Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Energetica

A.a. 2022/2023

Sessione di Laurea Marzo 2023

Verifica della precisione di un modello, basato su previsioni meteorologiche, per la potenza oraria di impianti fotovoltaici

Tesi di laurea magistrale

Relatore:
Prof. Filippo Spertino
Co-relatori:
Ing. Alessandro Ciocia

Candidato:
Alessandro Gasperoni

Ringrazio innanzitutto il professore Filippo Spertino che mi ha prima appassionato alla materia e poi permesso di lavorare a questo progetto

di tesi. Ringrazio poi l'ingegnere Alessandro Ciocia, che pazientemente e con tanta disponibilità mi ha aiutato in questo percorso.

Ringrazio infinitamente la mia famiglia; mia mamma, mio babbo e Valentina per il loro supporto. Oltre a questo nucleo più stretto, ho avuto anche un appoggio fondamentale da parte degli zii Giancarlo, Eleonora, Giovanni, e mia nonna Anna.

Ringrazio infine gli amici, che in maniera più o meno simile hanno condiviso questo tragitto con me.

Contents

Introduction

Global warming and the resulting climate change are some of the biggest challenges and problems of today. In this context, a central role is played by energy transition, that is the transition from energy sources that release larger amounts of carbon dioxide, such as fossil fuels, towards clean energy sources. In this perspective, the contribution of photovoltaic energy is fundamental, in particular in Italy, where its abundance is relatively high compared to European average. In addition, renewable energy sources, being inherently decentralised, can contribute to energy independence, which is a very important topic after the huge increase in gas and electricity prices in Europe last year and recent geopolitical issues. On the other hand, renewable sources also have disadvantages; among these the main one is undoubtedly the intermittency of the resource and the consequent difficulty to forecast the production profile, which is determined by external events, while the production output of traditional plants is simply determined by energy demand. Electrical storage can be a solution as a compensation of the variability of the energy resource, however, it would require higher costs and materials, whereby it is not feasible on a very large scale. Forecasting photovoltaic production, which is the topic of this thesis, is a complementary alternative to storage and is essential to integrate renewables into the electricity grid and minimize costs.

The aim of this thesis is to utilize a photovoltaic power calculation model whose result is on an hourly production profile, with the use of weather forecast variables as input data to predict the produced electrical power. The quality of the model is tested on a PV plant for which electrical power and total irradiance are measured. The above-mentioned models are optimized in a second step to reduce the final error. As forecasts nature incorporates high uncertainties, a main focus of this thesis is represented by decoupling weather forecast error from power calculation model error, and by evaluating forecasts quality. More in detail:

- In Chapter 1 an overview of photovoltaics is presented, which describes the technology used, its contribution to the electricity grid, the importance that has recently acquired in the context of global warming. In particular the equation of PV cells which correlates current, voltage and power with weather parameters like irradiance and air temperature. Finally, a review of the literature of some photovoltaic prediction models., with the focus on the difference between physical models, that use the correlation between meteorological and electrical variables,

and statistical models which are based on statistical methods and artificial intelligence.

- Chapter 2 contains the description of the calculation model that is used in this thesis, which belongs to the category of physical models, and how it will be optimized in a second step.
- Chapter 3 is dedicated to the description of the methodology used to acquire data. In detail, in the study there will be the contribution of measured irradiation data, electrical power data measured and expected meteorological data (in particular irradiance components, but also temperature and wind speed). The greatest attention is paid to this last type of data, since they are forecast data, they will have to be continuously updated and they introduce the highest source of uncertainty. Some daily irradiation profiles will be analyzed in this section.
- Chapter 4 will report the results of the model. Here, after a rapid introduction on error calculation, models performance is displayed firstly on daily basis in order to have a clear visual comparison between forecasted and actual value, in particular it shown the effect of optimization and of forecast updating under a set of different circumstances. The chapter continues with the results of the power calculation on the whole period of simulation to estimate the global performance of the model. The last part of the fourth chapter is dedicated to the evaluation of the quality of irradiance forecast data over the 2022 year for four Italian regions.

1 General context

1.1 Renewable energy resources (RES)

Renewable power sources are harnessed from renewable and abundant resources that can continuously replenish themselves. This provides significant benefits for environmental conservation and human wellbeing, and they serve as a substitute for non-renewable energy sources in electricity generation. Renewable energy sources (RES) play a pivotal role in the decarbonization processes. These sources can be classified into:

- solar energy, which is divided between solar thermal and solar photovoltaic energy.
- wind power.
- hydroelectric power.
- biomass energy.
- geothermal energy.

Categorization of RES can be based on programmability, with two classifications: programmable and non-programmable. Programmable RES, including hydroelectric and biomass energy, possess the flexibility to regulate their production to meet grid demands. The power output can be managed and scheduled to match the need for electricity. For example, hydroelectric power plants can release or retain water to match demand, while biomass power plants can adjust the amount of fuel burnt. Conversely, non-programmable RES, such as photovoltaic and wind energy, are influenced by weather patterns and are difficult to regulate. They cannot be easily scheduled or controlled, and their power generation is influenced by natural resources such as sunlight or wind. However, techniques like curtailment and re-dispatching can be utilized to integrate non-programmable RES into the grid. Curtailment implies disconnection of non-programmable RES from the grid when the power they produce exceeds demand. On the other hand, re-dispatching necessitates adjusting the output of conventional power plants to balance the grid in case of surplus power from non-programmable RES.

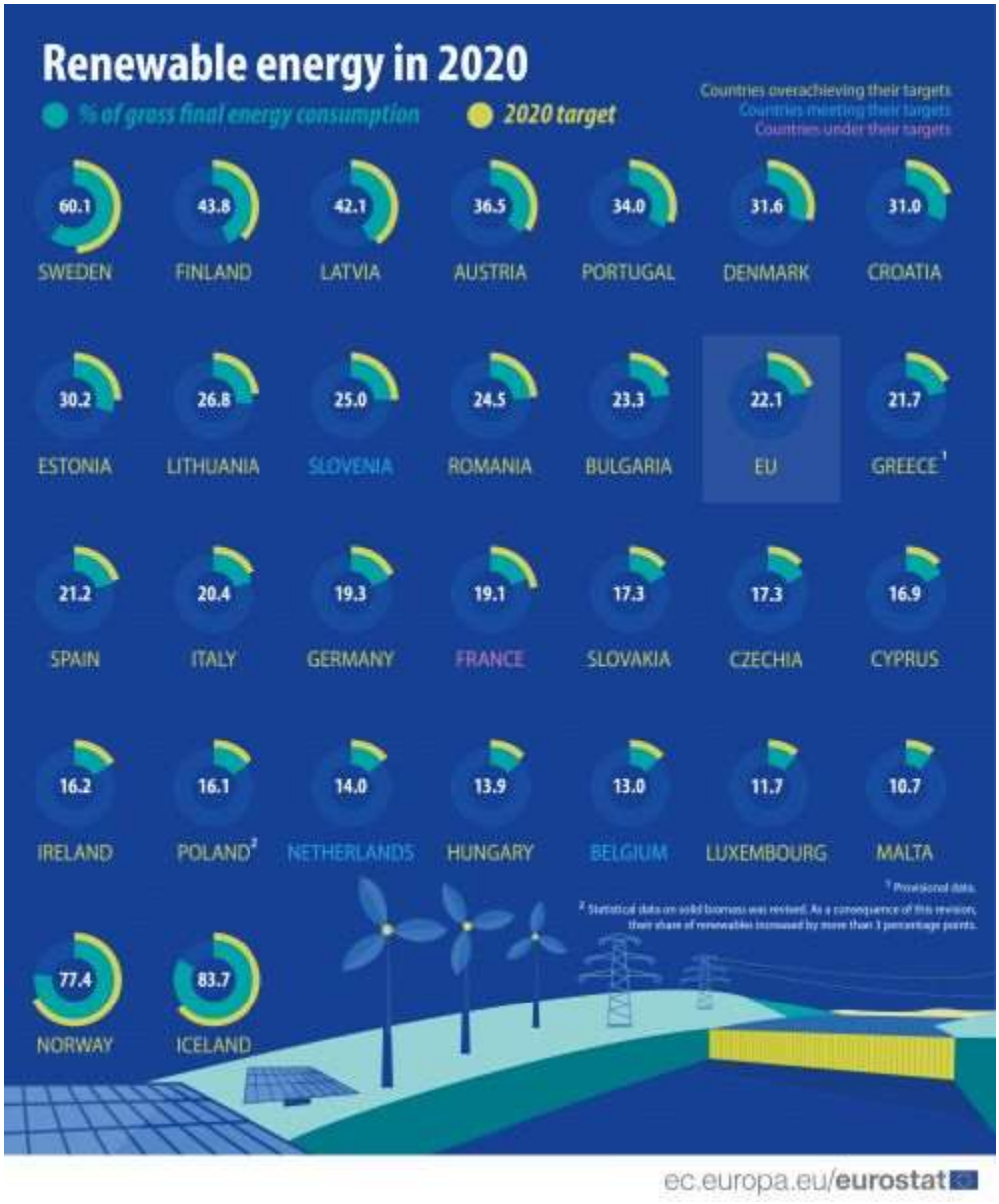

Figure 1-1: RES production share in 2020 and relative target [1].

In Europe, overall, RES cover about 20% of demand, although the overall share of energy from renewable sources is not distributed equally among the various states, as shown in Figure 1.1. In it can be noted that some States such as Sweden, Iceland and Norway exceed 50% of the total share of electricity from renewable energy sources (due to the preponderant share of hydroelectric energy). While in Italy, RES cover about 20% of the annual needs, even if the percentage is destined to rise in recent years thanks to the economic incentives issued by the Government.

Focusing on power production sector only, the renewable share increases. In particular in 2022 the renewable energy production accounted for 41% in the total mix, where the highest contribution comes from hydroelectric (10.6% of the total and 29.6% among renewables) [2].

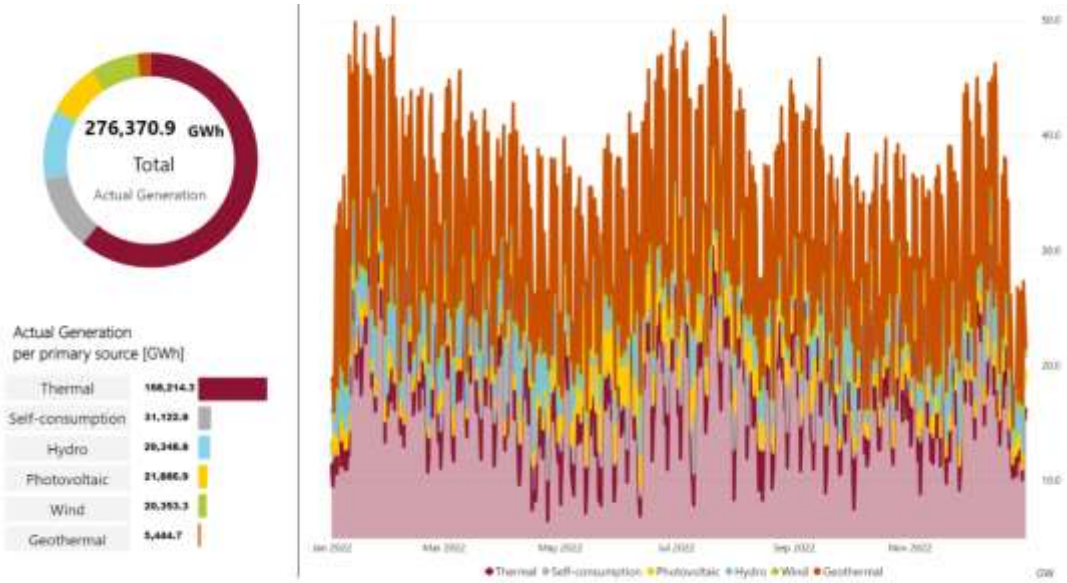

Figure 1-2: Electricity generation in Italy, 2022. Source: Terna [2].

Photovoltaic energy is immediately second to hydroelectric, with a total of 27557.2 GWh of energy produced in 2022 leading to 7.9% of the total and 27.4% among renewables. Figure 1-2 and Figure 1-3 summarize these data.

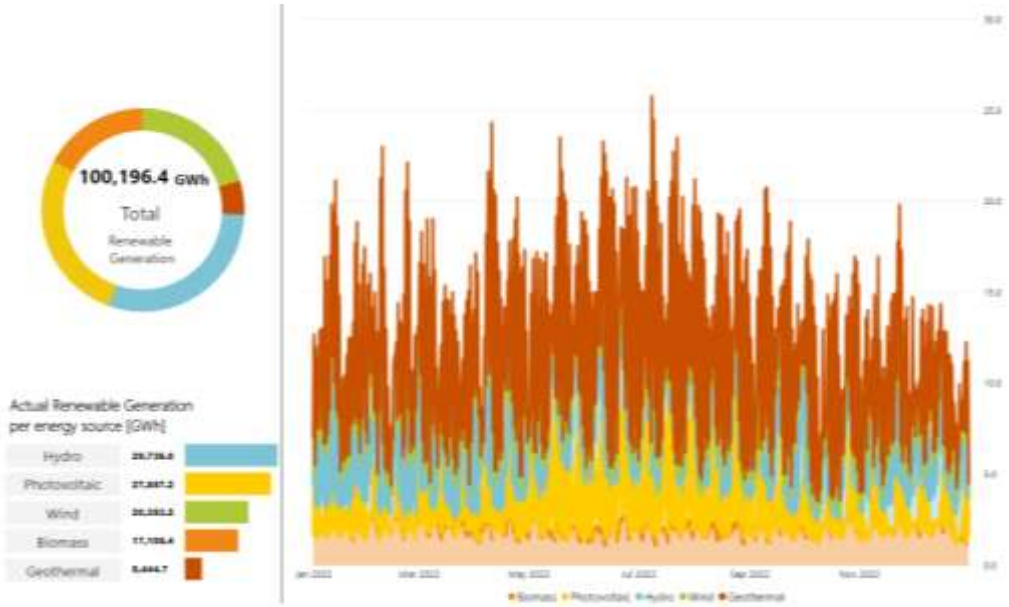

Figure 1-3: Renewable energy generation in Italy, 2022. Source: Terna [2]

Photovoltaic energy is even more relevant in terms of installed capacity, in fact, as shown in the following figure, renewable plants make up 51% of installed power, of which photovoltaics with 24.2 GW of installed power is in first place.

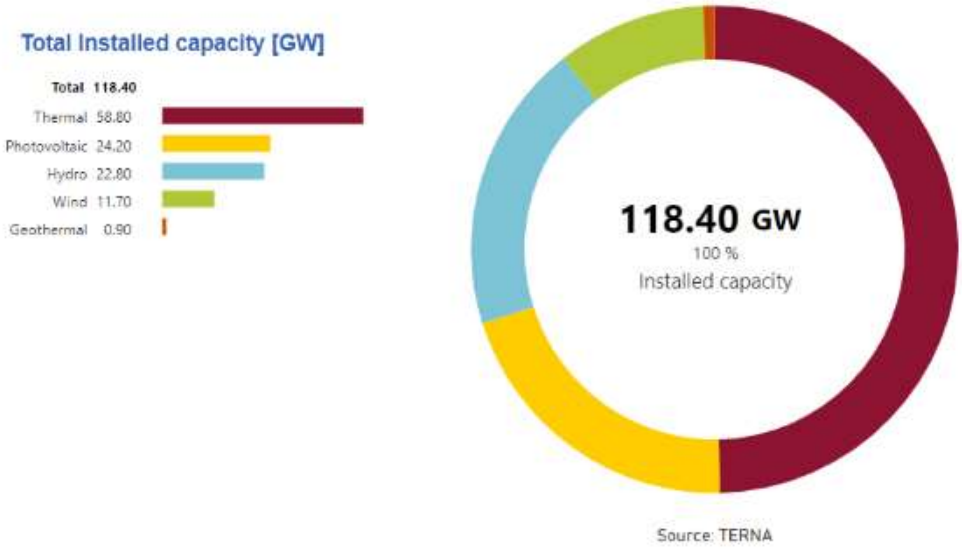

Figure 1-4: Installed capacity in Italy, 2022. Source: Terna [2].

1.1.1 Climate change and counteractions

Climate change is a complex issue, but one of the key drivers is the increasing levels of CO_2 in the atmosphere. CO_2 is a greenhouse gas, which means that it absorbs and traps heat from the sun, making the Earth's surface warmer. As human activity, such as the burning of fossil fuels, has exponentially increased over the last decades, the levels of CO_2 in the atmosphere have also risen, contributing to a warming of the Earth's surface. This phenomenon, highlighted in the Figure 1-5, is known as "Climate Change" or "Global Warming". The relationship between CO_2 emissions and temperature increase is not linear, but there is a clear correlation between the two. As CO_2 emissions have increased, the Earth's average surface temperature has been progressively rising. The Intergovernmental Panel on Climate Change (IPCC) has found that the Earth's average surface temperature has increased by about 1.0°C since the pre-industrial period, with most of the warming occurring over the past 35 years [3].

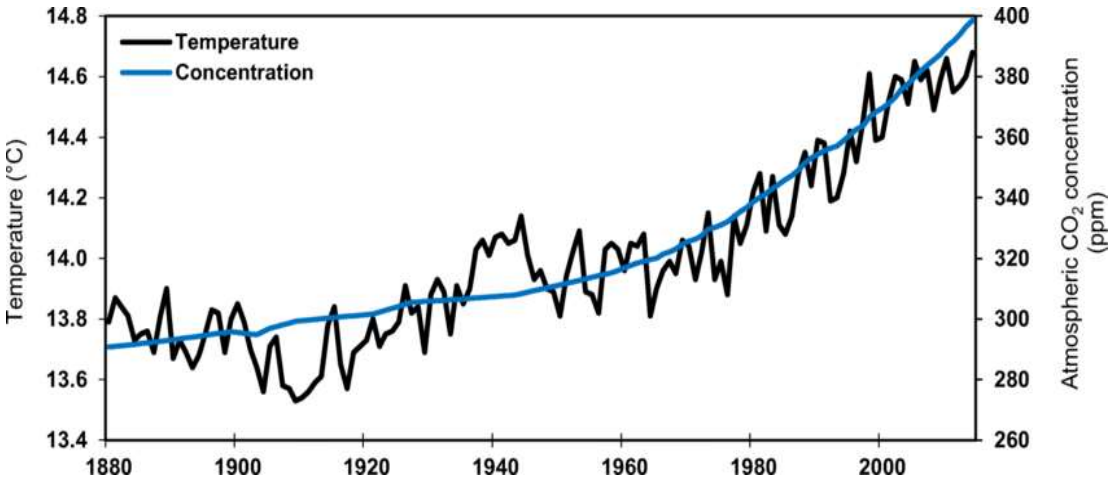

Figure 1-5: Temperature anomalies and CO2 concentration correlation [3].

The United Nations (UN) established the United Nations Framework Convention on Climate Change (UNFCCC) in response to the issue of climate change. This legally binding treaty, which went into effect in 1994, aims to address the issue of greenhouse gas emissions, which are the main contributor to the rising global average temperature. The UNFCCC is in place to encourage cooperation and action among the 197 industrialized countries that are party to it. Each year, the UNFCCC holds the Conference of the Parties (COP) to assess the progress and challenges in tackling climate change.

Figure 1-6: Roadmap of the most important European and global energy policies [4].

Figure 1-6 shows the main EU and international energy policies. It starts with the Kyoto Protocol; it is very important because it is the first legally binding international agreement to set targets for the reduction of greenhouse gas emissions. Prior to the adoption of the Kyoto Protocol, there were no binding international commitments to reduce greenhouse gas emissions, and climate change was largely seen as a voluntary issue [5]. The adoption of the protocol marked a significant shift in the global approach to climate change, and it set the stage for further international action on the issue. It is published in 1997, on the occasion of *COP3* between more than 180 countries, and it enters into force on 16 February 2005 after ratification by Russia. As of May 2013, 192 States have acceded to and ratified the Protocol. It included an obligation to reduce emissions of pollutants (carbon dioxide and five other greenhouse gases) by a non-inferior amount to 8.65% compared to the emissions recorded in 1990 (considered as the base year) in the period 2008-2012. The results of the Kyoto Protocol have been mixed. On the one hand, the protocol has played a significant role in raising awareness about the issue of climate change and has helped to galvanize international action on the issue. It has also led to the adoption of several important measures to reduce greenhouse gas emissions, such as the promotion of renewable energy sources, the adoption of energy efficiency measures, and the development of carbon-neutral technologies. On the other hand, the overall impact of the Kyoto Protocol on greenhouse gas emissions has been limited. While the protocol has helped to reduce emissions in some countries, global emissions have continued to rise, and the Earth's average surface temperature has continued to increase [5].

Following the footsteps of the Kyoto Protocol, The Renewable Energy Directive (RED) is a 2009 European Union directive that sets out binding targets for the use of renewable energy in the EU. The EU has made a commitment, by 2020 to reduce its emissions by 20% below 1990 levels to increase the share of energy derived from renewables to 20% (and 10% in transport sector and to reduce energy consumption by 20% by means of improving energy efficiency [6].

Future objectives were introduced at *COP 21* in 2015 among 196 states, where the *Paris Agreement*, which entered into force in 2016, was defined. In November 2018, 195 members of the UNFCCC signed the agreement and 183 decided to join it. The long-term goal of the Paris Agreement is to limit the increase in global average temperature to well below the threshold of 2 °C above pre-industrial levels, and to limit this increase to 1.5 °C, as this would substantially reduce the risks and effects of climate change.

1.1.2 RES – Integration into electric grids

Following the Bersani decree of 1999, the energy market in Italy was liberalized. Thus, as shown in Figure 1-7, the electrical system is divided into four distinct segments:

- ***Production,*** conversion of energy from a primary source into electricity, through various power plants (thermoelectric, hydroelectric, nuclear, wind, solar); in a competitive framework of the electricity market in many jurisdictions.
- ***Transmission,*** i.e., dispatching of a large amount of electricity (indicatively up to 1000 MVA of power) through HV power lines (220, 380 kV) that have the function of transporting the electrical power from the plants to the load centres (points where the power required by several aggregate loads is concentrated). The lines are interconnected on a continental scale and managed by a single TSO (*Transmission System Operator*), in Italy Terna.
- ***Distribution***, transmission of an inferior quantity of electricity (approximately up to 10 MVA of power) through power lines mainly in MV (sometimes also in LV and HV). The lines are connected and managed on a local scale (tens of km) by a single DSO (*Distribution system operator*).
- ***Sale and use,*** sale of electricity to private and public consumers, in many national regulations is carried out in a competitive structure.

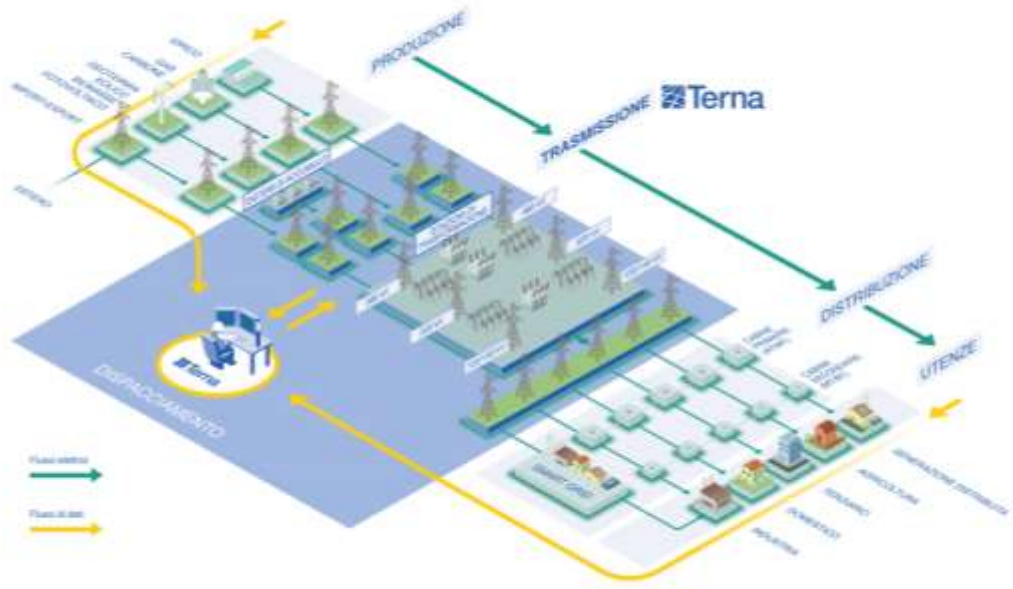

Figure 1-7: Italian electricity system [7].

Terna, in accordance with the provisions of the Autorità di Regolazione per Energia Reti e Ambiente (ARERA) and of Ministero dello Sviluppo Economico (MISE), deals with:

- manage and maintain the high-voltage national electricity grid, investing in infrastructure (transmission system) to improve safety, quality and efficiency.
- maintaining the balance between energy demand and supply (dispatching services).
- contribute to the energy transition by promoting the use of RES in market compatibility.

The use of renewable energy sources (RES) plays a crucial role in promoting sustainability. However, incorporating them into the electricity grid can create challenges related to the security and stability of the system. One way to integrate RES is through large-scale connections to the grid, which allows for the harnessing of these energy sources on a large scale. Another option is through the use of microgrids, which are localized groups of energy and storage sources that can operate both connected and disconnected from the main electricity grid depending on the conditions. Microgrids provide a higher level of security and continuity of

power supply during grid outages, but also make the system more vulnerable and unstable as it is isolated from the main network.

The increased use of renewable energy sources (RES) in the electricity system can lead to instability and challenges, particularly with non-programmable sources where regulation is more difficult. The main challenges can be broadly categorized as:

- bidirectional power flow, it implies new strategies and regulations for various issues such as fault detection and extinction.
- regulation of frequency and active power, in order to make the system stable to perturbations.
- regulation of voltage and reactive power, to maintain the appropriate voltage level in the various nodes of the system;
- The use of inverters tends to reduce the inertia of the system compared to rotating machines, so the same level of stability is not guaranteed.

1.2 RES and energy security

The necessity of having clean energy has been the main driver for the energy transition, but not the only one. The possibility of producing more energy on its own territory reduces indeed energy dependence on foreign countries, thus being more vulnerable to price increase. The issue of the dependence of energy imports and its consequences is an extremely topical topic following the latest trends in electricity and natural gas costs in Europe. Figure 1-8: Early 2021 steep gas price increases across EU . shows the steep increase in gas prices since early 2021 and the consequent effect on electricity prices. Italy has a very poor availability of fossil fuels; therefore, it relies for the vast majority of its energy needs from gas and oil imports, whose price can be very inconsistent. On the other hand, Italy has a great abundance of solar energy, although it is not homogeneously distributed and subjected to intermittencies.

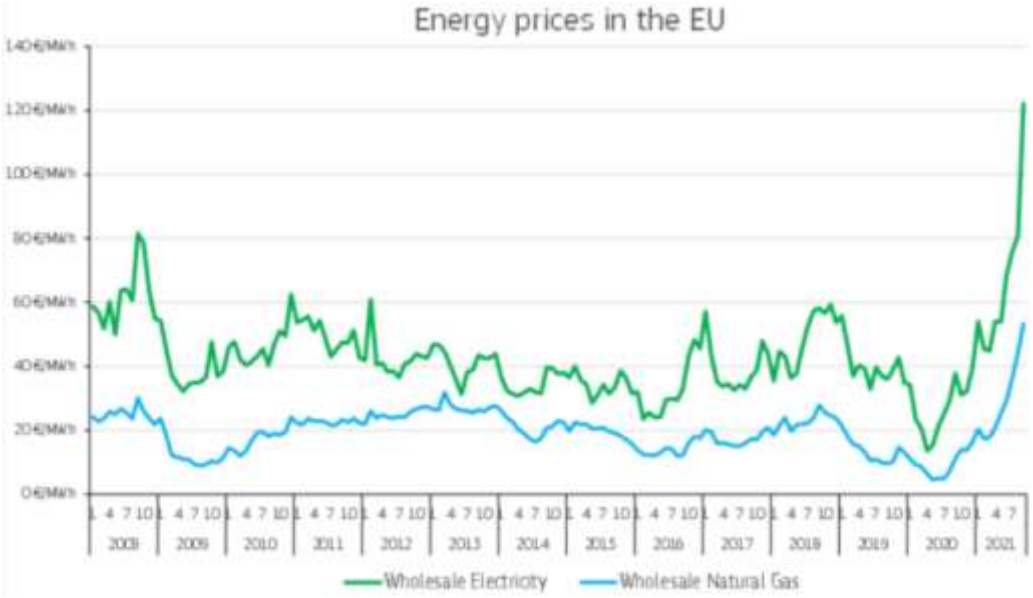

Figure 1-8: Early 2021 steep gas price increases across EU [1].

Following the events of Russia's invasion of Poland last February, the cost of energy has further increased. These geopolitical events remind us once again how crucial it is to invest in renewable energy sources, which, even without considering the environmental aspect, are dispersed sources in contrast to fossil fuels.

1.3 Solar photovoltaic systems

1.3.1 Solar radiation and photovoltaic effect

PV effect is a physical process in which a PV cell converts the solar irradiance into electricity. From a radiative point of view, the Sun can be approximated as a black body at a temperature of 5800 K, therefore its emission spectrum, i.e., the distribution of the energy intensity of the radiation as a function of the wavelength, is described by Plank's law. Wavelengths are associated with the energy carried by the individual quantum of which solar radiation is composed. As it passes through the atmosphere, a fraction of the sun's rays is absorbed or scattered as a result of collisions with molecules in the atmosphere (including water vapor, clouds and aerosols). Scattering affects all wavelengths of the spectrum and causes an attenuation of the spectrum, while absorption is a discrete phenomenon that has resonances in certain wavelengths determined by the molecule in question. As a result, the electromagnetic spectrum assumes an irregular profile. Figure 1-9 shows

the irradiance spectrum on top of atmosphere (Air Mass 0) and after being attenuated by the atmosphere (Air >Mass 1).

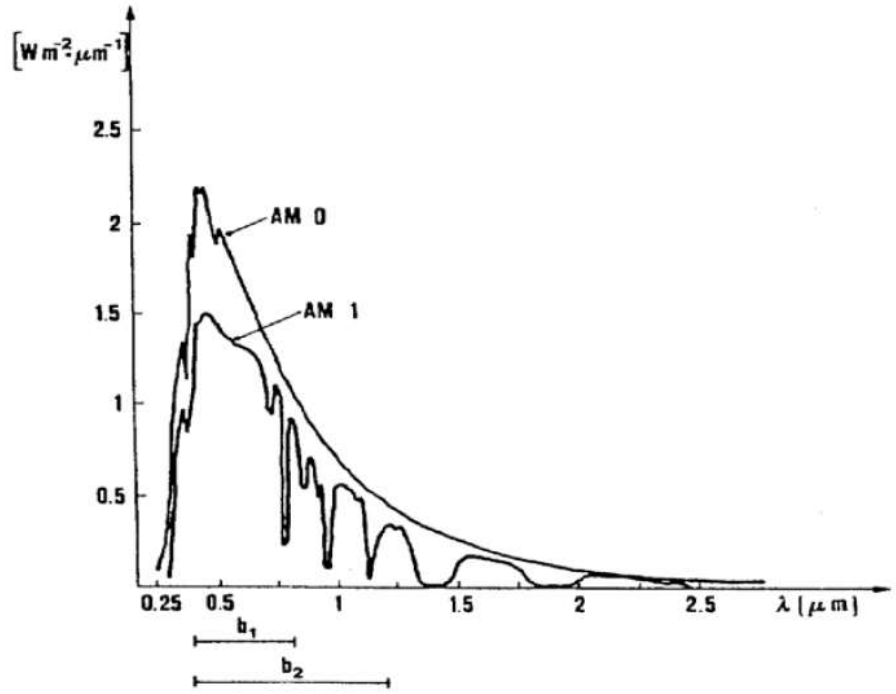

Figure 1-9: Solar spectrum, for AM = 0 and AM = 1 [8].

Each photon reaching has an amount of energy determined by its wavelength:

$$E = \frac{hc}{\lambda}$$

where c is the speed of the light, h is Planck's constant and λ is the wavelength of the photons. The photovoltaic effect is based on the characteristics of semiconductors, materials in which there are a valence band, composed of electrons involved in chemical bonds, and a conduction band, consisting of electrons with a higher energy level to allow their movement in the lattice. The energy required by an electron to be promoted by the conduction valence band is called the Energy Gap. The photons of the solar spectrum with an energy higher than the energy gap of the semiconductor material are able to promote an electron in the conduction band, while in the valence band a hole is left, creating electron-hole pairs. These electron-hole pairs create an electric flow, that is, electric current through an external electric circuit connected to the PV cell terminals. The built-in electric field

that is a specific electric feature of the PV cells provides the voltage potential difference that drives the current through an external load. Two layers of different semiconductor materials are placed in contact with each other in order to induce the built-in electric field within a PV cell. The first layer that is n-type has abundance of electrons; the other layer that is p-type has abundance of holes. The abrupt change in doping and in the lattice activate diffusion mechanism, with the result that excess electrons move from the n-type side to the p-type side. As a result, a positive charge is built up along the n-type side of the interface and negative charge along the p-type side. Thus, an electric field is created at the surface where the layers meet, called the p/n junction.

1.3.2 PV technologies

The single photovoltaic cell delivers a current density of 25÷35 mA/cm^{2}and generates a voltage level of 0.5÷0.6 V. The commercial unit of the photovoltaic system consists of the module, consisting of the connection in series of several cells. The most used semiconductor in the realization of PV modules is silicon in the crystalline and multi-crystalline versions, as reported inFigure 1-10.

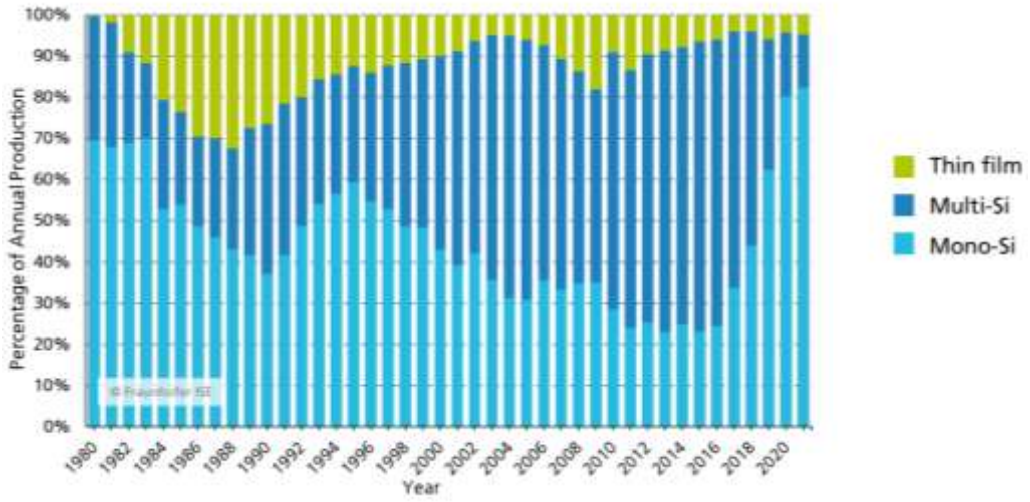

Figure 1-10: market trend shares of PV technologies [9].

Monocrystalline silicon has a higher efficiency (up to 22%) thanks to its better spectral response, in particular it has the widest bandwidth for which it can generate current, and therefore a higher percentage of photons can be exploited to create electron-hole pairs. Poli-crystalline silicon on the other hand has a lower efficiency (14-16%) but lower production costs. As shown in the figure, these two

technologies make up more than 90% of the annual production of recent years. Thin film technologies can be based on silicon (in the monocrystalline, multi-crystalline and amorphous version) or on non-silicon-based semiconductors. The most common technologies in the thin film version are the following:

- Cadmium telluride (CdTe): with a commercial efficiency of 13-15%.
- Copper, indium, gallium, selenide (CIGS): with a commercial efficiency of 11-13%.
- Amorphous silicon (a-Si) and tandem cell (a-Si/ μc-Si): where a layer of microcrystalline silicon (μc-Si) is combined with amorphous silicon. The efficiency is around 6-10%.

1.3.3 Equivalent circuit of a solar cell

An ideal solar cell may be modelled by a current source in parallel with a diode, whose current output is proportional to the irradiation. However, in practice there are losses, so a shunt resistance and a series resistance component are added to the model. The resulting five-parameters equivalent circuit is represented in Figure 1-11. Series resistance R_s accounts for resistances of electrodes and semiconductors, while R_{sh} is due to alternative paths for the current and it is reduced by improving edge insulation.

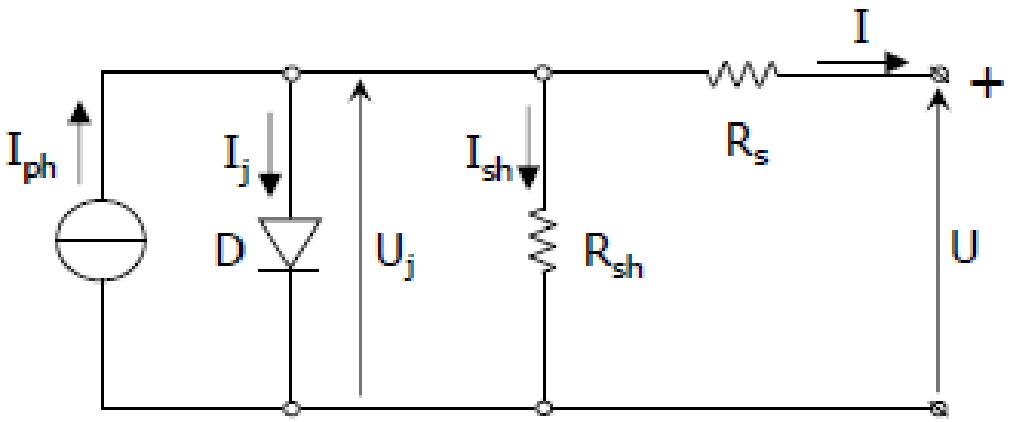

Figure 1-11: 5-parameters equivalent circuit of a PV cell [10]

The current balance equation is the following:

$$I = I_{ph} - I_j - I_{sh} \text{ (1-1), where:}$$

- I = output current.
- I_{ph} = photogenerated current. Since it depends on the material and it is proportional to the irradiance G and the area of the cell A, it can be expressed with the following equation:

$$I_{ph} = k_{mat} \cdot G \cdot A$$

- I_j = diode current, it takes into account electron-hole recombination effects occurring inside the cell. It is modelled using the Shockley equation for an ideal diode:

$$I_j = I_0 \left[exp\left(\frac{qU_j}{mkT}\right) - 1 \right]$$

  I_0 is the saturation current of the diode, q is the elementary charge, U_j is the voltage across both diode, k the Boltzmann constant, T the absolute temperature and m the diode quality factor.
- I_{sh} = shunt current, which is calculated simply by applying the Ohm law across the shunt resistor terminals.

Combining all the previous equations, the resulting I-V curve, is shown in the following equation and figure. I_{ph}, I_0, m, R_s and R_{sh} are the five parameters of this equivalent circuit.

$$I(V) = I_{ph} - I_0 \cdot exp\left[q\left(\frac{V + R_s \cdot I}{mkT} - 1\right)\right] - \frac{V + R_s \cdot I}{R_{sh}}$$

Figure 1-12 contains the I-V curve described by the previous equation. The most important three points of the curve are highlighted:

- The short-circuit current, I_{SC}, is the current through the solar cell when the voltage across the solar cell is zero (i.e., when the solar cell is short circuited). The short-circuit current is due to the generation and collection of light-generated carriers. The short-circuit current is the largest current which may be drawn from the solar cell [10].
- The open-circuit voltage, V_{OC}, is the maximum voltage available from a solar cell, and this occurs at zero current. The open-circuit voltage corresponds to the amount of forward bias on the solar cell due to the bias of the solar cell junction with the light-generated current. The open-circuit voltage is shown on the I-V curve [10].
- The maximum power point, P_{MPP}, is the point on the I-V curve of a solar module, where the product of current and voltage reaches its maximum

level. The points on the I and V axes which associated to this curve are named I_{MPP} (current at maximum power point) and V_{MPP} (voltage at maximum power point). PV electronics is equipped with maximum power point tracking (MPPT), a technology used with variable power sources to adjust the working point on the I-V curve in order to extract the maximum amount of energy under different ambient conditions [10].

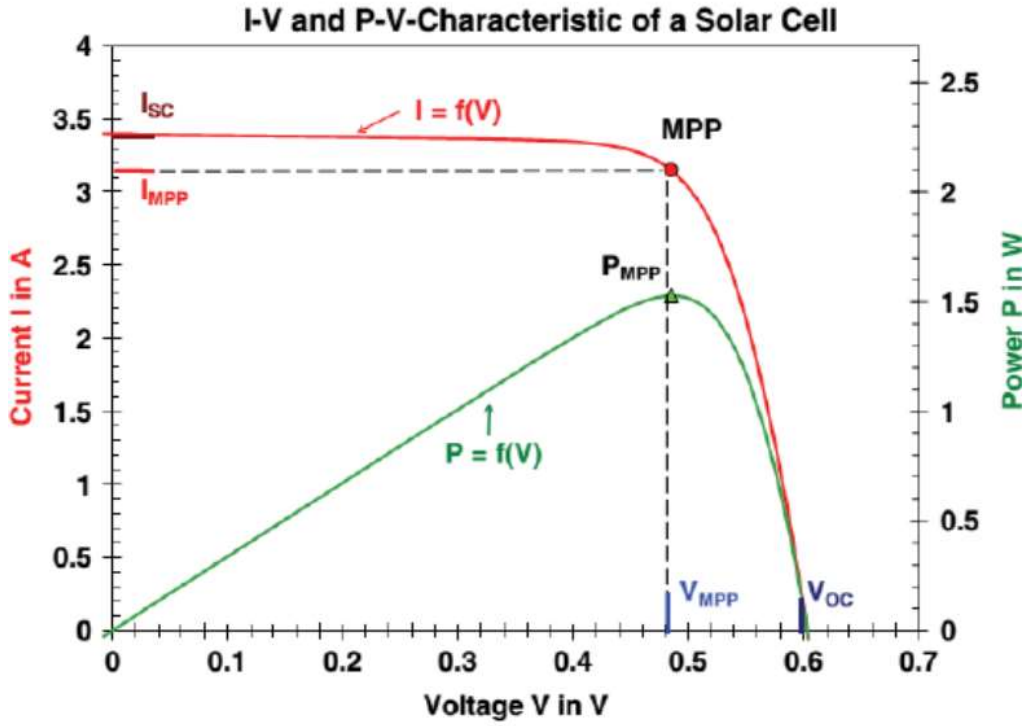

Figure 1-12: I-V curve of a PV cell [8].

1.3.4 Dependence on irradiance and temperature

With the same given amount of irradiance, an increase in cell temperature corresponds to a slight increase in I_{PH} and I_{SC}, while the decrease in U_{OC} is more remarkable, as reported in Figure 1-13.

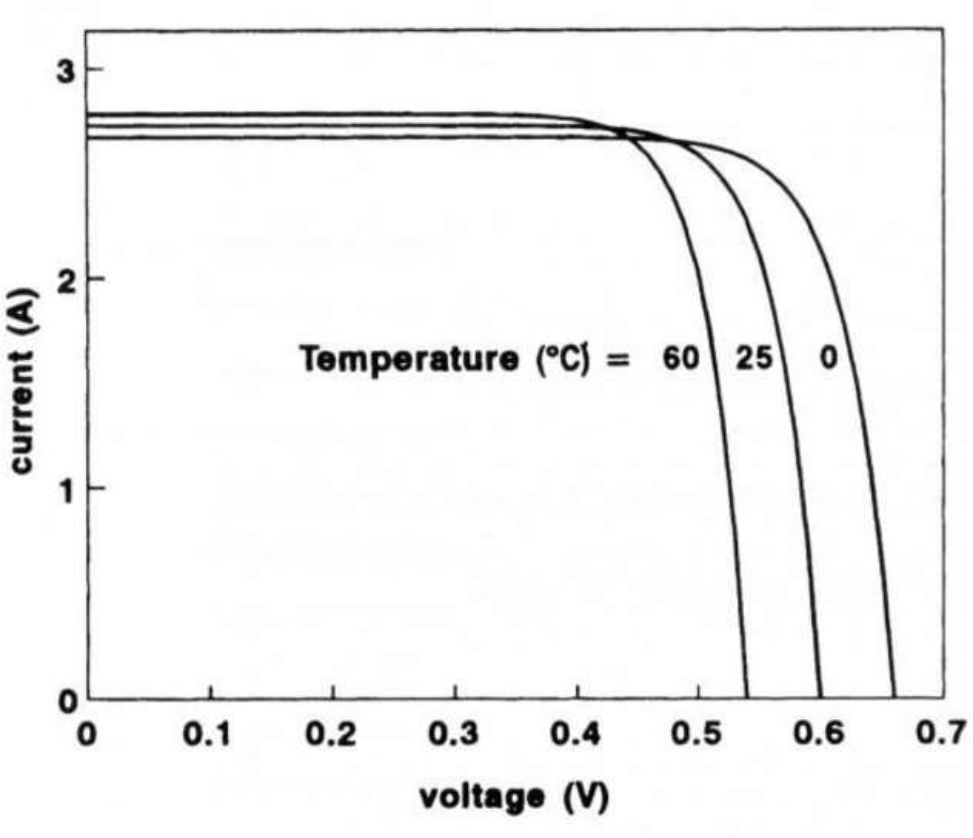

Figure 1-13: effect of temperature on I-V curves [8].

The effect of irradiance, at constant cell temperature, is instead much more marked in the case of current, which depends linearly on the irradiance G. With a larger irradiance value, a greater number of photons that can generate current is indeed available. The open circuit voltage U_{OC} increases too, but at a logarithmic rate. The effect of irradiance is shown in Figure 1-14.

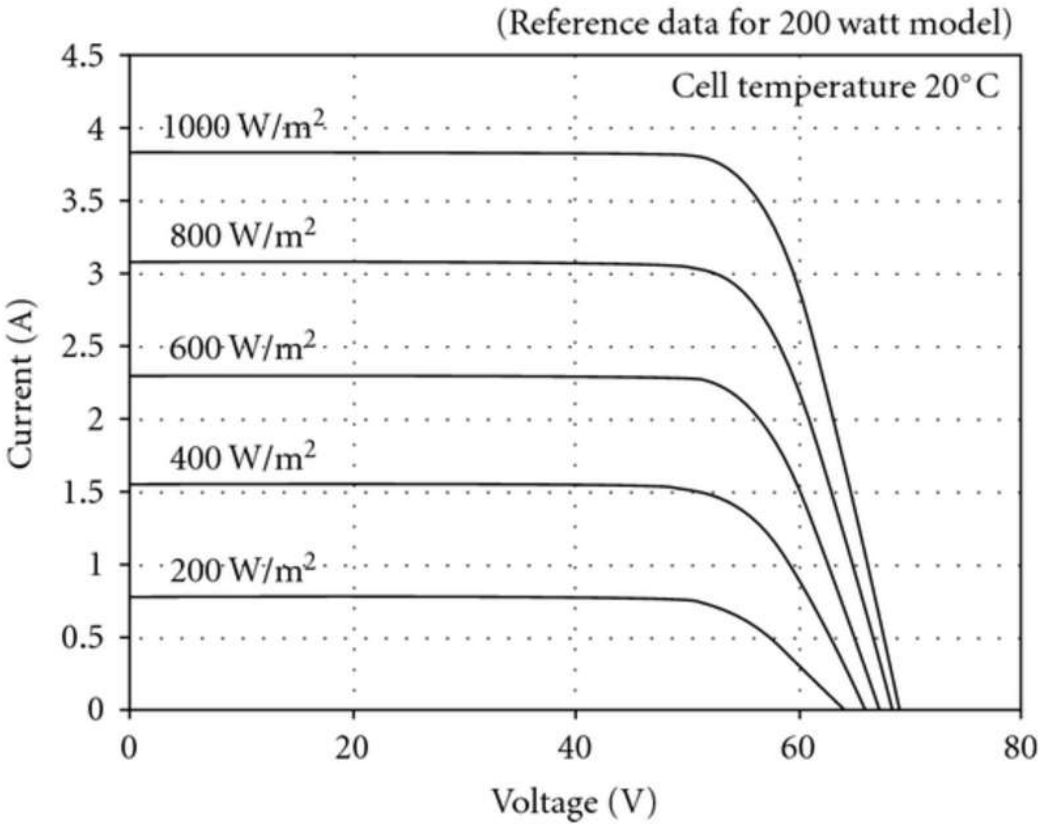

Figure 1-14: Effect of irradiance on I-V curves [8].

In most applications, it can be assumed that I_{SC} and current depend only on irradiance, while voltage and U_{OC} depend only on temperature. The resulting power produced increases linearly with the irradiance and decreases when the temperature of the cell increases. In many applications, in order to account for voltage reductions due to high temperature, the power thermal coefficient γ_{th} is used. It is defined as:

$$\gamma_{th} = \frac{dP_{Max}}{dT_{PV}} \cdot \frac{1}{P_{Max}}$$

Its value depends on the material used. In the case of crystalline silicon it is equal to 0,5 % $°C^{-1}$.

1.3.5 Series and parallel connection of solar cells

To meet the power demands of electric loads, multiple solar cells must be connected in series or in parallel to produce higher voltage and current levels than

a single cell can provide. By using multiple cells, the overall power output of a solar panel or system can be increased to match the requirements of the load.

A series connection of PV cells is a method of connecting multiple cells together in order to increase the overall voltage output of a PV system. In this configuration, the current flows through each cell in a single path. The voltage of the overall system is equal to the sum of the individual cell voltages, while the current remains the same. However, there are also some disadvantages to using a series connection. One of the main issues is that the overall system is more sensitive to shading or failure of a single cell. If one cell in the series is shaded or fails, the entire circuit is affected, and the voltage output of the entire system is reduced. This can result in a significant reduction in the overall power output of the PV system.

In parallel connection configuration, cells or modules are connected in parallel, with the positive terminal of one cell or module connected to the positive terminal of the next, and the negative terminal of one cell or module connected to the negative terminal of the next. The voltage of the overall system remains the same as that of a single cell or module, while the current is equal to the sum of the current produced by each PV module. As highlighted in the next paragraph, parallel connection is commonly used only for modules strings already in series connection. One of the main issues of using parallel connection is that the overall system is more sensitive to mismatches in the performance of the individual cells or modules. If one cell or module has a lower voltage or current output than the others, it will "drag down" the entire circuit, resulting in a reduction in the overall power output of the PV system. A blocking diode is connected in series to groups of strings to prevent defective cells working as a load with reverse current.

Mismatch losses refer to the reduction in power output that occurs when the individual PV cells or modules in a PV system have slightly different performance characteristics. This can happen due to variations in the manufacturing process, changes in environmental conditions, or aging of the cells or modules. Mismatch losses can occur in both series and parallel connections of PV cells and modules, but they are generally more pronounced in parallel connections. In the overall PV system, losses due to mismatch are around 2-4% [10].

1.3.6 PV field structure

The photovoltaic system, also known as field, is suitably formed starting from single cells that form a module and then through the electrical connection of

multiple modules in strings (or arrays) with powers ranging from kW to MW. Figure 1-15 shows the steps for the realization of a photovoltaic field starting from the single cell.

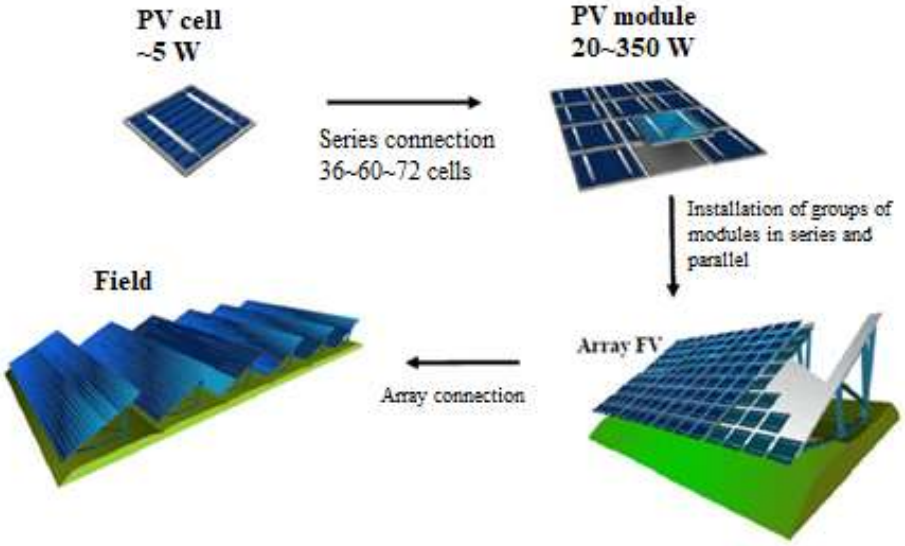

Figure 1-15:PV field composition [10].

In turn, photovoltaic generation systems can be:

1. Isolated or stand-alone systems, if the field is not connected to the mains and feeds a load directly. In this case, a storage system is used in order to use the energy accumulated in the moments of non-production.
2. Systems connected to the grid or grid connected in case the field is connected to the power grid.

In addition, in the case of connection of the photovoltaic system in the grid, in Italy, the following rules must be complied with:

1. CEI 0-16 standard, for connections to HV and MV networks.
2. CEI 0-21 standard, for LV connections.

The general network pattern is highlighted in figure 1.16.

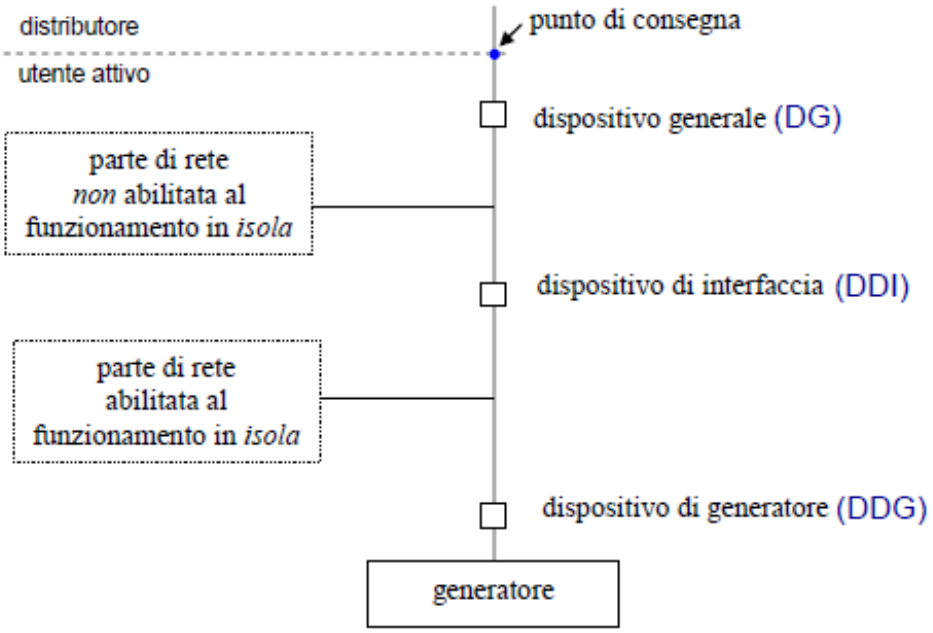

Figure 1-16: general scheme of grid connection. Source: [11].

1.4 Power production forecasts

In accordance with the objective of increasing the production of electricity from renewable sources such as photovoltaic production, it becomes necessary to be able to adequately estimate the production profiles of PV plants. The ability of precisely forecasting the energy produced by PV systems is of great importance and has been identified as one of the key challenges for massive PV integration. It is decisive for grid stability, since deviations between forecasted and produced energy must be supplied by the rest of technologies that form the energy portfolio. Some of the units that build the electric system act as operating reserve generators. Moreover, also from an economic point of view, a proper PV forecast would be able to lower the number of units in hot standby and, consequently, reduce the operation costs.

Power forecasting, therefore, is fundamental for non-programmable RES such as wind and photovoltaic to avoid imbalances. The main way in which forecasts can be classified is according to time horizon as summarized in table 1-1.

Table 1-1: Temporal horizon forecasts classification and related purpose

Type of forecast	Time Base	Application
Intra-daily	5÷60 min	Adjustments/Dispatching Market clearing Contingency analysis
short term	1÷6 h	Programming Congestion management
Medium term	Days	Programming/Reserves Congestion management Trading

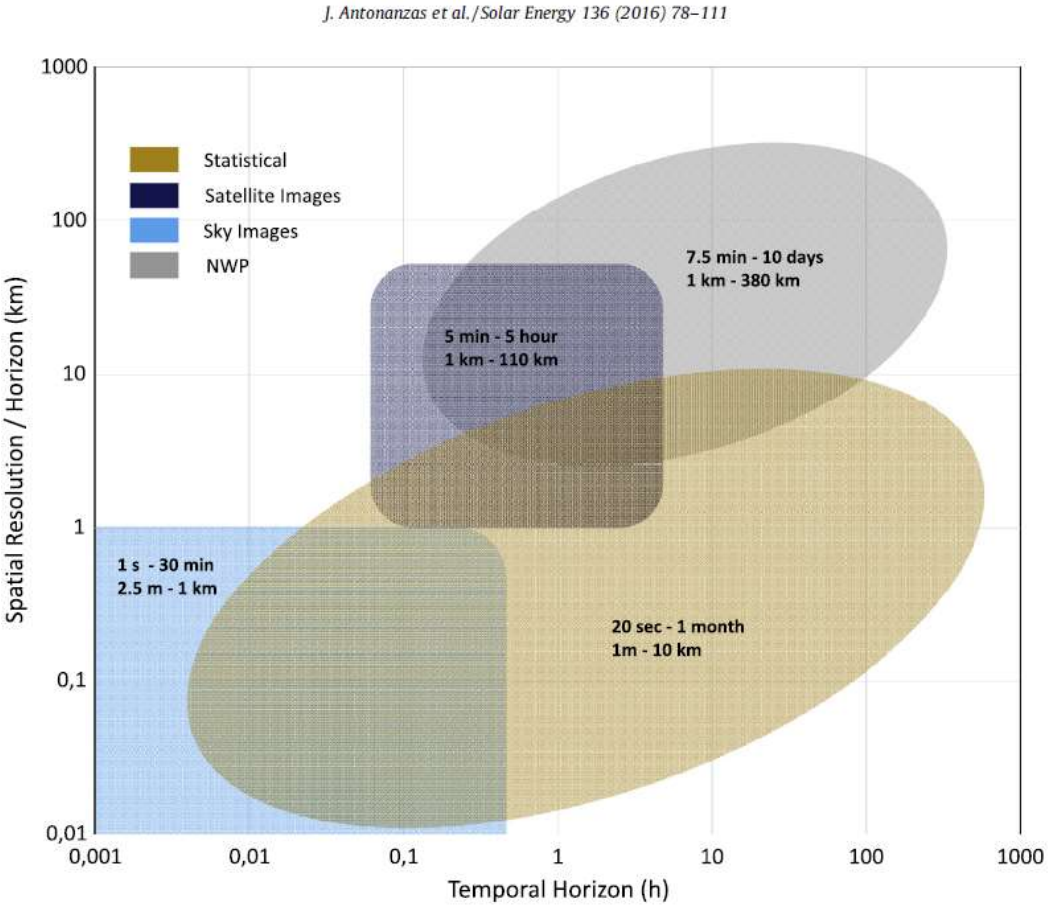

Figure 1-17: Distribution of different techniques and sources of inputs with respect to their spatial resolution/horizon and the temporal horizon for which they are used in power forecasts [12].

Another important distinction among forecast power models is according to the type of model which is used. In particular they can be physical, statistical or hybrid models.

Physical models use the so-called exogenous data (temperature, wind speed and direction, irradiance cloud cover, …), which may come from local measurements, information from total sky imagers, satellites images, numerical weather prediction (NWP), values from other meteorological databases and neighbouring plants. Then a PV performance model with its analytical equations is applied to generate PV power forecasts. For these reasons this approach is also referred as "white box" method [12].

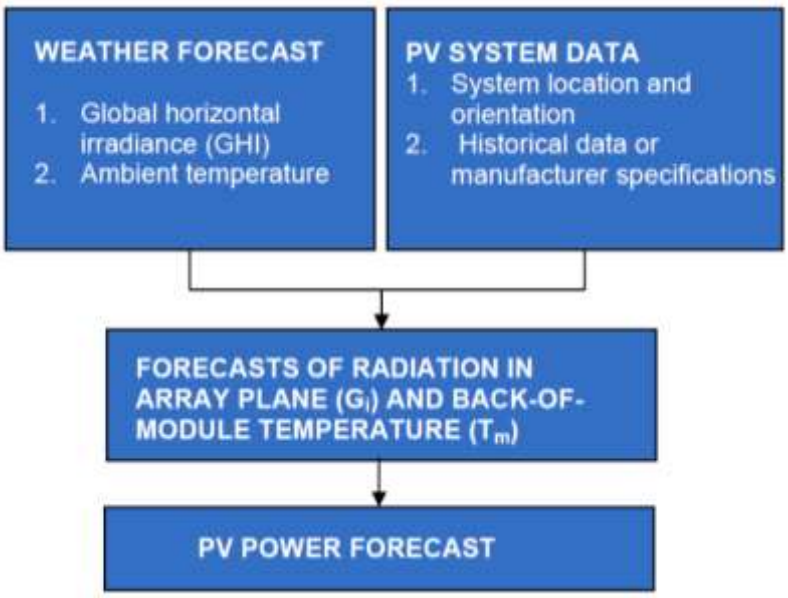

Figure 1-18: PV performance model concept [13].

Figure 1-18 shows in a very simplified way the PV physical models approach. In [12]a general comparison between physical and performance models is made. With reference to the previous figure, the output is a PV forecasts as a function of relevant meteorological variables and PV plant specifications (like azimuth and tilt of PV modules). PV output power is affected by two main variables: the irradiance in the plane of the PV array, Gi, and the temperature at the back of the PV modules, Tm. The type of PV determines the relevant irradiance; non-concentrating PV requires global irradiance in the array plane, while concentrating PV requires direct normal irradiance. While some PV models consider other variables, like the

incidence angle of beam irradiance and the spectral distribution of irradiance, accurate results have been achieved without these factors. PV models can be developed using historical data or manufacturer information, depending on the availability of data. Weather forecasts do not provide Gi or Tm, so solar and PV models are used to calculate them based on PV system specifications and weather forecasts, like global horizontal irradiance (GHI) and ambient temperature. These models serve as an intermediary step. Temperature of PV modules can be calculated using PV system specifications, irradiance, ambient temperature and wind speed if available [14].

The PV performance method has a crucial advantage over the statistical approach in that it enables power output predictions for a plant before construction, without relying on historical data. However, a significant disadvantage of parametric models is their heavy dependence on NWP, which is often characterized by insufficient spatial and temporal resolution and is a major source of error in this approach. Notably, while the error reported for plant modelling was 1.2%, incorporating irradiance predictions raised the error to 10% [14]. To address such errors, Model Output Statistics (MOS) can be utilized to enhance the temporal resolution of weather forecasts. Nonetheless, MOS requires historical weather data, which may not be readily available, thereby reducing the benefits of PV performance models. Furthermore, since each PV performance model is specific to a site and relies on certain technical equipment specifications, simplifications must be made, which can impact the model's accuracy [14].

Statistical models rely primarily on endogenous data formed by past data to train models, with little or no reliance on solar and PV models. This is a methodology that relies on data to extract patterns from past records in order to forecast future plant behaviour. The power output can be directly calculated without the need for meteorological predictions, hence it is sometimes called the "direct method." Unlike physical models, statistical techniques are utilized, resulting in a "black box" model. Therefore, high-quality historical data is necessary for precise predictions. In contrast to the parametric approach, a large historical dataset is usually required, assuming that the plant has already been operational for a while. This approach has the advantage of correcting systematic errors related to input measurement. Selecting an appropriate training dataset is critical for achieving high accuracy in the resulting model [12].

Hybrid models approach: to improve accuracy, it is often preferable to combine different techniques, resulting in hybrid, blended, combined, or ensemble models.

This approach can be divided into two subcategories: two or more statistical techniques (hybrid-statistical) can be combined or incorporating a statistical technique into a PV performance model (hybrid-physical) [12].

Many statistical models rely on Artificial Intelligence, that can be described as simulation of human intelligence processes by computer systems. These processes include the ability to learn from data, react to external inputs and based on these data, make decisions and adapt to new situations. AI is a broad field that includes multiple sub-disciplines, including machine learning, that involves the development of algorithms and statistical models that enable computers to learn from data and improve their performance. It is a method of teaching computers to learn from examples, patterns, and experiences, instead of being explicitly programmed with a set of rules.

Artificial Neural Networks (ANN) are the most used machine learning techniques in solar power forecasting [12]. An Artificial Neural Network (ANN) is a computational model inspired by the structure and function of the human brain. It consists of layers of interconnected "neurons," which process and transmit information. These neurons are organized into layers, with the input layer receiving data, one or more hidden layers processing the data, and the output layer providing the final result or prediction [15]. The basic building block of an ANN is the artificial neuron, also known as a node or unit, which is a mathematical function that receives input, performs a computation, and produces an output. The input is passed through a set of weights, which are adjusted during the training process to optimize the performance of the network. The training process of an ANN involves adjusting the weights and biases of the network based on a training dataset, which helps the network to learn how to produce the desired output for a given input. The training process is done using supervised learning, where the network is provided with examples and the goal is to learn to produce correct outputs for the inputs. Each ANN is made of an input layer, a number of hidden layers and an output layer. The main function of the hidden layer is to extract features or representations of the input data that are useful for the task at hand. It does this by applying a series of mathematical operations, known as activation functions, to the input data. Activation functions are non-linear transformations that introduce non-linearity into the network, allowing it to model complex relationships between the inputs and outputs [15]. Figure 1-19 shows how the three layers are structured.

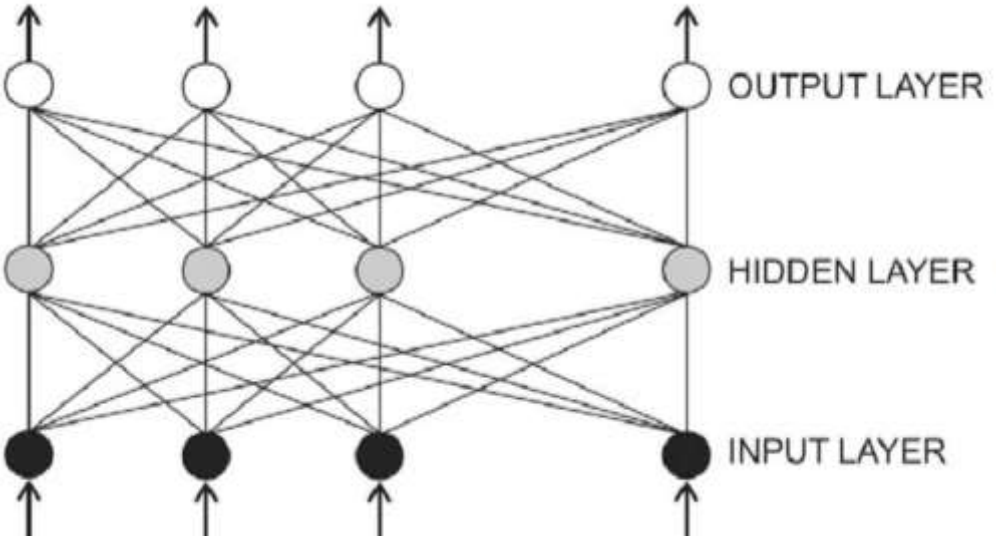

Figure 1-19: Simple forward ANN [15]

A large number of ANNs topologies exists and the classification by the number of hidden layers is the main one.

Additionally, forecasts can also be made on different spatial basis. Forecasts can be made for a single PV system or for an ensemble of them. Normally, grid operators prefer regional forecasts since they are more useful to keep the balance between demand and supply in the electric system. Variability in power output is reduced when an ensemble of plants is considered, since the forecast error increase with the variability of signal to forecast.

For example, in [16] they studied an approach to predict regional PV power output based on irradiance forecasts provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). They evaluated the forecast error for the singe site, an ensemble covering the area of 220x220 km and the whole area of Germany. A temporal averaging procedure is then applied. They tried different approaches; best results were obtained combining the forecast data with a clear sky model and then corrected systematic deviations with a bias correction. The evaluation of the PV power prediction scheme resulted in an RMSE of 0.11 kW/kW_{peak} for single systems. For the ensemble power prediction for an area of 220 km x 220 km an RMSE of 0.06 kW/kW_{peak} was found, and for a larger ensemble covering the area of Germany the RMSE of 0.05 kW/kW_{peak} [16].

In a similar way, in [17] a study to propose a method to obtain one-day ahead hourly regional forecasts of photovoltaic power for a regional area. Two kinds of data are

necessary for the application of the forecast methods. Measured regional PV power generation used in the training and configuration of the forecast methods and the data used as input information for the forecasts. For this study, up to 33h ahead forecasts done at 03h UTC were used as input data to forecast the regional PV power from 06h to 19h of the next day [17]. Forecasts for the air temperature, relative humidity and cloudiness (in three levels) provided by the GPV-MSM weather forecast system were used as input data. The simplified equation of time was used to calculate the extraterrestrial solar irradiance. The aim of the study was to assess the efficacy of the proposed method which utilized principal component analysis (PCA) and support vector machines. PCA was employed to select appropriate weather forecast data on a regional scale, and it proved to be a valid option, as it produced better results than persistence and non-PCA methods in evaluations conducted in four regions of Japan. Additionally, the results indicated that the use of PCA reduced the forecast error in cases where there were a high number of PV systems and input variables, which often led to learning problems for the forecast algorithm. This approach also offered the advantage of requiring only regional monitoring of PV power generation, rather than individual PV systems [17]. The performance of the proposed method was similar to that of a method which utilized regional smoothing, but was outperformed only when data from the largest region with a variety of climates was used. In such cases, it may be more appropriate to obtain regional forecasts from individual PV power generation monitoring, if such data aggregation is feasible. Future studies should explore the potential gains in forecast accuracy resulting from the combination of PCA and smoothing techniques in large size regions. Additionally, other methods for obtaining regional forecasts should be investigated to improve the understanding of the applicability and validity of the proposed method. Finally, the evaluations showed that the use of PCA yielded low bias forecasts and good forecast accuracy; the annual RMSEn was near to 0.07 kWh/kW_{rated} in the 4 regions studied [17].

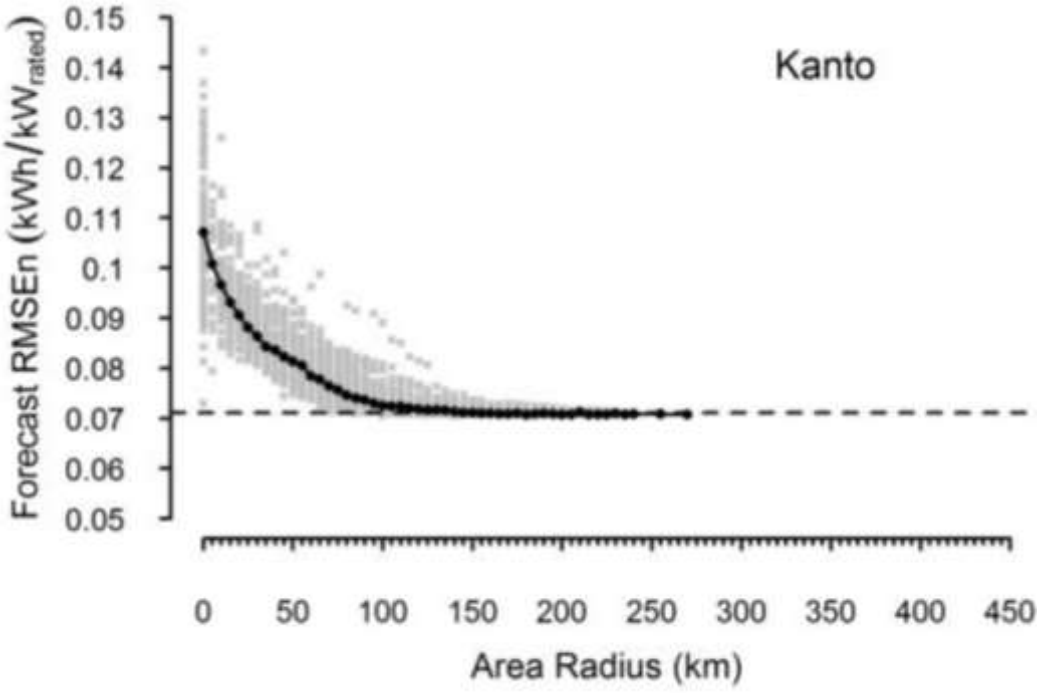

Figure 1-20: smoothing effect of power forecast for an ensemble of PV plants for a regional scale [17].

Finally, independently on the metrics used to assess the performance of the proposed models, there are some other factors that hamper comparisons among studies:

- Climatic variability: High climatic variability normally leads to higher forecast errors than areas with a more stable climate.
- Day/night values and normalization: To make a fair comparison between studies it is important to state clearly which time frame has been taken into consideration and whether only daylight values.
- Spatial aggregation: it reduces the ensemble error, so regional results cannot be compared to single site results.
- Testing period: Some authors tested over a long period of time covering all sky conditions, while other authors tested their models on either only sunny days or only cloudy days, which also increases difficulty to perform comparisons.

1.4.1 Forecast error consequences

Forecasting models are essential in minimizing the discrepancy between projected and actual power profiles. Accurate forecasts benefits have a significative impact both on:

- The Transmission System Operator (TSO), which would have a safer and easier grid management. Moreover, the high penetration of solar energy with its intrinsic variability, leads to the necessity of operating reserves and induced ramps, associated to higher costs and lower system stability [12].
- The plant owner, as with more accurate forecasts they can generate more precise bids and the avoid the risk of PV curtailment.

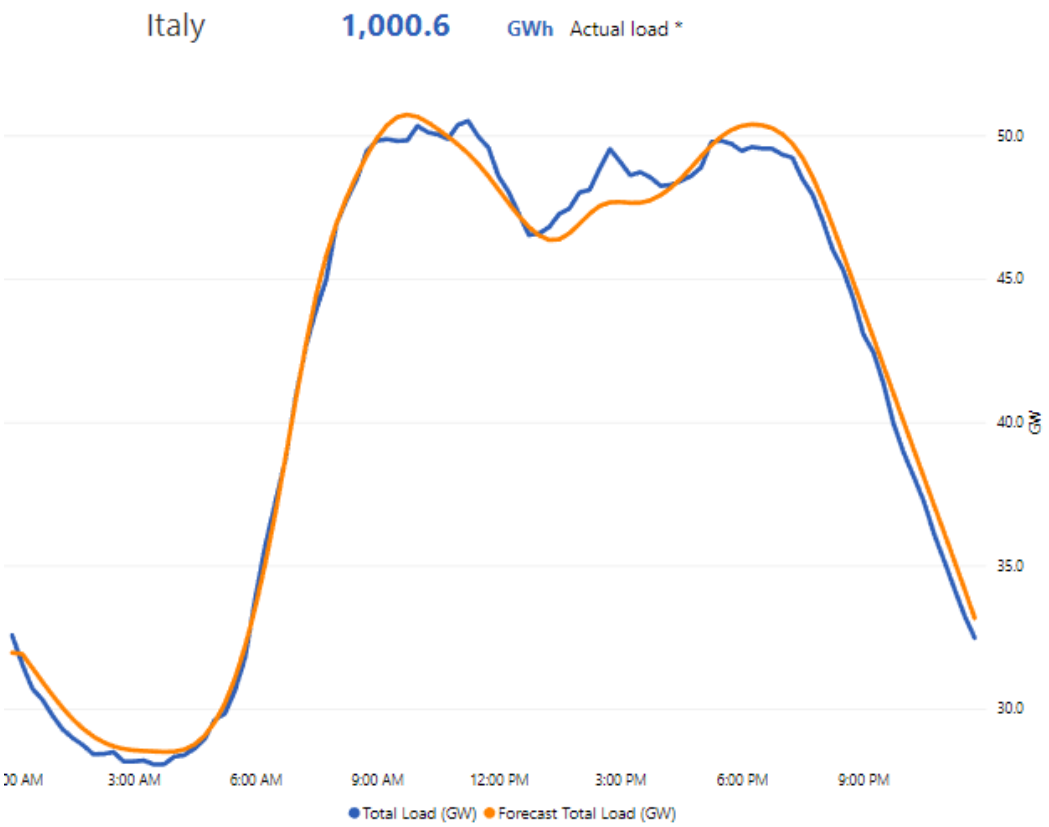

Figure 1-21: day ahead forecast load for Italy [18].

Forecasts are generated every day for the following day, by the TSO and GSE (Gestore Servizi Energetici). A load profile on the national grid is predicted as shown in Figure 1-21. It is possible to notice that actual load is relatively close to

actual value, where you can appreciate how the prediction is almost perfect, both under "*base* load" and "*peak load*" conditions. However, there may be slight imbalances between the two profiles, due to variations in energy consumption by consumers and/or fluctuations in energy production from non-programmable sources like photovoltaics.

In most cases the error in load forecasts is lower than uncertainties in energy production, when renewables energy resources represent an important share in the energy mix. Neglecting the load forecast error in comparison to forecast error, the power imbalance between actual and expected profile can be written as:

$$\Delta P = P_{forecast} - P_{Geff} = (P_{G,forecast} - P_{G,eff}) - (P_{L,forecast} - P_{L,eff}) \cong -\Delta P_G$$

$$\Delta P_G = P_{Gprev} - P_{Geff} \cong -\Delta P$$

Where subscript $_L$ refers to load and subscript $_G$ refers to generation. For simplicity we will consider only PV generation.

Consequently, there will be two cases:

- $\Delta P_G > 0$, i.e., overestimation in the forecast of photovoltaic production.
- $\Delta P_G < 0$, i.e., underestimation in the forecast of photovoltaic production.

The two separate cases have different consequences in network operation and pricing in energy markets.

In the case of forecasts overestimations:

- The TSO, the next day, will dispatch a smaller amount of power from PV generation in the network than expected. Therefore, in order to maintain the equilibrium of the net network load, production from programmable sources will be increased.
- The GSE organized the commercial trading the day before, based on a supply expected from photovoltaic higher than that actually available the next day. This leads to a lower price than the one with the correct forecast.

In the case of underestimation of production:

- The TSO, the next day, will dispatch more power generated from PV plants than expected. Thus, in order to maintain the equilibrium of the net load in the network, production from programmable sources must be decreased.
- GSE organized the trade trading the day before, based on an expected supply of photovoltaic lower than that actually available the next day. This leads to a higher price than the one with the correct forecast.

2 Hourly power calculation model

As discussed in the previous chapter, models can be categorized based on the type of incoming data. The model adopted in this thesis is deterministic, that is, it uses as input data all the meteorological data of the analysed location and the physical parameters of the photovoltaic system. The output is an hourly power profile that will then be compared with the measured electrical power data. This chapter is dedicated to the description of this model; more precisely a total number of 2 non-optimized models and 5 optimized models are reported.

The first paragraph is dedicated to non-optimized models, while in the second paragraph it is described how some parameters are varied to better match forecast and actual PV production and reduce the final error.

2.1 Hourly power calculation model

Figure 2-1 shows in a very simple way how a PV plant is connected to the electric grid, and it will be used for the starting equation of produced power. Figure 2-2: PV performance model visual representation.

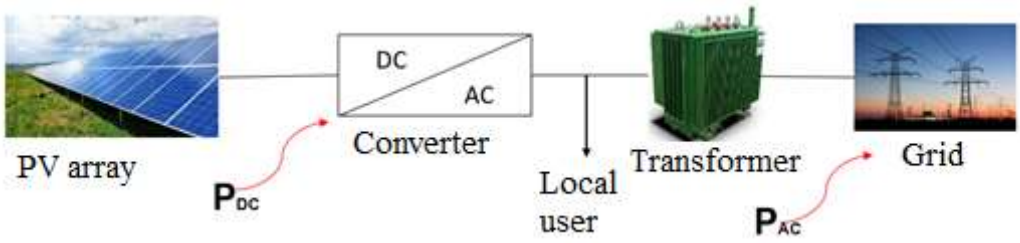

Figure 2-1: connection scheme of a PV plant connected to the grid.

More in detail, Figure 2-2 contains all the steps of the hourly power calculation model for photovoltaic systems. Each colour corresponds to a specific type of data or intermediate step. In particular:

- Blue boxes are the input data.
- Green boxes are the meteorological variables. They can be considered as an input, but unlike the other inputs contained in the blue boxes, they represent the most critical input or step of the model, because, as they refer to

forecasts, they incorporate the biggest source of error. Chapter 3 will be focused on the processing of these data.

- Yellow boxes are important quantities referring to PV power calculation steps obtained as intermediate steps within the model.
- Orange boxes represent the physical models used to move from one step to another.
- Finally, the purple box represents the power output from the system, as well as the final output on which the analysis of the results will be focused.

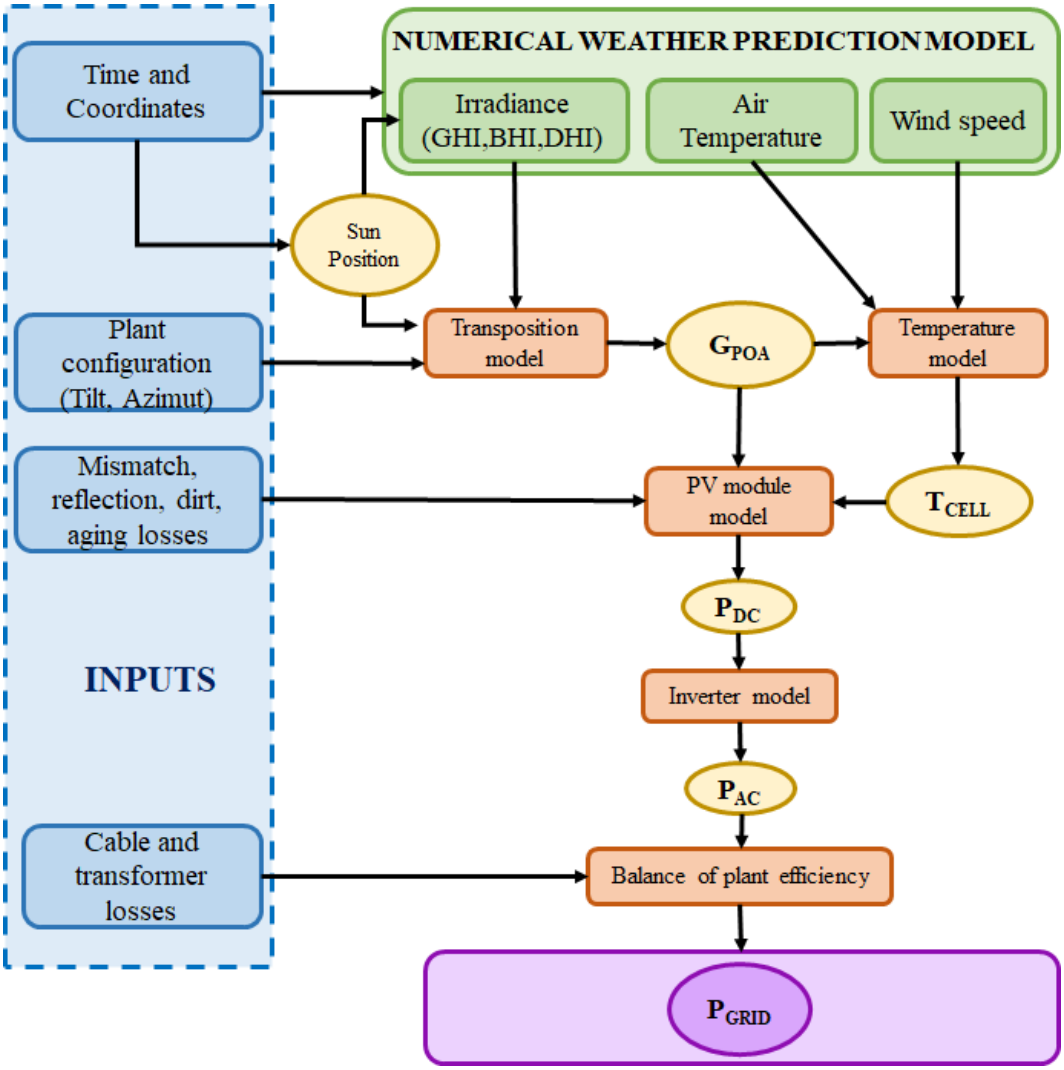

Figure 2-2: PV performance model visual representation.

With reference to Figure 2-1, the starting equation for calculating output power is the following:

$$P_{AC} = P_{DC} \cdot \eta_{CONV} \cdot \eta_{TRASF} =$$

$$= (P_{STC} \cdot \frac{G - \boldsymbol{G_0}}{G_{STC}} \cdot C_T \cdot \eta_G \cdot \boldsymbol{C_A}) \cdot (1 - f_g) \cdot \eta_{CONV} \cdot \eta_{TRASF} \quad (2.1)$$

The terms in red in the previous equation and in some of the following equations, refer to parameters that will undergo an optimization process in order to minimize the difference between the measured energy and that estimated by the model. Components of equation are then defined:

- P_{DC}: power generated by the PV plant in direct current [kW];
- η_{CONV}: inverter efficiency.
- η_{TRASF}: transformer efficiency.
- P_{STC}: plant nominal power measured Standard Test Condition [kW].
- G: solar irradiance on the plane of array (POA Irradiance) [kW/m^2].

Irradiance data is available on the horizontal plane; therefore they must be transposed on the plane of array by means of the following equation from ASHRAE model:

$$G = \frac{BHI}{\cos(\theta_z)} \cdot \cos(\theta) + DHI \cdot F_{CS} + \rho \cdot GHI \cdot (1 - F_{CS}) \quad (2.2)$$

The terms of equation 2.2 are:

- BHI: direct or beam horizontal irradiance. It is the irradiance component reaching the ground on a horizontal plane without being reflected nor absorbed by atmosphere. It can reach a maximum value of 1000 W/m^2.

- DHI: diffuse horizontal irradiance. It is the irradiance component reaching the ground on a horizontal plane after being reflected by atmosphere.

- GHI: global horizontal irradiance. It is the sum of BHI, DHI and the G_r, the latter defined as the reflected irradiance by the surrounding ground (usually it is a negligible contribute).

$$GHI = BHI + DHI + G_r$$

- θ_z: Solar zenith angle. It is the angle between the and the Zenith axis (perpendicular to the ground).

- θ: sun rays' angle of incidence. It is the angle between the perpendicular to the plane of array and the Earth-Sun line.

- F_{CS}: Earth-sky view factor.

- ρ: albedo coefficient. It takes into account the contribution of the reflected irradiance from surrounding environment.

In this thesis the ASHRAE model is not applied in its entirety. ASHRAE model indeed refers to clear sky conditions and calculates irradiance components as a function of the geographical coordinates and the time of the year. Forecasted data are instead fed into the model, while ASHRAE model is limited to the transposition of irradiance on the plane of array.

- G_0 solar irradiance threshold. It is the minimum irradiance required by the plant to be able to produce energy [kW/m^2].

- G_{STC}: irradiance in Standard Testing Conditions, equal to 1000 kW/m^2.

- C_T: Temperature coefficient [-]. It is calculated as:

$$C_T = 1 + \boldsymbol{\gamma_T} \cdot (T_{cella} - T_{STC}) \qquad (2\text{-}1)$$

γ_T% is the power thermal coefficient [%/°C], it depends on the technology of the PV modules, and it accounts for power deviations for temperature values different for standard testing conditions of 25 °C. As the case study has crystalline silicon modules, the chosen value for power thermal coefficient is:

$$\gamma_{T\%} = -0.50\% \quad \left[\frac{\%}{°C}\right];$$

- η_G: global efficiency [-]. It is the product of the following efficiencies.

$$\eta_g = \eta_{life} \cdot \eta_{dirt} \cdot \eta_{reflection} \cdot \eta_{mismatch} \cdot \eta_{cable} \qquad (2\text{-}2)$$

  - η_{life} accounts for aging of the PV plant. Its value depends on the number n of years passed since the beginning of operation of the plant, and it is calculated as:

$$\eta_{life} = 1 - \gamma_{life\%} \cdot n \qquad (2\text{-}3)$$

  - η_{dirt} consider the presence of dust particles that may deposit on the modules, it depends on environmental pollution and proximity to excessively dusty sites.

- $\eta_{reflection}$ takes into account the reflection phenomena of the front glazing of the modules, therefore a small fraction of the irradiance will not reach the PV cells.
- $\eta_{mismatch}$ accounts for mismatch losses due to the interconnection of modules which have not identical electrical performance.
- η_{cable} is for losses in the cables due to Joule effect.

- C_A: coefficient of adaptation [-]. During optimization it is used to compensate for inaccuracies of other terms, kept constant.

Some of these parameters are constant and their dependence on the plant can be neglected. For these parameters bibliographical values and typical construction values are used. Some of the values used in the model are reported below:

- $\gamma_{life\%} = 1.22\%$;
- $\eta_{dirt} = 0.976$;
- $\eta_{reflection} = 0.973$;
- $\eta_{mismatch} = 0.97$;
- $\eta_{cable} = 0.99$;
- $\eta_{TRASF} = 0.99$.

Regarding the efficiency of the converter, an equivalent converter is considered for the generic PV plants by means of a quadratic model of the inverter. The use of a quadratic model arises from the need to determine a single yield curve for $\eta_{CONV}$the different plants of the sample. Actually, each system is characterized by its nominal power and a specific type of converter. The types of inverters used for the conversion from direct to alternating in photovoltaics are:

- *Central inverter*, i.e. single inverter for all modules. It is used for medium-high powers 20-250 kW and has high efficiencies and low costs.
- *String Inverter*, i.e., a single Inverter for each string of modules in the system. This type is mainly used for residential applications from 1.5 to 5 kW of installed power and maximizes the optimization of the extracted power for each string;
- *Multi string Inverter*, multiple DC/DC converters are used with MPPT independent of each string and single inverter;
- *AC Module,* in this case the optimization via MPPT is maximum since each module has its own converter. However, monitoring is complex in this case.

The model used in thesis work is intended for the generic PV plant, therefore it is not possible to know precisely the type of converter used for each system. Especially for large plants, there may be multiple converters in string configuration or multi-string inverter to extrapolate the maximum power and therefore the productions on the DC sides can be different as well as the yields of the individual converters. To overcome this problem, with the quadratic model we make approximately a corresponding converter, where the hourly value of the yield will be obtained by interpolation between the yield curve of a quadratic model, the consequent DC power obtained by the model and the calculated DC power. The model is constructed by setting a maximum percentage value of 100 for AC power and using 100 points in steps of 1 to obtain a model yield curve using equation 2.6. Then the percentage power curve in DC of the model is determined through the efficiency curve obtained and finally the latter is multiplied by the nominal power of each system to obtain the absolute value of the DC power that will be interpolated with the calculated DC power. [19]

$$\eta_{CONV} = \frac{P_{AC}}{P_{DC}} = \frac{P_{AC}}{P_{AC} + P_{AC,losses}} \qquad (2.5)$$

Further explaining the term relating to power losses, we therefore obtain:

$$\eta_{CONV} = \frac{P_{AC}}{P_{AC} + \boldsymbol{a_p} + \boldsymbol{b_p} \cdot P_{AC} + \boldsymbol{c_p} \cdot {P_{AC}}^2} \qquad (2.6)$$

The denominator of equation 2.6 is three terms relative to losses, for no-load losses, linear losses, and quadratic losses. These terms will undergo an optimization process.

The quadratic model of the converter has a maximum value of about 0.979 so the power produced in DC can be at maximum $1.02 \cdot P_{STC}$.

Cell temperature, essential to determine the thermal coefficient, is calculated using two models: NOCT and WIND, which will be described in the next section.

NOCT model (*Normal Operating Cell Temperature):*

It is calculation based on the meteorological parameters of:

- T_a: air temperature at the installation site of the PV plant.
- G : incident irradiance on modules.

$$T_{cella} = T_a + \frac{\boldsymbol{NOCT} - T_{NOCT}}{G_{NOCT}} \cdot G \qquad (2.7)$$

Where:

- T_a is air temperature.
- G è is the POA irradiance.
- T_{NOCT} is the ambient temperature at NOCT conditions, equal to 20 °C.
- G_{NOCT} is the incident solar irradiance on the modules at NOCT conditions, equal to 800 W/m^2.
- NOCT is the cell temperature under NOCT conditions, that is under the conditions of T_{NOCT}, G_{NOCT}, wind speed of a 1 m/s, stable conditions of open circuit and 45° tilted module. The value of NOCT used for the purpose of this thesis, considering the multi-crystalline silica technology, is assumed equal to:

  NOCT = 47 °C

WIND velocity method:

In WIND velocity model the peculiar meteorological parameters are wind velocity *v*, compared to NOCT model which uses just are G and T_a. The temperature equation is the following:

$$T_{cella} = \boldsymbol{a_{Tc}} \cdot T_a + \boldsymbol{b_{Tc}} \cdot G - \boldsymbol{c_{Tc}} \cdot v + d \qquad (2.8)$$

with the following coefficients:

- a = 0.943
- b = 0.028 $\left[\frac{°C\ m^2}{W}\right]$
- c = 1.528 $\left[\frac{°C\ s}{m}\right]$
- d = 4.3 [°C]

The formula is based on the values of air temperature, irradiance and wind speed, each of which is multiplied by a respective coefficient. Usually, the values of the coefficients are fixed experimentally. However, in order to reduce the gap between actual measurement and the calculation of the model, they will be optimized.

Finally, it represents the corrective factor of the day and has a significant impact only for the calculation of energy for the winter semester, where there are more cloudy days. This coefficient was added to the model in order to reduce the overestimation, detected by the results of previous students, during the winter semester. The corrective factor of the day is calculated, for each implant and for each f_g i-th day, with the following quadratic relation:

$$f_{g,i} = \boldsymbol{a_f} \cdot CCI^2 + \boldsymbol{b_f} \cdot CCI + \boldsymbol{c_f} \qquad (2.9)$$

where *a, b* and *c* are coefficients of quadratic expression, while the CCI (cloud cover index) takes into account the cloudiness of the i-th day and derives from the comparison between the total daily irradiation calculated in clear sky conditions (CSI) using the Moon model and the total daily measured radiation. Here is his expression:

$$CCI = \frac{(ICS - IRR_{\text{mis}})}{ICS} \qquad (2.10)$$

In the thesis work *scarto giornaliero* is defined as:

$$scarto\ giornaliero_i = 1 - \frac{\left(ICS_{giorno,i} - IRR_{giorno,i}\right)}{ICS_{giorno,i}} \qquad (2.11)$$

Therefore:

$$(1 - scartogiornaliero_i) = CCI = \frac{(ICS - IRR_{\text{mis}})}{ICS} \qquad (2.12)$$

So that the equation (3. 9) can be rewritten as:

$$f_{g,i} = \boldsymbol{a_f} \cdot (1 - scartogiornaliero_i)^2 + \boldsymbol{b_f} \cdot (1 - scartogiornaliero_i) + \\ + \boldsymbol{c_f} \qquad (2.13)$$

For each plant in the sample, the hourly irradiation profile at clear skies for the entire 2018 is then calculated (Calculation of Moon Profiles) and then the daily deviation for each i-th day of the year is determined by comparing it with the irradiation profiles that are already available, i.e., those extrapolated from the *SoDa* database. This operation is carried out with the help of the Matlab *Estrapolazione_scarto_giornaliero.mlx* script that exploits the calculation of Moon Profiles.

2.2 Calculation models and parameters optimization

The objective of optimizing the parameters, for the various models used, is to reduce the gap between the actual production profiles and the profiles obtained from the calculation models. The goal is to compensate for the inaccuracies present in the model by estimating, indirectly, the most correct values for some of the parameters present in the calculation expression.

The optimization chosen for the model is nonlinear multi-variable. It is based on the minimization of the objective function considering some inequality constraints of the variables to be optimized:

$$min_x\, f(\boldsymbol{x})$$

$$s.t: \quad \boldsymbol{lb} \leq \boldsymbol{x} \leq \boldsymbol{ub}$$

Where:

- $\boldsymbol{x}$ is the vector of optimization variables that minimize the objective function, i.e., the parameters.
- $\boldsymbol{lb}$ is the vector of the lower limits of the optimization variables.
- $\boldsymbol{ub}$ is the vector of the upper limits of the optimization variables.

$f(\boldsymbol{x})$ is the objective function to be minimized. It is the average quadric error normalized for the nominal power of each plant.

$$f(\boldsymbol{x}) = \sum_{j=1}^{J} \frac{\sqrt{\frac{1}{T}\sum_{i=1}^{T}\left(P_{AC,prev,i} - P_{mis,i}\right)^2}}{P_{STC}}$$

A quadratic deviation is used, rather than in absolute value, to give more importance to hourly deviations that exceed unity. For each system, the standard deviation of the difference between the i-th power values provided by the model and those measured over a given period of time is calculated. Finally, the waste is normalized with respect to the size of the implant.

Ultimately, 7 combinations between calculation models and optimizations will be analyzed, which consider the calculation of the cell temperature according to the NOCT model and according to the wind speed model, with different parameter optimizations. Note that the number in the various model names refers to the

optimized parameters except for the inverter loss parameters (which will be optimized for each model). The models analyzed are:

1. NOCT model - not optimized with bibliographic parameters.
2. WIND model - not optimized with bibliographic parameters.
3. Model NOCT 3 parameters - single optimization (on the summer semester);G_0, γ_T, C_A *and* a_p, b_p, C_p
4. Model NOCT 4 parameters - double optimization (on the summer semester and coefficients for day factors on the winter semester);$G_0, \gamma_T, C_A, NOCT$ *and* $a_p, b_p, C_p\ a_f, b_f, C_f$
5. Model WIND 6 parameters - double optimization (on the summer semester and coefficients for day factors on the winter semester);$G_0, \gamma_T, C_A, a_{Tc}, b_{Tc}, C_{Tc}$ *and* $a_p, b_p, C_p\ \ a_f, b_f, C_f$
6. Model WIND 4 parameters coeff. A - Double optimization (on the summer semester and coefficients for day factors on the winter semester);$G_0, \gamma_T, C_A, a_{Tc}$ *and* $a_p, b_p, C_p\ \ a_f, b_f, C_f$
7. Model WIND 4 parameters coeff. B - Double optimization (on the summer semester and coefficients for day factors on the winter semester);$G_0, \gamma_T, C_A, b_{Tc}$ *and* $a_p, b_p, C_p\ \ a_f, b_f, C_f$

3 Data resource: acquisition and processing of forecast parameters, measured irradiance and electrical power

The following chapter shows how data is obtained and organized. In the introduction a brief interaction among data is explained. The first paragraph is dedicated to forecast data, with the main focus on the irradiance as it has the major influence on PV energy production. In particular:

- It starts with data source description.
- Then the procedure of data acquisition and MATLAB codes used to gather it are presented.
- After that, it is reported how these data are pre-processed and organized, in order to create different time leads forecasts profiles.
- Finally, some GHI plots are shown.

The second paragraph describe measurement methods used for irradiance and electric power.

To sum up, this study consists of the acquisition, processing and comparison of four types of data:

- Forecast weather data: the meteorological parameters of interest for photovoltaic applications are downloaded, and pre-processed for the hourly power calculation model.
- Forecast power data: obtained applying power calculation model to weather forecast data.
- Measured electrical power data: These are used to calculate the error of the photovoltaic power prediction calculation model.
- Measured irradiation data: which are used to compare determine whether the final deviation is due to model or forecast inaccuracies.

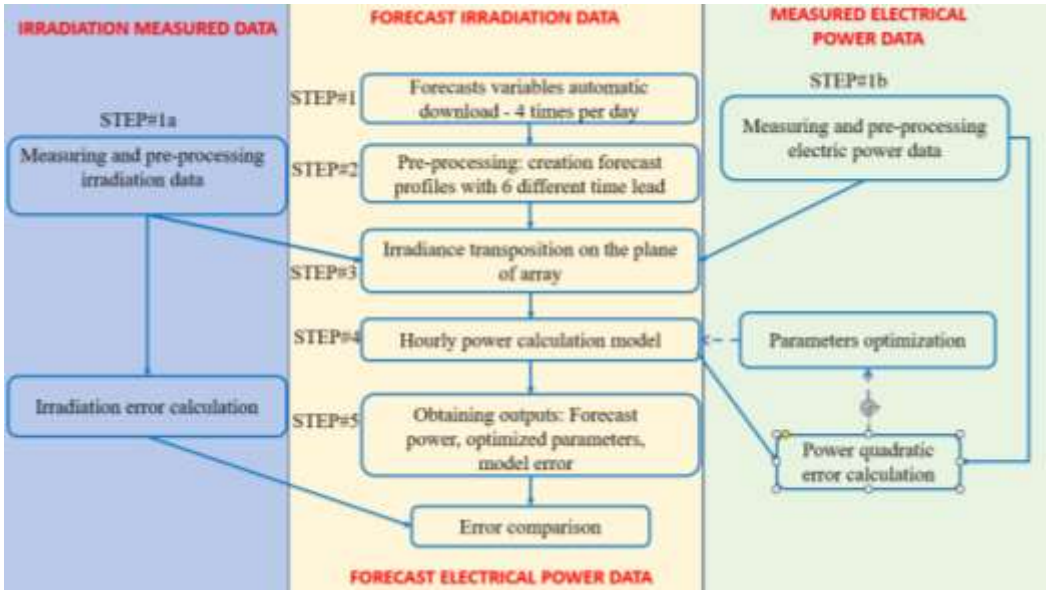

Figure 3-1: Main steps of the thesis, focusing on distinction among the three different data sources.

3.1 Forecast data

Meteorological variables that influence PV energy production (solar radiation, wind speed, temperature…) at the Earth's surface can be accessed in three forms: measurements from ground-based instruments (e.g., radiometers, anemometers, or thermometers), remote-sensing retrievals, and output of dynamical weather models. These three forms of information, though describing the same quantities, should be regarded as complementary, rather than substitutive [20]. Forecast used in the model of this thesis utilize a combination and elaboration of these techniques depending on the time horizon. Due to the sheer size of NWP data and the number of queries that are typically received, operational NWP forecasts only stay online for a few days, therefore forecast data are gathered every day with an Application Programming Interface (API).

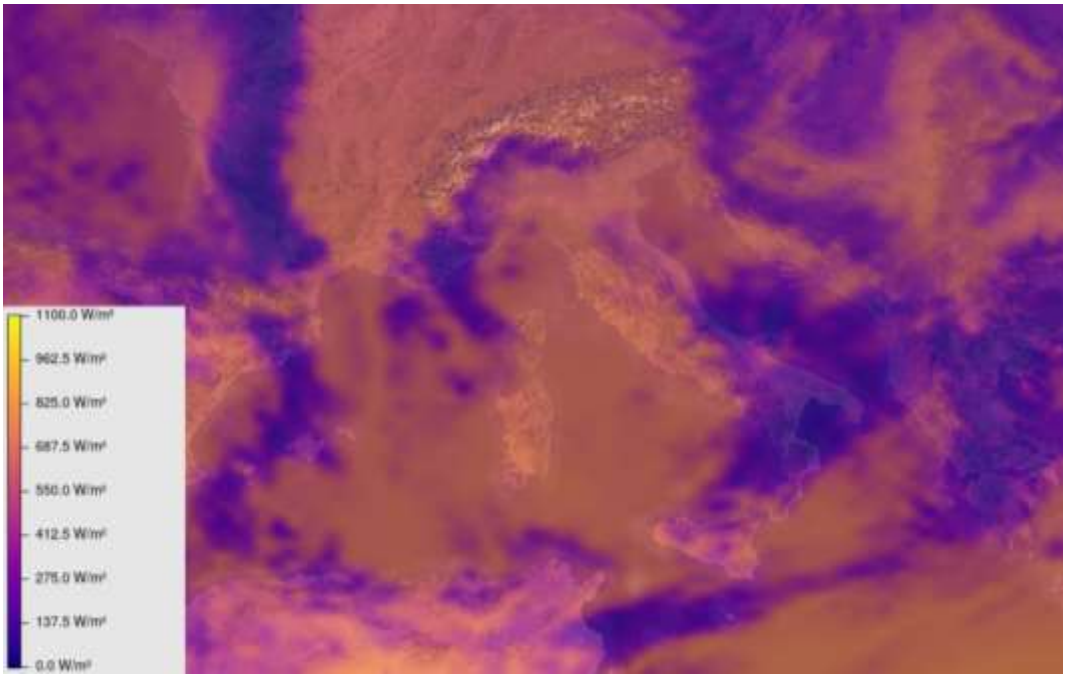

Figure 3-2: Heatmap of global horizontal irradiation over Italy surface [20].

Forecasts are requested to the weather service provider Meteomatics [20] [21], a Swiss company with other offices also in Germany and England, that provides an extensive number of meteorological parameters over multiple time horizons and spatial resolutions. In the case of Central Europe in particular, many usable models are available. Meteomatics processes the model data from the leading Numerical Weather Prediction (NWP) providers such as Met Office, ECMWF, NOAA (US), MeteoSwiss and satellite data providers [20]. The most important advantage of Meteomatics consist in the possibility to access meteorological parameters from different models using Meteomatics single API endpoint. Models and sources are combined into an intelligent blend, in a way that the best data source is chosen for each time and location.

Downloaded raw data cannot be processed directly, therefore a series of scripts is applied in this order:

1. `MAIN_V4.mlx`: it downloads meteorological parameters of a series of locations for a certain time period. It is built as a MATLAB live script, but then it is converted into a stand-alone application with application compiler function. It is described in the paragraph " 3.1.1- Weather API construction" and located in folder "programmi".

2. `unione_giorni_IRR_meteomatics_V2.mlx`: it generates for each location and parameter a matrix containing all the timesteps and all the different time-horizon forecast profile. It is described in paragraph “3.1.2 - Data organization” and located in folder “elaborazione_dati_in_locale”.

3. `PRE_PROCESSING_piano_inclinato.mlx`: it is used to produce a unique matrix containing all the plane of array irradiances of any timestep. It generates as many matrixes as the number of forecasts time horizon, but it stores all the locations together. It is described in paragraph “3.1.5 – irradiation profiles on a tilted plane” and it is located in folder “pre_processing”.

4. `Estrapolazione_scarto_giornaliero.mlx`: for any forecast time horizon, it generates the file `scarto_giornaliero.mat` which is needed during the optimization part. It described in the introduction of chapter 0 and itis located in folder “pre_processing”.

3.1.1 Weather API construction

In this paragraph the procedure used to obtain a forecast is explained. For each query it is mandatory to include:

1. User credentials.
2. `Start_date`: date of the first day of period of requested forecasts. It is set as the day previous the current day, in order to also obtain data of the forecasts elaborated with the most recent updates using sky images. The unit of measurement is the day:

```
start_date = floor(now)-1;
```

3. `end_date`: date of the midnight of the last day of period of requested forecasts. It is set in order to get forecasts up to the end of the third day after the current time:

```
end_date = start_date+5;
```

4. `resolution`: the timestep of the format of the forecasts. It defines the timestep of the forecasts:

```
resolution = 1/24;
```

A resolution of 1/24 on a daily time basis, corresponds to a timestep equal to 1 hour. It is possible to select timesteps even shorter than 1 hour, however these values would then be interpolated starting from the hourly values, therefore 1/24 is the shortest significative timestep available.

5. `Forecast_parameters`: list of meteorological variables that are requested. They are described more in detail in the next page of this paragraph. If the number of parameters is larger than 10, they must be allocated into two different blocks:

```
parameters1 = [...
    'clear_sky_rad:W,' ...
    'direct_rad:W,' ...
    'diffuse_rad:W,' ...
    'global_rad:W,' ...
    't_2m:C,' ...
    'wind_speed_10m:ms,' ...
    'low_cloud_cover:p,' ...
    'medium_cloud_cover:p,' ...
    'high_cloud_cover:p,' ...
    'total_cloud_cover:p'];

parameters2 = [...
    'pressure_000m:Pa,' ...
    'relative_humidity_2m:p,' ...
    'sun_azimuth:d,' ...
    'sun_elevation:d'];
```

6. Coordinates: latitude and longitude. They are not a single value, but a vector containing a number of element equal to the maximum number of locations that can be requested each query, depending on the subscription plan as indicated below in this paragraph.
7. Model, they difference among the various option are related to:
    a. Spatial resolution: it depends on the model, it can be up to 0.0012° (~90 m at European latitudes),
    b. Temporal resolution: usually hourly resolution, or a multiple of 1 hour.
    c. Updates per day: usually 2 or 4 times per day, reason why in this study forecasts are downloaded 4 times per day.
    d. Maximum lead time: usually in the range of 2-5 days.
8. Output format.

On Meteomatics different subscription plans are available, therefore the following specifications will refer to the plan that was used for this application. The subscriptions used for this thesis is characterized by the following parameters:

- number of queries per day = #50.
- number of queries per minute = #10.
- number of locations per query = #10.
- number of meteorological parameters per query = #10.
- available models.
- option to access to historical data reanalysis, which was not activated.

The forecast model needs as input the following meteorological parameters:

1. Plane of Array irradiance.
2. Air temperature.
3. Wind speed.

Air temperature and wind speed are directly available, while irradiance need more parameters to be processed in order to be obtained. Therefore, downloaded parameters are the following:

- Beam Horizontal Irradiance (BHI) [W/m^2].
- Direct Horizontal Irradiance (DHI) [W/m^2].
- Global Horizontal Irradiance (GHI) [W/m^2].
- Clear sky irradiance [W/m^2].
- Air Temperature [°C].
- Wind speed [m/s].
- Solar Zenith angle [°].
- Solar azimuth angle [°].
- Low cloud cover [%]: they are formed below 2000 m in height.
- Medium cloud cover [%]: formed between 2000 and 8000 m altitude.
- High cloud cover [%]: formed between 8000 and 14000 m altitude.

- Total cloud cover [%]: determined by a combination of the previous three layers of clouds. In particular low clouds are responsible of the highest contribute to total cloud cover, while high clouds are the least relevant.

The description of some of these parameters in the list, which are averaged on an hourly basis, is absent because these variables are explained more in detail in the chapter of description of the model. Variables like irradiation components, temperature, wind speed are needed to estimate photovoltaic power output and are forecasted data, while solar angles are determined only by the time of the year and plant location. Although solar angles could be calculated without using Meteomatics, these parameters are still maintained in the API because provide useful details to understand the trend of the daily profile and possibly apply systematic corrections. Additional variables like pressure and humidity are available, however they are excluded from this model because their contribution is marginal compared to the final error, since predicted data are used instead of effective reanalysis or measurements.

Flowchart in Figure 3-3 shows the structure of the script `MAIN_V4.mlx`. It generates as many files as the number of the locations, each time a new query is made.

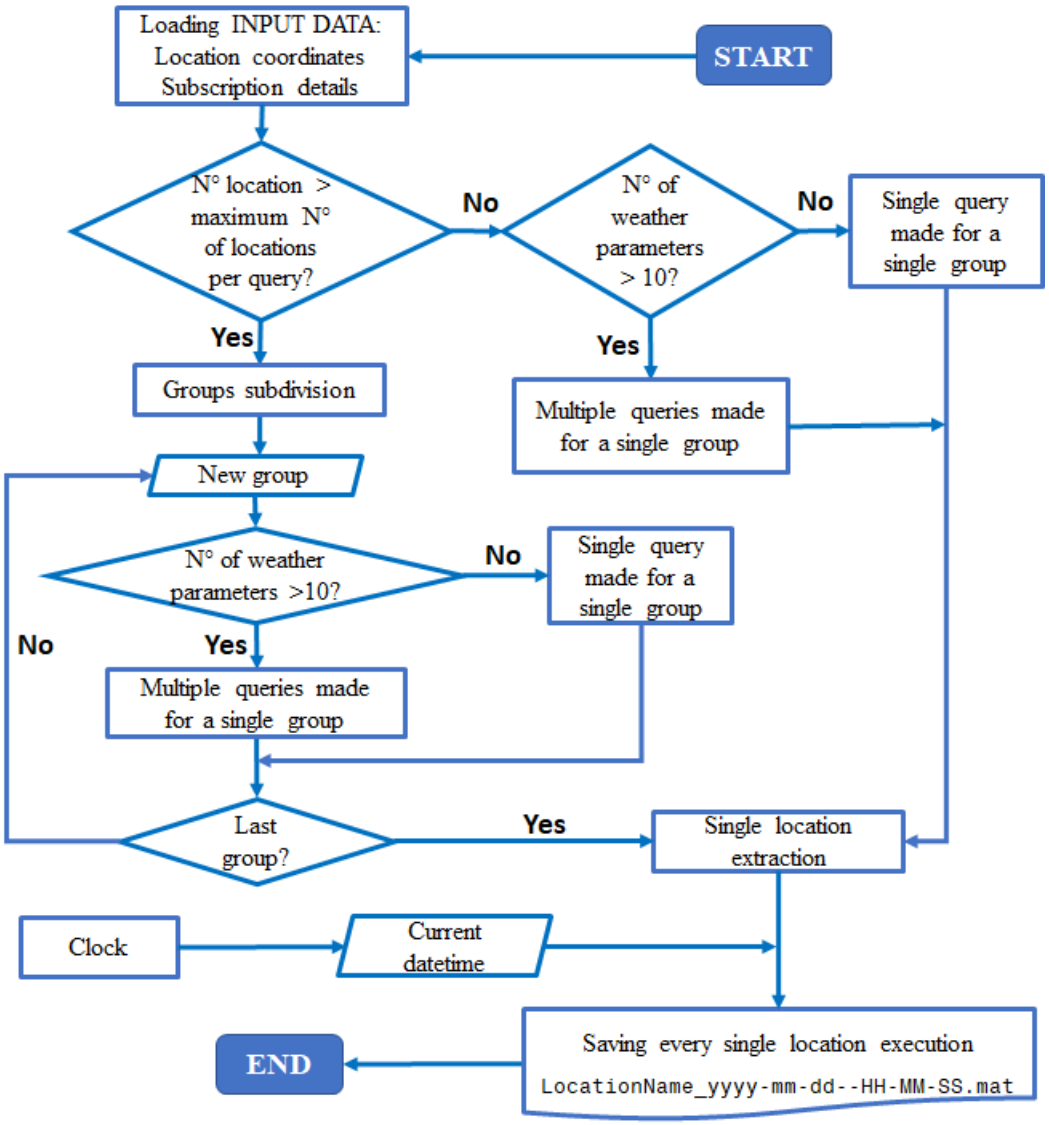

Figure 3-3: Flowchart of the code acquiring meteorological variables.

Finally, considering that many of NWP models are updated four times per day (at 00:00 UTC, 06:00 UTC, 12:00 UTC and 18:00 UTC), forecasts are therefore queried four times per day to have always the most update data. In order to make this process automatic, the MATLAB script which has the function of downloading weather forecast data has been converted to a stand-alone MATLAB application and configured with the Windows task scheduler.

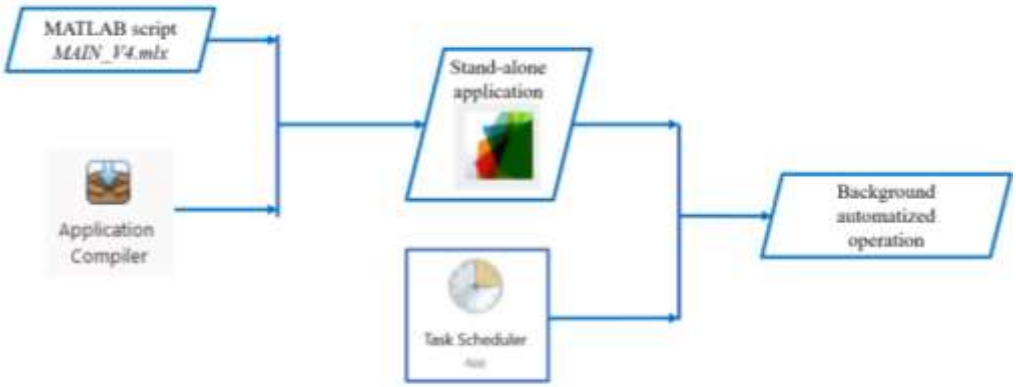

Figure 3-4: automatization of forecast download program.

3.1.2 Data organization

The number of generated raw files is equal to the product of:

- Number of locations, for which forecasts are requested.
- Number of weather parameters downloaded for each site.
- Number of queries per day
- Numbers of day.

Data downloaded from Meteomatics are not suitable for direct processing into the main script. `unione_giorni_IRR_meteomatics_V2.mlx`. script is then applied. It produces as many matrixes as the number of the locations; each matrix as a number of rows equal to the total number of timesteps (hours) and number of columns equal to the number of forecasts profile that are going to be built. Raw data pre-processing is summarized in the flowchart shown in Figure 3-5.

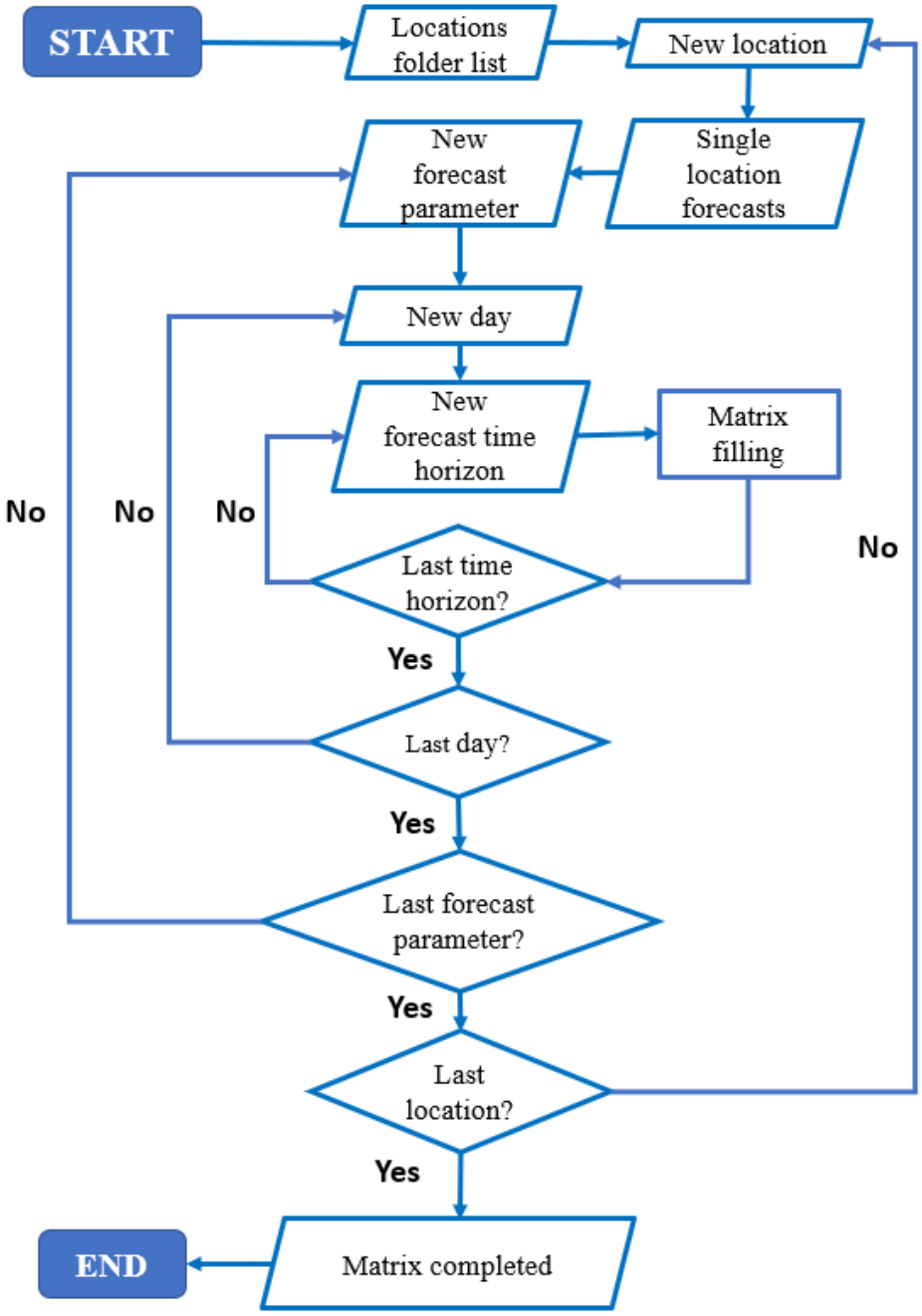

Figure 3-5: Flowchart of the code which creates a unique file for each location containing all days.

3.1.3 Theoretical and effective lead times

In paragraph 3.1.1 it is specified that forecasts must be downloaded within a time range. If we consider a forecast period of 2 days, the first timesteps will have a time lead of only a couple of hours, while the last timestep will have a 48 hours time lead. Therefore, it is important to create daily profiles with the same time lead. The time lead of a timestep is determined by how many hours there are between the time of forecast acquisition and the timestep itself, plus the hours needed by the NWP models to produce the result of the simulation. Figure 3-6.Figure 3-6 shows the difference between the effective forecast lead time of the simulation, and the perceived forecast lead time from the current time.

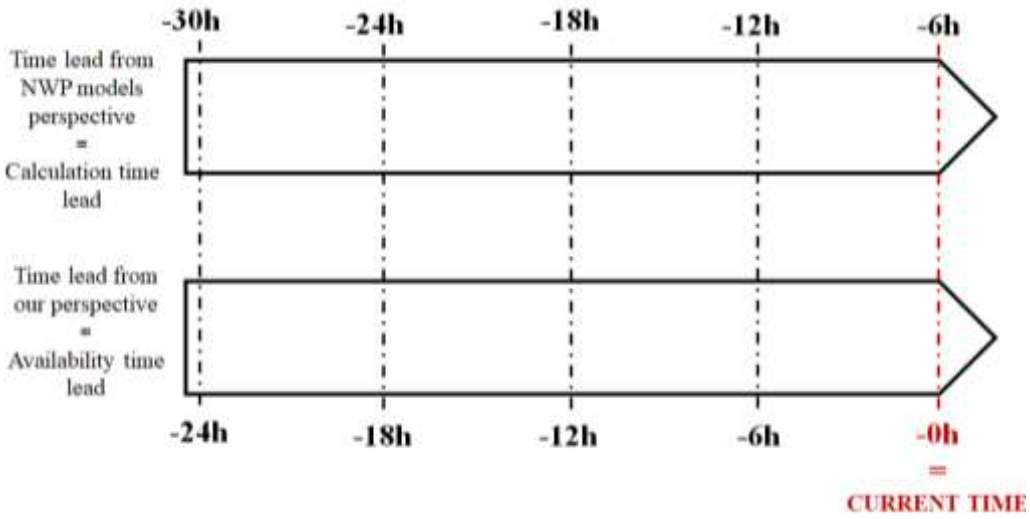

Figure 3-6.

With reference to the previous figure, in this thesis six different time leads profiles will be analyzed:

- Six hours time lead profile, consisting in forecasts with a lead time of 6 hours and available at the current time. This profile will be referred as “t-6h”.
- Twelve hours time lead profile, consisting in forecasts with a lead time of twelve hours and available six hours before the current time. This profile will be referred as “t-12h”.
- Eighteen hours time lead profile, consisting in forecasts with a lead time of eighteen hours and available twelve hours before the current time. This profile will be referred as “t-18h”.

- Twenty-four hours time lead profile, consisting in forecasts for the hour twenty-four hours after the current time and available eighteen hours before the current time. This profile will be referred as “t-24h”.
- Thirthy hours time lead profile, consisting in forecasts with a lead time of thirthy hours and available twenty-four hours before the current time. This profile will be referred as “t-30h”.
- Profile coming from the results of the NWP model using the final, most updated inputs, mainly coming from sky images. However, as the model takes 6 hours to produce outputs, these weather data refer to a time interval prior to the moment of weather forecasts. In conclusion these last data are are not of practical use as they do not refer to a future time. They are anyway utilized as a meter of how much forecasts can improve as they are more updated. This profile will be called “t-Post” profile.

However, although forecasts could be theoretically queried every hour, they got updated only four times per day. Therefore, taking as an example the -6h profile, forecasts are queried with six hours advance only for the time in which the query is executed, that is 00:00, 6:00, 12:00 and 18:00. As a result, what is called -6h profile contains forecasts with an advance between 6 and 11 hours. In general, the time lead of each profile is not constant, but it may have an increase of maximum amount of 5 hours. This amount only depends for each timestep on how many hours are passed between the execution time (00:00, 6:00,12:00 and 18:00) and the hour corresponding to that timestep a certain time.

Table 3-1 shows for each time of the day and for each profile what is the real advance the forecasts have been requested. This table is used at the beginning of the study to build daily irradiation profiles of the next paragraph, but not fed into further Matlab codes.

Table 3-1: Actual forecasts lead time for the five time-horizon profiles.

Time	-6h	12h	-18h	-24h	-30h
00:00	-6	-12	-18	-24	-30
01:00	-7	-13	-19	-25	-31
02:00	-8	-14	-20	-26	-32
03:00	-9	-15	-21	-27	-33
04:00	-10	-16	-22	-28	-34
05:00	-11	-17	-23	-29	-35
06:00	-6	-12	-18	-24	-30
07:00	-7	-13	-19	-25	-31
08:00	-8	-14	-20	-26	-32
09:00	-9	-15	-21	-27	-33
10:00	-10	-16	-22	-28	-34
11:00	-11	-17	-23	-29	-35
12:00	-6	-12	-18	-24	-30
13:00	-7	-13	-19	-25	-31
14:00	-8	-14	-20	-26	-32
15:00	-9	-15	-21	-27	-33
16:00	-10	-16	-22	-28	-34
17:00	-11	-17	-23	-29	-35
18:00	-6	-12	-18	-24	-30
19:00	-7	-13	-19	-25	-31
20:00	-8	-14	-20	-26	-32
21:00	-9	-15	-21	-27	-33
22:00	-10	-16	-22	-28	-34
23:00	-11	-17	-23	-29	-35

3.1.4 Irradiation profiles on horizontal plane

As mentioned before, decoupling forecast error and PV power model error is important to evaluate the performance of the PV model. For example, a final error of +30%, at first sight could seem a poor performance of the PV power calculation model, but if also in forecasts an overestimation in the range of 25-35% is detected, the responsibility of the bad performance must be attributed to the quality of

forecasts and not to flaws in the calculation model. Therefore, in the following figures, measured and forecast irradiances will be compared. For each plot, a synthetic table with the daily residual is attached. The following plots refer to the variable GHI, which has the main influence on the PV energy production. In this paragraph three plot examples are reported:

1. Very cloudy day
2. Clear-sky sunny day
3. Sunny day, but with 2 hours characterised by high variability in the GHI trend.

Global horizontal irradiance measures are explained in paragraph 3.2.1.

Figure 3-7: Progressively improvement in forecast for a cloudy day. shows the example of a day which was at first forecasted as a sunny day and it turned out to be very cloudy. More precisely it displays the six forecast profiles described in paragraph 3.1.3 and actual measurements of GHI. This is a day which perfectly represents the benefits of updating the forecast more times per day, because each updated forecast for that day is producing a more accurate output, decreasing from an initial error of almost 300% to a final error of 60%, as summarized in table 3.2. The error reported in tables 3.2-4 is calculated on a daily basis as:

$$Energy\ deviation = \frac{Forecasted\ data - Measured\ data}{Measured\ data} \cdot 100$$

Table 3-2: progressive residual improvement for a cloudy day.

	Post	t-6	t-12	t-18	t-24	t-30	Measured
Irradiation [kWh/m²]	0.64	0.88	0.93	1.11	1.44	1.59	0.40
Deviation [%]	59%	119%	132%	176%	257%	294%	-

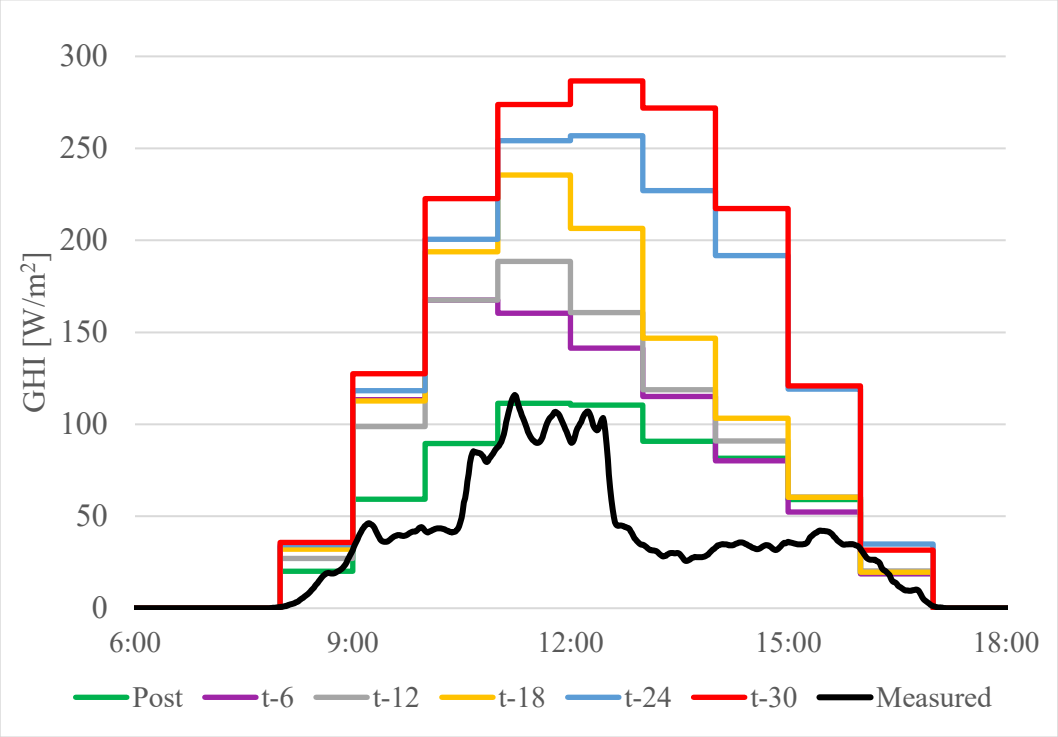

Figure 3-7: Progressively improvement in forecast for a cloudy day.

As the majority of sunny days do not present a large variation in updated forecasts, with the result that more profiles tend to collapse into a single one, in the following figures only the most recent and the less recent forecast will be reported for sake of having more readable plots. An example of a day characterised by a negligible variability in GHI update is reported in Figure 3-8Figure 3-7. Table 3.3 also shows the minimal variation of the residual calculated on a daily basis with the different forecast profiles.

Table 3-3: negligible daily residual improvement for a sunny day.

	Post	t-6	t-12	t-18	t-24	t-30
Irradiation [kWh/m²]	2.18	2.19	2.19	2.19	2.19	2.19
Deviation [%]	18.8%	19.4%	19.2%	19.1%	19.2%	19.3%

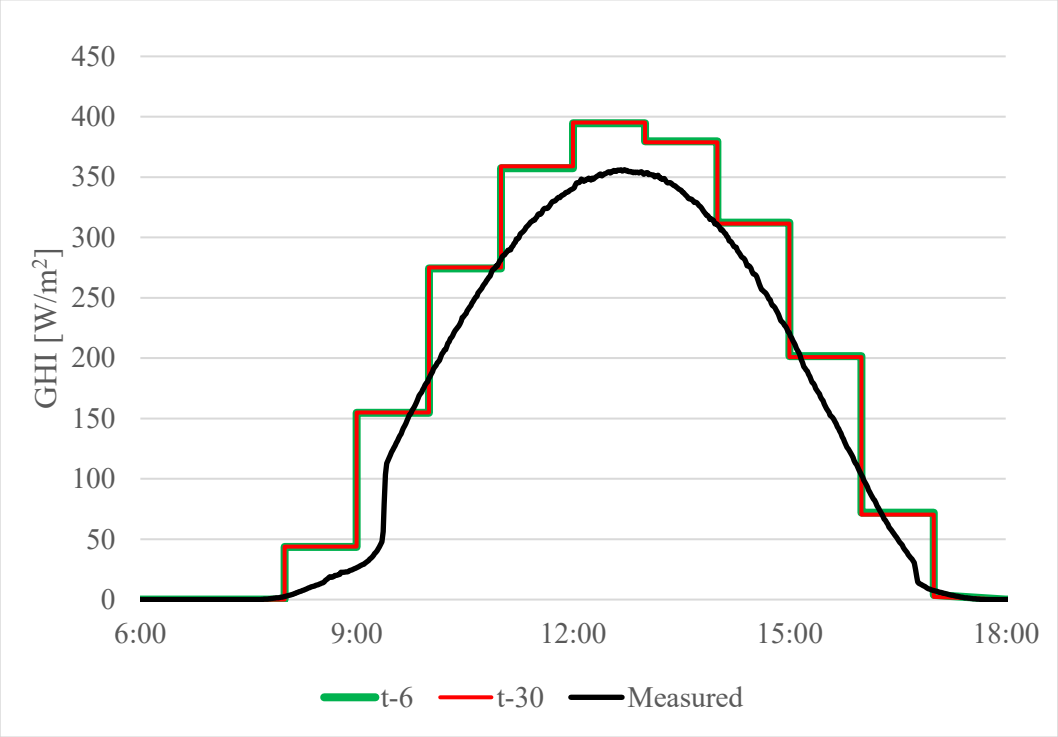

Figure 3-8: irradiation profiles for a clear sky day with extremely low variability.

Benefits from updated forecasts are not limited to very cloudy days, but in general when an element characterised by an elevated rate of variability and uncertainty appears. Figure 3-9 for example, represents what is mainly a clear sky day, with the exception of some clouds for a couple of hours in the morning. In this case the updated forecast profile cannot perfectly match the real profile due to the high rate of variability of cloud cover. Results consist in a generalized decrease of irradiation profile around the time interval in which clouds appear.

Table 3-4: residuals calculated with least and most updated forecasts, for a sunny day with a couple of cloudy hours.

	Post	t0	t-6	t-12	t-18	t-24	Measured
Irradiation [kWh/m²]	2.03	2.07	2.07	2.08	2.08	2.12	99
Deviation [%]	23%	25%	26%	26%	27%	28%	-

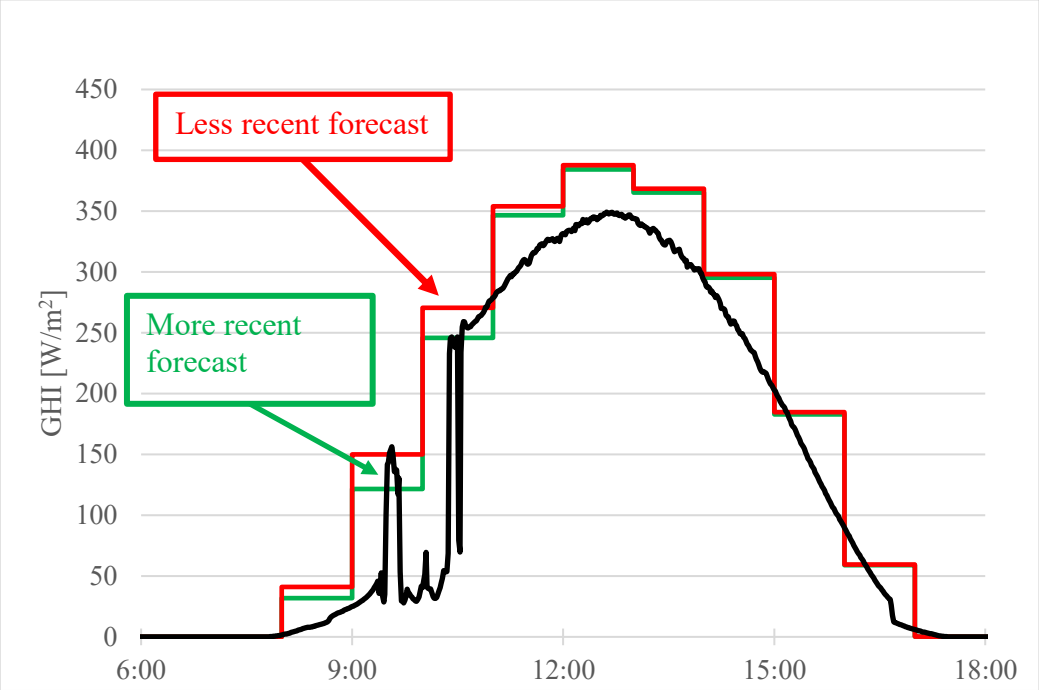

Figure 3-9: Most and least updated profiles for a sunny day with a couple of hours characterized by a high cloud variability.

3.1.5 Irradiation Profiles on a tilted plane

As previously introduced, irradiation components downloaded by Meteomatics are available only on the horizontal plane, therefore they must be translated on the plane of array as explained by the model in the paragraph 2.1. In particular script *PRE_PROCESSING_piano_inclinato.mlx* is applied, which uses the transposition model used described in equation 2.2. it requires the following inputs:

- Latitude and longitude of each plant.
- Solar irradiation components: they must be in an 8760xN°_of_pants format.
- Differently from previous thesis works, the number of time horizons forecasts is introduced too. In particular, the original core is inserted into an external cycle which for any forecast produces the final POA irradiation matrix. For simplicity, considering the similarity in the dimensions of data and of the files pathway of the other variables (temperature and wind speed), needed into the power calculation model, also temperature and wind speed matrixes are saved in the appropriate folder when running this script, but they do not need any further processing.

In order to assess whether the final error of the model is caused by the translation model or by the forecast error, measured irradiation da are collected both on the horizontal plane and on a tilted (37°) plane facing towards east. Results, here reported in graphical and tabular form, show the trend of the relative error of translated POA irradiation. This error is non negligible (for the day considered 17 %), but the error caused in forecasts is almost preserved. More in detail, forecasted irradiance on the horizontal plane GHI appears in general to be over-estimated, and as a consequence the POA irradiance tends to be higher than the real value as well, but with the same proportion. As this trend is verified also for the vast majority of days, it is reasonable to assume that the error introduced by the translation model is smaller than the uncertainties determined by the use of forecast data. Table 3-5 and Figure 3-10 shows these results.

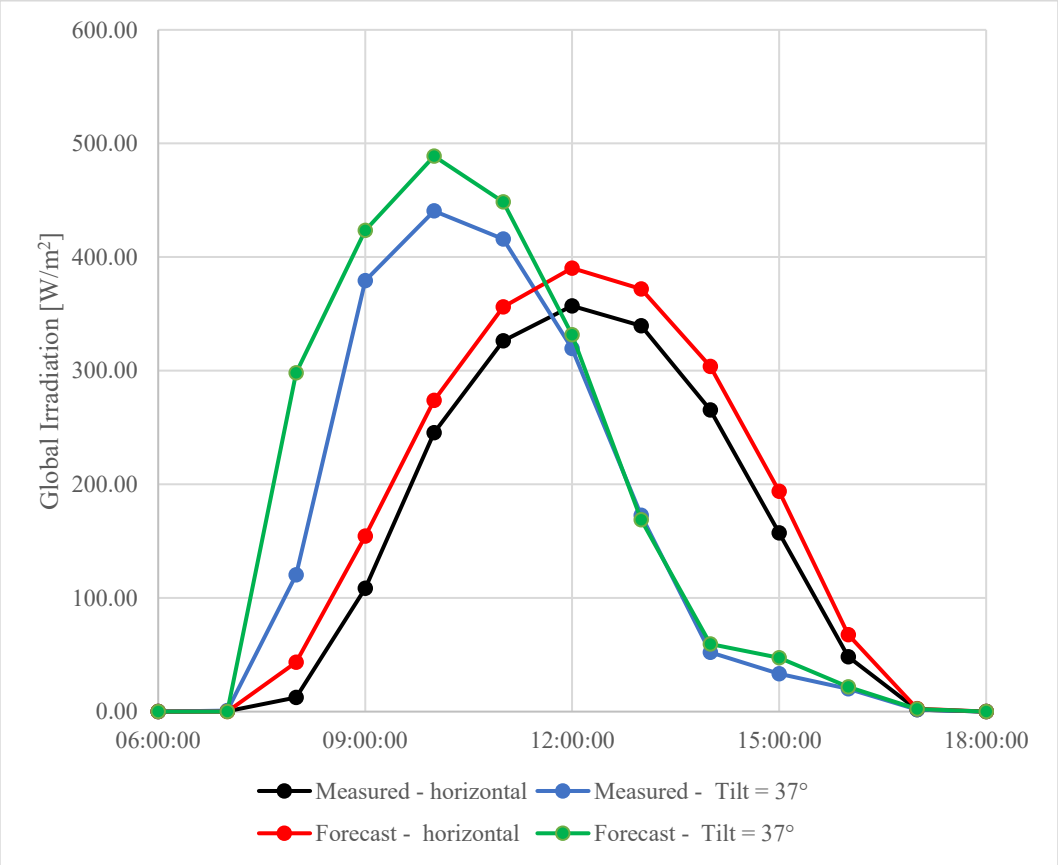

Figure 3-10: Comparison between forecasted and measured irradiation, for horizontal and tilted plane.

Table 3-5: comparison of the daily irradiation residual between GHI and POA irradiance.

	Measured horizontal	Measured Tilt = 37°	Forecast horizontal	Forecast Tilt = 37°
Irradiation [kWh/m²]	1.87	1.95	2.17	2.28
Deviation [%]	-	-	15.9%	17.0%

3.2 Measured data

The case study is related to a photovoltaic system installed in 2015 with a nominal power of 4,250 kW and polycrystalline silicon modules, equipped with calibrated irradiation sensors and monitoring systems. The plant is installed in the town of Rivalta di Torino. Figure 3-11: Case study plant location. shows PV plant case study location.

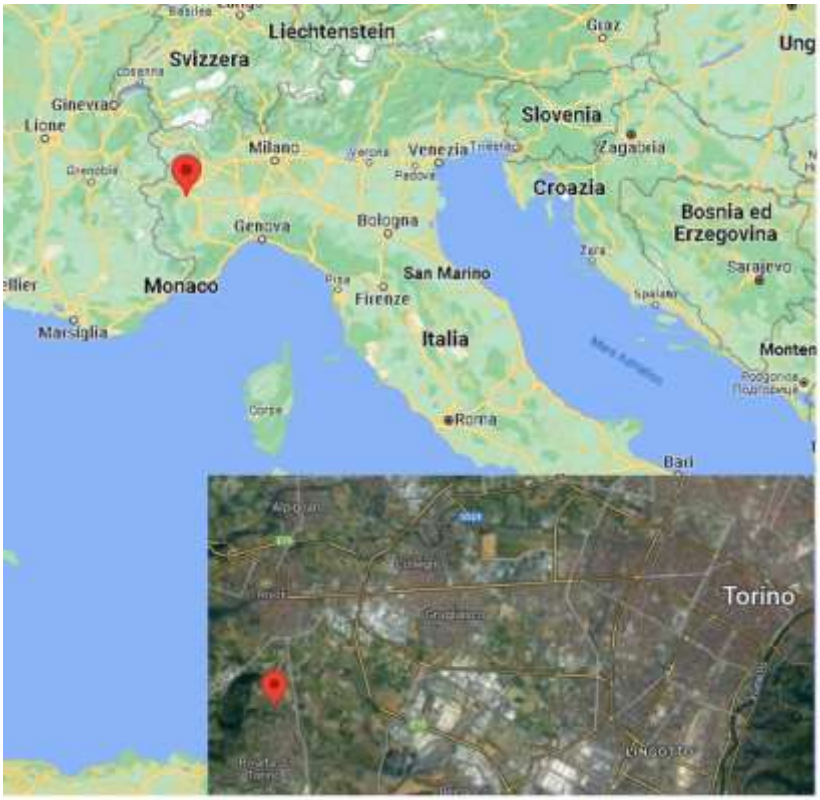

Figure 3-11: Case study plant location.

Measured data consist in:

- global irradiance on horizontal plane (GHI): it is needed to decouple the forecast error from the output electric power error;
- global irradiance on a tilted plane: it is used to evaluate possible errors when converting horizontal irradiance into incident irradiance on the plane of array;
- produced power by the PV plant to evaluate the final error and results of the model. It is also used inside the hourly calculation power model in the parameters optimization part.

3.2.1 Measured irradiance

Instruments sensor Spektron 210 and ICP DAS I-7017 data acquisition and control module. The Spektron 210 provides a voltage proportionally to the intensity of the solar irradiation.

Table 3-6: Spektron 210 specifications

Model	**Spektron 210**
Sensor type	Monocrystalline cell (13 mm / 33 mm)
Measuring range	0 – 1500 W/m²
Sensor accuracy	± 5 % (annual mean)
Outlet	approx. 75 mV at 1000 W/m²
Calibration	Sun Simulator Solar Constant 1200 with Reference sensor calibrated by the ISE
Design of the sensor	Measuring cell laminated in novaflon and EH foil
Casing	Z-profiled aluminium plate, connection encapsulated
Dimensions	118 mm x 50 mm x 44 mm
Protection mode	IP65
Weight	250 g (incl. cable)

The I-7017 is an analog input module that provides 8 differential input channels, whose datasheet is reported in Table 3-7. Considering that the output voltage from the irradiation sensor is around 75 mV at 1000 W/m², the voltage input of the acquisition model is selected to ±150 mV range.

Table 3-7: ICP DAS – I7017 technical specifications.

MODEL		I-7017
Channels		8
Wiring		6-channel Differential and 2-channel Single-ended, or 8-channel Differential.
Input range	**Voltage**	±150 mV, ±500 mV, ±1 V, ±5 V, ±10 V
	Current	±20 mA, 0 ~ 20 mA, 4 ~ 20 mA (Require an optional external 125 Ω resistor)
Resolution		16-bit
accuracy		0.10%
Sampling rate		10 Hz
Input Impedance		20 MΩ
Common Voltage protection		±15 V_{DC}
Overvoltage Protection		±35 V_{DC}

Data are then accessed with EZ Data Logger, a PC based data logging, data acquisition, control and monitoring software. It allows to remotely control and configure the ICPDAS data acquisition hardware. Moreover, it sores data in a database and can be exported to spreadsheets. The availability

3.2.2 Measured Power

Electrical power is measured by means of HT Solar300, located downstream the inverter. HT Solar300N is a multifunction device for verification of single-phase and three-phase PV system efficiency and Power Quality analysis. For the sake of this thesis, only the measured output power is used, with a timestep of 1 minute.

Figure 3-12: solar 300N configuration

Table 3.9 contains a brief datasheet of HT SOLAR 300N multimeter. For all the specifications, it is available at https://emin.com.mm/ht-instrumentssolar300n-hv00300n-0201-ht-instrument-solar300n-hv00300n-0201-power-quality-analyzer-for-checking-the-efficiency-of-photovoltaic-systems-without-cts-myanmar-33634/pr.html.

Table 3-8: HT SOLAR 300N datasheet.
FS = Full Scale

MODEL		**HT SOLAR 300N**
DC Voltage – PV systems	Range (V)	0.0 ÷ 1000.0
	Resolution (V)	0.1
	Accuracy	± (0.5%rdg + 2dgt)
	Input impedance	10MW
AC TRMS Voltage – PV systems Single/Three phase plants	Range (V)	0.0 ÷ 600.0
		0.0 ÷ 1000.0 (P-P)
	Resolution (V)	0.1
	Accuracy	± (0.5%rdg + 2dgt)
	Input impedance	10MW
DC and AC TRMS Current with external transducers	Range (mV)	0.0 ÷ 1000.0
	Resolution (mV)	0.1
	Accuracy	± (0.5%rdg + 0.06%FS)
	Input impedance	510 k
Environmental conditions	Reference temperature:	23°C ± 5°C
	Working temperature (°C)	0 ÷ 40
	Working humidity	<80%RH

HT SOLAR 300N has a very high acquisition frequency, for simplicity its outputs are exported on a minute basis. Forecasts are instead available on an hourly basis; therefore, they need to be averaged on hourly intervals in order to match the array dimensions of forecast data.

Irradiances outputs from ICP DAS I-7017 and power outputs from HT SOLAR 300N, are available on the identical time basis of one minute. As they undergo the same averaging process, for convenience the pre-processing of these data is made in the same script, called

PRE_PROCESSING_Potenza_misurata_media_oraria.mlx, located in folder "pre_processing". The outputs are the following files:

1. *potenza_media_oraria_misurata.mat*: which is used in the code to calculate the energy deviation in the optimization part. In addition, it is used in the post-processing data analysis to evaluate the final results.
2. *IRR_medio_orario_misurato.mat*: which is used to quantify forecast error, in order to account the weight of irradiance forecast error in the final power calculation model error. It is not used directly in the code, but it is however needed as an input because in the last lines it is anyway overwritten in an export excel file.

4 Hourly power calculation model performance: test on year 2022, results and daily power profiles

The following chapter covers the results of the model. In the first paragraph error and performance metrics are described. In the second paragraph some daily profiles are reported in order to analyse the results in a punctual way under a series of different circumstances. The third paragraph shows in a synthetic way the global results considering the error over the whole period. Finally, in the last paragraph the behaviour of the weather variable GHI is analysed over the whole year for four Italian regions, in order to evaluate the forecasts quality.

The following results were obtained with the *MAIN_GASPERONI.mlx* and *MAIN_GASPERONI_Cloud_Model_V2.mlx* scripts in folder "". For each model, the input data of the Matlab program are:

- Annual hourly irradiation matrix (IRR) (kWh/h/m2), 8760 x M (systems obtained from filtering for each region), data already reported on the inclined plane (downstream of the development of the ASHRAE model);
- Annual ambient temperature matrix 8760xM (°C).
- Wind speed matrix 8760xM (m/s).
- Matrix of sorted filtered data, i.e. table containing master data information (name, address, operating data, etc.) for sample plants.
- Filtered annual production matrix 8760xM (kWh/h), note that each row contains the annual production profiles of the plant of the same row as the matrix of filtered data ordered;
- Column vector Number of plants available for each class, #Classi x 1 (11x1);
- Column vector G0_iniziale (G00), #Classi x 1 (11x1), derives from the clustering procedure in order to avoid that the optimization parameters obtained depend on the initial values.

- Matrix annual daily deviation between the daily profiles of irradiance in clear sky and the real irradiance profiles measured, 365xM, in order to determine the day factor (to be considered non-zero only for optimization in the winter semester).

The last matrix refers to the MATLAB *script Estrapolazione_Scarto_giornaliero*. In it, the radiance profiles at the clear sky (Moon Profiles) are first obtained and then compared with the actual irradiance measured on the inclined plane in order to obtain the daily deviation. For Estrapolazione*_Scarto_giornaliero script,* the required input data are:

- provinces -> String with name of provinces to each region #province x 1.
- Hb -> Matrix with direct solar irradiation values, daily average monthly in a year, #province x 12. To be taken from pdf file "Data - dati_climatici_uni103491";
- Hd -> Matrix with diffuse solar irradiation values, daily average monthly in a year, #province x 12; To be taken from pdf file "Data - dati_climatici_uni103491";
- Sample plant data table containing main information (province, azimuth, tilt, latitude, longitude);
- Annual hourly irradiation matrix (IRR) (Wh/m2), 8760 x M (systems obtained from filtering for each region), data already reported on the inclined plane (downstream of the development of the ASHRAE model). It is compared with Moon profiles.

4.1 Error and performance metrics

Error is calculated on hourly, daily and global basis. The energy deviation is evaluated in percentage terms, the definition is highlighted in the following equation:

$$D_{\%} = \frac{\sum_i E_{Model} - \sum_i E_{Measured}}{\sum_i E_{Measured}} \qquad (4.1)$$

As it is defined, a deviation greater than zero implies an overestimation of the model compared to measurements, while a deviation less than zero involves an underestimation. It gives a rapid and general overview of the model. As positive and negative errors tend to compensate each other, the deviation produces global results with a very low error. In order to evaluate the performance of the model and of the forecasts, also Mean Absolute Percentage Error (MAPE) is introduced. It is calculated as described in chapter 8 of [12]:

$$MAPE = \frac{100}{N} \frac{\sum_{i=1}^{N} |P_{forecast,i} - P_{measured,i}|}{P_{measured,avg}} \quad (4.2)$$

Where N is the number of hours during the plant is working and $P_{measured,avg}$ the average measured power.

4.2 Production profiles comparison

Since the global horizontal irradiance (GHI) is the most important meteorological parameter to calculate the yield of a photovoltaic system, the calculation model was performed twice, using as a value of GHI:

1. Simply the value provided by the Meteomatics query (see paragraph 3.1.1.). it will be referred as "base model".
2. Starting from the power that would be produced in conditions of perfectly clear sky, corrected with the coefficient of cloudiness:

$$P_{\text{forecast}} = P_{\text{clear,sky}} \cdot F(t)$$

With $F(t)$ expressed as:

$$F(t) = c \cdot \frac{100 - low(t)}{100} + d \cdot \frac{100 - mid(t)}{100} + e \cdot \frac{100 - high\,(t)}{100} + f$$

Where the coefficients *c, d, e, f* are semi-empirical coefficients and low(t), mid(t) and high(t) represent the percentage of cloudiness at every hour, for the three different cloud levels The results of this second approach will be labeled as reactive to the "Cloud Model".

In this chapter a more detailed insight into some specific days is presented. In particular it is highlighted the effect of:

1. Effect of the general category of the model (that is the differences between NOCT and wind velocity-based models), both in case of:
   a. Base, non-optimized model
   b. The optimized version of the model which has produced the best results.
2. Effect of the optimization: the focus is on the comparison between the base and optimized versions, both for NOCT and wind velocity-based models.

3. Effect of the weather: in which daily profiles are displayed under different meteorological conditions of:
   a. Regular sunny day.
   b. Intermediately cloudy day
   c. Highly cloudy day.

In the case of days presenting high variability in irradiance and therefore in power output (in general cloudy days) the effect of the model is much less important than the forecast time lead. Therefore, for this case the comparison between different time horizons is presented.

4.2.1 Comparison between NOCT and Wind velocity models

In this paragraph the two main categories of models are compared, which are: NOCT models and Wind velocity based models.

NON-OPTIMIZED MODELS: NOCT AND WIND

Wind and NOCT models are different just by the way cell temperature is calculated. Cell temperature affects the PV production, but with a minor impact compared to the irradiance; therefore NOCT and Wind non-optimized models produces very similar results. Figure 4-1 shows the daily profile of the electric power calculated with NOCT and WIND non optimized models, compared with actual measured data. As explained at the beginning of this paragraph, the profiles of these two models tend to collapse into a single curve. On the same plot also forecast and measured irradiances are reported, to keep into account the input error introduced by forecasts. The point of maximum power production is shifted to the right with respect to the maximum irradiance point because the monitored plant faces towards west. It is possible to notice that the overestimation in irradiance is smoothed by the models, which still produce a positive energy deviation. Table 4-1 contains on a daily basis the energy deviation of profiles shown in fFigure 4-1. These are the oldest forecast profiles (t-30), and they refer to 20/01/2022.

Table 4-1

	Irradiation [kWh/m^2]		Electric Energy [kWh]		
	Measured GHI	Forecast GHI	Measured Energy	NOCT	Wind
Daily Total	1.91	2.31	7.97	8.2	8.2
Deviation [%]	-	21.06	-	3.3	3.1

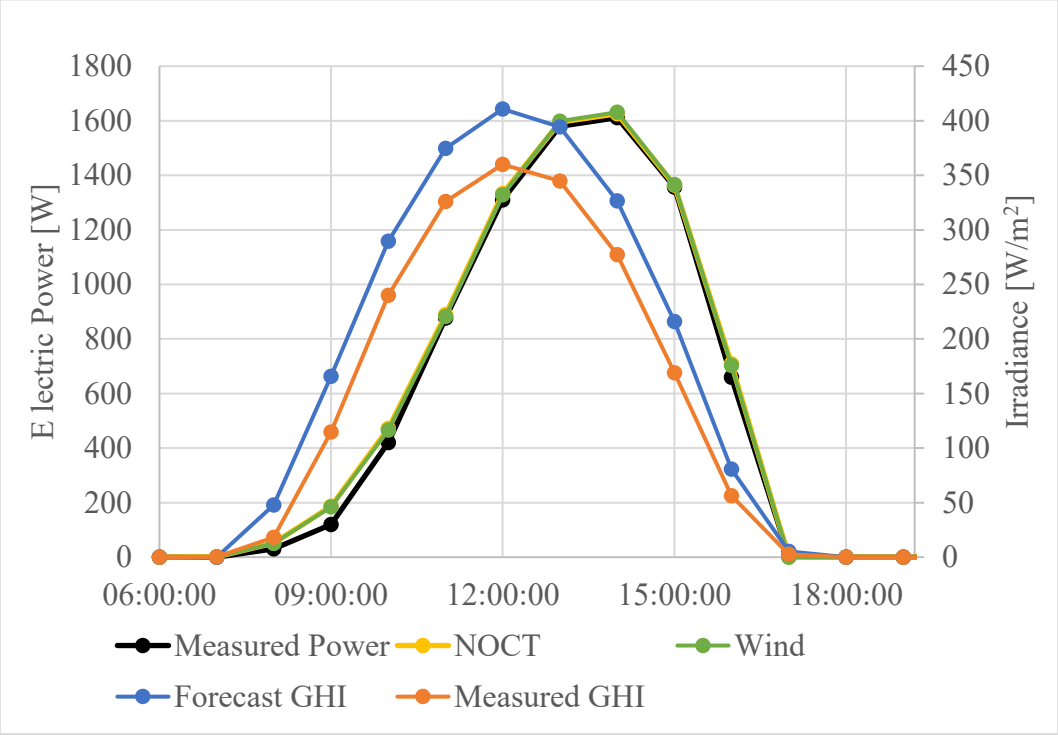

Figure 4-1: comparison between forecasted and measured irradiances, and measured and forecasted (for base NOCT and WIND model) electric power.

OPTIMIZED MODELS: NOCT VS WIND

In Figure 4-2 the optimized versions of NOCT and WIND models are reported, in particular they are the NOCT 4 PAR and WIND 3 PAR models. Also in this figure measured power, measured irradiance and forecast irradiance are plotted, for the same day of 20/01/2022. In this case the two models still produce very similar results, but differently from fFigure 4-1, they do not collapse into a single curve. As it can be seen in this figure and it will be discussed in paragraph 4.3, WIND 3 PAR tends to have a negative energy deviation (the prediction underestimates the actual production), while NOCT 4 PAR 2 STEP tends to have a positive energy deviation and it globally has the best performance. Table 4-2 contains on a daily basis the energy deviation of profiles shown in figure 4-2. The following data refer to the oldest forecast profiles (t-30).

Table 4-2

	Irradiation [kWh/m²]		Electric Energy [kWh]		
	Measured GHI	Forecast GHI	Measured Energy	NOCT 4 PAR 2 STEP	WIND 3 PAR
Daily Total	1.91	2.31	7.97	8.15	7.77
Deviation [%]	-	21.1	-	2.32	-2.42

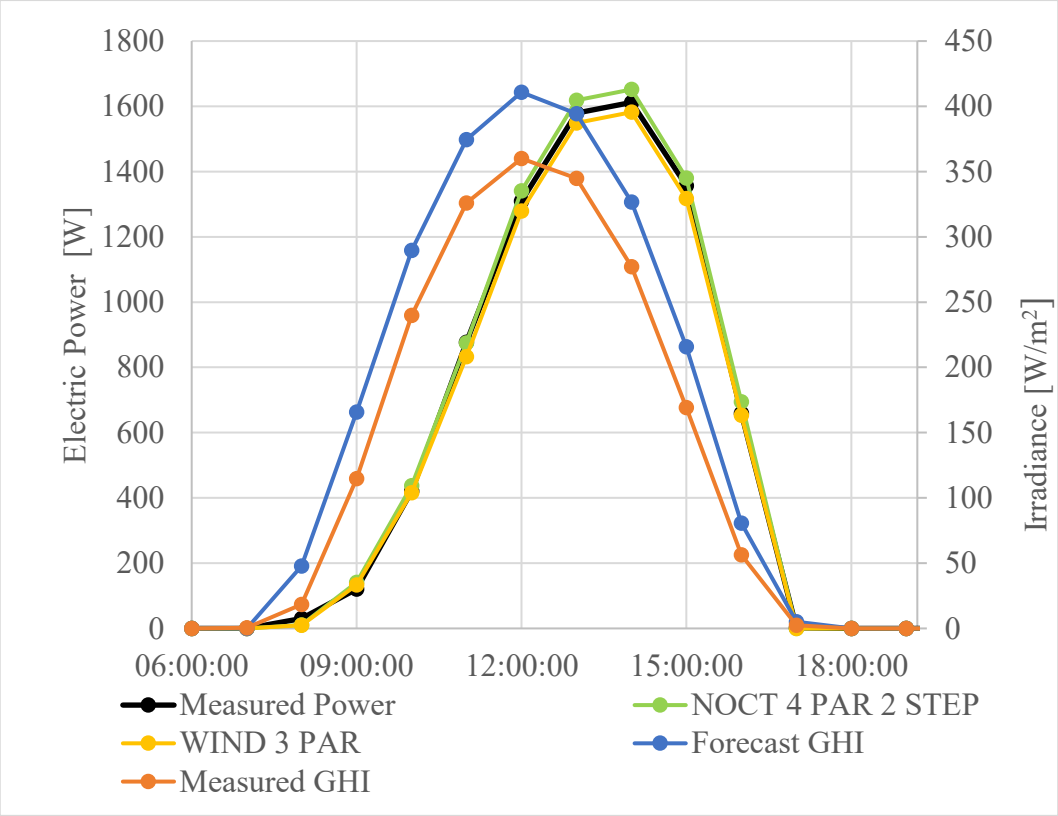

Figure 4-2: comparison between forecasted and measured irradiances, and measured and forecasted electric power for optimized models.

4.2.2 Effect of the optimization

In this sub-paragraph non optimized model profiles and their optimized versions will be compared.

NOCT MODELS: BASE VS OPTIMIZED

Figure 4-3 contains NOCT model, NOCT 4 PAR 2 STEP model and measured produced power. First of all, there is a general overestimation of produced power, that is coherent with the general irradiance deviation. Secondly, even if trends are quite close to each other, the optimized profile is slightly decreased, with the largest improvements in the first hours in the morning. In the below figure the oldest forecast profile (t-30) is considered. Table 4-3 contains on a daily basis the energy deviation of profiles shown in figure 4-3. Data refers to the day of 24/01/2022.

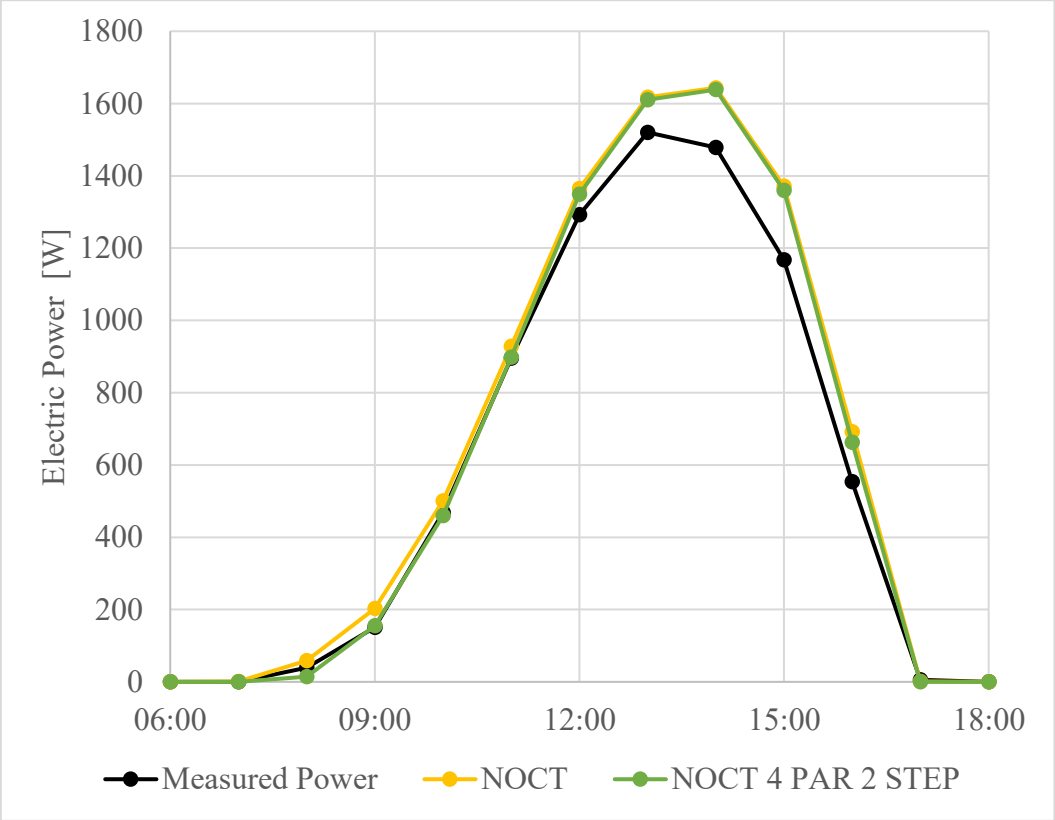

Figure 4-3: comparison between NOCT models: optimized and non-optimized.

Table 4-3

t-30	Measured	NOCT	NOCT 4 PAR 2 STEP
Energy [kWh]	7.57	8.38	8.15
Energy deviation [%]	-	10.7	7.64

WIND MODELS: BASE VS OPTIMIZED

Figure 4-4 contains WIND model, WIND 3 PAR model and measured produced power. The optimization is not punctual but global, therefore it may happen that in a limited number of points the optimized profile is more distant to the measurements than the non-optimized profile. In this case, the two central hours have been slightly penalized, but the whole profile during the whole day produced better results. Non-optimized model in this case always overestimates the production, while the optimized model slightly overestimates the production except from the 2 central hours. Data refers to the day of 19/01/2022, for which the oldest forecast profiles (t-30) are considered.

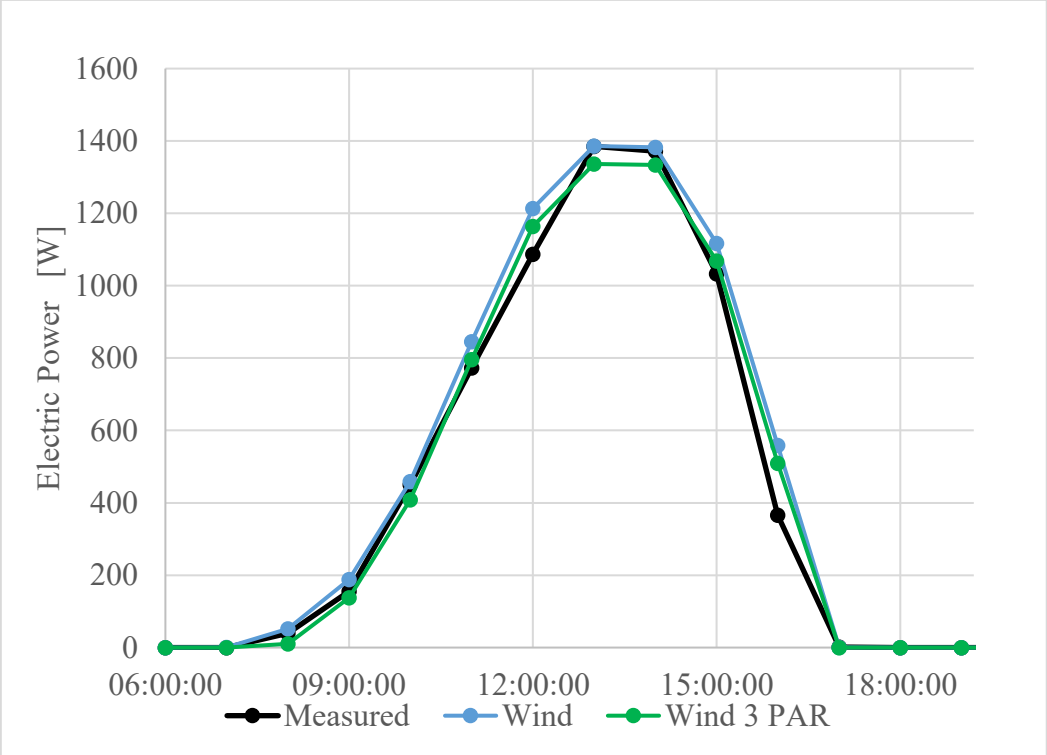

Figure 4-4: comparison between WIND models: optimized and non-optimized.

Table 4-5 reports the energy deviation calculated over the entire day associated to profiles shown in Figure 4-4.

Table 4-4

t-30	Measured	WIND	WIND 3 PAR
Energy [kWh]	6.55	7.20	6.76
Energy deviation [%]	-	8.15	1.55

4.2.3 Comparison between different meteorological conditions

In the following section, several profiles under different weather conditions will be compared. In the case of days with high variability or improvements in forecasts, also the most recent forecast is plotted and compared with the oldest.

EXAMPLE OF A REGULAR SUNNY DAY

Figure 4-5 shows NOCT and NOCT 4 PAR 2 STEP profiles for a generic sunny day and compared to measured values. The plot in the figure below refers to the oldest forecast profile, the so-called t-30 forecast profile, for the day of 15/01/2022, but it is representative of any sunny, clear sky day.

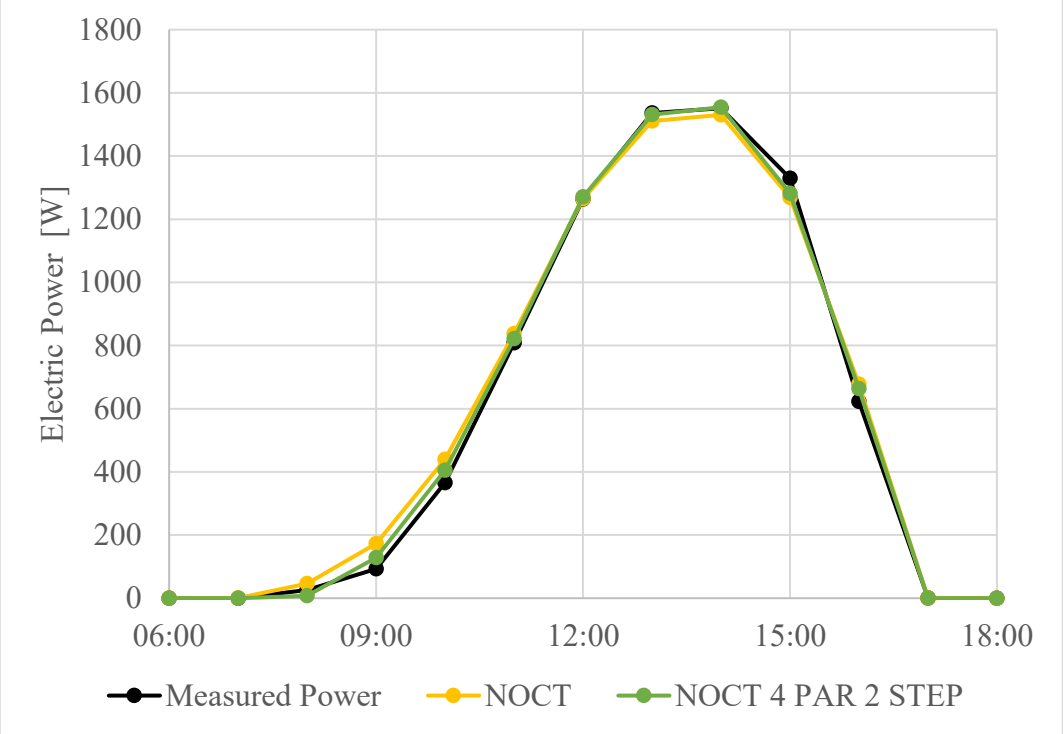

Figure 4-5: profiles example for a sunny day.

Table 4-5 contains on a daily basis the energy deviation of profiles shown in. Figure 4-5.

Table 4-5

t-30	Measured	NOCT	NOCT 4 PAR 2 STEP
Energy [kWh]	7.60	7.75	7.67
Energy deviation [%]	-	2.01	0.92

Figure 4-6 shows for the same day of 15/01/2022, the NOCT non optimized profiles for the 6-hours and 30-hours time leads. As introduced with the irradiance analysis in chapter 3, for very sunny days the benefit of having updated forecasts is almost negligible as the oldest and newest forecasts collapse into a single curve. In this case for example, a 24 hours more recent update just brings a 0.45% improvement in the residual.

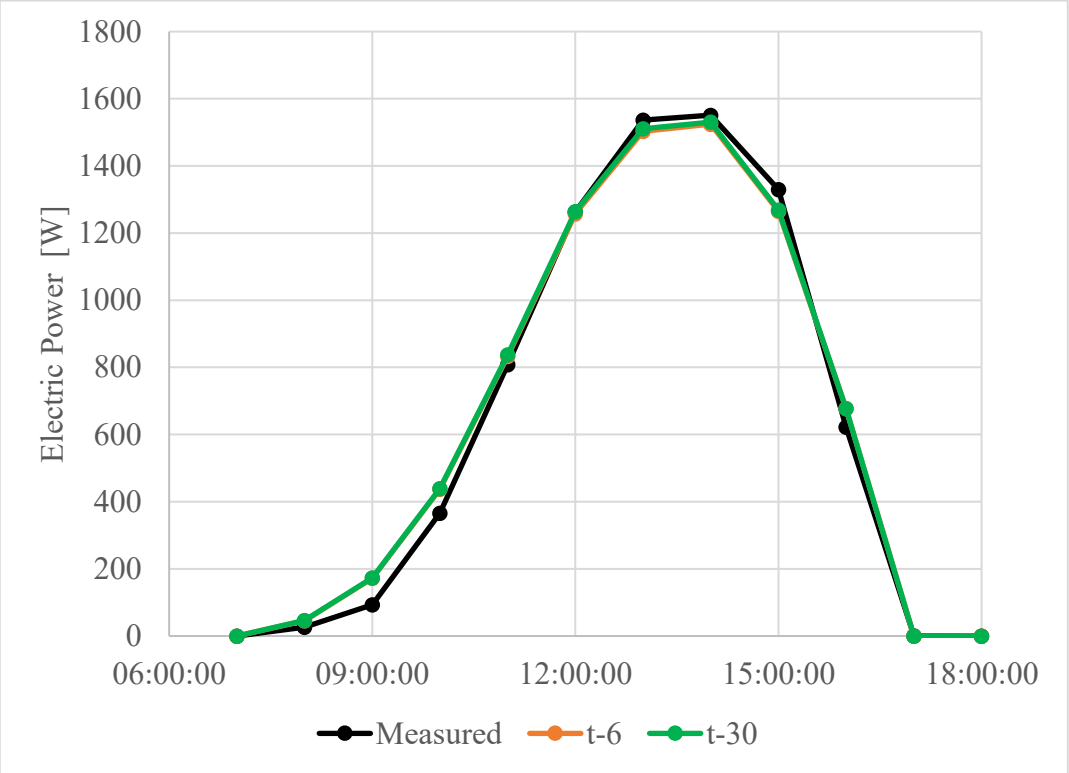

Figure 4-6: profile example for a sunny day; comparison between oldest and newest forecasts.

Table 4-6 reports the energy deviation calculated over the entire day associated to profiles shown in Figure 4-6

Table 4-6

		Measured	NOCT
t-6	Energy [kWh]	7.60	7.71
	Energy deviation [%]	-	1.56
t-30	Energy [kWh]	7.60	7.75
	Energy deviation [%]	-	2.01

Comparing therefore Figure 4-5 and Figure 4-6, it is possible to affirm that during sunny days characterized by modest or null forecast updating, the effect of model optimization affects more the final profile than the use of a more recent update. In paragraph 4.3 the low variability of forecast during sunny days will be more extensively indagated and discussed.

EXAMPLE OF A DAY WITH LARGE FORECASTS UNDERSTIMATION

Figure 4-7 contains the oldest forecast (t-30) for NOCT and NOCT 4 PAR 2 STEP models, for what it was predicted as an intermediately cloudy day. 31 January 2022 is chosen as a day initially forecasted as intermediately cloudy during the central hours of the day, and then corrected as sunny. The effect of optimization produces an almost identical power profile, even a little further from the real one. This behaviour is due to the fact that the optimization is implemented over the whole period, and as in general forecasts are optimistic, it has been noticed that the optimization tends to decrease in general the output profile. As a consequence, if forecasts strongly underestimate the irradiation and therefore the energy production, there could be situations like this one where the optimized model fails compared to non-optimized version. Table 4-7Table 4-5 summarizes on a daily basis the energy deviation of models during the day of 31/01/2022.

Table 4-7

t-30	Measured	NOCT	NOCT 4 PAR 2 STEP
Energy [kWh]	10.2	7.63	6.74
Energy deviation [%]	-	-25.0	-33.7

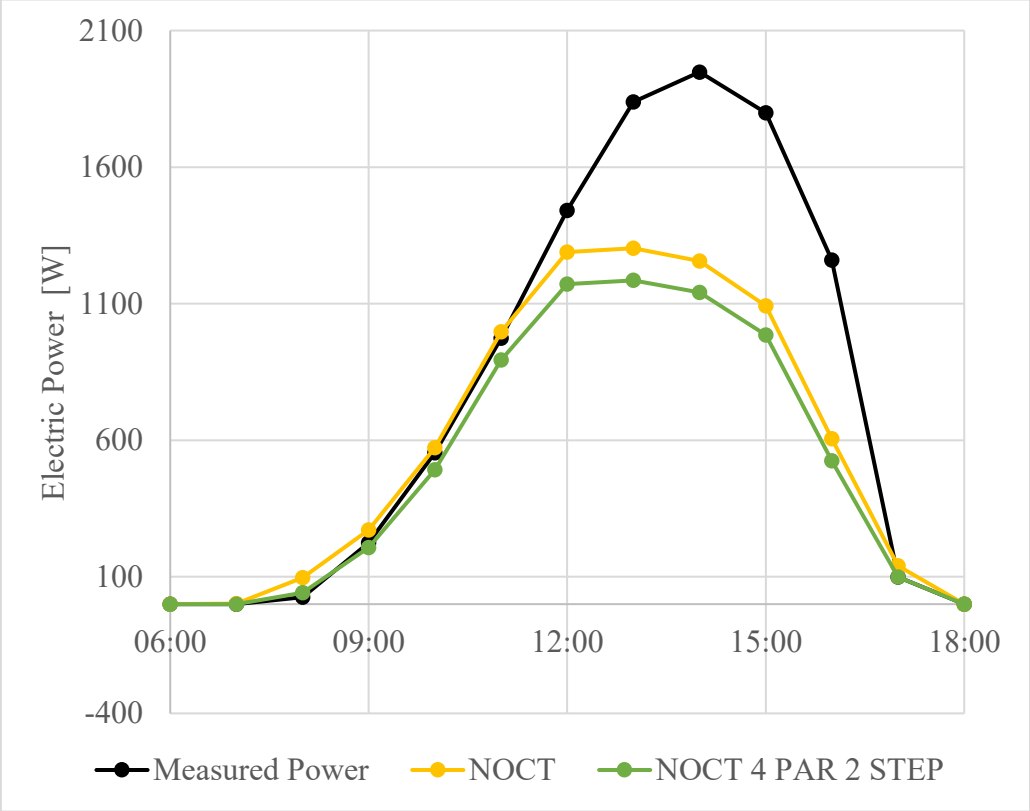

Figure 4-7: plot examples for a day whose initial guess had a high error; oldest forecast.

Figure 4-8 shows the same day and the same models, but with the most recent forecast profile (t-6). The values of the irradiance are still far from being correct, but the improvement from the original guess is remarkable. Furthermore, in this case the optimized model gives a slightly worse energy deviation, but this is due to the initial overestimation in the morning, while it better matches the actual production profile. Table 4-8 reports daily energy deviations related to the previously commented profiles.

Table 4-8

t-6	Measured	NOCT	NOCT 4 PAR 2 STEP
Energy [kWh]	10.2	8.62	8.48
Energy deviation [%]	-	-15.2	-16.6

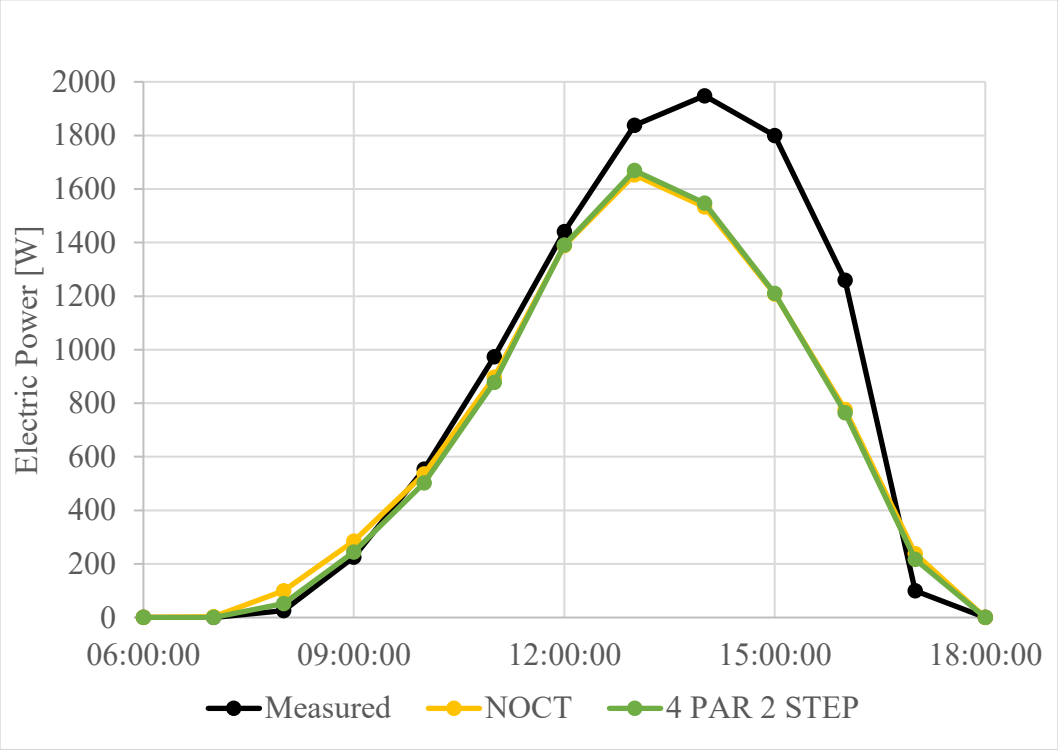

Figure 4-8: plot examples for a day whose initial guess had a high error; newest forecasts.

Figure 4-9 finally compares for the day of 31 January the oldest (t-30) and newest (t-6) profiles with actual measurements. As the focus is on the forecast update, only NOCT non-optimized model is plotted. Table 4-9 shows that for this model the energy deviation is improved from an initial -25% to a final -15.2% when updating forecasts. It emerged therefore that in the case of this types of days, characterized by an important irradiance underestimation, the optimization has no effect, while forecast update has the most relevant influence on results.

Table 4-9

		Measured	NOCT
t-6	Energy [kWh]	10.16	8.62
	Energy deviation [%]	-	-15.21
t-30	Energy [kWh]	10.16	7.63
	Energy deviation [%]	-	-25.0

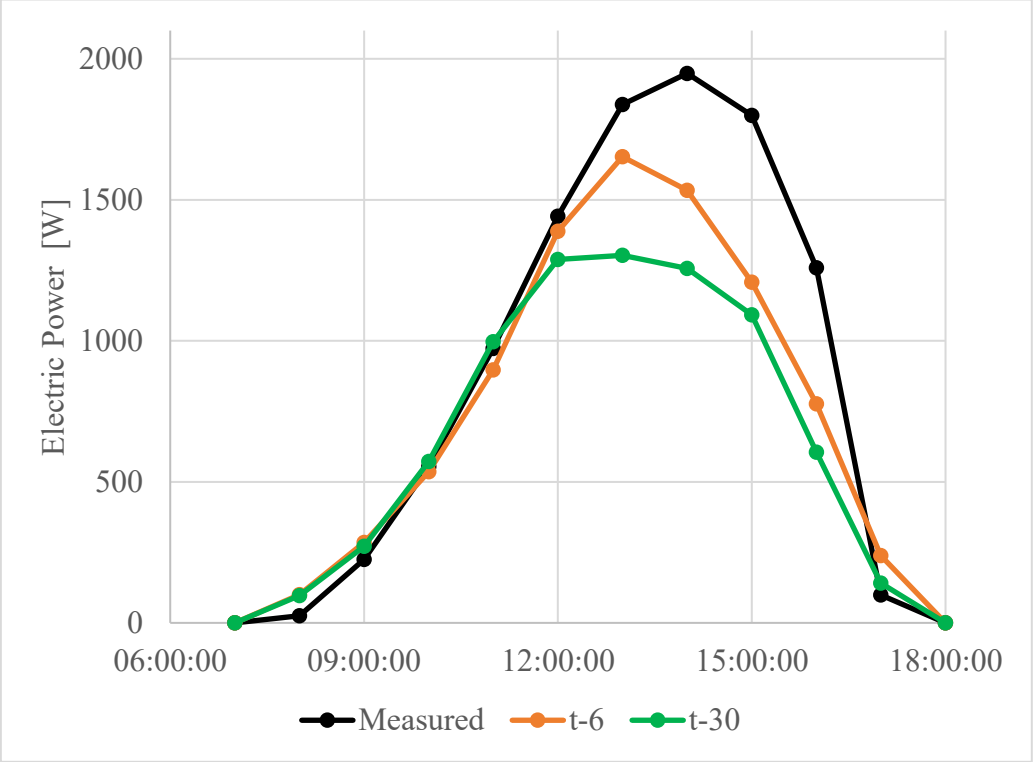

Figure 4-9: plot example for a day with significant improvement determined by forecasts update.

EXAMPLE OF A DAY WITH LARGE FORECASTS OVERSTIMATION

Finally, 6 February 2022 is considered as a very cloudy day example. Figure 4-10 shows the oldest profiles (t-30) of NOCT and NOCT 4 PAR 2 STEP models are compared to measured data. The optimized version improves the result of the NOCT model, but it is still distant from the measured data, with an energy deviation that is reduced from 134% just to 114%, as reported in Table 4-10. This large overestimation is determined by an upstream error in irradiance forecasts, which predicted this day to be sunny.

Table 4-10

t-30	Measured	NOCT	NOCT 4 PAR 2 STEP
Energy [kWh]	3.77	8.82	8.07
Energy deviation [%]	-	134	114

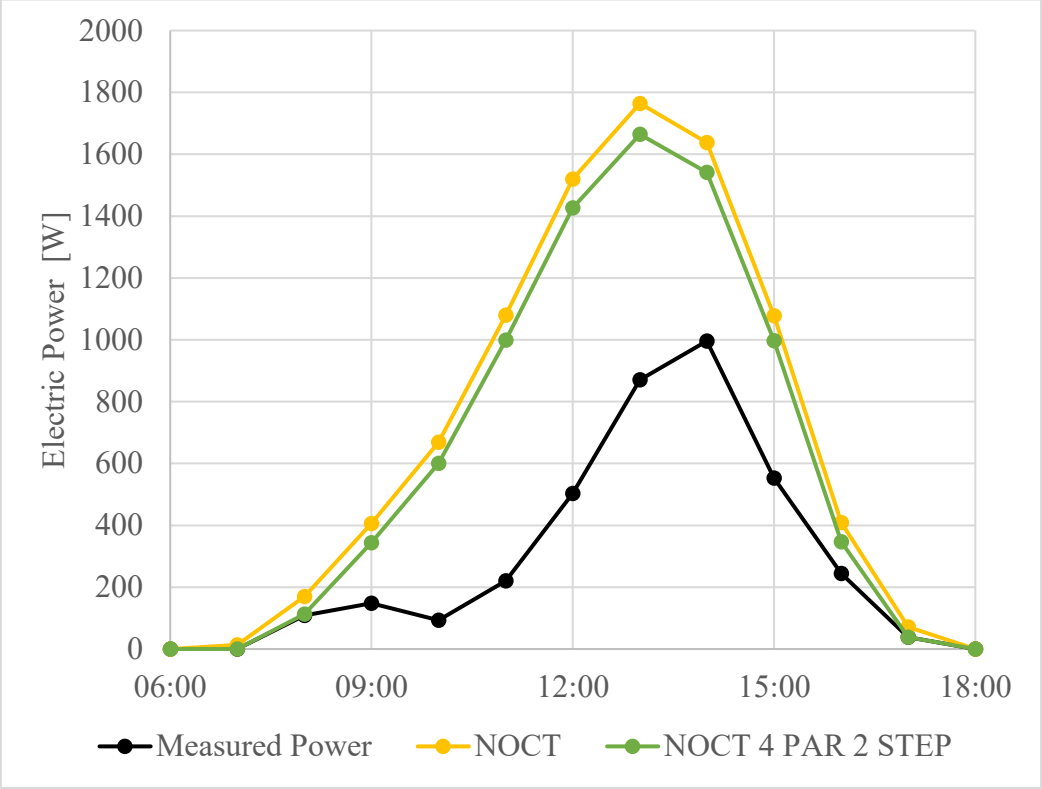

Figure 4-10: plot examples for a cloudy day, oldest forecast.

Figure 4-11refers to the same models and day of previous picture, but with the most recent (t-6) forecast profiles. As summarized in Table 4-11, the optimized NOCT 4 PAR 2 STEP model has better energy deviation calculated on the day, but it must be noted that the overestimation in the morning partially compensates the afternoon underestimation, leading to a global better performance. Considering the performance on the single timesteps, avoiding the daily smoothing effect, in some hours NOCT model is more suitable, and vice versa. Moreover, the 9:00-13:00 frame is characterized by a very high variability in irradiance and measured power, therefore it becomes unlikely for the forecast to exactly match that complicate profile.

Table 4-11

t-6	Measured	NOCT	NOCT 4 PAR 2 STEP
Energy [kWh]	3.77	5.08	3.17
Energy deviation [%]	-	34.7	-16.0

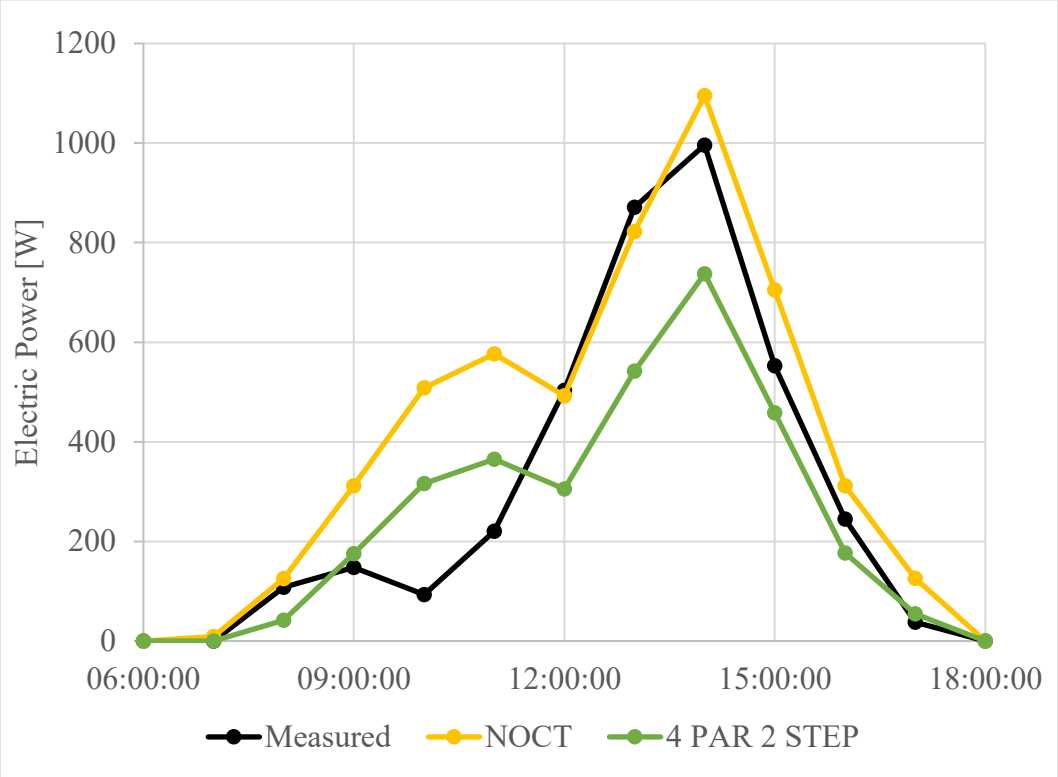

Figure 4-11: plot examples for a cloudy day, newest forecast.

Finally, Figure 4-12 summarize the previous 2 figures. Less recent (t-30) and more recent (t-6) profiles are compared with measured data, for the day of 06/02/2022. As the focus is on the forecast update, only NOCT non-optimized model is plotted. Table 4-12 summarizes the energy deviation of plotted profiles on a daily basis. Even if the energy deviation of the most recent profile (35%) is larger than the typical deviation of other days, it is a massive improvement compared to the initial guess, that led to an energy overestimation of 134%.

Table 4-12

		Measured	NOCT
t-6	Energy [kWh]	3.77	5.08
	Energy deviation [%]	-	34.7
t-30	Energy [kWh]	3.77	8.82
	Energy deviation [%]	-	134

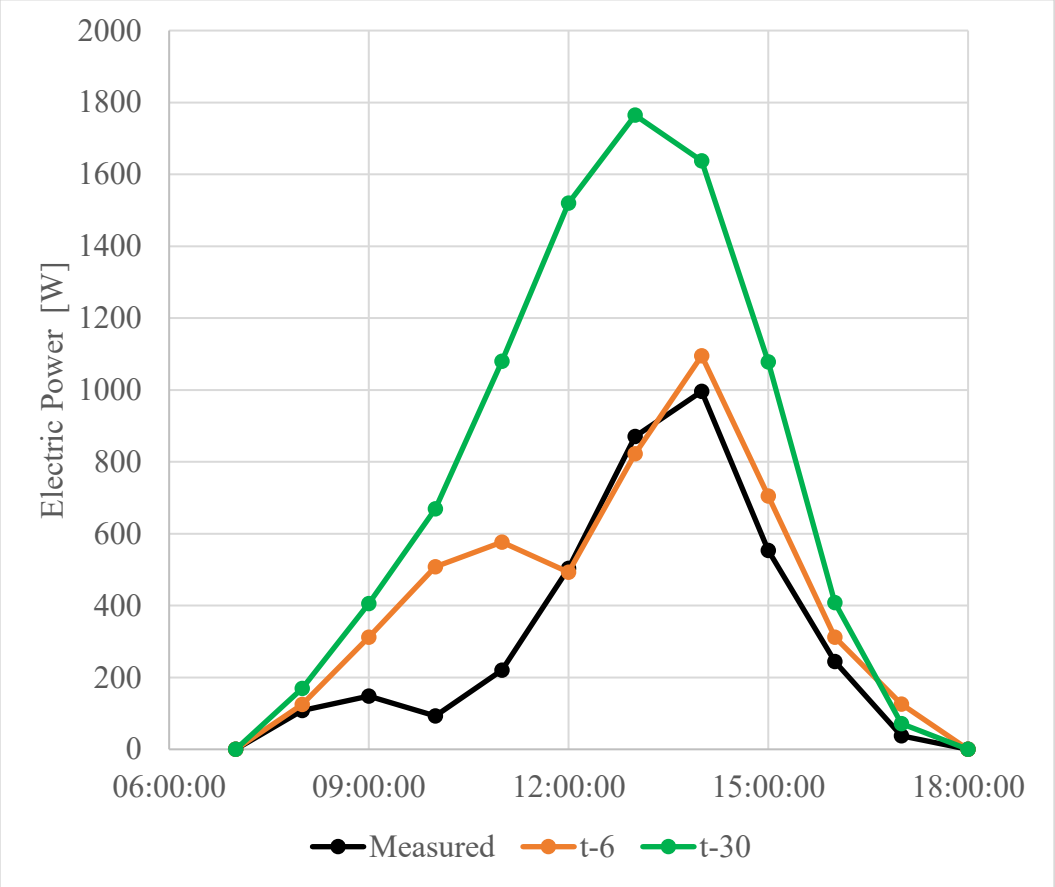

Figure 4-12: plot examples for a cloudy day, comparison between oldest and newest forecast.

In conclusion, in the case of very cloudy days, the effect of using more updated forecasts is predominant to the benefits brought by the optimization of the model. The optimization is indeed meant to produce an overall better matching between forecasts and measured data, without overturning completely the results. It is an expected results as if the initial forecast contains a very large error, the optimization process cannot modify the parameters value beyond certain limited boundaries.

4.2.4 Comparison between cloud models

The goal of this paragraph is to highlights the difference between cloud model and given GHI model. The focus is therefore on the nature of the input data used and not on the optimization. In this perspective, only the NOCT model will be analyzed.

CLOUD MODEL VS GIVEN GHI MODEL: SUNNY DAY

Figure 4-13 reports the oldest profile (t-30) obtained with NOCT model for the sunny day of 15/01/2022. In this picture the given GHI model, cloud model and actual power measurements are shown. For this day, cloud model performs better until 11:00, while in 12:00-15:00 given GHI models produces better results. Overall, results are very similar during a sunny day.

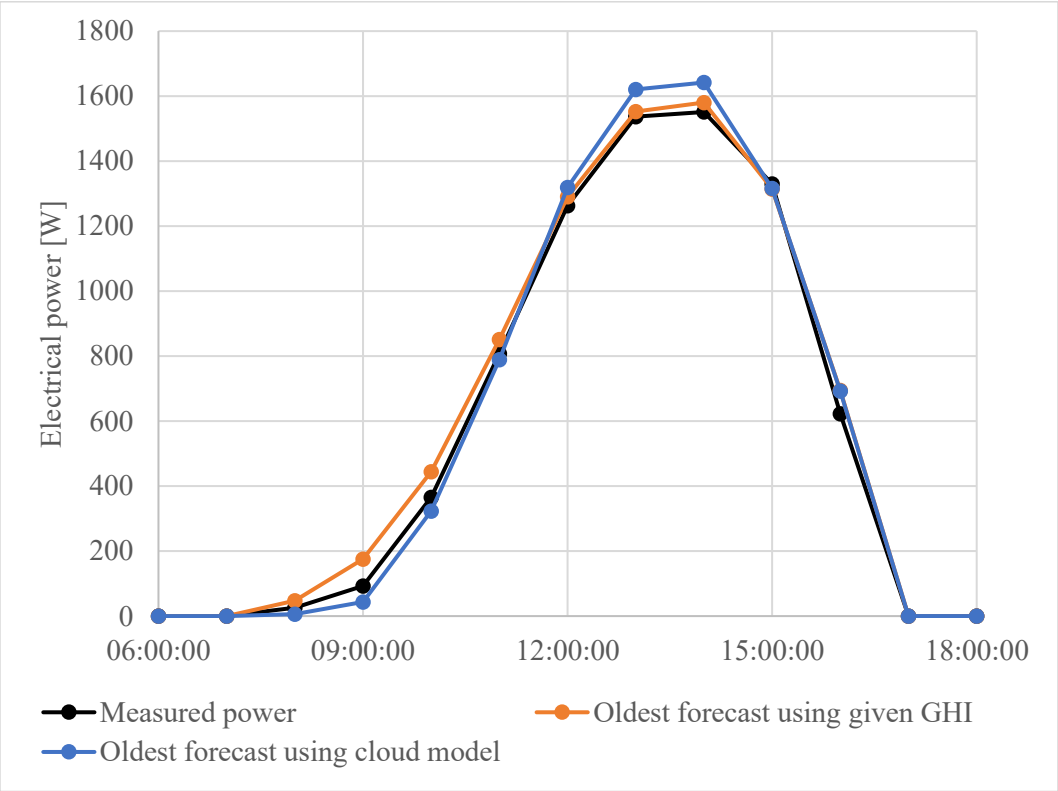

Figure 4-13: comparison between given GHI and cloud model, sunny day.

Table 4-13shows the energy deviation of the previous profiles calculated on the whole day. From this table it could seem that the cloud model has a better performance, but as it will be explained in paragraph 4.3, given GHI model has a lower mean absolute percentage error. On a daily basis, it may happen that positive and negative error compensate, leading to extremely low energy deviation, like in this case for cloud model.

Considering the absolute errors over the whole period, paragraph 4.3 will show that given GHI model produced better results than cloud model.

Table 4-13

t-24 - NOCT	Measured	Given GHI	Cloud Model
Energy [kWh]	7.60	7.71	7.65
Energy deviation [%]	-	3.94	0.71

CLOUD MODEL VS GIVEN GHI MODEL: CLOUDY DAY

Figure 4-14 Figure 4-13reports the oldest profile (t-30) obtained with NOCT model for the very cloudy day of 06/02/2022. In this picture the given GHI model, cloud model and actual power measurements are shown. The main difference is in the profiles shape: given GHI model output is steeper and more intermittent with respect to the cloud model one. Secondly, for this day the given GHI profile showed the best performance.

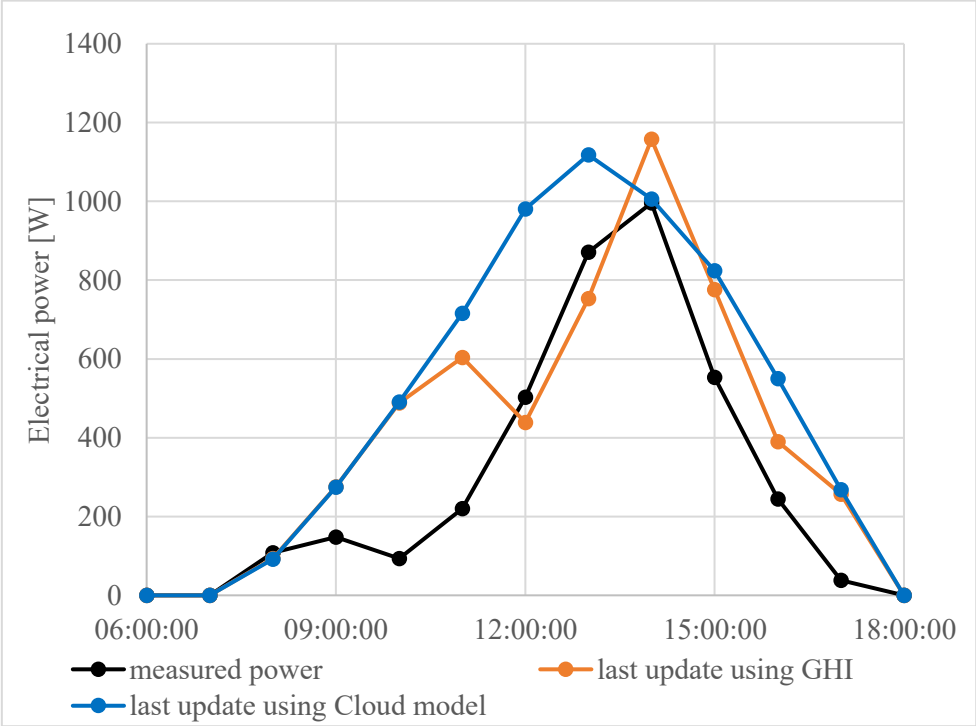

Figure 4-14: comparison between given GHI and cloud model, cloudy day.

Table 4-14 contains the daily energy deviation of the previous profiles. The error is fairly large, but it must be considered that it is a cloudy day.

Table 4-14

t-6 - NOCT	Measured	Given GHI	Cloud Model
Energy [kWh]	3.77	5.08	6.30
Energy deviation [%]	-	34.7	73.5

4.2.5 Daily error profile

As expected, the highest values of the relative error are at the sunset and at the sunrise. anticipated in paragraph 4.2, in first and last hours of the day very high relative errors were obtained, as shown in f**Errore. L'origine riferimento non è stata trovata.**. The figure below contains the relative and absolute error for each timestep for the sunny day of 15/01/2022. These errors refer to the results obtained with NOCT model and less recent (t-30) forecast profile.

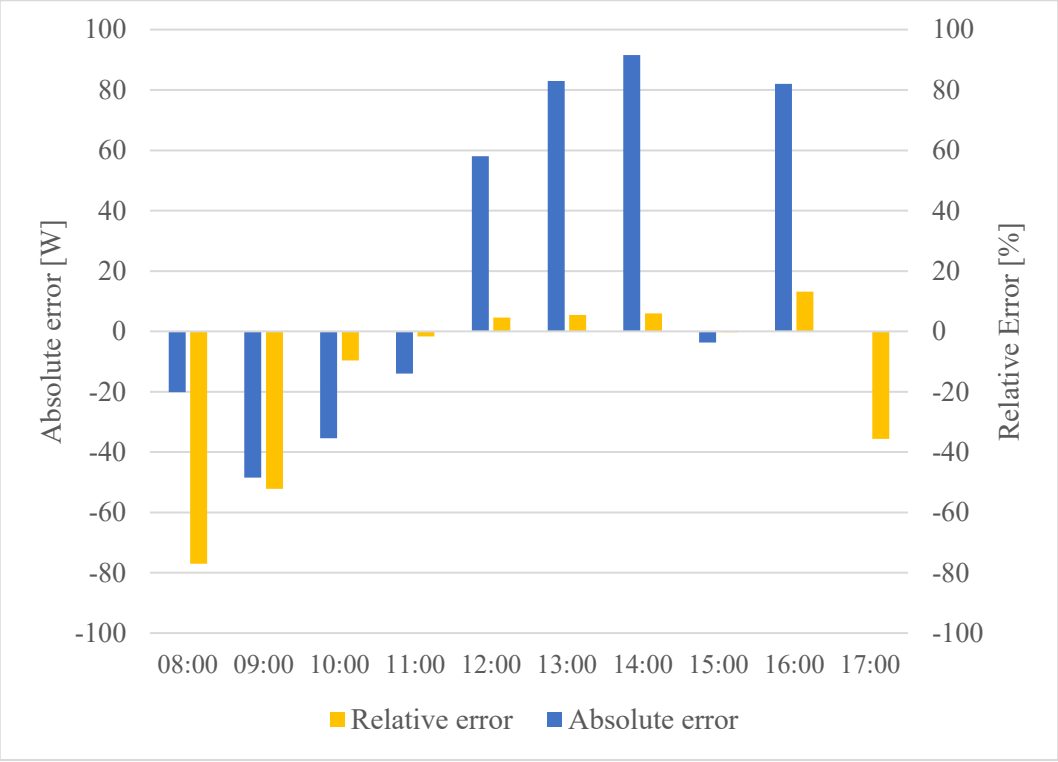

Figure 4-15: relative and absolute error for each timestep during a typical day

The high relative error at 8:00 and at 17:00 has almost no relevance, because the production at sunrise and sunset is almost equal to zero, so the denominator of the relative error is quite low, as a result the relative error increases by orders of magnitude. The highest absolute errors are located in the central hours of the day, but the lowest relative errors are associated to these timesteps. It is possible to notice that the central hours reflect the general overestimation in irradiance already extensively discussed in this chapter.

4.3 Energy deviation and MAPE tables

Given GHI model

Table 4-15 contains the energy deviation for the given GHI model. The first row is dedicated to irradiation deviation (GHI), while all the rest of the table refers to produced electrical energy deviation. The first row of the table highlights that irradiation forecasts are in general optimistic, with an irradiation overestimation in the range of 18-25%, with some improvements on the shorter time leads. On the other hand, electric energy results show a lower deviation, as positive and negative errors tend to compensate each other.

Table 4-15: energy deviation for all models and time leads.

Energy deviation [%]	t Post	t-6	t-12	t-18	t-24	t-30
Irradiation (GHI)	18.3%	18.6%	24.5%	23.9%	23.3%	23.8%
NOCT non-OTT	1.2%	-0.7%	4.7%	4.7%	4.3%	3.9%
Wind non-OTT	0.8%	-1.1%	4.3%	4.1%	3.9%	3.5%
NOCT 3 PAR	-4.6%	-6.7%	-0.7%	-0.9%	-1.1%	-1.5%
Wind 3 PAR	-5.0%	-7.1%	-1.1%	-1.3%	-1.1%	-2.0%
NOCT 4 PAR	0.9%	0.9%	0.9%	0.9%	0.8%	0.8%
Wind 6 PAR	6.7%	6.0%	6.0%	6.1%	6.2%	5.9%

With some models an energy deviation close to 0% is achieved, despite the error in forecasts is much higher (range of 18-24 %). To avoid the smoothing effect over the whole period, Table 4-16 introduces the MAPE for all the models and time

leads. As for the previous table, the first row is dedicated to irradiation deviation (GHI), while all the rest of the table refers to produced electrical energy deviation. In the case of MAPE, the deviation is slightly higher regarding irradiance forecasts, while the deviations concerning electrical power are in a range much closer to the input error introduced by forecasts. As expected, the post elaboration produced by far the best results, while unexpectedly, from that point of view the t-6 profile had the worst performance. The NOCT 4 PAR model produced the best result, while WIND 6 PAR was not able to reduce the error.

Table 4-16: MAPE for all models and time leads

MAPE [%]	t Post	t-6	t-12	t-18	t-24	t-30
Irradiation (GHI)	19.5%	25.4%	25.1%	24.7%	24.9%	25.1%
NOCT non-OTT	11.5%	17.7%	14.7%	14.6%	14.4%	13.7%
Wind non-OTT	11.4%	17.5%	14.5%	14.4%	14.2%	13.5%
NOCT 3 PAR	12.0%	18.1%	14.8%	14.6%	14.5%	13.8%
Wind 3 PAR	12.2%	18.2%	14.9%	14.7%	14.5%	13.9%
NOCT 4 PAR	10.1%	18.4%	13.3%	12.9%	12.7%	12.0%
Wind 6 PAR	21.8%	29.4%	23.4%	22.7%	23.2%	22.9%

Cloud model

In a similar way, energy deviation of all models and time leads is reported for cloud model in Table 4-17. The first difference is that there is not the first row of forecast GHI error because, for cloud models, the output power is calculated on the basis of clear sky radiation, and then that power is corrected by cloud cover coefficients. In this case, the electric energy produced is always overestimated, in the range of +(1.8%-10%). Optimized models determine a reduction of energy deviation, with the best performance achieved by NOCT 4 PAR. For energy deviation, the forecast lead has a weak influence on the results.

Table 4-17: linear normalised error for all models and time horizons – cloud model

Energy deviation [%]	t Post	t-6	t-12	t-18	t-24	t-30
NOCT non-OTT	9.5%	9.9%	9.7%	9.7%	9.8%	9.7%
Wind non-OTT	9.5%	9.9%	9.8%	9.9%	9.9%	9.8%
NOCT 3 PAR	5.4%	5.5%	5.7%	5.9%	5.8%	5.7%
Wind 3 PAR	5.4%	5.5%	5.7%	5.9%	5.8%	5.7%
NOCT 4 PAR	2.1%	1.8%	2.0%	2.0%	1.9%	2.0%
Wind 6 PAR	3.2%	2.5%	2.7%	2.7%	2.8%	2.7%

As introduced for given GHI model, MAPE is introduced to avoid the compensation of underestimations and overestimations of energy produced. Table 4-18 shows MAPE for all models and time leads for cloud model. In general, the results are very similar to given GHI model. In both cases Wind 6 PAR produces the highest error, while for all the other models the MAPE of cloud model is slightly higher (in the range of 14.5-19%) than given GHI model (in the range of 10-18%). Moreover, also in this case the best results were obtained with NOCT 4 PAR. The main difference is determined by the modest variations in MAPE for the different time leads, even for the post elaboration.

Table 4-18: MAPE for all the models and time leads – cloud model

MAPE [%]	t Post	t-6	t-12	t-18	t-24	t-30
NOCT non-OTT	18.4%	18.9%	18.8%	18.9%	18.7%	18.8%
Wind non-OTT	18.6%	19.1%	19.0%	19.1%	18.9%	19.0%
NOCT 3 PAR	18.1%	18.7%	18.4%	18.4%	18.3%	18.3%
Wind 3 PAR	18.2%	18.8%	18.5%	18.5%	18.4%	18.4%
NOCT 4 PAR	14.5%	17.5%	17.2%	17.2%	17.1%	17.0%
Wind 6 PAR	28.1%	30.1%	29.1%	29.1%	29.3%	29.3%

From the tables in the previous chapter, it could seem that the effect of the update of forecasts does not bring a relevant improvement. At the same time, it is found out that within a single day timespan, the updating of forecasts can remarkably improve the result with some extreme cases of energy error reduction from ~300% to ~60% as shown in Figure 3-7. This is due to the fact that, even if considering winter period, the majority of days is sunny, and it has been observed that in the case of sunny days, the different time lead profiles tend to collapse into one single profile that shows a negligible variation. As the majority of days of the considered period is sunny, the large improvements during cloudy days are smoothed by the largely more abundant clear sky days.

Figure 4-16 shows in a graphic way what described previously. Despite the low irradiance values due to winter period, the most part of the days are sunny, therefore already the first forecast profile is close to the real profile, while the last update brings a little or null improvement. In the figure below it is plotted the measured value of GHI for the January-February period to show that in the considered period, sunny days were largely predominant. Moreover, for some days the relative error is reported; the red caption refers to oldest profile while the green caption represents the newest forecast update. It is possible to notice how relevant is the improvement during cloudy days, while for the regular sunny days the benefit of updated profile is lost. It follows that on the whole period, the global result will be much more affected by the majority of these sunny days.

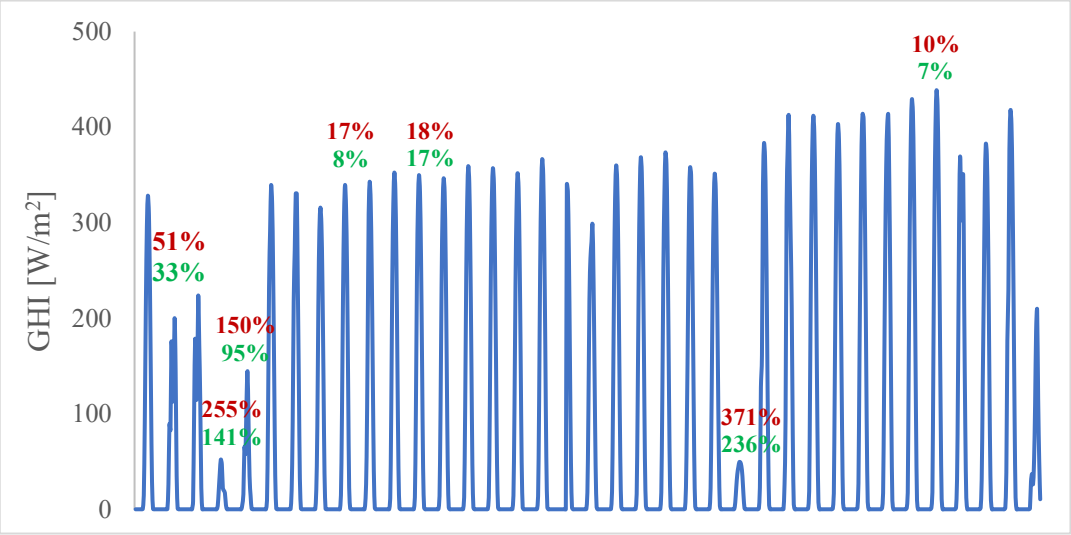

Figure 4-16: Daily profiles during 2 winter months.

Table 4-19 reports MAPE calculated on a daily basis, for the same period of figure 4-16.

Table 4-19: daily absolute mean percentage error [%].

Date	t-post	t-6	t-12	t-18	t-24	t-30
01/01/2022	8.8	7.0	8.0	7.8	8.7	7.8
02/01/2022	33.3	47.5	33.3	32.2	47.7	51.2
03/01/2022	27.1	34.7	35.0	37.5	33.1	34.9
04/01/2022	141.0	212.4	254.8	205.9	163.7	180.0
05/01/2022	94.9	131.7	165.7	176.1	159.4	149.6
06/01/2022	18.3	17.9	17.1	18.6	18.7	19.6
07/01/2022	17.9	16.7	17.1	16.5	18.8	19.6
08/01/2022	7.7	10.8	6.7	14.0	20.2	18.6
09/01/2022	17.0	16.4	12.7	14.4	12.8	7.5
10/01/2022	23.3	25.1	25.7	26.1	26.9	28.3
11/01/2022	16.2	16.6	17.2	17.1	16.9	16.9
12/01/2022	17.4	17.4	17.2	16.8	16.9	17.6
13/01/2022	16.5	17.3	17.4	17.3	17.4	17.2
14/01/2022	13.5	14.1	13.9	14.1	14.4	14.3
15/01/2022	15.2	15.7	15.7	15.6	15.7	15.9
16/01/2022	18.9	19.4	19.3	19.2	19.3	19.3
17/01/2022	14.9	14.9	14.7	14.8	14.8	15.0
18/01/2022	92.8	90.9	90.6	91.1	92.0	92.3
19/01/2022	33.4	37.9	37.1	32.2	22.5	29.7
20/01/2022	20.4	20.7	21.1	21.3	21.3	21.1
21/01/2022	18.5	18.5	18.4	18.5	18.7	18.7
22/01/2022	13.4	17.1	17.3	17.4	17.5	17.8
23/01/2022	21.5	24.6	24.9	25.2	25.4	25.3
24/01/2022	23.8	27.5	28.1	26.4	26.8	27.5
25/01/2022	235.9	276.0	320.0	400.0	431.3	371.3
26/01/2022	15.4	14.2	17.1	17.3	15.6	16.8
27/01/2022	16.1	13.7	18.8	19.5	18.2	16.5
28/01/2022	7.5	9.0	8.2	12.0	12.9	14.6
29/01/2022	12.2	14.4	14.0	12.5	14.5	15.1
30/01/2022	10.5	12.8	13.9	14.0	14.1	14.3
31/01/2022	7.9	27.0	15.1	15.9	17.5	17.1
01/02/2022	7.2	11.0	10.9	11.1	9.8	10.3
02/02/2022	6.1	6.7	6.4	6.5	11.1	8.7
03/02/2022	18.2	15.0	18.2	20.5	21.2	25.4
04/02/2022	18.5	23.3	23.1	17.7	23.0	22.8
05/02/2022	13.2	14.0	14.4	13.6	16.9	16.7

The trend is very discontinuous, but it can be seen that in general, for cloudy days, that are the ones with highest MAPE, the error tends to improve as forecasts lead decreases

4.4 Error profiles

Commenting the results of the previous tables, it emerged that the update of forecast does not bring a remarkable error reduction. This paragraph is then dedicated to the analysis of the quality of forecasts. In particular it has been studied, for four locations. the evolution of global horizontal irrigation profiles. Each location is located in a different Italian region: Piedmont, Lombardy, Latium and Sicily.

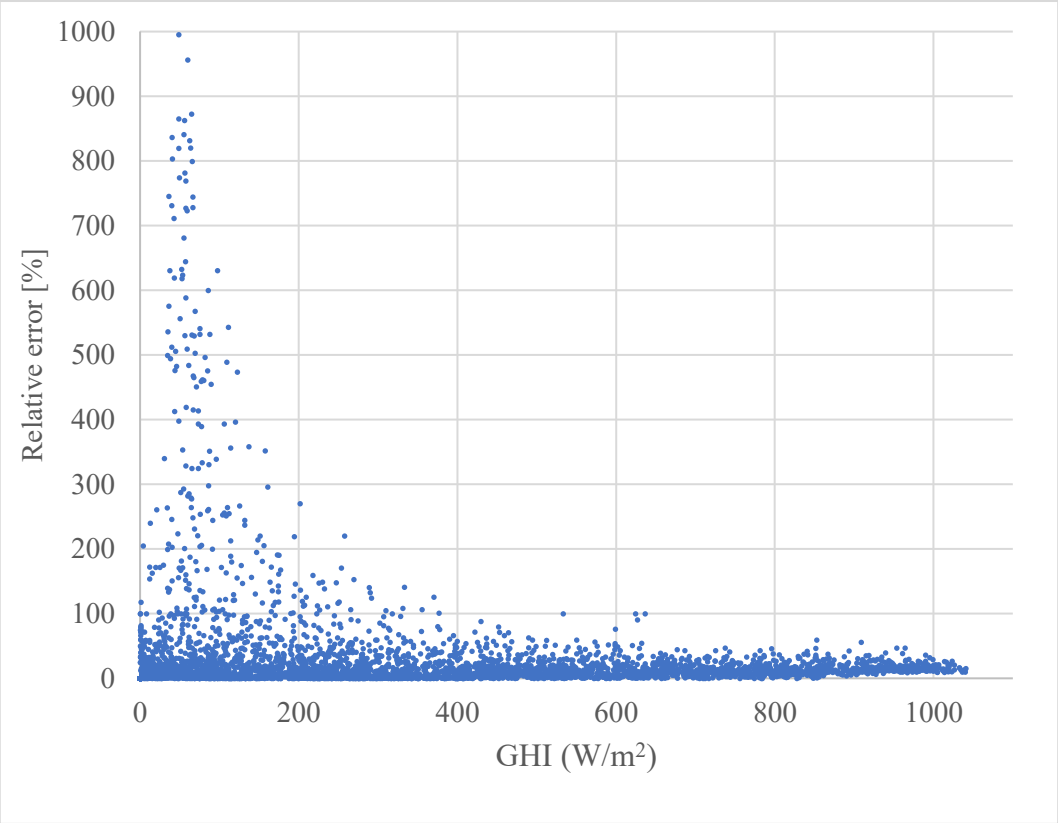

Figure 4-17: Error distribution profile.

For each time lead and each timestep the linear relative error is calculated, then it is averaged over the whole period. In PV applications it is common that the highest relative errors occurs at sunrise and sunset (see Figure 4-15) and in general when in presence of low values of irradiance, as shown in Figure 4-17. It can be clearly

seen that errors over 100% are almost entirely confined for GHI values inferior to 300 W/m^2, and that the extremely large errors all occur below 100 W/m^2. The following graph is obtained considering the less recent (t-30) forecast profile considering location Rivalta for region Piedmont.

The forecast lead time alone is therefore not sufficient to describe in a detailed way the quality of forecasts. It is indeed introduced the further distinction among irradiation classes. This subdivision is made because it is expected to have very high errors during these times of the day with low irradiance, but from a energetic point of view on the whole period, these timesteps will have a reduced effect. The error is then monitorized along two axis: on the time axis (from oldest to most recent forecast) and on the class axis, that takes into account the absolute value of the quantity on which the error is calculated. More in detail, all the timesteps are classified by the value of GHI, each class has a 100W/m^2 wide interval, like shown in tables 4.19-22.

Table 4-20 contains result for the location of Rivalta, Torino. The error is located mainly in the first two classes (71,4-75,1% for 0-100 W/m^2 class and 32-37,5% for 100-200 W/m^2 class). In the interval 0-200 W/m2, the improvement of updated forecasts (around 5%) is much less relevant compared to the improvement by moving from one class to the next one (around 40% from first to second and 17% from second to third. In the interval 300-1000 W/m^2, in general the error keeps reducing, but at a slower rate and not in a monotonous way. Moreover in this range, in general, forcasts still keeps improving from oldest to newest but not in every single situation. Finally, the last class (GHI>1000 W/m^2) is not very signifivative because it contains a very limited number of points.

Table 4-20

Piedmont region		Forecast lead				
Class		t-6	t-12	t-18	t-24	t-30
GHI (W/m^2)	0-100	71.4	72.2	72.6	77.4	75.1
	100-200	32.0	35.1	34.1	37.0	37.5
	200-300	17.2	17.6	17.6	18.1	19.1
	300-400	13.6	13.3	13.5	14.5	15.0
	400-500	16.6	15.7	15.6	15.1	15.2
	500-600	16.3	16.7	15.9	16.4	16.5
	600-700	15.6	14.4	14.5	14.3	15.0
	700-800	12.7	12.6	13.6	13.0	13.3
	800-900	16.0	13.9	14.1	14.4	14.6
	900-1000	15.0	16.9	17.3	17.4	17.2
	>1000	13.1	15.8	16.5	16.0	16.0

Table 4-21 shows results for province of Viterbo, located in Latium region, which are very similar to Piedmont region trends. The errors in the first two classes are even larger compared to Piedmont case (116%-124% for 0-100 W/m^2 class and 43,2%-46,6% for 100-200 W/m^2 class), while for values of GHI > 200 W/m^2 the values converge into results collacted in a similar interval to the one of the previous region.

Table 4-21

Latium region		Forecast lead				
Class		t-6	t-12	t-18	t-24	t-30
GHI (W/m^2)	0-100	115.9	111.4	122.3	123.8	124.0
	100-200	43.2	42.6	44.9	45.7	46.4
	200-300	19.6	20.7	22.2	22.6	22.8
	300-400	16.1	15.1	15.8	15.6	15.6
	400-500	14.2	13.9	13.6	13.7	13.5
	500-600	15.0	13.8	14.4	14.0	14.0
	600-700	13.2	13.6	14.9	15.5	15.4
	700-800	12.3	13.0	12.7	13.0	12.6
	800-900	13.7	14.0	14.1	14.5	14.2
	900-1000	12.8	15.1	14.9	14.7	14.2
	>1000	11.3	14.1	14.1	13.9	13.6

Table 4-22 contains results for province of Cremona, located in Lombardy region. The results are very similar to what described for Piedmont. GHI > 1000 W/m^2 class exhibits in general an absolute error comparabel to 200-300 W/m^2 class, but it is not very significative because it contains only a very limited number of points. Except from this last row, the general trend confirms the reduction of the error when GHI increases and forecast time lead decreases.

Table 4-22

Lombardy region		Forecast lead				
Class		t-6	t-12	t-18	t-24	t-30
GHI (W/m^2)	0-100	81.0	88.2	88.0	104.8	101.3
	100-200	31.8	32.7	35.3	35.3	35.4
	200-300	17.2	17.5	17.9	18.7	18.5
	300-400	17.0	16.4	18.1	18.0	17.7
	400-500	18.2	17.8	17.8	17.9	17.5
	500-600	15.1	15.7	15.3	15.8	16.0
	600-700	15.1	15.7	15.8	15.6	16.5
	700-800	13.6	13.9	14.3	13.9	14.0
	800-900	14.2	15.0	15.3	15.8	15.4
	900-1000	13.5	15.4	15.9	16.5	15.8
	>1000	10.8	17.1	18.1	18.5	18.6

Finally, table 4-23 shows results for Sicily region, which are very similar to previous cases. Except from the cell 800-900 W/m^2 class – t-6 forecast, we can see aprogressyve improvement when increasing the value of GHI and reducing the time lead of forecasts. In all four cases the most accurate class is the one between 700 and 800 W/m^2.

Table 4-23

Sicily region		Forecast lead				
Class		t-6	t-12	t-18	t-24	t-30
GHI (W/m2)	0-100	81.7	84.4	86.1	92.3	90.3
	100-200	32.6	35.4	34.9	38.0	38.0
	200-300	18.7	18.8	19.1	19.5	20.1
	300-400	15.7	15.4	15.8	16.5	17.4
	400-500	15.8	15.3	15.3	15.2	14.9
	500-600	14.5	14.9	14.2	14.6	14.7
	600-700	15.1	14.0	14.1	13.8	14.4
	700-800	12.8	12.6	13.6	13.1	13.4
	800-900	16.0	13.9	14.1	14.4	14.8
	900-1000	15.0	16.9	17.3	17.4	17.2
	>1000	13.1	15.8	16.5	16.0	16.0

Conclusions

In the thesis work, chapter 4 contains the results obtained from a physical forecast model for the calculation of the hourly power for photovoltaic systems, on the basis of weather forecasts and construction data of the plants under analysis. Using forecast data, it is essential to identify how accurate forecasts are in order to understand if the final error is mainly determined by the uncertainty of the input data (forecasts) or if it depends on the power calculation model, described in chapter 2, applied afterwards. Even if more meteorological parameters are used, the focus is on irradiation as it is the factor with the major influence on the PV production. For the following conclusions two different time periods will be considered: the January-February period which contains also all the electric power measures, and the whole 2022 year in which only the part of irradiation forecasts is analysed.

For what it concerns the irradiation during the winter period, both in chapter 3 and 4, it was found that forecasts are in general optimistic, with an overestimation of irradiance, averaged on the whole period, around 20% (18,3% for the most update profile and 24% for the oldest profiles). These input error will be reflected downstream the process. Furthermore, calculating the average error of the different time lead profiles, it emerged that, except from the last update profile, there is not large improvement when considering more recent forecasts. On the other end, focusing on single days, there are some cases in which the newer profiles progressively correct a wrong initial forecast, decreasing from a starting 300% relative error to a final 60% relative error. By the way, from an energetical point of view, that is averaging on the whole period, despite being during winter, the large majority of days is composed of sunny days. As a result, the weight of these cloudy days are negligible on the final result, which are dominated by the sunny days leading to a weak influence of updated forecasts on the final result.

As second step, again regarding irradiation, forecasts were analysed on the whole 2022 year, in four different points associated to four different Italian regions. In this case the error was calculated in function of the time lead of the forecasts and of the value of GHI itself. For all the four regions, it emerged that there is a gradual improvement when the time lead decreases (on average it reduces by 1/8 – 1/10), and a more remarkable error reduction for higher values of GHI. More precisely there are very large errors in the class 0-100 W/m2, which vary among the regions

but are all in the range of 70-120%, with some reductions as the time lead decreases. The error still remains very high in the class 100-200 W (in the range of 30-40%), and then keeps reducing while the class increases but at a slower rate, always ranging between 10 and 19%. As expected, the highest values of the relative error are associated to the first classes, which in general correspond to sunset and sunrise timesteps, associated to high errors in PV and solar applications.

Concerning the results of electric power, they refer to the January-February 2022 period. As anticipated, these results are affected by the irradiance forecast input. All the models produced a mean absolute percentage error in the range of 19-27%, an interval comparable to the input irradiation overestimation of 21-33%. In a similar way to what it has been noticed for forecasts, for the majority of sunny days the oldest power profile does not undergo relevant improvements (in the range of 0.5-1%), therefore for these types of days the effect of optimization is predominant over forecast updating. On the other hand, for cloudy days newest forecast are decisive for the model performance, with an energy deviation that can drop from 134% to 33%. For these types of day, which represent a small minority, the effect of optimization (from 134% to 120 %) is an order of magnitude lower compared to the massive improvement brought by forecasts update.

In the next thesis work, a similar study will be extended to a larger number of plants for a more extended period.

5 Bibliografia

1] «https://ec.europa.eu/eurostat/web/energy/data/database,» [Online].

2] «https://www.terna.it/it/sistema-elettrico/transparency-report,» [Online].

3] [Online]. Available: https://gml.noaa.gov/ccgg/trends/.

4] *Dispense del corso "Energy Economics" - Politecnico di Torino - D. Chiaramonti.*

5] [Online]. Available: https://ec.europa.eu/clima/policies/strategies/progress_it,.

6] [Online]. Available: https://www.eea.europa.eu/highlights/eu-achieves-20-20-20.

7] [Online]. Available: https://www.terna.it/it/sistema-elettrico/ruolo-terna/come-funziona-sistema-elettrico.

8] *Markvart, Tom, and Luis Castañer. "Principles of solar cell operation." McEvoy's Handbook of Photovoltaics. Academic Press, 2018. 3-28..*

9] [Online]. Available: https://www.ise.fraunhofer.de/content/dam/ise/de/documents/publications/studies/Photovoltaics-Report.pdf.

Dispense del corso "Solar Photovoltaic systems".

10]

Dispense del corso "Distribuzione e utilizzazione dell'energia elettrica" -
11] *Politecnico di Torino- G. Chicco- 2020.*

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison,
12] *F. J., & Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. Solar energy, 136, 78-111..*

Pierro, M., Bucci, F., De Felice, M., Maggioni, E., Perotto, A., Spada, F.,
13] *... & Cornaro, C. (2017). Deterministic and stochastic approaches for day-ahead solar power forecasting. Journal of Solar Energy Engineering, 139(2)..*

[Online]. Available:
14] https://meteonorm.com/assets/publications/Photovoltaic_and_Solar_Forecasting_State_of_the_Art_REPORT_PVPS__T14_01_2013.pdf.

Karabacak, Kerim, and Numan Cetin. "Artificial neural networks for
15] *controlling wind–PV power systems: A review." Renewable and Sustainable Energy Reviews 29 (2014): 804-827..*

Schneider, M., Beyer, H., Heinemann, D., Karampela, G., Hurka, J., &
16] *Lorenz, E. (2008). Qualified Forecast of Ensemble Power Production by Spatially Dispersed Grid-Connected PV Systems..*

Junior, Joao & Oozeki, Takashi & Ohtake, Hideaki & Shimose, Ken-ichi
17] *& Takashima, Takumi & Ogimoto, Kazuhiko. (2014). Regional forecasts and smoothing effect of photovoltaic power generation in Japan: An approach with principal component analysis. Renewab.*

[Online]. Available: https://www.terna.it.

18]

19] *Tesi di Laurea Magistrale, Modello previsionale di produzione di impianti fotovoltaici e applicazione statistica su larga scala, Alba Giuseppe..*

20] [Online]. Available: https://www.meteomatics.com/en/about-us/.

21] *TYang, D., Wang, W., & Hong, T. (2022). A historical weather forecast dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) for energy forecasting. Solar Energy, 232, 263-274..*

22] *Tesi di Laurea Magistrale, Modelli per la previsione della potenza oraria prodotta da impianti fotovoltaici: analisi delle prestazioni ed ottimizzazione, Iemmola Roberto.*