







Multivariate Models for Photovoltaic Power Forecasting with Non-climatic Exogenous Variables

Isidro Fraga Hurtado , Julio R. Gómez Sarduy , Zaid García Sánchez , *Member, IEEE*, Roy Reyes Calvo , Yuri U. López , *Senior Member, IEEE*, and Enrique C. Quispe , *Senior Member, IEEE*

Abstract—Forecasting electricity generation from renewable resources is crucial for the efficient planning and operation of power systems. The development of forecasting models based on local meteorological variables is common, however, sometimes this information is unavailable. This study explores the use of multivariate models that do not incorporate meteorological variables, but use historical power-generated data from eight PV plants located in the same region to predict the future value of a target plant. This allows for improved forecasting when meteorological variables are unavailable and the only information available is the generation of the PV plants. The performance of LSTM and BiLSTM networks is compared for different time horizons, considering various lags of the power series itself for estimating future values. The main contributions of this study include the introduction of power time series from other plants as model inputs, the use of spatial interpolation to fill in missing data and the application of causality tests between time series for the selection of predictor variables, and the uncertainty associated with the predictions is analyzed using quantile regression techniques.

Link to graphical and video abstracts, and to code: <https://latam.ieceer9.org/index.php/transactions/article/view/9697>

Index Terms: multivariate models, photovoltaic power, forecasting, renewable energy.

I. INTRODUCCIÓN

A nivel mundial, las últimas décadas se han caracterizado por un rápido crecimiento de las tecnologías de energía renovable. Fundamentalmente, la instalación de sistemas fotovoltaicos (PV) a nivel global ha aumentado sustancialmente, con una capacidad acumulada instalada de 623 GW en 2019 [1]. La principal desventaja del incremento en este tipo de fuentes se debe a la variabilidad en la entrega de energía y su efecto. A nivel mundial, los estudios demuestran que un aumento en la penetración de estas fuentes de energía renovable introduce varios desafíos técnicos en la operación del Sistema Eléctrico de Potencia (EPS), siendo uno

The associate editor coordinating the review of this manuscript and approving it for publication was Giner Alor-Hernández (*Corresponding author: Enrique C. Quispe*).

I. Fraga, and J. R. Gómez are with Centro de Estudios de Energía y Medio Ambiente, Universidad de Cienfuegos, Cuba (e-mails: ifraga@ucf.edu.cu, and jgomez@ucf.edu.cu).

Z. García is with Higher Polytechnic School, University of the Balearic Islands, Islas Baleares (e-mail: zaid.garcia@uib.cat).

R. Reyes is with the Department of Mechanics, Faculty of Engineering, Universidad de Cienfuegos, Cuba (e-mail: royr@ucf.edu.cu).

Y. U. López and Enrique C. Quispe are with Energy Research Group, Universidad Autónoma de Occidente, Cali, Colombia (e-mails: ylopez@uao.edu.co, and ecquispe@uao.edu.co).

de los principales el control de frecuencia [2], [3]. La estabilidad de frecuencia emerge entonces como una de las principales limitaciones a analizar al planificar la operación del EPS.

No solo la variabilidad en la entrega de energía afecta la operación de la red, sino que estas tecnologías basadas en energías renovables, especialmente aquellas basadas en energía solar y algunas tecnologías eólicas, reducen la inercia de los EPSs, comprometiendo su operación estable. Se estudian medidas para reducir el riesgo de inestabilidad de frecuencia en los EPSs, estando la mayoría dirigidas a conectar sistemas de almacenamiento que restauren rápidamente el impacto causado en los EPSs [4], [5], [6], [7]. Estos problemas se agravan a medida que aumenta la participación de estas tecnologías en la generación, siendo este aspecto incluso el factor limitante fundamental de la penetración de fuentes renovables en los EPSs.

Se debe prestar especial atención a los EPSs aislados, ya que estos generalmente consisten en unidades de generación de baja inercia, y por lo tanto, la penetración de fuentes de energía renovable puede deteriorar la robustez de los EPSs [2], [3]. Otros estudios recientes se centran en lograr una precisión adecuada en la predicción del comportamiento de la generación variable y considerarlo en estrategias para la cobertura económica del gráfico de carga.

Por otro lado, equipar al operador de la red eléctrica con herramientas para predecir el comportamiento de generación renovable a corto plazo se vuelve esencial en la investigación relacionada con los EPSs.

En [8] se sugiere que la predicción de la producción de energía por parte de las plantas solares fotovoltaicas (PVSP) puede reducir eficazmente los efectos adversos en el sistema causados por la conexión de tales fuentes a la red y ayudar a los operadores del sistema a ajustar la planificación en tiempo real y coordinar la generación de PVSP con otras fuentes de generación para garantizar la seguridad de operación del EPS. Por lo tanto, predecir la generación de electricidad a partir de sistemas solares fotovoltaicos, tanto en grandes redes eléctricas como en micro redes, desempeña un papel clave en la operación eficiente, económica, estable y sostenible del suministro de electricidad.

La penetración a gran escala de recursos de energía renovable está impulsando las políticas para establecer la previsión obligatoria de un día de la generación fotovoltaica en muchos países [9] debido a que la previsión precisa de la potencia fotovoltaica permite la planificación y programación de la red, la gestión de energía y la minimización de costos, garantizando una operación segura con calidad y el equilibrio adecuado entre la oferta y la demanda. Típicamente, se utilizan modelos de

previsión a corto plazo (unos pocos minutos hasta unas pocas horas) para fines de control y programación del flujo de energía entre cargas, fuentes y sistemas de almacenamiento. Estos modelos con horizontes de tiempo corto son esenciales para controlar y diseñar sistemas de gestión de energía en micro redes [10], [11].

Predecir la potencia de salida de los sistemas solares fotovoltaicos es un desafío porque está altamente relacionado con circunstancias externas como la radiación solar y otros parámetros climáticos cambiantes rápidamente en un lugar. Una revisión de la literatura relacionada con el tema revela que desde 2011, las publicaciones de nuevos modelos de predicción fotovoltaica y estudios sobre sus características han aumentado significativamente [12]. En todos los casos, las estrategias utilizadas para el desarrollo de modelos son similares a las utilizadas para predecir otras magnitudes de sistemas eléctricos como la carga horaria, la carga máxima y el precio de la energía en el mercado [13], [14], [15].

La tendencia actual es desarrollar modelos de previsión que utilicen nuevas técnicas o la combinación de herramientas de tecnologías disruptivas, por ejemplo, herramientas de Aprendizaje Automático o Aprendizaje Profundo para la predicción fotovoltaica, formando modelos híbridos que logren una precisión adecuada [16], [9], [17], [8], [18] [19], [20], [21].

Dependiendo del horizonte de previsión, es mejor utilizar predictores numéricos del clima (NWP) o modelos basados en datos históricos. Los primeros se utilizan principalmente para horizontes temporales entre 24 y 48 horas, mientras que los últimos son más efectivos para marcos temporales cortos como minutos o unas pocas horas [12].

Mellit, et al, plantean que la mayoría de los modelos que se reportan en la literatura se enfocan en el pronóstico horario o diario y que no se ha abordado bien la predicción a muy corto plazo a pesar de ser muy importante para la gestión de energía y el control de microrredes [10]. Esos autores usaron solamente datos históricos de la potencia fotovoltaica generada en una pequeña planta de 4 kWp y no incluyeron ninguna variable exógena (temperatura del aire, velocidad del viento, nubosidad, etc.). Otros investigadores, para mejorar la exactitud del pronóstico, han desarrollado modelos multivariados, incluyendo variables meteorológicas [22], [23], [24].

En la revisión bibliográfica realizada solo encontramos un reporte [25], donde se propone por primera vez usar como variables exógenas, para el pronóstico preciso de la potencia generada por una planta fotovoltaica, las series de potencia de salida de plantas adyacentes sin usar variables meteorológicas. En su trabajo ellos concluyen que, en investigaciones futuras deben investigar la relación que existe entre la distancia geográfica de las plantas y la precisión del pronóstico ya que su trabajo se basa en plantas que se encuentran ubicadas muy cercanas entre sí, en un radio de 2 km. Esa fue la motivación para nuestra investigación.

En este trabajo nos enfocamos entonces en el problema de pronosticar la salida de una planta fotovoltaica sin emplear datos meteorológicos e incorporando información de otras plantas ubicadas en la misma región climática pero distantes varias decenas de kilómetros unas de otras. Los datos de series históricas de salidas de otras plantas de la región, que estén disponibles, se utilizarán como entradas al modelo propuesto para predecir la salida de la planta objetivo. Se compararán

modelos basados en redes LSTM y BiLSTM para diferentes horizontes de tiempo (5 min, 15min, 30 min y 60 min) y considerando diferente cantidad de retrasos para estimar el valor del próximo paso.

Las contribuciones de este trabajo se resumen en:

- Se introducen como entradas a un modelo multivariable de predicción de la salida de una planta fotovoltaica, las series de tiempo de potencias de otros parques ubicados en la misma región climática, pero distantes a decenas de kilómetros. Esto toma en cuenta la no disponibilidad o pérdida de información meteorológica que es una realidad en muchas instalaciones y se aplica de manera innovadora a parques mayores de 2 MWp conectados al sistema eléctrico nacional.
- El preprocesamiento de la información considera también la potencia generada en todos los parques en la aplicación del método de interpolación espacial denominado Inverse Distance Weighting para la imputación de datos para completar la información.
- Para la selección de los parques a incluir como entradas del modelo no solo se analiza la correlación de Pearson, como es costumbre en la selección de variables predictoras influyentes en los modelos, sino que se realiza un test de Granger como método estadístico recomendado para establecer causalidad entre series temporales.

- Se cuantifica la incertidumbre asociada con el pronóstico de la potencia fotovoltaica a partir de la regresión cuantílica que considera la variabilidad en diferentes percentiles de la distribución de los errores.

Se trabaja sobre un caso de estudio que es el parque solar de 5.5 MWp ubicado en Yaguaramas, en la provincia de Cienfuegos, Cuba. Todos los parques involucrados en el estudio se encuentran ubicados geográficamente en la región central de Cuba en un área que comprende tres provincias.

El trabajo está estructurado como sigue: Una sección introductoria que plantea brevemente el problema a resolver y que resume la contribución del mismo a partir de un análisis de la literatura consultada. La segunda sección expone la metodología y los modelos propuestos. En la tercera sección se muestran los resultados y la discusión, basados en el caso de estudio y en la cuarta sección se dan las conclusiones de este trabajo.

II. MATERIALES Y MÉTODOS

El esquema general de la metodología seguida para la elaboración de los modelos de pronóstico se muestra en la Figura 1. El primer paso es la obtención y procesamiento de la información para disponer de una base de datos que contenga las mediciones de la potencia generada por los parques fotovoltaicos a considerar. Esta base de datos es adquirida a partir de los Sistemas de Supervisión, Control y Adquisición de Datos (SCADA) que se encuentran activos en los parques fotovoltaicos de una región determinada.

Para el preprocesamiento de los datos se elaboró un software que tiene la función de automatizar el proceso de limpieza y completamiento de la base de datos la cual se importa desde una hoja de cálculo de Excel. En todos los casos se detectan y eliminan los registros repetidos y se crean los registros faltantes.

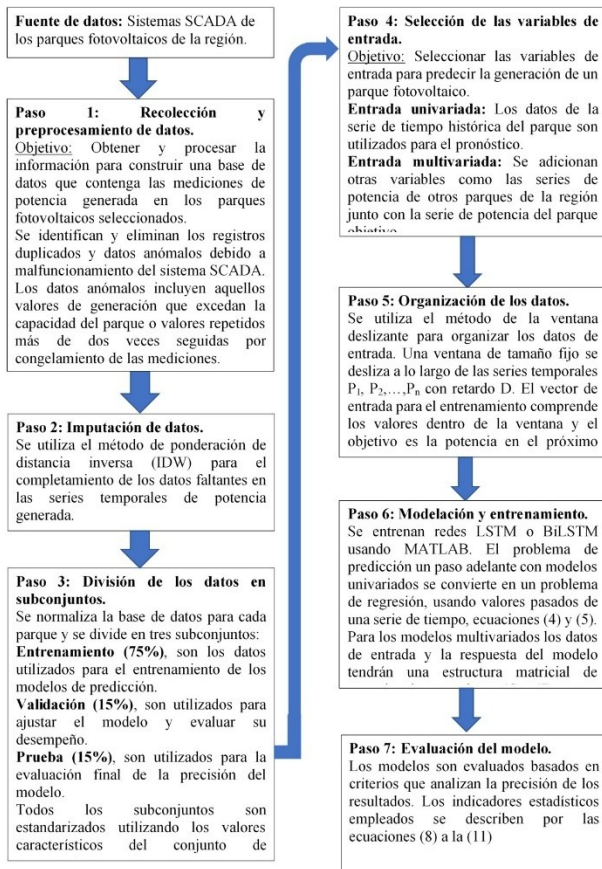


Fig. 1. Esquema del proceso de desarrollo de los modelos de pronóstico.

Como resultado de mal funcionamiento del sistema SCADA o falla durante el proceso de recolección de datos, aparecen registros con datos faltantes o anómalos. Los datos faltantes y anómalos deben ser tratados para completar adecuadamente la base de datos original. En esta investigación se consideran datos anómalos aquellos valores que superan la capacidad máxima del parque y aquellos que se repiten más de dos veces en registros consecutivos producto de congelamiento de las mediciones. Todos estos valores son tratados como datos faltantes y son completados mediante un procedimiento de imputación de datos faltantes. En todo este proceso de completamiento de los datos se trabaja con valores normalizados a la potencia máxima del parque con el objetivo de poder aplicar el método de imputación de datos seleccionado.

Una vez completada la base de datos, y desnormalizada para cada parque, se divide la misma en un conjunto de entrenamiento, un conjunto de validación y un conjunto de prueba. El primer conjunto de datos comprende el 75% de los registros. Un 15% de los datos se toma para conjunto de validación y el 10% restante se utiliza como conjunto de prueba. Además, antes de entrenar los modelos, todos los valores estos subconjuntos, para cada parque, se estandarizan usando los valores característicos de los datos de entrenamiento de acuerdo a la expresión siguiente:

$$P_{std_{i,j}} = \frac{P_{i,j} - \mu_j}{\sigma_j} \quad (1)$$

donde $P_{std_{i,j}}$ es el valor estandarizado del i -ésimo registro de la potencia del parque j , $P_{i,j}$ es el valor real del i -ésimo registro de la potencia del parque j , μ_j es el valor medio de las mediciones de potencia correspondiente al conjunto de entrenamiento para el parque j y σ_j es la desviación estándar las mediciones de potencia correspondiente al conjunto de entrenamiento para el parque j .

Para el completamiento de los datos faltantes en series temporales existen diferentes métodos descritos en la literatura [26], [27], [28], [29]. En esta investigación se utiliza el método de ponderación de distancia inversa (IDW por sus siglas del inglés Inverse Distance Weighting) [26], [27]. Este es un método determinista de interpolación espacial basado en la distancia entre la ubicación para la cual se debe interpolar un valor y las ubicaciones de las observaciones registradas. Se fundamenta en el supuesto de que la potencia que se genera en un parque fotovoltaico en una ubicación particular presenta una elevada correlación con los valores de potencia generada en parques ubicados en otras localidades de la misma región. Por lo tanto, se puede estimar la potencia generada en cualquier parque por medio de una combinación lineal de los valores medidos de potencia generada en los parques vecinos. Los valores faltantes en la serie de potencia de un parque se sustituyen por valores calculados por la siguiente expresión:

$$P_E = \frac{\sum_{i=1}^n W(r_{i,E}) P_i}{\sum_{i=1}^n W(r_{i,E})} \quad (2)$$

$$W(r_{i,E}) = \frac{1}{r_{i,E}^p} \quad (3)$$

donde P_i es el registro de la potencia en el sitio i , con $i = 1, 2, \dots, n$, $W(r_{i,E})$ es la función de ponderación del i -ésimo sitio, $r_{i,E}$ es la distancia en km entre el parque fotovoltaico i -ésimo y el parque fotovoltaico de estimación y p es el exponente utilizado en la interpolación. La literatura consultada [27] propone generalmente $p = 2$ para proporcionar aún más peso a los lugares más cercanos. La altitud no se considera para la estimación de los valores faltantes porque todas las estaciones se encuentran prácticamente a la misma altura sobre el nivel del mar.

El método de imputación de datos descrito se fundamenta en el supuesto de que existe una elevada correlación entre la generación de potencia del parque objetivo y la generación de potencia de otros parques de la región. Basado también en esta correlación, otros investigadores han empleado información de parques vecinos como variables de entrada en el modelo de predicción del parque objetivo. Zhen et al, propusieron incorporar, como variables de entrada, la información de siete parques adyacentes para predecir la generación en un parque objetivo [25], lo cual mejoró el modelo de pronóstico.

Motivados por sus resultados nosotros vamos a investigar el efecto de introducir en el modelo, no solo los valores u observaciones pasadas del mismo parque, sino de otros que se encuentran ubicados en la región central.

En este sentido, cuando se trata de definir si una variable es útil para predecir la otra, no solo se puede establecer a partir del

coeficiente de correlación, sino que una forma más útil es el test de causalidad de Wiener-Granger o causalidad de Granger.

El test de causalidad de Wiener-Granger, o causalidad de Granger, es una prueba estadística empleada para determinar si una serie temporal puede predecir a otra. Para ello se basa en la idea de que si una serie temporal X causa otra Y, los modelos de Y en los que se emplean datos retrasados de X e Y deben funcionar mejor que los basados únicamente en datos retrasados de Y permitiendo identificar en series temporales en las que se observa una correlación que variable antecede a la otra.

El test realiza la siguiente hipótesis nula: la serie de la primera columna del argumento no es causa de la segunda con un cierto retraso. Así, si se obtiene p-valores por debajo del nivel de significancia se puede rechazar la hipótesis concluyendo que la causalidad existe. Para la realización de esta prueba, las series temporales deben no deben tener datos faltantes y ser estacionarias. La primera condición se cumple con el completamiento de los datos por imputación.

Para verificar la estacionariedad se realizó una prueba de Dickey-Fuller aumentada (ADF). Esta es un tipo de prueba estadística que determina si una raíz unitaria está presente en los datos de series de tiempo. Las raíces unitarias pueden causar resultados impredecibles en el análisis de series de tiempo. Se forma una hipótesis nula en la prueba de raíz unitaria para determinar qué tan fuertemente los datos de series temporales se ven afectados por una tendencia. Esta prueba tiene como hipótesis nula que la serie bajo estudio tiene al menos una raíz unitaria [30]. Al aceptar la hipótesis nula, aceptamos la evidencia de que los datos de la serie temporal no son estacionarios. Al rechazar la hipótesis nula o aceptar la hipótesis alternativa, aceptamos la evidencia de que los datos de la serie de tiempo son generados por un proceso estacionario. Este proceso también se conoce como tendencia estacionaria. Los valores de la estadística de prueba ADF son negativos. Los valores más bajos de ADF indican un rechazo más fuerte de la hipótesis nula. Los resultados de la prueba ADF se muestran en la tabla I.

TABLA I
PRUEBA DE DICKEY-FULLER AUMENTADA PARA CADA SERIE DE TIEMPO

Serie de tiempo	h	Decisión	Valor-P	Estadístico	Valor Crítico
Cagaguas	1	"Rechazar H0"	0.001	-933.875	-19.416
Frigorífico	1	"Rechazar H0"	0.001	-968.578	-19.416
Guasimal	1	"Rechazar H0"	0.001	-792.494	-19.416
Marrero	1	"Rechazar H0"	0.001	-874.163	-19.416
Mayajigua 1	1	"Rechazar H0"	0.001	-741.631	-19.416
Mayajigua 2	1	"Rechazar H0"	0.001	-860.585	-19.416
Venegas	1	"Rechazar H0"	0.001	-799.035	-19.416
Yaguaramas	1	"Rechazar H0"	0.001	-752.793	-19.416

Como se observa, el rechazo de la hipótesis nula en esta prueba (indicado por "Rechazar H0") significa que hay suficiente evidencia para afirmar que las series de tiempo en cuestión son estacionarias. Esto es importante porque permite utilizar

técnicas avanzadas de análisis y modelado en estos datos sin preocuparse por problemas asociados con series no estacionarias.

A. Estructura de las Variables Predictoras y la Respuesta del Modelo

La estructura de los datos depende del horizonte de predicción, así como del número de variables y de retrasos a considerar. La determinación de las variables de entrada para la predicción de la potencia fotovoltaica generada por un parque es crucial, y depende no solo de su disponibilidad, sino de su correlación con las respuestas [25].

En esta investigación se exploran modelos univariados, es decir, aquellos donde la predicción se hace a partir de los datos históricos de la propia serie y modelos multivariados donde se incorporan como variables, además de la propia serie, la generación de potencia de otros parques vecinos.

De igual manera los modelos pueden incorporar varios retrasos (D) en la secuencia de entrada para predecir un valor futuro. Para verificar cuales son los retrasos que deben incluirse se hace una prueba de autocorrelación.

La manera de estructurar la muestra de datos a suministrar a la red es mediante el método de ventana deslizante como se ilustra en la Figura 2. Una ventana de tamaño fijo se desliza por las series de tiempo de potencia P1 y P2 que tiene igual tamaño (D). Los valores en la ventana se toman como vector de entrada de entrenamiento y la potencia en el siguiente instante es el objetivo. Durante la prueba o la validación, los valores de la ventana deslizante son utilizados como entradas del modelo para predecir la potencia en el siguiente instante de tiempo. La longitud de la ventana es igual a los retrasos que se consideran para predecir el próximo valor.

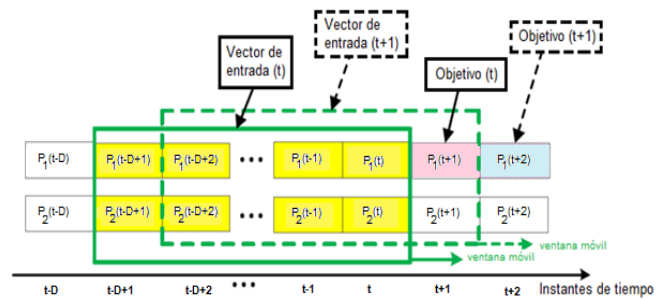


Fig. 2. Ventana móvil sobre las series de tiempo.

Este método se implementa en Matlab para introducir los datos como arreglos numéricos bidimensionales, donde cada columna de la matriz representa un vector de entrada. Considere, por ejemplo, el caso de un modelo univariado-onestep, este problema de predicción se puede expresar como un problema de regresión utilizando muestras de valores pasados de una única serie de tiempo bajo análisis. Si $P_1 = \{P_1[t]\}$, $t > 0$, es la serie de tiempo de los valores de potencia generada en el parque objetivo y se conoce el valor de la potencia en el instante de tiempo t y todos los valores anteriores, entonces, el valor a predecir para un horizonte de tiempo h será $P_1[t + h]$, $h > 0$.

La predicción $\hat{P}_1[t + h]$ se obtiene determinando los parámetros de un modelo de regresión en el que el vector de

entrada está formado por D registros pasados de la serie de tiempo:

$$x_1[t] = [P_1[t] \quad P_1[t-1] \quad \dots \quad P_1[t-D+1]]^T \quad (4)$$

Entonces,

$$\hat{P}_1[t+h] = f_u(x_1[t]; \mathfrak{X}_{t-1}) \quad (5)$$

donde f_u es el modelo de regresión univariado con los parámetros propios determinados mediante un proceso de entrenamiento dirigido por datos, \mathfrak{X}_{t-1} es la información pasada asociada a los estados internos de cualquier modelo de redes neuronales dinámicas y relacionadas con la serie de tiempo hasta t-1.

Para el caso de modelos multivariados, considerando que las variables exógenas a considerar son las potencias de n parques de la región, se puede aplicar la expresión (4) para cada variable de manera que se tiene:

$$\begin{aligned} x_1[t] &= [P_1[t] \quad P_1[t-1] \quad \dots \quad P_1[t-D+1]]^T \\ x_2[t] &= [P_2[t] \quad P_2[t-1] \quad \dots \quad P_2[t-D+1]]^T \\ &\vdots \\ x_n[t] &= [P_n[t] \quad P_n[t-1] \quad \dots \quad P_n[t-D+1]]^T \end{aligned} \quad (6)$$

Para este caso la ecuación (5) queda:

$$\hat{P}_1[t+h] = f_u(x_1[t]; x_2[t]; \dots; x_{n1}[t]; \mathfrak{X}_{t-1}) \quad (7)$$

B. Redes Neuronales Empleadas en la Modelación

En esta investigación se emplearon redes LSTM (Long Short-Term Memory) que son un tipo especializado de redes neuronales recurrentes (RNN) compuestas por unidades de memoria llamadas celdas. Estas celdas están diseñadas para recordar y olvidar información a lo largo del tiempo. Las redes LSTM tienen una estructura interna compleja y utilizan puertas para regular el flujo de información permitiendo que la red aprenda a retener u olvidar información relevante a lo largo de secuencias temporales.

Una extensión de las redes LSTM estándar son las redes Bidireccionales LSTM (BiLSTM) que permiten procesar la información en ambas direcciones: hacia adelante en el tiempo (como en una LSTM tradicional) y hacia atrás en el tiempo.

C. Métricas para Evaluar la Calidad de los Modelos

Los indicadores estadísticos más usuales para evaluar el funcionamiento de un modelo de predicción son el porcentaje de error absoluto medio (MAPE), el error absoluto medio (MAE), la raíz del error medio cuadrático (RMSE) y el coeficiente de correlación (r). Estos indicadores de error, utilizados en esta investigación, se calculan como:

$$MAPE = \frac{100}{N} \cdot \sum_{t=1}^N \left| \frac{P(t) - \hat{P}(t)}{P(t)} \right| \quad (\%) \quad (8)$$

$$MAE = \frac{100}{N} \cdot \sum_{t=1}^N |P(t) - \hat{P}(t)| \quad (\%) \quad (9)$$

$$RMSE = \sqrt{\sum_{t=1}^N \frac{1}{N} \cdot (P(t) - \hat{P}(t))^2} \quad (kW) \quad (10)$$

$$r = \frac{\sum_{t=1}^N (P(t) - \bar{P}) \sum_{t=1}^N (\hat{P}(t) - \bar{\hat{P}})}{\sqrt{\sum_{t=1}^N (P(t) - \bar{P})^2 \sum_{t=1}^N (\hat{P}(t) - \bar{\hat{P}})^2}} \cdot 100\% \quad (11)$$

donde N es el número de observaciones, P(t) es valor real (medido) de la potencia fotovoltaica generada en el instante t, en kW, $\hat{P}(t)$ es la predicción de la potencia fotovoltaica generada para el instante t, en kW, \bar{P} es el valor medio de la potencia generada, en kW y $\bar{\hat{P}}$ es el valor medio de la potencia pronosticada, en kW.

III. RESULTADOS Y DISCUSIÓN

Como se mencionó anteriormente se tomó como parque objetivo el parque fotovoltaico ubicado en Yaguaramas, provincia de Cienfuegos. La serie de potencia de este parque en lo adelante se denominará como P₁. Se tienen además los datos de las series de potencia de otros siete parques ubicados en la región central de Cuba. Los mismos se corresponden con: Caguaguas (P₂), Firgorífico (P₃), Guasimal (P₄), Marrero (P₅), Mayajigüa-1 (P₆), Mayajigüa-2 (P₇) y Venegas (P₈). Las mediciones de potencia generada contenidas en el paquete son agrupadas por años y están dadas con una resolución de 5 min. Por tanto, para esta investigación se dispone de los datos de potencia generada en MW correspondientes al año 2021, lo que comprende 105 120 registros de potencia por cada parque que agrupan 288 registros diarios. El 75% de los registros se toma como datos de entrenamiento, mientras el 15% se utiliza para validación y el restante 10% para prueba.

Se empleó el software Matlab® R2021b tanto para el tratamiento inicial de los datos como para la obtención y validación de los modelos de regresión usando Deep Learning. El desarrollo de los diferentes modelos se ejecutó en una laptop HP EliteBook 855 G8 Notebook PC con procesador AMD Ryzen 7 PRO 5850U a una frecuencia 1,901 GHz y una memoria RAM de 16 GB.

Primeramente, se verificó la relación existente entre las series de potencia de los diferentes datos para comprobar la factibilidad de incorporar la información de otros parques como variables predictoras en los modelos. Para ello se determinó el coeficiente de correlación de Pearson cuyos resultados se muestran en la Figura 3. Se observa que existe una fuerte correlación positiva entre las potencias generadas por los parques considerados, en el caso del parque de Yaguaramas por encima de 0.72 con relación a los demás.

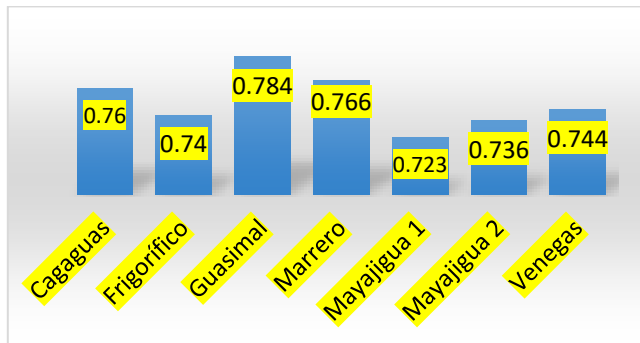


Fig. 3. Coeficiente de correlación de Pearson entre las series de potencia generada por cada parque con relación a la serie de potencia del parque de Yaguaramas.

Se verificó que todas las series son estacionarias de acuerdo a los resultados de la prueba de Dickey-Fuller aumentada y no se rechaza la hipótesis alternativa de incluir las potencias generadas en los parques de la región en los modelos del parque de Yaguaramas de acuerdo a los resultados de la prueba de causalidad de Granger. El estudio del efecto de incluir la información de otros parques de la región en los modelos de pronóstico es el enfoque principal de este trabajo.

Además, para añadir características temporales a las entradas, es decir, incluir retrasos de la propia serie en las variables de entrada para predecir un valor futuro, se calculó la función de autocorrelación. Para el caso de la serie de tiempo estocástica univariada para la potencia generada en el parque fotovoltaico de Yaguaramas existe una alta autocorrelación por encima de 0.8 hasta los primeros 6 retrasos. Cada retraso equivale a 5 min correspondiente al paso de tiempo de los registros de datos. Los correlogramas de todas las series de tiempo consideradas presentan, en todos los casos, hasta 10 retrasos con autocorrelación por encima de 0.7.

A. Modelos Analizados

Con el objetivo de analizar el impacto de considerar diferentes entradas sobre la calidad del pronóstico en el parque objetivo, se desarrollaron varios modelos empleando redes LSTM y BiLSTM. Se desarrollaron modelos univariados y se desarrollan también modelos multivariados, donde las variables exógenas son las potencias generadas en otros parques de la región.

El modelo univariado-onestep es el tipo de predicción más simple que se puede realizar. En este tipo de configuración se usa una variable a la entrada del modelo y se tiene una variable de salida y la predicción se realiza tan sólo un instante de tiempo a futuro dentro de la serie. Además, generalmente la variable de entrada es la misma variable a predecir. En los modelos multivariados-onestep se ingresa a la red múltiples variables y la predicción se hace sobre una sola variable y tan sólo para un instante de tiempo a futuro. La incorporación de otras variables predictoras como entradas del modelo se realiza para mejorar el pronóstico, y normalmente las variables a considerar son variables meteorológicas. Sin embargo, en esta investigación se propone emplear las potencias generadas en otros parques de la región como se explicó anteriormente.

Los modelos se concibieron para aceptar un registro histórico de la potencia generada cada 5 minutos y aprenderá a predecir el comportamiento que tendrá esa misma variable potencia en un instante de tiempo futuro dado por el horizonte de tiempo que se determine. Se evaluaron modelos para diferentes horizontes de tiempo. Los horizontes de tiempo que se consideraron fueron 5, 30 y 60 minutos.

Como se demostró, existe autocorrelación entre los valores pasados de la serie y el valor actual, por lo tanto, se van a explorar diferentes modelos con varios retrasos de tiempo para ver su efecto sobre la calidad de la predicción. Debido a que para cada caso se pueden realizar decenas de pruebas con distintas configuraciones, en este artículo solo se muestran los modelos que arrojan los mejores resultados u ofrecen información relevante.

Con el objetivo de apreciar el efecto de variar las entradas, los retrasos y el horizonte de predicción se mantuvo para todos los casos la estructura y los hiperparámetros de las redes. Estos son: Arquitectura: LSTM/BiLSTM → DropoutLayer → fullyConnectedLayer → regressionLayer; Tasa de aprendizaje inicial: 0.005; Épocas para actualización de la tasa de aprendizaje: 10; Razón de actualización de la tasa de aprendizaje: 0.1; Número de épocas: 200; Tamaño de batch: 128; Dropout: 0.5; Función de activación: tanh; Algoritmo de optimización: Adam y Función de pérdidas: MAE

Los casos estudiados con diferentes entradas y arquitecturas se muestran en la Tabla II.

TABLA II
ARQUITECTURA Y ENTRADAS DE LOS DIFERENTES MODELOS

Modelo	Arquitectura	Entradas	Retrasos (D)	Horizonte (h)
Modelo 1	LSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁	1	1
Modelo 2	LSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁	6	1
Modelo 3	BiLSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁	6	1
Modelo 4	BiLSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁	6	6
Modelo 5	BiLSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁	6	12
Modelo 6	BiLSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁ , P ₂ , P ₃ , P ₄ , P ₅ , P ₆ , P ₇ , P ₈	6	1
Modelo 7	BiLSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁ , P ₂ , P ₃ , P ₄ , P ₅	6	1
Modelo 8	BiLSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁ , P ₂ , P ₃ , P ₄ , P ₅	6	6
Modelo 9	BiLSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁ , P ₂ , P ₃ , P ₄ , P ₅	6	12
Modelo 10	BiLSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁ , P ₄	6	12
Modelo 11	BiLSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁ , P ₄	6	1
Modelo 12	BiLSTM Units=128; Dropout=0.5; fullyConnected=1	P ₁ , P ₄	6	6

Los primeros 5 modelos son univariados, es decir, solo tienen como entrada la serie de potencia del parque objetivo Yaguaramas (P₁). El resto de los modelos son multivariados. El modelo 6 considera todos los parques, incluyendo el parque objetivo y los tres modelos restantes solo los parques que mayor

correlación tienen con el parque objetivo, incluyendo este, es decir, P_1 , P_2 , P_4 y P_5 .

En la Tabla III se muestran los indicadores estadísticos para evaluar la calidad de los modelos obtenidos para diferentes horizontes de tiempo. Estas estadísticas de error son la base de comparación entre los diferentes modelos.

TABLA III
ESTADÍSTICAS DE ERROR DE LOS DIFERENTES MODELOS

Modelo	Horizonte de tiempo	RMSE (kW)	MAE (kW)	MAPE (%)	r (%)
Modelos univariados					
Modelo 1	5 min	445.28	148.83	12.52	94.09
Modelo 2	5 min	416.06	169.12	12.91	95.00
Modelo 3	5 min	407.24	138.80	12.68	95.08
Modelo 4	30 min	507.59	212.95	28.08	92.27
Modelo 5	60 min	584.40	266.61	47.39	89.70
Modelos multivariados					
Modelo 6	5 min	404.29	162.44	14.15	95.17
Modelo 7	5 min	404.12	141.96	12.99	95.16
Modelo 8	30 min	498.56	217.77	23.67	92.46
Modelo 9	60 min	593.83	294.03	37.25	89.44
Modelo 10	60 min	557.65	267.59	41.19	90.71
Modelo 11	5 min	403.60	147.65	12.97	95.17
Modelo 12	30 min	490.61	207.99	23.84	92.80

Se puede apreciar que, a medida que se incrementa el horizonte de tiempo las estadísticas de error se incrementan con relación al pronóstico de un paso adelante (5 min). No obstante, para un mismo horizonte de tiempo, estas son prácticamente del mismo orden de magnitud por lo que se dificulta la comparación. Por ejemplo, el coeficiente de correlación r no cambia mucho y se encuentra en un rango entre 94.09% y 95.17% para horizonte de 5 min, para 30 min está entre 92.27% y 92.8% y para 60 min cae entre 89.31% y 90.44%.

Entonces, para comparar modelos con el mismo horizonte de tiempo vamos primeramente a comprobar cuáles son diferentes entre sí. Para ello realizamos la Prueba de Tukey, también conocida como procedimiento de comparaciones múltiples de Tukey, que es una técnica estadística utilizada para comparar las medias de múltiples grupos y determinar si existen diferencias significativas entre ellos. Esta prueba es útil cuando se realizan comparaciones entre tres o más grupos, y es comúnmente utilizada en experimentos o estudios donde se quiere identificar qué grupos difieren entre sí en términos de una variable de interés. Para realizarla se utilizó la función *multcompare* de Matlab. En este caso la variable a utilizar para la comparación es el error o residuo entre el pronóstico dado por el modelo y la medición de potencia real. Al interpretar los resultados de la Prueba de Tukey, si el intervalo de confianza para la diferencia entre dos medias no incluye cero y el valor p asociado es menor que el nivel de significancia ($\alpha=0.05$), entonces se concluye que hay una diferencia significativa entre los modelos en términos del error de la predicción. Para el caso de horizontes de tiempo de 5 min la prueba de Tukey arrojó los resultados que se muestran en la Tabla IV.

TABLA IV
RESULTADOS DE LA PRUEBA DE TUKEY PARA LOS MODELOS CON HORIZONTE DE TIEMPO DE 5 MINUTOS

Modelo 1	Modelo 2	-75.5976	-59.3814	-43.1651	-0.0000
Modelo 1	Modelo 3	0.9228	17.1390	33.3553	0.0311
Modelo 1	Modelo 6	-57.0258	-40.8095	-24.5933	-0.0000
Modelo 1	Modelo 7	-2.0132	14.2030	30.4193	0.1252
Modelo 1	Modelo 11	-29.7291	-13.5128	2.7034	0.1651
Modelo 2	Modelo 3	60.3041	76.5204	92.7367	0.0000
Modelo 2	Modelo 6	2.3556	18.5718	34.7881	0.0140
Modelo 2	Modelo 7	57.3681	73.5844	89.8006	0.0000
Modelo 2	Modelo 11	29.6523	45.8685	62.0848	0.0000
Modelo 3	Modelo 6	-74.1648	-57.9486	-41.7323	0.0000
Modelo 3	Modelo 7	-19.1523	-2.9360	13.2802	0.9956
Modelo 3	Modelo 11	-46.8681	-30.6159	-14.4356	0.0000
Modelo 6	Modelo 7	38.7963	55.0126	71.2288	0.0000
Modelo 6	Modelo 11	11.0805	27.2967	43.5130	0.0000
Modelo 7	Modelo 11	-43.9321	-27.7159	-11.4996	0.0000

Las dos primeras columnas muestran los grupos comparados. La cuarta columna muestra la diferencia entre las medias estimadas de los grupos. Las columnas tercera y quinta muestran los límites inferior y superior de los intervalos de confianza del 95% para la diferencia de medias real. La sexta columna contiene el valor p para una prueba de hipótesis de que la diferencia de medias correspondiente es igual a cero. Todos los valores p inferiores a 0.05 indican que los modelos comparados difieren en términos de error.

Los modelos 1, 2 y 3 son univariados y difieren entre ellos como se demostró con la prueba de Tukey. Si se comparan con un enfoque basado en las métricas de error (RMSE, MAE y MAPE) y el coeficiente de correlación (r) de cada modelo para determinar cuál tiene un desempeño más favorable en términos de precisión y ajuste de los datos, entonces el Modelo 3 es superior entre los modelos univariados con horizontes de tiempo de 5 min.

Para el Modelo 3, el MAPE (12.68%) es superior al del Modelo 1 e inferior al del Modelo 2, tiene un RMSE = 407.24 kW y un MAE = 138.80 kW que son inferiores a los de los Modelos 1 y 2. Además, el coeficiente de correlación r = 95.08% es ligeramente más alto, lo cual indica una mejor capacidad de ajuste.

Por otro lado, los modelos 6, 7 y 11 son multivariados, también para horizontes de tiempo de 5 min. Entre estos, el Modelo 11 tiene valores de RMSE = 400.63 kW, inferior al resto. Aunque el MAE del Modelo 11 es ligeramente superior al del Modelo 7 (147.66 kW vs. 141.96 kW), es menor que el del Modelo 6 (147.66 kW vs. 162.44 kW) y su coeficiente de correlación es similar. En este caso se prefiere el Modelo 11 sobre los demás por ser menos complejo ya que solo involucra dos variables de entrada y tener un RMSE menor. Esto representa menores esfuerzos en la preparación de los datos, costo computacional y tiempo de entrenamiento.

Entonces para horizontes de tiempo de 5 min, se tiene el Modelo 3 que es univariado y el Modelo 11 que es multivariado. Ambos son basados en redes BiLSTM e incorporan seis retrasos de cada serie temporal en las características de entrada. Estos dos modelos son diferentes entre sí según se constata de la Tabla 3. Al evaluar estos modelos se observa que, el Modelo 3

tiene menor MAE y MAPE que el Modelo 11, pero estas métricas no penalizan demasiado los errores grandes. El RMSE es sensible a valores atípicos y penaliza errores grandes más fuertemente que otras métricas. Un valor menor de RMSE indica un modelo mejor en este aspecto. El Modelo 11 tiene un RMSE ligeramente menor (400.63 kW vs 407.24 kW), lo que lo hace mejor en este criterio. Por otro lado, el coeficiente de correlación del Modelo 11 es mayor y por lo tanto su coeficiente de determinación R². Un coeficiente de determinación más alto indica que el modelo explica una mayor proporción de la variabilidad en los datos, lo que generalmente sugiere un mejor ajuste del modelo. Por lo tanto, aunque ambos modelos tienen métricas muy similares, el Modelo 11 es preferible debido a su menor RMSE y mayor coeficiente de correlación.

En la Figura 4 se muestra la comparación entre el pronóstico con el Modelo 11 para un día nublado y un día claro. Note como la predicción se ajusta mejor para días claros, mientras que, para un día nublado, con bruscas variaciones de la generación, el resultado es ligeramente peor. En la Fig. 5 se muestra la correlación que existe entre los valores medidos y pronosticados por el modelo, note que el R²=0.9058 lo que indica que el 90% de la variabilidad en la variable dependiente es explicada por este.

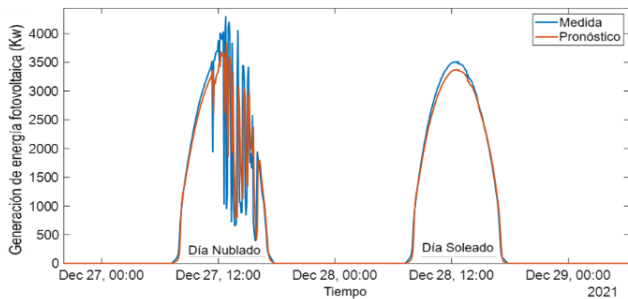


Fig. 4. Pronóstico para 5 min con un día claro y un día nublado según el Modelo 11.

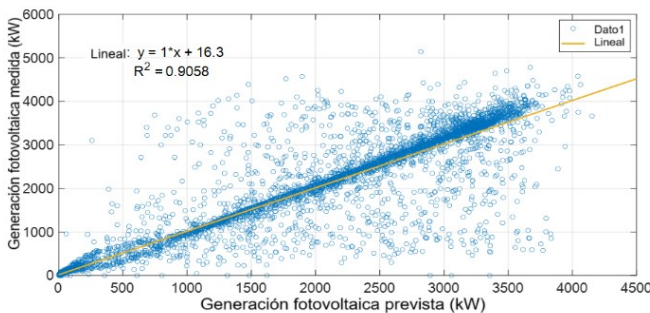


Fig. 5. Mediciones y pronóstico de la generación fotovoltaica con el Modelo 11 para el conjunto de datos de prueba.

Como se evidencia en la Tabla II, a medida que se incrementa el horizonte de tiempo la exactitud del pronóstico es menor. Los modelos 4, 8 y 12 son para horizontes de tiempo de 6 pasos adelante (30 min) y se acepta que los modelos 4 y 8, así como los modelos 8 y 12, son diferentes entre sí según la prueba de Tukey realizada (Tabla V).

TABLA V

PRUEBA DE TUKEY PARA LOS MODELOS CON HORIZONTE DE TIEMPO DE 30 Y 60 MINUTOS

Modelo 4	Modelo 8	-34.8888	-18.7523	-2.6157	0.0177
Modelo 4	Modelo 12	-10.0405	6.0960	22.2326	0.6495
Modelo 8	Modelo 12	8.7118	24.8483	40.9848	0.0009
Modelo 5	Modelo 9	-105.9158	-87.2942	-68.6725	-0.0000
Modelo 5	Modelo 10	-49.2002	-30.5786	-11.9569	0.0003
Modelo 9	Modelo 10	38.0939	56.7156	75.3373	-0.0000

De igual manera sucede para los modelos 5, 9 y 10 que se corresponden a horizontes de 12 pasos adelante (60 min) y para los cuales se acepta la hipótesis de que todos son diferentes entre sí.

Para 30 min adelante, el Modelo 12 es superior, tiene el menor valor de RMSE (490.61 kW), menor valor del MAE (207.99 kW) y el coeficiente de correlación r ligeramente superior al resto de los modelos de 30 min. Este Modelo 12 es multivariado, pero solo se ha considerado el parque cuya potencia tiene una mayor correlación con la del parque objetivo. Este resultado indica que, si bien la incorporación de parques de la región puede mejorar el pronóstico, a medida que se incrementa el horizonte de tiempo, estas variables y sus retrasos no explican bien la potencia que se va a generar en el parque. Por otro lado, en las Figuras 6 y 7 se constata que para un horizonte de tiempo de 30 min la capacidad de ajuste del modelo empeora con relación a horizonte de tiempo de 5 min. Específicamente la Figura 7 muestra la correlación que existe entre los valores medidos y los pronosticados por el modelo con un R²=0.8613; lo que indica que el 86.13% de la variabilidad en la variable dependiente se explica por este modelo.

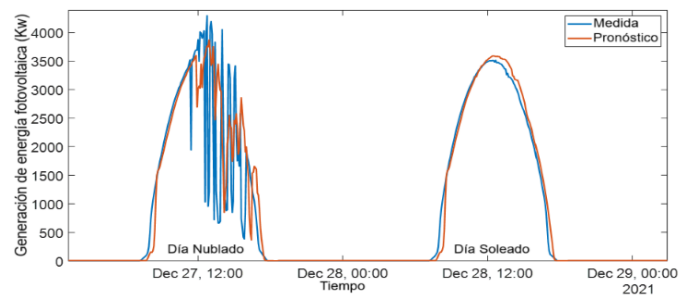


Fig. 6. Pronóstico para 30 min con un día claro y un día nublado según el Modelo 12.

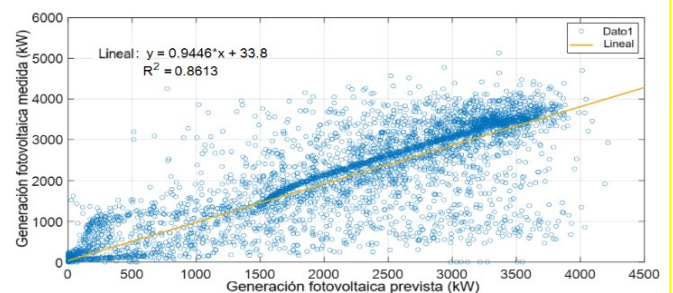


Fig. 7. Mediciones y pronóstico de la generación fotovoltaica con el Modelo 12 para el conjunto de datos de prueba.

B. Cálculo de Intervalo de Pronóstico

La cuantificación de la incertidumbre asociada con el pronóstico de la potencia fotovoltaica es esencial para la gestión y el control de las plantas por lo que es necesario establecer intervalos para el valor pronosticado [10]. Para cuantificar la incertidumbre en nuestros pronósticos y proporcionar intervalos de confianza más robustos, se implementa la regresión cuantílica, como en [31], que permite obtener intervalos de pronóstico más amplios y flexibles.

A modo de ejemplo se muestra, en la Fig. 8, el pronóstico realizado con el Modelo 11, multivariado y para un horizonte de tiempo de 5 min para los dos días considerados. Se puede apreciar que, en el caso de día nublado, donde hay bruscas variaciones de la generación, la mayoría de los valores de potencia caen dentro del intervalo de pronóstico, mientras que para días soleados, todos los valores de potencia generada caen dentro del intervalo de pronóstico.

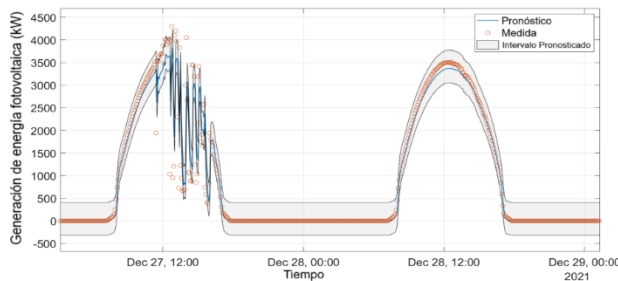


Fig. 8. Intervalo de pronóstico para el Modelo 11 con horizonte de tiempo de 5 min.

IV. CONCLUSIONES

La existencia de una fuerte correlación positiva entre las potencias generadas por los diferentes parques fotovoltaicos considerados en el estudio indica que la información de otros parques de la región puede ser útil como variables predictoras en los modelos de pronóstico del parque objetivo. Esto permite mejorar el pronóstico cuando no se cuenta con variables meteorológicas y la única información disponible es la generación de los parques fotovoltaicos de la región.

Se verificó que todas las series de potencia son estacionarias según la prueba de Dickey-Fuller aumentada, lo que valida su uso en modelos de pronóstico. Además, la prueba de causalidad de Granger no rechaza la hipótesis de incluir las potencias generadas en otros parques en los modelos del parque de Yaguaramas.

Se desarrollaron varios modelos empleando redes LSTM y BiLSTM, tanto univariados como multivariados. Los modelos se evaluaron para diferentes horizontes de tiempo y se mantuvieron constantes los hiperparámetros de las redes para una generación más precisa. Se observa que, a medida que se incrementa el horizonte de tiempo, la exactitud del pronóstico disminuye como es de esperar. También se comprobó que la inclusión de retrasos mejora el pronóstico realizado.

Para comparar los modelos entre sí se utilizó la prueba de Tukey y se encontraron diferencias significativas en términos del error de la predicción. Se identificaron los modelos más adecuados para diferentes horizontes de tiempo, considerando métricas como RMSE, MAE, MAPE y el coeficiente de

correlación. Para el caso de la predicción con horizonte de tiempo de 5 min, el mejor modelo es el Modelo 11 con 6 retrasos de tiempo e incluyendo como entradas, además de la generación de potencia del propio parque, la del parque Guasimal.

Para cuantificar la incertidumbre asociada con el pronóstico de la potencia fotovoltaica se implementó la regresión cuantílica. Este enfoque proporciona intervalos de confianza más robustos, considerando la variabilidad en diferentes percentiles de la distribución de los errores lo que contribuye a una toma de decisiones informada y a la evaluación de riesgos en la operación de las plantas. En trabajos futuros se puede investigar cómo influye la distancia entre los distintos parques fotovoltaicos en la exactitud del pronóstico incorporando series de potencia de otros parques más cercanos.

REFERENCIAS

- [1] R. Haas, N. Duic, H. Auer, A. Ajanovic, J. Ramsebner, J. Knapek y S. Zwickl-Bernhard, «The photovoltaic revolution is on: How it will change the electricity system in a lasting way.» *Energy*, vol. 265, p. 126351, 2023. DOI: 10.1016/j.energy.2022.126351.
- [2] M. Godoy, Z. García, H. Hernández, J. A. Cueto, J. I. Silva y G. Crespo, «Frequency response analysis under faults in weak power systems.» *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, n° 2, pp. 1077–1088, April 2022. DOI: 10.11591/ijece.v12i2.pp1077-1088.
- [3] V. Babu, J. Roselyn, C. Nithya y P. Sundaravadivel, «Development of Grid-Forming and Grid-Following Inverter Control in Microgrid Network Ensuring Grid Stability and Frequency Response.» *Electronics*, vol. 13, 2024. DOI: 10.3390/electronics13101958.
- [4] A. Saleh, E. Ozkop, R. J. Meng, Z. García y O. A. Betancourt, «Selecting Locations and Sizes of Battery Storage Systems Based on the Frequency of the Center of Inertia and Principle Component Analysis.» *IEEE Transactions on Industry Applications*, vol. 56, n° 2, pp. 1040–1051, 2020. DOI: 10.1109/TIA.2019.2960003.
- [5] O. A. Betancourt, Z. García, A. Saleh, E. F. Hill, X. Zhao y F. P. Sánchez, «Battery Energy Storage Systems for Primary Frequency Regulation in Island Power Systems.» en *Proceedings of the 2020 IEEE/IAS 56th International and Commercial Power Systems Technical Conference (I&CPS)*, Las Vegas, 2020. DOI: 10.1109/ICPS48389.2020.9176784.
- [6] A. K. Hasan, M. H. Haque y S. M. Aziz, «Enhancing Frequency Response Characteristics of Low Inertia Power Systems Using Battery Energy Storage.» *IEEE Access*, vol. 12, pp. 116861–116874, 2024. DOI: 10.1109/ACCESS.2024.3291234.
- [7] A. Criollo, L. Minchala-Avila, D. Benavides, P. Arévalo, M. Tostado-Véliz, D. Sánchez-Lozano y F. Jurado, «Enhancing Virtual Inertia Control in Microgrids: A Novel Frequency Response Model Based on Storage Systems.» *Batteries*, vol. 10, n° 1, p. 18, 2024. DOI: 10.3390/batteries10010018.
- [8] X. Yang, M. Xu, S. Xu y X. Han, «Day-ahead forecasting of photovoltaic output power with similar cloud space fusion based on incomplete historical data mining.» *Applied Energy*, vol. 206, pp. 683–696, 2017. DOI: 10.1016/j.apenergy.2017.08.180.
- [9] M. Pierro, F. Bucci, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada y C. Cornaro, «Multi-Model Ensemble for day ahead prediction of photovoltaic power generation.» *Solar Energy*, vol. 134, pp. 132–146, 2016. DOI: 10.1016/j.solener.2016.04.040.
- [10] A. Mellit, A. Massi Pavan y V. Lughi, «Deep learning neural networks for short-term photovoltaic power forecasting.»

- Renewable Energy, vol. 172, pp. 276–288, 2021. DOI: 10.1016/j.renene.2021.02.166.
- [11] J. Lee, J. Kang, S. Lee y H. Oh, “Ultra-Short Term Photovoltaic Generation Forecasting Based on Data Decomposition and Customized Hybrid Model Architecture,” *IEEE Access*, vol. 12, pp. 1–1, 2024. DOI: 10.1109/ACCESS.2024.3362234.
- [12] A. Muñoz, “Modelos de predicción a corto plazo de la generación eléctrica en instalaciones fotovoltaicas,” Tesis doctoral, Universidad de La Rioja, 2014. Disponible en: <https://dialnet.unirioja.es/descarga/tesis/44089.pdf>.
- [13] K. K. M. Sajjad y F. Keynia, “Applying Wavelet to ANN based short-term load forecasting: A case study of Zanjan power system,” *International Journal of Electrical and Electronics Engineering Research*, vol. 3, no. 2, pp. 209–216, 2013. DOI: 10.24247/ijeeerjun20133.
- [14] J. R. Gómez, K. G. Di Santo y M. A. Saidel, “Linear and non-linear methods for prediction of peak load at University of São Paulo,” *Measurement*, vol. 78, pp. 187–201, 2016. DOI: 10.1016/j.measurement.2015.09.053.
- [15] J. A. Peña, P. R. Viego, J. R. Gómez y E. A. Padrón, “Pronóstico del consumo pico para la gestión energética de la Universidad de Cienfuegos,” *Revista Universidad y Sociedad*, vol. 11, no. 4, pp. 220–228, 2019. Disponible en: <https://rus.ucf.edu.cu/index.php/rus/article/view/1297>.
- [16] P. Ramsami y V. Oree, “A hybrid method for forecasting the energy output of photovoltaic systems,” *Energy Conversion and Management*, vol. 95, pp. 406–413, 2015. DOI: 10.1016/j.enconman.2015.02.005.
- [17] M. A. Gómez, J. R. Gómez, J. V. Lorenzo, R. Fonte y Z. García, “Pronóstico de la generación eléctrica de sistemas fotovoltaicos. Un inicio en Cuba desde la universidad,” *Universidad y Sociedad*, vol. 13, no. 1, pp. 253–265, 2021.
- [18] I. Fraga, M. A. Gómez, J. R. Gómez y Z. García, “Predicción de la generación fotovoltaica usando Deep learning,” *Revista Universidad y Sociedad*, vol. 15, no. S1, pp. 266–275, 2023.
- [19] H. Mauladdawilah, E. J. Gago, H. Balfaqih y M. C. Pegalajar, “Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies,” *Heliyon*, vol. 10, no. e33419, 2024. DOI: 10.1016/j.heliyon.2024.e33419.
- [20] M. Gupta, A. Arya, U. Varshney, J. Mittal y A. Tomar, “A review of PV power forecasting using machine learning techniques,” *Progress in Engineering Science*, vol. 2, p. 100058, 2025. DOI: 10.1016/j.proengs.2025.100058.
- [21] S. Al-Dahidi, M. Madhjarasan, L. Al-Ghussain, A. Abubaker, A. Ahmad, M. Alrbai, M. Aghaei, H. Alahmer, A. Alahmer y P. Baraldi, “Forecasting Solar Photovoltaic Power Production: A Comprehensive Review and Innovative Data-Driven Modeling Framework,” *Energies*, vol. 17, p. 4145, 2024. DOI: 10.3390/en17114145.
- [22] F. Succetti, A. Rosato, R. Araneo y M. Panella, “Deep learning networks for multivariate prediction of photovoltaic time series,” *IEEE Access*, vol. 8, pp. 211490–21505, 2020. DOI: 10.1109/ACCESS.2020.3039999.
- [23] N. H. A. Rahman, M. Z. Hussin, S. I. Sulaiman, M. A. Hairuddin y E. H. M. Saat, “Univariate and multivariate short-term solar power forecasting of 25MWac Pasir Gudang utility-scale photovoltaic system using LSTM approach,” en 7th International Conference on Renewable Energy and Conservation (ICREC), París, Francia, 2022. DOI: 10.1109/ICREC.2022.1234567.
- [24] M. S. Hossain y H. Mahmood, “Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast,” *IEEE Access*, vol. 8, pp. 172524–172533, 2020. DOI: 10.1109/ACCESS.2020.3025194.
- [25] H. Zhen, D. Niu, K. Wang, Y. Shi, Z. Ji y X. Xu, “Photovoltaic power forecasting based on GA improved Bi-LSTM in microgrid without meteorological information,” *Energy*, vol. 231, p. 120908, 2021. DOI: 10.1016/j.energy.2021.120908.
- [26] Z. Sen y A. D. Sahin, “Spatial interpolation and estimation of solar irradiation by cumulative semivariograms,” *Solar Energy*, vol. 71, no. 1, pp. 11–21, 2001. DOI: 10.1016/S0038-092X(01)00028-1.
- [27] M. C. Crespo, “Imputación de datos faltantes en redes de distribución de baja tensión: Aplicación a edificios de pública concurrencia,” Tesis de maestría, Universidad de Sevilla, 2018. Disponible en: <https://idus.us.es/handle/11441/83537>.
- [28] J. Adaro, J. Marchesi y J. Zizzias, “Tratamiento para Datos Faltantes en Series de Radiación Solar,” *Revista Tecnología y Ciencia*, vol. 12, no. 25, pp. 60–65, 2014. Disponible en: <https://rtyc.utn.edu.ar/index.php/rtyc/article/download/468/321/1822>.
- [29] A. C. Arias, “Propuesta y evaluación de una estrategia para imputación múltiple y multivariada de valores faltantes en series de tiempo del campo meteorológico utilizando aprendizaje automático,” Tesis de maestría, Instituto Tecnológico de Costa Rica, 2022. Disponible en: <https://repositoriotec.tec.ac.cr/handle/2238/14060>.
- [30] J. Afriyie, S. Twumasi-Ankrah, K. Gyamfi, D. Arthur y W. Pels, “Evaluating the performance of unit root tests in single time series processes,” *Mathematics and Statistics*, vol. 8, no. 6, pp. 656–664, 2020. DOI: 10.13189/ms.2020.080605.
- [31] M. Massaoudi, I. Chihi, L. Sidhom, M. Trabelsi, S. S. Refaat, H. Abu-Rub y F. S. Oueslati, “An effective hybrid NARX-LSTM model for point and interval PV power forecasting,” *IEEE Access*, vol. 9, pp. 36571–36588, 2021. DOI: 10.1109/ACCESS.2021.3062776.



Isidro Fraga Hurtado received his B.Sc. degree in Electrical Engineering from the Universidad de Camagüey, Cuba, in 2001, and his M.Sc. degree in Electrical Engineering from the Universidad Central “Marta Abreu” de las Villas, Cuba, in 2019. He is currently pursuing a Ph.D. in Sustainable Energy and Industrial

Transformations at the Universidad de Cienfuegos, Cuba. Since 2022, he has been with the Universidad de Cienfuegos, where he serves as a Full Professor and Researcher at the Center for Energy and Environment Studies (CEEMA). His research interests include power quality, energy efficiency, energy management, and the application of artificial intelligence to energy systems and equipment.



Julio R. Gómez Sarduy received his B.Sc. degree in Electrical Engineering from the Universidad Central “Marta Abreu” de las Villas, Cuba, in 1986, and his M.Sc. and Ph.D. degrees in Electrical Engineering from the same university in 1997 and 2007, respectively. Since 2004, he has been with the Universidad de Cienfuegos, Cuba,

where he currently serves as a Full Professor, Researcher, and Director of the Master’s Program in Energy Efficiency at the Center for Energy and Environment Studies (CEEMA). His research interests include the analysis of electrical machines and drives, energy efficiency, energy management, and the

application of artificial intelligence to energy systems and equipment.



Zaid García Sánchez (Member, IEEE) received his B.Sc., M.Sc., and Ph.D. degrees in Electrical Engineering from the Universidad Central de las Villas, Santa Clara, Cuba, in 2002, 2004, and 2011, respectively. He was a Visiting Scholar at the Emera & NB Power Research Center for Smart Grid Technologies, University of

New Brunswick, Canada, from 2018 to 2019. Currently, he is a Postdoctoral Researcher at the University of the Balearic Islands through the Vicent Mut program, funded by the Ministry of Education and Universities of the Balearic Islands and the European Social Fund (FSE+). His research interests include power system modeling, stability analysis, planning, and protection, as well as the integration and operation of renewable energy sources and the development of smart grid functionalities.



Roy Reyes Calvo received his B.Sc. degree in Automatic Engineering from the Universidad Central “Marta Abreu” de las Villas, Cuba, in 2004, and his M.Sc. degree in Automatic and Informatics Systems from the same institution in 2007. He is currently pursuing a Ph.D. in Sustainable Energy

and Industrial Transformations at the Universidad de Cienfuegos, Cuba. Since 2004, he has been with the Universidad de Cienfuegos “Carlos Rafael Rodríguez,” where he serves as a Full Professor at the Faculty of Engineering. His research interests include the design, modeling, simulation, and control of renewable energy systems—both on-grid and off-grid—as well as the integration of renewable sources and power quality enhancement in electrical networks.



Yuri U. López Castrillón (Senior Member, IEEE) received his B.Sc. degree in Electrical Engineering from the Universidad Autónoma de Occidente, Colombia, in 1995, and his M.Sc. and Ph.D. degrees in Renewable Energies and Energy Efficiency from the Universidad de Zaragoza, Spain, in 2008 and 2010, respectively.

Since 1997, he has been with the Universidad Autónoma de Occidente in Cali, Colombia, where he currently serves as a Full Professor at the Faculty of Engineering and Basic Sciences. His research interests include the design, modeling, and simulation of renewable energy systems—both on-grid and off-grid—as well as the integration of renewable energy and power quality improvement in electrical networks.



Enrique C. Quispe (Senior Member, IEEE) received his B.Sc. degree in Electrical Engineering from the Universidad Nacional de Ingeniería, Peru, in 1980, and his M.Sc. and Ph.D. degrees in Electrical Engineering, along with an M.Eng. in Industrial Automation, from Universidad del Valle, Colombia, in 1994,

1997, and 2012, respectively. In 2019, he completed a postdoctoral research stay at the Institute of Systems and Robotics of the University of Coimbra, focusing on the optimization of electric motor systems. From 1982 to 1991, he worked as a research and development engineer in the Electric Motors Department of DELCROSA S.A., Lima, Peru. Since 1992, he has been with the Universidad Autónoma de Occidente, Cali, Colombia, where he is currently a Full Professor, Director of the Energy Research Group (GIEN), and Coordinator of the Energy Line in the Doctorate in Engineering program at the Faculty of Engineering and Basic Sciences. His research interests include the analysis of electrical machines and drives, energy efficiency and management, power quality, and energy transition challenges. He also serves as Chairman of Technical Committee 131 on Rotating Electrical Machines of ICONTEC–Colombia and Chairman of Study Committee A1 on Power Generation and Electromechanical Energy Conversion of CIGRÉ–Colombia.