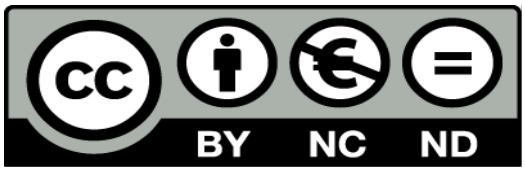
UNIVERSIDAD DE LA RIOJA

TESIS DOCTORAL

Título
New methodologies and improved models in the estimation of solar irradiation
Autor/es
Fernando Antoñanzas Torres
Director/es
Francisco Javier Martínez de Pisón Ascacíbar y Oscar Perpiñan Lamigueiro
Facultad
Escuela Técnica Superior de Ingeniería Industrial
Titulación
Departamento
Ingeniería Mecánica
Curso Académico

Tesis presentada como compendio de publicaciones. La edición en abierto de la misma NO incluye las partes afectadas por cesión de derechos



New methodologies and improved models in the estimation of solar irradiation

Fernando Antoñanzas Torres

A thesis submitted in

fulfilment of the requirement for the award of the

Degree of Doctor of Engineering

DEPARTMENT OF MECHANICAL ENGINEERING

UNIVERSITY OF LA RIOJA

APRIL 2016

I hereby declare that this thesis entitled "New methodologies and improved models in the estimation of solar irradiation" is the result of my own research except as cited in the references. This thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :

Student : Fernando Antoñanzas Torres

Date : April 2016

Supervisors : Dr. Francisco Javier Martínez de Pisón Ascacíbar

Dr. Óscar Perpiñán Lamigueiro

To the Sun

Acknowledgements

Along this four-year trail I cannot avoid looking behind and thank all the people that have walked next to me.

Oscar showed me the trailhead for knowing better our king star and he did all his best to transmit his open-source way of understanding research and life.

Francisco Javier trusted me and gave me the support in those days I did not even have a job. He was essential in making this hike a pleasant experience. I appreciate very much the great working atmosphere he creates within his group.

Andres has been moving from one place to the other, but always present. He opened the door of the research and helped me leave a life that did not belong me.

I am very thankful to the whole EDMANS group: Julio, Alpha, Javier, Ruben, Enrique and Manuel for accompanying me along different journeys of this thesis and for making me never feel alone.

Special thanks to Jesus Polo for hosting me at CIEMAT, and for making this research experience so pleasant.

I also want to acknowledge professor Carlos Coimbra at University of California San Diego for letting me collaborate within his group during my research stay in the EEUU.

My sweetest gratitude is for my family. Dad, mom, Javier, Irene and grandpas, thank you for made me a curious person, I love you. I also want to express my love to Bea for her patience and support. My family and Bea were in charge of providing the assistance without which this thesis would never have been real.

Fernando Antoñanzas Torres, Logroño, Spain

Abstract

The wide development of solar energy, technical agriculture and climate monitoring happening in these years requires a better knowledge of incoming solar irradiation. Although solar irradiation can be measured with pyranometers with high accuracy if correctly maintained, there is a lack of these sensors in most of the countries and regions. Besides, the high spatial and temporal variability of solar irradiation make measurements from relatively nearby stations not reliable for certain applications. As a result, solar irradiation must be frequently modeled and estimated. Many different approaches have been proposed in the last decades for generating solar irradiation out of other commonly measured meteorological variables such as temperatures, rainfall and sunshine duration. More recently, other techniques using sensors onboard satellites are able to provide solar irradiation with a higher spatial coverage. Furthermore, novel machine learning techniques can generate accurate estimations of solar irradiation. However, despite the massive development of all these techniques, still there are some drawbacks and issues in the estimation of solar irradiation directly affecting accuracy.

In this context, this thesis focuses on two main blocks of studies: the temporal and the spatial methods for the estimation of solar irradiation. Beginning by traditional models, models were benchmarked based on the errors and robustness and their capacity of spatial generalization was also evaluated. From this point, more complex techniques such as support vector regression with a special optimization procedure were proposed and results were compared to parametric models. To end the block of temporal models, a wide range of satellite-based models were studied to evaluate the sources of uncertainty and error in the estimation of global and also beam irradiation under different scenarios. Regarding the spatial methods, satellite-based estimations were compared to on-ground measurements and then combined to generate more accurate maps of solar irradiation, not only for global horizontal irradiation but also for the effective irradiation on three different tilted angles. Furthermore, a very precise downscaling method for satellite-based estimations was proposed taking into account the topography

and geostatistics using some on-ground records. The results of these proposed methods show a useful insight on the improvement of the estimation of solar irradiation.

Some of the methods proposed in this thesis were provided as free `R` programming code, available as supplementary material in the articles. This code was generated with the aim of being useful for future replications and applications of the proposed methods in different regions and was one the most relevant final products of this thesis.

To conclude, the contributions presented in this thesis prove the great field of improvement in the temporal and spatial estimation of the main energy input in our planet, the solar irradiation.

Resumen

El gran desarrollo que se está viviendo en la actualidad y en los últimos años en el campo de la energía solar, de la agricultura tecnificada y del análisis climático requiere de un mejor conocimiento de la irradiación solar que recibe nuestro planeta. Aunque la radiación solar se puede medir con alta precisión con piranómetros, si estos están correctamente mantenidos, existe gran escasez de estos sensores en la mayoría de los países y regiones. Además, debido a la alta variabilidad espacial y temporal de la radiación solar, aquellos datos de estaciones cercanas al punto de interés pueden tener un alto grado de incertidumbre y no ser fiables para determinadas aplicaciones. Por lo tanto, la irradiación solar debe ser modelada y estimada en numerosas ocasiones. En las últimas décadas se han propuesto una gran variedad de técnicas para estimar la radiación solar a partir de otras variables meteorológicas comúnmente medidas como las temperaturas, la lluvia y la duración solar. Además, haciendo uso de los datos generados por sensores instalados en los satélites se han desarrollado modelos que son capaces de estimar la radiación solar con una mucho mayor covertura espacial. Además, con el desarrollo de las técnicas de inteligencia artificial, también denominadas de *artificial intelligence*, se pueden generar estimaciones con un alto grado de precisión. Sin embargo y a pesar del gran número de estudios e innovaciones en este campo, todavía existen algunos problemas e inconvenientes en todos estos métodos, que afectan directamente al error en las estimaciones de radiación solar.

En este contexto, esta tesis se ha centrado en dos bloques de estudios principales: los modelos temporales y los modelos espaciales para la estimación de la irradiación solar. Inicialmente, se comenzó con modelos paramétricos tradicionales, que fueron comparados por su error y robustez, así como por su capacidad de generalización espacial. A partir de este punto, otras técnicas más complejas de inteligencia artificial como las máquinas de vector soporte con un procedimiento especial de optimizado se desarrollaron y los resultados se compararon con los de los modelos tradicionales. Para terminar con el bloque de modelos temporales, un gran número de modelos de satélite y de cielo claro

fueron evaluados para analizar el origen de la incertidumbre y de los errores en estos modelos y su modo de propagación. Este análisis se realizó para la irradiación global y también para la directa en distintos escenarios dependiendo de la covertura nubosa. En relación a los modelos espaciales, se estudiaron comparativamente los errores de las estimaciones de satélite con las medidas en tierra y se unieron ambas fuentes de datos para realizar mapas más precisos de irradiación solar, no solo para la irradiación global horizontal, sino también para la irradiación efectiva en tres planos comúnmente usados para la energía solar fotovoltaica. En este bloque también se desarrolló una metodología para la mejora de la resolución espacial usando estimaciones de satélite y teniendo en cuenta la topografía del terreno y datos de varias estaciones con técnicas geoestadísticas. Los resultados de los métodos que se han propuesto muestran claramente una mejora respecto a los modelos anteriores allí donde se han evaluado.

Algunos de los métodos propuestos en esta tesis han sido facilitados como código libre programado en el lenguaje `R` y se han facilitado como material suplementario de los artículos. Este código se generó con un claro objetivo de compartir los métodos con la comunidad científica y que fuesen útiles para futuros estudios y aplicaciones de los métodos en regiones diferentes. Por lo tanto, este código libre es uno de los productos finales más relevantes de esta tesis.

Para concluir, las contribuciones presentadas en esta tesis demuestran el gran campo de mejora que existía y todavía existe en la estimación temporal y espacial del principal input energético de nuestro planeta, la irradiación solar.

Contents

List of Figures

List of Tables

Nomenclature

A	Correction based on variable importance
$AERONET$	Aerosol Robotic Network
AOD	Aerosol optical depth
BHI	Beam horizontal irradiance
$Bo0_d$	Daily extraterrestrial irradiation on the horizontal plane
$BSRN$	Baseline Surface Radiation Network
$CMSAF$	Climate Monitoring Satellite Application Facility
CSP	Concentrated solar power
DHI	Diffuse horizontal irradiance
DHI_{ani}	Anisotropic diffuse horizontal irradiation
DHI_{iso}	Isotropic diffuse horizontal irradiation
DNI	Direct normal irradiance
ΔT_i	Range of temperatures of day i
$ESRA$	European Solar Radiation Atlas
GHI	Global horizontal irradiance
GHI_d	Daily global horizontal irradiation
GHI_{down}	Downscaled global horizontal irradiation
H	Daily average relative humidity
M	Binomial variable of daily rainfall
$MACC$	Monitoring Atmospheric Composition and Climate
MAE	Mean absolute error
MAE_{val}	Mean absolute error of validation
MAE_{test}	Mean absolute error of testing
MBE	Mean bias error
R^2	Coefficient of determination

$rMAE$	Relative mean absolute error
$rRMSE$	Relative root mean square error
P	Daily rainfall
PV	Photovoltaic energy
SVR	Support vector regression
v	Variable
W	Daily average wind speed

Chapter 1

Introduction

1.1 Background

Energy and matter are the main components of the universe and both of them are limited.

The Industrial Revolution set the beginning of a new civilization based on high-quality energy sources: coal, oil and natural gas, which propitiated a never seen growth in the global population. However, the fact that natural resources and particularly fossil energy sources are limited poses a tremendous challenge of sustainability for our civilization. In history, sustainability has been traditionally approached via an increase of complexity for problem resolution (Tainter et al., 2006). In these days and age, if we want a civilization energetically sustainable in the long term and at a manageable cost, we need to make the transition from fossil energy sources to a more varied mix of renewable energy sources.

In the XXth century and beginning of the XXIst, a wide range of renewable energy technologies were explored and developed. Some of these technologies perished and others, mainly wind, solar photovoltaic and solar thermal prevailed and become commercial. Among the latest sources, the photovoltaic technology stands out due to its simplicity and scalability from small and isolated rooftop installations, to more complex multi-megawatt grid-connected power plants. The global installed photovoltaic capacity has evolved from 3.7GW in 2004, to more than 177GW at the end of 2014 (REN21, 2015). Furthermore, this technology has seen a remarkable cost-reduction derived from the economy of scale in photovoltaic panels manufacturing, also accompanied by a notable increase on the efficiency of solar cells.

The photovoltaic development has occurred in the majority of countries with different penetration depending on the energy solutions provided. Rural

electrification with photovoltaic technology in areas isolated from the grid poses as one of the most deployed solutions in developing countries. Besides, solar pumping of water for human consumption and irrigation, and district illumination are as well common initiatives in many African countries. Nevertheless, most of the global installed photovoltaic capacity is accounted in on-grid installations, as a way to reduce the external dependency of countries on fossil fuels, which are mainly produced in very instable areas of our planet. In this case, countries have made a strong effort with regulation and subsidies (feed in tariffs, green certificates and fiscal benefits) to estimulate the installation of this technology. It is noteworthy that the grid parity of photovoltaic technology, stage when it is competitive with traditional fossil energy sources, has already occurred in many areas and still it is expected a 40% cost reduction over the next 4-5 years (DB, 2015).

The concentrated solar power technology has also experienced an outstanding growth in the last ten years, basically structured in parabolic trough, central tower and linear Fresnel technologies, and accounting at the end of 2014 a global installed capacity of 4.4GW (REN21, 2015). The manageability of this energy, which can be stored as latent heat in molten salts, and therefore adapt generation to demand, is one of its key benefits. However, the high complexity of this technology that only gets advantage of the beam component of solar irradiation, and also the fact that the condenser cooling is best achieved expending a great amount of water (it can also be performed with dry cooling at a much higher cost) limits the potential areas of installation. Regarding solar thermal technologies, the solar thermal energy for water heating, mainly for homes, has accomplished the greatest growth from 100GW-thermal in 2004 to 406GW-thermal at the end of 2014 (REN21, 2015).

This outstanding development of solar energy technologies is still tiny when looking at big numbers. For instance, in 2014 only the 0.9% of the total electricity generation was produced by the photovoltaic technology and at the same time, electricity only represented around the 18.1% of the total final energy consumption (IEA, 2014). As a result, in a scenario of switching from fossil fuels to renewables, electrification of energy and ”solarization” of electricity are expected to have a tremendous growth in the following decades.

This coming development of solar energy technologies requires significant improvements in capital costs and in the technologies itself, but also, high quality estimations regarding the available solar resource in order to better plan energy policies, select optimal locations and reduce uncertainties of these investments. As a result, the generation of accurate and long-term datasets and maps of solar radiation is crucial for this upcoming ”solarization”.

1.2 Problem statement and motivation of this thesis

Solar radiation is the main energetic input in the Earth, affecting life and most of the climatic and meteorological processes occurring within our atmosphere. Since the very origin of the gender Sapiens, we have been curious about the solar geometry and the variations of the incoming solar radiation: sunrises, sunsets, cloud passing phenomena and solar eclipses. Several old civilizations achieved an outstanding knowledge of solar geometry, such as the Mayans, Greeks and Egyptians among others, but it has not been until the last century when solar radiation started to be quantified due to the requirement of numerous scientific areas (i.e. solar energy, climate monitoring and agriculture management). However, the cost and complexity derived from measuring solar radiation with sensors and the local validity of these measurements due to spatial and temporal variations have made solar radiation estimates an essential source of data.

Solar irradiation was firstly estimated by Angstrom (1924) from measurements of sunshine duration recorded with heliographs. Since then, many other techniques have been proposed for the same purpose studying the atmosphere as a filter that mitigates the extraterrestrial incoming irradiation. Traditionally, one of the most deployed methods has been the atmospheric-transmittance parameterization from commonly measured and related-to-solar irradiation variables. Hargreaves (1981) and Bristow and Campbell (1984) proposed relatively simple parametric models to estimate daily solar irradiation using only the daily range of temperatures and the extraterrestrial irradiation. This way, a cleaner atmosphere, which usually presents a greater daily range of temperatures, would provide higher irradiation. Other proposals corrected these models by adding information about the cloud cover using another simple to record variable: the rainfall (Jong and Stewart, 1993). Although a massive number of variations of these models have been proposed, the accuracy of these models is highly spatially dependent on the location of validation. Besides, novel algorithms of machine learning techniques have also been applied in the estimation of solar irradiation, i.e. artificial neural networks, support vector regression, random forests and extreme learning machines, among others. The common feature of these techniques is the requirement of sufficiently long records of measured irradiation and meteorological variables for model training, validation and testing to avoid over-fitting. Nevertheless, a higher capacity of generalization, robustness and improvement in errors has been reported when compared to classic parametric models, although still facing the issue of low spatial generalization (Urraca et al., 2015).

Furthermore, another branch of models omitted the influence of clouds and tried to estimate solar irradiance under clear sky conditions using data of atmospheric parameters such as aerosols, water vapor and ozone content (Rigollier et al., 2000; Ineichen, 2008; Gueymard, 2008). These models are optimally combined with cloud-cover data from satellite images and are frequently useful in the spatial estimation of solar irradiation. However, the low spatial resolution of atmospheric inputs, in the range of hundreds of kilometers, and satellite images, in the range of kilometers, propagate errors in the estimation and therefore, these models also need to be locally validated. In this sense, strong efforts are being made in the downscaling of solar irradiation to obtain high-resolution maps with the highest fidelity.

Despite the important advances undertaken in the last years to better describe the solar radiation, the estimation of the solar resource still faces two big issues:

- The development of more robust and spatially generalist models to generate long and reliable solar irradiation datasets.
- The development of high quality and high-resolution maps of solar irradiation.

1.3 Scope of research and objectives

In the framework of solar irradiation modeling is of primary importance to improve the accuracy of models for local estimation and for spatial mapping. The main objective of this thesis has been the development of different methods to obtain spatially robust models for global solar irradiation estimation and mapping.

The research was structured in order to address the following objectives:

1. Achieve a better knowledge on the limitations of classic parametric models for the estimation of global solar irradiation and provide a general method for a robust calibration of their parameters.
2. Development of models with machine learning techniques capable to overcome the limitations of parametric models.
3. Sensitivity analysis of clear-sky and satellite models for solar irradiance estimation to the accuracy of their inputs and understand the propagation of errors.

4. Development of a solar irradiation mapping methodology capable to overcome errors of current methods.
5. Development of a method in order to generate very high spatial resolution solar irradiation maps taking into account the topography of terrain.

1.4 Contributions presented in the thesis

The main contributions of this thesis are collated into five scientific publications in journals listed in the Journal Citation Reports®. The common intersection among these publications is the estimation of solar irradiation with sufficient accuracy to be used for solar energy planning, climate monitoring and agriculture.

Publication I (Antonanzas-Torres et al., 2013). Index factor JCR (2013): 3.361 (Q2 23/83, Energy & Fuels).

Antonanzas-Torres, F., Sanz-Garcia, A., Martinez-de-Pison, F.J., Perpinan-Lamigueiro, O., 2013. Evaluation and improvement of empirical models of global solar irradiation based on temperature and rainfall in northern Spain. *Renewable Energy* **60**, 604-614.

In this article it is done the most comprenhensive review to date on solar irradiation parametric models using temperatures and rainfall as explanatory variables, accounting a total number of twenty-two different models and other two new models that were proposed for the study area. A new methodology of model evaluation based on bootstrapping is proposed to calibrate and obtain more robust models, which are therefore less sensitive to the specificity of the training period. Furthermore, a new methodology for parametric model development is described, leading to generate optimized models to the study area depending on the importance of explanatory variables. Eventually, two different models were created with this methodology for the case of study in La Rioja region (Spain). Results of models created with the new methodology strikingly improved the errors obtained with the review of twenty-two models in the same region. Furthermore, it has been reported that models proposed in this publication also overcome the errors of other parametric models not only in this area of study (Antonanzas-Torres et al., 2015; Urraca et al., 2015).

The author of this thesis worked on all stages of this project and wrote entirely this publication. A. Sanz-Garcia and F.J. Martinez-de-Pison collaborated with him in the development of the method-

ology for model evaluation. O. Perpiñan-Lamigueiro contributed in the programming using the free-environment R software and also in the meteorological data-cleaning process.

Publication II (Antonanzas-Torres et al., 2015). Index factor JCR (2014): 4.380 (Q1 14/89, Energy & Fuels; Q1 3/127, Mechanics; Q1 3/21, Physics, Nuclear; Q1 3/55, Thermodynamics).

Antonanzas-Torres, F., Urraca, R., Antonanzas, J., Fernandez-Ceniceros, J., Martinez-de-Pison, F.J., 2015. Generation of daily global solar irradiation with support vector machines for regression. *Energy Conversion and Management* **96**, 277-286.

This publication presents a different approach to classic parametric models in the estimation of daily global solar irradiation. Support vector machines for regression were described with an optimized method for variable selection with genetic algorithms to develop spatially generalist models with a high level of accuracy. An evaluation between general and locally trained models was also presented, concluding that models trained with meteorological samples from different locations were able to be more parsimonious and robust than those trained with data from a single location. Furthermore, this study was also compared with the equivalent using two different parametric models from Bristow and Campbell (1984) and Antonanzas-Torres et al. (2013). While the applicability of parametric models was more recommended at the location of training, support vector machines were able to learn from data from different locations and therefore showed a higher capacity of spatial generalization. It is remarkable that support vector regression trained with the method proposed achieved a striking reduction in the mean absolute error of 41.4% and 19.9% as compared to the afore-mentioned parametric models. One of the most remarkable findings of using this technique is that in 33.6% of cases it was comparable to the use of a First Class pyranometer (uncertainty within 5%) and that regarding annual sums of solar irradiation only in 3 out of the 14 sites studied in Spain the error was higher than 5%. This study was the root for another study in which solar irradiation was mapped using the estimations of a general model built on support vector regression that predicted solar irradiation in stations in which it was not recorded (Antonanzas et al., 2015).

The author of this thesis contributed in all stages of this study. F.J. Martinez-de-Pison, J. Fernandez-Ceniceros and R. Urraca contributed

in the development of the optimized methodology for support vector regression training. Additionally, J. Antonanzas and R. Urraca collaborated in the analysis of results.

Publication III (Polo et al., 2014). Index factor JCR (2014): 3.476 (Q1 20/89, Energy & Fuels)

Polo, J., **Antonanzas-Torres, F.**, Vindel, J.M., Ramirez, L., 2014. Sensitivity of satellite-based methods for deriving solar radiation to different choice of aerosol input and models. *Renewable Energy* **86**, 785-792.

This article presents a sensitivity assessment of satellite-derived models for generating solar irradiance direct and diffuse components. The evaluation takes into account the current main global databases of aerosols and water vapor: AERONET (on-ground measurements), MODIS and MISR (based on satellites measurements) and MACC (reanalysis). Three well-known clear-sky models for solar irradiance estimation were considered (ESRA, SOLIS and REST2), as well as three global to direct conversion models (Louche, DirInt and DirIndex) to explore the sensitivity in the case of direct normal irradiance estimations. The work introduced a novel analysis of uncertainty propagation of each of these models in the framework of a satellite-based model. It was observed the impact of the error in aerosol optical depth into the final estimation of global horizontal irradiance and also direct normal irradiance. The study also considered the sensitivity analysis in the specific range of direct normal irradiance for concentrated solar power technology (400-900 W/m2). The study was performed using ancillary solar irradiance measurements at the "Plataforma Solar de Almería" (Almeria, Spain) belonging to the Baseline Surface Radiation Network and also records from the Almeria-PSA station belonging to the AERONET network.

The author of this thesis contributed in all stages of the sensitivity assessment in collaboration with J. Polo. The article was written in collaboration with J. Polo, J.M. Vindel and L. Ramirez-Espinosa.

Publication IV (Antonanzas-Torres et al., 2013). Index factor JCR (2013): 5.51 (Q1 6/83, Energy & Fuels).

Antonanzas-Torres, F., Cañizares, F., Perpiñán, O., 2013. Comparative assessment of global irradiation from a satellite estimate model (CM SAF) and on-ground measurements (SIAR): a Spanish case study. *Renewable and Sustainable Energy Reviews* **21**, 248-261.

This article presents a novel methodology for combining satellite-derived estimates and on-ground measurements for solar global irradiation mapping using the geo-statistical technique of kriging with external drift. Initially, global horizontal irradiation estimates from the Satellite Application Facility for Climate Monitoring (CM SAF) and on-ground measurements from 301 meteorological stations in Spain were compared at the location of these stations. On a second step, the analysis was compared with estimations of effective irradiation on three different tilted planes that are common in photovoltaics (fixed, two axis tracking and north-south horizontal axis), using estimated irradiation from these two databases. Eventually, different maps of annual global irradiation in the horizontal plane and in the three afore-mentioned titled planes were developed using kriging with external drift for Spain and compared against the CM SAF maps. The methods of this article were implemented in the free-environment R software and provided as supplementary material for future replications and applications of the study and methodology.

The author of this thesis contributed in the data collection and analysis of the results. The programming in R was performed by O. Perpiñan.

Publication V (Antonanzas-Torres et al., 2014). Index factor JCR (2014): 0.904 (Q3 61/89, Energy & Fuels).

Antonanzas-Torres, F., Martinez-de-Pison, F. J., Antonanzas, J., Perpinan, O., 2014. Downscaling of global solar irradiation in complex areas in R. *Journal of Renewable and Sustainable Energy* **6**, 063105. doi: 10.1063/1.4901539.

This article introduces a new downscaling method for solar irradiation in order to generate maps of high resolution (200x200m). Data from the CM SAF satellite database in 0.05x0.05º (˜5km grid) was downscaled taking into account the topography effects of shading and sky view factor, which were not considered in the satellite database. On a second step, kriging with external drift was useful to accurately adjust the downscaled map with a few on-ground measurements from pyranometers. The method proved useful in a case of study in La Rioja region, known for its complex topography. The code was programmed in free-environment R software and freely provided for future replications and applications of the study in other regions.

This study was partially financed by a project of the "Instituto de Estudios Riojanos" (IER- 2013). The author of this thesis contributed in all stages of this project. O. Perpiñan helped in the R code development. J. Antonanzas and F.J. Martinez-de-Pison collaborated in the article writing.

1.4.1 Thematic unit

The thematic unit involving the publications of this thesis is the solar irradiation modeling and mapping. All five publications answer to different objectives and approaches of solar irradiation estimation. However, the common intersection among these publications is the estimation of solar irradiation with sufficient accuracy to be used for solar energy planning, climate monitoring and agriculture. Due to the importance in this study of atmospheric monitoring, soft-computing, mathematical modeling, statistics and geo-statistics, they must be mentioned in this section.

1.5 Thesis outline

This thesis dissertation is divided in nine chapters. This first chapter briefly introduces the topic of solar radiation modeling, explaining the motivation of the study, as well as its scope and objectives. Chapters 2 through 6 contain the five scientific publications contributing to this thesis. The first three publications focuse on the temporal estimation of solar irradiation, whereas the other two emphasize the spatial estimation. Chapter 7 is a summary of the most significant results found in these studies. Chapter 8 collates a general discussion of previous findings. Chapter 9 presents the main conclusions of this dissertation and finally, chapter 10 collates the future works.

Chapter 2

PUBLICATION I

Antonanzas-Torres, F., Sanz-Garcia, A., Martinez-de-Pison, F.J., Perpinan-Lamigueiro, O., 2013. Evaluation and improvement of empirical models of global solar irradiation based on temperature and rainfall in northern Spain. *Renewable Energy* **60**, 604-614.

Chapter 3

PUBLICATION II

Antonanzas-Torres, F., Urraca, R., Antonanzas, J., Fernandez-Ceniceros, J., Martinez-de-Pison, F.J., 2015. Generation of daily global solar irradiation with support vector machines for regression. *Energy Conversion and Management* **96**, 277-286.

Chapter 4

PUBLICATION III

Polo, J., **Antonanzas-Torres, F.**, Vindel, J.M., Ramirez, L., 2014. Sensitivity of satellite-based methods for deriving solar radiation to different choice of aerosol input and models. *Renewable Energy* **86**, 785-792.

Chapter 5

PUBLICATION IV

Antonanzas-Torres, F., Cañizares, F., Perpiñán, O., 2013. Comparative assessment of global irradiation from a satellite estimate model (CM SAF) and on-ground measurements (SIAR): a Spanish case study. *Renewable and Sustainable Energy Reviews* **21**, 248-261.

Chapter 6

PUBLICATION V

Antonanzas-Torres, F., Martinez-de-Pison, F. J., Antonanzas, J., Perpinan, O., 2014. Downscaling of global solar irradiation in complex areas in R. *Journal of Renewable and Sustainable Energy* **6**, 063105. doi: 10.1063/1.4901539.

Chapter 7

Results

This chapter summaries and discusses the most relevant results included in the publications associated with this thesis.

The general framework of this section is structured into five subsections, one for each publication. The first three subsections refer to the improvement of solar radiation models for the generation of solar radiation time-series: (i) classic parametric models based on temperatures and rainfall, (ii) complex models based on support vector regression and (iii) sensitivity of clear-sky and satellite models to their atmospheric inputs.

The last two subsections refer to the spatial estimation of solar irradiation: (iv) spatial mapping with geostatistics in different tilted angles and (v) for the downscaling of solar irradiation derived from satellite images considering the topography.

7.1 Results in Publication I

All the results described in this section are collated in the paper "Evaluation and improvement of empirical models of global solar irradiation: Case study northern Spain" (Antonanzas-Torres et al., 2013).

The objectives of this study were proving the validity of the method proposed to create locally adapted models based on the variable importance and then, evaluating parametric models according to their robustness. Eventually, these objectives were fulfilled in this study by developing on-site calibrated and more robust models than the other twenty-models reviewed.

The evaluation of variable importance seeks to improve the performance of the well known Bristow and Campbell (1984) general model with the local and specific relationships between predictors and the estimated daily global horizontal irradiation using the equation 7.1.

$$GHI_d = a \cdot (1 - exp\,(-b \cdot \Delta T^c)) \cdot Bo0_d \cdot A + p_{n+1} \tag{7.1}$$

$$A = 1 + \sum_{j=1}^{n} p_j \cdot v_j \tag{7.2}$$

where A (equation 7.2) is the correction of the Bristow and Campbell (1984) model based on the variable importance, p the parameter related to the variable v and n the number of total variables to be considered.

As a result, it is possible to correct this model with other meteorological variables recorded at the location based on their importance. The evaluation of variable importance was performed using the locally weighted polynomial regression, denoted as the loess smoother fit model (Cleveland, 1979). Each instance of the dataset was fitted with a low degree polynomial adjusted with weighted least squares. Hence, instances more related to the output received more weight, which was determined with equation 7.3.

$$w\,(x) = \left(1 - |x^3|\right) \tag{7.3}$$

Finally, the coefficient of determination (R^2) was calculated for this model against the intercept only null model as a measure of variable importance.

The assessment of variables importance was evaluated in seventeen different meteorological stations in La Rioja region (Spain) under different conditions and terrain during the period 2007-2010. Figure 7.1 depicts meteorological sta-

tions considered in the study from two different databases (SOS Rioja in blue circles and SIAR in red triangles with pyranometers within 5% and 10% of daily tolerance, respectively). Maximum and minimum daily temperatures, daily rainfall (P), daily average wind speed (W), daily average relative humidity records (H) and daily global horizontal irradiation (GHI_d) were obtained and filtered from spurious. On a second step, other variables were created from the previous: the binomial variable of rainfall (M), equal to 1 with some rain recorded and 0 with no rainfall, and the daily range of temperatures (ΔT). Since the influence of the past conditions the present and the present conditions the future, these variables were created for the previous and following days. With this, Table 7.1 collates the variable importance.

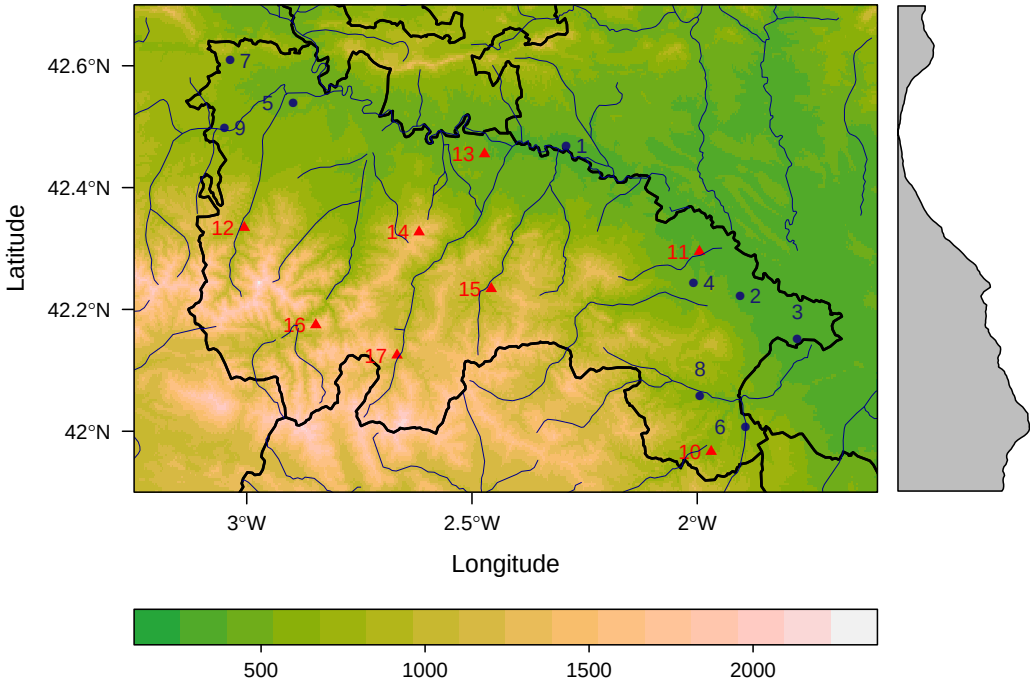


Figure 7.1: *Location of the meteorological stations selected in the region of La Rioja. The color band represents elevation (m). SIAR stations are shown by blue circles and SOS Rioja stations by red triangles*

Table 7.1: *Summary of variable importance results related to each variable v for the improvement of parametric models*

v	P_i	P_{i+1}	P_{i-1}	M_i	M_{i+1}	M_{i-1}	ΔT_i
R^2	0.0056	0.012	0.016	0.153	0.068	0.059	0.533
v	ΔT_{i+1}	ΔT_{i-1}	ΔT_{i+2}	ΔT_{i-2}	ΔT_{i+3}	ΔT_{i-3}	W_i
R^2	0.359	0.340	0.301	0.172	0.206	0.167	0.089
v	W_{i+1}	W_{i-1}	H_i	H_{i+1}	H_{i-1}	H_{i+2}	H_{i-2}
R^2	0.076	0.071	0.465	0.344	0.251	0.251	0.199

It is noticeable that the influence of M is greater than P, which means that GHI_d is more sensitive to the fact that any rainfall is recorded than the quantity of this rain. Consequently, a whole battery of models with different combinations of variables was built. Eventually, two different models were selected based on the robustness assessment explained latter on in this section: one using only temperatures and rainfall and the other including relative humidity and wind speed as per equations 7.4 and 7.5.

$$GHI_d = Bo0_d \cdot a\left(1 - \exp\left(-b \cdot \Delta T^c\right)\right) \cdot \left(1 + d \cdot M_{j-1} + e \cdot M_j + f \cdot M_{j+1} + g \cdot \Delta T_{j+1} + h \cdot \Delta T_{j-1}\right) + l \quad (7.4)$$

$$GHI_d = Bo0_d \cdot a\left(1 - \exp\left(-b \cdot \Delta T^c\right)\right) \cdot \left(1 + d \cdot M_{j-1} + e \cdot M_j + f \cdot M_{j+1} + g \cdot \Delta T_{j+1} + h \cdot \Delta T_{j-1} + l \cdot W_j + m \cdot H_j\right) + n \quad (7.5)$$

where, lower case letters stand for parameters to be calibrated.

On a second step, it is performed a robustness assessment in order to evaluate the stability of models when trained under many different conditions. As a result, the most suitable model is not only selected based on the lowest errors of testing (Lawrence et al., 1996), but based on the robustness. The training datasets (2007-2010) are divided into sampled 80% for calibration and 20% for validation. Bootstrapping was found useful to obtain a high level of knowledge from these relatively short datasets (Crawley, 2005; Mora and Mora-Lopez, 2010). This technique was repeated 100 times in each station for each model to eventually obtain the confidence intervals of the model parameters and errors.

Figure 7.2 depicts the confidence intervals (95%, n=100) of the mean absolute errors of validation (MAE_{val}) with grey vertical lines. The mean absolute error of testing (MAE_{test}) using the dataset of 2011 was represented with blue crosses. Low values of MAE_{val} accompanied of a narrow confidence interval indicate accuracy and stability of the model in that station. On the other side, a low MAE_{test} implies a high capacity of generalization with unseen data.

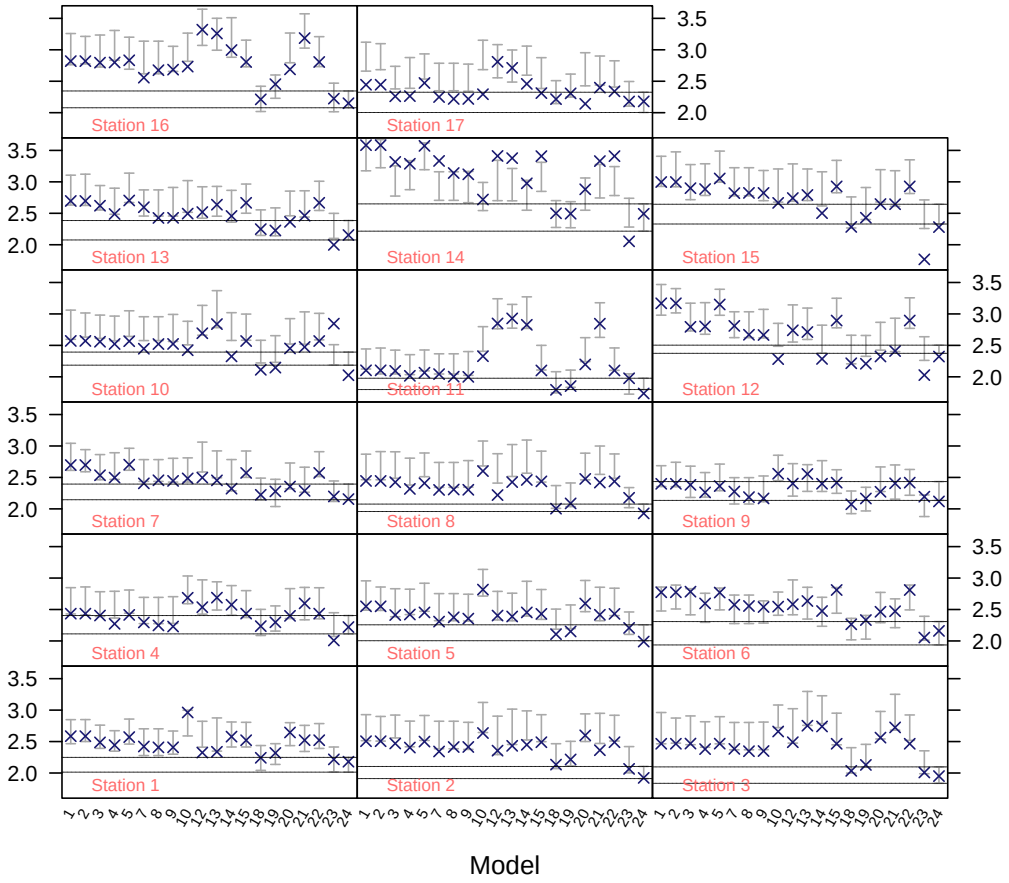


Figure 7.2: *Confidence intervals (95% C.I., n = 100) of MAE_{val} (grey vertical lines) and MAE_{test} (blue crosses) (MJ/m^2day) of the 24 parametric models evaluated. Note that some of the values of models 11, 16 and 17 lie outside the range of the figure*

The proposed methodology was useful to discriminate those models that generate wide confidence intervals of high MAE_{val} and at the same low MAE_{test}. Thus, the model can perform with low errors for a specific period of testing but very likely increase the errors in a different scenario. For instance, several models such as the 12, 13 and 14 at station 8, models 10 and 12 at station 12, models 12 and 13 at station 1 and models 12 and 20 at station 17, among others, behaved in that way. In the bibliography parametric models were always evaluated calibrating

with a specific dataset and testing with other, which migh result in this situation of low stability and at the same low errors of testing.

Following this method, stable and accurate models such as model 24 (proposed and built in this article based on the variable importance) should be selected at station 17 instead of model 20, although the latter generates lower MAE_{test}. The applicability of this method is of special interest with short datasets available for the evaluation of models. In order to select the most suitable model for the study region, four different variables were created for this purpose, the mean confidence interval width of MAE and RMSE ($\overline{R_{MAE,val}}$ and $\overline{R_{RMSE,val}}$) and the average MAE and RMSE ($\overline{MAE_{val}}$ and $\overline{RMSE_{val}}$). Table 7.2 collates the values of these variables for each of the models. It is noticeable that the proposed models (23 and 24) outperformed the other models with $\overline{R_{MAE,val}}$ as low as 0.360 and 0.261 MJ/m2day, respectively. Thus, the model 23 achieved the best errors by means of rainfall and temperatures. Nevertheless, a striking improvement was obtained in terms of robustness and accuracy by including two additional variables: the wind speed and relative humidity, used in the proposed model 24. In this case, $\overline{MAE_{val}}$ of 2.195 MJ/m2day and $\overline{RMSE_{val}}$ of 2.879 MJ/m2day were achieved, significantly lower values than in the other 23 models.

The unpaired t-test was also evaluated to determine whether averages of MAE_{val} were statistically different between pairs of models within each station. According to this test, models 23 and 24 built with the proposed method of variable importance were statistically more robust (lower MAE_{val}) than any other model in all stations. Regarding the other models, the original Bristow and Campbell (1984) model was superior to other derived from the original models, which was also the case with the original Hargreaves (1981) model. This fact proves the extremely local validity of these models, since where they were originally validated surpassed the general models but not in other regions such as La Rioja. Therefore, it is evident the benefits of using a method to adjust original models according to the local conditions, for instance, the proposed model adjustment based on the variables importance.

Table 7.2: *Summary of statistics in MJ/m^2day of the 24 parametric models tested*

$Model$	1	2	3	4	5	6	7	8
$\overline{MAE_{val}}$	2.81	2.81	2.70	2.68	2.80	2.77	2.53	2.62
$\overline{R_{MAE,val}}$	0.44	0.42	0.43	0.43	0.41	0.43	0.42	0.42
$\overline{RMSE_{val}}$	3.57	3.56	3.48	3.45	3.54	3.49	3.41	3.30
$\overline{R_{RMSE,val}}$	0.56	0.55	0.60	0.57	0.55	0.54	0.58	0.61
$Model$	9	10	11	12	13	14	15	16
$\overline{MAE_{val}}$	2.61	2.79	4.43	2.79	2.80	2.75	2.72	6.30
$\overline{R_{MAE,val}}$	0.42	0.42	0.76	0.53	0.49	0.49	0.44	0.76
$\overline{RMSE_{val}}$	3.40	3.59	5.87	3.83	3.80	3.71	3.49	7.38
$\overline{R_{RMSE,val}}$	0.60	0.58	1.00	0.75	0.72	0.69	0.58	0.80
$Model$	17	18	19	20	21	22	23	24
$\overline{MAE_{val}}$	3.40	2.32	2.33	2.68	2.73	2.72	2.25	2.20
$\overline{R_{MAE,val}}$	0.50	0.39	0.39	0.44	0.50	0.43	0.36	0.26
$\overline{RMSE_{val}}$	4.26	3.02	3.08	3.46	3.69	3.50	3.00	2.88
$\overline{R_{RMSE,val}}$	0.65	0.55	0.54	0.61	0.70	0.58	0.54	0.36

The proposed model 24 was able to achieve a high level of certainty when compared to the tolerance of pyranometers, with which solar irradiation is measured. In the case of stations belonging to SIAR (10% of pyranometers tolerance) and SOS Rioja (5% of tolerance), 41.65% and 20.12% of the daily residuals fell within those tolerance and therefore, being equivalent to a pyranometer. These differences were lower attending to the monthly sums of daily irradiation with 91.7% and 45.8% of the cases in SIAR and SOS Rioja within the tolerance. But regarding to the annual sums of daily irradiation, value that is generally assumed for solar irradiation mapping to characterize a territory in terms of solar irradiation, all stations were strikingly within a 5% of divergence (maximum value of 4.8%). Thus, this proposed parametric model and a pyranometer of 5% tolerance were equivalent for annual sums of daily irradiation.

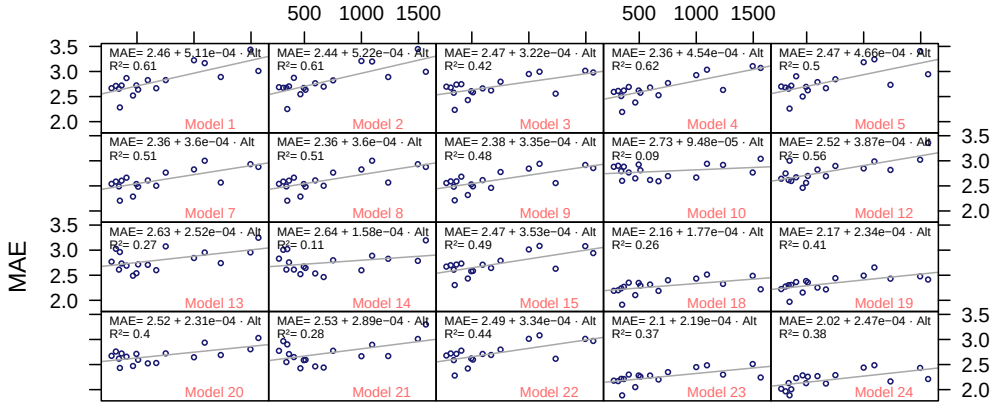


Figure 7.3: *Relation between elevation (m) and median of the MAE_{val} (MJ/m^2day)) of the 24 parametric models evaluated. Models 11, 16 and 17 are not shown due to their high MAE_{val}*

The influence of elevation on the performance of parametric models was observed in most models as shown in Figure 7.3, in which stations located at higher altitude than 1000m presented higher errors. This can be explained by the greater meteorological variability at higher altitude, generally associated in this region with more nubosity, situation in which parametric models fail more. For instance, the differences in MAE for the model 24 were on average 11.3% lower discriminating between non rainy days and rainy days for the whole set of stations. This finding was also proved with the other parametric models and it is explained by the fact that solar irradiation is more complex to be estimated on rainy and overcast days that in perfectly clear sky days Jong and Stewart (1993). Furthermore, it can explain the low MAE_{test} observed in Figure 7.2 by the fact that 2011 was a especially dry year in La Rioja with on average 19.7% less rainfall than in the training period (2007-2010).

7.2 Results in Publication II

All results in this section are collated in the paper "Generation of daily global solar irradiation with support vector machines for regression" Antonanzas-Torres et al. (2015).

The main objective of this study was proving the benefits of an automatic procedure for training a general support vector regression (SVR) model for solar irradiation estimation based on commonly measured meteorological variables.

The general model performance was compared against local models, one trained in each station. The high difficulties of selecting important variables for solar irradiation estimation and avoiding over-fitting were surpassed with a wrapper-scheme with genetic algorithms for input selection. The method was proved useful in Spain using 14 meteorological stations depicted in Figure 7.4, under different climates and terrain.

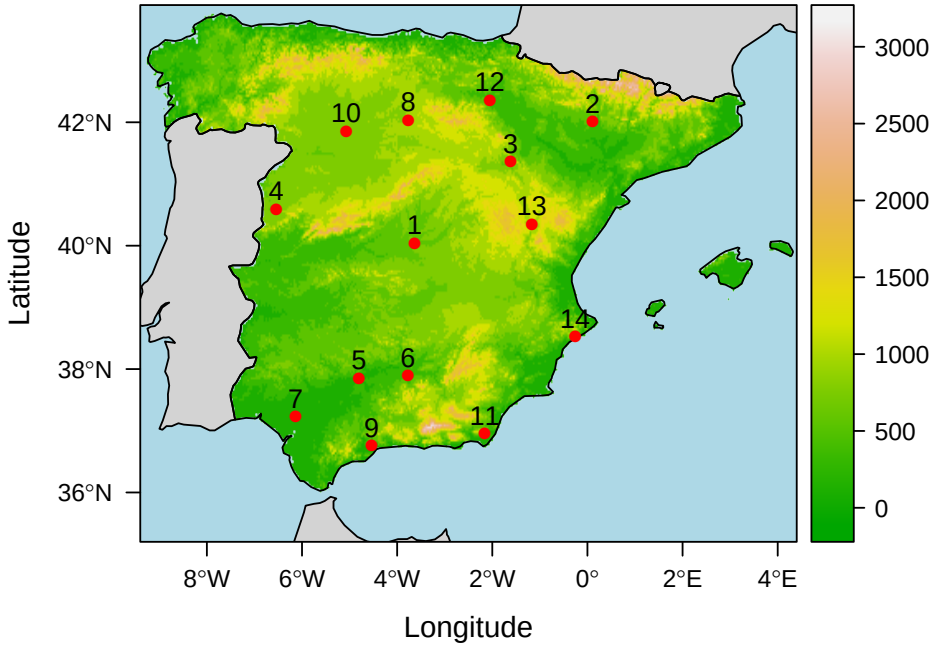


Figure 7.4: *Topographic map of the positions of meteorological stations (elevation in m) considered for the study*

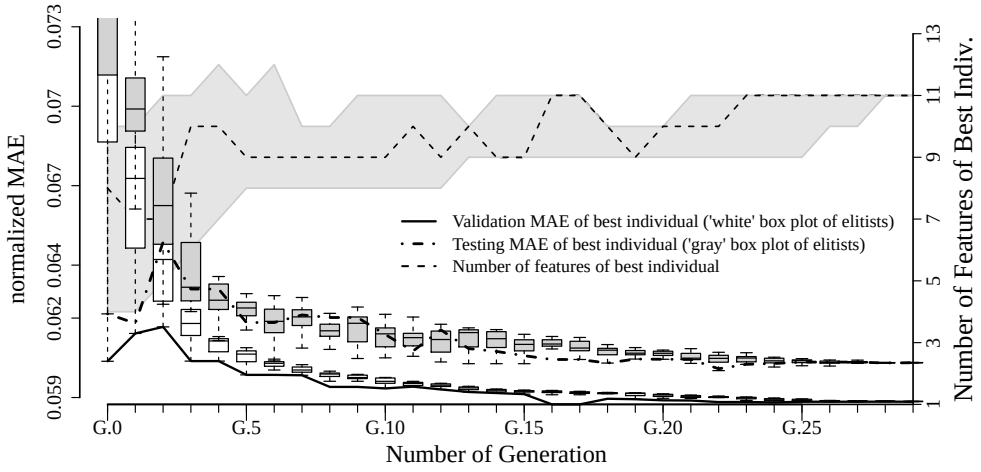


Figure 7.5: *Evolution of MAE_{val}, MAE_{test} and number of selected variables along the different generations of the GA-optimization procedure for the general SVR*

	$Feat.$	T_{med}	T_{max}	T_{min}	HR	WS	R	$R_{a,i}$	ΔT_i	ΔT_{i-1}	ΔT_{i+1}	M_i	M_{i-1}	M_{i+1}	C	ϵ	γ	$Time$
SVR_1	10	x	x	x	x	x		x	x	x	x	x			1.47	0.05	0.08	154
SVR_2	7	x		x	x	x	x	x	x						3.09	0.00	0.13	146
SVR_3	9	x	x	x	x	x		x	x		x	x			2.09	0.06	0.07	21
SVR_4	9	x	x		x	x	x	x	x	x	x				2.22	0.12	0.08	228
SVR_5	7	x		x	x	x		x	x	x					1.47	0.06	0.14	226
SVR_6	10	x	x	x	x	x		x	x	x	x	x			1.94	0.06	0.10	276
SVR_7	9	x	x		x	x	x	x	x	x	x				2.13	0.04	0.10	261
SVR_8	9	x	x		x	x		x	x	x	x	x			1.21	0.07	0.11	241
SVR_9	8	x	x		x	x	x	x	x					x	1.76	0.05	0.13	270
SVR_{10}	10	x	x	x	x	x		x	x	x	x	x			1.56	0.07	0.07	137
SVR_{11}	10	x	x	x	x	x	x	x	x			x		x	2.91	0.07	0.08	235
SVR_{12}	9	x	x		x	x		x	x	x	x	x			1.83	0.04	0.08	153
SVR_{13}	10	x	x	x	x	x		x	x		x	x	x		6.02	0.07	0.04	116
SVR_{14}	8	x	x	x	x	x		x	x			x			1.05	0.09	0.11	289
$SVR_{gen.}$	11	x	x	x	x	x		x	x	x	x	x		x	1.96	0.05	0.07	629

Table 7.3: *Parameters of the best individual for local SVRs (per station) and for the general SVR obtained with the GA-based optimization process. $Feat.$ stands for the number of inputs selected by the methodology. The last column lists the computational time in minutes*

On a first step, daily meteorological time-series were normalized (0-1) and divided into two datasets, one for training and other for testing. The testing dataset comprised the year 2013, while the training dataset collated 2000-2012. In order to reduce the high computational cost associated to the development of the general SVR, it was performed sampling a proportional stratified 10 % of the training dataset. From the best combination of variables and parameters of the SVR obtained applying the methology proposed, again a new general SVR model was trained using all the data from the training dataset. Figure 7.5 depicts the training process showing of the normalized MAE (average of 10 cross validation (CV)) for validation MAE_{val} and testing MAE_{test} and the number of variables selected. Although convergence of errors was almost achieved in the first generations using genetic algorithms, the convergence of number of variables took more generations (around 23). The normalized MAE_{test} and MAE_{val} reached similar values in the range of 6% with 11 variables, which is related with no model over-fitting. These variables are shown in Table 7.3 both for the general SVR model and also for a SVR trained with the same method in each of the stations. It is noticeable that both the general and local models required relatively similar variables (8-11 variables). The extraterrestrial irradiation, the daily range of temperatures, relative humidity and temperatures were selected in most of the models. These variables are also selected in most of the parametric models in the literature.

The SVR models were able to significantly improve the results of the parametric models analyzed (Bristow and Campbell, 1984; Antonanzas-Torres et al., 2013), with average MAE_{test} of 1.78 MJ/m^2day and 1.81 MJ/m^2day for local and general SVR models, respectively. This is around twice the tolerance of

the First Class pyranometers (5 % of daily uncertainty) used in the study, which is about 0.9 MJ/m^2day. These results proved the high capacitiy of generalization of SVR models, which were also superior compared to other techniques for the same region (2.05 MJ/m^2day with a locally trained parametric model (Almorox et al., 2011) and 2.33 MJ/m^2day with a general model built on an artificial neural network (Moreno et al., 2011)). It is also noteworthy that the general SVR performed better in half of the stations. This was explained by the higher capacity of generalization of the general SVR model to unseen events in a location that might have happened in other sites. Nevertheless, due to the simplicity of parametric models, they performed better with locally trained models than with a general model.

A strong trend of higher errors in the winter time was detected in contrast with the summer, which was previously found by Antonanzas-Torres et al. (2013) and explained by the higher cloud cover, which leads to higher errors (as explained in the following Chapter 7.3). The proportion of days in which the estimation was within the tolerance of pyranometers strikingly ranged between 21-47%. In these cases, the general SVR model and a First Class pyranometer were equivalent for recording daily global solar irradiation. Furthermore, regarding the annual sums of solar irradiation, commonly required for solar energy planning and mapping, the SVR general model performed within the 5% error at 11 out of 14 stations, significantly better than using the same inputs but using parametric models (5 and 8 stations for (Bristow and Campbell, 1984; Antonanzas-Torres et al., 2013), respectively). The highest annual error with the general SVR was as low as 7.4%.

One of the key aspects of developing a general model for estimating solar irradiation is the idea that it might perform better when it is necessary to use it in a different site from the training. Thus, the capacity of spatial generalization was evaluated with general SVR models and parametric models. Figure 7.6 depicts these differences in spatial MAE between a general SVR model and locally trained SVR models. The average cross validation MAE for SVR, obtained from training general models in 13 stations and testing in the station left, was 1.86 MJ/m^2day, very similar to the testing MAE of the general SVR (1.81 MJ/m^2day). This finding also occurred with parametric models, which proved something that was expectable: general models are better for spatial estimation than locally trained models. The MAE in stations in the center and South of Spain were lower compared to those in the North. This can be explained by the higher meteorological complexity of this area. Compared to general models, local models tend to fail at predicting irradiation in new locations due to the over-fitting to locally specific conditions.

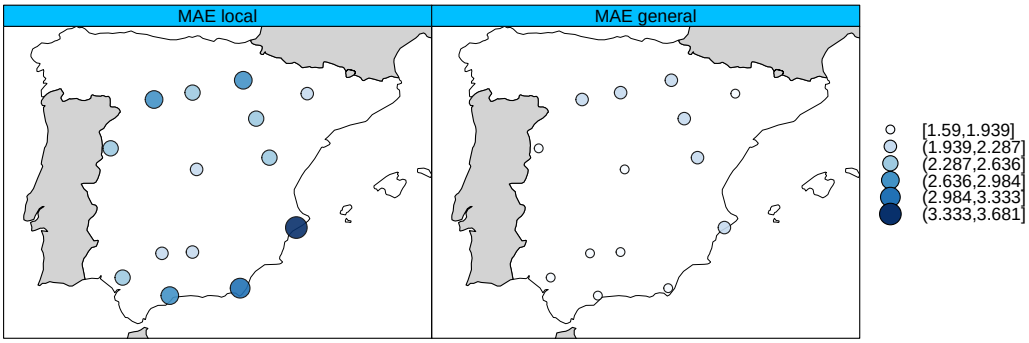


Figure 7.6: *Bubble plot of the MAE of locally-trained (left) and general (right) models. In the left plot, the average MAE of 13 locally trained models are evaluated at the 14th station. In the plot on the right, the MAE of a general model trained with a 13-station database is assessed at the 14th station*

7.3 Results in Publication III

All results in this section are collated in the paper "Sensitivity of satellite-based methods for deriving solar radiation to different choice of aerosol input and models" (Polo et al., 2014).

This study was motivated by the great range of information sources for obtaining aerosol and other atmospheric data, the variety of models to estimate solar irradiance under clear sky conditions and also models to compute direct normal irradiance from global irradiance. The objective was to quantify the range of errors under different choices of models and data and the way uncertainties propagate in the framework of a satellite-based method for estimating solar radiation. Although most clear sky models calculate the three components of solar irradiance, the reality is that satellite-based methods only calculate global irradiance, and then direct normal irradiance is calculated on a second step for CSP applications. Table 7.4 shows all combinations considered. R1, R2 and R3 serve for the assessment of the sensitivity of clear sky models (ESRA (Rigollier et al., 2000), SOLIS (Ineichen, 2008) and REST2 (Gueymard, 2008)) with accurate attenuants input from the Aerosol Robotic Network (AERONET). A1, A2 and A3 evaluate the uncertainty due to the choice in the source of aerosol data (Moderate Resolution Imaging Spectroradiometer (MODIS); Multi-angle Imaging

SpectroRadiometer (MISR); Monitoring Atmospheric Composition and Climate (MACC)). C2 and C3 evaluate the uncertainty due to the clear sky model selected and finally, D2 and D3 assess the uncertainty due to global-to-direct conversion models (Louche (Louche et al., 1991), DirInt and DirIndex (Perez et al., 1992)).

Table 7.4: *Cases considered for the sensitivity study regarding the kind of atmospheric data used, the clear-sky model and the global-to-direct model*

	Atmospheric				Clear			Global-to-direct		
Case	AERONET	MODIS	MISR	MACC	ESRA	SOLIS	REST2	Louche	DirInt	DirIndex
R1	x				x			x		
R2	x					x		x		
R3	x						x	x		
A1		x			x			x		
A2			x		x			x		
A3				x	x			x		
C2				x		x		x		
C3				x			x	x		
D2				x					x	
D3				x						x

The study was performed with data from the Solar Platform of Almeria (PSA) in an area of special interest for solar power plants (PV and CSP). In order to benchmark the results, ancillary measurements of solar irradiance (global, diffuse and beam) were collected from the Baseline Surface Radiation Network (BSRN) at the PSA. Furthermore, high quality atmospheric components from on-ground records were obtained for the same location from AERONET. Results are firstly presented for all kind of sky conditions and on a second step, disaggregated into clear-sky and non clear sky conditions. Besides, it is performed an analysis of direct normal irradiance specifically for the operational range for CSP. Eventually, a novel evaluation on the propagation of uncertainties from aerosol optical depth to irradiation was performed for clear-sky models.

Table 7.5: *Summary of errors and statistics for clear-sky conditions for the different cases analyzed in the sensitivity study*

Case	R1	R2	R3	A1	A2	A3	C2	C3	D2	D3
GHI										
MBE (Wm^{-2})	-0.023	0.004	-0.035	0.012	-0.008	-0.008	0.047	-0.018	-0.008	-0.008
rMAE (%)	7.43	9.04	9.14	7.67	8.51	7.64	10.74	9.9	7.64	7.64
RMSE (Wm^{-2})	42.29	45.67	47.74	35.17	41.5	41.23	49.31	45.05	41.23	41.23
R^2	0.986	0.981	0.987	0.989	0.985	0.985	0.983	0.985	0.985	0.985
DNI										
MBE (Wm^{-2})	-0.051	-0.063	-0.017	0.054	0.02	-0.004	0.029	0.028	-0.069	-0.001
rMAE (%)	12.41	14.79	10.11	14.29	17.88	13.62	13.56	14.06	13.15	13.14
RMSE (Wm^{-2})	201.14	148.71	116.39	125.51	155.83	154.37	126.11	128.19	128.81	132.16
R^2	0.579	0.733	0.79	0.806	0.69	0.694	0.791	0.784	0.78	0.764

Table 7.6: *Summary of errors and statistics for cloudy and overcast conditions for the different cases analyzed in the sensitivity study*

Case	R1	R2	R3	A1	A2	A3	C2	C3	D2	D3
GHI										
MBE (Wm^{-2})	0.046	0.058	0.049	0.084	0.029	0.061	0.126	0.062	0.061	0.061
rMAE (%)	17.44	18.14	17.58	18.14	17.5	17.79	19.55	18.04	17.19	17.79
RMSE (Wm^{-2})	127	129.82	127.26	128.16	127.21	127.12	133.37	127.49	127.12	127.12
R^2	0.817	0.809	0.82	0.817	0.816	0.817	0.813	0.819	0.817	0.817
DNI										
MBE (Wm^{-2})	0.033	0.05	0.066	0.155	-0.007	0.082	0.233	0.107	0.104	0.152
rMAE (%)	20.77	22.53	22.79	24.47	21.52	22.57	27.89	24.34	22.58	24.06
RMSE (Wm^{-2})	222.65	230.68	228.58	230.05	234.02	225.49	237.29	232.52	294.84	233.34
R^2	0.563	0.53	0.546	0.541	0.524	0.549	0.548	0.531	0.511	0.523

Table 7.5 shows the errors for all scenarios under clear-sky conditions. It is noticeable the low bias in estimations both for GHI and DNI in all cases. Errors were higher in DNI than in GHI due to the higher complexity of estimating DNI (derived from the difficulties of cloud impact calculation). The relative mean absolute error (rMAE) ranged 7-11% for GHI and 10-15% in DNI. The best performance in GHI was obtained using high quality atmospheric data from AERONET as input and the ESRA clear sky model, while for DNI was for REST2 with AERONET data. Regarding the comparison with other low spatial resolution aerosol datasets, results using MACC showed an acceptable performance. Besides, the REST2 model only was able to obtain accurate results with high quality data, while a more simple model, such as ESRA, performed better with satellite or reanalysis gridded aerosol inputs. DirInt and DirIndex generated better results that the specific global-to-direct conversion of clear sky models.

Table 7.6 shows errors for cloudy and overcast conditions. These errors kept higher in DNI than GHI, same as for clear sky conditions. rMAE for not clear-sky conditions ranged 17-18% and 20-28% for GHI and DNI, respectively. The R^2 was in the range of 0.8 and 0.55 for GHI and DNI. Both MISR and MACC obtained better results than MODIS for GHI and DNI. As for clear-sky conditions, the ESRA model showed a slightly better performance than the other clear sky models.

Furthermore, it is analyzed the particular case of DNI estimation for the operational range for the concentrated solar power (CSP), 400-900 W/m^2, the range of irradiance for which CSP plants are designed. The REST2 model using data from AERONET yielded the best results, while results with MACC were superior than using MODIS or MISR. DirIndex yield to better results than DirInt.

Finally, the uncertainty of clear sky models is evaluated synthetically by perturbing the aerosol optical depth (AOD) on a ±50%, and evaluating the

output of ESRA, SOLIS and REST2 models while the other inputs are fixed at constant values.

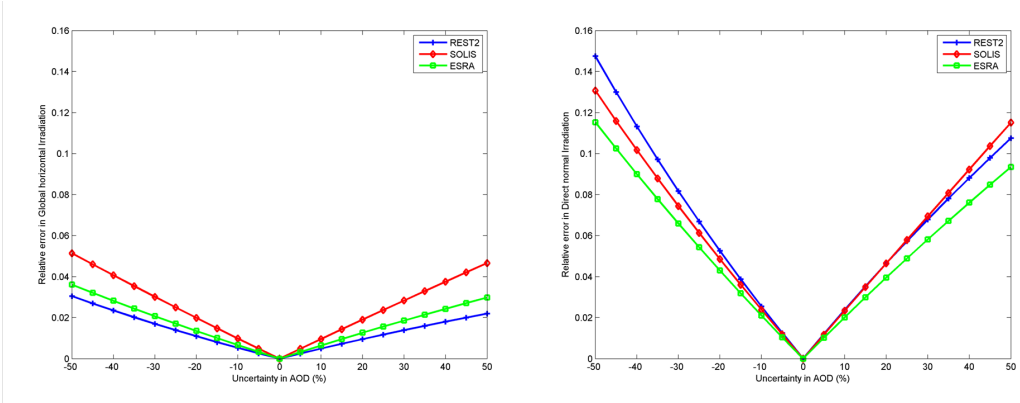


Figure 7.7: *Impact of the uncertainty of aerosol optical depth on the clear-sky models output. Assuming zero error for a given initial condition of AOD, the plot shows the evolution of the error of each model when AOD is increasing or decreasing from the starting point. The left plot corresponds to the global horizontal irradiance and the right plot to the direct normal irradiance*

Figure 7.7 depicts the errors of the clear-sky model errors on a daily basis as a function of the uncertainty in the AOD. The impact of the AOD uncertainty was significantly higher for DNI than for GHI. REST2 was the least sensitive to the uncertainty on GHI, but on the contrary, it was very sensitive for DNI. The ESRA model showed more parsimonious for DNI than SOLIS and REST2. All models showed a bit more sensitive to under-estimation of AOD than to over-estimation.

The previous figure of the synthetic analysis must be complemented with Figure 7.8, which shows the propagation of errors for the real case of study in the PSA. The comparison was performed with real measurements of GHI and DNI. Thus, minimum errors are located in different positions for each models in this figure. While the propagation of errors was quite soft for GHI, for DNI was strikingly strong. REST2 was generally the most accurate model for DNI. However, ESRA yielded to higher errors than SOLIS and REST2 with under-estimation of AOD, while it was the most accurate for over-estimation of AOD >30%. This finding explains the noticeable good performance with MACC retrievals in DNI, which were found to over-estimate around 30 % the AOD. As a result, the different behavior of clear-sky models to the errors in AOD explains the previous results of the sensitivity scenarios.

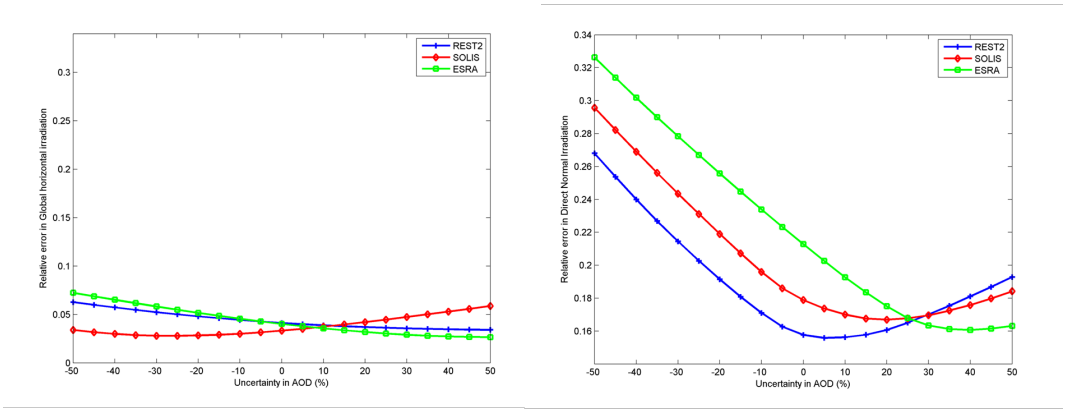


Figure 7.8: *Impact of the uncertainty of aerosol optical depth on the clear-sky models output. Assuming zero error for a given initial condition of AOD, the plot shows the evolution of the error of each model when AOD is increasing or decreasing from the starting point. The left plot corresponds to the global horizontal irradiance and the right plot to the direct normal irradiance*

The results presented in this publication show the different sources of uncertainties that affect satellite-derived models of solar radiation. Thus, the first and most important uncertainty came from the impact of the cloud cover in the estimation of solar irradiation. This was explained by the general and remarkable differences between GHI and DNI errors. Other second order factors were the sources of aerosol inputs and global-to-direct models. Regarding the source of aerosols, the most accurate and available database (AERONET), using sun-photometers, generated the best results with DNI, the component of solar irradiation in which aerosols are more important. Nevertheless, since the presence of sun-photometers in meteorological stations is very rare, solar irradiation is normally approached via gridded datasets from satellite or reanalysis. Using these gridded databases, the uncertainty of these variables that govern solar radiation have a different impact basically depending on the clear sky model used and the component of solar irradiation.

7.4 Results in Publication IV

All the results described in this section are extensively described in the paper "Comparative assessment of global irradiation from a satellite estimate model (CM SAF) and on-ground measurements (SIAR): a Spanish case study" Antonanzas-Torres et al. (2013).

The main objective of this article was the development of a methodology to improve the accuracy of solar irradiation maps in different tilted planes with

special interest in photovoltaics (effective irradiation on a fixed plane, north-south horizontal axis tracking and two axis tracking on elevation and azimuth) by combining satellite-derived irradiation data and on-ground measurements. All the programming code in `R` environment was freely released for the implementation of this methodology in other regions.

Initially, the satellite-derived database of solar irradiation by the Climate Monitoring Satellite Application Facility (CMSAF, 2015) was compared against on-ground GHI measurements in 301 meteorological stations in continental Spain. The 71% of these stations were in yearly values of GHI with a difference lower than 5% between both sources: CMSAF and measurements. This 5% represents the intrinsic daily uncertainty of the First Class pyranometers installed in these stations. Furthermore, it is noticeable that 96.5% of the yearly values outside the 5% band were greater in CMSAF than in SIAR. This might be explained by several factors such as: the topographic attenuation by shades and sky view factor (both analyzed in Section 8.5) and not considered in the CMSAF, the effect of atmospheric events not taken into account by the CMSAF and also also due to the lack of some measurements and soiling on sensors in the meteorological stations. Regarding these yearly differences in the effective irradiation on three different tilted planes, the values increase with the complexity of the plane (lower differences for the fixed plane and greater for two-axis tracking), which might be due to the decomposition model used for extracting diffuse and beam components from global irradiation (Collares-Pereira and Rabl, 1979). This finding was consistent with previous studies (Suri et al., 2007; Perpinan, 2009). In this line, a seasonal pattern was found on the divergence between data sources, with a wider range in the winter time and lower in the summer, which is explained by the higher cloud cover in the winter time in Spain that leads to higher errors in the estimation (Urraca et al., 2015).

Secondly, the performance of the kriging with external drift (KED) using irradiation data from the CMSAF as explanatory variable was evaluated. The KED technique did not perform as an interpolation function in this study since the nugget effect of the semi-variogram was not null and therefore, there was an intrisic variability independent from the distance between meteorological stations. The solution achieved with the KED was a smoothing funcion of the on-ground measurements using CM SAF as the spatially known variable. In the case of GHI and fixed plane, a pure nugget variogram model was the most indicated, which was also considered in other studies (Antonanzas et al., 2015). Regarding the other tracking cases, a spherical variogram was considered. The spatial divergences between the map of yearly sums of daily GHI and effective irradiation on the fixed plane generated with KED and the homologous maps by the CMSAF showed that

KED provided higher values than CMSAF in the Cantabric coast (43.1ºN, 1W to 6W), while in Andalucia (southern part of continental Spain) the CMSAF overestimated the KED. Nevertheless, this spatial pattern changes in the two-tracking systems, in which in general terms the effective irradiation by the CMSAF was greater than with KED, although generally within the 5% uncertainty band.

7.5 Results in Publication V

All the results described in this section are collated in the paper "Downscaling of global solar irradiation in complex areas in R" Antonanzas-Torres et al. (2014).

The main objective of this paper was to develop a useful methodology to downscale and generate solar radiation maps of high spatial resolution in areas with complex topography and provide open-source code for replications of the study in other regions under similar conditions. The idea of releasing free open-source code in R environment for solar radiation downscaling deeply contrasts with other "black-box" approaches of the literature (Meteonorm, 2015; SolarGis, 2015). This method was proved useful with a striking 46.75% and 43.38% reduction in MAE and $RMSE$, compared to the original gross resolution database in a very mountainous region of Spain (La Rioja).

Although it is well known the impact of topography on solar radiation, satellite-derived solar radiation databases generally omit this attenuation. As a result, in these models it is considered a perfectly flat Earth surface and the shading is only taken into account as a variation of the pixel albedo in satellite images. However, the low spatial resolution of these images, generally in the range of kilometers, soften the impact of shadows and mask their impact when shades are sufficiently small. In the present study, the effect of topography was evaluated and integrated via shade analysis and sky-view factor (portion of visible sky). The method was applied to the satellite-derived solar radiation datasets from the Climate Monitoring Satellite Application Facility (CM SAF) in order to reduce the spatial resolution from the original 0.05º (˜5 x 5 km) to 200 x 200 m.

Although the method proposed is indeed part of the methodology and not specifically a result, it has been presented in this section since this method is one of most remarkable results of this publication, given that several calculation approaches were investigated until the best performing algorithm was obtained.

Figure 7.9 depicts the final calculation procedure. Initially, hourly values of global horizontal irradiation (GHI) and beam horizontal irradiation (BHI)

from the CM SAF were disaggregated into the final resolution (200 x 200 m), and the diffuse component (DHI) was obtained from the difference of GHI and BHI. On a second step the DHI was decomposed using the Hay and McKay (1985) model into the isotropic (DHI_{iso}) and anisotropic (DHI_{ani}) components; the isotropic corresponding to the solar radiation coming from all sky directions and the anisotropic from the Sun surroundings.

On a second step, the impact of horizon block was evaluated on the DHI_{ani} and BHI. As a result, the horizon was divided into 72 portions of 5º each and the horizon angle, angle between the point of study and the horizon, was calculated and compared against the solar elevation. In case the solar elevation angle was lower than the horizon angle it was assumed the effect of horizon block. It must be noted that for the Northern hemisphere only the 180º corresponding to the South direction would affect these components, but it was calculated for the whole horizon to extract the sky view factor, ratio of visible sky. The sky view factor affects the DHI_{iso} in a way that for instance, areas located at the bottom of a valley would receive a reduced portion of the DHI_{iso} and in the case of the highest peak in the area DHI_{iso} would be totally received.

The sum of all these components was denoted as GHI_{down}. However, the fact that GHI_{down} lowered the values of the original gross-resolution GHI within each pixel in complex topographic areas introduced a bias only in these regions. But, as previously stated satellite-derived irradiation considers someway shades based on the lower albedo registered as an average in the pixel. Thus, GHI_{down} was considered a valid explanatory variable of *kriging with external drift* (KED) to interpolate on-ground measured data from pyranometers. This step was denoted as *post-processing* of GHI_{down}. Based on the findings explained in the Results in Publication IV (Antonanzas-Torres et al., 2013) a pure nugget effect was considered for the semi-variogram of the KED, which is generally associated with intrinsic variability of solar radiation and with no spatial autocorrelation on residuals.

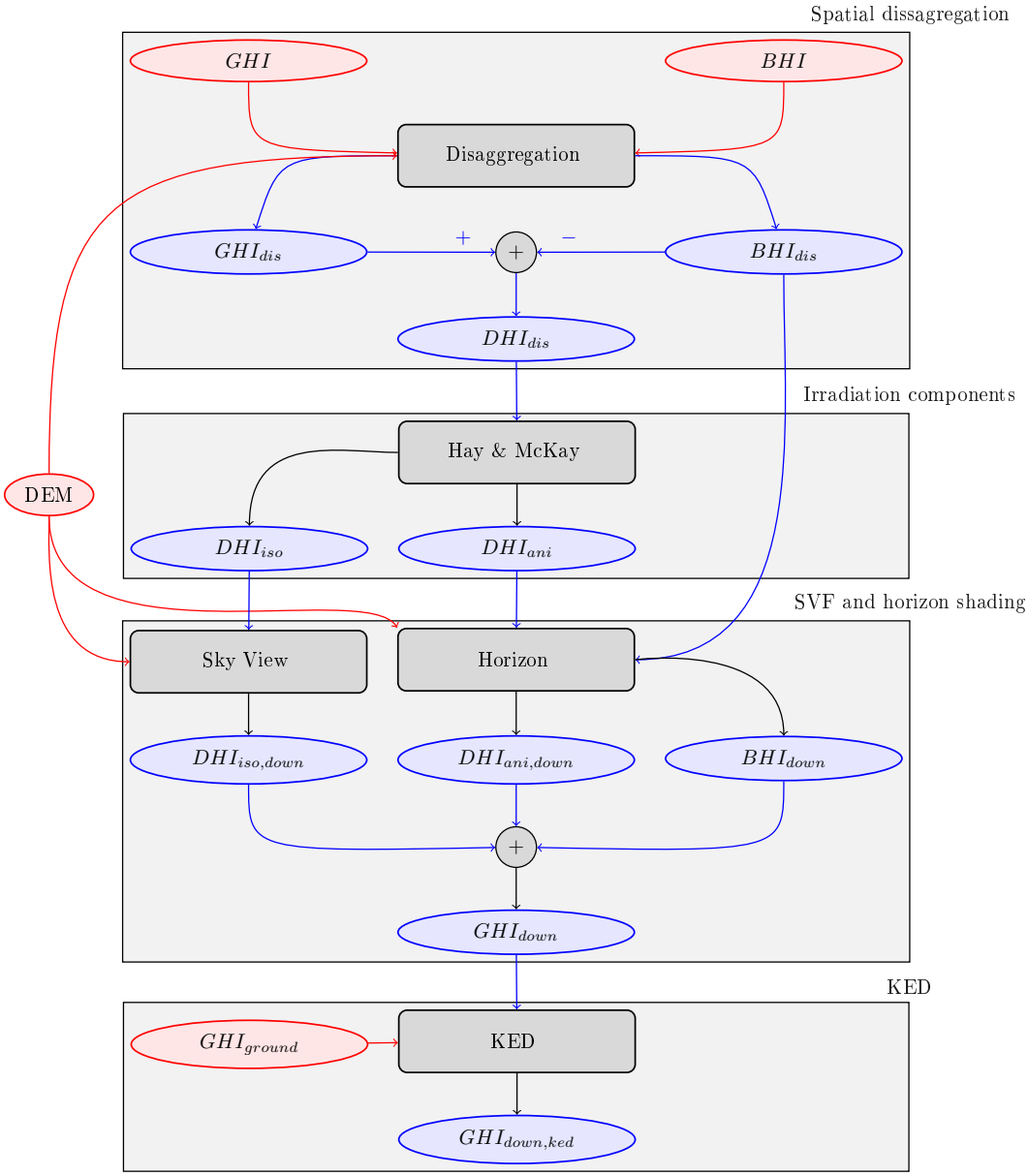


Figure 7.9: *Methodology of downscaling: this figure uses red ellipses and lines for data sources, blue ellipses and lines for derived rasters (results), and black rectangles and lines for operations. GHI and BHI stand for global and direct horizontal irradiation, respectively. DHI stands for diffuse horizontal irradiation. Subscripts dis stand for disaggregated, iso for isotropic, ani for anisotropic, down for downscaled and ground for measured values. KED and DEM stand for kriging with external drift and digital elevation model, respectively*

Figures 7.10 and 7.11 depict the original gross resolution solar radiation map as per CM SAF and the downscaled map obtained with the method proposed, respectively.

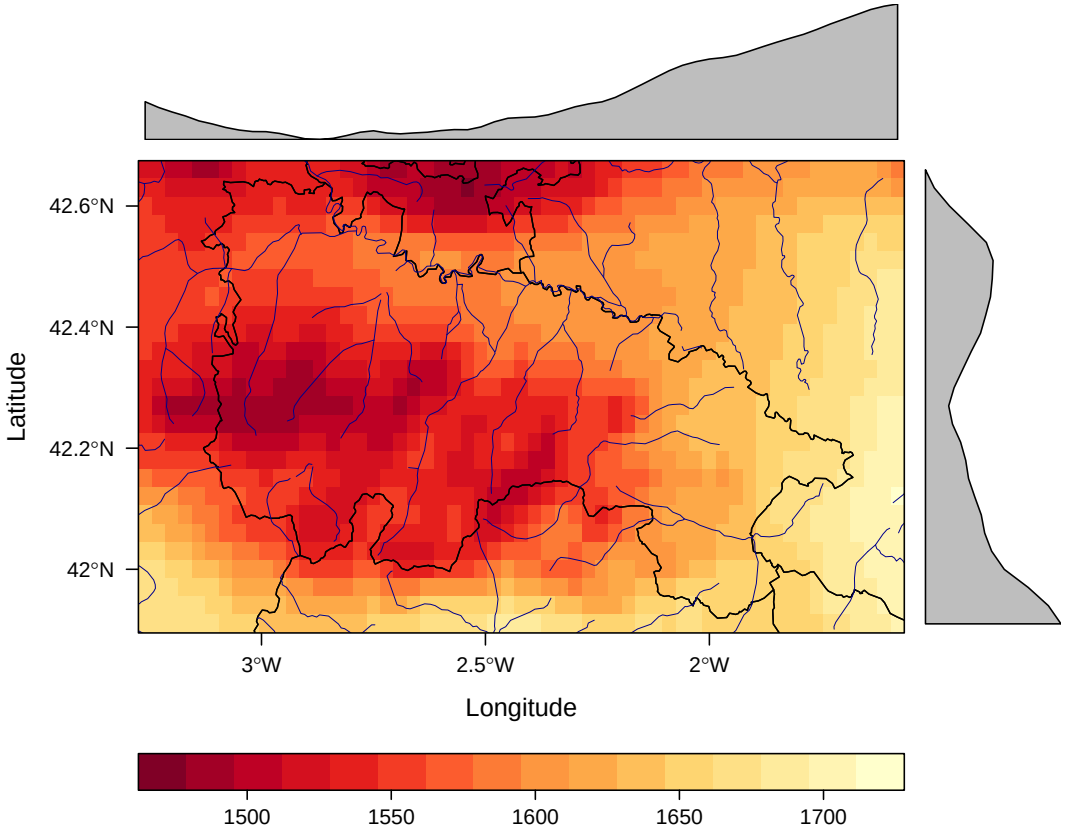


Figure 7.10: *Annual GHI of 2005 from CM SAF estimates (0.05 x 0.05º) in La Rioja (kWh/m^2)*

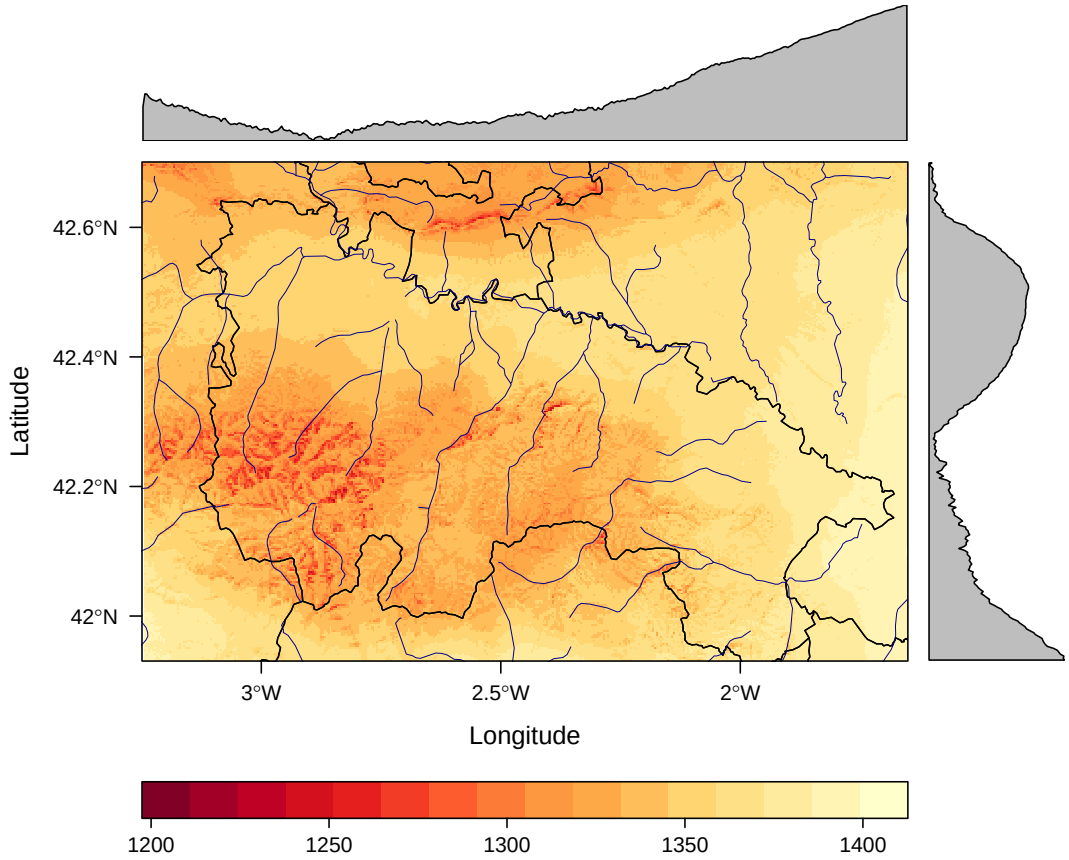


Figure 7.11: *Annual GHI of 2005 downscaled (200 x 200 m) in La Rioja (kWh/m^2)*

This methodology leads to an improvement in estimates. The MAE was strikingly down by 46.75% and the $RMSE$ by 43.38% compared to CM SAF. Table 7.7 shows the errors obtained with CM SAF and with the methodology proposed. Furthermore, the downscaling proposed obtains lower MBE (91.92 kWh/m^2) than the validation threshold by CM SAF (131.4 kWh/m^2) (Posselt et al., 2012). Nevertheless, this threshold value is surpassed in the case of CM SAF (175.62 kWh/m^2) in testing meteorological stations. Both, CM SAF and downscaling under-estimated GHI proved by the negative MBE.

Table 7.7: *Summary of testing errors obtained in kWh/m^2 for CM SAF and for the downscaling proposed*

	GHI_{cmsaf}	$GHI_{down,ked}$
MAE	172.62	91.92
$RMSE$	175.02	99.08
MBE	-172.62	-91.92

The downscaling with *KED* was able to improve the results of the CM SAF in four stations at an average range of 11.2% (including sites of testing and also used in the *KED*. Figure 7.12 depicts the zonal variability of solar radiation within each pixel due to the topography. It is noticeable that in some areas of the study region solar radiation can vary up to 60% within a pixel size of 0.05º. This finding is clearly indicative of the importance of considering the topography effect as attenuator of solar irradiation in areas of complex topography and therefore, downscale solar irradiation in these areas.

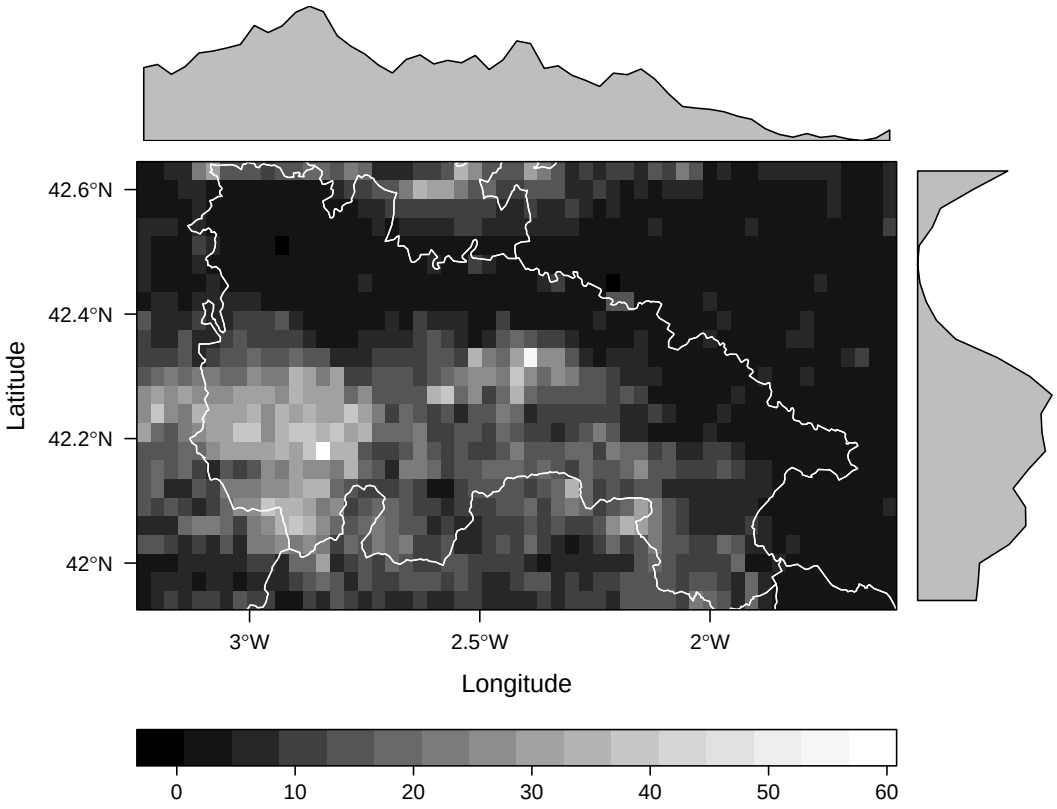


Figure 7.12: *Difference of zonal standard deviations between the downscaling $GHI_{down,ked}$ and the CM SAF GHI in per one units*

Chapter 8

General Discussion

This chapter presents the main implications and limitations of this work following the same structure of the Results (Chapter 7).

The main goal of this thesis was the estimation of solar irradiation with as much accuracy as possible without the use of pyranometers. Different methods and approaches were proposed combining a wide range of atmospheric records and data sources.

Traditionally, solar irradiation was estimated with models related to other commonly measured meteorological variables such as air temperatures and rainfall. In order to have an overview of the range of errors of these classic models, it was performed a wide review accounting twenty-two models. Throughout the evaluation of these models in seventeen sites of a small region of Spain (La Rioja), it was found a disagreement between the errors of models and sites. Thus, the accuracy of classic models was very dependent on each site condition, and also to the period used for calibration. It was noticed that errors obtained in this region were significative higher than those obtained where these models were originally proposed. This was explained in Antonanzas-Torres et al. (2013) due to the different importance of variables in each site and region. As a result, the rainfall has a different model importance in arid or humid areas, which also occurs with the temperatures. Furthermore, the calibration data assumes an important role in the model performance. So, if models are calibrated along a specifically sunny period and then used to estimate solar irradiation during a more cloudy period, higher errors are expected. Based on the importance of meteorological variables and the calibration period, a method to generate locally adapted models to the specific conditions of each site was proposed. Furthermore, the robustness of models was evaluated via training models with bootstrapping seeking for models with less variation in their parameters and errors. The new models generated

significantly improved the results of previous models in the literature, and despite their simplicity around 20% of the days were estimated within a 5% of error. It was also found that other commonly measured variables such as relative humidity and wind speed, scarcely considered in previous studies, also led to an improvement in errors. The MAE ranged 2.195 MJ/m^2day and the robustness of models was as low as $0.261MJ/m^2day$.

Nevertheless, despite the effort taken in selecting the most relevant variables in this area, the errors were still high to be acceptable for solar energy applications. So, the limits of the accuracy of solar irradiation estimation were investigated with much more complex machine learning techniques considering the same meteorological variables used in the previous parametric models. In a specific study for La Rioja region (Spain) (not included in this thesis) (Antonanzas-Torres et al., 2014), different techniques such as the multilayer perceptron, the M5P regression trees, the instance based learning (IBK) and the support vector regression (SVR) were compared with the results of the best performing parametric models and it was concluded that the support vector regression led to a significant improvement in errors compared to other techniques. From this initial insight of improvement in the solar irradiation estimation for a small region it came the idea of developing a general model capable of being more parsimounious and generalistic to new events and thus, avoiding the problem of local dependency.

As a result, the region selected was enlarged to the continental Spain area, including a higher diversity of meteorological events. In Antonanzas-Torres et al. (2015) it was found that the errors were improved with a general SVR model trained with data from several locations when evaluated in a not seen location, than a local model trained with a data from a single location and evaluated in different sites. Furthermore, the errors with the general SVR model were comparable to the local model when evaluated where trained. This implies that the SVR was able to learn from the different meteorological events. However, due to the simplicity of parametric models, local models were superior than the general model.

The SVR estimates with the general model were equivalent to a First Class pyranometer (5% of daily error) in 20-46 % of the instances of the stations, which in the case of parametric models it was around the 20%. These findings were the root for a further study (Antonanzas et al., 2015), in which solar radiation maps were generated with kriging out of a great number of meteorological stations without pyranometers, that were used with SVR as estimators of solar radiation for increasing the density of sites with solar radiation data.

The fact that some solar irradiation records are necessary to train models based on meteorological variables and, that these records are not always available makes satellite-derived solar irradiation a suitable source for these situations. However, these satellite estimations rely on inputs that are available in very low spatial resolutions (in the range of degrees) and therefore can be sensitive to these inputs. This was the main goal of our study (Polo et al., 2014). The cloud cover was the main factor affecting the accuracy of solar irradiation estimated with satellite images. Thus, GHI can be estimated with a higher level of certainty than the DNI, in which the impact of the cloud cover is deeper.

The source of the aerosols data and the global-to-beam models accounted as second order factors in the errors. Aerosol data from the sun-photometers of AERONET led to the lowest errors in DNI estimation. However, it must be stated that these AERONET records have many gaps, short registers and consequently, cannot be considered for solar resource assessments in most of the cases. Therefore, although higher errors are obtained, it is necessary to rely on gridded datasets from satellite or reanalysis for aerosols and water vapor. It was noticeable a great disaggrement between AERONET records and estimated values from MACC, MODIS and MISR. This disagreement can be due to the low spatial resolution of these estimates. Therefore, a higher resolution and accuracy in the estimation of these aerosols would lead to a better estimation of solar irradiation. This was the motivation of a different study developed by the author of this thesis in which the aerosol optical depth was downscaled for a better solar resource estimation using other exogenous variables (Antonanzas-Torres et al., 2014).

One of the limitations of the study by Polo et al. (2014) is that it was only analyzed for one location in the Solar Platform of Almeria (Spain) due to the complexity of the scenarios considered. Thus, other factors such as the influence of elevation on aerosols and in clear-sky models performance was not evaluated. This was the motivation of a different study (Antonanzas-Torres et al., 2016) developed in a very specific location in the Tenerife island (Canary Islands, Spain), in which two meteorological stations vertically separated 2200m and horizontally 11km were used for this study. The influence of elevation was noticeable when using aerosol data from the upper sites into the lower, with a strong trend to the over-estimation. One of the limitations we found in these studies is the difficulty in obtaining water vapor content data, required for estimating solar clear-sky irradiance in many models. Although the water vapor content is recorded in AERONET sites, these registers are generally short and with many gaps complicating the long term estimation of solar irradiation.

The fourth contribution of this thesis (Antonanzas-Torres et al., 2013) was the comparison between datasets from pyranometers and satellite estimations from the CM SAF in continental Spain. We found a noticeable good agreement between both sources on a daily and annual basis. The relative mean absolute difference on a daily basis was in most of the stations within a 5% and the relative root mean square difference within a 7%. Indeed, the yearly difference between both sources was at 71% of the sites within 5% (the pyranometers uncertainty), and outside this 5%, 96.5% of the pyranometers provided lower irradiation than the CM SAF. Regarding, the irradiation incident on the inclined planes it was comprised between 5% and 6%. We used kriging with external drift (KED) using the spatially known irradiation data from CM SAF as exogenous variable, as a tool to attenuate differences between both sources and somehow adjust the CM SAF estimations to real measurements. Thus, differences between KED estimates and measurements were lowered to a mean absolute difference of less than 3.8% for all planes. This method can be of high applicability in other regions where some pyranometers are available, since nowadays most of the world is covered with CM SAF free satellite data. We released the code in `R` environment and also the data for future replications of the study. The main limitation of this study is that we neglected the effect of topography on surface incoming solar irradiation. However, due to the fact that solar energy power plants are generally installed on flat areas this method provides a relatively high confidence for this purpose.

The fifth contribution (Antonanzas-Torres et al., 2014) was motivated by considering the effect of topography on solar irradiation. In this case, we downscaled satellite irradiation from CM SAF to a very fine spatial resolution (200x200m) considering the shades and the sky-view factor. Finally, we applied kriging with external drift using the downscaled raster layer (considering shades and sky-view factor) as exogenous variable. The study was performed in La Rioja region (northern Spain) and noticeable differences were found between downscaled irradiation and CM SAF in the mountainous areas (up to 22% on a yearly basis). The method was freely released for future replications of the study. The main drawback for its applicability is the high computational cost of calculating the horizon angle for each of the pixels. Although in the northern hemisphere the shades are only calculated for the south direction, the sky view factor must be calculated for 360º. For this reason, parallel-multicore computing was utilized to speed up the process.

Chapter 9

Conclusions

In the framework of solar radiation modeling, different approaches were explored and improved, providing useful methods and code for future replications of the study in other regions. The results revealed an outstanding number of possibilities for estimating solar radiation with a relatively high level of certainty without the necessity of pyranometers.

The contributions of this thesis were organized into two main blocks of studies. The first block collates the temporal models seeking to generate accurate solar irradiation time series. The second block was oriented to the spatial models for the estimation and mapping of solar irradiation.

The research presented in this thesis provided the following main findings:

1. Parametric models for solar irradiation estimation are not universal. The usage of parametric models should be accompained of an onsite adjustment of these models not only based on the calibration of parameters, but also on the relative importance of available meteorological variables with solar irradiation. Due to the high diversity of climates in our planet, the importance of these meteorological variables differs from one location to other. Furthermore, it must be taken into account the robustness of these models to be calibrated with different data from the same location (in other words, how much parameters and errors vary depending on the training period).
2. The performance of support vector machines for regression trained with the method based on feature selection and parameter tuning with genetic algorithms proved superior to parametric models. This technique showed a extremely high capacity of spatial generalization when trained with data from different sites, so the errors were improved in some cases against locally trained models. Since, most of the stations performed within the 5% of an-

nual error (tolerance of First Class pyranometers), this methodology is very promising for solar resource mapping, wherein annual sums are normally considered.

3. Regarding the study of uncertainties in satellite-derived irradiation, the most important source of uncertainty comes from the impact of clouds. Other second order factors are due to the aerosol information and the clear-sky model selected. Although some clear-sky models can lead to a better estimation of solar irradiation with a satellite-based method than others, this is a function of the quality/uncertainty of the aerosol data used. For this reason, the choice of a specific clear-sky model in a satellite-based method clearly depends on the knowledge of the general uncertainty of the aerosol data to be used as input.
4. A method based on kriging with external drift using data from pyranometers and the satellite-based CM SAF as exogenous variable led to smooth and adjust CM SAF estimations to measured data. This method proved very promising for solar radiation mapping on the horizontal plane and effective irradiation on three different tilted planes (fixed, one axis tracking, two axis tracking). Furthermore, this method can be utilized for other meteorological variables mapping and downscaling if provided as spatially gridded data.
5. A method for solar irradiation downscaling using spatial data from a satellite-based method and taking into account the topography effect and kriging with external drift was proved useful in mountainous regions. We observed that areas with more complex topography suffered from a much higher divergence respect satellite-based data from CM SAF. Although the topography impact is generally neglected by satellite-based models, considering this factor clearly contributes to a significant improvement in the spatial estimation of solar irradiation.

Chapter 10

Future works

All in all, the main findings of this thesis contribute to a better knowledge of solar irradiation modeling. These methods can be of useful application for solar resource assessments in those areas with lack of pyranometers, or to reconstruct and generate long time series and maps where desired.

During the progress of this thesis, we found some exciting questions and issues that could be a matter of research in the future:

1. Study of solar irradiation in areas with high albedo. We have found that satellite estimates frequently confuse the ground albedo with the cloud cover in areas with high albedo, such as deserts and snow covered regions leading to errors. Particularly, due to this albedo confusion the CM SAF generates not available (NA) datasets in some of these regions(Riihela et al., 2015). Furthermore, deserts and snow covered regions usually present a high aerosol content due to the propagation of dust and snow particles with the wind. However, the access to high quality aerosol data for clear-sky solar irradiance estimation is limited and normally, low spatial resolution gridded estimates are considered, which also introduces uncertainty in the estimation of solar irradiation. Additionally, there is a lack of pyranometers in these regions and some of the time series are not complete due to the extreme conditions that in some cases make unable the measurements. In the line of some of the studies we have performed in the EDMANS group using soft-computing techniques for the estimation and mapping of solar irradiation (Antonanzas-Torres et al., 2015; Antonanzas et al., 2015; Urraca et al., 2015), it could be very interesting to adapt these studies to the conditions of high albedo areas and evaluate the errors. Furthermore, exogenous stations without pyranometers but with records of other meteorological variables could be considered as remote sensors, estimating solar irradiation with a

previously trained model in the other stations with pyranometers, using the methodology by Antonanzas et al. (2015). Besides, kriging could be tested and compared in case of the availability of a sufficiently high grid of pyranometers.

2. Benchmarking of methods for the estimation of solar radiation in different regions. It would be interesting to benchmark different parametric models, soft-computing techniques, satellite-derived models and interpolation techniques in different regions. Most of the studies focus on techniques belonging to the same method (for instance, benchmarking of soft-computing techniques), but not to the range of errors that could be obtained using other methods. Given that satellite estimates can generate remarkable errors in some areas and also long gaps in the estimation, it could be useful to have a ranking of other methods and techniques as a backup for these situations. These studies could be divided as follows: desert regions, snow covered regions, tropical regions, coastal locations and islands and finally, continental areas. Furthermore, the distance between stations could be varied in order to have more scenarios.
3. Study of the impact of local aerosol sources on solar irradiation. There are several sources of atmospheric aerosols, such as dust, snow particles, smoke, and anthropogenic pollution, among others. However, as previously commented, satellite estimates normally rely on gridded estimates of aerosol content in low spatial resolution and therefore, these sources neglect most local events such as dust derived from agriculture, wildfires, and smog from cities, which clearly affect incoming solar irradiation. Although some studies have been done regarding the effect of these local aerosols on solar radiation (Perry and Troccoli, 2015; Khodakarami and Ghobadi, 2016), there is still a wide field for research. It would be very interesting to quantify the impact of some of these aerosols out of the potential radiation with a clean atmosphere, understanding clean as the available aerosol content previous to the studied event. This way, it would be possible to have a range of uncertainty in the estimation of solar irradiation in areas with frequent events of local aerosols.
4. Study of solar irradiation in areas with high atmospheric water vapor content. A similar study to the previous point for aerosols could be developed for water vapor. Tropical regions are known for the high water vapor content, which affects solar irradiation. Once again, satellite estimates rely on low spatial resolution estimates and on-ground records of water vapor are very rare. This introduces uncertainty in the estimation of clear-sky solar irradiation. In my opinion, a deeper research in the estimation and down-

scaling of water vapor could be performed with a consequent improvement in clear-sky solar irradiance estimates.

Bibliography

Almorox, J., C. Hontoria, and M. Benito (2011). Models for obtaining daily global solar radiation with measured air temperature data in Madrid (Spain). *Applied Energy 88*(5), 1703–1709.

Angstrom, A. K. (1924). Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Royal Meteorology Society 50*(210), 121–126.

Antonanzas, J., R. Urraca, F. Martinez-de Pison, and F. Antonanzas-Torres (2015). Solar irradiation mapping with exogenous data from support vector regression machines estimations. *Energy Conversion and Management 100*, 380–390.

Antonanzas-Torres, F., J. Antonanzas, R. Urraca, M. Alia-Martinez, and F. Martinez-de Pison (2016). Impact of atmospheric components on solar clear-sky models at different elevation: case study Canary Islands. *Energy Conversion and Management 109*, 122–129.

Antonanzas-Torres, F., F. Canizares, and O. Perpinan (2013). Comparative assessment of global irradiation from a satellite estimate model (CM SAF) and on-ground measurements (SIAR): a Spanish case study. *Renewable and Sustainable Energy Reviews 21*, 248–261.

Antonanzas-Torres, F., F. J. Martinez-de Pison, J. Antonanzas, and O. Perpinan (2014). Downscaling of global solar irradiation in complex areas in R. *Journal of Renewable and Sustainable Energy 6*, 063105.

Antonanzas-Torres, F., A. Sanz-Garcia, J. Antonanzas-Torres, O. Perpinan-Lamigueiro, and F. Martinez-de Pison-Ascacibar (2014). Current status and future trends of the evaluation of solar global irradiation using soft-computing-based models. In I. Global (Ed.), *Soft Computing Applications for Renewable Energy and Energy Efficiency*.

Antonanzas-Torres, F., A. Sanz-Garcia, F. Martinez-de Pison, J. Antonanzas, O. Perpinan-Lamigueiro, and J. Polo (2014). Towards downscaling of aerosol

gridded dataset for improving solar resource assessment. Application to Spain. *Renewable Energy 71*, 534–544.

Antonanzas-Torres, F., A. Sanz-Garcia, F. J. Martinez-de Pison, and O. Perpinan-Lamigueiro (2013). Evaluation and improvement of empirical models of global solar irradiation based on temperature and rainfall in northern Spain. *Renewable Energy 60*, 604–614.

Antonanzas-Torres, F., R. Urraca, J. Antonanzas, J. Fernandez-Ceniceros, and F. J. Martinez-de Pison (2015). Generation of daily global solar irradiation with support vector machines for regression. *Energy Conversion and Management 96*, 277–286.

Bristow, K. L. and G. S. Campbell (1984). On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology 31*(2), 159–166.

Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistics Association 74*(368), 829–836.

CMSAF (2015). Climate Monitoring Satellite Application Facility. http://dx.doi.org/10.5676/EUM_SAF_CM/RAD_MVIRI/V001.

Collares-Pereira, M. and A. Rabl (1979). The average distribution of solar radiation: correlations between diffuse and hemispherical and between daily and hourly insolation values. *Solar Energy 22*, 155–164.

Crawley, M. J. (2005). *Statistics: an introduction using R*. John Wiley.

DB (2015). Solar grid parity in a low oil price era. Technical report, Deutsche Bank. https://www.db.com/cr/en/docs/solar_report_full_length.pdf.

Gueymard, C. A. (2008). REST2: high-performance solar radiation model for cloudless sky irradiance, illuminance, and photosynthetically active radiation - validation with a benchmark dataset. *Solar Energy 82*, 272–285.

Hargreaves, G. H. (1981). Responding to tropical climates. Technical report, 1980-81 Food and Climate Review. The Food and Climate Forum. Aspen Institute for Humanistic Studies, Boulder, Colorado.

Hay, E. and D. C. McKay (1985). Estimating solar irradiance on inclined surfaces: A review and assessment of methodologies. *International Journal of Solar Energy 3*, 203–240.

IEA (2014). Key world energy statistics. Technical report, International Energy Agency. http://www.iea.org/publications/freepublications/publication/KeyWorld2014.pdf.

Ineichen, P. (2008). A broadband simplified version of the Solis clear sky model. *Solar Energy 82*(8), 758–762.

Jong, R. D. and D. W. Stewart (1993). Estimating global solar radiation from common meteorological observations in western Canada. *Can. J. Plant Sci 73*(2), 509–518.

Khodakarami, J. and P. Ghobadi (2016). Urban pollution and solar radiation impacts. *Renewable and Sustainable Energy Reviews 57*, 965–976.

Lawrence, S., C. L. Giles, and A. C. Tsoi (1996). What size neural network gives optimal generalization? Convergence properties of backpropagation. UMIACS-TR-96-22 and CS-TR-3617. MD 20742. Technical report, Institute for Advanced Computer Studies. University of Maryland College Park.

Louche, A., G. Notton, P. Poggi, and G. Simonnot (1991). Correlations for direct normal and global horizontal irradiation on a French Mediterranean site. *Solar Energy 46*, 261–266.

Meteonorm (2015). Handbook Part I: Software. Version 7.1. Technical report, Meteotest.

Mora, J. and L. Mora-Lopez (2010). Comparing distributions with bootstrap techniques: an application to global solar radiation. *Math Comput Simulat 81*(4), 811–819.

Moreno, A., M. Gilabert, and B. Martinez (2011). Mapping daily global solar irradiation over Spain: a comparative study of selected approaches. *Solar Energy 85*, 2072–2084.

Perez, R., P. Ineichen, E. Maxwell, R. Seals, and A. Zelenka (1992). Dynamic global-to-direct irradiance conversion models. *ASHRAE Trans 98*, 354–3698.

Perpinan, O. (2009). Statistical analysis of the performance and simulation of a two-axis tracking PV system. *Solar Energy 83*(11), 2074–2085.

Perry, M. and A. Troccoli (2015). Impact of a fire burn on solar irradiance and PV power. *Solar Energy 114*, 167–173.

Polo, J., F. Antonanzas-Torres, J. Vindel, and L. Ramirez (2014). Sensitivity of satellite-based methods for deriving solar radiation to different choice of aerosol input and models. *Renewable Energy 86*, 785–792.

Posselt, R., R. W. Mueller, R. Stockli, and J. Trentmann (2012). Remote sensing of solar surface radiation for climate monitoring? The CM SAF retrieval in international comparison. *Remote Sensing Environment 118*, 186–198.

REN21 (2015). Renewables 2015- global status report. Technical report, Renewable Energy Policy Network for the 21st Century. http://www.ren21.net/wp-content/uploads/2015/07/REN12-GSR2015_Onlinebook_low1.pdf.

Rigollier, C., O. Bauer, and L. Wald (2000). On the clear sky model of the ESRA - European Solar Radiation Atlas with respect to the heliosat method. *Solar*

Energy 68(1), 33–48.

Riihela, A., T. Carlund, J. Trentmann, R. Muller, and A. Lindfors (2015). Validation of CM SAF surface solar radiation datasets over Finland and Sweden. *Remote Sensing 7*, 6663–6682.

SolarGis (2015). SolarGis database version 1.8 satellite derived solar radiation and meteorological data. Report, GeoModel.

Suri, M., T. Huld, E. Dunlop, and J. Hofierka (2007). *Solar resource modelling for energy applications: digital terrain modelling*. Springer.

Tainter, A. J., T. Allen, and T. Hoekstra (2006). Energy transformations and post-normal science. *Energy 31*(1), 44–58.

Urraca, R., J. Antonanzas, F. Martinez-de Pison, and F. Antonanzas-Torres (2015). Estimation of solar global irradiation in remote areas. *Journal of Renewable and Sustainable Energy 7*, 023136.