

# Fuzzy Time Series Methods Applied to Short-Term Photovoltaic Power Forecasting

V. M. S. Ardila, J. N. Maciel and O. H. Ando Junior.

**Abstract**— Solar photovoltaic energy has shown a significant growth in the last decade. In the face of this growth, there are challenges to consider for the high penetration rates of solar photovoltaic, since this type of energy generation is variable. Photovoltaic generation forecasting is one of the alternatives to face such challenges. In this work we propose the realization of indirect forecasts of photovoltaic generation using Fuzzy Time Series. Considering short-term horizon predictions, two Fuzzy Time Series methods are used and evaluated to obtain a Global Horizontal Irradiance value. The first order called Multivariate WEIGHTED-FTS method applies linear chronological weights and, the second is FIG-FTS, Fuzzy Information Granular method, a higher order multivariate method which works as a wrapper that transforms real multivariate time series values into fuzzy univariate time series. The Global Horizontal Irradiance values obtained from both methods, were submitted to the spatial smoothing process, obtaining spatial irradiance on which a first-order low pass filter was applied considering the physical parameters of the photovoltaic simulated power system. This proposed indirect forecasting of photovoltaic generation was statistically evaluated and the results showed good model performance. Error statistics, such as RMSE and MBE, show that the higher order FIG-FTS method performs better than the WEIGHTED-FTS method in forecasting the GHI. Using the direct proportionality of the power of an FTS with the Global Horizontal Irradiance, spatial smoothing is performed by applying the low-pass filter considering the system area, resulting in a power simulation, also evaluated with statistical metrics. These results are analyzed and it is concluded that the application of the method is suitable for photovoltaic generation forecasting.

**Index Terms**— fuzzy time series, photovoltaic energy prediction, short-term forecasting

## I. INTRODUCCIÓN

LA Energía Solar Fotovoltaica (ESF) se ha posicionado como una fuente de energía que aporta en buena medida a la diversificación de la matriz energética mundial. Según estadísticas de la Agencia Internacional de Energía Renovable (IRENA) en el año 2010 la capacidad instalada en Sistemas Fotovoltaicos (SFV) era de 40 GW y en el 2020 esa capacidad instalada a nivel mundial alcanzó los 714 GW lo que representa un aumento de 94% en la última década [1].

Fecha de presentación: 14 de diciembre de 2021.

V.M.S. Ardila, graduanda del curso de Ingeniería de Energía, Universidad Federal de Integración Latinoamericana – UNILA, email: [vanesa.ardila@aluno.unila.edu.br](mailto:vanesa.ardila@aluno.unila.edu.br)

J.N. Maciel, candidato a Doctor del Programa Pós-Graduação Interdisciplinar en Energía & Sustentabilidad - PPGIES, Universidade Federal de Integración Latino-Americana - UNILA, e-mail: [joylan.maciel@unila.edu.br](mailto:joylan.maciel@unila.edu.br).

O.H Ando Junior Profesor del Departamento de Ingeniería Eléctrica de la Universidad Federal Rural de Pernambuco (UFRPE) y del Programa de Posgrado en Energía y Sostenibilidad (PPGIES) de la Universidad Federal de Integración Latinoamericana – UNILA, e-mail: [eng.oswaldo@gmail.com](mailto:eng.oswaldo@gmail.com)

En el caso suramericano según IRENA en el 2010 la capacidad instalada de SFV era de 43 MW y en el 2020 ese número se elevó para 12 GW representando un aumento del 99% en la región [1].

Frente a este crecimiento aparecen varios desafíos a considerar para las altas tasas de penetración de SFV, siendo que, este tipo de generación de energía es variable. Se tiene que lidiar con la existencia de un límite máximo de generación que cambia con el tiempo, de segundos a minutos a horas lo que se conoce como variabilidad y, además, este límite no se conoce con perfecta precisión que se denomina como incertidumbre [2]. El movimiento en torno al sol genera variabilidad que se consigue predecir, ya aquella asociada a las nubes puede ser momentánea y difícil de prever [3], así como la incertidumbre debida a dificultades para pronosticar el comportamiento del clima [2].

Los pronósticos de generación de la ESF son fundamentales para enfrentar los desafíos que traen la variabilidad y la incertidumbre, además son una gran herramienta para la gestión de las redes eléctricas, su seguridad y la comercialización de la energía solar [4]. Tanto las generadoras como las proveedoras que trabajan con SFV requieren previsiones para la toma de decisiones de carácter operativo y de planificación [5]. Dada también la alta dependencia de la ESF de las condiciones climáticas, su naturaleza es inestable y puede afectar la confiabilidad y calidad de la red eléctrica, causando fluctuaciones de frecuencia y voltaje [6]. Una forma de predecir la intermitencia durante la generación de energía solar fotovoltaica es aplicar métodos de predicción, con precisión y desempeño aceptables, en horizontes predefinidos [7].

Considerando este contexto, el objetivo principal de este trabajo consiste en aplicar, investigar y evaluar dos métodos de Series de Tiempo Difusas (STD) en la predicción de generación de ESF de corto plazo con datos históricos de una base de datos única, obtenida en Florianópolis, Santa Catarina en Brasil, la base de datos fue cedida por el Laboratorio de Fotovoltaica [8].

Partiendo de lo general que presenta el objetivo principal, se definen los objetivos específicos de este estudio: (1) Comparar la precisión del pronóstico de Irradiancia Global Horizontal de dos métodos STD, el método WEIGHTED de primer orden con ponderaciones cronológicas y otro método de orden superior que trabaja con Gránulos de Información Difusa, FIG, que tienen aprendizajes diferentes; (2) Posteriormente evaluar el uso de STD en una simulación de potencia mediante el método de suavización espacial [9], a partir de los pronósticos de irradiancia; (3) Transversal a los dos primeros objetivos se pretende comparar el rendimiento de las STD con diferentes

horizontes de predicción de corto plazo, siendo 5, 15 y 30 minutos.

Las STD se destacan por permitir la flexibilidad del sistema al considerar circunstancias naturales lidiando con el conocimiento vago e impreciso en datos de las series de tiempo [10]. Dentro de las herramientas posibles, los métodos de STD son implementados a través de la biblioteca de Python pyFITS [11], que desarrolla las etapas del método propuesto por [10] para la realización de pronósticos con STD. Esta biblioteca es el resultado de un trabajo del laboratorio MINDS (*Machine Intelligence and Data Science Lab*) que investiga inteligencia computacional y aprendizaje automático, optimización, visualización de datos y toma de decisiones [12].

En este estudio se evalúa el desempeño de los métodos de STD combinado con el método de suavización espacial para resolver el problema de pronósticos de generación de ESF considerando la dimensión espacial y características del sistema en cuestión. Para dicha evaluación se utiliza la recopilación hecha por [3] donde se recomiendan tres criterios para medir la precisión del modelo: el sesgo general, la dispersión y la capacidad para reproducir distribuciones estadísticas. Las métricas más recomendadas para cuantificar estos criterios son, respectivamente, el Error de Sesgo Medio (MBE), el Error Cuadrático Medio (RMSE) y el test de Kolmogorov-Smirnov (KS). Además, será empleado el Coeficiente de Determinación ( $R^2$ ) que expresa el ajuste de los datos del modelo pronosticado a los datos originales [13].

Este trabajo presenta en la sección II, el fundamento teórico, donde son detallados los enfoques de pronóstico y es expuesto el método de las STD, cómo son creadas las relaciones difusas y, a partir de esta, como se genera un pronóstico. En la sección III se expone el proceso de implementación del modelo, se describen la base de datos usada, la implementación de los hiperparámetros y el proceso de aprendizaje de cada método. En la sección IV se muestran los resultados obtenidos de las predicciones de radiación y de las posteriores simulaciones de potencia, así como una discusión de los resultados obtenidos. Finalmente, las conclusiones son mostradas en la sección V.

## II. FUNDAMENTO TEÓRICO

En esta sección se detallan los enfoques de pronóstico y las escalas temporales y se examinan brevemente los diferentes métodos de pronóstico existentes en la literatura. Se expone el método de las STD, el proceso de creación de las relaciones difusas y las etapas para generar un pronóstico.

### A. Pronóstico de Energía Solar Fotovoltaica

Se define inicialmente, esto según los datos disponibles, el recorte temporal sobre el cual se trabajarán los pronósticos, i.e, el enfoque y las escalas temporales. Sobre los enfoques temporales, en [14] se encuentran los principales en el pronóstico de la producción de SFV: enfoque directo e indirecto. Las previsiones con enfoque indirecto primero predicen la irradiancia solar y luego, utilizando un modelo de rendimiento fotovoltaico de la planta, obtienen la energía producida; las previsiones con enfoque directo calculan directamente la producción de energía de la planta [14]. De igual manera es útil distinguir las técnicas de pronóstico según las escalas temporales, dada la importancia que adquiere la

definición específica del horizonte de predicción [7]; Son especificados dos tipos de pronósticos según las necesidades de este trabajo: el pronóstico a corto plazo se ocupa de los horizontes de tiempo a partir de minutos hasta horas y el pronóstico a largo plazo que se ocupa de los horizontes temporales desde horas a unos días por delante [4].

Son resaltados los enfoques de pronósticos y las escalas temporales primero ya que se consideran transversales en los métodos de pronósticos. Ahora son enunciados varios de los métodos más usados, desde los estadísticos hasta los que usan inteligencia artificial.

Dentro de los métodos de pronóstico que se considera pertinente resaltar se tienen inicialmente los métodos estadísticos, que extraen relaciones con los datos históricos precedentes para hacer previsiones de comportamiento futuro de la planta [14]. Aquí se encuadran métodos de los más simples como el método de persistencia y los autorregresivos comparados en [15] como ARMA y ARIMA. En los métodos estadísticos la calidad de los datos históricos es esencial para un buen pronóstico [13]. Existen también los métodos físicos que a diferencia de los métodos estadísticos utilizan la configuración específica del sistema y la información meteorológica para la realización de pronósticos, métodos basados en imágenes del cielo donde son considerados la cobertura de las nubes y su profundidad óptica, y los métodos basados NWP (*Numerical Weather Prediction*) o de predicción numérica del clima [16]. Métodos basados en Inteligencia Artificial (IA) como las Redes Neuronales Artificiales (RNA) que trabajan imitando el proceso de aprendizaje de un cerebro humano, aprenden la relación entre los parámetros de entrada y las variables de salida mediante el estudio de los datos registrados anteriormente. No necesitan información característica sobre el sistema, esto hace que el uso de RNA sea ideal para modelar datos no lineales, dinámicos, ruidosos y sistemas complejos [17].

Finalmente, se hace un énfasis particular en el método de las Series de Tiempo Difusas (STD), método utilizado en este estudio. Las STD son conceptuadas como siendo parte del ramo de la Inteligencia Computacional [18] y se destacan por permitir la flexibilidad del sistema al considerar circunstancias naturales, lidiando con el conocimiento vago e impreciso en datos de las series de tiempo, de esta manera posibilita disminuir el problema de la incertidumbre y la variabilidad de los datos del recurso solar [19].

Una vez que se define el método de pronóstico se muestra la forma en la que será evaluado su rendimiento, con el intuito de padronizar los resultados y hacerlos comparativos con otros trabajos. Para dicha evaluación de rendimiento, se utiliza la recopilación hecha por [3] donde se recomiendan tres criterios para medir la precisión del modelo como son: el sesgo general, la dispersión y la capacidad para reproducir distribuciones estadísticas. Las métricas recomendadas para cuantificar estos criterios son, respectivamente, el error de sesgo medio (*MBE*), el error cuadrático medio (*RMSE*) normalizado para conseguir comparar los modelos donde las muestras tienen diferentes tamaños como indicado en [13] y el test de normalidad de Kolmogorov-Smirnov (KS) que se define como el valor máximo de la diferencia absoluta entre las dos Funciones de Probabilidad Acumulada [3].

El *MBE* y *RMSE* proveen información sobre un rango esperado de errores en una dada geografía o estación, así mismo

el *RMSE* representa de buena manera la dispersión horaria o sub horaria de los valores [3]. Además se utiliza el coeficiente de determinación ( $R^2$ ) que indica cuán próximo se encuentra los datos que fueron predichos y los datos de test [20].

$$MBE = \frac{1}{M} \sum_{t=1}^N (c_i - m_i) \quad (1)$$

$$RMSE = \left[ \frac{1}{M} \sum_{t=1}^N (c_i - m_i)^2 \right]^{\frac{1}{2}} \quad (2)$$

$$KS = \text{MAX } |\varphi(c_i) - \varphi(m_i)| \quad (3)$$

$$R^2 = 1 - \frac{\sigma^2(c_i - m_i)}{\sigma^2(m_i)} \quad (4)$$

Aquí  $N$  representa el número total de datos,  $M$  se refiere a media,  $c_i$  es el  $i$ -ésimo valor predicho y  $m_i$  representa el  $i$ -ésimo valor observado [20]. En las ecuaciones  $\sigma$  representa la varianza en la base de datos. El test de Kolmogorov-Smirnoff (KS), referido en (3), se obtiene buscando máxima la diferencia absoluta entre las distribuciones de frecuencia acumulativa de GHI  $\varphi(c_i)$  modeladas y  $\varphi(m_i)$  observadas [21].

### B. Series de Tiempo Difusas

La construcción teórica que fundamenta el método de las STD, su desarrollo a partir de los conceptos de la Lógica Difusa (LD) son presentados en esta sección

Las STD fueron introducidas por [10] como métodos no paramétricos para el pronóstico de series de tiempo basados en la teoría de Lógica Difusa que proveen una representación diferente de series de tiempo. Se destacan por permitir la flexibilidad del sistema al considerar circunstancias naturales lidiando con el conocimiento vago e impreciso en datos de las series de tiempo [19].

La principal diferencia entre las series de tiempo convencionales y las series de tiempo difusas es que las observaciones de las primeras son números reales, mientras que de las segundas son conjuntos difusos [10]. Dichos conjuntos difusos conforman el universo del discurso, entonces, dada una serie temporal  $Y \in \mathbb{R}^1$  y sus valores individuales  $y(t) \in Y$  para  $t = 0, 1, \dots, T$ , el Universo del Discurso ( $U$ ) es delimitado por los valores máximos y mínimos de  $Y$  tal que  $U = [\min(Y), \max(Y)]$  en el cual los conjuntos difusos  $f_i(t)$ , ( $t = 1, 2, 3 \dots$ ) son definidos y  $F(t)$  es la colección de  $f_i$ , así  $F(t)$  es llamada Serie de Tiempo Difusa en  $Y(t)$  [10].

Siendo el Universo del Discurso ( $U$ ) con  $U = \{u_1, u_2, \dots, u_i\}$ , donde  $u_i$  son posibles valores lingüísticos de  $U$ , entonces se define un conjunto difuso de variables lingüísticas  $A_i$  de  $U$  donde  $\mu_{A_i}$  es la función de membresía del conjunto difuso  $A_i$ , tal que  $\mu_{A_i}: U \rightarrow [0, 1]$ . Si  $u_i$  es miembro de  $A_i$ , entonces  $\mu_{A_i}(u_i)$  es el grado de pertenencia de  $u_i$  a  $A_i$  [22].

#### 1) Relaciones de Lógica Difusa

Con la definición del conjunto de datos que representan el universo del discurso se establece una Relación de Lógica Difusa (RLD) que crea la relación entre los datos de entrada y los datos estimados de salida. A continuación, se muestra la definición de las RLD hecha por [10], tanto para modelos de primer orden como de orden superior:

Suponiendo  $F(t)$  es causada solo por  $F(t - 1)$  entonces hay una relación difusa entre  $F(t)$  y  $F(t - 1)$  y puede ser expresada como:

$$F(t) = F(t - 1) \circ R(t, t - 1) \quad (5)$$

Aquí " $\circ$ " es el operador de la función Max-Min. La relación  $R$  es llamada modelo de primer orden de  $F(t)$ .

Si  $F(t)$  es causada por más conjuntos difusos,  $F(t - 1)$ ,  $F(t - 2)$ , ..., y  $F(t - p)$  ( $p > 0$ ) simultáneamente, la relación difusa es de orden superior (orden  $p$ ) y se representada por

$$F(t) = (F(t - 1) \times F(t - 2) \times \dots \times F(t - p)) \circ R(t, t - p) \quad (6)$$

Los pasos principales propuestos por [10] con las RLD, generan una demanda computacional compleja para el cálculo de cada una de las relaciones difusas. Es por esto que [23] propone una mejoría del algoritmo mediante una simplificación de las operaciones aritméticas, siendo menos compleja que la función Max-Min, creando Grupos de Relaciones de Lógica Difusa (GRLD). Considerando  $F(t - 1) = A_i$  y  $F(t) = A_j$ , la Relación de lógica difusa (RLD) puede ser definida como  $A_i \rightarrow A_j$  donde  $A_i$  y  $A_j$  son llamados lado izquierdo (LI) y lado derecho (LD) de la RLD. Tenemos que las RLDs con el mismo LI son reunidas en GRLD. Los grupos LI indican el valor de entrada del modelo y las LD corresponden a las salidas que fueron estimadas [23].

#### 2) Etapas del Algoritmo STD

En esta sección son descritos los pasos a seguir para la construcción de los modelos STD que detallan y explican el algoritmo utilizado para obtener los pronósticos de irradiancia solar y posterior simulación de los valores de potencia. Aquí (Fig. 1) está representado el método propuesto por [10] y desarrollado a través de la biblioteca de Python pyFITS, con los hiperparámetros y parámetros pertinentes al modelo.

##### a) Preprocesamiento de datos.

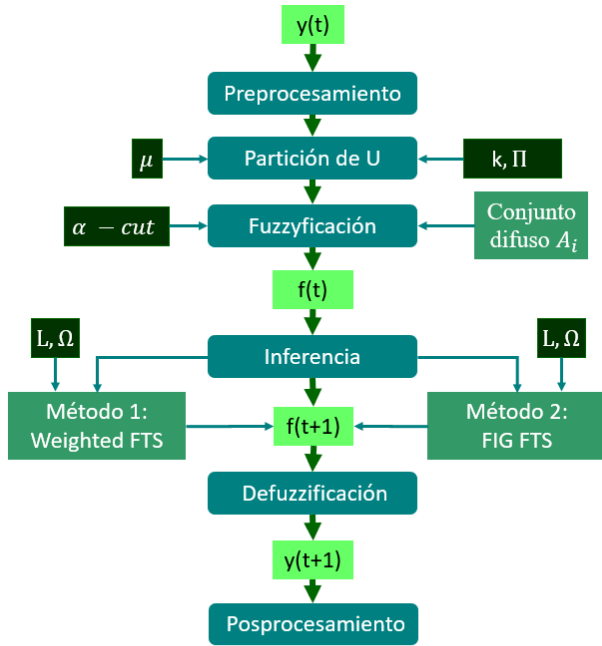
El primer paso necesario en cualquier problema de pronóstico es verificar el conjunto de datos de antemano y preprocesarlo si es necesario [19]. Aquí se manejan y adecuan los datos que puedan contener campos faltantes o información incorrecta, así como medidas repetidas. Considerando que en muchas series de tiempo se hace necesario, e.g., escalar datos dentro de un intervalo, eliminar tendencia y/o estacionalidad y entender el comportamiento de los ciclos o la aleatoriedad, que puedan interferir en los procesos de validación del modelo.

La función de Membresía, el número de particiones del universo del discurso y el método de particiones están considerados dentro de los hiperparámetros del modelo. Los hiperparámetros de los métodos STD proporcionan más versatilidad y flexibilidad, permitiendo el control de la sensibilidad del proceso de fuzzyficación, el número de reglas generadas y principalmente la precisión del modelo [24]. Los hiperparámetros comunes de los procesos STD son mostrados

en la Tabla I, donde se explica cómo influyen en el modelo; la Tabla I fue extraída de [22].

b) *Fuzzyficación de los datos*

Una vez definido el esquema de partición y el número de particiones, durante el proceso de fuzzyficación cada elemento de la Serie de Tiempo Numérica  $Y(t)$  será reemplazado por el conjunto difuso con un valor de membresía máximo, entonces será creada la Serie de Tiempo Difusa  $F(t)$  [11]. En este paso influye el hiperparámetro  $\alpha$  – cut, que es el grado mínimo de membresía a ser considerado en el proceso de fuzzyficación.



● Datos ● Hiperparámetros ● Procesos ● Parámetros

Fig. 1. Algoritmo del proceso de pronóstico con las STD. Fuente: Adaptado de [22].

TABLA I  
HIPERPARÁMETROS DEL MODELO

Símbolo	Parámetro	Descripción
$k \in \mathbb{N}$	Número de particiones	Número de conjuntos difusos que será creado en la variable lingüística
$\mu : U \rightarrow [0,1]$	Función de Membresía (FM)	Mide el valor de membresía de un valor $y \in U$ al conjunto difuso
$\Pi$	Método de partición	Define cómo será la división del universo del discurso
$\alpha \in [0,1]$	El $\alpha$ – cut,	Grado mínimo de membresía a ser considerado en el proceso de fuzzyficación
$\Omega \in \mathbb{N}$	Orden del modelo, número de atrasos	Numero de atrasos usados en el precedente de casa regla difusa
$L \in \Omega \times \mathbb{N}$	Índice de atrasos temporales	Vector de índices con longitud $\Omega$ y $1 \leq L[i] < L[i + 1]$ para $i = 0, \dots, \Omega$
$\kappa \in \mathbb{N}$	k-Nearest Neighbor (kNN)	El número de vecinos cercanos que el índice espacial busca en FIG durante la fuzzyficación

c) *Generación de las reglas difusas y proceso de Inferencia.*

La generación de las reglas difusas tiene una componente sintáctica que genera valores lingüísticos y la componente semántica que asocia cada termino lingüístico a su significado. Estas reglas dependen del método y sus características, así como de los hiperparámetros del modelo de orden ( $\Omega$ ) y los índices de retraso ( $L$ ) [22].

El proceso de Inferencia depende de método de aprendizaje escogido. Los métodos WEIGHTED-FTS y FIG-FTS, se utilizan para que el modelo aprenda y represente los patrones temporales que se encuentran en los datos que han sido fuzzyficados. El objetivo de este proceso es producir una  $f(t + 1)$  con los conjuntos difusos, para representar el valor futuro que está siendo pronosticado  $y(t + 1)$  [22].

Los métodos utilizados son clasificados como multivariados, dentro del grupo de variables se definen la variable endógena, es la que se desea predecir, también conocida como variable objetivo o variable dependiente, y las demás variables que se denominan exógenas, explicativas o independientes [22].

(1) Método WEIGHTED FTS

Este método es multivariable de primer orden propuesto por [25] que aplica ponderaciones cronológicas lineales y produce pronósticos más precisos que el método de series de tiempo difusas propuesto Chen [23] donde todas las FLR tienen el mismo peso durante el proceso de pronóstico. El método de [25] asigna los pesos apropiados a las relaciones difusas. Dicho método, como propuesto, es descrito a seguir:

Suponiendo el pronóstico de  $F(t)$  es  $A_{j1}, A_{j2}, \dots, A_{jk}$ . Los pesos correspondiente para  $A_{j1}, A_{j2}, \dots, A_{jk}$  dichos  $w_1, w_2, \dots, w_k$  son especificados. Sin embargo, después de formar la matrix de pesos con aquellos que fueron asignados se tiene  $W(T) = [w'_1, w'_2, \dots, w'_k]$  debe satisfacer la siguiente condición:

$$\sum_{h=1}^k w'_h = 1 \quad (7)$$

Por lo tanto, esos pesos  $w_1, w_2, \dots, w_k$  deben ser estandarizados. Se obtiene así la siguiente matrix de pesos:

$$W(T) = [w'_1, w'_2, \dots, w'_k] = \left[ \frac{w_1}{\sum_{h=1}^k w_h}, \frac{w_2}{\sum_{h=1}^k w_h}, \dots, \frac{w_k}{\sum_{h=1}^k w_h} \right] \quad (8)$$

donde  $w_h$  es el peso correspondiente para  $A_{jh}$

(2) Método de Información Granular Difusa (FIG)

El método de FIG (*Fuzzy Information Granular*) es un método multivariable de orden superior. Funciona como un *wrapper* que transforma valores reales de series de tiempo multivariable en series de tiempo difusas y univariadas [24].



La serie de tiempo resultante  $F$  queda compuesta por puntos de datos  $f(t) \in F$  que representan la secuencia de Gránulos de Información Difusa  $\mathcal{G}_i$ . Cada gránulo contiene un conjunto difuso de variables lingüísticas  $\tilde{\mathcal{V}}_i$  relativo a cada variable  $\mathcal{V}_i$  [24]. Existe una variable lingüística global que es la unión de todos los Gránulos de Información Difusa  $\mathcal{G}_i$ , los cuales a su vez son la combinación de uno de los conjuntos difusos por cada variable, tal que  $\mathcal{G}_i = \{A_j^{v_i}\}$ ,  $\forall \mathcal{V}_i \in \mathcal{V}$  y su función de membresía es dada por  $\mu_{\mathcal{G}_i} = \cap \mu_{A_j^{v_i}}$ , donde  $\cap$  es el mínimo T-norm. El conjunto FIG es indexado por puntos medios de sus conjuntos difusos internos. Con la variable lingüística FIG el proceso de fuzzyficación transforma cada punto de datos multivariable  $y(t) \in Y$  en  $\mathcal{G}_i \in FIG$  tal que  $f(t) = \mathcal{G}_i$  [24].

Para cada uno de estos métodos para el proceso de pronóstico se aplica la entrada al modelo y se calcula la salida que a su vez será el valor previsto [10]. El horizonte del pronóstico es el número de valores posteriores que se desean predecir, o el número de retrasos para predecir después del último valor de entrada  $y(t - L(0)), \dots, y(t - L(\Omega))$  [24].

d) *Defuzzificación*

En este paso se transforman los valores del pronóstico difuso que han sido predichos con la definición del horizonte del discurso  $H$ , los cuales son valores lingüísticos en números reales [11]. El objetivo del proceso es transformar  $f(t + 1)$  en un valor numérico estimado  $y(t + 1)$ .

e) *Posprocesamiento de Datos.*

En este último paso del proceso de pronóstico, se ejecuta, el método de suavización espacial propuesto en [9], donde el punto clave es el filtro de paso bajo de primer orden y su respectivo polo cuyo valor es una función del área del SFV [9].

Este modelo se basa en la proporcionalidad directa de la potencia de un SFV con la irradiancia incidente [26] y en que las variaciones de irradiancia en un determinado punto tienden a ser suavizadas considerando el efecto espacial, fenómeno llamado de suavización espacial [9]. Cuando el filtro es aplicado a la irradiancia pronosticada mediante STD, teniendo en cuenta el área de la planta, se obtiene a la salida del filtro, un valor suavizado llamado irradiancia espacial  $G_s(t)$  y es usado para representar la irradiancia en una superficie,  $A$  [9].

La aplicación del modelo como descrito en [26] inicia con la obtención de la serie de tiempo de irradiancia espacial  $G_s(t)$  a partir de los valores de la serie de tiempo de irradiancia que fue pronosticada  $G(t)$

$$\frac{G_s(t)}{G(t)} = \frac{1}{\tau s + 1} \quad (9)$$

Donde  $s$  es la variable de la transformada de Laplace,  $t$  es el tiempo y  $\tau$  es la constante de tiempo del filtro, aproximada considerando el valor de la frecuencia de corte,  $f_c$ , obtenida a partir del ajuste de la curva para las frecuencias de corte *cut-off* de los espectros de potencia de varias plantas SFV en función de la superficie  $A$  [Ha] del sistema [9].

$$\tau = \frac{\sqrt{A}}{2\pi \cdot f_c} \quad (10)$$

Una vez obtenida la irradiancia espacial es posible obtener mediante el uso de (11), una potencia simulada para el sistema, mediante el producto de la irradiancia espacial con la potencia instalada ( $P^*$ ) del SFV dada por [27]. Además, es considerada en esta simulación que la elevación de la temperatura del módulo reduce su eficiencia [27], es por esto que en (11) se incluye un factor en función de la temperatura del módulo a lo largo del día. Este factor aumenta o disminuye un 0,4% de la potencia como indicado en la mayoría de las hojas de datos de los módulos de SFV para cada grado que se desvía de las condiciones estándares de test [27].

$$P_{sim}(t) = \frac{G_s(t) \cdot P^*}{G_{STC}} [1 - 0,004(T(t) - T_{STC})] \quad (11)$$

Donde  $G_{STC} = 1000 \text{ W/m}^2$  es la irradiancia en condiciones estándar de test, al igual que la temperatura de contacto  $T_{STC}$  [27].

C. *Principales Estudios de la Literatura Académica*

En los últimos años los pronósticos usando STD han tenido avances significativos [13]. Se consideran trabajos como el de [4] donde se utiliza una base de datos de irradiancia registrados en intervalos de 30 minutos, genera pronósticos utilizando dos métodos de STD, mostrando el óptimo rendimiento cuando comparado con otros métodos de pronóstico. Ya en [28] se desarrolla un modelo basado en Lógica Difusa (LD) de predicción a corto plazo para predecir una hora delante de producción de energía solar y en [29] y [30] se aplica la lógica difusa para hacer predicciones de carga de corto plazo. Se destacan [19], [24] donde se proponen métodos de mejoría del método de STD como el pronóstico probabilístico y el método de gránulos de información en orden de simplificar el proceso con modelos multivariable.

Son encontradas publicaciones donde se muestra el avance del estudio de las STD como en [31] donde se introducen las Series de Tiempo Difusas no Estacionarias (NSFTS, por sus siglas en inglés) que es capaz de adaptar dinámicamente sus conjuntos difusos para reflejar los cambios en los procesos estocásticos subyacentes basados en errores residuales. Es observado en [22] la presentación de pronósticos con intervalos e probabilísticos con el tratamiento de simulaciones estocásticas o *ensembles*, lidian con los inconvenientes de las predicciones por punto, esto también es trabajado en [19] donde el problema de la distribución de probabilidad es abordado utilizando la estimación de la densidad de Kernel. Otro tema que muestra el avance en el estudio de las STD es abordado en [32], aquí se presenta un método de optimización de hiperparámetros para STD de Alto Orden Ponderadas que automatiza la generación de modelos acurados e parsimoniosos utilizando Algoritmos Genéticos.

El método de STD es utilizado junto con otros métodos de pronóstico y se evidencian resultados interesantes de este acoplamiento. En [33] combinan las STD y las Redes Neuronales Convolucionales (RNC) para el pronóstico de carga

a corto plazo, señalan buena eficiencia del método; similar lo hacen en [34] donde la propuesta es de mejorar un modelo de pronóstico solar basado en un Red Neurona Artificial con preprocesamiento de lógica difusa.

De manera particular en [29] se resalta la capacidad de los métodos STD para lidiar con variaciones repentinas en la temperatura, considerando su influencia en los SFV, de una manera simple y robusta.

### III. MATERIALES Y MÉTODOS

A continuación, se describe la implementación del modelo analítico y herramientas utilizadas para la realización de la evaluación experimental con base en el algoritmo señalado en la sección II.

En ese sentido los experimentos fueron conducidos con la finalidad de comparar dos métodos de STD, WEIGHTED-FTS y FIG-FTS a través de tres horizontes temporales de corto plazo de 5, 15 y 30 minutos. Posteriormente, como posprocesamiento, se aplica el método de la suavización espacial de [9] para obtener una simulación de potencia generada.

El procedimiento realizado es mostrado (Fig. 2) para dar un entendimiento general de lo que se aborda en esta sección. Así mismo resaltar que la simulación de potencia es consecuencia del pronóstico de irradiancia realizado a través de dos métodos STD.

#### A. Herramientas, Tecnologías y Base de Datos

Los experimentos fueron realizados en una máquina virtual creada mediante Google Colab [35], un ambiente colaborativo de programación de fuente abierta. Una herramienta gratuita para escribir y ejecutar códigos en Python. Utilizando la Unidad de Procesamiento Gráfica (GPU) que brinda la plataforma Nvidia K80s con 12GB de memoria.

La base de datos de irradiancia usada para entrenar y testar el modelo STD y los datos de potencia usados para validar el modelo del filtro de paso bajo en la simulación de potencia, son extraídos de un SFV de 2,2 kWp operando en el Centro de Pesquisa e Capacitação em Energia Solar Fotovoltaica da UFSC ubicado en la ciudad de Florianópolis/SC Brasil, donde se encuentra el sensor de irradiancia un piranómetro - SMP22 Kipp & Zonen y Sensor de temperatura y humedad - PTB110 VAISALA. Consta de 12 meses desde 01/01/2018 a 31/12/2018 aplicados para el entrenamiento del modelo y 12 meses 01/01/2019 a 31/12/2019 para testar como se observa en la Tabla II.

TABLA II  
DIVISIÓN DE LA BASE DE DATOS Y NÚMERO DE REGISTROS

Horizonte Temporal	Entrenamiento (Año 2018)	Test (Año 2019)
5 minutos	105.121	105.121
15 minutos	35.041	35.041
30 minutos	17.521	17.521

Las observaciones de dos años representan una proporción 50/50 para entrenamiento y test. Se considera como entrenamiento un año con la intención de proporcionarle al

aprendizaje del modelo STD, el comportamiento de los perfiles de estacionalidad [4] lo mismo que para el periodo de test.

En la tabla III son mostradas las variables que componen la base de datos y la salida del modelo. La base de datos está compuesta por cuatro columnas, *data* (donde se reseña la fecha, hora y minuto de cada registro) de donde se extraen las variables *min* (minuto), *hora*, *mes*; *temp\_amb* (temperatura ambiente); *ghi* (Irradiancia Global Horizontal) seleccionada como variable endógena; *temp\_contacto* (temperatura de operación del módulo) que no consta en la base de datos disponibilizada por [8] por lo que es calculada considerando temperatura de operación nominal de la célula [36]; *pot* (potencia generada por el SVF) y *pot\_modelada* (potencia modelada que resulta del proceso de aplicar el filtro de paso bajo) como salida del método.

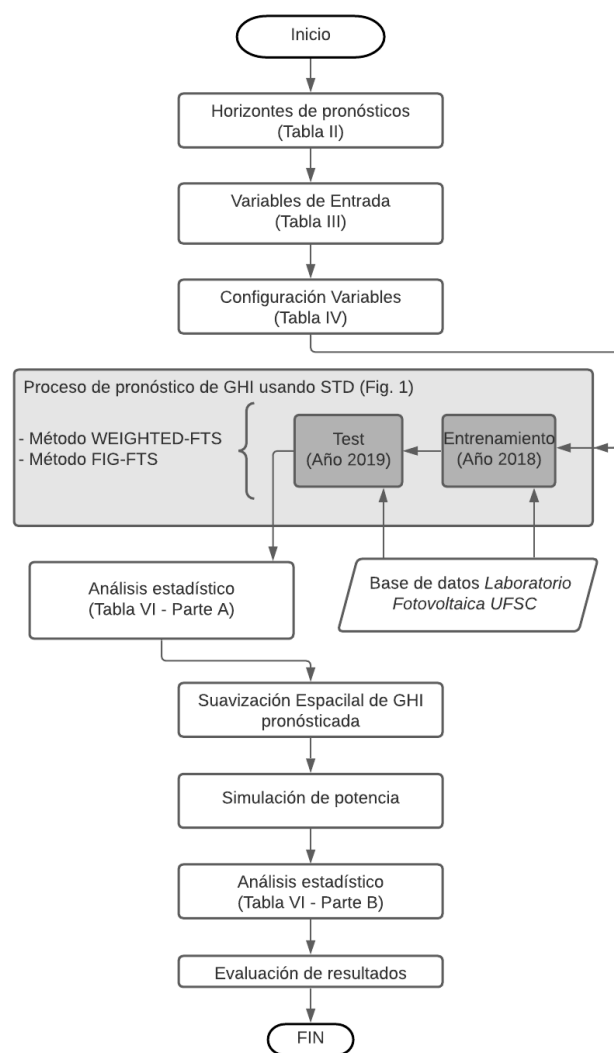


Fig. 2. Algoritmo del método de pronóstico de irradiancia y posterior simulación de potencia de este trabajo. Fuente: Elaboración propia.

La irradiancia solar tiene dos componentes de estacionalidad: anual y diaria. Estas dos componentes pueden ser extraídas de la columna *data*. La inclusión de variables meteorológicas no representa un aumento significativo en la precisión de los resultados [37], por esta razón solo es considerada la

temperatura ambiente, siendo que esta interfiere en rendimiento de los módulos solares.

TABLA III  
VARIABLES UTILIZADAS EN EL MODELO

Parámetro	Descripción de la variable	Tipo
<i>data</i>	Variable que contiene la información de estacionalidad, de aquí se extraen además las variables <i>min</i> (minuto), <i>hora</i> , <i>mes</i>	Entrada
<i>tem_amb</i>	Temperatura Ambiente	Entrada
<i>ghi</i>	Irradiancia Global Horizontal	Entrada/Salida
<i>pot</i>	Potencia generada por el sistema	Entrada
<i>temp_contacto</i>	Temperatura del módulo	Entrada
<i>pot_modelada</i>	Potencia obtenida de la suavización espacial	Salida

B. Método y Configuración Experimental

A continuación, son mostrados los hiperparámetros seleccionados, como se crean las reglas difusas y se da el proceso de aprendizaje del modelo, así como el pronóstico de GHI y posterior simulación de potencia.

1) Hiperparámetros

Una vez escogida la base sobre la cual se realiza el pronóstico de irradiancia, son definidos los hiperparámetros. Se resaltan dos cosas, la primera de ellas es el hecho de que la elección del valor de los mismos es empírica y dependiente de los datos [19]. La segunda se refiere al hecho de que el modelo es multivariable, entonces, para cada variable es necesario definir los hiperparámetros para entrenar al modelo.

En la Tabla IV se muestran los valores seleccionados para cada variable. A las variables exógenas se les asignó el mismo valor de  $\alpha - cut$  y la misma función de membresía (triangular), ya en el caso del número de particiones, para las variables relacionadas con el tiempo se respetó su división temporal y para la temperatura se seleccionó el número de particiones que estuviera cerca de su valor de variación. Con la variable endógena se experimentó usando una función de membresía diferente (gaussiana) con el intuito de captar de mejor manera la transición entre los conjuntos difusos, así mismo, el valor del grado mínimo de membresía,  $\alpha - cut$ , es un poco menor que en las demás variables, intentando incluir más valores al momento de realizar el proceso de fuzzyficación.

Como fue mencionado anteriormente, la selección de estos hiperparámetros deviene del conocimiento heurístico del proceso. La base de elección de los diferentes valores fue

TABLA IV  
CONFIGURACIÓN DE LAS VARIABLES UTILIZADAS EN EL MODELO

Variable	Función de Membresía ( $\mu$ )	Particiones ( $k$ )	Grado mínimo de membresía ( $\alpha - cut$ )
<i>Minuto</i>	Triangular	60	0,30
<i>Hora</i>	Triangular	24	0,30
<i>Mes</i>	Triangular	12	0,30
<i>Temperatura</i>	Triangular	24	0,30
<i>Irradiancia</i>	Gaussiana	35	0,25

tomada, también, con el apoyo de trabajos de pronóstico de irradiancia solar mediante STD como en [19], [22], [38].

Se muestra (Fig. 3) la representación gráfica de la función de membresía escogida para las variables, es mostrado el grafico de una FM triangular para las variables exógenas y en el caso de la variable GHI, variable endógena, que es la variable que se pretende predecir, se utiliza la FM gaussiana y se muestran los valores lingüísticos específicos que son utilizados.

2) Reglas Difusas, Aprendizaje y Pronóstico

Las particiones de la variable endógena (irradiancia) se separan por niveles de valores lingüísticos, PP – ‘muy pequeño’, P – ‘pequeño’, M – ‘medio’, G – ‘grande’, GG – ‘Muy grande’, cada uno de los valores lingüísticos con 7 subniveles. Lo cual generará reglas como el ejemplo en [38] :

$$PP5, P6 \rightarrow P6, M0, M1$$

Esta regla puede ser leída como: **SI**  $y(t - 1)$  **ES** Muy Pequeño (Subnivel 5) **Y**  $y(t)$  **ES** Pequeño (Subnivel 6) **ENTONCES**  $y(t + 1)$  **será** Pequeño (Subnivel 6) **O** Medio (Subnivel 0) **O** Medio (Subnivel 1).

El proceso creación de reglas en modelo que se describe en este trabajo, con los dos métodos WEIGHTED-FTS de primera orden ( $\Omega = 1$ ) y FIG-FTS de segunda orden ( $\Omega = 2$ ) y k-NN dos ( $\kappa = 2$ ), son el resultado del proceso de aprendizaje de cada uno de los modelos utilizando los datos de *train data*.

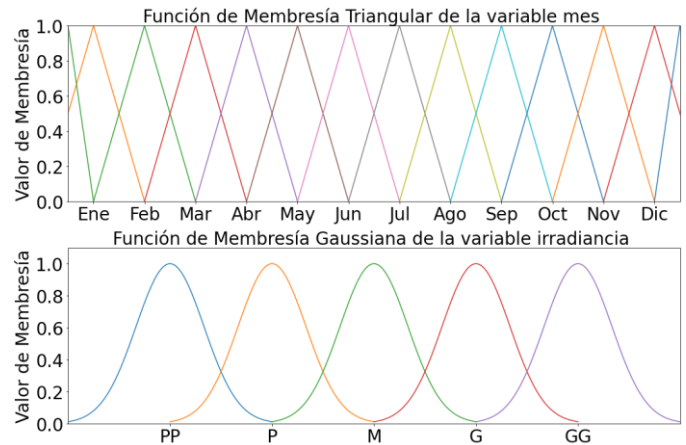
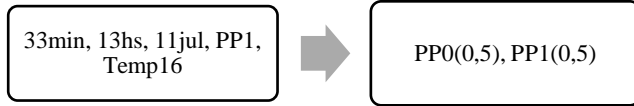


Fig. 3. Ejemplo de las particiones con Función de Membresía triangular y Gaussiana de las variables. Fuente: Elaboración propia.

Está representado (Fig. 4) el proceso de generación de reglas de los modelos WEIGHTED-FTS (Orden 1) y FIG-FTS (Orden 2) que tienen el formato Precedente  $\rightarrow$  Consecuente, que vienen de un patrón temporal que indica dos conjuntos difusos que aparecieron en secuencia en la serie temporal difusa creada [38]. Donde el precedente indica un conjunto en el tiempo  $t$  y el consecuente el conjunto que apareció en el tiempo  $t + 1$  [38]. Aquí (Fig.4) se evidenciada la diferencia del orden, cómo funcionan los gránulos de información del método FIG y como se manifiestan el orden de cada uno de los métodos. Recordando que el orden es la cantidad de *lags* (valores pasados) que son usados por los modelos, i.e., cuanta información pasada es necesaria para describir acontecimientos futuros [38]. Se resalta las diferencias entre el método de WEIGHTED-FTS que es de orden 1 y realiza ponderaciones

cronológicas para el resultado y los gránulos de información generados en FIG-FTS siendo este un método de orden superiores, ambos multivariables. Una vez terminado el proceso de aprendizaje de los modelos se realiza el proceso de pronóstico, aquí son utilizados los datos de la base de datos correspondientes a *test data*.

*WEIGHTED-FTS (Orden 1):*



*FIG-FTS (Orden 2):*

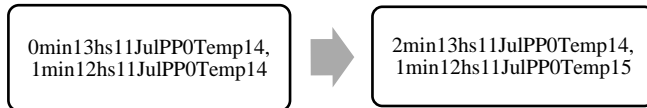


Fig. 4. Resultado del proceso de aprendizaje, generación de reglas para cada uno de los métodos STD utilizados. Fuente: Elaboración propia.

### 3) Filtro de Paso Bajo y Simulación de Potencia

Según [39] la energía fotovoltaica es aproximadamente proporcional a un perfil de irradiancia espacial. Así, una vez obtenido el pronóstico de irradiancia, obtenemos un perfil de irradiancia espacial considerando las especificaciones del SFV que cuenta con 20 módulos de 110 kWp organizados en 4 *strings* de 5 módulos y un Inversor de potencia ABB UNO de 2,5 kW. Y se procede a la aplicación del filtro mediante la función de *butter* de la biblioteca de Python *Scipy*. Finalmente, mediante la aplicación de (11). Se obtiene un valor de potencia simulada para cada uno de los valores según el horizonte de predicción.

TABLA V  
CONFIGURACIÓN EXPERIMENTAL DE LOS MODELOS STD, VARIABLES Y HORIZONTES DE PREDICCIÓN.

Desarrollo del modelo	WEIGHTED-FTS	FIG-FTS
Horizontes de predicción (minutos)	5,15 y 30	5,15 y 30
Entrada	Data, Temperatura ambiente, GHI, Potencia generada, temperatura de contacto	Data, Temperatura ambiente, GHI, Potencia generada, temperatura de contacto
Hiperparámetros de los modelos para el aprendizaje	Primera orden ( $\Omega=1$ )	Segunda orden ( $\Omega=2$ ) y k-Nearest Neighbor ( $\kappa=2$ )
Salida parcial	GHI	GHI
Parámetros del filtro de paso bajo	Temperatura del módulo, potencia pico instalada del sistema, GHI_STC	Temperatura del módulo, potencia pico instalada del sistema, GHI_STC
Salida	Potencia Modelada	Potencia Modelada

En la Tabla V se reúne la configuración experimental aplicada con cada uno de los métodos, así como las respectivas entradas y salidas.

## IV. RESULTADOS Y DISCUSIÓN

En esta sección se muestran los resultados obtenidos de las predicciones de radiación y de las posteriores simulaciones de potencia. El análisis experimental se realizó con dos métodos de pronósticos WEIGHTED-FTS y FIG FTS, con tres horizontes de predicción de corto plazo de 5, 15 y 30 minutos como mostrado en la Tabla V. En cada método son introducidas como variables exógenas mes, día, hora, minuto y temperatura ambiente ya como variable endógena, la Irradiancia Global Horizontal (GHI). Posteriormente al pronóstico de irradiancia obtenido se le aplica la suavización espacial mediante las especificaciones técnicas y físicas del SFV para obtener la simulación de potencia generada. En el entorno de Google Colab el promedio de entrenamiento varío de 10 minutos a 20 minutos según el horizonte de predicción.

### A. Estadísticas de error

Los experimentos se realizaron a partir de los conceptos y pasos relatados en la Sección II y el método presentado en la Sección III. Los resultados con respecto a errores de predicción de los dos modelos STD son reunidos en la Tabla VI.

En este estudio se utilizó las siguientes métricas estadísticas: *MBE* para cuantificar el sesgo general y detecta si el modelo está produciendo una sobreestimación ( $MBE > 0$ ) o una subestimación ( $MBE < 0$ ) de los datos pronosticados y captura las tendencias generales en lugar de la variabilidad [13]; *RMSE* (*nRMSE*, normalizado) como medida de incertidumbre [3] además es muy sensible a grandes errores individuales y captura la variabilidad en lugar de las tendencias generales [13]; y el Coeficiente de Determinación ( $R^2$ ) que expresa el ajuste de los datos del modelo predicho en relación a los datos originales [20]. Además, el Test de Kolmogorov-Smirnov (KS) que muestra la capacidad para reproducir distribuciones estadísticas [3]. Por fin, se aplicó el Test de KS supone que cuanto mayor sea el número de datos experimentales más cerca estará la distribución modelada de la distribución medida [3].

En la Tabla VI son mostrados los resultados de la estadística de error. En la parte A de la tabla se encuentra el pronóstico de irradiancia, con el intuito de comparar la precisión de los dos métodos STD, ya en la parte B las métricas mostradas son utilizadas para evaluar el uso de STD en una simulación de potencia mediante el método de suavización espacial. Así mismo se señala en las dos partes de la tabla el análisis de los diferentes horizontes temporales en cada uno de los métodos STD utilizados.

Para el test estadístico KS es utilizado un nivel de significación de  $\alpha = 0,05$ . Con el cual se establece un Valor Crítico ( $V_c$ ) para cada uno de los horizontes de pronóstico, esto considerando el tamaño de las muestras que cada uno ofrece [21]. Se comprueba si ambas muestras tienen la misma distribución. Para ello es realizado un contraste de hipótesis, suponiendo que ambas muestras tienen la misma distribución, la hipótesis nula es aceptada, contra la hipótesis alternativa de que son distintas, hipótesis nula rechazada [21]. Del resultado de test de KS tenemos que para  $KS < V_c$  la distribución de frecuencias modeladas es similar a la distribución de frecuencias entonces se acepta la hipótesis nula ( $H_0$ ) y para  $KS > V_c$  se tiene la hipótesis alternativa ( $H_1$ ) donde se establece que la distribución de frecuencias medida no es consistente con



la distribución modelada, i.e., un mal ajuste de los datos pronosticados [3]. La hipótesis alternativa equivale al rechazo de la hipótesis nula.

Se observa en la Tabla VII el test de normalidad aplicado a todas las muestras obtenidas tanto en el pronóstico de GHI como la posterior simulación de potencia. La prueba KS determina si dos conjuntos de datos difieren significativamente [21]. Con esta prueba se compara la distribución de un conjunto de datos pronosticados con una distribución de datos observados y se evalúan las diferencias [21].

TABLA VI  
MÉTRICAS ESTADÍSTICAS DE ERRORES DE PREDICCIÓN

Método STD	Horizonte de Predicción	MBE	RMSE	$n$ RMSE	$R^2$
Parte A - Pronóstico de Irradiancia con los métodos STD					
WEIGHTED-FTS	5 min	4,68	36,32	0,08	0,97
	15 min	3,97	36,59	0,25	0,968
	30 min	7,05	43,98	0,61	0,96
FIG-FTS	5 min	3,67	28,14	0,06	0,98
	15 min	1,82	36,27	0,25	0,97
	30 min	-1,98	53,19	0,74	0,94
Parte B - Simulación de Potencia a Partir de Pronóstico de Irradiancia Aplicando Filtro de Paso Bajo					
WEIGHTED-FTS	5 min	7,57	143,70	0,33	0,93
	15 min	11,31	160,24	1,11	0,91
	30 min	26,63	197,78	2,76	0,87
FIG-FTS	5 min	5,54	139,12	0,32	0,93
	15 min	6,64	134,40	0,93	0,94
	30 min	6,76	177,75	2,48	0,90

### B. Análisis y discusión

En esta sección son presentados los análisis estadísticos y las comparaciones de los errores de predicción para los modelos de STD aplicados en este trabajo, tanto para el pronóstico de irradiancia solar cuanto para la simulación de la potencia generada.

Inicialmente, como fue señalado en los objetivos específicos de este trabajo, se evalúa y compara la precisión del pronóstico de Irradiancia Global Horizontal de dos métodos STD. El método FIG-FTS manifiesta mejores resultados en el horizonte temporal de 5 minutos y 15 minutos, ya para el del 30 min el método de WEIGHTED-FTS tiene mejores resultados como se ve en la Tabla VI. El método FIG-FTS de orden superior tiene un desempeño mejor cuando se trabaja con menores escalas temporales, esto dado que el FIG-FTS realiza previsión multivariable, actuando como un método de entrada/salida múltiple (MIMO), en el que todas las variables son a la vez objetivos y variables explicativas [22]. Por esto, se considera que para realizar previsiones con métodos multivariables es necesario contar con una base de datos con redundancia mínima y lo más detallada posible y que la utilización de métodos de orden 1 como el WEIGHTED-FTS es preferible cuando la base

de datos está más tiene una escala temporal mayor, lo que implica un menor número de datos.

La prueba KS determina si dos conjuntos de datos difieren significativamente [21]. Entonces, como es detallado en la Tabla VII, los valores de KS son menores para el método FIG-FTS en los horizontes de tiempo de 5 y 15 minutos, i.e., que dicho método tiene más capacidad de reproducir las distribuciones de frecuencia observadas con escalas temporales menores. La Tabla VII muestra como la simulación de potencia no tiene la capacidad de reproducir las distribuciones estadísticas normales de los datos por lo que en la tabla se muestra el rechazo de la hipótesis nula.

El resultado del pronóstico aquí mostrado (Fig. 5) ejemplifica tres días de pronósticos con los datos de test relativos a los días 01, 02 y 03/05/2019 para los dos métodos de STD y los 3 horizontes de predicción empleados.

La comparación de la precisión de los dos métodos de STD utilizados para el pronóstico de Irradiancia Global Horizontal por horizonte de predicción es aquí presentado (Fig. 6) donde, de manera gráfica, es tratado el Coeficiente de Determinación ( $R^2$ ). Esto cuando se trata del pronóstico de irradiancia mediante las STD. Con la comparación aquí hecha (Fig. 6) se hace visible el mejor desempeño del método FIG-FTS en el horizonte temporal de 5 y 15 minutos al compararlo con el método WEIGHTED-FTS, el cual resulta en un mejor  $R^2$  cuando el modelo realiza previsiones en el horizonte temporal de 30 minutos.

TABLA VII  
RESULTADO DEL TEST DE KOLMOGÓROV-SMIRNOV (KS)

Test de Normalidad - KS	Vc	KS	Resultado
Parte A - Pronóstico de Irradiancia con los métodos STD			
WEIGHTED-FTS / 5 min	0,1979	0,1195	H0 Aceptada
WEIGHTED-FTS / 15 min	0,2381	0,2325	H0 Aceptada
WEIGHTED-FTS / 30 min	0,3368	0,2646	H0 Aceptada
FIG-FTS / 5 min	0,1979	0,1057	H0 Aceptada
FIG-FTS / 15 min	0,2381	0,1878	H0 Aceptada
FIG-FTS / 30 min	0,3368	0,2464	H0 Aceptada
Parte B - Simulación de Potencia a Partir de Pronóstico de Irradiancia Aplicando Filtro de Paso Bajo			
WEIGHTED-FTS / 5 min	0,1979	0,3093	H0 Rechazada
WEIGHTED-FTS / 15 min	0,2381	0,3498	H0 Rechazada
WEIGHTED-FTS / 30 min	0,3368	0,4786	H0 Rechazada
FIG-FTS / 5 min	0,1979	0,2768	H0 Rechazada
FIG-FTS / 15 min	0,2381	0,3146	H0 Rechazada
FIG-FTS / 30 min	0,3368	0,4082	H0 Rechazada

Se expone ahora, la evaluación del uso de STD en una simulación de potencia mediante el método de suavización espacial [9], presentado en la Sección II en la etapa f de posprocesamiento del pronóstico de GHI con las STD.

La potencia fue calculada a partir de los pronósticos de irradiancia que fueron realizados con los métodos WEIGHTED-FTS y FIG-FTS. Las métricas de error correspondientes a este parte del proceso que es presentada en este trabajo esta mostradas de igual manera en la Table VI parte

b. Se observa que los resultados de potencia obtenidos a partir de los pronósticos realizados con el método FIG-FTS de orden 2 muestran mejor desempeño que los resultados obtenidos con el método de orden 1. Esto se observa aquí (Fig. 8) donde se compara el valor de  $R^2$ . Sin embargo, es necesario señalar que el uso de un orden de modelo mayor no implica un aumento de su desempeño. Ya que entre mayor el orden más conjuntos difusos serán generados demasiados conjuntos difusos generan un sobreajuste, haciendo que el modelo empiece a aprender el ruido de los datos, análogamente, conjuntos con orden menor generan un subajuste, debido a la simplificación excesiva de la señal [38]. Para la elección adecuada del orden es posible realizar procesos de optimización de hiperparámetros como señalado en [32].

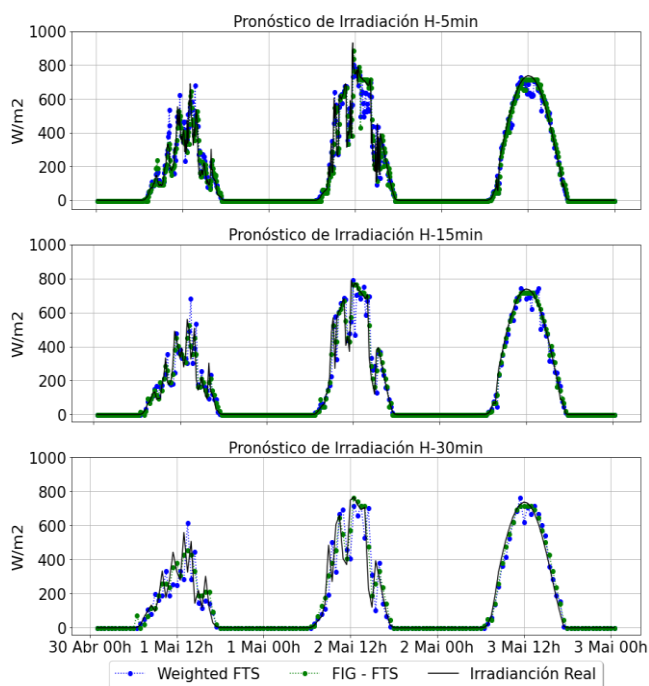


Fig. 5. Ejemplo de predicción de GHI utilizando los métodos STD con 3 horizontes temporales. Fuente: Elaboración propia.

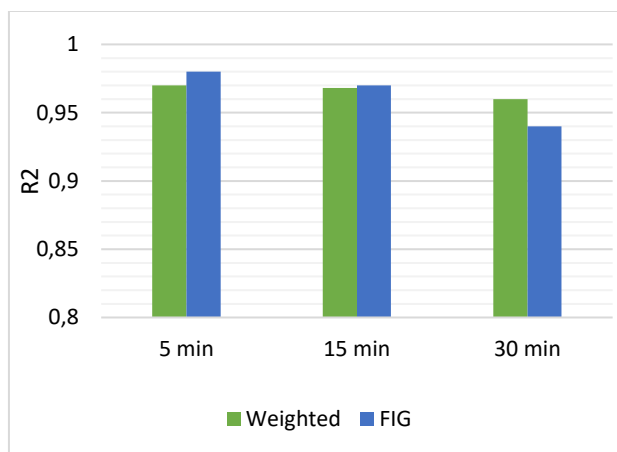


Fig. 6. Comparación del coeficiente de determinación para los métodos implementados y comparación según el horizonte temporal para el pronóstico de GHI. Fuente: Elaboración propia.

Se observa, también, en la Tabla VI en la parte B, que los valores de RMSE y MBE son mayores que los presentados en los pronósticos de GHI. Esto se explica dado que además del error que se presenta en la realización del pronóstico de GHI se adiciona el error que presenta en la simulación de potencia, se considera aquí también el factor de corrección con la temperatura de contacto que disminuye la eficiencia del módulo fotovoltaico, como se muestra en (11). Un ejemplo de la simulación de potencia es mostrado aquí (Fig. 7) donde aparecen el valor de GHI obtenido con el método FIG-FTS a partir del cual se realiza la simulación de la potencia, se muestra también el valor de potencia real observado, con el cual se comparan los resultados.

Como ya fue dicho a partir del pronóstico de GHI se hace una simulación de los valores de potencia generada, utilizando el método de suavización espacial y posterior paso por el filtro de paso bajo de primera orden [9]. Un ejemplo de la simulación de potencia es mostrado aquí (Fig. 7), donde se simula a partir del método FIG bajo el horizonte temporal de 5 minutos 3 días de potencia generada.

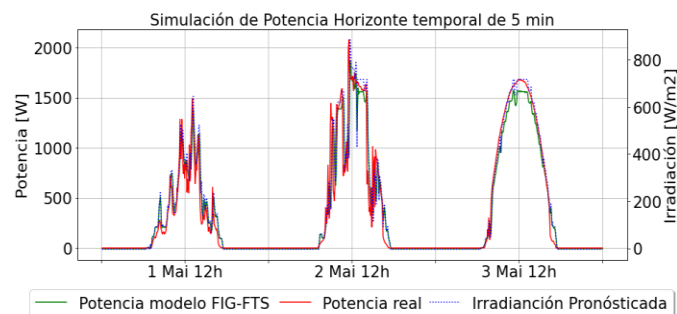


Fig. 7. Simulación de potencia para el horizonte de 5 minutos con el método FIG-FTS. Fuente: Elaboración propia

Las métricas de error correspondientes a este parte del proceso que es presentada en este trabajo son mostradas de igual manera en la Tabla VI. Se muestra la tendencia de mejores resultados para el método FIG y el horizonte temporal de 5 minutos cuando analizados el MBE y el RMSE. Se obtienen estos mismos en el análisis de coeficiente de determinación, como mostrado aquí (Fig. 8).

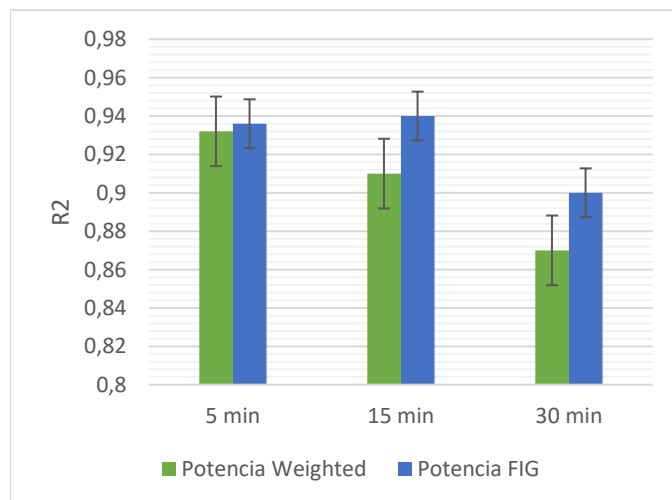


Fig. 8. Comparación del coeficiente de determinación para la simulación de potencia. Fuente: Elaboración propia

Finalmente se resalta, como se ve en la Tabla VI para los pronósticos de irradiancia a medida que se disminuye la escala temporal se manifiesta un aumento de la precisión del método. Vemos que las métricas de RMSE y MBE arrojan mejores resultados en el horizonte de 5 minutos comparados con el de 15 y 30 minutos. Estos resultados son esperados, ya que el error de previsión medido por el nRMSE aumenta con el aumento de la escala temporal de la previsión.

## V. CONCLUSIONES

En este trabajo se realizó el desarrollo y aplicación de dos métodos multivariantes de STD para la predicción indirecta de generación de energía solar fotovoltaica, a través de horizontes temporales de corto plazo. La aplicación de la metodología de las STD sumado a la suavización espacial para la simulación de potencia permiten una configuración experimental controlada, posibilitando el acompañamiento de todo el proceso, fundamentalmente de cómo ocurre el aprendizaje de cada uno de los modelos y la creación de las reglas difusas como mostradas aquí (Fig. 4).

A cada uno de los métodos utilizados, tanto WEIGHTED-FTS como FIG-FTS, se le asignó como variable endógena la GHI y como variables exógenas minuto, hora, día, fecha y temperatura ambiente trabajadas a través de los horizontes temporales de pronóstico de 5, 15 y 30 minutos. A pesar de que existen trabajos donde son aplicados los métodos STD utilizados, una contribución de este estudio es haber empleado el método de suavización espacial con la aplicación de un filtro de paso bajo, sumado al método como posprocesamiento de los datos, para simular valores de potencia donde son consideradas las características particulares de SFV. Para fines de análisis fueron utilizados datos obtenidos del SFV del 2,2 kWp de la estación solarimétrica instalada en el laboratorio de Fotovoltaica de la Universidad Federal de Santa Catarina - UFSC, localizada en la ciudad de Florianópolis en Brasil.

- i. En el análisis de ambos métodos, se encontró que el método FIG-FTS, de orden superior, presenta mejores resultados en el pronóstico de GHI a través de los horizontes de 5 y 15 minutos lo cual se percibe en los resultados estadísticos de la Tabla VI y en el análisis de la estadística de KS representando la capacidad del modelo de reproducir la función de distribución acumulativa de los datos observado. Sin embargo, los valores de GHI pronosticados con el método FIG en el horizonte de 30 minutos (el más amplio de los horizontes trabajados), presenta subestimación, i.e.,  $MBE < 0$  de los valores pronosticados. En el horizonte de tiempo de 30 min, las mejores estadísticas son observadas al aplicar el método WEIGHTED-FTS como se puede ver aquí (Fig. 6). Lo cual indica que para horizontes de predicción más amplios se muestran más indicados métodos de primera orden.
- ii. Una vez realizada la simulación de potencia con el filtro de paso bajo, considerando las especificaciones del SFV, se evidenció que el acoplamiento del método de FIG resulta en mejores índices estadísticos. Se percibe también por las Tabla VII que horizontes de tiempo más cortos, como de 5 minutos, mejoran los valores

obtenidos de las estadísticas como el RMSE y los valores del test de KS.

- iii. El análisis estadístico da cuenta de valores mayores de error, así como la incapacidad de reproducir la distribución estadística de las muestras en los datos obtenidos en la simulación de potencia. Esto dado que la aplicación del filtro de paso bajo, al suavizar los picos rápidos de irradiancia, elimina también valores de irradiancia con grandes variaciones típicas de este tipo de medidas.
- iv. A pesar de que el método de orden superior tiene mejores resultados es necesario resaltar que el aumento del valor de orden del método indiscriminadamente no implica un aumento de su desempeño. Ya que entre mayor el orden más conjuntos difusos serán generados y demasiados conjuntos difusos generan un sobreajuste, haciendo que el modelo empiece a aprender el ruido de los datos, análogamente, conjuntos con orden menor generan un subajuste, debido a la simplificación excesiva de la señal [38].
- v. Se sugiere, con base en los resultados, que el uso de las Series de Tiempos Difusas (STD) proporcionan resultados con buena precisión. Se resalta que la implementación del STD a través de la biblioteca de Python, pyFTS, es un proceso confiable ya que la misma permite el acceso a su código fuente. Esta biblioteca es una buena herramienta para abordar el proceso de pronóstico a través de la Lógica Difusa. Se resalta también que la misma ha sido objeto de mejoras que aumentan su rendimiento tal como la inclusión de modelos de predicción de intervalos y optimización de hiperparámetros.
- vi. También es importante resaltar que el proceso de pronóstico con STD implica la utilización de base de datos que cuenten con una buena integridad de los datos por eso es importante el preprocesamiento de los mismos. La base de datos que fue cedida por el laboratorio de Fotovoltaica de la Universidad Federal de Santa Catarina, fue de vital importancia.

Finalmente, se resalta otra contribución relativa a la originalidad de este estudio es que los resultados de predicción tienen dos enfoques: directo e indirecto, siendo que este abordaje no fue observado en la literatura [14], [40] sobre predicción de energía solar fotovoltaica. Además, este estudio permitió evaluar dos métodos de STD con configuraciones experimentales distintas y evaluar su rendimiento al usarlo junto con la suavización espacial aplicada con el filtro de paso bajo, lo cual demostró ser una buena alternativa. Futuramente se propone incluir la predicción con optimización de hiperparámetros [32] así como la realización de pronósticos probabilísticos [19] que tengan más en consideración las componentes estacionales de las variables de entrada.

## VI. REFERENCIAS

- [1] "IRENA – International Renewable Energy Agency." [Online]. Available: <https://www.irena.org/>. [Accessed: 26-Nov-2020].
- [2] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere, "Photovoltaic and Solar Forecasting : State

- of the Art,” no. October, p. 40, 2013.
- [3] Jan Kleissl, *Solar Energy Forecasting and Resource Assessment*. 2013.
- [4] C. A. Severiano, P. C. L. Silva, H. J. Sadaei, and F. G. Guimarães, “Very short-term solar forecasting using fuzzy time series,” *IEEE Int. Conf. Fuzzy Syst.*, no. July, 2017.
- [5] Q. Song, “Seasonal forecasting in fuzzy time series,” *Fuzzy Sets Syst.*, vol. 107, no. 2, pp. 235–236, 1999.
- [6] Z. Ziadi, S. Taira, M. Oshiro, and T. Funabashi, “Optimal power scheduling for smart grids considering controllable loads and high penetration of photovoltaic generation,” *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2350–2359, 2014.
- [7] J. N. Maciel, J. Javier Gimenez Ledesma, and O. Hideo Ando Junior, “Forecasting Solar Power Output Generation: A Systematic Review with the Proknow-C,” *IEEE Lat. Am. Trans.*, vol. 19, no. 4, pp. 612–624, 2021.
- [8] “Fotovoltaica | Grupo de Pesquisa Estratégica em Energia Solar Fotovoltaica.” [Online]. Available: <https://fotovoltaica.ufsc.br/sistemas/fotov/>. [Accessed: 30-Aug-2021].
- [9] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, “From irradiance to output power fluctuations: The PV plant as a low pass filter,” *Prog. Photovoltaics Res. Appl.*, vol. 19, no. 5, pp. 505–510, 2011.
- [10] Q. Song and B. S. Chissom, “Fuzzy time series and its models,” *Fuzzy Sets Syst.*, vol. 54, no. 3, pp. 269–277, 1993.
- [11] MINDS, “pyFTS Quick Start — pyFTS 1.6 documentation.” [Online]. Available: <https://pyfts.github.io/pyFTS/build/html/quickstart.html#what-are-fuzzy-time-series-fts>. [Accessed: 06-Sep-2020].
- [12] “MINDS.” [Online]. Available: <https://minds.eng.ufmg.br/>. [Accessed: 06-Sep-2020].
- [13] R. Blaga, A. Sabadus, N. Stefu, C. Dughir, M. Paulescu, and V. Badescu, “A current perspective on the accuracy of incoming solar energy forecasting,” *Prog. Energy Combust. Sci.*, vol. 70, pp. 119–144, 2019.
- [14] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, “Review of photovoltaic power forecasting,” *Sol. Energy*, vol. 136, pp. 78–111, 2016.
- [15] Tte. Javier Jaraíz Corrales, “Trabajo Fin de Grado,” *Zaguan.Unizar.Es*, pp. 5, 6 y 9, 2019.
- [16] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, “Photovoltaic and solar power forecasting for smart grid energy management,” *CSEE J. Power Energy Syst.*, vol. 1, no. 4, pp. 38–46, 2016.
- [17] J. Mubiru, “Predicting total solar irradiation values using artificial neural networks,” *Renew. Energy*, vol. 33, no. 10, pp. 2329–2332, 2008.
- [18] E. P. Klement, “Fuzzy Logic in Arti cial Intelligence CD { Technical Report 94 / 67 Technische Universitat Wien Institut fur Informationssysteme Abteilung fur Datenbanken und Expertensysteme,” no. July, 1997.
- [19] P. C. L. Silva, M. A. Alves, C. A. S. Junior, G. L. Vieira, F. G. Guimarães, and H. J. Sadaei, “Probabilistic Forecasting with Seasonal Ensemble Fuzzy Time-Series,” pp. 1–12, 2019.
- [20] C. A. Gueymard, “A review of validation methodologies and statistical performance indicators for modeled solar radiation data : Towards a better bankability of solar projects,” *Renew. Sustain. Energy Rev.*, vol. 39, pp. 1024–1034, 2014.
- [21] B. Espinar *et al.*, “Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations,” *Sol. Energy*, vol. 83, no. 1, pp. 118–125, 2009.
- [22] P. C. de L. e Silva, “Scalable Models For Probabilistic Forecasting With Fuzzy Time Series,” p. 187, 2019.
- [23] S. M. Chen, “Forecasting enrollments based on fuzzy time series,” *Fuzzy Sets Syst.*, vol. 81, no. 3, pp. 311–319, 1996.
- [24] G. Silva, Petrônio Cândido de Lima, Severiano Jr., Carlos Alberto, Marcos Antônio, Alves, Miri Weiss, Cohen, Federico Gadelha, “A New Granular Approach for Multivariate Forecasting.” Springer Nature Switzerland AG 2019, 2019.
- [25] H. K. Yu, “Weighted fuzzy time series models for TAIEX forecasting,” *Phys. A Stat. Mech. its Appl.*, vol. 349, no. 3–4, pp. 609–624, 2005.
- [26] J. Schnabel and S. Valkealahti, “Energy Storage Requirements for PV Power Ramp Rate Control in Northern Europe,” *Int. J. Photoenergy*, vol. 2016, 2016.
- [27] V. N. S. Diaz, “Avaliação de Desempenho das Estratégias de Controle para Suavização de Potência Ativa de Sistemas Fotovoltaicos com Armazenamento de Energia,” pp. 1–36, 2019.
- [28] A. Chugh, P. Chaudhary, and M. Rizwan, “Fuzzy logic approach for short term solar energy forecasting,” *12th IEEE Int. Conf. Electron. Energy, Environ. Commun. Comput. Control (E3-C3), INDICON 2015*, 2016.
- [29] T. Hong and P. Wang, “Fuzzy interaction regression for short term load forecasting,” *Fuzzy Optim. Decis. Mak.*, vol. 13, no. 1, pp. 91–103, 2014.
- [30] A. H. Etemadi, E. J. Davison, and R. Iravani, “A decentralized robust control strategy for multi-DER microgridspart II: Performance evaluation,” *IEEE Trans. Power Deliv.*, 2012.
- [31] P. C. de Lima e Silva, C. A. Severiano, M. A. Alves, R. Silva, M. Weiss Cohen, and F. G. Guimarães, “Forecasting in non-stationary environments with fuzzy time series,” *Appl. Soft Comput. J.*, vol. 97, p. 106825, 2020.
- [32] P. de Oliveira e Lucas, P. Silva, and F. G. Guimarães, “Otimização Evolutiva de Hiperparâmetros para Modelos de Séries Temporais Nebulosas,” no. October, 2019.
- [33] H. J. Sadaei, P. C. de Lima e Silva, F. G. Guimarães, and M. H. Lee, “Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series,” *Energy*, vol. 175, pp. 365–377, 2019.
- [34] B. Sivaneasan, C. Y. Yu, and K. P. Goh, “Solar Forecasting using ANN with Fuzzy Logic Pre-processing,” *Energy Procedia*, vol. 143, pp. 727–732,

- 2017.
- [35] E. Bisong, “Building Machine Learning and Deep Learning Models on Google Cloud Platform,” *Build. Mach. Learn. Deep Learn. Model. Google Cloud Platf.*, 2019.
  - [36] L. De Oliveira, S. Ufc, and P. C. M. C. Ufc, “Modelos Para a Temperatura De Operação De Módulos Fotovoltaicos : Uma Revisão Das Correlações E Variáveis Pertinentes Modelos Para a Temperatura De Operação De Módulos Fotovoltaicos : Uma Revisão Das Correlações E Variáveis.”
  - [37] J. N. Maciel, V. H. Wentz, J. J. G. Ledesma, and O. H. Ando Junior, “Analysis of Artificial Neural Networks for Forecasting Photovoltaic Energy Generation with Solar Irradiance,” *Brazilian Arch. Biol. Technol.*, vol. 64, no. spe, pp. 1–14, 2021.
  - [38] P. C. de L. Silva, “A short tutorial on Fuzzy Time Series — Part II | by Petrônio Silva | Towards Data Science.” [Online]. Available: <https://towardsdatascience.com/a-short-tutorial-on-fuzzy-time-series-part-ii-with-an-case-study-on-solar-energy-bda362ecca6d>. [Accessed: 07-Sep-2020].
  - [39] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, “Power output fluctuations in large scale pv plants: One year observations with one second resolution and a derived analytic model,” *Prog. Photovoltaics Res. Appl.*, vol. 19, no. 2, pp. 218–227, 2011.
  - [40] P. Li, K. Zhou, and S. Yang, “Photovoltaic Power Forecasting: Models and Methods,” in *2nd IEEE Conference on Energy Internet and Energy System Integration, EI2 2018 - Proceedings*, 2018.