

Universidad Politécnica de Madrid  
E. T. S. de Ingenieros en Topografía, Geodesia y Cartografía  
Departamento de Ingeniería Topográfica y Cartografía



**Predicción espacio-temporal de la irradiancia solar  
global a corto plazo en España mediante  
geoestadística y redes neuronales artificiales**

**TESIS DOCTORAL**

**Federico Vladimir Gutiérrez Corea**

Directores:

Dr. Miguel Ángel Manso Callejo  
Departamento de Ingeniería Topográfica y Cartografía  
ETSI en Topografía, Geodesia y Cartografía, UPM

Dr. Francisco Serradilla García  
Departamento de Inteligencia Artificial  
ETSI de Sistemas Informáticos, UPM

Madrid, España, Diciembre de 2014





El Tribunal de la Tesis Doctoral, nombrado por el Magnífico y Excelentísimo Sr. Rector de la Universidad Politécnica de Madrid, con fecha

integrado por:

Presidente:

Secretario:

Vocal 1º:

Vocal 2º:

Vocal 3º:

realizado el acto de defensa y lectura de tesis el día

decide otorgar a la tesis doctoral presentada la calificación de:

EL PRESIDENTE

EL SECRETARIO

EL VOCAL 1º

EL VOCAL 2º

EL VOCAL 3º

*A Elia, Vlady, mi padre,  
familia y amigos*

## **AGRADECIMIENTOS**

La culminación la presente tesis doctoral no hubiese sido posible sin la ayuda de muchas personas.

Deseo agradecer a mis directores de tesis: Dr. Miguel Ángel Manso Callejo por guiarme en todos los aspectos del largo trayecto que conlleva la formación de capacidades científicas, su gran dedicación en las revisiones, sugerencias y comprensión y al Dr. Francisco Serradilla García por orientarme en el camino de la investigación mediante Redes Neuronales Artificiales y trasladarme su entusiasmo y conocimiento en esta temática.

Deseo agradecer a todos los compañeros del grupo de investigación MERCATOR -Tecnologías de la GeoInformación- de la Universidad Politécnica de Madrid (UPM) desde donde se realizaron acuerdos de investigación junto a Red Eléctrica de España (REE) que han permitido financiar esta tesis. Agradezco especialmente al Dr. Miguel Ángel Bernabé Poveda quien me introdujo al grupo de investigación MERCATOR/UPM.

Agradezco a Elia su completo apoyo y amor, a mis padres por enseñarme desde edades tempranas la importancia de los estudios, a mi abuelita por haber estado siempre pendiente y a toda mi familia por el apoyo, en especial a mi hermano Julio. Agradezco a los amigos hechos en Madrid por haber compartido con migo y mi familia este tiempo en España, así como a mis amigos de Nicaragua por el constante interés tanto en mi familia como en mis avances.

Agradezco a los dirigentes de INETER y del gobierno de Nicaragua por otorgarme el permiso y apoyo para cursar los estudios de doctorado, a mis compañeros de INETER por los ánimos y por introducirme en las ciencias geoespaciales.



## RESUMEN

El enriquecimiento del conocimiento sobre la Irradiancia Solar (IS) a nivel de superficie terrestre, así como su predicción, cobran gran interés para las Energías Renovables (ER) - Energía Solar (ES)-, y para distintas aplicaciones industriales o ecológicas. En el ámbito de las ER, el uso óptimo de la ES implica contar con datos de la IS en superficie que ayuden tanto, en la selección de emplazamientos para instalaciones de ES, como en su etapa de diseño (dimensionar la producción) y, finalmente, en su explotación. En este último caso, la observación y la predicción es útil para el mercado energético, la planificación y gestión de la energía (generadoras y operadoras del sistema eléctrico), especialmente en los nuevos contextos de las redes inteligentes de transporte. A pesar de la importancia estratégica de contar con datos de la IS, especialmente los observados por sensores de IS en superficie (los que mejor captan esta variable), estos no siempre están disponibles para los lugares de interés ni con la resolución espacial y temporal deseada. Esta limitación se une a la necesidad de disponer de predicciones a corto plazo de la IS que ayuden a la planificación y gestión de la energía.

Se ha indagado y caracterizado las Redes de Estaciones Meteorológicas (REM) existentes en España que publican en internet sus observaciones, focalizando en la IS. Se han identificado 24 REM (16 gubernamentales y 8 redes voluntarios) que aglutinan 3492 estaciones, convirtiéndose éstas en las fuentes de datos meteorológicos utilizados en la tesis. Se han investigado cinco técnicas de estimación espacial de la IS en intervalos de 15 minutos para el territorio peninsular (3 técnicas geoestadísticas, una determinística y el método HelioSat2 basado en imágenes satelitales) con distintas configuraciones espaciales. Cuando el área de estudio tiene una adecuada densidad de observaciones, el mejor método identificado para estimar la IS es el Kriging con Regresión usando variables auxiliares -una de ellas la IS estimada a partir de imágenes satelitales-. De este modo es posible estimar espacialmente la IS más allá de los 25 km identificados en la bibliografía. En caso contrario, se corrobora la idoneidad de utilizar estimaciones a partir de sensores remotos cuando la densidad de observaciones no es adecuada.

Se ha experimentado con el modelado de Redes Neuronales Artificiales (RNA) para la predicción a corto plazo de la IS utilizando observaciones próximas (componentes espaciales) en sus entradas y, los resultados son prometedores. Así los niveles de errores disminuyen bajo las siguientes condiciones: (1) cuando el horizonte temporal de predicción es inferior o igual a 3 horas, las estaciones vecinas que se incluyen en el modelo deben encontrarse a una distancia máxima aproximada de 55 km. Esto permite concluir que las RNA son capaces de aprender cómo afectan las condiciones meteorológicas vecinas a la predicción de la IS.

## **ABSTRACT**

The enrichment of knowledge about the Solar Irradiance (SI) at Earth's surface and its prediction, have a high interest for Renewable Energy (RE) - Solar Energy (SE) - and for various industrial and environmental applications. In the field of the RE, the optimal use of the SE involves having SI surface to help in the selection of sites for facilities ES, in the design stage (sizing energy production), and finally on their production. In the latter case, the observation and prediction is useful for the market, planning and management of the energy (generators and electrical system operators), especially in new contexts of smart transport networks (smartgrid). Despite the strategic importance of SI data, especially those observed by sensors of SI at surface (the ones that best measure this environmental variable), these are not always available to the sights and the spatial and temporal resolution desired. This limitation is bound to the need for short-term predictions of the SI to help planning and energy management.

It has been investigated and characterized existing Networks of Weather Stations (NWS) in Spain that share its observations online, focusing on SI. 24 NWS have been identified (16 government and 8 volunteer networks) that implies 3492 stations, turning it into the sources of meteorological data used in the thesis. We have investigated five technical of spatial estimation of SI in 15 minutes to the mainland (3 geostatistical techniques and HelioSat2 a deterministic method based on satellite images) with different spatial configurations. When the study area has an adequate density of observations we identified the best method to estimate the SI is the regression kriging with auxiliary variables (one of them is the SI estimated from satellite images). Thus it is possible to spatially estimate the SI beyond the 25 km identified in the literature. Otherwise, when the density of observations is inadequate the appropriateness is using the estimates values from remote sensing.

It has been experimented with Artificial Neural Networks (ANN) modeling for predicting the short-term future of the SI using observations from neighbor's weather stations (spatial components) in their inputs, and the results are promising. The error levels decrease under the following conditions: (1) when the prediction

horizon is less or equal than 3 hours the best models are the ones that include data from the neighboring stations (at a maximum distance of 55 km). It is concluded that the ANN is able to learn how weather conditions affect neighboring prediction of IS at such Spatio-temporal horizons.

# ÍNDICE

<b>AGRADECIMIENTOS</b> .....	<b>iii</b>
<b>RESUMEN</b> .....	<b>vii</b>
<b>ABSTRACT</b> .....	<b>ix</b>
<b>ÍNDICE</b> .....	<b>xi</b>
<b>LISTADO DE FIGURAS</b> .....	<b>xiv</b>
<b>LISTADO DE TABLAS</b> .....	<b>xvi</b>
<b>LISTADO DE ACRÓNIMOS Y SIGLAS</b> .....	<b>xviii</b>
<b>1 INTRODUCCIÓN</b> .....	<b>1</b>
1.1 MOTIVACIÓN.....	3
1.2 OBJETIVOS DE INVESTIGACIÓN.....	4
1.3 OBJETIVOS ESPECÍFICOS.....	4
1.4 HIPÓTESIS.....	5
1.5 PREGUNTAS DE INVESTIGACIÓN.....	6
1.6 METODOLOGÍA.....	7
1.7 ESTRUCTURA DE LA TESIS.....	8
<b>2 ESTADO DE LA CUESTIÓN</b> .....	<b>11</b>
2.1 DISPONIBILIDAD DE DATOS METEOROLÓGICOS Y DE IRRADIANCIA SOLAR Y EN TIEMPO CUASI REAL POR ESTACIONES.....	11
2.2.1 <i>Evolución de los registros meteorológicos</i> .....	13
2.2.2 <i>Observaciones Meteorológicas de altruistas</i> .....	14
2.2 LA RADIACIÓN SOLAR.....	16
2.2.1. <i>Irradiancia Solar Extraterrestre</i> .....	17
2.2.2. <i>Componentes de la Irradiancia Solar</i> .....	18
2.3 ESTIMACIÓN ESPACIAL DE LA ISG INTRA-HORARIA.....	19
2.3.1. <i>Método - Inverse Distance Weighting (IDW)</i> .....	19
2.3.2. <i>Método - Ordinary Kriging (OK)</i> .....	20
2.3.3. <i>Método - Regression Kriging (RK)</i> .....	21
2.3.4. <i>Estimación mediante Sensores Remotos</i> .....	23
2.4 PREDICCIÓN DE LA ISG MEDIANTE REDES NEURONALES ARTIFICIALES.....	25
2.4.1. <i>La Neuron Artificial</i> .....	26
2.4.2. <i>Perceptrón Multicapa</i> .....	27
2.4.3. <i>Modelización de la IS a corto plazo con RNA</i> .....	29
2.5 RESUMEN.....	30

<b>3</b>	<b>EVALUACIÓN DE LA DISPONIBILIDAD DE DATOS METEOROLÓGICOS EN TIEMPO CUASI-REAL EN ESPAÑA.....</b>	<b>33</b>
3.1	METODOLOGÍA DE EVALUACIÓN DE LA DISPONIBILIDAD DE DATOS METEOROLÓGICOS.....	33
3.2	DISCUSIÓN DE LOS RESULTADOS .....	37
3.3	RESUMEN .....	56
<b>4</b>	<b>ESTIMACIÓN ESPACIAL DE LA ISG .....</b>	<b>59</b>
4.1	ÁREA DE ESTUDIO Y FUENTES DE DATOS .....	59
4.1.1	Área de estudio .....	59
4.1.2	Fuentes de Datos:.....	61
4.2	METODOLOGÍA .....	63
4.2.1	Colecta de datos y adaptación temporal.....	63
4.2.2	Preparación de variables explicativas .....	64
4.2.3	Agrupación de los datos fuente .....	68
4.2.4	Estimación espacial de la ISG a intervalos de 15 minutos .....	71
4.2.5	Primera aproximación a la validación de las estaciones de voluntarios.....	74
4.3	RESULTADOS Y DISCUSIÓN .....	75
4.3.1	Puntuación de los métodos de estimación.....	75
4.3.2	Análisis detallado de los métodos ganadores.....	80
4.3.3	Análisis de una primera aproximación a la validación de estaciones de .. voluntarios.....	89
4.4	RESUMEN .....	93
<b>5</b>	<b>PREDICCIÓN ESPACIO-TEMPORAL DE LA ISG.....</b>	<b>95</b>
5.1	SITIO DE ESTUDIO Y DATOS EXPERIMENTALES .....	95
5.1.1	Sitio de estudio .....	96
5.1.2	Datos experimentales .....	97
5.2	METODOLOGÍA .....	98
5.2.1	Preparación de los datos.....	98
5.2.2	Normalización y selección de datos de entrenamiento y de test: .....	100
5.2.3	Parámetros de entrenamiento de las RNA .....	100
5.2.4	Modelos de RNA y estadísticos para medir la calidad de las predicciones .....	101
5.3	RESULTADOS Y DISCUSIÓN.....	104
5.3.1	Parámetros para la selección de las arquitecturas de RNA.....	104
5.3.2	Predicciones de la ISG a corto plazo .....	108
5.4	RESUMEN .....	111

<b>6</b>	<b>CONCLUSIONES Y DEMOSTRACIÓN DE LA HIPÓTESIS .....</b>	<b>113</b>
6.1.	CONCLUSIONES.....	113
6.1.1.	<i>Respecto a la disponibilidad de datos meteorológicos en tiempo cuasi-... real en España.....</i>	<i>113</i>
6.1.2.	<i>Respecto a la estimación espacial de la IS: .....</i>	<i>116</i>
6.1.3.	<i>Respecto a la predicción espacio-temporal de la IS:.....</i>	<i>119</i>
6.2	APORTACIONES ORIGINALES.....	120
6.3	REVISIÓN DE LAS PREGUNTAS DE INVESTIGACIÓN E HIPÓTESIS .....	121
6.4	TRABAJOS FUTUROS .....	123
<b>7</b>	<b>BIBLIOGRAFÍA .....</b>	<b>125</b>
<b>8</b>	<b>ANEXOS.....</b>	<b>143</b>
8.1	LISTA DE REDES DE ESTACIONES METEOROLÓGICAS Y SUS URLS DE ACCESOS.	143
8.2	PUBLICACIONES CIENTÍFICAS GENERADAS EN EL PERIODO DE TESIS .....	147
8.2.1.	<i>Artículos en revistas indexadas por la ISI Web of Science .....</i>	<i>147</i>
8.2.2.	<i>Próximo a enviarse a revista indexada por la ISI Web of Science .....</i>	<i>147</i>
8.3.3.	<i>Capítulos de libros y artículos publicados en congresos.....</i>	<i>147</i>
8.3.4.	<i>Posters y presentaciones en congresos .....</i>	<i>149</i>

## LISTADO DE FIGURAS

<b>Figura 1.</b> Comparativa de tamaños a escala verdadera entre el Sol y la Tierra, .... sin representar la distancia entre ambos .....	1
<b>Figura 2.</b> Componentes de la radiación solar .....	19
<b>Figura 3.</b> Resolución espacial del satélite Meteosat.....	25
<b>Figura 4.</b> Representación de una neuronal biológica y una neurona artificial ...	27
<b>Figura 5.</b> RNA de tipo perceptrón multicapa.....	28
<b>Figura 6.</b> Ilustración simplificada de la arquitectura de sistemas para la ..... recolección y almacenamiento de observaciones meteorológicas desde . distintas fuentes.....	36
<b>Figura 7.</b> Distribución espacial de las redes y estaciones meteorológicas..... identificadas .....	38
<b>Figura 8.</b> Número de estaciones según sus características .....	42
<b>Figura 9.</b> Comportamiento de la cantidad de observaciones que otorgaron ..... diariamente las estaciones .....	44
<b>Figura 10.</b> Diagrama de barras en 3D con el promedio porcentual de la cantidad.. de observaciones diarias.....	45
<b>Figura 11.</b> Diagrama de dispersión de la cantidad de observaciones diarias para . seis estaciones .....	49
<b>Figura 12.</b> Diagramas de estadísticos de las modas de las observaciones diarias.... por estación agrupadas por redes.....	51
<b>Figura 13.</b> Diagramas de estadísticos de las medias de las observaciones diarias... por estación agrupadas por redes.....	52
<b>Figura 14.</b> Mapa lagunaridad por tipo de estaciones en el territorio español .....	56
<b>Figura 15.</b> Distribución de las estaciones meteorológicas con sensores de IS de ..... las redes (AEMet, CyL y Meteoclimatic) .....	60
<b>Figura 16.</b> Preparación y colecta de datos para la estimación espacial de ISG ...	68
<b>Figura 17.</b> Seis formas distintas de agrupar los datos para el análisis de las ..... estimaciones espaciales .....	70
<b>Figura 18.</b> Flujo de trabajo de los procesos de análisis de los métodos de ..... estimación espacial .....	75

<b>Figura 19.</b> <i>Relación entre la ISG estimada por distintos métodos y la ISG observada por las estaciones en superficie</i> .....	80
<b>Figura 20.</b> <i>%RMSE de las interpolaciones en relación a los vecinos para el método RK2 en CYL</i> .....	83
<b>Figura 21.</b> <i>%RMSE en relación a los vecinos para el método RK2 en grupo VI (CYL + 9 AEMet)</i> .....	88
<b>Figura 22.</b> <i>El mejor método/ área de estudio como la combinación de dos sub-grupos</i> .....	90
<b>Figura 23.</b> <i>Relación entre el Kt y %RMSE diarios</i> .....	92
<b>Figura 24.</b> <i>Límite administrativo de la comunidad Autónoma Junta de Castilla y León y REM automáticas pertenecientes a la Red InfoRiego</i> .....	97
<b>Figura 25.</b> <i>Esquema de organización de los datos de entrada a las RNA de forma horizontal</i> .....	99
<b>Figura 26.</b> <i>Ilustración del método de ventana deslizante como forma de entrada a las RNA</i> .....	102
<b>Figura 27.</b> <i>Representación simplificada de una RNA que para la predice predicción la ISG para en un horizonte “t” futuro, en base a datos de... distintas estaciones vecinas y de sus variables en diferentes instantes... de tiempo</i> .....	103

## LISTADO DE TABLAS

<b>Tabla 1.</b>	<i>Niveles de errores en trabajos relacionados .....</i>	30
<b>Tabla 2.</b>	<i>Redes de estaciones meteorológicas y sus tamaños relativos en..... términos del número de estaciones .....</i>	38
<b>Tabla 3.</b>	<i>Promedio porcentual de la cantidad de observaciones diarias..... agrupadas por red y periodicidad.....</i>	45
<b>Tabla 4.</b>	<i>VARIABLES meteorológicas incluidas en todas las estaciones .....</i>	46
<b>Tabla 5.</b>	<i>VARIABLES meteorológicas incluidas en partes de las estaciones .....</i>	47
<b>Tabla 6.</b>	<i>Media, moda y desviación estándar de las modas de la cantidad de ..... observaciones diarias por estación agrupadas por red.....</i>	49
<b>Tabla 7.</b>	<i>Media, moda y desviación estándar de las medias de la cantidad de .... observaciones diarias por estación agrupadas por red.....</i>	50
<b>Tabla 8.</b>	<i>Promedio porcentual de las modas de la cantidad de observaciones .... diarias por estación agrupadas por red.....</i>	50
<b>Tabla 9.</b>	<i>Promedio porcentual de las medias de la cantidad de observaciones ... diarias por estación agrupadas por red.....</i>	50
<b>Tabla 10.</b>	<i>Cantidad de estaciones potencialmente duplicadas por tipo (GUB /..... RMEV) basadas en 4 valores de buffers .....</i>	53
<b>Tabla 11.</b>	<i>Clasificación de los métodos sin discriminar los valores de irradiancia (agregación diaria) de acuerdo a los 6 grupos de datos .....</i>	76
<b>Tabla 12.</b>	<i>Clasificación de los métodos para valores de irradiancia mayores a.... 200 (agregación diaria) de acuerdo a los 6 grupos de datos .....</i>	76
<b>Tabla 13.</b>	<i>Distancia media hacia la estación Vecina Más Cercana (VMC) y ..... distancia media entre todas las estaciones de cada grupo .....</i>	78
<b>Tabla 14.</b>	<i>El mejor método por grupo y por estadístico, sin discriminar valores .. de irradiancia (agregación para el número total de días) .....</i>	78
<b>Tabla 15.</b>	<i>El mejor método por grupo y por estadístico, discriminando valores de irradiancia mayores de 200 (agregación para todos los días).....</i>	79
<b>Tabla 16.</b>	<i>Estadísticos para el método RK2 en el grupo IV (CYL) .....</i>	81
<b>Tabla 17.</b>	<i>Estadísticos para el método RK2 en el grupo V (CYL + 9 AEMet)....</i>	86
<b>Tabla 18.</b>	<i>Estadísticos para las Estimaciones de la ISG derivadas de imágenes de satélite en el grupo VI (10 AEMet) .....</i>	89

<b>Tabla 19.</b>	<i>Identificación de parámetros para RNA que predicen la ISG: Pruebas con horizonte de 15 minutos .....</i>	<i>105</i>
<b>Tabla 20.</b>	<i>Identificación de parámetros para RNA que predicen la ISG: Pruebas con horizonte de 1 hora .....</i>	<i>106</i>
<b>Tabla 21.</b>	<i>Identificación de parámetros para RNA que predicen la ISG: Pruebas con horizonte de 2 horas.....</i>	<i>107</i>
<b>Tabla 22.</b>	<i>Identificación de parámetros para RNA que estiman la ISG sin..... variables de instantes previos.....</i>	<i>108</i>
<b>Tabla 23.</b>	<i>Resumen de resultados de las validaciones cruzadas de las RNA que... estiman la ISG a corto plazo para horizontes desde 1 a 6 horas .....</i>	<i>109</i>
<b>Tabla 24.</b>	<i>Resumen de las redes GUB con área de influencia mayor o igual al de región (comunidad autónoma).....</i>	<i>143</i>
<b>Tabla 25.</b>	<i>Resumen de las redes GUB con área de influencia menor al de región (comunidad autónoma).....</i>	<i>143</i>
<b>Tabla 26.</b>	<i>Resumen de las REM con área de influencia mayor al de región..... (comunidad autónoma).....</i>	<i>144</i>
<b>Tabla 27.</b>	<i>Resumen de las REM con área de influencia menor al de región..... (comunidad autónoma).....</i>	<i>145</i>
<b>Tabla 28.</b>	<i>URL de acceso a cada red .....</i>	<i>145</i>

## LISTADO DE ACRÓNIMOS Y SIGLAS

- %MAE:** Error Medio Absoluto Porcentual (*Percentage Mean Absolute Error*).
- %RMSE:** Error Medio Cuadrático Porcentual (*Percentage Root Mean Square Error*).
- AEMet:** Agencia Estatal de Meteorología.
- AC:** Ángulo Cenit.
- BBox:** Rectángulo Contenedor.
- CS:** Constante Solar.
- CWOP:** Programa de Observación Meteorológico Ciudadano (*Citizen Weather Observation Program*).
- CYL:** Castilla y León.
- DGT:** Dirección General de Tráfico.
- DV:** Dirección del viento.
- EBRO:** Cuenca del Ebro.
- EISG:** Estimación de la Irradiancia Solar Global.
- EM:** Estaciones Meteorológicas.
- ER:** Energías Renovables.
- ES:** Energía Solar.
- GeoDB:** Base de Datos Geográfica.
- GUB:** Gubernamentales.
- H0d:** Irradiación Extraterrestre Diaria sobre una superficie Horizontal.
- HC3v2:** HelioClim3 versión 2.
- HGd:** Irradiación Global Horizontal Diaria.
- HR:** Humedad Relativa.
- HRA:** Humedad Relativa del Aire.
- IA:** Inteligencia Artificial.
- IDW:** Distancia Inversa Ponderada (*Inverse Distance Weighting*).
- IEA:** Agencia Internacional de Energía (*International Energy Agency*).
- IoT:** Internet de las Cosas (*Internet of Things*).
- IS:** Irradiancia Solar.
- ISE:** Irradiancia Solar Extraterrestre.
- ISEH:** Irradiancia Solar Extraterrestre Horizontal.
- ISEN:** Irradiancia Solar Extraterrestre Normal.
- ISG:** Irradiancia Solar Global.
- JCYL:** Junta de Castilla y León.
- KT:** Índice de Claridad Diario.
- KTi:** Índice de Claridad Instantáneo.
- MADIS:** Asimilación e Ingesta de Datos (*Meteorological Assimilation Data Ingest System*).
- MAE:** Error Medio Absoluto (*Mean Absolute Error*).
- METEO:** Meteoclimatic.

- MLP:** Perceptrón Multicapa (*Multi Layer Perceptron*).
- MLRM:** Modelos de Regresiones Lineales Múltiples (*Multiple Linear Regression Models*).
- MSG:** Meteosat de Segunda Generación (*Meteosat Second Generation*).
- NA:** Neurona Artificial.
- NB:** Neurona Biológica.
- nRMSE:** Error Medio Cuadrático Normalizado (*Normalized Root Mean Square Error*).
- UN:** Sin umbral.
- OGC:** Consorcio Geoespacial Abierto (*Open Geospatial Consortium*).
- OGD:** Datos Gubernamentales Abiertos (*Open Government Data*).
- OK:** Kriging Ordinario (*Ordinary Kriging*).
- OMM:** Organización Meteorológica Mundial.
- P:** Presión.
- V:** Red Eléctrica de España.
- REM:** Redes de Estaciones Meteorológicas.
- REMV:** Redes de Estaciones Meteorológicas de Voluntarios.
- RK:** Kriging con Regresión (*Regression Kriging*).
- RMSE:** Error Medio Cuadrático (*Root Mean Square Error*).
- RNA:** Redes Neuronales Artificiales.
- RS:** Radiación Solar.
- SALMON:** System for Automated Learning Modeling Operative Networks.
- SAT:** Satélite (utilizado para indicar los valores obtenidos desde la base de datos HelioClim3).
- SGBDR:** Sistema de Gestión de Base de Datos Relacional.
- SHC:** Programa de Calentamiento y Enfriamiento Solar (*Solar Heating and Cooling Programme*).
- SWE:** Sensores en la Web (*Sensor Web Enablement*).
- T:** Temperatura.
- TA:** Temperatura Ambiente.
- UTC:** Tiempo Universal Coordinado (*Coordinated Universal Time*).
- VGI:** Información Geográfica de Voluntarios (*Volunteer Geographic Information*).
- VV:** Velocidad del Viento.



# 1 INTRODUCCIÓN

En las sociedades modernas la utilización de los Recursos y Energías Renovables (RER) como la energía solar, hidroeléctrica, eólica o geotérmica entre otras, han tomado fuerza como una de las soluciones a largo plazo con mayor potencial, sostenible y más adecuado para ayudar a minimizar el impacto ambiental y ayudar en la solución de los actuales problemas climáticos (Kaya 2006). La importancia de las fuentes alternativas de energía en el consumo energético mundial, a lo largo de los años, se evidencia en el incremento superior al 100 % del consumo de estas fuentes de energías en los últimos 40 años, pasando del 1.6 % en 1973, al 3.5 % en 2012 (IEA 2014). Este valor aún se puede considerar bajo, si se tienen en cuenta las capacidades de las fuentes de energía renovables existentes, como es el caso del Sol.

El Sol, a una distancia 158 millones de kilómetros, posee un diámetro de aproximadamente 109 veces el de la Tierra, superándola en volumen en más de 1.33 millones de veces (SOHO 2011). Estas dimensiones y otras características hacen que la Tierra reciba del Sol una potencia de energía que se estima equivalente en más de 10.000 veces el consumo de todo el planeta (Moebius 2005; BBC 2013).

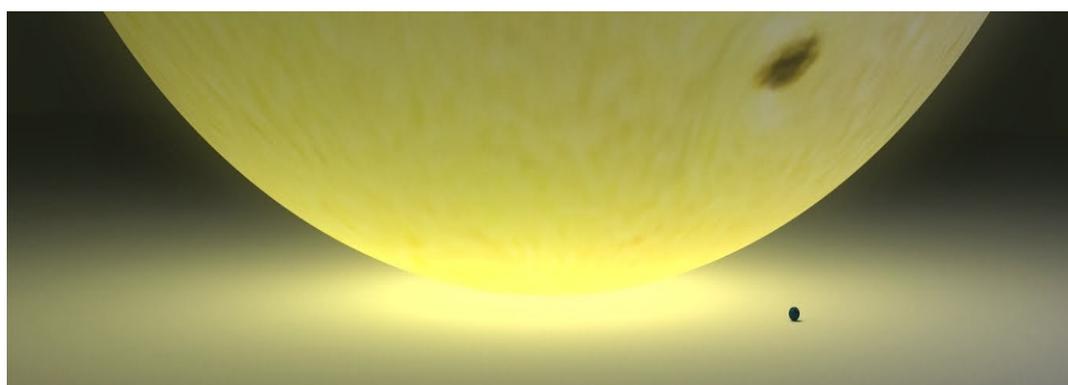


Figura 1. Comparativa de tamaños a escala verdadera entre el Sol y la Tierra, sin representar la distancia entre ambos (Thurston 2014).

Se estima que la población mundial actual consume en 15 teravatios de potencia y se espera que para 2050 este valor se duplique (BBC 2013; The Economist 2008).

Según Morebius (2006), para cubrir la demanda energética actual, bastaría con un área total de celdas solares inferior al 1% de la superficie total del planeta. Esta estimación la hace considerando celdas solares con el 10% de efectividad, tomando en cuenta el día, la noche, la interacción de la cobertura nubosa y otras componentes meteorológicas.

Conscientes de la necesidad de aprovechar mejor los RER y promover el desarrollo sostenible, las grandes economías (EEUU, UE, China entre otras) han definido sus objetivos de uso a medio-largo plazo. La Comisión de Reforma y Desarrollo Nacional de China menciona un objetivo de hasta 100 gigavatios de capacidad fotovoltaica instalada para 2020, aunque el umbral a alcanzar aún está en discusión, es un hecho el creciente aumento de China en este aspecto (600 % en 3 años), pasando de 1 a 7 gigavatios entre 2010 y finales de 2012 (EC 2013). EEUU tiene como meta conseguir que en el año 2030, el 20 % de la energía eléctrica consumida proceda de la energía eólica (Toole *et al.* 2010). De forma similar, la directiva europea relativa al fomento del uso de energía procedente de fuentes renovables (EC 2009a) define como objetivo que para el año 2020, el 20 % de la energía que se consuma en la Unión Europea proceda de fuentes renovables (EE News 2010, p. 11).

En base a estos objetivos, en España se prevé que para el año 2020, el 22.7 % del consumo eléctrico proceda de fuentes renovables como el viento y el Sol (MITYC 2009; IDAE 2010). En este sentido el conocimiento y la predicción de los valores de la Irradiancia Solar (IS) desempeñan un papel importante en el desarrollo y explotación de los sistemas basados en Energía Solar (ES) como son las plantas termo-solares y las fotovoltaicas (Mellit y Pavan 2010), para las cuales la información de la irradiancia es necesaria en todas las etapas, desde la decisión de la selección de sitios idóneos para la construcción de nuevas plantas de ES, pasando por su fase de diseño, hasta finalmente llegar a la etapa de explotación, donde estos datos sirven para la predicción de la producción energética (Perpiñan 2008, p. 21). En este último caso, los datos de la IS y sus predicciones en distintos horizontes temporales son de utilidad para la planificación y administración de la energía tanto para empresas generadoras como para las operadoras del sistema eléctrico (Voyant 2013).

A pesar de la importancia de la información de la IS para investigaciones relacionadas con la energía solar (y otros tipos de estudios), estos datos no siempre están disponibles en los sitios de interés debido a la ausencia de estaciones meteorológicas (Yadav y Chandel 2013). En este sentido, los sensores de Irradiancia Solar Global (ISG) son especialmente relevantes como fuente de información para el cálculo de numerosos procesos ecológicos e industriales, por ejemplo la fotosíntesis, evapotranspiración, producción de energía solar, monitorización de ciudades inteligentes (Hernández *et al.* 2012) y también tiene otras aplicaciones incluyendo la verificación de modelos, la inferencia de datos y la asimilación de datos en modelos y contextos como la meteorología, el clima o la hidrología, entre otros campos de estudio.

Las razones anteriores han motivado los tres aspectos que se abordan en la presente tesis doctoral. (1) Evaluar la disponibilidad de las redes de estaciones meteorológicas que ponen a disposición del público en general los datos de sus observaciones meteorológicas; (2) investigar la estimación espacial de la Irradiancia Solar en intervalos de 15 minutos en el territorio español, y (3), Experimentar el modelado de la predicción temporal de la Irradiancia Solar a corto plazo mediante Redes Neuronales Artificiales (RNA) utilizando observaciones vecinas (componente espacial) como entrada a estos modelos.

## **1.1 Motivación**

La motivación de la presente tesis doctoral, nace de la necesidad de mejorar la densificación espacial del conocimiento de la Irradiancia Solar Global a intervalos de 15 minutos en la superficie así como de predecir su valor en un futuro próximo. De allí surge el interés por investigar y caracterizar las redes de estaciones meteorológicas que publican libremente las observaciones de sus estaciones meteorológicas en Internet (tanto las oficiales como las de voluntarios). La irregular distribución espacial de las estaciones motivó el interés por investigar los métodos de estimación espacial de la ISG en intervalos de 15 minutos en otros puntos del territorio peninsular español. La predicción futura de la ISG mediante Redes Neuronales Artificiales (RNA) nace del interés del grupo de investigación

Mercator de la UPM (Tecnologías de la Geoinformación) en los métodos y tecnologías de ayuda en la toma de decisiones, relacionadas con esta variable, en entornos inteligentes, véase ciudades o redes inteligentes (*smartcities, smartgrid*), mejorando los actuales niveles de incertidumbre.

## **1.2 Objetivos de investigación**

El objetivo principal es investigar la predicción espacial y temporal a corto plazo de la Irradiancia Solar Global (ISG) en el territorio peninsular español mediante Redes Neuronales Artificiales (RNA) y geoestadística con el propósito de generar mejores modelos predictivos que ayuden en la estimación de la producción de energía renovable procedente del Sol. El interés de conocer la realidad sobre la ISG de la forma más continua posible (espacial y temporalmente) ha implicado la experimentación con observaciones de la ISG registradas por estaciones meteorológicas, tanto oficiales como de voluntarios, así como con los valores de la ISG derivados a partir de imágenes satelitales.

## **1.3 Objetivos específicos**

1. Inventariar la disponibilidad de observaciones de redes de estaciones meteorológicas (REM) automáticas en el territorio español, que publiquen datos libremente en internet, y analizar su distribución espacial: tanto la redundancia de estaciones como la lagunaridad espacial de sus localizaciones.
2. Caracterizar cualitativa y cuantitativamente las redes identificadas (p. ej. cantidad de estaciones de cada red, periodicidad de las observaciones, latencia en la entrega de las observaciones, protocolos de comunicaciones de acceso a las observaciones, formatos de datos, etc.).
3. Analizar la disponibilidad de observaciones para tres REM con distintas características: dos de tipo gubernamental (AEMet con cobertura nacional y Castilla y León a nivel regional) y una no oficial (observaciones de estaciones meteorológicas de voluntarios) Meteoclimatic (con cobertura nacional).

4. Analizar la bondad de las predicciones espaciales de la IS, a intervalos de 15 minutos, de cinco métodos en el territorio peninsular español (3 métodos de interpolación geoestadística, uno determinista y el método HelioSat2 a partir de imágenes de satélite) aplicados en 6 configuraciones espaciales de los datos, con el objetivo de analizar los efectos zonales (locales) de una mayor o menor densidad de Estaciones Meteorológicas (EM).

5. Propuesta de inclusión de las observaciones de la IS procedentes de los sensores de las EM de voluntarios como una fuente auxiliar que ayuden a densificar espacial y temporalmente la información de la IS existente en España, tras realizar sobre las mismas un proceso de validación.

6. Predecir temporalmente, a corto plazo (*nowcasting*), la IS mediante el modelado de distintos modelos de redes neuronales artificiales e incluyendo la componente espacial como entrada de los modelos.

## **1.4 Hipótesis**

La presente tesis está compuesta por dos hipótesis:

H1: El uso de los valores de la Irradiancia Solar, derivados a partir de imágenes de satélite, como variable auxiliar en los métodos de predicción espacial de la misma permitirá generar predicciones con errores inferiores al de los propios valores derivados de imágenes de satélite. De este modo se podrá determinar con menor error si las observaciones de la ISG realizadas por las estaciones meteorológicas de aficionados / voluntarios son válidas y, en su caso, caracterizar su error.

H2: Es posible mejorar la predicción a corto plazo (*nowcasting*) de la ISG con modelos basados en Redes Neuronales Artificiales (RNA) mediante el uso de la mayor cantidad de variables de entrada y la inclusión de la componente espacial.

La innovación de la investigación reside en el desarrollo de modelos para predecir espacial y temporalmente la IS a corto plazo en el territorio peninsular español con

mejores niveles de errores a los identificados en la bibliografía para métodos similares. En el caso de la estimación espacial, se innova con la exploración de la fortaleza del uso de distintas fuentes de información disponibles para esta variable climática, como son: EM oficiales, EM de voluntarios y valores de la IS derivados desde sensores remotos. En el caso temporal, la innovación se logra a través de la experimentación e inclusión de componentes espaciales en las entradas de los modelos de redes neuronales artificiales y los análisis independientes de los niveles de errores para los distintos horizontes temporales. Una segunda innovación consiste en evidenciar la existencia, y sus características, de una cantidad importante de EM de voluntarios que se proponen validar para usar sus observaciones de IS, como un medio que ayude en la densificación espacial de esta variable en el ámbito de estudio.

## **1.5 Preguntas de investigación**

Para alcanzar el objetivo fundamental de ésta investigación y comprobar/descartar las hipótesis, se plantean las siguientes cuestiones:

- 1) ¿Existen Estaciones Meteorológicas (EM) que ofrecen datos de forma abierta en el territorio español en más de una red de estaciones?
- 2) ¿Cómo están geográficamente distribuidas las REM en España?
- 3) ¿Es posible interpolar determinística o estadísticamente la IS en el territorio nacional a intervalos de 15 minutos más allá de los 25 km con errores inferiores al 25 %?
- 4) ¿Están las observaciones de la IS de las EM de la red Meteoclimatic (voluntarios) dentro del margen de error del modelo de referencia, para poder considerarlas válidas?
- 5) ¿Se mejora el error de la predicción temporal de la IS a corto plazo en base a RNA con la inclusión de la mayor cantidad de datos de entradas en los modelos?

6) ¿Cómo se relacionan el horizonte espacial y temporal en la predicción de la IS a corto plazo con las RNA?

## **1.6 Metodología**

1. Elaboración del estado del arte de los modelos que estiman la IS, tanto en el ámbito espacial como temporal, identificando las variables de entrada que utilizan, sus incertidumbres y cotas de error, así como sus granularidades espaciales y temporales.

2. Análisis de la disponibilidad y caracterización de las observaciones de la ISG procedentes de distintas REM en superficie identificadas (públicas y voluntarias) en el territorio español, con el propósito de conocer lo mejor posible las estaciones que cubren la geografía del área de estudio.

3. Obtención de los valores de la IS global de las estaciones identificadas previamente, tanto los valores de sus observaciones como las estimaciones de la IS derivada a partir de imágenes de satélite para su emplazamiento.

4. Desarrollo y evaluación de métodos de estimación espacial de la IS global a intervalos de 15 minutos para el territorio Peninsular español y el archipiélago Balear, considerando la densidad y la distribución geográfica de las EM.

5. Validación de la ISG observada por las EM de aficionados / voluntarios.

6. Experimentación de la predicción a corto plazo de la IS global con Redes Neuronales Artificiales (RNA) mediante el diseño de distintos modelos de tipo Perceptrón Multicapa en base a la variación de la cantidad de capas, neuronas y la inclusión de la componente espacial y distintas variables de entrada, que permitan determinar cuáles son las mejores combinaciones que generan las mejores predicciones de la ISG.

## **1.7 Estructura de la tesis**

Esta tesis doctoral está estructurada en 6 capítulos, constituyendo el núcleo de la misma los capítulos 3, 4 y 5 en los que se responde a las preguntas de investigación planteadas. A continuación se enumeran y describen muy brevemente cada uno de los capítulos.

Este primer capítulo de introducción es en el que se plantean los aspectos científicos que dan forma a la tesis doctoral. Se indican las motivaciones, se presentan los objetivos, se plantean las hipótesis y las 6 preguntas de investigación y, finalmente, se describe la metodología utilizada para realizar la tesis.

El segundo capítulo presenta el estado de la cuestión para cada uno de los bloques/capítulos que conforman esta tesis doctoral. El capítulo ha ayudado a formular los aspectos de investigación de la tesis que se plantean en el primer capítulo. Esta revisión ha permitido identificar trabajos relacionados y problemas, trabajar sobre los mismos e identificar las posibles aplicaciones prácticas de los resultados de investigación.

En el capítulo tres, se evalúa la disponibilidad de datos meteorológicos en tiempo cuasi real en el territorio español, indicando las redes y estaciones que existen así como muchas de sus características que facilitan el acceso mediante robots web. En este capítulo se abordan las preguntas de investigación 1 y 2.

En el capítulo cuatro se estudia y se evalúan los métodos de estimación de la Irradiancia Solar Global a intervalos de 15 minutos para el territorio Peninsular español y el archipiélago Balear. Se explora la posibilidad de generar estimaciones espaciales mediante interpolaciones que sobrepasen la distancia límite indicada por estudios relacionados y, en su conjunto, el capítulo responde a las preguntas de investigación 3 y 4.

En el capítulo cinco se investiga la predicción temporal de la Irradiancia Solar Global a corto plazo, experimentando con modelos basados en Redes Neuronales Artificiales que incluyen entre sus entradas, una amplia cantidad de datos

procedentes de las distintas variables en diferentes instantes de tiempos previos, así como las observaciones de estaciones vecinas en paralelo (componente espacial), respondiendo así a las preguntas de investigación 5 y 6.

En el capítulo 6 se presentan las conclusiones de la presente tesis, así como las principales contribuciones científicas. Se argumentan las respuestas a las preguntas de investigación que permiten comprobar las hipótesis y se finaliza el capítulo indicándose los trabajos futuros y líneas de investigaciones que abre la presente tesis. Los últimos dos capítulos se corresponden con la bibliografía y los anexos respectivamente.



## **2 ESTADO DE LA CUESTIÓN**

En algunos casos la complejidad de las aplicaciones, relacionadas con la estimación y predicción de la energía solar, requieren un extenso registro de observaciones meteorológicas, disminuir los intervalos de muestreo de las mismas o reducir la latencia en la entrega de los reportes de observaciones (Bocco *et al.* 2006; Running *et al.* 1987; Salcedo-Sanz *et al.* 2009). Sin embargo estos datos no siempre están disponibles en los lugares de interés debido a la ausencia de estaciones meteorológicas en la zona (Yadav y Chandel 2013).

En base a lo anterior, y focalizando en el hecho de que el cálculo de la energía solar depende principalmente de los valores de la Irradiancia Solar, se divide el capítulo en secciones. Así en la sección 2.1 se describen las consideraciones asociadas a la disponibilidad de datos meteorológicos en España, en la sección 2.2, se describen los componentes de la radiación solar. En la sección 2.3 se indican los métodos para la modelización de la estimación espacial de la Irradiancia Solar Global en periodos inferiores a una hora, y finalmente, en la sección 2.4, se presentaran los modelos existentes para la predicción temporal de la ISG a corto plazo utilizando Redes Neuronales Artificiales.

### **2.1 Disponibilidad de datos meteorológicos y de Irradiancia Solar y en tiempo cuasi real por estaciones**

Existen distintas formas de obtener valores de Irradiancia Solar (IS) en la superficie terrestre. Estos pueden obtenerse tanto por observación directa in-situ mediante sensores de IS instalados en Estaciones Meteorológicas (EM), como también por estimaciones indirectas a través de distintos métodos. Entre los métodos de estimación indirecta destacan los que estiman los valores de la IS a partir de imágenes de satélite. Los modelos lo hacen a partir de otras variables meteorológicas y las interpolaciones espaciales a partir de valores puntuales, entre otros (Badescu 2008; Zelenka *et al.* 1999; NREL 1992).

Independientemente de la variedad de fuentes de valores de IS, se sabe que los

mejores valores de la IS en la superficie terrestre son los observados por las EM con sensores de IS (Polo *et al.* 2008). Así lo precisan algunas de las aplicaciones o procesos relacionados con la IS, que no sólo son especialmente relevantes para el cálculo de la producción de energía solar, sino que además, sirven como fuente de información para calcular numerosos procesos ecológicos e industriales: fotosíntesis, evapotranspiración, monitorización en las ciudades Inteligentes (Hernández *et al.* 2012). Los datos de IS también tienen otras aplicaciones como la verificación de modelos, inferencia de datos o asimilación de datos en modelos usados en contextos tales como la meteorología, climatología o hidrología, entre otros campos de estudio. En algunos casos, la complejidad de estas aplicaciones (como por ejemplo la generación y validación de modelos de cortas resoluciones espacio-temporales) requiere un amplio registro de observaciones con una densificación espacial y temporalmente de las observaciones del fenómeno meteorológico estudiado, y una menor latencia de entrega entre sus reportes (NOAA 2012; Roebber *et al.* 2004; Running *et al.* 1987; Salcedo-Sanz *et al.* 2009).

Sin embargo, a pesar de la importancia de las observaciones de la IS mediante EM antes mencionada, éstas presentan la desventaja de su poca densificación espacial. Según Badescu (2008), la relación entre EM oficiales con sensores de IS y temperatura en USA es de 1:100, y en el resto de mundo es de 1:500.

Todo lo anterior implicaría que la densificación espacial de las estaciones meteorológicas en superficie o sensores que observen la IS fuese óptima. Sin embargo, de acuerdo con Antonanzas-Torres *et al.* (2013), la escasa densificación se debe a la inversión necesaria para su instalación y mantenimiento. Hecho que tiene un efecto disuasivo al considerar la opción de densificar las redes de EM de observación de la IS. Por esta razón, en esta tesis doctoral, se indaga y se caracterizan las Redes de EM en tiempo cuasi-real, que mediante Internet publican libremente sus observaciones. Se incluyen además de las REM Gubernamentales, las REM de voluntarios como propuesta de mecanismo alternativo que ayude en la densificación espacial de la observación de la IS por estaciones en superficie. Ésta propuesta se sustenta en la evolución de internet, que ha hecho posible que no sólo

las agencias oficiales publiquen datos meteorológicos sino que también lo hagan aficionados / voluntarios, incrementándose así la cantidad de datos disponibles (Knights 2007 *et al.* 2007; Yanes 2011).

### **2.2.1 Evolución de los registros meteorológicos**

La disponibilidad, almacenamiento y accesibilidad de las observaciones meteorológicas ha evolucionado con el tiempo. Los primeros registros climatológicos eran observaciones no instrumentales de eventos climáticos catastróficos o listados cronológicos del clima diario (Landsberg 1980). Las primeras series de observaciones captadas por estaciones meteorológicas instrumentales son las series de observaciones de la temperatura en el centro de Inglaterra desde 1659 (Manley 1974) o las series de precipitaciones de París desde 1688, entre otras (Slonosky 2002; Slonosky y NOAA/NCDC 2010; Brázdil *et al.* 2005). Uno de los primeros medios utilizados para la notificación rápida de información meteorológica fue el correo tradicional, cuando en EEUU Benjamín Franklin investigaba el progreso de un huracán en 1753. Los inicios de las telecomunicaciones, permitieron el uso del telégrafo en 1846 para la alerta por tormentas en EEUU y el envío de reportes meteorológicos diarios en 1848 en Nueva York y en 1849 en Londres (Miller 1931).

El número de observaciones meteorológicas creció, pasando de ser documentadas por pocos humanos, hasta formar redes más densas automatizadas en el siglo XX. La primera estación meteorológica automática operativa en EEUU data de 1941. De forma similar el desarrollo de los radares meteorológicos también permitió ofrecer observaciones en tiempo real (Fiebrich 2009). La era espacial introdujo el uso de imágenes de satélite y las técnicas de teledetección para la determinación de valores climáticos. El primer satélite meteorológico fue lanzado en 1960 por EEUU (Fiebrich 2009; Conway 1997). A partir de toda la información anterior, ha sido común encontrar información meteorológica en distintas organizaciones gubernamentales y científicas (fuentes oficiales de información meteorológica). En España, existe la Agencia Estatal de Meteorología (AEMet) e internacionalmente la Organización Mundial de Meteorológica (OMM, 2012).

A finales de los 90 con el auge de la web, se produjo un punto de inflexión respecto a la accesibilidad de la información (Knights 2007). Mucha información meteorológica de los organismos oficiales comenzó a estar disponible en la Web. Un ejemplo claro es la Red Automática de Información Meteorológica de Florida que desde 1998 comparte en internet su información (Lusher *et al.* 2009). Otro ejemplo más reciente en España es el caso de AEMet que abrió sus observaciones meteorológicas por un periodo de dos años a partir de noviembre de 2010 (AEMet 2010), alineándose en ese periodo, con las políticas de compartir abiertamente la información generada por las administraciones (Obama 2009; European Commission 2009b).

### **2.2.2 Observaciones Meteorológicas de altruistas**

El descenso en los precios de las EM está permitiendo que aficionados, particularmente, adquieran su propia estación (Williams *et al.* 2009). La eclosión de internet, sumado a la aparición de la tecnología conocida como “Internet de las Cosas” (*Internet of Things: IoT*) que facilitan compartir información desde distintos tipos de dispositivos, sensores, y particularmente EM en redes globales (Gubbi *et al.* 2013; Hartley 2012; Miorandi *et al.* 2012). Estos son factores que pueden asociarse a la creciente incorporación en la web de mucho contenido de tipo meteorológico generado por usuarios (Williams *et al.* 2009; Knights 2007; Schroth y Janner 2007).

El movimiento social consistente en generar/compartir datos georreferenciados en Internet, se denomina de forma general como “Información Geográfica de Voluntarios” (*Volunteer Geographic Information: VGI*). De acuerdo con Goodchild (2007b), la VGI ha emergido como una forma de Web 2.0. WikiMapia (WikiMapia 2013) y OpenStreetMap (OpenStreetMap 2013) son buenos ejemplos de VGI (Neis *et al.* 2011). En los últimos años se han realizado varios estudios para evaluar la calidad de los datos VGI (Al-Bakri y Fairbairn 2012; Neis *et al.* 2011; Finn *et al.* 2010; Girres y Touya 2010), constituyendo un hecho que revela el latente interés por la utilidad de este tipo de información. Este movimiento

también es activo en el campo de la meteorología, con las REM de voluntarios. En este ámbito los usuarios asignan una localización y metadatos a sus EM, mientras que utilizan la tecnología de IoT para compartir las observaciones realizadas por sus estaciones automáticas en redes de usuarios (Hartley 2012).

Estos movimientos (VGI) están haciendo que las agencias oficiales de este tipo de información (geográfica y meteorológica) dejen de ser las únicas productoras de datos; permitiendo a los usuarios pasar del usual rol de consumidor a asumir también del de productor (*prosumer*) (Yanes 2011; Goodchild 2007a; Goodchild 2007b). Alrededor del mundo existen muchas Redes de Estaciones Meteorológicas de Voluntarios (REMV), como por ejemplo: Programa de Observación Meteorológico Ciudadano (*Citizen Weather Observation Program: CWOP*) con más de 8000 estaciones (CWOP 2012), WeatherLink con 7922 (WeatherLink 2012), Meteoclimatic con 1635 (Meteoclimatic 2014), Weather Underground (Weather Underground 2012a) con más de 17000 incluyendo las estaciones personales del Sistema de Asimilación e Ingesta de Datos (*Meteorological Assimilation Data Ingest System: MADIS*) de la agencia gubernamental estadounidense para la Administración Nacional Oceánica y Atmosférica (*National Oceanic and Atmospheric Administration, NOAA*) (Weather Underground 2012b). En España, además de AEMet y las redes de voluntarios, se han desplegado varias REM oficiales manejadas por la administración autonómica, universidades y ministerios (relacionados con agricultura, tráfico e investigación del clima) que difunden en la web sus observaciones.

En el ámbito público, existe la iniciativa de Datos Gubernamentales Abiertos (*Open Government Data: OGD*). Se trata de una iniciativa global para hacer que los datos oficiales de los gobiernos se ofrezcan libremente en internet para el público en general (Kalampokis *et al.* 2011). En España, como en otros países, el OGD ha sido dotado de un marco legal (mediante el RD 1945/2011) desde noviembre de 2011 (BOE 2011). Esta ley pretende que los datos, generados por el desempeño de las funciones propias de autoridades públicas, estén disponibles para todos los ciudadanos para fomentar su reutilización en general.

De acuerdo con Mitchell (Mitchell 2010), “la iniciativa OGD apareció a través de internet desde que el presidente de EEUU, Barack Obama, en el año 2009, realizó su primera acción ejecutiva para la apertura y transparencia”. Algunos gobiernos de las mayores economías mundiales han desarrollado sitios webs con catálogos que listan los puntos de accesos a los datos generados por sus administraciones. Por ejemplo el sitio web de EEUU ([www.data.gob](http://www.data.gob)) o los sitios webs de algunos miembros de la Unión Europea como Reino Unido ([www.data.gov.uk](http://www.data.gov.uk)) y España ([www.datos.gob.es](http://www.datos.gob.es)). En España también se pueden enumerar varias fuentes oficiales de información meteorológica que ponen a disposición sus datos, por ejemplo: AEMet (Abella 2011) a nivel estatal, Euskadi mediante EUSKALMET (Gobierno Vasco 2012a) como parte del OpenData Euskadi (Gobierno Vasco 2012b) y METEOGALICIA (Junta de Galicia 2012a) como parte del OpenData de la Junta de Galicia (Junta de Galicia 2012b).

## **2.2 La Radiación Solar**

La Radiación Solar (RS) es el término utilizado para describir la energía que por fusión nuclear se libera desde el centro del Sol cuando este convierte átomos de hidrogeno en helio, generándose así una pérdida de masa y una liberación de energía (Duffie y Beckman 1991; Mayo 2013; NASA 2012). Esta energía (RS), viaja desde el Sol en forma de ondas electromagnéticas, con longitudes de onda del espectro electromagnético comprendidas entre los 300 nm a 4000 nm y son aprovechadas, al llegar a la Tierra, por las tecnológicas renovables de tipo solar (Gueymard y Myers 2008).

La Irradiancia Solar que llega a la tierra puede verse en dos etapas. La primera es la IS extraterrestre, que puede ser calculada de forma determinista, y la segunda es la Irradiancia Solar Global (ISG) que llega a la superficie de la Tierra y a la que le afectan las condiciones estocásticas de la atmósfera.

### 2.2.1. Irradiancia Solar Extraterrestre

La Irradiancia Solar que recibe la Tierra en el límite exterior (tope) de la atmósfera, es conocida como Irradiancia Solar Extraterrestre (ISE). Se pueden distinguir dos formas de ISE: (i) la proyectada sobre un plano normal a los rayos del Sol, denominada “Irradiancia Solar Extraterrestre Normal” (ISEN), y (ii) la “Irradiancia Solar Extraterrestre Horizontal” (ISEH), que es la Irradiancia que llega al exterior de la atmósfera a un plano horizontal paralelo a la superficie de la Tierra que tiene por debajo. En ambos casos la ISE se caracteriza por depender de factores deterministas de tipo astronómico y geográfico (constante solar, distancia Tierra-Sol a lo largo del año, movimientos de rotación y traslación de la Tierra, la latitud y longitud de ubicación).

La Constante Solar (CS) es un valor de la ISE que se calcula en base a la distancia media entre el Sol y la Tierra, y que corresponde a un plano de un metro cuadrado de superficie ubicado en el tope de la atmósfera y orientado de forma perpendicular a los rayos del Sol. Según recomendaciones de la Organización Mundial de Meteorológica el valor de la CS es de 1367 W/m<sup>2</sup> (Harrouni 2008; Gueymard 2004). La ISEH, ISEN y la CS se relacionan entre sí a través de las ecuaciones (1) y (2) (Duffie y Beckman 1991).

$$ISEN = CS \left( 1 + 0.033 \cos \left( \frac{360n}{365} \right) \right) \text{ [W/m}^2\text{]} \quad (1)$$

donde “n” es el día del año. En (1) se puede apreciar que el valor de la ISEN varía en aproximadamente +/- 3% del valor de la CS en función del día del año “n”. Llegando a obtener un valor máximo de ISEN en el perihelio (a inicios del año cuando la Tierra está más cercana al Sol), y el valor mínimo en el afelio (a mediados del año).

$$ISEH = ISEN \cos \theta_z \text{ [W/m}^2\text{]} \quad (2)$$

donde  $\theta_z$ , es el ángulo de los rayos del Sol incidente (ángulo cenital) sobre un plano horizontal en el tope de la atmósfera para cualquier instante de tiempo. Este

ángulo, puede describirse en función de otros ángulos (Duffie y Beckman 1991, p. 13). Si se reescribe la ecuación (1) en (2), y se considera  $\theta_z$  en función a la latitud de observación ( $\phi$ ), la declinación solar ( $\delta$ ) y el ángulo de la hora solar ( $\omega$ ), la ecuación de ISEH queda de la siguiente forma (3):

$$ISEH = CS \left( 1 + 0.033 \cos \left( \frac{360n}{365} \right) \right) (\cos \phi \cos \delta \cos \omega + \sin \phi \sin \delta) \text{ [W/m}^2\text{]} \quad (3)$$

La ISEH presentada en la ecuación (3) adopta especial interés para esta tesis doctoral, puesto que ha sido implementada y utilizada como datos de entrada en los experimentos del modelado de la predicción temporal a corto plazo de la Irradiancia Solar.

### **2.2.2. Componentes de la Irradiancia Solar**

La ISG que llega a la superficie de la Tierra es el resultado de la ISEH que se transmite desde la periferia de la atmósfera hasta la superficie y que como resultado de este viaje se ve afectada por las condiciones estocásticas asociadas a los movimientos atmosféricos. De esta forma, una parte de la ISEH primero es reflejada por la atmósfera, después atenuada, absorbida, dispersada por los aerosoles, vapor de agua y por varios factores asociados a las nubes como son la cantidad, forma y espesor de las mismas (Wang 2012). Con todo ello la IS restante siempre es menor que su valor de ISEH correspondiente en un lugar y tiempo dado. A la proporción existente entre ambos valores se la conoce como Índice de Claridad, que puede ayudar a identificar las condiciones meteorológicas.

La interacción de la atmósfera y la IS, hace que se distinga entre los distintos componentes de la IS que llega a la superficie de la Tierra: (1) Irradiancia Solar Directa (es aquella que llega a un punto directamente de los rayos del Sol sin haber sido afectada por la atmósfera), (2) Irradiancia Solar Difusa (es la irradiancia que procede desde la atmósfera por los efectos de dispersión, reflejo y absorción), (3) Irradiancia Solar de Albedo (es aquella reflejada por el suelo debido a edificios, masas de agua, montañas y etc.) y, (4) La Irradiancia Solar Global en un plano horizontal que es la suma de los tres anteriores (Gueymard y Myers 2008).

En la figura 2 se ilustran la incidencia de los tres componentes de la IS sobre un árbol (las tres flechas que llegan al árbol). La figura también ilustra cómo parte de la IS es dispersada, reflejada y absorbida por la atmósfera y cómo otra parte es reflejada por las nubes.

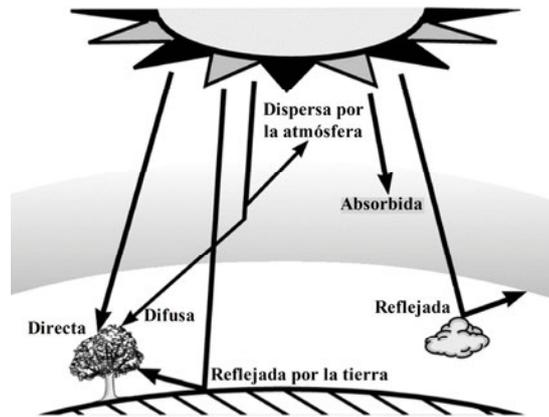


Figura 2. Componentes de la Radiación Solar. Figura adaptada desde [www.newport.com](http://www.newport.com)

La ISG es la componente más común en las bases de datos de IS. En la presente tesis doctoral, se utiliza exclusivamente la ISG que incide sobre un plano horizontal, por ser la IS que nos ofrecen las estaciones meteorológicas usadas en los experimentos, obtenida de las distintas fuentes.

## 2.3 Estimación espacial de la ISG intra-horaria

Como ya se describió antes, los valores de la IS en la superficie terrestre pueden obtenerse por observación directa *insitu* o estimada de un modo indirecto. En esta sección del estado del arte, se presentan los métodos de interpolación espacial utilizados en la estimación de superficies de ISG haciendo énfasis en intervalos inferiores a una hora.

### 2.3.1. Método - Inverse Distance Weighting (IDW)

El método de Distancia Inversa Ponderada (*Inverse Distance Weighting*: IDW), es un método de interpolación espacial determinístico, la influencia entre

observaciones más cercanas y más lejanas del punto a interpolar está definida de un modo determinista/analítico. Los pesos que se le asignan a las observaciones de Irradiancia Solar de cada estación son inversamente proporcionales a la distancia entre el punto a estimar (interpolar) y las estaciones de referencia. Por tanto al tomarse únicamente en cuenta la cercanía de las observaciones respecto al punto a interpolar, implica que mientras menor sea dicha distancia, mayor será la influencia que ejercen los valores de ISG observados por las estaciones. Distintos estudios han utilizado IDW para interpolar la IS (Pons y Ninyerola 2008; Zelenka *et al.* 1999; Perez *et al.* 1997). La fórmula general del método de interpolación IDW se presenta en la ecuación (4):

$$\hat{z}_{idw(s_0)} = \frac{\sum_i^n \frac{Z_i}{d_{i0}^\beta}}{\sum_i^n \frac{1}{d_{i0}^\beta}} \quad (4)$$

donde  $\hat{z}_{idw(s_0)}$ , es el valor estimado en el nodo  $s_0$ ,  $n$ , es el número de puntos observados,  $Z_i$ , el valor de la observación  $i$ ,  $d_{i0}$  es la distancia desde el punto  $i$  al nodo  $s_0$  a estimar. Tanto el denominador principal como el valor de ponderación  $\beta$  tienen como finalidad controlar la forma en que los pesos disminuyen con la distancia.

### 2.3.2. Método - Ordinary Kriging (OK)

Tanto el método Kriging Ordinario (*Ordinary Kriging*: OK) como el siguiente, Kriging con Regresión (*Regression Kriging*: RK), son métodos geoestadísticos. La geoestadística modela un semivariograma, que considera tanto la distancia entre los puntos observados como la variación de sus valores, para definir objetivamente los pesos a utilizar en la interpolación. El método OK también ha sido utilizado en distintas ocasiones para la estimación de la IS (Bojanowski *et al.* 2013; Alsamamra *et al.* 2009). La fórmula general de OK se presenta en la ecuación (5) y la fórmula del semivariograma experimental en la ecuación (6).

$$\hat{z}_{ok}(s_0) = \sum_{i=1}^n w_i(s_0) z(s_i) \quad (5)$$

donde  $\hat{z}_{ok}(s_0)$  es el valor estimado en el nodo  $s_0$ ,  $n$  es la cantidad de puntos observados utilizados,  $w_i$  es el vector de pesos Krigings, y  $Z_i$  el valor de la observación  $i$ .

$$\gamma(h) = \frac{1}{2} \sum [(z(s_i) - z(s_i + h))^2] \quad (6)$$

aquí,  $z(s_i)$  es el valor de observación en un punto  $s_i$  dado,  $z(s_i + h)$  es el valor de la observación vecina ubicada a una distancia  $h$ ; de ésta forma se asignan los pesos a las distintas observaciones respecto al nodo a interpolar.

Una vez procesado el semivariograma experimental se procede a calcular el modelo de variograma, que puede ser de tipo: lineal, esférico, exponencial, circular, *Gaussian*, *Bessel*, *power* o similares (Hengl 2009). En la presente tesis doctoral se utilizó el paquete *automap* para R, tanto para el cálculo del semivariograma experimental como para el procesamiento y la selección del semivariograma teórico de mejor ajuste.

### 2.3.3. Método - Regression Kriging (RK)

Kriging con Regresión (*Regression Kriging*: RK) es un método geoestadístico que incluye información adicional a la variable objeto de la interpolación (en forma de variables auxiliares explicativas), con el propósito de auxiliar al proceso de interpolación con datos que estén mejor distribuidos espacialmente. Así las variables auxiliares funcionan como apoyo la variable objeto ya que poseen mejor dispersión espacial y de este modo permiten compensar la relativa poca distribución espacial de la variable objetivo (Alsamamra *et al.* 2009; Bojanowski *et al.* 2013).

El método RK ya ha sido utilizado en distintas ocasiones para estimar la IS (Antonanzas-Torres *et al.* 2013; Alsamamra *et al.* 2009; Evrendilek y Ertekin

2007; Ertekin y Evrendilek 2007), así como para interpolar otras variables meteorológicas y climáticas (Hengl *et al.* 2012; Hengl 2009; Pebesma 2006; Hudson 1993).

Al igual que en OK, la base del RK reside en el modelo de variograma usado para generar las interpolaciones; además, RK está compuesto por la suma de la parte estocástica y de la determinista de la variación espacial. Ambas partes pueden ser modeladas de forma independiente y procesadas en dos pasos, como se muestra en las ecuaciones (7) y (8).

En los experimentos de estimación espacial de esta tesis doctoral que utilizan RK, primero se localiza la parte determinista mediante un ajuste por Modelos de Regresiones Lineales Múltiples (*Multiple Linear Regression Models: MLRM*) entre la Irradiancia Solar Global y las variables auxiliares. Seguidamente se procesa la parte estocástica, que consiste en la ejecución de OK sobre los residuos del MLRM. La suma de ambos valores genera la estimación final de la Irradiancia Solar Global en un punto dado.

$$\hat{z}(s_0) = \hat{m}(s_0) + \hat{\epsilon}(s_0) \quad (7)$$

La variable objeto de la estimación es función de la suma de los modelos determinista y estocástico, primer y segundo término de la ecuación respectivamente. Los componentes de la ecuación (7) son derivados en la ecuación (8):

$$\hat{z}(s_0) = \sum_{k=0}^p \hat{\beta}_k q_k(s_0) + \sum_{i=1}^n \lambda_i \epsilon(s_0) \quad (8)$$

En la expresión,  $p$  es la cantidad de variables auxiliares,  $\hat{\beta}_k$  es la estimación de los coeficientes por el MLRM,  $q_k(s_0)$  son las variables auxiliares en los puntos  $s_0$  dados,  $\lambda_i$  son los pesos Kriging determinados por la estructura de la dependencia espacial de los residuos, y  $\epsilon(s_0)$  son los residuos en los puntos  $s_0$  dados.

### 2.3.4. Estimación mediante Sensores Remotos

La base para la estimación de la Irradiancia Solar mediante sensores remotos es la relación inversa que existe entre la reflectividad en el exterior de la atmósfera y la Irradiancia Solar, asociada a la nubosidad (Wielicki *et al.* 1996). Esta relación fue derivada a partir de la correlación que en la década de los 60's encontraron Fritz *et al.* (1964) entre las medidas de Irradiancia Solar por estaciones meteorológicas y la reflectividad obtenida por el satélite TIROS III. A partir de esa fecha, son varios los estudios que han implicado el uso de imágenes de satélite para la estimación de la Irradiancia Solar (Posselt *et al.* 2012; Laszlo *et al.* 2008; Pinker y Laszlo 1992; Gautier *et al.* 1980; Tarpley 1979).

Entre las ventajas de los sensores remotos destacan el permitir generar superficies continuas de valores de Irradiancia Solar con una continuidad espacial que abarca grandes áreas del planeta, esto permite conocer fácilmente las diferencias relativas entre zonas (Bojanowski 2014, p. 7). Sus desventajas son la resolución espacial y temporal, la espacial está ligada al tamaño del pixel de las imágenes que toma el satélite y la distorsión con la latitud y longitud (en el caso de los satélites geoestacionarios) (Ineichen *et al.* 2009). Otra desventaja está asociada al tiempo requerido para obtener y procesar la imagen de satélite para generar las estimaciones de la Irradiancia Solar a partir los valores de sus pixeles (ubicación y color de los mismos).

En Europa, a partir de la familia de satélites geoestacionarios Meteosat, se han generado distintas bases de datos de Superficies de Irradiancia Solar. Entre ellas destacan la base de datos HelioClim3 (SoDA-is 2011) que es utilizada en el capítulo 4 de la presente tesis. HelioClim3 se genera a partir del método HelioSat-2 (Rigollier *et al.* 2004), que estima la Irradiancia Solar en la superficie terrestre usando dos imágenes del espectro visible (600 nm y 800 nm) obtenidas con una periodicidad de 15 minutos por el satélite MSG (Meteosat Second Generation) (Blanc *et al.* 2011; SoDA-is 2013). De forma resumida el método HelioSat-2 sigue los siguientes pasos: (1) mezcla las dos imágenes (600 nm y 800 nm) normalizando el resultado de acuerdo al valor registrado por el sensor y a la elevación instantánea del Sol, (2) genera una nueva imagen en la que el pixel más

blanco es el más nublado, (3) calcula la proporción de nubosidad en cada pixel de la imagen anterior y la compara con el mismo valor de pixel cuando las condiciones son de cielo despejado, (4) genera, finalmente, los valores estimados de la Irradiancia Solar.

Actualmente en Soda-is están disponibles las bases de datos de Irradiancia Solar de HelioClim1 (HC1), HelioClim3 (HC3) y se anuncia que pronto estará a disposición la versión 4 (HC4) (SoDA-is 2014b). HC1 contiene información de IS desde 1985 hasta 2005, periodos de tiempo que no fueron de interés para la presente tesis doctoral. HC3 contiene información de la IS desde 2004 hasta la actualidad generando nueva información cada día. Soda-is provee valores de HC3 para los distintos componentes de la IS (Global, Directa, Difusa) y en diferentes formas (como Irradiancia e Irradiación). La resolución temporal también se ofrece con varios niveles de granularidad: anual, mensual, semanal, diario, hora y subhoras. Los valores de la IS pueden obtenerse para distintas orientaciones del plano (horizontal, ángulo indicado de inclinación y normal). La resolución espacial de la estimación es aproximadamente de 4 km x 4 km, en promedio, para el territorio español (SoDA-is 2011). En esta tesis doctoral se utilizan los valores de Irradiancia Solar Global de HC3 para un plano horizontal, estos valores se obtuvieron mediante una conexión de internet a los servicios webs del sistema Soda-is (SoDA-is 2013).

En la figura 3 se muestra la cobertura de la base de datos HelioClim3. A simple vista se aprecia que cubre ampliamente la geografía española. La figura 3 (a) muestra la Irradiancia Solar Global sobre el plano horizontal durante el año 2010, de tal manera que los colores más cercanos al azul son valores mínimos mientras que los más cercanos al rojo indican valores máximos. En ella se aprecia, fácilmente, el potencial de la Irradiancia Solar en España respecto de otros países europeos. La figura 3 (b) muestra la resolución espacial de las estimaciones de la IS debidas a las distorsiones ópticas de captura de imagen del satélite Meteosat. Se puede apreciar que el pixel central de la imagen tiene una resolución espacial 3 km x 3 km, mientras que en los extremos dicha resolución alcanza valores superiores a 12 km x 12 km. Esto se debe a la característica geoestacionaria del satélite

Meteosat y a la diferencia de latitudes geográficas.

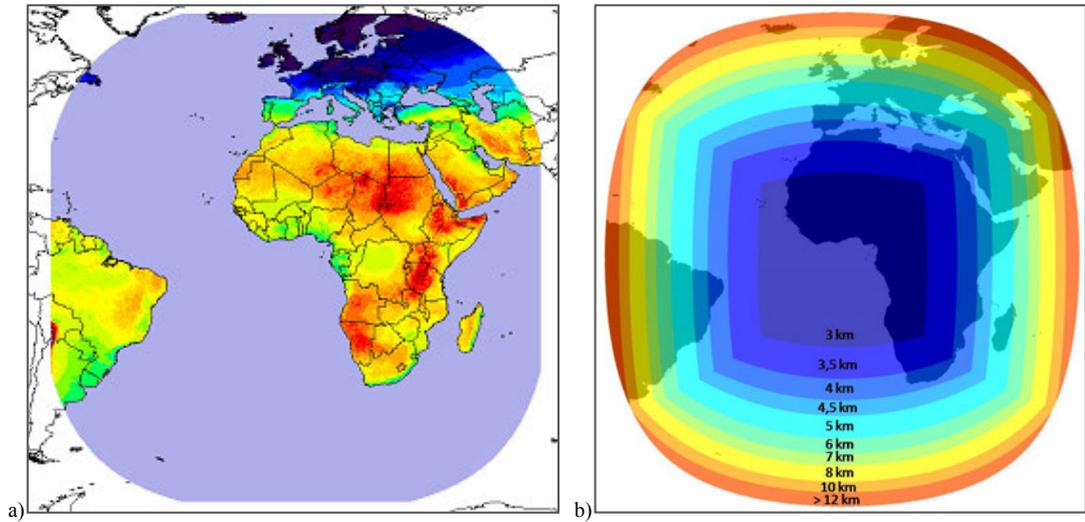


Figura 3. Resolución espacial del satélite Meteosat. (a) Máscara rectangular con esquinas redondeadas que muestra la cobertura de la base de datos HelioClim. (b) Variación de la resolución espacial de forma más o menos concéntrica a medida que se aleja de la correspondiente posición central del satélite en la superficie. Fuente: © 2014 MINES ParisTech / ARMINES / Transvalor.

## 2.4 Predicción de la ISG mediante Redes Neuronales Artificiales

La predicción de la Irradiancia Solar Global (ISG) a distintos horizontes temporales es fundamental para la estimación de la generación eléctrica en plantas solares ya sean térmicas o fotovoltaicas. Los sistemas de pronóstico son capaces de proveer aproximaciones realistas de la dinámica de la atmósfera a través de métodos numéricos que garantizan su estabilidad mediante los procesos de asimilación de datos (Kretschmar *et al.* 2004), sin embargo estos no son adecuados cuando se trata de la predicción a corto plazo de la IS Global (ISG) que llega a la superficie de la Tierra. La ISG a corto plazo presenta la complejidad de poseer características no lineales que se amplifican ante condiciones meteorológicas cambiantes y por ser una variable a la que le afectan las nubes, los gases y otras partículas de la atmósfera como el polvo (Wang *et al.* 2012; Badescu 2008, p. vii). Con el propósito de dar soporte a la planificación y administración de

los sistemas de generación y distribución eléctricos, existe una demanda creciente desde el sector energético para la generación de pronósticos a corto plazo más exactos de los recursos de energía solar (Martins *et al.* 2012; Guarnieri *et al.* 2012).

Las Redes Neuronales Artificiales (RNA) forman parte del área de conocimiento de la Inteligencia Artificial (IA), emulando mediante programas informáticos, la capacidad humana de aprender, memorizar y encontrar relaciones. Las RNA en particular intentan reproducir de manera extremadamente simple el comportamiento de las redes neuronales biológicas (Hagan *et al.* 1996). La habilidad para aprender relaciones no lineales y su capacidad de modelar sistemas complejos, de los que se desconocen las reglas de sus comportamientos, las han convertido en una herramienta útil en distintos ámbitos científicos (Reed y Marks 1999; Yadav y Chandel 2013; Paoli *et al.* 2010).

#### **2.4.1. La Neurona Artificial**

La unidad básica de las RNA es la Neurona Artificial (NA), que es una abstracción matemática simplificada del comportamiento de una Neurona Biológica (NB). Las RNA se componen de un gran número de NA agrupadas en capas y altamente conectadas entre sí, para trabajar conjuntamente en la solución de un problema. En la figura (4) se ilustran ambos tipos de Neuronas.

La figura 4 (a) muestra una Neurona Biológica y en ella se pueden apreciar las ramas del árbol de dendritas, que son las encargadas de contactar y coleccionar información desde otras neuronas, a través del axón. El núcleo de la célula es el encargado de integrar las señales recolectadas y generar una respuesta que se transfiere, de nuevo a través del axón, hacia otras neuronas. De forma análoga en la figura 4 (b) se presenta la NA. Al igual que el árbol de dendritas de la NB, la NA recibe un conjunto de señales mediante las entradas,  $x_1, x_j, x_n$ . Cada entrada es amplificada o reducida por su correspondiente valor de peso  $w_1, w_j, w_n$  (que se corresponden con las sinapsis en la NB), de tal forma que los productos de esos dos valores (señal por el peso) son sumados en  $u$ , y su resultado enviado a la

función de transferencia  $f(u)$ , que se corresponde metafóricamente con el núcleo de la NB. Esta función procesa y genera la salida y de la NA (Reed y Marks 1999; Mubiru 2008; Cortez *et al.* 2002).

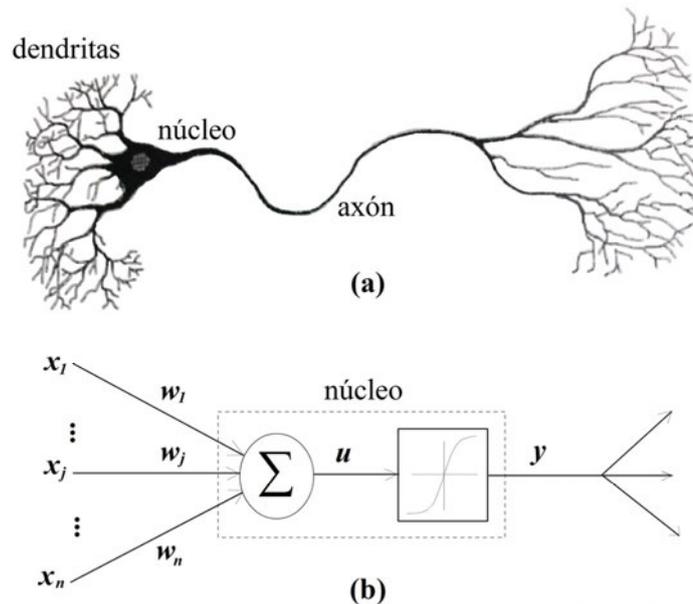


Figura 4. Representación de una Neurona Biológica (a) y una Neurona Artificial (b) Figura adaptada desde Reed y Marks (1999).

La ecuación (9) representa el modelo matemático de la Neurona Artificial correspondiente a la figura 4, y que a continuación se describe.

$$y = f\left(\sum_{j=1}^n w_j x_j\right) \quad (9)$$

En esencia, el conocimiento que adquieren las RNA es almacenado en los pesos de las interconexiones entre neuronas (sinapsis en la NB). En Hagan *et al.* (1996) se presenta una descripción más profunda de la Neurona Artificial.

### 2.4.2. Perceptrón Multicapa

El Perceptrón Multicapa (*Multi Layer Perceptron*: MLP) es la arquitectura de RNA más popular empleada en la solución de problemas científicos (Tymvios *et*

al. 2008), también se le conoce como el *aproximador* universal, por la capacidad demostrada para aproximar relaciones no lineales entre entradas y salidas con cualquier grado de exactitud (Reed y Marks 1999, p. 37). Una RNA con arquitectura de tipo MLP cuenta con, al menos, una capa de neuronas ocultas, y cada capa se conecta hacia adelante con todas las neuronas de la siguiente capa.

En la figura 5 se ilustra una RNA de tipo MLP con  $n$  valores de entrada, una capa oculta con  $N_i$  neuronas artificiales y una sola neurona “o” en la capa de salida. En la figura también se aprecian los pesos ( $w$ ) asociados a cada conexión entre neuronas. Los superíndices  $L1$  o  $L2$  indican la capa a la cual pertenece el peso; el primer subíndice indica la entrada o neurona desde donde parte la conexión y el segundo subíndice indica la neurona hacia la que se dirige la conexión. La salida “o” del MLP de esta figura responde al sistema de ecuaciones (10) y (11).

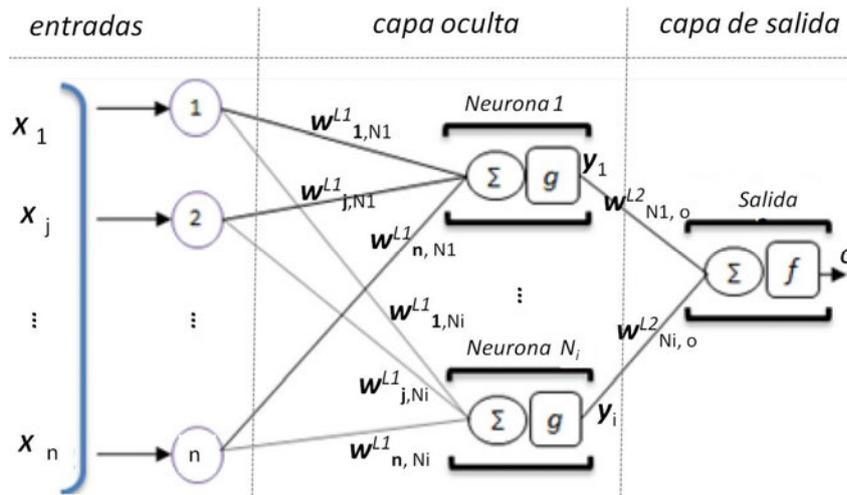


Figura 5. RNA de tipo Perceptrón Multicapa (MLP) con  $n$  entradas,  $i$  neuronas en la capa oculta y una neurona “o” en la capa de salida. Figura adaptada desde Voyant *et al.* (2013).

$$o = f \left( \sum_{i=1}^{i=i} (y_i) * (w^{L2}_{Ni,o}) \right) \quad (10)$$

$$y_i = g \left( \sum_{j=1}^n (x_j) * (w^{L1}_{j, Ni}) \right) \quad (11)$$

### 2.4.3. Modelización de la IS a corto plazo con RNA

Las RNA han sido ampliamente utilizadas en la modelización de la predicción de Irradiancia Solar en distintos horizontes horarios y en distintos lugares del mundo como España (Linares-Rodríguez *et al.* 2011; Bosch *et al.* 2008; Hontoria *et al.* 2002), Italia (Mellit y Pavan 2010), Isla de Córcega en Francia (Voyant *et al.* 2014; Voyant *et al.* 2011; Paoli *et al.* 2010), Turquía (Koca *et al.* 2011), Arabia Saudita (Mohandes *et al.* 1998), China (Wang *et al.* 2012), India (Yadav y Chandel 2012), Argelia (Mellit *et al.* 2006), Uganda (Mubiru y Banda 2008; Mubiru 2008), entre otros (Mellit *et al.* 2008; Mellit y Kalogirou 2008).

En la presente tesis doctoral se selecciona el uso del MLP para experimentar con la predicción a corto plazo de la IS y probar la segunda hipótesis planteada. Esta selección se basa en las investigaciones de Voyant *et al.* (2014), Wang *et al.* (2012) y Mellit y Pavan (2010), que focalizan directamente en la predicción a corto plazo de la IS, utilizando este tipo de RNA para sus modelos. A continuación se destacan las características más relevantes de estas tres investigaciones como trabajos relacionados con la presente tesis:

1) Para predecir a corto plazo la IS, sólo utilizan datos de su misma ubicación geográfica.

2) Sobre los niveles de error:

Para cuantificar los niveles de error en la predicción de la ISG, los autores trabajan con distintos estadísticos, Error Medio Cuadrático Normalizado (*Normalized Root Mean Square Error: nRMSE*), Error Medio Absoluto Porcentual (*Percentage Mean Absolute Error: %MAE<sup>1</sup>*), Error Medio Cuadrático (*Root Mean Square Error: RMSE*) y Error Medio Cuadrático Porcentual (*Percentage Root Mean Square Error: %RMSE*). Algunos autores diferencian estos estadísticos entre estaciones climáticas (primavera, verano, otoño e invierno) y otros solo diferencian días claros y nublados. En la siguiente tabla se resumen las discrepancias más relevantes y algunos estadísticos calculados entre los valores reales y estimados de la ISG.

---

<sup>1</sup> También conocido como MAPE

Autores	Estadísticos	Periodo de cálculos	Nivel de agregación	Niveles de errores
Voyant <i>et al.</i> (2014)	nRMSE	Estimaciones en las horas de luz.	promedios	(27.3 %)
			verano	21.7 %
			invierno	42.4 %
Wang <i>et al.</i> (2012)	RMSE	Estimaciones para las 24 horas.	promedios	63.47
			claros	42.29 W/m <sup>2</sup>
			nubosos	84.65 W/m <sup>2</sup>
Mellit y Pavan (2010)	nRMSE	Estimaciones en las horas de luz.	promedios	(62.5 %)
			claros	58 % (promedio).
			nubosos	67 %

Tabla 1. Niveles de errores en trabajos relacionados.

3) Variables de entradas y las salidas: Los tres autores estiman la IS con intervalos temporales de 1 hora, utilizando variables de entrada endógenas: valores previos, valores medios, la derivada de tercer orden del valor máximo, diferencias discretas normalizadas de la ISG y, en algunos casos, usan variables exógenas: presión atmosférica, nubosidad, temperatura o el nº del día en el año. En todos los casos los métodos utilizan estos valores en una ventana de tiempo deslizante conforme el tiempo avanza.

## 2.5 Resumen

El estado de la cuestión presentado en este capítulo ha servido como base fundamental para la identificación de los aportes a la ciencia que se pueden hacer en materias tanto de la estimación espacial en intervalos de 15 minutos de la IS, como de su predicción temporal a corto plazo. También ha permitido identificar las posibles aplicaciones prácticas que ésta investigación puede llegar a tener.

Así en la sección 2.1, se han evidenciado la importancia de la disponibilidad de datos meteorológicos y de la Irradiancia Solar observada por estaciones en superficie en España, indicándose cómo ha evolucionado la disposición de esta información. También cómo los avances tecnológicos y la web han favorecido que se pueda acceder remotamente a estos datos de un modo ágil utilizando internet, lo

que facilita que ya no sólo las agencias oficiales puedan ofrecer información meteorológica sino que los propios usuarios pueden asumir el rol de productor. Seguidamente en la sección 2.2, se describió la Radiación Solar y sus componentes, como base para poder tratar la estimación espacial y la predicción temporal de la Irradiancia Solar.

La sección 2.3, trata sobre los métodos de estimación espacial de la Irradiancia Solar en la superficie de la Tierra, haciendo énfasis en los métodos de interpolación espacial y los sensores remotos en intervalos de tiempo inferiores a una hora, indicándose las investigaciones realizadas con ellos. Los estudios relacionados indican que a partir de distancias de interpolación de observaciones realizadas por estaciones en superficie superiores a 25 km, los métodos de interpolación presentan errores medios cuadráticos superiores al 25 % y que por lo tanto, a partir de este umbral es recomendable el uso de sensores remotos para la obtención de superficies de Irradiancia Solar en intervalos temporales inferiores a una hora.

En la sección 2.4 se presentan las investigaciones relacionadas con la predicción temporal de la Irradiancia Solar a corto plazo utilizando Redes Neuronales Artificiales, a la vez que se hace una introducción a esta técnica de modelización propia de la Inteligencia Artificial y el Aprendizaje Automático. Se han encontrado varios trabajos de investigación relacionados con la predicción de la Irradiancia Solar a corto plazo utilizando Redes Neuronales Artificiales, sin embargo se considera que la consistencia y robustez de estos modelos aun pueden ser mejorados si se analizan los errores de los distintos horizontes temporales de predicción por separado y si se incluyen otras variables de entrada endógenas que den soporte a la componente espacial como es la ISG en estaciones vecinas.



### **3 EVALUACIÓN DE LA DISPONIBILIDAD DE DATOS METEOROLÓGICOS EN TIEMPO CUASI-REAL EN ESPAÑA**

Este capítulo de la tesis presenta el estudio realizado sobre la disponibilidad de datos meteorológicos, en tiempo cuasi real, en las múltiples redes de estaciones instaladas en España hasta el año 2012. Los objetivos han sido: (i) chequear la disponibilidad de observaciones meteorológicas obtenidas por estaciones automáticas en España, (ii) caracterizar las redes identificadas de forma cualitativa y cuantitativa, (iii) llevar a cabo análisis espaciales y temporales de la disponibilidad de observaciones colectadas por tres redes de estaciones con distintas características [dos Gubernamentales (GUB): la Agencia Estatal de Meteorología (AEMet) con una cobertura a nivel nacional, la red InfoRiego de la junta de Castilla y León (CASTILLA Y LEÓN) a nivel regional, y una tercera REMV que opera a nivel nacional: Meteoclimatic], y (iv) llevar a cabo el análisis de redundancia y lagunaridad espacial de las estaciones en el territorio.

Los análisis realizados en este capítulo de la tesis, han estado dirigidos a responder las siguientes preguntas de investigación:

- ¿Existen Estaciones Meteorológicas (EM) que ofrecen datos de forma abierta en el territorio español en más de una red de estaciones?
- ¿Cómo están geográficamente distribuidas las REM en España?

#### **3.1 Metodología de evaluación de la disponibilidad de datos meteorológicos**

La metodología aplicada para este fin consta de 8 etapas. Algunas etapas se han llevado a cabo de forma manual, otras fueron automatizadas de manera total o parcial, describiéndose a continuación cada una de las etapas.

### **Descubrimiento de las URLs:**

Este proceso consistió en la exploración de los sitios webs de las agencias meteorológicas así como de las REMV, realizando búsquedas mediante términos meteorológicos claves y áreas geográficas. El objetivo fue encontrar los puntos de accesos a las distintas redes presentes en el área de interés.

### **Determinación de las localizaciones de las estaciones que conforman la red:**

En esta etapa se exploran las web o directorios de archivos remotos identificados en el paso anterior, para encontrar el listado de las EM que la conforman. Con los listados se construye una Base de Datos Geográfica (GeoDB) para disponer de una visión espacial de todas las estaciones.

### **Registro de las características de las estaciones que conforman las redes:**

Al tratarse de redes heterogéneas, se observaron de forma independiente para poder identificar similitudes o diferencias entre sus características (protocolo, formato, periodicidad, latencia de entrega y tipo de tiempo) tanto de las redes como de sus estaciones.

### **Análisis de las redes y estaciones meteorológicas:**

Una vez registradas las características de cada red, se analizaron estadísticamente para determinar sus relevancias relativas (porcentajes de cada característica).

### **Recolección periódica de las observaciones meteorológicas:**

Basado en la información de las etapas anteriores, el siguiente hito consistió en acceder a las observaciones para almacenarlas en la GeoDB.

Se seleccionaron las EM de las redes de AEMet, JCYL y Meteoclimatic como prueba de concepto. AEMet contiene la principal red de datos meteorológicos

gubernamentales de España. Meteoclimatic es la red principal basada EM de voluntarios. Mientras que JCYL es la red regional de la comunidad autónoma de Castilla y León. Esta fue escogida, sobre las otras redes regionales oficiales, porque permite recuperar sus observaciones históricas y por presentar ciertas similitudes con la red de AEMet en lo que respecta al protocolo de acceso y formato de datos, lo que supone ahorro de tiempo y esfuerzo en la programación de los robots web recolectores de las observaciones de ambas redes. Además, estas tres REM tienen distintas características y pueden ser consideradas representativas. Por ejemplo: dos son redes gubernamentales y una es de voluntarios, dos son redes con un alcance geográfico nacional y una a nivel de comunidad autónoma, dos tienen acceso FTP y sus datos están en formato CSV y la tercera es de acceso HTTP y formato de datos HTML.

La figura 6 ilustra la arquitectura del sistema implementado. Esta arquitectura está diseñada para ser escalable, mediante un controlador (engranaje grande) que permitirá agregar futuros recolectores (robots web) para otras redes. A la derecha aparecen las tres redes. A la izquierda, como engranajes pequeños, se representan los recolectores. Cada recolector está especializado en las características propias e identificadas de cada red.

### **Análisis de las observaciones recolectadas por las 3 redes:**

Una vez recolectadas las observaciones registradas por las EM, se procedió a determinar la periodicidad de sus observaciones en régimen estacionario y en función de la cantidad de observaciones reportadas diariamente por cada estación. La moda de ésta cantidad es el estadístico elegido para conocer la periodicidad de las estaciones. La desviación estándar de las modas indica lo estables que son las EM respecto al análisis de sus observaciones.

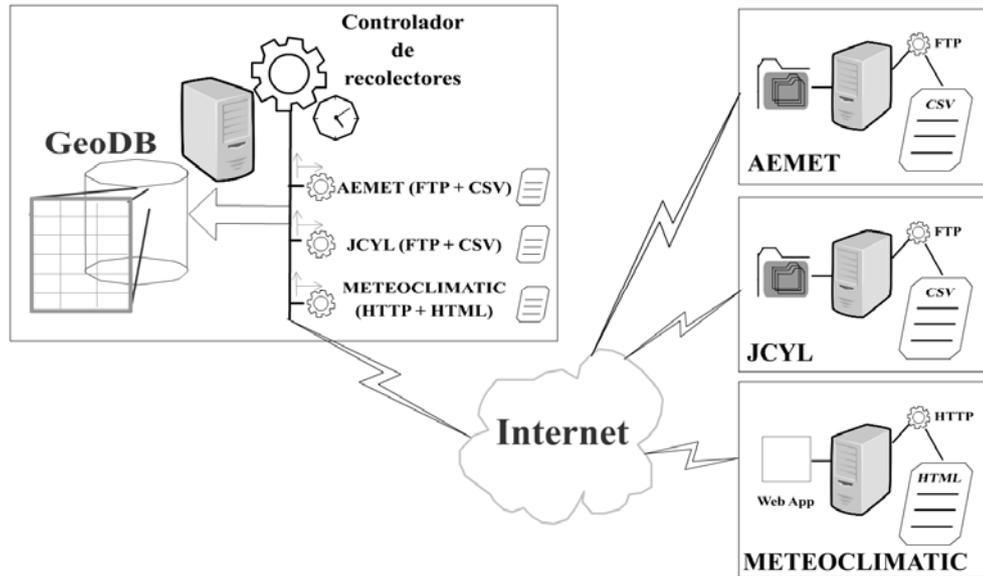


Figura 6. Ilustración simplificada de la arquitectura de sistemas para la recolección y almacenamiento de las observaciones meteorológicas desde distintas fuentes.

### **Análisis de redundancia:**

Las estaciones que se encuentran muy próximas entre sí fueron identificadas utilizando consultas espaciales de tipo área de influencia (*buffer*) con un radio de 100 m en torno a cada estación. También se han explorado otros valores para los *buffers* (150, 200 y 250 m) con el objetivo de detectar redundancia de EM. Para las REMV se asumió que todas las estaciones contenidas dentro de un mismo *buffer* eran redundantes. Esto también ayuda a evitar los errores propios de los redondeos o las transformaciones/conversiones de coordenadas para estas estaciones. Con el fin de mejorar la detección de estaciones redundantes entre las redes Gubernamentales, además del análisis espacial (*buffer* de 100 m), se compararon los parámetros registrados y los valores de sus observaciones en los mismos intervalos de tiempo.

### **Análisis de lagunaridad de estaciones:**

La lagunaridad es un método que permite describir patrones geográficos de dispersión espacial (Plotnick *et al.* 1996). En ésta fase se realiza un análisis de regiones sin observaciones (lagunaridad). El área geográfica de estudio fue

dividida en celdas de dimensiones proporcionales al largo y ancho del territorio (en este caso la península ibérica), con la condición de que cada celda tuviera un área equivalente a la relación existente entre la superficie total de España y la cantidad de estaciones.

### **3.2 Discusión de los Resultados**

En España existen varias redes meteorológicas, cada una de ellas agrupa un conjunto de estaciones de acuerdo a su área de influencia. Se han identificado 3908 estaciones meteorológicas desigualmente repartidas (mayor concentración en las grandes ciudades, levante y territorio insular). Estos resultados incluyen una vista general de las REM en España, los que a continuación se presentan en tres partes. Primero, se muestran los resultados cuantitativos y la interpretación de los valores agregados de las redes y sus estaciones según sus características de interés del estudio. A continuación se presentan los resultados del análisis de las observaciones de las tres redes (AEMet, Junta de Castilla y León (JCYL) Meteoclimatic). Finalmente se presentan los resultados de los análisis de redundancia y lagunaridad. La figura 7 muestra una visión espacial de las estaciones identificadas.

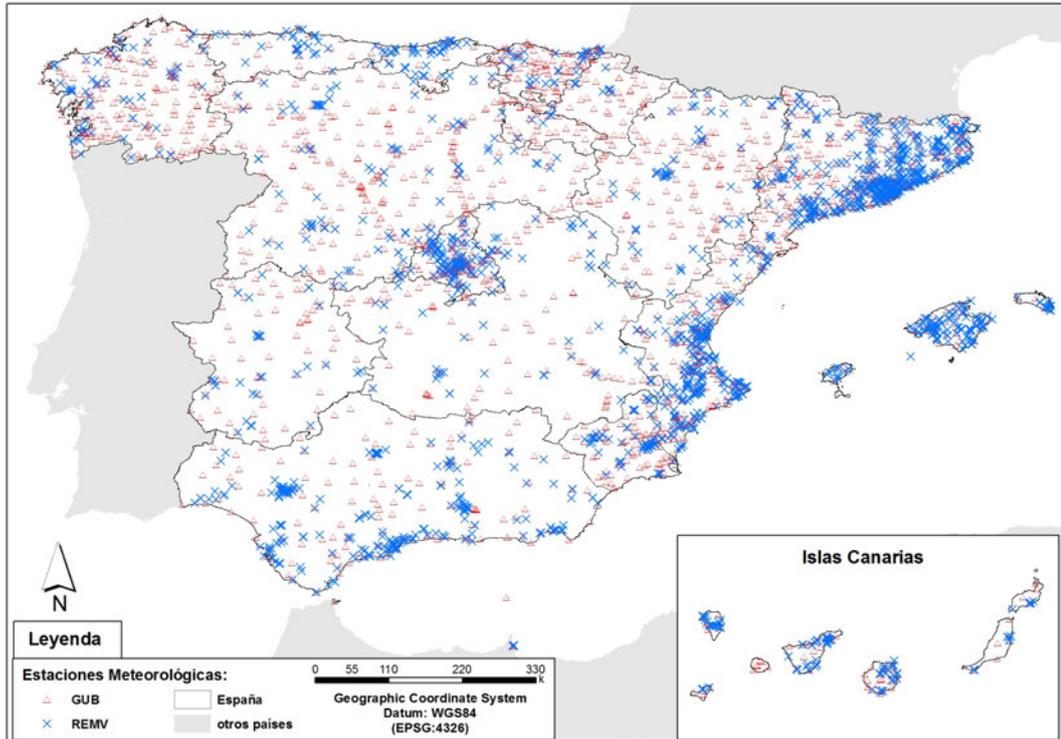


Figura 7. Distribución espacial de las redes y estaciones meteorológicas identificadas.

### Inventario de las REM en España y sus características:

En total, se encontraron 24 redes de información meteorológica que libremente comparten sus datos en internet, con un total de 3908 estaciones con características variadas. En la tabla 2 se presentan estas redes indicándose sus tamaños relativos en términos del número de estaciones. En el anexo 1 se listan los accesos web para cada una de las redes y se describe brevemente la afiliación de cada red.

<i>Red</i>	<i>Número de estaciones</i>		<i>Red</i>	<i>Número de estaciones</i>	
Meteoclimatic	1'344	(34.90 %)	ARAGÓN	48	(1.22 %)
AEMet	706	(19.44 %)	MEDITERRANEO-VALENCIANO	45	(1.15 %)
AWEKAS	366	(9.36 %)	MURCIA	44	(1.12 %)
WUNDER	304	(7.77 %)	EBRO	38	(0.97 %)
DGT	216	(5.52 %)	MUNICIPALIDAD DE MADRID	24	(0.61 %)
CATALUÑA	168	(4.29 %)	MADRID	24	(0.61 %)
CWOP	112	(2.86 %)	NAVARRA	19	(0.48 %)
WeatherLink	103	(2.63 %)	LA RIOJA	16	(0.40 %)

<i>Red</i>	<i>Número de estaciones</i>		<i>Red</i>	<i>Número de estaciones</i>	
<b>GALICIA</b>	85	(2.17 %)	<b>CASTILLA LA MANCHA</b>	12	(0.30 %)
<b>PAÍS VASCO</b>	63	(1.61 %)	<b>MeteoSal</b>	8	(0.20 %)
<b>VALENCIA</b>	52	(1.33 %)	<b>MeteoComillas</b>	5	(0.12 %)
<b>CASTILLA Y LEÓN</b>	51	(1.30 %)	<b>UNIMURCIA</b>	1	(0.02 %)

Tabla 2. Redes y volumen de estaciones.

Listado y caracterización de las redes por su naturaleza y área de influencia:

### **Redes gubernamentales:**

Fueron identificadas un total de 16 redes gubernamentales, 3 de las 16 disponen de una área de influencia mayor al de una comunidad autónoma: La Agencia Estatal de Meteorología (AEMet) distribuye sus estaciones por todo el territorio español, la Dirección General de Tráfico (DGT) sobre las carreteras españolas y la cuenca del Ebro (EBRO) sobre la geografía de dicha cuenca (véase tabla 24, anexo 1). 12 de las 16 son administradas por comunidades autónomas (ARAGÓN, CASTILLA LA MANCHA, CATALUÑA, MEDITERRANEO-VALENCIANO, MADRID, VALENCIA, PAÍS VASCO, GALICIA, LA RIOJA, CASTILLA Y LEÓN, MURCIA y NAVARRA, y la MUNICIPALIDAD DE MADRID. Esta última red presenta un alcance geográfico a nivel municipal (véase tabla 25, anexo 1).

### **Redes de voluntarios:**

Se han identificado 8 REMV y 5 de ellas poseen un área de influencia mayor o igual que el estado español (AWEKAS, CWOP, Meteoclimatic, WeatherLink, WUNDER), ver tabla 26 del anexo 1. Las tres redes restantes (MeteoComillas, MeteoSal, UNIMURCIA) poseen un área de influencia menor al estatal del tipo: comunidad autónoma, municipal y nivel local (véase tabla 27, anexo 1).

Interpretación de los resultados a nivel agregado:

**Tipo de Red:**

Ésta característica hace posible jerarquizar la confiabilidad de las observaciones, por los controles de calidad y validación que el responsable aplica a los instrumentos de observación. El 42.60 % (1665) de las estaciones son gubernamentales (GUB) y el 57.40 % (2243) son de REMV.

**Periodicidad:**

Cuantifica en término medio el tiempo entre dos observaciones consecutivas de la misma estación a nivel de red. El 74.84 % (2925) de las estaciones registran observaciones a intervalos de tiempo inferiores a la hora (sub-horas). De ellas, el 53.35 % (2085) son REMV y 21.49 % (840) son GUB. El 15.20 % (594) de las estaciones registran sus observaciones cada 30 minutos, de estas el 1.80% (70) son REMV y el 13.40 % (524) son GUB. El 7.49 % (293) registran datos cada hora, siendo el 2.25 % (88) REMV y el 5.24 % (205) GUB. El restante 2.45 % (96) registran sus observaciones diario, siendo todas de tipo GUB. En el caso de las 3 redes que se les ha analizado sus observaciones (AEMet, JCYL y Meteoclimatic) la periodicidad fue calculada en base a la cantidad de observaciones diarias reportadas por cada estación. La figura 8(a) resume los valores anteriores.

**Latencia de entrega:**

Con independencia de la periodicidad, muchos factores pueden causar retraso en la publicación de sus observaciones. Con el propósito de simplificar las estadísticas asociadas a la latencia de entrega, se han realizado 3 grupos/categorías similares a los de la periodicidad. La categoría de sub-horas agrupa a todas aquellas estaciones que entregan sus observaciones con latencia de entre 0 y 59 minutos. En esta categoría se encuentran el 73.02 % de las estaciones (2854), de las cuales el 15.63 % (611) son GUB y el 57.39 % (2243) son todas las estaciones REMV. Las restantes de las categorías son solo estaciones GUB. El 23.51 % (903) reportan con latencia de “hora” y el restante 3.86 % (151) con latencias de “día anterior”. En la figura 8(b) se resumen los valores anteriores.

### **Formatos de datos.**

Se han identificado 7 formatos de datos diferentes. En muchos casos las redes ofrecen más de un formato. Predominan las páginas webs y por lo tanto el formato HTML, usado por todas las redes excepto por CASTILLA Y LEÓN (GUB) y sus 51 estaciones, las 23 redes restantes representan el 98.69 % (3857) de las estaciones. El segundo formato más popular es CSV (valores separados por coma), con un 32.29 % (1262) de estaciones, de las que 10.64 % (416) son REMV y el 21.64 % (846) son GUB. Un 34.39 % de las estaciones (1344) ofrecen sus datos en formato RSS (XML de sindicación), correspondiéndose exclusivamente a la red Meteoclimatic. En formato Flash es usado por la red AWECAS (REMV), representando un 9.35 % (366) estaciones. En formato txt y de hojas de cálculos “Spread Sheet” tan solo lo ofrecen un 2.17 % (96) y 0.89 % (35) de las estaciones respectivamente, principalmente la red GALICIA (GUB). La figura 8(c) resume los valores anteriores.

### **Protocolos de acceso:**

Se han identificado 4 protocolos de acceso a los datos (HTTP\_GET, HTTP\_POST, SOA y FTP). Al igual que ocurre con el HTML en los formatos de datos, la mayoría de las redes orientan sus protocolos al acceso de usuarios. Sólo algunas redes están orientadas al acceso por software, permitiendo invocar servicios webs y obtener como respuesta documentos XML. La mayoría de las redes (todas las REMV y 13 de las 16 GUB) utilizan HTTP\_GET, lo que representa el 93.16 % (3641) de las estaciones repartido como 57.39 % (2243) VVO y 35.77 % (1398) GUB. El segundo protocolo de acceso, por su cantidad, es SOA con el 49.98 % (1836) mayoritariamente de tipo REMV 44.80 % y el restante 2.17 % de una única red GUB. El 20.75 % (811) ofrecen acceso mediante el protocolo FTP, en su totalidad son estaciones de redes GUB (AEMet y CASTILLA Y LEÓN). Tan solo el 5.52 % de las estaciones ofrecen acceso con el método HTTP\_POST pertenecientes a dos redes GUB (ARAGÓN y CATALUÑA). En la figura 8(d) se resumen los valores anteriores.

La figura 8 muestra diagramas de barras anidadas con los valores de las 3 últimas características descritas. En todos los casos la barra blanca corresponde al total de la categoría sin importar si es GUB o REMV. Las barras interiores de color negro representan el total de la categoría de tipo GUB y la barra de color gris el total de la categoría de tipo REMV.

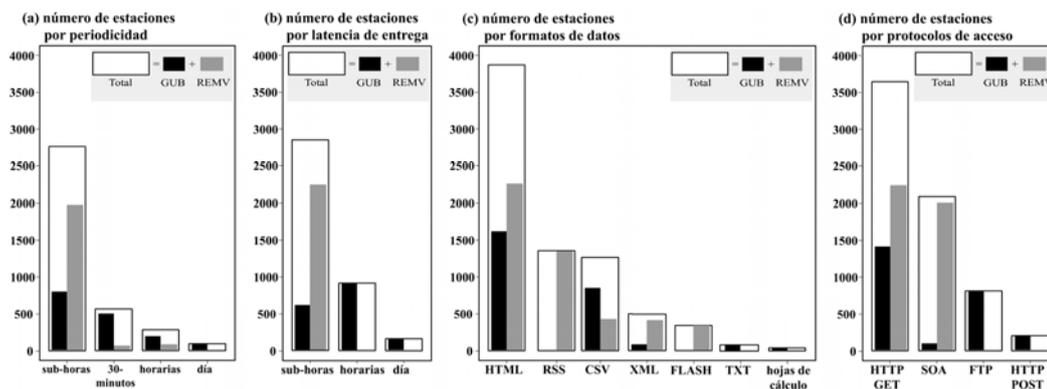


Figura 8. Número de estaciones según sus características. Gráfico de barras anidadas por: (a) periodicidad, (b) latencia de entrega, (c) formato de datos, y (d) protocolo de acceso.

### Tipo de Estaciones:

Las estaciones pueden ser de dos tipos: automática o manual, siendo la mayoría de tipo automático.

### Tipo de Tiempo:

Para interpretar correctamente las observaciones en el tiempo, se necesita conocer su sistema de referencia temporal. Se ha identificado el uso de: (i) Tiempo Universal Coordinado (*Coordinated Universal Time: UTC*) y (ii) Local, tiempo dependiente del país, zona horaria o región. En el caso de estudio (España) existen dos husos horarios que cambian 2 veces al año.

Se han seleccionado, para el estudio las redes: AEMet, CASTILLA Y LEÓN, Meteoclimatic, por la disponibilidad en tiempo real de sus observaciones así como por poseer características diferenciadas entre sí y por ser representativas de las 24 redes identificadas. Estas tres redes fueron seleccionadas con el propósito de

analizar la disponibilidad de sus observaciones en tiempo cuasi-real. Se enfatizó en su característica cuasi-real, por recoger y almacenar la información de forma previa a su publicación, hecho que introduce un retardo significativo en la disponibilidad de la información, ofreciendo a cambio un histórico más o menos extenso de observaciones (desde los 7 días de AEMet, toda la historia CASTILLA Y LEÓN, o sin histórico en Meteoclimatic).

### **Observaciones recolectadas desde sus fuentes:**

La recolección de observaciones de cada red se inició en fechas diferentes. En el caso de la red CASTILLA Y LEÓN desde Enero de 2009 hasta la actualidad con 51 estaciones. Para Meteoclimatic en Mayo de 2010 registrándose desde entonces 1344 estaciones (a pesar de que la red informa de la existencia de aproximadamente 1700 estaciones), esta varianza es debida al hecho de que Meteoclimatic es una red de estaciones dinámica y el sitio web indica el número total de estaciones registradas, estén o no estén publicando datos actualmente. En el caso del presente estudio, sólo aquellas estaciones de las que se recolectaron datos durante este periodo fueron incluidas en la base de datos. Además, esta red incluye estaciones localizadas tanto en Portugal como en Francia, que caen fuera del alcance del presente estudio. AEMet por dos años desde Noviembre de 2010, fecha en que esta red empezó a compartir sus observaciones de forma abierta, hasta que se volvió a cerrar el acceso abierto de estos datos.

En la figura 9 se presentan tres gráficas (una para cada red) en las que se cuantifican las estaciones que han registrado observaciones con distinta periodicidad a lo largo del tiempo (cantidad de observaciones al día por estación). Las curvas punteadas de color negro cuantifican el número de estaciones con observaciones para cada día. Las líneas de color rojo cuantifican el número de estaciones que han ofrecido más de 48 observaciones al día o más de 2 observaciones por hora (sub-horas), las líneas de color azul cuantifican el nº de estaciones con una cantidad de observaciones mayor a 24 pero menor igual a 48 (30-minutales) y finalmente las líneas de color verde cuantifican el nº de estaciones que han entregado al menos 24 observaciones al día (horarias).

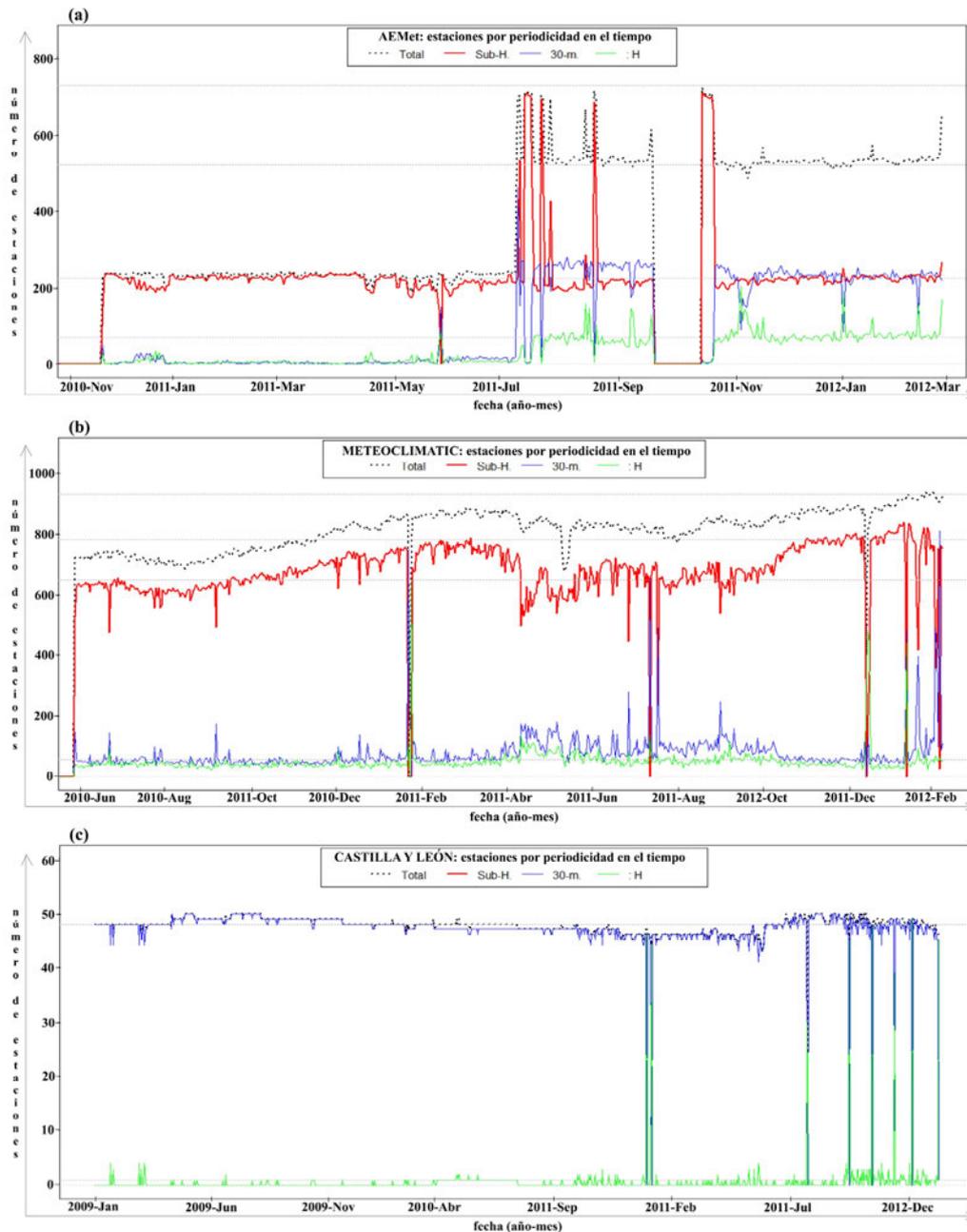


Figura 9. Comportamiento de la cantidad de observaciones que reportaron diariamente las estaciones: (a) AEMet, (b) Meteoclimatic, (c) JCYL, a lo largo del tiempo en función a la cantidad de observaciones que registró diariamente cada estación.

En el caso de AEMet (figura 9 (a)), se aprecia que a mediados de 2011 se incrementó considerablemente la cantidad de estaciones, pasando de 200 aproximadamente hasta casi 800. En el caso de Meteoclimatic (figura 9 (b)) se muestra la tendencia de la variación en el n° de estaciones activas incrementando su n° desde las 750 hasta las 930 (180 estaciones más en casi 2 años), de las cuales

130 ofrecen observaciones en periodos de tiempo inferiores a la hora. En el caso de la red de CASTILLA Y LEÓN (figura 9 (c)) se puede apreciar la estabilidad tanto en nº de estaciones como en los periodos de observación (30-minutales).

También se ha realizado un análisis agregado de estos datos. En la tabla 3 y en la figura 10 se muestran el promedio porcentual de la cantidad de observaciones diarias registradas por las estaciones, agrupadas por red y por tres niveles de temporalidad. En ambos casos se puede apreciar que la red Meteoclimatic ha registrado más observaciones sub-horas (83 %), mientras que AEMet un (62 %).

<i>RED/ Periodo:</i>	<i>AEMet</i>	<i>Meteoclimatic</i>	<i>JCYL</i>
Sub-Hora			
[# Obs. X Día] > 48	62 %	83 %	0 %
30-Minutos			
24 < [# Obs. X Día] <= 48	29 %	10 %	100 %
Hora			
[# Obs. X Día] <= 24	9 %	7 %	0 %

Tabla 3 Promedio porcentual de la cantidad de observaciones diarias agrupadas por red y periodicidad.

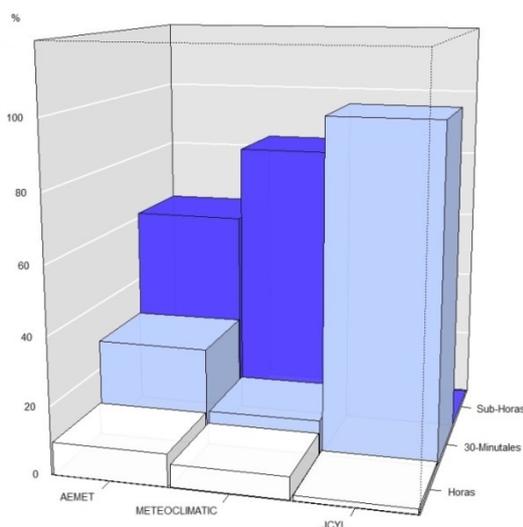


Figura 10. Diagrama de barras en 3D con el promedio porcentual de la cantidad de observaciones diarias. Barras agrupadas por red y periodicidad: Sub-horas [barras más oscuras], 30-minutales [barras grises] y Horarias [barras blancas]

### Frecuencia de las variables meteorológicas:

Las tablas 4 y 5 muestran las variables meteorológicas observadas por cada una de las tres redes, así como el número de estaciones que incluyen dichas variables. Se cuantificaron un total de 16 variables: Temperatura (T), T Máxima, T Mínima, Humedad Relativa (HR), HR Máxima, HR Mínima, Presión (P), P Máxima, P Mínima, Precipitación, Irradiancia Solar (IS), Velocidad del Viento (VV), Dirección del viento (DV), Velocidad Máxima del Viento, Velocidad Mínima del Viento y Visibilidad. Como se puede observar no todas las estaciones observan todas las variables. La tabla 4 indica sólo aquellas variables que son observadas por todas las estaciones en al menos una de las redes. En esta tabla es posible notar cómo el porcentaje de estaciones de acuerdo a su frecuencia es equivalente a los porcentajes de las redes mostrados en la tabla 3. La tabla 5 lista aquellas variables que han sido observadas por algunas de las estaciones.

<i>VARIABLE</i>	<i>AEMet</i> <i>Número de estaciones</i>	<i>CASTILLA Y LEÓN</i> <i>Número de estaciones</i>	<i>Meteoclimatic</i> <i>Número de estaciones</i>
Estaciones totales	760 (100 %)	51 (100 %)	1344 (100%)
Temperatura (T)	760 (100 %)	51 (100 %)	1344 (100%)
T máxima	760 (100 %)	- -	1344 (100%)
T mínima	760 (100 %)	- -	1344 (100%)
Humedad relativa (RH)	760 (100 %)	51 (100 %)	1344 (100%)
HR máxima	- -	- -	1344 (100%)
RH mínima	- -	- -	1344 (100%)
Presión máxima	- -	- -	1344 (100%)
Presión mínima	- -	- -	1344 (100%)

Tabla 4. Variables meteorológicas incluidas en todas las estaciones.

<i>VARIABLE</i>	<i>AEMet</i> <i>Número de estaciones</i>	<i>CASTILLA Y LEÓN</i> <i>Número de estaciones</i>	<i>Meteoclimatic</i> <i>Número de estaciones</i>
Estaciones totales	760 (100 %)	51 (100 %)	1'344 (100%)
<b>Presión (P)</b>	169 (23 %)	- -	1'344 (100%)
[# obs. x día] > 48	167 (22 %)	- -	1'121 (83 %)
24 < [# obs. x día] <= 48	1 (0.1 %)	- -	134 (10 %)
[# obs. x día] <= 24	1 (0.1 %)	- -	89 (7 %)
<b>Precipitación</b>	760 (100 %)	51 (100 %)	1'333 (99 %)
[# obs. x día] > 48	468 (62 %)	51 (100 %)	1'177 (88 %)
24 < [# obs. x día] <= 48	219 (29 %)	0 (0 %)	70 (5 %)
[# obs. x día] <= 24	73 (9 %)	0 (0 %)	86 (7 %)
<b>Irradiancia solar</b>	25 (3 %)	51 (100 %)	229 (17 %)
[# obs. x día] > 48	25 (3 %)	51 (100 %)	198 (15 %)
24 < [# obs. x día] <= 48	0 (0 %)	- -	12 (1 %)
[# obs. x día] <= 24	0 (0 %)	- -	9 (1 %)
<b>Dirección del viento</b>	665 (88 %)	51 (100 %)	1'340 (99.3 %)
[# obs. x día] > 48	391 (52 %)	- -	1'183 (83 %)
24 < [# obs. x día] <= 48	183 (24 %)	- -	70 (10 %)
[# obs. x día] <= 24	91 (12 %)	- -	87 (7 %)
<b>Velocidad del viento</b>	665 (88 %)	51 (100 %)	1'335 (99 %)
[# obs. x día] > 48	391 (52 %)	- -	1'116 (83 %)
24 < [# obs. x día] <= 48	183 (24 %)	- -	134 (10 %)
[# obs. x día] <= 24	91 (12 %)	- -	86 (7 %)
<b>Viento máximo</b>	665 (88 %)	- -	1'341 (99 %)
[# obs. x día] > 48	391 (52 %)	- -	1'183 (89 %)
24 < [# obs. x día] <= 48	183 (24 %)	- -	70 (5 %)
[# obs. x día] <= 24	91 (12 %)	- -	88 (7 %)
<b>Viento mínimo</b>	665 (88 %)	- -	- -
[# obs. x día] > 48	391 (52 %)	- -	- -
24 < [# obs. x día] <= 48	183 (24 %)	- -	- -
[# obs. x día] <= 24	91 (12 %)	- -	- -
<b>Visibilidad</b>	86 (11 %)	- -	- -
[# obs. x día] > 48	86 (11 %)	- -	- -
24 < [# obs. x día] <= 48	0 (0 %)	- -	- -
[# obs. x día] <= 24	0 (0 %)	- -	- -

Tabla 5 Variables meteorológicas observadas parcialmente por las estaciones.

Al compararse ambas tablas se puede concluir que la red de CASTILLA Y LEÓN ofrece el menor número de variables (7 de 16) y que esta red no registra los valores agregados mínimos ni máximos para ninguna de las variables, tampoco incluye la Presión atmosférica ni la Visibilidad. Además, todas las estaciones de esta red

registran el mismo número de variables (la red es homogénea). La red Meteoclimatic registra 14 de las 16 variables (excepto Velocidad mínima del Viento y la Visibilidad), también se observa que la mayoría de las variables en esta red están incluidas en el 99 % de sus estaciones, con excepción de la Irradiancia Solar que sólo está incluida en el 17 % de las estaciones. AEMet registra 12 de las 16 variables, excluyendo los valores máximos y mínimos para la Presión y la Humedad Relativa. En el caso de AEMet, 5 de las 12 variables son registradas por todas las estaciones de esta red, mientras que más del 88 % de las estaciones registran otras 4 variables relacionadas al viento y un menor porcentaje registran la Presión atmosférica, la Visibilidad y la Irradiancia Solar. Un detalle notorio es que aquellas estaciones de AEMet que registran más de 48 observaciones al día también registran la Presión y la Visibilidad.

#### **Principales características de la periodicidad de las estaciones:**

Con el propósito de cuantificar la disponibilidad de observaciones en tiempo real de cada estación, se analizó estadísticamente la cantidad de observaciones diarias registradas por cada estación (moda, media y desviación estándar (stdv)).

En la figura 11 se presentan los diagramas de dispersión, con los respectivos estadísticos, para las observaciones de 6 estaciones representativas de las 3 redes en el número promedio de sus observaciones diarias. Las redes y estaciones seleccionadas para esta ilustración, han sido aquellas que adquieren sus observaciones en frecuencias irregulares, es por esta razón que las observaciones hechas por la red de CASTILLA Y LEÓN han sido excluidas. La columna izquierda son 3 estaciones de AEMet y en la columna derecha 3 estaciones de Meteoclimatic.

En las dos primeras gráficas de cada red (figuras 11. a, b)) se puede apreciar que las estaciones han registrado más de 48 observaciones por día (sub-horas). Las figuras 11 (c) y (d) dos casos de estaciones a 30 minutos. Finalmente las figuras 6 (e) y (f) presentan a dos estaciones horarias.

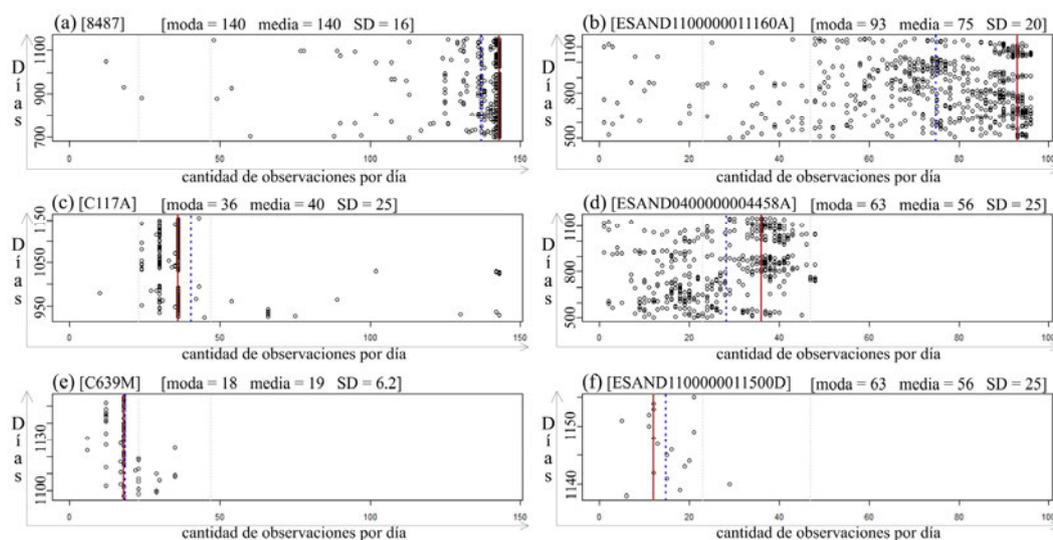


Figura 11. Diagrama de dispersión de la cantidad de observaciones diarias para seis estaciones.

Principales características de la periodicidad de observaciones a nivel de red:

Con el propósito de describir la periodicidad de las estaciones de cada red, se analizaron estadísticamente las modas, el promedio y la desviación estándar agregada de las observaciones registradas diariamente por sus estaciones. Además se calculó el porcentaje de estaciones cuyas modas o promedios están comprendidos en los intervalos de periodicidad definidos: sub-hora ([modas o medias] > 48), 30-minutales ( $24 < [\text{modas o medias}] \leq 48$ ) y horarias (modas o media] <= 24). En las tablas 6 y 7 se presentan los resultados de este análisis.

<i>Red</i>	<i>Moda</i>	<i>Media</i>	<i>Stdv</i>	<i>Estaciones</i>
<b>AEMet</b>	140	93	57	760
<b>Meteoclimatic</b>	92	81	24	1344
<b>CASTILLA Y LEÓN</b>	47	47	0	51

Tabla 6. Media, Moda y desviación estándar de las modas de la cantidad de observaciones diarias por estación agrupadas por red.

<i>Red</i>	<i>Moda</i>	<i>Media</i>	<i>Stdv</i>	<i>Estaciones</i>
<i>AEMet</i>	83	75	40	760
<i>Meteoclimatic</i>	78	68	17	1344
<i>CASTILLA Y LEÓN</i>	47	46	2.7	51

Tabla 7. Media, Moda y desviación estándar de las medias de la cantidad de observaciones diarias por estación agrupadas por red.

También se realizó el análisis estadístico de los porcentajes de las modas y promedios de las estaciones de cada red, y se categorizaron en los siguientes intervalos: sub-hora, 30-minutales y horarias, presentándose los resultados en las tablas 8 y 9.

<i>Red</i>	<i>Sub-Hora</i>	<i>30-Minutos</i>	<i>Hora</i>
<i>AEMet</i>	56 %	31 %	13 %
<i>Meteoclimatic</i>	88 %	5 %	7 %
<i>CASTILLA Y LEÓN</i>	0 %	100 %	0 %

Tabla 8. Promedio porcentual de las modas de la cantidad de observaciones diarias por estación agrupadas por red.

<i>Red</i>	<i>Sub-Hora</i>	<i>30-Minutos</i>	<i>Hora</i>
<i>AEMet</i>	58 %	40 %	3 %
<i>Meteoclimatic</i>	87 %	9 %	4 %
<i>CASTILLA Y LEÓN</i>	0 %	100 %	0 %

Tabla 9. Promedio porcentual de las medias de la cantidad de observaciones diarias por estación agrupadas por red.

De las tablas 6 y 7 se aprecia que las estaciones de AEMet presentan mayores valores de desviación estándar, para las modas (57) y para las medias (40). Esto se debe a la mayor varianza entre ambos tipos de población. Lo anterior se aprecia de forma gráfica en la figura 12 (a) donde se observa una distribución (bi modal) de las modas. En esta figura se aprecia que el 44% de los puntos se concentra en torno al intervalo (1..48) de la moda, mientras que el 54% restante se concentra en torno

al valor 144. Las medias presentan una tendencia similar a las modas respecto al primer intervalo (1..48) para el 43 % de las medias, sin embargo el 58 % restante se agrupa en 2 grandes intervalos, el primero (75..95) y el segundo (110..140), como se muestra en las figuras 13 (a) y 13 (b).

En las estaciones de Meteoclimatic se aprecian una diferencia en torno a 10 unidades entre las medias y modas de las modas (92 y 81), y medias y modas de las medias (78 y 68). La desviación estándar adopta valores próximos al 20 (+-4), ya sea de la media de las medias (24) o la media de las modas (17) (ver tablas 6 y 7). Las figuras 12 (c y d) de las modas y 13 (c y d) de las medias presentan una distribución mono modal, observándose una concentración de los valores en la parte derecha de los diagramas que representa las estaciones sub-horas. La presencia de una cola larga en todos los casos pero más fina en el caso de las modas representa a las estaciones con periodicidad más gruesa.

Las estaciones de la CASTILLA Y LEÓN presentan una moda y media aproximada a 47 y una desviación estándar cercana a cero con una sola moda, como se presenta en las tablas 6 y 7 y se representa gráficamente en las figuras 12 (e, f) y 13 (e, f).

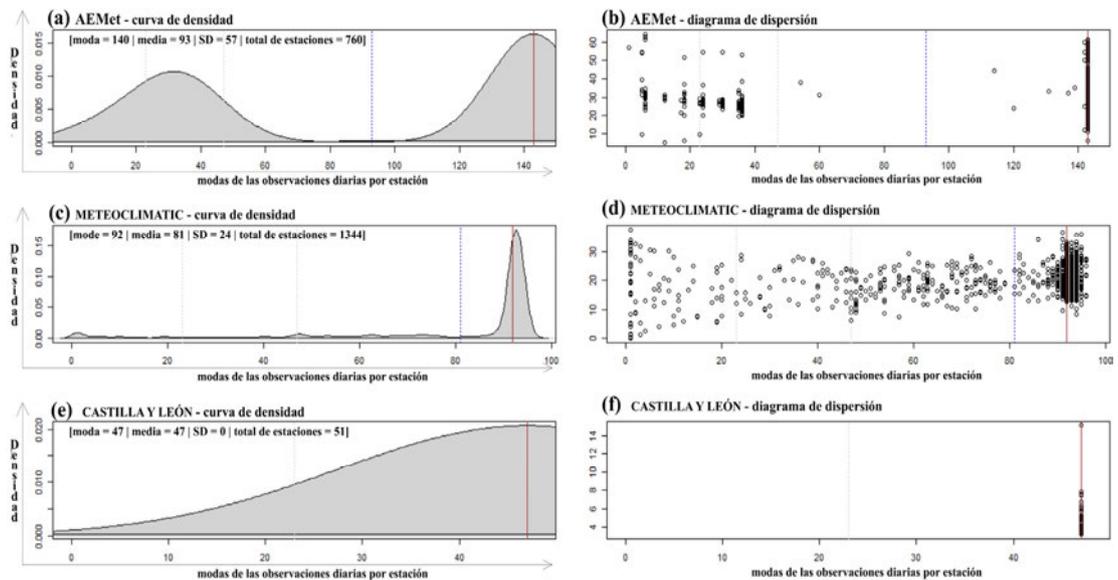


Figura 12 Diagramas estadísticos de las modas de las observaciones diarias por estación agrupadas por redes. Red AEMet (a) y (b). Red Meteoclimatic (c) y (d). Red CASTILLA Y LEÓN (e) y (f).

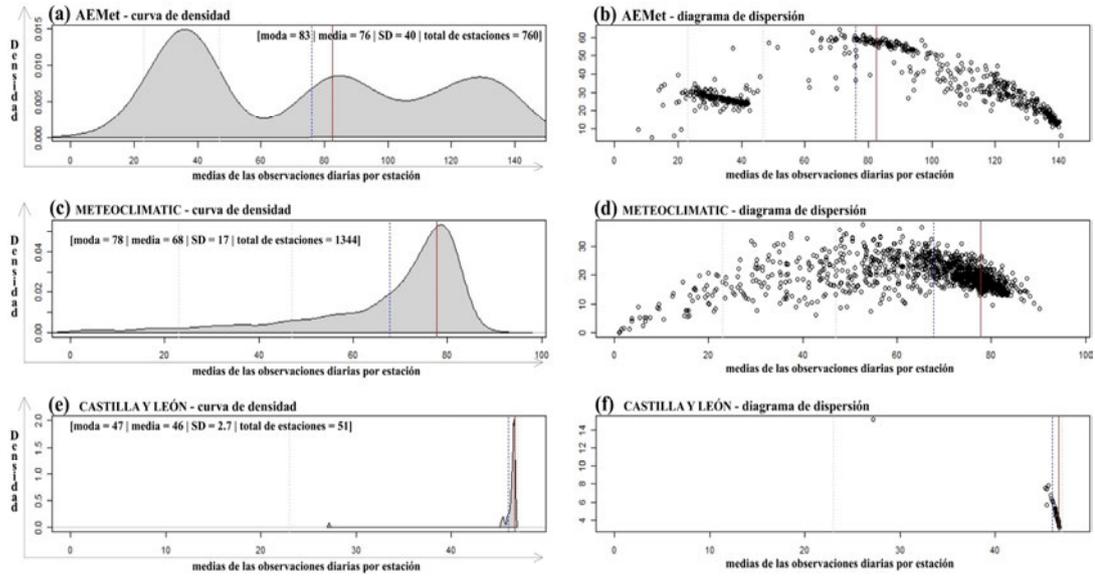


Figura 13. Diagramas estadísticos de las medias de las observaciones diarias por estación agrupadas por redes. Red AEMet (a) y (b). Red Meteoclimatic (c) y (d). Red CASTILLA Y LEÓN (e) y (f).

### Análisis espacial de redundancia y lagunaridad:

Al observarse el mapa en la figura 7, se evidencia que existen zonas en las que se concentran muchas estaciones. Un acercamiento en estas zonas permite observar la escasa distancia que separa algunas estaciones. Por este motivo y para detectar EM repetidas en distintas redes se ha realizado un análisis espacial. Por el contrario, también se puede identificar en la misma figura zonas sin estaciones (con lagunas) y por esta razón se ha realizado un análisis de lagunaridad.

### Análisis de redundancia:

El propósito del análisis de redundancia es aproximar el número de estaciones únicas existentes. Se ha utilizado el peor escenario para restar las estaciones duplicadas, es decir, si en los *buffer* existe más de una estación, se restó el exceso a uno de estaciones.

En la tabla 10 se presenta el número total de estaciones que caen dentro de las zonas de *buffer*. Aquí es posible notar que el número de estaciones incluidas en las zonas de *buffer* tiende a crecer cuando aumenta el tamaño de este, además se puede

apreciar que la diferencia entre el número estimado de estaciones GUB entre la zona de buffer mayor y menor es de 13, mientras que para las estaciones REMV es de 61. También se puede notar que en todos los casos (excepto uno en 250 m.), la columna para 6 estaciones en la zona de *buffer* reporta cero. Estas observaciones fueron utilizadas como criterios de selección de las distancias de zonas de *buffer* de 100, 150, 200 y 250 m sobre cada estación, y para evitar así incluir estaciones evidentemente distintas pero que coincidirían en buffer de mayor tamaño.

<i>Tamaño Buffer / Tipo estación</i>	<i>1 estación en buffer</i>	<i>2 estaciones en buffer</i>	<i>3 estaciones en buffer</i>	<i>4 estaciones en buffer</i>	<i>5 estaciones en buffer</i>	<i>6 estaciones en buffer</i>	<i>Estimaciones (peor caso)</i>
<b>100 m (GUB)</b>	1645	20	0	0	0	0	1655
<b>150 m (GUB)</b>	1632	32	3	0	0	0	1649
<b>200 m (GUB)</b>	1628	34	3	0	0	0	1646
<b>250 m (GUB)</b>	1620	42	3	0	0	0	1642
<b>100 m (REMV)</b>	1418	490	237	80	15	0	1765
<b>150 m (REMV)</b>	1384	504	243	96	15	0	1744
<b>200 m (REMV)</b>	1346	514	258	104	20	0	1719
<b>250 m (REMV)</b>	1322	524	270	104	20	6	1704

Tabla 10. Cantidad de estaciones potencialmente duplicadas por tipología (GUB / RMEV) basadas en 4 valores de Buffer.

Para verificar si las redundancias identificadas con este método son reales en las redes gubernamentales, se han comprobado los 10 casos posibles en el *buffer* a 100 m, y se han encontrado las siguientes situaciones:

En 9 casos son estaciones distintas a pesar de su proximidad, al comprobarse que registran valores diferentes en las observaciones para un mismo intervalo de tiempo y no todas captan las mismas variables meteorológicas. 7 de estos 9 casos son estaciones de la red AEMet. 6 de los 7 son parejas de estaciones que siguen un patrón en su codificación, una posee un código completamente numérico y el segundo un código finalizado en letras (por ejemplo 8353 y 5038Y). Se detectó que las segundas registran principalmente valores cada diez minutos aunque en intervalos irregulares de tiempo que por lo general son de 1 hora, precedido de un silencio de 3 horas. Sin embargo en fechas concretas todas ellas han registrado

datos sin silencio. Este indicio sugiere que dichas estaciones pueden estar configuradas para tener ese comportamiento. En el 7º caso ambas estaciones presentan códigos finalizados en letras, sin embargo la primera ha dejado de registrar observaciones a finales del 2011, seguidamente la segunda empezaba a registrar. Este hecho sugiere que se ha sustituido la estación antigua por una nueva.

El 8º caso (estación 8168 de AEMet y la estación 16 de ARAGÓN) no sigue los patrones anteriores y para el 9º caso (estación DGT 815 y la estación P01 de CASTILLA Y LEÓN) no se dispone de registros de observaciones para realizar este tipo de análisis.

El 10º caso corresponde a dos estaciones (001 y 031) de la red MEDITERRANEO-VALENCIANO. Al revisar su página web se encontró que la estación 001 ya no existe, y la identificada como 031 se localiza en las mismas coordenadas.

Aunque no se ha realizado una revisión exhaustiva de las estaciones REMV, se han encontrado algunos patrones en los nombres o identificadores de las estaciones que se localizan en una posición similar. En la red Meteoclimatic aparecen identificadores que se distinguen en el último dígito y suelen ser correlativos como por ejemplo: las estaciones con nombres: “Madrid-Parque Avenidas”, “Madrid-ciemat42”, “Madrid\_Ciemat42Lab” con códigos ESMAD2800000028040A, ESMAD2800000028040B, ESMAD2800000028040C. Estas mismas estaciones se parecen también a las estaciones de la red AWECAS con nombres “Madrid – Ciemat”, “Madrid-Ciemat42” y con códigos 7963, 8055 respectivamente. Estos y otros casos similares, sugieren que los propietarios de las estaciones REMV tratan de mantener la identidad de sus estaciones a pesar de suscribirlas en distintas redes. Esta consideración junto a la proximidad son criterios aplicables para identificar estaciones de REMV repetidas.

### **Análisis de lagunaridad:**

Tras la detección de estaciones repetidas (análisis de redundancia), se realizó el análisis de lagunaridad, excluyendo dichas estaciones. Se han unido las 1765 estaciones REMV diferentes arrojadas por el análisis sobre el *buffer* de 100 m con las 1664 estaciones de tipo GUB (también tras el análisis sobre el *buffer* de 100 m) y las 9 estaciones excluidas por el análisis que no estaban repetidas (comprobación anterior), sumando un total de 3429 estaciones en todo el territorio español. Como criterio de exclusión en las REMV se ha hecho prevalecer las estaciones de la red más densa (Meteoclimatic). Como consecuencia de este hecho se han excluido 479 estaciones consideradas redundantes para un cálculo de lagunaridad más acertado.

El análisis de lagunaridad se ha realizado sobre una malla de 3786 celdas. Malla generada con celdas proporcionales al Rectángulo Contenedor (BBox) del territorio peninsular español. La diferencia de 357 celdas respecto a la cantidad de estaciones, es equivalente a 10.41 % de unidades porcentuales y se debe a irregularidad del perímetro de las costas. A continuación se listan los parámetros utilizados:

- Superficie aprox. de España = 505000 km<sup>2</sup>.
- Área aproximada del BBox para el territorio peninsular español = 917777 km<sup>2</sup>.
- Proporción de incremento aproximado (Ancho / Alto) del BBox= 1.184. Este Ancho/Alto se calculó como la proporción entre la longitud incremental desde Galicia hasta Menorca y la latitud incremental desde Ceuta y Melilla hasta el País Vasco.
- Área aproximada de las celdas para una relación de una estación por cada celda = 147.27 km<sup>2</sup> (505000 km<sup>2</sup> / 3429 estaciones).
- Altura aproximada de cada celda = 11.15 km.
- Ancho aproximada de cada celda = 13.20 km.

El mapa de la figura 14 presenta la distribución de las celdas generadas y el resultado gráfico del análisis de lagunaridad. El 60.72 % de las celdas están vacías (lagunas). Existen celdas que solo contienen estaciones GUB (19.86 %), solamente

estaciones REMV (8.42 %) y celdas con ambos tipos de estaciones (11.01 %). También se encontró que en las grandes ciudades existe alta densidad de estaciones por celda, por ejemplo Madrid, Barcelona y Valencia contienen celdas de hasta 61, 51 y 31 estaciones respectivamente.

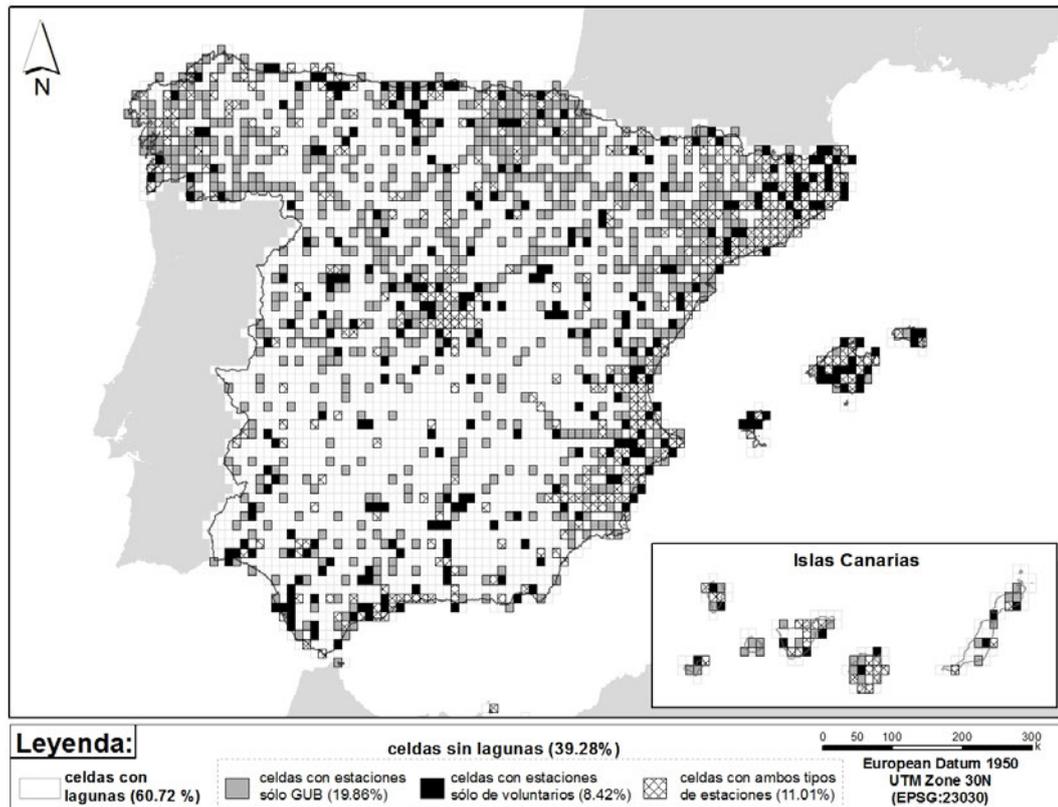


Figura 14. Mapa de lagunaridad de estaciones meteorológicas para el territorio español: Malla de 3786 celdas con dimensiones de 11.15 km de alto por 13.20 km de ancho.

### 3.3 Resumen

Esta parte de la investigación cubrió el objetivo de inventariar y caracterizar la disponibilidad y distribución geográfica de observaciones realizadas por REM automáticas en el territorio español que publiquen datos libremente en internet, así como el análisis de la disponibilidad de observaciones para tres REM con distintas características: dos de tipo gubernamental (AEMet con cobertura nacional y JCYL a nivel regional) y una no oficial (REMV) Meteoroclimatic (con cobertura nacional). La evaluación de la disponibilidad de datos meteorológicos en tiempo cuasi-real,

se realizó principalmente explorando de forma general la web para el descubrimiento de las Redes de Estaciones Meteorológicas (REM) España, después se exploró con detalle cada uno de los sitios de las REM a medida que se generaba una Base de Datos de cada una de sus estaciones donde se apuntaron las distintas características asociadas a los formatos y parámetros de accesos, formatos de datos y otros, que permitieron la creación de robots web para la descarga periódicas de los datos de estas redes. Finalmente se realizó un análisis de lagunaridad y redundancia, que permitió disponer de una mejor visión de la distribución de todas estas estaciones en el territorio español.

Los resultados han revelado la existencia de 24 REM automáticas desplegadas en España (16 GUB y 8 de voluntarios). Las REMV presentan características heterogéneas que pudieran ayudar a complementar a las GUB. Se presentan diferencias dentro y entre las distintas redes (GUB y de REMV) en cuanto a la frecuencia de adquisición de los datos, variación en los retrasos de las publicaciones, y diferencias semánticas en las observaciones realizadas. El análisis realizado sobre 3 redes ha revelado que el 88 % de las estaciones en la red de voluntarios graban más de 48 observaciones por día (más de dos observaciones por hora). El 56 % de las estaciones de AEMet analizadas reportan más de 48 observaciones por día, el 31 % reportan entre 25 y 48 observaciones al día (más de una pero a lo sumo 2 observaciones por hora). El análisis de redundancia revela la existencia de 3429 estaciones no duplicadas. El análisis de regularidad y de distribución espacial de las estaciones ha permitido concluir que 60.72 % de las celdas en que fue dividida el área de estudio, no contienen estaciones meteorológicas.

Los datos descargados por robots web desarrollados gracias a la evaluación de las REM, forman parte de los datos utilizados en los experimentos de los siguientes capítulos de la presente tesis doctoral.



## **4 ESTIMACIÓN ESPACIAL DE LA ISG**

El objetivo principal en este capítulo de la tesis es investigar la bondad de las estimaciones espaciales de la IS a intervalos de 15 minutos de cinco métodos en el territorio Peninsular español y el archipiélago Balear (3 métodos de interpolaciones geoestadísticos, uno determinista y HelioSat2 que se basa en el uso de imágenes de satélite) aplicados en 6 configuraciones espaciales de los datos, con el objetivo de analizar los efectos zonales (locales) de una mayor o menor densidad de EM. Como objetivo secundario, se propone la futura inclusión de las observaciones de la IS procedentes de los sensores de las EM de voluntarios realizándose una primera aproximación de un proceso de validación. Una vez validadas éstas estaciones auxiliarían en la densificación espacial y temporal de la información de la IS existente en España.

La comparación de las distintas técnicas de estimación espacial en las distintas formas de agrupación de los resultados tratados en el presente capítulo de la tesis, han estado dirigidos a responder las siguientes preguntas de investigación:

- ¿Es posible interpolar determinística o estadísticamente la IS en el territorio nacional a intervalos de 15 minutos más allá de los 25 km con errores inferiores al 25 %?
- ¿Están las observaciones de la IS de las EM de la red Meteoclimatic (voluntarios) dentro del margen de error del modelo de referencia, para poder considerarlas válidas?

### **4.1 Área de Estudio y Fuentes de Datos**

#### **4.1.1 Área de estudio**

La extensión geográfica es el territorio Peninsular español y el archipiélago Balear, que abarca un área de 497167 km<sup>2</sup> (492175 + 4992) km<sup>2</sup>, y cubre 16 de las 17 regiones españolas, excepto el archipiélago Canario y las ciudades autónomas de

Ceuta y Melilla ubicadas al norte de África.

En la figura 15 se muestra el mapa con el área de estudio, así como la distribución de las estaciones meteorológicas en el año 2011 procedentes de las tres redes de estaciones en superficie involucradas. Las guías de la rejilla distan 200 km entre sí. Los puntos de color rojo representan las estaciones ISG de la AEMet, pudiéndose apreciar que están distribuidas por todo el territorio aunque con muy baja densidad. Los elementos de color azul (puntos y cruces) representan las estaciones de la red GUB de CASTILLA y LEÓN (CYL). Se puede apreciar que ésta red es más densa que la de AEMet, aunque sus estaciones se limitan a la comunidad de Castilla y León. Finalmente como puntos en color verde están representadas las estaciones de la REMV Meteoroclimatic (METEO) ubicadas en el área de estudio. Se puede apreciar que las estaciones se encuentran ubicadas principalmente cerca de la costa mediterránea y en las grandes ciudades.

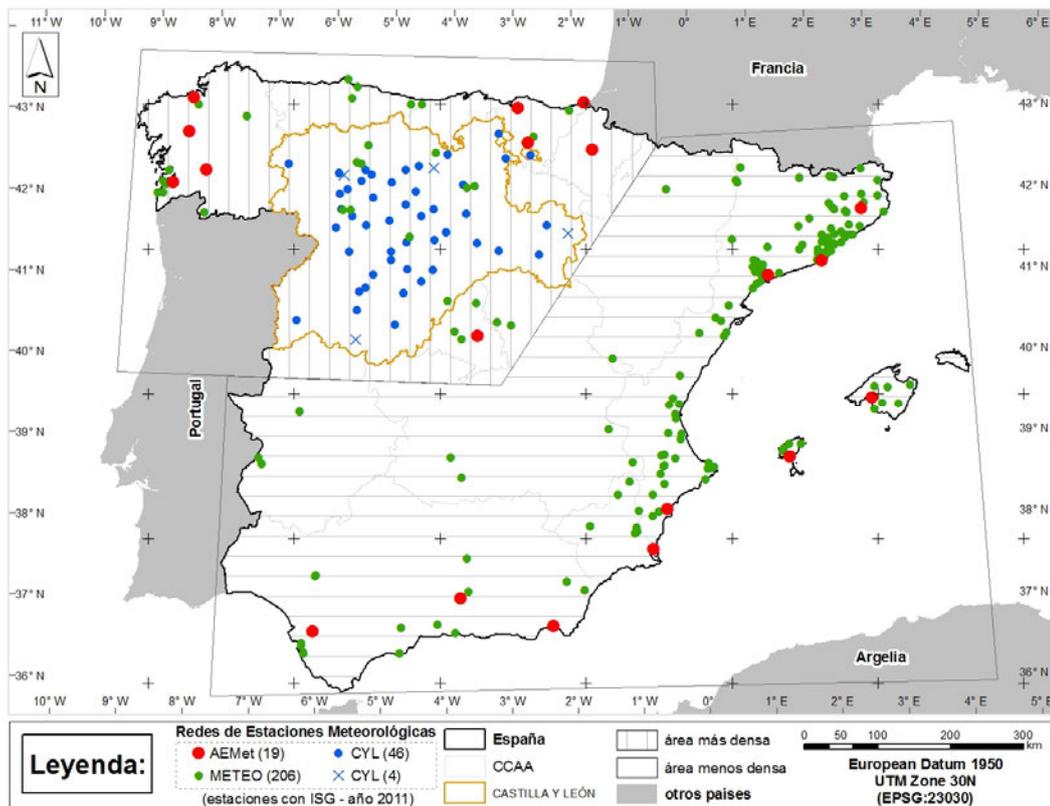


Figura 15. Distribución de las estaciones meteorológicas con sensores de IS de las redes (AEMet, CYL y Meteoclimatic).

En el mapa también se muestran 3 divisiones de interés en relación a la densidad y tipo de estaciones que agrupan. Estas áreas se van a utilizar en posteriores procesos de análisis. Las sub-áreas están representadas por tramas de líneas horizontales y verticales, y la región de Castilla y León. Las dos primeras sub-áreas son divisiones artificiales. La tercera área se corresponde con la división política de España que delimita la región de Castilla y León.

#### **4.1.2 Fuentes de Datos:**

Se utilizaron 4 fuentes de datos distintas, tres son REM terrestres que registraron observaciones de ISG con periodicidades sub-horas en el año 2011 y la última fuente es la base de datos de superficies de ISG HelioClim3v2 estimadas a partir de imágenes del satélite Meteosat. A continuación se describe cada una de ellas.

##### **AEMet:**

Se han utilizado los datos descargados para el trabajo de investigación descrito en el anterior capítulo de 2 años (desde Noviembre de 2010 hasta Octubre de 2012). Durante este periodo, se identificaron 23 estaciones que recolectan información de ISG a intervalos de 10 minutos. Para este propósito sólo se han utilizado los datos de 2011 de 19 estaciones (excluyendo 4 estaciones localizadas en las Islas Canarias) y que pasaron sus filtros. La AEMet hace un chequeo básico de las observaciones en tiempo cuasi real de la ISG (AEMet 2014). Los criterios de este control básico incluyen: (i) Que la ISG no sea mayor que un porcentaje de la constante extraterrestre; (ii) que la IS difusa no sea mayor que la ISG; (ii) en un mismo plano la  $ISG = IS \text{ difusa} + IS \text{ directa}$ , entre otras. Además la agencia también chequea y calibra sus radiómetros según los estándares de la Organización Meteorológica Mundial (OMM) (AEMet 2014).

## **CASTILLA y LEÓN (CYL):**

Esta red de estaciones meteorológicas automáticas pertenece al gobierno autonómico de Castilla y León, que ofrece las observaciones de sus estaciones meteorológicas en internet (ITACyL 2012a) a través del Instituto Tecnológico Agrario de Castilla y León (ITACyL 2012b). Para esta parte de la investigación se utilizaron las observaciones de las 50 estaciones que registraron valores de ISG en el año 2011, con una periodicidad de 30 minutos. Los sensores de esta red son calibrados una vez al año y sus observaciones son chequeadas regularmente usando la norma UNE 500540:204 (AENOR 2004) y los siete niveles de control de calidad indicados en la misma (Estévez y Gavilán 2008).

## **Meteoclimatic:**

Hasta junio de 2013, la red de estaciones meteorológicas de Meteoclimatic (METEO) reportó un total de 1921 estaciones registradas repartidas por la península Ibérica, un hecho que indica la importancia de ésta REMV. Esta red ofrece la información de sus estaciones meteorológicas en Internet mediante páginas web (Meteoclimatic 2014) o a través de hilos de suscripción RSS (Meteoclimatic 2011a). Para esta parte de la investigación se utilizaron las observaciones de las 206 estaciones que registraron valores de ISG en el año 2011 en España. Los controles de calidad en esta red, a diferencia con las anteriores, no son estándares ni obligatorios. Los administradores de la red, así como los usuarios otorgan un sello de calidad basado en sus experiencias y una serie de sugerencias sobre como instalar el sensor (Meteoclimatic 2011b) y se consideran los errores aleatorios encontrados por los propios usuarios (Meteoclimatic 2014). Hasta junio de 2013, Meteoclimatic reportaba que el 34.8 % de sus estaciones habían seguido este proceso.

## **HelioClim3v2:**

Las estimaciones de ISG derivadas a partir de imágenes de satélite utilizadas provienen de la base de datos HelioClim3 versión 2 (HC3v2) (SoDA-is 2011). Esta

base de datos es creada con el método HelioSat2 (Rigollier *et al.* 2004), que utiliza imágenes tomadas cada 15 minutos por el satélite Meteosat de Segunda Generación (*Meteosat Second Generation: MSG*) para estimar la Radiación Solar en la superficie de la Tierra (Blanc *et al.* 2011; SoDA-is 2013). Las estimaciones de ISG desde HC3v2 fueron obtenidas a través de los servicios webs del sistema Soda-is (SoDA-is 2013).

El control de calidad de la base de datos HC3v2 lo han llevado a cabo sus creadores a través de 29 estaciones en tierra. La incertidumbre de la ISG la estiman siguiendo un proceso de *benchmarking* recomendado por la tarea 36 del Programa de Calentamiento y Enfriamiento Solar (*Solar Heating and Cooling Programme: SHC*) de la Agencia Internacional de Energía (*International Energy Agency: IEA*) (IEA-SHC 2010), este *benchmark* define el uso de umbrales (0.1; 10; 50 y 200) W/m<sup>2</sup>, resultando en errores cuadráticos medios de 22.9 %, 22 %, 20.4 % y 16.3 % respectivamente (SoDa 2012b).

## **4.2 Metodología**

### **4.2.1 Colecta de datos y adaptación temporal**

La información de ISG (observaciones) fueron descargadas de forma periódica desde varias fuentes (AEMet, CYL, METEO, HC3v2) mediante programas recolectores para la web (robots web). Estos web-robots fueron desarrollados según las particularidades de cada uno los orígenes como: protocolos de acceso, formato de datos, periodos y latencia de entrega de las redes como se describe en el capítulo 3.

Todas las observaciones se transformaron a un marco temporal común: tiempo UTC y los valores de la ISG en los minutos múltiplos de 15 (minutos: 0, 15, 30 y 45). AEMet y HC3v2 registran sus observaciones en hora UTC por lo que no han necesitado ninguna transformación, mientras que para METEO el tiempo UTC se calcula restando 1 u 2 horas al tiempo local de España (según el periodo del año invierno/verano); para CYL este se calcula sumando el *off-set* del meridiano local en que se encuentra cada estación respecto al meridiano UTC.

Los valores de ISG en los minutos múltiplos de 15 se estimaron por interpolación lineal de las observaciones consecutivas, registrando en un atributo adicional la separación en minutos de dichos vecinos. Esta diferencia de tiempo se utilizó posteriormente en los procesos para discriminar observaciones, de tal forma que sólo se toman en cuenta, los valores de ISG observados y los valores interpolados a partir de observaciones vecinas que no disten más de 30 minutos. En el caso de HC3v2 no fue necesaria ninguna transformación ya que ofrece sus estimaciones en los minutos múltiplos de 15. Para CYL se interpolaron la mayoría de sus observaciones, esto fue debido a que el tiempo de cada estación es dependiente del meridiano de su ubicación (la longitud de su ubicación actúa como su propio meridiano cero), provocando que las observaciones originales se desplazasen en el tiempo unos minutos al transformar a tiempo UTC. Estas características implicaron que la mayoría de las observaciones de las estaciones de esta red fueran interpoladas.

También en el caso de METEO se interpolaron la mayoría de las observaciones en los minutos múltiplos de 15, y en este caso se debe a la variabilidad con la que los propietarios configuran los periodos de observación de sus estaciones. Las observaciones de AEMet se interpolaron sólo los minutos 15 y 45, a partir de las observaciones registradas en los minutos (10 y 20) y (40 y 50) respectivamente, ya que las observaciones son 10-minutales.

#### **4.2.2 Preparación de variables explicativas**

Según Alsamamra *et al.* (2009), las variables explicativas deben ser seleccionadas de tal forma que aporten información adicional al método de interpolación Kriging con Regresión (*Regression Kriging*: RK) y por lo tanto su selección es un factor crítico. Además, su uso correcto permite compensar la escasa distribución y densificación geográfica de la variable a interpolar. Estas variables sólo se aplican en los métodos RK.

Son varios los factores que pueden influir en la cantidad de ISG percibida en una posición de la superficie terrestre. Distintos estudios toman en consideración

diferentes factores, así en Alsamamra *et al.* (2009) utilizan variables derivables del Modelo de Elevación Digital del Terreno para la estimación de la ISG media mensual; Antonanzas-Torres *et al.* (2009) estiman valores anuales de irradiación para tres planos distintos utilizando como variables auxiliares de entrada los valores de irradiación para distintos ángulos de los planos. Evrendilek y Ertekin (2007) estiman el promedio de la RS mensual mediante un amplio número de variables auxiliares: latitud, longitud, altitud, *aspect*, distancia al mar, temperaturas mínimas y máximas, humedad relativa, temperatura del suelo, nubosidad, precipitación, evapotranspiración, longitud del día, duración del brillo del Sol, promedios diarios de radiación solar, mes del año; Mientras que en Ertekin y Evrendilek (2007) sólo utilizan la duración del brillo del Sol como variable auxiliar en la comparación de 18 modelos empíricos para la estimación de la ISG promedio mensual; En Moreno *et al.* (2011) utilizan las temperaturas mínimas, temperaturas máximas y precipitación para estimar la ISG diaria. Bojanowski *et al.* (2013) calibran los coeficientes de tres modelos para la RS diaria utilizando como variable auxiliar la duración del brillo del Sol, la nubosidad y la temperatura del aire. Alsamamra *et al.* (2009) y Kumar *et al.* (2009), modelan la variación topográfica de la RS utilizado como variables auxiliares la estación climática, el ángulo de inclinación del Sol, las condiciones atmosféricas y las sombras originadas por las variaciones en alturas en el terreno. Sin embargo ninguno de los estudios anteriores ha focalizado la estimación de los valores de la ISG cada 15 minutos.

Este trabajo de investigación que tiene como propósito contribuir a mejorar los métodos de interpolación con información adicional que expliquen los cambios en la ISG en periodos de 15 minutos, y para ello se han seleccionado tres variables auxiliares. La primera variable explicativa es una abstracción de las horas de luz durante el día y los cambios continuos en los ángulos del Sol, que simula el movimiento del Sol a lo largo del día y a través de los años. Esta variable se estructuró como una distancia estandarizada calculada para cada observación de la ISG desde su Tiempo Solar Verdadero (*True Solar Time*: TST) hasta su Medio Día Solar (MDS) con respecto al amanecer y el atardecer, para cada estación y para cada día en cuestión. La variable se ha denominado Distancia al Medio Día Solar

(DMDS), no tiene unidades y sus valores están comprendidos entre -1 y 1. El valor -1 es el momento en el que el TST es equivalente al amanecer, el valor de 0 (cero) ocurre cuando el TST es equivalente al MDS (no hay diferencia entre ambos y es el tiempo ideal para observar el mayor valor de la IS), y el valor 1, que ocurre cuando el TST es equivalente al anochecer. Los valores DMDS fuera de este rango implican que el Sol se encuentra bajo el horizonte y por lo tanto los valores de la ISG de la estación son equivalentes a cero. Este hecho resulta interesante porque permite filtrar el número de observaciones a procesar reduciéndolas aproximadamente en un tercio (correspondiente a los periodos nocturnos).

Por lo tanto, la variable DMDS fue creada para que su valor sea independiente de la localización de cada estación, independiente de la fecha y la hora. El cálculo de DMDS implicó la consideración de un total de ocho componentes que se deben conocer o derivar previamente; Dos son variables existentes: (i) el tiempo UTC de cada observación y (ii) la longitud del emplazamiento de cada estación; Estas dos variables sirven de entrada al procesamiento de las siguientes: (iii) amanecer, (iv) atardecer, (v) TST, (vi) MDS (vii) hora decimal y (viii) ecuación del tiempo. Las formulas astronómicas de estos últimos componentes se detallan en (NOOA. 2013). La variable DMDS es una de las variables auxiliares utilizadas en el método de Kriging de Regresión de las siguientes secciones.

La segunda variable es la Estimación de la Irradiancia Solar Global (EISG) desde sensores remotos. Esta variable fue seleccionada siguiendo la tendencia de otros estudios que utilizan variables auxiliares cuyos valores se derivan de imágenes de satélite como es el caso de Hengl *et al.* (2012) que utiliza imágenes del satélite MODIS para las temperaturas.

En esta tesis doctoral, la EISG se obtiene directamente desde la base de datos HC3v2. Cada valor de la EISG se asocia a su respectiva observación de la ISG en la superficie (AEMet, CYL y METEO). Además de utilizar la EISG como variable auxiliar para el cuarto método de interpolación espacial (sección 3.3.4), también se utilizó para comparar las observaciones de la RS en superficie y las estimaciones de la RS a partir de satélite.

La tercera variable utilizada es la latitud, esta variable se obtuvo directamente de las coordenadas de cada estación. Se utiliza en los dos métodos de Kriging de Regresión y su selección se basó en el hecho conocido de que bajo condiciones de cielo claro, los lugares más cercanos al Ecuador terrestre tienen mayores valores de IS que los más lejanos (considerando un mismo nivel de altura), independiente de la temporada del año, día o tiempo (Kumar *et al.* 1997). Así, por ejemplo, en Evrendilek y Ertekin (2007) encontraron una tendencia global Sur-Norte en Turquía.

En la figura 16, se representa gráficamente cada una de las etapas del método descritas hasta este punto (1, 2, 3 y 4) para esta parte de la investigación. También se refleja una última fase (5) que representa la matriz de datos de entrada para los próximos pasos de la metodología conformada por la información colectada y procesada: ISG, EISG, DMDS, TST, Amanecer, Atardecer, tiempo UTC, tipo de valor ISG (observada/interpolada), ubicación de la estaciones (longitud, latitud) así como la información adicional de la estación (Identificador de la estación, Nombre Estación, Origen, y etc.). En la figura, las distintas etapas están representadas mediante una metáfora de engranaje. Cada etapa tiene una o más entradas, ubicadas a la izquierda. Así las primeras fuentes de información proceden de los servidores de internet y la última la conforma la base de datos con la matriz resultante hasta este punto. En esta figura, la salida del proceso 1 ilustra un conjunto de puntos con cierto desorden, representando los distintos tipos de tiempo y temporalidades de las fuentes de datos. El proceso 2 toma la información anterior y la hace coincidir en un mismo marco temporal. El proceso 3 ilustra los cálculos de la primera variable auxiliar (DMDS). El proceso 4 representa el enlace entre las observaciones ISG en superficie y su respectiva estimación por satélite EISG (segunda variable auxiliar) basándose en el tiempo UTC y la ubicación de las estaciones.

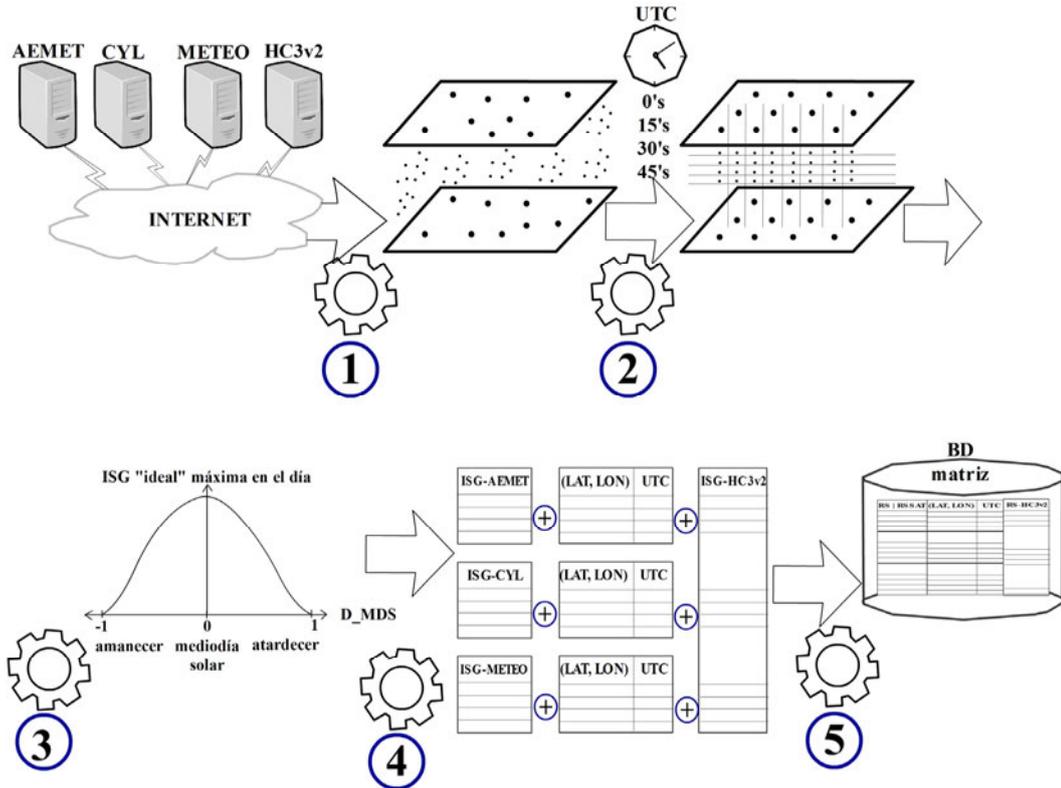


Figura 16. Preparación y colecta de datos para la estimación espacial de la ISG.

### 4.2.3 Agrupación de los datos fuente

El análisis se realizó agrupando los datos de 6 configuraciones geográficas distintas, de tal manera que considerasen los efectos zonales con mayor o menor densidad de estaciones. La figura 17 ilustra en seis mapas las diferentes agrupaciones.

Las figuras 17 (a, b, c) consideran el uso de la totalidad del área de estudio para, posteriormente, validar todas las estaciones voluntarias contenidas. Estas tres agrupaciones se diferencian en la manera de mezclar las estaciones. Así sólo el Grupo I (AEMet + CYL), figura 17 (a), mezcla todas las estaciones oficiales del área de estudio para generar las estimaciones espaciales sobre las que luego se validarán las estaciones voluntarias. En este grupo, la relación en el número de estaciones voluntarias respecto a las oficiales es 2.98:1.

El Grupo II (AEMet + 4 CYL), figura 17 (b), utiliza dos fuentes oficiales: de AEMet considera todas las estaciones, mientras que de CYL solo considera 4 estaciones que están aproximadamente equidistantes entre sí. Se han seleccionado 4 de las 50 estaciones para romper el clúster que generan todas las estaciones de la región de CYL en el área de estudio. Esta elección penaliza la proporción de estaciones oficiales respecto a las voluntarias en una razón de 8.95:1.

En la figura 17 (c), se presenta el Grupo III (AEMet), que solamente incluye las estaciones de AEMet para toda el área de estudio por estar regularmente distribuida en dicha región. En este caso la razón (voluntarias: oficiales) es 10.84:1.

Las figuras 17 (d, e, f) representa las agrupaciones que consideran sub-áreas del ámbito de estudio, considerando únicamente las estaciones que caen dentro de las mismas sin importar su fuente. La figura 17 (d) representa al Grupo IV (CYL), que sólo considera la región de CYL. En esta figura se puede observar que no se incluye ninguna estación de AEMet. La proporción es favorable a las estaciones oficiales 1:5 (voluntarias: oficiales).

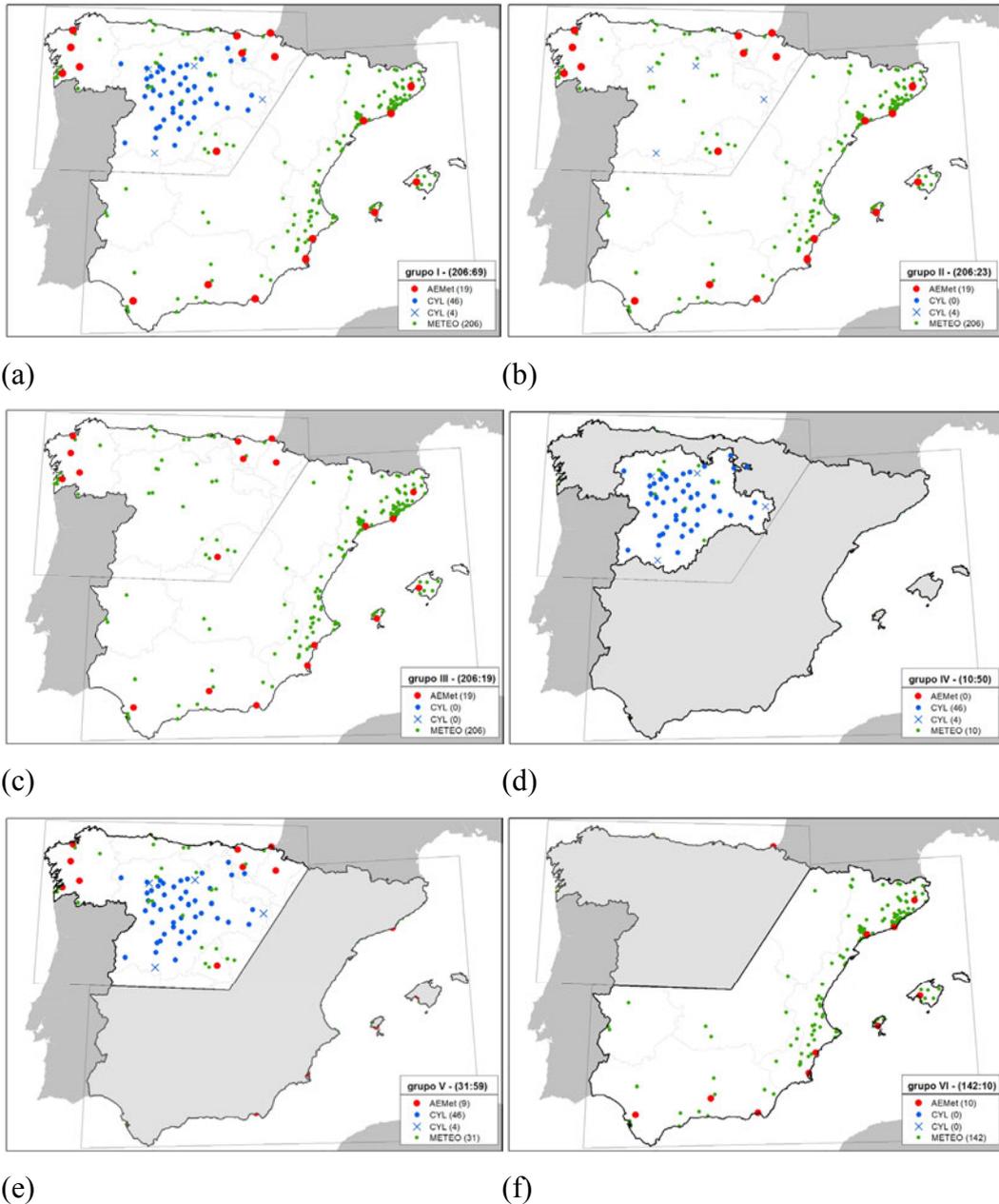


Figura 17. Seis agrupaciones los datos para las estimaciones espaciales. **(a)** Grupo I (AEMet + CYL); **(b)** Grupo II (AEMet + 4 CYL); **(c)** Grupo III (AEMet); **(d)** Grupo IV (CYL); **(e)** Grupo V (CYL + 9 AEMet); **(f)** Grupo VI (AEMet).

El Grupo V (CYL + 9 AEMet), ilustrado en la figura 17 (e), en el que se considera el área Noroeste del territorio peninsular de España. Esta área presenta una razón de 1:1.9 (voluntarias: oficiales). Por el contrario, la figura 17 (f) muestra al Grupo VI (10 AEMet), que es el área más pobre en relación a la cantidad de estaciones oficiales, con solamente 10 estaciones AEMet, además de ser el área que más estaciones voluntarias posee, con una razón desproporcionada de 14:1 (voluntarias: oficiales).

#### 4.2.4 Estimación espacial de la ISG a intervalos de 15 minutos

Los siguientes procedimientos se ejecutaron de forma independiente para cada una de las 6 agrupaciones de datos explicados en la sección anterior. Las interpolaciones (estimaciones) espaciales se realizaron agrupando las observaciones de la ISG en periodos de tiempo de 15 minutos y se procesaron de forma independiente, esto evita tener que enfrentarse a la variabilidad de la ISG a lo largo del día y a lo largo del año. Todos los procedimientos y cálculos de la metodología a partir de este punto se realizan con el paquete informático R (R Development Core Team 2012). Los procesos de interpolación utilizan el paquete geoestadístico para R *gstat* (Pebesma 2004), con excepción de la generación de los semivariogramas en el que se utilizan el paquete para R *automap* (Automap 2011).

#### Detección de valores atípicos

La detección de los valores atípicos (*outliers*) se realizó mediante el método de BoxPlot (Croarkin 2010) en base a las diferencias entre las observaciones y las estimaciones, denotadas ( $Y$ ) e ( $\hat{Y}$ ) respectivamente, en los distintos procesos del análisis. BoxPlot ordena los resultados de las diferencias con el propósito de obtener el primer (Q1) y tercer cuartil (Q3) de dichos valores, y con estos calcula el valor inter-cuartil ( $IQ = Q3 - Q1$ ). Seguidamente define los límites inferior y superior a partir de los cuales los distintos valores son considerados *outliers* leves (en caso de estar fuera de los límites) o valores válidos (en caso de estar dentro de los mismos). Estos límites están ubicados a -1.5 del valor IQ a la izquierda del Q1 y a 1.5 del valor IQ a la derecha de Q3.

Para evitar que las observaciones de las estaciones oficiales introduzcan ruido en los procesos de interpolación, a pesar de haber pasado los controles sistemáticos de las organizaciones, se ha realizado un análisis previo de los datos en el que se ha detectado estaciones con comportamiento anómalo a lo largo del día, para así obviarlas en los futuros procesamientos de cada día en cuestión. En este caso las observaciones ( $Y$ ) son los valores cada 15 minutos de la ISG desde las estaciones, y las estimaciones ( $\hat{Y}$ ) son las obtenidas a partir de las imágenes satelitales (EISG). Sus diferencias se agruparon por día y por estación, y se agregaron usando dos

estadísticos: Error Medio Cuadrático (*Root Mean Square Error*: RMSE) (ecuación 12) y Error Medio Absoluto (*Mean Absolute Error*: MAE) (ecuación 13), a los que se les aplicó la detección de *outliers* por BoxPlot. Indicar, finalmente, que se consideraron *outliers* aquellas estaciones identificadas por este método tanto si lo fueron por el estadístico RMSE como por el MAE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (12)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (13)$$

Identificadas las estaciones oficiales válidas para cada día, se procedió a utilizar sus observaciones en los cuatro métodos de interpolación espacial descritos el capítulo 2.3. IDW, OK y las dos formas de RK. Ambas formas de RK (denominadas RK1 y RK2 respectivamente) se distinguen por tener diferente la primera de sus dos variables auxiliares (DMDS en el caso de RK1 y EISG en el caso de RK2), sin embargo ambas comparten como segunda variable auxiliar la “Latitud” geográfica.

No se juntaron ambas variables (EISG y DMDS) en un mismo método de RK por el hecho de que ambas presentan valores de multicolinealidad casi extrema. Esto se comprobó separando todas las observaciones entre *ante meridiem* “a.m.” (DMDS  $\leq 0$ ) y *post meridiem* “p.m.” (DMDS  $> 0$ ), el valor de correlación de a.m. es 0.7825 con un coeficiente de determinación  $R^2$  de 0.6124, y en p.m. el valor de correlación es -0.8102 con  $R^2$  de 0.6566. Al ser ambas variables conceptualmente distintas y con  $R^2 > 0.6$ , se adoptó la decisión de no utilizarlas juntas y así evitar que los MLRM sean más sensibles a errores menores (Allison 1998, pp. 141, 150).

### **Validación de los métodos**

Para determinar la precisión de los modelos se realizaron validaciones cruzadas de tipo *leave-one-out*. Esto permitió identificar cuál de los cuatro métodos de interpolación es el más adecuado. Esto se llevó a cabo utilizando los datos de todas

las estaciones menos una (oculta), para luego estimar el valor en la ubicación de la estación oculta y compararlo con su valor observado. La precisión del modelo HelioSat2 fue obtenida desde las estimaciones directas ubicadas en la base de datos HC3v2, calculándose las diferencias entre las observaciones en tierra y las estimaciones.

En todos los casos se calcularon los estadísticos RMSE relativo (%RMSE) y el MAE relativo (%MAE) en base al ratio entre el error estadístico y la media de la variable observada multiplicada por 100, como se puede observar en las ecuaciones (14) y (15) respectivamente:

$$\%RMSE = \frac{RMSE}{\frac{1}{n} \sum_{i=1}^n (Y_i)} \cdot 100 \quad (14)$$

$$\%MAE = \frac{MAE}{\frac{1}{n} \sum_{i=1}^n (Y_i)} \cdot 100 \quad (15)$$

Estos indicadores %RMSE y %MAE se calcularon independientemente para cada una de las 6 formas de agrupar geográficamente los datos, y para cada conjunto de observaciones 15-minutales procesadas y agrupadas por día.

Con el fin último de poder comparar los resultados de los distintos modelos con el *benchmarking* de referencia en (SoDa-is 2013b), se calculan los errores para 5 umbrales de la ISG observada en tierra (NU, 0.1, 10, 50 y 200) W/m<sup>2</sup>, estos umbrales también son utilizados por (SoDa-is 2013b) para el cálculo de la incertidumbre del modelo HelioSat2. La aplicación del sin umbral (NU) y el umbral (0.1), han permitido destacar el aporte que hacen a los *outliers* los valores muy pequeños (ISG < 0.1), mientras que los otros umbrales permiten identificar la respuesta de los distintos métodos de estimación ante la eliminación de valores de ISG bajos.

#### **4.2.5 Primera aproximación a la validación de las estaciones de voluntarios**

Una vez encontrada la mejor combinación del método de estimación y de la forma de agrupar los datos en función al mejor valor del error y su desviación estándar (stdv), el siguiente paso ha consistido en estimar los valores de la ISG en los emplazamientos de las estaciones de voluntarios. Conociendo el valor del error de las interpolaciones y su desviación estándar (stdv) se procedió a clasificar las observaciones ISG de dichas estaciones entre aquellas que se encuentran dentro de este rango y las que se caen fuera, para finalmente cuantificar estas clasificaciones por día y por estación y de ésta manera saber qué estaciones se comportaron de una forma “válida” o “inválida” respecto a los errores de los modelos de interpolación o la estimación por satélite.

En la figura 18, se representa el flujo de trabajo, que resumidamente incluye: por cada grupo de observaciones 15-minutales y por cada una de las 6 agrupaciones geográficas de las estaciones, la ejecución de proceso de generación de MLRM (para los casos de RK), seguidamente se ilustran los procesos de estimación de variogramas (para los casos que aplican), los procesos de estimaciones mediante interpolaciones, procesos de validaciones cruzadas de los métodos de interpolación, la interpolación del mejor método/forma de agrupación en la posición de las estaciones voluntarias y la clasificación de las voluntarias que se encuentran dentro del margen de error de las estimaciones.

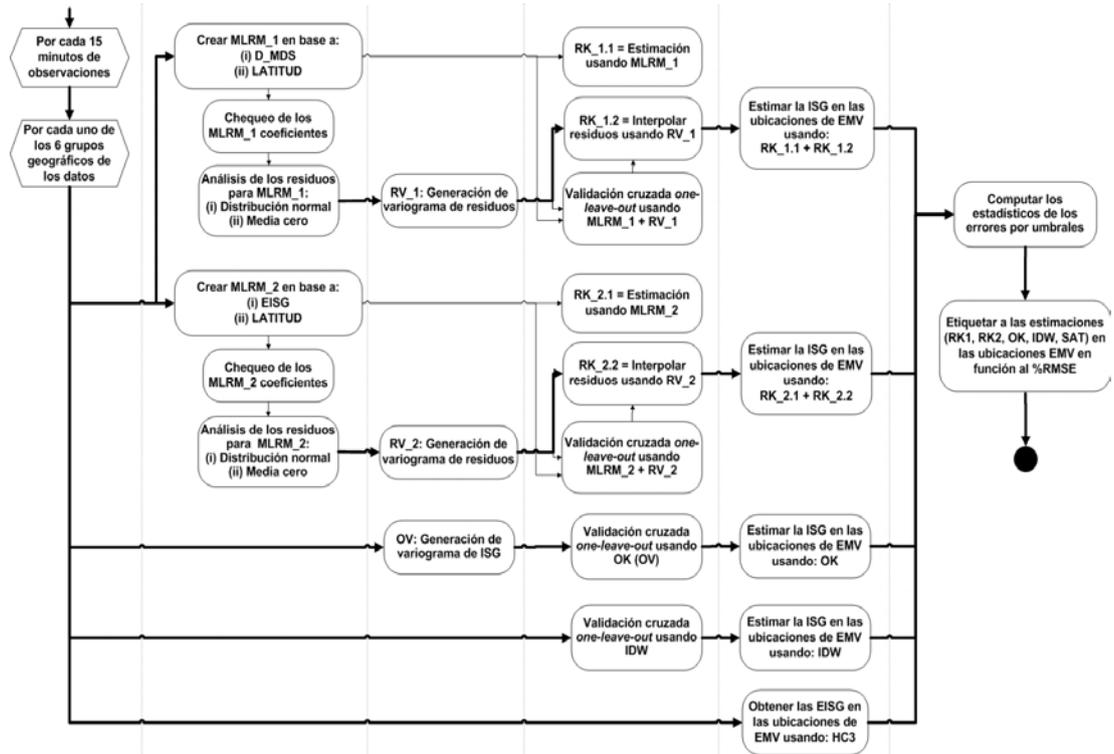


Figura 18. Flujo de trabajo de los procesos de análisis de los métodos de estimación espacial.

## 4.3 Resultados y discusión

### 4.3.1 Puntuación de los métodos de estimación

La precisión y puntuación de los métodos de estimación en cada una de las agrupaciones geográficas se obtuvo calculando agregaciones diarias de los estadísticos %RMSE y %MAE en base a las diferencias 15-minutales (ISG – EISG) y descartando los días identificados como atípicos por el método BoxPlot. Estas puntuaciones se presentan en las tablas 11 y 12; La tabla 11 presenta ambos estadísticos sin umbralizar los valores registrados en las observaciones ISG, mientras que la tabla 12 presenta los dos estadísticos umbralizando la IS con valores mayores a  $200 \text{ W/m}^2$ . Esta umbralización ha permitido evidenciar el efecto que produce la estimación de la IS cuando los valores son bajos (tabla 11). Si se excluyen dichos valores ( $IS \leq 200 \text{ W/m}^2$ ) se observa que los distintos métodos de estimación producen mejores resultados (tabla 12).

<i>Puntuación</i>	<i>I</i>		<i>II</i>		<i>III</i>		<i>IV</i>		<i>V</i>		<i>VI</i>		<i>Estadístico</i>
<b>1º</b>	<b>RK2</b>	<b>23.75</b>	<b>RK2</b>	<b>27.09</b>	<b>SAT</b>	<b>26.93</b>	<b>RK2</b>	<b>21.77</b>	<b>RK2</b>	<b>23.27</b>	<b>SAT</b>	<b>23.12</b>	%RMSE
<b>2º</b>	SAT	29.09	SAT	27.49	RK2	28.02	RK1	25.25	RK1	28.59	RK2	29.82	
<b>3º</b>	RK1	30.59	RK1	39.28	IDW	39.86	OK	25.41	IDW	28.86	IDW	34.16	
<b>4º</b>	IDW	30.79	IDW	39.50	RK1	41.17	IDW	25.86	OK	28.92	OK	36.75	
<b>5º</b>	OK	31.61	OK	40.56	OK	43.22	SAT	28.95	SAT	29.94	RK1	37.91	
<b>1º</b>	RK2	16.68	RK2	19.24	SAT	18.75	RK2	15.19	RK2	16.30	SAT	16.29	%MAE
<b>2º</b>	RK1	20.51	SAT	19.46	RK2	19.44	RK1	17.30	RK1	19.26	RK2	19.54	
<b>3º</b>	IDW	20.92	RK1	28.01	IDW	28.04	OK	17.53	OK	19.52	IDW	24.95	
<b>4º</b>	OK	21.18	IDW	28.60	RK1	29.48	IDW	17.95	IDW	19.70	OK	27.28	
<b>5º</b>	SAT	21.74	OK	29.42	OK	31.67	SAT	21.81	SAT	22.74	RK1	27.40	

Tabla 11. Clasificación de los métodos sin discriminar los valores de irradiancia (agregación diaria) de acuerdo a los 6 grupos de datos: I (AEMet + CYL); II (AEMet + 4 CYL); III (AEMet); IV (CYL); V (CYL + 9 AEMet); VI (10 AEMet).

<i>Puntuación</i>	<i>I</i>		<i>II</i>		<i>III</i>		<i>IV</i>		<i>V</i>		<i>VI</i>		<i>Estadístico</i>
<b>1º</b>	<b>RK2</b>	<b>17.45</b>	<b>SAT</b>	<b>18.78</b>	<b>SAT</b>	<b>18.29</b>	<b>RK2</b>	<b>16.50</b>	<b>RK2</b>	<b>17.37</b>	<b>SAT</b>	<b>17.06</b>	RMSEp
<b>2º</b>	SAT	19.90	RK2	19.18	RK2	19.97	RK1	19.04	SAT	20.63	RK2	21.98	
<b>3º</b>	RK1	22.25	RK1	27.50	IDW	28.38	OK	19.12	IDW	20.94	IDW	25.68	
<b>4º</b>	IDW	22.49	OK	27.84	RK1	29.07	IDW	19.57	RK1	21.04	OK	26.32	
<b>5º</b>	OK	22.73	IDW	27.96	OK	29.77	SAT	20.82	OK	21.18	RK1	28.56	
<b>1º</b>	RK2	13.12	SAT	14.09	SAT	13.62	RK2	12.41	RK2	13.12	SAT	12.80	MAEp
<b>2º</b>	SAT	15.56	RK2	14.38	RK2	14.56	RK1	14.28	RK1	15.51	RK2	15.52	
<b>3º</b>	RK1	16.43	RK1	21.02	IDW	21.51	OK	14.50	IDW	15.81	IDW	19.39	
<b>4º</b>	IDW	16.88	IDW	21.49	RK1	21.97	IDW	14.89	OK	16.09	OK	20.65	
<b>5º</b>	OK	17.13	OK	21.79	OK	23.14	SAT	16.46	SAT	16.31	RK1	21.27	

Tabla 12. Clasificación de los métodos para valores de irradiancia mayores a 200 W/m<sup>2</sup> (agregación diaria) de acuerdo a los 6 grupos de datos: I (AEMet + CYL); II (AEMet + 4 CYL); III (AEMet); IV (CYL); V (CYL + 9 AEMet); VI (10 AEMet).

De la observación de los valores de ambas tablas se puede deducir que los métodos RK2 y SAT son siempre los ganadores, sin importar el tipo de estadístico o el umbral de discriminación. Siendo el método SAT siempre el ganador en las formas

de agrupamiento (III y VI) y RK2 siempre el ganador en las formas de agrupamiento (I, IV y V). Cuando se consideran todas las observaciones, el método ganador para el agrupamiento II es RK2. Sin embargo cuando solo se consideran las observaciones de ISG mayores a  $200 \text{ W/m}^2$  el método ganador es SAT, sugiriéndose así la mejora del método SAT ante valores grandes de ISG.

Las formas de agrupamiento (III y VI) en las que SAT es ganador, son aquellas en las que sólo participan las estaciones de AEMet, por lo que poseen el menor número de estaciones por área de trabajo y por lo tanto una menor densidad de estaciones (ver tabla 13). También se puede apreciar que, siempre que el método SAT es el ganador, el segundo mejor método es RK2.

Las formas de agrupamiento (I, IV y V) en las que RK2 es ganador se caracterizan porque siempre contienen todas las estaciones de CYL más otras estaciones de AEMet. Esto implica una mayor cantidad de estaciones y por lo tanto mayor densidad. En ellas, las formas de agrupamiento (IV y V), y a pesar de que el método RK2 utiliza las estimaciones por satélite como variable auxiliar, presentan a SAT como el peor método. Estos agrupamientos (IV y V) disponen de mayores densidades de estaciones por lo que, en término medio las estaciones están más próximas (hay menor distancia promedio entre todas las estaciones) como se puede observar en la tabla 13. En estos casos el segundo mejor método es RK1, con excepción del agrupamiento V bajo el estadístico %RMSE sin limitación de umbral. En todas las otras formas de agrupamiento sin importar el umbral o el estadístico, los segundos mejores métodos son RK2 o SAT.

Considerando la relación que pueda existir entre los métodos ganadores y la cantidad de estaciones que posee cada tipo de agrupación, así como las distancias medias a los vecinos, en la tabla 13 se presentan los estadísticos de cada una de las formas de agrupación. Se puede apreciar como los grupos que incluyen todas las estaciones de CYL (IV y V) presentan estadísticos más adecuados (promedio de distancia entre los vecinos más bajos), al contrario de lo que ocurre en los grupos en los que sólo contienen estaciones AEMet (III y VI) que cubren más área con menos estaciones.

<i>Grupo</i>	<i>Distancia media al VMC (km)</i>	<i>Distancia media (vecinos) (km)</i>	<i>Número de Estaciones</i>	<i>Área del Grupo</i>	<i>km<sup>2</sup> por cada estación</i>
I (AEMet + CYL)	43.69	288.93	69	497167	7205
II (AEMet + 4 CYL)	100.16	470.30	23	497167	21615
III (AEMet)	97.43	495.86	19	497167	26166
IV (CYL)	25.94	137.42	50	94226	11884
V (CYL + 9 AEMet)	30.03	178.82	59	195064	3306
VI (10 AEMet)	111.26	373.37	10	303839	30383

Tabla 13. Distancia media hacia la estación Vecina Más Cercana (VMC) y distancia media entre todas las estaciones de cada grupo.

El mejor método seleccionado por cada forma de agrupamiento, fue aquel que presentó el valor agregado de la media más bajo. En las tablas 14 y 15 se presentan estos métodos ganadores junto con otros valores de relevancia (media, desviación estándar, porcentajes de días descartados porque el método presentó estimaciones atípicas, porcentaje de días en que el método obtuvo mejores valores de medias que los demás métodos en su mismo grupo). La tabla 14 muestra los casos en los que se han considerado todos los valores de ISG mientras que en la tabla 15 sólo se muestran los casos con el umbral  $ISG > 200 \text{ W/m}^2$ .

<i>Grupo</i>	<i>mejor método</i>	<i>Media</i>	<i>Stdv</i>	<i>% Días descartados</i>	<i>% Días como N°1</i>	<i>Estadístico</i>
I (CYL + AEMet)	RK2	23.75	11.38	3.35	91.62	%RMSE
II (AEMet + 4 CYL)	RK2	29.09	10.02	2.67	49.85	
III (AEMet)	SAT	26.93	10.03	3.86	69.44	
<b>IV (CYL)</b>	<b>RK2</b>	<b>21.77</b>	<b>11.57</b>	<b>0.00</b>	<b>84.27</b>	
V (CYL + 9 AEMet)	RK2	23.27	11.63	1.40	92.16	
VI (10 AEMet)	SAT	23.12	10.53	4.42	83.91	
I (CYL + AEMet)	RK2	16.68	7.90	1.12	89.94	%MAE
II (AEMet + 4 CYL)	RK2	19.24	7.66	2.08	48.96	
III (AEMet)	SAT	18.75	7.63	4.45	64.39	
<b>IV (CYL)</b>	<b>RK2</b>	<b>15.19</b>	<b>7.99</b>	<b>0.00</b>	<b>79.49</b>	
V (CYL + 9 AEMet)	RK2	16.30	8.12	0.28	88.80	
VI (10 AEMet)	SAT	16.29	8.08	5.05	81.39	

Tabla 14. El mejor método por grupo y por estadístico, sin discriminar valores de irradiancia (agregación para el número total de días).

Grupo	mejor método	Media	Stdv	% Días descartados	% Días como N°1	Estadístico
I (CYL + AEMet)	RK2	17.45	7.23	3.07	81.01	%RMSE
II (AEMet + 4 CYL)	SAT	18.78	5.74	0.29	72.40	
III (AEMet)	SAT	18.29	5.35	1.78	84.27	
<b>IV (CYL)</b>	<b>RK2</b>	<b>16.50</b>	<b>7.96</b>	<b>0.28</b>	<b>74.16</b>	
V (CYL + 9 AEMet)	RK2	17.37	7.90	1.40	81.23	
VI (10 AEMet)	SAT	17.06	6.24	1.26	89.27	
I (CYL + AEMet)	RK2	13.12	5.85	1.39	82.12	%MAE
II (AEMet + 4 CYL)	SAT	14.09	4.73	0.89	69.73	
III (AEMet)	SAT	13.62	4.50	2.08	79.53	
<b>IV (CYL)</b>	<b>RK2</b>	<b>12.41</b>	<b>6.40</b>	<b>0.28</b>	<b>73.03</b>	
V (CYL + 9 AEMet)	RK2	13.12	6.46	0.28	80.67	
VI (10 AEMet)	SAT	12.80	5.30	0.95	85.80	

Tabla 15. El mejor método por grupo y por estadístico, discriminando valores de irradiancia mayores de 200 W/m<sup>2</sup> (agregación para todos los días).

En ambas tablas 14 y 15, se aprecia que aplicando el método RK2 al grupo IV (CYL) se obtiene el menor valor medio de error (el mejor resultado), sin embargo, este método/grupo aún no puede ser seleccionado como el único ganador, pues no cubre todo el área de estudio. Aquí también se aprecia que el grupo VI (10 AEMet) con el método SAT tiene el segundo valor más bajo de media, pero este grupo tampoco cubre todo el área. La unión de las áreas de los dos mejores grupos/métodos (IV y V) tampoco cubre toda el área de estudio, por lo que se ha de considerar un tercer grupo/método. En este caso la selección es el Grupo (V) con el método RK2.

Este tercer grupo/método (Grupo V: KR2) presenta el siguiente mejor valor de media, además su extensión incluye parte del área de interés que aún faltaba por cubrir. Otra característica a destacar de los tres grupos seleccionados, es que presentan valores de desviación estándar entre 6.24 % y 7.96 % cuando se umbraliza la ISG con valores superiores a 200 W/m<sup>2</sup>. Estos valores en las validaciones cruzadas no se alejan demasiado del máximo de 5 % que en según Zelenka *et al.* (1999) es el aceptado para estaciones con observaciones horarias que tienen un mantenimiento adecuado. Es plausible, por tanto, que la diferencia

presentada en este estudio se deba al hecho de trabajar con observaciones intra-hora. Por otro lado, los tres grupos restantes no seleccionados (I, II y III) así como sus métodos, presentan la característica de contemplar toda el área de estudio, sin embargo los tres presentan los tres valores de media de los errores más elevados.

#### 4.3.2 Análisis detallado de los métodos ganadores

Una vez definida la combinación de los tres grupos/métodos que cubren el área de estudio, se procedió a analizar cada uno de ellos. En la figura 19 se presentan los diagramas de dispersión, línea de regresión, así como los coeficientes de correlación ( $R$ ) y los coeficiente de determinación ( $R^2$ ) para cada uno de ellos. A nivel ilustrativo en la figura 19 sólo se muestra una nube de puntos que representa el 5 % de los valores, seleccionados de forma aleatoria, de entre aquellos valores medidos/estimados que no son *outliers* (según el método BoxPlot de sus discrepancias en la estimación). Sin embargo los valores de los coeficientes así como la línea de regresión si están basadas en todo el universo y sin discriminar *outliers*. De la figura 19 se puede concluir que los tres métodos ganadores en sus respectivos grupos presentan una fuerte correlación lineal, valores  $R$  de 0.97, 0.92 y 0.95. Además la figura 19 muestra valores  $R^2$  altos para los tres métodos (0.94, 0.86 y 0.90).

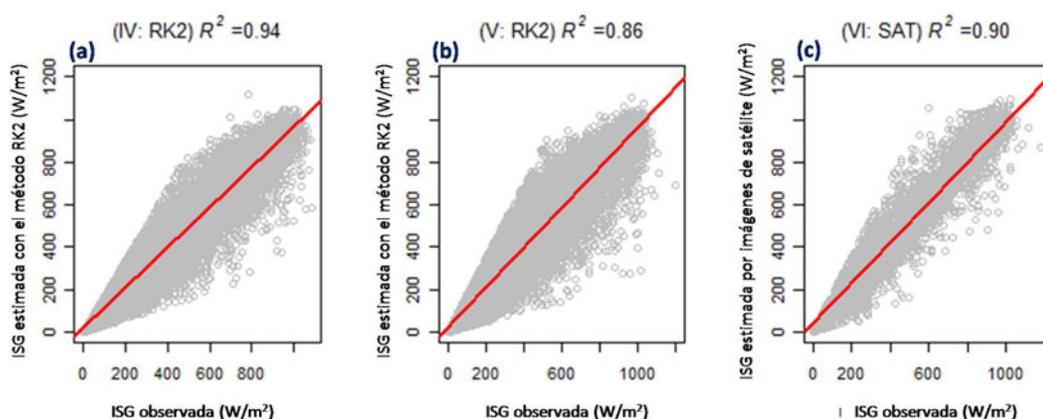


Figura 19. Relación entre la ISG estimada por distintos métodos y la ISG observada por las estaciones en superficie. (a) RK2: grupo IV (CYL). (b) RK2: grupo V (CYL + 9 AEMet). (c) SAT: grupo VI (10 AEMet).

A continuación se procede a analizar ciertas características y limitaciones destacables de cada uno de los métodos ganadores, teniendo en cuenta la consistencia entre los estadísticos dependientes ya sea el RMSE como de MAE, si bien es el primer estadístico el comúnmente utilizado en la literatura de referencia para comparar.

### **RK2 en la forma de agrupación IV (CYL)**

La tabla 16 presenta los estadísticos para el método RK2 en la forma de agrupamiento (IV- CYL) para 6 umbrales de ISG (todo, 0.1, 10, 50, 200 W/m<sup>2</sup>). Se puede apreciar que los estadísticos porcentuales (%RMSE y %Stdv) van disminuyendo a medida que se umbralizan los valores de la ISG, desde una media de 22.46 y stdv de 2.87 para el umbral (0.1), hasta aproximadamente la mitad de esos valores (11.26 y 1.35 respectivamente) para el umbral (200 W/m<sup>2</sup>). Por el contrario los estadísticos absolutos presentan una leve tendencia creciente (68.95 y 10.36 W/m<sup>2</sup>) para el umbral (0.1), y (72.21 y 12.97 W/m<sup>2</sup>) para el umbral.

<i>Umbral</i>	<i>%RMSE</i> (%)	<i>Stdv %RMSE</i> (%)	<i>RMSE</i> (W/m <sup>2</sup> )	<i>Stdv RMSE</i> (W/m <sup>2</sup> )	<i>Rango RMSE</i> (W/m <sup>2</sup> )	<i>%RMSE</i> (All)	<i>Stdv RMSEp</i> (All)
<i>NA</i>	22.46	2.87	68.95	10.36	59 - 79	247.31	508.06
<i>0.1</i>	22.46	2.87	68.95	10.36	59 - 79	148.74	362.46
<i>10</i>	21.03	2.75	68.97	10.59	59 - 81	51.90	25.22
<i>50</i>	17.51	2.45	72.86	11.51	61 - 84	32.73	9.15
<i>200</i>	11.26	1.35	75.21	12.97	62 - 88	17.88	3.23

Tabla 16. Estadísticos para el método RK2 en el grupo IV (CYL).

Los valores %RMSE y RMSE mostrados en la tabla 16 son más bajos que los valores respectivos de (69, 72, 77, 83) W/m<sup>2</sup> y (22.9, 20.2, 20.04, 16.3) % y para umbrales (0.1, 10, 50 y 200) W/m<sup>2</sup> del *benchmarking* en (SoDa-is 2013b).

También se puede notar que no hay diferencia entre el umbral (0.1) y no umbral (NA) para los estadísticos sin *outliers*. Sin embargo para esos mismos niveles de umbral si hay una gran disminución entre los estadísticos %RMSE (All) y Stdv (All) del 40 % (149.74 / 247.31) y el 45 % (362.46 / 508.06) respectivamente. La

importancia de esta evidencia es que permite destacar como los valores muy pequeños de la ISG ( $ISG < 0.1$ ) contribuyen a los *outliers*. También se puede apreciar que esa tendencia continua, aunque en menor medida, en los valores de umbral bajos (10 y 50). Esto evidencia una limitación del método ante la estimación de valores bajos de la ISG (amanecer y anochecer o días nublados).

La figura 20 presenta los gráficos del %RMSE en función de la distancia que separa el punto a interpolar de los vecinos. Esto demuestra la fortaleza del método aquí desarrollado y permite comparar los resultados de la presente investigación con los de Zelenka *et al.* (1999) y Perez *et al.* (1997). Estas gráficas, al igual que las de la figura 21, incluyen una línea de regresión lineal orientativa (recta azul), así como su respectivo intervalo de confianza (zona gris). Las flechas en el eje “Y” apuntan al 25 %RMSE, y las del eje “X” a los límites de 25 y 34 km indicados en los estudios antes citados.

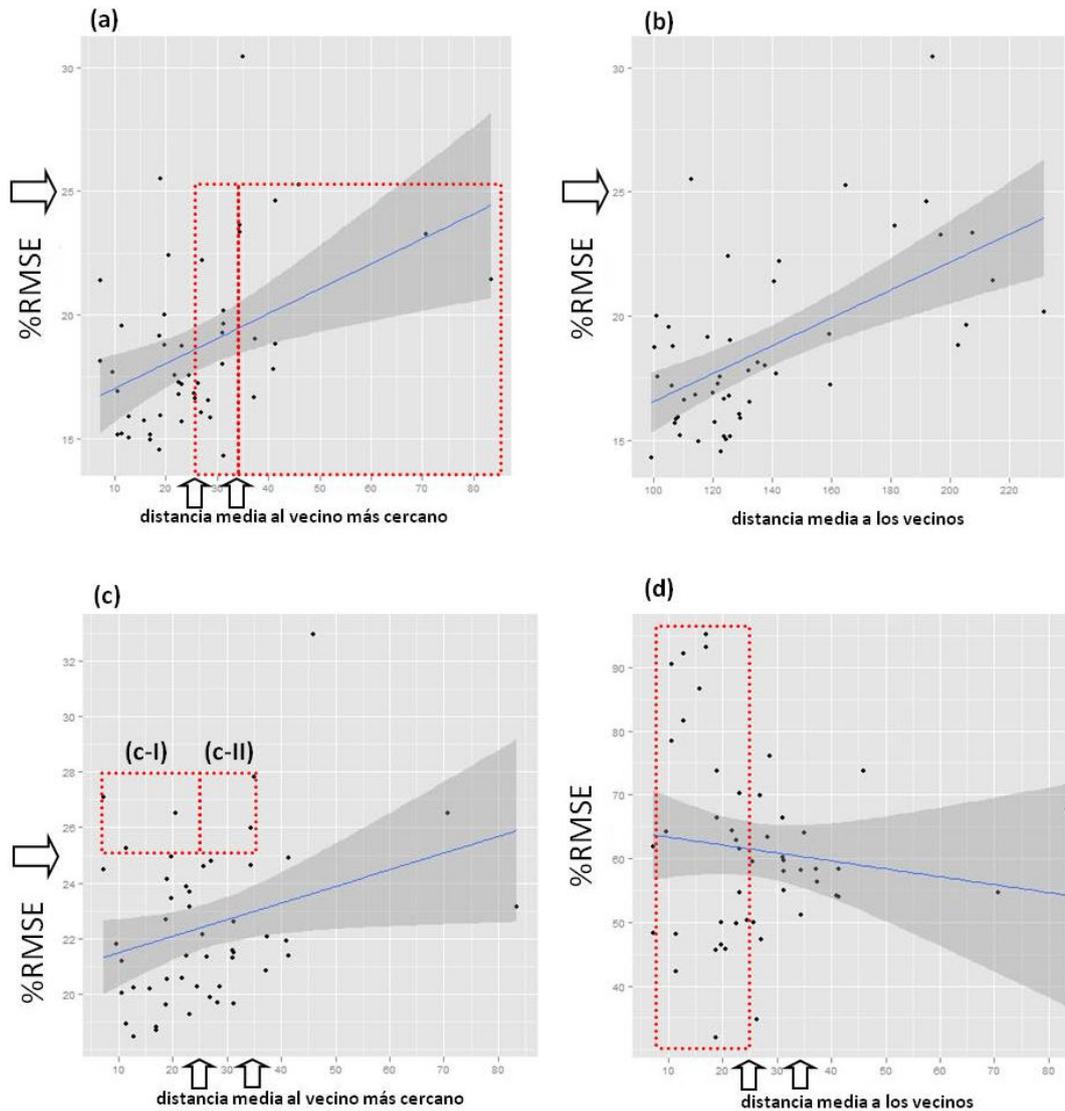


Figura 20. %RMSE de las interpolaciones en relación a los vecinos para el método RK2 en CYL. (a) %RMSEp en relación al vecino más cercano para umbral 200 W/m<sup>2</sup>. (b) %RMSEp en relación a la distancia media entre los vecinos para umbral 200. (c) %RMSE en función al vecino más cercano sin umbralizar. (d) %RMSE de las estimaciones derivadas por satélites en función al vecino más cercano sin umbralizar.

La figura 20 (a) muestra el %RMSE con umbral  $> 200 \text{ W/m}^2$ . En esta gráfica se aprecia un efecto “nugget” del orden de 15-17 %, similar al reportado por Zelenka *et al.* (1999) y Perez *et al.* (1997). En estos estudios se sugiere que el efecto “nugget” se debe a la naturaleza discontinua de la estructura espacial de la ISG horaria (cielo azul con nubes). Cabe destacar que a pesar de esta y posteriores similitudes, dichos estudios utilizaron como método de interpolación a una variante del IDW, también utilizaron datos de ISG horarios, así como valores ISG

con rango RMSE entre los 80 y los 100 W/m<sup>2</sup>, mientras que el rango RMSE de los datos 15 minútales usados en la presente investigación está entre los 62 y los 88 W/m<sup>2</sup>.

La principal aportación de la figura 20 (a) es que el %RMSE es inferior o igual al 25.5 %, independientemente de la distancia que separa el punto a interpolar del vecino más cercano, o independientemente de la distancia media a todos los vecinos, en el caso de la figura 20 (b). Lo anterior hace destacar al método RK2, ya que a pesar de trabajar con observaciones ISG intra-horas, no superan el límite de 25 % de %RMSE para observaciones horarias indicado por Zelenka *et al.* (1999) y Pérez *et al.* (1997). En ellos se indica que por encima de los 25 km (en el primero) y 34 km (en el segundo) las interpolaciones a partir de observaciones en superficie sobrepasan el 25 % de %RMSE y por lo tanto a partir de esas distancias las estimaciones por satélite son más precisas por lo que es preferible su uso. Los rectángulos rojos en la figura 20 (a) representan estaciones que según los estudios antes citados deberían haber presentado en sus interpolaciones %RMSE mayores que el 25 %. Por otro lado, los resultados aquí mostrados también presentan otra mejora respecto de las observaciones intra-horarias presentadas en Zelenka *et al.* (1999) donde el efecto “nugget” es de 22 % y aproximadamente desde los 20 km el %RMSE supera el 30 %, llegando a 40 % aproximadamente a los 80 km.

La figura 20 (c), al igual que en la figura (a), se presenta la relación entre el %RMSE para cada estación y su vecino más cercano. La diferencia con las figuras 20 (a y b), reside en que no se umbralizan los valores de la ISG. En esta figura se aprecia como la inclusión de valores de la ISG menores a 200 W/m<sup>2</sup> aumenta el efecto “nugget” hasta 20-22.5 %. También se aprecia en los rectángulos rojos (c-I y c-II), que existen 6 estaciones que sobrepasan el umbral del 25 % aunque no llegan al 28 %. Sin embargo 5 de 6 de estas estaciones se encuentran dentro de los 34 km, y 3 de esas 5 dentro de los 25 km, por lo que es plausible asumir que haber superado el umbral del 25 % de %RMSE es consecuencia de los bajos valores de la ISG y no de la distancia a su vecino más cercano. La única excepción en todo lo anterior lo presenta la estación (BU101) con una distancia a su vecino más cercano de 35 km y un %RMSE de 35 %.

La figura 20 (d) es ilustrativa del comportamiento del método SAT, con estimaciones de la ISG derivadas a partir de imágenes de satélite (EISG). Para este grupo IV se encontró que este método ocupa el último puesto en todos los casos (ver tablas 11 y 12). Como es de esperar, en esta figura se aprecia que no existe una relación entre la separación de los puntos estimados y el %RMSE de la EISG; A simple vista se observa que existe una amplia dispersión de puntos (contenidos dentro del recuadro rojo) en todo el eje “Y” entre los 7 y 20 km; De igual forma se aprecia que estos valores de %RMSE son mayores que los del método RK2 en la figura 20 (c).

### **RK2 en la forma de agrupación V (CYL + 9 AEMet)**

En la tabla 17 se presentan los estadísticos resultantes de aplicar el método RK2 en la forma de agrupamiento V (CYL + 9 AEMet) para 6 umbrales de la ISG (todo, 0.1, 10, 50, 200). De aquí cabe destacar que los valores %RMSE para los umbrales 50 y 200 (18.95 y 12.29) %, son mejores que los respectivos valores de (20.04 y 16.3) % del *benchmarking* en (SoDa 2012b); mientras que los valores RMSE a partir de los umbrales 10 (70.49, 68.95 y 63.39) W/m<sup>2</sup>, son mejores que los equivalentes del citado *benchmarking*. De forma general se aprecia el mismo comportamiento que en (RK2 grupo IV) y tabla 16 respecto a la disminución de los %RMSE y su stdv a medida que los valores de umbrales son mayores. Sin embargo en este caso, los valores de las medias relativas y sus desviaciones estándar también disminuyen, contrario a lo que ocurre con (RK2 grupo IV). Otra diferencia a destacar respecto al método anterior, es que allí siempre presenta mejores valores que el citado *benchmarking* para todos los umbrales.

Umbral	%RMSE (%)	Stdv %RMSE (%)	RMSE (Wm <sup>2</sup> )	Stdv RMSE (Wm <sup>2</sup> )	Rango RMSE (Wm <sup>2</sup> )	%RMSE (All)	Stdv %RMSE (All)
NA	24.70	4.26	71.23	23.38	48 - 95	533.20	1'114.30
0.1	24.69	4.26	71.23	23.38	48 - 95	357.53	933.02
10	22.90	3.83	70.49	22.96	48 - 93	63.97	53.70
50	18.95	3.22	68.95	19.11	50 - 88	37.98	17.61
200	12.29	2.01	63.39	12.25	51 - 76	22.42	13.76

Tabla 17. Estadísticos para el método RK2 en el grupo V (CYL + 9 AEMet).

En la figura 21, se presentan las relaciones entre el %RMSE y los vecinos cercanos. La figura 21 (a) muestra la relación entre el %RMSE y el vecino más cercano para el umbral ISG mayor de 200. En este caso, se aprecia que el efecto “nugget” es de aproximadamente del 11 % de %RMSE, y en toda el área de trabajo, el valor de %RMSE se mantiene en término medio por debajo del 20.22 %, independientemente de la distancia al vecino más cercano (desde los 7.3 km hasta los 108.4 km). Además la mayoría de los puntos (a excepción de 3), se encuentran por debajo el 15 % de %RMSE.

Para este umbral ( $ISG > 200 \text{ W/m}^2$ ), la relación entre el %RMSE y la distancia media de todas las estaciones vecinas sigue una tendencia similar al de la relación entre %RMSE y la distancia a la estación vecina más cercana, al igual que ocurrió para el caso de la figura 20 (b). Este es uno de los motivos por los que aquí la figura 21 (b) muestra la relación entre el %RMSE y el vecino más cercano para el umbral  $> 50 \text{ W/m}^2$ . El otro motivo es que todos los puntos (a excepción de dos) se encuentran en término medio por debajo del 23.68 % de %RMSE independiente de la distancia al vecino más cercano.

Cabe destacar que para los umbrales 50 y 200  $\text{W/m}^2$ , el método RK2 para el grupo V (CYL + 9 AEMet), no supera el límite de 25 % de %RMSE después de los 25 km y 34 km para observaciones horarias indicado por Zelenka *et al.* (1999) y Pérez *et al.* (1997).

Todos estos valores antes mencionados, representan una mejora respecto al método (RK2 grupo IV), evidenciando la importancia de disponer más estaciones para realizar las interpolaciones (59 estaciones para el grupo V contra 50 para el grupo IV, ver tabla 13).

La figura 21 (c), al igual que en la figura 20 (c), presenta la relación entre el %RMSE para cada estación y su vecino más cercano sin umbralizar. En esta figura se aprecia como la inclusión de valores de la ISG menores a  $200 \text{ W/m}^2$  o menores a  $50 \text{ W/m}^2$ , aumentan el efecto “nugget” hasta 21-23 %. En la figura 21 se presentan un amplio número de puntos (el 42 % de estos), contenidos entre los rectángulos rojos, que sobrepasan el umbral del 25 % de %RMSE. De esos puntos, la mitad (el 22 %) mostrados en el rectángulo (c-I) se encuentran dentro de los 34 km, y a su vez, la mayoría de estos (el 19 %) se encuentran dentro de los 25 km. Por lo que al igual que para (RK2 en grupo IV), es plausible asumir que haber superado el umbral del 25 % de %RMSE depende de los valores bajos de la ISG y no de la distancia a su vecino más cercano.

La figura 21 (d) es ilustrativa del comportamiento del método SAT, ya que son las estimaciones de la ISG derivadas por satélite (EISG) para el grupo V. En este caso el método SAT ocupa el último puesto para los estadísticos sin umbralizar (ver tablas 11 y 12). Como es de esperar, en esta figura se aprecia que no existe una relación entre la separación de los puntos y el %RMSE de la EISG. De igual forma se aprecia que estos valores %RMSE son mayores que los del método RK2 en la figura (c).

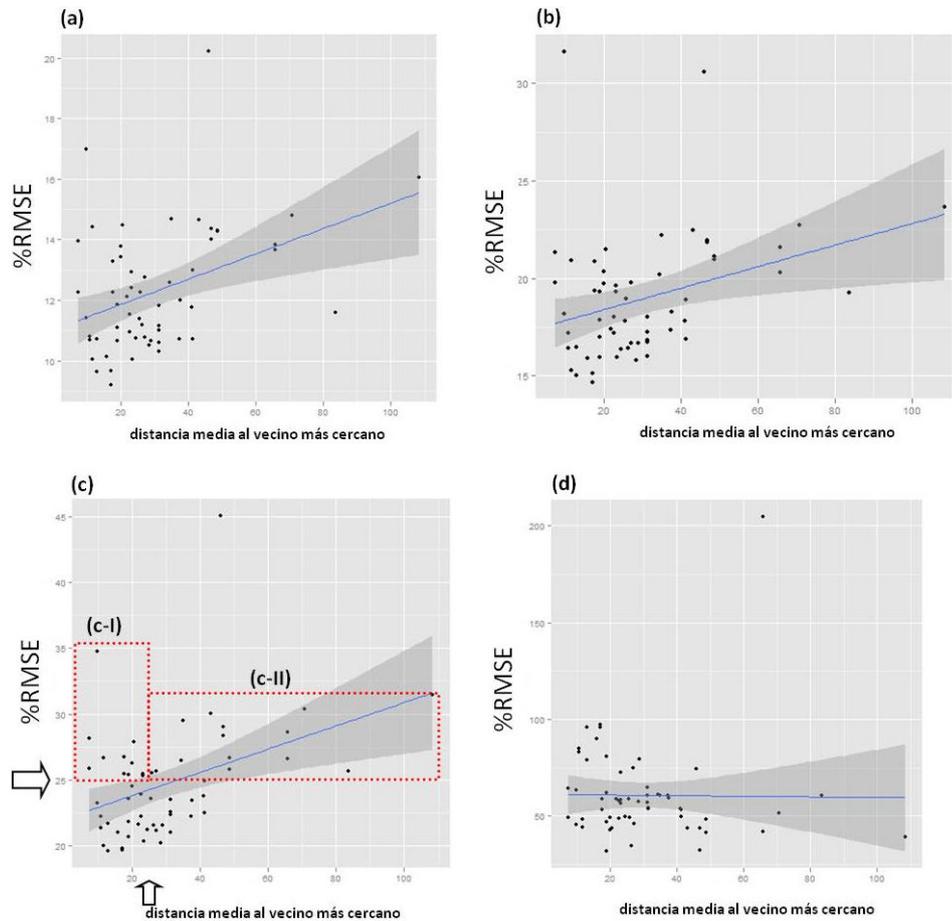


Figura 21. %RMSE en relación a los vecinos para el método RK2 en grupo VI (CYL + 9 AEMet). (a) %RMSE en función al vecino más cercano para umbral 200 W/m<sup>2</sup>. (b) %RMSE en función a la distancia media entre los vecinos para umbral 200. (c) %RMSE en función al vecino más cercano sin umbralizar. (d) %RMSE de las Estimaciones derivadas por satélites en función al vecino más cercano sin umbralizar.

### SAT en la forma de agrupación VI (10 AEMet)

En la tabla 18 se presentan los estadísticos para el método HelioSat2 en la forma de agrupamiento (VI- 10 AEMet) para los 6 umbrales de la ISG (todo, 0.1, 10, 50, 200) W/m<sup>2</sup>. Se puede apreciar, de modo similar a los métodos de RK2 anteriores que las medias relativas disminuyen a medida que aumentan los niveles de umbral; Sin embargo, con estas estimaciones, las desviaciones estándar además de ser las más bajas, aumentan a medida que suben los niveles de umbral (contrario a lo que ocurre en los métodos RK2 de las secciones anteriores). A pesar de ello el valor más alto de stdv (1.27) % es menor que el más bajo de los casos anteriores, 1.35 % (RK2 grupo IV). Respecto al *benchmarking* en (SoDa-is 2013b), se puede destacar

que aunque los %RMSE para umbrales 0.1 y 10 W/m<sup>2</sup> son muy parecidos, estos valores para los umbrales 50 y 200 W/m<sup>2</sup> son más bajos que los del citado *benchmarking* y parecidos a los de los casos RK2 de las agrupaciones IV y V.

<i>Umbral</i>	<i>%RMSE</i> (%)	<i>Stdv %RMSE</i> (%)	<i>RMSE</i> (Wm <sup>2</sup> )	<i>Stdv RMSE</i> (Wm <sup>2</sup> )	<i>Rango RMSE</i> (Wm <sup>2</sup> )	<i>%RMSE</i> (All)	<i>Stdv %RMSE</i> (All)
NA	22.85	0.87	105.13	15.67	89-121	713.09	823.02
0.1	22.84	0.87	105.13	15.68	89 -121	500.27	476.25
10	21.55	1.03	104.64	15.78	89 -120	84.05	33.26
50	18.04	1.27	103.59	15.12	88 - 119	46.06	8.14
200	11.46	1.25	101.21	15.91	85 - 117	23.13	2.64

Figura 18. Estadísticos para las Estimaciones de la ISG derivadas de imágenes de satélite en el grupo VI (10 AEMet).

Aunque el método ganador el grupo VI (10 AEMet) es SAT, es previsible que el %RMSE de las estimación de la ISG derivadas por satélite (EISG) no dependan de las distancias entre estaciones, esto se debe a que las EISG del Método HelioSat2 no dependen de las observaciones en tierra. Al igual que ocurre para los grupos IV y V (ver figuras 20 (d) y 21 (d)).

Sin embargo, respecto a los valores %RMSE, se puede destacar que las EISG presentan parecidos resultados a los del método RK2 en los grupos IV y V en todos los umbrales. Una posible causa del parecido es que en este grupo sólo se incluyen las estaciones de la red AEMet.

#### 4.3.3 Análisis de una primera aproximación a la validación de estaciones de voluntarios

El análisis detallado del mejor método de estimación para los tres grupos (IV, V y VI) en la sección anterior, ha permitido descartar al método/grupo IV en favor del método/grupo V. Esto se debe a dos motivos:

- (1) El método RK2, en el grupo V, mejora la relación %RMSE respecto a la distancia al vecino más cercano al incluir además de las 50 estaciones de CYL las 9 estaciones de AEMet.
- (2) El área V contiene en su interior al área del grupo IV. Por lo tanto para la validación de las estaciones voluntarias en toda el área de estudio se utiliza como el mejor método/grupo, la unión entre el método RK2 en el grupo V, y al método SAT en el grupo VI, como se muestra en la ecuación 16.

$$\text{método/grupo para toda el área} = \text{RK2}(\text{en grupo V}) \cup \text{SAT}(\text{en grupo VI}) \quad (16)$$

En la figura 22 se presenta la combinación del mejor método/grupo, con el método RK2 en el grupo V representado con líneas verticales y el método SAT en el grupo VI por líneas horizontales. En esta figura también se aprecian las estaciones que forman parte de las interpolaciones para el grupo V, que son las estaciones de AEMet (símbolos rojos) más las estaciones de JCYL (símbolos azules). Para ambos grupos también se aprecian las estaciones de voluntarios que serán validadas con sus respectivos métodos ganadores (símbolos verdes).

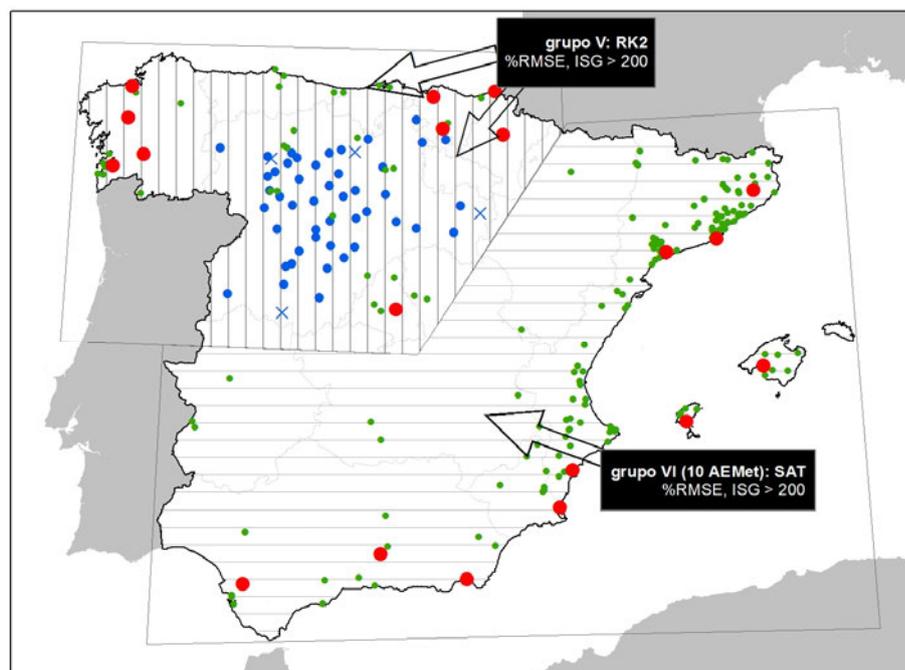


Figura 22. El mejor método/área de estudio como la combinación de dos sub-grupos.

Después de identificar el método/grupo a aplicar en toda el área de estudio (ver ecuación 16), las EM de voluntarios fueron validadas considerando dos restricciones bajo las cuales los métodos de estimación presentan resultados más precisos. Estas restricciones son:

- (1) Consideración de valores de la IS con umbral  $ISG > 200 \text{ W/m}^2$ : Esto ha sido hecho en base a los resultados de los experimentos previos, en los que se encontró que en este nivel de umbral los métodos de estimación presentan los mejores niveles de error y desviación estándar.
- (2) Validación sólo en días claros de acuerdo a la clasificación de “tipos de días” propuestos por (Iqbal 1983). Esta clasificación utiliza la magnitud del índice de claridad diario, denotado  $K_T$  (Liu y Jordan 1960), en la ecuación (17), en el cual se define un día claro como aquel donde  $K_T > 0.7$  (Janetz y Kudish 2008), con  $K_T$  definido como la proporción entre la Irradiación Global Horizontal Diaria ( $H_G^d$ ) y la Irradiación Extraterrestre Diaria sobre una superficie Horizontal ( $H_0^d$ ).

$$K_T = \frac{H_G^d}{H_0^d} \quad (17)$$

La  $H_G^d$  fue integrada a intervalos de 15 minutos utilizando el método trapezoidal implementado en el paquete *pracma* para R. Los cálculos subyacentes con respecto a la  $H_0^d$ , han sido realizados con el paquete para R *SolarR* (Perpiñán 2012).

Se ha identificado una fuerte correlación negativa entre  $K_T$  y %RMSE en los métodos de estimación espacial de la ISG:

- El método RK en el Grupo V presenta una fuerte correlación negativa con un valor de  $R = -0.928$ .
- El método SAT en el Grupo VI también presenta una fuerte correlación negativa con un valor de  $R = -0.81$ .

La figura 23 presenta diagramas de dispersión entre  $K_T$  y %RMSE, así como las respectivas curvas e índices de regresión. La figura 23(a) presenta el caso del método RK2 en el grupo V, mientras que la figura 23(b) ilustra el caso del método SAT en el grupo VI.

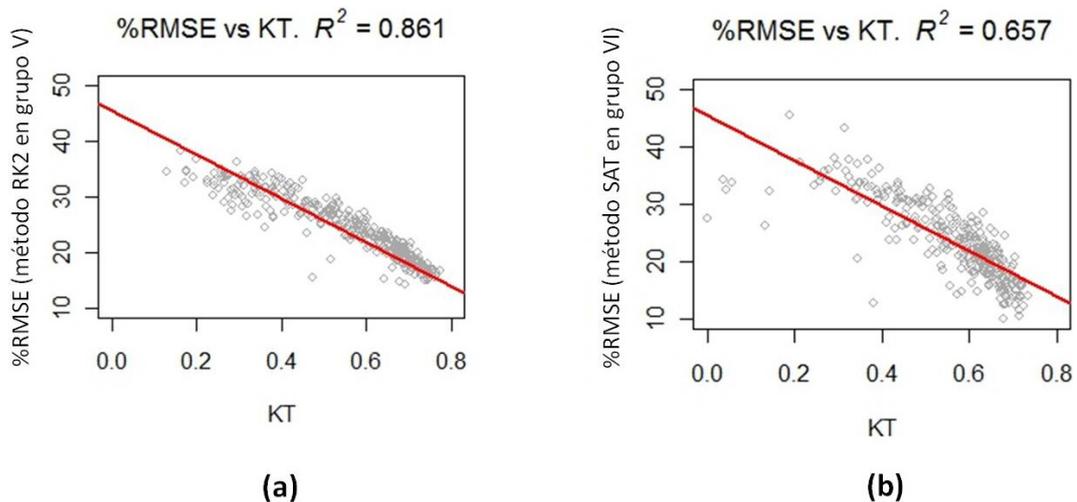


Figura 23. Relación entre el  $K_t$  y %RMSE diarios.

Una vez aplicadas las dos condiciones de validación, se procedió a analizar la cantidad de días en los que las estaciones REMV presentan niveles de error entre la media y  $\pm 2$  \* desviación estándar (mean, 2 stdv) y  $\pm 3$  \* desviación estándar (mean, 3 stdv) respecto a los modelos de estimación utilizados.

De las 31 estaciones EM de voluntarios incluidas en el método RK2 del grupo V, se encontró lo siguiente:

- 8 estaciones (26 %) presentan un nivel de error contenido dentro de 2 stdv.
- 11 estaciones (35 %) presentan un nivel de error contenido dentro de 3 stdv.

De las 177 estaciones EM de voluntarios incluidas en el método SAT en el grupo VI, que contiene se encontró que:

- 130 estaciones (74 %) presentan un nivel de error contenido dentro de 2 stdv.
- 135 estaciones (77 %) presentan un nivel de error contenido dentro de 3 stdv.

## 4.4 Resumen

Esta parte de la investigación cubrió el objetivo de identificar el mejor método para estimar la ISG en periodos de 15 minutos en el territorio Peninsular español y el archipiélago Balear, basado en interpolaciones aplicadas a las observaciones de dos redes de estaciones oficiales y contemplando la mejor forma de agrupar las estaciones en el área de estudio. Una aplicación práctica de este objetivo de investigación, ha sido comparar las observaciones de la ISG hechas por las EM de voluntarios (Meteoclimatic) con el modelo de estimación aplicado; esto ha sido realizado para identificar las EM que están dentro del margen de error (media + - 2 desviaciones estándar).

Para realizar los experimentos, se utilizaron los datos descargados que se corresponden al año 2011 en el territorio Peninsular español y en las Islas Baleares procedentes de tres REM descritas en el capítulo 3, dos REM oficiales (AEMet y Castilla y León) y una REMV (Meteoclimatic) que suman un total de 244 estaciones. Adicionalmente se utilizan las estimaciones de la IS derivadas a partir de imágenes de satélite, procedentes de la base de datos HelioClim3. Debido a su estatus de oficiales, las dos primeras REM fueron utilizadas para el propósito de identificar el mejor método para la estimación IS Global en el área de estudio y la tercera se utilizó en una aplicación práctica de mejor método identificado.

Con el propósito de encontrar el mejor método de estimación espacial de la IS para periodicidad de 15 minutos en el área de estudio y al considerarse la amplia extensión geográfica del estudio, se analizaron cinco técnicas de predicción espacial en seis configuraciones geográficas que agruparon a las EM, generándose así un total de 30 combinaciones experimentales. Los criterios que se plantearon para la selección de las seis agrupaciones fueron, la densidad de estaciones, y la mezcla/separación de estaciones pertenecientes a distintas redes. De las cinco técnicas de predicción espacial, 4 son interpolaciones que se han de calcular y la quinta serán las estimaciones de la IS obtenidas desde la base de datos HelioClim (estos últimos datos también se utilizarán como variable auxiliar de dos de las técnicas de interpolación). Las cuatro técnicas de interpolación seleccionadas fueron: Distancia Inversa Ponderada (*Inverse Distance Weighting*: IDW), Kriging

Ordinario (*Ordinary Kriging*: OK), y dos formas de Kriging con Regresión (*Regression Kriging*: RK) que se diferenciaron en las variables auxiliares utilizadas para cada una.

Se definió el mejor método de estimación ISG basado en el Error Medio Cuadrático Porcentual de las diferencias entre los valores observados y los estimados. Se encontró que el mejor método para estimar la ISG cada 15 minutos es el método de interpolación de Kriging de Regresión que utilice entre sus variables de entrada a la IS estimada a partir de imágenes de satélite, siempre y cuando el área de trabajo cuente con una densidad adecuada de estaciones. En caso contrario, cuando la densidad de las estaciones es baja, el mejor método es la estimación directa de la IS a partir de imágenes de satélite.

Una vez conocido el margen de error del mejor método y teniendo en cuenta que las observaciones registradas por las EM de voluntarios, en especial la ISG no han sido validadas, se procedió realizar una primera aproximación para proponer validarlas, llevándose a cabo una comparación entre las ISG observada por las EM de voluntarios y el mejor método de estimación espacial de la ISG identificado, etiquetando a aquellas estaciones que presentaban niveles de error dentro del método de referencia. Esto permitió identificar que el 67% de las EM de voluntarios analizadas (163 estaciones) presentan valores dentro del margen de error (media  $\pm$  2 desviaciones estándar), lo cual representan más del doble de las 69 estaciones oficiales utilizadas.

## 5 PREDICCIÓN ESPACIO-TEMPORAL DE LA ISG

Cuando se trabaja con contextos espaciales reducidos y en el corto plazo de tiempo (pequeña escala espacio-temporal), la Irradiancia Solar Global (ISG) se ve afectada por las condiciones dinámicas de la atmósfera y un alto ruido debido principalmente a la nubosidad. Estas condiciones presentan relaciones no lineales que vuelven compleja la predicción a corto plazo de la ISG. Las Redes de Neuronas Artificiales (RNA) forman parte de la Inteligencia Artificial (IA) y han demostrado su utilidad para investigar modelos asociados a procesos naturales que aproximan funciones no lineales, como es el caso de la predicción de la ISG a corto plazo (ver sección 2.4).

El objetivo principal de este capítulo de tesis ha sido investigar la predicción de la ISG con horizontes temporales de hasta 6 horas, mediante el modelado de RNA, incluyendo la componente espacial como entrada a los modelos (información de las estaciones vecinas). Los experimentos realizados mediante distintos modelos de RNA y la inclusión de distintas entradas se han diseñado para responder a las siguientes preguntas de investigación:

- ¿Se mejora el error de la predicción temporal de la IS a corto plazo en base a RNA con la inclusión de la mayor cantidad de datos de entradas en los modelos?
- ¿Cómo se relacionan el horizonte espacial y temporal en la predicción de la IS a corto plazo con las RNA?

### 5.1 Sitio de estudio y datos experimentales

Al igual que los trabajos relacionados (Voyant *et al.* 2014; Wang *et al.* 2012; Mellit y Pavan 2010), en esta investigación se realiza la predicción de la ISG a corto plazo sobre un punto fijo (objetivo de predicción), sin embargo la innovación del método reside en incluir la componente espacial en los datos de entrada indirectamente mediante el uso de datos en paralelo procedentes de estaciones

vecinas al punto objetivo. Esta innovación y por tanto característica diferenciadora ha supuesto un trabajo adicional en la preparación de los datos de entrada para el entrenamiento de las RNA.

### **5.1.1. Sitio de estudio**

La selección de las Estaciones Meteorológicas (EM) a utilizar ha sido condicionada al uso de Redes de Estaciones Meteorológicas (REM) oficiales, a la distribución geográfica y a la densidad de estaciones con sensores de la ISG en ellas contenidas. Esto implicó descartar el uso de la red Meteoclimatic, dejando la decisión entre la REM de la Junta de Castilla y León (CYL) sobre la red de AEMet. El criterio de distribución geográfica y la densidad se basó en el cálculo de densidad de superficie promedio que cubre cada estación de las distintas redes presentado en la tabla 13 (ver mapa en la figura 15). Por todo lo anterior se seleccionó la REM de CYL como la fuente de datos a utilizar en este tipo de predicciones.

Una vez seleccionada la REM de CYL, el siguiente paso fue la selección de la EM objetivo de la predicción. Las consideraciones fueron las mismas que para la selección de la red (densidad y distribución geográfica). Para esto se seleccionaron 10 estaciones cercanas al centro de gravedad de CYL (donde a simple vista se observa que las estaciones están más cercanas entre sí), seguidamente se analizó la cantidad de estaciones vecinas a cada una de ellas en un radio máximo de 100 km. Se encontró que dos tuvieron el número máximo de 30 estaciones vecinas, con lo cual se procedió a seleccionar la más cercana al centro de gravedad.

En la figura 24, se presenta un mapa con la REM de CYL. Como mapa base (fondo cartográfico) se puede apreciar el relieve de esta comunidad autónoma así como que la mayor parte de ella está conforma por una meseta (con una altura media de 800 m.). En el mapa, la estación objetivo de la predicción está representada por el círculo de color verde, mientras que el cuadrado azul representa la segunda estación con 30 estaciones en un *buffer* de 100 km, el resto de las estaciones cercanas al centro de gravedad están representadas por los

triángulos de color rojo y el resto de las estaciones están representadas por los círculos de color azul. Los círculos verdes y azules, representan un *buffer* de 100 y 55 km respectivamente en torno al objetivo de la predicción.

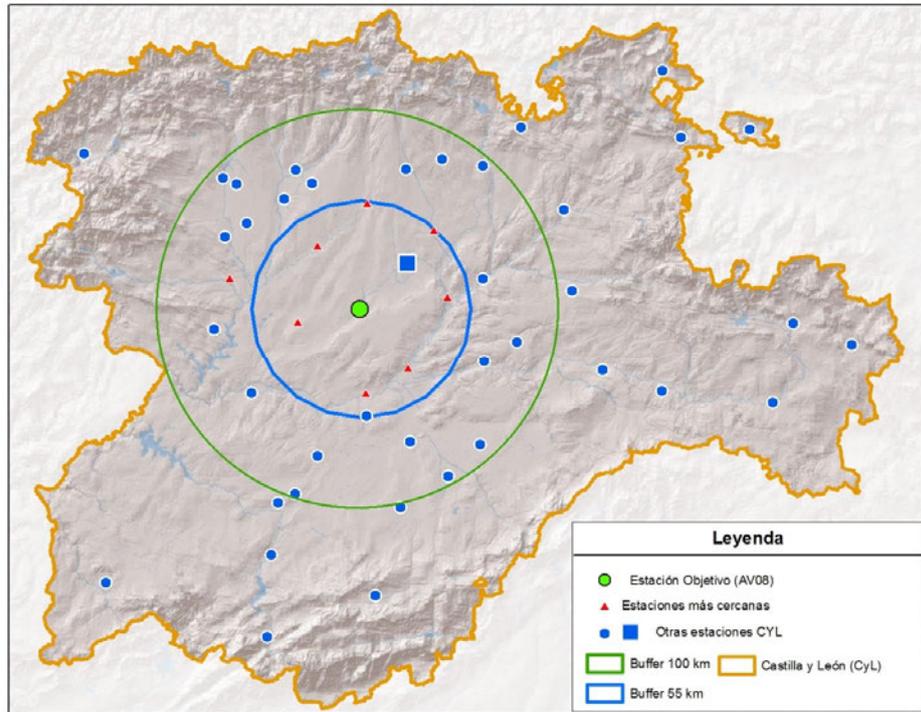


Figura 24. Límite administrativo de la comunidad Autónoma Junta de Castilla y León y REM automáticas pertenecientes a la Red InfoRiego.

### 5.1.2. Datos experimentales

Los datos experimentales utilizados son las observaciones del año 2011 obtenidos desde esta REM de CYL (en la sección 4.1.2 se explicó de forma general las características de esta red). De todas las observaciones disponibles se utilizan cinco variables meteorológicas: (1) Irradiancia Solar Global (ISG), (2) Temperatura Ambiente (TA), (3) Humedad Relativa del Aire (HRA), (4) Dirección del Viento (DV) y (5) Velocidad del Viento (VV). Adicionalmente se agregaron 4 variables relacionadas calculadas de un modo determinista referidas al instante de tiempo exacto de la toma de las observaciones meteorológicas, siendo éstas: (6) la Irradiancia Solar Extraterrestre Horizontal (ISEH), (7) el Índice de Claridad Instantáneo (KTi), (8) Distancia al Medio Día Solar (DMDS) y (9) el Ángulo Cenit

(AC). Estas variables están relacionadas entre sí. Por ejemplo el KT se puede obtener desde la ISG y la ISEH; de igual forma el mayor valor de la ISG y la ISEH se van a dar cuando la DMDS sea cercana a cero, y a la vez los valores de esas tres variables tienen correlación con el Ángulo Cenit. Sin embargo este es uno de los objetivos de los experimentos, usar la mayor cantidad de variables posibles, por dos razones: primero porque los estudios de referencia han focalizado en el uso de menos variables o de la inclusión de un procesamiento previo de sus estadísticos, y segundo porque las RNA lo permiten. Por tanto la variante propuesta en esta tesis puede servir para futuras comparaciones.

Los datos experimentales se encuentran almacenados en un Sistema de Gestión de Base de Datos Relacional (SGBDR) de tipo Oracle 11.2 con extensión espacial, en el que han sido cargadas las observaciones desde los robots webs y se organizan según los estándares de normalización de bases de datos. Así entre otras características, cada registro (fila) representa las observaciones de las diferentes variables meteorológicas (columnas) para un instante de tiempo de una estación.

## **5.2 Metodología**

Los pasos metodológicos utilizados para investigar las predicciones de la ISG a corto plazo mediante RNA han sido influenciados por las características asociadas al objetivo de experimentar con la inclusión de un gran número de entradas en estas RNA. Se ha intentado mantener como entradas la mayor cantidad de datos procedentes de distintas variables, en diferentes instantes de tiempos previos, así como las observaciones en paralelo de estaciones vecinas.

### **5.2.1. Preparación de los datos**

La obtención de conjuntos de datos para una ubicación en concreto, junto a las 9 variables antes mencionadas más otras características opcionales asociadas a la estación (por ejemplo: latitud, longitud y altitud) es facilitado por las capacidades de manipulación y almacenamiento que permiten los SGBDR. Sin embargo, la

obtención de juegos de entrenamiento y test con todas las entradas de las distintas estaciones en sus distintos instantes temporales organizados en forma de variables independientes (representados en una sola fila) no es una tarea trivial.

Esta etapa de la metodología implicó el diseño de un esquema en el que se representase a todos los datos como un solo juego de datos de salida (una sola fila de salida) y su implementación mediante un conjunto de vistas. El resultado final sigue el esquema presentado en la figura 25:

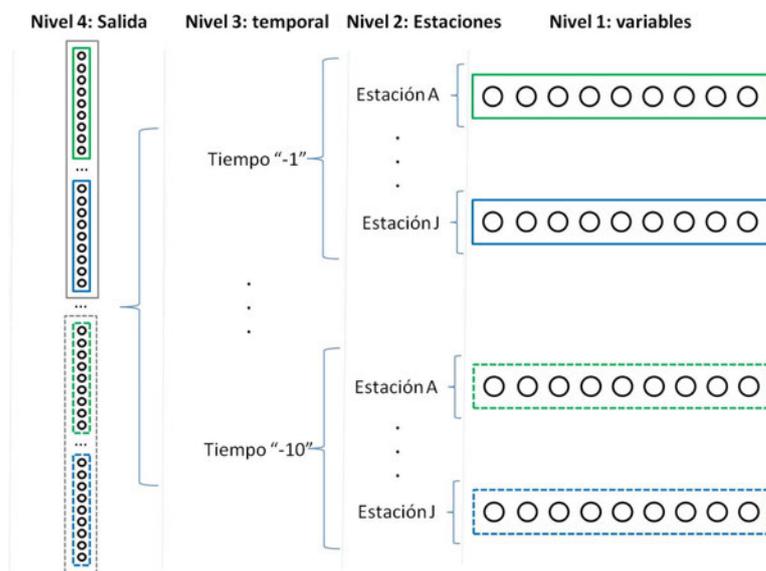


Figura 25. Esquema de organización de los datos de entrada de forma horizontal.

En la figura 25 se pueden apreciar los distintos niveles de agrupamiento implementados para alcanzar una salida que represente a todos los datos de entrada de forma unificada. El primer nivel se corresponde al de las variables de cada estación; este es el nivel más básico y se obtiene directamente desde las tablas de almacenamiento en la Base de Datos. El segundo nivel agrupa las variables de las 10 estaciones en un orden predefinido (primero la estación A, hasta la J), este orden se mantiene en los siguientes niveles de agrupamiento. El tercer nivel se corresponde con la agrupación temporal en la que todas las observaciones de todas las estaciones para un mismo instante de tiempo se agrupan según el orden predefinido previamente.

### 5.2.2. Normalización y selección de datos de entrenamiento y de test:

La normalización de los datos es un requerimiento necesario para la convergencia de los modelos de RNA y mejora su rendimiento. Todos los datos se normalizaron a  $[0, 1]$  según la ecuación (18). Donde,  $x \in [x_{min}, x_{max}]$  y,  $y \in [y_{min}, y_{max}]$ ; donde “ $x$ ” es el dato original y “ $y$ ” es el su valor normalizado; Asumiendo los valores de  $y_{min} = 0$ ;  $y_{max} = 1$ .

$$y = \frac{(y_{max} - y_{min})(x - x_{min})}{x - x_{min}} + y_{min} \quad (18)$$

Todos estos procesos se realizaron desarrollando funciones con el lenguaje de programación y software estadístico R (R Development Core Team, 2012). La carga de los datos desde la Base de Datos se realizó con el paquete *rodbc* (Ripley y Lapsley 2014). Los datos de entrenamiento y test se dividieron en 70% y 30% respectivamente. Para extraer un porcentaje de datos para el entrenamiento y otro para los test se utilizó el paquete *caret* (Kuhn 2014) que selecciona una muestra al azar los datos de cada subconjunto considerando la distribución de una variable, utilizándose como variable la ISG.

### 5.2.3. Parámetros de entrenamiento de las RNA

Con el propósito de determinar la mejor configuración de RNA aplicada a una arquitectura de tipo Perceptrón Multicapas (*Multilayer Perceptron*: MLP), manteniendo la mayor cantidad de entradas posibles, se intentaron estudiar algunos parámetros que influyen en el aprendizaje y convergencia de la red como lo son: (i) el número de capas ocultas, (ii) la cantidad de neuronas en cada capa, (iii) el intervalo de normalización de los datos, (iv) el ratio de aprendizaje, (v) la cantidad de veces que se entregan todos los ejemplos a la RNA para su entrenamiento (*Epochs*). Otros parámetros se dejaron fijos, basados en los buenos resultados que estos ya han demostrado en estudios previos: (vi) el algoritmo de entrenamiento utilizado, ha sido *BackPropagation*, (vi) la función de comparación seleccionada ha sido la “suma de los errores al cuadrado”.

La selección de los parámetros adecuados se llevó a cabo mediante un proceso iterativo utilizando el software (*System for Automated Learning Modeling Operative Networks: SALMON*) (Serradilla 2013). SALMON es un programa desarrollado por uno de los directores de esta tesis especializado en el entrenamiento y simulación de RNA que ofrece una interfaz de usuario potente, hecho que facilitó sustancialmente el proceso de entrenamiento de las redes. Aunque las RNA ya quedan identificadas mediante SALMON, el siguiente paso fue elaborar las mismas con el paquete *nerualnet* (Fritsch y Guenther 2014) envolviéndolas mediante funciones desarrolladas con lenguaje R (R Development Core Team 2012) que facilitasen su entrenamiento, análisis e interpretación de los errores generados por los modelos de RNA. Esta segunda implementación alternativa de RNA bajo R se justifica en el hecho de que, al embeberse dentro de R, se habilita el uso directo de las funciones estadísticas y gráficos de este que facilitan la interpretación de las capacidades de entrenamientos. Por otro lado, se intentó encontrar los parámetros de las RNA directamente desde *nerualnet*, y aunque es posible hacerlo, se prefirió continuar con el uso de ambos programas, porque con el segundo el proceso inicial del entrenamiento de RNA presenta algunos inconvenientes, entre los que cabe destacar que si el algoritmo de entrenamiento no converge, o si se detiene la red por estar utilizando mucho tiempo en el entrenamiento, la RNA no se crea ni siquiera para el nivel de error en que se encontraba en el momento de su detención.

#### **5.2.4. Modelos de RNA y estadísticos para medir la calidad de las predicciones**

Los modelos de RNA creados se concentraron en la predicción futura de la ISG en horizontes temporales desde 1 hasta 6 horas, creándose de forma independiente una RNA por cada horizonte temporal. En todos los casos, los datos de entrada utilizaron información de las distintas variables en instantes de tiempo previos, lo que se conoce como el método de ventana deslizante, como se puede apreciar en la figura 26.

En esta figura se muestra como la predicción del valor futuro de la ISG depende de un conjunto de variables de entrada correspondientes a una ventana de tiempo previa. En intervalo temporal ( $\Delta t$ ) que se deja entre el valor futuro a predecir y el instante previo más cercano al  $t$  de predicción, es el que define el horizonte temporal de la predicción.

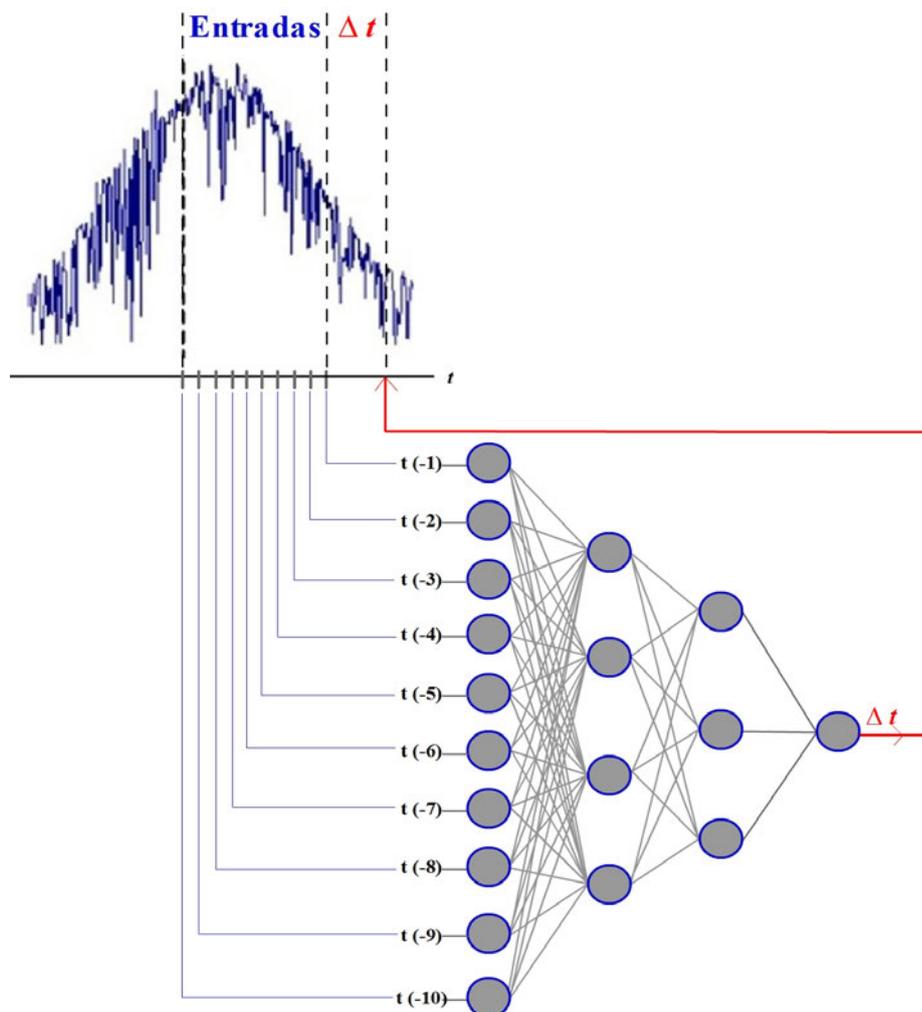


Figura 26. Ilustración del método de ventana deslizante como forma de entrada a las RNA. Figura adaptada desde Voyant *et al.* (2013)

Los modelos de RNA aquí creados se pueden agrupar en dos tipos, que se diferencian entre sí por la utilización, o no, de datos procedentes de estaciones vecinas. En ambos casos, el esquema de datos de entrada es el mismo que se definió en el paso metodológico en 5.2.1. La diferencia reside en no utilizar los datos de estaciones vecinas, en cuyo caso se discriminan las entradas que no se corresponden con la estación objetivo.

En la ilustración de la figura 27 se intenta representar la arquitectura general simplificada de las RNA creadas para la predicción a corto plazo de la ISG en el punto objetivo utilizando como entrada información espacio-temporal. Aquí se puede apreciar cómo llegan a las entradas de la RNA datos de estaciones vecinas, y valores de sus variables para distintos intervalos de tiempo mediante (método de ventana deslizante), también se aprecia que éstas RNA poseen dos capas ocultas de neuronas, y como salida se obtiene el valor de la ISG predicho para un horizonte  $t_i$ .

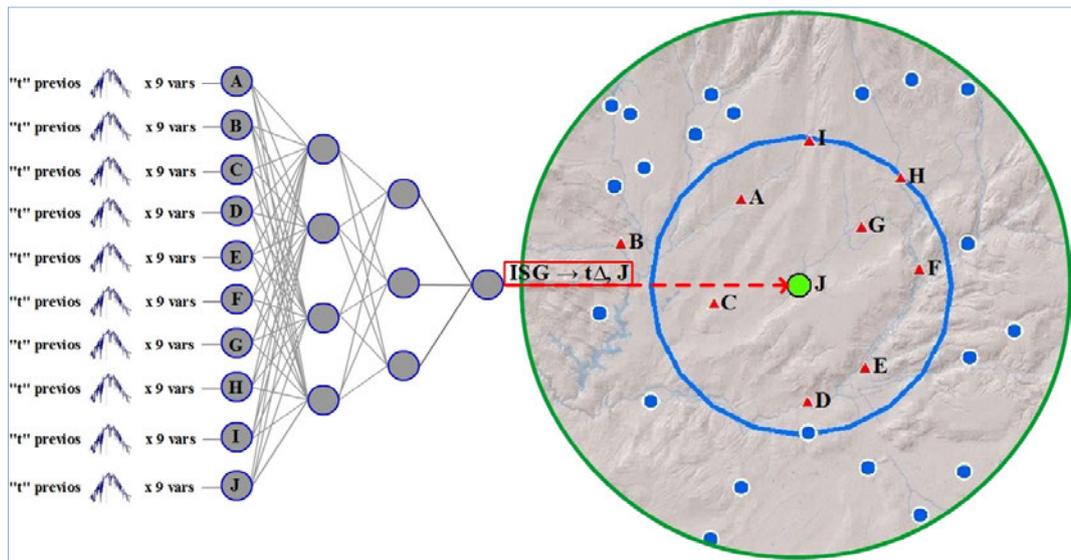


Figura 27. Representación simplificada de una RNA para la predicción la ISG en un horizonte “ $\Delta t$ ” futuro, en base a datos de distintas estaciones vecinas y de sus variables en diferentes instantes de tiempo.

La calidad de las predicciones de las distintas RNA creadas, se llevó a cabo utilizando cuatro estadísticos: RMSE, %RMSE, %MAE, nRMSE. De estos, los tres primeros ya han sido definidos en la sección 4.2.2, en las formulas 6, 7 y 9 respectivamente. El nRMSE se define en la ecuación 19. Donde  $Y_i$  es el valor observado de la ISG, mientras que  $\hat{Y}_i$  es el valor predicho por la RNA. Se seleccionaron estos estadísticos para poder hacer comparaciones con los estudios previos.

$$nRMSE = \sum_{i=1}^n \sqrt{\frac{(Y_i - \hat{Y}_i)^2}{Y_i^2}} \quad (19)$$

## 5.3 Resultados y discusión

### 5.3.1. Parámetros para la selección de las arquitecturas de RNA

Una de las actividades más complejas en la creación de estos modelos de RNA además de la organización de los datos de entrada, ha sido el proceso de identificación de los parámetros adecuados para el correcto entrenamiento de la red. Este fue un proceso iterativo en el que se probaron distintas arquitecturas y configuraciones de red (ver sección 5.2.3). En esta sección se presentan los resultados del Error Medio Cuadrático Porcentual (*Percentage Root Mean Square Error: %RMSE*) de las validaciones cruzadas (utilizando los datos en test) de 50 de estos experimentos, distribuidos en distintas tablas.

En este caso el objetivo ha sido la identificación de los parámetros más prometedores para posteriormente utilizarlos en el entrenamiento de nuevas RNA que optimizaran el resultado de sus predicciones (los resultados de dichas RNA optimizadas se presentan en la siguiente sub-sección). Dicho de otra forma, estos experimentos permitieron descartar los parámetros de entrenamiento y arquitecturas de RNA que no generaban resultados adecuados.

#### **Parámetros y RNA para horizontes temporales muy cortos (15 minutos)**

En la tabla 19 se presentan 6 RNA para la predicción de un horizonte temporal de 15 minutos. Las RNA (1-3) lo hacen mediante la inclusión de estaciones vecinas mientras que las RNA (4-6) realizan la predicción sólo en base a las observaciones de instantes previos de la misma estación objetivo. En ambos casos se puede apreciar, a simple vista, que el %RMSE es en todos los casos inferior al 10%. También se ve que las RNA que incluyen datos de estaciones vecinas presentan una leve mejora en los niveles de errores, sin embargo la gran similitud entre los niveles de errores entre ambos tipos de redes llevan a interpretar que para este horizonte temporal tan inmediato predominan en los modelos de RNA los datos de instantes previos de la misma estación objetivo.

Otra característica a destacar es que la RNA de tipo Perceptron Simple (sin necesidad de agregar capas ocultas), es la que presenta los mejores resultados. Esto implica que en este horizonte temporal, la predicción de la ISG es un problema lineal con respecto a las entradas. Sin embargo, cabe destacar que a pesar de las relaciones entre las distintas variables independientes (sección 5.1.2) se observa como las RNA son capaces de generar modelos de predicción adecuados, distinto a lo que ocurriría con el uso de Modelos de Regresiones Lineales Múltiples donde primero se tendría que eliminar las variables con altos niveles de colinealidad.

Red #	Arquitectura de Red	%RMSE en TEST
1	<b>PERCEPTRON</b>	<b>6.26%</b>
2	20HL1	6.44%
3	50HL1	9.24%
4	<b>PERCEPTRON</b>	<b>6.44%</b>
5	10HL1	<b>7.01%</b>
6	20HL1	<b>9.77%</b>

Tabla 19. Identificación de parámetros para RNA que predicen la ISG: Pruebas con horizonte de 15 minutos.

### **Parámetros y RNA para horizonte temporal de 1 hora**

Los resultados de las 15 RNA que realizan predicciones de la ISG a un horizonte temporal de 1 hora se presentan en la tabla 20. En esta tabla, la primeras RNA (7-14) realizan la predicción utilizando datos de estaciones vecinas, mientras que las últimas (15-21) lo hacen con sólo observaciones de instantes previos de la misma estación objetivo. En ambos casos, se puede apreciar que los mejores niveles de error se encuentran con un MLP de una capa oculta. También se observa cómo el uso de dos capas ocultas asemeja los niveles de errores al del Perceptron Simple, sin embargo cabe destacar que tanto el Perceptron simple como las otras configuraciones de RNA mantienen niveles de errores no muy alejados entre sí, lo cual puede ser un indicio de que para este horizonte temporal el problema aún tiene una fuerte relación lineal.

Red #	Otros parámetros	Arquitectura de Red	% RMSE en TEST
7	Delta-Deriv = 0.03	PERCEPTRON	24.62%
8	Delta-Deriv = 0.05	PERCEPTRON	25.46%
9	Delta-Deriv = 0.1	PERCEPTRON	25.78%
<b>10</b>	<b>Delta-Deriv = 0.1</b>	<b>10HL1</b>	<b>23.99%</b>
11	Delta-Deriv = 0.1	50HL1	25.70%
12	Delta-Deriv = 0.05	10HL1	25.70%
13	Delta-Deriv = 0.05	50HL1	28.91%
14	Delta-Deriv = <b>0.05</b>	10HL1-5HL2	25.10%
15	Delta-Deriv = 0.03	PERCEPTRON	26.25%
16	Delta-Deriv = 0.02	PERCEPTRON	26.08%
17	Delta-Deriv = 0.05	PERCEPTRON	26.29%
<b>18</b>	<b>Delta-Deriv = 0.025</b>	<b>30HL1</b>	<b>25.77%</b>
19	Delta-Deriv = 0.05	30HL1	25.97%
20	Delta-Deriv = 0.025	10HL1-5HL2	25.82%
21	Delta-Deriv = 0.025	10HL1-5HL2	25.87%

Tabla 20. Identificación de parámetros para RNA que predicen la ISG: Pruebas con horizonte de 1 hora.

### Parámetros y RNA para horizontes de más de 1 hora

A continuación se presentan las RNA con las que se experimentó la predicción futura de la ISG para un horizonte temporal de 2 horas, mostrándose los resultados en la tabla 21. Se pueden apreciar características importantes para la selección de los parámetros de las RNA en estos horizontes temporales. La primera es que las mejores predicciones se alcanzan con MLP de dos capas ocultas, y la segunda es que los Perceptrones Simples no logran converger y alcanzar niveles de errores aceptables, lo que implica que la predicción de la ISG para un horizonte de 2 o más horas deja de ser un problema con relaciones lineales. En esta tabla, no se agregan resultados de los experimentos para horizontes temporales de 3-6 horas puesto que el patrón encontrado fue similar a lo antes mencionado.

Red #	Otros parámetros	Arquitectura de Red	%RMSE en TEST
22	Delta-Deriv = 0.03	PERCEPTRON	148.68%
23	Delta-Deriv = 0.05	PERCEPTRON	148.68%
24	Delta-Deriv = 0.01	PERCEPTRON	148.68%
25	Delta-Deriv = 0.05	10HL1	26.97%
26	Delta-Deriv = 0.035	10HL1	27.50%
27	Delta-Deriv = 0.05	100HL1	148.68%
28	Delta-Deriv = 0.03	100HL1	148.68%
29	Delta-Deriv = 0.03	150HL1	148.68%
30	Delta-Deriv = 0.035	50HL1-25HL2	26.81%
31	Delta-Deriv = 0.055	50HL1-25HL2	24.32%
32	Delta-Deriv = 0.025	50HL1-25HL2	No convergió
<b>33</b>	<b>Delta-Deriv = 0.05</b>	<b>100HL1-10HL2</b>	<b>24.23%</b>
34	Delta-Deriv = 0.05	100HL1-10HL2	26.13%
35	Delta-Deriv = 0.035	300HL1	148.68%
36	Delta-Deriv = 0.035	300HL1-25HL2	27.67%
37	Delta-Deriv = 0.05	300HL1-25HL2	29.35%
38	Delta-Deriv = 0.1	300HL1-25HL2	29.99%
39	Delta-Deriv = 0.035	100HL1-10HL2	28.58%
40	Delta-Deriv = 0.055	50HL1-25HL2	No convergió
41	Delta-Deriv = 0.055	50HL1-25HL2	No convergió

Tabla 21. Identificación de parámetros para RNA que predicen la ISG: Pruebas con horizonte de 2 horas.

Las RNA (42-44) utilizan datos de estaciones vecinas para estimar el valor de la ISG en el punto objetivo (estimación puramente espacial). Se puede apreciar que se alcanza un %RMSE de 14.9 % y que este nivel de error se obtiene con una RNA de tipo Perceptron Simple. Las RNA (45-50) predicen la ISG “instantánea” utilizando otras variables de entrada para cualquiera de las EM pertenecientes a la REM de CYL para predecir los valores de ISG “instantánea”, en este tipo de problema se aprecia que un MLP con 1 capa oculta presenta los mejores resultados.

Red #	Arquitectura de Red	% RMSE en TEST
42	<b>PERCEPTRON</b>	<b>14.09%</b>
43	MLP (30)	15.50%
44	MLP (30, 15)	15.75%
45	PERCEPTRON	30.72%
<b>46</b>	<b>MLP (8)</b>	<b>29.11%</b>
47	MLP (30)	29.50%
48	MLP (30, 15)	30.29%
49	MLP (30, 15)	31.58%
50	MLP (30, 30)	33.88%

Tabla 22. Identificación de parámetros para RNA que estiman la ISG sin variables de instantes previos.

### Identificación final de parámetros para las RNA

Los experimentos de carácter exploratorio llevados a cabo en esta sub-sección, permitieron identificar que para modelar la predicción a corto plazo de la ISG con estaciones vecinas, los mejores resultados se pueden alcanzar con MLP con dos capas ocultas con un número de neuronas de entre 50-300 en la primera capa y unas 25 neuronas en la segunda, mientras que estas mismas predicciones utilizando sólo datos del mismo punto genera mejores resultados con un MLP de una sola capa oculta con entre 10-50 neuronas en dicha capa.

#### 5.3.2. Predicciones de la ISG a corto plazo

En esta sub-sección se presentan los resultados definitivos de modelar la predicción a corto plazo de la ISG mediante un modelo de RNA para cada horizonte temporal. Se analizan los errores de los resultados en base a los estadísticos nRMSE y RMSE y %MAE nos permite hacer comparaciones con los trabajos previos.

Una vez identificada la arquitectura RNA y parámetros de entrada para éstas en la sub-sección anterior, se generaron otros experimentos para la optimización de las RNA definitivas utilizadas. Se identificó que la predicción a corto plazo de la ISG

con estaciones vecinas los mejores resultados se alcanzan con 100 neuronas en la primera capa oculta y 10 neuronas en la segunda. En el caso de no utilizar datos de estaciones vecinas, se encontró que los mejores resultados los genera un MLP con una capa oculta con 30 neuronas, siendo estos los parámetros definitivos utilizados.

### **Análisis global de los resultados**

En la tabla 23 se presentan los resultados globales (sin desagregar) de las predicciones de la ISG calculados a partir de de las validaciones cruzadas (utilizando los datos en test) y agrupadas según los distintos horizontes de predicción (1-6 horas) y por la inclusión de estaciones vecinas en los datos de entrada.

<b>Nombre de Red</b>	<b>Arquitectura de Red</b>	<b>%RMSE</b>	<b>%MAE</b>	<b>RMSE W/m<sup>2</sup></b>	<b>nRMSE</b>
RNA que utilizan datos de estaciones vecinas así como valores de instantes de tiempo previos de las distintas variables					
ANN1	100-10 (D-0.05)	<b>22.63</b>	<b>35.55</b>	<b>43.33</b>	<b>12.88</b>
ANN2	100-10 (D-0.035)	<b>24.24</b>	<b>38.15</b>	<b>47.49</b>	<b>13.80</b>
ANN3	100-10 (D-0.05)	<b>28.07</b>	<b>37.52</b>	<b>54.99</b>	<b>15.98</b>
ANN4	100-10 (D-0.05)	31.52	46.96	61.75	17.94
ANN5	100-10 (D-0.05)	35.25	45.32	69.04	20.07
ANN6	100-10 (D-0.05)	34.47	43.95	67.52	19.62
<b>Promedios:</b>		29.36	41.24	57.36	16.72
RNA que sólo utilizan datos de de instantes de tiempo previos de la propia estación objetivo de predicción					
ANN1-B	30 (D-0.05)	23.05	26.21	46.04	13.38
ANN2-B	30 (D-0.05)	26.06	44.52	51.05	14.84
ANN3-B	30 (D-0.05)	28.81	33.78	56.43	16.04
ANN4-B	30 (D-0.055)	<b>30.49</b>	<b>44.02</b>	<b>59.74</b>	<b>17.36</b>
ANN5-B	30 (D-0.055)	<b>30.34</b>	<b>46.74</b>	<b>59.43</b>	<b>17.27</b>
ANN6-B	30 (D-0.055)	<b>32.09</b>	<b>47.99</b>	<b>62.86</b>	<b>18.27</b>
<b>Promedios:</b>		28.95	40.92	56.69	16.47

Tabla 23. Resumen de resultados de las validaciones cruzadas de las RNA que estiman la ISG a corto plazo para horizontes desde 1 a 6 horas.

Los resultados mostrados en la tabla 23 muestran cómo en todos los casos los modelos de RNA creados logran predecir la ISG a corto plazo con niveles de

errores inferiores al 20% de nRMSE. Además desde la tabla se destaca cómo las RNA que predicen la ISG para horizontes temporales de 1 a 3 horas y que incluyen en sus entradas los datos de estaciones vecinas presentan mejores valores de nRMSE (12.88 %, 13.80 %, 15.98 %), contrario a lo que ocurre con las RNA que sólo usan datos de instantes previos del mismo punto objetivo de predicción (13.38 %, 14.84 %, 16.04 %). También es importante destacar que para los horizontes temporales de entre 4-6 horas los resultados se invierten, siendo mejores el segundo tipo de RNA con valores de nRMSE de (17.36 %, 17.27 % y 18.27 %) contra los valores de (17.94 %, 20.07 % y 19.62 %) del primer tipo de RNA.

Estos resultados se potencian por el hecho de utilizar la mayor cantidad de datos de entrada, así como la inclusión de datos de estaciones vecinas y a la capacidad que tienen las RNA de aprenderse las relaciones no lineales entre este gran número de entradas y la salida.

Los resultados alcanzados en esta investigación presentan un valor promedio en nRMSE de 16.45 %. Este valor es mejor y a la vez consistente con los valores los valores alcanzados en los trabajos de Voyant *et al.* (2013) y Paoli *et al.* (2010)<sup>2</sup>, en los que obtuvieron valores de nRMSE de 27.8 % y ~20 % respectivamente.

El valor promedio del estadístico RMSE resultante es de 57.02 W/m<sup>2</sup>, este es un mejor valor que el promedio entre días claros y nuboso de 63.47 W/m<sup>2</sup> obtenido por Wang *et al.* (2012). La comparación de los resultados con el otro estadístico común (%MAE) y el citado trabajo previo, no ha podido ser establecida, puesto que presentan discrepancias importantes. Es posible que estas diferencias se deban a las distintas localizaciones de las áreas de estudio, estando Wang *et al.* (2012) ubicado en China, mientras que tanto la investigación aquí desarrollada como las de Paoli *et al.* (2010) y Voyant *et al.* (2013) están ubicadas en el Sur de Europa. Otra diferencia es el hecho de que en Wang *et al.* (2012) el análisis de los errores se hace mezclando las predicciones en periodos de día como en periodos nocturnos.

---

<sup>2</sup> valor promedio de los resultados generales indicados en esa investigación.

El hecho de que tanto nuestra investigación como los trabajos previos sean de distintos contextos geográficos (España, Francia, Italia y China), implica que las comparaciones realizadas sean principalmente de carácter orientativo.

## **5.4 Resumen**

En este capítulo de la tesis se ha investigado la predicción a corto plazo de la Irradiancia Solar Global (ISG) para horizontes temporales de 15 minutos, 1, 2, 3, 4, 5 y 6 horas mediante Redes Neuronales Artificiales (RNA) y la relación que existe entre los niveles de error de las predicciones para los distintos horizontes temporales y la distancia máxima para incluir observaciones de estaciones vecinas un máximo de 55 km. Los datos experimentales fueron obtenidos desde las estaciones de la REM de CYL por ser la red oficial utilizada en esta tesis que presenta la mayor densidad de estaciones.

Los pasos metodológicos implicaron la organización de los datos de entrada de una forma “horizontal”, así como la experimentación con distintas RNA que permitiesen identificar los parámetros y arquitecturas de las RNA que generan las mejores predicciones de la ISG en los diferentes horizontes temporales. Finalmente se optimizaron una serie de RNA (una por cada horizonte temporal) de tal forma que alcanzaran los mejores niveles de errores en las predicciones.

La innovación ha consistido en el uso de la mayor cantidad de datos de entrada posible mediante el uso directo de valores de instantes previos de las distintas variables y la inclusión de la componente espacial en las entradas de las RNA. Esta innovación ha implicado el reto de experimentar con el modelado de RNA con hasta 900 entradas (altas dimensiones).

Los resultados de esta investigación han permitido generar modelos de RNA que predicen la ISG a corto plazo con niveles de errores inferiores al 20 % del nRMSE en todos los casos, además los resultados permiten confirmar que la inclusión de los datos de estaciones vecinas pueden mejorar los niveles de errores de las predicciones, confirmándose que para horizontes temporales de 1-3 horas la

inclusión de estos datos de estaciones vecinas ubicados a una distancia máxima de 55 km ha permitido generar mejores predicciones que las de las RNA que sólo utilizan datos de instantes previos de la misma estación objetivo de predicción.

Las redes entrenadas han mostrado una buena capacidad de generalización, ya que la comparación de los errores obtenidos con los datos de entrenamiento no discrepa mucho de los errores obtenidos con el conjunto de test empleado en el proceso de evaluación cruzada.

## **6 CONCLUSIONES Y DEMOSTRACIÓN DE LA HIPÓTESIS**

### **6.1. Conclusiones**

#### **6.1.1. Respecto a la disponibilidad de datos meteorológicos en tiempo cuasi-real en España**

La necesidad de indagar la cantidad de estaciones meteorológicas automáticas y el tipo de observaciones disponibles en el territorio Español como Datos Abiertos (Open Data) en Internet, motivó estudiar la disponibilidad de datos meteorológicos en tiempo cuasi-real en España. El principal reto planteado en la realización del análisis surgió de la heterogeneidad de las fuentes de datos meteorológicos identificadas, la falta de información armonizada sobre éstas y la falta de mecanismos sencillos que habiliten el uso interoperable de los datos (observaciones).

Se demostró que las Redes de Estaciones Meteorológicas de Voluntarios (REMV) aportan un número significativo de estaciones en el territorio español y que ayudan, en cierta medida, a complementar la distribución geográfica de las observaciones realizadas por las estaciones oficiales gubernamentales (GUB). Esta conclusión pretende llamar la atención sobre las observaciones realizadas por las REMV, similar a la expectación existente actualmente por la Información Geográfica generada por Voluntarios (*Volunteer Geographic Information: VGI*).

No es posible comparar al movimiento VGI con el de las REMV que comparten las observaciones de sus estaciones en Internet. VGI implica que la persona que hace la observación tome la iniciativa de generar y compartir los datos geográficos, mientras que con las REMV, el dueño de la estación sólo tiene que decidir unirse a una red para publicar y compartir los datos, el resto (la captura de las variables meteorológicas) se realiza de forma automática por la estación. Sin embargo, la disponibilidad de las observaciones meteorológicas georreferenciadas posibilita la creación de capas de información geográfica temática a partir de la información generada por los voluntarios, así como personalizar o generar servicios de

información más especializados, o simplemente ofrecer un *mashup* de esta información sobre mapas georreferenciados.

Esta parte de la investigación suple la carencia de un catálogo general que describa las redes de estaciones disponibles en el territorio español y sus características, a la vez que describe un método para la identificación de redes de estaciones que pudiera ser utilizado en otros países. Tanto el catálogo, como la descripción de la metodología usada para describir cómo se obtuvieron las observaciones, pueden ser utilizados para desarrollar un conjunto de programas de software de ejecución periódica para la recolección e integración de los datos en un repositorio común que haría posible ofrecer un servicio unificado de acceso y descarga de observaciones meteorológicas, preferiblemente siguiendo estándares tales como Open Linked Data o el *framework* de facilitación de acceso a datos de Sensores en la Web (*Sensor Web Enablement*: SWE) definido por el Consorcio Open GeoSpatial (OGC)

No se puede garantizar que el extenso inventario llevado a cabo sea exhaustivo, puesto que algunas organizaciones pueden tener sus propias redes trabajando internamente (intranet) sin compartir los datos en Internet.

Se han identificado 16 redes gubernamentales (GUB) con 1858 estaciones distribuidas en la geografía de un modo bastante regular. También se han identificado 8 REMV con 2558 estaciones distribuidas desigualmente por el territorio nacional español. Este inventario es una instantánea tomada en una fecha específica. Se considera que este inventario es más acertado en el contexto de las REM gubernamentales, debido a que los cambios en este tipo de redes tienden a ser menores y están previstas de antemano, mientras que en el caso de las redes de tipo REMV los cambios, la conectividad de las estaciones o el seguimiento de los incidentes es menos predecible; sin embargo el comportamiento histórico de estas redes sugiere que una tendencia de crecimiento.

Esta parte de la investigación ha demostrado la heterogeneidad en todas las características de las distintas redes y sus estaciones (formatos de datos, protocolos

de acceso, intervalos en la captura de observaciones, retardo en la publicación de las mismas y en la propiedad y responsabilidad de las redes). Aunque los formatos de datos y los protocolos de acceso varían de forma similar, se identificó que la mayoría de las redes tienden a utilizar interfaces orientadas a un usuario básico con páginas webs en HTML y HTTP\_GET (ya sea de forma principal o alternativa).

La determinación de la periodicidad de observación de las estaciones se hizo computando las observaciones registradas por día ya que esta información no la proporcionan muchas redes en sus metadatos. En ese contexto, se estudiaron 3 redes, una REMV y dos GUB (nacional y regional). Las redes GUB presentan mayor regularidad en las observaciones diarias pero la latencia de entrega de datos pudiera imposibilitar su uso en tiempo real. Por el contrario, las estaciones REMV presentan mayor dispersión estadística en lo relativo al periodo entre observaciones (cada dueño define este y otros criterios de funcionamiento de su estación) siendo los periodos de observación variados pero próximos a la decena de minutos. A cambio los datos están disponibles en tiempo real, hecho que facilita a las aplicaciones que requieran observaciones meteorológicas así, pueden utilizar datos de las REM de voluntarios teniendo en cuenta que estos datos no han sido validados y que las estaciones de distintas redes pueden ofrecer los datos en instantes de tiempo, intervalos y latencias distintas.

La cantidad de estaciones redundantes detectadas asciende a 479 (12.25 %), quedando el n° estimado de estaciones únicas en 3429. La identificación de redundancia en las EM de tipo GUB es más complicada ya que no se aprecian similitudes en los identificadores y precisa de una revisión individualizada tras detectar la proximidad. Por el contrario la identificación de redundancia en las estaciones REMV puede ser más sencilla y automatizable ya que los identificadores y nombres registrados en las distintas redes suelen ser parecidos. El hecho de que las estaciones REMV puedan estar ubicadas en más de una red puede generar inconvenientes para los usuarios de más de una, puesto que tendrían que pre-procesar los datos para eliminar duplicados. Sin embargo, por otro lado, estos duplicados también se pueden considerar como una fortaleza en caso de que una de las redes deje de funcionar temporalmente.

El análisis de lagunaridad realizado ha revelado que el 39.28 % de las celdas contienen al menos una estación meteorológica y el restante 60.72 % están vacías. Las celdas que solapan con las grandes ciudades / áreas urbanas tienen mayor número de estaciones. El aporte de las estaciones REMV en la densificación espacial de las observaciones meteorológicas está dado por el 8.42 % de las celdas que contiene sólo este tipo de estaciones. Para aplicaciones que requieran datos en tiempo real se puede considerar que el 11.01 % de las celdas contienen estaciones tanto de tipo GUB como REMV, habilitándose la opción de seleccionar estaciones en función de sus características. Aproximadamente la mitad de las celdas no vacías (19.86 % del total) contienen únicamente estaciones GUB.

Como se esperaba, el análisis realizado probó que existen más EM de voluntarios en las regiones con mayor población. La utilidad de las estaciones REMV se hace relevante para las aplicaciones que requieren datos en tiempo cuasi-real, aspecto esencial para la toma de decisiones dirigidas a la preservación del medio ambiente, como las aplicaciones de Energía Verde (renovable) y las Ciudades Inteligentes.

La existencia de estaciones gubernamentales calibradas cerca de los emplazamientos de las EM de voluntarios, habilitan la posibilidad de diseñar y llevar a cabo estudios dirigidos a la validación de las observaciones generadas por las EM de voluntarios.

El catalogo “datos.gob.es” para OpenData España aún no apunta a las fuentes de datos autonómicas que forman parte del movimiento OpenData. La búsqueda de “estaciones meteorológicas automáticas” no arroja ningún resultado, a diferencia de lo que sí ocurre por ejemplo con [www.data.gov.uk](http://www.data.gov.uk).

### **6.1.2. Respecto a la estimación espacial de la IS:**

La selección del mejor método de estimación espacial de la IS en intervalos de 15 minutos para el territorio español depende de la densidad de estaciones que observan esta variable meteorológica en una determinada área. Cuando el área de

interés está debidamente cubierta por sensores en la superficie, el mejor método es el Kriging con Regresión (*Regression Kriging*: RK) soportado por los valores de la IS estimados desde satélite y la latitud geográfica de la localización de la estación como variables auxiliares (KR2).

Se encontró que el método KR2 permite predecir espacialmente (interpolación) la IS en superficie más allá de los 25 km, límite que indica la bibliografía de referencia para el estado del arte de la predicción espacial de la IS en intervalos de 15 minutos. También que el método KR2 ha sido probado hasta los 108 km, distancia para la que el %RMSE nunca alcanzó el 25 % indicado como límite por la bibliografía de referencia. Por lo tanto, el método KR2 desarrollado en esta tesis aprovecha las fortalezas de ambas fuentes de información (observaciones en superficie y valores derivados de imágenes de satélite); esto permite cubrir distancias de interpolación más allá del límite que implica el hecho de sólo utilizar las observaciones de las estaciones en tierra pero con un nivel de error menor al de sólo utilizar valores derivados desde imágenes de satélite.

Los hechos anteriores habilitan algunas aplicaciones prácticas. Por ejemplo: (1) la validación de sensores de la IS de las REMV utilizadas en esta investigación; (2) la futura generación de superficies de IS en intervalos de 15 minutos con el propósito de estimar el potencial de producción de energía generada por el Sol en lugares donde no hay observaciones directas de la IS y con un error inferior al 25 % RMSE.

La desventaja de utilizar el método RK2 es la necesidad de disponer de valores de IS derivadas desde imágenes de satélite. En el caso de aplicaciones en tiempo cuasi-real, a esta necesidad también se le agrega el requerimiento de que los valores IS deben ser obtenidos de forma rápida. Si bien esto no es un problema, existen servicios de pago, que en un tiempo aproximado de 3 minutos entre la captura de las imágenes de satélite y el procesamiento, pueden entregar valores IS (SoDA-is 2014).

Por otro lado, cuando la densidad de estaciones con sensores IS es muy baja, se encuentren muy dispersos entre sí, o si el área de trabajo está poco densificada, entonces lo mejor (sobre cualquier método de interpolación) es la estimación de la ISG desde imágenes de satélite, resultado que concuerda con lo expuesto en la literatura.

Para cualquiera de los métodos (Interpolaciones o Estimación por Satélite), los *outliers* encontrados son mayores cuando se trata de estimar la IS en franjas de tiempo donde los valores de la IS son bajos (amanecer/anochece). A medida que se estiman valores IS más altos, la desviación estándar y las medias de errores van disminuyendo (la pendiente de la variación en el tiempo decrece y pensamos que se debe a los problemas de deriva en los relojes de las estaciones se minimizan).

En esta parte de la investigación, además del objetivo de estimar espacialmente la IS a intervalos de 15 minutos, se llevaron a cabo interpolaciones temporales basadas en las observaciones de cada estación con el objetivo de: (i) alinear las observaciones de las distintas estaciones a los minutos 0, 15, 30 y 45 de cada hora (independientemente de la fuente); (ii) obviar datos donde existen más observaciones que las necesarias (usualmente en la noche). Esta interpolación temporal se ha realizado siempre que la ventana de tiempo entre dos observaciones consecutivas fuese inferior o igual a 30 minutos. Estas características en los datos han evitado tener que utilizar otros métodos de interpolación temporal más sofisticados y complejos.

La aplicación práctica de haber investigado la estimación espacial la IS a 15 minutos, bajo condiciones de cielo claro y para valores  $IS > 200 \text{ W/m}^2$ , permitió determinar que el 67 % (de un total de 206 estaciones) de las observaciones de IS realizadas por los sensores de EM de voluntarios, analizadas estuvieron dentro del margen de error de los modelos evaluados. Esos valores no deberían ser subestimados a pesar de ser el resultado de un enfoque sencillo, debido a que en valores absolutos este 67 % representa 138 estaciones de tipo REMV. Estas 138 estaciones REMV son, comparativamente, aproximadamente el doble de las 69 estaciones oficiales utilizadas.

### **6.1.3. Respecto a la predicción espacio-temporal de la IS:**

La inclusión en los modelos de RNA de la mayor cantidad de datos de entrada posible, mediante el uso directo de valores de instantes previos de las distintas variables y la inclusión de la componente espacial, en las entradas de las RNA ha implicado el reto de experimentar con el modelado de RNA que tengan hasta 900 entradas (altas dimensiones).

Los resultados han permitido generar modelos de RNA que predicen la ISG a corto plazo con niveles de exactitud inferiores al 20 % del nRMSE en todos los horizontes temporales.

Los resultados también han permitido identificar el aporte y los límites de la inclusión de estaciones vecinas en los datos de entrada a los modelos de RNA; así se encontró que existe una relación entre la distancia de los vecinos y el aporte que hacen al horizonte temporal de predicción. Los resultados concluyen que para horizontes de predicción de hasta 3 horas, lo mejor es utilizar datos de las estaciones vecinas (distancias inferiores a 55 km de distancia) respecto a la estación objetivo de predicción, mientras que para horizontes de entre 4 y 6 horas es mejor no utilizar dichas estaciones en ese rango de distancias.

La relación entre las predicciones en distintos horizontes temporales y las distancias en que se encuentran las estaciones vecinas se puede explicar por el traslado del efecto meteorológico local de éstas hacia la estación objetivo de predicción, de tal forma que estaciones a 55 km o menos de lejanía y las condiciones metrológicas temporalmente cercanas (1-3 horas) influirán en la estación que es objetivo de predicción, mientras que los efectos de las observaciones de estas estaciones ya habrán pasado para horizontes de (4-6) horas, hecho que dificulta más que beneficia a las predicciones de este horizonte.

Las diferencias entre los niveles de errores promedio tanto para los modelos de RNA que incluyen datos de las estaciones vecinas, como las que sólo incluyen datos de las los instantes previos de las variables de la misma estación objetivo, presentan una mejora entre sí de 1.06 % en promedio. A pesar que esta mejora no

es sustancial, es probable que para tomadores de decisiones asociados a las grandes generadoras pueda significar una ganancia significativa respecto a la cantidad vatios generados (aprovechados) a lo largo del tiempo.

## **6.2 Aportaciones originales**

Se genera un inventario de 24 Redes de Estaciones Meteorológicas (REM) automáticas en España evidenciando sus principales características que permitan la recolecta de datos desde ellas mediante métodos automáticos como robots webs.

Se presenta la disposición y complementariedad espacial de todas las estaciones de las 24 REM agrupadas entre estaciones oficiales, de voluntarios y de ambos tipos, lo cual facilitará la planificación de instalación de nuevas estaciones por parte de los organismos oficiales, así como la selección de éstas estaciones por parte de los usuarios.

Se realiza una primera aproximación a la validación de las estaciones meteorológicas de voluntarios con sensores de Irradiancia Solar (IS), como un medio para ayudar en la densificación de esta variable en la superficie.

Se propone una metodología para la estimación de la IS en periodos de 15 minutos basados en geoestadística que combina las observaciones realizadas por las estaciones en superficie y el uso de las estimaciones de la Irradiancia Solar (E-IS) proveniente desde sensores remotos, permitiendo alcanzar mejores niveles de errores que los indicados por los trabajos previos.

Se propone la modelización de la predicción espacio-temporal de la IS a corto plazo mediante Redes Neuronales Artificiales (RNA) que incluyan la mayor cantidad de entradas y de instantes previos de las variables, así como la inclusión de datos desde estaciones vecinas. Estas RNA permiten una mejora en los niveles de errores previamente encontrados por investigaciones relacionadas.

## **6.3 Revisión de las preguntas de investigación e hipótesis**

Para facilitar la revisión de las preguntas de investigación se vuelven a presentar las cuestiones planteadas en el capítulo 1 a medida que se van argumentando las respuestas.

**1)** ¿Existen Estaciones Meteorológicas (EM) que ofrecen datos de forma abierta en el territorio español en más de una red de estaciones?

El inventario de las Redes de Estaciones Meteorológicas (REM) realizado en el capítulo 3 ha permitido identificar la existencia de hasta 24 REM con un total de más de 3908 estaciones. Sin embargo un análisis de redundancia permite bajar este número hasta 3429 estaciones, evidenciándose así que muchas estaciones publican sus datos en más de una REM.

**2)** ¿Cómo están geográficamente distribuidas las REM en España?

El análisis de lagunaridad realizado sobre las 24 REM, han permitido identificar que el 60.72 % de las celdas en las que se dividió el territorio español no contienen ningún tipo de EM en su interior (presentan lagunas), el 19.65 % de las celdas sólo contienen estaciones oficiales, el 8.42 % sólo de voluntarias y el 11 % contiene ambos tipos de estaciones.

**3)** ¿Es posible interpolar determinística o estadísticamente la IS en el territorio nacional a intervalos de 15 minutos más allá de los 25 km con errores inferiores al 25 %?

La metodología desarrollada en el capítulo 4 para la estimación espacial de la IS en periodos de 15 minutos ha permitido identificar que es posible generar interpolaciones geoestadísticas de la IS con ambas redes que reducen los errores atribuidos a las estimaciones de la IS a partir de sensores remotos bajo las siguientes condiciones: i) Las estaciones deben presentar una adecuada densidad/distribución, ii) Se deben de utilizar como variable auxiliar las estimaciones de la IS a partir de sensores remotos al usar el método geoestadístico.

4) ¿Están las observaciones de la IS de las EM de la red Meteoclimatic (voluntarios) dentro del margen de error del modelo de referencia, para poder considerarlas válidas?

Una aplicación de la metodología desarrollada en el capítulo 5 como una primera aproximación para la validación de las EM de voluntarios permitió identificar que el 67 % de estas estaciones presentan valores dentro del margen de error ( $\pm 2$  desviaciones estándares) cuando se comparan las estimaciones realizadas con el método aplicado y las propias observaciones de la IS en condiciones normales: de cielo claro e IS mayor a  $200 \text{ W/m}^2$ .

5) ¿Se mejora el error de la predicción temporal de la IS a corto plazo en base a RNA con la inclusión de la mayor cantidad de datos de entradas en los modelos?

La inclusión directa de las 9 variables de entradas por cada instante temporal y por cada estación vecina agregada a los modelos permitió, en el capítulo 5, generar RNA que predican la IS a corto plazo con niveles de errores inferiores a los previamente identificados por trabajos relacionados.

6) ¿Cómo se relacionan el horizonte espacial y temporal en la predicción de la IS a corto plazo con las RNA?

Los experimentos realizados con RNA en el capítulo 5 han permitido identificar la correlación existente entre la inclusión de estaciones vecinas y el horizonte temporal de predicción. De esta forma, se mejoran las predicciones de hasta 3 horas de horizonte temporal si se incluyen las observaciones de estaciones vecinas (hasta 55 km de distancia), sin embargo para horizontes de 4 a 6 horas es mejor no utilizarlas y predecir sólo con las variables de instantes previos del mismo punto objetivo.

En base a la respuesta a las preguntas de investigación se ha podido confirmar la doble hipótesis plantada en esta tesis doctoral. De tal forma que ha quedado demostrado que el uso de valores de la Irradiancia Solar, derivados a partir de

imágenes de satélite, como variable auxiliar en los métodos de estimación espacial permite generar interpolaciones con errores inferiores al de los propios valores derivados de imágenes de satélite. También quedó demostrado que es posible mejorar la predicción a corto plazo (*nowcasting*) de la ISG con modelos basados en Redes Neuronales Artificiales (RNA) mediante el uso de más variables de entrada y la inclusión de la componente espacial, sin embargo la componente espacial se debe usar considerando la relación identificada entre los horizontes temporales de la predicción y la distancia de influencia de las estaciones vecinas.

## **6.4 Trabajos futuros**

El diseño y desarrollo de un conjunto de programas para coleccionar e integrar observaciones desde diferentes redes meteorológicas en una base de datos que haga posible utilizar estos datos de una forma más simple e interoperable. En este sentido, ya existe un avance, la base de datos está diseñada y ya se han creado programas recolectores para algunas de las redes, sin embargo faltaría por desarrollar más recolectores para el resto de redes.

No se debería desestimar el uso del 67 % de las estaciones REMV que registran la IS y cuyas observaciones en los análisis preliminares de validación realizados presentar valores dentro del  $\pm 2\%$  de error para densificar espacialmente la disponibilidad de observaciones en superficie de la IS. Esto motiva el iniciar una nueva investigación enfocada directamente en la validación de las observaciones IS procedentes de las estaciones REMV, considerando distintos factores que pueden influenciar las interpolaciones de la IS en los emplazamientos de las estaciones REMV como por ejemplo: (i) La altitud “Z” de la estación respecto al nivel del mar, (ii) el relieve circundante a la estación, (iii) edificios cercanos a la estación, entre otros.

También se debe investigar la aplicación de métodos avanzados para la regeneración de las series de observaciones cuando por circunstancias se interrumpe el acceso a los datos o se pierden observaciones en determinados instantes temporales. El objetivo es conseguir alinear las observaciones en un

mismo marco temporal (minutos múltiplos de 15) como por ejemplo el algoritmo de preservación de la media propuesto por (Rymes y Myers 2011) y así determinar si se logra alguna mejora respecto al método de interpolación lineal entre dos observaciones continuas que se utilizó.

Se debe profundizar en la investigación de las influencias en los distintos horizontes temporales de la predicción a corto plazo de la IS mediante RNA y los criterios de distancia a considerar para filtrar estaciones y sus observaciones, como datos de entrada de los modelos, por ejemplo selección de distancias que formen círculos concéntricos entorno al lugar para el que se pretende entrenar la RNA. Esto, por ejemplo, permitirá determinar si la inclusión de estaciones ubicadas a distancias del punto de interés de entre 55 y 100 km permitía mejorar los niveles de errores de las predicciones temporales en un horizonte de 4 a 6 horas.

También es interesante investigar la relevancia de los valores de las distintas variables de entrada en distintos instantes de tiempo, así como la agrupación de entradas de las estaciones vecinas por cuadrantes con el objetivo de identificar la posibilidad de crear RNA que sean capaces de predecir la IS en cualquier punto.

## 7 BIBLIOGRAFÍA

Abella, A. (2011). Reutilización de información pública y privada en España. *Rooter Press* **2011**, Madrid, Spain.  
[http://rooter.es/documents/PAPER\\_REUTILIZACION\\_INFORMACION\\_PUBLICA\\_PRIVADA\\_OPENDATA.pdf](http://rooter.es/documents/PAPER_REUTILIZACION_INFORMACION_PUBLICA_PRIVADA_OPENDATA.pdf) (visitado 01/11/2014).

AEMet (2011). AEMet establece el acceso libre y gratuito a todos sus datos por medios electrónicos. Madrid, España.  
[http://www.aemet.es/es/zona\\_portada\\_destacada/politicadatos](http://www.aemet.es/es/zona_portada_destacada/politicadatos) (visitado 01/11/2014).

AEMet (2011). La Radiación Solar.  
[http://www.aemet.es/documentos/es/eltiempo/observacion/radiacion/Radiacion\\_Solar.pdf](http://www.aemet.es/documentos/es/eltiempo/observacion/radiacion/Radiacion_Solar.pdf) (visitado 01/11/2014).

AENOR (2004). Redes de estaciones meteorológicas automáticas: Directrices para la validación de registros meteorológicos procedentes de redes de estaciones automáticas (Validación en tiempo real). (UNE 500540:2004).

Al-Bakri, M y Fairbairn, D (2012). Assessing Similarity Matching for Possible Integration of Feature Classifications of Geospatial Data from Official and Informal Sources. *International Journal of Geographical Information Science* **2012**, 26, 1437-1456. DOI:10.1080/13658816.2011.636012

Allison, P. (1998). Multiple regression: A primer. *Pine Forge Press* **1998**, Thousand Oaks, California, USA.

Alsamamra, H.; Ruiz-Arias, J.A.; Pozo-Vázquez, D.; Tovar-Pescador, J. (2009). A Comparative Study of Ordinary and Residual Kriging Techniques for Mapping Global Solar Radiation Over Southern Spain. *Agric. For. Meteorol.* **2009**, 149, 1343–1357. DOI: 10.1016/j.agrformet.2009.03.005

Antonanzas-Torres, F.; Cañizares, F. and Perpiñán, O. (2013). Comparative Assessment of Global Irradiation from a Satellite Estimate Model (CM SAF) and on-Ground Measurements (SIAR): A Spanish Case Study. *Renewable and Sustainable Energy Reviews* **2013**, 21248-261. DOI: 10.1016/j.rser.2012.12.033

Automap (2011). Automatic Interpolation Package. <http://cran.r-project.org/web/packages/automap> (visitado 01/11/2014).

Badescu, V. (ed.) (2008). Modeling Solar Radiation at the Earth's Surface. *Springer-Verlag Berlin Heidelberg* **2008**. ISBN 978-3-540-77455-6

BBC (2013). Horizons – Energy. London, UK.  
<http://www.bbc.com/specialfeatures/horizonsbusiness/wp-content/uploads/2013/04/Energy1.pdf> (visitado 01/11/2014).

Blanc, P.; Gschwind, B.; Lefèvre, M. and Wald, L. (2011). The HelioClim Project: Surface Solar Irradiance Data for Climate Applications. *Remote Sensing* **2011**, 3(2), 343-361. doi:10.3390/rs3020343

Bocco, M.; Ovando, G. and Sayago, S. (2006). Development and Evaluation of Neural Network Models to Estimate Daily Solar Radiation at Córdoba, Argentina. *Pesquisa Agropecuária Brasileira* **2006**, 41(2), 179-184.

BOE (2011). Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, para el ámbito del sector público estatal. Madrid, Spain.

Bojanowski, J.S. (2014). Quantifying solar radiation at the earth surface with meteorological and satellite data. Enschede, *University of Twente Faculty of Geo-Information and Earth Observation (ITC)* **2014**. ITC Dissertation 242. Skidmore, A.K. (Promoter) and Vrieling, A. (assistant promoter). ISBN: 978-90-6164-371-5.

Bojanowski, J.S.; Vrieling, A.; Skidmore, A.K. (2013). Calibration of Solar Radiation Models for Europe using Meteosat Second Generation and Weather Station Data. *Agric. For. Meteorol.* **2013**, *176*, 1–9. DOI: 10.1016/j.agrformet.2013.03.005

Bosch, J. L., Lopez, G., & Batlles, F. J. (2008). Daily solar irradiation estimation over a mountainous area using artificial neural networks. *Renewable Energy* **2008**, *33*(7), 1622-1628.

Brázdil, R.; Pfister, C.; Wanner, H.; Storch, H.V. and Luterbacher, J. (2005). Historical Climatology in Europe—the State of the Art. *Climatic Change* **2005**, *70*(3), 363-430. 10.1007/s10584-005-5924-1

Conway, E.D. (ed.) (1997). An introduction to satellite image interpretation. *The Johns Hopkins Univ Press* **1997**. Baltimore, Maryland, USA.

Cortez, P.; Rocha, M.; Allegro, F. S. and Neves, J. (2002). Real-time forecasting by bio-inspired models. In: *Proceeding of the Artificial Intelligence and Applications* **2002**, Málaga, Spain.

Croarkin, C.; Tobias, P. (2010). NIST/SEMATECH e-Handbook of Statistical Methods. *The National Institute of Standards and Technology* **2010**, Gaithersburg, MD, USA.

CWOP. Citizen Weather Observer Program. <http://www.wxqa.com> (visitado 01/11/2014).

Duffie, J. A. and Beckman, A. (1991). Solar Engineering of Thermal Processes, Third Edition. *Wiley Press* **1991**, ISBN-10: 0471698679.

Estévez, J. and Gavilán, P. (2008). Procedimientos de Validación de datos de estaciones meteorológicas automáticas. Aplicación a la red de información agroclimática de Andalucía. *Instituto de Investigación y Formación Agraria y Pesquera de Andalucía, Consejería de Agricultura y Pesca de la Junta de Andalucía* **2008**.

<http://www.juntadeandalucia.es/agriculturaypesca/ifapa/servifapa/contenidoAlf?id=52392b22-6860-4efd-8b9a-32dfb0275646> (visitado 01/11/2014).

Ertekin, C.; Evrendilek, F. (2007). Spatio-Temporal Modeling of Global Solar Radiation Dynamics as a Function of Sunshine Duration for Turkey. *Agric. For. Meteorol.* **2007**, 145, 36–47. DOI: 10.1016/j.agrformet.2007.04.004

European Commission (2009a). Directiva 2009/28/CE del Parlamento Europeo y del Consejo. Relativa al fomento del uso de energía procedente de fuentes renovables. *EUR-LEX* **2009**, Brussels, Belgium. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:140:0016:0062:es:PDF>

European Commission (2009b). Ministerial Declaration on eGovernment. In *European Commission* **2006**. Malmö, Sweden. <http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/ministerial-declaration-on-egovernment-malmo.pdf>

European Commission (2013). Joint Research Center Scientific and Policy Report. PV Status Report 2013. *Publication office of the European Union* **2013**. Ispra, Italy. [http://iet.jrc.ec.europa.eu/remea/sites/remea/files/pv\\_status\\_report\\_2013.pdf](http://iet.jrc.ec.europa.eu/remea/sites/remea/files/pv_status_report_2013.pdf)

European Economy News (2010). SMART, SUSTAINABLE AND INCLUSIVE GROWTH. Europe's "moment of truth": How to the Europe 2020 strategy must transform the union. *EUR-LEX* **2010**. <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52010DC2020&from=EN> (visitado 01/11/2014).

Evrendilek, F.; Ertekin, C. (2007). Statistical Modeling of Spatio-Temporal Variability in Monthly Average Daily Solar Radiation Over Turkey. *Sensors* **2007**, 7, 2763–2778. doi:10.3390/s7112763

Fiebrich, C.A. (2009). History of Surface Weather Observations in the United States. *Earth-Science Reviews* **2009**, 93(3-4), 77-84. DOI: 10.1016/j.earscirev.2009.01.001

Finn, P.G.; Udy, N.S.; Baltais, S.J.; Price, K. and Coles, L. (2010). Assessing the Quality of Seagrass Data Collected by Community Volunteers in Moreton Bay Marine Park, Australia. *Environmental Conservation* **2010**, 37(01), 83-89. DOI: <http://dx.doi.org/10.1017/S0376892910000251>

Gautier, C.; Diak, G. and Masse, S. (1980). A Simple Physical Model to Estimate Incident Solar Radiation at the Surface from GOES Satellite Data. *J. Appl. Meteor.* **1980**, 19, 1005–1012.  
doi: [http://dx.doi.org/10.1175/1520-0450\(1980\)019<1005:ASPMTE>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1980)019<1005:ASPMTE>2.0.CO;2)

Girres, J.F. and Touya, G. (2010). Quality Assessment of the French Openstreetmap Dataset. *Transactions in GIS* **2010**, 14(4), 435-459. DOI: 10.1111/j.1467-9671.2010.01203.x

Gobierno Vasco. Euskalmet. Agencia Vasaca de Meteorología. España. <http://www.euskalmet.euskadi.net/s07-5853x/es/meteorologia/home.apl?e=5> (visitado 01/11/2014).

Gobierno Vasco. Open Data EUSKADI. España. <http://opendata.euskadi.net/w79-home/es> (visitado 01/11/2014).

Goodchild, MF (2007a). Citizen as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research* **2007**, 224-32.

Goodchild, M.F. (2007b). Citizens as Sensors: The World of Volunteered Geography. *GeoJournal* **2007**, 69(4), 211-221. DOI 10.1007/s10708-007-9111-y

Goodchild, M.F. and Li, L. (2012). Assuring the Quality of Volunteered Geographic Information. *Spatial Statistics* **2012**, 1110-120. DOI: 10.1016/j.spasta.2012.03.002

Gubbi, J.; Buyya, R.; Marusic, S. and Palaniswami, M. (2013). Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions. *Future Generation Computer Systems* **2013**, 29(0), 1645-1660. DOI:10.1016/j.future.2013.01.010

Gueymard C.A. (2004). The sun's total and spectral irradiance for solar energy applications and solar radiation models. *Solar Energy* **2004**, 76(4):423-53. DOI: 10.1016/j.solener.2003.08.039

Gueymard, C.A. and Myers D.R. (2008). Solar radiation measurement: Progress in radiometry for improved modeling. In Modeling Solar Radiation at the Earth Surface (2008). *Springer-Verlag Berlin Heidelberg* **2008**. ISBN 978-3-540-77455-6. DOI: 10.1007/978-3-540-77455-6\_1

Hagan, M. T.; Demuth, H. B. and Beale, M. H. (1996). Neural network design (Vol. 1). *Boston: Pws* **1996**.

Hartley, B. (2012). The Internet of Things - Weather Monitoring Too. In *WMO technical conference on meteorological and environmental instruments and methods of observation* **2012**, Belgium, Brussels.

Harrouni, S. (2008). Fractal classification of typical meteorological days from global solar irradiance: Application to five sites of different climates. In Modeling Solar Radiation at the Earth's Surface: pp 29-54. *Springer-Verlag Berlin Heidelberg* **2008**. ISBN 978-3-540-77455-6. DOI: 10.1007/978-3-540-77455-6\_2

- Heipke, C. (2010). Crowdsourcing Geospatial Data. *ISPRS Journal of Photogrammetry and Remote Sensing* **2010**, 65(6), 550-557. DOI: 10.1016/j.isprsjprs.2010.06.005
- Hengl, T. (2009). A Practical Guide to Geostatistical Mapping of Environmental Variables; *University of Amsterdam* **2009**. Amsterdam, The Netherlands. pp. 291.
- Hengl, T.; Heuvelink, G.B.; Tadić, M.P. and Pebesma, E.J. (2012). Spatio-Temporal Prediction of Daily Temperatures using Time-Series of MODIS LST Images. *Theoretical and applied climatology* **2012**, 107(1-2), 265-277. DOI: 10.1007/s00704-011-0464-2
- Hernández, L.; Baladrón, C.; Aguiar, J.M.; Calavia, L.; Carro, B.; Sánchez-Esguevillas, A.; Cook, D.J.; Chinarro, D. and Gómez, J. (2012). A Study of the Relationship between Weather Variables and Electric Power Demand Inside a Smart grid/smart World Framework. *Sensors* **2012**, 12(9), 11571-11591. DOI:10.3390/s120911571
- Hontoria, L.; Aguilera, J. and Zufiria, P. (2002). Generation of hourly irradiation synthetic series using the neural network multilayer perceptron. *Solar Energy* **2002**, 72(5), 441-446.
- Hudson, G. Kriging temperature in Scotland using the external drift method. In Geostatistics Troia'92; Soares, A., Ed.; *Springer Netherlands* **1993**, Dordrecht, The Netherlands; Volume 1, pp. 577–588. DOI: 10.1007/978-94-011-1739-5\_45
- Ianetz, A. and Kudish, A. (2008). A method for determining the solar global and defining the diffuse and beam irradiation on a clear day. In Modeling Solar Radiation at the Earth's Surface, pp 93-113. *Springer-Verlag Berlin Heidelberg* **2008**. ISBN 978-3-540-77455-6. DOI: 10.1007/978-3-540-77455-6\_4

IDAE (2010). Informe de previsión. Plan de Energías Renovables 2011-2020. <http://www.idae.es/index.php/mod.pags/mem.detalle/relcategoria.1153/id.501/reلمenu> (visitado 01/11/2014).

IEA (2014). 2014 key world energy statistics. <http://www.iea.org/publications/freepublications/publication/KeyWorld2014.pdf> (visitado 01/11/2014).

IEA-SHC (2010). Task 36 Objectives. <http://archive.iea-shc.org/task36/objectives.htm> (visitado 01/11/2014).

ITACyL. Geoportal: Acceso a Datos. España. <ftp://ftp.itacyl.es> (visitado 01/11/2014).

ITACyL. ¿Qué es el Instituto Tecnológico Agrario de Castilla y León? España. [http://www.itacyl.es/opencms\\_wf/opencms/itacyl/quienes\\_somos/que\\_es\\_itacyl/index.html](http://www.itacyl.es/opencms_wf/opencms/itacyl/quienes_somos/que_es_itacyl/index.html) (visitado 01/11/2014).

Iqbal, M. (ed.) (1983). An introduction to solar radiation. *Elsevier Science & Technology Book* **1983**.

Junta de Galicia. Meteo Galicia. España. <http://www.meteogalicia.es/web/index.action> (visitado 01/11/2014).

Junta de Galicia. Portal de Open Data de la Xunta de Galicia. <http://abertos.xunta.es/mapa-do-portal> (visitado 01/11/2014).

Kalampokis, E.; Tambouris, E. and Tarabanis, K. (2011). Open Government Data: A Stage Model. *Electronic Government* **2011**, 235-246.

Knights, M. (2007). Web 2.0. *Communications Engineer* **2007**, 5(1), 30-35.

Koca, A., Oztop, H. F., Varol, Y., & Koca, G. O. (2011). Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey. *Expert Systems with Applications* **2011**, 38(7), 8756-8762.

Kretzschmar, R.; Eckert, P.; Cattani, D. and Eggimann, F. (2004). Neural network classifiers for local wind prediction. *Journal of Applied Meteorology and Climatology* **2004**, 43 (5). pp. 727–738.

Kuhn, M. (2014) caret: Classification and Regression Training. <http://cran.r-project.org/web/packages/caret/index.html> (visitado 01/11/2014).

Kumar, L.; Skidmore, A.K.; Knowles, E. (2009). Modelling Topographic Variation in Solar Radiation in a GIS Environment. *Int. J. Geogr. Inf. Sci.* **1997**, 11, 475–497.

Landsberg, H.E. (1980). Past Climates from Unexploited Written Sources. *The Journal of Interdisciplinary History* **1980**, 10(4), 631-642.

Laszloa, I.; Cirenb, P.; Liub, H.; Kondraguntaa, S.; Tarpleya, J. D.; Goldberga, M.D. (2008). Remote sensing of aerosol and radiation from geostationary satellites. *Advances in Space Research* **2008**, 41 (11). pp. 1882–1893. DOI: 10.1016/j.asr.2007.06.047

Linares-Rodríguez, A., Ruiz-Arias, J. A., Pozo-Vázquez, D., & Tovar-Pescador, J. (2011). Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks. *Energy* **2011**, 36(8), 5356-5365.

Liu, B.Y. and Jordan, R.C. (1960). The Interrelationship and Characteristic Distribution of Direct, Diffuse and Total Solar Radiation. *Solar Energy* **1960**, 4(3), 1-19.

Lusher, W.; Jackson, J. and Morgan, K. (2009). The Florida Automated Weather Network (FAWN): Ten Years of Providing Weather Information to Florida Growers. University of Florida IFAS Extension, Publication #SL 298.

Mayo, L. (2013). The Sun: 101!. Sun-Earth Days, NASA.  
[http://sunearthday.nasa.gov/2013/articles/thesun\\_101.php](http://sunearthday.nasa.gov/2013/articles/thesun_101.php) (visitado 01/11/2014).

Manley, G. (1974). Central England Temperatures: Monthly Means 1659 to 1973. *Quarterly Journal of the Royal Meteorological Society* **1974**, 100(425), 389-405.

Martins, F. R.; Pereira, E. B.; Guarnieri, R. A. (2012). Solar Radiation Forecast Using Artificial Neural Networks. *International Journal of Energy Science* **2012**, 2(6). <http://www.ijesci.org/paperInfo.aspx?ID=6069>

Mellit, A., Benghanem, M., & Kalogirou, S. A. (2006). An adaptive wavelet-network model for forecasting daily total solar-radiation. *Applied Energy* **2006**, 83(7), 705-722.

Mellit, A. and Kalogirou, S. A. (2008). Artificial intelligence techniques for photovoltaic applications: A review. *Progress in Energy and Combustion Science* **2008**, 34 (5), pp. 574–632. DOI: 10.1016/j.pecs.2008.01.001

Mellit, A.; Kalogirou, S. A.; Hontoria, L. and Shaari, S. (2008). Artificial intelligence techniques for sizing photovoltaic systems: A review. *Renewable and Sustainable Energy Reviews* **2008**, 13 (2), pp. 406-419. DOI: 10.1016/j.rser.2008.01.006

Mellit A. and Pavan A. M. (2010). A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Solar Energy* **2010**, 5(84), 802-821, DOI: 10.1016/j.solener.2010.02.006.

Meteoclimatic. Meteoclimatic - España. <http://www.meteoclimatic.com> (visitado 01/11/2014).

Meteoclimatic. Subscripcion RSS. España.  
[http://www.meteoclimatic.com/index/wp/rss\\_es.html](http://www.meteoclimatic.com/index/wp/rss_es.html) (visitado 01/11/2014).

Meteoclimatic. Sello «Meteoclimatic» a la calidad. España.  
[http://www.meteoclimatic.com/index/quamet\\_es.html](http://www.meteoclimatic.com/index/quamet_es.html) (visitado 01/11/2014).

Meteoclimatic. Karma. España.  
[http://www.meteoclimatic.com/index/quamet\\_es.html](http://www.meteoclimatic.com/index/quamet_es.html) (visitado 01/11/2014).

Miller, E.R. (1931). The Evolution of Meteorological Institutions in the United States. *Monthly Weather Review* **1931**, 591.

MITYC (2009). Informe en relación con la cumplimentación del artículo 4.3 de la directiva 2009/28/ce, de 23 de abril de 2009, relativa al fomento del uso de energía procedente de fuentes renovables. *Subdirección General de Relaciones Energéticas Internacionales; secretaría de estado de energía 2009* (Informe).  
<http://www.minetur.gob.es/energia/desarrollo/EnergiaRenovable/Documents/InformeFomentoRenovables.pdf> (visitado 01/11/2014).

Miorandi, D.; Sicari, S.; De Pellegrini, F. and Chlamtac, I. (2012). Internet of Things: Vision, Applications and Research Challenges. *Ad Hoc Networks* **2012**, 10(7), 1497-1516.

Mitchell, H. (2010). Opening Up about Open Government: Do Open Government Initiatives Measure Up?. *EContent* **2010**, 33(6), 24-29.

Moebius, E. (2005). Amount of Energy the Earth Gets from the Sun. NASA's Cosmicopia. [http://helios.gsfc.nasa.gov/qa\\_sun.html#sunenergymass](http://helios.gsfc.nasa.gov/qa_sun.html#sunenergymass) (visitado 01/11/2014).

- Moebius, E. (2006). Solar Wind Electricity?. NASA's Cosmicopia.  
[http://helios.gsfc.nasa.gov/qa\\_sun.html#swelectric](http://helios.gsfc.nasa.gov/qa_sun.html#swelectric) (visitado 01/11/2014).
- Mohandes, M.; Rehman, S. and Halawani, T. O. (1998). Estimation of global solar radiation using artificial neural networks. *Renewable Energy* **1988** 14(1), 179-184.
- Miorandi, D.; Sicari, S.; De Pellegrini, F. and Chlamtac, I. (2012). Internet of things: vision, applications and research challenges. *Ad Hoc Netw* **2012**, 10(7):1497–1516.
- Moreno, A.; Gilabert, M. and Martínez, B. (2011). Mapping Daily Global Solar Irradiation Over Spain: A Comparative Study of Selected Approaches. *Sol. Energy* **2011**, 85, 2072–2084.
- Mubiru, J. (2008). Predicting total solar irradiation values using artificial neural networks. *Renewable Energy* **2008**, 33(10), 2329-2332.
- Mubiru, J. and Banda, E.J.K.B. (2008). Estimation of monthly average daily global solar irradiation using artificial neural networks. *Solar Energy* **2008**, 82(2), pp 181-187. DOI: 10.1016/j.solener.2007.06.003
- Neis, P.; Zielstra, D. and Zipf, A. (2011). The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011. *Future Internet* **2011**, 4(1), 1-21.
- Fritsch, S. and Guenther, F (2014). Neuralnet package. <http://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf> (visitado 01/11/2014).
- NASA (2012). The Sun. Sun-Earth.  
[http://www.nasa.gov/mission\\_pages/sunearth/multimedia/Sunlayers.html](http://www.nasa.gov/mission_pages/sunearth/multimedia/Sunlayers.html) (visitado 01/07. 2013).

NOAA. General Solar Position Calculations. USA.  
<http://www.esrl.noaa.gov/gmd/grad/solcalc/solareqns.PDF> (visitado 01/11/2014).

NOAA. Meteorological Assimilation Data Ingest System. USA.  
<http://madis.noaa.gov> (visitado 01/11/2014).

NREL (ed.) (1992). National Solar Radiation Data Base (1961-1990) User's Manual Asheville, North Carolina, USA.: National Climatic Data Center.

Obama, B. (2009). Transparency and Open Government. Memorandum for the Heads of Executive Departments and Agencies.

OMM. Organización Meteorológica Mundial. <http://www.wmo.int> (visitado 01/11/2014).

OpenStreetMap. OpenStreetMap. <http://www.openstreetmap.org> (visitado 01/11/2014).

Paoli, C., Voyant, C., Muselli, M., & Nivet, M. L. (2010). Forecasting of preprocessed daily solar radiation time series using neural networks. *Solar Energy* **2010**, 84(12), 2146-2160.

Pebesma, E.J. (2004). Multivariable Geostatistics in S: The Gstat Package. *Comput. Geosci.* **2004**, 30, 683–691.

Pebesma, E.J. (2006). The Role of External Variables and GIS Databases in Geostatistical Analysis. *Trans. GIS* **2006**, 10, 615–632.

Perpiñan, O. (2008). Grandes centrales fotovoltaicas: Producción, seguimiento y ciclo de vida. Tesis Doctoral, *Escuela Técnica Superior de Ingenieros Industriales, Universidad Nacional de Educación a Distancia* **2008**. Madrid, España.

- Perpiñán, O. (2012). SolaR: Solar Radiation and Photovoltaic Systems with R. *Journal of Statistical Software* **2012**, 501-32.
- Perez, R.; Seals, R.; Zelenka, A. (1997). Comparing Satellite Remote Sensing and Ground Network Measurements for the Production of site/time Specific Irradiance Data. *Sol. Energy* **1997**, 60, 89–96.
- Pinker, R. T. and Laszlo, I. (1992). Modeling Surface Solar Irradiance for Satellite Applications on a Global Scale. *J. Appl. Meteor.* **1992**, 31, 194–211.  
doi: [http://dx.doi.org/10.1175/1520-0450\(1992\)031<0194:MSSIFS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1992)031<0194:MSSIFS>2.0.CO;2)
- Polo, J.; Zarzalejo, L.F. and Ramírez, L. (2008). Solar radiation derived from satellite images. In Badescu, V. (ed.). Modeling Solar Radiation at the Earth's Surface. *Springer-Verlag Berlin Heidelberg* **2008**. ISBN 978-3-540-77455-6. pp. 449-462.
- Pons, X. and Ninyerola, M. (2008). Mapping a topographic global solar radiation model implemented in a GIS and refined with ground data. *Int. J. Climatol.* **2008**, 28: 1821–1834. DOI: 10.1002/joc.1676
- Posselt, R.; Mueller, R.; Stöckli, R. and Trentmann, J. (2011). Spatial and Temporal Homogeneity of Solar Surface Irradiance across Satellite Generations. *Remote Sens.* **2011**, 3(5), 1029-1046; doi:10.3390/rs3051029
- Plotnick, R.E.; Gardner, R.H.; Hargrove, W.W.; Prestegard, K. and Perlmutter, M. (1996). Lacunarity Analysis: A General Technique for the Analysis of Spatial Patterns. *Physical Review E.* **1996**, 53(5), 5461.
- R Development Core Team (2012). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* **2012**. Vienna, Austria.

Rigollier, C.; Lefèvre, M. and Wald, L. (2004). The Method Heliosat-2 for Deriving Shortwave Solar Radiation from Satellite Images. *Solar Energy* **2004**, 77(2), 159-169.

Ripley, B.; Lapsley, M. (2014). RODBC: ODBC Database <http://cran.r-project.org/web/packages/RODBC/index.html> (visitado 01/11/2014).

Reed, R. D. and Marks, R. J. (1998). Neural smithing: supervised learning in feedforward artificial neural networks. *Mit Press* **1998**.

Roebber, P.J.; Schultz, D.M.; Colle, B.A. and Stensrud, D.J. (2004). Toward Improved Prediction: High-Resolution and Ensemble Modeling Systems in Operations. *Weather and Forecasting* **2004**. pp. 19936-949.

Roopa, T.; Iyer, A.N. and Rangaswamy, S. (2013). CroTIS-Crowdsourcing Based Traffic Information System. *In Big Data Congress IEEE International Congress* **2013**. pp. 271-277.

Running, S.W.; Nemani, R.R. and Hungerford, R.D. (1987). Extrapolation of Synoptic Meteorological Data in Mountainous Terrain and its use for Simulating Forest Evapotranspiration and Photosynthesis. *Canadian Journal of Forest Research* **1987**, 17(6), 472-483.

Rymes, M. D. and Myers, D. R. (2001). Mean preserving algorithm for smoothly interpolating averaged data. *Solar energy* **2001**, 71(4), 225-231.

Salcedo-Sanz, S.; Pérez-Bellido, Á.M.; Ortiz-García, E.G.; Portilla-Figueras, A.; Prieto, L. and Correoso, F. (2009). Accurate Short-Term Wind Speed Prediction by Exploiting Diversity in Input Data using Banks of Artificial Neural Networks. *Neurocomputing* **2009**, 72(4-6), 1336-1341.

Serradilla, F. (2014). Manual de salmon. <http://aicu.eui.upm.es/wordpress/productos/manual-de-salmon> (visitado 01/11/2014).

Slonosky, V.C. (2002). Wet Winters, Dry Summers? Three Centuries of Precipitation Data from Paris. *Geophysical Research Letters* **2002**, 29(19). DOI: 10.1029/2001GL014302

Slonosky, V.C. and NOAA/NCDC. DATA SET: Paris Monthly Precipitation 1688-2009. <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/historical/europe/france/paris-precip-slonosky2002.txt> (visitado 01/11/2014).

Schroth, C. and Janner, T. (2007). Web 2.0 and SOA: Converging Concepts. Enabling the Internet of Services. *IT professional* **2007**, 9(3), 36-41.

SoDA-is (2011). Main Properties of the HelioClim Databases. France. [http://www.soda-is.com/eng/helioclim/helioclim\\_main\\_properties.html](http://www.soda-is.com/eng/helioclim/helioclim_main_properties.html) (visitado 01/11/2014).

SoDA-is (2013). HelioClim, Overview. France. [http://www.soda-is.com/doc/summary\\_helioclim.pdf](http://www.soda-is.com/doc/summary_helioclim.pdf) (visitado 01/11/2014).

SoDa-is (2013b). HelioClim-3 Uncertainty. France. [https://web.archive.org/web/20130511070410/http://www.soda-is.com/eng/helioclim/helioclim3\\_uncertainty\\_eng.html](https://web.archive.org/web/20130511070410/http://www.soda-is.com/eng/helioclim/helioclim3_uncertainty_eng.html) (visitado 01/11/2014).

SoDA-is (2014). SoDa – Leadtimes. <http://www.soda-is.com/eng/helioclim/leadtime.html#map> (visitado 01/11/2014).

SoDA-is (2014b). SoDa - HelioClim Info. <http://www.soda-is.com/eng/helioclim/index.html> (visitado 01/11/2014).

Tarpley, J. D. (1979). Estimating Incident Solar Radiation at the Surface from Geostationary Satellite Data. *J. Appl. Meteor.* **1979**, 18, 1172–1181.  
doi: [http://dx.doi.org/10.1175/1520-0450\(1979\)018<1172:EISRAT>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1979)018<1172:EISRAT>2.0.CO;2)

The Economist (2008). Special report: The future of energy – The power and the glory, June 2008. <http://www.economist.com/node/11565685> (visitado 01/11/2014).

Thurston, T. (2014). Relative size of the Sun and Earth. The accessible universe. <http://accessibleuniverse.blogspot.com.es> (visitado 01/11/2014).

Toole, G. L.; Fair, M.; Berscheid, A. and Bent, R. (2010). Electric power transmission network design for wind generation in the western united states: Algorithms, methodology, and analysis. *Transmission and distribution conference and exposition 2010*. IEEE PES. pp. 19-22. DOI: 10.1109/TDC.2010.5484549.

Tymvios, F. S., Michaelides, S. C., & Skouteli, C. S. (2008). Estimation of surface solar radiation with artificial neural networks. In *Modeling Solar Radiation at the Earth's Surface 2008*. Springer Berlin Heidelberg. pp. 221-256.

Voyant, C., Muselli, M., Paoli, C., & Nivet, M. L. (2011). Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. *Energy 2011*, 36(1), 348-359.

Voyant, C.; Randimbivololona, P.; Nivet, M.L.; Paoli, C. and Muselli, M. (2014). Twenty four hours ahead global irradiation forecasting using multi-layer perceptron. *Meteorological Applications 2014*, 21(3). DOI: 10.1002/met.1387.

Wang, F.; Mi, Z.; Su, S. and Zhao, H. (2012). Short-Term Solar Irradiance Forecasting Model Based on Artificial Neural Network Using Statistical Feature Parameters. *Energies 2012*, 5(5). 1355-1370. doi:10.3390/en5051355.

WeatherLink. weatherlink. <http://www.weatherlink.com/map.php> (visitado 01/11/2014).

Weather Underground. Weather Underground. <http://www.wunderground.com> (visitado 01/11/2014).

Weather Underground. About Our Data.

<http://www.wunderground.com/about/data.asp> (visitado 01/11/2014).

Wielicki, B.A.; Barkstrom, B. R.; Harrison, E.F.; Lee, R. B.; Smith, G.L. and Cooper, J. E. (1996). Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment. *Bull. Amer. Meteor. Soc.* **1996**, 77, 853–868. doi: [http://dx.doi.org/10.1175/1520-0477\(1996\)077<0853:CATERE>2.0.CO;2](http://dx.doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2)

WikiMapia. WikiMapia. <http://wikimapia.org> (visitado 01/11/2014).

Williams, M.; Cornford, D.; Bastin, L.; Jones, R. and Parker, S. (2009). Automatic Processing, Quality Assurance and Serving of Real-Time Weather Data Over Lightweight Protocols. *In StatGIS spatial statistics and GIS 2009*, Milos Island, Greece.

Yadav, A. K. and Chandel, S. S. (2012). Artificial Neural Network based Prediction of Solar Radiation for Indian Stations. *International Journal of Computer Applications* **2012**, 50(9):1-4, July 2012.

Yadav, A.K. and Chandel, S.S. (2014). Solar radiation prediction using Artificial Neural Networks Techniques: A review. *Renewable and Sustainable Energy Reviews* **2014**, 33 (2014), 772-781.

Yanes, A. (2011). OpenWeather: A Peer-to-Peer Weather Data Transmission Protocol. *Aalto University – School of Electrical Engineering* **2011**. Otaniemi, Espoo.

Zelenka, A.; Perez, R.; Seals, R. and Renné, D. (1999). Effective Accuracy of Satellite-Derived Hourly Irradiances. *Theoretical and Applied Climatology* **1999**, 62(3), 199-207.

## 8 ANEXOS

### 8.1 Lista de redes de estaciones meteorológicas y sus URLs de accesos.

<i>RED</i>	<i>Periodicidad</i>	<i>Latencia de reportes</i>	<i>Protocolo de acceso</i>	<i>Formato de datos</i>	<i>Tipo de tiempo</i>
AEMet (760)	Sub_H=429 30-Min=291 H = 40	120 min	FTP HTTP_GET	CSV HTML	UTC
Agencia Estatal de Meteorología					
DGT (216)	Sub_H=216	5 min	HTTP_GET	HTML	Local
Dirección General de Trafico					
EBRO (38)	30-Min=38	60 min	HTTP_GET	HTML	Local
Confederación hidrográfica del Ebro					

Tabla 24. Resumen de las redes GUB con área de influencia mayor o igual al de región (comunidad autónoma).

<i>RED</i>	<i>Periodicidad</i>	<i>Latencia de reportes</i>	<i>Protocolo de acceso</i>	<i>Formato de datos</i>	<i>Tipo de tiempo</i>
ARAGÓN (48)	30-Min=48	720 min	HTTP_POST	HTML	UTC
Oficina del irrigador del gobierno de Aragón					
CASTILLA LA MANCHA (12)	H=12	5 min	HTTP_GET	HTML	Local
Red de control y vigilancia de la calidad del aire de Castilla la Mancha					
CATALUÑA (168)	30-Min=63 H=105	10 min	HTTP_POST	HTML	UTC
Servicio meteorológico de Cataluña					
VALENCIA-MEDITERRANEAN	30-Min=33 D =12	60 min	HTTP_GET	HTML	UTC
Centro de estudios ambientales de mediterráneo					
MADRID (24)	H=24	60 min	HTTP_GET	HTML	UTC
Red de calidad del aire – Comunidad de Madrid					
VALENCIA (52)	D=52	Día	HTTP_GET	HTML	—
Sistema de información agroclimático para la irrigación – Comunidad Valenciana					
BASQUE COUNTRY (63)	Sub_H=63	5 min	HTTP_GET	HTML	UTC
Agencia Vasca de meteorológica					
GALICIA (85)	Sub_H= 85	5 min	HTTP_GET SOA	XML TXT HTML	UTC
MeteoGalicia					
LA RIOJA (16)	Sub_H=16	15 min	HTTP_GET	CSV HTML HOJA DE CÁLCULO	Local
Estaciones del gobierno de la Rioja					

<b>RED</b>	<b>Periodicidad</b>	<b>Latencia de reportes</b>	<b>Protocolo de acceso</b>	<b>Formato de datos</b>	<b>Tipo de tiempo</b>
CASTILLA Y LEON (51)	30-Min=51	between 6 y 24 hours	FTP	CSV	Local
Red InfoRiego – Comunidad de Castilla y León					
MADRID MUNICIPALITY (24)	H=24	60 min	HTTP_GET	HTML	Local
Red de calidad del aire – Municipalidad de Madrid					
MURCIA (44)	Sub_H=12 D=32	0 min	HTTP_GET	HTML	UTC
Sistema de información agrario – Región de Murcia					
NAVARRA (19)	Sub_H=19	20 min	HTTP_GET	HTML CSV	UTC
Meteorología y climatología de Navarra					

Tabla 25. Resumen de las redes GUB con área de influencia menor al de región (comunidad autónoma).

<b>RED</b>	<b>Periodicidad</b>	<b>Latencia de reportes</b>	<b>Protocolo de acceso</b>	<b>Formato de datos</b>	<b>Tipo de tiempo</b>
AWEKAS (336)	Sub_H=336	10 min	HTTP_GET	HTML FLASH	Local
Sistema de mapa meteorológico automático ( <i>Automatic Weather Map System, AWEKAS</i> )					
CWOP (112)	Sub_H=112	5 min	HTTP_GET	CSV HTML	UTC
Programa de observación meteorológica por ciudadanos ( <i>Citizen Weather Observer Program, CWOP</i> )					
Meteoclimatic (1344)	Sub_H=1186 30-Min = 70 H=88	15 min	HTTP_GET SOA	HTML GeoRSS	UTC
Red Meteoclimatic					
WeatherLink (103)	Sub_H=103	1 min	HTTP_GET SOA	HTML	Local
Sistema <i>Weather Link</i>					
WUNDER (304)	Sub_H=304	1 min	HTTP_GET SOA	HTML CSV XML	Local
<i>Weather Underground</i>					

Tabla 26. Resumen de las REMV con área de influencia mayor al de región (comunidad autónoma).

<i>RED</i>	<i>Periodicidad</i>	<i>Latencia de reportes</i>	<i>Protocolo de acceso</i>	<i>Formato de datos</i>	<i>Tipo de tiempo</i>
MeteoComillas <sup>a</sup> (5)	Sub_H=5	5 min	HTTP_GET	HTML	Local
Red de Comillas					
MeteoSal <sup>b</sup> (8)	Sub_H=8	5 min	HTTP_GET	HTML	UTC
Red meteorológica de aficionados de Cantabria					
UNIMURCIA <sup>c</sup> (1)	Sub_H=1	5 min	HTTP_GET	HTML	Local
Estación meteorológica - Universidad de Murcia					

a- Municipal.

b- Regional.

c- Local.

Tabla 27. Resumen de las REMV con área de influencia menor al de región (comunidad autónoma).

<i>RED</i>	<i>URL</i>
AEMet	<a href="ftp://ftpdatos.aemet.es/datos_observacion/observaciones_diezminutales">ftp://ftpdatos.aemet.es/datos_observacion/observaciones_diezminutales</a>
DGT	<a href="http://infocar.dgt.es/etraffic/Buscador?SensoresMeteorologico=true&amp;IncidenciasMETEOROLOGICA=true&amp;provincia=30&amp;version=mapa&amp;pagina=null&amp;accion_buscar=Buscar">http://infocar.dgt.es/etraffic/Buscador?SensoresMeteorologico=true&amp;IncidenciasMETEOROLOGICA=true&amp;provincia=30&amp;version=mapa&amp;pagina=null&amp;accion_buscar=Buscar</a>
EBRO	<a href="http://195.55.247.237/saihebro/index.php?url=/datos/mapas/tipoestacion:EM">http://195.55.247.237/saihebro/index.php?url=/datos/mapas/tipoestacion:EM</a>
ARAGÓN	<a href="http://servicios.aragon.es/oresa/datosMeteorologicos.inicio.do?javascript=true&amp;sm=2060&amp;cpath=S">http://servicios.aragon.es/oresa/datosMeteorologicos.inicio.do?javascript=true&amp;sm=2060&amp;cpath=S</a>
CASTILLA LA MANCHA	<a href="http://pagina.jccm.es/medioambiente/rvca/estaciones/ultimos/temporales_gen.htm">http://pagina.jccm.es/medioambiente/rvca/estaciones/ultimos/temporales_gen.htm</a>
CATALONIA	<a href="http://www.meteo.cat/xema/AppJava/TaulaEMAs.do">http://www.meteo.cat/xema/AppJava/TaulaEMAs.do</a>
VALENCIA-MEDITERRANEAN	<a href="http://www.ceam.es/ceamet/observaciones/ceam/torres_ceam.html">http://www.ceam.es/ceamet/observaciones/ceam/torres_ceam.html</a>
MADRID	<a href="http://gestiona.madrid.org/azul_internet/html/web/DatosRedAccion.icm?ESTADO_MENU=2_1">http://gestiona.madrid.org/azul_internet/html/web/DatosRedAccion.icm?ESTADO_MENU=2_1</a>
VALENCIA	<a href="http://estaciones.ivia.es/MapaCV.htm">http://estaciones.ivia.es/MapaCV.htm</a>
BASQUE COUNTRY	<a href="http://www.euskalmet.euskadi.net/s07-5853x/es/meteorologia/lectur.apl?e=5">http://www.euskalmet.euskadi.net/s07-5853x/es/meteorologia/lectur.apl?e=5</a>
GALICIA	<a href="http://www2.meteogalicia.es/galego/observacion/estacions/estacionsHistorico.asp?Nest=10152&amp;prov=A%20Coru%F1a&amp;tiporede=automaticas&amp;red=102&amp;idprov=0">http://www2.meteogalicia.es/galego/observacion/estacions/estacionsHistorico.asp?Nest=10152&amp;prov=A%20Coru%F1a&amp;tiporede=automaticas&amp;red=102&amp;idprov=0</a>
LA RIOJA	<a href="http://www.larioja.org/npRioja/default/defaultpage.jsp?idtab=442821">http://www.larioja.org/npRioja/default/defaultpage.jsp?idtab=442821</a>
CASTILLA Y LEON	<a href="ftp://ftp.itacyl.es/Climatologia/DatosHorarios">ftp://ftp.itacyl.es/Climatologia/DatosHorarios</a>
MADRID MUNICIPALITY	<a href="http://www.mambiente.munimadrid.es/svca/index.php?lang=es">http://www.mambiente.munimadrid.es/svca/index.php?lang=es</a>
MURCIA	<a href="http://siam.imida.es/apex/f?p=101:110:1792104273985054">http://siam.imida.es/apex/f?p=101:110:1792104273985054</a>

<b>RED</b>	<b>URL</b>
NAVARRA	<a href="http://meteo.navarra.es/estaciones/mapasdatostiemoreal.cfm">http://meteo.navarra.es/estaciones/mapasdatostiemoreal.cfm</a>
AWEKAS	<a href="http://www.awekas.at/es/temp.php?nid=8">http://www.awekas.at/es/temp.php?nid=8</a>
CWOP	<a href="http://www.wxqa.com">http://www.wxqa.com</a>
Meteoclimatic	<a href="http://www.meteoclimatic.com">http://www.meteoclimatic.com</a>
WeatherLink	<a href="http://www.weatherlink.com/map.php">http://www.weatherlink.com/map.php</a>
WUNDER	<a href="http://www.wunderground.com/weatherstation/ListStations.asp?selectedCountry=Spain&amp;MR">http://www.wunderground.com/weatherstation/ListStations.asp?selectedCountry=Spain&amp;MR</a>
MeteoComillas	<a href="http://www.meteocomillas.com/index.php?option=com_content&amp;view=article&amp;id=121&amp;Itemid=146">http://www.meteocomillas.com/index.php?option=com_content&amp;view=article&amp;id=121&amp;Itemid=146</a>
MeteoSal	<a href="http://meteosal.com/index.php/tiempo-real.html">http://meteosal.com/index.php/tiempo-real.html</a> [Note: Each station has its own URL access]
UNIMURCIA	<a href="https://estacion.um.es/estacion/servlet/estacion.EstacionInit">https://estacion.um.es/estacion/servlet/estacion.EstacionInit</a>

Tabla 28. URL de acceso a cada red.

## **8.2 Publicaciones científicas generadas en el periodo de Tesis**

### **8.2.1. Artículos en revistas indexadas por la *ISI Web of Science***

Gutiérrez-Corea, F.V.; Callejo-Callejo, M.A.; Moreno-Regidor, M.P. and Velasco-Gómez, J. (2014). Spatial Estimation of Sub-Hour Global Horizontal Irradiance Based on Official Observations and Remote Sensors, *Sensors* **14(4)**, DOI: 10.3390/s140406758, April 2014, pp 6758-6787, ISSN: 1424-8220.  
[Science Citation Index (SCI)]

Gutiérrez-Corea, F.V.; Manso-Callejo, M.A. and Vázquez-Hoehne, A. (2013). Assessment of the Availability of Near-Real Time Open Weather Data Provided by Networks of Surface Stations in Spain, *Earth Science Informatics* **6(3)**, DOI: 10.1007/s12145-013-0120-8, September 2013, pp 145-163, ISSN: 1865-0473  
[Science Citation Index (SCI)]

### **8.2.2. Próximo a enviarse a revista indexada por la *ISI Web of Science***

Forecasting of Short-Term Solar Irradiance Based on Artificial Neural Networks and Data from Neighbouring Meteorological Stations. (*To Send*) *Solar Energy* (2015), ISSN: 0038-092X.  
[Science Citation Index (SCI)]

### **8.3.3. Capítulos de libros y artículos publicados en congresos**

Moreno, F.J.; Gutiérrez-Corea, F.V., Manso-Callejo M.A. (2012). Using 3D GeoDesign for Planning of new Electricity Networks in Spain in *Computational Science and Its Applications (ICCAS 2012)*. B. Murgante, O. Gervasi, S. Misra, N. Nedjah, A. Rocha, D.Taniar, B. Apduhan (Eds.), *Lecture Notes in Computer Science (LNCS)*, DOI10.1007/978-3-642-31125-3\_35, Springer Computer Science Editorial. Volume 7333, ISBN: 978-3-642-31074-4  
[Conference Proceedings Citation Index- Science (CPCI-S)]

Gutiérrez-Corea, F.V.; Malinowski, E. and Rodríguez, A. (2012). Lenguajes Geoespaciales en *Fundamentos de las Infraestructuras de Datos Espaciales (IDE)*; M. A. Bernabe-Poveda, C. Lopez (Eds). ISBN: 987-84-939196-6-7. pp. 275-285. Madrid, Spain. Editorial Universidad Politécnica de Madrid (UPM-PRESS), Serie Científica UPM.

Gutiérrez-Corea, F.V., Manso-Callejo, M.A.; Moreno, F.J. and Castaneda, E. (2012). Integrated Geoprocessing for Generation of Affected Assets and Rights Reports for High Voltage Electrical Infrastructures. *GEOProcessing 2012*. The Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services, p. 59-65, ISSN: ISBN: 978-1-61208-178-6, Valencia, Spain, February 2012.

Gutiérrez-Corea, F.V. and Manso-Callejo, M.A. (2012). Análisis espacio-temporal de la disponibilidad de observaciones meteorológicas en tiempo real: Casos de estudio: AEMET, JCYL y METEOCLIMATIC. *Jornadas Científicas de la Asociación Meteorológica Española (AME)*, Museo de la Ciencia COSMOCAIXA, Alcobendas, Madrid, Spain (2012), ISSN: ISBN10 = 84-695-6430-7; ISBN13 = 978-84-695-6430-1.

Castañeda, E.; Manso-Callejo, M.A.; Moreno, F.J. and Gutiérrez-Corea, F.V. (2012). E-Report generator supporting communications and fieldwork: a practical case of electrical network expansion projects. ISBN: 978-1-61208-178-6.

Moreno, F.J.; Gutiérrez-Corea, F.V.; Manso-Callejo, M.A. and Castaneda E. (2011). Geoservicios-Webs en la comunicación de afecciones por obras lineales: Líneas eléctricas de alta tensión. *Jornadas Ibéricas de Infraestructuras de Datos Espaciales 2011*, Barcelona, Spain, November 2011.

#### **8.3.4. Posters y presentaciones en congresos**

Gutiérrez-Corea, F.V. and Manso-Callejo M.A., (2012). Geospatial Standardized Services for Integration of Weather Data Coming From Public and Voluntary Stations. *EuroGEOSS Advancing the vision for GEOSS*, January Madrid, Spain , January 2012 (Poster).

Gutiérrez-Corea, F.V.; Moreno, F. J. and Manso-Callejo, M. A. (2012). GeoProcesamiento integrado con ArcObjects para generar afecciones por infraestructuras eléctricas de alta tensión. *Conferencias ESRI - España 2012*, Madrid, Spain, October 2012 (Presentation).