Características espacio-temporales de la producción fotovoltaica en el área Euro-Mediterránea: análisis climático

Claudia Gutiérrez Escribano

Preface

Applied science has made societies evolved trough the linked between the basic science and the 'ingenios' that implement the structural knowdlege to the individual dailylife. The bridge between the fundamental studies and its evolution or transformation into something applied (useful) for the societies development is been a matter of interpretation or 'translation' (so interpreters/translator) that have been able to find the accurate language to exchange this knowdlege.

Transversality between the fundamental disciplines of physiscs, chemistry or biology and their different applied branches, most of them under the umbrella of ingenieering, end unavoidably into a transformation from the abstract ion to the tanglible with the premise or ideal of improving wellbeing of present and future societies.

However, it is very likely that this conception of the scientific application had lead to the most challeging problem that humanity has to face from its existance, climate change, being this anthropocentrism part of the problem and not of the solution.

In his book "El pensador intruso", Jorge Wasengber wrote about the evolution of science and the underlying progress what follows:

'There are above all two vices that tend to set 'pre-cooked' ideollogy into science. First is base on different kinds of anthropocentrism and consist in put the knowdlege subjet into the cosmos's center. The history of knowdlege is the witness: each time we 'barremos' the 'I' from the spot, the knowdlege progresses and only because of it.'

Contents

Abstract

Resumen de la tesis

Part I

Introduction

1

Introduction

> *"Begin at the beginning," the King said gravely, "and go on till you come to the end: then stop."*
>
> — Lewis Carroll, *Alice in Wonderland*

1.1   A changing world

Ideas a desarrollar:

* Un mundo en constante evolución. * Cambio climático * Transición energética global * Cambios sociales/políticos. * Migraciones

* Entrelazamiento de todas estas cuestiones. * El conocimiento puro + el conocimiento aplicado: obligación de la ciencia a ser en la medida de lo posible aplicada a solucionar problemas de las distintas sociedades. Justicia social. Más si estos problemas han sido creados por el ser humano.

1.2   Renewable Energies

Podemos definir el concepto físico de energía como la capacidad de un sistema para realizar un trabajo, clasificando las distintas formas de energía en dos grandes grupos: la energía cinética, relacionada con el movimiento del sistema, o la energía potencial, que atiende a la posición de mismo.

En esta clasificación fundamental encontramos que la energía mecánica, la energía electromagnética o la energía térmica son formas de energía que se engloban

dentro de la energía cinética, mientras que la energía química (de la cual es un sub-tipo la energía nuclear) o la energía potencial gravitacional (como es el caso de las centrales hidroeléctricas) se encontrarían dentro de la energía potencial.

Es elemental para nuestro estudio delimitar la diferencia entre formas y fuentes de energía. Estas útimas son definidas como aquellas a partir de las cuales la energía 'útil', aquella que es directamente aplicable para su finalidad sea esta el movimiento, la producción de electricidad o los distintos procesos metabólicos en el ser humano, puede ser extraída directamente o mediante algun proceso de transformación. Así, los combustibles fósiles son una fuente de energía, puesto que su energía 'útil' debe ser extraída. Es decir, extraemos la energía química contenida en los combustibles fósiles y necesitamos ciertos procesos de transformación entre la energía potencial de una masa de agua elevada que termina pasando a través de una turbina para obtener electricidad. Algunos ejemplos de fuentes de energía son los combustibles fósiles, la energía solar o la biomasa.

Aparece de manera natural la clasificación de estas fuentes de energía según su origen renovable, como la energía solar, eólica o geotérmica y no renovable como los combustibles fósiles. La energía renovable, es por lo tanto, aquella que tiene su ori-gen en estas fuentes inagotables de energía. El sol, a pesar de su indiscutible finitud, es considerado una fuente inagotable de energía debido a la diferencia entre la escala temporal de la vida humana y la vida de la estrella a la que nos refereimos, muchos órdenes de magnitud mayor.

El uso de la energía renovable ha acompañado el desarrollo de la humanidad desde tiempos remotos. Desde el uso de la biomasa para generar energía térmica para calentarse, hasta la utilización de la energía eólica transformada en energía mecánica en los molinos tradicionales o en la navegación. Sólo hacia la mitad del siglo 19, con la invención de la máquina de vapor, comenzó la utilización de los combustibles fósiles y con ello lo que se denominó la primera revolución industrial.

A pesar de constituir un periodo de gran desarrollo en términos tanto tecnológi-cos como de las sociedades y su bienestar, este lapso de tiempo es pequeño com-parado con la historia de la humanidad. Además, este desarrollo, a pesar de haber conseguido que los niveles de vida alcancen los más altos estándares en nuestra his-toria, este hecho no ha sido ni mucho menos global ni 'isótropo' (homogéneo).

1.3 Links between climate and renewable energy

1.4 Photovoltaic Energy

1.5 Climate change and the Mediterranean area

1.6 General scientific question: spatiotemporal behaviour of solar resource and photovoltaic production

1.7 Organization of the manuscript

2

State of knowdlege

Part II

Data & Methods

3

Data

In the present chapter, the datasets used in this manuscript are described. Different type of data is used in each of the studies, due to the fact that different problems are addressed.

It is well-known that the lack of a well spread solar radiation measurements network has been a constrain not only for the development of solar forecasting or resource assessment techniques, but also for the study of the whole atmospheric/climatic processes in which solar radiation takes places. However, the progress and the improvement in the satellite dataset has helped to overcome some of this issues.

3.1 Solar radiation measurements: *Baseline Surface Radiation Network, BSRN*

Solar radiation data from ground stations are not common at the publicly available metheorological stations network. One of the main sources of high quality data that can be easily obtain is the 'Baseline Surface Radiation Network', BSRN [**Konig-Langlo2013**].

The aim of this project is to detect changes in the Earth's radiation field that could be a consequence of climate changes [ref web]. The monitoring network provides high-quality and high-frequence data of short and log-wave radiation fluxesnfrom different stations around the globe which correspond to different climatic zones. In figure 3.1 all the available stations from BSRN are displayed.

The World Climate Research Programme (WCRP) Radiative Fluxes Working Group initiated the Baseline Surface Radiation Network (BSRN) to support the research projects of the WCRP and other scientific programs related to solar radiation.

3.2 Satellite data: *Climate Monitoring Satellite Application Facility, CM-SAF*

It has been previously commented that it is a well known problem the scarcity of available ground-based solar radiation measurements. As the preference alternative, satellite datasets has become one of the main sources of solar radition data because its high spatial and temporal resolution as well as its increasing accuracy.

a partir de aquí copiado

The Satellite Application Facility on Climate Monitoring (CM SAF) develops, produces, archives and disseminates satellite-data-based products in support to climate monitoring. The product suite mainly covers parameters related to the energy water cycle and addresses many of the Essential Climate Variables as defined by GCOS (GCOS 138).

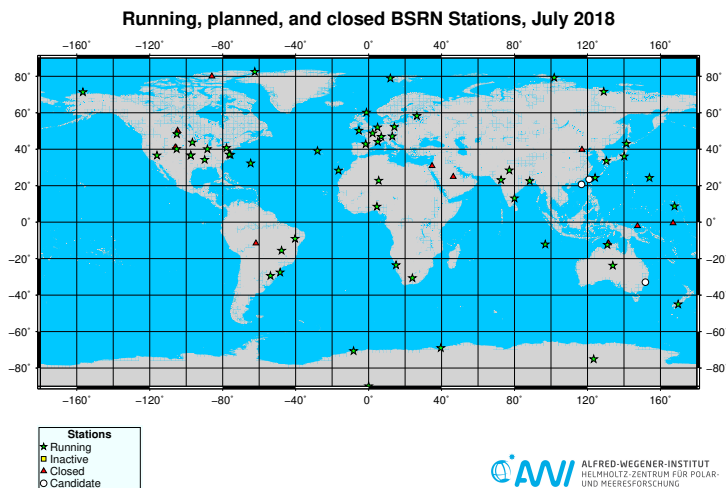The CM SAF produces both Enviromental Data Records and Climate Data Records.

FIGURE 3.1:

Environmental Data Records (EDR) are time-tagged earth-located geophysical parameters produced from sensor data. EDRs are derived in low to medium latency not fulfilling strictest climate requirements.

Climate Data Records (CDR) are time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change.

The CM SAF products are derived from several instruments on-board operational satellites in geostationary and polar orbit, i.e., the Meteosat and NOAA satellites, respectively. The existing data sets will be continued using data from the instruments on-board the new EUMETSAT Meteorological Operational satellite (MetOP). The products have mostly been validated against several ground-based data sets both in situ and remotely sensed. The accomplished accuracy for products derived in near real time is sufficient to monitor variability on diurnal and seasonal scales. Products based on intercalibrated radiance data can also be used for climate variability analysis up to inter-annual scale. A central goal of the recently started 2nd Continuous Development and Operations Phase of the CM SAF (2012–2017) is to further improve all CM SAF data sets to a quality level that allows for studies of inter-annual variability.

3.3   Modelling solar radiation

Atmospheric variables can be obtain as a product of meteorological or climate models.

3.4   Photovoltaic prodution data

3.5   Other datasets

- ECAD dataset for temperature
- Bases de datos de eurocordex y medcordex

4

Methods

In order to fullfill the objectives of the present work it is necessary to apply different methodologies in each of the results chapters, all of them has been explained in detail at the corresponding section.

However, in the present chapter the different tools needed for the development of each study are described in a general manner. The three result's chapters are based in a modeling chain that includes a photovoltaic model that is composed of two basic steps. First, the transposition to the POA of the components of the solar radiation, which implies also to assume some simplification models of the atmosphere and a model for the electrical performance of the system. Besides, the input of this PV model can be solar radiation measurments, satellite radaition stimations o as we saw in the previous chapter, the output of atmospheric models.

In chapter 5 it is analysed the interannual variability and complementarity of solar resource over the Iberian Peninsula and for that purpose we apply **clustering techniques**. The regionalization allow us to simplify the spatio-temporal analysis of solar resource. The evaluation of the productivity, defined as the amount of energy produced by a PV system normalized by the power capacity of the system, implies the **modelization of the photovoltaic system**.

The chapter 6 analyses the impact of **aerosols** in photovoltaic productivity over the Euro-Mediterranean area. Some climate simulations are used as the input of the photovoltaic model. The modelling chain allows to make a sensitivity test to quantify the role of aerosols in the area.

Finally, the photovoltaic energy potential is analysed in the future, using **climate projections** from different climate models. Trends and anomalies with respect to a reference period are evaluated. In this case, a **multi-model analysis** through different RCMs simulations allow us to evaluate the solar resource under climate change scenarios. In this point it is also consider the representation of aerosols in the climate projection as a fundamental variable for the shortwave downward radiation projections, SSR, and PV productivity.

4.1 Clustering algorithm applied to climate data

In a data-driven world, **pattern recognition** is being applied to many disciplines from biology to finanze through social science, in order to obtain relevant information from different datasets. However, if the similarity measurements can not be define first, due to unknown previous labeled data, another approach will be implemented. In that case, the task will consist in identifying the similarities afterwards, from a dataset of features, applying clustering algorithms for that purpose.

4.1.1 Clustering algorithms

It is not under the scope of this work to analayse in detail all the clustering methods available, due to the vast number of them and its increasing complexity [ref:data clusterng a review]. However, it is worth it to give an overview of its classification and application in order to better understand the choice made in the study of our problem.

There are different types of clustering algorithms based on its application and the criteria applied to construct the clusters. The two main categories are divided in **hierarchical** and **non-hierarchical algorithms**, but there are other taxonomies based on the algorithm construction that can be used to classified different methods [ref].

The **hierarchical clustering algorithms** create a number of nested clusters and they can be also divided themselves in agglomerative or divisive. The first one, obtain a smaller number of cluster in each step, whereas the divisive algorithms work on the other direction.

Most of these hierarchical algorithms are variants of the single-link or complete-link algorithms. They have a different way to measure similarity. In the first case, the distance between two clusters is the minimum of all the pairwise distances measured between the objects of the two clusters, in contrast, for the complete-link algorithm the distance is considered the maximum of all the pairwise distances.

On the orther hand, the non-hierarchical or **partitional clustering algoritm**, create a number of non-overlapping clusters. These methods could be useful when the amount of data involved is large and the construction of a dendogram, produced by the hierarchical algorithm could be computational expensive.

The partitional methods create the group of clusters using an **optimization function** which defines the similarity. The main inconvenient of these algorithms is the in-advanced definition of the number of clusters. Usually, the algorithm is implemented several times for a wide number of partitions and the optimum number of clusters is defined afterwards using validity index criteria.

4.1.2   Clustering climate data

As previously commented, the use of pattern recognition and in particular, the clustering analysis, has been widely used across many different disciplines. The use of these techniques to group together atmospheric variables that could help in environmental classifications is a more recent research field.[ref]

Historically, the climatic divisions were based mostly on the differences in vegetation types around the globe. The Koppen-Terawata classification [ref] is the most widely known climate classification and bases its divisions in obtaining the vegetation thresholds from precipitation and temperature data. Some authors had already pointed the limitations of this classification method mostly due to two aspects: the first one is that vegetation thresholds are not well defined and that could be an issue when higher spatial resolution scales want to be defined. On the other hand is the fact that not only precipitation and temperature are influencing the vegetation species but also other atmospheric variables like solar irradiation, as well as other environmental factors that could be related to antropoghenic emissions or waste.

As these classical division are not adapted to the necessities of different fields and could be even biased, another way to geo-spatial classification is needed. In this sense it has become frequent to sucessfully use data-driven classification of different variables if it is necessary to give an spatially resolved answer. [ref]

4.1.3   Applied clustering method

In this work a clustering method is applied to classify solar irradiation due to the fact that classical climate divisions are only based on temperature an precipitation. The spatial pattern tha could be extract from the Koppen-Terawatta classification, can not be used for our pourpose.

A partitional commonly used clustering method has been selected to classify solar irradiation of the area. The **'k-means'** algorithm is easy to implement, due to the fact tha has been largely used over the literature. Moreover, the combination of a **principal component analysis** of the dataset previous to the k-means algorithm application has been proved to be an useful tool to reduce the data-dimensionality and apply the algorithm in a more efficient manner.[ref]

K-Means algorithm

The k-means algorithm is a partitional clustering method that provide a set of partitions from the application of an optimization function, usually the euclidean distance, 5.1. The algorithm initiates from a pre-defined number of clusters, 'k', and randomly selects a group of centroids equal to the number of clusters. The optimization function is applied to assign each object to one of the clusters, depending on the similarity (distance) to each of the centroids. This proccess is applied until the algorithm converges.

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} ||x_i - c_j||^2 \tag{4.1}$$

One of the limitations of these method is its dependency on the first selection of the centroids, that could lead to a local optimum instead to a global optimum. To solve these difficulty, the algorithm can be run many times to test the sensitivity of the algorithm to a different initial conditions. Another option is to use an initialization technique to select the centroids.

As the number of clusters 'k' is not normally known in advance, the algorithm is applied from 2 to 'n' times, with 'n' high enough to get the whole variability of the dataset and the optimum number of clusters is defined afterwards. This optimum 'k' will define the number of clusters that explain most of the variability and will assure that the increase in the number of clusters does not improve the variance representation. [ref]

Principal Component Analysis

The Principal Component Analysis, PCA [ref], consists in a decomposition of the dataset into a number of vectors whose linear combination represents the original data. The transformation of the original data into a lower dimensional space is a reduction of the dimensionality retaining the maximum of the data variance.

The new orthogonal system has in its first coordinate axis, the projected values of the original data that preserve most of the variance in the dataset, and it is called the principal component. The rest of principal components decrease the amount of variance explained consecutively.

The PCA has been applied in advance to K-means clustering algorithm among the literature [ref]. It was explained by [ref] that the reduction in dimensionality is directly linked to K-means due to the fact that the clustering membership indicators are actually the eigenvectors given by the PCA. For this reason, it has been applied before the K-means algorithm fastening the algorithm.

Validity Index

Clustering validation through the validation or validity index measures the goodness of the clustering results. As well as the clustering techniques, there is a classification for the methods used to validate the clustering results.

External validation techniques use information outside the dataset involved, whereas the internal validation techniques only need the information that is present in the data. The external methods knows the number of optimal clusters in advance, so they are applied to select the best clustering algorithm. On the other hand, the internal validation methods will give us the optimum number of clusters after the application of the selected algorithm and only use the information that rely on the data.

Two different index are used in this study, the **Calinski-Harabasz**[1] index (Eq.:4.2), the **Davies-Bouldien**[2] index (Eq.:4.3), and the L-method proposed by Salvador and Chan, 2005 [ref]. Similar results are obtained with them. In order to analyse the optimum partiton, is important to consider the nature of the variables that we are analysing. Most of the atmospheric variables, are continious variables that could be closely related to other variables such as latitude. For that reason, the regionalization procedure can not be applied like for other discrete or non-continuous data. That characteristics should be considerd when the results have to be evaluated.

---

[1]In the equation 4.2: The 'BCSM' is the Between-Cluster-Scatter-Matrix, and its trace is the sum of squares distances between each cluster center c_i and the global centroid vector of all the objects of the dataset. The term 'WCSM' is the Within-Cluster-Scatter-Matrix and the trace of this matrix is the sum of squares of the distances between the objects inside each cluster and the centroid. 'k' is the number of clusters

[2]In the equation 4.3: d_i is the averaged distance between the data clasiffied to class i and the clauster center c_i and $d(c_i, c_j)$ is the distance between the different cluster centers. 'k' is the number of clusters.

$$CH = \frac{trace_{BCSM}}{trace_{WCSM}} \times \frac{n-k}{k-1} \tag{4.2}$$

$$DB = \frac{1}{k} \sum_{i-1}^{k} \max_{i=1,\dots i \neq j} \frac{d_i + d_j}{d(c_i, c_j)} \tag{4.3}$$

4.2 Simulating a photovoltaic system

The simulation process of a photovoltaic system has two main steps. First, it is necessary to estimate the amount of energy that reach solar cells to be transformed into electricity. This energy will be defined as global effective irradiation, $G_{eff}(\alpha, \beta)$. Secondly, the amount of energy that the system is able to give depends on the electrical performance of its components, which is the second part of the modeling process.

Both stages involve some modeling assumptions and each one would be explained in detail in the following sections. In general, due to the fact that it is very unlikely to obtain solar radiation measurements in the plane-of-array irradiarion ($G(\alpha, \beta)$), global horizontal irradiation ($G(0)$) is the most common variable measured or modeled that can be obtain from different sources. In order to estimate the amount of energy that reach the generator surface, $G(\alpha, \beta)$, it is necessary to do a transposition of the different radiation components from the horizontal plane $G(0)$ to the plane of the array (POA).

Once the solar irradiation components at the generator surface are calculated, it will be assessed how much of that energy can be transformed into electricity. The relative position between the generator and the sun makes that some optical losses could appear. After considering the optical losses the gplane of array irradiation becomes the global effective irradiation $G_{eff}(\alpha, \beta)$.

The second step that will calculate the energy output will indicate the performance of the electrical components from the global effective irradiation and other variables that can influence on cells, like ambiance temperature. The whole photovoltaic system includes the generator, the inverter and the transmission elements and wires.

All the processes involved in the stage of simulating a photovoltaic system in this thesis have been made using solaR [ref], an R package that implements all the necessary functions to estimate the energy provided by the sytem.

4.2.1 Global effective irradiation

Assessing the amount of energy reaching cells of the photovoltaic generator requires to compute trackers movements, as well as the relative position between sun and panels throughout the year, a detailed description of these methods can be found in Perpinan.Marcos.ea2013. Once these equations are computed, the procedure is described below:

- First, global irradiation on the horizontal plane is decomposed in two components, **direct** and **diffuse** irradiation, Eq.4.4. To estimate these quantities it is necessary to consider equations proposed by [**Liu1960**] to characterize solar irradiation. Definition of *clearness index*, Eq.4.6, is the ratio between global irradiation and extra-terrestrial irradiation at the horizontal plane. They also proposed to relate that index with the *diffuse fraction*: the ratio of diffuse to global irradiation in Eq.4.5. This relation varies depending on the time scale. For daily values, we estimate correlation between the clearness index and the diffuse fraction using equations in [**Aguiar1992**]. After that, the diffuse component is obtained with the definition of the *diffuse fraction*, Eq.4.5.

$$G_d(0) = B_d(0) + D_d(0) \tag{4.4}$$

$$F_{D,d} = \frac{D_d(0)}{G_d(0)} \tag{4.5}$$

$$K_{Td} = \frac{G_d(0)}{B_{0d}(0)} \tag{4.6}$$

- Secondly, daily irradiance has to be estimated from irradiation values. Considering low variability of solar irradiance, it is assumed that average irradiance in a short interval coincides with irradiation in that interval. Regarding equations proposed by [**Aguiar1992**], the ratio of the diffuse irradiance to diffuse irradiation is assumed to be equivalent to the ratio of extraterrestrial irradiance to extraterrestrial irradiation Eq.4.7, and the ratio of global irradiance to daily global irradiation is followed from the same reference Eq.4.8.

$$r_D = \frac{D(0)}{D_d(0)} = \frac{B_0(0)}{B_{0d}(0)} \tag{4.7}$$

$$r_G = \frac{G(0)}{G_d(0)} = r_D \cdot (a + b \cdot \cos(\omega)) \tag{4.8}$$

- The third step considers only geometrical criteria to compute the direct and diffuse irradiance components at the inclined plane. Diffuse irradiance is calculated with the anisotropic model proposed in [**hay1985estimating**].

- The last step estimates the effective irradiance incident on a generator subtracting dust and angle of incidence losses from the incident irradiance with the model proposed in [**Martin2001**].

In figure 4.1 the steps to assess effective irradiation at the plane-of-array is summarized.

4.2.2 Photovoltaic energy yield

Once the effective irradiation that reach solar cells is being assessed, the transformation into power output depends on some factors regarding the photovoltaic system. In order to estimate potential for photovoltaic production, the term yield is defined as the energy produced by the power installed [$\frac{k}{k}$]. That energy, comes from the integration in each time step of the power output of the photovoltaic system.

Considering that reach photovoltaic generator is composed of modules, the generator nominal power output is calculated by multiplying the power output of a module by the number of modules, assuming the same electric performance of all modules.

$$P_{out} = I_m \cdot V_m \tag{4.9}$$

Ambient temperature influence cells performance. The assumption used to this assessment consider a linear relationship between cell temperature and global effective irradiation, Eq.4.10. NOCT in equation 4.10 is considered constant being the temperature of a cell when it works in determined conditions: irradiance of 800 [$\frac{W}{2}$] and ambient temperature of 20žC.

$$T_c = T_a + G_{ef} \cdot \frac{NOCT - 20}{800} \tag{4.10}$$

After considering all these factors, the power output of the whole system is assessed by the consideration of a common inverter for transforming DC current into AC, also arrangement losses of the generator are included. Other systems factors that influence the performance of the photovoltaic system are shadows over the generator due to the positions of the PV modules over the land. This factor is not consider in our calculations assuming that we look for an estimation of the potential yield of an area, not the product of a real PV plant. Detailed description of the software employed to the assessment, solaR, is in [**Lamigueiro2012**]. The process for the photovoltaic
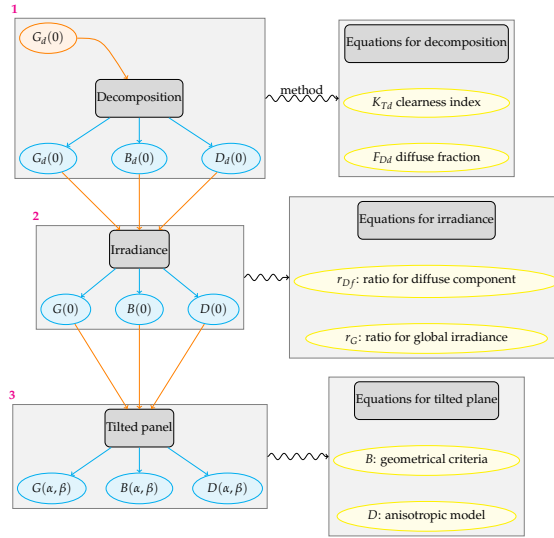
FIGURE 4.1: *Algorithm scheme: steps on the calculation of global effective irradiation, incident irradiation at the inclined plane. Orange color means input of the calculation and blue color output or results. If a result in a previous step is used in the next one, arrows linking steps are orange. Right side of the scheme represents the method and equations needed.*

output is summarized in table 4.1 including all the steps and elements involved as explained in [**Perpinan2009**].

4.3 Using Regional Climate Models

La cadena de modelado necesaria para obtener una estimación de la producción fotovoltaica puede incrementarse introduciendo como input una estimación o una salida de otro modelo (meteorológico, climático...) en lugar de observaciones. Este paso aumentará la incertiumbre de la potencia entregada por el sistema, pero merece la pena tomarla en consideración teniendo en cuenta la escasez de estaciones de medida que proporcionan datos de radiación.

En nuestro contexto, en el cual la longitud de las series temporales es clave[3], los modelos climáticos son una herramienta que permite evaluar el recurso presente, su

---

[3]el estudio climático es la evaluación del comportamiento medio de las distintas variables a lo largo del tiempo, su variabilidad y tendencias. Para poder definir una climatología se considera en consenso que son necesarios 30 años da datos.

Element	Method
PV generator	Identical modules with $dV_{oc}/dT_c = 0,475\frac{\%}{°C}$ and $NOCT = 47°C$. The MPP point calculated as in [**garcia2005caracterizacion**]).
Inverter	Efficiency equation proposed in [**jantsch1992results**]: $$\eta_{inv} = \frac{p_o}{p_o + k_0^o + k_1^o p_o + k_2^o p_o^2} \qquad (4.11)$$ where $p_o = P_{ac}/P_{inv}$ is the normalized output power of the inverter. The characteristic coefficients of the inverters are: $k_0^o = 0.01$, $k_1^o = 0.025$, $k_2^o = 0.05$.
Other losses	- Average tolerance of the set of modules, 3%. - Module parameter dispersion losses, 2%. - Joule losses due to the wiring, 1.5%. - Average error of the MPP algorithm of the inverter, 1%. - Losses due to the MV transformer, 1%. - Losses due to stops of the system, 0.5%.

TABLE 4.1: *Calculation procedure for the estimation of energy produced by a PV system from daily global horizontal irradiation data. Left column represents the element of the PV system and the right column the equations and methods used in each case for the efficiency of the elements.*

evolución futura así como server de input en la cadena de modelado para estimar la potencia entregada por un sistema fotovotaico.

The history of Climate modeling has been linked to the computational development since its beginning. The mathematical representation of the climate system was a consequence of the advance in numerical weather prediction models that took place for the first time in Princeton around 1950.

Climate modeling is based on a 3-D representation of the whole climatic system, where the Earth is divided in a 3-D spatial grid whose size or resolution has evolved in time with computational power. Equations of motion, momentum and conservation laws are solved for each gridbox. As some of the atmospheric processes occurs in a finer spatial scale than the resolution of the model, parametrizations are needed for some processes such as convection. These models were first named as General Circulation Models, GCM, as their aimed was to represent main circulation flows in the atmosphere, but the development of more complex models that would start to include ocean and land, would name later these models as **Global Climate Models, GCMs**.

It has been almost X years since the first run of a climate model. In this time it has been a diversification and more research groups has developed its own models. Also, many international framework initiatives has unify efforts to understand better the future evolution of our climate with the intensification and promotion of research collaboration programs that include all the models (WCRP, IPCC, CLIVAR...)

Although horizontal resolution of GCMs has evolved since its origin from X° to the new generation of models that are close to X°, for many years GCMs were not suitable enough to study some regional climate processes that affect in a particular manner to smaller areas.

It was in the early 90s (Dickinson 1989 and Giorgi 1990) when it was proposed to use global models as the necessary boundary conditions to force a 'limited area model LAM'. Those LAMs had been used for numerical weather prediction forecast, but they were applied to predict the weather just few days in advance. The idea of run these simultaions for longer periods will result in the evolution of **Regional Climate Modeling**, what would provide regional climate information of processes that occurs in a spatial scale not resolved by a global models.

In a similar way than the GCMs, RCMs community has expanded in last decades and the number of research groups has increased. Also, some international initiatives try to engage different groups and modellers to develop better models and simula-

tions (PRUDENCE, ENSEMBLES,CORDEX, EURO-CORDEX, MED-CORDEX).

In the present work we make use of RCMs focused on the Mediterranean area, considering the models and simulations included in the EURO-CORDEX and MED-CORDEX projects.

These RCMs can provide the necessary atmospheric variables for the analysis of renewable energy resources. Due to its complexity and the need of use parametrizations for some atmospheric processes, they can have sistematic bias that make difficult to use them for resource assessment. However, due to the low frequency variability observed in some variables and to the importance of considering the future projections, they are a valuable tool for analysing resources and its evolution. Besides, the use of climate models allows to study specific climatic events and to understand its mechanisms, linking the implications of those situations with renewable generation.

For chapter 7, "Impact of aerosols in photovoltaic energy production", only one RCMs is used. Three different simulations are used in this chapter in order to quantify the sensitivity of PV energy production to changes in atmospheric aerosols content. The simulations are centered in the Mediterranean area, with the domain described by the MED-CORDEX initiative (picture?).The simulations length and characteristics of the aerosols dataset included are explained in detailed in the corresponding chapter.
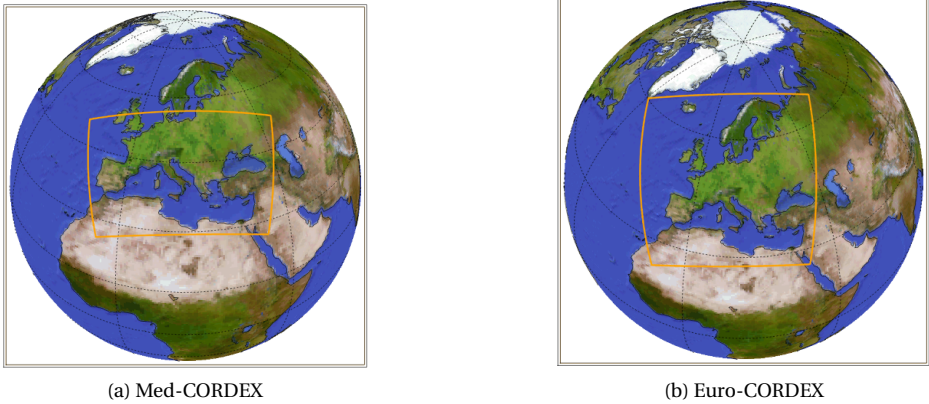

(a) Med-CORDEX   (b) Euro-CORDEX

FIGURE 4.2: *Domains from Med-CORDEX and Euro-CORDEX. Images from wRCP-CORDEX website.*

4.4 Future projections and scenarios

Behind the comprehensive study of the climate system it is not only la voluntad del conocimiento puro, sino también la preocupación por un sistema cambiante cuyos elelmentos influyen en el bienestar de los distintos ecosistemas y especies en la Tierra.

The Intergovernmental Panel on Climate Change, IPCC, was launched in 1992 and it is an international organism that was founded by the World Meteorological Asociation, WMO, and the United Nations Environmental Program, UNEP. Its purpose was to gather together all the scientific information available about the anthropogenic climate change. This information will summarize our knowdlege about climate change, its impacts as well as some mitigation strategies were elaborated.

In order to elaborate the climate scenarios, which means the possible evolution of the climate system based on the actual climate conditions and models, it is necessary also to generate different plaussible scenarios of emissions. The development of this emissions of greenhouse gases or aerosols will be based on the assumptions about technological changes or demographic criterias as well as the socioeconomic development of each region. Those emissions scenarios, with different options for the future evolution, will be the driving force for the climate simulations in order to elaborate the climate projections.

The first emissions scenarios were elaborated in the 90s and they have chaged since then. For the 5th AR, that was released 2014 were defined the RCP scenarios. They are different concentrations pathaways.

For chapter 8, "Future projections for PV technology", several RCMs and its correspondin GCMs are used. The simulations include the historical period and the future scenario RCP8.5 or RCP4.5. The analysis of the results will be done with respect to a reference historical period.

In table 4.2 there is a summary of the climate models and the simulations that are used in chapter 7 and 8.

In table 1. there are summarized the RCMs and different simulations used in the thesis. For chapter 7, the same RCMs is use with different representation of the aerosols. In chapter 8 it is analysed the future projections of different RCMs for the variables involved in photovoltaic energy production.

Study	Climate Model		Domain	Simulation
	GCM	RCM		
Aerosols' impact	CNRM-CM5	CNRM-ALADIN53	MED-CORDEX	AER
				NO-AER
				TREND
Future projections	CNRM-CM5	ALADIN53	EURO-CORDEX	HIST/RCP85
		RCA4		HIST/RCP85
		CCLM4		HIST/RCP85
	ICHEC-EC-EARTH	RACMO	EURO-CORDEX	HIST/RCP85
		RCA4		HIST/RCP85
		CCLM4		HIST/RCP85

TABLE 4.2: *Summarize of the climate models (global, GCM, or regional, RCM) used in chapters 7 and 8. The name of the global model includes the institute that elaborate the simulation and afterwards the global models's name. The domain of the simulations is MED-CORDEX or EURO-CORDEX, described in the referenced paper.*

Part III

Results

5

Long-term characteristics of solar resource over the Iberian Peninsula

Abstract

33

The spatio-temporal characteristics of different renewable resources like solar radiation, wind and precipitation are very different among them and a detailed understanding of each is important for an adequate planning and management of the electricity system. In this chapter, we elaborate a comprehensive methodology to analyse variability and complementarity of PV production that can be applied for long-term energy questions and for climate scales, but it is not limited to those time-scales or to solar resource thanks to the flexibility of the method.

The work is focused on solar irradiation as source of energy for photovoltaic (PV) generation over the Iberian Peninsula, IP, and on photovoltaic productivity, which is defined as the amount of energy generated normalized by the installed capacity at each location. The IP is an interesting site due to the fact that it is a large area with limitations in the interconnectivity with the rest of the European continent. It is also a complex climatic area [] due to its geographical location and its orograhy, which makes it interesting for testing the methodology.

The main steps of the method are the application of an objective clustering algorithm for performing a regionalization of the whole domain, the analysis of the temporal variability of solar radiation and photovoltaic energy yield, and the intercomparison of the obtained clusters for examining their complementarity.

The approach applied in this chapter means a different perpective to analyse the long-term issues in solar resource as the interannual variability, the influence of the large-scale circulation modes, trends or the synoptic patterns related to its variability. The use of the multi-step scheme allows to simplify the spatial analysis and answer the questions from a more applied point of view.

Data from a satellite dataset of 30 years, which is usually considered as the time for defining a climatology, is used inthis chapter and a long-term overview of solar radiation variability is analyzed over the Iberian Peninsula (IP). The spatial distribution and variability of photovoltaic (PV) power yield is calculated for different tracking systems. The variability is analyzed on an interannual time scale, which is relevant for energy supply security and year-to-year price stability. It shows robustness and stability of solar radiation and PV production on average for the whole domain, but with significant differences among clusters that could allow spatial compensation of PV production. The relationship between the variability of solar irradiation and of PV yield is not uniform among the different clusters. Areas where solar irradiation is higher are more sensitive to tracking type.

The whole process described in this chapter provides the information of how solar resource and the PV energy yield perform in a limited area and provide the tools to analyze the relationships between sub-areas and their variability. In this sense,

this method can be applied for isolated or nearly isolated electric systems located in regions with a variety of climates, or for interconnected systems involving several countries.

5.1 Introduction

The natural variability of renewable energy resources like solar radiation and wind presents some challenges for the management of electricity systems, which were designed for conventional technologies like nuclear or thermal power plants. For that reason a thorough knowledge and understanding of space-time features of solar radiation is needed. In the case of solar PV energy, its variability Widen2015 can be studied from the perspective of the resource or from the perspective of the PV power output, that includes some aspects of the PV generators involved in variability, like inverters or tilted and tracking panels which increase the complexity of the assessment. There are many studies focused on the short-term variability Zamo.Mestre.ea2014 that analyzed PV production ramps due to changes in solar incident irradiation associated with cloud motion Cros2014, IEA-PVPS-T14-1.32015. Also, the smoothing effect that a well-spread site planning has on the PV production is being investigated Marcos2012, Perpinan.Marcos.ea2013.

Not only short-term scales are important to address renewable resources intermittency but also longer time scales are relevant in order to make the system more efficient and reliable Davy2012. Policymakers, transmission system operators and investment companies, need an accurate evaluation of resources availability in present and future climate conditions for their mid and long-term planning. The analysis of interannual variability has a particular importance in order to assess stability of the resource and the financial viability of renewable energy plants pryor2006inter, as well as the likelihood of strong electricity price oscillations like the ones associated for example with the large interannual variations of hydroelectric production.

Regarding that perspective of long-term variability of solar resources, there are studies focused on long series from stations Sanchez-Lorenzo2009, Sanchez-Lorenzo2013, vazquez2012interannual or reanalysis data that identify low frequency changes in solar radiation, as the "dimming" and "brightening" periods Wild2005, that show relationship between solar irradiation and anthropogenic aerosols Nabat2014a. Some studies examine the influence of large-scale circulation atmospheric modes like the NAO (North Atlantic Oscillation) on solar radiation Pozo-Vazquez2004, Jerez2013, while others study the spatial variability instead of the temporal variability Gueymard.Wilcox2011a.

Some authors make use of regionalization techniques for atmospheric variables in order to analyze them from a climatological point of view Argueso2011 or for solar energy purposes, mainly for operation and short-term assessment Zagouras2013, Zagouras2014, Zagouras.Pedro.2014.

In this chapter a multi-step scheme that systematizes the time-space comparison of solar irradiation or photovoltaic productivity among sub-regions of the target area

is described, taking into account a large set of factors involved in PV production that could affect the variability of photovoltaic energy yield. Spatial complementarity between clusters is analyzed through correlation coefficients of solar irradiation and PV energy yield time series, showing possibilities for compensating PV production shortages in certain clusters.

The analysis over the Iberian Peninsula show that most of the results obtain in previous studies can be also seen with this approach. In addition, the methodology allow to give a clearer spatial answer to the questions as well as to sistematize the procedure.

This chapter is organized as follows: in first place, a description of the methodolgy is presented. Each stage of the multi-step scheme is summarized in section **??** and explained in more detail in the Methodology chapter. After that, the clustering algorithm is applied to the irradiation over the Iberian Peninsula and main results are shown.

5.1.1 Methods

The whole comprehensive scheme followed is represented in figure 5.1, showing the 4 stages and procedures that will be explained in detail in the following section.

From data of solar irradiation on the horizontal plane as the starting point of the method, the most common variable obtained from different data sources, the scheme will provide different outputs that can be used to evaluate resource and PV production:

- Regionalization of the area to facilitate the spatial analysis.

- Resource and energy yield aggregated by areas.

- Interannual variability of the resource and the PV production.

- Evaluation of the complementarity of the resource and PV production among areas.

Regionalization by clustering

Regionalization procedures provide the ability of extracting general information of the areas that could be treated as a coherent unit, facilitating the analysis and not
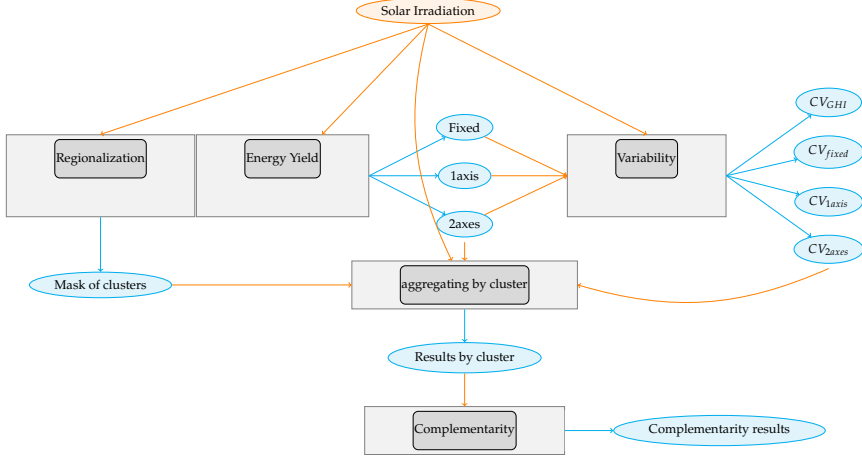
FIGURE 5.1: *Scheme: Each gray block represents each of the operations needed to get the variability and complementarity results. Orange ellipses are the data employed and blue ellipses are the results of each stage. If the results of one of the blocks are used as input for another stage, connectors are represented in orange color.*

considering those characteristics that are not under study. As it was explained in the methodological chapter, classical climatological classifications have some grade of subjectivity due to the fact that they rely on arbitrary assumptions Kottek2006 and their criteria are based on temperature and precipitation trewartha1980koppen. For our purpose, objective and data-derived criteria are more suitable due to the fact that a different variable is analyzed and its classification does not match classical climate divisions. Objective methods based on clustering techniques have been applied successfully over the literature for the analysis of renewable energy resources Polo2015, Zagouras2013, Zagouras2014, Zagouras.Pedro.2014, gomez2015characterization and atmospheric variables Argueso2011, garcia2012seasonal.

A commonly applied regionalization methodology includes the Kmeans algorithm after preprocessing the data through Principal Component Analysis Ding2004. This two-step method first reduces redundant information by a Principal Component Analysis that decreases dimensionality of the original dataset. After that, k-means algorithm is applied to the reduced data to find the optimal partition of clusters, which is based on similarity between each element or object inside the cluster and its centroid. This is considered as the most representative element of the cluster, and similarity is

measured by an objective function defined in the cluster algorithm.

This method presents some problems regarding the random selection of the cluster centroids in the first step. Different initial centroids can lead to different solution or a local optimum could be found. Also, there could be some computational problems if many iterations are needed to get the final partition.

The procedure used in [**Argueso2011**] and in [**Zagouras2014**] is adapted to get the optimal partition in our scheme from a combined clustering grouping and avoiding the above mentioned problems: the **k-means partitional algorithm is initialized with a hierarchical clustering solution of a the dimension-reduced data by a Principal Component Analysis.** For the particular case applied in this work, vectors of daily solar irradiation are used for the regionalization. The following steps are needed to get the optimal partition of clusters in the area:

- **To reduce data dimensionality**. Principal components are eigenvectors of an orthogonal matrix after applied a singular value decomposition (SVD) to the original data, daily solar irradiation vectors, whose initial dimension is reduced to the first eigenvectors that retain 95% of the variance. Considered that, a linear combination of these eigenvectors represents the initial data.

- **Hierarchical clustering to initialize k-means**. A hierarchical clustering method classifies data based on a hierarchy. If it is agglomerative, it will start with a cluster for each observation of the data and observations will group together recursively by similarity using the "complete linkage" method. Once the hierarchy is obtained, centroids can be calculated for each emerged partition with a number of clusters between 2 and n, where n is a high enough selected number of clusters. Centroids will be the initial seed for the Kmeans algorithm, avoiding the computational problems and favoring reproducibility.

- **Kmeans algorithm**. The k-means algorithm is a partitional clustering method that minimizes an objective function that defines similarity among the elements of each cluster. In our case we made use of the Euclidean distance, Eq. 5.1 between the objects or elements in the cluster and its centroid as the objective function. The number of clusters in which the data is divided into has to be known beforehand. In order to overcome the inconvenience, the algorithm is run from 2 to n clusters and the optimum number is determined by making use of a clustering validity index after that.

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} ||x_i - c_j||^2 \qquad (5.1)$$

- **Validity index**. In order to determine the optimal partition, validity clustering techniques are applied. There are two types of validation for the clustering methods. First, external clustering validation that make use of external information out of the data; and secondly, there are internal clustering validation methods that rely only on information from the data 5694060. The latter are used to preserve objectivity as much as possible and are based on two criteria: compactness and separation of the clusters emerged. We use one of the most applied validity index, the Calinski-Harabasz index CalinskiH, CH, that evaluates the average between and within cluster sums of squares.

- **L-method**. CH index is calculated for every partition from 2 to n clusters. The resulting CH graph in figure 5.2 for the Iberian Peninsula regionalization is shown for a number of clusters between 2 and 70 as an example. Theoretically, the partition with the maximum CH is the optimum, but the graph shows a decreasing trend which leads to imprecision in finding the optimum. The large number of data and the continuous variable analyzed are responsible for that. For that reason, the L-method is applied Salvador2004. This method selects the intersection of two best-fit lines in the graph CH vs. k, where k is the number of clusters of the partition Zagouras2013. All possible pairs of lines that fit linearly to the left and right sequence of data points are created. Each line has at least two points. The total root mean squared error is calculated as in Eq.:5.2:

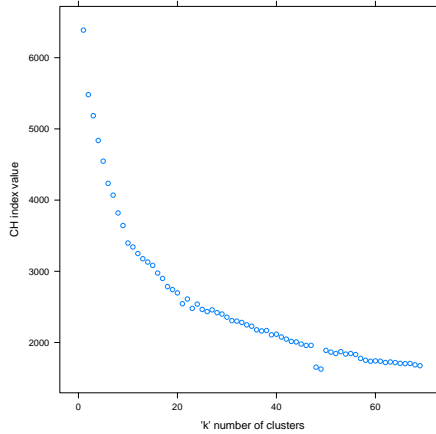$$RMSE_T = \frac{c-1}{k-1} RMSE_{left} + \frac{k-c}{k-1} RMSE_{right} \qquad (5.2)$$

FIGURE 5.2: *Calinski-Harabasz index by 'k' number of clusters*

Where c is the number of clusters where the graph is split into the two fit-lines, k is the total number of clusters. The "total root mean square error" is a weighted error with two terms, one for each side of c in the graph. Each side has a heavier weight depending on the points involved in the fitting. The minimum of $RMSE_T$ gives us the optimum number of clusters of the data Zagouras2013 which are used in the following steps.

Photovoltaic energy yield

The simulation of a photovoltaic energy system is described in a previous chapter. Here it is summarized in order to do not miss the coherence of the text.

The assessment of the power output of a photovoltaic system is carried out in two main steps:

1. In first place, global irradiation on the horizontal plane $G(0)$, which is the most common variable obtained from data sources, has to be transformed into the plane-of-array irradiation, $G(\alpha, \beta)$, where α is the azimuth angle and β the inclination angle of the generator plane. Due to optical losses (reflection, angle of incidence, and dust), the irradiation available is reduced for the photovoltaic cells inside the panels and the plane-of-array irradiation is then denoted as effective irradiation on the PV generator $G_{eff}(\alpha, \beta)$.
   Three different types of tracking types are considered for the photovoltaic generator that influences on the tilt of the panels:

   - **Fixed** panels with an optimum angle of inclination that depends on the latitude of the place.
   - **North-South** oriented panels that track the sun daily varying the azimuth angle, we will refer to them as "one axis"
   - **Two-axes** tracking system that allows variation of the azimuth and inclination angles, we will refer to them as "two axes".

2. Once the effective irradiation that reach solar cells has been assessed, second step is the transformation into power output that depends on the photovoltaic system. The photovoltaic system is composed of a PV generator, consisting of several PV modules, and an inverter to transform the DC current output from the generator into AC current to be integrated into the network. In order to estimate potential for photovoltaic production, the term *yield* is defined as the system energy produced divided by the power installed [$\frac{k}{k}$].

Variability and complementarity

The metric to analyze interannual variability is the coefficient of variation, CV Eq.5.3, which is defined as:

$$CV = \frac{\sigma}{\overline{X}} \tag{5.3}$$

In this equation, σ is the standard deviation of the variable analyzed and it is divided by the mean of the variable in the period of the study. Sometimes CV is represented in percentage. This measure is dimensionless and can be applied in different time scales, which is helpful for comparisons.

To assess complementarity of the solar resource in the area of study, the Pearson's correlation coefficient between the time series of pairs of clusters Eq.5.4 is calculated:

$$\rho_{i,j} = \frac{\sigma_{c_i,c_j}}{\sigma_{c_i}\sigma_{c_j}} \tag{5.4}$$

In this equation, c_i and c_j are the time series corresponding to the clusters i and j. The concept of complementarity is associated with negative correlations between sub-regions of a wider area. If there is such a complementarity, a positive change in the variable for one of the clusters will be associated to a negative change for the other one. This is relevant for identifying spatial compensation possibilities and reducing overall variability in a network with high penetration of solar PV energy.

Complementarity can occur in different scales, either spatial or temporal and to understand it sometimes is a matter of balance. Spatial resolution for the complementarity assessment must be high enough to make sense of the comparison between zones, due to the fact that it is clear that geographically dispersed areas, far from each other, will have very different evolution of atmospheric variables, but may not be interesting from the electricity generation point of view. On the other hand, if the area of study is too small, atmospheric variables and therefore renewable resources will evolve in a very similar way. For such small areas, local complementarity between different resources can be analyzed, but not spatial complementarity of one resource. The methodology presented here addresses this issue by applying the inter-cluster comparison, that ensures homogenity within each cluster and differences between them.

Correlation coefficient for a long time series may hide changes in complementarity for shorter sub-periods with higher frquency correlations. For that reason a moving window is applied to the time series calculating the correlation coefficient and providing an indication of how complementarity between clusters varies during the

whole period. Width of the window depends on the length of the time-series analysed.

In order to obtain the more important cluster pairs regarding complementarity, the median of the correlation coefficient series is calculated for each pair. After that, the cluster pairs are reordered from lower to higher values of this median correlation.

5.1.2 Results

The optimal partition after having applied the clustering method is represented in figure 5.3. The CH validation procedure gives an optimum number of 19 clusters for the area, where each of the clusters has an homogeneous time evolution of solar irradiation. Due to the nature of clustering techniques, there is not an unique/best method to select the optimum partition. Another index (Davies-Boudin, [**davies1979cluster**]) has been applied for comparison, and the obtained optimum number of clusters was of the same order than for CH index.

Variability and complementarity results

After regionalization, it is performed an analysis of solar irradiation on the horizontal plane and PV yield by tracking system, including their temporal variability.

Regarding interannual variability, we have calculated the CV of two time-aggregated means of solar irradiation and PV yield:

- On one hand it is applied for the yearly mean of daily irradiation $G_{d,y}(0)$ and yearly PV yield. This metric gives the variation of the of energy from one year to another and if it is low, general stability of the solar resource and PV production is guaranteed.

- On the other hand the interannual variability of the monthly time series $G_{d,m}(0)$ and monthly energy yield is also investigated in order to quantify differences in the annual cycle.

The CV is also aggregated by cluster, in order to facilitate the intercomparison among areas.

Power from the PV generator depends quasi-linearly with solar irradiation at the plane-of-array ($G_{eff}(\alpha,\beta)$), besides second order effects (spectrum, wind, etc) Perpinan2007 . Due to that the fixed typology is the one with lower yield because the amount of irradiation reaching cells is lower than the amount of energy reaching panels when trackers have one or two axes movements.
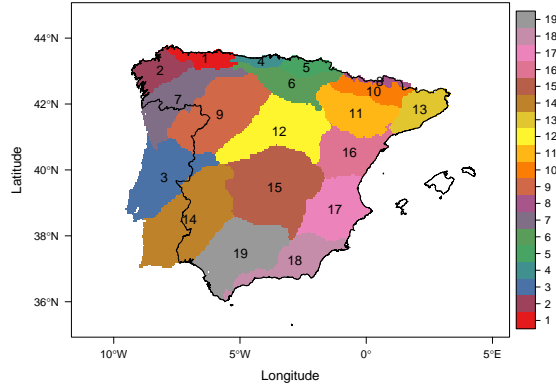
FIGURE 5.3: *Optimal partition of 19 Clusters after applied the algorithm and the validity index*

For areas where solar irradiation is higher, yield differences between trackers are higher. This can be seen in figure 5.4 where yearly mean yield for the 30-years period is aggregated by cluster and tracking system, and clusters are sorted vertically from less to more energy yield. A noteworthy result is that yield increase from fixed panels to one-axis panels is non-linear. This increase ranges between 17% for the clusters with less solar irradiation, located at the northern coast (clusters 4, 5), and 30% for the southern clusters with more solar irradiation (clusters 18, 19). In contrast, enegy yield increase from one-axis to two-axes panels is almost constant, around 12% for all clusters. A consequence of the non-linear PV yield increase from fixed to one-axis panels is that the energy yield differences between clusters are much higher for tracking than for fixed systems. While for fixed panels PV energy yield varies between 1000 and 1450 $[\frac{k}{k}]$, for two-axes systems it varies between 1350 and 2100 $[\frac{k}{k}]$. These average values are coherent with results obtained in [**Antonanzas-Torres2013**] when considering a value of 0.75 for the system performance ratio.

Electricity price variations significantly depend on the variations of the monthly renewable electricity production from year to year. This time-scale is also the most influenced by the large scale circulation modes for solar potential in the Iberian Peninsula Jerez2013a. The winter half of the year, from October to March, is especially variable.
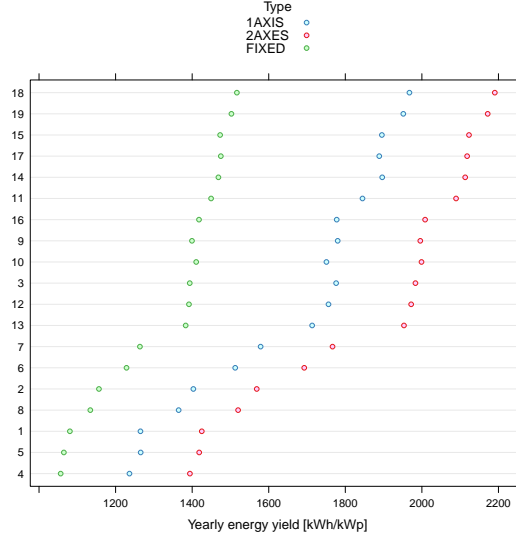
44

FIGURE 5.4: *Yearly mean of PV yield by cluster and for each tracking system* $[\frac{k}{k}]$. *Values are sort from lower to higher yield values.*

The interannual variability for monthly yield is higher than for the irradiation at the horizontal plane, as it occurred for yearly values which results are in the appendix. In winter months, these differences in CV are much higher than in summer. This behavior is more pronounced in northern areas.

In order to quantify these differences in variability between solar irradiation and solar power output, the ratio between variability of yield by tracking system and solar irradiation is represented in figure 5.5 for each month and cluster. If CV of energy yield is higher than CV of solar irradiation, values are above one. On the other side values will be below one if CV of solar irradiation is higher than CV of energy yield.

The highest ratios are obtained between CV_{2axes} and CV_{G0}. The ratio of CV_{1axis} is clearly lower in winter months, but is very similar in summer to the ratio of CV_{2axes}. Yield with an 'horizontal' axis tracker and 'two-axes' trackers increase the variability between 20% in summer and more than 80% in some areas in winter. The fixed typology ratio, CV_{fixed}/CV_{G0} has a much wider range in the whole year. In winter months, it has values between 1.2 and 1.6, depending on the cluster, and is not far from the other two typologies. In contrast, this ratio decreases rapidly in summer months, reaching values below one between May and August. This means that for that period, variability of the 'fixed yield' is smaller than variability of solar irradiation at the horizontal plane.
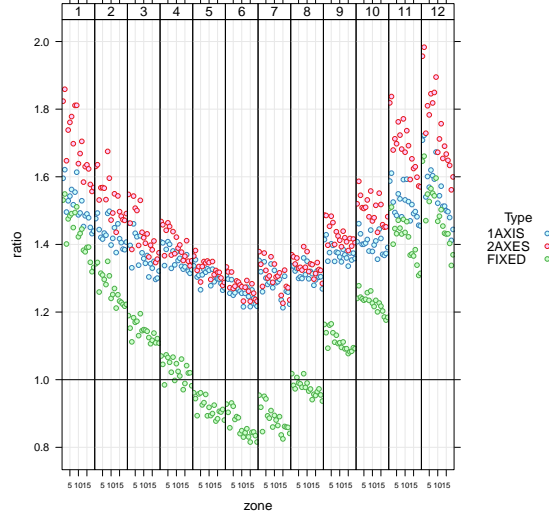
FIGURE 5.5: *CV ratios between each type of tracking system and solar irradiation at the horizontal plane, grouped together by month in the graph. Ratios are calculated for each cluster, represented in the x axis. "Fixed" represents $\frac{CV_{fixed}}{CV_{G0}}$, "1 axis" is $\frac{CV_{1axis}}{CV_{G0}}$ and "2 axis" is $\frac{CV_{2axes}}{CV_{G0}}$*

The results of CV show that variability of PV energy yield at tilt panels is higher than variability of solar resource at horizontal plane in most cases, explained by the nature of solar irradiation at tilt panels and its dependency of solar irradiation at horizontal plane, [**Perpinan2009**].

The monthly time series are also selected for the complementarity analysis.

Regarding solar power complementarity, opposite-evolving time-series for different areas would strongly increase the reliability of the whole electric system, as short-falls of solar irradiation in certain areas could be compensated by above-normal irradiation in others. However, this ideal situation is difficult to find in a rather limited area like the IP, at least for monthly timescales over a long time period of 30 years. In this case, the absence of correlation becomes also important, as it avoids simultaneous shortfalls or simultaneous above-normal values, and therefore softens the overall power production. The correlation matrix for the 30-year period and each month is in the appendix, showing the results for the whole period. In most cases, the correlation coefficient is highly positive, specially between pairs in the northern part of the Iberian Peninsula. For the southern part, the correlation coefficient is also positive but it decreases in July and August. There are some exceptions between the north-

ern clusters 4 and 5 and the southern clusters 14 to 19, these pairs are only slightly correlated, not correlated or even slightly anticorrelated for some months.

Overall, southern and eastern clusters are uncorrelated at least during part of the year with northern and northwestern clusters. In some cases, the absence of correlation is found between nearby clusters: in winter months, the north-eastern cluster 11 (central Ebro valley) is uncorrelated to the closely-lying clusters 4, 5 and 8 (in the northern coast and Pyrenees). This is probably related to persistent atmospheric situations with north to northwesterly winds, that cause cloudiness in the windward clusters and clear skies in the leeward Ebro cluster, due to a foehn effect. This result points out the selective character of the clustering method.

It could be that the obtained clusters present higher complementarity in shorter sub-periods. We have divided the whole 30-year period in sub-periods of consecutive 15 years. The correlation coefficients have been calculated again for the resulting 15-year moving window, for each pair of clusters and for each month. In this way, we obtain how each correlation coefficient evolves during the 30 year period. The analysis has been applied for the four variables in the study: solar irradiation at the horizontal plane and PV energy yield for each tracking system.

The median value of the correlation coefficient series is calculated for each pair. Each serie comprises 12 monthly values for each of the 15-year moving windows. After that, the cluster pairs are reordered from lower to higher values of this median correlation, and the first 15 pairs (with the lowest median correlation) are selected as the most representative of complementarity. These cluster pairs are represented in figure5.6, showing the time-evolution of its correlation coefficient. These results are for solar irradiation on the horizontal plane. The corresponding figures for PV yield by tracking system are not shown due to the similar results obtained for this analysis.

The most relevant pairs in terms of complementarity include a northern (1, 2, 4 or 5) and a south-eastern cluster (17 or 18), as can be seen in figure 5.6. The negative correlations for these pairs reach values below -0.6, at least in some 15-year sub-periods. These clusters are negatively correlated in most cases. Clusters 19 and 14 (southern and south-western IP) are also negative correlated with clusters in the north, although with lower values than the south-eastern clusters.

It is important to notice the appearance of cluster pairs 3-4 and 3-5 in this figure. These clusters are closer than the previously commented cases, which highlights the adequacy of the clustering method. All three are Atlantic coast clusters, but while cluster 3 includes part of the western coast, clusters 4 and 5 are northern coast areas. This fact, together with the position of the main mountain ranges, can explain their partially complementary behaviour.

In order to highlight the months with maximum anti-correlation, figure 5.7 presents, for the same 15 cluster pairs as above, the minimum values of the monthly correlation coefficient (where the minimum for each month is calculated over all 15-year sub-periods). Differences between months are clearly observed in this graph. Only two months (March and June) show consistently positive values of this parameter, and
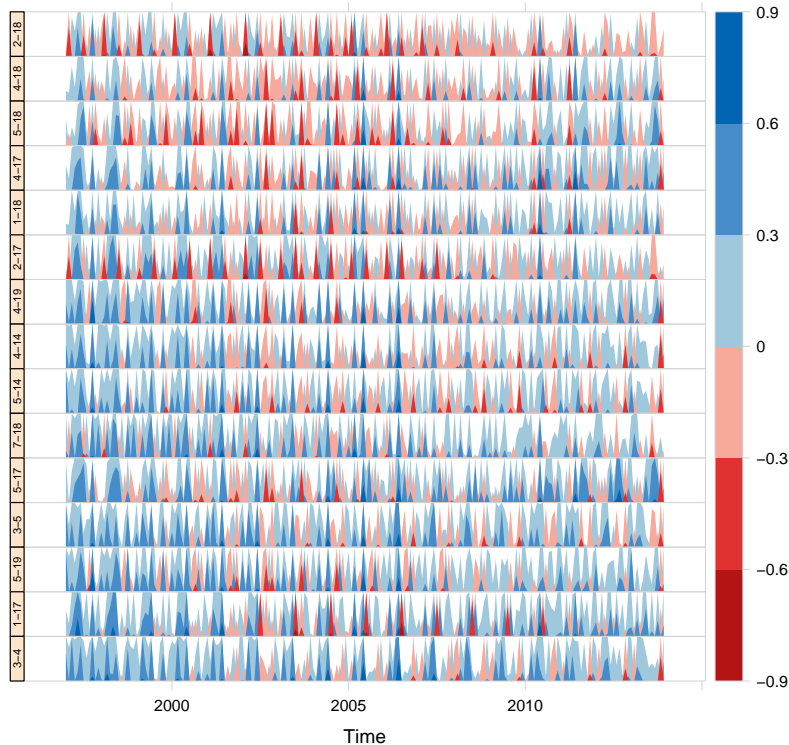
FIGURE 5.6: *Correlation coefficient of solar irradiation at the horizontal plane: evolution of a 15-year moving window of monthly values, for the 15 cluster pairs showing the smallest median correlation values. Monthly negative correlations (in red) and monthly positive correlations (in blue) are represented in the same axis to facilitate the comparison of the multiple time-series. Also, higher correlation values overlap with lower ones, enabling a compact presentation of all the information in a narrower plot. The cluster pairs are indicated on the left, while the year in the x-axis indicates the end of each 15-year moving window.*

therefore a low complementarity. In the other months, this parameter has predominantly negative values, revealing a certain degree of complementarity. July, August and September show rather low values and relatively high complementarity, which is important as these are months with a high productivity and also include the summer demand peak. In general, for the second half of the year, the values of this minimum correlation value are lower than for the first half of the year.
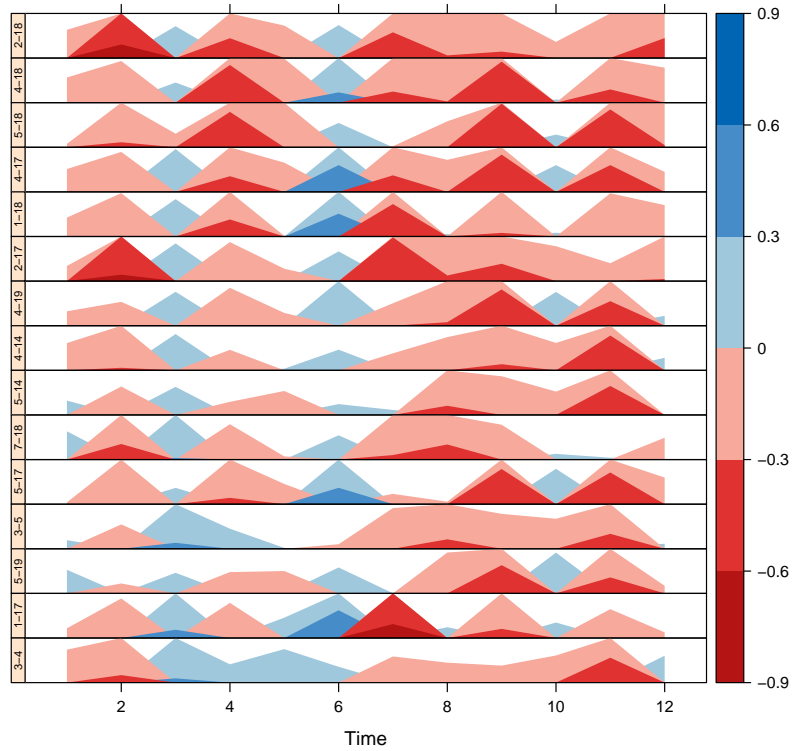
5.1.3 Conclusion

5.1.4 Summary

FIGURE 5.7: *Correlation coefficient of solar irradiation at the horizontal plane: minimum values of the monthly correlation coefficient, for the same cluster pairs as figure 5.6. The minimum is calculated over all 15-year subperiods. The type of representation of correlation values is the same as in figure 5.6. The cluster pairs are indicated on the left, while the numbers in the x-axis indicate the months.*