

# Predicción de nubes a corto plazo para una planta solar a partir de datos históricos

Rafael Caballero<sup>1</sup>, Luis F. Zarzalejo<sup>2</sup>, Álvaro Otero<sup>1</sup>, Luis Piñuel<sup>1</sup> and Stefan Wilbert<sup>3</sup>

<sup>1</sup> *University Complutense de Madrid, 28040 Madrid, Spain*

{rafacr, alvama06, lpinuel}@ucm.es

<sup>2</sup> *Renewable Energy Division. Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), 28040 Madrid, Spain.*

lf.zarzalejo@ciemat.es

<sup>3</sup> *Institute of Solar Research, German Aerospace Center (DLR), 04200 Tabernas, Spain.*

stefan.wilbert@dlr.de

## Abstract <sup>1</sup>

Es este trabajo se aborda el problema de la predicción de radiación global sobre superficie horizontal con alta resolución espacial y temporal (5 minutos) a partir de los datos registrados durante un año en la red radiométrica de alta resolución ubicada en la Plataforma Solar de Almería. En particular se muestra un método capaz de predecir el valor de radiación en los siguientes minutos a partir de los valores de los minutos anteriores. El método emplea el tipo de red neuronal recurrente conocido como LSTM, capaz de aprender patrones complejos y predecir el próximo elemento de una serie temporal. Los resultados muestran una mejora apreciable en la precisión del método con respecto a la predicción basada en el último valor conocido.

**Keywords:** Solar radiation nowcast, LSTM, supervised machine learning.

## 1. Introducción

Desde hace ya décadas, la generación de electricidad a partir de energías renovables se ha convertido en una necesidad cada vez más acuciante por la problemática medioambiental asociada a residuos y emisiones que los sistemas de producción convencional ocasionan. Esta producción resulta cada vez más competitiva respecto a los métodos de generación tradicionales y, dentro de del conjunto de energías limpias, la energía solar ha adquirido un papel más relevante con el incremento del número de plantas solares en operación.

El constante aumento del número de sistemas de producción eléctrica a partir de energía solar y su integración a las redes de distribución energética tradicionales conlleva la necesidad de prever la energía que se va aportar a la red. Por otro lado, los gestores de las plantas solares también requieren de esa previsión para participar en el mercado de la energía y planificar las operaciones de mantenimiento. En definitiva, la capacidad de predecir el recurso solar disponible es crítica para los operadores de plantas solares, ya que esto puede afectar la gestión de la planta y, en consecuencia, la generación de electricidad resultante. Es este trabajo se aborda el problema de la predicción de radiación global sobre superficie horizontal con alta resolución espacial y temporal (pocos minutos) a partir de los datos recogidos durante un año en la red radiométrica de alta resolución ubicada en la Plataforma Solar de Almería (www.psa.es). El objetivo es encontrar un método

que:

- Disminuya el error de métodos tradicionales, tales como proponer como predictor el último valor conocido o sobre métodos basados en técnicas estadísticas como ARIMA.
- Dado que el objetivo a medio plazo es integrar el modelo de predicción en una planta solar real, el método propuesto debe ser capaz de definir un modelo y predecir el siguiente valor en un tiempo breve. Aunque se quisieran predecir valores a varios minutos vista, sería deseable que la predicción se realizara en menos de un minuto, ya que en otro caso estamos desaprovechando un resultado nuevo. Buscamos por tanto un modelo de predicción ligero, capaz de mejorar en particular la predicción constante, que, como veremos resulta una muy buena aproximación a muy corto plazo.

Los modelos de predicción de radiación se suelen agrupar en tres tipos. El primer tipo lo forman los *modelos físicos* basados en ecuaciones matemáticas

que describen la física de la atmósfera [1], por ejemplo utilizando los vectores de movimiento de nubes obtenidos a partir de los vectores de movimiento de nubes [2-3]. En una línea diferente, los *modelos estadísticos* [4-5], establecen relaciones entre las observaciones pasadas y las futuras predicciones. La tercera línea dentro de la que se encuentra nuestra propuesta, también emplea los datos históricos pero a través de modelos de aprendizaje automático tales como las redes neuronales [6-8].

La principal diferencia de nuestro trabajo con los ya mencionados es que éstos se centran en predicciones normalmente de partir de 30 minutos [8], nosotros buscamos predicciones a muy corto plazo, de entre 1 y 10 minutos (nowcasting). Es importante notar que cualquier mejora en la predicción, por pequeña que sea, puede tener un impacto económico, al permitir estimar con más precisión la radiación solar recibida y, por tanto, la energía que se va a aportar a la red.

En la siguiente sección comenzamos presentando el conjunto de datos al que hemos aplicado el método y el preprocesado de estos datos. En la sección 3 buscaremos una posible segmentación que permita establecer una distinción entre periodos con y sin nubes. La sección 4 indica los métodos que hemos elegido para comparar con nuestra propuesta, y el tipo de red neuronal elegida. Los resultados de los experimentos se explican en la sección 5. Para terminar, la sección 6 presenta las conclusiones finales.

## 2. Conjunto de datos inicial

### 2.1. Estaciones radiométricas

Las 7 estaciones radiométricas utilizadas en este trabajo, forman parte de la red de estaciones ubicadas en la Plataforma Solar de Almería (PSA-CIEMAT). En 2014 dentro del contexto del Proyecto DNICast (predicción de la radiación solar directa con alta resolución espacio-temporal, <http://www.dnicast-project.net>) se propuso la creación de una red de estaciones que incorporase un total de 19 estaciones radiométricas de alta calidad que cubrieran una superficie de aproximadamente 0.5 km<sup>2</sup> (Fig. 1).



Fig. 1 Localización de las estaciones radiométricas en la PSA

En particular, partimos de los datos de radiación global sobre superficie horizontal registrados por 7 de estas estaciones equipadas con piranómetros termoelectrónicos Kipp & Zonnen. La localización de cada una de estas estaciones se recoge en la Tabla 1.

Tabla 1: Datos de las estaciones

Nombre	Frec.	Lat.	Long.	Alt.
BSRN	60 s	37.092	-2.363	490.6
ARFRISOL	60 s	37.094	-2.357	499.6
DISS	5 s	37.098	-2.359	504.4
TSA	1 s	37.093	-2.357	499.1
KONTAS	1 s	37.095	-2.355	505.8
CESA1	60 s	37.095	-2.361	503.4
PSA-HP	10 s	37.091	-2.358	500.0

La configuración de la red permite que diariamente se generen los ficheros con los datos de cada estación (las variables registradas son generalmente: irradiancia global horizontal, difusa, directa normal, temperatura, humedad relativa, velocidad y dirección de viento), y éstos son volcados al servidor central vía FTP de forma automatizada.

En este estudio emplearemos los datos de radiación global horizontal correspondientes año 2015 completo. Los datos se han preparado para su

análisis en una secuencia de 4 pasos.

## 2.2. Ajustes y truncado temporal

Dado que las estaciones tienen distintas frecuencias de registro, tal y como indica la Tabla 1, optamos por considerar solo la mayor entre ellas, que corresponde a una frecuencia de un valor por minuto.

A continuación, procedimos a sincronizar los relojes de cada estación. En algún caso se producían desincronizaciones, que pudieron ser corregidas, excepto en el caso de la estación DISS que tuvo que ser descartada para el resto del estudio.

La siguiente tarea consistió en delimitar una franja horaria de trabajo común para todos los meses, minimizando así el efecto de la altura solar y la variación del número de horas de sol durante el año. Se ha optado por filtrar los datos en el horario entre las 9 y las 16 horas, incluyendo las 9 horas pero excluyendo a partir de las 16h. Esto reduce nuestro conjunto de datos a 7 horas x 60 minutos vectores de 6 valores al día, totalizando 153 300 vectores en el año.

## 2.3. Datos ausentes

El segundo paso ha consistido en tratar los datos ausentes. Aunque escasos (<0.01%), la ausencia de datos puede generar un mal funcionamiento o incluso impedir la obtención de resultados en algunos métodos de predicción. Las alternativas usuales para corregir esta situación consisten en o bien borrar las filas defectuosas, o bien completarlas a partir del resto de los datos. En nuestro caso hemos optado por la segunda opción, haciendo que el valor perdido sea reemplazado por la media del resto de los sensores en ese periodo.

## 2.4. Modelo de Cielo Despejado

El tercer paso ha consistido en aplicar un *modelo de cielo despejado* que permita normalizar la variable radiación solar corrigiendo los efectos del movimiento solar aparente. En nuestro caso hemos seguido el método propuesto en [10]. La Figura 2 muestra un día concreto, inestable en las primeras horas y despejado a partir de las 12:30h. En el eje vertical se representa el índice de cielo claro (cociente entre la radiación global y la radiación de cielo despejado) frente a la hora del día. Para facilitar la visualización se han modificado ligeramente las coordenadas Y de cada estación, que de otra manera quedarían superpuestas.

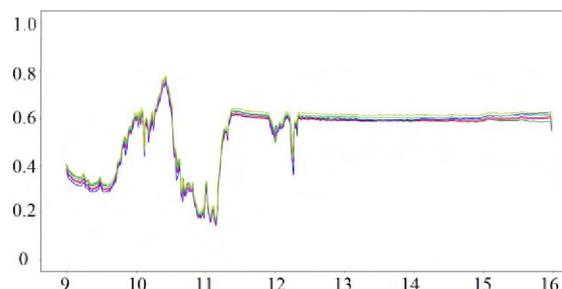


Fig. 2 Gráfica de radiación del día 24 de enero de 2015

## 2.5. Valores erróneos

Tras aplicar el modelo de cielo despejado, todavía se aplicó una cuarta y última fase de detección de errores ocasionales en las estaciones y su corrección, de nuevo sustituyendo el valor defectuoso por la media de los otros valores.

## 3. Búsqueda de cielo nublado mediante segmentación

Un primer análisis de interés consiste en comprobar si los datos admiten alguna segmentación, que permita agrupar mediciones similares y de esta forma determinar un número pequeño de posibles escenarios a analizar.

### 3.1. Número idóneo de clústers

Uno de los métodos más comunes para este propósito es el método propuesto por T. Caliński y J. Harabasz [11]. En este método, los valores a segmentar se consideran puntos dentro de un espacio n-dimensional euclídeo, para los que hay que encontrar una partición óptima. En nuestro caso tendríamos puntos de un espacio de 6 dimensiones, donde cada dimensión corresponderá al valor de una de las estaciones.

La definición de partición óptima en este método se corresponde a aquella partición que permitan minimizar la suma de cuadrados dentro de cada segmento

Los detalles de este algoritmo están más allá del ámbito de este artículo. En todo caso se encuentra implementado en forma de librerías de los lenguajes Python (librería *sklearn*), y R (librería *vegan*). En ambos casos nos da como número óptimo el valor  $k=2$ . Por ejemplo, en R si suponemos que los valores de las estaciones se encuentran en el *dataframe* *d*:

```
fit <- cascadeKM(scale(d, center = TRUE, scale = TRUE), 1, 10, iter = 1000)
```

```
calinski.best <-
as.numeric(which.max(fit$results[2,]))
cat("K óptimo:", calinski.best, "\n")
```

### 3.2. Proceso de segmentación, e interpretación de los resultados

Los centros de cada segmento no se obtienen directamente con este método, pero una vez obtenido que el K óptimo es 2, podemos emplear algún método de clasificación como k-means. En R podemos escribir simplemente `kmeans(d,2)`, y obtenemos los resultados de la Tabla 2.

Tabla 2: centro de los clúster para cada estación

	Cluster 1	Cluster 2
ARFRISOL	0.259	0.682
CESA	0.264	0.679
PSA	0.269	0.685
TSA	0.268	0.690
KONTAS	0.277	0.703
BSRN	0.286	0.703

La primera columna indica el nombre de cada una de las estaciones, mientras que las otras dos indican el valor central para cada uno de los dos clústers.

La interpretación más natural es que el primer grupo corresponde a una medición de nubes, mientras que el segundo se correspondería a un estándar de cielo despejado. Un vistazo al histograma de frecuencias para cada una de las estaciones parece confirmar esta interpretación. En particular, la Figura 3 contiene el histograma de la estación ARFRISOL.

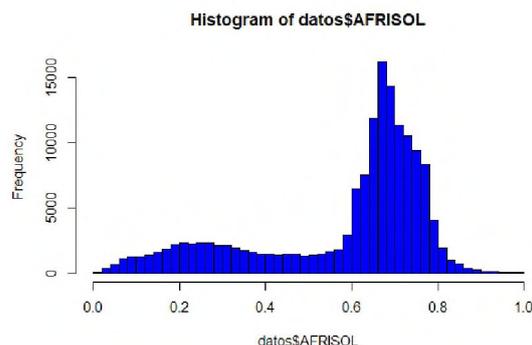


Fig. 3 Histograma de frecuencias para las distintas radiaciones registradas en la estación ARFRISOL

En efecto, la figura parece mostrar una combinación de dos distribuciones. Los centros del clúster 1 se corresponden con el centro de la distribución situada a la izquierda, y por tanto de menor radiación. Esto indicaría que se trata de los valores que corresponden a un cielo con nubes. En cambio, la segunda distribución, más apuntada, corresponde al cielo despejado, que como se aprecia en el histograma es más frecuente en las estaciones estudiadas.

## 4. Predicción de radiación solar global

El método de segmentación descrito en la sección anterior puede emplearse para detectar de forma automática la presencia de nubes, y constituye nuestra primera aportación. En el resto del trabajo presentamos nuestra propuesta para la predicción de radiación solar global horizontal a corto plazo.

### 4.1. Estacionalidad

Antes de comenzar el análisis conviene comprobar si nuestro conjunto de datos sigue es una serie temporal estacionaria. Se llama de esta manera a aquella serie temporal cuyas propiedades no dependen del momento en el que se observan [12]. Este tipo de series son más difíciles de predecir, y si se da el caso, conviene transformar la serie intentando evitar la estacionalidad.

Para comprobar si nos encontramos ante una serie hemos utilizado el método aumentado de Dickey–Fuller (ADF) [13]. En nuestro caso, el método que indica que, en efecto nos encontramos ante un conjunto de datos estacionario, tanto considerando el conjunto de datos anual como las series de datos de cada día. Para evitar la estacionalidad, se suele recurrir a la utilización de las diferencias [14], es decir se reemplaza cada valor por el incremento, positivo o negativo, con respecto al dato anterior. Con esta transformación se pierde un dato por día (el primero de la mañana, donde la

diferencia no tiene sentido), pero en nuestro caso esto carece de importancia dado el elevado volumen de datos del que disponemos.

Tras esta transformación se puede comprobar que ahora se obtienen datos estacionarios, de forma más clara en los datos diarios, y por muy escaso margen en el total anual. Esto nos lleva a decidir que emplearemos datos diarios para nuestros modelos.

#### 4.2. Métodos directos

Vamos a utilizar dos métodos sencillos de predicción referenciados habitualmente en la bibliografía.

- Predicción naïve: La predicción es simplemente el último valor conocido.
- Media: Se propone como predicción la media del último intervalo conocido.

A menudo, sobre todo en datos con alta estacionalidad, estos métodos sencillos son a menudo los más efectivos a corto plazo. En nuestro caso, donde consideramos horizontes de muy pocos minutos, veremos que el método naïve resulta muy difícil de superar, tal y como veremos en la sección de experimentos.

#### 4.3. ARIMA

Los modelos Auto Regressive Integrated Moving Average (ARIMA) [13], utilizan una combinación de  $p$  valores antiguos de la variable para predecir el siguiente valor a calcular, combinados con los  $q$  errores observados en las anteriores predicciones. A estos dos parámetros  $p$  y  $q$  se les añade un tercer parámetro  $d$ , que indica el número de diferencias de valores que deben realizarse. Este número de diferencias es el que permite convertir el modelo en no estacional, y que mencionamos en la sección 4.1. Para el cálculo de los valores  $(p,d,q)$  hemos utilizado la biblioteca de R forecast, y en concreto el método `auto.arima`, obteniendo unos valores de  $p=5$ ,  $d=1$  y  $q=0$ .

#### 4.4. Redes neuronales LSTM

Una red neuronal artificial (ANN por sus siglas en inglés) [15], es un mecanismo de cómputo compuesto por *neuronas artificiales*. Cada neurona recibe una señal de entrada, que procesa y cuya salida reenvía a otras neuronas conectadas a ella.

El número de neuronas de una red, el grafo de conexiones y la función de procesado de cada neurona se deciden durante el diseño de la red. Sin embargo, tanto las funciones como las conexiones entre las neuronas tienen pesos, que se ajustan durante el proceso de aprendizaje.

La red aprende con respecto a un conjunto de entrenamiento, en nuestro caso los datos recogidos de las estaciones durante el día en curso, ajustando los pesos para minimizar una determinada función de coste o de pérdida que relaciona la salida de la red con la salida esperada, en nuestro caso la predicción para un cierto horizonte futuro  $f$ .

En nuestro caso, hemos programado la red mediante la librería Python Keras (<https://keras.io/>) funcionando sobre la librería TensorFlow (<https://github.com/tensorflow/tensorflow>). Keras admite numerosas funciones de coste (parámetro *loss*), entre las que hemos elegido la función `'mean_squared_error'`, dado que emplearemos la raíz del error cuadrático medio (RMSE por sus siglas en inglés) para determinar el error de cada método.

En las redes neuronales, las neuronas se agrupan en capas. Cada capa realiza algún tipo de transformación sobre su entrada, y a su vez pasa su salida a la siguiente capa. Por tanto, las señales viajan desde la capa de entrada hasta la capa de salida, posiblemente tras transitar por cierto número de capas ocultas.

En nuestro caso, partimos de una capa de entrada que recibe una única señal, y tenemos una única capa oculta, una capa de tipo LSTM (Long Short-Term Memory networks [16], el tipo de red neuronal recurrente (esto es, con ciclos) utilizada a menudo para la predicción de series temporales [17].

Hemos encontrado que el número de neuronas adecuado depende del horizonte considerado. En particular, en los experimentos, hemos empleado una red LSTM de 4 neuronas.

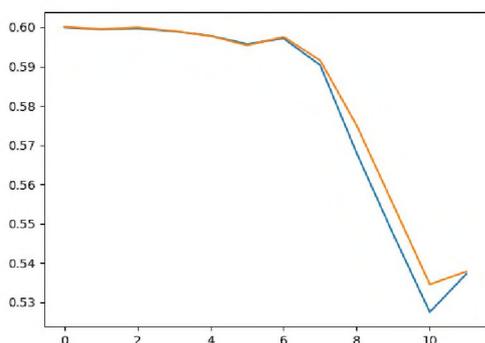


Fig. 4 Predicción por LSTM para los 11 minutos anteriores a las 12h del día 24 de enero de 2015

La Figura 4 muestra un ejemplo de predicción mediante LSTM (línea naranja) comparado con los valores reales (línea azul) para un horizonte de 11 minutos, a partir de los datos recogidos entre las 9 y las 11:49 del día 24 de enero de 2015.

## 5. Resultados experimentales

La Tabla 3 muestra el resultado de nuestros experimentos.

Tabla 3: RMSE de la predicción LSTM modificada, junto con el % de incremento del RMSE del resto de métodos con respecto a LSTM modificada

$f$ (min)	LSTM (RMSE)	AVG (% inc)	Naïve (% inc)	ARIMA (% inc)
1	0.054	6.59	7.08	7.10
2	0.070	7.46	2.03	1.17
3	0.093	2.31	1.14	1.27
4	0.11	3.72	1.05	1.15
5	0.12	3.02	1.03	1.16
10	0.15	2.47	0.96	1.17
20	0.21	1.37	0.93	1.02

La primera columna indica distintos horizontes de predicción en minutos. La segunda columna (RMSE) muestra la raíz del error cuadrático medio para la red neuronal basada en LSTM. El resto de las columnas muestra el % de diferencia entre el RMSE para el resto de los métodos y el obtenido para la red LSTM. En particular, la tercera columna muestra el incremento de error muestra que el método de la media (AVG) presenta un error de entre el 7.46% (predicción a 2 minutos) y 2.47% (predicción a 10 minutos) el error con LSTM. La cuarta columna corresponde al incremento de error para el método Naïve. Este método, el más sencillo, es el de menor error a partir del minuto 10. De todas formas, en los primeros minutos se ve superado por nuestra propuesta, la red LSTM, aunque por valores de tan solo el 3% ya en el minuto 5. Por otra parte, la predicción basada en ARIMA se sitúa en alrededor del 15% de incremento de error.

Los datos se han obtenido como la media de 3 experimentos por cada día:

- Predicción de radiación en el minuto 12h+f a partir de los datos obtenidos entre las 9 a 12 (f en minutos y con los valores indicados en la tabla 3).
- Predicción de radiación a las 13h+f a partir de los datos obtenidos entre las 12 y las 13 horas.
- Predicción de la radiación a las 15h+f a partir de los datos obtenidos entre 13 y las 15 horas.

Hemos comprobado que los datos no varían sensiblemente al considerar como conjunto de entrenamiento valores superiores a la hora. Es decir, no hay variación significativa entre considerar predicciones a partir de datos acumulados de la hora anterior y considerando, por ejemplo, las 3 horas anteriores. Esto es positivo porque indica que se pueden comenzar a predecir valores a partir de las 10h de cada día. En cambio, con datos de menos de una hora sí se aprecian aumentos significativos del error.

En resumen, la tabla muestra la media de un total de 365 (días) x 3 (predicciones por día) x 6 (estaciones) = 6570 tests.

Para comprobar si los datos de la tabla suponen diferencias significativas hemos llevado a cabo la prueba de los rangos con signo de Wilcoxon [18], comprobando que, en efecto, las diferencias de la media de RMSE entre LSTM y cada una de las

otras técnica muestra diferencias estadísticamente significativas.

Por último, y en relación con el objetivo de lograr un método predictivo eficiente en términos de tiempo, debemos señalar que los modelos se generan en media en un tiempo de 23 segundos en un ordenador XPS 13 9350, con 4 CPUs a 2.20 gigahercios y 16 gigabytes de RAM, por tanto cumpliendo nuestro objetivo de ser capaz de general los modelo en un tiempo menor al minuto.

## 6. Conclusiones

En este artículo hemos considerado el problema de la predicción de radiación directa a muy corto plazo sobre un conjunto de datos concreto recopilados en la Plataforma Solar de Almería durante un periodo de un año. Hemos comprobado que, mediante la utilización de redes LSTM, es posible mejorar, aunque por un margen estrecho, la predicción basada en la repetición del último valor conocido, método simple pero muy efectivo en los primeros minutos, y que sí mejora los resultados de otras técnicas habituales como la predicción basada en la media, o incluso de otras más complejas como ARIMA.

En cuanto a trabajo futuro, debemos señalar que nuestra red solo considera cada estación individualmente, es decir la predicción está basada en los datos de la propia estación durante la hora anterior. Pensamos que el desarrollo de modelos multidimensionales que empleen los resultados de todas las estaciones para predecir los de una en concreto pueden llevarnos a predicciones más precisas. La intuición es que la radiación en un punto puede venir anticipada por la de estaciones cercanas sobre las que, por ejemplo, ya ha comenzado a pasar una nube.

## 7. Acknowledgements

Este trabajo ha sido subvencionado parcialmente por el proyecto nacional español TIN2015-66471, por el proyecto Santander-UCM PR26/16-21B-1.

*The authors declare that no competing interests exist.*

## 8. References

[1] D. Renne, *Semi-Annual Status. Task 36: Solar. Resource Knowledge Management*. Solar Resource Knowledge Management. 2009.

[2] E Lorenz, A Hammer, and D Heinemann. *Short term forecasting of solar radiation based on satellite data*. In EURO SUN 2004 (ISES Europe Solar Congress), pages 841- 848, 2004.

[3] Bosch, J. L. y Kleissl, J. Cloud motion vectors from a network of ground sensors in a solar power plant. *Solar Energy* 95(1), 13-20, 2013.

[4] G. Reikard. *Predicting solar radiation at high resolutions: A comparison of time series forecasts*. *Solar Energy*, 83(3):342-349, 2009.

[5] Martín, L., Zarzalejo, L. F., Polo, J., Navarro, A., Marchante, R. y Cony, M. Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning. *Solar Energy* 84(10), 1772-1781, 2010.

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

[7] A. Mellit and A. M. Pavan. *A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy*. *Solar Energy*, 84(5), 807-821, 2010.

[8] M. Bou-Rabee, S. A. Sulaiman, M. Saleh, and S. Marafi. *Using artificial neural networks to estimate solar radiation in Kuwait*. *Renewable and Sustainable Energy Reviews*, 72, 434-438. 2017.

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

[10] I. A. Walter, R. G. Allen, R. Elliott, M.E. Jensen, D. Itenfisu, B Mecham, ... and T Spofford. *ASCE's standardized reference evapotranspiration equation*. In *Watershed Management and Operations Management 2000* (pp. 1-11). 2000.

[11] T. Caliński, and J. Harabasz. *A dendrite method for cluster analysis*. *Communications in Statistics-theory and Methods*, 3(1), 1-27. 1974

[12] D. Kwiatkowski, P. C B Phillips, P. Schmidt, and Y. Shin. 1992. *Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?* *Journal of Econometrics* 54 (1-3): 159-78. 1992.

[13] R. J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice*. OTexts. 2018.

[14] S. Makridakis, S.C. Wheelwright, and R.J. Hyndman. *Forecasting: methods and applications*. John Wiley & Sons, 1998.

- [15] S. Haykin. *Neural Networks: A Comprehensive Foundation Upper*. Saddle River, NJ, USA, pp. 1–842. 1998.
- [16] S. Hochreiter and J. Schmidhuber. *Long Short-Term Memory*. *Neural Computation*. 9 (8): 1735–1780. 1997
- [17] F. A. Gers , J. Schmidhuber and F. Cummins. *Learning to Forget: Continual Prediction with LSTM*. *Neural Computation* Volume 12-10. p.2451-2471. 2000
- [18] F. Wilcoxon. *Individual Comparisons by Ranking Methods*. *Biometrics* 1, 80-83. 1945