UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

MITIGATION CONTROL AGAINST PARTIAL SHADING EFFECTS IN LARGE-SCALE PHOTOVOLTAIC POWER PLANTS USING AN IMPROVED FORECASTING TECHNIQUE

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL ELÉCTRICA

CAROLINA DE LOS ÁNGELES MAYOL COTAPOS

PROFESOR GUÍA:

CLAUDIA RAHMANN ZÚÑIGA

MIEMBROS DE LA COMISIÓN:
JANNIK HAAS
RODRIGO PALMA BENHKE
FELIPE VALENCIA ARROYAVE

Este trabajo fue parcialmente financiado por el proyecto CONICYT/FONDAP/15110019 “Solar Energy Research Center” SERC-Chile y el Instituto de Sistemas Complejos de Ingeniería (ISCI).

SANTIAGO DE CHILE
2017

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA ING. MENCIÓN ELÉCTRICA, Y DE LA MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL ELÉCTRICA

POR: Carolina de los Ángeles Mayol Cotapos

FECHA: 01/2017

PROFESOR GUÍA: Claudia Rahmann Zúñiga

MITIGATION CONTROL AGAINST PARTIAL SHADING EFFECTS IN LARGE-SCALE PHOTOVOLTAIC POWER PLANTS USING AN IMPROVED FORECASTING TECHNIQUE

En un trabajo previo se propuso un control de mitigación de efecto nube que permitía disminuir los efectos nocivos de la nubosidad parcial sobre parques fotovoltaicos en la frecuencia de sistemas eléctricos de potencia. Esto último sin la necesidad del uso de acumuladores de energía. La estrategia se basa en la operación sub-óptima de los parques (operación en *deload*) con tal de disponer de reservas de potencia. A pesar que la implementación del sistema nombrado mejoró la frecuencia del sistema de forma significativa en comparación al caso base (sin el sistema de control), la operación en *deload* de los parques implica una gran cantidad de energía que no se está aprovechando, lo que no se consideró en la metodología. Con tal de mejorar esto, el siguiente trabajo propone un control de mitigación de efecto nube en parques fotovoltaicos de gran escala basado en una herramienta de pronóstico de radiación. Esto último permite disminuir las pérdidas de energía junto con mitigar los efectos de la nubosidad parcial, mediante la determinación de un nivel de *deload* en los parques fotovoltaicos usando dicho pronóstico.

En primer lugar, esta tesis presenta una revisión bibliográfica y discusión del estado del arte de las técnicas de pronóstico en parques fotovoltaicos. Se muestra que la selección de la técnica de pronóstico depende en la información disponible y la ventana de tiempo del pronóstico, es decir, dependerá del caso de estudio. Dicho esto, se propone el uso de una técnica de pronóstico basada en redes neuronales en el Sistema Interconectado del Norte Grande (SING) de Chile. El pronóstico sirve para determinar el nivel de *deload* en el parque fotovoltaico para los siguientes 10 minutos, en función de una rampa de radiación.

Los resultados muestran que la implementación de la técnica de pronóstico no solo mejora la respuesta en frecuencia del sistema, sino que también disminuye las pérdidas energéticas de forma significativa.

ABSTRACT OF THE THESIS TO OBTAIN THE: Professional degree of Electrical Engineer and the degree of Master of Science in Electrical Engineering

BY: Carolina de los Ángeles Mayol Cotapos

DATE: 01/2017

THESIS ADVISOR: Claudia Rahmann Zúñiga

MITIGATION CONTROL AGAINST PARTIAL SHADING EFFECTS IN LARGE-SCALE PHOTOVOLTAIC POWER PLANTS USING AN IMPROVED FORECASTING TECHNIQUE

A previous investigation proposed a control strategy to allow partially shaded photovoltaic power plants (PV-PPs) to mitigate the detrimental effects on the frequency of power systems without the need for energy storage. That strategy was based on a sub optimal operation of the plants (deloaded operation) in order to have power reserves available. Even though the system's frequency improved significantly with the inclusion of the aforementioned system, the deloaded operation implies energy losses in the power system, which was not considered in the methodology. To improve this, the following work proposes a mitigation control against partial shading effects in large PV-PPs based in on an irradiance forecasting tool. This allows to determine the deload level in the PV-PPs, to lower the energy losses in addition to mitigate the effects of partial shading.

First, this thesis presents a literature review and discussion of the state-of-the-art forecasting techniques in PV-PPs. It is shown that the selection of a forecasting technique depends on the available information and the timeframe of the forecast, and therefore the case study. Based on this, a forecasting technique based on artificial neural networks (ANNs) is proposed and implemented for the Northern Interconnected System (NIS) of Chile. The forecast is for determining the deload level in the PV-PPs for the next 10 minutes, based on the forecasted radiation ramp.

The results show that the implementation of the forecasting technique not only improves the system's frequency response, but also decreases the energy losses in a significant way.

Acknowledgements

En primer lugar quisiera agradecer a mi profesora guía, Claudia Rahmann, por su gran dedicación y apoyo en la realización de este trabajo. En este mismo punto también quiero agradecer a mis amigos del laboratorio de postgrado, con quienes aprendí mucho.

Por otra parte me gustaría agradecer a mis amigos OPAMPS, que hicieron de mi estadía en eléctrica un tiempo muy grato desde el primer curso de circuitos. También agradecer a mis amigos “con más experiencia” – Pau, Pipe, J, Yerko, Pudú, Inbox, Joshua, Manu, Smith – por la misma razón y por hacerme pasar por muchos procesos de aprendizaje.

Por último me gustaría agradecer a mi familia. Primero a mis papás por hacerme lo que soy hoy y apoyarme en todas mis cosas – por muy locas que fueran – desde muy temprano. A mis abuelos y tías por abrirme las puertas de su casa, donde siempre me sentí parte. Y, finalmente, a Felipe por darme tantas razones para seguir adelante, su apoyo y amistad.

Table of Contents

List of Tables

List of Figures

Chapter 1.

Introduction

Grid-connected photovoltaic (PV) power plants have been the renewable energy source with the highest annual growth rate in the world, reaching 30% in 2014 [1] and a total installed capacity of 177 GW by the end of the same year [2]. However, the inherent variability and uncertainty of PV generation, as well as the conventional operation at the maximum power point (MPP) of PV power plants (PV-PPs), necessarily leads to a reduction of power systems capability to deal with frequency deviations [3]. These are still the most relevant challenges in the operation and control of power systems with high penetration of PV systems [4].

The variability of the output power of PV-PPs has two main causes: 1) changes in the sun's position throughout the day and 2) partial or total shading due to moving clouds. The rising and setting of the sun regularly leads to 10-13% changes in PV output over a period of 15 minutes for single-axis tracking PV-PPs [5]. On the other hand, clouds are driven by stochastic processes and are largely responsible for rapid changes in the output power of PV-PPs that concern transmission system operators (TSOs). Changes in solar radiation in one location due to passing clouds can exceed 60% of the peak radiation in a matter of seconds [5].

Considering the day-ahead power system operation planning, the geometry of the sun is well known and thus easily managed by TSOs [3]. On the other hand, cloud-induced variability and irradiation uncertainty can be managed through: 1) increasing operating reserves in the power system or 2) incorporating new control strategies in PV-PPs to allow these technologies to participate in frequency regulation. Operating experience with large PV plants has demonstrated that large, rapid changes in the output of PV plants are possible [6]. This is the case of multi-MW PV plants in the Southwestern U.S. and southern California, where it has been reported that PV output can change more than 70% in a timeframe of 5 to 10 minutes [6], [7]. Depending on the system operating conditions and PV penetration, large changes as the described before could exhaust the ramping reserves available in conventional generation units, and thus threaten the power balance of the system. Furthermore, the efforts of conventional power plants to follow net load changes can lead to an increase in the cycling of synchronous generators, which could cause extra wear and tear on the generating equipment [3]. Several studies have reported that these frequency issues could constrain the integration of large-scale PV-PPs in power systems [8], [9]. These issues occur especially in systems with slow

secondary frequency response, with reduced capabilities for frequency response, or in low-inertia power systems. Secondary control is a combination of automatic generation control (AGC) and manual dispatch actions to maintain energy balance and scheduled frequency. Other possibilities can include the use of energy storage and the deloaded operation of renewables. The secondary frequency response covers a time frame of tens of seconds to tens of minutes, being the thermal units the slowest to respond due to the thermodynamics of those generation units.

Chile has an installed capacity of over 1.2 GW in PV technology [10], [11] and by 2017 it will rise to over than 2.5 GW [11]. This makes Chile a top 10 country in PV installed capacity worldwide [2]. On the other hand, considering the high solar potential in northern Chile, it is expected to have a significant amount of PV-PPs concentrated there. Due to the low inertia, limited ramp rates and slow reaction times of the available conventional units in the Northern Interconnected System (NIS) of Chile, this level of PV penetration makes of Chile a critical case study from a frequency perspective.

In [3] a mitigation control against partial shading effects (MICAPAS) for PV-PPs is proposed. This control is based on a deloaded operation of the PV arrays, meaning that the panels do not operate in the maximum power point (MPP). This control allows the PV-PP to deploy their active power reserves when shaded sections exist. However, the presented strategy uses a constant “deload level”, without any kind of economic consideration. The deloaded operation in PV panels means that part of the total energy is lost and so the PV-PPs owners lose money if there is no economic retribution for ancillary services. An improved control strategy would decide the deload level for each timeframe (1 hour, 10 min or 5 min depending on the power system) in a dynamic way, based on the forecast radiation levels and system conditions.

In order to improve the control response between maximizing PV energy production and minimizing the effects of the output power variability, this work proposes a control strategy to determine the PV-PPs deload level based on an improved forecasting technique.

1.1. Research hypothesis

It has been proven that a mitigation control against partial shading effects using a deloaded operation of PV-PPs improves the frequency response and performance of a power system. In this context, to use a forecasting technique to determine a dynamic deload in a PV-PP will not only improve the system’s performance but also minimize the energy losses.

1.2. Objectives

General objective

To design a mitigation control against partial shading effects in large-scale photovoltaic power plants based on an improved forecasting tool.

Specific objectives

1. Study and analyze the state-of-the-art in forecasting techniques for radiation and PV-PPs output power.
2. Develop a forecasting tool suitable for different type-days.
3. Implement the aforementioned forecasting technique and get the forecasted radiation data per PV-PP.
4. Define a deload function depending on the experimenter power ramps by the system.
5. Design and implement the mitigation control against partial shading effects considering a dynamic deload level.
6. Analyze the dynamic power system response under different scenarios and evaluate the proposed model performance.

1.3. Thesis structure

This thesis is structured into the following chapters:

- Chapter 2 presents a literature review and discussion of forecasting methods for grid-connected photovoltaic systems.
- Chapter 3 presents the structure of both control models: forecasting and for mitigation of partial shading.
- Chapter 4 presents the methodology of this work.
- Chapter 5 presents the case study used to test both forecast and control models.
- Chapter 6 presents and analyze the results.
- Chapter 7 contains a summary of this work, the final conclusions and also some references to possible evolutions and future work.

Chapter 2.

Forecasting Methods for Grid-Connected Photovoltaic Systems

In this section, some of the existing forecasting methods are reviewed. In general, there are two ways to forecast PV output power: indirect and direct forecasting methods [12]. For indirect forecasting methods, historic solar irradiance and weather information is used in order to predict first the solar radiation and then the associated output power. Direct forecasting methods use historical data of existing PV-PPs output power to predict future injections.

This chapter presents different forecasting methods for grid-connected PV systems. In order to get a general impression of all forecasting methods, wind power forecasting methods have been also included.

2.1. Uncertainty and variability of the output power in PV-PPs

Variability and uncertainty are inherent characteristics of power systems. Constant load and generation changes are events that may arise in a sudden way during the day, especially when there is a significant penetration of variable generation technologies. These changes must be compensated in order to ensure frequency regulation and therefore the power system's frequency stability.

Generation units change their power output to follow load changes throughout the day. To do this, they keep a specific amount of operating reserves to track the frequency grid changes.

The increase of variable generation in power systems will increase the variability and uncertainty that must be managed by TSOs and system planners [5]. Different studies have shown that using forecasts of variable generation and decreasing the time between the dispatches done by system operators during the day can greatly facilitate the integration of these technologies in power systems [13].

The output power of PV-PPs is variable because the sun changes its position throughout time. The rising and setting of the sun leads to 10-13% changes in PV output over a period of 15 minutes for single-axis tracking plants [5]. Moreover, the most important source of short-term variability of PV generation is the cloud-induced because it is given by a stochastic process [6] and it can change the peak insolation by more than 60% in a matter of seconds [5]. On the contrary, changes induced by the sun's geometry

are known as a deterministic process, and then they can be well estimated. Figure 2.1 shows the global radiation under cloudy conditions, and under a clear sky. It shows that under cloudy conditions, significant changes happen in a timescale of minutes. In order to attend the partial shading problem in PV-PPs, the implemented forecast has to work in a time scale of seconds to minutes (and not more). The chosen time scale will depend on the different characteristics of the available data and the forecasting errors that result from this data.

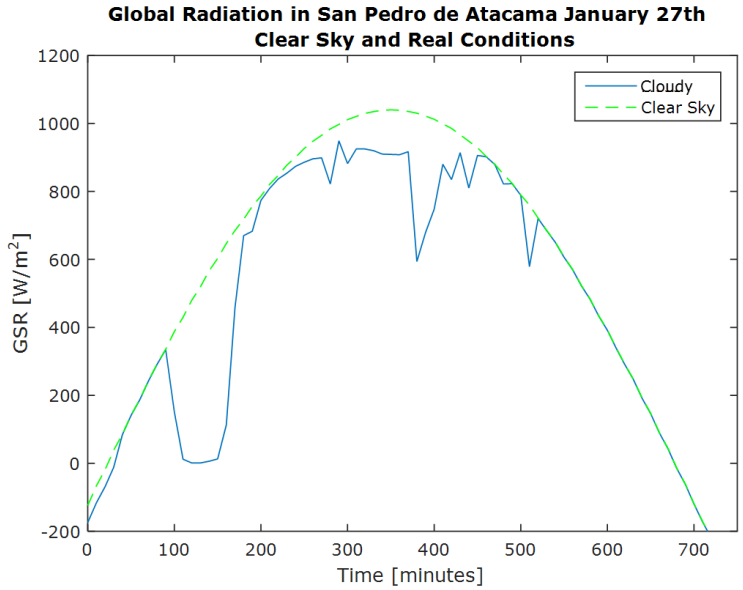


Figure 2.1: Global solar radiation under clear sky and real conditions.

2.2. Forecasting methods in PV-PPs

This section presents a review of the most relevant forecasting techniques used in PV-PPs nowadays: 1) numerical methods [12], [14], [15], 2) artificial neural network (ANN) based methods [16], [17], [18], 3) auto-regressive models [19], [20], [21]; and 4) probabilistic techniques that lead to synthetic series [22], [23], [24].

Numerical methods

Numerical weather predictions (NWPs) are weather predictions based on mathematical models, given by several mesoscale simulation systems. They are physically-based weather forecasts and are mostly used for local weather forecasting in regions or cities [15]. The temporal development of the atmosphere's state is modeled by differential equations that describe the physical laws governing the weather [25].These kind of models have the disadvantage of a coarse spatial resolution. The spatial resolutions are usually between 5 [15], 12.5 [26] and 25 km^2 grids [14]. The temporal resolution (between 3-6 hours) is also too wide for the purposes of this work.

Post-processing methods have been used in order to 1) reduce systematic errors; 2) account for local effects; and 3) to calculate other variables such as PV power forecast. Most of the post-processing methods that are used to forecast PV power output have temporal interpolation and spatial averaging [14]. There are also physical post-processing, introducing new input parameters for better results [15], [18].

In [26], an extensive analysis of the North American Mesoscale (NAM) forecast in California is presented. Unlike [23] and [24], a cloud probability is added to the NAM forecast using satellite imagery. These results were compared to measured GHI in stations all around California. For clear sky conditions, i.e. in cases where the clear-sky index k_t was over 0.8, the forecast interval contained 89.5% of the measurements. For very cloudy conditions ($k_t < 0.6$), the forecast contained between 25% and 80% of the measurements. Similar percentages were contained by Lorenz *et al.* in [14], but with wider forecast intervals.

In [15], the poor performance of NWP in cloudy days is improved by adding the water path[1] participation in the linear function that describes solar irradiance in that research.

[1] The water path is a measure of the total amount of liquid water present between two points in the atmosphere.

ANN-based methods

One effective method for PV forecasting are ANN-based methods [16], [17], [18], [27]. ANNs are computational models inspired on human neurons. In contrast to conventional mathematical logic, the main characteristics of human thinking process is imprecise, fuzzy, but adaptive [28]. ANNs learn from examples by constructing an input-output mapping without explicit derivation of the model equations.

An ANN consists of many interconnected artificial neurons that interact with each other in a certain way. Neurons are simple processing units, and each connection has a weight factor [18]. Every neuron in the network sums its weighted inputs and compares this sum with a threshold to produce an internal activity level a_i. This process is shown in equation (1). Here, w_{ij} is the weight associated to the connection between input j to neuron i, x_{ij} is the input signal number j to neuron i, and w_{i0} is the weight associated to the connection between an input clamped at -1 to neuron i [18], or the threshold value. The weights of the connections are adjusted during the training process in order to get the desired input/output relation of the neural network (NN).

$$a_i = \sum_{j=1}^{n} w_{ij} \cdot x_{ij} - w_{i0} \tag{1}$$

The internal activity is passed through a nonlinear function φ to produce the output y_i of the neuron (equation 2).

$$y_i = \varphi(a_i) \tag{2}$$

Several nonlinear functions can be used in order to produce the neuron's output, being the most used ones the sigmoid functions [16], [18], [29]. Sigmoid functions are S-shaped functions, as shown in equation (3).

$$\varphi(a_i) = \frac{1}{1 + e^{-a_i/p}} \tag{3}$$

In equation (3), a_i is the activation into neuron i and p is a number that controls the shape of the curve, usually set to 1. Higher values of p give a flatter response curve and lower values of p produce a steeper curve (Figure 2.2).

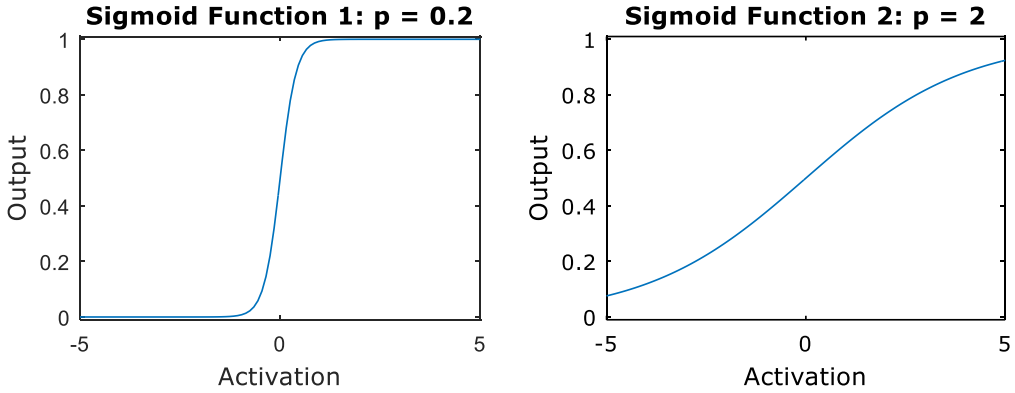


Figure 2.2: Different sigmoid response curves.

There are two different network models: feedforward and feedback (or recurrent) networks [28]. The first ones consist of a network where signals travel only in one way: from input to output. Feedback networks, on the other hand, have signals traveling in both directions by introducing loops within the network as applied in [30]. In PV power forecast and most of the applications, however, multi-layer feedforward networks are normally used [16], [17], [18].

Multi-layer feedforward networks consist of an input layer composed by single layer of neurons following an arbitrary number of hidden layers, each one composed by a specific number of neurons. The input signal propagates throughout the network in a forward direction. An example of a three-layer feedforward neural network is shown in Figure 2.3.

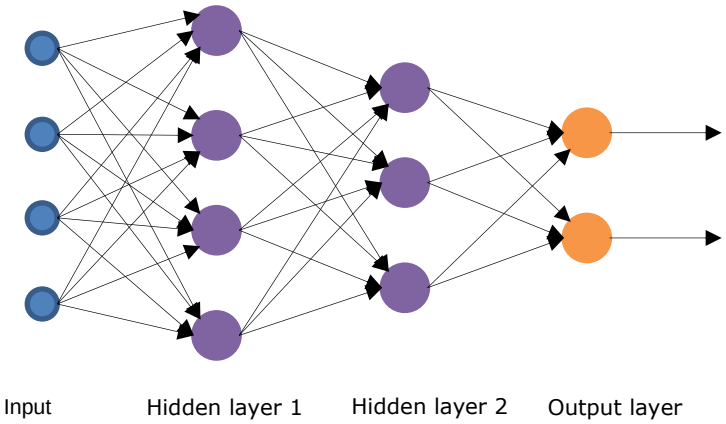


Figure 2.3: Three-layer feedforward neural network.

There can be any number of hidden layers within a feed-forward network. However, this number is not an issue because of the universal approximation theorem. This theorem states that a feedforward network with a single hidden layer can represent any function, given the right parameters [28]. This behavior is confirmed by empirical results [29]. The number of hidden neurons, on the other hand, is determined by the complexity of the

problem and the size of the training data [27]. Most of the research done until now determines this number by a heuristic process [16], [17], [18], [27].

In a given high dimensional input-output dataset, neural networks are able to provide a promising modelling service. The dimension of the dataset plus the number of hidden neurons determine the generalization capacity and the accuracy of the neural network. An ANN that is too small will fail to formulate a precise mapping. A network that is too big, on the other hand, may trigger an overfitting problem that tends to deteriorate the generalization ability of the model [27]. This problem can be solved by dividing the dataset and cross-validating the results.

The next step is to choose an algorithm in order to estimate the weights. There are various methods to estimate the weights in an ANN. The most used method in forecasting applications is the backpropagation method [4], [16], [17], [21], [27]. Another method is the Neuro-Evolution of Augmenting Topologies (NEAT), which consists in a genetic algorithm that changes the weights and number of nodes of the network simultaneously [31].

Autoregressive methods

Autoregression is a regression of a variable against itself, i.e. the forecast of a given variable is done by using a linear combination of past values of the same variable. Hence, an autoregressive model is a model that estimates the variable in time t given past values of the same variable in times $t-1$, $t-2$, etc.

It is important to note that for these kind of models, the process y_t has to be *stationary*. Loosely speaking, a stochastic process is stationary if its statistical properties do not change with time. Even knowing that the PV output power is a nonstationary process, it can be made stationary with various techniques such as normalization or post-processing [20], [32],[33], [34].

An autoregressive model of order p can be written as [35]:

$$y_t = c + \phi_1 \cdot y_{t-1} + \phi_2 \cdot y_{t-2} + \cdots + \phi_p \cdot y_{t-p} + e_t \quad (4)$$

where c is a constant, e_t is white noise and ϕ_x are fixed constants. This model is called an $AR(p)$ model. This kind of models are used generally for identification, but they are also useful for time series development.

In [20], an AR model is used in order to predict a future time series of irradiance. The model uses past samples of global horizontal irradiance (GHI) and forecasts hourly irradiance values. The GHI is normalized by using clear-sky index in order to create a clearer stationary time series. The method uses 15-min observations of solar power resolution and was used to predict hourly values of solar power.

Exogenous inputs such as a forecast predictions $\hat{y}_{t|t-1}^{forecast}$, ambient temperature or relative humidity, for example, can be added to an AR model. These kind of models are called ARX (autoregressive with exogenous input) and may provide a better estimation for future values since it incorporates more information. The results vary depending on the correlation between the exogenous input variables and the forecasted variable. An example of this is shown in [19], where GHI and satellite data is combined and used as input in an $AR(5)$ model. The model uses satellite data as an exogenous input because it adds more information concerning cloud coverage, and therefore predicts cloudy days with more precision. In [20], the AR model is enhanced by using a NWP forecast as an exogenous input, creating an ARX model.

An ARX model can be written as follows:

$$y_t = c + \phi_1 \cdot y_{t-1} + \phi_2 \cdot y_{t-2} + \cdots + \phi_p \cdot y_{t-p} + K_1 \cdot \hat{y}_{t|t-1}^{forecast} + e_t \quad (5)$$

Rather than using past values of the forecast variable in a regression, a moving average model (MA) uses past forecast errors (f_e) in a regression-like model [36]:

$$y_t = c' + e_t + \theta_1 \cdot f_{e_{t-1}} + \theta_2 \cdot f_{e_{t-2}} + \cdots + \theta_q \cdot f_{e_{t-q}} \quad (6)$$

where e_t is white noise and θ_x are unknown coefficients (being $x = 1,2,\ldots,p$). This model is referred as a $MA(q)$ model. Here, y_t can be understood as a weighted moving average of the past forecast errors.

When combining the MA model with the derivative AR, an ARIMA model is obtained. ARIMA stands for autoregressive integrated moving average model. The full model can be written as [37]:

$$y'_t = \bar{c} + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t \quad (7)$$

where y'_t is the differenced series. This model is said to be an $ARIMA(p,d,q)$ model or process, where p is the order of the AR part, d is a non-negative integer defining the differential part and q is the order of the MA part. This equation can also be written as [38], [34]:

$$(1-B)^d Y_t = \mu + \frac{\theta(B)}{\phi(B)} e_t \quad (8)$$

It is worth noting that many of the discussed models in the literature are special cases of the ARIMA model. For instance, an $ARIMA(p,d,q)$ with $d = 0$ is an $ARMA(p,q)$ model. An ARMA model is stationary for appropriate ϕ and θ. An ARMAX model is an ARMA model with exogenous input. In solar forecasting applications the exogenous input could be a NWP-based forecast as in [20], precipitation amount, humidity, insolation [34], among others.

Many of these models are already implemented in *MATLAB.* Hence, their application is straightforward having consideration the limitations, like stationarity.

Synthetic series

Synthetic series are used to make statistical data based on observations. This kind of series are important for PV power forecasting because usually 1-hour series are available while the resolution of partial shading in PV-PPs in the order of minutes, thus the creation of synthetic series can help to broad the resolution when measured data is missing.

In [22], a second order Markov model is proposed. Markov models are representations of stochastic processes. A Markov process X_t with a set of m allowed states is said to be in state j at time t if $X_t = j$. In a first order, given that the process is in state i at time $t-1$, the probability that it will be in state j at time t is given by a fixed probability P_{ij} written mathematically as:

$$P_{ij} = P(X_t = j \,|X_{t-1} = i, X_{t-2} = i_{t-2}, \dots, X_0 = i_0) = P(X_t = j \mid X_{t-1} = i) \qquad (9)$$

P_{ij} is known as the transition probability from state i to j, and is independent of the states of the process at other times. In a second order Markov process, the transition not only depends on $t-1$ but also on $t-2$. The second order transition matrix is constructed as follows:

$$P^2 = \begin{bmatrix} P_{111} & P_{112} & \dots & P_{11m} \\ P_{121} & P_{122} & \dots & P_{12m} \\ \vdots & \vdots & \ddots & \vdots \\ P_{211} & P_{212} & \dots & P_{21m} \\ P_{221} & P_{222} & \dots & P_{22m} \\ \vdots & \vdots & \ddots & \vdots \\ P_{mm1} & P_{mm2} & \dots & P_{mmm} \end{bmatrix}$$

Another way to generate synthetic series is shown in [23]. First, a cluster depending on atmospheric conditions is determined by cloud conditions using a modified clear sky index. Then, the synthetic irradiation is split into two parts: a deterministic component and a stochastic component, as shown in equation (10).

$$I^i_{10m} = I^i_{10m\_i3} + sign(r) \cdot A \qquad (10)$$

The deterministic component $I^i_{10m\_i3}$ is determined by the mean value on a defined time frame. Then, for the stochastic component, random numbers are generated from an uniform distribution curve [0,1]. The inverse beta value corresponding to that probability and sky condition is calculated and multiplied by the maximum standard deviation to

generate A. Finally, r is defined by random numbers from a normal distribution curve with zero mean and unit standard [23].

Kernel density estimation method

Since there is little research on probabilistic active power forecasting of large-scale PV-PPs, wind power forecasting methods have also been reviewed. In [39], a Kernel density estimation (KDE) is proposed.

A probability density function (pdf) of a random variable Y can be estimated knowing that the explanatory variable X is equal to x. This is shown in equation (11) where $x_{t+k\,|\,t}$ are the explanatory variables for time $t+k$ that exist in time t, p_{t+k} is the wind power forecasted at $t+k$, $f_{P,X}$ is the multivariable density function and f_X is the pdf of X.

$$\hat{f}_P(p_{t+k} \mid X = x_{t+k\,|\,t}) = \frac{f_{P,X}\left(p_{t+k}, x_{t+k\,|\,t}\right)}{f_X(x_{t+k\,|\,t})} \tag{11}$$

The pdf of X has to be estimated, and it is done by a kernel density estimation given by

$$\hat{f}_X(x) = \frac{1}{N \cdot h}\sum_{i=1}^{N} K\left(\frac{x - X_i}{h}\right) \tag{12}$$

where N is the number of samples, K is a kernel function and h is the bandwidth parameter.

Kernels functions are compared in order to choose the most useful one. This is the most important step in the method, because it determines the accuracy. The explanatory variables used in this case are lineal combinations of wind speed, direction, hour of the day and other measured values.

2.3. Discussion and proposal

Different solar forecasting methods were presented in this section. Although there is a lot of research in this area, there is still much to improve in the forecasting process of intra-hour solar irradiance.

NWP methods allows to model regional irradiance forecast accurately, being their most accurate spatial resolution within the range of kms (1 to 100 kms). Their most accurate temporal resolution is between 1h to days and thus these kind of models are the most attractive ones for long-term forecasting. On the other hand, these forecast techniques are less accurate under cloudy conditions, even after post-processing, and therefore NWP methods are usually used as an input of other empirical methods to improve intra-hour forecasting.

ANN methods offers a nonlinear approximated model for power output, independent of physical characteristics such as tilt angle. These models have successful modeled the PV output power within intra-hour time horizons, and so it works well for this work purposes. Other advantage of ANN is that they are easy to manipulate online once trained offline. The drawback of NN-based models is that they have to be re-trained as long as the natural conditions evolve (as it happens with global warming).

Autoregressive linear methods have the advantage of generating forecast time series in a simple way. They are statistical models with relevant researches and works in different time frames. It only depends of the available information as in ANN. The disadvantage is that one important condition of autoregressive models is the stationary characteristic of the forecasted variable, and therefore there is a need to pre-process the used data in order to fulfil the model requirements.

Synthetic series methods have the capacity of accurately describing irradiance behavior when given sufficient data. Hence, the correct construction of Markov series in a 10 minute range, for example, depends directly on the availability of data in a 10 minutes timeframe. The same happens with other way to generate this kind of series. If synthetic series are generated, a key issue is to determine a clear classification and clustering of data. The drawback of this probabilistic forecasting (and any other as NN-based methods or autoregressive) is the availability of datasets in such a small timeframe. On the other hand, the analysis requires a lot of computational processing.

A summary-diagram of the spatial and temporal resolutions for each reviewed forecasting technique is shown in Figure 2.4. It is worth noting that every forecasting technique can be used for other temporal or spatial resolutions. However, the resolution shown in the figure are those for which the best results have been obtained.

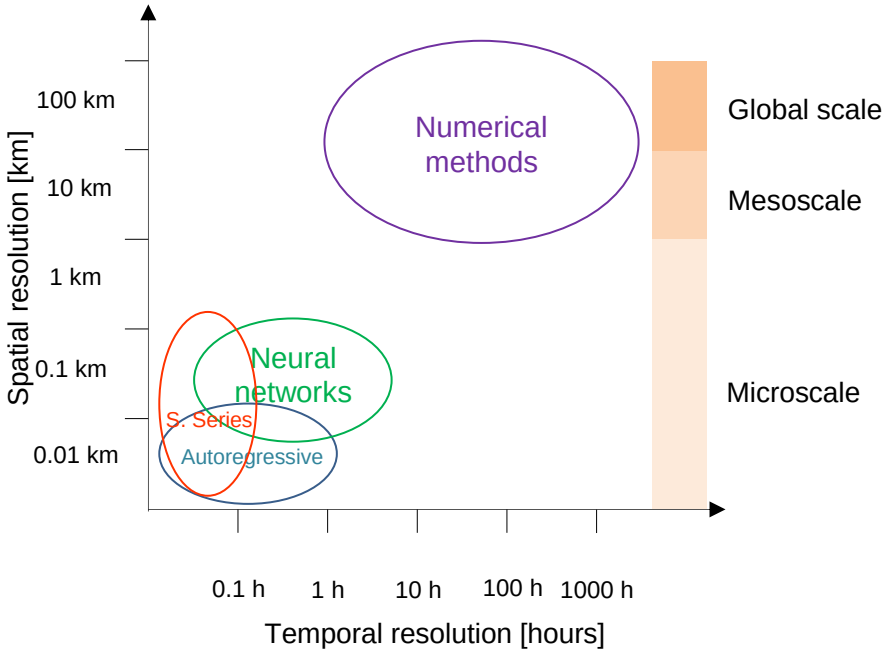


Figure 2.4: Summary-diagram of the spatial and temporal resolutions for each reviewed forecasting technique.

Kernel estimation is good for short-term forecasts. It has the advantage of not assuming a probability density function a priori, because it is determined depending on the given explanatory variables. The problem can get more complex because of the use of a big number of explanatory variables.

The available meteorological data for this work includes information for every 10 minutes. It contains temperature, relative humidity, global, direct and diffuse solar radiation and wind speed. On the other hand, previous work has shown that different type-days (in terms of cloudiness) have dissimilar forecasting results and behavior, and usually cloudy days have higher forecasting errors. Consequently, this work proposes a two-stage, NN-based forecast, being the first one a classification stage to determine the kind of day of the next day, and the second stage forecasts the radiation every ten minutes. The kind of day is determined with a clear-sky index, which is described in Appendix A.

Chapter 3.

Models

This work consists on the implementation of two local models at PV-PP level: forecasting and control. The forecasting model is based on 4 NN: 1 NN to forecast the kind of day and 3 NN to forecast intra-hour radiation. The control model is based on 3 stages: the first one calculates the PV-PP deload level, the second one is a centralized PV-PP control which determines the available reserves of the power plant and then delivers a signal to the third part, which is a control scheme that modifies the power output per section accordingly.

Figure 3.1 shows the Northern Interconnected System (NIS) of Chile and a zoom into the local control of a generic PV-PP. The bottom of the figure shows a schematic of the proposed two stage model and each of the main functions per local model.

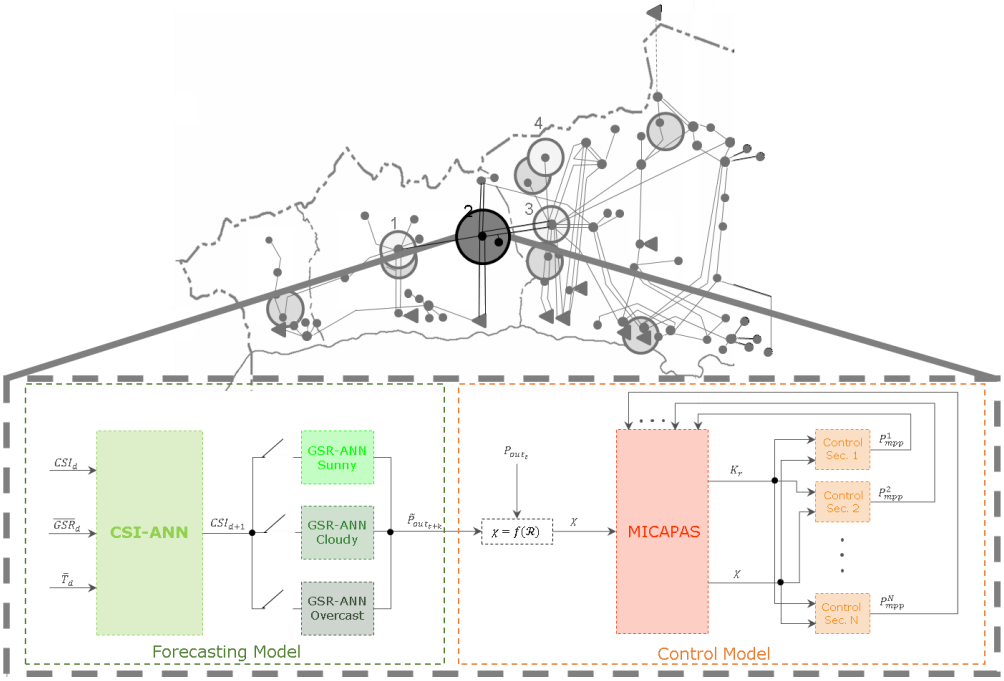


Figure 3.1: Diagram of the work-related models.

3.1. Forecasting model

Various research have shown that to forecast radiation during cloudy days is more challenging than other type of days. Due to the stochastic nature of passing clouds, forecast accuracy decreases strongly with forecast sampling time [32]. As the main focus of this work are partial shading conditions, counting with the best forecast results during cloudy days is crucial.

Even though different types of climate – tropical, arid, mediterranean, continental or polar – have different trends, it can be stated that different kind of days – sunny, cloudy and overcast – exhibit similar behavior [40]. Many researchers have concluded that a classification stage prior to the radiation forecast provides better results than just using the same forecasting technique for every kind of day ([12], [17], [21], [41]). Because of this, it is proposed to split the data into different types of day depending on their CSI, and train a different ANN per type of day. This improves the network's response to each kind of day, because the overall behavior of each day is more similar and then the network training is more straight-forward. The drawback of using this method is that more data is needed in order to get good results with each network. Hence, the forecasting model of this work includes a classification stage, where the type of day of the next day is forecasted to then choose between three neural networks, one for each kind of day: sunny, cloudy and overcast. The implemented model uses global solar radiation (GSR) as input because the GSR data is easier to get than the PV-PP output power. Nevertheless, if PV-PP power output data is available, then it is recommended to use this information because of the noise generated by dust, wind, and other conditions that make the GSR/Power output conversion less accurate.

The forecasting model is split into two stages. The first stage forecasts the type of day of the next day and selects the pertinent ANN. The second sub-stage uses an ANN (depending on the type of day) in order to forecast the solar radiation and power output every 10-min during a certain day. The ANN used to forecast the next day's type of day is called CSI-ANN, because it selects the kind of day by its clear-sky index (CSI). The ANNs that are used to forecast radiation within the day, every 10 minutes, are called GSR-ANN (where GSR stands for Global Solar Radiation). They are different depending on the type of day: 1) Sunny, 2) Cloudy and 3) Overcast.

A general forecasting model scheme is shown illustratively in Figure 3.2 where, based on the output of the CSI-ANN, one of the three GSR-ANN is activated. It shows that the kind of day of the next day (day $d + 1$) is determined by the CSI-ANN, which has 3 inputs: CSI of day d, mean GSR of day d and mean temperature of day d.

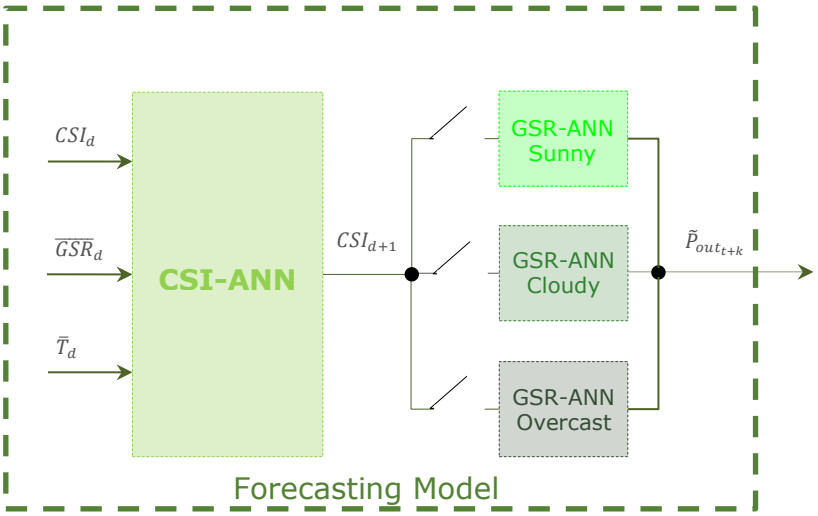


Figure 3.2: Schematic diagram of forecasting model.

The clear-sky index (CSI) is a good way to classify the day since it represents a comparison between the actual solar radiation and the one under clear-sky conditions (i.e. without clouds). This means that the classification depends mostly on the amount of clouds in the region.

The used CSI is calculated as shown in equations (13) and (14). Here, the current global radiation (I stands for irradiance) is compared with the extraterrestrial radiation. For more information about the calculation of CSI refer to Appendix A.

$$K_t = \frac{I_{global}}{I_{ext}} \tag{13}$$

$$I_{ext} = I_0 \cdot \left[1 + 0.033 \cdot \cos\left(\frac{2\pi}{365} \cdot DOY\right)\right] \cdot \cos(\theta_t) \tag{14}$$

Considering the classification done by Larrañeta in [23], where the type of days are given by CSI intervals, the days can be organized into 8 different types. Each type of day is determined by the thickness of the clouds and their fluctuations during the day and can be identified as one of 5 sky conditions (totally covered, mostly covered, partly covered, mostly clear and totally clear).Owing to the usual lack of data to train the NN, a classification of 8 type of days seems overwhelming. Furthermore, it is possible to reduce the 5 sky conditions proposed in [23] into three different groups: Sunny, Cloudy and Overcast. The proposed classification and description is shown in Table 3.1.

The CSI-ANN is trained with all the available data and the GSR-ANN are trained with a dataset that corresponds to the respective kind of day.

Table 3.1: Classification and description by clear-sky index.

TYPE OF DAY	DESCRIPTION	CSI
SUNNY	MOSTLY OR TOTALLY CLEAR DAY. NO CLOUDS WITH OR WITHOUT TURBIDITY.	(0.6,1]
CLOUDY	MOSTLY COVERED DAY WITH GREAT FLUCTUATIONS OR THIN CLOUDS WITH FLUCTUATIONS.	(0.3,0.6]
OVERCAST	TOTALLY COVERED DAYS AND MOSTLY COVERED DAYS WITH SOME FLUCTUATIONS.	[0,0.3]

CSI-ANN

Once the data is classified by type of day, a correlation analysis is done in order to determine the ANN inputs to get the CSI for the next day. The objective of the CSI-ANN is to forecast the CSI of the next day (CSI_{d+1}). The purpose of the correlation analysis is to study the strength of that relationship with available statistical data (Table 3.2).

Table 3.2: Available data for the CSI-ANN construction.

AVAILABLE DATA	NOMENCLATURE
CSI OF DAY d	CSI_d
MEAN TEMPERATURE OF DAY d	$\bar{T}_d$
MEAN GLOBAL SOLAR RADIATION OF DAY d	$\overline{GSR}_d$
DAY OF THE YEAR OF DAY d	DoY_d
MEAN RELATIVE HUMIDITY OF DAY d	$\overline{RH}_d$

For the correlation analysis, the Spearman rank correlation is used because this kind of correlation does not make assumptions about the distribution of the data, as opposed to the Pearson correlation, and it captures de monotonic relationship between the random variable [42], which is relevant while working with sudden changes in solar radiation. Figure 3.3 shows the results of the Spearman correlation between CSI_{d+1} versus $\bar{T}_d$, $\overline{GSR}_d$, DoY_{d+1} and $\overline{RH}_d$. It is shown that the $\bar{T}_d$ and $\overline{GSR}_d$ are the most correlated variables with CSI_{d+1}, and then are good choices for inputs in the CSI-ANN.

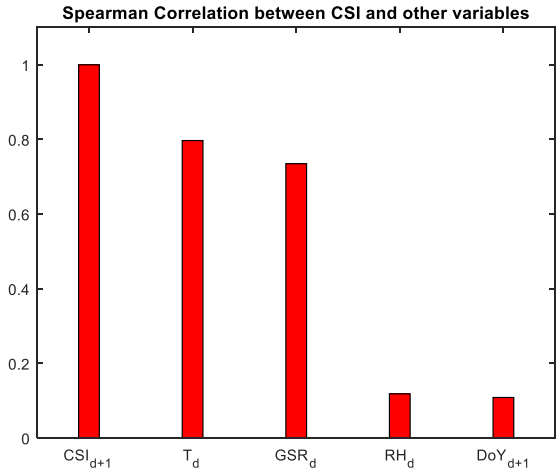


Figure 3.3: Spearman rank correlation between CSI and other variables.

Finally, with the aim of getting a more accurate forecast model, it is proposed to add autoregressive factors in the CSI-ANN input. To get the functional lags, the partial autocorrelation of the clear sky index is calculated. The partial autocorrelation, is the correlation between two variables after being adjusted for a common factor that may be affecting them, such as stationarity. The autocorrelation function is not calculated because it measures stationarity [43] and, as it was explained before, solar radiation is a stationary process and then the autocorrelation function would reflect the usual rising and setting of the sun as a correlation.

The partial autocorrelation function is shown in Figure 3.4. Even though the first 8 lags are relevant statistically speaking, the first lag is used because it is significantly more relevant than the other ones, reaching a 0.97 correlation compared to a 0.2 of the third.

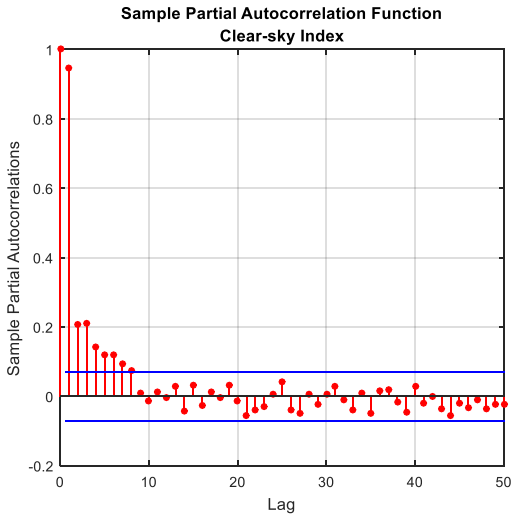


Figure 3.4: Clear-sky index sample partial autocorrelation function.

Once the correlation analysis is done, an ANN is trained using Lavenberg-Marquardt backpropagation with the Neural Network Toolbox of MATLAB. Lavenberg-Marquardt is used because of its training velocity. Cross-validation is not necessary at this stage because, first, three subsets are used (training, validation and test) and, second, it serves as a classification neural network and so the CSI does not have to be too accurate, but to be near the most probable day. When dividing the data into three subsets, each one serves one purpose. The training set is used to fit the models, the validation set is used to estimate the prediction error for model selection and the test set is used for assessment of the generalization error of the chosen model [44]. The sets are selected in order to have a proportional type of day representation. Therefore, the data is selected without breaking days apart and having the same (or similar) proportion of types of days.

Figure 3.5 shows the performance of the designed ANN data, i.e., training, validation and test, and the overall performance. The best overall performance is given by the ANN of 8 hidden nodes and so is the one chosen.

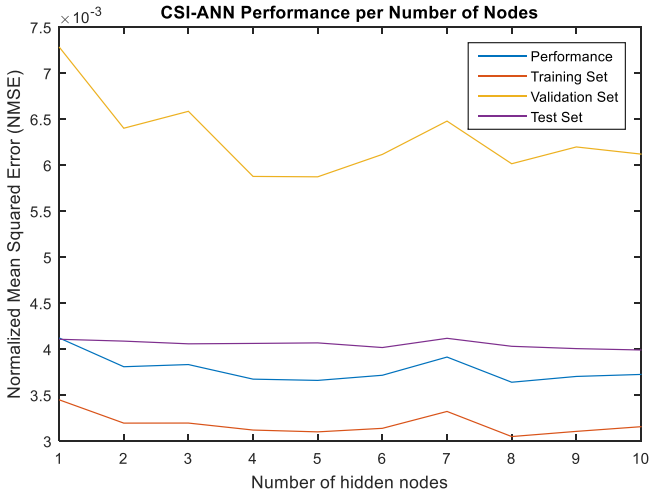


Figure 3.5: CSI ANN performance in the training, validation and test sets.

In Figure 3.6, a schematic of the CSI-ANN is shown. This includes the 3 inputs (CSI_d, $\bar{T}_d$ and $\overline{GSR}_d$), the 8 nodes in the hidden layer chosen previously, and the output (CSI_{d+1}).

When the whole data is tested in terms of the quantity of days that are not well forecasted (for example, a cloudy day is forecasted as a sunny day) the error is equal to an 8.76%. This means that a wrong GSR-ANN is going to be used during almost 32 days a year. This is why it is important that all three GSR-ANNs have a good generalization capability. As a consequence, the error can be compensated by the GSR-ANN.

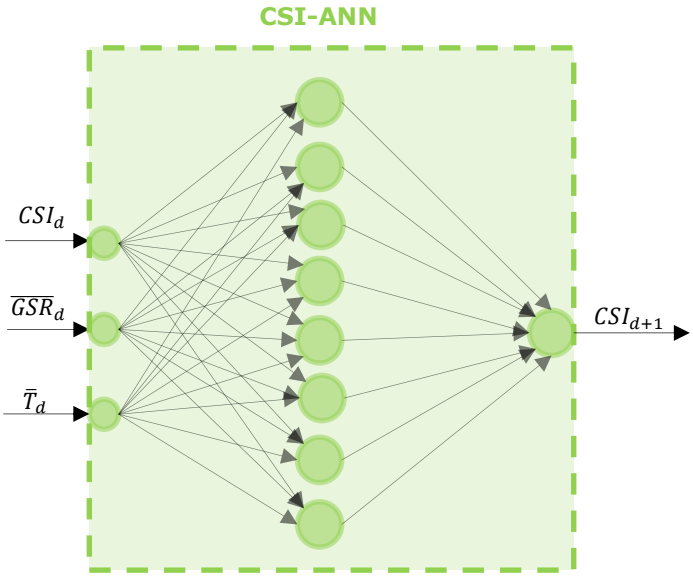


Figure 3.6: CSI-ANN diagram.

GSR-ANN

In the second forecasting stage, three different neural networks have to be designed. To do this, the overall data is divided into three datasets: 1) Sunny, 2) Cloudy and 3) Overcast, which are then used to train each different neural network. To do this, the CSI of every day of the available data is calculated as shown in Appendix A, and the information of each kind of day is arranged into three datasets.

The objective of the GSR-ANN is to forecast the global solar radiation of the next 10 minutes (GSR_{t+10}). In this case, the available data for the correlation analysis is the one shown in Table 3.3.

Table 3.3: Available data for the GSR-ANN construction.

AVAILABLE DATA	NOMENCLATURE
GSR MEASURED IN TIME t	GSR_t
TEMPERATURE MEASURED IN TIME t	T_t
GLOBAL SOLAR RADIATION MEASURED IN TIME t	GSR_t
DAY OF THE YEAR OF ACTUAL DAY (d)	DoY_d
RELATIVE HUMIDITY MEASURED IN TIME t	RH_t

A new correlation analysis has to be done to determine the inputs in these new neural networks. To do this, the available data shown above is compared with the GSR_{t+10}. Figure 3.7 shows the Spearman Rank Correlation between GSR and other variables in sunny days and the partial autocorrelation function of GSR. It can be seen that the GSR

at time $t + 10$ is slightly correlated with the temperature and inversely correlated with the relative humidity (RH) at time t. It does not correlate with the day of the year and so this variable is not to be considered as an input.

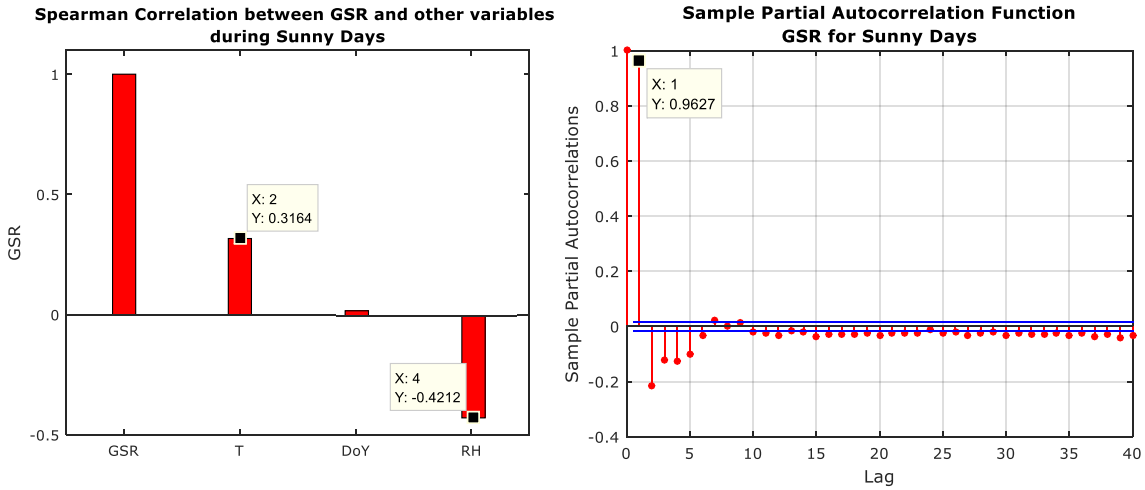


Figure 3.7: Spearman Rank Correlation between GSR and other variables (left) and sample partial autocorrelation (right) for sunny days.

Figure 3.8 shows the Spearman rank correlation between GSR and other variables under cloudy and overcast conditions on the left, and the partial autocorrelations functions on the right. It is shown that overcast days have a similar performance to sunny days but with a lower GSR/T correlation and higher GSR/RH correlation. The GSR/RH correlation is lower than GSR/T only in the case of cloudy days. This shows that cloudy days have different behavior than overcast or sunny days.

The performed analysis shows that in every kind of day there is a very significant partial autocorrelation between GSR in time t and GSR in time $t + 10\,min$. Thus, every ANN should consider at least one lag. As the second lag is not significant comparing it to the first one, only one lag is considered.

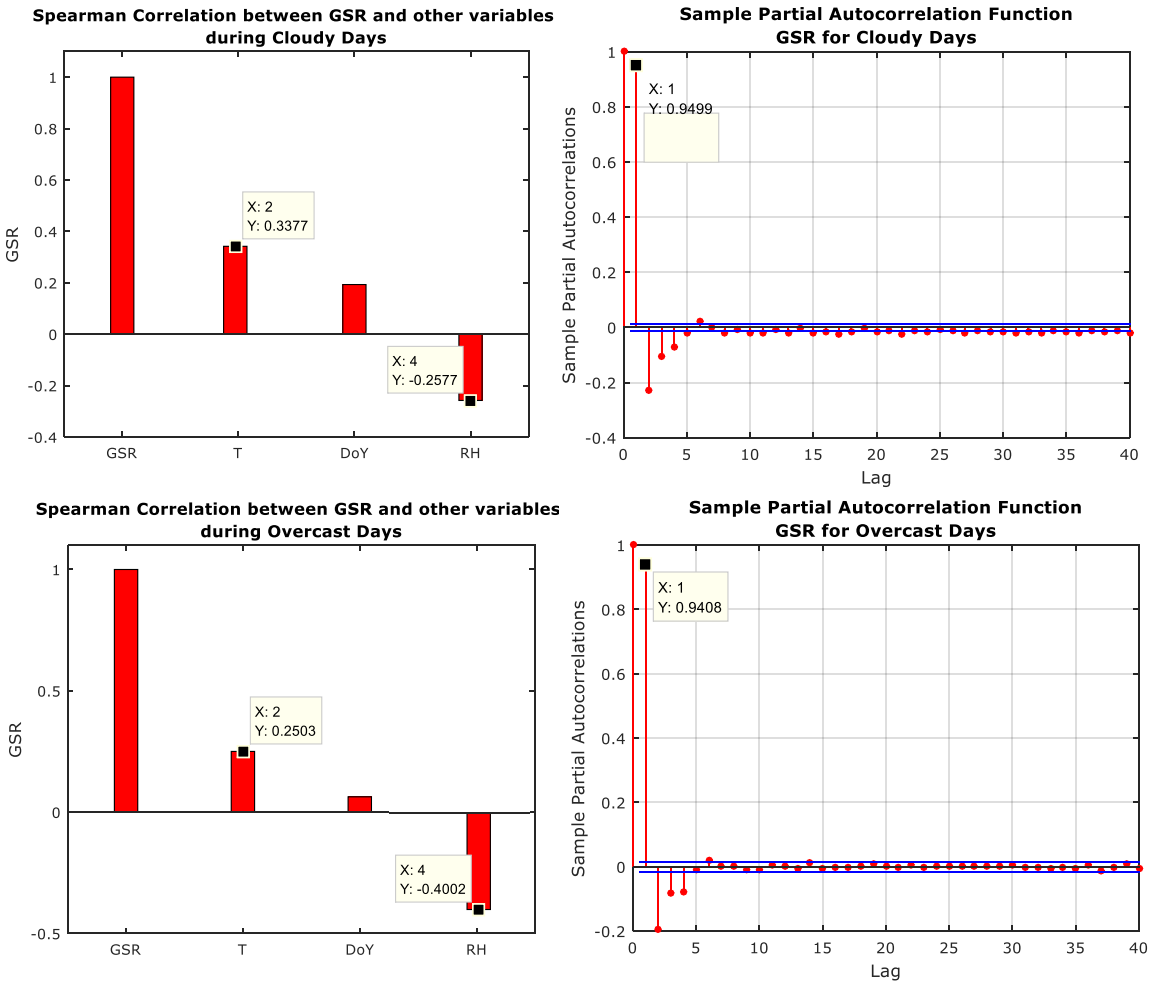


Figure 3.8: Spearman rank correlation between GSR and other variables (left) and sample partial autocorrelation (right). Cloudy days on top and overcast days in the bottom.

The correlation analysis shows that the best way to design a neural network for GSR is considering three inputs:

1) Forecasted mean temperature of the next lag (10-min ahead)
2) Forecasted mean relative humidity of the next lag (10-min ahead)
3) Actual GSR

For each of the three neural networks, the Bayesian Regularization algorithm is used for training, because the forecasting of GSR is more complex than the forecast of CSI, and then the computational complexity of Levenberg-Marquardt is too heavy to train this network [45].

The next step for the GSR ANN construction is to determine the number of nodes in the hidden layer. The number of nodes can be determined by heuristic, genetic or other

optimization algorithm processes. Because of the complexity of the problem and the data quantity, a heuristic approach is implemented. The process starts with one neuron/node in the hidden layer and then increases iteratively until reaching 16 nodes in the hidden layer. A maximum of 16 is chosen because neural networks in these applications usually use less than that [17], [24], [30].

A 10-fold cross-validation is done for every number of nodes, in order to avoid overfitting in the neural network and have a better generalization capability in the final model. An example of a cross-validation process considering three sets (training, validation and test) and 5 is shown in Figure 3.9. In that figure, each circle represents radiation, relative humidity and temperature time-series data, and the data is split in three sets – training, validation and test – in 5 different ways. In the case of this work, the data is split in 10 different ways (folds).

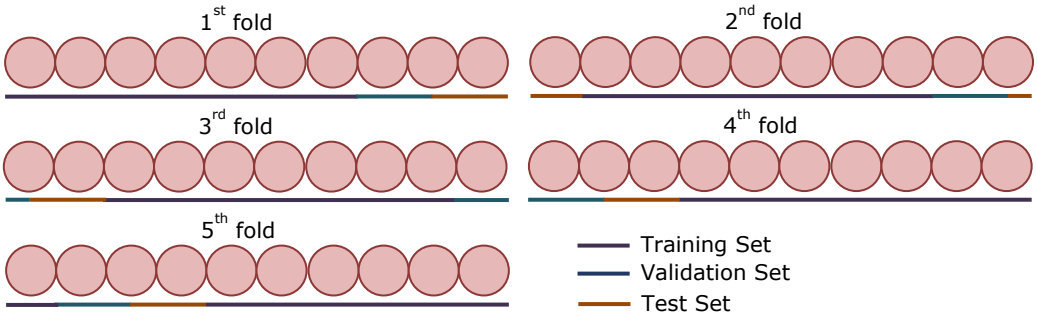


Figure 3.9: 5-fold cross-validation process with two sets: training and test.

Once the data is splitted, each fold is used to train one neural network. Again, an iterative process is used to train the ANN and determine the quantity of nodes in the hidden layer. In order to determine the model that best fits in each kind of day, the mean and standard deviation of the training and test data performances (given by the NMSE) are analized to find the best alternative. This is shown in Figure 3.10 for every kind of day. In each type of day, the number of nodes is determined by the lowest mean NMSE and standard deviation.

In overcast days (above of Figure 3.10), the best alternative is given by 11 nodes considering mean and standard deviation of the NMSE. The best alternative for cloudy days is given by 15 nodes and for Sunny days is 9 nodes, because these cases have the lower mean and standard deviation of NMSE.

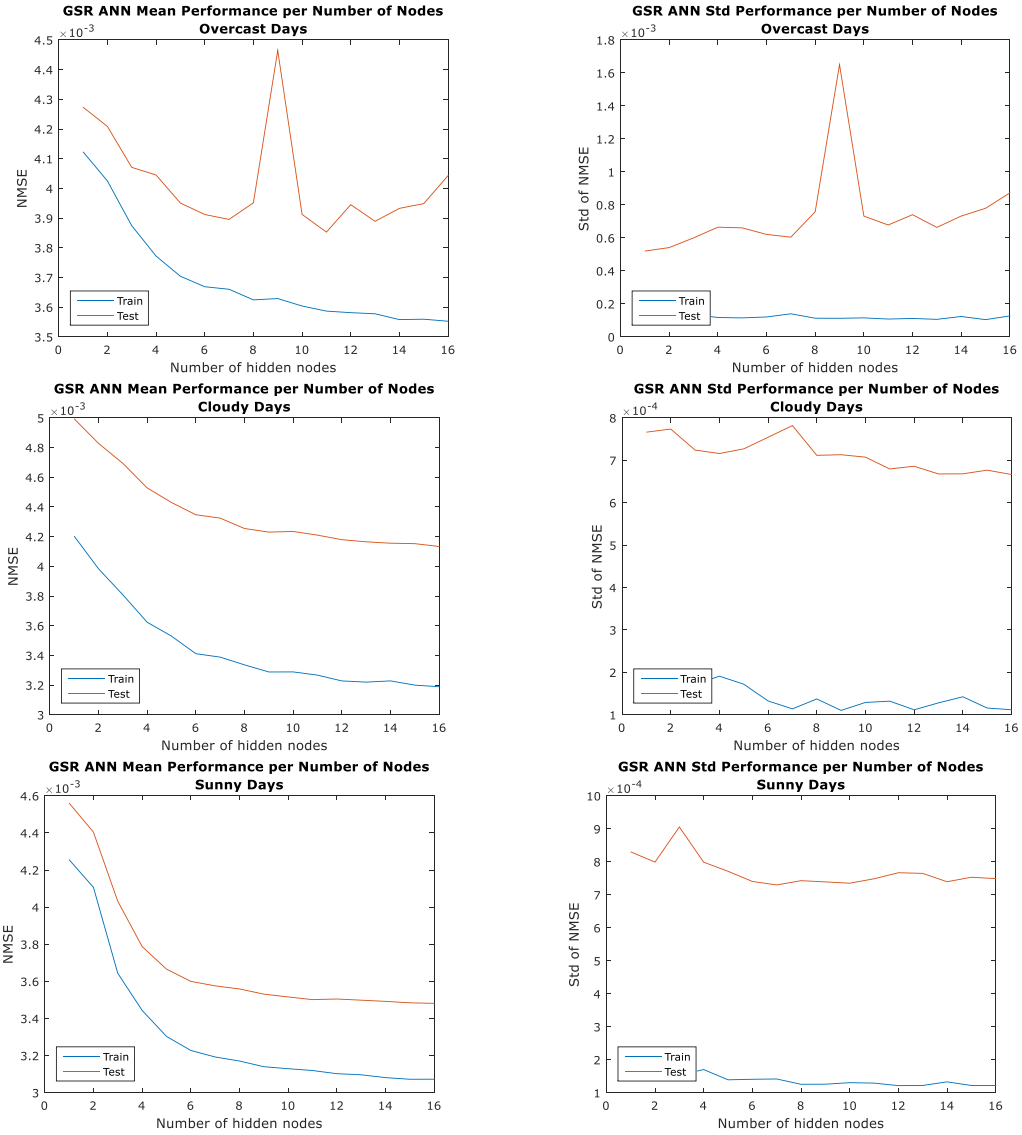


Figure 3.10: Training and test performances of GSR ANN for Overcast (top), Cloudy (middle) and Sunny (bottom) days. The mean performance given the 10-fold cross-validation per node (left) and the standard deviation (std) of the performance per node (right).

The next step is to determine exactly what ANN to use, because in each fold of the cross-validation process a different neural network was designed. For this, the validation data performance is calculated in order to know the overall performance of the neural networks.

Figure 3.11 shows 3 graphs, one for each kind of day. Each graph shows the best NN per kind of day, depending on the number of nodes in the hidden layer. As it was

explained before, each fold serves to train a different NN, and then it has to be determined the best NN not only for its number of hidden nodes but the correspondent fold. In Figure 3.11, the NMSE of validation, train and test data is shown considering every given fold.

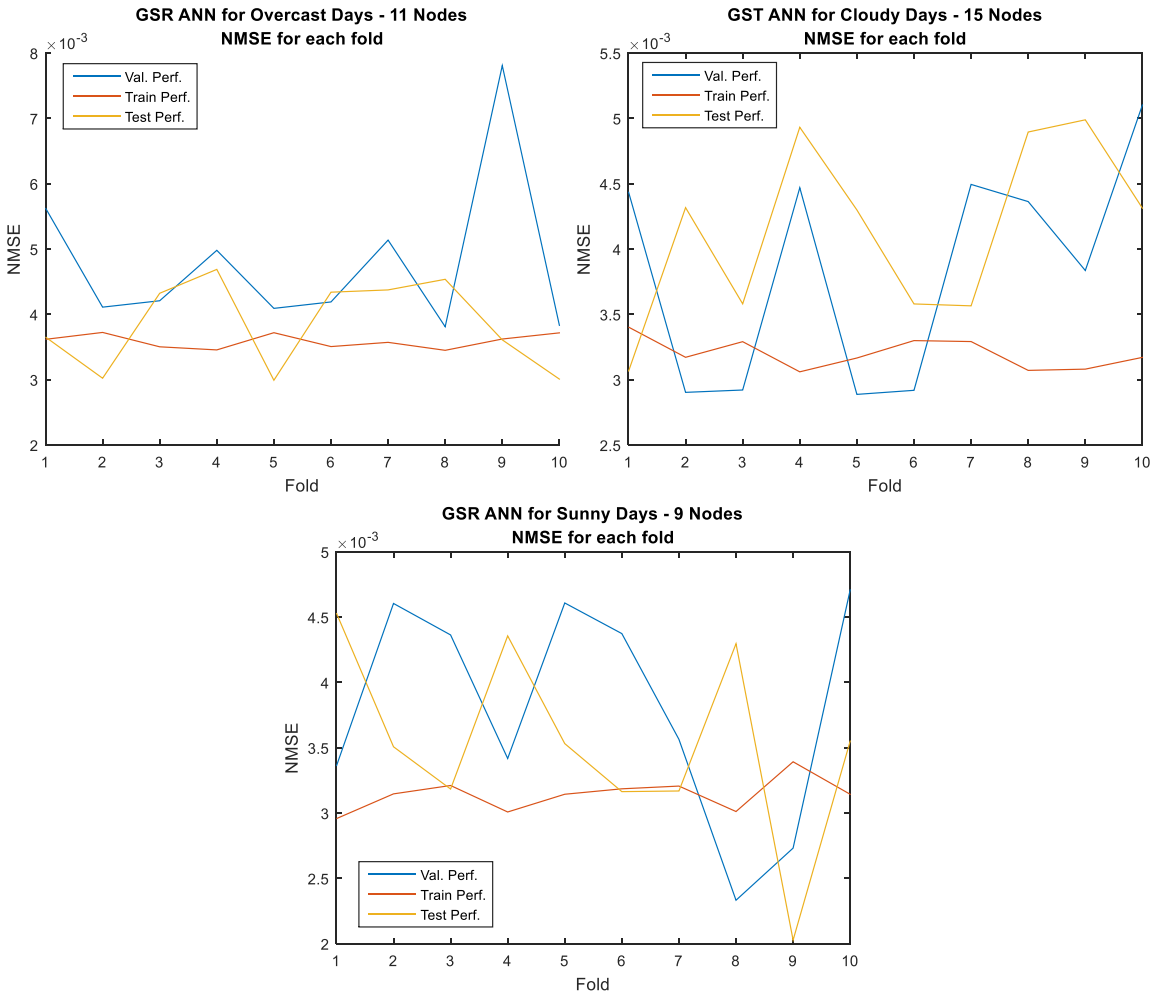


Figure 3.11: NMSE for validation, train and test performances for each fold. From top to bottom and left to right are shown the best fit GSR ANNs for overcast, cloudy and sunny days.

The best GSR ANN for each kind of day are the ones shown in Table 3.4. Different performances given by the NMSE are shown too.

Table 3.4: Best neural networks for each kind of day.

Type of day	Number of nodes	Number of folds	NMSE of Validation Data	NMSE of Training Data	NMSE of Testing Data
Overcast	11	10	0.0038	0.0037	0.0030
Cloudy	15	3	0.0029	0.0033	0.0036
Sunny	9	9	0.0027	0.0034	0.0020

Finally, the GSR-ANNs are shown in Figure 3.12.

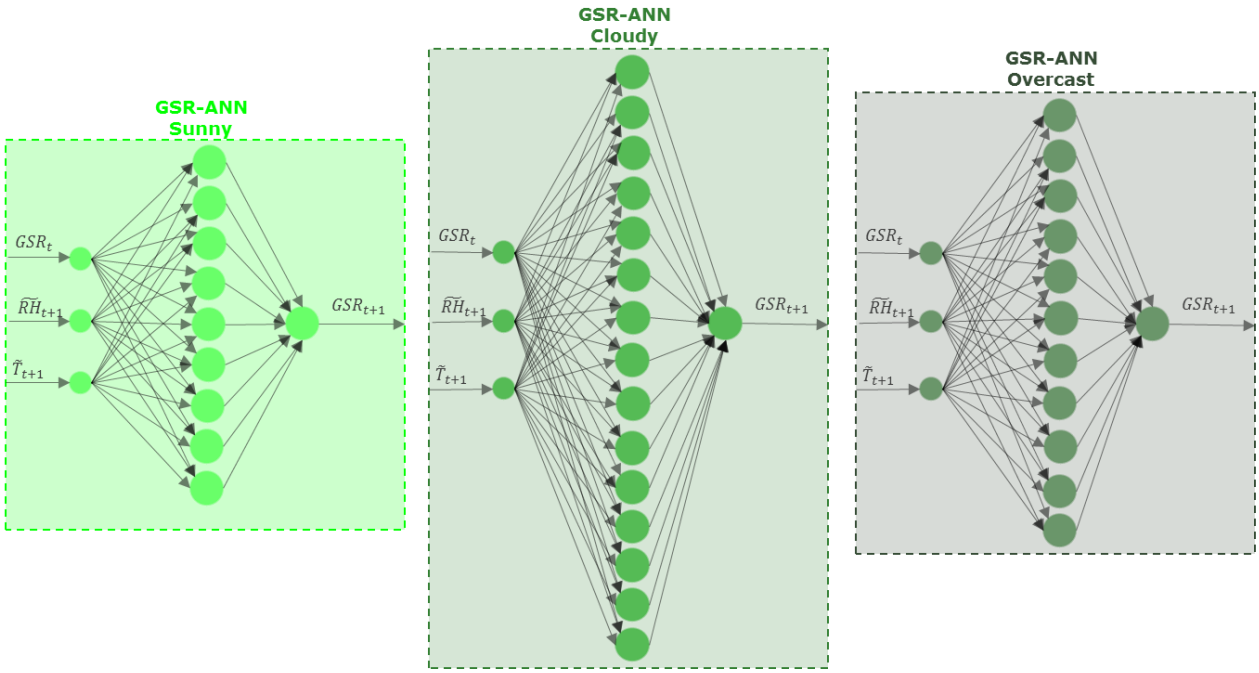


Figure 3.12: GSR-ANNs diagrams.

3.2. Control Model

The control model involves the calculation of a deload level depending on the forecasted GSR in $t + 10\,min$ given by the first stage ($\widetilde{GSR_{t+10}}$) and the measured GSR in t (GSR_t). The deload level is the percentage in which the P_{mpp} is decreased. This is calculated by a function that is determined by the best and worst radiation ramps the system could suffer. After this calculation, the MICAPAS control model proposed in [3] is implemented.

Figure 3.1 shows a schematic of the control model. It first calculates the deload level χ for the PV-PPs based on the forecasted power in $t + 10$ ($\widetilde{P_{out}^{t+10}}$) and its current value P_{out_t}. The deload level χ enters a centralized controller of the PV-PP and computes the available reserves of the power plant and the power deficit between sections.

The strategy considers a division of PV-PP into N sections, each one given by a different micro-inverter. Each section has a certain reserve level which depends on the incident radiation and deload level χ. With the information given by the N PV-PP sections, the MICAPAS calculates a control signal which is used to deploy reserves when a partial shading of the PV-PP is calculated. Each section of the PVPP includes a *d-q* power controller to change the MPP reference. Hence, their power output is altered as needed by the MICAPAS.

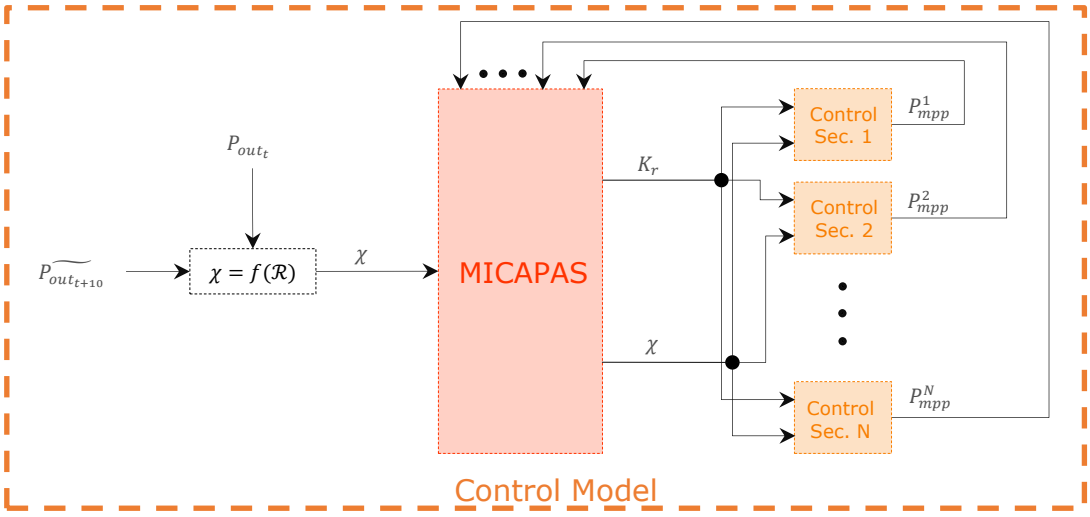


Figure 3.13: Diagram of control model.

MICAPAS

The frequency control strategy to allow partially shaded PV-PPs to mitigate the detrimental effects on system is based in a coordinated deloaded operation of the PV arrays that conform the PV-PPs. The strategy divides the PV-PP into *N* sections, each section with a specific reserve level given by the deload level that was calculated before.

MICAPAS requires a PV-PP architecture in which the MPP of each section can be tracked and independently controlled. Thus, a parallel-connected microconverter, series-connected microconverter, or a microinverter architecture are possible choices [46]. In this study, a microinverter architecture is employed; where each PV section can be composed by several panels. An example of a PV-PP with a microinverter architecture is shown in Figure 3.14.

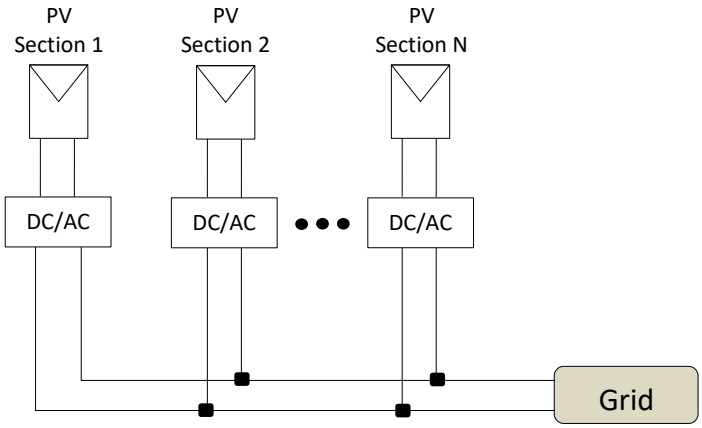


Figure 3.14: PV-PP with a microinverter architecture.

MICAPAS is continually monitoring each PV section of the PV-PP, in particular receiving the MPP of each section i (P_{mpp_i}), which depends on the GSR and temperature T at t. Based on these values, MICAPAS is constantly calculating a deload value (given by the function described in equations (19) and (20)). When one or more sections are under shaded conditions and there are reserves to deploy (i.e. the deload level is different from zero), the controller orders the unshaded sections to deploy their active power reserves. A general block diagram is shown in Figure 3.15.

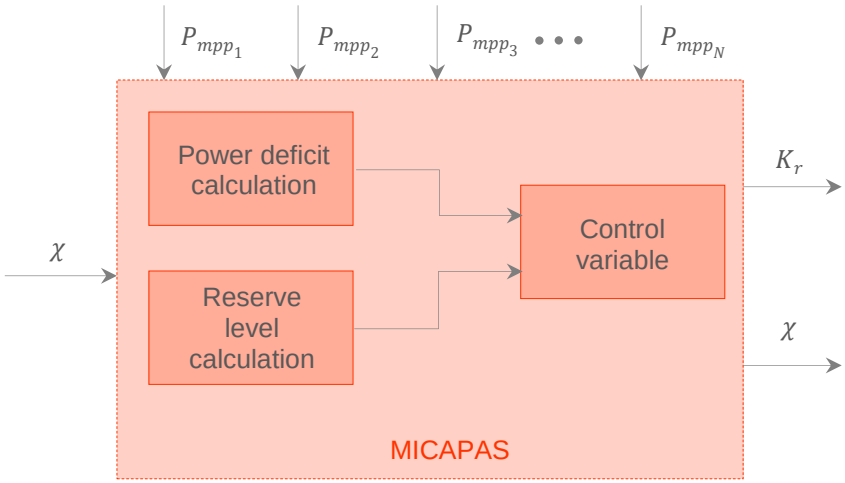


Figure 3.15: MICAPAS schematic.

The block "Power deficit calculation" in Figure 3.15 computes the total power deficit due to a partial-shading condition in the PV-PP based on:

$$\Delta P = \left[\sum_{i=1}^{N} \max\left\{P_{mpp_i}^{t}\right\}_{i=1,\dots,N} - P_{mpp_i}^{t}\right] \cdot (1-\chi) \tag{15}$$

Where χ is the deload level defined in the previous section. As can be seen from (15), ΔP is different from zero only when at least one PV section is not under shading conditions.

The block "Reserve level calculation" estimates the total amount of operating reserves available in the PV-PP according to:

$$\mathrm{R_t}(\chi) = \sum_{i=1}^{N} P_{mpp_i} \cdot \chi \tag{16}$$

Finally, MICAPAS calculates a control variable in order to deploy the power reserves in case of shading conditions. This signal enters to the control of each PV section making the active power controller to act. The control signal is defined as:

$$K_r = \left(1 - \frac{\Delta P}{R_t(\chi)}\right) \tag{17}$$

The signal K_r takes values between 0 (when the power deficit ΔP is equal to the total reserves) and 1 (when ΔP is zero). If $\Delta P > R_t(\chi)$, the value of K_r is limited to its lower value, 0.

As a summary, when $K_r = 1$, no control action is carried out and the operation of the PV section is sustained and when $0 < K_r < 1$, the control signal deploys the active power reserves of the PV-PP.

Deload level

The MICAPAS uses a deload level χ, which is calculated using both the forecasted ($\widetilde{P_{out_{t+10}}}$) and measured ($P_{out_t}$) power. In particular, this work uses the forecasted and measured radiations ($\widetilde{GSR_{t+10}}$ and GSR_t, respectively) to calculate a forecasted ramp $\mathcal{R}$ that represents the magnitude of the fall or rise of radiation in 10 minutes. The ramp $\mathcal{R}$ is calculated as follows:

$$\mathcal{R} = \frac{\left|\widetilde{GSR}_{t+10} - GSR_t\right|}{GSR_t} \tag{18}$$

Based on the forecasted PV power ramp, the idea is to find a "sound" deload level for the PV-PP. Thus, the main idea is to select a function $\chi(\mathcal{R})$, where χ is the deload level and $\mathcal{R}$ is the radiation ramp. To do this, two functions are compared: a ramp-type function and a step-type function.

To define the $\chi(\mathcal{R})$ functions, it is important to have a notion of what a "bad" and a "good" ramp is. This is determined by historical measured data, were the maximum, minimum and mean radiation ramps are analyzed. The maximum and minimum deload levels also have to be determined to define the $\chi(\mathcal{R})$ functions. The minimum deload level is given by the forecasting error and the maximum deload level is given by the one that is needed to maintain the system's frequency between its operational limits in [3].

The first step to determine $\chi(\mathcal{R})$ is to define the minimum radiation ramp level upon which the control scheme should work at minimum deload level. Since the rising and setting of the sun produces periodic radiation ramps during a day, the minimum ramp level should not be zero. Furthermore, the minimum ramp is calculated as the maximum ramp during a day without clouds in a year. For this purposes, a clear sky model of a summer day in northern Chile is calculated. Hence, the radiation of a day without clouds is determined. This is because the radiation ramps during a summer day are the highest of the year, and then, the steepest ramps during a clear-sky day through a summer day are the worst possible radiation ramps for a non-cloudy situation. The clear sky model is defined according to [47].

Figure 3.16 (a) shows the calculated clear-sky day and Figure 3.16 (b) shows a histogram of the radiation ramps during that day. It can be seen that the maximum

radiation ramps during a summer day are 0.05 (5%), therefore this ramp level is set to be the lowest in the deload function $\chi(\mathcal{R})$.

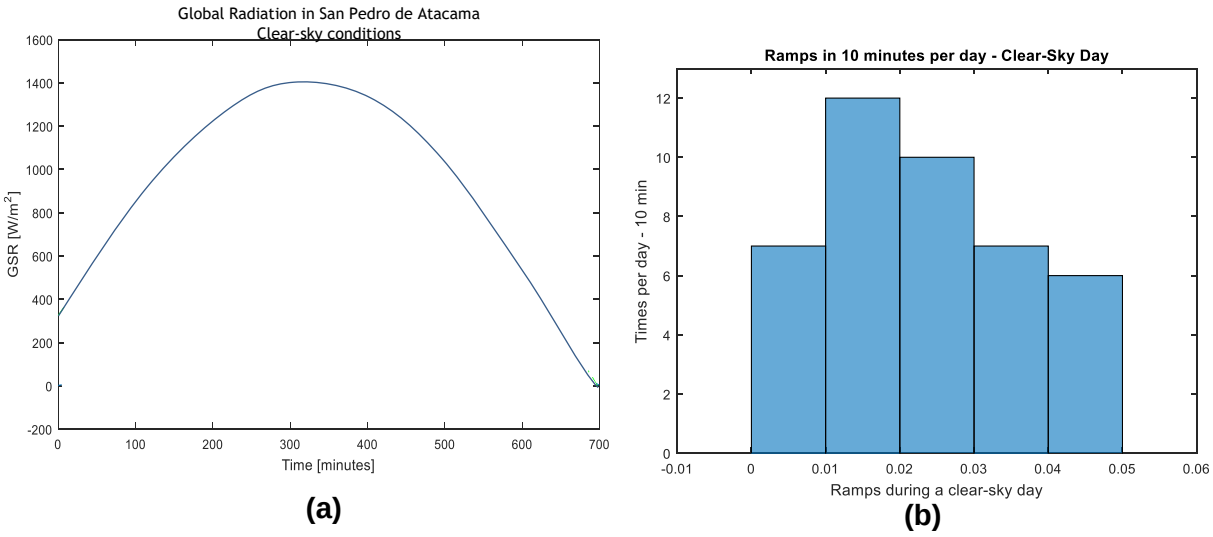


Figure 3.16: (a) Clear-sky day conditions during a summer day in northern Chile and (b) 10-minute ramps during a clear-sky day.

The second step to determine the ramp function is to define the maximum radiation ramp level in which the control should work at maximum deload level. To define this, the radiation ramps are calculated in each PV-PP. A histogram of radiation ramps every 10 minutes during a year (in each PV-PP) are shown in Figure 3.17. Each color in the histogram bars show a different PV-PP.

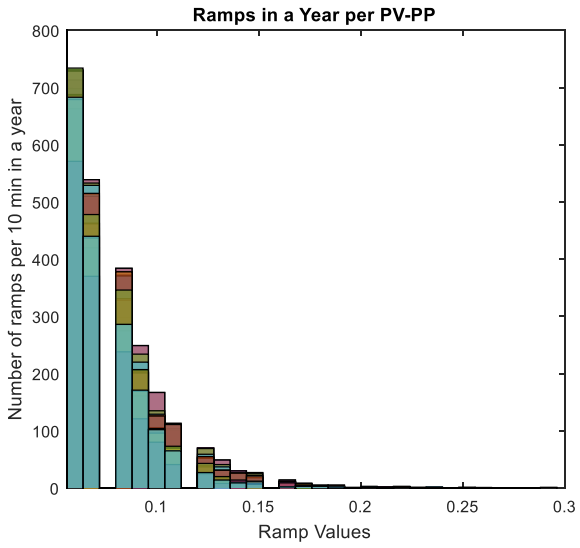


Figure 3.17: Radiation ramps during a year.

Figure 3.17 shows that the steepest ramp during a year, considering every PV-PP, is equal to 28%. This level is used as the maximum ramp possible for the deload function $\chi(\mathcal{R})$.

Finally, the minimum deload level is set as the root mean squared error (RMSE) of the forecast and the measured radiations, divided by the minimum radiation level between 11:00 and 17:00 of the corresponding day of the forecasting, which in this case corresponds to an approximate 2% of error. The reason to choose this error is that between those hours the PV-PP penetration is maximum, and so a forecasting error would be more relevant than at other times in the day.

The maximum deload level is set as 15%, because it proves to give good results in [3], leaving the system frequency in the normal range of operation. On the other hand, the step-type $\chi(\mathcal{R})$ function will start and end in the same deload level than the ramp-type of function, but the deload level will change every 5% of ramp. The equations for both functions are shown below.

$$\chi(\mathcal{R})[\%] = \begin{cases} 2 & \mathcal{R}[\%] \leq 5\% \\ 0{,}6 \cdot \mathcal{R}[\%] - 1 & 5\% < \mathcal{R}[\%] < 28\% \\ 15 & \mathcal{R}[\%] \geq 28\% \end{cases} \tag{19}$$

$$\chi(\mathcal{R})[\%] = \begin{cases} 2 & \mathcal{R}[\%] \leq 5\% \\ 5 & 5\% < \mathcal{R}[\%] < 10\% \\ 10 & 10\% < \mathcal{R}[\%] < 15\% \\ 15 & \mathcal{R}[\%] \geq 15\% \end{cases} \tag{20}$$

Both deload functions are shown in Figure 3.18. The saturation of both functions is proposed to avoid unnecessary energy losses in the system.

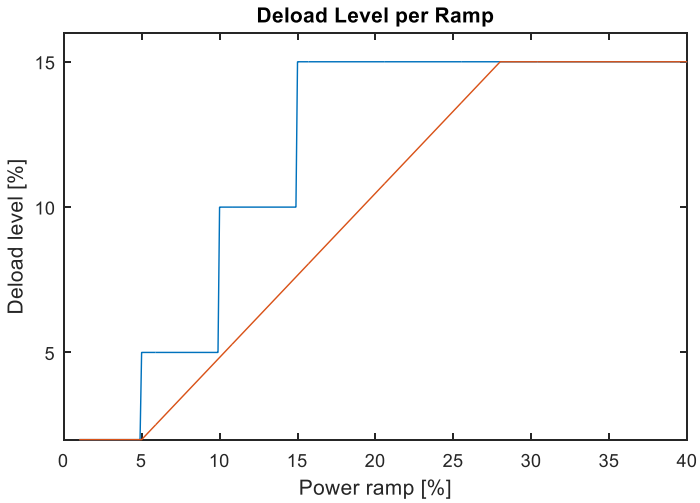


Figure 3.18: Deload level versus radiation ramp in 10 minutes.

Control of PV section

The PV section control, or local control scheme of each PV section, receives the signal K_r and χ from MICAPAS to change the power output if it is requested by the central controller (see Figure 3.19). It also delivers the $P^N_{mpp_i}$ from the MPPT of section N to MICAPAS. For this control scheme to work, a cascaded DC/DC converter connection of photovoltaic modules has to be considered, as it is shown in [48].

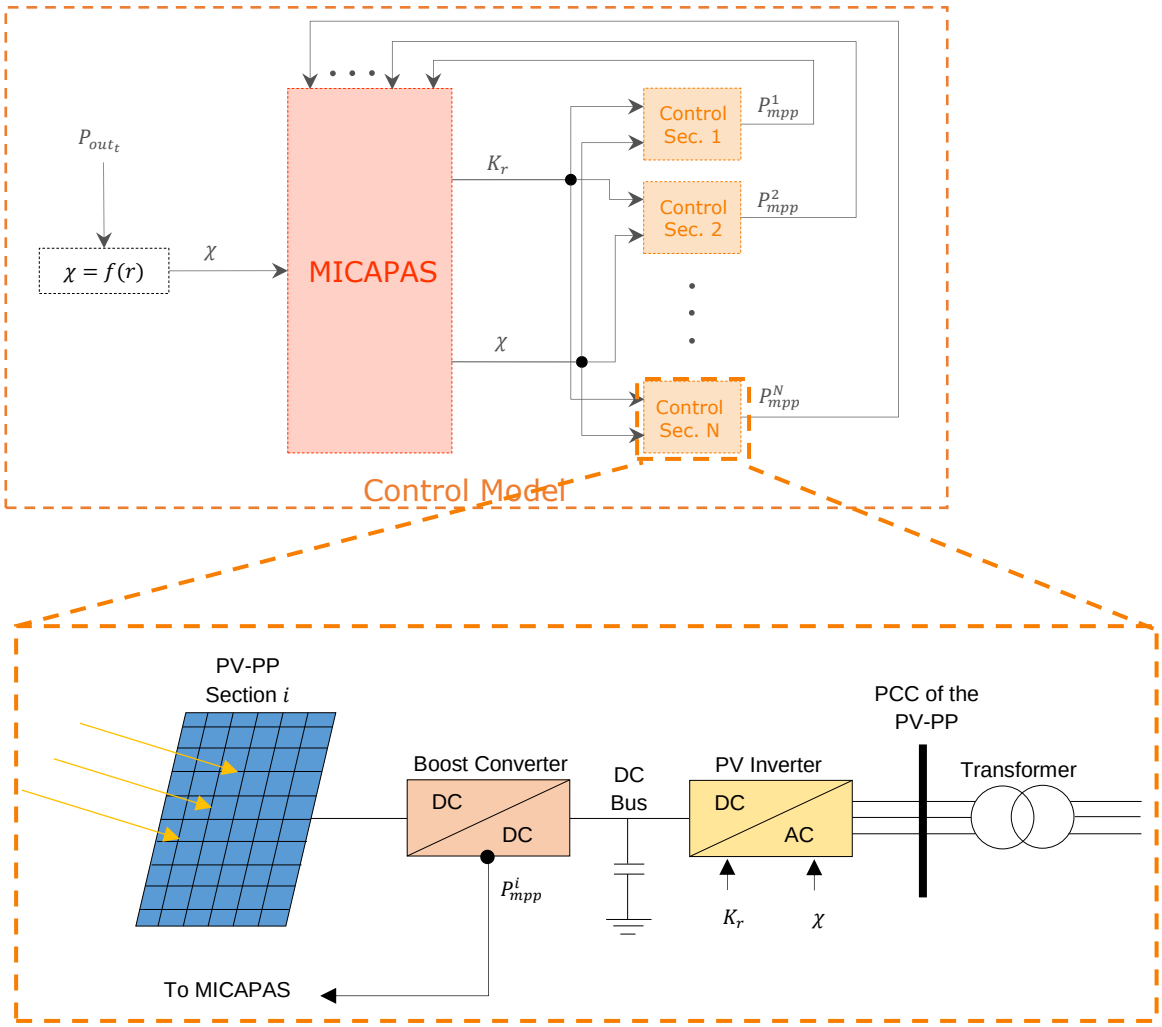


Figure 3.19: PV section configuration.

A cascaded DC/DC converter configuration is represented in Figure 3.19, where a boost converter is connected to a DC bus, providing galvanic isolation and performing MPPT control before connecting the PV section to the grid via a PV inverter. This kind of configuration is widely used in grid-connected photovoltaic systems [49], [50], [51] and decouples the PV system operating point from the PV inverter grid control. The MPP of the PV section (P^i_{mpp}) is calculated in the boost converter and additionally sent to the MICAPAS.

In this study, the deloaded operation of the PV section is accomplished by controlling the PV inverter of each PV-PP section and operating it at a lower DC voltage than the one calculated by the MPPT controller. Although an increased DC voltage would also result in an output power reduction of the PV-PP, a reduced DC voltage is selected due to the higher efficiency of the converter [52]. A general scheme of the section controller implemented in the PV inverter is shown in Figure 3.20.

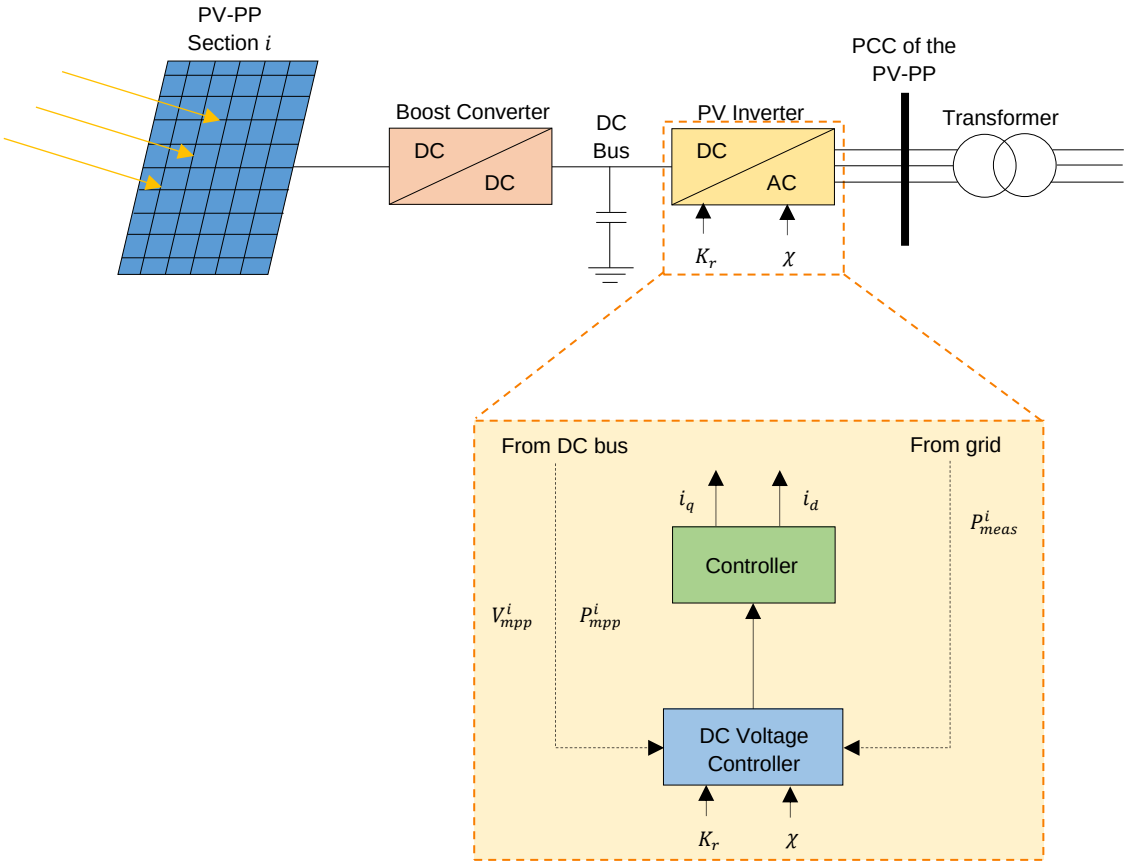


Figure 3.20: PV Section Control scheme, for a certain section i of the PV-PP.

The "PV-PP Section i Array" of Figure 3.20 calculates the output power from a radiation measurement given the following expressions [53]:

$$P_{panel} = \frac{E_{g,pv} \cdot P_{pk,panel} \cdot \eta_{rel} \cdot \eta_{inv}}{E_{STD}} \quad (21)$$

$$P_{system} = P_{panel} \cdot num_{panels} \quad (22)$$

Where P_{panel} is the active power output of the panel in kW, P_{system} is the system's active power in kW, num_{panels} is the number of panels per inverter, $E_{g,pv}$ is the GSR, E_{STD} is the standard irradiance value of $1000\ W/m^2$, $P_{pk,panel}$ is the total rated peak power of the solar panel in kW, η_{rel} is the relative efficiency of the panel and η_{inv} is the efficiency factor of the inverter.

The MPPT of each section is implemented in the boost converter and therefore the signals of voltage and power at the MPP (V^i_{mpp} and P^i_{mpp}, respectively) are calculated disjointedly from the PV control scheme. The P^i_{mpp} signal is sent to the central controller

(MICAPAS) and the local DC voltage controller in the PV-PP section. The P_{mpp}^{i}signal is used to calculate the supplementary signal ΔV_{dc_i} of the DC voltage controller.

In this work, the deloaded operation is accomplished by operating the array at a lower level than the optimal point V_{mpp}^{i} given by the MPPT. This is done by calculating a new signal ΔV_{dc_i} in the DC voltage controller. The DC voltage controller compares the measured power output P_{meas}^{i} with the reference power output P_{i}^{ref}. The error signal is sent to a PI controller subsequently.

The block “DC Voltage Controller” of Figure 3.20 is a PI controller that adjusts the active power considering the deload level. A control diagram of this block is shown in Figure 3.21.

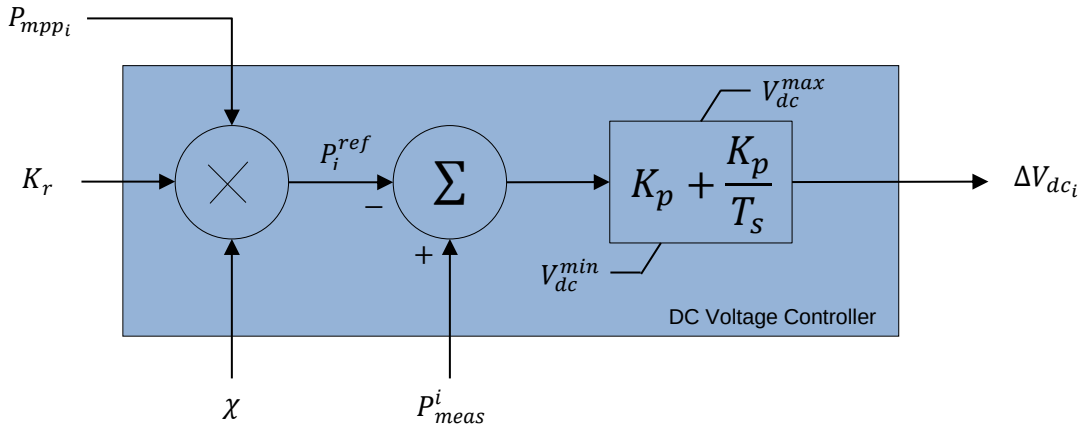


Figure 3.21: DC voltage controller block diagram.

The reference power P_{i}^{ref} is calculated as:

$$P_{i}^{ref} = P_{mpp_i} \cdot K_r \cdot \chi \tag{23}$$

The control compares P_{i}^{ref} to the measured output power P_{meas}^{i}. This enters the PI control that delivers a value of ΔV_{dc_i} in order to operate in a deloaded level. If the deload level is zero, then $P_{i}^{ref} = P_{meas}^{i}$ and ΔV_{dc_i} is zero.

The V_{mpp_i} signal enters to the $d-q$ controller as the optimal DC voltage to operate the section. The difference between the measured value, V_{dc}^{meas} and the optimal value V_{mpp_i} is compared to the additional signal ΔV_{dc_i} and the error is sent to a PI controller thereafter to generate the d-axis current reference value (Figure 3.22). However, a delay block is added in order to compare the V_{dc}^{meas} and V_{mpp_i} signals in which the delay time

corresponds to the action time of the MPPT. Finally, the d-axis current then regulates the active power of the converter.

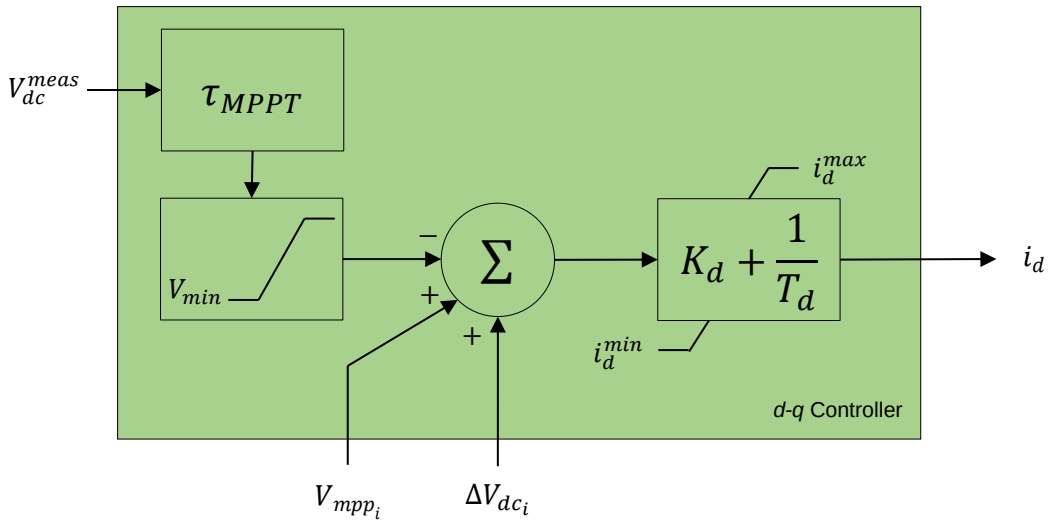


Figure 3.22: Block diagram of the *d* current controller.

Chapter 4.

Methodology

This chapter describes the methodology of this work. The methodology is shown in general in Figure 4.1.

.

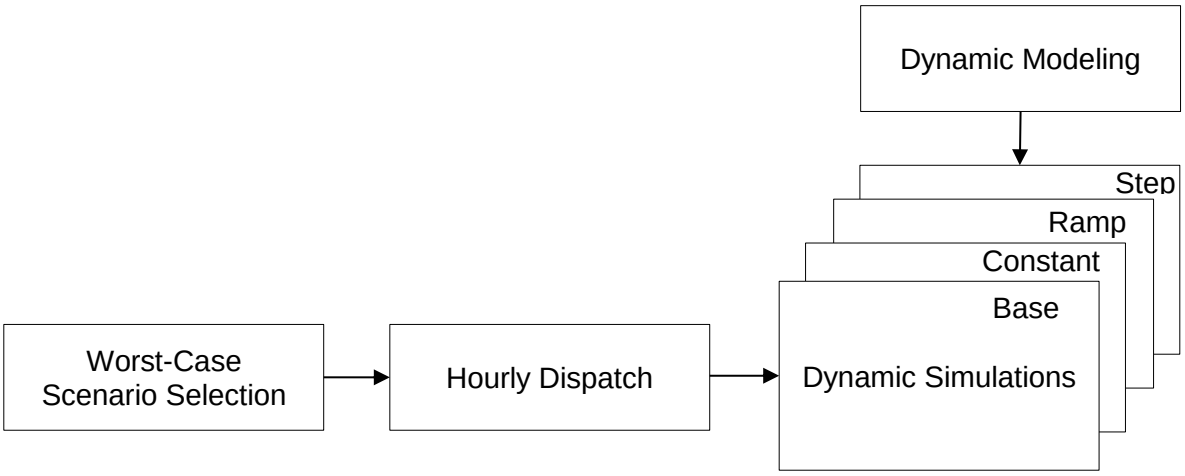


Figure 4.1: Methodology of this work.

Dynamic modeling

This part consists in the implementation of the models of the power system and the PV-PPs in the simulation tool DigSILENT PowerFactory 15.1. The simplified model of the Northern Interconnected System of Chile (NIS) has 120-buses, 68 generation units and 101 loads for the year 2020. Most of the generation units are gas-based or carbon-based thermal units. The vast majority of the loads correspond to industrial loads, particularly induction motors associated with the mining industry.

Worst-case scenario selection

As the main objective of this control strategy is to act when partial shading is occurring in a PV-PP, the worst-case scenario consider a cloudy day with the steepest ramps.

Because the computational burden of simulating a whole day is very high, the interval between 11:00 to 17:00 hours is selected for the dynamic simulations.

To define worst ramps the day with the smallest standard deviation with respect to clear sky model of the same day was first selected. This day (called reference day) is January 5th. Afterwards, the difference between the other days and this day was calculated as follows:

$$\Delta Rad_i^k = \left| Rad_i^k - Rad^{ref} \right| \tag{24}$$

Where $\Delta Rad_{i,j}^k$ is the difference of radiation between day i of (PV-PP k) and the reference day (Rad^{ref}). Then, the sum of ramps in a day is calculated using the following equation:

$$\mathrm{R_{index}} = \sum_{k=1}^{M} \left| \frac{\Delta Rad_i^k(n) - \Delta Rad_i^k(n-1)}{10} \right| \tag{25}$$

Where $\mathrm{R_{index}}$ is the index that reflects the ramping level in a day. M is the number of radiation samples in a day and is divided by ten because the ramps are caused in 10 minutes. Finally, $\mathrm{R_{index}}$ is normalized as follows:

$$\widehat{\mathrm{R_{index}}} = \frac{\mathrm{R_{index}}}{\max(\mathrm{R_{index}})} \tag{26}$$

Then, the worst day (more variable) is the day with an index equal to 1. The worst days are selected by choosing the maximum $\widehat{\mathrm{R_{index}}}$ per kind of day (sunny, cloudy and overcast). These are shown in Figure 4.2.

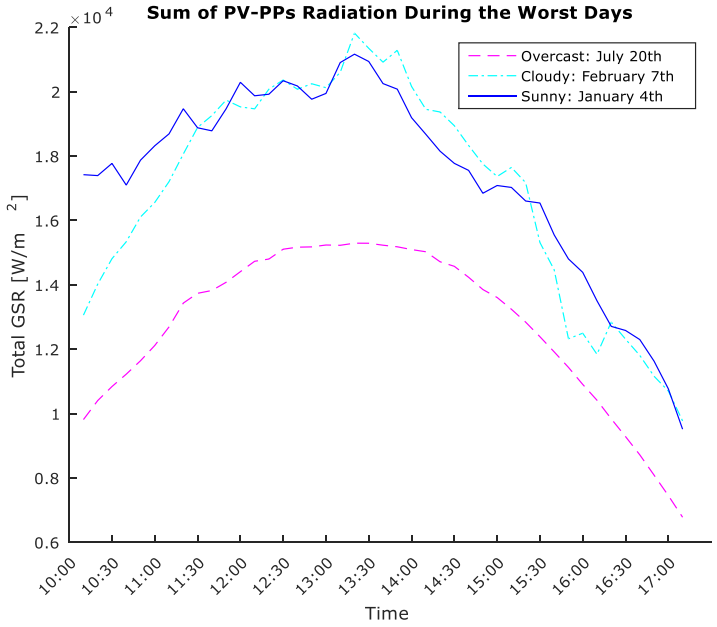


Figure 4.2: Worst days per classification of kind of day.

Finally, the case study is February 7th, because is the one that shows worst results ($\widehat{R_{index}}$ near 1) in almost every PV-PP and it is also the day with the maximum PV penetration in the power system. This means this day is a worst case not only from a ramp perspective, but also from a frequency stability point of view.

Hourly dispatch

The hourly dispatch is defined by an approximation of the electricity demand on February 7th 2020. This dispatch is determined by using historical demand data and a merit list based on the marginal costs of the conventional generators in the year 2020.

Since the PV-PPs have a relevant penetration in the power system only when the radiation is significant, the day was not fully simulated. Hence, the worst hours in a radiation perspective were selected. The selected hours in that day were from 10:30 and 18:30, because those are the hours of maximum penetration of PV-PPs. On the other hand, an 88.2% of the projected demand of the NIS for 2020 corresponds to the industrial sector [54], which in the case of the NIS is mainly mining industry. This means that the demand on the NIS is almost constant, and so the considered demand is equal to 2832.36 [MW].

Dynamic simulations

In this work, 4 different dynamic simulations are performed:

- Base case: This simulation corresponds to the system without deload. Thus, in this case there is no mitigation control against partial shading.

- Simulation S15c: This dynamic simulation considers a constant deload level of 15%, because it is shown in [3] to be the best case scenario and maintaining the frequency between a normal operation range.

- Simulation S15Ramp: This simulation uses a ramp-type deload function $\chi(\mathcal{R})$.

- Simulation S15Step: This simulation uses a step-type deload function $\chi(\mathcal{R})$.

Chapter 5.

Case study

The power system under study is the Northern Interconnected System (NIS) of Chile. This system is used both for analyzing the forecasting performance of the neural networks and the dynamic performance of the whole control strategy. The NIS is located in the middle of the Atacama Desert, a place with one of the greatest solar potentials in the world [55].Considering the high solar potential in northern Chile, and the estimations of 2.5 GW of installed capacity in PV technology by 2017 in Chile [11], it is expected to have a significant amount of PV-PPs concentrated in the NIS.

The NIS is a 50-Hz-isolated electricity system and is characterized by a thermal generation mix based on coal, oil and natural gas. The most relevant characteristic of the conventional units in the NIS from a frequency perspective are slow reaction times, low inertia levels and limited ramp rates.

The projected installed capacity for 2020 is 5800 MW and its projected peak load 3300 MW. The system load is characterized, by 88.2% of industrial load, corresponding to the mining industry, while the remaining 11.8% corresponds to residential customers. The PV capacity estimation of the NIS is based on the available information of future projects corresponding to private initiatives in Chile [56]. The PV-PPs are distributed at four distinctive locations throughout the Atacama Desert.

A simplified diagram of the NIS is shown in Figure 5.1.The black dots are electrical substations, the black triangles are thermal power plants, the light blue circles represent the meteorological stations in the NIS and the yellow circles represent PV-PPs sectors. In sector 1, five power plants are represented, sector 2 includes one power plant, sector 3 three power plants and sector 4 six. The maximum power per zone in each simulation scenario is shown in Table 5.1. Sector 2 is not included in the dynamic simulations because the presence of that power plant (and others) increases the PV penetration from more than 37% to almost a 50%, which made the system very unstable.

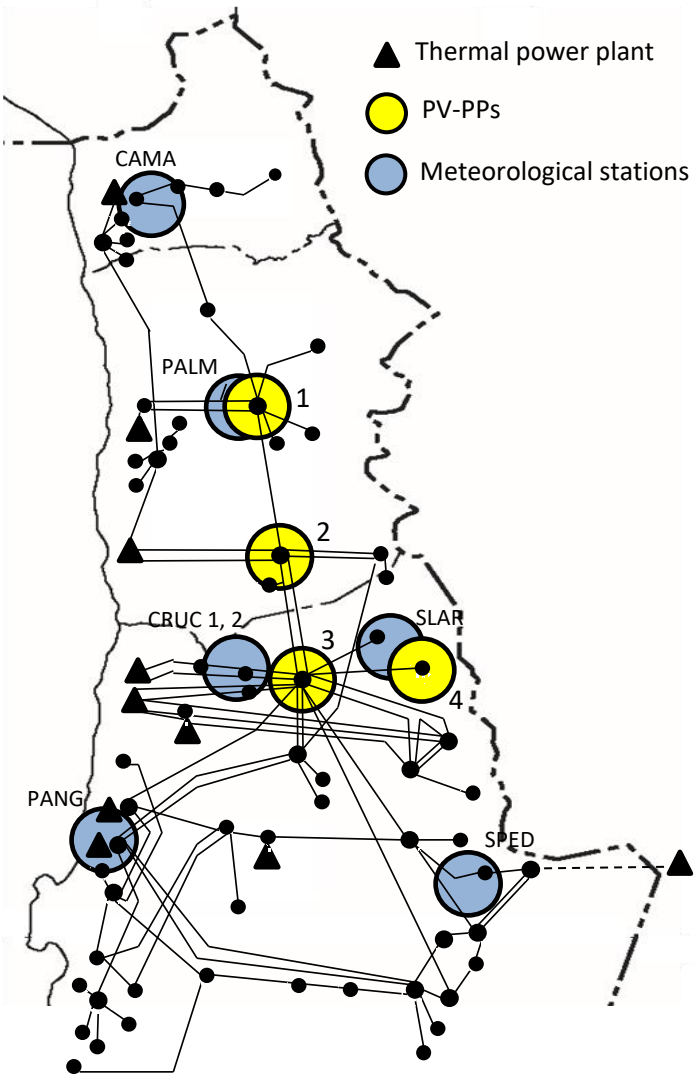


Figure 5.1: Simplified diagram of the NIS of Chile.

Table 5.1: Maximum power of PV-PPs during the different operation scenarios.

PHOTOVOLTAIC POWER PLANT ZONES	*MAXIMUM POWER BASE CASE*	*MAXIMUM POWER CONSTANT 15%*	*MAXIMUM POWER RAMP*	*MAXIMUM POWER STEPS*
ZONE 1	*347.12*	*297.13*	*342.41*	*345.39*
ZONE 2	*N/C*	*N/C*	*N/C*	*N/C*
ZONE 3	*386.94*	*331.23*	*381.70*	*385.02*
ZONE 4	*165.84*	*141.94*	*163.6*	*165.04*
TOTAL	***899.91***	***770.33***	***887.71***	***895.41***

In an ideal situation, solar data or output power of PV-PPs would be available and then the forecasting models would be determine by using that data. However, there was no data for the radiation or the output power per PV-PP, and therefore the CSI-ANN and GSR-ANNs for each PV-PP were trained using the data of the nearest meteorological station. The solar radiation data per PV-PP is calculated as it is shown in Appendix B using available information from meteorological stations nearby. Since the generated data per PV-PP is constituted by only one year, this data is not enough to train the ANNs. Therefore, different ANNs are trained depending on their sector. For sector 1, PALM meteorological station is used, for sector 2 a ponderation between PALM and CRUC 1, for sector 3 CRUC 1 and for sector 4 SLAR. The corresponding meteorological station for each PV-PP sector is shown in Table 5.2.

Table 5.2: Used ANNs for each PV-PP sector.

PV-PP SECTOR	*ANN METEOROLOGICAL STATION*
1	*PALM*
2	*PALM AND CRUC 1*
3	*CRUC 1*
4	*SLAR*

Nevertheless, each section of the PV-PPs has to have a different radiation level depending on whether there are clouds or not. To mimic this effect, three out of four sections of the PV-PPs is considered to have the aforementioned radiation data and the other one is considered to have a clear sky model radiation.

The next sections of this chapter explain the case study used in the forecasting tool and the control model in separate ways for better understanding.

5.1. Forecasting case study

Due to the lack of data per PV-PPs, the ANNs were trained, validated and tested based on the available information of three meteorological stations: Pozo Almonte (PALM), Crucero (CRUC) and Salar (SLAR). The places are shown in Figure 5.2 and the coordinates and elevation of the places are shown in Table 5.3. This table shows the available data in each meteorological station. The complete dataset corresponds to 10-min measures of global horizontal radiation, diffuse radiation, direct normal radiation, temperature, relative humidity and wind speed.

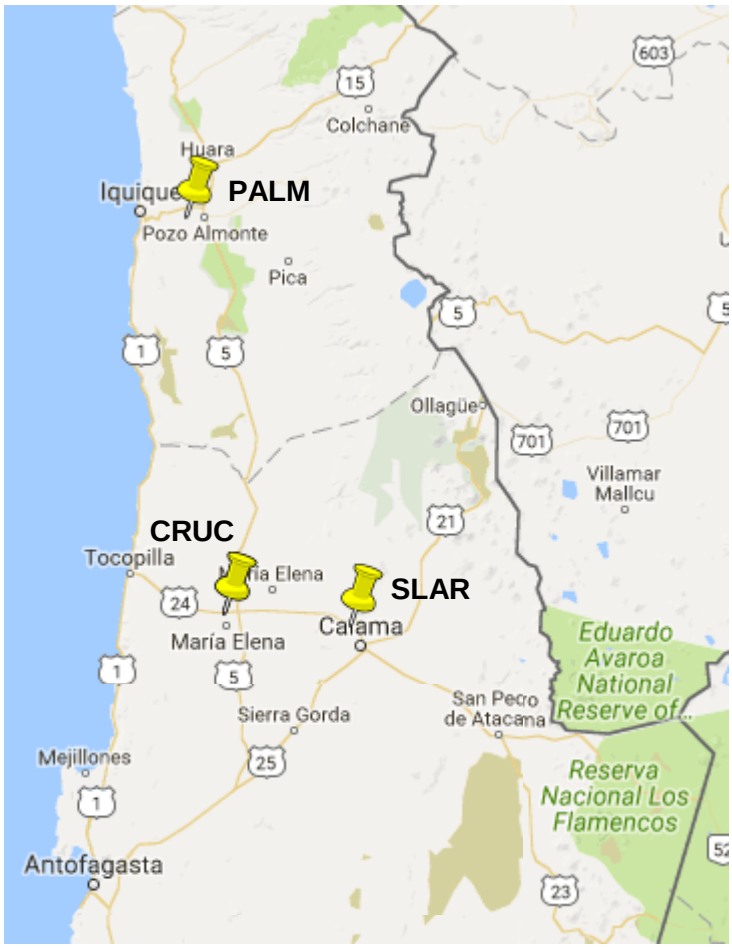


Figure 5.2: PALM, CRUC and SLAR meteorological stations.

Table 5.3: Sites characteristics.

	PALM	CRUC	SLAR
LATITUDE	20°15'24.48" S	22°16'28.56" S	22°20'27.24" S
LONGITUDE	69°46'30.00" W	69°33'57.96" W	68°52'35.76" W
ELEVATION	1024 m.a.s.l.	1185 m.a.s.l.	2526 m.a.s.l.
AVAILABLE DATA	08/01/2008 – 10/29/2015	08/28/2009 – 11/11/2014	05/20/2010 – 12/31/2012

For the CSI-ANN training, per day measures are needed, therefore the 10-minute data was averaged in order to get mean values of temperature, relative humidity and GSR per day.

For the GSR-ANNs, the 10-minute data was split into the different kind of days – Sunny, Cloudy and Overcast – depending on their CSI classification. The classification was done considering the calculation of CSI shown in Appendix A and the classification that was shown in Table 3.1. Each group of data – Sunny, Cloudy and Overcast – was used to train the respective GSR-ANN. The total data for each kind of day per meteorological station is shown in Table 5.4.

Table 5.4: Total 10-min data for kind of day.

METEOROLOGICAL STATION	*KIND OF DAY*	*TOTAL 10-MIN DATA*
PALM	*SUNNY*	*48942*
	CLOUDY	*94215*
	OVERCAST	*41178*
CRUC	*SUNNY*	*44626*
	CLOUDY	*69157*
	OVERCAST	*28163*
SLAR	*SUNNY*	*19250*
	CLOUDY	*21910*
	OVERCAST	*11078*

5.2. Control Case Study

The dynamic performance of the control model was tested for a critical day in terms of solar variability. The solar variability was measured as shown in Chapter 4.

The proposed control was tested for four different scenarios regarding their deload level, all of them with the same total PV capacity (see Table 5.5).

Table 5.5: PV-PPs in dynamic simulations and maximum power.

PHOTOVOLTAIC POWER PLANT	***MAXIMUM POWER BASE CASE***
ATACAMA SOLAR	*272.90*
PV COYA	*87.11*
SAN PEDRO DE ATACAMA I	*34.91*
SAN PEDRO DE ATACAMA II	*34.91*
SAN PEDRO DE ATACAMA III	*34.91*
SAN PEDRO DE ATACAMA IV	*34.91*
CALAMA SOLAR 1	*13.10*
CALAMA SOLAR 2	*13.10*
CRUCERO ESTE	*139.11*
CRUCERO OESTE	*160.72*
SALAR DE HUASCO	*34.91*
POZO ALMONTE SOLAR 1	*8.74*
POZO ALMONTE SOLAR 2	*13.10*
POZO ALMONTE SOLAR 3	*17.47*
TOTAL	***899.91***

The difference among the scenarios is the nature of the deload level, χ. The first two scenarios consider a constant deload level over the day (0% and 15%). The third scenario considers the proposed scheme with a deload level defined based on a ramp function, and the fourth scenario considers the proposed scheme with a deload level depending on a step function. This is shown in Table 5.6.

.

Table 5.6: Considered scenarios for the study.

SCENARIO	*CHARACTERISTICS*	*DELOAD LEVEL*
S0C	CONSTANT	$\chi = 0$
S15C	CONSTANT	$\chi = 15$
S15RAMP	VARIABLE	$\chi = \begin{cases} 2 & \mathcal{R}[\%] \leq 5\% \\ 0{,}6 \cdot \mathcal{R}[\%] - 1 & 5\% < \mathcal{R}[\%] < 28\% \\ 15 & \mathcal{R}[\%] \geq 28\% \end{cases}$
S15STEP	VARIABLE	$\chi = \begin{cases} 2 & \mathcal{R}[\%] \leq 5\% \\ 5 & 5\% < \mathcal{R}[\%] < 10\% \\ 10 & 10\% < \mathcal{R}[\%] < 15\% \\ 15 & \mathcal{R}[\%] \geq 15\% \end{cases}$

Chapter 6.

Results

In this work, a mitigation control against partial shading effects is designed including an improved forecasting technique to determine the deload level in the PV panels. This section presents first the results of the forecasting model and then the results of the control model, considering the calculated deload level using the forecasting tool.

6.1. Forecasting

As seen in previous sections, three different GSR-ANN were designed after a CSI-ANN, in each meteorological station. The CSI-ANNs have a NMSE between 1.8% and 2.4%, which makes it a very accurate classification stage. The NMSE of each CSI-ANN, plus the number of nodes in the hidden layer in each, is shown in Table 6.1.

Table 6.1: NMSE per CSI-ANN.

METEOROLOGICAL STATION	*NUMBER OF NODES IN THE HIDDEN LAYER*	*NMSE OF THE CSI-ANN*
PALM	6	2.43 %
CRUC	14	1.83 %
SLAR	13	1.91 %

The number of nodes in the hidden layer was determined by the best overall results following a trial-and-error process. The overall performance was calculated as the NMSE of all the sets (training, validation and test). This analysis is shown in Figure 6.1. It is worth noting that the CSI-ANN may calculate a decimal number, thus the result of the CSI-ANN has to be round to its nearest integer.

The results show that the error convergence between the performance of each subset in the CRUC meteorological station is better, which may suggest that the CRUC data is less variable than the one in the other meteorological stations, because every set of data has more similar results. On the other hand, the SLAR station has higher errors in the test set. This is because SLAR has less information than the other meteorological stations, therefore the generalization capacity of the CSI-ANN in SLAR is worse than the other stations.

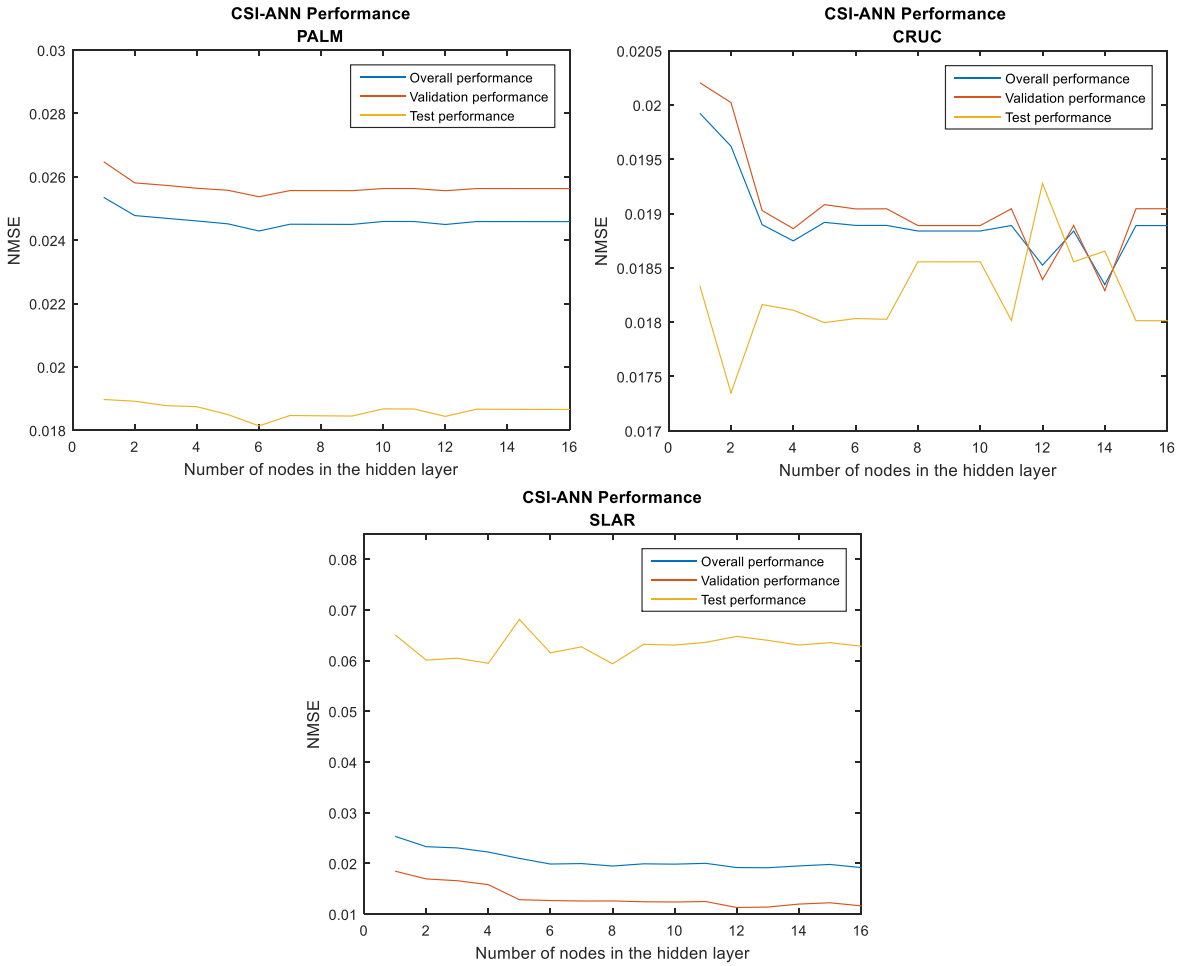


Figure 6.1: Overall, validation and test performance in the CSI-ANN per meteorological station.

When addressing the error as the number of times in a year were the type of day is wrongly forecasted, the result is 28 days, thus a 7.7% of the year. This means that the GSR-ANNs have to be evaluated for all kind of days, in case the CSI-ANN wrongly forecasts the type of day.

The other three GSR-ANN per meteorological station passed through a more complex design process, where a 10-fold cross-validation was done. This process suggests a larger computational burden in the training process than the CSI-ANN, were there was no cross-validation. The distribution and computational burden for each GSR-ANN in each meteorological station is shown in Table 6.2.

Table 6.2: Number of measures and computational burden in each kind of day.

METEOROLOGICAL STATION	KIND OF DAY	COMPUTATIONAL BURDEN (HOURS)
PALM	SUNNY	6.9
	CLOUDY	14.6
	OVERCAST	6.0
CRUC	SUNNY	6.3
	CLOUDY	10.7
	OVERCAST	4.1
SLAR	SUNNY	2.7
	CLOUDY	3.4
	OVERCAST	1.6

As the CSI-ANN can wrongly forecast 28 days in a year, it is important to show the forecasting error of the GSR-ANNs when evaluating not only the corresponding type of day data, but also the data of the other type of days. The NMSE of every GSR-ANN, when evaluating them with the data of every type of day, is shown in the following table:

Table 6.3: NMSE between forecasted and real values per GSR-ANN for every type of day.

METEOROLOGICAL STATION	GSR-ANN / DATA	OVERCAST	CLOUDY	SUNNY
PALM	OVERCAST	0.02 %	0.02 %	0.03 %
	CLOUDY	0.05 %	0.02 %	0.01 %
	SUNNY	0.05 %	0.02 %	0.01 %
CRUC	OVERCAST	0.02 %	0.04 %	0.06 %
	CLOUDY	0.08 %	0.03 %	0.02 %
	SUNNY	0.08 %	0.03 %	0.02 %
SLAR	OVERCAST	0.03 %	0.05 %	0.21 %
	CLOUDY	0.09 %	0.03 %	0.44 %
	SUNNY	0.10 %	0.03 %	0.02 %

When evaluating the GSR-ANNs with their corresponding type of day data, the NMSE is always smaller than 3% in every meteorological station, which means the forecast has good results when the kind of day is well forecasted. On the other hand, when the day is wrongly forecasted, the NMSE in PALM can reach a 4.7% error, in CRUC a 8.3% and in SLAR a 44.0%. SLAR is a critical case because of the lack of data in that place (less than 2 years), and therefore an important conclusion is that the lack of data affects significantly the overall performance of the proposed forecasting model.

In order to visually show the result of a forecasting, the worst and best-case scenarios were selected in terms of NMSE per day in the PALM meteorological station.

The worst-case scenario in sunny days in terms of bias was at January 27th 2014. The maximum bias in this day was of 451.30 W/m^2. On the other hand, the best-case scenario in sunny days was at January 5th 2010, reaching a maximum bias of 11.52 W/m^2. The forecast is shown in Figure 6.2 and it can be seen that the maximum bias is between 14:00 and 16:00 hours.

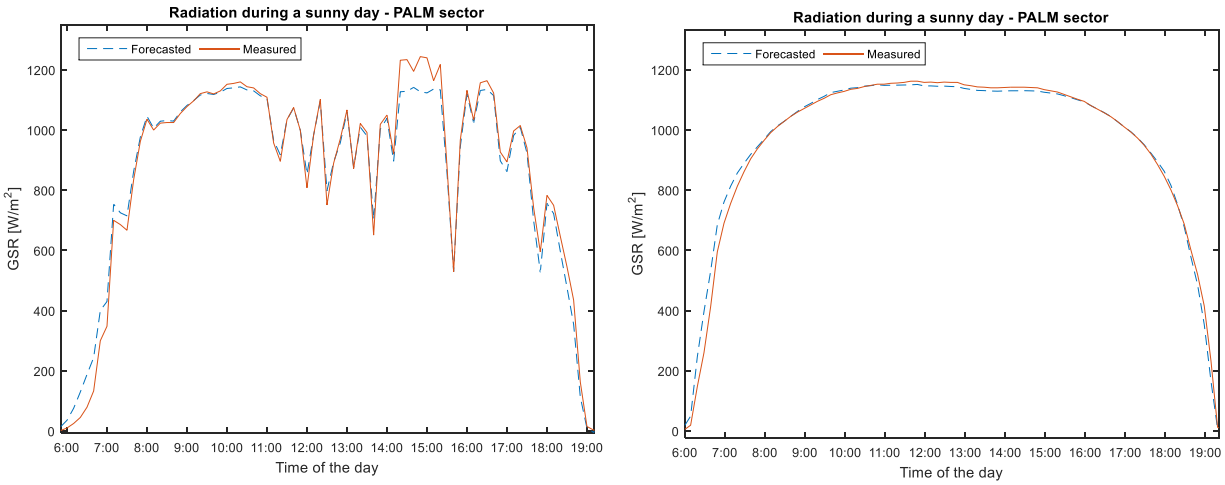
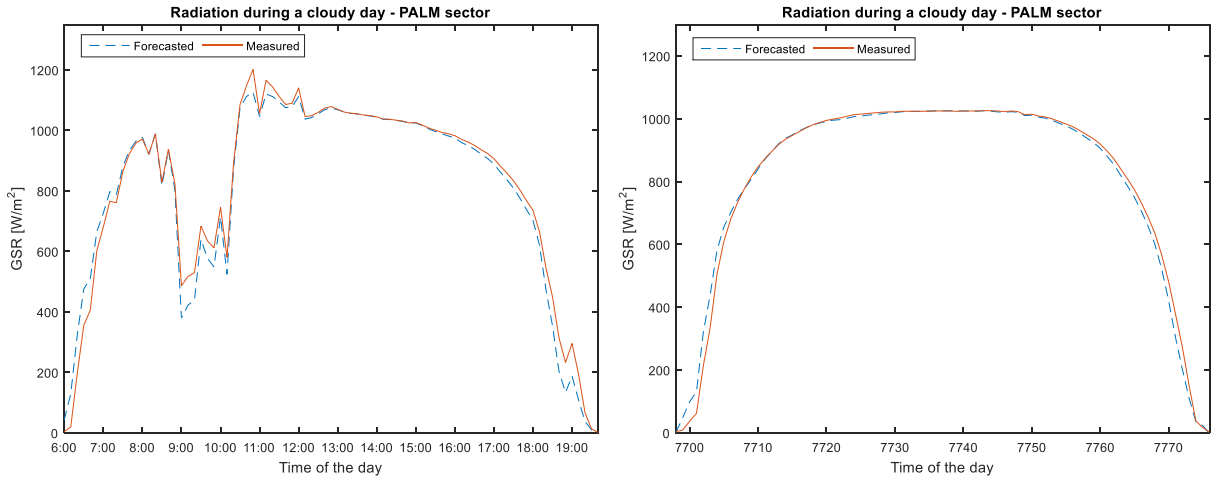


Figure 6.2: Forecast of the worst and best-case day during sunny days in PALM meteorological station. In (a) is shown the worst-case during January 27th 2014 and in (b) is shown the best-case during January 5th 2010.

The worst-case scenario in terms of bias for cloudy days was at February 11th 2014. The maximum bias is 406.73 W/m^2. On the other hand, the best-case scenario in cloudy days was at September 20th 2013, reaching a maximum bias of 6.35 W/m^2. Figure 6.3 shows the ANN forecast of both days and the measurements. This figure shows that the forecast underestimates the real radiation almost every time during the maximum penetration times, which is best for the mitigation control application because it allows to save all the necessary reserve without overestimating the deload level.



(a) (b)

Figure 6.3: Forecast of the worst and best-case day during cloudy days in PALM meteorological station. In (a) is shown the worst-case during February 11th 2012 and in (b) is shown the best-case during September 20th 2013.

Finally, the worst-case scenario for overcast days is at February 13th 2012, were the maximum bias reached is 273.31 W/m^2. On the other hand, the best-case scenario in overcast days was at May 15th 2012, reaching a maximum bias of 9.87 W/m^2. Figure 6.4 shows the ANN for both days and it shows the same underestimation as in case of cloudy days.

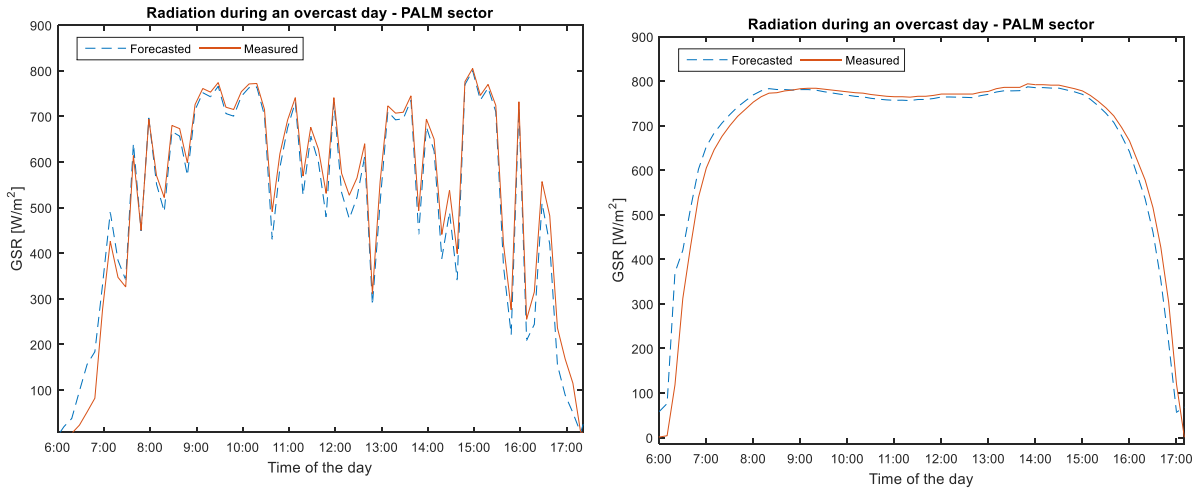


Figure 6.4: Forecast of the worst and best-case day during overcast days in PALM. In (a) is shown the worst case on February 13th 2012 and (b) shows the best day in July 4th 2011.

Overall is fair noting that, even though sunny days exhibit a higher bias, the total radiation of the day is also higher than cloudy and overcast days. This means that, in percentage, the difference is lower in sunny days because the overall radiation is higher.

6.2. Control

The control model was tested in a simplified model of the NIS with 120-buses for the year 2020 in the software DigSILENT PowerFactory. Four different scenarios were simulated for comparison purposes:

1. S0c: Base case. No control action is taken.

2. S15c: Best-case in [3]. The model is tested for a constant deload of 15%.

3. S15Ramp: The model is tested for a variable deload level $\chi(\mathcal{R})$ with a ramp-type function.

4. S15Step: The model is tested for a variable deload level $\chi(\mathcal{R})$ with a step-type function.

For scenarios S15Ramp and S15Step, the deload level was calculated previous to the model implementation in the power system simulation tool. To do this, the radiation data was forecasted depending on the type of day and the proximity to the corresponding

meteorological station. After the forecast is done, the radiation ramps and deload level are calculated for each deload-variable scenario.

Figure 6.5 shows the power system's frequency for all of the aforementioned operation scenarios. The straight lines at 49.8 and 50.2 [Hz] indicates the lowest and highest allowable levels for the system frequency according to the Chilean grid code during normal operation [57]. It is shown that there is a significant improvement of the system's frequency when applying the MICAPAS controller (in comparison to the base case). It is also shown that a frequency-related normal operation of the power system is achieved during all the day in the cases were the control scheme is operating. When the data is analyzed in detail, it shows that there is an abnormal operation of the power system in both deload-variable scenarios.

The frequency peaks that can be identified between 11:00 and 12:00, and between 17:00 and 18:00 can be attributed to a forecasting error in predicting the steepest drop or rise. The other option would be that the control scheme is deploying reserves as the deload level is changing, which is the case on the rest of the day in the variable-deload scenarios.

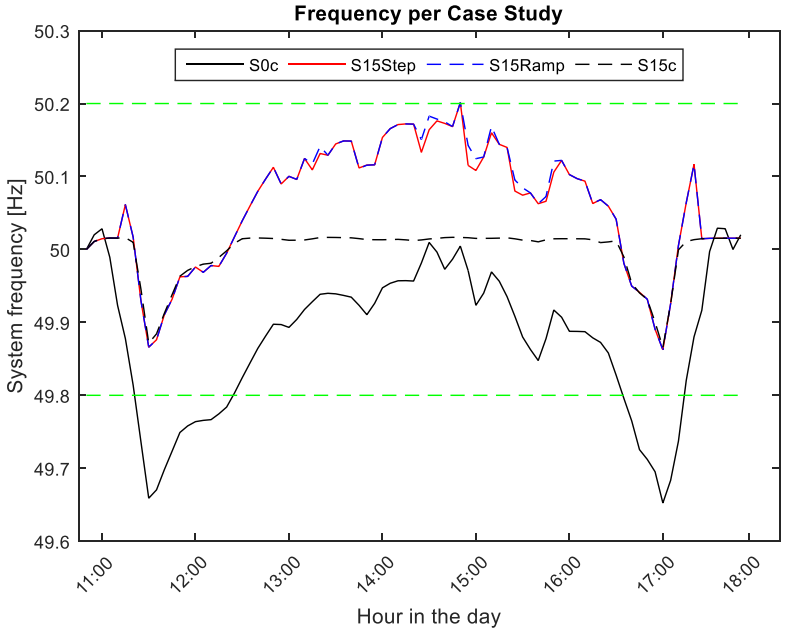


Figure 6.5: System frequency for scenarios S0c, S15c, S15Ramp and S15Step.

In scenarios S15Ramp and S15Steps, the system operates over the minimum normal operation frequency throughout all time, and above the maximum normal operation frequency a 1.2% of the time, which is not clear from the figure but is shown in Table 6.4. This issue is solved by the active power reduction loop, that is mandatory for PV-PPs as shown in the Chilean grid code [57]. It can also be argued that this control

scheme may increase the energy shedding from PV-PPs, but that would be the case only if the frequency is under or above the normal operation band, which is not the case a 98.8% of the time during a worst-case day. Therefore, the energy shedding would not be necessary a significant amount of times during a year. Furthermore, the Chilean grid code mandates PV-PPs to shed energy only if the frequency is out of the normal operation band during 3 hours, which is not the case. In addition, the figure also shows very few differences between both deload-variable scenarios. This is because both deload levels are very similar between a power ramp of zero and 10%, as it was shown in Figure 3.18, and the ramps during this day are between 0 and 10% more than a 91% of the time. Moreover, a 53% of the ramps during the day are equal to 5% or less, thus the deload level is the same for both scenarios.

In conclusion, the frequency response under this control scheme is better than the base case and very similar to a case with a constant deload level of 15%, because it stays between a normal operation band a 98.8% of the time and the partial shading effect is mitigated.

Table 6.4 shows the maximum PV penetration in each operation scenario, the percentage of the time that the system frequency is below 49.8 Hz and the percentage of the time that the system frequency is over 50.2 Hz.

Table 6.4: Maximum PV penetration, amount of time below 49.8 Hz and amount of time above 50.2 Hz for every operation scenario.

Operation scenario	*Maximum PV penetration*	*Amount of time (%) below 49.8 Hz*	*Amount of time (%) above 50.2 Hz*
S0c	*37.89%*	*24.00%*	*0.00%*
S15c	*31.76%*	*0.00%*	*0.00%*
S15Ramp	*36.53%*	*0.00%*	*1.18%*
S15Step	*36.84%*	*0.00%*	*1.18%*

It is clear from the table that the maximum PV penetration is given by the base case, which is reasonable because the base deload level for both S15Ramp and S15Step is 2%. Nevertheless, the maximum PV penetration is higher in the variable cases than in the constant 15% case, for at least a 4%. A higher PV penetration implies a more difficult situation for frequency regulation in the system, and so the amount of time above 50.2 Hz for both variable scenarios is understandable and it could also be considered be low.

Another important objective to this work is to prove that this forecasting technique decreases the energy losses during a day without worsening the frequency performance. The second point is proven by the aforementioned analysis, while the energy losses are shown in Figure 6.6. It is clear that there is an important difference between constant and variable deload level. The total energy losses during the day in S15c are equal to a 7%, whereas the energy losses in S15Ramp and S15Step are 2.04% and 2.17%, respectively. Therefore, the energy losses are decreased in at least 3 times when adding the

forecasting technique to the control system. It is also worth noting that during a variable day, which is the case in this simulation, energy losses are less than during a better day as it comes to cloudiness. This is because during a very variable day the power reserves are being deployed in a more consistent manner than during a less variable day, and therefore the energy losses during a better day in terms of clouds variability are higher than in this case. Hence this control scheme should be avoid during clear-sky days.

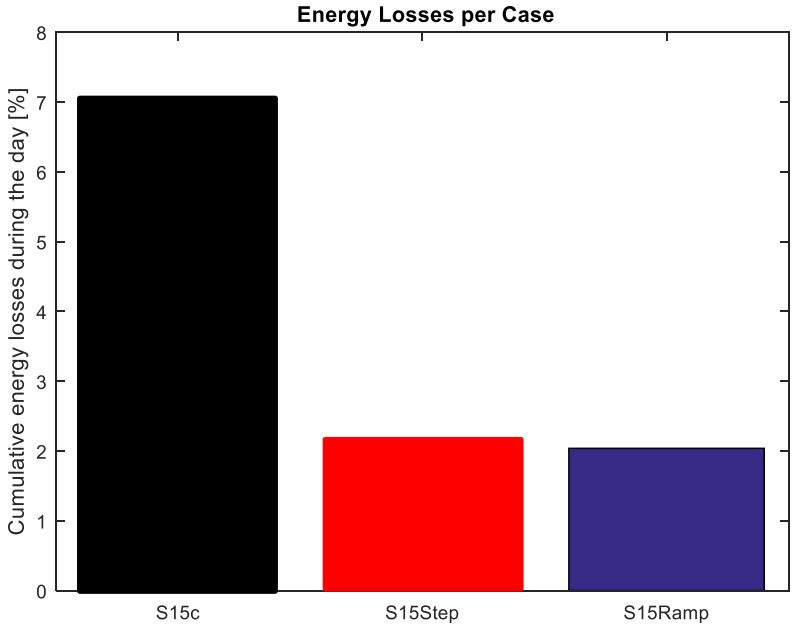


Figure 6.6: Cumulative energy losses for scenarios S15c, S15Ramp and S15Step, in comparison with the base case.

Other important issue to address are the control efforts from the centralized controller when a variable deload level is applied. This is important because the MICAPAS may add variability of power flows between the different sections of a PV-PP. Figure 6.7 shows the changes in the MICAPAS main signal, K_r, which changes the deload level in the PV sections. It is shown that, meanwhile in scenario S15C the signal K_r is constantly changing its reference, in scenarios S15Ramp and S15Steps it changes only in the first great fall and somewhere in the middle of the day, which is a better response for the controller.

Regarding the controller efficiency, it can be concluded that the ramp-type function for the deload level is more efficient, because in this case K_r is zero most of the time which means that all the power reserves are being deployed from the PV-PP.

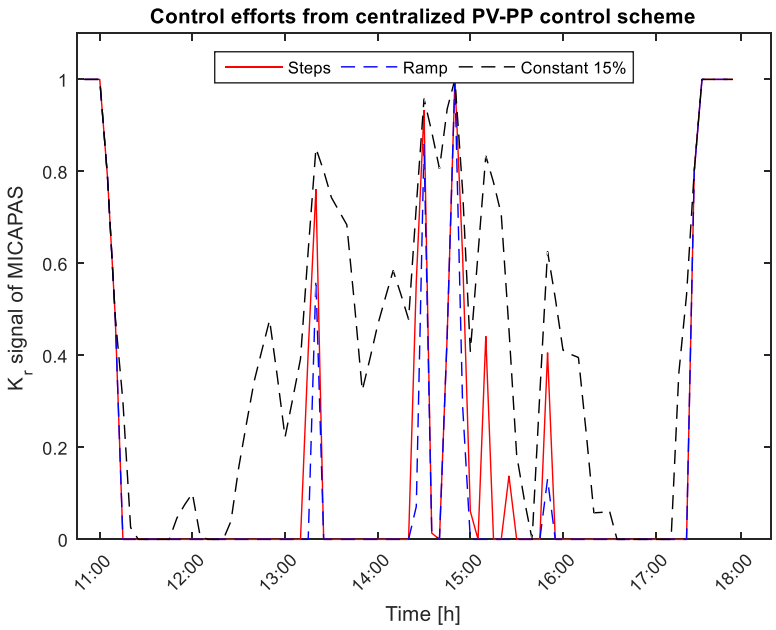


Figure 6.7: MICAPAS signal K_r during the day.

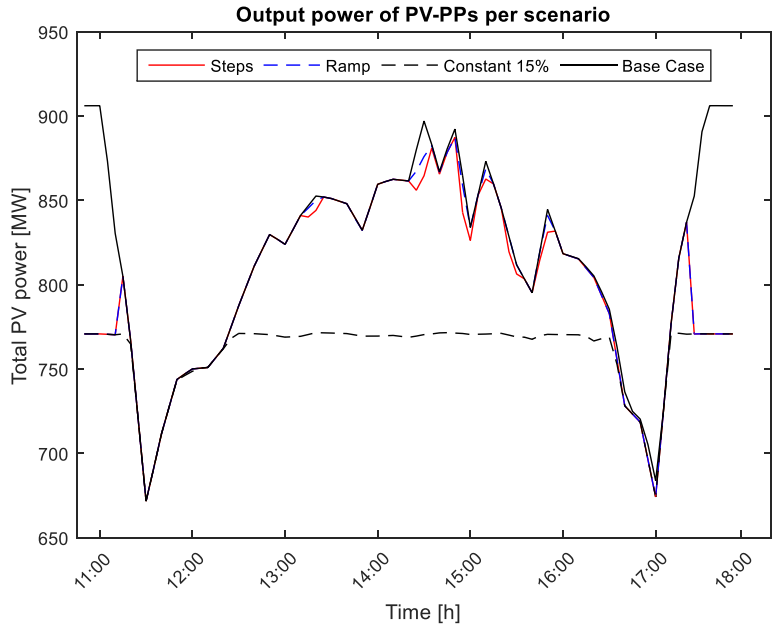


Figure 6.8: Total PV-PPs output power during the day per scenario.

What is shown in Figure 6.7 regarding the system's efficiency is confirmed by Figure 6.8 because it shows that, when a ramp-type function is used, the power differences are lower than in the step-type function. It is shown that in both variable scenarios most of the power reserves are being deployed constantly, while in the constant

level of deload the power losses are significant because a 15% of deload is not necessary all the time.

Finally, Figure 6.9 shows the 10-minute power ramps of the aggregated power of the conventional generation units. The NIS operator demand a ramping capacity, for each conventional unit of their system, of 2.7 MW/min [56]. This means that the 10-min power ramps of each conventional generation unit has to be 27 MW/10min. It is clear from the figure that the aggregated power of the conventional generation units surpass this ramping level in both variable scenarios, but the needed ramps from each unit is not, unless only one unit is capable of regulating frequency in the system. In S15Ramp the conventional units are over a 27 MW/10min ramping level a 5.1% of the time and in S15Step a 3.3% of the time, which means that there are not ramping problems when implementing this control strategy.

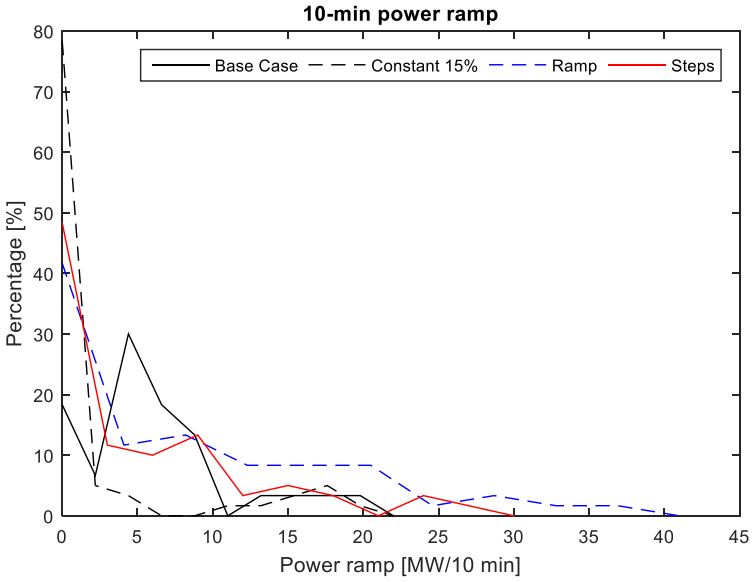


Figure 6.9: 10-min power ramps of conventional generation units.

Chapter 7.

Conclusions

This work develops an improvement to the MICAPAS scheme proposed in [3]. The proposed improvement included a deload level calculation every 10 minutes depending on a forecasting technique based on ANNs.

The first part of this work included the design and implementation of the ANNs for the forecasting technique to determine the deload level of the PV-PPs. The forecasting model was divided in two different stages: the first one forecasts the type of the next day by a CSI and the second one uses this information and forecasts the intra-hour radiation level. The results show an accurate forecast, with a NMSE between 1.8% and 2.4%. From the literature review it can be concluded that using one ANN per type of day is better than using only one ANN per PV-PP. It is also concluded that the needed data to have good results has to be composed by, at least, 5 years of data every 10 minutes. This is shown by the larger NMSE of the SLAR meteorological station, which only had 2 years of data available.

Since the CSI-ANN has forecasting errors, the GSR-ANNs have to take care of those errors. It is shown that the GSR-ANNs that were trained with enough data (>4 years) have good results in forecasting other type of days. This can be a particular conclusion in the case study because it corresponds to a desert climate with few abnormities caused by the Bolivian winter phenomena, but could be valid in other places if tested with enough data.

The second part of this thesis included the design and implementation of a control model in PV-PPs. The dynamic simulations showed that, when no control is implemented, the power system's frequency was outside the normal band a 24% of the day, whereas when a constant deload level is applied the system's frequency remained within the normal operation band throughout the day. When a variable deload level is implemented, the system's frequency remained between normal operation limits a 98.8% of the time, being the 1.2% an over-frequency. The over-frequency issue is solved by implementing an active power reduction loop in the controller, which is mandatory for some grid codes, as the Chilean one [57].

Regarding the energy losses, the simulations show that using a variable deload level harnesses solar energy by at least 3 times in comparison with the use of solar energy when a constant level is applied. Since the simulations are done during a worst-case scenario day regarding radiation ramps, it is clear that more reserves are being deployed throughout the day because there are differences between sections at almost all time. Therefore, a conclusion is that energy losses during a better day are even less in a

variable-deload control strategy because less reserves are being deployed during that day and only a 2% of deload is applied, contrary to the 15% of a constant-deload strategy. It is also clear from the MICAPAS signal results and PV-PPs output power that both variable-deload scenarios are more efficient than a constant deload level in energy usage, because the variable deload cases deploy all their reserves most of the time, contrary to the constant deload level scenario.

7.1. Future work

As future work, it is proposed to implement a previous step in the local controller which forecasts the next day's irradiance variability. Depending on the severity of irradiance variability the local control scheme of the PV-PP is enabled or disabled and therefore the energy losses are minimized.

Another interesting work is to implement the same control system but using energy storage instead of a deloaded operation of the PV-PPs, and analyze the economic impact.

Finally, this control scheme could generate power flows issues between the PV-PPs. Hence it is proposed to analyze these effects when applying this control strategy or not.

List of Acronyms

ANN	Artificial Neural Network
CSI	Clear-Sky Index
DC	Direct Current
DoY	Day of Year
GSR	Global Solar Radiation
GW	Giga-Watts
KDE	Kernel Density Estimation
MICAPAS	Mitigation Control Against Partial Shading
MPP	Maximum Power Point
MPPT	Maximum Point Power Tracking
MSE	Mean-Squared Error
MTM	Markov Transition Matrices
MW	Mega-Watts
NAM	North American Mesoscale
NEAT	Neuro-Evolution of Augmenting Topologies
NIS	Northern Interconnected System
NMSE	Normalized Mean Squared Error
NN	Neural Network
NWP	Numerical Weather Prediction
pdf	Probability density function
PV	Photovoltaic
PV-PP	Photovoltaic Power Plant
RH	Relative Humidity
SNR	Signal-to-Noise Ratio
STD	Standard Deviation
T	Temperature
TSO	Transmission System Operator

Bibliography

[1] T. Foley, K. Thornton, R. Hinrichs-rahlwes, S. Sawyer, M. Sander, R. Taylor, S. Teske, H. Lehmann, M. Alers, and D. Hales, *Renewables 2015 Global Status Report*. 2015.

[2] I. I. E. Agency, “Trends 2015 in Photovoltaic Applications: Survey Report of Selected IEA Countries between 1992 and 2014,” 2015.

[3] C. Rahmann, V. Vittal, J. Ascui, and J. Haas, “Mitigation Control Against Partial Shading Effects in Large-Scale PV Power Plants,” *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 173–180, 2016.

[4] F. Golestaneh and S. Member, “Batch and Sequential Forecast Models for Photovoltaic Generation,” in *IEEE Power & Energy Society General Meeting*, 2015, pp. 4–8.

[5] A. Mills, M. Ahlstrom, M. Brower, A. Ellis, R. George, T. Hoff, B. Kroposki, C. Lenox, N. Miller, M. Milligan, J. Stein, and Y. Wan, “Understanding variability and uncertainty of photovoltaics for integration with the electric power system,” Lawrence Berkeley National Laboratory, Berkeley, CA, USA, Rep. LBNL-2855E, 2010.

[6] A. D. Mills and R. H. Wiser, “Implications of geographic diversity for short-term variability and predictability of solar power,” *Power Energy Soc. Gen. Meet. 2011 IEEE*, pp. 1–9, 2011.

[7] A. Mills and R. Wiser, “Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power,” *Energy*, no. September, 2010.

[8] R. Shah, N. Mithulananthan, and K. Y. Lee, “Impact of Large-Scale PV Penetration on Power System Oscillatory Stability,” *Power Energy Soc. Gen. Meet. 2010 IEEE*, vol. 11, no. 1, pp. 1–7, 2010.

[9] J. Von Appen, M. Braun, T. Stetz, and D. Geibel, “Time in the Sun - The Challenge of High Pentration in the German Electric Grid,” *IEEE Power and Energy Magazine*, no. March/April 2013, pp. 55–64, 2013.

[10] Comisión Nacional de Energía, “Capacidad Instalada por Sistema Eléctrico Nacional,” 2016.

[11] Comisión Nacional de Energía, “Estadísticas en Electricidad - Proyectos de generación y transmisión en construcción SIC-SING (Planilla Excel).” Gobierno de Chile, 2015.

[12] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Pai, “A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output,” *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 917–926, 2014.

[13] B. Kirby and M. Milligan, “Facilitating Wind Development: The Importance of Electric Industry Structure,” *Electr. J.*, vol. 21, no. 3, pp. 40–54, 2008.

[14] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, “Irradiance forecasting for the power prediction of grid-connected photovoltaic systems,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 2, no. 1, pp. 2–10, 2009.

[15] A. Shah, H. Yokoyama, and N. Kakimoto, “High-precision forecasting model of solar irradiance based on grid point value data analysis for an efficient photovoltaic system,” *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 1–8, 2015.

[16] J. Liu, W. Fang, X. Zhang, and C. Yang, “An Improved Photovoltaic Power Forecasting Model With the Assistance of Aerosol Index Data,” *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 434–442, 2015.

[17] C. Chen, S. Duan, T. Cai, and B. Liu, “Online 24-h solar power forecasting based on weather type classification using artificial neural network,” *Sol. Energy*, vol. 85, no. 11, pp. 2856–2870, 2011.

[18] S. Rehman and M. Mohandes, “Artificial neural network estimation of global solar radiation using air temperature and relative humidity,” *Energy Policy*, vol. 36, no. 2, pp. 571–576, 2008.

[19] R. Dambreville, P. Blanc, J. Chanussot, and D. Boldo, “Very short term forecasting of the Global Horizontal Irradiance using a spatio-temporal autoregressive model,” *Renew. Energy*, vol. 72, pp. 291–300, 2014.

[20] P. Bacher, H. Madsen, and H. A. Nielsen, “Online short-term solar power forecasting,” *Sol. Energy*, vol. 83, no. 10, pp. 1772–1783, Oct. 2009.

[21] J. Shi, W. J. Lee, Y. Liu, Y. Yang, and P. Wang, “Forecasting power output of photovoltaic systems based on weather classification and support vector machines,” *IEEE Trans. Ind. Appl.*, vol. 48, no. 3, pp. 1064–1069, 2012.

[22] B. O. Ngoko, H. Sugihara, and T. Funaki, “Synthetic generation of high temporal resolution solar radiation data using Markov models,” *Sol. Energy*, vol. 103, pp. 160–170, 2014.

[23] M. Larrañeta, S. Moreno-Tejera, M. A. Silva-Pérez, and I. Lillo-Bravo, “An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series,” *Sol. Energy*, vol. 122, pp. 517–528, 2015.

[24] L. Hontoria, J. Aguilera, and P. Zufiria, “Generation of hourly irradiation synthetic series using the neural network multilayer perceptron,” *Sol. Energy*, vol. 72, no. 5, pp. 441–446, May 2002.

[25] F. D. Stacey and P. M. Davis, *Physics of the Earth*. Cambridge: Cambridge University Press, 2008.

[26] P. Mathiesen, J. M. Brown, and J. Kleissl, “Geostrophic wind dependent probabilistic irradiance forecasts for coastal california,” *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 510–518, 2013.

[27] K. Methaprayoon, C. Yingvivatanapong, W. J. Lee, and J. R. Liao, “An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty,” *IEEE Trans. Ind. Appl.*, vol. 43, no. 6, pp. 1441–1448, 2007.

[28] T. W. S. Chow and S.-Y. Cho, *Neural Networks and Computing: Learning Algorithms and Applications*. London: Imperial College Press, 2007.

[29] M. Buckland, *AI Techniques for Game Programming*, 1st ed. Cincinnati, Ohio: Stacy L. Hiquet, 2002.

[30] S. H. Ling, F. H. F. Leung, S. Member, H. K. Lam, Y. Lee, S. Member, and P. K. S. Tam, “A Novel Genetic-Algorithm-Based Neural Network for Short-Term Load Forecasting,” *IEEE Trans. Ind. Electron.*, vol. 50, no. 4, pp. 793–799, 2003.

[31] K. O. Stanley and R. Miikkulainen, “Evolving neural networks through augmenting topologies.,” *Evol. Comput.*, vol. 10, no. 2, pp. 99–127, 2002.

[32] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere, “Photovoltaic and Solar Forecasting: State of the Art,” *Int. Energy Agency Photovolt. Power Syst. Program. Rep. IEA PVPS T14*, pp. 1–40, 2013.

[33] N. Troldborg and J. Sørensen, “A simple atmospheric boundary layer model applied to large eddy simulations of wind turbine wakes,” *Wind Energy*, vol. 17, no. April 2013, pp. 657–669, 2014.

[34] Y. Li, Y. Su, and L. Shu, “An ARMAX model for forecasting the power output of a grid connected photovoltaic system,” *Renew. Energy*, vol. 66, pp. 78–89, 2014.

[35] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. New York: Springer-Verlag, 1996.

[36] R. Weber, “Time Series,” *University of Cambridge*, 2007. [Online]. Available: http://www.statslab.cam.ac.uk/~rrw1/timeseries/t.pdf.

[37] J. Grandell, *Time series analysis*. 2014.

[38] H. T. C. Pedro and C. F. M. Coimbra, “Assessment of forecasting techniques for solar power production with no exogenous inputs,” *Sol. Energy*, vol. 86, no. 7, pp. 2017–2028, Jul. 2012.

[39] R. J. Bessa, V. Miranda, a. Botterud, J. Wang, and E. M. Constantinescu, “Time adaptive conditional kernel density estimation for wind power forecasting,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 660–669, 2012.

[40] D. C. Ahrens, “Middle-Latitude Cyclones,” in *Meteorology Today: An Introduction to Weather, Climate, and the Environment*, Tenth Edit., Brooks/Cole, 2013, pp. 313–335.

[41] C. Voyant, M. Muselli, C. Paoli, and M. L. Nivet, “Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation,” *Energy*, vol. 39, no. 1, pp. 341–355, 2012.

[42] G. Papaefthymiou, P. H. Schavemaker, L. van der Sluis, W. L. Kling, D. Kurowicka, and R. M. Cooke, *Integration of stochastic generation in power systems*, vol. 28, no. 9 SPEC. ISS. 2006.

[43] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*. Wiley, 2008.

[44] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Second., vol. 1. 2009.

[45] G. Lera and M. Pinzolas, “Neighborhood based Levenberg-Marquardt algorithm for neural network training,” *IEEE Trans. Neural Networks*, vol. 13, no. 5, pp. 1200–1203, 2002.

[46] A. Bidram, A. Davoudi, and R. S. Balog, “Control and circuit techniques to mitigate partial shading effects in photovoltaic arrays,” *IEEE J. Photovoltaics*, vol. 2, no. 4, pp. 532–546, 2012.

[47] M. J. Reno, C. W. Hansen, and J. S. Stein, “Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis,” *SANDIA Rep. SAND2012-2389*, no. March, pp. 1–66, 2012.

[48] G. R. Walker and P. C. Sernia, “Cascaded DC-DC converter connection of photovoltaic modules,” *IEEE Trans. Power Electron.*, vol. 19, no. 4, pp. 1130–1139, 2004.

[49] S. Kouro, J. I. Leon, D. Vinnikov, and L. G. Franquelo, “Grid-Connected Photovoltaic Systems: An Overview of Recent Research and Emerging PV Converter Technology,” *Ind. Electron. Mag. IEEE*, vol. 9, no. 1, pp. 47–61, 2015.

[50] J. M. Carrasco, L. G. Franquelo, J. T. Bialasiewicz, S. Member, E. Galván, R. C. P. Guisado, S. Member, M. Ángeles, M. Prats, J. I. León, and N. Moreno-alfonso, “Power-Electronic Systems for the Grid Integration of Renewable Energy Sources : A Survey,” *Ieee Trans. Ind. Electron.*, vol. 53, no. 4, pp. 1002–1016, 2006.

[51] Q. Li and P. Wolfs, “A review of the single phase photovoltaic module integrated converter topologies with three different DC link configurations,” *IEEE Transactions on Power Electronics*, vol. 23. pp. 1320–1333, 2008.

[52] C. Rahmann and A. Castillo, “Fast Frequency Response Capability of Photovoltaic Power Plants: The Necessity of New Grid Requirements and Definitions,” *Energies*, vol. 7, pp. 6306–6322, 2014.

[53] D. Gmbh, “DIgSILENT PowerFactory Technical Reference Documentation PV System.” DIgSILENT GmbH, 2014.

[54] Comision Nacional de Energia Electrica, “Informe de Prevision de Demanda 2015-2030,” p. 50, 2015.

[55] S. S.r.o, “Solargis.” [Online]. Available: http://solargis.info/.

[56] D. S. J and J. C. A. T, “Estudio de Transmisión para la Inserción ERNC- Escenarios año 2018 Y 2021,” 2015.

[57] CNE, “Norma Técnica de Seguridad y Calidad de Servicio,” pp. 1–179, 2016.

[58] R. Perez, P. Ineichen, R. Seals, and A. Zelenka, “Making full use of the clearness index for parameterizing hourly insolation conditions,” *Sol. Energy*, vol. 45, no. 2, pp. 111–114, 1990.

[59] Task Committee on Standarization of Reference Evapontranspiration, “The ASCE Standarized Reference Evapotranspiration Equation,” 2005.

[60] K.-N. Liou, *An Introduction to Atmospheric Radiation*, Second Edi. Elsevier Science: Academic Press, 2002.

[61] M. de Energía, GIZ, and FCFM, “Campaña de medición del recurso Eólico y Solar.” [Online]. Available: http://walker.dgf.uchile.cl/Mediciones/.

[62] M. de Energía and F. Departamento de Geofísica, “Explorador de Energía Solar.” [Online]. Available: http://walker.dgf.uchile.cl/Explorador/Solar3/.

Appendix A.

Concepts and definitions

Clear-sky and clearness indices

The clear sky index k_t is defined as the ratio of the global irradiance I_{glob} to the expected irradiance under clear sky conditions $I_{clearsky}$ [14]:

$$k_t = \frac{I_{global}}{I_{clearsky}} \tag{A.1}$$

The clearness index K_t is similarly defined, but includes a normalization of the ground level irradiance with respect to the extraterrestrial irradiance [58], [22], as it is shown in the following equation:

$$K_t = \frac{I_{global}}{I_{ext}} \tag{A.2}$$

Extraterrestrial irradiance is easier to model than the clear sky irradiance due to the absence of physical atmospheric fluctuations. Extraterrestrial irradiance, or radiation, is the radiation that reaches the outer part of earth's atmosphere, and varies slightly throughout the year [47]. It can be defined as shown in equation (A.3) [47], [59].

$$I_{ext} = I_0 \cdot \left[1 + 0.033 \cdot \cos\left(\frac{2\pi}{365} \cdot DOY\right)\right] \cdot \cos(\theta_t) \tag{A.3}$$

where $I_0 = 1367\ W/m^2$ is the solar constant, DOY is the day of the year and θ_t is the solar zenith angle at time t.

The solar zenith angle at time t can be calculated by using the equations in [60]. The important variables are shown in Figure A.1.

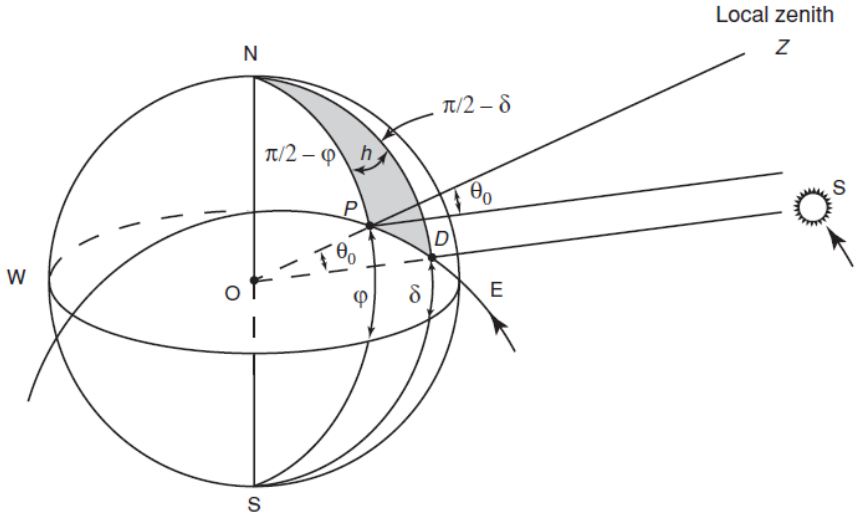


Figure A.1: Relationship of the solar zenith angle θ_0 to the latitude φ, the solar inclination angle δ, and the hour angle h. P and D are the point of the observation and the point directly under the sun, respectively [60]**.**

The solar zenith angle is given by equation (A.4). Here, the hour angle is given by $h = 2\pi t/\Delta t_{\odot}$, where $\Delta t_{\odot} = 86400$ sec and the latitude φ is in rads.

$$\cos\theta_t = \sin\varphi \sin\delta + \cos\varphi \cos\delta \cos h \tag{A.4}$$

On the other hand, a proper approximation of the sun's declination is given by [60]:

$$\delta = \sum_{n=0}^{3} (c_n \cos nt + d_n \sin nt) \tag{A.5}$$

Where the coefficients are:

Table A.1: Equation (A.5) coefficients.

n	c_n	d_n
0	0.006918	0
1	-0.399912	0.070257
2	-0.006758	0.000907
3	-0.002697	0.000148

Appendix B.

Generation of PV-PPs radiation data

In this work, the radiation data per PV-PP is generated with the available information of meteorological stations nearby. The information is public and it can be found in [61]. As every meteorological station were put into service in different years, there are meteorological station with as few data as 2 years. Because of this, the radiation data in every year, per meteorological station was pondered to get a one-year radiation time-series.

The nearby meteorological stations (the one that are placed in northern Chile) are seven. Figure 5.1 shows a simplified diagram of the Northern Interconnected System (NIS) of Chile, and the relative position of the PV-PPs and meteorological stations. There are 20 PV-PPs, but there are placed in 4 clearly defined zones (the ones shown in the figure).

Distance measures will be useful in order to generate different radiation series per PV-PP by defining a weight value that depends on the distance between the PV-PP and every one of the seven meteorological stations in the zone. Then, the meteorological station data every 10-min is multiplied by the corresponding weight and the sum of weights and values of the meteorological stations will generate a whole new series. This will be the PV-PP series.

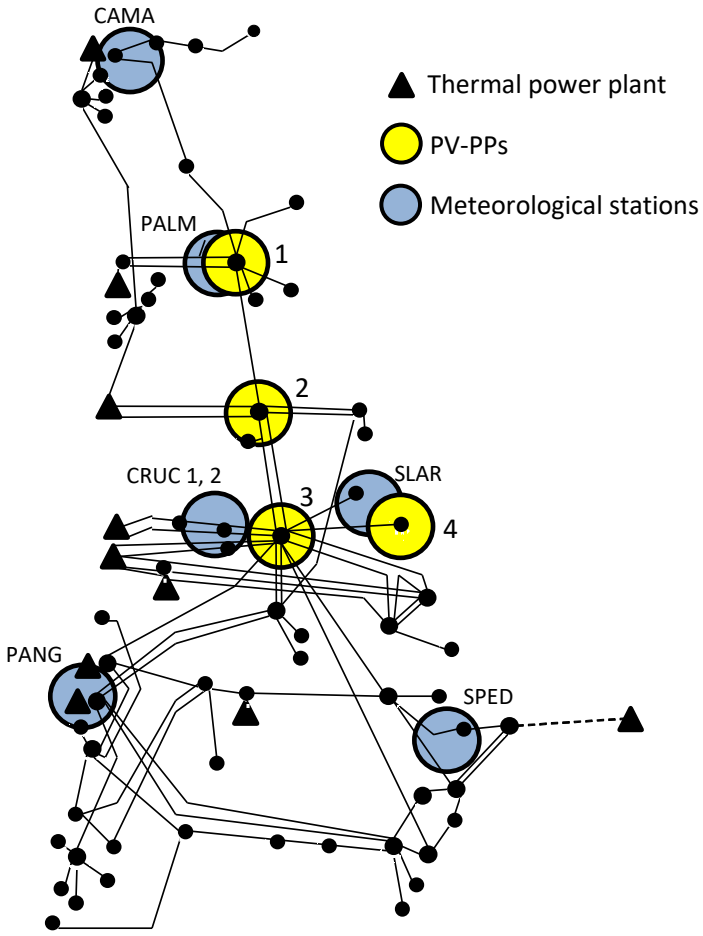


Figure B.1: Simplified diagram of the Northern Interconnected System (NIS) of Chile, including the estimated position of meteorological stations nearby.

For the distance measures to be computed, the UTM coordinates were calculated. Then, the distance (in UTM) between the PVPP number j ($PVPP_j$) and the meteorological station i (MS_i) is defined as:

$$d_{MS_i}^{PVPP_j}[UTM] = \sqrt{\left(N_{UTM}^{MetStat_i} - N_{UTM}^{PVPP_j}\right)^2 + \left(E_{UTM}^{MetStat_i} - E_{UTM}^{PVPP_j}\right)^2} \quad (27)$$

This distances are in UTM, and so they have to be normalized. As the data of the nearer meteorological stations is more similar to the PV-PP than the ones that are more far away, the normalization is done as follows:

$$d_{MS_i}^{PVPP_j}[p.u.] = 1 - \frac{d_{MS_i}^{PVPP_j}}{\max\left(d_{MS_i}^{PVPP_j}\right)} \quad (28)$$

Finally, the weights of each meteorological station per PV-PP have to sum 1. Then, they are defined as

$$W_{MS_i}^{PVPP_j} = \frac{d_{MS_i}^{PVPP_j}[p.u.]}{\sum_{i=1}^{7} d_{MS_i}^{PVPP_j}} \quad (29)$$

Being $GSR_{meas}^{MS_i}$ the measurements every 10-min of the meteorological station i, the radiation data of PV-PP number j ($GSR_{gen}^{PVPP_j}$) is defined as:

$$GSR_{gen}^{PVPP_j} = GSR_{meas}^{MS_i} \cdot W_{MS_i}^{PVPP_j} \quad (30)$$

Finally, this data was compared with a mesoscale numerical model that uses satellite images in order to ensure the pertinence of this methodology.

Although an ideal situation would be to have access to historical radiation and/or power output data of PV-PPs, that is difficult to achieve. That is why other methods to generate radiation data have to be used.

In order to know how well the proposed methodology works, a correlation analysis was done, where the generated data was compared with the data given by a mesoscale model that uses satellite information to get solar radiation in Chile. That model can be found in [62].

First, in order to ensure the pertinence of the mesoscale model, the measured data in the San Pedro de Atacama meteorological station was compared with the modeled data at the exact same place. The Spearman correlation between this data is shown in Figure B.1. The correlation between model and measured data is very high (over 0.9), and then the pertinence is confirmed.

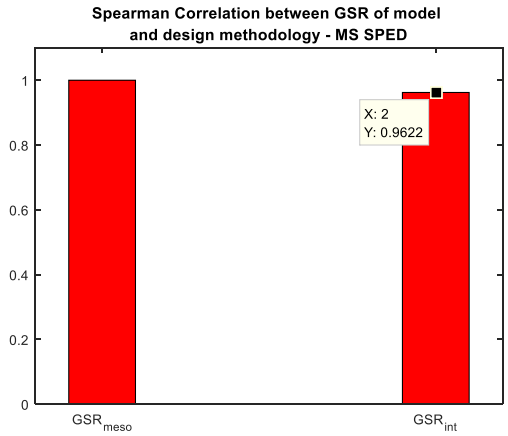


Figure B.1: Spearman correlation between GSR of mesoscale model and design methodology in San Pedro de Atacama meteorological station.

The second step was to ensure the correlation between the data of the mesoscale model and the one that was generated by the proposed methodology. The PV-PP Atacama Solar was used for this purposes, and the results are shown in Figure B.2. It is shown that the correlation is higher than the one shown in San Pedro de Atacama. This happens because the proposed methodology calculates the mean values of every year in order to get values of just one year, and then the errors regarding a mesoscale model are decreased.

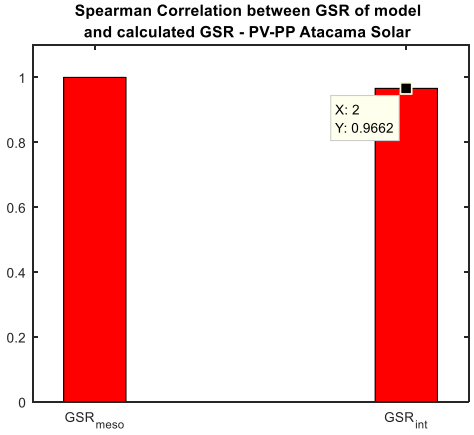


Figure B.2: Spearman correlation between GSR of mesoscale model and design methodology in PV-PP Atacama Solar.