



UNIVERSIDAD DE LAS PALMAS  
DE GRAN CANARIA

Departamento de Ingeniería Eléctrica

# Modelo Predictivo de Radiación Solar mediante técnicas de Machine Learning. Aplicación a la isla de Gran Canaria

Solar Radiation Forecasting Model using Machine Learning  
techniques. Application to Gran Canaria Island

## Tesis Doctoral

**Luis Mazorra Aguiar**

Las Palmas de Gran Canaria  
Octubre 2015





UNIVERSIDAD DE LAS PALMAS  
DE GRAN CANARIA

Departamento de Ingeniería Eléctrica  
Programa de doctorado: Tecnología Industrial

# Modelo Predictivo de Radiación Solar mediante técnicas de Machine Learning. Aplicación a la isla de Gran Canaria

Solar Radiation Forecasting Model using Machine Learning  
techniques. Application to Gran Canaria Island

## Tesis Doctoral

Autor

Director

Director

Luis Mazorra Aguiar

Felipe Díaz Reyes

Philippe Lauret

Las Palmas de Gran Canaria  
Octubre 2015



# Agradecimientos

*Aprovecho este momento para agradecer la ayuda y el apoyo recibido durante la elaboración de esta tesis.*

*En primer lugar me gustaría agradecer la dedicación y apoyo recibida de mis directores Felipe Díaz Reyes y Philippe Lauret durante la elaboración de este trabajo. En especial querría agradecer al Dr. Felipe Díaz la confianza depositada en mi para escribir esta tesis. Su ayuda ha sido constante desde la organización del trabajo hasta la búsqueda de financiación para poder realizar la estancia en la Universidad de la Reunión. No sólo ha sido un apoyo académico durante la investigación sino un amigo en quien confiar durante estos meses de duro trabajo. Mención especial también a la inestimable ayuda del Dr. Philippe Lauret, no sólo durante mi estancia en la Universidad de la Reunión sino siempre que la diferencia horaria nos lo permitiera. Durante mi estancia ofreció sin condiciones toda su experiencia y trabajo en este campo y guió mi investigación para conseguir este resultado final independientemente del tiempo y esfuerzo que conllevara. Además quisiera agradecerle su compañía durante mi estancia, su tiempo para enseñarme su maravillosa isla y que me recibiera con los brazos abiertos en su casa.*

*No quisiera olvidarme de todas las personas a las que conocí en el Laboratoire de Physique et Ingénierie Mathématique pour l’Energie et l’environnement (PIMENT), a Philippe Lauret, Mathieu David, Emeric Tapaches, Thierry Mara y Malik Mamode por compartir su tiempo y sus conocimientos conmigo y tener la paciencia de hablar en inglés. En especial agradecer a Mathieu David su ayuda con los datos del ECMWF y sus consejos y a Thierry Mara por su compañía durante mi estancia.*

*Por otra parte, quisiera agradecer su apoyo a mis compañeros del Departamento de Ingeniería Eléctrica y a los miembros del grupo de investigación del IUSIANI. Quisiera destacar sobre todo su generosidad al aceptarme en el equipo y sus consejos a la hora de elaborar esta tesis.*

*Quiero agradecer también la capacidad de trabajo, ayuda con los códigos y aportación de ideas a Brais Pereira y Raquel Pérez.*

No puedo olvidarme de los datos de radiación sin los que no podría haber realizado esta tesis. Quiero agradecer la ayuda al Instituto Tecnológico de Canarias por los datos de radiación de las estaciones de medida de las Islas Canarias, en especial a Antonio Ortegón por su atención y su trabajo sin condiciones. Agradecer también al Dr. Mathieu David y al PIMENT por los datos extraídos del modelo numérico ECMWF y al Dr. Philippe Blanc, Responsable des activités de recherche sur l'évaluation des ressources énergétiques renouvelables of MINES ParisTech / ARMINES por la base de datos satelitales.

Agradecer también el apoyo económico recibido de la Cátedra Endesa Red de la Universidad de Las Palmas de Gran Canaria en su Convocatoria del curso 2014-2015 de ayudas para realizar tesis doctorales y al Vicerrectorado de Internacionalización y Cooperación de la Universidad de Las Palmas de Gran Canaria, en el marco de la Convocatoria 2014/15 del Programa de apoyo a PFC, TFT y Tesis Doctorales de la ULPGC definidas en el ámbito de la Cooperación Internacional para el Desarrollo.

Por último pero no por ello menos importante agradecer el cariño de mis hijos Irene y Pablo. De manera muy especial quiero agradecer a Elena por su paciencia, su apoyo sin condiciones a mi trabajo, su ayuda infinita durante estos meses en los que he estado ausente y su compañía. También agradecer la ayuda y apoyo recibida de mis padres para llegar hasta aquí.

Esta tesis ha sido desarrollada en el marco del siguiente proyecto subvencionado:

- *Integración de nuevas metodologías en simulación de campos de vientos, radiación solar y calidad del aire.* Subvencionado por la Convocatoria 2014 – Proyectos I+D+I – Programa Estatal de Investigación, Desarrollo e Innovación orientada a los retos de la sociedad. Referencia: CTM2014-55014-C3-1-R

# Índice general

<b>Abstract</b>	<b>1</b>
<b>1. Introducción</b>	<b>3</b>
<b>2. Fundamentos y datos de radiación solar</b>	<b>15</b>
2.1. Fundamentos de radiación solar . . . . .	15
2.1.1. Naturaleza de la radiación solar . . . . .	15
2.1.2. Principios del movimiento solar . . . . .	17
2.1.3. Radiación exoatmosférica sobre superficie horizontal . . . . .	23
2.2. Modelo de Cielo despejado . . . . .	24
2.3. Datos terrestres de radiación . . . . .	27
2.3.1. Estaciones de medida en la isla de Gran Canaria . . . . .	28
2.3.2. Equipos de medida . . . . .	32
2.3.3. Tratamiento de los datos de radiación solar . . . . .	33
2.3.4. Análisis de los datos de radiación solar . . . . .	35
2.4. Adquisición de datos del ECMWF . . . . .	43
2.4.1. Datos de radiación solar del ECMWF . . . . .	44
2.4.2. Datos de viento y humedad del ECMWF . . . . .	46
<b>3. Radiación solar satelital</b>	<b>49</b>
3.1. Datos obtenidos del modelo de radiación Heliosat . . . . .	50
3.2. Análisis de los datos de radiación satelitales . . . . .	52
3.3. Análisis espacio temporal de los datos . . . . .	54
3.3.1. Correlación con el índice de cielo despejado . . . . .	56
3.3.2. Correlación con la variación del índice de cielo despejado . . . . .	63

<b>4. Modelos de predicción de radiación solar</b>	<b>67</b>
4.1. Introducción . . . . .	67
4.2. Modelos estadísticos de referencia . . . . .	68
4.2.1. Modelo de predicción Persistente . . . . .	68
4.2.2. Modelo de predicción Smart-Persistence . . . . .	69
4.2.3. Modelo de predicción Climatológico . . . . .	69
4.3. Modelos lineales . . . . .	70
4.3.1. Modelo lineal Autorregresivo AR . . . . .	71
4.3.2. Modelo lineal autorregresivo de medias móviles ARMA . . . . .	71
4.3.3. Estudio de la complejidad de los modelos lineales . . . . .	72
4.4. Redes Neuronales Artificiales . . . . .	75
4.4.1. Introducción a las redes neuronales . . . . .	76
4.4.2. Aplicaciones . . . . .	78
4.4.3. Fundamentos teóricos de las redes neuronales . . . . .	79
4.4.4. La neurona simple . . . . .	83
4.4.5. El Perceptrón Multicapa . . . . .	85
4.4.6. Regla de aprendizaje. Backpropagation . . . . .	89
4.4.7. Algoritmo de gradiente conjugado escalado . . . . .	94
4.4.8. Técnicas de optimización de la arquitectura de la red . . . . .	95
4.5. Redes Neuronales Bayesianas . . . . .	96
4.5.1. Aproximación probabilista del aprendizaje en RNA . . . . .	97
4.5.2. Optimización bayesiana de los parámetros de control . . . . .	99
4.5.3. Selección bayesiana de la arquitectura de una RNA . . . . .	100
4.5.4. Técnica bayesiana para la selección automática de entradas relevantes	101
<b>5. Aplicación de los modelos de predicción</b>	<b>105</b>
5.1. Introducción . . . . .	105
5.2. Predicción a partir de datos terrestres y del ECMWF . . . . .	107
5.2.1. Modelos lineales de predicción . . . . .	108
5.2.1.1. Estudio de la complejidad de los modelos lineales . . . . .	108
5.2.1.2. Resultados obtenidos con los modelos lineales . . . . .	115
5.2.2. Modelos de predicción basados en Redes Neuronales Artificiales . . .	120
5.2.2.1. Estudio de la complejidad de las RNAs . . . . .	122
5.2.3. Resultados obtenidos y comparación entre los modelos . . . . .	126

5.2.3.1. Resultados anuales y trimestrales . . . . .	130
5.2.3.2. Resultados según el tipo de día . . . . .	140
5.3. Predicción a partir de datos satelitales y otras variables . . . . .	153
5.3.1. Selección de los datos de radiación Satélite . . . . .	154
5.3.2. Estudio de la predicción utilizando otros datos meteorológicos del ECMWF . . . . .	156
5.3.3. Resultados obtenidos y comparación entre los modelos . . . . .	158
5.3.3.1. Resultados anuales y trimestrales . . . . .	159
5.3.3.2. Resultados según el tipo de día . . . . .	168
5.4. Producción de energía eléctrica . . . . .	186
<b>6. Conclusiones y líneas futuras</b>	<b>189</b>
<b>Conclusions</b>	<b>195</b>
<b>Listado de Figuras</b>	<b>199</b>
<b>Listado de Tablas</b>	<b>207</b>
<b>Bibliografía</b>	<b>211</b>



# Abstract

Solar forecasting has become an important issue for power systems planning and operating, especially in islands grids. Power generation and grid utilities need day ahead, intra-day and intra-hour Global Horizontal solar Irradiance forecasts for operations. A model based in geometric considerations to estimate solar radiation for every places in Gran Canaria was developed by F. Díaz. The model used solar radiation data from ground measurement stations and from numerical weather prediction model (MM5).

To improve Solar Irradiance forecasting used in geometrical model developed by F. Diaz was established as the aim of this Thesis to continue the research line. The work is focused on intra-day solar forecasting with forecast horizons ranging from 1 hour to 6 hours ahead with hourly granularity. Depending on the forecast horizon and granularity, forecasting results are used for operational planning, switching sources or re-scheduling of means of production, programming backup, planning for reserve usage, and peak load matching. Different input data are used for forecasting methods according to time horizon. Statistical models, such as autoregressive (AR), autoregressive moving average (ARMA) and artificial neural networks (ANN), are used for short-time forecasts from 5 min. to 6 h. Different statistical models are proposed to forecast Global Horizontal solar Irradiance using ground measurement data, numerical weather prediction model data (from ECMWF) and satellite data (from Helioclim-3) as inputs.

During this Thesis the following issues have been worked out:

- A clear sky model was computed for each site. Clear sky models compute global solar horizontal irradiance expected for a clear sky condition at any location and time. The Bird model (Bird & Hulstrom, 1981), used in this survey, is well known to provide accurate results with only a few meteorological inputs (Badescu, et al., 2013), Aerosol Optical Depths (AOD500 mm and AOD380 mm), water vapour and Ozone atmospheric content.
- Ground hourly datasets used in this thesis were obtained from six stations in Gran Canaria Island. An essential step in any solar radiation survey is a quality assessment of the data. All measurements series have been treated with SERI-QC control software for only global solar radiation. In this case, negative data and data over top of atmosphere (ToA) were filtered out. The hourly GHI average was computed for hours with more than 50 % of data available. On the other hand, in global solar radiation forecasting, it is a general practice to remove night hours from the data before building and testing the forecasting models.

- In order to characterize each site, ground measurement data were classified using mean intraday clear sky index and intraday variability for each day of the studied year. According to both parameters we could divide days of the year in different types, see Fig. (2). Intraday mean is divided in A, B and C days, where A type days represent heavily cloudy days and C type stand for much clearer days. Numbers stands for the stability and so a type III day has a great variability and type I is a steady one.
- HelioClim-3 images provide an estimation of the global solar irradiation at ground level for each pixel from the spaceborne Meteosat Second Generation satellite SEVIRI sensor. The main reason for using satellite data is its great spatial information and the possibility to see the evolution of the surroundings of the desired location. To assess the satellite raw data, an interface was created to visualize a particular behaviour of a known meteorological pattern, such as static cloud cover on summer time over the city of Las Palmas, on the northeast of the island.
- In this Thesis we propose improving solar radiation forecasting accuracy based on ANN with only ground data, introducing information from vicinity locations based on satellite-derived data. The selection of the number of satellite inputs is an important decision in the forecasting performance. We studied the importance of the satellite-gridded data using a time-lagged correlation analysis between the satellite-derived data on big grid around the island and ground data for the different sites. Therefore we can deduce the best closest reactions in the area, which will account as additional information that can be selected in order to improve further predictions.
- Different lineal forecasting models, AR and ARMA, were studied in this thesis. The complexity of both models was estimated using temporal series of clear sky index.
- The accuracy of ANNs to approach continuous functions depends on the network structure. In this paper, Bayesian regularization framework was used to control model complexity. In our case, we trained different ANNs using several number of hidden units and inputs, both ground and exogenous data, and Bayesian framework provides us the best one.
- During this work wind speed and relative humidity were used as inputs in the ANN. ECMWF provided us both data for different atmospheric altitude. Therefore, the optimal altitude for each ground station was estimated comparing forecasting results obtained with different input data.
- Application of different statistical models for solar radiation forecasting in several ground stations in Gran Canaria Island. Models were proved using solar radiation ground data, satellite data and ECMWF data.

# Capítulo 1

## Introducción

La radiación electromagnética que emite el Sol es la responsable de muchos de los procesos naturales que permiten la vida en la Tierra. La energía proporcionada por el Sol es la fuente de la que provienen directa o indirectamente muchas de las formas de energía utilizadas por el ser humano. Desde los combustibles fósiles, producto de la transformación de la vida en la Tierra, hasta la energía eólica, son el fruto de la energía recibida desde el Sol. Esta energía obtenida de la radiación solar ha sido utilizada de manera directa desde la antigüedad. En estos casos los ejemplos de aprovechamiento se basaban en la energía solar térmica.

Es en 1839 cuando el científico francés Edmond Becquerel [Becquerel39] observa cómo un material produce energía eléctrica al ser expuesto a la luz solar. Se describe por primera vez el efecto fotovoltaico. En 1905, Albert Einstein [Einstein05] estudió este fenómeno y publicó un artículo sobre el efecto fotoeléctrico por el que se le concedió el Premio Nobel en 1921. En este trabajo describe el proceso por el cual se pueden liberar electrones en un material metálico al ser expuesto a la luz solar. Las primeras pruebas experimentales obtenidas con una célula fueron realizadas por Robert Milikan en 1916. Sin embargo, no es hasta 1954 cuando Daryl Chapin, Calvin Fuller y Gerald Pearson [Chapin54] desarrollan las primeras células fotovoltaicas de Silicio. La carrera espacial en los años 60, cuando se empezó a demostrar como una fuente de energía muy útil en los satélites, y la crisis energética de los años 70 favorecieron el desarrollo de la energía solar como alternativa.

Desde entonces las instalaciones de producción de energía eléctrica a partir de energía solar se han ido desarrollando y aumentando. Ya en los años 80 se instalaron en California (EEUU) un grupo de nueve centrales termoeléctricas de captadores cilindro parabólicos llamadas Solar Energy Generating Systems, con una potencia total instalada de 354 MW entre todas. Mientras que en la actualidad nos podemos encontrar por ejemplo, con la central de energía solar termoeléctrica Ivanpah Solar Electric Generating System conformada por 300.000 heliostatos que enfocan la luz del sol a la parte superior de las torres. El Sistema de Generación de Energía Solar Ivanpah está instalado en California y, en su capacidad máxima, el trío de torres produce un total de 392 megavatios (MW) de energía solar. En España, por su parte, destacan las instalaciones de Solnova Solar Power Station en Sevilla con 150 MW, la de Andasol en Granada también con 150 MW y las centrales de Extresol en Extremadura con 100 MW.

En 1982, se instala en California la primera central fotovoltaica con un potencia instalada superior a 1 MW. El desarrollo de esta tecnología ha logrado incrementar de manera considerable la potencias de estas centrales. Así, en la actualidad existen instalaciones como la de Solar Star I, II y III ubicada en las proximidades de Rosamond (California), una central fotovoltaica de 579 MW y finalizada en junio de 2015, la planta fotovoltaica Desert Sunlight en el Desierto de Mojave (California), con una capacidad instalada de 550 megavatios, la central de Charanka Solar Park en India, con una potencia instalada de 214 MW se encuentra dentro del complejo de Gujarat Solar Park que pretende conseguir 605 MW ol central de Longyangxia Hydro- Solar PV Station en la provincia china de Qinghai, con 320 MW instalados.

El incremento de la instalación de fuentes de energía eléctrica renovables han conducido a un mayor interés por estudiar los problemas técnico económicos derivados de su inyección en la red eléctrica. Las fuentes de energía renovables son, por definición, inestables por lo que las compañías generadoras y los operadores del sistema deben lidiar con estas fluctuaciones. Este nuevo reto no es necesario con las centrales convencionales, que se ajustan a la demanda en todo momento. Por lo tanto, una predicción fiable de las condiciones de radiación solar pueden conducir a un aprovechamiento más eficiente de la energía producida en las centrales solares. La energía proveniente de fuentes convencionales es fácilmente regulable, por lo que disponer de los patrones de carga y producción de energía solar con adelanto permitirá a los operadores un mejor balance entre la demanda y la generación de potencia [Heinemann06a, Wittmann08].

En los sistemas insulares, como es el caso de la Isla de Gran Canaria, estos problemas se agudizan al no encontrarse conectados con la red general eléctrica del continente ni entre las islas. En estos sistemas eléctricos la energía se debe producir y consumir en el propio territorio. Un alto grado de penetración de fuentes de energía fluctuantes puede conducir a comportamientos inestables en las redes eléctricas. Además, la radiación solar presenta una alta variabilidad espacio temporal en las Islas Canarias, por lo que una predicción fiable es una herramienta sumamente importante [Diagne14, Lauret15]. Esta variación está muy influenciada tanto por la climatología de la zona como por la escarpada orografía, que hace muy complicado establecer relaciones fiables en los niveles de radiación solar de las diferentes zonas de la isla [DR13, Badosa15].

Dependiendo del horizonte temporal de predicción de la radiación solar, los resultados se utilizarán para la planificación operacional, alternar las fuentes, reprogramar los medios de producción, programar las reservas, planificar el uso de las mismas o atender o prever los picos de demanda. Kostylex & Pavlovski [Kostylev11] identifican una serie de horizontes de predicción a corto plazo:

- **Intrahorario**, desde 15 min. hasta las 2 horas con intervalos que van desde 30 segundos hasta 5 minutos. Estas predicciones se suelen relacionar con eventos concretos y operaciones relacionadas con la variabilidad del sistema.
- **Horaria**, desde una 1 a 6 horas en adelanto con intervalos de una 1 h. Estas predicciones se suelen utilizar para realizar un balance equilibrado entre la producción y la demanda de energía eléctrica.

- **Diario**, hasta 3 días en adelante con intervalos horarios de predicción. Estas predicciones están relacionadas con la programación de las centrales y operaciones de planificación en el mercado de la electricidad.

La investigación en el campo de la predicción de radiación solar es un campo muy amplio y con una actividad muy alta en los últimos años. Los métodos más importantes de predicción utilizados en la bibliografía son en general:

- **Métodos estadísticos**, se basan en la predicción utilizando series temporales de datos de radiación de la zona en cuestión. Para trabajar con estos métodos se debe tener un número suficiente de datos históricos. Entre los métodos estadísticos más importantes nos encontramos con modelos lineales, como los modelos Autorregresivos (AR) y Autorregresivos de Medias Móviles (ARMA), y modelos de *Machine Learning* como las Redes Neuronales Artificiales (RNAs).
- **Modelos de predicción basados en imágenes hemisféricas del cielo**, se trata de imágenes del cielo tomadas con cámaras con un ángulo de visión de 180° situadas en el suelo. Estas imágenes permiten obtener un conocimiento de la nubosidad de la zona con una resolución temporal de segundos. Los equipos implicados en este tipo de predicciones se han desarrollado de forma importante en los últimos años.
- **Modelos basados en imágenes satelitales**, en la actualidad los satélites meteorológicos geoestacionarios disponen de imágenes de la atmósfera terrestre alrededor de la Tierra con resoluciones temporales menores a una hora. Este hecho los convierte en una herramienta muy útil para realizar predicciones de radiación solar con horizontes temporales de hasta horas de adelanto.
- **Modelos numéricos de predicción meteorológica (NWP)**, se emplean para la predicción del estado de la atmósfera desde horas hasta 15 días de antelación. La predicción de los cambios en la atmósfera, incluyendo la formación y disolución de la nube, se basan en modelos físicos. Estos modelos físicos se describen mediante ecuaciones diferenciales básicas que se resuelven utilizando de métodos numéricos.

Los datos de entrada utilizados en cada modelo de predicción serán diferentes según el horizonte temporal deseado [Sengupta15]. Para predicciones intrahorarias, que requieren gran precisión espacial y temporal, los modelos estadísticos obtienen buenos resultados con una resolución temporal desde unos pocos minutos. Los métodos estadísticos utilizan datos de irradiancia medida en las estaciones terrestres. La alta correlación observada en las series temporales de datos de irradiancia hace de estas herramientas una buena solución. Sin embargo son incapaces de predecir los cambios en las condiciones de nubosidad. Los modelos de predicción basados en la información de la nubosidad y su movimiento obtenida con imágenes hemisféricas del cielo muestran buenos resultados para horizontes temporales menores a 1 h [Lorenz12].

Por otro lado, para obtener información sobre la radiación en un rango de predicción horario, los modelos basados en el movimiento de la nube (*cloud motion vectors*) a partir de imágenes provenientes de satélites geoestacionarios han sido ampliamente utilizados

con resultados precisos [Hammer99, Perez10]. Estas imágenes satelitales se encuentran disponibles con resoluciones espaciales entre 1 km y 5 km y temporales entre 15 min y 30 min para los actuales satélites. Los modelos numéricos NWP se utilizan para realizar predicciones desde las 6 horas hasta los días de horizonte temporal.

Los modelos estadísticos de predicción se muestran en general eficaces para todos los horizontes temporales de predicción. Así, se obtienen buenos resultados a muy corto plazo utilizando modelos para series temporales basadas únicamente en datos terrestres medidos en la localidad en cuestión. También se obtienen predicciones fiables de para horizontes diarios de predicción utilizando métodos estadísticos para refinar los resultados previos de un modelo numérico de predicción [Diagne13]. Existe un gran potencial de mejora en las predicciones realizando simulaciones con la combinación de datos diferentes en conjunto con diversos modelos estadísticos basados en el aprendizaje.

El objetivo de esta tesis es el estudio de distintos métodos de predicción de la irradiancia solar global horizontal en el horizonte temporal horario, de 1 a 6 horas en adelante con intervalos horarios. Como ya se ha comentado, según el horizonte temporal de predicción existen diversas técnicas en constante desarrollo. En general, los modelos de predicción estadísticos, como los métodos Autorregresivos (AR), los métodos Autorregresivos de Media Móviles (ARMA) y las Redes Neuronales Artificiales (RNAs), son los más apropiados para horizontes de predicción entre 5 min. y 6 h [Lorenz12]. Durante esta tesis se trabajará con estos modelos y datos de entrada provenientes de estaciones de medida terrestres, de imágenes satelitales y del modelo numérico ECMWF.

## Estado del arte

En el presente trabajo se analizarán principalmente diferentes modelos estadísticos de predicción de radiación solar, así como métodos para la mejora de la predicción utilizando datos provenientes de otros modelos de predicción. En el primer caso nos centraremos en los trabajos realizados con modelos lineales y RNAs en la predicción de radiación solar. Los datos exógenos utilizados serán los obtenidos por medio de métodos numéricos de predicción y datos de radiación basados en imágenes satelitales.

Los modelos lineales de predicción (AR y ARMA) han sido ampliamente utilizados desde los años setenta en el campo de la radiación solar. Los modelos de series lineales se han utilizado para describir su comportamiento, para generar series sintéticas de valores o para definir Años Meteorológicos Típicos [Mazorra10], *Typical Meteorological Year (TMY)* en inglés, y para realizar predicciones de radiación solar, a escala tanto horaria como diaria. J. Boland, [Boland95, Boland08], realizó una descripción de la radiación solar diaria y horaria, un estudio del modelado de series temporales de radiación solar y un método de estimación de la radiación solar difusa en Australia a partir de modelos Autorregresivos (AR), de medias móviles (MA) y Autorregresivos de medias móviles (ARMA). Estos trabajos se basan en la Metodología de Box-Jenkins, [Box98], que describe un proceso iterativo para identificar el modelo óptimo y luego utilizarlo en las predicciones. G. Reikard [Reikard09] se basó en modelos Autoregresivos Integrados de medias móviles (ARIMA) para seis estaciones en los Estados Unidos. P. Bacher [Bacher09] por su parte, evaluó para estaciones en Dinamarca el comportamiento de modelos Autoregresivos (AR) simples,

con datos pasados de la serie temporal y añadiendo datos exógenos (ARX) provenientes de modelos numéricos de predicción meteorológica, (NWP) por sus siglas en inglés. S. SAFI, [Safi02], estudió distintos modelos de medias móviles (MA) para localidades en Marruecos. Aguiar & Collares-Pereira realizaron trabajos con modelos ARMA a partir de datos del índice de claridad horario [CP89]. En los últimos años nos encontramos con numerosas metodologías basadas en modelos estadísticos autorregresivos (AR) a partir de datos de radiación solar, tanto de estaciones de medida como provenientes de satélites [Dambreville14b, Dambreville14a, Zagouras15]. Por otro lado, en [Lauret12] se propone un método híbrido entre redes neuronales y ARMA basados en técnicas de decisión Bayesianas, mientras que en [David14] se propone el uso de un modelo ARMA recurrente para la predicción de radiación solar horaria en una isla de clima Tropical, Isla Reunión.

En esta tesis las Redes Neuronales Artificiales utilizadas para predecir la radiación solar es el Perceptrón Multicapa. Desde que en 1957 Frank Rosenblatt [Rosenblatt58] desarrolló este tipo de redes y en 1986 Rumelhart, Hinton y Williams presentan su trabajo en el que se desarrolla un algoritmo de aprendizaje conocido como *retropropagación (backpropagation)* [Rumelhart86] para redes neuronales multicapa, el número de trabajos sobre RNA se han multiplicado. Se han desarrollado un gran número de aportaciones en los métodos de aprendizaje y tipo de estructuras. Las RNAs es una herramienta capaz de reproducir una relación no lineal entre un conjunto de datos de entrada y otro de salida [Bishop95], lo cual lo hace una herramienta muy atractiva. Las RNAs se han utilizado con éxito para realizar predicciones con las series temporales de radiación solar en forma del índice de cielo despejado. Estos modelos estadísticos pueden trabajar únicamente basándose en datos históricos de radiación solar. Se han realizado trabajos de predicción con RNAs para datos de irradiancia solar desde horizontes temporales horarios, como en los trabajos de [Hontoria02, Lauret06b, Mellit10, Inman13], hasta para predecir la irradiancia solar con 24 horas de antelación [Bosch08, Rehman08].

Los modelos estadísticos basados en RNAs permiten añadir otro tipo de variables como datos de entrada. Así, se puede trabajar con RNAs combinando datos de radiación históricos y otras variables meteorológicas. De esta manera podemos encontrar a Rehman [Rehman08], utilizando datos de temperatura y humedad relativa terrestre para predecir la irradiancia diaria. Kemmoku [Kemmoku99] predice la irradiancia diaria con una aplicación en serie de varias RNAs partiendo de datos de presión atmosférica y otros datos meteorológicos. Sfetsos & Coonick [Sfetsos00] introducen datos de temperatura, velocidad del viento y presión, además de los datos de radiación solar, para predecir valores de irradiancia global horaria. De la misma manera, se pueden encontrar diversos trabajos de predicción con diferentes combinaciones de datos meteorológicos como la longitud del día (horas de sol), temperatura media, humedad relativa, latitud y longitud para obtener datos tanto de irradiancia horaria como de irradiancia diaria [Mohandes98, Ghanbarzadeh09, Mellit10].

Además de las RNAs existen otros tipos de modelos estadísticos no lineales ampliamente utilizados en la predicción de la radiación solar. Estos modelos también se basan en técnicas de aprendizaje por lo que requieren de una base de datos históricos. En esta tesis se utilizarán las RNAs basadas en las técnicas probabilistas bayesianas según los trabajos de [MacKay03]. Estas técnicas permiten entre otros aspectos mejorar el proceso de aprendizaje, además de estudiar la complejidad del modelo a definir [Lauret08, BS10].

En los últimos años se han desarrollado otras técnicas de Machine Learning con las que se están obteniendo buenos resultados de predicción. De esta manera encontramos trabajos de predicción basados en Support Vector Machine [Zeng13, FJ13, Wolff13] en los que se exponen los resultados obtenidos con esta técnica, y trabajos basados en los Gaussian Process [Sun14, Lauret15]. Como variante a las RNAs convencionales se han desarrollado en los últimos años redes neuronales con funciones de activación basadas en la función de ondas (wavelet function) [Mellit10, Cao08]. Estas redes se conocen como Wavelet Neural Networks (WNNs) o Wavelet Networks (WNs).

Los modelos numéricos de predicción son adecuados para predecir distintas variables atmosféricas hasta un horizonte temporal de unos 15 días. Las variaciones en el estado de la atmósfera se modelan en base a unas ecuaciones diferenciales que describen los procesos físicos. Los modelos NWP globales se encuentran actualmente operados por 15 diferentes agencias meteorológicas mundiales. Como ejemplo tenemos el Global Forecaste System (GFS) utilizado por la US National Oceanic and Atmospheric Administration (NOAA) y el Integrated Forecast System (IFS) operado por la European Centre for Medium-Range Weather Forecasts (ECMWF). Por otro lado los modelos de mesoescala se encuentran disponibles únicamente para algunas zonas del globo terrestre pero ofrecen una resolución espacial mayor que los globales. Entre estos modelos encontramos el MM5 desarrollado por la Pennsylvania State University y el National Centre for Atmospheric Research (NCAR) o el modelo WRF diseñado como un modelo. En los últimos años se han realizado estudios de comparación con las predicciones realizadas por estos modelos locales [Heinemann06b, Perez11].

La precisión de estos modelos de predicción varía según la escala temporal utilizada y zona geográfica en la que se trabaje. Heinemann [Heinemann06b] muestra que se pueden obtener datos de radiación para cielos despejados sin prácticamente ninguna desviación. Una comparación de los resultados obtenidos por estos modelos para estaciones de Estados Unidos, Canadá y Europa se describen en [Perez07, Perez10, Perez13] mostrando errores en la predicción de radiación horaria de 38 % rRMSE. En Europa, se han obtenido resultados del orden del 40 % de error para estaciones del centro de Europa y del 30 % de error para estaciones en España. Además han sido analizados los resultados con respecto a diferentes propiedades relevantes para su aplicación en la producción fotovoltaica [Lorenz09b, Lorenz09a, Lorenz11].

Las series temporales de radiación solar tienen una parte determinista diaria y anual, pero también existe una componente aleatoria debido entre otros aspectos a la presencia de nubes en el cielo. Así, determinar la nubosidad de un determinado lugar podrá ofrecer una información valiosa para predecir la radiación. Para horizontes de predicción superiores a las horas, el cambio de la nubosidad está fuertemente influenciado por el movimiento de las nubes. Las imágenes satelitales y las imágenes hemisféricas del cielo proporcionan la posibilidad de predecir la presencia de nubes, extrapolando el movimiento de las mismas en las horas anteriores al horizonte temporal deseado.

Los modelos de predicción basados en imágenes satelitales e imágenes hemisféricas del cielo detectan el movimiento de las nubes utilizando técnicas de seguimiento de los vectores del movimiento *Cloud Motion Vectors* [Lorenz12]. Las predicciones a corto plazo basadas en imágenes del cielo es un campo relativamente nuevo con diversos trabajos para horizontes temporales intrahorarios principalmente [Chow11, Urquhart13].

Las predicciones basadas en los vectores de movimiento de las nubes obtenidas a partir de imágenes satelitales logran unos resultados del orden del 17%rRMSE para predecir el índice de nubosidad en horizontes temporales de 30 minutos y un 30%rRMSE hasta 2 horas [Hammer99]. Por otro lado en [Lorenz09a] se realizó una comparación de varios métodos basados en los vectores de movimiento de las nubes a partir de las imágenes del Meteosat para predecir irradiancia solar con horizontes temporales superiores a una hora. Mientras que [Perez10] muestra los resultados de las predicciones de irradiancia basados en las imágenes del Geostationary Operational Environmental Satellite (GOES).

Los datos satelitales utilizados en esta tesis se obtuvieron de la base de datos del **Helioclim-3**, en particular, de la **versión 5 (HC3v5)**. Los datos de radiación se estiman a partir de las imágenes satelitales del Meteosat con el método del Heliosat-2 [Rigollier04, Blanc11b]. Esta versión le proporcionó al Helioclim una mejor resolución temporal (15 minutos) y espacial (3 km nadir). Las imágenes se obtienen en tiempo real en la estación de recepción del METEOSAT y se realizan los cálculos cada 15 minutos. La base de datos HelioClim estima la radiación a cielo despejado con el modelo de McClear, que usa los datos de AOD, Ozono y vapor de agua del proyecto MACC [Lefevre13]. Con las últimas actualizaciones se ha mejorado el error obtenido por el modelo Heliosat-2 tanto para días despejados como para días nubosos [Eissa15] y se ofrece, además de la estimación de la radiación ocurrida, una predicción para las próximas horas [Thomas15].

En los últimos años han surgido diversos estudios basados en la utilización conjunta de datos históricos de radiación medidos en estaciones terrestres, datos obtenidos por algún modelo numérico y datos de radiación obtenidos a partir de datos satelitales. Estos datos se utilizan como variables de entrada en diversos modelos estadísticos de predicción. Así, con los datos provenientes de imágenes satelitales nos encontramos con modelos autorregresivos (AR) [Dambreville14b, Dambreville14a, Zagouras15], redes neuronales artificiales [Marquez13, MA15] y algoritmos genéticos para elegir la información relevante de toda la base de datos del satélite [Zagouras15]. La elección de los píxeles de interés para mejorar la predicción de los datos terrestres es uno de los puntos más importantes del trabajo. En este caso nos encontramos con [Dambreville14a] que propone utilizar las correlaciones entre las series temporales de la variación del índice de cielo despejado cada 15 minutos, tanto de los datos terrestres como satelitales. Por otra parte [Zagouras15, MA15] proponen utilizar las correlaciones entre las series temporales del índice de cielo despejado horarias. En ambos casos las correlaciones se realizan con los datos satelitales desfasados con respecto a los terrestres. Por otro lado, para tiempos de predicción superiores a 6 h los métodos numéricos de predicción (NWP) post-procesados con un modelo estadístico y datos terrestres de radiación muestran los mejores resultados [Diagne13].

## Objetivos y metodología

Con los antecedentes descritos, el conocimiento de la radiación se revela fundamental para diversas actividades relacionadas con la producción y la gestión de la energía eléctrica, sobre todo en regiones insulares. Por lo tanto, contar con modelos de predicción de radiación solar fiables para ser utilizados en diversos campos resulta de clara aplicación. F. Díaz [DR13] desarrolla un modelo basado en las consideraciones geométricas de la isla

de Gran Canaria para establecer los niveles de radiación solar en cada punto. Los datos de radiación solar utilizados se basaban en datos históricos y datos obtenidos a partir de un modelo numérico de predicción meteorológica (MM5). En esta tesis, para continuar con la línea de investigación, se ha decidido mejorar las predicciones de radiación utilizadas por el modelo geométrico descrito por F. Díaz [DR13], además de conseguir un horizonte temporal de predicción horario.

Para cumplir este objetivo se han trabajado los siguientes aspectos:

- Un modelo de radiación a cielo despejado que contemple las condiciones físicas de atmósfera y la geometría solar.
- Recopilación de los datos de radiación terrestres de las estaciones de medida en la isla de Gran Canaria y tratamiento de los mismos. Se ha realizado un filtrado de los datos según el modelo SERI QC y un ángulo cenital concreto. Las medias horarias se han calculado a partir de los datos recogidos cada minuto únicamente en aquellas horas en las que se disponía del 50% de los datos.
- Análisis de los datos de radiación solar realizando una distribución de los tipos de días en función de la media diaria del índice de cielo despejado y la desviación estándar de la variación del mismo.
- Estudio de la cuadrícula de datos de radiación solar obtenidos alrededor de la isla de Gran Canaria a partir de imágenes satelitales, en concreto del HelioClim-3. Se ha desarrollado una herramienta que permite visualizar los datos de irradiancia solar y el índice de cielo despejado cada 15 minutos durante todo el año de medida estudiado. De esta manera se permite visualizar las variaciones intradiarias de la radiación en la geografía insular.
- Cálculo de las correlaciones entre los índices de cielo despejado de los datos terrestres de las estaciones de medida y cada uno de los píxeles de la cuadrícula de datos satelitales obtenida. Las correlaciones se han calculado con los datos satelitales desfasados con respecto a los datos terrestres para poder estudiar la relación entre las estaciones de la isla y su alrededor en distintos instantes.
- Desarrollo de una herramienta que permite visualizar los resultados de estas correlaciones para cada estación y para cada desfase temporal entre ambas series. Estos resultados se pueden observar para todo el conjunto de datos anuales y según la estación del año elegida (invierno, primavera, verano y otoño). La correlación nos permite calcular la relación entre cada píxel y la estación de medida, con lo que nos dará una pista de los datos más significativos para obtener mejores resultados de predicción.
- Estudio de los modelos lineales de predicción AR y ARMA. Se calcula la complejidad de ambos modelos a partir las series temporales del índice de cielo despejado y se decide el modelo lineal óptimo para realizar las predicciones de radiación solar.
- Estudio del modelo óptimo de Redes Neuronales Artificiales para realizar las predicciones de radiación solar a partir de datos de radiación terrestres y datos

exógenos. Se decide el número de neuronas ocultas de la capa intermedia y el número de entradas relevantes para cada caso.

- Elección de los píxeles de la cuadrícula de datos satelitales que consiguen mejores resultados de predicción en conjunto con los datos terrestres como entradas de las RNAs.
- Elección de la altitud atmosférica a la que extraer los datos de velocidad del viento y humedad relativa del modelo numérico de predicción. Se estudian diferentes altitudes y se comparan los resultados de predicción de radiación solar obtenidos.
- Aplicación de los diferentes modelos estadísticos para la predicción de la radiación en las diferentes estaciones de la isla de Gran Canaria. Los modelos se han estudiado utilizando datos terrestres de radiación solar y datos exógenos como los datos satelitales o los datos del modelo numérico de predicción elegido.

Para realizar el tratamiento de datos de radiación solar se ha partido del cálculo de modelo de cielo despejado siguiendo los trabajos de Bird & Hulstrom [Bird81] utilizando los valores de aerosoles AOD500 nm, AOD380 nm y el contenido en vapor de agua de la columna vertical obtenidos de la red AERONET [AERONET14, Holben98]. Una vez calculado el índice de cielo despejado se ha realizado un filtrado de los datos disponibles en cada estación siguiendo el método SERI-QC [Maxwell93, Younes05] para bases de datos de radiación global horizontal únicamente. Por otro lado, para trabajar en los modelos estadísticos de predicción con series temporales continuas se filtran las horas nocturnas. Este filtro se basa en el ángulo cenital de cada medida. El criterio seguido en esta tesis establece la frontera de datos válidos en  $80^\circ$ , a partir del cual se considera que los datos son horas nocturnas.

El análisis de la climatología de la isla a partir de los datos horarios de radiación solar de las seis estaciones de medida se realiza a partir de la caracterización del comportamiento meteorológico de cada estación. Para caracterizar cada estación, la base de datos de radiación solar horaria se ha clasificado según una distribución de los tipos de días en función de la media diaria del índice de cielo despejado y la desviación estándar de la variación del mismo en los años estudiados [Dambreville14b, Dambreville14a]. Los días se dividen en nueve tipos, donde la media se divide en días que presentan medias diarias bajas, por lo que se consideran días nubosos, hasta los días que presentan media altas de radiación, por lo que representan los días despejados. La variabilidad se divide en días que presentan una variabilidad bastante baja, por lo que se consideran días con los valores de radiación estables, y días que presentan variabilidades altas.

Los datos satelitales utilizados en esta tesis se obtuvieron de la base de datos del Helioclim-3, en particular, de la versión 5 (HC3v5). El análisis de los datos satelitales ofrecidos por la cuadrícula obtenida alrededor de la isla nos permite confirmar los efectos de la meteorología de la zona en la radiación solar. Con la herramienta de visualización de los datos desarrollada se puede analizar la variación de la radiación cada 15 minutos. También se ha realizado la correlación entre las dos bases de datos (datos satelitales y datos terrestres de cada estación) utilizando los índices de cielo despejado para estudiar la relación entre ambos [Dambreville14b]. Para poder evaluar la correlación entre las series

temporales en diferentes momentos se establecieron desfases temporales entre ambas series. En cada estación se dispone de las correlaciones entre cada píxel de los datos satélite desfasado hasta 3 horas y el dato terrestre correspondiente a la propia estación sin desfasar. De esta manera se pretende conocer y valorar la relación entre cada punto geográfico de los alrededores y la estación en la que se desea realizar la predicción.

Los modelos estadísticos utilizados en esta tesis establecen una relación entre unos datos de entrada y una variable de salida esperada. Para obtener la relación entre ambas series temporales es necesario un entrenamiento a partir de un conjunto de datos históricos recogidos por las estaciones de medida. Durante este entrenamiento se debe estudiar además la complejidad de dichos modelos estadísticos. En el caso de los modelos lineales, esto conlleva decidir el orden de los mismos, es decir, el número de parámetros relevantes para realizar la predicción. En esta tesis se ha estudiado la complejidad de los modelos lineales siguiendo el método descrito por J. Boland [Boland95, Boland08], utilizando las Funciones de Autocorrelación y Autocorrelación Parcial, ACF y PACF, y los criterios de decisión bayesiana, BIC. En cuanto a las RNAs, la complejidad del modelo consiste en estudiar el número de variables de entrada y número de neuronas ocultas relevantes para la predicción. En esta tesis las teorías que se van a utilizar están basadas en la interpretación probabilista del algoritmo de retropropagación para el perceptrón multicapa realizado por Mackay [MacKay03, MacKay92]. Como ya se ha comentado, las RNAs se utilizan partiendo de diferentes conjuntos de datos de entradas. Para cada simulación que se realice durante el trabajo se ha estudiado la complejidad del modelo según las técnicas probabilistas bayesianas.

En los últimos años se han realizado diversos trabajos de predicción de radiación solar partiendo de datos terrestres en conjunto con datos satelitales o datos de un modelo numérico de predicción [Marquez13, Dambreville14b, Diagne14, Zagouras15]. Uno de los aspectos más importantes destacados en estos trabajos es la elección de los datos exógenos relevantes para la predicción. En este caso, se han utilizado los datos satelitales con mayores valores de correlación con respecto a la estación de medida para mejorar los resultados obtenidos con los datos terrestres. Utilizando la herramienta desarrollada para visualizar las correlaciones calculadas se realizaron diversas simulaciones cambiando la elección de píxeles. En cuanto a los datos de viento y humedad relativa del modelo numérico ECMWF se ha trabajado con el módulo y dirección de la velocidad y la humedad relativa a una altitud atmosférica determinada. Para determinar la altitud a la que extraer los datos referidos se utilizaron tres diferentes criterios en cada estación [Lave13, DA96, Badosa15].

Por último se ha realizado la aplicación de los modelos estadísticos de predicción a los datos de Gran Canaria para establecer el modelo que mejor se ajusta a las condiciones climáticas. Los resultados obtenidos con todos los modelos se han estudiado teniendo en cuenta los valores anuales de radiación, pero también se han estudiado los resultados según la estación del año y el tipo de día (siguiendo la distribución de nueve días obtenida). De esta manera se pretende conocer el comportamiento general de los modelos y su fiabilidad para distintas condiciones meteorológicas.

## Contenido de la tesis

En el Capítulo 2 se presentan los datos terrestres de radiación solar para cada una de las estaciones de la isla de Gran Canaria. Se exponen los resultados del tratamiento y análisis de los datos, estableciendo una primera relación entre las estaciones de medida según su ubicación. También se describen los datos de radiación solar, viento y humedad relativa obtenidos del modelo numérico de predicción ECMWF.

Los datos de radiación solar obtenidos a partir de imágenes satelitales se presentan en el Capítulo 3 de esta tesis. Se describen las herramientas desarrolladas para poder visualizar los datos del índice de cielo despejado y los valores obtenidos en las correlaciones entre cada píxel satelital y el dato terrestre. Además se presenta el análisis de los datos observados, tanto anualmente como por cada estación meteorológica del año.

Los modelos estadísticos utilizados durante la tesis se exponen en el Capítulo 4. Es en este capítulo donde se explican los modelos naïve (Persistence, Smart-Persistence), el modelo Climatológico, los modelos lineales (AR y ARMA) y los modelos basados en Redes Neuronales Artificiales (RNAs). Además se describe la metodología seguida para establecer una relación entre los datos de entrada y la salida deseada en cada caso. En el caso de los modelos lineales y las RNAs se exponen las técnicas utilizadas para establecer la complejidad de los modelos.

Por último, en el Capítulo 5 se presentan los resultados obtenidos para las estaciones de medida de la isla con cada uno de los modelos. Las simulaciones se han dividido en dos grandes grupos según el número de años de medida con los que se pueda trabajar. En un primer momento, se ha trabajado con un año de medida para el entrenamiento de los modelos y otro año de medida para comprobar el ajuste de los modelos entrenados. En este grupo se han simulado los modelos estadísticos con datos terrestres y datos de radiación solar del ECMWF como variables de entrada.

Los datos de radiación solar satelitales y los datos de viento y humedad relativa del ECMWF están únicamente disponibles para el año 2005. Así, en un segundo grupo se han realizado las simulaciones de los modelos estadísticos utilizando únicamente un año de medida. En este caso se debe dividir el año 2005 en dos conjuntos de datos, uno para realizar el entrenamiento y otro para el test.

Finalmente, en el Capítulo 6 se presentan las principales conclusiones y las líneas futuras de investigación del presente trabajo.



## Capítulo 2

# Fundamentos y datos de radiación solar

En este capítulo se realizará una breve introducción sobre la naturaleza de la radiación solar que llega a la atmósfera terrestre, de manera que se pueda entender la forma de proceder con los datos disponibles. Se explicarán los distintos factores que influyen en la radiación solar disponible en la atmósfera terrestre, la descomposición de la misma al atravesar la atmósfera y los ángulos de incidencia según la localización y la época del año. Para trabajar con los modelos de predicción de radiación solar elegidos en esta tesis se utilizarán las series temporales del índice de cielo despejado, por lo que se deberá calcular el modelo de cielo despejado para cada estación de medida.

Además también se presentarán los datos disponibles para la isla de Gran Canaria en las diferentes estaciones de medida del Instituto Tecnológico de Canarias, así como los aparatos de medida utilizados en dichas estaciones. Cuando se trabaja con una base de datos tan amplia, se considera básico contar con unos datos fiables. En este capítulo también se describirá el proceso de tratamiento de datos seguido en esta tesis y el posterior análisis de los mismos para realizar una caracterización del clima local.

Por último, se presentarán los datos obtenidos del modelo numérico de predicción solar, en concreto del European Centre for Medium-Range Weather Forecasts, *ECMWF*.

### 2.1. Fundamentos de radiación solar

#### 2.1.1. Naturaleza de la radiación solar

El sol es una esfera de gas caliente con un diámetro de  $1,39 \times 10^9$  m y está, de media, a  $1,5 \times 10^{11}$  m. de la Tierra. Visto desde la Tierra, el Sol gira sobre su eje una vez cada cuatro semanas, aunque en realidad no gira como una masa sólida. El ecuador gira en 27 días aproximadamente y las regiones polares se toman unos 30 días. Tiene una temperatura efectiva de 5777 °K (su núcleo puede llegar a los 15 millones °K), y una densidad estimada de 100 veces la del agua. El Sol es de hecho un gran reactor de fusión constituido por gases retenidos por fuerzas gravitacionales.

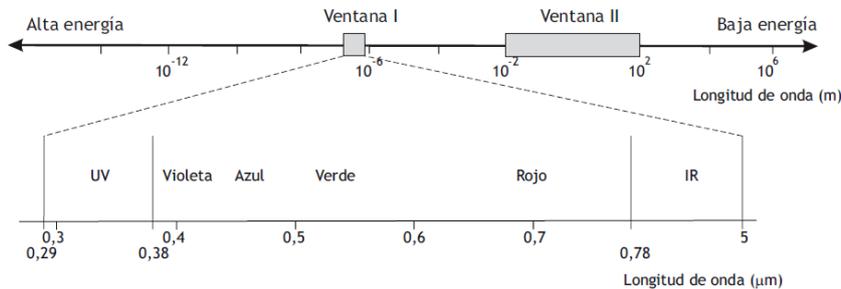


Figura 2.1: Espectro de longitudes de onda de la radiación solar

La radiación solar emitida es el resultado conjunto de varias capas, **área convectiva**, **fotosfera**, **cromosfera**, que emiten y absorben radiación de varias longitudes de onda. La radiación resultante y su espectro de distribución ha sido medida por varios métodos en numerosos experimentos.

La excentricidad de la órbita de la Tierra es tal que la distancia con el Sol varía un 1.7% a lo largo del año. La radiación emitida por el Sol y su relación con el espacio entre él y la Tierra resulta en una radiación de intensidad casi fija en la atmósfera terrestre. La constante solar,  $I_{sc}$ , es la energía procedente del Sol por unidad de tiempo, recibida sobre una superficie unitaria perpendicular a la dirección de propagación de la radiación fuera de la atmósfera. Según la consideración de diversos autores [Page86, Lorenzo06], existe una variación en la energía solar debida a que la distancia solar varía a lo largo del año y a las diferencias en la radiación emitida directamente por el Sol. Pero esta variación se considera despreciable para medidas energéticas, así que se toma un valor constante,  $I_{sc} = 1367 \text{ W/m}^2$ .

La radiación solar se mueve en un amplio espectro de longitud de ondas, con una cantidad específica de energía para cada longitud de onda, pero no toda esa energía es recibida en la superficie terrestre debido al filtrado de la atmósfera. El ojo humano es sensible a la radiación electromagnética con longitudes de onda comprendidas entre 380 y 780 nm. aproximadamente, margen que se denomina *luz visible*. Las longitudes de onda más cortas del espectro visible corresponden a la *luz violeta* y la más larga a la *luz roja*, y entre estos extremos se encuentran todos los colores del arco iris. Las ondas electromagnéticas con longitudes de onda ligeramente inferiores a las de la luz visible se denominan *rayos ultravioleta*, y las que poseen longitudes de onda ligeramente superiores, se conocen como *ondas infrarrojas*. La mayoría de la energía recibida corresponde al ancho de banda de la luz visible, ver Figura 2.1.

El espectro de la radiación de onda filtrada al atravesar la atmósfera, debido a fenómenos de absorción y dispersión, se ve modificado antes de alcanzar la superficie terrestre. En la Figura 2.2, se observa la diferencia entre la radiación exoatmosférica, considerada para  $AM0$ , y la radiación en la superficie terrestre, considerada para una  $AM1.5$ .

Los componentes del atmósfera actúan sobre la radiación cuando la atraviesa. De esta manera, parte es reflejada por las nubes, otra absorbida por ozono, oxígeno, anhídrido

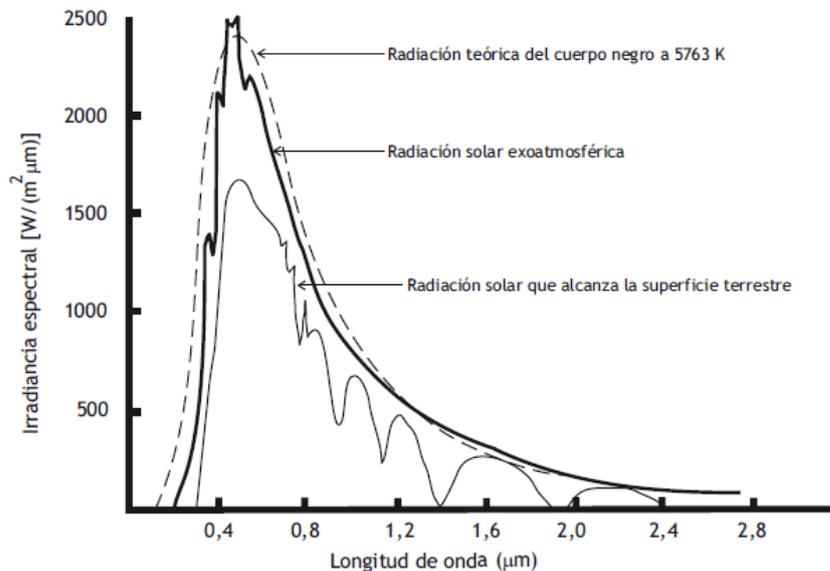


Figura 2.2: Distribución espectral de la radiación solar

carbónico, vapor de agua, etc. y por último dispersada por moléculas, gotas de agua, polvo en suspensión, etc. Una vez llega al suelo es a su vez absorbida por el propio suelo y parte reflejada por éste otra vez hacia la atmósfera. Como resultado podemos hablar de la radiación solar global descompuesta en tres componentes bien diferenciadas, ver Figura 2.3:

- **Radiación solar directa**, es la radiación recibida del Sol sin ser dispersada en la atmósfera. Está constituida por los rayos que se reciben directamente del Sol.
- **Radiación solar difusa**, es la que procede del cielo pero no de los rayos del Sol, proviene de la radiación dispersada en la atmósfera. Las propiedades direccionales dependen en gran medida de la posición, forma y composición de las nubes, cuya modelización es una función compleja y variable en el tiempo.
- **Radiación solar albedo**, procede del suelo y es debida a la reflexión de la radiación incidente en él. La cuantía depende de la naturaleza del suelo. En algunos casos particulares los alrededores de los receptores se preparan de forma deliberada para incrementar el albedo y que contribuya a la radiación global.

### 2.1.2. Principios del movimiento solar

Uno de los aspectos básicos para definir la radiación solar es conocer la dirección de los rayos del Sol al incidir en la superficie terrestre. La Tierra realiza dos movimientos principales en su desplazamiento por la órbita solar, que determinan la cantidad de energía

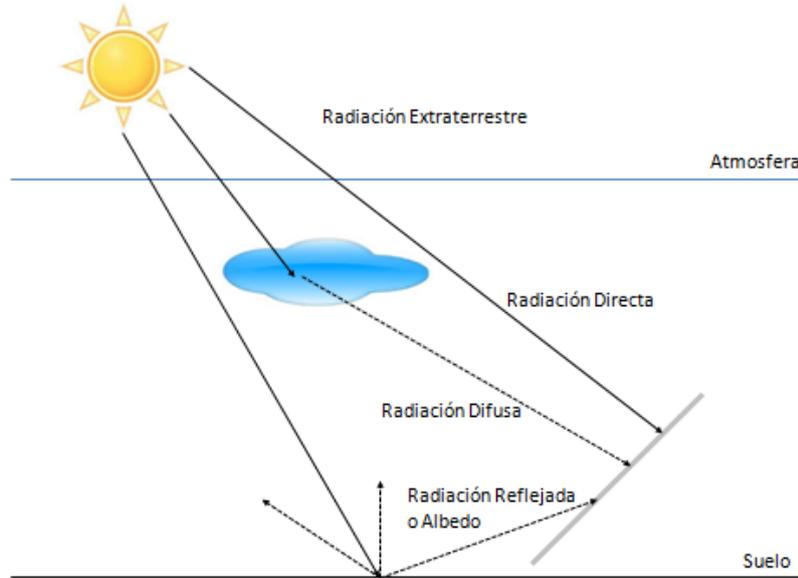


Figura 2.3: Componentes de la radiación solar incidente en una superficie receptora

que recibimos a lo largo del día y año. Un movimiento de *rotación* sobre su propio eje y un movimiento de *traslación* alrededor del Sol siguiendo la órbita terrestre.

En el movimiento de traslación la Tierra describe una órbita elíptica alrededor del Sol, que se encuentra en uno de los focos. Se denomina plano de la eclíptica al plano que contiene dicha órbita y se ha definido como un año el tiempo que la Tierra tarda en recorrerla, ver Figura 2.4.

La excentricidad de la órbita elíptica es de sólo 0.017, por lo que la distancia en el Sol y la Tierra varía a lo largo del año. Debido a que esta variación no es significativa se suelen utilizar unas expresiones aproximadas basadas en considerar la velocidad angular constante. Teniendo en cuenta este aspecto la distancia entre el Sol y la Tierra se ha definido de manera aproximada por la siguiente ecuación (2.1),

$$r = r_0 \left( 1 + 0.0017 \sin \left( \frac{360(n + 93)}{365} \right) \right) \quad (2.1)$$

donde  $r$  es la distancia entre el Sol y la Tierra,  $r_0$  es la distancia promedio definida como  $1.49 \times 10^8 \text{ km}$  y  $n$  es el número del día del año en el que se quiere calcular el radio, siendo 1 para el día 1 de enero y 365 para el 31 de diciembre. A partir de esta definición se obtiene la variación de la distancia entre el Sol y la Tierra según la expresión (2.2).

$$\epsilon_0 = \left( \frac{r}{r_0} \right)^2 = 1 + 0.0033 \cos \left( \frac{360n}{365} \right) \quad (2.2)$$

Por otro lado, si se tiene en cuenta que la velocidad angular de la Tierra en su camino alrededor del Sol no es constante, se ajusta a las leyes de Kepler, *los planetas barren áreas*

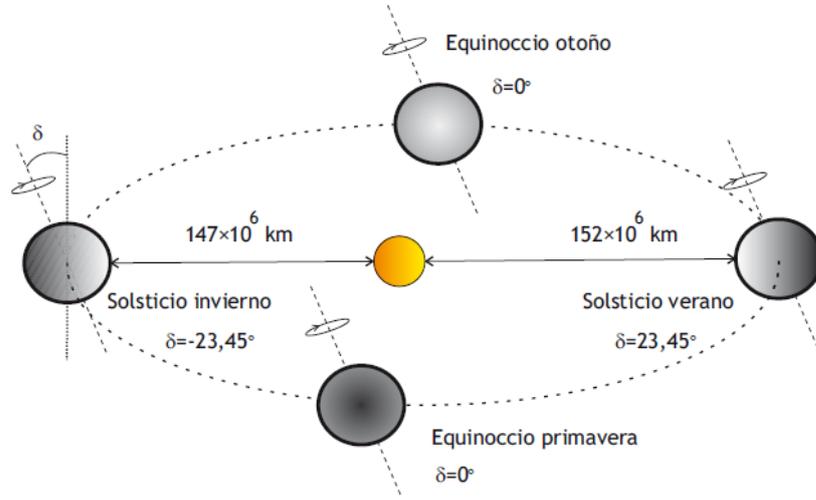


Figura 2.4: Órbita de la Tierra alrededor del Sol. Fuente [DR13]

*iguales en tiempos iguales*. La expresión de la variación de la distancia en el Sol y la Tierra utilizada finalmente es la siguiente (2.3).

$$\epsilon_0 = 1.00011 + 0.034221 \cos \tau + 0.001280 \sin \tau + 0.000719 \cos 2\tau + 0.000077 \sin 2\tau \quad (2.3)$$

$$\tau = \frac{2\pi(n-1)}{365} \quad (2.4)$$

El movimiento de rotación consiste en el giro que realiza la Tierra alrededor de su eje central, llamado eje polar. El tiempo que tarda en completar un giro se conoce como un día. El plano de la eclíptica, que define el movimiento de traslación, no es perpendicular a este eje de rotación. El eje polar tiene una inclinación de  $23.45^\circ$  con respecto a este plano, manteniendo una orientación constante.

La inclinación del eje es la causante de que el Sol se encuentre más alto en el cielo al mediodía durante el verano que en invierno y por lo tanto, que el ángulo de inclinación de los rayos solares varíen. De la misma manera, para una latitud dada, el ángulo entre el ecuador terrestre y la línea que une los centros de la Tierra y el Sol varía a lo largo del año. Este ángulo se conoce como declinación  $\delta$ , y se considera constante para un mismo día en los cálculos que estamos realizando, ver Figura 2.5.

La declinación solar es cero en los equinoccios de primavera (20/21 marzo) y de otoño (22/23 septiembre). Estos días el Sol se encuentra en el ecuador, haciendo que la duración de los días sea igual a la de las noches en toda la Tierra y que el Sol salga por el Este y se ponga por el Oeste. En el solsticio de verano (21/22 junio) la declinación es igual a  $23.45^\circ$ , y el Sol se encuentra en el Trópico de Cáncer, provocando que en el hemisferio Norte se tenga el día más largo y la noche más corta del año y que la salida y puesta del Sol se desplacen

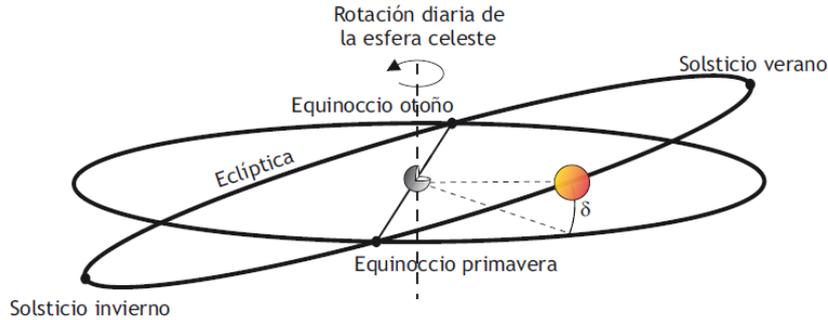


Figura 2.5: Plano de la eclíptica. Fuente [DR13]

al NE y NO respectivamente. En el solsticio de invierno (21/22 diciembre) la declinación es igual a  $-23.45^\circ$ , y el sol se encuentra en el Trópico de Capricornio, provocando en el hemisferio Norte la noche más larga y el día más corto del año. En este caso la salida y puesta del Sol se desplazan al SE y SO respectivamente. En el hemisferio Sur ocurre al revés que en el Norte en estos casos, ver Figura 2.6.

En esta tesis se utilizará una expresión matemática para hallar la declinación para cada día del año. La ecuación sólo depende del número de orden del día dentro del año  $n$ .

$$\delta = (0.006918 - 0.399912 \cos \tau + 0.070257 \sin \tau - 0.006758 \cos 2\tau + 0.000907 \sin 2\tau - 0.002697 \cos 3\tau + 0.00148 \sin 3\tau) \frac{180}{\pi} \quad (2.5)$$

Es muy común representar el cielo situando la Tierra fija en el centro de una esfera, ver Figura 2.7. De esta manera, cada un de los puntos de la esfera representan una dirección del cielo vista desde la Tierra. La esfera se conoce como *esfera celeste*. El ecuador celeste queda definido por la intersección del plano del ecuador con dicha esfera, mientras que los polos celestes son la intersección con el eje polar.

A partir de esta representación, el movimiento de la Tierra alrededor del Sol se describe como un movimiento del Sol alrededor de la Tierra siguiendo el mayor círculo que forma un ángulo con el ecuador celeste igual a la inclinación del eje polar,  $23.45^\circ$ . Este plano representa el plano de la eclíptica. Los dos movimientos se representarán a través de un movimiento del Sol recorriendo la eclíptica una vez al año y un giro una vez al día de la esfera celeste alrededor de la Tierra. De esta manera, el Sol cada día describe un círculo alrededor de la Tierra que varía según el orden del día del año, siendo máximo en los equinoccios y mínimo en los solsticios. El giro de la esfera celeste alrededor de la Tierra y del sol alrededor de la eclíptica es de sentido contrario.

La posición relativa del sol respecto a un plano situado en la superficie terrestre, con cualquier orientación y en cualquier momento, se puede describir utilizando varios ángulos en un sistema de coordenadas solidario con un punto de la Tierra, ver Figura 2.8.

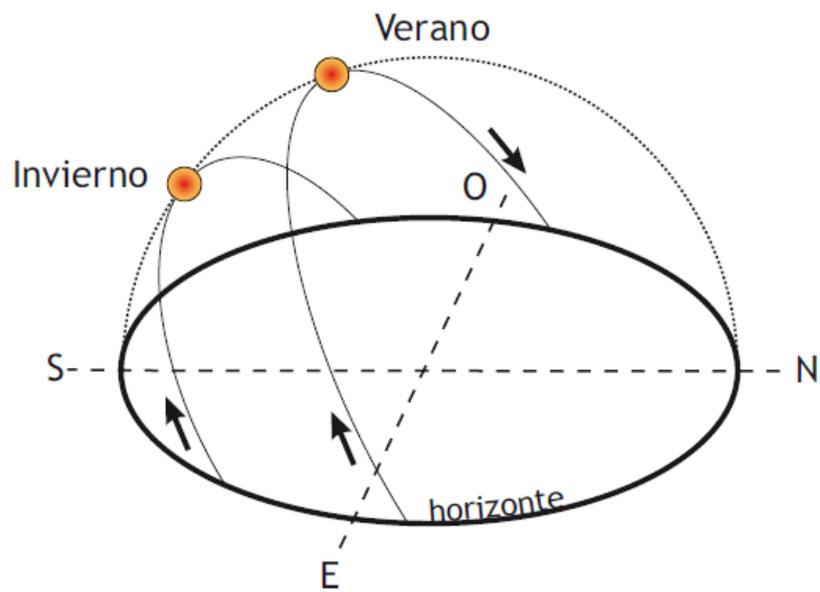


Figura 2.6: Trayectoria solar diaria para verano e invierno. Fuente [DR13]

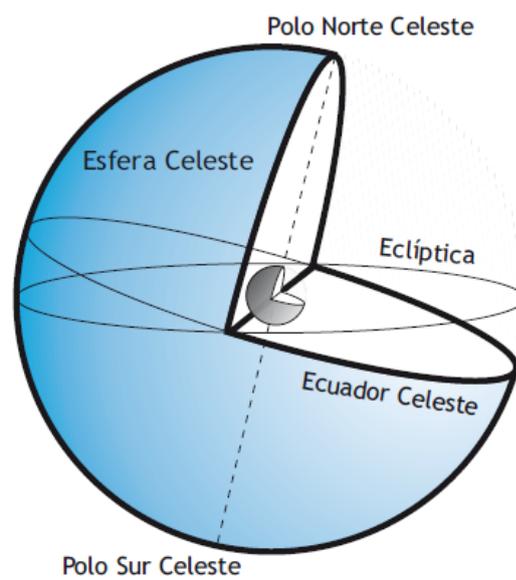


Figura 2.7: Esfera celeste y plano de la eclíptica. Fuente [DR13]

- **Latitud**  $\Phi$ , si se tiene en cuenta que la vertical de un lugar específico de la tierra intersecta a la esfera celeste en dos puntos llamados cenit (Norte) y nadir (Sur). El complementario del ángulo que forma esta recta con el eje polar es la latitud (posición angular norte o sur respecto del ecuador). Se denomina meridiano al círculo máximo de la esfera terrestre que contiene los polos, al cenit y al nadir;  $-90^\circ < \Phi < 90^\circ$ . Se considera positiva en el hemisferio Norte y negativa en el hemisferio Sur.
- **Declinación**  $\delta$ , la posición angular del sol cuando éste se encuentra en el meridiano del lugar (mediodía solar) con respecto al plano del ecuador y representa el ángulo entre el ecuador terrestre y la línea que une los centros de la Tierra y el Sol; oscila a lo largo del año entre  $-23.45^\circ < \delta < 23.45^\circ$ .
- **Acimut solar**  $\gamma_s$ , ángulo desviación de la proyección en el plano horizontal del meridiano del Sol con el meridiano del lugar (el Sur en el hemisferio norte y el Norte en el hemisferio sur). En el Este es negativo y positivo en el Oeste;  $-180^\circ < \gamma_s < 180^\circ$ . Para latitudes, norte y sur, entre  $23.45^\circ$  y  $66.45^\circ$ , el acimut solar va a estar entre  $-90^\circ$  y  $90^\circ$  para días de menos de 12 horas de duración; si los días duran más entre la salida y la puesta de sol el acimut solar tendrá valores mayores de  $90^\circ$  y  $-90^\circ$ . El ángulo azimut solar marca el recorrido solar a lo largo del día.
- **Ángulo de incidencia**  $\theta$ , es el ángulo entre la radiación directa sobre la superficie y la normal a esa superficie.
- **Ángulo cenital**  $\theta_{zs}$ , es el ángulo entre la vertical y la línea hacia el Sol, es decir, el ángulo de incidencia de la radiación directa sobre una superficie horizontal. La relación entre la cantidad de atmósfera que tiene que atravesar la radiación directa y la que tendría que atravesar si el Sol estuviera en su cenit se denomina cantidad de aire  $m$  (*air mass*). Al nivel del mar  $m = 1$  si el Sol está en su cenit, y  $m = 2$  si tenemos un ángulo cenital de  $60^\circ$ . La parte de la radiación exoatmosférica que llega a la superficie terrestre dependerá de este valor.
- **Ángulo de altitud solar o elevación**  $h_0$ , ángulo entre la horizontal y la línea hacia el sol, complementario del anterior. La elevación es el ángulo con el que vemos al sol si miramos en su dirección, tomando como origen la horizontal (suelo).

Definiendo la latitud como positiva en el hemisferio Norte y negativa en el hemisferio Sur, se define el ángulo cenital para una localidad y una época del año según la expresión (2.6).

$$\cos(\theta_{zs}) = \sin(h_0) = \sin \delta \sin \Phi + \cos \delta \cos \Phi \cos \omega \quad (2.6)$$

El ángulo horario  $\omega$  se define como el desplazamiento angular, partiendo de Este a Oeste, del meridiano local según la rotación terrestre. Este valor se considera cero al mediodía, negativo durante la mañana y positivo por la tarde. El movimiento de rotación de la Tierra se considera a una velocidad de  $15^\circ$  por hora. La expresión utilizada para calcular el ángulo horario es (2.7),

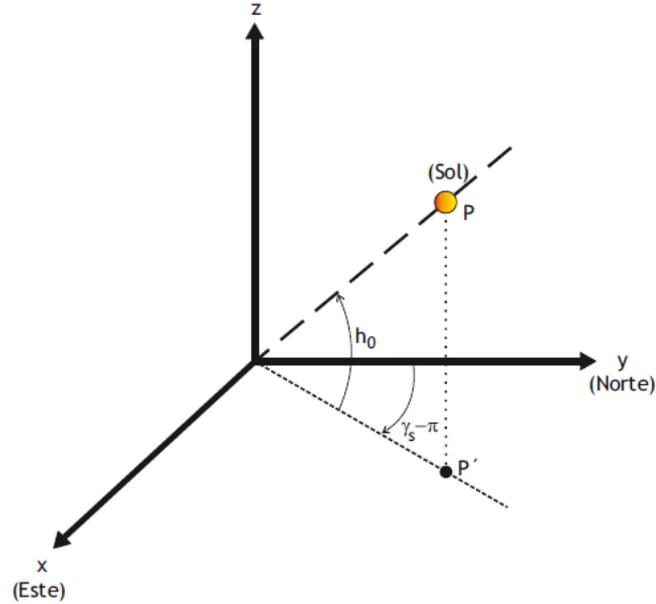


Figura 2.8: Sistema de referencia sobre la superficie terrestre. Fuente [DR13]

$$\omega(\circ) = 15((Hr - 12) + (Lo - 15Tz)\frac{4}{60} + ET/60) \quad (2.7)$$

donde  $Lo$  es la longitud local,  $Tz$  es el tiempo de adelanto respecto al meridiano de Greenwich (meridiano origen del ángulo horario) del meridiano correspondiente al huso horario local y  $ET$  es la ecuación del tiempo que varía de un día a otro según la ecuación (2.8).

$$ET(min) = (0.000075 + 0.001868 \cos(\tau) - 0.032077 \sin(\tau) - 0.014615 \cos(2\tau) - 0.040849 \sin(2\tau))229.18 \quad (2.8)$$

El ángulo horario es el tiempo basado en el movimiento angular aparente del sol sobre el cielo, con el mediodía solar el sol cruza el meridiano local y la ecuación del tiempo tiene en cuenta las perturbaciones en la rotación de la tierra que afecta a la hora en la que el sol cruza el meridiano.

### 2.1.3. Radiación exoatmosférica sobre superficie horizontal

La radiación exoatmosférica sobre un plano normal a la radiación directa proveniente del Sol para cualquier día del año, se puede calcular a partir de una expresión sencilla teniendo en cuenta la ligeras variaciones de la distancia entre el sol y la tierra.

$$I_{0n} = I_{sc}\epsilon_0 \quad (2.9)$$

En esta ecuación, como ya se explicado anteriormente,  $I_{sc}$  representa la energía procedente del Sol por unidad de tiempo, recibida sobre una superficie unitaria perpendicular a la dirección de propagación de la radiación fuera de la atmósfera. Esta energía se conoce como constante solar y se toma un valor constante,  $I_{sc} = 1367W/m^2$ . Por otra parte,  $\epsilon_0$  es la variación de la distancia entre el Sol y la Tierra lo largo del año y se calcula según la ecuación (2.3).

Teniendo en cuenta la relación entre la radiación sobre una superficie inclinada y una horizontal al plano terrestre, se puede expresar la radiación exoatmosférica sobre una superficie horizontal en  $W/m^2$  como (2.10).

$$I_0 = I_{sc}\epsilon_0 \cos(\theta_{zs}) \quad (2.10)$$

Sustituyendo la ecuación del ángulo cenital (2.6) en la ecuación (2.10) se obtiene la expresión final de la radiación exoatmosférica.

$$I_0 = I_{sc}\epsilon_0(\sin \delta \sin \Phi + \cos \delta \cos \Phi \cos \omega) \quad (2.11)$$

A menudo es necesario para los cálculos de irradiación solar diaria tener la integral diaria de la radiación exoatmosférica sobre una superficie horizontal,  $H_0$ . Ésta se obtiene integrando la ecuación anterior en el periodo desde la salida hasta la puesta de sol, donde  $\omega_s$  es el ángulo de la puesta de sol.

$$H_0 = \frac{24 \cdot 3600}{\pi} I_{sc}\epsilon_0 \left( \frac{\pi\omega_s}{180} \sin \delta \sin \Phi + \cos \delta \cos \Phi \cos \omega_s \right) \quad (2.12)$$

En este caso la radiación exoatmosférica diaria viene definida en  $J/m^2 dia$ .

## 2.2. Modelo de Cielo despejado

Las series temporales de radiación solar no se consideran series estacionarias, ya que están influenciadas por la variabilidad diaria y anual del movimiento de la Tierra. Cuando se trabaja con modelos estadísticos de predicción solar, se debe partir de series estacionarias. Las series estacionarias son aquellas que no muestran una tendencia estacional y tienen una varianza constante, presentando una auto correlación constante en todo el intervalo temporal de la misma [Chatfield13].

La radiación solar incidente en una superficie horizontal en una localidad concreta para un tiempo específico depende del ángulo cenital. Para trabajar con modelos estadísticos se considera adecuado tratar las influencias deterministas dependientes de la geometría solar por separado de las influencias no deterministas generadas por fenómenos atmosféricos [Diagne13]. Así, para conseguir transformar las series temporales de radiación en series estacionarias se han introducido dos variables, el índice de claridad  $k$  y el índice de cielo despejado  $k^*$ .

El índice de claridad se considera como el resultado de dividir la radiación solar en una superficie horizontal  $I$  entre la radiación exoatmosférica en la misma localidad para una

superficie horizontal  $I_0$ , según la ecuación (2.13). El índice de claridad consigue extraer la influencia determinista de la geometría solar, ya que la influencia del ángulo cenital se encuentra incluido en el cálculo de la radiación exoatmosférica. Este índice se utiliza ampliamente para tratar la influencia determinista.

$$k = \frac{I}{I_0} \quad (2.13)$$

La segunda variable introducida, el índice de cielo despejado, es ampliamente utilizada en la bibliografía como método para evitar el carácter estacional de las series temporales de radiación solar horaria. En esta tesis se ha optado por utilizar las series temporales del índice de cielo despejado como variable en los modelos de predicción. De esta manera se consiguen series temporales estacionarias mediante la siguiente fórmula (2.14),

$$k^* = \frac{I}{I_{cs}} \quad (2.14)$$

donde  $I$  es la radiación solar horaria sobre una superficie horizontal medida, por ejemplo, en una estación de medida. Mientras que  $I_{cs}$  es la radiación solar calculada para un día a cielo despejado. Los modelos de radiación a cielo despejado, *clear sky models*, estiman la radiación que incide en cualquier superficie en cualquier instante de tiempo considerando unas condiciones de cielo limpio. Algunos de los modelos de cielo despejado consultados en la bibliografía se basan en diferentes variables climáticas como datos de entrada [Reno12, Younes07]. En la isla de Gran Canaria se han probado diversos modelos de cielo despejado, mostrando resultados satisfactorios y similares entre sí [DR13, BM01, Šúri04, Hofierka02, Bird81]. El modelo que se utilizará en esta tesis, por su precisión y sencillez de aplicación, es el modelo de Bird & Hulstrom [Bird81]. Este modelo es ampliamente conocido y utilizado porque proporciona excelentes resultados a partir únicamente de una serie de datos meteorológicos [Badescu13].

Las variables meteorológicas que utiliza el modelo de cielo despejado son los espesores ópticos de aerosol para longitudes de onda de 500 nm y 380 nm (*Aerosol Optical Depths AOD500 nm and AOD380 nm*), y el contenido de vapor de agua y ozono en la columna vertical de la atmósfera. AODs miden la atenuación de la radiación solar como resultado de la dispersión y absorción de la luz solar en la columna vertical de la atmósfera.

Los valores AOD500 nm, AOD380 nm y el contenido en vapor de agua de la columna vertical se han obtenido de la red AERONET [AERONET14, Holben98]. De esta fuente se han podido extraer para las Islas Canarias las medias mensuales de los datos desde año 2008 hasta el año 2014. A partir de este conjunto de datos se calculó la media climatológica de la zona en cuestión y se mantuvieron los valores obtenidos para la estimación del modelo de cielo despejado para todo el año. La medición de AERONET para las Islas Canarias se realiza en la isla de Tenerife.

El contenido en Ozono de la atmósfera se ha obtenido del *World Ozone Monitoring Mapping* perteneciente al Gobierno Canadiense [Canada's15]. Al igual que en el caso anterior existe un dato disponible para las Islas Canarias y se han resumido todos los datos obtenidos en una media climatológica anual.

		C0-Pozo Izquierdo	C1-Las Palmas
Proveedor datos		ITC	ITC
Periodo datos		2003 - 2005	2002 - Jan-Jun 2003 - Jul-Dec 2004 - 2005
Intervalo tiempo toma datos		1 min	1 min
	Longitud (°)	-15.4244	-15.4269
Coordenadas	Latitud (°)	27.8175	28.1108
Geográficas	Zona horaria (h)	-1	-1
	Altitud (m)	47	17
Media presión atmosférica (Pa)		101202	101202
Columna Ozono (cm)	Media	0.3	0.3
Columna vapor agua (cm)	Media	1.815	1.815
AOD500nm	Media	0.159	0.159
AOD380nm	Media	0.184	0.184
Ba (Asymmetric factor)		0.84	0.84
Nº horas de cielo despejado [Ineichen06]		3224	1660
rRMSE %		3.97 %	3.98 %
Variabilidad del lugar [Hoff12]		0.14	0.18

Tabla 2.1: Variables meteorológicas del modelo de cielo despejado y resultados de su estimación.

El modelo de Bird [Bird81] estima la radiación solar global en una superficie horizontal para un día despejado. A partir de los datos anteriores, se calculan una serie de variables como la dispersión de Rayleigh, las absortancias del Ozono, gases como el oxígeno y dióxido de carbono, el vapor de agua y la absortancia y dispersión de los aerosoles. Mediante estas variables se estimará finalmente la radiación global incluyendo la radiación exoatmosférica y el ángulo cenital en el momento y lugar especificados.

En esta tesis se ha estimado el modelo de cielo despejado de Bird para todas las horas del año en cada una de las estaciones de medida del Instituto Tecnológico de Canarias. En la Tabla 2.1, se pueden observar los datos utilizados y algunos resultados obtenidos al estimar el modelo de cielo despejado para dos estaciones de medida en la isla de Gran Canaria. En este caso se ha optado por mostrar dos estaciones significativas de los dos climas claramente diferenciales en la isla, Pozo Izquierdo-C0 en el sureste de la isla y Las Palmas-C1 en el noreste de la isla.

Para comprobar la precisión del modelo de cielo despejado de Bird no se pueden comparar los resultados obtenidos con todos los datos disponibles, ya que en muchos casos no corresponden a condiciones de cielo limpio. La comprobación se realizará únicamente comparando las horas de radiación correspondientes a cielo despejado. En esta tesis se ha utilizado el modelo de Ineichen [Ineichen06] para reconocer los datos correspondientes a condiciones de cielo limpio entre todo el conjunto de datos. Finalmente, se calculará el

	C0	C1	C2	C4	C5	C6
$N^\circ$ horas de la muestra	17520	26280	14592	17544	17520	17520
$N^\circ$ horas de cielo despejado [Ineichen06]	3224	1660	3159	1675	1448	2765
Precisión modelo Bird rRMSE %	3.97	3.98	4.06	4.93	4.23	4.01
Variabilidad del lugar [Hoff12]	0.14	0.15	0.13	0.16	0.18	0.20

Tabla 2.2: Estimación de la precisión del modelo de cielo despejado de Bird.

error cuadrático medio relativo (**%rRMSE**) resultante de comparar el modelo de cielo despejado estimado con el modelo de Bird y los datos reales correspondientes a cielo despejado.

En la Tabla 2.2 se puede observar el error global obtenido al aplicar el modelo de cielo despejado de Bird para cada hora del año en todas las estaciones de medida de la isla de Gran Canaria. El error obtenido para todas las estaciones se encuentra alrededor del 4%, por lo que el modelo se considera adecuado para nuestro propósito. Estos errores %rRMSE se encuentran en mismo orden que el modelo de McClear, que usa los datos de AOD, Ozono y vapor de agua del proyecto MACC [Lefevre13]. Esto significa que la variabilidad de estos parámetros se puede considerar no relevante y por lo tanto utilizar las medias es suficiente. En la Tabla 2.2 no se observa ninguna diferencia apreciable en el error obtenido entre las estaciones del Sur y del Norte. Por otro lado, el número de horas de sol en condiciones de cielo despejado detectadas por el método de Ineichen es, en general, mayor en las estaciones del Sur. Las estaciones del Norte de la isla presentan mayor número de días nublados por la presencia de nubes en verano debido a los vientos Alisios.

En este apartado también se propone una variable para estimar la variabilidad de la radiación solar en cada estación. Para cada estación se ha calculado la variabilidad en los datos de radiación solar horarios por el método propuesto por Hoff & Perez [Hoff12]. La variable propuesta por Hoff & Perez es la desviación típica de la variación del índice de cielo despejado  $k^*$  para dos horas consecutivas. Un lugar que presente un valor superior a 0.2 en esta variables se considera que tiene condiciones climáticas inestables. En el caso de Gran Canaria, como se observa, todas las estaciones presentan valores inferiores, aunque siempre menores en las estaciones del Sur cuyo clima es más estable.

### 2.3. Datos de radiación solar disponibles en las estaciones de medida

Los modelos de predicción que se utilizarán en esta tesis necesitan un conjunto de datos de radiación solar medidos para poder realizar el *aprendizaje*. Por lo tanto, es muy importante disponer de una base de datos suficientemente amplia y fiable para que el proceso pueda ofrecer buenos resultados. Antes de comenzar a utilizar los datos se deberá realizar un tratamiento y análisis de los mismos, de manera que la base de datos finalmente obtenida constituya un conjunto lo más fiable posible.

En esta tesis se llevarán a cabo los siguientes pasos para conseguir este propósito:



Figura 2.9: Distribución geográfica de la estaciones de medida del Instituto Tecnológico de Canarias en la isla de Gran Canaria

- Estudio del conjunto de datos y del porcentaje de datos sin huecos de los que se dispone en cada estación.
- Tratamiento de los datos disponibles para filtrar aquellos datos considerados incorrectos.
- Analizar los datos disponibles.

### 2.3.1. Estaciones de medida en la isla de Gran Canaria

Los datos de radiación solar horaria utilizados en esta tesis provienen de las estaciones de medida del Instituto Tecnológico de Canarias (ITC) en la isla de Gran Canaria. En las Islas Canarias el ITC dispone de 23 estaciones de medida, con las que se ha realizado el mapa solar de la región [Díaz12]. En la isla de Gran Canaria existen siete estaciones de medida repartidas por la geografía insular, ver Figura 2.9. Se dispone de los datos de radiación de dichas estaciones desde año 1998 hasta 2010, aunque, dependiendo de la estación el período puede ser menor.

Aunque se disponen de siete estaciones, únicamente se va a trabajar en esta tesis con seis ya que la estación de Gáldar no dispone de datos suficientes para conseguir un conjunto fiable. Para reconocer las estaciones éstas se han nombrado con una 'C' y un

Código estación	Latitud	Longitud	Altitud
C0	27.82	-15.42	47
C1	28.11	-15.42	17
C2	27.99	-15.79	197
C4	27.77	-15.58	265
C5	28.03	-15.49	525
C6	27.88	-15.72	300

Tabla 2.3: Coordenadas geográficas de las estaciones de medida del ITC en la isla de Gran Canaria.

código numérico, de manera que en la isla de Gran Canaria se encuentran las siguientes estaciones:

- C0-Pozo Izquierdo
- C1-Las Palmas.
- C2-La Aldea
- C4-Maspalomas
- C5-Sta. Brígida
- C6-Mogán
- C7-Gáldar (no se utilizará en esta tesis)

Se puede observar que las estaciones cubren la mayoría del territorio insular. Las estaciones C1 y C5 (Las Palmas y Sta. Brígida) en la zona Norte, mientras que las estaciones C0, C2, C4 y C6 (Pozo Izquierdo, La Aldea, Maspalomas y Mogán) se encuentran en la vertiente sur de la isla. Por otro lado la mayoría de las estaciones se encuentran prácticamente en la costa, excepto C6 y C5 que se encuentran en el interior y a mayor altitud sobre el nivel del mar.

En la Tabla 2.3 se pueden observar la coordenadas geográficas de latitud, longitud y altitud de cada estación. Estas coordenadas se utilizarán para realizar cualquier cálculo o estimación de radiación solar.

En las seis estaciones que finalmente se van a utilizar en esta tesis se recogen datos de irradiancia solar en potencia  $W/m^2$ . Los datos de radiación se recogen cada minuto o cada 5 minutos, dependiendo de la estación y del año en concreto. A partir de los datos almacenados en la base de datos se realizan los cálculos de la radiación en  $Wh/m^2$  horaria y diaria integrando los datos de potencia. Aunque en esta tesis, únicamente se van a utilizar los datos de radiación global horizontal, en las estaciones de medida se dispone de equipos para medir otras variables meteorológicas. Dependiendo de la estación y del año, nos encontramos con datos de presión atmosférica, temperatura ambiente, humedad relativa, horas de sol o radiación difusa.

Año	ene	feb	mar	abr	may	jun	jul	ago	sep	oct	nov	dic
2003	100	99.7	96.6	99.7	93.5	100	100	100	100	93.5	100	100
2005	100	96.4	93.5	96.7	100	96.7	96.8	96.8	83.3	100	90	90.3

Tabla 2.4: Porcentaje de datos correctos disponibles en los años elegidos para la estación de C0-Pozo Izquierdo.

Como ya se ha dicho, las estaciones disponen de medidas registradas desde 1998 hasta 2010, excepto C0-Pozo Izquierdo cuyas medidas comenzaron a tomarse a partir de 2001. En cualquier caso, la base de datos muestra lagunas de datos derivadas del mal funcionamiento de los equipos o que se encuentren en revisión.

Los datos horarios cedidos por el ITC ya habían sido previamente analizados. El ITC asigna un código de error negativo a todos los datos recogidos según si se observa un fallo en los instrumentos de medida, fallo en el sensor o en el almacenado de la información. Cuando se va a proceder a la integración de los datos horarios, ésta no se realiza para los datos marcados con un código de error negativo. Los datos horarios cedidos por el ITC exigen un 50% de datos correctos para calcular las estadísticas, en caso contrario se considera que no existe dato de medida en esa hora.

Los modelos de predicción que se utilizan en esta tesis se basan en el aprendizaje a partir de un conjunto de datos históricos. La predicción de datos futuros se basa en la observación de un conjunto de datos pasados. En cualquier proceso de aprendizaje se debe disponer de un conjunto de datos de entrenamiento, con el que se definen los modelos teóricos, y un conjunto de test, con el que se validan los modelos obtenidos. Ambos conjuntos deben ser comparables y contener una distribución similar de todas las posibles situaciones climáticas que se pueden observar en la isla. En esta tesis, se ha decidido trabajar con un año de datos de entrenamiento y un año de datos para la validación, de manera que en ambos casos se encuentren representadas todas las estaciones del año.

Los modelos estadísticos de series temporales reproducen mejores resultados de predicción utilizando bases de datos con una mayor continuidad en los mismos. Es por ello, que una vez analizados los datos se decidió trabajar únicamente con años prácticamente completos, en la que todos los meses presentaran un porcentaje de datos perdidos los más bajo posible. En algunas estaciones no se encontraron años completos en los que todos los meses presentaran una continuidad suficiente, por lo que se procedió a diseñar años ficticios como mezcla de dos años distintos. Una vez revisados los datos de todas las estaciones se muestran en las tablas (2.4-2.9) los porcentajes de datos de radiación solar IGH de los años finalmente elegidos para este estudio.

Por último en la tabla 2.10 se muestran los conjuntos de datos finalmente elegidos para el estudio. Se ha dividido la base de datos en un conjunto de entrenamiento y otro de test para cada estación.

Año	ene	feb	mar	abr	may	jun	jul	ago	sep	oct	nov	dic
2002	99.9	100	100	96.5	100	100	100	77.1	80	79.6	100	97.8
2003	99.9	100	99.8	100	99.9	78.75	0	0	0	6.5	96.5	100
2004	100	100	96.6	83.3	79	67.4	98.8	95.2	100	100	100	93
2005	96.6	99.4	77.3	80	98.1	100	100	99.9	100	100	93.1	78.9

Tabla 2.5: Porcentaje de datos correctos disponibles en los años elegidos para la estación de C1-Las Palmas.

Año	ene	feb	mar	abr	may	jun	jul	ago	sep	oct	nov	dic
2001	100	99.3	96.8	83.2	93.7	100	100	100	99.9	100	62.8	87
2002	98.8	100	99.9	100	95	100	100	100	99.9	100	89.4	38.4

Tabla 2.6: Porcentaje de datos correctos disponibles en los años elegidos para la estación de C2-La Aldea.

Año	ene	feb	mar	abr	may	jun	jul	ago	sep	oct	nov	dic
2000	100	100	100	100	100	100	0	0	0	0	0	0
2003	99.7	89.1	98.9	99.9	100	100	99.1	91.7	96.3	99.1	93.3	91.1
2005	99.9	71.4	99.7	96.8	73.5	38.2	100	100	96.7	100	99.6	98.7

Tabla 2.7: Porcentaje de datos correctos disponibles en los años elegidos para la estación de C4-Maspalomas.

Año	ene	feb	mar	abr	may	jun	jul	ago	sep	oct	nov	dic
2001	63.2	42.9	0	0	92.1	100	100	99.3	0	96.8	99.4	98.8
2002	99.6	99.9	83.3	99.4	83.9	6	2	80.4	99.9	99.7	99.4	99.1
2003	99.6	100	99.1	100	100	100	100	99.3	96.7	98.9	98.5	98.9

Tabla 2.8: Porcentaje de datos correctos disponibles en los años elegidos para la estación de C5-Sta. Brígida.

Año	ene	feb	mar	abr	may	jun	jul	ago	sep	oct	nov	dic
2002	100	100	100	100	100	100	100	100	100	63.4	100	96.8
2003	100	100	100	100	100	100	100	67.7	96.7	100	100	100
2005	100	100	100	100	35.5	96.7	100	100	96.7	100	99.6	96.1

Tabla 2.9: Porcentaje de datos correctos disponibles en los años elegidos para la estación de C6-Mogán.

Estación	Periodo de estudio
<b>C0</b>	entrenamiento: ene-dic 2003 test: ene-dic 2005
<b>C1</b>	entrenamiento: ene-dic 2002; ene-jun 2003; jul-dic 2004 test: ene-dic 2005
<b>C2</b>	entrenamiento: ene-oct 2001 test: ene-oct 2002
<b>C4</b>	entrenamiento: ene-dic 2003 test: ene-jun 2000; jul-dic 2005
<b>C5</b>	entrenamiento: ene-dic 2003 test: ene-may 2002; jun-jul 2001; ago-dic 2002
<b>C6</b>	entrenamiento: ene-dic 2002; ene-dic 2003 test: ene-dic 2005

Tabla 2.10: Conjunto de datos utilizados según la estación de medida.

### 2.3.2. Equipos de medida

Los equipos de medida utilizados por las estaciones del I.T.C. son piranómetros. Los piranómetros son instrumentos para medir la radiación solar global (directa más difusa), comúnmente sobre superficies horizontales aunque se pueden instalar inclinados respecto a la superficie terrestre, ver Figura 2.10. Los detectores de estos instrumentos tienen que responder independientemente de la longitud de onda del espectro de la radiación solar y, además, deben responder independientemente del ángulo de incidencia de la radiación solar. Los detectores de la mayoría de los instrumentos están cubiertos por cristales semiesféricos para protegerlos del viento y otros efectos del medio-ambiente; las cubiertas deben ser muy uniformes en grosor para evitar la distribución desigual de la radiación.

El Instituto Tecnológico de Canarias utiliza los piranómetros Kipp & Zonnen CMP-11, con una precisión menor del 3% de error relativo para el total diario de radiación solar. Estos equipos se diseñaron para medir la radiación solar global (flujo de radiación, W/m<sup>2</sup>) en una superficie plana, que resulta de la suma de la radiación solar directa y la radiación solar difusa incidente sobre el hemisferio superior.

Los piranómetros Kipp & Zonnen CMP-11 está provisto de un detector térmico. Este tipo de detectores responden al total de la potencia absorbida y teóricamente no discriminan ningún rango del espectro de la radiación. Esto implica que el detector térmico desnudo es también sensible a la radiación solar infrarroja, de longitud de onda larga ( $\lambda > 3000 \text{ nm}$ ). La radiación es absorbida por un disco negro. El calor generado fluye a través de una resistencia térmica hasta el cuerpo del piranómetro. La diferencia de temperatura a lo largo de la resistencia del disco se convierte en un voltaje. La diferencia de temperatura se ve fácilmente afectada por el viento, la lluvia y la temperatura ambiental. Así que, el detector está protegido por dos semiesferas de cristal. Éstas permiten una igual transmitancia de la componente directa de la radiación solar global para cualquier posición del sol. El rango espectral está limitado por la transmitancia del cristal.

Si se utiliza un anillo para hacer sombra y evitar la radiación directa, el piranómetro



Figura 2.10: Piranómetros Kipp & Zonnen. Cortesía del ITC

mide la radiación difusa, ver Figura 2.11. El anillo permite medir la radiación difusa continuamente sin tener que preocuparse por el mantenimiento del aparato, los ajustes se realizan para cambiar la declinación cada 15 días. El anillo también sombrea parte de la radiación difusa y se tiene que calcular una corrección para esta sombra y aplicar a la radiación difusa observada. Las correcciones se basan en observaciones de la distribución de la radiación difusa en el cielo y habitualmente son factores de 1.05 a 1.2 según la época del año.

### 2.3.3. Tratamiento de los datos de radiación solar

A la hora de realizar cualquier tipo de estudio con datos de radiación solar, un paso esencial es estudiar la calidad de los datos del conjunto disponible. La evaluación de la calidad de los datos es necesaria para impedir que los modelos de predicción elegidos no converjan o conduzcan a resultados poco fiables.

Las estaciones de medida de la isla de Gran Canaria utilizan equipos de medida ampliamente reconocidos y pasan una revisión y calibración periódica. La base de datos cedida por el ITC ya ha sido previamente filtrada, descartando los valores con un código de error negativo.

Además se ha realizado un segundo filtrado de los datos aplicando técnicas reconocidas en la bibliografía [Maxwell93, Younes05, Moradi09, Roesch11]. En este caso, todas las medidas han sido evaluadas siguiendo el método SERI-QC [Maxwell93, Younes05] para bases de datos de radiación global horizontal únicamente. En este caso el filtro se basará



Figura 2.11: Piranómetro con anillo de sombra para eliminar la radiación directa. Cortesía del ITC

en el valor del índice de claridad  $k$ . Se filtran los valores cuyo índice de claridad sea inferior a cero ( $k < 0$ ) o sea superior a uno  $k > 1$ .

Por otro lado, en los procesos de entrenamiento y validación de modelos de predicción de radiación solar, es común trabajar sin las horas de noche. Este filtro se basa en el ángulo cenital de cada medida. El criterio seguido en esta tesis establece la frontera de datos válidos en  $80^\circ$ , a partir del cual se considera que los datos son horas nocturnas. De esta manera se pretende evitar los momentos del atardecer donde los sensores de los piranómetros pueden llevar asociados errores en la lectura y la estación puede estar afectada por el efecto de sombreado (*shadow effects*). El *shadow effect* se ha estudiado para cada una de las estaciones de Gran Canaria, ver Figura 2.12, y se ha considerado  $80^\circ$  como un límite aceptable [Blanc11a, RA10].

En las figuras 2.13 y 2.14 se representan el conjunto de datos de dos años una vez filtrado para las estaciones de C0-Pozo Izquierdo y C5-Sta. Brígida. Los colores cálidos indican los niveles más altos de radiación y los colores fríos los valores más bajos, siendo el azul oscuro el correspondiente a radiación nula o a un dato filtrado. En común en ambas figuras, se puede observar como las noches ocupan las zonas superior e inferior del área (correspondiente a las horas previas al amanecer y posteriores al anochecer). También se puede observar que los días son simétricos respecto al mediodía solar (13 horas de Greenwich en las Islas Canarias), siendo en este momento el instante de mayor radiación. A lo largo del año la longitud de los días varía, siendo más larga durante los meses centrales (verano) que en invierno. En la estación de C0-Pozo izquierdo, en la cara Sur de la isla, se observa una mayor cantidad de datos con radiaciones altas, debido a que en esta zona

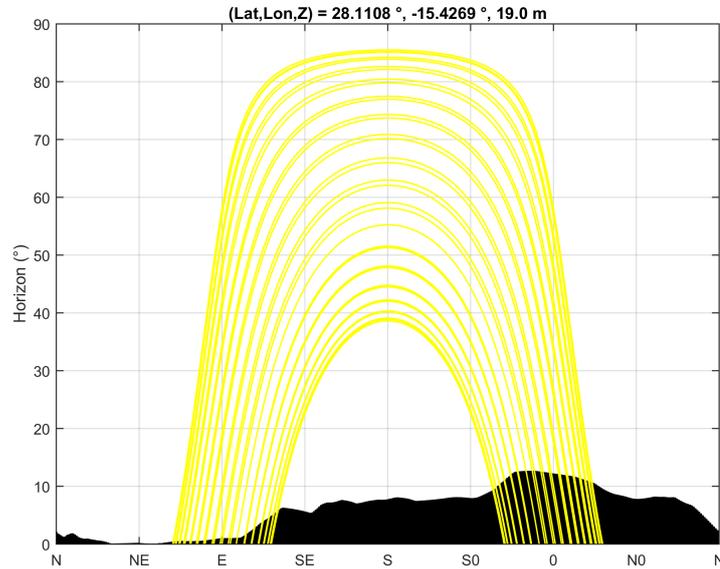


Figura 2.12: Shadow effect sobre la estación de medida de C1-Las Palmas. Cortesía del Mines ParisTech

	C0	C1	C2	C4	C5	C6
Datos iniciales	17520	26280	14592	17544	17520	17520
Datos filtrados	10177	15450	8139	10057	10140	15432
Datos utilizados	7343	10830	6453	7487	7380	10848

Tabla 2.11: Resultados del número de datos filtrados.

el clima es más estable, predominando los cielos despejados.

En la Tabla 2.11 se expone un resumen con el número de datos que se utilizarán en los modelos de predicción para llevar a cabo el entrenamiento y la validación de los mismos.

#### 2.3.4. Análisis de los datos de radiación solar

En este apartado se busca realizar un análisis de la climatología de la isla a partir de los datos horarios de radiación solar de las seis estaciones de medida. La caracterización del comportamiento meteorológico de cada estación permitirá establecer una tendencia general en los niveles de radiación, conociendo por ejemplo el número de días despejados del año, y estudiando la variabilidad que presenta dicho lugar a las condiciones estándar.

Para caracterizar cada estación, la base de datos de radiación solar horaria se clasificó según la media diaria del índice de cielo despejado frente a la variabilidad diaria de los años estudiados [Dambreville14b, Dambreville14a]. La media diaria del índice de cielo despejado se calcula según la ecuación (2.15), donde  $H_d$  se define como el número de horas

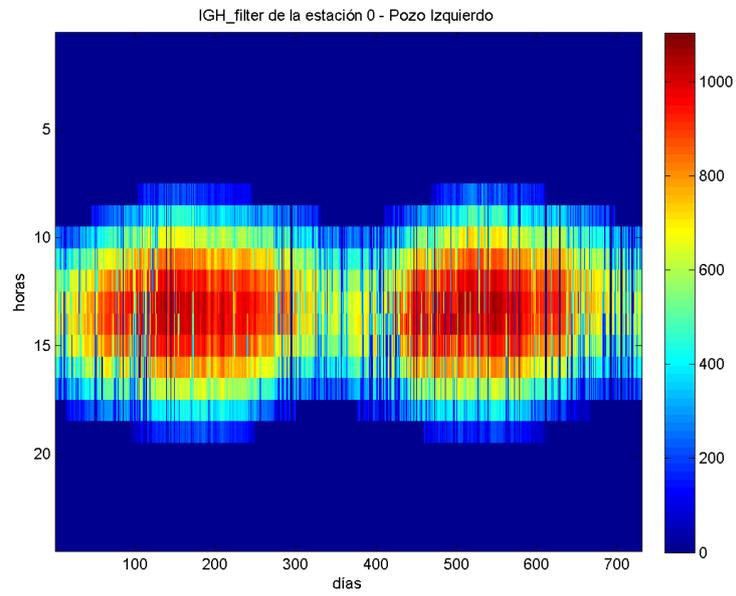


Figura 2.13: Conjunto de datos de la estación C0-Pozo Izquierdo tras realizar el filtro

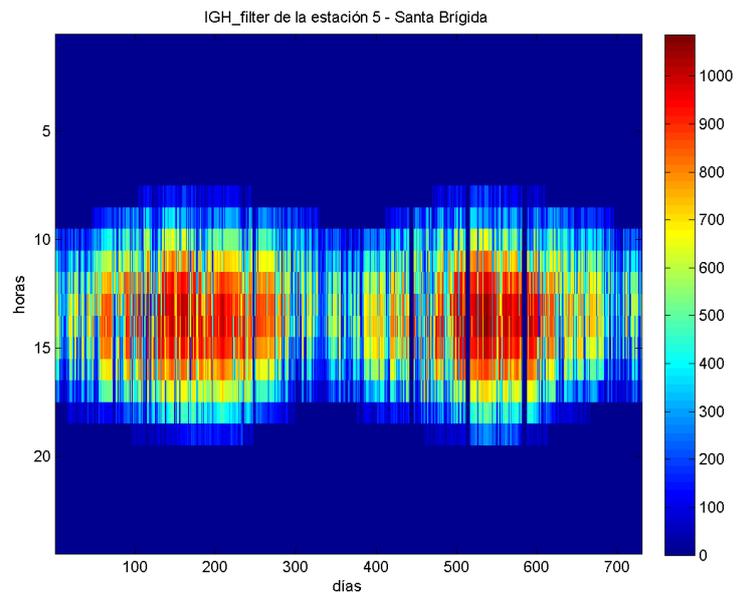


Figura 2.14: Conjunto de datos de la estación C5-Sta. Brígida tras realizar el filtro

Tipo de datos	Condición
A	$\bar{k}^* < 0.5$
B	$0.5 < \bar{k}^* < 0.9$
C	$\bar{k}^* > 0.9$
I	$std(\Delta k^*) < 0.05$
II	$0.05 < std(\Delta k^*) < 0.15$
III	$std(\Delta k^*) > 0.15$

Tabla 2.12: Condiciones a cumplir por cada uno de los subconjuntos de datos.

con un ángulo cenital menor de  $80^\circ$ . La variabilidad se estudiará siguiendo la desviación estándar de la variación hora a hora del índice de cielo despejado, según las ecuaciones (2.16) y (2.17).

$$\bar{k}^* = \frac{\sum_{t=1}^{Hd} k^*(t)}{Hd} \quad (2.15)$$

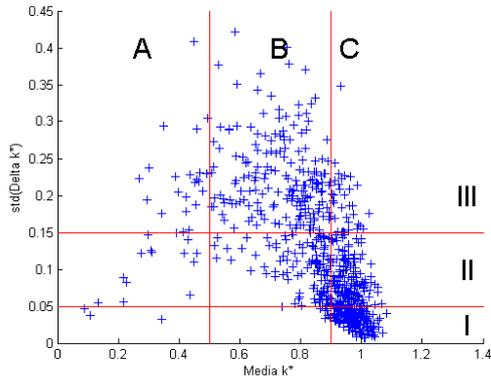
$$std(\Delta k^*) = \sqrt{\frac{\sum_{t=1}^{Hd} (\Delta k^*(t) - \bar{\Delta k^*})^2}{Hd}} \quad (2.16)$$

$$\Delta k^* = k^*(t+1) - k^*(t) \quad (2.17)$$

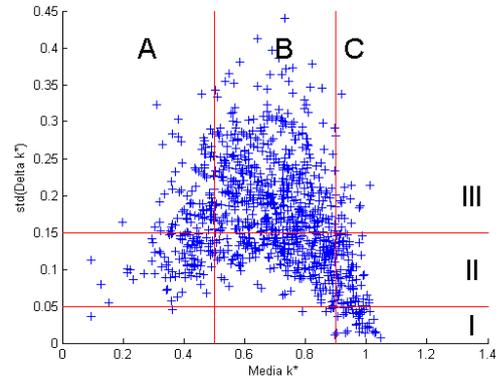
Los datos obtenidos se observan en la Figura 2.15 para las seis estaciones de medida del ITC. Cada punto de las gráficas representa un día concreto, definido por su media diaria del índice de cielo despejado de la horas del día y la variabilidad del mismo día. La tendencia obtenida en todas las estaciones muestra una distribución similar, independiente de la ubicación de la misma. Se observa que la variabilidad es mínima para los días con una alta nubosidad,  $\bar{k}^*$  muestra valores bajos, y para días muy despejados,  $\bar{k}^*$  muestra valores cercanos e incluso superiores a 1. Mientras que los días centrales de la muestra presentan las variaciones más altas de radiación solar durante el día. Por otro lado, se puede observar que la acumulación de la mayoría de los días se concentra en diferentes condiciones según la estación.

Para poder realizar un estudio de las diferencias climáticas de cada estación según la distribución de los días siguiendo este método, se ha realizado una división de los tipos de días según el valor de cada una de las variables. En el eje x, la media diaria del índice de cielo despejado se ha dividido en días tipo A, B o C. Donde los días tipo A presentan medias diarias baja por lo que se consideran días nublados, los días B presentan valores medios y los días tipo C presentan media altas de radiación, por lo que representan los días despejados. En el eje y, la variabilidad la hemos dividido en días tipo I, II o III. Donde los días tipo I presentan una variabilidad bastante baja, por lo que se consideran días con los valores de radiación estables, mientras que los días tipo III presentan variabilidades altas, ver Tabla 2.12 y Figura 2.15.

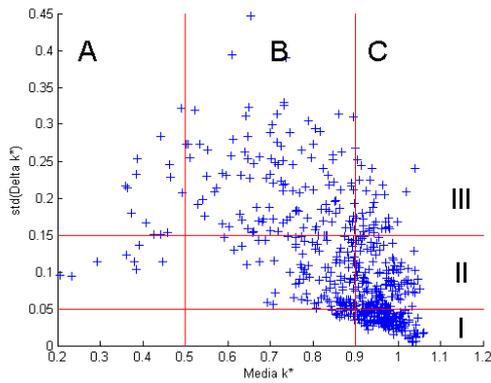
De esta manera, se puede dividir la nube de puntos en nueve tipo de días diferentes según sus condiciones de radiación. Un día muy despejado con una media de radiación



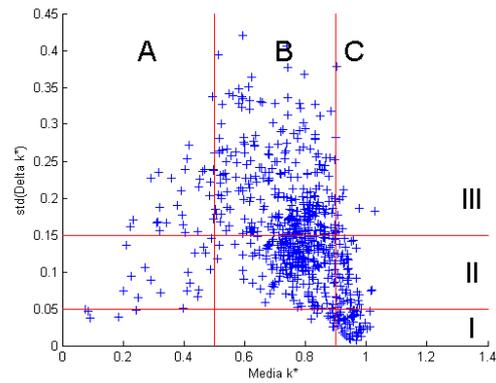
(a) C0



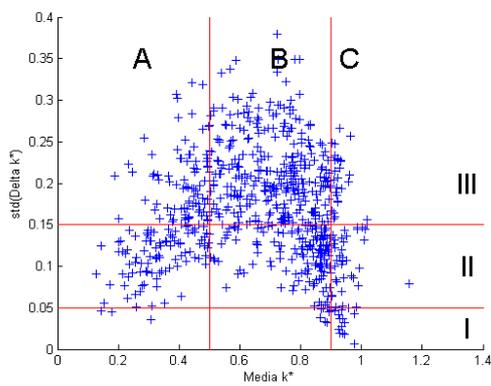
(b) C1



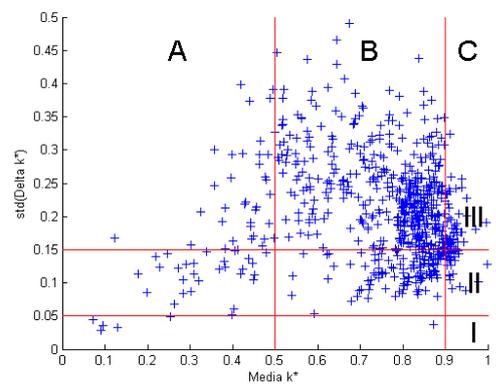
(c) C2



(d) C4



(e) C5



(f) C6

Figura 2.15: Distribución de los datos diarios del índice de cielo despejado para las estaciones de Gran Canaria en función de la media y la variabilidad diaria.

$\overline{k^*}$  alta y una variabilidad baja  $std(\Delta k^*)$  se representa dentro del conjunto de días CI. Por otro lado, un día CIII presentaría una media diaria de radiación solar alta pero con alta variabilidad, mostrando inestabilidad en los niveles de radiación. Los días tipo A, por contra, ofrecen medias del índice de cielo despejado bajas y alejadas de la curva diaria estimada con el modelo de Bird, por la presencia de nubosidad en la zona.

En la Figuras 2.16 y 2.17, se observan, para las estaciones de C0-Pozo Izquierdo y C1-Las Palmas, un ejemplo de nueve días distintos correspondientes a las nueve clasificaciones presentadas. En azul se representa la curva diaria de irradiación estimada con el modelo de cielo despejado de Bird ( $IGH_{cs}$ ), y con línea roja se observa la curva real de datos de irradiación global horizontal ( $IGH$ ) medidos en la estación. Así, por ejemplo en un día CI ambas curvas son muy similares debido a en este conjunto se incluyen los días despejados. Los días CIII muestran medias altas de radiación, por lo que la curva alcanza los niveles máximos de radiación estimados por el modelo de Bird pero aparecen irregularidades en la curva debido a la alta variabilidad. Por otro lado, los días tipo A, independientemente de si tienen o no alta variabilidad, la curva de radiación siempre se encuentra muy por debajo de la estimación de cielo despejado.

Una vez se ha realizado la división del número de días en cada estación, es interesante estudiar la distribución en cada conjunto para establecer las condiciones climáticas de cada estación y conocer de manera general el comportamiento de la isla en su conjunto. Los resultados obtenidos muestran consistencia con la observación empírica del clima en la isla y confirman que la elección arbitraria de los límites de cada tipo de día [Dambreville14a] se puede utilizar en Gran Canaria.

En las Tablas (2.13-2.18), se muestra la división de días según los distintos tipo elegidos en cada estación de medida. En la isla de Gran Canaria existen un mayor número de días despejados en la parte Sur lo que se traduce en un mayor número de días tipo C en las estaciones C0 y C2, mientras que en el resto de estaciones prevalecen los días tipo B. En el norte de la isla, donde se observan una mayor número de días nublados debido al efecto de los vientos Alisios en su encuentro con las montañas, las estaciones C1 y C5 presentan una mayoría de días tipo B y III. Como norma general, en la isla se producen un bajo porcentaje de días tipo I, con baja variabilidad, predominando los días tipo II y III. De la misma manera, en ninguna estación de medida existe un porcentaje significativo de días tipo A con baja radiación.

Total días	A: alta nubosidad	B: nubosidad media	C: baja nubosidad
707	6 %	44 %	50 %
III: alta variabilidad 34 %	3	26	5
II: variabilidad media 40 %	2	17	21
I: baja variabilidad 26 %	1	1	24

Tabla 2.13: Distribución porcentual del número de días de cada tipo para la estación de C0-Pozo Izquierdo.

Esta distribución de días en nueve tipos distintos según sus características climáticas se utilizará posteriormente para comparar los resultados de las predicciones. Los modelos de

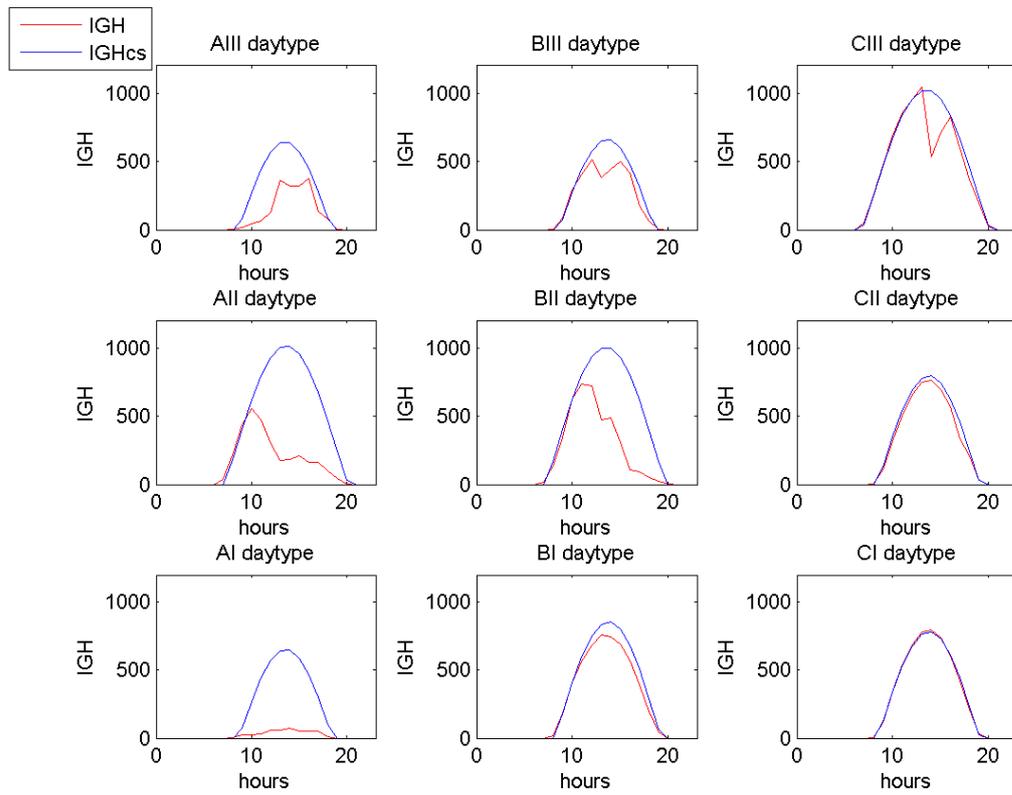


Figura 2.16: Representación de la curva diaria de radiación IGH (rojo) frente a la curva diaria de cielo despejado (azul) para cada uno de los tipos de días en la estación de C0-Pozo Izquierdo

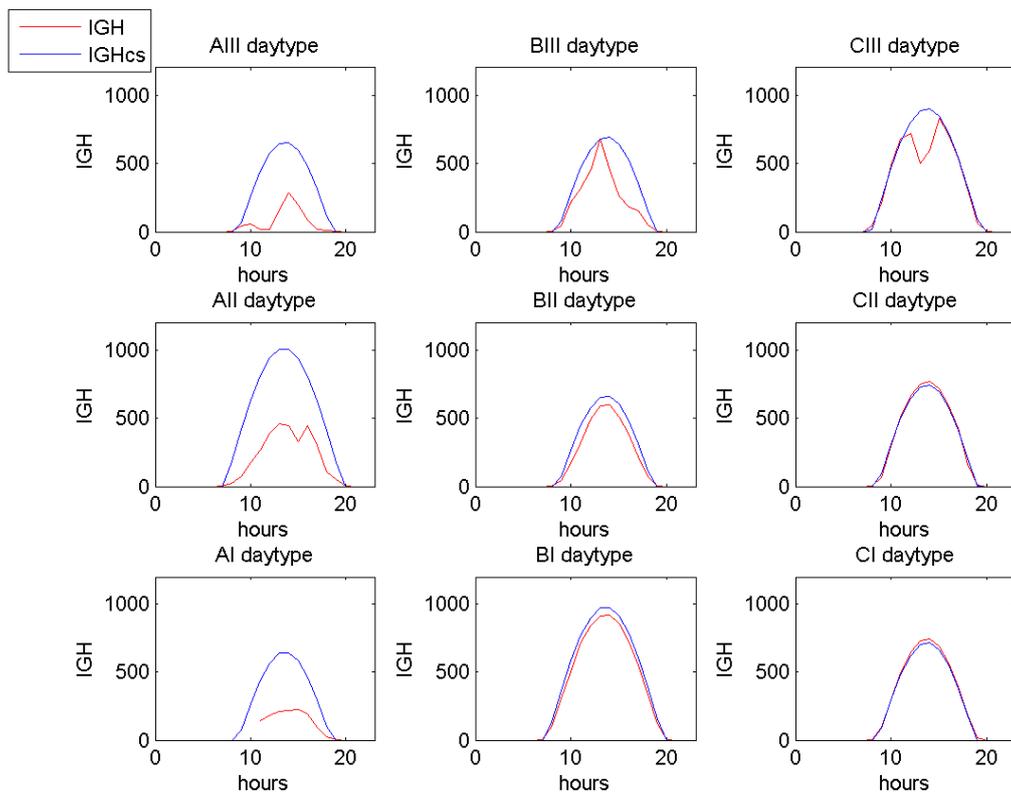


Figura 2.17: Representación de la curva diaria de radiación IGH (rojo) frente a la curva diaria de cielo despejado (azul) para cada uno de los tipos de días en la estación de C1-Las Palmas

Total días	A: alta nubosidad	B: nubosidad media	C: baja nubosidad
1045	19 %	69 %	12 %
III: alta variabilidad 56 %	9	46	2
II: variabilidad media 39 %	9	23	7
I: baja variabilidad 5 %	1	1	3

Tabla 2.14: Distribución porcentual del número de días de cada tipo para la estación de C1-Las Palmas.

Total días	A: alta nubosidad	B: nubosidad media	C: baja nubosidad
600	3 %	40 %	57 %
III: alta variabilidad 27 %	2	19	6
II: variabilidad media 44 %	1	19	24
I: baja variabilidad 29 %	0	2	27

Tabla 2.15: Distribución porcentual del número de días de cada tipo para la estación de C2-La Aldea.

Total días	A: alta nubosidad	B: nubosidad media	C: baja nubosidad
721	7 %	77 %	16 %
III: alta variabilidad 46 %	4	41	1
II: variabilidad media 44 %	2	36	7
I: baja variabilidad 10 %	1	1	8

Tabla 2.16: Distribución porcentual del número de días de cada tipo para la estación de C4-Maspalomas.

Total días	A: alta nubosidad	B: nubosidad media	C: baja nubosidad
713	26 %	63 %	11 %
III: alta variabilidad 58 %	12	42	4
II: variabilidad media 38 %	13	20	5
I: baja variabilidad 4 %	1	1	2

Tabla 2.17: Distribución porcentual del número de días de cada tipo para la estación de C5-Sta. Brígida.

predicción estimarán la radiación para los distintos tipos de días presentados y así poder estudiar los métodos según las condiciones climáticas del año.

Total días 1049	A: alta nubosidad 11 %	B: nubosidad media 76 %	C: baja nubosidad 13 %
III: alta variabilidad 72 %	6	57	9
II: variabilidad media 26 %	4	18	4
I: baja variabilidad 2 %	1	1	0

Tabla 2.18: Distribución porcentual del número de días de cada tipo para la estación de C6-Mogán.

## 2.4. Adquisición de datos de un método numérico de predicción (NWP)

Los modelos numéricos de predicción '*NWP*' (*Numerical Weather Prediction*) se emplean para la predicción del estado de la atmósfera desde horas hasta 15 días de antelación. La predicción de los cambios en la atmósfera, incluyendo la formación y disolución de las nubes, se basan en modelos físicos. Estos modelos físicos se describen mediante ecuaciones diferenciales básicas que se resuelven utilizando de métodos numéricos [Lorenz12, Diagne13].

Los NWP globales calculan el futuro estado de la atmósfera para una determinada superficie, partiendo de unas condiciones iniciales conocidas mediante las agencias meteorológicas. Los modelos NWP globales se encuentran actualmente operados por 15 diferentes agencias meteorológicas mundiales. Como ejemplo tenemos el Global Forecaste System (GFS) utilizado por la US National Oceanic and Atmospheric Administration (NOAA) y el Integrated Forecast System (IFS) operado por la European Centre for Medium-Range Weather Forecasts(ECMWF).

Los modelos globales de predicción tienen normalmente una resolución reducida para determinar las variaciones en una distribución espacial muy localizada. Los NWP distribuyen la superficie terrestre en una cuadrícula cuya precisión varía. Las predicciones de los NWP cubren la superficie con una resolución aproximada desde 0.125 grados hasta 0.5 grados actualmente dependiendo del modelo. La resolución temporal varía de 1 hora para los modelos típicos regionales y de 3 a 6 horas para los modelos globales.

Los cálculos realizados por los modelos NWP globales comienzan basándose en unos datos conocidos del estado de la atmósfera. Los modelos globales obtienen esta información de la red de medidas meteorológicas realizadas por las distintas agencias internacionales. Las variables clave con las que llevan a cabo estos cálculos son la temperatura, humedad, la presión atmosférica, las condiciones espaciales del viento en las tres dimensiones o la temperatura de la superficie oceánica entre otras.

Los modelos NWP globales cubren la superficie completa de la Tierra. Muchos procesos físicos ocurren para escalas espaciales mucho menores que la precisión de los modelos globales, como podrían ser la dispersión y absorción de la radiación o los procesos de condensación y convección que influyen en la formación de nubes. Para mejorar la predicción local, tanto espacial como temporalmente, se han desarrollado modelos locales y regionales que reducen la distribución de la cuadrícula obteniendo superficies de 3 a 10

$km^2$  de precisión para datos horarios, aunque se pueden conseguir resoluciones superiores. Estos modelos sólo cubren una parte de la superficie terrestre y utilizan como datos de partida los datos obtenidos por los modelos NWP globales como condiciones de contorno. Estos modelos son normalmente desarrollados por las distintas agencias nacionales de meteorología o por compañías privadas.

La precisión de estos modelos se está mejorando actualmente utilizando modelos estadísticos de predicción. Estos modelos, que se basan en el aprendizaje del comportamiento de los datos históricos registrados para una localidad, están consiguiendo buenos resultados para diferentes horizontes de predicción combinando los datos históricos con datos provenientes de modelos numéricos de predicción. En general, se están desarrollando mejoras en los modelos de predicción estadísticos combinando datos históricos de medida, modelos de predicción numéricos o datos provenientes de la predicción satélite.

#### 2.4.1. Datos de radiación solar del ECMWF

Como se ha explicado, existen diferentes modelos numéricos de escala global desarrollados por diferentes agencias meteorológicas. En esta tesis se han empleado valores de predicción del Integrated Forecast System (IFS) operado por la European Centre for Medium-Range Weather Forecasts (ECMWF).

Este modelo proporciona las predicciones, incluyendo la irradiancia en la superficie terrestre y diferentes parámetros de las nubes, con una precisión espacial horizontal de alrededor de 16 km x 16 km y 137 niveles verticales de resolución. Los datos de salida están disponibles con una resolución temporal de 3 horas, quedando disponible una resolución horaria para trabajos de investigación. Las predicciones del ECMWF han demostrado su calidad tanto para datos de radiación solar como de condiciones de viento.

En esta tesis se han utilizado datos de predicciones realizadas por el ECMWF para la isla de Gran Canaria gracias al Laboratoire de Physique et Ingénierie Mathématique pour l'Énergie et l'environnement (PIMENT) de la Universidad de la isla de La Reunión. En concreto se obtuvieron datos para una superficie que cubre la isla con un rango de 27.5° a 28.5° de latitud Norte y de 15° a 16° de longitud Oeste y una resolución de 0.125 ° tanto en longitud como en latitud. En la Figura 2.18 se puede observar la distribución de la cuadrícula sobre la isla y la posición en rojo de cada una de las estaciones del ITC.

Para cada elemento de la cuadrícula se disponen de todos los datos de la predicción solicitada para los años entre el 2000 y 2005. Durante estos años el ECMWF únicamente puede ofrecer una resolución temporal de 3 horas. Así, se disponen de las predicciones realizadas por el modelo numérico de predicción ECMWF para la distribución espacial especificada cada tres horas para el día siguiente (horizonte de un día de adelanto). La gestión de las redes eléctricas necesita habitualmente de datos de radiación horaria y además, en esta tesis en concreto, se ha decidido trabajar con datos de radiación horarios. En la bibliografía se encuentran diversas técnicas de interpolación de los datos obtenidos por el ECMWF cada 3 horas para convertirlos en datos horarios. En este trabajo se ha utilizado la técnica descrita por el Proyecto Endorse [Espinar11] en el que se tiene en cuenta la conservación de la energía solar en la integración.

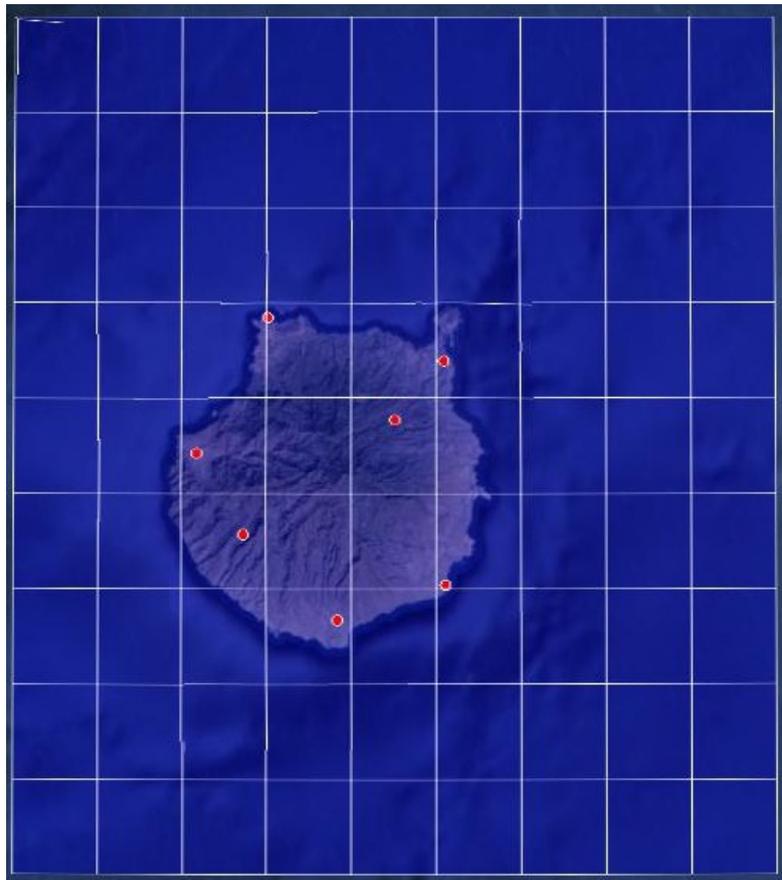


Figura 2.18: Distribución geográfica de los datos obtenidos del modelo de predicción numérica ECMWF para la isla de Gran Canaria.

Una vez descargados los datos predecidos por el ECMWF para cada uno de los elementos de la cuadrícula en los que se encuentren las estaciones de medida del ITC se interpolarán. De esta manera, se dispondrá de una base de datos horaria para los mismos años en los que se disponía de datos terrestres en las estaciones de medida. Las variables que se descargaron para dichas fechas y con las que se trabajará en esta tesis son las siguientes:

- Latitude, ofrece la latitud central de cada elemento de la cuadrícula.
- Longitude, ofrece la longitud central de cada elemento de la cuadrícula.
- Time, marca el momento al que corresponde cada medida.
- TCC Total Cloud Cover, se obtiene el índice de nubosidad entre 0 (sin nubes) y 1 (totalmente nuboso) derivado del estudio de todos los niveles de altitud.
- SSRD Surface Solar Radiation Downwards, en este dato se dispone de la radiación solar horizontal acumulada entre dos instantes sucesivos en  $J/m^2$ .

#### 2.4.2. Datos de viento y humedad del ECMWF

Además de los datos de radiación, en esta tesis se estudiarán los efectos de incluir los datos de velocidad y dirección del viento y la humedad relativa de la atmósfera. En las estaciones de medida no se dispone de estos datos por lo que se optó por incluir las predicciones que realiza el ECMWF de estas variables. Así, de la misma manera que en el caso anterior se obtuvieron las siguientes medidas para la superficie terrestre con idéntica resolución espacial y temporal. En este caso los datos extraídos del programa únicamente están disponibles para el año 2005, por lo que su contribución a la predicción de la radiación solar se estudiará en las estaciones de C0-Pozo Izquierdo y C1-Las Palmas, como estaciones representativas del clima en el sur y norte de la isla respectivamente.

- Latitude, ofrece la latitud central de cada elemento de la cuadrícula.
- Longitude, ofrece la longitud central de cada elemento de la cuadrícula.
- Time, marca el momento al que corresponde cada medida.
- LCC Low Cloud Cover, se obtiene el índice de nubosidad entre 0 (sin nubes) y 1 (totalmente nuboso) derivado del estudio de los niveles entre la superficies y el 0.8 de la presión en la superficie.
- MCC Medium Cloud Cover, se obtiene el índice de nubosidad entre 0 (sin nubes) y 1 (totalmente nuboso) derivado del estudio de los niveles entre el 0.8 y el 0.45 de la presión en la superficie.
- HCC High Cloud Cover, se obtiene el índice de nubosidad entre 0 (sin nubes) y 1 (totalmente nuboso) derivado del estudio de los niveles entre el 0.8 de la presión en la superficie y el nivel más alto estudiado por el modelo.

En este caso, no interesaba únicamente los datos de viento y humedad en la superficie terrestre. Se busca incluir los datos de viento y humedad a distintas altitudes de la atmósfera por lo que se extrajeron los datos del ECMWF para todos los niveles disponibles, obteniéndose diferentes medidas entre las altitudes de 1000-1 hPa de presión.

- R Relative Humidity, se define con respecto a la saturación sobre el hielo por debajo de 23°C y con respecto a la saturación sobre el agua.
- U-velocity, velocidad del viento para la coordenada U del eje de coordenadas.
- V-velocity, velocidad del viento para la coordenada V del eje de coordenadas.

Aunque se realizó la extracción de los datos para todas las altitudes se utilizarán los correspondientes a ciertas altitudes en los modelos estadísticos elegidos. En concreto en esta tesis se decidió incluir los datos de viento y humedad para las siguientes altitudes de la atmósfera:

- Altura de la base de las nubes, como en el ECMWF para el año 2005 no se disponía de este dato, se obtuvo esta altitud como aquella a la cual la humedad relativa es mayor o igual al 95 % [Lave13]. En este caso por lo tanto únicamente se dispone en cada instante de los datos de módulo y dirección del viento.
- Altura de 700 HPa, se obtuvieron los datos de velocidad y dirección del viento y de humedad relativa para esta altitud siguiendo un estudio en la Isla Reunión donde se indicaba esta altitud como la más influyente en la predicción de la radiación solar [Badosa15]. Esta altitud se define como aquella a la cual la humedad relativa presenta una mayor desviación típica.
- Altura de la inversión térmica en las Islas Canarias según el estudio [DA96], se obtuvieron los datos de velocidad y dirección del viento y de humedad relativa para la altitud en que se produce la inversión térmica en cada mes del año.

El ECMWF ofrece los datos del viento según sus coordenadas (u,v) pero en los modelos de predicción estos valores se utilizaron como el módulo y dirección del vector correspondiente.



## Capítulo 3

# Radiación solar obtenida a partir de imágenes satelitales

Los métodos estadísticos de predicción utilizados en esta tesis se basan en el aprendizaje a partir de datos históricos, aprendizaje supervisado. En concreto, se ha propuesto utilizar un modelo basado en *redes neuronales artificiales (RNA)* para predecir la radiación desde 1 hora hasta 6 horas de adelanto. Las RNAs son un método de *machine-learning* capaz de establecer una relación entre un conjunto de datos de entrada y otro de salida, hecho que lo hace muy relevante [Bishop95]. Con las RNAs la radiación solar se trata como una serie temporal estacionaria, por lo que hace falta estimar el índice de cielo despejado para toda la base de datos histórica. En la bibliografía se encuentran numerosas referencias con los resultados de las RNAs en la predicción de la radiación solar únicamente partiendo de datos de radiación solar históricos [Hontoria02, Lauret06a, Mellit10].

Las predicciones basadas en los vectores de movimiento de las nubes (*cloud motion vectors*) obtenidos a partir de imágenes satélites muestran buenos resultados para predicciones hasta 6 h de adelanto [Hammer99, Perez10]. Mientras que los métodos basados en imágenes terrestres del cielo muestran mejores resultados para predicciones con menos de una hora de adelanto [Lorenz12]. Por otro lado, para tiempos de predicción superiores a 6 h los métodos numéricos de predicción (NWP) muestran los mejores resultados [Diagne13].

Los métodos estadísticos constituyen una herramienta fiable y ampliamente probada realizando el aprendizaje con datos históricos previos de radiación solar. En los últimos años han surgido diversas metodologías basadas en la utilización conjunta de datos históricos de radiación medidos en estaciones terrestres y datos de radiación obtenidos a partir de datos satelitales. Estos datos se utilizan como variables de entrada en diversos modelos estadísticos de predicción. Así nos encontramos con modelos autorregresivos (AR) [Dambreville14b, Dambreville14a, Zagouras15], redes neuronales artificiales [Marquez13] y algoritmos genéticos para elegir la información relevante de toda la base de datos del satélite [Zagouras15]. En esta tesis, de forma análoga a dichos autores, se propone mejorar la predicción de la radiación solar obtenida con RNAs y una base de datos históricas, introduciendo información de la superficie terrestre alrededor del punto en cuestión. Esta información se obtendrá a partir de datos de radiación solar en la superficie terrestre

derivados de imágenes satelitales.

La decisión más importante será la elección de la información satelital que es realmente relevante para la predicción en cada localidad. En esta tesis se estudia la importancia de cada elemento de la cuadrícula de datos satelitales (píxeles) utilizando la correlación entre dichos datos y los datos de dicha localidad desfasados en el tiempo [Dambreville14b, Zagouras15]. Así, el análisis de la correlación nos permitirá seleccionar los píxeles más importantes según el desfase entre el dato satelital y el terrestre.

En este capítulo se presentarán los datos de radiación derivados de imágenes satelitales obtenidos para la isla de Gran Canaria. Al igual que para el caso de los datos terrestres, se llevará a cabo un análisis de los datos para establecer las condiciones climáticas de la isla. En la última parte de este capítulo se estudiarán las correlaciones entre los datos terrestres y los datos satelitales para establecer los datos relevantes en cada localidad.

### 3.1. Datos obtenidos del modelo de radiación Heliosat

Los datos satelitales utilizados en esta tesis se obtuvieron de la base de datos del **Helioclim-3**, en particular, de la **versión 5 (HC3v5)**. Los datos de radiación se estiman a partir de las imágenes satelitales del Meteosat. Con el método del Heliosat-2 la información de la radiación solar para el área seleccionada se obtiene procesando dichas imágenes. Los datos han sido cedidos por el Dr. Philippe Blanc, *Responsable des activités de recherche sur l'évaluation des ressources énergétiques renouvelables de MINES ParisTech/ARMINES*.

El método Heliosat fue desarrollado por el *Centre Energétique et Procédés (CEP) de MINES ParisTech/ARMINES*, para transformar la información de las imágenes satelitales obtenidas por el Meteosat en datos de radiación solar en la superficie terrestre. La primera versión [Cano86] se basaba en la relación entre la capa de nubes y la radiación global en la superficie terrestre.

En los años siguientes se introdujeron mejoras paulatinamente hasta llegar a la versión Heliosat-2 [Rigollier04, Blanc11b] en la que se basa la base de datos del Helioclim3. Las mejoras introducidas fueron una nueva formulación de albedos o la utilización de coeficientes de calibración para el sensor entre otras. Esta versión le proporcionó al Helioclim una mejor resolución temporal (15 minutos) y espacial (3 km nadir). Las imágenes se obtienen en tiempo real en la estación de recepción del METEOSAT y se realizan los cálculos cada 15 minutos. Este modelo proporciona datos desde 2004 hasta la actualidad.

Debido a que anteriormente a 2005, la base de datos de Helioclim3 no dispone de un año completo con medidas horarias y que ninguna de nuestras estaciones de medida terrestres disponen de años posteriores a 2005, se solicitó únicamente datos para dicho año en un área alrededor de la isla de Gran Canaria. De esta manera, cuando se trabaje con datos provenientes de imágenes satelitales únicamente estudiaremos las estaciones de C0-Pozo Izquierdo y C1-Las Palmas, Tabla 2.10. Al trabajar con un sólo año de medidas se deberán dividir los datos en dos conjuntos, uno de entrenamiento y otro de validación. La división debe ser adecuada para que ambos conjuntos representen toda la gama de valores de las distintas estaciones del año.

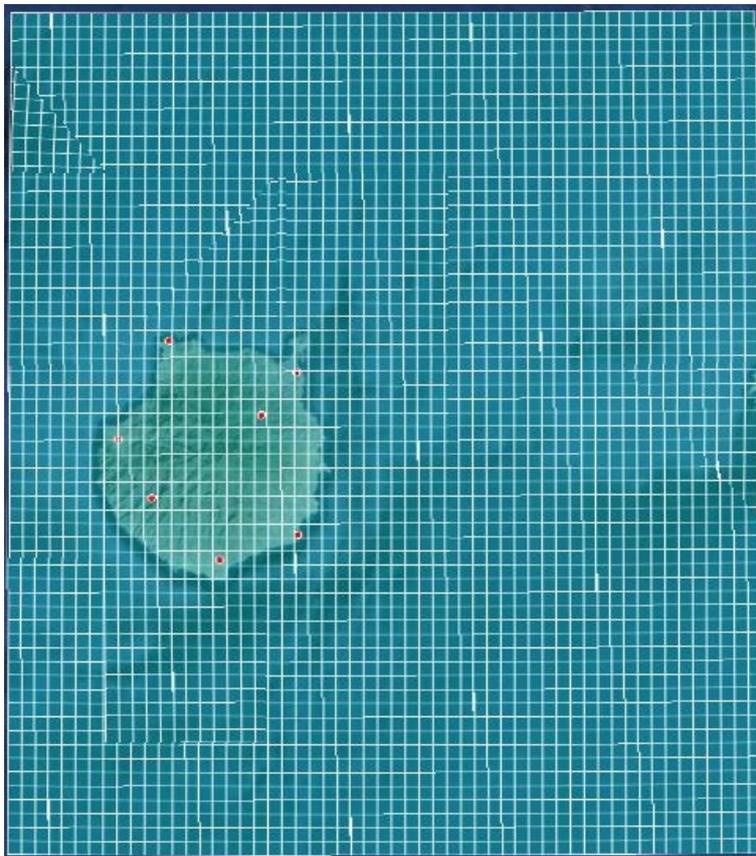


Figura 3.1: Distribución geográfica de los datos obtenidos del modelo basado en imágenes satelital HC3v5 para la isla de Gran Canaria

El área seleccionada incluye la totalidad de la isla de Gran Canaria, además de una superficie significativa en el noreste de la isla para tener información de la zona de la que provienen los vientos predominantes, ver Figura 3.1. Esta decisión está motivada por el conocimiento de la influencia de los vientos Alisios en la formación de nubes de la zona norte de las Islas Canarias. Las coordenadas del área de estudio, en grados, cubren un rango de 27.25 a 28.75 ° de latitud Norte y de 15 a 14.5 ° de longitud Oeste y una resolución de 0.025 ° en latitud y 0.0278 ° en longitud. El área completa se divide por lo tanto en una cuadrícula con 61 x 55 píxeles de información.

La resolución espacial de cada píxel es de 3 x 3 km y cada píxel contiene toda la información solicitada. Esta información incluye la radiación global horizontal en la superficie (GHI), la radiación a cielo despejado en la superficie terrestre horizontal (CLS) y la radiación exoatmosférica (ToA). Todos los datos obtenidos se encuentran en  $Wh/m^2$  y una resolución temporal de 15 minutos. Los datos de las estaciones de medida terrestres tienen una resolución temporal de 1 h, por lo tanto es necesario realizar el cómputo de la radiación horaria a partir de los datos provenientes de las imágenes satelitales para poder realizar la comparación entre ambas bases de datos.

### 3.2. Análisis de los datos de radiación satelitales

La razón principal para utilizar datos satelitales es la posibilidad de disponer de información de una distribución espacial, en lugar de un punto de medida donde se encuentre instalada la estación meteorológica. En los próximos ejemplos se observará que, de esta manera, se podrá estudiar la evolución de la radiación en los alrededores de la localidad elegida.

Para establecer una comparación entre los datos terrestres y satelitales es necesario estudiar si existe congruencia entre ambas series de datos. Por lo tanto, al igual que para los datos terrestres, se realizó la distribución de los distintos tipos de días en base a su media y variabilidad diaria del índice de cielo despejado horario. En las Figuras 3.2 y 3.3 se observa la comparación de la distribución de tipos de días obtenidas en ambas estaciones además de un ajuste de la tendencia general en una curva. Ambas bases de datos presentan una distribución y una tendencia similar, aunque los datos terrestres muestran mayor variabilidad en general en ambas estaciones. Las curvas, cuya regresión se realizó con una polinomial de tercer grado, muestran también un comportamiento muy similar. En las Figuras 3.2 y 3.3 se observa fácilmente que la principal diferencia radica en el máximo de ambas curvas de regresión, es decir, en la variabilidad de ambas muestras. Este fenómeno viene del hecho de que los datos satelital representan un área mayor ( $9km^2$ ) que los datos terrestres, que representan el punto en el que esté instalado el piranómetro. Los datos terrestres se verán afectados en mayor medida por las variaciones meteorológicas locales. En general, ambas muestras presentan una tendencia general muy similar, lo que nos indica que se podría intentar mejorar las predicciones con la utilización conjunta de los mismos.

Debido a la gran cantidad de datos disponibles, cada 15 minutos en una cuadrícula de 61 x 55 píxeles durante todo un año, se creó una interfaz gráfica para visualizar los datos satelitales. Dicha interfaz nos permite estudiar el comportamiento particular de fenómenos

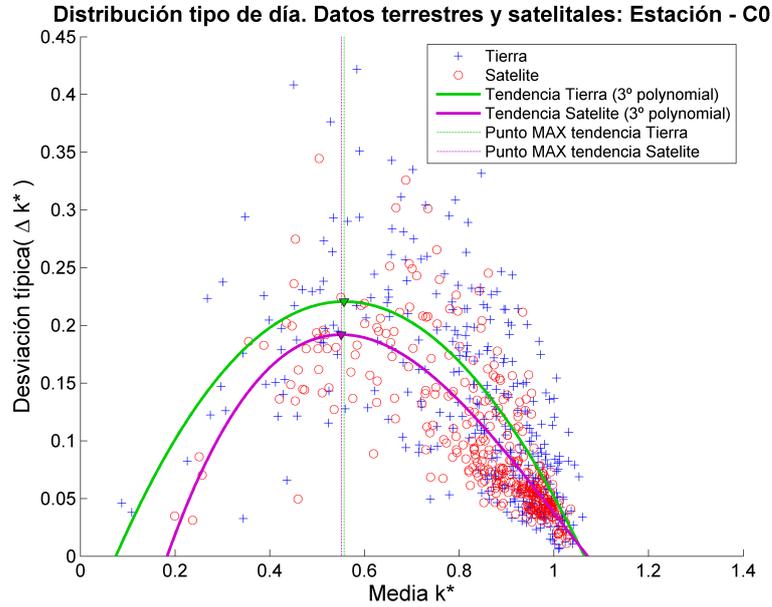


Figura 3.2: Distribución de los tipos de días utilizando tanto datos satelitales como terrestres en base a la media y la variabilidad diaria del índice de cielo despejado para la estación C0-Pozo Izquierdo

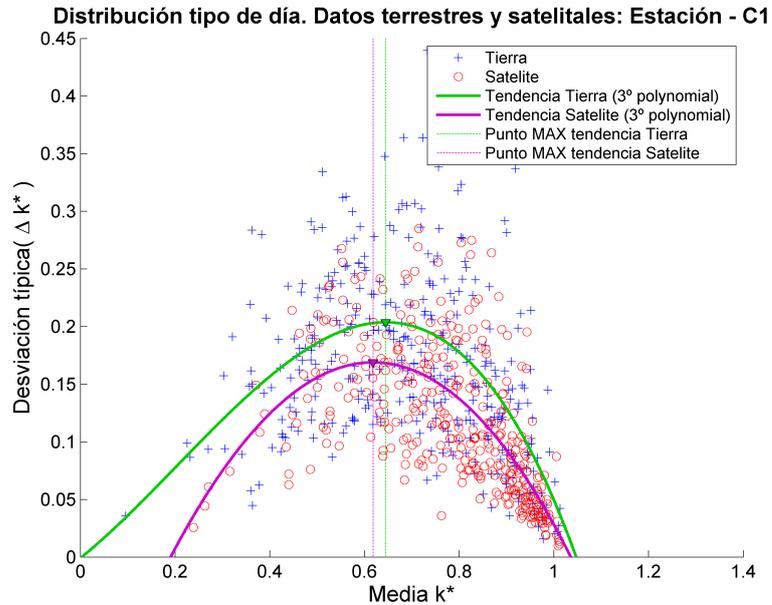


Figura 3.3: Distribución de los tipos de días utilizando tanto datos satelitales como terrestres en base a la media y la variabilidad diaria del índice de cielo despejado para la estación C1-Las Palmas

meteorológicos conocidos como la presencia estática de nubosidad sobre la ciudad de Las Palmas, en el noreste de la isla, durante el verano. Este fenómeno se representa en dos diferentes escenarios, en uno se observa la acumulación de nubes aparentemente debidas al movimiento de los vientos Alisios predominantes, ver Figura 3.4, y en otro se observa la formación local de las nubes sobre la ciudad sin que aparentemente éstas hayan sido provocadas debidas al movimiento, ya que no se observa su presencia durante las primeras horas, ver Figura 3.5.

En las Figuras 3.4 y 3.5, la secuencia superior muestra la evolución durante un día cada 30 minutos del índice de cielo despejado para toda la superficie de datos satelitales disponible. Mientras que la secuencia inferior muestra la secuencia de datos de radiación solar para la misma superficie. La presencia de nubes se puede inferir estudiando la secuencia de datos del índice de cielo despejado. En la Figura 3.4 se puede observar claramente la gran influencia de los vientos Alisios predominantes en la formación de nubes sobre todo el frente norte de la isla. Esta nubosidad permanece estable durante la mañana, incluso con la presencia de fuertes vientos, debido al fenómeno de inversión térmica en conjunto con la orografía escarpada de la zona norte [Díaz12]. Se observa que las nubes siguen un movimiento de norte a sur principalmente debido a los vientos predominantes en la zona donde no se ven interrumpidas por la orografía de la isla.

Por otro lado, en la Figura 3.5, se muestra la creación de nubosidad estática de nubes local sin ninguna aparente contribución de nubes provenientes de la dirección principal de los vientos Alisios. Esta formación de nubes aparecen y crecen gracias a la acumulación de partículas de agua presentes en las corrientes de aire y que permanecen indetectables para las imágenes satelitales. Esta observación nos llevar a pensar que los métodos basados únicamente en el movimiento de las nubes no van a proporcionar toda la información necesaria para predecir la formación de nubes en las zonas montañosas.

Aunque estos dos sucesos aparentemente muestran diferentes condiciones climáticas, ambos son provocados por el mismo fenómeno principal, la inversión térmica en el norte de la isla y los vientos Alisios. Estos ejemplos son un simple vistazo a la complejidad en la predicción meteorológica cuando la disipación, creación y evolución de las masas de vapor de agua están sujetas a gran cantidad de variables.

### 3.3. Análisis espacio temporal de los datos

Los datos satelitales se ha estudiado para utilizarlos, junto con los datos terrestres, a la hora de mejorar las predicciones de radiación solar. Para estudiar la relación con el clima y establecer algún tipo de relación útil entre ambas bases de datos, se optó por utilizar la correlación de Pearson. Con este estudio se pretende establecer qué elementos de la cuadrícula de datos satélites tienen mayor relación con la radiación medida en las dos estaciones terrestres. Hay que recordar que se va a trabajar con las estaciones de C0-Pozo Izquierdo y C1-Las Palmas, por lo que siempre se estudiará por separado la relación de todos los píxeles satelitales con ambas estaciones.

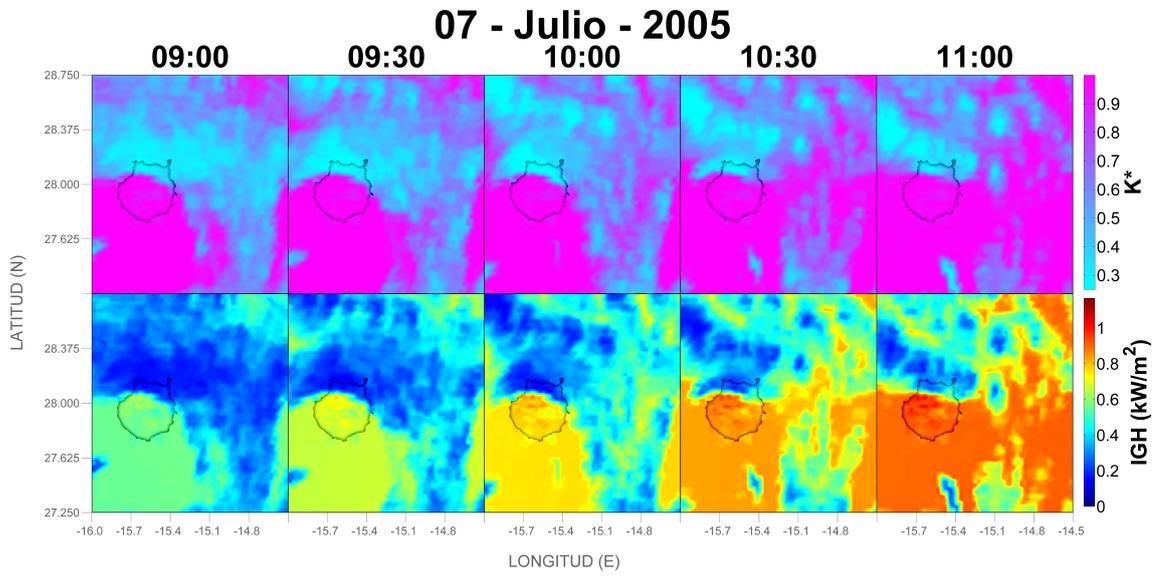


Figura 3.4: Evolución intradiaria de la radiación global horizontal y del índice de cielo despejado de los datos satelitales para cada 30 minutos. Día 07/07/2005.

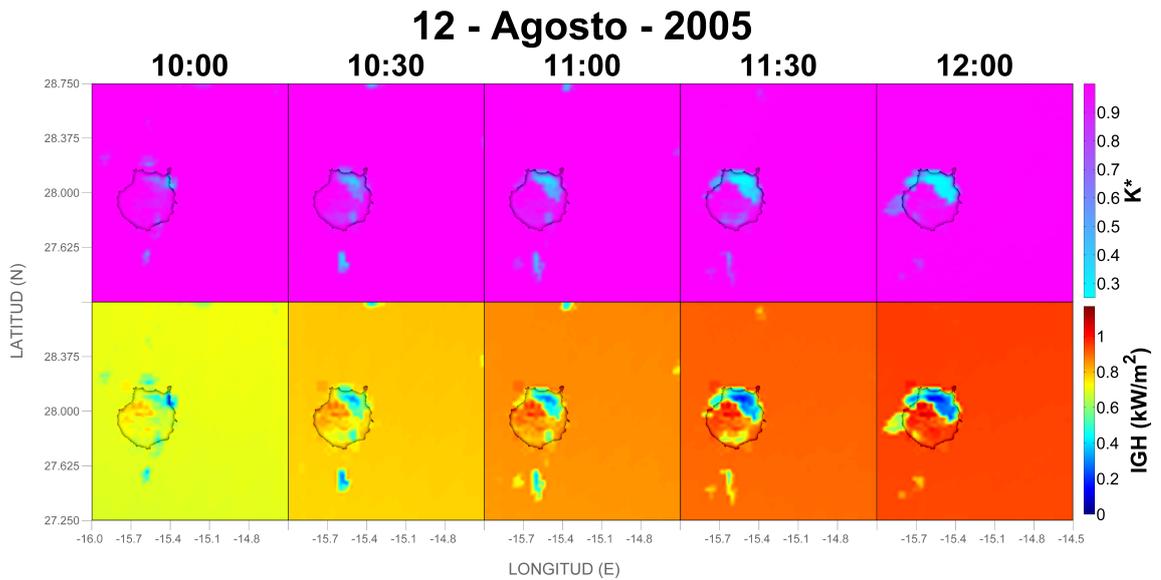


Figura 3.5: Evolución intradiaria de la radiación global horizontal y del índice de cielo despejado de los datos satelitales para cada 30 minutos. Día 12/08/2005.

### 3.3.1. Correlación con el índice de cielo despejado

Primero se realizó la correlación entre las dos bases de datos (datos satelitales y datos terrestres de cada estación) utilizando los índices de cielo despejado para estudiar la relación entre ambos [Dambreville14b]. Para poder evaluar la correlación entre las series temporales en diferentes momentos se establecieron desfases temporales entre ambas series. Los desfases temporales seleccionados se mueven desde  $h = 0$ , ambas series sin ninguna diferencia temporal, y  $h = 3$ , la serie de datos satelitales retrasada 3 horas respecto a los datos terrestres, ecuación (3.1). De esta manera se puede establecer una relación entre los datos satélites en los instantes pasados y el dato medido en la superficie terrestre de la estación objeto de estudio. En cada estación se dispondrá de las correlaciones entre cada píxel de los datos satélite desfasado hasta 3 horas y el dato terrestre correspondiente a la propia estación sin desfasar.

$$C_{k^*}(i, j)_h = \text{corr}(k_{ground}^*(t), k_{satellite}^*(t - h)) \text{ para } h = 0, 1, 2 \text{ \& } 3 \quad (3.1)$$

En las Figuras 3.6 y 3.7 se muestran los resultados de las correlaciones anuales para cada estación. Luego con estas imágenes se pueden deducir los píxeles con una mayor relación con nuestra estación, y que por tanto aportarán una información más relevante para mejorar la predicción. En ambas estaciones se han obtenido resultados que se parecen al comportamiento esperado.

En la estación C1-Las Palmas, Figura 3.7, las correlaciones más altas se obtienen con los píxeles de la zona norte de la isla, mientras que en la estación C0-Pozo Izquierdo la relación es con la zona sur, Figura 3.6. En ambos casos, los valores obtenidos en las correlaciones decaen conforme aumenta el desfase temporal entre los datos satélites y terrestres. Un vistazo más exhaustivo muestra correlaciones inversamente proporcionales entre el norte y el sur, dividiendo la isla en dos sectores. Este factor es crucial para permitir una mejor comprensión del microclima de la zona. Otro fenómeno importante que queda reflejado en las imágenes es que el norte de la isla está influenciado por los vientos Alisios, mientras que las montañas protegen el sur del efecto de los mismos. Todos estas observaciones encajan perfectamente con el conocimiento que se tiene de que el sur de la isla tiene un mayor número de días despejados que el norte.

En las correlaciones anuales la correlación se realiza incluyendo toda la base de datos satelitales y terrestres, con lo que quedan mezclados en los resultados todas las relaciones existentes en las diferentes estaciones del año. Así, para poder observar las posibles diferencias en las relaciones entre los datos satélites en las distintas estaciones del año, se calcularon correlaciones trimestrales independientes [Zagouras15]. Estas correlaciones trimestrales están divididas en los trimestres meteorológicos para dividir los datos según su parecido climático.

Los resultados obtenidos y representados en las imágenes de las correlaciones trimestrales para todos los desfases temporales son coherentes con las condiciones climáticas observadas. La estación más representativa es el verano por la presencia de los vientos Alisios, como se observa en las Figuras 3.8 y 3.9. Sobre todo en las costa este, se puede observar claramente como los vientos Alisios provocan anomalías sobre la superficie ocupada por el océano, mientras que la zona sur de la isla se mantiene protegida

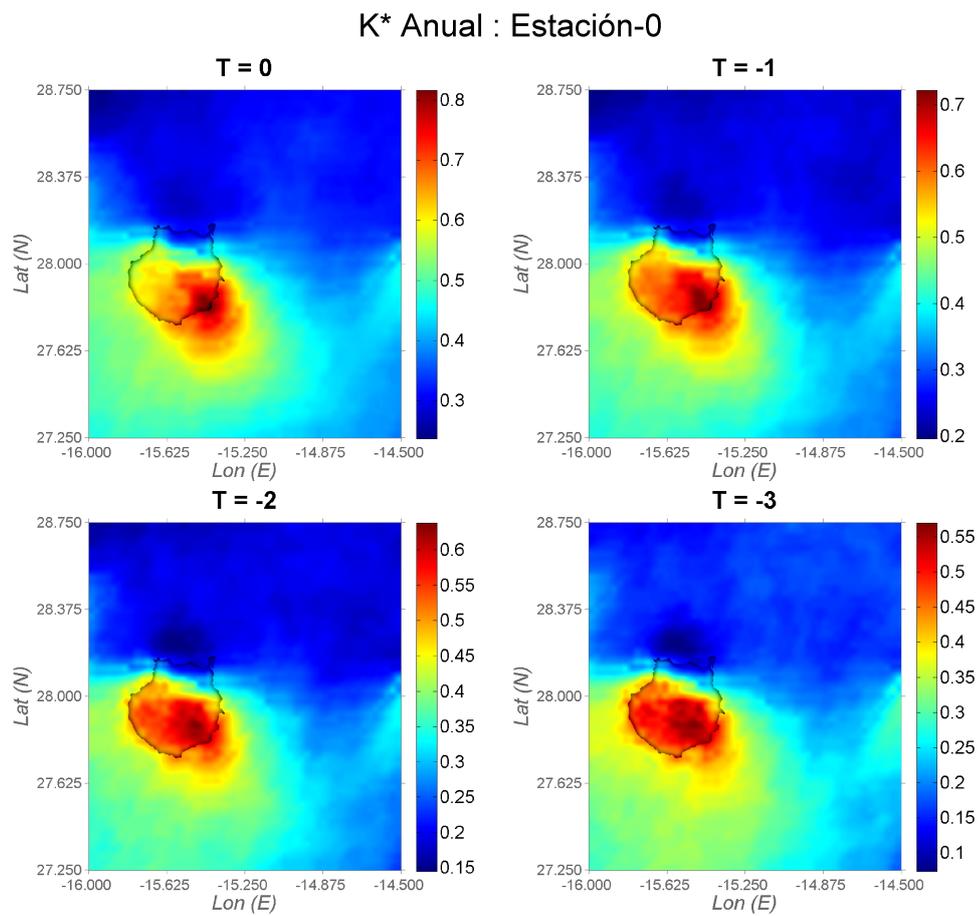


Figura 3.6: Mapa de intercorrelación anual para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de  $h = 0, 1, 2$  &  $3$  horas en la estación C0-Pozo Izquierdo.

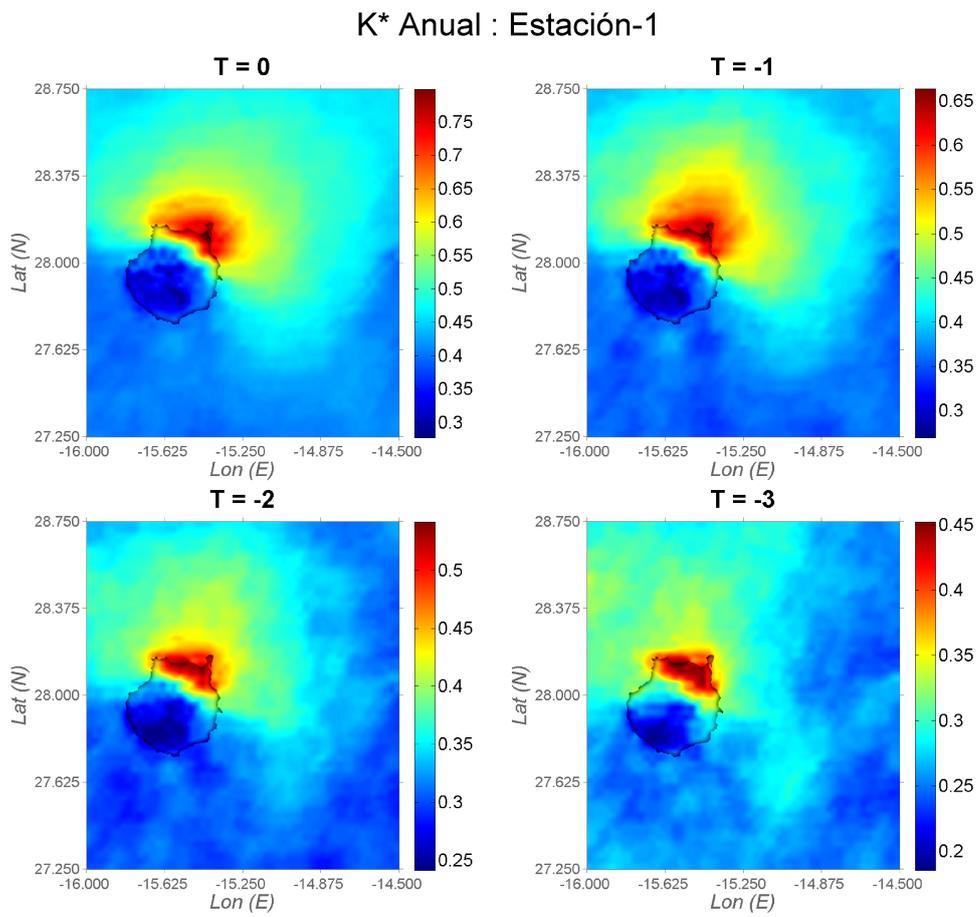


Figura 3.7: Mapa de intercorrelación anual para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de  $h = 0, 1, 2$  &  $3$  horas en la estación C1-Las Palmas.

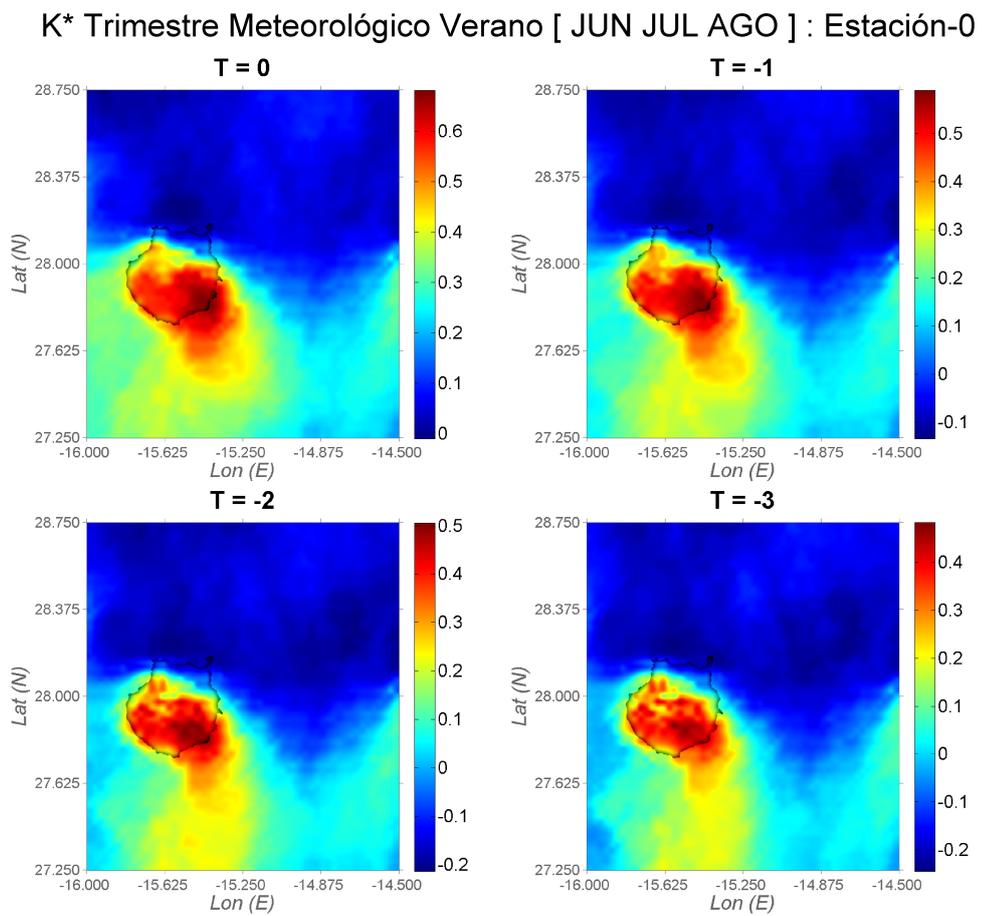


Figura 3.8: Mapa de intercorrelación en verano para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de  $h = 0, 1, 2$  &  $3$  horas en la estación C0-Pozo Izquierdo.

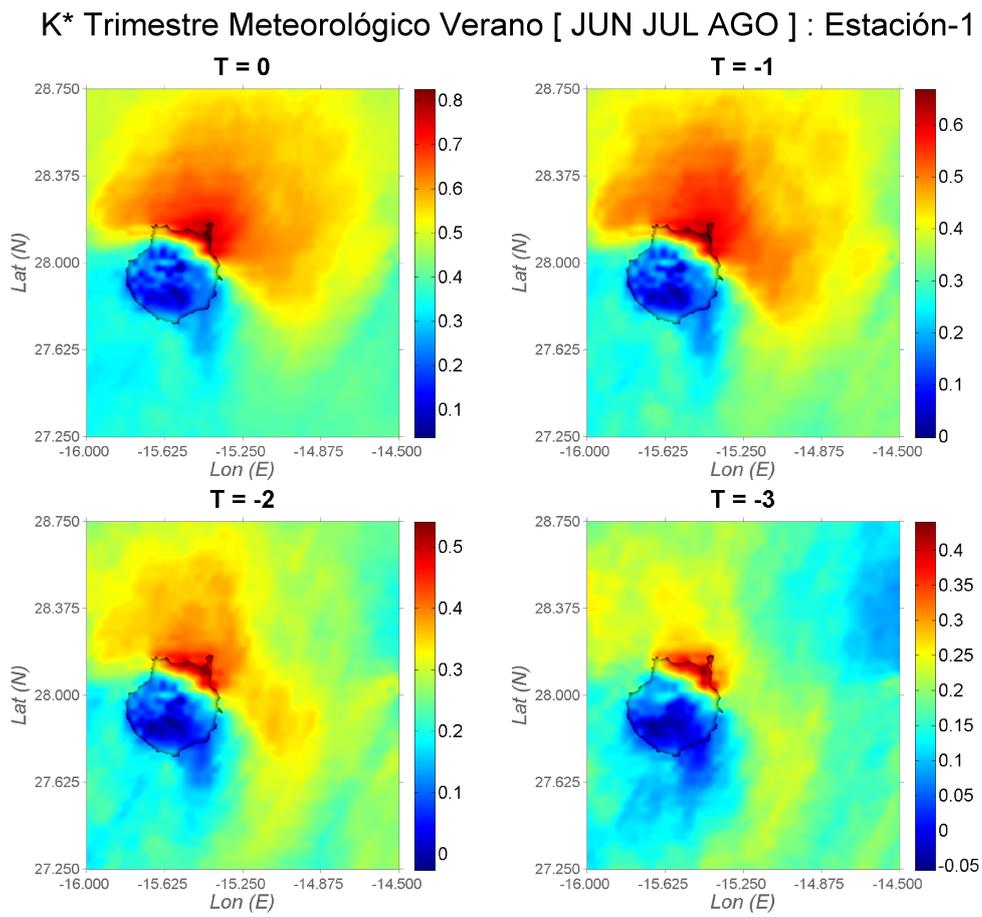


Figura 3.9: Mapa de intercorrelación en verano para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de  $h = 0, 1, 2$  &  $3$  horas en la estación C1-Las Palmas.

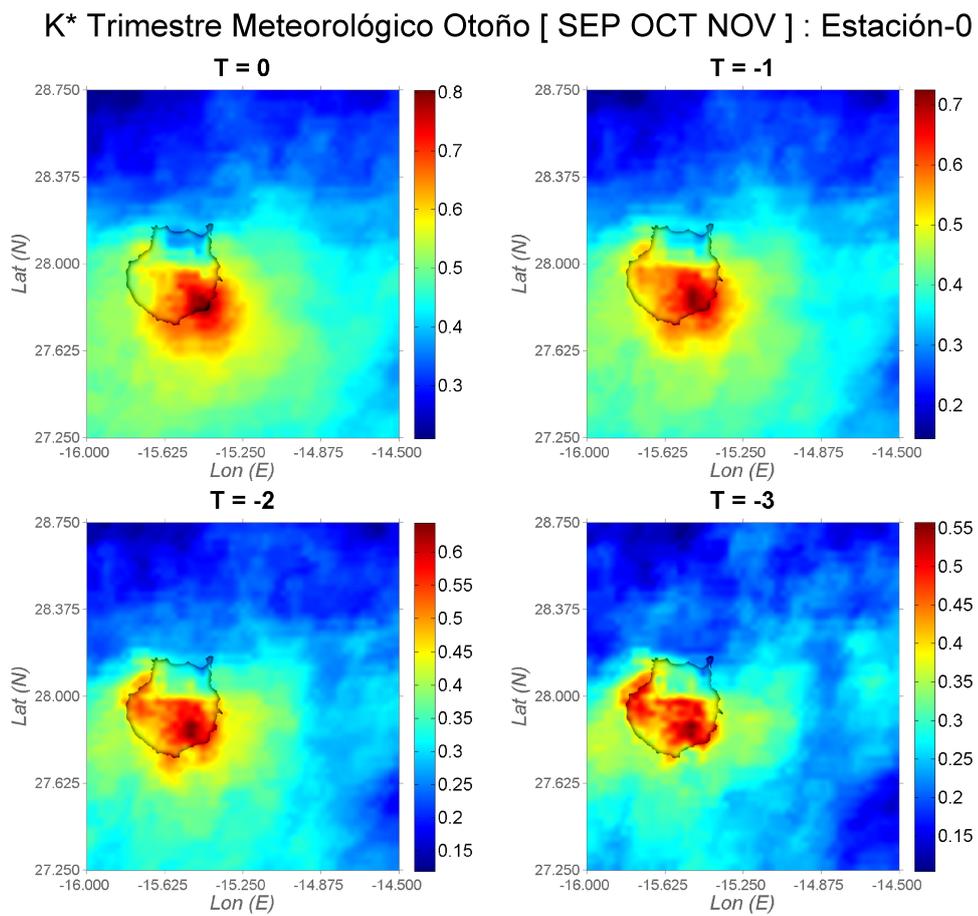


Figura 3.10: Mapa de intercorrelación en otoño para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de  $h = 0, 1, 2$  &  $3$  horas en la estación C0-Pozo Izquierdo.

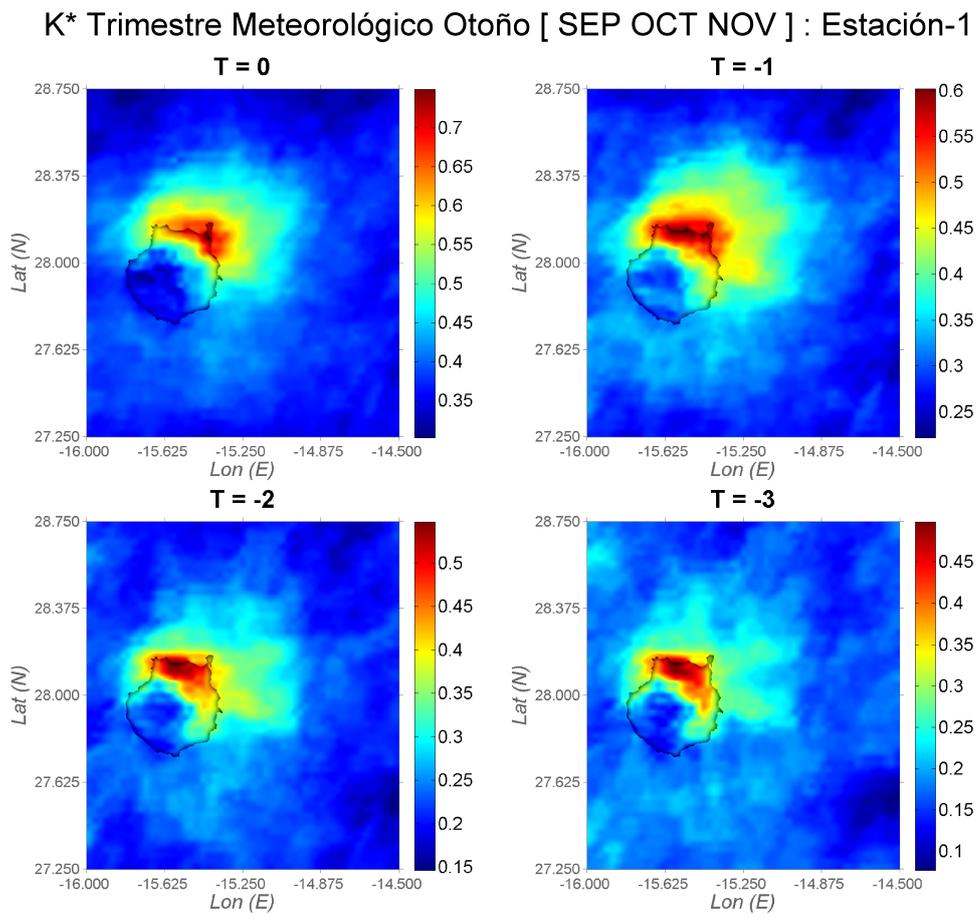


Figura 3.11: Mapa de intercorrelación en otoño para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de  $h = 0, 1, 2$  &  $3$  horas en la estación C1-Las Palmas.

por las montañas. En la Figura 3.8 incluso se puede observar como existe un área de la imagen en la derecha con una relación mayor provocada por el mismo efecto en la isla de Fuerteventura. Por último, se observa claramente, al igual que anualmente, la relación inversamente proporcional entre la estación del sur con los píxeles del norte de la isla y viceversa.

En contraste con las imágenes del verano, se observa en las Figuras 3.10 y 3.11, que representan el otoño, como la superficie de píxeles con mayor correlación se concentra alrededor de las estaciones. Se conserva la división norte-sur de la isla pero no muestra una influencia de los vientos predominantes en los alrededores.

Todos estos análisis conducen a la conclusión de que ambas correlaciones estudiadas, tanto la cuatrimestral como la anual, presentan información adicional al estudio de la radiación. Además del estudio espacial, nos permite tener información de la relación en diferentes horizontes temporales. Al demostrar que existe una relación fuerte entre los datos satélites con distintos desfases temporales y el dato terrestre, nos permite pensar que la inclusión de algunos píxeles satelitales mejorarán la predicción. Este estudio será utilizado en el Capítulo 5 para realizar la elección de los píxeles más relevantes. A diferencia del estudio de Dambreville [Dambreville14b], que selecciona únicamente 9 píxeles independientemente del horizonte temporal, se ha optado por no limitar el número de píxeles y variar el número en cada desfase temporal dependiendo de las correlaciones obtenidas.

### 3.3.2. Correlación con la variación del índice de cielo despejado

De la misma manera que se procedió con las correlaciones entre los índices de cielo despejado de las bases de datos terrestres y satelitales, se intentó conseguir información adicional con una correlación similar utilizando la variación horaria entre los índices de cielo despejado [Dambreville14b]. Esta información debería aportar la dirección de los sucesos climatológicos que están llegando a la isla. Las fórmulas que describen el proceso seguido se observan en las Ecuaciones (3.2) y (3.3). Además, como en el caso descrito en el apartado 3.3.1, también se realizó un análisis anual y trimestral de las bases de datos disponibles para el año 2005.

$$\Delta k^* = k^*(t + 1) - k^*(t) \quad (3.2)$$

$$C_{\Delta k^*}(i, j)_h = \text{corr}(\Delta k_{ground}^*(t), \Delta k_{satellite}^*(t - h)) \text{ para } h = 0, 1, 2 \text{ \& } 3 \quad (3.3)$$

En ambos casos, anual y trimestral, esta correlación basada en la variación del índice de cielo despejado se descartó porque los resultados obtenidos aportaban información errática. Además, esta información no se consideró relevante debido a los bajos valores obtenidos en las correlaciones. Como se observa en la Figura 3.12, para cada desfase temporal, los máximos valores de correlación obtenidos son menores de 0.1. Un análisis más profundo de la información, realizando una aproximación a una distribución normal de los valores para cada desfase temporal, Figura 3.13, se concluye que las curvas obtenidas no ofrecen información relevante ya que todos los valores se concentran muy cerca del cero, excepto

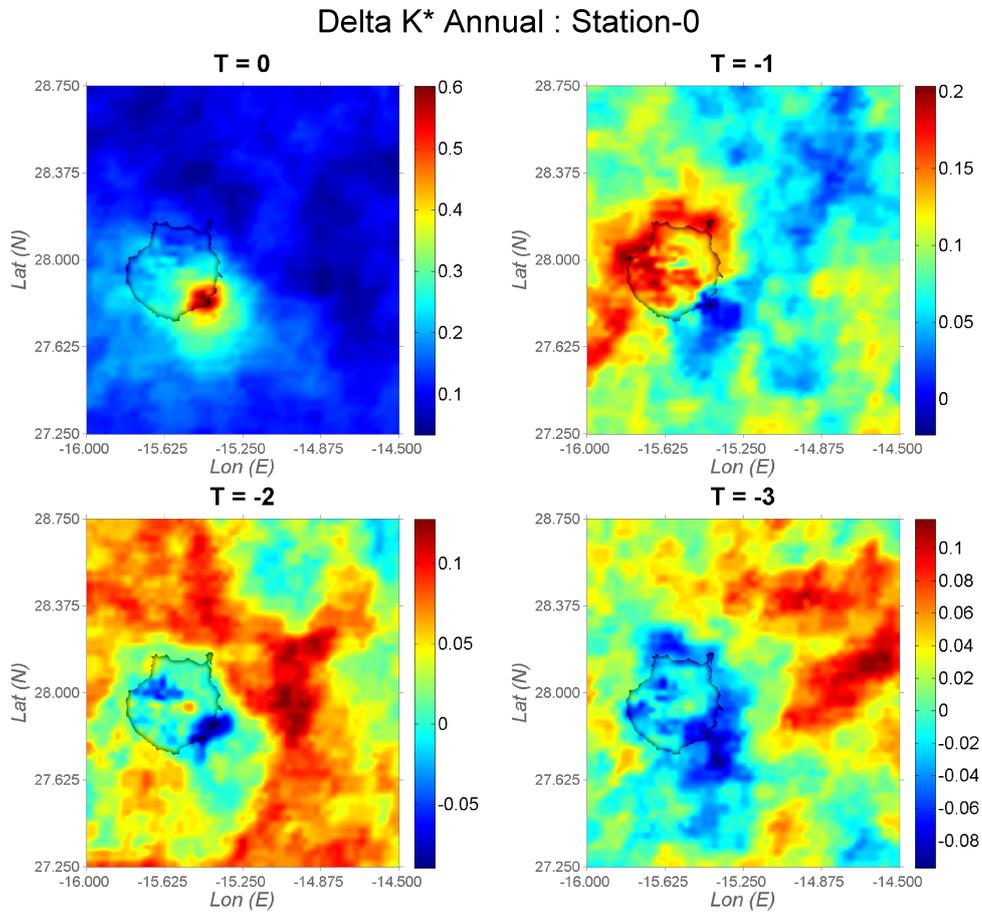


Figura 3.12: Mapa de intercorrelación anual para la variación índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de  $h = 0, 1, 2$  &  $3$  horas en la estación C0-Pozo Izquierdo.

para el caso de desfase temporal  $h = 0$ . En esta tesis se descarta el uso de esta variable porque no ofrece información suficientemente relevante y la elección de los píxeles satélites que se utilizarán para mejorar las predicciones se obtendrán a partir de las correlaciones descritas en el apartado 3.3.1.

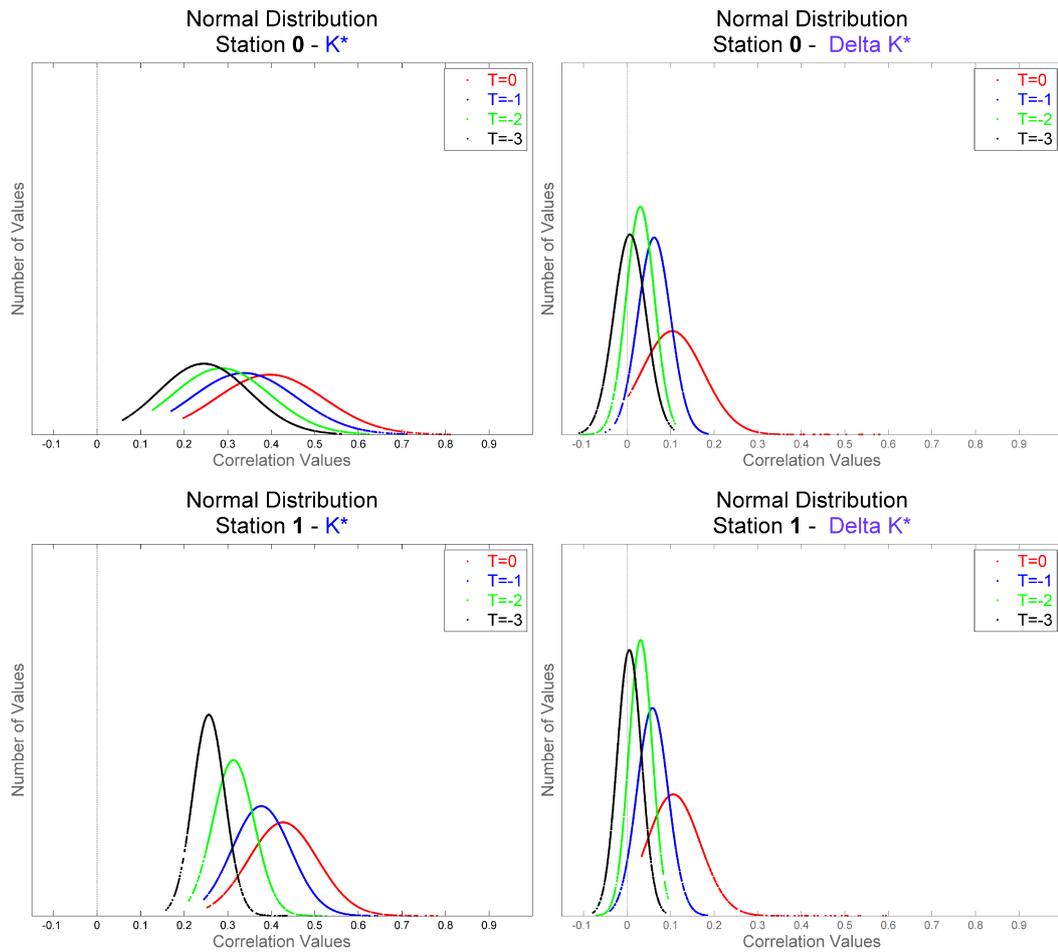


Figura 3.13: Distribución normal de la correlación anual para la variación del índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de  $h = 0, 1, 2$  &  $3$  horas en la estación C0-Pozo Izquierdo.



## Capítulo 4

# Modelos de predicción de radiación solar

### 4.1. Introducción

El objetivo de esta tesis es el estudio de distintos métodos de predicción de la radiación solar global horizontal en el horizonte temporal horario, de 1 a 6 horas en adelante. Como ya se ha comentado, según el horizonte temporal de predicción existen diversas técnicas en constante desarrollo. Los modelos estadísticos, como los métodos Autorregresivos (AR), los métodos Autorregresivos de Media Móviles (ARMA) y las Redes Neuronales Artificiales (RNAs), son los más apropiados para horizontes de predicción entre 5 minutos y 6 h [Lorenz12].

En esta tesis se utilizarán, como modelos principales, diferentes métodos estadísticos ampliamente reconocidos en la bibliografía especializada. En un principio, estos modelos estadísticos utilizarán únicamente datos de las series temporales de datos de radiación global horizontal terrestre. Entre los métodos utilizados nos encontramos con métodos *naïve*, métodos estadísticos lineales y métodos no lineales, como las Redes Neuronales Artificiales, **RNAs**. Los métodos lineales y las RNAs son capaces de *descubrir y aprender* secuencias repetitivas dentro de la serie histórica de datos terrestres. Posteriormente, con el objetivo de mejorar las predicciones, en este trabajo se utilizarán datos provenientes de métodos numéricos de predicción [Lauret15], en concreto del European Centre for Medium-Range Weather Forecasts, **ECMWF**, y datos de radiación derivados de imágenes satelitales, en concreto del HelioClim3 [Zagouras15, Dambreville14b].

Tanto si los modelos de predicción utilizados se basan en datos de radiación terrestre únicamente o también en datos exógenos, los cálculos se realizarán con el índice de cielo despejado,  $k^*$ . Cuando se hayan realizado los procesos de entrenamiento de cada uno de los modelos, el cálculo y discusión de los errores se llevará a cabo en términos de irradiancias horarias. La radiación solar terrestre obtenida en la predicción  $\hat{I}_g$  se obtiene según la ecuación (4.1). La función general que relaciona los parámetros de entrada y salida es la ecuación (4.2),

$$\hat{I}_g = (\hat{k}^*) \cdot I_{cs} \quad (4.1)$$

$$\hat{k}^*(t+h) = F [k_g^*(t), \dots, k_g^*(t-i), k_{e1}^*(t), \dots, k_{e1}^*(t-j), \dots, k_{en}^*(t), \dots, k_{en}^*(t-j)] \quad (4.2)$$

donde  $\hat{k}^*(t+h)$  es el índice de cielo despejado calculado para un horizonte de predicción  $h$ , en nuestro caso  $h = 1, 2, \dots, 6$  horas,  $k_g^*(t-i)$  es el índice de cielo despejado para las  $i$  horas pasadas obtenido a partir de los datos terrestres medidos en la estación y  $k_{en}^*(t-j)$  corresponde índice de cielo despejado para las  $j$  horas pasadas obtenido a partir del dato exógeno  $en$ . Una de las decisiones más importantes a tener en cuenta es la adecuada elección del número de datos pasados que se utilizarán, ya que corresponderán a las entradas de los modelos de predicción. La utilización de datos irrelevantes como entradas aumentará la complejidad del modelo y, como consecuencia, podría afectar a su eficacia.

En cada caso, la función de transferencia  $F$  se definirá durante el proceso de entrenamiento. Esto supone disponer de un grupo de datos de entrenamiento con un número  $N$  de pares de datos de entrada y salida,  $D = [X_i, t_i]_{i=1}^N$ . En nuestro caso, la matriz de datos de entrada,  $X_i$ , contiene todas las variables utilizadas, tanto los índices de cielo despejado terrestres como los provenientes de datos exógenos, por cada par de datos  $N$ . Por otro lado, el vector  $t_i$  contiene el índice de cielo despejado medido en las diferentes estaciones para el tiempo horizonte de predicción, es decir, el valor de radiación que queremos reproducir. Por lo tanto, durante el proceso de entrenamiento la función  $F$  tratará de reproducir con el mayor ajuste posible las salidas medidas  $t_i$  a partir de las entradas  $X_i$ .

## 4.2. Modelos estadísticos de referencia

El objetivo al desarrollar un método sofisticado de predicción es mejorar de forma importante la calidad del ajuste realizado por los llamados métodos *naïve o de referencia*. En esta tesis, nos hemos planteado comparar las predicciones realizadas por los diferentes métodos estadísticos con dos modelos de referencia. Dichos modelos serán el **Persistente**, **Persistence** en inglés, y el modelo **Climatológico**. Estos modelos únicamente utilizan las series temporales históricas de radiación solar como variables de entrada y su utilidad ha sido ampliamente descrita en la bibliografía [Lauret15, Diagne13, Lorenz12, Sengupta15, Beyer09].

### 4.2.1. Modelo de predicción Persistente

En esta tesis se han probado dos métodos *naïve* de predicción de la radiación solar global horizontal para todos los horizontes temporales de predicción. El primer método planteado es el modelo *Persistence (Pers)* [Lauret15, Beyer09, Sengupta15], definido por la ecuación (4.3), ya que es el método más comúnmente utilizado para la predicción a corto plazo. Este modelo asume que el índice de cielo despejado para cada horizonte

temporal  $h$  sólo depende del valor en el instante  $t$ , lo cual implica que las condiciones meteorológicas se mantengan invariantes entre los tiempos  $t$  y  $t + h$ .

$$\widehat{k}^*(t + h) = k^*(t) \quad (4.3)$$

En esta tesis, acorde con la bibliografía, se utilizará el modelo *Persistence* como referencia para situar las posibilidades de mejora del resto de modelos que se prueben. Generalmente es un modelo que ajusta muy bien la predicción de la hora siguiente, pero su eficiencia disminuye muy rápidamente conforme se incrementa el horizonte temporal de predicción. Por otra parte, consigue muy buenos resultados en días despejados, mientras que pierde precisión en los días nublados.

#### 4.2.2. Modelo de predicción Smart-Persistence

El siguiente modelo presenta una sencilla mejora respecto del modelo Persistente, que consiste en utilizar la media de los  $h$  valores anteriores del índice de cielo despejado para predecir el índice de cielo despejado para el horizonte temporal  $t + h$  [Hoff12]. El modelo, conocido por su nombre en inglés *Smart-Persistence* (*Smart-Pers*), queda definido por la ecuación (4.4).

$$\widehat{k}^*(t + h) = \text{mean}[k^*(t), \dots, k^*(t - h)] \quad (4.4)$$

Con este modelo, como se podrá comprobar en los resultados para todas las estaciones de la isla de Gran Canaria, se consigue una clara mejora para los horizontes temporales de predicción superiores, sobre todo a partir de la cuarta hora. Esto es debido a que para la predicción a una hora no existe diferencia alguna respecto al *Persistence*, mientras que a medida que se aumenta el horizonte temporal se añade información de la serie temporal de valores anteriores.

#### 4.2.3. Modelo de predicción Climatológico

El último modelo de referencia será el *Climatológico* [Lorenz12], el cual predice el valor del índice de cielo despejado para cualquier horizonte temporal  $h$  según el valor obtenido al realizar la media de todos los valores históricos anteriores de la serie temporal (4.5). En este caso, el grupo de valores históricos corresponderá al conjunto de valores de entrenamiento.

$$\widehat{k}^*(t + h) = \text{mean}[k^*(t)]_{\text{training data}} \quad (4.5)$$

Por la propia definición y formulación matemática, este modelo proporcionará un valor constante de índice de cielo despejado para todos los instantes de tiempo, cuyo valor representa el valor climático más esperado en cada una de las estaciones. La precisión, por tanto, de este tipo de modelos se mantendrá también constante para cualquier horizonte temporal de predicción. Los modelos estadísticos de predicción tenderán a igualar la precisión del *Climatológico* conforme aumente el horizonte temporal. En esta tesis se tratará de mejorar este nivel de ajuste con las técnicas de Machine Learning elegidas.

### 4.3. Modelos lineales

La radiación solar está considerada un proceso estocástico, cuyos niveles esperados dependen de los valores anteriores de la serie temporal y de las variaciones meteorológicas que se puedan producir en la zona de estudio, en un rango temporal que puede ir desde los segundos hasta los días. Las series temporales que describen fenómenos físicos se pueden definir como estacionarias o no estacionarias. Las series estacionarias se mantienen constantes respecto a su tendencia general y las fluctuaciones pueden aparecer de manera completamente aleatoria. Como ya se ha explicado, las series temporales de radiación solar se transforman en una serie estacionaria dividiendo por el valor de radiación estimado por un modelo de cielo despejado [Diagne13].

Los modelos estadísticos lineales han sido utilizados en el modelado de series temporales durante años en diversos campos, como economía, demografía, meteorología o medio ambiente. Entre los modelos lineales más utilizados nos encontramos:

- *Modelos Autorregresivos (AR)*: Se trata de un modelo de regresión lineal basado únicamente en datos pasados de la serie temporal que se estudia. El orden de modelo viene definido por la variable  $p$ , que representa el número de parámetros de los que consta la función de transferencia.
- *Modelos de medias móviles (MA)*: En este caso el orden de modelo viene definido por la variable  $q$ , que representa el número de parámetros de los que consta la función de transferencia.
- *Modelo Autorregresivos de medias móviles (ARMA)*: Se trata de un modelo formado por dos partes, una basada en el modelo Autorregresivo (AR) y otra basada en el modelo de medias móviles (MA). La complejidad del modelo vendrá definida por los parámetros  $p$  y  $q$ , que representan el orden de la parte (AR) y (MA) respectivamente.
- *Modelo Autorregresivos Integrados de media móviles (ARIMA)*: Se trata de un modelo estadístico utilizado para encontrar patrones para una predicción futura a partir de variaciones y regresiones de datos estadísticos. La complejidad del modelo vendrá definida por los parámetros  $p$ ,  $d$  y  $q$ , que representan el orden de la parte (AR), la parte de integración (I) y (MA) respectivamente.
- *Modelo lineales con datos exógenos*: En este caso los modelos utilizan, además de los datos pasados de la serie temporal, otros provenientes de series temporales exógenas a la que se pretende modelar. Se encuentran técnicas autorregresivas (ARX), Autorregresivas de medias móviles (ARMAX) o Autorregresivas Integradas de medias móviles (ARIMAX).

Estas técnicas han sido ampliamente utilizadas desde los años setenta en el campo de la radiación solar. Los modelos de series lineales se han utilizado para generar series sintéticas de valores o para definir Años Meteorológicos Típicos [Mazorra10] y para realizar predicciones de radiación solar. Aguiar & Collares-Pereira [CP89] realizaron trabajos con modelos ARMA a partir de datos del índice de claridad horario y J. Boland

[Boland95, Boland08] realizó una descripción de la radiación solar diaria y horaria, un estudio del modelado de series temporales de radiación solar y un método de estimación de la radiación solar difusa en Australia. En esta tesis, como se podrá ver en el capítulo 5, se han evaluado dos modelos lineales basados únicamente en valores pasados, el modelo (AR) y el modelo (ARMA).

#### 4.3.1. Modelo lineal Autorregresivo AR

En un modelo Autorregresivo (AR) [Chatfield13], la predicción de un valor de la serie temporal en un horizonte temporal  $h$ , se asume vendrá representada por una combinación lineal de valores pasados de la propia serie, según la ecuación (4.6),

$$\widehat{k}^*(t+h) = \sum_{i=0}^{p-1} [\Phi_{i+1} k^*(t-i)] + \epsilon_{t+h} \quad (4.6)$$

donde  $\widehat{k}^*(t+h)$  representa el valor de radiación en el horizonte temporal  $h$  y  $\epsilon_t$  es un ruido blanco con varianza  $\sigma^2$ . Los valores  $k^*(t-i)$  son los datos pasados de la serie temporal que se han elegido para establecer una relación lineal con el valor a predecir, mientras que los parámetros del modelo son  $p$ , que representa el orden de AR o, lo que es lo mismo, el número de valores pasados utilizados, y  $\{\Phi_i\}_{i=1,2,\dots,p}$ , que muestra los parámetros de autorregresión obtenidos a partir de los datos de la muestra durante el entrenamiento. Para optimizar el modelo AR, una de las claves del desarrollo del mismo es determinar el número de orden del mismo,  $p$ . En esta tesis, se proponen métodos basados en el estudio de la muestra de Funciones de Autocorrelación Parcial [Boland95, Boland08], *PACF* por sus siglas en inglés (*Partial Autocorrelation Function*), y el Criterio de Información Bayesiana, *BIC* por sus siglas en inglés (*Bayesian Information Criterion*).

#### 4.3.2. Modelo lineal autorregresivo de medias móviles ARMA

En los modelos Autorregresivos de medias móviles ARMA, los valores futuros de la serie temporal están basados en dos modelos básicos, un modelo Autorregresivo (AR) y un modelo de medias móviles (MA), como una combinación lineal de una cantidad determinada de valores pasados de la serie y de errores, según la ecuación (4.7)

$$\widehat{k}^*(t+h) = \sum_{i=0}^{p-1} [\Phi_{i+1} k^*(t-i)] + \epsilon_{t+h} + \sum_{j=0}^{q-1} [\Theta_{j+1} \epsilon_{t-j}] \quad (4.7)$$

donde  $\widehat{k}^*(t+h)$  representa el valor del índice de cielo despejado en el horizonte temporal  $h$ . Los valores  $k^*(t-i)$  son los datos pasados de la serie temporal que se han elegido para establecer una relación lineal con el valor a predecir y  $\epsilon_t$  es una serie de ruido blanco con media cero y varianza  $\sigma^2$ . En este caso, los parámetros del modelo son  $p$ , que representa el orden de AR o, lo que es lo mismo, el número de valores pasados utilizados,  $\{\Phi_i\}_{i=1,2,\dots,p}$ , que muestra los parámetros de autorregresión obtenidos a partir de los datos de la muestra durante el entrenamiento,  $q$ , que indica orden del modelo MA,

y  $\{\Theta_i\}_{i=1,2,\dots,q}$  que representan los parámetros que acompañan a la serie de errores  $\epsilon_t$ . Los parámetros de los modelos, tanto AR como MA, se obtienen mediante una regresión por mínimos cuadrados resultante de la comparación del conjunto de datos pasados utilizados como entrada, y de datos futuros que se desean obtener [Box98]. Al igual que para el modelo AR, existen diversas técnicas para establecer los órdenes del modelo AR,  $p$ , y del modelo MA,  $q$ , con lo que el modelo quedará definido como ARMA( $p,q$ ). En esta tesis, se proponen métodos basados en el estudio de la muestra de Funciones de Autocorrelación Parcial, *PACF* por sus siglas en inglés (*Partial Autocorrelation Function*), la muestra de Funciones de Autocorrelación Simple, *ACF* por sus siglas en inglés (*Autocorrelation Function*), y el Criterio de Información Bayesiana, *BIC* por sus siglas en inglés (*Bayesian Information Criterion*)[Boland95, Boland08].

La popularidad de las técnicas ARMA radica es su flexibilidad para representar diferentes tipos de series temporales según el orden de los modelos. Las series temporales deben ser series estacionarias para poder realizar ajustes adecuados con estos modelos [Hamilton94], habiendo demostrado ser adecuadas para la predicción de valores futuros. En la mayoría de los casos, se ha demostrado que para series estacionarias los modelos óptimos obtenidos tienen órdenes  $p$  y  $q$  no superiores a dos [Box98].

### 4.3.3. Estudio de la complejidad de los modelos lineales

En los modelos lineales utilizados en esta tesis una de las decisiones más importantes será el orden de los modelos AR y MA respectivamente. Para ello, siguiendo el método descrito por J. Boland [Boland95, Boland08], utilizaremos las Funciones de Autocorrelación y Autocorrelación Parcial, ACF y PACF, y los criterios de decisión bayesiana, BIC.

Para entender el cálculo de estos parámetros, primero realizaremos una serie de definiciones. Partiendo de dos series temporales  $X$  e  $Y$  con medias  $\mu_X$  y  $\mu_Y$  respectivamente, se define la covarianza de  $X$  e  $Y$  según la ecuación (4.8)

$$Cov(X, Y) = E(X - \mu_X)(Y - \mu_Y) \quad (4.8)$$

donde  $E(x)$  representa el valor esperado de una variable. Si las series  $X$  e  $Y$  no son independientes la covarianza toma valores positivos o negativos. En el caso de que para valores altos de  $X$  tiendan a coincidir valores altos de  $Y$ , la covarianza será positiva, mientras que si  $Y$  muestra valores bajos en el mismo instante, la covarianza será negativa. El coeficiente de correlación se obtiene dividiendo la covarianza por el producto de la desviaciones estándar de ambas series (4.9)

$$\rho = Corr(X, Y) = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sqrt{E(X - \mu_X)^2(Y - \mu_Y)^2}} \quad (4.9)$$

Si en lugar de dos series temporales aleatorias, se trabaja con dos series temporales que representan el mismo proceso estocástico en diferentes momentos,  $X_t$  y  $X_{t+\tau}$ , la expresión anterior se conoce como el coeficiente de autocorrelación para el instante  $\tau$  (4.10)

$$\rho_\tau = \text{Corr}(X_t, X_{t+\tau}) = \frac{E[(X_t - \mu_t)(X_{t+\tau} - \mu_{t+\tau})]}{\sqrt{E(X_t - \mu_t)^2 E(X_{t+\tau} - \mu_{t+\tau})^2}} \quad (4.10)$$

En una serie temporal se podría calcular, variando el valor de  $\tau$ , un coeficiente de autocorrelación entre dos instantes temporales cualesquiera. Por otro lado, en el caso de que la serie temporal  $X_t$  se considere un proceso estacionario, las desviaciones típicas de todas las variables en el tiempo son idénticas y por lo tanto la expresión de la *Función de Autocorrelación* (ACF) es según (4.11)

$$\rho_\tau = \text{Corr}(X_t, X_{t+\tau}) = \frac{E[(X_t - \mu_t)(X_{t+\tau} - \mu_{t+\tau})]}{E(X_t - \mu)^2} \quad (4.11)$$

La autocorrelación se puede calcular para todos los valores de la serie temporal  $X_t = X_1, X_2, \dots, X_N$ , y así obtener la autocorrelación muestral **SACF** para cada diferencia temporal  $\tau$  (4.12)

$$\hat{\rho}_\tau = \frac{\sum_{t=1}^{N-\tau} [(x_t - \mu)(x_{t+\tau} - \mu)]}{\sum_{t=1}^N (x_t - \mu)^2} \quad (4.12)$$

donde  $\mu$  representa la media de toda la muestra de la serie temporal  $X_t$ , compuesta por  $N$  valores, y  $\hat{\rho}_\tau$  será el valor de cada una de las Autocorrelaciones de la muestra. De esta manera, la Autocorrelación muestral SACF es una medida de la relación lineal que existe entre dos instantes cualesquiera dentro de la muestra, separados por un tiempo  $\tau$ . La autocorrelación puede tomar valores entre -1 y +1, siendo la relación lineal más fuerte cuanto más se acerque a ambos extremos. En el caso de valores positivos, la relación entre los distintos valores de la serie temporal será directamente proporcional. Además, si el valor de correlación es alto para una diferencia de tiempo  $\tau = 1$ , se está indicando una fuerte relación entre  $X_t$  y  $X_{t-1}$ ,  $X_{t-1}$  y  $X_{t-2}$ , y así hasta  $N$ .

Los valores de autocorrelación nos indican la relación lineal existente entre los instantes temporales de la serie. Por lo tanto, los valores significativos de la muestra de ACF representan el número de valores pasados que pueden ser necesarios para generar un modelo lineal. En el caso que nos ocupa, considerando un proceso estacionario gaussiano y que para valores altos de  $\tau$  la autocorrelación tiende a cero, se puede considerar que el 95 % se encuentran dentro del intervalo  $[-1.96 \text{std}(\rho_\tau), +1.96 \text{std}(\rho_\tau)]$ .

La autocorrelación parcial PACF establece la correlación entre dos instantes de tiempo de la serie con un retardo  $\tau$ , sin tener en cuenta la dependencia creada por los retardos intermedios existentes entre ambos instantes, es decir, muestra la correlación pura. La muestra de Autocorrelación Parcial **SPACF** para todos los retardos de tiempo  $\tau$  se puede obtener a partir de la **SACF** por medio de las ecuaciones de Yule-Walker. Los parámetros  $\hat{\phi}_{\tau\tau}$  ofrecen la estimación de SPACF para cada retardo a partir la ecuación (4.13)

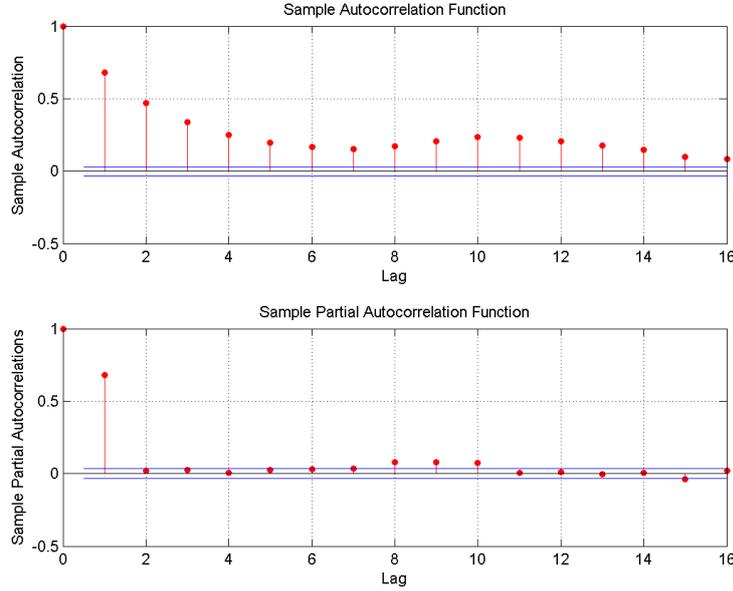


Figura 4.1: Ejemplo de SACF y SPACF para serie temporal estacionaria aleatoria. En rojo se representan los intervalos de elección

$$\hat{\phi}_{\tau\tau} = \frac{\rho_{\tau} + \sum_{t=1}^{\tau-1} \hat{\phi}_{\tau-1,t} \hat{\rho}_{\tau-1}}{1 + \sum_{t=1}^{\tau-1} \hat{\phi}_{\tau-1,t} \hat{\rho}_t} \quad (4.13)$$

De la misma manera que en el caso de SACF, se busca identificar los valores de SPACF significativos para establecer el retardo a partir del cual la relación entre las variables es prácticamente nula, ver Figura 4.1. Se puede considerar que el 95% de los valores se encuentran dentro del intervalo  $(-1.96/\sqrt{N}, +1.96/\sqrt{N})$ .

Una vez calculados las muestras de SACF y SPACF para la serie temporal estacionaria, se puede utilizar esta información para estimar el orden de los modelos lineales AR(p), MA(q) y ARMA(p,q). El criterio general, según J. Boland [Boland95, Boland08] es:

- Cuando el SACF decae paulatinamente y el SPACF contiene picos significativos hasta un valor de retardo  $\tau = p$ , se considera que la muestra se puede ajustar con modelo Autorregresivo AR(p).
- Cuando el SPACF decae paulatinamente y el SACF contiene picos significativos hasta un valor de retardo  $\tau = q$ , se considera que la muestra se puede ajustar con modelo de medias móviles MA(q).
- Cuando ambas muestras, SACF y SPACF, decaen paulatinamente, se considera que la muestra se puede ajustar con modelo Autorregresivo de medias móviles

ARMA( $p,q$ ). En este caso,  $p$  y  $q$  se deberán aumentar gradualmente hasta obtener un modelo óptimo.

En algunas ocasiones es muy difícil decidir entre los diferentes modelos, por ejemplo entre un AR(6) o un ARMA(2,1). Es por ello, que se han desarrollado unas técnicas de decisión basada en la teoría bayesiana [Lauret12, Lebarbier04, LR12], donde se tiene en cuenta el error cometido por el modelo y la complejidad del mismo a la hora de decidir el orden. En esta tesis, se ha optado por el Criterio de Información Bayesiana, *BIC* por sus siglas en inglés (*Bayesian Information Criterion*), ya que penaliza más el número de parámetros del modelo que otras técnicas existentes como el *Akaike Information Criterion AIC*,

$$BIC = -2\ln(\hat{L}) + m_k \ln(N) \quad (4.14)$$

donde  $\hat{L}$  es el valor de máxima verosimilitud del modelo,  $m_k$  es el número de parámetros a definir del modelo, en nuestro caso depende de los órdenes  $p$  y  $q$ , y  $N$  es el número de valores de la serie temporal en la ecuación (4.14). Para realizar una simplificación en el cálculo de la máxima verosimilitud, se asume que la distribución es gaussiana y por lo tanto la ecuación finalmente utilizada para definir el BIC es (4.15),

$$BIC = \ln(\hat{\sigma}^2) + m_k \frac{\ln(N)}{N} \quad (4.15)$$

donde  $\sigma^2$  es la varianza de los errores. Para tomar una decisión del modelo óptimo a utilizar, se utilizarán las gráficas de SPACF y SACF para identificar el número máximo de orden de los modelos AR y MA respectivamente. Estos valores  $p$  y  $q$  no serán los valores finalmente utilizados para ajustar la serie, sino que se realizarán los cálculos del parámetro BIC para todas las combinaciones posibles de modelos variando dichos parámetros  $\{1, 2, \dots, p\}$  y  $\{1, 2, \dots, q\}$ . El modelo óptimo a utilizar teniendo en cuenta la complejidad del mismo será el que muestre un valor BIC menor.

El capítulo 5, donde se muestran los resultados obtenidos para la Isla de Gran Canaria, se observará que finalmente, además del estudio del *Bayesian Information Criterion BIC*, se estudiaron los valores del error cuadrático medio relativo de modelo de ajuste. En muchas ocasiones, el modelo óptimo señalado por el BIC no ofrecía una mejora sustancial del error con respecto a modelo más sencillo, en el que se utilizaban un menor número de valores pasados de la serie temporal.

## 4.4. Redes Neuronales Artificiales

En la mayoría de los casos, la respuesta inmediata a cuál es la diferencia más significativa entre el ser humano y el resto de los animales, sería sin lugar a dudas nuestra capacidad de raciocinio. Esta característica nos ha permitido desarrollar una tecnología capaz de emular nuestras capacidades. En los últimos años, con el desarrollo de la ciencia y la tecnología, se ha manifestado la necesidad de realizar el tratamiento de gran cantidad de información, lo que ha conducido a enfrentarnos al reto de diseñar sistemas inteligentes

capaces de emular el funcionamiento del cerebro humano. Este es el objetivo de una rama científica conocida como Inteligencia Artificial (IA).

La Inteligencia Artificial se suele dividir en dos grandes áreas, la IA Simbólica y la IA Subsimbólica [Isasi04]. En la primera, los sistemas que se utilizan para resolver el problema se diseñan siguiendo esquemas previamente fijados. En la IA se dice que estos sistemas siguen un esquema de arriba hacia abajo, ya que necesitan disponer de una solución aproximada. Por el contrario, en la IA Subsimbólica no se utilizan esquemas diseñados previamente, sino que se parte de sistemas genéricos que irán modificándose y generándose mediante mecanismos de aprendizaje hasta formar un sistema capaz de resolver el problema. Es decir, desde esta perspectiva se estudian los mecanismos de los sistemas nerviosos del cerebro que nos hacen inteligentes, para poder diseñar sistemas basados en su estructura, funcionamiento y características que se adapten a los problemas que se deseen resolver.

Las técnicas de Machine Learning, o de aprendizaje automático en español, se encuentran dentro del ámbito de la Inteligencia Artificial Subsimbólica. Estas técnicas trabajan en el desarrollo y estudio de algoritmos capaces de resolver un problema o tomar decisiones a partir de un conjunto de datos conocido. Por ejemplo, para conseguir crear una hipótesis y ofrecer una respuesta a una situación no conocida, los algoritmos generan un modelo capaz de aprender de los datos de partida. El modelo generado se ajustará a la realidad en la medida en que pueda realizar predicciones de datos nuevos, que no se hayan presentado durante el entrenamiento. Por lo tanto, el conjunto de datos observados de un fenómeno concreto se deberá dividir en un conjunto de entrenamiento y otro de test, de manera que el primero nos permita generar un modelo y el segundo se utilice para validar el ajuste del mismo. La validación del modelo se calculará estableciendo una desviación sobre la realidad.

Entre las técnicas de Machine Learning más conocidas se encuentran las **Redes Neuronales Artificiales (RNA)**, un sistema de aprendizaje basado en la estructura de las neuronas biológicas, las **Redes Bayesianas**, en las que los nodos de la red se encuentran asociados a una función de probabilidad, los **Support Vector Machines**, utilizados para resolver problemas de clasificación y regresión, los **Gaussian Process**, útiles para el estudio de análisis espacial, etc. Aunque todos estos métodos han sido utilizados satisfactoriamente en el campo de la radiación solar [Diagne13, Mellit08], en esta tesis nos hemos centrado en el trabajo con **Redes Neuronales Artificiales y Redes Bayesianas**.

#### 4.4.1. Introducción a las redes neuronales

En general las redes neuronales tratan de emular el funcionamiento de una neurona como el elemento más simple en la estructura del cerebro. En la bibliografía especializada nos encontramos con diferentes definiciones, como la que realiza Mackay *"Muchos investigadores les gustaría crear máquinas que puedan aprender, reproducir patrones o descubrir patrones en los datos"* [MacKay03].

En la mayoría de las definiciones se establece una analogía con las neuronas del cerebro humano, aunque, como se verá más adelante, no todas las RNA emulan una

determinada estructura neuronal. Todas las redes neuronales realizan operaciones en una serie de elementos básicos que se conocen como neuronas por analogía. Estas unidades o *neuronas* están interconectadas entre sí por unos *pesos sinápticos*, los cuales variarán con el tiempo durante el periodo de aprendizaje. En general siempre que se describa una RNA se especificarán los siguientes tres aspectos [MacKay03]:

- **Arquitectura.**- Se debe tener en cuenta el número de neuronas involucradas y la forma de relación entre cada una de ellas.
- **Regla de activación.**- En esta parte se define en qué manera las neuronas responden a la interconexión entre ellas.
- **Método de entrenamiento.**- Los métodos de entrenamiento definen la manera en la que los pesos cambian durante el mismo.

Aunque se pueden situar las primeras investigaciones sobre Redes Neuronales Artificiales en el siglo XIX, no es hasta la segunda mitad del siglo XX en que se comienza realmente a desarrollar esta herramienta, gracias entre otras cosas al desarrollo del hardware. El primero que intentó realizar avances en computación a partir del estudio del cerebro fue Alan Turing en 1936. Warren McCulloch y Walter Pitts desarrollaron en 1943 una teoría sobre cómo trabajar con neuronas modelando una red simple a partir de circuitos eléctricos [McCulloch43, González95]. Posteriormente, en 1957 Frank Rosenblatt comenzó el desarrollo de uno de los tipos de redes que utilizamos en esta tesis, llamado *Perceptrón*. Éste era capaz de reconocer una serie de patrones que se le hubieran presentado anteriormente, pero presentaba una serie de limitaciones a la hora de clasificar clases no separables linealmente [Rosenblatt58].

En 1960 Widrow y Hoff presentaron el *ADALINE*, *Adaptive Linear Element*, el cual, a través de un algoritmo de entrenamiento sencillo denominado LMS (Least Mean Square), conseguía un sistema adaptativo que aprendía de forma más precisa que el perceptrón [Widrow60, Haykin96].

Los inconvenientes que presentaba el perceptrón a la hora de resolver problemas no separables linealmente, llevaron a Minsky y Papert en 1969 a publicar su libro *Perceptrons* [Minsky88]. En éste se exponían con mucha claridad los problemas que era capaz de resolver el perceptrón, pero se incidía en los inconvenientes que presentaban en la mayoría de los casos. Además también expusieron sus opiniones en contra del uso de las extensiones del perceptrón, como el multicapa, hecho por el que fueron criticados posteriormente ya que sus conjeturas parecen haber sido erróneas. En cualquier caso, este trabajo detuvo el desarrollo de las Redes Neuronales Artificiales durante diez años. Durante estos años, se sucedieron algunas investigaciones por parte de científicos diversos como Kohonen y Aderson, que proponen independientemente un modelo similar de memoria asociativa, *Asociador Lineal*.

En 1982, con la publicación de John Hopfield [Hopfield82], se recuperó el interés por este tipo de herramientas. En este trabajo, el autor describe de manera muy clara una variante del *Asociador Lineal*. Así mismo, Hopfield mostró cómo esta variante es capaz de trabajar. El uso principal que se le ha dado a estas redes es como memorias y

para resolver problemas de optimización. En ese mismo año, Teuvo Kohonen [Kohonen88] publica un trabajo sobre mapas autoorganizativos mediante reglas simples, cuyo sistema de aprendizaje es de tipo no supervisado, mientras que al año siguiente, Fukushima, Miyake e Ito presentan un dispositivo capaz de realizar reconocimiento de patrones con éxito, el *Neocognitrón* [Fukushima83].

En 1986, se presenta el trabajo de Rumelhart, Hinton y Williams, en el que se desarrolla un algoritmo de aprendizaje conocido como *retropropagación (backpropagation)* [Rumelhart86] para redes neuronales multicapa. A partir de esta publicación, el número de trabajos sobre RNA se han multiplicado apareciendo un gran número de aportaciones en los métodos de aprendizaje y tipo de estructuras.

#### 4.4.2. Aplicaciones

Las RNA tienen múltiples aplicaciones que se pueden dividir según el problema a resolver y el campo de conocimiento en el que se aplican. Así, para resolver problemas de **Clasificación** de elementos nos encontramos los siguientes trabajos:

- **Medicina.**- En el campo del diagnóstico médico, se han desarrollado trabajos en detección de cardiopatías o en detección de tumores cancerígenos, en la que una red neuronal analiza la posible presencia de tumores en una imagen.
- **Farmacia.**- Se utilizan para diagnosticar posibles efectos adversos al administrar un determinado fármaco para el tratamiento de cardiopatías o tratamientos oncológicos.
- **Procesado de señal.**- En este campo las redes neuronales se han desarrollado en un amplio espectro. Por ejemplo en la ecualización de canales de comunicación se ha mostrado más efectiva que otros métodos. Se emplea también en el reconocimiento de patrones en imágenes o en el reconocimiento de voz.
- **Economía.**- Debido a la necesidad de tomar decisiones entre una gran cantidad de opciones las RNA son bastante aplicables en la concesión de créditos, detección de posibles fraudes o la posibilidad de la quiebra de un banco.

Por otro lado también se pueden aplicar para resolver problemas de **Modelización**. Se pueden encontrar diversos trabajos con RNAs en los siguientes campos:

- **Medicina.**- Se relacionan con el modelado de diferentes tipo de señales como el electrocardiograma (FEGG), el electromiograma (EMG), electroencefalograma (EEG), etc.
- **Farmacia.**- Determinar la concentración de un determinado fármaco en sangre.
- **Procesado de señal.**- En este campo se han hecho trabajos en la eliminación activa de ruido y en el control de sistemas.

- **Economía.**- Las RNA han sido utilizadas para intentar predecir comportamientos futuros que permitan valorar el éxito de una operación. En concreto encontramos trabajos en predicción del gasto eléctrico, cambio de moneda, tendencias a corto y medio plazo en la bolsa, predicción de stocks, etc.
- **Medio Ambiente.**- Las variaciones en el medio ambiente no son lineales y dependen de muchas variables por lo que lo hacen un campo muy apetecible para utilizar las RNA. En concreto, en la predicción de la radiación solar que nos ocupa en esta tesis, se encuentran multitud de trabajos al respecto. Estos trabajos se han desarrollado partiendo de datos de radiación como variable de entrada [Hontoria01, Lauret15] o de la mezcla éstos con otros parámetros meteorológicos como la temperatura, humedad relativa, horas de sol, velocidad y dirección del viento [AA98, Rehman08, Mellit10, Ghanbarzadeh09]. Por otro lado, también se han aplicado en la predicción de niveles tóxicos de ozono en zonas urbanas y rurales o en la predicción de temperatura [Almonacid13]

Así, el objetivo de las Redes Neuronales Artificiales es crear un sistema con una serie de unidades, de manera que el comportamiento global trate de emular el funcionamiento del sistema neuronal humano. Esto hace imprescindible el estudio de las características y funcionamiento de una neurona simple.

#### 4.4.3. Fundamentos teóricos de las redes neuronales

El sistema nervioso humano es el encargado de realizar las comunicaciones entre los órganos de los sentidos, que reciben los estímulos del exterior, y los órganos diana, que realizan las acciones (músculos, glándulas). Así pues, se encarga de recoger la información del exterior, transmitirla, procesarla y enviarla una vez elaboradas. En este proceso nos encontramos con los *receptores*, células encargadas de recoger la información del exterior o del propio interior del cuerpo, *el sistema nervioso*, que recoge, elabora y envía esta información, y los *órganos diana*, que reciben la información y la interpretan.

La transmisión de toda esta información se realiza por medio de unas células específicas conocidas como neuronas. Estas células forman redes en las cuales se elabora y transmite la información mediante señales electro-químicas. Por lo tanto, una parte de la red de células neuronales debe estar conectada a los órganos receptores, donde reciben la información del exterior o del interior del cuerpo. La información recibida de los receptores se denomina *estímulo* y pone en funcionamiento la red. Las neuronas de la red conducen la información hasta las conexiones con los órganos diana. La red neuronal no es más que multitud de neuronas simples interconectadas entre sí. Por lo tanto, una neurona puede recibir información de un órgano receptor o de otra neurona, y la transmite a otra neurona o a un órgano diana [Isasi04, González95]. Los componentes principales de una célula neuronal son los siguientes:

- **Dendritas**, la neurona recibe la información del exterior y la transmiten al interior de la célula a través de estas ramificaciones de entrada.

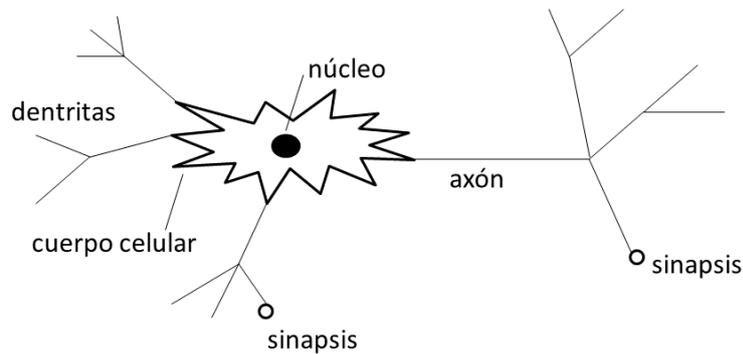


Figura 4.2: Reproducción de una célula neuronal [Isasi04]

- **Cuerpo**, también conocido como núcleo, donde se procesa la información hasta generar una respuesta que propaga al axón.
- **Axón**, transmite la señal de salida de la neurona.

La conexión entre los axones de salida de una neurona y las dendritas de entrada de la neurona siguiente se realiza mediante unos contactos especiales llamados *sinapsis*. Las señales que se transmiten son de carácter electro-químico. La señal generada por la neurona y transportada por el axón es un impulso eléctrico, mientras que la señal que se transmite entre el axón y las dendritas de la siguiente neurona es de origen químico. La unión entre ambas neuronas es un espacio líquido denominado sinapsis donde existen determinadas concentraciones de elementos ionizados. Dependiendo de la concentración de estos iones cambia la propiedad conductiva del espacio, de manera que la sinapsis puede actuar como potenciador o inhibidor de la señal que viene de los axones. Esta señal es de suma importancia, ya que la suma de todas las señales recibidas por las dendritas activará a la neurona, siempre que supere un cierto umbral.

El funcionamiento general de una neurona comienza con la recepción de la información de las dendritas de la neurona a través de la sinapsis. La información proviene de las células vecinas a las que pueda estar conectada la neurona en cuestión. Las dendritas transmiten la información recibida al núcleo o cuerpo de la neurona, donde se procesa y genera una respuesta. Esta respuesta se propaga al axón, donde se ramifica y se transmite a las dendritas de las neuronas conectadas, ver Figura 4.2.

Las Redes Neuronales Artificiales (RNA) intentan reproducir el comportamiento de las redes de neuronas biológicas. Por lo tanto, la estructura más simple de estas redes,

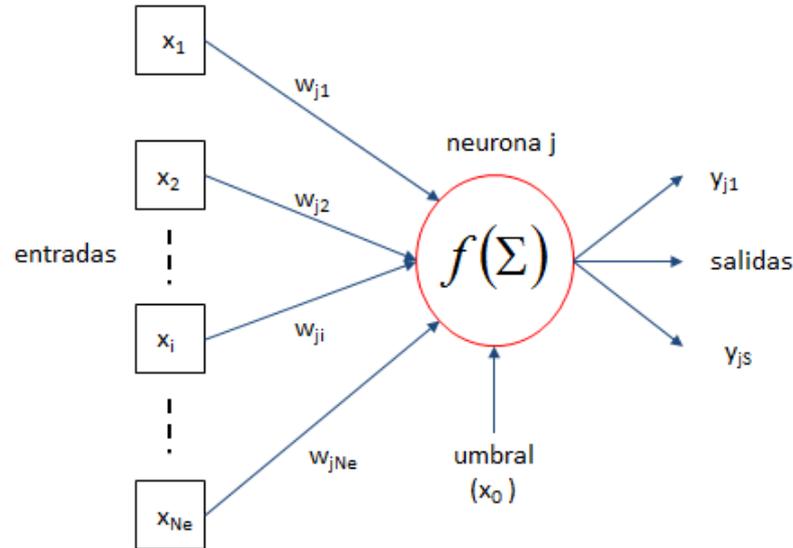


Figura 4.3: Estructura general una neurona artificial

la neurona artificial, debe también emular las características de la neurona biológica. En la Figura 4.3, se establece la analogía entre ambos tipos de neuronas. Las señales  $\{x_1, x_2, \dots, x_i, \dots, x_{N_e}\}$  serán los estímulos que entran en las neuronas a través de las dendritas. Las sinapsis *potencian o inhiben* estas entradas según la concentración del medio, para lo cual se utilizan unos parámetros  $\{\omega_{j1}, \omega_{j2}, \dots, \omega_{ji}, \dots, \omega_{jN_e}\}$ , conocidos por analogía como *pesos sinápticos*. Todas señales de entrada pueden excitar la neurona si superan un umbral específico para cada una. Si la suma de las entradas ponderadas por los pesos sinápticos iguala o supera este umbral, la neurona se activa y genera una serie de salidas  $\{y_{j1}, \dots, y_{js}\}$ , mientras que en caso contrario permanece en letargo. La salida estará afectada por una función propia de la neurona que simula el tratamiento de la información que se produce en el núcleo. El aprendizaje de las neuronas se basa en la capacidad que se tenga para modificar los pesos sinápticos y así conseguir influir en la importancia de cada entrada en la célula.

Una RNA está conformada por un conjunto de neuronas simples interconectadas entre sí. Estas neuronas se distribuyen dentro de la propia red conformando capas. La entradas de la mismas pueden provenir del exterior, *estímulos*, o de otras neuronas según en la capa en la que se encuentran. La capas de una red neuronal son la *capa de entrada*, recibe la información del exterior, la *capa oculta*, son capas intermedias que se conectan únicamente con otras neuronas, y la *capa de salida*, que obtiene la salida final de la red. Estas estructuras neuronales se pueden clasificar de diferentes maneras según el criterio utilizado.

Así, según el **número de capas**:

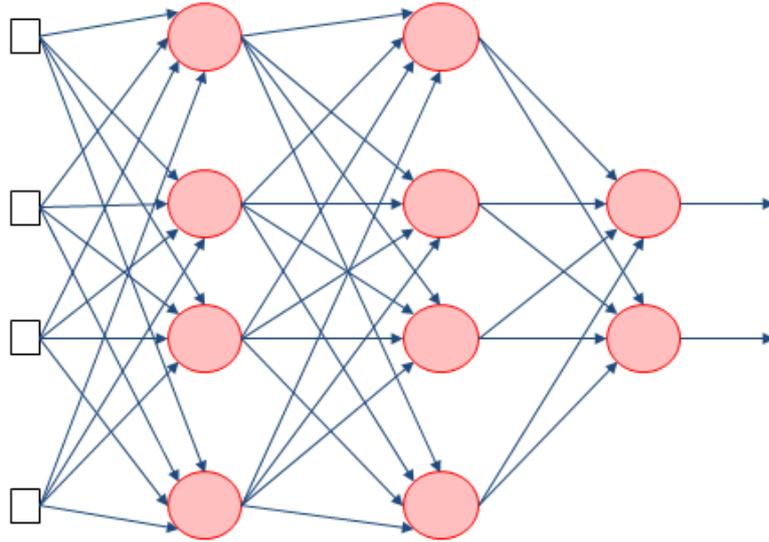


Figura 4.4: Estructura general una red multicapa

- *Redes neuronales monocapa*, disponen de una capa de entradas y una capa de neuronas donde se realizan los cálculos. La capa de entrada no se cuenta, ya que no realiza ningún cálculo y sólo transmite las variables de entrada.
- *Redes neuronales multicapa*, Es una generalización de la anterior, pero se incluyen una serie de capas intermedias (ocultas) entre la entrada y la salida, ver Figura 4.4.

Según el **tipo de conexiones** entre las neuronas de la red:

- *Redes neuronales no recurrentes*, en este tipo de redes la señales únicamente se propagan en un sentido, no existe la posibilidad de realimentación.
- *Redes neuronales recurrentes*, en este caso existen lazos de realimentación. Estos lazos se pueden dar entre neuronas de la misma capa o de capas distintas. Las estructuras recurrentes se muestran muy adecuadas para sistemas no lineales.

Según el grado de conexión de las capas:

- *Redes neuronales totalmente conectadas*, en este caso todas las neuronas de una capa se encuentran conectadas con las de la capa siguiente, si no son recurrentes, y con las de la anterior también si son recurrentes.
- *Redes neuronales parcialmente conectadas*, no se conectan la totalidad de las neuronas de las capas.

Como ya se ha comentado, la capacidad de las redes neuronales de modificar los pesos sinápticos forma parte del proceso de aprendizaje. Al igual que se debe definir la estructura de la red según el problema que se desea abordar, es necesario definir el procedimiento por el cual las conexiones del sistema varían para proporcionar la solución adecuada. La división entre algoritmos *supervisados* y *no supervisados* es la más importante según el método de aprendizaje. En el caso de los algoritmos no supervisados, la red no conoce la salida final que debe dar y agrupa las entradas según sus características. Por el contrario el aprendizaje supervisado enseña a la red la salida real que debe conseguir durante el entrenamiento. En este caso, la red compara su salida con la que pretende conseguir, evalúa la diferencia y modifica los pesos sinápticos para minimizar esta diferencia. A su vez, este último se puede dividir en aprendizaje por refuerzo, en el que únicamente se puede evaluar si la salida es igual o no a la salida deseada, y aprendizaje por corrección, en el que sí se conoce la diferencia entre ambas salidas y se puede cuantificar.

#### 4.4.4. La neurona simple

El modelo de una neurona simple tiene cuatro elementos básicos en cualquiera de las arquitecturas posibles, ver Figura 4.5:

- **Pesos sinápticos**, son las variables  $\omega_{ji}$  que ponderan las conexiones entre las neuronas. Durante el entrenamiento los pesos se irán modificando hasta encontrar la mejor combinación para conseguir la salida deseada.
- **Sumador**, realiza la suma ponderada de la entradas a la neurona.
- **Función de activación**, una vez realizada la suma, si se supera el umbral y la neurona se activa, se aplica la función de activación. A continuación se muestran algunas de las funciones más utilizadas en la bibliografía:
  - *Función de signo*, la primera de las funciones propuesta. Únicamente separa los datos de entrada según superen o no un límite.

$$f(x) = \begin{cases} -1 & x < 0 \\ 1 & x \geq 0 \end{cases} \quad (4.16)$$

- *Función lineal*, en la que la salida es igual a la entrada.

$$f(x) = x \quad (4.17)$$

- *Función sigmoideal*, una derivada de la función signo. Muchos métodos de aprendizaje necesitan una función de activación derivable. Esta función cumple ese requisito, mientras que la función signo no.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4.18)$$

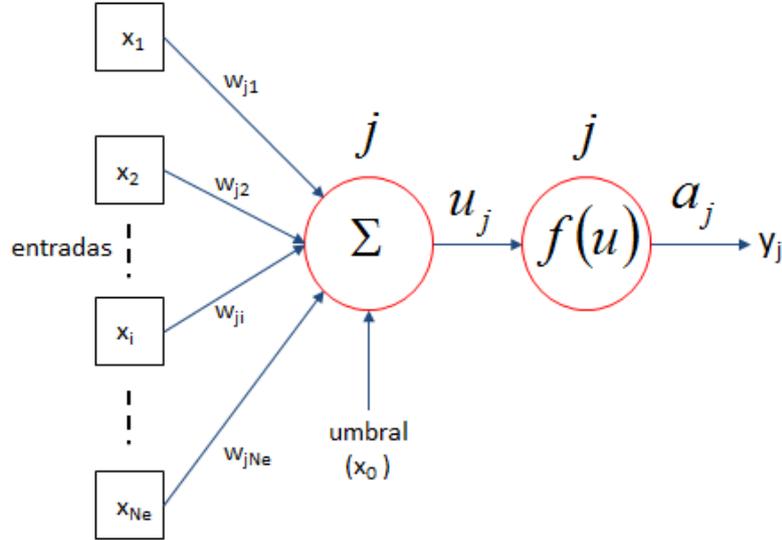


Figura 4.5: Estructura general una neurona artificial sin memoria

- *Función tangente hiperbólica*, una función bipolar de la anterior en la que la salida se limita entre -1 y +1. Esta es la función que se utilizará en las redes neuronales de esta tesis.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4.19)$$

- *Función gaussiana*, función que tiende a cero a partir de un valor. Comúnmente e utilizada como función de base radial (RBF).

$$f(x) = K_1 e^{-K_2 x^2} \quad (4.20)$$

- **Umbral**, determina el nivel al cual se activa la neurona.

La expresión matemática de la neurona de la Figura 4.5, supone que cada entrada  $x_i$  es un valor numérico, cuya sumatoria ponderada por los pesos  $\omega_{ji}$  conformará la entrada total a la neurona. El umbral se suele representar como un peso sináptico que acompaña a una entrada de valor la unidad.

$$U_j = \sum_{i=1}^{N_e} (\omega_{ji} x_i) \quad (4.21)$$

Así pues, la salida  $y$  de una neurona simple con una función de activación  $f$  vendrá dada por la expresión (4.22)

$$y_j = f\left[\sum_{i=1}^{N_e} (\omega_{ji}x_i) + \omega_0\right] \quad (4.22)$$

La primera estructura neuronal que se desarrolló, el **Perceptrón**, se basa en una neurona simple con una función de activación de signo. Se deduce, por tanto, que este elemento es un clasificador binario ya que puede determinar la pertenencia de la entrada a dos clases diferentes. El método de aprendizaje es supervisado, por lo que durante el entrenamiento se le deben presentar a la red un conjunto  $n$  de parejas de valores entrada-salida  $\{X_i, t\}_n$ , cada uno de los cuales se denominará *patrón de entrenamiento*. Estos patrones aportarán la información necesaria para que la estructura encuentre la combinación que mejor se adapte al problema. En la práctica primero se inician los pesos aleatoriamente y se calcula la salida con el primer conjunto de entradas  $X_i$ . La salida obtenida  $y$  se compara con la salida objetivo del patrón de entrenamiento  $t$  y, si no coinciden se modifican los valores de la conexiones, es decir, los pesos  $\omega_{ij}$ .

$$\omega_{n+1} = \omega_n + \alpha(t - y)x_n \quad (4.23)$$

A partir del estudio de esta sencilla estructura se desarrollaron el resto de redes neuronales. Una de las mejoras más importantes, por el número de aplicaciones y porque introduce la manera de aprendizaje posterior, es el *ADALINE*, *Adaptive Linear Element*. Este sistema sigue la misma estructura que el Perceptrón pero modificando el método de actualización de los pesos. En este caso, se realiza la comparación de la salida obtenida  $y$  con la deseada  $t$  antes de aplicar la función de activación signo. De esta manera, la comparación es numérica y, por lo tanto, la estructura de pesos óptima es aquella que minimice la función de error. En la mayoría de ocasiones, el cálculo del error se realiza con la función cuadrática (4.24).

$$E(n) = error^2(n) = (t(n) - y(n))^2 \quad (4.24)$$

Durante todo el aprendizaje se buscará realizar una modificación de pesos que minimice la función de error. El procedimiento usado para localizar este mínimo se definirá como el algoritmo de aprendizaje, entre los que destacan los de descenso por gradiente utilizados en esta tesis.

#### 4.4.5. El Perceptrón Multicapa

En general, una estructura de RNA multicapa como la de la Figura 4.6 dispone de una capa de neuronas de entrada, al menos una capa oculta y un capa de neuronas de salida. Las entradas a la red  $x_i$ , que conforman la capa de entrada, no se consideran formalmente neuronas ya que no realizan ninguna función sino propagar la información a las neuronas de la capa oculta. Cada una de las neuronas de la capa de entrada  $x_i$  está conectada con cada una de las neuronas  $j$  de la capa oculta. La conexión se realiza a través de los pesos sinápticos, siendo  $\omega_{ji}^1$  la conexión entre la entrada  $i$  y la neurona  $j$ . Así, matemáticamente en la neurona  $j$  se genera una salida intermedia  $a_j$  según la expresión (4.25)

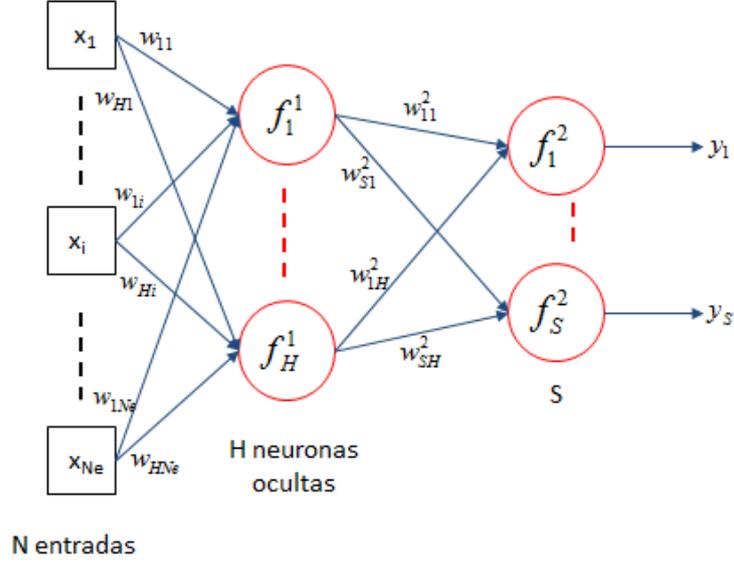


Figura 4.6: Ejemplo de los elementos de una red multicapa general

$$a_j = f_j \left[ \sum_{i=1}^{N_e} (\omega_{ji}^1 x_i) \right] \quad (4.25)$$

donde  $\omega^1$  representa una matriz con todos los pesos sinápticos de conexión entre la capa de entrada y la primera capa oculta. Esta matriz contendrá en cada fila todos los pesos de conexión entre cada una de las entradas y una neurona específica de la capa oculta. Por contra, en cada columna nos encontramos la conexión entre una entrada cualquiera y cada una de las neuronas de la capa oculta (4.26).

$$\omega^1 = \begin{bmatrix} \omega_{11}^1 & \omega_{12}^1 & \dots & \omega_{1N_e}^1 \\ \omega_{21}^1 & \omega_{22}^1 & \dots & \omega_{2N_e}^1 \\ \dots & \dots & \dots & \dots \\ \omega_{j1}^1 & \dots & \omega_{ji}^1 & \omega_{jN_e}^1 \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (4.26)$$

La función de activación de cada neurona de la capa oculta está representada por  $f_j$ , que se activa si la suma ponderada de las entradas supera el umbral. Las salidas intermedias  $a_j$  se propagan hacia la siguiente capa oculta o hacia la capa de salida de la misma manera que entre las capas anteriores. Las conexiones entre ellas se realizarán mediante los pesos  $\omega^2$ . Así en la capa de salida final, la neurona de  $s$  produce una salida según la ecuación (4.27),

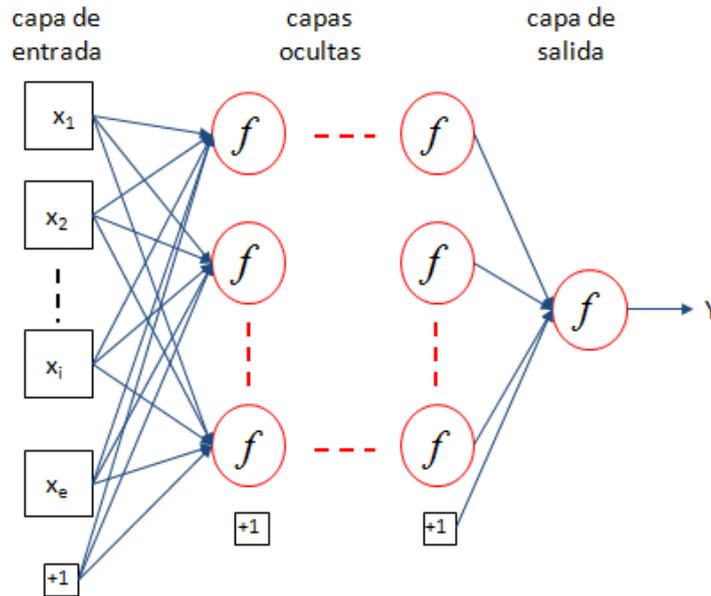


Figura 4.7: Arquitectura del Perceptrón Multicapa (MLP)

$$y_s = f_s \left[ \sum_{j=1}^H \omega_{sj}^2 f_j \left[ \sum_{i=1}^{N_e} (\omega_{ji}^1 x_i) \right] \right] \quad (4.27)$$

donde  $H$  muestra el número de neuronas de la capa intermedia y  $N_e$  muestra el número de entradas a la RNA. En este caso,  $\omega_{sj}^2$  representa el peso de conexión de la neurona de la capa intermedia  $j$  con la neurona de la capa de salida  $s$ , donde el superíndice dos indica que nos encontramos en el segundo conjunto de pesos. La función  $f_s$  es la función de activación de la neurona de salida  $s$ .

El *Perceptrón* simple mostraba limitaciones para resolver problemas de separaciones no lineales. Por este motivo, se desarrolló el *Perceptrón Multicapa*, en inglés *Multilayer Perceptron (MLP)*, como una combinación de varias unidades simples. Algunos autores han demostrado que el MLP es un aproximador universal de cualquier función continua en el espacio  $\mathbb{R}^n$  [Hornik89]. Al igual en que se explicó para las RNA multicapa, el MLP dispone las neuronas entre la capa de entrada, al menos una capa oculta o intermedia y la capa final de salida. En la Figura 4.7 se representa un MLP con una capa oculta y una salida con una única neurona. Las entradas se representan por las variables  $x_i$ , los pesos  $\omega^1$  son los pesos sinápticos entre la capa de entrada y la capa intermedia, mientras que  $\omega^2$  son las conexiones entre la capa intermedia y la única neurona de salida, y por último la salida final de la red se representa por  $y$ .

Las conexiones del MLP siempre son hacia adelante, *feed-forward*, es decir que cada capa se conecta con la capa siguiente. Generalmente, la red está conectada totalmente, con lo que todas las neuronas de una capa están conectadas con todas las neuronas de la capa siguiente. Por otro lado, cada neurona tiene asociado un umbral de activación,

que en el caso del MLP se suele tratar como una conexión más asociada a una entrada de valor la unidad.

Las funciones de activación  $f$  de cada neurona más utilizadas son la función sigmoideal (4.18), con un rango de valores continuo entre  $[0, 1]$  y la función tangente hiperbólica (4.19), con un rango de valores continuo entre  $[-1, 1]$ . El diseñador elegirá el tipo de función de activación basándose en los valores que desee que tomen las salidas de las neuronas. Por contra, las neuronas de salida tienen una función de activación distinta y, es muy común que ésta sea la función lineal (4.17).

Matemáticamente, un MLP con una capa intermedia con función de activación  $f_j$  en la neurona  $j$  y una única neurona de salida con una función de activación lineal se puede representar con la siguiente formulación. Las salidas de la capa intermedia se obtienen según la expresión (4.28).

$$a_j = f_j \left[ \sum_{i=1}^{N_e} (\omega_{ji}^1 x_i) + \omega_0^1 \right] \quad (4.28)$$

Las salidas  $a_j$  de la capa intermedia se propagarán hacia la única neurona de salida y producirán la salida  $y$  según la ecuación (4.29).

$$y = \sum_{j=1}^H \omega_{sj}^2 f_j \left[ \sum_{i=1}^{N_e} (\omega_{ji}^1 x_i) + \omega_0^1 \right] + \omega_0^2 \quad (4.29)$$

Cuando se utiliza un MLP para resolver un problema, uno de los pasos más importantes a realizar es la elección de la arquitectura del mismo. Este diseño implica la elección del número de neuronas de cada capa y el número de capas. Algunos de estos parámetros los debe seleccionar el diseñador y otros vienen impuestos por la propia naturaleza del problema.

El número de neuronas de la capa de salida viene definido por las variables que se están trabajando. En el caso de esta tesis, la salida que se pretende es única porque se busca la predicción de la radiación solar. Sin embargo, el número de entradas no siempre viene definido por el problema. Se puede dar el caso, de que existan algunas entradas que no sean relevantes para la solución del problema y la información que aportan es irrelevante, por lo que su uso complica la red neuronal sin obtener ningún beneficio. Para evitar estas situaciones, es muy conveniente hacer un estudio de la importancia de cada una de las entradas para resolver el problema y descartar previamente las que no sean relevantes. Existen multitud de técnicas para realizar este tipo de análisis (algoritmos genéticos, análisis de sensibilidad, etc). Para definir el número de capas ocultas y el número de neuronas de cada una de ellas no existe ninguna regla teórica definida. Así, el diseñador realizará la elección basándose en técnicas de ensayo y error. El número de neuronas ocultas debe ser suficiente para que exista el número de pesos adecuado para resolver el problema. Sin embargo, un número elevado de conexiones complicaría la estructura de la red, elevaría la carga computacional derivada del proceso de entrenamiento e impediría la generalización de la solución, ya que se ajustaría demasiado a los datos que se le muestran en el entrenamiento pero no a otro conjunto de datos diferentes. El número de neuronas

intermedias puede influir en el comportamiento pero, generalmente, no es un parámetro significativo, ya que el MLP puede mostrar muy diferentes arquitecturas para resolver un mismo problema. En general con una sola capa intermedia el MLP puede resolver la mayoría de los problemas, por lo que en esta tesis únicamente se estudiará el número de neuronas de una única capa intermedia.

#### 4.4.6. Regla de aprendizaje. Backpropagation

Con todo lo explicado, la importancia de las RNA radica en su capacidad de aprendizaje conforme se le muestran las relaciones entre los datos de entrada y salida. Así, la parte más importante del proceso de resolución del problema es el aprendizaje. Éste consiste en general en la modificación paulatina de los pesos sinápticos de las conexiones entre neuronas, de manera que se capacite a la red para resolver de la mejor manera posible el problema. Las RNA son sistemas que aprenden a partir de ejemplos, por lo que éstos deben ser suficientes en número, para darle tiempo a la red a adaptar los pesos, y representativos, ya que si los ejemplos mostrados son específicos de un tipo de evento la red se especializará en resolver este tipo de problema. A la hora de preparar el conjunto de datos de entrenamiento es importante que todas las regiones significativas del espacio estén representadas en el mismo.

El proceso de aprendizaje de una red neuronal presenta una serie de pasos, Figura 4.8. Primero se establece el conjunto de datos de entrenamiento, conformado por datos de entrada  $x_i$  y datos de salidas esperadas  $t$ , posteriormente se realiza un inicio aleatorio de los pesos de la RNA, se calcula la salida y se compara con la salida deseada, modificando los pesos según la diferencia que exista hasta conseguir un determinado criterio de convergencia. Mientras no se cumpla dicho criterio, se repetirá el proceso introduciendo todos los ejemplos del conjunto de entrenamiento. Cuando se cumple el criterio se dice que ha concluido el proceso de aprendizaje. La modificación de los pesos puede realizarse después de introducir cada ejemplo del conjunto, *aprendizaje on-line*, o una vez introducidos todos los ejemplos, *aprendizaje Batch*.

El criterio de convergencia dependerá del tipo de problema a resolver y del tipo de estructura elegido. La finalización del proceso de aprendizaje se puede determinar:

- *Mediante un número de ciclos fijos*, se fija el número de iteraciones que se van a realizar de antemano. Cuando se alcance se detiene el aprendizaje y se acepta la RNA obtenida.
- *Cuando el error descienda por debajo de una cantidad*, se define antes de comenzar un error mínimo que se desea alcanzar. Puede darse el caso de que la red no consiga reducir el error hasta dicho objetivo, en ese caso se debe introducir otro criterio de parada (como puede ser el número de ciclos). Si ocurre esto se dice que no se ha conseguido encontrar la solución y deberán cambiarse algunos parámetros para probar otra vez.
- *Cuando la modificación del error sea insignificante en cada iteración*, puede ocurrir que, independientemente del valor nominal del error, éste no se disminuya de manera

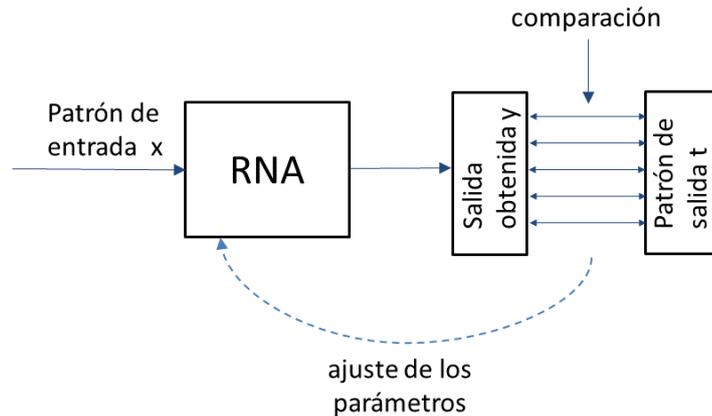


Figura 4.8: Esquema de aprendizaje supervisado de una RNA

significativa con cada iteración. Se fija de antemano un valor mínimo de modificación del error entre una iteración y la siguiente.

- Cuando la modificación de los pesos sea irrelevante, si en el proceso de entrenamiento llega el momento en que ya no se varían los valores de los pesos en cada iteración, se da por finalizado el entrenamiento.

El método de aprendizaje es el proceso seguido para ir adaptando los parámetros de la red. En esta tesis el criterio que se va a utilizar se basa en el estudio del error cometido por la red. Existen diversos métodos para minimizar la función de error elegida. En el caso del Perceptrón Multicapa (MLP), los más utilizados son los algoritmos por descenso de gradiente que se basan en la minimización de una determinada función de error entre la salida deseada  $t$  y la obtenida en la predicción  $y$ . La función de error  $E_D$  ampliamente utilizada en la bibliografía es el error cuadrático medio [Bishop95], según la ecuación (4.30),

$$E_D(\omega) = \frac{1}{2} \sum_{i=1}^N (t_i - y_i)^2 \quad (4.30)$$

donde  $N$  es el número total de patrones del conjunto de datos de entrenamiento. Cada patrón se considera una pareja de valores entrada-salida relacionados entre sí  $\{x_i, t_i\}$ . El algoritmo de aprendizaje que utilizará el MLP es el de RetroPropagación, conocido por su nombre en inglés Backpropagation. En su caso, es un algoritmo de descenso por gradiente que actualiza los pesos en cada iteración retropropagando la señal desde la capa salida hasta la capa de entrada. El algoritmo primero propaga la señal hacia adelante, desde

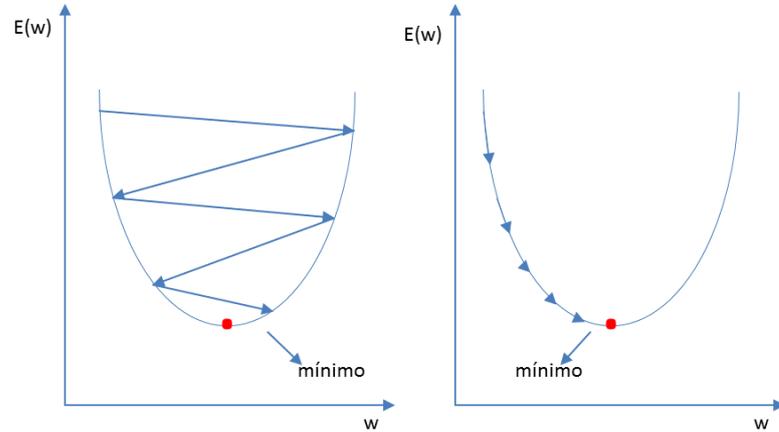


Figura 4.9: Esquema del avance en la optimización. En la izquierda se toma un valor alto de coeficiente de aprendizaje y en la derecha valores pequeños.

la entrada a la salida, calculando la salida  $y$  a partir de los conjuntos de pesos actuales  $(\omega^1, \omega^2)$  y estimando el error entre esta salida obtenida y la deseada  $t$ . Posteriormente, en función del error obtenido en la salida, el algoritmo actualiza los valores de los pesos sinápticos que determinan las conexiones entre las neuronas partiendo de la capa de salida a la capa de entrada.

La minimización de la función de error es un problema no lineal y, como consecuencia, tienen que utilizarse técnicas de optimización no lineales para su resolución. En el caso del Perceptrón la regla de minimización es ajustar los parámetros de la red siguiendo la dirección negativa del gradiente de la función de error. Por tanto, aplicando el método de descenso de gradiente, cada peso  $\omega$  de la RNA se modifica en cada iteración con la siguiente regla de aprendizaje general (4.31),

$$\omega_{k+1} = \omega_k - \eta \frac{\partial E}{\partial \omega_k} \quad (4.31)$$

donde  $k$  y  $k + 1$  representan las sucesivas iteraciones durante el entrenamiento. Se debe tener en cuenta que en cada entrenamiento los pesos se modifican y se vuelve a calcular el error. El parámetro  $\eta$  se conoce como el coeficiente de aprendizaje e influirá en la velocidad de convergencia durante el entrenamiento. Si se elige un valor de  $\eta$  muy alto, en cada iteración se produce una modificación grande de los pesos y se desplazará rápidamente por la superficie de la función de error. De esta manera se corre el riesgo de pasar por encima de un punto mínimo y oscilar alrededor de él sin alcanzarlo. Por el contrario, si se elige un valor muy pequeño las variaciones de pesos serán muy pequeñas y el descenso será muy paulatino, por que serán necesarias muchas iteraciones, ver Figura 4.9.

La actualización de los pesos en el *algoritmo de backpropagation* se inicia por la capa de salida y se va propagando hasta la capa de entrada. En esta tesis, se trabajará en todos los casos con redes de una sola capa intermedia con función tangente hiperbólica y una

salida única con función lineal. La regla general de aprendizaje para este tipo de redes conlleva pues, la actualización del conjunto de pesos que une la capa oculta con la de salida,  $\omega^2$ , y la los pesos que unen la capa de entrada con la capa oculta,  $\omega^1$ . Teniendo en cuenta estos aspectos, la actualización de los pesos de la capa de salida queda definida según (4.32).

$$\omega_{1j}(k+1) = \omega_{1j}(k) - \eta \frac{\partial E_D}{\partial \omega_{1j}(k)} \quad (4.32)$$

Teniendo en cuenta la función de error (4.30) y que la salida deseada  $t$  no depende de los pesos y por lo tanto es una constante, la derivada del error respecto de los pesos de salida queda según la ecuación (4.33).

$$\frac{\partial E_D}{\partial \omega_{1j}(k)} = -(t(k) - y(k)) \frac{\partial y(k)}{\partial \omega_{1j}(k)} \quad (4.33)$$

Llegados a este punto se deberá calcular la derivada de la función de salida  $y$  respecto del peso  $\omega_{1j}$ . Como se explicó, la función de salida  $y$  queda definida según la ecuación (4.29). Al aplicar la regla de la cadena se debe tener en cuenta que el único término de la ecuación (4.29) cuya derivada es distinta de cero es  $\omega_{1j}a_j$ . El término  $a_j$  está definido por la ecuación (4.25), por lo que La derivada de la salida  $y$  respecto de los pesos quedará según la ecuación (4.34).

$$\frac{\partial y(k)}{\partial \omega_{1j}(k)} = f_j \left[ \sum_{i=1}^N (\omega_{ji}^1 x_i) + \omega_0^1 \right] = a_j \quad (4.34)$$

Siendo común definir  $\delta$  como el término de actualización de la salida en cuestión para la iteración de entrenamiento  $k$  según la ecuación (4.35)

$$\delta(k) = -(t(k) - y(k)) \quad (4.35)$$

Finalmente se concluye que la actualización de los pesos del conjunto de salida se realiza según la ecuación (4.36).

$$\omega_{1j}(k+1) = \omega_{1j}(k) + \eta(t(k) - y(k))a_j = \omega_{1j}(k) + \eta\delta(k)a_j \quad (4.36)$$

Se puede observar que la actualización del peso de conexión entre la neurona  $j$  de la capa oculta y la salida únicamente dependerá de la salida intermedia  $a_j$  de dicha neurona y del error cometido por la salida obtenida  $y$  respecto a la salida deseada  $t$ , expresado en la función  $\delta(k)$ . Siguiendo el mismo método se puede obtener la expresión de actualización del umbral de la neurona de salida  $\omega_0^2$  según la ecuación (4.37).

$$\omega_0^2(k+1) = \omega_0^2(k) + \eta\delta(k) \quad (4.37)$$

Una vez se han actualizado los pesos de la capa de salida, se continua con la retropropagación y se realizará la actualización de los pesos que unen las capa de entrada

con la capa oculta,  $\omega^1$ . Así, la actualización de la conexión entre la entrada  $i$  y la neurona intermedia  $j$  queda definida según (4.38).

$$\omega_{ji}(k+1) = \omega_{ji}(k) - \eta \frac{\partial E}{\partial \omega_{ji}(k)} \quad (4.38)$$

Al igual que en el caso anterior se utiliza la regla de la cadena para hallar las expresiones generales de la derivada. Se debe tener en cuenta que el peso  $\omega_{ji}$  influye en la activación de la neurona  $j$ ,  $a_j$ , y que el resto de activaciones de la capa oculta no dependen de dicho peso,

$$\omega_{ji}(k+1) = \omega_{ji}(k) - \eta \delta_j(k) x_i \quad (4.39)$$

donde  $x_i$  es la entrada correspondiente al peso que se está actualizando y  $\delta_j(k)$  viene dado por la expresión (4.40) para la neurona  $j$  en la iteración  $k$ ,

$$\delta_j(k) = f'_j \left[ \sum_{i=1}^N (\omega_{ji}^1 x_i) + \omega_0^1 \right] \delta(k) \omega_{1j} \quad (4.40)$$

donde  $f_j$  es la función de activación de la neurona intermedia  $j$ , cuya expresión es la tangente hiperbólica en el caso que estamos explicando. Por lo que su derivada se puede expresar de forma sencilla según (4.41).

$$f'_j = 1 - f_j^2 \quad (4.41)$$

Por último, de la misma manera que en el caso anterior, se obtiene la regla de actualización de los umbrales de las neuronas de la capa oculta.

$$\omega_0^1(k+1) = \omega_0^1(k) + \eta \delta_j(k) \quad (4.42)$$

El método del descenso de gradiente, para un MLP como el descrito, es un algoritmo simple de optimización pero presenta una serie de inconvenientes. Se puede dar el caso del método alcance un mínimo local y por lo tanto no consiga minimizar la función hasta el valor global. Por otro lado, en las zonas donde la pendiente de la función de error sea casi nula, el número de iteraciones necesario para alcanzar el mínimo podría ser exageradamente grande.

Para resolver estos problemas se han presentado numerosas variantes y modificaciones del algoritmo de entrenamiento. Una de la variantes más comunes es aplicar un *factor de momentum* o *factor de inercia*. En este caso, se le añade a la función de actualización un nuevo término según la ecuación (4.43).

$$\omega_{k+1} = \omega_k - \eta \frac{\partial E}{\partial \omega_k} + \beta_m (\omega(k) - \omega(k+1)) \quad (4.43)$$

En general, la adición de este parámetro disminuye el número de iteraciones necesarias para alcanzar la solución y tiende a evitar los mínimos locales. De esta manera, cuando

el incremento de pesos es alto, en la siguiente iteración lo será aún más. Por otro lado, en caso de que los incrementos oscilen entre positivo y negativo, el incremento efectivo final se reduce. En los últimos años [Bishop95, BS10], para evitar los problemas derivados de este tipo métodos, se han desarrollado diversos métodos de optimización de la función de error. En esta tesis se utilizará el algoritmo de gradiente conjugado escalado, considerado como un algoritmo de segundo orden en cuya formulación interviene la *matriz Hessiana*. Son métodos más robustos que aceleran la velocidad de convergencia, aunque aumentan la complejidad del modelo y la carga computacional.

#### 4.4.7. Algoritmo de gradiente conjugado escalado

El algoritmo de gradiente conjugado escalado, en inglés *scale conjugate gradient*, es también un algoritmo de aprendizaje supervisado cuya principal ventaja es que converge más rápido que el método de descenso del gradiente [Bishop95]. La diferencia más notable con respecto a este último es que no utiliza una única dirección para minimizar la función ni un valor constante del coeficiente de aprendizaje  $\eta$ . En cada iteración se busca la dirección óptima entre un conjunto de direcciones conjugadas y se calcula el valor óptimo del coeficiente de aprendizaje en esa dirección [BS10].

En general, el proceso de entrenamiento para la iteración  $k$  se inicia con la definición aleatoria del primer conjunto de pesos  $\omega(k)$  de las conexiones de la RNA y se evalúa el gradiente según dichos parámetros,

$$g(k) = \vec{\nabla} E \quad (4.44)$$

y se toma la dirección inicial  $d(k)$  de búsqueda según el mismo criterio que el algoritmo de descenso de gradiente,

$$d(\vec{k}) = -g(\vec{k}) \quad (4.45)$$

y el coeficiente de aprendizaje  $\eta$  queda definido según la ecuación (4.46),

$$\eta(k) = -\frac{d(\vec{k})^t g(\vec{k})}{d(\vec{k})^t H(k) d(\vec{k})} \quad (4.46)$$

donde  $H(k)$  representa la matriz hessiana en la iteración  $k$  definida según la expresión (4.47).

$$H(k) = \frac{\partial^2 E}{\partial \omega_i \partial \omega_j} \quad (4.47)$$

Así, en cada iteración se debe calcular y almacenar los valores correspondientes a la matriz hessiana, lo que conlleva mayor complejidad de cálculo. Una vez se dispone de estos parámetros se procede a actualizar los pesos según la ecuación (4.31). Con los nuevos pesos se estudia el resultado de la RNA con el grupo de pesos  $\omega(k+1)$  y se comprueba que se haya alcanzado el criterio de convergencia fijado. Si no se alcanza este criterio se deberá

realizar el cálculo del nuevo gradiente con los nuevos parámetros  $\vec{g}(k+1)$  y evaluar la nueva dirección de búsqueda usando las siguientes ecuaciones.

$$\vec{d}(k+1) = -\vec{g}(k+1) + \beta_m(k)\vec{d}(k) \quad (4.48)$$

$$\beta_m(k) = -\frac{\vec{g}^t(k+1)(\vec{g}(k+1) - \vec{g}(k))}{g^t(k)g(k)} \quad (4.49)$$

Este proceso se repetirá en cada iteración hasta alcanzar el criterio de convergencia. En esta tesis se ha utilizado este algoritmo según se ha implementado en la herramienta NETLAB en el entorno MATLAB [Bishop95, Nabney02].

#### 4.4.8. Técnicas de optimización de la arquitectura de la red

La elección de la red más sencilla posible capaz de resolver el problema es una de las decisiones más importantes, ya que ayuda a simplificar los cálculos y permite una mejor generalización del resultado. Algunos de los métodos más utilizados son los *métodos de poda*, que consisten en la eliminación de pesos o neuronas innecesarios durante el aprendizaje de la red. El algoritmo consistirá en realizar una elección de una arquitectura más compleja de la necesaria y que sea el algoritmo el que se encargue de podar la conexiones innecesarias. Además de permitir reducir la carga computacional, nos ayudará a encontrar posibles entradas que no aporten nada al cálculo de la salida, ya que los pesos correspondientes a dichas entradas serán *podados*.

Los métodos basados en aplicar términos de penalización son los métodos más simples y utilizados en la práctica. Se basan en incluir un término adicional de regularización  $E_\omega$  en la ecuación general del error (4.30), con el objetivo de eliminar los pesos innecesarios. Este término provoca un decaimiento en estos pesos de manera que se puede optar por eliminar aquellos que, al final del entrenamiento, queden por debajo de un umbral determinado.

$$S(\omega) = E_D(\omega) + \mu E_\omega(\omega) \quad (4.50)$$

$$E_\omega(\omega) = \frac{1}{2} \sum_{j=1}^m (\omega_j^2) \quad (4.51)$$

donde  $m$  representa el número de parámetros de la red y  $\mu$  es el parámetro de regularización o penalización, que determina la importancia de los términos de regularización. Utilizar el valor correcto de este parámetro es uno de los aspectos importantes a la hora de definir una RNA. Un valor pequeño del mismo puede conducir a un sobreajuste, en inglés *overfitting*, con lo que la función representaría correctamente los datos del conjunto de entrenamiento pero no sería capaz de ajustar un conjunto no conocido de datos, es decir, no se conseguiría la generalización de la solución, Figura 4.10. Mientras que si el valor es muy grande puede conducir a no ajustar los datos, en inglés *underfitting*.

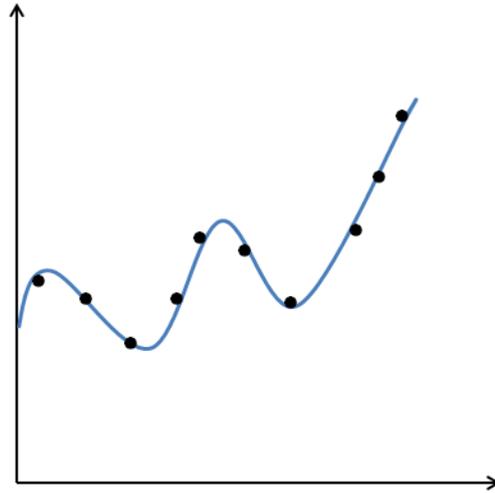


Figura 4.10: Función excesivamente ajustada al conjunto de datos mostrado (overfitting)

Para hallar el valor óptimo de este parámetro de penalización una técnica muy utilizada se conoce como la *técnica de validación cruzada de  $K$  iteraciones*. Este método es muy útil para realizar la elección de diversos parámetros de la RNA utilizando para ello un conjunto de datos de entrenamiento y otro de validación. De esta manera se consigue que la RNA realice el entrenamiento no solo con los datos de un conjunto sino que se le muestran otros datos diferentes. En este caso todo el conjunto de datos se debe dividir en  $K$  subconjuntos iguales respetando que cada uno de estos subconjuntos esté representado por toda la variedad de datos de la muestra.

## 4.5. Redes Neuronales Artificiales basadas en las teorías probabilistas Bayesianas. Redes Bayesianas

A la hora de construir un modelo basado en RNA, la arquitectura de la misma y los parámetros de entrenamiento son elementos básicos a la hora de conseguir resultados que eviten, por ejemplo, el overfitting. Las técnicas de validación cruzada consiguen resultados satisfactorios, pero obligan a realizar una separación del conjunto de datos. Estas técnicas presentan inconvenientes en procesos donde no se disponga de datos suficientes y por otro lado aumentan la carga computacional y la complejidad del modelo matemático. Mediante el uso de las teorías probabilistas Bayesianas se consigue resolver estos aspectos de manera automática durante el proceso de entrenamiento [Bishop95].

En esta tesis las teorías que se van a utilizar están basadas en la interpretación probabilista del algoritmo de retropropagación para el perceptrón multicapa realizado por Mackay [MacKay03, MacKay92]. En la teoría clásica se consigue la configuración óptima de pesos optimizando la función de error, mientras que la inferencia bayesiana considera una distribución de probabilidad de los pesos. Esta distribución de probabilidad se inicia a priori y, tras observar los datos, se estima la probabilidad posterior mediante el uso del

Teorema de Bayes. Esta distribución posterior se puede utilizar, entre otras cosas, para realizar predicciones con la RNA a partir de nuevos datos de las variables de entrada.

El teorema de Bayes relaciona la probabilidad a priori de un suceso  $P(h)$  con la probabilidad a posteriori del mismo suceso en base a unas observaciones  $P(h|O)$ . La probabilidad a priori se entiende como la probabilidad inicial sin conocer nada sobre el proceso. Posteriormente se obtienen una serie de observaciones relacionadas con el proceso que nos lleva a modificar esa probabilidad. Así, el teorema expresa la probabilidad condicional de una hipótesis aleatoria  $h$  dado unas observaciones  $O$ ,  $P(h|O)$ , según la distribución de probabilidad condicional de las observaciones  $O$  dada la hipótesis  $h$ ,  $P(O|h)$ , la probabilidad a priori de las observaciones  $P(O)$  y la probabilidad a priori de la hipótesis  $p(h)$ .

$$P(h|O) = \frac{P(O|h)P(h)}{P(O)} \quad (4.52)$$

La probabilidad condicionada de las observaciones dados los parámetros  $P(O|h)$  se conoce como la verosimilitud y la probabilidad  $P(h|O)$  es la probabilidad a posteriori de la hipótesis tras conocer los datos. La aplicación de esta teoría a las RNAs conduce a interpretar el resultado como la distribución más verosímil de las hipótesis  $h$ , siendo la solución deseada la que maximice la probabilidad a posteriori de la hipótesis en función de las observaciones  $P(h|O)$ .

#### 4.5.1. Aproximación probabilista del aprendizaje en RNA

La aplicación de la teorías bayesianas a las Redes Neuronales Artificiales lleva a considerar que el proceso de aprendizaje busca encontrar los pesos que maximicen la probabilidad de los datos disponibles en el conjunto de entrenamiento  $P(D|\omega)$ .

La distribución inicial de los pesos de la red neuronal se puede realizar basándose en una distribución de probabilidad a priori  $P(\omega)$ , la cual será modificada tras observar los datos obteniendo la probabilidad a posteriori de los pesos  $P(\omega|D)$ .

$$P(\omega|D) = \frac{P(D|\omega)P(\omega)}{P(D)} \quad (4.53)$$

La distribución de probabilidad a priori de los pesos se puede expresar como una distribución gaussiana con una varianza alta (4.54)

$$P(\omega|\alpha) = \frac{1}{Z_\omega(\alpha)} e^{(-\alpha E_\omega)} \quad (4.54)$$

donde  $\alpha$  representa la inversa de la varianza del conjunto de pesos y umbrales y se le conoce como un *hiperparámetro*, ya que controla la distribución de otros parámetros. Por otro lado,  $Z_\omega(\alpha)$  representa la constante de normalización de la función de probabilidad a priori. Para conseguir una interpretación probabilista del término regularizador de la función de error  $E_\omega$ , ecuación (4.50), y simplificar el análisis se ha elegido una distribución

gaussiana de los pesos a priori. Así, la constante de normalización se puede expresar según (4.55), donde  $m$  es el número total de parámetros de la RNA.

$$Z_\omega(\alpha) = \int e^{(-\alpha E_\omega)} d\omega = \left(\frac{2\pi}{\alpha}\right)^{m/2} \quad (4.55)$$

El razonamiento de la función de verosimilitud está ligada a la definición del ruido del modelo. El objetivo de una red neuronal es establecer una relación  $F$  entre el conjunto de entradas  $x_i$  y las salidas deseadas  $t_i$ , dado un conjunto de datos de entrenamiento con un número  $N$  de patrones o pares de entrada-salida,  $D = \{X_i, t_i\}_{i=1}^N$ . Tanto el fenómeno a estudiar como la relación obtenida están afectados por ciertas incertidumbres, por lo que dicha relación se puede expresar basada en la presencia de una desviación (4.56)

$$t_i = F(x_i) + \varepsilon_i \quad (4.56)$$

donde  $\varepsilon_k$  es una expresión de las desviaciones existentes debidas a la incertidumbre. De esta manera, para el patrón de entrenamiento  $i$ , se puede decir que la función  $F$  representada por la RNA obtiene la variable de salida deseada  $t_i$  a partir de un vector de entradas  $x_i$  más un ruido Gaussiano. Además, asumiendo que los errores tienen una distribución normal con media cero y varianza  $\sigma^2 = 1/\beta$ , la distribución del ruido queda (4.57).

$$P(\varepsilon_i|\beta) = \sqrt{\frac{\beta}{2\pi}} \exp\left(-\frac{\beta}{2}\varepsilon_i^2\right) \quad (4.57)$$

Teniendo en cuenta que la distribución de ruidos son independientes entre sí, la probabilidad conjunta de la totalidad de la muestra  $N$  se puede escribir (4.58).

$$P(\varepsilon|\beta) = P(\varepsilon_1, \dots, \varepsilon_N) = \prod_{i=1}^N P(\varepsilon_i|\beta) = \left(\frac{\beta}{2\pi}\right)^{N/2} \exp\left(-\frac{\beta}{2} \sum_{i=1}^N \varepsilon_i^2\right) \quad (4.58)$$

Teniendo en cuenta que el ruido se puede representar como la diferencia entre la salida deseada y la salida obtenida,  $\varepsilon_i = t_i - y_i$ , se puede obtener la función de verosimilitud de la forma (4.59),

$$P(D|\omega, \beta) = \frac{1}{Z_D(\beta)} e^{(-\beta E_D)} \quad (4.59)$$

de la misma manera,  $\beta$  es un hiperparámetro y se representa como la inversa de la varianza del ruido de las salidas deseadas. La interpretación gaussiana también nos permite calcular la constante de normalización  $Z_D$  según (4.60).

$$Z_D(\beta) = \left(\frac{2\pi}{\beta}\right)^{N/2} \quad (4.60)$$

Una vez conocidas la función de probabilidad a priori y la función de verosimilitud, se puede utilizar el teorema de bayes para inferir la probabilidad a posteriori [Lauret06a, Lauret08],

$$P(\omega|D) = \frac{P(D|\omega)P(\omega)}{P(D)} = \frac{1}{Z_\omega} e^{(-\alpha E_\omega)} \frac{1}{Z_D} e^{(-\beta E_D)} = \frac{1}{Z_s(\alpha, \beta)} e^{(-S(\omega))} \quad (4.61)$$

donde  $S(\omega) = \beta E_D + \alpha E_\omega$  y la constante de normalización  $Z_s(\alpha, \beta) = \int e^{(-S(\omega))} d\omega$ . El objetivo de la RNA es conseguir la distribución de pesos que maximice la función de probabilidad a posteriori,  $\omega_{MP}$ . Como las constantes de normalización se consideran independientes de los parámetros, se puede conseguir minimizando el logaritmo negativo de la ecuación (4.61), que es equivalente a utilizar la función (4.62).

$$S(\omega) = \frac{\beta}{2} \sum_{i=1}^N [y(x_i, \omega) - t_i]^2 + \frac{\alpha}{2} \sum_{j=1}^m \omega_j^2 = \frac{\beta}{2} E_D + \frac{\alpha}{2} E_\omega \quad (4.62)$$

Así, la teoría bayesiana nos conduce a una función  $S(\omega)$  que deberá ser minimizada durante el período de entrenamiento. Se observa que esta función (4.62) es similar a la función de error (4.50), teniendo en cuenta que el parámetro de regularización o penalización depende únicamente del ratio  $\eta = \alpha/\beta$  [Bishop95].

#### 4.5.2. Optimización bayesiana de los parámetros de control

Las técnicas bayesianas se utilizan como un método alternativo a las técnicas clásicas de optimización. Así, a partir del conjunto de datos de entrenamiento completo se consigue, no sólo optimizar los pesos de la RNA, sino obtener los parámetros de regularización óptimos. Además de controlar el sobreajuste, al aplicar las técnicas bayesianas se podrá estudiar la arquitectura de la red y la selección de las entradas importantes.

El valor óptimo de los hiperparámetros se estima durante el entrenamiento a partir de una suposición inicial. Con el conjunto de datos de entrenamiento se pretende hallar el conjunto de pesos óptimo,  $\omega_{MP}$ , mediante un proceso iterativo. De la misma manera y simultáneamente se podrán encontrar los valores óptimos de los hiperparámetros  $\alpha_{MP}$  y  $\beta_{MP}$ . A continuación se describe el proceso iterativo que se seguirá en esta tesis descrito por [Lauret08]:

1. Se inician los hiperparámetros  $\alpha$  y  $\beta$  con valores bajos. Con estos valores se estiman los pesos de la red mediante la distribución de probabilidad a priori. Estamos en la iteración  $k$ , con los pesos obtenidos  $\omega^k$  y los hiperparámetros definidos  $\alpha^k$  y  $\beta^k$  se puede estimar la función de error  $S^k(\omega)$ , ecuación (4.62).
2. Mediante un algoritmo de optimización como el gradiente conjugado escalado se obtiene el conjunto de pesos óptimos  $\omega_{MP}^{k+1}$  que minimiza el error  $S^k(\omega)$  realizando el número de iteraciones que determine el criterio de convergencia. Con estos pesos se calcula el valor de  $E_\omega^{k+1}$  y  $E_D^{k+1}$ .
3. En este paso se recalculan los hiperparámetros  $\alpha^{k+1}$  y  $\beta^{k+1}$  siguiendo los siguientes pasos:

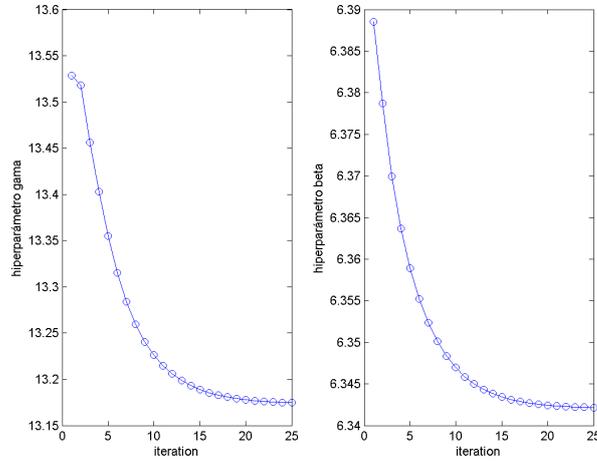


Figura 4.11: Convergencia de los hiperparámetros con cada iteración

- a)  $\gamma^{k+1} = \sum_{p=1}^m \left( \frac{\lambda_p}{\lambda_p + \alpha^k} \right)$  donde  $\lambda_p$  son los valores propios de la matriz Hessiana del error sin tener en cuenta el término regularizador,  $H = \beta^k \nabla \nabla E_D$ .
- b)  $\alpha^{k+1} = \frac{\gamma^{k+1}}{2E_\omega^{k+1}}$ .
- c)  $\beta^{k+1} = \frac{N\gamma^{k+1}}{2E_D^{k+1}}$

4. Se volverá al paso número 2 utilizando los parámetros  $\omega^{k+1}$ ,  $\alpha^{k+1}$  y  $\beta^{k+1}$  para calcular los nuevos valores hasta alcanzar el criterio de convergencia.

La convergencia de este procedimiento se produce cuando el error regularizado  $S_\omega$  es igual a la mitad del número de datos  $N/2$ . Así, cuando  $S_\omega = N/2$  entonces  $\alpha = \alpha_{MP}$  y  $\beta = \beta_{MP}$ . A través del estudio del valor de los distintos hiperparámetros en cada iteración  $k$  se puede observar también si ha convergido cuando el valor se vuelve constante, Figura 4.11.

### 4.5.3. Selección bayesiana de la arquitectura de una RNA

Las técnicas bayesianas permiten, por tanto, estimar el conjunto de pesos y de hiperparámetros óptimos. Además, atendiendo a la probabilidad de cada modelo, mediante este tipo de técnicas se puede aproximar el modelo óptimo. Esta probabilidad, también conocida como evidencia del modelo (evidence of the model), se puede utilizar para elegir entre una serie de redes neuronales distintas.

Partiendo de un conjunto de modelos  $M_i$ , las probabilidades a posteriori de los diferentes modelos dados los datos  $D$  se estiman siguiendo el teorema de Bayes

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{P(D)} \quad (4.63)$$

donde  $P(M_i)$  es la probabilidad a priori asignada a cada modelo  $M_i$ ,  $P(D)$  es la constante de normalización y  $P(D|M_i)$  es llamado la evidencia de cada modelo. Asumiendo una aproximación gaussiana de la distribución de pesos a posteriori se obtiene la expresión (4.64) calculando el logaritmo de la evidencia, conocido como *log of evidence*. En esta tesis, se entrenarán diferentes RNA con diferente número de neuronas ocultas  $h$  y aquella con mayor *log of evidence* será considerada la óptima para el conjunto de datos disponibles, Figura 4.12. En la bibliografía se encuentran diferentes fórmulas de *log of evidence* del modelo [Bishop95, MacKay03, Penny99] y la expresión finalmente elegida se expresa en la ecuación (4.64). En la práctica, se calcula el *log of evidence* para seleccionar un rango de número de neuronas ocultas más probable y se realiza el entrenamiento para comprobar cuál proporciona el menor error respecto al conjunto de datos del test para decidir la arquitectura final.

$$\begin{aligned} \log P(M_i|D) = & -\alpha_{MP} E_{\omega}^{MP} - \beta_{MP} E_D^{MP} - \frac{1}{2} \log |A| + \frac{m}{2} \log \alpha_{MP} + \\ & + \frac{N}{2} \log \beta_{MP} + \frac{1}{2} \log \left( \frac{2}{\gamma} \right) + \frac{1}{2} \log \left( \frac{2}{N - \gamma} \right) \end{aligned} \quad (4.64)$$

Donde  $N$  es el número de entradas a la red,  $m$  es el número total de parámetros a estimar,  $\gamma$  es el número de parámetros considerados necesarios al final de las iteraciones, es decir, el número de pesos que no se han aproximado a cero, y  $|A|$  es el determinante de la matriz Hessiana de la función de error total  $S(\omega)$ .

#### 4.5.4. Técnica bayesiana para la selección automática de entradas relevantes

En algunos casos, conviene estudiar el número de entradas a la red que son relevantes para resolver el problema. Un gran número de entradas puede elevar la complejidad del modelo y la carga computacional, pues implica arquitecturas de gran tamaño y conectividad. Las técnicas bayesianas, durante las iteraciones del proceso de entrenamiento, conducen a valores próximos a cero los pesos cuyas conexiones se consideran irrelevantes para la resolución del problema.

Aún así, puede ser conveniente eliminar algunas de estas conexiones para simplificar el cálculo. La técnica bayesiana de selección automática de entradas relevantes, en inglés se conoce como *Automatic Relevance Determination (ARD)*, ofrece una forma simplificada para determinar las entradas importantes [Bishop95, MacKay03, Lauret08]. Esta consiste en asignar el mismo hiperparámetro de regularización a todos los pesos asociados a una entrada, en lugar del mismo a toda la red, de manera que cada conjunto de pesos se encuentra definido por este valor. Este hiperparámetro de grupo  $\alpha_g$  es una distribución gaussiana de media cero y varianza  $1/\alpha_g$ . Al final de las iteraciones de entrenamiento, los pesos con valor alto de este parámetro se suponen que están próximos a cero. En este caso,

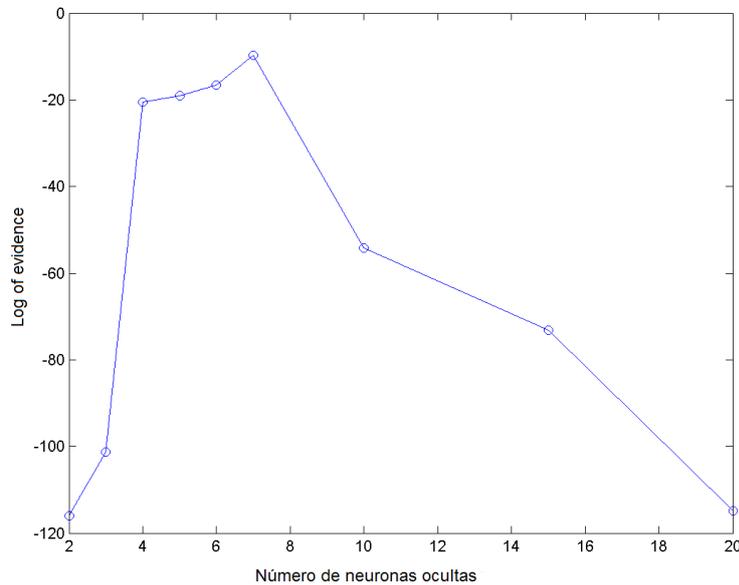


Figura 4.12: Resultados de la técnica de log of evidence para decidir el número de neuronas ocultas óptimo para resolver un problema. Ejemplo de la estación C6 en Gran Canaria.

la entrada correspondiente al conjunto de pesos regido por el hiperparámetro en cuestión se considera no relevante para la red y puede ser eliminada.

En la práctica se muestran figuras de barras como la Figura 4.13, donde cada barra muestra el valor de la varianza de los hiperparámetros, es decir la inversa del valor del hiperparámetro  $\alpha_g$ . Las entradas que corresponden con una barra que muestre un valor bajo con respecto al resto de entradas se considera irrelevante y se puede probar a realizar una poda de dicha entrada. Esta técnica muestra una estimación de la probabilidad de la importancia de cada entrada. Así que se recomienda comprobar si el error disminuye al realizar una *poda* de las entradas señaladas como poco relevantes.

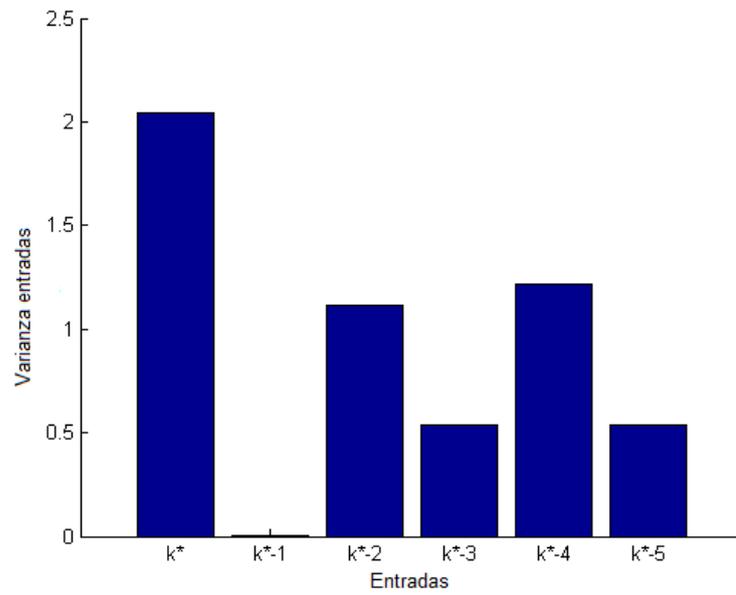


Figura 4.13: Resultados de la técnica de selección automática del número de entradas necesarias para resolver un problema. Ejemplo de la estación C5 en Gran Canaria.



## Capítulo 5

# Aplicación de los modelos de predicción de Radiación Solar

### 5.1. Introducción

En este capítulo se expondrán los resultados obtenidos en la predicción de los valores de radiación solar horario para la isla de Gran Canaria. El objetivo de esta tesis es el estudio de distintos métodos de predicción de la radiación solar global horizontal con horizontes temporales desde  $h = 1$  hasta  $h = 6$  horas.

Los modelos de predicción elegidos para realizar este estudio son, como ya se ha comentado en el Capítulo 4, distintos métodos estadísticos. En un primer momento se estudiará la complejidad de dichos modelos para obtener, en cada caso, una solución óptima adaptada a nuestros datos. Se ha realizado el estudio de la complejidad de los modelos lineales Autorregresivos (AR), Autorregresivos de Medias Móviles (ARMA) y de las Redes Neuronales Artificiales (RNAs). Este estudio se realizó tomando como datos de partida los datos históricos de radiación solar recogidos por las estaciones de medida terrestres de la isla de Gran Canaria.

Posteriormente, se estudiará si las predicciones de radiación solar mejoran aportando diferentes datos de partida. De esta manera, además de los datos históricos de radiación solar, se expondrán simulaciones de predicción a partir de datos de radiación satelitales, datos de radiación del modelo de predicción numérico ECMWF y datos de viento y humedad relativa del ECMWF. En este caso, el modelo estadístico utilizado serán las RNAs ya que aporta los mejores resultados de predicción.

Las predicciones de radiación solar a partir de métodos estadísticos necesitan disponer de suficientes datos históricos para realizar el *entrenamiento* de los modelos. En esta tesis se ha optado por trabajar con un año de medidas para realizar el *entrenamiento* y otro para comprobar el ajuste del modelo obtenido, *test*. Los errores finales que se utilizarán para comprobar la bondad de los modelos se obtendrán realizando las predicciones con el año del *test*. Con esto se pretende evitar el sobreajuste (overfitting) de los modelos a unos datos en concreto. Los años de medida de radiación solar, tanto para el entrenamiento como para el test, deben tener el menor porcentaje de huecos de medidas posible, ya

que se pretende tener un conjunto de datos continuo que represente todas las diferencias climáticas anuales. En todas las estaciones de medida terrestres disponibles en la isla de Gran Canaria se pueden establecer dos años de media completos para trabajar con los modelos estadísticos. El modelo numérico de predicción meteorológico ECMWF también ofrece la posibilidad de contar con los dos mismos años de medida en cada estación. Como se comenta en el Capítulo 2, entre los años 2000 y 2005 el ECMWF únicamente puede ofrecer una resolución temporal de 3 horas. Así, se disponen de las predicciones realizadas por el modelo numérico de predicción ECMWF para la isla de Gran Canaria cada tres horas para el día siguiente (horizonte de un día de adelanto).

Por otro lado, los datos satelitales se obtuvieron de la base de datos del **Helioclim-3**, en particular, de la **versión 5 (HC3v5)**. La base de datos **Helioclim-3** no dispone datos horarios para un año completo anteriores a 2005. Los datos de radiación terrestres disponibles en la isla de Gran Canaria pertenecen a años entre 2000 y 2005 dependiendo de la estación, ver Tabla 2.10. De esta manera, cuando se trabaje con datos provenientes de imágenes satelitales únicamente se puede utilizar el año 2005. Los datos de velocidad y dirección del viento y la humedad relativa de la atmósfera obtenidos del ECMWF también están disponibles únicamente para el año 2005 en la isla de Gran Canaria.

Así, se han realizado dos conjuntos de simulaciones con diferentes modelos estadísticos y diferentes variables de entrada. En un primer caso, se utilizarán dos años de medida para realizar el entrenamiento y el test y las variables de entrada se reducen a los datos históricos de radiación terrestre y los datos de radiación obtenidos de las predicciones del ECMWF. Los resultados de estas simulaciones se muestran en la Sección 5.2. Mientras que al incluir como variables de entrada los datos satelitales y los datos de viento y humedad del ECMWF se trabajará únicamente con el año 2005. En este caso, el año de medida se dividirá en un conjunto para el entrenamiento y otro para el test. Los resultados de estas simulaciones se muestran en la Sección 5.3.

Para evaluar el ajuste de las predicciones de cada método, se utilizarán las medidas de error estándar. En esta tesis se han elegido el Error Cuadrático Medio, en inglés **Root Mean Square Error (RMSE)**, y el Error Medio Absoluto, en inglés **Mean Absolute Error (MAE)**. Ambos errores han sido ampliamente referenciados en la bibliografía especializada [Lauret15, Willmott05], al igual que sus correspondientes medidas relativas (**rRMSE** y **rMAE**). Las medidas relativas se obtienen dividiendo el error absoluto entre la media del conjunto de datos de radiación solar de la zona en cuestión. En esta tesis, para decidir la calidad de los diferentes métodos de predicción se utilizarán los errores relativos. Además, para comparar la mejora relativa de los diferentes modelos con respecto al modelo más sencillo, *Persistence Model*, se ha calculado un parámetro llamado **%SKILL** para cada modelo [Coimbra13]. Un valor positivo y alto de este parámetro supone que dicho modelo mejora sustancialmente el *Persistence Model*, mientras que un valor negativo significa que el modelo simulado no mejora los resultados ya obtenidos.

$$RMSE_{modelo} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{I}_{g,i} - I_{measure,i})^2} \quad (5.1)$$

$$MAE_{modelo} = \frac{1}{N} \left| \sum_{i=1}^N (\hat{I}_{g,i} - I_{measure,i}) \right| \quad (5.2)$$

$$SKILL(\%) = \left( 1 - \frac{RMSE_{modelo}}{RMSE_{persistence}} \right) \times 100 \quad (5.3)$$

Donde  $N$  representa el número de datos horarios de la muestra sobre la que se está calculando el error,  $\hat{I}_{g,i}$  es el valor de radiación solar para la hora  $i$  calculado por el modelo de predicción,  $I_{measure,i}$  es el valor de radiación solar obtenido por las estaciones de medida terrestres y  $RMSE_{persistence}$  es el error obtenido por las predicciones realizadas con el modelo *Persistence* para dicha estación.

Las predicciones en los modelos estadísticos se realizarán con el índice de cielo despejado, ( $k^*$ ). Cuando se hayan realizado los procesos de entrenamiento de cada uno de los modelos, el cálculo y discusión de los errores se llevará a cabo en términos de irradiancias horarias en  $W/m^2$ , según  $\hat{I}_g = (k^*) \cdot I_{cs}$ .

## 5.2. Predicción de Radiación Solar a partir de datos históricos terrestres y del ECMWF

En esta Sección se describirán los cálculos de predicción de la radiación solar en la isla de Gran Canaria utilizando datos históricos y datos de radiación solar obtenidos por el modelo numérico ECMWF. En la Sección 2.3 se describieron las estaciones de medida utilizadas en esta tesis. Se pueden observar los años de medida que se utilizarán en cada caso, Tabla 2.10, y el porcentaje de datos mensuales disponibles. Todos los modelos estadísticos descritos en esta Sección se simularán con los datos disponibles de cada una de las estaciones de la isla:

- C0-Pozo Izquierdo
- C1-Las Palmas
- C2-La Aldea
- C4-Maspalomas
- C5-Sta. Brígida
- C6-Mogán

En un primer momento se estudiarán las predicciones únicamente utilizando la base de datos históricos de cada una de las estaciones. Como ya se ha comentado, para estudiar la complejidad del modelo y obtener el modelo óptimo para realizar las predicciones se utilizará el año definido como de *entrenamiento*. Mientras que la comprobación de los errores obtenidos por dicho modelo se obtendrán con el año definido como de *test*. En esta tesis se estudiarán los siguientes modelos estadísticos:

- Modelo *Persistence*
- Modelo *Smart-Persistence*
- Modelo *Climatológico*
- Modelo lineal Autorregresivo *AR*
- Modelo lineal Autorregresivo de Medias Móviles *ARMA*
- Modelo de Redes Neuronales Artificiales *RNAs*

### 5.2.1. Modelos lineales de predicción

Los modelos estadísticos lineales han sido utilizados en el modelado de series temporales durante años. En esta tesis se estudiará la capacidad de predicción de la radiación solar de los dos siguientes modelos lineales:

- *Modelos Autorregresivos (AR)*: Se trata de un modelo de regresión lineal basado únicamente en datos pasados de la serie temporal que se estudia. El orden de modelo viene definido por la variable  $p$ , que representa el número de parámetros de los que consta la función de transferencia.
- *Modelos Autorregresivos de medias móviles (ARMA)*: Se trata de un modelo formado por dos partes, una basada en el modelo autorregresivo (AR) y otra basada en el modelo de medias móviles (MA). La complejidad del modelo vendrá definida por los parámetros  $p$  y  $q$ , que representan el orden de la parte (AR) y (MA) respectivamente.

#### 5.2.1.1. Estudio de la complejidad de los modelos lineales

A la vista de lo anterior, parece obvio que estudiar la complejidad del modelo lineal que mejor se ajusta a nuestros datos es un primer paso muy importante. La variable  $p$  representará en ambos casos el orden del modelo AR, definiendo el número de parámetros del mismo. Por tanto, la definición de esta variable señalará el número de datos pasados de la serie temporal necesarios para establecer una relación lineal con el valor a predecir.

Como se ha explicado, la variable utilizada para trabajar con los modelos de cielo despejado es el índice de cielo despejado  $k^*$ . Así, los datos pasados de la serie temporal se refieren a los valores del índice de cielo despejado para las horas anteriores al momento en el que se quiere realizar la predicción  $t$ .

#### Modelo autorregresivo AR

El modelo autorregresivo para predecir un índice de cielo despejado en el horizonte temporal  $h$  respecto al momento actual queda definido con la ecuación 4.6.

$$\widehat{k}^*(t+h) = \sum_{i=0}^{p-1} [\Phi_{i+1} k^*(t-i)] + \epsilon_{t+h}$$

Los valores  $k^*(t-i)$  son los datos pasados de la serie temporal, mientras  $p$  representa el orden de AR. Por otro lado,  $\{\Phi_i\}_{i=1,2,\dots,p}$  muestra los parámetros de autorregresión obtenidos a partir de los datos de la muestra durante el entrenamiento.

Para tomar una decisión del modelo óptimo a utilizar, es decir el orden  $p$  del mismo, se utilizarán las gráficas de SPACF para identificar el máximo orden de los modelos AR. En cada estación se realizará la muestra de Autocorrelación Parcial **SPACF** para un conjunto de retardos de tiempo. La autocorrelación parcial PACF establece la correlación entre dos instantes de tiempo de la serie con un retardo determinado. De esta manera, para identificar los valores de SPACF significativos y establecer el retardo a partir del cual la relación entre las variables es prácticamente nula se ha considerado que el 95% de los valores se encuentran dentro del intervalo  $(-1.96/\sqrt{N}, +1.96/\sqrt{N})$ . Utilizando este límite se establecerá para cada estación el valor máximo de retardo significativo, que representará el número de datos pasados necesarios para realizar una predicción óptima y, por lo tanto, el orden del modelo  $p$ .

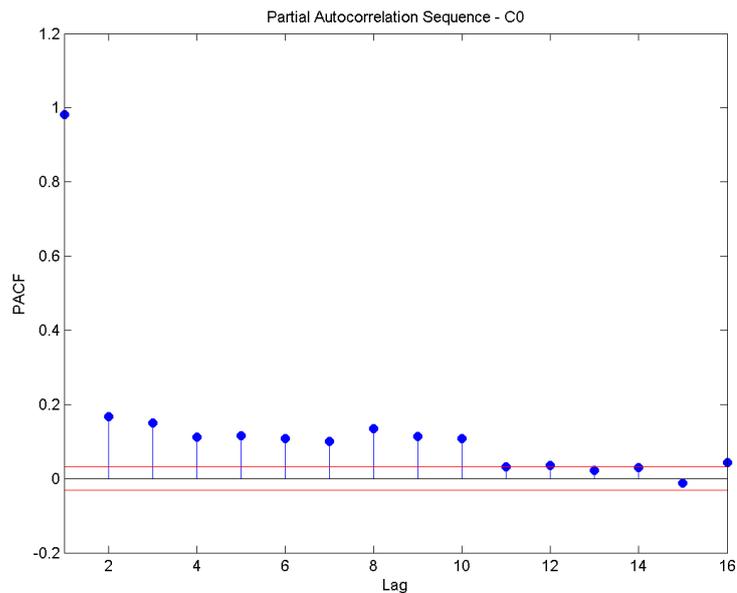


Figura 5.1: SPACF de la serie temporal del índice de claridad horario para la estación de C0-Pozo Izquierdo. En rojo se representan los intervalos de elección

Como se observa para las estaciones de C0 y C5, el primer valor tiene una importancia muy superior a los siguientes, cuyo valor de autocorrelación parcial va decayendo hasta superar el límite marcado, Figuras 5.1 y 5.2. En la estación de C0 el valor máximo se encuentra entorno al 10, mientras que en C5 se puede entender que el valor máximo está en 14. Para realizar unas simulaciones homogéneas en todas las estaciones se ha elegido como valor máximo del orden del modelo  $p = 14$ .

Estos valores de  $p$  no serán los valores finalmente utilizados para ajustar la serie, sino que se realizarán los cálculos del parámetro  $BIC$ , ecuación 4.15, para todos los

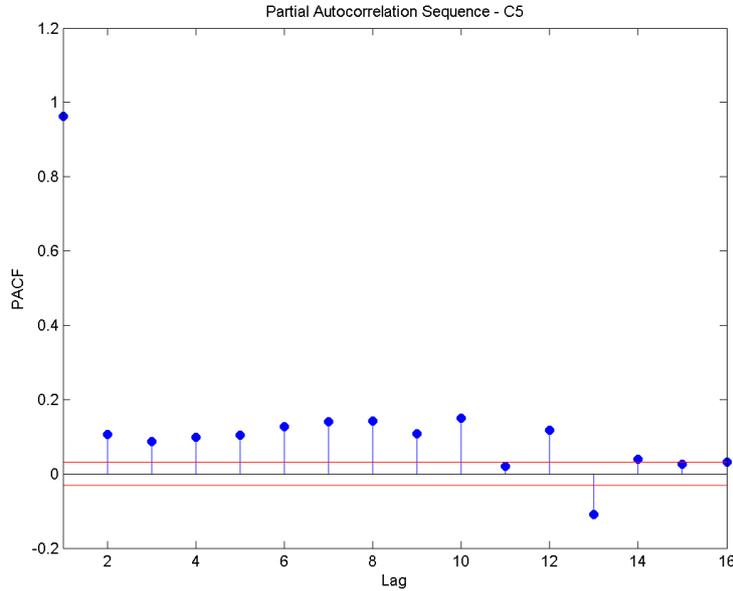


Figura 5.2: SPACF de la serie temporal del índice de claridad horario para la estación de C5-Sta. Brígida. En rojo se representan los intervalos de elección

órdenes,  $p = 1 \dots 14$ , hasta el máximo seleccionado para todos los horizontes temporales de predicción,  $h = 1 \dots 6h$ . El parámetro *BIC* (*Bayesian Information Criterion*) tiene en cuenta la complejidad del modelo a la hora de decidir el óptimo. El modelo óptimo a utilizar será el que muestre un valor BIC menor.

$$BIC = \ln(\hat{\sigma}^2) + m_k \frac{\ln(N)}{N}$$

Además del estudio del *Bayesian Information Criterion BIC*, se estudiaron los valores del error cuadrático medio relativo de modelo de ajuste **%rRMSE**. En muchas ocasiones, el modelo óptimo señalado por el BIC no ofrecía una mejora sustancial del error con respecto a un modelo más sencillo, en el que se utilizaban un menor número de valores pasados de la serie temporal.

Se puede observar en las Figuras 5.3 y 5.4, que el valor del parámetro *BIC* desciende significativamente conforme se aumenta el orden del modelo AR para todos los horizontes temporales. Este descenso se detiene entorno al orden  $p = 10$ , por lo que el modelo óptimo según este criterio se encontrará entorno a este orden. Dependiendo de la estación y del horizonte temporal se calculó el número de orden que proporcionaba el *BIC* más bajo. Por ejemplo, para estación C0-Pozo Izquierdo se obtuvieron los órdenes  $p = \{9, 9, 9, 10, 10, 10\}$  para los horizontes temporales de  $h = \{1, 2, 3, 4, 5, 6\}$  respectivamente.

Por otro lado, en las Figuras 5.5 y 5.6, se muestran los errores obtenidos en la predicciones realizadas por los distintos modelos AR. En este caso, se observa en las estaciones del sur, como C0, que la diferencia de error comienza a ser significativa a partir

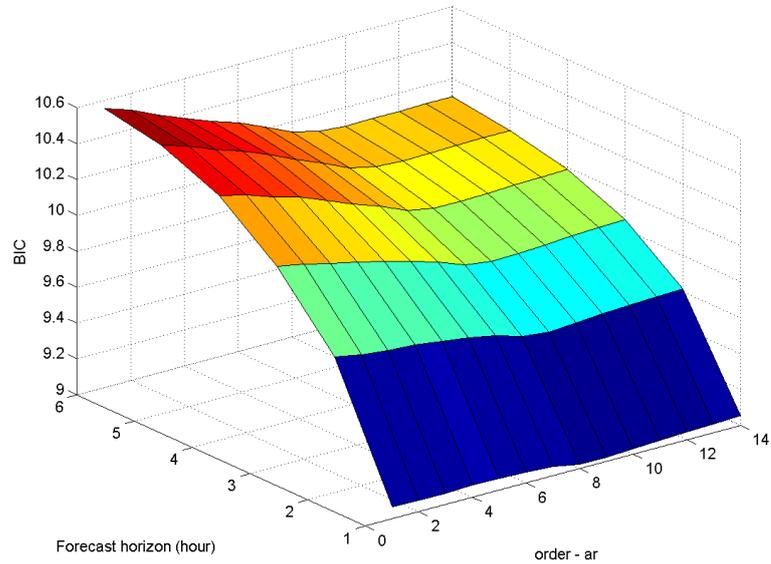


Figura 5.3: Valores del parámetro BIC obtenidos para la estación de C0-Pozo Izquierdo. Se muestran las simulaciones de un modelo AR con órdenes desde 1 hasta 14 para todos los horizontes temporales de predicción  $h = 1 \dots 6 h$

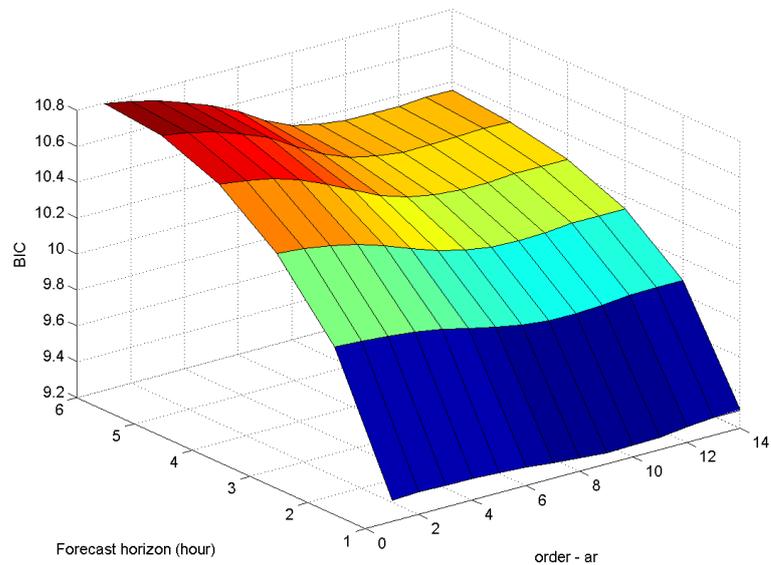


Figura 5.4: Valores del parámetro BIC obtenidos para la estación de C5-Sta. Brigida. Se muestran las simulaciones de un modelo AR con órdenes desde 1 hasta 14 para todos los horizontes temporales de predicción  $h = 1 \dots 6 h$

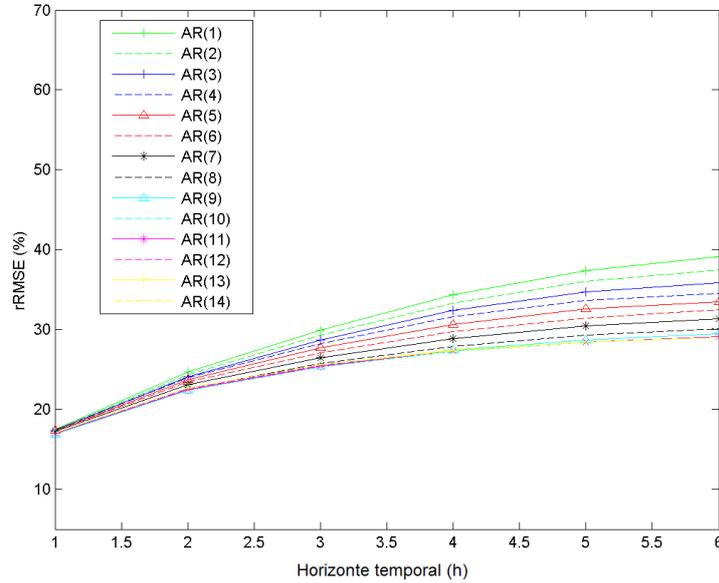


Figura 5.5: Error cuadrático medio relativo  $\%rRMSE$  para la estación de C0-Pozo Izquierdo. Se muestran las simulaciones de un modelo AR con órdenes desde 1 hasta 14 para todos los horizontes temporales de predicción  $h = 1 \dots 6$  h

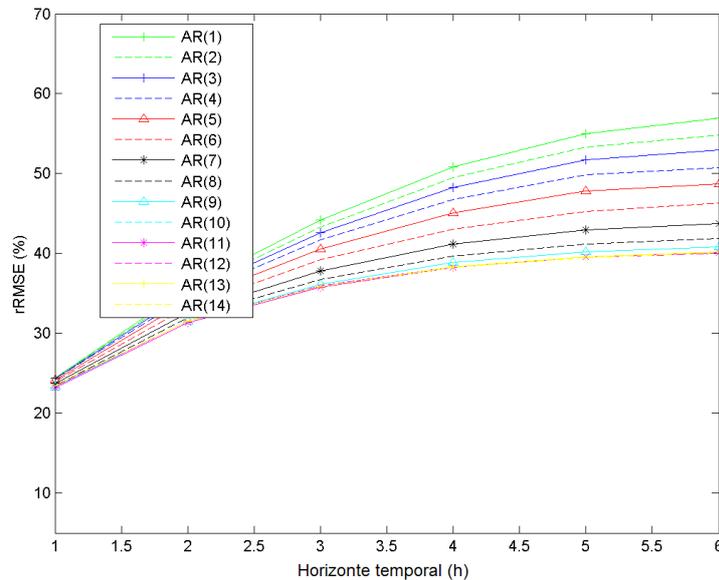


Figura 5.6: Error cuadrático medio relativo  $\%rRMSE$  para la estación de C5-Sta. Brigida. Se muestran las simulaciones de un modelo AR con órdenes desde 1 hasta 14 para todos los horizontes temporales de predicción  $h = 1 \dots 6$  h

de la tercera o cuarta hora, aunque el *BIC* haya señalado el orden 9 como el óptimo. Mientras que en las estaciones del norte, como C5, sí se empieza a observar una diferencia significativa desde la segunda hora de predicción.

Esto nos lleva a pensar que, atendiendo al *%rRMSE*, las mejoras introducidas por aumentar el orden del modelo AR no son significativas para las primeras horas de predicción. Se debe tener en cuenta que el orden del modelo representa el número de valores históricos necesarios para realizar la predicción. Un número alto implica la utilización de valores del índice de cielo despejado muy anteriores al momento que se desea predecir. Cuando se preparó la serie temporal de los datos disponibles se filtraron los valores nocturnos de radiación solar, por lo que un orden alto puede implicar utilizar valores de días anteriores. En cualquier caso, hasta que se compare con los resultados del modelo ARMA, se utilizarán como órdenes óptimos para cada estación aquellos que hayan proporcionado un *BIC* más bajo en cada estación y en cada horizonte temporal.

### Modelo autorregresivo de medias móviles ARMA

En el modelo Autorregresivo de Medias Móviles ARMA, los valores futuros de la serie temporal están basados en dos modelos básicos, un modelo Autorregresivo (AR) y un modelo de medias móviles (MA), como una combinación lineal de una cantidad determinada de valores pasados de la serie y de errores, según la ecuación (4.7).

$$\hat{k}^*(t+h) = \sum_{i=0}^{p-1} [\Phi_{i+1}k^*(t-i)] + \epsilon_{t+h} + \sum_{j=0}^{q-1} [\Theta_{j+1}\epsilon_{t-j}]$$

Los valores  $k^*(t-i)$  son los datos pasados de la serie temporal y  $\epsilon_t$  es una serie de ruido blando con media cero y varianza  $\sigma^2$ . En este caso, los parámetros del modelo son  $p$ , que representa el orden de AR o, lo que es lo mismo, el número de valores pasados utilizados,  $\{\Phi_i\}_{i=1,2,\dots,p}$ , que muestra los parámetros de autorregresión obtenidos a partir de los datos de la muestra durante el entrenamiento,  $q$ , que indica orden del modelo MA, y  $\{\Theta_i\}_{i=1,2,\dots,q}$  que representan los parámetros que acompañan a la serie de errores  $\epsilon_t$ .

Para tomar una decisión del modelo óptimo a utilizar, es decir los órdenes  $p$  y  $q$  del mismo. Al igual que para el modelo AR, se utilizarán las gráficas de SPACF para identificar el máximo orden de la parte autorregresiva AR del modelo, el orden  $p$ . Mientras que para decidir el orden máximo  $q$  de la parte de media móviles MA, se utilizarán las gráficas de SACF. En cada estación se realizará la muestra de Autocorrelación Simple **SACF** para un conjunto de retardos de tiempo. La Autocorrelación muestral SACF es una medida de la relación lineal que existe entre dos instantes cualesquiera dentro de la muestra. Por lo tanto, los valores significativos de la muestra de ACF representan el número de valores pasados que pueden ser necesarios para generar un modelo lineal. En el caso que nos ocupa, se puede considerar que el 95 % se encuentran dentro del intervalo  $[-1.96std(\rho_\tau), +1.96std(\rho_\tau)]$ . Utilizando este límite se establecerá para cada estación el valor máximo del orden del modelo  $q$ .

Como se observa para las estaciones de C0 y C5, Figuras 5.7 y 5.8, los valores de autocorrelación simple van decayendo hasta superar el límite marcado. Una vez analizadas

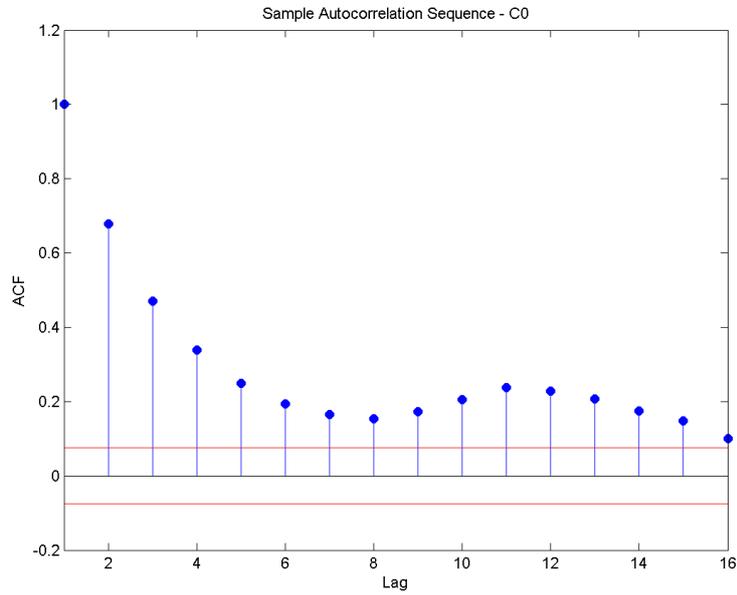


Figura 5.7: SACF de la serie temporal del índice de claridad horario para la estación de C0-Pozo Izquierdo. En rojo se representan los intervalos de elección

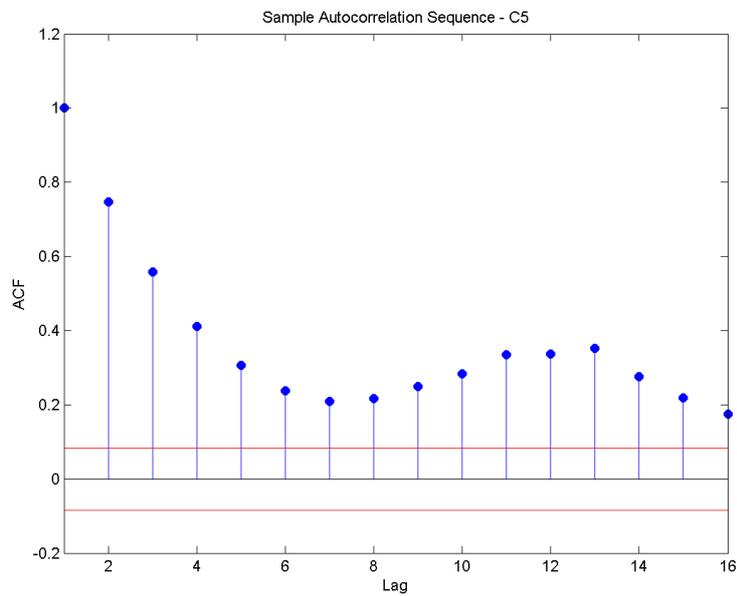


Figura 5.8: SACF de la serie temporal del índice de claridad horario para la estación de C5-Sta. Brígida. En rojo se representan los intervalos de elección

todas las estaciones de media se observó que los valores de autocorrelación dejaban de ser significativos para un orden de la parte MA  $q = 16$ . Mientras que, como ya se comentó observandolas gráficas de SACF, el orden máximo de la parte AR será  $p = 14$ .

Estos valores  $p$  y  $q$  no serán los valores finalmente utilizados para ajustar la serie, sino que se realizarán los cálculos del parámetro BIC para todas las combinaciones posibles de modelos variando dichos parámetros  $\{1, 2, \dots, 14\}$  y  $\{1, 2, \dots, 16\}$ . El modelo óptimo a utilizar teniendo en cuenta la complejidad del mismo será el que muestre un valor BIC menor. Además del estudio del *Bayesian Information Criterion BIC*, se estudiaron los valores del error cuadrático medio relativo de modelo de ajuste **%rRMSE**.

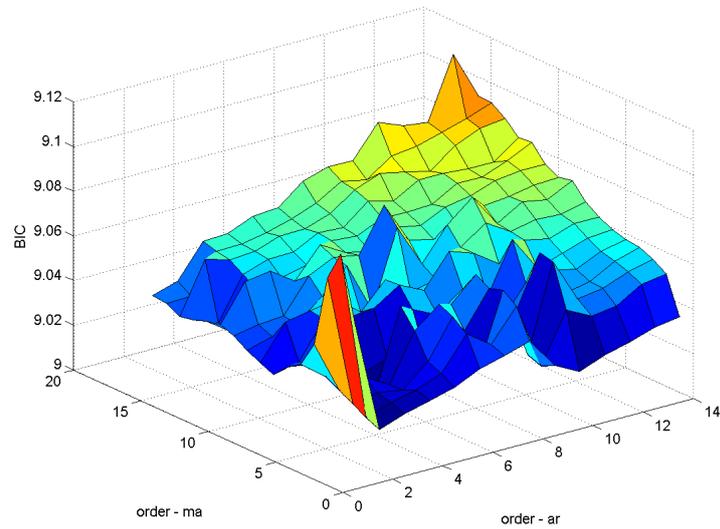
En las Figuras 5.9-5.12 se muestran los resultados de los valores de BIC y %rRMSE para las estaciones C0 y C5 respectivamente para los horizontes temporales de  $h = 1$  y  $h = 6$ . Una vez estudiadas todas las estaciones se observó que la información ofrecida por el BIC difería significativamente según la estación y el horizonte temporal. Así, para la estación C0 se obtuvo un modelo ARMA{2,1} para  $h = 1$  y un modelo ARMA{4,5} para  $h = 2, \dots, 6$ . Mientras que en la estación C5 se obtuvo un ARMA{7,4} para todos los horizontes temporales y para C1 se obtuvo un ARMA{11,1}. Comparando estos resultados con la información ofrecida por el %rRMSE, se observó que las mejoras proporcionadas por los modelos ARMA con órdenes altos no eran relevantes con respecto a un modelo ARMA{2,1} sencillo. Sobre todo para las primeras horas de predicción el error obtenido se mantiene bastante estable.

El orden  $p$  representa los valores pasados que se utilizan para realizar la predicción. Hay que recordar que un número alto implica la utilización de valores del índice de cielo despejado muy anteriores al momento que se desea predecir. En la serie temporal de los índices de cielo despejado este hecho puede implicar utilizar valores de días anteriores. En esta tesis se optó por tanto por simplificar la utilización del modelo ARMA y trabajar con órdenes  $p = 2$  y  $q = 1$ . Los resultados de este modelo fueron satisfactorios para todas las estaciones y de esta manera se utiliza un modelo común para todas.

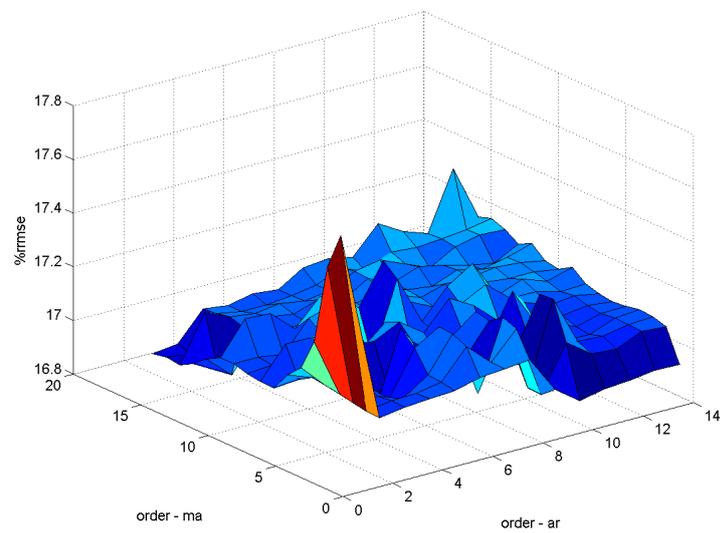
### 5.2.1.2. Resultados obtenidos con los modelos lineales

Una vez se estudiaron ambos modelos lineales se decidió realizar la comparación entre ambos antes de continuar con el resto de modelos. El modelo AR elegido será aquel cuyo orden corresponda con el óptimo ofrecido por el BIC para cada estación y cada horizonte temporal de predicción en horas. En la Tabla 5.1 se observan todos los órdenes para cada estación de medida. Por otro lado, como ya se ha descrito, el modelo ARMA elegido para todas las estaciones y con el que se realizará la comparación es aquel con órdenes  $p = 2$  y  $q = 1$ .

En la Figura 5.13, se puede observar la comparación entre el error %rRMSE cometido por el modelo AR y ARMA respectivamente. Este error ha sido calculado a partir del conjunto de datos del año denominado de *test*. En las primeras horas de predicción la diferencia entre ambos modelos es prácticamente imperceptible, pero el modelo ARMA se muestra más eficaz para los horizontes temporales desde  $h = 3$  hasta  $h = 6$ . Excepto en la estación C4 donde el resultado es similar durante todo el rango horario.

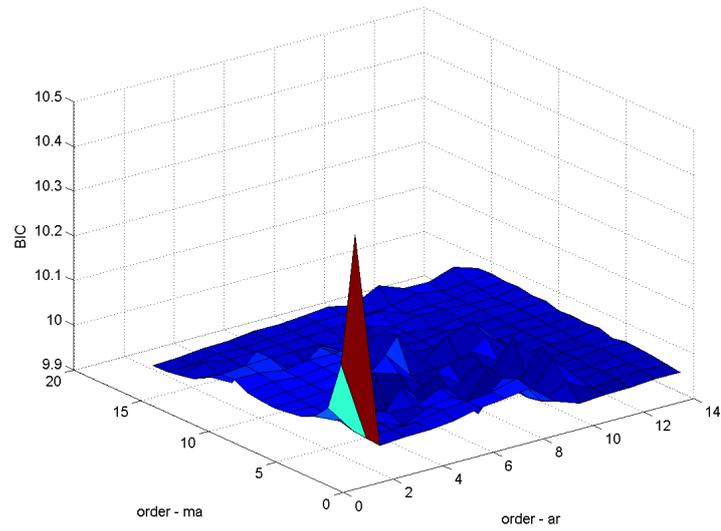


(a)

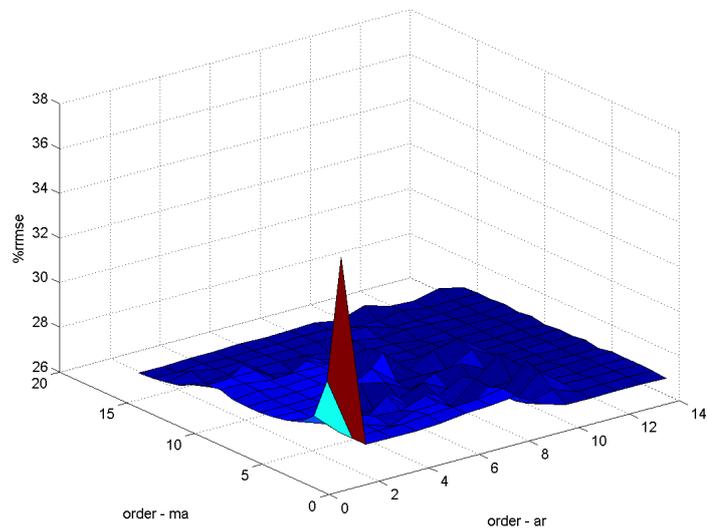


(b)

Figura 5.9: Error cuadrático medio relativo %rRMSE y BIC para la estación de C0-Pozo Izquierdo. Se muestran las simulaciones de un modelo ARMA para todas las combinaciones posibles de  $p$  y  $q$ . En cada parte se observa a) BIC 1h, b) %rRMSE 1h

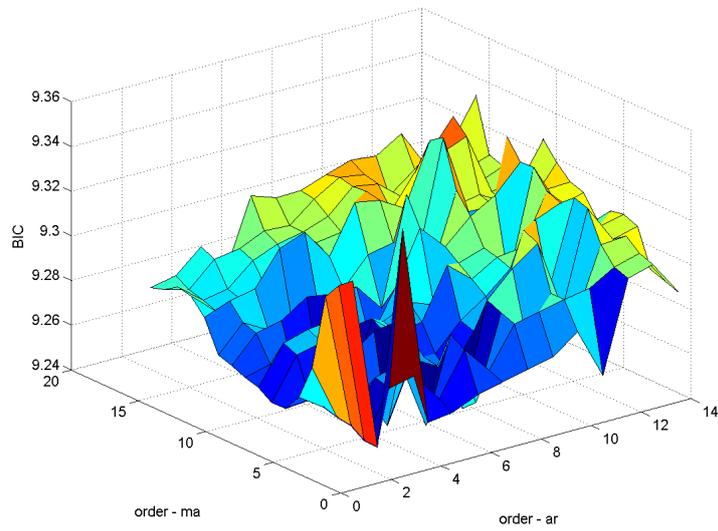


(a)

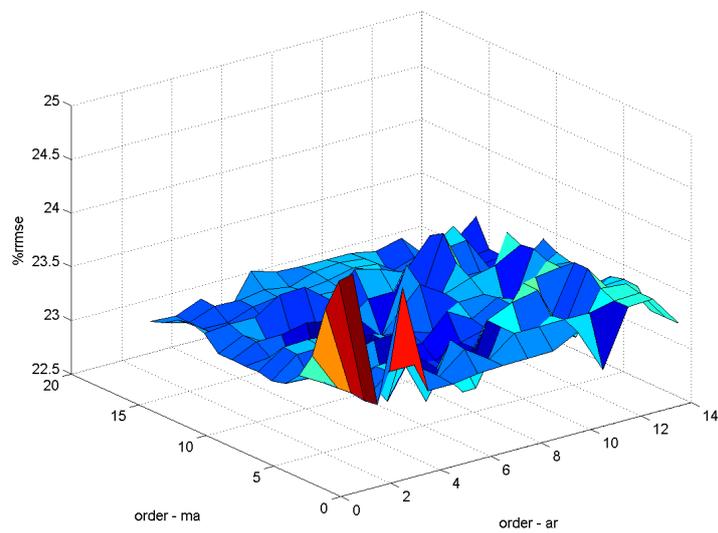


(b)

Figura 5.10: Error cuadrático medio relativo %rRMSE y BIC para la estación de C0-Pozo Izquierdo. Se muestran las simulaciones de un modelo ARMA para todas las combinaciones posibles de  $p$  y  $q$ . En cada parte se observa a) BIC 6h, b) %rRMSE 6h

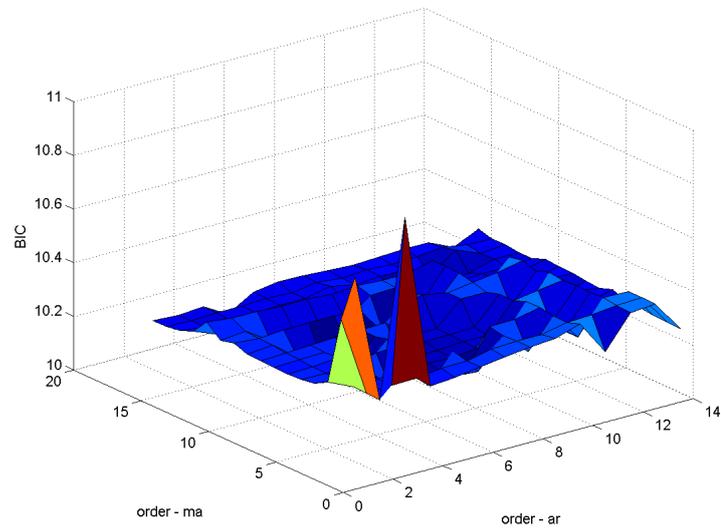


(a)

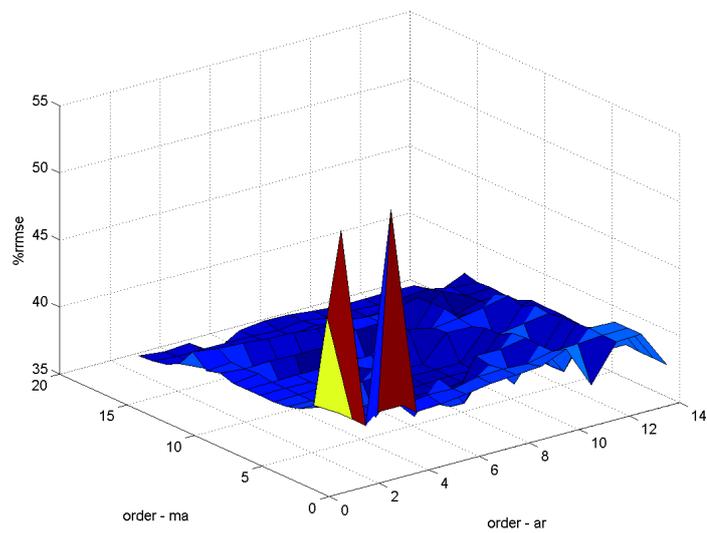


(b)

Figura 5.11: Error cuadrático medio relativo %rRMSE y BIC para la estación de C5-Sta. Brígida. Se muestran las simulaciones de un modelo ARMA para todas las combinaciones posibles de  $p$  y  $q$ . En cada parte se observa a) BIC 1h, b) %rRMSE 1h



(a)



(b)

Figura 5.12: Error cuadrático medio relativo %rRMSE y BIC para la estación de C5-Sta. Brígida. Se muestran las simulaciones de un modelo ARMA para todas las combinaciones posibles de  $p$  y  $q$ . En cada parte se observa a) BIC 6h, b) %rRMSE 6h

Estaciones	1 h	2 h	3 h	4 h	5 h	6 h
C0	9	9	9	10	10	10
C1	11	11	12	14	14	14
C2	8	8	9	11	11	11
C4	10	10	11	11	11	11
C5	9	9	10	10	10	10
C6	7	8	9	10	11	11

Tabla 5.1: Órdenes del modelo AR elegido para cada estación y horizonte temporal  $h$

Otro dato importante y que podemos ir observando desde este momento, es que en las estaciones C0, C2 y C6 los resultados de predicción son claramente mejores que en el resto de las estaciones. Estas estaciones se encuentran en el sur de la isla y poseen, como ya se describió en el Capítulo 2, un clima más estable y con un mayor número de días despejados.

A la vista de estos resultados, a partir de este momento se utilizará el modelo lineal ARMA para comparar con el resto de modelos de predicción. Los resultados obtenidos son ligeramente mejores que con el modelo AR utilizando un número menor de datos pasados de radiación solar.

### 5.2.2. Modelos de predicción basados en Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNAs) [Bishop95] son modelos matemáticos de predicción que encuentran una relación óptima entre un conjunto de datos de entrada y una salida deseada. La relación se establece mediante un proceso de *entrenamiento*. Las RNAs están compuestas por unidades, llamadas neuronas, que reciben una entrada de otra neurona o de una fuente externa. Todas las conexiones entre neuronas se representan con un peso asociado que se irá modificando durante el entrenamiento.

Cada neurona recibe una entrada afectada por esos pesos y realiza la suma de todas las entradas para producir una nueva salida hacia la neurona siguiente. La suma se ve afectada por una *función de activación* o *función de transferencia* para limitar la salida de cada neurona. En esta tesis, todas las neuronas se calcularán utilizando una función tangente hiperbólica.

El Perceptrón Multicapa (MLP) es una de las arquitecturas de RNAs más utilizada por la bibliografía. Las RNAs utilizadas durante este trabajo tienen la estructura de un MLP. En concreto, se optó por una capa de entrada, una capa intermedia o capa oculta y una capa de salida, sin conexiones de retroalimentación ni conexiones laterales entre las citadas capas. La capa oculta está formada por varias neuronas con una función de activación tangente hiperbólica. La capa de entrada representa los datos de la serie temporal de índices de cielo despejado, siendo  $T$  el número de datos pasados elegidos para predecir la salida, mientras que la capa de salida representa el índice de cielo despejado predicho para el horizonte temporal  $h$ ,  $\hat{k}^*(t+h)$ .

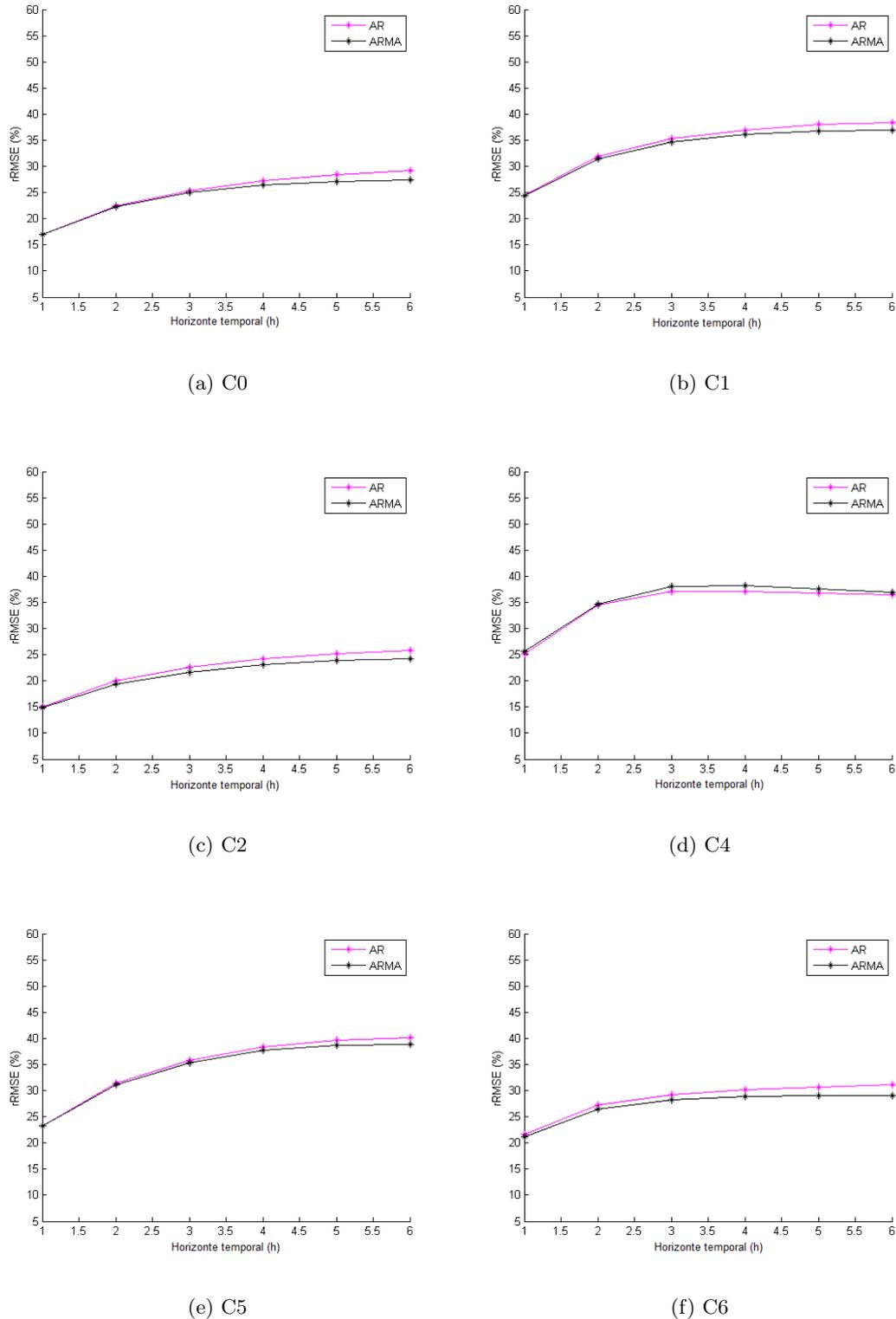


Figura 5.13: Error cuadrático medio relativo %rRMSE para todas las estaciones de medida y todos los horizontes temporales de predicción  $h = 1...6h$ . El modelo AR utilizado corresponde al óptimo señalado por el BIC, mientras que el modelo ARMA corresponde a un orden  $p = 2$  y  $q = 1$

$$\hat{k}^*(t+h) = \sum_{j=1}^H \omega_{sj}^2 f_j \left[ \sum_{i=0}^{T-1} (\omega_{ji}^1 k^*(t-i) + \omega_0^1) \right] + \omega_0^2 \quad (5.4)$$

Los pesos de la RNA se inician aleatoriamente y son luego optimizados durante el período de entrenamiento. Este proceso se realiza minimizando una función de error (*función de coste*) mediante el algoritmo de retroalimentación (*backpropagation*). La función de error más utilizada es función del error cuadrático medio entre la salida obtenida por la red,  $\hat{k}^*(t+h)$ , y la salida deseada  $k^*(t+h)$  (en nuestro caso el índice de cielo despejado medido en las estaciones), *target set*. La optimización de la RNA se realiza utilizando el conjunto de datos de entrenamiento.

La precisión de las RNAs para aproximar las series temporales depende en gran medida de la estructura de la red. En muchas ocasiones se consigue realizar una buena aproximación de los datos de entrenamiento, pero peores resultados cuando se utilizan datos nuevos (este problema se conoce como sobreajuste). Para evitar este problema se debe optimizar la estructura de la red utilizando también el conjunto de datos de test. En la bibliografía existen diversas técnicas de regularización para estudiar la complejidad del modelo [Bishop95, Lauret08, Lauret06b]. En esta tesis se ha optado por las técnicas Bayesianas de regularización de RNAs [MacKay03]. La aproximación Bayesiana considera una función de densidad de probabilidad en el espacio de los pesos. Los valores óptimos del conjunto de pesos corresponden a aquellos que hacen máximo el valor de la función de probabilidad. Las técnicas Bayesianas introducen dos nuevos *hiperparámetros*  $\alpha$  y  $\beta$  a la función de error para controlar la complejidad del modelo.

$$S(\omega) = \frac{\beta}{2} \sum_{i=1}^N [k_i^*(t+h) - \hat{k}_i^*(t+h)]^2 + \frac{\alpha}{2} \sum_{k=1}^m \omega_k^2 = \frac{\beta}{2} E_D + \frac{\alpha}{2} E_\omega \quad (5.5)$$

Para realizar todos los cálculos de entrenamiento y optimización de la complejidad del modelo se ha utilizado la herramienta NETLAB, desarrollada en el programa MATLAB para desarrollo de RNAs [Nabney02].

### 5.2.2.1. Estudio de la complejidad de las RNAs

Como se ha mencionado, uno de los aspectos más importantes a la hora de realizar una predicción es decidir la complejidad de la RNA. En el caso del MLP se controlará el número de entradas y el número de neuronas de la capa oculta. El número de entradas representa, en el caso de las predicciones de radiación solar, el número de datos pasados de radiación solar terrestre.

Las técnicas Bayesianas ofrecen diferentes posibilidades para estudiar la complejidad del modelo. Para valorar la importancia de cada entrada, se dividen los pesos en un grupo por cada entrada, un grupo para los pesos de la segunda capa y un grupo para los umbrales de la primera y segunda capa por separado. Cada grupo es controlado por un hiperparámetro independiente  $\alpha_g$ , así que todos los pesos relacionados al mismo grupo están afectados por el mismo hiperparámetro. Esta técnica, conocida como Automatic

Relevance Determination (ARD), asigna un coeficiente de regularización diferente a cada entrada y por lo tanto determina aquéllas que son más relevantes durante el proceso de entrenamiento. Al final del entrenamiento, los pesos con un gran valor de  $\alpha_j$  son cercanos a cero. Las entradas, cuyos pesos correspondientes son cercanos a cero, se consideran no relevantes para la RNA y podrían ser eliminadas.

Además, las técnicas Bayesianas también nos permiten elegir el número de neuronas óptimo. En este caso se calcula la probabilidad del modelo (*evidence of the model*) y se utiliza esta medida para elegir el mejor de los modelos propuestos. La bibliografía ofrece diferentes fórmulas para calcular el *log of evidence* del modelo [Bishop95, MacKay03, Penny99]. En esta tesis se utilizó finalmente la ecuación 4.64.

$$\begin{aligned} \log P(M_i|D) = & -\alpha_{MP} E_{\omega}^{MP} - \beta_{MP} E_D^{MP} - \frac{1}{2} \log |A| + \frac{m}{2} \log \alpha_{MP} + \\ & + \frac{N}{2} \log \beta_{MP} + \frac{1}{2} \log \left( \frac{2}{\gamma} \right) + \frac{1}{2} \log \left( \frac{2}{N - \gamma} \right) \end{aligned}$$

Primero, se estudió la información ofrecida por la técnica ARD para elegir el número de entradas óptimo para la RNA. El estudio se realizó con el número máximo de 6 entradas, igual al horizonte temporal máximo. Como se explicó, las técnicas Bayesianas asignan durante el entrenamiento, valores de cercanos a cero a los pesos que se consideran no relevantes para la red. En las Figuras 5.14 y 5.15 se pueden observar los resultados de la técnica ARD para todos los horizontes temporales en las estaciones C0 y C5. La información ofrecida no es constante para todos los horizontes temporales ni para los distintos números de neuronas ocultas utilizadas. En general, se puede considerar que las seis entradas elegidas aportan un grado de relevancia en algunos de los casos estudiados. En cualquier caso, se optó por *podar* aquellas entradas que la técnica ARD consideraba no relevantes en cada simulación y estudiar el error obtenido con el conjunto del test. Los resultados obtenidos indicaban que siempre era mejor utilizar las seis entradas y, durante el entrenamiento, permitir a las técnicas Bayesianas establecer la importancia de los pesos. Es decir, finalmente se utilizarán seis entradas de datos pasados del índice de cielo despejado y la RNA establecerá los valores de los pesos según la relevancia de las entradas mediante las técnicas Bayesianas.

Una vez quedó establecido el número de datos pasados que se utilizarán como entradas, se estudió el número de neuronas ocultas. Para cada estación y cada horizonte temporal de predicción se simuló varios modelos con distinto número de neuronas ocultas y la decisión se tomó en base al valor del *log of evidence* de cada modelo simulado. La mayoría de los resultados obtenidos conducen a RNAs con número de neuronas ocultas reducido. En la Figura 5.16 se observan los resultados para la estación C0 y los horizontes temporales  $h = 1$  y  $h = 3$ . En el primer caso se puede observar que el máximo se sitúa en 4 neuronas ocultas, mientras que en el segundo el máximo se sitúa en 3 neuronas, aunque con una diferencia muy pequeña con respecto a utilizar 2 ó 4 neuronas. Como ya se indicó en la Sección 4.5.3, la información ofrecida por el *log of evidence* se complementa con un estudio del %rRMSE de las predicciones con respecto al conjunto de datos del test. Así, a partir del resultado óptimo observado en la Figura 5.16, se eligen diversos modelos con un número de

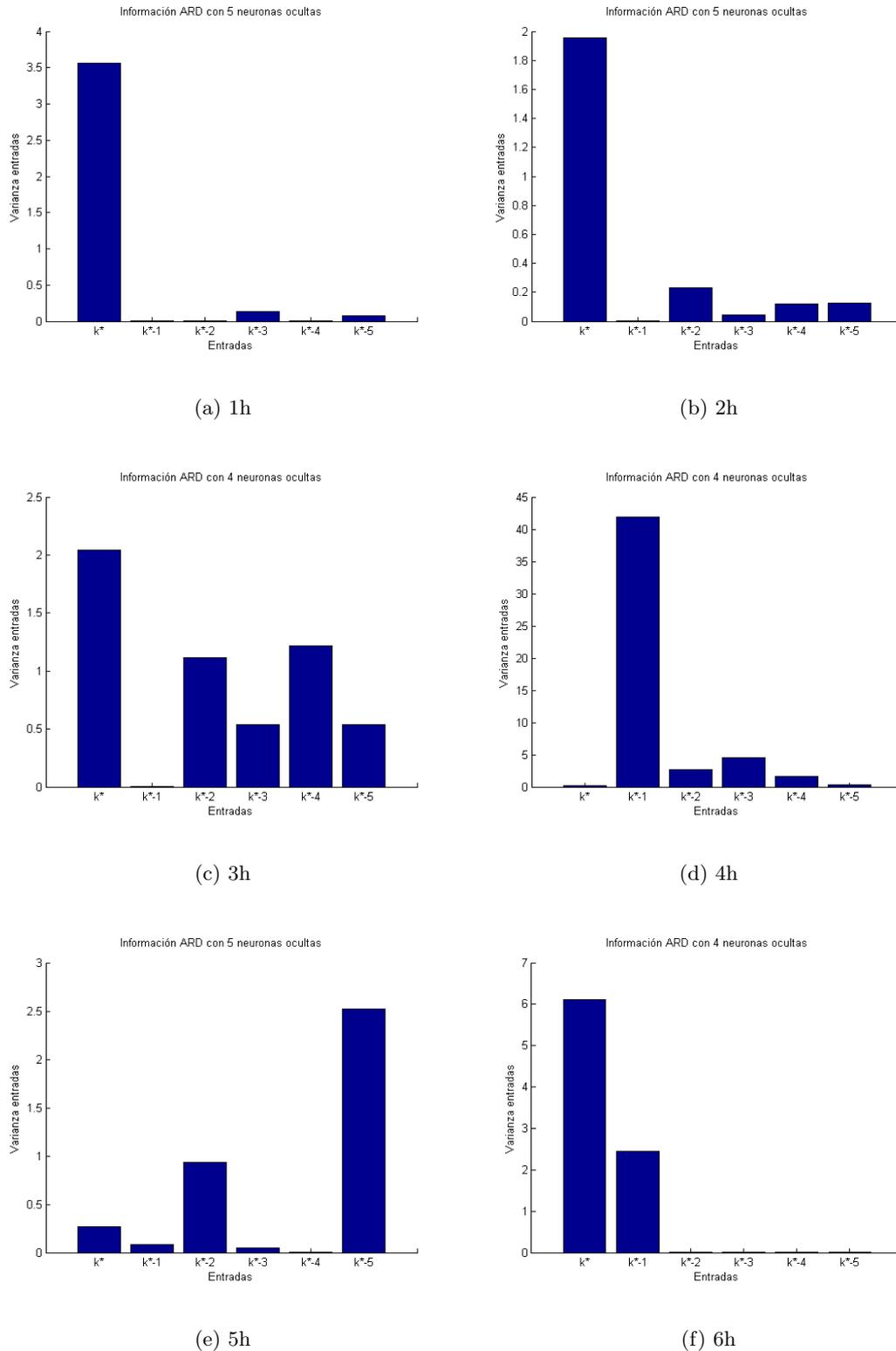


Figura 5.14: Resultado del ARD utilizando el número óptimo de neuronas para cada horizonte temporal en la estación de C0-Pozo Izquierdo. Cada entrada representa los datos terrestres pasados de la serie temporal

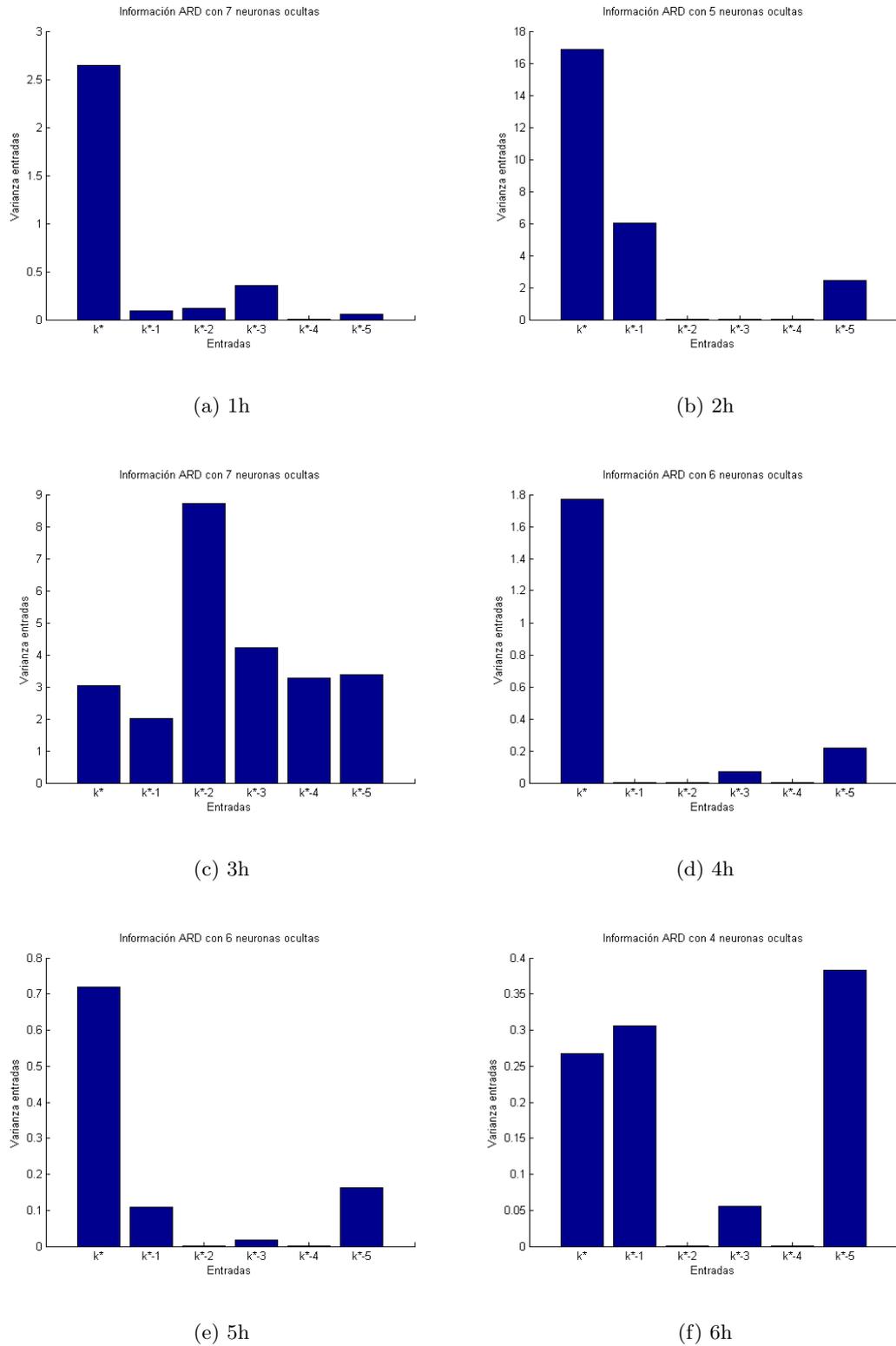


Figura 5.15: Resultado del ARD utilizando el número óptimo de neuronas para cada horizonte temporal en la estación de C5-Sta. Brígida. Cada entrada representa los datos terrestres pasados de la serie temporal

neuronas ocultas	%rrmse entrenamiento	%rmae entrenamiento	%rrmse test	%rmae test
2	14.96	9.74	17.37	11.91
3	14.86	9.65	17.45	12.00
4	14.74	9.47	17.39	11.62
5	14.56	9.19	17.36	11.40
6	14.55	9.23	17.58	11.73
7	14.60	9.35	17.67	11.75
10	14.42	9.09	17.95	11.73
15	14.03	8.71	18.75	11.94
20	13.60	8.46	20.07	12.15

Tabla 5.2: Errores del conjunto de datos de radiación terrestres de entrenamiento y de test utilizando distintos números de neuronas ocultas en la estación de C0-Pozo Izquierdo para el horizonte temporal 1 h.

neuronas alrededor del mismo. Con estos modelos se realiza el estudio del error %rRMSE para decidir el número final de neuronas ocultas.

Para la estación C0 y los horizontes temporales  $h = 1$  y  $h = 3$  se muestran los resultados obtenidos en las Tablas 5.2 y 5.3. Aunque los errores %rRMSE y %rMAE observados con el conjunto de datos del test no difieren significativamente, se eligió el número de neuronas cuya predicción ofrecía un menor error en cada caso. De esta manera, para un horizonte temporal  $h = 1$  se obtuvo una RNA con 5 neuronas y para  $h = 3$  se obtuvo una RNA con 4 neuronas.

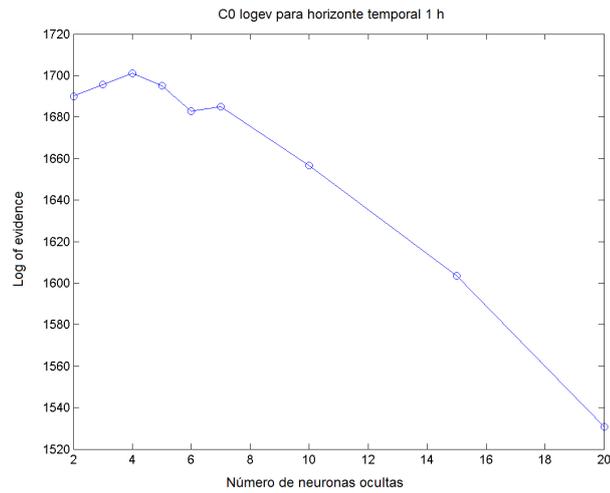
### 5.2.3. Resultados obtenidos y comparación entre los modelos

En este apartado se estudian los resultados obtenidos en las predicciones para los distintos horizontes temporales en cada estación de medida. El objetivo es encontrar una relación entre las variables de entrada, datos históricos del índice de cielo despejado, y el valor futuro a predecir. La función general que relaciona los parámetros de entrada y salida es la ecuación (4.2),

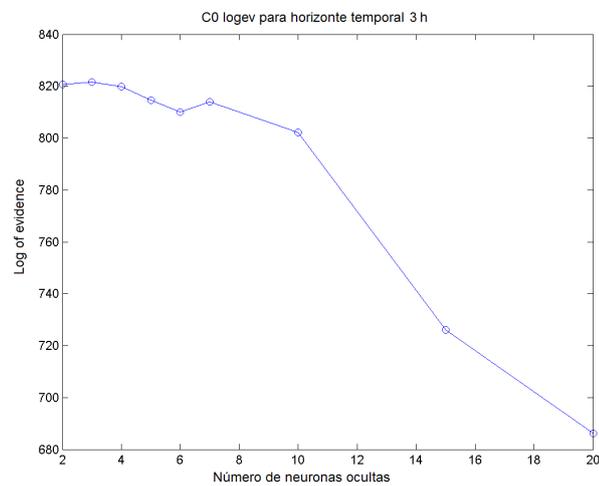
$$\widehat{k}^*(t+h) = F [k_g^*(t), \dots, k_g^*(t-i), k_{e1}^*(t), \dots, k_{e1}^*(t-j), \dots, k_{en}^*(t), \dots, k_{en}^*(t-j)]$$

donde  $\widehat{k}^*(t+h)$  es el índice de cielo despejado calculado para un horizonte de predicción  $h$ , en nuestro caso  $h = 1, 2, \dots, 6$  horas,  $k_g^*(t-i)$  es el índice de cielo despejado para las  $i$  horas pasadas obtenido a partir de los datos terrestres medidos en la estación y  $k_{en}^*(t-j)$  corresponde índice de cielo despejado para un las  $j$  horas pasadas obtenido a partir del dato exógeno  $en$ .

Si los modelos naïve o modelos de referencia predicen con mayor precisión que los modelos propuestos no resulta interesante desarrollar metodologías sofisticadas. De esta manera, en este trabajo, se propone comparar las predicciones obtenidas con las RNAs



(a) 1h



(b) 3h

Figura 5.16: Resultado del Log of evidence para decidir el número de neuronas óptimo en la capa oculta para la estación de C0-Pozo Izquierdo

neuronas ocultas	%rrmse entrenamiento	%rmae entrenamiento	%rrmse test	%rmae test
2	20.72	14.60	24.88	18.28
3	20.67	14.55	25.09	18.53
4	20.47	14.28	24.81	17.89
5	20.39	14.14	25.12	18.13
6	20.17	13.98	25.08	18.11
7	20.21	14.10	25.25	18.19
10	19.83	13.64	26.24	18.69
15	19.49	13.49	26.63	18.71
20	19.19	13.29	27.21	19.01

Tabla 5.3: Errores del conjunto de datos de radiación terrestres de entrenamiento y de test utilizando distintos números de neuronas ocultas en la estación de C0-Pozo Izquierdo para el horizonte temporal 3 h.

con dos modelos *naïve* y el modelo *climatológico*. Estos modelos únicamente utilizan las series temporales históricas de radiación solar como variables de entrada y su utilidad ha sido ampliamente descrita en la bibliografía [Lauret15, Diagne13, Lorenz12].

Además también se compararán los resultados con los errores obtenidos con el modelo de predicción numérico ECMWF para el día siguiente y el error cometido por la estimación de los datos satélites obtenidos por el Helioclim-3. Estos modelos realizan una estimación de los niveles de radiación independiente del horizonte temporal.

### Modelo Persistence

El primer modelo de referencia utilizado es el Persistence (Pers) definido según la ecuación (4.3). Este modelo asume que el índice de cielo despejado para cada horizonte temporal  $h$  sólo depende del valor en el instante  $t$ , lo cual implica que las condiciones meteorológicas se mantengan invariantes entre los tiempos  $t$  y  $t + h$ .

$$\widehat{k}^*(t + h) = k^*(t)$$

### Modelo Smart-persistence

Este modelo representa una sencilla mejora con respecto al Persistence y se denomina por su nombre en inglés, Smart-Persistence. El modelo consiste en predecir el índice de cielo despejado para cada horizonte temporal  $h$  como la media de los  $h$  valores previos del índice de cielo despejado. El modelo queda definido según la ecuación (4.4).

$$\widehat{k}^*(t + h) = \text{mean}[k^*(t), \dots, k^*(t - h)]$$

### Modelo Climatológico

El último modelo de referencia será el *Climatológico* [Lorenz12], el cual predice el valor del índice de cielo despejado para cualquier horizonte temporal  $h$  según el valor obtenido al realizar la media de todos los valores históricos anteriores de la serie temporal, ecuación (4.5). En este caso, el grupo de valores históricos corresponderá al conjunto de valores de entrenamiento.

$$\widehat{k}^*(t+h) = \text{mean}[k^*(t)]_{\text{training data}}$$

### Modelo autorregresivo de medias móviles ARMA

En la Sección 5.2.1.2 se expone el modelo ARMA con órdenes  $p = 2$  y  $q = 1$  como el mejor modelo lineal para todas las estaciones. La comparación con el resto de modelos de predicción se realizará con el citado modelo ARMA. En este caso, la estimación del modelo ARMA óptimo se estudiará a partir del conjunto de dato de entrenamiento. Los datos de entrada del modelo serán únicamente los índices de cielo despejado previos al momento de la predicción.

### Modelo basado en RNAs con datos históricos de radiación solar

Por último serán las RNAs serán el modelo de predicción estudiado para mejorar los datos obtenidos por los modelos anteriores. En la Sección 5.2.2.1 se explicó cómo obtener el número de entradas y el número de neuronas ocultas óptimos para realizar la predicción de datos del índice de cielo despejado. Con la RNA, en un primer momento, únicamente se partirá de datos históricos del índice de cielo despejado recogido por las estaciones terrestres y se denominará como (NN). En esta tesis, se ha considerado trabajar con seis datos previos al momento de la predicción como datos de entrada de la RNA. Las redes elegidas serán Perceptrones Multicapa con una única capa oculta.

### Modelo basado en RNAs con datos de radiación del ECMWF

Una vez se estableció un método para realizar una predicción de la radiación solar mediante RNAs se estudió la relevancia de añadir como variables de entrada otros datos. En esta Tesis, se ha trabajado con los datos obtenidos con el modelo numérico de predicción ECMWF. Así, en un primer momento, además de los seis datos pasados del índice de cielo despejado recogidos por las estaciones terrestres se añadirán las siguientes variables:

- Coseno del ángulo cenital.
- TCC Total Cloud Cover, se obtiene el índice de nubosidad entre 0 (sin nubes) y 1 (totalmente nuboso) derivado del estudio de todos los niveles de altitud.
- SSRD Surface Solar Radiation Downwards, en este dato se dispone de la radiación solar horizontal acumulada entre dos instantes sucesivos en  $J/m^2$ . Como en el caso de los datos terrestres se calculará el índice de cielo despejado para trabajar.

Las RNAs que utilizan datos históricos terrestres y datos del ECMWF se denominarán (NN+ECMWF). En todos las simulaciones realizadas durante esta Tesis con RNAs se

siguió el mismo procedimiento para estimar la complejidad del modelo descrito en la Sección 5.2.2.1. Se debe tener en cuenta que para la estación C4 uno de los años de medida de las estaciones terrestres es la primera mitad del año 2000, Tabla 2.10. En este año no se dispone de datos del modelo numérico ECMWF, por lo que para esta estación se realizará el entrenamiento con el año 2003 y el test con la segunda mitad del año 2005. Los resultados obtenidos en esta estación no serán por tanto concluyentes ya que únicamente se encuentran representadas las estaciones meteorológicas de la segunda mitad del año.

### 5.2.3.1. Resultados anuales y trimestrales

En esta Sección se presentan los resultados obtenidos por los diferentes métodos de predicción de radiación solar desde  $h = 1$  hasta  $h = 6$  horas de horizonte temporal. Los resultados se mostrarán para el conjunto de datos anual y para todas las estaciones de la isla de Gran Canaria. Los errores obtenidos han sido calculados con el conjunto de datos del test.

Además de los resultados obtenidos con el año completo de datos, para estudiar las variaciones estacionales de los modelos, se estudiaron las predicciones para cada trimestre del año. En este caso las estaciones se han dividido según el siguiente esquema:

- **Invierno**, incluye los meses de diciembre, enero y febrero.
- **Primavera**, incluye los meses de marzo, abril y mayo.
- **Verano**, incluye los meses de junio, julio y agosto.
- **Otoño**, incluye los meses de septiembre, octubre y noviembre.

Los resultados anuales para cada estación y cada horizonte temporal se pueden observar en las Tablas 5.4-5.6. Los resultados en función del RMSE nos ofrecen la dimensión del error en las unidades en las que se trabaja la radiación solar horaria, en este caso  $Wm^{-2}$ , y se puede comparar con la media del año 2005 en cada estación mostrada en las tablas. Por otro lado el %rRMSE nos permite evaluar el error relativo cometido por cada modelo comparado con la media anual. Por último, se estudiará el parámetro %SKILL, que nos permitirá establecer una relación de la diferencia de predicción de cada modelo con respecto al modelo Persistence. En este caso, un valor positivo indicará que el modelo de predicción simulado mejora la predicción realizada por el Persistence, mientras que un valor negativo indicará que dicho modelo no mejora este modelo de predicción sencillo y por lo tanto no representa una opción a seguir. En las Figuras 5.17 y 5.18 se pueden observar para cada estación la evolución gráfica de los errores relativos %rRMSE y %rMAE respectivamente.

En la Tabla 5.5 se pueden estudiar los errores anuales en términos de %rRMSE. Todos los modelos muestran en general un aumento en el error de predicción conforme aumenta el horizonte temporal, excepto el modelo Climatológico ya que su predicción se basa en el histórico de datos independientemente del horizonte temporal. En el modelo Persistence este aumento es el más considerable, incrementando el error entre la primera y la última hora entre un 15 % y un 20 % para las estaciones del sur, mientras que las estaciones del norte muestran un aumento entre un 25 % y un 30 %. Por otro lado, en los modelos ARMA

o los basados en RNAs (NN y NN+ECMWF) la diferencia del error entre la primera y la última hora es de aproximadamente un 10%. Por ejemplo, en la estación de C0-Pozo Izquierdo el %rRMSE para el mejor modelo de predicción (NN+ECMWF) aumenta con el horizonte temporal de predicción desde 17.00% para  $h = 1$  hasta 25.73% para  $h = 6$  y en C1-Las Palmas varía entre 23.99% y 33.71%.

De manera mayoritaria, en las Tablas 5.4-5.6 y en la Figura 5.17, se aprecia que el peor modelo de predicción para las tres primeras horas es el Climatológico, y para las últimas horas el modelo Persistence reproduce los peores errores de predicción. Esto es debido a que el modelo Climatológico permanece constante independientemente del horizonte de predicción. En todas las estaciones y para todos los horizontes de predicción los modelos ARMA y las redes neuronales con datos históricos terrestres (NN) muestran resultados similares. Las redes neuronales presentan mejores resultados sobre todo a partir de la tercera hora de predicción, mientras que en las dos primeras los resultados son muy similares e incluso algo mejor el ARMA para las estaciones de C0, C2 y C5. En cualquier caso, al utilizar las redes neuronales con los datos provenientes del modelo numérico ECMWF (NN+ECMWF), además de los datos históricos, los resultados son claramente mejores en cualquier estación y horizonte temporal de predicción.

Existe una gran diferencia entre las estaciones del sur y norte de la isla en referencia a los resultados anuales de predicción, recogiéndose diferencias de hasta un 12% para el horizonte temporal  $h = 6$ . Como ya se esperaba y se explicó en el análisis de los datos previos, la formación de nubes debidas a los vientos Alisios provoca unas variaciones meteorológicas más difícilmente predecibles en el norte. Además de observarse en los errores generales de predicción, esta diferencia también se denota en las mejoras obtenidas por los nuevos métodos de predicción.

Para las estaciones del sur como C0 y C2 principalmente, donde se han observado una mayoría de días despejados, el modelo NN+ECMWF mejora el modelo Persistence alrededor de un 1% y un 12% para los horizontes temporales  $h = 1$  y  $h = 6$  respectivamente. Sin embargo, las estaciones del norte de la isla como C1 y C5, donde existe una mayor formación de nubosidad, el modelo NN+ECMWF conduce a unas mejoras entre un 2% y un 17%.

La comparación del modelo NN+ECMWF con respecto a los modelos NN y ARMA en términos del %rRMSE muestra, para las estaciones del sur, una mejoría menor a un 1% para las tres primeras horas de predicción. Para los horizontes temporales de predicción a partir de la cuarta hora, esta diferencia va aumentando paulatinamente hasta alcanzar alrededor de un 1.4% con respecto al NN y un 2% con respecto al ARMA para  $h = 6$ . Por otro lado, en las estaciones del norte, la mejoría de este modelo muestra desde la segunda hora de predicción unos niveles entorno al 2%. De la misma manera que en el sur, esta mejoría aumenta progresivamente hasta alcanzar alrededor de un 2.5% con respecto al NN y casi un 4% con respecto al ARMA para  $h = 6$ . La mayor diferencia se observa en la estación de C5-Sta. Brígida, con un 4% respecto al ARMA, mientras que en C1-Las Palmas la diferencia se sitúa en el 3%.

Las diferencias entre los modelos se observan gráficamente en las Figuras 5.17 y 5.18 de manera muy sencilla. El modelo Persistence muestra, en general, los peores resultados a partir de la tercera hora y continua aumentando el error conforme se incrementa

estación	métodos	1 h	2 h	3 h	4 h	5 h	6 h
C0	Persistence	94.38	134.32	161.79	183.57	196.41	201.86
	Smart Persistence	94.38	135.02	157.48	166.13	163.32	159.14
	Climatológico	149.84	149.86	149.87	149.90	149.92	149.94
	ARMA	90.51	118.80	132.63	140.48	144.07	145.60
	NN	92.29	119.84	131.93	139.29	141.32	144.40
	NN+ECMWF	90.39	119.40	131.67	135.24	135.85	136.81
	media IGH = 531.74 $Wm^{-2}$						
C1	Persistence	114.92	160.44	187.67	205.28	217.33	221.53
	Smart Persistence	114.92	162.80	184.39	190.77	186.70	178.41
	Climatológico	166.27	166.28	166.29	166.30	166.31	166.31
	ARMA	107.79	139.32	153.00	159.36	162.46	163.17
	NN	107.06	139.08	152.69	157.08	158.35	158.20
	NN+ECMWF	106.05	133.99	143.33	146.89	148.91	148.98
	media IGH = 441.98 $Wm^{-2}$						
C2	Persistence	84.72	117.40	139.41	158.27	171.24	177.03
	Smart Persistence	84.72	118.73	139.44	149.26	150.79	148.30
	Climatológico	138.48	138.43	138.28	138.20	138.18	137.99
	ARMA	81.88	106.61	119.22	127.20	131.57	133.52
	NN	82.88	108.38	118.92	126.56	133.77	129.54
	NN+ECMWF	80.38	104.59	116.62	118.92	121.65	122.19
	media IGH = 551.93 $Wm^{-2}$						
C4	Persistence	123.32	192.77	238.61	257.06	262.51	255.64
	Smart Persistence	123.32	202.16	225.88	219.40	198.51	175.52
	Climatológico	152.75	152.60	152.43	152.51	152.58	152.65
	ARMA	113.01	153.15	167.84	168.46	166.16	162.96
	NN	108.16	138.44	144.56	146.12	146.01	147.38
	NN+ECMWF	101.24	129.78	144.21	140.19	142.36	141.30
	media IGH = 441.49 $Wm^{-2}$						
C5	Persistence	106.26	152.96	185.92	210.75	225.70	230.97
	Smart Persistence	106.26	157.34	185.45	194.52	190.90	184.06
	Climatológico	185.53	185.55	185.58	185.62	185.63	185.61
	ARMA	102.19	137.42	156.00	166.11	170.38	171.34
	NN	103.82	139.18	155.18	164.37	165.65	165.88
	NN+ECMWF	98.45	129.29	143.16	150.46	152.28	154.08
	media IGH = 440.93 $Wm^{-2}$						
C6	Persistence	113.29	159.74	186.47	203.24	207.51	207.82
	Smart Persistence	113.29	154.95	167.00	166.93	160.74	154.82
	Climatológico	156.04	156.05	156.05	156.06	156.08	156.08
	ARMA	104.22	129.81	138.41	141.67	142.43	142.87
	NN	99.52	125.23	135.51	138.56	138.44	140.81
	NN+ECMWF	91.88	111.61	119.73	122.27	124.36	127.19
	media IGH = 491.36 $Wm^{-2}$						

Tabla 5.4: Errores RMSE para todos los horizontes temporales  $h = 1...6$  en todas las estaciones de medida.

estación	métodos	1 h	2 h	3 h	4 h	5 h	6 h
C0	Persistence	17.75	25.26	30.43	34.52	36.94	37.96
	Smart Persistence	17.75	25.39	29.62	31.24	30.71	29.93
	Climatológico	28.18	28.18	28.19	28.19	28.19	28.20
	ARMA	17.02	22.34	24.94	26.42	27.09	27.38
	NN	17.36	22.54	24.81	26.19	26.58	27.16
	NN+ECMWF	17.00	22.45	24.76	25.43	25.55	25.73
	media IGH = 531.74 $Wm^{-2}$						
C1	Persistence	26.00	36.30	42.46	46.45	49.17	50.12
	Smart Persistence	26.00	36.83	41.72	43.16	42.24	40.37
	Climatológico	37.62	37.62	37.62	37.63	37.63	37.63
	ARMA	24.39	31.52	34.62	36.06	36.76	36.92
	NN	24.22	31.47	34.55	35.54	35.83	35.79
	NN+ECMWF	23.99	30.32	32.43	33.24	33.69	33.71
	media IGH = 441.98 $Wm^{-2}$						
C2	Persistence	15.35	21.27	25.26	28.67	31.03	32.07
	Smart Persistence	15.35	21.51	25.26	27.04	27.32	26.87
	Climatológico	25.09	25.08	25.05	25.04	25.04	25.00
	ARMA	14.84	19.32	21.60	23.05	23.84	24.19
	NN	15.02	19.64	21.55	22.93	24.24	23.47
	NN+ECMWF	14.56	18.95	21.13	21.55	22.04	22.14
	media IGH = 551.93 $Wm^{-2}$						
C4	Persistence	27.93	43.66	54.05	58.23	59.46	57.90
	Smart Persistence	27.93	45.79	51.16	49.70	44.96	39.76
	Climatológico	34.60	34.56	34.53	34.54	34.56	34.58
	ARMA	25.60	34.69	38.02	38.16	37.64	36.91
	NN	24.50	31.36	32.74	33.10	33.07	33.38
	NN+ECMWF	22.93	29.40	32.66	31.75	32.25	32.01
	media IGH = 441.49 $Wm^{-2}$						
C5	Persistence	24.10	34.69	42.17	47.80	51.19	52.38
	Smart Persistence	24.10	35.68	42.06	44.12	43.30	41.74
	Climatológico	42.08	42.08	42.09	42.10	42.10	42.09
	ARMA	23.18	31.17	35.38	37.67	38.64	38.86
	NN	23.55	31.56	35.19	37.28	37.57	37.62
	NN+ECMWF	22.33	29.32	32.47	34.12	34.54	34.94
	media IGH = 440.93 $Wm^{-2}$						
C6	Persistence	23.06	32.51	37.95	41.36	42.23	42.30
	Smart Persistence	23.06	31.53	33.99	33.97	32.71	31.51
	Climatológico	31.76	31.76	31.76	31.76	31.76	31.76
	ARMA	21.21	26.42	28.17	28.83	28.99	29.08
	NN	20.25	25.49	27.58	28.20	28.18	28.66
	NN+ECMWF	18.70	22.71	24.37	24.88	25.31	25.89
	media IGH = 491.36 $Wm^{-2}$						

Tabla 5.5: Errores %rRMSE para todos los horizontes temporales  $h = 1...6$  en todas las estaciones de medida.

estación	métodos	1 h	2 h	3 h	4 h	5 h	6 h
C0	Persistence	-	-	-	-	-	-
	Smart Persistence	0.0000	-0.52	2.67	9.50	16.85	21.16
	Climatológico	-58.79	-11.5864	7.35	18.34	23.67	25.72
	ARMA	4.10	11.55	18.03	23.48	26.65	27.87
	NN	2.21	10.77	18.45	24.10	28.03	28.44
	NN+ECMWF	4.22	11.1	18.61	26.317	30.81	32.20
	media IGH = 531.74 $Wm^{-2}$						
C1	Persistence	-	-	-	-	-	-
	Smart Persistence	0.00	-1.47	1.75	7.07	14.10	19.46
	Climatológico	-44.70	-3.65	11.39	19.00	23.50	24.95
	ARMA	6.21	13.16	18.47	22.37	25.25	26.34
	NN	6.88	13.34	18.66	23.49	27.15	28.61
	NN+ECMWF	7.75	16.52	23.64	28.45	31.49	32.77
	media IGH = 441.98 $Wm^{-2}$						
C2	Persistence	-	-	-	-	-	-
	Smart Persistence	0.00	-1.13	-0.03	5.69	11.94	16.23
	Climatológico	-63.50	-17.96	0.74	12.61	19.23	21.96
	ARMA	3.35	9.19	14.48	19.63	23.17	24.58
	NN	2.06	7.56	14.57	19.90	21.74	26.70
	NN+ECMWF	5.01	10.79	16.22	24.74	28.83	30.86
	media IGH = 551.93 $Wm^{-2}$						
C4	Persistence	-	-	-	-	-	-
	Smart Persistence	0.00	-4.87	5.34	14.65	24.38	31.34
	Climatológico	-23.93	20.77	36.04	40.62	41.86	40.30
	ARMA	8.36	20.55	29.66	34.46	36.70	36.25
	NN	12.32	28.25	39.51	43.27	44.51	42.48
	NN+ECMWF	17.93	32.74	39.66	45.57	45.89	44.85
	media IGH = 441.49 $Wm^{-2}$						
C5	Persistence	-	-	-	-	-	-
	Smart Persistence	0.00	-2.87	0.25	7.70	15.42	20.31
	Climatológico	-74.64	-21.35	0.15	11.90	17.72	19.60
	ARMA	3.83	10.16	16.09	21.18	24.51	25.82
	NN	2.24	8.95	16.47	21.94	26.54	28.12
	NN+ECMWF	7.30	15.41	22.94	28.54	32.47	33.24
	media IGH = 440.93 $Wm^{-2}$						
C6	Persistence	-	-	-	-	-	-
	Smart Persistence	0.00	3.00	10.44	17.87	22.54	25.50
	Climatológico	-37.77	2.28	16.29	23.20	24.78	24.89
	ARMA	8.01	18.73	25.78	30.29	31.36	31.25
	NN	12.15	21.59	27.31	31.79	33.25	32.21
	NN+ECMWF	18.90	30.12	35.78	39.81	40.04	38.77
	media IGH = 491.36 $Wm^{-2}$						

Tabla 5.6: Comparación con el modelo Persistence mediante el parámetro %SKILL para todos los horizontes temporales  $h = 1...6$  en todas las estaciones de medida.

el horizonte temporal. Con el Smart-Persistence se consigue mejorar los resultados al aumentar el horizonte de predicción porque cada vez utiliza mayor cantidad de datos pasados para estimar el valor futuro de radiación. Este modelo tiende a igualar el modelo Climatológico de predicción, como es obvio por su formulación. Los modelos ARMA y los basados en las redes neuronales en las primeras horas muestran una ligera mejora con respecto a los modelos naïve, sin embargo, conforme aumentan las horas esta mejoría muestra valores relevantes. Además, tanto el ARMA como las NN tienden a acercarse al modelo Climatológico conforme aumenta el horizonte de predicción, aunque siempre mostrando una mejoría. Al añadir los datos del ECMWF a las redes neuronales se consigue establecer una mejoría respecto a todos los modelos en todo el horizonte de predicción, incluyendo una mejoría respecto al modelo Climatológico para las últimas horas.

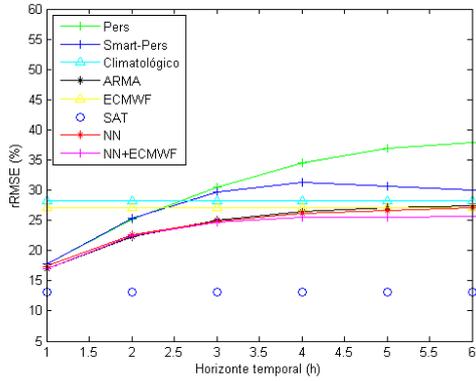
En la Tabla 5.6 se puede estudiar la mejoría de cada uno de los modelos con respecto al Persistence (%SKILL). En ella se observa que los modelos mejoran las predicciones respecto al Persistence (cuanto mayor sea el parámetro %SKILL) conforme aumenta el horizonte temporal de predicción, lo cuál quiere decir que la idoneidad de estos métodos respecto a los métodos sencillos es mayor conforme nos alejamos en el tiempo. A partir de la tercera hora de predicción ya no se observa ningún valor negativo, con lo que todos los modelos muestran mejores resultados que el Persistence.

Se concluye que tanto en las estaciones del sur como en el norte, los mejores resultados de predicción se consiguen con el modelo NN+ECMWF para todas las estaciones y todos los horizontes temporales. Por otro lado, aunque los resultados en cuestión de errores de predicción absolutos y relativos son mejores en las estaciones del sur, la mejoría respecto al resto de modelos es independiente de la estación de la isla. En general, en todas las estaciones de la isla el modelo NN+ECMWF ofrece una mejoría entorno al 30 %SKILL.

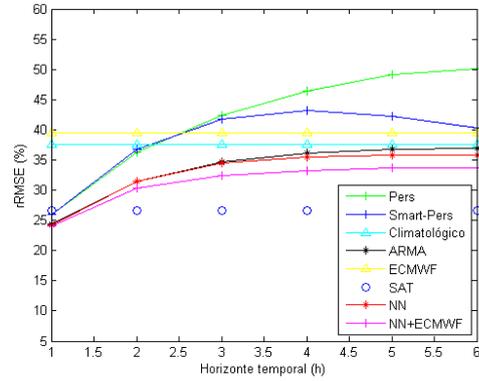
En el análisis de los datos de radiación se explicó la diferencia entre las estaciones de la isla de Gran Canaria según las estaciones del año. Además de los resultados de predicción anuales se decidió estudiar el ajuste de los modelos según la época del año. En las Tablas 5.7 y 5.8 se muestran los resultados de predicción para un horizonte temporal de  $h = 1$  y  $h = 6$  respectivamente. Los errores se presentan en términos del %rRMSE para el año completo (Anual) y para cada una de las estaciones del año (Invierno, Primavera, Verano y Otoño).

En general se puede observar que en las estaciones del sur del isla, C0 y C2, existe una diferencia en los resultados de la predicción entre los meses de verano y los meses de invierno. Así por ejemplo, utilizando el mejor método de predicción (NN+ECMWF), los valores de %rRMSE obtenidos en C0 para el verano oscilan entre 10-19% para todo el horizonte temporal de predicción, mientras que en invierno esta oscilación se produce entre 25-38%. En las estaciones del norte de la isla la diferencia entre el verano y el invierno no es tan acusada, debido a la formación de nubosidad durante los meses de verano por el efecto continuo de los vientos Alisios. En la estación de C1 por ejemplo, los errores para el verano oscilan entre 22-33% para todo el horizonte temporal y en invierno oscilan entre 27-37%. En los meses de verano, para la primera hora de predicción, el método de redes neuronales con datos del ECMWF consigue unos errores incluso menores del 10% en algunas estaciones del sur, mientras que en C1 el error no baja del 20%.

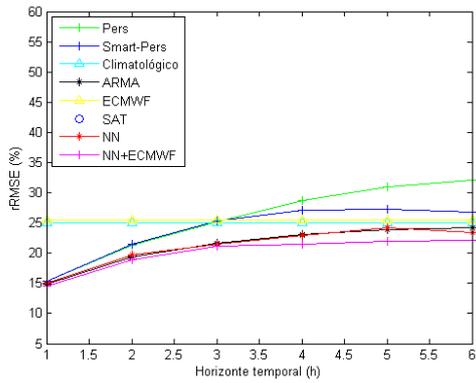
Entre las estaciones del norte de la isla, C1-Las Palmas y C5-Sta. Brígida, los errores



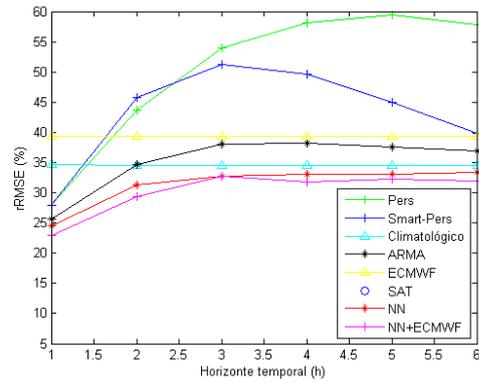
(a) C0



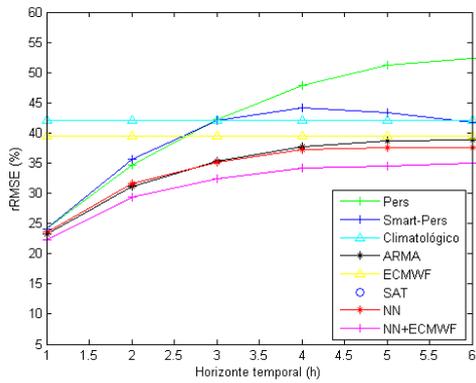
(b) C1



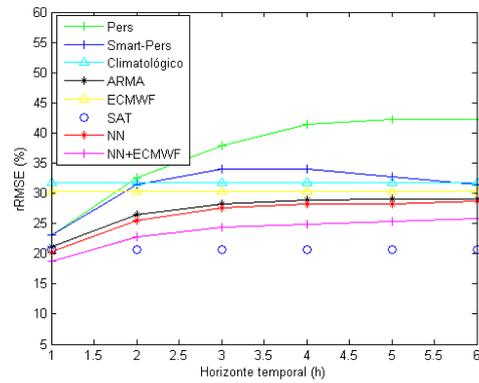
(c) C2



(d) C4

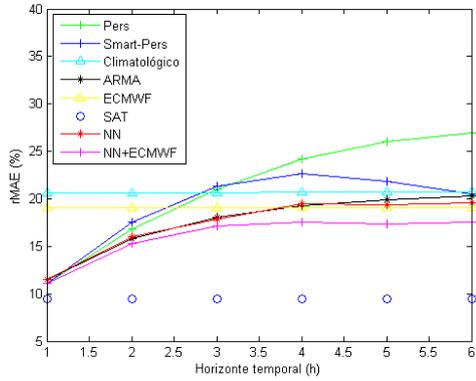


(e) C5

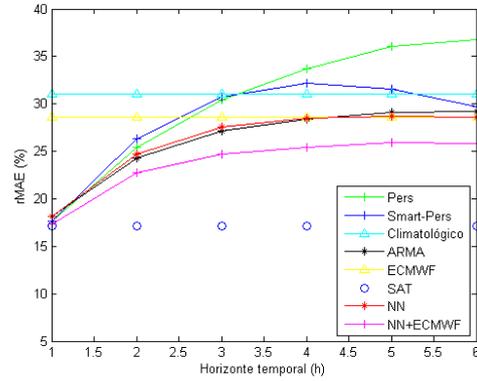


(f) C6

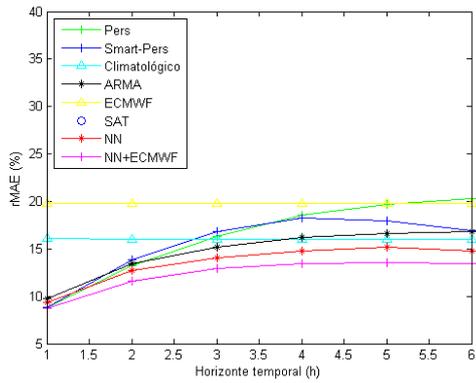
Figura 5.17: Evolución del %rRMSE en función del tiempo horizonte  $h = 1...6$  para cada modelo de predicción



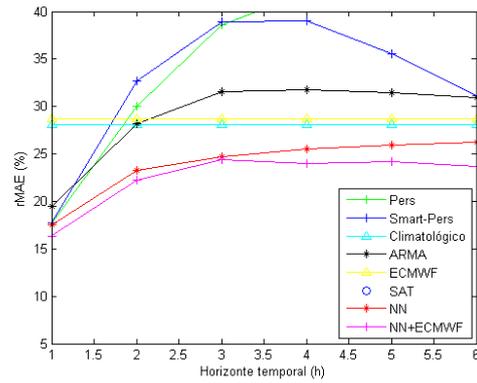
(a) C0



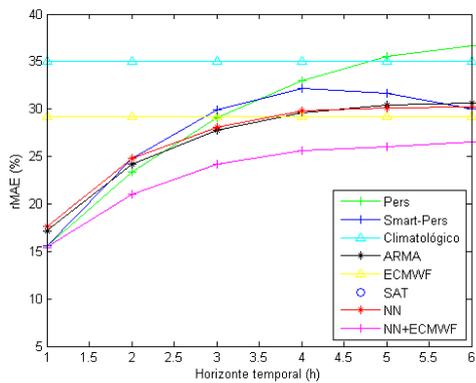
(b) C1



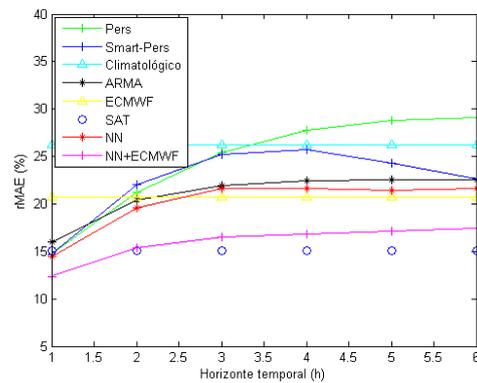
(c) C2



(d) C4



(e) C5



(f) C6

Figura 5.18: Evolución del %rMAE en función del tiempo horizonte  $h = 1...6$  para cada modelo de predicción

estación	métodos	Anual	Invierno	Primavera	Verano	Otoño
C0	Persistence	17.75	25.37	17.60	10.93	22.44
	Smart Persistence	17.75	25.37	17.60	10.93	22.44
	Climatológico	28.18	45.39	24.34	20.52	33.20
	ARMA	17.02	24.47	16.67	10.97	21.12
	NN	17.36	26.42	16.77	10.51	21.53
	NN+ECMWF	17.00	25.30	16.62	10.09	21.41
C1	Persistence	26.00	29.20	23.04	24.20	29.61
	Smart Persistence	26.00	29.20	23.04	24.20	29.61
	Climatológico	37.62	42.11	36.74	35.98	37.39
	ARMA	24.39	27.36	21.76	22.76	27.23
	NN	24.22	27.88	21.53	22.43	27.06
	NN+ECMWF	23.99	26.95	21.26	22.42	27.10
C2	Persistence	15.35	19.45	18.80	8.95	16.16
	Smart Persistence	15.35	19.45	18.80	8.95	16.16
	Climatológico	25.09	30.76	26.66	20.47	26.92
	ARMA	14.84	18.60	17.76	9.28	15.85
	NN	15.02	18.70	18.04	9.49	15.64
	NN+ECMWF	14.56	17.76	17.48	9.30	15.31
C4	Persistence	27.93	37.81	18.61	22.87	30.81
	Smart Persistence	27.93	37.81	18.61	22.87	30.81
	Climatológico	34.60	44.56	26.76	28.00	39.03
	ARMA	25.60	34.48	17.85	20.71	28.17
	NN	24.50	30.97	18.88	20.88	26.93
	NN+ECMWF	22.93	30.64	-	17.93	26.37
C5	Persistence	24.10	25.68	30.84	17.53	25.28
	Smart Persistence	24.10	25.68	30.84	17.53	25.28
	Climatológico	42.08	46.15	46.36	37.45	40.39
	ARMA	23.18	25.20	29.54	16.52	24.55
	NN	23.55	25.37	29.65	18.02	23.72
	NN+ECMWF	22.33	24.05	28.97	15.84	23.50
C6	Persistence	23.06	38.61	21.21	13.77	28.48
	Smart Persistence	23.06	38.61	21.21	13.77	28.48
	Climatológico	31.76	50.04	28.76	23.55	36.13
	ARMA	21.21	35.20	19.55	12.60	26.23
	NN	20.25	31.88	21.19	11.62	24.25
	NN+ECMWF	18.70	30.96	18.94	10.10	22.66

Tabla 5.7: Errores %rRMSE anuales y por trimestres para el horizonte temporal  $h = 1$  en todas las estaciones de medida.

estación	métodos	Anual	Invierno	Primavera	Verano	Otoño
C0	Persistence	37.96	44.59	34.69	32.35	46.90
	Smart Persistence	29.93	42.03	29.15	24.26	32.07
	Climatológico	28.20	45.59	24.34	20.47	33.37
	ARMA	27.38	39.63	25.24	20.57	32.40
	NN	27.16	42.83	24.44	19.65	31.40
	NN+ECMWF	25.73	38.17	23.84	19.06	29.78
C1	Persistence	50.12	53.22	45.80	50.19	51.49
	Smart Persistence	40.37	45.01	39.17	39.41	38.70
	Climatológico	37.63	42.24	36.78	35.80	37.52
	ARMA	36.92	41.06	34.60	35.31	37.25
	NN	35.79	40.23	34.72	33.88	35.67
	NN+ECMWF	33.71	36.57	31.86	33.02	34.23
C2	Persistence	32.07	36.44	38.52	23.27	31.42
	Smart Persistence	26.87	34.86	29.61	20.51	27.50
	Climatológico	25.00	30.10	26.74	20.51	26.40
	ARMA	24.19	29.88	26.84	18.84	24.78
	NN	23.47	28.27	26.41	18.34	23.52
	NN+ECMWF	22.14	26.63	25.05	16.25	23.43
C4	Persistence	57.90	57.93	36.60	59.71	54.12
	Smart Persistence	39.76	47.74	31.41	38.36	39.71
	Climatológico	34.58	44.83	26.81	27.91	39.25
	ARMA	36.91	44.93	26.32	32.72	38.66
	NN	33.38	47.62	27.90	26.03	37.83
	NN+ECMWF	32.01	40.27		26.98	35.56
C5	Persistence	52.38	53.38	60.90	45.28	52.11
	Smart Persistence	41.74	45.37	49.95	33.39	43.088
	Climatológico	42.09	46.23	46.50	37.52	40.44
	ARMA	38.86	44.31	46.65	29.80	40.21
	NN	37.62	42.10	43.71	30.64	38.78
	NN+ECMWF	34.94	37.07	41.78	27.67	37.58
C6	Persistence	42.30	58.17	40.06	33.08	49.22
	Smart Persistence	31.51	48.45	29.49	22.22	37.80
	Climatológico	31.76	50.05	28.77	23.49	36.32
	ARMA	29.08	46.59	26.86	18.28	35.73
	NN	28.66	45.23	27.07	19.07	34.72
	NN+ECMWF	25.89	43.26	24.16	15.82	31.57

Tabla 5.8: Errores %rRMSE anuales y por trimestres para el horizonte temporal  $h = 6$  en todas las estaciones de medida.

en verano tampoco siguen una misma tendencia, encontrándose unos errores claramente superiores en la estación C1 para todos los horizontes de predicción. Esto se debe a que la estación C5 se encuentra a mayor altitud sobre el nivel del mar (C1 está en la costa) y no se ve afectada en gran medida por la formación de nubes típica de esta época. Por otro lado, entre las estaciones del sur, C2 muestra unos errores similares en las estaciones de invierno, primavera y otoño, con errores entre el 15-18 % para  $h = 1$  y 23-26 % para  $h = 6$ , mientras que en la estación de C0 en invierno y otoño los valores del %rRMSE son superiores, superando incluso el 30 % para  $h = 6$ . Aunque ambas estaciones muestran un clima más estable que en el norte y con mayor número de días despejados, C2-La Aldea en el suroeste de la isla muestra un clima más estable debido a que se encuentra completamente opuesta a la dirección de los vientos predominantes de la isla.

La diferencia del mejor modelo de predicción (NN+ECMWF) respecto al modelo Persistence aumenta con el horizonte de predicción en las estaciones del sur. Para las primeras horas de predicción prácticamente no se consigue una mejora relevante, incluso en la estación C2 el modelo Persistence consigue resultados inferiores al 9 %. En el horizonte temporal  $h = 6$  para dichas estaciones se consiguen mejoras que oscilan entre 8-15 %. Por otro lado, en las estaciones del norte, las mejoras se consiguen desde la primera hora de predicción, donde se observa entorno a un 2 %, hasta la última  $h = 6$ , donde oscila entre un 15-20 %.

Al añadir datos exógenos del ECMWF (NN+ECMWF) las predicciones alcanzan un %rRMSE menor para todas las estaciones de año independientemente de la ubicación. Los resultados obtenidos, para un horizonte temporal  $h = 1$ , muestran una diferencia de error que no supera el 1 % en comparación con los modelos ARMA y NN. En cambio, para  $h = 6$ , en las estaciones del sur el método NN+ECMWF logra unas mejoras entorno al 1-2 % dependiendo de la estación del año, y en las estaciones del norte estas mejoras alcanzan hasta un 4-5 % en los meses de invierno. Entre C1 y C5, en el norte ambas, se vuelve a observar una diferencia en verano, ya que en la primera la mejora se sitúa en torno al 1 % debido a la nubosidad.

### 5.2.3.2. Resultados según el tipo de día

En la Sección 2.3.4 se realizó el análisis de los datos de radiación solar en cada estación de medida en función de la media y la variabilidad diaria del índice de cielo despejado. Los días se dividieron en nueve tipos de días según la Tabla 5.9, donde por ejemplo un día CI representa un día completamente despajado con una alta radiación y un día tipo AIII será un día con alta variabilidad y una media de radiación baja.

En las Tablas 5.10-5.13 se muestran los resultados obtenidos para todos los tipos de días con cada uno de los modelos de predicción en las estaciones de C0, C1, C2 y C5. Se debe tener en cuenta la cantidad de datos disponible para cada tipo de día y estación de medida. Los tipos de días cuyo número en el conjunto de datos del test no superen el 5 % no se mostrarán en las tablas ya que los resultados no se consideran concluyentes. Así, por ejemplo en las estaciones del sur de la isla no se muestran en las tablas días tipo A debido al escaso número de los mismos.

Los resultados para la predicción de días tipo I presentan los mejores valores, ya que son

Tipo de datos	Condición
A	$\bar{k}^* < 0.5$
B	$0.5 < \bar{k}^* < 0.9$
C	$\bar{k}^* > 0.9$
I	$std(\Delta k^*) < 0.05$
II	$0.05 < std(\Delta k^*) < 0.15$
III	$std(\Delta k^*) > 0.15$

Tabla 5.9: Condiciones a cumplir por cada uno de los subconjuntos de datos.

los días con menor variabilidad. En general, para todas las estaciones de medida, los días tipo C, cuya media de radiación es la más alta, presentan los mejores resultados, mientras que los peores días son los tipo A, con medias de radiación muy baja. En cuanto a los diferentes métodos de predicción, en los días tipo B y C los modelos de redes neuronales con datos de radiación del ECMWF (NN+ECMWF) han obtenido los mejores resultados para todos los horizontes temporales en todas las estaciones de la isla. La diferencia de los resultados de estos modelos no lineales respecto a los modelos Persistence o ARMA es significativamente menor en los días tipo A. En concreto, en los días tipo AII los mejores modelos para las primeras horas de predicción son los modelos Persistence y en los días tipo AIII el modelo ARMA presenta los mejores resultados.

En la estación C0-Pozo Izquierdo, Tabla 5.10, el 94 % de los días del conjunto del test son tipo B y C. En ambos tipos de días el mejor modelo de predicción es el NN+ECMWF para todos los horizontes temporales de predicción. Se puede observar claramente que las mejoras más importantes se establecen para los días tipo C. En estos días, desde la segunda hora de predicción se consiguen mejoras del orden del 2 % respecto al modelo ARMA. Conforme se aumenta el horizonte de predicción se alcanzan mejoras más relevantes, llegando al 7 % respecto al modelo ARMA y al 2-3 % respecto a las redes neuronales sin datos exógenos (NN). La diferencia entre el modelo NN y NN+ECMWF en los días BIII es prácticamente nula para todas las horas, mientras que para los días BII a partir de la tercera hora aumenta paulatinamente hasta el 2 %. En ambos tipos de días se mejora el modelo ARMA para todas las horas.

En la estación C2-La Aldea Tabla 5.12, con un 97 % de días tipo B y C, los resultados son muy similares a los observados para la estación C0. Los días tipo C muestran una clara mejoría de los resultados con el modelo NN+ECMWF desde la segunda hora de predicción, sobre todo con respecto al modelo ARMA (mejoras de hasta el 8 % para  $h = 6$ ). En los días BII no se advierte una diferencia notable entre los modelos NN y NN+ECMWF, pero ambos modelos mejoran el ARMA. En cambio, en los días tipo BIII los tres modelos prácticamente muestran resultados similares.

En las estaciones C1-Las Palmas y C5-Sta. Brígida, Tablas 5.11 y 5.13 respectivamente, la mayoría del conjunto de días presentan un perfil tipo BII o BIII, hasta un 62 % en la estación C1 y un 68 % en la C5. Al igual que en las estaciones del sur de la isla, los días tipo C presentan los menores errores y la mejora obtenida por el modelo NN+ECMWF es la más significativa. En los días tipo BII y BIII este modelo sigue presentando mejoras notables sobre todo para las últimas horas de predicción, ya que la diferencia con los

tipo día	modelo	1 h	2 h	3 h	4 h	5 h	6 h
<b>BIII</b>	Persistence	32.93	44.60	49.88	53.89	55.86	55.66
	SmartPersistence	32.93	43.29	46.94	47.65	45.42	43.75
	ARMA	29.87	36.61	38.23	38.99	39.06	38.77
	NN	30.08	35.71	37.03	36.42	37.73	37.48
	NN+ECMWF	30.33	36.54	37.48	37.74	37.85	37.87
<b>CIII</b>	Persistence	17.51	23.57	31.34	37.15	41.32	44.40
	SmartPersistence	17.51	24.19	31.29	35.66	37.76	38.26
	ARMA	16.15	20.39	24.20	26.13	27.03	27.38
	NN	15.69	18.62	23.21	23.64	21.83	22.61
	NN+ECMWF	15.01	17.47	19.94	20.22	20.39	20.57
<b>BII</b>	Persistence	14.46	22.63	29.46	34.82	37.85	38.67
	SmartPersistence	14.46	24.21	30.23	31.90	30.57	28.18
	ARMA	14.62	20.22	23.00	24.19	24.33	23.97
	NN	14.47	19.53	21.40	22.59	21.19	22.53
	NN+ECMWF	14.30	19.95	21.41	21.88	20.94	20.15
<b>CII</b>	Persistence	8.53	13.58	18.46	22.04	24.71	26.17
	SmartPersistence	8.53	14.07	17.72	19.35	19.30	18.48
	ARMA	8.16	11.69	14.21	15.68	16.60	17.08
	NN	8.28	12.25	13.59	15.64	15.49	14.77
	NN+ECMWF	7.68	10.47	12.59	12.53	12.37	12.49
<b>CI</b>	Persistence	6.12	11.02	15.53	19.12	21.52	22.58
	SmartPersistence	6.12	11.68	15.41	16.70	16.10	14.99
	ARMA	6.41	10.03	12.43	13.89	14.70	15.07
	NN	5.72	9.51	11.40	13.79	13.48	13.70
	NN+ECMWF	5.55	8.72	10.43	10.60	10.47	10.61

Tabla 5.10: Errores %rRMSE en los distintos tipos de días para todos los horizontes temporales en la estación C0-Pozo Izquierdo

tipo día	modelo	1 h	2 h	3 h	4 h	5 h	6 h
<b>AIII</b>	Persistence	47.81	64.67	73.16	78.94	82.67	86.42
	SmartPersistence	47.81	64.08	70.60	74.59	76.60	76.57
	ARMA	44.50	56.34	61.46	64.76	66.94	68.66
	NN	44.22	57.15	62.36	65.51	68.06	68.14
	NN+ECMWF	44.56	56.97	62.63	65.57	67.96	67.64
<b>BIII</b>	Persistence	32.81	45.20	51.53	54.80	56.80	56.99
	SmartPersistence	32.81	45.31	49.46	49.86	47.74	44.47
	ARMA	29.98	37.61	39.94	40.41	40.37	39.94
	NN	29.57	36.91	38.90	38.69	38.04	37.61
	NN+ECMWF	29.63	36.46	37.76	37.57	37.44	37.16
<b>AII</b>	Persistence	28.79	40.36	49.53	57.71	64.40	68.81
	SmartPersistence	28.79	42.04	52.25	58.54	62.17	64.75
	ARMA	29.56	40.29	47.10	51.88	55.21	57.46
	NN	31.08	42.69	50.03	55.53	59.89	62.47
	NN+ECMWF	30.55	44.14	52.59	58.59	60.86	62.49
<b>BII</b>	Persistence	15.37	24.03	30.90	36.40	40.07	41.08
	SmartPersistence	15.37	25.99	32.90	35.19	34.14	31.89
	ARMA	15.40	22.01	25.69	27.67	28.49	28.46
	NN	14.82	21.51	25.29	26.66	26.97	26.73
	NN+ECMWF	14.70	19.95	22.20	23.47	23.93	24.36
<b>CII</b>	Persistence	10.82	15.86	20.84	24.20	27.12	28.55
	SmartPersistence	10.82	16.68	20.42	21.77	21.27	20.26
	ARMA	11.55	16.52	19.99	21.92	23.20	23.87
	NN	12.67	18.71	22.67	24.10	24.67	24.73
	NN+ECMWF	10.37	13.12	15.11	15.73	16.89	16.06
<b>CI</b>	Persistence	5.07	8.90	12.70	15.74	18.05	20.11
	SmartPersistence	5.07	9.30	12.26	14.62	17.71	22.11
	ARMA	8.28	13.75	17.62	20.20	21.94	23.16
	NN	9.36	16.48	21.08	23.63	24.67	25.58
	NN+ECMWF	6.65	10.58	12.67	13.85	15.82	15.30

Tabla 5.11: Errores %rRMSE en los distintos tipos de días para todos los horizontes temporales en la estación C1-Las Palmas

tipo día	modelo	1 h	2 h	3 h	4 h	5 h	6 h
<b>BIII</b>	Persistence	31.13	41.42	45.98	49.22	50.84	50.65
	SmartPersistence	31.13	40.59	43.76	44.42	43.83	42.82
	ARMA	28.40	34.57	36.22	37.01	37.19	37.01
	NN	29.03	34.85	36.01	37.11	39.22	37.77
	NN+ECMWF	28.37	34.04	35.98	36.36	36.71	37.68
<b>CIII</b>	Persistence	13.11	19.14	24.64	29.57	31.76	31.72
	SmartPersistence	13.11	19.94	24.69	25.21	23.15	20.28
	ARMA	12.37	16.26	18.54	19.86	19.95	19.49
	NN	11.81	15.11	16.06	16.83	17.04	15.79
	NN+ECMWF	11.02	12.88	13.78	15.61	15.29	13.48
<b>BII</b>	Persistence	11.49	17.73	23.48	27.89	30.84	32.10
	SmartPersistence	11.49	19.37	24.98	27.25	27.14	25.89
	ARMA	11.70	16.50	19.45	21.00	21.72	21.89
	NN	10.81	14.78	14.99	16.78	16.76	16.72
	NN+ECMWF	10.44	13.55	15.77	16.37	16.89	17.17
<b>CII</b>	Persistence	8.11	12.74	17.18	20.67	23.10	24.21
	SmartPersistence	8.11	13.63	17.72	19.55	19.40	18.60
	ARMA	8.94	13.10	15.76	17.26	18.03	18.33
	NN	7.72	10.29	11.50	11.96	12.63	12.36
	NN+ECMWF	6.99	9.07	9.81	10.72	10.66	10.30
<b>CI</b>	Persistence	3.60	6.82	10.05	13.00	15.39	16.86
	SmartPersistence	<b>3.60</b>	7.70	11.21	13.34	14.21	14.50
	ARMA	5.12	8.32	10.42	11.75	12.54	12.92
	NN	4.15	6.47	8.38	8.68	8.95	8.57
	NN+ECMWF	3.58	5.17	5.64	6.12	6.09	6.12

Tabla 5.12: Errores %rRMSE en los distintos tipos de días para todos los horizontes temporales en la estación C2-La Aldea

tipo día	modelo	1 h	2 h	3 h	4 h	5 h	6 h
<b>AIII</b>	Persistence	52.48	70.87	80.85	86.82	89.86	93.75
	SmartPersistence	52.48	70.41	77.08	81.96	86.62	89.85
	ARMA	49.89	63.94	70.41	73.99	76.11	78.13
	NN	50.30	66.28	72.11	76.49	77.60	80.02
	NN+ECMWF	51.22	67.08	73.56	77.82	78.18	78.62
<b>BIII</b>	Persistence	27.78	39.78	48.17	54.51	58.24	58.88
	SmartPersistence	27.78	41.09	48.56	50.34	48.03	44.53
	ARMA	26.07	34.35	38.34	40.25	40.68	40.17
	NN	25.74	32.95	36.12	37.27	37.09	36.39
	NN+ECMWF	24.74	32.14	33.96	35.14	34.64	34.47
<b>CIII</b>	Persistence	11.79	17.89	24.46	30.22	34.79	37.09
	SmartPersistence	11.79	18.82	24.72	27.45	26.81	25.03
	ARMA	13.44	20.03	24.95	28.08	29.90	30.66
	NN	15.64	22.97	25.55	28.20	27.74	27.11
	NN+ECMWF	11.39	16.38	20.33	24.02	23.99	25.01
<b>AII</b>	Persistence	35.68	50.99	63.01	74.00	82.21	89.90
	SmartPersistence	35.68	52.81	65.29	74.63	81.78	88.28
	ARMA	37.97	53.85	64.45	72.23	77.56	81.66
	NN	39.70	57.14	69.92	78.93	86.03	90.51
	NN+ECMWF	38.62	53.74	71.10	75.89	82.85	84.48
<b>BII</b>	Persistence	12.97	20.66	27.00	31.67	34.22	34.87
	SmartPersistence	12.97	21.81	26.74	27.88	26.78	25.33
	ARMA	13.07	19.20	22.91	24.87	25.61	25.74
	NN	13.05	19.06	22.22	24.10	24.19	24.14
	NN+ECMWF	11.85	16.04	18.54	19.55	20.15	20.53
<b>CII</b>	Persistence	9.60	16.17	20.42	23.47	25.54	26.05
	SmartPersistence	9.60	16.44	19.24	19.94	19.42	19.01
	ARMA	10.17	15.67	18.51	20.09	20.93	21.20
	NN	13.83	21.49	23.11	25.68	24.85	24.29
	NN+ECMWF	9.86	12.24	14.62	15.76	16.84	19.08

Tabla 5.13: Errores %rRMSE en los distintos tipos de días para todos los horizontes temporales en la estación C5-Sta. Brígida

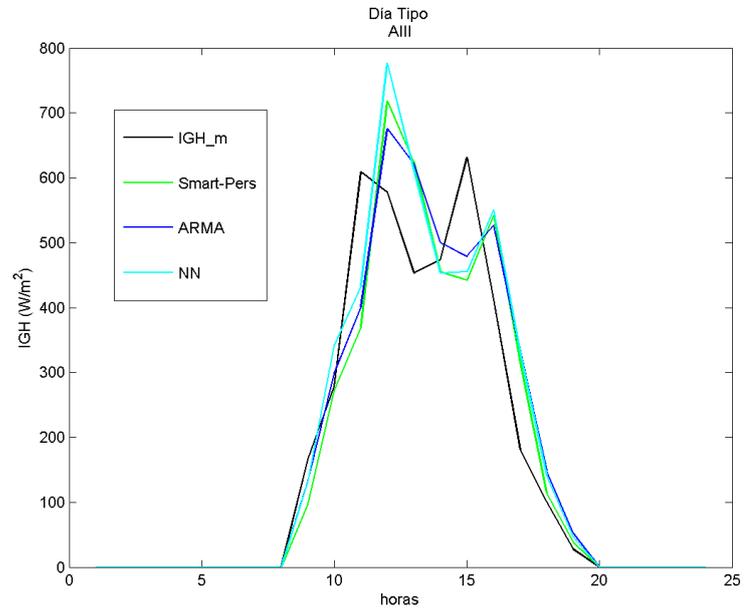
modelos NN y ARMA crece con el horizonte de predicción. En los días BIII para la estación C1 los modelos NN y NN+ECMWF presentan perfiles muy similares. En estas estaciones, el modelo ARMA ha obtenido los mejores resultados para los días tipo AIII, aunque las diferencias con el resto de modelos no son nada significativas, menores al 1%. Los días tipo AII en cambio muestran unos mejores resultados con los modelos Persistence en las primeras horas y con el modelo ARMA para las últimas horas.

De manera general se puede concluir que para los días tipo C el modelo NN+ECMWF muestra los mejores resultados para toda la isla. En los días tipo B este modelo presenta también los mejores resultados, aunque se observa que la influencia de los datos del ECMWF sobre las predicciones es menor. En cambio para los días tipo A de las estaciones del norte el mejor modelo en general es el ARMA. Se debe tener en cuenta que este tipo de días representan una minoría y los resultados de predicción son los peores.

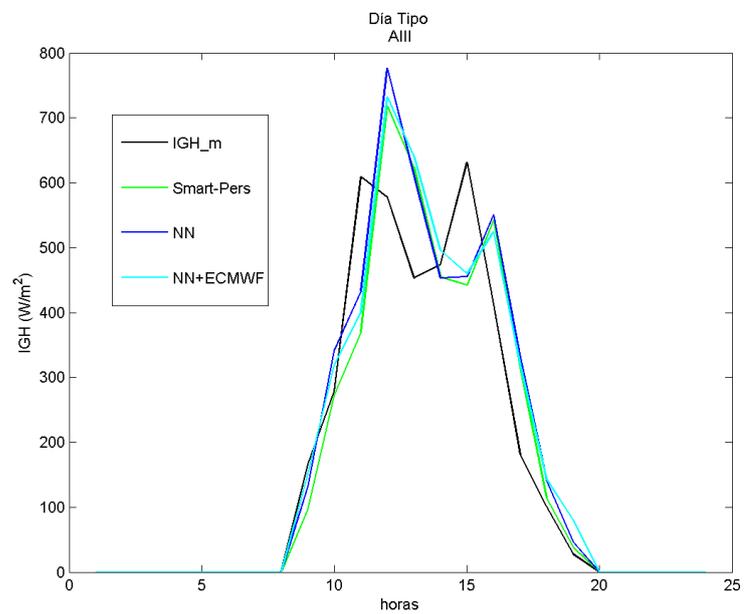
En las Figuras 5.19-5.24 se muestran los ejemplos de las predicciones hora a hora para días concretos del año. En dichas figuras se observan los perfiles diarios de los distintos modelos de predicción con respecto al valor perfil de radiación solar medido en las estaciones ( $IGH_m$ ). Los modelos no se han puesto en una sola gráfica para facilitar la visión. Las imágenes que se muestran corresponden a días tipo AIII, BIII y CIII por ser ejemplos de días representativos de los distintos conjuntos de datos.

En las Figuras 5.19-5.20 se muestran las predicciones para un día tipo AIII en la estación C5 para un horizonte de predicción  $h = 1$  y  $h = 6$  respectivamente. Este tipo de días tienen un perfil con una media baja y una alta variabilidad. Para  $h = 1$  todos los modelos consiguen reproducir la variabilidad que presentan las medidas reales aunque no se consigue evitar un retraso de una hora en las predicciones propias del Persistence. En cambio para  $h = 6$  se observa que los modelos Persistence pierden precisión, las NN sobreestiman la radiación aunque son capaces de representar las variaciones y los modelos ARMA y NN+ECMWF reproducen de manera general mejor los resultados pero sin seguir las variaciones del día.

En las Figuras 5.21-5.22 se muestran las predicciones para un día tipo BIII en la estación C0. Para  $h = 1$  se consigue un perfil diario similar a las medidas y prácticamente no existe diferencia entre los modelos Persistence, NN y NN+ECMWF, mientras que el modelo ARMA subestima la radiación. En cambio para  $h = 6$  el modelo Persistence y ARMA claramente subestiman las medidas de radiación solar, mientras que los modelos basados en redes neuronales no se distinguen significativamente. El modelo NN+ECMWF siempre tiende a suavizar los resultados alrededor de la curva de tendencias media del día. Por último, En las Figuras 5.23-5.24 se muestran las predicciones para un día tipo CIII en la estación C0. Los resultados obtenidos muestran en general las mismas características que el caso anterior, aunque para  $h = 6$  se puede observar una ligera diferencia entre NN y NN+ECMWF ya que esta última muestra un perfil suavizado.

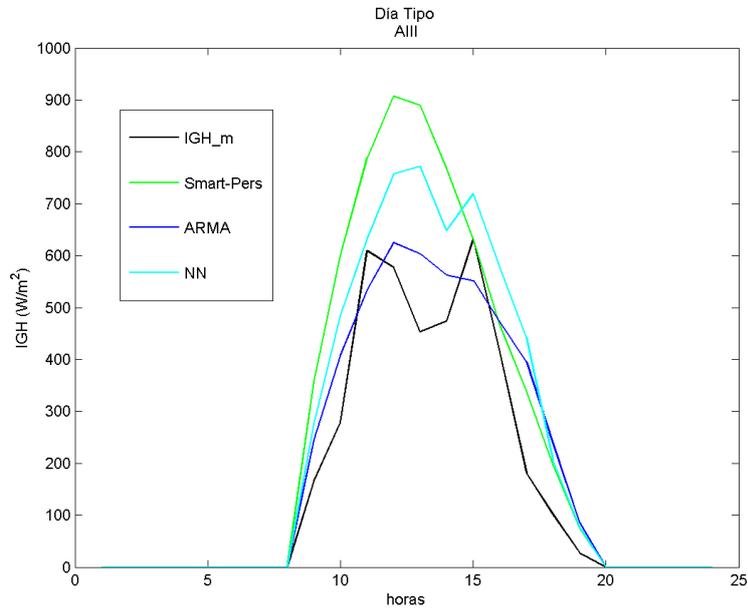


(a)

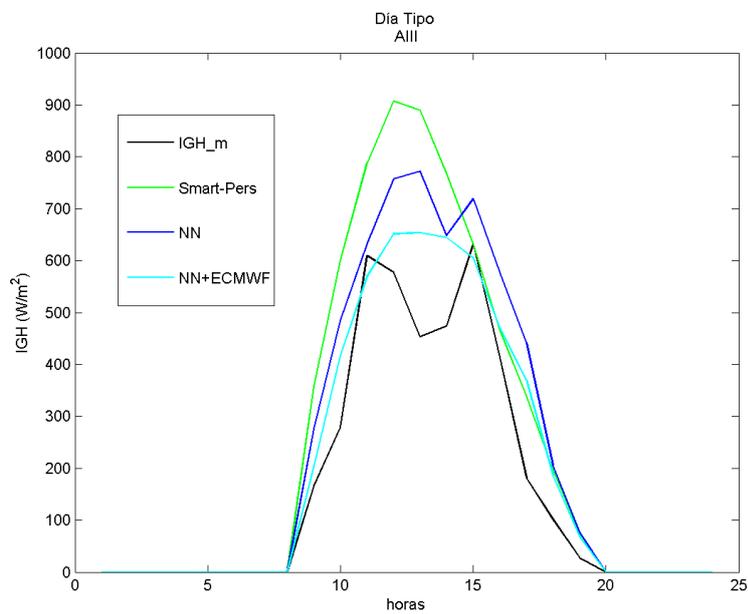


(b)

Figura 5.19: Predicción de la radiación diaria IGH con diferentes modelos en la estación C5 para un horizonte de  $h = 1$  en un día tipo AIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF

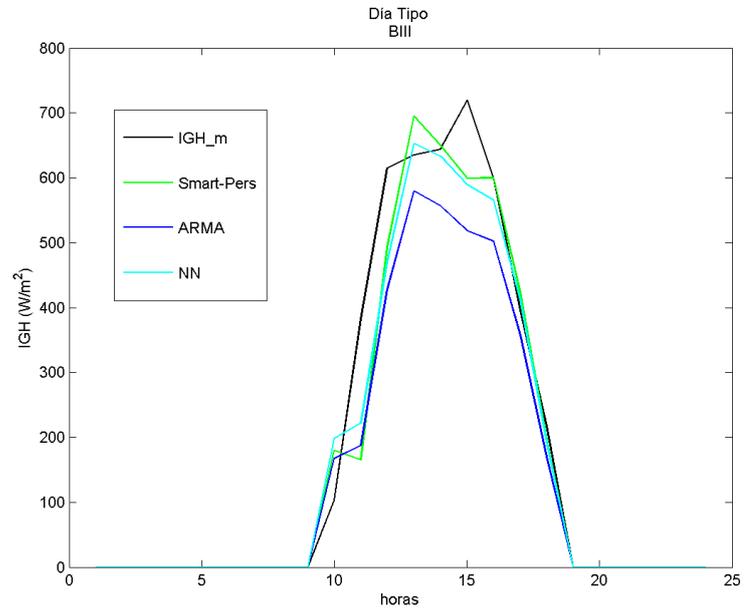


(a)

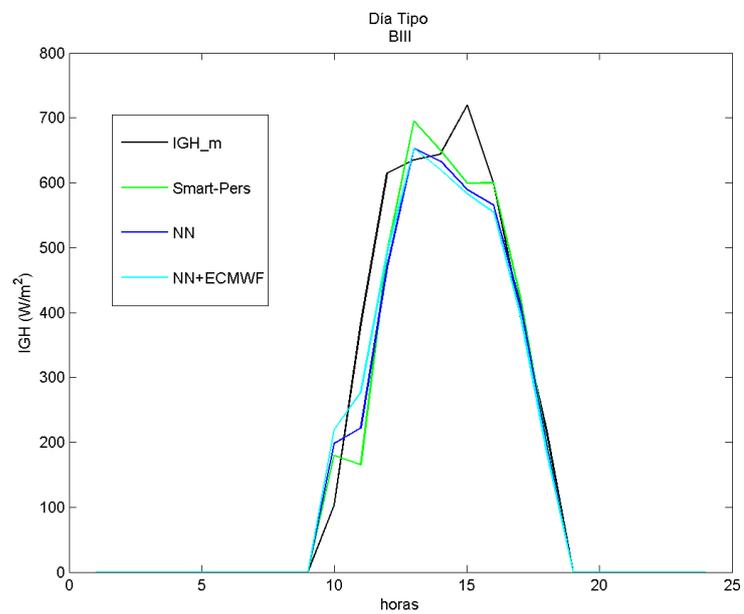


(b)

Figura 5.20: Predicción de la radiación diaria IGH con diferentes modelos en la estación C5 para un horizonte de  $h = 6$  en un día tipo AIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF

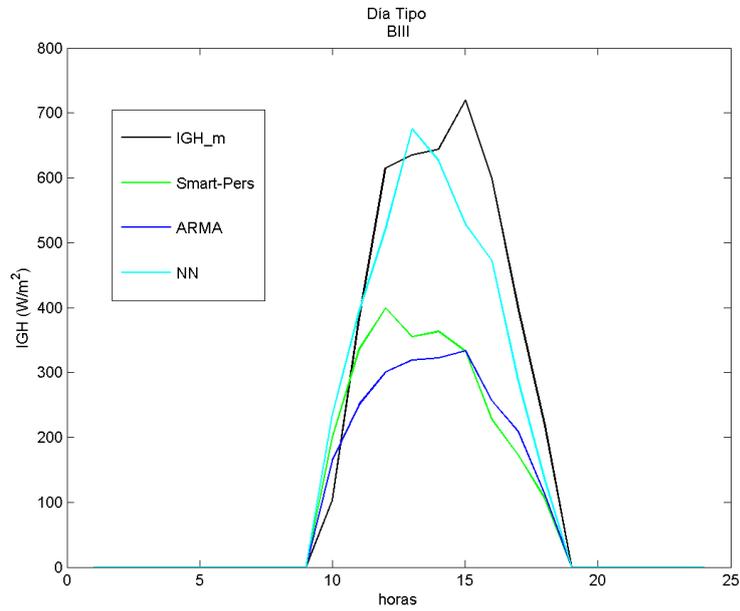


(a)

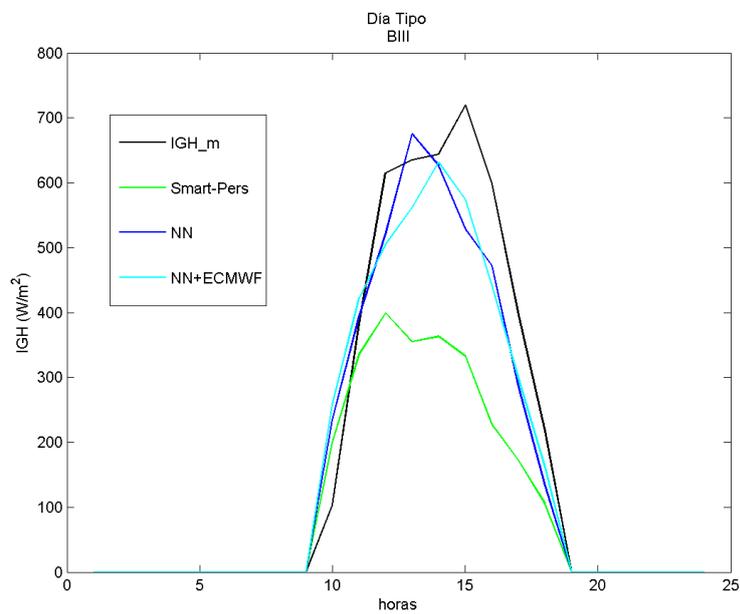


(b)

Figura 5.21: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo BIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF

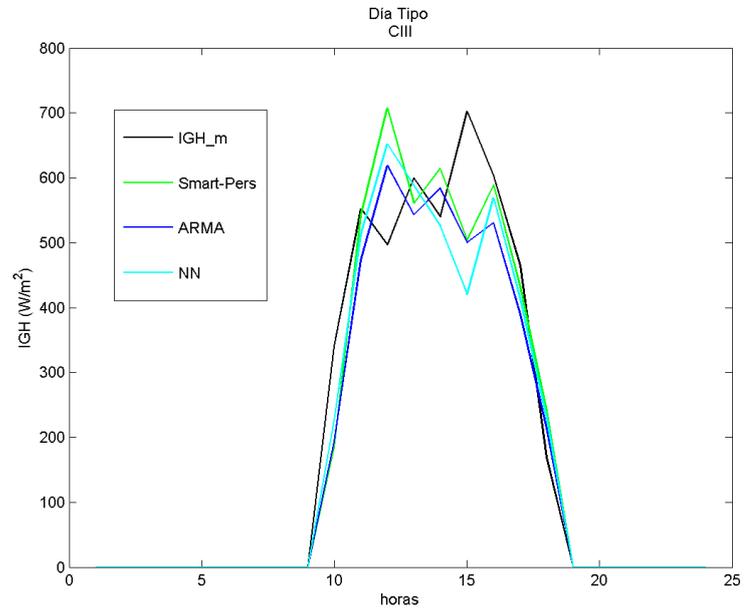


(a)

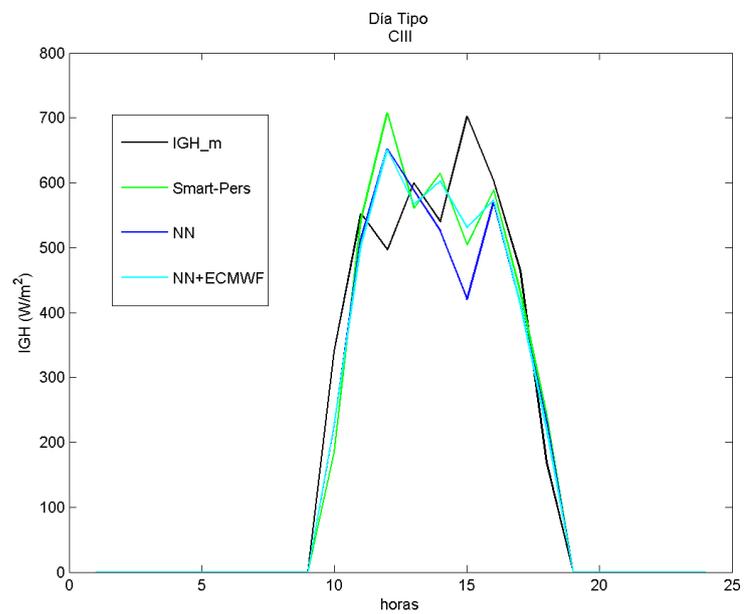


(b)

Figura 5.22: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo BIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF

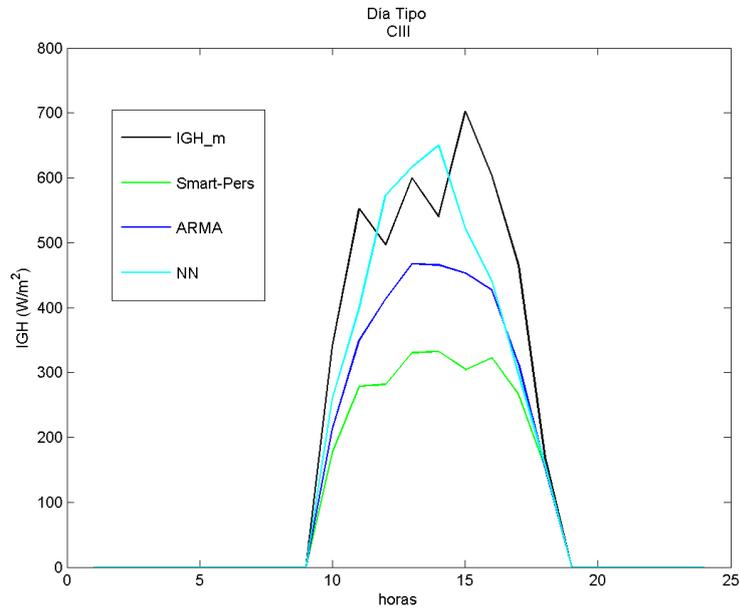


(a)

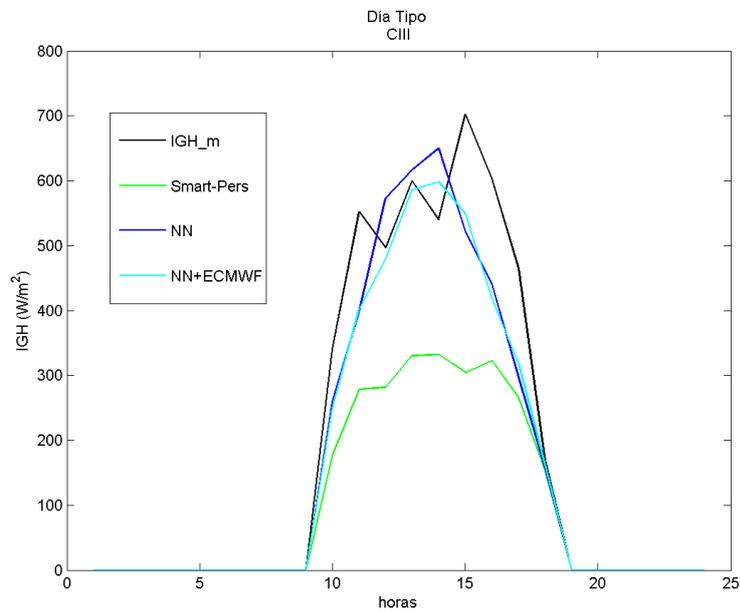


(b)

Figura 5.23: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo CIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF



(a)



(b)

Figura 5.24: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo CIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF

### 5.3. Predicción de Radiación Solar a partir de datos satelitales y otras variables

En esta Sección se describirán los cálculos de predicción de la radiación solar añadiendo datos exógenos a la serie de datos históricos del índice de cielo despejado utilizando RNAs. En esta tesis se ha decidido trabajar, por un lado, con los datos de radiación solar obtenidos a partir de las imágenes satelitales proporcionados por el Helioclim-3. Por otro lado también se simularán predicciones de radiación solar, utilizando como entradas a las RNAs, datos de velocidad del viento y humedad relativa proporcionados por el modelo numérico ECMWF. En todos los casos, siempre se utilizarán los datos históricos obtenidos en las estaciones de medida terrestres además de los datos exógenos.

Tanto los datos satelitales como los datos de velocidad del viento y humedad relativa disponibles corresponden únicamente al año 2005. Por lo tanto, para trabajar en conjunto entre los datos terrestres y los datos exógenos se trabajará sólo con este año. En este caso, se ha decidido trabajar con las estaciones de C0-Pozo Izquierdo y C1-Las Palmas como representación de los climas del sur y el norte de la isla respectivamente. Para poder realizar una comparación entre los modelos que se probaron con dos años de medida y los nuevos modelos con los datos exógenos, se volverán a calcular todas las predicciones realizadas con cada modelo partiendo con el año 2005 únicamente. En los resultados se mostrarán las predicciones obtenidas con los siguientes modelos:

- Modelo *Smart-Persistence*
- Modelo *Climatológico*
- Modelo de Redes Neuronales Artificiales con datos terrestres - *NN*
- Modelo de Redes Neuronales Artificiales con datos terrestres y datos satelitales - *NN+SAT*
- Modelo de Redes Neuronales Artificiales con datos terrestres y datos de radiación del ECMWF - *NN+ECMWF*
- Modelo de Redes Neuronales Artificiales con datos terrestres, datos de radiación del ECMWF y datos de velocidad del viento y humedad relativa del ECMWF - *NN+ECMWF+WIND*
- Modelo de Redes Neuronales Artificiales con datos terrestres, datos satelitales y datos de radiación del ECMWF - *NN+ECMWF+SAT*
- Modelo de Redes Neuronales Artificiales con datos terrestres, datos satelitales, datos de radiación del ECMWF y datos de velocidad del viento y humedad relativa del ECMWF - *NN+ECMWF+WIND+SAT*

En el caso anterior se disponía de dos años de datos completos, sin embargo en este caso sólo se dispone de un año de medida completo para obtener un conjunto de datos de entrenamiento y test. Ambos conjuntos deben representar muy bien todas las épocas

del año para poder obtener los modelos y estudiar los errores sobre diferentes situaciones climáticas. Así, se ha decidido obtener el conjunto de datos de entrenamiento como las tres primeras semanas de cada mes y el conjunto de datos del test como la última semana de cada mes. El conjunto de datos de entrenamiento representará el 75 % de los datos del año, mientras que el conjunto de datos del test será el 25 %.

### 5.3.1. Selección de los datos de radiación Satélite

Los resultados obtenidos con las predicciones realizadas con las RNA a partir de datos terrestres se pretenden mejorar utilizando datos exógenos provenientes de la base de datos satelitales HelioClim-3. Estos datos satelitales se añadirán como nuevas entradas a la red neuronal en conjunto con los datos históricos terrestres. Debido a la complejidad de los cálculos durante el entrenamiento de las RNAs, se ha decidido añadir un máximo de 30 entradas provenientes de los datos satélites. La técnica elegida para decidir los píxeles con la información más relevante dentro de la cuadrícula, y por lo tanto los que se utilizarán como entradas, será la correlación entre los datos satelitales y los datos terrestres de la estación en cuestión, ver Sección 3.3.1.

En cada estación se dispondrá de las correlaciones entre cada píxel de los datos satélite desfasado desde  $t = 0$  hasta  $t = 3$  horas y el dato terrestre correspondiente a la propia estación sin desfasar. En cada estación, por tanto, se dispone de cuatro imágenes con los valores de las correlaciones correspondientes a cada desfase temporal. Los píxeles con una mayor correlación serán los elegidos para trabajar con las RNAs. Para facilitar el cálculo, debido a la gran cantidad de píxeles, se aplicó un filtro por cada ventana de  $3 \times 3$  píxeles. Se simplificaron las imágenes resumiendo la información cada  $3 \times 3$  píxeles en un *Super-píxel* con el valor mediano.

Se ha elegido trabajar con un número máximo de 30 píxeles, por lo que se deberán elegir éstos entre las cuatro imágenes de los desfases temporales. Así, para decidir cuántos píxeles de cada imagen se introducían finalmente en la RNAs se realizaron seis test diferentes. En cada uno de estos test se elegía un porcentaje diferente de píxeles de cada imagen según los siguientes criterios:

- **Test-1**, la primera aproximación consiste en contar el número de píxeles con una correlación superior a 0.5 en cada imagen correspondiente a un desfase temporal [Zagouras15]. Luego, se calcula el porcentaje de píxeles con un valor superior al citado en cada imagen. El mismo procedimiento se siguió para ambas estaciones y se eligió trabajar con el porcentaje medio de ambas. Los 30 píxeles elegidos son aquellos con mayor correlación en cada imagen de acuerdo con los porcentajes hallados. En nuestro caso, al aplicar el Test-1, se deben seleccionar un 56 % de píxeles en la imagen  $t = 0$ , 30 % en  $t = -1$ ; 8 % en  $t = -2$  y 6 % en  $t = -3$ , lo que se traduce en extraer 17, 9, 2 y 2 píxeles respectivamente, ver Figura 5.25.
- **Test-2 y Test-3**, debido a que la mayoría de los píxeles añadidos en el Test-1 correspondían a las dos primeras imágenes ( $t = 0$  y  $t = -1$ ) se considero realizar dos nuevas distribuciones porcentuales. En ambas distribuciones se aumentó el número de píxeles seleccionados en el resto de imágenes. Así, en el Test -2 se extrajeron 12, 9, 6 y 3 píxeles de cada imagen y en el Test-3 se extrajeron 9, 9, 8 y 4 píxeles.

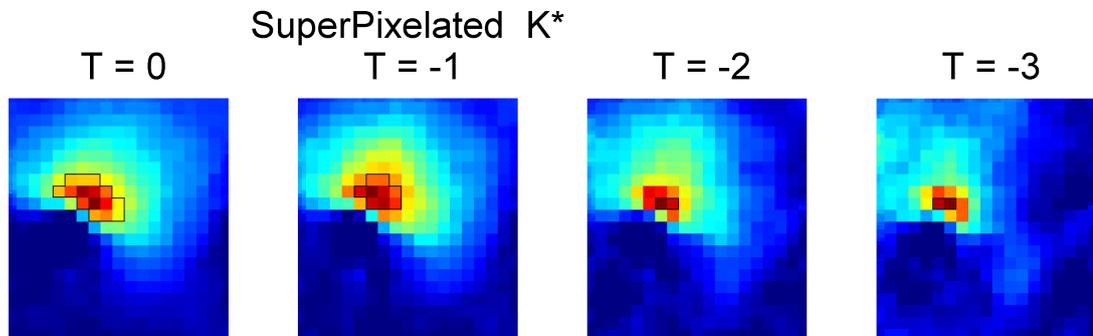


Figura 5.25: Selección de los Superpixel (3x3) en la estación C1-Las Palmas utilizando el Test-1 para elegir el número en cada imagen desfasada  $t = 0, 1, 2$  & 3 horas. El área negra muestra los Superpíxeles seleccionados. Cada píxel representa la intercorrelación de  $k^*$  entre los datos terrestres y satelitales

- **Test-4**, en el estudio se observó que las correlaciones más altas siempre se obtenían en las dos primeras imágenes. En este test se decidió realizar una distribución teniendo en cuenta únicamente el número de píxeles superior a 0.5 de dichas imágenes. De esta manera, se extrajeron 19 y 11 píxeles respectivamente de las imágenes  $t = 0$  y  $t = -1$ .
- **Test-5**, esta simulación realiza la misma suposición que en el Test-1 pero se calcula una distribución diferente por cada estación del año. Como se comentó en la Sección 3.3.1, las correlaciones estacionales también ofrecían información coherente con el clima de la isla, así que se decidió realizar una elección de píxeles diferente en cada estación.
- **Test-6**, por último, una vez se simularon los cuatro primeros tests, se decidió escoger la distribución porcentual del que mejor resultados obtuviera y realizar una selección de píxeles trimestral con dicha distribución.

Los resultados obtenidos con los diferentes tests mostraron errores similares en ambas estaciones para todos los horizontes temporales. En general, el Test-5 y el Test-6, con una distribución trimestral de píxeles, obtuvieron peores resultados. Aunque no se obtuvo para cada horizonte temporal de predicción el mismo resultado, se decidió trabajar con un solo Test para cada estación para facilitar las predicciones en el futuro. En cada estación se calculó el mejor Test basándonos en %rRMSE obtenido en las predicciones realizadas con el conjunto de datos del test.

Así, en la estación C0 se eligió el Test-3 y en la estación C1 se eligió el Test-2. A la hora de estudiar cada Test también se realizó el mismo procedimiento de elección de la arquitectura óptima de la RNA. Cuando se estudien los resultados obtenidos por el método de predicción de RNA con datos satelitales se hará referencia en cada estación a estos Tests señalados aquí.

### 5.3.2. Estudio de la predicción utilizando otros datos meteorológicos del ECMWF

Además de los datos de radiación, en esta tesis se estudiarán los efectos de incluir los datos de velocidad y dirección del viento y la humedad relativa a distintas altitudes de la atmósfera. En esta tesis se optó por incluir las predicciones que realiza el ECMWF de estas variables. En este caso los datos extraídos del programa únicamente están disponibles para el año 2005, por lo que su contribución a la predicción de la radiación solar se estudiará en las estaciones de C0-Pozo Izquierdo y C1-Las Palmas.

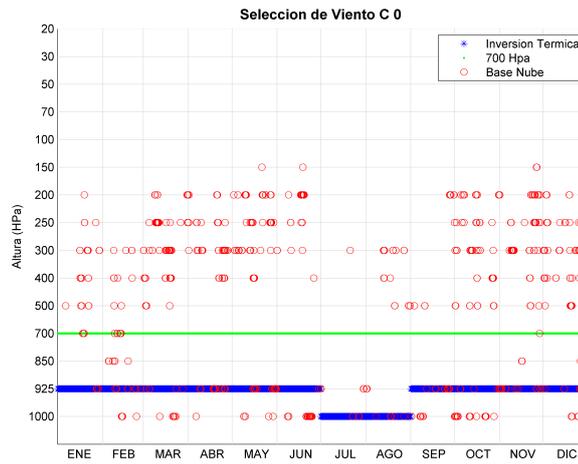
Como ya se explicó en la Sección 2.4.2, se busca incluir los datos de viento y humedad a distintas altitudes de la atmósfera por lo que se extrajeron los datos del ECMWF para todos los niveles disponibles. En este caso se obtuvieron diferentes medidas entre las altitudes de 1000-1 hPa de presión. A continuación se describen las variables empleadas en este caso:

- R Relative Humidity, se define con respecto a la saturación sobre el hielo por debajo de 23°C y con respecto a la saturación sobre el agua.
- U-velocity, velocidad del viento para la coordenada U del eje de coordenadas.
- V-velocity, velocidad del viento para la coordenada V del eje de coordenadas.

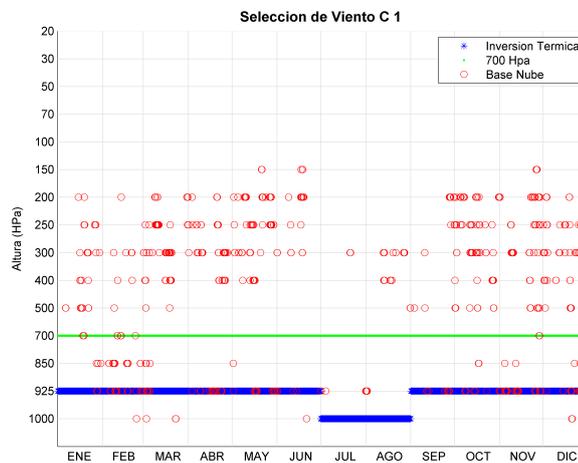
En las RNAs se decidió incluir estos datos correspondientes a la altitud atmosférica que se estime más relevante para la radiación solar. En la bibliografía se encontraron diversas referencias sobre esta altitud. En concreto en esta tesis se decidió incluir los datos de viento y humedad para las siguientes altitudes de la atmósfera:

- **ECMWIND cbh**, datos a la altitud de la base de las nubes. Como en el ECMWF para el año 2005 no se disponía de este dato, se obtuvo esta altitud como aquella a la cual la humedad relativa es mayor o igual al 95 % [Lave13]. En este caso por lo tanto únicamente se dispone en cada instante de los datos de módulo y dirección de la velocidad del viento.
- **ECMWIND 700 Hpa**, datos a la altitud de 700 HPa. Se obtuvieron los datos de velocidad y dirección del viento y de humedad relativa para esta altitud siguiendo un estudio en la Isla Reunión donde se indicaba esta altitud como la más influyente en la predicción de la radiación solar [Badosa15]. Esta altitud se define como aquella a la cual la humedad relativa presenta una mayor desviación típica.
- **ECMWIND TI**, datos a la altitud de la inversión térmica en las Islas Canarias según el estudio [DA96]. Se obtuvieron los datos de velocidad y dirección del viento y de humedad relativa para la altitud en que se produce la inversión térmica en cada mes del año.

Para poder estudiar la aportación que realiza cada conjunto de variables se decidió estudiar los resultados de las RNAs paso a paso. Primero se repitieron los cálculos



(a) C0



(b) C1

Figura 5.26: Altitud a la que se extrajeron los datos de velocidad del viento y humedad relativa según cada criterio de selección en las estaciones de C0 y C1

únicamente con RNAs y los datos terrestres como datos de entrada (*NN*), pero, a diferencia de los cálculos de la Sección 5.2.3, utilizando únicamente el año 2005. Posteriormente se añadieron los datos de radiación solar y nubosidad obtenidos por las predicciones del ECMWF (*NN+ECMWF*). Por último, se añadieron a los anteriores los datos del módulo y dirección de la velocidad del viento y la humedad relativa atmosférica obtenidos por el ECMWF (*NN+ECMWF+WIND*).

En el último caso, se realizaron las simulaciones con los datos de viento y humedad según los tres criterios de altitud descritos, ver Figura 5.26. En cada simulación se estudió la arquitectura óptima de la RNAs y los resultados de la predicción respecto al conjunto de datos del test. En el caso de la estación C0 se obtuvo el mejor resultado utilizando los datos de viento a la altitud de 700 hPa, mientras que en la estación C1 el mejor resultado se consigue utilizando la altitud de la nube. Como ya se ha comentado, en las estaciones del Norte de la isla, como C1, la nubosidad es más frecuente por lo que la información del viento a la altitud de la base de las nubes aporta una información más relevante.

En la Sección 5.3.3 se presentarán los resultados obtenidos para el año 2005 con cada uno de estos modelos. En estos resultados se muestran, para los datos de viento, las predicciones obtenidas con la altitud óptima obtenida para cada estación.

### 5.3.3. Resultados obtenidos y comparación entre los modelos

En este apartado se estudiarán los resultados obtenidos al realizar las predicciones de radiación solar con las Redes Neuronales Artificiales a partir de diferentes conjuntos de entradas. Como ya se comentó, se estudiarán las predicciones utilizando datos históricos de las medidas terrestres, datos de radiación obtenidos a partir de imágenes satelitales con el HelioClim-3, datos de radiación, velocidad del viento y humedad relativa provenientes de las predicciones del día siguiente del ECMWF. Los modelos con datos terrestres (*NN*) se irán combinando con los diferentes datos exógenos expuestos para estudiar la relevancia de cada uno de ellos en la predicción.

Como en la Sección 5.2.3, también se exponen los resultados con dos modelos sencillos de predicción, Smart-Persistence y Climatológico. Aunque en este caso, como ya se ha probado que las redes neuronales ofrecen una clara mejoría en las predicciones respecto a estos últimos, nos centraremos en describir las diferencias entre los modelos de RNAs entre sí.

En este caso cabe recordar que sólo se dispone del año 2005 para realizar el entrenamiento y el test de los modelos. El conjunto del test únicamente dispone del 25 % de los valores horarios del año. Aún así, se realizó también el estudio según las distintas estaciones del año y los nueve tipos de días descritos. Las diferencias en los resultados obtenidos en términos de RMSE y MAE son muy similares entre los modelos por lo que, y para no saturar el texto de datos, se ha preferido hacer referencia en esta Sección únicamente a los valores de RMSE.

### 5.3.3.1. Resultados anuales y trimestrales

En esta Sección se presentan los resultados obtenidos por los diferentes métodos de predicción de radiación solar desde  $h = 1$  hasta  $h = 6$  horas de horizonte temporal. Los resultados se mostrarán para el conjunto de datos anual y para todas las estaciones de la isla de Gran Canaria.

Los resultados anuales para cada estación y cada horizonte temporal se pueden observar en las Tablas 5.14-5.16. Los resultados se presentan en términos del error RMSE, %rRMSE y el parámetro %SKILL, que nos permitirá establecer una relación de la diferencia de predicción de cada modelo con respecto al modelo Persistence. En las Figuras 5.27-5.28-5.29 se pueden observar para cada estación la evolución gráfica de los errores relativos %rRMSE desde  $h = 1$  hasta  $h = 6$  horas de horizonte temporal.

En la Tabla 5.15 se pueden estudiar los errores anuales en términos de %rRMSE. Todos los modelos muestran en general un aumento en el error de predicción conforme aumenta el horizonte temporal. En el modelo Smart-Persistence este aumento entre la primera y la última hora conduce desde un error en torno al 15 % hasta un error del 26 % en las estaciones del sur, mientras que en el norte oscila entre 27 % y 42 % para  $h = 1$  y  $h = 6$  respectivamente. Por su parte los modelos de predicción basados en RNAs conducen a resultados inferiores al 22 % en las estaciones del sur y entorno al 34 % en las estaciones del norte para la última hora de predicción.

De manera mayoritaria, se aprecia que el peor modelo de predicción para las tres primeras horas es el Climatológico, y para las últimas horas el modelo Smart-Persistence reproduce los peores errores de predicción. Al igual que en la Sección 5.2.3.1 se observa una gran diferencia entre las estaciones del sur y norte de la isla en referencia a los resultados anuales de predicción, recogándose diferencias de hasta un 12 % para el horizonte temporal  $h = 6$ . Para las estaciones del sur como C0 los modelos basados en redes neuronales mejoran el modelo Smart-Persistence alrededor de un 1.5 % y un 4 % para los horizontes temporales  $h = 1$  y  $h = 6$  respectivamente. Sin embargo, las estaciones del norte de la isla como C1, donde existe una mayor formación de nubosidad, los modelos basados en redes neuronales conducen a unas mejoras entre un 3 % y un 6 %. Las estaciones del sur muestran mejoras inferiores a las del norte respecto a los modelos sencillos de predicción.

Los resultados de los distintos modelos de RNAs con datos exógenos se comentarán por separado para poder explicar la influencia de cada uno de ellos en las predicciones. En la Figura 5.27 se puede ver la evolución de los modelos NN, NN+SAT, NN+ECMWF y NN+ECMWF+SAT, para estudiar la influencia de los datos satelitales y del ECMWF por separado y en combinación. En la Figura 5.28 se comparan los modelos NN, NN+ECMWF y NN+ECMWF+WIND, para estudiar la influencia de los diferentes datos aportados por el ECMWF, tanto de radiación como de viento. Los modelos de RNAs con datos exógenos mejoran los resultados del resto de modelos para cualquier horizonte temporal.

En la estación C0 tanto la aportación satelital como el ECMWF mejoran la predicción con respecto a las NN. El modelo NN+SAT obtiene mejores resultados respecto al NN+ECMWF sobre todo en las tres primeras horas, con mejoras superiores al 1 %. Este resultado es consistente con el hecho de que los datos satelitales incluidos en las redes neuronales correspondían a las imágenes obtenidas con retardos horarios desde las cero

estación	métodos	1 h	2 h	3 h	4 h	5 h	6 h
C0	Smart Persistence	92.47	124.64	140.10	144.44	141.50	138.69
	Climatológico	145.13	145.18	145.25	145.33	145.42	145.50
	NN	88.20	113.39	125.13	127.12	131.68	130.03
	NN+ECMWF	86.40	112.93	117.70	116.94	120.01	120.88
	NN+ECMWF+WIND	85.93	108.09	110.40	115.13	116.16	116.80
	NN+ECMWF+SAT	84.00	106.17	110.51	114.93	118.89	120.43
	NN+ECMWF+WIND+SAT	84.47	104.99	108.43	112.85	115.032	118.778
	NN+SAT	83.58	104.58	113.51	116.15	117.31	120.60
	media IGH = 543.10 $Wm^{-2}$						
C1	Smart Persistence	118.95	169.11	190.69	195.34	190.21	182.18
	Climatológico	163.64	163.70	163.78	163.86	163.95	164.03
	NN	110.63	143.90	157.06	162.11	162.09	162.878
	NN+ECMWF	110.32	139.15	148.96	149.10	148.03	148.30
	NN+ECMWF+WIND	108.98	136.76	145.84	147.93	148.42	148.70
	NN+ECMWF+SAT	104.75	134.37	142.82	145.41	147.31	147.88
	NN+ECMWF+WIND+SAT	105.36	134.43	142.57	145.24	147.92	149.22
	NN+SAT	105.34	136.59	147.31	151.94	156.71	157.03
	media IGH = 433.79 $Wm^{-2}$						

Tabla 5.14: Errores RMSE para todos los horizontes temporales  $h = 1...6$  en las estaciones de medida C0 y C1 para los datos de 2005

horas hasta las tres horas anteriores. En las últimas horas de predicción y sobre todo en la  $h = 6$ , el modelo NN+ECMWF iguala en incluso supera la aportación de los datos satelitales. Si se combinan ambos modelos NN+ECMWF+SAT se considera que se obtienen los mejores resultados en todos los rangos horarios. Se podría decir que en las tres primeras horas los datos satelitales son muy útiles y las últimas tres horas es más importante la aportación del ECMWF. Así que la combinación de ambos modelos puede ser una buena opción y será la que se utilice para comparar con el resto de modelos. El modelo (NN+ECMWF+SAT) consigue una mejoría respecto a las NN desde casi un 1% hasta un 2% para la primera y la última hora de predicción.

En la estación C1 la aportación de los datos satelitales se muestra eficaz únicamente en las tres primeras horas. El modelo NN+ECMWF muestra en las últimas horas una mejoría de hasta un 2% respecto al modelo NN+SAT. En cambio, la combinación de ambos modelos NN+ECMWF+SAT consigue las mejores predicciones para todos los horizontes temporales de predicción. Este modelo es, por lo tanto, el que se considera como mejor opción en la estación de C1. La comparación con el modelo de NN muestra unas predicciones con una diferencia desde el 1.5% hasta el 3.5% en la primera y la última hora de predicción a favor del modelo propuesto.

En la Figura 5.28 se pueden observar las distintas aportaciones de los datos obtenidos del modelo numérico ECMWF. Con el modelo NN+ECMWF, es decir, únicamente añadiendo los datos de radiación y nubosidad del ECMWF, se mejoran los resultados de las redes neuronales con datos terrestres NN. Este resultado ya se conoce de los datos obtenidos en la Sección 5.2.3.1. Al añadir los datos de velocidad y dirección del viento

estación	métodos	1 h	2 h	3 h	4 h	5 h	6 h
C0	Smart Persistence	17.03	22.95	25.80	26.60	26.05	25.54
	Climatológico	26.72	26.73	26.74	26.76	26.78	26.79
	NN	16.24	20.88	23.04	23.41	24.25	23.94
	NN+ECMWF	15.91	20.79	21.67	21.53	22.10	22.26
	NN+ECMWF+WIND	15.82	19.90	20.33	21.20	21.39	21.51
	NN+ECMWF+SAT	15.47	19.55	20.35	21.16	21.89	22.17
	NN+ECMWF+WIND+SAT	15.55	19.33	19.97	20.78	21.18	21.87
	NN+SAT	15.39	19.26	20.90	21.39	21.60	22.21
	media IGH = 543.10 $Wm^{-2}$						
C1	Smart Persistence	27.42	38.98	43.96	45.03	43.85	42.00
	Climatológico	37.72	37.74	37.76	37.77	37.79	37.81
	NN	25.50	33.17	36.21	37.37	37.37	37.55
	NN+ECMWF	25.43	32.08	34.34	34.37	34.12	34.19
	NN+ECMWF+WIND	25.12	31.53	33.62	34.10	34.21	34.28
	NN+ECMWF+SAT	24.15	30.98	32.92	33.52	33.96	34.09
	NN+ECMWF+WIND+SAT	24.28	30.99	32.86	33.48	34.10	34.40
	NN+SAT	24.28	31.49	33.96	35.03	36.13	36.20
	media IGH = 433.79 $Wm^{-2}$						

Tabla 5.15: Errores %rRMSE para todos los horizontes temporales  $h = 1...6$  en las estaciones de medida C0 y C1 para los datos de 2005

estación	métodos	1 h	2 h	3 h	4 h	5 h	6 h
C0	Smart Persistence	0.00	2.65	6.53	14.06	19.68	21.76
	Climatológico	-56.94	-13.38	3.09	13.53	17.46	17.91
	NN	4.68	11.51	16.54	24.32	25.15	26.51
	NN+ECMWF	6.62	11.86	21.49	30.38	31.78	31.68
	NN+ECMWF+WIND	7.08	15.58	26.34	31.50	34.07	34.11
	NN+ECMWF+SAT	9.16	17.08	26.27	31.62	32.51	32.06
	NN+ECMWF+WIND+SAT	8.65	18.00	27.65	32.86	34.71	32.99
	NN+SAT	9.67	18.38	24.29	30.85	33.31	31.84
	media IGH = 543.10 $Wm^{-2}$						
C1	Smart Persistence	0.00	-1.24	2.29	8.46	15.35	20.16
	Climatológico	-37.57	1.99	16.08	23.21	27.04	28.11
	NN	7.16	14.02	19.63	24.04	27.80	28.51
	NN+ECMWF	7.42	16.85	23.78	30.14	34.06	34.91
	NN+ECMWF+WIND	8.38	18.12	25.27	30.68	33.95	34.83
	NN+ECMWF+SAT	11.94	19.55	26.81	31.86	34.44	34.61
	NN+ECMWF+WIND+SAT	11.45	19.51	26.95	31.94	34.17	34.61
	NN+SAT	11.44	18.23	24.52	28.80	30.26	31.18
	media IGH = 433.79 $Wm^{-2}$						

Tabla 5.16: Errores %SKILL para todos los horizontes temporales  $h = 1...6$  en las estaciones de medida C0 y C1 para los datos de 2005

y de humedad relativa del ECMWF (NN+ECMWF+WIND) los resultados mejoran al anterior pero muy ligeramente. En la estación de C0, en el sur, esta mejoría no supera el 1%, aunque está presente en todo el horizonte temporal de predicción, en cambio en la estación de C1 la mejoría es más ligera y únicamente en las tres primeras horas. Por lo tanto, añadiendo datos del ECMWF se pueden conseguir resultados con unas mejorías de hasta el 2% en la estación del sur y el 3% en las del norte respecto al modelo NN.

Los resultados son positivos pero no presentan una mejoría significativa con los obtenidos al añadir únicamente las predicciones de radiación solar y nubosidad del ECMWF. Se observa que la presencia de datos de viento y humedad pueden aportar alguna mejoría pero se debe estudiar más en profundidad la capa de la atmósfera que se aporta a la red neuronal.

Por último, en la Figura 5.29 se pueden observar los modelos NN+ECMWF+WIND, NN+ECMWF+SAT y NN+ECMWF+WIND+SAT, para comparar los dos mejores modelos descritos anteriormente con una combinación de los datos satelitales con los datos de radiación y viento del ECMWF. Las diferencias entre los tres modelos no resultan en absoluto significativas. En la estación C1 el mejor modelo de predicción es el NN+ECMWF+SAT, por lo que los datos de viento no aportan ninguna mejoría. En cambio en C0, el modelo NN+ECMWF+WIND+SAT muestra en general los mejores resultados sobre todos en las últimas horas.

En la estación C0 los datos de viento parecen mostrar resultados mejores, aunque muy ligeramente. Este hecho nos conduce a pensar que se debe estudiar más en profundidad la altitud atmosférica a la que se extraen los datos de viento y humedad. En el futuro se debe estudiar la relación entre los datos satelitales y el viento en las diferentes capas de la atmósfera. Con los datos de viento disponibles en la actualidad, el modelo que se recomendaría para trabajar en toda la isla sería el NN+ECMWF+SAT.

En la Tabla 5.16 se puede estudiar la mejoría de cada uno de los modelos con respecto al Persistence (%SKILL). Todos los modelos de predicción basados en redes neuronales consiguen mejorar al Persistence para todos los horizontes temporales de predicción. En general los resultados en términos del error %rRMSE demostraron que en las estaciones del sur éstos eran menores, obteniéndose predicciones entre el 15% y el 22% de error, que en el norte, donde oscilan entre el 24% y el 34% de error. En cambio, en ambas estaciones de la isla las mejoras con respecto al modelo Persistence que nos ofrece el %SKILL muestra valores muy similares.

También se realizó el estudio de los resultados según las estaciones del año (invierno, primavera, verano y otoño). En las Tablas 5.17 y 5.18 se presentan los errores %rRMSE para todos los datos anuales y los datos de cada estación del año para los horizontes de predicción  $h = 1$  y  $h = 6$  respectivamente. En las Figuras 5.30 y 5.31 se muestran los valores del error por estaciones del año en términos de %rRMSE y %SKILL para el modelo NN+ECMWF+SAT tanto en C0 como en C1.

En la estación C0 se puede observar en general que los errores en los meses de invierno son los mayores para todas las horas. Los valores de %rRMSE se mantienen prácticamente constantes entorno al 35%. Por otro lado, las diferencias con los meses de primavera y verano son muy significativas, ya que estos últimos ofrecen errores entre el 15-20%. La estación C1 en el norte por contra muestra valores de errores más similares en todas las

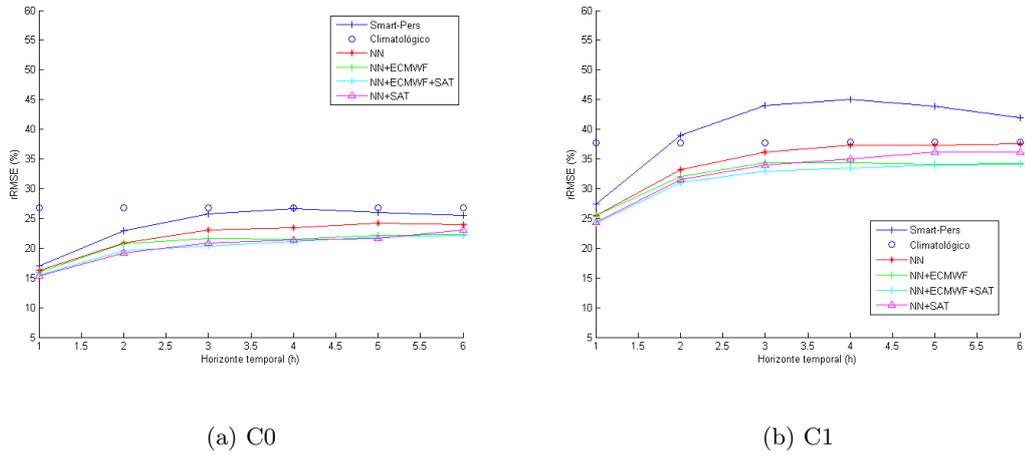


Figura 5.27: Evolución del %rRMSE en función del tiempo horizonte  $h = 1..6$  para los modelos de predicción basados en datos satélites y datos de radiación del ECMWF

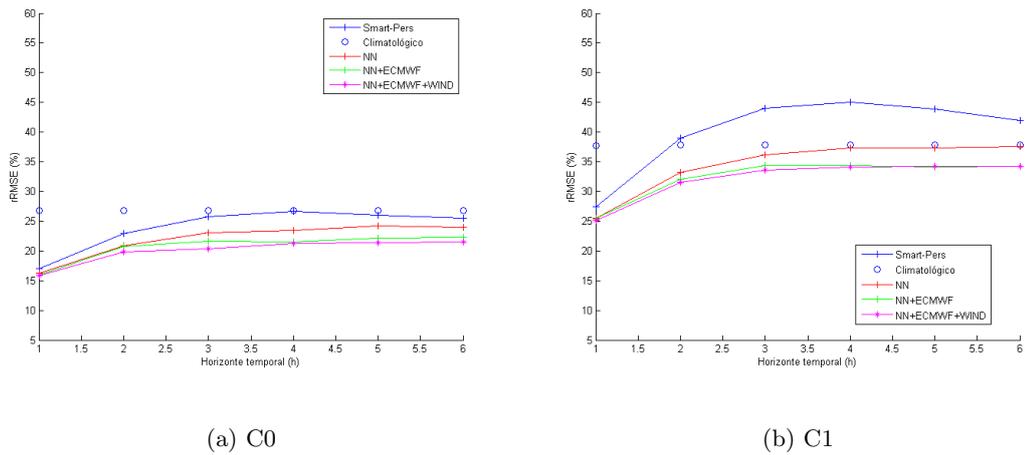


Figura 5.28: Evolución del %rRMSE en función del tiempo horizonte  $h = 1..6$  para los modelos de predicción basados en datos de radiación y datos de viento del ECMWF

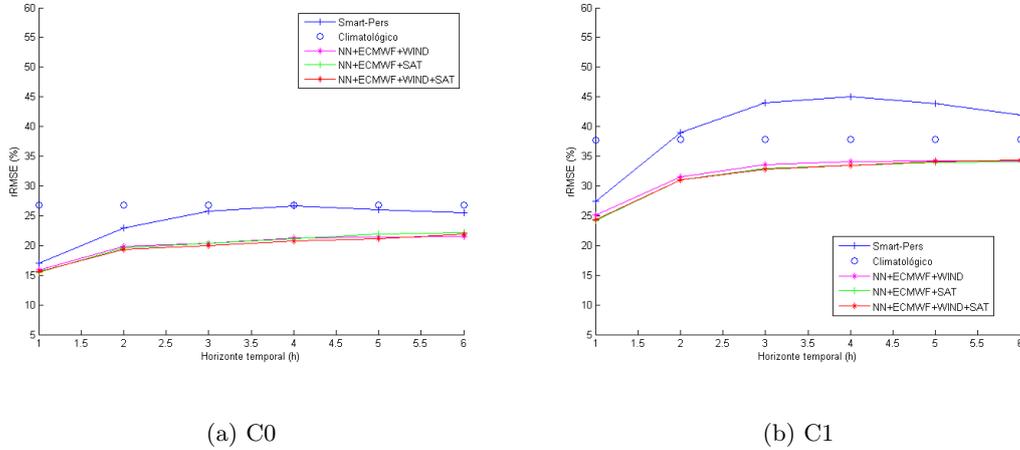


Figura 5.29: Evolución del %rRMSE en función del tiempo horizonte  $h = 1..6$  para los modelos de predicción basados en datos satélites, datos de radiación y datos de viento del ECMWF

estaciones del año. Como ya se ha dicho, los meses de verano presentan una nubosidad acusada por el efecto de los vientos Alisios, por lo que las predicciones no son tan eficaces como en el sur. En los meses de verano y primavera los errores se sitúan entorno al 30 % para todo el año, mientras que en invierno y otoño oscilan entre 35-40 % generalmente.

Los mejores resultados en invierno para la estación C0 se consiguen con el modelo NN+ECMWF, mientras que los datos de viento no aportan ninguna mejoría al modelo. En primavera los resultados son similares con todos los modelos ofreciendo los mejores resultados los modelos con datos satelitales (NN+SAT) y la combinación con datos de radiación del ECMWF (NN+ECMWF+SAT). En verano los datos satelitales consiguen los mejores errores de predicción, mientras que en otoño los datos de viento consiguen similares resultados.

En la estación de C1, los errores en invierno y primavera muestran menores valores para el modelo NN+SAT en las tres primeras horas y con el modelo NN+ECMWF en las siguientes. El modelo que combina ambos datos NN+ECMWF+SAT consigue los mejores resultados en general para todas las horas. Los resultados en verano con el modelo NN+ECMWF+WIND son los menores aunque muy similares al resto de modelos. Por contra, en otoño es claramente mejor el modelo que utiliza sólo datos satelitales NN+SAT.

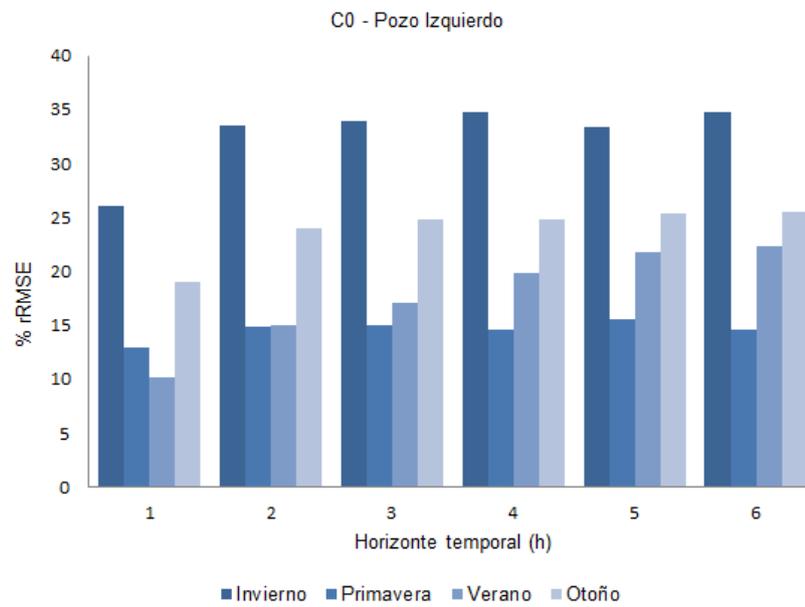
El %SKILL presenta un crecimiento paulatino conforme se aumenta el horizonte de predicción para todas las estaciones del año, excepto en invierno para la estación de C0. En la estación del sur C0 se consiguen mejoras significativas respecto al Persistence sobre todo en los meses de primavera, verano y otoño. En cambio en C1 la diferencia con el invierno no es tan acusada y consiguen peores valores en primavera.

estación	métodos	Anual	Invierno	Primavera	Verano	Otoño
C0	Smart Persistence	17.03	27.91	14.18	10.47	21.35
	Climatológico	26.72	43.69	20.59	22.99	29.03
	NN	16.24	26.38	13.32	10.72	21.11
	NN+ECMWF	15.91	26.22	13.42	10.43	19.85
	NN+ECMWF+WIND	15.82	27.39	13.48	10.77	19.62
	NN+ECMWF+SAT	15.47	26.02	12.93	10.11	19.02
	NN+ECMWF+WIND+SAT	15.55	27.81	13.75	10.70	18.85
	NN+SAT	15.39	26.50	12.55	9.99	19.03
C1	Smart Persistence	27.42	34.18	22.40	24.54	30.81
	Climatológico	37.72	42.35	36.46	32.63	41.79
	NN	25.50	31.09	21.58	22.23	29.02
	NN+ECMWF	25.43	30.70	20.96	22.61	29.48
	NN+ECMWF+WIND	25.12	30.35	20.92	21.96	29.49
	NN+ECMWF+SAT	24.15	28.65	20.05	21.42	28.46
	NN+ECMWF+WIND+SAT	24.28	28.75	20.07	21.75	28.44
	NN+SAT	24.28	28.24	20.36	21.96	28.37

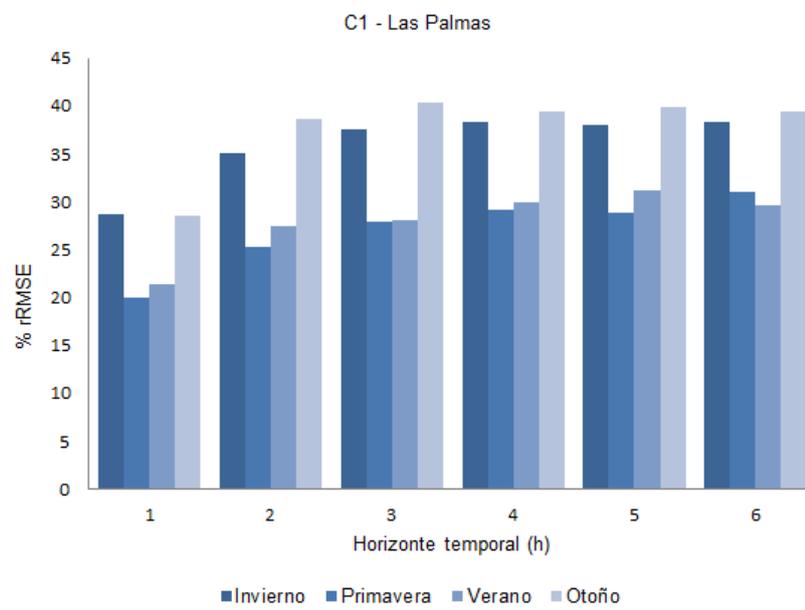
Tabla 5.17: Errores %rRMSE anuales y por trimestres para el horizonte temporal  $h = 1$  las estaciones C0 y C1 para el año 2005

estación	métodos	Anual	Invierno	Primavera	Verano	Otoño
C0	Smart Persistence	25.54	36.32	15.68	27.62	29.82
	Climatológico	26.79	44.44	20.64	23.20	29.47
	NN	23.94	36.79	16.61	23.49	27.69
	NN+ECMWF	22.26	33.66	14.57	22.85	26.40
	NN+ECMWF+WIND	21.51	35.75	15.48	22.25	24.98
	NN+ECMWF+SAT	22.17	34.82	14.60	22.37	25.56
	NN+ECMWF+WIND+SAT	21.87	40.15	14.57	22.44	25.35
	NN+SAT	22.21	35.55	14.68	21.82	25.80
C1	Smart Persistence	42.00	48.68	39.47	37.25	42.79
	Climatológico	37.81	41.99	36.78	32.41	42.49
	NN	37.55	42.24	35.34	32.90	40.68
	NN+ECMWF	34.19	37.56	31.23	29.66	40.64
	NN+ECMWF+WIND	34.28	37.64	31.33	29.57	40.83
	NN+ECMWF+SAT	34.09	38.27	31.04	29.61	39.35
	NN+ECMWF+WIND+SAT	34.40	38.44	31.38	30.05	39.82
	NN+SAT	36.20	41.70	35.72	30.37	38.32

Tabla 5.18: Errores %rRMSE anuales y por trimestres para el horizonte temporal  $h = 6$  las estaciones C0 y C1 para el año 2005

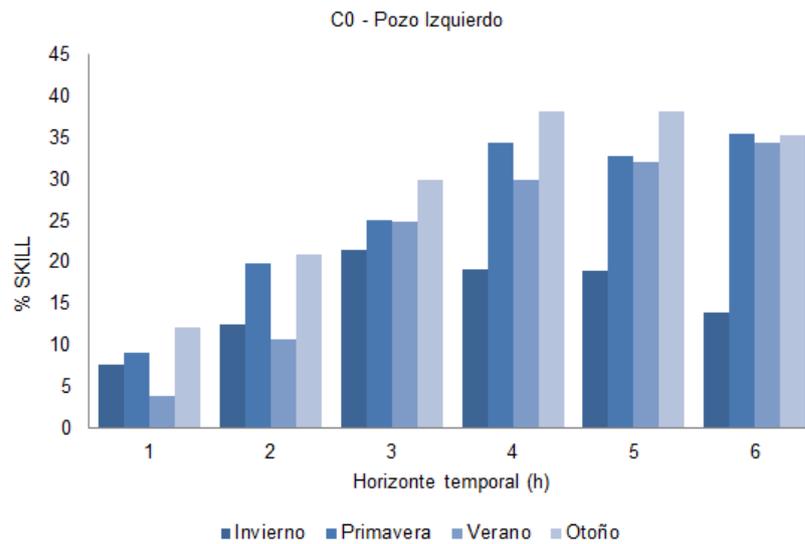


(a) C0

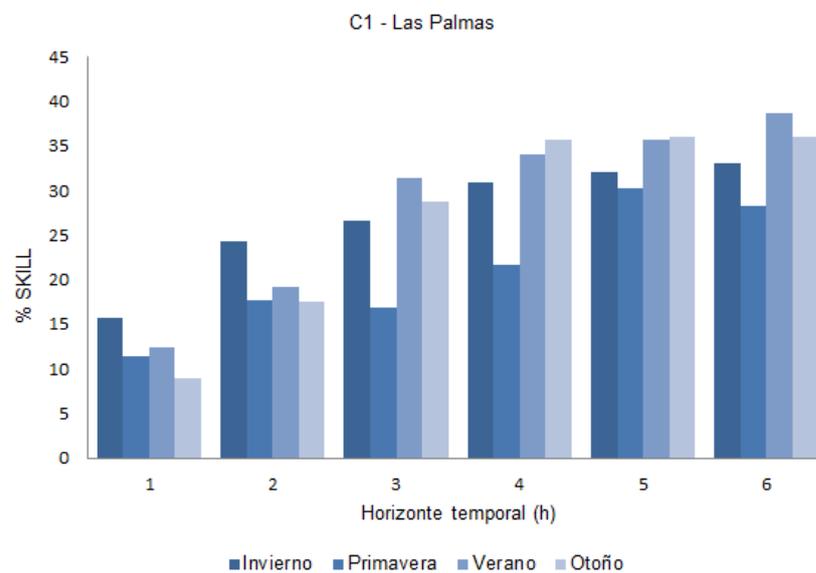


(b) C1

Figura 5.30: Evolución del %rRMSE en función del tiempo horizonte  $h = 1...6$  para el modelo basado en datos de radiación satélite y del ECMWF por cada trimestre



(a) C0



(b) C1

Figura 5.31: Evolución del %SKILL en función del tiempo horizonte  $h = 1..6$  para el modelo basado en datos de radiación satélite y del ECMWF por cada trimestre

### 5.3.3.2. Resultados según el tipo de día

En la Sección 2.3.4 se realizó el análisis de los datos de radiación solar en cada estación de medida en función de la media y la variabilidad diaria del índice de cielo despejado. Los días se dividieron en nueve tipos de días según la Tabla 5.9.

En las Tablas 5.19-5.21 se muestran los resultados obtenidos para todos los tipos de días con cada uno de los modelos de predicción. Los tipos de días cuyo número en el conjunto de datos del test no superen el 5% no se mostrarán en las tablas ya que los resultados no se consideran concluyentes.

En general en ambas estaciones los tipo I presentan los mejores resultados ya que son los días con menor variabilidad de la radiación solar. Las predicciones de los días tipo C, que representan los días con mayores medias de radiación solar, obtienen los mejores resultados en términos del %rRMSE, no superando el 18% en ninguna de las estaciones con el modelo NN+ECMWF+SAT. Los peores resultados se obtienen con los días nublados, tipo A, superando incluso el 90% ó 60% en la estación C0 y C1 respectivamente.

En la estación C0, observando exclusivamente la influencia de los modelos NN+SAT y NN+ECMWF, se puede concluir que para los días tipo B en general la aportación de los datos satelitales proporciona los mejores resultados para todos los horizontes temporales de predicción. La combinación de ambos métodos, NN+ECMWF+SAT, obtiene mejores resultados para los días BIII excepto para la última hora de predicción, mientras que en los días BII este modelo es ligeramente peor que la utilización exclusiva de los datos satelitales. En los conjuntos de días con mayores medias diarias de radiación y poca variabilidad, CI y CII, los resultados con los tres métodos son muy similares y no se observa que la combinación de los datos satelitales con el ECMWF mejore los resultados obtenidos por separado. En cambio en los días CIII, despejados pero con alta variabilidad, sí se observa una clara ventaja al utilizar el método NN+SAT en las primeras horas y el NN+ECMWF en las segundas. La combinación de ambos métodos no ofrece ninguna ventaja respecto a ambos modelos, aunque presenta en general un resultado medio para todos los horizontes temporales.

Los datos de viento y humedad relativa, NN+ECMWF+WIND, mejoran en general en la estación de C0 al modelo NN+ECMWF para todos los tipo de días, excepto para los días CI. En cambio únicamente en los días tipo BII y CII su combinación con los datos satelitales NN+ECMWF+WIND+SAT ofrece mejores resultados que el resto de modelos en general, mientras que en los días BIII el modelo NN+ECMWF+WIND obtiene mejores resultados excepto en las dos primeras horas.

Se debe recordar que en esta estación el 52% de los días son tipo C. La disparidad de resultados y la poca cantidad de datos en cada conjunto hace difícil y poco consistente tomar una decisión de los modelos que mejor predicen en cada tipo de día. En general se concluye que para la estación C0 la aportación de los datos de viento es representativa para los días tipo B y CII, que suponen un 69% de los días. En cambio para los días tipo CI y CIII la aportación de los datos satelitales y del ECMWF muestran mejores resultados.

En la estación C1 la mayoría de los días, un 77%, pertenecen los conjuntos tipo BII y BIII, con valores medio de radiación diaria intermedios y variabilidades medio altas. Los resultados de comparar las aportaciones de los datos satelitales y del ECMWF por

tipo día	métodos	1 h	2 h	3 h	4 h	5 h	6 h
AIII	Smart Persistence	65.09	82.47	81.32	80.81	81.43	83.95
	Climatológico	116.25	116.24	116.22	116.20	116.19	116.17
	NN	65.19	80.54	86.65	87.63	89.32	92.72
	NN+ECMWF	63.03	88.25	88.72	85.36	84.53	88.34
	NN+ECMWF+WIND	64.79	90.38	86.30	85.48	83.22	85.05
	NN+ECMWF+SAT	60.69	80.46	80.86	85.11	84.71	90.01
	NN+ECMWF+WIND+SAT	62.72	86.18	87.52	89.54	86.74	93.50
	NN+SAT	59.80	78.00	87.80	85.10	86.80	91.21
BIII	Smart Persistence	33.01	39.47	38.50	36.40	32.97	31.71
	Climatológico	30.57	30.57	30.57	30.57	30.57	30.56
	NN	29.73	34.56	32.38	29.95	29.37	28.42
	NN+ECMWF	29.92	34.51	31.24	28.93	29.31	28.67
	NN+ECMWF+WIND	29.24	32.00	28.33	28.25	28.13	27.81
	NN+ECMWF+SAT	28.13	32.81	30.22	28.44	28.81	28.87
	NN+ECMWF+WIND+SAT	27.98	31.48	28.71	29.50	29.64	28.71
	NN+SAT	28.93	32.76	32.27	29.88	28.84	27.75
CIII	Smart Persistence	19.13	23.43	28.57	26.49	22.67	20.56
	Climatológico	22.41	22.41	22.42	22.42	22.43	22.44
	NN	16.62	17.98	22.17	22.04	22.46	19.78
	NN+ECMWF	16.94	20.19	21.23	18.99	18.82	18.29
	NN+ECMWF+WIND	16.88	19.17	19.44	19.86	18.85	18.32
	NN+ECMWF+SAT	15.45	17.05	15.73	18.74	21.10	17.61
	NN+ECMWF+WIND+SAT	15.78	16.30	16.78	17.49	19.76	19.22
	NN+SAT	14.95	16.89	16.90	17.96	20.21	18.49

Tabla 5.19: Errores %rRMSE para los días tipo AIII, BIII y CIII para todos los horizontes temporales  $h = 1...6$  en la estación C0 con los datos de 2005

tipo día	métodos	1 h	2 h	3 h	4 h	5 h	6 h
BII	Smart Persistence	13.28	18.83	23.17	24.77	23.44	20.98
	Climatológico	18.52	18.53	18.53	18.53	18.53	18.54
	NN	13.47	17.00	18.35	19.05	18.86	18.55
	NN+ECMWF	12.64	16.87	17.56	17.25	17.46	17.58
	NN+ECMWF+WIND	12.09	15.50	16.40	16.19	16.20	16.06
	NN+ECMWF+SAT	12.35	15.17	16.49	16.45	17.31	16.79
	NN+ECMWF+WIND+SAT	11.77	14.11	14.82	15.85	15.05	15.60
	NN+SAT	11.98	14.14	16.02	16.32	16.61	16.91
CII	Smart Persistence	7.97	13.68	17.46	19.35	20.25	20.25
	Climatológico	19.69	19.70	19.71	19.72	19.72	19.73
	NN	7.79	12.46	16.11	16.52	18.23	17.40
	NN+ECMWF	7.44	11.04	12.68	13.22	13.35	13.54
	NN+ECMWF+WIND	7.54	10.36	11.11	12.21	12.57	12.75
	NN+ECMWF+SAT	8.13	10.81	12.72	13.04	14.01	13.45
	NN+ECMWF+WIND+SAT	8.35	10.41	11.96	11.84	11.97	12.28
	NN+SAT	7.54	10.71	12.61	13.25	12.76	13.67
CI	Smart Persistence	6.51	11.66	15.01	16.56	16.24	15.05
	Climatológico	18.89	18.90	18.93	18.96	19.00	19.03
	NN	6.34	10.12	13.09	14.35	14.69	15.40
	NN+ECMWF	5.86	9.42	10.57	11.25	11.40	11.42
	NN+ECMWF+WIND	6.17	8.88	11.02	11.89	12.38	12.73
	NN+ECMWF+SAT	6.91	8.94	10.15	10.90	11.07	11.79
	NN+ECMWF+WIND+SAT	6.81	9.57	10.77	11.22	12.09	12.89
	NN+SAT	6.36	9.32	9.73	11.08	11.82	11.64

Tabla 5.20: Errores %rRMSE para los días tipo BII, CII y CI para todos los horizontes temporales  $h = 1...6$  en la estación C0 con los datos de 2005

tipo día	métodos	1 h	2 h	3 h	4 h	5 h	6 h
AIII	Smart Persistence	44.51	52.49	56.05	61.70	63.78	62.68
	Climatológico	61.41	61.39	61.37	61.34	61.32	61.30
	NN	41.50	50.51	52.47	55.08	57.55	58.48
	NN+ECMWF	41.28	48.31	48.70	49.45	49.57	49.65
	NN+ECMWF+WIND	40.47	45.63	41.49	44.97	45.73	46.71
	NN+ECMWF+SAT	41.89	46.56	48.92	46.19	46.59	48.60
	NN+ECMWF+WIND+SAT	41.53	47.15	43.23	44.04	47.81	46.22
	NN+SAT	42.67	51.08	50.97	51.76	55.57	61.58
BIII	Smart Persistence	32.28	46.17	50.23	49.47	46.36	42.49
	Climatológico	37.28	37.28	37.29	37.29	37.29	37.29
	NN	29.10	37.36	39.63	39.75	38.38	37.60
	NN+ECMWF	29.27	36.39	38.06	37.74	36.52	36.57
	NN+ECMWF+WIND	28.97	36.20	38.17	37.69	36.94	37.32
	NN+ECMWF+SAT	27.31	35.07	36.45	36.54	36.54	36.89
	NN+ECMWF+WIND+SAT	27.53	35.23	36.69	36.54	36.80	36.90
	NN+SAT	27.38	35.32	37.11	37.12	37.84	36.50
BII	Smart Persistence	15.99	26.54	34.59	38.02	38.17	36.90
	Climatológico	26.69	26.72	26.73	26.74	26.74	26.75
	NN	15.62	22.51	26.36	28.16	28.92	29.46
	NN+ECMWF	14.91	21.14	23.66	24.56	25.18	24.61
	NN+ECMWF+WIND	15.23	23.50	27.81	25.60	26.73	26.50
	NN+ECMWF+SAT	14.93	20.47	23.36	24.95	24.19	24.18
	NN+ECMWF+WIND+SAT	14.97	20.64	24.20	25.52	25.81	25.54
	NN+SAT	15.06	21.25	24.92	26.87	27.29	26.63
CII	Smart Persistence	11.65	17.64	23.15	24.94	23.16	21.14
	Climatológico	32.03	32.04	32.06	32.07	32.08	32.10
	NN	12.68	17.86	21.71	23.40	24.11	24.96
	NN+ECMWF	11.97	15.16	18.06	18.40	18.62	19.03
	NN+ECMWF+WIND	11.95	17.70	21.58	19.44	22.42	21.25
	NN+ECMWF+SAT	11.02	12.52	16.51	17.78	18.02	18.32
	NN+ECMWF+WIND+SAT	11.39	13.40	17.38	18.49	19.45	19.58
	NN+SAT	10.93	14.53	16.55	19.22	19.21	22.05

Tabla 5.21: Errores %rRMSE para los distintos tipos de días para todos los horizontes temporales  $h = 1...6$  en la estación C1 con los datos de 2005

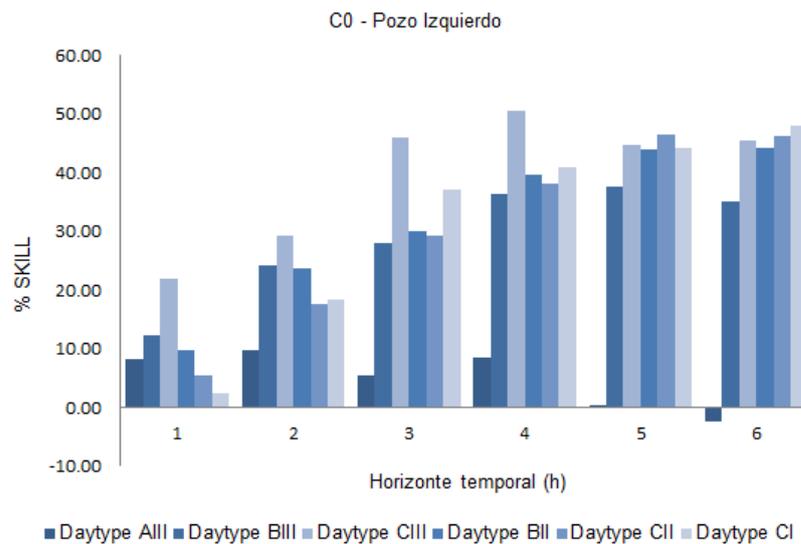
separado son más concluyentes. En los días tipo AIII y BII la mejor opción nos la aporta el modelo con datos del ECMWF, mientras que la aportación de los datos satelitales es más eficaz para los días de más variabilidad BIII. En los días despejados tipo CII, las tres primeras horas los datos satelitales son más eficaces mientras que para las últimas horas el modelo del ECMWF ofrece mejores resultados. En cualquier caso, la combinación de ambos NN+ECMWF+SAT consigue los mejores resultados en general para los cuatro tipos de días.

Los datos de viento no aportan grandes mejoras en la estación C1. Únicamente en los días tipo AIII parece ser la mejor opción de predicción, pero estos días son los que peores resultados se consiguen en general y el conjunto de datos no es significativo. En el resto de días la aportación de los datos de viento y humedad relativa no mejora los resultados obtenidos por el NN+ECMWF+SAT.

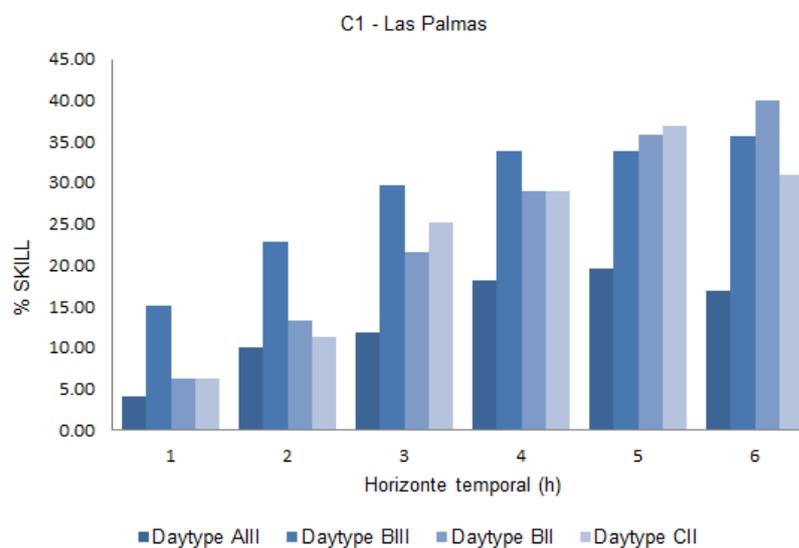
En la Figura 5.32 se pueden observar los resultados en términos del %SKILL para el modelo NN+ECMWF+SAT. En general se puede decir que los días tipo A obtienen los peores resultados, incluso decreciendo conforme aumentan las horas de predicción. En cambio el resto de tipos días los valores de comparación con el Persistence mejoran con el horizonte temporal. El conjunto de días con una mejoría más significativa en general es el tipo CIII. En la estación C1 los días tipo A siguen mostrando un peor resultado pero el resto de días mantienen unos valores de %SKILL similares.

En las Figuras 5.33-5.44 se muestran los ejemplos de las predicciones hora a hora para días concretos del año. En dichas figuras se observan los perfiles diarios de los distintos modelos de predicción con respecto al valor del perfil de radiación solar medido en las estaciones ( $IGH_m$ ) y modelo de cielo despejado ( $IGH_{cs}$ ). Los modelos no se han puesto en una sola gráfica para facilitar la visión. Las imágenes que se muestran corresponden a días tipo BIII, BII y CII por ser ejemplos de días representativos de los distintos conjuntos de datos. En cada tipo de día se representan las predicciones para un horizonte de predicción  $h = 1$  y  $h = 6$  respectivamente. Para los días tipo BII, con un perfil con una media intermedia y una alta variabilidad, las predicciones muestran una buena estimación para la primera hora, mientras que en las últimas horas la predicción de las variaciones no se consigue representar. Para  $h = 1$  todos los modelos consiguen reproducir la variabilidad que presentan las medidas reales aunque no se consigue evitar un retraso de una hora en las predicciones propias del Persistence. Para  $h = 6$  el modelo ECMWF realiza un suavizado de la curva y sobreestima el resultado, mientras que los datos satelitales permiten mantener algo más la variabilidad pero también sobreestiman.

En los días tipo BII, cuya variabilidad es menor, los resultados son bastante satisfactorios para todas las horas de predicción. Las pequeñas variaciones intradiarias siguen sin ser estimadas por los modelos de predicción, aunque en general generan una curva de predicción con la misma tendencia que la curva real. Los días tipo CII muestran unos resultados similares a los anteriores y se puede observar cómo los datos satelitales y del ECMWF mejoran las predicciones de las redes neuronales con datos terrestres únicamente.

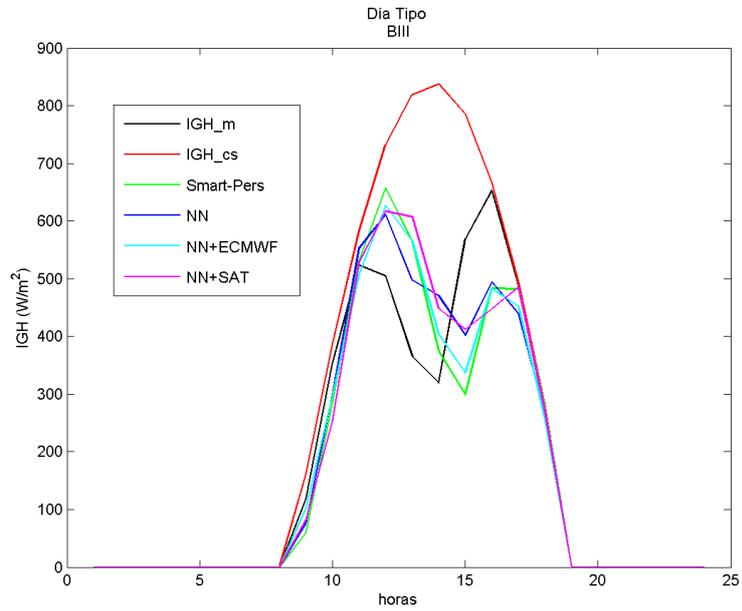


(a) C0

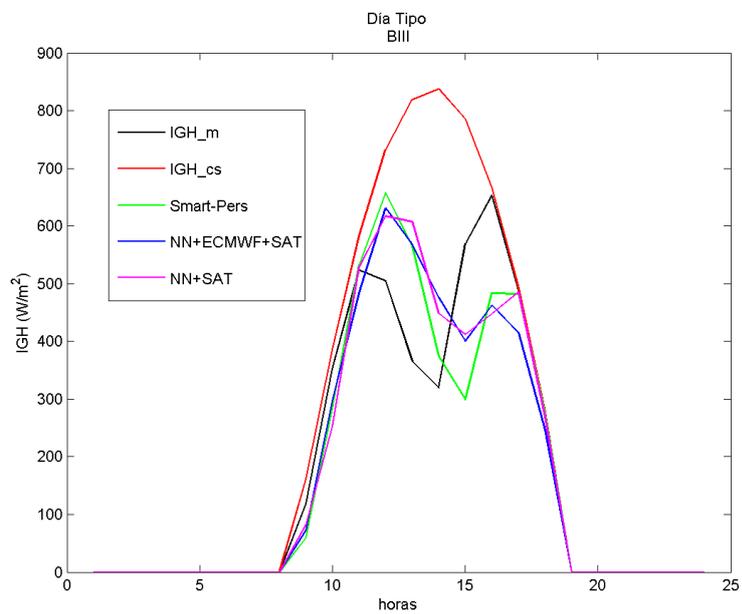


(b) C1

Figura 5.32: Evolución del %SKILL en función del tiempo horizonte  $h = 1..6$  para el modelo basado en datos de radiación satélite y del ECMWF según los distintos tipos de días

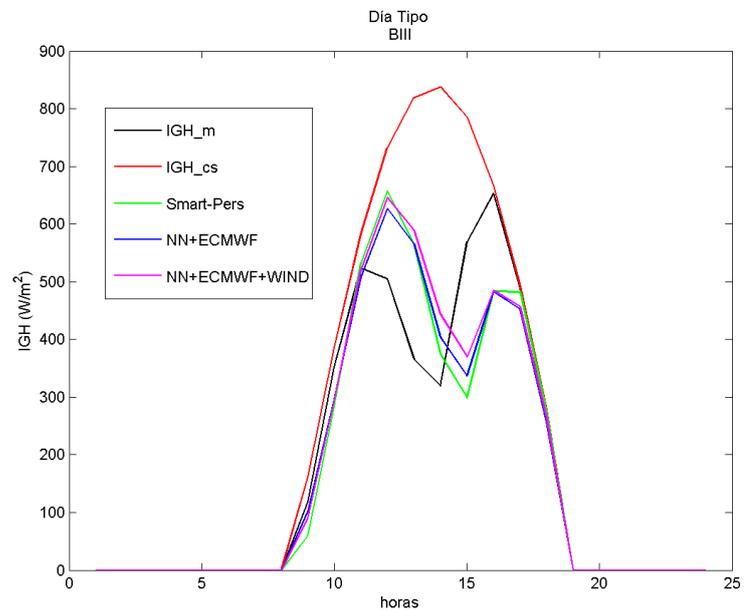


(a)

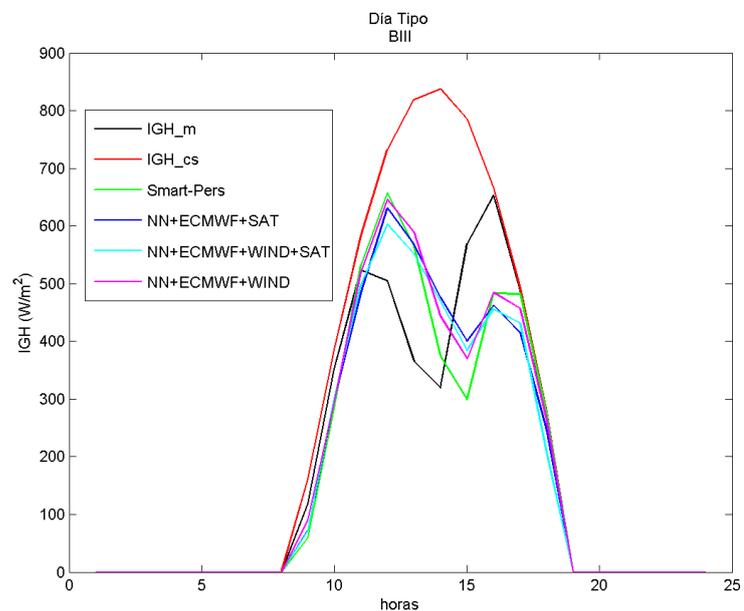


(b)

Figura 5.33: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo BIII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT

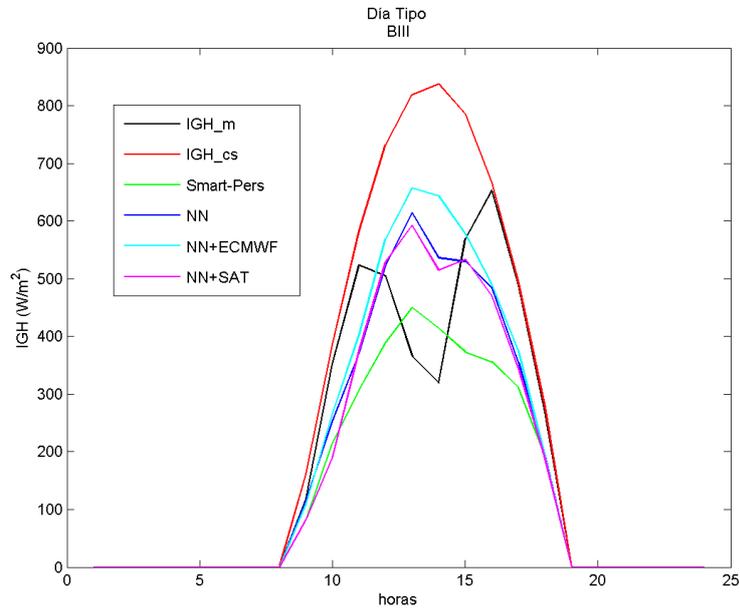


(a)

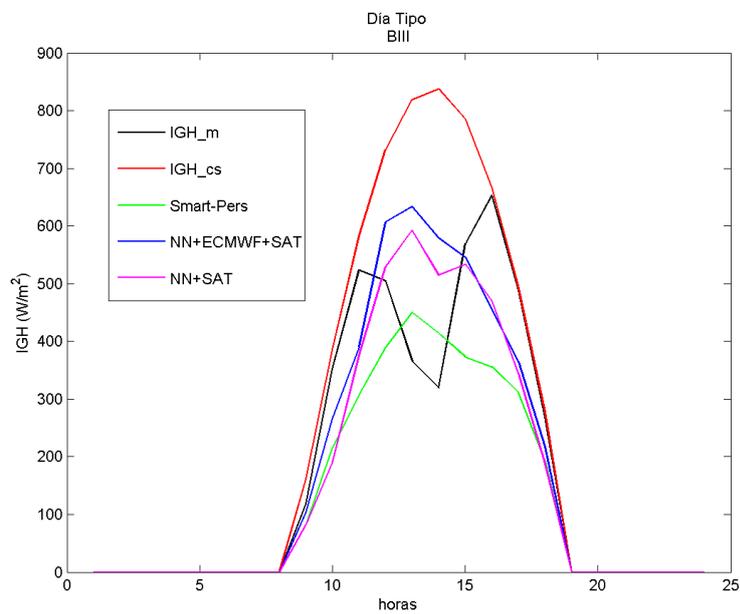


(b)

Figura 5.34: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo BIII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND+SAT y NN+ECMWF+WIND

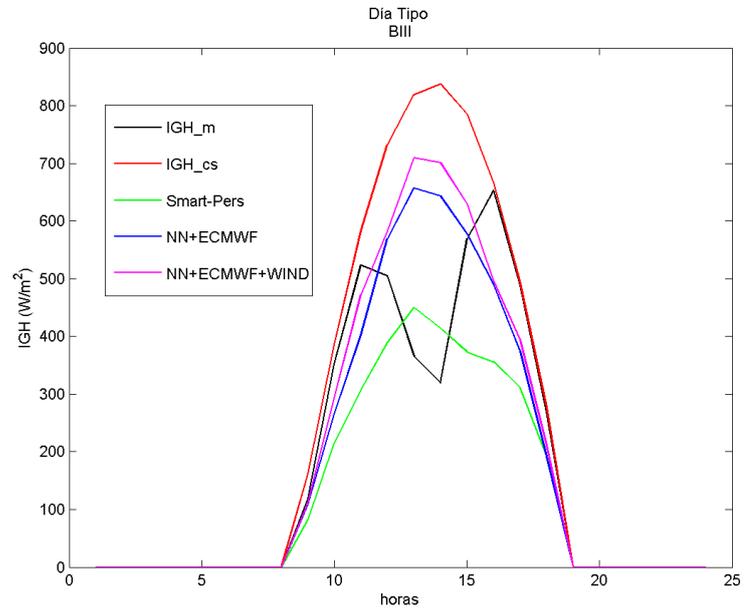


(a)

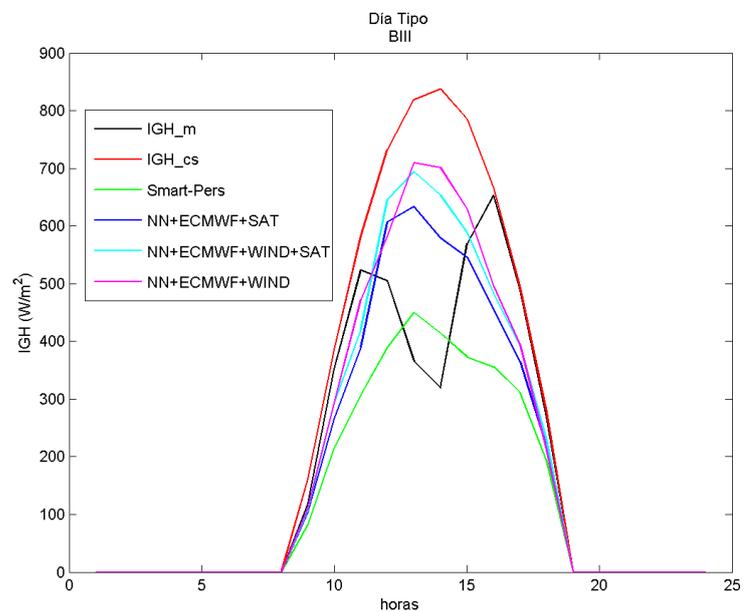


(b)

Figura 5.35: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo BIII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT

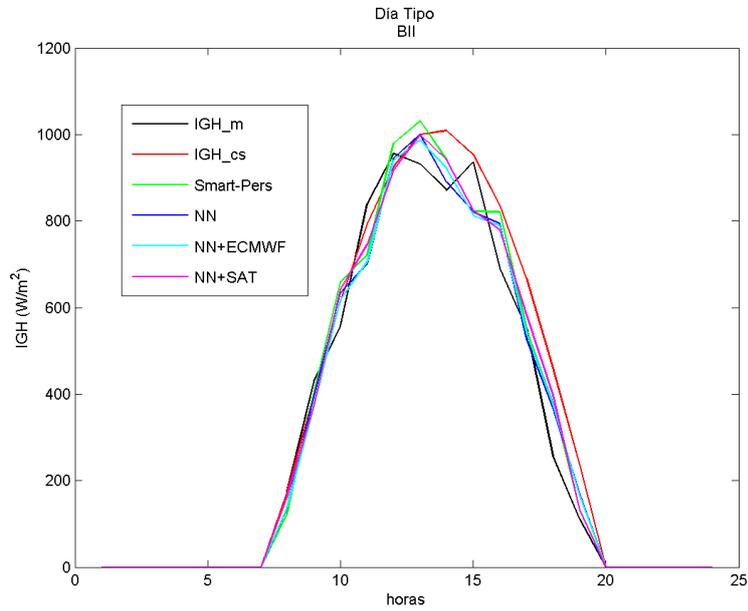


(a)

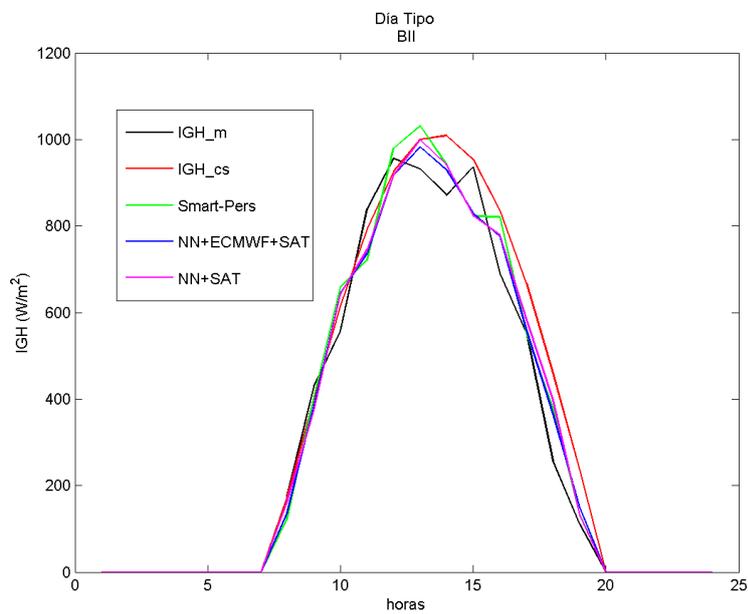


(b)

Figura 5.36: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo BIII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND+SAT y NN+ECMWF+WIND

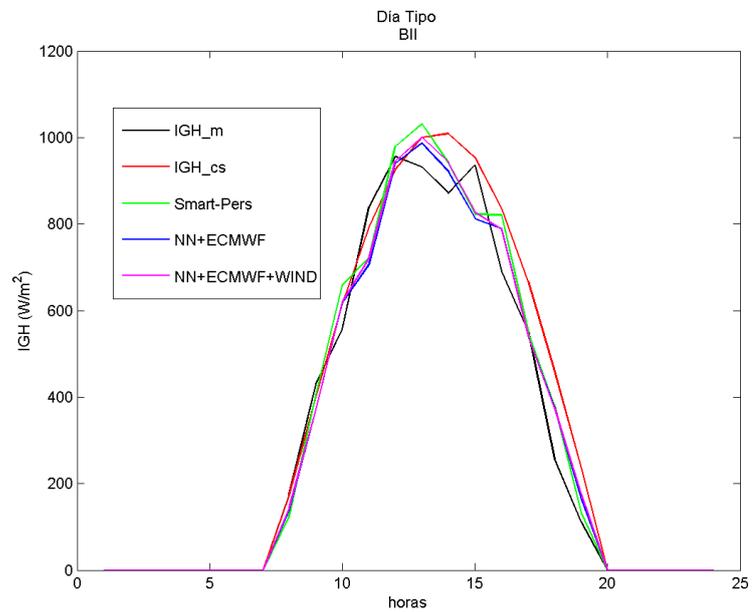


(a)

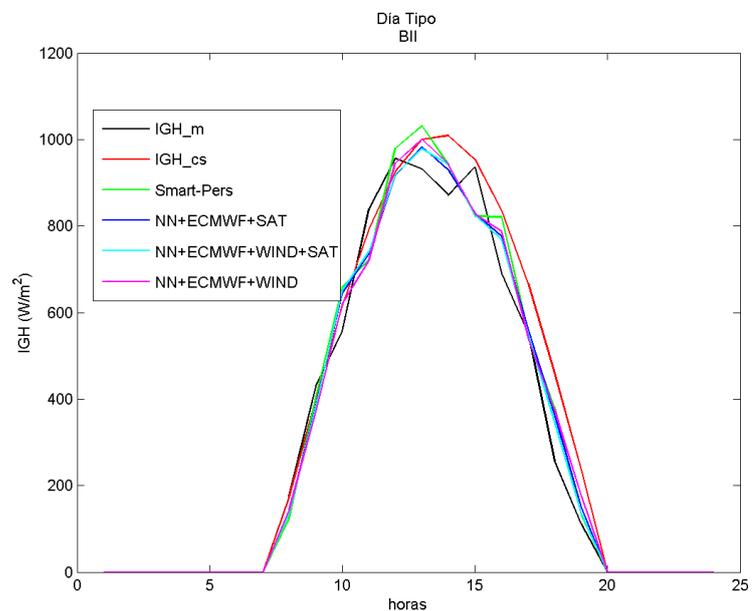


(b)

Figura 5.37: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo BII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT

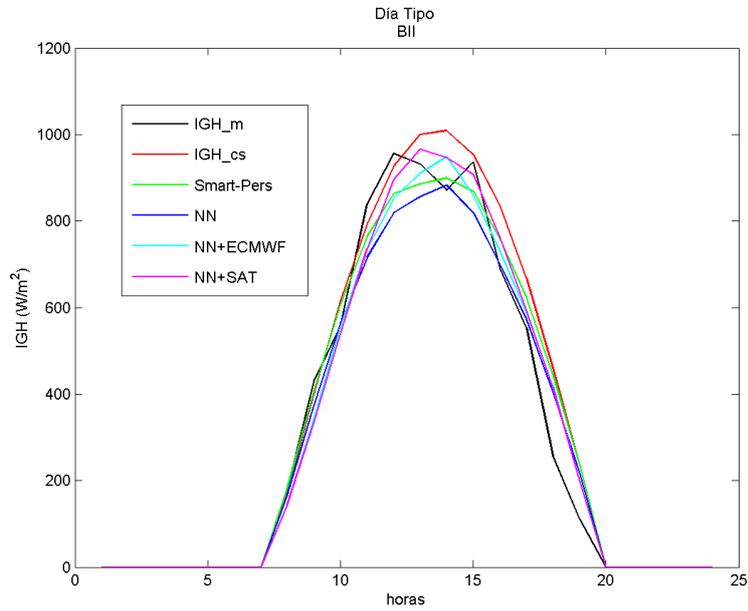


(a)

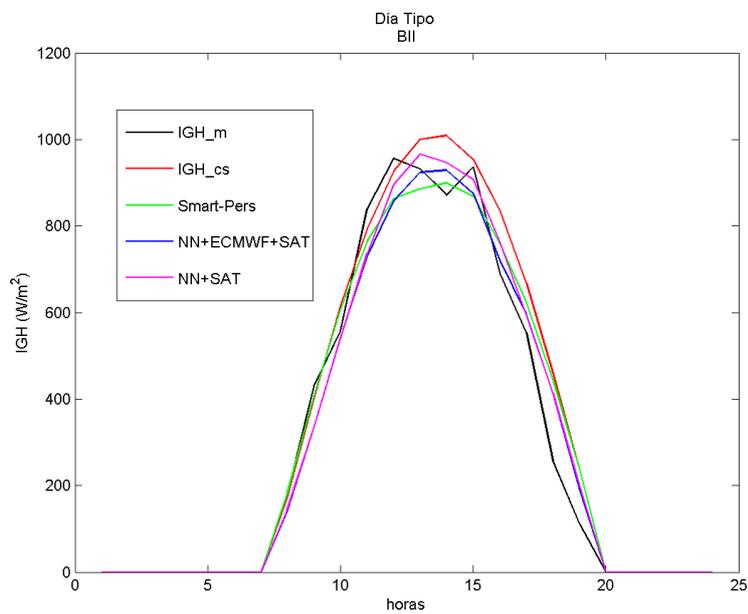


(b)

Figura 5.38: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo BII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND+SAT y NN+ECMWF+WIND

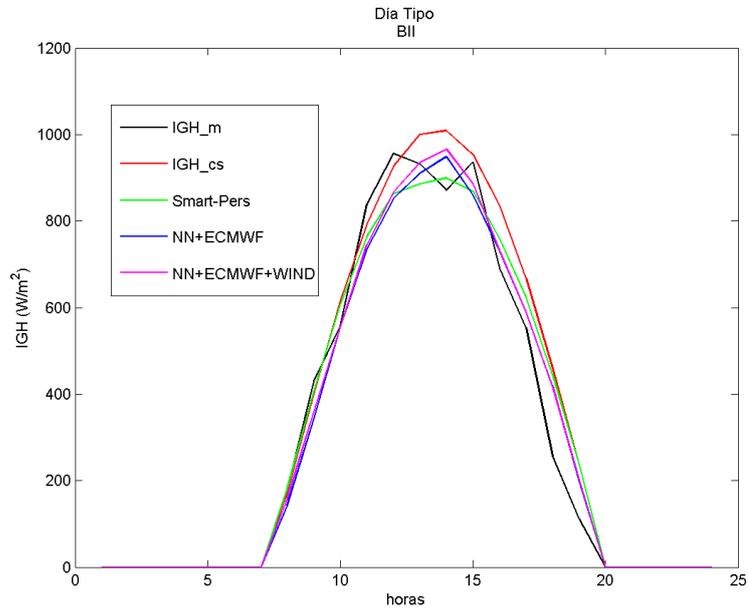


(a)

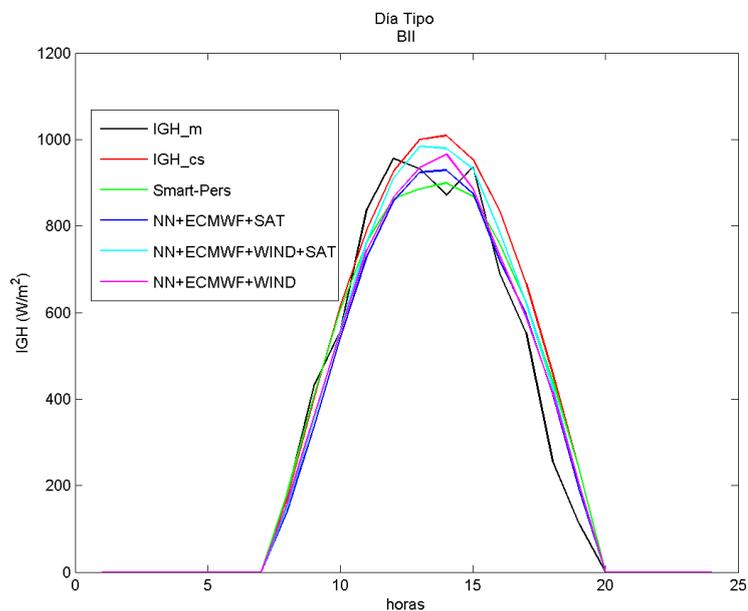


(b)

Figura 5.39: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo BII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT

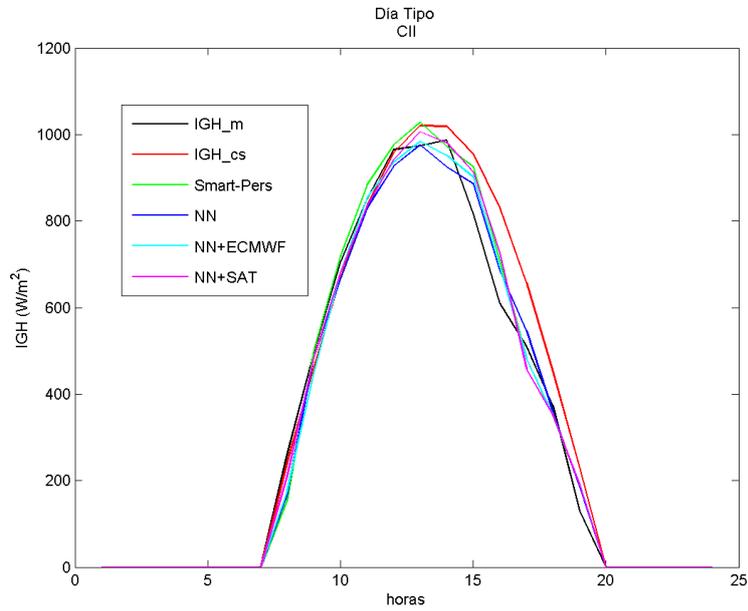


(a)

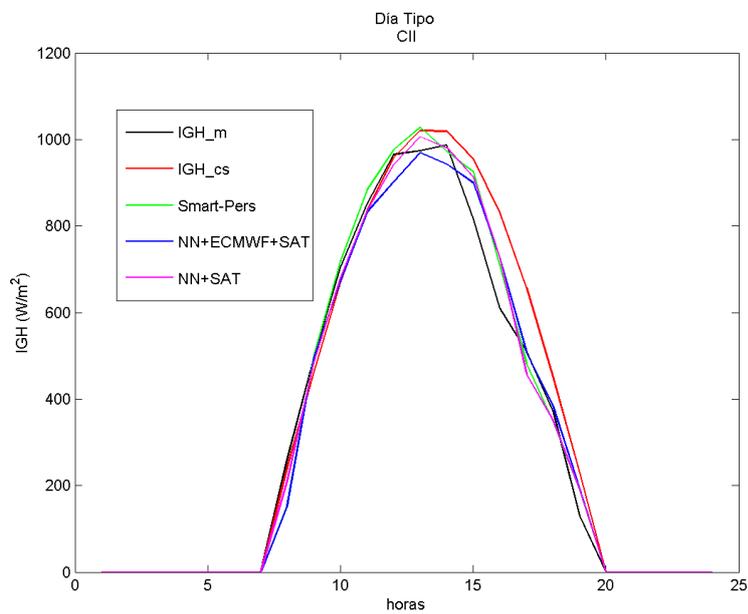


(b)

Figura 5.40: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo BII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND+SAT y NN+ECMWF+WIND



(a)



(b)

Figura 5.41: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo CII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT

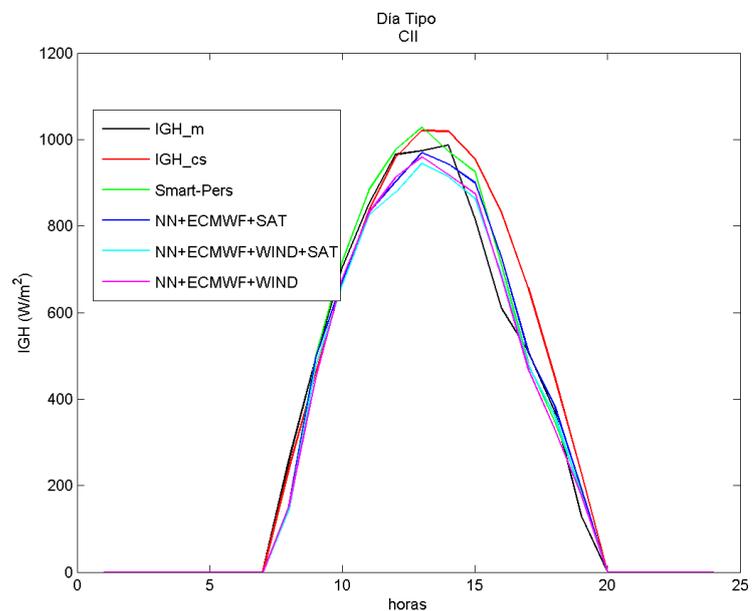
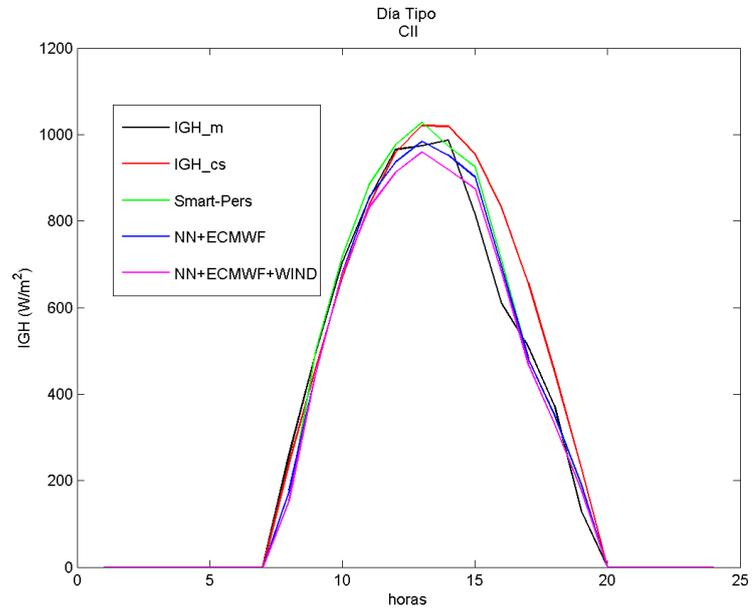
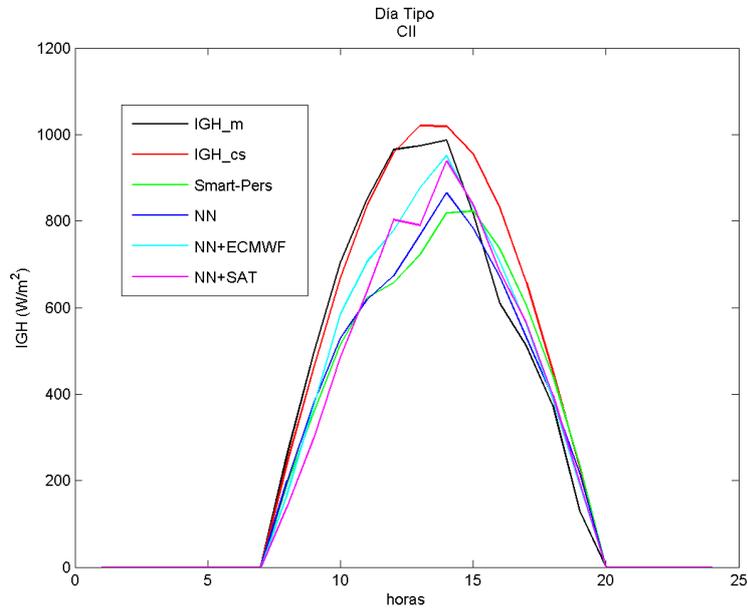
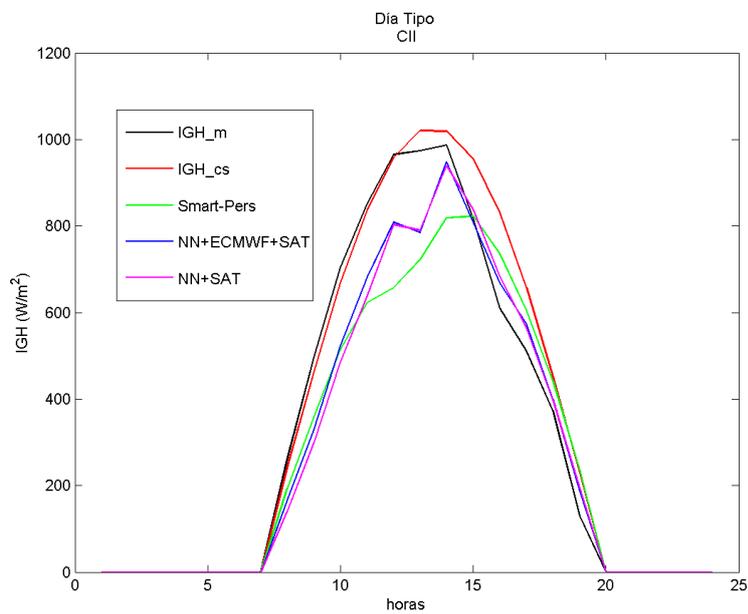


Figura 5.42: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo CII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND+SAT y NN+ECMWF+WIND

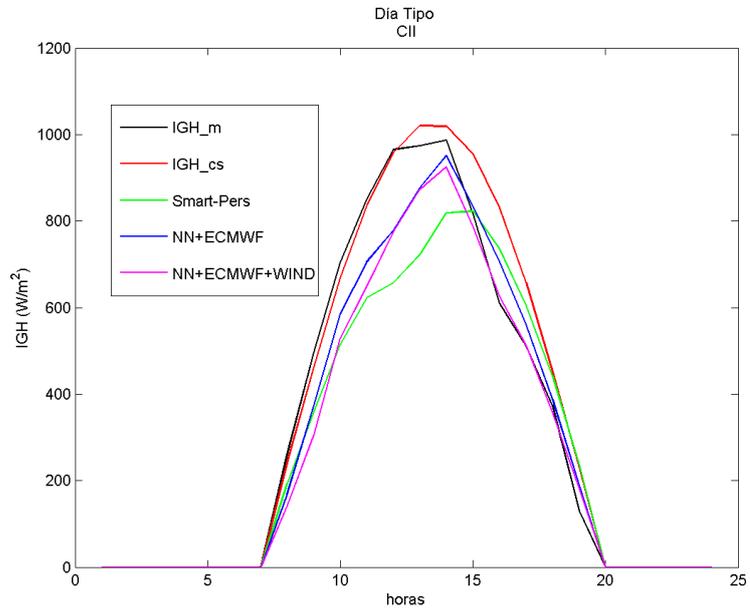


(a)

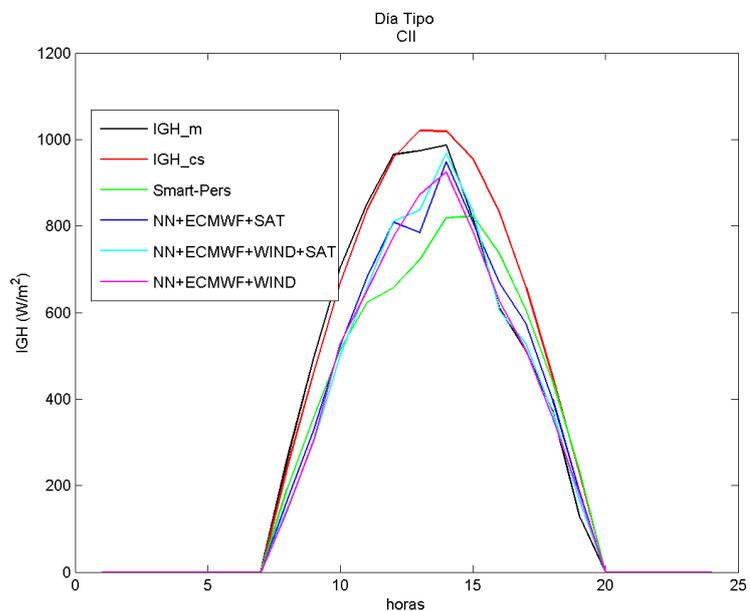


(b)

Figura 5.43: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo CII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT



(a)



(b)

Figura 5.44: Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo CII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND+SAT y NN+ECMWF+WIND

## 5.4. Producción de energía eléctrica a partir de las predicciones

De todos es conocido el interés del conocimiento de la radiación solar como fuente primigenia de generación eléctrica mediante generadores de tipo fotovoltaico o mediante centrales solares termoeléctricas. De hecho, el crecimiento de estas fuentes energéticas en el mix de los sistemas eléctricos de potencia actuales, hace que la investigación en la mejora de la precisión de las predicciones de la radiación solar a corto plazo, sea prioritaria. Esta tesis se enmarca en esta realidad.

Las predicciones de radiación solar horaria permitirán a los operadores del sistema eléctrico, estimar la producción eléctrica en los distintos puntos de la isla de Gran Canaria en los que se haya instalado centros de producción solar. Los modelos de estimación de la producción de energía eléctrica a partir de los datos de radiación nos permiten obtener los valores de energía para cada hora del día. En esta tesis, para mostrar la utilidad de las predicciones de radiación solar, se ha utilizado un modelo de generación eléctrica a partir de paneles fotovoltaicos [DR13]. Este modelo estima la producción de energía eléctrica a partir de los datos radiación solar, temperatura media y las características de los paneles fotovoltaicos.

En la Figura 5.45 se puede observar la comparación entre la producción estimada a partir de las predicciones y la producción basada en las medidas de radiación. El ejemplo muestra una instalación fotovoltaica de 480 paneles MSK modelo TCF200 situada en la estación de C0-Pozo Izquierdo. La producción en ambos casos se ha calculado utilizando el modelo fotovoltaico descrito en [DR13]. En un caso se muestra la curva obtenida a partir de las estimaciones de la radiación solar obtenidas con el modelo de predicción NN+ECMWF+SAT para los horizontes de predicción  $h = 1$  y  $h = 6$ . Mientras que la otra curva muestra la energía producida a partir de la radiación solar medida en las estaciones. El día tipo representado pertenece al conjunto definido como CII.

Los resultados obtenidos con el modelo de predicción recomendado se muestran como una herramienta eficaz a la hora de estimar la producción eléctrica de una instalación. Como ya se ha comentado, la predicción para un horizonte de una hora obtiene mejores resultados que para un horizonte de 6 horas. En la Figura 5.45 se pueden observar las diferencias de estimación intradiarias. En cuestión de energía total generada en el día, de acuerdo con las medidas la instalación produce  $234.23 \text{ kWh}$  mientras que la predicción a una hora estima una producción de  $234.33 \text{ kWh}$ , lo que supone un error de un  $0.04 \%$  en energía para predicción a una hora. Por otro lado, la predicción a seis horas estima una producción de  $209.70 \text{ kWh}$ . En este caso, el error en términos de energía es de  $10.47 \%$ .

Del análisis de las curvas representadas en la Figura 5.45, se puede concluir que la herramienta desarrollada es útil para la mejora de la gestión del funcionamiento de los sistemas eléctricos por parte del operador del sistema, si bien, sigue siendo necesaria la búsqueda de nuevos modelos que mejoren las predicciones.

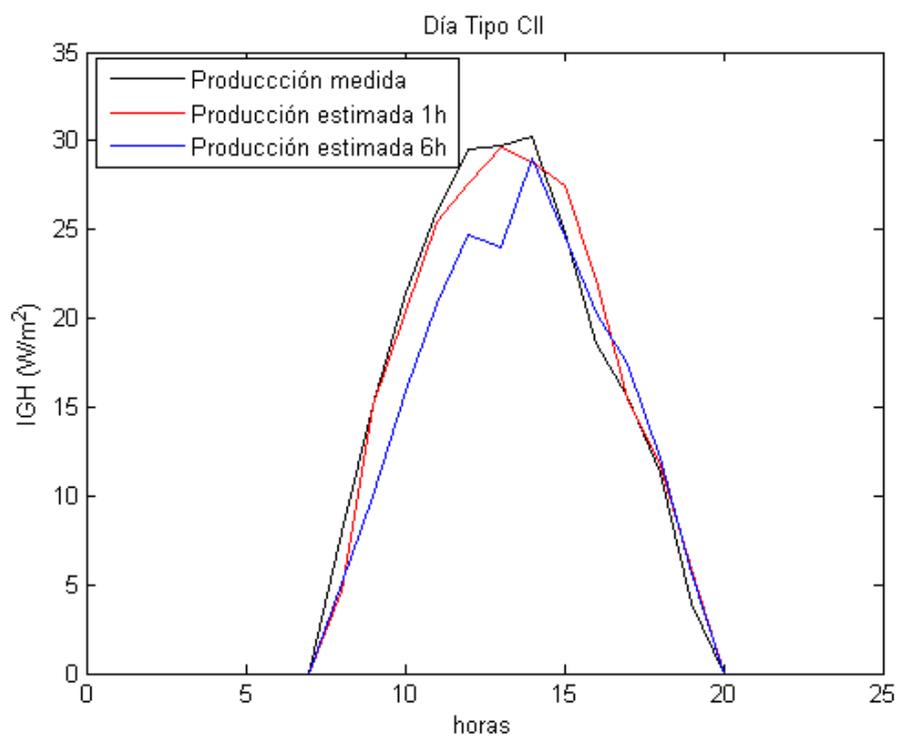


Figura 5.45: Comparación de la predicción de la energía eléctrica producida por una planta fotovoltaica obtenida a partir de la radiación estimada por el modelo NN+EMWF+SAT y medida en la estación de C0-Pozo Izquierdo.



## Capítulo 6

# Conclusiones y líneas futuras

Durante la realización de esta tesis se han ido obteniendo una serie de conclusiones que se exponen a continuación:

Los datos de radiación solar horaria disponibles para cada una de las estaciones de la isla de Gran Canaria nos permitieron establecer una caracterización del comportamiento meteorológico. Antes de trabajar con los datos de radiación solar se realizó un tratamiento y filtrado de los mismos y se calculó el índice de cielo despejado con el modelo de Bird & Hulstrom [Bird81] en todas las estaciones de medida, con un error medio del 4 %rRMSE. Se estableció una distribución de los diferentes tipos de días en función de la media diaria del índice de cielo despejado frente a la variabilidad diaria [Dambreville14b]. De esta manera se establecieron nueve tipos de días para cada estación de medida. Los resultados obtenidos nos permitieron realizar una primera división entre las estaciones del sur y del norte de la isla. En el sur las estaciones mostraron un mayor porcentaje de días tipo C, mientras que en el norte existe una gran cantidad de días más nublados tipo B y tipo III. En general en ninguna estación de la isla existe un número significativo de días con radiación muy baja tipo A. Estos datos permiten concluir que la zona de la isla con mejores condiciones de radiación solar es en general el sur, en concreto el suroeste, y la peor es el noreste. Esta división de la isla coincide con la dirección predominante de los vientos Alisios.

Para estudiar la cuadrícula de datos satelitales obtenidos del HelioClim-3, se creó una interfaz que permite visualizar los comportamientos climáticos particulares conocidos, como es la presencia de nubes durante los meses de verano en las estaciones del norte de la isla. Esta interfaz permite estudiar la evolución intradiaria de la radiación solar en cada píxel de la cuadrícula de datos satelitales. De esta manera se pueden observar escenarios diferentes provocados por los vientos Alisios y la inversión térmica en el norte de la isla, como pueden ser la aparición de nubes sin que hubiera presencia de las mismas anteriormente o la acumulación de nubes por el movimiento de las mismas por efecto del viento.

Las correlaciones entre la cuadrícula de datos satelitales obtenidos del HelioClim-3 y los datos de las estaciones de medida C0-Pozo Izquierdo y C1-Las Palmas, nos permiten tener una mejor comprensión de los microclimas de la zona. Las correlaciones muestran el grado de interrelación entre los datos terrestres de cada estación y cada uno de los píxeles satelitales disponibles. Esta correlación se realizó desfasando los datos satelitales desde

la hora  $h = 0$  hasta  $h = -3$  respecto al dato terrestre. Las correlaciones muestran altos valores de relación entre los píxeles del norte y la estación de C1, mientras que los píxeles del sur muestran mayor relación con la estación C0. Las imágenes obtenidas permiten realizar una clara división de los datos satelitales según su relación con cada estación, dividiendo la isla en dos. En los resultados que ofrecen las imágenes se puede ver reflejado el efecto de los vientos Alisios en la nubosidad sobre el norte de la isla. Se puede observar claramente la formación de nubes en el noreste y la protección que la elevación central de la isla ejerce sobre el suroeste de la misma. El estudio trimestral de las correlaciones también se mostró coherente con el conocimiento del clima de la zona. Se concluye que las correlaciones entre ambos conjuntos de datos es una herramienta válida para mostrar la climatología de la isla. Así pues, esta herramienta nos permite estimar los píxeles satelitales que presentan mayor relación con las estaciones de medida y por lo tanto se consideran más útiles para mejorar las predicciones de las Redes Neuronales Artificiales.

El estudio de los diferentes modelos de predicción de los datos de radiación nos condujo a realizar la comparación entre dos modelos lineales, el Autorregresivo (AR) y el Autorregresivo de Media Móviles (ARMA). Se estudió la complejidad de ambos modelos mediante las técnicas de autocorrelación parcial entre los datos de las series temporales y BIC (Bayesian Information Criterion). El modelo ARMA se mostró más eficaz a la hora de predecir la radiación que el modelo AR. El modelo ARMA consigue mejores resultados de predicción utilizando únicamente dos entradas de datos pasados de radiación solar, mientras que el modelo AR necesitaba de hasta 11 entradas para conseguir resultados óptimos.

En un primer momento, para estudiar los resultados de los diferentes modelos, se trabajó con dos años de medida por cada estación. Para todas las estimaciones realizadas con RNAs se estudió la complejidad del modelo, realizando la elección del número de neuronas y de entradas para cada caso particular. En este caso se disponía únicamente de datos de radiación terrestre recogida en las estaciones de medida y datos de radiación provenientes del ECMWF. Al utilizar las redes neuronales con los datos provenientes del ECMWF (NN+ECMWF) además de los datos terrestres, los resultados son claramente mejores en cualquier estación y horizonte temporal de predicción. Las predicciones mostraron diferentes resultados en las estaciones del norte y sur de la isla. Así, por ejemplo, en las estaciones del Sur este modelo mejora al modelo Persistence alrededor de un 1 % y un 12 % para los horizontes  $h = 1$  y  $h = 6$  respectivamente, mientras que en las estaciones del norte esta mejora varía entre un 2 % y un 17 %. Se concluye que la aportación de los datos de radiación del ECMWF mejoran los resultados de predicción en todas las estaciones de la isla.

A la hora de estudiar las predicciones horarias comparándolas con el perfil de radiación medido, el modelo NN+ECMWF reproduce los mejores resultados en general. En el muy corto plazo, no hay diferencia notable con los modelos naïve, afectados por el retraso del Persistence. En cambio, para las últimas horas de predicción los modelos que utilizan redes neuronales se ajustan mejor a la curva real. Se observa que el modelo con NN+ECMWF suaviza mucho la curva y no recoge las variaciones intradiarias, aunque capta mejor la tendencia general del día.

Por último se añadieron otros datos exógenos a las redes neuronales para analizar su comportamiento. En esta tesis se utilizaron los datos satelitales de radiación, la radiación

del ECMWF y el viento y humedad relativa del ECMWF. En este caso únicamente se disponía del año 2005 para las estaciones de C0 y C1.

Para establecer las variables exógenas más influyentes se estudiaron los píxeles de datos satelitales más adecuados para mejorar las predicciones. Se desarrolló una herramienta que permite realizar una selección de los datos satelitales en base a la intercorrelación con los datos terrestres. A la hora de probar los resultados, se demostró que en la estación C0 el mejor resultado se conseguía con un conjunto de píxeles diferente a la estación C1.

De la misma manera, se estudió la altitud atmosférica óptima de la que se introducirían los datos de viento y humedad para cada estación. Se establecieron tres simulaciones con diferentes altitudes, basadas en la bibliografía existente. De nuevo, los resultados obtenidos en C0 y C1 fueron diferentes. Los datos de viento y humedad calculados a 700 hPa fueron la mejor opción para C0 y los datos de viento a la altitud con humedad relativa superior al 95 % para C1.

Para poder explicar la influencia de cada uno de los datos exógenos en las predicciones se estudió primero la influencia de los datos satelitales y del ECMWF por separado y en combinación. En la estación de C0, para las tres primeras horas de predicción la aportación de datos satelitales (NN+SAT) consigue mejores resultados comparada con la aportación del ECMWF (NN+ECMWF). En cambio, en las tres últimas horas, la aportación del ECMWF consigue resultados similares e incluso mejores. En este caso la combinación de ambos modelos (NN+ECMWF+SAT) consigue los mejores resultados comparados con las NN. En la estación C1, se puede observar de manera más acusada el mismo efecto. En las tres primeras horas el mejor modelo es el que utiliza los datos satelitales (NN+SAT) y en las últimas horas el modelo con los datos del ECMWF (NN+ECMWF). De la misma manera, la combinación de estos modelos (NN+ECMWF+SAT) consigue los mejores resultados. Se concluye que para ambas estaciones la combinación de datos exógenos satelitales y del ECMWF proporcionan en general el mejor resultado de predicción.

Se ha estudiado también cómo afecta cada uno de los datos obtenidos del ECMWF, tanto de radiación como de viento y humedad, en las predicciones de radiación solar con RNAs. En la estación de C0, el modelo con los datos de viento y humedad (NN+ECMWF+WIND) muestra una ligera mejoría menor al 1 % en todos los horizontes temporales respecto al modelo (NN+ECMWF). En el caso de la estación C1, esta mejoría es igualmente ligera y únicamente en las tres primeras horas. Los resultados obtenidos al añadir el viento y la humedad son positivos pero no mejoran significativamente la aportación de los datos de radiación del ECMWF únicamente.

Por último, la combinación de los datos satelitales y los datos de radiación, viento y humedad relativa del ECMWF obtiene resultados dispares en ambas estaciones. En la estación C1 el mejor modelo de predicción es el NN+ECMWF+SAT, por lo que los datos de viento no aportan ninguna mejoría. En cambio en C0, el modelo NN+ECMWF+WIND+SAT muestra en general los mejores resultados sobre todo en las últimas horas.

A la vista de estos resultados, la presencia de datos de viento y humedad puede aportar alguna mejoría pero se debe estudiar más en profundidad la capa de la atmósfera que se aporta a la red neuronal. Se debe desarrollar una herramienta que permita hacer una elección de la altitud atmosférica y los píxeles satelitales que mejor se adapten a los datos

disponibles. Teniendo en cuenta las ligeras mejoras obtenidas con este último modelo y los datos de viento disponibles en la actualidad, el modelo que se recomendaría para trabajar en toda la isla sería el NN+ECMWF+SAT. Con este modelo, en la estación de C0-Pozo Izquierdo el %rRMSE aumenta con el horizonte temporal desde 15.47% para  $h = 1$  hasta 22.17% para  $h = 6$  y para C1-Las Palmas varía entre 24.15% y 34.09%. Por otra parte, este modelo mejora al modelo smart-Persistence alrededor de un 1.5% y un 3.5% para los horizontes  $h = 1$  y  $h = 6$  respectivamente, mientras que en las estaciones del norte esta mejora varían entre un 3% y un 8%. Se concluye que la aportación de los datos de radiación del ECMWF mejoran los resultados de predicción en todas las estaciones de la isla.

A partir de los datos de radiación obtenidos en las predicciones se puede obtener una estimación de la producción de energía eléctrica para cada hora del día. Las mejoras obtenidas en las predicciones para los horizontes temporales entre una y seis horas son muy importantes para el sistema eléctrico. De esta manera se consiguen reducir los costes de operación y mantenimiento del sistema, ya que ofrecen mayor control del mismo y reducen la incertidumbre. Para los generadores de energía eléctrica estas aproximaciones se pueden utilizar para mejorar las intervenciones en el mercado intradiario y permitir actualizar las decisiones tomadas en la gestión de la instalación. Se debe tener en cuenta que en el *Real Decreto 738/2015, de 31 de julio, por el que se regula la actividad de producción de energía eléctrica y el procedimiento de despacho en los sistemas eléctricos de los territorios no peninsulares*, el coste por los desvíos en los que incurran las instalaciones generadoras deberá ser abonado por sus titulares.

A la hora de estudiar el perfil diario de radiación solar horaria con los distintos modelos de predicción se demuestra que los datos del ECMWF tienden a suavizar las curvas y sobreestimar los resultados. Por otro lado, los datos satelitales consiguen reflejar ligeramente mejor las variaciones intradiarias de la radiación solar. Los resultados obtenidos con los modelos recomendados reproducen un perfil diario de radiación que consigue predecir en general la energía diaria. Sin embargo, los modelos no reproducen exactamente las variaciones horarias de la radiación solar. Se deben estudiar diferentes técnicas de predicción para conseguir mejores resultados de irradiancia a corto plazo.

Por último, como líneas futuras de investigación se establece la necesidad de realizar una selección de los píxeles satelitales de una manera más eficaz. La información que aportaba la cuadrícula de datos satelitales mostraba las variaciones climatológicas de la isla, pero la mejora obtenida al incluirlos en las redes neuronales fue ligera comparada con otros métodos.

De la misma manera los datos de viento y humedad han aportado ligeras mejoras respecto a los resultados conseguidos con los datos satelitales y del ECMWF en la estación C0. Se propone realizar un estudio en detalle del viento y la humedad en las diferentes capas de la atmósfera y encontrar aquella que ofrezca mejores resultados.

Las redes neuronales mejoran los resultados al añadir datos exógenos a la predicción pero se debería probar con otros modelos de predicción. En concreto es conveniente intentar modelizar el movimiento [Marquez13] y formación de la nubes, ya que uno de los problemas más importantes es la nubosidad en el norte de la isla. El estudio ha permitido establecer unos resultados diferentes según las estaciones y las épocas del años, por lo que una opción

a seguir sería la combinación de los modelos [Lauret12] óptimos en cada caso.



# Conclusions

During the course of this thesis project, the following conclusions were reached:

Based on the hourly solar radiation data available for each of the stations on the island of Gran Canaria, we were able to determine characteristic weather patterns. Before working with solar radiation ground data a quality assessment was completed and the clear sky index was estimated with Bird & Hulstrom [Bird81] model for every locations, obtaining a 4 %rRMSE average error. The various types of days were categorized based on comparing the mean intraday clear sky index with daily variability [Dambreville14b]. This allowed us to establish an average of nine types of days per station. The results obtained allowed us to make a preliminary division of stations in the southern and northern parts of the island. Stations in the south evidenced a higher percentage of Type C days, while those in the north registered a greater number of Type B and Type 3 cloudy days. In general, none of the stations on the island had a significant number of very low radiation Type A days. Based on this data, we may conclude that the southern part of the island, and particularly the southeast, has the best conditions in terms of solar radiation, whereas the northeast has the worst conditions. This division of the island coincides with the predominant direction of the trade winds.

To assess the satellite raw data, an interface was created to visualize a particular behaviour of a known meteorological pattern, such as static cloud cover on summer time over the city of Las Palmas. This interface showed us two different scenarios, such as accumulation of clouds brought by the predominant trade winds and the formation of local clouds without apparent influence from any actual clouds prior of the formation. Whereas these events apparently present different weather conditions, both of them are caused by the same background phenomena, the local thermal inversion at the north of the island and the Trade Winds.

The correlations between the satellite data grid obtained from HelioClim-3 and data from the C0-Pozo Izquierdo and C1-Las Palmas measurement stations allow for a better understanding of microclimates in the area. The correlations show the degree of interrelation between each station's ground data and each available satellite pixel. This correlation was carried out by lagging the satellite data from  $h = 0$  hours to  $h = -3$  with respect to the ground datum. The correlations indicate high relationship between pixels in the north and Station C1, while pixels in the south have a stronger correlation with Station C0. The images obtained allow for a clear division of satellite data based on their correlation with each station, dividing the island in two. The results offered by these images reflect the effect of the trade winds on cloudiness in the northern part of the island.

Cloud formations may clearly be observed in the northeast whereas in the southeast it is possible to observe the protection offered by the elevation in the central part of the island. The quarterly study of correlations was also coherent with knowledge about the climate in the area. It is possible to conclude that the correlations between both sets of data is a valid tool for demonstrating the island's climatology. This tool thus allows us to estimate which satellite pixels have a greater correlation with the measurement stations and may therefore be considered more useful for improving the forecast results of Artificial Neural Networks.

After studying the various radiation data forecasting models, we compared two lineal models: Autorregressive (AR) and Autorregressive Moving Average (ARMA). The complexity of these two models was studied using partial autocorrelation techniques between the time series data and BIC (Bayesian Information Criterion). The ARMA model proved more efficient than the AR model in forecasting radiation. The ARMA model achieved better forecast results using only two solar radiation data inputs, while the AR model required up to 11 inputs in order to achieve optimal results.

In order to study the various models, we initially worked with an average of two years of measurements for each station. During this thesis, number of hidden units and inputs were studied for every simulations. In this case, the only data available was the ground radiation registered at the measurement stations and radiation data from ECMWF. Using neural networks with data from ECMWF (NN+ECMWF) in addition to the ground data clearly improved forecast results for any station or forecast horizon. Forecast results were different for stations in the northern and southern parts of the island. Thus, for instance, for stations in the south this model improves upon the Persistence model by between 1% to 12% for the  $h = 1$  and  $h = 6$  forecast horizons respectively, whereas, for stations in the north, this improvement fluctuated between 2% and 17%. The conclusion is that inputting ECMWF radiation data improves forecast results in every station on the island.

When studying hourly forecasts and comparing them with the measured radiation profile, the NN+ECMWF model produced the best results overall. In the very short term, there is no notable difference compared to naïve models affected by the lag in Persistence. However, for the last hours of the forecast, models that rely on neural networks more closely followed the actual curve. It can be observed that the NN+ECMWF model decidedly softens the curve and does not reflect intraday variations, although it more accurately reflects the day's overall trend.

Lastly, other exogenous data was added to the neural networks in order to analyze their behavior. This thesis project used the satellite radiation data as well as radiation, wind and relative humidity data from ECMWF. In this case, the only data available for stations C0 and C1 was for the year 2005.

In order to determine which exogenous variables had the greatest influence, a study was made of the satellite data pixels best suited for improving forecasts. A tool was developed that makes it possible to select satellite data based on their intercorrelation with ground data. Tests of the results proved that the best result for station C0 was obtained using a different set of pixels than for station C1.

Similarly, the optimal atmospheric height from which the wind and humidity data should be inputted for each station was also studied. Based on the existing bibliography,

we set up three simulations using different heights. Again, the results obtained for C0 and C1 were different. Wind and humidity data estimated at 700 HPa were the best option for C0, whereas wind data at height with relative humidity above 95 % were best for C1.

In order to explain the influence of each exogenous datum on forecasts, satellite data and data from ECMWF were first studied separately and in combination. At Station C0, for the first three hours forecast, inputting satellite data (NN+SAT) obtained better results compared to inputting data from ECMWF (NN+ECMWF). However, in the last three hours, similar and even better results were obtained by including ECMWF data. In this case, the best results were obtained by combining both models (NN+ECMWF+SAT) as compared with NN. In Station C1, this same effect is more clearly observable. In the first three hours, the best model is the one that uses satellite data (NN+SAT) and in the last hours the best model is the ECMWF data model (NN+ECMWF). Similarly, the best results were obtained by combining these models (NN+ECMWF+SAT). The conclusion is that, in general combining exogenous satellite data and ECMWF data provides the best forecast results for both stations.

The effects of each datum obtained from ECMWF, including radiation, wind and humidity data, on solar radiation forecasts made using RNA's was also studied. In Station C0, the wind and humidity data model (NN+ECMWF+WIND) shows a slight improvement of less than 1 % in all forecast horizons compared to the (NN+ECMWF) model. In the case of Station C1, this improvement is also slight and only applies to the first three hours. The results obtained from adding wind and humidity are positive but do not offer a significant improvement compared inputting only ECMWF radiation data.

Lastly, combining satellite data with radiation, wind and relative humidity data from ECMWF produced uneven results in the two stations. In Station C1, the best forecasting model was NN+ECMWF+SAT, so the wind data did not offer any improvement. However, in C0, the NN+ECMWF+WIND+SAT model generally produced the best results, especially in the last hours.

In view of these results, the presence of wind and humidity data may offer some improvement, but it is necessary to study the layer of the atmosphere inputted in the neural network. It is necessary to develop a tool that will make it possible to select the atmospheric height and satellite pixels that are best suited to the available data. Taking into account the slight improvement obtained using this last model and the wind data currently available, we would recommend working with the NN+ECMWF+SAT model on the entire island. With this model, at the C0-Pozo Izquierdo station, %rRMSE increases with the forecast horizon from 15.47 % for  $h = 1$  to 22.17 % for  $h = 6$ , and for C1-Las Palmas, it fluctuated between 24.15 % and 34.09 %. On the other hand, this model improves upon the smart-Persistence model by between 1.5 % and 3.5 % for the  $h = 1$  and  $h = 6$  forecast horizons respectively, while, for stations in the north, this improvement fluctuated between 3 % and 8 %. The conclusion is that inputting ECMWF radiation data improves forecast results in every station on the island.

Based on the radiation data obtained in the forecasts, it is possible to obtain an estimate of electric power production for each hour of the day. The improvements obtained in forecasts for forecast horizons between one and six hours are very important for the electric power system. This helps reduce operating and maintenance costs for the system by

offering greater control and reducing uncertainty. These approaches may be used by electric power generators in order to improve interventions in the intraday market and allow them to revise management decisions with respect to the facilities. It is important to bear in mind that Royal Decree 738/2015, dated July 31st, which regulates power production as well as clearance procedures in electrical systems not located on the peninsula, specifies that any cost in which generating facilities may incur due to variability must be paid by the owners.

When studying the daily solar radiation profile using the various forecasting models, it has been proven that ECMWF data tend to soften the curve and overestimate results. On the other hand, satellite data manage to offer a slightly more accurate reflection of the intraday variations in solar radiation. The results obtained using the recommended models reproduce a daily radiation profile that can generally predict daily energy production. Nevertheless, the models do not reproduce hourly variations in solar radiation precisely. It is necessary to study the various forecasting techniques in order to obtain better results for short-term irradiance.

Lastly, future lines of research should carry out a more effective selection of satellite pixels. The information in the satellite data grid indicated climatological variations on the island, yet the improvement obtained by inputting these data in the neural networks was slight compared to other methods.

Similarly, wind and humidity data resulted in slight improvements with respect to the results obtained using satellite data and ECMWF data in Station C0. We propose carrying out a detailed study of wind and humidity in the various layers of the atmosphere in order to determine which layer offers the best results.

Neural networks improve results by adding exogenous data to the forecast, but other forecasting models should be tested. Specifically, it would be desirable to model [Marquez13] movement and cloud formation, as one of the key problems is cloudiness in the northern part of the island. The study allowed us to determine different results depending on the stations and the times of year, so one possible option would be combining the optimal models for each case [Lauret12].

# Índice de figuras

2.1. Espectro de longitudes de onda de la radiación solar . . . . .	16
2.2. Distribución espectral de la radiación solar . . . . .	17
2.3. Componentes de la radiación solar incidente en una superficie receptora . .	18
2.4. Órbita de la Tierra alrededor del Sol. Fuente [DR13] . . . . .	19
2.5. Plano de la eclíptica. Fuente [DR13] . . . . .	20
2.6. Trayectoria solar diaria para verano e invierno. Fuente [DR13] . . . . .	21
2.7. Esfera celeste y plano de la eclíptica. Fuente [DR13] . . . . .	21
2.8. Sistema de referencia sobre la superficie terrestre. Fuente [DR13] . . . . .	23
2.9. Distribución geográfica de la estaciones de medida del Instituto Tecnológico de Canarias en la isla de Gran Canaria . . . . .	28
2.10. Piranómetros Kipp & Zonnen. Cortesía del ITC . . . . .	33
2.11. Piranómetro con anillo de sombra para eliminar la radiación directa. Cortesía del ITC . . . . .	34
2.12. Shadow effect sobre la estación de medida de C1-Las Palmas. Cortesía del Mines ParisTech . . . . .	35
2.13. Conjunto de datos de la estación C0-Pozo Izquierdo tras realizar el filtro . .	36
2.14. Conjunto de datos de la estación C5-Sta. Brígida tras realizar el filtro . . .	36
2.15. Distribucion de los datos diarios del índice de cielo despejado para las estaciones de Gran Canaria en función de la media y la variabilidad diaria. . . . .	38
2.16. Representación de la curva diaria de radiación IGH (rojo) frente a la curva diaria de cielo despejado (azul) para cada uno de los tipo de días en la estación de C0-Pozo Izquierdo . . . . .	40
2.17. Representación de la curva diaria de radiación IGH (rojo) frente a la curva diaria de cielo despejado (azul) para cada uno de los tipo de días en la estación de C1-Las Palmas . . . . .	41
2.18. Distribución geográfica de los datos obtenidos del modelo de predicción numérica ECMWF para la isla de Gran Canaria . . . . .	45

3.1.	Distribución geográfica de los datos obtenidos del modelo basado en imágenes satelital HC3v5 para la isla de Gran Canaria . . . . .	51
3.2.	Distribución de los tipos de días utilizando tanto datos satelitales como terrestres en base a la media y la variabilidad diaria del índice de cielo despejado para la estación C0-Pozo Izquierdo . . . . .	53
3.3.	Distribución de los tipos de días utilizando tanto datos satelitales como terrestres en base a la media y la variabilidad diaria del índice de cielo despejado para la estación C1-Las Palmas . . . . .	53
3.4.	Evolución intradiaria de la radiación global horizontal y del índice de cielo despejado de los datos satelitales para cada 30 minutos. Día 07/07/2005. . .	55
3.5.	Evolución intradiaria de la radiación global horizontal y del índice de cielo despejado de los datos satelitales para cada 30 minutos. Día 12/08/2005. . .	55
3.6.	Mapa de intercorrelación anual para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de $h = 0, 1, 2$ & 3 horas en la estación C0-Pozo Izquierdo. . . . .	57
3.7.	Mapa de intercorrelación anual para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de $h = 0, 1, 2$ & 3 horas en la estación C1-Las Palmas. . . . .	58
3.8.	Mapa de intercorrelación en verano para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de $h = 0, 1, 2$ & 3 horas en la estación C0-Pozo Izquierdo. . . . .	59
3.9.	Mapa de intercorrelación en verano para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de $h = 0, 1, 2$ & 3 horas en la estación C1-Las Palmas. . . . .	60
3.10.	Mapa de intercorrelación en otoño para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de $h = 0, 1, 2$ & 3 horas en la estación C0-Pozo Izquierdo. . . . .	61
3.11.	Mapa de intercorrelación en otoño para el índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de $h = 0, 1, 2$ & 3 horas en la estación C1-Las Palmas. . . . .	62
3.12.	Mapa de intercorrelación anual para la variación índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de $h = 0, 1, 2$ & 3 horas en la estación C0-Pozo Izquierdo. . . . .	64
3.13.	Distribución normal de la correlación anual para la variación del índice de cielo despejado entre los datos de tierra y cada píxel de datos satelitales con un retraso de $h = 0, 1, 2$ & 3 horas en la estación C0-Pozo Izquierdo. .	65
4.1.	Ejemplo de SACF y SPACF para serie temporal estacionaria aleatoria. En rojo se representan los intervalos de elección . . . . .	74
4.2.	Reproducción de una célula neuronal [Isasi04] . . . . .	80
4.3.	Estructura general una neurona artificial . . . . .	81

4.4. Estructura general una red multicapa . . . . .	82
4.5. Estructura general una neurona artificial sin memoria . . . . .	84
4.6. Ejemplo de los elementos de una red multicapa general . . . . .	86
4.7. Arquitectura del Perceptrón Multicapa (MLP) . . . . .	87
4.8. Esquema de aprendizaje supervisado de una RNA . . . . .	90
4.9. Esquema del avance en la optimización. En la izquierda se toma un valor alto de coeficiente de aprendizaje y en la derecha valores pequeños. . . . .	91
4.10. Función excesivamente ajustada al conjunto de datos mostrado (overfitting) . . . . .	96
4.11. Convergencia de los hiperparámetros con cada iteración . . . . .	100
4.12. Resultados de la técnica de log of evidence para decidir el número de neuronas ocultas óptimo para resolver un problema. Ejemplo de la estación C6 en Gran Canaria. . . . .	102
4.13. Resultados de la técnica de selección automática del número de entradas necesarias para resolver un problema. Ejemplo de la estación C5 en Gran Canaria. . . . .	103
5.1. SPACF de la serie temporal del índice de claridad horario para la estación de C0-Pozo Izquierdo. En rojo se representan los intervalos de elección . . .	109
5.2. SPACF de la serie temporal del índice de claridad horario para la estación de C5-Sta. Brígida. En rojo se representan los intervalos de elección . . . .	110
5.3. Valores del parámetro BIC obtenidos para la estación de C0-Pozo Izquierdo. Se muestran la simulaciones de un modelo AR con órdenes desde 1 hasta 14 para todos los horizontes temporales de predicción $h = 1..6 h$ . . . . .	111
5.4. Valores del parámetro BIC obtenidos para la estación de C5-Sta. Brigida. Se muestran la simulaciones de un modelo AR con órdenes desde 1 hasta 14 para todos los horizontes temporales de predicción $h = 1..6 h$ . . . . .	111
5.5. Error cuadrático medio relativo %rRMSE para la estación de C0-Pozo Izquierdo. Se muestran la simulaciones de un modelo AR con órdenes desde 1 hasta 14 para todos los horizontes temporales de predicción $h = 1..6 h$ . . . . .	112
5.6. Error cuadrático medio relativo %rRMSE para la estación de C5-Sta. Brigida. Se muestran la simulaciones de un modelo AR con órdenes desde 1 hasta 14 para todos los horizontes temporales de predicción $h = 1..6 h$ . . . . .	112
5.7. SACF de la serie temporal del índice de claridad horario para la estación de C0-Pozo Izquierdo. En rojo se representan los intervalos de elección . . .	114
5.8. SACF de la serie temporal del índice de claridad horario para la estación de C5-Sta. Brígida. En rojo se representan los intervalos de elección . . . .	114
5.9. Error cuadrático medio relativo %rRMSE y BIC para la estación de C0-Pozo Izquierdo. Se muestran la simulaciones de un modelo ARMA para todas las combinaciones posibles de $p$ y $q$ . En cada parte se observa a) BIC 1h, b) %rRMSE 1h . . . . .	116

5.10. Error cuadrático medio relativo %rRMSE y BIC para la estación de C0-Pozo Izquierdo. Se muestran la simulaciones de un modelo ARMA para todas las combinaciones posibles de $p$ y $q$ . En cada parte se observa a) BIC 6h, b) %rRMSE 6h . . . . .	117
5.11. Error cuadrático medio relativo %rRMSE y BIC para la estación de C5-Sta. Brígida. Se muestran la simulaciones de un modelo ARMA para todas las combinaciones posibles de $p$ y $q$ . En cada parte se observa a) BIC 1h, b) %rRMSE 1h . . . . .	118
5.12. Error cuadrático medio relativo %rRMSE y BIC para la estación de C5-Sta. Brígida. Se muestran la simulaciones de un modelo ARMA para todas las combinaciones posibles de $p$ y $q$ . En cada parte se observa a) BIC 6h, b) %rRMSE 6h . . . . .	119
5.13. Error cuadrático medio relativo %rRMSE para todas las estaciones de medida y todos los horizontes temporales de predicción $h = 1..6h$ . El modelo AR utilizado corresponde al óptimo señalado por el BIC, mientras que el modelo ARMA corresponde a un orden $p = 2$ y $q = 1$ . . . . .	121
5.14. Resultado del ARD utilizando el número óptimo de neuronas para cada horizonte temporal en la estación de C0-Pozo Izquierdo. Cada entrada representa los datos terrestres pasados de la serie temporal . . . . .	124
5.15. Resultado del ARD utilizando el número óptimo de neuronas para cada horizonte temporal en la estación de C5-Sta. Brígida. Cada entrada representa los datos terrestres pasados de la serie temporal . . . . .	125
5.16. Resultado del Log of evidence para decidir el número de neuronas óptimo en la capa oculta para la estación de C0-Pozo Izquierdo . . . . .	127
5.17. Evolución del %rRMSE en función del tiempo horizonte $h = 1..6$ para cada modelo de predicción . . . . .	136
5.18. Evolución del %rMAE en función del tiempo horizonte $h = 1..6$ para cada modelo de predicción . . . . .	137
5.19. Predicción de la radiación diaria IGH con diferentes modelos en la estación C5 para un horizonte de $h = 1$ en un día tipo AIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF . . .	147
5.20. Predicción de la radiación diaria IGH con diferentes modelos en la estación C5 para un horizonte de $h = 6$ en un día tipo AIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF . . .	148
5.21. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de $h = 1$ en un día tipo BIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF . . .	149
5.22. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de $h = 6$ en un día tipo BIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF . . .	150

5.23. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo CIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF . . . 151

5.24. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo CIII. Modelos a) Smart-Persistence, ARMA y NN, b) Smart-Persistence, NN y NN+ECMWF . . . 152

5.25. Selección de los Superpixel (3x3) en la estación C1-Las Palmas utilizando el Test-1 para elegir el número en cada imagen desfasada  $t = 0, 1, 2$  & 3 horas. El área negra muestra los Superpíxeles seleccionados. Cada píxel representa la intercorrelación de  $k^*$  entre los datos terrestres y satelitales . . 155

5.26. Altitud a la que se extrajeron los datos de velocidad del viento y humedad relativa según cada criterio de selección en las estaciones de C0 y C1 . . . . 157

5.27. Evolución del %rRMSE en función del tiempo horizonte  $h = 1..6$  para los modelos de predicción basados en datos satélites y datos de radiación del ECMWF . . . . . 163

5.28. Evolución del %rRMSE en función del tiempo horizonte  $h = 1..6$  para los modelos de predicción basados en datos de radiación y datos de viento del ECMWF . . . . . 163

5.29. Evolución del %rRMSE en función del tiempo horizonte  $h = 1..6$  para los modelos de predicción basados en datos satélites, datos de radiación y datos de viento del ECMWF . . . . . 164

5.30. Evolución del %rRMSE en función del tiempo horizonte  $h = 1..6$  para el modelo basado en datos de radiación satélite y del ECMWF por cada trimestre 166

5.31. Evolución del %SKILL en función del tiempo horizonte  $h = 1..6$  para el modelo basado en datos de radiación satélite y del ECMWF por cada trimestre 167

5.32. Evolución del %SKILL en función del tiempo horizonte  $h = 1..6$  para el modelo basado en datos de radiación satélite y del ECMWF según los distintos tipos de días . . . . . 173

5.33. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo BIII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT . . . . . 174

5.34. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo BIII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND y NN+ECMWF+WIND+SAT . . . . . 175

5.35. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo BIII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT . . . . . 176

- 5.36. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo BIII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND y NN+ECMWF+WIND+SAT . . . . . 177
- 5.37. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo BII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT . . . . . 178
- 5.38. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo BII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND y NN+ECMWF+WIND+SAT . . . . . 179
- 5.39. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo BII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT . . . . . 180
- 5.40. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo BII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND y NN+ECMWF+WIND+SAT . . . . . 181
- 5.41. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo CII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT . . . . . 182
- 5.42. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 1$  en un día tipo CII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND y NN+ECMWF+WIND+SAT . . . . . 183
- 5.43. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo CII con el año 2005. Modelos a) Smart-Persistence, NN y NN+ECMWF y NN+SAT, b) Smart-Persistence, NN+SAT y NN+ECMWF+SAT . . . . . 184
- 5.44. Predicción de la radiación diaria IGH con diferentes modelos en la estación C0 para un horizonte de  $h = 6$  en un día tipo CII con el año 2005. Modelos a) Smart-Persistence, NN+ECMWF y NN+ECMWF+WIND, b) Smart-Persistence, NN+ECMWF+SAT, NN+ECMWF+WIND y NN+ECMWF+WIND+SAT . . . . . 185

5.45. Comparación de la predicción de la energía eléctrica producida por una planta fotovoltaica obtenida a partir de la radiación estimada por el modelo NN+EMWF+SAT y medida en la estación de C0-Pozo Izquierdo. . . . . 187



# Índice de tablas

2.1. Variables meteorológicas del modelo de cielo despejado y resultados de su estimación. . . . .	26
2.2. Estimación de la precisión del modelo de cielo despejado de Bird. . . . .	27
2.3. Coordenadas geográficas de las estaciones de medida del ITC en la isla de Gran Canaria. . . . .	29
2.4. Porcentaje de datos correctos disponibles en los años elegidos para la estación de C0-Pozo Izquierdo. . . . .	30
2.5. Porcentaje de datos correctos disponibles en los años elegidos para la estación de C1-Las Palmas. . . . .	31
2.6. Porcentaje de datos correctos disponibles en los años elegidos para la estación de C2-La Aldea. . . . .	31
2.7. Porcentaje de datos correctos disponibles en los años elegidos para la estación de C4-Maspalomas. . . . .	31
2.8. Porcentaje de datos correctos disponibles en los años elegidos para la estación de C5-Sta. Brígida. . . . .	31
2.9. Porcentaje de datos correctos disponibles en los años elegidos para la estación de C6-Mogán. . . . .	31
2.10. Conjunto de datos utilizados según la estación de medida. . . . .	32
2.11. Resultados del número de datos filtrados. . . . .	35
2.12. Condiciones a cumplir por cada uno de los subconjuntos de datos. . . . .	37
2.13. Distribución porcentual del número de días de cada tipo para la estación de C0-Pozo Izquierdo. . . . .	39
2.14. Distribución porcentual del número de días de cada tipo para la estación de C1-Las Palmas. . . . .	42
2.15. Distribución porcentual del número de días de cada tipo para la estación de C2-La Aldea. . . . .	42
2.16. Distribución porcentual del número de días de cada tipo para la estación de C4-Maspalomas. . . . .	42

2.17. Distribución porcentual del número de días de cada tipo para la estación de C5-Sta. Brígida. . . . .	42
2.18. Distribución porcentual del número de días de cada tipo para la estación de C6-Mogán. . . . .	43
5.1. Órdenes del modelo AR elegido para cada estación y horizonte temporal $h$ .	120
5.2. Errores del conjunto de datos de radiación terrestres de entrenamiento y de test utilizando distintos números de neuronas ocultas en la estación de C0-Pozo Izquierdo para el horizonte temporal 1 h. . . . .	126
5.3. Errores del conjunto de datos de radiación terrestres de entrenamiento y de test utilizando distintos números de neuronas ocultas en la estación de C0-Pozo Izquierdo para el horizonte temporal 3 h. . . . .	128
5.4. Errores RMSE para todos los horizontes temporales $h = 1..6$ en todas las estaciones de medida. . . . .	132
5.5. Errores %rRMSE para todos los horizontes temporales $h = 1..6$ en todas las estaciones de medida. . . . .	133
5.6. Comparación con el modelo Persistence mediante el parámetro %SKILL para todos los horizontes temporales $h = 1..6$ en todas las estaciones de medida. . . . .	134
5.7. Errores %rRMSE anuales y por trimestres para el horizonte temporal $h = 1$ en todas las estaciones de medida. . . . .	138
5.8. Errores %rRMSE anuales y por trimestres para el horizonte temporal $h = 6$ en todas las estaciones de medida. . . . .	139
5.9. Condiciones a cumplir por cada uno de los subconjuntos de datos. . . . .	141
5.10. Errores %rRMSE en los distintos tipos de días para todos los horizontes temporales en la estación C0-Pozo Izquierdo . . . . .	142
5.11. Errores %rRMSE en los distintos tipos de días para todos los horizontes temporales en la estación C1-Las Palmas . . . . .	143
5.12. Errores %rRMSE en los distintos tipos de días para todos los horizontes temporales en la estación C2-La Aldea . . . . .	144
5.13. Errores %rRMSE en los distintos tipos de días para todos los horizontes temporales en la estación C5-Sta. Brígida . . . . .	145
5.14. Errores RMSE para todos los horizontes temporales $h = 1..6$ en las estaciones de medida C0 y C1 para los datos de 2005 . . . . .	160
5.15. Errores %rRMSE para todos los horizontes temporales $h = 1..6$ en las estaciones de medida C0 y C1 para los datos de 2005 . . . . .	161
5.16. Errores %SKILL para todos los horizontes temporales $h = 1..6$ en las estaciones de medida C0 y C1 para los datos de 2005 . . . . .	161
5.17. Errores %rRMSE anuales y por trimestres para el horizonte temporal $h = 1$ las estaciones C0 y C1 para el año 2005 . . . . .	165

5.18. Errores %rRMSE anuales y por trimestres para el horizonte temporal $h = 6$ las estaciones C0 y C1 para el año 2005 . . . . .	165
5.19. Errores %rRMSE para los días tipo AIII, BIII y CIII para todos los horizontes temporales $h = 1...6$ en la estación C0 con los datos de 2005 . . . . .	169
5.20. Errores %rRMSE para los días tipo BII, CII y CI para todos los horizontes temporales $h = 1...6$ en la estación C0 con los datos de 2005 . . . . .	170
5.21. Errores %rRMSE para los distintos tipos de días para todos los horizontes temporales $h = 1...6$ en la estación C1 con los datos de 2005 . . . . .	171



# Bibliografía

- [AA98] S. Al-Alawi, H. Al-Hinai. An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation. *Renewable Energy*, vol. 14, num. 1, pags. 199–204, 1998.
- [AERONET14] AERONET. AERONET AEROSOL ROBOTIC NETWORK. url<http://aeronet.gsfc.nasa.gov/>, 2014. Accedido 2015.
- [Almonacid13] F. Almonacid, P. Pérez-Higueras, P. Rodrigo, L. Hontoria. Generation of ambient temperature hourly time series for some Spanish locations by artificial neural networks. *Renewable Energy*, vol. 51, pags. 285–291, 2013.
- [Bacher09] P. Bacher, H. Madsen, H. A. Nielsen. Online short-term solar power forecasting. *Solar Energy*, vol. 83, num. 10, pags. 1772–1783, 2009.
- [Badescu13] V. Badescu, C. A. Gueymard, S. Cheval, C. Oprea, M. Baciú, A. Dumitrescu, F. Iacobescu, I. Milos, C. Rada. Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania. *Renewable Energy*, vol. 55, pags. 85–103, 2013.
- [Badosa15] J. Badosa, M. Haefelin, N. Kalecinski, F. Bonnardot, G. Jumaux. Reliability of day-ahead solar irradiance forecasts on Reunion Island depending on synoptic wind and humidity conditions. *Solar Energy*, vol. 115, pags. 306–321, 2015.
- [Becquerel39] A.-E. Becquerel. Mémoire sur les effets électriques produits sous l’influence des rayons solaires. *Comptes Rendus*, vol. 9, num. 567, pags. 1839, 1839.
- [Beyer09] H. G. Beyer, J. Polo Martinez, M. Suri, J. L. Torres, E. Lorenz, S. C. Müller, C. Hoyer-Klick, P. Ineichen. Report on Benchmarking of Radiation Products. *MESOR-038665 Management and Exploitation of Solar Resource Knowledge*, , , 2009.
- [Bird81] R. E. Bird, R. L. Hulstrom. Simplified clear sky model for direct and diffuse insolation on horizontal surfaces. Technical report, Solar Energy Research Inst., Golden, CO (USA), 1981.

- [Bishop95] C. M. Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.
- [Blanc11a] P. Blanc. Atlas du Potentiel Solaire Photovoltaïque et Thermodynamique en region PACA. Technical report, Centre Energetique et Procèdes, MINES ParisTech/ARMINES, 2011.
- [Blanc11b] P. Blanc, B. Gschwind, M. Lefèvre, L. Wald. The HelioClim project: Surface solar irradiance data for climate applications. *Remote Sensing*, vol. 3, num. 2, pags. 343–361, 2011.
- [BM01] M. Blanco-Muriel, D. C. Alarcón-Padilla, T. López-Moratalla, M. Lara-Coira. Computing the solar vector. *Solar Energy*, vol. 70, num. 5, pags. 431–441, 2001.
- [Boland95] J. Boland. Time-series analysis of climatic variables. *Solar Energy*, vol. 55, num. 5, pags. 377–388, 1995.
- [Boland08] J. Boland. *Time series modelling of solar radiation*. Springer, 2008.
- [Bosch08] J. Bosch, G. Lopez, F. Batlles. Daily solar irradiation estimation over a mountainous area using artificial neural networks. *Renewable Energy*, vol. 33, num. 7, pags. 1622–1628, 2008.
- [Box98] G. Box, G. Jenkins. *Time series analysis, forecasting and control*. Wiley, 1998.
- [BS10] J. Bosch Saldaña. *Modelización del recurso solar utilizando redes neuronales artificiales y su aplicación a la generación de mapas topográficos de radiación*. Tesis Doctoral, Universidad de Almería, 2010.
- [Canada’s15] E. Canada’s. World Ozone Monitoring Mapping. [urlhttp://es-ee.tor.ec.gc.ca/e/ozone/ozoneworld.htm/](http://es-ee.tor.ec.gc.ca/e/ozone/ozoneworld.htm/), 2015. Accedido 2015.
- [Cano86] D. Cano, J.-M. Monget, M. Albuissou, H. Guillard, N. Regas, L. Wald. A method for the determination of the global solar radiation from meteorological satellite data. *Solar Energy*, vol. 37, num. 1, pags. 31–39, 1986.
- [Cao08] J. Cao, X. Lin. Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks. *Energy Conversion and Management*, vol. 49, num. 6, pags. 1396–1406, 2008.
- [Chapin54] D. M. Chapin, C. Fuller, G. Pearson. A new silicon p-n junction photocell for converting solar radiation into electrical power. *Journal of Applied Physics*, , num. 25, pags. 676–677, 1954.
- [Chatfield13] C. Chatfield. *The analysis of time series: an introduction*. CRC press, 2013.

- [Chow11] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields, B. Washom. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy*, vol. 85, num. 11, pags. 2881–2893, 2011.
- [Coimbra13] C. Coimbra, J. Kleissl, R. Marquez. Overview of solar forecasting methods and a metric for accuracy evaluation. *Solar Resource Assessment and Forecasting*, edited by: Kleissl, J., Elsevier, Waltham, Massachusetts, , pags. 171–194, 2013.
- [CP89] M. Collares-Pereira, R. Aguiar. The modeling of daily sequences of hourly radiation through autoregressive methods. *Clean and Safe Energy Forever. Proceedings of the 1989 Congress of the International Solar Energy Society*. ISES, 1989.
- [DA96] P. J. Dorta Antequera. Las inversiones térmicas en Canarias. *Investigaciones geográficas*, , num. 15, pags. 109–126, 1996.
- [Dambreville14a] R. Dambreville. *Prévision du rayonnement solaire global par télédétection pour la gestion de la production d'énergie photovoltaïque*. Tesis Doctoral, Grenoble, 2014.
- [Dambreville14b] R. Dambreville, P. Blanc, J. Chanussot, D. Boldo. Very short term forecasting of the Global Horizontal Irradiance using a spatio-temporal autoregressive model. *Renewable Energy*, vol. 72, pags. 291–300, 2014.
- [David14] M. David, P. Lauret, M. Diagne, T. P.J. Day ahead and intra day forecasting applied to an insular site. *Colloque Martinique Energie Environnement*. CM2E, 2014.
- [Diagne13] M. Diagne, M. David, P. Lauret, J. Boland, N. Schmutz. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, vol. 27, pags. 65–76, 2013.
- [Diagne14] M. Diagne, M. David, J. Boland, N. Schmutz, P. Lauret. Post-processing of solar irradiance forecasts from WRF model at Reunion Island. *Solar Energy*, vol. 105, pags. 99–108, 2014.
- [Díaz12] F. Díaz, G. Montero, J. M. Escobar, E. Rodríguez, R. Montenegro. An adaptive solar radiation numerical model. *Journal of computational and applied mathematics*, vol. 236, num. 18, pags. 4611–4622, 2012.
- [DR13] F. Díaz Reyes. *Modelo numérico para la generación de mapas de radiación solar y su aplicación al aprovechamiento de energía solar fotovoltaica y termoeléctrica*. Tesis Doctoral, Universidad de Las Palmas de Gran Canaria, 2013.
- [Einstein05] A. Einstein. The photoelectric effect. *Ann. Phys*, vol. 17, num. 132, pags. 4, 1905.

- [Eissa15] Y. Eissa, M. Korany, Y. Aoun, M. Boraiy, M. M. Abdel Wahab, S. C. Alfaro, P. Blanc, M. El-Metwally, H. Ghedira, K. Hungershofer et al. Validation of the surface downwelling solar irradiance estimates of the HelioClim-3 database in Egypt. *Remote Sensing*, vol. 7, num. 7, pags. 9269–9291, 2015.
- [Espinar11] B. Espinar, L. Wald, P. Blanc, C. Hoyer-Klick, M. Schroedter-Homscheidt, T. Wanderer. Report on the Harmonization and Qualification of Meteorological Data: Project ENDORSE-Energy Downstream Service: Providing Energy Components for GMES-Grant Agreement No. 262892. *Paris, France: Armines. Accessed November*, vol. 11, pags. 2014, 2011.
- [FJ13] J. G. d. S. Fonseca Jr, O. Takashi, O. Hideaki, S. Ken-ichi, T. Takumi, O. Kazuhiko. Analysis of different techniques to set support vector regression to forecast insolation in Tsukuba, Japan. *Journal of International Council on Electrical Engineering*, vol. 3, num. 2, pags. 121–128, 2013.
- [Fukushima83] K. Fukushima, S. Miyake, T. Ito. Neocognitron: A neural network model for a mechanism of visual pattern recognition. *Systems, Man and Cybernetics, IEEE Transactions on*, , num. 5, pags. 826–834, 1983.
- [Ghanbarzadeh09] A. Ghanbarzadeh, A. Noghrehabadi, E. Assareh, M. Behrang. Solar radiation forecasting based on meteorological data using artificial neural networks. *Industrial Informatics, 2009. INDIN 2009. 7th IEEE International Conference on*, pags. 227–231. IEEE, 2009.
- [González95] J. R. H. González, V. J. M. Hernando. *Redes neuronales artificiales: fundamentos, modelos y aplicaciones*. Ra-ma, 1995.
- [Hamilton94] J. D. Hamilton. *Time series analysis*, volumen 2. Princeton university press Princeton, 1994.
- [Hammer99] A. Hammer, D. Heinemann, E. Lorenz, B. Lückehe. Short-term forecasting of solar radiation: a statistical approach using satellite data. *Solar Energy*, vol. 67, num. 1, pags. 139–150, 1999.
- [Haykin96] S. Haykin. Adaptive filter theory. part xv. matched exponents for the representation of signals, 1996.
- [Heinemann06a] D. Heinemann, E. Lorenz, M. Girodo. Forecasting of solar radiation. *Solar energy resource management for electricity generation from local level to global scale*. Nova Science Publishers, New York, , , 2006.
- [Heinemann06b] D. Heinemann, E. Lorenz, M. Girodo. Forecasting of solar radiation. *Solar energy resource management for electricity generation from local level to global scale*. Nova Science Publishers, New York, , , 2006.

- [Hoff12] T. E. Hoff, R. Perez. Modeling PV fleet output variability. *Solar Energy*, vol. 86, num. 8, pags. 2177–2189, 2012.
- [Hofierka02] J. Hofierka, M. Suri et al. The solar radiation model for Open source GIS: implementation and applications. *Proceedings of the Open source GIS-GRASS users conference*, pags. 1–19, 2002.
- [Holben98] B. Holben, T. Eck, I. Slutsker, D. Tanre, J. Buis, A. Setzer, E. Vermote, J. Reagan, Y. Kaufman, T. Nakajima et al. AERONET-A federated instrument network and data archive for aerosol characterization. *Remote sensing of environment*, vol. 66, num. 1, pags. 1–16, 1998.
- [Hontoria01] L. Hontoria, J. Aguilera, J. Riesco, P. Zufiria. Recurrent neural supervised models for generating solar radiation synthetic series. *Journal of Intelligent and Robotic Systems*, vol. 31, num. 1-3, pags. 201–221, 2001.
- [Hontoria02] L. Hontoria, J. Aguilera, P. Zufiria. Generation of hourly irradiation synthetic series using the neural network multilayer perceptron. *Solar Energy*, vol. 72, num. 5, pags. 441–446, 2002.
- [Hopfield82] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, vol. 79, num. 8, pags. 2554–2558, 1982.
- [Hornik89] K. Hornik, M. Stinchcombe, H. White. Multilayer feedforward networks are universal approximators. *Neural networks*, vol. 2, num. 5, pags. 359–366, 1989.
- [Ineichen06] P. Ineichen. Comparison of eight clear sky broadband models against 16 independent data banks. *Solar Energy*, vol. 80, num. 4, pags. 468–478, 2006.
- [Inman13] R. H. Inman, H. T. Pedro, C. F. Coimbra. Solar forecasting methods for renewable energy integration. *Progress in energy and combustion science*, vol. 39, num. 6, pags. 535–576, 2013.
- [Isasi04] P. Isasi, I. Galván. Redes neuronales artificiales—un enfoque práctico, 2004.
- [Kemmoku99] Y. Kemmoku, S. Orita, S. Nakagawa, T. Sakakibara. Daily insolation forecasting using a multi-stage neural network. *Solar Energy*, vol. 66, num. 3, pags. 193–199, 1999.
- [Kohonen88] T. Kohonen. An introduction to neural computing. *Neural networks*, vol. 1, num. 1, pags. 3–16, 1988.
- [Kostylev11] V. Kostylev, A. Pavlovski et al. Solar power forecasting performance—towards industry standards. *1st International Workshop on the Integration of Solar Power into Power Systems Aarhus, Denmark*, 2011.

- [Lauret06a] P. Lauret, M. David, E. Fock, A. Bastide, C. Riviere. Bayesian and sensitivity analysis approaches to modeling the direct solar irradiance. *Journal of solar energy engineering*, vol. 128, num. 3, pags. 394–405, 2006.
- [Lauret06b] P. Lauret, E. Fock, T. A. Mara. A node pruning algorithm based on a Fourier amplitude sensitivity test method. *Neural Networks, IEEE Transactions on*, vol. 17, num. 2, pags. 273–293, 2006.
- [Lauret08] P. Lauret, E. Fock, R. N. Randrianarivony, J.-F. Manicom-Ramsamy. Bayesian neural network approach to short time load forecasting. *Energy conversion and management*, vol. 49, num. 5, pags. 1156–1166, 2008.
- [Lauret12] P. Lauret, A. Rodler, M. Muselli, M. David, H. Diagne, C. Voyant. A Bayesian model committee approach to forecasting global solar radiation. *arXiv preprint arXiv:1203.5446*, , , 2012.
- [Lauret15] P. Lauret, C. Voyant, T. Soubdhan, M. David, P. Poggi. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Solar Energy*, vol. 112, pags. 446–457, 2015.
- [Lave13] M. Lave, J. Kleissl. Cloud speed impact on solar variability scaling—Application to the wavelet variability model. *Solar Energy*, vol. 91, pags. 11–21, 2013.
- [Lebarbier04] E. Lebarbier, T. Mary-Huard. Le critère BIC: fondements théoriques et interprétation. Technical report, INRIA, 2004.
- [Lefevre13] M. Lefevre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques*, vol. 6, num. 9, pags. 2403–2418, 2013.
- [Lorenz09a] E. Lorenz, J. Hurka, D. Heinemann, H. G. Beyer. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 2, num. 1, pags. 2–10, 2009.
- [Lorenz09b] E. Lorenz, J. Remund, S. C. Müller, W. Traunmüller, G. Steinmaurer, D. Pozo, J. A. Ruiz-Arias, V. L. Fanego, L. Ramirez, M. G. Romeo et al. Benchmarking of different approaches to forecast solar irradiance. *24th European photovoltaic solar energy conference, Hamburg, Germany*, volumen 21, page 25, 2009.
- [Lorenz11] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, C. Kurz. Regional PV power prediction for improved grid integration. *Progress in*

- Photovoltaics: Research and Applications*, vol. 19, num. 7, pags. 757–771, 2011.
- [Lorenz12] E. Lorenz, D. Heinemann. Prediction of solar irradiance and photovoltaic power. *Comprehensive Renewable Energy*. Elsevier, Oxford, , pags. 239–292, 2012.
- [Lorenzo06] E. Lorenzo. Electricidad Solar Fotovoltaica, Volumen II, Radiación Solar y Dispositivos Fotovoltaicos. *Ed. Progensa. Sevilla*, , , 2006.
- [LR12] J. M. Lujano-Rojas, R. Dufo-López, J. L. Bernal-Agustín. Optimal sizing of small wind/battery systems considering the DC bus voltage stability effect on energy capture, wind speed variability, and load uncertainty. *Applied Energy*, vol. 93, pags. 404–412, 2012.
- [MA15] L. Mazonra Aguiar, B. Pereira, M. David, F. DÁaz, P. Lauret. Use of satellite data to improve solar radiation forecasting with Bayesian Artificial Neural Networks. *Solar Energy*, vol. Accepted, , 2015.
- [MacKay92] D. J. MacKay. A practical Bayesian framework for backpropagation networks. *Neural computation*, vol. 4, num. 3, pags. 448–472, 1992.
- [MacKay03] D. J. MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.
- [Marquez13] R. Marquez, H. T. Pedro, C. F. Coimbra. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. *Solar Energy*, vol. 92, pags. 176–188, 2013.
- [Maxwell93] E. Maxwell, S. Wilcox, M. Rymes. Users manual for seri qc software, assessing the quality of solar radiation data. *Solar Energy Research Institute, Golden, CO (available at <http://www.osti.gov/bridge>)*, , , 1993.
- [Mazonra10] L. Mazonra, F. Díaz, G. Montero, R. Montenegro. Typical meteorological year (TMY) evaluation for power generation in Gran Canaria Island, Spain. *Proceedings of 25th European Photovoltaic Solar Energy Conference and Exhibition, Valencia, Spain*, pags. 4726–4728, 2010.
- [McCulloch43] W. S. McCulloch, W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, vol. 5, num. 4, pags. 115–133, 1943.
- [Mellit08] A. Mellit, S. A. Kalogirou. Artificial intelligence techniques for photovoltaic applications: A review. *Progress in energy and combustion science*, vol. 34, num. 5, pags. 574–632, 2008.
- [Mellit10] A. Mellit, A. M. Pavan. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a

- grid-connected PV plant at Trieste, Italy. *Solar Energy*, vol. 84, num. 5, pags. 807–821, 2010.
- [Minsky88] M. Minsky, S. Papert. *Perceptrons: an introduction to computational geometry* (expanded edition), 1988.
- [Mohandes98] M. Mohandes, S. Rehman, T. Halawani. Estimation of global solar radiation using artificial neural networks. *Renewable Energy*, vol. 14, num. 1, pags. 179–184, 1998.
- [Moradi09] I. Moradi. Quality control of global solar radiation using sunshine duration hours. *Energy*, vol. 34, num. 1, pags. 1–6, 2009.
- [Nabney02] I. Nabney. *NETLAB: algorithms for pattern recognition*. Springer Science & Business Media, 2002.
- [Page86] J. Page. *Prediction of Solar Radiation on Inclined Surfaces* (D Reidel, Dordrecht). , , 1986.
- [Penny99] W. D. Penny, S. J. Roberts. Bayesian neural networks for classification: how useful is the evidence framework? *Neural Networks*, vol. 12, num. 6, pags. 877–892, 1999.
- [Perez07] R. Perez, K. Moore, S. Wilcox, D. Renné, A. Zelenka. Forecasting solar radiation—Preliminary evaluation of an approach based upon the national forecast database. *Solar Energy*, vol. 81, num. 6, pags. 809–812, 2007.
- [Perez10] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker, D. Renné, T. E. Hoff. Validation of short and medium term operational solar radiation forecasts in the US. *Solar Energy*, vol. 84, num. 12, pags. 2161–2172, 2010.
- [Perez11] R. Perez, M. Beauharnois, K. Hemker, S. Kivalov, E. Lorenz, S. Pelland, J. Schlemmer, G. Van Knowe. Evaluation of numerical weather prediction solar irradiance forecasts in the US. *ASES Annual Conference. Raleigh, NC, USA*, volumen 1721, 2011.
- [Perez13] R. Perez, E. Lorenz, S. Pelland, M. Beauharnois, G. Van Knowe, K. Hemker, D. Heinemann, J. Remund, S. C. Müller, W. Traummüller et al. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. *Solar Energy*, vol. 94, pags. 305–326, 2013.
- [RA10] J. A. Ruiz-Arias, T. Cebecauer, J. Tovar-Pescador, M. Šúri. Spatial disaggregation of satellite-derived irradiance using a high-resolution digital elevation model. *Solar Energy*, vol. 84, num. 9, pags. 1644–1657, 2010.

- [Rehman08] S. Rehman, M. Mohandes. Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy*, vol. 36, num. 2, pags. 571–576, 2008.
- [Reikard09] G. Reikard. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, vol. 83, num. 3, pags. 342–349, 2009.
- [Reno12] M. J. Reno, C. W. Hansen, J. S. Stein. Global horizontal irradiance clear sky models: Implementation and analysis. *SANDIA report SAND2012-2389*, , , 2012.
- [Rigollier04] C. Rigollier, M. Lefèvre, L. Wald. The method Heliosat-2 for deriving shortwave solar radiation from satellite images. *Solar Energy*, vol. 77, num. 2, pags. 159–169, 2004.
- [Roesch11] A. Roesch, M. Wild, A. Ohmura, E. G. Dutton, C. N. Long, T. Zhang. Assessment of BSRN radiation records for the computation of monthly means. *Atmospheric Measurement Techniques*, vol. 4, num. 2, pags. 339–354, 2011.
- [Rosenblatt58] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, vol. 65, num. 6, pags. 386, 1958.
- [Rumelhart86] D. Rumelhart, G. Hinton, R. Williams. Learning Internal Representations by Error Propagation, Parallel Distributed Processing, Explorations in the Microstructure of Cognition, ed. DE Rumelhart and J. McClelland. Vol. 1. 1986, 1986.
- [Safi02] S. Safi, A. Zeroual, M. Hassani. Prediction of global daily solar radiation using higher order statistics. *Renewable Energy*, vol. 27, num. 4, pags. 647–666, 2002.
- [Sengupta15] M. Sengupta, A. Habte, S. Kurtz, A. Dobos, S. Wilbert, E. Lorenz, T. Stoffel, D. Renné, C. Gueymard, D. Myers et al. Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications. , , 2015.
- [Sfetsos00] A. Sfetsos, A. Coonick. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy*, vol. 68, num. 2, pags. 169–178, 2000.
- [Sun14] A. Y. Sun, D. Wang, X. Xu. Monthly streamflow forecasting using Gaussian process regression. *Journal of Hydrology*, vol. 511, pags. 72–81, 2014.
- [Šúri04] M. Šúri, J. Hofierka. A new GIS-based solar radiation model and its application to photovoltaic assessments. *Transactions in GIS*, vol. 8, num. 2, pags. 175–190, 2004.

- [Thomas15] C. Thomas, L. Saboret, E. Wey, P. Blanc, L. Wald. Preliminary assessment of a new SoDa service for real-time estimates and short-term forecasts of the solar radiation. *15th EMS Annual Meeting/12th ECAM*, page 335, 2015.
- [Urquhart13] B. Urquhart, M. Ghonima, D. Nguyen, B. Kurtz, C. Chow, J. Kleissl. Sky imaging systems for short-term forecasting. *Kleissl, J., Elsevier, Waltham, Massachusetts, , , 2013*.
- [Widrow60] B. Widrow, M. E. Hoff. Adaptive switching circuits. In 1960 WESCON Convention Record Part IV, pages 96–104. Reprinted in JA Anderson and E. Rosenfeld. *Neurocomputing: Foundations of Research, , , 1960*.
- [Willmott05] C. J. Willmott, K. Matsuura. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, vol. 30, num. 1, pags. 79, 2005.
- [Wittmann08] M. Wittmann, H. Breitzkreuz, M. Schroedter-Homscheidt, M. Eck. Case studies on the use of solar irradiance forecast for optimized operation strategies of solar thermal power plants. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 1, num. 1, pags. 18–27, 2008.
- [Wolff13] B. Wolff, E. Lorenz, O. Kramer. Statistical learning for short-term photovoltaic power predictions. *Proceedings of DARE 2013 Workshop on Data Analytics for Renewable Energy Integration, September, volumen 23, 2013*.
- [Younes05] S. Younes, R. Claywell, T. Muneer. Quality control of solar radiation data: present status and proposed new approaches. *Energy*, vol. 30, num. 9, pags. 1533–1549, 2005.
- [Younes07] S. Younes, T. Muneer. Clear-sky classification procedures and models using a world-wide data-base. *Applied energy*, vol. 84, num. 6, pags. 623–645, 2007.
- [Zagouras15] A. Zagouras, H. T. Pedro, C. F. Coimbra. On the role of lagged exogenous variables and spatio-temporal correlations in improving the accuracy of solar forecasting methods. *Renewable Energy*, vol. 78, pags. 203–218, 2015.
- [Zeng13] J. Zeng, W. Qiao. Short-term solar power prediction using a support vector machine. *Renewable Energy*, vol. 52, pags. 118–127, 2013.