UNIVERSITY OF LJUBLJANA

Faculty of Electrical Engineering

Albert Gili Selga

Modelling photovoltaic system for a home energy management control

Master Thesis

Mentor: prof. dr. Marko Čepin

Ljubljana, 2021

Acknowledgments

This master thesis could not be possible without the student exchange program known as Erasmus+ and the agreement between my host and receiving institution: *Universitat Politècnica de Catalunya* (UPC) and *Univerza v Ljubljani* (UL), respectively.

The collaboration with the *Institut Jozef Stefan* (IJS) and COMSENSUS, where I have worked as a student practice, has also been crucial for the development of it.

Additionally, I want to make a special mention to Marko Čepin, who has been my tutor, for all the support and the guidelines given along the thesis.

Finally, I would like to give special recognition to my family and friends who have helped and supported me in difficult times. Without them, the thesis would not be possible either.

Table of contents

List of figures

List of tables

List of symbols used

Symbol	Unit	Meaning
$c.f._i$	ϕ	correction factor
d	day	day
Div_i	ϕ	ratio of predicted power with historical power
$G_{f,i}$	W/m^2	irradiance forecasted
$G_{horizon,i}$	W/m^2	irradiance received on the panel considering the horizon of the location
$G_{no\_horizon,i}$	W/m^2	irradiance received on the panel without considering the horizon of the location
G_{NOCT}	W/m^2	irradiance in normal operating cell conditions test
G_{STC}	W/m^2	irradiance in standard test conditions
lat	º	latitude
lon	º	longitude
$NOCT$	ºC	normal operating cell temperature
$P_{h,i}$	W	historical power
P_{nom}	W	nominal power
$P_{out,i}$	W	power output
$P_{p,i}$	W	predicted power
SL	ϕ	system losses
$T_{amb,i}$	ºC	ambient temperature
T_{NOCT}	ºC	temperature in normal operating cell conditions test
$T_{cell,i}$	ºC	panel's temperature
T_{STC}	ºC	temperature in standard test conditions
t_{UTC}	hour	coordinated universal time
W_i	ϕ	weight parameter
α	º	elevation angle
β	º	tilt of the panels
γ	$ºC^{-1}$	photovoltaic temperature coefficient
δ	º	declination
η_{system}	ϕ	efficiency of the system
θ_{AOI}	º	angle of incidence
θ_{az}	º	solar azimuth
θ_{az_panel}	º	azimuth of the panels
θ_z	º	solar zenith
ρ	ϕ	albedo coefficient
ω	º	hour angle
σ	ϕ	variance

List of acronyms used

Acronym	Meaning
AOI	Angle of Incidence
DER	Distributed Energy Resources
DHI	Diffuse Horizontal Irradiance
DNI	Direct Normal Irradiance
EMS	Energy Management System
EoT	Equation of Time
GHI	Global Horizontal Irradiance
GHI	Global Horizontal Irradiance
GTI	Global Tilted Irradiance
HEMS	Home Energy Management System
HVAC	Heating, Ventilating and Air Conditioning
JSON	JavaScript Object Notation
LST	Local Solar Time
LSTM	Local Standard Time Meridian
LT	Local Time
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
NOCT	Normal Operating Cell Temperature
POA	Plane of Array
PV	Photovoltaic
R2	Coefficient of determination
RMSE	Root Mean Squared Error
STC	Standard Test Conditions
TMY	Typical Meteorological Year
UTC	Coordinated Universal Time

Abstract

Electrical energy demand is expected to increase in the upcoming years making more unsustainable the traditional ways to get it. For this reason, the world is increasing the share of renewable energy. However, the energy transition comes with some challenges due to the uncertainties of the natural resource grid as wind and solar. This change of paradigm is a reality for big-scale systems like the electricity grid and small-scale systems like a Home Energy Management System (HEMS).

In essence, the objective of the thesis is to design a forecasting photovoltaic (PV) model versatile and generic for being able to adapt to the specifications of every system under investigation. The model is designed by considering and using parametric and non-parametric methods. This type of modelling is popularly known as grey-box modelling. The core of the model is defined by physics equations, so it is the parametric part of the model. However, this part has some assumptions and idealizations with it. For compensating them, two data-driven corrections are done following non-parametrical techniques. Therefore, the model takes advantage of both procedures.

In order to validate the model, two case studies that consist of two PV systems located in different countries are defined. From their historical power output data and the forecast made by the PV model, it is possible to calculate their mismatch and verify the design model. In the end, the calculations show that the model is robust when it comes to forecasting even if the systems are located at different points, but it loses accuracy on those days that the sky is fully covered.

Resum

Es preveu que la demanda d'energia elèctrica augmenti en els propers anys fent més insostenibles les formes tradicionals de obtenir-la. Per aquesta raó, el món està augmentant la presencia d'energia renovable instal·lada. No obstant, la transició energètica ve acompanyada d'alguns, elèctrica degut a les incerteses de fonts naturals com la generació d'energia eòlica i la solar. Aquest canvi de paradigma és una realitat tant per als sistemes a gran escala, com la xarxa elèctrica, com per als sistemes a petita escala, com un sistema de gestió de l'energia a la llar (HEMS).

L'objectiu d'aquesta tesis és dissenyar un model de previsió fotovoltaica (FV) versàtil i genèric per poder adaptar-se a les especificacions de cada sistema fotovoltaic a estudiar. El model es dissenya considerant i utilitzant mètodes paramètrics i no paramètrics. Aquest tipus de disseny és popularment conegut com modelització de caixa gris. El nucli del model està definit per equacions físiques, pel que és la part paramètrica de el model. No obstant, aquesta part porta amb si algunes suposicions i idealitzacions per no afegir dificultats innecessàries al model. Per compensar-les, es realitzen dues correccions a partir de dades disponibles del sistema seguint tècniques no paramètriques. Per tant, el model aprofita i busca els avantatges d'aquests dos mètodes.

Per validar el model, es defineixen dos casos d'estudi que consisteixen en dos sistemes fotovoltaics situats a diferents països. A partir de les seves dades històriques de producció d'energia i de la previsió realitzada pel model fotovoltaic, és possible calcular el seu desajust i verificar el model de disseny. Al final, els càlculs mostren que el model és robust quan es tracta de pronosticar encara que els sistemes estiguin situats en punts diferents, però perd precisió en els dies en què el cel està totalment cobert.

1. Introduction

1.1 Background

Over the last centuries, the world's population has increased, and along with it, energy production and consumption. Effects of the massive emissions of greenhouse gases, and the limited natural sources, such as oil, have evidenced an imperious necessity of a transition to new energy sources.

Renewable energy systems have the main role in this energy transition. Especially, the generation system based on photovoltaic power has a significant impact as it is developing rapidly in the energy sector. This progression appears in Figure 1.1 which shows the cumulative PV systems capacity installed for the last years [1].

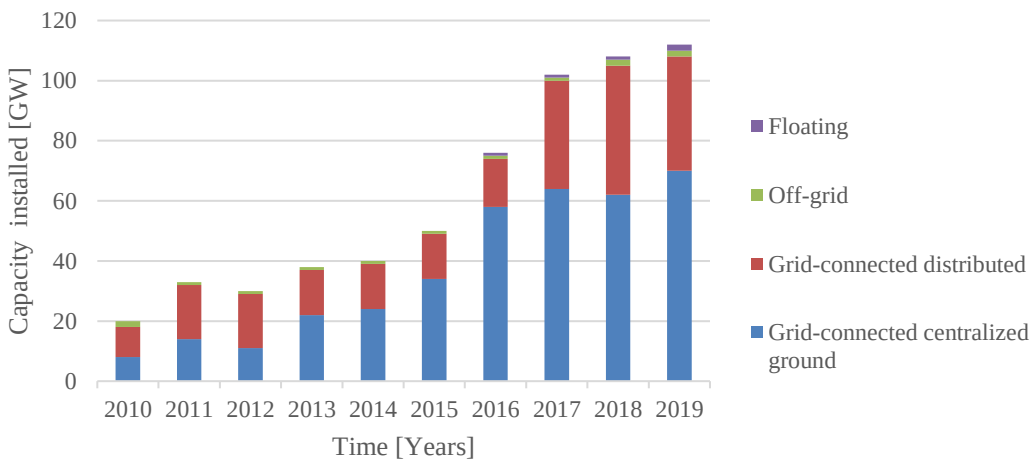


Figure 1.1: Cumulative PV systems installed in the world by years

The energy transition is not something trivial and brings some new issues that have to be solved. Contrary to traditional electrical lines, distributed grids are more likely to congest. One of the main reasons is the appearance of the prosumer (end-users that produce and consume electricity) that adds new complex scenarios to predict. However, this congestion problem can be solved with good management of the prosumer's system. Furthermore, a proper management system can be converted into valuable assets for the electrical system since they can adapt their consumption thanks to their flexibility.

Therefore, not only it is important to increase the presence of renewable energy-based systems, but also the management of these systems, more known as energy management systems (EMS).

1.2 Objectives

The objective of this thesis is the creation of one model capable of behaving as a photovoltaic system (PV). The following features are set to realize the objectives and to evaluate the work progress:

- Develop a PV forecasting model with *python* language.
- The model has to be generic. In other words, the goal is to define a model capable of forecasting the output for a PV system located in any part of the world, does not matter whether it is in Spain or Slovenia for instance.
- The model has to be adaptable and flexible to any system and its available data.
- The model has to forecast the power output of the study systems for one day and with a frequency of 15 minutes.
- Verify the proper performance of the model with case studies. In the end, it poses challenges for future work.

1.3 Scope of the thesis

In this thesis, a PV forecasting model will be built for a home energy management system (HEMS). For this purpose, computer-aided tools and open data are used for its development and testing. *Python* is the programming language adopted, which comes with large software libraries such as *pandas* that works well for data treatment.

Although the model is designed for working within a HEMS, it is out of the thesis scope to create one and integrate the model in it. HEMS is described in order to understand what kind of PV model is needed.

Since the data used for the model's testing is from open sources, it is not possible to check their veracity, or neither how the measurements were taken. It is out of the thesis scope the discussion of their quality as well.

2. Theoretical background and overview of the literature

This section wants to explain what is specifically Home Energy Management Systems (HEMS) and why are they a decisive issue for flexibility. Also, it aims to explore literature regarding the system modelled forthcoming in the report.

2.1 HEMS

As mentioned in 1.2, the main goal of this thesis is to implement one tool capable of forecasting the outputs of any PV system, so it can be part of a HEMS.

2.1.1 Concept of HEMS

HEMS is a demand response tool that shifts and curtails demand to improve the energy consumption and production profile of a house according to electricity price and consumer comfort. It can communicate with household devices, and receive external information (like electricity prices) to improve the energy consumption and production schedule of household devices [2].

Therefore, HEMS may be defined as a home that offers energy management services for efficient monitoring and management of electricity generation, power conservation, as well as energy storage methods [3].

2.1.2 HEMS architecture

In order to achieve its purpose, HEMS counts with internal detectors and actuators, which are sending information to a controller regarding the behaviour and the forecast of loads that has the house, as it illustrates in the diagram of Figure 2.1.

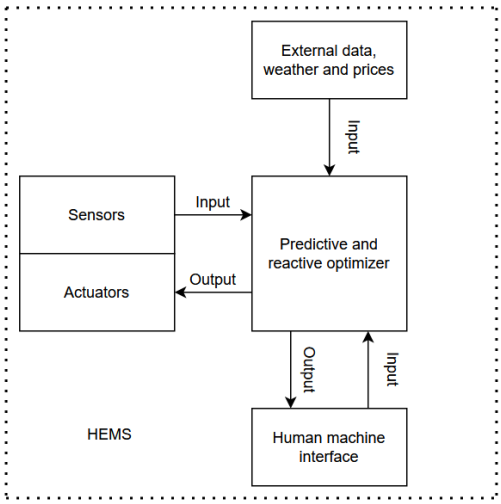


Figure 2.1: General schema of a HEMS

So, HEMS is a total of smaller systems or components working for a specific objective. The most common ones are:

- Typical house appliances: fridge, washing machine.
- Sensors: smart meters, thermometers.
- Actuators: smart plugs, valves.
- Controllers: They can act locally, or use a gateway that allows the network between HEMS and the outside world, to facilitate remote access via Internet [4].
- Distributed Energy Resources (DER): PV system, wind power generation system.
- Batteries.

Figure 2.2 shows a schematic design of a typical HEMS architecture and how each part is connected with another.

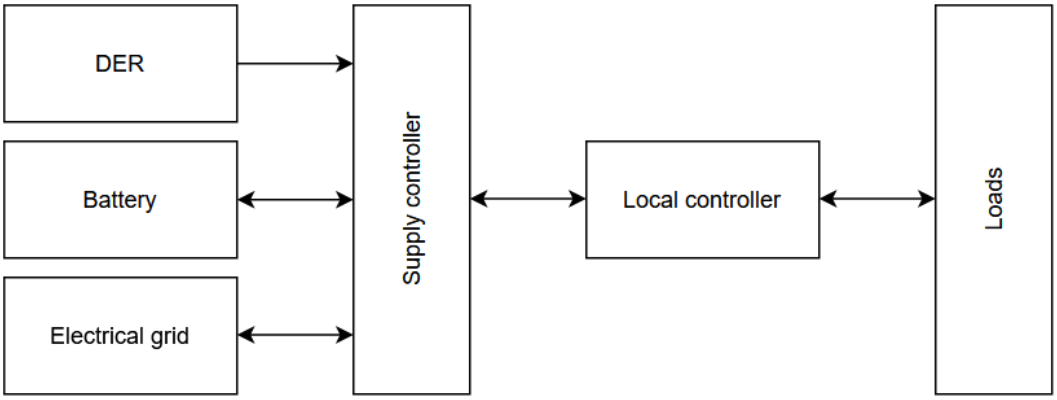


Figure 2.2: Common HEMS architecture

Depending on which components have the architecture of the HEMS, how they are connected, and how the controller is programmed, HEMS can provide different energy

services. In [5], up to eleven different services have been distinguished. Table 2.1 defines each one of them:

Table 2.1: Services of different architectures

Services	**Definition**
Energy management service	It serves to improve the optimisation of energy use and the efficiency of electricity use in the home. To provide it, end device control and metering of energy consumption are required.
Renewable energy system management service	The HEMS is connected to the DER and the battery management system. It allows the management of the electrical energy generation system used in the home.
Energy storage management service	It consists of storing energy directly or indirectly. It is envisaged within renewables and DER. The energy storage system would benefit from generation to the consumer, but the resources need to be properly battery capacity for storage. It responds to real-time prices. When the electrical load is low, the electrical energy is stored and vice versa.
Energy Information service	It provides data information containing the status of energy storage, the amount of electrical energy consumed by end devices, and the electrical energy generated by DER.
Premise energy status display service	It provides information on the energy status and the amount of energy consumption to the consumer by displaying the energy status of devices operating at a consumer's premises. The consumer can save energy and control devices based on it.
Remote monitoring and control service using a handheld device	It allows remote monitoring and control by the consumer using the personal computer and smartphone from the outside.
Home grid alarm service	It indicates the abnormal status of the home network through effective monitoring of devices. It can detect electricity theft, electrical short circuit, and price signal in real-time.
Condition monitoring of home appliance service	It provides on/off switching of existing appliances and controls the different operating modes according to the specification of each appliance.
Home appliance management service	It provides a periodic check of the appliances and can check their physical condition. It can check if there is a fault in the appliance and the replacement and repair cycle.
Plug-in electric vehicle and battery management service	It provides the status information of the battery connected to the battery management system of electric vehicles and HEMS.
Real-time pricing related service	It is used to effectively respond to the real-time price mechanism of the main grid. This service allows electricity to be stored in the reservoir during the period of relatively lower prices, and then the stored electricity must be used during the period of higher main grid electricity prices.

2.1.3 Functionalities of HEMS

Each HEMS model can have a different purpose: that is to say, different objective functions depending on what is sought to minimize or maximize. Despite the infinite services that they can provide; any HEMS should accomplish the following five specific functionalities [3]:

- Monitoring: The monitoring process makes real-time information on the energy use pattern accessible. It facilitates energy utilization and makes the user focus on electricity conservation. It can provide a visualization of the services for several of its operating modes along with the energy status of each of the household devices.
- Logging: is the process of collecting and storing information on the data relating to the unit of electricity consumed by each appliance, generated from DERs and energy conservation states. This functionality includes demand response analysis for real-time prices.
- Control: it can be divided into direct control and remote control. Direct control can be defined as the control that is exercised over both the electronic device and the control system, while remote controls mean that customers have access to, monitor, and control the consumption patterns of their electronic devices through other systems and devices.
- Management: is one of the main functions of HEMS, which improves both the optimization and efficiency of electricity utilization within the home. It can also encompass operations including renewable energy system management service, energy storage management service, home appliance management service, and plug-in electric vehicle and battery management service.
- Alarm: Alarms are generated here and transmitted to the intelligent HEMS control panel, which contains information on fault locations, types, etc.

2.1.4 Forecasting in a HEMS

As seen so far, HEMS is an interdisciplinary tool that provides different services, depending on how it is architected and what its objective function is. One of these services might be the forecast of some of the parts that form HEMS. However, it does not mean that all of them count with this specific service.

Forecasting is interesting to manage those systems that, along with diversification within various devices and goals to be attained, there is a dominant issue of uncertainty. It is the case of houses that include PV panels, Battery Energy Storage (BES), among others. Besides, forecasting is interesting for defining the behaviour of the occupants of the dwelling, and thus, predicting their energy use patterns [5]. Many works with HEMS can be developed focusing on the scheduling and forecasting provided through the external source.

Nowadays, a wide range of forecasting techniques, such as machine learning (ML) or regression techniques, can be included in HEMS. The level of accuracy of forecasts depends to a large extent on the procedures used and the methods of error measurement [6].

Before implementing a forecast model in HEMS, some facts should take into account. It may be advantageous for HEMS since it can coordinate in parallel DER systems and appliances accordingly to what will happen. However, it comes with the cost of designing a scheduling

algorithm apart from the forecast model. In article [7], a forecasting model is applied in a HEMS, and the prosumer's bill of the case study sees a reduction by 16-25% of his total bill, where is clear that forecasting supports the HEMS.

2.2 PV system

As it is well known, a photovoltaic system is a power system that obtains electricity from sunlight cause of the properties of the semiconducting materials, without any intervening heat engine or rotating equipment (conventional techniques). Therefore, the nature of a PV system is closely linked to the solar irradiance it receives and the relative position of the Sun from its location.

Solar PV technology has become one of the most mature and rapidly evolving renewable energy technologies and is expected to play an important role in the necessary energy transition for cleaner energy sources. Consequently, for the last years, PV systems have seen rapid global growth not only for research but also in the market. According to International Energy Agency (IEA) [8], Solar PV generation increased by 22% in 2019 and represented the second-largest absolute generation growth of all renewable technologies, slightly behind wind and ahead of hydro. Moreover, it is expected to grow even more for a sustainable development scenario as can be seen in Figure 2.3.

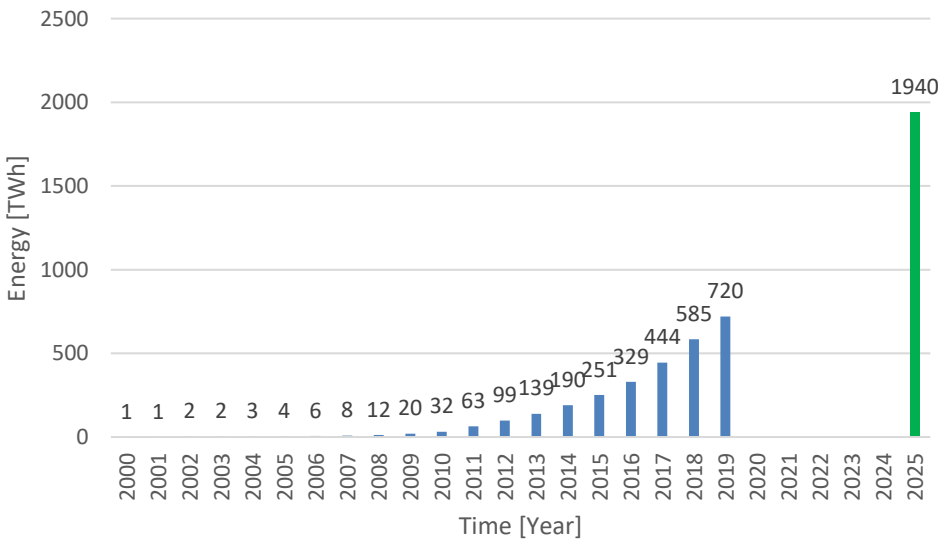


Figure 2.3: Solar PV generation growth and expectations for a sustainable development scenario

This market growth and the maturation of this technology have decreased the price of the PV module due to the increasing scale of production and ongoing manufacturing innovations, which is significant since it is half of the price of the whole PV system [9]. Since the growth of the installation of the PV systems and their proliferation, there has been a diversification among them.

2.2.1 Type of systems

PV systems can be classified according to various considerations, the most common ones are two: classifying by their applications and by their cell materials and structure also known as solar cell technologies.

2.2.1.1 Application types

There are three main types of solar PV systems by considering their application [10] and they may represent as Figure 2.4:

- **Grid-tied solar PV system:** This is a basic solar installation with a standard grid-connected inverter. The grid-connected solar PV system does not have a battery bank for storage. From a grid-tied system, energy can be generated and used only during the day. This system is very cost-effective, simple to design, easy to manage, and requires less maintenance. The main objective of the grid-connected system is to reduce the energy bill.
  The disadvantage of a direct-to-grid system is that it can only be used during the day. The energy cannot be stored for future use, such as during power outages. However, this disadvantage can be overcome by using a battery bank, but this configuration will eventually increase the system's cost.
  These PV systems often produce more electricity than the loads need. Therefore, this excess electricity is fed back into the grid instead of being stored in batteries.
- **Standalone or off-grid solar PV system:** Off-grid solar PV systems use batteries for storing the generated electricity during the daytime for use in the future or during any emergency case like a cloudy day or night. Sometimes such systems have back-up generators if they have to face long periods without seeing the Sun. The advantage of this system is that it gives sufficient energy to a household, and it can power places that are far away from the grid. Off-grid systems are comparatively costlier and expensive than grid-tied solar PV systems.
- **Hybrid PV system:** This system gives the advantages of both grid-direct and off-grid systems. It can help in lowering the customers' utility bills through widespread incentives. Furthermore, these systems can serve as a back-up to the grid in the event of a power failure, as the power from the battery bank can be used during such an emergency.

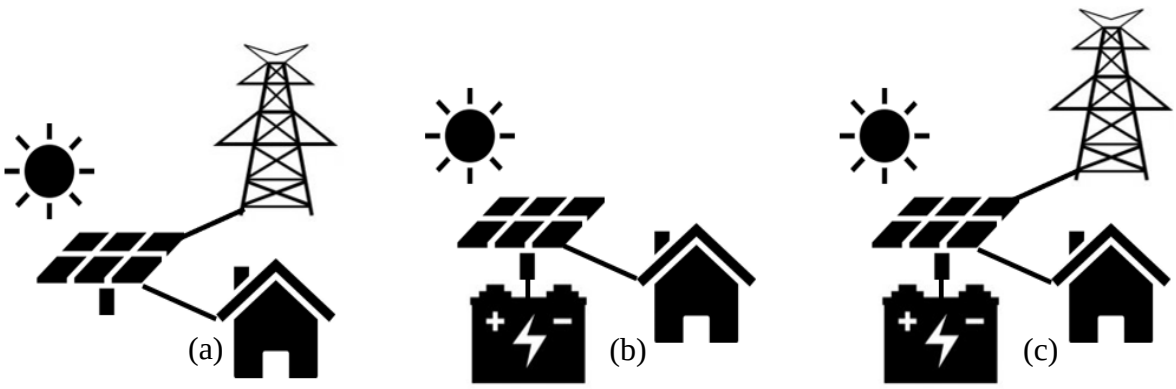


Figure 2.4: (a) Grid-tied PV system (b) Off-grid PV system (c) Hybrid PV system

2.2.1.2 Solar cell technologies

Solar cell technologies are broadly classified into wafer-based crystalline silicon solar cell technology, thin-film solar cell technology, and other new emerging technologies, Figure 2.5 shows schematically their classification. Of the existing solar PV technologies, there are two basic commercial PV module technologies available on the market nowadays that are used by the solar PV industry [9] [11]:

- Wafer-based Silicon panels: Solar cells made of crystalline Silicon in the form of single or polycrystalline wafers:
  - Monocrystalline: these panels have a uniform crystalline structure, and have the highest efficiency ratings to date, and perform better than other types of panels in low light conditions. They have the highest initial cost; however, the energy savings over time can make the cost worthwhile.
  - Polycrystalline: The silicon used in these panels is not homogenous; which means that the crystal structure can be different in various areas of the panel. As a result, polycrystalline solar panels are less efficient than monocrystalline solar panels. However, they are less expensive than the monocrystalline ones.
- Thin-film panels: typically incorporate very thin layers of photovoltaic active material placed on a glass or metal substrate using vacuum deposition manufacturing techniques similar to those used in the architectural glass coating. Thin-film solar panels are less efficient than wafer-based silicon panels and have a shorter lifetime. However, their costs are much lower due to simple manufacturing methods compared to crystalline solar panels. Thin-film solar panels can also be made flexible, whereas crystalline solar panels are much more fragile and will crack if bent. Different semiconductors can be used with this technology:
  - Amorphous Silicon
  - Compound Semiconductors: CdTe, CIS.

On the other hand, there are emerging technologies that still are not economically feasible in the PV market, but in the not distant future, they may have an important role in it [9]. An example of these technologies is an Organic PV cell that uses conductive organic polymers or small organic molecules for getting electricity by the photovoltaic effect.

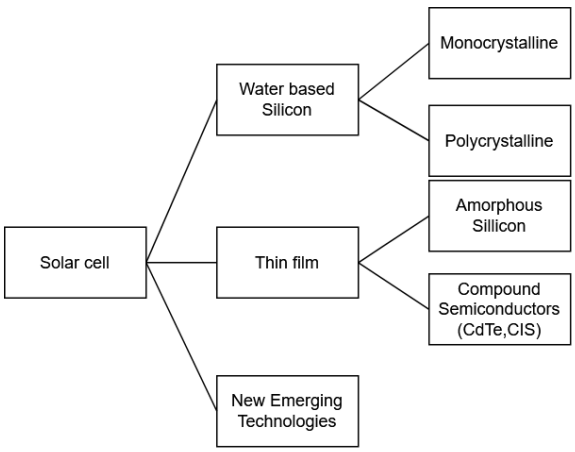


Figure 2.5: Solar cell technologies classification

2.2.2 Components and their losses

Although there are such different systems that can be classified by different assumptions like it has been seen in section 2.2.1, PV systems usually have the same components for converting the raw sunlight to electricity either directed current (D.C.) or alternating current (A.C.) shaped depending on the function of the photovoltaic system. A basic PV architecture looks like Figure 2.6, and its components are [12]:

- PV arrays: The responsible for converting the solar irradiance to D.C. electricity by the photovoltaic effect.
- Charge regulator: It protects against overcharge and deep discharge to the battery or the load of the system.
- Batteries: The responsible for storing the energy generated by a solar array. It is the component that gives more flexibility to the system.
- Inverter: The responsible to converts D.C. electricity to A.C. electricity. It can be used or not, depending on the nature of the load.
- Cables: The responsible for the energy transportation and connection between the other components.
- Load: It is the part of the system which consumes the active energy produced and transported. The load defines which capacity of PV is necessary to install in order to satisfy its demand, and whether the inverter is necessary or not depending on its nature.

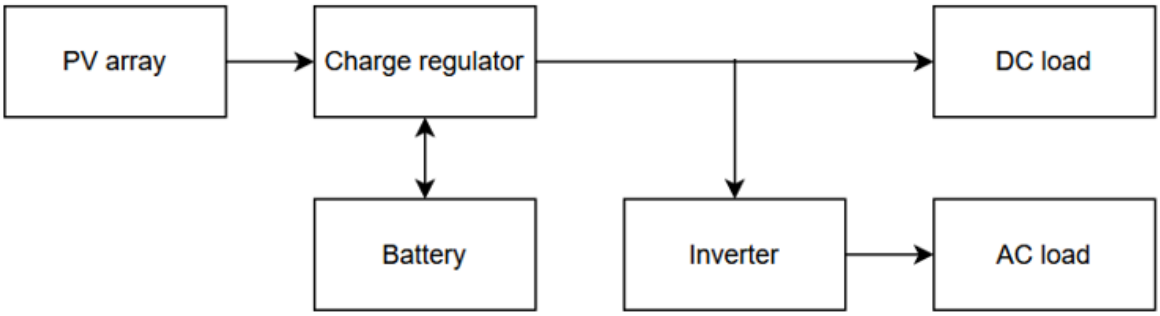


Figure 2.6: Basic PV system architecture and its connections

The PV systems have inherent losses due to the components that constitute them. Firstly, the charge regulator is responsible for the availability of the system and even can regulate the quantity of the energy to transform for the system.

PV arrays have intrinsic losses for their solar cell technologies and their operating field conditions, such as soiling or shading are impeding the solar irradiance to reach the panel's surface or which is the temperature of the cell.

The cables also present losses while they are transporting the electricity due to the Joule effect which part of the electricity is transformed into heat unintentionally, also there are losses in the connections between the components.

Inverters are not an exception, and part of the electricity that transforms is lost, each manufacturer defines their efficiency. Figure 2.7 shows how the initial power obtained is

being lost when the electricity is passing through the system due to the main reasons for the PV losses.

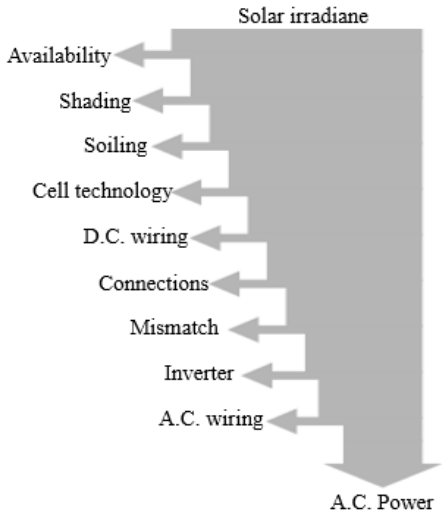


Figure 2.7: Common system losses of a PV system

Components are not the only ones responsible for PV system losses. The degradation or aging of the system, which is the gradual loss of performance over many years, is another kind of loss that affects the system output power.

Aging between the PV modules deforms the electrical characteristics of the whole PV array, resulting in mismatch power losses, also considered in Figure 2.7. The main problem of this degradation cause uneven distribution over time. Indeed, non-uniform aging causes mismatching among PV modules, while mismatching leads to faster aging, which is a common issue in PV systems [13]. Being non-uniform losses makes it more difficult to forecast a PV system performance.

2.2.3 Solar irradiance and Sun position

Irradiance is a primary input for a PV performance model and nearly always is the largest source of uncertainty in it. The cause is all the number of factors that can affect the total irradiance that reaches the system. Such as are atmospheric variability, non-normal incidence, soiling, shading, wind effects, among others. All of them make it almost impossible to define the output of the system with precision. Otherwise, with uniform light conditions, the PV system responds quite predictably. According to Joshua S. Stein, there are three options for getting irradiance values [14]:

- Typical Meteorological Year (TMY): Irradiance is mostly modelled from other measurements for the same given geographical location. Commonly, it has data from a long period, ten years or more, according to the Joint Research Centre, and the TMY selects or approximates the irradiance more typical for a given time and location.
- Satellite modelled data: Irradiance is an indirect measurement. It has the advantage that it can be obtained from everywhere.

– Ground measurements from the site: Irradiance is a direct measurement. It is accurate if it is maintained, but it is ambitious to think that it is for any location.

When working with irradiances is required, some astronomical facts regarding the Earth movement and Sun are necessary to keep in mind. Earth, as is well known, has two main motions: translation and rotation. Both of them affects directly to the received irradiances over time. Since the orbit of the Earth has an elliptical shape, the total extra-terrestrial irradiance can modify between 6% or 7% of its total [14]. Furthermore, extra-terrestrial irradiance is also affected by Earth's rotation due to is tilted. From these two facts, one key parameter known as declination angle is calculated with Equation (2.1) [15]. This angle represents the angle between the solar arrays and the equator plane, and it changes over the year depending on which day (d) is [16], as illustrated in Figure 2.8.

$$\boldsymbol{\delta = -23.45 \cdot \cos\left(\frac{360}{365} \cdot (d + 10)\right)} \tag{2.1}$$

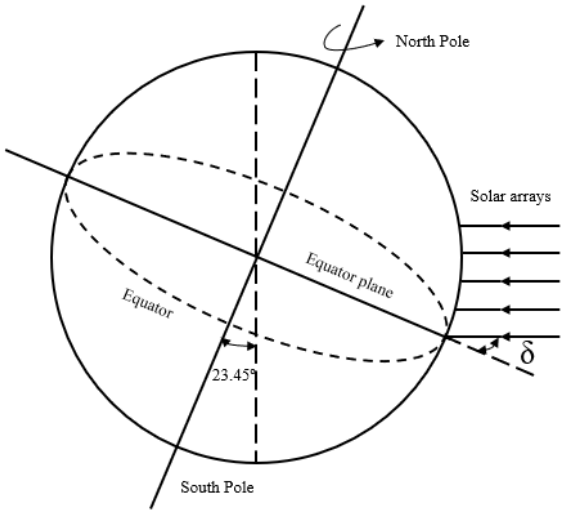


Figure 2.8: Representation of the declination angle

Another parameter that helps to determine Sun's relative position from an Earth's location is the hour angle which is the angle measured from the zenith line to the horizon. It is zero at noon, and with a positive value throughout the morning. At the sine shine, it equals 90º, and it is distributed equally with the hours of the day half. The same graduating afternoon but in negative values [17]. It can be represented in Figure 2.9 and described in Equation (2.2) where LST is the Local Solar Time.

$$\boldsymbol{\omega = 15 \cdot (LST - 12)} \tag{2.2}$$

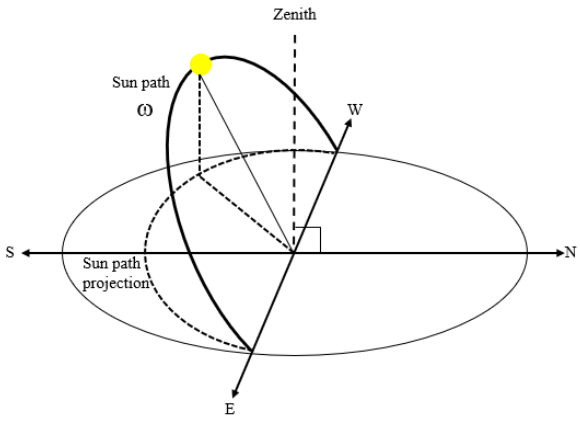


Figure 2.9: Representation of the hour angle (ω)

LST can be found by using corrections to adjust the Local Time (LT). One of these corrections is the equation of time (EoT) which is an empirical equation that corrects for the eccentricity of the Earth's orbit and the Earth's axial tilt, due to some disturbances in Earth rotation as predicted by Kepler's laws. An approximation accurate to within ½ minute expressed is Equation (2.3) [18]:

$$\boldsymbol{EoT = 9.87 \cdot \sin(2 \cdot B) - 7.53 \cdot \cos(B) - 1.5 \cdot \sin(B)} \tag{2.3}$$

$$\boldsymbol{B = \frac{360}{365} \cdot (d - 81)} \tag{2.4}$$

With the EoT, the LT, and the Local Standard Time Meridian (LSTM), which is a reference meridian used for a particular time zone and can be calculated with the LT and the UTC by Equation (2.5), Time Correction Factor (TC) can be calculated, and with it can be obtained LST as it has been said before. The TC (in minutes) accounts for the variation of the LST within a given time zone due to the longitude variations within the time zone and also incorporates the EoT above [19].

$$\boldsymbol{LSTM = 15 \cdot (LT - t_{UTC})} \tag{2.5}$$

$$\boldsymbol{TC = 4 \cdot (lon - LSTM) + EoT} \tag{2.6}$$

$$\boldsymbol{LST = LT + \frac{TC}{60}} \tag{2.7}$$

With all the previous equations, it is possible to determine Sun's position over a day. It can always be expressed, referred to as an observation point, by a pair of angles: the elevation angle (α) and the azimuth angle (θ_{az}). The elevation angle, or also known as altitude angle, is the angular height of the Sun in the sky measured from the horizontal. It can go from 0º

to 90º when the Sun is in the noon. Whereas the azimuth angle is the compass direction from which the sunlight is coming. It can have a range of 180º. Both of them are defined in Equations (2.8) and (2.9) [19] and represented in Figure 2.10.

$$\boldsymbol{\alpha = x + lat - \delta} \begin{cases} \boldsymbol{x = 90^{\circ}\ northern\ hemisphere} \\ \boldsymbol{x = 180^{\circ}\ southern\ hemisphere} \end{cases} \tag{2.8}$$

$$\boldsymbol{\cos(\theta_{az}) = \frac{\sin(\delta) \cdot \cos(lat) - \cos(\delta) \cdot \sin(lat) \cdot \cos(\omega)}{\cos(\alpha)}} \tag{2.9}$$

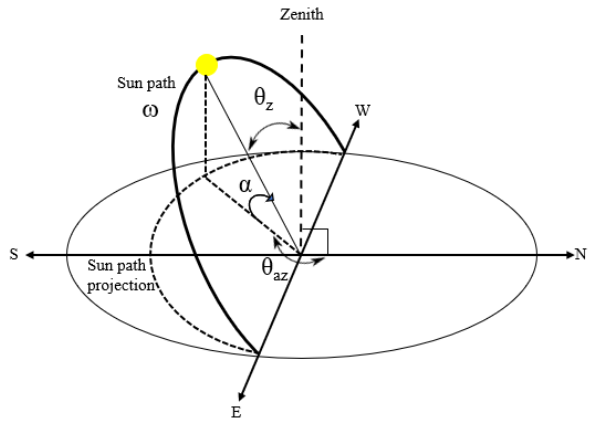


Figure 2.10: Representation of the Elevation, Azimuth, and Zenith angle

With the solar position calculated, it is possible to determine how much irradiance is the surface of the PV panel receives.

2.2.4 Solar irradiance on a tilted surface

The quantity of short-wave radiant energy emitted by the sun passing through a unit horizontal area is referred generally to as global solar irradiance. This incoming solar irradiance is modified as it travels through the atmosphere and again once it hits the surface. So the three main components of solar irradiance are the direct, the diffused, and the reflected solar radiation [20]:

- Direct Normal Irradiance (DNI): the amount of irradiance is intercepted unimpeded, in a direct line from the sun.
- Diffuse Horizontal Irradiance (DHI): is that irradiance scattered by atmospheric constituents, such as clouds and dust.
- Ground-Reflected Irradiance or albedo: It is the irradiance that is reflected from surface features. Each surface has its coefficient albedo (ρ), which is the proportion of light reflected from the total light received by the surface.
- Global Horizontal Irradiance (GHI): It is the sum of DNI, DHI, and ground-reflected radiation. However, because ground-reflected radiation is usually insignificant compared to direct and diffuse, for all practical purposes global radiation is said to be the sum of direct and diffuse radiation only using the zenith angle, which is the elevation complementary angle, also represented in the previous Figure 2.10.

$$\boldsymbol{GHI = DHI + DNI \cdot \cos(90 - \alpha)} \tag{2.10}$$

PV panels generate electricity with any irradiance incident on its plane of the array (POA), regardless of its nature. This total amount of irradiance is called Global Tilted Irradiance (GTI). It is a reference for PV applications and can sometimes be affected by shading. GTI can be measured, but as it is not easy to measure for every system, it is often modelled from DNI, DHI, and ρ adjusting its values according to the study system, Equation (2.11) [21], [22], [23].

$$\boldsymbol{GTI = DNI' + DHI' + \rho'} \tag{2.11}$$

The DNI corrected is calculated by taking into account the positional parameters of the Sun, the characteristics of the panel, such as the azimuth of the panel and its tilt, and by considering the Angle of Incidence (AOI), which is the angle between the direction of the solar panels and the normal vector on the panel's surface, Equations (2.12) and (2.13) Figure 2.11 illustrates AOI as well as the characteristics of the panel.

$$\boldsymbol{DNI' = DNI \cdot \cos(\theta_{AOI})} \tag{2.12}$$

$$\boldsymbol{\cos(\theta_{AOI}) = \cos(\theta_z) \cdot \cos(\beta) + \sin(\theta_z) \cdot \sin(\beta) \cdot \cos(\theta_{az} - \theta_{az,panel})} \tag{2.13}$$

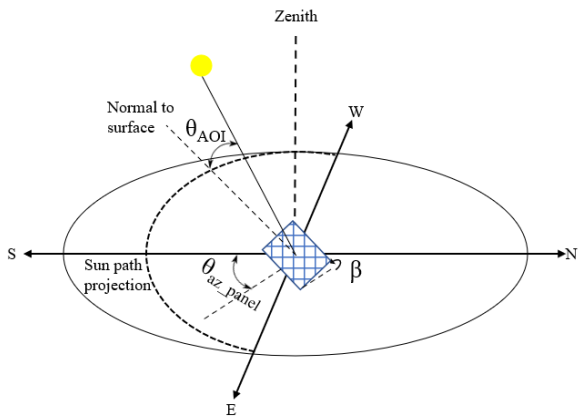


Figure 2.11: Representation of AOI, panel's azimuth, and its tilt

Several sky diffuse models try to define DHI corrected and the ground-reflected irradiance corrected for a tilted surface, such as the Hay Davies model used in [21]. One that is often used for its simplicity is to consider that the atmosphere is isotropic so that these irradiances are also isotropic. In other words, the intensity of the diffuse radiation from the sky is assumed to be uniform, and the reflection on the ground as well [23]. The isotropic model defines them through Equations (2.14) and (2.15).

$$\boldsymbol{DHI' = DHI \cdot \frac{(1 + \cos(\beta))}{2}} \tag{2.14}$$

$$\boldsymbol{\rho' = GHI \cdot \rho \cdot \frac{(1 - cos(\beta))}{2}} \tag{2.15}$$

2.2.5 Model

For modelling, in any kind of system one vital step is defining which parameters are the input and which are the output of the model. In the case of the PV system, as it has been said in the previous section 2.2.1, one vital input is the solar position and its irradiances. However, defining more inputs for the model will depend on how it is configured.

Two types of modelling for any PV system are differentiated considering which kind of data and methodologies are used in them [24]:

- Parametric model: conceives the PV system as a white box where each subsystem can be modelled the parameters and physical equations.
- Non-parametric model or data-driven model: conceives the PV system as a black box. This approach does not presume any knowledge of the internal characteristics and processes of the system. Instead, it is a data-driven model that estimates the behaviour of the system from a historical time series of inputs and outputs.

2.2.5.1 Parametric model

The parametric approach is a multi-step model that requires a detailed characterisation of the performance of each significant component of the PV plant and, in order to obtain the operating conditions of the PV generator, it is necessary to know the correlation model that fits the site correctly. As not all this information is always available, some simplifications and preliminary assumptions are often made, leading to uncertainty in the results later on. Therefore, the accuracy and precision of the estimates of a parametric model depend not only on the performance of its sub-models but also on the accuracy of the parameters [24].

A parametric methodology can also be used to estimate the power before the construction of a PV plant, during the thesis and planning phases, using, for example, the nameplate characteristics of the PV plant components and generic decomposition/translation models.

The PV parametric models usually define tests to obtain parameters since the accurate measurement of a module is quite challenging. With them, it is easier to characterize the module performance. Two of the most known are the Standard Test Condition (STC) and the Normal Operating Cell Condition (NOCT).

STC is a test to define which is the output power of the PV generators under what is considered ideal conditions. These conditions are the input data collected in Table 2.2 for a normal incidence of the irradiance to the panel. This particular test permits not only to see the efficiency of the PV panels in standard conditions but also to compare between them by using this fixed set of conditions, even if the operating conditions of each generator are different [25].

Table 2.2: Standard Test Conditions parameters and values

Parameter	**Symbol**	**Value**	**Units**
Normal irradiance	G	1000	[W/m^2]
Cell temperature	T_{cell}	25	[ºC]
Air mass	AM	1.5	[ϕ]

Nevertheless, when PV systems operate in the field, they tend to operate at higher temperatures and somewhat lower insolation conditions than the STC, and vice versa. To determine the power output of the solar cell, it is important to determine the expected operating temperature of the PV module. NOCT is defined as the temperature reached by open-circuited cells in a module under the conditions defined in Table 2.2 [19] [25]. With this temperature, the cell temperature at operating conditions can be defined and, with it, the total power output that can be made by the study system can be adjusted.

Table 2.3: Normal Operating Cell Temperature test conditions

Parameter	**Symbol**	**Value**	**Units**
Normal irradiance	G	800	[W/m^2]
Cell temperature	T_{cell}	20	[ºC]
Wind velocity	v_w	1	[m/s]

2.2.5.2 Non-parametric model

The non-parametric approach avoids the need for simplifying assumptions and precise parameters by using historical time series of meteorological variables (inputs) and power measurements (outputs). Therefore, the accuracy of a non-parametric model depends mainly on the quality of these data. However, this feature also comes with its main disadvantage, which is the PV plant must exist and be operational for some time so that the relevant input/output information is available. An interesting advantage of a non-parametric model is the possibility to compensate for systematic errors (biases) associated with the inputs [24].

Working with a non-parametric model is a synonym to work with historical data. However, the historical time series may have different lengths, which eventually leads to different results. Selecting methods are designed to define smaller time series lengths and what to consider to choose them. Therefore, the training set of the model can be composed of N days, selected from a larger database, which is the historical data. Three diverse selecting methods are distinguished [13] [24]:

- Previous: This method selects those N days immediately before the day to forecast. As a consequence, the database must be complete up to the day before the prediction.
- KT: This method selects N days based on the absolute difference between the clearness index of the day to be predicted and the clearness index of each day included in the database. The N days with the smallest absolute difference are chosen to form the training set.

- KS: This method selects N days based on the similarity between the empirical distribution function of the irradiance forecast for the day to be predicted and the empirical distribution function of the irradiance forecast for each day included in the database. The N days with the lowest differences with the day to forecast are the ones to conform to the training set.

2.2.5.3 Overview of the two models

Two ways of modelling PV systems are available for the thesis, and both of them presents some advantages and disadvantages, Table 2.4, that have to be considered in order to formulate the PV model.

Table 2.4: Comparision of parametric and non-parametric models

	Advantages	**Disadvantages**
Parametric models	– No need for the existence of the PV plant – In general, faster with calculations.	– Detailed characterisation – Assumptions – Depends on the accuracy of the parameters
Non-parametric models	– No need for assumptions – Compensates biases	– The PV plant must be built, and provide historical data. – Depends on the quality of data.

3. Methodology

This chapter explains and justifies how the PV model is defined, and which methodologies are considered to accomplish it. Furthermore, it describes the study cases that are going to be used for chapter 5 and discusses which evaluation metrics are going to be used as well for evaluating the model performance.

3.1 PV Model

As previously stated in section 2.2., PV systems have a vast range of types depending on how they are classified. Since one of the main objectives is to make the model the most generic possible, it is required some assumptions to homogenize any system.

First of all, this model is centred only on fixed-axis systems. Indeed, single or two-axis tracking systems have recently become more and more attractive for ground-mounted systems [1], but, when it comes to PV panels for rooftops, it is more common to find fixed-axis kind [26], probably because they are more cost-effective than the other two.

Secondly, the PV system is modelled as a mixed model, using parts of the parametric models and non-parametric models explained in section 2.2.5. The core of the model is designed as a parametrical model. Only essential data that could be found easily in any PV installation technical sheet is used. In this way, each study system may be characterized by its features and the forecast output will be more accurate. Moreover, open data like the weather forecast is used for defining the atmosphere conditions, as can be seen in Figure 3.1.

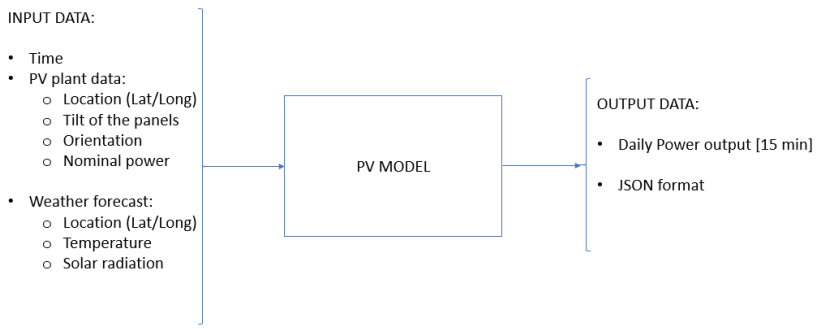


Figure 3.1: Diagram of input and output data of the model

Notwithstanding, only taking into account these few variables does not make the model very robust, and the accuracy of the results would be far from being satisfying. Variables that

could fix the robustness of it, like system losses or shades over the panels, are challenging to determine because they change over time. The solution proposed is adjusting the obtained results of the parametrical model with historical values using data-driven techniques. In this way, it is not vital to know those mentioned variables, and at the same time, the model gains robustness.

3.1.1 Parametrical part of the model

Any PV system follows Equations (3.1) and (3.2). Its generation power is determined by the nominal power (P_{nom}) of the installation, its system losses, and also the solar irradiance received on the panels [27], [28]. The STC defines system efficiency by comparing the output for the ideal conditions against the real ones.

$$\boldsymbol{P_{out,i} = P_{nom} \cdot \frac{G_{f,i}}{G_{STC}} \cdot \eta_{system}} \tag{3.1}$$

$$\boldsymbol{\eta_{system} = 1 - SL} \tag{3.2}$$

The previous Equation (3.1) is improved by taking into account the temperature conditions of the system that gets the form of Equation (3.3), the module temperature can be calculated using Equation (3.4) [27], [28].

$$\boldsymbol{P_{out,i} = P_{nom} \cdot \frac{G_{f,i}}{G_{STC}} \cdot \eta_{PV} \cdot \left(1 + \gamma \cdot \left(T_{cell,i} - T_{STC}\right)\right)} \tag{3.3}$$

$$\boldsymbol{T_{cell,i} = T_{amb,i} + \frac{G_{f,i}}{G_{NOCTC}} \cdot (NOCT - T_{NOCTC})} \tag{3.4}$$

The next subsections will discuss which type of parameters the model has for a better definition of its parametrical part. These parameters are differentiated by:

- Test conditions
- Weather data
- Study system

Furthermore, it will assign the default value for each parameter and which assumption is done, if it is required.

3.1.1.1 Test parameters

As can be seen in Table 3.1, some of the variables of the model take the values according to its test conditions. These variables should not change over time or neither by changing the PV study system, so they are constants with default values.

Table 3.1: Default values according to the test conditions

Parameters	Default value	Units
G_{STC}	1000	[W/m^2]
T_{STC}	25	[ºC]
G_{NOCTC}	800	[W/m^2]
T_{NOCTC}	20	[ºC]

3.1.1.2 Weather parameters

$G_{f,i}$, and $T_{amb,i}$ are variables that change over time. Both are weather data, so their accuracy is subject to the source where they are taken. For the results and the discussion of this thesis, the open-source used for getting both variables is *Dark Sky* [29]. This open-source not only gives the forecast for both variables but also it is possible to get historical data from them, which is important for the following parts of the model.

The vast majority of the open data weather forecasts do not give information about the solar irradiances, and if they do, it is without taking into account the weather conditions. In other words, they forecast a clear sky solar irradiance. However, *Dark Sky* gives solar irradiances taking into account the coverage of the sky. Figure 3.2 illustrates the differences between both cases.

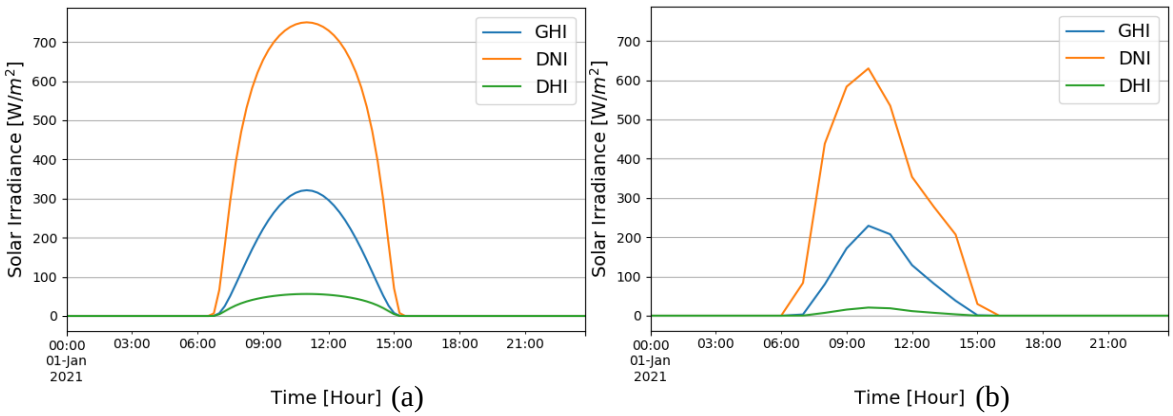


Figure 3.2: (a) Solar irradiances forecast for a clear sky (b) Solar irradiances forecast taking into account sky coverage

Weather parameters do not have default values since they can be obtained through an open-source. So, it means that they will always be required as an input of the model. Also, along with them, the location of the PV system where the data comes from.

3.1.1.3 System parameters

The remaining parameters to discuss are those intrinsic to the PV panel to study. They are the nominal power of the system (P_{nom}), NOCT, the temperature coefficient (γ), and the system losses (SL).

The nominal power is a regular variable to know of a PV system, and it is the elemental variable for the formulation proposed in Equation 3.1 and Equation 3.3. As a consequence, the nominal power does not have a default value, and it is always a must parameter to introduce as an input variable.

In contrast to the nominal power, NOCT and the temperature coefficient (γ) are not always available data of a PV installation. In these cases, the model adopts default values for them. NOCT is equal to 48ºC when it is unknown because according, to C. Honsberg and S. Bowden [30], it is a mean value for a typical module. And γ is equal to 0,38%/ºC when it is unknown. The model assumes that the PV panel is made from crystalline silicon since it is the most common material used in solar cells [31]. With this material, the γ can have a range of [0,27%/ºC-0,51%/ºC] [32]. If these system parameters are known and are different from the default values, the model can change its values.

SL are considered ideal and constant over time, but as explained in section 3.1, it does not happen like that. The total SL are the sum of smaller losses that the system has from the conversion of the sun-light until the A.C. output power.

SL can be measured or also estimated for each system. However, if the system to study does not have information about them, the model considers that the SL are equal to 0,2. This value is taken from [33], and according to it the overall losses for a system and considering the typical values of the smaller losses, the efficiency of the system is equal to 0,804 approximately, Table 3.2.

Table 3.2: Typical values and ranges for the losses of a PV system

Item	**Typical**	**Range**
PV module D.C. rating	1	0,85-1,05
Light Induced Degradation (LID)	0,98	0,90-0,99
D.C. wiring	0,98	0,97-0,99
Diodes and connections	0,995	0,99-0,997
Mismatch	0,98	0,97-0,995
Inverter	0,96	0,93-0,97
Transformer	0,97	0,96-0,99
A.C. wiring	0,99	0,98-0,993
Soiling	0,95	0,75-0,995
Shading	1	0,00-1
Availability of system	0,98	0,00-0,995
Overall at STC	0,804	0,62-0,964

In the end, if the information of the PV study system is available it is used otherwise if it is unknown, the model will attribute the default values of Table 3.3.

Table 3.3: Default values of the system parameters

Parameters	Default value	Units
P_{nom}	-	[W]
NOCT	48	[ºC]
γ	0,38	[%/ºC]
SL	0,2	[ϕ]

With the previous equations, the model is capable of forecasting the output power for the following day. However, assumptions done with SL and the usage of external data can lead the model to miscalculations that make the results to be far away from the real values of the power output. For this reason, the model uses historical data from the location and the PV system of study for adjusting the forecast results with the real values.

3.1.2 Correction of the model with historical data from location

The location, the orientation, and the tilt of the PV panels are crucial not only for calculating which quantity of irradiance falls perpendicularly to them but also for considering which horizon they are facing. Horizon tells about shadings that inversely affect proportionally the power generation of the plant. The more shadows, the less energy is generated.

In general, weather forecasts do not have information about horizons, and unfortunately, the *Dark Sky* is not an exception. It should not be a critical problem for those PV plants that are well located and do not have any shadows on their horizon. But in the real world, nothing is ideal, so even for those PV plants, it is interesting to consider the horizon if the information is available.

PVGIS [34], an online open-data has the database of the horizon for a given location. *PVGIS* is a project from the European Commission Joint Research Centre that focuses on solar resource assessment, PV performance studies, and the dissemination of knowledge and data about solar radiation and PV performance.

PVGIS calculates the effect of local hills or mountains that block the light of the sun during some periods of the day, by using data about ground elevation with a resolution of around 90 m. However, this calculation does not take into account shadows from nearby obstacles such as houses or trees.

Figure 3.3 shows how two close locations can have such a different horizon and how they affect their irradiances. Moreover, it shows how it can change the shape over the year. It is interesting seeing how few meters of distance might drop the plant output. The irradiances showed in Figure 3.3 are an average of previous years calculated by using the Typical Meteorological Year (TMY) methodology explained in Section 2.2.3.

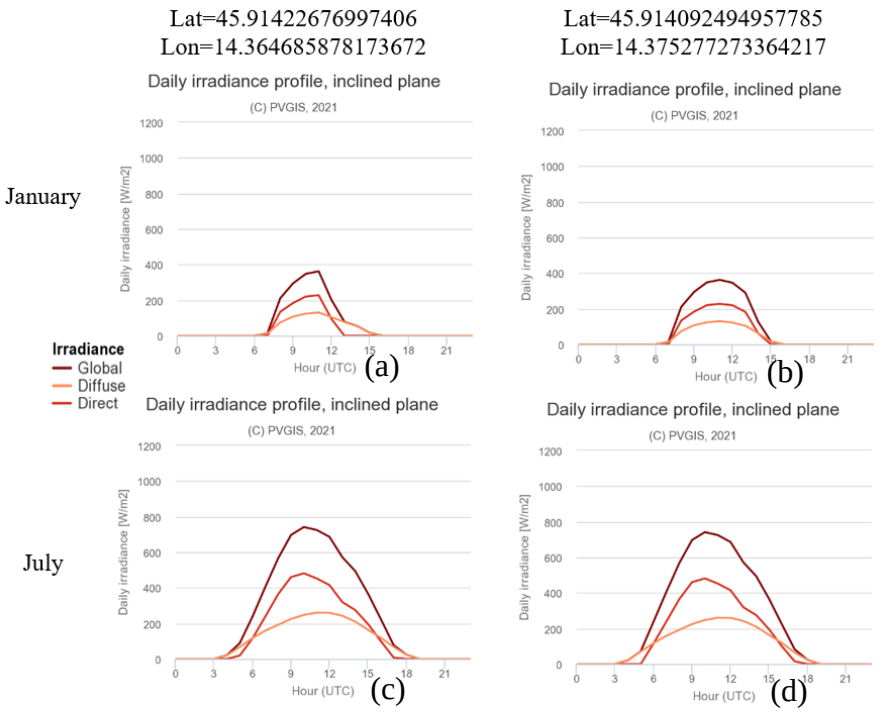


Figure 3.3: (a) Averaged irradiances received on January (b) Averaged irradiances received on July (c) Averaged irradiances received on January (d) Averaged irradiances received on July for the given locations

A correction factor is proposed for taking profit from this available information. It consists of the relation between the irradiance that could arrive on the surface with a clear horizon, and the irradiance that can arrive on the surface taking into account the horizon that faces it. This correction factor multiplies the power output obtained in Equation (3.3) for approaching the model's results to reality.

$$\boldsymbol{P_{out,i}}' = \boldsymbol{P_{out,i}} \cdot \frac{\boldsymbol{G_{horizon,i}}}{\boldsymbol{G_{no\_horizon\ ,i}}} \tag{3.5}$$

Building the correction factor like it is built-in Equation (3.5) ensures that its value is between 0 and 1. If the correction factor is equal to 0, it means that there is some obstruction between Sun and the panels. And if the correction factor is equal to 1, it means that all the possible irradiance that can arrive at the PV panels is reaching them. In other words, the horizon is clear of obstructions. Figure 3.4 shows in one two-axis plot the irradiances reaching PV panels with and without horizon and the ratio between both of them, which is the correction factor used for correcting the output power.

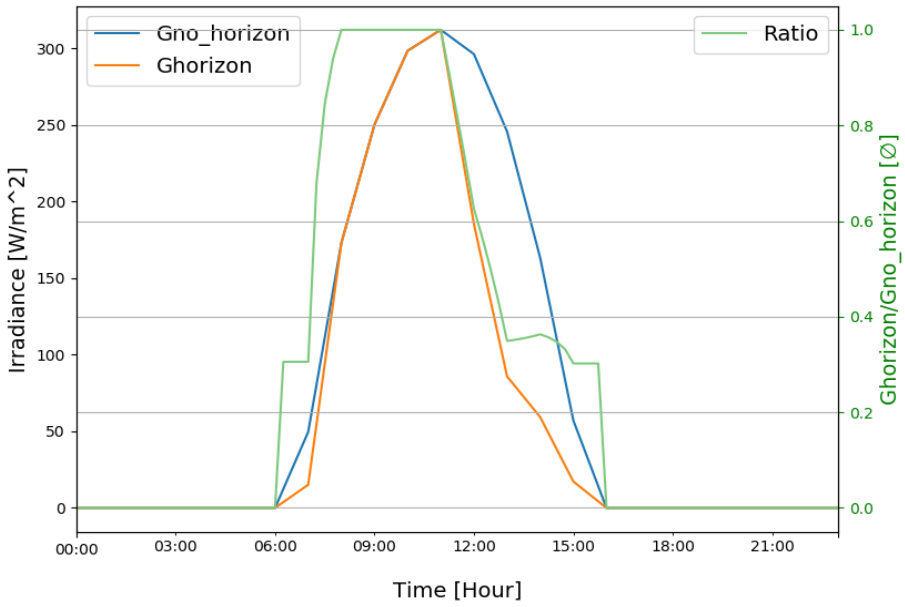


Figure 3.4: Irradiances received with and without horizon and the ratio between them for one day

Moreover, in Figure 3.4, it can be seen how when the irradiances do not match, the ratio of the correction factor drops, and when they converge, it is equal to 1. For this specific location, it could assume that there are obstructions in the horizon during the first hours of the day, and also, they are from midday until 16 o'clock approximately.

3.1.3 Correction of the model with historical data of PV plant

Historical data of a PV system is a valuable resource of information of the study system. Not only does it illustrate how much power is generating every day for real, but also it can help to compensate for all the assumptions made previously for building the PV model. By using regression methods can take care of the biased results or poor input data. Also, by using historical data, it can be detected if there are some regular shades nearby the PV system caused by a house or a tree, the fact that was not contemplated with the correction of the horizon, in section 3.1.2.

This part of the model proposes to take advantage of the historical data by creating a correction factor considering the prior days of the day to be forecast. The days are chosen by the selecting method called “previous” explained in section 2.2.5.2.

Unfortunately, working with historical data is seldom ideal. Data can present some noise produced by the constant change of the sky or measurement equipment. Figure 3.5 is an example of a sample of historical data from one Slovenian PV system, where it shows the noise in it.

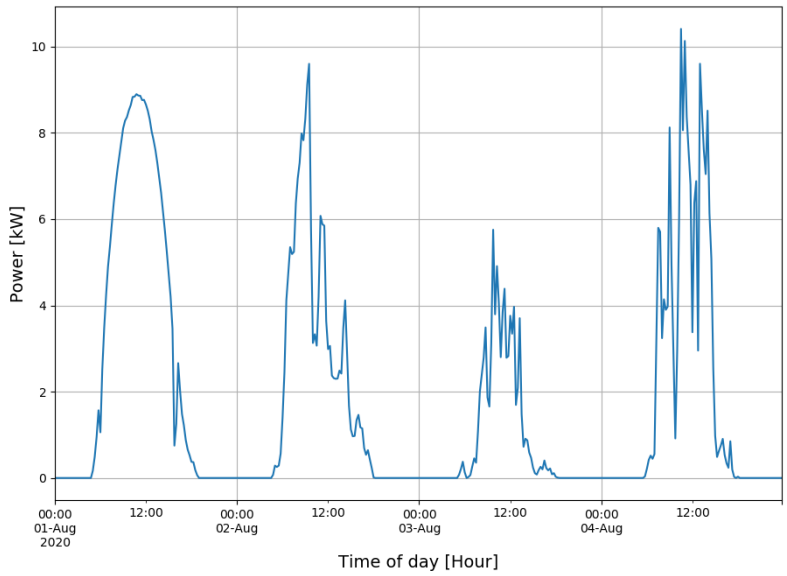


Figure 3.5: Plot of the historical power output of a Slovenian PV plant

This inherent noise of the historical data can add some uncertainties in the correction, which is desirable of being avoided. For that reason, it requires a transformation using Equation (3.6). It consists of a Simple Moving Averaging (SMA) of the data. SMA uses a sliding window to take the average over a set number of periods. In this case, each value is averaged with the two previous values, itself, and the two future values. In other words, the sliding window is centred, using five periods. Figure 3.6 shows the result of the transformation and how the power output curve is smoothed after it.

$$\boldsymbol{P_{h,i}}' = \sum_{i-2}^{i+2} \frac{\boldsymbol{P_{h,i}}}{\mathbf{5}} \tag{3.6}$$

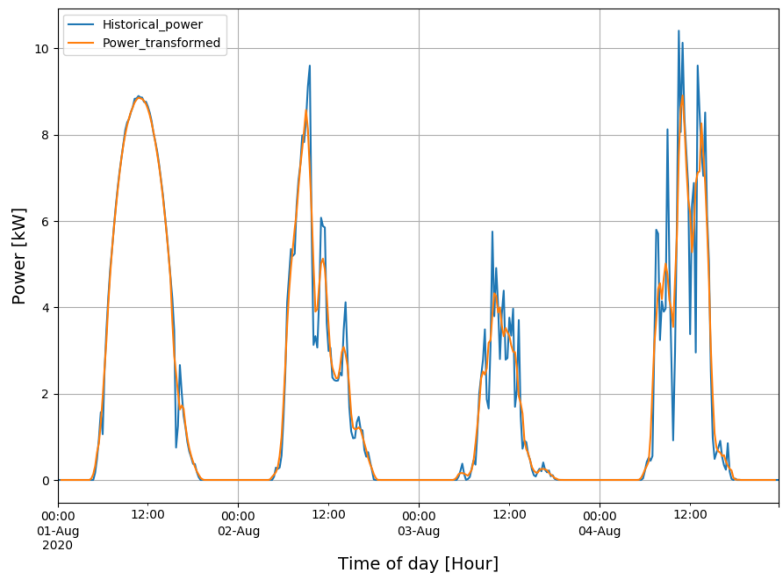


Figure 3.6: Plot with historical power output and its transformation

Continuedly, the model proceeds to calculate the mismatch between the averaged data with the historical prediction of the PV model. The historical prediction of the PV model results for using the PV basic formulation with historical weather and the horizon's location as an input. The mismatch is a division between the two data sets over time Equation (3.7). With this ratio, it displays how far the obtained results are from the real ones. Figure 3.7 shows the representation of it.

$$\boldsymbol{Div_i} = \frac{\boldsymbol{P_{p,i}}}{\boldsymbol{P_{h,i}}'} \tag{3.7}$$

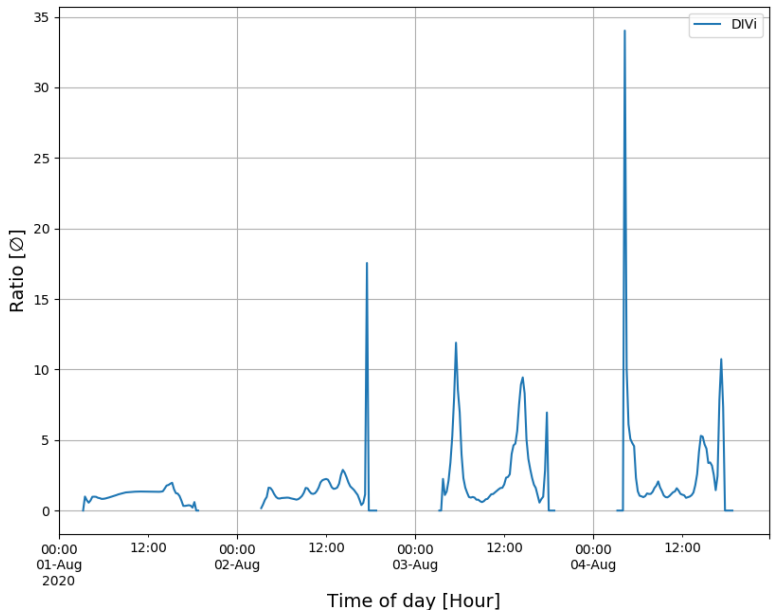


Figure 3.7: Division of the historical data with the predicted data for four days

This ratio between the prediction and the historical data is not enough for correcting the model. On the one hand, the gigantic spikes seen in Figure 3.7 would create outliers in the forecast. On the other hand, the aim of this regression is the behaviour of the system of study. So, it is necessary to separate the daily singularities, as it may be one not predicted cloud, from the ones that affect the system every day, as it may be the system degradation. For addressing these two issues, it proposes a weighted regression with a filter parameter.

The filter parameter is mainly for getting rid of huge spikes. It takes into consideration the output power of the prediction and the historical data. If one of them is below the threshold created by the multiplication of the filter and the nominal power, the ratio for that time will be considered zero.

The weighted function consists of the Gauss distribution function Equation (3.8). The reason is for giving more importance to those times that ratio is more nearby to one. In this way, without losing any information, it can be separate the daily singularities that can affect the system from the regular system working performance.

$$\boldsymbol{W_i} = \frac{\boldsymbol{e^{\left(-(Div_i-1)^2 / 2\cdot\sigma^2\right)}}}{\boldsymbol{\sqrt{2\cdot\pi\cdot\sigma^2}}} \tag{3.8}$$

With all the previous considerations, it creates the correction factor. Also, considering as many days as required and for each same hour, it is aggregated the weight function to the previous calculated ratio Equation (3.9).

$$c.f_{\cdot i} = \frac{\sum_{d=1}^{total_{days}} Div_{i,d} \cdot W_{i,d}}{\sum_{d=1}^{total_{days}} W_{i,d}} \begin{cases} Div_{i,d} = 0 \; if \; P_{p,i} \; or \; P'_{h,i} < P_{nom} \cdot filter \\ Div_{i,d} = Div_{i,d} \; otherwise \end{cases} \tag{3.9}$$

In the end, it obtains for a frequency of fifteen minutes, the correction factor that corrects the forecast of the model Equation (3.10). Figure 3.8 shows an example of a correction factor and how changing the values of the filter affects it.

$$P'_{out,i} = P_{out,i} / c.f_{\cdot i} \tag{3.10}$$

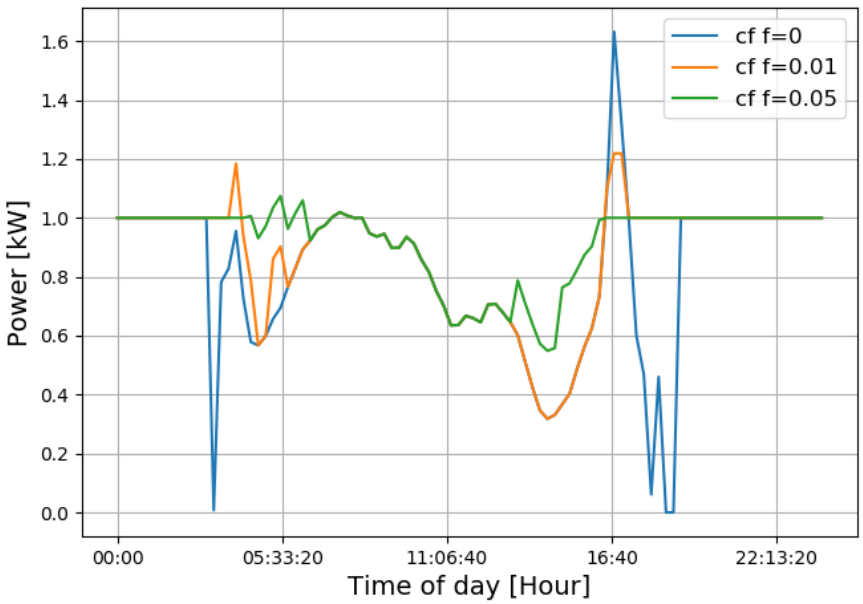


Figure 3.8: Correction factor for the same case but using different filters

3.1.4 Overview of the model

This section has the purpose of seeing an overview of the complete PV model. Figure 3.9 is a conceptual diagram of how is the model by taking into all the considerations made in previous parts of the report.

Firstly, using the historical weather data and considering the location's horizon as input data to the PV basic formulation, it gets the supposed historical forecast. Secondly, the historical data is transformed by averaging it.

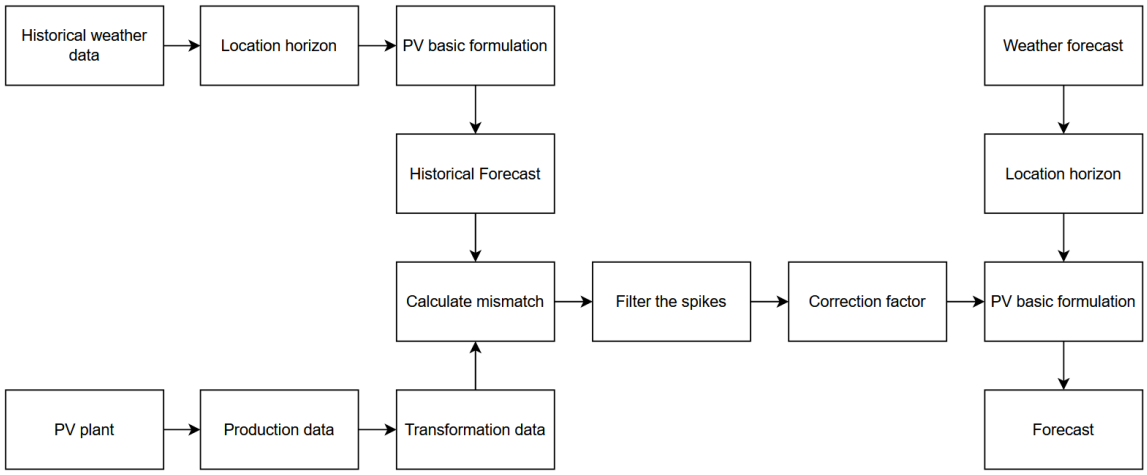


Figure 3.9: Conceptual diagram of the PV model

Then, the two paths unite by calculating the mismatch between the two obtained data sets. The mismatch correction is consolidated by getting rid of its spikes and also by giving more weight to observations that are interesting for the proper operation of the model. By doing that, a correction factor is calculated, ready to correct the forecast.

Simultaneously, the model calculates the forecast using the horizon of the location and the weather forecast as input data to the PV basic formulation. To conclude, the model corrects this forecast by multiplying it with the correction factor got it previously.

3.2 Model evaluation metrics

Once the PV model is developed and outputs forecasting results, it is necessary to verify its accuracy. The way to do it is by employing evaluation metrics. There are countless evaluation metrics, and each one has its pros and cons. As this thesis consists of a regression problem, this section discusses only regression evaluation metrics [35], [36], [37], [38].

Mean Square Error (MSE) is a widely used regression metric. It essentially finds the average squared error between the actual and predicted values. The smaller the MSE, the closer the fit is to the data. The mathematical expression is described as Equation (3.11).

$$\boldsymbol{MSE} = \frac{\mathbf{1}}{\boldsymbol{N}} \cdot \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{y_i} - \boldsymbol{\hat{y}_i})^{\mathbf{2}} \tag{3.11}$$

Root Mean Square Error (RMSE) is also one of the most widely used metrics for regression problems. It is the square root of the MSE presented before, making them very similar. RMSE has the same unit as the input data as it undoes the square applying square root, which is useful because it is directly interpreted with the got results. The mathematical expression is shown in Equation (3.12).

$$\boldsymbol{RMSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{N}} \cdot \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} (\boldsymbol{y_i} - \widehat{\boldsymbol{y}}_{\boldsymbol{i}})^{\mathbf{2}}} \tag{3.12}$$

Mean absolute error (MAE) is the absolute difference between the real value and the predicted value. This technique is considered more robust to outliers than the MSE metric, as this last one penalizes it more by applying the square of the error. Also, for this reason, it is not suitable for applications where the goal is to pay more attention to the outliers. The mathematical expression of this metric is defined in Equation (3.13).

$$\boldsymbol{MAE} = \frac{\mathbf{1}}{\boldsymbol{N}} \cdot \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} |\boldsymbol{y_i} - \widehat{\boldsymbol{y}}_{\boldsymbol{i}}| \tag{3.13}$$

The coefficient of Determination (R^2) represents the proportion of variance that has been explained by the independent variables in the model. It provides an indication of goodness of fit and therefore a measure of how well-unseen samples are likely to be predicted by the model, through the proportion of explained variance. As such variance is dataset dependent, R^2 may not be meaningfully compared across different datasets. The best possible score is 1. A constant model that always predicts the expected value of y, disregarding the input features, would get an R^2 score of 0. R^2 can have negatives values if the model does not follow the trend of the data. It is expressed by Equation (3.14).

$$\boldsymbol{R^2} = \mathbf{1} - \frac{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{y_i} - \widehat{\boldsymbol{y}}_{\boldsymbol{i}})^{\mathbf{2}}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{y_i} - \overline{\boldsymbol{y}})^{\mathbf{2}}} \tag{3.14}$$

Recapitulating, MSE and RMSE give practically the same information, so using both of them for evaluating the model could be redundant. Probably, RMSE is more interesting than MSE since its measurement units are the same as the results, and hence, it makes it easier to interpret the magnitude of the error. Alternatively, MAE may not be the best evaluation metric for the model since it does not penalize the outliers. One outlier can wholly vary the total amount of power output forecasted for the next day. On the contrary, R^2 is interesting to see how well the model fits with the system. However, it will not be used for comparing different cases of studies as mentioned before. Consequently, neither MAE nor MSE are used to evaluate the model, while the selected evaluation metrics are RMSE and R2.

3.3 Case studies

In this section, the stages of the discussion stages of the study cases are defined. Moreover, some assumptions and restrictions for a proper and realistic operation of the model are defined as well. Assumptions commonly found in the case study are described.

Bearing in mind that one of the main objectives of the model is to be generic, it would not make sense to analyse only one case of study. It is required for at least two PV systems from different locations for its verification.

Conveniently, the two countries chosen for the cases of study are Spain and Serbia, because the historical data of these two places were available, so it can evaluate the model in its totality. In order to get historical data of PV systems in Spain, an open-data resource called pvoutput has been used, it is a free online service used for sharing and comparing photovoltaic solar panel output data [30]. It is possible to download data for a PV system that shares its information for the previous months. However, for the Serbian PV system, the data have been provided by *Institutu Mihajlo Pupin*. Not only the historical data from the system's power output but also the direct measurements done by the same system.

3.3.1 Case study of Spain

The Spanish case study is a PV system located in Llíber, Alacant. All data used to calculate the subsequent results and discussion are obtained from the open-sources mentioned earlier: *pvoutput* and *Dark sky*.

3.3.1.1 Input data

The study PV installation is on the rooftop of a private house. It consists of 24 solar panels with nominal power equal to 250 W for each one. So, the total nominal power of the system is 6.000 W. Moreover, these PV panels are fixed with a tilt of 30º, and their surface orientation is 180º to the south. Additionally, the brand of the solar panels is *Zytech.* The datasheet of them defines their NOCT of 45ºC±2ºC and their temperature coefficient is equal to 0,40%/ºC. The datasheet can be found in Annex A: Datasheet of Zytech.

Table 3.4 collects all the input data that defines the study system. The test parameters adopt the default values defined in section 3.1.1.1. Also, the system losses of this case study are considered constant and with a default value equal to 0.2. The correction factor obtained of the historical data has the objective to compensate for this assumption.

Table 3.4: Input data of the PV system from Spain

Test parameters	**Symbol**	**Values**	**Units**
STC irradiance	G_{STC}	1000	[W/m^2]
STC temperature	T_{STC}	25	[ºC]
NOCTC irradiance	G_{NOCTC}	800	[W/m^2]
NOCTC temperature	T_{NOCTC}	20	[ºC]
Weather parameters	**Symbol**	**Values**	**Units**
Irradiance	G_i	*Dark sky*	[W/m^2]
Ambient temperature	$T_{amb,i}$	*Dark sky*	[ºC]
System parameters	**Symbol**	**Values**	**Units**
Latitude	lat	38.725384	[º]
Longitude	lon	-0.002660	[º]
Nominal power	P_{nom}	6000	[W]
Tilt	β	30	[º]
Surface orientation	θ_{az_panel}	180 (South)	[º]
NOCT	NOCT	45ºC	[ºC]
Temperature coefficient	γ	0,004	[ºC^{-1}]
System losses	SL	0,2	[ϕ]

The historical data of the Spanish PV system has a frequency of 15 minutes and starts from 1/11/2020 and ends on 19/01/2021. Therefore, it has in total 80 days on how the power performance of the system was, which can be used in the correction factor of the model by considering historical data.

3.3.1.2 Procedure

Before starting forecasting, it examines the historical data using visualisation tools to analyse its information and also to identify if there are missing values or outliers in the historical data.

Afterwards, for seeing how the model adjusts in the Spanish case, two different days from the historical data are considered to forecast their power output: one day with an apparently clear sky and another one that was a cloudy day. The clear sky day is where the PV model should have the best forecast since it is when fewer irregularities happen with the PV generator. On a cloudy day, the opposite is true. The model should have more difficulty in predicting the performance of the PV panels. However, it is important to see if the model is able to have accurate forecasts for these days. For the two days of study, it is going to apply four steps for seeing their results.

The first step consists of just forecasting the output power of those days using only the parametrical part of the model, which in theory is enough for getting the first results and comparing them with the historical data for that day.

The second step is to apply the correction of the previous outcomes by taking into account the obstructions in the horizon of the PV system. Performing these separately two steps allow for a discussion on whether the horizon is relevant for this particular location and how it influences the results.

The third step is the discussion of how many days should have been in consideration to correct the results with the historical data. An iterative process of how the evaluation metrics are changed by adding each time more previous days to consider is used for analysing from when there is a convergence on the results.

Finally, the last step is the calculation of the power output using the whole model and considering the total of the previous days decided in the third step. Again, by doing the other steps separately, it can be discussed whether historical data improves the model or vice versa.

3.3.2 Case study of Serbia

The case study in Serbia is a PV system located on the rooftop of the *Mihajlo Pupin Institute* in Belgrade. Unlike the case study in Spain, the historical data and meteorological data in this study come from direct measurements. Therefore, the open sources of *pvoutput* and *Dark sky* are not used in this case.

3.3.2.1 Input data

This system is not designed for domestic services, such as the Spanish one. However, it has been considered appropriate to use it as a case study for two reasons. Firstly, the historical data provided are more extensive than those available from open sources. Secondly, it has the direct meteorological measurements of the location, irradiance is obtained by the pyranometer which is positioned with the same slope as the panels. It might be interesting to compare both case studies to whether the results obtained are accurate considering the input data obtained comes from sources of different nature.

The system consists of 180 solar panels with nominal power equal to 280 W for each one. So, the total nominal power of the system is 50.000 W. Moreover, these PV panels are fixed with a tilt of 31º, and their surface orientation is 180º to the south.

For this study case is not necessary to know the NOCT value of the panels because it is known the $T_{panel,i}$ from the direct measurements of the PV system, so in this case, it can be used the Equation (3.3) directly without having to use the Equation (3.4) of the section 3.1.1. Unfortunately, the type of PV panels installed is not known, so the temperature coefficient must be assumed with the default value of 0.38%/ºC defined in section 3.1.1.3.

For the test parameters and system losses, the same assumptions have been made, as in the Spanish study case. Table 3.5 collects all the input data that defines the study system.

Table 3.5: Input data of the PV system from Serbia

Test parameters	**Symbol**	**Values**	**Units**
STC irradiance	G_{STC}	1000	[W/m^2]
STC temperature	T_{STC}	25	[ºC]
NOCTC irradiance	G_{NOCTC}	800	[W/m^2]
NOCTC temperature	T_{NOCTC}	20	[ºC]
Weather parameters	**Symbol**	**Values**	**Units**
Irradiance	G_i	Direct measurements	[W/m^2]
Ambient temperature	$T_{amb,i}$	Direct measurements	[ºC]
Variable	**Symbol**	**Values**	**Units**
Latitude	lat	46,398445	[º]
Longitude	lon	15,865453	[º]
Nominal power	P_{nom}	50000	[W]
Tilt	β	31	[º]
Surface orientation	θ_{az_panel}	180 (South)	[º]
NOCT	NOCT	-	[ºC]
Temperature coefficient	γ	0,0038	[ºC^{-1}]
System losses	SL	0,2	[ϕ]

The historical data of the Serbian PV system has a frequency of 15 minutes and starts from 01/06/2016 and ends on 30/06/2018. Having more than two years of data is interesting for

seeing how works the model over a year. Besides, it is interesting for analysing whether the correction factor is affected over time, and if so, how it is affected.

3.3.2.2 Procedure

As it is done in the Spanish case, the first thing to do when it is working with data is to do an insight in it to detect if there can be errors such as missing data and correct them if applicable.

For this study case, only one day is studied in detail due to two reasons. The first one is because the Spanish case is enough for seeing how differently the model adjusts on a clear and cloudy day. The second reason is that after the analysis in detail of the day chosen, it is forecasted all the days by following the same steps for seeing an overview of the evaluation metrics over a year, and for analysing how the correction factor changes over time.

For the analysis of a day of this case of study one part of the model is not necessary. Since the input data of the model comes from direct measurements, the horizon correction of the model is not needed because the pyranometer measuring the received irradiance is affected directed by the horizon as well.

Therefore, the study of the particular day is done in three steps instead of four. Calculating the power output only using the parametrical part of the model. Discuss how many days should be considered for using the historical data. Finally, correct the previous outcome by calculating the correction factor of the historical data by taking into account the previous days discussed.

After analysing the mentioned day, a forecast for the whole days of the historical data is done, as it has been said previously. It can be discussed if the model is robust over the year by studying the evaluation metrics values.

With these two different case studies, it is possible to see if the model can work in distant locations and with systems of distinct sizes and parameters. Furthermore, it is possible to compare if the model can adapt either if the input data is from open sources or direct measurements and whether its results are still good enough for both cases

4. Model architecture

The model is built by programming and considering all the methodology explained in Chapter 3 in order to have the output power forecast for each case study independently of its initial conditions.

The model is programmed by using the *Python* language, because, among other reasons, it is open-source with a wide range of software libraries. Furthermore, it is a portable language which means that it can be used regardless of any kind of operative system of the computer, which is interesting if future projects are built from it.

Two files have been created. One file is for calling the weather open-source data if applicable in the study case. The second file is the PV model itself, which will be the one explained in the further points of this chapter. They have been separated because if for any reason it is wanted to use another weathercast in future works, it is still possible to use the main file without having to change it.

4.1 Software libraries

In programming, a software library consists of a pre-written code with a collection of files, programs, routines, scripts, or methods referenced in programming code. In general, it is designed to help both the programmer and the compiler of the programming language to build and run the software.

Typically, a developer can manually add a software library to a program to achieve more functionality or to automate a process without having to write code for it.

Open software libraries have been used inside the PV model to address some specific issues that could be problematic and laborious to solve by coding from scratch. Table 4.1 shows that not only the software libraries but also which performance has them inside the code of the model.

Table 4.1: Software libraries and their use in the model.

Software libraries	**Usage**
Pandas	Analysis and manipulation of large data
NumPy	Versatile vectorization, indexing, and broadcasting concepts.
requests	Call requests from open sources of the internet abstracting the complexities of making requests.
pvlib	Set of classes and methods for defining the solar position as well as determine total on surface irradiance
matplotlib	Comprehensive library for creating static visualizations in Python.
sklearn.metrics	Module implements function assessing prediction error.
json	Read and write JSON data.

4.2 PV model architecture

The architecture of the PV model file consists of a Class format with methods in it. Defining the model as a Class provides some advantages over programming some different independent functions from each other.

Firstly, the input data of the Class can be reused and called in any function of it. So entering the data only once is enough to use it, which means that it is more optimal for the code. Another advantage of working with a class of methods is that it is easier to troubleshoot problems within the code, as a malfunctioning method can quickly be found because it is independent of the others, even if they are in the same Class.

Furthermore, the use of a Class allows the user to create objects, which have the initial functions and variables. The initial parameters are those necessary for building the object, and that's why any Python Class needs the *__init__* function that takes care of that task.

As a result of this architecture, the rest methods of the Class, apart from the init function, are created separately from each other. However, they still can use the initial parameters of the Class, as well as other particulars of the method.

For the good management of the Class, it is important to classify specific functions of the model in different methods for each one. The more encapsulated the Class is with more functions, the easier it is to detect errors. Besides, these functions can be used as a set of constructs. Depending on the system under study, it is possible to use the functions in one order or another or even not to use the ones that are not necessary for that case. This architecture makes the model more flexible and adaptable to any case study.

The code of the PV model is built with the methods showed in Table 4.2. These methods are the essential ones that define the PV model. Moreover, Table 4.2 explains which are the input parameters of each method and what is the return of the same ones. However, other methods have also been built for assisting functions like the visualization of the outcomes by using graphs.

Table 4.2: Main methods of the PV model

Methods name	**Input parameters**	**Return**
__init__	– latitude – longitude – surface_tilt surface_azimuth – Wp – site – start_date – end_date	Creation of the object of the class and initialization of its parameters.
solar_position	– Init parameters	Series of Sun position.
poa_irradiance	– Init parameters – weather_data: GHI DNI and DHI – solar_position: solar zenith and azimuth.	Series with POA: GHI, DNI, and DHI.
solar_power	– Init parameters – t_amb: Series with the ambient temperature – irradiance: POA GHI – system_losses: the constant value of SL – NOCT: NOCT – coef_t: coefficient temperature – stc_irradiance: G_{STC}	Series with the output power of the PV system.
pvgis	– Init parameters – usehorizon: parameter to consider the horizon or not for the location.	Data from PVGIS application.
power_location	– Init parameters – power_output: Output power of the PV system.	Series with the output power considering the system location.
correction_factor	– Init parameters – prediction: Historical output predicted data – real: Historical output data – filter: Power threshold to not consider.	Series of the correction factor considering the historical data.
correct_forecast	– Init parameters – forecast: Power output forecast – corr_factor: correction factor	Series with the output power considering the historical data.
accuracy	– Init parameters – series1: Series of results – series2: Series to compare	RMSE and R2 between the two series.
to_JSON	– Init parameters – Output: Series to transform	Transform output to JSON format.

5. Results and discussion

This chapter aims to present the conclusions obtained from the constructed model, such as the results of the model in different case studies and their analysis. It also intends to discuss the results obtained for particular conditions individually and then compare them collectively to see an overview of the model's performance.

5.1 Case study of Spain

As it has been explained in section 3.3, the case study of Spain is a PV system located on the rooftop of a private house from Spain. It has historical data of its output power for a total of 80 days. Furthermore, this case study uses weather open-data for its calculations.

According to section 3.3.1.2, the analysis of this case study starts with an overview of the historical data to better understand the behaviour of the system and to detect the days that are going to be used for testing the model's accuracy. Figure 5.1 presents the energy production history of the Spanish system over the eighty days of data. It shows that during two days of December (9 and 10), the PV system was off or did not measure the power production data.

Since the idea is to use historical data to correct the results, a zoom-in on the data has been done in January month for discussing which days to use in the model's evaluation. Figure 5.2 shows cloudy and what are assumed to be clear days. From the 12th of January onwards, it looks like the system is less affected by the clouds. However, from the way the power drops after midday, it could be assumed that the system is affected every day by some obstacle at the same time. It is still premature to discuss whether this obstruction is either because of the horizon's relief or because of the shadow of a near element.

In the further sections, two specific days are going to be discussed for evaluating the model. The 10th and the 15th of January are studied as cloudy and clear days, respectively.

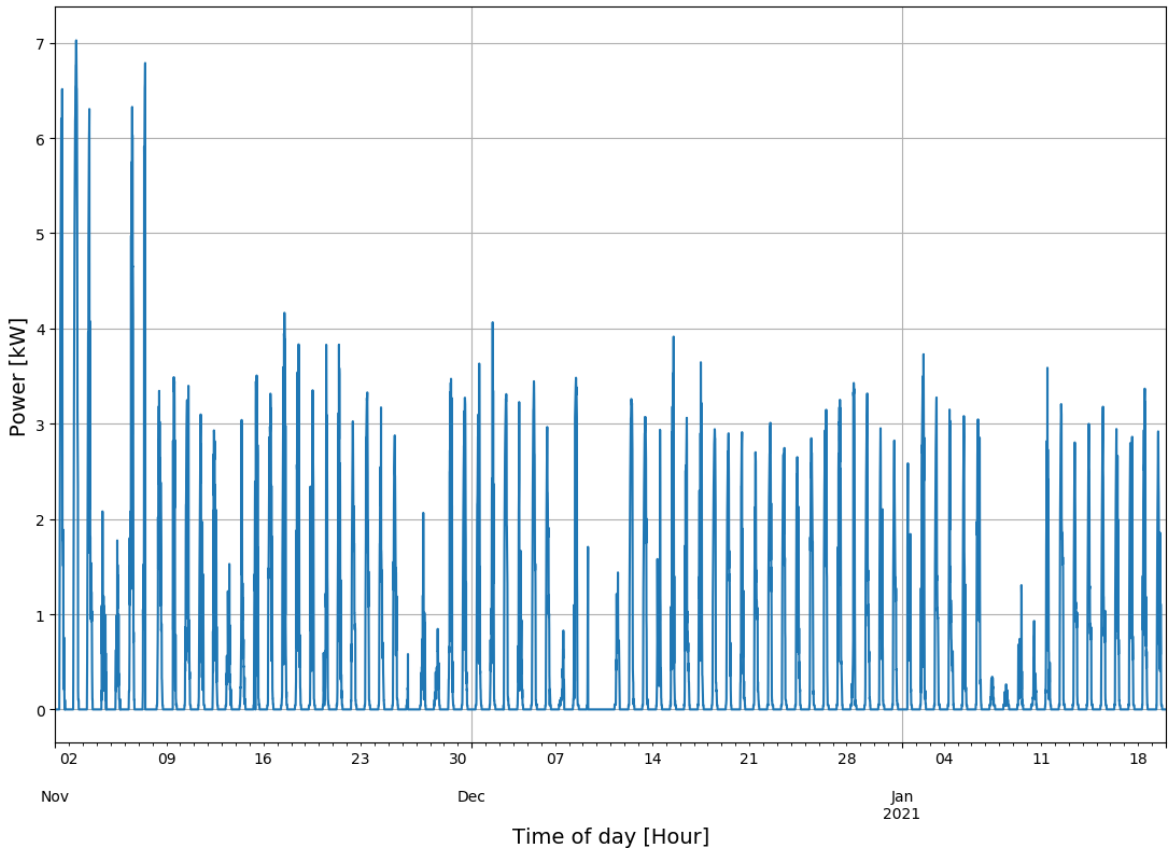


Figure 5.1: Historical data of the Spanish study case

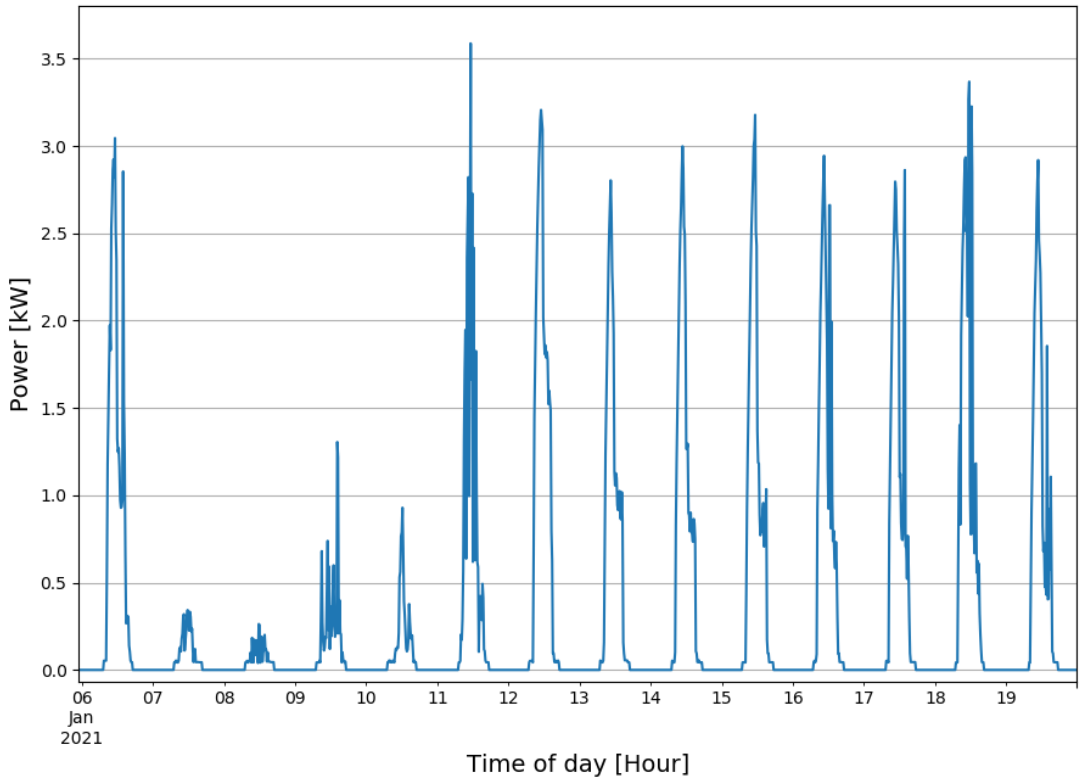


Figure 5.2: January data of Spanish study case

5.1.1 Forecast with the model's parametrical part

In this section, only the parametric part of the PV model is used to predict the power output of the plant, leaving the data-driven corrections for the following steps. As the case has information on its panels for the Spanish study, most of the input data is known. So, the results of this part will be tightly bound to the quality of the meteorological data.

5.1.1.1 Cloudy day

On the 10th of January of 2021, it was a cloudy day, and the PV system with a nominal power of 6 kW did not deliver 1 kW during the day like can be seen in Figure 5.3. This type of day is the most troublemaker for forecasting PV models because countless unpredictable events make the model lose its robustness.

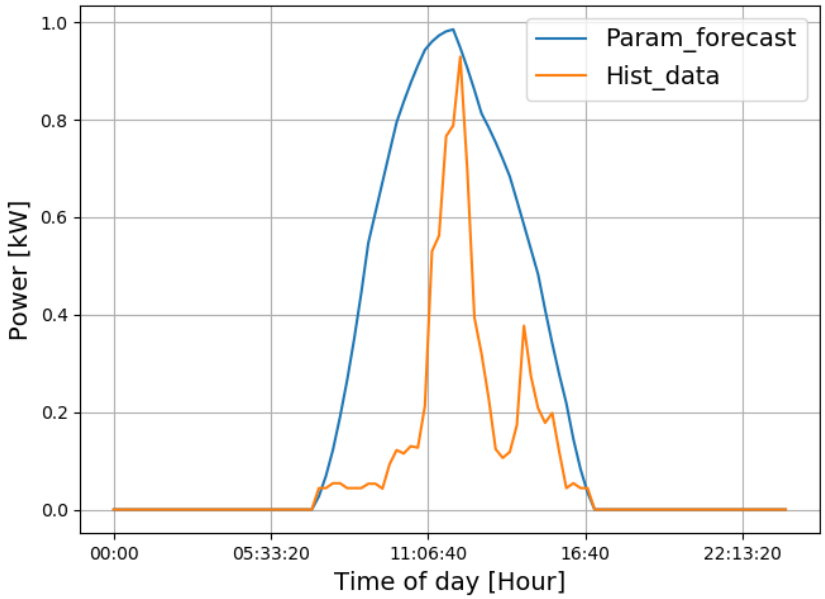


Figure 5.3: Historical data and forecast of the 10th of January of 2021(cloudy day) in Spain by using the model's parametrical part

For this day, the meteorological data was accurate in terms of the maximum solar irradiance value that the PV system reached. However, what it could not successfully predict was how the coverage sky would change over the day, causing the output power to drop. The accuracy of the model can also be seen in the results of Table 5.1, where it has an RMSE of 0,275 kW, a bit high considering that the maxim outcome of this day was near 0,9 kW. Moreover, the R2 equals 0,366, which is a positive value but still far from 1, the maximum value it can have.

Table 5.1: Evaluation metrics of the parametrical part of the model of a cloudy day

	RMSE [kW]	R2 [ϕ]
Parametrical_forecast vs Historical_data	0,275	0,366

5.1.1.2 Clear day

The 15th of January of 2021 was a clear day in the PV system. In this case, the model should perform the forecast better. However, as shown in Figure 5.4, the forecasting and the historical data have a divergence from 11 am onwards, approximately. As seen in Figure 5.2, the same seems to be true for other days with clear skies. It is not surprising that the model could not predict it based on meteorological data alone, as the reason for the occurrence is likely to be some shadowing affecting the system. This study case shows how with only the parametric part the model is too simple for forecasting and needed to be adjusted with other considerations.

This divergence between actual and predicted values is also observed in the evaluation metrics, as shown in Table 5.2. RMSE is even higher than a cloudy day, and the R2 is still far from 1. Yet, they are likely to be improved with further corrections.

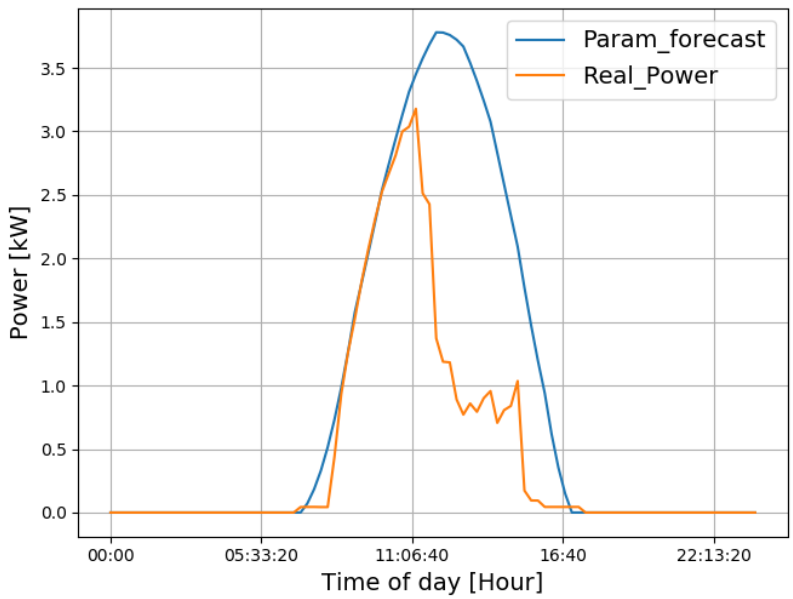


Figure 5.4: Historical data and forecast of the 15th of January of 2021 (clear day) in Spain by using the model's parametrical part

Table 5.2: Evaluation metrics of the parametrical part of the model of a clear day

	RMSE [kW]	R2 [ϕ]
Parametrical_forecast vs Historical data	0,917	0,527

5.1.2 Forecast considering the horizon

This section aims to correct the results of the parametrical part of the PV model using the historical location and orientation data. With this input data, the model can call the mean irradiances that the position receives monthly by taking into account the horizon if applicable. The ratio to correct the previous forecast can be calculated by using them.

In the case of Spain, the two study days are in January. So both days have the same ratio to correct their horizon. Figure 5.5 is a two-axis graph showing the average irradiances received by the site and the ratio between the two over the day. The left axis contains the power in kW of the irradiances. The right axis is the ratio values, and they can take values from 0 to 1.

For this particular case, it appears that the system should be affected by the horizon in the early hours of the day and also after about 15. It is in this time window that the Ghorizon and the Gno_horizon do not have the same values. As a consequence, the value of the ratio decreases or increases depending on the distance between the two irradiances.

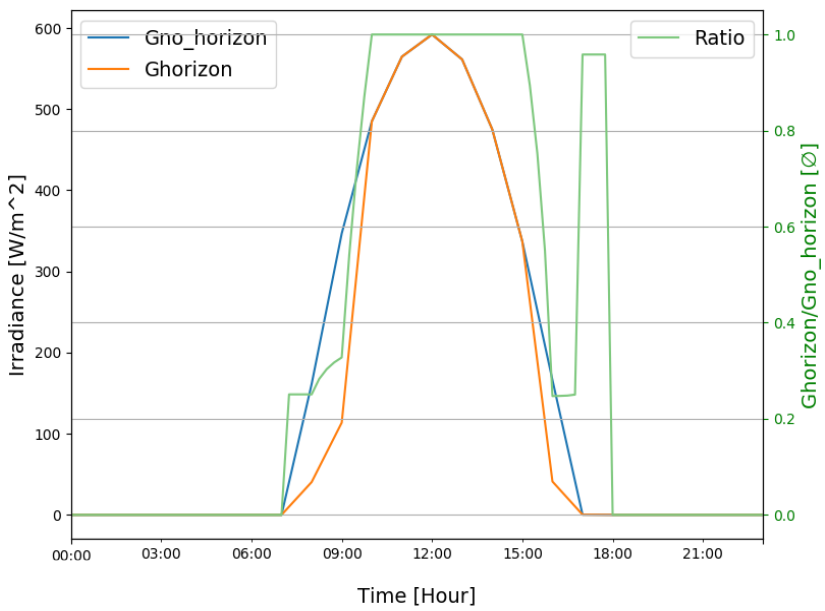


Figure 5.5: Irradiances considering horizon and not, and their ratio in Spain

5.1.2.1 Cloudy day

Figure 5.6 graphs three concepts: the forecast done by the parametrical part of the model, the correction of it with the ratio calculated accordingly to its horizon, and the historical data. It seems that the model fits better by considering the horizon, especially in the afternoon hours. It also translates into an improvement in the evaluation metrics of Table 5.3, where the RMSE decreases and the R2 increases.

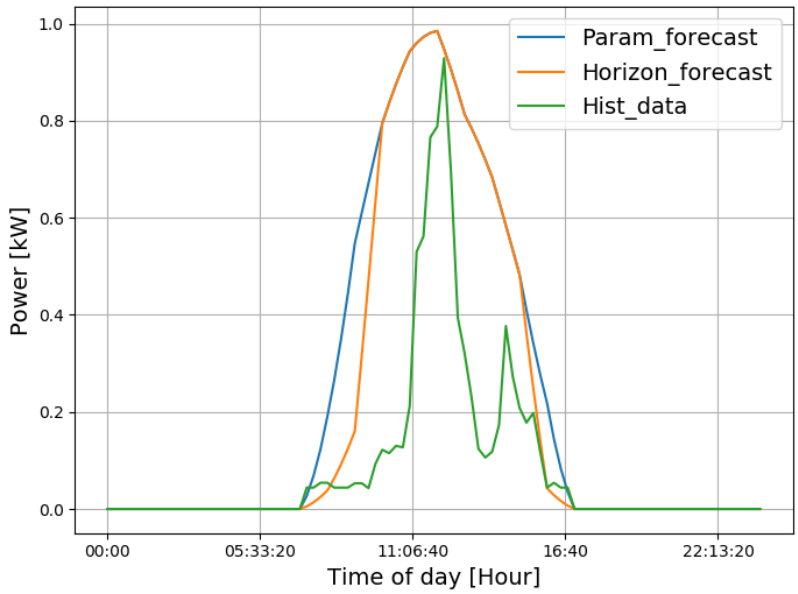


Figure 5.6: Forecast considering the horizon, parametrical forecast, historical data of the 10th of January of 2021 in Spain (cloudy day)

Table 5.3: Evaluation metrics of the model's parametrical part and its correction of a cloudy day

	RMSE [kW]	R2 [ϕ]
Parametrical_forecast vs Historical_data	0,275	0,366
Horizon_forecast vs Historical_data	0,251	0,457

5.1.2.2 Clear day

For this case, the results are not as good as the cloudy day. In the early hours of the morning, the results are a little tighter, but during the rest of it, the correction is more distant than the previous results. However, in the afternoon, the correction again helps to adjust the results to the historical data, Figure 5.7.

The improvement that the model sees in the afternoon is undervalued by the worsening results of the morning. Table 5.4 shows how the global results for the model are almost the same by using the correction. Furthermore, by looking again at Figure 5.7, the main problem of the study case is not due to the horizon. The hours when the output power decreases are the ones when the ratio horizon value is still 1.

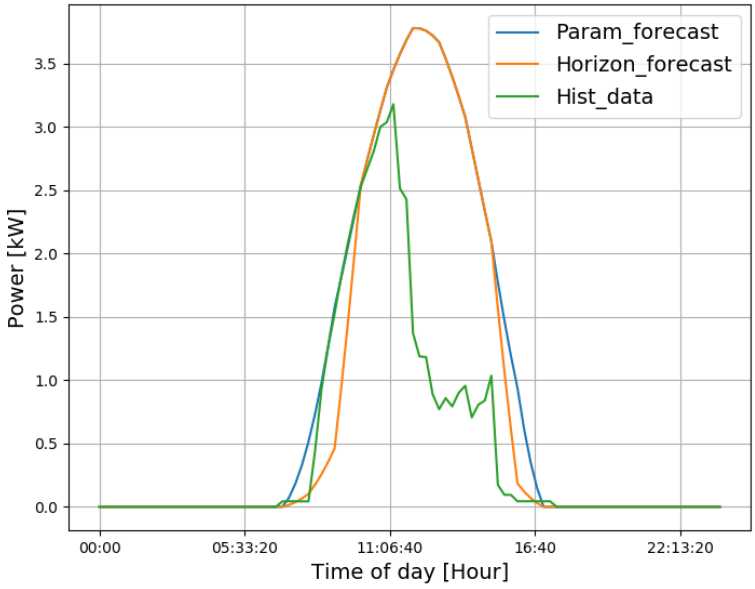


Figure 5.7: Forecast considering the horizon, parametrical forecast, historical data of the 15th of January of 2021 in Spain (clear day)

Table 5.4: Evaluation metrics of the model's parametrical part and its correction of a clear day

	RMSE [kW]	R2 [ϕ]
Parametrical_forecast vs Historical data	0,917	0,527
Horizon_forecast vs Historical_data	0,915	0,528

5.1.3 Discussing previous days

For correcting the result by using the historical data, one parameter has to be defined: the previous days to consider for calculating the correction factor.

On the one hand, if many days are considered, the more biased the correction factor will be, as the path of the sun changes daily and, with it, the sunrise and sunset. On the other hand, if too few days are considered for the correction, the model would lose robustness. The reason is that the correction factor would be more linked to meteorological phenomena than to the behaviour of the system under study. For example, assuming that the model corrected the forecast obtained for a clear day by taking into account only the previous day, which was a cloudy day, the correction would not meet expectations making worse results.

For discussing how many days should be considered for the study case, one day is taken from the historical dataset. It is calculated the variation of the evaluation metrics by changing the total of days considered in the correction. The results are in Figure 5.8, whereby having 1 or 2 days makes the RMSE and the R2 worse. If more days are considered it makes their values better, but in the long run, it also makes them worse.

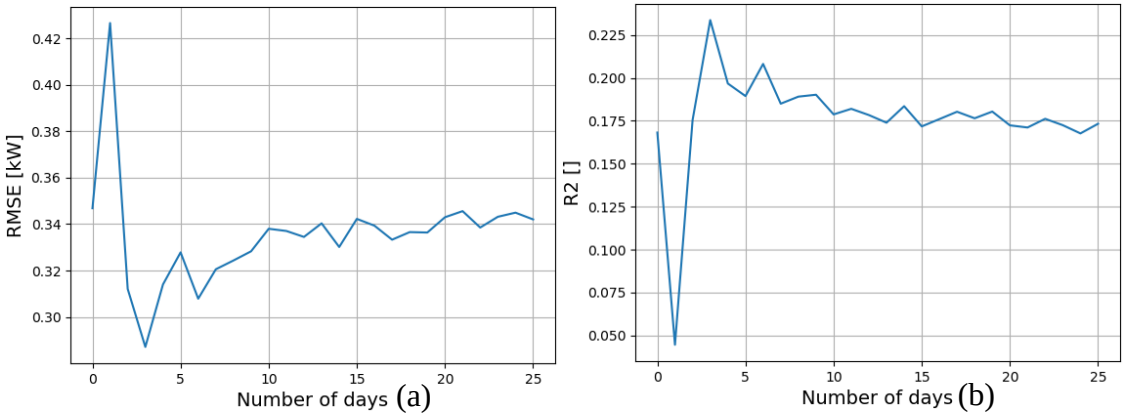


Figure 5.8: (a) Evolution of RMSE (b) Evolution of R2 depending on the previous days in Spain

5.1.4 Forecast considering the historical data

In this section, for the correction of the two days in this case study, only three previous days are taken into account to correct the results with the historical data. Moreover, it changes the filter for seeing how the results vary by not considering the small power output values that may cause spikes.

5.1.4.1 Cloudy day

With the historical correction, the cloudy day does not improve how the model fits with the system since the values decrease in Table 5.5 from the previously obtained. However, the distance between the forecast and the historical data is closer since the RMSE value also decreases. It also can be seen in Figure 5.9 where the corrected forecast and the historical data are not at the same distance as the horizon forecast with the historical data, but still, their shape is completely distinct.

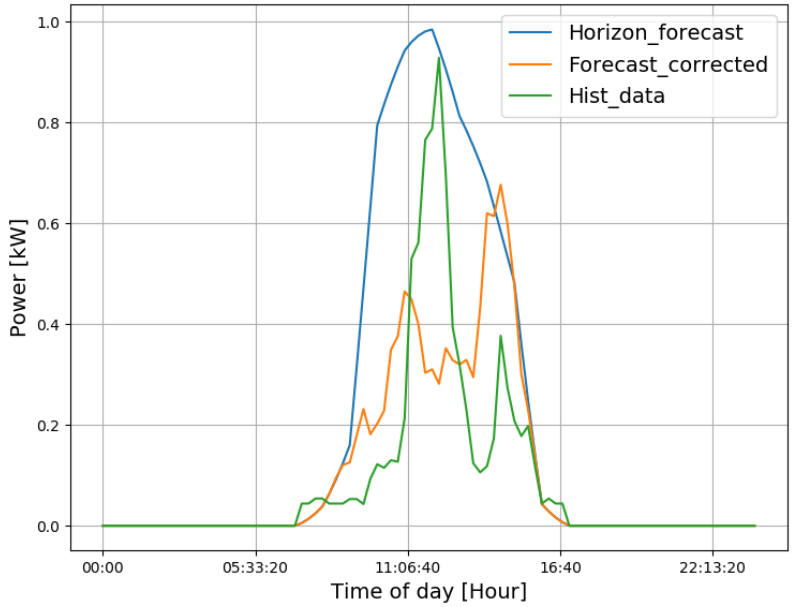


Figure 5.9: Forecast corrected using 3 days, parametrical forecast, historical data of the 10th of January of 2021 in Spain (cloudy day)

Table 5.5: Evaluation metrics of the model's performance on a cloudy day

	RMSE [kW]	R2 [ϕ]
Parametrical_forecast vs Historical data	0,275	0,366
Horizon_forecast vs Historical_data	0,251	0,457
Forecast_corrected vs Historical data (filter=0%)	0,151	0,263
Forecast_corrected vs Historical data (filter=1%)	0,150	0,268
Forecast_corrected vs Historical data (filter=5%)	0,213	0,461

5.1.4.2 Clear day

For the sunny day, the results obtained are wholly different. The use of historical data not only reduces the distance between the corrected forecast and the historical data, but the obtained shape of the corrected forecast is similar to the historical data, Figure 5.10. It translates into an improvement in the two-evaluation metrics, RMSE and R2, which show a decrease and an increase in their values, respectively, Table 5.6. The previous days did not have large spikes that affected the performance of the model, but it is still improved by not considering data below the threshold of 1% of nominal power.

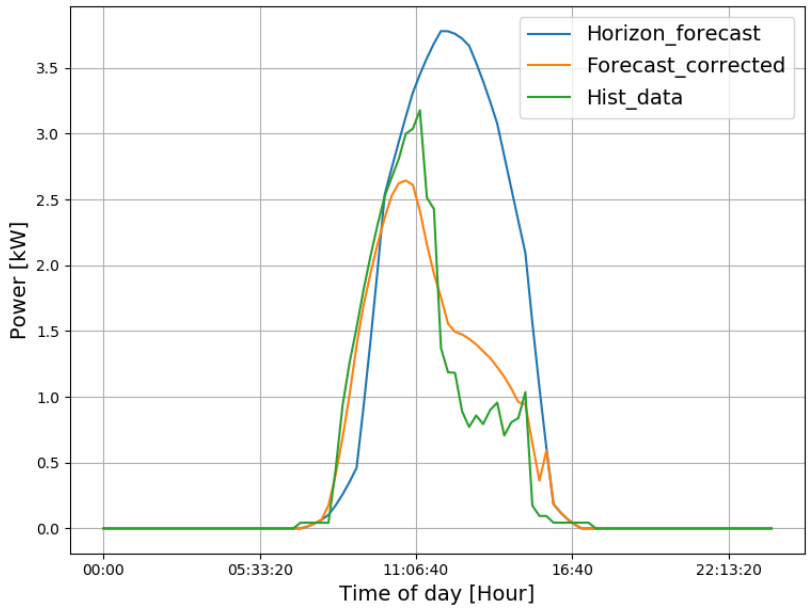


Figure 5.10: Forecast corrected using 3 days, parametrical forecast, historical data of the 15th of January of 2021 in Spain (clear day)

Table 5.6: Evaluation metrics of the model's performance on a clear day

	RMSE [kW]	R2 [ϕ]
Parametrical_forecast vs Historical data	0,917	0,527
Horizon_forecast vs Historical_data	0,915	0,528
Forecast_corrected vs Historical data (filter=0%)	0,214	0,929
Forecast_corrected vs Historical data (filter=1%)	0,208	0,934
Forecast_corrected vs Historical data (filter=5%)	0,263	0,895

The reason why the historical data affect a clear day and a cloudy day differently is due to the fact that a clear day is free of unexpected weather events. So, their correction achieves to adapt the forecast with the behaviour of the system. However, a cloudy day is unique due to unpredictable infinity situations. Although the study system is still the same, and the distance error is improved, the model cannot foresee are the sky conditions in the historical data. Therefore, the fitness of the model with the historical data gets worse.

5.2 Case study of Serbia

As it has been explained in section 3.3.2, the Serbian case study (with 50kW) is a PV system larger than the Spanish case study (with only 6kW) with the purpose of researching. The data of it is not only the historical output power but also the direct measurements of the meteorological parameters took it from the same place, as the panel's temperature or the solar irradiance on a tilted surface.

The total amount of data of this system is immense. Having it from 01/06/2016 to 30/06/2018 with a frequency of 15 minutes conforms to a dataset of 72.960 observations. Unlike the Spanish system, it is impossible graphing all of them in the same chart. However, it is still possible to detect missing data in the dataset by using scan functions that check all rows of the dataset.

For the historical data has been detected that 2.181 non-consecutive rows have missing values. It represents not even 3% of the total dataset. For this reason, the missing values, in this case, will be filled with values of 0 even though some information is being lost.

For this case, only one day is studied in detail, going through the steps of the model to get the result. Afterward, it is done the same for all the days for getting their evaluation metrics and see the model's performance over two years. The day selected is the 3rd of August of 2017, Figure 5.11, for seeing the model's behaviour in another season than the Spanish case.

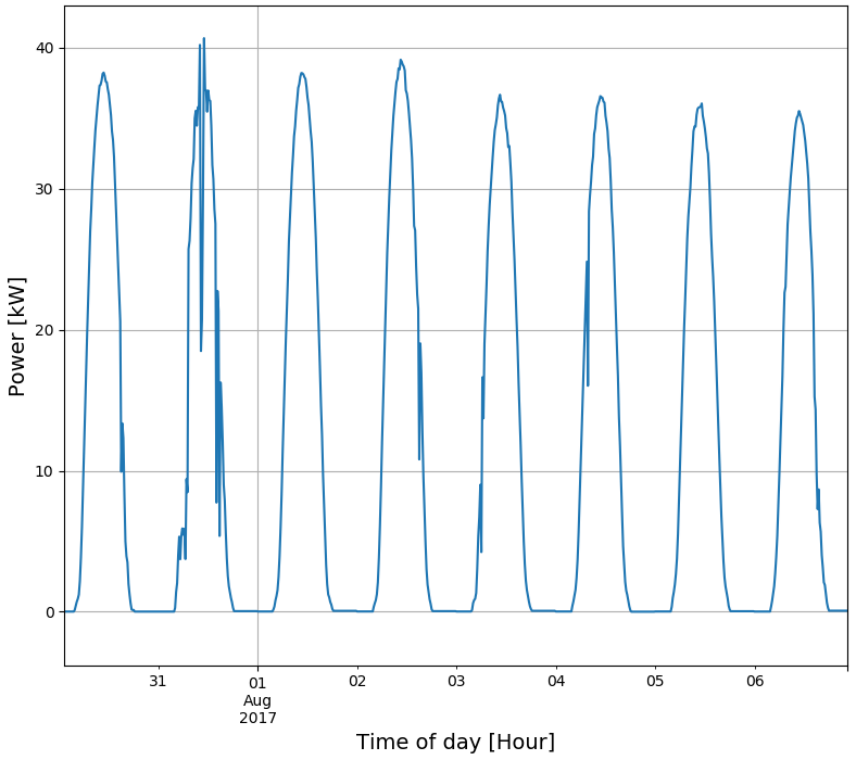


Figure 5.11: Sample of historical output data of the Serbian study case

5.2.1 Forecast with the model's parametrical part

The chosen day for the analysis was clear as the majority of August in Serbia. The results obtained are quite good. Not only do the direct measurements give the model accurate information about the change in solar irradiances for each instant, but it also appears that the assumed system losses are very close to those of the system on that particular day. Figure 5.12 shows how good is the forecast for the day, and it can also be seen in Table 5.7.

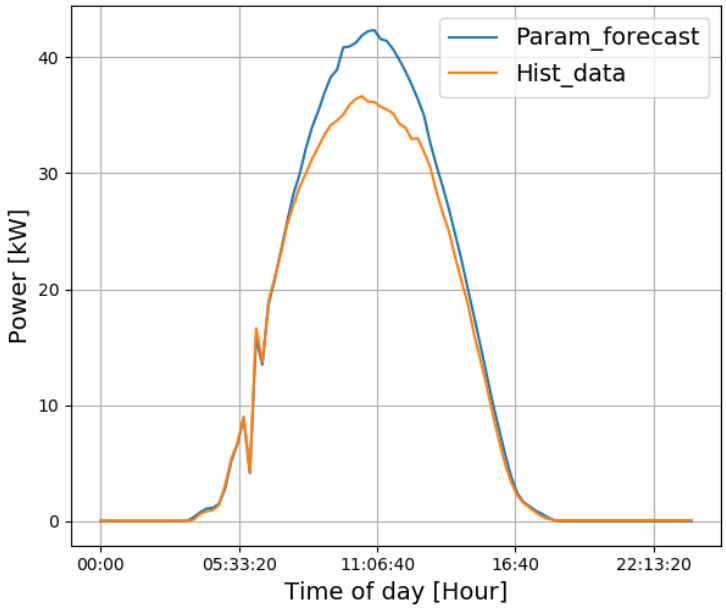


Figure 5.12: Historical data and forecast of the 3rd of August of 2017 in Serbia by using the model's parametrical part

Table 5.7 Evaluation metrics of the parametrical part of the model of the study day

	RMSE [kW]	R2 [ϕ]
Parametrical_forecast vs Historical_data	2,265	0,979

5.2.2 Discussing previous days

Despite having such good results only with the parametric part of the model, it is interesting whether using historical data still can improve the results. Therefore, it has to be discussed how many days should be considered for the forecast correction.

The same methodology that has been done in the Spanish case is applied for this case. An analysis of how the evaluation metrics change depending on how many days are considered in the historical correction is performed to define which days are optimal for it. The results of how the evaluation metrics change by taking into account different days are shown in Figure 5.13. If only a few days are considered, the system lost robustness, and the evaluation metrics got worsen. If more days are taken, it appears that the corrected results improve because the correction factor is adjusted more to the system performance than to the meteorological phenomena. After a threshold, the results got worsen again due to the sun path's changing over the year.

For the correction of the parametric prediction done previously, it is considered the twenty previous days of the forecasting day, which are the extremes for the RMSE and the R2 in Figure 5.13.

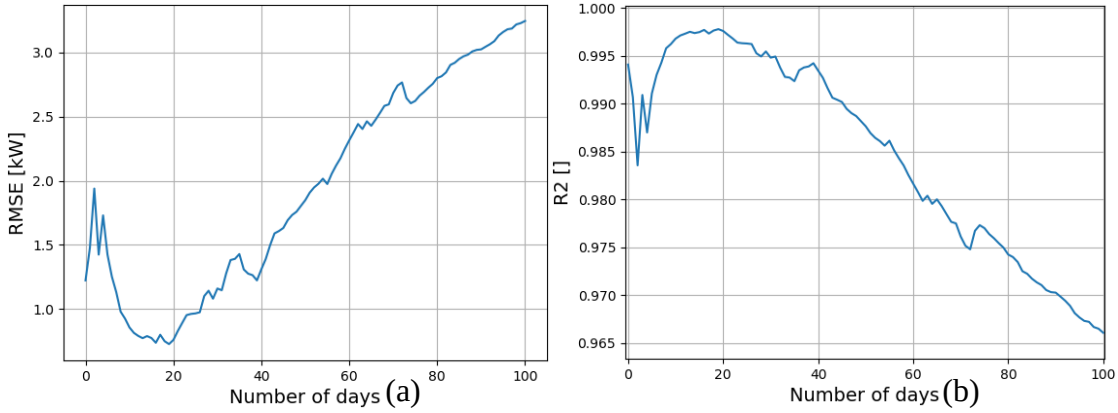


Figure 5.13: (a) Evolution of RMSE (b) Evolution of R2 depending on the previous days in Serbia

5.2.3 Forecast considering the historical data

As was expected for the previous results, the corrected forecast of this specific day has been improved by taking into account the historical data as shown in Figure 5.14. The corrected outcomes may seem to have more noise than the previous ones, but the evaluation metrics in Table 5.8 are still better than the first results.

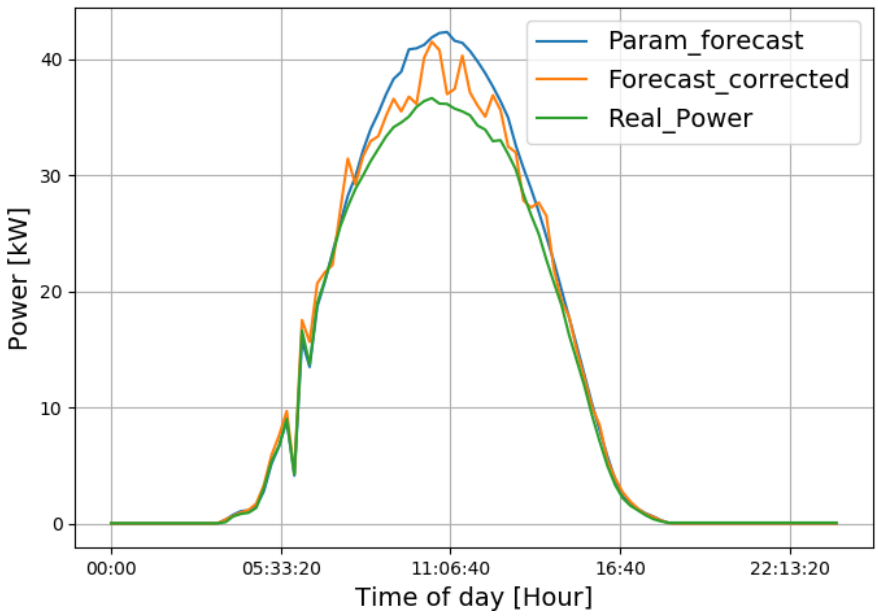


Figure 5.14: Forecast corrected using 20 days, parametrical forecast, historical data of the 3rd of August of 2017 in Serbia

In this case, the use of the threshold to filter out powers below a certain percentage of the nominal power of the PV system avoids spikes in the results. Setting the filter to 1% of the nominal power is sufficient to eliminate peaks. And by increasing the threshold, the correction is adjusted worse as more information is being lost from the historical.

Table 5.8: Evaluation metrics of the model's performance of the study day

	RMSE [kW]	R2 [ϕ]
Parametrical_forecast vs Historical_data	2,265	0,979
Forecast_corrected vs Historical data (filter=0%)	NaN	NaN
Forecast_corrected vs Historical data (filter=1%)	1,440	0,991
Forecast_corrected vs Historical data (filter=5%)	1,390	0,990

5.2.4 Results for one year

This section aims to extrapolate the same analysis done for one day to all the days of one year. The year chosen is 2017 because it is the only year in the historical dataset that has information for the whole year.

Like it has been done for the analysed day of the previous sections, each day of the year is calculated its forecast by using only the parametrical part of the model and correcting it by considering the twenty days before that day. Table 5.9 shows the monthly average got of the evaluation metrics for each day of the year 2017.

Table 5.9 Evaluation metrics monthly averaged of the Serbian system

2017	RMSE [kW]	RMSE with historical data [kW]	R2 [Ø]	R2 with historical data [Ø]
January	2,038	1,945	0,679	0,677
February	2,953	2,533	0,796	0,886
March	1,576	2,172	0,951	0,954
April	1,099	1,353	0,978	0,980
May	1,493	1,992	0,983	0,978
June	1,649	1,685	0,984	0,983
July	1,841	1,273	0,984	0,989
August	1,713	1,228	0,982	0,985
September	1,596	1,533	0,972	0,972
October	4,083	3,601	0,783	0,847
November	1,898	1,835	0,850	0,903
December	1,860	1,805	0,807	0,824
Annual mean	1,983	1,913	0,896	0,915

The fact that the system counts with its direct measurements and the fact that the system does not have any apparent obstruction makes the parametrical part reliable. Even though it does not make a big difference to use or not the historical data for this specific system, the annual mean of the RMSE and R2 is improved.

The model achieves to predict with an RMSE around to 2 kW of one system of 50 kW, which represents the 4% of the total. Looking at the results monthly seems that the model is more

robust in those months that are from spring and summer because the RMSE is lower and the R^2 is closer to 1 like can be seen in Figure 5.15 and Figure 5.16. The reason probably is the meteorological conditions of these months that tend to have more clear days, so there are fewer disturbances in the historical data.

Although there are three months when the RMSE by using historical data is higher than using the parametric RMSE, it is positive to see that the same parameter is better when the conditions are worse, like in October or February, Figure 5.15. The same happens with the R2, which may be worse in those months when the parametric part value is almost one but still improves those months when the R2 parametric value decreases Figure 5.16.

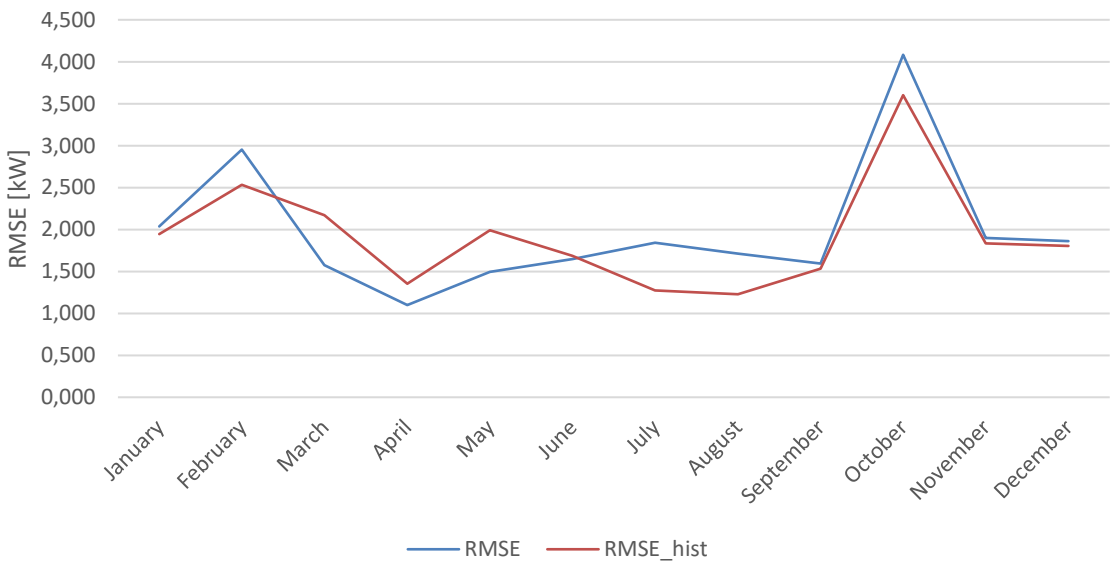


Figure 5.15: RMSE values over 2017

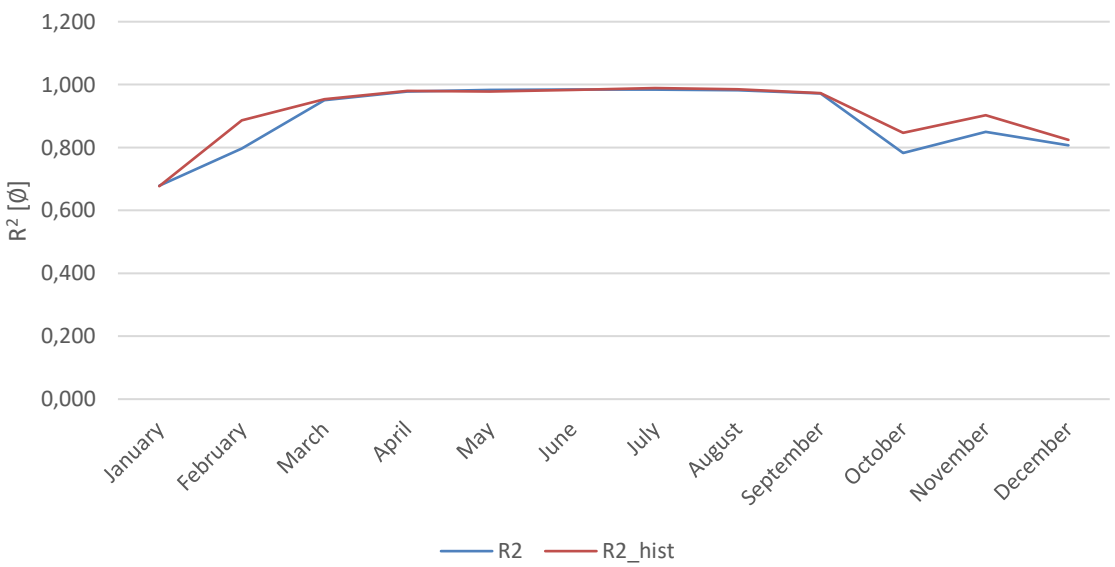


Figure 5.16: R^2 values over 2017

5.3 Comparison of both cases

The adjustment of the model results with the historical data is quite positive for both cases. The model has had to change its configuration to adapt to each case's necessities and initial conditions, and yet it has been able to calculate the key parameters and the wanted outcomes for both analyses as well.

The Spanish case study has operated with open data sources for its analysis, and it has been used for comparing a cloudy and a clear day. The differences between a clear and a cloudy day are not few. The clear day benefits of a considerable improvement when it takes into account the historical days for changing its results. However, the cloudy day losses capacity of forecasting by using historical data, even though the RMSE is still shrinking. Furthermore, the fact that the PV system has an obstruction not contemplated in the historical data has evidenced that only using the parametrical part proposed for this model is not enough for achieving satisfactory results. The horizon and the historical correction may be crucial in other studies like it has been for this one.

The Serbian case study has operated with direct measurements for its analysis. It has been used to analyse a day of a different season of the Spanish case (studying a summer day instead of the winter day), and to see the performance of the model for a whole year. The analysis of the summer day has been like it was expected. The energy production of the PV system was close to its P_{nom}. The historical data has helped to improve the forecast probably because in summer there are more clear days without unforeseen weather events than in winter. Moreover, not only the model's results were satisfactory for the day under analysis but also for the rest of the year on average. However, there are some specific days or months that the model does not get improving results by considering the historical data, so it can not be stated that this part of the model will always improve the results of the parametrical part.

Nevertheless, the Serbian PV system has better results than the Spanish one. The reason may be due to the initial conditions for both cases were not the same, so it is not something that can be attributed to the fact that the systems are located differently. The most likely reason for the differences in the fit for the two cases is that the Serbian case had direct measurements, especially solar irradiance. It could explain why with only the parametrical part of the model, its outcomes were exceptional. Also, another possible reason that affects the results is that the Serbian case, differently from the Spanish case, had not an obstruction that drops the efficiency of the PV system study.

One other noteworthy point when comparing the two cases is the difference in the number of previous days needed to correct the results using historical data. The Spanish study has only been required three days, whereas the Serbian study has been required twenty days. One assumption that can be done is because that the Spanish system was affected by some obstruction. As it has been said, the sun's path changes daily, but this fact should affect both systems more or less equally. However, having an obstacle as in the Spanish system can cause the change of the radiation received more rapidly each day.

6. Conclusions

This thesis aimed to develop and validate a PV forecasting model capable of adjusting its performance according to the system to be studied. Throughout this document, it has been shown all the details related to the design of the model. The following conclusions can be drawn:

- The PV model has been designed by using parametrical and non-parametrical techniques. This fact has helped to use the advantages of both ways of modelling and to compensate for their disadvantages.
- Designing the model as a class of methods is a good way to allow the user more flexibility when case studies have different initial conditions, which was one of the main objectives of the thesis. Furthermore, using the *python* programming language allows the user to use it in further studies despite her operating system.
- The two case studies have demonstrated that it is possible not only to use the model for different locations and still have satisfactory results in both of them but also the great adaptability of the model for each of them.
- The obtained results indicate that the model performs better when it has direct measurements at its disposal rather than relying on open data resources and trying to correct the data for the horizon of the case study. However, data-driven corrections to the model are still necessary, as considering only its parametric part is too simple for the ideal assumptions considered, as the system losses are constant.
- The historical values of the PV system output do not always improve the results of the model. As it has been seen, the study system can be under unpredictable conditions like cloudy days. Nevertheless, the annual analysis done in the Serbian case study, the global average is positive.

In conclusion, with growing concern for the environment and increasing installations of PV systems on homes, forecasting models are becoming increasingly important, not only to save a little money at the end of the month but also to help create a sustainable future. This research tries to be a humble contribution to this future.

6.1 Recommendations for future research

Despite that the established objectives are accomplished, there are future interesting work lines based on the current thesis developments. For future work, the following proposals are suggested:

- Use the current model in other PV systems with larger data if it is available (for example 5 or 10 years), and analyse if there is some direct relation between the PV system degradation and the correction factor obtained.
- Use another selection method on how many and which days should be considered in the model correction with historical data. And the ensuing discussion of which is the best result for the model.
- Do a global optimization of the model's code for having a better response speed of the software. It would be interesting if the model is going to be used in studies that use a large amount of data.
- Implement the designed model in a home energy management system and study whether its performance is adequate or whether some adjustments need to be made to improve it.

7. Bibliography

[1] G. Masson in I. Kaizuka, „Trends in photovoltaic applications 2020," IEA PVPS, 2020. p 8-17.

[2] M. Beaudin in H. Zareipour, „Home energy management systems: A review of modelling," Journal Elsevier, Calgary, 2015. p 2-14.

[3] B. Mahapatra in A. Nayyar, „Home energy management system (HEMS): concept, architecture, infrastructure, challenges and energy management schemes," Springer, Germany, 2019 p. (3-10).

[4] Y.-S. Son, T. Pulkkinen, K.-D. Moon in C. Kim, „Home Energy Management System based on Power Line Communication," IEEE Transactions on Consumer Electronics , 2010 p.(1382).

[5] J. I. Lee, C.-S. Choi, W.-K. Park, J.-S. Han in I.-W. Lee, „A Study on the use cases of the smart grid home energy management system," IEEE, Korea, 2011 p. (745-748).

[6] P. Danesea in M. Kalchschmidt, „The role of the forecasting process in improving forecast accuracy andoperational performance," ELSEVIER, Italy, 2011. pp. (204–214).

[7] A. Arbaz, M. Waqas, K. Shehzad in A. Mahmood, „Home Energy Management and Knapsack Technique in Smart Grid Environment," Research gate, Pakistan, 2015. pp. (1-5).

[8] International Energy Agency (IEA), „iea," [Elektronski]. Available: https://www.iea.org/. [Poskus dostopa January 2021].

[9] P. Mohanty, T. Muneer, E. J. Gago in Y. Kotak, Solar Photovoltaic System Applications: A Guidebook for Off-Grid Electrification, Springer, 2016.

[10] A. Awasthi, A. K. Shukla, M. M. S.R., C. Dondariya, K. Shukla, D. Porwal in G. Richhariy, „Review on sun tracking technology in solar PV system," Elsevier Energy Reports, India, 2020. pp. (392–405).

[11] The Fuel Cell Store, „fuellcellstore," 1999. [Elektronski]. Available: https://fuelcellstore.com. [Poskus dostopa January 2021].

[12] A. Joycea, C. Rodriguesa in R. Mansob, „Modelling a PV system," Elsevier Renewable energy, Portugal, 2001. pp. (275-280).

[13] A. Refaat, M. H. Osman in N. V. Korovkin, „Current collector optimizer topology to extract maximum power fromnon-uniform aged PV array," Elsevier Energy, 2020. pp. (1-11).

[14] J. S. Stein, „PV System Performance Modeling," International Energy Agency, United States of America, 2017.

[15] P. I. Cooper, „The absorption of radiation in solar stills,“ Solar Energy, 1969 vol. 12, p. 333 - 346.

[16] A. Karafil, H. Ozbay, M. Kesler in H. Parmaksiz, „Calculation of Optimum Fixed Tilt Angle of PV Panels Depending on Solar Angles and Comparison of the Results with Experimental Study Conducted in Summer in Bilecik, Turkey,“ Research Gate, Turkey, 2015.

[17] E. Al-Hilaly in I. J. Muhsin, „Gaussian Equation to Describe the Percent of Shadow Length in Satellite Image,“ Research Gate, Iraq, 2018 p 7.

[18] R. M. Milne, „Note on the Equation, of Time,“ The Mathematical Gazette, 1921 vol. 10, no. 155, p. 372 - 375.

[19] C. Honsberg in S. Bowden, „pveducation,“ [Elektronski]. Available: www.pveducation.org. [Poskus dostopa October-January 2020-2021].

[20] C. Chalkias, A. Faka in K. Kalogeropoulos, „Assessment of the Direct Sun-Light on Rural Road Network through Solar Radiation Analysis Using GIS,“ Research Gate, Greece, 2013. pp. (224-231).

[21] M. M. S. Akter in M. A. Shoeb, „A Novel Model to Calculate Global Tilted Irradiation (GTI) from Solar Variables Using Netcdf and Rstudio,“ International Journal of Scientific & Engineering Research, Australia, 2015. pp. (462-469) .

[22] W. F. Holmgren, C. W. Hansen in M. A. Mikofski, „pvlib python: a python package for modeling solar energy systems,“ Journal of Open Source Software, 2018.

[23] P. Loutzenhisera, H. Manz, C. Felsmann, P. Strachan, T. Frank in G. Maxwell, „Empirical validation of models to compute solarirradiance on inclined surfaces for building energy simulatio,“ Elsevier Solar Energy , 2007. pp. (254–267).

[24] M. P. Almeidaa, M. Muñozb, I. d. l. Parrab in O. Perpiñán, „Comparative study of PV power forecast using parametric and nonparametric PV models,“ Elsevier Solar Energy, Brazil-Spain, 2017. pp. (854-866).

[25] J. Montes-Romero, M. Torres-Ramírez, J. d. l. Casa, A. Firman in M. Cáceres, „Software tool for the extrapolation to Standard Test Conditions (STC) from experimental curves of photovoltaic modules,“ IEEE, Spain, 2016 . pp.(1-6).

[26] House Energy, „Advantages and disadvantages of fixed, adjustable and rotating solar rack systems (trackers),“ [Elektronski]. Available: https://www.house-energy.com/Solar/Solar-Racks.html. [Poskus dostopa 3 1 2021].

[27] Y. Riffonneau, S. Bacha, F. Barruel in S. Ploix, „Optimal Power Flow Management for Grid Connected PV Systems With Batteries,“ IEEE, France, 2011.

[28] H. Fontenot in B. Dong, „Modeling and control of building-integrated microgrids for optimal energy management – A review,“ ELSEVIER, Syracuse, 2019.

[29] Dark Sky, „darksky,“ [Elektronski]. Available: https://darksky.net/dev. [Poskus dostopa November-February 2020-2021].

[30] C. Honsberg in S. Bowden, „pvoutput,“ PVoutput, [Elektronski]. Available: www.pvoutput.org. [Poskus dostopa 12-1 2020-2021].

[31] M. Danylenko, „Matmatch,“ May 2018. [Elektronski]. Available: https://matmatch.com/blog/solar-panels-materials-efficiency-cost/. [Poskus dostopa January 2021].

[32] K. Emery, J. Burdick, Y. Caiyem, D. Dunlavy, H. Field, B. Kroposki, T. Moriarty, L. Ottoson, S. Rummel, T. Strand in M. Wanlass, „Temperature dependence of photovoltaic cells, modules, and systems," IEEE, USA, 1996.

[33] B. Marion, J. Adelstein, K. Boyle, H. Hayden, B. Hammond, T. Fletcher, B. Canada, D. Narang, A. Kimber, L. Mitchell, G. Rich in T. Townsend, „Performance Parameters for Grid-Connected PV Systems," IEEE, USA, 2005. pp.(1601-1606).

[34] C. E. C. J. Research, „PVGIS," Research, [Elektronski]. Available: https://re.jrc.ec.europa.eu/pvg_tools/es/#MR. [Poskus dostopa (10-02) (2020-2021)].

[35] A. Swalin, „Medium," April 2018. [Elektronski]. Available: https://medium.com/usf-msds/choosing-the-right-metric-for-machine-learning-models-part-1-a99d7d7414e4. [Poskus dostopa January 2021].

[36] D. Mishra, „towardsdatascience," December 2019. [Elektronski]. Available: https://towardsdatascience.com/regression-an-explanation-of-regression-metrics-and-what-can-go-wrong-a39a9793d914. [Poskus dostopa January 2021].

[37] V. Karbhari, „Medium," December 2018. [Elektronski]. Available: https://medium.com/acing-ai/how-to-evaluate-regression-models-d183b4f5853d. [Poskus dostopa January 2021].

[38] Scikit-learn, „Scikit-learn," 2019. [Elektronski]. Available: https://scikit-learn.org/stable/modules/model_evaluation.html. [Poskus dostopa January 2021].

8. Annex A: Datasheet of Zytech

ZT Series Polycrystalline

Solar Module

Zytech

250P/255P/260P

Efficiency

High Module Conversion Efficiencies using three or four bus bars per cell*

Warranty

Peace of Mind Guaranteed with a twenty-five Year Linear Power Warranty and Product Quality Ensured for ten years

Certifications

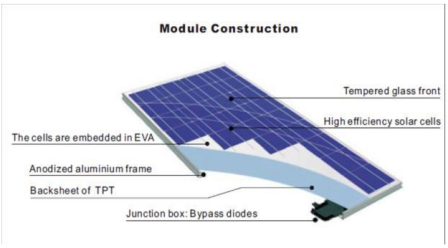
IEC 61215:2005, IEC 61730:2004, UL 1703-3rd Ed: 2014, ULC/ORD C1703-01:2014, ISO 9001:2008, PSK 024:2008

Tolerance

Strict quality control guarantees higher average
power output according to tolerance -3% to +3%

Optimized Strength

Minimum standard for wind 2700Pa to a Mechanical load capacity up to 5400Pa meet customer's needs for durability on high mountains, sea shores, & paths between buildings



EMPOWERING You

To Change The Future

Zytech was founded in Zaragoza (Spain) in 2005. Since then the group has progressively increased its infrastruc-ture and production capacity to become a global power with offices and headquarters in Spain, Germany, France, Italy, BENELUX, Mexico, United States of America, Korea,
and Malaysia.
Zytech takes pride in their R&D department which specia-lizes in product enhancement, state-of-the-art machinery and rigorous quality control that guarantees an European quality product at the best price.

www.zytechsolar.com

ZT Series Polycrystalline

Solar Module

250P/255P/260P

Cell Data	
Technology	Polycrystalline Silicon
Number Per Module	60
Dimension	156 × 156mm（6 inches）
Orientation	6 × 10

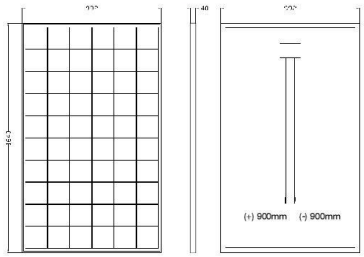
Thermal Data	
Nominal Operating Cell Temperature（NOCT）	45°C ± 2°C
Temperature Coefficient of Voc	- 0.275% / °C
Temperature Coefficient of Isc	+0.023 % / °C
Temperature Coefficient of Power Pmax	- 0.40 % / °C

Electrical Data（STC）		ZT250P	ZT255P	ZT260P
Maximum power（W）	Pmax	250	255	260
Power Output Tolerance（%）		±3%		
Maximum Power Voltage（V）	Vmpp	30.39	30.41	30.43
Maximum Power Current（A）	Impp	8.23	8.28	8.54
Open Circuit Voltage（V）	Voc	36.97	37.10	37.12
Short Circuit Current（A）	Isc	9.29	9.34	9.65
Module Efficiency（%）		15.40%	15.71%	16.01%

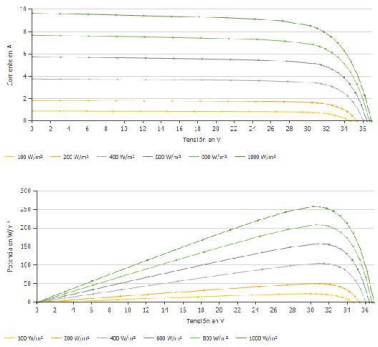
* At Standard Conditions（STC）Irradiance 1000 watt/m², spectrum AM 1,5 at a cell temperature of 25°C

System Integrated Parameters	
Maximum System Voltage SCII	1000 VDC（UL1000V）
Maximum Reverse Current	Do not apply external voltages larger than Voc to the module
Operating Temperature	-40~+85°C
Max Series Fuse Rating	15A

Front Side Rear



IU and PU Curve



Physical Characteristics	
Module Dimension（L×W×H）	1640×992×40mm. Code: PV30017 (ZT260P)
Weight	18 kg
Module IP Grade / J-Box	IP67 Module Rated / 3 Bypass Diodes IP67
Connector	MC4 or MC4 Compatible
Glass	3.2mm (0.13 in), High Transmission, AR Coated Tempered Glass
Cable	4 mm² PV cable, 900mm
Frame	Silver/ Black Anodized Aluminum Alloy

Packing Configuration	
Modules Per Pallet	26
Pallet Per 40' HQ Container	28（728 pcs）
Packing Box Dimension（L×W×H）	1700×1150×1040 mm
Packing Orientation	Side

Zytech Solar reserves the right to change specifications without notice *According to cells brand

USA Corporate Headquarters
Zytech Solar INC
11817 20th Street San Diego
CA 92127 USA
Phone: +1 619 878 2730
usa@zytechgroup.co

EUR Corporate Headquarters
Pol. Industrial Cortrovia
R. Janeiro, 12, E 50196
La Muela (Zaragoza) SPAIN
Phone: +34 976 14 1819
info@zytechgroup.co

GERMANY
Zytech GmbH
Str. Martin Str. 1 57714 Oberkirchbach
Tel: +499738 858282
Fax: +499738 858279
info@zytechsolar.com

INDIA PLANT
Zytech Solar India Pvt Ltd
Plot 841, Chandragiri Colony
Timsigerry, Secunderabad
Telangana, India 500016
info@zytechsolar.co.in

MEXICO PLANT
Zytech Group Mexico
Colonia Jardines del Pedregal
CP 01900 Mexico D.F.
Tel (52) 55 4162 5907
mexico@zytechgroup.co

CHINA PLANT
Zytech Engineering Technology Co. Ltd.
High Technology
Science Park 122 Zhuzhou Road
266 101 Qingdao, China
qingdao@zytechsolar.com