the PV installations, but it cannot be taken for granted that it is known with high accuracy, or it can even be unknown in some cases. Nevertheless, this data is only used for a normalization process, and the rest of the method has been developed in such a way that the results are not affected its value. In practice, if this value is known, it is convenient to use it because it contributes to producing results that are more intuitive. If no data is available, the peak power can be taken from best guess or set by default as 1 kWp.

The CUF represents the fraction of the energy produced by the PV system relative to the energy that would have been produced if the PV system had been producing at peak power during the whole time interval.

The comparison between the CUF of neighboring PV installations provides the starting point to the determination of their degree of similarity. For a focus installation and each one of its neighbors, the ratio between the CUF of both installations is calculated, which produces a Capacity Utilization Ratio (CUR) for each focus-peer couple, expressed as:

$$CUR = \frac{CUF_{focus}}{CUF_{peer}}$$

Two neighboring twin installations without any performance problems have the same CUF during any time interval, the temporal variation in their CUF is the same, and the ratio between their two CUF - their CUR - is equal to 1 at any time. Therefore, the typical fluctuation in the CUR between a focus and a peer installation provides an idea of the degree of correlation that exists between their energy outputs. A high dispersion indicates a low correlation (Elsinga and Van Sark, 2015; Taylor et al., 2015b). A neighboring installation constitutes a good peer if its energy output is highly correlated to the focus, and therefore if the fluctuation of the associated CUR is low. The degree of fluctuation of the CUR depends on the time resolution at which it is calculated. Therefore, the choice of this time resolution has a strong influence on the evaluation of the goodness of a peer. The ideal time interval is such that the CUR is stable enough so that most of its fluctuations are useful to represent the goodness of a peer, but at the

same time fluctuates enough as a function of the key operating conditions so that it provides enough information for the evaluation of the goodness parameter. The correlation between the energy outputs of two neighboring PV systems tends to be lower during shorter time intervals because of the dynamics of solar irradiance (Marcos et al., 2011). At shorter time intervals, in the order of the minutes, the fluctuation in the *CUR* is noisy, and it is not suitable to obtain a clear idea of the general correlation trends that are the most relevant when analyzing system performance, but it instead provides an image of the dynamics of the cloud cover. During longer time intervals, these correlations tend to be higher, and they remain so over larger distances between the two installations (Marcos et al., 2012). On the other hand, it is not desirable to choose long time intervals, because this makes it more difficult to assess whether the correlation stands for most of the operating conditions of the PV systems. Furthermore, the time interval has to be short enough so that each time value can roughly be associated with a solar position. The developments have shown that the hourly time step is convenient because it makes it possible to reflect the change in solar position and the global change in cloudiness, while it filters out the short-term fluctuations that are of little interest and that would blur the important results.

The time period during which the fluctuation in the *CUR* is evaluated is also a dilemma. It needs to be long enough so that it is statistically representative of a reasonable amount of operating conditions, while it has to be short enough to take into account any fundamental change in these operating conditions. Additionally, it is often necessary that the data can be analyzed as soon as possible without the need for long periods of historical data. In practice, a time period between one week and one month leads to a good compromise. This work has evaluated the fluctuations in *CUR* based on one-month time periods.

The degree of correlation between focus and peer can be quantified through the standard deviation σ of the hourly *CUR* of the couple focus-peer during a period of one month. Nevertheless, the standard deviation is affected by outliers, and its value is often driven by the abnormalities instead of the central trend that is of interest for the assessment of the focus-peer correlations. Therefore, it is better to use a robust estimator of the dispersion. For this reason, the dispersion is calculated through the Median Absolute Deviation (MAD) of *CUR*, which is expressed as (Gauss, 1816; Walker, 1929):

$$MAD(CUR) = median(|CUR_i - median(CUR_i)|)$$

There is a direct relationship between the MAD and the standard deviation σ of the central and gaussian part of a statistical population, given as:

$$\sigma \approx 1.4826 \cdot MAD$$

Therefore, the MAD can be used as a robust estimator of the standard deviation.

The MAD(CUR) between a focus and a neighboring installation tends to be lower if the correlation between the energy outputs of these two installations is higher, and therefore if the neighbor is a better peer. For a given focus installation, a weight w can be assigned to each peer as a function of the MAD of the CUR focus-peer. The relationship between MAD(CUR) and w has the general form:

$$w = \frac{1}{[MAD(CUR)]^\alpha}$$

The determination of the best exponent α follows several considerations.

If n independent observations are carried out whose associated standard deviation is σ_i , and if the distribution of the errors is gaussian, then the resulting standard deviation σ_{TOT} associated with the combination of these observations is given by:

$$\sigma_{TOT} = \frac{\sigma_i}{\sqrt{n}}$$

Therefore, e.g., a peer installation whose $\sigma(CUR)$ is of 1% is as good in terms of contribution to the global indicator as four peers whose $\sigma(CUR)$ is 2%. More generally, a peer whose $\sigma(CUR) = x$ is as good as n^2 peers whose $\sigma(CUR) = nx$. Therefore, the goodness g of one peer installation is proportional to the inverse of the square of its associated standard deviation:

$$g = K \frac{1}{\sigma^2}$$

Because $\sigma(CUR)$ is proportional to $MAD(CUR)$, the same form of relationship also stands true for the MAD, and g can be defined as:

$$g = \frac{1}{MAD^2}$$

The $P2P$ is constructed by interpolation of the energy output from its peers, and these peers are located around the focus over a two-dimensional surface. The interpolation method that is derived from the estimation of g through MAD(CUR) bears analogy to the inverse distance weighting interpolation method (Shepard, 1968), where the weighting factor is obtained from the goodness parameter instead of obtaining it from the distance. The inverse distance interpolation has the following general form:

$$w_{di}(x) = \frac{1}{d(x, x_i)^p}$$

Mathematically, the exponent p must be ≥ 2 for the method to converge. Therefore, the number 2 has been chosen as the exponent:

$$w = g^2$$

This exponent choice finally leads to:

$$w = \frac{1}{MAD^4}$$

The analyses applied to the installations indicate that this exponent of 4 produces good results, although the sensitivity around this value is moderate and other values close to 4 could also be considered. These weighting factors are normalized for each peer so that the sum of the weighting factors of all the peers of one focus installation is equal to unity:

$$\lambda_i = \frac{w_i}{\sum w_i}$$

where λ_i is the relative weighting factor for the peer i .

In practice, the construction of the P2P is limited to the 50 best peers for each focus, in order to limit the calculation times, and also because the quality of P2P does not significantly improve anymore beyond a few dozen peers.

Figure 42 represents the distances between one particular focus installation and all its peers located at less than 15 km. The area of each circle is proportional to the weight of each peer. The successive ellipses represent 1, 2, 3, 4, 5, 10 and 15 km. The best peers tend to be close to the focus, but the distance is not the only key parameter. In particular, the similarity in the orientations also plays a major role.

Figure 42: Distance focus-peers on a map (lat, long) with points size proportional to weights.

3.3. Constructing the P2P

The P2P is constructed by comparison between the CUF of the focus installation (CUF_{focus}) and a reference CUF (CUF_{ref}), calculated as a weighted median of the CUF of all its peers, using the weighting factors λ :

$$CUF_{ref} = percentile_{50}(CUF, \lambda)$$

The weighted median of a partition divides it into two halves such that the sum of weights in each partition is as equal as possible (Edgeworth, 1887; Edgeworth, 1888). Other percentiles than the percentile 50 (the median) are possible, but the median is robust, and it generally leads to a P2P that is more stable than with most other percentiles.

The P2P is calculated as the ratio of the CUF of the focus and the reference CUF:

$$P2P = \frac{CUF_{focus}}{CUF_{ref}}$$

A P2P equals to one (or 100%) means that on average, the PV system that is analyzed performs as well as the median performance of its neighboring installations. This definition makes it possible for the P2P of a PV installation to reach values that are higher than 100%, i.e., 110%, in particular if for some reason the focus installation has a higher performance than most of its peers. Conversely, in some cases, the overall value of the P2P of an installation that is operating without any particular fault can be somewhat lower than 100%, i.e., 90%. Furthermore, because the peak power is not always known with accuracy, the P2P of some installations can greatly

differ from 100%. The absolute value of the P2P is not relevant because the fault detection procedures are based on the variations in this P2P. Nevertheless, it is often more intuitive to set the typical value of the P2P of a PV installation around 100%. Therefore, this P2P is usually normalized so that its median value is 100%.

The P2P is relatively stable, even when the energy output of one or several of the peers is unstable. Figure 43 shows the P2P of a PV system, at a 1-hour time resolution and during a day, along with the CUF of its 5 best peers, sorted by descending goodness order. The day was globally characterized by clear-sky conditions, except for the central hours of the day, during which a light cloud cover induced a decrease in the energy output of the neighboring systems at different moments. The P2P remained very stable most of the time, and the CUF_{ref} remained very close to the CUF_{focus} , even when the CUF of several peers departed from one another. Nevertheless, the P2P shows much higher variations during the first and last moments of the day, which correspond to low solar elevations. Under these conditions, it is more likely that two neighboring installations will not operate under the same conditions, and small differences in the absolute value of the energy outputs can lead to important differences between the CUF of two peers. In consequence, the P2P cannot remain stable under all circumstances, and the fault detector needs to determine when it can be trusted or not, which is the object of the next section.

Figure 43: P2P of a PV system, at a 1-hour time resolution and during a day, along with the CUF of its 5 best peers.

Figure 44 shows the daily P2P associated with the PR of the installation introduced in Figure 1, and it illustrates the difference between the two performance indicators in terms of stability. The P2P makes it possible to identify that this installation was affected by faults during three distinct periods of time. Nevertheless, the situation is not clear for one particular day in September. The next section presents a systematic approach that makes it possible to detect these faults by distinguishing between normal and faulty operation.

Figure 44: Daily PR and P2P for the installation already introduced in Figure 1. The P2P is more stable than the PR. Three faulty periods are clearly distinguishable, while the situation is not clear for one particular day in September.

3.4. Applying the P2P to fault detection

The automatic fault detection procedure is carried out through the establishment of a minimum threshold value of P2P, below which the performance is considered as abnormal. This threshold can be set statically or dynamically.

The establishment of a static fault detection threshold is relatively simple. It is illustrated in Figure 45, which shows the histogram of the daily P2P of one installation in Belgium for the year 2012. Two distinct statistical populations appear. Most of the P2P values (in green) are clustered around 1 and are distributed following a somewhat normal distribution. These P2P values are representative of normal functioning. Other P2P values (in red), are found between 0 and 0.9. These values correspond to performance problems (faults). The fault detection threshold needs to separate these two different P2P populations. The expert eye would set this separation around P2P values of 0.93. Previous works have presented procedures that lead to the establishment of such a static threshold (Leloux et al., 2014; Tsafarakis et al., 2018; Vergura, 2018a, 2018b).

Figure 45: Histogram of daily P2P on one PV system during 2012. The green bars are representative of the normal operation. The red bars represent faults.

Nevertheless, this static fault detection procedure bears limitations. The uncertainty around the P2P is not constant over time, and it depends on the prevailing weather conditions (Shahandashti et al., 2018). For example, the uncertainty is lower during clear-sky conditions, and higher during fast-evolving cloud conditions, or at low solar elevation angles. Therefore, a dynamic fault detection threshold has been developed that is able to adjust automatically as a function of the uncertainty surrounding the *P2P*.

For a given time interval, e.g., during one hour or one day, the P2P of the focus installation is more reliable if the majority of its best peers agree on its value. A proxy to this agreement is the similarity of the CUR values between the focus and each one of its peers. The absolute value of the CURs can differ from one another. Nevertheless, once normalized by a value representative of their respective long-term trends, the dispersion between them can be used to evaluate the reliability of the P2P.

The CUR values are normalized as:

$$CUR_{NORM} = \frac{CUR}{\mu_{1/2}(CUR)}$$

where $\mu_{1/2}(CUR)$ is the median of the CUR during a period of time that is long enough to be statistically representative of the current overall level of performance of the installation, e.g., one month.

The reliability of the P2P at a moment t is evaluated as the weighted median absolute deviation (MAD_w) of CUR_{NORM} :

$$\Phi_{P2P} = MAD_w(CUR_{NORM}, \lambda, t)$$

where the weighting factors of the peers, λ , give more importance to the CUR of the best peers in the calculation of the MAD on CUR_{NORM} .

A confidence interval around P2P is deduced as:

$$CI_{P2P} = (z \cdot K_{\sigma/MAD} \cdot \Phi_{P2P} + \varepsilon_{P2P}) \cdot \mu_{1/2}(P2P)$$

The factor z represents the number of standard deviations by which the value is away from the mean value. Under a Gaussian distribution hypothesis, the factor z can be adjusted as a function of the confidence level that is desired. A value of $z = 3$ is very common in quality control and fault detection procedures. As explained before, σ can be deduced from MAD . The factor ε_{P2P} is introduced to provide a tolerance margin to the confidence interval so that even if the calculated ϕ_{P2P} is very low, the CI_{P2P} still possesses a minimum width. This is important to avoid some false alarms during the fault detection that correspond to hypothetical faults whose energetical importance is small. For example, this value of ε_{P2P} can be set to 5%. The factor $\mu_{1/2}(P2P)$ represents the median of P2P over a longer period, e.g., one month, and the multiplication by this factor makes it possible to normalize the CI_{P2P} to the absolute values of the P2P for this focus installation.

Once the confidence interval around P2P has been calculated, the fault detection threshold is set as:

$$\theta_{P2P} = \mu_{1/2}(P2P) - CI_{P2P}$$

The application of this procedure to the data of Figure 43 is illustrated in Figure 46, where the P2P is represented with its confidence interval (the grey zone) and the fault detection threshold (the red line). An alternative representation of the same reality is shown where the CUF_{focus} is represented within a confidence interval applied around the CUF_{ref} . This case is representative of an installation whose P2P is very stable, owing to peer installations that are close and similar to the focus. The P2P and its confidence interval do not fluctuate significantly, even in the presence of a light cloud cover in the middle of the day. Nevertheless, the reliability of the P2P rapidly decreases at low solar elevations, but the associated energy production is often irrelevant. For this reason, the fault detections applied to hourly and sub-hourly data have been restricted to solar elevation angles higher than 10° . This is visualized through the exclusion of the early morning and late afternoon from the confidence interval around P2P. In this particular case, no data has been identified as a fault because no P2P value is below the fault detection threshold.

Figure 46: CUF, P2P, confidence intervals, and fault detection threshold corresponding to a PV installation, at 1-hour time resolution.

The application of this fault detection procedure to the example of the installation of Figure 1 and Figure 44 makes it possible to obtain a systematic categorization of energy production as normal operation or fault, as illustrated in Figure 47. The fault detector confirms that the installation has gone through three periods of several days affected by faults. It also confirms that another day in September presented a fault. The P2P under normal operating conditions has been relatively stable, and the fault detection threshold itself has remained stable. Nevertheless, some days were marked by a significant decrease in the fault detection threshold, which means that even on daily data, the P2P has not been reliable during some moments. This further illustrates the need for a dynamic fault detection threshold.

Figure 47: Fault detection applied to daily P2P data. The installation has been affected by faults during several periods of time. The confidence interval around the P2P shows important fluctuations, which illustrates the need for a dynamic fault detection threshold.

Once the faults have been detected, it is possible to assign them an energy loss. The energy losses associated with a fault can be calculated as:

$$\Delta E_{\text{fault}} = \frac{\mu_{1/2}(P2P) - P2P_{\text{fault}}}{\mu_{1/2}(P2P)}$$

where $P2P_{\text{fault}}$ represents the value of the P2P corresponding to a fault.

Many days convey periods of time of some minutes to some hours that can greatly differ from the previous one or the next ones. For sub-hourly time resolutions, this variability is further increased, and the hour is generally a good compromise between time resolution and stability, and it is not always possible nor advisable to go beyond it. Nevertheless, some faults can be better detected and diagnosed at higher time resolutions, and such monitoring data are becoming increasingly available.

Most of the time, at sub-hourly resolutions, it is advantageous to limit the fault detections to clear-sky or overcast conditions because cloud transitions induce noise that is not related to any particular faulty mechanism that is intrinsic to the operation of the PV system. From sub-hourly time resolution data, it is possible to automatically detect the presence of clear-sky or overcast conditions on the sole basis of the shape of the CUF, because the evolution of the solar irradiance, and hence the CUF, is much smoother. The exact procedure to be applied to the data is out of the scope of this work, but it has been described in previous works (Reno et al., 2012).

Figure 48 illustrates the incorporation of a clear/overcast sky filtering on CUF at a 10-min time resolution during a period of five consecutive days. The first two days and the last one are predominantly characterized by a fast-evolving cloud cover, and most of the P2P values are excluded from the fault detection, which is visible through the gaps in the confidence interval around P2P. On the contrary, the third day is exclusively characterized by clear-sky conditions, and the fourth day is mostly characterized by clear-sky conditions. During these two days, a pattern emerges where faults are detected in the early morning and in the late afternoon. The same pattern repeats itself over most of the days of summer in the presence of clear-sky conditions, and it is due to shading.

Figure 48: Incorporation of a clear/overcast sky filtering on CUF at a 10-min time resolution.

Figure 49 presents a closer view on a day of summer that is characterized by clear-sky conditions in the morning and late afternoon, and by fast-evolving cloud conditions during the middle of the day. Two distinct groups of faulty operating conditions are visible. The evolution of P2P suggests that the shading that affects the installation in the morning is provoked by an object whose angular azimuthal size is relatively large but whose impact on the shading fraction is relatively low, while the shading in the afternoon is cast by an object that is of smaller azimuthal extent but of higher impact on the shading fraction.

Figure 49: Closer view on a day of summer that is characterized by clear-sky conditions in the morning and late afternoon, and by fast-evolving cloud conditions during the middle of the day. Two distinct groups of faulty operating conditions are visible.

3.5. Fault diagnosis

The topic of fault diagnosis, generally speaking, is out of the scope of this work, because each type of fault obeys to a proper pattern and requires a specific procedure. An increasing amount of work is being published on fault detection, but much remains to be done in terms of the proper identification of fault patterns (Dhimish et al., 2017; Gallardo-Saavedra et al., 2019; Livera et al., 2019). This section illustrates the application of the P2P to the diagnosis of the shading affecting the installation presented in the previous section. For that installation, a PV expert is capable of inducing that shading problems are present by looking at the evolution of the 10-min P2P over time, but a time series analysis is not the best approach to reach a systematic diagnosis. That problem is better tackled through a representation of the P2P as a function of the solar position. Figure 50 presents the 10-min P2P values accumulated over one complete year on a solar azimuth plot. The solar position area is considered as shaded when its P2P is more than 3 standard deviations under the median value obtained between all the boxes, and under clear-sky conditions. The application of this procedure in the case of this PV installation leads to the consideration that a solar position is characterized by shading when its P2P is lower than 0.93.

Figure 50: Shading diagnosis on a solar azimuth plot using $P2P$ as the performance indicator.

More advanced diagnosis and quantification of the shading can be attained through the representation of the $P2P$ on a solar position plot, as a function of both solar azimuth and elevation angles, as illustrated in Figure 51 for the same installation. The $P2P$ values have been grouped as a function of the solar azimuth and elevation angles, using an angular step of 10° , and the value associated with each solar position is obtained by the median of all the $P2P$ values that fall within each solar position box. The value of the $P2P$ is represented through a heatmap, where red-shifted colors correspond to lower $P2P$ values, and hence to higher shading losses. This visualization also provides a more intuitive representation of the angular shape of the surrounding objects that cast shadows on the PV generator.

Figure 51: Heatmap of the shading diagnosis obtained from the 10-min $P2P$ data of one PV installation for one year. The value of the $P2P$ is represented through a heatmap, where red-shifted colors correspond to lower $P2P$ values, and hence to higher shading losses.

3.6. Application of P2P to PV system fleets

This section presents the global lessons learned through the application of this fault detection procedure to the energy output data measured at more than 6,000 PV installations.

During the reference year 2015, every two weeks, a fault detection report has been generated from the faults that were detected in all the PV installations analyzed. The reports contained the identification of faults that were observed on these installations. Faults have been detected on more than 500 of these installations, i.e. slightly less than 10% of the database. This does not imply that no other installation has been affected by a fault, but it is merely what this procedure has been able to detect. It is not possible to determine the exact amount or nature of the faults that each one of these installations has undergone, because this would require to install complex measurement hardware at each one of these installations and to possess all the necessary tools to analyze these data. Nevertheless, a validation campaign has been attempted. During the summer of 2015, a random sample of 100 installations with faults was chosen and their owners were contacted. Feedback has been received from 30 of the system owners. If the fault had been observed by the owner, then more information was asked on its nature. The objective was to verify whether the faults that were detected by the model were real and whether the model was able to detect all the major faults hindering the performance of the installations. The installation owners were therefore asked whether they had been aware of some other problem on their installation that the fault detection procedure could have failed to notice.

On 29 out of the 30 installations, it was possible to confirm that the faults that were detected were real. Only once the fault detection resulted to be a false alarm, due to a problem related to IT rather than the P2P itself. The faults that were detected using the P2P had different origins. Approximately half of them concerned the inverters, which suffered from problems relating to their switch, their internal clock, their ground-fault circuit interrupter (GFCI), or their insulation. Other problems were related to general power outages from the grid, voltage problems at the interface between the grid and the inverter, faulty contacts, defective bypass diodes on the PV modules, humidity problems in the contacts of the PV modules, or shading.

Some global trends have been observed concerning the uncertainties around PR and P2P. A comparison between the uncertainty on PR and P2P was carried out on the basis of the hourly data because the hourly PR and solar irradiation data were provided by the monitoring service that provided the energy output data. The median MAD(PR) on all the PV installations for the whole year 2015 for hourly data is 0.17, with a corresponding relative standard deviation of 25.5%, to be compared with an MAD(P2P) of 0.10. The standard deviation for P2P does not have any significance because it is driven almost exclusively by the extreme values that are filtered out during the fault analyses. The same considerations apply, in lesser proportions, to the PR, and it is, therefore, better to compare both performance indicators on the basis of the MAD. In these regards, the P2P is generally much more reliable than the PR for fault detection. This higher uncertainty surrounding the PR is not only the result of the uncertainties on the solar irradiation but also of the metadata on the orientation and tilt angle of the installations.

Without an accurate knowledge of these metadata, the associated solar irradiation data makes little sense, while P2P does not need these data, which makes it a more robust performance indicator.

Despite its relative robustness, the stability of the P2P varies from one installation to another. It has also been shown to be greatly affected by the weather conditions themselves, leading to important seasonal variations in the MAD(P2P). Figure 52 presents the boxplot of the MAD(P2P) applied to the hourly P2P values calculated for all the PV installations located in Belgium, and it shows that the uncertainty is higher in winter than in summer.

Figure 52: Boxplots of hourly P2P as a function of the month of the year, showing that the uncertainty is higher in winter than in summer.

Figure 53 shows the corresponding seasonal evolution of the MAD(P2P) applied to daily P2P data. The overall level of uncertainty is much lower, but the seasonal trends remain.

Figure 53: Boxplots of daily P2P as a function of the month of the year, showing that the uncertainty is higher in winter than in summer.

It has also been observed that the stability of the P2P of a focus installation greatly depends on the proximity and similarity of its peers, up to a certain limit. If a focus installation has many peers close to it, its P2P tends to be of better quality, translating into a lower MAD(P2P). Figure 54 shows the evolution of the MAD(P2P) as a function of the number of peers located at less than x km from the focus, with $x = 1, 5$, or 10 km. These statistics are represented for the month of March 2015, because at the spring equinox the level of uncertainty is intermediate between summer and winter. The trends show that focus installations that possess many peers close to them can usually benefit from a P2P that is more stable. Nevertheless, this relationship is not straightforward, and in practice, the P2P is generally stable enough if the focus PV installation possesses several peers that are similar to it located at less than 5 km, or if it possesses a dozen of similar peers located at less than 10 km, with intermediate cases also visible.

Figure 54: *MAD(P2P) as a function of the number of peer installations located at less than x km from the focus, with $x = 1, 5$ or 10 km.*

4. SOLAR IRRADIATION FROM PV ENERGY PRODUCTION

The measurement of the PV energy output can constitute an indicator of the amount of solar irradiation received by a PV system. However, using the energy production registered at PV systems as a solar irradiation sensor is not straightforward (Killinger et al., 2016; Elsinga et al., 2017). It requires first to derive the solar irradiation from the energy production of the PV system (knowing that the power output of a PV system is not directly proportional to the solar irradiance that it receives). Second because modules are installed at a tilt angle close to local latitude to maximize array output (or at some minimum tilt to ensure self-cleaning by rain) this requires to convert the retrieved tilted global solar irradiance to horizontal.

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

4.1. Methodology

1) Estimating the solar radiation received by each PV system from their energy production

The PV modules are widely used as calibrated references providing solar irradiance measurements from short circuit current. There is a relationship between the power output (P) of a PV system and its operating conditions, in particular, the incident solar irradiance (G) and the cell temperature (T_c):

$$P = f(G, T_c)$$

The energy production registered at PV systems can similarly be used as a solar radiation sensor. Solar irradiance has the units of a power density, expressed in W/m², whereas solar radiation has the units of an energy density, expressed in Wh/m². The two concepts are, thus, fundamentally different. Nevertheless, while working with hourly data, and if they are expressed in these units, solar irradiance and solar radiation can be substituted. The relationship between hourly energy output (E_h), hourly incident solar radiation (G_h) and cells temperature is:

$$E_h = f(G_h, T_c)$$

Nevertheless, this relationship is not as straightforward as in the case of calibrated PV modules, because, the power output of a PV system is not directly proportional to the solar irradiance that it receives. Other factors are to be taken into account to establish a relationship between operating conditions and power output, such as the losses related to DC/AC conversion, the losses due to the temperature of the cells, or the losses due to the balance of the system. We have developed a simple model to take these additional second-order effects into account, and thereby to make use of PV systems to obtain solar irradiance data in the same way that calibrated PV modules are used.

The first step is to calibrate these PV systems in some way so that their energy production provides us solar radiation data. This is realized through the determination of a parameter that permits to link these variables between them. We are looking for a parameter ($f_{Calibration}$) that provides the following relationship:

$$G_h = f_{Calibration} \cdot E_h$$

This parameter should as much as possible be in line with the sensitivity of a PV system to the external conditions. The performance ratio (PR) of a PV system is the most frequently used performance indicator in the PV community, and it is defined by:

$$PR = \frac{E_h / G_h}{P_p / G^*}$$

Where P_p is its nominal (or peak) DC power, and G^* is the solar irradiance corresponding to the Standard Test Conditions (STC) under which the peak power of the PV modules was rated. This PR is not constant where the PV system experiences a variation in its operating conditions. Therefore, we need to construct a normalized version of PR to be able to use it as a calibration factor.

The PR of a PV system is generally lower than 1 (or 100%), because of several sources of losses in the PV system. We have lumped these losses into four main categories:

- f_G : PV module's yield as a function of incident irradiance level;
- f_{TEMP} : PV module's yield as a function of cell temperature;
- f_{DCAC} : Yield of the conversion DC/AC;
- f_{BOS} : Yield of the Balance Of System (BOS).

Therefore, if these 4 parameters explain all the difference spreading between the real value of the PR of the PV system and 100%, the following expression stands:

$$\frac{1}{f_G \cdot f_{TEMP} \cdot f_{DCAC} \cdot f_{BOS}} \cdot PR = 1$$

Three of these four losses parameters can be expressed analytically:

$$f_G = f(G, \eta_{PDC, EUR}, T_c, \theta_s, T_\theta)$$

where $\eta_{PDC, EUR}$ is the “European Yield” of the PV modules, i.e., its yield curve under varying G conditions, T_c is the cell temperature of the PV modules, θ_s is the angle of incidence of solar irradiance of the PV modules, and T_θ is the angular transmission function of the PV modules.

T_c can be deduced by:

$$T_c = T_a + G \frac{(NOCT - 20)}{800}$$

where T_a is the air temperature, and $NOCT$ is the Nominal Operation Cells Temperature, which is available from the PV modules manufacturers’ datasheet, and whose typical value is 48 °C.

$$f_{TEMP} = 1 + \gamma(T_c - T_c^*)$$

where γ is the coefficient of variation of PV power with cells temperature, which is available from the PV modules manufacturers’ datasheet, and whose typical value is -0.45%/°C.

$$f_{DCAC} = f(\eta_{INV, EUR}, \frac{P_{AC}}{P_p})$$

where η_{INV} is the yield of a typical inverter, $\eta_{INV, EUR}$ is the “European Yield” of a typical inverter, P_{AC} is the nominal rated power of the inverter, P_p is the nominal rated power of the PV generator.

The fourth parameter, f_{BOS} , cannot be estimated, because the real energetic behavior of each PV system is unknown a priori. Nevertheless, its determination is not compulsive in our case because it is possible to restate the equations as follows:

$$f_{PDC} = \frac{P_p}{G^*}$$

And therefore, we can reformulate equation (5) as:

$$\frac{1}{f_G \cdot f_{TEMP} \cdot f_{DCAC} \cdot f_{BOS} \cdot f_{PDC}} \cdot \frac{E_h}{G_h} = 1$$

This fifth losses factor f_{PDC} has been extracted from PR, and it accounts for the difference between the real nominal DC power of the PV modules under STC conditions, and the nominal power as rated by the PV module manufacturer. As for f_{BOS} factor, this factor cannot be estimated because the real power of the PV systems modules is unknown unless carefully calibrated, which would not be feasible in the present case. But f_{PDC} and f_{BOS} can be lumped together into a new losses factor, called f_{PERF} , accounting for the global variation in performance between the real PV systems and a very high-quality PV system.

$$\frac{1}{f_G \cdot f_{TEMP} \cdot f_{DCAC} \cdot f_{PERF}} \cdot \frac{E_h}{G_h} = 1$$

This f_{PERF} parameter sums up all the performance losses that, on the one hand, could be avoided, and on the other hand, and above all, that cannot be modeled through a simple and general analytical expression. But this parameter can be estimated through the following rearrangement:

$$f_{PERF} = \frac{1}{f_G \cdot f_{TEMP} \cdot f_{DCAC}} \cdot \frac{E_h}{G_h}$$

The f_{PERF} parameter can be estimated for each PV system from historical data of E and G . The E data directly comes from measurements taken by the energy meters, and the G data can be obtained from a solar radiation database, or by other sources, such as the combination of clear-sky models and cloud fraction data, explained later in this text. In this work we have explored both ways to estimate f_{PERF} from G , and we concluded that the results were very similar, which is very good news, because it demonstrates that f_{PERF} is only characteristic on the PV system and its operating conditions and that it does not depend of the solar radiation database that is chosen to estimate it.

Once f_{PERF} has been estimated from a learning process, the relationship between E_h and G_h is straightforward:

$$G_h = \frac{1}{f_G \cdot f_{TEMP} \cdot f_{DCAC} \cdot f_{PERF}} \cdot E_h$$

2) Estimating the tilted solar radiation in any plane $G(\alpha_2, \beta_2)$ from another plane $G(\alpha_1, \beta_1)$

Each PV system is characterized by its orientation from South, noted α , and its tilt angle from the horizontal plane noted β . The solar irradiance that can be deduced from the energy produced by the PV systems is thus their tilted solar radiation data. Nevertheless, we are looking for a method able to generate tilted solar radiation data on any tilted plane, i.e. for any combination of α and β . We, therefore, need to be able to translate any tilted solar irradiance data from one plane to another. We will note $G(\alpha_1, \beta_1)$ the solar irradiance in the plane of the PV system (1) from which it has been estimated. It is our reference plane for the present exercise. We now note $G(\alpha_2, \beta_2)$ the tilted solar irradiance that we desire to obtain, on another plane (2). We now need to find a general relation of the type $G(\alpha_2, \beta_2) = f(G(\alpha_1, \beta_1))$. We apply decomposition models to do so. We have chosen decomposition models that are widely used in the solar energy community, such as the Perez model, or the Hay model. The text that follows explains each step of the model that we use in the case of the Hay model.

To obtain $G(\alpha_2, \beta_2)$ by a decomposition model, we first need to decompose $G(\alpha_1, \beta_1)$ into its direct (or beam), diffuse and reflected components, respectively written as, $D(\alpha_1, \beta_1)$, and $R(\alpha_1, \beta_1)$.

By integration of these solar irradiance data along an entire day, we can calculate the corresponding daily solar radiation data $G_d(\alpha_1, \beta_1)$, as follows:

$$G_d(\alpha_1, \beta_1) = \int_d G(\alpha_1, \beta_1)$$

where the “ d ” indicates “day”.

The Clarity Index $K_{T,d}$ takes the form:

$$K_{T,d} \approx \frac{G_d(\alpha_1, \beta_1)}{B_{0,d}(\alpha_1, \beta_1)}$$

where $B_{0,d}(\alpha_1, \beta_1)$ is the extra-atmospheric daily tilted solar radiation, which is similarly obtained by integration of the extra-atmospheric tilted irradiance:

$$B_{0,d}(\alpha_1, \beta_1) = \int_d B_0(\alpha_1, \beta_1)$$

This extra-atmospheric tilted irradiance can, in turn, be estimated from pure geometrical considerations:

$$B_0(\alpha_1, \beta_1) = B_0(n) \cdot \max(0, \cos(\theta_{s1}))$$

where the angle θ_{sl} is the angle of incidence of the solar irradiance, and $B_0(n)$ is the extra-atmospheric normal irradiance:

$$B_0(n) = B_0 \cdot \varepsilon_0$$

where ε_0 is a correction factor of the Sun-Earth distance that takes into account the eccentricity along the Earth's orbit around the Sun.

Similarly, the Diffuse Factor $F_{D,d}$ takes the form:

$$F_{D,d} \approx \frac{D_d(\alpha_1, \beta_1)}{G_d(\alpha_1, \beta_1)}$$

There is an empirical relationship between $K_{T,d}$ and $F_{D,d}$:

$$F_{D,d} = 0.952 \quad \text{for } K_{T,d} \leq 0.13$$

$$F_{D,d} = 0.868 + 1.335K_{T,d} - 5.782K_{T,d}^2 + 3.721K_{T,d}^3 \quad \text{for } 0.13K < K_{T,d} \leq 0.8$$

$$F_{D,d} = 0.115 \quad \text{for } K_{T,d} > 0.8$$

Therefore, it is now possible to deduce the diffuse component of solar irradiance:

$$D_d(\alpha_1, \beta_1) = F_{D,d} \cdot G_d(\alpha_1, \beta_1)$$

Strictly speaking, the $K_{T,d}$ and $F_{D,d}$ are relationships between horizontal irradiances, and not tilted irradiances as we wrote in these formulas. Nevertheless, it is an approximation that we made in our calculations. We are aware that this approximation can introduce inaccuracies, and we are now working on possible alternatives to this formula.

Previous works have also demonstrated that this relationship stands true:

$$r_D = \frac{D(0)}{D_d(0)} = \frac{B_0(0)}{B_{0,d}(0)}$$

which now makes it possible to determine the hourly diffuse component on the plane 1:

$$D(\alpha_1, \beta_1) = D_d(\alpha_1, \beta_1) \cdot \frac{B_0(\alpha_1, \beta_1)}{B_{0,d}(\alpha_1, \beta_1)}$$

This diffuse component is generally not distributed uniformly over the whole sky. Hay and Davis have proposed a model that assimilates the diffuse radiation as a sum of two distinct and physically opposed kinds, which are the circumsolar (CIR) and isotropic (ISO) components:

$$D(\alpha_1, \beta_1) = D^{CIR}(\alpha_1, \beta_1) + D^{ISO}(\alpha_1, \beta_1)$$

Each of these components can be defined as:

$$D^{ISO}(\alpha_1, \beta_1) = D(0) \cdot (1 - k_a) \frac{1 + \cos(\beta_1)}{2}$$

$$D^{CIR}(\alpha_1, \beta_1) = D(0) \cdot k_a \cdot \frac{\max(0, \cos(\theta_{s1}))}{\cos(\theta_{zs})}$$

where k_a is the anisotropy coefficient, defined as:

$$k_a = \frac{B(n)}{B_0(n)}$$

The reflected component can be estimated from the global solar irradiance as:

$$R(\alpha_1, \beta_1) = \rho G(\alpha_1, \beta_1) \cdot \frac{1 - \cos(\beta_1)}{2}$$

where $\rho = 0.2$ is a typical value for the ground albedo.

And the remaining component, the direct irradiance, can be deduced from a simple subtraction:

$$B(\alpha_1, \beta_1) = G(\alpha_1, \beta_1) - D(\alpha_1, \beta_1) - R(\alpha_1, \beta_1)$$

It is now necessary to transpose all of the three components from plane 1 to plane 2.

The transposition between the direct components is a simple matter of geometry:

$$B(\alpha_2, \beta_2) = B(\alpha_1, \beta_1) \frac{\max(0, \cos(\theta_{s2}))}{\max(0, \cos(\theta_{s1}))}$$

The transposition of the reflected components is also straightforward:

$$R(\alpha_2, \beta_2) = R(\alpha_1, \beta_1) \cdot \frac{1 - \cos(\beta_2)}{1 - \cos(\beta_1)}$$

The transposition of the diffuse components is a little more complex, but it can be achieved through:

$$DTI^{CIR}(\alpha_2, \beta_2) = DTI^{CIR}(\alpha_1, \beta_1) \cdot \frac{\max(0, \cos(\theta_{s1}))}{\max(0, \cos(\theta_{s2}))}$$

$$DTI^{ISO}(\alpha_2, \beta_2) = DTI^{ISO}(\alpha_1, \beta_1) \cdot \frac{(1 + \cos(\beta_1))}{(1 + \cos(\beta_2))}$$

$$D(\alpha_2, \beta_2) = D^{ISO}(\alpha_2, \beta_2) + D^{CIR}(\alpha_2, \beta_2)$$

The combination of the previous formulas can now allow estimating $D(\alpha_2, \beta_2)$ from $D(\alpha_1, \beta_1)$:

$$D(\alpha_2, \beta_2) = D(\alpha_1, \beta_1) \cdot \frac{(1 - k_a) \cdot \frac{1 + \cos(\beta_1)}{2} + k_a \cdot \frac{\max(0, \cos(\theta_{s1}))}{\cos(\theta_{zs})}}{(1 - k_a) \cdot \frac{1 + \cos(\beta_2)}{2} + k_a \cdot \frac{\max(0, \cos(\theta_{s2}))}{\cos(\theta_{zs})}}$$

The global irradiance on plane 2 can now be deduced by addition of its three components:

$$G(\alpha_2, \beta_2) = B(\alpha_2, \beta_2) + D(\alpha_2, \beta_2) + R(\alpha_2, \beta_2)$$

3) Use of Kriging to obtain the solar radiation at a given location from the solar irradiance data available at neighboring locations

We have developed a formula to obtain $G(\alpha_2, \beta_2)$ from the energy production of any PV system. We now need to develop a method that provides $G(\alpha_2, \beta_2)$ at a given location, defined by its latitude and longitude, from the $G(\alpha_2, \beta_2)$ provided by all the available neighboring PV systems. The method belongs to the family of simple kriging.

The first step is to decide which neighboring PV systems are eligible as a valid source of information to obtain the desired solar radiation data. At an hourly level, previous works have demonstrated that the solar radiation that is measured at one location is generally representative of the solar radiation observed at another location if the distance between these two points is lower than 20-30 km. We have therefore decided to exclude the information provided by any PV system distant more than 30 km.

Obviously, the reliability of the solar radiation data provided by a neighboring PV system is higher when it is closer to the location where the data is required. The data provided by each of the neighboring PV systems is, therefore, weighted in consequence, taking into account the relation between uncertainty and distance. This is done through the following formula:

$$w_i = \frac{1}{\sqrt{d_0^2 + d_i^2}}$$

where w_i is the weight of the data provided by the PV system i , noted d_i , is the distance between this PV system and a given location and d_0 is a constant that represents a distance below which there is no substantial improvement of the accuracy of the data. This factor accounts for the “nugget” effect for the kriging interpolation. We have observed that in most cases this nugget distance is of approximately 1 km, which is the value that we have retained.

$$d_0 = 1km$$

All of the weighting factors are normalized before applying any kriging interpolation. This is done through the following formula:

$$\lambda_i = \frac{w_i}{\sum w_i}$$

where λ_i is the relative weighting factor for the PV system i .

We then interpolate the solar radiation data available at each neighboring PV system using a weighted median as the mathematical interpolator function:

$$G_K(\alpha_2, \beta_2) = \mu_{1/2, \lambda} [G_i(\alpha_2, \beta_2), \lambda_i]$$

where $G_K(\alpha_2, \beta_2)$ is the solar radiation data obtained by interpolation from all the neighboring PV systems, and $\mu_{1/2, w}$ represents the weighted median operation.

4.2. Results

A) Application to PV system fleet in Belgium

These procedures to generate solar irradiation data from the PV output data were applied to the 6,000 PV systems monitored in Belgium by Rtone. Globally, these results have shown that this method can be very promising, along with some important limitations. One method that was used to assess the quality of the solar irradiation data that were generated was to generate the solar irradiation data from the PV output data, and then to use these data as an input in PV simulations to deduce the PV power that should have been produced by a neighboring installation. If the data and the models are of good quality, then the PV energy output measured at the neighboring installation and the corresponding PV energy output deduced from the solar irradiation data and the simulation should provide similar numbers, which is what was observed for about one fourth of the PV installations in the database. Figure 55. shows the example of the retrieval of the hourly solar irradiation data for one year on one PV installation in a case that was satisfactory.

Figure 55: Comparison between the PV energy output measured and retrieved from solar irradiation at one PV system in Belgium.

Unfortunately, for the other three fourths, the results were of lesser quality and for one fourth of them the results were of very poor quality. This led to the conclusion that primary to the application of this method, it is necessary to obtain accurate information on the metadata. This can be done through some procedures developed in other works and that have already been discussed in Chapter 3.

B) Comparisons with ground measurements in Belgium

The solar irradiation data derived from the PV output data has also been compared to solar irradiation measurements at weather stations in Belgium. These comparisons have been carried out on the basis of comparisons between the irradiation data derived from the PV power data and the solar irradiation data measurements taken at ground stations maintained by The Royal Meteorological Institute of Belgium (RMI). RMI has a long term experience with ground-based measurements of solar radiation in Belgium (uninterrupted 30 min average measurements in Uccle since 1951, in Oostende since 1958, and in Saint-Hubert since 1959). Uccle is one of the 22 Regional Radiation Centres established within the WMO Regions. The incoming global horizontal solar irradiance is currently measured in 13 Automatic Weather Stations (AWS) in addition to the measurements performed in the main/ reference station in Uccle. At the reference station in Uccle, measurements of the global horizontal solar irradiance are performed by Kipp & Zonen CM11-secondary standard pyranometers. At the 13 RMI's AWS, GHI observations for the year 2014, were made with a Kipp & Zonen CNR1 net radiometer. It consists of a pyranometer (model type CM3 complying with the ISO Second Class Specification) and a pyrgeometer (model type CG3) pair that faces upward and a complementary pair that faces downward. The pyranometers and pyrgeometers measure shortwave and far-infrared radiation, respectively.

All ground measurements are made with a 5-s time step and then integrated to bring them to a 10-min time step. The 10-min data have undergone a series of automated quality control procedures (Journée and Bertrand, 2011) prior to being visually inspected and scrutinized in depth by a human operator for more subtle errors. Because the radiometric station in Saint-Hubert (AWS 6476) was known to be subject to operating problems in 2014, all GHI measurements from this station were discarded. Because the data quality control revealed that the CNR1 net radiometer installed in the Buzenol station (i.e. AWS 6484) has only performed well intermittently during the year 2014, GHI measurements from this station were not used for validation purpose. As a result, 1470 systems have been selected and considered in this study after several preliminary checks. The geographical location of these PV installations and the remaining 13 ground measurement sites is provided in Figure 56, together with the selected residential PV systems. It is worth pointing out that due to availability reasons, most of the data comes from Wallonia and Brussels. The vast majority of the selected PV generators have a tilt angle between 20 and 50, which generally corresponds to the slope of the roofs on which they are installed, as discussed in Chapter 2.

Figure 56: Location of the selected 1470 residential PV systems within the Belgian territory. Red dots indicate the location of the RMI's ground radiometric stations considered in this study.

Statistical error indexes were computed between in-situ hourly irradiance measurements and the estimations computed from the hourly energy production of residential PV systems surrounding the measurement stations. An initial radius of 5 km centered on the station was considered to select the residential PV systems for the validation purpose. When less than 4 PV installations were found within the delimited area, the radius was extended to 10 km. No PV system was found in the vicinity of the Sint-Katelijne-Waver, Retie and Mont-Rigi stations and 3 other stations only have one surrounding residential PV system. At the opposite, the maximum number of installations surrounding a station is 37 for Ernage.

Figure 57 presents an example of daily solar surface irradiation over the Belgian territory as computed from the interpolation of ground measurements using the ordinary kriging (OK) method (Wackernagel, 1995), the MSG satellite-derived estimation and the PV systems method. Clearly, due to the sparsity of the ground stations networks, interpolating ground data generates only a coarse distribution of the solar surface irradiation: large-scale variations of the solar irradiation (such as the south–east to north–west positive gradient) are identified but local fluctuations remain unseen (panel a). Satellite-derived estimates, on the other hand, provide a global coverage and are therefore able to account for clouds induced small-scale variability in surface solar radiation (panel b). Regarding the PV systems method, the daily solar irradiation estimated at the PV clusters level were then interpolated by OK method to cover the entire Belgian territory (panel c).

Figure 57: Comparison between the daily spatial distribution of surface solar global irradiation (in Whm^{-2}) over Belgium as computed (a) by the interpolating ground measurements, (b) by the MSG satellite retrieval method and, (c) by the PV systems power output method. Illustrations are for the 20 June, 2014. Black dots in panel (a) indicate the location of the RMI's in-situ measurements sites. Black dots in panel indicate the location of the 150 clusters of PV systems.

As we see, the southeast to northwest positive gradient is well apparent as well as some of the regional specificities. For instance, the Gaume region (area in the south-east of Belgium) located on the south side of the Ardenne (hilly mass) and that enjoys longer sunshine time appears clearly on the mapping. In general, the PV systems method provides small-scale patterns partly supported by the MSG derived mapping. These results, therefore, show that this method is promising. Nevertheless, while some first results were positive, the quality of the hourly solar irradiation obtained was globally worse with the PV system method than with MSG. Some important limitations have also appeared and they have made it impossible to obtain a clear and generalized result for the moment. In particular, the accuracy of the solar irradiation data was limited by the uncertainty on the metadata, mostly tilt and azimuth angles of the PV generators. As already discussed in Chapter 2 and Chapter 3, these metadata are not known with a high degree of reliability. Indeed, metadata are prone to errors on the part of the owner, by entering incorrect system parameters but also in some cases due to installer's conventions. As an example, tilt angle can be erroneously reported as 0 simply because 0 is used as default value in case of missing information. Another identified limitation is that many angles values are rounded as multiples of 5. This directly translates into important errors in the derived data. Therefore, the results that were obtained were of very variable quality, depending on the quality of the metadata.

The lack of accuracy encountered in the information on the orientation and/or the inclination of the PV installations does not allow to retrieve reliable solar irradiance data. It was found that the provided orientation and tilt angles can typically bear inaccuracies up to 5-10%. However, tilt angle and surfaces orientation have been found to have a large impact on the accuracy of the global horizontal solar irradiance calculation. Increasing the number of PV installations involved in the computation process allows smoothing the estimation to some extent.

Another limitation is that there are certain sun positions for which the method fails to produce a valid estimation. As an example, unsuccessful tilt to horizontal conversions occurs at low solar elevation, irrespective of the number of PV systems involved in the conversion process. Validation results computed on an hourly basis provide a mean RMSE value of about 40% when considering a group of neighboring PV installations in the estimation process while values as large as 60% are reported when using a single installation. By comparison, satellite-based global horizontal irradiance estimation exhibits a better performance with an associated overall RMSE of about 20%. By contrast the overall performance of both methods does not differ significantly on a daily basis. However, spatially the distribution of the solar surface irradiation computed from PV systems outputs presents small-scale pattern artifacts signature of erroneous estimations. Globally, the development of a procedure able to assess the true azimuth and tilt angle of a PV generator from the sole knowledge of its hourly energy output data would be largely beneficial for the method. Towards this objective, it is worth mentioning that while the identification of a PV system's location and orientation from performance data has received limited attention in the past, Williams et al., 2012 have introduced a method for automatic detection of PV system configuration. And even more recently, Haghdadi et al., 2017 have proposed a method to estimate the location and orientation of distributed photovoltaic systems from their generation output data. The conversion of PV system energy production to

tilted global solar irradiance can certainly be improved by taking into account the wind influence on the operating temperature of the solar cell as an example or by performing clear-sky radiative transfer computations with a more sophisticated model, etc. But such refinements will only have a limited influence on the final result since the main limitation is due to erroneous/invalid transposition of the titled global solar irradiance to the horizontal because of inaccurate information about the PV systems.

5. CONCLUSION

The work has attempted to contribute to the development of the state of the art in three specific topics:

- The assessment of the overall performance of the distributed PV systems in Europe;
- The development of a novel fault detection method applied to fleets of distributed PV systems;
- The development of a method makes it possible to deduce solar irradiation that is received on the plane of distributed PV systems from the measurement of their PV energy output.

This work has reviewed the state of the art of residential PV systems in Europe, with a focus on Belgium and France, and in a lesser measure the UK, with some comparisons made with PV plants in Spain.

In Belgium, the PV market in Wallonia and Brussels developed towards residential PV systems as a consequence of limiting the most interesting public financial support to systems of several kW, and from the surface typically available on the roofs. The PV market in France developed towards residential PV systems as a consequence of limiting the most interesting public financial support to systems of a maximum of 3 kW_p.

The PV industry (manufacturers of PV modules and inverters) is dominated by a reduced number of actors, while an important fraction of PV installations is realized by small installers, working at a regional scope.

We have analyzed the data from more than 31,000 PV installations in Europe. The mean Energy Yield of the PV systems located in each of the four reference countries is 1115 kWh/kW_p for France, 898 kWh/kW_p for the UK, 908 kWh/kW_p for Belgium, 1450 kWh/kW_p for the PV plants in Spain mounted on a static structure, and 2127 kWh/kW_p for those mounted on a solar tracker in Spain. As a whole, the orientation of residential PV causes energy productions to be some 6-7% inferior to optimally oriented PV systems. These losses due to orientation are generally low enough to ensure that the PV systems installed on buildings are a viable alternative to solar plants optimally oriented.

The quality of the PV systems is quantified using the Performance Ratio (PR), and the Performance Index (PI). After a mean exposure time of 2 years, the mean value of Performance Ratio is 78% in Belgium and 76% in France, and the mean Performance Index of the PV systems is 85% in both countries, which implies that the typical real PV system produces 15% less than a very high-quality PV system (or reference PV system). On average, the real power of the PV modules falls some 5% below their corresponding nominal power announced on the manufacturer's datasheet. However, some modules show a real power more than 15% below the nominal power announced by their manufacturer. A brief analysis by PV module technology has led to relevant observations about two technologies. On the one hand, the PV

systems equipped with HIT modules show performances higher than average. On the other hand, the systems equipped with the CIS modules show a real power that is 16 % lower than nominal value.

We have shown that the distribution of the yearly integrated PR can be modeled well using a Weibull distribution for PR values ranging from 0.6 to 0.9. This range of values represents the majority of the PV systems, and we suggest that they are representative of the state-of-the-art for PV systems in Europe. We suggest that the typical PR value for the PV systems installed in Europe over the last few years is 0.79 and that the value for the PV systems installed in 2015 is 0.81.

We have observed that the performance of the PV systems tends to increase when the peak power of the PV systems increases, in particular for PV systems whose peak power is lower than 4 kW_p.

We have grouped the PV systems as by inverter manufacturer, and we found significant performance differences, up to 5%, between the best performer and the worst. Similarly, we have grouped the PV systems as a function of the PV module manufacturer, and have found significant performance differences, up to 6%, between the best performer and the worst. We have found significant differences in performance as a function of the PV module/cell technology. We have found an increase in performance in the yearly integrated PR around 3 to 4% over the last seven years, which represents an increase of about 0.5% per year. We have qualitatively observed some performance degradation of the PV systems over the last few years, but we were not able to reliably quantify it without further investigation.

This work has also developed a performance indicator that constitutes an alternative to other performance indicators, such as the PR, for the detection of faults affecting PV installations, and that can be constructed on the sole basis of the energy production data measured at the energy meters. This procedure does not require the input of operating conditions data, such as solar irradiation, air temperature, or wind speed. The resulting performance indicator has been designated as Performance to Peers (*P2P*) because it is based on comparisons between neighboring and similar installations, i.e., *peer* PV systems.

This fault detection procedure has been developed in the context of the automatic fault detection applied to more than 6,000 PV installations in Europe. Faults were detected on more than 500 of these installations during the year 2015. A validation campaign was set up to verify whether the faults that were detected were real and whether the method was able to detect all the major faults hindering the performance of the installations. From the feedback that was received from 30 PV installations owners, the fault detection procedure has produced a proper fault detection on 29 out of the 30 PV systems. This work confirms that it is possible to carry out automatic fault detection procedures on the basis of the energy output comparisons between neighboring installations. The P2P has been shown to be more stable than the PR, in particular in the presence of sub-par metadata, and thus constitutes a more robust performance indicator for fault detection procedures.

This work has also explored the possibility to generate solar irradiation data from the PV energy output data. Some first promising results have been shown, but important limitations have also been revealed, in particular, the high sensitivity to the accuracy of the metadata, mostly on the tilt and azimuth angles of the PV generators.

Over the course of this work, we have identified several promising future lines of research.

Even though we have analyzed more than 31,000 PV systems in Europe, these were mostly located in four countries. We have analyzed many PV modules and inverters manufacturers and technologies, but not all of them. Furthermore, new technologies have recently appeared in the market, and it would be interesting to analyze them in order to update our results. The uncertainties around many of our analyses are important, and further development may enable us to reduce them in order to obtain more accurate and more in-depth conclusions. We still have many gaps to cover to achieve a comprehensive understanding of the state-of-the-art of PV systems in Europe. With this in mind, we welcome any future collaboration that might improve our outlook, whether by completing our database with other countries and technologies or by comparing methods and results. By definition, a review of the state-of-the-art of PV systems requires regular iterations in order to continue to incorporate the latest advances.

This work is among the first to use correlations between neighboring PV systems for the fault detection applied to large PV system fleets. The first results show promising potential, but much remains to be done. In particular, several pathways for future investigation works seem promising.

This P2P method could be combined with some metadata correction procedures (Saint-Drenan et al., 2015; Haghdadi et al., 2017; Killinger et al., 2017) for the translation of the PV output from different orientations to generate virtual peers and improve the whole process (Killinger et al., 2016b). These procedures also revealed to be essential prior to the application of the generation of solar irradiation data from PV energy output data.

The P2P and other performance indicators and methods are not mutually exclusive but rather complementary. Analogously, there is a trend towards merging ground and satellite data to achieve better overall results (Journée and Bertrand, 2010; Journée et al., 2012; Riley et al., 2016). The integration of this novel P2P approach into the already existent fault detection procedures could provide added value to them.

Finally, much remains to be done in terms of proper fault diagnosis. In this regard, it would be convenient to combine the P2P with scatter plots, sometimes also called stamp collections, to apply the P2P to the pattern recognition process that is particular to each type of fault.

CONTRIBUTIONS OF THIS PHD THESIS

Publications in international scientific journals (5)

Jonathan Leloux, Luis Narvarte, Adrien Desportes, David Trebosc, Performance to Peers (P2P): A benchmark approach to fault detections applied to photovoltaic system fleets, *Solar Energy* (In review process).

Jesus Robledo, Jonathan Leloux, Eduardo Lorenzo, Christian A. Gueymard, From video games to solar energy: 3D shading simulation for PV using GPU, *Solar Energy*, 2019 (In press).

Cédric Bertrand, Caroline Housmans, Jonathan Leloux, Michel Journée, Solar irradiation from the energy production of residential PV systems, *Renewable Energy*, Volume 125, 2018, Pages 306-318, ISSN 0960-1481, <https://doi.org/10.1016/j.renene.2018.02.036>.

Jonathan Leloux, Eduardo Lorenzo, Beatriz García-Domingo, Jorge Aguilera, Christian A. Gueymard, A bankable method of assessing the performance of a CPV plant, *Applied Energy*, Volume 118, 2014, Pages 1-11, ISSN 0306-2619, <https://doi.org/10.1016/j.apenergy.2013.12.014>.

Jonathan Leloux, Luis Narvarte, David Trebosc, Review of the performance of residential PV systems in France, *Renewable and Sustainable Energy Reviews*, Volume 16, Issue 2, 2012, Pages 1369-1376, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2011.10.018>.

Jonathan Leloux, Luis Narvarte, David Trebosc, Review of the performance of residential PV systems in Belgium, *Renewable and Sustainable Energy Reviews*, Volume 16, Issue 1, 2012, Pages 178-184, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2011.07.145>.

Presentations at international scientific conferences (21)

Jonathan Leloux, Jamie Taylor, Rodrigo Moretón, Luis Narvarte, David Trebosc, Adrien Desportes, Monitoring 30,000 PV systems in Europe: Performance, Faults, and State of the Art, 31st European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, 2015.

Jamie Taylor, Jonathan Leloux, Lisa M.H. Hall, Aldous M. Everard, Julian Briggs, Alastair Buckley, Performance of distributed PV in the UK: a statistical analysis of over 7000 systems, 31st European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, 2015.

Nikolay Tyutyundzhiev, Miroslav Petrov, Francisco Martínez-Moreno, Jonathan Leloux, Luis Narvarte, Advanced testing of PV plants using portable SCADA tool, 31st European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, 2015.

Nikolay Tyutyundzhiev, Konstantin Lovchinov, Francisco Martínez-Moreno, Jonathan Leloux, Luis Narvarte, Advanced PV modules inspection using multirotor UAV, 31st European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, 2015.

Jamie Taylor, Jonathan Leloux, Aldous M Everard, Julian Briggs, Alastair Buckley, Monitoring thousands of distributed PV systems in the UK: Energy production and performance, PV SAT conference, Leeds, 2015.

Jamie Taylor, Jonathan Leloux, Aldous M Everard, Julian Briggs, Alastair Buckley, Quantifying PV fleet output variability in the UK: Consequences for Distribution Network Operators, PV SAT conference, Leeds, 2015.

Nikolay Tyutyundzhiev, Francisco Martínez-Moreno, Jonathan Leloux, Luis Narvarte, Equipment and procedures for on-site testing of PV plants and BIPV, 29th European Photovoltaic Solar Energy Conference and Exhibition, Amsterdam, 2014.

Rodrigo Moretón, Eduardo Lorenzo, Jonathan Leloux, José Manuel Carrillo, dealing in practice with hot-spots, 29th European Photovoltaic Solar Energy Conference and Exhibition, Amsterdam, 2014.

Jonathan Leloux, Luis Narvarte, Loreto González, A free real-time hourly tilted solar irradiation data Website for Europe, 29th European Photovoltaic Solar Energy Conference and Exhibition, Amsterdam, 2014.

Jesús Robledo, Jonathan Leloux, Eduardo Lorenzo, 3D simulation of complex shading affecting PV systems taking benefit from the power of graphics cards developed for the video game industry, 29th European Photovoltaic Solar Energy Conference and Exhibition, Amsterdam, 2014.

Luis Narvarte, Alberto Luna, Adrien Desportes, Automatic fault detection on BIPV systems without solar irradiation data, Jonathan Leloux, 29th European Photovoltaic Solar Energy Conference and Exhibition, Amsterdam, 2014.

Jonathan Leloux, BIPV performance diagnosis, Intersolar Conference, Munich, 2014.

Jonathan Leloux, Eduardo Lorenzo, Beatriz García-Domingo, Jorge Aguilera, A bankable method for the field testing of a CPV plant, 28th European Photovoltaic Solar Energy Conference and Exhibition, Paris, 2013.

Jonathan Leloux, Luis Narvarte, Alberto Luna, Adrien Desportes, Automatic detection of PV systems failures from monitoring validated on 10,000 BIPV systems in Europe, 28th European Photovoltaic Solar Energy Conference and Exhibition, Paris, 2013.

Jonathan Leloux, Luis Narvarte, David Trebosc, Detección automática de fallos de operación en instalaciones fotovoltaicas domésticas, XV Congreso Ibérico y X Iberoamericano de Energía Solar, Vigo, 2012.

Jonathan Leloux, Luis Narvarte, David Trebosc, Performance Analysis of 10,000 Residential PV Systems in France and Belgium, 26th European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, 2011.

Jonathan Leloux, Dolores Pachón, Gabriel Sala, Spectral solar radiation measurement and simulation for CPV modules production estimation, 6th International Conference on Concentrating Photovoltaic Systems, Freiburg, 2010.

Jonathan Leloux, Gabriel Sala, Towards the consolidation of a photovoltaic observatory in Wallonia and Brussels (Belgium), 24th European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, 2009.

Víctor Rico, Alberto Martín, Francesca Rubio, Jonathan Leloux, Dolores Pachón, Gerald Juengst, Tobias Gerstmaier, First Results of the Spanish National Project “Espectro Solar”, Stefan Mau, Yolanda Sanchez, 24th European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, 2009.

Gabriel Sala, Dolores Pachón, Jonathan Leloux, M. Victoria, A. W. Bett, P. Banda, A. Gombert, V. Diaz, M. A. Vazquez, A. El Moussaoui, F. A. Abulfotuh NACIR: A European Initiative Dedicated to Cooperation with Mediterranean Partner Countries in the Field of Photovoltaic Concentration, 24th European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, 2009.

Gabriel Sala, Dolores Pachón, Jonathan Leloux, Marta Victoria, Andreas Bett, Pedro Banda, Andreas Gombert, Vicente Diaz, Miguel Angel Vazquez, Abdelhakim El Moussaoui, Fuad Ahmed Abulfotuh, NACIR: A New Project on CPV's Funded by the European Commission Under 7th FP, 34th IEEE Photovoltaic Specialists Conference, Philadelphia, 2009.

Communications in national trade journals (3)

Jonathan Leloux, Luis Narvarte, Alberto Luna, Adrien Desportes, Detecção de falhas em sistemas integrados a edificações, Eletricidade Moderna, Brazil, 2014.

Jonathan Leloux, Luis Narvarte, David Trebosc, , Instalaciones fotovoltaicas domésticas: Detección automática de fallos de operación, Era Solar, Spain, 2014.

Jonathan Leloux, Analyse de la performance de 1000 systèmes photovoltaïques en Belgique en 2010, Renouvelable, Webmag n°37, Belgium, 2011.

Patents (3)

Jonathan Leloux, Luis Narvarte, Eduardo Lorenzo, Rodrigo Moretón, Método de detección automática y diagnóstico de fallos de operación en instalaciones solares fotovoltaicas distribuidas, basado en la comparación de sus producciones de energía, Spanish Patent Office, 2014.

Jonathan Leloux, Luis Narvarte, Rodrigo Moretón, Eduardo Lorenzo, Método de generación de datos de irradiación solar a partir de datos de producción energética de instalaciones solares fotovoltaicas, Spanish Patent Office, 2014.

Jesús Robledo, Jonathan Leloux, Eduardo Lorenzo, Procedimiento de evaluación de pérdidas energéticas provocadas por sombras proyectadas sobre paneles solares fotovoltaicos, Spanish Patent Office, 2014.

Business creation (1)

WebPV was created in 2013 in the context of the European FP7 Research Project PV CROPS, as a startup and spin-off company of the Instituto de Energía Solar - Universidad Politécnica de Madrid, to provide specialized automatic fault detection and diagnosis services to the solar energy industry (PV plants, distributed PV, BIPV).

Awards (2)

Best runner-up paper presented at PVSAT 2015

We received the award "Best runner-up paper presented at PVSAT 2015" for our publication entitled "Monitoring thousands of distributed PV systems in the UK: Energy production and performance".

Best Business Plan ActuaUPM 2012

After two rounds in a contest with 474 ideas presented, the Panel has selected WebPV as the best idea for a startup company, with a prize of €15,000 from UPM in order to set-up the company.

REFERENCES

- Almonacid F., Rus C., Pérez-Higueras P., Hontoria L., Calculation of the energy provided by a PV generator. Comparative study: Conventional methods vs. artificial neural networks, *Energy*, Volume 36, Issue 1, January 2011, Pages 375-384, ISSN 0360-5442, <http://dx.doi.org/10.1016/j.energy.2010.10.028>.
- Anis W.R., Nour M.A.S., Energy losses in photovoltaic systems, *Energy Conversion and Management*, Volume 36, Issue 11, 1995, Pages 1107-1113, ISSN 0196-8904, [https://doi.org/10.1016/0196-8904\(94\)00085-E](https://doi.org/10.1016/0196-8904(94)00085-E).
- Baltus C.W.A., Eikelboom J.A., Van Zolingen R.J.C., Analytical monitoring of losses in PV systems, 14th European Photovoltaic Solar Energy Conference, 1997.
- Behnke, Erdman, and Whitaker Engineering, Inc., Commonwealth Energy Biogas/PV Mini-Grid, Renewable Resources Program, Interim Exposure and Operations Report for the Comprehensive Large PV System Comparison, 2005.
- Blaesser G., PV system measurements and monitoring the European experience, *Solar Energy Materials and Solar Cells*, Volume 47, Issues 1–4, October 1997, Pages 167-176, ISSN 0927-0248, [http://dx.doi.org/10.1016/S0927-0248\(97\)80008-6](http://dx.doi.org/10.1016/S0927-0248(97)80008-6).
IEA PVPS, Task 1, Strategic PV analysis and Outreach, Report IEA PVPS T1-35:2019, Snapshot of global PV markets.
- Blanc P., Gschwind B., Lefèvre M., Wald L., Remote Sensing, 2072-4292, 3, 2, 343-361, The HelioClim Project: Surface Solar Irradiance Data for Climate Applications, 2011, doi:10.3390/rs3020343.
- Bletterie B., Sensitivity of photovoltaic inverters to voltage sags — Test results for a set of commercial products, *e & i Elektrotechnik und Informationstechnik*, Volume 122, Number 10, 349, 2007.
- Bonsignore L., Davarifar M., Rabhi A., Tina G. M., Elhajjaji A., Neuro-Fuzzy Fault Detection Method for Photovoltaic Systems, *Energy Procedia*, Volume 62, 2014, Pages 431-441, ISSN 1876-6102, <http://dx.doi.org/10.1016/j.egypro.2014.12.405>.
- Brecl K., Topič M., Apparent performance ratio of photovoltaic systems - A methodology for evaluation of photovoltaic systems across a region, *Journal of Renewable and Sustainable Energy*, <http://dx.doi.org/10.1063/1.4955088>, 2016.
- Bressan M., El Basri Y., Galeano A. G., Alonso C., A shadow fault detection method based on the standard error analysis of I-V curves, *Renewable Energy*, Volume 99, December 2016, Pages 1181-1190, ISSN 0960-1481, <http://dx.doi.org/10.1016/j.renene.2016.08.028>.
- Caamaño E. et al., Quality control of wide collections of PV modules: lessons learned from the IES experience, *Progress in Photovoltaics: research and applications* 7, 137-149, 1999.
- Capozzoli A., Lauro F., Khan I., Fault detection analysis using data mining techniques for a cluster of smart office buildings, *Expert Systems with Applications*, Volume 42, Issue 9, 1 June 2015, Pages 4324-4338, ISSN 0957-4174, <http://dx.doi.org/10.1016/j.eswa.2015.01.010>.
- Chao K. H., Ho S. H., Wang M. H., Modeling and fault diagnosis of a photovoltaic system, *Electric Power Systems Research*, Volume 78, Issue 1, January 2008, Pages 97-105, ISSN 0378-7796, <http://dx.doi.org/10.1016/j.epsr.2006.12.012>.
- Charles J.P., Hannane F., El-Mossaoui H., Zegaoui A., Nguyen T.V., Petit P., Aillerie M., Faulty PV panel identification using the Design of Experiments (DoE) method, *International Journal of Electrical Power & Energy Systems*, Volume 57, May 2014, Pages 31-38, ISSN 0142-0615, <http://dx.doi.org/10.1016/j.ijepes.2013.11.037>.
- Chilès J.P., Delfiner P., *Geostatistics: Modeling Spatial Uncertainty*, John Wiley & Sons, Sep 25, 2009.
- Chine W., Mellit A., Pavan A.M., Kalogirou S.A., Fault detection method for grid-connected photovoltaic plants, *Renewable Energy*, Volume 66, June 2014, Pages 99-110, ISSN 0960-1481, <http://dx.doi.org/10.1016/j.renene.2013.11.073>.

Chouder A., Silvestre S., Automatic supervision and fault detection of PV systems based on power losses analysis, *Energy Conversion and Management*, Volume 51, Issue 10, October 2010, Pages 1929-1937, ISSN 0196-8904, <http://dx.doi.org/10.1016/j.enconman.2010.02.025>.

Chouder A., Silvestre S., Sadaoui N., Rahmani L., Modeling and simulation of a grid connected PV system based on the evaluation of main PV module parameters, *Simulation Modelling Practice and Theory*, Volume 20, Issue 1, January 2012, Pages 46-58, ISSN 1569-190X, <http://dx.doi.org/10.1016/j.simpat.2011.08.011>.

Chouder A., Silvestre S., Taghezouit B., Karatepe E., Monitoring, modelling and simulation of PV systems using LabVIEW, *Solar Energy*, Volume 91, May 2013, Pages 337-349, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2012.09.016>.

Christian N.J. et al, Influence of spectral effects on the performance of multijunction amorphous silicon cells, *PV Conference and Exhibition*, Rome, 2002.

Coello J. at al., Quality control during the supply of PV modules: Fundamental key to guarantee the profitability of PV installations, *Proceedings 23rd European Photovoltaic Solar Energy Conference*, 2008.

Colantuono G., Everard A., Hall L.M.H., Buckley A.R., Monitoring nationwide ensembles of PV generators: Limitations and uncertainties. The case of the UK, *Solar Energy*, 2014.

Collares-Pereira M., Rabl A., The average distribution of solar radiation-correlations between diffuse and hemispherical and between daily and hourly insolation values, *Solar Energy*, Volume 22, Issue 2, 1979, Pages 155-164, ISSN 0038-092X, [http://dx.doi.org/10.1016/0038-092X\(79\)90100-2](http://dx.doi.org/10.1016/0038-092X(79)90100-2).

Daliento S., Chouder A., Guerriero P., Monitoring, Diagnosis, and Power Forecasting for Photovoltaic Fields: A Review, *International Journal of Photoenergy*, vol. 2017, Article ID 1356851, 13 pages, 2017. doi:10.1155/2017/1356851.

Decker B., Jahn U., Performance of 170 grid connected PV plants in Northern Germany—Analysis of yields and optimization potentials, *Solar Energy*, Volume 59, Issue 4, Pages 127-133, ISSN 0038-092X, [http://dx.doi.org/10.1016/S0038-092X\(96\)00132-6](http://dx.doi.org/10.1016/S0038-092X(96)00132-6), 1997.

De Miguel A., Bilbao J., Aguiar R., Kambezidis H., Negro E., Diffuse solar irradiation model evaluation in the North Mediterranean Belt area, *Solar Energy*, Volume 70, Issue 2, 2001, Pages 143-153, ISSN 0038-092X, [http://dx.doi.org/10.1016/S0038-092X\(00\)00135-3](http://dx.doi.org/10.1016/S0038-092X(00)00135-3).

Dhimish M., Holmes V., Fault detection algorithm for grid-connected photovoltaic plants, *Solar Energy*, Volume 137, 1 November 2016, Pages 236-245, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2016.08.021>.

Dhimish M., Holmes V., Dales M., Parallel fault detection algorithm for grid-connected photovoltaic plants, *Renewable Energy*, Volume 113, 2017, Pages 94-111, ISSN 0960-1481, <https://doi.org/10.1016/j.renene.2017.05.084>.

Díez-Mediavilla M., Dieste-Velasco M.I., Rodríguez-Amigo M.C., García-Calderón T., Alonso-Tristán C., Performance of grid-tied PV facilities: A case study based on real data, *Energy Conversion and Management*, Volume 76, December 2013, Pages 893-898, ISSN 0196-8904, <http://dx.doi.org/10.1016/j.enconman.2013.08.035>.

Dincer F., The analysis on photovoltaic electricity generation status, potential and policies of the leading countries in solar energy, *Renewable and Sustainable Energy Reviews*, Volume 15, Issue 1, January 2011, Pages 713-720, ISSN 1364-0321, DOI: 10.1016/j.rser.2010.09.026.

Drews A., De Keizer A.C., Beyer H.G., Lorenz E., Betcke J., Van Sark W.G.J.H.M., Heydenreich W., Wiemken E., Stettler S., Toggweiler P., Bofinger S., Schneider M., Heilscher G., Heinemann D., Monitoring and remote failure detection of grid-connected PV systems based on satellite observations, *Solar Energy*, Volume 81, Issue 4, April 2007, Pages 548-564, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2006.06.019>.

Drews A., Beyer H.G., Rindelhadt U., Quality of performance assessment of PV plants based on irradiation maps, *Solar Energy*, Volume 82, Issue 11, November 2008, Pages 1067-1075, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2008.04.009>.

Edgeworth F.Y., On observations relating to several quantities, *Hermathena* 6, no. 13, 1887: 279-85. <http://www.jstor.org/stable/23036355>.

Edgeworth F.Y., On a new method of reducing observations relating to several quantities, *Philosophical Magazine Series* 5, 25:154, 184-191, 1888, DOI: 10.1080/14786448808628170.

Elkholy A., Fahmy F.H., El-Ela A.A., Nafeh A.E.A., Spea S.R., Experimental evaluation of 8 kW grid-connected photovoltaic system in Egypt, *Journal of Electrical Systems and Information Technology*, Volume 3, Issue 2, September 2016, Pages 217-229, ISSN 2314-7172, <http://dx.doi.org/10.1016/j.jesit.2015.10.004>.

Elsinga B., Van Sark W.G.J.H.M., Inter-system time lag due to clouds in an urban PV ensemble, 2014 IEEE 40th Photovoltaic Specialist Conference (PVSC), Denver, CO, 2014, pp. 0754-0758, doi: 10.1109/PVSC.2014.6925029.

Elsinga B., Van Sark W., Spatial power fluctuation correlations in urban rooftop photovoltaic systems. *Prog. Photovolt: Res. Appl.*, 23: 1390–1397. doi: 10.1002/pip.2539, 2015.

Elsinga B., van Sark W., Ramakers L., Inverse photovoltaic yield model for global horizontal irradiance reconstruction; vol. doi: 10.1002/ese3.162. 2017.

Eltawil M.A., Zhao Z., Grid-connected photovoltaic power systems: Technical and potential problems—A review, *Renewable and Sustainable Energy Reviews*, Volume 14, Issue 1, January 2010, Pages 112-129, ISSN 1364-0321, <http://dx.doi.org/10.1016/j.rser.2009.07.015>.

Engerer N.A., Mills F.P., KPV: A clear-sky index for photovoltaics, *Solar Energy*, Volume 105, July 2014, Pages 679-693, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2014.04.019>.

Erbs D., Klein S., Duffie J., Estimation of the diffuse radiation fraction for hourly, daily and monthly-averaged global radiation. *Solar Energy* 1982; 28:293–302.

Erge T., Hoffmann V.U., Kiefer K., The German experience with grid-connected PV-systems, *Solar Energy*, Volume 70, Issue 6, 2001, Pages 479-487, ISSN 0038-092X, [https://doi.org/10.1016/S0038-092X\(00\)00143-2](https://doi.org/10.1016/S0038-092X(00)00143-2).

Essah E.A., Rodriguez A., Glover N., Assessing the performance of a building integrated BP c-Si PV system, *Renewable Energy*, Volume 73, January 2015, Pages 36-45, ISSN 0960-1481, <http://dx.doi.org/10.1016/j.renene.2014.04.002>.

Firman A., Toranzos V., Vera L., Busso A., De la Casa J., Passive Monitoring of the Power Generated in Grid Connected PV Systems., *Energy Procedia*, Volume 57, 2014, Pages 235-244, ISSN 1876-6102, <http://dx.doi.org/10.1016/j.egypro.2014.10.028>.

Firth S.K., Lomas K.J., Rees S.J., A simple model of PV system performance and its use in fault detection, *Solar Energy*, Volume 84, Issue 4, April 2010, Pages 624-635, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2009.08.004>.

Forero N., Hernández J., Gordillo G., Development of a monitoring system for a PV solar plant, *Energy Conversion and Management*, Volume 47, Issues 15–16, September 2006, Pages 2329-2336, ISSN 0196-8904, <http://dx.doi.org/10.1016/j.enconman.2005.11.012>.

Fraunhofer-Institut Für Solare Energiesysteme ISE, Annual Report, Achievements and Results, 2007.

Fuentes M., Vivar M., Burgos J.M., Aguilera J., Vacas J.A., Design of an accurate, low-cost autonomous data logger for PV system monitoring using Arduino™ that complies with IEC standards, *Solar Energy Materials and Solar Cells*, Volume 130, November 2014, Pages 529-543, ISSN 0927-0248, <http://dx.doi.org/10.1016/j.solmat.2014.08.008>.

Gallardo-Saavedra S., Hernández-Callejo L., Duque-Pérez O., Quantitative failure rates and modes analysis in photovoltaic plants, *Energy*, Volume 183, 2019, Pages 825-836, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2019.06.185>.

Gaiddon B., Fraisse S., Saade A., Stettler S., Toggweiler P., Schilter C., Remund J., Monitoring of large groups of PV systems with GIS, 25th European Photovoltaic Solar Energy Conference and Exhibition, 2010.

Garoudja E., Harrou F., Sun Y., Kara K., Chouder A., Silvestre S., Statistical fault detection in photovoltaic systems, *Solar Energy*, Volume 150, 2017, Pages 485-499, ISSN 0038-092X, <https://doi.org/10.1016/j.solener.2017.04.043>.

Gauss C.F., Bestimmung der Genauigkeit der Beobachtungen, *Zeitschrift für Astronomie und verwandte Wissenschaften*, 1: 187–197, 1816.

Gokmen N., Karatepe E., Celik B., Silvestre S., Simple diagnostic approach for determining of faulted PV modules in string based PV arrays, *Solar Energy*, Volume 86, Issue 11, November 2012, Pages 3364-3377, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2012.09.007>.

Gokmen N., Karatepe E., Silvestre S., Celik B., Ortega P., An efficient fault diagnosis method for PV systems based on operating voltage-window, *Energy Conversion and Management*, Volume 73, September 2013, Pages 350-360, ISSN 0196-8904, <http://dx.doi.org/10.1016/j.enconman.2013.05.015>.

Golnas A., Bryan J., Wimbrow R., Hansen C., Voss S., Performance assessment without pyranometers: Predicting energy output based on historical correlation, in *Photovoltaic Specialists Conference (PVSC)*, 2011 37th IEEE , vol., no., pp.002006-002010, 19-24 June 2011, doi: 10.1109/PVSC.2011.6186347.

Hammond R. et al., Effects of soiling on PV module and radiometer performance, *AZ Photovoltaic Specialists Conference*, 1997.

Hay J.E., Calculation of monthly mean solar radiation for horizontal and inclined surfaces, *Solar Energy*, Volume 23, Issue 4, 1979, Pages 301-307, ISSN 0038-092X, [http://dx.doi.org/10.1016/0038-092X\(79\)90123-3](http://dx.doi.org/10.1016/0038-092X(79)90123-3).

Hay J.E., Study of shortwave radiation on non-horizontal surfaces. Rep. No. 79-12, Atmospheric Environment Service: Downsview, Ontario; 1979.

Hay J.E., Calculating solar radiation for inclined surfaces: Practical approaches, *Solar Energy*, Elsevier, 2003.

Hachana O., Tina G.M., Hemsas K.E., PV array fault Diagnostic Technique for BIPV systems, *Energy and Buildings*, Volume 126, 15 August 2016, Pages 263-274, ISSN 0378-7788, <http://dx.doi.org/10.1016/j.enbuild.2016.05.031>.

Haghdadi N., Copper J., Bruce A., MacGill I., A method to estimate the location and orientation of distributed photovoltaic systems from their generation output data, *Renewable Energy*, Volume 108, 2017, Pages 390-400, ISSN 0960-1481, <https://doi.org/10.1016/j.renene.2017.02.080>.

Herteleer B., Huyck B., Catthoor F., Driesen J., Cappelle J., Normalised efficiency of photovoltaic systems: Going beyond the performance ratio, *Solar Energy*, Volume 157, 2017, Pages 408-418, ISSN 0038-092X, <https://doi.org/10.1016/j.solener.2017.08.037>.

Hosseinzadeh M., Salmasi F.R., Determination of maximum solar power under shading and converter faults—A prerequisite for failure-tolerant power management systems, *Simulation Modelling Practice and Theory*, Volume 62, March 2016, Pages 14-30, ISSN 1569-190X, <http://dx.doi.org/10.1016/j.simpat.2016.01.011>.

Hu Y., Gao B., Song X., Yun Tian G., Li K., He X., Photovoltaic fault detection using a parameter based model, *Solar Energy*, Volume 96, October 2013, Pages 96-102, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2013.07.004>.

Huld T., Gottschalg R., Beyer H.G., Topič M., Mapping the performance of PV modules, effects of module type and data averaging, *Solar Energy*, Volume 84, Issue 2, February 2010, Pages 324-338, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2009.12.002>.

Hussin M.Z., Hasliza N., Yaacob A., Zain Z.M., Omar A.M., Shaari S., A development and challenges of grid-connected photovoltaic system in Malaysia, *IEEE Control and System Graduate Research Colloquium*, Shah Alam, Selangor, 2012, pp. 191-196. doi: 10.1109/ICSGRC.2012.6287160.

IEA, PVPS Task 13: performance and reliability of photovoltaic systems (<http://www.iea-pvps.org/index.php?id=57>).

Isermann R., Supervision, fault-detection and fault-diagnosis methods — An introduction, Control Engineering Practice, Volume 5, Issue 5, May 1997, Pages 639-652, ISSN 0967-0661, [http://dx.doi.org/10.1016/S0967-0661\(97\)00046-4](http://dx.doi.org/10.1016/S0967-0661(97)00046-4).

Jahn, U., Nasse, W., Operational performance of grid-connected PV systems on buildings in Germany. Progress in Photovoltaics: Research and Applications, 2004, 12: 441–448. doi: 10.1002/pip.550.

Jordan D. C., Kurtz S. R., Reliability and Geographic Trends of 50,000 Photovoltaic Systems in the USA, 29th European PV Solar Energy Conference and Exhibition, doi:10.4229/EUPVSEC20142014-5AO.7.5, 2014.

Jordan D. C., Kurtz S. R., VanSant K., Newmiller J., Compendium of photovoltaic degradation rates, Progress in Photovoltaics, 24: 978–989, 2016, doi: 10.1002/pip.2744.

Journée M., Bertrand C., Improving the spatio-temporal distribution of surface solar radiation data by merging ground and satellite measurements, Remote Sensing of Environment, Volume 114, Issue 11, 15 November 2010, Pages 2692-2704, ISSN 0034-4257, <http://dx.doi.org/10.1016/j.rse.2010.06.010>.

Journée M., Bertrand C., Quality control of solar radiation data within the rmib solar measurements network. Solar Energy 2011; 85:72–86.

Journée M., Müller R., Bertrand C., Solar resource assessment in the Benelux by merging Meteosat-derived climate data and ground measurements, Solar Energy, Volume 86, Issue 12, December 2012, Pages 3561-3574, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2012.06.023>.

Kausika B.B., Moraitis P., Van Sark W.G.J.H.M., 2018, Energies, 1996-1073, 11, 6, 1330, Visualization of Operational Performance of Grid-Connected PV Systems in Selected European Countries doi:10.3390/en11061330.

Kiefer K., Reich N.H., Dirnberger D., Reise C., Quality assurance of large scale PV power plants, Photovoltaic Specialists Conference (PVSC), 2011 37th IEEE , 2011, doi: 10.1109/PVSC.2011.6186343.

Killinger S., Müller B., Saint-Drenan Y. M., McKenna R., Towards an improved nowcasting method by evaluating power profiles of PV systems to detect apparently atypical behavior, IEEE 43rd Photovoltaic Specialists Conference (PVSC), 2016.

Killinger S., Braam F., Müller B., Wille-Haussmann B., McKenna R., Projection of power generation between differently-oriented PV systems, Solar Energy, Volume 136, 15 October 2016, Pages 153-165, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2016.06.075>.

Killinger S., Engerer N., Müller B., QCPV: A quality control algorithm for distributed photovoltaic array power output, Solar Energy, Volume 143, February 2017, Pages 120-131, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2016.12.053>.

Killinger S., Lingfors D., Saint-Drenan Y.M., Moraitis P., Van Sark W.G.J.H.M., Taylor J., Engerer N.A., Bright J.M., On the search for representative characteristics of PV systems: Data collection and analysis of PV system azimuth, tilt, capacity, yield and shading, Solar Energy, Volume 173, 2018, Pages 1087-1106, ISSN 0038-092X, <https://doi.org/10.1016/j.solener.2018.08.051>.

Koubli E., Palmer D., Rowley P., Gottschalg R., Inference of missing data in photovoltaic monitoring datasets, IET Renewable Power Generation, 2016, DOI: 10.1049/iet-rpg.2015.0355.

Leloux J., Narvarte L., Trebosc D., Review of the performance of residential PV systems in Belgium, Renewable and Sustainable Energy Reviews, Volume 16, Issue 1, January 2012, Pages 178-184, ISSN 1364-0321, <http://dx.doi.org/10.1016/j.rser.2011.07.145>.

Leloux J., Narvarte L., Trebosc D., Review of the performance of residential PV systems in France, Renewable and Sustainable Energy Reviews, Volume 16, Issue 2, February 2012, Pages 1369-1376, ISSN 1364-0321, <http://dx.doi.org/10.1016/j.rser.2011.10.018>.

Leloux J., Narvarte L., Luna A., Desportes A., Automatic fault detection on BIPV systems without solar irradiation data, 29th European Photovoltaic Solar Energy Conference and Exhibition, 2014.

Leloux J., Taylor J., Moretón R., Narvarte L., Trebosc D., Desportes A., Monitoring 30,000 PV systems in Europe: Performance, Faults, and State of the Art, 31st European Photovoltaic Solar Energy Conference and Exhibition, 2015.

Liu B., Jordan R.C., The interrelationship and characteristic distribution of direct, diffuse, and total solar radiation, *Solar Energy* 1960;4:1–9.

Livera A., Theristis M., Makrides G., Georghiou G.E., Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems, *Renewable Energy*, Volume 133, 2019, Pages 126-143, ISSN 0960-1481, <https://doi.org/10.1016/j.renene.2018.09.101>.

Lonij V.P.A., Jayadevan V.T., Brooks A.E, Rodriguez J.J., Koch K., Leuthold M., Cronin A.D., Forecasts of PV power output using power measurements of 80 residential PV installs, in Photovoltaic Specialists Conference (PVSC), 2012 38th IEEE , vol., no., pp.003300-003305, 3-8 June 2012, doi: 10.1109/PVSC.2012.6318280.

Madeti S. R., Singh S.N., Monitoring system for photovoltaic plants: A review, *Renewable and Sustainable Energy Reviews*, Volume 67, January 2017, Pages 1180-1207, ISSN 1364-0321, <http://dx.doi.org/10.1016/j.rser.2016.09.088>.

Marion B., A method for modeling the current–voltage curve of a PV module for outdoor conditions, *Progress in Photovoltaics: Research and Applications*, Volume 10, Issue 3, pages 205–214, 2002.

Martin N., Ruiz J.M., Annual angular reflection losses in PV modules, *Progress in Photovoltaics*, 2006.

Martínez-Moreno F. et al., Testing of PV arrays in large grid-connected PV plants installed in Spain, 23rd EUPVSEC, 2008.

Martinez-Moreno F., Munoz J., Lorenzo E., Experimental model to estimate shading losses on PV arrays, *Solar Energy Materials and Solar Cells*, Volume 94, Issue 12, December 2010, Pages 2298-2303, ISSN 0927-0248, DOI: 10.1016/j.solmat.2010.07.029.

Martínez-Moreno F., Lorenzo E., Muñoz J., Moretón, R., On the testing of large PV arrays. *Progress in Photovoltaics: Research and Applications*, 2011, 19: n/a. doi: 10.1002/pip.1102.

Mellit A., Kalogirou S.A., Artificial intelligence techniques for photovoltaic applications: A review, *Progress in Energy and Combustion Science*, Volume 34, Issue 5, October 2008, Pages 574-632, ISSN 0360-1285, <http://dx.doi.org/10.1016/j.pecs.2008.01.001>.

Mellit A., Pavan A.M., Performance prediction of 20 kWp grid-connected photovoltaic plant at Trieste (Italy) using artificial neural network, *Energy Conversion and Management*, Volume 51, Issue 12, December 2010, Pages 2431-2441, ISSN 0196-8904, <http://dx.doi.org/10.1016/j.enconman.2010.05.007>.

Mukai T., Kawamoto S., Ueda Y., Saijo M., Abe N., Residential PV system users' perception of profitability, reliability, and failure risk: An empirical survey in a local Japanese municipality, *Energy Policy*, Volume 39, Issue 9, September 2011, Pages 5440-5448, ISSN 0301-4215, <http://dx.doi.org/10.1016/j.enpol.2011.05.019>.

Muneer T., *Solar Radiation and Daylight Models*, Elsevier, 2004.

Muñoz J., Martínez-Moreno F. and Lorenzo E., On-site characterisation and energy efficiency of grid-connected PV inverters. *Progress in Photovoltaics: Research and Applications*, 2011, 19: 192–201. doi: 10.1002/pip.997.

Munro D.K., Blaesser G., The performance of PV systems and components in the THERMIE programme, *Renewable Energy*, Volume 5, Issues 1-4, Climate change Energy and the environment, August 1994, Pages 172-178, ISSN 0960-1481, DOI: 10.1016/0960-1481(94)90368-9.

Myong S.Y., Park Y.C., Jeon S.W., Performance of Si-based PV rooftop systems operated under distinct four seasons, *Renewable Energy*, Volume 81, September 2015, Pages 482-489, ISSN 0960-1481, <http://dx.doi.org/10.1016/j.renene.2015.03.055>.

- Oozeki T., Yamada T., Otani K., Takashima T., Kato K., Performance trends in grid-connected photovoltaic systems for public and industrial use in Japan, *Progress in Photovoltaics*, 18: 596–602, doi:10.1002/pip.992, 2010.
- Pelland S., Abboud I., Comparing Photovoltaic Capacity Value Metrics: A Case Study for the City of Toronto, *Progress in Photovoltaics*, 16: 715–724. doi:10.1002/pip.864, 2008.
- Perez R., Kmiecik M., Herig C., Renné D., Remote monitoring of PV performance using geostationary satellites, *Solar Energy*, Volume 71, Issue 4, 2001, Pages 255-261, ISSN 0038-092X, [http://dx.doi.org/10.1016/S0038-092X\(01\)00050-0](http://dx.doi.org/10.1016/S0038-092X(01)00050-0).
- Phinikarides A., Shimitra C., Bourgeony R., Koumparou I., Makrides G., Georgiou G. E., Development of a Novel Web Application for Automatic Photovoltaic System Performance Analysis and Fault Identification, Conference: 43rd IEEE Photovoltaic Specialists Conference, PVSC, 2016, DOI: 10.1109/PVSC.2016.7749921.
- Popovic I., Radovanovic I., Methodology for detection of photovoltaic systems underperformance operation based on the correlation of irradiance estimates of neighboring systems, *Journal of Renewable and Sustainable Energy*, 053701, 10, 5, 2018, doi:10.1063/1.5042579.
- Reich N.H., Mueller B., Armbruster A., Van Sark W. G. J. H. M., Kiefer K., Reise C., Performance ratio revisited: is PR > 90% realistic?, *Progress in Photovoltaics*, 20: 717–726, doi:10.1002/pip.1219, 2012.
- Reno M.J., Hansen C.W., Stein J.S., Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis, Sandia Technical Report, SAND2012-2389, 2012.
- Riley D., Johnson J., Photovoltaic Prognostics and Health Management using Learning Algorithms, Conference: 38th Photovoltaic Specialists Conference (PVSC), IEEE, 2012, DOI: 10.1109/PVSC.2012.6317887.
- Riley E., Lave M., Wu E., Dise J., Tirumalai T., Bosch J.L., Tammineedi C., On the ability of ground based global horizontal irradiance measurements to reduce error in satellite derived plane of array irradiance data for fixed tilt photovoltaic power plants, 2016, IEEE 43rd Photovoltaic Specialists Conference (PVSC), doi: 10.1109/PVSC.2016.7749597.
- Saint-Drenan Y.M., Bofinger S., Fritz R., Vogt S., Good G.H., Dobschinski J., An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation, *Solar Energy*, Volume 120, October 2015, Pages 479-493, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2015.07.024>.
- Saint-Drenan Y.M., Good G.H., Braun M., Freisinger T., Analysis of the uncertainty in the estimates of regional PV power generation evaluated with the upscaling method, *Solar Energy*, Volume 135, October 2016, Pages 536-550, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2016.05.052>.
- Schmid J., Schmidt H., Inverters for photovoltaic systems, Proceedings 5th Contractor's Meeting of the Photovoltaic Demonstration Projects, Ispra, 1991.
- Seme S., Sredenšek K., Štumberger B., Hadžiselimović M., Analysis of the performance of photovoltaic systems in Slovenia, *Solar Energy*, Volume 180, 2019, Pages 550-558, ISSN 0038-092X, <https://doi.org/10.1016/j.solener.2019.01.062>.
- Shahandashti M., Ashuri B., Mostaan K., Automatic fault detection for Building Integrated Photovoltaic (BIPV) systems using time series methods, *Built Environment Project and Asset Management*, 2018, ISSN: 2044-124X.
- Shepard D., 1968, A two-dimensional interpolation function for irregularly-spaced data, 23rd ACM national conference (ACM 68), 1968, <http://dx.doi.org/10.1145/800186.810616>.
- Silvestre S., Chouder A., Karatepe E., Automatic fault detection in grid connected PV systems, *Solar Energy*, Volume 94, August 2013, Pages 119-127, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2013.05.001>.
- Silvestre S., Aires da Silva M., Chouder A., Guasch D., Karatepe E., New procedure for fault detection in grid connected PV systems based on the evaluation of current and voltage indicators, *Energy Conversion and Management*, Volume 86, October 2014, Pages 241-249, ISSN 0196-8904, <http://dx.doi.org/10.1016/j.enconman.2014.05.008>.

Silvestre S., Mora-López L., Kichou S., Sánchez-Pacheco F., Dominguez-Pumar M., Remote supervision and fault detection on OPC monitored PV systems, *Solar Energy*, Volume 137, 1 November 2016, Pages 424-433, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2016.08.030>.

Simon M., Meyer E.L., The effects of spectral evaluation of c-Si modules, *Progress in Photovoltaics*, 2011.

Skoplaki E., Palyvos J.A., On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations, *Solar Energy*, Volume 83, Issue 5, 2009, Pages 614-624.

Solórzano J., Egido M.A., Automatic fault diagnosis in PV systems with distributed MPPT, *Energy Conversion and Management*, Volume 76, December 2013, Pages 925-934, ISSN 0196-8904, <http://dx.doi.org/10.1016/j.enconman.2013.08.055>.

Sopitpan S., Changmuang P., Panyakeow S., Monitoring and data analysis of a PV system connected to a grid for home applications, *Solar Energy Materials and Solar Cells*, Volume 67, Issues 1–4, March 2001, Pages 481-490, ISSN 0927-0248, [http://dx.doi.org/10.1016/S0927-0248\(00\)00317-2](http://dx.doi.org/10.1016/S0927-0248(00)00317-2).

Spataru S., Sera D., Kerekes T., Teodorescu R., Diagnostic method for photovoltaic systems based on light I–V measurements, *Solar Energy*, Volume 119, September 2015, Pages 29-44, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2015.06.020>.

Spertino F., Corona F., Monitoring and checking of performance in photovoltaic plants: A tool for design, installation and maintenance of grid-connected systems, *Renewable Energy*, Volume 60, December 2013, Pages 722-732, ISSN 0960-1481, <http://dx.doi.org/10.1016/j.renene.2013.06.011>.

Spiliotis E., Legaki N.Z., Assimakopoulos V., Doukas H., El Moursi M.S., Tracking the performance of photovoltaic systems: a tool for minimising the risk of malfunctions and deterioration, *IET Renewable Power Generation*, vol. 12, no. 7, pp. 815-822, 21 5, 2018, doi: 10.1049/iet-rpg.2017.0596.

Stein J.S., McCaslin S., Hansen C.W., Boyson W.E., Robinson C.D., Measuring PV system series resistance without full IV curves, 2014 IEEE 40th Photovoltaic Specialist Conference (PVSC), Denver, CO, 2014, pp. 2032-2036. doi: 10.1109/PVSC.2014.6925326.

Stein J.S., Novel Strategies for PV System Monitoring, PV-Tech Power, Technical Report, February, 2015.

Stettler S., Toggweiler P., Remund J., SPYCE satellite photovoltaic yield control and evaluation, 21st European Photovoltaic Solar Energy Conference, 2006.

Tadj M., Benmouiza K., Cheknane A., Silvestre S., Improving the performance of PV systems by faults detection using GISTEL approach, *Energy Conversion and Management*, Volume 80, April 2014, Pages 298-304, ISSN 0196-8904, <http://dx.doi.org/10.1016/j.enconman.2014.01.030>.

Takashima T., Yamaguchi J., Otani K., Oozeki T., Kato K., Ishida M., Experimental studies of fault location in PV module strings, *Solar Energy Materials and Solar Cells*, Volume 93, Issues 6–7, June 2009, Pages 1079-1082, ISSN 0927-0248, <http://dx.doi.org/10.1016/j.solmat.2008.11.060>.

Taylor J., Leloux J., Everard A.M., Briggs J., Buckley A.R., Monitoring thousands of distributed PV systems in the UK: Energy production and performance, 11th Photovoltaic Science, Application and Technology Conference (PVSAT-11), 2015.

Taylor J., Leloux J., Everard A.M., Briggs J., Buckley A.R., Quantifying PV fleet output variability in the UK: Consequences for Distribution Network Operators, 11th Photovoltaic Science, Application and Technology Conference (PVSAT-11), 2015.

Tsafarakis O., Van Sark W.G.J.H.M., Analysis and comparison of performance of photovoltaic systems during the “Counting the Sun” public awareness campaign on PV performance, 31st European Photovoltaic Solar Energy Conference and Exhibition, 2015.

Tsafarakis O., Sinapis K., Van Sark W.G.J.H.M., PV System Performance Evaluation by Clustering Production Data to Normal and Non-Normal Operation, 2018, *Energies*, 1996-1073, 11, 4, 977, doi:10.3390/en11040977.

- Tyutyundzhiev N., Petrov M., Martínez-Moreno F., Leloux J., Narvarte L., Advanced testing of PV plants using portable SCADA tool, 31st European Photovoltaic Solar Energy Conference and Exhibition, 2015.
- Ueda Y., Kurokawa K., Kitamura K., Yokota M., Akanuma K., Sugihara H., Performance analysis of various system configurations on grid-connected residential PV systems, *Solar Energy Materials and Solar Cells*, Volume 93, Issues 6–7, June 2009, Pages 945-949, ISSN 0927-0248, <http://dx.doi.org/10.1016/j.solmat.2008.11.021>.
- Van Sark W.G.J.H.M., Reich N.H., Müller B., Armbruster A., Kiefer K., Reise C., World Renewable Energy Forum, WREF, doi: 10.13140/2.1.2138.7204, 2012.
- Vargas J.P., Goss B., Gottschalg R., Large scale PV systems under non-uniform and fault conditions, *Solar Energy*, Volume 116, June 2015, Pages 303-313, ISSN 0038-092X, <http://dx.doi.org/10.1016/j.solener.2015.03.041>.
- Vaz A.G.R., Elsinga B., Van Sark W.G.J.H.M., Brito M.C., An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands, *Renewable Energy*, Volume 85, January 2016, Pages 631-641, ISSN 0960-1481, <http://dx.doi.org/10.1016/j.renene.2015.06.061>.
- Vergura S., Hypothesis Tests-Based Analysis for Anomaly Detection in Photovoltaic Systems in the Absence of Environmental Parameters, *Energies*, 1996-1073, 11, 3, 485, 2018, doi:10.3390/en11030485.
- Vergura S., A Statistical Tool to Detect and Locate Abnormal Operating Conditions in Photovoltaic Systems, 2071-1050, 10, 3, 608, 2018, doi:10.3390/su10030608.
- Wackernagel H., *Multivariate geostatistics: an introduction with applications*, Berlin:Springer-Verlag; 1995.
- Walker H.M., *Studies in the history of statistical method: With special reference to certain educational problems*, Williams & Wilkins Co, <http://dx.doi.org/10.1037/13379-000>, 1929.
- Wiles J.C., et al., PV installations, a progress report, IEEE Photovoltaic Specialists Conference, 2002.
- Williams M., Kerrigan S., Thornton A., Energy L., Automatic Detection of PV System Configuration. Presented at the World Renewable Energy Forum: 5 pp, 2012.
- Wittkopf S., Valliappan S., Liu L., Ang K.S., Cheng S.C.J., Analytical performance monitoring of a 142.5 kWp grid-connected rooftop BIPV system in Singapore, *Renewable Energy*, Volume 47, November 2012, Pages 9-20, ISSN 0960-1481, <http://dx.doi.org/10.1016/j.renene.2012.03.034>.
- Woyte A., Richter M., Moser D., Mau S., Reich N.H., Jahn U., Monitoring of Photovoltaic Systems: Good Practices and Systematic Analysis, 28th European Photovoltaic Solar Energy Conference and Exhibition, doi: 10.4229/28thEUPVSEC2013-5CO.6.1, 2013.
- Woyte A., Richte M., Moser D., Reich N.H., Green M., Mau S., Beyer H.G., Analytical monitoring of grid-connected photovoltaic systems - Good Practices for Monitoring and Performance Analysis, International Energy Agency Photovoltaic Power Systems Programme, IEA PVPS Task 13, Subtask 2, Report IEA-PVPS T13-03, ISBN 978-3-906042-18-3, 2014.
- Yagi Y., Kishi H., Hagihara R., Tanaka T., Kozuma S., Ishida T., Waki M., Tanaka M., Kiyama S., Diagnostic technology and an expert system for photovoltaic systems using the learning method, *Solar Energy Materials and Solar Cells*, Volume 75, Issues 3–4, 1 February 2003, Pages 655-663, ISSN 0927-0248, [http://dx.doi.org/10.1016/S0927-0248\(02\)00149-6](http://dx.doi.org/10.1016/S0927-0248(02)00149-6).
- Yang D., Solar radiation on inclined surfaces: corrections and benchmarks, *Solar Energy* 2016;136:288–302.