# Modelling and Forecasting of Photovoltaic Generation for Microgrid Applications: from Theory to Validation

**Thèse N° 9039**

## ENRICA SCOLARI

**2019**

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

The good life is the one inspired by love,
and guided by knowledge.
— Bertrand Russel

Ai miei nonni…

# Acknowledgements

First of all, I would like to thank my supervisor Prof. Mario Paolone for giving me this unique opportunity and for being always present during these years. He has been a reference for my research, from both a technical and educational point of view. His way of leading our group, his energy, and his force have represented a great source of motivation and inspiration for me. Then, thanks to Fabrizio for your support from the very beginning and for being the pillar of my PhD. I have learned so much from you, ranging from power system to computer science, from skiing to cables soldering. I will always be very grateful. A thank also to my co-supervisor Dr. Alexandre Oudalov for his support and suggestions during the years and for bringing the industry perspective into my research. A special thank goes to Dr. Rachid Charkaoui for bringing the positiveness and wisdom that each laboratory would need, and to Sophie for taking care of us every day!

And then a thank to who, by simply being part of my life, have contributed to the final outcome of my thesis. A special thank goes to Asja and Asia, who made my years here unique as only friends can do. You are two incredible girls and an example of women in engineering. To my officemates and friends, Andreas and Lorenzo, and to Emil. They were always very sensitive, helpful and trustful. To all my friends here and back home: the Zaphiro, all the DESL colleagues now and in the past, my flatmates (Irene, Elena, Mathieu), Chiara F., Enrico, the GVP, my forever friends Alessia, Sara F., Greta, Calle, Sara P., Gloria, Andrea, all the UNIBO group, the Californian team, Chia, Daddo, Mattia, and Becky.

The greatest thank goes to my family (nonni, mamma and papà, Alessandra and Oscar, Laura and Christian, Martina and Gaia) for the continuous support, confidence and for constantly believing in me: you are my source of happiness.

Finally, a special thank to Paolo, simply because without you and your intelligence I would not have achieved this goal.

# Abstract

The penetration of stochastic renewable generation in modern power systems requires to reconsider conventional practices to ensure the reliable operation of the electrical network. Decentralized control schemes for distributed energy resources (DERs) have gained attention to support the grid operation. In order to cope with the uncertainties of the DERs, predictive schemes that leverage on forecast of renewable generation recently came into prominence. The period of the control action typically depends on the availability of the reserve in the grid. For the case of microgrids, their limited physical extension and the lack of spatial smoothing imply fast power fluctuations and the necessity of coupling energy management strategies with real-time control. Among the DERs, small-scale photovoltaic (PV) systems are expected to represent most of the future available capacity, and consequently, solar resource assessment and power forecasting are of fundamental importance.

This thesis focuses on developing forecasting methods and generation models to support the integration of photovoltaic systems in microgrids, considering short-term temporal horizons (below one hour) and fine spatial resolution (single site installations). In particular, we aim at computing probabilistic prediction intervals (PIs), i.e., we include information accounting for the intrinsic uncertainty of the prediction. In this respect, nonparametric tools to deliver PIs from sub-second to intra-hour forecasting horizons are proposed and benchmarked. They forecast the AC power and/or the global horizontal irradiance (GHI) by extracting selected endogenous influential variables from historical time-series. The methods are shown to outperform available state-of-the-art techniques and are able to capture the fastest fluctuations of small-scale PV plants. Then, we investigate how the inclusion of features from ground all-sky images can be used to improve time-series-based forecasting tools, thanks to identifying clouds movement. In this respect, we define a toolchain that allows predicting the cloud cover of the sun disk, through image processing and cloud motion identification. Furthermore, a methodology to estimate the irradiance from all-sky images is proposed, investigating the possibility of using an all-sky camera as an irradiance sensor. Next, we consider the problem of having power measurements that are corrupted by exogenous control actions (e.g. curtailment) and, therefore, not representative of the true potential of the PV plant. We propose a model-based strategy to reconstruct the maximum power production of a PV power plant thanks to integrating measurements of the PV cell temperature, system DC voltage, and current. The strategy can improve time-series-based direct power forecasting techniques when the production of the PV system is curtailed and thus the measured power does not correspond to the maximum available. The proposed methods to model and forecast

the PV generation are then integrated into a single chain that allows delivering power PIs that are able to account for the overall uncertainty of a PV system at a predefined confidence level. In the last part of the thesis, the proposed methods are experimentally validated in a real microgrid by considering possible applications in modern power systems. First, we show how the stochastic behaviour of a PV plant can be accounted in a control framework specifically designed for the real-time control of microgrids. Then, we focus on an energy management problem and we discuss how PV plants can be used as flexible resources to track a specific dispatch plan set the day before operation.

# Résumé

La pénétration toujours croissante des ressources énergétiques distribuées (REDs) dans les systèmes électriques modernes oblige à reconsidérer les pratiques conventionnelles pour assurer le fonctionnement fiable des réseaux électriques. En effet, afin de faire face aux incertitudes liées aux REDs, caractérisé par une production stochastique, les schémas de contrôle prédictif décentralisé ont récemment pris de l'importance. Ces schémas s'appuyant sur les prévisions de consommation et de production d'énergie renouvelable. La durée de l'action de contrôle dépend généralement de la disponibilité de la réserve de production dans un réseau électrique donné. Dans le cas des micro-réseaux, leur extension physique limitée et l'absence de lissage spatial impliquent des fluctuations de puissance rapides et la nécessité de coupler des stratégies de gestion de l'énergie au contrôle en temps réel. Parmi les REDs, les systèmes photovoltaïques (PV) à petite échelle devraient représenter la majeure partie de la capacité de production future disponible. Par conséquent, une évaluation précise des ressources solaires et une prévision de leur puissance sont d'une importance fondamentale. Cette thèse porte sur le développement de méthodes de prévision et de modèles de génération pour soutenir l'intégration de systèmes photovoltaïques dans des micro-réseaux, en prenant en compte des horizons temporels à court terme (moins d'une heure) et des horizons spatiaux limités (installations à site unique). En particulier, nous cherchons à calculer des intervalles de prédiction probabilistes (IP), c'est-à-dire que nous incluons des informations rendant compte de l'incertitude intrinsèque de la prédiction. À cet égard, des méthodes non paramétriques permettant de fournir des IPs à des horizons de prévision inférieurs à la seconde ou à l'heure sont d'abord proposés et comparés. Ils peuvent être utilisés pour prévoir les puissances en AC et/ou le rayonnement horizontal global en extrayant certaines variables influentes endogènes à partir de séries temporelles historiques. Il est démontré que les méthodes proposées surpassent les techniques de pointe disponibles et permettent de capturer les fluctuations les plus rapides des installations photovoltaïques à petite échelle. Ensuite, nous étudions comment l'intégration d'entités à partir d'images du ciel (prises par un sky imager orienté vers le haut qui prend des photos hémisphériques) peut être utilisée pour améliorer les outils de prévision basés sur des séries temporelles, grâce à l'identification du mouvement des nuages. À cet égard, nous définissons une chaîne d'outils permettant de prédire la couverture nuageuse du disque solaire, par le traitement de l'image et l'identification du mouvement des nuages. De plus, une méthodologie pour estimer l'irradiance à partir d'images du ciel est proposée. Elle permet d'étudier la possibilité d'utiliser le sky imager comme capteur d'irradiance. Ensuite, nous examinons le problème des mesures de puissance qui sont corrompues par des actions

# Contents

# Contents

# List of Figures

# List of Tables

# List of Tables

# Introduction

## Motivation

The increasing proportion of distributed generation (DG) in the electrical grid requires redefining conventional practices to mitigate the impact of stochastic resources like photovoltaic (PV) systems [6]. The main issues are related to the operational limits of the distribution grid, in terms of quality of supply and line congestions, [7].

A paradigm increasingly advocated in the recent technical literature to cope with the variability of stochastic generation is the development of robust and predictive control strategies. They take advantage of the short-term forecast of renewable generation in order to plan adequate counteractions to prevent or mitigate operational issues related to renewables' power fluctuations. Examples include the dispatchability of the distributed resources, achieving self-consumption of locally generated electricity, and the short-term redispatching of conventional generation units, e.g. [8, 9]. In this context, knowing the expected solar power and its associated uncertainty is a fundamental aspect, [10]. The main challenges of predicting the solar energy output are related to its volatile characteristics and the temporal and spatial dependencies of the irradiation patterns, [11]. Indeed, the PV output has both a deterministic and a stochastic component. The deterministic component depends on the sun position and can be easily calculated. The stochastic component depends on the cloud shading effects and represents the main source of uncertainty. In particular, for localized power plants, cloud-induced power dynamics can be extremely high, with fluctuations at sub-second time-scale [12]. Since conventional solar forecasting methods, like those based on satellites and numerical weather predictions, lack of the spatial and temporal resolution required by modern power system applications to cope with these fast dynamics, new short-term forecasting methods based on local monitoring are required, [2]. These methods generally rely on analytic of historical time-series and are based on machine learning and statistical algorithms, [13]. The inputs are diverse: local measurements of the solar irradiance and temperature, ground sky-images, or electrical measurements of the PV plant like currents and voltages, e.g. [14]. Since the solar power can be curtailed, the real-time direct measurements of the power might not reflect the true potential of the PV system and therefore physical models of the plant should also be integrated into the forecasting chain.

There is a general gap in the literature about methodologies able to quantify the overall

uncertainty of a PV plant operation at very short-term and low aggregation level. In this context, the thesis provides and benchmarks different methods to predict the power output of a PV plant at short-term horizons (from sub-second to several minutes). The forecasting is provided in terms of a prediction interval (PI) which is the interval where the future observation is expected to fall, with a given probability. Accounting for the uncertainty of the prediction is important since it allows defining robust control policies. The developed methods are integrated into control frameworks operating in experimental testbeds where photovoltaic plants are able to provide flexibility to the grid in coordination with other resources, e.g. battery storage systems.

## Thesis Outline

The thesis is organized as follows:

**Chapter 1** summarizes the main challenges associated with the integration of renewable resources into the electrical grid and introduces some main concepts, e.g., *active distribution network* (ADN) and *microgrid*. It describes recently proposed predictive-based approaches for microgrids control and presents the state-of-the-art on solar generation modelling and forecasting.

**Chapter 2** focuses on the direct AC active power forecast of small-scale photovoltaic (PV) plants for sub-second time horizons. A nonparametric method for irradiance forecast, called Dynamic Interval Predictor (DIP) originally proposed in [12], is extended to compute prediction intervals for PV AC active power from 100 ms up to 500 ms, by considering influential variables other than the derivative of the irradiance.

**Chapter 3** deals with global horizontal irradiance (GHI) forecasting based on historical time-series. It describes a nonparametric tool for computing short-term PIs of the GHI using clustering of historical observations of the clear-sky index. The main novelty of the proposed method is the ability to deliver PIs from sub-second (500 ms) to intra-hour (5 min) forecasting horizons. Results are presented by comparing the proposed method with state-of-the-art prediction models.

**Chapter 4** investigates the use of an all-sky camera (installed at ground level) to infer cloud motion and to estimated the GHI value. Image processing, feature selection, and machine learning-based methods are combined and benchmarked at this scope to provide both probabilistic and point forecast of the GHI.

**Chapter 5** proposes model-based methods to reconstruct the maximum power of a PV plant starting from measurements of the DC voltage, current, and temperature. The estimation performance is evaluated by using measurements from an experimental setup and benchmarked against estimations from a GHI sensor (pyranometer).

**Chapter 6** discusses a gray-box method to assess the overall flexibility and associated uncer-

tainties of a small-scale PV plant. The proposed method integrates a model for the PV modules, a model for the converter, and a probabilistic short-term forecasting tool, i.e. it considers the overall chain that allows obtaining the AC maximum power. For different forecast horizons and for different levels of curtailed energy, we identify which sub-model is the most critical in terms of uncertainty.

**Chapter 7** provides application examples of the methods presented in the previous chapters. In particular, the importance of knowing the future behavior of a PV plant is demonstrated, by considering the real-time control of a low voltage microgrid, and an energy management problem to track a specific dispatch plan set the day before operation.

**Chapter 8** contains a summary of the main outcomes and future perspectives.

## Contributions

A summary of the contributions of the thesis is the following:

- Development and experimental validation of a new probabilistic method to forecast the short-term AC power availability of a small-scale PV plant in a direct way, namely by simply learning from historical measurements of the AC power. It is shown how the PV power can vary more than 20% in one second, this requiring the forecasting tool to be computationally efficient. It is discussed how direct power forecasting methods can encounter some limitations when the power measurements are not representative of the true potential of the PV plant (e.g., because of some curtailment control actions). The importance of reconstructing the maximum power availability is quantified in terms of improved performance of the learning process of a selected forecasting tool.

- Development and experimental validation of new probabilistic methods to estimate and forecast the global horizontal irradiance (GHI) at short-term (from sub-second to 5-10 min). Machine learning-based models using different inputs are considered and benchmarked: from pyranometers, electrical measurements of voltages and currents, all-sky images, and satellites. The comparison is performed in terms of predicted maximum power, i.e. also considering the uncertainty associated with the conversion from the GHI to the AC power.

- Deployment of the prediction methods in embedded industrial systems to provide real-time prediction intervals of the generation of monitored PV plants. The methods are shown to be reliable (i.e., to guarantee a given coverage probability) and computationally compliant with the fastest dynamics that characterize small-scale inertia-less power systems (namely, sub-second time scale). An all-sky camera prototype has been developed. It makes use of machine learning and image processing techniques to provide information on the cloud motion and improve the forecasting with respect to the sole use of GHI measurements.

- Application of the developed methods to provide PV flexibility when controlling a real-scale test-bed microgrid and for the energy dispatching of an electrical feeder. The added value coming from integrating information on the PV expected maximum power is shown by controlling the PV power output of a PV roof installation.

# 1 Inclusion of Stochastic Resources into Predictive Control Frameworks

## 1.1 Chapter Highlights and Summary

In this Chapter, we introduce the main topics and define the principal concepts of the thesis, and we give an overview of the current literature.

We first define the concepts of *active distribution network* (ADN) and *microgrid* and we introduce the related challenges with a special insight on the control of stochastic resources. Since the thesis focuses on the integration of photovoltaic (PV) into the electrical grid, we then give an overview on how PV plants are generally modelled, from the single PV cell to the whole system. Finally, an overview of the PV *variability* and power forecasting is presented, also serving as an introduction to the following Chapter.

## 1.2 Active Distribution Network and Microgrid: Main Definitions

In the last two decades, distribution networks have evolved from passive networks (i.e. supplying local loads) to active distribution networks. As defined by CIGRE, [15], *"ADNs have systems in place to control a combination of distributed energy resources (DERs), defined as generators, loads, and storage. Distribution system operators (DSOs) have the possibility of managing the electricity flows using a flexible network topology. DERs take some degree of responsibility for system support, which will depend on a suitable regulatory environment and connection agreement."*
The integration of DERs is mainly driven by environmental factors, however their implementation is such that a large number of technical and regulatory issues need to be accounted for, [16]. Technical issues are related to:

- Power quality: deviations in magnitude and frequency of the voltage and current waveforms from the specified ranges, creating problems for the customer. The basic types of power quality disturbances are transients, over and under voltages, harmonic distortion, voltage flickers, voltage sags and swells, electrical noise.

- Stability: voltage and frequency are strongly coupled in a microgrid and therefore the common classification that distinguishes between *voltage instability* and *frequency instability* in common power systems is more difficult for microgrids. Stability issues for microgrids should be divided into control system stability (electric machine and converter stability) and power supply and balance stability (loss of a generation unit, violation of DERs power limits, poor power sharing among multiple DERs, wrong selection of slack(s) resources, and/or involuntary no-fault load tripping). A review of the main challenges related to the stability of a microgrid can be found in [17].

- Protection: a first problem is related to DERs location since they are distributed rather than central, and therefore the short circuits might originate from directions that are not considered in conventional protection schemes. A second consideration is that DERs are connected to the grid via inverters, and the short circuit capacity of grids with inverters is lower than that of grids with rotating machines of the same rating; furthermore, the inverter short circuit current has a different time characteristic determined by the inverter control scheme. These considerations lead to what is called protection blinding, i.e. the grid contribution to the short-circuit current could be below the pick-up current of the feeder relay. Third, the DER should be now able to provide short circuit capacity before and after the fault and the Fault Ride Through (FRT) capability is of major concern. A detailed description of the main issues related to protections can be found in [18].

Traditional regulations also need to be reviewed. In particular, distribution system operators (DSOs) will have a crucial role, as they are expected to carry the main investment burden, interact with transmission operators, and with the market players, see [19, 20].

In order to cope with the aforementioned technical issues, modern communication technologies and smart electronic devices are needed in ADNs, leading to what is called a *smartgrid*. In this context, the concept of *microgrid* is defined, which is *a group of DERs, including Renewable Energy Resources (RES) and Energy Storage Systems (ESS), and loads, that operate together locally as a single controllable entity*, [21]. The main advantage is that the microgrid can disconnect itself from the main grid (islanding) when a power quality issue or a fault occurs in the local grid, [22]. Furthermore, by locally controlling the available resources, a microgrid can facilitate the integration of stochastic resources by solving the technical problems in a decentralized fashion, without requiring to refurbish the regional/national distribution systems which would be more complex and expensive. Finally, when in grid-connected mode, the microgrid can provide ancillary services to the main grid by dispatching the available resources, [23].

## 1.3 Active Distribution Network and Microgrid: Control

### 1.3.1 Traditional Control

The available literature summarizing practices for microgrid control (e.g [6, 24]) distinguishes between three levels of control, similarly to the frequency control of bulk traditional power systems: primary, secondary, and tertiary control. These levels differ for the speed of response and the infrastructure requirements. The primary control is essential to provide independent active and reactive power sharing controls for the DER, and features the fastest response (second time-scale). The secondary control is based on a longer term energy management to compensate for the voltage and frequency deviations caused by the operation of the primary controls. Ultimately, the tertiary control manages the power flow between the microgrid and the main grid and facilitates an economically optimal operation. In view of the challenges caused by the highly fluctuating DERs, we focus here on the primary control that is conventionally based on droop control. It generally uses local measurements and local control laws to directly control the DERs output. Droop-based methods are conventionally used in power systems and are based on the relationship between active power/frequency and reactive power/voltage to emulate the droops characteristics of synchronous generators, [25, 26, 27, 28, 29, 30]. Synchronous machines and their well-known dynamics are progressively substituted by power electronics-interfaced resources and the droop principle can be integrated into power electronics interfaces between the sources and the microgrid, namely voltage-source inverters (VSI). The main advantage of droop is that it is only based on local measurements without any need of communication, this conferring flexibility. However, these techniques have some main challenges:

- in a microgrid the interconnecting impedances are featured by a non-negligible resistive component so that the real power is affected more by the voltage magnitude, and the reactive power by phase angle differences. There are two main approaches to face this problem. The first is to introduce virtual resistive impedances at the converter output through closed-loop control as in [30]. This impedance modifies the output voltage reference based on current feedback. The second technique is to decouple the voltage and frequency droop controls by introducing current and voltage control scheme able to compensate the effect of the line impedances (e.g. [26]). As an example, PV inverters used to interface the resource to the grid, are generally used for droop-based voltage regulation by controlling the reactive power injections [31]. However, since power distribution networks are characterized by lines parameters with large $R/X$ (ratio of the components of their longitudinal impedances), there is generally the need of coupling the reactive power control with active power curtailment (APC) strategies [32, 33, 34];

- distributed generation has generally an intermittent nature and does not contribute to the system inertia. This implies the lack of an energy buffer and the electrical decoupling of the resource to the grid. The lower the inertia, the more nervous is the grid frequency in responding to abrupt changes in load and generation patterns. In this case, enhanced

droop control strategies are proposed, based for example on mimicking the inertia of conventional generation (fossil fuel-based) using storage systems, [35].

### 1.3.2 Predictive Control

The main alternative to conventional droop-based methods is the definition of predictive control-oriented approaches that are capable of optimizing the microgrid operation by leveraging consumption and generation forecasting. The use of predictions of the future behaviour of the system leads to a more performing exploitation of the available resources with respect to conventional myopic strategies, thanks to the optimal management of the available resources. The main drawback is that these methods generally rely on communication whose failure might lead to suboptimal operating conditions.

Regarding the control architectures, we can define between two main approaches, i.e. centralized and decentralized. A centralized control architecture consists in a central controller that, informed of the states of the available resources in the microgrid, determines an appropriate dispatch according to selected objectives. The central controller can use either on-line or off-line calculations to take decisions of the optimal operation. Examples of centralized approaches for scheduling the energy resources can be found in [36]. The Authors propose a scheduling based on two stages: a day-ahead scheduler for the optimization of distributed resources production during the following day and an intra-day scheduler that every 15 min adjusts the scheduling in order to take into account the operation requirements and constraints of the distribution network. In [4, 37] a new control framework based on hierarchical agents is proposed to control a power grid. Each resource is equipped with a resource agent (RA). A RA translates the internal state and constraints of its resource into an abstract, device-independent format to a grid agent (GA), which controls a subsystem consisting in the local electrical grid interconnecting a number of resources. On the contrary, decentralized approaches intend to solve the problem in a way that decisions on the control variables are taken locally. In the literature, decentralized secondary control is based on the concept of Multi-Agent System (MAS). A MAS is a system composed of intelligent agents that, provided with local information, interact with each other to reach several goals. State estimation theory in power system can be integrated in a MAS. The use of multi-agent systems to control a distributed smart grid in a simulated environment is described in [38] and [39].

### 1.3.3 Inclusion of Uncertainties into Predictive Control

In the above-described context, stochastic resources (mainly photovoltaic and wind) play a major role, and accounting for uncertainties has become a challenge at all the different control levels. In particular, the growth of photovoltaic has been exponential and its worldwide cumulative capacity reached about 302 GW by the end of 2016, [40], with small-scale PV systems playing a major role. For the case of microgrids, their limited geographical extension joined with the capability to function autonomously determine the presence of very fast (sub-seconds) power fluctuations and thus the need for very fast control actions. Furthermore, the

scheduling and operation of microgrids also imply the need to account for their intermittent behaviour on a longer time-scale (from minutes to day-ahead). For the above-mentioned reasons, microgrids control is asked to deal with uncertainties at different time-scales. Decision making under uncertainty is a well-known problem in modern control theory. One method to deal with uncertainty is *stochastic control*. In stochastic control, uncertainties in the observations are modelled with probability distribution functions (pdfs), [41]. Alternatively, *robust control* methods try to bound the uncertainty rather than defining it with a distribution function. Provided that the uncertain parameters are bounded, the control is said to be robust if the results meet the control system requirements in all cases. Therefore, robust control theory might be stated as a worst-case analysis method, [42]. Several control schemes have been proposed in the literature to account for the stochastic nature of load and generation. In [43] the scheduling problem of a building is considered where the objective function is to minimize the overall cost of electricity and natural gas for a building operation over a time horizon while satisfying the energy balance. The uncertainties are captured and their impact is analyzed by the scenario tree method. Authors of [44] consider the load dispatch and optimal reconfiguration of a microgrid, by also accounting for the stochasticity of renewable resources using a forecasting tool based on support vector regression techniques. In [45], a two-stage adaptive robust model for the unit commitment problem in the presence of nodal injection uncertainty is proposed. Authors of [46] propose an energy management strategy for grid-connected microgrid with renewable sources in order to minimize the net monetary cost; a robust formulation is proposed accounting for the worst-case amount of harvested renewable energy.

In this context, the necessity of modelling the PV resource variability and uncertainty in a probabilistic fashion becomes of fundamental importance. The next paragraphs give an introduction of how this can be achieved, by modelling and forecasting the solar irradiance and the PV generation.

## 1.4 Modelling of Photovoltaic Plants

The main inputs to model a photovoltaic plant are weather information and *solar irradiance* data. This last quantity is an instantaneous measurement of solar power over a given plane; unless differently specified, in this work we refer to terrestrial irradiance, namely inside the atmosphere. When the extraterrestrial solar radiation traverses the atmosphere, some photons are scattered by interacting with the atmosphere, producing what is called diffuse irradiance component. The global irradiance ($I$) received by a horizontal surface includes both the beam (or direct) component ($B$) and the diffuse one ($D$). For a horizontal surface is possible to write:

$$I = B cos(\theta_z) + D, \tag{1.1}$$

where $\theta_z$ is the solar zenith angle relative to an observer on the surface of the Earth. The main solar angles used to describe the solar position are shown in Fig. 1.1 and are the solar elevation ($\theta_{el}$), the azimuth ($\theta_A$), and the zenith ($\theta_z$). When referring to a tilted surface, a further irradiance component should be considered that is the one reflected from the ground and is indicated with R. Thus, for tilted surfaces we can write:

$$I = Bcos(\theta_z) + D + R. \tag{1.2}$$

Fig. 1.2 shows possible components of the solar irradiance on different surfaces.
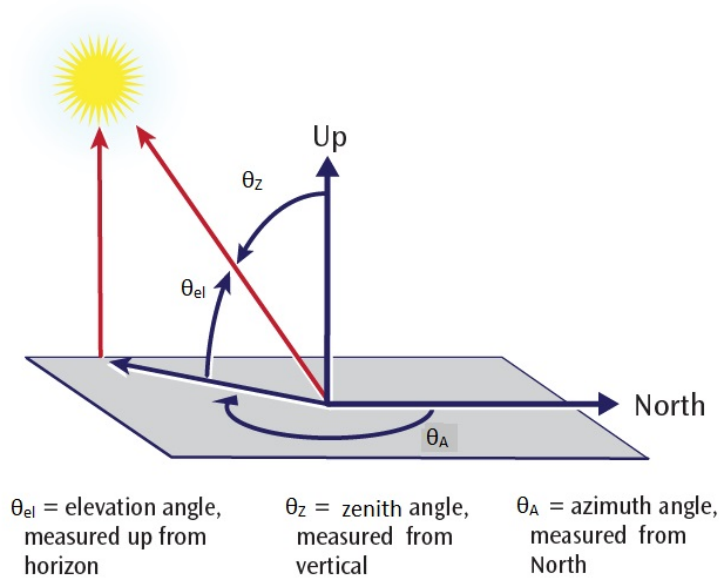


*Figure 1.1 – Solar angles, [1].*

Modelling PV systems generally includes the estimation of the irradiance on the tilted plane (plane of array, POA). This implies the necessity of knowing at any time i) the sun position and the ii) the orientation of the array. Furthermore, modelling the POA irradiance from the *global horizontal irradiance* (GHI) involves two steps, namely the decomposition of GHI into its *direct* and *diffuse* components and the transposition of these components to the POA of the modules. Decomposition models are generally empirical and use the *clear-sky index* (that is the ratio of the global horizontal irradiance and the theoretical clear-sky one), e.g. [47]. The most used transposition models are the Hay/Davies and the Perez models, [48, 49]. An overview of the performance of the different transposition methods can be found in [50]. They differ in the way they treat the diffuse component which is considered as the largest potential source of error.

The irradiance on the PV panel surface can be monitored by installing specific instruments in the proximity of the plant. *Pyranometers* are generally used to measure the GHI since they have
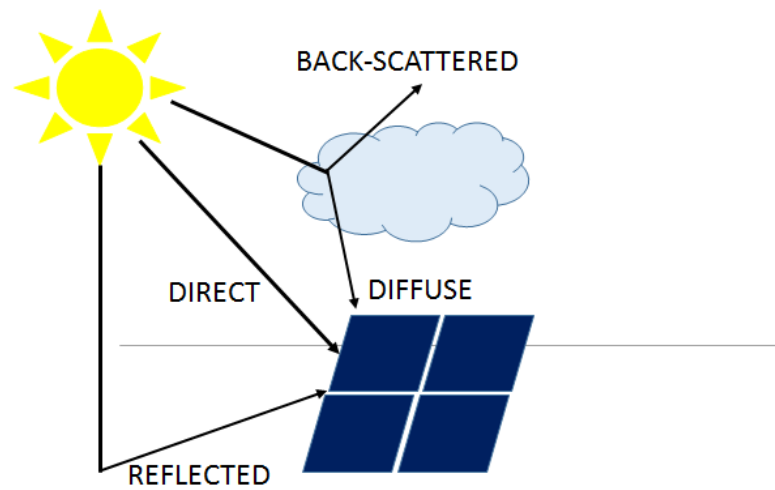
*Figure 1.2 – Solar radiation components.*

a 180° field of view. They can also be used to measure the diffuse component, by blocking out the beam radiation with a disk placed over the instrument. It is possible to distinguish between two types of pyranometers. *Thermopile pyranometers* have a black surface that is heated by the incident solar radiation. The resulting increase of temperature is measured via thermocouples connected series-parallel to make a thermopile. Alternatively, solid-state silicon *photodiodes* generate a photocurrent that is proportional to the incident flux. In general, photodiode pyranometers are less accurate in presence of clouds but have a higher response time than thermopile pyranometers. An example of a photodiode pyranometer is shown in Fig. 1.3.



*Figure 1.3 – Apogee All-Season Heated Pyranometer.*

Besides the irradiance, another main input for PV modelling is the cell temperature. The cell temperature depends on the air temperature, wind speed, solar irradiance, and type of PV module, [51]. The most common procedure to measure the cell temperature is to measure

(a) i-v curves for different irradiance values.

(b) i-v curves for different temperature values.
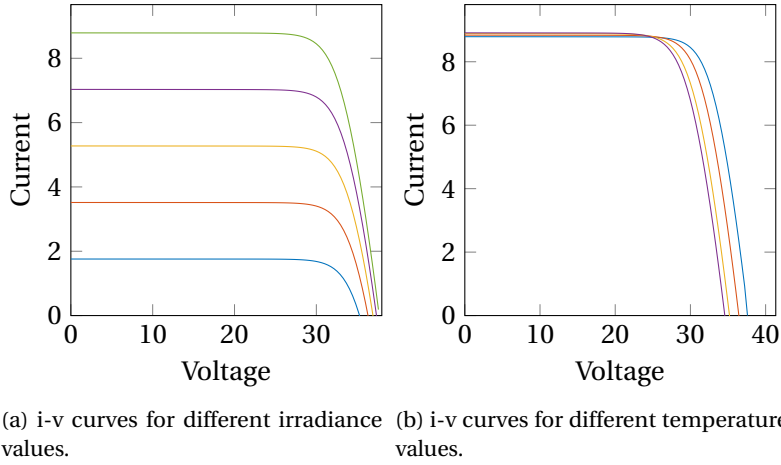
*Figure 1.4 – PV cell characteristics.*

the backside temperature of the module and assume that is equal to the cell one. In order to consider the temperature difference between the rear surface of the module and the cell, an offset can be estimated. E.g., the work in [52] proposes a difference of 3° for an irradiance of 1000 W/m$^2$ that is then linearly scaled with the irradiance. Alternatively, the cell temperature can be estimated from measured weather quantities using empirical models. For example, if the irradiance $I$ and the air temperature $T_{air}$ are known by measurements, it is possible to approximate the cell temperature $T$ as:

$$T = T_{air} + \frac{\text{NOCT} - 20}{80} I \tag{1.3}$$

where the Nominal Operating Cell Temperature (NOCT) is a parameter of the module and can be generally found in the PV module datasheet.

The values of the irradiance and temperature define the operating current-voltage (i-v) curve of a PV cell. The i-v characteristics of a generic PV cell are shown in Fig. 1.4, where different colours refer to different irradiance and temperature values (Figures 1.4a and 1.4b, respectively). In order to model the i-v curves (and so the behaviour of the panels), equivalent circuit models can be used. The parameters of the models can be found by using datasheet information or by fitting the available measurements. Fig. 1.5 shows the circuit for an ideal PV cell. From the semiconductor theory, $i_p$ is the current generated by the incident light and is proportional to the irradiance, and $i_D$ is the leakage current of the diode (represents the voltage-dependent current lost to recombination), [53]. The governing equation is the Shockley diode equation:

$$i = i_p - i_D \left[ \exp\left( \frac{vq}{n_r kT} \right) - 1 \right], \tag{1.4}$$

where $k, q$ are physical constants and stand for diode Boltzmann constant and electron charge

respectively, and $n_r$ is the diode ideality factor. In order to account for the real behaviour
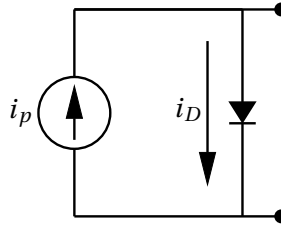


*Figure 1.5 – Circuital model of an ideal PV cell.*

of a PV cell, shunt and series resistances are added, leading to the well-known one-diode 5-parameter model shown in Fig. 1.6.
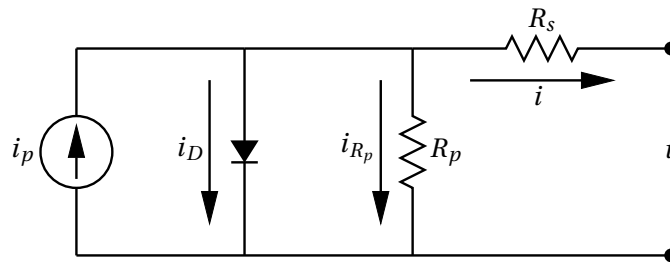


*Figure 1.6 – Circuital model of real PV cell including the series and parallel resistances.*

The main equation in this case is:

$$i = i_p + -\frac{v + R_s i}{R_p} - i_D \left[ \exp\left( q \frac{v + R_s i}{n_r k T} \right) - 1 \right], \tag{1.5}$$

where $R_p$ and $R_s$ are the shunt and series resistance, respectively. The model is derived from physical principles (e.g. [54, 3]). The series resistance considers the losses due to the current flow resistance in the silicon material, in the electrodes, and in their contact. The shunt resistance accounts for the leakage current in the p–n junction and depends on the temperature. The single-diode model represents a good compromise between simplicity and accuracy and is therefore the most adopted one. The model requires to compute the 5 parameters $R_s, R_p, i_p, i_D, n_r$. Normally they can be extracted from the equations solved at different operating points (open circuit, short circuit, maximum power point) and using information from the PV module datasheet. Real measurements, if available, can also be used to fit the i-v curves; however this requires to cover all the different operating conditions for a sufficient number of irradiance and temperature combinations.

More complex models have been proposed in the literature to improve the single diode model. For example the two-diode model allows accounting for recombination current losses which is higher at low voltage, see [55]. A review of the possible models can be found in [56] while a review on the parameter estimation of solar photovoltaic (PV) cells can be found in [57].

A PV module refers to a number of cells connected in series. Modules are connected in series and in parallel to form an array. The PV model can be extended from cell to module and then

to an entire plant by knowing how the elements are connected (series or parallel) as described in [58]. This allows computing the DC power output of a PV plant when its operational point on the i-v curve is identified. For a PV panel composed of $n_p$ cells in parallel and $n_s$ in series, Eq. 1.5 becomes:

$$i = i_p n_p + -\frac{v + R_s i \frac{n_s}{n_p}}{R_p \frac{n_s}{n_p}} - i + -i_D n_p \left[ \exp\left( q \frac{v + R_s i \frac{n_s}{n_p}}{n_r k T n_s} \right) - 1 \right].$$ 
(1.6)

The last important step in modelling a PV plant is the conversion of the DC power into an AC one, to account for the losses of the inverter. The inverter efficiency is the ratio of the usable AC power and the input DC power, and it is mainly a function of the DC power and DC voltage. Two main models are available in the literature to model the behaviour of PV inverters. The Sandia model, [59], predicts the AC output power from DC input power and voltage while the Driesse model, [60], estimates the power losses in the converter. Both the models are based on empirical coefficients that need to be fitted using real measurements from the converter. An example of the efficiency of a PV converter as a function of the AC power (in % of the rated one) is shown in Fig. 1.7.
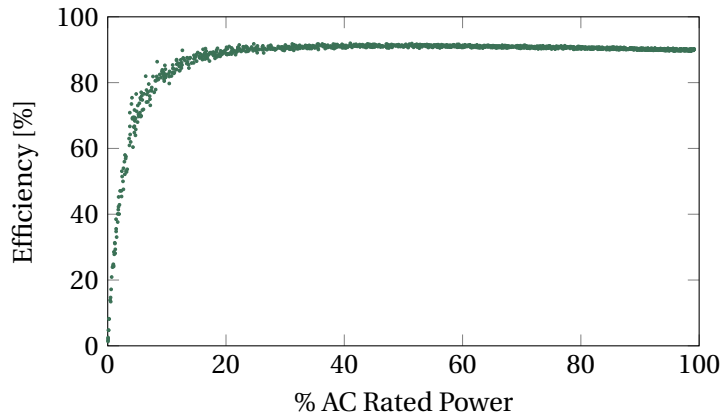


*Figure 1.7 – PV Converter Efficiency.*

## 1.5  Solar Resource Variability

The variability of PV generation mostly depends on the irradiance that is impinging the PV panels. Irradiance variability includes a deterministic component, that can be computed by knowing the solar geometry, and a cloud-related variability component. In order to remove the deterministic component, the clear-sky index is introduced. It is defined as the ratio between the measured GHI, $I$, and the clear-sky irradiance at ground level, $I_{cs}$ (i.e., the irradiance that would reach the ground in clear-sky conditions):

$$K = \frac{I}{I_{cs}}. \tag{1.7}$$

The clear-sky irradiance can be calculated using clear-sky models; a review of the available methods can be found in [61]. The solar position and the extraterrestrial radiation are common inputs to all the clear-sky models and can be calculated by knowing the site's location (latitude and longitude) and the local time; more sophisticated models also account for properties of the atmosphere (e.g., aerosols, water vapour..).

Irradiance cloud-induced variability should be considered from both a spatial and temporal perspective. As shown in [11], in a given location, the solar energy variability decreases when the temporal integration increases. Furthermore, increasing the solar generation from a single unit to a larger number of dispersed units also reduces the overall variability (this phenomenon is known as smoothing effect). In order to understand the spatial and temporal effects on variability, we first introduce the clear-sky index variability:

$$\text{Clear-sky index variability} = \sqrt{\text{Var}[\Delta K_t]} \tag{1.8}$$

where Var is the variance and $\Delta K_t = K_t - K_{t-1}$, namely is the difference between adjacent elements of the time series. Hoof and Perez, [62], estimate that for a fleet of N locations with uncorrelated clear-sky index time-series, the fleet nominal variability is:

$$\sigma_t^{fleet} = \frac{\sigma_t}{\sqrt{N}} \tag{1.9}$$

where $\sigma_t$ is the variability of a single plant. In other words, nearby plants, that are highly correlated, exhibit the same variability of a single plant while dispersed plants, for which the correlation of the time series is lower, shows a reduced variability (of a factor $\sqrt{N}$). Therefore, the key factor is to assess the correlation among the available PV plants. The correlation is generally considered as a function of three factors: the distance between two locations, the considered time scale, and the speed of the clouds. Authors of [63] shows that the zero correlation threshold is 500 m, 1 km, 4 km and 10 km for fluctuation time scales of 20 s, 1 min, 5 min, and 15 min, respectively. This information can be used for planning PV installations such that natural smoothing naturally occurs. For example, the same work considers dispersed solar resource in a metropolitan area (40 × 40 km) and shows that the high frequency (20 s) variability experienced by a single small system is reduced by a factor of 80 when considering the entire metropolitan area. The variability reduction would become smaller as the considered frequency increases: the metropolitan variability would be reduced by respectively 40, 10, and 4 for fluctuation time scales of 1, 5, and 15 min. In [64], the Authors show that the level of variance of a 10 MWp plant is four times higher than that one of 1.1 MWp plants dispersed across a 15 km area. In other words, resource spreading can reduce the issues at substation level by reducing the variability of the aggregated power when considering an entire network.

However, local transients at the single plant level still persist, and issues such as voltage flickers and inverter tripping are not eliminated. There might be local installations that violate the grid constraints, and therefore local monitoring is necessary to undertake appropriate counteractions.

From a temporal perspective, short-term variability is to be accounted for small-scale PV plants and decreases with the considered spatial footprint. For small and medium scale PV plants (from meters to few kilometers) variations in the order of seconds and minutes can be relevant and create issues (e.g. voltage violations). On the contrary, for large areas and regional transmission, fluctuations above 15-30 min should instead be considered to guarantee the grid balance between generation and consumption.

In the present work, we focus on localized PV plants and microgrids where natural mitigation due to spatial smoothing does not always apply. In this context, short-term solar forecasting and local solar assessment are extremely important to mitigate the operational issues caused by fast power fluctuations. As an example, Figures 1.8a and 1.8b show the irradiance and AC power and their derivatives measured on a 10 kW rooftop installation at EPFL during a day characterized by high volatility. Irradiance fluctuations of 300  $W/m^2s$ are registered, meaning a variation of 30% in one second. This reflects in a power fluctuation of almost 3 kW in one second, for a PV plant with rated power of 10 kW. As a further example, Figures 1.9a and 1.9b respectively show daytime global horizontal irradiance measurements (GHI, recorded at the EPFL campus by using a pyranometer) and the power consumption of a group of five EPFL buildings equipped with a 95 kWp PV-roof system. As visible, GHI variations (which varies up to 85% in magnitude in less than two minutes) cause very steep fluctuations of the aggregated power. The availability of high-quality short-term GHI forecast enables the possibility of taking preemptive control actions and mitigating the effect of its fluctuations.



(a) Irradiance and its derivative

(b) AC power and its derivative
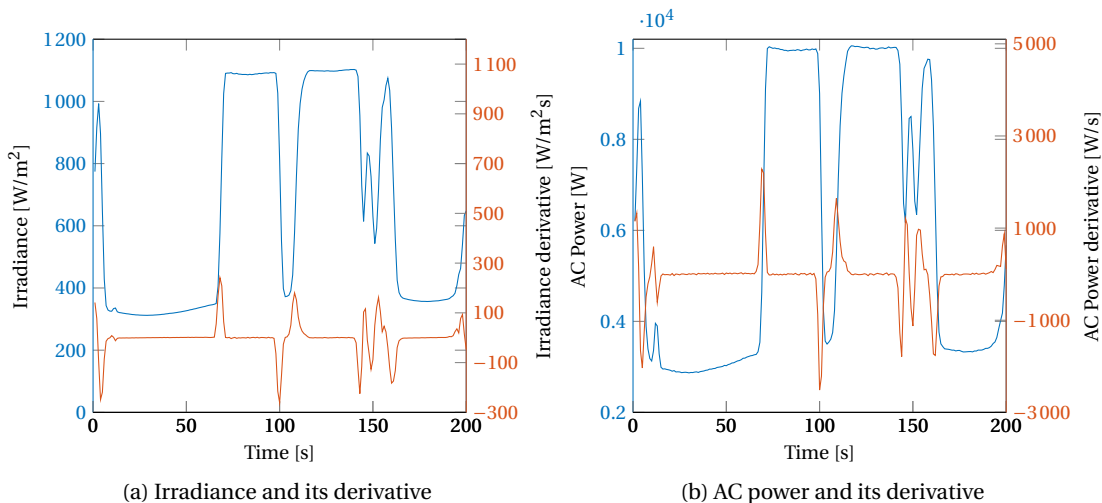
*Figure 1.8 – Irradiance and AC power variations during a partly cloudy day in July 2017, registered for a 10 kW rooftop PV installation at EPFL.*

(a) Measured GHI                    (b) Aggregated Power Profile
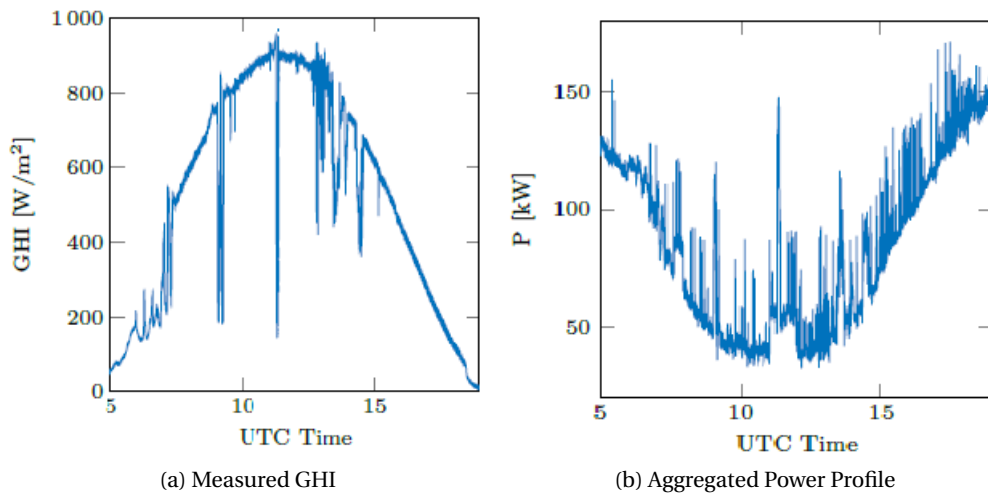
*Figure 1.9 – GHI and aggregated power profile (load and PV production) as a function of the UTC Time, registred at the EPFL campus on the 15$^{th}$ of May 2016.*

## 1.6    Solar Forecast: Definitions and Classification of the Methods

In this thesis, we refer to the term *forecast horizon* as the amount of time between the present and the effective time of the prediction, always considering one-step-ahead forecast. The *time resolution* refers instead to the sampling frequency of the time-series of reference. A vast number of papers have been proposed to forecast the maximum PV power available from a PV plant, in different perspectives.

A first classification distinguishes between direct and indirect models. In the case of direct forecast, the PV output power is predicted directly using past data, e.g. by correlating historical measurements (meteorological or electrical) with the PV power, e.g. [65]. On the contrary, indirect forecast refers to the case where the irradiance (and the temperature) are first predicted and then they are given as input to a model of the PV plant to obtain the power value, e.g. [66]. On one side, direct forecast might be a more performing approach since it does not depend on the PV plant modelling error. On the other side, the direct use of power measurements might be incorrect when, due to disturbances or control actions (e.g. active power curtailment), the measured power is not representative of the maximum availability of the plant and therefore not correlated with the weather conditions.

A second classification is based on the temporal/spatial scale targeted by the forecasting method. In general, the choice of the forecast method is strictly related to the target forecast horizon and geographical scale, [10]. Fig. 1.10 shows the different techniques adopted for different spatial resolutions and temporal horizons. Day-ahead regional irradiance forecasting relies on satellite observations and numerical weather predictions (NWPs), e.g. [67]. Statistical methods are generally applied from very short-term (second) to larger horizons (day-ahead) and low spatial resolution (< 10 km). They refer to time-series-based approaches that are able

17
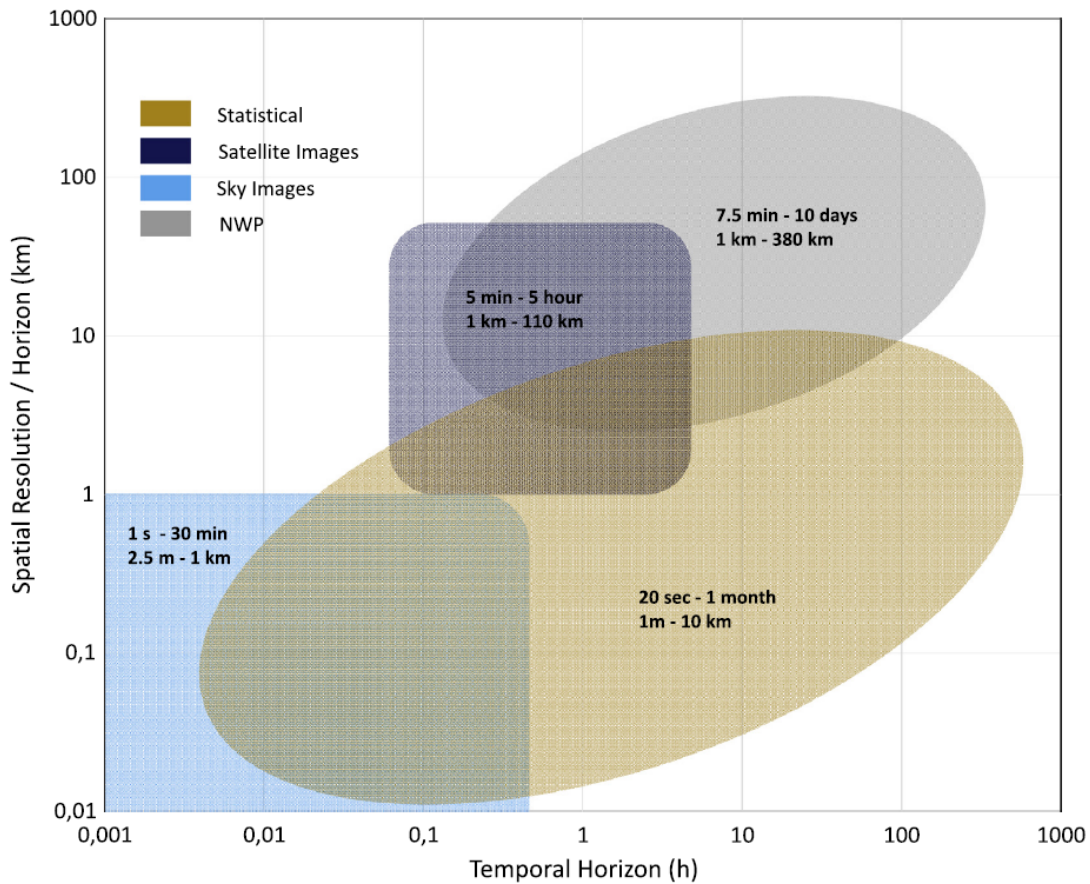
*Figure 1.10 – Forecasting techniques at different spatial/temporal scales, from [2].*

to extract relations from past data to predict the future behaviour of the plant. In general, the relation of interest is between the target variable (the PV power) and selected influential variables, also called predictors. Most of these methods are based on machine learning techniques, like artificial neural network (ANN) e.g. [68], and support vector machines (SVM) e.g. [69]. The fast advancement in the field of machine learning has led to an increasing number of available tools that, coupled with the promising performance, have made these methods particularly appealing for practitioners in the solar forecasting field. The main drawbacks are related to the need of a training set that has to be representative and reliable (and thus based on high quality data), the computational efficiency, and the tuning of the set of inputs required to be set before the training (also referred as hyperparameters). Satellites and ground images represent a valid alternative or complementary information for short-term forecast. The temporal and spatial resolutions of satellite-based estimations are generally site specific and depend on the characteristics of the available fleet of geostationary satellites. E.g., for Europe, Africa, and Indian Ocean, Meteosat-10 and 11 satellites provide real-time imagery with 1-3 km spatial resolution, and 5-15 min temporal resolution. For Asia and the Pacific region, Himawari-8 provides full disk scans with a spatial resolution in the range 0.5

to 2 km and 5 min temporal resolution. Examples of short-term satellites-based forecast can be found in [70, 71]. The use of all-sky cameras is a promising solution for intra-hour forecast horizons, [14, 72]. Apart from cloud detection and motion, sky images contain more information impacting the GHI prediction: examples are the cloud cover and the type of clouds. This kind of information can be combined with machine learning methods to compute the forecast, e.g. [73, 74]. Forecasting tools can also be classified based on the origin of the inputs. Input data can be endogenous (meaning that they are based on current and/or lagged time-series of the production of a PV plant), or exogenous (meaning that the inputs come from other measurements like temperature, wind speed, sky images, neighboring PV plants). A further classification concerns how the forecasting information is delivered: a first kind (deterministic) considers only the point forecast while the second one (probabilistic) is in form of prediction interval (PIs) and includes information accounting for the intrinsic uncertainty of the prediction. As discussed above, it is more appropriate when dealing with control and decision making in modern power systems, [75]. Regarding GHI point predictions, the simplest forecast model is the persistent one, which is commonly used as a benchmark for performance evaluation. It assumes that the GHI remains constant with the forecast horizon. In general, most of the point forecast techniques are based on AI methods. A more deterministic approach consists in detecting the position of the clouds, deducing clouds motion and calculating the time when a cloud covers the sun, e.g. [14, 76]. Several works address the problem of probabilistic forecast and propose PIs computation models. As detailed in the next paragraph, PIs give a range of possible values in which the future realization is expected to lie with a given confidence level $\alpha$, [77]. Probabilistic solar power forecast is proposed in [78, 79] where a set of likely predictions (i.e. an ensemble) is provided using a historical set of variables and deterministic meteorological models. Authors of [78] use a distance criterion to retrieve similar past forecasts, under the assumptions that their errors are likely to be similar to the errors of the current forecast. These methods refer to 0-72 hours forecast horizons, considering hourly power data. In [80], a hybrid model is proposed, integrating Support Vector Machine (SVM), ANN and sky imaging techniques to deliver real-time PIs for direct normal irradiance (DNI) for 5, 10, 15, 20 minutes ahead. At each time step, the computational time is less than 5 seconds. Another stochastic approach in [81] proposes the design of a k-nearest neighbors (KNN) algorithm. The KNN algorithm is used to predict the GHI and DNI and their uncertainty intervals, for time horizons from 5 to 30 minutes. More recently, Authors of [82] proposed a data-driven method to construct GHI probability densities for one hour-ahead predictions, using nonparametric bootstrap and a map of solar position. The developed method has low computational complexity, requiring 0.56 seconds on a personal computer. In [83], point forecasts are generated using AutoRegressive Moving Average (ARIMA) and the associated PI is calculated using a Generalized AutoRegressive Conditional Heteroskedasticity model (GARCH), considering a prediction horizon from 10 minutes to 6 hours. The use of recursive formulas, to update the model parameters in real-time, allows reducing the computational complexity of the method. The next paragraph is meant to give an insight on PIs since probabilistic forecast is mainly targeted in the thesis.

### 1.6.1 Probabilistic Forecast

Let $x_1...x_{N+1}$ be a set of random variables, a prediction interval (PI) at level $\alpha$ is defined by the upper and lower bounds $\left(I^{\downarrow\alpha}_{N+1}, I^{\uparrow\alpha}_{N+1}\right)$ such that:

$$P\left(I^{\downarrow\alpha}_{N+1}(x_1...x_N) \leq x_{N+1} \leq I^{\uparrow\alpha}_{N+1}(x_1...x_N)\right) \geq \alpha. \tag{1.10}$$

For example, if the distribution is normal and with known parameters, a prediction interval at a level $\alpha = 0.95$ would be $\mu \pm 1.96\sigma$ where $\mu$ and $\sigma$ are the mean and the standard deviation. However, there is some additional uncertainty since often the distribution is not known and the parameters need to be estimated, [84]. In general, most of the literature focuses on having a symmetric target distribution (e.g. the normal one) where the upper and lower bounds of the PI are centered at the point forecast. However, when the hypothesis that the point prediction errors follow a known distribution appears to be weak, an alternative solution is to develop *nonparametric* approaches for estimating the PIs. Distribution-free approaches do not rely on any assumption on the error-generating process, i.e. they are not based on a particular model, [77]. This is the case for photovoltaic power where empirical analysis showed that the distribution of the point errors does not generally follow a known distribution, [83, 85]. As an example, Fig. 1.11 shows the statistical distribution of the persistent point forecast error on the clear-sky index measured at EPFL, for 1 min ahead. We can see that the distribution of the errors shows a peak around the mean and asymmetry. In order to avoid assumptions on the
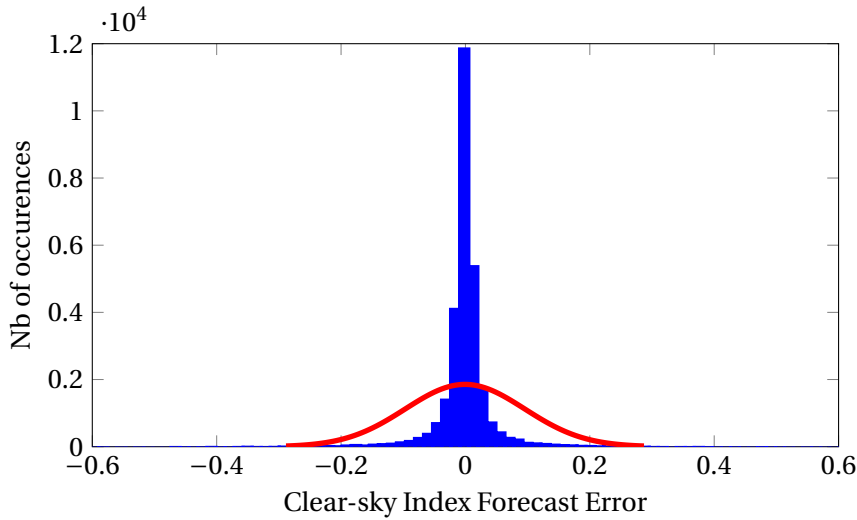


*Figure 1.11 – Statistical distribution of the persistent point forecast error.*

shape of predictive distributions, in this thesis nonparametric forecast methods are targeted. The proposed methods are then benchmarked against models assuming a given distribution of the point forecast error.

# 2 Direct Power Forecast of Photovoltaic Generation via Time-Series Analysis

## 2.1 Chapter Highlights and Summary

As mentioned in Chapter 1, direct power forecast of PV plants refers to methods that directly calculate the power output of the plant without firstly predicting the irradiance. In this Chapter, we focus on the direct AC active power forecast of small-scale photovoltaic (PV) plants for sub-second time horizons. Indeed, for power distribution systems integrating distributed PV plants, the volatility of the solar irradiance in time scales below a second can impact the grid operation (e.g. [12]). These dynamics can involve high value of the time derivative of the injected AC active power if the PV panels are coupled with inverters equipped with a Maximum Power Point Tracking (MPPT) control capable to react with time scales similar to the solar irradiance variations, e.g. [86].

This Chapter presents a method to compute prediction intervals (PIs) for PV AC active power from 100 ms up to 500 ms (ultra-short-term) and relies on the content of [87]. The original method, called Dynamic Interval Predictor (DIP), was first developed in [12] for sub-second irradiance forecast and it is based on the experimental evidence that the error of a given point forecast method is correlated with the irradiance time derivative. The DIP is here extended. First, the method is applied to directly forecast the PV AC active power by only using past AC power measurements and without relying on the irradiance measurements. Second, we consider the absolute point forecast error instead of the relative one to quantify the uncertainty of the prediction. For the sake of clarity, we name this extended version eDIP (extended DIP). Furthermore, we consider two versions of the eDIP. In the first case (Algorithm I) the explanatory variable expected to influence the prediction uncertainty is the derivative of the PV AC active power as proposed in [12]. The the second case (Algorithm II) we consider the value of the AC active power as a further explanatory variable that is expected to influence the error caused by the point forecast.

In the results section, the two algorithms are compared against benchmark models that assume instead a parametric distribution of the point forecast error and a reliability analysis of the method is also proposed. Finally, we describe how the algorithm can be embedded into

an industrial microcontroller to be adopted within a real-time control framework.

## 2.2 Ultra-Short-Term Prediction Intervals of Photovoltaic AC Active Power

In this Section the experimental setup, the observed correlations, and the method are described in details.

### 2.2.1 Experimental Setup and Observed Power Fluctuations

The experimental setup used for the validation of the proposed methodology is located at the following GPS coordinates: 46.518°N, 6.565°E. The PV plant consists of 255 W Polycrystalline modules connected to a 10 kW solar converter as follows: 2 branches with 14 modules each and one branch with 11 modules. In each branch the modules are connected in series. The DC/AC three-phase converter is equipped with DC/DC converters implementing the maximum power point tracking (MPPT). Three LEM CV 3-100 Voltage Transducers are used to measure the AC voltages while three LEM LF 205-S Current Transducers are used as AC current sensors. The output of these sensors is sampled at 50 kHz and AC phasors extraction is used to estimate the PV output power with a pace of 50 ms using the algorithm in [88]. Three different downsampling times are considered to generate series of the PV AC active power: 100 ms, 250 ms, and 500 ms. The time-series are obtained by considering instantaneous values (simple downsample by decimation). In what follows data collected between August and September 2015 are used as a validation set considering a total period of one month. Only hours with daylight are included in the evaluation. The whole time-series is shown in Fig. 2.1. Significant AC active power fluctuations are observed at sub-second time scale for our PV installation. In order to provide an example, Fig. 2.2 shows the time evolution of the AC active power along with its time derivative for a time window of 150 s, considering a period of high solar variability. For our installation, the observed fluctuations can reach time derivatives of 2 kW/s. Considering that our PV plant has a rated power of 10 kW, this represents a fluctuation of 20% of the PV capacity in about one second. In this respect, we underline that the PV power fluctuation strongly depends on the variation of the solar irradiance and on the action of the MPPT controller adopted by the PV converter.

### 2.2.2 Experimentally Observed Correlations

Similarly to [12], we show here the experimental correlation between the PV AC active power, its derivative, and the errors caused by a given point forecast method. As mentioned, the method is based on the investigation of the correlation between the error of the point forecast computation $e_{t+\Delta t}$ at $t + \Delta t$ and the derivative of AC power $\dot{P}_t$ at $t$, where $\Delta t$ is the considered
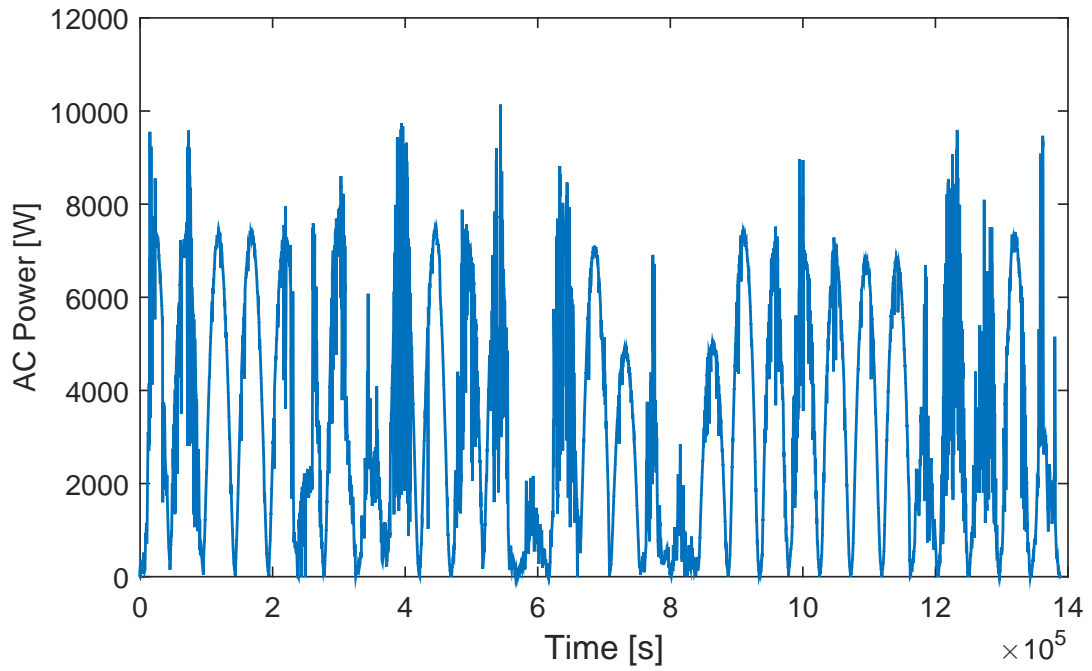
*Figure 2.1 – PV AC active power measured between August and September 2015 (30 days).*
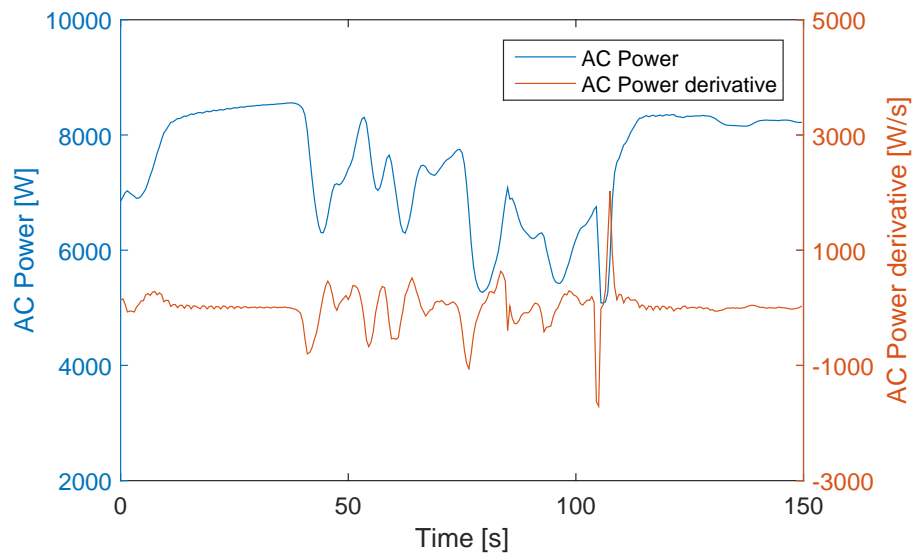


*Figure 2.2 – Example of measured PV variations: AC active power and its time derivative. Time derivatives are calculated for a time horizon of 500 ms.*

forecast horizon. The error $e_{t+\Delta t}$ is defined as:

$$e_{t+\Delta t} = P_{t+\Delta t} - \hat{P}_{t+\Delta t}, \tag{2.1}$$

where $P_{t+\Delta t}$ is the measured AC power value at $t + \Delta t$ and $\hat{P}_{t+\Delta t}$ the predicted AC power (one-step ahead point forecast). As anticipated, using the relative error, defined as $e_{t+\Delta t}/P_{t+\Delta t}$, can be misleading. Indeed, high values of the relative error can be generated by low AC power values rather than by high power fluctuations. This means that we might have significant errors even if the AC power derivative is negligible. The improvement obtained by using the absolute error instead of the relative one is shown in Section 2.3.2. The discrete time derivative is calculated as:

$$\dot{P}_t = \frac{P_t - P_{t-\Delta t}}{\Delta t}. \tag{2.2}$$

In what follows, in order to simplify the notation, we use $t \pm 1$ since the time discretization is done at constant $\Delta t$ and we refer to one-step ahead forecasts. Fig. 2.3 illustrates the 2.5%-97.5% quantile intervals of the absolute error (computed at $t + 1$) associated to different ranges of the power derivative (computed at $t$). In this example, a time horizon of 250 ms is selected. To avoid redundancy, the figure shows the absolute value of the power derivative (indeed, a symmetric behavior for negative and positive derivatives was observed). It is worth noting that different quantiles of the error are associated to different power derivative ranges. Consequently, we are interested in investigating the possibility to statistically quantify the error made by the point forecast computation (at $t + 1$) as a function of the AC power time derivative (at $t$). This consideration is the starting point of the eDIP method, presented in the next Section (Algorithm I).
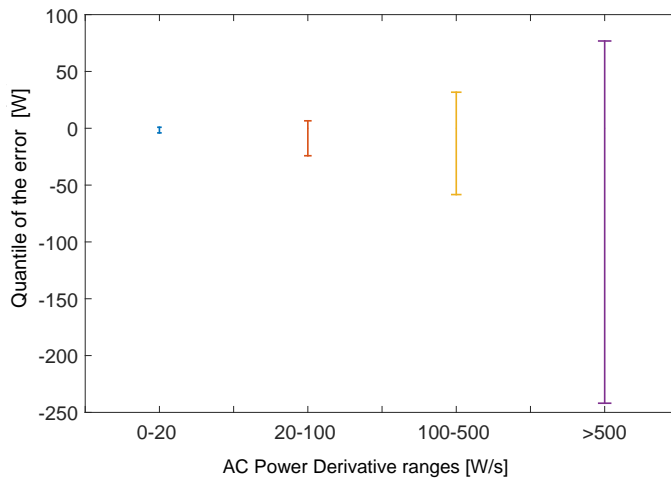


*Figure 2.3 – 2.5%-97.5% quantile intervals of the error caused by the point forecast at $t + 1$ for different ranges of the AC power time derivative at $t$. Time horizon=250 ms.*

As mentioned, we propose here an improved version of the eDIP that considers the value of the AC active power as a further explanatory variable. The method is presented in the next Section (Algorithm II). In this context, Fig. 2.4 illustrates the 2.5%-97.5% quantile intervals of
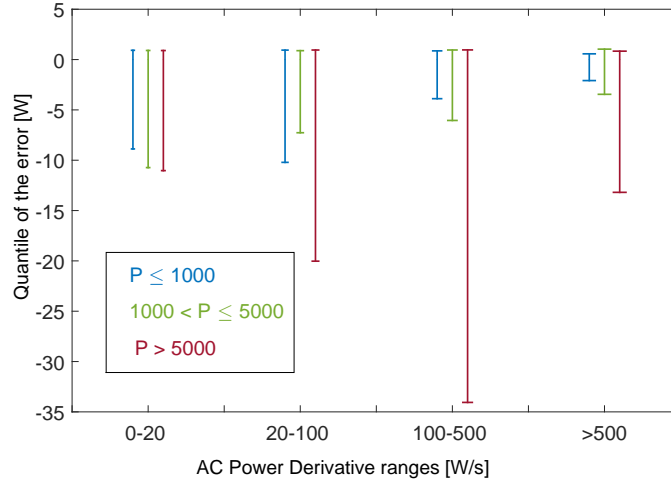


*Figure 2.4 – 2.5%-97.5% quantile intervals of the error caused by the point forecast at $t + 1$ for different ranges of the AC power time derivative at $t$. Measurements are clustered in three groups based on the AC active power value. Time horizon=250 ms.*

the absolute error associated to different ranges of the power derivative, considering different ranges of the AC active power value (for clearness we show here only three ranges of AC active power, highlighted with different colors). Data are first clustered based on the power value and then, for each group, the quantiles of the error are calculated for different ranges of the power derivative. We can observe that, for a given derivative range, the error distribution depends on the considered power range, this suggesting that a further clustering in this direction can improve the performance. This is the main assumption behind the development of Algorithm II.

### 2.2.3 Algorithm I

In this first part, we summarize the procedure for computing the PIs using only the AC power derivative as predictor. For more details, the reader is referred to the original DIP formulation in [12]. The main idea of the DIP is to use the correlations defined by the conditional distribution matrix **R** to compute the PIs. We define $e(n)$ as the $n^{th}$ discretized value of errors and $\dot{p}(m)$ as the $m^{th}$ discretized value of derivatives. Then, $R(n, m)$ corresponds to the probability that the next forecast error is $e(n)$, given that the derivative of the AC active power is $\dot{p}(m)$. The process to compute the PI is the following.

1. Define the nominal confidence level $\alpha$;

2. At the generic time step $t$, calculate $\dot{P}_t$ via Eq. (2.2);

3. Create an array of discretized values of the power derivative and of the absolute error (i.e. divide the range of the derivative and the error into uniform bins indexed $m$ and $n$, respectively). [1]

4. Find column $m$ for which the discretized value $\dot{p}(m)$ is the closest to $\dot{P}_t$;

5. Column $m$ of matrix $\mathbf{R}$ is considered as the histogram of a probability distribution that has a discrete cumulative distribution function $F()$. The matrix $\mathbf{R}$ returns the values of $F()$ at the points $e(n)$ as:

$$F(e(n)) = \sum_{n'=1}^{n} R(n', m). \tag{2.3}$$

6. Compute the upper and lower quantiles of the distribution of errors corresponding to column $m$ of matrix $\mathbf{R}$, associated to the confidence level $\alpha$. The values $e_{t+1}^{\uparrow \alpha}$ and $e_{t+1}^{\downarrow \alpha}$ are obtained by the conditions:

$$\begin{aligned} F(e_{t+1}^{\downarrow \alpha}) &= \frac{1 - \alpha}{2}, \\ F(e_{t+1}^{\uparrow \alpha}) &= \frac{1 + \alpha}{2}. \end{aligned} \tag{2.4}$$

7. The PI widths are then obtained by:

$$\begin{aligned} P_{t+1}^{\downarrow \alpha} &= e_{t+1}^{\downarrow \alpha} + \hat{P}_{t+1} \\ P_{t+1}^{\uparrow \alpha} &= e_{t+1}^{\uparrow \alpha} + \hat{P}_{t+1}. \end{aligned} \tag{2.5}$$

The PI computation is based on the knowledge of the conditional distribution matrix $\mathbf{R}$. The matrix $\mathbf{R}$ is updated at each time step and is fed by the raw sample data of the PV AC active power. The proposed update method is based on the following. Let assume that, at time $t$, we have a new observation, and let $e(n_0)$ and $\dot{p}(m_0)$ be the closest discretized values to the error $e_{t+1}$ and to the derivative $\dot{P}_t$, respectively. The update equations for matrix $\mathbf{R}^{NEW}$ are:

$$\begin{aligned} R^{NEW}(n_0, m_0) &= \frac{N^{OLD}(n_0, m_0) + 1}{N^{OLD}(m_0) + 1}, \\ R^{NEW}(n, m_0) &= \frac{N^{OLD}(n, m_0)}{N^{OLD}(m_0) + 1}, n \neq n_0, \\ R^{NEW}(n, m) &= R^{OLD}(n, m), n \neq n_0, m \neq m_0, \end{aligned} \tag{2.6}$$

with:

- $N^{OLD}(m_0)$ is the number of observations having a derivative in the range of $\dot{p}(m_0)$;

---

[1] The power derivative and the error ranges are determined using historical observations.

- $N^{OLD}(n_0, m_0)$ is the number of observations having an error in the range of $e(n_0)$ and a derivative in the range of $\dot{p}(m_0)$.

The superscripts *NEW* and *OLD* refer to the updated and the old version of the **R** matrix, respectively.

### 2.2.4 Algorithm II

In Algorithm II we add a third dimension to matrix **R**. In particular we introduce $p(l)$ as the $l^{th}$ discretized value of AC active powers that the PV plant can generate. Then, $R(n, m, l)$ corresponds to the probability that the next forecast error is $e(n)$, given that the derivative of the power is in the range of $\dot{p}(m)$ and the measured AC active power is in the range of $p(l)$. Then, the process to compute the PIs is:

1. Define the nominal confidence level $\alpha$;

2. Create an array of discretized values of the power indexed as $l$. [2]

3. Find the layer $l$ for which the discretized power value $p(l)$ is the closest to the AC power measurement $P_t$.

4. Then, the procedure corresponds to the one described for Algorithm I (numbers 2-7), but applied to the the selected layer $l$ of matrix **R**.

We assume that, at time $t$, we have a new observation and $p(l_0)$ is the closest discretized AC power value to the measurement $P_t$. Then, Let $e(n_0)$ and $\dot{p}(m_0)$ be the closest discretized values to the error $e_{t+1}$ and to the derivative $\dot{P}_t$, respectively. The update equations for matrix $\mathbf{R}^{NEW}$ are:

$$R^{NEW}(n_0, m_0, l_0) = \frac{N^{OLD}(n_0, m_0, l_0) + 1}{N^{OLD}(m_0, l_0) + 1},$$ (2.7)

$$R^{NEW}(n, m_0, l_0) = \frac{N^{OLD}(n, m_0, l_0)}{N^{OLD}(m_0, l_0) + 1}, n \neq n_0,$$

$$R^{NEW}(n, m, l) = R^{OLD}(n, m, l), n \neq n_0, m \neq m_0, l \neq l_0.$$

## 2.3 Results and Discussion

First, we show the improvement obtained by using the absolute error definition instead of the relative one to deliver our PIs. Furthermore, the reliability of the PIs is shown, proving that the method is able to guarantee the target coverage probability (i.e. the percentage of successful PIs is close to the target confidence level). Then, the eDIP is applied to three different point

---

[2]The power range varies from 0 to a maximum value, that is the PV plant rated power.

forecast techniques to prove that the algorithm can be coupled with different point forecast methods while maintaining good performance. Additionally, the eDIP is compared with two benchmark methods used to calculate the PIs. To ensure a fair comparison, the Holt Winter (HW) method is selected as point forecaster for this second analysis. The HW is based on exponential smoothing and has a simple model formulation, [89]. These first results are all obtained by implementing Algorithm I but the same conclusions can be drawn when applying Algorithm II. Then, a further analysis is carried out to compare the two proposed algorithms (I and II). For the testing results shown hereafter, PIs are constructed at a nominal confidence level of 95%. The evaluation of the performance is done by using power measurements covering the period of 30 days shown in Fig. 2.1. In order to analyze our results, it is necessary to define specific performance metrics.

### 2.3.1 Metrics

We use three standardized metrics from the existing literature to evaluate the performance of the proposed methods [90]. The first metric is the PI coverage probability (PICP) which counts the number of times that the realization falls inside the PI for a given confidence level $\alpha$:

$$\text{PICP} = \frac{1}{L} \sum_{t=1}^{L} c_t \tag{2.8}$$

where $L$ is the total number of forecast instances of the testing dataset and

$$c_t = \begin{cases} 1, & P_{t+1}^{\downarrow\alpha} \le P_{t+1} \le P_{t+1}^{\uparrow\alpha} \\ 0, & \text{otherwise.} \end{cases} \tag{2.9}$$

Then, to account for the fact that the wider the PI, the easier it is to have a realization falling inside it, we measure the PI normalized averaged width (PINAW):

$$\text{PINAW} = \frac{1}{LP_{max}} \sum_{t=1}^{L} (P_{t+1}^{\uparrow\alpha} - P_{t+1}^{\downarrow\alpha}), \tag{2.10}$$

where, for our specific plant, $P_{max} = 10$ kW. The third metric quantifies the trade-off between having a large coverage probability and small interval width. It is called coverage width-based criterion (CWC):

$$\text{CWC} = \text{PINAW}(1 + \gamma(\text{PICP})e^{-\mu((\text{PICP})-\mu)}) \tag{2.11}$$

where

$$\gamma = \begin{cases} 0, & \text{PICP} \ge \alpha \\ 1, & \text{PICP} < \alpha. \end{cases} \tag{2.12}$$

The parameter $\mu$ can be tuned based on how much bad PIs are to penalize, see [90]. Accordingly to the literature, we select here $\mu = 50$ to highly penalize wrong PIs. PIs should have high PICP (higher or equal to the target confidence level) coupled with a low value of PINAW.

### 2.3.2 Error Definition

Table 2.1 shows the performance improvement obtained by using the absolute error instead of the relative one used in [12] to build the algorithms and compute the PIs. Results are shown for Algorithm I but an equivalent improvement was found for Algorithm II (not shown here to avoid redundancy). Different columns refer to different time horizons of 100, 250 and 500 ms, respectively. We can see that, when using the relative error, the PICP is below the nominal confidence level even with a higher value of PINAW. This holds for all the considered forecast horizons. All the results shown hereafter are obtained considering the absolute error distribution to build the matrix **R** and deliver our PIs.

*Table 2.1 – Performance comparison: absolute and relative error.*

| PICP-PINAW-CWC % | | | |
|---|---|---|---|
| Time Horizon | 100 ms | 250ms | 500 ms |
| Abs. Error | 95.15-0.059-0.059 | 95.47-0.12-0.12 | 95.44-0.27-0.27 |
| Rel. Error | 94.68-0.38-0.79 | 94.75-0.55-1.13 | 94.70-0.72-1.15 |

### 2.3.3 Reliability of the PIs Delivered by the eDIP

In this Section, we further investigate the reliability of the PIs delivered by the eDIP. A method is considered reliable when the PICP is close to the target confidence level. We consider Algorithm I and a forecast horizon of 500 ms (however results can be generalized for all the horizons).

From Fig. 2.5 we can conclude that the eDIP is able to guarantee a coverage probability that is close to the target confidence level.

### 2.3.4 Independence of the Point Forecast Method

In this Section we show that the proposed eDIP can be coupled with generic point forecast methods. For this first analysis we use Algorithm I. The eDIP is applied to three different point forecast techniques:

1. Holt Winters (HW), [91].

2. The simple linear interpolation (where the linear interpolant is used to compute the
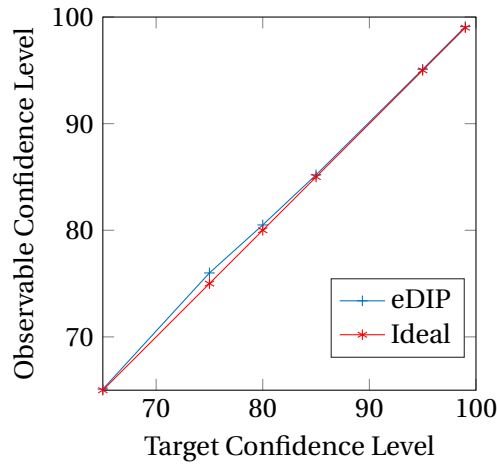
*Figure 2.5 – Reliability Diagram for the eDIP. The observed confidence level (measured as PICP) is close to the ideal case (1:1 line)*

.

prediction at the next time step).

3. The Adaptive Neuro Fuzzy Inference System (ANFIS) with a Gaussian function as mother function, [92]. For this purpose we have adopted a toolbox available in Matlab.

Table 2.2 shows the performance metrics of the eDIP applied to the above listed point forecast methods.

*Table 2.2 – Performance of the eDIP coupled with the different point forecast methods.*

| PICP-PINAW-CWC % | | | |
|---|---|---|---|
| Time Horizon | 100 ms | 250 ms | 500 ms |
| HW | 95.15-0.059-0.059 | 95.47-0.12-0.12 | 95.44-0.27-0.27 |
| Linear Interpolation | 95.04-0.14-0.14 | 94.62-0.087-0.19 | 95.20-0.24-0.24 |
| ANFIS | 95.82-0.13-0.13 | 95.76-0.13-0.13 | 96.00-0.16-0.16 |

By observing Table 2.2 it is possible to derive the following conclusions:

1. For a time horizon of 100 ms the eDIP coupled with the HW is the most performing one in terms of trade-off between high coverage probability and low interval width.

2. For 250 ms, the HW and the ANFIS return comparable performance.

3. The ANFIS point forecast method coupled with the eDIP is outperforming the other methods for forecast horizons of 500 ms.

4. In general, we have better performance for lower forecast horizons due to the lower variability. The only case where the PICP is lower than the nominal confidence level is the one using linear interpolation at 250 ms forecast horizon.

### 2.3.5 Comparison with Benchmark Methods for PI Computation

In this Section, we compare the proposed eDIP with different conventional interval predictors. For this second analysis, we use Algorithm I. We consider the following cases:

1. HW forecast method with the proposed eDIP.

2. HW forecast method with PIs based on the variance computation with Gaussian distribution of the error associated with the forecast computation.

3. HW forecast method with PIs based on the basic Bootstrap (BS), [93]. Namely, we sample (in a bootstrap kind of way) the point forecast error time-series, and we use the empirical quantiles of the error distribution (associated to $\alpha$) to compute the PIs.

The error distributions of cases 2) and 3) are updated at each additional sample.

*Table 2.3 – Performance comparison of different PIs computation methods.*

| PICP-PINAW-CWC % | | | |
|---|---|---|---|
| Time Horizon | 100 ms | 250 ms | 500 ms |
| eDIP | 95.15-0.059-0.059 | 95.47-0.12-0.12 | 95.44-0.27-0.27 |
| GAUSS | 98.30-0.35-0.35 | 97.60-0.54-0.54 | 97.71-1.80-1.80 |
| BS | 92.22-0.12-0.56 | 90.75-0.15-1.30 | 92.63-0.27-1.10 |

Table 2.3 shows that:

1. If we simply look at the CWC value (that combines the other two metrics) we can conclude that, for every time horizon, the eDIP coupled with the HW point forecast method is the most performing.

2. The HW coupled with the Gaussian error distribution has high coverage probability but this is counterbalanced by higher values of the PI width.

3. The HW coupled with the Bootstrap has low values of PINAW but PIs are penalized by a coverage probability which is lower than the nominal confidence level.

Fig. 2.6 gives a more comprehensive evaluation of the PIs obtained with the different interval predictors. Due to the large amount of data, parallel box plots are shown for the comparison.
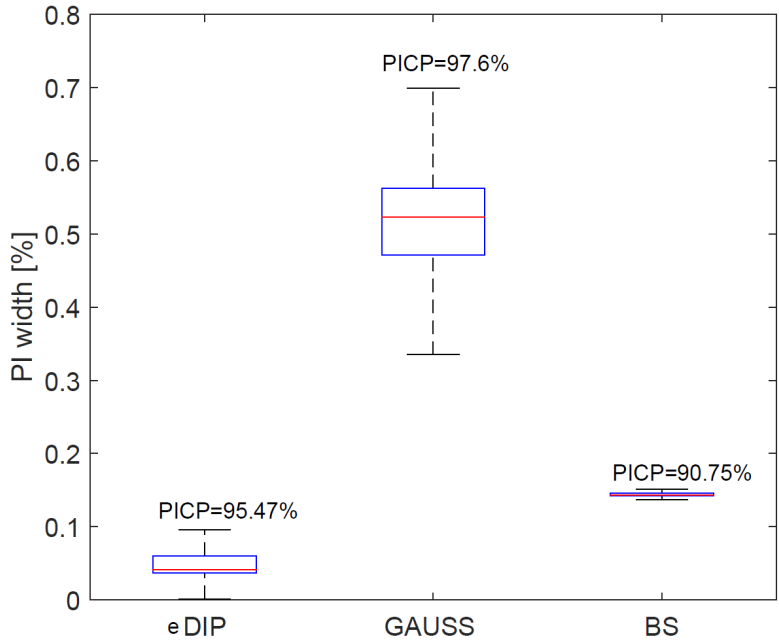
*Figure 2.6 – Box plot of the normalized interval width. 30 days of power data are considered for a time horizon of 250 ms.*

The plotted interval widths are computed such as each distance is normalized with respect to the maximum PV power output. Each box refers to one of the three different interval predictors, considering again the whole time-series of Fig. 2.1 and a time horizon of 250 ms. For each bound predictor, the red central mark represents the median, the two blue edges of the box are the $25^{th}$ and $75^{th}$ percentiles and the whiskers extend to the maximum and minimum values of the normalized PI width not considered as outliers. For the sake of completeness, the value of the PICP associated to each case is added above the relative box plot. The presented eDIP is characterized by the lowest values of the median, quantiles, and whiskers but by a higher number of outliers compared to the other two cases (outliers are not plotted here for the sake of clarity). For the eDIP, 12% of the points are outliers, reaching a maximum value of the normalized PI width of 20%. For the Gaussian case, 8% of the points are outliers with a maximum normalized PI width of 0.96%. For the Bootstrap case, 11% of the points are outliers with a maximum normalized PI width of 0.41%. The Gaussian method is characterized by a larger interquartile range, which means that the values of the normalized PI are more spread. The PIs are characterized by a higher variability and by more extreme values (represented by the whiskers). However, this is counterbalanced by a high coverage probability. For the BS case the interval widths are less dispersed but, as discussed, we have a poorer coverage probability.

### 2.3.6 Comparison between Algorithm I and II

Table 2.4 shows a comparison between algorithms I and II. We select the HW as point forecast method and the AC active power value is discretized in 10 equally spaced intervals ($l$=10).

Table 2.4 shows that, for each time horizon, the addition of a third dimension accounting

*Table 2.4 – Performance comparison between Algorithm I and II.*

| PICP-PINAW-CWC % | | | |
|---|---|---|---|
| Time Horizon | 100 ms | 250 ms | 500 ms |
| Alg. I | 95.15-0.059-0.059 | 95.47-0.12-0.12 | 95.44-0.27-0.27 |
| Alg. II | 95.14-0.055-0.055 | 95.46-0.11-0.11 | 95.31-0.24-0.24 |

for the AC active power as further influential variable, allows improving the performance, in terms of trade-off between PICP and PINAW. From a posteriori analysis, we noticed that performance improves when increasing the number of discretization intervals up to $l$=200, see Table 2.5. Increasing $l$ leads to a lower PINAW while keeping high value of PICP (higher than $\alpha$). However, for $l$ higher than 200 we have that PICP decreases below the value of $\alpha$, this reducing the quality of the PIs.

The improvement introduced by Algorithm II is particularly evident when encountering high dynamics. This is firstly suggested by Fig. 2.4 and then confirmed by experimental evidence. As an example, Fig. 2.7 illustrates the time evolution of the proposed eDIP during high PV AC active power fluctuations, for the two algorithms. In both cases the eDIP tries to adapt the width of the PI in order to keep the AC power measurements inside the interval. It is possible to see that Algorithm II succeeds and returns PIs that are narrower than those given by Algorithm I.

*Table 2.5 – Performance of Algorithm II for time horizon of 500 ms and different discretizations of the AC active power.*

| PICP-PINAW-CWC % | | | | |
|---|---|---|---|---|
| $l = 5$ | $l = 50$ | $l = 100$ | $l = 200$ | $l = 400$ |
| 95.3-0.26-0.26 | 95.5-0.21-0.21 | 95.5-0.18-0.18 | 95.1-0.16-0.16 | 94.4-0.12-0.27 |

### 2.3.7 Deployment into an Industrial Hardware Platform

In order to discuss the applicability of the proposed eDIP, the method (Algorithm II) is deployed into an industrial microcontroller. In particular, the eDIP has been working with a real-time measurements system since July 2015 as a forecasting tool for the PV plant described in the experimental setup. The predictor is deployed in a National Instrument CompactRIO 9068, composed by a reconfigurable Artix-7 FPGA, a dual-core ARM Cortex-A9 processor, a 512 MB DDR3 memory and equipped with a NI Linux Real-Time OS. Real-time measurements are acquired by using analog input modules (NI-9215), characterized by an input range of ± 10 V and with a maximum sampling frequency of 100 kHz. The **R** matrix is updated at each sample and there is no need to store past data. When a new power measurement is ready, the matrix is updated and it is available for the following PI computations. At each iteration, the whole

(a) Algorithm I        (b) Algorithm II
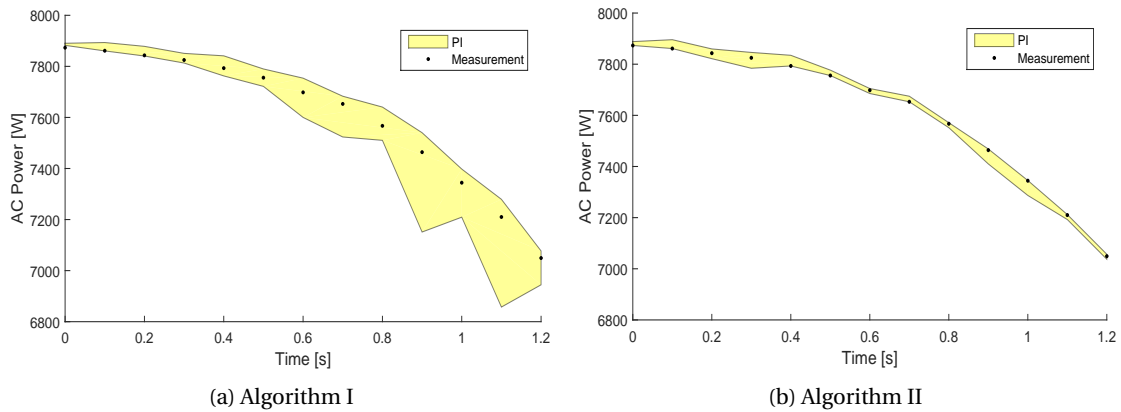
*Figure 2.7 – PIs (yellow shading) and AC active power measurements (black dots) are plotted for the two algorithms and for high AC power derivative.*

process including the update of the matrix and the calculation of the PI takes less than 1 ms.

## 2.4 Chapter Conclusions

This Chapter focuses on the computation of sub-second PIs to directly forecast the AC active power output of PV panels equipped with MPPT controllers. The proposed method, based on the theory developed in [12], uses the experimentally observed correlation between the AC active power derivative and the error caused by a generic point predictor. An improved approach, accounting for the value of the AC active power as additional explanatory variable, is also presented. The validation dataset consists of an experimental time-series of PV power measurements of 30 days. As in [12], the method does not require any hypothesis on the distribution of the error nor any specific point forecast technique and it is able to account for high AC active power fluctuations by adapting the PI width and while maintaining the target coverage probability. The eDIP is compared with other benchmark methods for PIs (i.e., Gaussian and Bootstrap). Results from this comparison show that the eDIP exhibits the best performance in terms of trade-off between high coverage probability and low interval width. A further analysis proves that accounting for the AC active power value, as additional influential variable, allows improving the performance in particular when encountering high variations. The proposed method has been deployed into an industrial hardware platform and efficiently operates at sub-second update speed for the real-time computation of PIs of a rooftop PV plant.

# 3 Solar Irradiance Forecast via Time-Series

## 3.1 Chapter Highlights and Summary

Forecasting the global horizontal irradiance (GHI) is usually the first step for predicting the power generated by the PV plant. Then, the GHI information can be converted into power values by using transposition models (if the PV panels are tilted) and physical models of the PV plants. In particular, indirect maximum power models are preferable to direct ones (described in Chapter 2) when the power measurements are uncorrelated with the true PV maximum power and, consequently, with the true irradiance (e.g. due to the action of curtailment strategies). The direct use of production measurements in time-series-based forecasting tools might lead to mistreaning the algorithm, as later discussed in Chapter 5.

This Chapter focuses on GHI forecasting based on time-series. It describes a nonparametric tool for computing ultra-short and short-term prediction intervals (PIs) of the global horizontal irradiance (GHI) based on the clustering of historical observations of the clear-sky index. The main novelty of the proposed method is the ability to deliver probabilistic PIs from sub-second (500 ms) to intra-hour (5 min) forecasting horizons, needed by microgrid applications. At first, the clear-sky model is introduced and its importance discussed. Then, the influential variables and the method to derive the PIs are described, and the advantage of differentiating the clear-sky index time-series is also presented. Results are presented by comparing the proposed method with state-of-the-art prediction models, including the eDIP described in the previous Chapter (this time applied to predict the solar irradiance rather than the PV AC power directly).

## 3.2 Clear-sky Model

GHI measurements are pre-processed in order to remove the daily and seasonal components due to changes of the sun position. This is achieved by introducing the clear-sky index $K$, which is defined as the ratio between the measured GHI and the clear-sky irradiance, respectively denoted by $I$ and $I_{cs}$. Clear-sky models allow deleting deterministic fluctuations

due to changes in the solar position. The clear-sky index is defined as the ratio between the measured GHI and the modeled clear-sky irradiance at ground level:

$$K = \frac{I}{I_{cs}}.$$  (3.1)

The clear-sky irradiance is the irradiance that would reach the ground in clear-sky conditions, i.e. absence of clouds. In this section, the clear-sky irradiance is obtained by applying the clear-sky model implemented in the geographical information system GRASS, which also accounts for topological shading [94, 95].

## 3.3 Influential Variables

The proposed PI estimation method consists in clustering historical data according to the value of certain influential variables, introduced in the following. The clusters are therefore used as empirical conditional probabilities of future realizations and used to compute the PI by calculating the quantiles according to a given confidence level. In particular, these influential variables should be representative of the irradiance fluctuations since it is the main cause of the uncertainty associated with solar forecasts. These variables, inputs of the clustering process, are selected according to the literature that considers the average and the variability of the clear sky-index as the most influential ones, [81, 96]. We consider a training dataset of historical clear-sky index observations $K_1, \ldots, K_N$, from which we extract the following influential variables:

- the average clear-sky index value on a mobile window of length $n$ considering the most recent data points:

$$M_i = \frac{1}{n} \sum_{j=i-n+1}^{i} K_j, \ i = n+1, \ldots, N$$  (3.2)

  of which we consider the normalized version $M_i^*$. Namely, we normalize the sequence $M_0, \ldots, M_1$ to a length of 1; [1]

- the clear-sky index variability:

$$V_i = \sqrt{\frac{1}{n} \sum_{j=i-n+1}^{i} (K_j - K_{j-1})^2}, \ i = n+1, \ldots, N$$  (3.3)

  which is a measurement of GHI fluctuations. As for the previous case, we consider the normalized version $V_i^*$.

---

[1] The normalized version of a vector $\mathbf{X}$ is a vector $\mathbf{X}^* = \mathbf{X}/||\mathbf{X}||$, where $||\mathbf{X}||$ is the 2-norm of $\mathbf{X}$.

Normalization of the influential variables is required to enable a fair comparison between parameters with different scale. For each observation, the vector $\mathbf{p_i}$ of influential variables is:

$$\mathbf{p_i} = (M_i^*, V_i^*), \quad i = n+1, \ldots, N. \tag{3.4}$$

The process to compute the PIs is performed in two ways:

- Method A: we cluster the original clear-sky index time-series;

- Method B: we cluster the differentiated clear-sky index time-series:

$$\Delta K_i = K_i - K_{i-1}, \ \ i = 2, \ldots, N, \tag{3.5}$$

    to verify if differencing leads to better prediction performance.

## 3.4  Clustering of the Training set

The k-means iterative algorithm is firstly used to classify historical observations of the clear-sky index according to predefined influential variables[2]. K-means clustering, [97], is a partitioning algorithm that allocates each observation into one and only one of the $k$ clusters, each one defined by a representative centroid. In particular, $k$ centroids are at first randomly selected (the first centroids are simply uniformly random observations). Then, each vector of the training dataset is assigned to the closest centroid, and the centroid is iteratively recalculated as the mean of the vectors of each class until convergence is reached (i.e., centroids do not change anymore between iterations).

### Method A

We apply the k-means algorithm to cluster the vectors $\mathbf{p_i}$ belonging to the training set, being $k$ the number of clusters. The algorithm assigns to each vector $\mathbf{p_i}$ a cluster index $l$ between 1 and $k$ and determines the centroids locations $\mathbf{c_l} = (M_l^*, V_l^*)$ for $l = 1, \ldots, k$. We denote the generic cluster $G_l$ as composed by all the clear-sky indexes $K_{i+1}$ for which $\mathbf{p_i}$ has index $l$.

### Method B

We apply the same clustering procedure described above for Method A. However, we denote the generic cluster $\Delta G_l$ as composed by all the differentiated clear-sky realizations $\Delta K_{i+1}$ for which $\mathbf{p_i}$ has index $l$.

---

[2]The k-means clustering has shown better performance when compared with other clustering techniques as DBSCAN

An example of the k-mean partitioning of the influential variables is shown in Fig. 3.1 where the normalized clear-sky index average and variability are clustered. It is worth noting that the k-means clustering of the training dataset can be performed off-line on historical data. This is a key aspect if considering the high reporting rate of predictions for microgrid applications since it allows reducing the computational complexity.
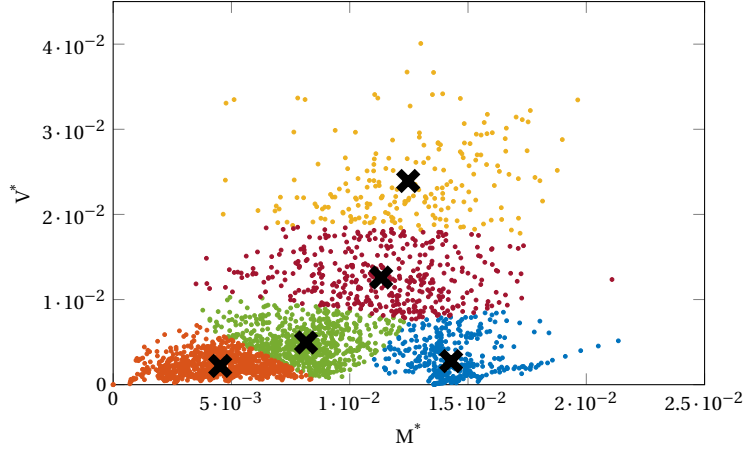


*Figure 3.1 – Example of k-means clustering obtained for k = 5. The x and y axis represent the normalized clear-sky index average and variability, respectively. The black marker signs the centroids.*

## 3.5 Prediction Intervals Computation

In this Section we describe how PIs are computed, distinguishing between the two proposed methods.

**Method A**

Starting from the clusters $G_1, \ldots, G_k$ defined in the previous section, the PIs at the target confidence level $\alpha$ can be computed as:

$$q_l^{\uparrow \alpha} = (1 + \alpha)/2 \text{ quantile of } G_l, \ l = 1, \ldots, k \tag{3.6}$$

$$q_l^{\downarrow \alpha} = (1 - \alpha)/2 \text{ quantile of } G_l, \ l = 1, \ldots, k. \tag{3.7}$$

For increased computational efficiency, we note that also this operation can be performed off-line, and the PIs for each class can be stored. Say being at time $t$, the objective is to perform the on-line computation of the PI for the next time interval $t + 1$. The vector of influential variables at $t$ is denoted by $\mathbf{p_t} = (M_t^*, V_t^*)$. It is calculated by normalizing the raw influential variables $M_t, V_t$ with respect to those available in the training dataset. The next step is the

calculation of the Euclidean distances between $\mathbf{p_t}$ and the centroids $\mathbf{c_l}$:

$$d_l = \|\mathbf{c_l} - \mathbf{p_t}\|^2, \quad l = 1,...,k \tag{3.8}$$

which is used as a similarity criterion to select the cluster representative of the future clear-sky outcome. We indicate with $\hat{l}$ the index corresponding to the cluster with minimum distance. It is used to select the quantiles used in the PI computation as:

$$K_{t+1|t}^{\uparrow\alpha} = q_{\hat{l}}^{\uparrow\alpha}. \tag{3.9}$$
$$K_{t+1|t}^{\downarrow\alpha} = q_{\hat{l}}^{\downarrow\alpha} \tag{3.10}$$

## Method B

Starting from the clusters $\Delta G_1, \ldots, \Delta G_k$ obtained from the differentiated GHI time-series, the PIs at the target confidence level $\alpha$ can be computed as:

$$q_l^{\downarrow\alpha} = (1-\alpha)/2 \text{ quantile of } \Delta G_l, \quad l = 1,...,k \tag{3.11}$$
$$q_l^{\uparrow\alpha} = (1+\alpha)/2 \text{ quantile of } \Delta G_l, \quad l = 1,...,k. \tag{3.12}$$

Also in this case, the quantiles extraction is computed off-line. The on-line computation of the PIs consists in finding the index $\hat{l}$ of the cluster with centroid at the minimum distance from $\mathbf{p_t}$. It is used to select the quantiles used in the PI computation as:

$$K_{t+1|t}^{\uparrow\alpha} = K_t + q_{\hat{l}}^{\uparrow\alpha}, \tag{3.13}$$
$$K_{t+1|t}^{\downarrow\alpha} = K_t + q_{\hat{l}}^{\downarrow\alpha}. \tag{3.14}$$

i.e., the current measurement is summed to the upper and lower quantiles of the differentiated time-series.

It is important to note that, so far, PIs are computed to forecast the clear-sky index. The last step consists in computing the PI for the GHI:

$$I_{t+1|t}^{\uparrow\alpha} = K_{t+1|t}^{\uparrow\alpha} I_{cs,t+1}, \tag{3.15}$$
$$I_{t+1|t}^{\downarrow\alpha} = K_{t+1|t}^{\downarrow\alpha} I_{cs,t+1}. \tag{3.16}$$

In the results section, performance of methods A and B are evaluated by comparing the estimated PIs as defined in (3.15)-(3.16) with the GHI measurements.

## 3.6 Selection of the Parameters

The hyper-parameters we need to specify when applying the k-means clustering procedure are:

- The number of samples $n$ used in equations (3.2) and (3.3);

- The number of cluster $k$ used for the partition of the training dataset;

- The length of the training dataset $N$.

The selection of the parameters values is a sensitivity process that is exhaustively evaluated in Section 3.9. The assessment is performed in a searching dataset, and then the selected values are applied in a testing dataset for performance evaluation. While it was seen from the results that variations of $n$ in the range from 2 to 5 do not, in general, alter modeling performance, the values of $k$ and $N$ are interdependent and the selection of these two parameters needs to be carried out simultaneously (i.e. we need to find the combination of $k$ and $N$ with the best performance). For each specific case, a diagnostic analysis should be performed. We consider here two main approaches to fix $k$ and $N$ as described in the next sections.

### 3.6.1 A-posteriori Selection of $k$ with Exhaustive Search (ES)

Several $(k, N)$ combinations are attempted in a searching dataset: the candidate combination is the one with best a-posteriori prediction performance. These values are then used to evaluate the performance of the testing dataset, supposing that it exhibits similar characteristics of the searching set. This approach can become computationally expensive, therefore motivating the development of methods for the a-priori selection of the free parameters, as described in the next Section.

### 3.6.2 A-priori Selection of $k$ with Silhouette Analysis (SA)

The objective is to use Silhouette Analysis, [98], to improve the partitioning of the training dataset, allowing for an a-priori selection of parameter $k$ and in order to avoid the exhaustive approach. The Silhouette Analysis consists in some main steps:

- a small value of $k$ is chosen (e.g. $k = 5$) and the clustering algorithm is run;

- the silhouette value for a generic point $i$ is calculated as:

$$s(i) = \frac{a(i) - b(i)}{\max(a(i), b(i))} \tag{3.17}$$

where $a(i)$ is the average distance from point $i$ to the other points in the same cluster, while $b(i)$ is the minimum average distance from instance $i$ to points in a different

cluster, minimized over clusters. Parameter $s$ is a measure of how close the instance is to the other instances in its cluster and how far it is to those in the other clusters. In general, a silhouette value close to 1 is desired because it means that the point is well clustered while a value close to -1 means misclassification;

- the mean of the silhouette values is computed. If most points have a high silhouette value, then the clustering is appropriate;

- the value of $k$ is augmented and it is evaluated if having more clusters allows for a better partitioning (higher mean of the silhouette values);

- the number of clusters is selected equal to the value $k$ above which we do not see any improvement in terms of increasing of the average silhouette value.

As it is shown in Section 3.9.3, the value of $N$ required to converge at constant $k$ is not sensitive to the characteristics of the dataset. In general, for each forecast horizon and given $k$, it is possible to identify a value of $N$ above which performance is close to convergence. Above this value, small oscillations are explained by the intrinsic stochasticity of the data. This feature is important for the modeler since it allows for the reduction of the parameters to be found.

## 3.7 Algorithms Time Complexity

In this Section, we evaluate the time complexity of the real-time computation of the proposed algorithms. This is an important aspect because they are designed with the stated objective of delivering PIs to real-time control processes for electrical power systems.
The algorithms consist of two parts, the training phase and on-line computation of PIs. The former does not have any real-time requirement and can be performed off-line. The latter phase is instead time critical, and it is to perform with a hard-real deadline. First, it consists in calculating the normalized influential variables, (3.2)-(3.3), an operation with constant time complexity, $O(1)$, which involves algebraic operations. Then, we have the computation of $k$ norms, $O(k)$, and a minimum search, which can be performed efficiently with a merge search, $O(k\log(k))$. Considering that $k$ is fixed by design, the time complexity of the real-time computation is constant time, $O(1)$, therefore denoting that complexity does not grow with the size of the problem (scalability). Statistics on the execution time of the algorithms are given in Section 3.9.8.

## 3.8 Data

### 3.8.1 Data Acquisition

Global horizontal irradiance measurements are collected at 50 ms resolution [3] by using an Apogee SP-230 all-seasons pyranometer which is located at the GPS location given in Section

---

[3] We note that the pyranometer response time is less than 1 ms.

2.2.1. Silicon-cell pyranometers are sensitive to the wavelength range of approximately 350-1100 nm. Spectral errors appear when measuring radiation in conditions that are different from calibration (namely, clear-sky conditions, solar zenith angle of approximately 45 degrees, and atmospheric air mass of 1.5). The Apogee pyranometer has a spectral error up to 6% at solar zenith angles of 75 degrees, up to 5% for air mass of 8, and up to 9% for cloudy conditions. We consider three datasets of 85 days each, corresponding to different periods of the year. The first contains irradiance measurements from July to September 2015 (Summer), the second from October to December 2015 (Autumn), and the third from January to March 2016 (Winter). Each dataset is divided into a searching subset of 55 days and a testing one composed of the remaining 30 days. The original time-series is downsampled to three different resolutions: 500 ms, 1 min, and 5 min. These series are used to compute one-step-ahead PIs for the corresponding forecast horizon. Downsampling is computed by averaging the intermediate samples. It is worth noting that applying the clear-sky normalization causes very high values of $K$ close to sunrise and sunset. Therefore, we consider only daylight values covering the period of the day for which the clear-sky index does not diverge.

### 3.8.2   Data Classification

Characterizing the dataset is important for performance comparison and evaluation. Indeed, the robustness of the method should be tested during periods of different irradiance volatility. In our case, we are interested in characterizing the three available datasets: Summer, Autumn, and Winter. First, we give an information regarding the weather of the selected period and location. In particular, we retrieve cloud cover data [4] from MeteoSwiss Idaweb services, from two weather stations in the vicinity of our installation. The average cloud okta values for the three seasons are: 3.86 okta in Summer, 4.67 okta in Autumn and 5.96 okta in Winter. Second, we introduce a criterion consisting in counting the percentage of periods with a volatility lower than a given threshold. For each timestep $t$, we calculate the per-unit difference as $\Delta I_t = (I_t - I_{t-1})/I_{max}$, where $I_{max} = 1000$ W/m$^2$. For each prediction horizon, we establish a threshold for $\Delta I_t$, above which the observation at time $t$ is considered with high volatility. The threshold is empirically computed as the 99% quantile of the $\Delta I$ time-series obtained by manually selecting a period of 3 clear-sky days. The values are shown in Table 3.1 for different forecast horizons.

*Table 3.1 – Thresholds.*

| Time Horizon | Thresholds |
|:---:|:---:|
| 500 ms | 0.0004 |
| 1 min | 0.011 |
| 5 min | 0.025 |

---

[4]Cloud cover corresponds to the fraction of the sky obscured by clouds when observed from a given location. The unit of measurement is the okta, ranging from 0 (completely clear-sky) to 8 (completely overcast).

*Table 3.2 – Percentage of periods with high irradiance volatility.*

| Season | Forecast Horizon | | |
|---|---|---|---|
| | 500 ms | 1 min | 5 min |
| Summer | 16 | 17 | 22 |
| Autumn | 5 | 9 | 12 |
| Winter | 13 | 15 | 20 |

The percentage of periods exceeding the threshold of GHI high volatility is shown in Table 3.2, for each dataset and for different forecast horizons. As it can be observed, the Summer period is characterized by the highest GHI volatility, followed by Winter and Autumn.

## 3.9 Results and Discussion

First, Section 3.9.2 shows the advantage given by the introduction of a clear-sky model at different forecast horizons. Then, in Section 3.9.3 the sensitivity of the performance with respect to the selection of the model parameters is discussed. In Sections 3.9.4-3.9.6 the performance of the proposed methods is benchmarked against existing techniques. First, we compare the proposed methodology with the symmetric quantile extraction, which is the simplest way to construct our intervals. We use the empirical quantiles extracted from the distribution of the time-series to build the PIs as in (6.4)-(6.5) and (6.7)-(6.8), respectively. It is important to highlight that the quantiles at time $t$ are extracted from the whole time-series, from $t = 0$ to $t - 1$. As a second benchmark, we compare our method with a model commonly used in forecasting. We first generate a point forecast using AutoRegressive Moving Average model (ARIMA), [99], with Double Exponential Smoothing. Then PIs are constructed assuming a Gaussian distribution of the point forecast error as:

$$K_{t+1|t}^{\uparrow \alpha} = \hat{K}_{t+1} + \eta_\alpha \sqrt{\sigma_t^2}, \tag{3.18}$$

$$K_{t+1|t}^{\downarrow \alpha} = \hat{K}_{t+1} - \eta_\alpha \sqrt{\sigma_t^2}. \tag{3.19}$$

where $\hat{K}_{t+1}$ is the point forecast obtained by the ARIMA model, $\eta_\alpha$ is the quantile of the normal distribution corresponding to the target confidence level $\alpha$ and $\sigma_t^2$ is the variance of the forecast error.

Unless otherwise indicated, the target confidence level used for the following analysis is fixed equal to 95%. Section 3.9.7 presents and discusses the reliability diagrams. Finally, statistics of the method execution time are provided in Section 3.9.8.

### 3.9.1 Metrics

We use the three standardized metrics defined in Section 2.3.1 to evaluate the PIs. In this case, the normalising factor is $I_{max} = 1000 \, \text{W/m}^2$. The value has been selected accordingly to the work in [81], that is later used as benchmark for performance comparison.

### 3.9.2 Clear-sky index and GHI Time-series Comparison

A first analysis aims at assessing the difference between using the measured irradiance (GHI) or the clear-sky index ($K$) time-series as inputs for the PIs computation method. In particular, we apply methods A and B and we increase the forecast horizon to evaluate when the inclusion of a clear-sky model becomes advantageous. Fig. 3.2 shows that the CWC is, in general,



(a) Method A               (b) Method B
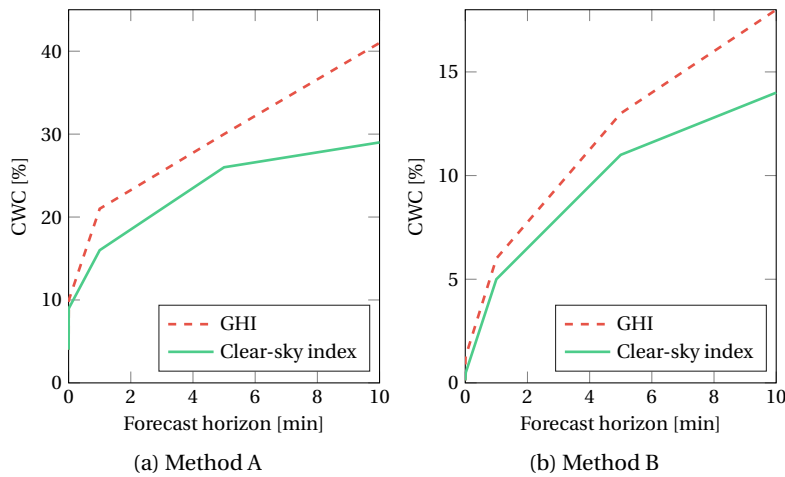
*Figure 3.2 – CWC as a function of the time horizon. Comparison between the use of the GHI time-series (original) and the clear-sky index one (normalized).*

lower (better performance) when using the clear-sky index. In particular, the advantage of using a clear-sky model becomes evident for time horizons longer than 1 min. As expected, the normalization of the time-series becomes more important at higher forecast horizons, when the effect of the sun position becomes more dominant. When referring to ultra-short-term horizons, fluctuations of solar irradiance are mainly related to cloud motion and the importance of a clear-sky model becomes marginal. Since the inclusion of a clear-sky model only leads to similar or better prediction performance, the proposed methods are applied to the clear-sky index time-series.

(a) 500 ms      (b) 1 min      (c) 5 min

(d) 500 ms      (e) 1 min      (f) 5 min
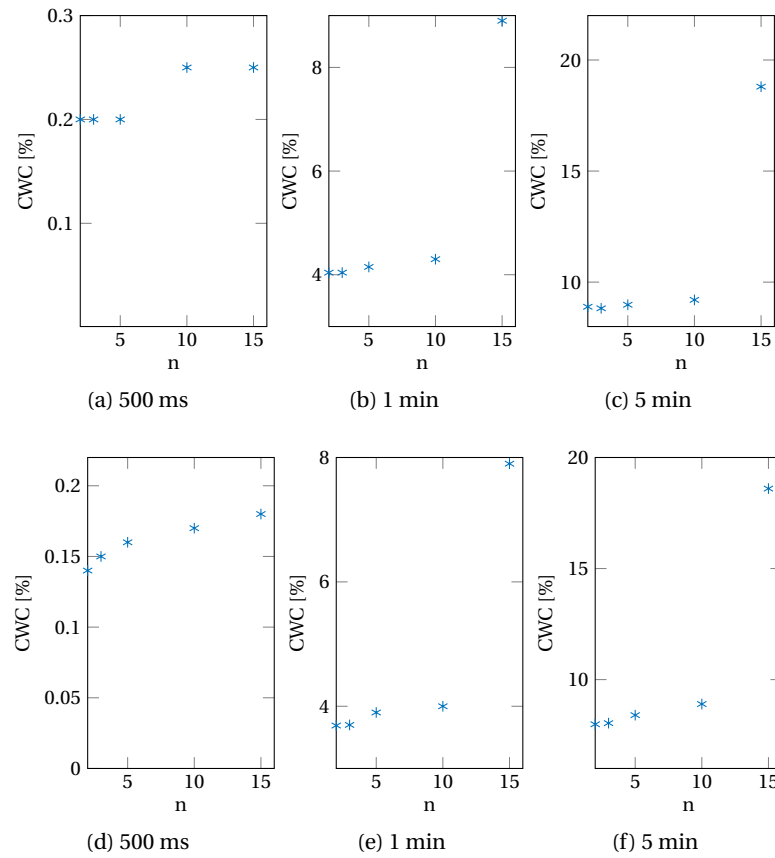
*Figure 3.3 – CWC as a function of n for different forecast horizons. N is equal to 30 days. k is equal to 5 for cases (a)-(c) and equal to 30 for case (d)-(f).*

### 3.9.3 Parameters Selection and Sensitivity Analysis

**Selection of Parameter $n$**

Fig. 3.3 shows the CWC metric as a function of $n$ (Eq. 3.2-3.3) obtained from a-posteriori analysis of the performance of the whole Autumn dataset. Method B is applied. The analysis considers different forecast horizons for both $k$=5, cases (a)-(c), and $k$=30, cases (d)-(f). It is possible to see that performance is not very sensitive to variations of $n$ in the range from 2 to 5. From an a-posteriori analysis of our datasets at different forecast horizons, we can conclude that $n$ can be fixed to a value between 2 and 5, for all the considered cases. Indeed, analogous conclusions can be inferred for Method A and different datasets (not shown here because of the similar behavior). These values of $n$ are a good trade-off between having enough significant measurements to compute the influential variables and avoiding to consider realizations that are too far from the actual conditions. The results presented in what follows are obtained with $n = 3$.

**Selection of Parameter $k$ and $N$**

As introduced in Section 3.6, two main procedures are proposed to assign parameters $k$ and $N$. In the case of the exhaustive search, we isolate 55 days of each dataset to perform the exhaustive searching. The number of clusters $k$ could potentially vary between one (single cluster) and the total number of training samples (each data is assigned to its own cluster). Since the computational effort of the k-means algorithm is linearly dependent to the number of clusters and to the number of data, we here limit $k$ to 1000 and $N$ to 30 days. The dashed lines in Figures 3.4, 3.5 and 3.6 show the value of $k$ that returns the best performance for a fixed $N$. This $k$ is obtained a-posteriori by applying methods A and B to the searching dataset of 55 days ($k_{ES,A}$ and $k_{ES,B}$, respectively). The three figures refer to different datasets. Each figure includes three sub-figures referring to the three forecast horizons, respectively. It is possible to note that, due to the heuristic nature of the methods, the optimal value of $k$ cannot be known a-priori and it varies among the different considered cases (namely, we do not have a global optimum). Therefore, we select $k$ and $N$ as the combination returning the best prediction performance (minimum CWC), found a-posteriori. These values found for the searching set of 55 days are then applied for performance evaluation in the remaining 30 days. In the case of the silhouette analysis, $k$ is calculated for different $N$ as the one maximizing the average silhouette of the training set and it is shown in Figures 3.4, 3.5 and 3.6 with the solid line. The value of $k$ that maximizes the average silhouette does not correspond to the one returning the best forecasting performance. However, we notice that its value does not vary with $N$ and can be selected independently. In order to determine $N$ in the case of the silhouette analysis, we fix $k$ equal to the one returned by the analysis ($k$=5) and we evaluate the prediction performance for a different number of training days. Figures 3.7, 3.8 and 3.9 show the CWC (in logarithmic scale) as a function of $N$. They refer to 500 ms, 1 and 5 min, respectively. Each figure consists of two plots, showing the performance for Method A and B, respectively. We can make the following observations:

- For each forecast horizon, it is possible to identify a first drop of CWC after which performance tends to stabilize. The value of $N$ that leads to performance stabilization is not sensitive to the dataset and can be fixed independently. On the contrary, as it is shown in the next sections, performance at convergence depends on the nature of the dataset and, in general, the behavior of the PIs depends on the volatility content of the dataset.

- For the sub-second time horizon and Method A we have a first drop of CWC after about one day, and then performance tends to stabilize. Method B reaches convergence after few hours of training with subsequent small CWC oscillations.

- For time horizon of 1 minute we have a first drop of CWC after about 5 days of training and then performance smooths out more slowly.

- For time horizon of 5 minutes we have a first drop of CWC after about 5 days of training and then a second drop after 10 days. Then, performance smooths out more slowly.

(a) 500 ms      (b) 1 min      (c) 5 min
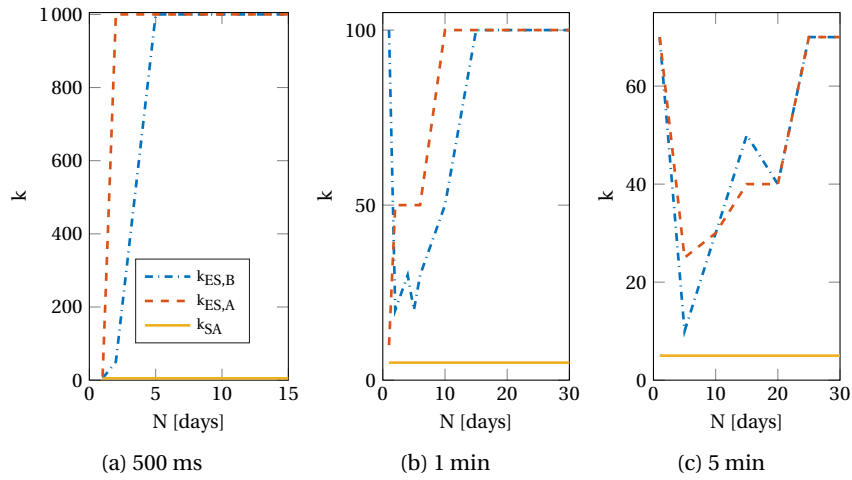
*Figure 3.4 – Number of clusters k as a function of the length of the training dataset. The Summer dataset is selected for the analysis. The dashed lines refer to the value of k corresponding to maximum performance for Method A and B, respectively. It is calculated a-posteriori by applying the ES. The solid line refers to the value of k from the SA.*



(a) 500 ms      (b) 1 min      (c) 5 min
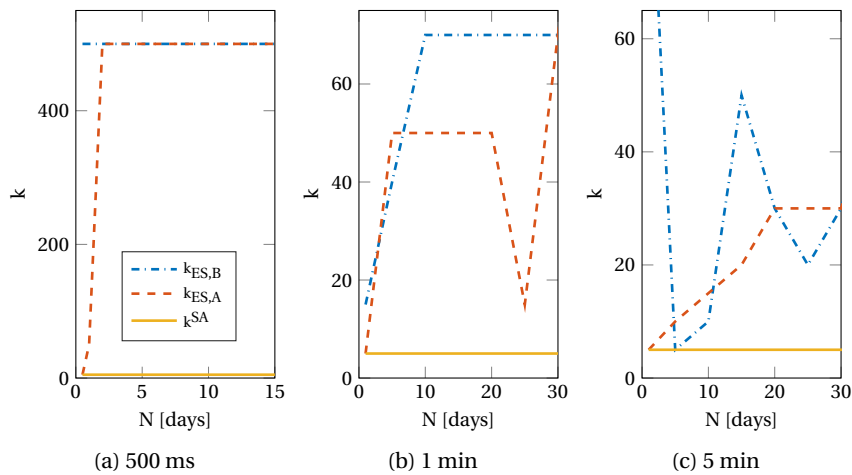
*Figure 3.5 – Number of cluster k as a function of the length of the training dataset. The Autumn dataset is selected for the analysis.*

In conclusion, when applying the exhaustive search we use the optimal combination of $k$ and $N$ found for the 55 days dataset as candidates for performance evaluation in the remaining 30 days. On the contrary, when applying the silhouette analysis approach, we select $k$=5 and $N$ equal to 1, 5 and 10 days respectively for three time horizons. These values are valid for all the three datasets. A comparison between the two approaches is presented in what follows.
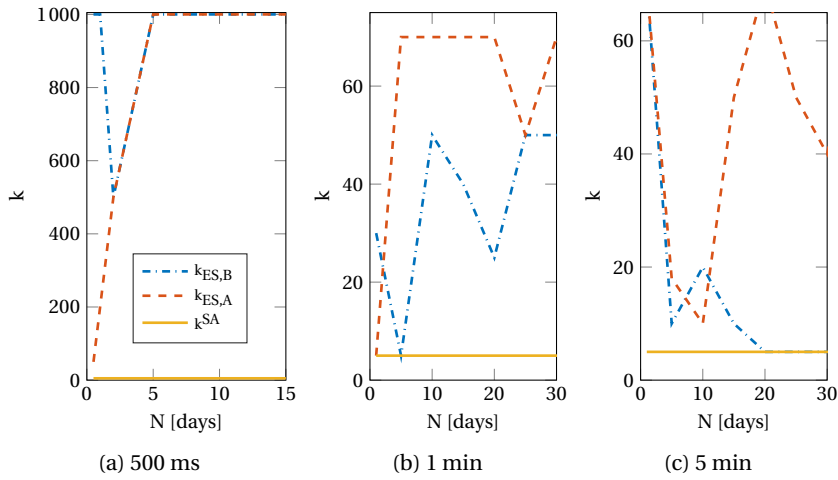
*Figure 3.6 – Number of cluster k as a function of the length of the training dataset. The Winter dataset is selected for the analysis.*

### 3.9.4   Ultra-short-term Forecasting: Performance Evaluation

We focus here on sub-second forecast horizon: one-step-ahead PIs at 500 ms. At first, we analyze the two proposed approaches to compute $k$ and $N$. Then, performance of methods A and B is evaluated, and therefore compared with existing methods from the technical literature.

**Exhaustive Search and Silhouette Analysis**

In this Section, we compare the performance obtained by applying the exhaustive search and the silhouette analysis to determine $k$ and $N$. Furthermore, results are compared with the optimal performance found a-posteriori to evaluate how far the estimations are from the optimum found a-posteriori. Evaluation is carried out in the testing set of 30 days, for each one of the three datasets.

Results are shown in Tables 3.3 for methods A and B. The comparison considers metric CWC. For 500 ms forecast horizon, the exhaustive search is outperforming the silhouette analysis and returns performance very close to the optimal a-posteriori. Indeed, when dealing with high sampling frequency, the large amount of data would require a number of clusters which is much higher than the one returned by the silhouette analysis. For sub-second time horizons, we proceed our analysis by using the exhaustive search method.

The following additional conclusions can be drawn from this analysis:

- For ultra-short-term forecast, Method B outperforms Method A for each considered case.

(a) Method A-500 ms    (b) Method B-500 ms
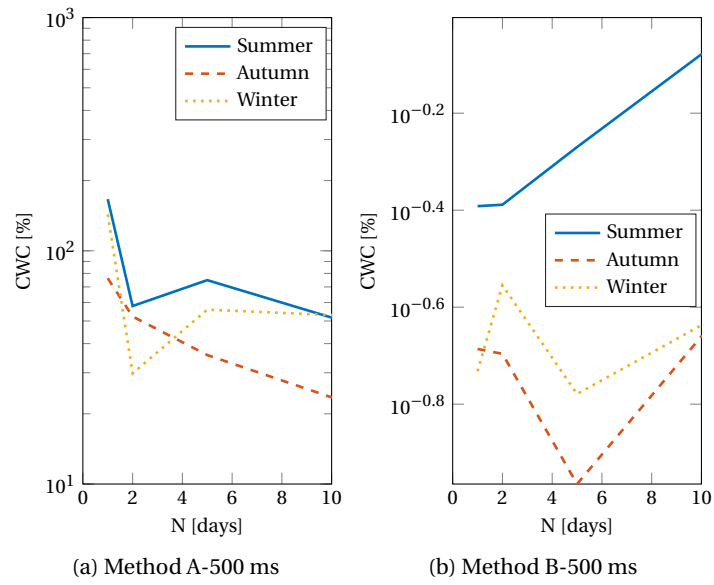
*Figure 3.7 – CWC as a function of the number of training days for 500 ms time horizon. CWC is shown in logarithmic scale.*



(a) Method A-1 min    (b) Method B-1 min
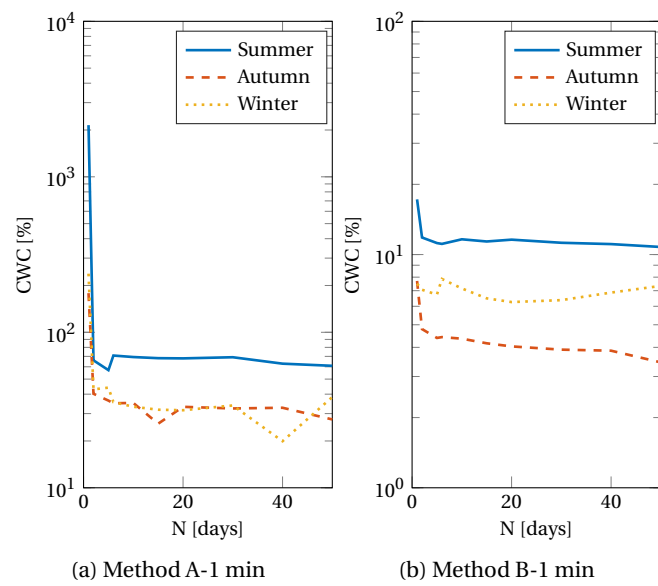
*Figure 3.8 – CWC as a function of the number of training days for 1 min time horizon. CWC is shown in logarithmic scale.*

49

(a) Method A-5 min
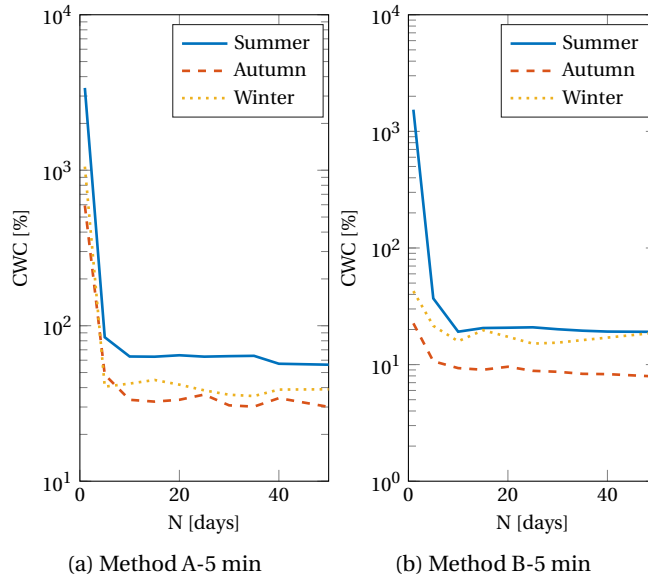
(b) Method B-5 min

*Figure 3.9 – CWC as a function of the number of training days for 5 min time horizon. CWC is shown in logarithmic scale.*

*Table 3.3 – CWC [%] for 500 ms, $\alpha$=95%.*

(a) Method A

| Method | Season | | |
|---|---|---|---|
| | Summer | Autumn | Winter |
| Optimal | 2.88 | 1.18 | 2.55 |
| Silhouette | 51 | 27.7 | 20.3 |
| Exhaustive Search | 4.69 | 2.41 | 4.03 |

(b) Method B

| Method | Season | | |
|---|---|---|---|
| | Summer | Autumn | Winter |
| Optimal | 0.24 | 0.046 | 0.13 |
| Silhouette | 0.37 | 0.12 | 0.27 |
| Exhaustive Search | 0.27 | 0.047 | 0.15 |

- For both the methods, the Summer period is characterized by worse performance and this is explained by the highest volatility content, as shown in Table 3.2. As expected, the Autumn period returns the best performance.

**Comparison with Benchmark Methods**

Table 3.4 shows the performance of the proposed methods compared with the above described benchmarks. First, we refer to the simple quantiles extraction (Quantiles A and Quantiles B), where the quantiles are computed from the original and differentiated time-series respectively. This comparison aims at showing the performance improving obtained by the k-mean clustering compared to the case where we extract the quantiles of the whole time-series without any clustering process. We can conclude that the k-means clustering is beneficial and leads to relevant performance improvement for all the analyzed cases.

The last row of Table 3.4 shows the results obtained by applying the ARIMA model and assuming a Gaussian distribution of the point forecast error (ARIMA + GAUSS). This method has to be compared with our proposed Method B since it requires a point forecast to compute the PIs. For 500 ms forecast horizon, the model is over confident with respect to the assumed normal distribution, returning PICP higher than 99% for $\alpha$=95%. Thus, to allow a fair comparison with our method, we empirically adjust the target confidence level (and so $\eta$) in order to obtain values of PICP similar to those given by the k-mean algorithm. Table 3.4 shows that, for the same coverage probability, Method B is characterized by lower PINAW.

*Table 3.4 – PICP-PINAW-CWC [%] for a time horizon of 500 ms, $\alpha$=95%.*

| Method | Season | | |
| --- | --- | --- | --- |
| | Summer | Autumn | Winter |
| Method A | 90.5-1.94-4.69 | 93.7-0.33-2.41 | 92.6-1.85-4.03 |
| Quantiles A | 94.6-57.4-113 | 93.3-29.5-62.6 | 95.7-35.1-35.1 |
| Method B | 97.0-0.27-0.27 | 96.1-0.047-0.047 | 98.2-0.15-0.15 |
| Quantiles B | 90.4-0.35-0.35 | 91.4-0.13-0.30 | 91.0-0.12-0.28 |
| ARIMA+GAUSS | 97.0-0.50-0.50 | 96.1-0.1-0.1 | 98.2-0.32-0.32 |

We refer to the extended Dynamic Interval Predictor (eDIP) described in Chapter 2. To allow a fair comparison, we compare Method B with the eDIP coupled with the persistent point forecast. From Table 3.5 we can conclude that the proposed method shows better performance when compared to the literature with respect to ultra-short-term horizons.

*Table 3.5 – PICP-PINAW-CWC [%].Performance comparison of the proposed Method B with the Dynamic Interval Predictor. $\alpha$=95%.*

| Method | Season | | |
| --- | --- | --- | --- |
| | Summer | Autumn | Winter |
| Method B | 97.0-0.27-0.27 | 96.1-0.047-0.047 | 98.2-0.15-0.15 |
| eDIP | 97.2-0.36-0.36 | 96.0-0.053-0.053 | 97.4-0.19-0.19 |

### 3.9.5 Short-term Forecasting

In this Section, we extend the proposed methods to higher forecast horizons (i.e. minutes).

**Exhaustive Search and Silhouette Analysis**

Tables 3.6 and 3.7 show metric CWC obtained by applying the exhaust searching, the silhouette analysis and the optimum a-posteriori, for 1 and 5 min forecast horizons. The silhouette analysis coupled with Method B shows here the best performance and is used for further comparison. On the contrary, when using the original time-series, the exhaustive search performs better and should be used to select $k$ and $N$.

*Table 3.6 – CWC [%] for 1 min, α=95%.*

(a) Method A

| Method | Season | | |
| --- | --- | --- | --- |
| | Summer | Autumn | Winter |
| Optimal | 21.7 | 8.50 | 22.9 |
| Silhouette | 58.0 | 31.6 | 33.0 |
| Exhaustive Search | 25.2 | 11.4 | 23.4 |

(b) Method B

| Method | Season | | |
| --- | --- | --- | --- |
| | Summer | Autumn | Winter |
| Optimal | 6.82 | 2.83 | 6.20 |
| Silhouette | 10.5 | 3.26 | 9.10 |
| Exhaustive Search | 14.0 | 3.81 | 10.3 |

*Table 3.7 – CWC [%] for 5 min, α=95%.*

(a) Method A

| Method | Season | | |
| --- | --- | --- | --- |
| | Summer | Autumn | Winter |
| Optimal | 34.2 | 16.7 | 24.0 |
| Silhouette | 54.7 | 25.7 | 31.5 |
| Exhaustive Search | 37.9 | 20.7 | 24.0 |

(b) Method B

| Method | Season | | |
| --- | --- | --- | --- |
| | Summer | Autumn | Winter |
| Optimal | 14.2 | 6.62 | 11.2 |
| Silhouette | 17.9 | 6.70 | 14.5 |
| Exhaustive Search | 16.2 | 15.1 | 23.4 |

**Comparison with Benchmark Methods**

First, we present the advantage given by the k-means clustering for forecasting of 1 and 5 min ahead. Table 3.8 shows the comparison with the simple quantiles extraction for the original and differentiated time-series (Quantiles A and B, respectively) for 1 min forecast horizon, as explained for the ultra-short-term analysis. Table 3.9 shows the same comparison for 5 minutes time horizon. For Method A, we apply the exhaustive search while for B we apply the silhouette analysis. For all the cases, we can see an improvement coming from the k-means clustering with respect to the simple quantile extraction.
The last row shows the results obtained by implementing the ARIMA model and assuming a Gaussian distribution of the point forecast error (ARIMA+GAUSS). For these forecast horizons, the PICP is slightly lower than the target confidence level, with values of PINAW that are however higher than those returned by Method B.

It is difficult to compare the proposed method with results available in the literature due to the different GHI measurements (characterized by dissimilar climatology). Similar results are

*Table 3.8 – PICP-PINAW-CWC [%] for a time horizon of 1 min, α=95%.*

|  | Season | | |
| --- | --- | --- | --- |
| **Method** | Summer | Autumn | Winter |
| Method A | 90.1-10.2-25.2 | 90.8-4.81-11.4 | 88.8-8.87-23.4 |
| Quantiles A | 94.7-56.7- 112 | 93.1-29.5-61.9 | 95.8-34.7-34.7 |
| Method B | 96.9-10.5-10.5 | 97.5-3.26-3.26 | 97.8-9.1-9.1 |
| Quantiles B | 89.7-13.8-34.8 | 90.6-6.1-14.6 | 91.7-6.73-15.3 |
| ARIMA+GAUSS | 93.4-19.2-40.3 | 94.0-8.13-16.5 | 95.6-10.6-10.6 |

*Table 3.9 – PICP-PINAW-CWC [%] for a time horizon of 5 min, α=95%.*

|  | Season | | |
| --- | --- | --- | --- |
| **Method** | Summer | Autumn | Winter |
| Method A | 91.5-16.5-37.9 | 86.7-6.96-20.7 | 96.1-24.0-24.0 |
| Quantiles A | 94.7-55.6-110 | 93.1-28.5-60.6 | 95.8-33.7-33.7 |
| Method B | 96.7-17.9-17.9 | 96.2-6.70-6.70 | 96.1-14.5-14.5 |
| Quantiles B | 89.4-25.9-66.4 | 89.5-13.6-34.0 | 91.2-16.0-36.0 |
| ARIMA+GAUSS | 91.7-28.9-65.8 | 92.3-14.1-31.4 | 95.0-19.0-19.0 |

obtained in [81] for 5 minutes ahead GHI forecast, where a probability coverage of $\approx 95\%$ and PINAW of $\approx 8\%$. In [81] a dataset of 1 year is considered. The percentage of periods of high volatility (i.e. with $\Delta K$ higher than 0.5) is $\approx 0.3 - 0.6\%$ while in our datasets is $\approx 0.9 - 1.5\%$.

### 3.9.6 From Ultra-short to Short-Term Forecasts

For sub-second time horizons, the best performance is obtained by applying the exhaustive search coupled with Method B. On the contrary, for higher forecast horizons, the silhouette analysis coupled with Method B is the most performing one. For all the considered horizons, differentiating the time-series has a positive effect on the final performance and allows having a PICP higher than or equal to $\alpha$. However, the improvement coming from the differentiation decreases with increasing forecast horizons. Indeed, performance of Method A worsens less than those of Method B when increasing the forecast horizon. To complete the analysis, Fig. 3.10 shows our metrics as a function of the forecast horizon and for different confidence levels: 85%, 95%, and 99%.We can see that PICP (left side) is always higher or equal to the target confidence level. Furthermore, the value of PINAW (right side) increases with the forecast horizon (to account for the higher uncertainty) and increases with $\alpha$, i.e. the method adapts the bound widths to guarantee the target coverage. Fig. 3.11 shows the PIs and the actual realizations obtained for 500 ms, 1 and 5 min forecast horizons, respectively. Method B is
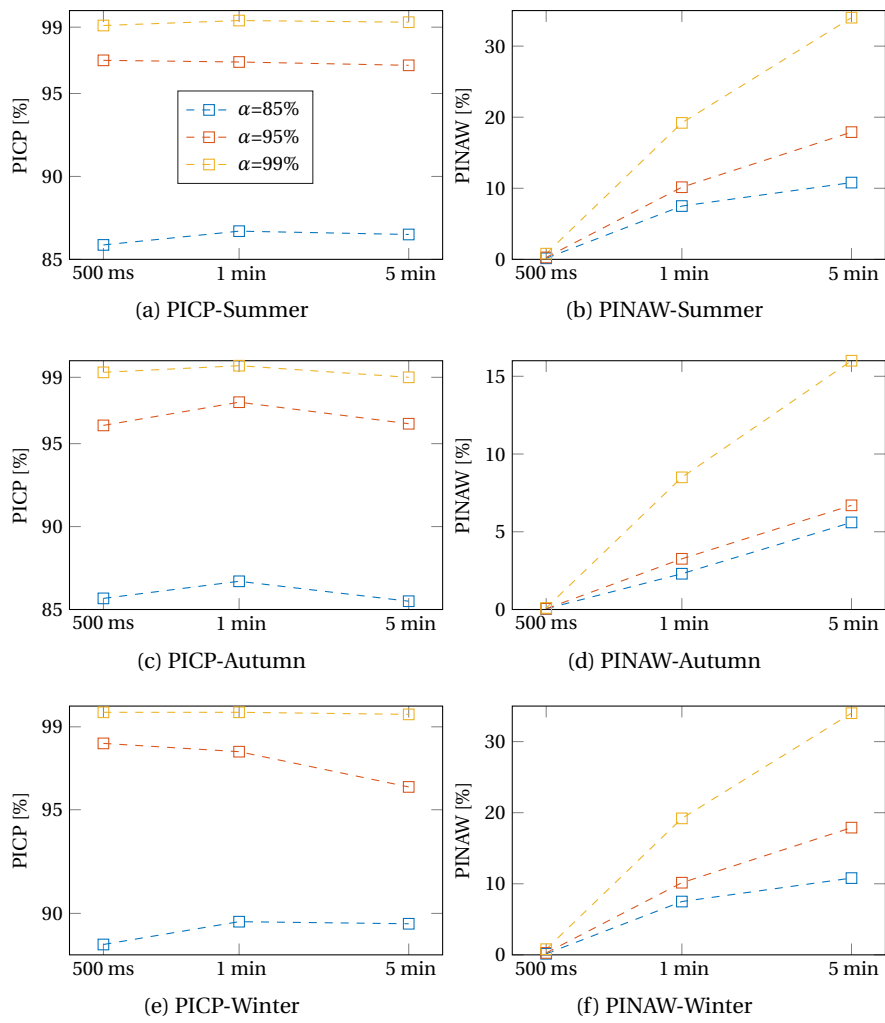
Figure 3.10 – PICP [%] and PINAW [%] are shown for the Summer, Autumn, and Winter periods and different target confidence levels α.

applied and the target confidence level is 99%. A day of high variability and a clear-sky day from the Winter period are selected for the comparison. The corresponding values of PINAW are shown in Table 3.10, for the two days respectively, showing that the PIs are narrower for the clear-sky day where the variability is lower. For the clear-sky day at 500 ms a zoom is added since the PIs and the points are not easily distinguishable. We can see that larger intervals are associated to higher time horizons, this reflecting the higher uncertainty. In Fig. 3.12 the same

*Table 3.10 – PINAW [%] is shown for the two days of Fig. 3.11.*

| | Day | |
|---|---|---|
| **Forecast Horizon** | High Volatility | Clear-sky |
| 500 ms | 0.22 | 0.18 |
| 1 min | 11.8 | 10.8 |
| 5 min | 24.9 | 23.8 |

days of Fig. 3.11 are selected and PIs are plotted for different confidence level (99%, 95% and 85%). It is interesting to notice that for the considered clear-sky day a target confidence level of 85% is enough to have all the measurements inside the PI.

### 3.9.7 Reliability Diagrams

The objective of this last analysis is to compare the target confidence levels with the observed ones, here represented by metric PICP.
The analysis is for the three considered periods and three forecast horizons and is shown in Fig. 3.13. We consider Method B and the ARIMA model with Gaussian distribution of the error and compare their performance with the ideal behaviour, namely when the target confidence level is identical to the observed one. As it can be seen, the confidence levels obtained for Method B exhibit an overall good matching with the target ones, proving the capability of the method to provide reliable predictions. In particular, Method B is always slightly over confident with the exception of the Autumn period where is under confident for low values of $\alpha$. The proposed benchmark method has lower reliability, it is over confident for sub-second time horizons while its behaviour for higher horizons depends on the value of $\alpha$. This mismatch suggests that parametric models, with the implicit assumption of a Gaussian distribution of the error, might not be suitable.

### 3.9.8 Execution Time Statistics

Execution times are computed adopting a Matlab 2016a implementation of the algorithm on an Intel Core i7-6600U CPU 2.60GHz machine. For the analysis, we consider the worst case scenario corresponding to the highest number of clusters. We select it equal to 1000 that is the maximum value of clusters returned by the analysis at 500 ms (above this value of $k$ we do not see any performance improvement). At each time step, the overall operational time

(a) High Volatility - 500 ms

(b) Clear-sky - 500 ms

(c) High Volatility - 1 min

(d) Clear-sky - 1 min

(e) High Volatility - 5 min

(f) Clear-sky - 5 min
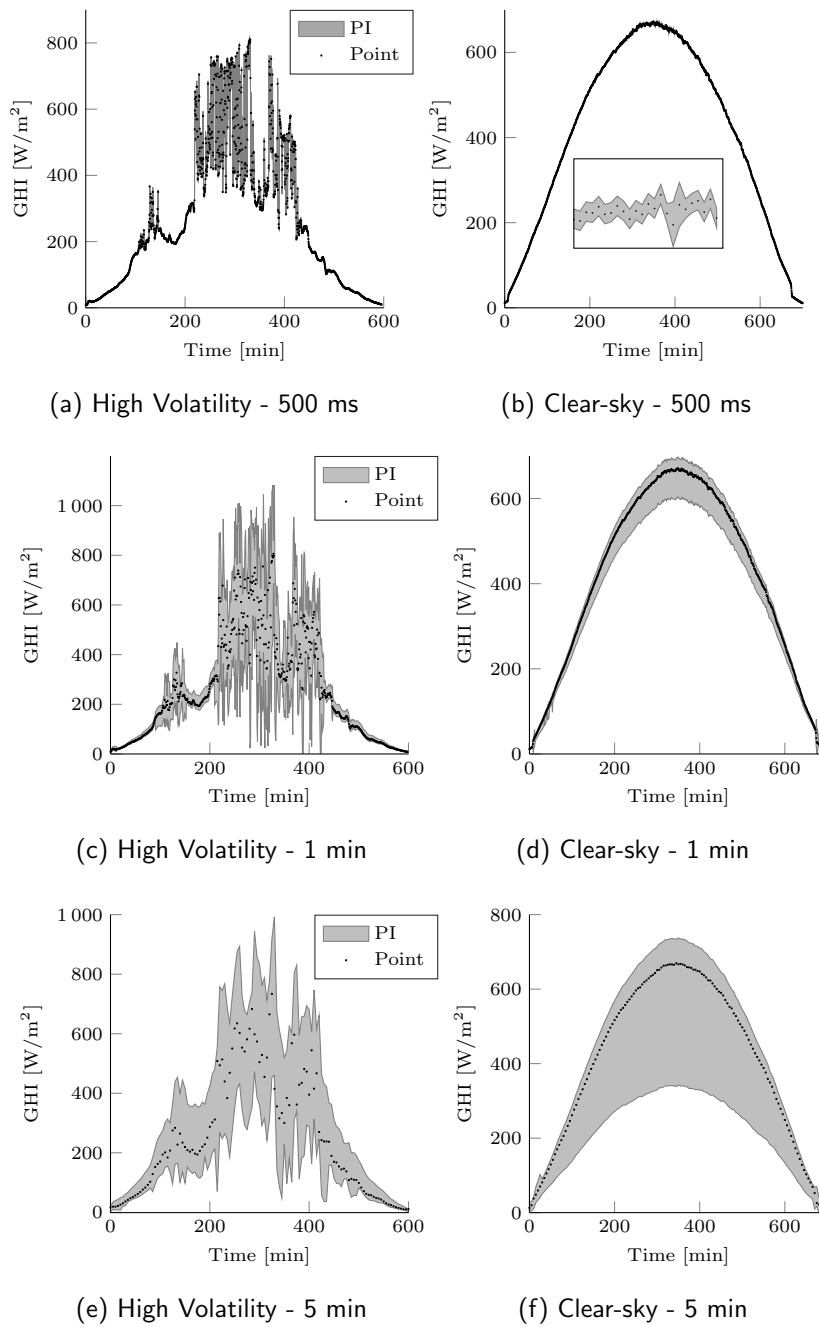
*Figure 3.11 – PIs and realizations are shown for different forecast horizons considering daylight hours, $\alpha = 99\%$ and Method B is applied. Two days with different weather conditions are selected from the Winter period.*

(a) High Volatility        (b) Clear-sky
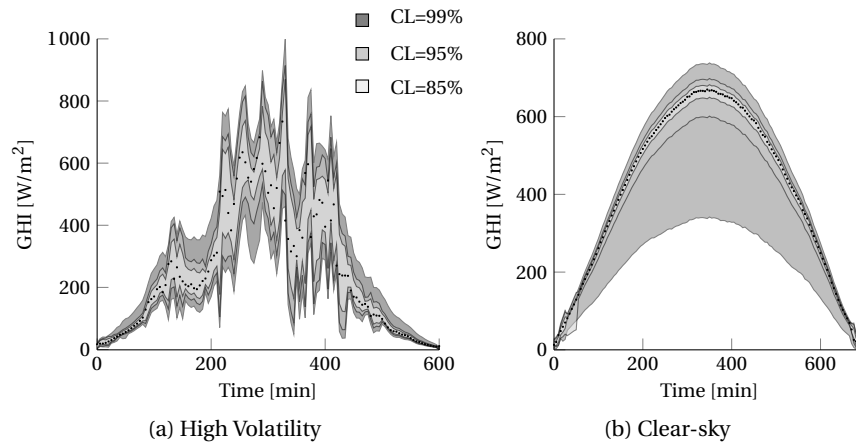
*Figure 3.12 – PIs and realizations are shown for 5 minutes time horizon and different target confidence levels, represented by different levels of shadings for the same data of Fig. 3.11.*

required to deliver the PI is always less than 0.5 ms. The mean and the standard deviation of the computational time at each time step are 0.35 ms and 1.14 ms, respectively. Therefore, the method can be used for the real-time computation of PIs at sub-second time scales.

## 3.10 Chapter Conclusions

The problem of quantifying the uncertainty associated with solar volatility is investigated in this Chapter, focusing on forecast horizons that are meaningful in microgrids control applications (i.e. from sub-second up to minutes). A simple method to deliver PIs for GHI is proposed and its performance assessed. The proposed technique extracts information from a limited training set: data are clustered off-line by using the well-known k-means algorithm and the quantiles of the obtained clusters are then used for PIs computation. The method does not rely on any specific point forecast technique and does not need any information from sky imaging. First, a clear-sky model is implemented. It is shown that the de-trending of the time-series is advantageous for time horizons higher than the minute time-scale, when the influence of the dynamics associated to solar position becomes non-negligible. We show that the algorithm outperforms the benchmark case with simple quantiles extractions and the benchmark case considering the ARIMA model Gaussian distribution of the point forecast error. Furthermore, performance is shown to be in line or improve those available in the literature, for all the considered forecast horizons and using a shorter and limited training set. The method is applied to the original and differentiated clear-sky index time-series. Results show that the benefit coming from the time-series differentiation decreases while increasing the forecast horizon. It is shown that the proposed method is able to adapt the widths of the PIs in order to guarantee the target coverage. In the next Chapter, we investigate if further information from sky images can improve the proposed method.
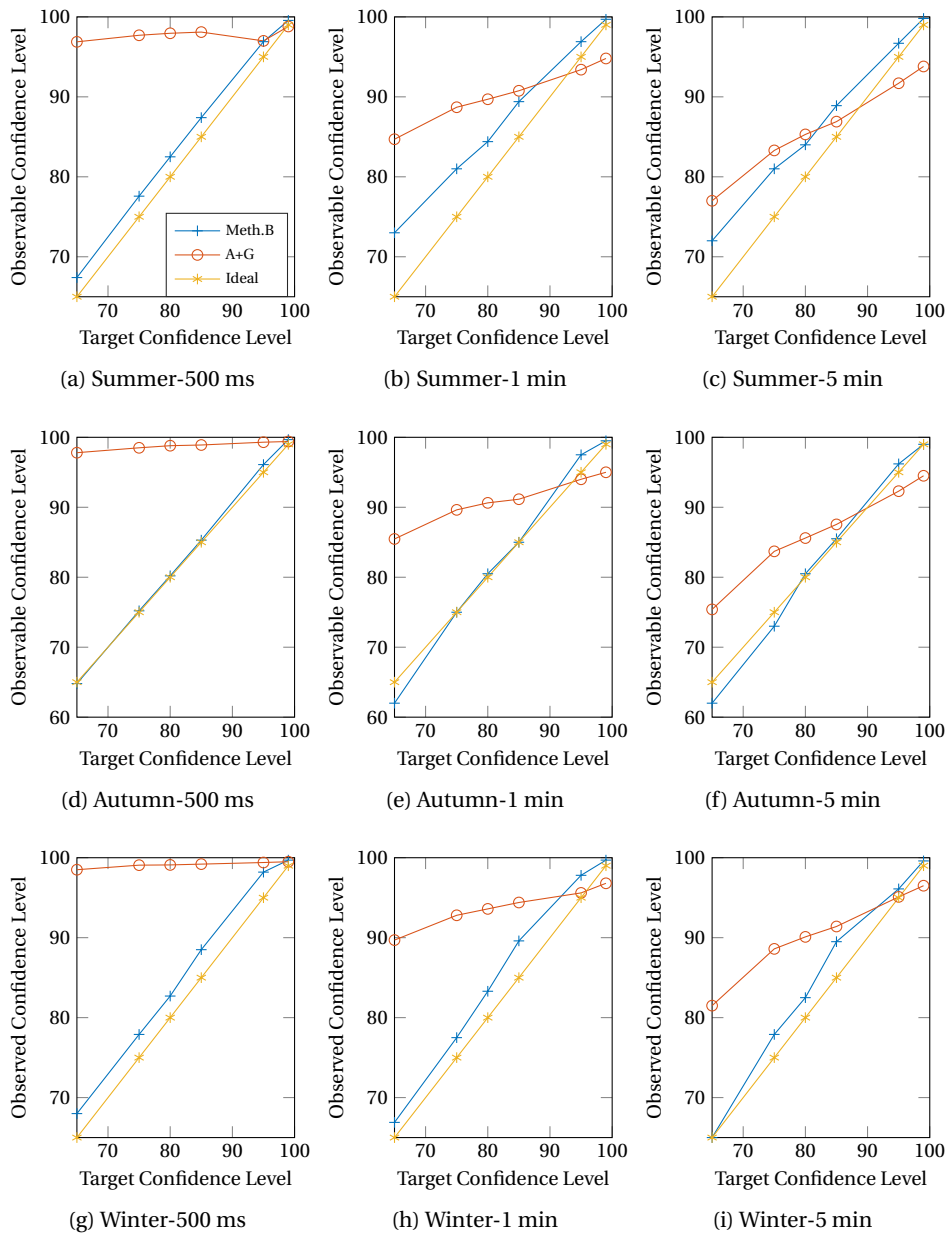
*Figure 3.13 – Reliability Diagrams for the three periods and forecast horizons.*

# 4 Solar Irradiance Forecast and Estimation via All-sky Camera

## 4.1 Chapter Highlights and Summary

Ground observations using all-sky cameras have recently gained recognition in the field of solar irradiance forecasting. Considering that the area seen by the hemispherical image is relatively small, the associated forecasting methods are interesting for single PV plants or small-scale distribution feeders. In this Chapter, we experimentally investigate how the inclusion of features from ground all-sky images can be used to: i) improve time-series-based GHI probabilistic forecasting tools as the one in Chapter 3 and ii) estimate the actual and the one-step-ahead GHI.

First, we propose a toolchain that allows predicting the cloud cover on the sun disk, through image processing and cloud motion identification. We investigate the benefit of integrating this information to improve probabilistic GHI forecasting methods, besides the sole use of the GHI time-series proposed in Chapter 3. This first analysis also includes a comparison of different cloud segmentation and cloud motion identification methods. Second, a methodology to estimate the irradiance from all-sky images using neural networks is proposed, investigating the possibility of using an all-sky camera as an irradiance sensor. The obtained estimations are benchmarked against satellites-based ones. The toolchain is also extended to consider one-step-ahead GHI point forecast.

## 4.2    Forecasting Toolchain

Since the solar irradiance is highly dependent on the cloud coverage, image-based solar cloud forecast has been largely studied by the recent literature, [14, 100]. This technique generally refers to short time horizons (the typical maximum horizon is in the order of 20 min, depending on the wind speed and the camera's field of view) and provides information on localised effects. Even if the available studies differ on how the single models are implemented, a generic toolchain usually adopted in the literature is shown in Fig. 4.1 and consists of the following steps:



*Figure 4.1 – Process of GHI forecast with all-sky images.*

1. Image acquisition. Fish-eye cameras are generally used to capture sky images at high resolution (second time scale).

2. Image pre-processing. The horizon (i.e. topological features and nearby objects) is removed from the image, and the distortion due to the fish-eye lens is corrected.

3. Cloud segmentation. It consists in detecting the clouds from the clear-sky.

4. Cloud motion or tracking. It consists in estimating the cloud movement.

5. Irradiance forecast. Information from the previous blocks are used to forecast the irradiance or irradiance variations. This may include cloud shadow mapping (i.e. determining the 3D positions of a cloudy pixels in a real world coordinate system) or empirical methods based on machine learning.

The following paragraphs detail how these steps are implemented in this work. We first describe the custom setup for image acquisition and how the images are pre-processed. Then, we focus on cloud segmentation and cloud motion methods. Our irradiance forecast model relies on the adoption of machine learning methods.[1] In particular, we predict the sun occlusion (i.e., the predicted *local cloud cover* in the sun disk) and we use it as influential variable to train the algorithm proposed in Chapter 3.

---

[1]Since our setup does not have a celiometer or similar instrumentations to calculate cloud height, cloud shadow mapping is not considered in the analysis.

## 4.3 Image Acquisition

Two cameras are installed on the roof of our laboratory building at EPFL (GPS coordinates 46.518°N, 6.565°E), both pointing at the zenith. The first camera is an ELP USB 2 megapixel camera with a CMOS sensor and a fish-eye lens. Frames are taken at 1 min resolution with manual exposure time to limit overexposure of the circumsolar area. The frame-grabber is a Raspberry PI embedded microcontroller. The second camera is a Basler USB 1.8 megapixel camera (acA1300-200uc) with a CMOS sensor and a 2/3" Format C-Mount fish-eye lens and it is shown in Fig. 4.2. A dome-shaped fish-eye protector is installed to protect the lens, characterized by a 92% light transmission. The frame-grabber is an Intel Nuc. Despite being more expensive, this second solution allows for a higher controllability of the exposure time. Frames are taken at 30 s resolution and at 3 different exposure times: automatic (from the camera internal control algorithm), underexposed, and overexposed. These three images are then combined in real-time to generate a High Dynamic Range (HDR) image using the PICCANTE image processing library, [101].



*Figure 4.2 – Basler all-sky camera installed on the rooftop PV plant at the EPFL DESL laboratory.*

All the images are timestamped and locally saved in a server at the EPFL DESL laboratory. We note that since the direct light is not blocked, the region around the sun is affected by glare which affects negatively the accuracy of the cloud segmentation. Results on the first part of the chapter (namely cloud segmentation, motion, and forecasting) are obtained with the Basler underexposed images, while results on the irradiance estimations (second part of the chapter) are obtained using the ELP camera. Indeed, the two setups showed similar performance, only one is selected for each analysis to avoid redundancy in the results.

## 4.4 Image Pre-processing

All-sky images are treated to remove the distortion introduced by the fish-eye lens with the procedure described in [102]. This requires the knowledge of the intrinsic, extrinsic and lens

distortion parameters, which are inferred with the method in [103], applied one-time prior operations in an off-line process. Moreover, all the pixels containing unnecessary information, like objects along the horizon, are removed by applying a bitmask, that is manually determined off-line since the position of the camera is stationary. Finally, the PVLib library, [104], is used to calculate the solar angles (azimuth, zenith, and elevation), by knowing the time and the GPS location of the camera, and then the position of the sun in the image is determined by applying the zenithal equal area projection method [105]. Fig. 4.3 shows an example of original and processed images.



(a) Original Image      (b) Pre-processed Image
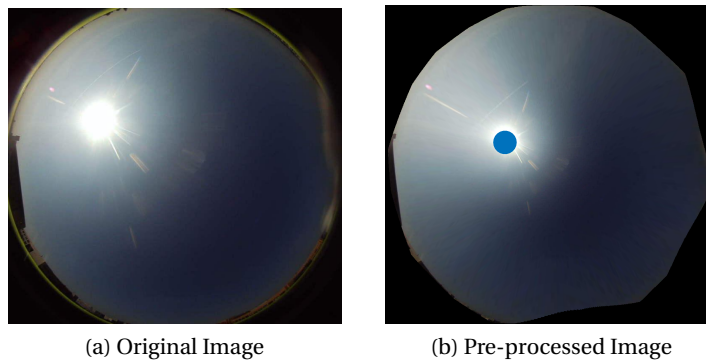
*Figure 4.3 – The original all-sky image (left) and pre-processed all-sky image (right) are shown. Fig. 4.3b is undistorted and cleaned of the horizon features, and the sun location is denoted by the blue disk.*

## 4.5   Cloud Segmentation

Cloud segmentation consists in determining if a pixel corresponds to a cloudy or a clear-sky point (binary classification). It is probably the most challenging part of the forecasting chain due to the fact that several types of clouds exist, combined with several varieties of lighting conditions.

Most of the literature presents a binary segmentation based on colour features. A common procedure for cloud detection consists in computing the red to blue ratio (RBR) of the image color channels for each pixel, and comparing it with a fixed threshold to detect whether the pixel corresponds to a cloud (high RBR) or a clear-sky (low RBR) pixel. The method was inspired by the Scripps Institution of Oceanography, [106]. Increased molecular Rayleigh scattering of shorter wavelengths causes the blue channel to be larger for clear-sky pixels. On the contrary, since clouds scatter the visible wavelengths more evenly, the red signal is similar to the blue signal for cloudy pixels. However, misclassification occurs due to very bright clear sky pixels (high RBR) in the circumsolar area and very dark clouds (low RBR). The method in [107] suggests an Hybrid Thresholding Algorithm (HYTA) based a normalized RBR. The HYTA first identifies the RBR as either unimodal or bimodal according to its standard deviation, and then the unimodal and bimodal images are treated by fixed and minimum cross entropy

(MCE) thresholding algorithms, respectively. Authors of [14] propose to use a clear-sky library and a sunshine parameter to improve the cloud detection. The clear-sky library is a database of clear-sky images captured for different solar zenith angles while the sunshine parameter is a dynamic RBR threshold that changes from image to image, and is typically small when the sun is obscured. Authors of [108] propose an augmented RBR definition. It makes use of a clear-sky library to achieve a better segmentation of the circumsolar pixels, and solar saturation and gray intensity level to account for dark clouds.

Alternatively to threshold-based methods, learning algorithms have recently gained importance. Authors of [109] propose a supervised segmentation method based on an analysis of different color spaces and using principal component analysis and partial least square regression, providing a probabilistic classification rather than a binary one. The method is shown to outperform other classical methods based on the RBR. The same Authors in [110] model the various labels/classes with a continuous-valued multi-variate distribution. Authors of [111] proposed an unsupervised learning method based on adaptive Gaussian mixture model and using as feature vectors different combinations of the red, blue and green channels.

### 4.5.1 Proposed Method for Cloud Segmentation based on the IBR

After an empirical analysis on the combination of several color channels (e.g.: red, blue, green, intensity) and of different segmentation methods (e.g., Otsu's thresholding [112], region growing [113]), the following procedure was found to be the most accurate:

- For each pixel we compute the intensity-blue ratio (IBR). The pixel intensity is computed by converting the image in grayscale.

- We compute the IBR for the corresponding clear-sky image (IBR$_\text{cs}$).

- Subtracting the IBR$_\text{cs}$ to the IBR of the original image leads to higher accuracy only if the sun is partly covered or clear-sky, but it misclassifies dark clouds as sky if the sun is entirely obscured. For this reason we proceed with the ratio subtraction only when the average pixel intensity in the sun disk ($I_{sun}$) is above a certain threshold $T_{sun}$:

$$\text{IBR} = \begin{cases} \text{IBR} - \text{IBR}_\text{cs}, & I_{sun} > T_{sun} \\ \text{IBR}, & \text{otherwise.} \end{cases} \tag{4.1}$$

where $T_{sun}$ is empirically selected.

- The optimal threshold, i.e. the value above which the pixel is considered as a cloud, is found by applying the Otsu's method on the training set, [112]. It is based on computing the histogram of an image, i.e. the distribution of the pixel values in the image. If the image is composed of objects on a background, then two peaks are present in the histogram and the optimal threshold is the value that minimises the intra-variance

inside each classes (object and background). The Otsu's method is implemented using the *graythresh* function in Matlab.

### 4.5.2 Segmentation Results: a Comparison of State-of-the-art Methods

In order to find the most appropriate method for the segmentation, we propose here a comparison of some of the methods available in the literature. We select 60 underexposed images. They are selected during September 2017 and to cover different conditions, from clear-sky to completely cloudy sky. These images are manually segmented and divided randomly into a training set consisting of 30 images and a testing set consisting of the remaining 30 images. The training set is used to compute the empirical parameters for the threshold-based methods or to train the algorithm in the supervised learning methods. All the results are presented by performing a two-fold cross-validation on the set of 60 images. The selected methods, their implementation, and the results of the comparison are described hereafter.

**Schmidt's segmentation**

The first method is the one proposed in [108]. The modified RBR is defined as:

$$R_{mod} = R_{orig} - R_{CSL}(a(S-1) - b(In-c)) \tag{4.2}$$

where $R_{orig}$ is the RBR of the considered image, $R_{CSL}$ is the corresponding RBR value on a clear-sky picture (selected from a clear-sky library such that the sun position is as close as possible to the target picture [2]), S is the saturation and refers to the percentage of saturated pixels in a circumsolar region with radius $r$, In is the gray intensity level, while $a$, $b$ and $c$ are tuning parameters. Finally, a pixel is marked as a cloud if its modified RBR value in Eq. 4.2 is larger than a threshold $R_{thresh}$. The free parameters $a$, $b$, $c$, $r$, $R_{thresh}$ involved in the cloud detection procedure are determined in order to give the best performance with our imaging setup. They are chosen with the following heuristic procedure: the training images are segmented by using different possible combinations of the parameters in the ranges shown in Table 4.1. Then, the segmented images are compared with the manually segmented ones and best values of the parameters are chosen as those leading to the best cloud map estimation.

*Table 4.1 – Segmentation parameter setting*

| Name | Selected value | Tested range |
|------|----------------|--------------|
| $a$ | 0.1 | 0 to 1, with a step of 0.1 |
| $b$ | 0.006 | 0 to 0.003, with a step of 0.001 |
| $c$ | 400 | 100 to 500, with a step of 10 |
| $R_{thresh}$ | 0.4 | 0.1 to 1, with a step of 0.1 |
| $r$ (pixels) | 30 | 30 to 100, with a step of 10 |

---

[2]The clear-sky library is a set of clear-sky pictures that is manually determined off-line by selecting clear-sky days from the same period of the testing images.

**Hyta**

The second method we propose to compare is the Hyta previously introduced. In our experiments, we set the standard deviation threshold to 0.3 (an image with a standard deviation greater than 0.3 is considered bimodal, otherwise is unimodal). The fixed threshold for unimodal images is instead fixed to 0.25. These parameters are selected as those leading to the best cloud map estimation in the training set.

**IBR Method**

For the proposed methods based on the IBR, we found that the best performance is obtained for a value of $T_{sun}$=230 ( considering pixels value of 8-bit integer giving a range of possible values from 0 to 255). The threshold above which the pixel is considered as cloudy is found equal to 0.73 and 0.0989 for the cases without and with sun subtraction, respectively.

**Supervised Learning Methods**

Finally, we implement four supervised segmentation methods:

- We use the method proposed in [110] to perform a binary cloud classification. The set of features (*SET 1*) is composed by the red channel, the RBR, and the saturation channel of the HSV color space. This set of features was found to be the most significant for cloud segmentation when performing an extensive analysis on 16 colour channels by the same Authors [114]. Assuming each feature to be a continuous random variable, Gaussian Discriminant Analysis is used to model the different class labels using multivariate Gaussian distribution (MVG), [115].

- The same set of features (*SET 1*) is used to classify the pixel using the binary support vector machine (SVM) classifier implementation in Matlab, [116]. We name this method as SVM1. [3]

- The same set of features (*SET 1*) is used to classify the pixel in cloudy or clear-sky using the multinomial logistic regression (MLR) implementation in Matlab, [117]. We name this method as MLR1. Given a set of predictors, the algorithm returns the predicted probabilities for the multinomial logistic regression model. In our case, the probability above which the pixel is considered as cloud is selected empirically on the training set equal to 0.72.

- A new set of features is considered as input to the SVM and the MLR. The second set of features (*SET 2*) is selected as proposed in [118]. The new features are: the RBR, the normalized RBR, the contrast and the entropy of the image, the red ad blue difference,

---

[3] An SVM classifies data by finding the best hyperplane that separates all data points of one class from those of the other class.

*Table 4.2 – Precision of several segmentation methods.*

| Method | Schmidt | Hyta | IBR | MVG | SVM1 | MLR1 | SVM2 | MLR2 |
|---|---|---|---|---|---|---|---|---|
| Precision [%] | 80.3 | 79.8 | <u>83.7</u> | 80.7 | 73.1 | 78.6 | 72.6 | 77.8 |

and the ratio of the red channel to the maximum of red and blue channel. We name these methods SVM2 and MLR2, respectively.

**Metrics and Results**

By comparing the obtained segmented image with the ground truth, we obtain the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) samples of a binary image. We compute the segmentation precision as:

$$\text{Precision} = TP/(TP + FP) \tag{4.3}$$

Results are shown in Table 4.2. From the table, we can conclude that the proposed IBR method is the most performing one, followed by the supervised MVG, and the Schmidt method.

An example of segmentation obtained using the proposed IBR method is shown in Fig. 4.4. We can see how also dark clouds are correctly classified.
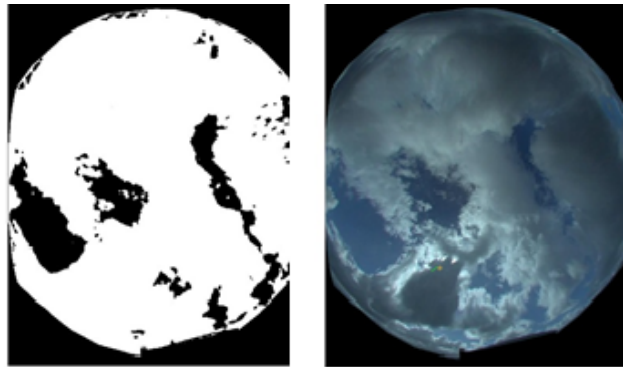


*Figure 4.4 – Example of segmentation using the IBR method.*

## 4.6 Cloud Motion

Cloud motion identification is the next important step to compute the irradiance forecast. Most of the literature compares two consecutive images to identify pixel correlations and cloud velocity fields. Then, assuming spatial homogeneity, a unique motion vector, here called *global motion vector*, is obtained. Finally, assuming a persistent cloud motion for the following time interval, the cloud map is advected at the speed and direction of the global motion vector to obtain the predicted cloud map. Several methods have been proposed in the literature

for cloud motion identification. Chow et Al., [14], use the cross-correlation method (CCM), applied to two consecutive images, to search the vector with the largest cross-correlation coefficient. Another technique for cloud tracking is the optical flow, [119], which consists of a collection of apparent velocities of objects in an image. It is applied for predicting sun occlusions in [120], using consecutive frames shot at 5 seconds distance. To stabilize the tracking process, the Kalman Filter is applied as a predictor-corrector algorithm. Authors of [121] propose to use the particle image velocimetry (PIV) and by then applying the k-means clustering on the obtained velocity vectors in order to obtain a representative velocity vector. Quesada-Ruiz et al. in [122] propose a method for cloud tracking applied to intra-hour direct normal irradiance forecast. A sector-based method is used to detect the direction of motion of potentially sun-blocking clouds, and an adjustable-ladder method focuses on sky regions that potentially affects DNI values. Finally, Bernecker et al. in [123] introduced non-rigid registration for detecting cloud motion. A sun occlusion probability is filtered by a Kalman filter to obtain continuous GHI forecasts for up to 10 min.

In our analysis three methods are considered to compute the motion fields:

1. Particle Image Velocimetry (PIV). In brief, PIV consists in comparing two consecutive pictures by evaluating the cross-correlation between portions of the images, called interrogation areas. This allows inferring the most likely particle displacement and to compute the motion vectors. We use the "MPIV" library in [124] and using an inter-rogation window of 32x32 pixels, the velocity vectors are calculated by the minimum quadratic difference algorithm, and with 0 overlapping between consecutive windows.

2. The optical flow (OF). OF is performed by implementing the Lucas-Kanade method is its Matlab formulation, [125]. It assumes that the flow is constant in a local neighbourhood of the pixel under consideration, and solves the basic optical flow equations for all the pixels in that neighbourhood, using the least squares criterion.

3. The persistent method. This method assumes that the clouds are persistent in a short-term horizon, and therefore the global cloud motion vector is zero.

Once the cloud fields are obtained from the PIV and OF, the global motion vector is obtained by spatially averaging the obtained vectors.

### 4.6.1   Local Cloud Cover Computation

The last step of our chain aims at predicting the local cloud cover at the following time step. It consists in translating the current cloud map according to the global motion vector, which is scaled in magnitude according to the forecasting horizon. This leads to the so-called forecasted cloud map.

An example of the procedure described until this stage is shown in Fig. 4.5. Fig. 4.5a shows
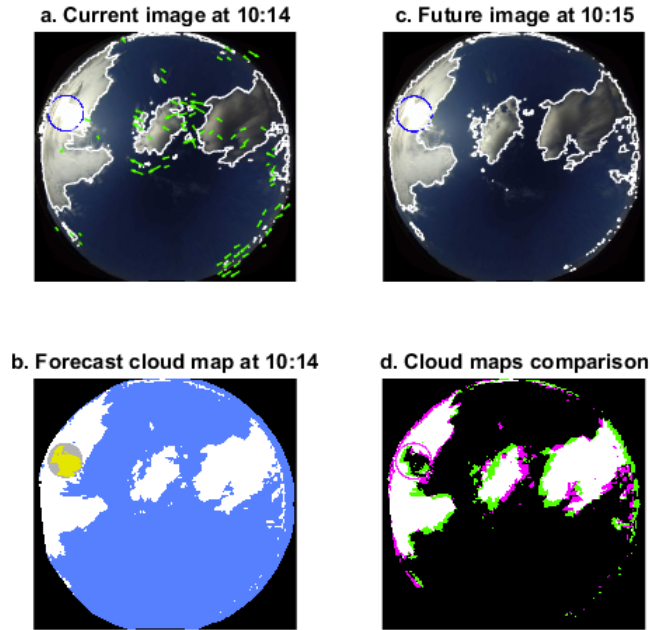
*Figure 4.5 – Example of the forecasted cloud map procedure.*

the undistorted view of the sky with the sun location (blue circle) and motion vectors (green arrows). The global motion vector, obtained by spatially averaging the vectors, is used to translate the cloud map obtained by segmenting the original image. The translated cloud map is shown in Fig. 4.5b, where the white color denotes cloudy pixels, blue the clear-sky, and yellow the circumsolar region. Fig. 4.5c shows the 1 min ahead realization. Fig. 4.5d compares the forecasted cloud map (purple color) against the future ground truth cloud map from Fig. 4.5c (green color). The white color denotes those pixels which are correctly classified as cloudy.

The forecasted local cloud cover is computed as the percentage of cloudy pixels in the forecasted cloud map in a region around the sun. The considered region is a disk centered at the sun position. We consider a circumsolar area rather than the whole picture since this is the region with the largest interest when considering short-term sun occlusions by clouds.

### 4.6.2 Results on Cloud Motion

In order to evaluate the performance of the cloud motion, we select 20 consecutive pictures during a partly cloudy period and we manually segment them such that the segmentation error is negligible in this analysis. Then, the three cloud motion methods (PIV, OF and Persistent) are applied to obtain the forecasted local cloud cover (which is computed on an area of 100 pixels around the sun position). We define the cloud motion error as the relative error between the forecasted local cloud cover and the ground truth one.

Results are shown in Table 4.3. The OF returns the lowest error and is therefore used in the

*Table 4.3 – Comparison on cloud motion.*

|                          | PIV | OF  | Persistent |
| ------------------------ | --- | --- | ---------- |
| Cloud Motion Error [%]   | 2.0 | 1.2 | 2.5        |

following analysis.

## 4.7  Improvement of Time-series-based Forecasting

In this Section, we investigate if integrating the local cloud cover into machine learning-based methods can lead to improve performance with respect to the case when only information from the irradiance time-series are used. We consider the probabilistic forecasting tool described in Chapter 3. Indeed, the clustering procedure is general enough to include heterogeneous influential variables (e.g. coming from the GHI time-series or the images). We augment the proposed algorithm by including the local cloud cover as an additional influential variable. To compute the forecasted local cloud cover and test its influence on the GHI forecasting tool, we choose the best performing segmentation and cloud motion algorithms from the previous analysis, namely the IBR method for segmentation and the optical flow for cloud motion.

We compare the k-means-based algorithm (formulated as Method B, i.e. by differentiating the clear-sky index time-series, see Chapter 3) for three different combinations of influential variables:

- average irradiance and its variability. This is the original method discussed in Chapter 3, which does not require all-sky camera information. We name this case *GHI measurements*;

- average irradiance, its variability, and the local cloud cover (henceforth called *Images + GHI measurements*);

- local cloud cover only (henceforth called *Images*).

The training and testing dataset consists of 22 and 8 days, respectively. GHI observations are from the pyranometer as described in Section 2.2.1. Images and GHI measurements are time synchronized at 1 min resolution. We consider 1, 2, 5, and 10 min as forecast horizons. Therefore, for the analysis at 2, 5, and 10 min the information are downsampled by simple decimation. We use the standardized metrics defined in Section 2.3.1 to evaluate the PIs, namely the prediction interval coverage probability (PICP) and the prediction interval normalized average width (PINAW). Results are summarized in Tables 4.4 and 4.5, and the PINAW is also shown in Fig. 4.6. They can be summarized as follows:

- from Table 4.4, all the considered cases have a coverage probability (PICP) close to the target confidence level (95%), denoting that the methods have a good reliability;

*Table 4.4 – PICP (%)*

| Forecast horizon | 1 min | 2 min | 5 min | 10 min |
|---|---|---|---|---|
| GHI measurements | 96.06 | 95.6 | 95.14 | 97.2 |
| Images + GHI measurements | 94.75 | 95.16 | 96.45 | 97.23 |
| Images | 94.92 | 94.56 | 94.75 | 95.95 |

- in any case, adding the forecasted local cloud cover to GHI measurements is beneficial as it improves the GHI forecast for all the considered forecasting horizons (or it does not impact negatively, as in the 1 min case);

- the use of the forecasted local cloud cover as only influential variable outperforms the other methods at 1, 2, and 5 min forecast horizons in terms of PINAW. However, the PICP is slightly lower.

*Table 4.5 – PINAW (%)*

| Forecast horizon | 1 min | 2 min | 5 min | 10 min |
|---|---|---|---|---|
| GHI measurements | 4.13 | 7.05 | 12.69 | 18.12 |
| Images + GHI measurements | 4.1 | 6.65 | 12.1 | 16.1 |
| Images | 3.5 | 5.6 | 11.7 | 17.1 |
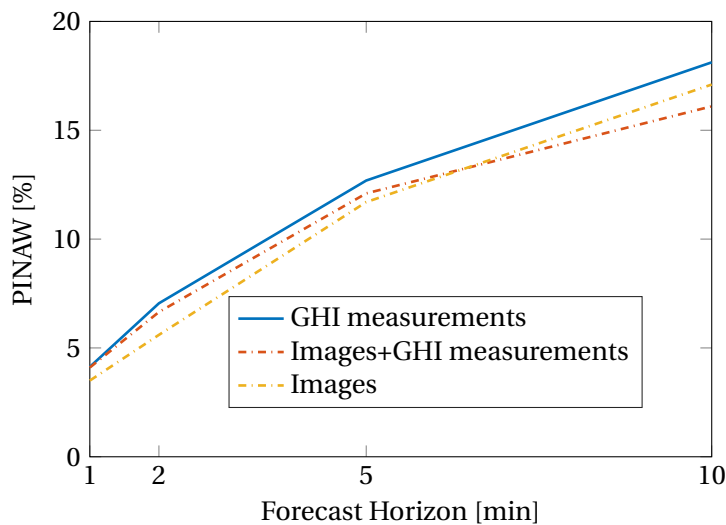


*Figure 4.6 – PINAW of the GHI prediction for different forecast horizons.*

In conclusion, results on GHI prediction intervals shows that including information of the local cloud cover is in general beneficial because it leads to get smaller PIs width. Future work might include the extension of the method to a larger dataset (e.g. one year). Furthermore, it is expected that improving the toolchain (e.g. having a more sophisticated segmentation and cloud motion methods) can lead to a more precise calculation of the local cloud cover, and finally to smaller PIs.

The main drawback of the method is that it still relies on GHI measurements. For this reason, the next step is to evaluate if, when correctly calibrated, the all-sky camera can be used as a GHI sensor itself, this allowing getting rid of the pyranometer or in general of a further sensing system.

## 4.8   All-sky-camera for GHI Estimation

Irradiance ground measurements are generally provided by thermopile or photodiode pyranometers; a comparison of the two technologies for PV-plant monitoring applications can be found in [126]. The use of all-sky cameras for irradiance estimation has recently come to prominence. The fundamental advantage over pyranometers is that all-sky images provide additional information, like the cloud map and movement, which can be relevant for irradiance forecasting processes. In contrast to satellites, the all-sky camera can deliver local high-frequency images (up to sub-second resolution) thanks to its reduced footprint and the possibility of being installed in the proximity of the PV panels.

Measurements of the solar irradiance over large geographical areas are usually based on satellite imaging. Heliosat-2, [127, 128], is the most established physical-based method. It converts observations made by geostationary meteorological satellites into estimates of the global irradiance at ground level, by combining a clear-sky model with a cloud index. The cloud index is calculated from a comparison of what is observed by the sensor and what is observed over that pixel in clear-sky conditions. Other methods using information from satellites normally use supervised regression techniques to map certain features extracted from satellite images to measured irradiance values. E.g., Authors of [129] use Meteosat SEVIRI thermal channels to train two artificial neural networks (ANN), one for estimating the direct normal irradiance (DNI) and one for the diffuse horizontal irradiance (DHI), which are then combined to estimate the GHI. One year of data is used for training and one year of data for testing. The normalized root mean square error (nRMSE) of the estimation at 10 min resolution is 12.4%. The temporal and spatial resolutions of satellite-based estimations are generally site specific and depend on the characteristics of the available fleet of geostationary satellites. E.g., for Europe, Africa, and Indian Ocean, Meteosat-10 and 11 satellites provide real-time imagery with 1-3 km spatial resolution, and 5-15 min temporal resolution. The coarse spatial and temporal resolutions of satellite-based estimations limit their applicability to regional electricity systems. This opens to the need for irradiance estimations using ground-based images.

Recently, Authors of [130] estimate both the direct normal irradiance (DNI), and the diffuse horizontal irradiance (DHI) using information from a camera system with a charge-coupled device (CCD) sensor. The former is estimated by using the pixels in the 16-bit high dynamic range (HDR), the latter by evaluating the intensity of the smear effect on CCDs caused by very bright light. Each of the components requires calibration, done three times during the testing period of 1 year. The estimation nRMSE of GHI is 6-9% compared to pyranometer

measurements and considering hourly averaged data. However, the method applies to CCD sensors only.

Few works focus on the use of image features to build regression models. Authors of [131] present an estimator of the direct, diffuse, and global irradiance using a full-colour sky camera system with a CCD sensor and a rotational shadow-band. Polynomials functions are fitted by selecting the best colour channels combinations to describe each solar radiation component. Four years of data are used in the analysis, showing a nRMSE for 1 min period GHI of about 9% (6% during clear-sky, 13% for partly-cloudy, and 12% for overcast conditions). The work in [132] proposes to estimate the direct, diffuse and global irradiance using a low-cost camera. The irradiance components are estimated by the use of specific radiation transfer functions established by linear regression. The average nRMSE of GHI estimation at 5 min resolution over a one month period is 17%. Authors of [133, 134] use image features as the average pixel intensity of the whole image and the circumsolar area, analysis of the gray-level co-occurrence matrix (GLCM), information of the RGB color space, and the cloud coverage to estimate normalized direct and diffuse irradiance. In particular, features are ranked with decision tree algorithm and then the k-nearest neighbour (k-NN) regression algorithm is applied. The average nRMSE for 10 s GHI estimations is 24%.

We propose here a new method to estimate the GHI from images delivered by an all-sky camera, based on an artificial neural network (ANN). At first, a number of features are extracted from historical all-sky images and from a clear-sky model. Then, feature selection is performed to determine the most relevant features by applying principal component analysis (PCA), [135]. The reduced features are used to perform a supervised training[4] of the ANN, where the target value is the clear-sky index (i.e. the ratio between the measured GHI and clear-sky GHI). At the time of delivering an estimation, the relevant features are extracted from the newly available all-sky image, and input to the ANN to estimate the current clear-sky index, which is finally converted to GHI by multiplying by the clear-sky GHI. The proposed method is denoted as *ASIs-based*. As a further contribution, the obtained results are benchmarked against methods that produce irradiance estimates from satellite images. In particular, we exploit the potentiality of the proposed method to integrate heterogeneous data sources to evaluate whether additional features from satellites images improve the GHI estimations. In the first case, Meteosat SEVIRI spectral channels are used as additional inputs to the proposed method, i.e. PCA feature selection and ANN (we refer to this case as *ASIs/S-based*). In the second case, only the spectral channels from the satellites are used as input to the proposed method, similarly to [129]. We refer to this last case as *S-based*. The GHI estimations obtained in the three cases (ASIs-based, ASIs/S-based, and S-based) are benchmarked against the Heliosat-2 method. The testing dataset covers different periods of the year and different time resolutions (1, 5 and 15 min). As a final analysis, the proposed ASI-s based method is extended to GHI forecasting, namely the output of the ANN is the one-step-ahead clear-sky index rather than the actual one. For this last analysis, results are compared with the persistent forecast.

---

[4]The training is considered supervised when the desired output value is known during the learning phase.

### 4.8.1 ASI-based GHI Estimation

We assume to have access to the following historical information: all-sky images taken at ground level, and GHI measurements used as ground truth.[5] The proposed procedure to estimate the GHI is summarized in Fig. 4.7 and detailed in the following paragraphs. During the training phase (performed off-line), historical all-sky images are pre-processed, and a large set of features is extracted from each image. Then, PCA is applied for feature selection and the selected subset is used for training an ANN, using the clear-sky index (calculated as described in Section 4.8.1) as a target. Once the ANN model is built, the on-line procedure to estimate the GHI can be followed: a new image is captured and pre-processed, and the set of selected features is used as input to the ANN to estimate the clear-sky index. The GHI is then obtained by multiplying the estimated clear-sky index by the clear-sky irradiance.
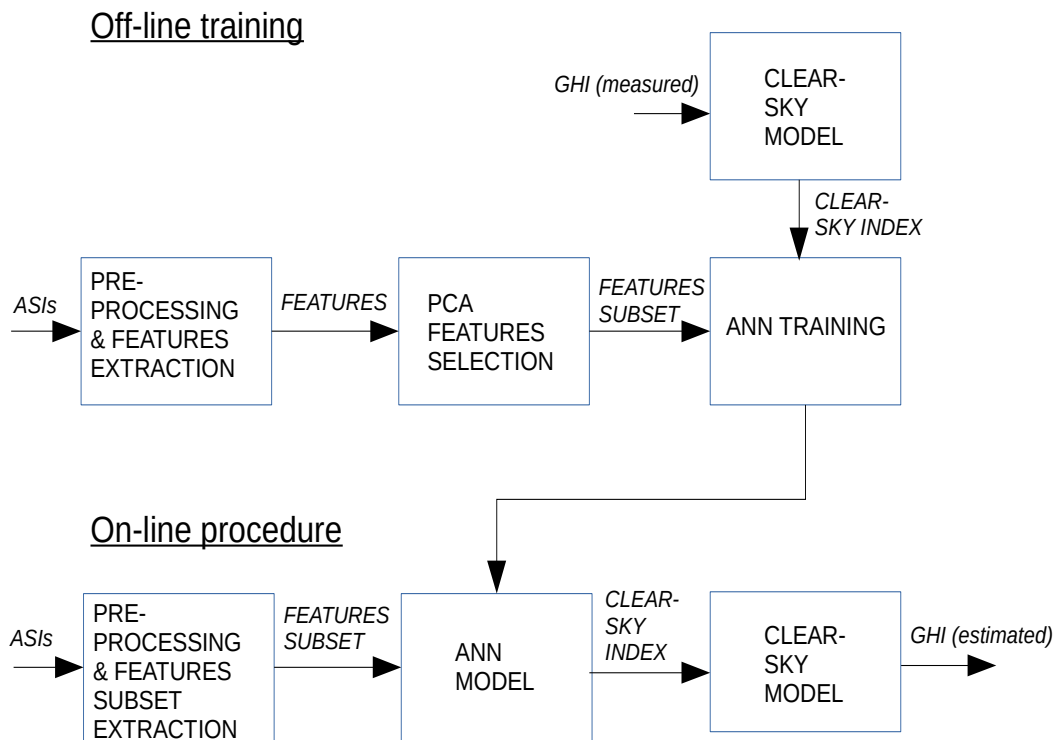


Figure 4.7 – Outline of the proposed method.

**Features Extraction**

The features extracted from the images are 26 in total and are extensively described in the following paragraphs. They are chosen empirically, or inspired from the existing literature,

---

[5]The measurements of the GHI are needed only during the training phase, and a possible final user would not need this further information during the on-line operation.

with the objective of capturing those quantities which might be related with the irradiance incident on the sensor. The considered features are: the average intensity of the pixels and the frequency of saturated pixels in four concentric areas around the sun, the mean of several colour channels representations over the whole image (red, green, blue, hue, saturation, value, luma and chrominance), image texture information (*contrast*, *homogeneity*, *correlation*, *energy*, *skewness*, *kurtosis*), the red-to-blue ratio (RBR), and the red-to-green-ratio (RGR). In addition to the 26 image features, the solar azimuth ($\theta_A$), solar elevation ($\theta_{el}$), and the clear-sky irradiance (GHI$_{cs}$) are used as additional inputs to the ANN and they are computed as described in [104].

Sun areas features: Inspired by [134], solid angles $\Omega$ of 5°, 10°, 15° and 20° are used to define four subtended areas around the sun location. For each area with a set $P$ of $m$ pixels, we compute the average intensity of the pixels as:

$$Average\,Intensity = \frac{1}{3m}\left(\sum_{R \in P} R + \sum_{B \in P} B + \sum_{G \in P} G\right), \tag{4.4}$$

where R, B, G are the red, blue and green pixel values, respectively. Pixel saturation happens when the color channels of the camera sensor assume the maximum value. The frequency of saturated pixels is calculated as:

$$Saturated\,Frequency = \frac{1}{3m}\left(\sum_{R \in P} c_R + \sum_{B \in P} c_B + \sum_{G \in P} c_G\right), \tag{4.5}$$

where for a generic color channel $i$

$$c_i = \begin{cases} 1, & i \geq \text{Threshold} \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

We select a threshold value of 253 assuming an 8-bit representation per color channel.

Color Features: It is possible to represent pixel values in different systems. The most common one is the RGB (Red, Green, Blue), that is physically used by the camera to capture the light information. An alternative system is the HSV (Hue, Saturation, Value), that is obtained through a non-linear transformation of the 3 dimensional space of RGB to an hexcon, see [136]. A third system is the YIQ system (where Y is the luma while I and Q represent the chrominance), that is obtained from the RGB space by applying the following linear transformation:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.701 & -0.587 & -0.114 \\ -0.299 & -0.587 & 0.886 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{4.7}$$

For the three color representation systems, we consider the mean of each color component

over the whole image. If the masked image's pixels set has dimension $M$, the mean values of a generic channel $X$ is defined as:

$$X_{mean} = \frac{1}{M} \sum_{X \in M} X.$$

(4.8)

Texture features: The image grey level co-occurrence matrix (GLCM), [137], can measure the texture of an image, i.e. it gives information about the spatial arrangement of the colors and intensities. In the GLCM matrix each element $(i, j)$ represents the number of times that the pixel valued $i$ occurs horizontally adjacent to a pixel valued $j$. We consider the normalized GLCM, where each element lies between 0 and 1 so that each element can be thought as a probability $p(i, j)$. In our analysis, the distance between the pixel of interest and its neighbour is considered as an offset of value 1 (namely, 0 row and 1 column). Contrast measures the intensity between a pixel and its neighbour over the whole image:

$$contrast = \sum_{i,j} |i - j|^2 p(i, j).$$

(4.9)

Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal:

$$homogeneity = \sum_{i,j} \frac{p(i, j)}{1 + |i + j|}.$$

(4.10)

Correlation measures how correlated a pixel is to its neighbour over the whole image:

$$correlation = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) p(i, j)}{\sigma_i \sigma_j},$$

(4.11)

where $\mu$ and $\sigma$ are the mean and the standard deviation of the row and column elements sums, respectively. Energy represents the smoothness of the image:

$$energy = \sum_{i,j} p(i, j)^2.$$

(4.12)

The entropy of the image:

$$entropy = - \sum_i p_i \log_2(p_i),$$

(4.13)

where $\log_2$ is the base 2 logarithm.

Finally, we add the skewness and the kurtosis computed for the grayscale image. Skewness measures the asymmetry around the distribution mean:

$$skewness = \frac{\frac{1}{M_{GL}} \sum_{I_{GL} \in M_{GL}} (I_{GL} - \bar{I}_{GL})^3}{\sum_{I_{GL} \in M_{GL}} \left( \sqrt{\frac{1}{M_{GL}} (I_{GL} - \bar{I}_{GL})^2} \right)^3},$$

(4.14)

where $M_{GL}$ is the number of pixels in the grayscale image, $I_{GL}$ is the pixel grayscale intensity, and $\bar{I}_{GL}$ the mean pixel intensity over the grayscale image. Kurtosis shows how outlier-prone the distribution is:

$$kurtosis = \frac{\frac{1}{M_{GL}}\sum_{I_{GL} \in M_{GL}}(I_{GL} - \bar{I}_{GL})^4}{\left(\frac{1}{M_{GL}}\sum_{I_{GL} \in M_{GL}}(I_{GL} - \bar{I}_{GL})^2\right)^2}. \tag{4.15}$$

RGB channel ratio: Two more features, usually adopted for cloud detection, are added: the red-to-blue ratio (RBR) and red-to-green-ratio (RGR). The ratios are defined as:

$$\text{RBR} = \frac{1}{M}\sum_{R,B \in M}\frac{R}{B}, \tag{4.16}$$

$$\text{RGR} = \frac{1}{M}\sum_{R,G \in M}\frac{R}{G}. \tag{4.17}$$

**Clear-sky index**

The target value of the proposed estimator is the clear-sky index, $K$, defined as the ratio between the measured GHI and the clear-sky irradiance $\text{GHI}_{cs}$, see 3.2.

The clear-sky irradiance is here calculated using the Ineichen/Perez model, [138, 139], as formulated in the PVLib library, [104]. We remind here that in the proposed method, the clear-sky index $K$ is used as the target of the ANN while the $\text{GHI}_{cs}$ is used as an additional influential variable.

**Principal Component Analysis**

Features selection is carried out to avoid over-fitting, curse dimensionality, and to reduce the computational time needed for training. One of the most common dimension reduction technique is Principal Component Analysis (PCA). PCA is a statistical method that transforms a set of possibly correlated variables into a new set of orthogonal variables, called principal components, so that the greatest variance lies on the few first principal components. In order to evaluate the significance of each original variable we combine the simple PCA with the procedure proposed in [140]. In particular:

- The original dataset consists of $N$ observations and $D$ features ($N$-by-$D$ matrix). We normalize the dataset by subtracting the mean across each feature dimension.

- We compute the covariance matrix of the dataset, along with its eigenvalues and eigenvectors ($V$).

- We determine the $L$ eigenvectors corresponding to the first $L$ largest eigenvalues. The number of principal component $L$ is here selected such that they contain 99.5% of the

total variance.

- We compute the contribution indexes of the $j^{th}$ component as:

$$C_j = \sum_{z=1}^{L} |V_{zj}| \quad j = 1...D. \tag{4.18}$$

- We sort the indices $C_j$ in descending order and select the first $L$ features that represent the reduced dataset in the original space. For example, if $C_x$ and $C_y$ are respectively the first and second largest among all the $C_j$, then we will have $D_1 = x$ and $D_2 = y$, meaning that the $x^{th}$ and $y^{th}$ feature components of the original dataset are the two most important features.

**Artificial Neural Network**

ANNs are used for both regression and classification problems in many fields, including irradiance modelling and forecasting. The general concept behind ANNs is that the input variables are connected to the output by sending signals through layers of units called neurons. In particular, the first layer contains the inputs, the last layer contains the output, and the middle layers contain a given number of hidden neurons. A feedforward network consists of connections that transfer the output of a neuron to the input of the next neuron, each connection characterized by a weight. Optimization algorithms, e.g. conjugate gradients, back-propagation, and Levenberg-Marquardt are used to adjust the weights, with the aim of minimizing a performance function on the training dataset. In our case, the Matlab implementation of the Levenberg-Marquardt algorithm is used to train the ANN, [141]. The performance function is the mean square error between the network predictions and the target outputs. The number of hidden layers is 1, and the number of neurons per layer is 10. These values are selected to avoid overfitting the data and to guarantee a limited computational time.

### 4.8.2 Satellite-based ANN

The same procedure described in Section 4.8.1 is followed, by applying PCA feature selection and then training the ANN. In this case, two possible set of input features are considered:

- Satellites spectral channels are added as features to those selected from the all-sky camera (ASIs/S-based method);

- Only satellites spectral channels are used as features (S-based method).

Rectified (level 1.5) Meteosat SEVIRI Rapid Scan image data are obtained from the EUMETSAT Data Centre at 5 min resolution, [142] and are used to retrieve the values of the spectral channels. High Rate SEVIRI image data consist of geographical arrays of various sizes of image pixels. Each pixel contains information on the received radiation in 12 spectral wavelength

channels. Of these 12 spectral channels, 11 (8 channels in the infrared and three in the solar spectrum) provide measurements with a resolution of 3 km while the $12^{th}$, the High Resolution Visible (HRV) channel, provide measurements with a resolution of 1 km. Spectral channels are retrieved using Pytroll for the pixel corresponding to our location, [143]. In particular, we extract those channels generally considered as sensitive to the constituents of the atmosphere and generally applied to observe wind fields, clouds, water vapor, etc., [129]: IR3.9 (3.9 μm), WV6.2 (6.25 μm), WV7.3 (7.35 μm), IR8.7 (8.7 μm), IR9.7 (9.7 μm), IR10.8 (10.8 μm), IR12.0 (12 μm), IR13.4 (13.4 μm), VIS0.6 (0.635 μm), VIS0.8 (0.81 μm). The first eight channels are converted into brightness temperatures as explained in [144].

### 4.8.3 Heliosat-2

Heliosat-2-based GHI estimations are obtained from the HelioClim-3 database, [145], from the SoDa Service, [146, 147]. We use the HelioClim-3 version 5 (HC3v5), based on the McClear clear-sky model, [148], with forecasted values of aerosols, water vapor, ozone from the Copernicus Atmospheric Monitoring Service. We retrieve the GHI time-series from the service at 1, 5, and 15 min time resolution. The spatial resolution corresponds to the Meteosat images (3 km at Nadir). The acquisition time step of Meteosat is 15 min so the 1 and 5 min sampled time-series are obtained by temporal interpolation. As suggested in [149], satellites-based estimations are refined through a calibration using on-site measurements. The calibration is performed accordingly to what proposed in [146], i.e. by fitting a regression between the GHI measurements and the ground truth daily value (e.g. the local measurements from the pyranometer). Daily GHI local measurements are retrieved for one year at one hour resolution, and time aligned with the Heliosat-2 estimations. The calibration is computed as:

$$\text{GHI}_{\text{cal}} = a\text{GHI}_{\text{day}} + bW^2 + c, \tag{4.19}$$

where $W = \frac{\text{dayofyear}}{365} - 0.5$, $\text{GHI}_{\text{day}}$ is the daily measured irradiance, and a,b, and c the fitting coefficients. The three coefficients are obtained from a regression analysis between the ground truth values and the HC3v5 values. In our case, the calibration allows reducing the annual bias error from 45 to 0.01 W/m$^2$.

### 4.8.4 Results

**Training and Testing Datasets**

All-sky images are selected from the ELP setup described in Section 4.3, frames are taken at 1 min resolution with fixed exposure time. The GHI is sampled at 1 s temporal resolution using the Apogee SP-230 all-season pyranometer described in Section 3.8, installed in the proximity of the camera (3 meters distant). GHI measurements, originally sampled at 1 s resolution, are downsampled to 1 min by averaging the samples. For the analysis, we consider four datasets of 15 days each from different periods of the year (April, July, September, and

*Table 4.6 – Characterization of the datasets.*

|           | Number of Data | GHI mean W/m$^2$ | GHI variability W/m$^2$ |
|-----------|----------------|------------------|-------------------------|
| April     | 5540           | 438              | 200                     |
| July      | 9720           | 412              | 270                     |
| September | 6395           | 304              | 240                     |
| December  | 4400           | 182              | 99                      |

December). Each dataset consists of: measurements of the GHI (considered as the ground truth, originally available at 1 s resolution and then downsampled by average to 1 min), ASIs (originally obtained at 1 min resolution), SEVIRI images from the satellites (originally available at 5 min resolution), and the Heliosat-2-based GHI estimations (originally available at 15 min resolution).The data are first processed and then time-aligned on a local server via Network Time Protocol (NTP).

A quality check procedure is performed for the GHI ground truth measurements, inspired by [150]. In particular:

- the values are set to NaN below a solar elevation angle of 10 degrees and if the measured GHI is negative;

- all the missing values are set to NaN;

- the values which exceed Extremely Rare Limits (ERL):

$$0.03\text{GHI}_T < \text{GHI} < \min(1.2\text{BNI}_T, 1.5\text{BNI}_T cos(\theta_z)^{1.2} + 100) \tag{4.20}$$

  and Physical Possible Limits (PPL):

$$0.03\text{GHI}_T < \text{GHI} < 1.2\text{BNI}_T cos(\theta_z)^{1.2} + 50 \tag{4.21}$$

  are discarded. $\text{GHI}_T$ and $\text{BNI}_T$ are the global and the beam components of the irradiance at the top of atmosphere, and $\theta_z$ the solar zenith angle.

- if the difference between two consecutive GHI measurements is higher than 1000 W/m$^2$ the values are set to NaN;

- any valid single value surrounded by NaN values is also set to NaN.

Table 4.6 shows for the minute samples time-series (April, July, September, and December) the number of data available after the quality check, the average measured GHI, and its variability (here defined as the standard deviation of the difference between the series and its moving average over a 15 min time interval).

GHI measurements from the pyranometer are downsampled by averaging the samples to obtain the time-series for the 5 and 15 min time resolutions. For the analysis at 5 and 15 min, ASIs estimations are also downsampled by averaging the estimations available at 1 min. For the analysis at 1 min, SEVIRI-based information are upsampled assuming they are piecewise constant in the 5 min interval. For the analysis at 15 min, SEVIRI-based information are downsampled by averaging the values. We note that the 5 and 15 min downsample of the data are performed only if at least 100% of the 1 min data are available in that time window. For the analysis at 1 and 5 min, Heliosat-2 estimations are upsampled by a smart temporal interpolation, which takes into account the sun position for each minute. Table 4.7 summarizes the upsampling/downsampling methods adopted for the different types of data.

*Table 4.7 – Data upsampling/downsampling methods.*

|  | 1 min | 5 min | 15 min |
|---|---|---|---|
| Pyranometer | downs. by average | downs. by average | downs. by average |
| ASIs | - | downs. by average | downs. by average |
| SEVIRI | piecewise constant upsample | - | downs. by average |
| Heliosat-2 | smart upsample | smart upsample | - |

It is worth noting that, since temporal information can be converted into spatial information via cloud speed (by simply considering cloud advection), a temporal average of the GHI measurements also represents a spatial average and vice-versa, [11].

The feature selection is performed on a reduced dataset containing 20 days at 1 min resolution. These days are randomly selected from the global dataset of 60 days. Training and testing are computed for each period separately and follow a three-fold cross-validation. Each of the three datasets is further split into three sub-sequences. For each sub-sequence, the ANN is trained on the first fold and tested on the remaining two. Each training fold contains 75% of the data and each testing the remaining 25%. The process is repeated for all the three folds, each time testing the ANN on the part of the data which is not used for the training. In total, the ANN is trained and tested three times for each dataset and for each time resolution. Each sub-sequence contains days with a mix of weather conditions: clear-sky, partly-cloudy, and overcast. The performance metrics reported below are the average of the metrics values obtained in each fold of the three-fold cross-validation.

**Selected Features**

All extracted features are listed in Table 4.8 for the three cases: features from images and solar position (ASIs), combination of ASIs with SEVIRI channels (ASIs/SEVIRI), and SEVIRI channels with solar position (SEVIRI). For all the cases, the clear-sky irradiance $GHI_{cs}$, the solar azimuth $\theta_A$, and the solar elevation $\theta_{el}$ are added as further influential variables. The most influential

features are selected using the procedure described in Section 4.8.1, and are underlined in Table 4.8. The number of selected features is 18 in the ASIs case, 19 in the ASIs/SEVIRI case (only one additional feature from satellites is selected by the PCA), and 7 in the SEVIRI case. Fig. 4.8 shows the percentage of the total variance explained by each principal component and its cumulative sum.

*Table 4.8 – Extracted Features.*

| ASIs | *Average Intensity* $\Omega = 5°$, $\underline{\Omega = 10°}$, $\Omega = 15°$, $\underline{\Omega = 20°}$, *Saturated Frequency* $\Omega = 5°$, $\Omega = 10°$, $\Omega = 15°$, $\underline{\Omega = 20°}$ $R_{mean}$, $\underline{G_{mean}}$, $B_{mean}$, $\underline{H_{mean}}$, $\underline{S_{mean}}$, $\underline{V_{mean}}$, $\underline{Y_{mean}}$, $\underline{I_{mean}}$, $Q_{mean}$, contrast, $\underline{homogeneity}$, $\underline{correlation}$, energy, $\underline{entropy}$, $\underline{skewness}$, kurtosis, $\underline{RBR}$, RGR, $GHI_{cs}$, $\theta_A$, $\underline{\theta_{el}}$ |
|---|---|
| ASIs/SEVIRI | *Average Intensity* $\Omega = 5°$, $\underline{\Omega = 10°}$, $\Omega = 15°$, $\underline{\Omega = 20°}$, *Saturated Frequency* $\Omega = 5°$, $\Omega = 10°$, $\Omega = 15°$, $\underline{\Omega = 20°}$ $R_{mean}$, $\underline{G_{mean}}$, $B_{mean}$, $\underline{H_{mean}}$, $\underline{S_{mean}}$, $\underline{V_{mean}}$, $\underline{Y_{mean}}$, $\underline{I_{mean}}$, $Q_{mean}$, contrast, $\underline{homogeneity}$, $\underline{correlation}$, energy, $\underline{entropy}$, $\underline{skewness}$, kurtosis, $\underline{RBR}$, RGR, $GHI_{cs}$, $\theta_A$, $\underline{\theta_{el}}$, $\underline{WV6.2}$, WV7.3, IR8.7,IR9.7, IR10.8, IR12.0, IR13.4, VIS0.6, VIS0.8 |
| SEVIRI | $GHI_{cs}$, $\theta_A$, $\underline{\theta_{el}}$, $\underline{WV6.2}$, $\underline{WV7.3}$, IR8.7,$\underline{IR9.7}$, IR10.8, IR12.0, $\underline{IR13.4}$, $\underline{VIS0.6}$,$\underline{VIS0.8}$. |

**Metrics**

Let $GHI_t$ be the ground truth GHI at time $t = 1, \ldots, T$, where $T$ is the number of samples, $\widehat{GHI}_t$ the estimation, $GHI_{mean}$ the mean irradiance over the considered period, and $GHI_{max} = 1000$ W/m$^2$. The metrics used to characterize the performance of the proposed techniques are: the per unit (pu) and normalized (n) mean squared error (RMSE), the mean absolute error (MAE), the mean bias error (MBE), and the correlation coefficient (CC):

$$\text{puRMSE} = \frac{1}{GHI_{max}} \cdot \sqrt{\sum_{t=1}^{T} \frac{(\widehat{GHI}_t - GHI_t)^2}{T}}, \tag{4.22}$$

$$\text{nRMSE} = \frac{1}{GHI_{mean}} \cdot \sqrt{\sum_{t=1}^{T} \frac{(\widehat{GHI}_t - GHI_t)^2}{T}}, \tag{4.23}$$

$$\text{MAE} = \sum_{t=1}^{T} \frac{|\widehat{GHI}_t - GHI_t|}{T}, \tag{4.24}$$

$$\text{MBE} = \sum_{t=1}^{T} \frac{\widehat{GHI}_t - GHI_t}{T}, \tag{4.25}$$

$$\text{CC} = \frac{\sum\limits_{t=1}^{T} (\widehat{GHI}_t - \overline{\widehat{GHI}_t})(GHI_t - \overline{GHI_t})}{\sqrt{\sum\limits_{t=1}^{T} (\widehat{GHI}_t - \overline{\widehat{GHI}_t})^2}\sqrt{\sum\limits_{t=1}^{T} (GHI_t - \overline{GHI_t})^2}} \tag{4.26}$$

where the bar symbol represents the average value of the time-series.

**GHI Estimation**

Results of the GHI estimations are shown in Tables 4.9, 4.10, 4.11, and 4.12 respectively for the four considered periods (April, July, September, and December), time resolutions, and investigated methods. The proposed ASIs-based estimations (ASIs) outperform the other cases in terms of all the considered metrics. The relative improvement with respect to the Heliosat-2 (H-2) method ranges between 20-45% in terms of RMSE. The improvement is observable also when considering a time resolution of 15 min (i.e., when satellites information is not upsampled). This is explained by the fact that the all-sky camera is able to capture fast local dynamics caused by cloud passing more accurately than the satellite method that is characterized by a spatial resolution of 3 km. This is also shown by the cross-correlation coefficient that is closer to 1 for the ASIs-based method. It is possible to see that even if the H-2 calibration allows deleting the yearly bias, a monthly bias is still present (MBE is in our case positive for April and July, and negative for September and December). From Tables 4.9-4.12 it is possible to verify that the ASIs/S-based method (ASIs/S) is not better than the ASIs-based, thus denoting that the additional information does not lead to increase the performance. Finally, the case with the SEVIRI features (S) shows the poorest performance with a puRMSE ranging from 11% to 22%. Fig. 4.9 shows the scatter plots of the estimates clear-sky index versus the measured one considering all the available estimations (i.e., all the four periods), and for 1, 5, and 15 min resolution. Left plots refer to the ASIs method and right plots to the H-2. Also from these plots is possible to see that for the ASIs cases the points are more distributed around the 1:1 line. The probability density of the GHI estimation error is shown for the whole considered period and 1 min resolution in Fig. 4.10 for the all-sky camera (ASIs) and Heliosat-2 (H-2) methods (the behaviour at 5 and 15 min resolution is similar and therefore not shown). The ASIs-based method has higher probability around the zero GHI estimation error compared to the H-2.

*Table 4.9 – GHI Estimations Comparison for 15 days in April.*

| | 1 min | | | | 5 min | | | | 15 min | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ASIs | ASIs/S | S | H-2 | ASIs | ASIs/S | S | H-2 | ASIs | ASIs/S | S | H-2 |
| puRMSE [%] | 6.50 | 7.17 | 17.4 | 9.92 | 5.75 | 8.42 | 16.1 | 9.11 | 5.48 | 7.33 | 15.0 | 8.26 |
| nRMSE [%] | 14.8 | 16.4 | 39.8 | 22.6 | 13.1 | 19.2 | 36.7 | 20.7 | 12.5 | 16.7 | 34.2 | 18.8 |
| MAE [W/m$^2$] | 41.2 | 69.4 | 126 | 82.8 | 37.5 | 74.4 | 112 | 75.2 | 38.7 | 47.9 | 114.7 | 70.2 |
| MBE [W/m$^2$] | -4.43 | -2.11 | 13.3 | 29.6 | -3.58 | -3.60 | 13.0 | 28.1 | -2.70 | -4.60 | 13.0 | 29.1 |
| CC | 0.95 | 0.94 | 0.71 | 0.93 | 0.96 | 0.87 | 0.79 | 0.95 | 0.96 | 0.94 | 0.8 | 0.96 |
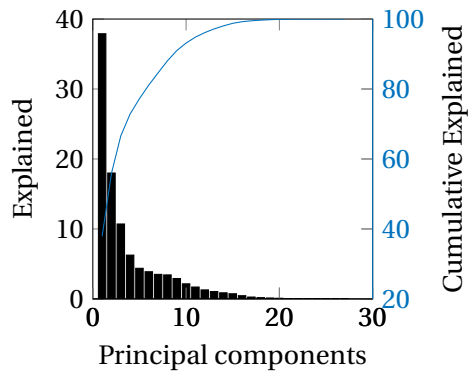
(a) ASIs
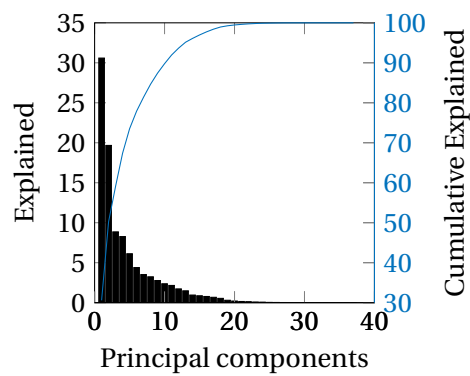


(b) ASsI/SEVIRI
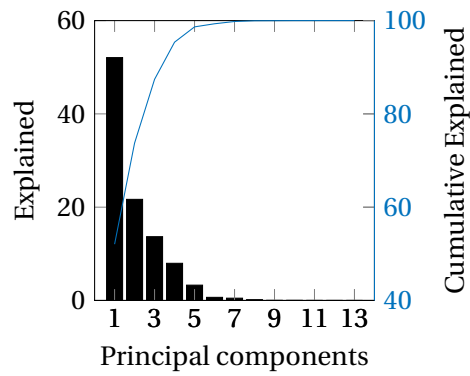


(c) SEVIRI

*Figure 4.8 – Percentage of variance that can be explained by each principal component alone (bat plot, left axis) and its cumulative sum (blue curve, right axis).*

*Table 4.10 – GHI Estimations Comparison for 15 days in July.*

| | 1 min | | | | 5 min | | | | 15 min | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ASIs | ASIs/S | S | H-2 | ASIs | ASIs/S | S | H-2 | ASIs | ASIs/S | S | H-2 |
| puRMSE [%] | 10.2 | 14.6 | 22.3 | 13.9 | 8.46 | 11.6 | 21.5 | 12.6 | 6.45 | 10.8 | 20.8 | 11.1 |
| nRMSE [%] | 24.8 | 35.5 | 54.2 | 33.7 | 20.5 | 28.2 | 52.1 | 30.5 | 15.6 | 26.3 | 50.5 | 26.8 |
| MAE [W/m$^2$] | 73.3 | 98.8 | 170 | 101 | 58.9 | 84.3 | 164 | 94.7 | 46.4 | 72.6 | 163 | 83.0 |
| MBE [W/m$^2$] | 6.42 | 12.0 | 11.0 | 13.7 | 4.6 | 13.7 | 11.3 | 14.5 | 7.3 | 12.1 | 11.2 | 14 |
| CC | 0.91 | 0.83 | 0.46 | 0.85 | 0.94 | 0.90 | 0.53 | 0.87 | 0.96 | 0.89 | 0.62 | 0.90 |

*Table 4.11 – GHI Estimations Comparison for 15 days in September.*

| | 1 min | | | | 5 min | | | | 15 min | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ASIs | ASIs/S | S | H-2 | ASIs | ASIs/S | S | H-2 | ASIs | ASIs/S | S | H-2 |
| puRMSE [%] | 9.20 | 10.7 | 23.4 | 11.8 | 7.03 | 8.14 | 23.6 | 10.8 | 6.97 | 7.9 | 23.9 | 9.67 |
| nRMSE [%] | 30.2 | 35.3 | 77.3 | 38.9 | 23.1 | 26.7 | 77.6 | 35.6 | 22.9 | 25.9 | 78.8 | 31.8 |
| MAE [W/m$^2$] | 55.8 | 68.9 | 87 | 85.8 | 44.6 | 64.9 | 85.0 | 79.5 | 54.0 | 62.3 | 86.1 | 71.1 |
| MBE [W/m$^2$] | -16.6 | -20.5 | -59.0 | -30.3 | -18.3 | -19.5 | -58.5 | -30.3 | -22.0 | -19.0 | -59.2 | -29.1 |
| CC | 0.89 | 0.81 | 0.23 | 0.80 | 0.90 | 0.89 | 0.23 | 0.81 | 0.88 | 0.87 | 0.24 | 0.81 |

*Table 4.12 – GHI Estimations Comparison for 15 days in December.*

| | 1 min | | | | 5 min | | | | 15 min | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ASIs | ASIs/S | S | H-2 | ASIs | ASIs/S | S | H-2 | ASIs | ASIs/S | S | H-2 |
| puRMSE [%] | 2.81 | 4.55 | 11.6 | 5.10 | 2.64 | 3.23 | 11.3 | 4.9 | 2.49 | 3.0 | 12.5 | 4.72 |
| nRMSE [%] | 15.4 | 25.1 | 63.8 | 27.9 | 14.5 | 17.7 | 62.2 | 27.0 | 13.6 | 16.4 | 69.0 | 25.9 |
| MAE [W/m$^2$] | 18.6 | 28.3 | 97.0 | 42.3 | 16.6 | 21.1 | 71.0 | 40.6 | 16.0 | 19.7 | 79.3 | 38.0 |
| MBE [W/m$^2$] | 3.50 | -5.5 | 30.1 | -19.6 | 4.52 | -4.30 | 25.0 | -19.5 | 5.93 | -5.0 | 22.7 | -20.0 |
| CC | 0.96 | 0.91 | 0.5 | 0.94 | 0.98 | 0.95 | 0.47 | 0.91 | 0.98 | 0.94 | 0.43 | 0.93 |

**Cloud-free and Partly-cloudy Analysis**

In this Section, the ASIs-based GHI estimations are compared against the Heliosat-2 on a testing set of selected days that differ in terms of weather conditions. The cases including the SEVIRI information are discarded from this analysis due to the poor performance.

Estimation performance is evaluated on eight prevalently cloud-free days and eight prevalently partly-cloudy days. The days are extracted from the whole period (i.e., two days in April, two days in July, two days in September, and two days in December), while all the remaining days are used to train the ANN [6]. The analysis refers to 1 min time resolution and results are in Table 4.13. We can see that, during clear-sky days, the H-2 method has similar performance of the ASI-s based method (with an identical CC), while during partly-cloudy days the proposed method improves the H-2 (relative improvement of 30% RMSE). This is also reflected in Figures 4.9 where, for the H-2 method, a high concentration of data points is visible close to an estimated clear-sky index equal to one, i.e. close to clear-sky conditions, while data are more sparse for clear-sky index values in the range 0.3-0.9 (i.e. intermediate overcast conditions).

---

[6] The selection of the cloud-free and clear-sky days is performed by visual inspections.
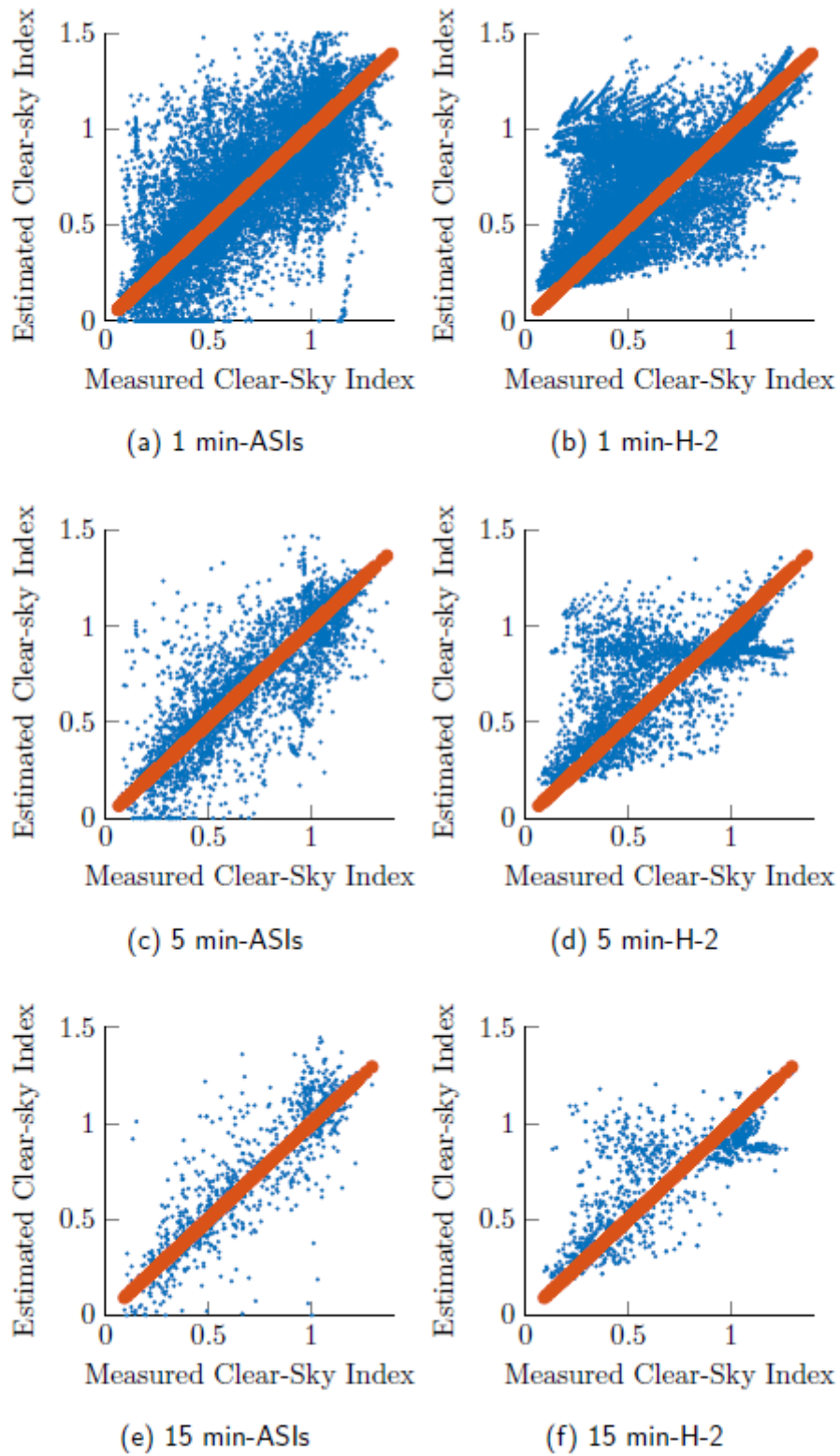
Figure 4.9 – Estimated vs Measured clear-sky index for the ASIs-based estimation (left plots) and the Heliosat-2 (right plots). The red line is the 1:1 line.
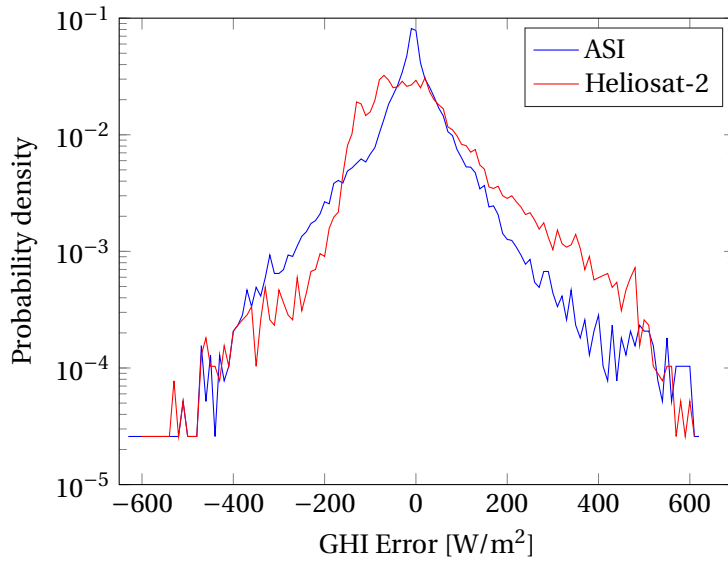
*Figure 4.10 – Probability density of the GHI estimation error, defined as the difference between the estimation and the measurement. The plot refers to the whole testing dataset (April, July, September, and December) and to 1 min time resolution. The y-axis is in logarithmic scale.*

The improvement from the ASI-based method is significant during periods of high variability when rapid irradiance fluctuations generate challenging conditions for the power system. Examples of measured and estimated GHI are shown in Fig. 4.11 for two typical weather conditions. During a cloud-free day the H-2 method matches the measurements slightly better than the ASIs-based, while during partly-cloudy conditions the proposed method is able to better follow the fast irradiance fluctuations, thanks to the availability of high-resolution ground images.

*Table 4.13 – GHI estimations comparison for selected days at 1 min.*

|  | Cloud-free | | Partly-cloudy | |
| --- | --- | --- | --- | --- |
|  | ASI | H-2 | ASI | H-2 |
| puRMSE [%] | 6.7 | 6.8 | 12.7 | 17.5 |
| nRMSE [%] | 13.3 | 13.6 | 34.9 | 47.9 |
| MAE [W/m$^2$] | 43.3 | 46.0 | 82.8 | 103 |
| MBE [W/m$^2$] | 3.7 | 5.0 | 22.0 | 6.1 |
| CC | 0.97 | 0.97 | 0.89 | 0.80 |

(a) Cloud-free day        (b) Partly-cloudy day
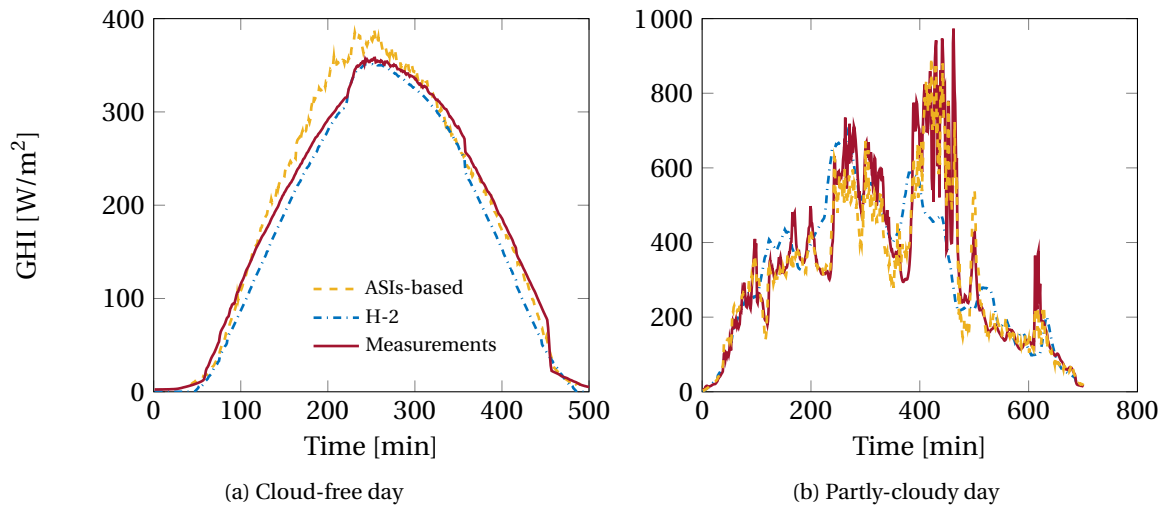
*Figure 4.11 – ASIs-based and the H-2 GHI estimations along with measurements from the pyranometer for a cloud-free and a partly-cloudy day.*

**Training set Analysis**

This Section investigates the importance of training the ANN using days from a similar period of the year of the testing days. We consider 30 days between June and July containing a mix of weather conditions: clear-sky, partly-cloudy, and overcast. The last 5 calendar days are used as a fixed testing set, while the training set can change accordingly to two case scenarios. In the first case, called *moving training*, the length of the training set is fixed to 15 days: the estimator is initially trained with the oldest data (with respect to the testing), and then, at each iteration, the time window of the testing data is progressively moved forward by one day. In the second case, called *increasing training*, the estimator is initially fitted with the oldest 15 days long set, and progressively a new day is added (one day each iteration). The procedures last until the last day of training corresponds to the day before the first day of testing. The two scenarios are illustrated in Fig. 4.12. The nRMSE values of the ASI-s based estimations are shown in Fig. 4.13 as a function of the time period between the last calendar day of training and the first calendar day of testing. Results are compared against the H-2 estimations which shows a constant nRMSE because the method does not require training on historical data. We can observe that when the time period between the last day of training and the first day of testing is lower than 15 days, the performance is stable and comparable to the results obtained in Section 4.8.4. However, for a time period larger than 15-20 days, performance starts deteriorating rapidly, and, after that, the proposed method performs poorer than the Heliosat-2. This analysis suggests that the training set should include days that are exemplary of the periods of the year under test. The similarity in the behaviour of the two training cases suggests that a single ANN model, trained on a set of days that are representative of different weather conditions, is sufficient to guarantee satisfactory performance of the method for that period. In order to show this, we evaluate the performance of the method when the training is
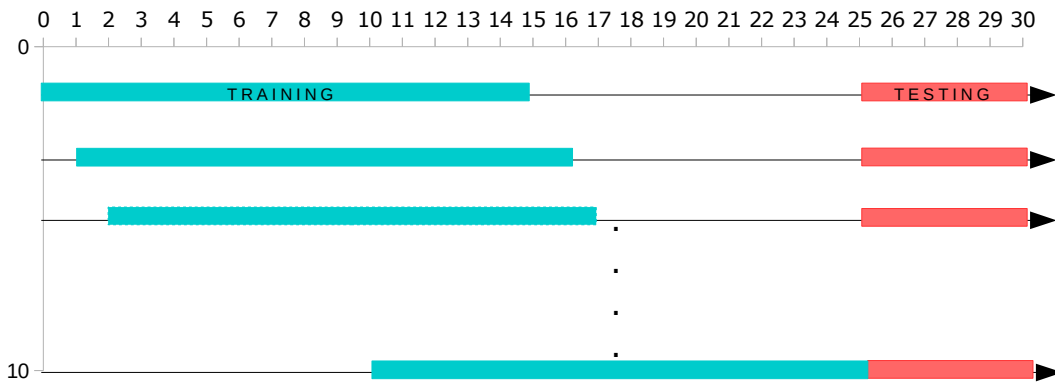
(a) Moving training
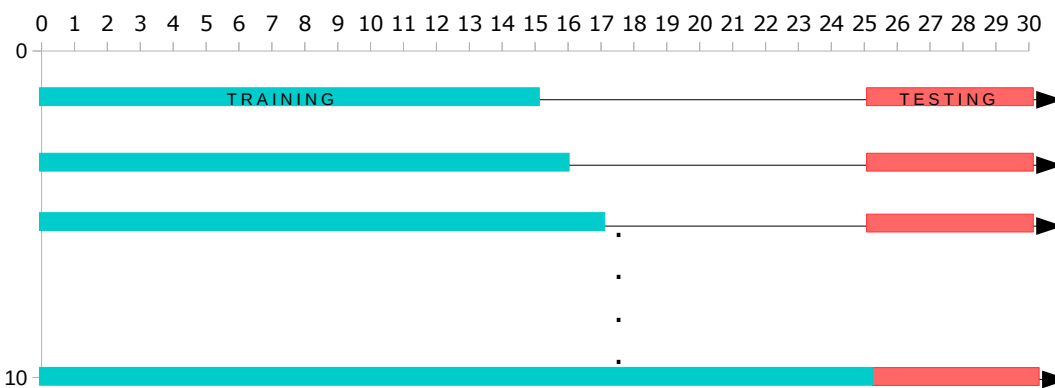


(b) Increasing training

*Figure 4.12 – Moving training and increasing training (blue bar) along a period of 30 days; the last 5 days are used as testing (red bar).*



Number of days between last day of training and first day of testing
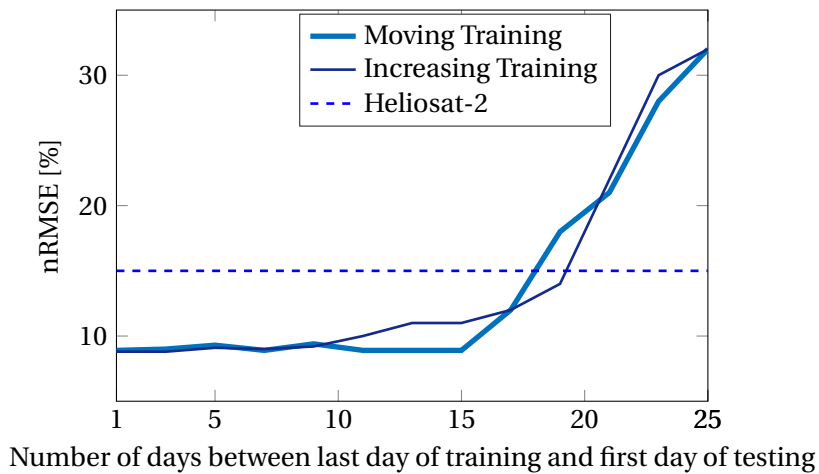
*Figure 4.13 – nRMSE is calculated on a fixed testing set of 5 days in July and for a changing training set. It is possible to observe that performance deteriorates when the time period between the two sets (training/testing) increases.*

*Table 4.14 – GHI estimations using the ASIs-based algorithm for July 2018.*

| | |
|---|---|
| puRMSE [%] | 9.6 |
| nRMSE [%] | 19.9 |
| MAE [W/m$^2$] | 69 |
| MBE [W/m$^2$] | -56 |
| CC | 0.97 |

done in one month in July 2017, and the testing in one month in July 2018. Results are shown in Table 4.14. It is possible to conclude that if the method is trained with days that are far in time (1 year) from the testing, but from the same period of the year (similar light conditions), the performance above-discussed is guaranteed. In other words, it is in principle possible to calibrate the algorithm with one year of data and then use the camera to sense the GHI in the following years, without the need of an intra-year retraining.

### 4.8.5 Computational Time

When a new image is available, the on-line procedure of the ASIs-based method (image pre-processing, features extractions and application of the ANN model to estimate the actual GHI) takes less that 2 s on a personal computer (Intel Core i7-6600 2.6 GHz, with 12 GB RAM). The proposed method can be implemented to capture local fast irradiance dynamics, which represents an advantage over satellite-based estimations characterized by a sparser temporal and spatial resolution.

### 4.8.6 Extension to Forecasting

In this Section, we evaluate the possibility of extending the proposed ASIs-based method to GHI point forecasting. The main differences with respect to the ASIs-based GHI estimation discussed so far, is that the target of the ANN is the one-step-ahead realization instead of the current clear-sky index. In other words, if the features are extracted at time $t$, the corresponding target is the clear-sky index at time $t + \Delta t$. For this analysis, we consider 1 month of data at 1 min resolution (August 2017), and we target different forecast horizons, namely $\Delta t$ can assume values 1, 5, 10, 15 min. Even in this case we perform a three-fold cross-validation and 75% of the data are used for training and the remaining 25% for testing. The persistent point forecast is used as benchmark (it means that the one-step-ahead forecast is assumed equal to the current GHI measurements from the pyranometer). Results are shown in Fig. 4.14 where the obtained puRMSE is shown for different forecast horizons. It is possible to see that the proposed method outperforms the baseline case (persistent forecast) in terms on puRMSE for forecast horizons higher than 1 min. This preliminary analysis confirms that the all-sky image contains meaningful information that can be used to improve forecasting performance with
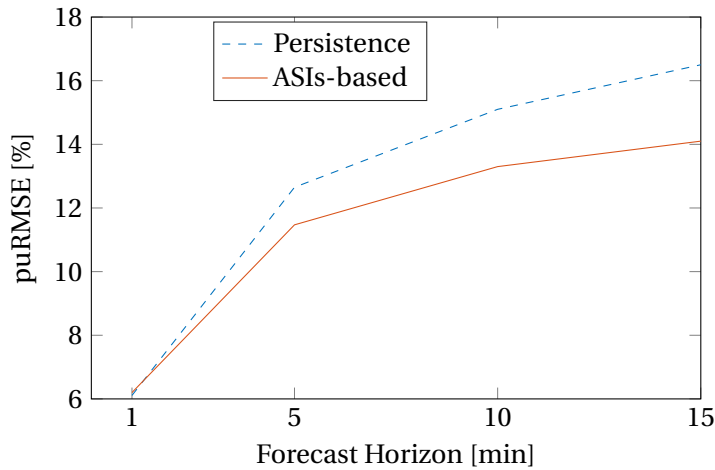
*Figure 4.14 – puRMSE as a function of the forecast horizon. The proposed ASIs-based forecast is benchmarked against the persistent method.*

respect to the baseline method. Future work might consider to include features that accounts for cloud dynamics for further improving the performance.

## 4.9   Chapter Conclusions

In the first part of the Chapter, a chain to predict the local cloud cover in the sun disk is proposed. This includes the analysis of several cloud segmentation and motion methods. Regarding binary cloud segmentation, a new method based on the intensity-to-blue ratio is proposed and proved to outperform available methods based on different color channels (e.g, the red-blue ratio) and supervised segmentation models. A preliminary analysis on different cloud motion algorithms has shown that the optical flow is the most promising method. Furthermore, we have shown that the information on the local cloud cover can be used to improve a machine learning based method to deliver GHI PIs, besides the sole use of past irradiance measurements proposed in the previous chapter. Since the main drawback of the proposed toolchain is that the prediction method is still dependent on the pyranometer, in the second part we evaluate the possibility of using the sky-camera as a GHI sensor itself. With this aim, a machine learning-based method to estimate the GHI using images from an all-sky camera is proposed. At first, a large set of features is extracted from the images, and then feature selection is performed using principal component analysis (PCA). The subset of selected features is then used as input to train an artificial neural network (ANN), considering the clear-sky index as a target. Furthermore, SEVIRI thermal channels are used alone and in combination with features from the camera as input to the same procedure (PCA and ANN), to evaluate if this information can improve the GHI estimation. Performance evaluation is presented using pyranometer measurements, collected in the same location of the camera, as ground truth. The analysis considers for different periods of the year and three different time resolutions (i.e., 1, 5, and 15 min). The GHI estimations are benchmarked against the

Heliosat-2, a well-established method to estimate the GHI from satellites (e.g. Meteosat in Europe). Results show that the all-sky camera-based GHI estimations proposed in this work outperform the Heliosat-2, with a RMSE relative improvement of 20-45%. In particular, this happens when fast irradiance dynamics are present (e.g. during partly-cloudy conditions), due to the fact that satellites-based models lack the spatial and temporal resolution needed to capture localised fluctuations. Results also suggest that the additional information from satellites thermal channels are not improving the estimation. The method has been extended to the case of GHI point forecast, showing an improvement with respect to a simple persistent method for forecast horizons higher than 1 min. Future work will evaluate how to integrate the information from the two best performing methods (ASIs-based and H-2), for example by distinguishing between clear-sky and partly-cloudy conditions beforehand.

# 5 Indirect Power Forecast of PV generation: from Irradiance to DC Power

## 5.1 Chapter Highlights and Summary

When estimations of the irradiance are available from sensors (e.g. pyranometers, sky-images, satellites) they can be used as input to a physical PV model to reconstruct the maximum power value.[1] Alternatively, a second approach consists in using measurements from the PV system itself to compute the irradiance on the plant. The two approaches are sketched in Fig. 5.1. An analysis on the first approach, showing the results on the DC maximum
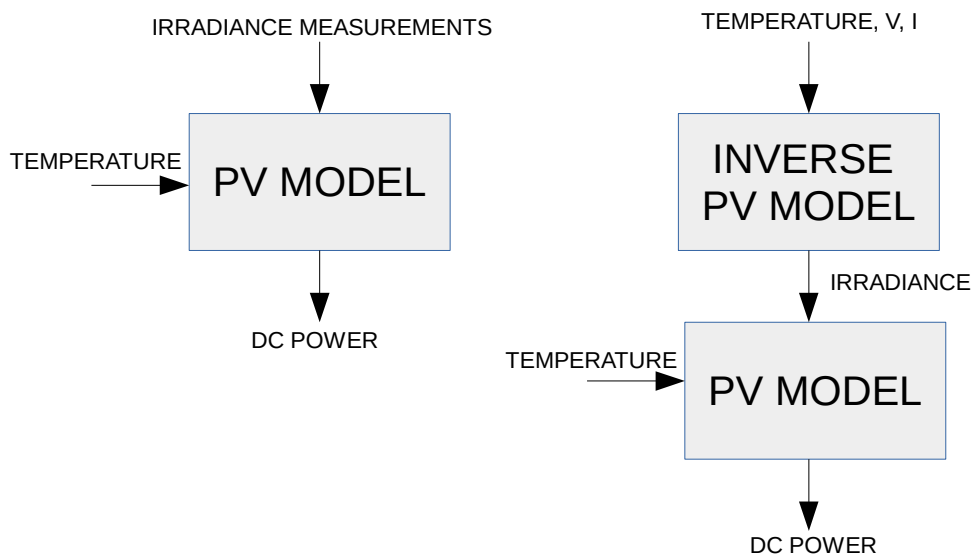


*Figure 5.1 – DC Power estimations. In the left case the irradiance is measured while in the right case it is estimated using an inverse model of the PV plant. Then the irradiance and the temperature are given as input to a PV model to estimate the PV maximum power.*

---

[1]We note that in this Chapter the term irradiance refers to the irradiance on the plane of array (POA) so that we do not need any transposition models. Transpositions model are presented in Appendix B.

power estimation starting from different sensing systems (pyranometer, sky images..) can be found in Appendix B. In this Chapter, we focus on the second approach and we propose three model-based methods to reconstruct the maximum power of a PV plant starting from experimental measurements of the DC voltage, current and temperature. Two formulations from the existing literature (the analytical from [151] and the Immersion & Invariance (I&I) from [152]) are considered and extended to estimate the theoretical maximum power of an entire PV array, irrespectively of its operating conditions. A third estimator, based on a Kalman Filter (KF), is proposed here and included in the performance assessment. More specifically the Chapter focuses on these contributions:

- a formal comparison of PV maximum power estimators is performed using experimental data from a 14.3 kWp PV-roof installation;

- performance is validated with measurements from a dedicated experimental setup which allows to account for MPPT and non-MPPT operating conditions;

- the maximum power estimations of the analyzed methods are compared with those obtained starting from pyranometer readings;

- the rejection to measurements noise and the bandwidth of the estimators are formally compared;

- a practical application of the methods to improve data-driven maximum power forecasting tools is proposed. We show that training the forecasting algorithm with historical data that are corrupted by exogenous disturbances might lead to a deterioration of the forecasting tool, in terms of its ability to predict the maximum power. In this context, we apply the analyzed methods to reconstruct the maximum point and we therefore make sure that the forecasting algorithm learns from values that are always representative of the true potential of the irradiance.

## 5.2 Introduction to DC Maximum Power Computation

The knowledge of the PV generation is normally achieved by directly measuring the power injected by the PV inverter into the grid. However, local measurements might be affected by the action of (unobserved) control actions, e.g. due to curtailment policies [153]. These exogenous disturbances make the observations uncorrelated with the true PV maximum power and, consequently, with the true irradiance. Due to their lack of representativeness, the use of these observations as proxy measurements, or as training data for PV forecasting, would have a detrimental impact on the estimation of the PV state and prediction performance, respectively. For this reason, indirect methods are advisable and mainly consist in inputting the measured irradiance and temperature into physical models of the PV system, to finally compute the PV injected power. As example, Fig. 5.2 depicts a situation where a PV plant operates in curtailment regime (star). The production at this working point is not representative of

the true irradiance potential, which corresponds instead to the maximum power point (dot). Fig. 5.3 shows real measurements of an entire day when the AC power is curtailed and therefore the measured power is not correlated with the corresponding irradiance.
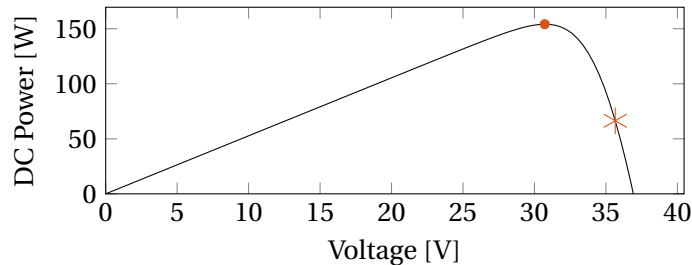


*Figure 5.2 – Example of measured power (star) and maximum power (dot) shown for one PV module.*



(a) AC Power Profile                    (b) Irradiance Profile
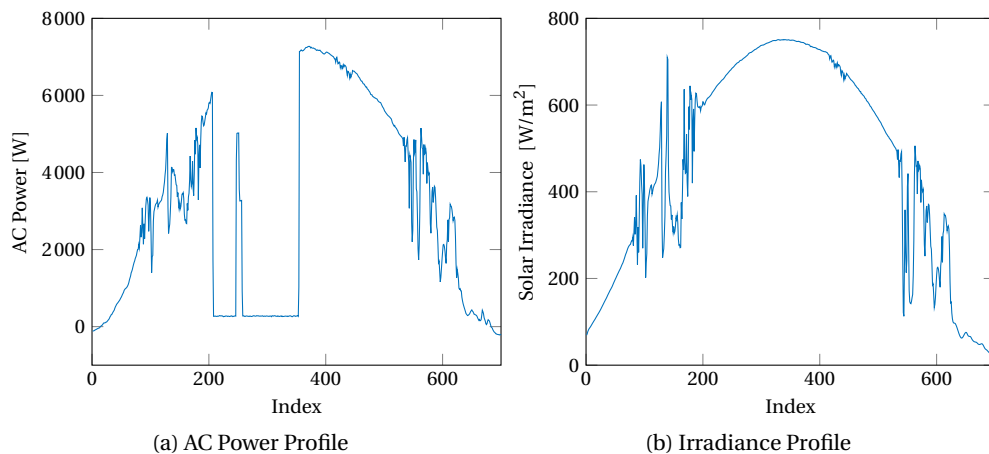
*Figure 5.3 – On the left plot the AC power of the PV plant is curtailed from index 200 to index 380. On the right the irradiance profile is plotted to show that the measured power and the solar irradiance are not correlated when curtailment happens.*

In this Chapter, we propose a novel methodology to reconstruct the maximum power production of a PV plant using model-based methods that need measurements of the array DC voltage, DC current, and cell temperature as inputs. Similar approaches have already been exploited in the existing literature to estimate the irradiance from a single PV module. Authors of [151] propose a closed-form analytical estimator of the irradiance based on the same inputs considered in this work. Similarly, Authors of [152] propose a globally convergent estimator based on the immersion and invariance (*I&I*) principle. In [154], temperature and DC electrical measurements are used to perform real-time estimation of the irradiance. The method has the drawback of requiring the PV system to move in three different states (panel under load, short circuit or open circuit), a feature that is not normally implemented in commercial PV systems. In [155], an approach based on neural network is also proposed to estimate the

irradiance starting from cell temperature and electrical DC information. The proposed model is implemented in a microcontroller and used to infer the irradiance. In [156], a model-based approach is used to implement a maximum power point tracking (MPPT) algorithm. The method requires a pyranometer for the identification of the model parameters and is validated for a single module. It is to note that there are some advantages in estimating the power output using measurements of electrical quantities rather than information from a pyranometer. On one hand, voltage and current measurements are generally available from the converter management system and the module temperature is easy to measure, [155]. On the other hand, irradiance sensors are sensitive to calibration, [157], and return measurements that are significative for a specific point rather than gathering the average irradiance conditions on the installation. For this last reason, several pyranometers might be necessary, this increasing the overall cost. Furthermore, when observations from a pyranometer are not available nor reliable, one might consider to estimate the irradiance and the PV production potential by exploiting the knowledge of the measured production of nearby PV installations. Also in this case, one should consider that the PV power output could be curtailed, thus the measured power would not be representative of the true irradiance potential and its maximum value should be reconstructed.

## 5.3 PV Plant Model

We introduce here the one-diode five-parameter model from [3], shown in Fig. 5.4 for a single PV cell. This model is accurate, simple to implement, and only requires datasheet information normally available from the panel manufacturer. The adopted model $f(\cdot)$ describes the relationship between the DC voltage $v$, current $i$, solar irradiance on the PV module plane $I^{\searrow}$, and cell temperature $T$, for a PV panel composed of $n_p$ cells in parallel and $n_s$ in series. The model is:

$$
f(v, i, T, I^{\searrow}) = 0 = i_p(T, I^{\searrow}) n_p + - \frac{v + R_s i \frac{n_s}{n_p}}{R_p(I^{\searrow}) \frac{n_s}{n_p}} - i +
$$
$$
- i_D(I^{\searrow}) n_p \left[ \exp\left( q \frac{v + R_s i \frac{n_s}{n_p}}{n_r k T n_s} \right) - 1 \right],
$$
(5.1)

where $k, q$ are physical constants and stand for diode Boltzmann constant and electron charge, respectively. The parameters $R_s, R_p, I_p, i_d, n_r$ respectively denote the series resistance, shunt resistance, light current, saturation current and diode ideality factor. Let the notation $*$ denote values at Standard Test Condition (STC), i.e. temperature $T^* = 25\,^\circ$C and irradiance

$I^* = 1000\,\text{W}\,\text{m}^{-2}$, the five parameters at different conditions $(T, I^{\searrow})$ are calculated as:

$$R_s = R_s^* \tag{5.2}$$

$$R_p = R_p^* \frac{I^*}{I^{\searrow}} \tag{5.3}$$

$$i_p = \left(i_p^* + \alpha\left(T - T^*\right)\right)\frac{I^{\searrow}}{I^*} \tag{5.4}$$

$$n_r = n_r^* \tag{5.5}$$

$$i_D = i_D^*[T/T^*]^3 \exp\left(E_g^*/kT^* - E_g/kT\right) \tag{5.6}$$

$$E_g = 1.17 - 4.73 \times 10^{-4}\frac{T^2}{T+636}, \tag{5.7}$$

where $E_g^*$ is the band gap energy (eV) at $T^*$, while $R_s^*, R_p^*, I_p^*, i_D^*, n_r^*$ are the parameters at STC, calculated with the procedure detailed in [158] by using the following datasheet information: the open circuit voltage $v_{OC}^*$, the short circuit current $i_{SC}^*$, the voltage and the current at the maximum power $v_{MP}^*$ and $i_{MP}^*$, the absolute temperature coefficients of the open circuit voltage $\beta$ and short circuit current $\alpha$.
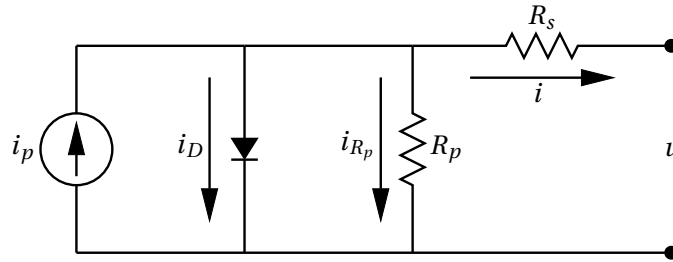


*Figure 5.4 – The adopted five parameter circuit model of a PV cell.*

## 5.4 Estimators of the Solar Irradiance

This Section describes three selected methods to estimate the irradiance received by a PV array (Analytical, Immersion and Invariance, and Extended Kalman Filter). The inputs are the DC voltage, DC current, and the cell temperature while the single diode model introduced in Section 5.3 is used to describe the PV system, for all the three cases. The output of the model (estimated irradiance, $I^{\searrow}$) is then used to reconstruct the maximum power as described in Section 5.5. We assume that the measured cell temperature is representative for the whole plant, in other words we assume a uniform temperature distribution. To test the fairness of this modelling assumption, two identical temperature sensors were installed in different parts of the plant. They recorded an average temperature difference lower than 0.4°C and a maximum value of ≈0.8 °C. Such a value determines a difference in the estimated maximum power of ≈0.5%, thus making the assumption acceptable for the considered rooftop PV installation. In case of larger scale PV plants, this assumption should be re-evaluated, and the single-diode model could be replicated considering more temperature sensing points. The sensitivity of

the models to temperature measurement errors is further discussed in Section 5.7.4.

### 5.4.1 Analytical Formulation

The irradiance is calculated analytically and in a closed-form by substituting equations (5.2)-(5.6) into (5.1) and solving for $I^\searrow$. Formally, it is:

$$I_A^\searrow = \frac{i + i_D^* n_p [T/T^*]^3 \exp\left(E_g^*/kT^* - E_g/kT\right)\left[\exp\left(q\frac{v+R_S^* i n_s/n_p}{n_r kT n_s}\right) - 1\right]}{\frac{1}{S^*}\left[n_p\left(I_p^* + \alpha(T - T^*)\right) - \frac{v+R_S^* i n_s/n_p}{R_p^* n_s/n_p}\right]} \tag{5.8}$$

where $v$, $i$ and $T$ are measured quantities and $I_A^\searrow$ is the inferred irradiance.

### 5.4.2 Immersion and Invariance

Authors of [152] design an estimator exploiting the fact that the $i$-$v$ characteristics described by Eq. (5.1) can be re-parametrized to show a monotonic behavior. The estimator is based on the principles of immersion and invariance, originally described in [159]. The re-parameterization, based on the model in Eq. (5.1), is as follows. They define a *measurable* signal $y(t)$:

$$y(t) = i(t) - F(i(t), v(t), T(t)) \tag{5.9}$$

where,

$$F(i, v, T) = I_0(T)\exp\left(\frac{C_1}{T}(v + C_2 i) - 1\right) \tag{5.10}$$

$$I_0(T) = -C_6 T^3 \exp\left(C_7 - \frac{C_8}{T} + \frac{C_9 T}{T + C_{10}}\right). \tag{5.11}$$

$C_i$ are constant values that can be calculated from single-diode equations and found in [152]. The only difference is in the definition of constant $C_3$ since Authors of [152] consider a proportional relation between $R_p$ and $I^\searrow$, while we here consider inverse proportionality, as expressed in Eq. (5.3).

Then, they express the nonlinear regression form as:

$$y(t) = \Phi(I^\searrow, t) \tag{5.12}$$

$$\Phi(I^\searrow, t) = I^\searrow (C_4 + C_5 T(t)) - C_3/S(v(t) + C_2 i(t)) \tag{5.13}$$

with $\Phi(I^\searrow, t)$ strictly monotonically increasing with $I^\searrow$. At this point, the immersion and invariance estimator states:

$$\dot{I_I^\searrow} = \gamma\left[y - \phi(I_I^\searrow)\right] \tag{5.14}$$

and $\gamma > 0$ ensures:

$$\lim_{t \to \infty} I_I^\searrow(t) = I^\searrow \tag{5.15}$$

where $I_I^\searrow$ is the estimated irradiance. Performance depends on the value of parameter $\gamma$ that should be selected as a trade-off between convergence speed and noise filtering.

### 5.4.3 Extended Kalman Filter

We propose to apply Kalman Filter to estimate the irradiance as a function of voltage, current, and temperature measurements[2]. The advantage of a KF over a conventional low pass filter is that, by exploiting the knowledge of the process model, it achieves to filter out system disturbances and measurements noise on the whole spectrum of the state variables. We consider a linear discrete-time system described by:

$$x_k = F_{k-1} x_{k-1} + G_{k-1} u_{k-1} + w_{k-1}, \tag{5.16}$$

$$y_k = H_k x_k + v_k, \tag{5.17}$$

$$E[w_k w_j^T] = Q_k \delta_{k-j}, \tag{5.18}$$

$$E[v_k v_j^T] = R_k \delta_{k-j}. \tag{5.19}$$

The goal is to estimate the state $x_k$, knowing some noisy measurements $y_k$, and the system dynamics. $F_{k-1}$ is the state-transition model, $H_k$ the observation model, and $u_{k-1}$ indicates external control variables. The noise $w_k$ and $v_k$ are white, zero-mean, uncorrelated and have as covariance matrices $Q_k$ and $R_k$, respectively. $\delta_{k-j}$ is the Kronecker delta function. When all the measurements including time $k$ are used to estimate $x_k$, the estimation is called a-posteriori:

$$\hat{x}_k^+ = E[x_k | y_1...y_k]. \tag{5.20}$$

For the a-posteriori state estimate and covariance we can write:

$$\hat{x}_k^+ = (I - K_k H_k)(F_{k-1} \hat{x}_{k-1}^+ + G_{k-1} u_{k-1}) + K_k y_k \tag{5.21}$$

$$P_k^+ = (I - K_k H_k)(F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1}) \tag{5.22}$$

$$K_k = (F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1}) H_k^T (H_k (F_{k-1} P_{k-1}^+ F_{k-1}^T + Q_{k-1}) H_k^T + R_k)^{-1} \tag{5.23}$$

where $P_k$ denotes the the covariance of the estimated error:

$$P_k^+ = E[(x_k - \hat{x}_k^+)(x_k - \hat{x}_k^+)^T], \tag{5.24}$$

---

[2]As known from the existing bibliography, Kalman estimation consists in reconstructing the state of a system with noisy measurements by integrating the knowledge of the process which generated them.

and $\boldsymbol{K_k}$ is the Kalman Gain. If $\boldsymbol{w_k}$ and $\boldsymbol{v_k}$ are white, zero-mean, uncorrelated, then the Kalman Filter is the optimal linear solution of the problem. In the Extended Kalman Filter (EKF), the state transition and observation models are not linear functions of the state. However, it is possible to linearize the non-linear function around the current estimate (e.g. by making a Taylor series expansion and dropping all but the constant and linear terms).

In what follow we describe how to apply the EKF to estimate the irradiance. The prerequisite to apply Kalman filtering is the knowledge of the system model and covariance matrices of system noise and measurements. To this end, we exploit the results from Chapter 3, where it is shown that the irradiance evolution in the few seconds time scale can be captured with a persistence model plus a random variation from an identifiable pdf (probability density function), which is function of certain data features. Let the state $x_k = I_k^{\nearrow}$ be the irradiance on the panel. At each discrete time $k$, the measurements $v_k, i_k$ and $T_k$ are linked to the state by the nonlinear relationship $f(\cdot)$ in Eq. (5.1). Let $f_1(\cdot), f_2(\cdot), f_3(\cdot)$ denote the function $f(\cdot)$ solved for voltage, current and temperature:

$$v_k = f_1(x_k, i_k, T_k) \tag{5.25}$$
$$i_k = f_2(x_k, v_k, T_k) \tag{5.26}$$
$$T_k = f_3(x_k, v_k, i_k). \tag{5.27}$$

The observation vector $\boldsymbol{y}_k = [v_k, i_k, T_k]^T$ is approximated as:

$$\boldsymbol{y}_k \approx \boldsymbol{H_k} x_k + \boldsymbol{D_k}, \tag{5.28}$$

where $\boldsymbol{H} = [H_1, H_2, H_3]^T$ and $\boldsymbol{D} = [D_1, D_2, D_3]^T$ are from first order Taylor expansions of $f_1, f_2, f_3$. For example, for the case of $f_1$, they are:

$$v_k \approx f_1(\cdot) + f_{1,x}(\cdot)(x_k - a_k) \tag{5.29}$$
$$H_1 = f_{1,x}(\cdot) \tag{5.30}$$
$$D_1 = f_1(\cdot) - f_{1,x}(\cdot) \cdot a_k \tag{5.31}$$

where $f_1(\cdot)$ and $f_{1,x}(\cdot)$ denotes the function and its first order derivative calculated in the point $a_k, i_k, T_k$, with $a_k$ is the irradiance value around which linearising (assumed as the last available estimate, i.e. $a_k = x_{k-1}$) and $i_k$ and $T_k$ are both from measurements.

The state-space formulation of the system model is:

$$x_k = F_{k-1} x_{k-1} + w_{k-1}, \qquad\qquad w_{k-1} \in N(0, Q_k) \tag{5.32}$$
$$\boldsymbol{y}_k = \boldsymbol{H_k} x_k + \boldsymbol{u}_k, \qquad\qquad \boldsymbol{u}_k \in N(\boldsymbol{0}, R_k). \tag{5.33}$$

where $F_{k-1} = 1$ is the (scalar) system matrix, $Q_k$ is the process noise variance, and $R_k$ the $3 \times 3$ measurement noise covariance matrix.

The variance $Q$ is computed by applying the method described in Chapter 3 and summarized

in the Appendix A. The covariance matrix of measurements noise $R$ is a diagonal matrix $R = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. Measurements are assumed to be uncorrelated. The variance components are calculated assuming that the tolerance of the sensors corresponds to the the 3-sigma level of a Gaussian distribution with zero mean, i.e. $\sigma_i = t_i/3, i = 1, 2, 3$, where $t_i$ is the tolerance of the instrument $i$. Once $\boldsymbol{H_k}$, $Q_k$ and $R_k$ are known from the procedures described above, the expected value $\hat{x} = \text{E}[x_k]$ and variance $P_k = \text{Var}[x_k]$ of the estimation are:

$$\hat{x}_k = (I_y - \boldsymbol{K_k}\boldsymbol{H_k})(F_{k-1}\hat{x}_{k-1}) + \boldsymbol{K_k}\boldsymbol{y}_k \tag{5.34}$$

$$P_k = (I_y - \boldsymbol{K_k}\boldsymbol{H_k})(F_{k-1}P_{k-1}F_{k-1}^T + Q_{k-1}), \tag{5.35}$$

where $I_y$ is the identity matrix, and $\boldsymbol{K_k}$ is the Kalman gain:

$$\boldsymbol{K_k} = P_{k-1}\boldsymbol{H_k}^T(\boldsymbol{H_k}P_{k-1}\boldsymbol{H_k}^T + R_k)^{-1}. \tag{5.36}$$

We note that the linearization of the observer equation leads to an Extended Kalman filter (EKF) formulation. Unlike the linear KF, the EKF violates the guarantee of optimality of the solution.

## 5.5 Maximum Power Computation

Once the irradiance is estimated from any of the methods proposed in Section 5.4 or measured by a pyranometer, it is then used, together with $T$, to compute the maximum power output according to the following procedure:

- as defined in [160], the open circuit voltage $v_{OC}$ is:

$$v_{OC} = v_{OC}^* \left(1 + \beta(T - T^*)\right) + V_t n_r n_s \ln\left(\frac{I^\search}{I^*}\right) \tag{5.37}$$

where $V_t = kT/q$ is the thermal voltage;

- we determine the $i$-$v$ curve by applying Eq. (5.1) for values of the DC voltage $v$ varying between 0 and $v_{OC}$[3];

- the maximum power of the PV module is computed as the maximum product $i \cdot v$ from the $i$-$v$ curve;

- the array maximum power is obtained by multiplying the module maximum power by the number of modules.

---

[3]The diode equation, (5.1), can be solved numerically or using the explicit formulation based on the Lambert function, described in [161], that gives an exact analytical solution for $i$ as a function of $v$ and it is computationally more efficient.

## 5.6 Experimental Setup

The considered experimental setup is a 14.3 kWp rooftop PV installation of the EPFL Romande Energie solar park in Switzerland and consists of two identical subsystems. Each subsystem is composed of one couple of strings (each with with 14 ECSOLAR 255 W Polycrystalline modules in series) connected to a three-phase DC/AC power converter. Panels are south facing and with 10 degrees tilt from the horizontal plane. The whole system is shown in Fig. 5.5, where the individual subsystems are marked with red and blue color. The two converters, denoted by C1 and C2, are of the same commercial model and work independently. They can operate in MPPT mode (using the Perturb and Observe method) or follow a specific active power external request (non-MPPT mode). This dedicated setup allows validating the capability of the proposed estimators also when the plant operates in non-MPPT mode, thanks to sending specific set-points requests to C1 while leaving C2 in MPPT mode (its output power is considered as the ground truth maximum power value). The equivalent behavior of the two converters has been tested experimentally by running both of them in MPPT mode: in this case, their power output differ of less than 0.2%. DC currents are measured with LEM LF 205-S current transducers with an accuracy of $\pm$ 0.2%. DC voltages are measured with LEM-CV 3-100 voltage transducers with an accuracy of $\pm$0.5%, as described in Section 2.2.1 for the AC measurements. The irradiance on the plane-of-array is measured with a silicon-cell device to limit the spectral mismatch due to the different spectral response between the modules and the sensor. Thus, an Apogee SP-230 all-season pyranometer is installed in the same location of the plant, as described in Section 3.8. The cell temperature is measured using a TSic303, a sensor with $\pm$0.5 °C accuracy, installed on the rear surface of the panel, as detailed in [162]. In order to correct the temperature readings accounting for the thermal resistance of the support material, we follow the procedure described in [163] and add a positive offset with magnitude $n \cdot 3$ °C, where $n$ is dimensionless irradiance (we here use $1000 \, \mathrm{W m^{-2}}$ as base quantity). The value of $n$ is calculated using the irradiance estimated at the previous time-step, which is a fair assumption since our estimations are at high time resolution and the temperature dynamics are slower than the irradiance ones. The location of the pyranometer and temperature sensor is marked with a white cross in Fig. 5.5. Analog measurements of voltage and current are acquired at 20 kHz with an 18 bit analog-to-digital converter (ADC, NI CompactRIO 9068 equipped with a 9215 module), while irradiance and temperature are sampled 20 kHz with a NI sbRIO 9625 with a 16 bit ADC. All the measurements are then downsampled by average at 1 s and saved in a database. The two acquisition devices mentioned above are synchronized, and the two groups of measurements are with a time jitter of 0.5 s at most. Only daylight data are selected for the analysis, i.e. for solar elevation larger than 3 degrees.

## 5.7 Results

In this Section, the performance of the methods in Section 5.4 is compared in terms of quality of the estimations of the maximum power they provide. Summarizing from the previous sections, this consists in i) computing the irradiance using the analytical (Analytical), immersion

*Figure 5.5 – PV installation at the EPFL laboratory (GPS coordinates 46.52-N, 6.56-E). The strings indicated with blue color are connected to C1 and the strings indicated with red color are connected to C2. The white cross indicates the panel where the temperature sensor and the pyranometer are installed.*

and invariance (I&I), and extended Kalman Filter (EKF) methods and ii) estimating the maximum power with the procedure described in Section 5.5. The quality of the maximum power estimations is assessed by comparing them against the measured power flow (ground truth value) of converter C2 that is always in MPPT mode. As a further benchmark, we include the maximum power estimation performed by using measurements from the pyranometer: this consists in feeding the procedure for the maximum power estimation in Section 5.5 with POA irradiance readings and cell temperature measurements. The Section is organized as follows. Section 5.7.1 presents the metrics used for the comparison. Sections 5.7.2 and 5.7.3 compare the performance of the proposed methods for maximum power estimation for non-MPPT and MPPT conditions, respectively. Finally, in Section 5.7.4 the robustness of the proposed estimators in terms of rejection against measurements noise is assessed by decreasing their signal-to-noise ratio. Two days are considered for the analysis: a clear-sky (9 of September 2016) and a partly cloudy day (12 of September 2016).

### 5.7.1 Metrics

Let $P_{dc,t}$ be the ground truth DC maximum power value at the time interval $t = 1,\ldots,T$, where $T$ is the number of samples, $\widehat{P}_{dc,t}$ the estimation, $\overline{P}_{dc}$ the average over time of the ground truth values in the interval $t = 1,\ldots,T$. Three metrics are used to characterize the performance of the proposed techniques, the normalized root mean squared error (nRMSE), the normalized

maximum error ($\text{Err}_{\max}$) and the normalized mean error (nME):

$$\text{nRMSE} = \frac{1}{\overline{P}_{dc}} \sqrt{\sum_{t=1}^{m} (\hat{P}_{dc,t} - P_{dc,t})^2 / m}, \tag{5.38}$$

$$\text{Err}_{\max} = \max\left\{ |\hat{P}_{dc,t} - P_{dc,t}| / \overline{P}_{dc}, \ t = 1, \ldots, m \right\}, \tag{5.39}$$

$$\text{nME} = \frac{1}{m} \sum_{t=1}^{m} (\hat{P}_{dc,t} - P_{dc,t}) / \overline{P}_{dc}. \tag{5.40}$$

### 5.7.2 Maximum Power Estimation in non-MPPT Conditions

Tables 5.1 and 5.2 show the performance of the proposed techniques and of the pyranometer-based estimations, when the power output of converter C1 is curtailed as shown in Fig. 5.6, for a clear-sky and partly cloudy day respectively. As explained in Section 5.4.2, performance of the I&I estimator depends on the value of parameter $\gamma$ that should be selected as a trade-off between good convergence and noise filtering. Authors of [152] suggest values in the range 0.1-1, thus we first fix $\gamma$ equal to 0.7. Performance denotes that the methods are able



(a) Clear-sky day

(b) Partly cloudy day
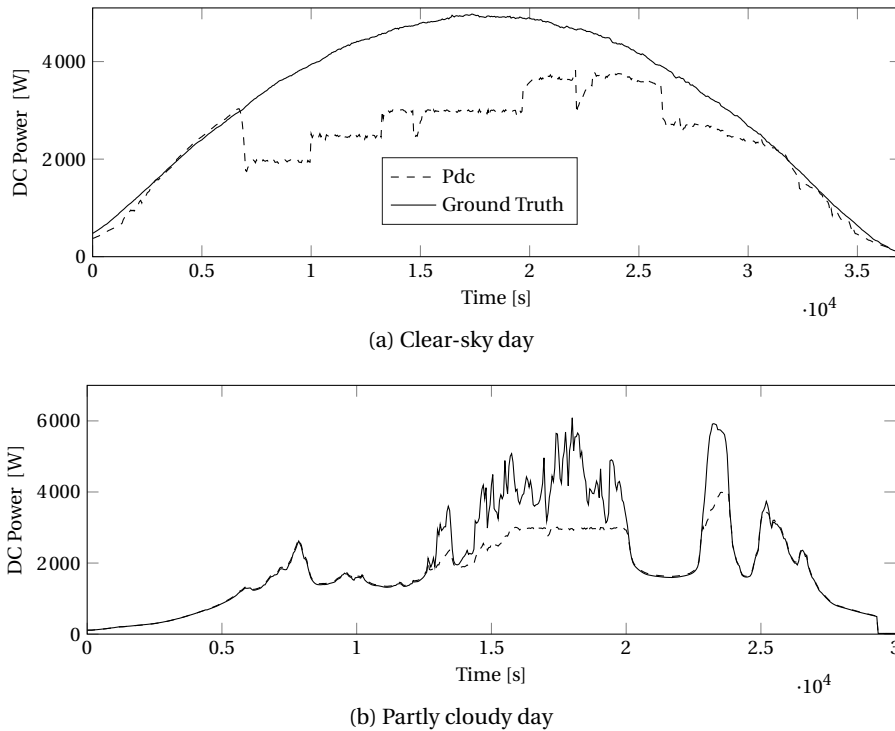
*Figure 5.6 – Maximum power (Ground Truth) and curtailed power (Pdc) are shown. Measurements come from two identical converters (same technology) that are working under equivalent conditions but different modes (i.e. MPPT for C2 and non-MPPT/curtailed for C1).*

to reconstruct the maximum output power starting from any operating point, as shown in Fig. 5.7. Indeed, it is possible to reconstruct the maximum power even when starting from
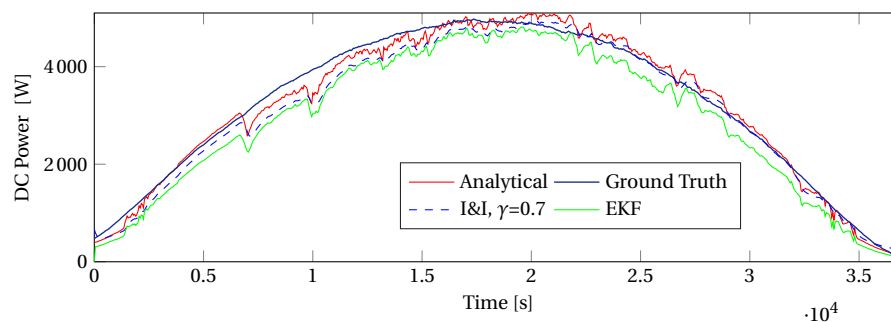
(a) Maximum Power Comparison/Clear-sky
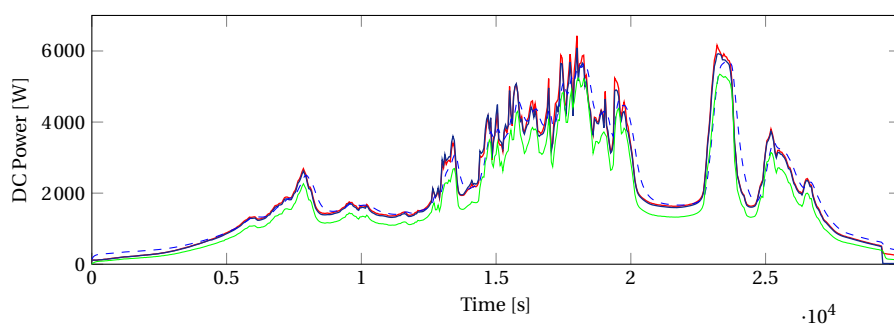


(b) Maximum Power Comparison/Partly Cloudy

*Figure 5.7 – Comparison between the ground truth maximum power and the reconstructed maximum power using the analytical, the EKF, and the I&I estimators.*

curtailed conditions. The analytical method outperforms the other estimators, during both clear-sky and partly cloudy conditions. For the pyranometer-based estimations, larger errors are likely generated when irradiance measurements are not representative of the conditions of the whole plant. Figures 5.8a and 5.8b compare the ground truth maximum power values against the analytical and pyranometer-based estimations for the clear-sky and partly cloudy day, respectively. For a more detailed comparison we refer to numerical results in Tables 5.1 and 5.2.

Tables 5.1 and 5.2 also show that the EKF and the I&I (with $\gamma$=0.7) estimators have worse performance than the others, especially in the partly cloudy case. This fact is justified by their smaller bandwidth. In support to this, we compare in Figures 5.9a and 5.9b the amplitudes of the fast Fourier transform (FFT) of the ground truth values and estimations, for the irradiance and maximum power respectively. We consider 6200 samples taken from a period with large power variations, in the central part of the cloudy day. As visible from Fig. 5.9a, only the analytical estimation is able to track the irradiance fluctuations registered by the pyranometer, especially at high frequency. Similarly, in Fig. 5.9b we observe that the EKF and I&I-based power estimations tend to filter out fast but significative dynamics, thus determining poorer performance. On the contrary, the analytical formulation is able to follow the ground truth series even at high frequencies. A further analysis, consists in gradually increasing the value of $\gamma$ in order to favor the I&I performance over the noise filtering. Results show that values of $\gamma$ in

the range 20-200 (for the partly cloudy day) and 5-200 (for the clear-sky day) allow achieving performance very close to the analytical case, thus denoting that the choice of $\gamma$ is critical and its selection may represent a drawback with respect to parameter-less methods as the analytical one. The worse results of the EKF are due to the non-stationary of the irradiance time series, even when differentiated, which makes difficult to identify an exact model of the process.

*Table 5.1 – Performance for a clear-sky day/non-MPPT.*

|  | nRMSE [%] | Err$_{max}$ [%] | nME [%] |
|---|---|---|---|
| Analytical | 5.40 | 20.4 | -0.8 |
| EKF | 11.53 | 21.2 | -10.4 |
| I&I, $\gamma$=0.7 | 7.46 | 20.1 | -4.57 |
| Pyranometer | 6.65 | 15.1 | -2.76 |

*Table 5.2 – Performance for a partly cloudy day/ non-MPPT.*

|  | nRMSE [%] | Err$_{max}$ [%] | nME [%] |
|---|---|---|---|
| Analytical | 5.6 | 7.15 | 2.53 |
| EKF | 19.9 | 30.0 | -13.9 |
| I&I, $\gamma$=0.7 | 19.9 | 23.0 | 5.17 |
| Pyranometer | 9.16 | 32.8 | 7.14 |

### 5.7.3 Maximum Power Estimation in MPPT Conditions

Tables 5.3 and 5.4 show the performance of the proposed techniques and of the pyranometer-based estimations when the system is in MPPT mode for the whole period, for a clear-sky and a partly cloudy day, respectively. The analysis confirms what observed for the non-MPPT case. Results from the pyranometer estimations are independent of the PV system state and are therefore identical for the MPPT and non-MPPT analysis. On the other hand, the analytical method is the most sensitive to the operating point conditions and its performance improves when the system is close to the maximum power point. As far as the I&I method is concerned, same results of Section 5.7.2 apply regarding the tuning of parameter $\gamma$.

### 5.7.4 Robustness Against Measurements Noise

The EKF and I&I-based estimations are expected to have improved rejection against measurements noise, thanks to integrating into the estimation process measurements which are progressively becoming available with time. In this Section, we deliberately decrease the signal-to-noise ratio of the measurements with the objective of showing the break-even

(a) Comparison with Pyranometer/Clear-sky



(b) Comparison with Pyranometer/Partly Cloudy
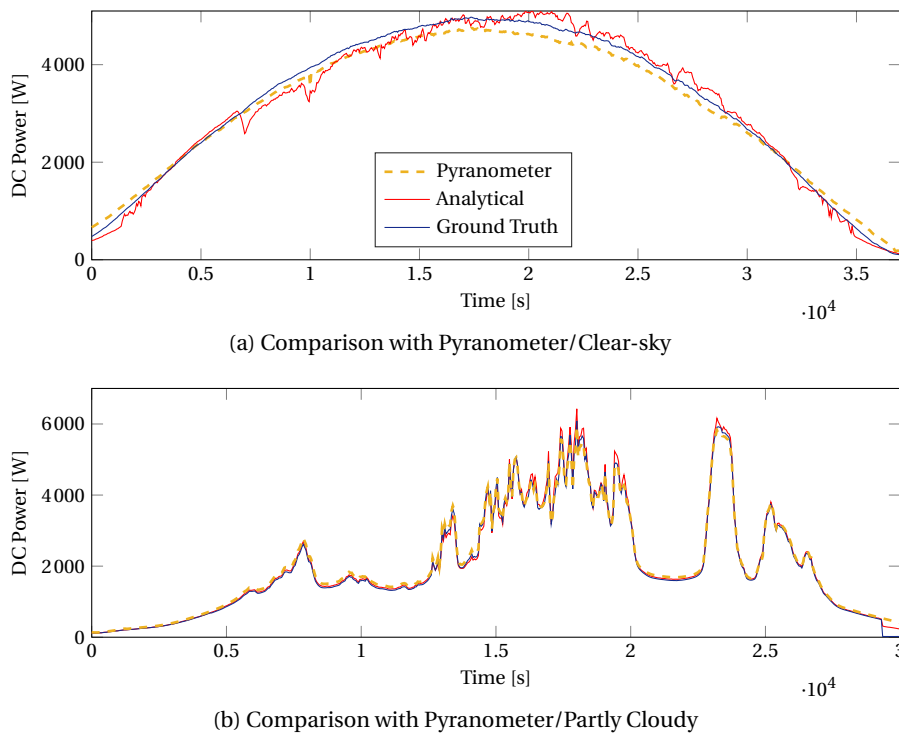
*Figure 5.8 – Comparison between the ground truth maximum power, the estimation using the analytical formulation and the one starting from the irradiance sensed by a pyranometer.*

*Table 5.3 – Performance for a clear-sky day/MPPT.*

|  | nRMSE [%] | $Err_{max}$ [%] | nME [%] |
|---|---|---|---|
| Analytical | 0.51 | 1.5 | 0.37 |
| EKF | 11.53 | 18 | -13.3 |
| I&I, $\gamma$=0.7 | 4.21 | 6.8 | -3.51 |
| Pyranometer | 6.65 | 15.1 | -2.76 |

*Table 5.4 – Performance for a partly cloudy day/MPPT.*

|  | nRMSE [%] | $Err_{max}$ [%] | nME [%] |
|---|---|---|---|
| Analytical | 3.7 | 5 | 1.51 |
| EKF | 19.9 | 35.1 | -14.1 |
| I&I, $\gamma$=0.7 | 19.9 | 22 | 4.06 |
| Pyranometer | 9.16 | 32.8 | 7.14 |

between the performance of the estimators. In other words, we want to investigate on the level of measurement noise after which the use of filter-based estimators is advisable compared to the analytical formulation. Original voltage, current and temperature measurements (which
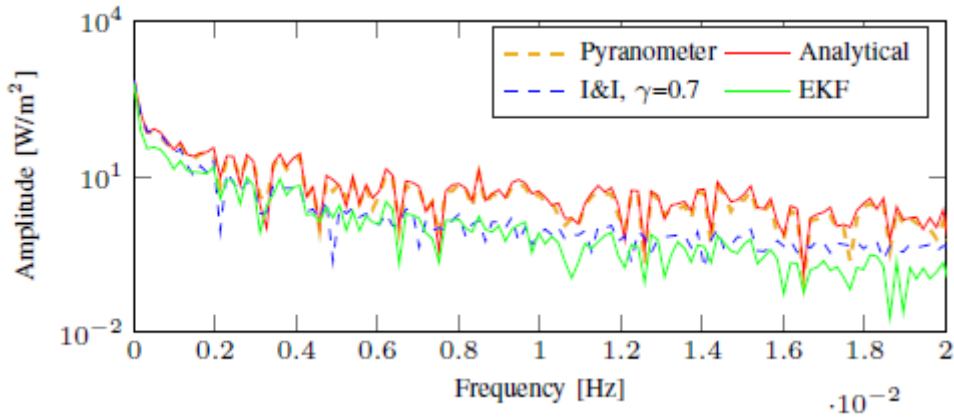
(a) FFT of Irradiance Series
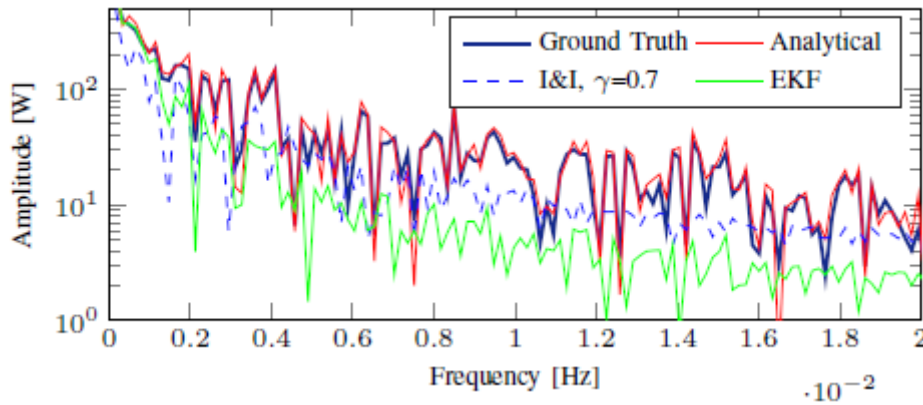


(b) FFT of Power Series

*Figure 5.9 – The FFT single-sided amplitude spectrum of the estimated and measured irradiances is shown as a function of the frequency in a semilogarithmic scale in Fig. 5.9a. The same spectrum is shown for the estimated and ground truth maximum power values in Fig. 5.9b.*

are already characterized by the original noise inherent the respective sensing devices) are corrupted with an additive independent and identically distributed (i.i.d.) Gaussian noise with increasing values of standard deviation (STD). This analysis is carried out for each measurement in a separate fashion, namely noise is added to a single measurement (voltage, current or temperature) while keeping the others to their original level of noise.[4] The original levels of noise and those artificial added are recapped in Table 5.5.

The analysis considering the additional noise on voltage, current and temperature are shown in Figures 5.10a, 5.11a and 5.12a for the MPPT case, and in Figures 5.10b, 5.11b and 5.12b for the non-MPPT case. The x-axis shows the total level of noise standard deviation (i.e. original noise

---

[4]In the EKF case, the element of matrix $R$ corresponding to the noisy measurement is augmented with the value of the variance of the fictitious additional noise. Assuming that the variables are uncorrelated, we can say that the variance of the sum equals the sum of the variances.

*Table 5.5 – Standard Deviations (STD) of the Input Measurements*

|       | Original noise (STD) | Added Noise (STD)    |
|-------|----------------------|----------------------|
| $i$   | 0.55                 | 0.3, 0.45, 0.55, 0.7 |
| $v$   | 0.23                 | 0.3, 0.71, 1         |
| $T$   | 0.4                  | 1.4, 2.23, 3.16      |

level plus additive i.i.d. noise) while the y-axis the nRMSE associated with the maximum power estimation. The partly cloudy day is considered for the analysis. For the I&I we consider two cases: a first with $\gamma$ fixed to 1 and a second case where $\gamma$ is tuned to optimize the performance ($\gamma_{opt}$). Figures 5.10, 5.11 5.12 show that the tuning of parameter $\gamma$ allows the I&I estimator to have stable performance even for high level of noise. When the noise augments, performance of the analytical method deteriorates while the I&I estimator with $\gamma_{opt}$ maintains a low nRMSE. For the other cases (EKF and I&I with $\gamma = 1$) we can see that break-evens with respect to the analytical performance happen for an STD of $\approx$0.68 for the current and $\approx$0.82 for the voltage on the MPPT case. For the non-MPPT case, the break-even occurs for an STD of $\approx$0.63 for the current $\approx$0.58 for the voltage. This break-even has, in general, a lower value in the curtailed case, for which the analytical estimation is more sensitive to the presence of measurement noise. Fig. 5.12 shows that the analytical estimation is less sensitive to temperature noise, especially when working close to MPPT conditions. The break-even is reached for very high STD, corresponding to an STD of $\approx$3.3 for the non-MPPT case.



(a) MPPT

(b) Curtailed
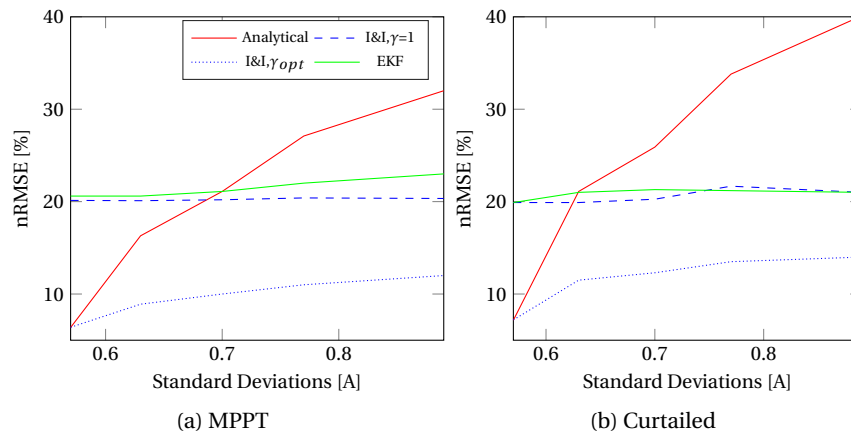
*Figure 5.10 – Noise is increased on the DC current.*

## 5.8  Improvement of Data-driven Forecasting Methods

We show how the discussed estimators are applied to treat historical production time series of a PV plant to filter out exogenous control action components (e.g. due to curtailment) from a training dataset, by reconstructing the maximum power value. The objective is to show how
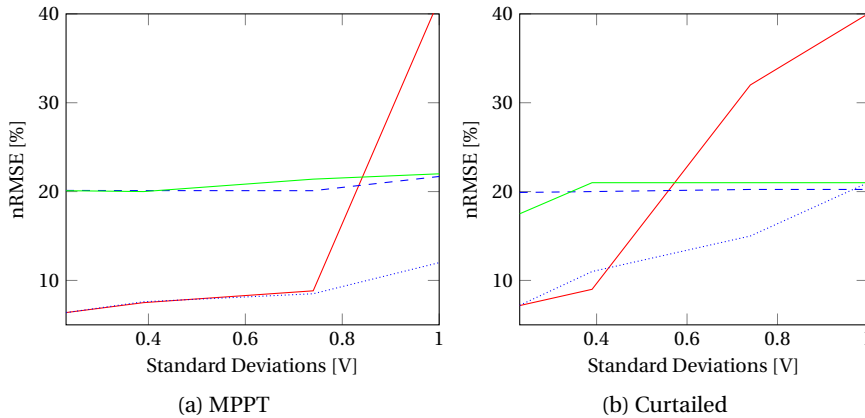
*Figure 5.11 – Noise is increased on the DC voltage.*
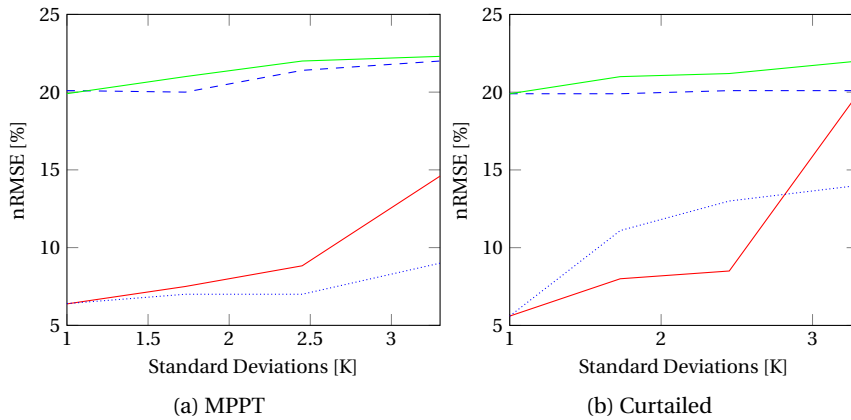


*Figure 5.12 – Noise is increased on the measured cell temperature.*

the proposed estimators can be used to improve the performance of a machine learning-based forecasting method for maximum power prediction. For this purpose, we select a scenario where the PV power is curtailed, thus historical power measurements are not representative of the maximum available power from the plant. As a forecasting tool, we implement an artificial neural network (ANN), a method often advocated in the existing literature in application to data-driven point predictions [2]. The fitting is performed considering 10 neurons and one hidden layer and using as training algorithm the Matlab implementation of the Levenberg-Marquardt back-propagation function, [164].

We consider its prediction performance in two cases. First, when the ANN is trained by using raw production measurements (direct forecast, DF); this means that the training can contain values that are not representative of the maximum available power. Second, the training time series is generated with the proposed estimators and thus the training data always approximate the maximum available power (filtered forecast, FF). This latter case implies the implementation of the analytical, EKF, I&I estimators, and the pyranometer-based

one to reconstruct the maximum power. For the I&I we fix $\gamma$=10. The considered prediction horizon is 5 minutes. The dataset consists of 35 days of experimental data (generated with the setup described in Section 5.6), 30 of which are used for the training phase and 5 for the testing. The training dataset contains 12 days where the PV power output is curtailed according to a random pattern by controlling the active power set-point of the experimental power converter. The same applies for 2 days of the testing dataset. For the remaining period, the power converter is left in MPPT mode. As an example, 3 days extracted from the testing time-series are shown in Fig. 5.13. For the first 2 days the power is curtailed while during the third day the plant works in MPPT.
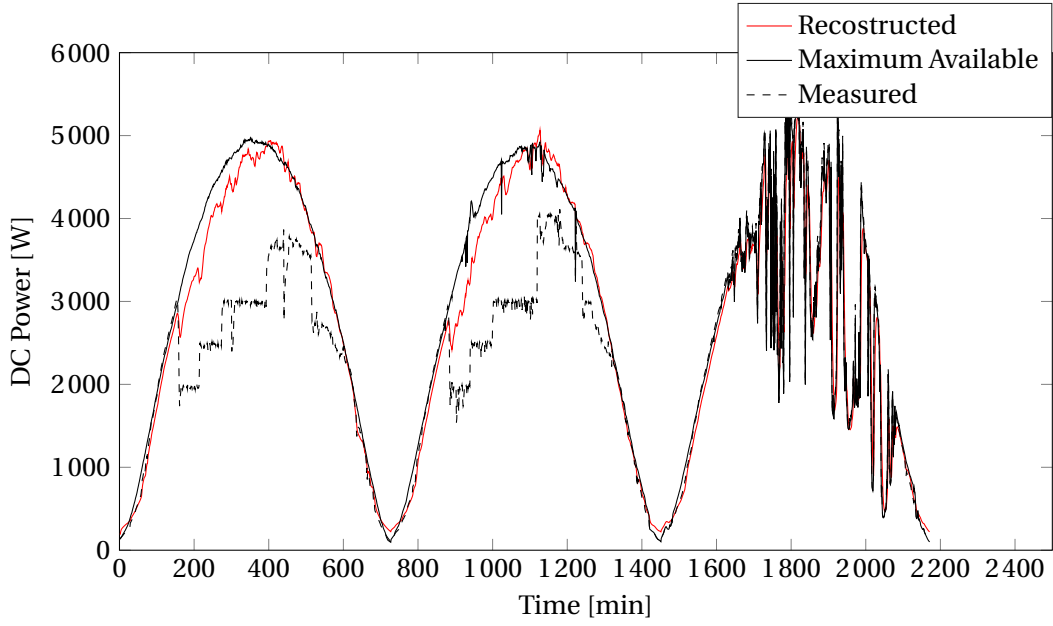


*Figure 5.13 – 3 days are shown from the testing set. The measured power of C1, shown with the dashed line, is curtailed during the first two days. The solid black line represents the maximum available DC power, measured from C2. The red line is the reconstructed maximum power using the I&I estimator.*

The metric used to evaluated point predictions performance is the normalized mean absolute error (nMAE):

$$\text{nMAE} = \frac{1}{M} \sum_{t=1}^{M} (\tilde{P}_{dc,t+1|t} - P_{dc,t+1}) / \overline{P}_{dc} \tag{5.41}$$

where $\tilde{P}_{dc}$ is the one-step-ahead prediction of the DC maximum power, $P_{dc}$ is the realization, $M$ is the number of measurements in the testing dataset and $\overline{P}_{dc}$ their average value. We remind here that the aim of this comparison is not to assess the skills of the forecasting method, rather showing the advantage introduced by the proposed pre-filtering approaches in improving time-series based forecasting tools. Results are summarized in Table 5.6. They show that reconstructing the maximum power with the proposed techniques is always beneficial

when curtailment strategies are adopted: the DF, that simply uses past row measurements of the output power, produces the largest nMAE. Performance of the other methods is in line with what obtained in the estimation comparison in Section 3.9.

*Table 5.6 – Normalized Mean Absolute Error*

|  | nMAE [%] |
|---|---|
| DF | 19.7 |
| FF, Analytical | 9.0 |
| FF, EKF | 14.7 |
| FF, I&I $\gamma = 10$ | 9.3 |
| FF, Pyranometer | 11.4 |

## 5.9   Chapter Conclusions

In this Chapter, we have analyzed three model-based methods to estimate the irradiance received by a PV system from measurements of the system DC current, voltage and cell temperature. The estimators are applied to reconstruct the maximum DC power output of a PV plant independently of the fact it operates in MPPT mode or under curtailment regimes. The estimation performance is evaluated by using measurements from an experimental setup and benchmarked against pyranometer estimations. Results show that:

- the considered estimators can reconstruct successfully the theoretical maximum power output of PV installations even when the plants operate in non-MPPT mode;

- when estimating the peak power of PV systems, the considered approach can outperform pyranometer-based estimations;

- for noise levels compatible with commercial sensors, the analytical estimator showed similar or better performance and bandwidth than the immersion and invariance and Kalman filter-based estimators, with the advantage of being parameter-less. If the available measurements are characterized by a high level of noise (STD higher than $\approx 0.6$ for voltage and current and than $\approx 3$ for the temperature), the use of filter-based strategies is advisable since they are able to delete noisy observations while accounting for the structure of the process. In particular, the I&I method is able to maintain low nRMSE but a preliminary tuning of $\gamma$ is required;

- when considering an artificial neural network-based point predictor the forecasting performance degrades if the training dataset contains operation points in curtailment regimes. Instead, the proposed estimators achieve to remove the components due to exogenous control actions by reconstructing the training dataset.

# 6 Uncertainty Assessment of the Output Power of Grid-Connected PV Plants

## 6.1 Chapter Highlights and Summary

PV power output forecasting is affected by an uncertainty that should be characterized in order to intelligently exploit the flexibility of the available resources in a microgrid. These uncertainties mostly depend on the external weather conditions, on the irradiance-to-power conversion models, and on the prediction methods and forecasting horizons.

In this Chapter, we propose a gray-box method to assess the overall flexibility and associated uncertainties of a small-scale PV plant. The contributions are the following. First, we describe how to accurately predict the one-step-ahead AC maximum available power, integrating a model for the PV modules, a model for the converter, and a probabilistic short-term forecasting tool. The sub-models are selected to be of easy reproducibility, and the methodology needs few inputs: i) information from the PV module datasheet, ii) real-time measurements of the PV plant voltages and currents, and iii) a temperature sensor. The available literature generally focuses on the individual sub-model accuracy, e.g. [165], while we here evaluate the overall chain that allows obtaining the AC maximum power. For different forecast horizons and for different levels of curtailed energy, we identify which sub-model is the most critical in term of uncertainty. The validation is made possible by a dedicated experimental setup that relies on a fully controllable converter, [166], which can track an external power reference of active and reactive power. Second, the uncertainty of the PV converter in tracking an external setpoint is accounted. Indeed, due to the internal characteristic of the converter (e.g. control loop), the implemented power injection can differ from the external request.

## 6.2 Problem Definition

We consider a generic PV system (Fig. 6.1), with the following characteristics (i) it is always connected to the grid as a grid-feeding converter, (ii) it allows to track an external active/reactive power setpoint $u = (P^*, Q^*)$ from a Grid Controller (GC). The PV system has an inherent *flexibility* to modify its power operating conditions, namely the capacity of curtailing the gen-

eration of this resource, along with the possibility of controlling its reactive-power injection.
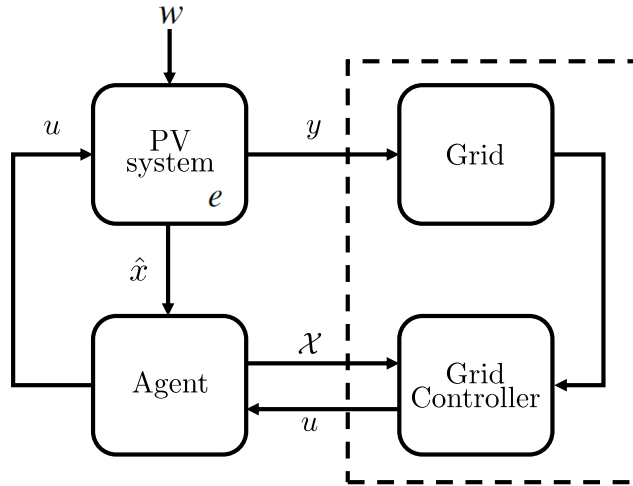


*Figure 6.1 – Problem context. We consider a generic PV plant connected to the grid. The resource is equipped with an agent that can communicate with a generic grid controller. The dashed lines delineate exogeneous elements to the resource.*

In order to allow the GC to exploit this flexibility, a dedicated PV control *agent* is introduced. It continuously receives an updated state $\hat{x}$ from the PV system (e.g., the measured power value) and sends to the GC a feedback $\mathcal{X}$ that represents the expected one-step-ahead operational flexibility and uncertainty. It is important to note that, since the converter can track external power setpoints, the measured AC power does not necessarily correspond to the maximum available one, and may not be representative of the production potential of the plant, that mostly depends on the external weather conditions $w$ (e.g., solar irradiance). Therefore, a physical model of the PV system is needed to reconstruct at each time-step the AC maximum available power, as already discussed in Chapter 5 and detailed in Section 6.4. These reconstructed power values are then used to train a time-series-based forecasting tool needed to predict the one-step-ahead AC maximum power. The uncertainty related to the delivered prediction is explained by both the stochasticity of the forecast and the PV/converter modeling errors. Furthermore, due to the converter internal operational constraints $e$, the external power reference $u = (P^*, Q^*)$ differs from the actual implemented power injection $y = (P, Q)$, this representing a further source of uncertainty. Consequently, a comprehensive uncertainty assessment is needed to account for these aspects: i) uncertainty on the one-step-ahead maximum available power computation, and ii) uncertainty on the setpoint tracking.

## 6.3 Experimental Setup

We anticipate in this Section the description of the experimental setup to introduce information about the observed state variables of the system we target to control. The PV system under test is detailed in Fig. 6.2. The power stage includes a PV-Side Converter (PVSC), that controls the power flow injected by the PV panels into the common DC-bus, and a Grid-Side Converter (GSC) responsible of controlling active and reactive power flows injected into the grid. We assume to have access to the measurements of the DC voltage, DC current, and average cell temperature of the PV array ($v$, $i$, $\theta$), and to the voltages and currents phasors at the fundamental frequency components from the three-phase AC grid ($\bar{V}, \bar{I}$)[1]. The global irradiance on the PV plant, in figure indicated with $G$, could be measured by a sensor (e.g. pyranometer, all-sky camera..), but in this analysis is inferred from the PV-array measurements as proposed in Chapter 5, and detailed in Section 6.4.
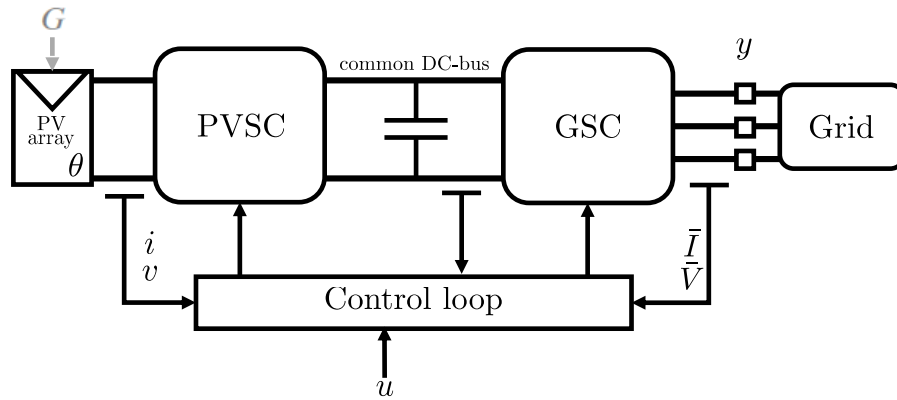


*Figure 6.2 – Detailed view of the PV system.*

## 6.4 Methods

In this Section, we propose a method that, regardless of the PV system operating point, is able to i) provide a probabilistic forecast of the one-step-ahead maximum available power, and ii) determine the uncertainty on tracking a power setpoint. For short-term horizons (from sub-seconds to minutes), the first goal can be achieved by using time-series based forecasting methods. Accordingly to Chapter 5, we first apply a model-based estimator (with parameters $\mathscr{P}_p$) to reconstruct the DC maximum available power, $\tilde{P}_{dc}$, from measurements of the DC voltage $v$, current $i$, and module temperature $\theta$. Then, we use a model of the converter (with parameters $\mathscr{P}_c$) to estimate the AC maximum power, $\tilde{P}$. The procedure is shown in Fig. 6.3. Historical values of $\tilde{P}$ are then used to train a forecasting tool. In what follows we

---

[1]These electrical quantities are commonly available in commercial PV converters, while the PV-array temperature can be easily measured.
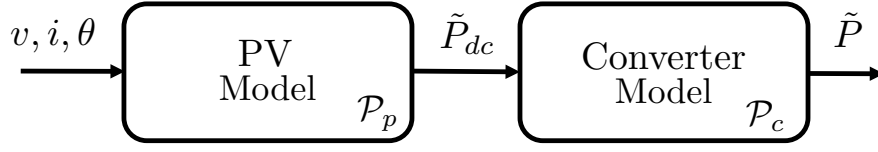
*Figure 6.3 – Maximum Available Power Estimator. The AC maximum power value can be reconstructed using the PV and converter model. The inputs are the measured DC voltage and current (v, i), and the cell temperature $\theta$.*

detail the overall methodology necessary to predict the PV plant power capability and the related uncertainty. The adopted models, used on-line, are presented in Section 6.4.1, while Section 6.4.2 describes how to off-line assess their parameters.

### 6.4.1  Models

The on-line process, depicted in Fig. 6.4, is as follows. At each time-step $t$, the measurements
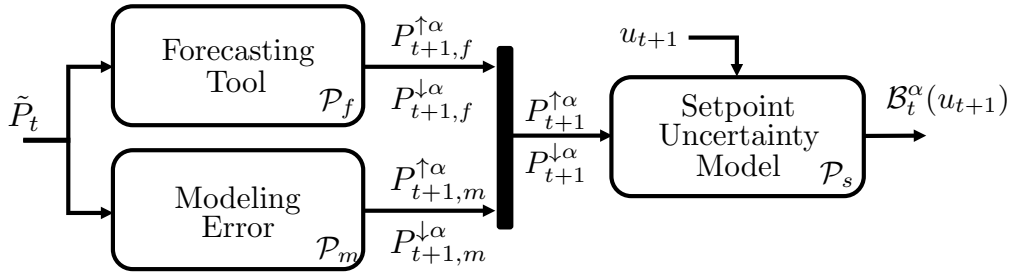


*Figure 6.4 – On-line modelling chain.*

$v_t$, $i_t$, $\theta_t$ are used as inputs to the maximum available power estimator in Fig. 6.3 to get $\tilde{P}_t$. This value is used as input to i) a forecasting tool (with parameters $\mathscr{P}_f$) to deliver the upper and lower bounds of the one-step-ahead prediction interval ($P_{t+1,f}^{\uparrow\alpha}$, $P_{t+1,f}^{\downarrow\alpha}$, with $\alpha$ target confidence level) and ii) a modeling error block (with parameters $\mathscr{P}_m$) to compute the upper and lower bounds accounting for the PV and converter model uncertainty ($P_{t+1,m}^{\uparrow\alpha}$, $P_{t+1,m}^{\downarrow\alpha}$). A conservative approach to compute the global uncertainty associated to the one-step-ahead prediction is to consider the extreme bounds between the model bounds and the forecasting intervals[2]:

$$P_{t+1}^{\uparrow\alpha} = \max(P_{t+1,f}^{\uparrow\alpha}, P_{t+1,m}^{\uparrow\alpha}) \tag{6.1}$$

$$P_{t+1}^{\downarrow\alpha} = \min(P_{t+1,f}^{\downarrow\alpha}, P_{t+1,m}^{\downarrow\alpha}). \tag{6.2}$$

These global bounds ($P_{t+1}^{\uparrow\alpha}$, $P_{t+1}^{\downarrow\alpha}$) are then used as an input for a block that models the ability of the power converter to track a power setpoint (with parameters $\mathscr{P}_s$). This block outputs a set-valued function, $\mathscr{B}_t^{\alpha}(u_{t+1})$, which represents the overall uncertainty of tracking a power

---

[2]A detailed analysis of these uncertainties is reported in Section 6.5.1.

setpoint $u_{t+1}$. Thus, $\mathcal{B}_t^\alpha(u_{t+1})$ include all the uncertainties related to the PV plant future operation. We next describe the different blocks involved in the methodology: forecasting tool, modeling error (PV model and converter model), and setpoint uncertainty model. We discuss the off-line assessment of the models parameters $\mathcal{P}$ in Section 6.4.2.

### PV Model

The PV model used to compute the maximum DC power is as described in Section 5.3. In particular, the analytical method is used to reconstruct in real-time the maximum available DC power. We indicate the parameters of the module model with $\mathcal{P}_p$.

### Converter Model

The converter model allows computing the AC power from the DC power and the DC voltage. Two main models are available in the literature, [59] and [60]. Both require fitting the parameters using datasheet and operational data. In our case, fitting a simple quadratic model led to a better accuracy:

$$\tilde{P}_t = \beta_0 + \beta_1 \tilde{P}_{dc,t} + \beta_2 v_t + \beta_{11} \tilde{P}_{dc,t}^2 + \beta_{12} \tilde{P}_{dc,t} v_t + \beta_{22} v_t^2, \tag{6.3}$$

where $\tilde{P}_t$ is the AC power (response variable), $\tilde{P}_{dc,t}$ the DC power, $v_t$ the DC voltage (the predictors), and $\mathcal{P}_c = \{\beta_0, \beta_1, \beta_2, \beta_{11}, \beta_{12}, \beta_{22}\}$ are the model coefficients.

For our converter, the values of the coefficients from Eq. (6.3) are shown in Table 6.1. Fig. 6.5

*Table 6.1 – Converter Model Coefficients*

| $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_{11}$ | $\beta_{12}$ | $\beta_{22}$ |
|---|---|---|---|---|---|
| -220 | 0.92 | -0.60 | $-2.89 \cdot 10^{-6}$ | $2.06 \cdot 10^{-5}$ | -0.01 |

compares the estimations from the fitting to the measured AC power values for an entire day of measurements, presenting an error of 1.6% (nRMSE).

### Forecasting Tool

To compute short-term prediction intervals (PIs), we adopt the method proposed in Chapter 3 directly applied to the AC power. It consists in clustering historical differentiated maximum power estimations using the k-means clustering algorithm. The clustering, performed off-line, is based on two influential variables: i) the average power and ii) the power variability, calculated on a rolling time window considering the last three AC maximum power estimations. At each time-step, when a new AC maximum power estimation $\tilde{P}$ is available, the influential variables are computed on-line. The Euclidean distance between the centroids and the actual influential variables is then used as the similarity criterion to identify the representative cluster.

117
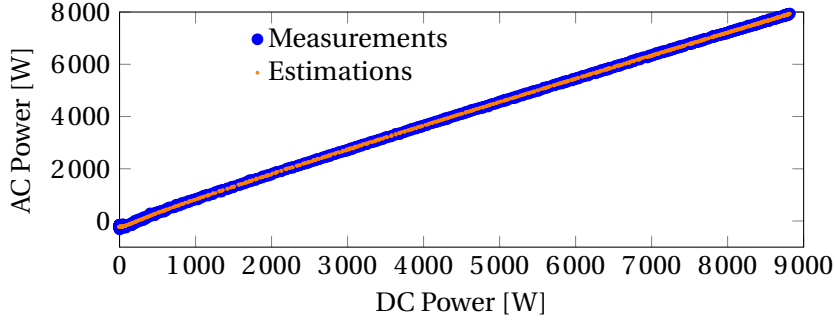
*Figure 6.5 – The AC power, estimated using the converter model described in Section 6.4.1, is compared with the measured AC power.*

Then, we use the upper and lower quantiles, $\mathscr{P}_f = \{q_f^{\uparrow\alpha}, q_f^{\downarrow\alpha}\}$, extracted from the representative cluster, to calculate the prediction bounds, at a given confidence level $\alpha$. In particular, we sum the quantiles to the estimated AC maximum power:

$$P_{t+1,f}^{\uparrow\alpha} = \tilde{P}_t + q_f^{\uparrow\alpha}, \tag{6.4}$$

$$P_{t+1,f}^{\downarrow\alpha} = \tilde{P}_t + q_f^{\downarrow\alpha}. \tag{6.5}$$

**Modeling Error**

The model-based strategy to reconstruct $\tilde{P}$ ensure that the forecasting tool learns from training data that are representative of the maximum power, even when the system is not operating in MPPT mode. However, the obtained prediction interval does not account for the error associated to the PV/converter models, needed in the chain to reconstruct the maximum available power. For this reason, a set of AC measurements, $\{\bar{V}, \bar{I}\}$, see Fig. 6.2, is used to account for the modeling error defined at each time step as:

$$e_t = \tilde{P}_t - P_t(\bar{V}_t, \bar{I}_t), \tag{6.6}$$

where $\tilde{P}_t$ is the AC maximum power estimation and $P_t$ the ground truth value, computed from $(\bar{V}_t, \bar{I}_t)$. This error captures all what is not captured by the forecasting model. We consider the modeling error distribution on a training set, and we extract the upper and lower quantiles of the set, $\mathscr{P}_m = \{q_m^{\uparrow\alpha}, q_m^{\downarrow\alpha}\}$, corresponding to the target confidence level $\alpha$. The uncertainty bounds associated to the model are defined as:

$$P_{t+1,m}^{\uparrow\alpha} = \tilde{P}_t + q_m^{\uparrow\alpha}, \tag{6.7}$$

$$P_{t+1,m}^{\downarrow\alpha} = \tilde{P}_t + q_m^{\downarrow\alpha}. \tag{6.8}$$

**Setpoint Uncertainty Model**

The uncertainty on tracking a power setpoint depends on the internal control law of the converter. In particular, for a given setpoint $u$, we have a set of possible AC active and reactive power injections $(P, Q)$, for which we can compute the distances $dP(u) = P^* - P(u)$ and $dQ(u) = Q^* - Q(u)$. We define $\left(\delta_P^{\uparrow\alpha}(u), \delta_P^{\downarrow\alpha}(u)\right)$ and $\left(\delta_Q^{\uparrow\alpha}(u), \delta_Q^{\downarrow\alpha}(u)\right)$ as the upper and lower quantiles extracted from the $dP(u)$ and $dQ(u)$ distributions, respectively. For a given setpoint $u_t = (P_t^*, Q_t^*)$, the implemented power lays in a *rectangular* set with a certain confidence level $\alpha$, i.e. $y_t \in \mathcal{R}^\alpha(u_t)$ where:

$$
\begin{aligned}
\mathcal{R}^\alpha(u_t) = \{ (P, Q) \in \mathbb{R}^2 \,| \\
\delta_P^{\downarrow\alpha}(u_t) \le P_t^* - P \le \delta_P^{\uparrow\alpha}(u_t), \\
\delta_Q^{\downarrow\alpha}(u_t) \le Q_t^* - Q \le \delta_Q^{\uparrow\alpha}(u_t) \},
\end{aligned}
\tag{6.9}
$$

$\mathcal{P}_s = \{\delta_P^{\downarrow\alpha}, \delta_P^{\uparrow\alpha}, \delta_Q^{\downarrow\alpha}, \delta_Q^{\uparrow\alpha}\}$ contains all the extracted quantiles computed off-line for different values of $u$.

The sets of parameters $\mathcal{P}_p, \mathcal{P}_c, \mathcal{P}_f, \mathcal{P}_m, \mathcal{P}_s$, are computed during an off-line phase, as described in the following.

## 6.4.2 Models Parameters: Off-line Assessment

The parameters of the module model $\mathcal{P}_p$ can be computed off-line using the procedure in [158] that only needs the module datasheet as input information. The converter model coefficients defining $\mathcal{P}_c$ are obtained by fitting a regression model on a set of data obtained by operating the converter from zero to the maximum rated power. To train the forecasting tool (and compute $\mathcal{P}_f$) we consider a set of $N$ historical measurements $\{v, i, \theta\}$ and use the maximum available power estimator, in Section 5.3, to get $\tilde{P}$ for each set of measurements. The obtained $N$ historical estimations of $\tilde{P}$ are used to build the clusters of the selected probabilistic forecasting tool, see Section 6.4.1. The clustered data are used to extract the quantiles that define $\mathcal{P}_f$. The quantiles of the modeling error distribution are used to define $\mathcal{P}_m$. In practice, for period of length $M$, we let the PV converter work in MPPT, and thus the maximum available power can be directly measured. The training set used to define the error should cover different irradiance and temperature conditions, for example by selecting few entire days characterized by different weather conditions. Finally, the setpoint uncertainty is assessed by requesting specific power setpoints to the converter, by sweeping the whole $PQ$ capability set. We then extract the quantiles from the distributions of the observed distances $\left(dP(u), dQ(u)\right)$, for all the requested setpoints $u$. This allows defining $\mathcal{P}_s$.

## 6.5   Results

The analysis refers to a 13 kWp rooftop PV installation of 51 polycrystalline modules.  The *agent* (Fig. 6.1) is implemented in a NI CompactRIO 9068 and communicates with the converter through a CAN bus interface. Currents, voltage, and temperature measurements are acquired as described in Section 5.6. Since our analysis includes curtailment periods, a second equivalent PV system is used as a reference installation to obtain the maximum PV power output when the test plant is operated in non-MPPT mode.[3]. In these conditions, the average absolute difference between the delivered AC active power values is less than 100 W. Therefore we can assume that the behaviours of the two systems are comparable.

### 6.5.1   Maximum Available Power Uncertainty Assessment

In this Section, we quantify the prediction uncertainty of the AC maximum power considering different forecast horizons (100 ms, 1 s, 1 min and 5 min), and different levels of curtailed energy (0%, 2% and 30%). The latter is computed as the percentage of curtailed energy over the maximum available in the testing period. The metrics are adapted from Section 2.3.1. The first metric is the bound Coverage Probability (CP) which counts the number of times the realization falls inside the PI for a given confidence level $\alpha$:

$$\text{CP} = \frac{1}{L} \sum_{t=1}^{L} c_t, \qquad\qquad c_t = \begin{cases} 1, & P_{t+1|t}^{\downarrow \alpha} \leq P_{t+1} \leq P_{t+1|t}^{\uparrow \alpha} \\ 0, & \text{otherwise.} \end{cases} \qquad (6.10)$$

where $L$ is the total number of testing instances.

The second one is the interval normalized Averaged Width (AW):

$$\text{AW} = \frac{1}{L P_{max}} \sum_{t=1}^{L} (P_{t+1|t}^{\uparrow \alpha} - P_{t+1|t}^{\downarrow \alpha}) \qquad (6.11)$$

where $P_{max} = 13$ kW is the rated power of the converter. Bounds are considered accurate when CP $\geq \alpha$ and the AW is low. The bounds are computed: i) as in Eqs. (6.1)-(6.2), i.e., considering the *worst case* between modeling and forecast uncertainty (*Global*), ii) considering only the forecast uncertainty (*Forecast*), and iii) considering only the modeling uncertainty (*Model*). Results are summarized in Table 6.2 and Table 6.3 for a target confidence level of 95%. In particular, we consider a training set of 4 days for the 100 ms and 1 s horizons, and of 10 days for the 1 and 5 min horizons. The testing set is of 2 days. The number of clusters is 100, 25, 15, and 5 for 100 ms, 1 s, 1 min, and 5 min, respectively. In our analysis, the same training set is used to extract the quantiles for the modeling error.  From Table 6.2 we observe that only the global prediction chain is able to deliver bounds with high coverage (CP is similar to the target

---

[3]This reference PV system is needed only to assess the performance of the proposed method. The considered models are based on measurements from the tested PV system and, as later explained in the text, the presence of a reference system is not necessary to the final user.

*Table 6.2 – Coverage Probability-CP [%], for α = 95%. Results are shown from 100 ms to 5 min forecast horizon. Underlines refer to global CP lower than α.*

| | Global Uncertainty | | |
|---|---|---|---|
| **Forecast Horizon** | **MPPT** | **2% Curt.** | **30% Curt.** |
| 100 ms | 98.1 | <u>93.2</u> | <u>86.0</u> |
| 1 s | 97.6 | <u>93.0</u> | <u>86.1</u> |
| 1 min | 98.0-8.7 | 97.0 | 96.7 |
| 5 min | 98.0-15 | 95.7 | 96.5 |

| Forecast Uncertainty | | | Model Uncertainty | | |
|---|---|---|---|---|---|
| **MPPT** | **2% Curt.** | **30% Curt.** | **MPPT** | **2% Curt.** | **30% Curt.** |
| 22.7-0.27 | 31 | 27 | 97.4 | 89 | 84 |
| 28.8-0.3 | 32.0 | 28 | 96.8 | 87 | 83 |
| 84.8 | 76.9 | 81.0 | 79.9 | 69.8 | 53 |
| 92.6 | 92.0 | 94.0 | 58 | 40.6 | 38.0 |

confidence level). At short-term horizons (below the minute scale), the forecast uncertainty is low (Table 6.3), and accounting for the model uncertainty is necessary to guarantee a decent coverage (higher than 85%). At horizons above the minute scale, the forecast uncertainty is instead dominant. As example, Fig. 6.6 shows forecast and model bounds for 1 s and 1 min forecast horizons, see Fig. 6.6a and 6.6b, respectively. It is possible to see that at 1 s horizon the model bounds are of bigger magnitude and are necessary to guarantee the target coverage probability. On the contrary, above 1 min horizon the forecast uncertainty is higher, and it has the highest contribution to the global coverage. From Table 6.2 we can see that the global CP is lower than $\alpha$ only at low forecast horizon (second and sub-second) with curtailment, highlighted with underlines. This is due to the fact that the selected maximum power estimator (based on measurements of the PV system) is very sensitive to the operating conditions, and a higher modeling error is expected when the system is far from MPPT conditions. However, in common plants, the modeling error can be computed only with respect to a ground truth value obtained by leaving the system in MPPT and comparing the measured and estimated power. Thus, the error defined in Eq. (6.6) only accounts for MPPT and not for curtailed conditions. This explains the lower coverage of the bounds when curtailment actions are adopted. Furthermore, this higher modeling error affects the global CP only at low horizons (lower than the minute scale), i.e. when accounting for the modeling error is more relevant.

To understand the higher error associated with the DC maximum power estimation when far from MPPT, we perform a sensitivity analysis with respect to the model's parameters. In particular, we change one-factor-at-a-time (OAT approach), [167]. Fig. 6.7 shows the sensitivity

*Table 6.3 – Normalized Average Width-AW [%], for α = 95%. Results are shown from 100 ms to 5 min forecast horizon.*

| | Global Uncertainty | | |
|---|---|---|---|
| **Forecast Horizon** | **MPPT** | **2% Curt.** | **30% Curt.** |
| 100 ms | 2.28 | 3.4 | 2.6 |
| 1 s | 2.1 | 3.2 | 2.5 |
| 1 min | 8.7 | 14.9 | 12 |
| 5 min | 15 | 25.4 | 20 |

| Forecast Uncertainty | | | Model Uncertainty | | |
|---|---|---|---|---|---|
| **MPPT** | **2% Curt.** | **30% Curt.** | **MPPT** | **2% Curt.** | **30% Curt.** |
| 0.27 | 0.2 | 0.78 | 2.15 | 2.2 | 2.2 |
| 0.3 | 0.3 | 0.75 | 2.15 | 2.2 | 2.2 |
| 6.7 | 11.9 | 9.4 | 2.15 | 2.2 | 2.2 |
| 15 | 24.1 | 20 | 2.15 | 2.2 | 2.2 |

of the DC maximum power estimation ($\tilde{P}_{dc}$) with respect to the parameters $i_D$, $n_r$, and $R_s$. We recall from Section 5.3 that, these parameters, together with $R_p$ and $I_p$, are computed from datasheet parameters at STC and then updated to account for their dependency of the temperature and irradiance. The sensitivity is done considering two values of irradiance (500 and 800 W/m$^2$) and constant temperature. For this assessment we consider that the initial operating point of the PV system can be MPPT, open circuit (OC), and curtailed (CUR, for this case we select a point halfway between MPPT and OC). We observe that when operating far from MPPT, the estimation is very sensitive to the values of $i_D$ and $n_r$. For example, an error on the estimation of $n_r$ of 10% can lead to an 80% error on the estimation of the DC maximum power when the system operates close to OC conditions. The value of the series resistance $R_s$ similarly affects the MPPT and curtailed conditions, and less affects OC conditions.

We can conclude that when referring to forecast horizons below 1 min, where the model uncertainty prevails in the definition of the global power prediction uncertainty, it is particularly important to invest in an accurate model of the PV model and precise equations that account for the dependency of the five parameters on the weather conditions.

### 6.5.2 Setpoint Tracking Uncertainty Assessment

In this Section we quantify the uncertainty produced by the power converter to track external setpoints. We recall from Sections 6.4.1 and 6.4.2 that for a given request $u = (P^*, Q^*)$, we have a set of possible injections $(P, Q)$, and a corresponding set of possible distances $dP(u) =$
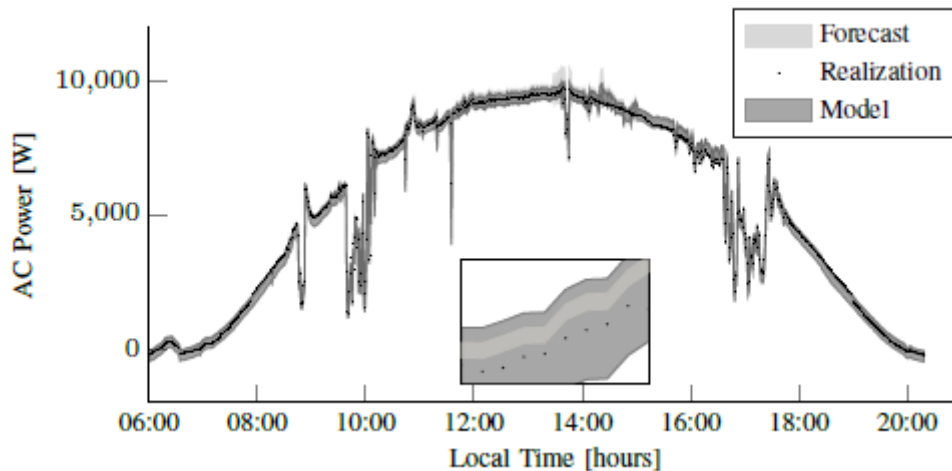
(a) Prediction Uncertainty, 1 s forecast horizon
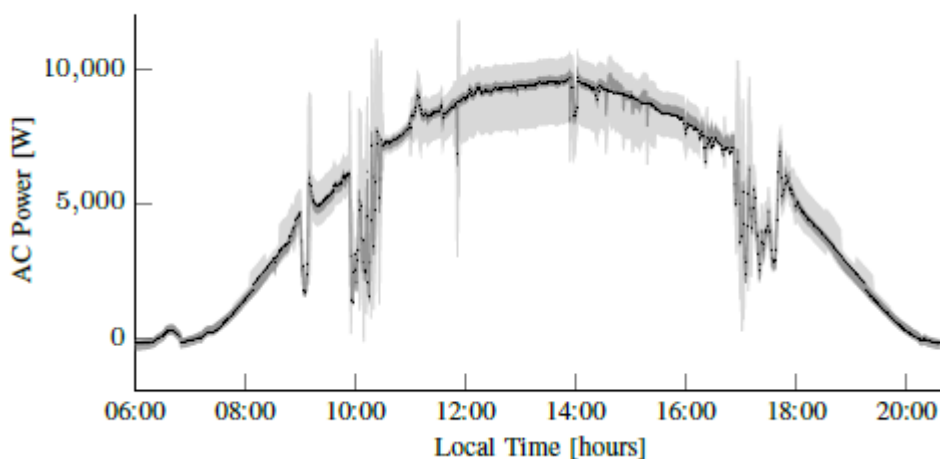


(b) Prediction Uncertainty, 1 min forecast horizon

*Figure 6.6 – Comparison of the forecast bounds, model bounds, and realizations. One day of active power generation is shown for the considered 13 kWp plant.*

$P^* - P(u)$ and $dQ(u) = Q^* - Q(u)$. In order to assess the uncertainty associated with the setpoint implementation, we perform several tests by sending to the converter different requests $u$, sweeping the whole $PQ$ admissible set, with a step size of 100 W/100 var, for the active and reactive power respectively[4]. The quantiles defining $\mathscr{P}_s$ are extracted from the $dP(u)$ and $dQ(u)$ distributions recorded during these off-line tests. As example, Fig. 6.8 shows the quantiles $(\delta_P^\downarrow, \delta_P^\uparrow)$ for a confidence level $\alpha = 95\%$ and for different couples $(P^*, Q^*)$. Similar results are obtained for $Q$. The quantiles can be stored in a 2D look-up table that returns in

---

[4]Each test lasts until the maximum available power is reached, accordingly to given irradiance and temperature conditions.

(a) Sensitivity on $i_D$

(b) Sensitivity on $n_r$

(c) Sensitivity on $R_s$
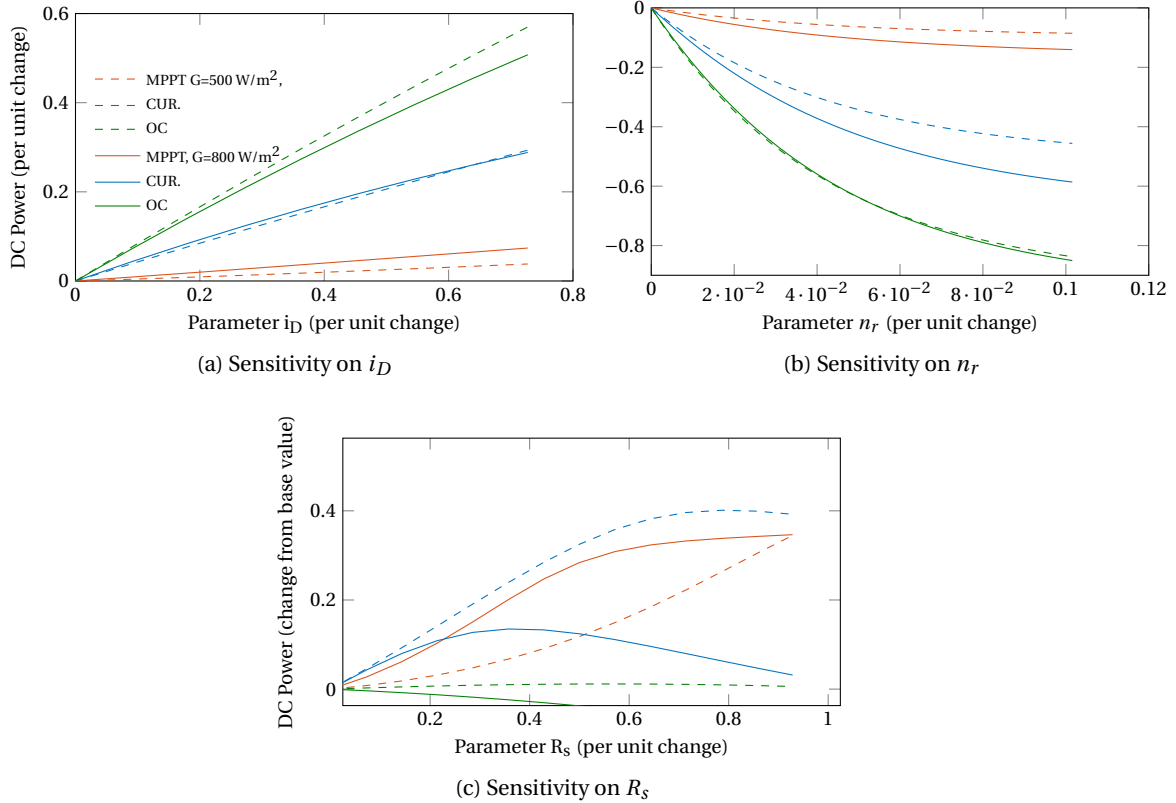
*Figure 6.7 – Per unit change of the DC maximum power estimation with respect to the per unit change of the input parameter. The base unit quantities of the DC maximum power and of the parameters are those originally obtained from the model in [3].*

real-time the quantiles $(\delta_P^\uparrow, \delta_P^\downarrow)$ and $(\delta_Q^\uparrow, \delta_Q^\downarrow)$ as a function of a given request $(P^*, Q^*)$. Fig. 6.8 shows that this uncertainty can reach magnitudes of 600 W that corresponds to $\approx 5\%$ of the plant rated power.

## 6.6 Chapter Conclusions

This Chapter tackles the problem of assessing the uncertainty associated with the operation of a controllable PV facility. The analysis includes two fundamental aspects: determining the PV generation potential for the next time interval (allowing identifying a region inside the capability curve of the power converter from where it is possible to pick a power setpoint), and characterizing the uncertainty associated to tracking a setpoint inside that region. The former element is determined by using a short-term forecasting method, whereas the latter depends on the converter internal dynamics and is evaluated with a data-driven approach. The proposed grey-box modelling relies on physical models as well as on measurements of the PV currents, voltages, and the cell temperature to compute the AC maximum power. The proposed method was tested considering different forecasting horizons, from 100 ms to 5 min
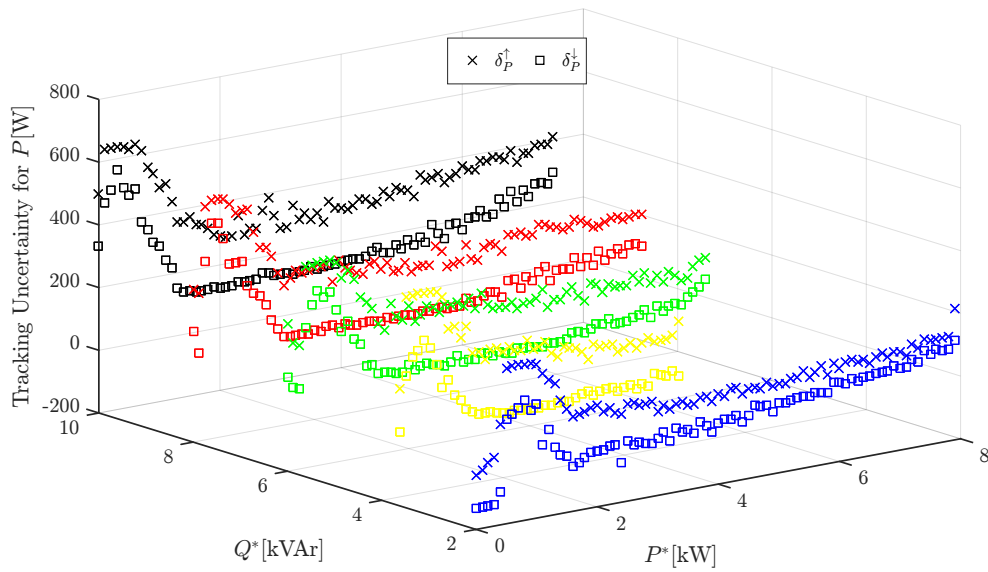
*Figure 6.8 – Setpoints Tracking Uncertainty on Active Power.*

ahead. Experimental results showed that, when considering forecasting horizons below 1 min, accurate physical models are the key to achieve reliable coverage of the prediction interval. On the other hand, for forecasting horizons larger than 1 min, the quality of the predictions is mostly affected by the uncertainty associated with the forecasting, which becomes dominant with respect to models performance.

Chapter 7 shows how these results can be used in a real-time control framework for microgrids applications to advertise the flexibility and uncertainty of a generic PV plant.

# 7 PV Modelling and Short-term Forecast: Applications

## 7.1 Chapter Highlights and Summary

In this Chapter, we show how direct and indirect prediction methods of PV generation can be used in real power system applications that include generation from photovoltaic plants.

In the first part, we describe a control framework, called *Commelec* (Composable Method for Real-Time Control of Active Distribution Networks with Explicit Power Setpoints), specifically designed in our laboratory at EPFL for the real-time control of microgrids. We focus on how PV systems are accounted is such a framework, and on how the developed probabilistic prediction tools can be used to provide information on the future power availability and associated uncertainty. In particular, we show how direct methods can be adopted for the case of an uncontrollable PV plant while indirect methods for a controllable one. In the second part, we focus on an energy management problem, called *dispatchable feeder*. We discuss how PV plants can be used as flexible resources in coordination with a battery storage system to track a specific dispatch plan set the day before operation.

## 7.2 The Commelec Framework

We refer here to a new control approach called Commelec for microgrids control, see [4, 37]. This framework specifically targets the real-time control of active distribution networks, and it is based on the idea that any power grid can be steered via explicit setpoints of active and reactive power of the available resources. A schematic describing the Commelec process is illustrated in Fig. 7.1. Software Agents (deployed in dedicated microcontrollers) are responsible for resources and subsystems (such as storage systems, PVs, loads etc.). A Grid Agent (GA, also called *leader agent*) is responsible for the control of a portion of a distribution network. The GA receives advertisements from each Resource Agent (RA, also called *follower agent*). An advertisement specifies the follower capabilities, expected behavior, and a simplified view of its internal state. The advertisements are device independent and use a common abstract framework; this is essential to ensure scalability and composability. It is important to notice

that the GA can itself be a follower of a higher level agent ($GA_0$) from which it can receive a requested power setpoint. The GA has to compute the requested setpoints to send to the followers and to aggregate information to send to its own leader. As explained in the next paragraph, the GA decision process has the goal of *steering* the electrical state of the grid such as to minimize the virtual costs of the RAs, keep the grid in a safe and feasible state of operation, and satisfy the target setpoint coming from the higher level GA.
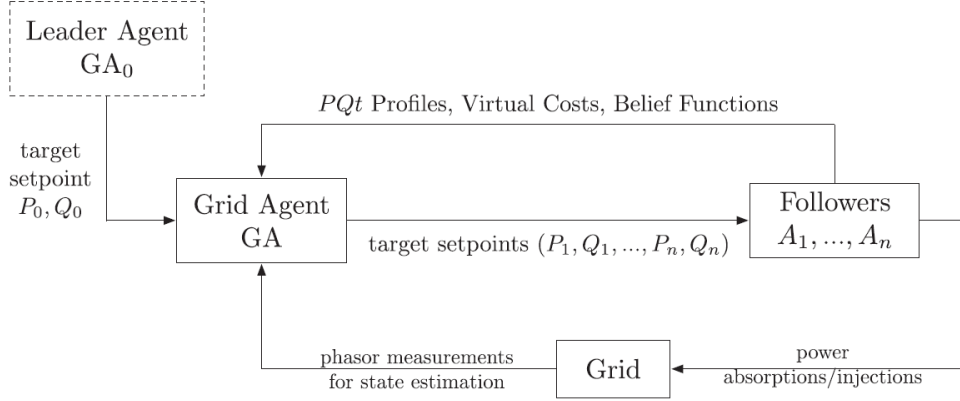


*Figure 7.1 – Schematic of the Commelec Structure, as given in [4].*

**Grid Agent**

The task of the GA is to minimize the following objective function:

$$F(U) = \sum_{i=1}^{n} \omega_i C_i(P_i, Q_i) + \omega_0 C_0\left(P_0(U), Q_0(U)\right) + J(U), \qquad (7.1)$$

where $U = (P_1, Q_1 ... P_n, Q_n)$ is the set of power setpoints to send to the $n$ followers. The function to minimize is the weighted sum of the normalized virtual costs $C_i$ of the resources, a penalty term $C_0$ on the power flow $(P_0, Q_0)$ at the point of connection with the upper-level grid, and a penalty term $J$ on the nodal voltages and line currents. The solution has to return a feasible solution, namely:

- it respects voltage and current limitations:

$$V_k \in [V_k^{nom} - \beta_k, V_k^{nom} + \beta_k], \quad I_l \leq I_l^{max} \qquad (7.2)$$

  where $\beta_k$ and $I_l^{max}$ are given threshold variables that depend on the grid, index $k$ indicates the nodes and $l$ the lines.

- The power injection at the slack is within a given region $R_0$.

All the terms of Eq. 7.1 are designed convex and, in order to solve the minimization at each time step, the single iteration of the gradient descent method is used.

**Resource Agents**

The main task of a RA (e.g., the PV Agent) is to translate its internal state into the abstract framework which consists of three main elements: *PQ profile, Virtual Cost* (VC), and *Belief Function* (BF):

- The PQ profile represents the region (of active and reactive power) where the subsystem can inject or absorb power.

- The VC quantifies the propensity of a system to stay in a particular zone of the PQ profile.

- The BF accounts for the uncertainty of the subsystem operation: it returns the set of all possible setpoints that the system might implement due to its own stochasticity.

The PQ profiles sent by the followers are assumed to be deterministic sets in the PQ plane, and so the GA performs a deterministic optimization under constraints defined by the belief function that accounts for the uncertainty related to the resources. It is worth noting that the real-time assessment of the BF, able to account for uncertainties of some resources (e.g. PV), is not a trivial aspect. For example, in the case of stochastic generation we need to provide high resolution prediction intervals (PIs) in few milliseconds. Furthermore, the RAs communicate to the GAs active/reactive power values that are measured at the point of connection with the grid; for this reason DC/AC converters are always considered as part of the resource. RAs are usually deployed in a microcontroller, specific for each resource. A general schematic of the process running in the microcontroller is shown in Fig. 7.2. The microcontroller communicates with the resource converter, the measurement board, and the GA. The requested setpoint coming from the GA is projected into the actual PQ state of the resource and then implemented by the converter. Voltages and currents values are acquired by a measurements system, processed to compute the AC active and reactive power and the DC power, logged and used by the RA. Besides projecting the request, the RA computes the advertisement. The advertisement (PQ, BF, CF) is then sent to the GA.

To summarize, a RA implements four consecutive tasks:

- Upon receiving a request setpoint, it retrieves the resource state and computes the current PQ profile. This set may differ from the one previously sent due to delays or inaccuracies.

- It performs an Euclidean projection of the requested setpoint onto the current PQ profile. The projected value is the command to send to the converter.

- The agent sends the setpoint to the actuator and waits until the setpoint is implemented. The corresponding delay depends on the resource's nature (e.g. on ramping constraints and controller parameters).

- After this delay, the RA computes the advertisement to be sent to the GA.
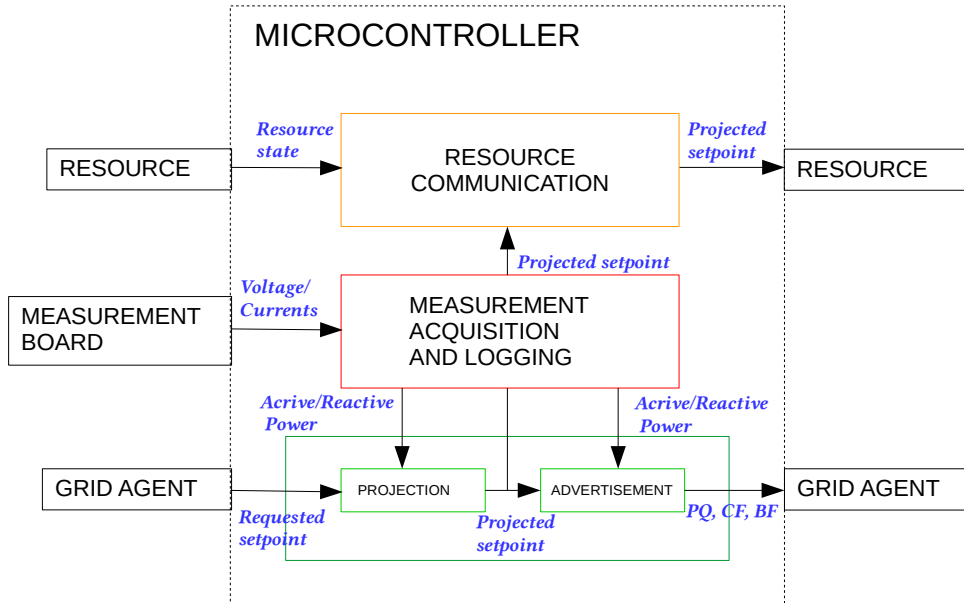
*Figure 7.2 – Structure of the software deployed in a microcontroller for each RA.*

A general overview on how to design the different RA (e.g. battery, supercapacitor, loads) can be found in [37, 5]. In the following sections we focus on how to design an uncontrollable and controllable PV agent, i.e. on how to advertise the behaviour of a photovoltaic plant that is part of the grid when the injected power cannot/can be controlled, respectively.

### 7.2.1 Uncontrollable PV Agent

In this paragraph, we consider the case of a PV plant that is not controllable. The PV converter works in maximum power point tracking (MPPT) conditions and thus we can assume that the measured power corresponds to the maximum available. This means that the uncontrollable PV agent (UPVA) does not provide any flexibility and is, for the Commelec control framework, a source of uncertainty. Fig. 7.3 shows the PQ profile and the BF of the uncontrollable PV agent (UPVA). The PQ profile is computed as the point forecast for the next time step while the BF is the associated prediction interval (PI, dashed lines). Since no control action is performed on the PV plant, we can directly apply time-series based direct power forecasting tool, as the one proposed in Chapter 2, to deliver the point forecast and the PI for the active power. In general, for grid connected PV plants, the converter tracks a power factor of one and therefore the uncertainty on the reactive power is simply a noise around the zero value. Since the plant is not controllable, we consider a zero cost function.
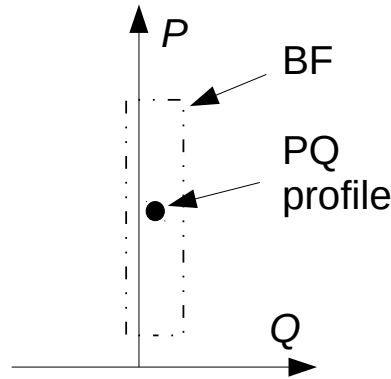
*Figure 7.3 – Representation of the advertisement sent by an UPVA.*

### 7.2.2  Controllable PV Agent

In this paragraph, we consider a PV plant that is controllable, namely we can ask for specific active/reactive power setpoints. We show here how it is possible to design an agent for a PV system that is meant to periodically inform the GA about the PV plant capabilities, independently of its actual operating point. These capabilities may continuously change as a function of the external perturbation and the converter limitations. Accordingly to the Commelec nomenclature and for the sake of clarity, we name here the three elements that represent the way the PV system can be controlled as (i) the PQ profile $\mathscr{A}$, (ii) the belief function $\mathscr{B}$, and (iii) the virtual cost $\mathscr{C}$. Therefore, the information sent to the GA can be defined as: $\mathscr{X} = (\mathscr{A}, \mathscr{B}, \mathscr{C})$. Note that, the PV Agent looks for expressing the controllability of the PV system at a given time-period in the future, i.e., at the time when it expects to receive a new power setpoint. We assume that, if the GA sends setpoints fast enough, the information generated by the PV Agent can be computed in discrete-time, considering that it remains constant between two consecutive time-steps. Therefore, we target to have a message update in the sub-second scale so that we can cope with the fastest dynamics of real PV systems. This serves to simplify the information sent to the GA without the need of explicitly including the time dependence of the external perturbations between consecutive time-steps. We next describe how to compute $\mathscr{X}$.

**PQ profile $\mathscr{A}$**

In general, we consider that at time $t$, the one-step-ahead power flexibility of a controllable PV system can be defined as a function of (i) the predicted maximum available power $P_{t+1}^{\uparrow \alpha}$, (ii) the converter rated power $S_r$, and (iii) a minimum power factor constraint $\text{PF}_{\min}$. This last is typically required to minimize the reactive power flows in the grid and to comply with quality-of-service norms, generally associated to voltage control. The maximum available power is defined by the upper bound of the global prediction bound, $P_{t+1}^{\uparrow \alpha}$. Since in this case the PV plant is controllable, there is no guarantee that the measured power is the maximum

131

available one and therefore direct forecasting methods are not advisable. On the contrary, the maximum power should be computed from the solar irradiance and cell temperature, and the procedure described in Chapter 6 used to deliver the predicted maximum power $P_{t+1}^{\uparrow\alpha}$.

Considering this, and $S = P + jQ$ as the complex power, we define $\mathscr{A}$ as the set in the $PQ$-plane where a GA can request a setpoint $u$ in the next time-step, i.e., $u_{t+1} \in \mathscr{A}_t$. This is,

$$
\begin{aligned}
\mathscr{A}_t = \{(P,Q) \in \mathbb{R}^2 \mid \\
0 \le P \le P_{t+1}^{\uparrow\alpha}, \frac{|P|}{||S||} \ge \mathrm{PF}_{\min}, ||S|| \le S_r\}
\end{aligned}
\tag{7.3}
$$

The graphical representation of $\mathscr{A}$ is shown in Fig. 7.4.
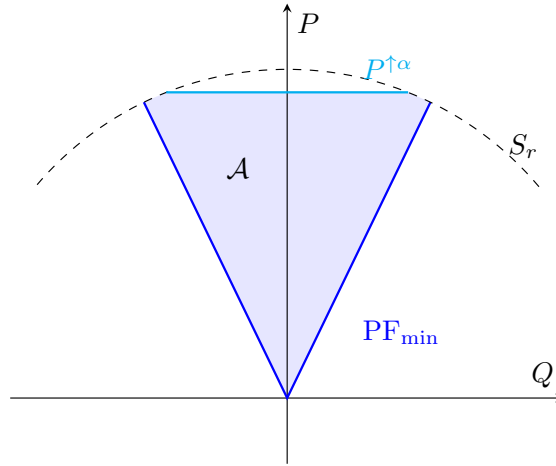


*Figure 7.4 – PV system Power Flexibility.*

**Belief Function $\mathscr{B}$**

The uncertainty of implementing a given setpoint is strongly dependent on two factors (i) the maximum available AC power (from forecast and model) and (ii) the setpoint tracking accuracy. Let us first focus on the former. The set defined by the uncertainty of the maximum available power, is a segment that depends on $\mathrm{PF}_{\min}$. The definition for any $u_{t+1} = (P_{t+1}^*, Q_{t+1}^*) \in \mathscr{A}_t$ is in Eq. (7.4).

$$
\mathscr{B}_{1,t}^{\alpha}(u_{t+1}) = \begin{cases} \{(P_{t+1}^*, Q_{t+1}^*)\} & 0 \le P_{t+1}^* \le P_{t+1}^{\downarrow\alpha}, \\ [(P_{t+1}^*, Q_{t+1}^*), (P_{t+1}', Q_{t+1}^*)] \cup [(P_{t+1}', Q_{t+1}^*), (P_{t+1}^{\downarrow\alpha}, Q_{t+1}')] & P_{t+1}^{\downarrow\alpha} \le P_{t+1}^* \le P_{t+1}^{\uparrow\alpha}, \end{cases}
\tag{7.4}
$$

where the global bounds $P_{t+1}^{\uparrow\alpha}$ and $P_{t+1}^{\downarrow\alpha}$ are as discussed in Chapter 6.

$P'_{t+1}$ and $Q'_{t+1}$ are expressed as:

$$P'_{t+1} = Q^*_{t+1} \tan(\arccos(\text{PF}_{\min})), \tag{7.5}$$

$$Q'_{t+1} = \text{sign}(Q^*_{t+1}) P^{\downarrow\alpha}_{t+1} \frac{\sqrt{1 - \text{PF}^2_{\min}}}{\text{PF}_{\min}}. \tag{7.6}$$

The graphical representation is shown in Fig. 7.5a.

The uncertainty of the power converter to deploy a power setpoint is defined by a *rectangular* set with a certain confidence level $\alpha$ as described in Eq. (6.4.1). The set defined by this uncertainty for any $(P^*_{t+1}, Q^*_{t+1}) \in \mathscr{A}_t$, shown in Fig. 7.5b, is thus:

$$\mathscr{B}^\alpha_{2,t}(u_{t+1}) = \mathscr{R}^\alpha(u_{t+1}). \tag{7.7}$$

The overall uncertainty on deploying a power setpoint $u_{t+1}$ can be, therefore, written as follows (see Fig. 7.5c):

$$\mathscr{B}^\alpha_t(u_{t+1}) = \bigcup_{(p,q) \in \mathscr{B}^\alpha_{1,t}(u_{t+1})} \mathscr{B}^\alpha_{2,t}(u_{t+1}). \tag{7.8}$$

**Virtual cost $\mathscr{C}$**

We assume that, a PV system operator would always look for maximizing the active power production and minimizing the reactive power magnitude. Thus, the virtual cost can be defined as a function of all power setpoint $(P, Q) \in \mathscr{A}$ as

$$\mathscr{C}(P, Q) = -P + Q^2. \tag{7.9}$$

Graphically, the virtual cost, defined in the domain of the $PQ$ profile, is shown in Fig. 7.6 where the red hue refers to higher cost.

### 7.2.3 Experimental Validation

The microgrid available at the DESL has been designed as a realistic scale of the CIGRE low voltage (400 V at 50 Hz) microgrid benchmark defined in [168]. We here describe its architecture as shown in Fig. 7.7, however the detailed description can be found in [5].

The currently available resources, framed with a dashed line in Fig. 7.7, are:

- Three controllable single-phase power converters to emulate the consumption of a 24 kW residential building that has electrical space-heating. In particular, any active/re-

(a) Maximum available power

(b) Setpoint tracking accuracy

(c) Overall uncertainty
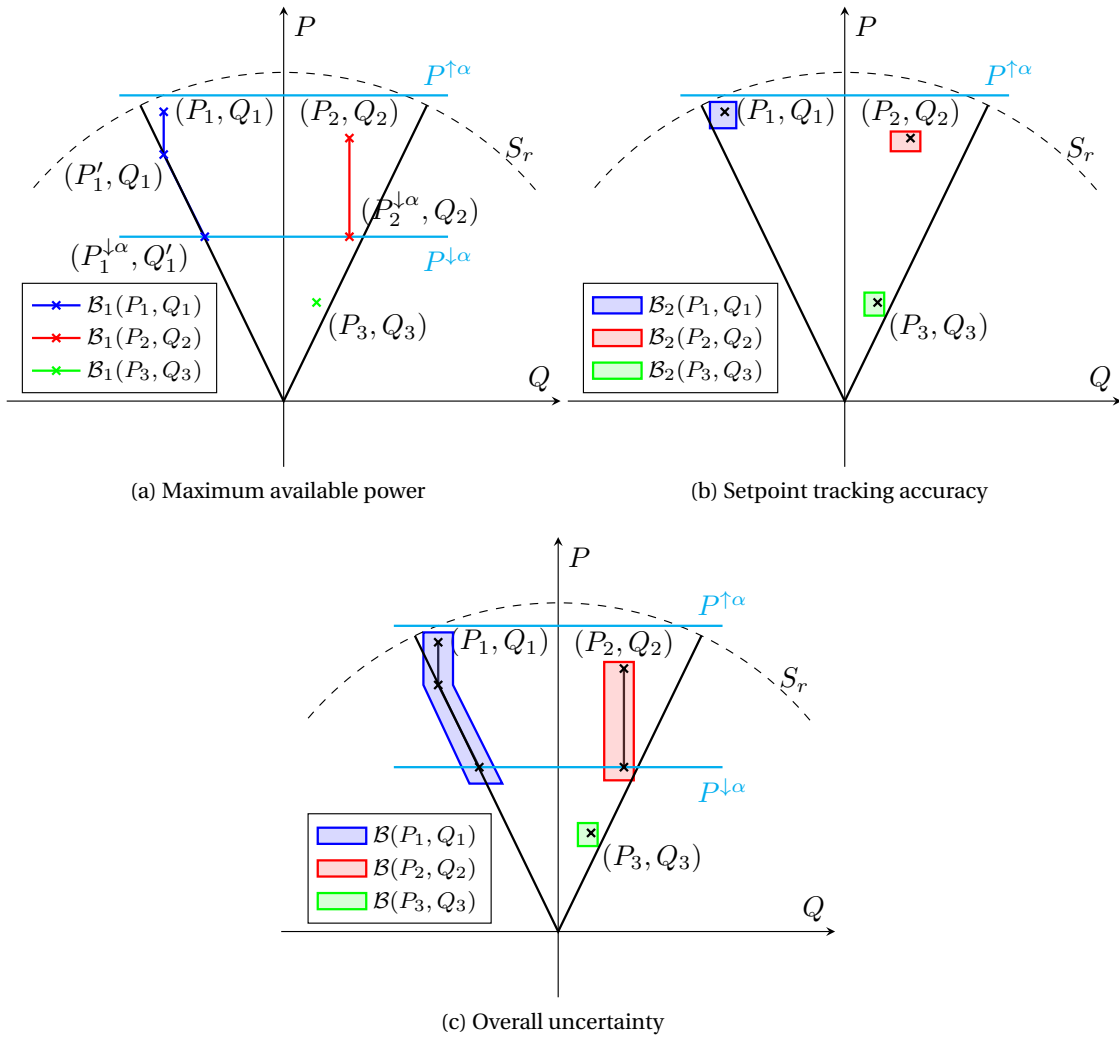
*Figure 7.5 – Examples of Belief Functions.  The subscripts represent three different possible setpoints.*

active power setpoint is translated into current amplitude and phase for the 3 phases. The load is indicated in Fig. 7.7 with L1.

- A battery energy storage system with a 25 kW/25 kWh power/energy rating, indicated in Fig. 7.7 with B. It is based on Lithium Titanate cells and is monitored by a local battery management system (BMS). The battery uses a 4-quadrants converter, which directly receives power setpoints.

- A 13 kW PV-roof installation (PV1), controllable using a a 4-quadrants converter as described in [169].

- A supercapacitor (SC) bank composed of 6 modules connected in series. It is a 50 kW - 0.8 kWh based a 2.7 V and 3000 F cell. Also this system used a 4-quadrants converter.
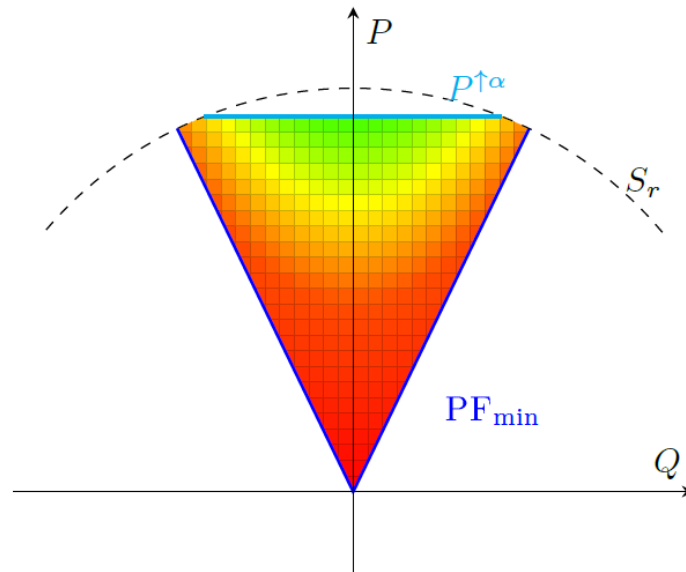
*Figure 7.6 – Cost Function*

- An hydrogen system composed of a fuel cell (15 kW) and a high-pressure (30 bar) electrolyzer (6 kW) based on a proton-exchange membrane (PEM) technology. The fuel cell and the electrolyzer are indicated with FC and EL, respectively.

- A 20 kW PV-roof installation (PV2) which is not controllable and always tracks the maximum power and a 7 kW PV façade installation (PV3) also not controllable.

Every resource in the microgrid is equipped with a resource agent (RA) deployed in a NI CRIO 9068 as described in Fig. 7.2. Phasor Measurements Units (PMU) and RAs send their data using IP multicast to both the grid agent (GA) and a Supervisory Control And Data Acquisition (SCADA) system.

In the following paragraph we consider two key examples to show how the Commelec control framework is able to exploit the predicted PV flexibility at two scopes: 1) to achieve the microgrid dispatchability and 2) to maintain the grid in a safe state of operation.

Since in this work we focus on the inclusion of stochastic generation (e.g. PVs) the considered setup is reduced, namely not all the resources are operating. Further examples of the Commelec framework capabilities can be found in [5].

### 7.2.4 Microgrid Real-Time Dispatchability

In this first experiment, we show the case where the microgrid is continuously requested to change its power flow at the grid connection point (GCP). For this purpose, the GA simply
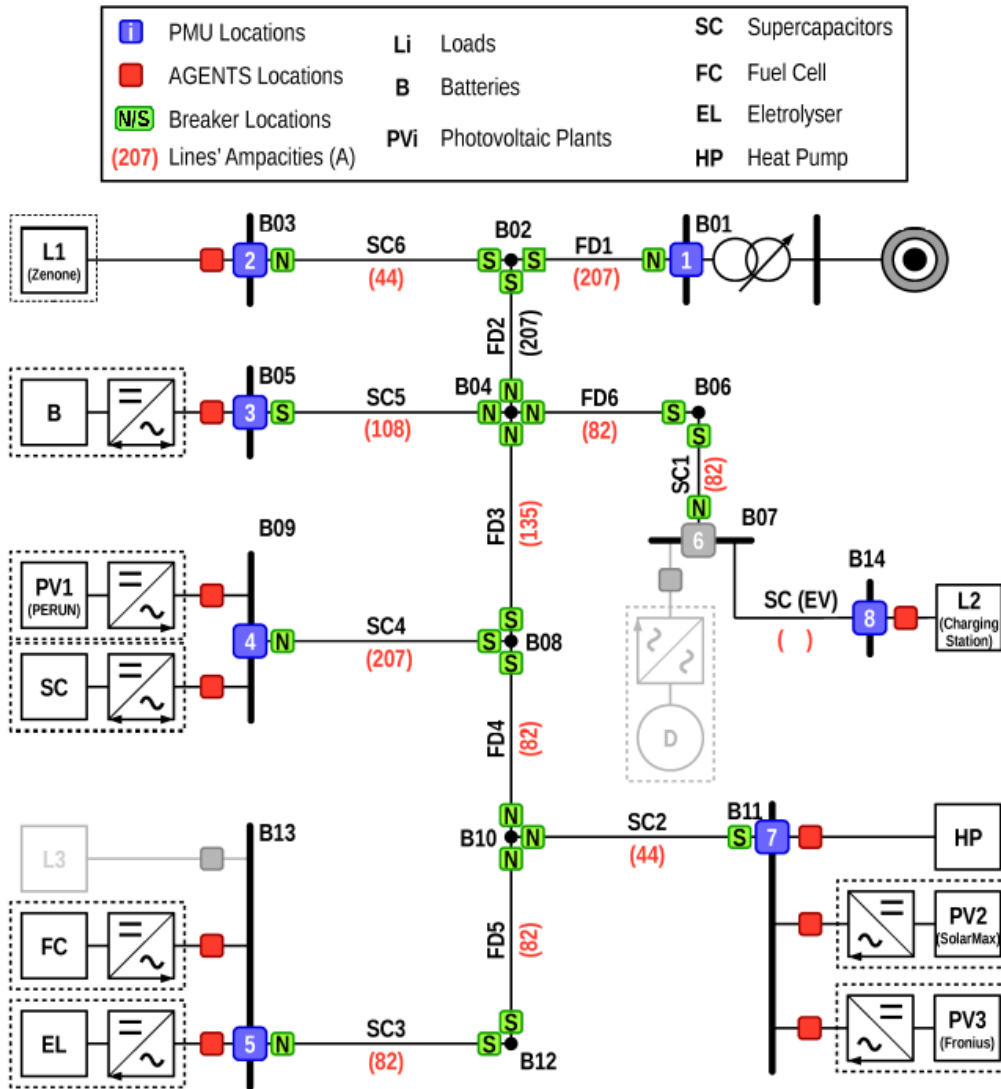
*Figure 7.7 – Microgrid Overall Architecture, [5].*

activates a power-tracking function in its objective ($C_0$ in Eq. (5)). We also set a constant reactive power request Q = 0 kvar. We consider to have in our microgrid only the controllable PV (PV1) and the uncontrollable one UPV (PV2). Thus, the only source of flexibility comes from the controllable PV. Fig. 7.8 shows how it is possible to use the PV flexibility to track the target request at the grid connection point (GCP). At first, both the PV and the UPV are working in MPPT. Then, at Time=50 s, we ask to decrease the power at GCP and so the controllable PV (yellow line) reacts accordingly. After this, a constant power of 15 kW is asked, even if the maximum available power is almost 20 kW. It is interesting to see that when a cloud arrives, at Time=290 s, the controllable PV is asked to react such that the setpoint at the GCP is tracked, i.e. the PV system is able to provide upward regulating power.
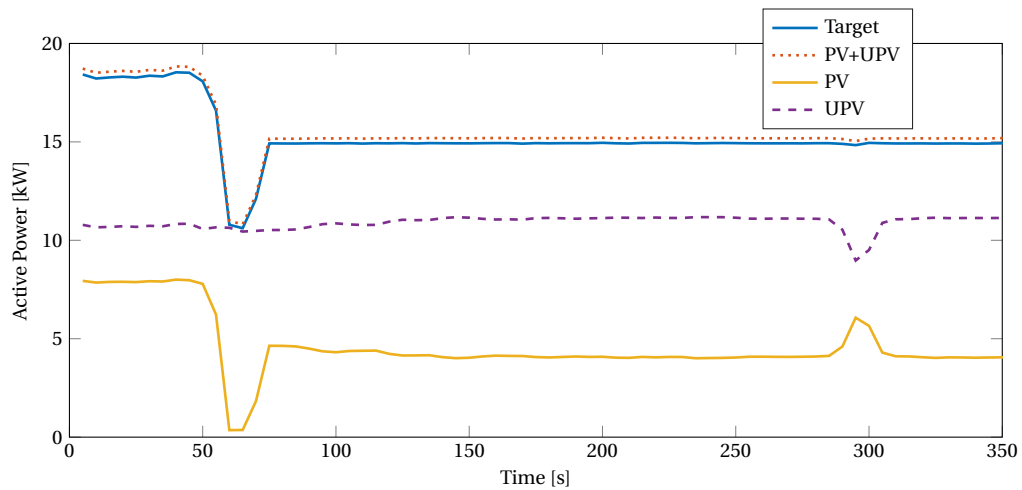
*Figure 7.8 – Real-Time Dispatchability using PV as source of flexibility.*

### 7.2.5   Line Congestion Management using PV Curtailment

In this second experiment, we show how the PV flexibility can be exploited by the GA to maintain the grid in a safe state of operation, avoiding line congestions and/or voltage violations. We consider in our microgrid only the controllable PV (PV1), the uncontrollable PV (PV2), and the Load (L1). The 24 kW residential building is modelled with discrete state-space equations as described in [170]. It consists of 8 rooms, each equipped with an electrical heater of 3 kW that can only be turned on or off. The controlled variable is the room temperature, which needs to remain within a given comfort zone. The PQ profile is therefore calculated to guarantee that the rooms' temperature are within the comfort bounds, the VC is to steer the rooms' temperature to the average temperature of the comfort interval, and the BF is the uncertainty of actuating a given power setpoint. The measured active power values are shown in Fig. 7.9a and the measured current at the GCP in Fig. 7.9b. The ampacity limit of this line is 20 A. At first, both the PV plants operate in MPPT, and the load is consuming active power in order to maintain the rooms temperature inside the comfort zone. At Time=160 s we reproduce the case when the load suddenly disconnects, causing the line current to increase. It is possible to see that the GA immediately curtails the PV power so to maintain the line current below the ampacity limit (20 A). Also in this case, the controllable PV can be used as a source of flexibility and it can be exploited to avoid line congestions.

## 7.3   An ADMM-based Coordination and Control Strategy for PV and Storage to Dispatch Stochastic Prosumers

In this second application, we focus on the integration of stochastic resources in the context of the secondary control/energy management of an electrical feeder. Recently, a new approach has been proposed and experimentally validated at EPFL to achieve the dispatchability of

(a)



(b)
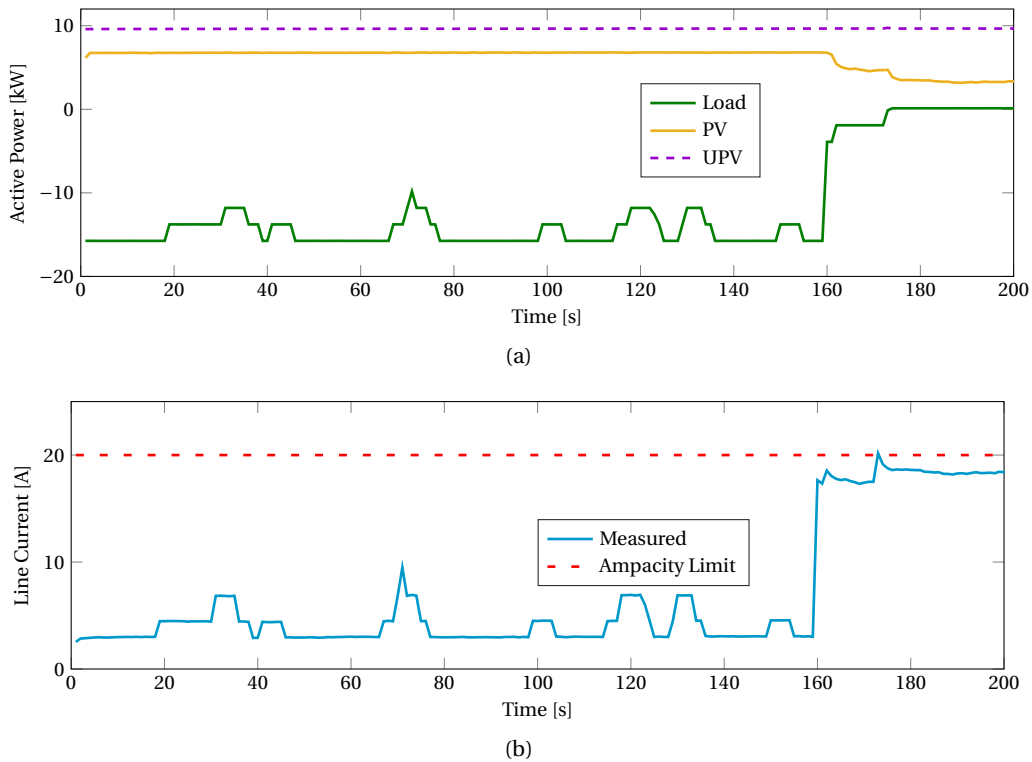
*Figure 7.9 – Line Congestion Management using PV as source of flexibility.*

distribution feeders with distributed resources by controlling utility-scale battery energy storage systems (BESSs), [8]. The process, here referred as *dispatchable feeder*, is divided in two parts:

- The day-ahead operation, where a scheduled trajectory, called  *dispatch plan*, is determined at the grid connection point (GCP) and at 5 minutes resolution, by implementing forecast of the local prosumption.

- The real-time operation, where the mismatch between the dispatch plan and the prosumption realization is corrected by adjusting the real power injections of the BESS converter with model predictive control (MPC). Both battery and consumption forecasting models are data-driven (namely identified from experimental measurements).

For the full discussion, the reader is referred to [8].

The idea of dispatching the active power flow at the GCP can be extended to the case where several flexible resources are available in the network. In particular, as described in Section 6, PV plants can also provide some down power flexibility if the converter is controllable. For this reason, the framework is extended for the case of a curtailable PV facility and battery energy storage system (BESS). Fig. 7.10 shows the experimental setup used to test and validate the proposed control: it consists of a radial distribution feeder interfacing a heterogeneous

mix of demand and generation (office buildings with uncontrollable rooftop PV installations) and two controllable resources, as later detailed in Section 7.3.1. In particular, the day-ahead
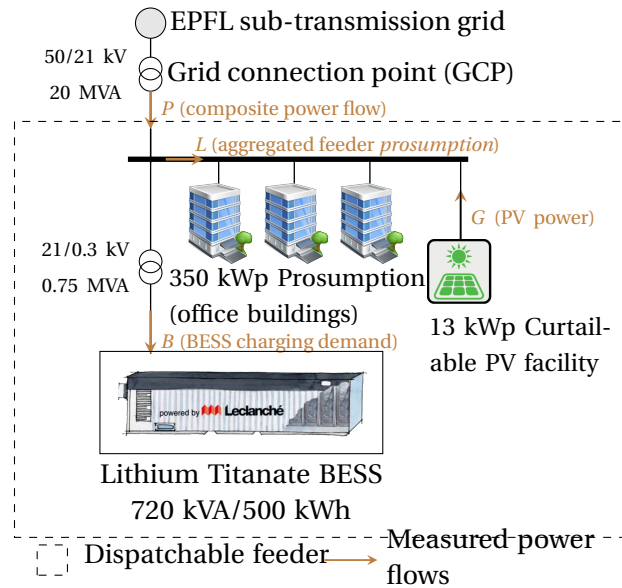


*Figure 7.10 – The EPFL's experimental setup used for the validation.*

operation is formulated as in [8] and it is here assumed as a given trajectory to track. Then, distributed optimization is applied every 5 min to coordinate PV and BESS and provide the real-time operation. The real-time operation consists in two algorithms executed at different paces:

- a lower level tracking problem, executed at 10 s resolution, to achieve a fine tracking of the dispatch plan by controlling the BESS active power. This is achieved with MPC with the formulation proposed in [8].

- an upper level coordination mechanism, running at 5 min resolution, to coordinate the operation between elements. Its role is essentially implementing an energy management strategy to, *i)*, make sure that enough power capacity is available for the faster control loop to compensate for the power mismatch, and, *ii)*, longer-term managing the BESS state-of-charge (SOC) such that enough flexibility is available to compensate for the energy error during the remaining part of the day.

  More in details, the ADMM (Alternating Direction Method of Multipliers, [171]) is applied to share the control objective between multiple resources. Each 5 min, the ADMM determines the trajectories of the battery injection and PV generation to achieve dispatch plan tracking by integrating short-term forecast of the demand and PV generation (allowing to estimate the PV curtailment potential).

The day ahead predictions of the solar potential are obtained from the MeteoTest meteorological web service, while the 5 min ahead are obtained by using a persistent forecast, based on

pyranometers and temperature measurements as inputs. Physical-based modeling tool-chain (transposition of the global horizontal irradiance and PV plant models) are used to compute the PV generation from the current irradiance and temperature. In particular, the transposition model is as described in Appendix B and the PV model as in Section 5.3. [1] The flow chart depicting the real-time operation procedure is sketched in Fig. 7.11. In the flow diagram, controllers running at two time samples: *i)* faster MPC controller executed each 10 s and *ii)* slower ADMM based coordination mechanism computing PV setpoints each 5 minute.
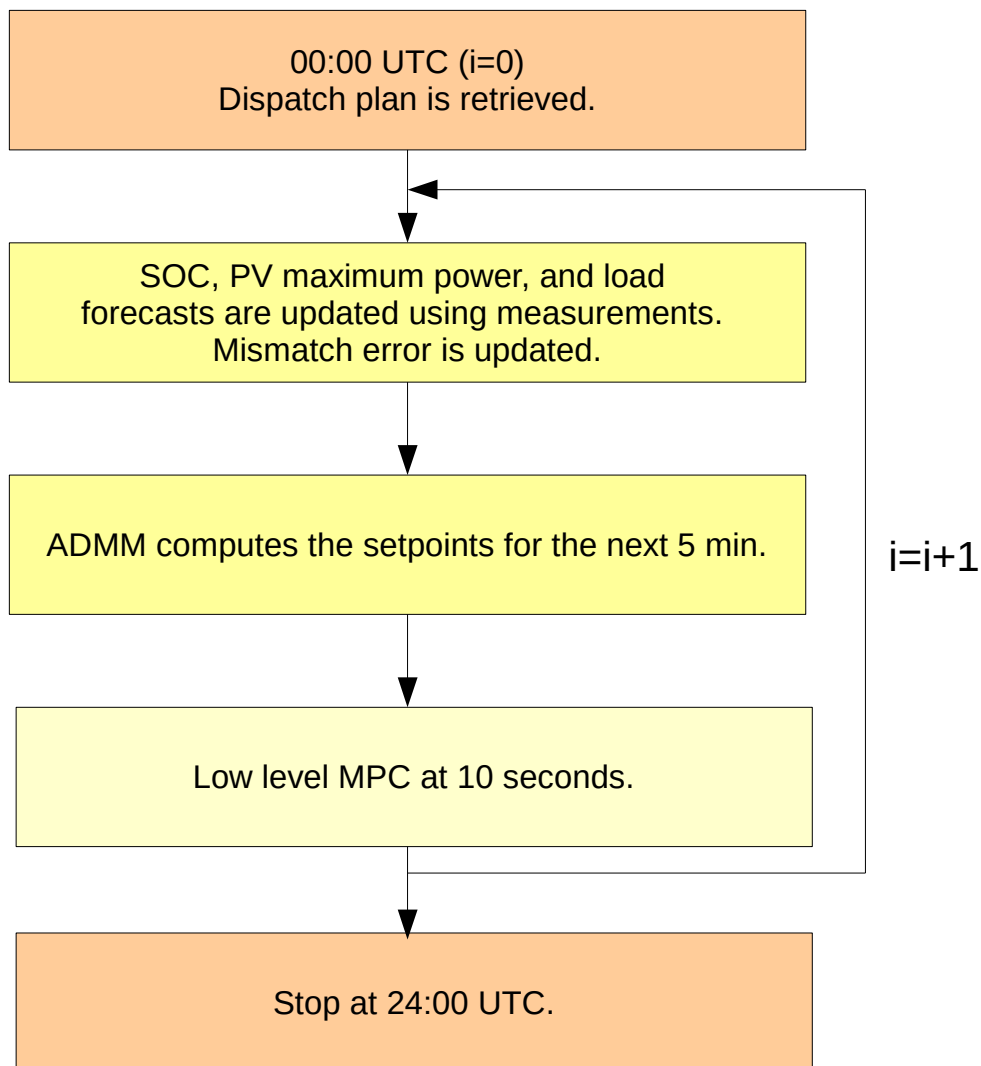


*Figure 7.11 – Flow chart showing real-time operation during 24 hours. The index i denotes the rolling current 5 minute interval.*

---

[1]The persistent forecast is here selected as point forecast method for simplicity. However, more sophisticated tools (e.g. machine learning based) can be integrated for improved performance of the control strategy.

For the sake of clarity, we give here the formulation of the optimization problem, as described in [172]. Let the index $i$ denote the rolling current 5 minute interval, $N$ the number of 5 minutes interval in 24 hours, $j = i,\ldots,N$ a 5 minute time index spanning from the current time until the end of the day, the sequence $\widehat{e}_j$ the forecasted deviation between the dispatch plan and forecasted realization for the remaining part of the day, $B_j^o$ and $G_j^o$ the battery and PV plant setpoint trajectories, respectively, and $\widehat{G}_j$ the maximum theoretical PV power plant output. The sequences $\widehat{e}_j$ and $\widehat{G}_j$ are calculated by applying forecasting tools and are assumed given in the following formulation. We focus here on the formulation of the optimization problem in a centralized manner. The objective is to compute the trajectories $B_j^o$ and $G_j^o$ for the remaining part of the day $j = i\ldots,N$ in order to compensate for the dispatch plan mismatch $e_j$, while performing the minimal amount of curtailment subject to PV and battery system constraints. The PV system constraint is that the active power should be within 0 and the theoretical maximum PV power production $\widehat{G}_j$. For the battery, the constraints are that the power should be within the four-quadrant apparent power converter capability, and the battery SOC within the bounds $(\mathrm{SOC}_j^{\min},\mathrm{SOC}_j^{\max})$.

Formally, the centralized optimization problem is:

$$\underset{\substack{G_i,\ldots,G_N \\ B_i,\ldots,B_N}}{\arg\min} \sum_{j=i}^{N} \left(G_j - \widehat{G}_j\right)^2 \tag{7.10}$$

subject to:

$$B_j + G_j = e_j \qquad\qquad j = i,\ldots,N \tag{7.11}$$
$$0 \le G_j \le \widehat{G}_j \qquad\qquad j = i,\ldots,N \tag{7.12}$$
$$B^{\min} \le B_j \le B^{\max} \qquad\qquad j = i,\ldots,N \tag{7.13}$$
$$\mathrm{SOC}_j^{\min} \le f(\mathrm{SOC}_i, \mathbf{B}_{i-1,j-1}) \le \mathrm{SOC}_j^{\max} \qquad\qquad j = i,\ldots,N. \tag{7.14}$$

It is possible to decompose the centralized problem into distributed ones, which can be solved iteratively until reaching a consensus on a coupling constraint that in this case is the dispatch constraint of Eq. 7.11.

Let $g_j(\cdot)$ be a barrier function with zero cost when the tracking error constraint (7.11) is respected and infinity otherwise:

$$g_j(G_j, B_j) = \begin{cases} 0 & B_j + G_j = e_j \\ \infty & \text{otherwise.} \end{cases} \tag{7.15}$$

Let $\mathcal{G}_j$, $\mathcal{B}_j$ be the variables that mimic the behavior of the original variables $G_j$, $B_j$, the so-called copied variables. The centralized objective can be re-written by moving the system

constraints into the objective function by using the barrier function $g_j(\cdot)$. It is:

$$\arg\min_{\substack{G_i,...,G_N \\ B_i,...,B_N}} \sum_{j=i}^{N} \left(G_j - \widehat{G}_j\right)^2 + \sum_{j=i}^{N} g_j(\mathscr{G}_j, \mathscr{B}_j) \tag{7.16}$$

subject to:

$$B_j - \mathscr{B}_j = 0 \qquad\qquad j = i,...,N \tag{7.17}$$

$$G_j - \mathscr{G}_j = 0 \qquad\qquad j = i,...,N. \tag{7.18}$$

Constraints (7.17) and (7.18) can be moved in the cost function by using two sequences of Lagrangian multipliers, denoted by $\boldsymbol{y_{Gi,N}}$ and $\boldsymbol{y_{Bi,N}}$, referred to as dual variables in the following. The augmented Lagrangian cost function is obtained by moving the equality constraints (7.17)-(7.18) in the cost function (7.16). It is:

$$\begin{aligned} L_\rho = &\sum_{j=i}^{N} \left(G_j - \widehat{G}_j\right)^2 + \sum_{j=i}^{N} g_j(\mathscr{G}_j + \mathscr{B}_j) + \\ &+ \frac{\rho}{2}\left(\left|\left|\boldsymbol{G}_{i,N} - \mathscr{G}_{i,N}\right|\right|_2^2 + \left|\left|\boldsymbol{B}_{i,N} - \mathscr{B}_{i,N}\right|\right|_2^2\right) + \\ &+ \boldsymbol{y_{G}}_{i,N}^T(\boldsymbol{G}_{i,N} - \mathscr{G}_{i,N}) + \boldsymbol{y_{B}}_{i,N}^T(\boldsymbol{B}_{i,N} - \mathscr{B}_{i,N}). \end{aligned} \tag{7.19}$$

For a complete tractation of the ADMM consensus and sharing problem the reader is referred to [172].

### 7.3.1 Results

Characteristics of the involved units are summarized in Table 7.1. The active power flow of the distribution network is monitored by sensing the consumption at the GCP with a PMU-based metering system. The BESS is controllable by sending active/reactive power setpoints to the power converter over a Modbus interface, and the PV plant is controllable by sending active power setpoints to the converter over CAN communication.

*Table 7.1 – Description of connected elements at GCP*

| Component | Parameter | Value |
|---|---|---|
| Grid-connected BESS | Nominal power | 720 kVA |
| | Energy capacity | 560 kWh |
| | Ramping rate | ±20 MW/s |
| Prosumption (office building + rooftop PV) | Peak active power demand | 350 kW |
| | Average demand | 101 kW |
| | Generation capacity | 82 kWp |
| Curtailable PV Plant | Generation capacity | 13 kWp |

The experimental results for one day of operation are shown in Fig. 7.12. In particular, Fig. 7.12a

shows the dispatch plan (in black), the prosumption realization (dashed red), and the realization at the GCP (gray shaded area): the prosumption realization differs from the dispatch plan due to forecasting errors, whereas the realization at the GCP, which is corrected by controlling the contribution of the battery and curtailable PV facility, achieves a good tracking of the dispatch plan. Fig. 7.12b shows the battery injection (in the upper panel), and the SOC evolution and its bounds (bottom panel, full and dashed line, respectively), which are an input of the decision problem. In this case, SOC bounds are constrained at time 20:00 UTC to reproduce a situation where the flexibility is constrained ( e.g., to simulate a situation where the battery capacity is saturated). It is noteworthy that SOC evolution is always within the allowed bounds. Fig. 7.12c shows the PV curtailment action, in particular the theoretical PV maximum power point (in red), the PV converter setpoint (black) and the measured PV active power. During the first part of the day, the PV setpoint corresponds to the maximum available power: the small differences between the two profiles are due to the fact that maximum power point operation is achieved by MPPT algorithms (maximum power point tracking), which normally rely on perturb-and-observe strategies, and not on a physical model-based toolchain, as for our case. As visible in Fig. 7.12c, the PV power is curtailed around midday. This is due to the distributed optimization policy, which decides to implement curtailment in order to satisfy the constraints of the optimization problem. To evaluate whether the curtailment action determined by the distributed optimization problem was necessary, we playback the experimental measurements of the prosumption realization in a ad-hoc simulation framework. where the feeder dispatch is enforced by controlling the battery only. In other words, we want to verify if battery SOC constraints are still respected in the same stochastic conditions but without leveraging the controllability of the curtailable PV facility. Fig. 7.13 compares the experimental SOC (dashed red, from the battery management system, which includes ADMM action), the simulated SOC with ADMM (shaded gray band), the simulated SOC without ADMM action, and the SOC upper and lower bounds (dashed blue and black lines). As visible in Fig. 7.13, the experimental SOC and simulated SOC without ADMM matches until approximately midday, the time when the PV curtailment action begins (from Fig. 7.12c). After midday, the two trajectories diverge because in the latter case the battery has to charge more in order to track the dispatch plan. Nevertheless, in the former case, the battery SOC respect the SOC upper bound constraint, whereas the latter strategy (without ADMM) fails at time 20:00 UTC. We can therefore conclude that the curtailment action determined by the formulation with distributed optimization is necessary to respect BESS constraints. Numerical results comparing the control performance with and without ADMM are shown in Table 7.2. They denote that the distributed optimization control, even if it sacrifices PV generation, it achieves to respect BESS capacity constraints and that it is able to coordinate the operation of the controllable resources to achieve to dispatch stochastic prosumption.

Table 7.3 shows the statistics on the tracking error for the case where there is no control at all (no dispatch), dispatch strategy without ADMM upper layer coordination strategy, and dispatch with ADMM. In this case, the dispatch strategy with achieves the best tracking performance with a RMS error less than 0.5 kW over 24 hours.

*Table 7.2 – PV generation, PV curtailments and SOC constraint violation.*

| Quantity | no ADMM | with ADMM |
|---|---|---|
| Max distance from SOC upper bound constraint (> 0 violation) | 3.09 % | -0.47 % |
| PV Generation | 33.8 kWh | 21.60 kWh |
| PV Curtailement | – | 12.2 kWh |

*Table 7.3 – Tracking error statistics without dispatch, with dispatch and no ADMM, and dispatch + ADMM (kW).*

| Scenario | RMSE | Mean | Max |
|---|---|---|---|
| No dispatch | 11.1 | -4.1 | 36.0 |
| Dispatch without ADMM | 1.60 | 0.53 | 7.8 |
| Dispatch + ADMM | 0.32 | $\leq 0.01$ | 2.27 |

## 7.4   Chapter Conclusions

In this Chapter, we have shown the importance of knowing the future available maximum power of a PV plant, considering two different applications and two experimental setups equipped with a controllable PV plant. The first application concerns explicit power-flow primary control in microgrids and targets dynamics at very short-time scale (sub-second). We have considered a real-time control framework that is able to control a set of heterogeneous resources in a real-scale microgrid deployed in our laboratory at EPFL. We have proved that it is possible to exploit the flexibility of a controllable PV plant if the one-step ahead maximum power and associated uncertainty are known and correctly advertised to a main controller. The results are shown considering two different scenarios, i.e. capability of the microgrid to follow a given dispatch profile and capability to avoid line congestion. The second application refers to a predictive control strategy for the energy management of distributed resources and targets the minute time scale. A curtailable PV facility and a battery energy storage system (BESS) are considered as flexible resources to dispatch stochastic prosumers. The proposed ADMM-based coordination mechanism successfully achieves to curtail PV generation in contingency situation (e.g., loss of battery capacity) and respect battery state-of-charge constraints.
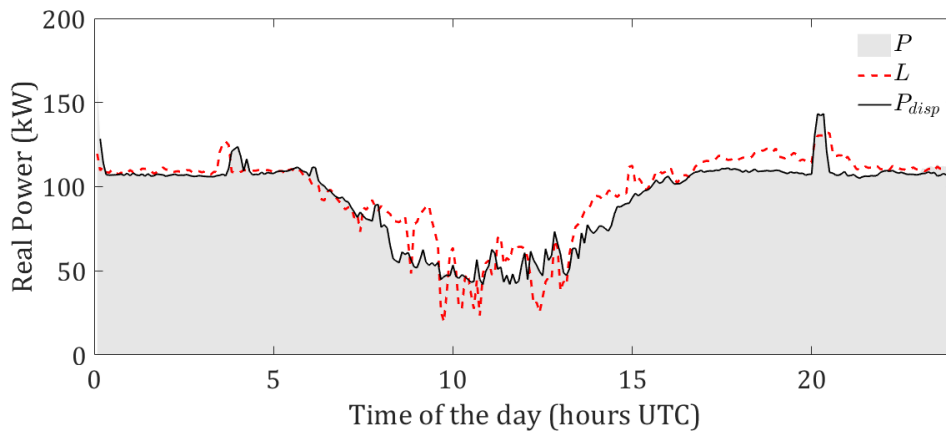
(a) Dispatch plan (black), measured prosumption realization (shaded area), measured active power flow at the GCP (dashed red).
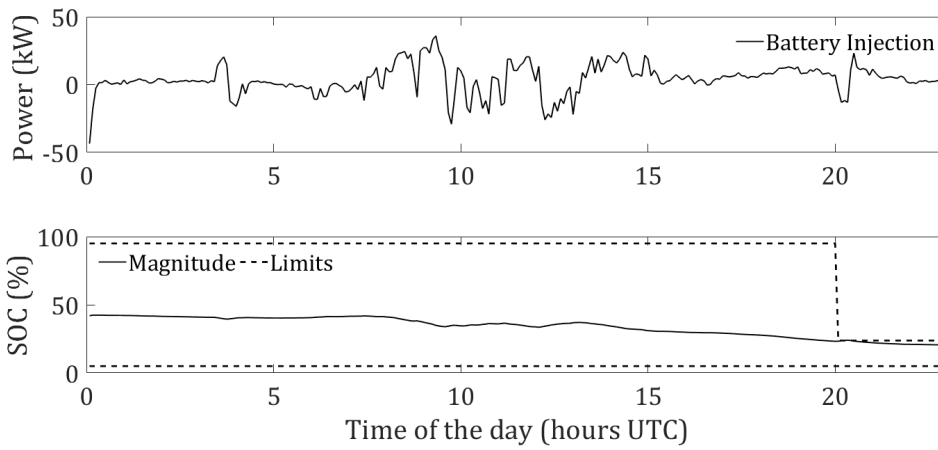


(b) Battery power injection (upper panel), and battery SOC evolution and respective limits (bottom panel).
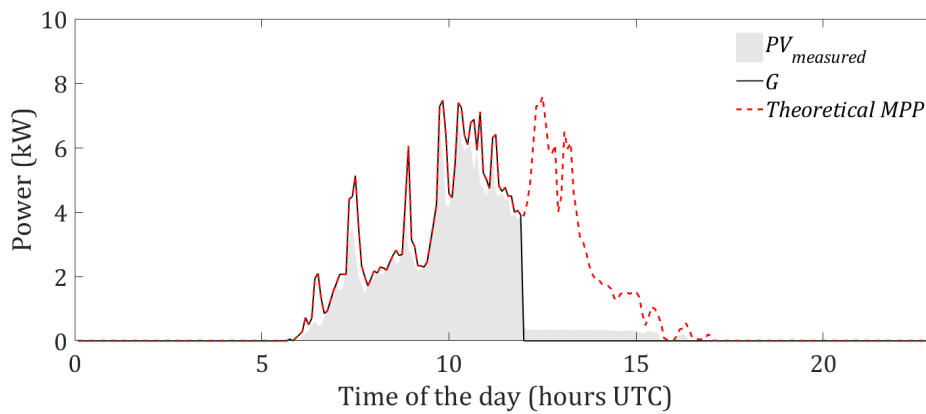


(c) Curtailed PV (black), theoretical maximum power point (MPP) PV (dashed red) and measured PV after implementation (shaded area)

*Figure 7.12 – Operation of the dispatchable feeder with ADMM strategy on 17 September 2017.*

145
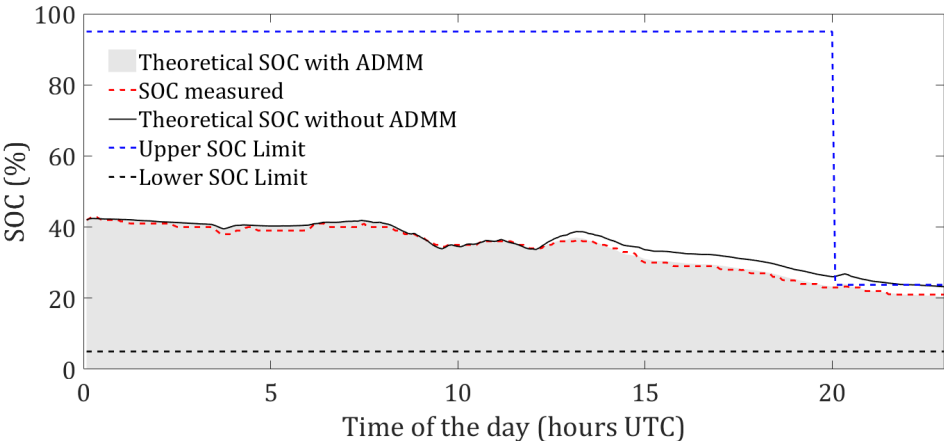
*Figure 7.13 – SOC evolution with and without ADMM. The former is experimental, whereas the latter is obtained by playing back into simulations experimental data. With ADMM, SOC constraints are respected, whereas they are not without ADMM.*

# 8 Conclusions and Future Work

## 8.1 Main Conclusions of the Thesis

The thesis developed, deployed, and experimentally applied new probabilistic methods that allow modern control strategies to account for the fastest dynamics of photovoltaic plants when connected to an electrical grid.

The case of an uncontrolled PV plant that always tracks the maximum power is considered, and a probabilistic method is investigated to directly forecast the PV AC active power output. The method is based on the experimentally observed correlation between the AC active power value, its derivative, and the error caused by a generic point predictor. Results show that the method is able to account for high AC active power fluctuations by adapting the PI width at different confident levels.

Since direct power forecast is not advisable if the power measurements are corrupted by curtailment actions, it might be necessary to first predict the irradiance and then convert it into power using physical models of the plant. For this reason, we have focused on the probabilistic forecasting of the GHI, targeting horizons that are meaningful for microgrids control applications (i.e., from sub-second to five minutes). A method to deliver PIs for GHI is proposed and its performance assessed. The technique extracts information from a limited training set: pyranometers readings are clustered off-line by using the k-means algorithm, and the quantiles of the obtained clusters are then used for the PIs computation. The method is applied to the original and differentiated time-series. Results show that the benefit coming from the time series differentiation decreases while increasing the forecast horizon. It is shown that the algorithm outperforms benchmark cases, e.g. the one using simple quantiles extractions and the one considering a Gaussian distribution of the point forecast error. Furthermore, performance is shown to be in line or improve those available in the literature, for all the considered forecast horizons and using a shorter and limited training set. Thanks to its computational inexpensiveness and good performance at different forecast horizons, the model can be useful for providing forecast of the GHI for the real-time control of microgrids.

Then, the use of ground images to support solar forecasting and sense the GHI is investigated. The main goal is to study if it is possible to detect cloud movements and sun occlusion and to use this information to reduce the uncertainty of predicting algorithms that are only based on past irradiance measurements. First, a chain to predict the percentage of cloudy pixels in the sun disk is proposed. It includes the development of cloud segmentation and motion algorithms. Regarding cloud segmentation, a new method based on the intensity-to-blue ratio is proposed and proved to outperform available models from the literature, including threshold-based methods that consider different color channels (e.g, the red-blue ratio) and supervised segmentation models. Furthermore, it is shown that the predicted local cloud cover can be used as a further feature to improve a machine learning based method to deliver PIs of the GHI, besides the sole use of past irradiance measurements. The possibility of using an all-sky imager as irradiance sensor is also investigated. At first, a large set of features is extracted from the images, and then feature selection is performed using principal component analysis (PCA). The subset of selected features is then used as input to train an artificial neural network (ANN). Training and performance evaluation is computed using pyranometer measurements, collected in the same location as the camera. The GHI estimations are benchmarked against the Heliosat-2, a well-established method to estimate the GHI from satellites (e.g. Meteosat in Europe). Results show that the all-sky camera-based GHI estimations proposed in this work outperform the Heliosat-2, with a RMSE relative improvement of 20-45%. In particular, this happens when fast irradiance dynamics are present due to the fact that satellites-based models lack the spatial and temporal resolution needed to capture localized fluctuations. The method was extended to the case of GHI point forecast, showing an improvement with respect to the baseline persistent method for forecast horizons larger than 1 min.

Then, the case when estimations of the irradiance are not available from sensors (e.g. pyranometers, sky-images, satellites) is considered. To cope with this issue, an alternative approach is proposed to reconstruct the maximum power of a PV plant: a gray-box model that leverage on measurements of the DC voltage, current, and cell temperature, and on a physical model of the PV plant. Three model-based estimators are investigated starting from the circuital equations of the PV plant, corresponding to three formulations: the analytical, the Immersion & Invariance (I&I), and the Kalman Filter (KF). Results showed that the considered estimators can reconstruct successfully the theoretical maximum power output of PV installations in all the operating conditions and outperform pyranometer-based maximum power estimations. For noise levels compatible with commercial sensors, the analytical estimator showed better performance and bandwidth than the I&I and KF-based estimators. However, if the available measurements are characterized by a high level of noise, the use of filter-based strategies is advisable since they are able to delete noisy observations while accounting for the structure of the process. An application of the maximum power estimators to improve data-driven maximum power forecasting tools is proposed. We showed that training the forecasting algorithm with historical data that are corrupted by curtailment actions leads to a deterioration of the learning stage and therefore to a worse performance of the forecasting method. In this context, we apply the analyzed methods to reconstruct the maximum point and we, therefore, make

sure that the forecasting algorithm learns from values that are always representative of the true potential of the plant.

Then, the problem of assessing the overall uncertainty associated with the operation of a controllable PV facility is tackled. The analysis includes two fundamental aspects: determining the PV generation potential for the next time interval (allowing identifying a region inside the capability curve of the power converter from where it is possible to pick a power setpoint), and characterizing the uncertainty associated to tracking a setpoint inside that region. The former element is determined by using the discussed gray-box model based on the analytical formulation, whereas the latter depends on the converter internal dynamics and is evaluated with a data-driven approach. Experimental results showed that, when considering forecasting horizons below 1 min, accurate physical models are the key to achieve reliable coverage of the prediction interval. On the other hand, for forecasting horizons larger than 1 min, the quality of the predictions is mostly affected by the uncertainty associated with the forecasting, which becomes dominant with respect to models performance.

The thesis concludes by showing two possible applications to prove how to profit from knowing the future availability of a PV plant. The first application concerns primary control in microgrids at a very short-time scale (sub-second). We have considered a control framework, called Commelec, that controls a set of heterogeneous resources (batteries, loads, PVs..) in a real-scale microgrid deployed at EPFL. Commelec consists of a main controller that, upon receiving information of the status of the available resources, is capable of managing power flows in the grid while guaranteeing its safe operation. We have proved that it is possible to exploit the flexibility of a controllable PV plant if the one-step-ahead maximum power and associated uncertainty are computed and correctly advertised to a main controller. Results are shown considering two different services, i.e. the capability of the microgrid to follow a given dispatch profile at the grid connection point and to avoid line congestion. The second application refers to a predictive control strategy, also deployed at EPFL, to dispatch the operation of a distribution feeder with heterogeneous prosumers according to a trajectory with 5 min resolution. A curtailable PV facility and a battery energy storage system are exploited as flexible resources to dispatch a set of stochastic prosumers. The proposed distributed coordination mechanism successfully achieved to curtail PV generation in contingency situation (e.g., loss of battery capacity) while respecting battery state-of-charge constraints.

## 8.2   Future Work

Overall, the methods proposed can be used to integrate PV plants into modern electrical grids more effectively and reliably. However, more research is still required in this direction. Possible interesting points that need further investigation are:

- Assess of the performance of the developed forecasting tools when considering larger geographical areas. This includes both the case of a single large PV plant and the case

of aggregated distributed PV plants. In particular, it would be interesting to asses for which level of aggregation or spatial extension, local estimations (from pyranometers, sky-cameras and monitored PV plants) are required to capture local irradiance dynamics or, viceversa, when satellite estimations, characterized by a lower resolution, are instead sufficient for this scope.

- Segmentation and cloud motion algorithms to detect cloud dynamics from the sky-images should be improved. Possible research includes considering more sophisticated segmentation methods, e.g. extending the binary segmentation to multiple regions, or considering *dynamic cloud segmentation* where information from contiguous pictures are incorporated.

- For the all-sky camera, all the discussed algorithms should be validated on a larger set to consider performance over an entire year.

- The ASIs-based GHI estimations proposed in Chapter 4 could be improved by including dynamic features (for example, the variation of the intensity during consecutive pictures, the variation of the cloud cover, etc.). It could be also interesting to implement more sophisticated deep learning techniques.

- Regarding the PV modelling chain, future research should also consider the effects of fault occurrence and degradation processes of the PV cells. For example, adaptive model identification with periodical re-training of model parameters can be implemented to account for degradation.

# A Appendix A

The following describes how to compute the variance $Q$ for the Estimated Kalman Filter-based irradiance estimation. In summary, it consists in grouping $N$ historical values of the differentiated irradiance time series ($\Delta I$) into clusters according to the value of selected data features:

- the average irradiance value on a mobile window of length $n$ considering the most recent data points:

$$M_i = \frac{1}{n} \sum_{j=i-n}^{i} \Delta I_j, \ i = n+1,\dots,N \tag{A.1}$$

- the irradiance variability:

$$V_i = \sqrt{\frac{1}{n} \sum_{j=i-n}^{i} (\Delta I_j - \Delta I_{j-1})^2}, \ i = n+1,\dots,N \tag{A.2}$$

The k-means iterative algorithm is used to classify historical observations using its formulation in Matlab, [97]. It returns $k$ clusters $G_1,\dots,G_k$ and their centroids $\mathbf{c_1},\dots\mathbf{c_k}$; the histograms of these clusters are assumed as the empirical pdfs of the variations with respect to the one-step-ahead irradiance realization. The number of clusters is chosen empirically with the objective of minimizing the variance of each cluster pdf. During real-time operation, the data features vector at time $t$, denoted by $\mathbf{p_t} = (M_t, V_t)$, is calculated. The next step is the calculation of the Euclidean distances between $\mathbf{p_t}$ and the centroids $\mathbf{c_l}$

$$d_l = \|\mathbf{c_l} - \mathbf{p_t}\|^2, \ l = 1,\dots,k \tag{A.3}$$

which is used as a similarity criterion to select the cluster representative of the future irradiance. We indicate with $\hat{l}$ the index corresponding to the cluster with minimum distance. The variance

of the cluster pdf is then used as the value for $Q$. In particular:

$$Q_{t+1|t} = \text{Var}(G_{\hat{i}}). \tag{A.4}$$

It is worth noting that while doing this, we do the approximation that clusters pdfs are normally distributed. Besides, we note that determining $Q$ requires past irradiance values, whereas previous methods do not.

# B Appendix B

In this Appendix, we compare the performance of three maximum power estimators. In particular, we consider that the GHI is firstly measured using three different systems explored in the thesis: a pyranometer, an all-sky camera, and satellites. For the first case, a pyranometer is installed in the vicinity of the PV plant, directly returning GHI measurements, as described in Chapter 3 (Section 3.8). In the second case, the *ASI-based* method in Section 4.8 is used to estimate the GHI by extracting meaningful influential variables from the images returned by the all-sky camera, also installed in the vicinity of the plant. In the last case, Heliosat-2 GHI estimations are retrieved from the Helioclim-3 database for the GPS location of the plant under test as described in Section 4.8.3. Then, all the GHI values are transformed into plane-of-array (POA) irradiance values using transposition models from the literature (see Section B.1 of this Appendix), and then the transposed irradiance is converted into DC power values using the 5-parameter PV model, as described in Sections 5.3 and 5.5. The cell temperature, used as second input of the model, is measured using a temperature sensor installed on the rear surface of the panel. The tested plant is a 3.7 kW rooftop PV installation, part of the plant introduced in Section 2.2.1. To evaluate the ability of the different methods in capturing fast power fluctuations, the estimated DC power values are compared with the measured DC power at 1 min time resolution.

## B.1   Transposition Model

In order to transpose the global horizontal irradiance ($I$) into the plane of array irradiance $I^{\searrow}$, we first need to decompose it into the beam and diffuse component, indicated with $B$ and $D$ respectively, [173]. By neglecting the reflected component of the irradiance, we can write $I = B + D$. First we can compute the direct irradiance by knowing the clear-sky index $K$ and

the clear-sky irradiance $I_{cs}$, using the adjustment proposed in [174]:

$$B = \begin{cases} 0 & K < 19/69 \\ B_{cs}(K - 0.38(1 - K))^{2.5}, & 16/69 \leq K \leq 1 \\ B_{cs}K, & K > 1. \end{cases} \tag{B.1}$$

where the clear-sky index is defined as the ratio between $I$ and the clear-sky irradiance $I_{cs}$:

$$K = \frac{I}{I_{cs}}. \tag{B.2}$$

The clear-sky values are calculated using the Ineichen/Perez model, [138, 139], as formulated in the PVLib library, [104]. Then, the global irradiance on the tilted plane $I^{\searrow}$ can be computed using the Klucher model, [175]:

$$\begin{aligned} I^{\searrow} = &D\left(\frac{1 + \cos\theta_T}{2}\right)\left(1 + F\sin^3\frac{\theta_T}{2}\right)(1.. + F\cos^2\theta_I\sin^3\theta_z) \\ &+ B\frac{\cos\theta_I}{\cos\theta_z}, \end{aligned} \tag{B.3}$$

where $F = 1 - (D/(B + D))^2$, $\theta_T$ is the panel tilt angle, $\theta_I$ the solar incidence angle normal to the angled plane.

## B.2   Results on DC Maximum Power Estimation from Different Sensors

Let $P_t$ be the ground truth maximum GHI value at the time interval $t = 1, \ldots, T$, where $T$ is the number of samples, $\widehat{P}_{dc,t}$ the estimation. Three metrics are used to characterize the performance of the proposed techniques, the normalized root mean squared error (nRMSE), the mean absolute error (MAE), and the mean bias error (MBE):

$$\text{nRMSE} = \frac{1}{P_{max}} \cdot \sqrt{\sum_{t=1}^{T} \frac{(\widehat{P}_{dc,t} - P_{dc,t})^2}{T}}, \tag{B.4}$$

$$\text{MAE} = \sum_{t=1}^{T} \frac{|\widehat{P}_{dc,t} - P_{dc,t}|}{T}, \tag{B.5}$$

$$\text{MBE} = \sum_{t=1}^{T} \frac{\widehat{P}_{dc,t} - P_{dc,t}}{T}. \tag{B.6}$$

where $P_{max}$ is the maximum power measured in the considered time period $T$. Results of the estimation performance are shown in Table B.1 for 15 days during January 2017. It is possible to see that the lowest error is returned by the local measurements (pyranometer and

all-sky camera). The relative improvement of the pyranometer and the ASIs with respect to the Heliosat-2 method is ≈40% and ≈25% in terms of nRMSE. The probability density of the estimation error is shown in Fig. B.1 for all the considered methods.

*Table B.1 – Estimations Comparison for 15 days in January.*

|  | Pyranometer | ASIs | Heliosat-2 |
|---|---|---|---|
| nRMSE [%] | 7.3 | 9.1 | 12 |
| MAE [W] | 72 | 113 | 151 |
| MBE [W] | 76 | 79 | 107 |
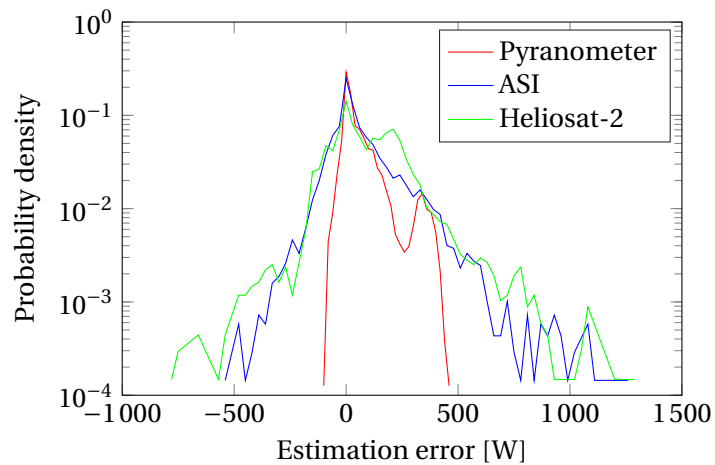


*Figure B.1 – Probability density of the pyranometer, ASIs and H-2 DC power estimation errors.*

As expected, the use of local measurements of the irradiance (namely pyranometer and ASI) leads to better estimations of the PV generation because they are representative of the local potential. On the other hand, satellite estimates of the GHI denote temporal and spatial averages, therefore they are less representative of local conditions. The pyranometer shows better performance than the ASIs-based.

# Bibliography

[1] J. S. Stein, "The photovoltaic performance modeling collaborative (pvpmc)," in *2012 38th IEEE Photovoltaic Specialists Conference*, June 2012, pp. 003 048–003 052.

[2] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. M. de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78 – 111, 2016.

[3] W. D. Soto, S. Klein, and W. Beckman, "Improvement and validation of a model for photovoltaic array performance," *Solar Energy*, vol. 80, no. 1, pp. 78 – 88, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X05002410

[4] A. Bernstein, L. Reyes-Chamorro, J.-Y. Le Boudec, and M. Paolone, "A Composable Method for Real-Time Control of Active Distribution Networks with Explicit Power Setpoints. Part I: Framework," *Electric Power Systems Research*, vol. 125, pp. 254–264, 2015.

[5] L. Reyes-Chamorro, A. Bernstein, N. J. Bouman, E. Scolari, A. Kettner, B. Cathiard, J.-Y. Le Boudec, and M. Paolone, "Experimental validation of an explicit power-flow primary control in microgrids," *IEEE Transactions on Industrial Informatics*, 2018.

[6] D. E. Olivares, A. Mehrizi-Sani, A. H. Etemadi, C. Canizares, R. Iravani, M. Kazerani, A. H. Hajimiragha, O. Gomis-Bellmunt, M. Saeedifard, R. Palma-Behnke *et al.*, "Trends in microgrid control," *Smart Grid, IEEE Transactions on*, vol. 5, no. 4, pp. 1905–1919, 2014.

[7] M. H. Bollen and F. Hassan, *Integration of distributed generation in the power system*. John wiley & sons, 2011, vol. 80.

[8] F. Sossan, E. Namor, D. Torregrossa, R. Cherkaoui, and M. Paolone, "Achieving the dispatchability of distribution feeders through prosumers data driven forecasting and model predictive control of electrochemical storage," Submitted.

[9] N. Troy, D. Flynn, and M. OMalley, "Multi-mode operation of combined-cycle gas turbines with increasing wind penetration," *Power Systems, IEEE Transactions on*, vol. 27, no. 1, pp. 484–492, 2012.

[10] R. H. Inman, H. T. Pedro, and C. F. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp.

535 – 576, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/
S0360128513000294

[11] R. Perez, M. David, T. E. Hoff, M. Jamaly, S. Kivalov, J. Kleissl, P. Lauret, M. Perez *et al.*,
"Spatial and temporal variability of solar energy," *Foundations and Trends® in Renewable
Energy*, vol. 1, no. 1, pp. 1–44, 2016.

[12] D. Torregrossa, J.-Y. Le Boudec, and M. Paolone, "Model-free computation of ultra-short-
term prediction intervals of solar irradiance," *Solar Energy*, 2015.

[13] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic
applications: A review," *Progress in Energy and Combustion Science*, vol. 34, no. 5, pp.
574 – 632, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/
S0360128508000026

[14] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, and B. Washom,
"Intra-hour forecasting with a total sky imager at the {UC} san diego solar energy
testbed," *Solar Energy*, vol. 85, no. 11, pp. 2881 – 2893, 2011. [Online]. Available:
http://www.sciencedirect.com/science/article/pii/S0038092X11002982

[15] C. D'Adamo, S. Jupe, and C. Abbey, "Global survey on planning and operation of active
distribution networks-update of cigre c6. 11 working group activities," in *Electricity
Distribution-Part 1, 2009. CIRED 2009. 20th International Conference and Exhibition on*.
IET, 2009, pp. 1–4.

[16] N. Hadjsaid, J.-F. Canard, and F. Dumas, "Dispersed generation impact on distribution
networks," *Computer Applications in Power, IEEE*, vol. 12, no. 2, pp. 22–28, Apr 1999.

[17] I. P. P. S. D. P. Committee, "Microgrid stability definitions, analysis, and modeling," 2018.

[18] C. J. W. G. B5/C6.26/CIRED, "Protection of distribution systems with distributed energy
resources," Tech. Rep., 2014.

[19] C. Cambini, A. Meletiou, E. Bompard, and M. Masera, "Market and regulatory factors
influencing smart-grid investment in europe: Evidence from pilot projects and
implications for reform," *Utilities Policy*, vol. 40, pp. 36 – 47, 2016. [Online]. Available:
http://www.sciencedirect.com/science/article/pii/S095717871630073X

[20] S. Ruester, S. Schwenen, C. Batlle, and I. Perez-Arriaga, "From distribution
networks to smart distribution systems: Rethinking the regulation of european
electricity dsos," *Utilities Policy*, vol. 31, pp. 229 – 237, 2014. [Online]. Available:
http://www.sciencedirect.com/science/article/pii/S0957178714000198

[21] R. H. Lasseter and P. Paigi, "Microgrid: A conceptual solution," in *Power Electronics
Specialists Conference, 2004. PESC 04. 2004 IEEE 35th Annual*, vol. 6. IEEE, 2004, pp.
4285–4290.

[22] N. Hatziargyriou, H. Asano, R. Iravani, and C. Marnay, "Microgrids," *IEEE power and energy magazine*, vol. 5, no. 4, pp. 78–94, 2007.

[23] A. G. Madureira and J. P. Lopes, "Ancillary services market framework for voltage control in distribution networks with microgrids," *Electric Power Systems Research*, vol. 86, pp. 1–7, 2012.

[24] A. Bidram and A. Davoudi, "Hierarchical structure of microgrids control system," *Smart Grid, IEEE Transactions on*, vol. 3, no. 4, pp. 1963–1976, 2012.

[25] F. Gao and M. R. Iravani, "A control strategy for a distributed generation unit in grid-connected and autonomous modes of operation," *Power Delivery, IEEE Transactions on*, vol. 23, no. 2, pp. 850–859, 2008.

[26] K. De Brabandere, B. Bolsens, J. Van den Keybus, A. Woyte, J. Driesen, and R. Belmans, "A voltage and frequency droop control method for parallel inverters," *Power Electronics, IEEE Transactions on*, vol. 22, no. 4, pp. 1107–1115, 2007.

[27] Y. Li, D. M. Vilathgamuwa, and P. C. Loh, "Design, analysis, and real-time testing of a controller for multibus microgrid system," *Power electronics, IEEE transactions on*, vol. 19, no. 5, pp. 1195–1204, 2004.

[28] F. Katiraei and M. R. Iravani, "Power management strategies for a microgrid with multiple distributed generation units," *Power Systems, IEEE Transactions on*, vol. 21, no. 4, pp. 1821–1831, 2006.

[29] J. M. Guerrero, M. Chandorkar, T.-L. Lee, and P. C. Loh, "Advanced control architectures for intelligent microgrids, part i: decentralized and hierarchical control," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 4, pp. 1254–1262, 2013.

[30] J. M. Guerrero, J. Matas, L. García De Vicuña, M. Castilla, and J. Miret, "Decentralized control for parallel operation of distributed generation inverters using resistive output impedance," *Industrial Electronics, IEEE Transactions on*, vol. 54, no. 2, pp. 994–1004, 2007.

[31] E. Demirok, P. C. Gonzalez, K. H. Frederiksen, D. Sera, P. Rodriguez, and R. Teodorescu, "Local reactive power control methods for overvoltage prevention of distributed solar inverters in low-voltage grids," *IEEE Journal of Photovoltaics*, vol. 1, no. 2, pp. 174–182, 2011.

[32] S. Ghosh, S. Rahman, and M. Pipattanasomporn, "Local distribution voltage control by reactive power injection from pv inverters enhanced with active power curtailment," in *PES General Meeting| Conference & Exposition, 2014 IEEE*. IEEE, 2014, pp. 1–5.

[33] E. Dall'Anese, S. V. Dhople, and G. B. Giannakis, "Optimal dispatch of photovoltaic inverters in residential distribution systems," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 2, pp. 487–497, 2014.

[34] S. Ghosh, S. Rahman, and M. Pipattanasomporn, "Distribution voltage regulation through active power curtailment with pv inverters and solar generation forecasts," *IEEE Transactions on Sustainable Energy*, 2017.

[35] A. Ulbig, T. S. Borsche, and G. Andersson, "Analyzing rotational inertia, grid topology and their role for power system stability," *IFAC-PapersOnLine*, vol. 48, no. 30, pp. 541 – 547, 2015, 9th IFAC Symposium on Control of Power and Energy Systems CPES 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405896315030785

[36] A. Borghetti, M. Bosetti, S. Grillo, S. Massucco, C. A. Nucci, M. Paolone, and F. Silvestro, "Short-term scheduling and control of active distribution systems with high penetration of renewable resources," *Systems Journal, IEEE*, vol. 4, no. 3, pp. 313–322, 2010.

[37] L. Reyes-Chamorro, A. Bernstein, J.-Y. Le Boudec, and M. Paolone, "A Composable Method for Real-Ttime Control of Active Distribution Networks with Explicit Power Setpoints. Part II: Implementation and Validation," *Electric Power Systems Research*, vol. 125, pp. 265–280, 2015.

[38] M. Pipattanasomporn, H. Feroze, and S. Rahman, "Multi-agent systems in a distributed smart grid: Design and implementation," in *Power Systems Conference and Exposition, 2009. PSCE'09. IEEE/PES.* IEEE, 2009, pp. 1–8.

[39] C. Rehtanz, *Autonomous systems and intelligent agents in power system control and operation.* Springer Science & Business Media, 2003.

[40] S. P. Europe, "Global market outlook for solar power 2016–2020," *Solar Power Europe: Bruxelles, Belgium*, p. 32, 2016.

[41] F. L. Lewis and F. Lewis, *Optimal estimation: with an introduction to stochastic control theory.* Wiley New York et al., 1986.

[42] J. Ackermann, *Robust control: Systems with uncertain physical parameters.* Springer Science & Business Media, 2012.

[43] X. Guan, Z. Xu, and Q. Jia, "Energy-efficient buildings facilitated by microgrid," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 243–252, Dec 2010.

[44] S. Tan, J. Xu, and S. K. Panda, "Optimization of distribution network incorporating distributed generators: An integrated approach," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2421–2432, Aug 2013.

[45] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 52–63, 2013.

[46] Y. Zhang, N. Gatsis, and G. B. Giannakis, "Robust energy management for microgrids with high-penetration renewables," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 4, pp. 944–953, Oct 2013.

[47] D. T. Reindl, W. A. Beckman, and J. A. Duffie, "Diffuse fraction correlations," *Solar energy*, vol. 45, no. 1, pp. 1–7, 1990.

[48] J. Hay and J. Davies, "Calculation of the solar radiation incident on an inclined surface. proceedings, first canadian solar radiation data workshop," *Toronto, Ontario, Canada*, 1978.

[49] R. Perez, R. Seals, P. Ineichen, R. Stewart, and D. Menicucci, "A new simplified version of the perez diffuse irradiance model for tilted surfaces," *Solar energy*, vol. 39, no. 3, pp. 221–231, 1987.

[50] M. Lave, W. Hayes, A. Pohl, and C. W. Hansen, "Evaluation of global horizontal irradiance to plane-of-array irradiance models at locations across the united states," *IEEE journal of Photovoltaics*, vol. 5, no. 2, pp. 597–606, 2015.

[51] F. Mavromatakis, E. Kavoussanaki, F. Vignola, and Y. Franghiadakis, "Measuring and estimating the temperature of photovoltaic modules," *Solar Energy*, vol. 110, pp. 656–666, 2014.

[52] D. King, W. Boyson, and J. Kratochvil, "Photovoltaic array performance model. sandia national laboratories 2004. report sand2004-3535."

[53] M. A. Green, "Solar cells: operating principles, technology, and system applications," *Englewood Cliffs, NJ, Prentice-Hall, Inc., 1982. 288 p.*, 1982.

[54] J. L. Gray, "The physics of the solar cell," *Handbook of photovoltaic science and engineering*, pp. 61–112, 2003.

[55] D. S. Chan and J. C. Phang, "Analytical methods for the extraction of solar-cell single-and double-diode model parameters from iv characteristics," *IEEE Transactions on Electron devices*, vol. 34, no. 2, pp. 286–293, 1987.

[56] V. J. Chin, Z. Salam, and K. Ishaque, "Cell modelling and model parameters estimation techniques for photovoltaic simulator application: A review," *Applied Energy*, vol. 154, pp. 500 – 519, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261915006455

[57] A. R. Jordehi, "Parameter estimation of solar photovoltaic (pv) cells: A review," *Renewable and Sustainable Energy Reviews*, vol. 61, pp. 354 – 371, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032116300016

[58] H. Tian, F. Mancilla-David, K. Ellis, E. Muljadi, and P. Jenkins, "A cell-to-module-to-array detailed model for photovoltaic panels," *Solar energy*, vol. 86, no. 9, pp. 2695–2706, 2012.

[59] W. E. Boyson, G. M. Galbraith, D. L. King, and S. Gonzalez, "Performance model for grid-connected photovoltaic inverters." Sandia National Laboratories, Tech. Rep., 2007.

[60] A. Driesse, P. Jain, and S. Harrison, "Beyond the curves: Modeling the electrical efficiency of photovoltaic inverters," in *2008 33rd IEEE Photovoltaic Specialists Conference*, May 2008, pp. 1–6.

[61] M. J. Reno, C. W. Hansen, and J. S. Stein, "Global horizontal irradiance clear sky models: Implementation and analysis."

[62] T. E. Hoff and R. Perez, "Quantifying pv power output variability," *Solar Energy*, vol. 84, no. 10, pp. 1782–1793, 2010.

[63] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr, and T. E. Hoff, "Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance," *Solar Energy*, vol. 86, no. 8, pp. 2170–2176, 2012.

[64] ARENA, "Report: Investigating the impact of solar variability on grid stability," 2015.

[65] C. Tao, D. Shanxu, and C. Changsong, "Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement," in *The 2nd International Symposium on Power Electronics for Distributed Generation Systems*, June 2010, pp. 773–777.

[66] E. Lorenz, D. Heinemann, H. Wickramarathne, H. Beyer, and S. Bofinger, "Forecast of ensemble power production by grid-connected pv systems," in *20th European PV Conference*. Milano, 2007, pp. 3–9.

[67] J. G. Silva Fonseca, T. Oozeki, T. Takashima, G. Koshimizu, Y. Uchida, and K. Ogimoto, "Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in kitakyushu, japan," *Progress in photovoltaics: Research and applications*, vol. 20, no. 7, pp. 874–882, 2012.

[68] M. Abuella and B. Chowdhury, "Solar power forecasting using artificial neural networks," in *North American Power Symposium (NAPS), 2015*. IEEE, 2015, pp. 1–5.

[69] M. Rana, I. Koprinska, and V. G. Agelidis, "Univariate and multivariate methods for very short-term solar photovoltaic power forecasting," *Energy Conversion and Management*, vol. 121, pp. 380 – 390, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0196890416303934

[70] R. Marquez, H. T. Pedro, and C. F. Coimbra, "Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to anns," *Solar Energy*, vol. 92, pp. 176 – 188, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X13000881

[71] E. Lorenz, A. Hammer, D. Heinemann *et al.*, "Short term forecasting of solar radiation based on satellite data," in *EUROSUN2004 (ISES Europe Solar Congress)*, 2004, pp. 841–848.

[72] J. Calbo and J. Sabburg, "Feature extraction from whole-sky ground-based images for cloud-type recognition," *Journal of Atmospheric and Oceanic Technology*, vol. 25, no. 1, pp. 3–14, 2008.

[73] Y. Chu, H. T. Pedro, and C. F. Coimbra, "Hybrid intra-hour {DNI} forecasts with sky image processing enhanced by stochastic learning," *Solar Energy*, vol. 98, Part C, pp. 592 – 603, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X13004325

[74] R. Marquez, V. G. Gueorguiev, and C. F. Coimbra, "Forecasting of global horizontal irradiance using sky cover indices," *Journal of Solar Energy Engineering*, vol. 135, no. 1, p. 011017, 2013.

[75] A. Bracale, P. Caramia, G. Carpinelli, A. R. Di Fazio, and G. Ferruzzi, "A bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control," *Energies*, vol. 6, no. 2, pp. 733–747, 2013.

[76] M. Ghonima, B. Urquhart, C. Chow, J. Shields, A. Cazorla, and J. Kleissl, "A method for cloud detection and opacity classification based on ground based sky imagery," *Atmospheric Measurement Techniques*, vol. 5, no. 11, pp. 2881–2892, 2012.

[77] P. Pinson, "Estimation of the uncertainty in wind power forecasting," Ph.D. dissertation, École Nationale Supérieure des Mines de Paris, 2006.

[78] S. Alessandrini, L. Delle Monache, S. Sperati, and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Applied Energy*, vol. 157, pp. 95–110, 2015.

[79] S. Sperati, S. Alessandrini, and L. Delle Monache, "An application of the ecmwf ensemble prediction system for short-term solar power forecasting," *Solar Energy*, vol. 133, pp. 437–450, 2016.

[80] Y. Chu, M. Li, H. T. Pedro, and C. F. Coimbra, "Real-time prediction intervals for intra-hour dni forecasts," *Renewable Energy*, vol. 83, pp. 234–244, 2015.

[81] H. T. Pedro and C. F. Coimbra, "Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances," *Renewable Energy*, vol. 80, pp. 770 – 782, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148115001792

[82] A. Grantham, Y. R. Gel, and J. Boland, "Nonparametric short-term probabilistic forecasting for solar radiation," *Solar Energy*, vol. 133, pp. 465–475, 2016.

[83] M. David, F. Ramahatana, P. Trombe, and P. Lauret, "Probabilistic forecasting of the solar irradiance with recursive {ARMA} and {GARCH} models," *Solar Energy*, vol. 133, pp. 55 – 72, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X16300172

[84] J.-Y. Le Boudec, *Performance evaluation of computer and communication systems.* Epfl Press, 2011.

[85] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term nonparametric probabilistic forecasting of renewable energy generation with application to solar energy," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3850–3863, Sept 2016.

[86] N. S. D'Souza, L. A. Lopes, and X. Liu, "An intelligent maximum power point tracker using peak current control," in *Power Electronics Specialists Conference, 2005. PESC'05. IEEE 36th.* IEEE, 2005, p. 172.

[87] E. Scolari, D. Torregrossa, J. . Le Boudec, and M. Paolone, "Ultra-short-term prediction intervals of photovoltaic ac active power," in *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Oct 2016, pp. 1–8.

[88] P. Romano and M. Paolone, "Enhanced interpolated-dft for synchrophasor estimation in fpgas: Theory, implementation, and validation of a pmu prototype," *Instrumentation and Measurement, IEEE Transactions on*, vol. 63, no. 12, pp. 2824–2836, 2014.

[89] R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting with exponential smoothing: the state space approach.* Springer Science & Business Media, 2008.

[90] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction intervals for short-term wind farm power generation forecasts," *Sustainable Energy, IEEE Transactions on*, vol. 4, no. 3, pp. 602–610, 2013.

[91] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Management science*, vol. 6, no. 3, pp. 324–342, 1960.

[92] J.-S. Jang, "Anfis: adaptive-network-based fuzzy inference system," *IEEE transactions on systems, man, and cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.

[93] A. C. Davison and D. V. Hinkley, *Bootstrap methods and their application.* Cambridge university press, 1997, vol. 1.

[94] GRASS Development Team, *Geographic Resources Analysis Support System (GRASS GIS) Software*, Open Source Geospatial Foundation, USA, 2015. [Online]. Available: http://grass.osgeo.org

[95] J. Hofierka, M. Suri *et al.*, "The solar radiation model for open source gis: implementation and applications," in *Proceedings of the Open source GIS-GRASS users conference*, 2002, pp. 1–19.

[96] R. Marquez and C. F. Coimbra, "Proposed metric for evaluation of solar forecasting models," *Journal of solar energy engineering*, vol. 135, no. 1, p. 011016, 2013.

[97] S. P. Lloyd, "Least squares quantization in pcm," *Information Theory, IEEE Transactions on*, vol. 28, no. 2, pp. 129–137, 1982.

[98] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis.* John Wiley & Sons, 2009, vol. 344.

[99] J. D. Hamilton, *Time series analysis*, vol. 2.

[100] S. R. West, D. Rowe, S. Sayeef, and A. Berry, "Short-term irradiance forecasting using skycams: Motivation and development," *Solar Energy*, vol. 110, pp. 188 – 207, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X14004150

[101] F. Banterle and L. Benedetti, "Piccante: An open and portable library for hdr imaging," 2014. [Online]. Available: http://vcg.isti.cnr.it/piccante

[102] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[103] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omni-directional cameras," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on.* IEEE, 2006, pp. 5695–5701.

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

[105] M. R. Calabretta and E. W. Greisen, "Representations of celestial coordinates in fits," *Astronomy & Astrophysics*, vol. 395, no. 3, pp. 1077–1122, 2002.

[106] R. W. Johnson, W. S. Hering, and J. E. Shields, "Automated visibility & cloud cover measurements with a solid state imaging system," SCRIPPS INSTITUTION OF OCEANOGRAPHY LA JOLLA CA MARINE PHYSICAL LAB, Tech. Rep., 1989.

[107] Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *Journal of atmospheric and oceanic technology*, vol. 28, no. 10, pp. 1286–1296, 2011.

[108] T. Schmidt, J. Kalisch, E. Lorenz, and D. Heinemann, "Evaluating the spatio-temporal performance of sky-imager-based solar irradiance analysis and forecasts," *Atmospheric chemistry and physics*, vol. 16, no. 5, pp. 3399–3412, 2016.

[109] S. Dev, Y. H. Lee, and S. Winkler, "Color-based segmentation of sky/cloud images from ground-based cameras," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 231–242, 2017.

[110] ——, "Multi-level semantic labeling of sky/cloud images," in *2015 IEEE International Conference on Image Processing (ICIP)*, Sept 2015, pp. 636–640.

## Bibliography

[111] F. Su, W. Jiang, J. Zhang, H. Wang, and M. Zhang, "A local features-based approach to all-sky image prediction," *IBM Journal of Research and Development*, vol. 59, no. 2/3, pp. 6–1, 2015.

[112] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[113] M. Kumar and K. K. Mehta, "A texture based tumor detection and automatic segmentation using seeded region growing method," *International Journal of Computer Technology and Applications*, vol. 2, no. 4, 2011.

[114] S. Dev, Y. H. Lee, and S. Winkler, "Systematic study of color spaces and components for the segmentation of sky/cloud images," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 5102–5106.

[115] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[116] J. F. Trevor Hastie, Robert Tibshirani, *The Elements of Statistical Learning*, S. S. in Statistics, Ed. Springer, 2008.

[117] P. McCullagh and J. A. Nelder, *Generalized linear models*. CRC press, 1989, vol. 37.

[118] S. Sun, J. Ernst, A. Sapkota, E. Ritzhaupt-Kleissl, J. Wiles, J. Bamberger, and T. Chen, "Short term cloud coverage prediction using ground based all sky imager," in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Nov 2014, pp. 121–126.

[119] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International journal of computer vision*, vol. 12, no. 1, pp. 43–77, 1994.

[120] S. Sun, E. Ritzhaupt-Kleiss, and T. Chen, "short term cloud coverage prediction using ground based all sky imager," *IEEE international conference on smart grid communications*, vol. 14, 2014.

[121] R. Marquez and C. F. Coimbra, "Intra-hour dni forecasting based on cloud tracking image analysis," *Solar Energy*, vol. 91, pp. 327–336, 2013.

[122] S. Quesada-Ruiz, Y. Chu, J. Tovar-Pescador, H. Pedro, and C. Coimbra, "Cloud-tracking methodology for intra-hour dni forecasting," *Solar Energy*, vol. 102, pp. 267–275, 2014.

[123] D. Bernecker, C. Riess, A. Elli, and J. Hornegger, "Continuous short-term irradiance forecasts using sky images," *Solar Energy*, vol. 110, pp. 303–315, September 2014.

[124] N. Mori and K.-A. Chang, "Experimental study of a horizontal jet in a wavy environment," *Journal of engineering mechanics*, vol. 129, no. 10, pp. 1149–1155, 2003.

166

[125] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/kanade meets horn/schunck: Combining local and global optic flow methods," *International journal of computer vision*, vol. 61, no. 3, pp. 211–231, 2005.

[126] L. Dunn, M. Gostein, and K. Emery, "Comparison of pyranometers vs. pv reference cells for evaluation of pv array performance," in *2012 38th IEEE Photovoltaic Specialists Conference*, June 2012, pp. 002 899–002 904.

[127] C. Rigollier, M. Lefevre, and L. Wald, "The method heliosat-2 for deriving shortwave solar radiation from satellite images," *Solar Energy*, vol. 77, no. 2, pp. 159 – 169, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X04001082

[128] C. Rigollier, "Vers un accès à une climatologie du rayonnement solaire: estimation de l'irradiation globale à partir d'images satellitales," Ph.D. dissertation, Université Nice Sophia Antipolis, 2000.

[129] Y. Eissa, P. R. Marpu, I. Gherboudj, H. Ghedira, T. B. Ouarda, and M. Chiesa, "Artificial neural network based model for retrieval of the direct normal, diffuse horizontal and global horizontal irradiances using seviri images," *Solar Energy*, vol. 89, no. Supplement C, pp. 1 – 16, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X12004318

[130] B. Kurtz and J. Kleissl, "Measuring diffuse, direct, and global irradiance using a sky imager," *Solar Energy*, vol. 141, no. Supplement C, pp. 311 – 322, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X16305722

[131] J. Alonso-Montesinos and F. Batlles, "The use of a sky camera for solar radiation estimation based on digital image processing," *Energy*, vol. 90, pp. 377–386, 2015.

[132] C. Gauchet, P. Blanc, B. Espinar, B. Charbonnier, and D. Demengel, "Surface solar irradiance estimation with low-cost fish-eye camera," in *Workshop on" Remote Sensing Measurements for Renewable Energy"*, 2012.

[133] T. Schmidt, "High resolution solar irradiance forecasts based on sky images," Ph.D. dissertation, 2017.

[134] T. Schmidt, J. Kalisch, and E. Lorenz, "Retrieving direct and diffuse radiation with the use of sky imager pictures," in *EGU General Assembly Conference Abstracts*, ser. EGU General Assembly Conference Abstracts, vol. 17, Apr. 2015, p. 13552.

[135] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[136] A. R. Smith, "Color gamut transform pairs," *ACM Siggraph Computer Graphics*, vol. 12, no. 3, pp. 12–19, 1978.

## Bibliography

[137] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov 1973.

[138] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, and F. Vignola, "A new operational model for satellite-derived irradiances: description and validation," *Solar Energy*, vol. 73, no. 5, pp. 307–317, 2002.

[139] P. Ineichen and R. Perez, "A new airmass independent formulation for the linke turbidity coefficient," *Solar Energy*, vol. 73, no. 3, pp. 151–157, 2002.

[140] F. Song, Z. Guo, and D. Mei, "Feature selection using principal component analysis," in *System science, engineering design and manufacturing informatization (ICSEM), 2010 international conference on*, vol. 1.   IEEE, 2010, pp. 27–30.

[141] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[142] G. EUMETSAT (Organization). Darmstadt-Eberstadt. Eumetsat :  The european organisation for meteorological satellites. [Online]. Available: www.eumetsat.int

[143] M. Raspaud, D. Hoese, A. Dybbroe, L. O. Rasmussen, P. Lahtinen, m4sth0, M. Itkin, oananicola, praerien, S. Macfarlane, and northaholic, "pytroll/satpy: v0.3.1," Jan. 2017. [Online]. Available: https://doi.org/10.5281/zenodo.252433

[144] J. Kaňák, "Overview of the ir channels and their applications."

[145] P. Blanc, B. Gschwind, M. Lefèvre, and L. Wald, "The helioclim project: Surface solar irradiance data for climate applications," *Remote Sensing*, vol. 3, no. 2, pp. 343–361, 2011.

[146] Soda service, helioclim-3 archive. [Online]. Available:  http://www.soda-pro.com/web-services

[147] C. Thomas, E. Wey, P. Blanc, L. Wald, and M. Lefevre, "Validation of helioclim-3 version 4, helioclim-3 version 5 and macc-rad using 14 bsrn stations," *Energy Procedia*, vol. 91, pp. 1059 – 1069, 2016, proceedings of the 4th International Conference on Solar Heating and Cooling for Buildings and Industry (SHC 2015). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1876610216303757

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

[149] C. Vernay, P. Blanc, and S. Pitaval, "Characterizing measurements campaigns for an innovative calibration approach of the global horizontal irradiation estimated by helioclim-3," *Renewable Energy*, vol. 57, pp. 339 – 347, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148113000876

[150] B. Espinar, P. Blanc, L. Wald, C. Hoyer-Lick, M. Schroedter-Homscheidt, and T. Wanderer, "Controlling the quality of measurements of meteorological variables and solar radiation. from sub-hourly to monthly average time periods," *European Geosciences Union General Assembly 2012*, 2012.

[151] A. Laudani, F. R. Fulginei, A. Salvini, M. Carrasco, and F. Mancilla-David, "A fast and effective procedure for sensing solar irradiance in photovoltaic arrays," in *2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*, June 2016, pp. 1–4.

[152] M. Carrasco, F. Mancilla-David, and R. Ortega, "An estimator of solar irradiance in photovoltaic arrays with guaranteed stability properties," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 7, pp. 3359–3366, 2014.

[153] S. Ghosh, S. Rahman, and M. Pipattanasomporn, "Distribution voltage regulation through active power curtailment with pv inverters and solar generation forecasts," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 1, pp. 13–22, Jan 2017.

[154] V. L. Vigni, D. L. Manna, E. R. Sanseverino, V. di Dio, P. Romano, P. di Buono, M. Pinto, R. Miceli, and C. Giaconia, "Proof of concept of an irradiance estimation system for reconfigurable photovoltaic arrays," *Energies*, vol. 8, no. 7, pp. 6641–6657, 2015.

[155] F. Mancilla-David, F. Riganti-Fulginei, A. Laudani, and A. Salvini, "A neural network-based low-cost solar irradiance sensor," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 3, pp. 583–591, March 2014.

[156] L. Cristaldi, M. Faifer, M. Rossi, and S. Toscani, "An improved model-based maximum power point tracker for photovoltaic panels," *IEEE transactions on instrumentation and measurement*, vol. 63, no. 1, pp. 63–71, 2014.

[157] D. R. Myers, "Solar radiation modeling and measurements for renewable energy applications: data and model quality," *Energy*, vol. 30, no. 9, pp. 1517 – 1531, 2005, measurement and Modelling of Solar Radiation and Daylight- Challenges for the 21st Century.

[158] A. Laudani, F. Mancilla-David, F. Riganti-Fulginei, and A. Salvini, "Reduced-form of the photovoltaic five-parameter model for efficient computation of parameters," *Solar Energy*, vol. 97, pp. 122 – 127, 2013.

[159] A. Astolfi, D. Karagiannis, and R. Ortega, *Nonlinear and adaptive control with applications.* Springer Science & Business Media, 2007.

[160] "IEC60891, Photovoltaic Devices. Procedures for Temperature and Irradiance Corrections to Measured IV Characteristics." IEC60891, 2010.

[161] A. Jain and A. Kapoor, "Exact analytical solutions of the parameters of real solar cells using lambert w-function," *Solar Energy Materials and Solar Cells*, vol. 81, no. 2, pp. 269–277, 2004.

[162] R. Smith, S. Kurtz, and B. Sekulic, "Back-of-module temperature measurement methods," *Solar Pro*, vol. 4, pp. 90–104, 2011.

[163] D. L. King, *Photovoltaic array performance model*.

[164] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[165] J. J. Roberts, A. A. M. Zevallos, and A. M. Cassula, "Assessment of photovoltaic performance models for system simulation," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 1104–1123, 2017.

[166] C. Zogogianni, L. Reyes-Chamorro, M. Paolone, V. Porobic, E. Adzic, and M. Vekic, "Real-time power-reference tracking method for pv converters," in *2017 IEEE Manchester PowerTech*, June.

[167] A. Saltelli, K. Chan, and E. Scott, "Sensitivity analysis wiley series in probability and statistics," *Willey, New York*, 2000.

[168] S. Papathanassiou, N. Hatziargyriou, K. Strunz *et al.*, "A benchmark low voltage microgrid network," in *Proceedings of the CIGRE symposium: power systems with dispersed generation*, 2005, pp. 1–8.

[169] C. Zogogianni, L. Reyes-Chamorro, M. Paolone, V. Porobic, M. Vekicc, and M. Vekic, "Real-time power-reference tracking method for pv converters," in *2017 IEEE Manchester PowerTech*, June 2017, pp. 1–5.

[170] G. T. Costanzo, A. Bernstein, L. E. Reyes Chamorro, H. W. Bindner, J.-Y. Le Boudec, and M. Paolone, "Electric Space Heating Scheduling for Real-time Explicit Power Control in Active Distribution Networks," in *IEEE PES ISGT Europe 2014*, 2014.

[171] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[172] R. Gupta, F. Sossan, E. Scolari, E. Namor, L. Fabietti, C. N. Jones, and M. Paolone, "An admm-based coordination and control strategy for PV and storage to dispatch stochastic prosumers: Theory and experimental validation," *CoRR*, vol. abs/1803.07641, 2018. [Online]. Available: http://arxiv.org/abs/1803.07641

[173] J. Bright, C. Smith, P. Taylor, and R. Crook, "Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data," *Solar Energy*, vol. 115, pp. 229 – 242, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X15001024

[174] R. Müller and J. Trentmann, "Algorithm theoretical baseline document direct irradiance at surface," Tech. rep., EUMETSAT Satellite Application Facility on Climate Monitoring, Tech. Rep., 2010.

[175] T. Klucher, "Evaluation of models to predict insolation on tilted surfaces," *Solar Energy*, vol. 23, no. 2, pp. 111 – 114, 1979. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0038092X79901105

## PERSONAL DETAILS

| | |
|---|---|
| *Birth* | June 7, 1989 |
| *Nationality* | Italian |
| *Phone* | (0041) 789553190 |
| *Mail* | `enrica.scolari@epfl.ch` |

.

## EDUCATION

### Ph.D Candidate in Electrical Engineering
<div align="right">2015-present</div>

*École Polytechnique Fédérale de Lausanne, EPFL, Switzerland.*

- Smart-grid technologies and control of microgrids including storage and photovoltaic;
- Application of machine learning in the energy sector;
- Statistical method and data analysis for solar forecasting.

### Master Degree in Energy Engineering
<div align="right">2012-2014</div>

*University of Bologna, Italy- 110/110 cum laude.*

- Solid knowledge of heat and mass transfer, applied thermo- and fluid-dynamics;
- Innovative electrical technologies;
- Environmental impact of energy systems.

### Bachelor Degree in Energy Engineering
<div align="right">2008-2011</div>

*University of Bologna, Italy- 110/110*

- Basics of maths, physics, and chemistry;
- Basics of thermodynamic and heat transfer;
- Basic knowledge of electric energy production and conversion, energy market, and renewables.

## SCHOLARSHIPS

### Technical University of Denmark, Copenhagen (DTU).
<div align="right">2014</div>

*Master Thesis*

- Multi-objective optimization of a Steam Rankine Cycle as a bottoming cycle for gas turbines in off-shore applications.

### University of California, San Diego (UCSD).
<div align="right">2012-2013</div>

*UCSD GPA: 4.0/4.0*

- Excellence scholarship during the Master Degree Program.

## WORK EXPERIENCE

### Doctoral Assistant

*Distributed Electrical System Laboratory, EPFL, Switzerland.*

- Researcher on the integration of stochastic resources in the electrical distribution grid (photovoltaic and battery storage systems);
- Supervisor of more than 10 master and bachelor students for thesis and semester projects;
- Teaching experience for the Energy Conversion Course at the EPFL.

### Proposal Engineer
2014

*Exergy ORC- Milan, Italy.*

- Proposal engineer to design competitive and efficient solutions based an Organic Rankine Cycles to meet the costumers needs.

### Internship
2013

*UCSD Center for Advanced Radiotherapy Technologies, CART.*

- Evaluation of cone beam CT dose in image guided radiation therapy for brain cancer patients via GPU-based Monte Carlo Simulations.

## SKILLS

| | |
|---|---|
| *Languages* | Italian (mother tongue) |
| | English (fluent) |
| | French (intermediate) |
| *Software* | MATLAB , Python, LabVIEW, Microsoft Office. |

## PUBLICATIONS

- E. Scolari, F. Sossan, M. Haure-Touzé, M. Paolone, *Local estimation of the global horizontal irradiance using an all-sky camera* (Solar Energy, 2018).

- E. Scolari, F. Sossan, and M. Paolone, *Experimental Comparison of Maximum Power Estimators for a Single Unit Photovoltaic Plant* (35th EU PVSEC Conference, Brussels 2018).

- E. Namor, F. Sossan, E. Scolari, R. Cherkaoui, M. Paolone, *Experimental Assessment of the Prediction Performance of Dynamic Equivalent Circuit Models of Grid-connected Battery Energy Storage Systems* (IEEE PES ISGT Europe, Sarajevo, October 21-25, 2018).

- R. Gupta, F. Sossan, E. Scolari, E. Namor, L. Fabietti, C. Jones, M. Paolone, *An ADMM-based Coordination and Control Strategy for PV and Storage to Dispatch Stochastic Prosumers: Theory and Experimental Validation* (PSCC XX Power Systems Computation Conference, Dublin 2018).

- E. Scolari, L. Reyes-Chamorro, F. Sossan, M. Paolone, *A Comprehensive Assessment of the Short-Term Uncertainty of Grid-Connected PV Systems* (IEEE Transaction on Sustainable Energy, 2018).

- L. Reyes-Chamorro, A. Bernstein, N. J. Bouman, E. Scolari, A. Kettner, B. Cathiard, J.-Y. Le Boudec and M. Paolone, *Explicit Real-Time Control of Power Flows in Microgrids* (IEEE Transaction on Industrial Informatics, 2018).

- E. Scolari, F. Sossan, M. Paolone, *Photovoltaic Model-Based Solar Irradiance Estimators: Performance Comparison and Application to Maximum Power Forecasting* (IEEE Transaction on Sustainable Energy, 2017).

- L.Magnone, F. Sossan, E.Scolari, M. Paolone, *A Heuristic Cloud Motion Detection Algorithm Based on All-Sky Images to Support Solar Irradiance Forecast* (44th IEEE PVSC Photovoltaic Specialists Conference, Washington DC, 2017).

- E. Scolari, F. Sossan, M. Paolone, *A Model-Based Filtering Strategy to Reconstruct the Maximum Power Generation of Curtailed Photovoltaic Installations: application to forecasting* (IEEE PES PowerTech Conference, Manchester 2017).

- E. Scolari, F. Sossan, and M. Paolone, *Irradiance prediction intervals for PV stochastic generation in microgrid applications* (Solar Energy, 2016).

- E. Scolari, D. Torregrossa, J. Y. Le Boudec, and M. Paolone, *Ultra-Short-Term Prediction Intervals of Photovoltaic AC Active Power.* (International Conference on Probabilistic Methods Applied to Power Systems, Beijing, China, 2016),

- L. Pierobon, A. Benato, E. Scolari, F. Haglind, A. Stoppato, *Waste heat recovery technologies for offshore platforms* (Applied Energy, 2014).

- D. Montanari, E. Scolari, C. Silvestri, Y.J. Graves, H. Yan, L. Cervino, X. Jia. *Comprehensive evaluations of cone-beam CT dose in image-guided radiation therapy via GPU-based Monte Carlo simulations* (Physics in Medicine & Biology, 2014).

## AWARDS

Best Poster Award at the 4th SCCER-FURIES Annual Conference (2017) for the work titled *All-sky Images-based Irradiance Forecast.*

## RESEARCH INTERESTS

- Machine learning applied to energy systems;
- Renewable sources and related forecasting methods;
- Solar energy;
- Data analysis.

## PERSONAL INTERESTS

My greatest passions are travelling and reading since through them I can leave my routine and challenge my-self in different environments and situations. In my free time, I volunteer for an Italian no-profit association.