

Trabajo Fin de Máster

Máster Universitario en Ingeniería Industrial

Predicción de la radiación global horizontal
mediante el uso de redes neuronales

Autor: Javier Rodríguez Lucena

Tutores: Isidoro Lillo Bravo

Elena Pérez Aparicio

Dpto. Ingeniería Energética
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2021



Trabajo Fin de Máster
Máster Universitario en Ingeniería Industrial

Predicción de la radiación global horizontal mediante el uso de redes neuronales

Autor:

Javier Rodríguez Lucena

Tutores:

Isidoro Lillo Bravo

Elena Pérez Aparicio

Dpto. de Ingeniería Energética
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2021

Proyecto Fin de Carrera: Predicción de la radiación global horizontal mediante el uso de redes neuronales

Autor: Javier Rodríguez Lucena

Tutor: Isidoro Lillo Bravo
Elena Pérez Aparicio

El tribunal nombrado para juzgar el Proyecto arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2021

El secretario del Tribunal

*A Dios, a mi familia, y a todos
los que me han acompañado en
este tramo del camino.*

Resumen

El aumento incesante de la demanda de electricidad mundial, tanto a nivel industrial como a nivel doméstico, unido a la necesidad de reducir las emisiones producidas en la producción de electricidad, ha llevado al ser humano a desarrollar nuevas metodologías asociadas a la obtención de energía eléctrica. Dentro de este grupo, se encuentran en auge en los últimos tiempos las llamadas energías limpias o energías renovables. Estas energías han sido objeto de múltiples estudios para mejorar la eficiencia tanto en su producción como en su distribución.

El presente documento estará encuadrado dentro del marco de la energía solar, y más concretamente se centrará en hacer una evaluación profunda los diferentes modelos existentes en la actualidad para la predicción de la radiación solar horizontal a corto plazo, haciendo hincapié en los métodos basados en redes neuronales artificiales. Con ello, se seleccionará una red neuronal en función de la disponibilidad de herramientas para su evaluación y se pondrá a prueba en un clima como el de Sevilla, evaluando diferentes espectros temporales en el aprendizaje de la red.

Del mismo modo, se analizarán los resultados obtenidos, y se hará una comparación de dichos resultados con los obtenidos en artículos científicos contrastados en climas similares al de la localidad de estudio, de tal manera que se pueda comprobar el desempeño de la red en las condiciones elegidas. Finalmente, se expondrá una breve conclusión de los resultados y se hará una reflexión de las posibles líneas de trabajo que puedan tomarse en el futuro.

Resumen	ix
Índice	xi
Índice de Tablas	xiii
Índice de Figuras	xv
1 Introducción	1
1.1. <i>Situación actual de la demanda eléctrica</i>	1
1.1.1. Situación actual de la demanda eléctrica mundial	1
1.1.2. Situación actual de la demanda eléctrica nacional	2
1.2. <i>Futuro de las energías renovables (2020 – 2050)</i>	4
1.3. <i>Revisión bibliográfica de los modelos de predicción</i>	5
2 Objeto del trabajo	13
3 Metodología	15
3.1. <i>Base de datos</i>	15
3.2. <i>Tratamiento de redes neuronales con MATLAB</i>	16
3.3. <i>Modelo utilizado</i>	17
3.3.1. Tratamiento de previo y flujo de datos	18
3.3.2. Unidad LSTM	18
3.3.3. Índices de precisión	19
3.4. <i>Implementación</i>	20
3.4.1. Parámetros de red	20
3.4.2. Algoritmo de entrenamiento	21
3.4.3. Clasificación de los datos de entrada	23
4 Resultados	25
4.1. <i>Análisis mensual</i>	25
4.2. <i>Análisis trimestral</i>	28
4.3. <i>Análisis anual</i>	29
4.4. <i>Análisis bianual y trianual</i>	30
4.5. <i>Análisis para días con alta variabilidad</i>	32
4.6. <i>Comparativa de resultados haciendo uso de la unidad GRU</i>	33
4.7. <i>Ruido en el problema multivariable</i>	34
5 Comparativa con artículos científicos	37
6 Conclusiones y líneas de trabajo futuras	39
7 Bibliografía	41

ÍNDICE DE TABLAS

Tabla 1. Tabla resumen de los modelos de predicción de la radiación solar.	12
Tabla 2. Tabla comparativa de iteraciones vs error cometido y tiempo de cálculo.	21
Tabla 3. Resumen de los argumentos del algoritmo neuronal.	23
Tabla 4. Resultados del RMSE mensuales por año.	27
Tabla 5. Resultados del RMSE trimestrales por año.	29
Tabla 6. Comparación del error para 120 y 200 iteraciones para el año 2015.	29
Tabla 7. Resultados del RMSE anual.	30
Tabla 8. Resultados del RMSE bianual y trianual.	30
Tabla 9. Tiempos de ejecución en función del período elegido.	32
Tabla 10. Comparativa de errores de LSTM frente a GRU.	33
Tabla 11. Error univariable vs error multivariable.	34
Tabla 12. RMSE según H. Zang <i>et al.</i> [12] vs resultados obtenidos.	37
Tabla 13. Comparación con los resultados con M. Sorkun <i>et al.</i>	38

ÍNDICE DE FIGURAS

Figura 1. Evolución de la producción eléctrica mundial.	1
Figura 2. Producción de electricidad mundial por fuente.	2
Figura 3. Producción de electricidad en España.	3
Figura 4. Producción de electricidad nacional por fuente.	3
Figura 5. Pronóstico de la producción renovable por escenario.	4
Figura 6. Precio de la energía renovable por escenario.	4
Figura 7. Red neuronal adoptada por Marzouq et al.	6
Figura 8. Arquitectura básica de una CNN.	7
Figura 9. Clasificación de los métodos de aprendizaje.	8
Figura 10. Ejemplo de una imagen tomada por una cámara de cielo.	9
Figura 11. GHI vs Tiempo.	15
Figura 12. Arquitectura de la red empleada.	17
Figura 13. Arquitectura básica de la unidad LSTM.	19
Figura 14. Proceso de entrenamiento para el mes de enero de 2019 y 120 iteraciones.	22
Figura 15. Proceso de entrenamiento para el mes de enero de 2019 y 200 iteraciones.	22
Figura 16. Proceso de entrenamiento para el mes de enero de 2019 y 250 iteraciones.	23
Figura 17. RMSE vs mes y año.	25
Figura 18. Predicción con volumen de datos de entrada mensual - enero 2018.	26
Figura 19. Predicción con volumen de datos de entrada mensual - abril 2018.	26
Figura 20. Predicción con volumen de datos de entrada mensual - julio 2018.	27
Figura 21. RMSE vs trimestre y año.	28
Figura 22. Predicción con volumen de datos de entrada trimestral– 2º trimestre 2019.	28
Figura 23. Predicción de la última semana de diciembre 2015 con volumen de datos mensual.	31
Figura 24. Predicción de la última semana de diciembre 2015 con volumen de datos anual.	31
Figura 25. Predicción con entrenamiento con días altamente variables.	32
Figura 26. Error RMSE de LSTM vs GRU.	33
Figura 27. Ejemplo del ruido introducido en la predicción.	35

1 INTRODUCCIÓN

En la actualidad, son muchos los estudios que se preocupan por la situación en la que se encuentran las diferentes alternativas para la producción de energía eléctrica. Estos “tipos de energía” incluyen desde la energía nuclear o la basada en el carbón, hasta la energía eólica o solar. Al mismo tiempo, todos estos métodos han experimentado innumerables intentos de mejora, tratando de conseguir el mayor rendimiento posible a un mínimo coste. La enorme preocupación por las emisiones derivadas de la producción de energía con recursos con alto contenido en carbono ha llevado a las energías renovables o “energías limpias” a un primer plano, con el objetivo principal de reducir el impacto provocado por dichas emisiones. La predicción de la radiación solar compone un pilar fundamental en las energías dependientes del Sol, ya que, una buena predicción, evitará desviaciones y hará más preciso el cálculo de la energía eléctrica producida en un período determinado de tiempo.

Antes de comenzar con el objeto del presente trabajo, es necesario tener una visión global de la situación eléctrica mundial y nacional, así como evaluar el papel que juegan las energías renovables tanto ahora como en el futuro próximo. Además, se debe realizar una revisión del estado del arte de los modelos de predicción que se usan actualmente, a fin de conocer la actualidad de los temas que se tratarán en adelante.

1.1. Situación actual de la demanda eléctrica

1.1.1. Situación actual de la demanda eléctrica mundial

La demanda de electricidad a nivel mundial no ha dejado de aumentar a lo largo de la historia. Esto se debe tanto al aumento de población como a la demanda creciente debido al desarrollo socioeconómico de cada país. Esta demanda de energía se cubre con diferentes fuentes de energía a través de diversas tecnologías. Necesariamente, la producción ha aumentado para poder cubrir esta demanda.

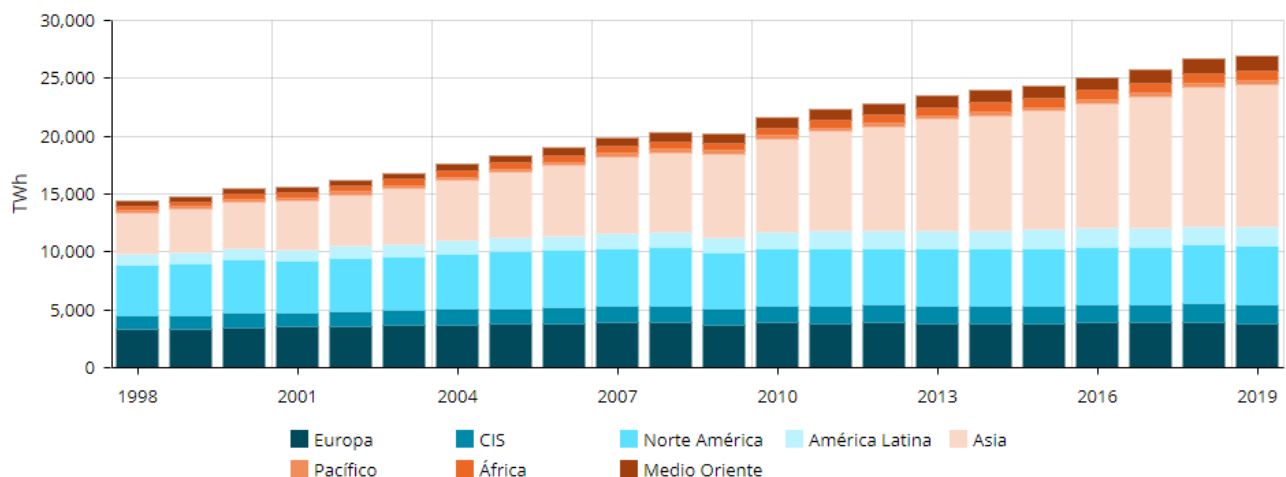


Figura 1. Evolución de la producción eléctrica mundial. (Enerdata [34])

En la figura 1 se expone como ha aumentado la producción de electricidad a nivel mundial en los últimos 20 años en torno a un 70%. Además, se hace notable el aumento en los países asiáticos, como puede ser China, que ha multiplicado por siete su producción (7482 TWh – 2019 frente a 1240 TWh – 1999) o la India triplicando su contribución (1614 TWh – 2019 frente a 546 TWh – 1999) [2]. Esto es debido al rápido desarrollo industrial que existe en estos países. También es notable la producción constante de Norte América, o Europa, países desarrollados con grandes núcleos cada vez más industrializados.

Según la IEA (*International Energy Agency*) [35], el consumo per cápita de electricidad se sitúa para el año 2018 en 3.3 MWh, mientras que en el año 1998 era de 2.2 MWh, reflejándose un aumento del consumo del 50%. Cabe preguntarse cómo se ha cubierto este consumo, o lo que es lo mismo, la procedencia de la energía producida a lo largo de los últimos años. En la figura 2 se puede observar como el carbón sigue siendo una fuente importante para la producción de electricidad a nivel mundial, y al mismo tiempo, como, a pesar de las contraindicaciones derivadas de sus emisiones, no merma su utilización de forma notable. Esto se debe a que la industria instalada está muy arraigada en las energías llamadas “*tradicionales*”, y a su uso en países en vías de desarrollo debido a que su precio sigue estando muy por debajo de las demás fuentes de energía. En segundo lugar, se encuentra el gas natural, con una tendencia de ascenso similar al carbón. Le sigue la energía hidráulica, con un ascenso notable desde el año 2000 en adelante, y la energía nuclear, estancada en una producción constante, debido a la peligrosidad de los residuos que genera, y a su coste de instalación y mantenimiento.

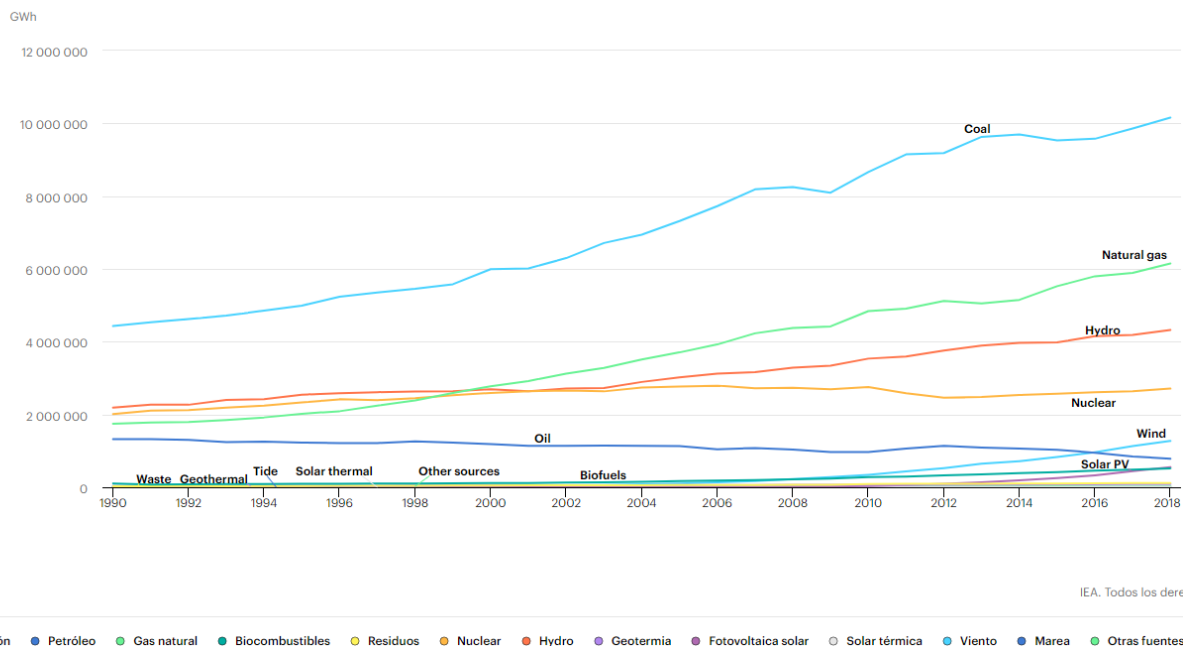


Figura 2. Producción de electricidad mundial por fuente. (IEA [35])

Por último, encontramos las llamadas energías renovables como la energía eólica y la energía fotovoltaica, con un repunte notable en los últimos años, debido a la gran inversión recibida con el objetivo de reducir las posibles emisiones causadas por las energías con alto contenido en carbono.

Si se evalúa la incidencia de las energías renovables para el año 2019, se constata el hecho de que el 27% de la producción eléctrica mundial se debe a fuentes renovables, por el 73% de las no renovables. El aumento de la capacidad eólica y solar, la implantación de políticas ambiciosas para ayudar a la energía renovable, y la reducción de los costes tecnológicos, ha hecho posible este aumento en el uso de dichas energías. En Europa las energías renovables ocupan el 35% de la producción energética, mientras que otros países como Estados Unidos, Rusia, o Japón, se sitúan por debajo de la media mundial, con un 21% de participación en su oferta energética.

1.1.2. Situación actual de la demanda eléctrica nacional

Si se evalúa el comportamiento de nuestro país en cuanto a demanda y producción eléctrica se refiere, se puede observar como la tendencia no es siempre al alza, como sucede en China o India.

En la figura 3 se puede observar una ralentización del crecimiento de la generación de energía eléctrica. El cambio en la tendencia se observa de forma clara en el año 2008, coincidente con la crisis económica en la que se sumió el país que ralentizó todo crecimiento industrial y económico del momento, y con ello, la demanda de

energía. Posteriormente, se observa cierta estabilización debido a la mejora de las condiciones económicas del país, y por ello, una reactivación del sector industrial.

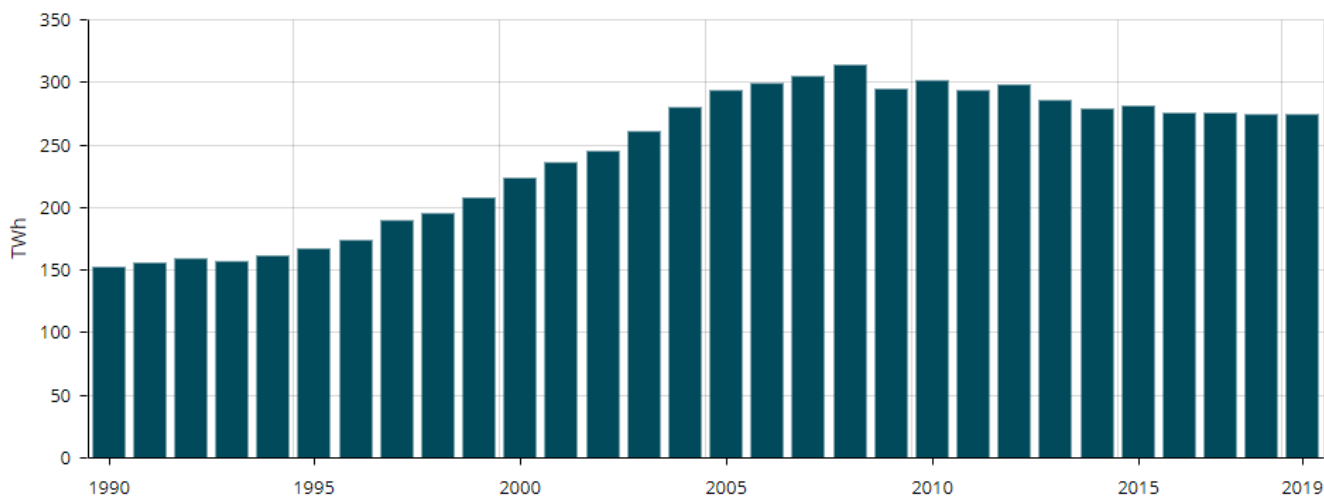
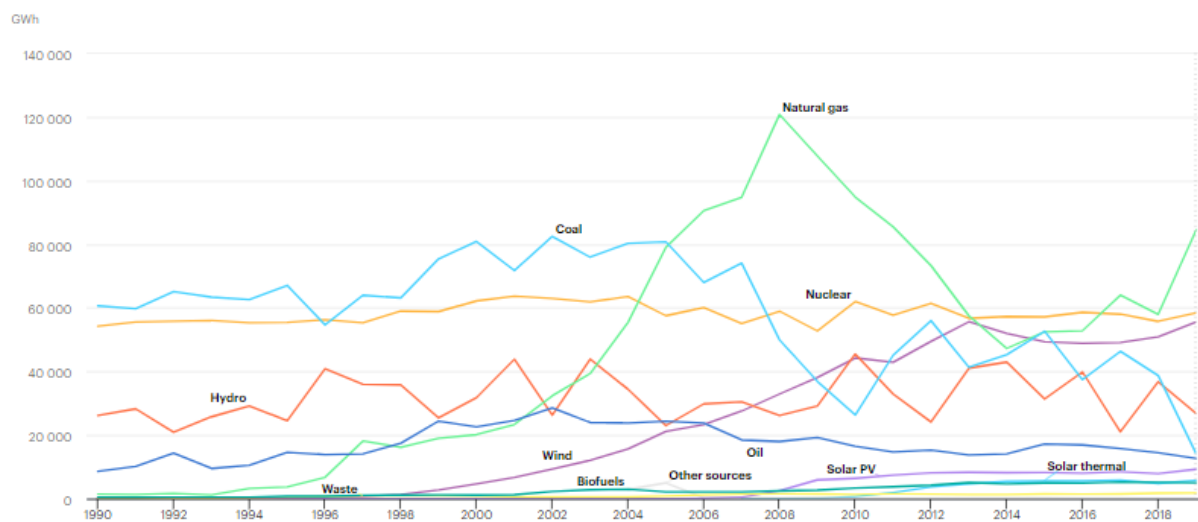


Figura 3. Producción de electricidad en España. (Enerdata [34])

Además, la tendencia de la producción en España en los últimos años coincide con el cierre de centrales de carbón por parte de las autoridades, y que se puede ver su efecto en la figura 4. Al mismo tiempo, como se ha comentado previamente para el panorama mundial, es notable el aumento de las energías renovables destacando sobre las demás la energía eólica, y en menor medida la energía solar fotovoltaica.



IEA. Reservados todos los derechos.

● Carbón ● Petróleo ● Gas natural ● Biocombustibles ● Desperdicio ● Nuclear ● Hydro ● Fotovoltaica solar ● Viento ● Otras fuentes ● Solar térmica

Figura 4. Producción de electricidad nacional por fuente. (IEA [35])

Las energías renovables supusieron en España un 38% de la producción de electricidad para el año 2019, frente las 62% atribuido a las energías no renovables. Se pone de manifiesto el comentado auge de las energías renovables gracias a las políticas llevadas a cabo por las autoridades. A pesar de que la tendencia es positiva, el cambio es lento respecto al cambio deseado.

1.2. Futuro de las energías renovables (2020 – 2050)

Debido a la gran importancia de las energías renovables en los últimos años, y, en consecuencia, al aumento de su implantación y explotación, existen en la actualidad diversidad de estudios dedicados a la previsión del comportamiento que tendrán este tipo de energías en los años venideros. Por ello, se ha considerado relevante evaluar algunos de ellos, con el fin de tener una idea de lo que puede suponer el futuro de las energías renovables en la oferta energética mundial.

En el análisis que realiza la compañía BP (*BP Energy Outlook* [31]), la tendencia de estas energías tiene tres posibles escenarios en función de las políticas que tome cada país con el objetivo de reducir las emisiones de gases de efecto invernadero y reducir así la huella de carbono. Estos escenarios son:

- **Rapid Transition Scenario (Rapid)**. Apuesta por unas políticas lideradas por el aumento del precio del carbón, y apoyo a sectores determinados que hagan que la energía derivada del carbono se reduzca de forma considerable.
- **Net Zero Scenario (Net Zero)**. Se asume que las decisiones tomadas en el escenario anterior se ven reforzadas por una mayor concienciación de la población y cambios considerables en su estilo de vida. Estas medidas reducirán aún más el impacto de las emisiones globales de carbono.
- **Business-as-usual (BAU)**. Se trata de continuar las políticas recientes, sin tomar mayores medidas, y con un avance lento en la reducción de las emisiones derivados del carbono.

En la figura 5 se puede ver la previsión de producción y precios que pueden tener los diferentes tipos de energías renovables desde la actualidad hasta el año 2050, en función de la clasificación comentada previamente.

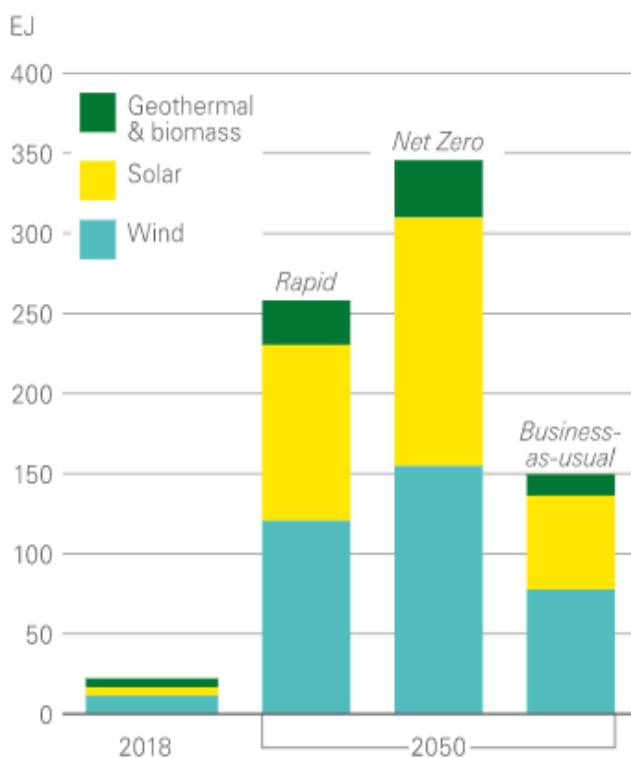


Figura 5. Pronóstico de la producción renovable por escenario. (BP Energy Outlook [31])

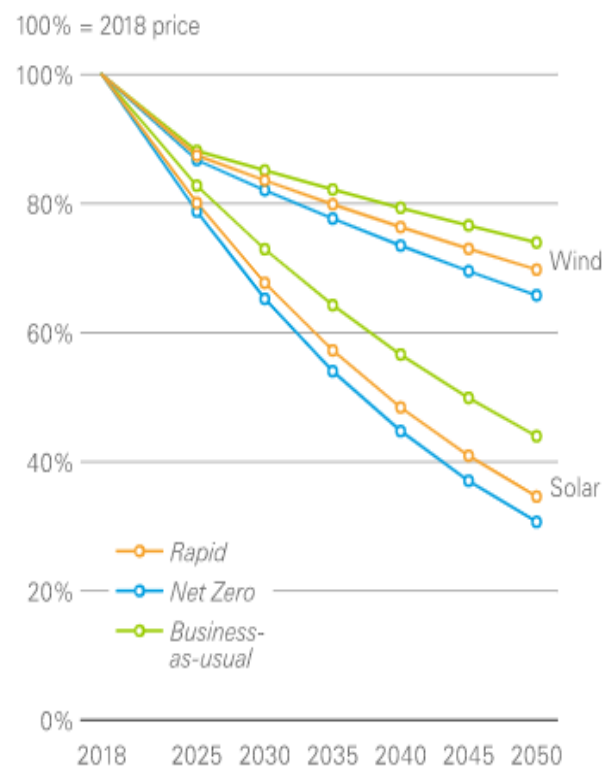


Figura 6. Precio de la energía renovable por escenario. (BP Energy Outlook [31])

El primer escenario muestra un incremento rápido de la producción a partir de energías renovables (superando los 250 EJ). Esto se debe al aumento del uso de la energía solar y eólica, respaldado por la caída de coste debido al descenso de sus curvas de aprendizaje (Figura 6). Además, si se consigue una rápida implantación, con rápida concienciación, y sin oposición a las decisiones tomadas por los distintos países, se llegaría al segundo escenario (hasta los 350 EJ). El último escenario representa un aumento del uso de las energías renovables según el mercado normal, BAU, llegando a 2050 EJ en el año 2050. Este aumento es menor que el de los otros dos escenarios, debido a la lentitud que supone la transición entre las energías renovables y las no renovables.

1.3. Revisión bibliográfica de los modelos de predicción

Una vez se tiene conocimiento de la situación energética actual, y de la importancia de las fuentes de energía renovables en la actualidad y en un futuro cercano, se puede evaluar con mayor conocimiento uno de los factores más importantes dentro de la energía solar: los modelos para la predicción de la radiación solar global. Estos modelos se han vuelto determinantes a la hora de evaluar la potencia que puede dar una planta fotovoltaica en un determinado momento. El impulso que ha tenido la energía solar, unida con el desarrollo de los paneles solares en los últimos años ha hecho necesario saber con precisión la radiación cada vez con una anticipación más ajustada. Si se sabe la radiación aproximada existente en un espacio temporal con suficiente anticipación, se podrá determinar con cierta precisión la potencia, pudiendo así concretar la energía suministrada por la planta a la red de manera más certera, evitando las fluctuaciones que puede haber si la predicción se realiza con un período de anticipación mayor, y por lo tanto reduciendo los costes de operación de la planta.

Son muchos los investigadores que se centran en la elaboración de modelos para la predicción de la radiación solar a corto plazo, dando lugar a diferentes metodologías de predicción, siendo algunas más predominantes que otras. La mayoría de ellas pueden englobarse en las siguientes categorías:

- **Modelos estadísticos.** Son los basados en el “Machine Learning”, también llamado “Deep Learning” o aprendizaje profundo, que se basa en la utilización de datos pasados o series temporales de datos (Time series) combinado con ecuaciones matemáticas lineales o no lineales. En este apartado es donde encontramos las llamadas redes neuronales artificiales o RNA (en inglés ANN – Artificial Neural Networks) o los métodos de lógica difusa o Fuzzy logic.
- **Modelos físicos.** Basados en imágenes y datos tomados por satélites, o modelos que se basan en imágenes de suelo, es decir, en imágenes tomadas desde los edificios. Estos métodos se suelen combinar con el estudio del movimiento de las nubes, para realizar una predicción más precisa. Otros modelos físicos se ayudan de modelos matemáticos para predecir la climatología, y con ello, mejorar la medición a partir de la interacción de las nubes. Dentro de este grupo se encuentran modelos como el WRF-Solar, o el CIADCast [49].
- **Modelos empíricos.** Son aquellas metodologías que buscan una relación empírica entre la predicción y los parámetros de lugar o variables específicas del clima. Dentro de estos modelos podemos encontrar el Angström-Prescott que se centra en las horas de sol o el modelo Hargreaves que se fija en la temperatura máxima y mínima. Este tipo de modelos se están abandonando a favor de otros más sofisticados que consiguen una mayor precisión en la predicción como los anteriormente descritos en la clasificación.

También es posible encontrar modelos que combinen varios de estos métodos buscando la mayor precisión en la predicción.

Haciendo un análisis de la literatura más reciente, se observa como las redes neuronales artificiales se han postulado como preferidas en la realización de los modelos de predicción de radiación solar global a corto plazo. El comportamiento básico de las redes neuronales artificiales es el mismo que el que tienen las neuronas dentro del cuerpo humano, es decir, realizar conexiones entre diferentes puntos de tal manera que se produzca un intercambio de información efectivo, dando lugar a un acto o proceso. En las redes neuronales artificiales estos puntos de conexión se dividen en capas, donde cada una de ellas alberga una información específica.

Para que la predicción sea efectiva, la red neuronal artificial debe tener unos parámetros de entrada basados en bases de datos completas, que sirvan de alimento a los diferentes niveles de información existentes en la red. Estos niveles de información son las llamadas capas ocultas o hidden layers, donde se realiza el intercambio de información. En estas capas se encuentran los llamados pesos, que son una manera de repartir la importancia de dicha información en función de la entrada recibida. Estos pesos serán los encargados de tomar la mejor decisión, y se determinarán gracias a un entrenamiento previo de la red. La capa de salida muestra la predicción de la variable deseada. En la figura 7 se puede encontrar un ejemplo básico de la arquitectura de una red neuronal. De esta manera, si se sigue el ejemplo de una neurona real, un estímulo genera una reacción en una neurona del sistema nervioso, y a través de un intercambio de información, se produce una respuesta a dicho estímulo. Cada red tiene una lógica diferente, lo que da lugar a muchos estudios basados en este tipo de redes.

Un estudio que hace uso de las redes neuronales artificiales es el realizado por M. Marzouq *et al.* [17]. En él, se desarrolla un modelo de red neuronal simple variando el horizonte de la predicción de 1 a 6 horas de anticipación, haciendo uso de los datos históricos de pronóstico. El emplazamiento del estudio es Marruecos, en este caso 28 ciudades, entre las que se encuentran Tetuán, Tánger, o Rabat, entre otras, con el objetivo de ver la precisión del modelo en diferentes tipos de climatología. Se propone un escenario predicción de carácter zonal, de tal manera que se pueda predecir la irradiación en lugares donde no hay suficientes datos meteorológicos. Al mismo tiempo, se realiza una comparación con otros dos escenarios propuestos como son el estudio de la ciudad por separado, y el estudio de todas las ciudades en su conjunto, con el fin de comparar la precisión de la predicción obtenida en los diferentes escenarios.

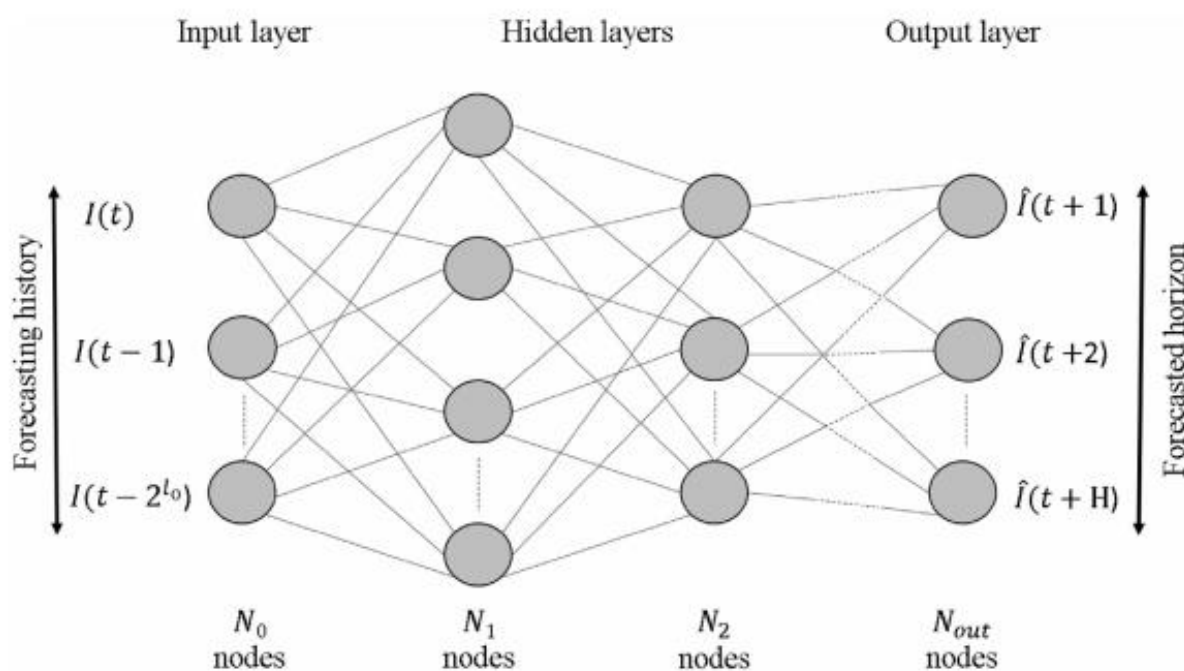


Figura 7. Red neuronal adoptada por Marzouq et al. [17]

Con el paso del tiempo, estas redes neuronales artificiales han sufrido variaciones a medida que ha aumentado el número de estudios basados en esta metodología, con el objetivo de aumentar la precisión del método. Es el caso de las redes neuronales recurrentes (RNN – Recurrent Neural Networks) donde se hace uso de bucles de realimentación para mejorar la precisión del método. El estudio realizado por Z. Pang *et al.* [19] trata de analizar la precisión que tiene el modelo clásico de redes neuronales artificiales (ANN) y el uso de la inteligencia artificial (IA) en la predicción de la radiación solar global en Alabama, EE. UU. con un horizonte temporal de una hora de antelación. Del mismo modo, se pone de manifiesto que si se hace uso de las redes neuronales recurrentes aumenta el rendimiento del algoritmo de predicción. Al igual que otros métodos, los datos meteorológicos necesarios como parámetros de entrada de los distintos modelos se recogen de una estación meteorológica local, situada en el mismo lugar donde se realiza el estudio. Se observa al mismo tiempo, que, si se reduce la ventana temporal a intervalos de diez minutos e introducción de distintos

escenarios móviles para la selección de datos, aumenta la predicción de la predicción. Obviamente, el tiempo de ejecución del modelo RNN es mayor que el del modelo ANN, y ambos tiempos se incrementan cuando se añaden las modificaciones comentadas con anterioridad. Por ello, se concluye que la única limitación que se puede atribuir al modelo RNN es su alto coste computacional, dando beneficios considerables en la predicción.

Con la idea de mejorar los modelos de predicción, surgen también modificaciones de las redes recurrentes. M. Sorkun *et al.* [22], se centran en desarrollar un modelo estadístico para predecir la radiación solar global a corto plazo con un horizonte temporal de una hora haciendo uso de una variación de la red neuronal recurrente (RNN), llamada unidad de memoria a corto plazo (LSTM – Long Short-Term Memory). Los parámetros de entrada son filtrados para localizar cuales son los óptimos para construir el modelo de la mejor manera posible, y con ello, el que mejor se ajuste a la predicción solar global. Por su parte, J. Wojtkiewicz *et al.* [26] tratan de predecir la radiación solar con una hora de anticipación en Phoenix (Arizona) haciendo uso de unidades recurrentes cerradas multivariantes (GRU – Gated Recurrent Units), un modelo basado en RNN y parecido al LSTM, pero más sencillo en implementación y ejecución (dos puertas lógicas para las celdas GRU por tres para LSTM).

También se puede hacer una hibridación de dos modelos que, aunque parecidos, no son del todo iguales, de tal manera que aumente la precisión de la predicción siempre que cada uno se emplee de forma correcta en el modelo que se vaya a proponer. En el artículo [28] Y. Yu *et al.*, proponen un modelo basado en las redes neuronales recurrentes (RNN) combinado con la unidad LSTM para la predicción de la radiación solar global a corto plazo bajo condiciones climatológicas difíciles en Atlanta y Hawái. Cuando el clima varía de forma intensa, el modelo RNN para la selección de las variables óptimas de entrada al modelo puede quedar relegado a un plano local, no siendo aplicable de forma continua en distintas predicciones. Por ello, el modelo LSTM hace que se recuerde la información contenida en las distintas secuencias temporales, eliminando los problemas causados por la RNN y mejorando la precisión de la predicción. El estudio se realiza para horizontes temporales comprendidos entre un día y una hora de antelación.

Esta hibridación también se puede realizar sustituyendo la red RNN por una red neuronal convolucional (CNN – Convolutional Neural Network). Esta visión la comparten S. Ghimire *et al.* [9] en su trabajo, donde desarrollan un modelo híbrido de predicción basado en el uso de una red neuronal convolucional para el reconocimiento de patrones con la red de memoria a corto plazo (LSTM), con el objetivo de predecir la radiación solar global cada media hora. La red convolucional reconoce los parámetros característicos necesarios extraídos de mediciones pasadas entre 2006 y 2018 en Alice Springs (Australia), y la red de memoria a corto plazo los toma para realizar la predicción de forma más precisa. De esta manera se consigue aunar las ventajas de ambos métodos.

H. Zang *et al.* [29], al igual que [9], desarrolla un método de predicción solar basado en la hibridación de una CNN y una red de memoria LSTM. De la misma manera, se aplica la red convolucional para extraer características espaciales de una matriz bidimensional compuesta por datos meteorológicos asociados a un lugar objetivo y los lugares colindantes. Posteriormente, se aplica la red de memoria a corto plazo para extraer los parámetros temporales históricos asociados al lugar de estudio. Con estos parámetros de entrada, se desarrollan ciertas correlaciones espaciotemporales para predecir la radiación horizontal global con una hora de anticipación. Se evalúa la fiabilidad de este método a lo largo de todo un año, para diversas estaciones y condiciones climatológicas (34 ubicaciones en Texas EE. UU).

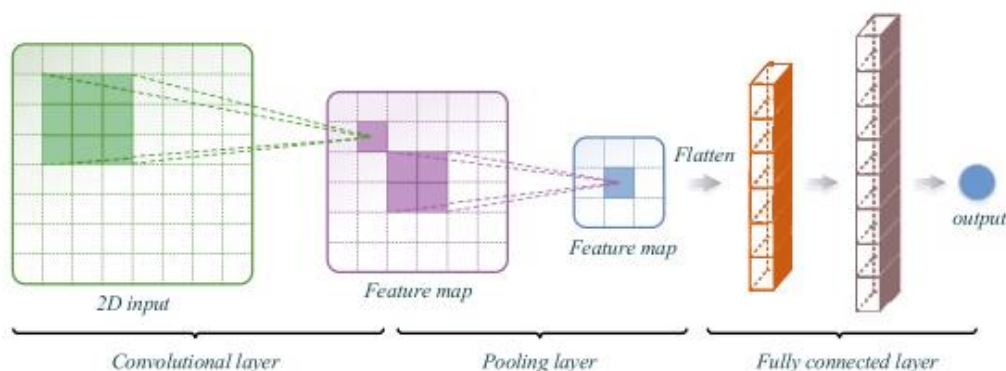


Figura 8. Arquitectura básica de una CNN. [29]

Otro estudio que hace uso de las CNN combinadas con el modelo LSTM es el escrito por J.-M. Yeom *et al.* [27]. Ayudándose de imágenes de radiación solar contemporáneas del sistema de imágenes meteorológicas de comunicaciones, océanos y satélites meteorológicos (COMS – Communication, Ocean, and Meteorological Satellite) para diseñar una red neuronal convolucional y un modelo predictivo LSTM. Este estudio se realiza en la península de Corea, y tiene por objetivo la predicción de la radiación solar global a corto plazo con una hora de antelación. Además, proporcionaba capturas de las variaciones temporales de las condiciones atmosféricas, lo que probablemente influya en el suministro de la energía solar y su estabilidad en general.

También se puede realizar combinaciones con otros modelos que no estén basados en redes neuronales. En su estudio, Y. Ge, Y. Nan y L. Bai [8] desarrollan un método de predicción de la radiación solar basado en una hibridación entre el modelo LSTM, y un método empírico de procesamiento de señales (EMD). El método EMD se utiliza para descomponer los datos de la secuencia temporal en una serie de secuencias de componentes estables. La predicción tiene por objetivo determinar con la mayor precisión posible la radiación solar con una hora de anticipación.

Las redes neuronales, tanto básicas como avanzadas, pueden ir precedidas por métodos de aprendizaje que logren identificar las variables más significativas y agruparlas de tal manera que mejore el comportamiento de las redes neuronales. En la figura 9 se puede ver una clasificación de los métodos de aprendizaje.

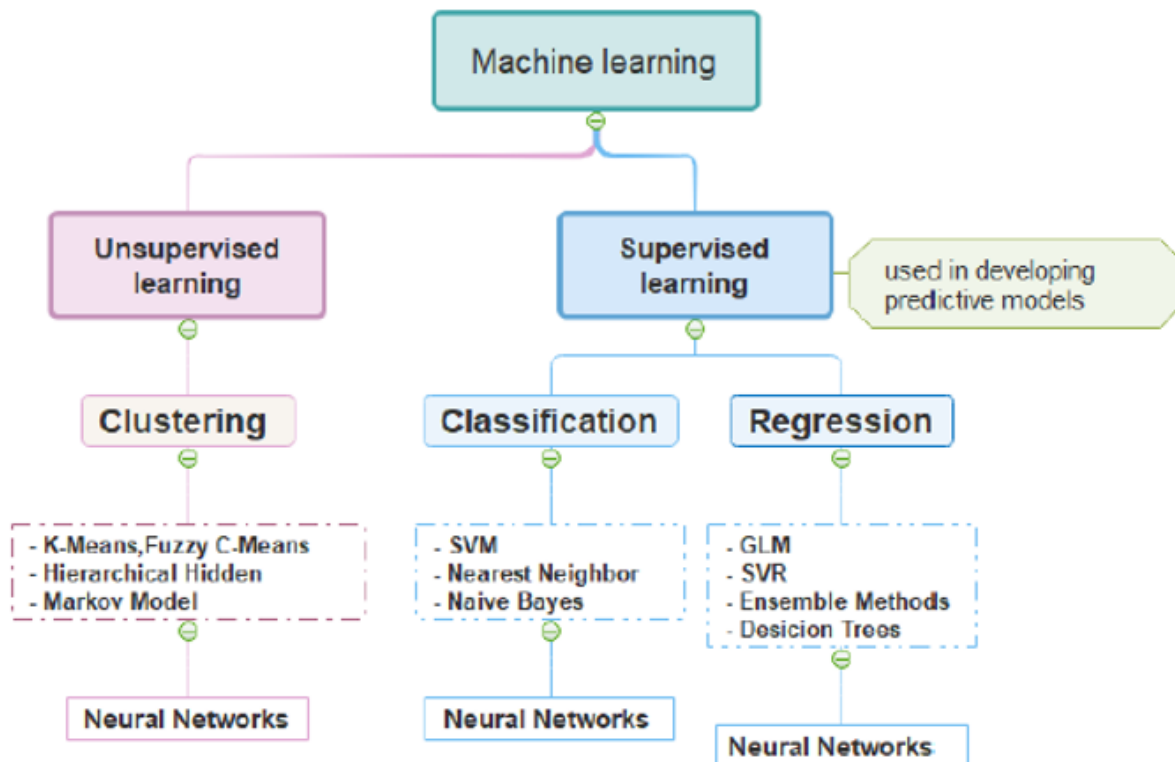


Figura 9. Clasificación de los métodos de aprendizaje. [7]

Existen dos tipos de métodos de aprendizaje: los supervisados, y los no supervisados. En el primero de los casos se trabaja con datos etiquetados, de tal manera que el algoritmo solo tiene que describir una función que, dada las variables de entrada, le asigna la salida más adecuada. en el caso de los datos no supervisados, no se conoce la relación entre las variables, por lo que el modelo debe definir una estructura de datos para facilitar y mejorar el análisis que se esté realizando.

De esta manera, O. El Alani *et al.* [7], mediante el uso de redes neuronales artificiales, tratan de predecir la radiación a corto plazo para un clima semiárido, exactamente en Benguerir, Marruecos. Como datos de partida extraen las múltiples variables meteorológicas (entre ellas se encuentra la humedad relativa, la temperatura, la presión o la velocidad del viento), así como la radiación solar global que le proporciona una estación situada en el mismo lugar donde se realiza el estudio. Para filtrar los datos de entrada hace uso del algoritmo LM

(Levenberg-Marquardt), encuadrado dentro de los algoritmos de aprendizaje supervisados de regresión. Con este conjunto de variables, se realiza una pequeña evaluación de cuál es el modelo arquitectónico de combinación de variables más eficiente para lograr la mayor precisión posible en la predicción.

Otros, como [8] o [28], clasifican los parámetros a través de la agrupación por “k-means”, en categorías diferenciadas. Las redes neuronales se emplean en este punto para ver si las diferentes aproximaciones pueden o no generalizarse.

Más allá de las redes neuronales, encontramos métodos que se basan en datos satelitales, y telemetría de suelo. En este apartado, encontramos el trabajo de investigación realizado por R. Alonso-Suárez, M. David, V. Branco y P. Lauret [1]. En él, se tiene por objetivo la construcción de tres modelos diferentes basados en líneas de regresión (LQR – Linear Quantile Regression) para producir pronósticos solares probabilísticos intradiarios (en el tiempo t la distribución de probabilidad de irradiancia solar futura) con un horizonte temporal desde diez minutos a tres horas con intervalos de diez minutos. Cada modelo tiene unos parámetros de entrada diferentes basados en datos satelitales y telemetría de suelo: el primero de ellos toma valores de mediciones de radiación solar en suelo actuales y pasadas; el segundo agrega la variabilidad de la radiación global tomada de las mediciones pasadas como parámetro para la predicción; y el tercero hace uso del albedo promedio espacial tomado satelitalmente. Se muestra como la adición de la variabilidad reduce el número de predicciones del intervalo, y cómo la información satelital mejora los pronósticos. El estudio se realiza para sitios con climas muy diferentes (dos secos: Desert Rock, Fort Peck; tres insulares: Oahu, Fouillole, Tampom; y dos subtropicales: Salto, y Sao Martinho da Serra), por lo que es necesario que la información sea fiable.

R. Alonso-Suárez investiga también otros métodos de predicción intradiaria haciendo uso de modelos físicos. En este caso junto a G. Giacosa [10], realizan un estudio sobre la predicción solar haciendo uso de la técnica de pronóstico de movimiento de nubes por satélite (CVM) propuesta por Lorenz et al. en el año 2004. Para estimar el movimiento de las nubes en cada píxel, el sistema evalúa una imagen de la nubosidad en una determinada área, y la compara con una imagen anterior de área superior involucrando las regiones colindantes. La predicción de la irradiancia solar global por hora se hace con una antelación de entre una y cinco horas, tanto para la ubicación donde se sitúa la estación como para la región en la que se encuentra.

Otro sistema que hace uso de las imágenes de cielo es el realizado por F. Wang *et al.* [25]. Se propone un método de predicción de la radiación solar con un horizonte temporal que varía entre uno y diez minutos, basado en un mapeo de la irradiancia de superficie en tiempo real. Debido a que la relación de mapeo entre la imagen del cielo y la irradiancia está cambiando dinámicamente, el modelo de mapeo está configurado para ser actualizado de acuerdo con los datos a tiempo real. Le siguen por el mismo camino S. Dev *et al.*, que en su estudio realizado en la Universidad de Singapur [6] evalúan las capturas en intervalos regulares usando cámaras con una lente del tipo ojo de pez. Las imágenes tomadas contienen información sobre la nubosidad existente en un preciso momento, y se combinan para dar lugar a las predicciones.



Figura 10. Ejemplo de una imagen tomada por una cámara de cielo. [37]

Todos los modelos pueden dar lugar a combinaciones y comparaciones entre ellos. F. J. Rodríguez-Benítez *et al.* [20] realizan un modelo de predicción solar a corto plazo para la península ibérica dividido en dos partes, donde se evalúa la capacidad de cuatro modelos para proporcionar pronósticos de radiación solar global y directa con hasta 6 horas de anticipación. Este estudio se realiza basándose en dos años de datos recopilados en cuatro estaciones meteorológicas. En su mayoría, los modelos siguen enfoques independientes: un modelo estadístico puro (Smart Persistence – SP), un modelo basado en CVM derivado de imágenes de satélite, un modelo fisicomatemático NWP (WRF-Solar) y un modelo híbrido satélite-NWP (CIADCast). Los resultados que determinan que el modelo SP tiene mejores resultados para un horizonte temporal bajo, mientras que los modelos basados en datos satelitales tienen mejores resultados en los horizontes intermedios y el WRF-Solar, al final del período temporal de pronóstico estudiado. El equilibrio entre los modelos satelitales y el WRF-Solar varía entre una y tres horas para la radiación solar global, y entre tres y cinco horas para la directa. Si evaluamos el rendimiento que tienen los distintos modelos, se puede observar que la predicción depende mucho de las condiciones topográficas y geográficas de la estación de medida, la variable que se esté evaluando, y el cielo y las condiciones meteorológicas.

Debido a la complementariedad de los diferentes modelos descritos en la primera parte del estudio, en la segunda parte [21] se evalúa las ventajas logradas con la combinación de dichos modelos. De esta manera, a través de un modelo SVM (Support Vector Machine), al igual que se hiciera en la selección de variables en los modelos basados en redes neuronales, se identifican las mejoras logradas en las previsiones de la radiación solar global. Se proponen dos enfoques diferentes para la combinación: el primero de ellos es combinar los métodos según el horizonte temporal de pronóstico; el enfoque general propone un modelo único válido para todos los horizontes. Al mismo tiempo se evalúa la influencia de combinar información sobre los distintos tipos de clima. La combinación da como resultado una mejora de la predicción, sin grandes diferencias entre los dos tipos de enfoque de combinación.

Otro estudio que combina varios métodos es el realizado por K. Skeie y A. Gustavsen [24] en el que se centran en la elaboración de un método que permita calcular la radiación solar global por hora en Trondheim, Noruega. Para ello toman las mediciones de suelo disponibles de nubosidad, temperatura, y precipitación. Para calcular la predicción de la radiación en los días nublados se sirven del método MAC, desarrollado por J.A. Davies y D.C. McKay (1982) [18]. En lugar de basarse en observaciones terrestres, como hace el método original, la cobertura de nubes en tres niveles y la cobertura total de nubes se utilizan como entrada en una parametrización del producto de nubes. La irradiancia horizontal global resultante se valida con la salida del modelo de predicción numérica del tiempo (NWP).

J. Jiang *et al.* en su artículo [13] establecen dos métodos de predicción de la radiación solar global en un plazo muy reducido de tiempo (intervalos comprendidos entre uno y diez minutos). Para ello, emplean la Regresión Lineal Ternaria (TLR) en el primero, y las redes neuronales de propagación con paso hacia atrás (BPN) en el segundo, combinado con la observación de imágenes de cielo y haciendo uso de un radiómetro colocado en una planta fotovoltaica en la ciudad de Dunhuang, China. De esta manera, toman una foto del cielo cada minuto, y procesan las nubes con diferentes puntos de medida, con el objetivo de generar un mapa bidimensional de la situación de las nubes en el cielo. Posteriormente, hacen uso del método de la velocidad de imagen de partículas (PIV), para analizar el movimiento de las nubes en el futuro cercano. Con se despliega un conjunto de parámetros de entrada para los dos modelos de predicción.

Prácticamente la totalidad de los estudios se realizaron tanto para días despejados como para días nublados, obteniendo distintos resultados de los índices seleccionados para evaluar la precisión, y obteniendo mejores resultados en los días claros respecto de los nublados. Esta situación es normal debido a la dificultad de predicción que tiene los días con alto índice de nubosidad, ya que la distinta tipología de cada nube entorpece la predicción. Es necesario que, aparte de ser un método seguro y fiable en situaciones triviales, como puede ser un día soleado sin nubes, el método responda bien ante situaciones con una gran densidad de nubes. La mayoría de los métodos de la literatura obtienen peores resultados para días nublados, ya que se incrementa la dificultad de forma considerable.

En la tabla 1 se encuentra un resumen de todos los métodos descritos en la revisión bibliográfica y sus características más importantes.

Ref.	Variable predicha	Modelo	Variables	Horizonte temporal	Localización
S. Ghimire <i>et al.</i>	GHI	CNN- LSTM	GHI histórica	30 min	Alice Springs (Australia)
R. Alonso-Suárez, M. David, V. Branco y P. Lauret	GHI	3 modelos LQR	Datos satelitales. Telemetría de suelo.	10 min - 3h	7 climas diferentes (Dos secos: Desert Rock, Fort Peck; tres insulares: Oahu, Fouillole, Tampom; y dos subtropicales: Salto, y Sao Martinho da Serra)
R. Alonso-Suárez, G. Giacosa	GHI	CVM	GHI. Imágenes satelitales.	1h - 5h	Uruguay
K. Skeie y A. Gustavsen	GHI	MAC-NWP	GHI histórica. Variables meteorológicas.	1h	Trondheim, Noruega
H. Zang <i>et al.</i>	GHI	CNN- LSTM	GHI y variables meteorológicas (históricas y previas a la predicción)	1h	Texas EE. UU
M. Sorkun <i>et al.</i>	GHI	LSTM	GHI y variables meteorológicas (históricas y previas a la predicción)	1h	Ajaccio, Córcega
Y. Ge, Y. Nan y L. Bai	GHI	EMD-LSTM	GHI y variables meteorológicas (históricas y previas a la predicción)	1h - 24h	Michigan, Alabama y Nevada
F. Wang <i>et al.</i>	GHI	Sky images mapping	GHI. Imágenes de cielo.	1min - 10 min	-
S. Dev <i>et al.</i>	GHI	Sky images mapping	Variables meteorológicas y GHI. Imágenes de cielo.	1min-6min	Singapur
F. J. Rodríguez-Benítez <i>et al.</i> vol. I	GHI	4 modelos: CVM, Smart persistence, NWP, satellite-NWP	GHI y variables meteorológicas (históricas y previas a la predicción). Imágenes satelitales.	Hasta 6h	Península ibérica

F. J. Rodríguez-Benítez <i>et al.</i> vol. II	GHI	Combinación Vol. 1 / Evaluación con SVM	GHI y variables meteorológicas (históricas y previas a la predicción). Imágenes satelitales.	Hasta 6h	Península ibérica
J. Wojtkiewicz <i>et al.</i>	GHI	GRU	GHI y variables meteorológicas (históricas y previas a la predicción). Datos de nubosidad.	1h	Phoenix
M. Marzouq <i>et al.</i>	GHI	ANN	GHI histórica.	Hasta 6h	Marruecos
Z. Pang <i>et al.</i>	GHI	ANN y RNN	Variables meteorológicas (históricas y previas a la predicción)	1h	Alabama
J. Jiang <i>et al.</i>	GHI	Sky images (TLR-BPN)	GHI. Imágenes de cielo.	1min - 10 min	Dunhuang, China
O. El Alani <i>et al.</i>	GHI	LM-ANN	Variables meteorológicas y geometría solar (históricas y previas a la predicción)	No definido	Benguerir, Marruecos
Y. Yu <i>et al.</i>	GHI	RNN-LSTM	GHI y variables meteorológicas (históricas y previas a la predicción). (Categorización de la nubosidad)	1h - 24h	Atlanta y Hawái
J.-M. Yeom <i>et al.</i>	GHI	CNN-LSTM	GHI (históricos y previos a la predicción). Datos satelitales.	1h	Corea del sur

Tabla 1. Tabla resumen de los modelos de predicción de la radiación solar. (Elaboración propia)

2 OBJETO DEL TRABAJO

Una vez se ha evaluado la situación de la producción eléctrica a nivel mundial y nacional, destacando el papel que juegan las energías renovables en el mix energético, y se ha hecho una revisión bibliográfica de los modelos de predicción de la radiación solar global sobre una superficie horizontal existentes en la actualidad, es conveniente valorar los objetivos propuestos en el presente proyecto.

El primero de ellos es encontrar la capacidad de realizar una predicción de la radiación fiable y precisa a través de las herramientas informáticas disponibles por parte de la comunidad universitaria. Al mismo tiempo, que la herramienta elegida ofrezca un amplio abanico de posibilidades de trabajo sin entrar en complejidades excesivas que hagan perder de vista el objetivo de la predicción.

El segundo objetivo es comprender la importancia que tienen los diferentes modelos de redes neuronales artificiales actualmente y cómo se sirven de las series temporales para lograr un rendimiento óptimo. Además, tener conciencia de que existen un gran número de modelos basados en otros conceptos, como pueden ser los basados en imágenes de cielo, que conviven con los modelos de redes neuronales artificiales, y que en ciertas situaciones también consiguen una precisión elevada en la predicción de la radiación, ampliando el espectro de estudio de los investigadores, y no cerrándose a una única línea de trabajo.

En tercer lugar, se tiene como objetivo evaluar la fiabilidad del modelo elegido en una localización como Sevilla con las herramientas disponibles. Se dispondrá de una amplia base de datos comprendida entre el 1 de enero del año 2000 y el 31 de diciembre del año 2020, lo que ofrecerá una amplia gama de posibilidades de trabajo sin necesidad de recurrir a fuentes externas de datos. Posteriormente, los resultados obtenidos se compararán con artículos científicos con el fin de contrastar la precisión del modelo de predicción elegido en el presente trabajo.

3 METODOLOGÍA

Una vez se ha estudiado la literatura existente de los diferentes tipos de modelos de predicción de la radiación solar global, será necesario elegir un modelo con el que realizar una predicción y, posteriormente, evaluar los resultados para ver su fiabilidad frente a diferentes situaciones meteorológicas.

En la primera parte de este apartado se expondrá la base de datos disponible para llevar a cabo la predicción. En segundo lugar, se describirá la herramienta elegida analizando las ventajas y los inconvenientes que presenta para reproducir un modelo de predicción consistente basado en redes neuronales. Por último, se desarrollará un análisis de los distintos modelos implementados, así como las limitaciones que se han encontrado a la hora de implementarlos.

3.1. Base de datos

Antes de entrar en el detalle de la metodología seguida para la predicción, será necesario tener conocimiento del lugar donde se va a realizar el estudio y la base de datos disponible para el mismo. Se ha decidido realizar la predicción en Sevilla, y más concretamente, en la Isla de la Cartuja. Esta elección viene motivada por la facilidad de acceder a los datos meteorológicos necesarios para realizar un rango amplio de los modelos estudiados previamente. Gracias a la estación de medida que el Grupo de Investigación de Termodinámica y Energías Renovables (GTER) posee en la Escuela de Ingenieros de Sevilla ($37^{\circ} 41' N$, $6^{\circ} 01' O$; Elevación: 12m), y que se pone a disposición para este trabajo, se puede acceder a diferentes variables meteorológicas como:

- Irradiancia solar global sobre una superficie horizontal mediante piranómetro Eko, modelo ms80. Viene determinada por la variable GHI (*Global Solar Irradiance*), y se define como los vatios que inciden sobre una superficie horizontal en un instante determinado (W/m^2). En la figura 11 se puede observar los valores de radiación referidos al período comprendido entre el 1 de enero de 2019 y el 15 de enero de 2019 con una granularidad de una hora.

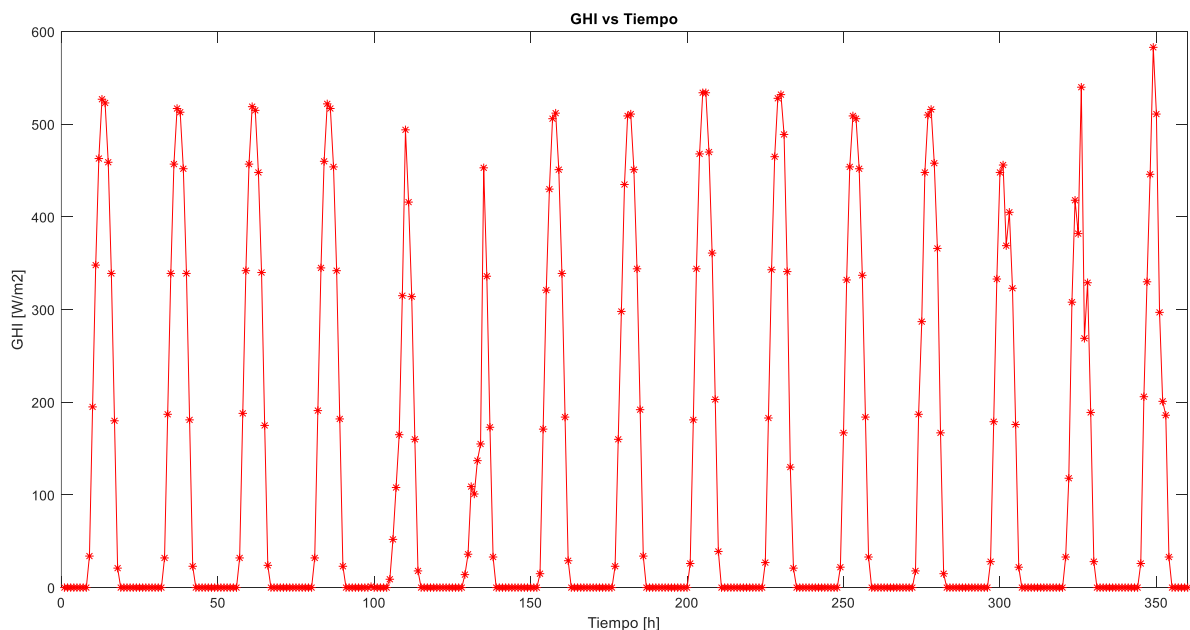


Figura 11. GHI vs Tiempo. (Elaboración propia)

- Irradiancia difusa medida con sistema de banda de sombra (W/m^2).
- Irradiancia global sobre superficie horizontal mediante piranómetro Eko, modelo ms80 (W/m^2).
- Irradiancia global sobre superficie inclinada 27° mediante piranómetro Hukseflux, modelo SR20 (W/m^2).
- Irradiancia difusa medida con sistema de sombreado de bolas (W/m^2).
- Irradiancia global sobre superficie inclinada 27° mediante célula calibrada Atersa (W/m^2).
- Irradiancia global sobre superficie horizontal mediante piranómetro Skye SP1110 photodiode (W/m^2).
- Irradiancia global sobre superficie inclinada 27° mediante piranómetro Skye SP1110 photodiode (W/m^2).
- Irradiancia de onda larga medida con pirgeómetro (W/m^2).
- Irradiancia global sobre superficie horizontal mediante célula calibrada Atersa (W/m^2).
- Irradiancia directa normal medida con pirheliómetro Eppley, modelo NIP (W/m^2).
- Irradiancia directa normal medida con pirheliómetro Kipp and Zonen, modelo CHP1 (W/m^2).
- Hora solar.
- Velocidad del viento. Medida en metros por segundo (m/s).
- Dirección del viento. Medido en grados ($^\circ$) respecto de la dirección norte, medidos en el sentido de las agujas del reloj.
- Humedad relativa. Medida en moles de vapor de agua entre moles de vapor de agua en aire saturado. Se expresa en porcentaje (%).
- Presión atmosférica mediante sonda barométrica (mBar)
- Temperatura ambiente. Medida en grados centígrados ($^\circ C$).

Haciendo uso de estas variables, podemos calcular otros datos relevantes como la temperatura de rocío ($^\circ C$) La granularidad de estas mediciones es de 5 segundos, y se disponen de mediciones desde el 1 de enero del año 2000, por lo que se pueden realizar predicciones en un amplio espectro temporal.

3.2. Tratamiento de redes neuronales con *MATLAB*

Dentro de los diferentes modelos existentes, aquellos que tienen una mayor proyección, a la vista de lo expuesto previamente en este documento, son las redes neuronales. Al mismo tiempo, la herramienta informática de la que se dispone es el software matemático *Matlab*. Este software dispone de una aplicación llamada “*Deep Network Designer*” donde se pueden elaborar distintos modelos de redes neuronales.

Las unidades de predicción para series temporales disponibles en la aplicación son la unidad LSTM, la unidad BiLSTM, y la unidad GRU. Al mismo tiempo permite desarrollar paso a paso las redes neuronales básicas como la ANN, o la CNN. Ante lo expuesto en la revisión bibliográfica, se ha decidido seguir con el estudio de las posibilidades que ofrecen estas redes neuronales con “memoria”, ya que todos los estudios que la utilizaron mejoraron sus resultados respecto a las redes primarias. Además, hasta la fecha *Matlab* no permite la consecución de distintas redes neuronales al mismo tiempo de forma que se pueda alcanzar una red compuesta, por lo que el estudio se limita a estas tres unidades neuronales simples. Esta limitación hace que no sea posible combinar las distintas variables meteorológicas para la predicción, dejando únicamente como variable de entrada la radiación solar horizontal. Esto se debe a que en los estudios que combinan distintas variables meteorológicas se hace uso de varias redes de forma simultánea, de tal manera que se puedan obtener los datos más relevantes de cada una de ellas. En nuestro caso, la implementación de una red única alimentada con un número mayor de variables únicamente lleva a la introducción de un mayor error en la predicción, lo que comúnmente se conoce como “ruido”. Los resultados se exponen en el apartado 4.7.

Por el contrario, esta aplicación ofrece claridad y simpleza para comenzar en el aprendizaje de cómo funcionan las redes neuronales. Además, ofrece la capacidad de representar los resultados y validarlos con sencillez sin necesidad de ser un programador experimentado en varios idiomas, ya que el lenguaje en el que se basa es C++, ampliamente conocido en la comunidad científica. Por último, *Matlab* es una herramienta accesible para muchos, y, en el caso de este trabajo, disponible en el servidor de la Universidad de Sevilla con el objetivo de usarlo con fines académicos.

Dentro de las redes disponibles, la más contrastada y que arroja mejores resultados en la predicción de la radiación solar horizontal según se explica en los artículos existentes es la unidad LSTM. Por otro lado, según se expuso en la revisión bibliográfica, la unidad GRU es una unidad más simple que la LSTM, mejorando solo el coste computacional de la red y sin resultados evidentes en la predicción de la radiación. Finalmente, la unidad BiLSTM no tiene aplicación en la predicción haciendo uso de las series temporales. Esta red se utiliza para clasificación de imágenes y datos.

En resumen, una vez evaluadas las ventajas y los inconvenientes de la herramienta *Matlab*, se ha decidido evaluar el desempeño de la unidad LSTM a fin de verificar su desempeño en la localización de estudio. La variable de entrada de los modelos será la radiación solar horizontal histórica. De la misma manera, se expondrán algunos resultados de la red haciendo uso de la unidad GRU, con el fin de contrastar las afirmaciones realizadas por los científicos en sus estudios.

3.3. Modelo utilizado

En este apartado se va a desarrollar el modelo que se ha empleado para la predicción de la radiación solar horizontal con una hora de anticipación.

En la figura 12 se puede observar la arquitectura de la red a estudiar. El primero de ellos es un bloque de entrada secuencial. Este bloque recoge los datos de un fichero determinado y lo pone a disposición del siguiente bloque. El segundo bloque es la unidad LSTM. Esta unidad, cuyo funcionamiento se desarrollará más adelante, es la unidad oculta de la red que se desea modelar. El número de unidades ocultas que le asignamos a nuestra red puede ser variable. El tercer bloque que compone la red es una capa de conexión múltiple. Esta capa toma todos los resultados y los aglutina de tal manera que reduzca el número de valores resultantes de la unidad anterior. Y por último se encuentra la capa de regresión, en la que se muestra el valor resultante de la predicción en un tiempo determinado.

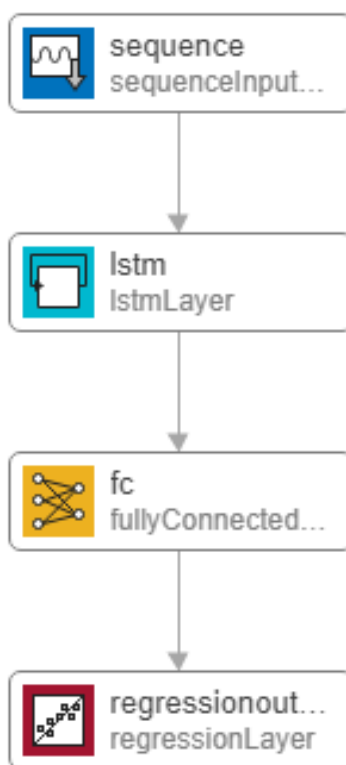


Figura 12. Arquitectura de la red empleada. (Elaboración propia)

3.3.1. Tratamiento de previo y flujo de datos

Antes de realizar la programación del modelo en cuestión, es necesario preparar los datos de entrada en el formato especificado para la red. En *Matlab*, las redes neuronales basadas en series temporales deben tener los valores de entrada contenidos en un formato específico para que se pueda llevar a cabo la lectura por parte de la red. Por ello, hay que convertir los datos archivados por día en archivos .txt, a una secuencia temporal archivada en un archivo .xlsx. Esto se realizará gracias a un bucle que recorre todos los archivos de entrada de la estación modificando su extensión. Para disminuir el tiempo de computación en la predicción, ya que el número de datos es muy elevado, se ha tomado una granularidad de 1 hora. De esta manera, ya que la granularidad de los datos es de 5 segundos, otro bucle selecciona las variables de cada hora, esto es, uno cada 720 registros, y las almacena en un Excel con un formato lineal.

Al mismo tiempo, hay que tratar los valores negativos que pueda haber en las mediciones en las horas nocturnas tomando valor 0 en esos casos. Por ello, se emplea un bucle que recorre la secuencia de datos completa poniendo a 0 los valores negativos que encuentre. Con ello se conseguirá tener una secuencia final completa de tal forma que se eviten datos negativos erróneos. Gracias a la formulación de la programación en *Matlab*, se puede realizar una modificación sencilla en el código de tal manera que el espectro temporal y la granularidad que se toman en el proceso puedan ser elegidos en función de las necesidades de la persona que lo vaya a utilizar. El rango de valores tratados se extiende desde el 1 de enero de 2015 al 31 de diciembre de 2019. Este volumen de valores se considera válido y suficientemente extenso para la aplicación en la predicción de la radiación solar horizontal. Además, los equipos de medida se han mejorado a lo largo del tiempo, por lo que se ha preferido coger un rango de valores más próximo al actual, donde las medidas son más precisas. El año 2020 se ha excluido debido a que los valores tomados para este año aún no estaban corregidos, por lo que los años anteriores tienen una mayor fiabilidad.

Una vez se han tratado los valores de entrada, debido a la naturaleza de la unidad LSTM, es necesario realizar una normalización de estos, de tal manera que los valores de entrada a la red estén comprendidos entre dos valores determinados. En una primera aproximación se ha decidido normalizar los valores haciendo uso de la media y la varianza de la distribución temporal tomada en el estudio de la red a través de la siguiente expresión:

$$X_{Nt} = \frac{X_t \cdot \mu}{\sigma}$$

siendo X_t el valor de la radiación solar global en un determinado tiempo t , μ la media de la distribución temporal elegida, σ la varianza de la distribución temporal elegida, y X_{Nt} el valor normalizado comprendido entre 0 y 1 de la radiación en un tiempo determinado t .

3.3.2. Unidad LSTM

Como referencia principal, se ha tomado el estudio realizado por H. Zang *et al.* [29], adaptándolo al emplazamiento elegido y a los datos disponibles. También se han tenido en cuenta otros documentos como [14] o [22]. El factor común de estos estudios es el uso de la unidad de memoria a corto plazo (LSTM – Long Short Term Memory).

Habiendo determinado la red neuronal, y antes de implementar el modelo, es necesario profundizar en su funcionamiento. Este tipo de redes, como se comentó anteriormente, es una modificación de una red neuronal recurrente (RNN). Estas redes hacían uso de bucles de realimentación dentro de las capas ocultas para mejorar la precisión del modelo. La principal ventaja de la unidad LSTM radica en su capacidad de “recordar” u “olvidar” información temporal intermedia del proceso, lo que mejora la desaparición de datos relevantes cuando las series temporales de entrada son demasiado extensas, y también facilita la extracción de datos de estas.

En la figura 13 se puede ver la arquitectura de la unidad, compuesta por tres puertas lógicas diferenciadas llamadas puerta de entrada (input), puerta de salida (output), y puerta de olvido (forget). Las operaciones que

se realizan en estas puertas se pueden desarrollar como:

$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$o_t = \sigma (W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f)$$

donde x_t se refiere a la variable de entrada en el tiempo t , h_{t-1} denota la salida de la predicción en el tiempo $t-1$, W_i, W_o, W_f se refiere a los pesos asignados a las diferentes puertas lógicas y b_i, b_o, b_f determinan los llamados errores de los cálculos correspondientes. Por otro lado, σ representa a la función de activación sigmoideal.

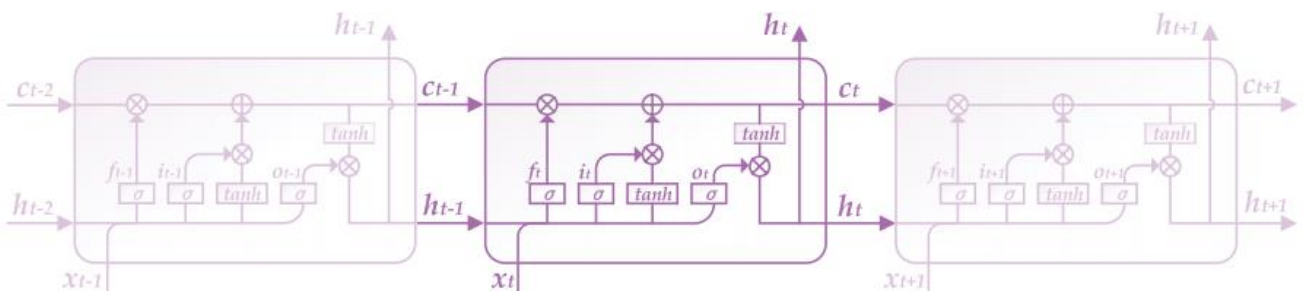


Figura 13. Arquitectura básica de la unidad LSTM (H. Zang *et al.* [29])

La salida de la unidad LSTM viene determinada por las variables c_t que es la celda de memoria en el tiempo t , y h_t que es la variable predicha en el tiempo t . Los cálculos que se realizan son:

$$\hat{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

$$c_t = c_{t-1} * f_t + i_t * \hat{c}_t$$

$$h_t = o_t * \tanh(c_t)$$

Donde $\tanh()$ es la función de activación de la tangente hiperbólica, W_c es la matriz de peso asociado a la celda de memoria, b_c es el error de cálculo, y \hat{c}_t es la celda de memoria temporal en el tiempo t .

3.3.3. Índices de precisión

Es necesario fijar unos índices para poder comparar qué estimación es mejor dentro de los distintos modelos propuestos. Los índices más usados en la actualidad son los siguientes:

- MAE (*Medium-Absolut-Error*). Conocido como el error absoluto medio, y utilizado en la comparación de dos variables continuas en una predicción. Su expresión es:

$$\frac{1}{N} \cdot \sum_{i=1}^N (y_{i,pred} - y_{i,med})$$

- RMSE (*Root-Mean-Square Error*). Es el denominado error cuadrático medio, y que se considera preciso cuando se está evaluando la precisión de un modelo frente al valor real medido, y se calcula como:

$$\sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (y_{i,pred} - y_{i,med})^2}$$

- nRMSE (*normalized Root-Mean-Square Error*). Es igual que el anterior, pero normalizando el resultado haciendo uso de la media de la predicción.

$$\frac{\sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (y_{i,pred} - y_{i,med})^2}}{y_a} \cdot 100\%$$

- R². Es el llamado coeficiente de determinación. Es un índice estadístico usado para determinar la precisión de una predicción, y viene determinado por la siguiente expresión:

$$\frac{|\sum_{i=1}^N (y_{i,pred} - y_{a,pred}) \cdot (y_{i,med} - y_{a,med})|}{\sqrt{[\sum_{i=1}^N (y_{i,pred} - y_{a,pred})^2][\sum_{i=1}^N (y_{i,med} - y_{a,med})^2]}}$$

Siendo N el número total de valores medidos, $y_{i,pred}$ el valor predicho en un tiempo determinado i , y $y_{i,med}$ el valor medido en ese mismo momento. y_a es el valor medio de la predicción o medición. En el caso del coeficiente de determinación, sus valores van entre 0% y 100% siendo más preciso cuanto mayor es el valor. Para el MAE y el RMSE, el valor de la predicción será mejor cuanto menor sea el valor del índice.

En el presente proyecto se ha decidido evaluar los resultados con el índice RMSE, ya que nos permitirá comparar con facilidad con otros artículos científicos que hagan uso de la misma red neuronal para predecir la radiación.

3.4. Implementación

En este apartado se va a definir cómo se va a hacer la distribución de los datos para que el estudio del modelo de predicción sea completo. Esta definición incluye los parámetros de la red neuronal, el algoritmo que se empleará para entrenar la red y la clasificación temporal de los datos de entrada.

3.4.1. Parámetros de red

Es necesario determinar previamente los parámetros de los que depende la red neuronal. El primero de ellos es el número de capas ocultas o *hidden layers* que tiene nuestra red. Como se vio en la revisión bibliográfica, la red se compone de una entrada, que en nuestro modelo elegido sería la unidad de entrada de datos secuenciales, una unidad de salida, que se corresponde con la capa de conexión o fully connected layer, y una unidad oculta, que viene determinada por unidad LSTM. Dentro de esta unidad podemos definir un número determinado de capas ocultas. El valor tomado en el presente trabajo es de 200 capas ocultas.

Por otro lado, tenemos las variables de entrada. En nuestro caso la variable de entrada será la radiación solar global histórica con una granularidad de una hora. Antes de cada predicción se representará la radiación del espectro temporal elegido a fin de verificar de forma visual que los valores tomados son correctos, y no hay

desviaciones o errores en los datos de entrada. Como variable de salida tendremos la predicción de la radiación global con una hora de anticipación.

3.4.2. Algoritmo de entrenamiento

Antes de realizar la predicción, la red neuronal debe ser entrenada con un cierto volumen de valores, de tal manera que se determinen los pesos en cada una de las capas de la red. Estos pesos harán que la predicción sea más o menos precisa. Existen varias opciones dentro de los algoritmos, por lo que no es tarea fácil determinar cuál será el mejor. En [31] se puede ver una clasificación de los algoritmos de entrenamiento, así como sus distintas funcionalidades. Al mismo tiempo, se realiza un cálculo del error asociado al entrenamiento en función del experimento que se ha realizado, y se muestra como para la predicción con series temporales, el algoritmo mejor posicionado es el ADAM (*Adaptive moment estimation*), por lo que será el algoritmo empleado para el entrenamiento de la red. Como valores de entrada del algoritmo de entrenamiento, se han tomado los que vienen por defecto en *Matlab*.

El número de iteraciones empleadas en el aprendizaje será de 200, ya que se ha observado que el error se estabiliza y permanece sin fluctuaciones y con un error bajo para este número de iteraciones. Conviene hacer hincapié en que el número de iteraciones no es una decisión baladí. Si se toman un número pequeño de iteraciones, el error será demasiado alto, ya que no se ha llegado a una estabilización en los pesos de la red que revelen una predicción fiable. Por el contrario, si se toma un valor demasiado alto, hará que las capas de la red tomen unos valores que darán el mismo valor de predicción cualquiera que sea la entrada debido a un “sobreentrenamiento”, o lo que es lo mismo, se perderá la flexibilidad que nos proporciona la red neuronal.

Además, el coste computacional también se verá incrementado. Si se evalúa la tabla 2, apoyándose en las figuras 14, 15, y 16, se pone de manifiesto dicha afirmación. Tomando como ejemplo el primer mes del año 2019, se ha implementado la misma red para tres supuestos diferentes. Para 120 iteraciones se observa como el entrenamiento no es suficiente teniendo un error mucho mayor que los otros dos casos. En el caso de 250 iteraciones, el error cometido en la predicción no mejora considerablemente respecto al ejemplo de 200 iteraciones, pero si aumenta el coste computacional, esto es, el tiempo empleado en realizar la predicción de la radiación solar global sobre una superficie horizontal.

Iteraciones	RMSE	Tiempo
120	84.016	45 seg
200	79.9231	1 min y 10 seg
250	78.4676	1 min y 23 seg

Tabla 2. Tabla comparativa de iteraciones vs error cometido y tiempo de cálculo. (Elaboración propia)

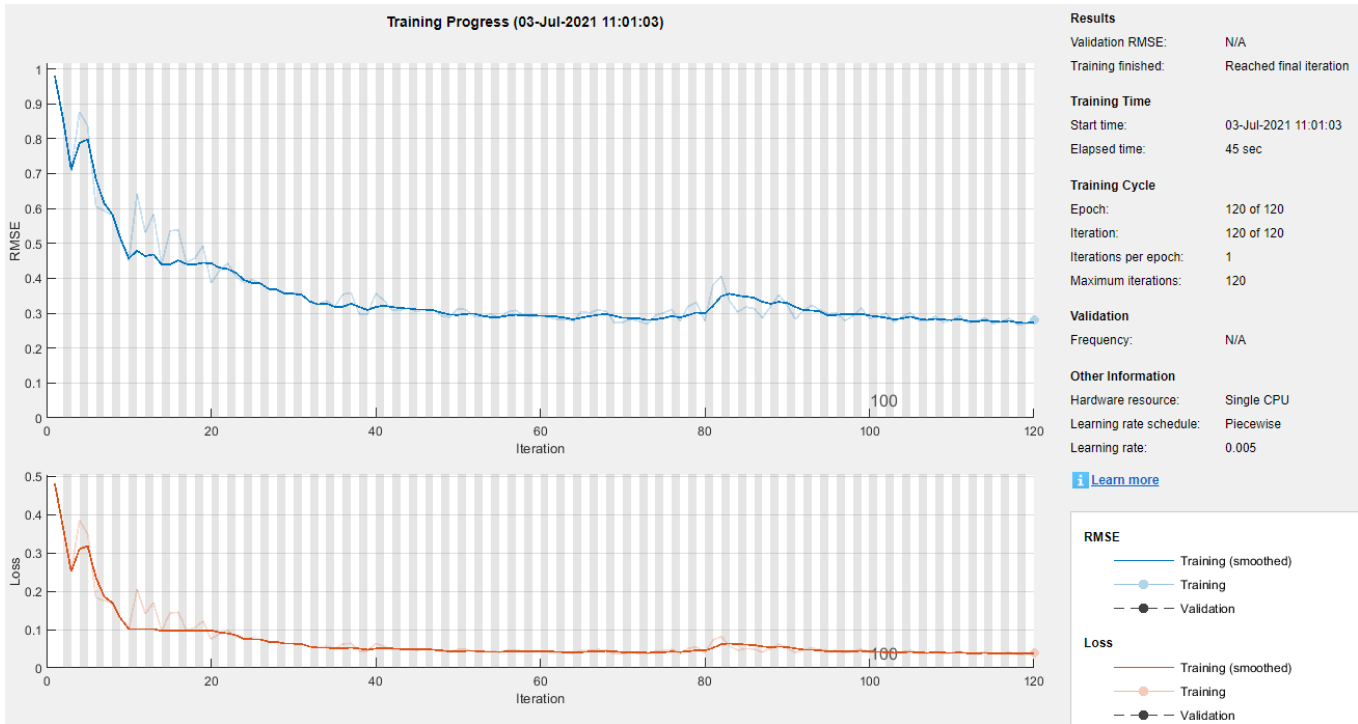


Figura 14. Proceso de entrenamiento para el mes de enero de 2019 y 120 iteraciones.

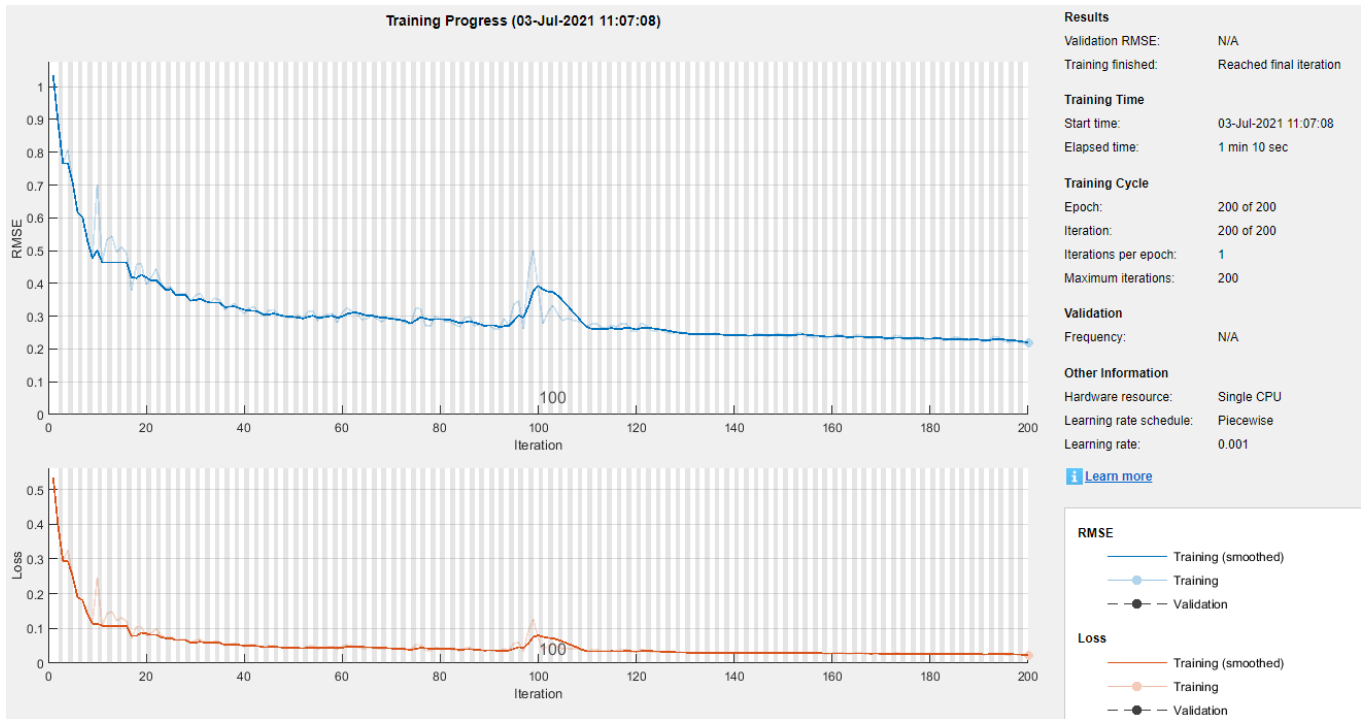


Figura 15. Proceso de entrenamiento para el mes de enero de 2019 y 200 iteraciones.

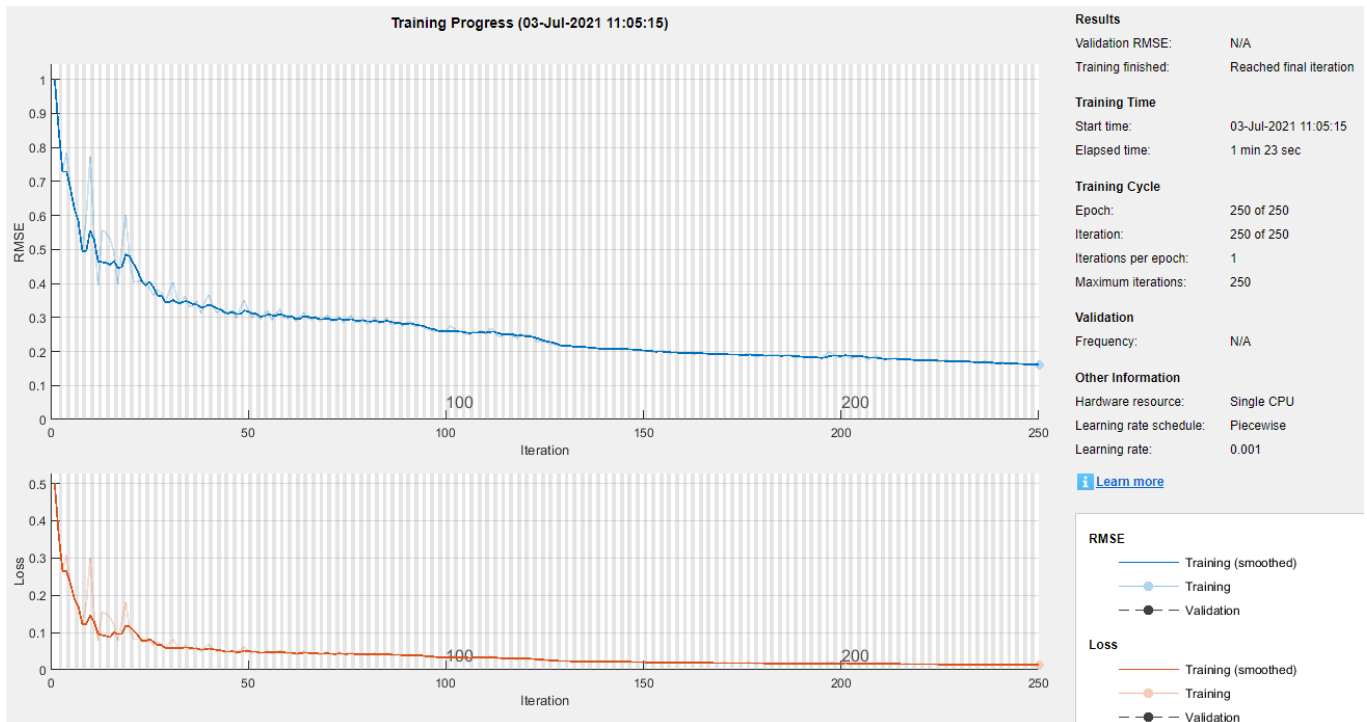


Figura 16. Proceso de entrenamiento para el mes de enero de 2019 y 250 iteraciones.

A la hora de realizar el entrenamiento de la red hay que tener en cuenta una consideración. Ningún entrenamiento es igual a otro realizado con las mismas hipótesis, por lo que el valor del error no será exactamente el mismo si consideramos dos supuestos exactamente iguales. En la tabla 3 se presenta una tabla resumen de los argumentos de entrada de la red neuronal.

Modelo	N.º de neuronas ocultas	N.º de iteraciones	Algoritmo de entrenamiento
LSTM	200	200	ADAM

Tabla 3. Resumen de los argumentos del algoritmo neuronal. (Elaboración propia)

3.4.3. Clasificación de los datos de entrada

Gracias a que se dispone de una gran base de datos, se tratará de ordenar los datos de forma que se exprima al máximo la capacidad de la red neuronal elegida. Por ello, se realizará el modelo para los siguientes intervalos temporales:

- Mensual. Se probará cada mes de cada año, a fin de obtener relaciones entre los todos los meses, ya sean el mismo mes de cada año o meses diferentes, observando claramente la tendencia de la tipología de día de cada mes por separado.
- Trimestral. El volumen trimestral nos permitirá ver qué tendencia estacional tiene la radiación y cuál es la dificultad en la predicción entre unos meses y otros.
- Anual. El volumen anual nos permitirá que la red aprenda de un valor amplio de valores, sin repetición de estacionalidad, de tal manera que se pueda observar si hay mejora o no al introducir en la red distintos parámetros con diferentes tendencias.
- Bianual. Se podrá observar si la introducción de un volumen de datos grande influye de manera positiva o negativa al error asociado a la predicción.

- Triannual. De la misma manera que con el bianual, se podrá determinar el efecto de introducir un volumen de datos aún mayor.
- Variabilidad 30 días. Una vez estudiados los casos anteriores, se verá el efecto que tiene la variabilidad en la predicción, de tal manera que se entrenará la red únicamente con los primeros 30 días con alta variabilidad del año 2019, y se comparará con la predicción realizada con todos los tipos de días.

En todos estos casos, el volumen de datos que se empleará para entrenar la red será del 75% del total, mientras que se empleará el 25% restante para realizar la prueba y calcular el error con el índice de precisión RMSE.

4 RESULTADOS

En este apartado se van a evaluar los resultados de la predicción de la radiación solar global haciendo uso de redes neuronales basadas en la unidad LSTM. La clasificación de los datos de entrada vendrá dispuesta como se especifica en el apartado anterior del presente proyecto.

4.1. Análisis mensual

Los resultados del análisis mensual vienen reflejados en la tabla 4. También se incluye la figura 17 para poder realizar un análisis más visual de los resultados. En ella, se puede observar cómo los meses que la variabilidad es mayor, el error aumenta de forma considerable. Estos meses se corresponden con las estaciones en las que hay un alto índice de días con nubes y claros. Esto hace que la variación de la radiación de un momento a otro sea brusca e impredecible, lo que dificulta el aprendizaje de la red neuronal, y con ello la calidad o exactitud del resultado. Salvo excepciones, los meses en los que ocurre este fenómeno son aquellos pertenecientes al entorno de la primavera y el otoño, esto es, los meses de marzo, abril y mayo y junio, y los meses de septiembre, octubre y noviembre. El resto de los meses el error es menor debido a que la radiación es mayor o menor, pero sigue un patrón diario similar. De la misma manera y salvando excepciones, el error mensual es similar entre un año y otro, debido a que la tendencia climatológica entre los meses de distintos años es siempre más o menos regular. Esta tendencia se puede ver claramente en el mes de enero o el mes de diciembre.

Este fenómeno se puede evaluar si se compara la figura 18, 19, y 20. La primera se corresponde con el mes de enero de 2018, donde se observan días con variabilidad y otros que no. La segunda de ellas se corresponde con el mes de abril del año 2018. En ella se puede ver como la variabilidad es elevada, y por ello el error aumenta. Aun así, la tendencia que sigue la predicción no es muy diferente a la que sigue la radiación real, lo que asegura en cierta manera la utilidad del modelo. Posteriormente veremos si el aumento de volumen de datos mejora estas tendencias. En la figura 20, se puede observar como para días claros el modelo es fiable, ya que la distribución horaria de la radiación es prácticamente igual todos los días. Esta comparativa confirma una obviedad. La dificultad de los modelos de predicción de la radiación solar radica en los días con alta variabilidad.

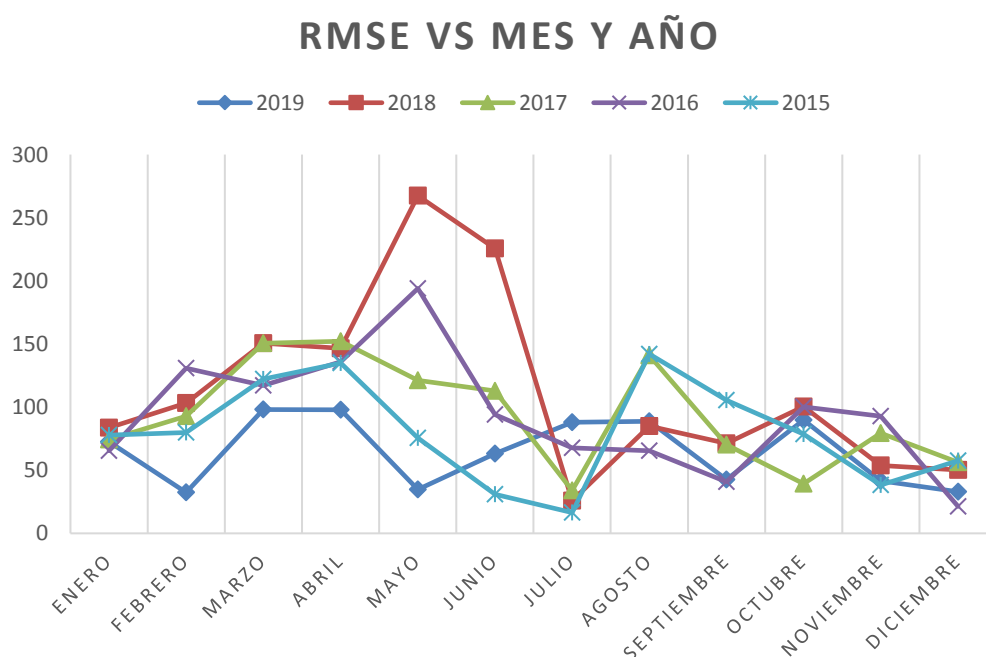


Figura 17. RMSE vs mes y año. (Elaboración propia)

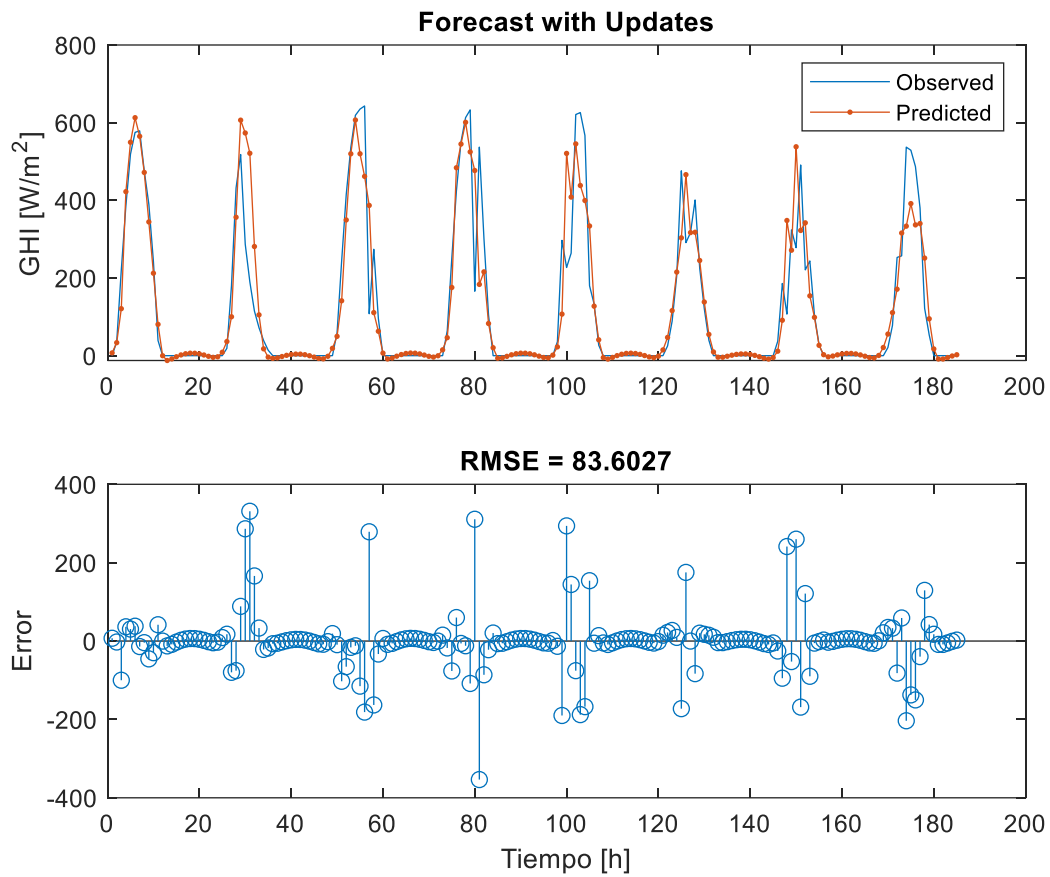


Figura 18. Predicción con volumen de datos de entrada mensual - enero 2018. (Elaboración propia)

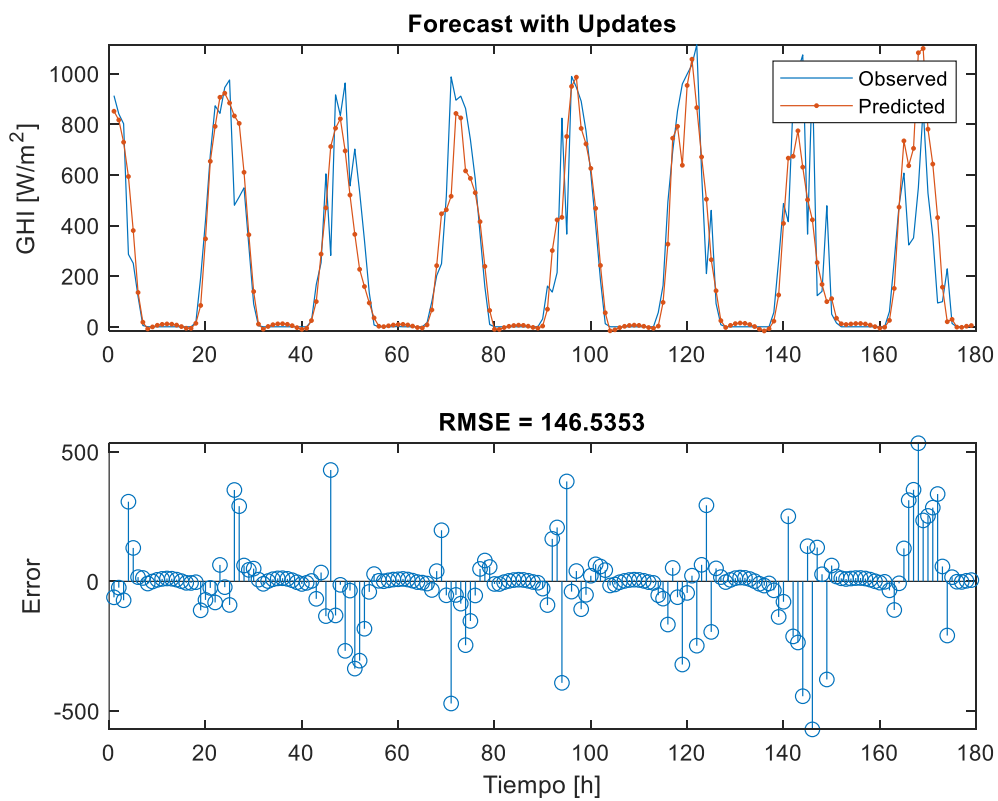


Figura 19. Predicción con volumen de datos de entrada mensual - abril 2018. (Elaboración propia)

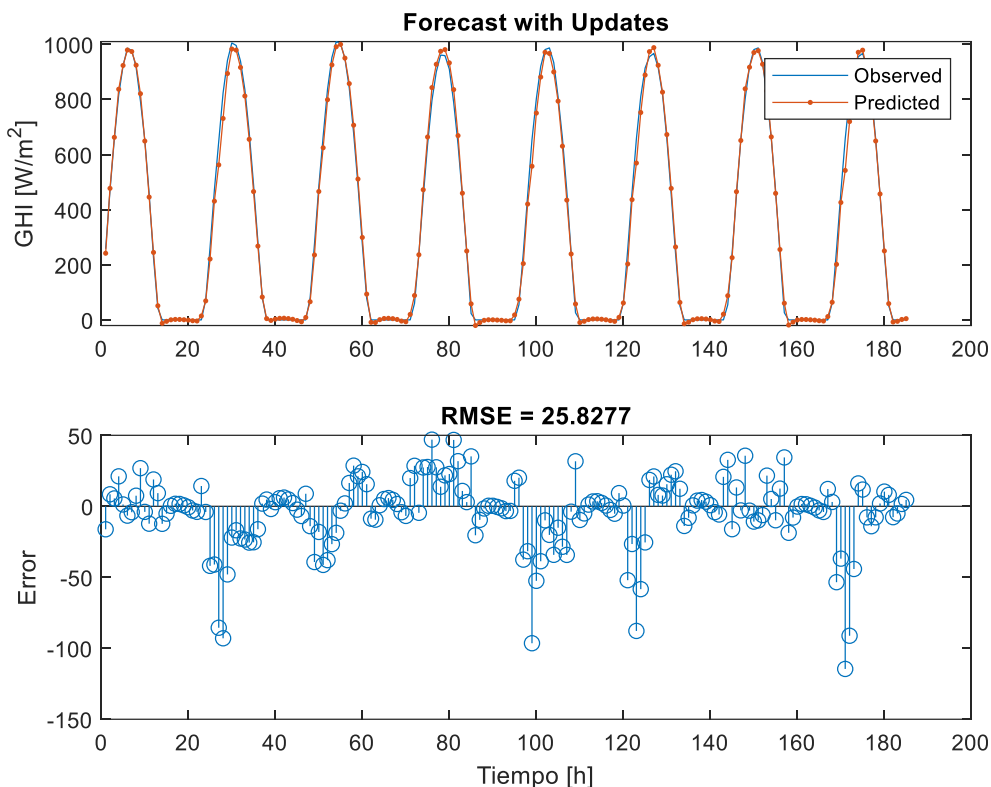


Figura 20. Predicción con volumen de datos de entrada mensual - julio 2018. (Elaboración propia)

Al ser un volumen de datos pequeño (entre 28 y 31 días), el error depende mucho de los días que se estén tomando para realizar la prueba. Un volumen pequeño de datos conlleva que el número de días en los que se determinará el error también sea pequeño. Con ello, si se trata de una semana con buen clima, el error será bajo, mientras que, si por el contrario la semana de prueba tiene un clima inestable, el error será elevado. Este problema se solucionará a medida que se escoja un volumen de datos mayor, ya que el rango de días tomados para hacer la prueba tendrá muchos tipos distintos de días, calculando un error medio entre todos ellos.

	2019	2018	2017	2016	2015
Mes	RMSE	RMSE	RMSE	RMSE	RMSE
Enero	72.6108	83.6	74.2484	65.5022	77.8192
Febrero	32.3768	103.1977	92.6573	130.7726	79.7776
Marzo	98.0387	150.6839	150.6188	117.2375	122.2087
Abril	97.733	146.5353	152.1226	135.8266	134.9718
Mayo	34.6113	267.3974	121.2266	193.9323	75.4739
Junio	62.9942	225.5308	112.7382	93.9619	31.0341
Julio	87.9127	25.8277	33.9037	67.7218	16.4337
Agosto	88.7843	84.9968	140.8601	65.4488	142.0408
Septiembre	42.476	71.3353	70.144	40.9047	105.4339
Octubre	89.7316	100.2933	39.1962	100.1747	78.4891
Noviembre	41.5898	53.8135	79.4234	92.852	38.351
Diciembre	32.9858	50.0654	56.1623	21.3503	57.4593

Tabla 4. Resultados del RMSE mensuales por año. (Elaboración propia)

4.2. Análisis trimestral

El análisis trimestral asegura un volumen de datos suficiente para que la predicción sea más fiable. La clasificación se ha hecho agrupando el año en trimestres, lo que se asemejaría de manera casi exacta a las estaciones del año. En la tabla 5 se exponen los resultados del error derivado de la predicción. Al mismo tiempo, se hace una representación gráfica de los resultados en la figura 21 a fin de facilitar la comprensión de los resultados al igual que se hizo en el apartado anterior.

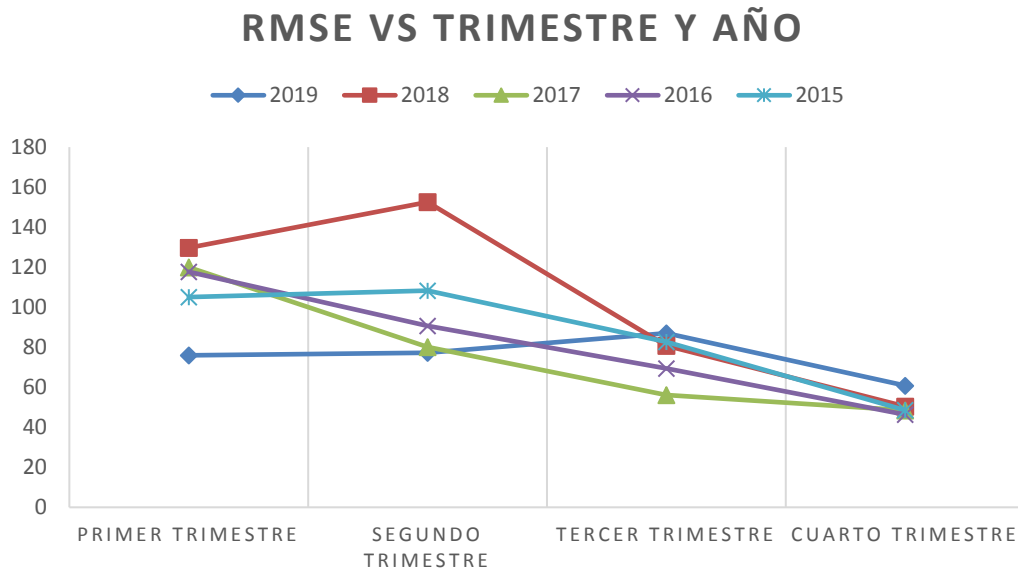


Figura 21. RMSE vs trimestre y año. (Elaboración propia)

Es interesante ver como mejora el error medio si se toma el volumen de datos de entrada trimestral. Esto es debido a que la diferencia de distribución de la radiación diaria es mayor, por lo que la red tiene un mayor conocimiento de la variabilidad de los datos, y también que el volumen de datos de prueba toma un rango más amplio de días, por lo que el cálculo del error cuadrático medio es más real.

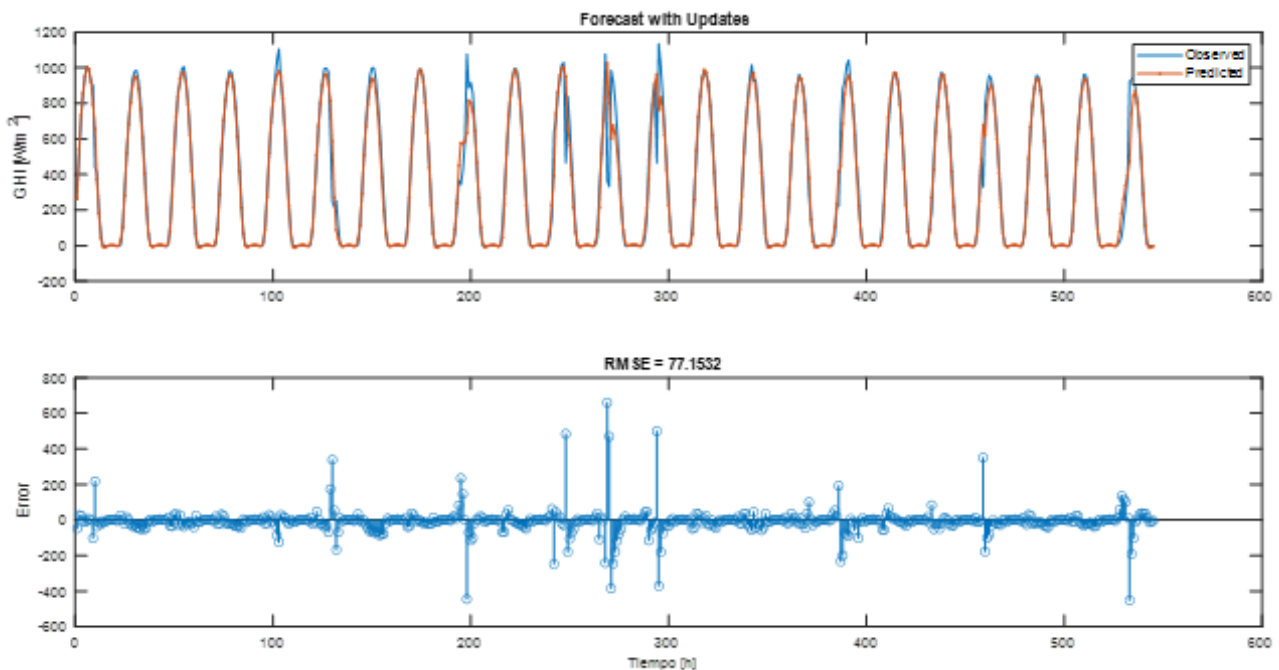


Figura 22. Predicción con volumen de datos de entrada trimestral– 2º trimestre 2019. (Elaboración propia)

Al igual que ocurría para el análisis mensual, salvo excepciones, el error de un año a otro es similar, siendo mayor en los meses con mayor variabilidad.

	2019	2018	2017	2016	2015
Mes	RMSE	RMSE	RMSE	RMSE	RMSE
Enero	76.0192	129.6195	119.9505	117.6891	105.1388
Febrero					
Marzo					
Abril	77.1532	152.5094	80.0685	90.6825	108.2931
Mayo					
Junio					
Julio	87.007	80.7552	56.1951	69.3678	82.6563
Agosto					
Septiembre					
Octubre	60.8299	50.4006	48.4793	46.3163	48.5507
Noviembre					
Diciembre					

Tabla 5. Resultados del RMSE trimestrales por año. (Elaboración propia)

4.3. Análisis anual

El análisis anual cobra más sentido en tanto en cuanto la variabilidad aumenta de forma considerable, dando a la red muchos tipos distintos de días. El coste computacional de esta operación también aumenta considerablemente. Aun así, la reducción de las iteraciones no se considera debido a que la disminución de estas supone un cambio importante en el error medio de la predicción de la radiación. En la tabla 6 se observa esta diferencia de predicción entre 120 y 200 iteraciones. Esta comparación lleva a la misma conclusión que se llegó en apartados anteriores del presente trabajo.

	120 iteraciones	200 iteraciones
RMSE	91.9832	63.2055

Tabla 6. Comparación del error para 120 y 200 iteraciones para el año 2015. (Elaboración propia)

De esta manera, aunque se tardará más en obtener los resultados, los valores serán más precisos. En la tabla 7 se puede observar el valor del error para el período comprendido entre 2015 y 2019. De manera general, se observa como el error disminuye respecto a los casos estudiados anteriormente. Además, este error se estabiliza, es decir, el error medio cometido para cada año es muy similar, salvo para el año 2017. Esto se debe a que, como se ha visto previamente en las figuras 18 y 21, el error cometido en la predicción para los últimos meses de este año es menor respecto a los demás, lo que se traduce en que hay una variabilidad menor, ya que se ha comprobado que son los días donde hay más dificultades de predecir la radiación solar horizontal con precisión. Por lo demás, este resultado se puede traducir en que el modelo funciona bien para un mismo tiempo de prueba sea cual sea la diferencia entre las variabilidades de entrada, ya que el modelo aprende de ellas de una manera parecida llegando con ello a un error medio del mismo orden de magnitud.

2019	2018	2017	2016	2015
RMSE	RMSE	RMSE	RMSE	RMSE
60.5174	66.3606	46.0745	69.6389	63.2055

Tabla 7. Resultados del RMSE anual (Elaboración propia)

En la figura 23 y 24 se expone una comparativa de la predicción de la radiación solar para la última semana del año 2015. A la izquierda se encuentra la predicción con un volumen de entrenamiento anual, mientras que a la derecha la predicción con un volumen de datos mensual. Como se puede observar, el error disminuye con un volumen de datos de entrada mayor, por lo que se puede asegurar que un aumento del volumen de datos de mensual a anual mejora la precisión de la predicción. Aun así, se observa que la predicción para los días con alto índice de variabilidad se mantiene prácticamente igual. Esto es debido a la dificultad de predecir la radiación con tiempo inestable. Dicho esto, la tendencia que toma la predicción para los días propuestos no es del todo insatisfactoria.

4.4. Análisis bianual y trianual

El análisis bianual arroja una duda en cuanto a la precisión de la predicción, ¿es un volumen de datos mayor necesariamente mejor que uno menor? Habiendo analizado los resultados mensuales, trimestrales, y anuales, y viendo la tendencia de descenso del error a medida que aumenta el volumen de datos, se puede llegar a tomar una conclusión precipitada.

Período	RMSE
2015-2016	62.928
2016-2017	52.503
2017-2018	66.7364
2018-2019	66.5083

Tabla 8. Resultados del RMSE bianual y trianual. (Elaboración propia)

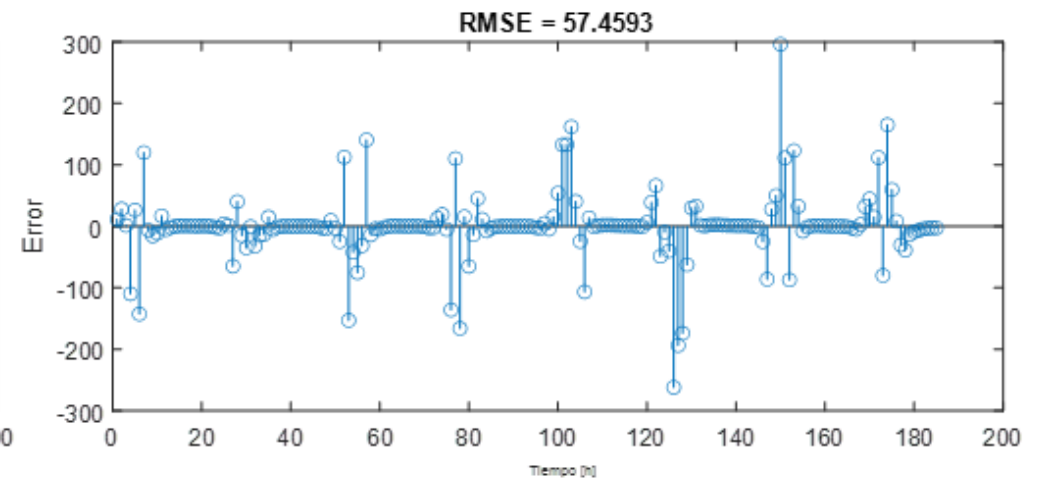
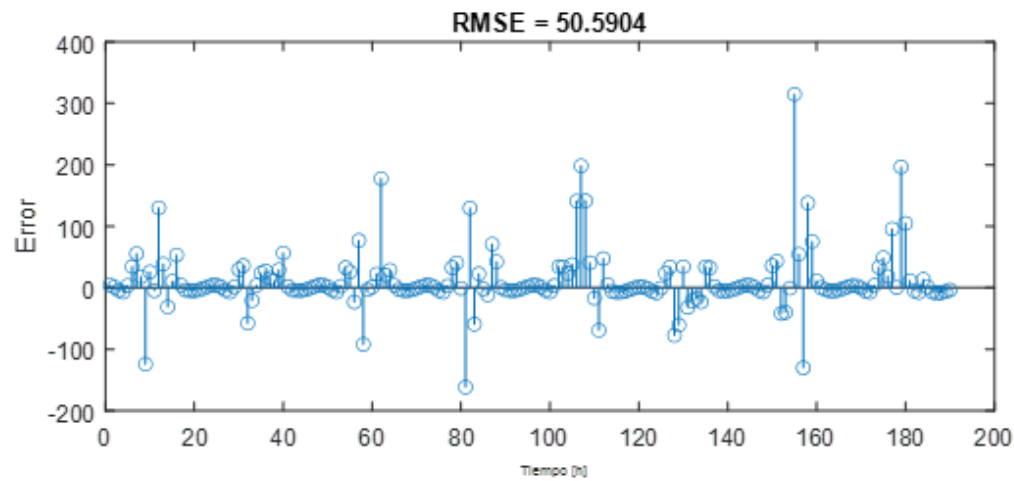
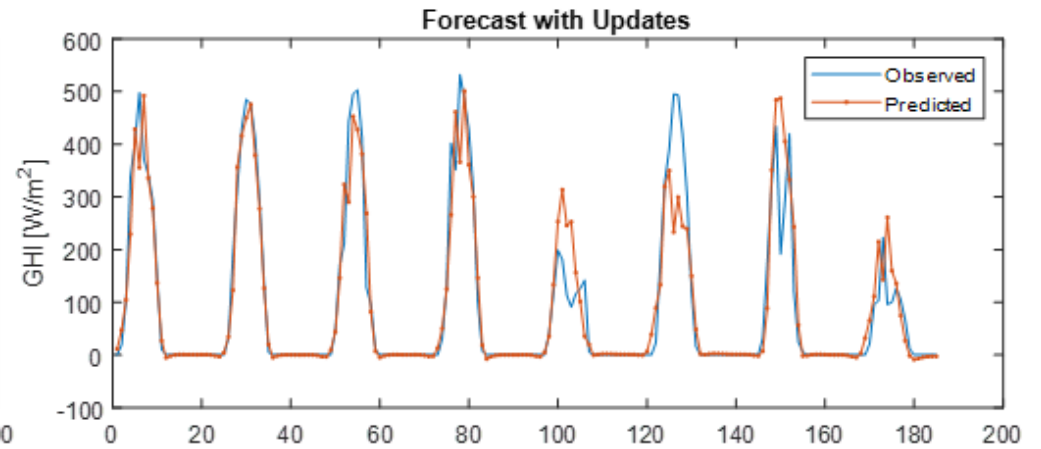
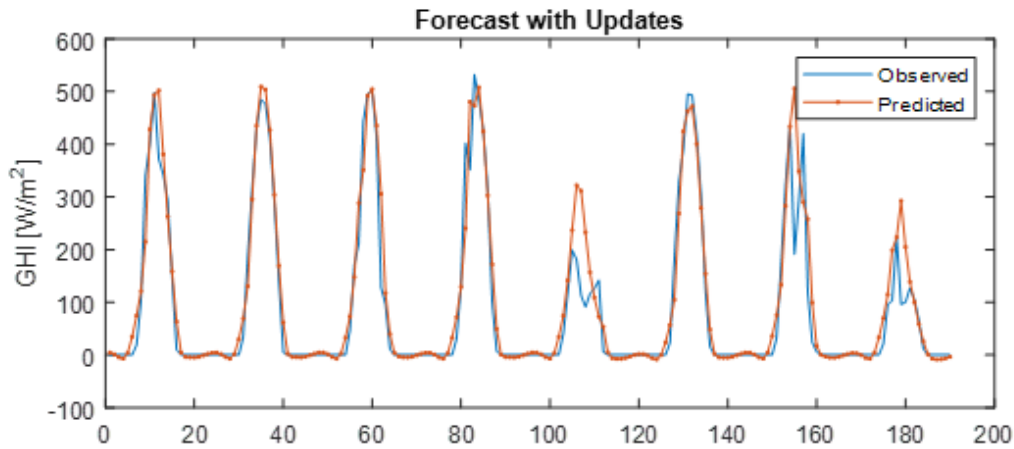


Figura 24. Predicción de la última semana de diciembre 2015 con volumen de datos anual (Elaboración propia)

Figura 23. Predicción de la última semana de diciembre 2015 con volumen de datos mensual (Elaboración propia)

En la tabla 8 se puede observar cómo un volumen de datos superior a un año hace que el error medio de la predicción, para un volumen de datos de prueba del 25% de la muestra, aumente levemente respecto a los errores anuales, pero, por otro lado, como se expone en la tabla 9, el tiempo de ejecución es muy superior si se compara con el tiempo anual (aproximadamente un 50%). Esto hace que se deba llegar a una solución de compromiso entre el tiempo que se pierde en el entrenamiento de la red frente a la mejora que ofrece en la predicción de la radiación.

Período	Tiempo de ejecución
Anual	9-11 min
Bianual	21-23 min

Tabla 9. Tiempos de ejecución en función del período elegido (Elaboración propia)

No ocurre lo mismo para un volumen de datos trianual. Se ha realizado una prueba para el intervalo comprendido entre el 01/01/2017 y el 31/12/2019. En este caso:

RMSE: 75.1621

Esto determina que para las condiciones de trabajo que se han tomado, un volumen superior a dos años no mejora la precisión media de la predicción de la radiación solar. Además, el tiempo de ejecución del entrenamiento supera los 30 minutos, por lo que tampoco es satisfactorio en ese aspecto. No se debe olvidar que en todo momento se ha utilizado un 75% de valores para realizar el entrenamiento y un 25% para realizar la prueba de precisión.

4.5. Análisis para días con alta variabilidad

Como análisis complementario, se ha decidido evaluar cuál sería la precisión del modelo si se entrenase únicamente con días con alto índice de variabilidad, o lo que es lo mismo con un tiempo inestable, en el que la línea temporal de la radiación durante un día completo tiene muchos altibajos. Para ello, se ha escogido los 30 primeros días con variabilidad del año 2019, siendo estos días alternos como cabe esperar, ya que la probabilidad de que en el emplazamiento elegido haya 30 días seguidos con tiempo inestable es muy baja.

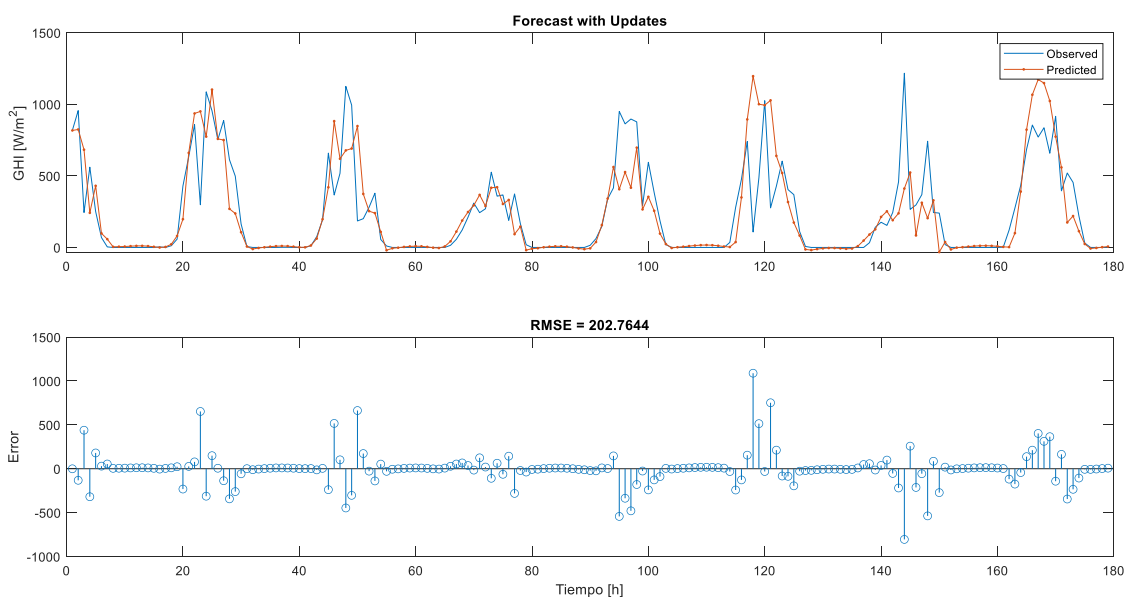


Figura 25. Predicción con entrenamiento con días altamente variables. (Elaboración propia)

En la figura 25 se puede observar cómo, aunque se entrene la red únicamente con días variables, la predicción no mejora. Esto se debe a que los días de este tipo no son de ninguna similares unos a otros, lo que conlleva un alto error en la determinación de los pesos asociados a la red neuronal que determina el valor de la predicción. Aun así, se puede ver que la tendencia que sigue los valores predichos frente a la radiación real sigue en líneas generales una tendencia parecida, lo que arroja un poco de luz frente a la alta dificultad que supone la predicción de la radiación solar global este tipo de días.

4.6. Comparativa de resultados haciendo uso de la unidad GRU

En este apartado se exponen los resultados de la unidad GRU para un volumen de datos de la prueba de un año, ya que es el rango de datos donde se ha conseguido mayor precisión de la predicción. Estos resultados se comparan con los resultados obtenidos con la unidad LSTM como se observa en la tabla 10. La tendencia que siguen los errores cometidos en la predicción, fácilmente evaluable en la figura 26, se asemeja a la comentada en la bibliografía, esto es, el error haciendo uso de la unidad GRU es mayor que el cometido por la unidad LSTM, pero el tiempo de ejecución es más reducido: seis minutos y medio de media para la unidad GRU frente a los diez minutos de media de la unidad LSTM. La cuestión sería evaluar si este error se puede considerar aceptable si primamos el tiempo de computación.

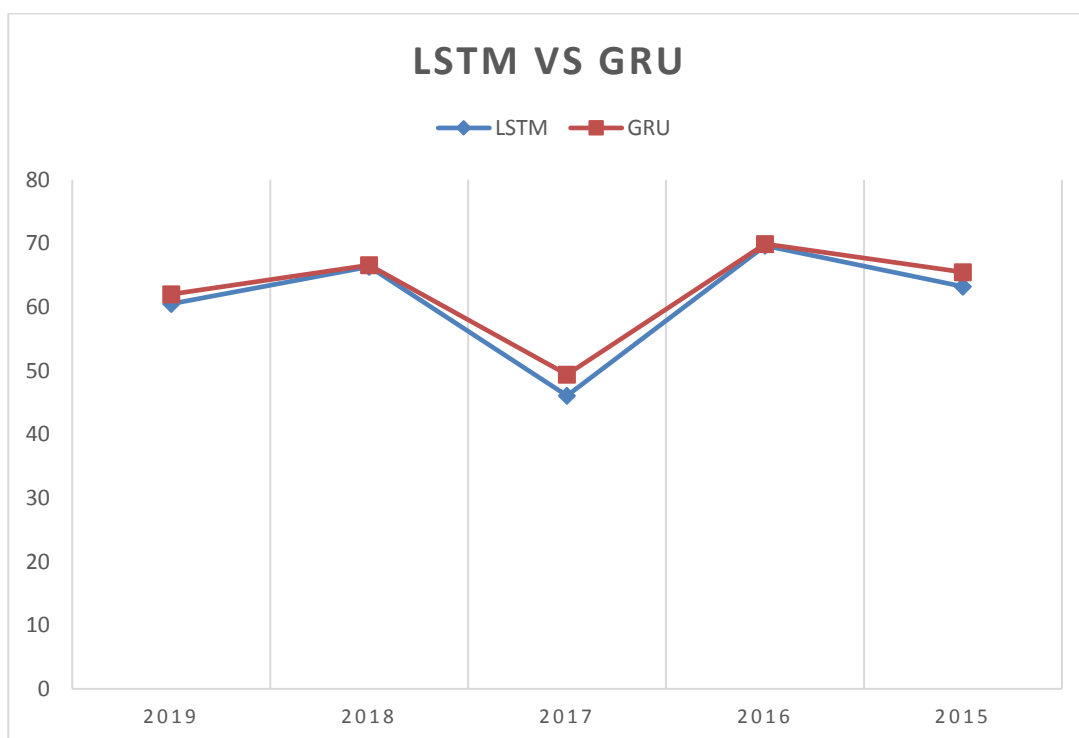


Figura 26. Error RMSE de LSTM vs GRU.

	2019	2018	2017	2016	2015
Unidad	RMSE	RMSE	RMSE	RMSE	RMSE
LSTM	60.5174	66.3606	46.0745	69.6389	63.2055
GRU	62.0107	66.6089	49.3991	69.9226	65.4665

Tabla 10. Comparativa de errores de LSTM frente a GRU. (Elaboración propia)

4.7. Ruido en el problema multivariable

Si se decide introducir un mayor número de variables para la predicción con la distribución actual de la red, el error aumenta respecto del calculado con la variable GHI histórica como única entrada. En la bibliografía se expone que un número mayor de variables hace que el error cometido sea menor, pero en estos casos las variables sufren un tratamiento previo y no se introducen en la red directamente. Este tratamiento previo a la introducción de las variables en la red se consigue con la unión de distintas redes. Como se vio en apartados anteriores, la herramienta elegida no dispone aún de la posibilidad de hacer estas redes compuestas, por lo que es de esperar que, si no tratamos las variables de entrada, estas modificarán las variables históricas introduciendo “ruido” en la predicción de la radiación solar horizontal.

En la tabla 11 se pone el error calculado para el año 2019 con el uso de la temperatura, la velocidad del viento, y la dirección del viento como variables adicionales. Además, se presenta la figura 27 a modo de ejemplo de las desviaciones cometidas en la predicción.

	2019
Variables	RMSE
Univariable	60.5174
Multivariable	75.1188

Tabla 11. Error univariable vs error multivariable. (Elaboración propia)

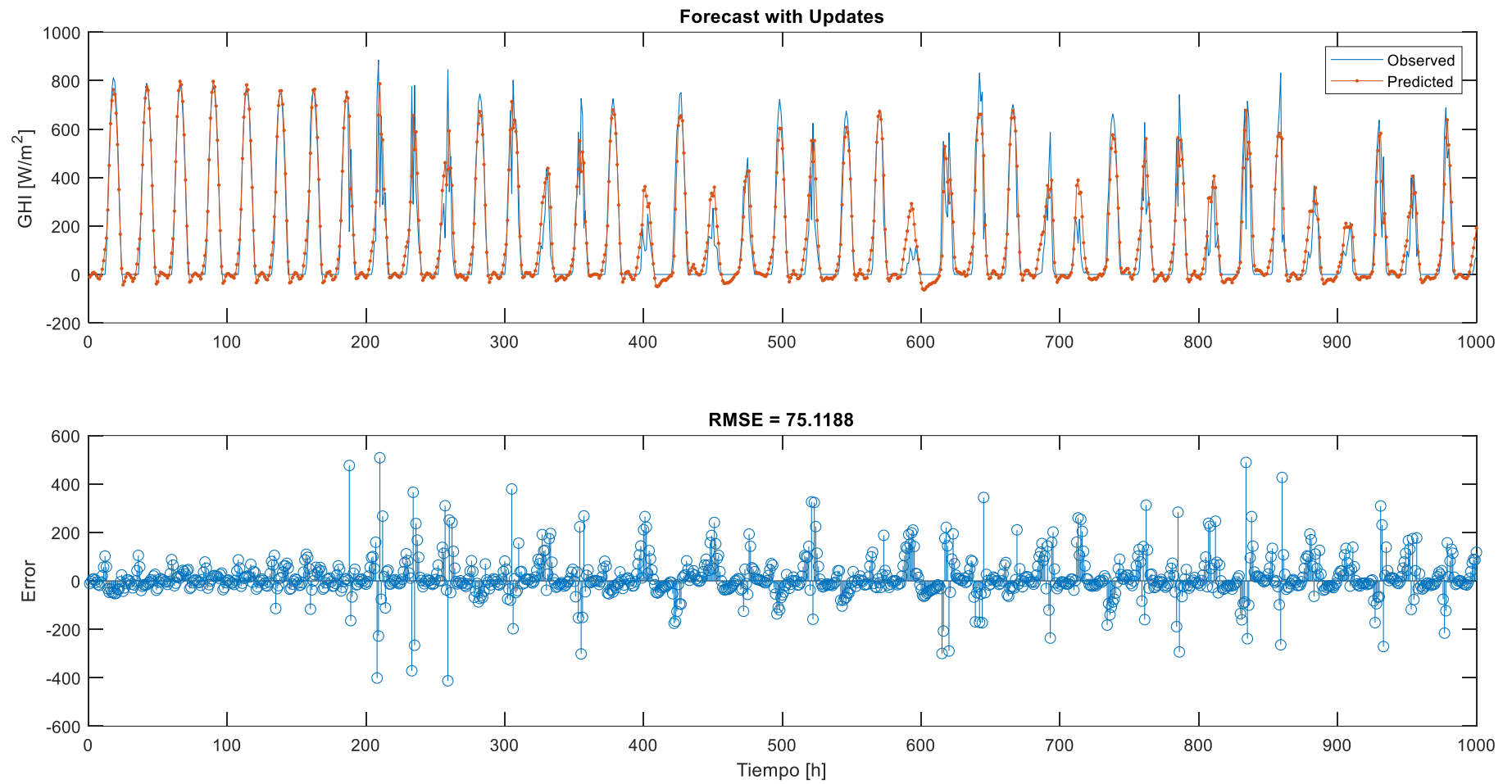


Figura 27. Ejemplo del ruido introducido en la predicción. (Elaboración propia)

5 COMPARATIVA CON ARTÍCULOS CIENTÍFICOS

Una vez evaluado el tipo de red que se ha tomado para realizar la predicción, es necesario contrastar los resultados obtenidos. Para ello, se va a realizar una breve comparativa con artículos contrastados que emplean la misma red neuronal en la predicción de la radiación en climas parecidos al del lugar elegido para llevar a cabo el presente proyecto.

El primero de ellos es el que se tomó como referencia en un primer momento para el desarrollo de la red, llamado “Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations” y llevado a cabo por Zang *et al.* [29]. Este estudio hace uso de una red neuronal compuesta por una red convolucional y red basada en la unidad LSTM. Aunque esta red no se pudo llevar a cabo por los distintos inconvenientes encontrados anteriormente, dispone de una comparativa de resultados donde se han ensayado los lugares elegidos para el estudio con diferentes modelos, entre ellos el modelo simple basado en la unidad LSTM.

Al existir diferentes lugares, no todos los climas serán válidos en la comparación, ya que el tipo de clima hace variar considerablemente el error de la predicción. De esta manera, antes de realizar la comparación, es necesario conocer el tipo de clima exacto que tiene la ciudad de Sevilla. Según [33], el clima en Sevilla se considera templado-cálido. Si se atiende a la clasificación climática según el baremo de Köppen-Geiger, se le asigna un lugar en el espectro Csa. Según la misma referencia consultada para determinar este clima, el clima de la ciudad de San Jacinto pertenece también a este espectro, por lo que se podrán comparar los resultados de ambas localizaciones.

Atendiendo al mismo índice de precisión, el artículo considera un período temporal que va desde el día 1 de enero de 2006 hasta el 31 de diciembre de 2011 para el entrenamiento de la red, mientras que el año 2012 se corresponde con el período de prueba, esto es, un 86% de los datos son para entrenar la red por un 14% para realizar la prueba. En la tabla 12 se expone los resultados obtenidos por el estudio en cuestión.

Período	RMSE H.Zang <i>et al.</i>	Intervalo RMSE Calculado
Anual	91.93	46.0745-69.6389
Invierno	78.56	76.0192-129.6195
Primavera	105.04	77.1532-152.5094
Verano	108.47	82.6563-87.007
Otoño	68.07	50.4006-60.8299

Tabla 12. RMSE según H. Zang *et al.* [12] vs resultados obtenidos (Elaboración propia)

Es cierto que tanto el período de entrenamiento como el período de prueba son superiores a los que se han estudiado en el presente proyecto. Desafortunadamente, no se dispone de una fuente de datos tan extensa que contenga datos corregidos y fiables, así como de una máquina capaz de trabajar con un volumen tan extenso de datos. Por otro lado, el orden de magnitud que tiene el error cuadrático medio [29] es similar al que se obtuvo en el apartado 4, llegando incluso a mejorarlo en algunos casos.

Esta diferencia puede ser debida a varios factores. Por un lado, el número de datos de prueba es superior que cualquier caso estudiado. Esto puede conllevar un aumento del error medio en términos generales debido a que hay un número mayor de días de prueba. También se puede deber a que el clima en Sevilla sea algo menos inestable que el clima en San Jacinto, por lo que los días de prueba haya un número menor de días con alta variabilidad. Por otro lado, al igual que ocurría en nuestra predicción, los errores son mayores aquellos meses que la variabilidad es más elevada.

Otro artículo que hace uso de redes neuronales es el escrito por M. Sorkun et al. [23], llamado “Time Series Forecasting on Solar Irradiation using Deep Learning”. En él se realiza una predicción de la radiación solar haciendo redes neuronales basadas en unidades GRU y LSTM. En el caso de la red LSTM tiene dos casos con suposiciones de parámetros de red diferentes. La base de estudio utilizada en este estudio es de 10 años (desde enero de 1998 hasta diciembre de 2007). La localización es Ajaccio (Francia) con un clima templado-cálido o Csa. En este caso, el índice de precisión empleado es el nRMSE. Ante la dificultad de encontrar lugares donde se hayan realizado estudios que hagan uso de la unidad LSTM en una localización que tenga un clima similar, se ha decidido transformar nuestro índice de precisión para poder hacer la comparativa. La tabla 13 expone los resultados obtenidos:

Período	nRMSE	nRMSE
2019	24.4824%	Caso 1 21.3652% Caso 2 21.1575%
2018	27.8602%	
2017	18.6421%	
2016	28.6084%	
2015	25.6411%	

Tabla 13. Comparación con los resultados con M. Sorkun *et al.* [23] (Elaboración propia)

Como no se especifica el período de prueba en el documento científico, y el resultado entre la prueba anual y bianual son muy similares, se ha decidido compararlo con el período anual. Se puede constatar que el orden de magnitud del error normalizado es muy parecido al error que se comete en la predicción del artículo. Por ello, el resultado obtenido en el apartado 4 queda totalmente contrastado, ya que en dos estudios realizados en dos localizaciones diferentes con un clima similar al de Sevilla el error cometido en la predicción es similar en todos los casos.

6 CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURAS

Del presente documento se pueden obtener diferentes conclusiones. En primer lugar, se puede hacer patente la posibilidad de usar softwares matemáticos como es el caso de *Matlab*, una herramienta de un relativo fácil acceso para la comunidad universitaria, y con un amplio rango de posibilidades a la hora de simular situaciones, que, aunque no sean las mejores soluciones posibles, ayudan a comprender al alumno los conceptos clave de las predicciones haciendo uso de las series temporales o *time series*.

En segundo lugar, se certifica la importancia de las redes neuronales en la predicción de la radiación solar, siendo estas las más usadas actualmente y las que arrojan mejores resultados en comparación con la realidad. Al mismo tiempo, la coexistencia de estas redes neuronales con otros modelos de predicción, desarrollados en mayor o menor medida, pero que ofrecen una alternativa a las redes neuronales y otro punto de mejora para el progreso de la comunidad científica.

Además, se puede afirmar que la unidad LSTM propuesta en este estudio es válida para el lugar de estudio elegido, arrojando resultados satisfactorios en función de los obtenidos por los artículos científicos actuales que comparten el mismo tipo de clima, por lo que es un punto de partida para el uso de las redes neuronales en esta localización. El mejor resultado se obtuvo para un espectro temporal de un año en el que en condiciones normales el error RMSE es igual a 60.7154. El año 2017 el error disminuye debido a las condiciones específicas del clima ese año a 46.0745. De esta manera, se obtiene el mismo orden de magnitud que artículos científicos contrastados. Este error en la predicción de la radiación solar global sobre superficie horizontal es inversamente proporcional al tiempo empleado en entrenar la red neuronal, por lo que el aprendizaje de estas redes se puede considerar satisfactorio con un volumen de datos superior a un año. Se ha observado también que, si el volumen de datos es superior a tres años, el error cometido no tiene una mejora considerable respecto del coste computacional asociado a la red en uso, entendiendo el coste computacional como el tiempo empleado en realizar la predicción incluyendo el entrenamiento previo de la red.

Finalmente, como líneas de trabajo para el futuro, se deja la posibilidad de usar una combinación de redes neuronales haciendo uso de otras herramientas informáticas más sofisticadas a fin de conseguir mejorar los resultados obtenidos en el presente proyecto. En la misma línea, probar con el uso de un mayor número de variables meteorológicas combinadas con usar una granularidad diferente de los datos con el fin de conseguir afinar en mayor medida el resultado de la predicción. Por último, hacer uso de un ordenador de mayor rendimiento a fin de conseguir reducir los tiempos de entrenamiento y el manejo, con mayor facilidad, de un volumen de datos mayor, de manera que se puedan contrastar por otra vía diferente las afirmaciones realizadas en apartados previos de este documento.

7 BIBLIOGRAFÍA

- [1] Alonso-Suárez, R., David, M., Branco, V., Lauret, P., 2020, “*Intra-day solar probabilistic forecasts including local short-term variability and satellite information*”, Elsevier, Renewable Energy, Vol 158, mayo.
- [2] Al-Dahidi, S., Ayadi, O., Adeeb, J., Louzazni, M., 2019, “*Assessment of Artificial Neural Networks Learning Algorithms and Training Datasets for Solar Photovoltaic Power Production Prediction*”, Frontiers in energy research, Vol. 7, Art. 130, noviembre.
- [3] Babatunde, O. M., Munda, J. L., Hamam, Y., 2020, “*Exploring the Potentials of Artificial Neural Network Trained with Differential Evolution for Estimating Global Solar Radiation*”, Energies, mayo.
- [4] Barrera, J. M., Reina, A., Maté, A., Trujillo, J. C., 2020, “*Solar Energy Prediction Model Based on Artificial Neural Networks and Open Data*”, MDPI, Sustainability, agosto.
- [5] Cifuentes, J., Marulanda, G., Bello, A., Resenes, J., 2020, “*Air Temperature Forecasting Using Machine Learning Techniques: A Review*”, Energies, agosto.
- [6] Dev, S., Savoy, F. M., Hui Lee, Y., Winkler S., 2019, “*Estimating solar irradiance using sky imagers*”, Copernicus Publications, octubre.
- [7] El Alani, O., Ghennioui, H., Ghennioui, A., 2019, “*Short term solar irradiance forecasting using artificial neural network for a semi-arid climate in Morocco*”, IEEE.
- [8] Ge, Y., Nan, Y., Bai, L., 2019, “*A Hybrid Prediction Model for Solar Radiation Based on Long Short-Term Memory, Empirical Mode Decomposition, and Solar Profiles for Energy Harvesting Wireless Sensor Networks*”, Energies, diciembre.
- [9] Ghimere, S., Deo, R. C., Raj, N., Mi, J., 2019, “*Deep solar radiation forecasting with convolutional neural network and longshort-term memory network algorithms*”, Elsevier, Applied Energy, Vol. 253, julio.
- [10] Giacosa, G., Alonso-Suárez, R., 2019, “*Performance assessment of intra-day solar irradiation forecast in uruguay using satellite cloud motion vectors*”, ISES Solar World Congress, International Solar Energy Society.
- [11] Guijo-Rubio, D., Durán-Rosal, A. M., Gutiérrez, P. A., Gómez-Orellana, A. M., Casanova-Mateo, C., Sanz-Justo, J., Salcedo-Sanz, S., Hervás-Martínez, C., 2020, “*Evolutionary artificial neural networks for accurate solar radiation prediction*”, Elsevier, Energy, Vol. 210, agosto.
- [12] Jiang, H., Lu, N., Tang, W., Yao, L., 2019, “*A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data*”, Elsevier, Renewable and Sustainable Energy Reviews, Vol. 114, agosto.
- [13] Jiang, J., Lv, Q., Gao, X., “*The Ultra-Short-Term Forecasting of Global Horizontal Irradiance Based on Total Sky Images*”, MDPI, Remote Sensing, noviembre.
- [14] LeCun, Y., Bengio, Y., Hinton, G., 2015, “*Deep Learning*”, Macmillan, Nature, Vol. 521, mayo.

- [15] Lee, J., Wang, W., Harrou, H., Sun, Y., 2020 “*Reliable solar irradiance prediction using ensemble learning-based models: A comparative study*”, Elsevier, Energy Conversion and Management, Vol. 208, febrero.
- [16] Li, Danny H. W., Chen, W., Li, S., Lou, S., 2019, “*Estimation of hourly global solar radiation using Multivariate Adaptive Regression Spline (MARS) e A case study of Hong Kong*”, Elsevier, Energy, Vol. 186, julio.
- [17] Marzouq, M., El Fadili, H., Zenkouar, K., Lakhliai, Z., Amouzg, M., 2020, “*Short term solar irradiance forecasting via a novel evolutionary multimodel framework and performance assessment for sites with no solar irradiance data*”, Elsevier, Renewable Energy, Vol. 157, abril.
- [18] McKay, D. C., Davies, J. A., 1982, “*Estimating solar irradiance and components*”, Pergamon Press, Solar Energy, Vol. 29, N° 1.
- [19] Pang, Z., Niu, F., O’neill, Z., 2020, “*Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons*”, Elsevier, Renewable Energy, Vol. 156, abril.
- [20] Rodríguez-Benítez, F. J., Arbizu-Barrena, C., Huertas-Tato, J., Aler-Mur, R., Galván-León, I., Pozo-Vázquez, D., 2019, “*A short-term solar radiation forecasting system for the Iberian Peninsula Part 1: Models description and performance assessment*”, Elsevier, Solar Energy, Vol.195, noviembre.
- [21] Rodríguez-Benítez, F. J., Arbizu-Barrena, C., Huertas-Tato, J., Aler-Mur, R., Galván-León, I., Pozo-Vázquez, D., 2019, “*A short-term solar radiation forecasting system for the Iberian Peninsula Part 2: Model blending approaches based on machine learning*”, Elsevier, Solar Energy, Vol.195, noviembre.
- [22] Sorkun, M. C., Durmaz İncel, Ö., Paoli, C., 2020, “*Time series forecasting on multivariate solar radiation data using deep learning (LSTM)*”, Turkish Journal of Electrical Engineering & Computer Sciences, Vol. 28, enero.
- [23] Sorkun, M. C., Durmaz İncel, Ö., Paoli, C., 2020, “*Time series forecasting on Solar Irradiation usin Deep Learning*”, Turkish Journal of Electrical Engineering & Computer Sciences, Vol. 28, enero.
- [24] Skeie, K., Gustavsen, A., 2020, “*Predicting solar radiation using a parametric cloud model*”, EDP Sciences, Vol. 172.
- [25] Wang, F., Xuan, Z, Zhen, Z., Li, Y., Li, K., Zhao, L., Shafie-khah, M., Catalão, J. P. S., 2020, “*A minutely solar irradiance forecasting method based on real-time sky image-irradiance mapping model*”, Elsevier, Energy Conversion and Management, Vol. 220, junio.
- [26] Wojtkiewicz, J., Hosseini, M., Gottumukkala, R., Chambers, T. L., 2019, “*Hour-Ahead Solar Irradiance Forecasting Using Multivariate Gated Recurrent Units*”, Energies, octubre.
- [27] Yeom, J., Deo, R. C., Adamowski, J. F., Park, S., Lee, C., 2020, “*Spatial mapping of short-term solar radiation prediction incorporating geostationary satellite images coupled with deepconvolutional LSTM networks for South Korea*”, IOP, agosto.
- [28] Yu, Y., Cao, J., Zhu, J., 2019, “*An LSTM Short-Term Solar Irradiance Forecasting Under Complicated Weather Conditions*”, IEEE, octubre.

[29] Zang, H., Liu, L., Sun, L., Cheng, L., Wei, Z., Sun, G., 2020, “*Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations*”, Elsevier, Renewable Energy, Vol. 160, junio.

[30] Zhu, T., Guo, Y., Wang, C., Ni, C., 2020, “*Inter-Hour Forecast of Solar Radiation Based on the Structural Equation Model and Ensemble Model*”, Energies, septiembre.

Enlaces Web:

[31] Blog de Luis Velasco (fecha de última consulta: 22/06/2021):

<https://velascoluis.medium.com/optimizadores-en-redes-neuronales-profundas-un-enfoque-pr%C3%A1ctico-819b39a3eb5>

[32] BP Energy Outlook (fecha de última consulta: 22/06/2021):

<https://www.bp.com/en/global/corporate/energy-economics/energy-outlook.html>

[33] Climate data (fecha de última consulta: 08/07/2021):

<https://es.climate-data.org/>

[34] Enerdata, 2020 “Anuario estadístico mundial de energía 2020” (fecha de última consulta: 24/06/2021):

<https://datos.enerdata.net/>

[35] International Energy Agency ((fecha de última consulta: 24/06/2021):

<https://www.iea.org/>

[36] Telefónica Tech (fecha de última consulta: 22/06/2021):

<https://empresas.blogthinkbig.com/que-algoritmo-elegir-en-ml-aprendizaje/>

Trabajos académicos:

[37] López, J., Vivas, C., 2018, “*Predicción de radiación solar a corto plazo mediante imágenes de cielo y visión por computador*”, Dpto. Ingeniería de Sistemas y Automática, Escuela Técnica Superior de Ingeniería, Universidad de Sevilla.

[38] De los Bueis, C., 2020, “*Análisis y corrección de la base de datos meteorológicos del GTER. Aplicación al periodo 2016-2019*”, Dpto. de Ingeniería Energética, Escuela Técnica Superior de Ingeniería, Universidad de Sevilla.