Universidad de Alcalá

Programa de Doctorado en Electrónica: Sistemas electrónicos avanzados. Sistemas inteligentes. (RD 99/2011) – D411

Prediction techniques adapted to the estimation of energy production of photovoltaic installations integrated in virtual power plants

Tesis Doctoral presentada por
Guillermo Moreno Baeza

Directores:

Dr. Pedro Martín Sánchez

Dr. Carlos Santos Pérez

Alcalá de Henares, 2022.

Agradecimientos

Primero agradecer a mis formadores, personas de gran sabiduría que se han esforzado en ayudarme hasta llegar al punto en el que me encuentro. Sencillo no ha sido el proceso y gracias a las ganas de transmitirme sus conocimientos he logrado grandes objetivos como culminar el desarrollo de mi tesis con éxito.

En segundo lugar, dar las gracias a mi pareja por ayudarme a hacer de este camino un trayecto más ameno, también a mis amigos, personas que me han apoyado en todo momento y me han enseñado que soy capaz de esto y mucho más.

Por último y no menos importante, agradecer a mi familia que siempre ha estado ahí, a pesar de la distancia y el tiempo dedicado a este trabajo que ya forma parte de mí. Sin ellos todo este camino no hubiese tenido ni siquiera un comienzo.

Resumen

La implantación de la energía solar fotovoltaica se encuentra en auge desde la pasada década, alentada por su considerable reducción de costes. Sin embargo, su integración a baja escala en la red eléctrica genera una serie de problemas que siguen sin resolverse. La creciente integración de las fuentes de carácter renovable sin una predicción de potencia eficiente dificulta el correcto funcionamiento de la red y desincentiva su uso. Para paliar este problema, se define y desarrolla el concepto de central eléctrica virtual, que permiten gestionar los recursos energéticos distribuidos con el objetivo de facilitar la integración de las fuentes de energía renovables. La presente tesis introduce una estrategia indirecta de predicción de potencia fotovoltaica a través de la irradiancia que permite facilitar la integración de las instalaciones fotovoltaicas en las centrales eléctricas virtuales, utilizando información gratuita de libre acceso junto a la de la propia instalación, incluso en lugares donde el histórico de irradiancia es exiguo o inexistente. Los resultados han sido contrastados con artículos de la literatura para demostrar su potencial. La estrategia ha sido probada en un nodo real, situado en la Universidad de Alcalá, y una central eléctrica virtual emulada con nodos localizados en la Comunidad de Madrid.

En la tesis se distinguen dos estrategias de predicción de irradiancia. Por una parte, una predicción híbrida de horizonte temporal diario que se compone de un modelo innovador basado en horas similares junto a otro basado en redes neuronales, cuyas predicciones se ponderan dinámicamente. En históricos reducidos, el modelo basado en horas similares produce las mayores precisiones, mientras que el modelo basado en redes neuronales reduce su error a lo largo del histórico hasta superar a este, a partir de un histórico de cuatro meses en el caso de estudio. Al convertir la instalación en un nodo dentro de un entorno de central eléctrica virtual, el modelo basado en horas similares obtiene especial relevancia, al reducir el error producido en mayor medida en comparación con otras técnicas. Por otra parte, la segunda estrategia tiene un horizonte temporal intradiario y está basado en redes neuronales recurrentes de memoria a corto plazo. En este caso, la escasez del histórico se evita usando información vía satélite de irradiancia, al ser la única información que necesita el modelo. La predicción se actualiza conforme se obtienen nuevas medidas de irradiancia en la instalación, optimizando su precisión en todo momento.

Una vez predicha la irradiancia bajo diferentes horizontes temporales, se utiliza un método analítico a partir de artículos de la literatura para conocer la potencia fotovoltaica generada en la instalación, obteniendo un error bastante reducido en el método. Seguidamente, se evalúa la potencia generada en cada nodo en una central eléctrica virtual, observando que el error de predicción se reduce considerablemente y se mantienen las conclusiones observadas en los nodos individuales. A partir de las predicciones de potencia fotovoltaica se generan intervalos de

predicción, ofreciendo información sobre el posible rango de potencia fotovoltaica generada en la instalación, bajo un nivel de confianza definido, considerando una distribución Laplaciana del error. Los resultados muestran que los intervalos representan adecuadamente el nivel de confianza seleccionado.

Estimada la potencia fotovoltaica producida bajo diferentes horizontes temporales, se unifican las predicciones para generar en todo momento la predicción de potencia óptima. Para comprobar el potencial de los resultados finales, se emplea la estrategia en un caso de estudio de un sistema de gestión energético de baterías en el nodo real de pruebas de la Universidad. Se observa que las baterías instaladas se amortizan a lo largo de su vida útil, permitiendo utilizar estos elementos sin una pérdida económica asociada dentro de un entorno de central eléctrica virtual, donde resultan esenciales.

Abstract

The implementation of photovoltaic solar energy is increasing since the last decade, encouraged by its considerable cost reduction. However, its low-scale integration into the electrical grid generates a series of problems that remain unresolved. The increasing integration of renewable sources without efficient power forecasting hinders the correct operation of the grid and discourages their use. In order to mitigate this problem, the concept of virtual power plant is defined and developed, which allows the management of distributed energy resources with the objective of facilitating the integration of renewable energy sources. The present thesis introduces an indirect strategy of photovoltaic power prediction through irradiance that allows to facilitate the integration of photovoltaic installations in virtual power plants, using free access data with information of the site, even in places where the irradiance historical data is scarce or nonexistent. The results are validated with articles from the literature to demonstrate its potential. The strategy is tested in a real node, located at the University of Alcala, and a virtual power plant emulated with nodes located in the Community of Madrid.

Two irradiance forecasting strategies are distinguished in the thesis. On the one hand, a hybrid day-ahead prediction that is composed of an innovative similar hours-based model together with an artificial neural network-based model, whose predictions are dynamically weighted. At reduced historical datasets, the similar hours-based model produces the highest accuracies, while the neural networks-based model reduces its error over the historical dataset until it outperforms the similar hours-based model, from a four-month historical dataset in the case study. By converting the site into a node within a virtual power plant environment, the similar hours-based model obtains special relevance, since it reduces the error produced to a greater extent compared to other techniques. On the other hand, the second strategy has an intra-day time horizon and it is based on long short-term memory recurrent neural networks. In this case, the shortage of historical data is avoided by using satellite irradiance information, as this is the only information needed by the model. The prediction is updated as new irradiance measurements are obtained at the facility, optimizing its accuracy during the day.

Once the irradiance is predicted under different time horizons, an analytical method is used from articles in the literature to determine the photovoltaic power generated in the site, obtaining a reduced error of the method. Afterwards, the power generated at each node in the virtual power plant is evaluated, noticing that the prediction error is considerably reduced and the conclusions observed in the individual nodes are maintained. From the photovoltaic power forecasts, prediction intervals are generated, providing information about the plausible range of photovoltaic power values obtained at the site, for a defined confidence level, considering a

Laplacian distribution of the error. The results show that the intervals properly represent the selected confidence level.

Once the photovoltaic power produced under different time horizons is estimated, the predictions are unified to generate the optimal power prediction at a defined time instant. To test the potential of the final results, the strategy is employed in a case study of a battery energy management system in the real virtual power plant node in the University. It is observed that the batteries are amortized over their lifetime, allowing the use of these elements without an associated economic loss in a virtual power plant environment, where they are essential.

Índice de la tesis

Índice de figuras 9

Índice de tablas 13

Lista de acrónimos 14

1 Introducción 16

1.1 Relativo al tema de estudio 16

1.2 Objetivos de la tesis 18

1.3 Metodología de trabajo y alcance 18

1.4 Estructura de la tesis 21

1.5 Publicaciones del autor 23

2 Estado del arte 25

2.1 Generación renovable en VPPs 25

2.2 Técnicas de predicción fotovoltaica 28

2.3 Conclusiones 34

3 Aspectos generales a considerar 36

3.1 Contexto del entorno de trabajo 36

3.2 Definiciones temporales 39

3.3 Descripción global del modelo 41

4 Variables de entrada 46

4.1 Información disponible en el modelo 46

4.2 Preprocesado de los mapas NWP 47

4.2.1 Obtención y análisis de los mapas 47

4.2.2 Factor de cobertura de nubes 49

4.3 Variables de relevancia 53

4.4 Clasificación del error 54

4.5 Sobremuestreo y faltas en la medida 57

4.6 Conclusiones 59
5 Estrategia de predicción diaria 60
5.1 Descripción de la estrategia de predicción 60
5.1.1 Predicción basada en horas similares 62
5.1.2 Predicción basada en redes neuronales 64
5.1.3 Hibridación de técnicas 67
5.2 Resultados 67
5.2.1 Evaluación del modelo para un nodo real de VPP 68
5.2.2 Evaluación del modelo para una VPP emulada 71
5.3 Conclusiones 75
6 Estrategia de predicción intradía 77
6.1 Descripción del modelo 77
6.2 Predicción de irradiancia y resultados 78
6.3 Conclusiones 83
7 Estimación de potencia 84
7.1 Método analítico 84
7.1.1 Irradiancia en el plano de los paneles 84
7.1.2 Irradiancia efectiva en el panel 86
7.1.3 Temperatura de los paneles 87
7.1.4 Potencia de los paneles 88
7.2 Resultados del método analítico 89
7.3 Precisión en la predicción de potencia 90
7.3.1 Predicción de potencia intradiaria 90
7.3.2 Predicción de potencia diaria 93
7.4 Conclusiones 94
8 Intervalos de predicción 96

8.1 Intervalos de predicción diario 97
8.2 Intervalos de predicción intradiario 99
8.3 Conclusiones 100
9 Análisis de las predicciones 102
9.1 Hibridación de estrategias 102
9.2 Caso de estudio: BEMS 104
9.3 Conclusiones 110
10 Conclusiones 112
10.1 Discusión de resultados 112
10.2 Contribuciones, limitaciones y líneas futuras 115
11 Conclusions 117
11.1 Discussion of results 117
11.2 Contributions, limitations and future lines of research 119
Anexos 121
Anexo I: Medidas del error cometido 121
Anexo II: Hojas de características 122
Anexo III: Radiación y nubosidad 127
Anexo IV: RNN-LSTM 128
Bibliografía 131

Índice de figuras

Figura 1.- Esquema eléctrico y de comunicaciones de una BTS transformada en un nodo de VPP. Página 36.

Figura 2.- Disposición de los paneles solares fotovoltaicos instalados en la Universidad de Alcalá, junto a su numeración y conexionado. Página 37.

Figura 3.- Estación meteorológica inalámbrica: (a) piranómetro; (b) PT100 fijado al módulo; (c) anemómetro; y (d) electrónica de adquisición. Página 38.

Figura 4.- Variables temporales utilizadas para definir las predicciones realizadas en la tesis. Página 41.

Figura 5.- Esquema general de la estrategia de predicción. Página 42.

Figura 6.- Proceso de identificación de la temperatura ambiente asociada a cada pixel. Comparando la imagen inicial de la izquierda con los colores de la leyenda se obtiene la imagen final de la derecha que reconoce adecuadamente esta variable. Página 48.

Figura 7.- NWP de nubosidad del 1 de marzo de 2020 a las 16:00 horas. Imagen descargada el 29 de febrero de 2020. Información procedente de la Agencia Estatal de Meteorología, AEMET©. Página 50.

Figura 8.- Posicionamiento horario del sol durante el equinoccio de primavera, definiendo las variables más relevantes en el plano terrestre para obtener el CCF. Página 51.

Figura 9.- Diagrama de bloques del procedimiento para obtener el CCF de los NWP. Página 52.

Figura 10.- Matriz de PCC entre todas las variables consideradas en el estudio, para la estación meteorológica de Juan Carlos I. Página 54.

Figura 11.- Diferencias horarias entre los mapas NWP de temperatura ambiente y las medidas reales de la estación de Juan Carlos I entre diciembre de 2019 y mayo de 2020. Página 55.

Figura 12.- Error cometido entre dos predicciones de temperatura ambiente el 1 de marzo de 2020 a las 16 horas en la Comunidad de Madrid, con una diferencia de tiempo de espera entre predicciones de 6 horas. Información elaborada utilizando, entre otras, la obtenida de la AEMET©. Página 56.

Figura 13.- Error cometido entre dos predicciones de temperatura ambiente el 1 de marzo de 2020 a las 16 horas en la Comunidad de Madrid, con una diferencia de tiempo de espera entre predicciones de 42 horas. Información elaborada utilizando, entre otras, la obtenida de la AEMET©. Página 56.

Figura 14.- Medidas de GHI de la instalación y del CAMS, junto a las medidas del error calculadas entre las dos bases de datos. Página 58.

Figura 15.- Medidas de temperatura ambiente del nodo real de pruebas y de las predicciones de la AEMET con una resolución temporal de 15 minutos, junto a las medidas del error calculadas entre las dos bases de datos. Página 58.

Figura 16.- Reconstrucción de las medidas de GHI y temperatura ambiente en el nodo real de pruebas situado en la Universidad de Alcalá. Página 59.

Figura 17.- Esquema general del modelo usado para predicciones con un horizonte temporal diario, que consiste en dos fases: una etapa previa de procesamiento de datos y la predicción híbrida. La estrategia de predicción está basada en ANNs y un modelo basado en horas similares, ponderando dinámicamente la salida en función del tipo de día. Página 61.

Figura 18.- Diagrama de bloques de la estrategia de horas similares para predecir la GHI con un horizonte temporal diario. Página 64.

Figura 19.- rMAE de la predicción basada en ANNs en función del algoritmo de aprendizaje seleccionado. Se observa que el SCG minimiza el error producido hasta 70 días de histórico, mientras que LM ofrece una mayor precisión a partir de ese instante. Página 65.

Figura 20.- Error de predicción de los modelos empleados en el nodo real de pruebas situado en la Universidad de Alcalá, en términos del (a) rMAE y (b) rRMSE. Página 69.

Figura 21.- Pesos usados en la ANN para el modelo híbrido en función del histórico del nodo real de pruebas situado en la Universidad de Alcalá. Página 70.

Figura 22.- Predicción de los valores de GHI en el nodo real de pruebas situado en la Universidad de Alcalá para los diferentes modelos en tres días consecutivos, mostrando diferentes tipos de día. Página 71.

Figura 23.- Localización de las estaciones a nivel de suelo en la Comunidad de Madrid empleadas en el presente estudio. Página 72.

Figura 24.- Error de predicción de los modelos empleados en la VPP emulada, en términos del (a) rMAE y (b) rRMSE. Página 73.

Figura 25.- Pesos usados en la ANN para el modelo híbrido en función del histórico de la VPP emulada. Página 74.

Figura 26.- Predicción de los valores de GHI en la VPP emulada para los diferentes modelos en tres días consecutivos, mostrando diferentes tipos de día. Página 75.

Figura 27.- Esquema general del modelo utilizado en predicciones intradiarias, distinguiendo: (i) adquisición de datos; (ii) variables de entrada y preprocesado; (iii) predicción de irradiancia; (iv) estimación de potencia; y (v) toma de decisiones y actuación del EMS. Página 78.

Figura 28.- Diagrama de flujo del modelo de predicción basado en RNN-LSTMs. Página 79.

Figura 29.- Medidas reales y predicciones intradiarias para un día concreto. Las líneas discontinuas son las predicciones a diferentes plazos intradiarios de

lanzamiento, mientras que las líneas de puntos son las predicciones a diferentes tiempos de espera. Al seleccionar ambos parámetros se define un único punto mostrado como un rombo rojo. Página 81.

Figura 30.- Matrices de error entre medidas reales y predichas de GHI en función del plazo intradiario de lanzamiento y el tiempo de espera: (a) MAE; (b) RMSE; (c) rMAE; y (d) rRMSE. Página 82.

Figura 31.- Comparación entre valores medidos de potencia y estimados con medidas reales de GHI. Los días seleccionados son: (a) día nublado (17 de mayo de 2021: $rMAE = 3.04\%$ $rRMSE = 3.90\%$); (b) día encapotado (1 de junio de 2021: $rMAE = 4.03\%$ $rRMSE = 5.95\%$); y (c) día soleado (4 de junio de 2021: $rMAE = 2.54\%$ $rRMSE = 3.44\%$). Página 89.

Figura 32.- Matrices del error de la potencia en el nodo real de pruebas en función del plazo intradiario de lanzamiento y el tiempo de espera: (a) MAE; (b) RMSE; (c) rMAE; y (d) rRMSE Página 91.

Figura 33.- Localización de las estaciones a nivel de suelo en la Comunidad de Madrid empleadas en el presente estudio Página 92.

Figura 34.- Matrices del error de la potencia en una VPP emulada de 23.76 kW_p, compuesta por ocho nodos situados en la Comunidad de Madrid, en función del plazo intradiario de lanzamiento y el tiempo de espera: (a) MAE; (b) RMSE; (c) rMAE; y (d) rRMSE Página 93.

Figura 35.- Error de predicción del modelo híbrido para el nodo real de pruebas y la VPP en términos del rMAE y el rRMSE. Página 94.

Figura 36.- Intervalos de predicción considerados en el nodo real de pruebas para un día específico y un nivel de confianza del 80%. Página 97.

Figura 37.- Distribución del error para cada grupo seleccionado en función del CCF. Para generar intervalos de predicción, se asume una distribución Laplaciana. Página 98.

Figura 38.- Distribución del error para cada grupo seleccionado, de acuerdo al tiempo de espera y el plazo intradiario de lanzamiento. Para generar intervalos de predicción, se asume una distribución Laplaciana. Página 99.

Figura 39.- Diferencia absoluta entre el PICP y el nivel de confianza para cada grupo seleccionado en función del tiempo de espera, el plazo intradiario de lanzamiento y el tipo de día: (a) soleado; (b) nublado; y (c) encapotado. Página 100.

Figura 40.- Matrices del error de la predicción de potencia con un horizonte temporal diario en una VPP emulada de 23.76 kW_p, compuesta por ocho nodos de situados en la Comunidad de Madrid, en función del plazo intradiario de lanzamiento y el tiempo de espera: (a) MAE; (b) RMSE; (c) rMAE; y (d) rRMSE. Página 103.

Figura 41.- Selección de la estrategia de predicción en función del tiempo de espera y el plazo intradiario de lanzamiento, para diferentes medidas del error consideradas: (a) rMAE; y (b) rRMSE Página 104.

Figura 42.- Planificación inicial del BEMS durante la noche del 16 de marzo de 2021, seleccionando el proceso de carga y descarga del BESS al usar la predicción de estrategia diaria. Página 105.

Figura 43.- Planificación del BEMS para un plazo intradiario de lanzamiento de (a) dos horas y (b) tres horas, el 16 de marzo de 2021, modificando el proceso de carga y descarga del BESS con la llegada de nuevas medidas. Página 106.

Figura 44.- Planificación del BEMS para un plazo intradiario de lanzamiento de cuatro horas, el 16 de marzo de 2021, momento a partir del que también se lanza la estrategia diaria del día siguiente. Página 107.

Figura 45.- Resto de planificación del BEMS el 16 de marzo de 2021, para un plazo intradiario de lanzamiento que va desde 5 horas en (a) hasta 10 horas en (f). Página 108.

Figura 46.- Resultados de la planificación final del BEMS durante 5 días consecutivos, del 14 al 18 de marzo de 2021. Se observa que las baterías realizan un ciclo de carga y descarga al día, optimizando el beneficio económico de la instalación. Página 109.

Figura 47.- Resultados obtenidos tras la gestión del BEMS durante los días mostrados en la Figura 46: (a) vida útil del BESS medida según su capacidad; (b) ahorro económico acumulado en €/kWh; y (c) ahorro económico acumulado en €. Página 110.

Figura 48.- Histograma del CSI (κ) con respecto a la GHI determinada para 401 días, medidas en la ciudad de Oahu, en Hawaii, en 2010 y 2011. Página 128.

Índice de tablas

Tabla 1.- PCC de las variables con respecto a la GHI para las diferentes estaciones meteorológicas utilizadas en el estudio. Página 53.

Tabla 2.- Análisis de sensibilidad de las entradas utilizadas en la ANN. Página 66.

Tabla 3.- Características del modelo basado en ANN-MLPs usado en el modelo híbrido de predicción con un horizonte temporal diario. Página 67.

Tabla 4.- Medidas del error del rMAE y el rRMSE tras 180 días de histórico. Página 74.

Tabla 5.- Parámetros seleccionados en la RNN-LSTM. Página 80.

Tabla 6.- Estudio comparativo entre los resultados de predicción intradiaria de GHI en la tesis y los desarrollados por otros autores en la literatura. Página 82.

Tabla 7.- Valores promedios de los coeficientes para paneles de diferentes tecnologías. Página 87.

Tabla 8.- rMAE y el rRMSE de la predicción de potencia tras 180 días de histórico. Página 94.

Tabla 9.- Errores medios cometidos, intervalos de predicción y PICP para cada uno de los grupos seleccionados en función del CCF. Página 98.

Lista de acrónimos

Acrónimo	*Definition*	Definición
$\boldsymbol{\mu}$	*Uncertainty Threshold*	Umbral de Incertidumbre
AEMET	*State Meteorological Agency*	Agencia Estatal de Meteorología
ANN	*Artificial Neural Network*	Red Neuronal Artificial
ARIMA	*AutoRegressive Integrated Moving Average*	Autorregresivo Integrado de Media Móvil
BC	*BlockChain*	Cadena de Bloques
BEMS	*Battery Energy Management System*	Sistema de Gestión Energético de Baterías
BESS	*Battery Energy Storage System*	Sistema de Almacenamiento Energético de Baterías
BPTT	*BackPropagation Through Time*	Propagación Inversa a Través del Tiempo
BR	*Bayesian Regularization*	Regularización Bayesiana
BTS	*Base Transceiver Station*	Estación Base para Comunicaciones Móviles
CAMS	*Copernicus Atmosphere Monitoring Service*	Servicio de Monitorización de Atmósfera Copernicus
CCF	*Cloud Cover Factor*	Factor de Cobertura de Nubes
CNN	*Convolutional Neural Network*	Red Neuronal Convolucional
CSI	*Clear Sky Index*	Índice de Cielo Despejado
DER	*Distributed Energy Resource*	Recurso Energético Distribuido
DG	*Distributed Generation*	Generación Distribuida
DNI	*Direct Normal Irradiance*	Irradiancia Normal Directa
DoD	*Depth of Discharge*	Profundidad de Descarga
ED	*Euclidean Distance*	Distancia Euclídea
EMS	*Energy Management System*	Sistema de Gestión Energético
ESS	*Energy Storage System*	Sistema de Almacenamiento de Energía
GHI	*Global Horizontal Irradiance*	Irradiancia Global Horizontal
GRU	*Gated Reunited Unit*	Unidades Recurrentes Cerradas
k-NN	*k–Nearest Neighbours*	k vecinos más próximos

LM	*Levenberg-Marquardt*	Levenberg-Marquardt
LSTM	*Long Short-Term Memory*	Memoria a corto plazo
MAE	*Mean Absolute Error*	Error Medio Absoluto
MAPE	*Mean Absolute Percentage Error*	Error Absoluto Medio Porcentual
MBE	*Mean Biased Error*	Error de Tendencia Medio
ML	*Machine Learning*	Aprendizaje Automático
MLP	*MultiLayer Perceptron*	Perceptrón Multicapa
MPPT	*Maximum Power Point Tracking*	Seguimiento del Punto de Máxima Potencia
NOCT	*Nominal Operating Cell Temperature*	Temperatura de Operación Nominal de Célula
NRMSE	*Normalized Root Mean Squared Error*	Raíz del Error Cuadrático Medio Normalizado
NWP	*Numerical Weather Prediction*	Predicción Numérica del Tiempo
OMIE	*Nominated Electricity Market Operator*	Operador del Mercado Ibérico de la Energía
PCC	*Pearson Correlation Coefficient*	Coeficiente de Correlación de Pearson
PICP	*Prediction Interval Coverage Probability*	Probabilidad de Cobertura del Intervalo de Predicción
ReLU	*Rectified Linear Unit*	Unidad Lineal Rectificada
RES	*Renewable Energy Source*	Fuente de Energía Renovable
rMAE	*relative Mean Absolute Error*	Error Medio Absoluto relativo
RMSE	*Root Mean Square Error*	Raíz del Error Cuadrático Medio
RNN	*Recurrent Neural Network*	Red Neuronal Recurrente
rRMSE	*Relative Root Mean Square Error*	Raíz del Error Cuadrático Medio relativo
SCG	*Scaled Conjugated Gradient*	Gradiente Escalado Conjugado
SoC	*State of Charge*	Estado de Carga
STC	*Standard Test Conditions*	Condiciones Test Estándar
SVM	*Support Vector Machine*	Máquina de Soporte Vectorial
UTC	*Universal Time Coordinated*	Tiempo Coordinado Universal
VPP	*Virtual Power Plant*	Central Eléctrica Virtual

1 Introducción

Este capítulo inicial pretende aportar las bases del documento con el fin de guiar al lector a través del contenido desarrollado en la tesis que se presenta. Para ello, en el apartado 1.1, se muestra de forma resumida los avances históricos y las motivaciones que han ayudado a la realización de la presente tesis. En el apartado 1.2 se describen los objetivos a alcanzar para su desarrollo y en el apartado 1.3 la metodología utilizada y su alcance. La estructura del documento se muestra en el apartado 1.4, desglosando los capítulos que contiene. Finalmente, las publicaciones del autor y su aportación a nivel de aprendizaje e innovación científica se han descrito en el apartado 1.5.

1.1 Relativo al tema de estudio

El campo científico relacionado con las redes eléctricas y sistemas de generación evoluciona a una velocidad sin precedentes apoyada por el progreso de la electrónica y la informática. La red eléctrica tradicional dispone de un sistema centralizado donde la energía es enviada a través de un sistema de transporte hacia los consumidores. La carga energética que hay que transportar en este caso es enorme y en constante crecimiento, pudiendo provocar cortocircuitos en cadena difíciles de controlar debido a las grandes potencias que se mueven actualmente. Estos sistemas energéticos tradicionales hacen uso principalmente de fuentes de carácter fósil y nuclear para generar un buen control de calidad de red, limitando el uso de energías renovables, que resultan difíciles de controlar y generan notables inconvenientes si no se complementan con almacenamiento y/o técnicas de adquisición de datos y predicción. La evolución tecnológica de los equipos tradicionales hacia los sensores inteligentes ha arraigado un nuevo enfoque eléctrico guiado por la necesidad de ofrecer sistemas más económicos, sostenibles, diversos y descentralizados. La generación distribuida (*Distributed Generation*, DG) fomenta el uso de fuentes de carácter renovable, ganando relevancia junto a otras como ciclo combinado o cogeneración. Esto permite a su vez aumentar el grado de explotación de estas fuentes y situarlas a la altura de las más tradicionales.

De entre las fuentes renovables, la que ha ofrecido una mayor evolución en los últimos años es la de carácter fotovoltaico. Cada año, surgen nuevas tecnologías y materiales que ofrecen mejoras en la eficiencia del sistema fotovoltaico, llegando las células solares actualmente a competir económicamente con fuentes energéticas de carácter fósil [1]. Este interés innovador es mayor al emplearse este tipo de fuente con equipos que piden un fuerte avance tecnológico, como es el desarrollo de las baterías, o han vislumbrado un nuevo enfoque hacia la innovación, por ejemplo, mediante el uso de microinversores a nivel tecnológico o, a nivel legislativo, aboliendo la normativa que establecía una gran restricción en este tipo de campos, y frenaba así el progreso investigador.

Actualmente, la generación eléctrica ha evolucionado a un campo más diversificado en el que nuevos conceptos salen a la luz y permiten el uso de todo tipo de fuentes generadoras. En la presente tesis, se busca dar cabida a la explotación de la tecnología fotovoltaica y la interoperabilidad de las partes para crear cierta independencia con la red eléctrica central, mediante las centrales eléctricas virtuales (*Virtual Power Plants*, VPPs) [2], las cuales permiten la agregación de unidades de DG, cargas controlables y equipos de almacenamiento, unidos a un clúster dentro de una entidad virtual, que es la responsable de gestionar los flujos energéticos dentro del mismo e intercambiar energía con la red eléctrica central. Por último, este concepto permite a todo consumidor unirse al mercado eléctrico al instalar fuentes de carácter renovable, y consigue adaptar dicho mercado al tradicional, incentivando la introducción de todo un nuevo concepto de red eléctrica mediante la creación de contratos inteligentes de forma segura en una cadena de bloques (*BlockChain*, BC), siendo una tecnología innovadora con gran potencial a nivel eléctrico y de comunicaciones aún por indagar [3]. La tecnología BC crea un mercado de energía local de forma descentralizada que dispone de una alta resiliencia al fallo y evita ser manipulada de forma maliciosa, ya que mantiene un histórico detallado de transacciones pasadas que no pueden ser modificadas de forma retrospectiva. Mediante el protocolo IEC-61850 se aporta un grado de estandarización a esta tecnología que permite globalizar todo este concepto [4].

La VPP conecta diferentes puntos geográficos, llamados nodos, con la red principal, permitiendo el soporte energético cuando sea necesario, de forma bidireccional. Mediante este concepto se crea un sistema desagregado en el que su generación es en buena parte de carácter renovable, apoyado mediante técnicas de predicción, y que permite la entrada al mercado de todos los elementos generadores, consumidores y prosumidores [5] que componen la red eléctrica. Es por ello que el recurso solar requiere de un gran conocimiento y desarrollo para participar de forma agregada al mercado eléctrico, al necesitar técnicas de predicción eficientes y económicas, considerando la coexistencia de una gran cantidad de nodos dentro de un entorno tanto público como privado.

De esta forma, surge la necesidad de generar técnicas de predicción fotovoltaica dentro del entorno VPP, difiriendo sustancialmente de otras metodologías en la literatura que asumen ciertas hipótesis, en parte de aplicabilidad utópica, entre estos: (i) la necesidad de recopilar variables exógenas para la estrategia que no se encuentran disponibles públicamente; (ii) no contemplar nuevas instalaciones que pueden no disponer de una base de datos suficientemente amplia; (iii) no generalizar la predicción para múltiples emplazamientos en diferentes puntos geográficos; (iv) no añadir intervalos de predicción que evalúen la incertidumbre de los resultados; entre otras.

En esta tesis se busca paliar problemas destacados previamente, buscando una estrategia fácilmente accesible que compita con los mostrados en la literatura y genere información valiosa, permitiendo su integración

dentro de una estructura agregada como es la VPP. La misión de la tesis reside en aportar técnicas de predicción de carácter fotovoltaico que permitan ayudar al desarrollo del concepto VPP, el cual divisa un futuro único en el mundo de la energía.

1.2 Objetivos de la tesis

En la presente tesis, a nivel general, se estudian diversas técnicas de predicción solar fotovoltaica que cumplan con los requisitos necesarios para su adaptación a las VPP. Entre estas labores: (i) diseñar una estrategia óptima que realice predicciones bajo varios horizontes temporales de la energía producida; (ii) incorporar intervalos que permitan evaluar la incertidumbre del sistema; (iii) facilitar el funcionamiento del sistema de gestión energético (*Energy Management System*, EMS); entre otras. Así mismo, para ofrecer una estrategia generalizada que no esté limitada por la privacidad de la información, la adquisición de datos necesaria para la predicción se obtiene de webs de acceso público ofrecidas por diversas agencias [6,7], siendo requerimiento único por parte del usuario incorporar un piranómetro en la instalación para la predicción. Esta estrategia toma como referencia la localización de cada instalación y se adapta a nuevas instalaciones fotovoltaicas, favoreciendo notablemente la precisión del sistema al trabajar bajo el concepto de VPP.

Todo este marco investigador se organiza en un entorno real de pruebas, mediante la aplicación de dichas técnicas a una instalación fotovoltaica existente en la Escuela Politécnica Superior de la Universidad de Alcalá (España) [8], adaptada a su funcionamiento como nodo VPP. Posteriormente, este caso se emula dentro de un conjunto de nodos, con el fin de comprobar que los resultados particulares de una única instalación se mantienen coherentes y generan un análisis favorable en la VPP.

Son varias las características que debe cumplir la estrategia propuesta. En primer lugar, se distinguen dos horizontes temporales diferentes: uno intradiario y otro diario, usando la hibridación cuando sea necesario unificar ventajas de varias técnicas. Así mismo, la información debe encontrarse lo más actualizada posible para garantizar una gran precisión, buscando una resolución temporal que se adapte a las necesidades de cada caso. Por último, la estrategia debe ser aplicable a instalaciones nuevas o que aún no disponen de la tecnología necesaria para la predicción, minimizando el error producido en comparación con modelos tradicionales o procedentes de la literatura.

1.3 Metodología de trabajo y alcance

El método de trabajo seguido para lograr los objetivos de la tesis consiste en dividir el problema general en problemas de menores dimensiones, resueltos de forma específica, integrándose conjuntamente. Se pueden asignar un total de cuatro fases, siendo, por orden de actuación: (i)

preparar el entorno de trabajo; (ii) identificar la necesidad; (iii) dividir el problema de investigación; y (iv) ensamblar las partes.

i. **Preparar el entorno de trabajo:** En primer lugar, se ha afrontado la disponibilidad de la información a la hora de desarrollar una estrategia, buscando un entorno experimental donde trabajar. Así mismo, también es de relevancia la demanda computacional necesaria para el test, al generar demoras innecesarias, siendo crítica en los primeros resultados o en situaciones de ensayo y error.
   Para obtener un entorno de trabajo realista y eficiente, se ha instalado en la Escuela Politécnica Superior de la Universidad de Alcalá un sistema fotovoltaico de $3\,kW_p$ de potencia, junto a una estación meteorológica, enviando las medidas recogidas a una nube. La estación meteorológica permite adquirir información de irradiancia, temperatura de célula, temperatura ambiente, velocidad y dirección del viento, presión atmosférica y humedad relativa. Los paneles fotovoltaicos dan alimentación a un grupo de cargas controlables, junto a un sistema de baterías y alimentación de red, que simulan un nodo de VPP con ayuda de un EMS, permitiendo diversificar la investigación dentro de este campo. Mediante este sistema y equipos de computación adecuados para la investigación se ha abordado este primer paso.

ii. **Identificar la necesidad:** Una vez dispuesto el entorno de trabajo, se han buscado trabas que impiden obtener predicciones de forma eficiente dentro de un entorno VPP. En este caso: (a) existe una escasez de variables relevantes para la predicción en un entorno de libre acceso, que impide generar predicciones en todo un territorio de forma eficiente y económica; (b) por otra parte, dichas predicciones no se pueden basar en un único modelo, al definirse diferentes horizontes y resoluciones temporales.

    a. Para obtener variables relevantes en un entorno de libre acceso, además de la estación meteorológica instalada, se han utilizado mapas de predicción numérica del tiempo (*Numerical Weather Prediction*, NWP) de temperatura ambiente y nubosidad. La base de datos del servicio de monitorización de atmósfera Copernicus (*Copernicus Atmosphere Monitoring Service*, CAMS) ha sido también de gran ayuda para aumentar la base de datos de medidas de irradiancia en casos requeridos.

    b. Las variables obtenidas en el apartado anterior han facilitado la predicción indirecta, consistente en predecir previamente la irradiancia para transformar dicha medida en potencia de los paneles mediante un modelo de la planta. Se han definido dos estrategias de predicción: una híbrida que predice de forma horaria la irradiancia al día siguiente; y otra basada en series temporales de irradiancia, que predice a nivel

intradiario, actualizándose a lo largo del día. Ambos horizontes temporales se unifican finalmente, para seleccionar en todo momento la estrategia más precisa en función del instante.

iii. **Dividir el problema de investigación:** Una vez encaminado el problema correctamente, el siguiente paso divide el mismo en varias etapas realizables. Se reconocen cinco partes independientes para esta fase: (a) lectura y procesado de mapas NWP; (b) predicción de la irradiancia al día siguiente; (c) predicción de la irradiancia intradiaria; (d) conversión de la irradiancia en potencia; y (e) generación de intervalos de predicción para cuantificar el grado de incertidumbre.

a. Los mapas NWP se han procesado al leer sus colores en RGB en cada instante de tiempo, desarrollando además un nuevo parámetro denominado factor de cobertura de nubes (*Cloud Cover Factor*, CCF), que indica la proporción de nubes que interfiere entre la estación y la posición del sol.

b. La predicción de horizonte temporal diario se ha realizado mediante una estrategia híbrida que tiene en cuenta el tipo de día. Este realiza una ponderación entre dos modelos: (i) uno denominado como modelo basado en horas similares, que genera una buena predicción bajo históricos escasos; y (ii) otro modelo basado en redes neuronales (*Artificial Neural Networks*, ANNs), con el que se complementa, optimizando la precisión de la estrategia conforme incrementa el histórico recopilado.

c. La predicción intradiaria se ha diseñado mediante una red neuronal recurrente de memoria a corto plazo (*Recurrent Neural Network – Long Short-Term Memory*, RNN-LSTM) y un histórico de irradiancia de varios años, al ampliarse con la base de datos del CAMS. Dicha predicción se actualiza periódicamente para minimizar el valor del error.

d. La conversión a potencia se ha realizado mediante parámetros de la planta conocidos. En primer lugar, se ha transformado la irradiancia de un plano horizontal al plano del panel, seguidamente, se ha determinado la irradiancia efectiva conociendo sus pérdidas angulares y espectrales, finalmente, la potencia se ha calculado mediante el modelo de Osterwald y el rendimiento de la instalación, considerando la presencia de sombras.

e. La generación de intervalos se ha construido en función del histórico de predicciones, dividiéndolo en varios grupos

independientes, y asumiendo una distribución Laplaciana del error.

iv. **Ensamblar las partes:** Una vez resueltas las diferentes partes de las que consta el problema, los resultados se han unificado, seleccionando en todo momento la predicción óptima. Esta predicción considera diferentes tiempos e intervalos de incertidumbre, siendo información de gran utilidad para el EMS de la VPP.

Mediante el cumplimiento de todas las partes, se ha conseguido un sistema eficiente de predicción fotovoltaica de gran utilidad en un entorno VPP e individual. Dicho sistema aporta información a futuro de la energía producida, conociendo su incertidumbre asociada, con información de libre acceso. Las limitaciones de este trabajo residen en su aplicación, ya que resulta inviable utilizar estas técnicas en zonas donde no se recoja información de libre acceso por parte de las agencias meteorológicas.

1.4 Estructura de la tesis

La tesis cuenta con los siguientes capítulos, divididos a su vez en apartados:

Capítulo 1: Introducción. Este capítulo presenta en el apartado 1.1 de forma general el contexto en el que se desarrolla la tesis y da una visión global en relación a las energías renovables. En el apartado 1.2 se detallan los objetivos que se persiguen a lo largo del documento. En el apartado 1.3 se describe el método de trabajo a usar, junto a sus limitaciones y logros alcanzados. El presente apartado 1.4 muestra la estructura global de la tesis. Finalmente, el apartado 1.5 presenta las publicaciones del autor y la relación que guardan con este documento.

Capítulo 2: Estado del arte. El estado del arte desarrollado en el apartado 2.1 muestra una revisión bibliográfica de los avances científicos relativos al ámbito las fuentes de carácter renovable en entornos VPP. Seguidamente, en el apartado 2.2, se revisan los modelos de predicción de potencia fotovoltaica más empleados y se ofrece un breve resumen de las categorías en la que se subdividen. Finalmente, el apartado 2.3 describe las conclusiones más destacadas y analiza las brechas existentes que han ayudado al desarrollo de la tesis.

Capítulo 3: Aspectos generales a considerar. En este capítulo se detallan conceptos importantes y se da una descripción general del entorno de trabajo y el modelo desarrollado. El apartado 3.1 describe el entorno real de pruebas a partir del cual se obtienen los resultados. El apartado 3.2 destaca ciertas descripciones de relevancia en el documento. Finalmente, el apartado 3.3 resume el modelo realizado mediante un esquemático para que su seguimiento en el documento sea más intuitivo.

Capítulo 4: Variables de entrada. Las variables de entrada a utilizar en la estrategia requieren de un apartado propio al tener que estudiarse en detalle. Primero, en el apartado 4.1, se describe la información disponible para el desarrollo de la estrategia. Seguidamente, en el apartado 4.2, se muestra cómo se ha obtenido la información de los mapas NWP de nubosidad y temperatura ambiente, para usarlos como entrada a la estrategia de predicción. El apartado 4.3 estudia las variables de mayor relevancia para la predicción. En el apartado 4.4 se calcula el error asociado a los mapas NWP. El apartado 4.5 explica cómo se realiza el sobremuestreo y se eliminan las faltas en el histórico. Finalmente, se especifican las conclusiones observadas en el apartado 4.6.

Capítulo 5: Predicción de horizonte temporal diario. En este capítulo se especifica la forma de obtener la predicción para un horizonte temporal de un día y una resolución horaria, mediante un modelo híbrido que se describe en el apartado 5.1. Concretamente, el subapartado 5.1.1 explica el primer modelo de predicción basado en horas similares, el subapartado 5.1.2 muestra el modelo de predicción basado en ANNs de perceptrón multicapa (*multilayer perceptron*, MLP), y el subapartado 5.1.3 especifica la forma de hibridación de ambos modelos. Los resultados del modelo se muestran en el apartado 5.2, dividiéndose entre los obtenidos para el nodo real de pruebas, en el subapartado 5.2.1, y los procedentes de una VPP emulada, en el subapartado 5.2.2. Finalmente, se describen las conclusiones destacadas en el apartado 5.3.

Capítulo 6: Predicción de horizonte temporal intradiario. En este capítulo se describe el modelo de predicción de irradiancia para un horizonte temporal intradiario y una resolución temporal de 15 minutos. El apartado 6.1 muestra el esquema general. Seguidamente, el apartado 6.2 explica la forma de diseñar la RNN-LSTM para generar la predicción, mostrando sus resultados. Finalmente, el apartado 6.3 enumera las conclusiones extraídas del capítulo.

Capítulo 7: Estimación de potencia. Una vez conocida la predicción de irradiancia, se calcula la potencia generada mediante un modelo de la planta. El apartado 7.1 desarrolla el método analítico utilizado para la conversión. El apartado 7.2 determina la precisión en el cálculo con datos reales de irradiancia y potencia en el nodo real de pruebas. Una vez conocido el error obtenido por el método, el apartado 7.3 genera los resultados en la predicción intradiaria, en el subapartado 7.3.1, y en la predicción al día siguiente, en el subapartado 7.3.2. Finalmente, se describen las conclusiones destacadas del capítulo en el apartado 7.4.

Capítulo 8: Intervalos de predicción. En este capítulo se determina la forma de calcular intervalos de incertidumbre en la predicción realizada. En el apartado 8.1 se muestran los relativos al horizonte temporal de un día, mientras que en el apartado 8.2 se muestran los obtenidos para un horizonte temporal intradiario. Finalmente, el apartado 8.3 especifica las conclusiones obtenidas en el capítulo.

Capítulo 9: Análisis de las predicciones. Una vez conocidas las predicciones bajo diferentes horizontes temporales, obteniendo intervalos de predicción que permitan evaluar la incertidumbre, el siguiente paso busca optimizar la predicción en todo momento, buscando minimizar el error producido. De esta forma, el apartado 9.1 selecciona en cada instante la estrategia de mayor precisión, evaluando finalmente los resultados obtenidos dentro de un caso de uso real de carga de baterías en el apartado 9.2. Las conclusiones del capítulo se enumeran en el apartado 9.3.

Capítulo 10: Discusión de resultados. En este capítulo se describen los resultados más relevantes obtenidos y se añaden unas conclusiones finales en el apartado 10.1. Finalmente, el apartado 10.2 analiza las contribuciones y se analizan las limitaciones y futuras líneas de investigación que complementan el desarrollo de la tesis.

Anexos. El documento cuenta con 4 anexos: el anexo I analiza la forma en la que se ha accedido al error en la predicción, el anexo II muestra las hojas de características de los equipos relevantes para la tesis en el entorno real de pruebas, el anexo III describe de forma general la radiación solar y el efecto de las nubes, finalmente, el anexo IV justifica el uso de las RNN-LSTM en el modelo.

1.5 Publicaciones del autor

Los artículos científicos de reconocido prestigio publicados por el autor, por orden cronológico, justificando su aportación en la presente tesis, son los siguientes:

i. *Martín, P., Moreno, G., Rodríguez, F. J., Jiménez, J. A., & Fernández, I. (2018). A hybrid approach to short-term load forecasting aimed at bad data detection in secondary substation monitoring equipment. Sensors, 18(11), 3947. (8 citaciones [9])*

En este primer artículo, el autor establece las bases necesarias para reconocer el modelo más adecuado de predicción en función del tipo de problema, estudia la relevancia a la hora de seleccionar las entradas al modelo, optimiza los parámetros de ajuste de cada modelo, obtiene resultados y formas de evaluar los mismos, y diseña un modelo híbrido que predice la carga a alimentar por varias subestaciones con el fin de corregir el error en la medida.

Así mismo, el autor construye bases en las publicaciones científicas, entre estas, expresarse en una lengua extranjera y enfrentarse a revisores expertos en la materia. Este artículo se considera de escasa relevancia con la tesis, de modo que su repercusión en el presente documento no va más allá de lo aquí expuesto.

ii. *Moreno, G., Martin, P., Santos, C., Rodríguez, F. J., & Santiso, E. (2020). A Day-Ahead Irradiance Forecasting Strategy for the Integration of*

Photovoltaic Systems in Virtual Power Plants. IEEE Access, 8, 204226-204240. (8 citaciones [10])

En este segundo artículo el autor se ha enfrentado a problemas iniciales, como reconocer variables relevantes para realizar la predicción fotovoltaica, investigar su accesibilidad de forma privada, estudiar la forma de adquisición de datos, aplicar un preprocesado de datos y generar nuevas variables relevantes a raíz de las obtenidas. Por otra parte, presenta un modelo híbrido para predecir la irradiancia recibida por los paneles con un horizonte temporal de un día y una resolución horaria, generando intervalos de incertidumbre para un mayor nivel de detalle.

Todo esto se realiza en un entorno real de pruebas y en una VPP emulada, siendo posible agregar nuevas instalaciones con poca información y minimizando el error en relación a otros modelos tradicionales de aplicación. Los capítulos 3, 4, 5 y 8 recogen información de este artículo para el desarrollo del documento.

iii. *Moreno, G., Santos, C., Martín, P., Rodríguez, F. J., Peña, R., & Vuksanovic, B. (2021). Intra-Day Solar Power Forecasting Strategy for Managing Virtual Power Plants. Sensors, 21(16), 5648. (2 citaciones [11])*

En este artículo el autor presenta una estrategia para obtener predicciones de irradiancia con un horizonte temporal intradiario y una resolución temporal de 15 minutos, actualizando la información de forma eficiente a lo largo del día y comparando los resultados con artículos relacionados en la literatura. Posteriormente, un modelo analítico permite convertir la irradiancia en potencia mediante información disponible en la propia instalación. Finalmente, al igual que en el artículo anterior, se obtienen intervalos de incertidumbre que aportan una mayor información a la estrategia desarrollada.

Este artículo también se realiza en un entorno real de pruebas y en una VPP emulada, obteniendo una precisión competente con artículos de la literatura relacionados. Los capítulos 3, 4, 6, 7 y 8 utilizan resultados procedentes de este artículo.

iv. Finalmente, para unificar los artículos publicados y concluir la tesis, el autor colabora en un artículo científico aún en revisión que permite incorporar al documento un caso de uso de baterías, utilizando las predicciones realizadas. El capítulo 9 de este documento muestra los resultados obtenidos en el artículo que guardan relación con el tema de la tesis.

2 Estado del arte

En este capítulo se revisa el estado del arte de los temas relacionados con la tesis. En concreto, las temáticas clave pueden ser diferenciadas en dos ramas: generación renovable en VPPs, que se analizan en el apartado 2.1, y técnicas de predicción fotovoltaica, que son tratadas en el apartado 2.2. Finalmente, el apartado 2.3 destaca las conclusiones más relevantes y especifica las contribuciones aportadas en la tesis.

2.1 Generación renovable en VPPs

La adopción de fuentes de carácter renovable como forma de generación de la energía se encuentra en un crecimiento mundial remarcable [12]. Dentro de estas fuentes, se destaca la tecnología fotovoltaica, siendo la base de la presente tesis. Las razones detrás de dicho crecimiento son: (i) el coste de la célula fotovoltaica se reduce anualmente, permitiendo amortizar los paneles solares en un tiempo cada vez más reducido [13]; (ii) el pico de generación fotovoltaica coincide con el momento de mayor demanda; (iii) la preocupación por el cambio climático es cada vez mayor a nivel mundial, jugando el sector de la electricidad un papel fundamental en esta tarea y, debido a ello, las instituciones aplican políticas que apoyan la investigación de recursos renovables [14]; (iv) los países desarrollados persiguen la búsqueda de la independencia energética, que tienen la generación renovable y el almacenamiento como pilares para lograr este propósito; y (v) el continuo progreso y la mejoría en la precisión de estrategias de predicción de potencia fotovoltaica.

En general, la predicción de la producción energética ha generado un gran desafío durante la última década. Este hecho toma mayor relevancia en fuentes no predecibles como la energía solar fotovoltaica, donde la generación de potencia fluctúa constantemente en función de diversos factores meteorológicos, entre ellos, la presencia de nubes, la temperatura ambiente, la humedad relativa o la velocidad del viento, que son estocásticos por naturaleza. Esta inherente incertidumbre ha impedido la integración a gran escala de fuentes de generación fotovoltaica [15,16]. Sin embargo, esta deficiencia puede ser resuelta mediante sistemas de almacenamiento que permiten una correcta gestión de la energía producida, aunque en contrapartida a su elevado coste asociado [17−19] Actualmente, existe una solución a esta problemática, mediante la agregación de varios sistemas fotovoltaicos en una VPP [2,20,21]. Este concepto permite a los prosumidores [5], involucrados tanto en la producción energética como en su consumo, maximizar sus oportunidades de beneficio al participar de forma activa en el mercado energético y formar parte en tareas de distribución de la energía u otros servicios que esta nueva tecnología puede aportar, como la regulación de voltaje y frecuencia [22−24].

El concepto de VPP fue concebido al inicio del presente siglo, descrito en [25], asociándose con la agregación de unidades de DG, cargas

controlables y equipos de almacenamiento, unidos a un clúster, dentro de una entidad virtual, que es la responsable de gestionar los flujos energéticos dentro del mismo e intercambiar energía con la red eléctrica central. El concepto tradicional de recurso energético distribuido (*Distributed Energy Resource*, DER) se diferencia de este enfoque al no ser su gestión concebida por el operador del sistema. Agrupando las unidades de DG de la VPP en una única entidad se permite el control y la monitorización, generando valor al mercado eléctrico [2]. Aunque las definiciones de VPP en la bibliografía varían, estas coinciden en disponer elementos de DG de diversa naturaleza con una gestión a nivel individual, trabajando en conjunto como una única planta eléctrica, y mejorando el funcionamiento del clúster [2,26−29]. Estrictamente, los autores de la literatura describen el concepto de VPP de forma diferente: "Una representación flexible de una cartera de DER que puede utilizarse para realizar contratos en el mercado mayorista y ofrecer servicios al operador del sistema" [2]. "Un sistema de comunicación e información, con control centralizado sobre una agregación de DGs, cargas controlables y dispositivos de almacenamiento" [26]. "Una agregación que incluye diferentes tecnologías DER, cargas de respuesta y dispositivos de almacenamiento que, cuando se integran, tienen una flexibilidad y controlabilidad similares a las grandes centrales eléctricas convencionales" [27]. "Un conjunto de unidades generadoras dispersas, cargas controlables y sistemas de almacenamiento, agregados para operar como una central eléctrica única. Los generadores pueden utilizar tanto fuentes de energía renovables (*Renewable Energy Sources*, RESs) como fósil. El corazón de una VPP es su EMS, que coordina los flujos de potencia procedentes de los generadores, las cargas controlables y el almacenamiento" [28]. "Una agregación de diferentes tipos de DERs que pueden dispersarse en diferentes puntos de las redes de distribución de media tensión" [29].

En la presente tesis se ha adoptado la siguiente definición: una VPP es un conjunto de elementos de DG, elementos de almacenamiento y cargas, que se apoyan en tecnologías de la información y comunicación para crear un único elemento virtual que se comporta idénticamente a una central eléctrica. Mediante la gestión, control y monitorización de los flujos energéticos en conjunto se minimizan los costes, se diversifica el mercado eléctrico, se maximizan los beneficios y se reduce la contaminación producida por recursos fósiles.

Una VPP dispone de tres elementos esenciales:

- Los DER que la contienen, que pueden diferenciarse según [2]:
    - Fuente de energía primaria: generadores que utilizan DER (paneles fotovoltaicos, aerogeneradores, pequeñas centrales hidroeléctricas, …) o energía fósil (cogeneración, biomasa, biogás, diésel, turbinas de gas, …).
    - Capacidad: de pequeña capacidad, conectándose a la VPP necesariamente para acceder al mercado, y de capacidad media o

gran escala, que pueden participar como VPP o individualmente en el mercado.

- Propiedad: pudiendo ser residencial, comercial e industrial (doméstica) o pública para apoyar la red. Si la empresa comercial pretende obtener beneficios de la venta se puede denominar productor independiente de energía.

- Naturaleza operativa [30]: de carácter estocástico, necesitando una forma de almacenamiento para operar fuera de esta naturaleza, o planificable, que compensa la falta de adaptación de la primera y se puede controlar.

La predicción fotovoltaica, núcleo de la presente tesis, forma parte de fuentes DER de pequeña capacidad y naturaleza estocástica. Conociendo en mayor detalle el comportamiento de estos recursos naturales se combaten los problemas derivados de su falta de control.

- Los sistemas de almacenamiento de energía (*Energy Storage System*, ESS) reducen la diferencia entre generación y demanda, siendo especialmente útiles en fuentes de carácter estocástico. Almacenando en periodos donde la oferta energética sea superior a la demanda puede ofrecer soporte en aquellos momentos que haya escasez de recursos [31].

- El EMS es el corazón del sistema de información y comunicación, gestionando el funcionamiento de los nodos de la VPP de forma bidireccional. Sus funciones son [32]:

  - Monitoriza el estado de cada elemento dentro de la VPP.

  - Realiza previsiones de fuentes DER, potencia de salida y cargas.

  - Coordina los flujos de potencia de los diferentes elementos.

  - Controla el funcionamiento de las unidades para minimizar costes, pérdidas energéticas y gases de efecto invernadero; mejorando la calidad energética, el perfil de tensión y el beneficio global.

La optimización de una VPP resulta esencial para mejorar su operación mientras que se minimizan los costes de producción. Esta optimización se puede enfocar en la estructura de sus componentes [33] o en la operación [34]:

- Según su estructura, la optimización depende del sistema eléctrico: si es nuevo este sistema, la VPP puede elegir la capacidad y ubicación de los nodos y los elementos que lo componen, además de las cargas, estrategias y programas de control; sin embargo, si ya existe, hay que atender a las limitaciones existentes en el entorno en el que nos encontramos.

- Para estos sistemas existentes, el funcionamiento óptimo podría enfocarse en su operación: determinando la generación de las unidades de DG, gestionando los ciclos de los ESS, estudiando la cantidad de energía a comprar en el mercado de la electricidad, entre otros.

2.2 Técnicas de predicción fotovoltaica

La predicción de potencia fotovoltaica, junto a la estimación de demanda o del precio de la energía, ofrece información de gran valor a nivel energético, especialmente cuando se mueven grandes cantidades de energía. Esta predicción debe ser considerada bajo diferentes resoluciones y horizontes temporales, incluso detallar su resolución espacial cuando se trabaja en un territorio amplio, como puede ser el entorno VPP. Los modelos de predicción fotovoltaica se pueden clasificar en función las siguientes características: la variable a predecir, las variables de entrada, el horizonte temporal, la resolución temporal, el modelo de predicción, la confianza de la predicción, entre otras.

- De acuerdo al parámetro a predecir, diferenciamos entre una estrategia directa [35], que predice la potencia fotovoltaica a través de históricos de generación producida y/o condiciones meteorológicas, o indirecta [36], que estima en primer lugar la irradiancia para calcular posteriormente la potencia fotovoltaica mediante un modelo de la instalación fotovoltaica.

- Tomando como criterio el horizonte temporal, la predicción se puede clasificar según cuatro grupos diferentes [37,38]: (i) predicción a muy corto plazo (*nowcasting*) que predice para horizontes temporales entre uno y varios minutos, usado por ejemplo para tareas de optimización del EMS a tiempo real; (ii) predicción a corto plazo, entre una y varias horas, para participar en el mercado eléctrico y tareas operativas de optimización de la instalación; (iii) predicción a medio plazo, entre varios días y varios meses; y (iv) predicción a largo plazo, que puede contemplar varios años.

- La resolución temporal depende del horizonte temporal y el propósito a lograr. Ejemplos de esta puede ser una resolución de un minuto durante la operación del mercado a tiempo real, cada 15 minutos para optimizar el sistema de almacenamiento energético de baterías (*Battery Energy Storage System*, BESS), de una hora para controlar el consumo producido, o incluso diaria y semanal para horizontes temporales mayores y permitir identificar tendencias [39]. Obviamente, la resolución temporal debe ser coherente al horizonte considerado, aumentando el rango temporal para mayores horizontes de tiempo. De esta forma se respeta la precisión del modelo y la toma de decisiones se realiza de manera consistente, por ejemplo, considerando rangos semanales al predecir la energía que se producirá al año siguiente.

- El modelo de predicción se puede categorizar dentro de cinco grupos [40]: (i) modelos estadísticos lineales basados en series temporales como el persistente o el ARMA; (ii) modelos estadísticos no lineales, entre ellos, basados en redes neuronales como las MLP-ANN o las RNN-LSTM; (iii) modelos indirectos basados en imágenes de nubes vía satélite o a nivel de suelo; (iv) estrategias de predicción físicas, también indirectas, a partir de mapas de NWP; y (v) modelos híbridos que combinan diferentes modelos con el fin de complementar estrategias, mejorar la precisión, optimizar los recursos, diversificar su utilidad, entre otros.

- De acuerdo a su incertidumbre, podemos diferenciar entre las predicciones deterministas o puntuales, que producen un único valor para cada instante de tiempo, y las predicciones probabilísticas, que informan sobre la confianza en un instante de tiempo en el que el valor se encuentra en un cierto rango definido por el modelo [41].

En la presente tesis se ha recurrido a una estrategia indirecta de predicción de irradiancia global horizontal (*Global Horizontal Irradiance*, GHI) para calcular la potencia fotovoltaica. Esto es debido a la correlación entre esta y las variables meteorológicas empleadas para la predicción, como la nubosidad o la temperatura, que son recogidas de los mapas NWP de la Agencia Estatal de Meteorología (*State Meteorological Agency*, AEMET), libres para acceder a través de su web [6]. Por otra parte, el CAMS dispone de un amplio histórico de irradiancia en su web [7], que permite incrementar la base de datos disponible, utilizada en el estudio para predicciones intradiarias. Así mismo, se usan estaciones meteorológicas de libre acceso repartidas por la Comunidad de Madrid donde se recopilan estas variables [42], lo que facilita generar resultados mediante una estrategia indirecta dentro de un entorno VPP.

Las técnicas de predicción empleadas se identifican dentro de dos modelos: (i) uno híbrido con un horizonte temporal de un día que usa una estrategia basada en horas similares junto a otra de ANN-MLPs; y (ii) una de predicciones intradiarias que utiliza un modelo de RNN-LSTMs. Por ello, es relevante destacar los avances bibliográficos de las estrategias que han sido de utilidad para el desarrollo de la tesis.

a) Modelos basados en horas similares

La predicción de horas similares es una de las aportaciones del documento y surge a raíz de la técnica de predicción basada en días similares. Esta técnica estima la GHI con un horizonte temporal diario a partir de similitudes entre predicciones meteorológicas de otras variables relacionadas con la GHI e históricos pasados de las mismas, tomando información selectiva de medidas de GHI para generar la predicción final. Existen varios artículos que utilizan este tipo de modelo de predicción de irradiancia, reconociendo la similitud entre predicción y medida a partir de la mínima distancia euclídea [43−48].

La predicción puede ser referida: (i) a la potencia fotovoltaica, obteniendo una predicción directa; o (ii) a la irradiancia, generando una predicción indirecta. La predicción directa es la más empleada para este tipo de técnicas. Esto es debido a que la potencia está altamente relacionada con la irradiancia y la temperatura, $P_{FV} = f(I, T)$, por lo tanto, mediante un histórico de predicciones de ambas variables se puede obtener un modelo bastante preciso [43−45,49,50]. Sin embargo, disponer de un histórico de NWP de irradiancia tiene un coste asociado que no se puede asumir sin una seguridad económica.

i. Modelos de predicción directa

Las predicciones directas del modelo basado en días similares muestran una bibliografía extensa. A continuación, se describen las más relevantes, usando expresiones encontradas en el anexo I para medir el error.

En [43] los autores emplean información de predicciones de libre acceso e históricos de temperaturas (máxima, mínima y media) para predecir la potencia fotovoltaica con un horizonte temporal diario por días similares, al minimizar la distancia euclídea de las temperaturas. Posteriormente, introducen esta predicción de potencia, las temperaturas del día a predecir y las del día similar escogido para ajustar el resultado en una ANN. Se realiza una distinción del histórico en función del tipo de día, obteniendo un error medio absoluto relativo (*relative Mean Absolute Error*, rMAE) horario desde un 10% para días soleados hasta un 19% en el caso de días nublados, mejorando otros modelos encontrados en la literatura.

En [49] se utilizan series temporales para predecir la potencia fotovoltaica al día siguiente a partir de medidas de GHI y temperatura (máxima, mínima y media). Por otra parte, en [51] se observa una reducción del error cometido bajo esta misma estrategia al reducir el horizonte temporal, lo que resulta coherente al contrastar ambos trabajos, ya que un modelo basado en históricos debe presentar una mayor precisión cuando se reduce su horizonte temporal, al ser menos propenso a cambios climáticos.

La técnica de días similares también puede ser usada para conocer la correlación entre la variable a predecir y las entradas a introducir en el modelo. Un ejemplo de evaluación de estas variables se observa en [44], con el fin de estimar mediante diferentes técnicas la potencia fotovoltaica disponible. El modelo propuesto justifica cuantitativamente la variable más destacada y la segunda variable de mayor relevancia, al comparar las distancias euclídeas entre la variable a predecir y los candidatos a predecir dicha variable. Aunque únicamente se seleccionan dos variables en el artículo, el proceso puede ser extrapolado para un número superior de variables, permitiendo identificar la correlación mediante métodos alternativos a otros tradicionales, como el coeficiente de correlación de Pearson (*Pearson Correlation Coefficient*, PCC).

Un aspecto adicional, realizado en el capítulo 8 de la tesis, es la predicción de intervalos asumiendo una distribución del error. Dichos resultados están basados en la técnica utilizada en [45], partiendo de una predicción de días similares con información de GHI, temperatura, humedad, nubosidad a diferentes alturas, entre otras. Finalmente se obtiene el intervalo mediante un histórico de días similares que forma la predicción, al usar una distribución Gaussiana o Laplaciana para este fin.

El artículo presentado en [50] emplea estadística basada en días similares para conocer la potencia al día siguiente a partir de valores de GHI y temperatura. El modelo supone una distribución Gaussiana de las variables, disponiendo adicionalmente de una predicción basada en la Teoría de Bayes para obtener finalmente una estimación probabilística. Los resultados confirman el buen comportamiento del modelo, llegando a una precisión de la raíz del error cuadrático medio (*Root Mean Square Error*, RMSE) horario de 284 kW, aproximadamente un 10% de la capacidad total, lo que genera grandes mejoras a artículos de la literatura relacionados.

ii. Modelos de predicción indirecta

Las predicciones indirectas de la potencia fotovoltaica no disponen de una bibliografía tan extensa en modelos basados en días similares debido a que no existe una correlación tan relevante como el caso directo, que encuentra una gran correlación con la irradiancia y la temperatura [52]. Sin embargo, si se procesan variables alternativas con gran correlación con la GHI [10], o al recopilar una gran cantidad de información [53], se pueden ofrecer resultados muy interesantes que permiten competir con las técnicas directas en este tipo de modelos.

En [53] se utiliza una gran cantidad de variables para generar una predicción de la GHI basada en series temporales de históricos recopilados. La predicción cuenta con 12 años de históricos y dividen las variables en tres grupos: históricos de irradiancia (global, directa y difusa), variables meteorológicas (temperatura, humedad, precipitación, turbulencia, cobertura de nubes, velocidad del viento, …) e imágenes de las nubes recogidas. Los autores realizan un preprocesado, eligiendo los tres días más similares, seguidamente, introducen todas las variables de entrada de estos tres días en una red neuronal profunda para obtener una GHI virtual, que es usada en un modelo de regresión para obtener la predicción final, determinista o probabilista, en función del horizonte temporal, también configurable. Para un horizonte temporal de 24 horas, el error absoluto medio (*Mean Absolute Error*, MAE) horario es de 89,81 W/m^2 en el modelo óptimo, similar al modelo basado en horas similares propuesto en la tesis.

En [46] la predicción de días similares se realiza para la siguiente medida a observar, dentro de una resolución temporal horaria. El modelo de predicción calcula las diferencias del histórico de GHI y determina el nuevo valor al añadir la diferencia esperada a la anterior medida. Los resultados

muestran un rMAE horario entre el 14% y el 24%, que muestra unos valores adecuados para ser un modelo de predicción innovador y de gran simpleza.

En [47], por otra parte, se realiza una predicción indirecta que usa el PCC para conocer las variables de mayor correlación con la GHI, destacando la temperatura y el viento. Este trabajo difiere del estudio realizado en la tesis, ya que no se han encontrado similitudes importantes entre la irradiancia y el viento, sin embargo, incita a volver a valorar las variables más relevantes al aplicar el modelo propuesto en otras zonas climáticas.

b) Modelos basados en redes neuronales

La existencia de diversos modelos de predicción genera la duda sobre cuál de ellos ofrece la mayor precisión, sobre todo en modelos estadísticos. Algunos estudios concluyen que los modelos estadísticos tradicionales mejoran a los de aprendizaje automático (*Machine Learning*, ML) [54], mientras que otros afirman lo contrario [55,56]. Sin embargo, esta interpretación no resulta adecuada si no se toma en cuenta el tamaño del histórico [57], la variable a predecir [58], el horizonte temporal [59] o la carga computacional [60]. Aunque los modelos estadísticos tradicionales han dominado en la última década, existe actualmente un cambio de rumbo hacia las estrategias de ML [41].

Con el fin de conocer el potencial en un grupo centralizado de nodos, en [61] realizan una predicción de la potencia fotovoltaica de 200 instalaciones en Alemania mediante mapas de irradiancia, mostrando que el error se reduce de $RMSE = 0.11\ kW/kW_p$, para una única instalación, a $RMSE = 0.06\ kW/kW_p$, para un área de $220\ km\ x\ 220\ km$ con múltiples instalaciones. Además, se obtiene como conclusión que el error se reduce en mayor medida conforme la distancia entre localizaciones aumenta. Esta estrategia ofrece una gran solución en el contexto VPP, ya que esta cuenta con múltiples nodos distribuidos a lo largo de un área, siendo la energía fotovoltaica la más sencilla y económica de explotar de las renovables, principalmente en viviendas, que pueden ser transformadas en cargas flexibles [62].

Las redes neuronales usadas en la tesis se dividen en dos grupos: las ANN-MLP, que estiman la GHI horaria para un horizonte temporal de un día, y las RNN-LSTM, que predice la GHI cada 15 minutos con un horizonte temporal intradiario. Estas se han seleccionado en función de la complejidad del modelo, la evaluación de patrones en el histórico, la disponibilidad de los datos, la carga computacional, las variables de entrada disponibles, entre otros.

A nivel general, las redes neuronales se caracterizan por ser las herramientas más usadas en predicción fotovoltaica [63,64]. De estas, las redes neuronales profundas ganan relevancia en la actualidad con la presencia de equipos computacionales más veloces, especialmente en secuencias de datos y series temporales, en los que se requieren técnicas

muy potentes. En este campo, las unidades recurrentes cerradas (*Gated Reunited Units*, GRUs), las RNN-LSTMs [65] y las redes neuronales convolucionales (*Convolutional Neural Networks*, CNNs)–LSTM son las técnicas más empleadas. En [66] los autores mostraron que estas tres técnicas generan mayores precisiones que las redes neuronales convencionales y las máquinas de soporte vectorial (*Support Vector Machines*, SVMs) en predicción de GHI a corto plazo. En [67] se propone una estrategia de predicción de potencia fotovoltaica basada en RNN-LSTMs, comparada con otras estrategias de predicción. Se deduce que las estrategias alternativas no son capaces de modelar la potencia fotovoltaica de la misma forma que la estrategia propuesta en el artículo.

Las RNN-LSTMs son las técnicas de aprendizaje recurrente más empleadas de todas [60]. Existe una cantidad extensa de artículos en la literatura que usan las RNN-LSTMs para predecir GHI, [60,64,66−78]. En [69] se concluye que aumentar notablemente el histórico de la variable a predecir genera una precisión en el modelo muy superior, siendo uno de los motivos por el que se utilizan las RNN-LSTMs con un histórico de 5 años para horizontes temporales intradiarios en la tesis.

En [68], los autores desarrollaron una RNN-LSTM para predecir de forma endógena la GHI con datos satélite, emulando la generación de varias plantas fotovoltaicas, con una resolución temporal de 3 horas. El artículo destaca que las RNN-LSTMs superan a los modelos alternativos, siendo la técnica más utilizada para este estudio. Aunque el uso de entradas exógenas aporta una gran precisión [69−73], las entradas endógenas son más fáciles de obtener, de modo que son más utilizadas en predicción intradiaria de irradiancia [60]. En [74] se usa la parte de LSTM para extraer características temporales y la CNN para características espaciales de la instalación de forma simultánea, obteniendo un modelo que mejora al convencional. Sin embargo, en la tesis, este tipo de redes supone una tecnología muy pesada que no mejora a su versión simple, ya que el caso de estudio solo cuenta con variables endógenas en la predicción [75].

En resumidas cuentas, las predicciones de irradiancia más simples requieren RNN-LSTMs [76], mientras que las CNN-LSTMs son necesarias cuando se aumenta el número de variables y la complejidad del modelo [77]. La resolución temporal seleccionada en el estudio es de 15 minutos, ya que es muy conveniente para las tareas de gestión, además, tiene la ventaja de reducir el error respecto a resoluciones temporales superiores [78].

c) Modelos de conversión a potencia

Las predicciones indirectas de potencia fotovoltaica necesitan modelos de la instalación para realizar la conversión final. Para ello, se usa una estrategia analítica que funciona adecuadamente ante valores de irradiancia y temperatura. El modelo de Osterwald [52] es el modelo tradicional más utilizado, que destaca además por su simplicidad, sin embargo, en la literatura se pueden encontrar estudios basados en este primero que buscan

mejorar su precisión bajo ciertas condiciones, por ejemplo, a bajas irradiancias [79,80]. Por otro lado, cuando el punto de operación de los paneles fotovoltaicos se puede conocer, existen modelos alternativos que tratan de mejorar la precisión del modelo [81,82], y existen otros estudios que utilizan la parametrización para simplificar el proceso [83].

Los modelos de conversión utilizan la irradiancia en el plano del panel, sin embargo, a veces la irradiancia de la instalación se mide en su plano horizontal, obteniendo la GHI. Principalmente, las medidas de GHI se recogen vía satélite, pero también pueden ser de utilidad en instalaciones con múltiples seguimientos del punto de máxima potencia (*Maximum Power Point Trackings*, MPPTs) o paneles con seguidores solares. Para resolver este inconveniente, se requiere un proceso que mida la irradiancia en el plano del panel: (i) mediante expresiones analíticas que separen la irradiancia en sus componentes principales, modificando su ángulo posteriormente; o (ii) mediante un modelo experimental que simplifique el proceso [84].

En la tesis se busca reducir la complejidad del modelo de conversión a potencia y evitar demandas computacionales excesivas, manteniendo la precisión del modelo. Por ello, se utiliza el modelo de Osterwald, ya que los valores de baja irradiancia son de poca relevancia y prácticamente inexistentes, simplificando así el proceso del entorno VPP. Al usar en la tesis información vía satélite de GHI, el sensor de la instalación se sitúa horizontalmente para realizar posteriormente una conversión en el plano del panel de forma analítica.

2.3 Conclusiones

En este capítulo se ha revisado el estado del arte de los temas tratados en la tesis, en primer lugar, las tecnologías de naturaleza renovable y su uso en un entorno de VPP, y seguidamente las técnicas de predicción fotovoltaica más relevantes utilizadas. Las decisiones tomadas han sido razonadas para el caso de estudio, de modo que el proceso desarrollado queda debidamente optimizado y justificado.

Aunque la definición de una VPP no forme parte esencial del trabajo, sí se trata de un concepto nuevo que es necesario resaltar, ya que la generación fotovoltaica forma un papel clave en este aspecto y, por lo tanto, la predicción eficiente de la misma bajo diferentes horizontes y resoluciones temporales. Las técnicas de predicción fotovoltaica también han sido definidas y desglosadas en función de diversos aspectos de importancia como el horizonte temporal, las variables de entrada, o el modelo de predicción.

La revisión del estado del arte se ha enfocado en los modelos de predicción indirecta utilizados en la tesis. En primer lugar, el modelo basado en horas similares, que se inspira en el modelo basado en días similares, y seguidamente las técnicas de ML, centrando el tema en el uso de redes neuronales.

Se usa el modelo basado en horas similares ya que el modelo tradicional basado en días similares utiliza rangos temporales de varias horas, que pueden relacionar instantes temporales con diferentes situaciones climáticas, mientras que el modelo propuesto tiene en cuenta la presencia de nubes, al definir un nuevo parámetro denominado CCF. Además, mediante el uso de la radiación extraterrestre como medida del tiempo, se permite seleccionar un mayor número de horas de relevancia para el modelo. La definición detallada del modelo basado en horas similares se aborda en el apartado 5.1.1.

Con respecto a las redes neuronales, se justifica en primer lugar su uso como modelo complementario del modelo basado en horas similares, y por otro lado se selecciona una RNN-LSTM para horizontes temporales intradiarios. En el campo de las redes neuronales no se busca generar innovación, si no seleccionar el modelo más apropiado en la literatura a nivel de precisión, simplicidad y carga computacional en función de las condiciones del problema.

Además de obtener los resultados dentro de un modelo experimental real, la tesis aborda otras contribuciones que caben destacar: (i) la implementación del modelo de predicción en un entorno VPP; (ii) el uso de variables meteorológicas de libre acceso que son publicadas regularmente en la web, permitiendo obtener predicciones de forma gratuita; (iii) el modelo basado en horas similares, que ofrece una gran precisión ante históricos escasos, encontrados en nuevas instalaciones o nuevos nodos VPP; (iv) la hibridación del modelo basado en horas similares con un modelo basado en redes neuronales que permite generar una elevada precisión de forma independiente al histórico, seleccionando el tipo de día de acuerdo a un nuevo parámetro llamado CCF; y (v) el uso de intervalos de predicción para definir la incertidumbre, seleccionando los grupos óptimamente en función del propio CCF y otras variables temporales consideradas.

3 Aspectos generales a considerar

Previo a la descripción detallada de cada uno de los pasos de la estrategia, es necesario explicar el contexto, definir ciertas variables temporales de relevancia y mostrar un esquema general. Para ello, el apartado 3.1 define el entorno de pruebas, el apartado 3.2 define ciertas expresiones usadas que facilitan la lectura, y el apartado 3.3 describe la estrategia brevemente para ofrecer una visión general del documento.

3.1 Contexto del entorno de trabajo

La presente tesis se ha desarrollado dentro del proyecto Helios, gracias a la financiación del Ministerio de Ciencia, Innovación y Universidades. En el mismo se ha diseñado y desplegado un prototipo de VPP a partir de la red de estaciones base para comunicaciones móviles (*Base Transceiver Stations*, BTSs). Este proyecto pone a disposición de la red eléctrica la capacidad de almacenamiento distribuido ya instalado, de forma coordinada, reconvirtiendo las BTSs en unidades DER con un bus de continua y generación fotovoltaica. Así, se consigue alimentar las cargas, recargar las baterías y/o inyectar energía en la red, optimizando la eficiencia energética, la rentabilidad económica y dando soporte a la red eléctrica.

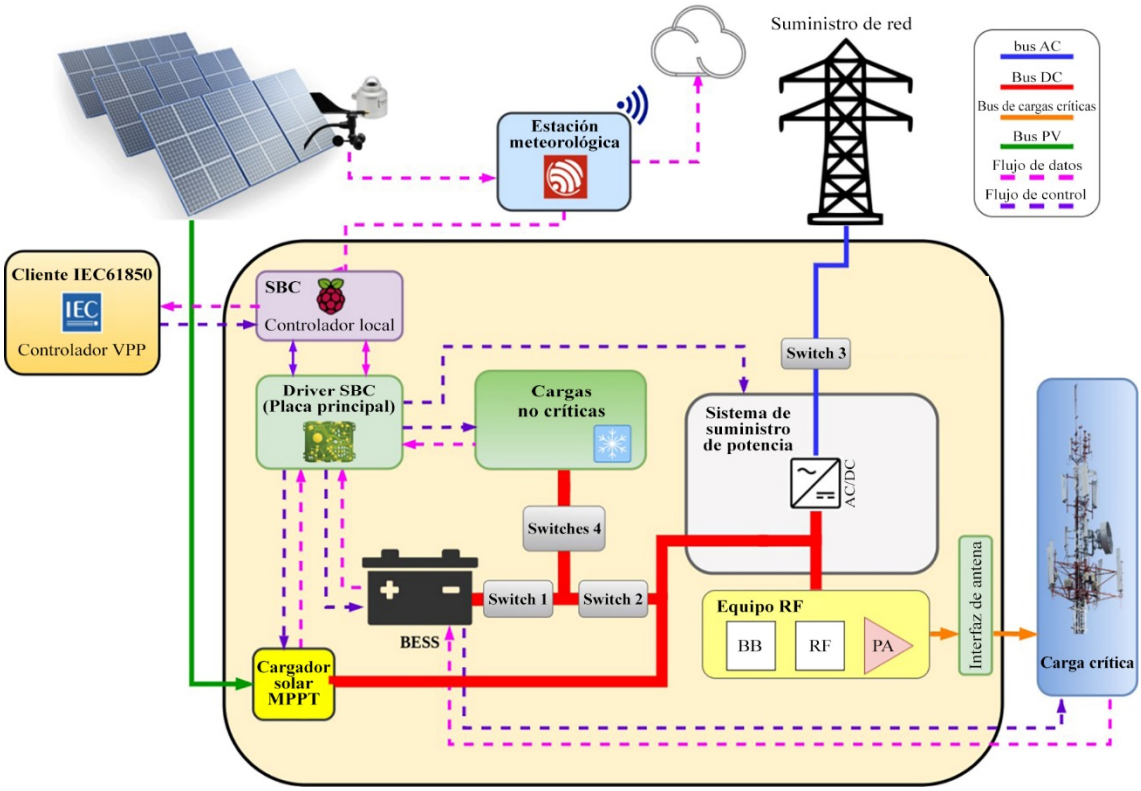


Figura 1.- Esquema eléctrico y de comunicaciones de una BTS transformada en un nodo de VPP.

El proyecto Helios ha permitido el diseño de un nodo real de VPP, emulando una BTS [8] y dando la posibilidad de llevar a cabo las predicciones tratadas en el presente documento. En la Figura 1 se muestra un esquema del diseño realizado, haciendo uso de la instalación fotovoltaica y la estación meteorológica para llevar a cabo las predicciones. En la Figura 2 se muestra la posición cenital de la instalación fotovoltaica, instalada en la Escuela Politécnica Superior de la Universidad de Alcalá, junto al agrupamiento de los paneles. Dispone de un total de 9 módulos solares policristalinos de 330 W_p con 72 células en serie. El grupo de paneles tienen una potencia pico de 2.97 kW_p, según tres paralelos de tres paneles en serie que alimentan un regulador MPPT en continua a 48 V de tensión nominal. En el anexo II se muestran las hojas de características de los equipos detallados.

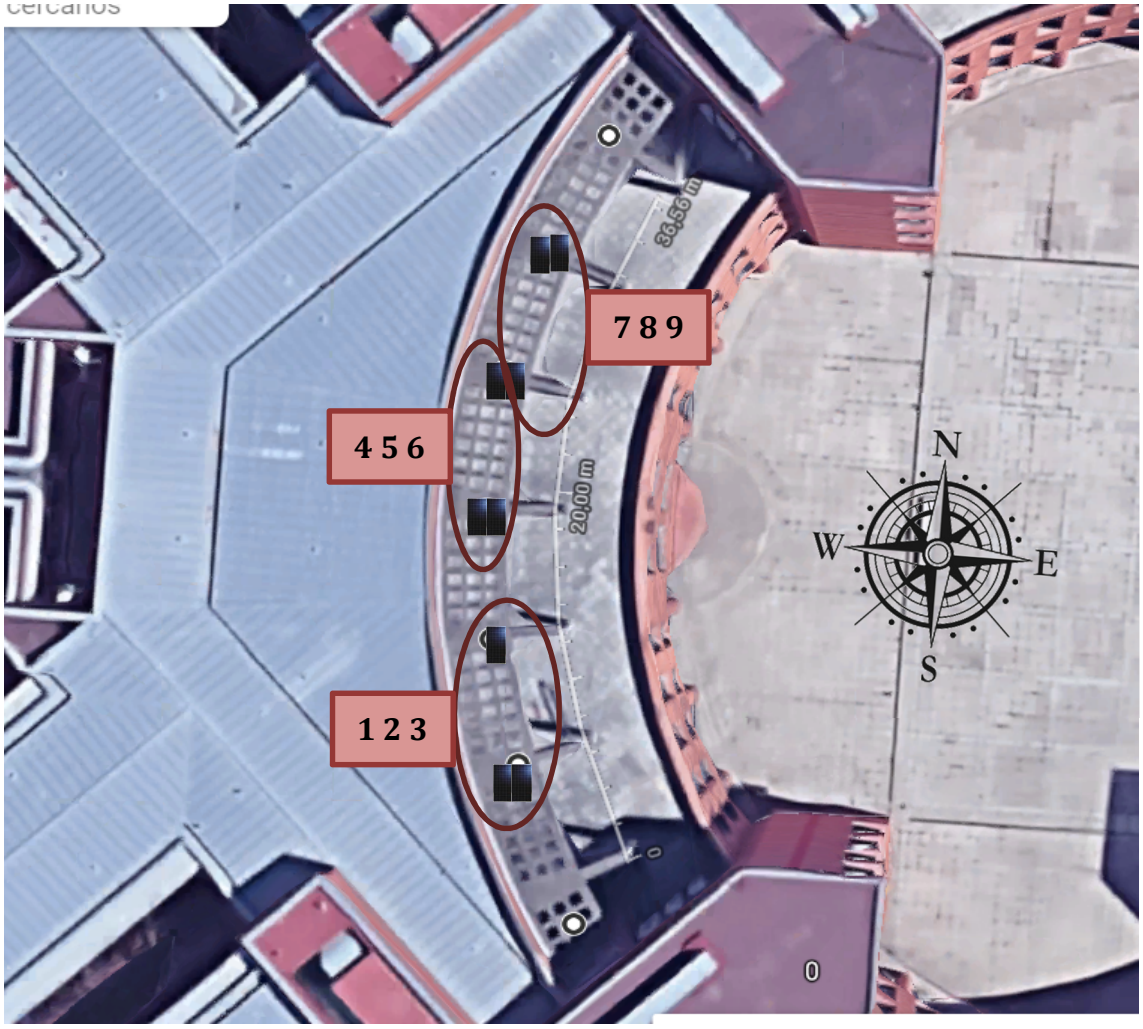
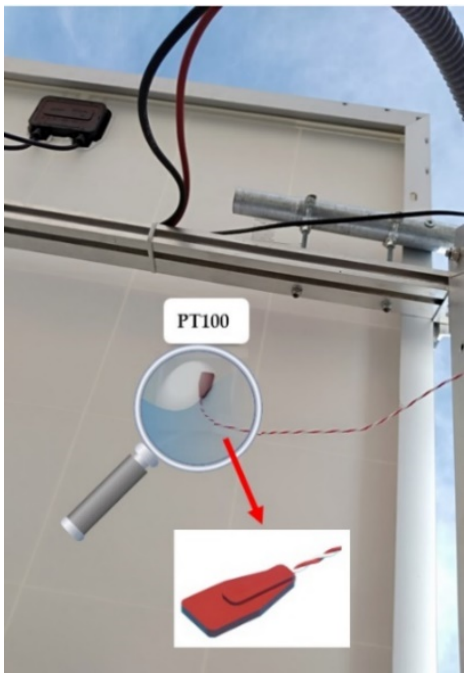


Figura 2.- Disposición de los paneles solares fotovoltaicos instalados en la Universidad de Alcalá, junto a su numeración y conexionado.

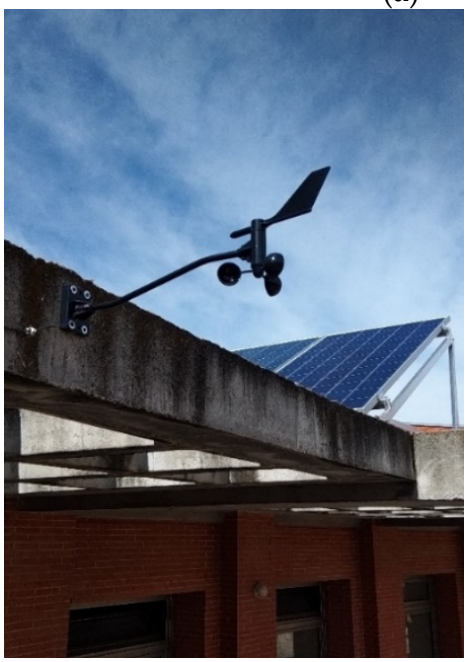
Por otra parte, la estación meteorológica inalámbrica se ha instalado cerca de los paneles solares para recolectar información meteorológica relevante, que se utiliza para la estrategia de predicción de la potencia fotovoltaica. Dichos elementos están formados por dos módulos, mostrados en la Figura 3: (i) un módulo de sensado para medir variables meteorológicas específicas de los paneles fotovoltaicos, Figura 3a a Figura 3c; y (ii) la electrónica de adquisición para obtener otras variables meteorológicas a nivel de suelo como la temperatura ambiente o la humedad, Figura 3d.



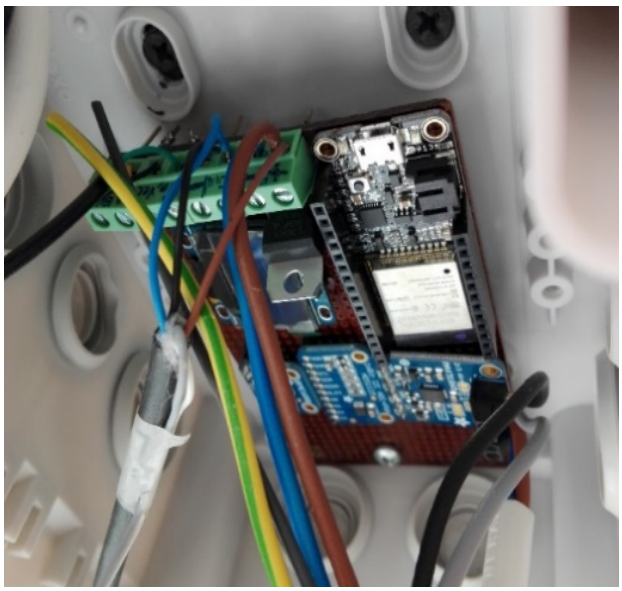
(a)



(b)



(c)



(d)

Figura 3.- Estación meteorológica inalámbrica: (a) piranómetro; (b) PT100 fijado al módulo; (c) anemómetro; y (d) electrónica de adquisición.

Todas estas variables se almacenan en Thingspeak [85] para que sean procesadas en tiempo real por la estrategia de predicción, permitiendo generar predicciones a lo largo del día para optimizar la precisión de la estrategia. El equipo que se ha usado para realizar las predicciones es un Intel Core i7-9750H 2.6GHz, con el software Matlab 2020a, suficiente para generar las predicciones dentro de un tiempo adecuado para el problema.

Además de la instalación descrita, con el fin de emular una VPP, es necesario tener como referencia un mayor número de nodos que ofrezcan información conjunta del sistema. Para solucionar este inconveniente, se recoge información de varias estaciones meteorológicas a nivel de suelo situadas a lo largo de la Comunidad de Madrid [42], que permite conocer las variables meteorológicas de interés para predecir desde otras zonas.

Por otra parte, considerando que estas estaciones meteorológicas no disponen de medidas de potencia fotovoltaica con las que contrastar los resultados, se toma el nodo real de pruebas situado en la Universidad de Alcalá como referencia. De esta forma se considera que todos los nodos utilizados para el entorno VPP tienen instalaciones fotovoltaicas con las mismas características a la de la universidad. Esto es posible al generar un modelo analítico de conversión a potencia fotovoltaica, que es función de la irradiancia y la temperatura ambiente medida en las diferentes zonas.

3.2 Definiciones temporales

Este apartado describe algunas definiciones de relevancia a la hora de analizar los resultados que se presentan en la tesis, en relación a su variable temporal. La medida del tiempo se puede referenciar de forma global definiendo así un instante de tiempo, p.ej. las 3 horas del tiempo universal coordinado (*Universal Time Coordinated*, UTC), o en función de una posición relativa trazando una extensión temporal, p.ej. 3 horas después del amanecer. A la hora de definir un instante, se hace uso de la hora solar, en lugar de usar la hora específica de un territorio, ya que puede dar lugar a confusiones por situación geográfica o por cambios de hora entre estaciones. Por otra parte, las extensiones temporales permiten analizar los resultados del documento, usando las siguientes definiciones:

- **Horizonte temporal (*time horizon*).** Es el periodo de tiempo total que comprende la predicción realizada. Por lo general, este horizonte de predicción se discretiza de acuerdo a la resolución temporal fijada por el diseñador del modelo. El horizonte temporal comprende toda la zona de interés a predecir. Específicamente, en la tesis se distingue entre un horizonte temporal diario, que finaliza al anochecer del día siguiente a la predicción, y otro intradiario asociado al anochecer de ese mismo día.

- **Resolución temporal (*time resolution*).** Caracteriza la separación temporal entre los instantes de tiempo concretos calculados dentro del horizonte temporal. Este parámetro representa una solución de

compromiso entre el número de instantes de predicción a calcular y la carga computacional y de memoria. Este parámetro suele ser constante para que sea más fácil caracterizar la evolución de la predicción. En este caso se utiliza una resolución temporal de una hora para el horizonte temporal diario y una resolución de 15 minutos para el intradiario.

- **Periodo de actualización (*update time*).** Indica el tiempo que trascurre desde que se realiza una predicción hasta que se realiza una nueva predicción. Cuanto más pequeño es este periodo, más reciente es la información que utiliza la predicción, pero mayor es el número de operaciones computacionales a realizar. Este parámetro se suele fijar en función de la frecuencia de adquisición de la información de relevancia. El periodo de actualización es de un día en el horizonte temporal diario, ya que esta predicción no se actualiza a lo largo del día, y de 15 minutos para el horizonte temporal intradiario.

- **Tiempo de espera (*lead time*).** Es la diferencia entre el instante en el que se lanza una predicción y el instante de tiempo a predecir, para un único valor de estudio. No se debe confundir con el horizonte temporal, que contempla la extensión máxima observable en la predicción. El tiempo de espera es variable, en función del dato que se quiere conocer y el instante en el que la predicción fue lanzada. Por ejemplo, si la predicción fue lanzada a las 11 horas y el dato a conocer se producirá a las 15 horas, el tiempo de espera asociado será de 4 horas.

- **Plazo intradiario de lanzamiento (*launch time*).** Es la diferencia entre el amanecer de un día concreto y el instante de ese mismo día en el que se lanza una predicción. Este parámetro es clave en las predicciones intradiarias, ya que indica el tiempo que se lleva recopilando medidas de irradiancia ese día. El plazo intradiario de lanzamiento es variable, en función del instante en el que la predicción fue lanzada y el amanecer. Según el ejemplo anterior, para el caso intradiario, si la predicción fue lanzada a las 11 horas y amaneció ese día a las 7 horas y media, el plazo intradiario de lanzamiento es de 3 horas y media.

En la Figura 4 se aprecia cada una de estas definiciones de forma gráfica, señalando el periodo que se asocia con cada concepto. La tesis utiliza las definiciones propuestas estrictamente, para reconocer fácilmente el concepto detrás de la terminología usada y simplificar su lectura.

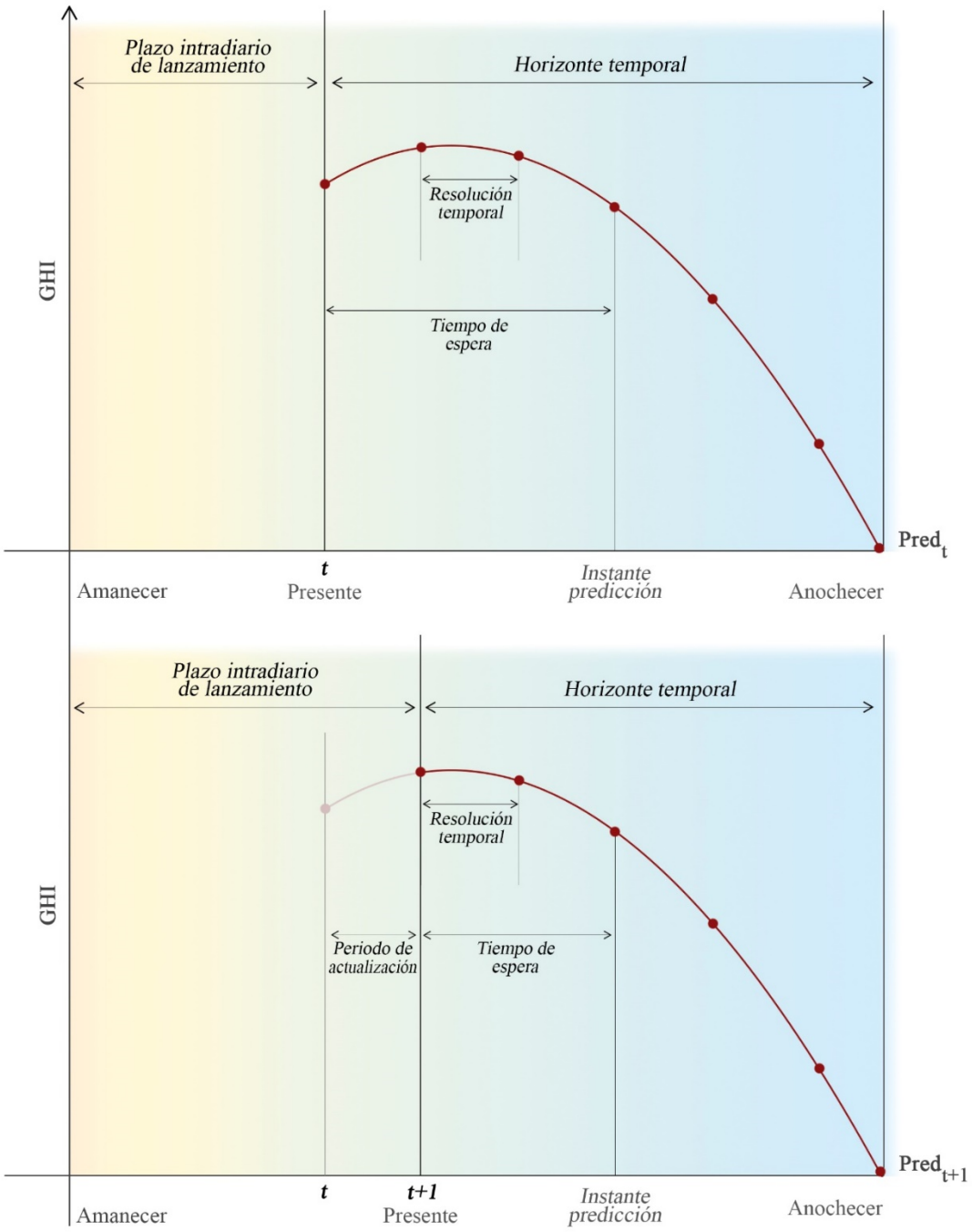


Figura 4.- Variables temporales utilizadas para definir las predicciones realizadas en la tesis.

3.3 Descripción global del modelo

El método de predicción se basa en el desarrollo de distintas estrategias que se complementan entre sí. Para facilitar la comprensión de las relaciones que existen entre ellos, en este capítulo se hace una descripción general de la estrategia de predicción mostrada en la Figura 5, distinguiendo los siguientes bloques:

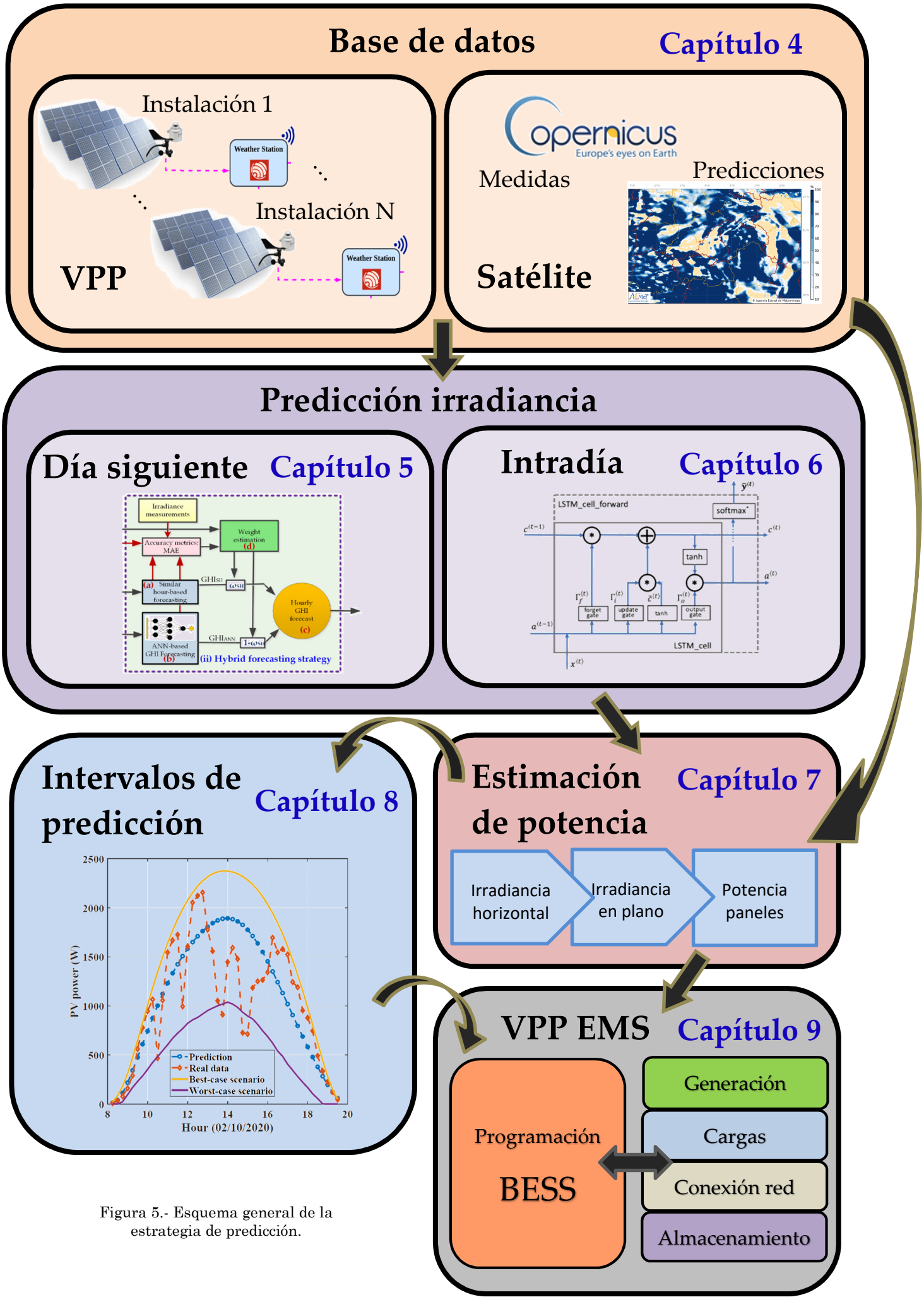


Figura 5.- Esquema general de la estrategia de predicción.

- **Base de datos.** La primera parte, analizada en el capítulo 4, describe la recopilación de variables, agrupadas en una base de datos. Aquí se exponen las variables disponibles, a usar por la estrategia de predicción de potencia fotovoltaica. Entre estas se distinguen dos agrupaciones:

    - Las variables recogidas en la instalación fotovoltaica. Entre estas variables hay que diferenciar: (i) las proporcionadas por la propia planta fotovoltaica, que da información de potencia de paneles, características de los equipos, pérdidas eléctricas asociadas al sistema, cargas a alimentar, suciedad y sombreado; y (ii) las que aporta la estación meteorológica de la instalación, que recoge información de GHI, temperatura ambiente, temperatura de paneles, velocidad y dirección del viento, presión y humedad relativa.

    - Información de libre acceso recogida vía satélite, a distinguir entre: (i) el histórico de medidas de irradiancia proporcionado por el CAMS; y (ii) los mapas de NWP de la AEMET, que muestra predicciones de nubosidad, temperatura ambiente, velocidad y dirección del viento, presión, precipitación, descargas eléctricas y rachas máximas.

- **Predicción de irradiancia.** Mediante la recopilación de variables en una base de datos se realiza la predicción de GHI, diferenciando dos modelos:

    - El primero de ellos genera una predicción de GHI con resolución temporal horaria para un horizonte temporal de predicción diario, analizado en el capítulo 5. Esto se realiza mediante un modelo híbrido que unifica un modelo basado en ANNs y otro modelo basado en horas similares. Asignando pesos a ambas predicciones, y distinguiendo entre tipos de días, se consigue hibridar ambos modelos, de forma que la predicción basada en horas similares predice óptimamente cuando el histórico disponible es escaso, mientras que el modelo basado en ANNs va incrementando su precisión conforme el histórico de la instalación aumenta. Este modelo emplea los NWP de la AEMET como variables de entrada para proporcionar información del sombreado de los paneles y la temperatura ambiente en el momento de la predicción.

    - El segundo modelo predice la GHI con un horizonte temporal intradiario y una resolución temporal de 15 minutos, descrito en el capítulo 6. Este modelo utiliza como entrada la base de datos de GHI de la propia instalación, ampliando esta medida con el histórico recogido desde la web del CAMS hasta disponer de un total de 5 años de información. Para ello se emplea una técnica basada en RNN-LSTMs que se va actualizando a lo largo del día.

- **Estimación de potencia.** El siguiente bloque, analizado en el capítulo 7, describe un método analítico que estima la potencia fotovoltaica, mediante la GHI predicha y un modelo de la instalación. El proceso de estimación de potencia puede diferenciarse en varios pasos, transformando la GHI medida al plano del panel y estimando después la potencia fotovoltaica generada:

    - En primer lugar, se calcula la irradiancia difusa y directa horizontal a partir de la correlación encontrada en [86], utilizando la predicción de GHI y el índice de claridad.

    - Posteriormente se determinan las tres componentes en el plano del panel: la directa se obtiene por simple geometría, el albedo usa un modelo isotrópico y la difusa emplea el modelo de [87] que la separa en una componente isotrópica y otra debida a la radiación circunsolar.

    - Conocidas las componentes en el plano del panel se determina la irradiancia efectiva, incluyendo pérdidas angulares y espectrales. El factor de corrección angular afirma que la reflexión aumenta con el ángulo de incidencia sobre el panel y se ha determinado empleando el método de [88]. El factor de corrección espectral dice que la fotocorriente cambia con el espectro de la luz solar para la misma irradiancia y se ha determinado empleando el método de [89].

    - Para calcular la potencia del panel resulta necesario estimar su temperatura, utilizando para ello la irradiancia efectiva y la temperatura ambiente obtenida de los mapas de NWP de la AEMET.

    - Finalmente, se calcula la potencia de los paneles mediante el modelo de Osterwald [52]. Este modelo utiliza las predicciones de la irradiancia efectiva y la temperatura de paneles, junto con información de la propia instalación: potencia pico, pérdidas eléctricas, coeficiente de temperatura y sombreado.

- **Intervalos de predicción.** En el capítulo 8, al conocer un histórico de predicciones de potencia y las medidas de la instalación, se pueden generar intervalos de predicción al asumir una distribución en el error [45]. En este caso, asumiendo una distribución Laplaciana, y distinguiendo entre tipos de días, se obtiene un rango de valores entre los que se estima que se encuentra la medida de potencia, bajo un nivel de confianza específico.

- **EMS de la VPP.** Finalmente, en el capítulo 9, se optimiza la estrategia de predicción para ofrecer en todo momento una alta precisión, en función del plazo intradiario de lanzamiento y el tiempo de espera.

Posteriormente, se emplea la información recopilada para planificar el uso de un sistema de gestión energético de baterías (*Battery Energy Management System*, BEMS) en el nodo real de pruebas, mejorando el aprovechamiento de la fuente de generación fotovoltaica. De esta forma, se consigue optimizar el beneficio económico del BESS, amortizando el uso de este a lo largo de su vida útil.

Todos estos pasos aquí enumerados se describen en los próximos capítulos de forma detallada, justificando el uso de variables y técnicas, analizando los resultados comparados con la literatura y obteniendo conclusiones relevantes de los mismos.

4 Variables de entrada

Previo a los modelos de predicción, resulta necesario realizar un estudio de las variables más relevantes, así como analizar la disponibilidad de las mismas dentro del presente caso de estudio. Por ello, en este capítulo, se evalúan las variables a utilizar para la predicción de la irradiancia. Así, el apartado 4.1 realiza una introducción inicial de la información disponible. El apartado 4.2 describe la forma de obtener los parámetros procedentes de mapas NWP. El apartado 4.3 realiza una evaluación de las variables más relevantes. El apartado 4.4 calcula el error cometido al utilizar información procedente de mapas NWP como entradas al modelo. El apartado 4.5 explica la fase de preprocesado previa a la predicción de irradiancia. Y, finalmente, el apartado 4.6 describe las conclusiones más destacadas a las que se han llegado.

4.1 Información disponible en el modelo

Tradicionalmente, un conocimiento exigente del modelo requiere de medidas a diferentes alturas de nubosidad, turbulencia, ozono, presión, albedo, luminosidad, temperatura, dirección y velocidad del viento, entre otras. Este tipo de predicciones emplean las ecuaciones de Navier-Stokes para resolver la irradiancia dentro de un mallado seleccionado [90]. Sin embargo, la predicción a partir de estos modelos implica un elevado coste que no es posible cubrir, considerando en la tesis una estrategia alternativa con la información que se describe a continuación. Por otra parte, no resulta impensable que técnicas más actuales, como las usadas en aprendizaje profundo, superen en precisión a modelos tradicionales, especialmente en instalaciones particulares y emplazamientos de pequeña escala [91].

Las variables utilizadas en la estrategia desarrollada en la tesis se pueden dividir en dos grupos en función de la procedencia de los datos: (i) datos privados adquiridos en las propias instalaciones fotovoltaicas; y (ii) información pública tomada de fuentes de libre acceso, tanto vía satélite como a nivel de suelo.

i. La información privada recogida en la instalación procede de las hojas de características de los equipos, condiciones físicas que afectan al rendimiento como el sombreado o las pérdidas por cableado, medidas de tensión y corriente y variables meteorológicas recopiladas. Por otra parte, para estimar la potencia fotovoltaica se usa información de los equipos, sombras en la instalación y pérdidas producidas. Además, una estación meteorológica situada en el emplazamiento aporta información de irradiancia, temperatura ambiente, temperatura de panel, presión, humedad relativa y velocidad y dirección del viento.

ii. La información pública recogida vía satélite procede de mapas NWP proporcionados por la AEMET [6], que predice la nubosidad, temperatura ambiente, presión, viento, precipitación, descargas eléctricas y rachas máximas en todo el territorio español; e información procedente del CAMS [7], que proporciona datos históricos de GHI, irradiancia normal directa (*Direct Normal Irradiance*, DNI), sus componentes directa y difusa, radiación extraterrestre y condiciones de cielo despejado desde 2004 hasta el presente. Por otra parte, la información pública a nivel de suelo procede de estaciones meteorológicas repartidas por la Comunidad de Madrid [42], que recogen medidas de irradiancia, temperatura ambiente, humedad relativa, presión, precipitación, velocidad y dirección del viento y radiación ultravioleta.
Los mapas NWP de temperatura ambiente se utilizan para estimar la temperatura de los paneles en el instante de la predicción, necesaria para calcular la potencia fotovoltaica producida. Los mapas NWP de nubosidad se usan para calcular el CCF, que se utiliza como indicador del tipo de día y se emplea para generar intervalos de incertidumbre y apoyar la actuación del EMS. La base de datos del CAMS se emplea para aumentar el histórico de GHI y rellenar faltas de medida en la instalación, recopilando 5 años para la predicción. La radiación extraterrestre se emplea en el modelo de predicción como un indicativo de la posición del sol. Finalmente, las estaciones a nivel de suelo permiten predecir la irradiancia en varios puntos, considerados como nodos, y emular una VPP a partir de la cual obtener conclusiones.

4.2 Preprocesado de los mapas NWP

Las variables de entrada a la estrategia de predicción recogidas de mapas de NWP requieren de un preprocesado que es necesario automatizar. En el apartado 4.2.1 se explica cómo obtener los mapas y la recopilación de la información mostrada a nivel de píxel. Por otro lado, en el apartado 4.2.2, se detalla la forma de obtener una nueva variable desarrollada en la presente tesis denominada CCF, con el objetivo de recopilar la información de la nubosidad de una forma más adecuada para predecir la GHI.

4.2.1 Obtención y análisis de los mapas

El modelo numérico de área limitada usado por la AEMET es el HARMONIE-AROME, que se emplea en predicciones meteorológicas, tanto en España como en otros países europeos [92,93]. Este realiza predicciones de variables meteorológicas en dos dominios: (i) Península Ibérica e Islas Baleares; y (ii) Islas Canarias. En España se utiliza desde junio de 2017, y dispone de una resolución temporal horaria y una resolución espacial de 2.5 km.

Estos modelos incluyen predicciones de GHI e DNI, necesarias para modelar la irradiancia en superficies inclinadas. Sin embargo, el coste asociado a la adquisición de esta información reduce considerablemente la

rentabilidad de las instalaciones fotovoltaicas. Por otra parte, la AEMET proporciona NWP de acceso gratuito en forma de mapas de nubosidad, temperatura ambiente, presión, velocidad y dirección del viento y precipitación, que se usa en la estrategia de predicción propuesta. La descarga se realiza desde la propia web [6], guardando en el equipo todas estas imágenes cronológicamente. Las imágenes se descargan diariamente y proporcionan predicciones con un horizonte temporal de 48 horas.

Los errores de GHI cometidos en el modelo numérico HARMONIE-AROME fueron estudiados en [94]. Los autores observaron una raíz del error cuadrático medio relativo (*relative Root Mean Square Error*, rRMSE), descrito en el anexo I, entre el 15% y el 30% para un tiempo de espera desde 15 minutos a 3 horas y media, respectivamente. El error cometido al predecir la radiación diaria también fue estudiado en diferentes estaciones y meses del año, observando un rRMSE entre el 5% en agosto y el 60% en enero.

Así, resulta muy interesante disponer de un mapa de predicción de toda la península como entrada al modelo, al ser favorable en el entorno VPP. Al disponer nodos en la VPP más distanciados entre sí se reduce el error global de la potencia fotovoltaica generada. De hecho, se observa en la literatura que si dichos nodos se encuentran en un área de unos 200 km de perímetro, la predicción del conjunto se optimiza [41,61].

Al descargar las imágenes desde la web de la AEMET, se observa que los pixeles no muestran el color RGB exacto de la leyenda. Por ello, la imagen debe ser tratada previamente antes de poder analizar la variable de interés dentro del mapa. Esto se observa en la imagen de la izquierda de la Figura 6, con respecto a la imagen procesada de la derecha, representando adecuadamente los colores recogidos de la leyenda situada debajo de dichas imágenes.

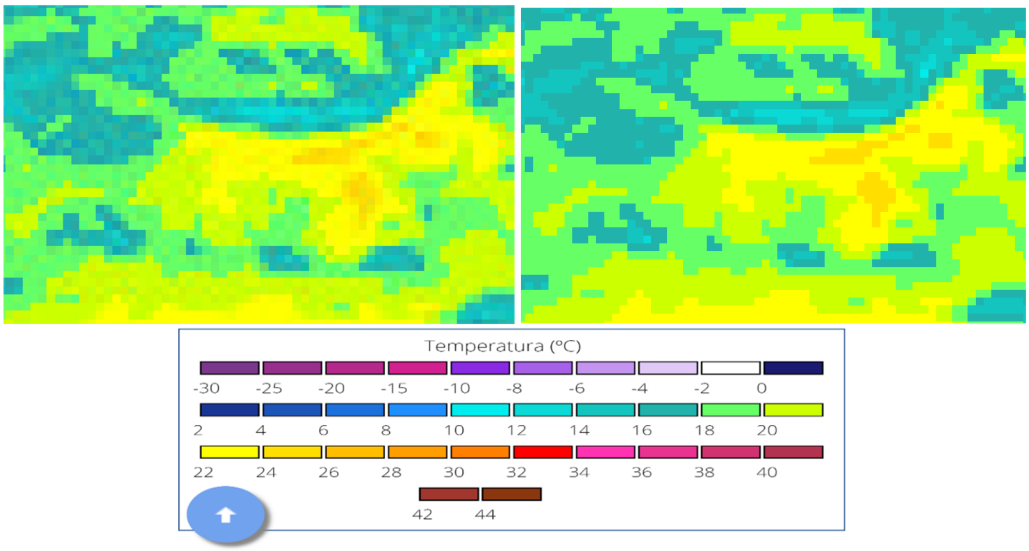


Figura 6.- Proceso de identificación de la temperatura ambiente asociada a cada pixel. Comparando la imagen inicial de la izquierda con los colores de la leyenda se obtiene la imagen final de la derecha que reconoce adecuadamente esta variable.

Para lograr esto, en primer lugar, se calcula la diferencia en términos de la intensidad de los colores primarios de la luz, RGB, entre cada píxel de la imagen y los colores de la leyenda. En caso de que el color de la imagen se corresponda exactamente con uno de los colores de la leyenda, las diferencias en R, G y B son cero. Sin embargo, esto no es habitual, por lo que habrá que cuantificar la similitud de colores para conocer la temperatura a la que hace referencia.

La expresión usada para determinar el color de temperatura ambiente de mayor similitud, T, es la siguiente:

$$Dif_i = \sqrt{\Delta R_i^2 + \Delta G_i^2 + \Delta B_i^2}$$

$$T = \min(Dif_{[1:N]})$$

Donde i toma valores de 1 a N, siendo N el número de colores existentes en la leyenda. ΔR, ΔG y ΔB representan la diferencia entre imagen y leyenda de los colores primarios rojo, verde y azul, respectivamente.

Se observa que la nueva imagen ha eliminado los colores intermedios y puede reconocer la temperatura ambiente con una resolución de píxel, que se corresponde con un cuadrado de lado $2.5\,km$ para dichos mapas. El proceso de filtrado de todo el mapa exige una elevada demanda computacional, por ello, únicamente se identifica la variable del pixel en el que se localiza el nodo o instalación, reduciendo considerablemente la carga computacional. Realizando este proceso para la totalidad de las imágenes recopiladas dentro de las fechas a estudiar, se obtiene finalmente el histórico de cada variable con resolución temporal horaria en la localización señalada.

4.2.2 Factor de cobertura de nubes

La nubosidad está muy relacionada con la GHI, de modo que resulta sencillo conocer las pérdidas de irradiancia que se producen en la superficie terrestre a partir de la transparencia de las nubes que ocultan el sol. Por lo tanto, resulta interesante usar los mapas de nubes de la AEMET como entradas al modelo, sin embargo, la posición exacta de las nubes en el cielo resulta difícil de estimar [95]. Debido a ello, para dar una precisión lo más fiable posible, se define una nueva variable denominada CCF, cuya función es indicar el sombreado en la instalación provocado por las nubes.

El CCF no tiene dimensiones y representa, dentro del mapa utilizado, la proporción de sombreado a nivel de suelo de las nubes en cada pixel para cada NWP de nubosidad. Este valor tiene una correlación negativa con la GHI, siendo cero cuando las nubes no bloquean al sol en la instalación y uno cuando lo cubren completamente.

Los mapas de nubosidad recogidos desde la web de la AEMET se expresan en tanto por ciento, variando entre el 0% para cielo despejado y el 100% para cielo completamente nublado, según se observa en la barra de colores de la Figura 7 para la Comunidad de Madrid.

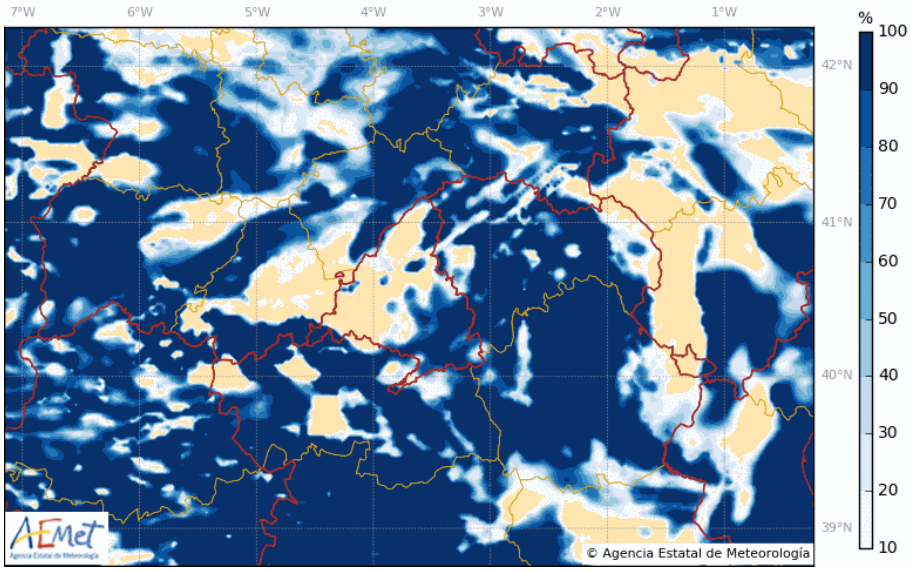


Figura 7.- NWP de nubosidad del 1 de marzo de 2020 a las 16:00 horas. Imagen descargada el 29 de febrero de 2020. Información procedente de la Agencia Estatal de Meteorología, AEMET©.

El siguiente paso para obtener una representación de las nubes que se interponen con el sol es conocer la posición de la estación meteorológica en el mapa. Cada pixel en el mapa tiene una resolución espacial de $2.5\,km$ de lado, aproximadamente. De esta forma, al ser una superficie bastante mayor que la propia estación, un único pixel puede representar la posición de la estación en el NWP. Para los NWP que aportan información a nivel de suelo, como el de temperatura ambiente, no resulta necesario realizar procesos posteriores, ya que el propio pixel contiene la información que se necesita. Sin embargo, las nubes se encuentran a diferentes alturas, llegando en ciertos casos a los 15 kilómetros [96], por lo que se requiere un proceso más elaborado para generar información relevante de la nubosidad y obtener así el CCF.

En este caso, se necesitan identificar las nubes que ocultan el sol en el mapa, ya que reducen la GHI al absorber la componente de irradiancia de mayor relevancia, la componente directa. El resto de nubes, al igual que las moléculas dispersas en la atmósfera, pueden incrementar la GHI observada, proporcionando radiación difusa adicional. Sin embargo, en el presente estudio no se tienen en cuenta, principalmente porque su efecto es de segundo orden y difícil de estimar con la información recopilada en el estudio. En [97] se puede observar dicho fenómeno bajo un día soleado, distinguiendo un incremento en la GHI, que supera la radiación en cielo despejado, en instantes de breve duración.

Para conocer las nubes involucradas, en la Figura 8 se representa la posición del sol en el cielo con frecuencia horaria para un día en particular,

correspondiente al equinoccio de primavera, con respecto a un punto situado en el hemisferio norte, representando también las variables más relevantes usadas para determinar el CCF. Como se puede observar, si la altura a la que se encuentra la nube y la posición del sol en elevación son conocidas, es posible calcular la distancia entre nube y piranómetro por simple trigonometría. Seguidamente, mediante la información del azimut, se determina su punto exacto en el plano terrestre. En el caso de los NWP de nubosidad de la AEMET, se puede realizar una representación idéntica a la Figura 8 en píxeles: al conocer la posición del píxel de la estación meteorológica se pueden determinar los píxeles del mapa que indican si hay nubes que ocultan el sol al sensor de irradiancia, para una altura definida.

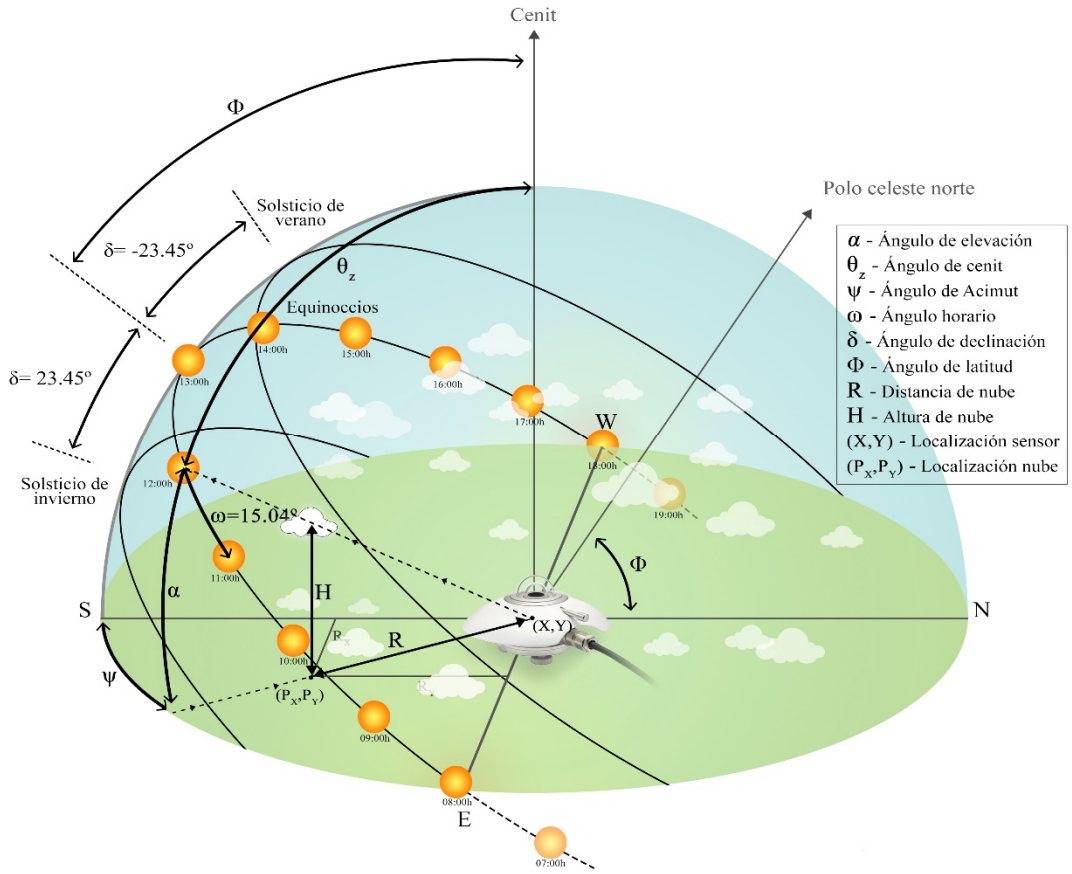


Figura 8.- Posicionamiento horario del sol durante el equinoccio de primavera, definiendo las variables más relevantes en el plano terrestre para obtener el CCF.

A continuación se representa la posición del píxel en el mapa, donde (R_X, R_Y) es la posición relativa del píxel con respecto al piranómetro y $S = 2.5\,km$ es la escala de píxel o distancia en el mapa que ocupa un lado de píxel.

$$P_X = X + R_X \approx \mathrm{X} + \frac{H}{S}\frac{sin(\psi)}{tan(\alpha)}$$

$$P_Y = Y + R_Y \approx \mathrm{Y} + \frac{H}{S}\frac{cos(\psi)}{tan(\alpha)}$$

Si se conocen los píxeles cuyas nubes sombrean el piranómetro, en todo el rango de alturas de nube, se dispone de una información de gran relevancia para la predicción de la GHI. La posición del sol desde la superficie terrestre es conocida, de forma que su azimut y elevación se pueden determinar de forma precisa, sin embargo, los NWP de nubosidad de la AEMET no dan información de la altura a la que se encuentran las nubes. Debido a ello, se recoge información de toda la zona en la que las nubes pueden provocar sombreado en el piranómetro.

La situación de las nubes en el mapa tiene un error asociado que depende de la precisión de las imágenes satélite usadas en los modelos NWP [98], la altura real a la que se encuentra la nube y la resolución espacial y temporal de los mapas. Debido a esta incertidumbre, los valores generados pueden ofrecer información poco precisa. Además, al considerar el estudio a nivel de pixel, se ha comprobado que el CCF es propenso a variar excesivamente a lo largo del tiempo.

De esta forma, se ha seleccionado un parámetro adicional denominado umbral de incertidumbre (*uncertainty threshold*, μ), que expande el área seleccionada para determinar el CCF en el mapa, suavizando su variación a lo largo del tiempo y permitiendo además que pueda ser utilizado como un indicador del tipo de día en la estación. μ se ha seleccionado incrementando progresivamente su valor hasta obtener una variación poco abrupta del CCF, observando así mismo que se reduce algo el error cometido en las predicciones, seleccionando $\mu = 20\,km$. En la Figura 9 se representa un flujograma que describe de una forma más intuitiva el procedimiento realizado para obtener el CCF.

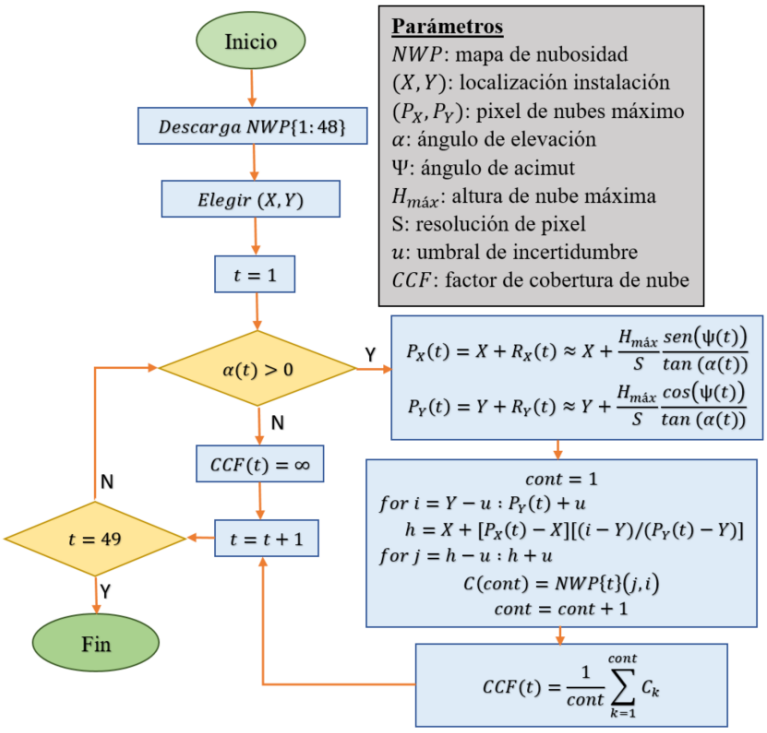


Figura 9.- Diagrama de bloques del procedimiento para obtener el CCF de los NWP.

4.3 Variables de relevancia

A la hora de elegir las variables más relevantes, se ha estudiado la correlación existente entre estas y la GHI, usando para ello el PCC, donde R y S son las variables a comparar, N es el número de medidas, μ representa la media y σ la desviación típica. El PCC indica una correlación completa cuando el valor resultante es ±1, no identificando ningún tipo de correlación entre variables cuando este resulta cero.

$$PCC\ (R,S) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{R_i - \mu_X}{\sigma_X}\right)\left(\frac{S_i - \mu_Y}{\sigma_Y}\right)$$

El estudio se ha realizado para varias estaciones meteorológicas situadas en la Comunidad de Madrid, concretamente las de Centro Municipal de Acústica, Hortaleza, Juan Carlos I, Moratalaz, Peñagrande y Villaverde [42]. Además de emplear información propia de la instalación, se han usado los mapas de NWP, obteniendo un histórico comprendido entre el 1 de junio de 2020 y el 31 de mayo de 2021, un año completo. Las estaciones consideradas tienen información de GHI, temperatura ambiente, humedad, viento, presión y precipitación, añadiendo el CCF estimado de los mapas de NWP y la radiación extraterrestre.

La radiación extraterrestre, b_{00}, es un parámetro que representa la radiación emitida por el sol previa al paso por la atmósfera.

$$b_{00} = b_0 \cdot \epsilon(n) \cdot \cos(\theta_s(n))$$

$$\epsilon(n) = 1.00011 + 0.034221 * cos(\mathrm{I}(n)) + 0.00128 * sin(\mathrm{I}(n)) + 0.000719 * cos(2\mathrm{I}(n)) + 0.000077 * sin(2\mathrm{I}(n))$$

$$\theta_s(n) = acos(sin(\delta(n)) * sin(\Phi) + cos(\delta(n)) * cos(\Phi) * cos(\omega))$$

Donde $b_0 = 1367\ W/m^2$ es la constante solar, $\epsilon(n)$ es la excentricidad de la órbita para el día del año n y $\mathrm{I(n)} = 2\pi(\mathrm{n}-1)/365$, y θ_s es el ángulo cenital del sol, en el que δ es la declinación y Φ la latitud. La Tabla 1 muestra los valores del PCC para las diferentes estaciones. Se observa que las variables de mayor correlación son la radiación extraterrestre, CCF, humedad y temperatura ambiente para todos los casos. En la Figura 10 se muestran las correlaciones entre todas las variables en la estación meteorológica de Juan Carlos I, que tiene un comportamiento similar al resto de estaciones.

Tabla 1.- PCC de las variables con respecto a la GHI para las diferentes estaciones meteorológicas utilizadas en el estudio.

PCC(GHI)	Extraterrestre	CCF	Temperatura	Humedad	Viento	Presión	Lluvia
CtroAcustica	**0.788**	**0.846**	**0.548**	**-0.563**	0.124	0.123	-0.074
Hortaleza	**0.737**	**0.814**	**0.474**	**-0.576**	0.132	0.048	-0.122
JuanCarlosI	**0.783**	**0.853**	**0.465**	**-0.534**	0.113	-0.008	-0.134
Moratalaz	**0.745**	**0.832**	**0.522**	**-0.624**	-0.056	0.221	-0.119
Peñagrande	**0.776**	**0.84**	**0.542**	**-0.562**	0.118	0.125	-0.121
Villaverde	**0.742**	**0.817**	**0.498**	**-0.619**	0.056	0.202	-0.127

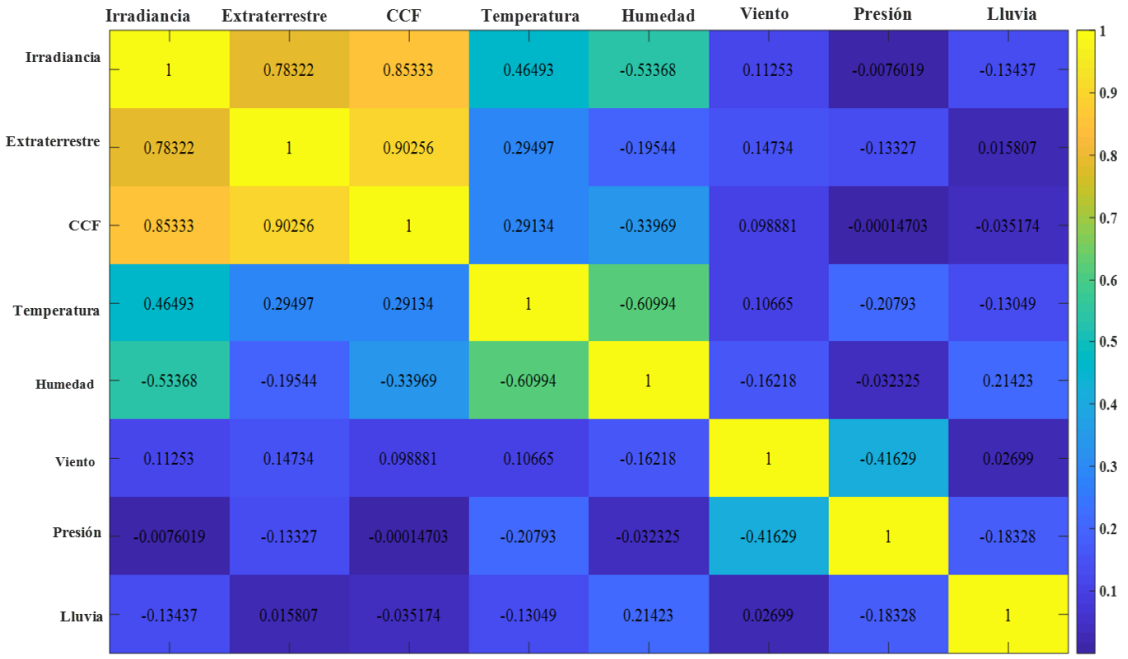


Figura 10.- Matriz de PCC entre todas las variables consideradas en el estudio, para la estación meteorológica de Juan Carlos I.

Como puede observarse, la GHI y la radiación extraterrestre tienen una alta correlación, $PCC = 0.783$, ya que la primera es el resultado de la segunda al cruzar la atmósfera terrestre. Al considerar el CCF, se observa una correlación mayor, $PCC = 0.853$, concluyendo que el CCF se trata de un gran predictor de la GHI.

Una vez analizados los resultados, se han tomado la radiación extraterrestre, el CCF y la temperatura ambiente como las variables más relevantes, usándose como entradas para la predicción de horizonte temporal diario. La humedad también tiene una correlación destacable, sin embargo, se descarta del estudio ya que la AEMET no dispone de esta variable en forma de mapas. Sí se puede conocer la predicción de humedad por municipios [99], sin embargo, puede generar limitaciones en el estudio.

4.4 Clasificación del error

Al introducir en el modelo de predicción información basada en predicciones meteorológicas de la AEMET, se asume que estas tienen un error. Por lo tanto, resulta apropiado estudiar la incertidumbre asociada, con el fin de considerar despreciable el efecto que las predicciones de los mapas meteorológicos generan en la estrategia de predicción desarrollada en la presente tesis.

En primer lugar, se evalúa el error entre las predicciones de temperatura ambiente de la AEMET y las medidas obtenidas en la estación de Juan Carlos I. En la Figura 11 se observan estas diferencias para los meses de diciembre de 2019 a mayo de 2020. Se aprecia que el error suele encontrarse dentro del rango de $\pm 2^{o}C$, lo que resulta adecuado para el caso

de estudio. Con respecto a las medidas del error del anexo I, se observa un MAE de 1.29^oC y un RMSE de 1.78^oC. El error máximo cometido es de $|\Delta T_{máx}| = 7^oC$, de modo que ciertos valores atípicos sí pueden reflejar un elevado error.

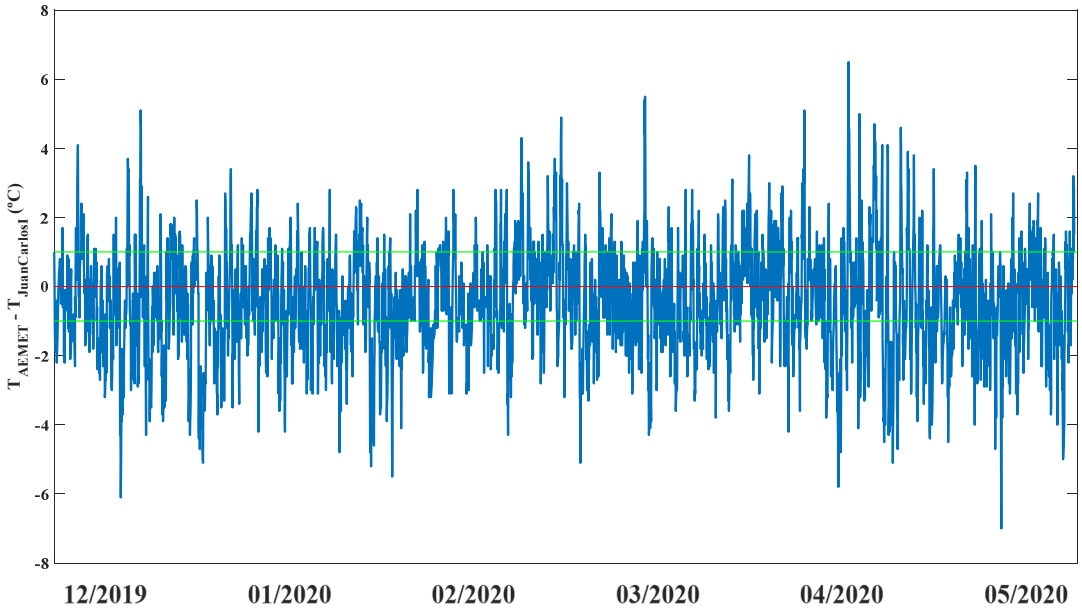


Figura 11.- Diferencias horarias entre los mapas NWP de temperatura ambiente y las medidas reales de la estación de Juan Carlos I entre diciembre de 2019 y mayo de 2020.

Además de las diferencias entre los mapas de NWP y las medidas reales, se debe considerar el error asociado al aumentar el tiempo de espera en los mapas. La Figura 12 muestra, en el propio mapa de NWP, las diferencias de temperatura ambiente el 1 de marzo de 2020 a las 16 horas, al contrastar dos imágenes de predicción: la primera calculada para un tiempo de espera de 2 horas, por lo tanto realizada a las 14 horas, superpuesta con otra calculada para un tiempo de espera de 8 horas, por lo tanto realizada a las 8 horas. Como recordatorio del apartado 3.2, el tiempo de espera es la diferencia entre el instante en el que se genera una predicción y el instante de tiempo que se quiere predecir.

En estas gráficas, el color azul representa que la predicción de menor tiempo de espera ha estimado una temperatura 2^oC más fría que el otro, mientras que el color rojo indica lo contrario. Los colores están escalonados, de forma que por cada tono más claro se ha cometido una diferencia adicional de $\Delta T = 2^oC$ en el error. Realizando un estudio en mayor detalle, se comprueba que un 72% del mapa queda inalterado y el 28% restante del mapa se modifica un valor de $\Delta T = \pm 2^oC$.

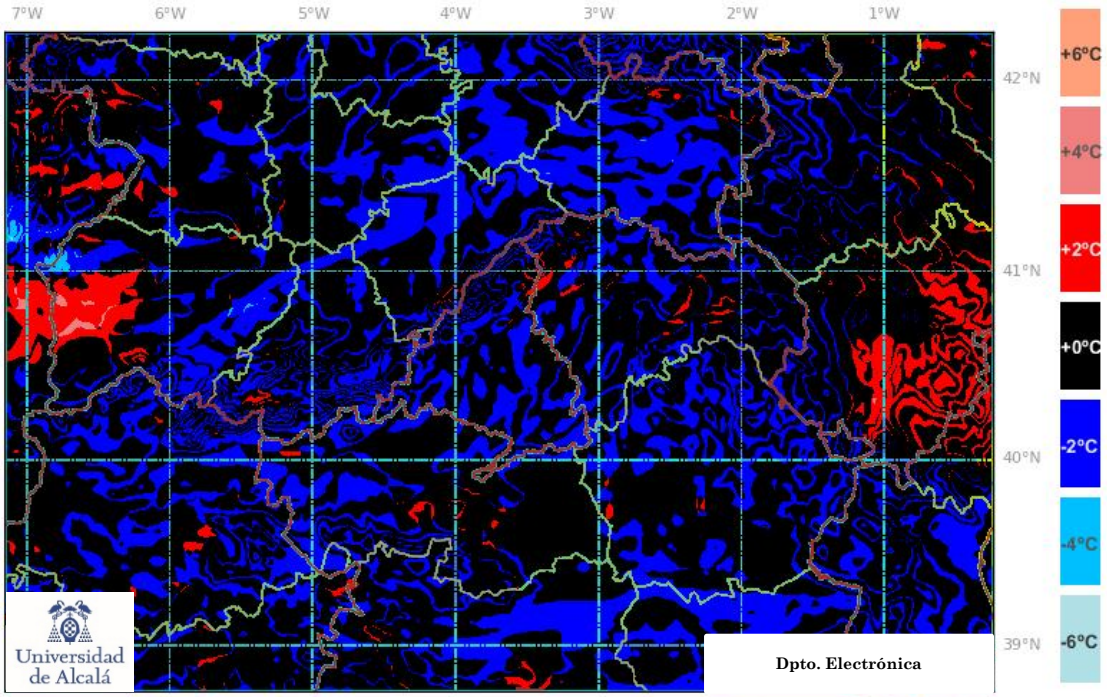


Figura 12.- Error cometido entre dos predicciones de temperatura ambiente el 1 de marzo de 2020 a las 16 horas en la Comunidad de Madrid, con una diferencia de tiempo de espera entre predicciones de 6 horas. Información elaborada utilizando, entre otras, la obtenida de la AEMET©.

En la Figura 13 se observa que estas diferencias se ven incrementadas al considerar una diferencia entre tiempos de espera mayor: la primera calculada para un tiempo de espera de 2 horas, por lo tanto realizada a las 14 horas, superpuesta con otra calculada para un tiempo de espera de 44 horas, por lo tanto realizada a las 20 horas de dos días atrás.

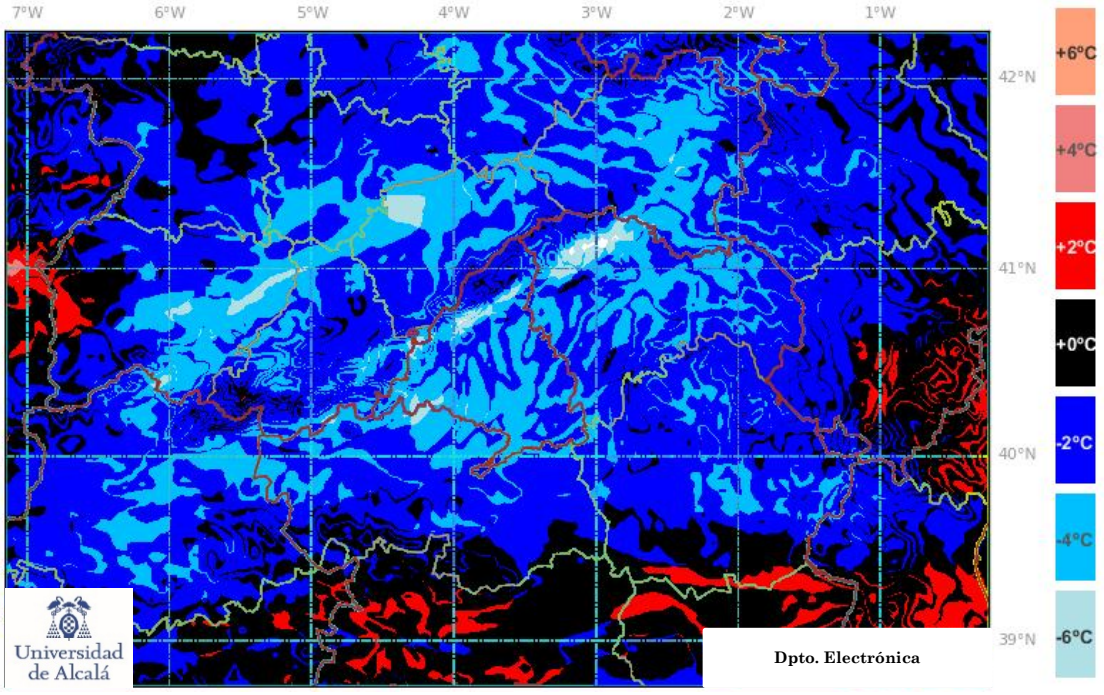


Figura 13.- Error cometido entre dos predicciones de temperatura ambiente el 1 de marzo de 2020 a las 16 horas en la Comunidad de Madrid, con una diferencia de tiempo de espera entre predicciones de 42 horas. Información elaborada utilizando, entre otras, la obtenida de la AEMET©.

Realizando un estudio en mayor detalle, se comprueba que el error aumenta considerablemente: un 35% del mapa queda inalterado, un 49% se modifica un valor $\Delta T = \pm 2^o C$, un 15% un valor $\Delta T = -4^o C$ y el 1% restante $\Delta T = -6^o C$. Así, se concluye que es necesario descargar los mapas de NWP lo más actualizados posible, minimizando de esta forma el error introducido a la estrategia de predicción.

4.5 Sobremuestreo y faltas en la medida

La GHI y la temperatura ambiente constituyen las variables de entrada a la estrategia de predicción intradiaria. La GHI obtenida del CAMS se usa para rellenar posibles faltas en la medida y ampliar el histórico de la instalación hasta obtener 5 años de información. Por otra parte, se realiza un sobremuestreo de la temperatura ambiente procedente de los NWP de la AEMET, para obtener una resolución temporal de 15 minutos.

a) Preprocesado de GHI

El histórico de GHI recogido en la instalación con respecto al recopilado de medidas satélite proceden de diferentes fuentes. Por una parte, las medidas de la instalación se sitúan a nivel de suelo y son recopiladas cada 30 segundos por un piranómetro. Por otra parte, las medidas del CAMS son recogidas desde el espacio con una resolución temporal de un minuto. Por tanto, resulta necesario un análisis comparativo para conocer si es posible unificar ambas bases de datos.

Para este fin, se calcula la correlación lineal entre ambas variables, haciendo uso del PCC similar al apartado 4.3. Se obtiene un valor de 0.985 entre las dos bases de datos, por lo que se asume que existe una correlación lineal, al encontrarse cerca del valor unidad. En la Figura 14 se muestran ambos históricos junto a los errores del anexo I. Se aprecia bastante similitud entre las bases de datos, con un error medio entre ellas de $MAE = 46.97\ W/m^2$.

b) Sobremuestreo de temperatura ambiente

En segundo lugar, es necesario aumentar la resolución temporal de la temperatura ambiente, ya que la información recogida de los mapas de NWP es horaria y la estrategia de predicción intradiaria se ha fijado en 15 minutos para proporcionar una información más actualizada al EMS. Para este fin, se recurre a una interpolación cuadrática del histórico, al suponer que la temperatura ambiente con resolución temporal horaria varía suavemente y de forma continua, cumpliendo con la propiedad de Darboux [100].

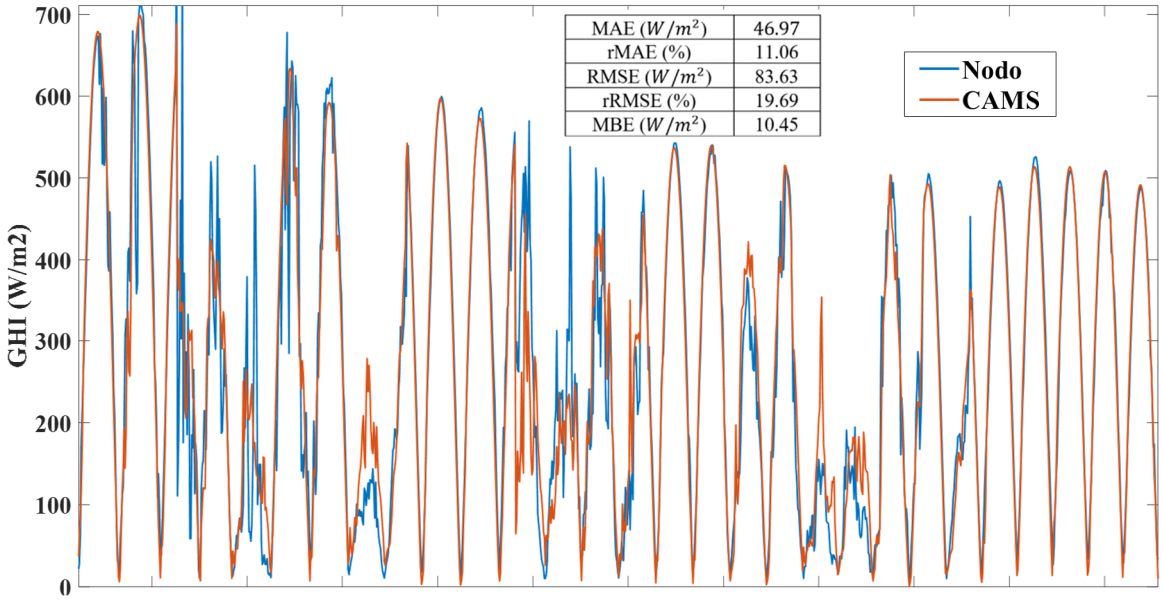


Figura 14.- Medidas de GHI de la instalación y del CAMS, junto a las medidas del error calculadas entre las dos bases de datos.

Para comprobar la precisión obtenida en los mapas, la Figura 15 muestra la temperatura ambiente medida en la instalación y la temperatura ambiente obtenida de los mapas de NWP con preprocesado. Se observa que la interpolación cuadrática sigue el gradiente de los valores medidos, obteniendo un error pequeño de $MAE = 2.12^{o}C$.

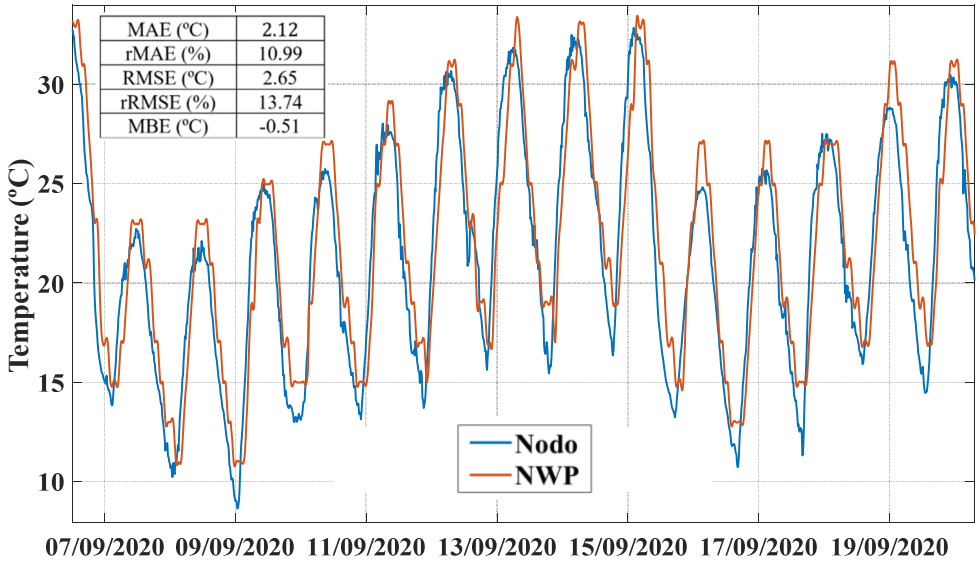


Figura 15.- Medidas de temperatura ambiente del nodo real de pruebas y de las predicciones de la AEMET con una resolución temporal de 15 minutos, junto a las medidas del error calculadas entre las dos bases de datos.

c) Rellenar faltas en la medida

Además de utilizar las variables del preprocesado como entrada a la estrategia de predicción, son útiles para rellenar posibles faltas de medida en la instalación. En la Figura 16 se observa la reconstrucción del histórico en momentos que se ha perdido información en la estación meteorológica del

nodo real de pruebas de la Universidad de Alcalá. Se aprecia que el proceso de construcción evita la generación de huecos en el histórico y se amolda adecuadamente a la información que se espera recibir.

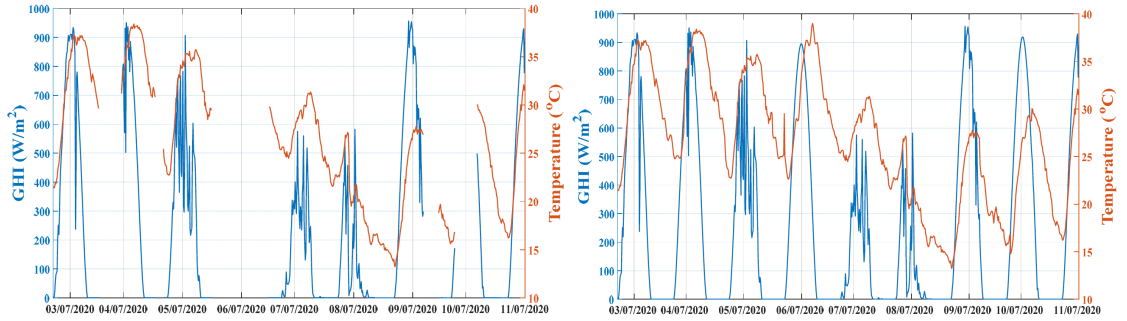


Figura 16.- Reconstrucción de las medidas de GHI y temperatura ambiente en el nodo real de pruebas situado en la Universidad de Alcalá.

4.6 Conclusiones

En este capítulo se han determinado las variables de entrada con mayor relevancia, analizando el error cometido entre medidas y realizando un preprocesado para la estrategia de predicción presentada en la tesis. A través de los mapas de nubosidad se obtiene el CCF, que permite determinar la proporción de sombreado a nivel de suelo de las nubes, siendo esta una contribución de la tesis. Al seleccionar las variables de relevancia, se ha cuantificado la correlación entre estas y la GHI mediante el PCC. Como resultado, se han seleccionado la radiación extraterrestre, el CCF y la temperatura ambiente, junto a la propia GHI.

Se ha comprobado que los mapas NWP de temperatura ambiente obtienen una precisión adecuada al compararse con las medidas recopiladas en las estaciones meteorológicas a nivel de suelo. Por otra parte, se han comparado mapas NWP a diferentes tiempos de espera, concluyendo que es necesario disponer de esta información lo más actualizada posible, ya que conforme aumenta el tiempo de espera, el error asociado también se ve incrementado. Finalmente, la GHI obtenida del CAMS se usa para rellenar posibles faltas en la medida y ampliar el histórico de la instalación hasta obtener 5 años de información. Por otra parte, se realiza un sobremuestreo de la temperatura ambiente de la AEMET para obtener una resolución temporal de 15 minutos, usándose también para rellenar faltas en la medida de temperatura ambiente.

5 Estrategia de predicción diaria

Una vez obtenidas las variables, y seleccionadas aquellas a usar como entrada a la estrategia, este capítulo se centra en las predicciones de irradiancia realizadas con un horizonte temporal de un día. El apartado 5.1 describe el modelo de predicción utilizado, describiendo los dos modelos usados, en los apartados 5.1.1 y 5.1.2, y su hibridación, en el apartado 5.1.3. El capítulo 5.2 muestra los resultados obtenidos al comparar el modelo híbrido desarrollado con otras técnicas tradicionales, tanto para un nodo real, en el apartado 5.2.1, como para una VPP emulada, en el apartado 5.2.2. Finalmente, el apartado 5.3 analiza las conclusiones más relevantes extraídas del capítulo.

5.1 Descripción de la estrategia de predicción

La estrategia de predicción con un horizonte temporal diario resulta de gran interés ya que permite visualizar la generación fotovoltaica producida con gran antelación y actuar en consecuencia. Gracias a ello, se facilitan labores muy importantes, como la gestión de la carga de las baterías durante la noche, la energía solar diaria a controlar o valorar si entrar al mercado eléctrico. Para obtener una precisión adecuada, se usan las predicciones meteorológicas proporcionadas por la AEMET, al aportar información a futuro que permite conocer relaciones entre la irradiancia a predecir y las variables exógenas, en este caso el CCF y la temperatura ambiente, en el momento de la predicción. Esto evita emplear históricos del pasado para predecir a horizontes temporales diarios, ya que las condiciones meteorológicas pueden verse modificadas, principalmente durante el transcurso de la noche.

Un requisito para obtener una precisión óptima es utilizar una base de datos suficientemente extensa. Sin embargo, en etapas iniciales o cuando se integran nodos a una VPP existente, hay una escasez de datos que puede reducir dicha precisión. Dentro del ámbito energético, esto puede repercutir en sanciones económicas impuestas por el operador de red. Según se ha comentado en el apartado 4.1, los NWP de irradiancia proporcionan una solución a esta problemática, sin embargo, obtener esta información acarrea un coste que también repercute en su rentabilidad, especialmente cuando el número de emplazamientos aumenta. Con el fin de lograr una alta precisión en históricos reducidos, sin coste asociado, en la Figura 17 se muestra el esquema general del modelo propuesto en la presente tesis, basado en un sistema híbrido que pondera dinámicamente, en función del tipo de día, una predicción basada en ANNs con un modelo innovador basado en horas similares.

El método dispone de dos partes diferenciadas: (i) fase de preprocesado y análisis de datos; y (ii) predicción implementada a través de una estrategia híbrida. El preprocesado resulta crucial para optimizar la información que se envía al modelo de predicción, sobre todo cuando se

dispone de un histórico de información escaso. En el presente caso, se usa información de nubosidad y temperatura ambiente recopilada de forma gratuita desde la AEMET en forma de mapas [6]. La información de nubosidad, junto a la posición del sol, se emplea para definir el CCF del apartado 4.2.2. Este factor contiene información del sombreado en la instalación causado por las nubes y caracteriza las mayores pérdidas de irradiancia, al conocer en cierto aspecto la radiación directa bloqueada por las nubes. La temperatura ambiente también es de relevancia, según se ha comprobado en el apartado 4.3.

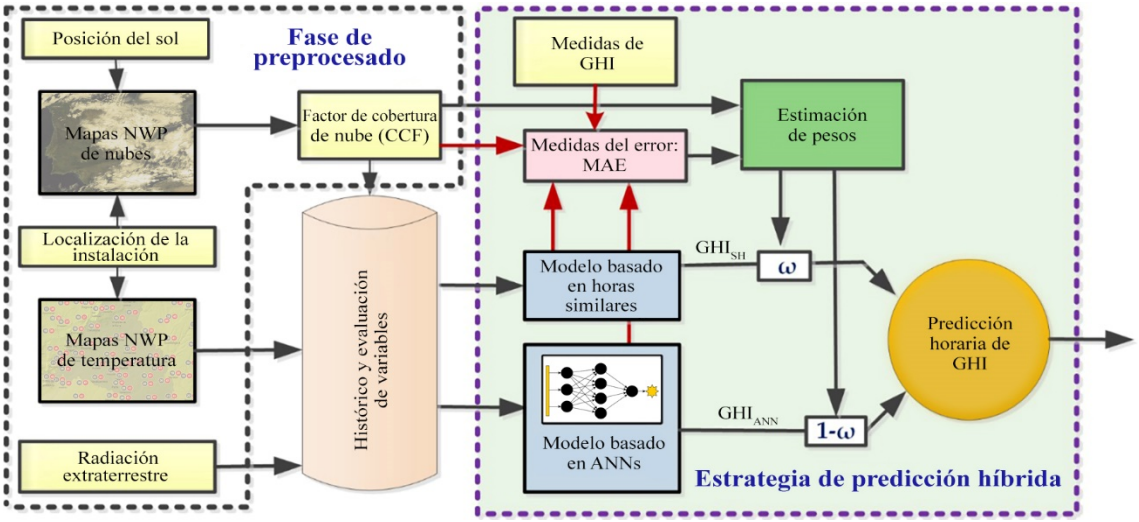


Figura 17.- Esquema general del modelo usado para predicciones con un horizonte temporal diario, que consiste en dos fases: una etapa previa de procesamiento de datos y la predicción híbrida. La estrategia de predicción está basada en ANNs y un modelo basado en horas similares, ponderando dinámicamente la salida en función del tipo de día.

La radiación extraterrestre ofrece información de la radiación emitida por el sol previa al paso por la atmósfera, de modo que también será de utilidad en el modelo. El anexo III describe de forma resumida la radiación solar y el efecto de las nubes al percibirse a nivel de suelo. El histórico de GHI se obtiene de varias estaciones meteorológicas [42], además de un piranómetro en un nodo real situado en la Escuela Politécnica Superior de la Universidad de Alcalá. Dado que las variables de entrada al modelo son mapas de nubes y temperatura ambiente de NWP, poseen información de cualquier zona del territorio español.

En la fase de predicción de la Figura 17 se desarrolla una estrategia híbrida que pondera entre modelos en función del histórico recopilado. El modelo basado en horas similares tiene un rendimiento destacable en históricos escasos, y la ANN mejora el rendimiento del mismo cuando el histórico de datos se va incrementando. Dichos modelos se complementan para ofrecer una predicción precisa en todo instante. Para combinar ambas estrategias se asignan pesos a cada modelo en función del tipo de día y el MAE de cada modelo. Se distinguen días de tipo soleado, nuboso y encapotado en función del CCF para formar los grupos. Los pesos se actualizan diariamente para minimizar el error producido en el modelo.

5.1.1 Predicción basada en horas similares

La estrategia de horas similares se inspira en la predicción basada en días similares [10,43−51,53]. Ya que la predicción basada en días similares se utiliza a nivel comparativo en el estudio, resulta interesante realizar una descripción breve de esta. La predicción basada en días similares supone que la GHI del día a predecir se puede estimar a partir de otras variables de alta correlación la misma, de las que se dispone información futura, usando información de días pasados para seleccionar el día de comportamiento más similar como predicción. En la tesis se utiliza la predicción de días similares, a nivel de contraste con el modelo basado en horas similares, de la siguiente forma: (i) se obtiene el CCF y la temperatura ambiente al día siguiente; (ii) se realizan comparaciones con históricos de días pasados de estas variables; (iii) se decide el día de mayor similitud; y (iv) se extrae su GHI medida y se utiliza como solución del día a predecir.

Es importante elegir un número específico de días pasados para la predicción. Si se toma un número de días excesivo, las estaciones del año tendrán influencia en la predicción, al cambiar las condiciones climáticas, y la GHI medida se verá afectada. Así mismo, si el número de días es pequeño, no se representarán adecuadamente los tipos de día y la predicción no dispondrá de suficiente información para realizar una buena predicción. Es de reseñar que, si se dispone de un histórico superior a un año, el número de días seleccionados aumenta proporcionalmente, al considerarse los mismos días de otros años, ganando potencial esta técnica. El número de días candidatos por año se ha seleccionado minimizando el RMSE en la predicción, seleccionando un total de 14 días. De estos 14 días se elige aquel con el CCF y la temperatura ambiente horaria más similar al día a predecir, usando la GHI medida de ese día como valor de la predicción. El modelo empleado en la tesis se inspira del trabajo realizado en [43], de la siguiente forma:

1. Al existir escalas diferentes entre la temperatura ambiente y el CCF, se normalizan las variables de entrada, considerando idénticos órdenes de magnitud.

2. Se calculan las diferencias horarias entre el día a predecir y cada uno de los 14 días anteriores seleccionados, tanto para el CCF como para la temperatura ambiente.

3. El día que genere menores diferencias para estas variables se usa como predicción, tomando su GHI medida como solución. Estas diferencias se calculan mediante la distancia euclídea para cada uno de los días considerados:

$$\mathrm{GHI\{D\}} = \mathrm{GHI}\left\{\min\left[ED_{[1:14]}\right]\right\} = \mathrm{GHI}\left\{\min\left[\sqrt{\sum_{i=1}^{24}\left(NWP_{i,0} - NWP_{i,[1:14]}\right)}\right]\right\}$$

Conocido el desarrollo de la predicción basada en días similares, se describe el modelo propuesto basado en horas similares. Este modelo desarrollado, que es una de las contribuciones aportadas en la tesis, ofrece una mayor precisión con los datos reales recopilados en la tesis y también bajo un entorno VPP, principalmente ante históricos escasos.

Una predicción basada en días similares considera las condiciones meteorológicas de todo un día, sin embargo, estas condiciones no siguen una tendencia única durante todo el día, siendo más razonable usar rangos de tiempos más reducidos. El modelo propuesto de horas similares considera un rango de tres horas para encontrar las similitudes: la hora para la cual se realiza la predicción y sus horas adyacentes.

Adicionalmente, si en lugar de utilizar momentos pasados con horas del día coincidentes, se seleccionan instantes con radiación extraterrestre similar, existirá una relación más directa en términos de irradiancia y el número de candidatos también puede ampliarse, lo que resulta muy interesante cuando se dispone de un histórico de pocos días en la instalación.

El modelo basado en horas similares se describe en el diagrama de bloques de la Figura 18, seleccionando las horas candidatas (c) del histórico de datos (z), y buscando similitudes entre candidatos, con respecto al instante a predecir (t), en términos del CCF y la temperatura ambiente (T). Los vectores de diferencias (D) de CCF y temperatura ambiente (T) se obtienen para evaluar la similitud de las horas candidatas (c) con respecto a la hora a predecir (t). Las horas adyacentes $(t-1)$ y $(t+1)$ también se comparan con sus adyacentes asociadas $(c-1)$ y $(c+1)$. Al haber 6 diferencias en total, resulta conveniente incorporar un peso (ω) a cada una de estas, para representar su importancia relativa en la predicción. Se seleccionan los pesos (ω) que minimizan el MAE en el histórico de predicciones pasadas, actualizando estos pesos diariamente.

Por último, se calcula la distancia euclídea (*Euclidean Distance*, ED) de las diferencias (D) ponderadas con los pesos (ω), para seleccionar los instantes de mayor relación con la hora a predecir (t). Se eligen los tres candidatos más similares, calculando su media como solución, reduciendo de esta forma los errores atípicos de predicción. Para optimizar la precisión del modelo, se realiza una corrección de la radiación extraterrestre (E), mejorando los resultados obtenidos. Este modelo realiza una iteración del proceso hasta obtener el conjunto de predicciones con un horizonte temporal diario, siempre que la radiación extraterrestre (E) no sea cero.

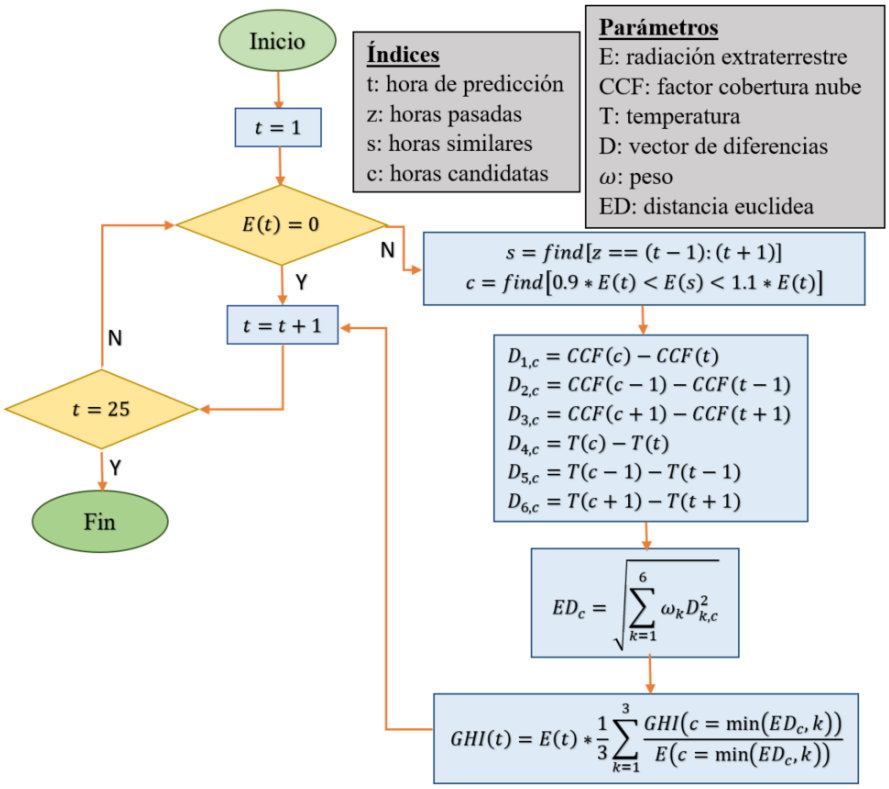


Figura 18.- Diagrama de bloques de la estrategia de horas similares para predecir la GHI con un horizonte temporal diario.

5.1.2 Predicción basada en redes neuronales

El incremento en la capacidad computacional de los procesadores actuales ha hecho que las redes neuronales se utilicen de forma generalizada, por su no linealidad y proporcionar una solución excelente sin necesidad de conocer en profundidad los modelos analíticos y estadísticos asociados a la variable a predecir. También se ha seleccionado esta técnica por ser la más usada en predicción de irradiancia [63,64]. De los diferentes tipos de redes neuronales existentes, las redes basadas en series temporales o recurrentes fueron descartadas, según se ha justificado al inicio del presente capítulo. Para el caso de estudio, resultan óptimas redes sencillas que saquen relaciones simples al problema y permitan una generalización de la solución en históricos escasos, con una elevada velocidad computacional. Por ello, se seleccionaron las ANN-MLP, debido a su simplicidad y velocidad de respuesta. Este tipo de redes tienen un buen rendimiento al relacionar entradas y salidas, utilizándose extensamente en irradiancia [40]. En cuanto a su estructura, la ANN-MLP utilizada puede dividirse de la siguiente forma:

1) El número de neuronas de la capa de entrada son 7, que se corresponden con: (i) CCF de la hora a predecir, t; (ii) CCF de la hora anterior a esta, $t-1$; (iii) CCF de la hora posterior a esta, $t+1$; (iv) temperatura de la hora a predecir, t; (v) temperatura de la hora anterior a esta, $t-1$; (vi) temperatura de la hora posterior a esta, $t+1$; y (vii) radiación extraterrestre como indicativo de la hora del día, t.

2) En la capa oculta se seleccionan un total de 10 neuronas, usando el RMSE como evaluador del error a minimizar. La función de activación empleada es logarítmica sigmoidal. Se emplea una única capa oculta al comprobar que incluir más capas no ofrece mejores resultados, teniendo como inconveniente el aumento de la carga computacional y el incremento en el riesgo de sobreentrenamiento.

3) Finalmente, la capa de salida tiene una neurona, que se corresponde con la GHI de la hora a predecir, t, con una función de activación lineal.

Para simular un nodo real de VPP en su fase inicial, la ANN se entrena diariamente, considerando para el primer día de predicción 7 días de histórico. Al entrenar con históricos reducidos y validar el algoritmo se observa que la ANN no generaliza adecuadamente la predicción, ya que el error cometido varía excesivamente entre diferentes predicciones. Para solucionar esto, se implementa una técnica comúnmente utilizada ante escasez de datos [69,101], que consiste en predecir con múltiples ANNs iguales en paralelo, realizando la media de todas las salidas como predicción de irradiancia. Se comprueba que realizando la media de 30 ANNs se elimina la variación del error cometido y la predicción se generaliza.

El algoritmo de aprendizaje utilizado ha sido seleccionado entre tres diferentes: Levenberg-Marquardt (LM), Regularización Bayesiana (*Bayesian regularization*, BR) y el Gradiente Escalado Conjugado (*Scaled Conjugate Gradient*, SCG). En la Figura 19 se comprueba el rMAE de cada uno de ellos, observando que el SCG genera una mayor precisión para históricos inferiores a 70 días, mientras que el LM genera una mayor precisión a partir de ese instante. Finalmente, se ha decidido aplicar el algoritmo de aprendizaje óptimo en función del histórico disponible, con el fin de generar la mayor precisión posible.

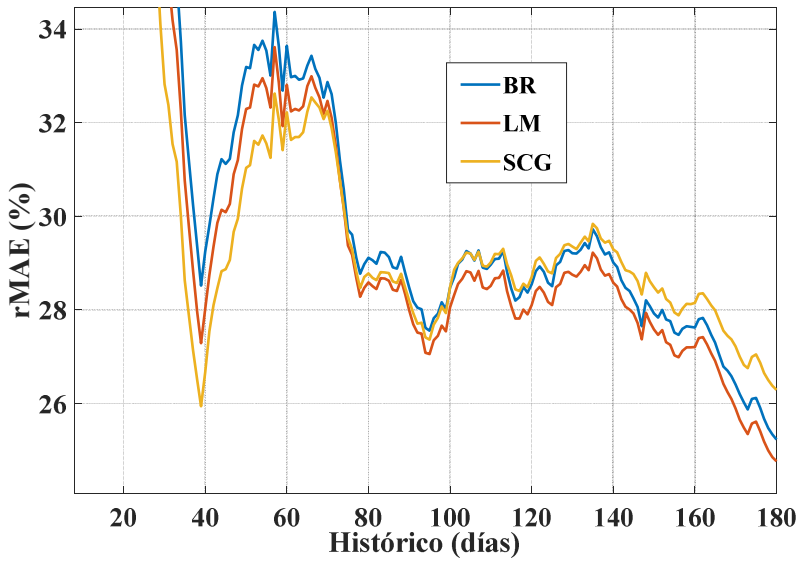


Figura 19.- rMAE de la predicción basada en ANNs en función del algoritmo de aprendizaje seleccionado. Se observa que el SCG minimiza el error producido hasta 70 días de histórico, mientras que LM ofrece una mayor precisión a partir de ese instante.

A la hora de cuantificar la importancia de cada una de las entradas seleccionadas, se realiza un análisis de sensibilidad de la siguiente forma: (i) se eliminan las neuronas una a una, prediciendo la GHI para todo el histórico recopilado; (ii) se calcula la diferencia absoluta entre dicha predicción y la obtenida con todas las neuronas activas; y (iii) se calcula el valor porcentual de cada entrada de la siguiente forma:

$$S_i(\%) = \frac{\Delta Pred_i}{\sum_{j=1}^{7} \Delta Pred_j}$$

Al comparar el análisis de sensibilidad con el PCC del apartado 4.3, es posible conocer si existe cierta coherencia entre la ANN y la relación lineal de esta: si las variables de entrada con mayor PCC son las que toman un mayor peso en el análisis de sensibilidad de la ANN, la relevancia de estas en la predicción de irradiancia se remarca.

Los resultados de este análisis se muestran en la Tabla 2. Se aprecia que la radiación extraterrestre es la entrada más relevante. El CCF en el instante t es la segunda variable de mayor importancia, seguido por la temperatura ambiente en ese mismo instante. La temperatura ambiente medida en los instantes $(t+1)$ y $(t-1)$ también se destaca, deduciendo que el gradiente de temperatura producido también aporta información a la ANN.

Tabla 2.- Análisis de sensibilidad de las entradas utilizadas en la ANN.

Entrada	Extraterrestre	CCF(t)	T(t)	CCF(t-1)	T(t-1)	CCF(t+1)	T(t+1)
Sensibilidad (%)	35.24	24.34	16.01	2.27	10.84	1.11	10.20

En el apartado 4.3, para la estación meteorológica de Juan Carlos I, se observa que la radiación extraterrestre tiene un PCC de 0.78, el CCF un valor de 0.85 y la temperatura ambiente un valor de 0.46. En comparación con el análisis de sensibilidad, se mantienen similares conclusiones, ya que la radiación extraterrestre y el CCF tienen un elevado peso, por encima de la temperatura ambiente. Se observa que la radiación extraterrestre toma mayor importancia en el análisis de sensibilidad de la ANN, sin embargo, hay que considerar que el PCC busca una correlación lineal, mientras que el análisis de sensibilidad indica la relevancia de las variables en la ANN, que tiene un comportamiento no lineal. De esta forma, la relación entre estudios no debe ser tomada estrictamente. Es de reseñar que la radiación extraterrestre es crucial en la ANN, al ser un indicativo del tiempo en la medida con respecto a la posición del sol.

Finalmente, en la Tabla 3 se resumen todas las características que describen la ANN seleccionada:

Tabla 3.- Características del modelo basado en ANN-MLPs usado en el modelo híbrido de predicción con un horizonte temporal diario.

Tipo de ANN	MLP
Entradas	T, CCF, E
Salida	GHI
Número de capas	3
Neuronas de entrada	7
Neuronas ocultas (una capa)	10
Neuronas de salida	1
Función activación capa oculta	Logística sigmoide
Función activación capa salida	lineal
Función objetivo	MAE
Algoritmo de aprendizaje	SCG - LM
Gradiente mínimo	1e-4
Iteraciones máximas	500

5.1.3 Hibridación de técnicas

El modelo híbrido calcula la predicción final como un valor ponderado de los dos modelos individuales, previamente descritos, según:

$$GHI = \omega_{SH}\, GHI_{SH} + (1 - \omega_{SH})\, GHI_{ANN}$$

El peso ω_{SH} se calcula minimizando el MAE del histórico de predicciones, mediante un simple barrido entre cero y uno. Este histórico se divide en función del tipo de día (soleado, nublado o encapotado), seleccionando el que coincide con el del momento de la predicción. Cada tipo de día se selecciona con el CCF, usando el método de los k vecinos más próximos (*k-Nearest Neighbours*, k-NN) para separar entre grupos de una forma sencilla y objetiva.

Los pesos se actualizan diariamente para cada tipo de día con el fin de optimizar la predicción final. La evaluación de los pesos se selecciona a partir de un histórico de hasta dos meses con la información más reciente. Hay dos razones por las que se limita este rango temporal a dos meses: (i) para evitar infravalorar a la ANN, que va mejorando su predicción conforme hay más datos disponibles; y (ii) para no considerar estaciones diferentes a la de la predicción, ya que las condiciones meteorológicas son diferentes y las entradas utilizadas dependen fuertemente de estas.

5.2 Resultados

Los resultados se analizan para dos situaciones diferentes. En primer lugar, se aplica el modelo de predicción para el nodo real de pruebas situado en la Escuela Politécnica Superior de la Universidad de Alcalá, que juega el

papel de un nodo de VPP. A continuación, se estudia el mismo caso para una agregación de instalaciones fotovoltaicas en forma de estaciones meteorológicas a nivel de suelo, formando una VPP.

Además del modelo híbrido desarrollado, se añaden otras técnicas de predicción a nivel comparativo: el modelo persistente, la predicción basada en días similares, la predicción basada en horas similares y la predicción basada en ANNs, para contrastar la efectividad del modelo propuesto.

5.2.1 Evaluación del modelo para un nodo real de VPP

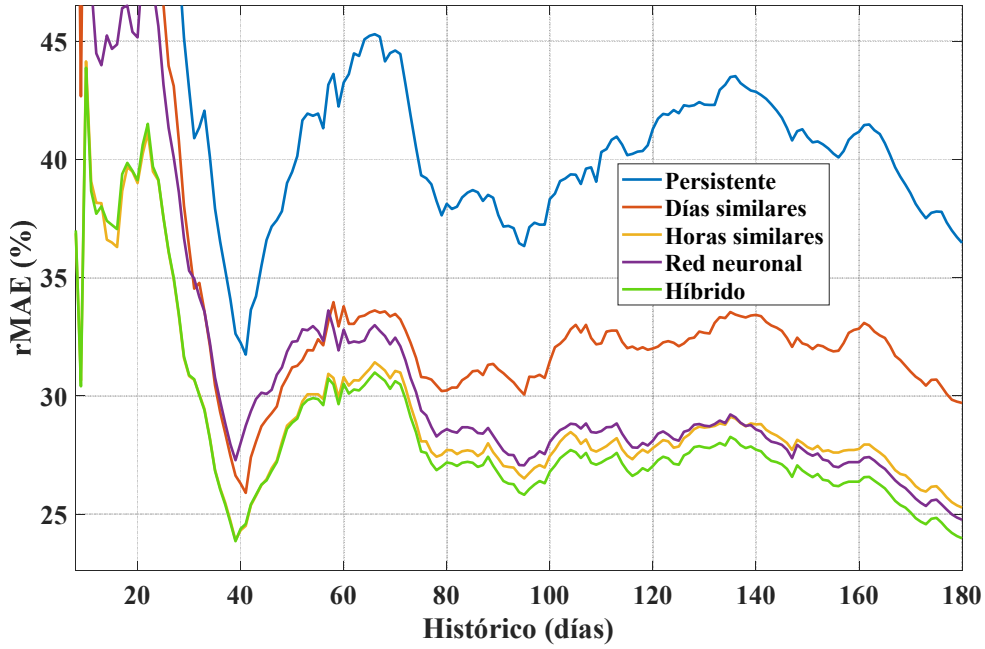
El nodo real de VPP recoge medidas tomadas en la instalación fotovoltaica de la Escuela Politécnica Superior de la Universidad de Alcalá. Las variables han sido recopiladas desde el 4 de diciembre de 2019 hasta el 31 de mayo de 2020, constituyendo un periodo total de 6 meses de histórico para la validación del modelo. El histórico recopilado se mantiene con respecto a los resultados ofrecidos en el artículo relacionado, [10], al corresponderse con una situación real de escasez de datos, justificando apropiadamente las características que debe tener el modelo de predicción. Más adelante, en el capítulo 9, se estudia el modelo propuesto para un histórico más actualizado, en contraste con el horizonte temporal intradiario.

La predicción inicial se realiza a la semana de histórico recopilado en la instalación, permitiendo predecir la GHI desde las etapas iniciales. Las medidas recopiladas se actualizan durante la noche, aumentando el histórico y volviendo a entrenar el modelo. De esta forma, se puede conocer la evolución de cada modelo de predicción y los pesos asociados a cada uno de ellos en el modelo híbrido.

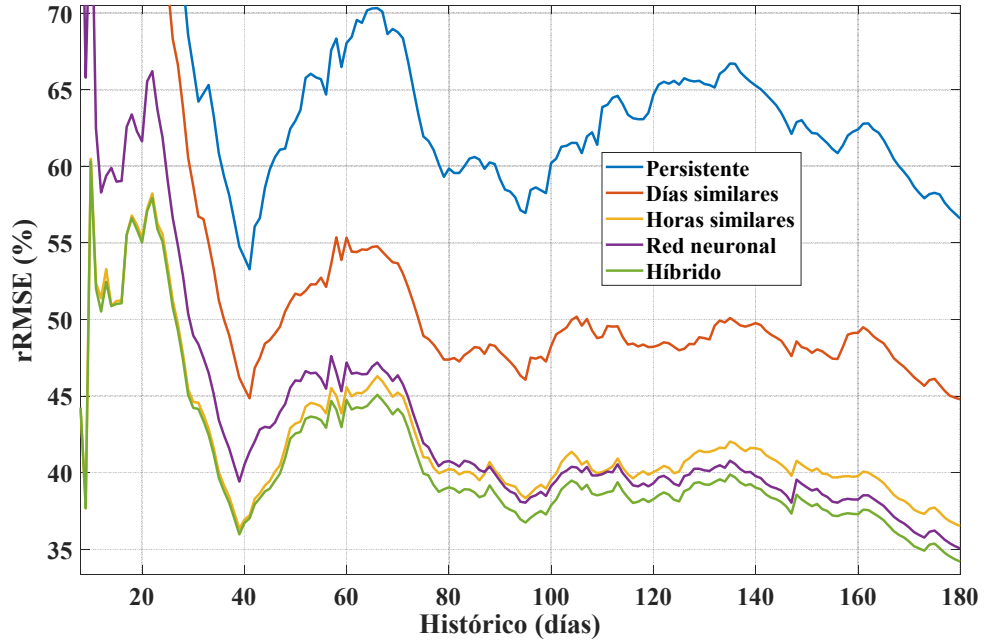
La medida del error se calcula según las expresiones del rMAE y el rRMSE del anexo I, ya que compara adecuadamente los resultados obtenidos a medida que aumenta el histórico de datos, evitando usar valores escalados que dependen de la estación del año, y ofreciendo una representación más intuitiva del error cometido.

Para validar los resultados del modelo de predicción, se realiza una comparación entre los siguientes modelos: (i) el modelo persistente, que asume que la predicción de GHI son las medidas recopiladas el día anterior [69]; (ii) predicción basada en días similares; (iii) predicción basada en horas similares; (iv) predicción basada en ANNs; y (v) modelo híbrido propuesto.

En la Figura 20 se muestran los errores de predicción cometidos en términos del rMAE y el rRMSE, a lo largo del histórico de datos recopilados. Para mejorar la visualización del gráfico, errores elevados de algunos modelos de predicción han sido eliminados de la figura. Estos errores descartados alcanzan un valor máximo de $rMAE\ (\%) = [96.8\ 66.2\ 69.7]$ y $rRMSE\ (\%) = [126.2\ 116.4\ 85.1]$ para el modelo persistente, la predicción basada en días similares y la predicción basada en ANNs, respectivamente.



(a)



(b)

Figura 20.- Error de predicción de los modelos empleados en el nodo real de pruebas situado en la Universidad de Alcalá, en términos del (a) rMAE y (b) rRMSE.

Se observa que el modelo persistente, por su simplicidad, es el que muestra el mayor error en la predicción. El modelo basado en días similares mejora considerablemente la predicción persistente, alcanzando un nivel de precisión similar al de las ANNs para históricos reducidos en términos del rMAE. Sin embargo, conforme el histórico aumenta, la ANN muestra una precisión cada vez mayor. El modelo basado en horas similares tiene una gran precisión ante históricos escasos al compararlo con los otros modelos, y la ANN va mostrando un mejor rendimiento conforme el histórico aumenta. Finalmente, el modelo híbrido es el que muestra la mayor precisión,

independientemente de la longitud del histórico, ya que combina las ventajas de los modelos basados en horas similares y ANNs.

La predicción híbrida calcula los pesos óptimos en función del histórico de predicciones recopilado, que tiene en cuenta el tipo de día. La evolución de estos pesos se muestra en la Figura 21. Se observa que, conforme el histórico aumenta, el modelo basado en ANNs va tomando más peso en la predicción, siendo coherente con lo observado en la Figura 20 y mejorando gradualmente con respecto al modelo basado en horas similares. Este efecto es más significativo en días nublados y encapotados, siendo el rendimiento del modelo de horas similares superior a la ANN en días de cielo despejado para la totalidad del histórico.

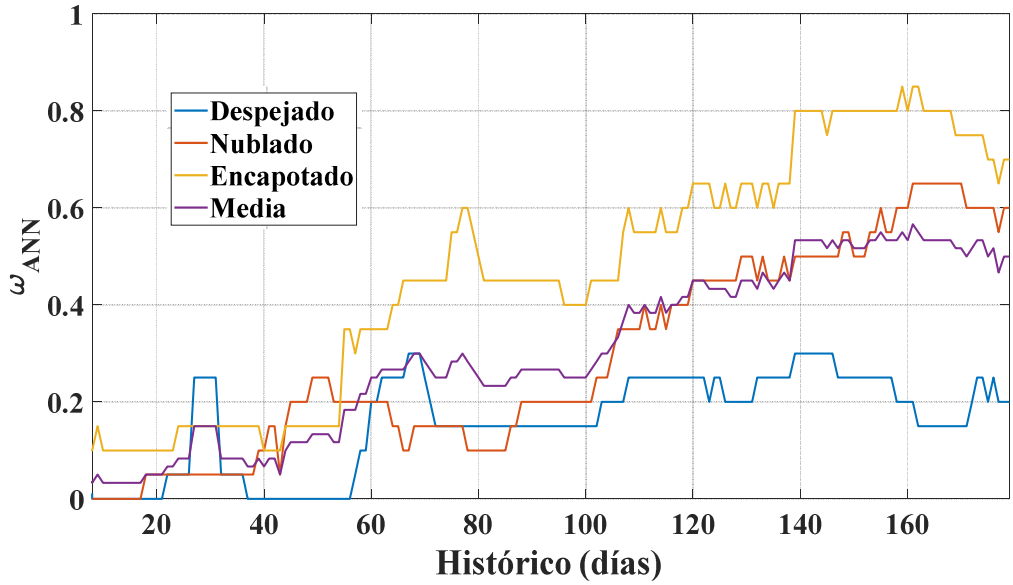


Figura 21.- Pesos usados en la ANN para el modelo híbrido en función del histórico del nodo real de pruebas situado en la Universidad de Alcalá.

La evolución de los pesos se hace más notable al analizar el valor medio de los tres tipos de día considerados. El modelo basado en ANNs va aumentando su peso a lo largo del histórico recopilado desde un valor de $\omega_{ANN} = 0.04$ hasta $\omega_{ANN} = 0.51$ cuando se alcanzan los seis meses de histórico.

Finalmente, en la Figura 22, se muestra la predicción de GHI para cada modelo considerado, del 14 al 16 de marzo de 2020. Estos días han sido seleccionados para mostrar los tres tipos de día considerados en la predicción del modelo híbrido. Para estos días seleccionados se emplea la mitad del histórico total aproximadamente (102 días), momento en el que los modelos basados en horas similares y ANNs ofrecen una precisión bastante similar.

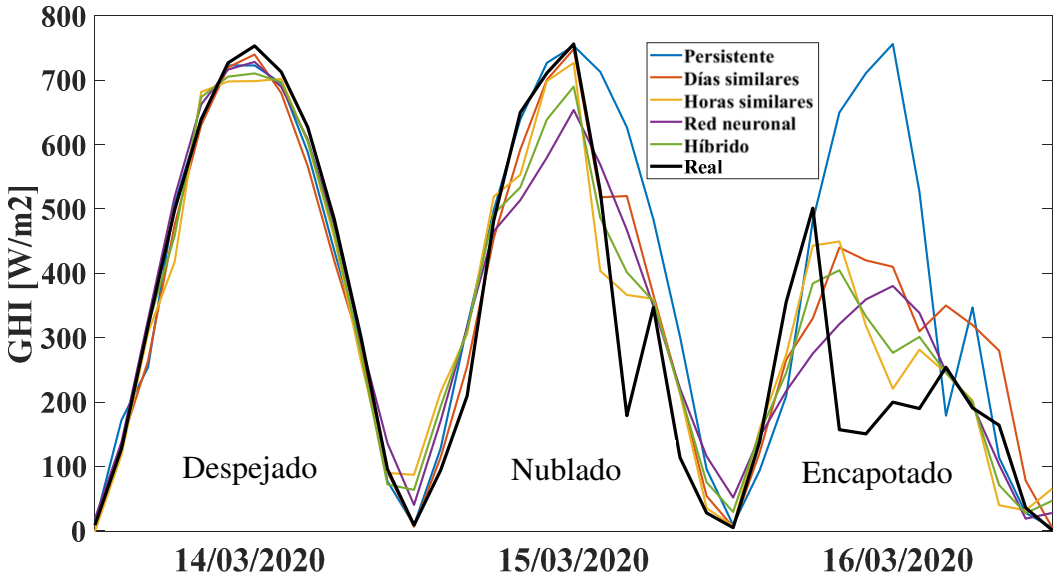


Figura 22.- Predicción de los valores de GHI en el nodo real de pruebas situado en la Universidad de Alcalá para los diferentes modelos en tres días consecutivos, mostrando diferentes tipos de día.

Se observa que el error producido en todos los modelos de predicción aumenta con la presencia de nubes. Esto es debido a que los días nublados son más complicados de predecir, ya que las nubes tienen un gran impacto en la GHI medida a nivel de suelo, siendo difícil predecir la situación exacta de las nubes en predicciones de un horizonte temporal diario. Sin embargo, se observa que la estrategia híbrida propuesta alcanza una mejora considerable en la predicción, particularmente en días nublados.

5.2.2 Evaluación del modelo para una VPP emulada

Una vez mostrado el potencial del modelo de predicción propuesto para una instalación fotovoltaica de un nodo real de pruebas, se realiza este estudio para un conjunto de plantas fotovoltaicas agrupadas. Al no disponer de otras instalaciones fotovoltaicas, se añaden las medidas de seis estaciones meteorológicas en la Comunidad de Madrid, situadas en las localizaciones de la Figura 23. Estas estaciones meteorológicas ofrecen toda la información necesaria para la predicción desde su web [42], utilizando además los mapas de NWP de la AEMET.

La predicción de la VPP es la suma de cada uno de los nodos considerados en el apartado, alcanzando potencias bastante superiores a los nodos individuales. Por lo tanto, para este caso resulta aún más conveniente representar el rMAE y el rRMSE en función del histórico recopilado, al mostrar el error porcentual y permitir comparar diferentes escalas.

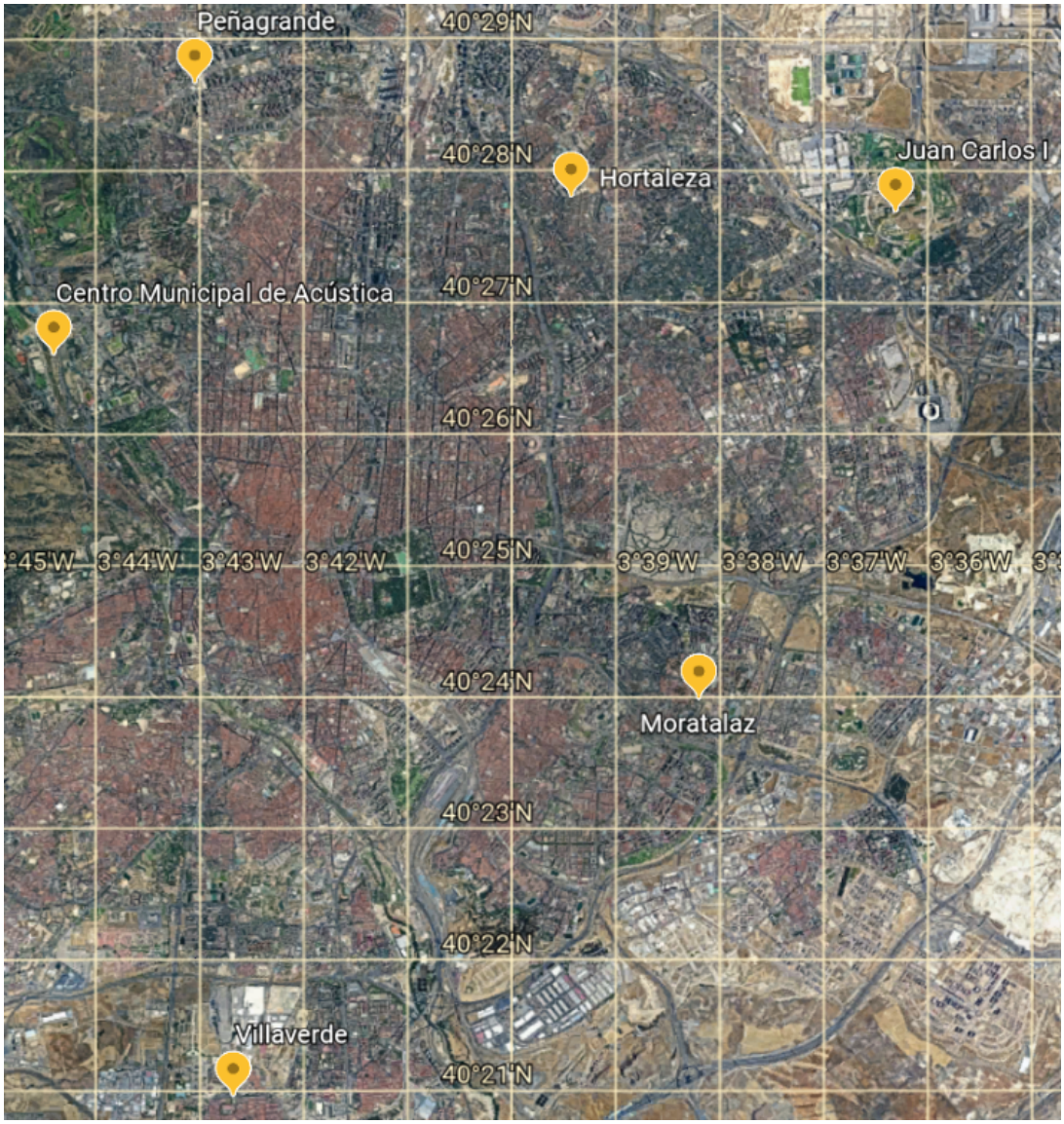
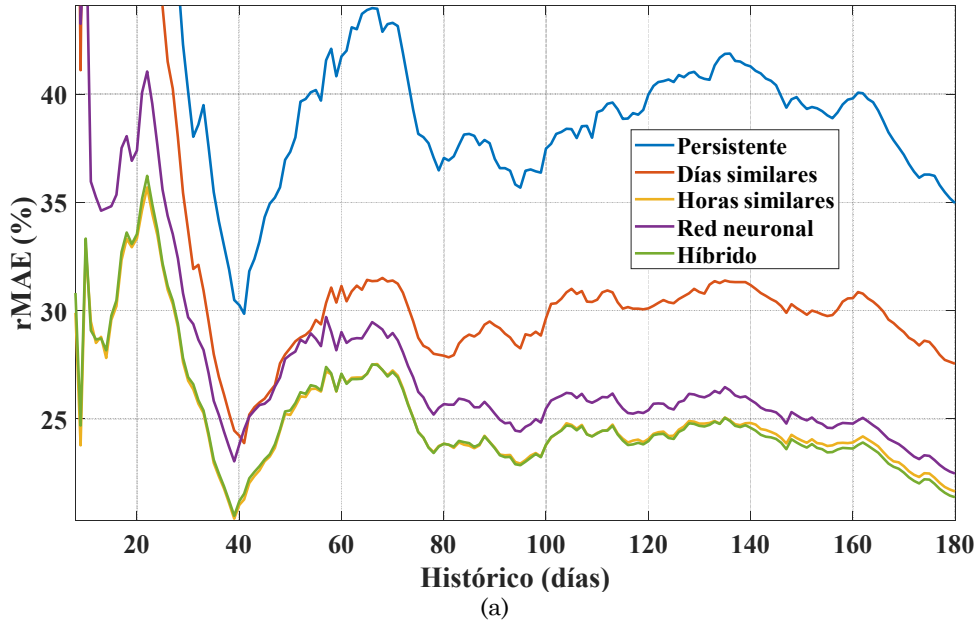
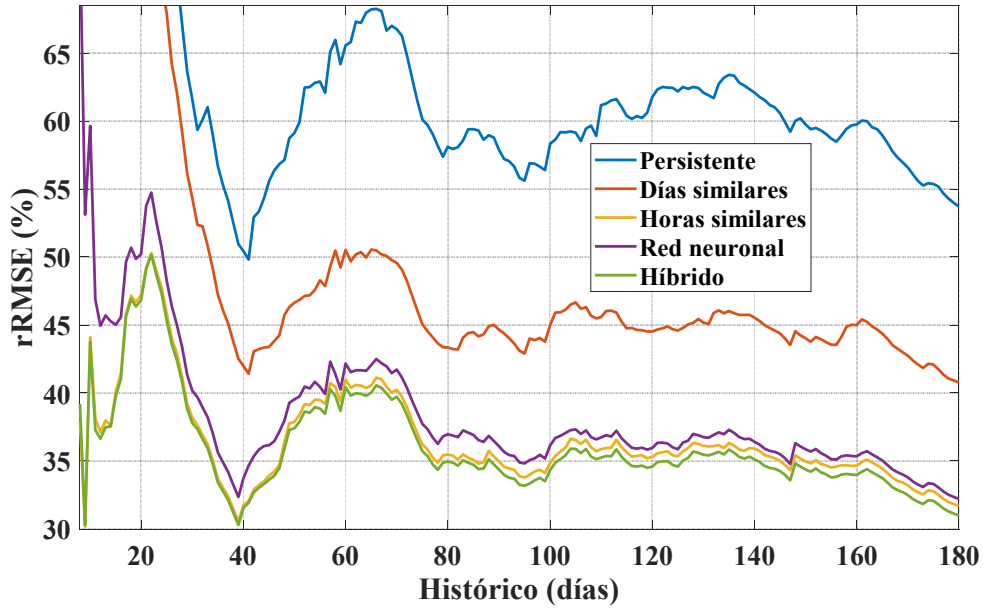


Figura 23.- Localización de las estaciones a nivel de suelo en la Comunidad de Madrid empleadas en el presente estudio.

La Figura 24 muestra la medida del error cometido en la predicción de GHI para la VPP, de acuerdo al rMAE y el rRMSE. Al igual que en la Figura 20, se ocultan errores excesivamente elevados. Los máximos alcanzados en estos errores son de $rMAE\ (\%) = [93.2\ 63.5\ 62.8]$ y $rRMSE\ (\%) = [119.7\ 110.7\ 74.5]$ para el método persistente, el modelo basado en días similares y el modelo basado en ANNs, respectivamente. Al igual que en el nodo real de pruebas, se observa que el modelo persistente muestra el peor rendimiento, seguido del modelo basado en días similares. El modelo basado en ANNs mejora conforme el histórico recopilado aumenta, sin embargo, no llega a mejorar la predicción del modelo basado en horas similares en todo el histórico. El modelo híbrido alcanza la mayor precisión nuevamente para todo el histórico contemplado.



(a)



(b)

Figura 24.- Error de predicción de los modelos empleados en la VPP emulada, en términos del (a) rMAE y (b) rRMSE.

La Tabla 4 muestra los resultados para cada nodo particular y para el entorno de VPP en los 180 días de histórico, Todos los nodos tienen una precisión similar al nodo real de pruebas de la Universidad de Alcalá. El modelo basado en horas similares alcanza medidas del error superiores a los otros dos en todos los nodos, sin embargo, se observa que este error se reduce en mayor proporción que el modelo basado en ANNs para el entorno de VPP. Esto es debido a que el modelo basado en horas similares compensa en mayor medida las desviaciones entre predicciones individuales, al utilizar el histórico de medidas de GHI de cada nodo para la predicción. Esto hace que ciertos errores entre nodos se compensen en mayor medida cuando se calcula la GHI total de la VPP, por lo que gana especial relevancia en este tipo de entornos.

Tabla 4.- Medidas del error del rMAE y el rRMSE tras 180 días de histórico.

Modelo	Nodos	Universidad	Ctro. Mpal. de Acústica	Hortaleza	Juan Carlos I	Moratalaz	Peñagrande	Villaverde	Media	**VPP**
Horas similares	rMAE (%)	25.27	25.40	25.71	24.28	25.00	24.63	24.37	24.95	**21.64**
	rRMSE (%)	36.51	37.26	39.15	36.60	36.45	36.52	37.11	37.09	**31.69**
ANNs	rMAE (%)	24.76	24.63	25.85	24.09	24.46	24.48	24.04	24.06	**22.48**
	rRMSE (%)	35.05	35.15	37.18	34.71	34.76	35.38	34.74	34.74	**32.21**
Híbrido	rMAE (%)	23.98	23.92	24.71	23.18	23.67	23.53	23.27	23.75	**21.37**
	rRMSE (%)	34.21	34.57	36.90	34.38	34.12	34.42	34.59	34.74	**30.99**

El modelo basado en horas similares reduce el rMAE del entorno VPP en un 13.27%, con respecto a la media de los errores de los nodos, y un 14.56% para el rRMSE. Esta reducción es de un 6.57% en el rMAE y un 7.28% en el rRMSE para el modelo basado en ANNs, y de un 10.02% en el rMAE y un 10.79% en el rRMSE para el modelo híbrido.

La Figura 25 muestra los pesos medios de todos los nodos considerados en la VPP en función del histórico de datos y el tipo de día. Estos pesos no se aplican directamente a la predicción, ya que se calculan individualmente, sin embargo, proporcionan cierta idea sobre cómo van evolucionando conjuntamente a lo largo del histórico recopilado. La evolución media de los pesos es bastante similar a la de un único nodo, dispuesto en la Figura 21, tomando inicialmente un valor de $\omega_{ANN} = 0.10$ hasta $\omega_{ANN} = 0.60$ para todo el histórico.

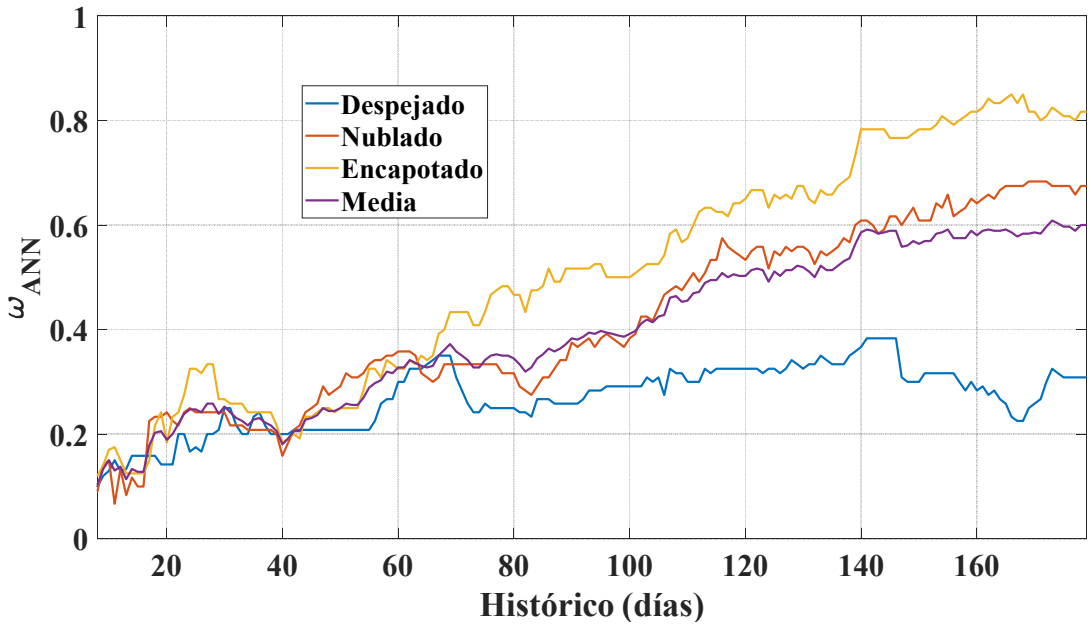


Figura 25.- Pesos usados en la ANN para el modelo híbrido en función del histórico de la VPP emulada.

Finalmente, en la Figura 26, se observa la predicción de GHI en la VPP completa para cada uno de los modelos evaluados, considerando los tres tipos de día estudiados, desde el 14 al 16 de marzo de 2020. Se observa, al igual que para un único nodo, que el error aumenta en función de la nubosidad, sin embargo, el modelo híbrido presenta los mejores resultados en todos los tipos de día considerados.

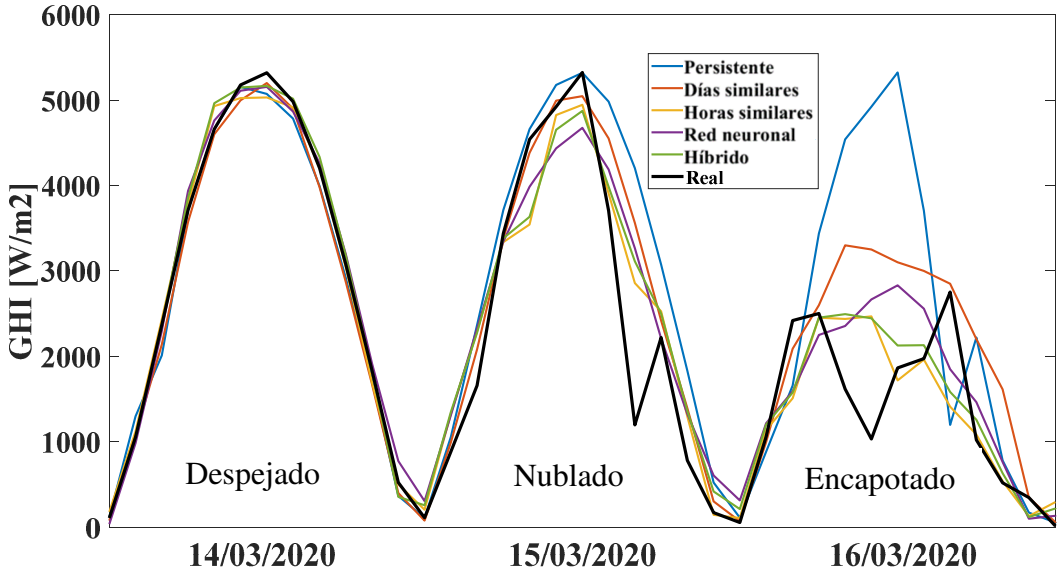


Figura 26.- Predicción de los valores de GHI en la VPP emulada para los diferentes modelos en tres días consecutivos, mostrando diferentes tipos de día.

5.3 Conclusiones

En este capítulo se ha presentado un modelo de predicción híbrido formado por dos modelos individuales: uno basado en horas similares y otro basado en ANNs. Las salidas de ambos modelos se ponderan dinámicamente para obtener una precisión óptima en la estrategia, de acuerdo al tipo de día y el error del histórico. Los modelos de predicción usan como variables de entrada mapas de NWP de la temperatura ambiente y la nubosidad, junto a la radiación extraterrestre. La precisión del modelo se ha calculado para un nodo real de pruebas situado en la Escuela Politécnica Superior de la Universidad de Alcalá, junto a un grupo de estaciones meteorológicas a nivel de suelo situadas en la Comunidad de Madrid que emulan una VPP.

En un nodo real de pruebas, tanto el modelo basado en horas similares como el modelo híbrido han demostrado una mayor precisión en históricos reducidos con respecto a otras técnicas tradicionales empleadas: el modelo persistente, el modelo basado en días similares y el modelo basado en ANNs. Por otra parte, conforme el histórico recopilado aumenta, el modelo basado en ANNs adquiere mayor precisión hasta mejorar al modelo basado en horas similares.

En la VPP emulada, el modelo basado en horas similares obtiene una mayor precisión que el modelo basado en ANNs, para un total de siete nodos considerados y 180 días de histórico. Esto es debido a que su error se reduce en mayor proporción al considerar un entorno VPP, debido a la forma que tiene de compensar los errores individuales de los nodos que la componen. Sin embargo, la predicción híbrida genera el mínimo error de todos los modelos considerados.

El modelo basado en horas similares reduce el rMAE de la VPP en un 13.27%, con respecto a la media de los errores de los nodos, y un 14.56% para el rRMSE. Esta reducción es de un 6.57% en el rMAE y un 7.28% en el rRMSE para el modelo basado en ANNs, y de un 10.02% en el rMAE y un 10.79% en el rRMSE para el modelo híbrido. En un histórico de un mes, el modelo basado en horas similares mejora al modelo basado en ANNs en un 16.32%, reduciendo esta mejora a un 9.07% en un histórico de dos meses.

6 Estrategia de predicción intradía

Una vez realizada la predicción con un horizonte temporal diario, al disponer de nuevas medidas en la instalación, es importante actualizar la predicción para un horizonte temporal intradiario, optimizando la precisión a lo largo del día. De esta forma, se logra ir minimizando el error cometido, realizando correcciones de la predicción en plazos menores de tiempo.

Este efecto se puede asemejar a la relación entre el mercado diario de la energía, donde se gestiona un 80% de la energía total, y el mercado intradiario, que realiza ajustes en un menor plazo de tiempo para cubrir eficientemente el 20% de energía restante.

Este capítulo describe la estrategia de predicción de irradiancia con un horizonte temporal intradiario. El apartado 6.1 describe de forma general el proceso de predicción. El apartado 6.2 muestra los resultados del nodo real de pruebas situado en la Universidad de Alcalá. Finalmente, el apartado 6.3 describe las conclusiones más relevantes.

6.1 Descripción del modelo

Las predicciones de irradiancia intradiarias actualizan las predicciones del día anterior y permiten tomar decisiones con una mayor fiabilidad. Estas predicciones se van actualizando en cuestión de minutos, lo que resulta de especial utilidad para el control en tiempo real de la instalación. Otro aspecto importante es el plazo intradiario de lanzamiento, que se fija en 15 minutos para proporcionar una solución de compromiso entre la carga computacional, propia de este tipo de modelos que usan bases de datos extensas, y una mayor exactitud en la predicción. De igual forma, se ha optado por una resolución temporal de 15 minutos, guardando coherencia con el plazo intradiario de lanzamiento y adaptándose a la actuación del BEMS de la VPP, exponiendo este caso de estudio en el apartado 9.2.

El uso de series temporales resulta bastante adecuado en modelos de predicción con un horizonte temporal intradiario, ya que las condiciones meteorológicas no varían excesivamente entre los datos más recientes y el momento de la predicción. Al ir realizando nuevas predicciones a lo largo del día para un instante determinado, se actualiza el plazo intradiario de lanzamiento y se reduce el tiempo de espera, mejorando la precisión. La recopilación de un histórico extenso y actualizado es crucial para minimizar el error y generalizar la predicción. Para no excluir a nuevas instalaciones o a nodos recién integrados en un entorno VPP, las medidas de irradiancia se complementan con información satélite de libre acceso proporcionada por el CAMS [7], usando un total de 5 años de histórico para la predicción.

La Figura 27 muestra el esquema general del modelo de predicción con un horizonte temporal intradiario, distinguiendo: (i) una fase previa de

adquisición de datos desde las instalaciones y las medidas satélite; (ii) un preprocesado de la información utilizada como entrada en el modelo; (iii) el modelo de predicción de irradiancia basado en series temporales mediante RNN-LSTMs; (iv) una fase de conversión a potencia, definiendo intervalos de incertidumbre en la predicción; y (v) el uso de la predicción por parte del EMS de la VPP.

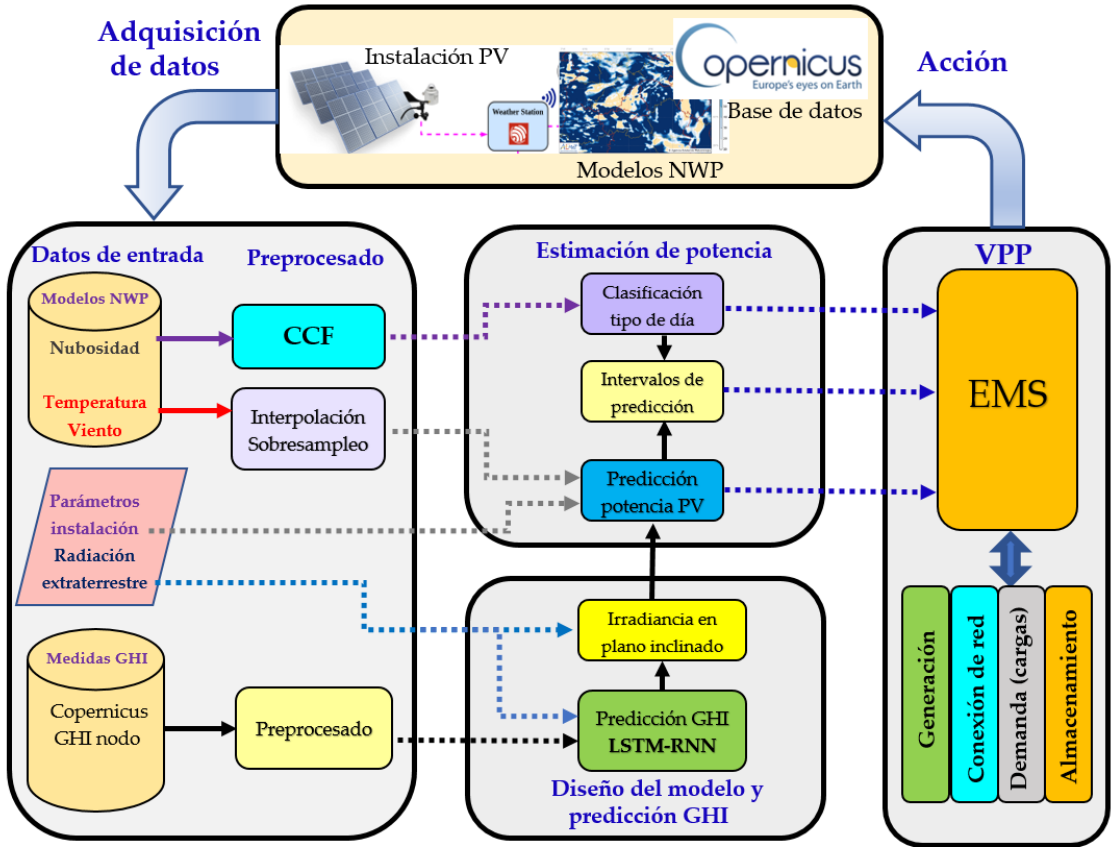


Figura 27.- Esquema general del modelo utilizado en predicciones intradiarias, distinguiendo: (i) adquisición de datos; (ii) variables de entrada y preprocesado; (iii) predicción de irradiancia; (iv) estimación de potencia; y (v) toma de decisiones y actuación del EMS.

Este capítulo se centra en la predicción intradiaria de irradiancia, posterior a la fase de preprocesado del capítulo 4. Seguidamente, en el capítulo 7, se muestran los resultados de estimación de potencia a partir de estas predicciones para el nodo real y su VPP emulada. Así mismo, en capítulos posteriores, se describe la generación de intervalos en función de los resultados obtenidos en las predicciones, junto a un caso de uso de gestión de baterías.

6.2 Predicción de irradiancia y resultados

Una vez procesadas las variables de entrada, se predice la GHI para estimar posteriormente la potencia generada por los paneles, enviando la información al EMS de la VPP. En este apartado se describe la estrategia de predicción de irradiancia usada, mostrando los resultados extraídos y

obteniendo unas conclusiones finales en relación a las variables temporales de la Figura 4.

La Figura 28 muestra la etapa de predicción de irradiancia y conversión a potencia de la Figura 27 con un mayor nivel de detalle. En primer lugar, las variables de entrada procedentes del preprocesado se utilizan para la predicción, tomando 5 años de GHI y radiación extraterrestre como entradas de una RNN-LSTM, encargada de predecir la GHI intradiaria. El proceso se actualiza en función del plazo intradiario de lanzamiento, que es de 15 minutos. El primer lanzamiento se realiza al amanecer y, al recibir nuevas medidas de la instalación en tiempo real, se obtienen nuevas predicciones que mejoran la precisión del modelo.

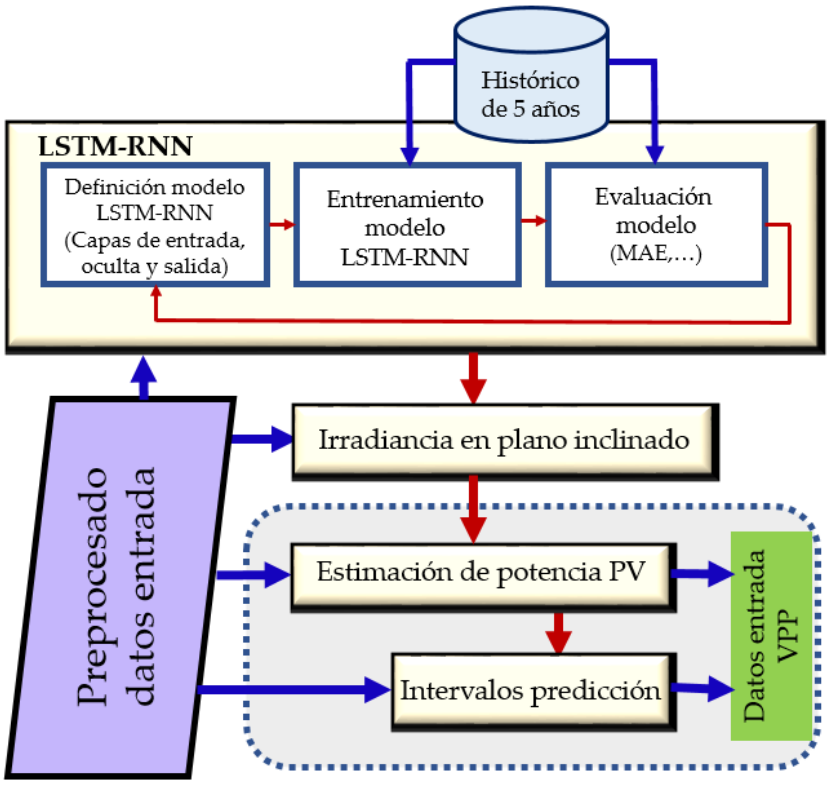


Figura 28.- Diagrama de flujo del modelo de predicción basado en RNN-LSTMs.

Para evitar una carga computacional excesiva, la RNN-LSTM no se reentrena cuando se reciben nuevas medidas en la instalación el mismo día. Por otra parte, para evitar perder precisión en el modelo, la red se entrena de forma diaria durante la noche, llegando así a una solución de compromiso entre eficiencia y carga computacional. Es de reseñar que, a la hora de diseñar los parámetros relacionados con la convergencia de la red, como el número de epochs o la selección del mini-batch, también se ha buscado esta solución de compromiso, tardando unos 110 segundos de media en reentrenar la red. Estrictamente es posible un reentrenamiento con cada nueva medida recibida, ya que la resolución temporal es de 15 minutos.

Una definición general de las RNN-LSTMs y su justificación dentro del modelo de predicción propuesto se aborda en el anexo IV. El proceso de predicción entrena la red al anochecer con 5 años de histórico de irradiancia, obteniendo la predicción de todo el día, actualizando la predicción cada 15 minutos según el plazo intradiario de lanzamiento.

A partir del histórico recopilado, se optimizan los hiperparámetros de la RNN-LSTM al realizar un barrido de cada uno de ellos de forma simultánea. Para reducir la carga computacional, tan solo se optimizan los hiperparámetros más relevantes de la RNN-LSTM: unidades en la capa oculta $[20, 50, 100]$, tamaño del mini-batch $[128, 256, 512]$, factor de regularización $[10^{-2}, 10^{-3}, 10^{-4}]$, índice de aprendizaje $[0.01, 0.05, 0.1]$ y el número de epochs $[30, 50, 70, 90]$. El resto, como el gradiente límite o las variables usadas en el optimizador Adam, se seleccionan según valores recomendados, al ser menos relevantes en la optimización y poder aumentar así la eficiencia en la convergencia de parámetros críticos.

La cuantificación del error del entrenamiento se realiza con el RMSE, con el fin de evitar errores muy elevados o atípicos que pueden afectar negativamente en tareas de gestión del EMS. La arquitectura de la RNN-LSTM se compone de dos unidades en la capa de entrada, una capa oculta con 50 bloques y una unidad en la capa de salida. En la Tabla 5 se muestran los parámetros seleccionados en la RNN-LSTM.

Tabla 5.- Parámetros seleccionados en la RNN-LSTM.

Número de entradas	2 (GHI, extraterrestre)
Unidades en capa oculta	50
Número de salidas	1
Tamaño del mini-batch	256
Factor de regularización	5.10^{-4}
Optimizador	Adam ($\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$)
Índice de aprendizaje inicial	0.01
Programa del índice de aprendizaje	Por partes (periodos)
Reducción del aprendizaje	0.5 cada 20 epochs
Epochs	70
Gradiente límite	1

Los resultados a nivel de test proceden de un nodo real de VPP dentro de un entorno de pruebas situado en la Escuela Politécnica Superior de la Universidad de Alcalá en el que se recogen medidas de GHI. El histórico de medidas usado para el test aborda todo un año, iniciando el 1 de junio de 2020 y finalizando el 31 de mayo de 2021. El proceso de predicción se ha realizado con una resolución temporal de 15 minutos, desde el amanecer hasta el anochecer, actualizando la predicción en función del plazo intradiario de lanzamiento.

En la Figura 29 se muestra de forma gráfica el conjunto de predicciones intradiarias realizadas a lo largo de un día en función del tiempo de espera y el plazo intradiario de lanzamiento. Cuando se fija el

plazo intradiario de lanzamiento, en la predicción se genera un vector de predicciones en función del tiempo de espera. Por otra parte, si ambos parámetros se seleccionan, se obtiene un único punto que se muestra como un rombo rojo en la Figura 29.

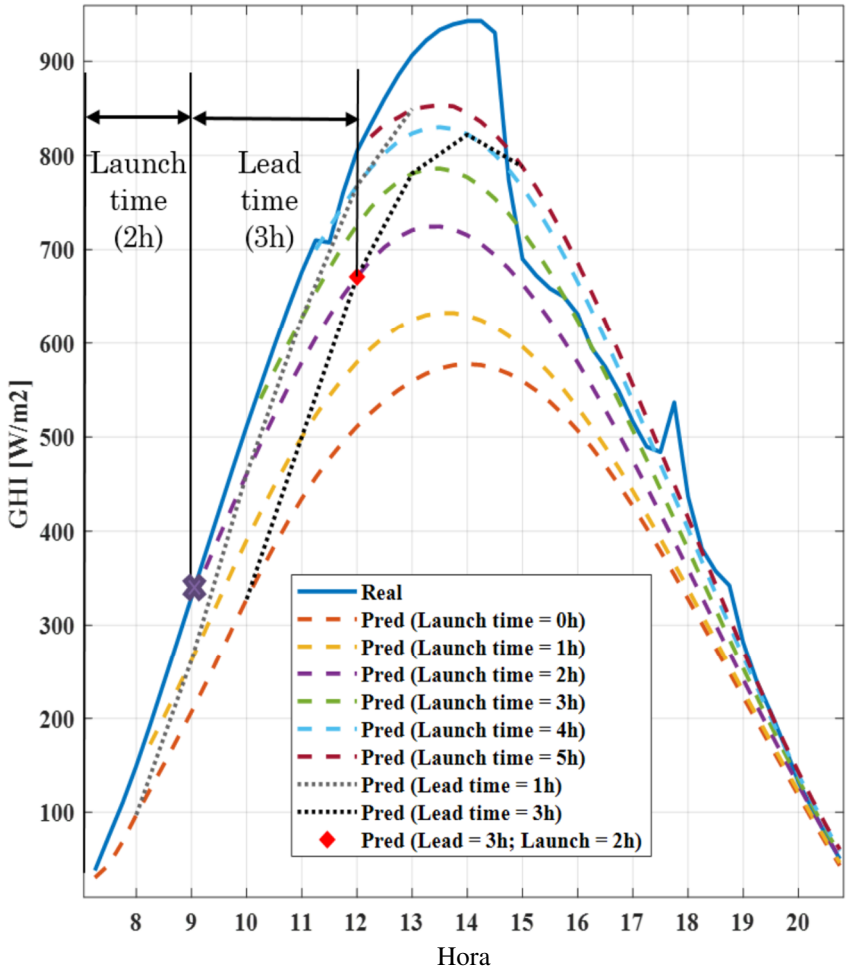


Figura 29.- Medidas reales y predicciones intradiarias para un día concreto. Las líneas discontinuas son las predicciones a diferentes plazos intradiarios de lanzamiento, mientras que las líneas de puntos son las predicciones a diferentes tiempos de espera. Al seleccionar ambos parámetros se define un único punto mostrado como un rombo rojo.

En la Figura 30 se muestran los errores cometidos al predecir la GHI en el nodo real de pruebas de la Universidad de Alcalá en función del plazo intradiario de lanzamiento y el tiempo de espera, según las definiciones del anexo I. Considerando errores escalados, en este caso el MAE y el RMSE, se observa un elevado error en pequeños plazos intradiarios de lanzamiento y tiempos de espera medios, lo que resulta lógico ya que la irradiancia es mayor. Sin embargo, al aumentar el plazo intradiario de lanzamiento, este error se reduce notablemente. Por otra parte, como es lógico pensar, cuando la GHI medida es menor, los errores escalados son pequeños y los errores porcentuales, en este caso el rMAE y el rRMSE, elevados. Sin embargo, en plazos intradiarios de lanzamiento de una hora y media o menos, el error porcentual es elevado independientemente del tiempo de espera.

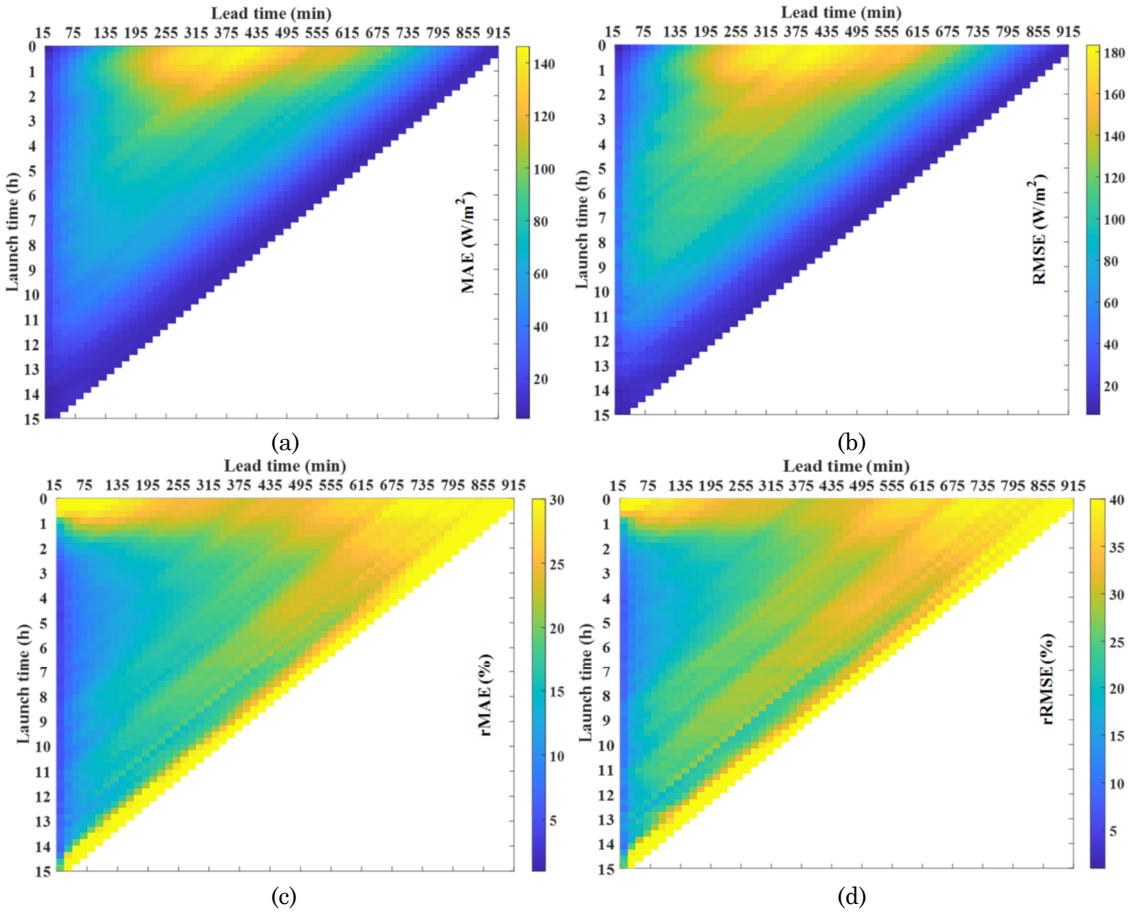


Figura 30.- Matrices de error entre medidas reales y predichas de GHI en función del plazo intradiario de lanzamiento y el tiempo de espera: (a) MAE; (b) RMSE; (c) rMAE; y (d) rRMSE.

Por último, se comprueba la precisión del modelo con respecto a otros estudios procedentes de la literatura en la Tabla 6. Es de reseñar que este análisis no debe ser considerado de forma estricta, ya que el uso de distintas bases de datos influye de diferente forma en los errores cometidos, sin embargo, ciertas conclusiones se pueden obtener de la tabla de forma generalizada.

Tabla 6.- Estudio comparativo entre los resultados de predicción intradiaria de GHI en la tesis y los desarrollados por otros autores en la literatura.

Modelo [Artículo]	Error	Horizonte temporal	Resolución temporal	Entradas	Resultados de la tesis
Smart pers. [102]	$rMAE = [8 - 18]$ %	6 h	15 min	GHI, GHI despejado, índice de nube, máxima altura nube, …	$rMAE = [4.17 - 17.73]$ %
CIAD Cast [102]	$rMAE = [11 - 20]$ %				
Satélite [102]	$rMAE = [10.5 - 19.5]$ %				
WRF-Solar [102]	$rMAE = [12 - 18]$ %				
SVM-Radial [102]	$rMAE = [7.5 - 15.5]$ %				
ARIMA [103]	$MAE = 71.48\ W/m^2$	1 h	1 h	GHI, GHI despejado, tipo nube, temperatura, humedad, precipitación, viento, …	$MAE = 41.88\ W/m^2$
RNN [103]	$MAE = 41.83\ W/m^2$				
LSTM [103]	$MAE = 31.86\ W/m^2$				
CNN-LSTM [67]	$MAE = 41.88\ W/m^2$	1 h	1 h	GHI, temperatura, viento, precipitación, humedad, acimut, …	$MAE = 41.88\ W/m^2$
CNN-LSTM [67]	$RMSE = 78.17\ W/m^2$				$RMSE = 72.54\ W/m^2$
CNN-LSTM [67]	$rMAE = 10.58$ %				$rMAE = 8.72$ %
CNN-LSTM [67]	$rRMSE = 19.75$ %				$rRMSE = 15.1$ %
LSTM [77]	$RMSE = [77 - 143]\ W/m^2$	8 h	1 h	GHI, humedad, nubosidad, temperatura, extraterrestre	$RMSE = [72 - 124]\ W/m^2$
LSTM [77]	$rRMSE = [18.4 - 33]$ %				$rRMSE = [15.1 - 29.2]$ %
GRU [104]	$RMSE = 67.29\ W/m^2$	1 h	1 h	GHI, cenit, humedad, temperatura	$RMSE = 72.54\ W/m^2$
LSTM [104]	$RMSE = 66.57\ W/m^2$				
GRU [105]	$RMSE = 58\ W/m^2$	30 min	1 min	GHI	$RMSE = 55.78\ W/m^2$
LSTM [105]	$RMSE = 55.29\ W/m^2$				

En primer lugar, teniendo en cuenta técnicas más tradicionales, en [102] se obtiene un rMAE mayor para un tiempo de espera de 15 minutos, sin embargo, esta diferencia se va reduciendo progresivamente hasta obtener similares rMAE para un tiempo de espera de 6 horas. Al comparar los resultados con el modelo autorregresivo integrado de media móvil (*AutoRegressive Integrated Moving Average*, ARIMA) de [103], también se observa un mayor error en tiempos de espera pequeños, sin embargo, la RNN-LSTM diseñada en [103] genera una mayor precisión que el modelo aquí desarrollado, observando una precisión similar con respecto a la RNN simple. En [103] se utilizan otras variables de alta correlación con la GHI, como la nubosidad, la temperatura y la humedad, justificando así esta precisión. Finalmente, comparando la estrategia presentada en la tesis con otras técnicas de aprendizaje profundo, en este caso modelos tipo GRU, RNN-LSTM y CNN-LSTM de [66,76,104,105], se observa una precisión entre modelos bastante similar.

A nivel de resumen, en tiempos de espera pequeños, el modelo de predicción presentado en esta tesis muestra una precisión superior con respecto a otros modelos tradicionales de la literatura. Sin embargo, conforme aumenta el tiempo de espera, la eficiencia del modelo propuesto se reduce progresivamente hasta que el error obtenido es similar a modelos tradicionales. Por otra parte, la RNN-LSTM de la tesis ofrece errores similares con respecto a otros modelos de aprendizaje profundo encontrados en la literatura. También es de reseñar que aportar otras variables meteorológicas de alta correlación con la GHI mejora la precisión del modelo, siendo una alternativa bastante recurrida, sin embargo, si estos recursos no están disponibles, estos modelos resultan irrealizables.

6.3 Conclusiones

Este capítulo presenta un modelo basado en RNN-LSTMs para predecir la GHI con un horizonte temporal intradiario y una resolución temporal de 15 minutos. Estos resultados se analizan en función del tiempo de espera y el plazo intradiario de lanzamiento, que muestran zonas con un error muy reducido. Por otra parte, cuando el plazo intradiario de lanzamiento es inferior o igual a una hora y media, el error porcentual cometido es elevado de forma independiente al tiempo de espera. Además, el error escalado se reduce en zonas de alta irradiancia al aumentar el plazo intradiario de lanzamiento, concluyendo que se necesita disponer de información lo más actualizada posible en la instalación para generar una precisión óptima.

En comparación con estudios procedentes de la literatura se alcanzan errores similares con respecto a otros modelos de aprendizaje profundo, mejorando a los más tradicionales. Por último, los modelos exógenos generan una mayor precisión que el modelo propuesto, sin embargo, se destaca la limitación existente a la hora de disponer de esta información.

7 Estimación de potencia

Una vez realizadas las predicciones de GHI a diferentes horizontes temporales se estima la potencia fotovoltaica, mostrando en este capítulo el método analítico usado, apoyado por recursos bibliográficos. Este método se explica en el apartado 7.1. El objetivo es conocer la potencia fotovoltaica mediante la predicción de GHI, la temperatura ambiente obtenida de los mapas de predicción de NWP e información de la propia instalación. El apartado 7.2 mide la precisión del método con medidas reales de la instalación. El apartado 7.3 calcula la precisión de los resultados con respecto a las predicciones de GHI, para un horizonte temporal intradiario en el apartado 7.3.1 y un horizonte temporal de un día en el apartado 7.3.2. Finalmente, el apartado 7.4 describe las conclusiones del capítulo.

7.1 Método analítico

El cálculo de la potencia del panel fotovoltaico a partir de la GHI puede desglosarse en 4 pasos diferentes: (i) determinar la irradiancia orientada al plano de los paneles; (ii) calcular las pérdidas angulares, por suciedad y espectrales para conocer la irradiancia efectiva en el plano de los paneles; (iii) estimar la temperatura de los paneles; y (iv) estimar la potencia final de los paneles en función de los parámetros anteriores.

7.1.1 Irradiancia en el plano de los paneles

Los pasos a seguir para calcular la irradiancia en el plano de los paneles son los siguientes: (a) calcular la irradiancia difusa horizontal y la irradiancia directa horizontal a partir de la GHI estimada; (b) a partir de estas, obtener las tres componentes de irradiancia en el plano de los paneles: la irradiancia difusa, la irradiancia directa y el albedo; y (c) finalmente, se calcula la irradiancia global en el plano de los paneles sumando las tres componentes.

Al calcular la irradiancia efectiva se necesita conocer sus tres componentes sin las pérdidas asociadas, en lugar de la irradiancia global, por lo que el paso (c) no se contempla en el apartado. En el anexo III se analiza la irradiancia y su efecto al atravesar la atmósfera, recibiéndose a nivel de suelo.

a) Irradiancia horizontal difusa y directa

Para calcular la irradiancia horizontal difusa (d_{hm0}), se requiere conocer la fracción de difusa horizontal (k_{dh}), donde g_{hm0} es la GHI:

$$d_{hm0} = k_{dh}\, g_{hm0}$$

La fracción de difusa horizontal se puede calcular a partir del índice de claridad horizontal (k_{th}), mediante correlaciones empíricas [86]. El índice de claridad horizontal se calcula como el cociente entre la GHI y la radiación extraterrestre (b_{00}), de la siguiente forma:

$$k_{th} = g_{hm0}/b_{00}$$

La radiación extraterrestre se calcula a partir de la excentricidad de la órbita y el ángulo cenital del sol, de la misma forma que en el apartado 4.3. Para determinar la fracción de difusa horizontal a partir del índice de claridad se utiliza la correlación de [86], de acuerdo a la siguiente expresión:

$$k_{dh} = 0.995 - 0.081k_{th} \qquad si\ k_{th} \leq 0.21$$

$$k_{dh} = 0.724 + 2.738k_{th} - 8.32k_{th}^2 + 4.967k_{th}^3 \qquad si\ 0.21 < k_{th} \leq 0.76$$

$$k_{dh} = 0.18 \qquad si\ k_{th} > 0.76$$

Finalmente, se calcula la irradiancia directa horizontal (b_{hm0}) a partir de la GHI:

$$b_{hm0} = g_{hm0} - d_{hm0} = g_{hm0} - k_{dh}g_{hm0} = g_{hm0}(1 - k_{dh})$$

b) Componentes en el plano de los paneles

La irradiancia directa en el plano de los paneles (b_{hmi}) se obtiene por consideraciones geométricas, donde θ_s es el ángulo cenital del sol en el momento de la medida:

$$b_{hmi} = b_{hm0}/\cos(\theta_s)$$

El albedo en el plano de los paneles (a_h) se calcula mediante un modelo isotrópico, donde r_0 es el coeficiente de albedo, que es de 0.2 en el nodo real de pruebas, al tener un suelo de hormigón, y β es el ángulo de inclinación de los paneles:

$$a_h = r_0 g_{hm0}(1 - \cos(\beta))/2$$

El cálculo de la irradiancia difusa en el plano de los paneles (d_{hi}) se realiza mediante el modelo de [87], descomponiendo en una componente isotrópica (d_{ii}) y otra debido a la radiación circunsolar (d_{ci}) :

$$d_{hi} = d_{ii} + d_{ci}$$

$$d_{ii} = d_{hm0}(1 - k_1)(1 + \cos(\beta))/2$$

$$d_{ci} = d_{hm0}k_1/\cos(\theta_s)$$

$$k_1 = b_{hm0}/b_{00}\ \ (cociente\ de\ anisotropía)$$

7.1.2 Irradiancia efectiva en el panel

Para el cálculo de la potencia suministrada por los paneles se necesita conocer la irradiancia efectiva (G_{panel}). Esta irradiancia incluye: (a) un factor de corrección angular, al aumentar la reflexión con el ángulo de incidencia sobre el panel; y (b) un factor de correlación espectral, al cambiar la fotocorriente con el espectro de la luz solar.

a) Pérdidas angulares y por suciedad

Para determinar las pérdidas angulares, que dependen del grado de suciedad de los paneles, se ha empleado el método de [88]:

$$b_{efect} = D_T b_{hi}(1 - F_B)$$

$$a_{efect} = D_T a_h(1 - F_A)$$

$$d_{efect} = D_T[d_{ci}(1 - F_B) + d_{ii}(1 - F_D)]$$

Donde D_T es la transmisión normal, característica del grado de suciedad de los paneles. Para un grado de suciedad de paneles medio se ha considerado $D_T = 0.97$. Las ecuaciones F_A, F_B y F_D se obtienen de [88]:

$$F_A(\beta) \cong \exp\left[-\frac{1}{a_r}\left(c_1\left(sin\beta + \frac{\beta - sin\beta}{1 - cos\beta}\right) + c_2\left(sin\beta + \frac{\beta - sin\beta}{1 - cos\beta}\right)^2\right)\right]$$

$$F_B(\alpha) = \frac{\exp\left(-\cos\left(\frac{a}{a_r}\right)\right) - \exp\left(-\frac{1}{a_r}\right)}{1 - \exp(-1/\alpha_r)}$$

$$F_D(\beta) \cong \exp\left[-\frac{1}{a_r}\left(c_1\left(sin\beta + \frac{\pi - \beta - sin\beta}{1 + cos\beta}\right) + c_2\left(sin\beta + \frac{\pi - \beta - sin\beta}{1 + cos\beta}\right)^2\right)\right]$$

Donde los coeficientes a_r y c_2 dependen de la tecnología del módulo. Para la plataforma experimental utilizada en la tesis, con células solares de silicio policristalino (p-Si), se toma los valores $a_r = 0.21$ y $c_2 = -0.049$. Finalmente, la irradiancia sobre los paneles, teniendo en cuenta las pérdidas angulares, es:

$$G_{ang} = b_{efect} + d_{efect} + a_{efect}$$

b) Pérdidas espectrales

Para determinar las pérdidas espectrales se ha empleado el método de [89], que corrige la irradiancia añadiendo un término denominado modificador de desajustes espectrales (f_{Ama}).

$$G_{panel} = f_{Ama} G_{ang} = f_{Ama}(b_{efect} + d_{efect} + a_{efect})$$

f_{Ama} depende de la masa de aire (AM), corregida para la altura del lugar (z), en forma de polinomio de cuarto orden:

$$f_{Ama} = a_0 + a_1 Ama + a_2 Ama^2 + a_3 Ama^3 + a_4 Ama^4$$

$$Ama = AM \cdot \exp(-0.0001184z)$$

$$AM = 1/(\cos(\theta_s) + 0.50572(96.07995 - 180\theta_s/\pi)^{-1.6364})$$

Los coeficientes a_0, a_1, a_2, a_3 y a_4 se obtienen para módulos concretos, según los valores promedio de la Tabla 7. Estos coeficientes se calcularon para una gran muestra de módulos [106], y para células solares de silicio policristalino (p-Si) estos valores son $a_0 = 0.93017$, $a_1 = 0.062992$, $a_2 - 0.012156$, $a_3 = 0.00094519$ y $a_4 = -2.6e^{-5}$.

Tabla 7.- Valores promedios de los coeficientes para paneles de diferentes tecnologías.

Coeficientes del método de Sandía						
Tecnología		a_0	a_1	a_2	a_3	a_4
2-a-Si	Media	8.8832e-01	1.5329e-01	-6.2913e-02	7.3055e-03	-2.9175e-04
	Moda	8.2392e-01	2.2748e-01	-8.8651e-02	1.1387e-02	-5.1542e-04
3-a-Si	Media	1.0470e+00	8.2115e-04	-2.5900e-02	3.1736e-03	-1.1026e-04
	Moda	1.0470e+00	8.2115e-04	-2.5900e-02	3.1736e-03	-1.1026e-04
CdTe	Media	9.4679e-01	5.8883e-02	-1.7872e-02	1.7536e-03	-6.7603e-05
	Moda	9.4170e-01	6.5160e-02	-2.0220e-02	2.1900e-03	-9.0700e-05
Si-Film	Media	9.2847e-01	7.2365e-02	-1.9025e-02	1.7040e-03	-4.9674e-05
	Moda	9.2800e-01	7.3144e-02	-1.9427e-02	1.7513e-03	-5.1288e-05
HIT-Si	Media	9.2715e-01	6.1518e-02	-9.8363e-03	6.1917e-04	-1.2734e-05
	Moda	9.2619e-01	6.2356e-02	-1.0119e-02	6.7031e-04	-1.6221e-05
EFG-mc-Si	Media	9.3600e-01	5.3645e-02	-7.9402e-03	5.2228e-04	-1.3142e-05
	Moda	9.3600e-01	5.3645e-02	-7.9402e-03	5.2228e-04	-1.3142e-05
CIS	Media	9.2100e-01	7.1817e-02	-1.4619e-02	1.2500e-03	-3.7403e-05
	Moda	9.2100e-01	7.1817e-02	-1.4619e-02	1.2500e-03	-3.7403e-05
mc-Si	Media	9.3017e-01	6.2992e-02	-1.2156e-02	9.4519e-04	-2.6000e-05
	Moda	9.2108e-01	7.5403e-02	-1.8481e-02	2.1538e-03	-1.0662e-04
c-Si	Media	9.3309e-01	5.9175e-02	-1.1066e-02	8.9574e-04	-2.8517e-05
	Moda	9.3800e-01	5.4230e-02	-9.9000e-03	7.3000e-04	-1.9100e-05

7.1.3 Temperatura de los paneles

Se puede considerar directamente la temperatura medida en la cara posterior de un panel como la temperatura de funcionamiento de las células. Esto tiene el problema de que el panel puede estar a la sombra y el resto al sol, lo que da una estimación poco fiable. Por ello, se ha estimado la temperatura de los paneles (T_{cell}) como:

$$T_{cell} = \frac{NOCT - 20^oC}{800W/m^2} G_{panel} + T_{amb}$$

Donde la temperatura de operación nominal de la célula (*Nominal Operating Cell Temperature*, NOCT) es la que da el fabricante, que para los módulos empleados en la instalación, encontrados en el anexo II, toma un valor de $TONC = 45^oC$. Esta expresión es de aplicación obligatoria en las predicciones, ya que la información conocida es la temperatura ambiente, y no la temperatura de panel.

Adicionalmente, se podría hacer una corrección en función del viento, mediante alguna de las expresiones de [107]. Sin embargo, el viento incidente en cada módulo puede ser diferente al medido en el anemómetro de la estación, siendo igualmente un efecto de segundo orden en la temperatura del panel.

7.1.4 Potencia de los paneles

Para determinar la potencia de la instalación fotovoltaica, se puede usar la siguiente expresión:

$$P_{DC} = SF \cdot \eta_{DC} \cdot P_{PV} = SF \cdot \left(\eta_{reg} \cdot \eta_{cable,DC} \cdot \eta_{otros}\right) \cdot P_{PV}$$

Donde SF son las pérdidas por sombras, que requiere analizar los obstáculos de la instalación a lo largo de todo el año, y η_{DC} es el rendimiento en continua, que considera la eficiencia del regulador instalado ($\eta_{reg} = 0.98$), el cableado en la zona de continua ($\eta_{cable,DC} = 0.97$) y otras pérdidas ($\eta_{otros} = 0.975$) como MPPT o falta de disponibilidad.

La instalación trabaja en continua, por lo que no aplica pérdidas típicas asociadas a la conversión a alterna. Para calcular la potencia de los paneles (P_{PV}), se prueban dos modelos: (a) el modelo clásico de Osterwald [42] y (b) una corrección del mismo a baja irradiancia [79].

a) Modelo de Osterwald [42]

Calcula la potencia a partir de la irradiancia efectiva y la temperatura de los paneles, utilizando un coeficiente de temperatura para la potencia:

$$P_{DC} = SF \cdot \eta_{DC} \cdot P_{peak} \cdot \frac{G_{panel}}{G_{STC}}\left(1 + \delta P_m\left(T_{cell} - T_{cell,STC}\right)\right)$$

Donde $P_{peak} = 2.97 kW_p$ es la potencia pico de todo el array de paneles en condiciones test estándar (*Standard Test Conditions*, STC), $G_{STC} = 1 kW/m^2$. $\delta P_m = -0.4\%/^oC$ es el coeficiente de temperatura para la potencia máxima, obtenido de la hoja de características del fabricante. η_{DC} incluye las pérdidas por cableado entre módulos, pérdidas por la tolerancia entre módulos y pérdidas por dispersión de parámetros, tomando un valor de $\eta_{DC} = 0.927$ en el nodo real de pruebas.

b) Modelo de Osterwald con corrección [79]

Este modelo tiene en cuenta que la eficiencia suele disminuir a baja irradiancia, corrigiendo el modelo de Osterwald para $G_{panel} \leq 125 W/m^2$:

$$P_{DC} = P_{DC(Osterwald)} \quad si\ G_{panel} > 125 W/m^2$$

$$P_{DC} = 0.008 \cdot G_{panel} \cdot P_{DC(Osterwald)} \quad si\ G_{panel} \leq 125 W/m^2$$

Ambos modelos son prácticamente idénticos en el caso de estudio, pues apenas existen valores de $G_{panel} \leq 125 W/m^2$, siendo ligeramente superior el modelo de Osterwald. Por lo tanto, se toma el modelo de Osterwald finalmente para la estimación de potencia de los paneles en el nodo real de pruebas.

7.2 Resultados del método analítico

Para comprobar la precisión de los resultados, se han contrastado medidas reales de potencia fotovoltaica del regulador de la instalación con valores estimados de potencia fotovoltaica, calculados con medidas reales de GHI de la estación meteorológica. La Figura 31 representa la medida de potencia y su estimación, seleccionando tres tipos de días para contemplar diferentes casos: soleado, nublado y encapotado. También se ha considerado el efecto de las sombras de un muro colindante, añadiendo una variación lineal desde $SF = 0.4$, a las 16:36 horas, hasta $SF = 0.95$, al anochecer. Los valores de SF dependen de la estación del año, debiendo corregirse periódicamente.

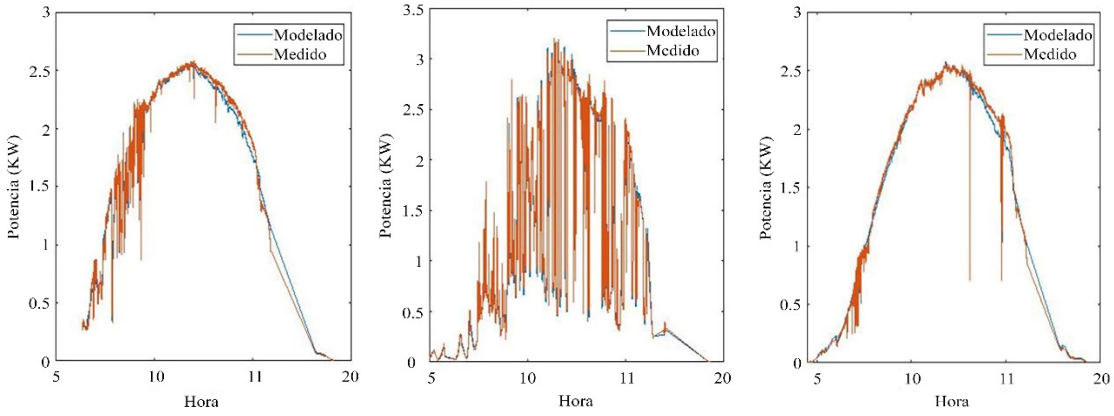


Figura 31.- Comparación entre valores medidos de potencia y estimados con medidas reales de GHI. Los días seleccionados son: (a) día nublado (17 de mayo de 2021: $rMAE = 3.04\%$ $rRMSE = 3.90\%$); (b) día encapotado (1 de junio de 2021: $rMAE = 4.03\%$ $rRMSE = 5.95\%$); y (c) día soleado (4 de junio de 2021: $rMAE = 2.54\%$ $rRMSE = 3.44\%$).

Los resultados muestran un error bastante reducido, similar a los reportados en otros trabajos relacionados [81,84]. El error medio obtenido es de $rMAE = 2.54\%$ en días soleados, $rMAE = 3.04\%$ en días nublados y $rMAE = 4.03\%$ en días encapotados. Con respecto a los errores cuadráticos,

se obtiene $rRMSE = 3.44\%$ en días soleados, $rRMSE = 3.90\%$ en días nublados y $rRMSE = 5.95\%$ en días encapotados.

Las medidas del regulador muestran que el MPPT tiene un punto de operación inestable en presencia de sombras parciales de nubes, reduciendo la potencia generada excesivamente en pequeños instantes de tiempo. Esto provoca que el error producido en días con presencia de nubes aumente. Sin embargo, esto no genera un problema al proceso de predicción, ya que la resolución temporal es de 15 minutos y se elimina este efecto.

7.3 Precisión en la predicción de potencia

Una vez conocida la precisión que ofrece el método analítico frente a medidas reales en la instalación, se utilizan las predicciones de GHI de capítulos previos para conocer el error cometido en el modelo desarrollado en la tesis. Los mapas NWP de la AEMET también se utilizan para predecir la temperatura ambiente en el nodo real de pruebas en el momento de la predicción. El apartado 7.3.1 calcula el error cometido para las predicciones con un horizonte temporal intradiario, mientras que el apartado 7.3.2 determina la precisión para la estrategia de predicción diaria.

7.3.1 Predicción de potencia intradiaria

En este apartado se evalúa el error obtenido en la predicción de potencia con un horizonte temporal intradiario. Este error debe ser coherente con el producido en la predicción de GHI, obteniendo las mismas conclusiones en función del tiempo de espera y el plazo intradiario de lanzamiento. En primer lugar, se evalúan los resultados para el nodo real de pruebas, posteriormente, se emula una VPP con las medidas de GHI obtenidas en las estaciones meteorológicas de la Comunidad de Madrid.

a) Potencia generada en el nodo real de pruebas

La Figura 32 muestra las matrices del error del nodo real de pruebas en las predicciones de potencia con un horizonte temporal intradiario en función del plazo intradiario de lanzamiento y el tiempo de espera. Se comprueba que se obtienen las mismas conclusiones que en la Figura 30: (i) el error escalado es elevado en plazos intradiarios de lanzamiento cortos y tiempos de espera medios, pero decrece significativamente al incrementar el plazo intradiario de lanzamiento; (ii) en plazos intradiarios de lanzamiento de aproximadamente una hora y media o menos, los errores porcentuales son elevados, independientemente del tiempo de espera; y (iii) los errores porcentuales son elevados en tiempos de espera superiores a 7 horas, de forma aproximada.

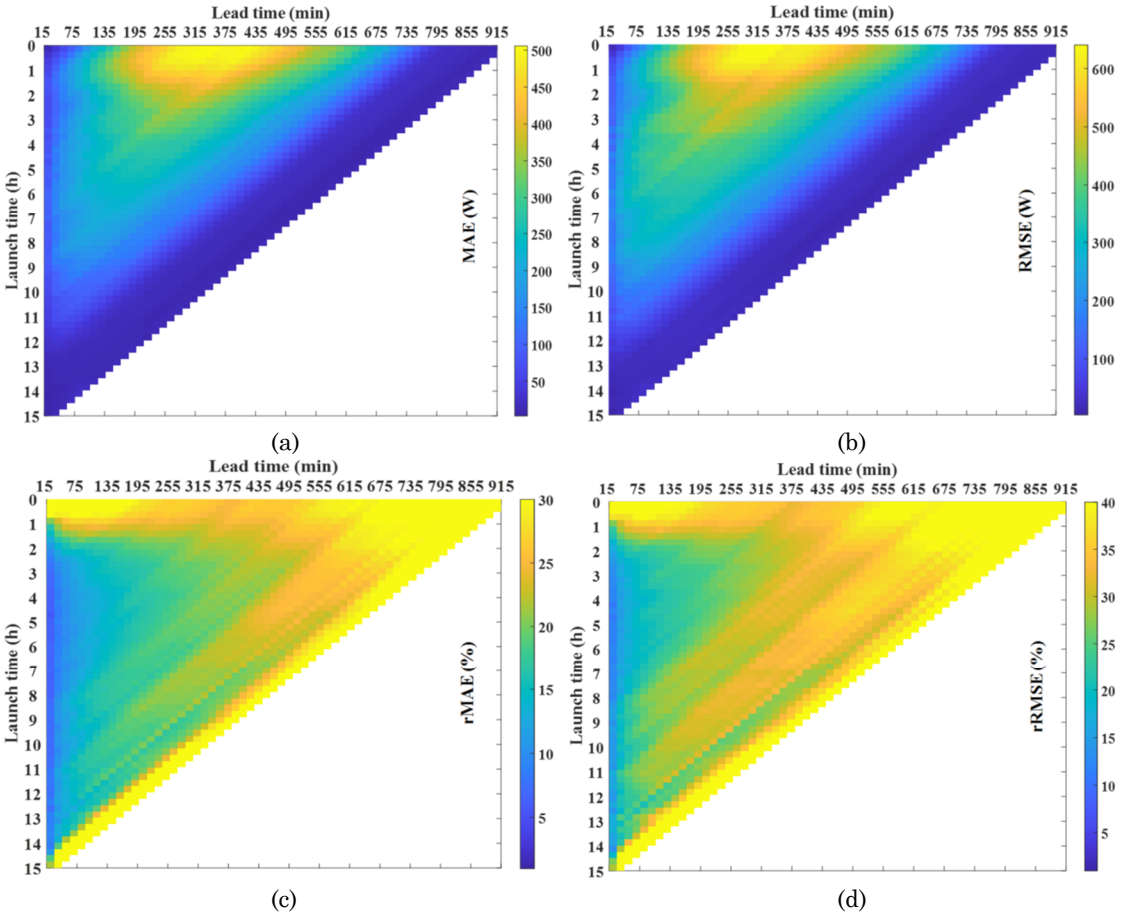


Figura 32.- Matrices del error de la potencia en el nodo real de pruebas en función del plazo intradiario de lanzamiento y el tiempo de espera: (a) MAE; (b) RMSE; (c) rMAE; y (d) rRMSE

b) Potencia generada en una VPP emulada

Una vez mostrado el potencial del modelo de predicción propuesto para una instalación fotovoltaica de un nodo real de pruebas, se realiza este estudio para un conjunto de plantas fotovoltaicas agrupadas. Al no disponer de otras instalaciones fotovoltaicas, se añaden las medidas de siete estaciones meteorológicas en la Comunidad de Madrid, situadas en las localizaciones de la Figura 33. Estas estaciones meteorológicas ofrecen toda la información necesaria para la predicción desde su web [42], utilizando además los mapas de NWP de la AEMET.

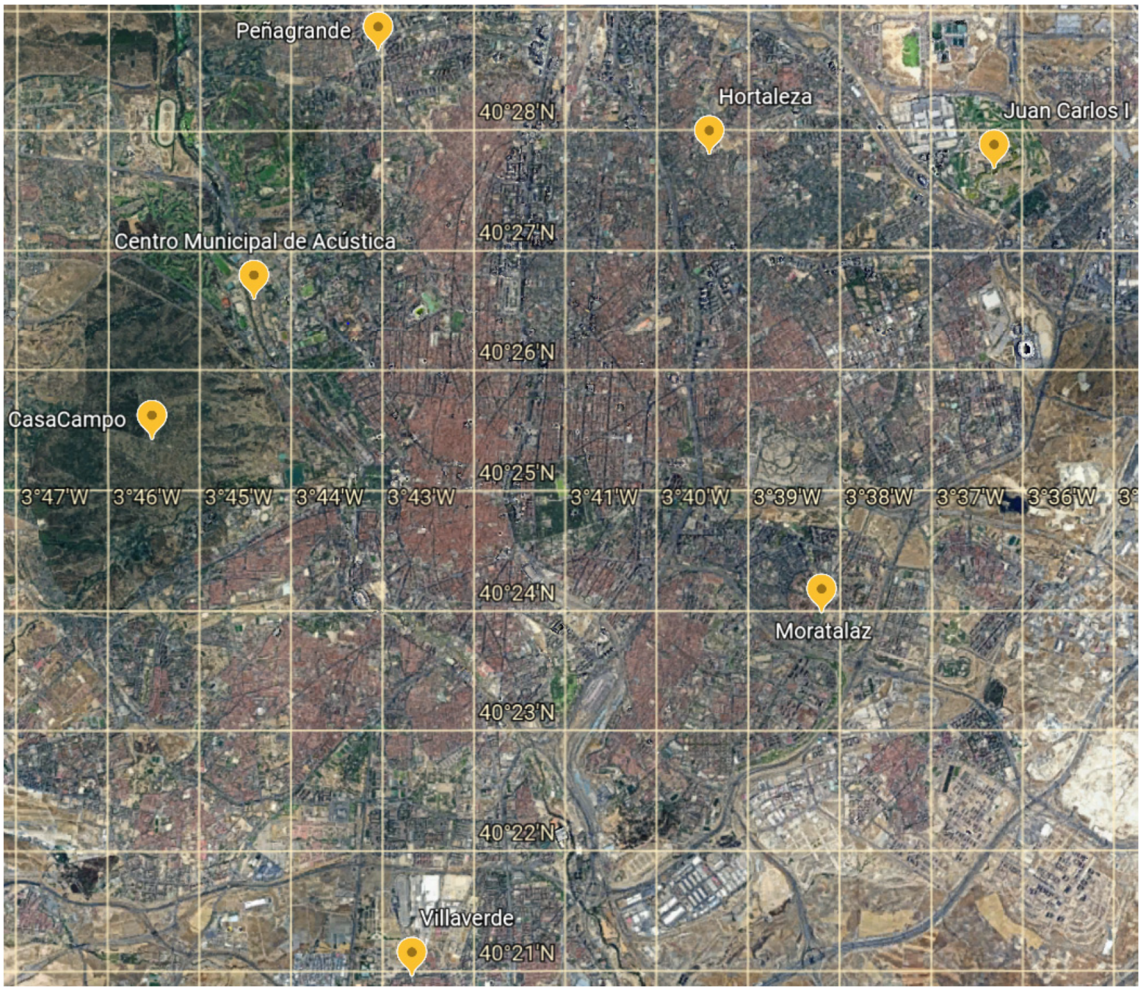


Figura 33.- Localización de las estaciones a nivel de suelo en la Comunidad de Madrid empleadas en el presente estudio

La Figura 34 muestra las matrices del error de la VPP en las predicciones de potencia con un horizonte temporal intradiario en función del plazo intradiario de lanzamiento y el tiempo de espera. Las predicciones se realizan para cada instalación de forma individual, sumando después todos los nodos para estimar la potencia generada por la VPP. Se pueden observar unos errores escalados superiores a los de la Figura 32, debido a que la potencia pico de la VPP considerada es ocho veces superior a la de una única instalación.

Por otra parte, analizando los valores porcentuales del error, se observa un mejor resultado en el caso de VPP con respecto a una única instalación. Dicha mejora se puede expresar como el error cometido en la VPP, menos la suma del error absoluto de las instalaciones individuales, dividido entre el error medio. Los resultados muestran una mejoría del 12.37% con respecto al rMAE y un 11.84% con respecto al rRMSE. También se pueden extraer las mismas conclusiones descritas en el nodo real de pruebas, por lo tanto, el potencial del modelo también es de aplicación en el entorno VPP.

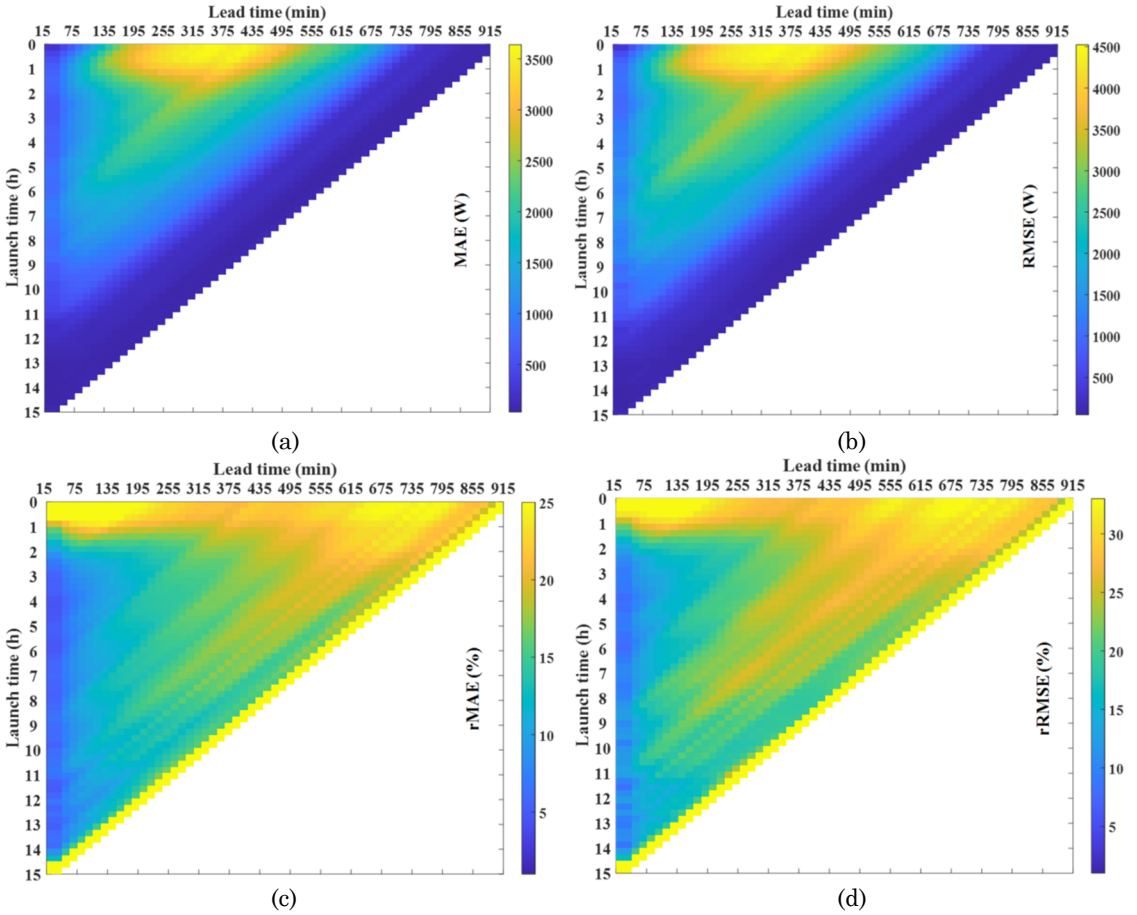


Figura 34.- Matrices del error de la potencia en una VPP emulada de 23.76 kW_p, compuesta por ocho nodos situados en la Comunidad de Madrid, en función del plazo intradiario de lanzamiento y el tiempo de espera: (a) MAE; (b) RMSE; (c) rMAE; y (d) rRMSE

7.3.2 Predicción de potencia diaria

En este apartado se evalúa el error obtenido en la predicción de potencia con un horizonte temporal diario, buscando obtener las mismas conclusiones que la predicción de GHI del capítulo 5. En este caso, se obtienen los resultados del nodo real de pruebas y la VPP considerando un total de siete nodos.

La Figura 35 muestra los errores porcentuales cometidos en función del histórico para el nodo real de pruebas y la VPP, observando un comportamiento muy similar al obtenido en la predicción de GHI de la Figura 20 y la Figura 24, respectivamente. La Tabla 8 muestra los errores porcentuales del modelo híbrido para los seis meses de histórico, en cada uno de los nodos considerados en el estudio del capítulo 5, comprobando que la precisión es muy similar entre nodos. Al igual que en la predicción de GHI de la Tabla 4, la precisión obtenida en la VPP es bastante superior al error medio de los nodos individuales. Esta mejora es de un 10.46% con respecto al rMAE y un 10.65% con respecto al rRMSE.

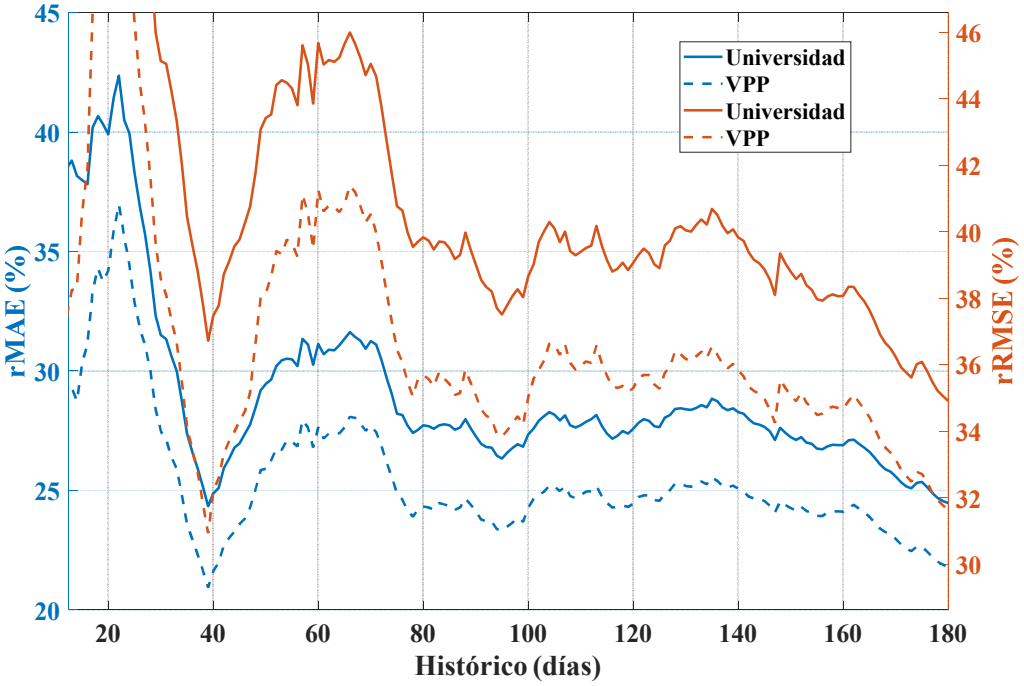


Figura 35.- Error de predicción del modelo híbrido para el nodo real de pruebas y la VPP en términos del rMAE y el rRMSE.

Tabla 8.- rMAE y el rRMSE de la predicción de potencia tras 180 días de histórico.

Modelo	Nodo	Universidad	Ctro. Mpal. de Acústica	Hortaleza	Juan Carlos I	Moratalaz	Peñagrande	Villaverde	Media	**VPP**
Híbrido	rMAE (%)	24.57	24.61	25.34	23.87	24.29	24.12	23.77	24.37	**21.82**
	rRMSE (%)	35.04	35.27	37.59	35.13	34.87	34.91	35.67	35.50	**31.72**

7.4 Conclusiones

En este capítulo se ha descrito el método analítico usado para estimar la potencia fotovoltaica a partir de medidas de GHI y temperatura ambiente, junto a las características de la propia instalación. Aspectos como las pérdidas angulares, la suciedad de los paneles, las pérdidas espectrales y el sombreado han sido considerados para llevar a cabo el estudio.

Los resultados muestran un $rMAE$ que fluctúa entre un 2.54% y un 4.03%, y un $rRMSE$ que oscila entre un 3.44% y un 5.95%, en función del tipo de día, obteniendo un menor error en días soleados que aumenta conforme se incrementa la presencia de nubes en la instalación. En relación a otros artículos de la literatura [81,84], se alcanzan errores bastante similares, validando el método analítico empleado en el estudio.

Además de calcular el error producido en el método analítico, utilizando medidas reales de la instalación, se ha estudiado la precisión obtenida con respecto a las predicciones de GHI realizadas en capítulos previos. Se ha realizado el estudio para el nodo real de pruebas y para la VPP, observando conclusiones similares al estudio de irradiancia y permitiendo validar el modelo desarrollado.

Al comparar la VPP con los nodos que la componen de forma individual, se observan mejoras en la precisión. Concretamente, en el horizonte temporal intradiario, la mejora es de un 12.37% con respecto al rMAE y un 11.84% con respecto al rRMSE. Por otra parte, para el horizonte temporal diario, la mejora es de un 10.46% con respecto al rMAE y un 10.65% con respecto al rRMSE.

8 Intervalos de predicción

La potencia fotovoltaica es de carácter estocástico, ya que está muy relacionada con las condiciones meteorológicas de la zona. Esto limita la habilidad de los modelos de predicción para aportar una gran precisión, obteniendo un nivel de incertidumbre que debe ser evaluado. Los intervalos de predicción ofrecen un rango de estimación para un valor futuro desconocido [41], expresando el grado de incertidumbre asociado a las predicciones puntuales y añadiendo un mayor nivel de confianza.

Estos intervalos proporcionan información adicional sobre el rango de posibles valores de generación fotovoltaica para un nivel de confianza seleccionado por el usuario. Esto puede evitar déficits energéticos inesperados o, por el contrario, excesos energéticos, aunque estos últimos son menos críticos que los primeros, ya que el inversor puede cambiar el punto de operación para producir la energía necesaria, a pesar de desperdiciar un recurso energético de utilidad.

Con el histórico de datos de predicciones fotovoltaicas, es posible calcular intervalos para nuevas predicciones. En la tesis se emplean intervalos estadísticos de predicción basados en el trabajo presentado en [45], utilizando un modelo de distribución Laplaciano con el histórico del error. Para medir la precisión obtenida al considerar una distribución Laplaciana, se emplea la probabilidad de cobertura del intervalo de predicción (*Prediction Interval Coverage Probability*, PICP) [108], que indica el porcentaje de valores predichos que permanecen dentro del intervalo seleccionado, debiendo ser similar al nivel de confianza (γ_L) utilizado.

$$PICP = \frac{1}{T}\sum_{t=1}^{T} \epsilon_t \, , where \; \epsilon_i = \begin{cases} 1 \; si \; x_i \in [L_i, U_i] \\ 0 \; si \; x_i \notin [L_i, U_i] \end{cases}$$

La Figura 36 muestra intervalos de predicción para un día específico, usando un nivel de confianza del $\gamma_L = 80\%$. Se observa que añadir esta información proporciona un mayor potencial a la estrategia, reconociendo un posible caso pesimista, en el intervalo inferior, y otro caso optimista, en el intervalo superior, entre los cuales se estima se encuentra la predicción.

El nivel de confianza (γ_L) es un indicativo del porcentaje medio de valores que quedan dentro del intervalo de la predicción, que será de utilidad si la distribución del error representa adecuadamente los resultados obtenidos. El nivel de confianza usado en la tesis es del 80%, sin embargo, este valor puede ser modificado por el usuario. Si se desea un rango de valores estrecho para actuar de forma precisa, el nivel de confianza se reduce, sin embargo, el riesgo que se asume es mayor, ya que habrá más probabilidad de que los valores queden fuera del rango.

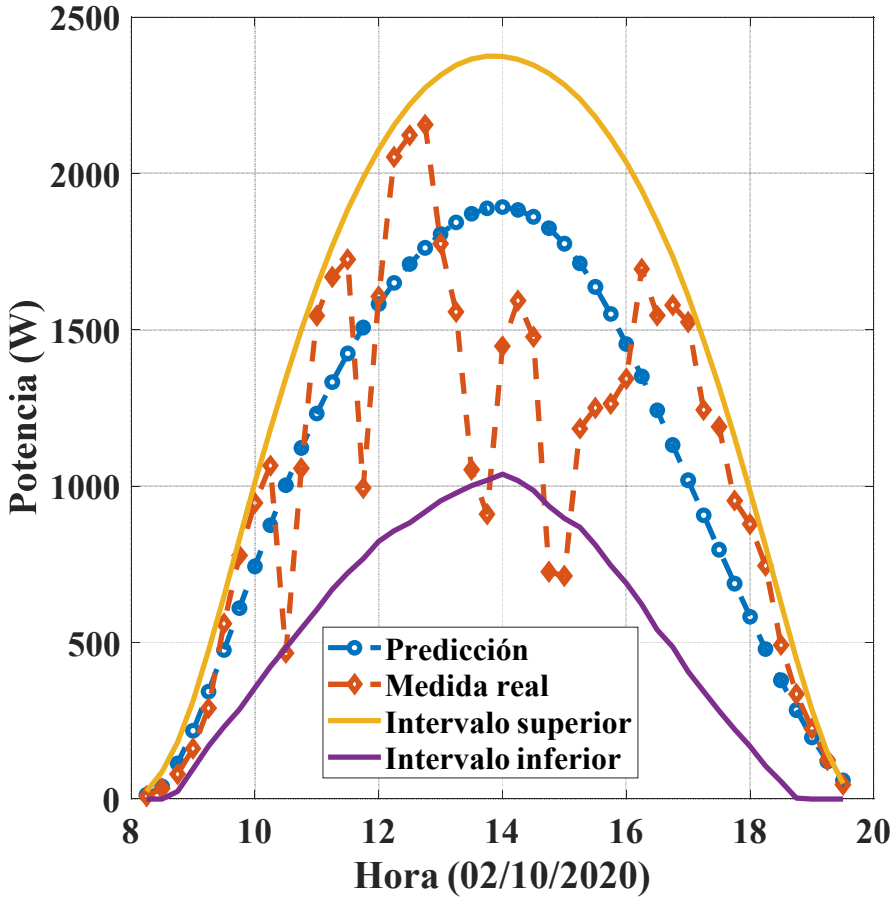


Figura 36.- Intervalos de predicción considerados en el nodo real de pruebas para un día específico y un nivel de confianza del 80%.

En los siguientes apartados se muestran los intervalos de predicción obtenidos en la tesis. El apartado 8.1 calcula los intervalos para la estrategia de predicción diaria. El apartado 8.2 determina los intervalos para predecir la potencia fotovoltaica con un horizonte temporal intradiario. Finalmente, el apartado 8.3 describe las conclusiones obtenidas en el estudio.

8.1 Intervalos de predicción diario

La generación de intervalos para un horizonte temporal de un día se realiza dividiendo el histórico de predicciones pasadas en 10 grupos, en función del CCF descrito en el apartado 4.2.2. El CCF indica la cantidad de nubes que generan sombra en la instalación, por lo que es un indicativo del tipo de día. A mayor presencia de nubes mayor es el error en la predicción, de forma que se dividen adecuadamente los intervalos, teniendo los días soleados un intervalo menor que los días nublados. En la Figura 37 se observa la distribución del error, siendo razonable asumir una distribución Laplaciana en los diferentes grupos.

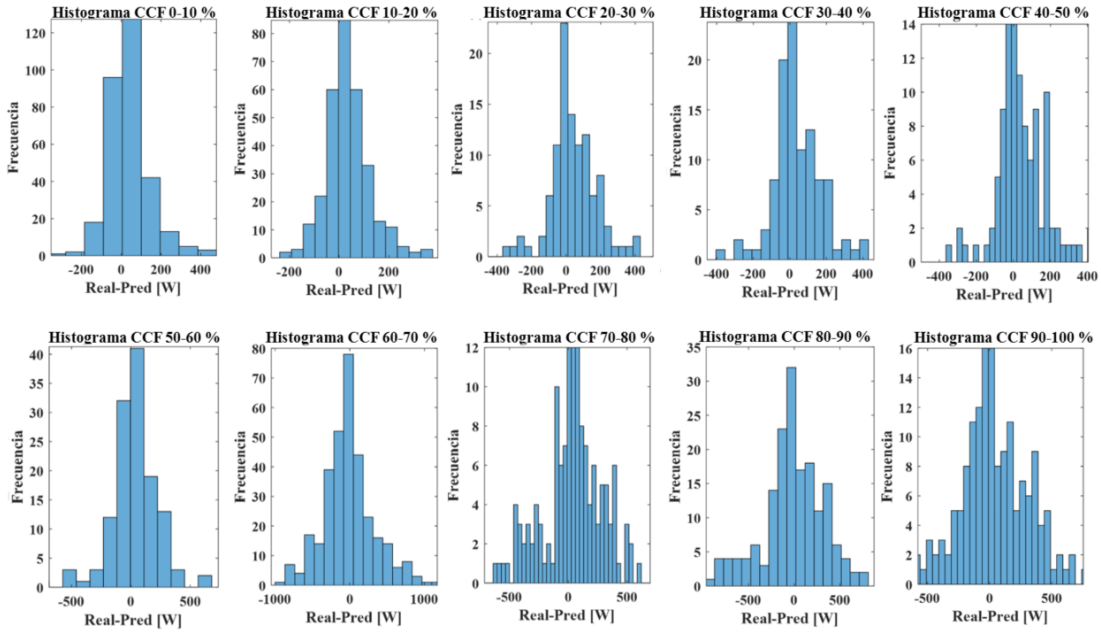


Figura 37.- Distribución del error para cada grupo seleccionado en función del CCF. Para generar intervalos de predicción, se asume una distribución Laplaciana.

Al asumir una distribución Laplaciana del error, es posible definir un intervalo de predicción para cada grupo, $G_{pred} \pm p_s$, en función del MAE y del percentil p de probabilidad $(1-s)$, sabiendo que para la distribución Laplaciana $p_s = \pm MAE \cdot \ln(2s)$.

La Tabla 9 muestra el error medio escalado y porcentual cometido en cada grupo. Se observa que el error aumenta con la presencia de nubes, sin embargo, cuando el CCF es superior al 60%, el error cometido no varía excesivamente. A la hora de evaluar la fiabilidad de la estimación, se hace uso del PICP para un nivel de confianza del 80% ($s = 0.1$), observando unos valores adecuados, ya que el PICP es similar al nivel de confianza.

Tabla 9.- Errores medios cometidos, intervalos de predicción y PICP para cada uno de los grupos seleccionados en función del CCF.

CCF	MAE $[W]$	rMAE [%]	p_s $[W]$	$PICP$
0 - 10 %	91.76	7.90	±147.68	0.84
10 - 20%	118.26	12.57	±190.33	0.82
20 - 30 %	145.60	16.75	±234.33	0.81
30 - 40 %	156.72	18.77	±252.23	0.81
40 - 50 %	194.19	20.73	±312.53	0.77
50 - 60 %	215.87	24.12	±347.42	0.77
60 - 70 %	234.48	26.61	±377.37	0.74
70 - 80 %	240.62	28.83	±387.25	0.79
80 - 90 %	248.91	30.35	±400.60	0.80
90 - 100 %	245.35	29.34	±394.87	0.76
HISTÓRICO	204.85	23.63	±329.69	0.81

8.2 Intervalos de predicción intradiario

Los resultados obtenidos en la predicción de horizonte temporal intradiario muestran que la precisión del modelo depende en gran medida del tiempo de espera y el plazo intradiario de lanzamiento. Por ello, además de agrupar el histórico en función del CCF, en este caso también se consideran estos dos parámetros en el agrupamiento. La Figura 38 muestra la distribución del error para diferentes valores del tiempo de espera y del plazo intradiario de lanzamiento, asumiendo una distribución Laplaciana.

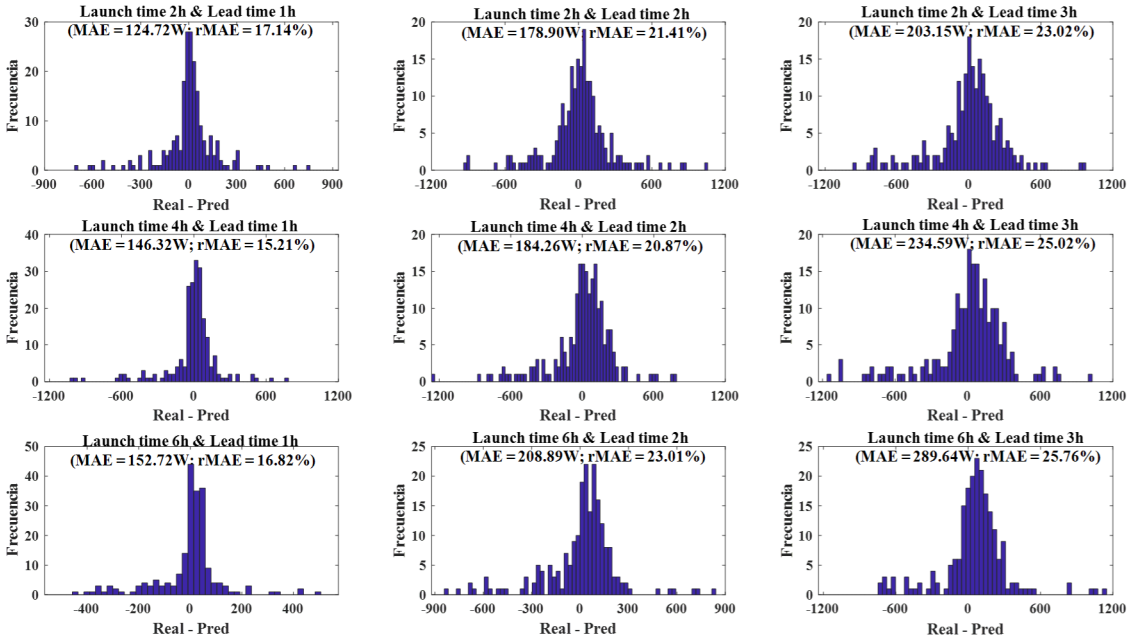


Figura 38.- Distribución del error para cada grupo seleccionado, de acuerdo al tiempo de espera y el plazo intradiario de lanzamiento. Para generar intervalos de predicción, se asume una distribución Laplaciana.

Al disponer los grupos de acuerdo a tres parámetros: (i) tiempo de espera; (ii) plazo intradiario de lanzamiento; y (iii) CCF, el número de particiones aumenta considerablemente, por lo que hay que limitar la cantidad de grupos para que el número de muestras sea representativo. De esta forma, en lugar de dividir el CCF similar al horizonte temporal diario, se han seleccionado 3 grupos en función del tipo de día: soleado, nublado y encapotado. La clasificación del tipo de día se ha realizado con el método de los k-NN, similar al apartado 5.1.3.

La Figura 39 muestra la diferencia absoluta entre el nivel de confianza seleccionado ($\gamma_L = 80\%$) y el PICP, para cada tipo de día. Si esta diferencia es cercana a cero, los intervalos están bien representados. En la Figura 39a, para días soleados, el PICP es cercano al nivel de confianza para todos los grupos, excepto en tiempos de espera elevados bajo plazos intradiarios de lanzamiento pequeños, donde dicha diferencia se ve incrementada. En la Figura 39b, para días nublados, el PICP difiere bastante del nivel de confianza en zonas cercanas al anochecer, sin embargo, se mantiene cercano a cero en el resto del área. Es de reseñar que esto no

supone un problema, ya que la predicción pierde relevancia en zonas cercanas al anochecer, al reducirse la energía producida significativamente. En la Figura 39c, para días encapotados, la diferencia entre el PICP y el nivel de confianza es mayor que en días soleados, pero este valor sigue siendo adecuado para emplearlo en la estrategia.

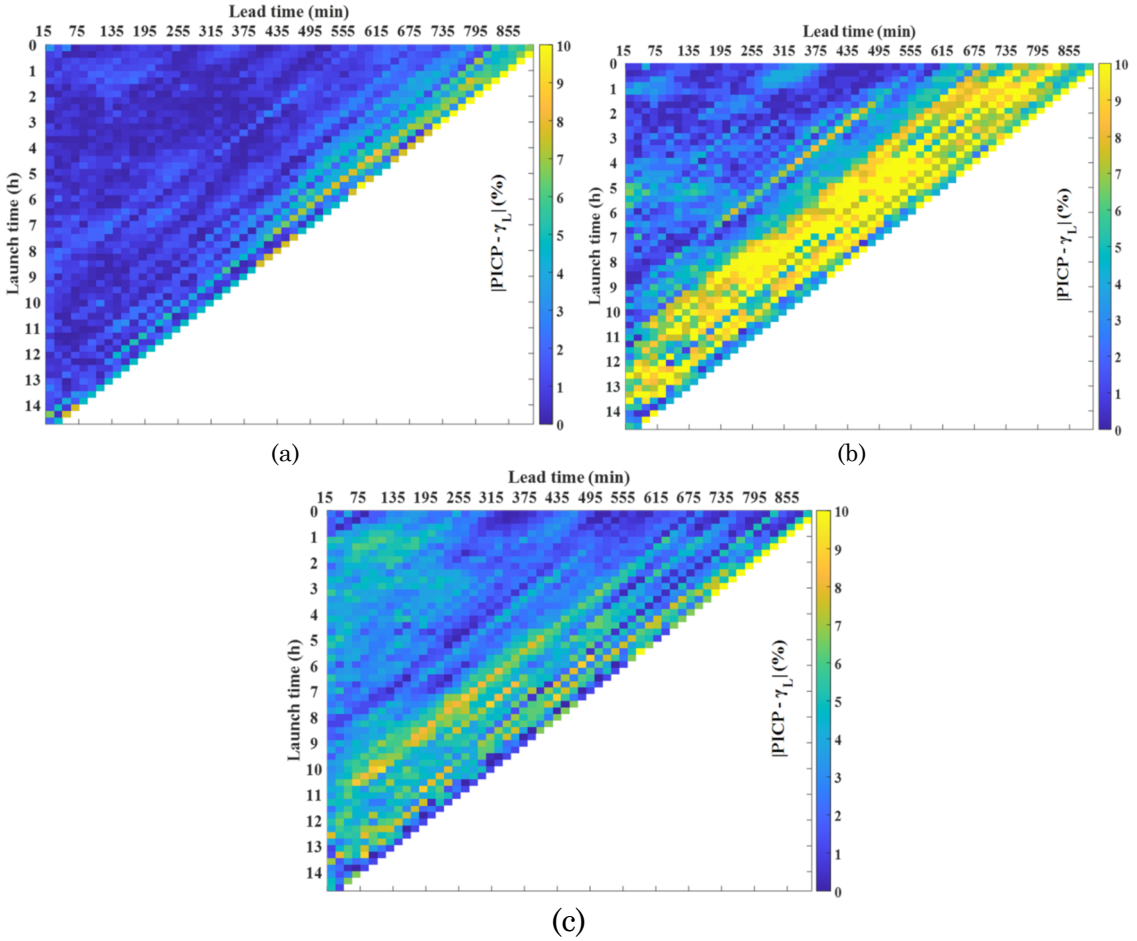


Figura 39.- Diferencia absoluta entre el PICP y el nivel de confianza para cada grupo seleccionado en función del tiempo de espera, el plazo intradiario de lanzamiento y el tipo de día: (a) soleado; (b) nublado; y (c) encapotado.

8.3 Conclusiones

En este capítulo se han seleccionado intervalos de predicción para cada horizonte temporal considerado en la tesis, ampliando la información ofrecida en cada instante de tiempo al considerar una distribución Laplaciana del error. Este proceso separa el histórico en función del CCF para el horizonte temporal diario, y en función del tiempo de espera, el plazo intradiario de lanzamiento y el tipo de día para el horizonte temporal intradiario. Se ha comprobado al calcular la diferencia entre el PICP y el nivel de confianza que el proceso es adecuado para representar los intervalos, empleándose por lo tanto en la estrategia.

En la estrategia de predicción intradiaria hay ciertas zonas con una diferencia notable entre el PICP y el nivel de confianza. Sin embargo, se corresponden con momentos de baja producción fotovoltaica cercanas al

anochecer y un error elevado, como puede observarse en la Figura 32. Por lo tanto, el intervalo de predicción en este caso no es relevante y se puede despreciar del estudio global.

9 Análisis de las predicciones

Este capítulo aúna los resultados obtenidos en las predicciones de potencia bajo diferentes horizontes temporales dentro del entorno VPP. Para ello, se hibridan las estrategias de predicción desarrolladas para una predicción actualizada en cada instante de tiempo que tenga la menor incertidumbre posible. Para validar la aplicabilidad de la hibridación realizada, se implementa un caso de uso real. El apartado 9.1 realiza una hibridación de la estrategia de predicción diaria e intradiaria, con el fin de minimizar el error. El apartado 9.2 desarrolla un caso de estudio real correspondiente a un BEMS basado en las predicciones. Finalmente, el apartado 9.3 analiza las conclusiones obtenidas.

9.1 Hibridación de estrategias

Después de presentar la predicción de la potencia fotovoltaica bajo diferentes horizontes temporales, obteniendo sus intervalos de predicción asociados, se realiza una hibridación entre técnicas para decidir qué estrategia ofrece la menor incertidumbre en función del plazo intradiario de lanzamiento y el tiempo de espera. Para lograr este objetivo, se comparan las estrategias de predicción diaria e intradiaria desarrolladas en capítulos anteriores, para seleccionar en todo momento el modelo que presente el mínimo error. El histórico de medidas utilizado aborda todo un año, iniciando el 1 de junio de 2020 y finalizando el 31 de mayo de 2021. La evaluación de los resultados se realiza en el entorno VPP, utilizando las predicciones de los siete nodos de la Figura 33 y las predicciones del nodo real de pruebas de la Universidad de Alcalá.

En primer lugar, se procede a evaluar el error cometido en la estrategia de predicción diaria, en función del plazo intradiario de lanzamiento y el tiempo de espera, al igual que en la Figura 34, analizando así qué estrategia utilizar en cada momento para minimizar el error. La Figura 40 muestra las matrices del error en la estrategia de predicción diaria. Como era de esperar, el error no depende del plazo intradiario de lanzamiento, ya que estas predicciones no se actualizan a lo largo del día, si no que se realiza una única predicción el día antes. Se observan resultados coherentes, al generar reducidos errores absolutos en zonas cercanas al amanecer y el anochecer, mientras que los errores porcentuales son elevados en dichas zonas. En la zona central del día, los errores absolutos son elevados, debido a que la cantidad de potencia a predecir es mayor, pero porcentualmente menores.

Una vez conocidas las matrices de errores para los horizontes temporales de las dos estrategias presentadas: la Figura 34 para la intradiaria y la Figura 40 para la diaria, se selecciona la predicción que presenta el mínimo error en función del plazo intradiario de lanzamiento y el tiempo de espera. De esta forma, la predicción final hibrida ambas estrategias a lo largo del tiempo, buscando la menor incertidumbre en la

predicción a medida que trascurre el día sobre el que se está trabajando. Es interesante destacar que, como era de esperar, los resultados obtenidos son coherentes con los de capítulos previos, observando que la estrategia de predicción diaria obtiene una mejor precisión que la intradiaria en las primeras horas del día, ya que la estrategia intradiaria aún no dispone de información de ese día, o cuando el tiempo de espera de la predicción dentro del mismo día es elevado.

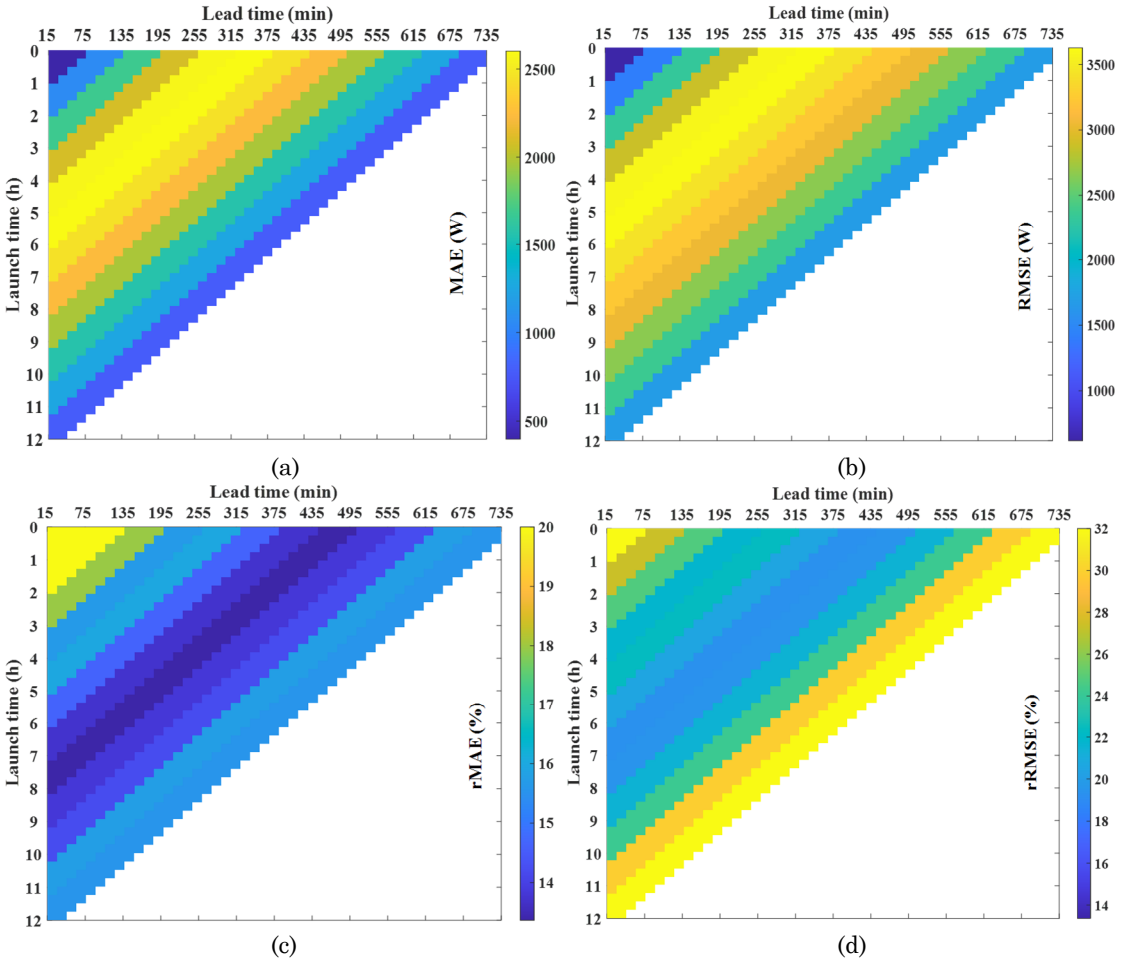


Figura 40.- Matrices del error de la predicción de potencia con un horizonte temporal diario en una VPP emulada de 23.76 kW_p, compuesta por ocho nodos de situados en la Comunidad de Madrid, en función del plazo intradiario de lanzamiento y el tiempo de espera: (a) MAE; (b) RMSE; (c) rMAE; y (d) rRMSE.

La Figura 41 muestra, utilizando los datos de todo el histórico, la elección entre ambas estrategias de predicción, para todos los posibles plazos intradiarios de lanzamiento y los posibles tiempos de espera, aplicando el criterio de error mínimo descrito. En la figura puede apreciarse como la predicción con la estrategia intradiaria gana relevancia cuando el plazo intradiario de lanzamiento es superior a una hora y media para tiempos de espera intermedios comprendidos entre los 15 y los 255 minutos. Por otra parte, se aprecia un mejor desempeño de la estrategia de predicción intradiaria ante errores cuadráticos, concluyendo que evita en mayor medida valores atípicos del error.

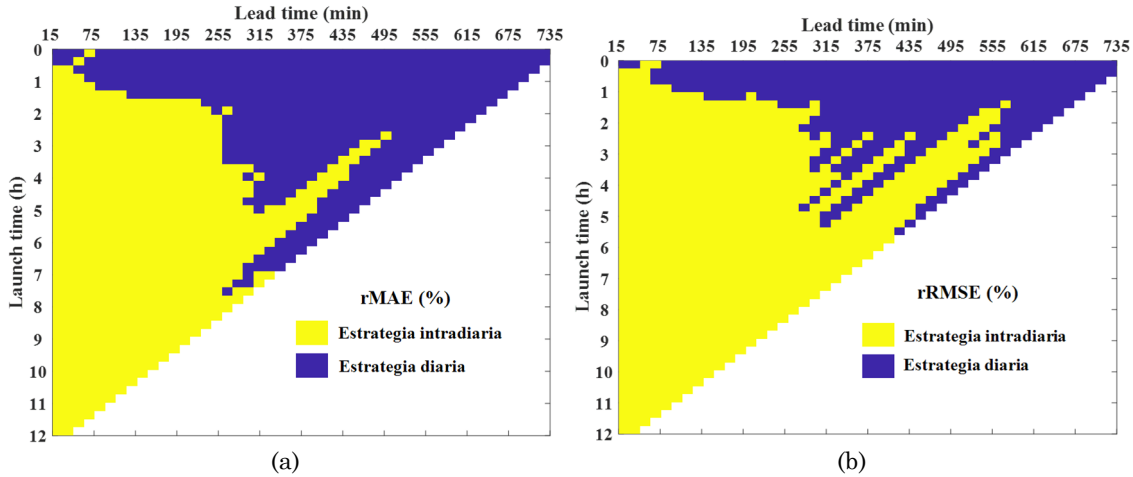


Figura 41.- Selección de la estrategia de predicción en función del tiempo de espera y el plazo intradiario de lanzamiento, para diferentes medidas del error consideradas: (a) rMAE; y (b) rRMSE

En función del evaluador del error a minimizar por el diseñador, la predicción final realiza una hibridación que selecciona entre ambas estrategias a lo largo del día para cada valor del plazo intradiario de lanzamiento contemplado. Es recomendable hacer uso del anexo I para ayudar en la decisión de qué evaluador es el más adecuado para cada problema en concreto, ya que cada caso de uso persigue un objetivo diferente y necesita minimizar una medida del error específica. También es importante remarcar que la Figura 41 muestra el error cometido para todo el histórico, sin embargo, existe la posibilidad de dividir las predicciones en tipos de días en función del CCF para optimizar la predicción en mayor medida si el problema a resolver lo requiere.

9.2 Caso de estudio: BEMS

Una vez descritas las predicciones de potencia bajo diferentes horizontes temporales, seleccionando la estrategia más adecuada en función del plazo intradiario de lanzamiento y el tiempo de espera, se propone el caso de estudio de un BEMS localizado en el nodo real de pruebas de la Universidad de Alcalá, detallado en el apartado 3.1. El BEMS utiliza las predicciones de potencia para gestionar el uso de las baterías, considerando como máximo un único ciclo de carga y descarga al día. Esto, junto al estado de carga (*State of Charge*, SoC) máximo y mínimo de las baterías, que se fija en $SoC_{máx} = 0.7$ y $SoC_{mín} = 0.2$ respectivamente, favorece a la vida útil de las baterías. El nodo de pruebas permite inyectar a red y se configura para una carga fija de 2 kW frente a los 3 kW_p de potencia fotovoltaica instalada, para fomentar que en ciertos momentos la producción fotovoltaica sea mayor que la carga. De esta forma, se implementa el caso más completo, en el que se dispone de excedentes de energía fotovoltaica que pueden ser utilizados bien para carga de las baterías o para inyectar a la red. La corriente máxima en el banco de baterías se fija en $I_{lím} = 30\ A$, trabajando a una tensión nominal de $U_{nom} = 50\ V$, lo que permite fijar un rendimiento farádico de carga igual a uno en la zona de trabajo de $SoC_{máx} = 0.7$ y $SoC_{mín} = 0.2$. En descarga sí se tiene en cuenta el rendimiento de la batería, que depende principalmente de

la corriente de descarga de la misma. Corrientes pequeñas proporcionan un rendimiento cercano a uno, mientras que corrientes mayores hacen que este rendimiento vaya disminuyendo. El precio de compra y venta de la energía utilizado es el del mercado eléctrico minorista español, este precio lo marca el Operador del Mercado Ibérico de la Energía (*Nominated Electricity Market Operator*, OMIE) y la Red Eléctrica de España (REE). A modo de ejemplo, a continuación, se expone la planificación del BEMS durante un día completo.

La Figura 42 muestra la planificación inicial del BEMS durante la noche del 16 de marzo de 2021, distinguiendo entre el precio de compra y venta de energía a la red, escalado por un factor 10 para facilitar su análisis respecto al resto de magnitudes, trazado en línea roja continua y discontinua respectivamente; la predicción de potencia fotovoltaica utilizada para la toma de decisiones en verde; la potencia planificada para ser inyectada a red mediante un área rosa; la potencia solicitada por las cargas fijas de la instalación en negro y la potencia planificada de carga y descarga de las baterías según áreas de color azul. Los instantes mostrados en el eje horizontal se referencian respecto al instante de máxima altitud solar, D, lanzando la primera predicción en un instante anterior a $D-6$, que se corresponde con un plazo intradiario de lanzamiento igual a 0 horas. Como veremos en figuras posteriores, la predicción del día siguiente se lanza a las 10:00 del día anterior en hora solar, sin embargo, en este caso se opta por iniciar el análisis una vez finalizadas las horas de sol del día anterior. Este caso de estudio utiliza una resolución temporal horaria, al considerarla sufrientemente precisa para realizar la optimización económica del desempeño de la batería con el conjunto de la instalación y no abordar problemas de calidad de red.

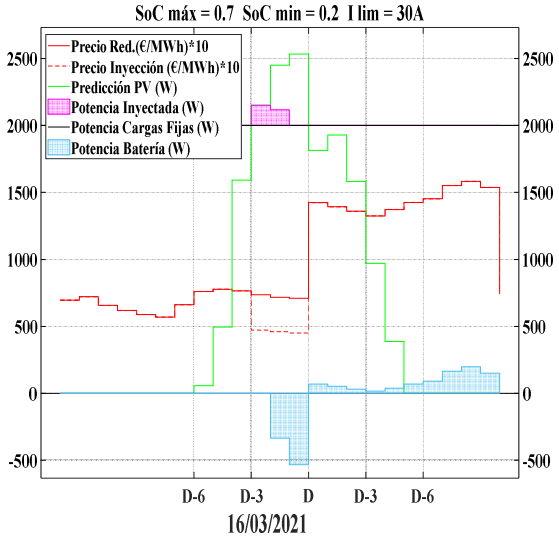


Figura 42.- Planificación inicial del BEMS durante la noche del 16 de marzo de 2021, seleccionando el proceso de carga y descarga del BESS al usar la predicción de estrategia diaria.

La estrategia de predicción de potencia mostrada es la diaria, puesto que aún no se dispone de medidas de irradiancia para ese mismo día que permitan lanzar, con una incertidumbre adecuada, la predicción intradiaria.

Esta predicción es utilizada por el BEMS para planificar la estrategia general de carga y descarga de la batería, teniendo en cuenta los precios de compra de la energía y la predicción de excedentes de potencia fotovoltaica esperados al día siguiente y su precio de venta a red. Una vez amanece, las estrategias diaria e intradiaria se utilizan simultáneamente de acuerdo a la Figura 41, permitiendo ajustar en tiempo real la estrategia del BEMS a posibles errores cometidos en la predicción.

En la Figura 42 se aprecia que las baterías se planifican para cargar en momentos en los que se dispone de excedentes, debido a que el precio de venta de energía es menor que cualquiera de los precios de compra de ese día. A continuación, se descargan aprovechando los momentos del día en el que el precio de la energía es más elevado. Cabe destacar que, en los momentos de exceso energético fotovoltaico, la estrategia optimizada del BEMS tiene en cuenta el precio de inyección a red de estos excedentes, ya que, si utiliza los excedentes para cargar el BESS en esas horas, el precio de compra de la energía se modificaría por el precio de inyección a red, que pasaría a ser la pérdida de oportunidad de inyección a red. De esta forma, la estrategia del BEMS puede considerar más eficiente económicamente cargar el BESS en esos momentos de excedentes, inyectar la energía a la red, o incluso realizar un escenario intermedio donde una parte de la energía se usa en la carga y otra parte se inyecta a red, según se observa en el día mostrado en el tramo $D-3$ a D.

Posterior al amanecer la estrategia va actualizando la predicción del día con la llegada de nuevas medidas, utilizando conjuntamente la estrategia diaria e intradiaria de acuerdo a la Figura 41. La Figura 43 muestra la actualización de estas predicciones a lo largo del día para un plazo intradiario de lanzamiento de dos y tres horas, que se corresponden con el instante $D-4$ y $D-3$ en el eje horizontal, respectivamente. Se observa que la estrategia del BEMS se ve modificada, ya que se esperaba un exceso energético en el instante $D-3$ que posteriormente es corregido, modificando la gestión energética del día a tiempo.

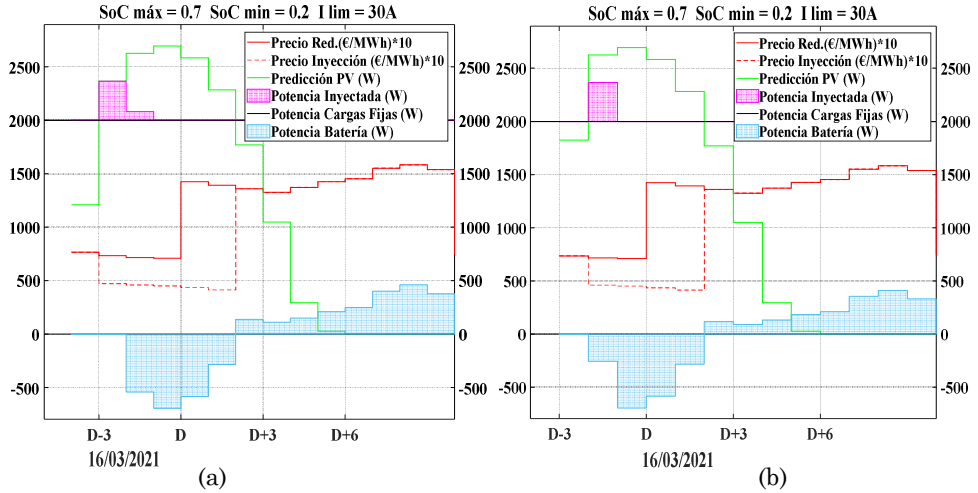


Figura 43.- Planificación del BEMS para un plazo intradiario de lanzamiento de (a) dos horas y (b) tres horas, el 16 de marzo de 2021, modificando el proceso de carga y descarga del BESS con la llegada de nuevas medidas.

Llegadas las 10:00 en hora solar, correspondiente al instante $D-2$ en las gráficas, el mercado diario de la energía realiza la subasta de los precios al día siguiente, permitiendo al BEMS ampliar el horizonte temporal de planificación del BESS. La Figura 44 muestra la actualización en la predicción para un plazo intradiario de lanzamiento de cuatro horas, que se corresponde con el instante $D-2$ en el eje horizontal, utilizando ahora también la estrategia diaria del día siguiente, obteniendo así la planificación del 17 de marzo de 2021. Esto permite a la instalación gestionar la carga, o descarga si se diera un precio inusualmente alto, durante la noche, obteniendo un primer análisis de la estrategia a realizar al día siguiente, al igual que se hizo en la Figura 42 para el 16 de marzo de 2021.

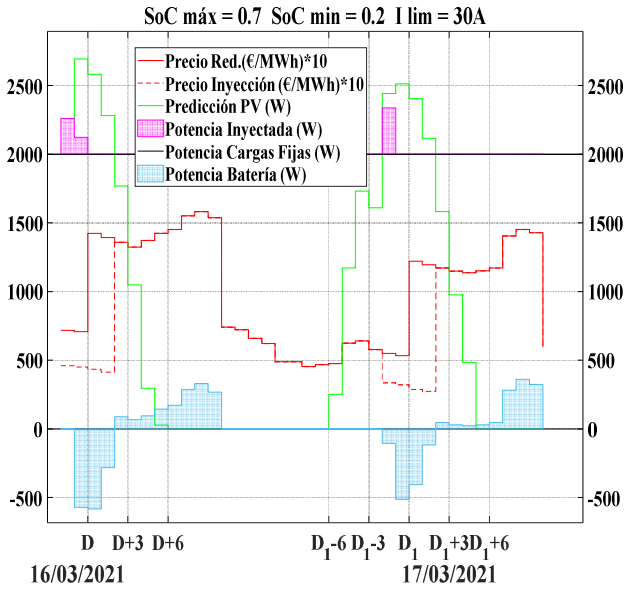


Figura 44.- Planificación del BEMS para un plazo intradiario de lanzamiento de cuatro horas, el 16 de marzo de 2021, momento a partir del que también se lanza la estrategia diaria del día siguiente.

A partir de este instante la planificación del BEMS se sigue realizando de forma idéntica hasta completar el día bajo estudio. En la Figura 45 se puede comprobar la evolución del resto del proceso de gestión del BEMS para el resto del día. La Figura 45a muestra la predicción de potencia fotovoltaica para un plazo intradiario de lanzamiento de 5 horas, la Figura 45b para un plazo intradiario de lanzamiento de 6 horas, así sucesivamente, hasta finalizar el día con la Figura 45f para un plazo intradiario de lanzamiento de 10 horas. Se aprecia el ajuste fino de la estrategia realizada a lo largo del día, modificando el planteamiento conforme llegan nuevas medidas y se comprueba que la estrategia previa se debe ajustar a las nuevas predicciones. Además, se observa que la estrategia de predicción para el día siguiente no se ve modificada, ya que en este caso la predicción de potencia se mantiene a lo largo del tiempo y se conserva el planteamiento inicial.

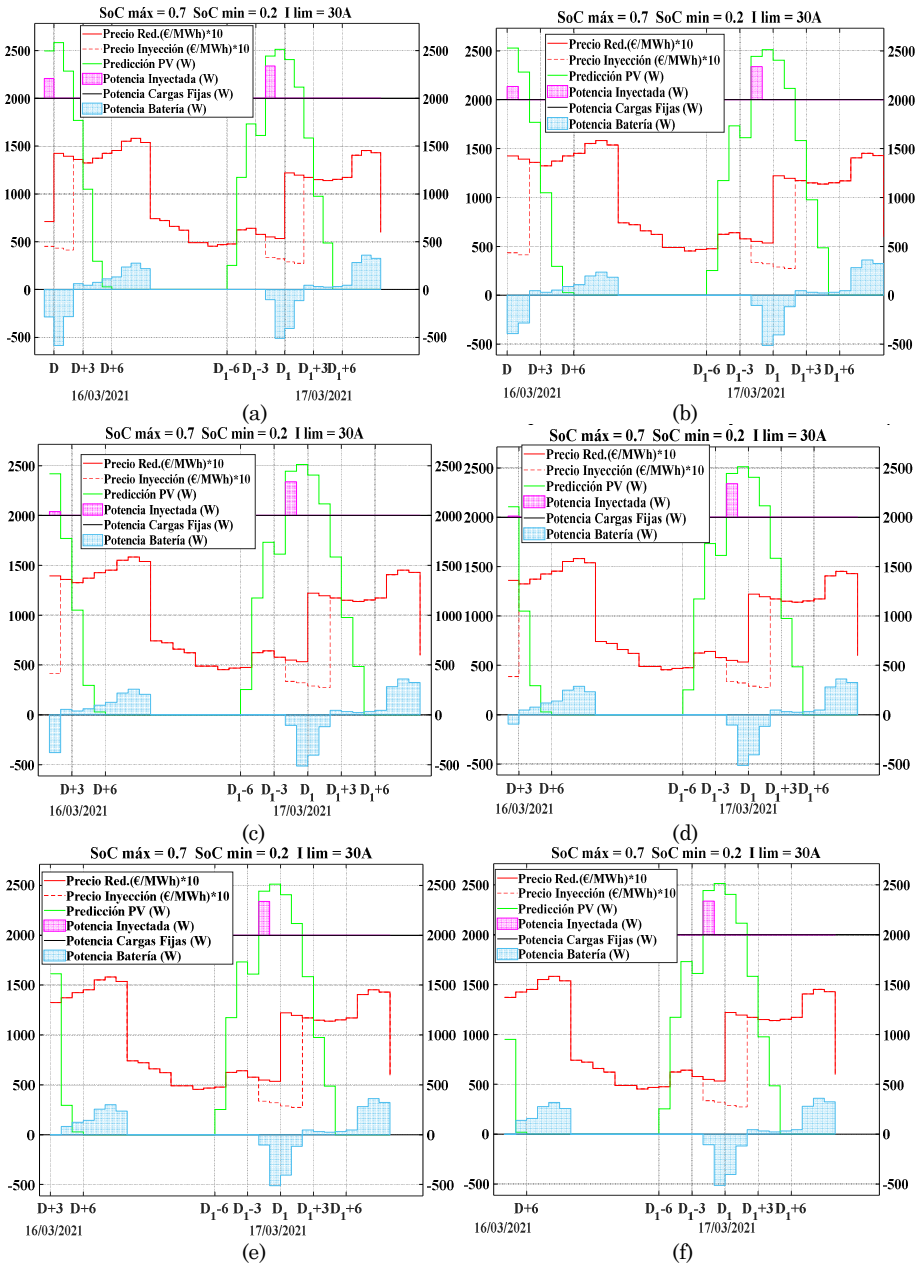


Figura 45.- Resto de planificación del BEMS el 16 de marzo de 2021, para un plazo intradiario de lanzamiento que va desde 5 horas en (a) hasta 10 horas en (f).

Finalmente, para comprobar la potencialidad de la propuesta, la Figura 46 muestra el resultado obtenido una vez realizado el proceso de planificación del BEMS mostrado previamente, durante cinco días consecutivos, del 14 al 18 de marzo de 2021. En este caso, la potencia fotovoltaica dibujada es la que se ha medido en la instalación. Se observan unos grandes resultados, consiguiendo el BEMS gestionar un ciclo diario de carga y descarga de baterías de forma eficiente al utilizar la predicción con menor incertidumbre en tiempo real. Apoyándose en la predicción puede decidir con anterioridad cómo gestionar la batería, maximizando el beneficio económico de la instalación.

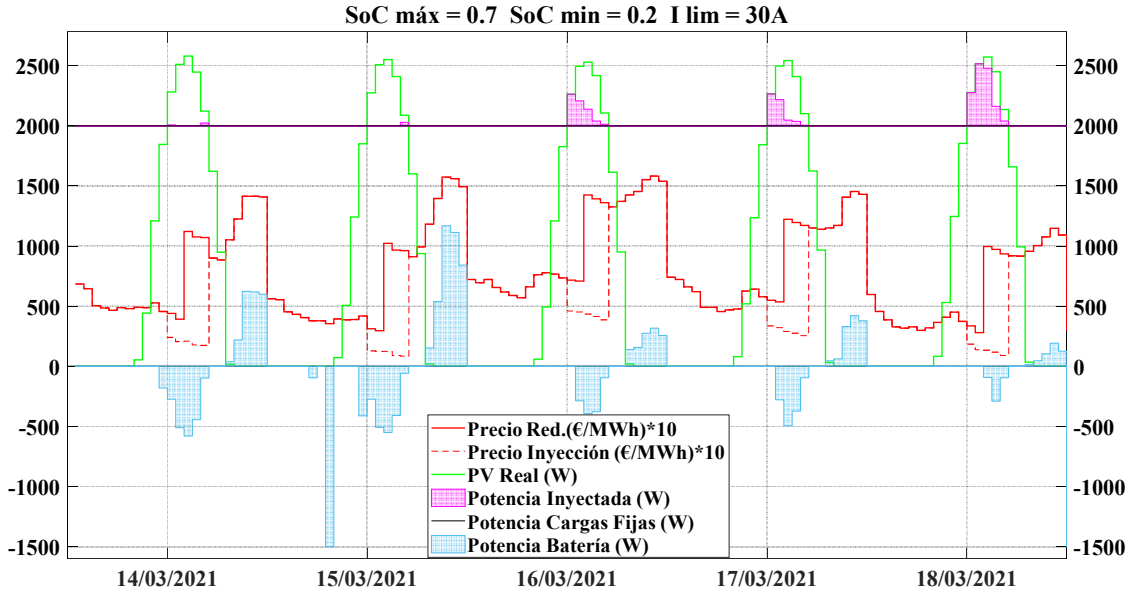


Figura 46.- Resultados de la planificación final del BEMS durante 5 días consecutivos, del 14 al 18 de marzo de 2021. Se observa que las baterías realizan un ciclo de carga y descarga al día, optimizando el beneficio económico de la instalación.

En todos los días seleccionados se observan excesos energéticos de dimensiones similares, sin embargo, la actuación del BEMS muestra diferencias en cada uno de ellos. El 14 de marzo apenas se inyecta energía a la red y el exceso de generación fotovoltaica se utiliza para cargar las baterías, descargando posteriormente esta energía en instantes donde el precio de la red es elevado. A la noche siguiente, el 15 de marzo, el precio de red se vuelve económico y el BEMS carga parcialmente las baterías, considerando también el aprovechamiento energético de la instalación fotovoltaica al día siguiente. El 16 de marzo se observa que el BEMS decide inyectar parte del exceso energético a red en instantes donde el precio de venta de la energía es elevado, almacenando parte de este exceso para utilizarlo en las horas más caras. Existen días que no utilizan toda la capacidad de las baterías, ya que el diferencial entre los precios de compra de red del día es pequeño y no resulta rentable utilizar la batería, al tener en cuenta la eficiencia del proceso de carga y descarga de la batería y el coste por envejecimiento asociado a la misma.

En la Figura 47a se observa la degradación de la capacidad total de la vida útil del BESS, a lo largo del proceso descrito en la figura anterior. Se comprueba que la reducción de la vida útil es muy baja, apenas un 0.159%, que resulta en una pérdida de un 11.60% anual. En el caso de estudio se asume que la vida útil máxima del BESS es de 7 años, ya que su propósito principal es servir de respaldo en caso de caída de la red eléctrica, así que pasado estos años se considera que la degradación de las baterías es tal que necesitan ser reemplazadas. Por ello, a lo largo de 7 años, el uso del BESS perdería un 81.20% de su vida útil, al seguir la estrategia del BEMS, quedando suficiente vida útil para realizar su función de respaldo cuando sea necesario. Dado que a los 7 años se va a reemplazar este BESS y debe estar presente obligatoriamente para su función de respaldo, no se tiene en cuenta el coste que supone usarlo hasta una determinada profundidad de

descarga (*Depth of Discharge*, DoD) en cada ciclo cada día. Superado ese umbral de DoD diario, el coste de uso de las baterías correspondiente a la superación del umbral es tenido en cuenta a la hora de calcular el ahorro diario en la estrategia del BEMS, ya que podría reducir el periodo de tiempo en el que el BESS tendría que ser remplazado.

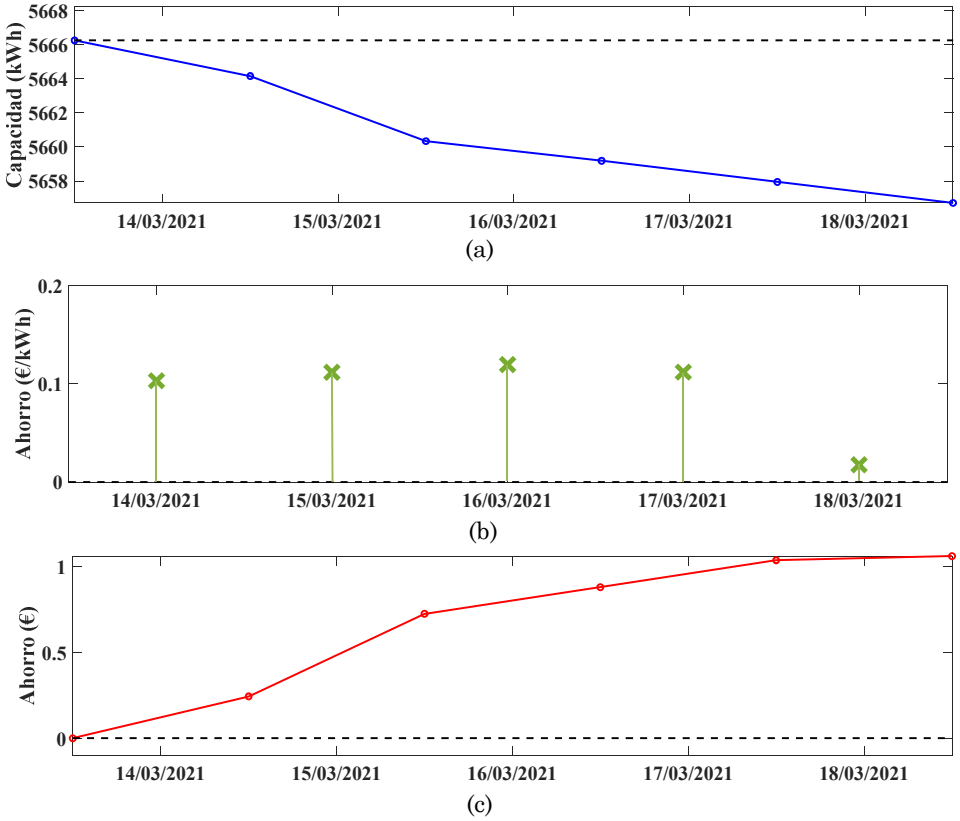


Figura 47.- Resultados obtenidos tras la gestión del BEMS durante los días mostrados en la Figura 46: (a) vida útil del BESS medida según su capacidad; (b) ahorro económico acumulado en $€/kWh$; y (c) ahorro económico acumulado en €.

La estrategia del BEMS trabaja en unidades relativas de $€/kWh$ para realizar la optimización, Figura 47b, lo que se traduce diariamente en un beneficio positivo en unidades absolutas de €, Figura 47c. Así, en la Figura 47c, se muestra el ahorro absoluto acumulado a lo largo de estos días, siendo de algo más de 1.06 €. Realizando unos simples cálculos se obtiene una estimación de ahorro acumulado medio a lo largo de los 7 años de vida útil del BESS de unos 541.66 €, que permite amortizar las baterías adquiridas. Es de reseñar que amortizar el BESS de una instalación eléctrica supone un reto en la mayoría de los casos, jugando un papel de gran importancia en un entorno VPP, de modo que los resultados mostrados son bastante favorables.

9.3 Conclusiones

Este capítulo hibrida los modelos diario e intradiario desarrollados en capítulos previos y proporciona ciertos resultados dentro de un entorno experimental. En primer lugar, en el apartado 9.1, se selecciona la estrategia de menor incertidumbre en función del plazo intradiario de lanzamiento y el tiempo de espera, seguidamente, en el apartado 9.2, se

realiza un caso de estudio de gestión de baterías por parte de un BEMS en función de las predicciones realizadas.

El estudio realizado muestra que la estrategia diaria es de utilidad en plazos intradiarios de lanzamiento pequeños y tiempos de espera elevados. La estrategia intradiaria muestra una gran precisión en plazos intradiarios de lanzamiento superiores a una hora y media, y domina ante medidas cuadráticas del error, por lo que se destaca su uso para evitar errores atípicos en la predicción.

El caso de estudio utiliza las predicciones de potencia fotovoltaica para gestionar las baterías del nodo de VPP, situado en la Universidad de Alcalá, optimizando el beneficio económico sin repercutir en su vida útil. Los resultados muestran cómo la planificación del BEMS se va modificando conforme se actualizan las predicciones a lo largo del día. Se comprueba que la estrategia propuesta consigue aprovechar los ciclos de carga y descarga, amortizando las baterías a lo largo de su vida útil y permitiendo disponer de este tipo de equipos en un entorno VPP sin coste adicional.

10 Conclusiones

Una vez explicada y analizada la estrategia de predicción de potencia fotovoltaica, se describen unas conclusiones finales a modo de resumen de los resultados obtenidos. Estas conclusiones se enfocan en los resultados más relevantes obtenidos, las contribuciones aportadas, las limitaciones del trabajo y las futuras líneas de investigación relacionadas con la tesis. El apartado 10.1 realiza una discusión de los resultados obtenidos en la estrategia de predicción desarrollada en la tesis, mientras que el apartado 10.2 se enfoca en puntualizar las contribuciones, limitaciones y futuras líneas de investigación.

10.1 Discusión de resultados

La predicción de potencia fotovoltaica es de gran importancia para la integración de instalaciones fotovoltaicas en entornos VPP de forma efectiva, adaptándose al funcionamiento del EMS para facilitar las tareas de gestión. Al conocer la energía producida en cada nodo de VPP, es posible optimizar el beneficio generado en los intercambios energéticos con el operador de red. Sin embargo, es difícil obtener grandes precisiones cuando existen factores que limitan la predicción de la potencia fotovoltaica: (i) la necesidad de recoger información en varios nodos que se encuentran dispersos en una amplia área; (ii) la disponibilidad de históricos extensos para la predicción en instalaciones que no tienen esta información, al ser nuevas o no estar adaptadas para este fin; y (iii) el acceso a la información con un coste asociado hace inviable su incorporación a los entornos VPP a medida que el número de nodos incrementa. No hay una manera eficiente de superar estas limitaciones con un modelo de predicción simple.

Por estos motivos, esta tesis presenta una estrategia para predecir la potencia fotovoltaica a diferentes horizontes temporales con información recogida de la propia instalación y medidas satélite gratuitas disponibles a lo largo del territorio español. Concretamente, las variables empleadas para las predicciones consisten en: (i) información de GHI y los equipos instalados en un nodo real de pruebas situado en la Escuela Politécnica Superior de la Universidad de Alcalá; (ii) medidas de GHI recogidas de un grupo de estaciones meteorológicas a nivel de suelo operando en la Comunidad de Madrid; (iii) mapas de predicción de temperatura ambiente y nubosidad generados por la AEMET vía NWP; y (iv) históricos de irradiancia en cualquier punto geográfico proporcionados por el CAMS. La información recopilada es evaluada para conocer su relevancia en la estrategia y se procesa para adecuarse a la predicción, generando nuevas variables como el CCF, encargado de identificar las nubes que producen sombra en la instalación fotovoltaica.

En primer lugar, se genera una estrategia de predicción indirecta que trabaja con un horizonte temporal diario con una resolución temporal horaria para conocer la GHI en el nodo real de pruebas y el entorno VPP.

Para ello, se emplea una estrategia híbrida que utiliza la predicción de dos modelos: (i) un modelo propuesto basado en horas similares; y (ii) un modelo basado en ANNs. Ambas predicciones son ponderadas dinámicamente con unos pesos que se ajustan en función del tipo de día, seleccionado a partir del CCF, y el MAE de las predicciones pasadas. En históricos reducidos, el modelo basado en horas similares produce las mayores precisiones, mientras que el modelo basado en ANNs reduce su error a lo largo del histórico hasta superar a este. En históricos de un mes, el modelo basado en horas similares mejora al modelo basado en ANNs en un 16.32%, mientras que en históricos de dos meses, la mejora se reduce a un 9.07%, esta mejora se invierte a partir de históricos de cuatro meses. La estrategia híbrida y el modelo basado en horas similares han demostrado una mayor precisión en históricos reducidos que otras técnicas relacionadas: el modelo persistente, el modelo basado en días similares y el modelo basado en ANNs.

Dentro de un entorno VPP emulado con siete nodos, el modelo basado en horas similares reduce el error en mayor medida que el resto, por lo que gana especial relevancia en este tipo de entornos. El modelo basado en horas similares reduce el rMAE del entorno VPP en un 13.27%, con respecto a la media de los errores de los nodos, y un 14.56% para el rRMSE. Esta reducción es de un 6.57% en el rMAE y un 7.28% en el rRMSE para el modelo basado en ANNs, y de un 10.02% en el rMAE y un 10.79% en el rRMSE para el modelo híbrido. Para todo el histórico contemplado de seis meses, el modelo híbrido genera una medida del error de $rMAE = 21.37\%$ y $rRMSE = 30.99\%$, ofreciendo una gran precisión en todo momento.

Una vez realizada la predicción bajo un horizonte temporal diario, en segundo lugar, se genera una estrategia de predicción indirecta que trabaja con un horizonte temporal intradiario y una resolución temporal de 15 minutos para predecir la GHI en la instalación. Esta predicción emplea un modelo basado en RNN-LSTMs para predecir la GHI a partir de un histórico de 5 años de irradiancia, usando medidas vía satélite del CAMS para completar el histórico. La predicción se actualiza conforme se obtienen nuevas medidas en la instalación para optimizar la precisión, definiendo el plazo intradiario de lanzamiento y el tiempo de espera para conocer las zonas donde el error se reduce considerablemente.

La estrategia muestra una gran precisión en plazos intradiarios de lanzamiento superiores a una hora y media, aumentando el error proporcionalmente conforme el tiempo de espera aumenta. Se concluye que se necesita disponer de información lo más actualizada posible en la instalación para generar una precisión óptima. En comparación con estudios procedentes de la literatura se alcanzan errores similares con respecto a otros modelos de aprendizaje profundo, mejorando a los más tradicionales. Por último, los modelos exógenos generan una mayor precisión que el modelo propuesto, sin embargo, existen limitaciones a la hora de disponer de toda la información que requieren estos modelos.

Una vez predicha la GHI bajo diferentes horizontes temporales, el siguiente paso utiliza un modelo de la planta y medidas de temperatura ambiente de los NWP de la AEMET para conocer la potencia fotovoltaica generada en la instalación, mediante un método analítico desarrollado a partir de artículos de la literatura. Los resultados se han contrastado con medidas reales de potencia e irradiancia en la instalación para conocer la precisión del método analítico. Se obtiene un error bastante reducido que varía entre el 2.54% y el 4.03% en términos del rMAE y entre el 3.44% y el 5.95% en términos del rRMSE, siendo consistente con otros artículos de la literatura. En cuanto a las predicciones de GHI, en el horizonte temporal intradiario, la mejora del entorno VPP con respecto al nodo real de pruebas es de un 12.37% con respecto al rMAE y un 11.84% con respecto al rRMSE. Por otra parte, para el horizonte temporal diario, la mejora es de un 10.46% con respecto al rMAE y un 10.65% con respecto al rRMSE. Las conclusiones observadas en las predicciones de GHI se mantienen al convertir estos valores a potencia, conservando el potencial del modelo.

Con estas predicciones de potencia fotovoltaica, es posible generar intervalos de predicción que ofrezcan un rango de potencias entre las que se encuentre la medida asociada, bajo un nivel de confianza definido. El modelo considera una distribución Laplaciana del error, dividiendo en grupos de baja dependencia para mejorar la precisión de los intervalos. En concreto, se separa el histórico en función del CCF para el horizonte temporal diario, y en función del tiempo de espera, el plazo intradiario de lanzamiento y el tipo de día para el horizonte temporal intradiario. El PICP se ha utilizado para conocer si los intervalos son adecuados para el nivel de confianza seleccionado del 80%, observando unos buenos resultados. Existen zonas en la estrategia intradiaria en los que el PICP muestra una mala representación de los intervalos, sin embargo, se corresponde con momentos de baja producción fotovoltaica cercanas al anochecer, por lo que puede despreciarse del estudio global.

Finalmente, la estrategia unifica la predicción de potencia fotovoltaica diaria con la intradiaria. El estudio realizado muestra que la predicción diaria es de utilidad en plazos intradiarios de lanzamiento pequeños y tiempos de espera elevados. La predicción intradiaria muestra una gran precisión en plazos intradiarios de lanzamiento superiores a una hora y media, y domina ante medidas cuadráticas del error, por lo que se destaca su uso para evitar errores atípicos en la predicción.

Toda la estrategia desarrollada se emplea en un caso de estudio que utiliza las predicciones de potencia fotovoltaica para gestionar las baterías del nodo de pruebas de la Universidad, optimizando el beneficio económico sin repercutir en su vida útil. Los resultados muestran cómo la planificación del BEMS se va modificando conforme se actualizan las predicciones a lo largo del día. Se observa que las baterías instaladas se amortizan a lo largo de su vida útil, permitiendo utilizar estos elementos sin una pérdida económica asociada dentro de un entorno VPP donde resultan esenciales.

10.2 Contribuciones, limitaciones y líneas futuras

Las contribuciones de la tesis se pueden desglosar en diferentes puntos. En relación a su entorno, (i) por una parte, implementar la estrategia en un nodo real ha permitido generar resultados experimentales que validan la eficiencia de la predicción, (ii) por otra parte, al emular una VPP se contrastan los resultados entre la fase experimental de forma individual y la agregación de nodos, obteniendo resultados cuantitativos de la mejora producida por la estrategia en dicho entorno. Con respecto a la recopilación de información, (iii) en primer lugar, destacar el uso de variables meteorológicas de libre acceso que son publicadas regularmente en la web, permitiendo generar predicciones de forma gratuita, (iv) además, la definición del CCF, que define la cantidad de nubes que ocultan el sol en la instalación y es un indicativo del tipo de día. De acuerdo a la estrategia de predicción diaria, (v) por una parte, el modelo basado en horas similares ofrece una gran precisión ante históricos escasos, encontrados en nuevas instalaciones o nuevos nodos VPP y, (vi) por otra parte, la hibridación del modelo basado en horas similares con un modelo basado en redes neuronales permite generar una elevada precisión de forma independiente al histórico. Finalmente, (vii) el uso de intervalos de predicción para definir la incertidumbre permite seleccionar los grupos óptimamente en función del CCF y otras variables temporales consideradas.

La mayor limitación de la tesis está relacionada con los mapas de NWP de la AEMET, ya que se centra en el territorio español. Los mapas de temperatura ambiente y nubosidad son fácilmente accesibles y gratuitos en la mayoría de países, sin embargo, no disponer de esta información eliminaría el potencial de la estrategia desarrollada. Por otra parte, la normativa vigente puede suponer un problema para la integración de este tipo de entornos, ya que depende de la situación político-social del territorio bajo estudio, entre otros factores cambiantes a lo largo del tiempo.

Los trabajos futuros de mayor interés a partir de la tesis desarrollada se enfocan en la generación de otros casos prácticos de estudio en el entorno de pruebas. La estrategia se ha desarrollado en gran detalle, pero su aplicabilidad no ha sido explotada. Este aspecto se encuentra actualmente en fase de desarrollo, abordando el caso de estudio del apartado 9.2 con un mayor nivel de complejidad, y publicando este trabajo en una revista de gran prestigio para generar visibilidad al potencial que puede ofrecer.

Por otra parte, los intervalos de predicción se han construido a partir de una predicción paramétrica que asume una distribución Laplaciana del error, definiendo rangos de potencia adecuadamente al separar el histórico en grupos de relevancia, en función del tipo de día y variables temporales. Sin embargo, a parte de este modelo, es recomendable utilizar predicciones probabilísticas no paramétricas que aportan una mayor cantidad de información y definen rangos bajo un mayor nivel de detalle. Por su

popularidad, se recomienda analizar modelos de regresión cuantílica, entre otros estudios destacados en la literatura [108]. Adicionalmente, este tipo de modelos permiten evaluar un mayor número de métricas del error, aportando información adicional al PICP, evaluando características como la fiabilidad, la nitidez y la resolución de la predicción.

Por último, la estrategia de predicción ha sido aplicada a un nodo de pruebas situado en la Universidad de Alcalá que cuenta con una potencia pico de 3 kW_p, alimentando una carga fija con la ayuda de la red y un BESS. Sin embargo, este caso experimental puede ser explotado en mayor medida al considerar diferentes situaciones que permitan abordar un mayor número de pruebas. Entre estas, se destaca aumentar la potencia pico de la instalación, regular las cargas de la instalación, simular cortes eléctricos de formas diferentes, entre otras. Por otra parte, el entorno VPP ha sido emulado con estaciones meteorológicas en la Comunidad de Madrid, sin embargo, disponer de un entorno experimental de varios nodos interconectados entre sí que simulen una VPP real permitiría obtener conclusiones más consistentes y evaluar un mayor número de casos de estudio.

11 Conclusions

Once the PV power prediction strategy is explained and analysed, some final conclusions are described as a summary of the results obtained. These conclusions focus on the most relevant results, the contributions, the limitations and the future lines of research related to the thesis. Section 11.1 discusses the results obtained in the forecasting strategy developed in the thesis, while section 11.2 focuses on the contributions, limitations and future lines of research.

11.1 Discussion of results

PV power forecasting is of great importance for the effective integration of PV installations in VPP environments, adjusting the operation of the EMS to facilitate management tasks. By knowing the energy produced at each VPP node, it is possible to optimize the profit generated in energy exchanges with the grid operator. However, it is difficult to obtain high accuracies when there are other aspects that limit the prediction of PV power, between them: (i) the need to collect information in several nodes that are scattered over a wide area; (ii) the availability of extensive historical datasets for the prediction in installations that do not have this information because they are new or has not been adapted for this purpose; and (iii) the access to information with an associated cost makes it unfeasible to incorporate it into VPP environments as the number of nodes increases. There is no efficient way to overcome these limitations with a simple prediction model.

For these reasons, this thesis presents a strategy to predict PV power at different time horizons with information collected from the installation and free satellite measurements available throughout the Spanish territory. Specifically, the variables used for the prediction consist of: (i) information of GHI and installed equipment from a real node located at the Polytechnic School of the University of Alcala; (ii) GHI measurements collected from a group of ground level weather stations operating in the Community of Madrid; (iii) ambient temperature and cloudiness forecast maps generated by the AEMET via NWP; and (iv) irradiance historical datasets at any geographical point provided by the CAMS. All this information is evaluated and processed due to its relevance in the strategy, generating new variables such as the CCF, which indicates the amount of clouds that produces shadows on the PV installation.

Firstly, an hourly indirect day-ahead forecasting model is generated to predict the GHI in the PV facility and the VPP environment. For this purpose, a hybrid strategy that uses two different models: (i) a similar-hour based model; and (ii) an ANN-based model, is employed. Both predictions are dynamically weighted, adjusting the values according to the type of day, selected from the CCF, and the MAE of past predictions. In reduced historical datasets, the similar hour-based model produces the highest

accuracies, while the ANN-based model reduces its error over the historical dataset until outperforming it. For a historical dataset of one month, the similar hours-based model improves over the ANN-based model by 16.32%, this improvement is reduced to 9.07% for a four-month historical dataset, and the ANN-based model outperforms the similar hours-based model for a historical dataset of four months onwards. The hybrid strategy and the similar hours-based model have shown a better accuracy in reduced historical datasets, compared to other related techniques: the persistent model, the similar days-based model and the ANN-based model.

For an emulated VPP with seven nodes, the similar hours-based model has a higher reduction of the error than the others, therefore, gaining special relevance in this type of environment. The similar hours-based model reduces the rMAE of the VPP by 13.27%, and by 14.56% in terms of the rRMSE. This reduction is 6.57% in terms of the rMAE and 7.28% in terms of the rRMSE for the ANN-based model, and 10.02% in terms of the rMAE and 10.79% in terms of the rRMSE for the hybrid model. Under a historical dataset of six months, the measurements of the error for the hybrid model is $rMAE = 21.37\%$ and $rRMSE = 30.99\%$, providing a high accuracy.

Once the day-ahead prediction is considered, an intra-day prediction strategy is generated with a time resolution of 15 minutes. This prediction employs a RNN-LSTM-based model to predict the GHI from a 5-year historical dataset of irradiance, using the CAMS satellite measurements to increase the dataset. The prediction is updated as new measurements are available at the facility to optimize accuracy, defining the launch time and the lead time to find the areas where the error is significantly reduced.

The strategy shows high accuracy at launch times greater than one hour and a half, increasing the error proportionally as the lead time increases. It should be noted that it is necessary to have the most up-to-date information available at the facility to generate an optimal accuracy. In comparison with studies from the literature, similar errors are obtained with respect to other deep learning models, improving the more traditional ones. Finally, exogenous models generate higher accuracies than the proposed model, however, there are some limitations to consider when collecting the information necessary in exogenous models.

Once the GHI is predicted under different time horizons, the next step uses a model of the PV facility and ambient temperature measurements from the NWP maps to obtain the PV power generated in the installation, using an analytical method developed from articles in the literature. The results are contrasted with real measurements of power and irradiance in the installation to study the accuracy of the analytical method. A reduced error is obtained, varying between 2.54% and 4.03% in terms of the rMAE and between 3.44% and 5.95% in terms of the rRMSE, being consistent with other articles from the literature. Considering the predicted GHI, for the intra-day solar power forecasting, the improvement of the VPP

environment with respect to the real node is 12.37% with respect to the rMAE and 11.84% with respect to the rRMSE. On the other hand, for the day-ahead strategy, the improvement is 10.46% with respect to the rMAE and 10.65% with respect to the rRMSE. The conclusions observed in the GHI predictions are maintained, preserving the potential of the model.

From the PV power forecasts, prediction intervals are generated, providing information about the plausible range of PV power values obtained at the site for a defined confidence level. The results show that the intervals properly represent the selected confidence level. The model considers a Laplacian distribution of the error, dividing into groups of a low dependence to improve the accuracy of the intervals. Specifically, the historical dataset is separated according to the CCF for the day-ahead strategy, and according to the lead time, the launch time and the type of day for the intra-day strategy. The PICP is used to evaluate if the intervals are adequate for the selected confidence level of 80%, obtaining good results. In the intra-day strategy there are some areas where the PICP shows a poor representation of the intervals, however, it corresponds to instants of low PV production, close to sunset, being discarded from the overall study.

The final step unifies the day-ahead strategy with the intra-day strategy. The study shows that the day-ahead strategy has a high accuracy on small launch times and large lead times. The intraday strategy shows a high accuracy on launch times higher than one hour and a half, and dominates on quadratic measurements of the error, highlighting its use to avoid atypical prediction errors.

The whole strategy is employed under a case study that uses the PV power predictions to manage the batteries of the real node located at the University, optimizing the economic benefit without influencing its lifespan. The results show how the BEMS scheduling is modified as the predictions are updated throughout the day. It is also observed that the batteries are amortized, allowing the use of these elements without an associated economic loss on a VPP environment, where they are essential.

11.2 Contributions, limitations and future lines of research

The contributions of the thesis can be divided into different points. In relation to its environment, (i) on the one hand, implementing the strategy in a real node has allowed to generate experimental results that validate the efficiency of the prediction, (ii) on the other hand, by emulating a VPP, the results between the experimental phase individually and the aggregation of nodes are contrasted, obtaining quantitative results of the improvement produced by the strategy in that environment. With respect to the dataset, (iii) on the one hand, the use of free access meteorological variables that are regularly published on the web, allowing to generate forecasts without a cost, (iv) on the other hand, the definition of the CCF, which defines the

amount of clouds that shade the installation, being an indicative of the type of day. According to the day-ahead prediction strategy, (v) on the one hand, the model based on similar hours offers high accuracy under data scarcity, typical in new installations or new VPP nodes and, (vi) on the other hand, the hybridization of the model based on similar hours with a model based on neural networks allows generating high accuracy independently of the historical data. Finally, (vii) the use of prediction intervals to define the uncertainty allows to select the groups optimally according to the CCF and other time variables considered.

The major limitation of the thesis is related to the NWP maps provided by the AEMET, since it focuses on the Spanish territory. Ambient temperature and cloudiness maps are easily accessible and of free access in most countries, however, not having this information would delete the potential of the developed strategy. On the other hand, regulations may pose a problem for the integration of this type of environments, since it depends on the social-political situation of the territory under study, among other factors that can change over time.

Future work focuses on the generation of other practical case studies in the real node environment. The present strategy is developed in great detail, but its applicability has not been exploited. This aspect is currently under development, addressing the case study in section 9.2 with a higher level of complexity, also publishing this work in a prestigious journal to generate visibility to the potential it can offer.

On the other hand, prediction intervals are defined with a parametric prediction that assumes a Laplacian distribution of the error, defining PV power ranges appropriately by separating the historical dataset into groups of relevance, in terms of the type of day, the lead time and the launch time. However, apart from this model, it is advisable to use nonparametric probabilistic predictions that provide a greater amount of information, defining ranges in a more detailed way. Due to their popularity, it is recommended to analyze quantile regression models, among other studies highlighted in the literature [108]. Additionally, this type of models can evaluate a larger number of error metrics, providing additional information to the PICP, evaluating characteristics such as reliability, sharpness and resolution.

Finally, the prediction strategy is applied to a real node located at the University of Alcala with a peak power of 3 kW_p, feeding a fixed load with the help of the grid and a BESS. However, this experimental case can be exploited to a greater extent by considering different situations that allow a larger number of tests to be addressed. These include increasing the peak power, regulating the loads of the installation, simulating different power outages, among others. Also, the VPP is emulated with weather stations in the Community of Madrid, however, an experimental environment with several interconnected nodes simulating a real VPP allows to generate more consistent conclusions and to evaluate a greater number of case studies.

Anexos

Los anexos se ordenan de la siguiente forma: el anexo I describe los índices usados al medir el error cometido en las predicciones; el anexo II muestra las hojas de características de los equipos del nodo real de pruebas; el anexo III analiza la radiación solar y el efecto de las nubes al percibirse a nivel de suelo; y el anexo IV da una visión de las RNN, enfocando la aplicación de las RNN-LSTM en la tesis.

Anexo I: Medidas del error cometido

La precisión en la predicción puede ser interpretada según diferentes evaluadores, eligiendo uno u otro en función de la variable involucrada y el propósito de lo que se desea representar. Las expresiones seleccionadas en la tesis son las más empleadas en predicción de irradiancia [109], siendo también muy utilizadas de forma generalizada. Se distingue, por una parte, las medidas del error basadas en porcentajes (rMAE y rRMSE) frente a las dependientes de la escala (MAE y RMSE), por otra parte, las medidas del error medio (MAE y rMAE) frente al cuadrático (RMSE y rRMSE). En función del resultado a representar, se usará un evaluador u otro para comparar adecuadamente los modelos de predicción.

Históricamente, las medidas cuadráticas del error se han empleado más por su importancia en modelos estadísticos [110]. En la tesis, al ser estas medidas más sensibles a errores atípicos, se utilizan principalmente en optimización de modelos. Por otra parte, las medidas medias del error se emplean para contrastar la precisión entre modelos, ya que estudian el error cometido en todo el histórico de una forma más generalizada e intuitiva. Las medidas dependientes de la escala son útiles cuando se contrastan predicciones en un mismo rango, pero no se pueden emplear entre predicciones de diferentes órdenes de magnitud. Por el contrario, los errores porcentuales tienen la ventaja de no depender de la escala, por lo que son útiles para comparar predicciones de órdenes de magnitud diferentes.

Para el caso de estudio, se descartan las expresiones del error porcentual que dividen con respecto al valor real de la medida en el instante seleccionado, ya que son muy sensibles a valores cercanos a cero. En este documento, estas expresiones se denominan error absoluto medio porcentual (*Mean Absolute Percentage Error*, MAPE) y raíz del error cuadrático medio normalizado (*Normalized Root Mean Squared Error*, NRMSE), sin embargo, su definición puede ser diferente en otros artículos de la literatura.

$$MAPE = \frac{100}{T}\sum_{t=1}^{T}\frac{|Y_t - \hat{Y}_t|}{Y_t}$$

$$NRMSE = 100\,.\,\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\frac{Y_t - \hat{Y}_t}{Y_t}\right)^2}$$

Donde Y_t es la medida real e $\widehat{Y_t}$ es la predicción de dicha medida, siendo T el número total de medidas existentes en la predicción. En la tesis, en lugar de utilizar estas medidas para evaluar el error porcentual, se emplean el rMAE y el rRMSE, que dividen con respecto a la media de las medidas reales observadas en el histórico considerado.

$$rMAE = 100.\frac{\frac{1}{T}\sum_{t=1}^{T}\left|Y_t - \widehat{Y_t}\right|}{\frac{1}{T}\sum_{t=1}^{T} \mathrm{Y_t}}$$

$$rRMSE = 100.\frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \widehat{Y_t}\right)^2}}{\frac{1}{T}\sum_{t=1}^{T} \mathrm{Y_t}}$$

Las medidas dependientes de la escala utilizadas en la tesis son el MAE y el RMSE. Al depender de la escala, no es recomendable utilizarlas cuando se analicen diferentes estaciones en un mismo histórico y/o al comparar los errores medidos entre diferentes tipos de día.

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|Y_t - \widehat{Y_t}|$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \widehat{Y_t}\right)^2}$$

Otra medida ampliamente utilizada es el error de tendencia medio (*Mean Biased Error*, MBE), que es un indicativo del sesgo producido en la predicción. Sin embargo, esta expresión no se evalúa en la tesis ya que su valor puede malinterpretarse. Este ha sido el caso en predicciones que ofrecen un error absoluto muy elevado, pero compensado, de modo que no se representa adecuadamente la información. Por otra parte, en ciertos casos, este evaluador indica un offset en la predicción que puede resultar interesante de analizar.

$$MBE = \frac{1}{T}\sum_{t=1}^{T} Y_t - \widehat{Y_t}$$

Anexo II: Hojas de características

Este anexo muestra las hojas de características de los equipos empleados en la predicción de potencia fotovoltaica: los paneles solares fotovoltaicos y el regulador de carga MPPT. Para evitar aportar información excesiva, no se adjuntan las hojas de características de los equipos del nodo real de pruebas que no son empleados directamente en la predicción.



TP672P - 320 / 325 / 330W

Polycrystalline Solar Module
72-Cell Series

KEY FEATURES

Maximize limited space
Maximum power output 330W

Excellent Anti-PID performance
2 times of industry standard Anti-PID test by TUV Rheinland

Highly reliable due to stringent quality control
In-house testing goes well beyond certification requirements

Certified to withstand the most challenging environmental conditions
2400 Pa wind load · 5400 Pa snow load · 25 mm hail stones at 82 km/h

IP68 junction box
The highest waterproof level

Lower temperature coefficients
Enhance power generation

SYSTEM & PRODUCT CERTIFICATES

- IEC 61215 / IEC 61730 / UL 1703
- ISO 9001 : 2008 Quality Management System
- ISO 14001 : 2004 Environment Mangement System
- OHSAS 18001 : 2007 Occupational Health and Safety Management System

QUALITY WARRANTY

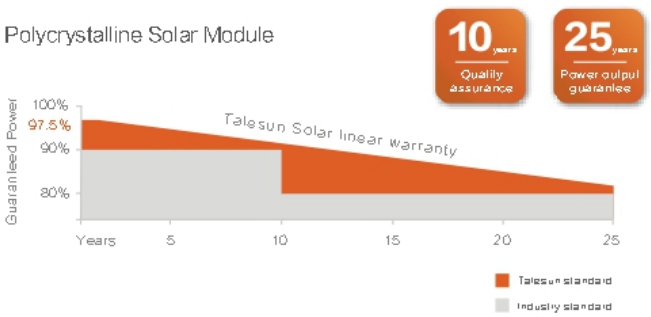
TALESUN guarantees that defects will not appear in materials and workmanship defined by IEC61215, IEC61730 or UL1703 under normal installation, use and maintenance as specified in Talesun's installation manual for 10 years from the warranty starting date.

ABOUT TALESUN SOLAR

TALESUN Solar is one of the world's largest integrated clean energy providers with 4 GW cell and 5 GW module production capacity globally. Its standard and high-efficiency product offerings are among the most powerful and cost-effective in the industry. Talesun Solar is committed to provide customers with customized; systematized and trustworthy turnkey solutions. Till now, Talesun Solar has accumulatively shipped more than 10 GW modules globally.

PERFORMANCE WARRANTY

Polycrystalline Solar Module



TALESUN

Web www.talesun.com
Tel +86 400 885 1098
Add No.1 Talesun Road, Shajiabang, Changshu, P.R.China

ELECTRICAL PARAMETERS

Performance at STC (Power Tolerance 0 – +3%)

Maximum Power (Pmax/W)	320	325	330
Operating Voltage (Vmpp/V)	37.1	37.4	37.7
Operating Current (Impp/A)	8.63	8.70	8.76
Open-Circuit Voltage (Voc/V)	45.5	45.7	45.9
Short-Circuit Current (Isc/A)	9.16	9.22	9.27
Module Efficiency ηm (%)	16.5	16.7	17.0

Performance at NOCT

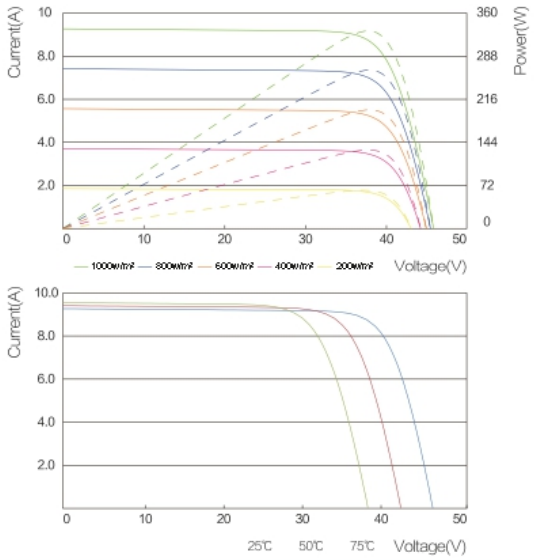
Maximum Power (Pmax/W)	236	240	243
Operating Voltage (Vmpp/V)	34.1	34.4	34.6
Operating Current (Impp/A)	6.92	6.98	7.04
Open-Circuit Voltage (Voc/V)	42.0	42.2	42.3
Short-Circuit Current (Isc/A)	7.42	7.46	7.51

STC: Irradiance 1000W/m², Cell Temperature 25° C, Air Mass AM1.5 NOCT: Irradiance at 800W/m², Ambient Temperature 20° C, Wind Speed 1m/s

MECHANICAL SPECIFICATION

Cell Type	Poly
Cell Dimensions	156.75*156.75mm(6inch)
Cell Arrangement	72(6*12)
Weight	22kg(48.5lbs)
Module Dimensions	1960*992*40mm(77.17*39.06*1.38inch)
Cable Length	1200mm(47.24inch)
Cable Cross Section Size	4mm²(0.006sq.in)
Front Glass	3.2mm High Transmission, Tempered Glass
No.of Bypass Diodes	3/6
Packing Configuration (1)	30pcs/Pallet,720pcs/40hq
Packing Configuration (2)	30pcs+5pcs/Pallet, 780pcs/40hq
Frame	Anodized Aluminium Alloy
Junction Box	IP68

I–V CURVE



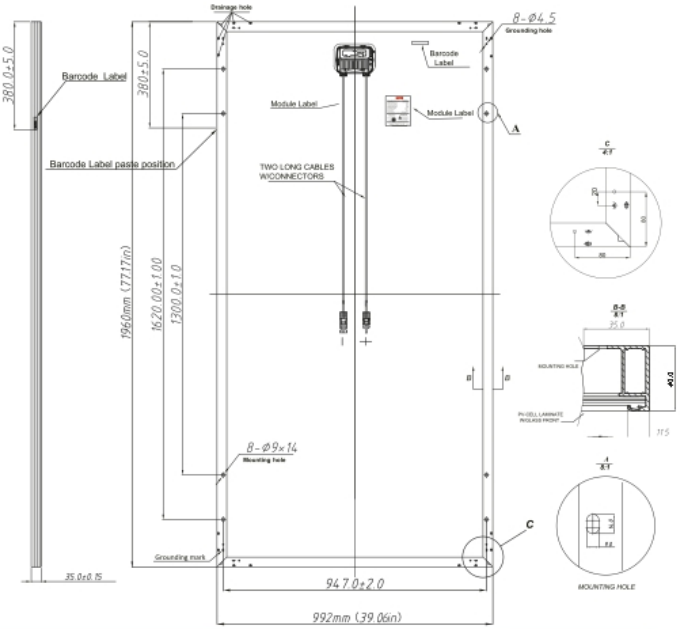
OPERATING CONDITIONS

Maximum System Voltage	1000V/DC(IEC)/1500V/DC(IEC)
Operating Temp	-40℃~+85℃
Maximum Series Fuse	15A
Static Loading	5400Pa
Conductivity at Ground	≤ 0.1Ω
Safety Class	II
Resistance	≥100MΩ
Connector	MC4 Compatible

TEMPERATURE COEFFICIENT

Temperature Coefficient Pmax	-0.40%/℃
Temperature Coefficient Voc	-0.31%/℃
Temperature Coefficient Isc	+0.06%/℃
NOCT	45±2℃

TECHNICAL DRAWINGS



The specification and key features described in this datasheet may deviate slightly and are not guaranteed. Due to ongoing innovation, R&D enhancement, Suzhou Talesun Solar Technologies Co., Ltd. reserves the right to make any adjustment to the information described herein at any time without notice. Please always obtain the most recent version of the datasheet which shall be duly incorporated into the binding contract made by the parties governing all transactions related to the purchase and sale of the products described herein. 201802EN

TALESUN

REGULADOR DE CARGA MPPT SR-ML 12/24/36/48V

60A

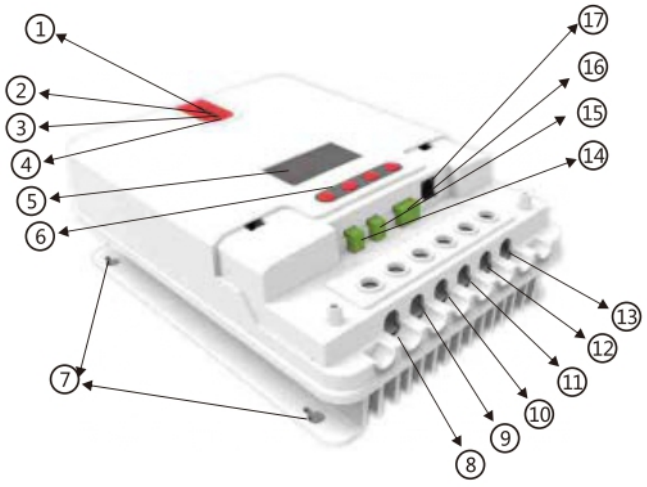
Características

- Gracias a la tecnología de seguimiento avanzado "multi-peak", el regulador es capaz de realizar un seguimiento MPPT aunque el panel solar tenga alguna sombra parcial que provoca picos rápidos de intensidad y voltaje.
- Un algoritmo integrado de seguimiento de máxima potencia que permite incrementar significativamente la eficiencia de producción de los sistemas fotovoltaicos. Este proceso aumenta la producción entre un 15 y un 20% en comparación a un sistema con regulación PWM convencional.
- Una combinación de varios algoritmos permiten un preciso seguimiento del punto óptimo de trabajo en la curva de intensidad y voltaje en un tiempo extremadamente corto.
- La eficiencia del seguidor MPPT alcanza hasta el 99.9%.
- Dispone de distintas opciones de carga en función del tipo de batería, siendo compatible con distintas tecnologías, incluyendo baterías de gel, selladas, abiertas, litio, etc.
- El controlador limita por corriente de carga. Cuando la potencia fotovoltaica excede la capacidad de carga del regulador, automáticamente se reduce la potencia de carga al límite que establece el regulador.
- Soporta reconocimiento automático de voltaje de batería.
- Indicadores LED de fallo y pantalla LCD que informa del funcionamiento anómalo para que el usuario pueda identificar rápidamente los fallos del sistema.
- Función de almacenamiento de datos histórico. Período máximo de 1 año.
- Se incluye una pantalla LCD que además de informar sobre el estado y funcionamiento del sistema, también permite modificar los parámetros del regulador.
- El controlador soporta el protocolo estándar Modbus, cubriendo las necesidades de comunicación.
- El controlador incorpora un sensor de exceso de temperatura. Cuando la temperatura sobrepasa el valor establecido, la corriente de carga del regulador se reducirá proporcionalmente hasta que la temperatura vuelva a parámetros aceptables para proteger el dispositivo para prevenir el sobrecalentamiento.
- Incorpora la función de compensación por temperatura, para que el regulador pueda ajustar la carga y la descarga de la batería para poder alargar al máximo su vida útil.
- Incorpora protección frente a rayos mediante diodos TVS.





Apariencia



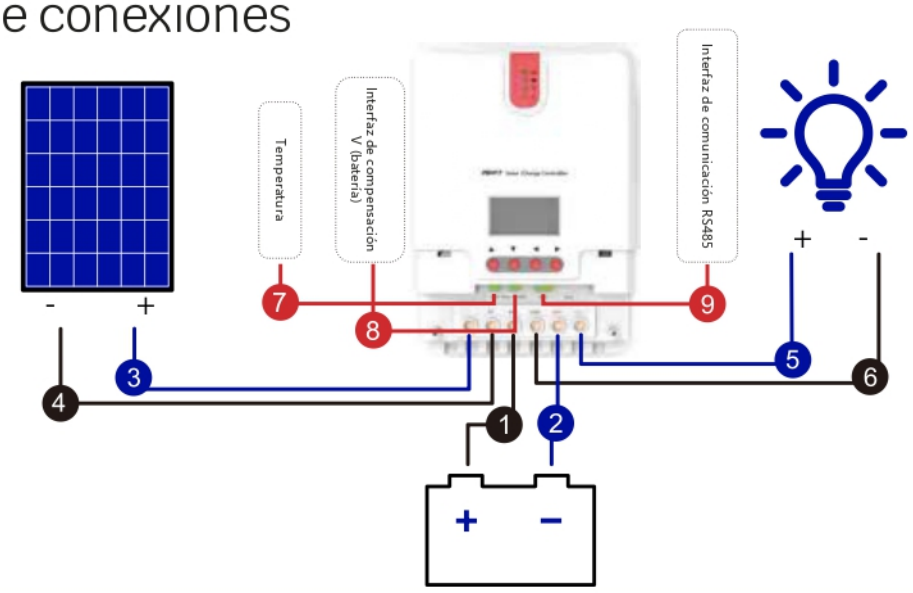
Nº	Item	Nº	Item
1	Indicador de carga en curso	10	Terminal negativo de batería
2	Indicador de batería	11	Terminal negativo de carga
3	Indicador de carga	12	Terminal positivo de batería
4	Indicador de funcionamiento anómalo	13	Terminal positivo de carga
5	Pantalla LCD	14	Interfaz de muestreo de temperatura externa
6	Botones de operación	15	Interfaz de compensación de voltaje de batería
7	Agujero de instalación	16	Interfaz de comunicación RS485
8	Terminal positivo de paneles	17	Interfaz de comunicación RS232
9	Terminal negativo de paneles		

REGULADOR DE CARGA MPPT SR-ML 12/24/36/48V 60A

Especificaciones técnicas

Parámetro	Valor
Modelo	4860
Voltaje del sistema	12V/24V/36V/48V Auto
Consumo sin carga	0.7W a 1.2W
Voltaje de batería	9V a 70V
Máx. voltaje de entrada	150V (25ºC) 145V (-25ºC)
Rango de voltaje MPPT máx.	Voltaje de batería +2V a 120V
Corriente de carga máx.	60A
Corriente de carga máx. CC	20A
Capacidad de carga máx.	10000uF
Entrada potencia fotovoltaica máx.	800W/12V; 1600W/24V; 2400W/36V; 3200W/48V
Eficiencia de conversión	≤98%
Eficiencia del MPPT	>99%
Compensación por temperatura	-3mv/ºC/2V (por defecto)
Rango de temperatura ambiente	-35ºC a +45ºC
Grado de protección	IP32
Peso	3.6Kg
Método de comunicación	RS232 | RS485
Altitud	≤3000m
Dimensiones	285*205*93mm

Diagrama de conexiones



Anexo III: Radiación y nubosidad

La transferencia de calor se produce en forma de conducción, convección y radiación. De estas, la radiación se debe únicamente a la temperatura del cuerpo emisor, siendo la única que no necesita contacto físico para producirse. La radiación emitida por un cuerpo caliente, en este caso de estudio el Sol, es distribuida en un rango de longitudes de onda, desde el infrarrojo hasta el ultravioleta, transformándose en un flujo de fotones con diferentes contenidos energéticos que llega hasta la Tierra. La radiación se pierde al interactuar con materia, siendo la transferencia energética final inversamente proporcional a la distancia entre cuerpos al cuadrado.

Mediante la ecuación de Stefan-Boltzmann se puede conocer el flujo de radiación medio fuera de la atmósfera terrestre, tomando el nombre de constante solar, para una temperatura de cuerpo negro del sol de $T = 5777\,K$, un radio solar de $r = 6.957 \cdot 10^8\,m$, una distancia Sol-Tierra de $l = 1.495 \cdot 10^{11}\,m$ y la constante de Boltzmann $\sigma = 5.670 \cdot 10^{-8}\,Wm^{-2}K^{-4}$. Sabiendo que, entre solsticios, la constante solar varía un $\pm 3.3\%$ de media, se define finalmente la radiación extraterrestre percibida [97].

$$G_l = \frac{\sigma T^4 \cdot 4\pi r^2}{4\pi l^2} = \sigma T^4 \left(\frac{r}{l}\right)^2 = 1367.6\,Wm^{-2}$$

Una vez la radiación alcanza la atmósfera terrestre, se atenúa de dos formas: dispersión y absorción. La dispersión ocurre al interactuar con las moléculas de aire, agua y polvo en la atmósfera y es especialmente efectiva en cortas longitudes de onda, explicando el color azulado del cielo, amarillento del sol o el enrojecimiento del anochecer. Gran parte de la luz dispersada es redirigida al espacio. Por otra parte, la absorción ocurre en el rango ultravioleta en las moléculas de ozono y en el rango infrarrojo en las moléculas de agua y el dióxido de carbono. En su proceso, la radiación se convierte en calor, emitiéndose como radiación de baja frecuencia por las partículas.

La atenuación de la radiación solar disminuye conforme se reduce el recorrido de los fotones en la atmósfera, justificando que el sol emita una radiación mayor conforme más se eleva en el cielo. El resultado final, a nivel de suelo, se conoce como irradiancia, en W/m^2, o irradiación, en Wh/m^2.

Por el efecto de dispersión, la radiación en la superficie terrestre se divide en dos componentes: radiación directa, puramente solar, y radiación difusa, dispersada por la atmósfera y enviada al suelo. Su proporción con respecto a la global depende de las condiciones atmosféricas: en días soleados entre un 10% y 20% de media es radiación difusa, mientras que en días encapotados prácticamente un 100% [111]. Existe un tercer tipo de radiación, denominado albedo, producido al reflejarse la radiación en cuerpos cercanos y el suelo, sin embargo, habitualmente supone una proporción muy pequeña, ya que depende de la reflectividad de los cuerpos,

que es baja para muchos materiales y terrenos, y es nula en superficies horizontales.

Para un día despejado, la irradiancia varía prácticamente en función de los movimientos interplanetarios y la latitud del lugar. El nivel de ozono, el polvo en el aire y otros factores atmosféricos afectan en menor medida si se asume una baja contaminación del aire. Por otra parte, en presencia de nubes, la irradiancia varía bruscamente, siendo de utilidad cuantificar este efecto mediante el índice de cielo despejado (*Clear Sky Index*, CSI), κ, que calcula la relación entre la GHI medida y su valor en cielo despejado (GHI_C).

$$\kappa = \frac{GHI}{GHI_C}$$

Cuando una nube oculta el sol la irradiancia se reduce drásticamente, sin embargo, si existen nubes distribuidas sin ocultar el sol, el CSI puede ser superior a uno, al reflejar las nubes colindantes la luz del sol. En la Figura 48 se representa un histograma de valores de CSI obtenidos a lo largo de 401 días en la ciudad de Oahu, en Hawaii, entre 2010 y 2011 [111]. Se observa que el máximo se da en días despejados, existiendo un segundo máximo relativo alrededor de $CSI = 0.3$. Por otra parte, se observan los valores superiores a uno que caracterizan el efecto mencionado.

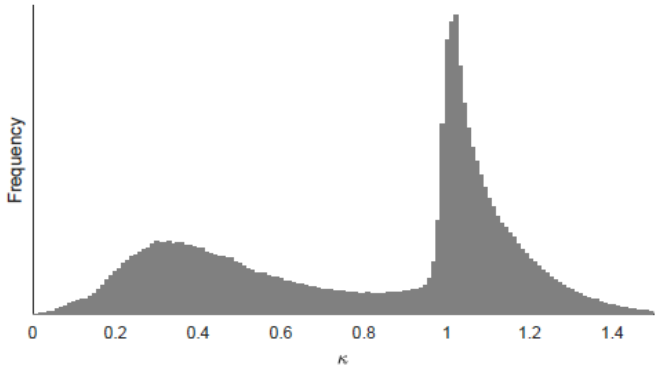


Figura 48.- Histograma del CSI (κ) con respecto a la GHI determinada para 401 días, medidas en la ciudad de Oahu, en Hawaii, en 2010 y 2011.

Anexo IV: RNN-LSTM

Las ANNs reconocen patrones de datos, aprendiendo de forma supervisada a partir de un entrenamiento que les permite predecir según clasificación y/o regresión. En la tesis, la red neuronal tradicional ANN-MLP se utiliza para la predicción de horizonte temporal diario y, por otra parte, la predicción de horizonte temporal intradiario emplea RNNs, al utilizar series temporales y disponer de un bucle de retroalimentación para usar la salida de una celda como la entrada de la siguiente.

Esto permite el proceso, el aprendizaje y la predicción de datos secuenciales, característica que no poseen las ANNs tradicionales, al disponer de entradas y salidas de tamaño fijo independientes entre sí. De

esta forma, las redes tradicionales no se ven afectadas por la salida anterior producida, mientras que las RNNs generan dependencia en este aspecto y los datos se relacionan con la información previa.

Al modificar el ciclo de retroalimentación, las RNNs procesan cada elemento secuencialmente por turnos y la información se conserva en el modelo. Esto resulta especialmente útil en tareas tales como la predicción del precio en el mercado eléctrico, el procesamiento de la señal, adivinar la siguiente palabra en una frase, escritura a mano, reconocimiento de voz o generar una pieza musical.

La primera RNN fue diseñada en la década de los 80 [112], tomando el nombre de redes Hopfield. En la década de los 90 estas redes experimentaron una gran mejoría con la llegada de las RNN-LSTMs [65], que resolvían el principal inconveniente de las RNNs tradicionales: la explosión y el desvanecimiento de gradientes [113]. Además, surgen las redes tipo GRU, que son más simples y veloces, en contrapartida a la precisión proporcionada.

Durante el entrenamiento, la RNN recibe las muestras y produce la estimación para el siguiente paso de tiempo, usando el algoritmo de retropropagación para su optimización. Esta propagación del error hacia atrás permite al modelo medir su gradiente en todas las conexiones. Después, el descenso de gradiente se emplea para ajustar los pesos y constantes con el fin de optimizar el proceso. Este proceso se repite con nuevas muestras hasta que el error se minimiza, debiendo asegurar la convergencia del proceso para generalizar el modelo.

A diferencia de las ANNs tradicionales, que usan pesos de forma separada para cada capa oculta, las RNNs comparten estos pesos en cada paso de tiempo. Al representar la red en función del tiempo, se considera que tiene una capa separada para cada paso temporal, compartiendo la misma matriz de pesos. Esto provoca que el gradiente del error en cada paso dependa del error producido en todos los pasos anteriores, conocido como propagación inversa a través del tiempo (*BackPropagation Through Time*, BPTT).

Durante la BPTT, al sumar las contribuciones de cada paso temporal al gradiente, especialmente en secuencias largas, se producen la explosión y el desvanecimiento de gradientes. El desvanecimiento de gradientes se atribuye a la presencia de pesos pequeños, produciendo que el siguiente gradiente se reduzca aún más hasta tender a cero. Al usar estos gradientes para actualizar los pesos de la red, la actualización de los parámetros se vuelve imperceptible y la red deja de aprender. En redes tradicionales esto no es un problema al disponer de pesos independientes entre capas. Por otra parte, si los pesos son grandes, se produce la explosión de gradientes, tendiendo en este caso al infinito. El problema de la explosión de gradientes puede ser resuelto eficientemente mediante un filtro. Por otra parte, el desvanecimiento de gradiente es más difícil de solucionar, sin embargo, con

una función de activación de unidad lineal rectificada (*Rectified Linear Unit*, ReLU) este problema resulta menos frecuente.

Para evitar este problema surgen las RNN-LSTMs, cuya estructura es similar modificando la función que calcula el estado oculto por una celda. Estas celdas reciben como entrada el estado anterior y el valor de datos actual y tienen la capacidad de mantener o eliminar información mediante el uso de tres puertas. Una de las puertas es la encargada de controlar qué información se olvida, conocida como puerta de olvido, la otra qué información de actualiza, conocida como puerta de entrada, y la otra crea un nuevo vector de valores candidatos que podrían agregarse al estado, conocida como puerta de salida. Durante la retropopagación, si la derivada de la salida comienza a converger a cero, los valores de la puerta de olvido pueden aumentar y evitar el desvanecimiento de gradientes, decisión tomada por la propia red durante el proceso de entrenamiento.

Además de la RNN-LSTM, las redes tipo GRU son una variante bastante popular desarrollada en 2014 [114], siendo una versión simplificada con un rendimiento comparable. Las redes tipo GRU tienen dos puertas, una puerta de reinicio y una puerta de actualización, similar a las usadas en la RNN-LSTM. La puerta de actualización es una combinación de las puertas de entrada y de olvido, y decide la cantidad del estado de celda anterior que persiste. La puerta de reinicio se aplica al estado anterior, y decide la combinación de los nuevos datos.

Las RNNs se adaptan al procesamiento de datos secuenciales, de forma que pueden ser combinadas con otras estrategias de gran potencial para aportar un modelo de excelentes características. Este es el caso de las CNN-LSTMs, que han logrado un rendimiento excelente en matrices con históricos de datos extensos [115]. Al extraer la salida de la CNN y utilizarla como entrada de la LSTM, se depura mejor la información, proporcionando variables de gran relevancia para el modelo.

Las RNNs y sus variantes son las técnicas recurrentes más usadas en series secuenciales [60]. En la presente tesis, en la predicción de irradiancia con horizonte temporal intradiario, se emplean RNN-LSTMs en lugar de redes tipo GRU porque pueden desarrollar patrones más complejos, evitando de una forma más efectiva el problema de desvanecimiento de gradientes. Por otra parte, las CNN-LSTMs pueden obtener precisiones superiores, en contrapartida a su carga computacional. Sin embargo, su principal potencial se centra en cantidades extensas de información, encontrando relaciones de una complejidad muy superior. Es de reseñar que únicamente dos variables de entrada son utilizadas en la predicción de irradiancia con horizonte temporal intradiario en la tesis. En [66] se comparan estas tres técnicas con otras más tradicionales, concluyendo que ofrecen precisiones muy superiores en predicción de GHI a corto plazo.

Bibliografía

[1] C. Hank, A. Sternberg, N. Köppel, M. Holst, T. Smolinka, A. Schaadt, C. Hebling y H. M. Henning, «Energy efficiency and economic assessment of imported energy carriers based on renewable electricity,» *Sustainable Energy & Fuels,* vol. 4, nº 5, pp. 2256-2273, 2020.

[2] D. Pudjianto, C. Ramsay y G. Strbac, «Virtual power plant and system integration of distributed energy resources,» *IET Renewable power generation,* vol. 1, nº 1, pp. 10-16, 2007.

[3] J. Abdella y K. Shuaib, «Peer to Peer Distributed Energy Trading in Smart Grid: A Survey,» *Energies,* vol. 11, nº 6, p. 1560, 2018.

[4] M. Gayo, C. Santos, F. Sánchez, P. Martin, J. Jiménez y M. Tradacete, «Addressing Challenges in Prosumer-Based Microgrids with Blockchain and an IEC 61850-Based Communication Scheme,» *IEEE Access,* vol. 8, pp. 201806-201822, 2020.

[5] R. Ritzer y N. Jurgenson, «Production, consumption, prosumption: The nature of capitalism in the age of the digital 'prosumer',» *Journal of consumer culture,* vol. 10, nº 1, pp. 13-36, 2010.

[6] «Agencia Estatal de Meteorología (AEMET) Modelo numérico de predicción meteorológica,» [En línea]. Available: http://www.aemet.es/es/eltiempo/prediccion/modelosnumericos/harmonie_arome_ccaa?opc2=mad. [Último acceso: 23 enero 2022].

[7] C. Europea, «Copernicus Europe's eyes on Earth,» [En línea]. Available: https://www.copernicus.eu/en. [Último acceso: 23 enero 2022].

[8] M. Tradacete, C. Santos, J. A. Jiménez, F. J. Rodríguez, P. Martín, E. Santiso y M. Gayo, «Turning Base Transceiver Stations into Scalable and Controllable DC Microgrids Based on a Smart Sensing Strategy,» *Sensors,* vol. 21, nº 4, p. 1202, 2021.

[9] P. Martín, G. Moreno, F. J. Rodríguez, J. A. Jiménez y I. Fernández, «A hybrid approach to short-term load forecasting aimed at bad data detection in secondary substation monitoring equipment,» *Sensors,* vol. 18, nº 11, p. 3947, 2018.

[10] G. Moreno, P. Martin, C. Santos, F. J. Rodríguez y E. Santiso, « A Day-Ahead Irradiance Forecasting Strategy for the Integration of Photovoltaic Systems in Virtual Power Plants,» *IEEE Access,* vol. 8, pp. 204226-204240, 2020.

[11] G. Moreno, C. Santos, P. Martín, F. J. Rodríguez, R. Peña y B. Vuksanovic, «Intra-Day Solar Power Forecasting Strategy for Managing Virtual Power Plants,» *Sensors,* vol. 21, nº 16, p. 5648, 2021.

[12] «Market Report Series: Renewables 2021. Analysis and forecasts for 2026. International Energy Agency (IEA),» 2021. [En línea]. Available: https://www.iea.org/reports/renewables-2021. [Último acceso: 23 enero 2022].

[13] M. Perez, R. Perez, K. R. Rábago y M. Putnam, «Overbuilding & curtailment: The cost-effective enablers of firm PV generation,» *Solar Energy,* vol. 180, pp. 412-422, 2019.

[14] B. Lin y J. Zhu, «The role of renewable energy technological innovation on climate change: Empirical evidence from China,» *Science of the Total Environment,* vol. 659, pp. 1505-1512, 2019.

[15] R. Shah, N. Mithulananthan, R. C. Bansal y V. K. Ramachandaramurthy, «A review of key power system stability challenges for large-scale PV integration.,» *Renewable and Sustainable Energy Reviews,* vol. 41, pp. 1423-1436, 2015.

[16] F. Ueckerdt, R. Brecha y G. Luderer, «Analyzing major challenges of wind and solar variability in power systems,» *Renewable energy,* vol. 81, pp. 1-10, 2015.

[17] X. Li, D. Hui y X. Lai, «Battery energy storage station (BESS)-based smoothing control of photovoltaic (PV) and wind power generation fluctuations,» *IEEE transactions on sustainable energy,* vol. 4, nº 2, pp. 464-473, 2013.

[18] M. Katsanevakis, R. A. Stewart y J. Lu, «Aggregated applications and benefits of energy storage systems with application-specific control methods: A review,» *Renewable and Sustainable Energy Reviews,* vol. 75, pp. 719-741, 2017.

[19] S. Grillo, M. Marinelli, S. Massucco y F. Silvestro, «Optimal management strategy of a battery-based storage system to improve renewable energy integration in distribution networks,» *IEEE Transactions on Smart Grid,* vol.

3, nº 2, pp. 950-958, 2012.

[20] S. Ghavidel, L. Li, J. Aghaei, T. Yu y J. Zhu, «A review on the virtual power plant: Components and operation systems,» de *IEEE International Conference on Power System Technology (POWERCON)*, Wollongong, NSW, Australia, 2016.

[21] H. Saboori, M. Mohammadi y R. Taghe, «Virtual power plant (VPP), definition, concept, components and types,» de *Asia-Pacific Power and Energy Engineering Conference*, Wuhan, China, 2011.

[22] R. Zhang y B. Hredzak, «Distributed Dynamic Clustering Algorithm for Formation of Heterogeneous Virtual Power Plants Based on Power Requirements,» *IEEE Transactions on Smart Grid,* vol. 12, nº 1, pp. 192-204, 2020.

[23] A. Khanjanzadeh, S. Soleymani y B. Mozafari, «A decentralized control strategy to bring back frequency and share reactive power in isolated microgrids with virtual power plant,» *Bulletin of the Polish Academy of Sciences. Technical Sciences,* vol. 69, nº 1, 2021.

[24] B. Marinescu, O. Gomis-Bellmunt, F. Dörfler, H. Schulte y L. Sigrist, «Dynamic virtual power plant: A new concept for grid integration of renewable energy sources,» *arXiv preprint,* p. arXiv:2108.00153, 2021.

[25] K. Dielmann y A. van der Velden, «Virtual power plants (VPP)-a new perspective for energy generation?,» de *Proceedings of the 9th International Scientific and Practical Conference of Students, Post-graduates Modern Techniques and Technologies*, Tomsk, Rusia, 2003.

[26] K. El Bakari y W. L. Kling, «Virtual power plants: An answer to increasing distributed generation,» de *2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, Gotemburgo, Suecia, 2010.

[27] C. Tarazona, M. Muscholl, R. Lopez y J. C. Passelergue, «Integration of distributed energy resources in the operation of energy management systems,» de *2009 IEEE PES/IAS Conference on Sustainable Alternative Energy (SAE)*, Valencia, España, 2009.

[28] P. Lombardi, M. Powalko y K. Rudion, «Optimal operation of a virtual power plant,» de *2009 IEEE Power & Energy Society General Meeting*, Calgary, Canadá, 2009.

[29] D. Hropko, J. Ivanecký y J. Turček, «Optimal dispatch of renewable energy sources included in virtual power plant using accelerated particle swarm optimization,» de *2012 ELEKTRO*, Rajecké Teplice, Eslovaquia, 2012.

[30] E. Mashhour y S. Moghaddas-Tafreshi, «Trading models for aggregating distributed energy resources into virtual power plant,» de *2nd International Conference on Adaptive Science & Technology (ICAST)*, Accra, Ghana, 2009.

[31] T. Sikorski, M. Jasiński, E. Ropuszyńska-Surma, M. Węglarz, D. Kaczorowska, P. Kostyla, Z. Leonowicz, R. Lis, J. Rezmer, W. Rojewski, M. Sobierajski, J. Szymanda, D. Bejmert, P. Janik y B. Solak, «A case study on distributed energy resources and energy-storage systems in a virtual power plant concept: Technical aspects,» *Energies,* vol. 13, nº 12, p. 3086, 2020.

[32] P. Lombardi, M. Stötzer, Z. Styczynski y A. Orths, «Multi-criteria optimization of an energy storage system within a Virtual Power Plant architecture,» de *2011 IEEE Power and Energy Society General Meeting*, Detroit, Míchigan, EE.UU., 2011.

[33] L. Nikonowicz y J. Milewski, «Virtual power plants-general review: structure, application and optimization,» *Journal of power technologies,* vol. 92, nº 3, p. 135, 2012.

[34] M. Salmani, S. Tafreshi y H. Salmani, «Operation optimization for a virtual power plant,» de *2009 IEEE PES/IAS Conference on Sustainable Alternative Energy (SAE)*, Valencia, España, 2009.

[35] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. Van Deventer, B. Horan y A. Stojcevski, «Forecasting of photovoltaic power generation and model optimization: A review,» *Renewable and Sustainable Energy Reviews,* vol. 81, pp. 912-928, 2018.

[36] M. Kudo, A. Takeuchi, Y. Nozaki, H. Endo y J. Sumita, «Forecasting electric power generation in a photovoltaic power system for an energy network,» *Electrical Engineering in Japan,* vol. 167, nº 4, pp. 16-23, 2009.

[37] M. Q. Raza y C. Ekanayake, «On recent advances in PV output power forecast,» *Solar Energy,* vol. 136, pp. 125-144, 2016.

[38] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin y Z. Hu, «Photovoltaic and solar power forecasting for smart grid energy management,» *CSEE Journal of Power and*

Energy Systems, vol. 1, nº 4, pp. 38-46, 2015.

[39] L. Massidda y M. Marrocu, «Smart meter forecasting from one minute to one year horizons,» *Energies,* vol. 11, nº 12, p. 3520, 2018.

[40] M. Diagne, M. David, P. Lauret, J. Boland y N. Schmutz, «Review of solar irradiance forecasting methods and a proposition for small-scale insular grids,» *Renewable and Sustainable Energy Reviews,* vol. 27, pp. 65-76, 2013.

[41] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison y F. Antonanzas-Torres, «Review of photovoltaic power forecasting,» *Solar energy,* vol. 136, pp. 78-111, 2016.

[42] «Datos meteorológicos horarios en la Comunidad de Madrid,» [En línea]. Available: https://datos.madrid.es/sites/v/index.jsp?vgnextoid=fa8357cec5efa610VgnVCM1000001d4a900aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD. [Último acceso: 23 enero 2022].

[43] M. Ding, L. Wang y R. Bi, «An ANN-based approach for forecasting the power output of photovoltaic system,» *Procedia Environmental Sciences,* vol. 11, pp. 1308-1315, 2011.

[44] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour y D. Wood, «Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators,» *IEEE Transactions on Smart Grid,* vol. 6, nº 5, pp. 2253-2262, 2015.

[45] J. G. D. S. Fonseca Junior, T. Oozeki, H. Ohtake, T. Takashima y O. Kazuhiko, «On the use of maximum likelihood and input data similarity to obtain prediction intervals for forecasts of photovoltaic power generation,» *Journal of Electrical Engineering and Technology,* vol. 10, nº 3, pp. 1342-1348, 2015.

[46] E. Akarslan y F. O. Hocaoglu, «A novel method based on similarity for hourly solar irradiance forecasting,» *Renewable Energy,* vol. 112, pp. 337-346, 2017.

[47] C. Du y Z. Liu, «Solar irradiance interval prediction based on set pair analysis theory,» *Energy Science and Applied Technology,* 2016.

[48] P. Mandal, T. Senjyu, N. Urasaki, T. Funabashi y A. K. Srivastava, «A novel approach to forecast electricity price for PJM using neural network and similar days method,» *IEEE Transactions on Power Systems,* vol. 22, nº 4, pp.

2058-2065, 2007.

[49] X. Zhang, B. Jiang, X. Zhang, F. Fang, Z. Gao y T. Feng, «Solar photovoltaic power prediction based on similar day approach,» de *2017 36th Chinese Control Conference (CCC)*, Dalian, China, 2017.

[50] C. Monteiro, T. Santos, L. A. Fernandez-Jimenez, I. J. Ramirez-Rosado y M. S. Terreros-Olarte, «Short-term power forecasting model for photovoltaic plants based on historical similarity,» *Energies,* vol. 6, nº 5, pp. 2624-2643, 2013.

[51] P. Bacher, H. Madsen y H. A. Nielsen, «Online short-term solar power forecasting,» *Solar energy,* vol. 83, nº 10, pp. 1772-1783, 2009.

[52] C. R. Osterwald, «Translation of device performance measurements to reference conditions,» *Solar cells,* vol. 18, nº 3-4, pp. 269-279, 1986.

[53] Z. Li, K. Wang, C. Li, M. Zhao y J. Cao, «Multimodal Deep Learning for Solar Irradiance Prediction,» de *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, Atalanta, Georgia, EE.UU., 2019.

[54] S. Makridakis, E. Spiliotis y V. Assimakopoulos, «Statistical and Machine Learning forecasting methods: Concerns and ways forward,» *PloS one,* vol. 13, nº 3, p. e0194889, 2018.

[55] F. Wang, Z. Zhang, J. Li, Z. Zhen y K. Li, «Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting,» *applied sciences,* vol. 8, nº 8, p. 1286, 2018.

[56] G. Li, H. Wang, S. Wang, J. Xin y H. Liu, «Recurrent neural networks based photovoltaic power forecasting approach,» *Energies,* vol. 12, nº 13, p. 2538, 2019.

[57] V. Cerqueira, L. Torgo y C. Soares, «Machine learning vs statistical methods for time series forecasting: Size matters,» *arXiv preprint,* p. arXiv:1909.13316, 2019.

[58] C. Voyant, G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte y A. Fouilloy, «Machine learning methods for solar radiation forecasting: A

review.,» *Renewable Energy,* vol. 105, pp. 569-582, 2017.

[59] V. Kostylev y A. Pavlovski, «Solar power forecasting performance–towards industry standards,» de *1st international workshop on the integration of solar power into power systems*, Aarhus, Dinamarca, 2011.

[60] R. A. Rajagukguk, R. A. Ramadhan y H. J. Lee, «A Review on Deep Learning Models for Forecasting Time Series Data of Solar Irradiance and Photovoltaic Power,» *Energies,* vol. 13, nº 24, p. 6623, 2020.

[61] E. Lorenz, J. Hurka, G. Karampela, D. Heinemann, H. Beyer y M. Schneider, «Qualified forecast of ensemble power production by spatially dispersed grid-connected PV systems,» de *23rd European Photovoltaic Solar Energy Conference*, Valencia, España, 2008.

[62] M. Nakamura, K. Takeno, R. Hisamitsu y M. Shoyama, «Bi-directional Multiport Converter for Utilizing Green Base Stations as Virtual Power Plant,» de *8th International Conference on Renewable Energy Research and Applications (ICRERA)*, Braşov, Rumanía, 2019.

[63] A. Mellit y S. A. Kalogirou, «Artificial intelligence techniques for photovoltaic applications: A review,» *Progress in energy and combustion science,* vol. 34, nº 5, pp. 574-632, 2018.

[64] Y. Y. Hong, J. J. F. Martinez y A. C. Fajardo, «Day-ahead solar irradiation forecasting utilizing gramian angular field and convolutional long short-term memory,» *IEEE Access,* vol. 8, pp. 18741-18753, 2020.

[65] S. Hochreiter y J. Schmidhuber, «Long short-term memory,» *Neural computation,* vol. 9, nº 8, pp. 1735-1780, 1997.

[66] H. Zang, L. Liu, L. Sun, L. Cheng, Z. Wei y G. Sun, «Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations,» *Renewable Energy,* vol. 160, pp. 26-41, 2020.

[67] M. Abdel-Nasser y K. Mahmoud, «Accurate photovoltaic power forecasting models using deep LSTM-RNN,» *Neural Computing and Applications,* vol. 31, nº 7, pp. 2727-2740, 2019.

[68] S. Srivastava y S. Lessmann, «A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data,»

Solar Energy, vol. 162, pp. 232-247, 2018.

[69] X. Qing y Y. Niu, «Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM,» *Energy,* vol. 148, pp. 461-468, 2018.

[70] D. Chandola, H. Gupta, V. A. Tikkiwal y M. K. Bohra, «Multi-step ahead forecasting of global solar radiation for arid zones using deep learning,» *Procedia Computer Science,* vol. 167, pp. 626-635, 2020.

[71] P. Kumari y D. Toshniwal, «Hourly solar irradiance prediction from satellite data using lstm,» 2019.

[72] M. S. Hossain y H. Mahmood, «Short-term photovoltaic power forecasting using an LSTM neural network,» de *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Washington D. C., EE.UU., 2020.

[73] M. Husein y I. Y. Chung, «Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach,» *Energies,* vol. 12, nº 10, p. 1856, 2019.

[74] K. Wang, X. Qi y H. Liu, «Photovoltaic power forecasting based LSTM-Convolutional Network,» *Energy,* vol. 189, p. 116225, 2019.

[75] S. Boubaker, M. Benghanem, A. Mellit, A. Lefza, O. Kahouli y L. Kolsi, «Deep Neural Networks for Predicting Solar Radiation at Hail Region, Saudi Arabia,» *IEEE Access,* vol. 9, pp. 36719-36729, 2021.

[76] X. Chen, X. Huang, Y. Cai, H. Shen y J. Lu, «Intra-day Forecast of Ground Horizontal Irradiance Using Long Short-term Memory Network (LSTM),» *Journal of the Meteorological Society of Japan. Ser. II,* 2020.

[77] B. Gao, X. Huang, J. Shi, Y. Tai y J. Zhang, «Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks,» *Renewable Energy,* vol. 162, pp. 1665-1683, 2020.

[78] N. Vakitbilir, A. Hilal y C. Direkoğlu, «Prediction of Daily Solar Irradiation Using CNN and LSTM Networks,» de *International Conference on Theory and Applications of Fuzzy Systems and Soft Computing*, 2020.

[79] D. F. Menicucci, «Photovoltaic array performance simulation models,» de *PV and Isolation Measurements Workshop*, Vail, Colorado, EE.UU., 1985.

[80] B. Marion, «Comparison of predictive models for photovoltaic module performance,» de *2008 33rd IEEE Photovoltaic Specialists Conference*, San Diego, California, EE.UU., 2008.

[81] R. Peña, A. M. Diez-Pascual, P. G. Díaz y L. V. Davoise, «A new method for current–voltage curve prediction in photovoltaic modules,» *IET Renewable Power Generation,* vol. 15, nº 6, pp. 1331-1343, 2021.

[82] B. Marion, S. Rummel y A. Anderberg, «Current–voltage curve translation by bilinear interpolation,» *Progress in Photovoltaics: Research and Applications,* vol. 12, nº 8, pp. 593-607, 2004.

[83] W. Heydenreich, B. Müller y C. Reise, «Describing the world with three parameters: a new approach to PV module power modelling,» de *23rd European PV Solar Energy Conference and Exhibition (EU PVSEC)*, Valencia, España, 2008.

[84] A. Takilalte, S. Harrouni, M. R. Yaiche y L. Mora-López, «New approach to estimate 5-min global solar irradiation data on tilted planes from horizontal measurement.,» *Renewable energy,* vol. 145, pp. 2477-2488, 2020.

[85] «ThingSpeak para proyectos IoT,» [En línea]. Available: https://thingspeak.com/. [Último acceso: 23 enero 2022].

[86] S. E. Berrizbeitia, E. Jadraque Gago y T. Muneer, «Empirical Models for the Estimation of Solar Sky-Diffuse Radiation. A Review and Experimental Analysis,» *Energies,* vol. 13, nº 3, p. 701, 2020.

[87] J. E. Hay y J. A. Davies, «Calculation of the Solar Radiation Incident on an Inclined Surface,» de *First Canadian Solar Radiation Data Workshop*, Toronto, Canadá, Ministry of Supply and Services, 1980, p. 59.

[88] N. Martin y J. M. Ruiz, «Calculation of the PV modules angular losses under field conditions by means of an analytical model,» *Solar Energy Materials and Solar Cells,* vol. 70, nº 1, pp. 25-38, 2001.

[89] J. A. Kratochvil, W. E. Boyson y D. L. King, *Photovoltaic array performance model,* Albuquerque y Livermore, EE.UU.: Sandia National Laboratories

(SNL), 2004.

[90] R. V. Rohli y C. Li, «The Seven Basic Equations in Weather Forecasting Models,» de *Meteorology for Coastal Scientists*, Springer, Cham, 2021.

[91] M. G. Schultz, C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari y S. Stadtler, «Can deep learning beat numerical weather prediction?,» *Philosophical Transactions of the Royal Society A,* vol. 379, nº 2194, p. 20200097, 2021.

[92] L. Bengtsson, V. Andrae, T. Aspelien, Y. Batrak, J. Calvo, W. de Rooy, E. Gleeson, B. Hansen, M. Homleid, M. Hortal, K. Ivarsson, G. Lenderink, S. Niemela, K. Nielsen, J. Onvlee, L. Rontu, P. Samuelsson, D. Santos, A. Subias, S. Tijm, V. Toll, X. Yang y M. Ø. Køltzow, «The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system,» *Monthly Weather Review,* vol. 145, nº 5, pp. 1919-1935, 2017.

[93] J. C. Sánchez, D. M. Pérez, G. M. Martín y S. V. Jimenez, «HARMONIE-AROME, modelo operativo de escala convectiva de AEMET. Parte I: Modelo de predicción y validación,» *Acta de las Jornadas Científicas de la Asociación Meteorológica Española,* vol. 1, nº 35, pp. 117-125, 2018.

[94] M. Martínez, A. Rodríguez y J. García-Moya, ««NowcRadiation» — Software de nowcasting para la predicción de la radiación solar,» de *Sexto simposio nacional de predicción - Memorial Antonio Mestre*, Madrid, España, 2018.

[95] J. Calvo, «HARMONIE-AROME: Modelo atmosférico de alta resolución de AEMET,» de *Seminario Cátedra de Ciencia de Datos y Aprendizaje Automático*, 2018.

[96] Y. Huang, S. Siems, M. Manton, A. Protat, L. Majewski y H. Nguyen, «Evaluating Himawari-8 cloud products using shipborne and CALIPSO observations: Cloud-top height and cloud-top temperature,» *Journal of Atmospheric and Oceanic Technology,* vol. 36, nº 12, pp. 2327-2347, 2019.

[97] J. A. Duffie y W. A. Beckman, Solar Engineering of Thermal Processes, University of Wisconsin-Madison: Wiley, 2013.

[98] A. Feijt y P. De Valk, «The use of NWP model surface temperatures in cloud detection from satellite,» *International Journal of Remote Sensing,* vol. 22, nº 13, pp. 2571-2584, 2001.

[99] «Agencia Estatal de Meteorología (AEMET) Predicción por municipios,» [En línea]. Available: http://www.aemet.es/es/eltiempo/prediccion/municipios/horas/tabla/alcala-de-henares-id28005. [Último acceso: 23 enero 2022].

[100] J. Malý, «The Darboux property for gradients,» *Real Analysis Exchange,* vol. 22, nº 1, p. 167–173, 1996.

[101] S. Haykin, «Neural networks and learning machines,» Hamilton, Ontario, Canadá, Pearson Prentice Hall, 2009, p. 175.

[102] J. Huertas-Tato, R. Aler, I. M. Galván, F. J. Rodríguez-Benítez, C. Arbizu-Barrena y D. Pozo-Vázquez, «A short-term solar radiation forecasting system for the Iberian Peninsula. Part 2: Model blending approaches based on machine learning,» *Solar Energy,* vol. 195, pp. 685-696, 2020.

[103] Y. Yu, J. Cao y J. Zhu, «An LSTM short-term solar irradiance forecasting under complicated weather conditions,» *IEEE Access,* vol. 7, pp. 145651-145666, 2019.

[104] J. Wojtkiewicz, M. Hosseini, R. Gottumukkala y T. L. Chambers, «Hour-ahead solar irradiance forecasting using multivariate gated recurrent units,» *Energies,* vol. 12, nº 21, p. 4055, 2019.

[105] K. Yan, H. Shen, L. Wang, H. Zhou, M. Xu y Y. Mo, «Short-term solar irradiance forecasting based on a hybrid deep learning methodology,» *Information,* vol. 11, nº 1, p. 32, 2020.

[106] S. Rodríguez, «Influencia de las variaciones espectrales en la corriente de cortocircuito de módulos fotovoltaicos. Trabajo de Fin de Grado de la Universidad de Alcalá en el departamento de Teoría de la Señal y Comunicaciones. Tutor: Peña, R.,» 2016. [En línea]. Available: https://github.com/sandialabs/MATLAB_PV_LIB/tree/master/Example%20Data. [Último acceso: 23 enero 2022].

[107] C. Schwingshackl, M. Petitta, J. E. Wagner, G. Belluardo, D. Moser, M. Castelli, M. Zebisch y A. Tetzlaff, «Wind effect on PV module temperature: Analysis of different techniques for an accurate estimation,» *Energy Procedia,* vol. 40, pp. 77-86, 2013.

[108] D. W. Van der Meer, J. Widén y J. Munkhammar, «Review on probabilistic forecasting of photovoltaic power production and electricity consumption,»

Renewable and Sustainable Energy Reviews, vol. 81, pp. 1484-1512, 2018.

[109] A. Teke, H. B. Yıldırım y Ö. Çelik, «Evaluation and performance comparison of different models for the estimation of solar radiation,» *Renewable and sustainable energy reviews,* vol. 50, pp. 1097-1107, 2015.

[110] R. J. Hyndman y A. B. Koehler, «Another look at measures of forecast accuracy,» *International journal of forecasting,* vol. 22, nº 4, pp. 679-688, 2006.

[111] J. Widén y J. Munkhammar, Solar Radiation Theory, Upsala, Suecia: Uppsala University, 2019.

[112] J. J. Hopfield, «Neural networks and physical systems with emergent collective computational abilities,» *Proceedings of the national academy of sciences,* vol. 79, nº 8, pp. 2554-2558, 1982.

[113] Y. Bengio, P. Simard y P. Frasconi, «Learning long-term dependencies with gradient descent is difficult,» *IEEE transactions on neural networks,* vol. 5, nº 2, pp. 157-166, 1994.

[114] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk y Y. Bengio, «Learning phrase representations using RNN encoder-decoder for statistical machine translation,» *arXiv preprint,* p. arXiv:1406.1078, 2014.

[115] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell y K. Saenko, «Sequence to sequence-video to text,» de *Proceedings of the IEEE international conference on computer vision*, Cambridge, Massachusetts, EE.UU., 2015.