

Elsevier Editorial System(tm) for Renewable & Sustainable Energy Reviews  
Manuscript Draft

Manuscript Number:

Title: Very Short-Term Photovoltaic Power Forecasting with Cloud Modeling: A Review

Article Type: Review Article

Section/Category: Photovoltaics

Keywords: Photovoltaic; Solar Power; Forecasting; Very Short Term; Nowcasting

Corresponding Author: Mr. Florian Barbieri,

Corresponding Author's Institution: Curtin University

First Author: Florian Barbieri

Order of Authors: Florian Barbieri; Arindam Ghosh, PhD; Sumedha Rajakaruna, PhD

**Abstract:** This paper endeavors to provide the reader with an overview of the various tools needed to forecast photovoltaic (PV) power within a very short-term horizon. The study focuses on the specific application of a large scale grid-connected PV farm. As a matter of fact, the solar resource is largely underexploited worldwide whereas it exceeds by far humans' energy needs. In the current context of global warming, PV energy could potentially play a major role to substitute fossil fuels within the main grid in the future. Indeed, the number of utility-scale PV farms is currently fast increasing globally, with planned capacities in excess of several hundred megawatts. This makes the cost of PV-generated electricity quickly plummet and reach parity with non-renewable resources. However, like many other renewable energy sources, PV power depends highly on weather conditions. This particularity makes PV energy difficult to dispatch unless a properly sized and controlled energy storage system (ESU) is used. An accurate power forecasting method is then required to ensure power continuity but also to manage the ramp rates of the overall power system. In order to perform these actions, the forecasting timeframe also called horizon must be first defined according to the grid operation that is considered. This leads to define both spatial and temporal resolutions. As a second step, an adequate source of input data must be selected. As a third step, the input data must be processed with statistical methods. Finally, the processed data are fed to a precise PV model. It is found that forecasting the irradiance and the cell temperature are the best approaches to forecast precisely swift PV power fluctuations due to the cloud cover. A combination of several sources of input data like satellite and land-based sky imaging also lead to the best results for very-short term forecasting.

Suggested Reviewers: Kashem Muttaqi PhD  
Associate Professor, University of Wollongong  
kashem@uow.edu.au

Dr Muttaqi has worked on significant projects related to renewable energy integration like "Integration of Solar, Wind and Storage Systems into Distribution Grids for Network Support"

John Boland PhD  
Professor of Environmental Mathematics, University of South Australia  
John.Boland@unisa.edu.au

Professor Boland is a specialist in the field of solar irradiance forecasting. He also co-wrote a review paper for this very journal in 2013.

Ahmad Zahedi PhD

Associate Professor, James Cook University

ahmad.zahedi@jcu.edu.au

Prof Zahedi has an extended experience on Power Engineering and Renewable Energy projects

Opposed Reviewers:

# Very Short-Term Photovoltaic Power Forecasting with Cloud Modeling: A Review

Florian Barbieri<sup>a,\*</sup>, Sumedha Rajakaruna<sup>a</sup>, Arindam Ghosh<sup>a</sup>

<sup>a</sup>Curtin University, Department of Electrical and Computer Engineering, Kent Street, Bentley WA 6102, Australia

## ARTICLE INFO

### Article history:

Received 00 December 00

Received in revised form 00 January 00

Accepted 00 February 00

### Keywords:

Photovoltaic; Solar Power; Forecasting;

Very Short Term; Nowcasting

## ABSTRACT

This paper endeavors to provide the reader with an overview of the various tools needed to forecast photovoltaic (PV) power within a very short-term horizon. The study focuses on the specific application of a large scale grid-connected PV farm. Solar resource is largely underexploited worldwide whereas it exceeds by far humans' energy needs. In the current context of global warming, PV energy could potentially play a major role to substitute fossil fuels within the main grid in the future. Indeed, the number of utility-scale PV farms is currently fast increasing globally, with planned capacities in excess of several hundred megawatts. This makes the cost of PV-generated electricity quickly plummet and reach parity with non-renewable resources. However, like many other renewable energy sources, PV power depends highly on weather conditions. This particularity makes PV energy difficult to dispatch unless a properly sized and controlled energy storage system (ESU) is used. An accurate power forecasting method is then required to ensure power continuity but also to manage the ramp rates of the overall power system. In order to perform these actions, the forecasting timeframe also called horizon must be first defined according to the grid operation that is considered. This leads to define both spatial and temporal resolutions. As a second step, an adequate source of input data must be selected. As a third step, the input data must be processed with statistical methods. Finally, the processed data are fed to a precise PV model. It is found that forecasting the irradiance and the cell temperature are the best approaches to forecast precisely swift PV power fluctuations due to the cloud cover. A combination of several sources of input data like satellite and land-based sky imaging also lead to the best results for very-short term forecasting.

© 2015 xxxxxxxx. Hosting by Elsevier B.V. All rights reserved.

## 1. Introduction

Solar power is globally underexploited whereas the sun can be seen as a giant natural fusion nuclear reactor [1] that provides the Earth with far more energy than the human kind needs [2][3] or will most probably ever need in the future [4]. Moreover, solar energy is the largest renewable energy resource available on our planet as well as the source of other resources like wind energy [5]. It is even reported that as the first energy input of the planet, it exceeds the second largest energy resource by a factor 104 [6]. As a gigantic power plant, the Sun basically does not need any maintenance nor any fuel supply. Additionally, the waste resulting from its activity does not need to be processed and does not pose any environmental threat. Also, the Sun will continue to send a steady quantity of solar energy to the surface of the Earth for a period of time far beyond the scale of human history. Yet, there is an urgent need for humanity to develop as soon as possible sustainable gas emission free power resources, for both economic and environmental reasons. Indeed, fossil fuels are becoming increasingly rare and expensive. Moreover, their

extensive use over the two past centuries has induced a significant global climate change [7].

PV generation is advantageous and valuable in various markets, as its peak generation matches with the timeframe of higher load demand. M. Morjaria *et al.* [8] mention that the cost of PV generation has significantly plummeted, increasing greatly the competitiveness of PV power. These authors have even seen that PV-generated electricity has reached grid parity in a large variety of markets, with a price equal or lower with respect to electromechanical non-renewable electricity. As a result, the number and the size of utility-scale PV farms have skyrocketed, as mentioned by Manz *et al.* [9]. The latter authors have even reported that some solar power plants are "planned to exceed several hundred megawatts of capacity."

Nonetheless, as it may be perceived as an inconsistent resource, PV power raises a grid integration concern [10] in particular due to the difficulty to dispatch that energy. With time, this concern is getting more and more serious as the globally installed PV generation capacity is fast increasing, reaching an estimated value close to 100 GW [11].

\* Corresponding author. Curtin University, Perth WA, Australia. Tel.: +61-892663134  
E-mail address: florian.barbieri@curtin.edu.au, florian.barbieri@gmail.com



Moreover, in terms of dynamics, PV plants' response to grid system disturbances conceptually differs from more traditional synchronous machines. It is however possible to improve ramp-rate controls along with the ability to dispatch PV power through sophisticated control algorithms. An adequate control of PV power can also provide the grid with a stable, reliable and effective operation. Additionally, to manage the variable generation of PV resources, grid operators need short-term forecasting methods as well as some support from alternative compensatory controllable resources. Indeed, the rapid progress and increasing accuracy of short-term solar generation forecast guarantees continued efficient and reliable system operations.

Yet, it is interesting to note that large capacity utility-scale grid-connected PV farms account for 38% of the overall PV generation [12]. PV farms can also use tracking devices to optimize the harvest of solar energy. Thus, such energy resources can be seen as a promising option to achieve an optimal penetration of PV power into the grid. This is why this study focuses on large scale grid connected PV farms. This review also tries to give an overview of all the tools needed to forecast the power of this type of power plant.

Although forecasting PV power may seem straight-forward at first glance, it must be noted that the dependence on weather conditions is an important obstacle to tackle. Indeed, since the emergence of PV power generation, the abundant and free solar energy has been difficult to harvest and to dispatch due to its entire dependency on uncertain and intermittent solar radiance. The stochastic discontinuities of sunlight intensity during the day create PV power fluctuations. Additionally, sunlight is not always available to harvest, as it is the case for instance during nighttime. Also, changing climate conditions may result in changing PV power output. As a result, the uncertainty of solar power makes it difficult to integrate into the grid at a high penetration level. This can be an important issue if solar generation is to become a major source of energy in the future [9]. While storing solar energy has long been seen as a solution to this problem, a precise forecast of the available energy is necessary for the previously mentioned control.

For the sake of simplicity, the solar radiance in the complete absence of clouds in the atmosphere can be modelled analytically. Indeed, the rotation of the Earth causes a daily disruption of the flux of extra-terrestrial energy received by the top of the atmosphere. Additionally, the tilt of the spin axis of our planet with respect to the Earth's plane of orbit around the Sun makes the global horizontal irradiance, or the rate of total incoming solar energy received on a unitary horizontal area will be higher in summer than in winter [13]. A unitary horizontal area is defined as measuring 1 square kilometer located on a horizontal plane at the Earth's surface. This effect is more dramatic for regions away from the equator, as illustrated in Figure 1. Additionally, the incidence angle of the sunlight when it goes through the atmosphere also plays in important role.

#### Nomenclature

G	solar irradiance ( $\text{W}/\text{m}^2$ )
t	time (s)
RH	relative humidity (%)
Lat	latitude (degrees)
$W_{\text{speed}}$	wind speed (m/s)
$W_{\text{direction}}$	wind direction (degrees)
$T_{\text{amb}}$	ambient air temperature (degrees Celsius)

Besides those obstacles to a consistent harvest, the light intensity emitted by the Sun varies on a regular basis, through magnetic activity cycles. On the other hand, it is not possible to model analytically the intermittent variations or disruptions of sunlight due to the humidity suspended in the atmosphere. Yet, the motion of clouds a dramatic impact on the performance of a solar generator [14] and thus need to be forecasted to avoid undesired issues and costs.

Due to changes in the cloud cover, it must be noted that some significant drops in PV generated power output can occur, in the range of up to 70% in a matter of 5 seconds [12, 15]. This may result in episodes of grid instability if not properly compensated. The cost of such weather-related power outages is far from negligible. As an example, in the comparable case of wind generators, errors in power prediction may cost up to 10% of the income of selling the generated energy [16].

Power forecasting is a smart option to tackle these problems. This approach can help to manage the ramp rate and running time of compensatory systems to ensure both power continuity and low operational costs. Diagne, Boland et al. (2013) [17] have argued in that sense, stating that an efficient integration of the variable energy output of PV systems and a better quality service both require an accurate and reliable forecast of irradiance. Indeed, reliable forecast information on the solar resource is needed to adjust the power to the expected load profile. In other words, PV forecasting may help to better manage the electrical of the electrical balance between power generation and consumption. However, PV power forecasting is still relatively recent. Such a forecast requires to take many weather parameters into consideration. In comparison, simpler wind power forecasting has been already better studied during the past years and is at a more mature stage of research. As a result, PV power forecasting could inspire from the methodologies already in use in wind power forecasting. This idea is supported by Manz et al., who state in [9] that "the integration of large-scale PV plants in the transmission system can follow the successful model already established by wind integration, with the consequential impact of variability treated in the same manner. Consequently, the second section of this paper examines the methodology used in wind power forecasting and see how useful it can be to PV power forecasting. The purpose of section 3 is to supply the reader with some information about the statistical methods that may be of use for PV power forecasting. Section 4 compares a few popular PV power forecasting methods. Section 5 presents the strategy that may be applied to model a PV Cell. This section starts with the model of the cell/module temperature, which is a critical step for the overall PV model. Section 6 put a stress on irradiance forecasting, considering a sky free of cloud as a first step and by taking clouds into account as a second step. The cloud coverage of the sky and the cloud classification are discussed in section 7, along with their related image processing methods. In section 8, mid-term PV output forecasting based on weather classification is studied while section 9 finally starts the actual topic of very short term PV power forecasting, considering and comparing various approaches.

## 2. Methods used in the reference field of wind power forecasting that can potentially be useful for PV power forecasting

Solar power forecasting is relatively new and far behind wind power forecasting. Wind forecasting can consequently be taken as good reference, as it has been further developed over the past years.

Comprehensive reviews conducted by S.S. Soman et al. (2010) [16] and by S.M. Lawan et al. (2014) [18] both give a good overview of the different horizons that are taken into account in wind forecasting. Soman et al. classified wind forecasting techniques into different categories according to their temporal resolution as well as according to their approach. Some studies like the one conducted by Espino and Hernandez (2011) [19] have compared several approaches within various time scales in order to define which method fitted the best for each horizon. The paper focuses on short (from 10 to 240 minutes) and very-short term wind forecasting using a statistical approach based on Support Vector Regression (SVR).

A quick overview of the different time horizons for wind speed and power forecasting is given in Table 1 [16, 18]. These categories are only a popular set and not the sole classification used in literature.

As illustrated in Table 1, the very short-term horizon can also be called nowcasting and ranges from a few seconds to 30 minutes ahead. It is the shortest timeframe considered in prediction, useful for immediate actions like electricity market clearing and pricing. Only little literature is available for the very-short term timeframe considering both wind and PV power forecasting.

The actual short-term forecasting horizon is comprised between 30 minutes and 6 hours ahead. This is the timeframe used for economic load dispatch as 30 minutes is the usual timeframe for grid operations and operating reserve. Most of the research related to wind forecasting has been done in this time scale.

The medium-term horizon corresponds to a time scale ranging between 6 hours and 1 day whereas the long-term horizon exceeds 1 day ahead.

These previously defined limits of forecasting terms are however not strict and may vary depending on the application of the prediction model. For example, in the application field of meteorology, the US National Weather Service considers 0 to 3 hours ahead forecasting as nowcasting. For some other agencies, forecasts up to six hours can also be called nowcasting. In particular, the National Oceanic and Atmospheric Administration (NOAA) provides the following definition in the glossary of their website [20]: “Nowcast: A short-term weather forecast, generally out to six hours or less [...] also called a Short Term Forecast”, “Short Term Forecast: A product used to convey information regarding weather or hydrologic events in the next few hours”, “Medium Range: In forecasting, (generally) three to seven days in advance”.

Besides various prediction time frames, several wind speed and power prediction models have already been used in scientific literature. The most popular ones are listed in table 2 [16].

When using the persistence model, also known as ‘naïve predictor’, it is assumed that the physical quantity that is forecasted, which is wind speed and direction, as it happens, remains unchanged during a time increment  $\Delta t$ . This means that for example the wind vector  $\vec{v}$  is the same at time  $t+\Delta t$  as it was at time  $t$ . This approach thus implies a very strong correlation between present and future value. This prediction method has been reported to be the most accurate for very short-term and short-term forecasts. This approach is even commonly used by meteorologists as a reference benchmark tool to assess the performance of another wind forecasting method.

The persistence model can advantageously supplement the physical model, which is mainly elaborated using Numerical Weather Prediction (NWP). This approach consists in taking into consideration a detailed physical description of the atmosphere. Indeed, the future wind speed and direction both strongly depend on other current meteorological quantities such as ambient temperature, solar irradiance, atmospheric pressure,

relative humidity, and dew point. The topology of the terrain is also considered when using a NWP for wind forecasting purposes. Due to the difficulty to gain live data and due to high costs, NWPs are run only a few times per day on powerful hardware resources, using complex mathematical models commonly based on kinematic physical equations. Consequently, this model is not appropriate for relatively short-term forecasting and is only accurate for medium and long-term. This method is thus suited for horizons exceeding 6 hours. It performs best with time scales comprised between 48 hours and 172 hours ahead.

Statistical approaches consist in training a model with measurement data and in comparing the predicted values with the actual values in immediate past to fine-tune the parameters. These models are accurate for relatively short-term horizons. The two main subclasses are respectively Time-series based approaches and Artificial Neural Network (ANN) based methods. Other statistical methods like fuzzy logic and genetic algorithm can also be used.

ANN models are powerful nonlinear data driven methods that are easy to model, inexpensive and less time consuming than other methods. Rather than being based on any predefined mathematical model, it is based on patterns. However, the accuracy of such prediction models quickly drops as the time horizon is extended. Several variants of Neural Networks are used like feed-forward neural networks (FNNs), multi-layer perceptrons (MLP), recurrent neural networks (RNNs) and time delay neural networks (TDNN) to name a few. More details are given in the next section.

Furthermore, some additional artificial intelligence (AI) techniques can also be successfully applied for short-term predictions, like for example genetic algorithms (GA), fuzzy logic, Markov chains (MC), etc.

Time series based forecasting methods have been very popular for wind prediction in the past years. The most commonly used model is the Auto-Regressive Moving Average (ARMA), which is an improvement of the regular predictive auto-regressive method (AR). However, this method can only be used for stationary time series. Several variations exist like Auto-Regressive Integrated Moving Average (ARIMA) and Auto-Regressive Moving Average with exogenous inputs (ARMAX). The coupled autoregressive and dynamical system (CARDS) is also a popular method. Other models like Bayesian Model Averaging (BMA) or Support Vector Regression (SVR) [19] are also applied.

Model performance is often assessed by the means of Root Mean Square Error (or deviation) (RMSE), normalized (or standardized or even relative) Root Mean Square Error (nRMSE or rRMSE), Mean Absolute Error (MBA) or Mean Bias Error (MBE) or other means.

The root mean square error (RMSE) and normalized root mean square error (nRMSE) are commonly used to evaluate the performance of an approach. If we consider the number  $N$  of samples, both errors are respectively defined by equations 1 and 2 [21]:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{forecast} - \text{measurement})^2} \quad (1)$$

$$nRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\text{forecast} - \text{measurement})^2}}{\frac{1}{N} \sum_{i=1}^N \text{measurement}} \quad (2)$$

The nRMSE equals the RMSE divided by the range of the variable that is measured. The nRMSE is usually expressed as a percentage.

The mean bias error (MBE) and mean absolute error (MAE) are other statistical quantities used to measure how close forecasts or predictions are to the eventual outcomes.

The MBE is an average of the algebraic errors while the MAE is an average of the absolute errors. Both are respectively given by equations (3) and (4):

$$MBE = \frac{1}{N} \sum_{i=1}^N (\text{forecast} - \text{measurement}) \quad (3)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\text{forecast} - \text{measurement}| \quad (4)$$

Additionally, it is also common practice to combine different approaches, like mixing physical and statistical approaches for example, or to combine different horizons, such as short-term and medium-term for instance. Such combinations are referred to as hybrid methods and aim at optimizing the short-term prediction values. Indeed, many researchers use a combination of methods and all indicate that hybrid methods outperform the individual methods [17]. Amongst the many possible combinations, mixing together NNs and fuzzy logic leads to the popular approach of Adapted Neuron Fuzzy Inference (ANFIS). Another popular technique consists in correlating the wind speeds at different sites.

All the discussed time scales and methods are also valid for PV power forecasting. Indeed, wind speed and direction is one factor amongst other meteorological quantities that greatly influence the prediction of PV power. As a matter of fact, a fast frontal wind may cool down the cell temperature within a PV array, which in turn increases the voltage output and by extension also the power output.

*Summary:* Although the following considerations are valid for wind power forecasting, they can also be used as a guideline for solar power forecasting, which is more complex since more variables are taken into account. As a first step, the forecasting horizon is defined according to the grid operation. Secondly, an approach is chosen on the basis of the horizon. Wind power forecasting studies show that the persistence approach is the best for very short-term horizons. On the other hand, time-series and ANN-based approaches are only reliable for short term horizons. NWP, for their part, prove to be only accurate for medium to long horizons.

### 3. Complementary information on statistical methods for PV Forecasting

Tools like neural networks (NNs) are very effective to deal with non-linear systems. Specific types of NNs are commonly used [2]:

- Expert Systems (ES)
- Artificial Neural Network (ANN)
- Genetic Algorithm (GA)
- Fuzzy Logic (FL)
- Computer Vision (CV)
- Hybrid Systems (HS)
- Many other methods

ESs interpret data to deal with knowledge processing and make decisions

for complex problems.

Genetic Algorithms utilize a fixed-sized population containing individual possible solutions to a given problem, which evolve in time. Genetic operators such as selection, crossover and mutation are applied to eliminate the poorest solutions and create new solutions from selected existing ones. These algorithms are very popular in machine learning.

ANNs are based on the emulation of a biological brain. It has the capacity to be trained and to learn. The architecture consists in multiple layers of interconnected artificial neurons. The neurons communicate through synapses. Each artificial neuron (AN) compares their inputs with a threshold value to produce an output. The weight of each input can be adjusted during a preliminary learning phase [22]. The ANNs support both supervised and unsupervised learning.

A multi-layer perceptron (MLP) is an ANN consisting in several fully interconnected layers of ANs. It is trained with a supervised learning technique. As a matter of fact, MLPs are the most popular form of ANNs. They are compound of an input layer, at least one hidden layer and an output layer. The hidden layers often use a hyperbolic tangent function of the input vector  $x$ , as shown in equation (5) [17].

$$f(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \quad (5)$$

Considering a MLP that processes a number  $n$  of input vectors  $x$ , is characterized with  $h$  hidden neurons and a single linear output  $y$ , the relationship between the input and the output is the following [17]:

$$y = \sum_{j=0}^h \left[ W_j f \left( \sum_{i=0}^n W_{ji} X_i \right) \right] \quad (6)$$

When used for one day-ahead forecasting, MLPs may lead to very accurate results. For example, Mellit and Pavan (2010) [23] have successfully used a MLP to forecast the solar radiance 24 hours ahead with a 98% precision during sunny days and less than 95% during cloudy days.

Fuzzy Logic is based on the utilization of logical statements and binary operators to model problem treated with a human-like reasoning. It is mainly used in control engineering.

These artificial intelligence methods have been successfully applied for the purpose of PV power output or irradiance forecasting [23]. According to Mellit et al. (2010), MLP and Radial Basis Functions (RBF) networks as well as fuzzy logic are particularly suited in the case of precise forecasts based on a significant number of meteorological and geographical input data. In other words, such methods are effective when the forecasted irradiance is expressed as follows:

$$G(t+1) = f(T_{amb}(t), RH(t), TCC(t), W_{speed}(t), W_{direction}(t), P(t), Lat, Long, G(t), \dots) \quad (7)$$

In equation (7),  $T_{amb}$  is the ambient temperature,  $RH$  is the relative humidity,  $TCC$  is the total cloud cover,  $W_{speed}$  and  $W_{direction}$  are respectively the wind speed and direction,  $P$  is the atmospheric pressure,  $Lat$  and  $Long$  are respectively the latitude and longitude.  $G$  is the irradiance.

On the other hand, recurrent Neural Networks (RNN), wavelet and wavelet-fuzzy networks are more suitable for forecasts based on past observed data only.

In mathematical terms, the forecast of the irradiance, for instance, is then expressed as:

$$G(t+1) = f(G(t), G(t-1), G(t-2), \dots) \quad (8)$$

Alternatively, ANFIS methods are suitable when the forecasted irradiance is modelled as a combination of the two previous models:

$$G(t+1) = f(G(t), G(t-1), G(t-2), \dots, T_{\text{amb}}(t), RH(t), TCC(t), w_{\text{speed}}(t), w_{\text{direction}}(t), P(t), \text{Lat}, \text{Long}, G(t)) \quad (9)$$

Computer vision deals with image processing and motion analysis. Many systems of this type have been developed for a large number of applications. However, the strategy employed depends mainly on the system that is analyzed.

Other methods may include Markov Random Filters (MRF), Probability Hypothesis Density filters (PHD), k-nearest neighbor (kNN) or State Vector Machine (SVM).

The kNN approach is used in particular by Heinle et al. [24] and Kazantzidis et al. [25] to classify clouds into cloud classes according to some of their features. This operation is performed by majority vote. However, kNNs can be slow to run and may require a large amount of memory.

A SVM is a nonlinear model that demonstrates faster computing features than an ANN and good convergence [26]. It requires nonetheless a large amount of data for regression.

For the purpose of solar power forecasting, Diagne, Boland et al. (2013) recommend the use of statistical methods for short forecasting horizons (intra-hour or intra-day), hence for high temporal resolutions (less than 10 hours) and high spatial resolutions (up to 500 meters). Even though hybrid models are more robust, single methods can present a particular strength within for a specific resolution. Indeed, ANNs may be recommended for temporal resolutions between 30 minutes and 10 hours as long as the targeted spatial resolution is between 10 and 500 metres. The persistence method is the best choice for very fine resolutions below 10 minutes and 10 meters. On the other hand, ARIMA, ARMA and CARDS are suitable for resolutions between the previous cases (from a few seconds to 1 hour and up to 12 meters).

*Summary:* A large number of statistical methods are available. ANNs can be accurate, especially when they are combined with other artificial intelligence methods like genetic algorithms. However, they show slow computational features. On the other hand, SVM-based method can be faster but require more data.

#### 4. Comparison of some popular PV power forecasting approaches

Horizons for PV power forecasting may be defined differently with respect to wind forecasting. Additionally, the choice of approaches may differ.

As an example, two different approaches have been confronted to each other in a study conducted by Yuehui et al. [27] by considering the case study of a 1 MW PV farm comprising 6006 panels. The two methods are respectively statistical and physical.

Firstly, this quick review of PV power forecasting methods defines the various time scales used in power system operation with different limits, compared to wind forecasting. The new definitions of the various horizons are summarized in Table 3 [27].

The very short-term horizon is defined as being in the order of several minutes to several hours, aiming for intraday real-time control. The short-term horizon ranges from a few hours to up to 3 days and is used for day-ahead economic dispatch. As an illustration of short-term horizon, Diagne, Boland et al. [17] mention that load patterns need to be forecasted 2 days ahead for scheduling of power plants and for planning transactions in the electricity market. The long-term forecast is defined as ranging from a week to a year and targets grid balance operations.

This classification of PV forecasting horizons is not unique and Kostylev and Pavlovski (2011) also distinguish 3 horizons that are intra-hour, intra-day and day ahead. This alternative discrimination of horizons has been used by Diagne, Boland et al. and is exposed in table 4 [17].

The effect of irradiance intermittency can be attenuated through the use of energy storage, spinning reserve and demand response [15]. All these methods need an accurate forecast on several timescales to be optimally operated. Indeed, day-ahead forecasting is required to determine pricing in the market. In contrast, sub-hour-ahead forecast is useful to schedule spinning reserves.

The comparison focuses on short-term horizon and compares the results of a physical approach with the ones of a statistical method by using the same days over a period of one month during each season of the year in a place located in the northern hemisphere. The physical model that is utilized for PV generation consists in a clear sky model and diode PV model. The statistical method uses a neural network (NN).

The measured output power of one representative month of each season is compared with the clear sky model power prediction, adjusted with NWP data. These measurements have been made in the northern hemisphere. Unsurprisingly, the best match is achieved during clear days of each month whereas the biggest mismatch is obtained during the month of October and December, which have been chosen to represent respectively autumn and winter. Both seasons are characterized by a measured power output much lower than the prediction. Overall, the nRMSE values range between 11.48% and 16.58%, as shown in Table 5 [27]. When the actually measured values of temperature and irradiance are additionally fed to the physical model, the precision increases dramatically and the nRMSE values drop down to 4.93%-8.78%. This makes the biggest discrepancy occur in spring for the cloudy month of March. This enhanced precision thus validates the physical model, provided it is fed with sufficient weather data.

The central column corresponds to a physical model using the measured irradiance  $G$  and temperature  $T$  as inputs. The nRMSE of the predictions from the best trained NN model are showed in the rightmost column of Table 5. The optimal NN is set to have 11 neurons and is fed with several weather data such as the ambient temperature, the cloud coverage, the irradiance and the position of the sun. Compared to the NWP physical model predictions, the results appear to be significantly better for the wintry month but only slightly more accurate for all other seasons.

Overall, there is only little improvement with respect to the NWP physical model. The study concludes that both approaches are valid for the considered horizon and that the precision mainly depends on the data input itself.

Another study, carried out by Guarnieri et al. (2008) [28] has demonstrated that the ANNs using training data can reduce the nRMSE of daily average global horizontal irradiance by 15% compared to a 12-18 hours ahead irradiance forecast based on a NWP approach as pointed by Diagne et al. in their review paper [17].

*Summary:* For solar power forecasting purposes, very short-term may be defined on a longer time frame than for wind power forecasting. Indeed, this horizon may be in range of several minutes rather than in range of several seconds. However, a forecasting horizon beyond very short-term is required for power dispatch and grid balance. For such horizons, the input data proves to be actually more important than the chosen method. Nonetheless, neural network based forecasting methods are more accurate than NWP-based approaches for periods of the year characterized by heavy rain and snow falls like winter.

## 5. Strategy to model a PV cell

Solar irradiance is the main driving parameter that influences PV power output [29]. Although high-efficiency multi-junction PV cells are under development, the large majority of the current PV installations use mono-junction modules. Also, most of the current studies concentrate on mono-junction modules. Consequently, this study only focuses on mono-junction modules.

With mono-junction modules, the operating temperature  $T_c$  greatly affects the energy efficiency of the conversion of light into electricity. In particular, the cell temperature  $T_c$  affects a great deal the value of the open-circuit voltage as shown in Figure 2 (a), with the example of a mono-junction polycrystalline module [30]. For its part, the short-circuit current is significantly affected by the solar irradiance, as illustrated in Figure 2 (b). The effects of both the cell temperature and the solar irradiance are obviously echoed on the output power curve [31].

Thus, it is important to know the operating temperature of photovoltaic modules to improve their energy performance. Amongst several empirical formulas, Evans' expression of efficiency summarizes the effect of both cell temperature and solar irradiance [22, 32]:

$$\eta = \eta_{ref} \cdot \left[ 1 - \beta \cdot (T_c - 25^\circ) + \gamma \cdot \log_{10} \left( \frac{G}{G_{ref}} \right) \right] \quad (10)$$

In equation (10),  $\beta = 4.5 \cdot 10^{-3} \text{ } ^\circ\text{C}^{-1}$  and  $\gamma = 0.12$  for crystalline silicon modules.

The cell temperature can either be modelled or alternatively assessed through the utilization of adaptive techniques such as artificial neural networks (ANNs). An adaptive technique can actually be used in an advantageous tool for a fast and precise forecast of the cell temperature and to guarantee a high performance.

Only 10 to 20% of sunlight is converted into a current by mono-junction modules whereas the rest is converted into heat that is partially transferred to the surrounding environment according to a specific coefficient. Consequently, the temperature of a module automatically rises as soon as it absorbs some irradiance, which in turn affects the operating point and

leads to a diminished electrical conversion of light [22]. Yet, a comprehensive analytical model would be excessively complex to use for calculations.

This is why the various temperature cell models found in literature assume some simplifications and do not consider accurately the full complexity of the irradiance absorption effect. However, an artificial neural network can help to deal with this complexity in a timely efficient manner. In addition, it must be noted that the current-voltage curves provided by manufacturers in their datasheets cannot be used to reliably assess the thermos-electrical behavior of a PV module as they are given for constant temperature or constant irradiance. Indeed, such datasheet curves do not consider the actual conditions in which the cell temperature changes along with the irradiance.

### 5.1 Modeling the cell/module temperature

Various empirical models are available in literature [22, 33]. They are listed in table 6 [14, 22, 33-46]. In this table, the instantaneous solar insolation is referred as  $G$  (kW/m<sup>2</sup>) whereas  $T_{amb}$  is the ambient temperature given in  $^\circ\text{C}$ . The NOCT conditions are defined as  $G_{NOCT}=800 \text{ W/m}^2$ ,  $w_{speed}=1 \text{ m/s}$  and  $T_{c, NOCT}=20^\circ\text{C}$ .

In any case, the two main parameters always remain the irradiance and the ambient air temperature. Amongst popular models, the Ross and Smokler formula from 1986 is widely used [47]:

$$T_c = T_{amb} + \frac{G}{G_{NOCT}} \cdot (T_{c, NOCT} - T_{NOCT}) \quad (11)$$

It must be noted that the Ross and Smokler model is only valid for a free standing module.

Another more precise example that considers 3 weather parameters is Chenni et al.'s model. The coefficients are average values of measurements performed on 6 different PV technologies (amorphous silicon, monocrystalline silicon, copper indium diselenide, EFG-polycrystalline silicon, polycrystalline silicon, and cadmium telluride) located on two different geographical sites [48]:

$$T_c = 0.943 \cdot T_{amb} + 0.028 \cdot G - 1.528 \cdot W_{speed} + 4.3 \quad (12)$$

Nevertheless, this still remains an imperfect model as it does not take heat losses into account.

The approach of assessing and predicting the cell/module temperature is particularly interesting for hybrid PV/thermal systems containing a fluid cooling the PV modules. Controlling the temperature of the PV modules help to increase the voltage output. Additionally, the heat that is extracted from the module can be in turn exploited to improve the efficiency of the complete system. As mentioned by Ciulla, Lo Brano, and Moreci [22], best results are achieved through the employment of a multilayer perceptron network.

In their publication, the MLP is composed of 2 non-linear threshold function blocks, 2 weight layers, 1 hidden layer, 1 error criterion block and 2 input sources that consider the following parameters:

- ambient temperature  $T_{amb}$
- wind speed  $W_{speed}$
- irradiance  $G$
- electrical power output of the module  $P_{elec} = I \cdot V$  (13)
- short circuit current of the module  $I_{sc}$
- open circuit voltage of the module  $V_{oc}$

The weather data comes from a weather station and the actual temperature of the panels of the experimental setup is measured with thermocouples. The electrical data are measured as well.

As a result, the confidence band of the evaluation is narrower than  $\pm 1^\circ\text{C}$ . For different cell temperature correlations, the mean absolute error (MAE) obtained with the ANN is around 0.1–0.2, according to the type of PV panel. This is much below the MAE of the analytical empirical correlations like Servant, Duffie-Beckman and Hove. These models display an MAE between 4 and 16, making the ANN superiorly accurate. Additionally, the authors suggest that the very short learning time enables to implement the ANN on a real-time system to evaluate the cell temperature in run-time conditions.

## 5.2 Further considerations of the PV model

Knowing the operating temperature of the cell helps to determine the maximum solar power per  $\text{m}^2$  as expressed in equation (14).

$$P_m = P_p \cdot \eta_t \cdot \frac{G}{G_{STC}} \quad (14)$$

Where  $P_p$  is the nominal power (W) given by the manufacturer at standard test conditions (STC) solar irradiance  $1000\text{W}/\text{m}^2$ , air mass of 1.5 and cell temperature of  $25^\circ\text{C}$ ,  $G_{STC}$  is the reference irradiance ( $1000\text{W}/\text{m}^2$ ) and the normalized efficiency  $\eta_t$  is defined by equation (15).

$$\eta_t = 1 + \gamma_{mp} \cdot (T_c - T_{STC}) \quad (15)$$

In equation 12,  $\gamma_{mp}$  is the maximum power thermal coefficient, typically about  $-0.5\%/^\circ\text{C}$  for crystalline modules. The  $\eta = \eta_0 \cdot \eta_t$  efficiency rate is where  $\eta_0$  is the nominal conversion rate in STC, which depends on the PV technology. The average nominal efficiency of the three main categories of mono-junction PV technologies is summarized in table 7 [49]. Monocrystalline PV have the highest nominal conversion rate whereas thin film amorphous silicon systems have the lowest.

The maximum power of an array characterized by an area  $A$  is given by equation (16) [50].

$$P_{m,A} = P_m \cdot A \quad (16)$$

The temperature affects current and voltage considered individually [5]:

$$V = V_{NOCT} \cdot [1 + \beta \cdot (T_{amb} - T_{NOCT})] \cdot \left[ 1 + \delta \cdot \ln\left(\frac{G}{G_{NOCT}}\right) \right] \quad (17)$$

$$I = I_{NOCT} \cdot \frac{G}{G_{NOCT}} \cdot [1 + \alpha \cdot (T_{c,NOCT} - T_{NOCT})] \quad (18)$$

In equation (17) and (18), the coefficients  $\alpha$ ,  $\beta$  and  $\delta$  must be determined from an initial condition. The two-diode model is the most precise for a

photovoltaic cell as shown in Figure 3 (a) [51].

In Figure 3 (a):

$$V = V_d - R_s \cdot I \quad (19)$$

$$I = I_{sc} - I_0 \cdot \left\{ e^{\frac{q \cdot V \alpha}{k \cdot T}} - 1 \right\} - \frac{V_d}{R_p} \quad (20)$$

It should be noted that at  $\frac{q \cdot V \alpha}{k \cdot T} = 38.9$ .  $25^\circ\text{C}$ , Additionally, the series resistance  $R_s$  is typically significantly smaller than the parallel resistance  $R_p$ .

When  $n$  cells are arranged as a module with all cells equally insulated, the voltage of the module becomes:

$$V_{mod\ ulte,\ insolated} = n \cdot (V_d - R_s \cdot I) \quad (21)$$

However, if only one of the cells gets shaded, the parallel resistance of the shaded cell will cause a dramatic voltage drop:

$$\Delta V = I \cdot (R_p - R_s) \cong R_s \cdot I \quad (22)$$

This effect is illustrated in Figure 3 (b) [51]. In that case, the output voltage then becomes:

$$V_{mod\ ulte,\ 1\_shaded\_cell} = \left( \frac{n-1}{n} \right) \cdot V_{mod\ ulte,\ insolated} - \Delta V \quad (23)$$

The voltage drop compared with a fully insulated module is

$$V_{mod\ ulte,\ 1\_shaded\_cell} - V_{mod\ ulte} = \frac{V_{mod\ ulte,\ insolated}}{n} - \Delta V \quad (24)$$

Using shunt diodes to compensate this effect is a common practice. It is also possible to work on the design to attenuate the effects of partial shading.

*Summary:* Ross and Smokler's expression of the cell temperature is very popular and widely used by manufacturers in their datasheets. However, Chenni et al.'s formula is the most precise to date although it can still be improved as it does not take the coefficient of heat dissipation into consideration. It must also be noted that the 2-diode model is the most precise to date for a mono-junction PV cell. From the point of view of the nominal conversion rate, monocrystalline PV cells and modules are the most productive, most expensive and also most sensitive to temperature gradient amongst mono-junction modules. Thin film modules are on the contrary the least efficient but cheapest and most robust ones.

## 6. Irradiance Forecasting

Diagne, Boland et al. have stated that forecasting global horizontal irradiance GHI is the first and necessary step for most of PV power prediction models. [17]. These authors also stress the fact that an accurate and efficient irradiance forecast can help grid operators to maintain grid balance. Two steps are needed for a comprehensive irradiance forecast. First, a clear sky irradiance needs to be modeled based on astronomical facts. The actual irradiance can be forecast in a second time by considering the effects of clouds on the clear sky solar radiance.

### 6.1 Clear Sky Model

The stochastic variations of global horizontal irradiance cannot be dealt with without a proper model of the clear sky solar radiation. Astronomical considerations are necessary to implement this model [4]. Such considerations relate mainly to cyclic variations. An important parameter to consider first when building a clear sky model is the calendar day number of the year, referred to as  $n$ . The number  $n$  hence represents how many days have passed since the 1<sup>st</sup> of January of the current year.

The Sun-Earth distance varies slightly by a factor of  $\pm 1.7\%$  throughout the year. This is because the trajectory of the Earth is not perfectly circular but slightly elliptical. The distance can be expressed in km in function of the calendar day number with equation (25).

$$d = 1.5 \cdot 10^8 * \left\{ 1 + 0.017 \cdot \sin \left[ \frac{360(n - 93)}{365} \right] \right\} \quad (25)$$

In particular, the equation of time models the elliptical orbit of the Earth can be formulated in minutes with equation (26).

$$E = 9.87 \cdot \sin \left[ 2 \cdot \frac{360}{364} \cdot (n - 81) \right] - 7.53 \cdot \cos \left[ \frac{360}{364} \cdot (n - 81) \right] - 1.5 \cdot \sin \left[ \frac{360}{364} \cdot (n - 81) \right] \quad (26)$$

The equation of time is very useful to calculate the solar time with respect to civil time, as given by equation (27).

$$\text{Solar time} = \text{Civil Time} + 4 \cdot [\text{Local Time Meridian} - \text{Local Meridian}] + E \quad (27)$$

The meridians are expressed in degrees. The local time meridian is defined accordingly to the time zone of the considered location. For example, the time zone for Perth, Western Australia, is UTC+8 and the corresponding local meridian is 120° East. The coefficient 4 min/degree features the rotation speed of the Earth. The solar noon can be deduced from the previous equation if solar time is 12 PM.

The hour angle is defined as  $H = 15^\circ \cdot h$ , where  $h$  is the decimal number of hours before solar noon, so that  $H$  is equal to  $+15^\circ$  at 11:00 A.M. solar time and  $H = -15^\circ$  at 1:00 P.M. solar time.

On a specific day of the year, the declination of the Earth  $\delta$  can be expressed in a simple manner with equation (28).

$$\delta = 23.45 * \sin \left[ \frac{360 \cdot (n - 81)}{365} \right] \quad (28)$$

Another expression of the declination of the Earth  $\delta$  has been formulated by in the form of Fourier series by Spencer in 1971 as shown in equation (29) [52].

$$\delta = 0.006918 - 0.399912 \cdot \cos(T_d) + 0.070257 \cdot \sin(T_d) - 0.006758 \cdot \cos(2 \cdot T_d) + 0.000907 \cdot \sin(2 \cdot T_d) - 0.002697 \cdot \cos(3 \cdot T_d) + 0.001480 \cdot \sin(3 \cdot T_d) \quad (29)$$

$T_d$  is the angular fraction of the year calculated by Spencer as per equation (30).

$$T_d = \frac{2 \cdot \pi \cdot (Jday - 1)}{365} \quad (30)$$

In equation (30),  $Jday$  is the Julian ephemeris day. The Julian calendar starts on the 1<sup>st</sup> of January -4712 at Greenwich mean noon and is

calculated as per equation (31) according to the Gregorian day with decimal time  $t_{dd}$ , month  $t_{mth}$ , and Gregorian year  $t_{yr}$ .

$$Jday = INT \left[ 365.25 \cdot (t_{yr} + 4716) \right] + INT \left[ 30.6001 \cdot (t_{mth} + 1) \right] + t_{dd} + 2 - A + INT \left[ \frac{A}{4} \right] - 1524.5 \quad (31)$$

In equation (31):

$$A = INT \left[ \frac{t_{yr}}{100} \right] \quad (32)$$

The operator INT[] returns the integer of the terms between the brackets. Adjusted from the local time zone  $h_{tz}$  to Greenwich Mean Time, the Julian day is expressed as per equation (33).

$$Jday, adj = Jday - \left[ \frac{h_{tz}}{24} \right] \quad (33)$$

The altitude angle of the sun  $\beta$  are respectively, both expressed in degrees:  $\beta = \sin^{-1} [\cos(Lat) \cdot \cos(\delta) \cdot \cos(H) + \sin(Lat) \cdot \sin(\delta)]$  (34)

In equation (34),  $\delta$  is the declination of the Earth,  $H$  is the hour angle and  $Lat$  is the latitude.

The altitude angle of the sun  $\beta$  might need to be corrected, as mentioned by Annear and Wells in their comparison of clear-sky solar radiation models [52]. An atmospheric refraction term is added, as shown in equation (35).

$$\beta_{corrected} = \beta + CF \quad (35)$$

The value of this correction term depends on the value of the altitude angle  $\beta$  as per table 8 [17].

At a given hour of the day, the solar azimuth angle is defined in degrees by equation (36).

$$\varphi_s = \sin^{-1} \left[ \frac{\cos(\delta) \cdot \sin(H)}{\cos(\beta)} \right] \quad (36)$$

It must be noted that  $|\varphi_s| > 90^\circ$  unless  $\cos(H) \geq \frac{\tan(\delta)}{\tan(Lat)}$  in which case  $|\varphi_s| \leq 90^\circ$ .

At solar noon, the altitude angle of the sun becomes  $\beta_N$ , expressed in degrees ( $^\circ$ ) by equation (37).

$$\beta_N = 90^\circ - Lat + \delta \quad (37)$$

In the case of fixed orientation, the PV collector tilt angle should be set up according to  $\beta_N$ :

$$\text{Tilt} = 90^\circ - \beta_N \quad (38)$$

There are few alternative ways to determine the position of the sun. A very exhaustive report has been published by Sandia National Laboratories [53] to describe clear sky models of various complexities and precisions.

The report mentions that NREL has developed the most accurate Solar Position Algorithm (SPA) to date, which can be used as the current reference. Based on location, date, and time inputs over a period of 8000 years, this algorithm show as little uncertainties as +/- 0.0003 degrees and has been described by Meeus [54] as well as by Reda and Andreas [55]. NREL has provided an online calculator version of this algorithm [56].

The extraterrestrial radiation emitted by the sun follows a few cycles. As per observation ranging from 1699 to 2008, some sunspot activities occur every 11 years in average, to which some rapid irradiance variations with even larger amplitude are superimposed [6]. When a sunspot peak happens, the extra-terrestrial solar insolation, which is the solar flux from outer space entering radially through the atmosphere, on any point at the top of the atmosphere, may become about 1.5% higher [4]. By neglecting the effect of sunspots, the extra-terrestrial solar radiation varies by about 6.8% during a year and can be expressed in  $W/m^2$  by:

$$ET = SC * \left[ 1 + 0.034 \cdot \cos\left(\frac{360 \cdot n}{365}\right) \right] \quad (39)$$

Where  $SC$  is the solar constant and  $n$  is the day number during the year. The value of  $SC$  is actually not constant and varies with time. The value of  $SC$  fluctuates about  $\pm 1 Wm^{-2}$  around a mean value during a typical 11-year cycle [57]. The latest and most accurate value is  $1360.8 \pm 0.5 W/m^2$  as per the latest measurement performed by NASA SORCE (Solar and Climate Experiment) during the 2008 solar minimum period. This new measurement is much lower than the canonical value of  $1365.4 \pm 1.3 W/m^2$  defined in the 1990s [6]. Often  $1366.1 W/m^2$  is used as an average solar constant, as defined by Gueymard in 2004 [57]. During the same year 2004, NASA defined the average of all the minimum and maximum data values collected from the 1970s as  $1367.4 W/m^2$  [52]. This is why the  $1367 W/m^2$  may also be often used in literature.

It must be noted, different clear sky irradiance models of various complexity and precision exist. We will try to review a simple one as well as the most accurate one, which is relatively complex.

It must also be noted that the total horizontal irradiance contains three components that are the horizontal beam, the diffuse and the reflective radiations. The total horizontal radiance  $I_{TH}$  is the radiation reaching the surface of the Earth on a flat horizontal plane. This quantity will be expressed firstly with simple common equations provided by G.M. Master (2004) [4], based on an empirical model developed by Threlkeld and Jordan (1958). The most precise model to date to express  $I_{TH}$  will be explained in a second time.

### 6.1.1. A relatively simple model

- **Direct beam irradiance:**

$$I_B = A \cdot e^{-k \cdot m} \quad (40)$$

$$I_{BH} = I_B \cdot \sin(\beta) \quad (41)$$

Where  $I_B$  is the direct beam radiation, normal to the rays while  $I_{BH}$  is the horizontal beam solar radiance. The horizontal beam irradiance is defined as the solar beam radiation reaching the surface of the Earth on a flat horizontal plane.  $A$  is the apparent extra-terrestrial solar flux,  $k$  is the optical depth of the atmosphere and  $m$  is the air mass ratio [4].

$$A = 1160 + 75 \cdot \sin\left[\frac{360}{365} \cdot (n - 275)\right] \quad (42)$$

$$k = 0.174 + 0.035 \cdot \sin\left[\frac{360}{365} \cdot (n - 100)\right] \quad (43)$$

$$m = \frac{1}{\sin(\beta)} \quad (44)$$

Recently, G.M. Masters has amended the expression of the air mass ratio with a more accurate equation that takes the spherical nature of the atmosphere into account [58]:

$$m = \sqrt{(708 \cdot \sin(\beta) + 1417)^2 + 1417} - 708 \cdot \sin(\beta) \quad (45)$$

- **Diffuse irradiance:**

Assuming the sky to be isotropic, the diffuse radiation on a horizontal plane on the surface of the Earth is proportional to the direct beam radiation:

$$I_{DH} = C \cdot I_B = 0.095 + 0.04 \cdot \sin\left[\frac{360}{365} \cdot (n - 100)\right] \cdot A \cdot e^{-k \cdot m} \quad (46)$$

In equation (46),  $C = 0.095 + 0.04 \cdot \sin\left[\frac{360}{365} \cdot (n - 100)\right]$  is

referred to as the sky diffuse factor.

- **Reflective irradiance:**

The reflective component of the total horizontal irradiance is often neglected. This component depends on the ground reflectivity or albedo  $\rho$  and the orientation of the PV collector. Typically,  $\rho$  ranges between approximately 0.1 for bitumen and gravel to 0.8 for fresh snow. The reflectance of grass is about 0.2. Considering a fixed collector with a tilt angle  $\Sigma$ , the reflective irradiance is defined as by equation (47).

$$I_{RC} = \rho \cdot (I_{BH} + I_{DH}) \cdot \left(\frac{1 + \cos(\Sigma)}{2}\right) \quad (47)$$

Using tracking devices enable to optimize the harvest of solar energy. However, the orientation of the PV collector with respect to the position of the sun must be carefully studied. If the direct beam of the sun strikes the PV collector with an incidence angle  $\theta$  with respect to a normal vector to the collector face, the beam and diffuse irradiance received by the collector are respectively:

$$I_{BC} = I_B \cdot \cos(\theta) \quad (48)$$

$$I_{DC} = I_{DH} \cdot \left(\frac{1 + \cos(\Sigma)}{2}\right) \quad (49)$$

As illustrated in Figure 4 [4], the PV collector may be tipped up with an angle  $\Sigma$  and oriented with an azimuth angle  $\phi_c$  with respect to the south in northern hemisphere or north in the southern Hemisphere.

The angle of incidence  $\theta$  is then expressed as shown in equation (50).

$$\cos(\theta) = \cos(\beta) \cdot \cos(\varphi_s - \varphi_c) \cdot \sin(\Sigma) + \sin(\beta) \cdot \cos(\Sigma) \quad (50)$$

### 6.1.2. Example of a very accurate model

Several models for calculating the position of the sun and the clear sky irradiance are available in literature. These models have been sometimes compared with one another [52]. This has led to the conclusion that the most complex models are also the most precise. In particular, the Meeus (1999) and Bird and Hulstrom (1981) model has been tested as the best option with RMS errors in average 6 W/m<sup>2</sup> lower than the other tested methods for 17 sites in North America over a period of 2,726 clear days. The Meeus and Bird and Hulstrom model takes several calibration parameters into account. The calibration parameters in question are the ratio of forward-scattered irradiance to the total scattered irradiance, the aerosol absorption, and the atmospheric turbidity.

Wunderlich (1972) formulated an even more precise first order approximation of the optical air mass ratio  $m$  than equation (44) on the basis of an expression found by Kasten (1964) modified in the light of the variations of atmospheric pressure according to the altitude  $z$  presented by List (1958). This expression of air mass ratio is used by Meeus and Bird and Hulstrom, with the sun altitude angle  $\beta$  calculated as per equation (34).

$$m = \frac{\left[ \frac{(288 - 0.0065z)}{288} \right]^{5.256}}{\sin(\beta) + 0.1500 \cdot (\beta + 3.885)^{-1.253}} \quad (51)$$

Klein (1948) identified two distinct effects of dust on the solar radiation that are the scattering  $d_s$  and the absorption  $d_a$ . However, the absorption is negligible, which means  $d \approx ds$ . Kimball (1930), Bolsenga (1964), along with Klein, have established a set of values for the effect of dust according to the environment as per table 5. The dust values range from 0 to 0.08 in remote areas and from 0.03 to 0.13 in urban areas.

In order to calculate the total irradiance, several parameters need to be taken into account. The relative position of the Earth if respect to the sun is one of them. Please note that all the following astronomical considerations are drawn from Meeus (1999). Based on the mean anomaly of the sun *MeanAn* and on the centre of the sun *CentSun*, the true anomaly of the sun expressed in degrees is by equation (52).

$$v = \text{MeanAn} + \text{CentSun} \quad (52)$$

Considering the current Julian century  $t$ , the current eccentricity of the Earth's orbit  $e$  is expressed without dimensions as:

$$e = 0.016708634 - t \cdot (0.000042037 + t \cdot 0.0000001267) \quad (53)$$

The current distance between the Earth and the Sun expressed in astronomical units (AU) both depend on  $v$  and  $e$ :

$$r = \frac{1.000001018 \cdot (1 - e^2)}{1 + e \cdot \cos(v)} \quad (54)$$

The eccentricity correction, which is dimensionless, depends on the current Earth-Sun distance  $r$  compared to the average distance  $r_0$ , which is 1 AU.

$$E_0 = \left( \frac{r}{r_0} \right)^2 \quad (55)$$

This leads to a new expression of the apparent extraterrestrial irradiance  $A$  that is function of the solar constant  $SC$  and of the uncorrected sun altitude angle from equation (34):

$$A = SC \cdot E_0 \cdot \sin(\beta) \quad (56)$$

This model takes into account 4 different causes of sunlight scattering.

The water vapor suspended in the atmosphere participates to scatter sunlight. Indeed, depending on the dew-point temperature  $T_{dpt}$ , the mean hourly content of water that can potentially precipitate  $w$ , expressed in cm, as per Reitan (1963) and Bolsenga (1965), is given by equation (57):

$$w = \exp(-0.0592 + 0.06912 \cdot T_{dpt}) \quad (57)$$

All the following formulas are based on the expression of the air mass ratio given by equation (51). The transmittance of water vapor  $X_w$  in cm is as follows:

$$X_w = w \cdot m \quad (58)$$

Bird and Hulstrom (1981) determined the transmittance of water vapor as the dimensionless expression:

$$T_w = 1 - \frac{2.4959 \cdot X_w}{(1 + 79.034 \cdot X_w)^{0.6828} + 6.385 \cdot X_w} \quad (59)$$

The ozone content in the atmosphere also plays a role in scattering sunlight.

The ozone content value  $U_0$  has been empirically measured as 0.31 cm for mid-latitude summer atmospheric conditions.

Still according to Bird and Hulstrom (1981), the amount of ozone in a slanted path  $X_0$  is:

$$X_0 = U_0 \cdot m \quad (60)$$

As a result, the transmittance of ozone content is given by equation (61).

$$T_0 = 1 - 0.1611 \cdot X_0 \cdot (1 + 139.48 \cdot X_0)^{-0.3035} - 0.002715 \cdot X_0 \cdot (1 + 0.044 \cdot X_0 + 0.0003 \cdot X_0^2)^{-1} \quad (61)$$

Uniformly mixed gases like carbon dioxide and oxygen are responsible for scattering sunlight as well. The transmittance of such gases is the following:

$$T_{UM} = \exp(-0.0127 \cdot m^{0.26}) \quad (62)$$

In addition, the transmittance of Rayleigh scattering is given by:

$$T_R = \exp(-0.0903 \cdot m^2 \cdot (1 + m - m^{1.01})) \quad (63)$$

Finally, aerosols scatter sunlight too. Considering the aerosol optical depth  $\tau_{A0,38}$  in a vertical path for a wavelength of 380 nm, for which there is no molecule absorption, and the aerosol optical depth  $\tau_{A0,50}$  at 500 nm wavelength, which corresponds to ozone absorption, the overall turbidity  $\tau_A$  is the following:

$$\tau_A = 0.2758 \cdot \tau_{A0,38} + 0.35 \tau_{A0,50} \quad (64)$$

The values of  $\tau_{A0,38}$  and  $\tau_{A0,50}$  vary significantly with the location and must be empirically measured. Their values may typically be between 0 and 1. Similarly, the overall atmospheric turbidity  $\tau_A$  may have values typically comprised between 0.02 and 0.5.

The transmittance of aerosol absorption and scattering  $T_A$  is function of the overall atmospheric turbidity  $\tau_A$  and the air mass ratio:

$$T_A = \exp(-\tau_A^{0.873} (1 + \tau_A - \tau_A^{0.7088}) \cdot m^{0.9108}) \quad (65)$$

Considering an empirical transmittance of aerosol absorption  $K_1$ , which recommended value is 0.1, even though the value should be empirically evaluated as it may reach 0.35 in urban areas, the transmittance of aerosol absorption  $T_{AA}$  is:

$$T_{AA} = 1 - K_1 \cdot (1 - m + m^{1.06}) \cdot (1 - T_A) \quad (66)$$

Considering the transmittance of empirical ratio of forward-scattered irradiance  $B_a$ , which recommended value is 0.84 according to Bird and Hulstrom (1981), the atmospheric albedo  $r_s$  is expressed as:

$$r_s = 0.0685 + (1 - B_a) \cdot \left(1 - \frac{T_A}{T_{AA}}\right) \quad (67)$$

The atmospheric albedo  $r_s$  depends on the transmittance of aerosol absorption and scattering and on the atmospheric turbidity.

Using the previous expressions from the equations (56), (59), (61), (62), (63), (65), and (67), the direct horizontal beam radiation is expressed as:

$$I_{BH} = 0.9662 \cdot A \cdot T_A \cdot T_w \cdot T_{UM} \cdot T_0 \cdot T_R \quad (68)$$

Similarly, it can be inferred from equations (50), (56), (59), (61), (62), (63), (65), and (66), that the diffuse radiation from atmospheric scattering is:

$$I_{DH} = 0.79 \cdot A \cdot T_{AA} \cdot T_w \cdot T_{UM} \cdot T_0 \cdot \left( \frac{0.5 \cdot (1 - T_R) + B_a \cdot \left(1 - \frac{T_A}{T_{AA}}\right)}{1 - m + m^{1.02}} \right) \quad (69)$$

$B_a$  is the empirical ratio of forward-scattered irradiance, already mentioned above.

Finally, the total horizontal irradiance  $I_{TH}$  is expressed as follows, on the basis of equations (67), (68), and (69):

$$I_{TH} = \frac{I_{BH} + I_D}{(1 - \rho \cdot r_s)} \quad (70)$$

$\rho$  is the ground reflectivity or albedo, and  $r_s$  is the atmospheric albedo.

## 6.2 Irradiance forecasting considering clouds

The precision of a PV power forecast depends greatly on the precision of the solar radiance forecast. So conclude E. Lorenz et al. [59]. In their study, these authors have managed to use one-day ahead irradiance forecasts at the scale of a country to predict regional PV power output up to three days ahead. Overall, the method consists in considering a regional observation, which greatly improves forecasts with respect to a single site observation. It must be noted that the precision of this method depends on the size of the observed region. Indeed, the larger the region where the PV systems are distributed, the more precise the forecast. As an illustration, the overall rRMSE for a day ahead for a single PV site is 37%. This value is decreased to 19% for a region of size 3° latitude high and 3° longitude wide. The rRMSE is further reduced to 13% for the whole of Germany, which dimensions are 9° × 10°. To describe the methodology that is used, a global numerical weather prediction (NWP) model of resolution 3 hours and 0.25° × 0.25° (which is approximately 27 km × (16-19) km in Germany) from the European Centre for Medium-Range Weather forecasts (ECMWF) is used as the source of forecasted irradiance data. A network of 11 PV system with a resolution of 200 × 120 km is also used to measure the power output. Finally, 200 weather stations spread over Germany are utilized to measure the irradiance over a period of 10 months. A spatial interpolation technique on a region of 100 km × 100 km is first utilized to refine the ECMWF global irradiance forecasts for a specific given site. Additionally, a temporal interpolation is performed by combining the 3-hourly forecast data *Gforecast,3h* provided by the ECMWF with a clear sky model. The 3-hourly mean value clear sky index  $k_{3h}$  is calculated as  $k_{3h} = G_{(forecast,3h)} / k_{3h}$ . The hourly mean value  $k_{1h}$  is then derived by linear interpolation. Alternatively, the mean hourly irradiance can be directly interpolated from the original 3-hourly data *Gforecast,3h*, but this method has proved to be less accurate in situations without clouds or with only a few. A correction may be necessary in some sky configurations. For example, in the case of the German region, the forecasted irradiance is underestimated in overcast situations with low irradiance ( $k < 0.2$ ) and overestimated in the case of variable cloud cover ( $0.3 < k < 0.8$ ). The values are thus replaced by the clear sky model irradiance in the case of a total cloud cover below 0.03.

In addition, a situation-specific bias is introduced for correction purposes. Other studies also use linear time series of global irradiance, possibly normalized with respect to a clear sky model, as an input with a two stage approach. During the first stage, either a PV system network [60] or satellite data [21] are utilized to forecast the irradiance.

In a second stage, some autoregressive models such as autoregressive moving average (ARMA) [21] or autoregressive with exogenous input (ARX) [60] are also used. However, such methods only focus on the irradiance and neglect the effect of temperature and wind on the performance of the systems.

Moreover, sudden changes in cloud cover may not be described in the model. As a matter of fact, the wind plays an import role in the abrupt changes of irradiance due to the passing of clouds. With an increasingly fine temporal resolution, the swift changes of cloud cover play a more and more important role.

Independently from the irradiance forecasting method, the source of data can be chosen on the basis of the desired temporal resolution. A finer resolution can always be obtained through interpolation.

Such temporal interpolations can be satisfactory during periods when the sky is completely clear or overcast. However, they may lead to significant errors in situations during which the sky cover varies stochastically, as it is the case in sub-hourly forecasts. Table 9 gives a broad overview of the various sources of data that may be used to forecast solar radiation. Hassanzadeh et al. (2010) suggest that satellite data well adapted for very short-term forecasting, whereas numerical weather prediction (NWP) data are better suited for long term forecasting, beyond two days. However, for fine temporal resolutions, real-time measured irradiance data can dramatically improve the performance of now-casting. In particular, Bacher et al. (2009) have demonstrated that direct measurements of the irradiance prevail in forecasts up to 2 hours ahead whereas NWP inputs are more valuable beyond that horizon.

On top of the input source, the targeted temporal resolution also defines the choice of forecasting methodology. For example, grid-connected photovoltaic plants require a horizon of at least 24 hours for integration purposes. Mellit et al. [23] have chosen to use an ANN-based approach for such a 24-hour irradiance forecast. More specifically, this team has utilized a multilayer perceptron to process the daily average irradiance and daily average temperature. The data consists in measurements taken in Italy over the course of several months. The choice of a method based on artificial intelligence is justified by the nature of the input data, as it is a time series containing several meteorological variables. With the help of a cross validation method, the team has managed to evaluate a performance above 98% during sunny days but below 95% during cloudy days. The authors suggest that adding more input parameters such as cloud, pressure, and wind speed and sunlight duration could possibly improve the accuracy of the forecast.

*Summary:* A clear sky model requires two main sub-models. The position of the sun needs to be precisely defined. The solar position algorithm developed by Reda and Andreas is the most precise to date. Additionally, the direct or global horizontal solar radiance also needs to be modeled. The Meeus and Bird and Mulstrom model is the most accurate with respect to measured insolation values. After modeling the irradiance in absence of clouds, the solar radiance must be modeled by considering the effect of clouds. Doing so requires to follow a methodology. The temporal resolution is chosen in accordance with the grid application that is sought. The selected temporal resolution influences the choice of the source of input data. A sky camera is the best source of input data for very short-term forecasting, whereas sensor networks and satellite images are both very suitable for short term. NWP data only performs well for longer horizons. In addition, the temporal resolution also conditions the forecasting method. For very-short horizons, persistence-based methods are the best, time series and ANN methods are the most appropriate for short term whereas an NWP approach is well-fitted for longer terms.

## 7. Cloud coverage modeling and cloud classification

Using Energy storage units (ESU) has long been proposed as a solution to dispatch PV power. Frequency domain has been previously utilized to specify PV power. Yet, this approach is not suitable for sizing an ESU. It is indeed necessary to know the peak energy amount corresponding to a charge or discharge of the battery [13]. Therefore, modeling and predicting cloud-related PV power intermittency is required for an efficient ESU control strategy with a deterministic scheduling approach. Studies have proven that two parameters can strongly impact the generation of PV power: the cloud coverage of the sky and the cloud type or class [25]. The cloud coverage can either be expressed as a percentage or in octas, also spelled oktas, which is a unit corresponding to 1/8th of the sky, i.e. 12.5%. Indeed, the size, thickness and altitude of the clouds covering the sky can greatly influence the direct solar irradiance received by a photovoltaic module [24, 29]. As a matter of fact, both the total coverage of the sky and the altitude of the cloud define the size of the shaded area on the ground, whereas the thickness of the cloud is linked to its opacity and by extension to its ability to block sunbeams. Consequently, considering a precise sky model is very relevant for PV power forecasting purposes.

The cloud coverage can be measured by analyzing the individual pixels of digital pictures obtained from a sky camera. Clear sky pixels are characterized with a high saturation of color and a predominance of blue. As opposite to cloud-free pixels, cloudy pixels display dull colorations featuring different shades of whites, possibly tinted of colorful hues ranging from red to blue.

These characteristics have been notably used by Martins et al. (2003) and Souza-Echer et al. (2006) to define a method to estimate the cloud coverage [25]. Alternatively, in the absence of a sky camera, the cloud cover can also be defined as the percentage of the ground area that is covered by cloud shadow [29]. This last approach implies however the use of an accurate cloud shadow model. This is an approach that can be used for example with a regional network of dispersed ground sensors. After correction of shading due to permanent obstacles that are not cloud, the clear-sky index  $K$  of a location can be defined as:

$$K = \frac{G}{G_{clear}} \quad (71)$$

In equation (72),  $G$  is the instantaneous irradiance and  $G_{clear}$  the predicted clear sky irradiance [15].

By convention, there is a total of 27 identified types of clouds, equally divided within three altitude levels also called étages. Each étage thus contains 9 types of clouds. It must be noted that the elevation of the base of each type of cloud depends on local atmospheric conditions. This elevation varies according to the season and the location. Table 10 illustrates some approximate heights [61].

The sunlight intensity  $I_0$  observed under a cloud that scatters and absorbs a clear-sky intensity of sunlight  $I$  depends on the optical depth or optical thickness of the cloud.

The optical depth  $\tau$  is expressed by the relationship [62]:

$$\frac{I}{I_0} = e^{-\tau} \quad (72)$$

From a shading point of view, it may not be relevant to deal with the whole complexity of all 27 types of clouds. Indeed, several types may cause similar shading effect. Heinle et al. have managed to automatically classify clouds into 7 genera [24, 25] summarized in Table 11 and illustrated in Figure 5. Some values of optical thickness defined by the International Satellite Cloud Climatology Project (ISCCP) [63] are given in Table 8. This quantity is directly related to the opacity of the clouds. However, some average values of the optical depth or thickness for each type of clouds are difficult to get, as per the authors' knowledge. A study is currently in progress to define simple and reliable values that could be used to define the transparency of clouds. Indeed, both the type and the thickness of a cloud directly affect the beam component of sunlight as mentioned by Chengrui et al. [29].

Clouds can be categorized in four main categories according to their altitude level.

Low level clouds occur below 6,500 feet i.e. 2,000 meters and are either named *cumulo-* (meaning 'heap') or *strato-* (which means 'layer') [64]. Stratus, cumulus and stratocumulus clouds are the only types that exclusively populate the lower level. Stratus (St) clouds are composed of a uniform and diffuse flat layer of gray cloud cover [65]. As they are partially transparent, the outline of the sun is clearly visible when seen through such clouds. This type of cloud may remain stationary from several hours to a couple of days but some layers may break up or dissipate, partially revealing some clear sky. They develop horizontally. In contrast, cumulus (Cu) clouds are detached and dense with a sharp outline. They grow vertically and eventually mature as cumulonimbus clouds prior to eventually dissipating. They are characterized with flat horizontal bottoms and rounded tops that resemble cauliflowers. The sunlit parts of such clouds are bright white while the base is darker. They usually develop during clear sky conditions, due to diurnal convection. Low-altitude cumulus clouds may move fast and cause significant sunlight discontinuity as these clouds have a clearly defined edge and cast a deep shadow [29]. On the other hand, stratocumulus (Sc) clouds are a hybrid type composed of several individual cumulus clouds distributed in characteristic stratus layers. They consist in a mix of thick and thin layers of gray or whitish patches. They also almost always have dark rounded masses resembling a honeycomb.

Mid-level clouds occur between 6,500 and 20,000 feet i.e. between 2,000 and 6,100 meters. Their names are given the prefix *alto-* (meaning 'high'). The two main types are altostratus and altocumulus. Altostratus (As) clouds are similar to stratus in structure and can be associated with them. They present gray or bluish layers of fibrous cloud sheets that may totally cover the sky. They are very thin and commonly reveal the sun as if it was seen through ground glass. They will thus not cause any significant decrease of insolation. They do not create any halo phenomena and do not cast any shadow on the ground. Similarly to stratocumulus clouds in low levels, altocumulus (Ac) clouds are constituted of clusters of cumulus type features in middle levels. They are the most common mid-level clouds, composed of thin semitransparent white or light gray layers of rounded shapes that may be partly diffuse and fibrous. Several layers may appear, often in presence of other types of clouds. A corona appears when such clouds pass in front of the sun. This ring has blue in the inside and red on the outside.

High level clouds appear above 20,000 feet i.e. 6,100 meters and are given

the prefix *cirro-* (which means 'wisp of hair'). Cirrus clouds (Ci) are wispy and feathery as they are entirely composed of ice crystals. They consist in delicate separate hair-like semitransparent white silky filaments and they hardly diminish the insolation. Cirrostratus (Cs) clouds form a widespread layer of whitish transparent veil-like clouds similar to low level stratus clouds. They almost always cover the whole sky with an extensive sheet cover. The sun produces a halo when hidden by a layer of cirrostratus.

Cirrocumulus (Cc) clouds are layered with small cumulus-like clouds. They correspond to a degraded stage of cirrus or cirrostratus. They are composed of more or less regularly arranged very small thin white grains, forming a 'mackerel sky' or 'buttermilk sky'.

Finally, some enormous clouds may be present in all three levels of the sky. This is the case for cumulonimbus (Cb) clouds, which correspond to a mature stage of vertical development of cumulus cloud. These dense clouds are responsible for storms, hail and heavy rains. The base is very dark and the overall cloud is very large and extremely opaque. The upper part is usually smooth and flat, reminding the shape of an anvil. Like cumulonimbus clouds, nimbostratus (Ns) clouds also occupy a large range of altitude, although not necessarily as high.

They are dark gray, very opaque and they obscure the sky, causing very low insolation. Nimbostratus clouds result from thickening dense stratus, altostratus or stratocumulus clouds. They cause episodes of continuous rain or snow.

A chart summarizing the various types of clouds is given in table 12.

It must be noted that to the knowledge of authors like Cai et al. [29], no universal algorithm can deal with all types of clouds due to the complexity of a comprehensive atmospheric model. Indeed, many occurrences may simultaneously happen on various levels of altitude, caused by unrelated parameters. The hereinabove mentioned authors recommend to focus on low level cumulus clouds as they are the type of clouds that are responsible for most of fluctuations. The same authors also propose to model the solar irradiance in a probabilistic manner. Climate science has established that the shape of cumulus cloud shadows contours have a fractal structure [12, 29].

Cai et al. use a midpoint displacement algorithm to model this structure. The algorithm used result in a 3-dimensional fractal output, which is only a mathematical artifact that enables to generate cloud shadows by intersecting a cross section of this output with a horizontal plane. Barnes et al. [12] point out that the size  $x$  a cumulus clouds shading a ground area for a period of time  $t$  follow the power law distribution where  $A$  and  $b$  are distribution parameters:

$$f(x) = A \cdot x^b \quad (73)$$

It has been inferred that the clear and shaded conditions distribution of the same ground area follow a generalized Pareto distribution where  $k$ ,  $\sigma$  and  $\theta$  are parameters of the distribution. The duration of the clear condition is truncated based on the length of the day. The distribution is as follows:

$$f(t | k, \sigma, \theta) = \frac{1}{\sigma} \cdot \left[ 1 + k \cdot \frac{(t - \theta)}{\sigma} \right]^{-1 - \frac{1}{k}} \quad (74)$$

In order to evaluate the area shaded by a cloud, it matters to know the altitude of the cloud base. This can be either done by using a ceilometer or indirectly by calculating the lifted condensation level. A laser ceilometer uses a laser beam that is reflected back by the base of clouds.

By using either time travel method or triangulation, the device can measure the distance of a light beam.

Generally speaking, the altitude range for such a device is 25,000 feet / 7,620 meters and clouds above this height are not detected. Alternatively, knowing the dew point temperature, the ambient temperature and the altitude of the weather station with respect to the sea level, it is possible to evaluate the altitude of the base of low level clouds like cumulus clouds. The lifted condensation level above sea level (ASL) corresponding to this altitude can be evaluated in meters with the expression [66]:

$$LCL = 400 \times 0.3048 \times (T_{amb} - T_{dew}) + h \quad (75)$$

In equation (75), LCL is given in meters. The variable  $h$  is the altitude above sea level (ASL) of the weather station, also expressed in meters, and  $T_{amb}$  and  $T_{dew}$  are respectively the ambient and the dew point temperature both expressed in °C, at the exact location of the weather station.

When the dew temperature value is not available, it is possible to approximate it based on the ambient temperature and the relative humidity ratio, as given in equation (76) with RH standing for the relative humidity, expressed as a percentage [67]:

$$T_{dew} \approx T_{amb} - \left( \frac{100 - RH}{5} \right) \quad (76)$$

Identifying accurately the type of cloud and classifying clouds also helps to get an idea of the altitude range. This is mainly useful for high level clouds.

A full classification of clouds helps to evaluate the cloud cover and hence cloud length as well as the type of clouds that are likely to cause shadows above a specific geographic site.

### 7.1 Cloud Classification Part 1: Assessment of cloud coverage

It is necessary to process images from a whole sky camera to evaluate the cloud coverage. This can be performed with a pixel by pixel classification and sorting pixels as either ‘clear sky’ or ‘cloudy’ pixels. As a first step, an image-mask must be applied onto the images in order to filter out any irrelevant information that are not related to the sky, like the ground or surrounding infrastructures. The solar disk and visible sunrays may also bias the estimation of cloud coverage and must be eliminated from the image as well. This can be done by using the fact that the related pixels in the image are usually completely white, revealing an overexposure of the camera’s sensor. In addition, a whitening effect of cloud-free pixels around the solar disk may lead to a misclassification of these pixels. After filtering all irrelevant pixels, the remaining pixels need to be classified as ‘clear sky’ or ‘cloudy’. An interesting property of the atmosphere can be used at this effect. As a matter of fact, in a clear atmosphere, which means in absence of natural (like fog or mist) or artificial (like haze or fog) aerosols, air molecules scatter more blue light than red light, making the sky appear blue to the human eye.

In contrast, clouds scatter blue and red light in similar proportion, leading us to see them light grey or even white [24, 25]. Kazantzidis, Tzoumanikas et al. (2012) give an excellent overview of the current cloud pixel discrimination criteria [25]. In particular, these authors point out the fact that Pfister et al. (2003) [68] and Long et al. (2006) [69] have

developed a method to process color sky images by comparing the RGB images with some blue and red thresholds in order to estimate the total cloud coverage. By using this method, pixels can be classified as cloudy or not cloudy according to the value of the red/blue (R/B) signal ratio. According to Long et al.’s method, pixels with values above a threshold value  $R/B = Th_1$  are classified as cloudy. Although Long et al. have used the value 0.6 for  $Th_1$ , this threshold value may vary a lot according to the camera that is used. Indeed, the R/B criterion lacks robustness and the threshold value needs to be recalibrated for each new set of images coming from a different camera. For example, Heinle et al. (2010) [24] used a value of 0.8 for  $Th_1$ . Additionally, some problems may arise to detect thick clouds or to classify circumsolar pixels.

To tackle these issues, Heinle et al. propose to use the difference  $R-B = Th_2$  as a criterion instead of the ratio  $R/B$ . The value  $Th_2 = 30$  is used by this team as an optimal reference threshold for their own set of images and for their own application. Though more robust than the  $R/B$  threshold  $Th_1$ , this threshold value  $Th_2$  is also relatively sensitive to the set of images it is applied to and may need adjustment.

Kazantzidis, Tzoumanikas et al. [25] have also verified that the difference  $R-B$  outperforms  $R/B$  for a sufficiently large set of images. Nevertheless, this team has discovered that a multi-color criterion combining R, G and B thresholds ( $(B > R + Th_1) \& (B > G + Th_2) \& (B > Th_3)$ ) used to detect clear sky pixels outperforms the R-B difference method in some common cases of cloud cover. This criterion is the most robust and accurate one so far. The reference values of  $Th_1 = 20$ ,  $Th_2 = 20$  and  $Th_3 = 60$  are optimum for the Canon Ixus used by this team and may require some tuning for other cameras.

However, no criterion at the authors’ knowledge can completely reliably discriminate thin cirrus or cumulus clouds close to the sun. This extreme case is often problematic and a source of significant uncertainty, leading to a possible underestimation of the cloud coverage.

### 7.2 Cloud Classification Part 2: Identification of cloud types

Heinle et al. (2010) [24] automated the classification of clouds as a further step following the *clear/cloudy* classification. On the basis of their Fisher Distances, a technique used in satellite imagery (Pankiewicz, 1995), 12 features have been selected to classify clouds. Seven of these features are spectral while 4 are textural and 1 is simply the total cloud cover. Indeed, the spectral features of an image do not provide any information about the spatial distribution of clouds. However, using colors solely is insufficient to distinguish altocumulus from stratocumulus clouds. This is why textural features must also be taken into consideration. In addition, the cloud cover is also used as a feature to tackle the issue of spatial distribution. In particular, estimating the total cloud cover helps to distinguish stratiform clouds from the rest. The color information of each cloud image is partitioned into a set of three monochromatic (also referred to as Grey Level) red (R), green (G) and blue (B) data sets. Each monochromatic (or grey level) partial image is treated separately as levels of R or G or even B. The average color and the tonal variation of the image both provide some useful information to distinguish thick and thin clouds, or even to tell high altitude bright clouds from other clouds.

The color component B proves to have the highest separation power due to the fact that it is the dominant color of the sky. Table 13 [24, 25] comprises some details related to the 12 hereinabove mentioned features. The color for which the feature is calculated is specified between the brackets of the name of the feature. In the expressions, N represents the total number of pixels in the image under consideration.

The values  $p_{R,i}$  and  $p_{B,i}$  are respectively the monochromatic intensity value of red and blue for an individual pixel of the respective red and blue grey level image referred to with the number  $i$  so that  $0 < i < N-1$ , N being the total number of pixels inside the grey level image. The variables  $a$  and  $b$  represent two different monochromatic intensity levels within the blue grey level dataset of the image. The values of these variables are comprised between 0 and  $G-1 = 255$ , where  $G = 256$  is the number of grey level intensities. The spatial relation between the pixel of intensity  $a$  and the pixel of intensity  $b$  is also called an offset. In particular, the spatial relation relationship [1,1] corresponds to a bottom right single pixel offset. The value  $n_{occ[1,1]}(a,b)$  represents the number of occurrences that a pixel with an intensity of  $a$  has a pixel with an intensity of  $b$  as its direct diagonal bottom right neighbor inside the blue grey scale dataset. It must be noted that the matrix of dimension  $G \times G$  containing all  $n_{occ[1,1]}(a,b)$  values with  $0 \leq a \leq G-1$  and  $0 \leq b \leq G-1$  is called the Grey Level Co-occurrence Matrix (GLCM).

A very accurate k-nearest-neighbor (kNN) classifier is used to discriminate sky images into one of the seven cloud conditions on the basis of the features mentioned in Table 13.

Heinle and al. used a sample of 1500 reference images for training purposes, containing about 200 independent images per cloud class, with a wide variety of cloud forms. After training the system and processing the input images, the Manhattan Distance is calculated to assess the difference between the normalized feature vector  $f$  of each element and the reference feature vector  $f_{ref}$ . On this purpose, the following formula is used:

$$d(f, f_{ref}) = \sum_{i=0}^{\dim} |f_i - f_{ref,i}| \quad (77)$$

The cloud visualized in the input images is classified according to the majority vote of the k closest matches to each known cloud class. The parameter k has been set as 3 by Heinle et al., leading to an average performance of 96% elements classified correctly over all 7 cloud classes. The performance is measured by using a Leave-One-Out Cross-Validation (LOOCV).

The authors noticed that the only difference between cumulus and altocumulus clouds is the size of individual cloudlets.

Additionally, some noticeable misclassification errors appear due to a confusion of stratiform clouds and thick rain clouds, such as nimbostratus. This is due to the fact that stratiform clouds such as stratus and stratocumulus can build up to form thick rain clouds that still display the same characteristics. Moreover, the presence of rain drops on the camera dome may distort images and lead to some further misclassifications.

To address these problems, Kazantzidis et al. (2012) [25] improved this classification method by introducing subclasses and by proposing a method to detect the presence of raindrops over the camera lens. A circle factor (CF) is used to evaluate how close to a perfect circle the fish-eye image of the camera is. In the presence of rain drops, the contour of the image is distorted, leading to a circle factor of about  $CF=0.6$  whereas this factor is close to  $CF=0.9$  with a dry dome. This factor does not only help to prevent a misinterpretation of texture values but can also detect the

presence of rain clouds. Furthermore, as there is only small differences of entropy feature among the various cloud classes, Kazantzidis et al. have decided not to use this feature. A set of 1050 images have been used to train the algorithm after visual inspection, while 1500 images are used for testing. A minimum of 150 images per cloud class has been used for the training, with only one cloud class per image. In order to tackle the problem of variability of cloud within the same cloud class, three new parameters have been introduced, as per Table 14 [25]. According to the values of these parameters, three tiers are defined.

These new parameters lead to the creation of a certain number of new subclasses for each cloud class as per Table 15 [25].

The accuracy of the overall classification ranges between 78% and 95%. In particular, cumulus clouds are detected with a success rate of about 92%. Despite these innovations, a small confusion remains between the cloud classes St-As and Cb-Ns. The misclassification of St-As engenders the lowest performance of 78%. However, a correct classification of clouds is very relevant for PV power forecasting when the optical depth of each cloud class is known.

*Summary:* The shorter the forecasting timeframe, the more dramatic the impact of clouds on the power output of PV systems. As it has been concluded in section 7, a sky imager is the best option of input data. This source of input images is indeed the only one that can help to distinguish the different types of clouds. Considering the dynamics of clouds, a sampling period of 1 second between each consecutive images is ideal to extract an accurate cloud motion vector. Adequate post-processing of the images is necessary to determine the cloud coverage as well as the type of clouds present in the sky. A multicolor criterion combining different thresholds for R, G and B data is the optimal method for this step. The images from a sky camera can also possibly be used to estimate the size of clouds are therefore the area of the cast shadow. After identifying the cloud masking the sun by comparison with a database, it is also possible to infer the opacity of this cloud and thus the extra scattering of sunlight. For this purpose, it is indeed vital to be able to identify each type of cloud among a set of 7 classes. Heinle's method performs very well to identify clouds accurately. It consists in working on monochromatic R, G and B images. A set of 12 features are extracted and discriminated through a kNN classifier with k ideal set to 3. A risk of misclassification nonetheless remain leading to a possible confusion between stratiform clouds and rain clouds, which both scatter sunlight differently. A solution proposed by Kazantzidis consists in introducing 3 more criteria and subclasses. However, to consider the effect of shading on PV systems, a sensor network remains the ideal option as it measures directly the area and shape of the shadow cast by clouds on the surface. For this reason, such networks are often utilized as reference data. Pyranometer networks are precise but expensive although it is possible to use an existing PV network instead to cut costs. The size and configuration of the network must nonetheless fit with the required spatial resolution. Independently from the source of input data, the altitude of the base of low altitude clouds, which are responsible for the fastest variations, can be assessed from the ambient temperature, the dew temperature and the elevation of the weather station.

## 8. Weather Classification for mid-term PV Output Forecasting

For many utility applications including power dispatch and grid balance, a short-term to long-term horizon suffices, as mentioned in Table 4. In this instance, PV power forecasting methods consist in focusing on a period of time that ranges between 30 minutes and several days ahead. Such methods may mix statistical and physical approaches with either PV networks or numerical weather prediction (NWP) as an input.

Peder Bacher et al. [60] have demonstrated that direct solar power observations are the most appropriate source of input data for forecasts up to 2 hours ahead whereas NWP is better suited beyond the horizon of 2 hours.

Weather classification is widely used for forecasting purposes for horizons exceeding 24 hours. For example, a day-ahead or week-ahead PV power output profile can be obtained with a classification of simple daily weather patterns such as rainy, cloudy, sunny (clear sky) and foggy days. Historical data of solar irradiance (for an indirect forecasting approach) [50] or PV output (for a direct forecasting approach) [26, 70] and meteorological forecasts are used as input data on this purpose. These two frequently used PV output forecasting methods can play a pivotal role to size the storage batteries required to complete the PV system under consideration. As an illustration, the case study of a sun-powered house in the Tokyo area, Japan, within the government supported smart city project, as described by Niimura et al. (2012) the text weather forecast from 9 years of historical data has been used to establish the classification. Table 16 [50] is an example of how the weather conditions can be reinterpreted in simple terms, based on various key words that may appear within the descriptive text of a weather forecast. Such a simplification of weather patterns is very suitable to forecast the power generation from a day to a week ahead.

The hourly profile of insolation is obtained based on the values that are statistically the most likely considering the time and weather conditions. The most likely insolation values from 4 am to 8 pm under various weather conditions are obtained from the probability distribution of insolation, which is based itself on hourly solar radiance data. The estimated radiation of a specific day and at a specific hour  $t$  is then defined based on the weather forecast by  $g(t) = G_{max} \cdot g_{most}(t) \cdot P_{precipitations}$  where  $G_{max}$  is the maximum monthly insolation,  $g_{most}(t)$  is the most likely normalized value of solar radiation at the hour  $t$  and  $P_{precipitations}$  is a factor that indicates the probability of rain. Finally, in that case study, the maximum seasonal Mean Absolute Error (MAE) of the forecast is about 0.33 for seven consecutive days. The major prediction errors occur during spring and summer in the northern hemisphere. The authors conclude that the major of the errors cause is the 80% of uncertainty of weather forecasts.

A support Vector Machine can be used for the same purpose of weather classification based PV power output forecasting for a one-day horizon. For example, Jie Shi and al. [26] select a simple weather model out of four possibilities (sunny, cloudy, rainy, foggy) on the basis of historical data and weather forecast of the next day. The historical input data are similar to the ones used by Niimura. The performance is the best for sunny days with mean Relative Error (MRE) of 4.85% whereas the MRE is 12.42% for cloudy days, with an average of 8.64% for all models.

*Summary:* For the purpose of power dispatch, SVM and weather classification prove to be extremely accurate for sunny or uniformly overcast days. Such an approach lacks precision for cloudy days though,

since the dynamic variation of the cloud cover. It can be applied when an average daily value of the solar radiance is sought but some significant errors may be expected during cloudy days. Therefore, a shorter term forecasting method is needed to ensure grid stability despite the changes of cloud cover during cloudy days.

## 9. Very short term PV power forecast

Accurate very-short term PV power output forecasting methods are required for safety and efficiency reasons. In particular, utility operators require a ten-minute warning to bring spinning reserves online. Besides ramp rates issues, an accurate forecast within a solar-aware smart grid can also help to prevent keeping industrial-scale loads such as water-pumps on when not needed. Additionally, a precise prevision enables to plan when to turn such loads on. Nevertheless, the main problem with very short-term PV power forecasting is the swift intermittency of sun radiation, which is essentially due to cloud motions. Tracking and forecasting such cloud movements is therefore necessary in order to deal with this problem. However, when it comes to deterministic prediction of PV output power, adequate temporal and space resolutions are critical [12, 29]. A first group of studies have favored the use of a numerical weather prediction (NWP) model whereas a second group prefers to use sky imagers, or geographically dispersed sensors or even satellite images to track clouds. A third group utilizes a random-sequence and time series models.

### 9.1 NWP approach

In the first group, NWP models can provide very valuable outputs such as forecasts of cloudy conditions, wind direction, wind speed and even sometimes irradiance with a large spatial and temporal resolution. Such a resolution fits well with the day-ahead forecasts needed to determine energy pricing or for energy dispatch purposes. However, the accuracy of such an approach is quite variable. In the case of a weather research and forecasting (WRF) numerical atmospheric model that would be made 30 hours in advance, the best result seem to be obtained in the occurrence of large-scale cirrus clouds or in the event of an entirely covered skies. Such a model may otherwise sometimes fail to predict the presence of clouds during a day or be mistaken in the timing of clouds. Overall, this approach is not suited to deal with small clouds in a partially cloudy day. Additionally, substantial computation time is required for NWP models. Indeed, 2 hours of computation may be needed to provide only 2 minutes of forecast [15].

### 9.2 Sky, satellite images and sensor array approach

With regard to the second group of studies, which use image processing and sensor analysis, three variables must be taken into account when tracking clouds: position, velocity (as a vector) and size [71]. Some bottlenecks faced by this approach are respectively the effectiveness of the feature extracting filter and the associated extensive computation time. Additionally, time resolution may sometimes be too coarse for a reliable forecast of power variation caused by clouds, as such fluctuations happen in matter of seconds. Figure 6 gives a rough idea of the temporal resolution with respect to the area that is observed for three different categories of input data. The range of observation must not be confused with the spatial resolution.

On the other hand, the spatial resolution may vary a lot among each type of input data source and the format of the spatial resolution may differ too much to make comparisons. For instance, sky imagers are commonly used to observe the base of clouds and are characterized by a resolution expressed in terms of pixels. In contrast, satellites observe the top of clouds or the ground with a precision expressed in meters or kilometers. Neighboring PVs or sensors of weather stations are often dispersed with an arbitrary and inconsistent distance.

An important challenge is to determine the velocity of clouds as ground-based measurements of the wind are not reliable for this purpose [15].

Indeed, the wind direction and speed may vary according to the altitude. The only relevant wind measurement would need to be done at the level of clouds.

### 9.2.1. Sky Images

In most cases, a fish-eye lens is mounted on whole sky cameras to give a panoramic field view of 180 degrees in every direction, so that the entire sky can be observed. However, the spatial resolution is limited to a local range consisting in a radius of a few kilometers around the camera. On the other hand, the time resolution of whole sky cameras is excellent as the time between two consecutive shots can be less than a second, which virtually enables real-time observations. This basically gives a significant advantage to whole sky imagers for a detailed examination of the cloud cover.

Additionally, sky cameras can be used to identify the types of clouds and small-scale variations of cloud cover over a specific spot. Combined with a solar radiation sensor, this enables to uncover the effect of specific types of clouds on the insolation. A 1-second sampling period to take and store snapshots of the sky is a popular choice [72] but slower periods ranging from 15s [24] to up to 5 minutes [25, 73-75] are also considered. The resolution of sky camera images vary within the range of 8 to 30-bit JPEG 640×480 [25, 72, 74, 75], 512×512 [73], 1024×1280 [76], or up to 3648×2736 pixels [24]. One or two axis sun trackers as well as to sun-masking techniques can be sometimes added. It must be noted that processing color images is an important step to classify clouds. Color [73] or infrared [75] filters can possibly be applied on the camera. In any case, due to the high amount of data, sky images must be automatically classified with high accuracy.

The classification is usually either based on a binary clear/cloudy basis or on an identification of cloud types. Various filtering methods are applied to the images after classification. The extraction of features lead to the determination of a cloud motion vector including cloud velocity, which is used in turn for prediction purposes. An example of complete processing flow is given in Figure 7. It must be noted that the methods used in each step of the flow may vary, especially the statistical method used prior to forecasting.

Optical flow based on Lucas-Kanade method has been successfully used by Wood-Bradley et al. [72] to process whole sky images with a prediction error from 5.3% (1 second off) to 21.2% (7 seconds off) to estimate the obstruction of the sun by a cloud, with times to shading ranging from 20 to 40 seconds. Based on jpg images, the method follows a flow of cloud detection, cloud movement tracking, cloud movement forecasting, and image processing through OpenCV functions. Unfortunately, the excessive computational time of this method may not permit to implement it online.

### 9.2.2. Satellite images

Satellite images offer a wider spatial range consisting in a few thousands of kilometers. The time resolution of forecasts based on velocimetry from satellite images is often comprised between 1 and 6 hours, which may however not be well suited to observe small-scale or fast cloud motion [15]. As an example, forecasts based on GOES satellite images have a 10 km spatial resolution and are updated every hour. Because of their limited temporal resolution, satellite-based forecasts can sometimes be outperformed by forecasts based on persistence models [71]. Additionally, low-resolution large-scale cloud observation from satellites does not permit cloud-radiation studies, e.g. determining the impact of a specific cloud type obstructing the sun on the solar irradiance reaching the ground. Nevertheless, satellite images are very useful to forecast Global Horizontal Irradiance (GHI) on a relatively short-term basis, for horizons shorter than 6 hours [77]. In particular, for up to 5 hours ahead GHI forecasts, satellite derived methods outperform considerably NWP based forecasts. It has even been demonstrated that satellite based methods are the best option for a temporal horizon comprised between 30 minutes and 6 hours [78]. This makes satellite images valuable to maintain dynamically the balance between solar energy supply and consumption. Satellite observation generally involve cloud motion vector field extraction from the images.

Some high quality services such as EUMETCAST [79] can deliver satellite images on a 15-minute basis for Europe, Africa and South America, which is an excellent temporal resolution. In particular, Dambreville et al. (2014) have used an estimation of the GHI forecast provided by HelioClim-3 images with a high resolution of 3 km × 3 km per pixel sampled every 15 minutes. In comparison, Hammer, Heinemann, Lorenz and Lückehe (1999) could observe cloudiness with 30 minutes precision and a resolution of 2.5 km × 2.5 km from METEOSAT satellite. Based on a year-long database, Dambreville et al. have implemented a 3 step method. First, a block matching algorithm is used to extract the cloud motion vector field after comparing two consecutive satellite images  $Im(t - T_s)$  and  $Im(T_s)$  where  $t$  is the time and  $T_s=1$  corresponds to a 15minutes sampling period. This cloud motion vector field represents the global movement observed between the images. As a second step, a region of interest (ROI) is found in the upstream direction. A third step consists in using statistical method or trained conversion within the ROI to define a minimum, a maximum and a mean forecast of the GHI at time  $t+T_s$ . Finally, the forecast is compared to ground measurements and corrected if needed. This flow is illustrated in Figure 8.

During the first step, the two consecutive images are divided into square-shaped pixels blocks characterized by a side  $b=20$  pixels. A research horizon  $rh=10$  pixels (corresponding to a wind speed of 120 km/h) is defined as the furthest possible translation (diagonally) of one block between the first and the second image. A block in the second image  $Im(t)$  is identified as the same block from the first image  $Im(t - T_s)$  through a maximum correlation coefficient. The vector of each block is defined by the difference of the coordinate of the block center between both images. The global motion vector is drawn from the average of the vectors within the 9 closest blocks from the site that is considered.

The region of interest is calculated on the basis of a window size expressed by:

$$ws = 1 + 2 \cdot T_s \quad (78)$$

In equation (78),  $T_s$  is the sampling period.

The atmospheric transmission  $k(t) = GHI(t,p)/GHI_{clear\_sky}(t,p)$  at the instant  $t$  is defined as the ROI's GHI normalized with respect to the clear sky  $GHI(t,p)$  at the same time  $t$  and for the same pixel  $p$ , based on the European Solar Radiation Atlas model. The value is defined as  $k = 1$  in clear sky conditions and  $k = 0$  in case of complete obstruction. The average forecasted value for  $k$  is used to get the average forecasted GHI, on the basis of a wind persistence hypothesis. The average forecasted GHI is:

$$\overline{GHI}(t + T_s, p) = \bar{k}(t + T_s) \cdot GHI_{clear\_sky}(t + T_s, p) \quad (79)$$

Similarly, the minimum and maximum values of the forecast  $k$  lead to a minimum and maximum forecasted GHI. However, comparisons with the ground measurements obtained with a sensor network show systematic errors. This is why a post-processing stage is introduced for correction. The performance of the forecast has been assessed against ground measurements. With an nRMSE between 23.5% and 30%, the corrected forecast outperforms slightly the persistence model for horizons between 15 and 45 minutes.

Variability poses a serious tracking and forecasting problem. As clouds are not plastic but continuously vary in shape and number, cloud tracking can be seen as a multi-target tracking (MTT) problem. A framework able to deal with variable number of targets is thus required. A probability hypothesis density filter (PHD) filter has been successfully applied to track cloud features (such as position, size and motion vectors) by Paolo Addesso, Roberto Conte, Maurizio Longo, Rocco Restaino and Gemino Vivone [71].

To complete the flow, a time-dependent penalty term feeds a Maximum a Posteriori Markov Random Fields (MAP-MRF) algorithm used for classification. Features are extracted after classification and used to update the multi-target state. The penalty term considers the classification of previous acquisitions for the current image classification, reducing the risk of misclassification. The purpose of such a method is to reach a trade-off between classification accuracy and reduced computation. The entire flow except the MAP-MRF takes less than a second to compute on single-core 3.4 GHz Intel CPU for each image. Nevertheless, the MAP-MRF takes 82 seconds to run in the same conditions. One of the main benefits is an increased classification accuracy compared with other methods. In particular, the accuracy has been significantly improved around the cloud borders, which often tends to generate a high misclassification rate, due to the fact that edges fade into the background with low contrast.

### 9.2.3. Sensor network

Geographically dispersed sensors or a largely distributed PV network can be another source of observation data for cloud cover. As it enables direct measurements, it is often used as a reference to assess the performance of an irradiance forecasting method. It provides with an intermediate and probably complementary spatial and temporal resolution with respect to sky imaging and satellite views. In particular, this approach offers a better resolution than forecasts based on satellite images. Indeed, the distributed range of the 83 residential rooftop PV network used by Jayadevan, Lonij, Cronin et al. [15, 80, 81] covers an area of  $50 \text{ km} \times 50 \text{ km}$  with an average spacing of 3 km between two neighbouring sensors. It records average AC power normalized by the system rating at 15-minutes intervals. A disadvantage of dealing with such a sensor network is the installation cost.

This problem can be tackled by using an existing infrastructure, as Lonij et al. did. Unlike sky or satellite images, the sun radiation is directly measured. There is thus no need to know about the optical properties of clouds or to use radiative models [15]. The infrastructure in question consists in a network of residential PV modules. The main limitation in precision is due to the relative small size of the network compared with time resolution. The sampling rate is also too low. Moreover, the same publication mentions that cloud edge velocity may not always be the same as wind velocity, which may lead to significant errors as a result. It must also be noted that Weather Research and Forecasting (WRF) numerical atmospheric models are better at forecasting the impact of slowly varying cirrus (high altitude, 9 to 12 km) clouds for up to 50 hours ahead whereas PV network is better to forecast the effects of quickly varying cumulus (low altitude, up to 5km) clouds.

### 9.2.4. Hybrid solutions

Satellite and sky images can be combined to optimize the resolution of nowcasting of clouds in near-medium future. This is a solution used by Gonzalez et al. [75] by interpolation. This team also use an ANN to detect clouds. A k-nearest-neighbour (kNN) algorithm is often used to classify cloud types [24].

Even all three approaches (NWP, satellite and sensor network) can be combined together to optimize even further the overall temporal and spatial resolution of the forecast. This is done by Lonij, Cronin et al. [15] to perform a 45-minute forecast.

## 9.3 Random-sequence and time series models

Considering the third group, the methodology involves acquiring and processing power output data commonly coming from a couple of 200-300 W PV panels with MPPT ability as well as irradiance data fed by a sensor like a pyranometer. Sampling these data at a 1-second interval is the most popular practice [12, 29]. However, it must be noted that high solar angles of incidence around sunrise time and sunset time usually result in a poor accuracy of the measurement of the power output, leading to corrupted samples. Nevertheless, little generation is expected during those times. Considering a dataset of valid samples, a threshold is usually defined to separate clear and shaded periods. Cai et al [29] consider the irradiance to define a constant threshold of 85% of the mean direct normal irradiance (DNI)  $G_{clear,avg}$ . Any period of time  $T_{shaded}$  characterized by a DNI below this threshold can be considered as shaded. With the help of a sky camera, it has been observed by this team that the most dramatic variations of the average DNI level  $G_{shaded}$  normalized with respect to  $G_{clear,avg}$  occur during short shaded periods of time, shorter than 200 s. Within these short shaded periods, the normalized irradiance may vary in range from 5% to 90% of the mean  $G_{clear,avg}$  DNI. With the help of a sky camera, it has been established that shaded periods longer than 200 s may be due to opaque and slow-moving cumulus clouds. Similarly, Barnes et al. [12] also classify the clear-sky irradiance profile into clear and shaded data by using a classification algorithm. This algorithm consists in performing a second-order curve fit on the irradiance profile completed by normalization of the irradiance profile and K-means clustering with two centroids. The probability for a PV generator to be shaded or fully insolated in the future is a crucial information to size the energy storage unit (ESU) needed to mitigate the effect of sunlight disruptions. It must be noted that only the discharge power of the ESU is taken into account to assess its size. The likelihood for a PV generator to be shaded is

calculated on the basis of the time series of binary states (either clear or shaded) considering the duration of each state. A semi-Markov discrete-time process has been modelled and applied. The maximum number of tap changes per day as well as the minimum clear-sky output power required to induce a tap change need to be established based on the lifetime of system elements and load-flow analysis.

For the specifically studied day, ESU throughput has been established as 2.625 MWh with 20 tap-changes per day. Nonetheless, this method is used by Barnes et al. for selecting an offline control strategy. The authors would recommend to consider a live feed of weather data to improve the performance. The results presented by the same authors related to this method are shown in Figures 9 (a) and (b) [12] (courtesy of A.K. Barnes, Dr. J.C. Barda and J.K. Hayes, university of Arkansas).

*Summary:* Optical flow is very precise but takes too long to compute. Hence, such a method can only be used off-line. Since the spatial resolution plays a pivotal role in irradiance forecasting, sky cameras can provide with the ideal space resolution to define very-short term cloud motion vectors. Sky cameras are also needed for cloud classification. A persistence approach also proves to outperform satellite-based methods. However, sensor networks are the reference source of input data since they can measure directly the irradiance and do not require complex cloud models. Such networks are nonetheless costly and may not have the adequate temporal and spatial resolution to track clouds precisely. Consequently, sky images sampled every second combined with a sensor network with a refresh of data every few minutes is an ideal combination for an optimal accuracy of irradiance forecasting. Indeed, the correlation of both systems enable to uncover the effect of shading of each type of cloud. Satellite images may also considered in addition to increase the spatial resolution and to observe the top of clouds.

## 10. Conclusions

As of now, PV power remains not fully predictable and hence relatively expensive and difficult to dispatch. However, a few approaches can be envisaged to improve the possibility to dispatch PV energy, to optimize its generation and to plan its storage, making it a more reliable energy resource. In particular, controlling the energy storage can mitigate the discontinuity of PV generation over various periods of time. However, an accurate forecast of the PV power output is necessary to control a system containing PVs, energy storage and additional machines. The temporal and spatial definition of the system must be defined as a first step, accordingly with the intended grid operation. In particular, to guarantee power continuity and safety, a very fine resolution is needed. Although, various approaches can address the problem, the persistence model tends to be the most accurate approach for very short-term forecasting. ANNs and SVM are appropriate approach for short-term horizons and NWP are better suited for longer horizons. While a probabilistic method based on historical data may be valuable for very long term forecasts, such an approach cannot take into consideration the complex variations of the cloud cover causing short-term sunlight disruptions. Only a deterministic atmospheric modelling approach can deal with the stochastic changes of solar radiance during the day. Within this type of model, NWP data-based models are well adapted for day ahead forecasts but suffer from a too coarse temporal resolution. On the other hand, sky imagers, geographically dispersed sensors or satellite images are excellent sources of data for cloud identification and tracking. Satellite images give a large-scale overview of cloud events, possibly with a good resolution. Sky

imagers are a precious tool to identify cloud types and anticipate the impact of the shading on PV power generation. A hybrid system combining at least a sensor network and a sky camera performs better than separate systems. Even combining all three solutions can outweigh the drawbacks of all individual solution. Low altitude clouds like cumulus clouds may cause fast occurring power discontinuity whereas thick rain clouds like nimbostratus clouds may cause long-lasting and severe power outting. Finally, it has been found that observing offline the effect of cloud-related shading on solar power generation over an adequate period of time can help to calculate the size of the required energy storage. This can notably be successfully done by using random sequence and clear/shaded classification. Some further improvements can be sought in the future. In particular, it would be valuable to elaborate algorithms that can calculate cloud coverage and classify clouds using online data and a fine sampling period. In addition, measuring precisely the effects of each type of cloud on the solar irradiance could greatly help. Finally, finding a method do distinguish reliably rain clouds from stratiform clouds would be valuable.

## ACKNOWLEDGEMENTS

The authors are grateful to the Australian Research Council (ARC) for their financial support through the ARC Discovery Grant DP110104554. The authors also would like to thank Arthur K. Barnes, Juan Carlos Barda, Jonathan K. Hayes, Paolo Adesso, Roberto Conte, Maurizio Longo, Rocco Restaino, Gemine Vivone, Panagiotis Tzoumanikas, Andreas Kazantzidis, and Anna Heinle for their assistance.

## REFERENCES

- Keller, B. and A. Costa, *A Matlab GUI for calculating the solar radiation and shading of surfaces on the earth*. Computer Applications in Engineering Education, 2011. **19**(1): p. 161-170.
- Mellit, A. and S.A. Kalogirou, *Artificial intelligence techniques for photovoltaic applications: A review*. Progress in Energy and Combustion Science, 2008. **34**(5): p. 574-632.
- <http://arena.gov.au/files/2013/08/Chapter-10-Solar-Energy.pdf>
- Masters, G.M., *Renewable and efficient electric power systems / Gilbert M. Masters*. Vol. Chapter 7. 2004, Hoboken, NJ: Hoboken, NJ : John Wiley & Sons.
- Chakraborty, S., *Power Electronics for Renewable and Distributed Energy Systems : A Sourcebook of Topologies, Control and Integration / edited by Sudipta Chakraborty, Marcelo G. Simões, William E. Kramer*, ed. M.G. Simões, W.E. Kramer, and SpringerLink. Vol. Chapter 3. 2013, London: London : Springer London : Imprint: Springer.
- Kopp, G. and J.L. Lean, *A new, lower value of total solar irradiance: Evidence and climate significance*. Geophysical Research Letters, 2011. **38**(1).
- Hansen, J., et al., *Target atmospheric CO<sub>2</sub>: Where should humanity aim?* arXiv preprint arXiv:0804.1126, 2008.
- Morjaria, M., et al., *A Grid-Friendly Plant: The Role of Utility-Scale Photovoltaic Plants in Grid Stability and Reliability*. Power and Energy Magazine, IEEE, 2014. **12**(3): p. 87-95.
- Manz, D., et al., *The Grid of the Future: Ten Trends That Will Shape the Grid Over the Next Decade*. Power and Energy Magazine, IEEE, 2014. **12**(3): p. 26-36.

10. Cai, T., S. Duan, and C. Chen. *Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement*. in *Power Electronics for Distributed Generation Systems (PEDG), 2010 2nd IEEE International Symposium on*. 2010.
11. Zhang, Y., et al., *RBF Neural Network and ANFIS-Based Short-Term Load Forecasting Approach in Real-Time Price Environment*. Power Systems, IEEE Transactions on, 2008. **23**(3): p. 853-858.
12. Barnes, A.K., J.C. Balda, and J.K. Hayes. *Modelling PV Clouding Effects Using a Semi-Markov Process with Application to Energy Storage*. in *World Congress*. 2014.
13. <http://www.bom.gov.au/climate/austmaps/solar-radiation-glossary.shtml>
14. Chenni, R., et al., *A detailed modeling method for photovoltaic cells*. Energy, 2007. **32**(9): p. 1724-1730.
15. Lonij, V.P.A., et al. *Forecasts of PV power output using power measurements of 80 residential PV installs*. in *Photovoltaic Specialists Conference (PVSC), 2012 38th IEEE*. 2012.
16. Soman, S.S., et al. *A review of wind power and wind speed forecasting methods with different time horizons*. in *North American Power Symposium (NAPS), 2010*. 2010.
17. Diagne, M., et al., *Review of solar irradiance forecasting methods and a proposition for small-scale insular grids*. Renewable and Sustainable Energy Reviews, 2013. **27**(0): p. 65-76.
18. Lawan, S.M., et al., *Different Models of Wind Speed Prediction; A Comprehensive Review*. International Journal of Scientific & Engineering Research, 2014. **5**(1): p. 1760-1768.
19. Espino, I. and M. Hernandez, *Nowcasting of Wind Speed using Vector Regression, Experiments with Time Series from Gran Canaria*. Renewable Energy and Power Quality Journal, 2011(N9).
20. <http://w1.weather.gov/glossary>
21. Hassanzadeh, M., M. Etezadi-Amoli, and M.S. Fadali. *Practical approach for sub-hourly and hourly prediction of PV power output*. in *North American Power Symposium (NAPS), 2010*. 2010.
22. Ciulla, G., V. Lo Brano, and E. Moreci, *Forecasting the Cell Temperature of PV Modules with an Adaptive System*. International Journal of Photoenergy, 2013. **2013**: p. 1-10.
23. Mellit, A. and A.M. Pavan, *A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy*. Solar Energy, 2010. **84**(5): p. 807-821.
24. Heinle, A., A. Macke, and A. Srivastav, *Automatic cloud classification of whole sky images*. Atmospheric Measurement Techniques, 2010. **3**(3): p. 557-567.
25. Kazantzidis, A., et al., *Cloud detection and classification with the use of whole-sky ground-based images*. Atmospheric Research, 2012. **113**: p. 80-88.
26. Jie, S., et al., *Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines*. Industry Applications, IEEE Transactions on, 2012. **48**(3): p. 1064-1069.
27. Yuehui, H., et al. *Comparative study of power forecasting methods for PV stations*. in *Power System Technology (POWERCON), International Conference on Power System Technology*. 2010.
28. Martins, F., E. Pereira, and R. Guarneri, *Solar Radiation Forecast Using Artificial Neural Networks*. International Journal of Energy Science, 2012. **2**(6): p. 217-217.
29. Chengrui, C. and D.C. Aliprantis, *Cumulus Cloud Shadow Model for Analysis of Power Systems With Photovoltaics*. Power Systems, IEEE Transactions on, 2013. **28**(4): p. 4496-4506.
30. <http://www.kyocerasolar.com/assets/001/5195.pdf>
31. Ishaque, K., Z. Salam, and H. Taheri, *Accurate MATLAB simulink PV system simulator based on a two-diode model*. Journal of Power Electronics, 2011. **11**(2): p. 179-187.
32. Evans, D., *Simplified method for predicting photovoltaic array output*. Solar energy, 1981. **27**(6): p. 555-560.
33. Jakhrani, A., et al., *Comparison of solar photovoltaic module temperature models*. World Applied Sciences Journal, 2011. **14**: p. 1-8.
34. Ross Jr, R. *Interface design considerations for terrestrial solar cell modules*. in *12th Photovoltaic Specialists Conference*. 1976.
35. Rauschenbach, H.S., *Solar cell array design handbook-The principles and technology of photovoltaic energy conversion*. 1980.
36. Risser, V. and M. Fuentes. *Linear regression analysis of flat-plate photovoltaic system performance data*. in *5th Photovoltaic Solar Energy Conference*. 1984.
37. Schott, T. *Operation temperatures of PV modules: a theoretical and experimental approach*. in *EC Photovoltaic solar energy conference*. 6. 1985.
38. Servant, J. *Calculation of the cell temperature for photovoltaic modules from climatic data*. in *Proceedings of the Ninth Biennial Congress of the ISES—Intersol*. 1985.
39. Ross Jr, R. and M. Smokler, *Flat-Plate Solar Array Project: Final Report: Volume 6, Engineering Sciences and Reliability*. 1986, Jet Propulsion Lab., Pasadena, CA (USA).
40. Lasnier, F., *Photovoltaic engineering handbook*. 1990: CRC Press.
41. Hove, T., *A method for predicting long-term average performance of photovoltaic systems*. Renewable Energy, 2000. **21**(2): p. 207-229.
42. Krauter, S.C., *Development of an integrated solar home system*. Solar energy materials and solar cells, 2004. **82**(1): p. 119-130.
43. Mondol, J.D., et al., *Long-term validated simulation of a building integrated photovoltaic system*. Solar Energy, 2005. **78**(2): p. 163-176.
44. Mondol, J.D., Y.G. Yohanis, and B. Norton, *Comparison of measured and predicted long term performance of grid a connected photovoltaic system*. Energy Conversion and Management, 2007. **48**(4): p. 1065-1080.
45. Mondol, J.D., Y.G. Yohanis, and B. Norton, *The effect of low insolation conditions and inverter oversizing on the long-term performance of a grid-connected photovoltaic system*. Progress in Photovoltaics: Research and Applications, 2007. **15**(4): p. 353-368.
46. Duffie, J.A., *William A. Beckman Solar Engineering of Thermal Processes*. 2006, JOHN WILEY & SONS, INC. NY.
47. *Solar electricity / edited by Tomas Markvart, with Klaus Bogus ... [et al.]*. 2nd ed.. ed, ed. K. Bogus and T. Markvart. 2000, Chichester: Chichester : John Wiley & Sons.
48. Tamizhmani, G., et al., *PHOTOVOLTAIC MODULE THERMAL/WIND PERFORMANCE: Long -Term Monitoring and Model Development For Energy Rating*. NCPV and Solar Program Review Meeting, 2003, 2003: p. 936 - 939.
49. Zieland, L., S. Rajakaruna, and D. Liyanage. *Modelling and performance assessment of large scale solar photovoltaic plants in rural Australia*. in *Power Engineering Conference (AUPEC), 2013 Australasian Universities*. 2013.
50. Niimura, T., et al. *Profiling residential PV output based on weekly*

- weather forecast for home energy management system. in *Power and Energy Society General Meeting, 2012 IEEE*. 2012.
51. Masters, G.M., *Renewable and efficient electric power systems / Gilbert M. Masters*. Vol. Chapter 8. 2004, Hoboken, NJ: Hoboken, NJ : John Wiley & Sons.
52. Annear, R.L. and S.A. Wells, *A comparison of five models for estimating clear-sky solar radiation*. Water resources research, 2007. **43**(10).
53. Reno, M.J., C.W. Hansen, and J.S. Stein, *Global horizontal irradiance clear sky models: Implementation and analysis*. SANDIA report SAND2012-2389, 2012.
54. Meeus, J.H., *Astronomical algorithms*. 1991: Willmann-Bell, Incorporated.
55. Reda, I. and A. Andreas, *Solar position algorithm for solar radiation applications*. Solar energy, 2004. **76**(5): p. 577-589.
56. <http://www.nrel.gov/midc/solpos/spa.html>
57. Paulescu, M., et al., *Weather modeling and forecasting of PV systems operation*. Vol. Chapter 2. 2012: Springer Science & Business Media.
58. Masters, G.M., *Renewable and efficient electric power systems*, in *Second Edition*. 2013, Hoboken, New Jersey : John Wiley & Sons Inc. p. 186-252.
59. Lorenz, E., et al., *Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems*. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 2009. **2**(1): p. 2-10.
60. Bacher, P., H. Madsen, and H.A. Nielsen, *Online short-term solar power forecasting*. Solar Energy, 2009. **83**(10): p. 1772-1783.
61. [http://oceanservice.noaa.gov/education/yos/resource/JetStream/synoptic/clouds\\_max.htm](http://oceanservice.noaa.gov/education/yos/resource/JetStream/synoptic/clouds_max.htm)
62. Raschke, R.A. and S.K. Cox, *The Determination of Cloud Optical Depth from Multiple Fields of View Pyrheliometric Measurements*. 1982: Department of Atmospheric Science, Colorado State University.
63. <http://isccp.giss.nasa.gov/products/fieldvars.html>
64. [http://www.crh.noaa.gov/lmk/?n=cloud\\_classification](http://www.crh.noaa.gov/lmk/?n=cloud_classification)
65. <http://www.srh.noaa.gov/srh/jetstream/clouds/cloudwise/types.html>
66. Great Britain. Meteorological, O., *Meteorological glossary / Meteorological Office compiled by D.H. McIntosh*. 5th ed., ed. D.H. McIntosh. 1972, London: London : H.M.S.O.
67. Lawrence, M.G., *The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications*. Bulletin of the American Meteorological Society, 2005. **86**(2): p. 225-233.
68. Pfister, G., et al., *Cloud coverage based on all-sky imaging and its impact on surface solar irradiance*. Journal of Applied Meteorology, 2003. **42**(10): p. 1421-1434.
69. Long, C.N., et al., *Retrieving cloud characteristics from ground-based daytime color all-sky images*. Journal of Atmospheric and Oceanic Technology, 2006. **23**(5): p. 633-652.
70. Chen, C., et al., *Online 24-h solar power forecasting based on weather type classification using artificial neural network*. Solar Energy, 2011. **85**(11): p. 2856-2870.
71. Adesso, P., et al., *MAP-MRF Cloud Detection Based on PHD Filtering*. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens., 2012. **5**(3): p. 919-929.
72. Wood-Bradley, P., J. Zapata, and J. Pye, *„Cloud tracking with optical flow for short-term solar forecasting “*. Solar Thermal Group, Australian National University, Canberra, Australia, 2012.
73. Linfoot, A. and R.J. Alliss, *J4. 2 A CLOUD DETECTION ALGORITHM APPLIED TO A WHOLE SKY IMAGER INSTRUMENT USING NEURAL NETWORKS*. 2008.
74. Chow, C.W., et al., *Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed*. Solar Energy, 2011. **85**(11): p. 2881-2893.
75. Gonzalez, Y., C. López, and E. Cuevas. *Automatic observation of cloudiness: Analysis of all-sky images*. in *WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation. Session*. 2012.
76. Jayadevan, V.T., et al. *Forecasting solar power intermittency using ground-based cloud imaging*. in *World Renewable Energy Forum 2012*. 2012.
77. [http://www.icem2011.org/presentations2011/4\\_Thursday/3C/1100\\_Heinemann.pdf](http://www.icem2011.org/presentations2011/4_Thursday/3C/1100_Heinemann.pdf)
78. Dambreville, R., et al., *Very short term forecasting of the Global Horizontal Irradiance using a spatio-temporal autoregressive model*. Renewable Energy, 2014. **72**(0): p. 291-300.
79. <http://www.eumetsat.int/website/home/Data/DataDelivery/EUMETCast/index.html#>
80. Lonij, V.P.A., et al., *Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors*. Solar Energy, 2013. **97**(0): p. 58-66.
81. Lonij, V.P., et al. *Analysis of 80 rooftop PV systems in the Tucson, AZ area*. in *Photovoltaic Specialists Conference (PVSC), 2012 38th IEEE*. 2012. IEEE.

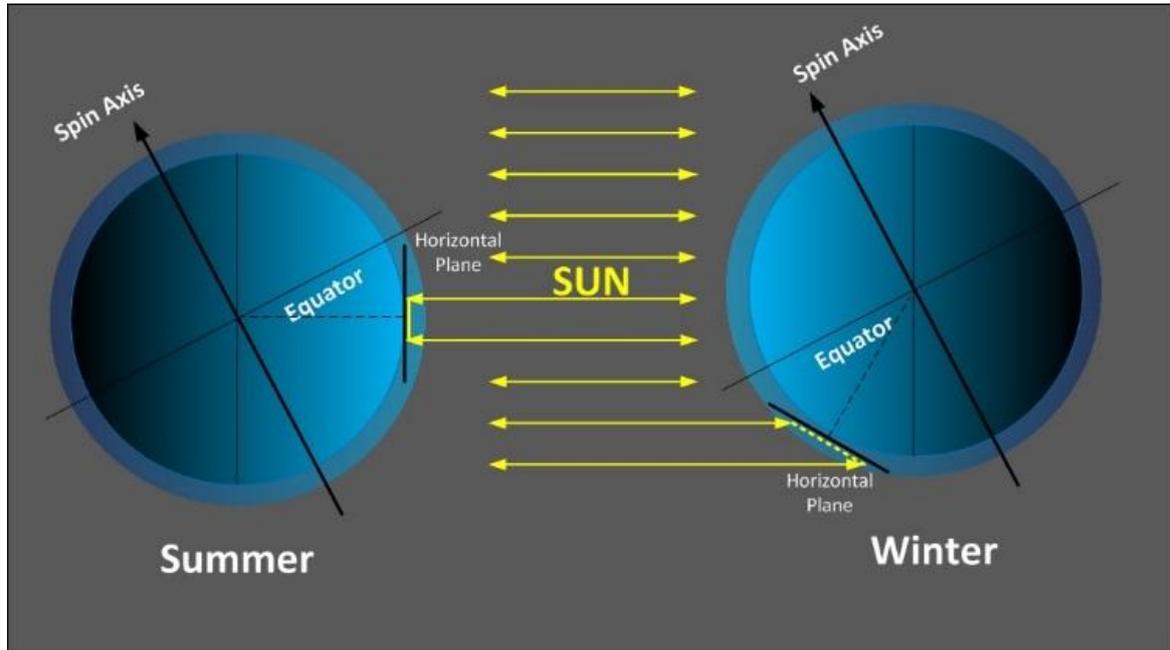


Fig 1 - Illustration of seasonal variations of global solar irradiance for a given place on the surface of the Earth.

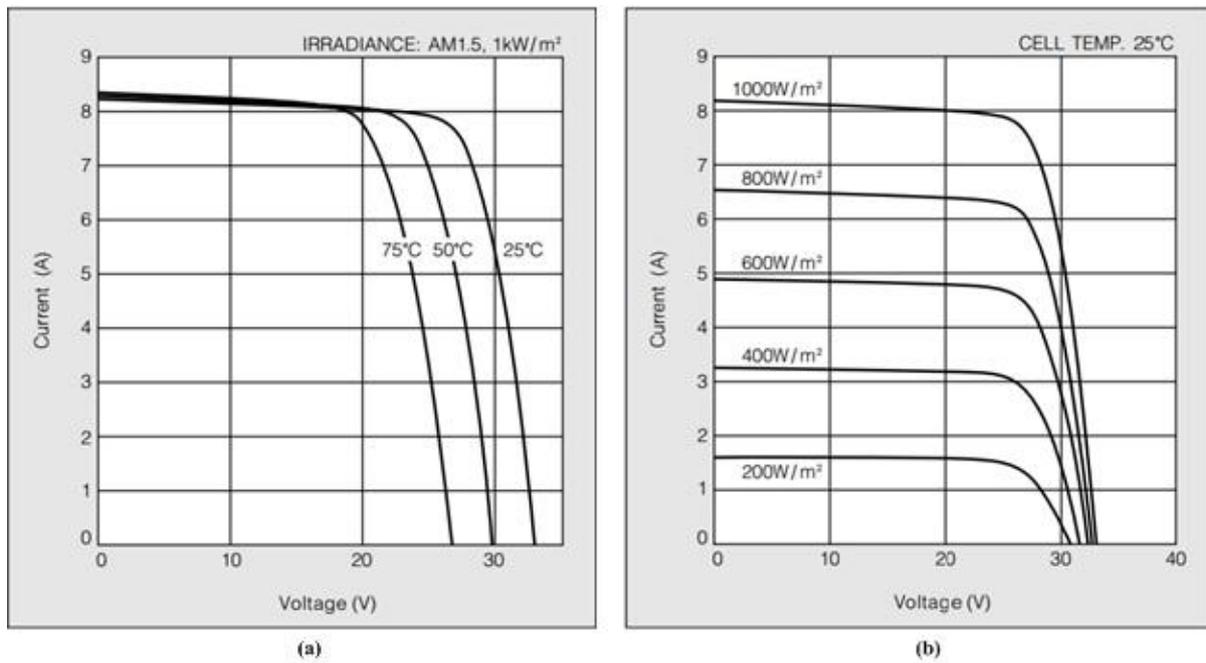


Fig. 2 - I-V curves of the module Kyocera KC200GT showing the impact of various temperature on the open circuit output voltage (a) and the effect of various solar radiance on the short-circuit output current (b) [30].

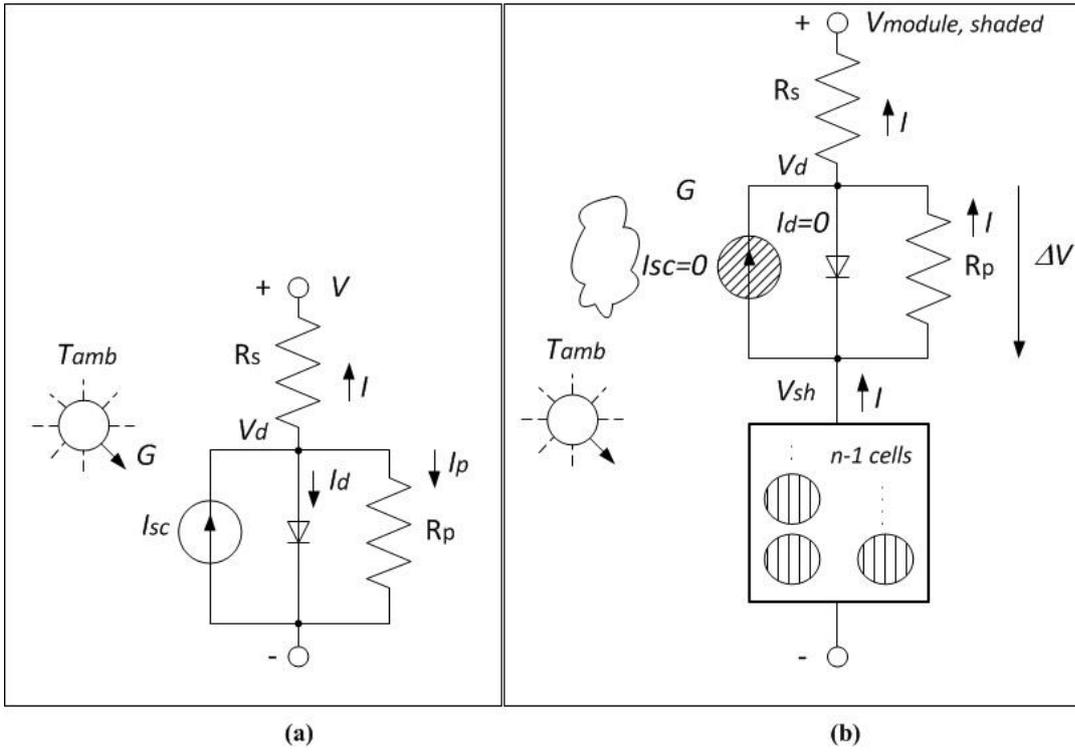


Fig. 3 - Two-diode model of a single PV cell (a) and of a module with partial shading of one cell among a module of  $n$  cells (b) [51].

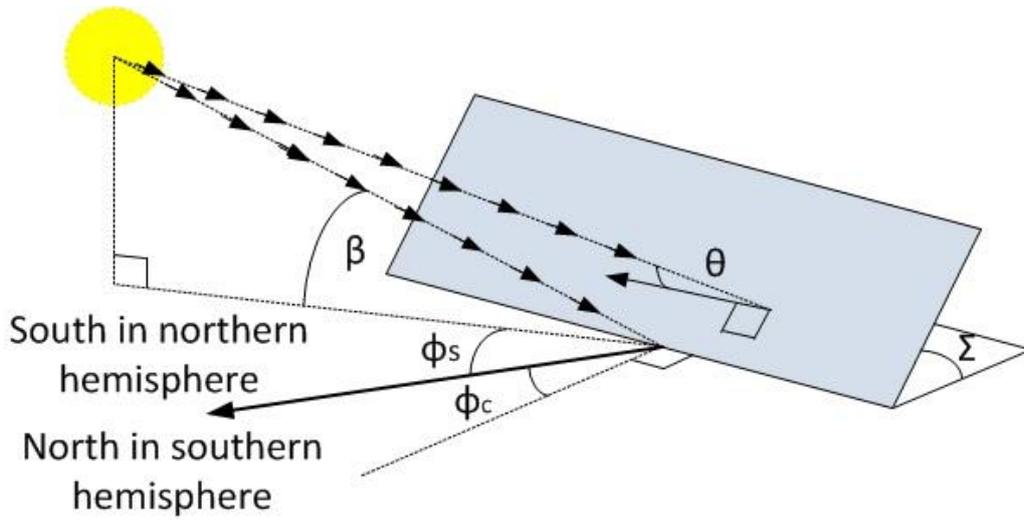


Fig. 4 – Angle of incidence on a PV Collector [4].

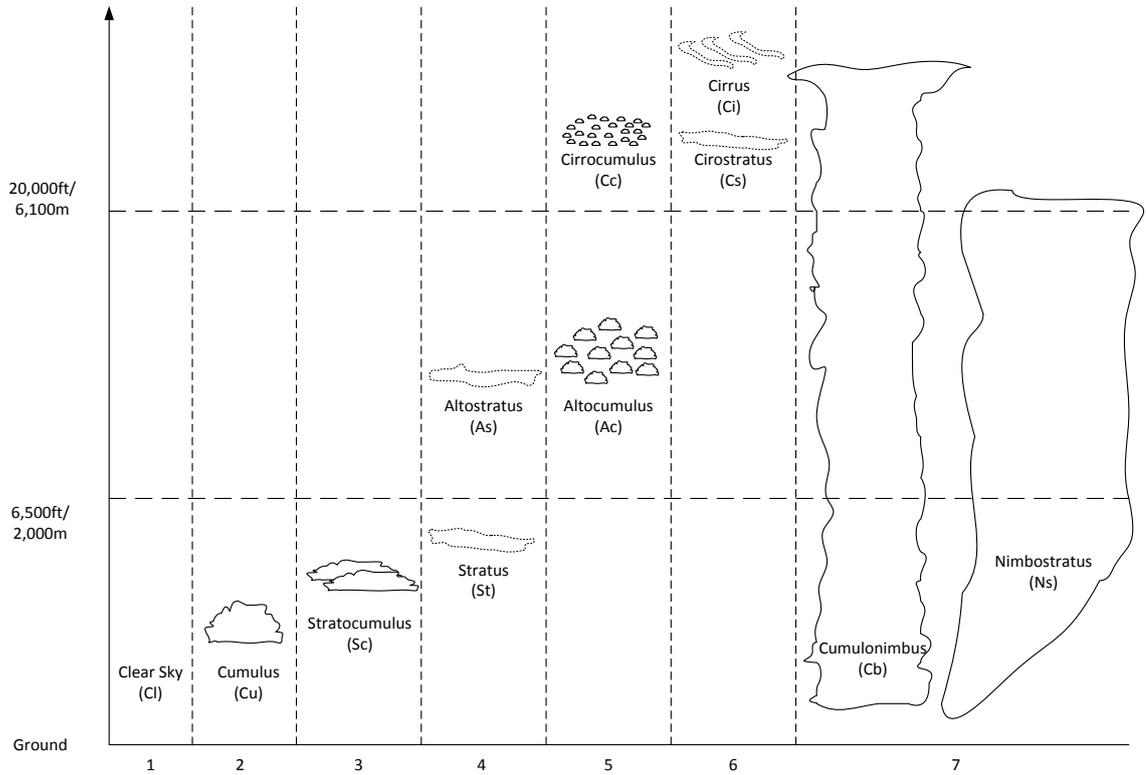


Fig. 5 – Illustration of the various types of clouds corresponding to Table 11.

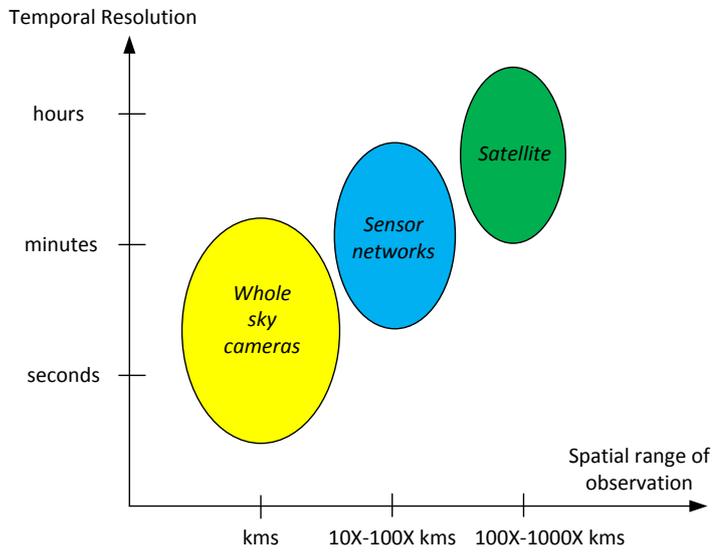


Fig. 6 – Temporal resolutions compared to the spatial range of observation for various types of input data.

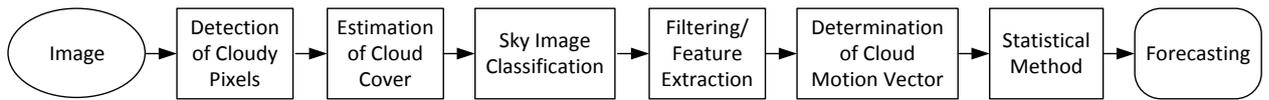


Fig. 7 – Typical Sky Imaging processing flow.

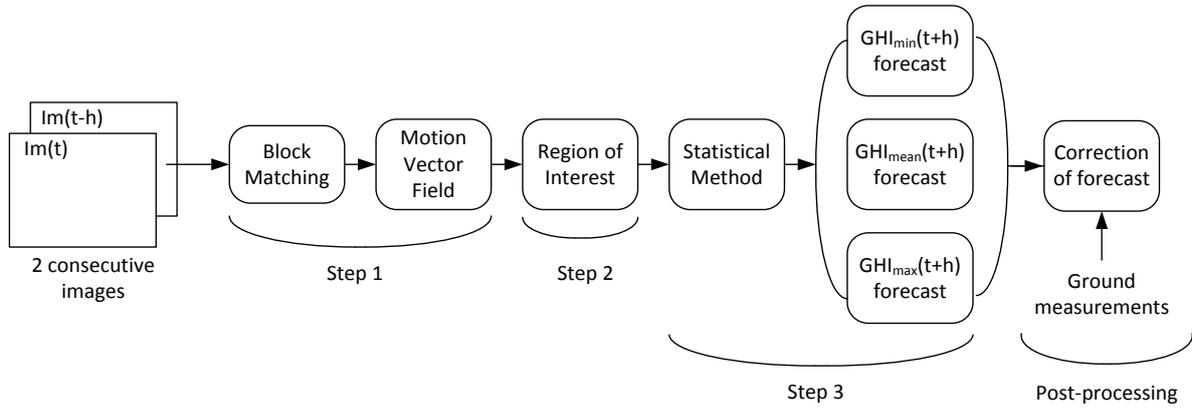
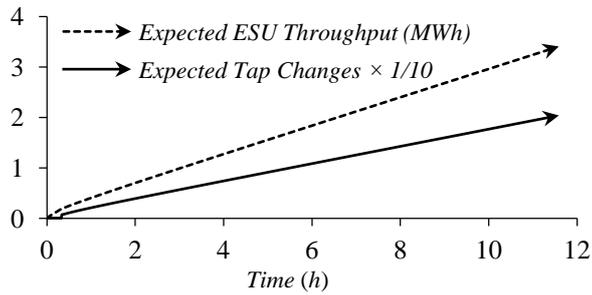


Fig. 8 – Satellite images processing flow.

(a)



(b)

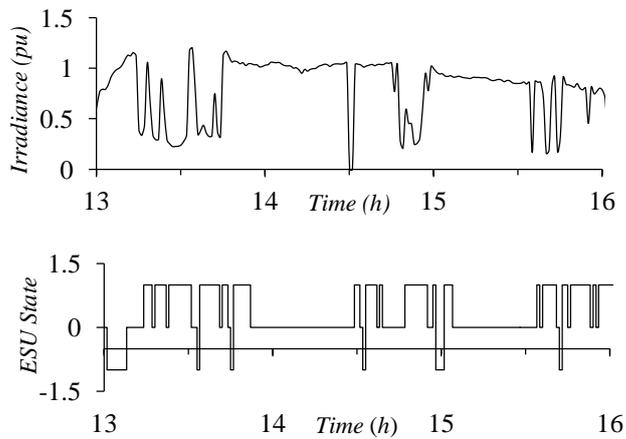


Fig. 9 – (a) Expected number of tap-changing operations and ESU throughput with respect to time. (b) Irradiance vs. time and corresponding ESU state vs. time. A state of +1 indicates discharging, -1 indicates charging, and 0 indicates standby [12].

**Table 1 - Wind speed/power forecasting time frames [16, 18]**

Time horizon	Range	Possible purpose
Very Short-term /Nowcasting	A few seconds to 30 minutes	Electricity market clearing
Short-term	30 minutes to 6 hours	Economic load dispatching
Medium-term	6 hours to 1 day	Operational security
Long-term	More than 1 day	Maintenance scheduling to optimize the operating cost

**Table 2 - Popular methods used for wind speed/power forecasting [16].**

Approach	Subclass/Tools	Examples	Characteristics
Persistence Method		$P(t+\Delta t)=P(t)$ if $k$ is small $\bar{v}(t + \Delta t) = \bar{v}(t)$	-Reference benchmark -Very accurate for very-short-term and short-term wind forecasting
Physical Approach/ Numerical Weather	Numeric Weather Prediction (NWP)	Global Forecasting System	Use of meteorological data. Accurate for long term wind forecasting.
Statistical Methods	Neural Network (NN)	Artificial Neural Network (ANN) Multilayer perceptron	Accurate for short-term
	Other techniques	Fuzzy Logic Genetic Algorithm Markov Chain SVR	Good for short-term
	Time-series Models	ARMA ARMAX ARIMA AR	Accurate for short-term
Hybrid Structures	- ANN+Fuzzy Logic = ANFIS - NWP+NN		-ANFIS is very good for very-short term forecast
			-NWP+NN is very accurate for medium and long term

**Table 3 - One possible definition of the various horizons for PV power forecasting [27].**

PV Horizon	Time resolution	Possible purpose
Very Short-term /Nowcasting	A few minutes to several hours	Real-time control
Short-term	Up to 3 days	Economic dispatch
Long-term	A week to a year	Grid balance

**Table 4 - Another definition of the various horizons for PV power forecasting [17].**

PV Horizon	Time resolution	Possible purpose
Intra-hour	15 minutes to 1 hours	Ramp rates control, variability management
Intra-day	1 to 6 hours	Load following
Day ahead	1 to 3 days	Transmission scheduling

**Table 5 - nRMSE expressed as a percentage for different forecasting methods. Data collected by Yuehui (2010) [27].**

Month/Season	Physical Model (NWP input)	Physical Model (measured G and T inputs)	Statistical Approach (NN)
March/Spring	11.48	8.78	11.05
June/Summer	11.54	5.17	10.45
October/Autumn	16.58	4.93	15.49
December/Winter	15.33	5.61	12.62
<b>Overall Year</b>	<b>12.45</b>	<b>5.51</b>	<b>10.5</b>

**Table 6 - Various empirical models available in literature to evaluate PV cell temperatures, as listed by Jakhriani et al. (2011) [32] and Ciulla et al. (2013) [22].**

Year	Authors	Empirical Model	Comments
1976	Ross [33]	$T_c = T_{amb} + k \cdot G$	$k = 0.02-0.04^\circ\text{C}\cdot\text{m}^2/\text{W}$
1980	Rauschenbach [34]	$T_c = T_{amb} + \frac{G}{G_{NOCT}} \cdot (T_{c,NOCT} - T_{amb,NOCT}) \cdot \left(1 - \frac{\eta_m}{\tau\alpha}\right)$	Considering wind speed $w_{speed} > 1\text{m/s}$ and a constant heat loss coefficient $U_L$
1983	Risser-Fuentes [35]	$T_c = 3.81 + 0.0282 \cdot 1.31 \cdot G \cdot T_{amb} - 165 \cdot W_{speed}$	Model verified with MPPT
1985	Schott [36]	$T_c = T_{amb} + 0.028 \cdot G - 1$	Model verified for $1\text{m/s} < V_w < 1.5\text{m/s}$ and $0^\circ\text{C} < T_{amb} < 35^\circ\text{C}$
1985	Servant [37]	$T_c = T_{amb} + \alpha \cdot G \cdot (1 + \beta \cdot T_{amb}) \cdot (1 - \gamma \cdot W_{speed}) \cdot (1 - 1.053 \cdot \eta_m)$	$A=0.0138, \beta=0.031, \gamma=0.042, T_{amb}$ is given in $^\circ\text{C}$ with a constant wind speed $V_w=1\text{m/s}$
1986	Ross-Smokler [38]	$T_c = T_{amb} + \frac{G}{G_{NOCT}} \cdot (T_{c,NOCT} - T_{NOCT})$	only valid for free standing modules with a constant wind speed $W_{speed}=1\text{m/s}$ and constant heat loss coefficient $U_L$
1990	Lasnier-Ang [39]	$T_c = 30.006 + 0.0175 \cdot (G - 300) + 1.14 \cdot (T_{amb} - 25)$	For p-Si only, $T_{amb}$ is given in $^\circ\text{C}$ whereas the wind speed and heat loss coefficient $U_L$ are not taken into account
2000	Hove [40]	$T_c = T_{amb} + \frac{G \cdot (\tau\alpha - \eta)}{U_L}$	$T_{amb}$ is given in $^\circ\text{C}$ , the ratio $\frac{\tau\alpha}{U_L}$ (with the transmittance $\tau$ , absorbance product $\alpha$ and heat loss coefficient $U_L$ ) is determined experimentally by assuming that the efficiency $\eta = 0$
2004	Krauter [41-43]	$T_c = T_{amb} + k \cdot G$ $T_c = T_{amb} + 0.028 \cdot G - 1$	$k=0.0058$ for lower modules, $0.012$ for upper modules and $0.03$ for usual modules
2005,2007	Mondol et al. [44]	$T_c = T_{amb} + 0.031 \cdot G - 0.058$	$T_{amb}$ is given in $^\circ\text{C}$ , $W_{speed} > 1\text{m/s}$ with constant heat loss coefficient $U_L$
2006	Duffie-Beckman [45]	$T_c = T_{amb} + \frac{G}{G_{NOCT}} \cdot \left( \frac{9.5}{5.7 + 3.8 \cdot W_{speed}} \right) \cdot (T_{c,NOCT} - T_{amb,NOCT}) \cdot (1 - \eta_m)$	The product $\tau\alpha$ of transmittance $\tau$ , absorbance product $\alpha$ is taken as $0.9$ whereas the heat loss coefficient $U_L$ is associated with wind speed
2007	Chenni et al. [14]	$T_c = 0.943 \cdot T_{amb} + 0.028 \cdot G - 1.528 \cdot W_{speed} + 4.3$	The coefficient of heat loss $U_L$ is not taken into account.

**Table 7 – Mean nominal conversion rate  $\eta_0$  by technology [49].**

Monojunction technology	Monocrystalline	Polycrystalline	Amorphous thin film
$\eta_0$ (%)	14.96	14.36	10.49

**Table 8 – Values recommended by the NOAA (2004) for the atmospheric refraction correction [17].**

$\beta$	Correction factor CF
$85^\circ$ to $90^\circ$	0
$5^\circ$ to $85^\circ$	$\frac{1^\circ}{3600} \left[ \frac{58.1''}{\tan(\beta)} - \frac{0.07''}{\tan^3(\beta)} + \frac{0.000086''}{\tan^5(\beta)} \right]$
$-0.575^\circ$ to $5^\circ$	$\frac{1^\circ}{3600} \left[ 1735'' - 518.2'' \beta + 103.4'' \beta^2 - 12.79'' \beta^3 + 0.711'' \beta^4 \right]$
$< -0.575^\circ$	$\frac{1^\circ}{3600} \left[ \frac{-20.774''}{\tan(\beta)} \right]$

**Table 9 – Possible sources of data.**

Type of forecast	Very short-term/Grid management			Utility Applications
<b>Approached based on</b>	Sky camera	Sensor network	Satellite	NWP
<b>Temporal resolution</b>	Seconds-minutes	Minutes	Minutes-hours	Day-days
<b>Comments</b>	Used to assess the cloud coverage and for cloud classification	As the only direct mean of measurement, often used as a reference to assess the precision of a forecasting approach	Used to assess the cloud coverage	Forecast of atmospheric conditions (ambient temperature, wind speed), useful for PV models

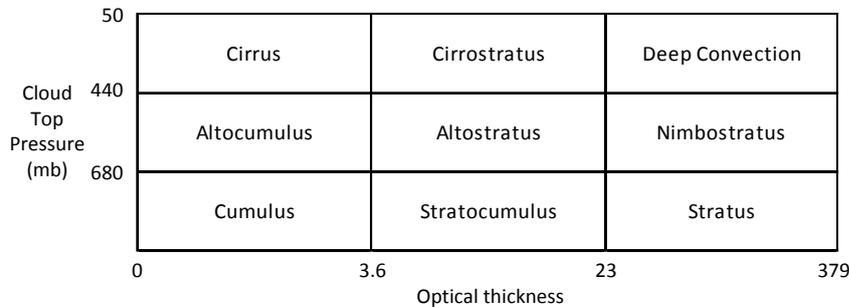
**Table 10 – Approximate heights of Cloud Levels.**

Level	Polar Region	Temperate Region	Tropical Region
<i>High Clouds</i>	10,000-25,000 feet (3-8 km)	16,500-40,000 Feet (5-13 km)	20,000-60,000 feet (6-18 km)
<i>Middle Clouds</i>	6,500-13,000 feet (2-4 km)	6,500-23,000 feet (2-7 km)	6,500-25,000 feet (2-8 km)
<i>Low Clouds</i>	Surface-6,500 feet (0-2 km)	Surface-6,500 feet (0-2 km)	Surface-6,500 feet (0-2 km)

**Table 11 – Cloud Classification for PV forecasting purposes.**

#	Cloud Classification	Level of the cloud	Altitude range of obstruction
1	Clear Sky (Cl)	Not applicable	Not applicable
2	Cumulus (Cu)	Low	Up to 6,500ft/2,000m
3	Stratocumulus (Sc)	Low	Up to 6,500ft/2,000m
4	Stratus-altostratus (St-As)	Middle - Low	Up to 20,000ft/6,100m
5	Cirrocumulus-altocumulus (Cc-Ac)	Middle - High	From 6,500ft/2,000m
6	Cirrus-cirrostratus (Ci-Cs)	High	Above 20,000ft/6,100m
7	Cumulonimbus-nimbostratus (Cb-Ns)	Low-middle-high	Very low base below 6,500ft / 2,000m and high top potentially above 20,000ft / 6,100m

**Table 12 – Cloud classification by altitude of top and optical thickness.**



**Table 13 – Relevant features to process images from sky cameras based on the works of Heinle (2010) [24] and Katzanidis (2012) [25].**

Type of Feature	Name	Expression	Comment
<b>Spectral</b>	Mean (R)	$M_R = \frac{1}{N} \cdot \sum_{i=0}^{N-1} p_{R,i}$	Useful to define the average tonal values
	Mean (B)	$M_B = \frac{1}{N} \cdot \sum_{i=0}^{N-1} p_{B,i}$	
	Mean (G)	$M_G = \frac{1}{N} \cdot \sum_{i=0}^{N-1} p_{G,i}$	
	Standard Deviation (B)	$SD_B = \sqrt{\frac{1}{N-1} \cdot \sum_{i=0}^{N-1} (p_{B,i} - M_B)^2}$	These features give some information about the tonal variations
	Skewness (B)	$SK_B = \frac{1}{N} \cdot \sum_{i=0}^{N-1} \left( \frac{p_{B,i} - M_B}{SD_B} \right)^3$	
	Difference (R-G)	$D_{RG} = M_R - M_G$	
	Difference (R-B)	$D_{RB} = M_R - M_B$	
	Difference (G-B)	$D_{GB} = M_G - M_B$	
<b>Textural</b>	Energy (B)	$EN_B = \sum_{a=0}^{G-1} \sum_{b=0}^{G-1} n_{\text{occ}[1,1]}(a,b)^2$	This shows the homogeneity of monochromatic level differences
	Entropy (B)	$ENT_B = \sum_{a=0}^{G-1} \sum_{b=0}^{G-1} n_{\text{occ}[1,1]}(a,b) \cdot \log_2(n_{\text{occ}[1,1]}(a,b))$	This measures how random the monochromatic level differences are
	Contrast (B)	$C_B = \sum_{a=0}^{G-1} \sum_{b=0}^{G-1} (a-b)^2 \cdot n_{\text{occ}[1,1]}(a,b)$	This measures local variations of monochromatic level differences
	Homogeneity (B)	$H_B = \sum_{a=0}^{G-1} \sum_{b=0}^{G-1} \frac{n_{\text{occ}[1,1]}(a,b)}{1 +  a-b }$	This measures how similar neighbor pixels are in terms of monochromatic levels.
<b>Total Cloud Cover</b>	Cloud Cover (R,G,B)	$CC = \frac{N_{\text{cloudy\_pixels}}}{N}$	This is a measure of the average cloud cover, useful to distinguish stratiform clouds

**Table 14 – Additional relevant image processing parameters [25].**

Criterion	Tier 1	Tier 2	Tier 3
Solar Zenith Angle $\phi_s$ (degrees)	<40	40-65	>65
Fraction of Cloud Coverage (octas)	<3	3-6	7-8
Visible Fraction of the Solar Disk (%)	<20	20-80	>80

**Table 15 – Additional sub-classes [25].**

Cloud Class	St-As	Cb-Ns	Cu	Cc-Ac	Sc	Ci-Cs	CI
Number of sub-classes	4	6	8	9	10	12	1

**Table 16 – Simple weather classification based on a textual forecast [48].**

<b>Key word</b>	<b>Classification</b>
Clear	Sunny
Fair	
Slightly overcast	
Heavily overcast	Cloudy
Cloudy	
Rain/Snow	Rainy
Fog/Mist	Foggy

### Highlights of this article:

- Very-short term photovoltaic (PV) power forecasting to manage grid stability and planning charges and discharges of an energy storage unit (ESU).
- Analysis of the tools and approaches used in the reference field of wind power forecasting. Introduction to the forecasting horizons and approaches that may similarly be applied to PV power forecasting.
- Review of some statistical methods that are relevant to PV forecasting.
- Case study comparing statistical and physical approaches.
- PV cell modelling including cell temperature modelling.
- Irradiance forecasting in clear sky conditions and considering the cloud coverage.
- Cloud coverage modelling and image processing for cloud classification.
- Weather classification method to forecast PV power output a day ahead.
- Comparison of the various sources of input data and methods for very short-term PV power output forecasting.