	GESTIÓN DE SERVICIOS ACADÉMICOS Y BIBLIOTECARIOS		CÓDIGO	FO-GS-15
	ESQUEMA HOJA DE RESUMEN		VERSIÓN	02
			FECHA	21/05/2020
			PÁGINA	1 de 90
ELABORÓ		REVISÓ	APROBÓ	
Jefe División de Biblioteca		Equipo Operativo de Calidad	Líder de Calidad	

RESUMEN TRABAJO DE GRADO

AUTOR:

NOMBRE(S): DIANA CAROLINA

APELLIDOS: BASTOS GUERRERO

FACULTAD: INGENIERÍA

PLAN DE ESTUDIOS: INGENIERÍA ELECTRÓNICA

DIRECTOR:

NOMBRE(S): SERGIO BASILIO

APELLIDOS: SEPÚLVEDA MORA

NOMBRE(S): MARIO JOAQUÍN

APELLIDOS: ILLERA BUSTOS

TÍTULO DEL TRABAJO (TESIS): ANÁLISIS Y ESTIMACIÓN DE LA RADIACIÓN SOLAR EN LA CIUDAD DE CÚCUTA USANDO MÉTODOS DE ESTADÍSTICA INFERENCIAL Y TÉCNICAS DE INTELIGENCIA ARTIFICIAL.

Debido al rápido incremento en la demanda de electricidad y la disminución de los recursos energéticos provenientes de fuentes no renovables, las personas han comenzado a considerar la idea de consumir energía eléctrica a partir de recursos renovables. La energía solar fotovoltaica es la fuente dominante de las energías renovables y, por lo tanto, la radiación solar que incide en un territorio es un factor muy importante para dimensionar la producción energética. En esta investigación, se presenta un análisis estadístico de la radiación solar en la ciudad de Cúcuta, Norte de Santander, Colombia; con el objetivo de brindar una descripción detallada de su variabilidad entre los años 2005 y 2015, y se emplea un sistema adaptativo de inferencia neuro-difusa (ANFIS) para estimar la radiación solar global. Los datos históricos de la base de datos del IDEAM con variables de entrada tales como (temperatura ambiente, brillo solar, velocidad del viento, humedad relativa y hora de medición) y de salida (radiación solar global solar) fueron usados para el entrenamiento y chequeo de la red. El modelo presenta un $R^2=0.911$ y un $RMSE= 124,23Wh/m^2$, evidenciando que presenta una precisión significativamente buena para la estimación de la radiación solar global en el territorio seleccionado.

PALABRAS CLAVES: ANOVA, ANFIS, RADIACIÓN SOLAR GLOBAL, MODELO DE ESTIMACIÓN INDIRECTO.

CARACTERÍSTICAS:

PÁGINAS: 90 PLANOS: 0 ILUSTRACIONES: 40 CD ROOM: 1

ANÁLISIS Y ESTIMACIÓN DE LA RADIACIÓN SOLAR EN LA CIUDAD DE CÚCUTA
USANDO MÉTODOS DE ESTADÍSTICA INFERENCIAL Y TÉCNICAS DE
INTELIGENCIA ARTIFICIAL

DIANA CAROLINA BASTOS GUERRERO

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER

FACULTAD DE INGENIERÍA

INGENIERÍA ELECTRÓNICA

SAN JOSÉ DE CÚCUTA

2020

ANÁLISIS Y ESTIMACIÓN DE LA RADIACIÓN SOLAR EN LA CIUDAD DE CÚCUTA
USANDO MÉTODOS DE ESTADÍSTICA INFERENCIAL Y TÉCNICAS DE
INTELIGENCIA ARTIFICIAL

DIANA CAROLINA BASTOS GUERRERO

Trabajo de grado presentado como requisito para optar por el título de:

INGENIERO ELECTRÓNICO

Director:

SERGIO BASILIO SEPÚLVEDA MORA

Ing. M.Sc. Ingeniería Electrónica

Codirector:

MARIO JOAQUIN ILLERA BUSTOS

Ing. M.Sc. Ingeniería Electrónica

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER

FACULTAD DE INGENIERÍA

INGENIERÍA ELECTRÓNICA

SAN JOSÉ DE CÚCUTA

2020

ACTA DE SUSTENTACIÓN DE UN TRABAJO DE GRADO

Fecha: CÚCUTA, 15 DE MAYO DE 2020
Hora: 16:00
Lugar: MODALIDAD REMOTO SINCRÓNICO
Plan de Estudios: INGENIERÍA ELECTRÓNICA
Título de la Tesis: "ANÁLISIS Y ESTIMACIÓN DE LA RADIACIÓN SOLAR EN LA CIUDAD DE CÚCUTA USANDO MÉTODOS DE ESTADÍSTICA INFERENCIAL Y TÉCNICAS DE INTELIGENCIA ARTIFICIAL"
Jurados: IE, PhD DINAEL GUEVARA IBARRA
IE, MSc SERGIO ALEXANDER CASTRO CASADIEGO
Director: IE, MSc SERGIO BASILIO SEPÚLVEDA MORA
Codirector: IE, MSc MARIO JOAQUIN ILLERA BUSTOS

Nombre del Estudiante	Código	Calificación
DIANA CAROLINA BASTOS GUERRERO	1161169	CINCO, CERO (5,0)

LAUREADA



DINAEL GUEVARA IBARRA



SERGIO A. CASTRO CASADIEGO



DINAEL GUEVARA IBARRA
Coordinador Comité Curricular
Ingeniería Electrónica



Vigilada Mineducación



**CARTA DE AUTORIZACIÓN DE LOS AUTORES PARA
LA CONSULTA, LA REPRODUCCIÓN PARCIAL O TOTAL Y LA PUBLICACIÓN
ELECTRÓNICA DEL TEXTO COMPLETO**

Cúcuta,

Señores,
BIBLIOTECA EDUARDO COTE LEMUS
Ciudad

Cordial saludo:

Yo, Diana Carolina Bastos Guerrero, identificado con la C.C. N° 1090514849, autor de la tesis y/o trabajo de grado titulado Análisis y estimación de la radiación solar en la ciudad de Cúcuta usando métodos de estadística inferencial y técnicas de inteligencia artificial presentado y aprobado en el año 2020 como requisito para optar al título de Ingeniero Electrónico; autorizo a la biblioteca de la Universidad Francisco de Paula Santander, Eduardo Cote Lamus, para que con fines académicos, muestre a la comunidad en general a la producción intelectual de esta institución educativa, a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios pueden consultar el contenido de este trabajo de grado en la página web de la Biblioteca Eduardo Cote Lamus y en las redes de información del país y el exterior, con las cuales tenga convenio la Universidad Francisco de Paula Santander.
- Permita la consulta, la reproducción, a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato CD-ROM o digital desde Internet, Intranet etc.; y en general para cualquier formato conocido o por conocer.

Lo anterior, de conformidad con lo establecido en el artículo 30 de la ley 1982 y el artículo 11 de la decisión andina 351 de 1993, que establece que **“los derechos morales del trabajo son propiedad de los autores”**, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

Diana Carolina Bastos Guerrero
C.C. 1090514849 de Cúcuta

Contenido	Pág
Introducción	12
1. Descripción del problema	14
1.1 Planteamiento del problema	14
1.2 Justificación	17
1.3 Beneficios Tecnológicos	18
1.4 Beneficios Económicos	19
1.5 Beneficios Sociales	19
1.6 Beneficios institucionales y empresariales	19
1.7 Objetivos	20
1.7.1 Objetivo general	20
1.7.2 Objetivos específicos	20
1.8 Delimitaciones	20
2. Marco Referencial	22
2.1 Antecedentes	22
2.2 Marco Teórico	24
2.2.1 Energía Solar Fotovoltaica.	24
2.2.2 Radiación Solar.	24
2.2.3 Piranómetro.	27
2.2.4 Análisis estadísticos de la radiación solar	27

2.2.5. Modelos de predicción de radiación solar.	30
2.2.6 Modelos de inteligencia artificial.	32
2.2.6.1 Redes Neuronales Artificiales (RNA)	32
2.2.6.2 Sistemas de inferencia difusa	34
2.2.6.3 Sistema adaptativo de inferencia neuro difusa (ANFIS).	37
3. Metodología	44
3.1 Adquisición de datos de radiación solar	44
3.2 Instrumentación para medición y registro de datos	45
3.3 Análisis estadístico de la radiación solar	49
3.4 Creación de la estructura ANFIS	50
3.4.1 Normalización	51
3.4.2 Creación de las redes ANFIS con diferentes FMs	51
3.4.3 Entrenamiento de la red.	54
3.4.4 Comparación y selección de la red.	55
3.4.5 Ejecución y validación de la red.	56
3.5 Evaluación del desempeño de la red.	56
4. Resultados	58
4.1 Adquisición de datos de radiación solar	58
4.2 Instrumentación para la medición y registro de datos	61
4.3 Análisis estadístico de la radiación solar	64

4.4 Creación de la estructura ANFIS	72
4.5 Evaluación del desempeño de la red	77
5. Conclusiones	82
Recomendaciones	84
Referencias	85

Lista de figuras	Pág
Figura 1. Componentes de la radiación solar global	25
Figura 2. Atlas de la radiación solar mundial como promedio anual	26
Figura 3. Piranómetro CMP6	27
Figura 4. Clasificación de los modelos de predicción de radiación solar	31
Figura 5. Arquitectura de una RNA multicapa	33
Figura 6. Detalle de una neurona en una RNA	33
Figura 7. Diagrama básico de un sistema de inferencia difusa	34
Figura 8. Tipos de sistemas de inferencia difusa	37
Figura 9. Topología de una red adaptable con nodos fijos y adaptables	38
Figura 10. Razonamiento difuso tipo 3	39
Figura 11. Estructura básica de una red ANFIS	40
Figura 12. Tipología de funciones de membresía	41
Figura 13. Grupo de investigación de microelectrónica aplicada y control (GIDMAC)	46
Figura 14. Diagrama representativo de la instrumentación requerida	47
Figura 15. Menú principal del software LoggerNet	48
Figura 16. Diagrama de conexiones del piranómetro y datalogger	49
Figura 17. Caracterización de la red ANFIS	52
Figura 18. Representación gráfica de la función de membresía en Matlab tipo: Triangular	53
Figura 19. Representación gráfica de la FM en Matlab tipo: Trapezoidal	53
Figura 20. Representación gráfica de la FM en Matlab tipo: Campana de Bell	53
Figura 21. Representación gráfica de la FM en Matlab tipo: Campana de Gauss	54
Figura 22. Representación gráfica de la FM en Matlab tipo: Campana de Gauss 2	54

Figura 23. Conteo de datos anuales	58
Figura 24. Conteo de datos del mes de enero	59
Figura 25. Datos medidos por el piranómetro a nivel diario para un día soleado	61
Figura 26. Datos medidos por el piranómetro a nivel diario para un día nublado	61
Figura 27. Resumen de la irradiación solar en el mes de febrero	62
Figura 28. Resumen de la última década de radiación solar global en Cúcuta	65
Figura 29. Resumen de la radiación solar en horas pico.	66
Figura 30. Evaluación de la variabilidad mensual de radiación solar global en el año 2010	69
Figura 31. Resumen de la radiación solar en el primer semestre del año de la última década	70
Figura 32. Resumen de la radiación solar en el segundo semestre del año de la última década	71
Figura 33. Resumen mensual de la radiación solar global	71
Figura 34. Entrenamiento de una red ANFIS con 2 FMs	72
Figura 35. Entrenamiento de una red ANFIS con 3 FMs	72
Figura 36. Modificación de la función de membresía	75
Figura 37. Comparación entre el desempeño del sistema y el conjunto de datos.	76
Figura 38. Relación entre los datos de entrada y salida del análisis cualitativo	76
Figura 39. Datos medidos vs. estimados para el modelo de estimación por hora en Cúcuta	79
Figura 40. Datos Medidos vs. Estimados para el modelo de estimación por día en Cúcuta	80

Lista de tablas

	Pág
Tabla 1. Tabla ANOVA	28
Tabla 2. Modelo matemático de las funciones de membresía	41
Tabla 3. Análisis de similitud de las bases de datos	59
Tabla 4. Resumen de irradiación solar para los meses de enero y febrero del 2020	63
Tabla 5. Análisis de la varianza de un solo factor	64
Tabla 6. Test de Bonferroni	67
Tabla 7. Porcentaje de precipitación en la región Andina, Colombia	68
Tabla 8. Análisis de la varianza de todos los meses del año	69
Tabla 9. Entrenamiento ANFIS con 2 funciones de membresía	73
Tabla 10. Entrenamiento ANFIS con 3 funciones de membresía	74
Tabla 11. Parámetros de consecuencia de la red ANFIS	75
Tabla 12. Evaluación del desempeño de la red ANFIS	77
Tabla 13. Características de modelos ANFIS de la literatura	81

Introducción

El Sol es la principal fuente primaria de luz y calor para el planeta Tierra. Este ente genera un constante flujo de energía, que se enmarca en el concepto de radiación solar, que es la energía radiante proveniente del Sol en forma de ondas electromagnéticas que se propagan en todas las direcciones a través del espacio y que determina la dinámica de los procesos atmosféricos y del clima de la Tierra; con el fin de establecer su impacto sobre la superficie terrestre, la potencia por unidad de área incidente de la radiación solar es definida como irradiación, y es medida en Wh/m^2 según el S.I. [1], [2].

Considerando lo anterior, la energía solar llega a ser una de las principales energías renovables del planeta, debido a que es virtualmente inagotable y posee un bajo impacto contaminante en el medio ambiente; adicionalmente, gracias a su fácil implementación representa una herramienta importante al llevar energía eléctrica a zonas de difícil acceso donde no llega la red de energía eléctrica convencional [3]. De esta forma, el conocimiento de la disponibilidad de la energía solar resulta indispensable para el correcto dimensionamiento de sistemas y tecnologías que la transforman, como por ejemplo, los sistemas fotovoltaicos para la producción de electricidad o sistemas térmicos para generar calor [4].

Por consiguiente, la adecuada caracterización del recurso solar termina generando un correcto dimensionamiento de los sistemas que usan esta fuente de energía para operar; esto representa un mejor comportamiento de estos sistemas y mejoras en la eficiencia de los diseños. Por esta razón, un estudio estadístico de los datos históricos de la radiación solar proporcionará información del comportamiento de la variable en tiempo futuro; asimismo, identificará los factores involucrados en su fluctuación, como una herramienta para definir criterios de su uso en sistemas fotovoltaicos o térmicos en una región específica.

En este estudio se desarrolló un modelo híbrido de inteligencia artificial definido como una red adaptativa de inferencia neuro-difusa (ANFIS) para estimar la radiación solar global en la ciudad de Cúcuta. Para el desarrollo del modelo se usaron datos de la última década de la estación meteorológica automática satelital Universidad Francisco de Paula Santander proporcionados por el IDEAM y se hizo uso de la herramienta computacional Matlab por medio de la Toolbox Fuzzy Logic Designer. El modelo propuesto permite la estimación de radiación solar global para completar la ausencia de los registros en la base de datos del IDEAM.

1. Descripción del problema

1.1 Planteamiento del problema

El cambio climático consiste fundamentalmente en el incremento gradual de la temperatura promedio del planeta, que tiene lugar como consecuencia del crecimiento del efecto invernadero debido a la acción humana en la búsqueda de energía. Este efecto lo ocasionan determinados gases, entre los que se destaca el dióxido de carbono (CO_2), producido principalmente durante la combustión de las fuentes de energía fósiles [5].

El consumo de energía eléctrica en el mundo en su mayoría es proveniente de fuentes energéticas no renovables como el petróleo, el gas natural, el carbón y los combustibles nucleares; los cuales son fuentes energéticas fósiles parcialmente abundantes en el planeta, pero finitas [6].

Este consumo de energía basado en combustibles fósiles está provocando que el efecto invernadero tenga un incremento acelerado afectando las condiciones ambientales, como el aumento de la temperatura de forma indiscriminada y el derretimiento de los polos, tanto en la zona norte como sur del planeta, generando una serie de eventos en efecto dominó que son perjudiciales en todos los ecosistemas vivientes en la actualidad. Esto ha generado en muchos países la necesidad de iniciar una transición hacia el uso de recursos energéticos de carácter renovable y ecológico.

Así es, como el concepto de energías renovables empieza jugar un rol importante en el sector energético de la sociedad. Las energías renovables son aquellas energías que se adquieren de fuentes naturales virtualmente inagotables ya sea por la gran cantidad de energía que tienen como el caso del Sol o porque tienen la capacidad de regenerarse por medios naturales. Este tipo de energías van de la mano con el medio ambiente lo cual, al ser aprovechadas y convertidas con

un correcto uso en energía eléctrica, generan una menor emisión de CO₂, disminuyendo los gases de efecto invernadero proveniente, como se mencionó, de fuentes de energía convencional. De esta manera, si se produce un cambio o diversificación de la matriz energética de la nación, se puede no solo reducir los efectos ambientales de las fuentes de generación tradicionales, sino que la red eléctrica será menos dependiente a eventos climáticos que generan una gran dificultad al sistema eléctrico actual, dominado ampliamente por generación hidroeléctrica. Con el Acuerdo para el Cambio Climático de París en 2015 quedó claro que el desarrollo económico del mundo debe moverse a partir de energías renovables, para dejar atrás la generación energética mediante combustibles fósiles [7].

Colombia está situada en las zonas tropicales cercanas al ecuador, con una gran variedad de climas que dependen de la altitud con respecto al nivel del mar. Un análisis preliminar determinó que Colombia tiene un potencial solar significativo con valores de irradiancia promedio más altos que el promedio mundial (de hasta 6,237 kW/m² en comparación con el promedio mundial de 3,9 kW/m²) y el promedio de los países del hemisferio norte y sur [6]. De esta manera, Colombia se convierte en un lugar ideal para la implementación de sistemas fotovoltaicos (FV) por los altos niveles de radiación solar de su zona geográfica.

En Colombia el sector eléctrico está constituido en un 66.6 % de energía hidráulica, 28.42 % de energía térmica y el porcentaje restante de fuentes de menor capacidad como las energías renovables entre las que se encuentran la energía solar fotovoltaica [8]. Estudios realizados muestran a detalle que Colombia cuenta con un recurso solar (radiación) promedio uniforme durante el año para todo el territorio del orden de 4,5 kWh/m², propicio para ser utilizado en generar y suplir las necesidades de energía eléctrica por medio de sistemas fotovoltaicos [9].

Estadísticas de la Unidad de Planeación Minero Energética (UPME) evidencian que, de las iniciativas radicadas para la implementación de nuevas fuentes de energía eléctrica con energías renovables, el 88,3% tienen que ver con energía solar, en donde 9 de cada 10 propuestas para generar energía, usarán paneles solares. Lo que representa un crecimiento del 32% de julio a noviembre de 2017 [6]. Esto indica que en el país es factible generar energía eléctrica por medio de sistemas fotovoltaicos.

Para el diseño e implementación de sistemas FV, es de suma importancia un conocimiento detallado y preciso de la radiación solar global sobre una superficie horizontal en una ubicación geográfica particular [10]. La radiación solar diaria es considerada el parámetro más relevante en meteorología, conversión solar y aplicaciones de energías renovables, especialmente para el dimensionamiento de sistemas FV autónomos [11]. El desarrollo de esta tecnología depende de la cuantificación de la radiación solar global en las zonas geográficas específicas en donde se planea la instalación de sistemas fotovoltaicos.

El instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) cuenta con 4451 estaciones meteorológicas a lo largo de Colombia. Solo 83 de éstas corresponden a estaciones automáticas satelitales que registran la radiación solar global [12].

Debido a que no se cuenta con la suficiente cobertura de la información de la radiación solar global para un correcto dimensionamiento de sistemas FV, es necesario diseñar modelos que utilicen técnicas de estimación con el objetivo de cuantificar esta variable en cualquier lugar específico de la geografía colombiana; lo anterior, a partir de otras variables meteorológicas y de datos históricos de radiación solar medidos en los sitios donde se encuentran disponibles, brindados por estaciones satelitales intercontinentales o por el IDEAM. De acuerdo con lo planteado anteriormente surge la siguiente pregunta de investigación:

¿Cómo analizar y estimar el comportamiento de la radiación solar en la ciudad de Cúcuta usando técnicas de estadística inferencial e inteligencia artificial como una fuente de información alternativa para esta variable?

1.2 Justificación

El acuerdo del cambio climático de París del 2015 promueve implementar las energías renovables como principal fuente energética, en donde la Energía Solar Fotovoltaica es la Fuente No Convencional de Energía Renovable (FNCER) con mayor potencial y proyección en el país. Teniendo en cuenta que el IDEAM no brinda la suficiente cobertura de información de la variable de radiación solar global, la UPME menciona en un reporte que hay una escasa caracterización del potencial del recurso solar en el país [6].

La radiación solar es uno de los parámetros más importantes para el desarrollo de aplicaciones e investigaciones relacionadas a energías renovables. Sin embargo, la adquisición de mediciones de radiación solar no siempre es posible por diferentes motivos (falta de estaciones meteorológicas) y es necesario contar con modelos que permitan estimarla.

Con el desarrollo de esta investigación se buscó satisfacer esa necesidad en la ciudad de San José de Cúcuta; se desarrolló un análisis estadístico del comportamiento de la variable radiación solar global identificando su variación en los últimos diez años, y se diseñó un modelo de estimación de la radiación solar utilizando técnicas de inteligencia artificial (ANFIS).

Con la obtención del modelo de inteligencia artificial se logra estimar la radiación solar a partir de otras variables meteorológicas que, de estar disponibles, permiten el diseño de futuros sistemas fotovoltaicos en zonas urbanas y en zonas aisladas con difícil acceso a la red eléctrica.

La implementación y aprovechamiento de estos nuevos sistemas tienen una gran acogida por ser considerados como una fuente limpia, y lo más importante, permiten el suministro de energía

disminuyendo los gases del efecto invernadero al no producir polución ni emisiones de dióxido de carbono, sin mencionar que también reducirían la alta dependencia a las fuentes hidrológicas de la nación. Con una inversión a largo plazo, el consumo de energía eléctrica a partir de fuentes renovables resulta ser menos costoso que comprarla de la compañía eléctrica.

Este tipo de proyecto se propuso como una estrategia para incrementar las fuentes de información del recurso solar en la región garantizando una alta precisión y confiabilidad de estos datos.

Una estimación precisa de la radiación solar permitirá el diseño y futura instalación de sistemas FV, dando soporte técnico–económico y asegurando la factibilidad de proyectos a ejecutar. Para el caso específico de Cúcuta, este tipo de modelos podrían ayudar a actualizar la base datos del IDEAM que tiene información hasta el 2015, y podría suministrar información perdida en años anteriores, teniendo en cuenta que las variables que se usan como entradas al modelo de estimación, sí están actualizadas en los periodos de tiempos en los que la variable irradiación solar no.

Finalmente, la realización de este proyecto complementa la formación académica y científica de la investigadora. Esto trae consigo una serie de beneficios que serán descritos a continuación. El desarrollo del proyecto fortalece la línea de investigación microelectrónica aplicada a sistemas de energía renovable, que pertenece al Grupo de Investigación y Desarrollo en Microelectrónica Aplicada y Control (GIDMAC).

1.3 Beneficios Tecnológicos

Al implementar este tipo de modelos de estimación utilizando técnicas de inteligencia artificial se puede llegar a actualizar la base de datos del IDEAM, proporcionando la información faltante de la radiación solar a partir de variables de entrada que sí están actualizadas entre el

periodo del 2005 al 2015, proporcionando alta precisión en el diseño e instalación de sistemas fotovoltaicos teniendo una buena estimación de la radiación solar.

1.4 Beneficios Económicos

Al usar un modelo basado en técnicas de inteligencia artificial se logró tener precisión con la obtención de los valores de la radiación solar como modelo de estimación, esto da partida a crear diseños de sistemas fotovoltaicos óptimos en el momento de hacer los análisis financieros, la toma de decisiones en la compra de los elementos y se podría obtener un retorno de inversión en un menor lapso de tiempo.

1.5 Beneficios Sociales

Al utilizar fuentes energéticas limpias para la creación de energía eléctrica, en este caso específico de energía solar fotovoltaica, se incentiva a la comunidad a generar un cambio en el consumo energético y concientizando al usuario que el consumo de energía eléctrica proveniente de fuentes fósiles genera un incremento en los gases de efecto invernadero afectando al medio ambiente.

La implementación de este tipo de modelos utilizando la inteligencia artificial para la estimación de la radiación solar a partir de otras variables de fácil acceso puede ser adecuado e implementado en zonas de difícil acceso de la red eléctrica proporcionando desarrollo sostenible y empleo local.

1.6 Beneficios institucionales y empresariales

Desde el ámbito institucional de la Universidad Francisco de Paula Santander, el desarrollo de proyectos permite fortalecer el grupo de investigación, que a su vez aumenta los indicadores de investigación para fines de re-acreditación de alta calidad del programa de Ingeniería Electrónica.

Desde el ámbito empresarial, las entidades que trabajan en el campo de las energías renovables, más específicamente con sistemas fotovoltaicos, pueden ver como una buena alternativa la estimación de la radiación solar sobre la ciudad, lo que les permite contar con un estudio más preciso a la hora de diseñar nuevos sistemas fotovoltaicos en la región y optimizar así sus propios productos para lograr un avance tecnológico en la región.

1.7 Objetivos

1.7.1 Objetivo general.

Analizar la variabilidad histórica de la radiación solar en Cúcuta usando estadística inferencial y modelar su comportamiento a través de técnicas de inteligencia artificial.

1.7.2 Objetivos específicos.

- Adquirir datos históricos disponibles de radiación solar y otras variables meteorológicas en la ciudad de Cúcuta de diferentes bases de datos de libre acceso.
- Instalar un pirómetro con estándar ISO 9060 y un datalogger para la adquisición de datos de radiación solar en la Universidad Francisco de Paula Santander.
- Determinar la variabilidad histórica de la radiación solar en la ciudad de Cúcuta mediante métodos de estadística inferencial.
- Construir un modelo para la estimación de la radiación solar en Cúcuta basado en métodos de inteligencia artificial.
- Evaluar la precisión del modelo propuesto para validar su aplicación en la ciudad de Cúcuta.

1.8 Delimitaciones

El proyecto investigativo se realizó dentro de las instalaciones de la Universidad Francisco de Paula Santander en los laboratorios del Departamento de Electricidad y Electrónica, disponiendo

del uso de un piranómetro CMP6 de norma ISO 9060 y un datalogger modelo CR300 Series, adquiridos por el Grupo de Investigación de Microelectrónica Aplicada y Control (GIDMAC) para la toma de mediciones de la variable radiación solar global. Este proyecto se realizó en un tiempo de doce meses para la investigación, diseño y desarrollo de un modelo de estimación basado en técnicas de inteligencia artificial. Para la implementación del algoritmo, se utilizó la licencia del software Matlab disponible en los computadores de los laboratorios del departamento. La variable de salida del modelo es radiación solar global, las variables de entrada consideradas fueron: temperatura, humedad relativa, brillo solar, velocidad del viento y hora.

2. Marco Referencial

2.1 Antecedentes

A continuación, se darán a conocer los antecedentes que sirvieron como guía para la elaboración del proyecto; se tuvieron en cuenta estudios que relacionaran procesos de estimación de la radiación solar, específicamente aquellos enfocados en técnicas de inteligencia artificial y análisis estadísticos.

Comparison of artificial intelligence methods in estimation of daily global solar [13]. En este trabajo, se compararon las estimaciones de radiación solar obtenidas de un satélite y un modelo numérico de predicción del tiempo en España continental. Los autores desarrollaron una metodología diferente a las que ya habían sido utilizadas en estudios similares, en la que las variabilidades temporales de ambas fuentes de estimación se compararon con la evolución anual de la radiación asociada a las diferentes zonas climáticas de estudio. Además, evaluaron la precisión de cada fuente de estimación comparándola con los datos obtenidos de las mediciones de radiación en tierra mediante piranómetros.

Statistical analysis of solar radiation in the city of Cúcuta [14]. En esta investigación se realiza un análisis de la radiación solar en la ciudad de Cúcuta, Colombia. La técnica de empleo fue el análisis estadístico con el objetivo de brindar una descripción detallada de la energía solar presente en el territorio. Como fuente de información utilizaron los datos compuestos por muestras de radiación solar global por 24 horas del día durante un periodo de 9 años brindados por una base de datos de la estación meteorológica del IDEAM. Los diferentes análisis estadísticos los hicieron mediante el uso del software libre especializado en estadística llamado R Studio. A partir de estos estudios realizados, por medio de diferentes tipos de gráficas se logra identificar los valores de la radiación solar para cada mes y para cada año completo.

Adicionalmente, los análisis realizados obtienen un valor de 5.335 horas solares pico en la ciudad con un error cuadrático medio del 0.35 %. Asimismo, el estudio tiene como enfoque encontrar una ecuación que prediga la radiación solar de acuerdo con los datos obtenidos para brindar desde el punto de vista descriptivo el recurso solar presente en la ciudad de Cúcuta.

Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria [15]. En este documento se investiga la precisión de una técnica de computación flexible para predecir la radiación solar cuyas variables de entrada fueron: temperatura mínima media mensual, temperatura máxima y duración de la luz solar obtenida de una estación meteorológica ubicada en Iseyin, Nigeria. El modelo estimación propuesto por los investigadores es el sistema adaptativo de inferencia neuro difusa (ANFIS) de tres variables de entrada y una de salida. Los resultados que obtuvieron fueron comparados con los datos medidos en la estación por medio de coeficientes y errores estadísticos. Las características estadísticas de $RMSE = 1.0854$ y $R^2 = 0.8544$ las obtuvieron en la fase de entrenamiento y $RMSE = 1.7585$ y $R^2 = 0.6567$ en la fase de prueba. Como resultado, consideran que el modelo propuesto utilizó técnicas eficientes para predecir la radiación solar global con fines prácticos.

An ANFIS based forecasting for solar radiation data from sunshine duration and ambient temperature [16]. Esta investigación presenta un nuevo enfoque para predecir y modelar los datos de radiación solar total a partir de la duración media del brillo solar y la temperatura del aire utilizando la técnica de inteligencia artificial de inferencia neuro difusa ANFIS. En el estudio toman sus registros de una base de datos en un periodo de 10 años (1981-1990) en Argel, Argelia. Utilizan el 90 % de sus datos para el entrenamiento de la red y el 10 % faltante para la evaluación de la misma. Obtuvieron resultados de un error relativo medio que no

excede el 1 % entre los datos reales y pronosticados; además obtuvieron un coeficiente de correlación R del 98 %.

2.2 Marco Teórico

El marco teórico constituye los aspectos técnicos requeridos y que sirvieron para orientar y conceptualizar el desarrollo de este proyecto de investigación.

2.2.1 Energía Solar Fotovoltaica. La energía solar es la fuente de vida principal del planeta Tierra, es aquella encargada de dirigir los ciclos biofísicos, geofísicos y químicos al igual que los ciclos del agua, carbono, oxígeno y del clima. La energía proveniente del Sol es aquella que origina e impulsa la creación de una serie de fenómenos naturales como lo son el movimiento del agua y el viento, al igual que el crecimiento de las plantas, por esta razón la energía solar es el pilar de las fuentes renovables de las cuales se destacan la hidroeléctrica, biomasa, eólica, corrientes marinas y de las olas, además de la solar. La potencia proveniente del Sol y recibida por el planeta Tierra una vez que atraviesa la atmósfera y si el día es despejado, está valorada a lo largo del año con una media de 340 W/m^2 [17].

A su vez, la energía solar fotovoltaica es aquella que transforma la radiación solar en energía eléctrica. Este fenómeno se produce a través de paneles fotovoltaicos. En estos paneles, la radiación solar provoca la excitación de los electrones de un dispositivo semiconductor en su mayoría de Silicio generando una pequeña diferencia de potencial [18].

2.2.2 Radiación Solar. La radiación solar es la transferencia de energía por ondas electromagnéticas y se produce directamente desde la fuente hacia fuera en todas las direcciones. Estas ondas no necesitan un medio material para propagarse y pueden atravesar el espacio interplanetario y llegar a la Tierra desde el Sol [19].

La atmósfera tiene un índice de refracción aproximadamente nulo a la radiación solar entrante. Teniendo en cuenta que a la atmósfera le llega un 100 % de la radiación solar, a la Tierra llega un 25 % de radiación solar directa a la superficie, un 26 % de la radiación es dispersada por la atmósfera denominada radiación solar difusa hacia la superficie, lo que indica que a la superficie terrestre llega un 51 % de radiación. El 49 % restante es dispersado, un 19% lo absorben los gases atmosféricos y las nubes, el otro 30 % es reflejado y perdido hacia el espacio, del cual las nubes reflejan un 20 %, la atmósfera dispersa un 6 % y el suelo refleja otro 4 % [20].

En función de cómo reciben la radiación solar los objetos situados en la superficie terrestre, mostrado en la Figura 1, se pueden distinguir tres tipos de radiación.

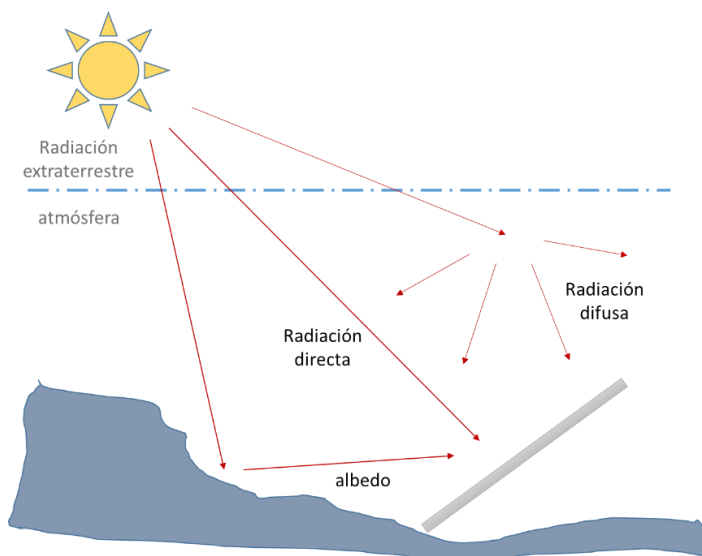


Figura 1. Componentes de la radiación solar global. Adaptado de [2]

Cuando la radiación solar entra en la atmósfera terrestre, una parte de la energía incidente es absorbida o dispersada por moléculas de aire, nubes y aerosoles. La radiación que no se refleja o dispersa y que llega a la superficie en línea directa desde el disco solar se conoce como radiación directa. La radiación que se dispersa y llega a la superficie en el suelo se llama radiación difusa.

Una pequeña parte de la radiación puede alcanzar la superficie donde se mide, después de reflejarse en el suelo; a ésta se le da el nombre de albedo [2].

En conclusión, existen tres tipos de radiación solar: difusa, directa y reflejada y radiación global definida como la suma de las tres. El trayecto de la radiación solar desde el Sol hasta la superficie terrestre está sometido bajo procesos de atenuación, reflexión y difusión los cuales alteran sus características principales, en especial la potencia. La reflexión en las nubes disminuye la radiación incidente en la superficie terrestre mientras que la absorción por vapor de agua, ozono y CO₂ produce una modificación de las características espectrales de la radiación [21].

En la Figura 2 se muestra la radiación solar global que incide en todo el mundo como promedio anual, de la cual se puede observar que los países que están sobre y cerca de la línea ecuatorial son sobre los que incide una mayor radiación, debido a que estos países no cuentan con el fenómeno de las estaciones. Cabe detallar que los indicadores de la radiación solar en el mapa están dados en watio por metro cuadrado (W/m²).

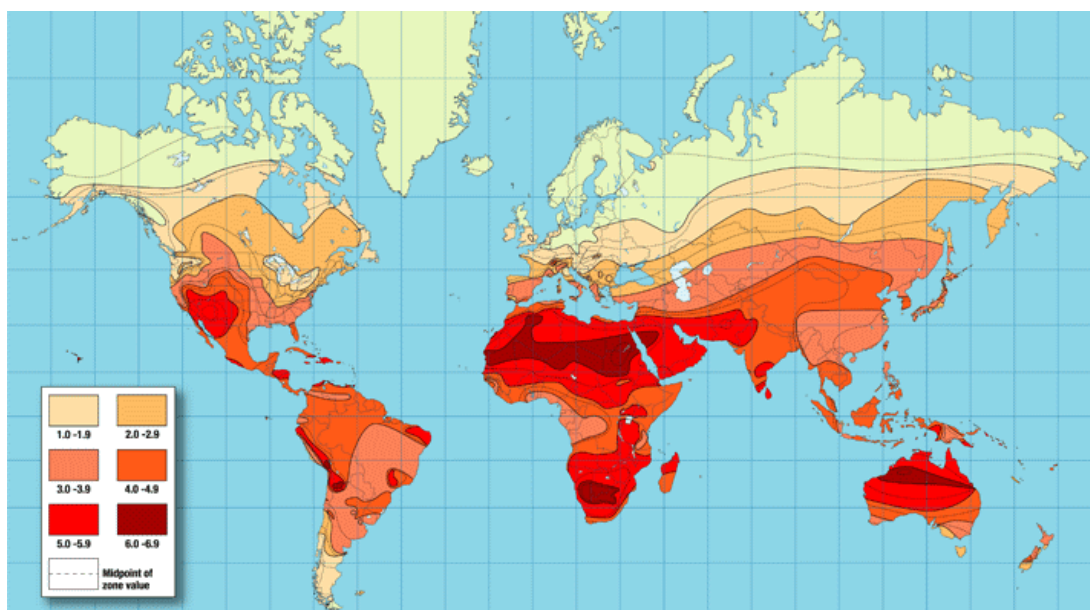


Figura 2. Atlas de la radiación solar mundial como promedio anual. Tomado de [22]

2.2.3 Piranómetro. El piranómetro es un instrumento utilizado para medir la radiación solar global sobre una superficie horizontal terrestre respondiendo de igual forma a todas las longitudes de onda del espectro de la luz blanca. El funcionamiento de este instrumento se basa en utilizar cuñas metálicas dentro de un disco las cuales están pintadas de blanco y negro de una manera intercalada; el disco es protegido por una o dos capas de vidrio principalmente en forma de domo. La diferencia de temperatura generada entre las cuñas es medida por múltiples termocuplas, las cuales, al calentarse producto de la radiación del Sol, emiten una pequeña fuerza electromotriz, tensión o diferencia de potencial, que pueden ser medidas por algún instrumento que guarde la información (integrador o datalogger). Si se desea obtener una lectura en unidades de W/m^2 , se debe multiplicar la tensión obtenida por el piranómetro por una constante del instrumento. Para este proyecto se utilizó un piranómetro CMP6 de la marca Kipp & Zonen con estándar ISO 6090 como se muestra en la Figura 3.



Figura 3. Piranómetro CMP6. Tomado de [23].

2.2.4 Análisis estadísticos de la radiación solar. El conocimiento de la disponibilidad de la energía solar resulta indispensable para el correcto dimensionamiento de sistemas y tecnologías que la transforman, por ejemplo, los sistemas fotovoltaicos para la producción de electricidad o sistemas térmicos para generar calor [4]. Por esta razón, un estudio estadístico de los datos

históricos de la radiación solar proporcionará información del comportamiento de la variable en tiempo futuro; asimismo, identificará los factores involucrados en su fluctuación, como una herramienta para definir criterios de su uso en sistemas fotovoltaicos o térmicos en una región específica. Por consiguiente, se presentan dos modelos de estadística inferencial que fueron utilizados para la evaluación del comportamiento de la radiación solar global en la ciudad de Cúcuta.

El análisis de la varianza de un solo factor (ANOVA) es un modelo de estadística inferencial que consiste en evaluar el comportamiento de una variable sometiéndola a una prueba de hipótesis para demostrar su varianza entre y dentro de diferentes tratamientos; su finalidad es determinar si existe una diferencia entre la media de valores según un factor categórico o tratamiento. El proceso consiste en someter a cada punto de la población en un par de pruebas de hipótesis, una de ellas es la hipótesis nula (H_0) la cual indica que las medias muestrales son iguales en todos los tratamientos, y la hipótesis alternativa (H_a) consiste en que por lo menos una de las medias muestrales varía significativamente entre los demás tratamientos.

Tabla 1.

Tabla ANOVA

ANOVA	Df	Sum of Sq	Mean Sq	F Value	P
Factor (YEAR)	-	-	-	-	-
Residuals	-	-	-	-	-

En la Tabla 1 se encuentran una serie de parámetros que definen el análisis ANOVA con relación a la investigación y se detallan a continuación [24]:

- Factor (YEAR): son los resultados de diferentes análisis que se hagan a cada año con respecto a todos los datos de la población.

- Residuals: hace referencia a los resultados de los análisis posibles entre cada año o tratamiento con respecto a las medias de cada uno.
- Df: son los grados de libertad del modelo, denotados en las siguientes ecuaciones:

$$Df(\text{Factor}(\text{YEAR})) = I - 1$$

$$Df(\text{Residuals}) = N - I$$

Donde I son los tratamientos (años) que están siendo analizados, y N es el número total de las muestras.

- Sum of Sq: La suma total de cuadrados de tratamiento hace referencia a la variación de las medias muestrales alrededor de la media muestral total y la suma de los cuadrados del error es la diferencia en los puntos muestrales individuales alrededor de su media muestral de acuerdo con cada tratamiento.
- Mean Sq: la media cuadrática del tratamiento y la media cuadrática del error, se definen por el cociente entre la suma total de cuadrados del tratamiento y los grados de libertad del tratamiento; y el cociente entre la suma de los cuadrados del error y los grados de libertad del error.
- F Value: es el cociente entre la media cuadrática del tratamiento y la media cuadrática del error, este valor indica si la hipótesis nula puede ser aceptada o rechazada. Cuando H_0 es verdadera, el numerador y denominador de F son, en promedio, del mismo tamaño, por lo que F tiende a estar cerca de 1.
- P: La columna con el índice "Pr" se refiere a la probabilidad de que la hipótesis nula sea aceptada o rechazada, es decir, se rechaza la H_0 si el valor p es mayor o igual al valor alfa asociado, que para este caso es igual a 0.05.

El análisis ANOVA se evalúa de acuerdo a dos índices muy importantes, el valor F y el valor de probabilidad del cual se denomina por un valor alfa de 0.05; según estos resultados se logra conocer si se acepta o rechaza la hipótesis nula y se llega a concluir si la variable puesta en estudio presenta cambios significativos con respecto a su media poblacional [25].

La prueba de Bonferroni es una continuación del modelo ANOVA con el fin de reforzar la toma de conclusiones con respecto a los resultados obtenidos, se basa en ajustar un nuevo valor de alfa, donde el alfa prima (α') es la división del alfa preestablecido como 0.05 y el número de pares posibles que se puedan hacer entre los tratamientos (m). Una vez se tiene el nuevo valor de alfa o alfa prima, se procede a hacer una prueba de hipótesis e identificar si existen valores de probabilidad estadísticamente significativos para proceder a rechazar la hipótesis nula [24].

2.2.5. Modelos de predicción de radiación solar. La radiación solar es una de las variables más influyentes para el desarrollo de aplicaciones e investigaciones relacionadas a energías renovables. Sin embargo, la adquisición de mediciones de radiación solar no siempre es posible por diferentes motivos (falta de estaciones meteorológicas) por lo que es necesario recurrir a modelos capaces de estimarla.

Para el diseño e implementación de sistemas fotovoltaicos, es de suma importancia un conocimiento detallado y preciso de la radiación solar global sobre una superficie horizontal en una ubicación geográfica en específico [26]. En países en vía de desarrollo, como Colombia, la red de estaciones meteorológicas que miden la radiación solar arroja escasos resultados además de imprecisos y dispersos. Cuando los datos de radiación no están disponibles, se pueden estimar a partir de otras variables meteorológicas utilizando modelos empíricos, modelos estadísticos, procesamiento de imágenes satelitales, técnicas de inteligencia artificial o modelos híbridos. En la Figura 4 se presenta la clasificación descrita anteriormente.

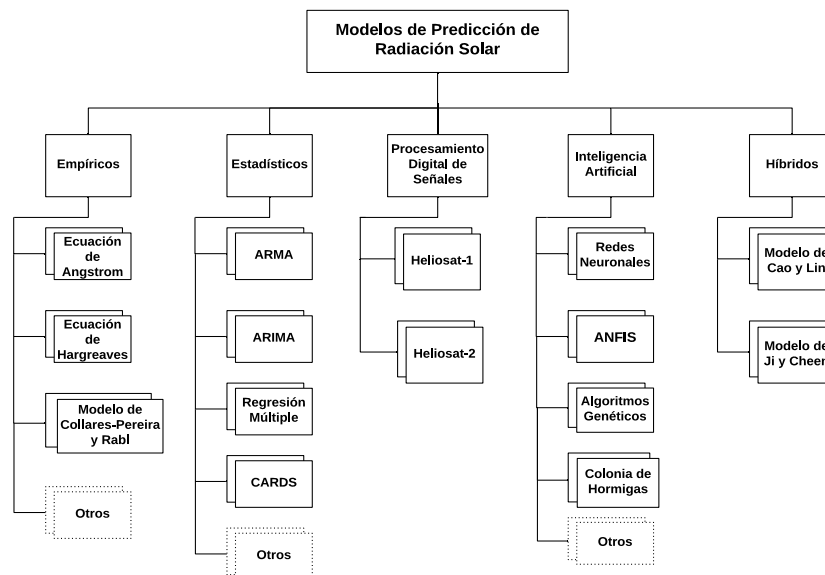


Figura 4. Clasificación de los modelos de predicción de radiación solar. Fuente: elaboración propia

Los modelos empíricos basan su estudio en proponer una ecuación que relacione algunas variables meteorológicas con la radiación solar. Los más usados en estudios relacionados son las ecuaciones de Angstrom y Hargreaves; ambos modelos hacen una estimación de la radiación solar por medio de las variables brillo solar y temperatura, se realizan mediante regresiones que se ajustan a partir de datos anteriormente medidos.

Los modelos estadísticos utilizan series de tiempo para resolver problemas en el campo de las matemáticas, finanzas y la ingeniería, utilizando ecuaciones matemáticas para codificar información extraída de los datos observados del pasado [27]. Los modelos estadísticos usados más frecuentemente para la predicción de radiación solar son ARMA y ARIMA.

Los modelos basados en imágenes satelitales determinan la radiación solar a partir del efecto que generan las nubes sobre la radiación en un sitio particular. Estos modelos utilizan procesamiento digital de señales y procesamiento de imágenes.

La inteligencia artificial (IA) es la combinación de algoritmos entrenados con el fin de recrear o emular la manera en que se comporta el ser humano, es decir, crear máquinas con las mismas capacidades del ser humano. Es una rama de la computación que trata con los símbolos y estrategias no algorítmicas para la solución de problemas [28], estas técnicas han sido utilizadas para predecir la radiación solar a partir de otras variables de entrada.

Un modelo híbrido es la combinación de dos o más modelos y/o técnicas de estimación de la radiación solar con el fin de obtener resultados más óptimos y precisos. Se han encontrado varias investigaciones en las que combinan técnicas estadísticas con IA, a su vez han empleado dos o más métodos de IA para obtener una mejora en sus estimaciones.

2.2.6 Modelos de inteligencia artificial. La inteligencia artificial, es una rama de las ciencias computacionales que se ocupa de los símbolos y métodos no algorítmicos para la resolución de problemas; se podría considerar que la IA es un dialecto simbólico constituido por cadenas de caracteres que representan conceptos del mundo real [28].

La IA está integrada con base en unos principios que pueden variar según la técnica aplicada, y depende de cual disciplina se esté trabajando, pero en general, tienen bases muy similares, lo único que las diferencia a unas de otras es la forma en que se va a afrontar o dar solución al problema [29].

2.2.6.1 Redes Neuronales Artificiales (RNA). Se definen como sistemas de mapeos no lineales cuya estructura se basa en los principios observados en los sistemas nerviosos de humanos y animales. Consiste en una gran cantidad de procesadores simples unidos por conexiones con pesos. Las unidades de procesamiento se llaman neuronas. Cada unidad recibe entradas de otros nodos y genera una salida simple que depende de la información local disponible, almacenada internamente o llegando a través de las conexiones con los pesos [28]. Es

decir, cada unidad de procesamiento (neurona) responde a una tarea simple: recibe entradas de otras unidades o fuentes externas y procesa la información para obtener una salida que se propaga a otras unidades. La Figura 5 representa la arquitectura de una RNA multicapa.

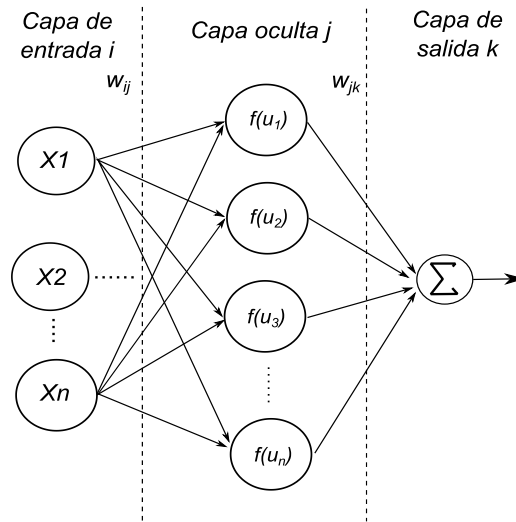


Figura 5. Arquitectura de una RNA multicapa. Tomado de: [30]

La Figura 5 muestra una estructura básica de una RNA con tres capas, donde la primera capa es la capa de entrada (i) que recibe la información, la segunda capa es la capa oculta (j) que puede estar constituida por varias capas (cada una con estructura similar o diferente) y determina el procesamiento de la información de entrada, y finalmente la tercera capa conocida como la capa de salida (k) donde los resultados se reciben y analizan para proporcionar la salida.

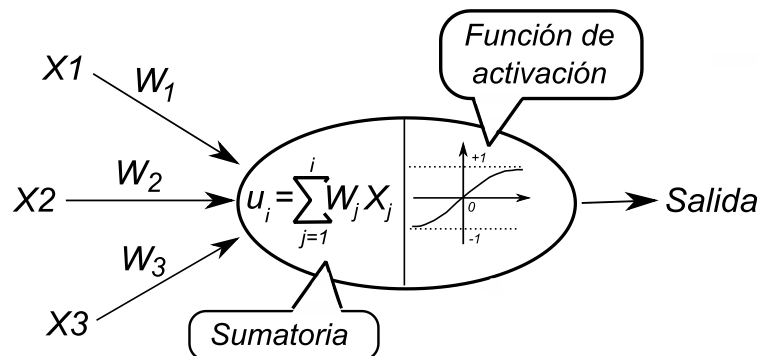


Figura 6. Detalle de una neurona en una RNA. Adaptada de: [28]

La Figura 6 representa el funcionamiento interno de una neurona o unidad de procesamiento de la información, donde cada capa está interconectada por su peso correspondiente ω_{ij} y ω_{jk} , y cada unidad suma sus entradas, agrega el término de sesgo o umbral y una función lineal o no lineal transforma la suma para producir una salida. Esta transformación se llama la función de activación del nodo, que con frecuencia es una función lineal en la capa de salida [30].

Una RNA realiza su proceso de entrenamiento utilizando los pares de datos entrada-salida por medio de un algoritmo de aprendizaje, en la mayoría de los casos se basa en el error encontrado entre la salida real y la obtenida en el momento, dependiendo del resultado la RNA ajusta los pesos de cada conexión para garantizar una salida más cercana a la real. De esta manera, la RNA optimiza la generación de sus salidas y está lista para resolver problemas reales.

2.2.6.2 Sistemas de inferencia difusa. Los sistemas de inferencia difusa también se conocen como sistemas basados en reglas difusas, modelos difusos, memorias asociativas difusas (FAM) o controladores difusos cuando se usan como controladores. Son una forma de representar conocimientos y datos inexactos en forma similar a como lo hace el pensamiento humano. Un sistema de inferencia difusa (FIS) define una correspondencia no lineal entre una o varias variables de entrada y una variable de salida [31].

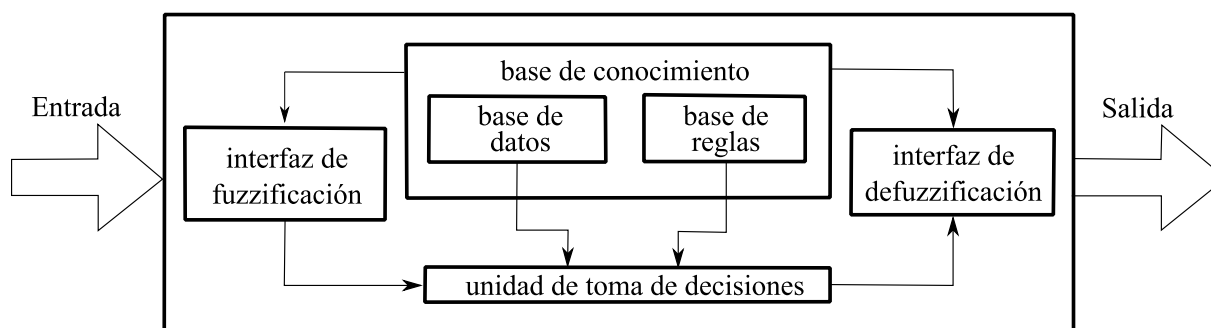


Figura 7. Diagrama básico de un sistema de inferencia difusa. Tomado de: [28]

Básicamente, un sistema de inferencia difuso se compone de cinco bloques funcionales mostrados en la Figura 7, este grupo de bloques contiene [31]:

- ✓ Una base de reglas que contiene una serie de reglas difusas si-entonces.
- ✓ Una base de datos que define las funciones de membresía (FM) de los conjuntos difusos utilizados en las reglas.
- ✓ Una unidad de toma de decisiones que realiza las operaciones de inferencia sobre las reglas.
- ✓ Una interfaz de fuzzificación que transforma las entradas nítidas (es decir, fijas) en grados de consecuencia con valores lingüísticos.
- ✓ Una interfaz de defuzzificación que transforma los resultados difusos de la inferencia en el mismo rango de la entrada.

Las funciones de membresía en un FIS son aquellas que indican el grado de pertenencia que tiene un dato con respecto al conjunto de datos. Los pasos del razonamiento difuso (operaciones de inferencia sobre reglas difusas si-entonces) realizadas por los sistemas de inferencia difusa son [31]:

1. Comparar las variables de entrada con las funciones de membresía en la parte de la premisa para obtener los valores de membresía (o medidas de compatibilidad) de cada etiqueta lingüística (este paso a menudo se denomina fuzzificación).
2. Combinar (a través de un operador específico de la norma T , usualmente multiplicación o min.) Los valores de membresía en la parte de la premisa para obtener la fuerza de disparo (peso) de cada regla.
3. Generar el consecuente calificado (difuso o nítido) de cada regla dependiendo de la fuerza de disparo.

4. Agregar los consecuentes calificados para producir una salida nítida (este paso se llama defuzzificación).

Dependiendo de los tipos de razonamiento difuso y reglas difusas de si-entonces empleadas, los sistemas de inferencia difusa se pueden clasificar en tres tipos [31]:

Tipo 1: La producción general es el promedio ponderado de la producción nítida de cada regla inducida por la fuerza de disparo de la regla (el producto o el mínimo de los grados de coincidencia con la parte de la premisa) y las funciones de membresía de salida. Las funciones de membresía de salida utilizadas en este esquema deben ser funciones monótonas.

Tipo 2: la salida difusa general se deriva aplicando la operación "ma" a las salidas difusas calificadas (cada una de las cuales es igual al mínimo de fuerza de disparo y la función de pertenencia de salida de cada regla). Se han propuesto varios esquemas para elegir la salida nítida final en función de la salida difusa general; algunos de ellos son centroide de área, bisectriz de área, media de máximos, criterio máximo, etc.

Tipo 3: se usan las reglas borrosas de Takagi-Sugeno si-entonces. La salida de cada regla es una combinación lineal de variables de entrada más un término constante, y la salida final es el promedio ponderado de la salida de cada regla.

La Figura 8 utiliza un sistema de inferencia difusa de dos reglas y dos entradas para mostrar diferentes tipos de reglas y razonamiento difusos mencionados anteriormente.

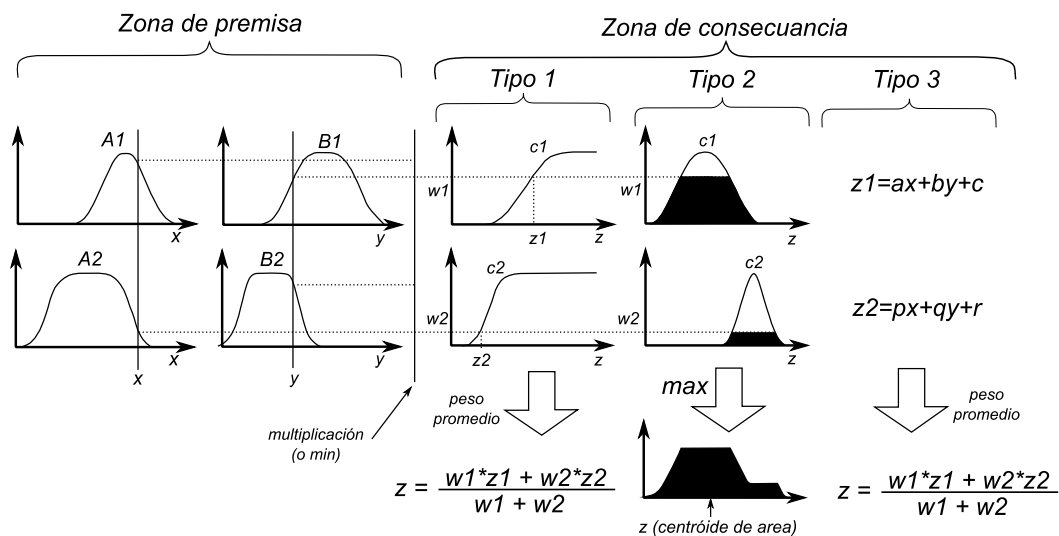


Figura 8. Tipos de sistemas de inferencia difusa. Adaptado de: [31]

2.2.6.3 Sistema adaptativo de inferencia neuro difusa (ANFIS). Es un modelo híbrido de inteligencia artificial el cual está comprendido por una red neuronal artificial con inferencia difusa. Este tipo de topología fue propuesto por R. Jang en su trabajo de tesis denominado neuro-fuzzy modeling en el año 1993. La idea principal del modelo consiste en que los nodos de las diferentes capas de una red neuronal convencional contienen parámetros difusos, siendo equivalente a los sistemas de inferencia difusa (FIS) con parámetros distribuidos [28].

ANFIS sigue el patrón de una red adaptativa comprendida por nodos fijos y nodos ajustables. En la Figura 9 se muestra una red de alimentación multicapa donde cada nodo realiza una función en específico (función del nodo) en las señales entrantes, además de los parámetros pertenecientes a cada nodo.

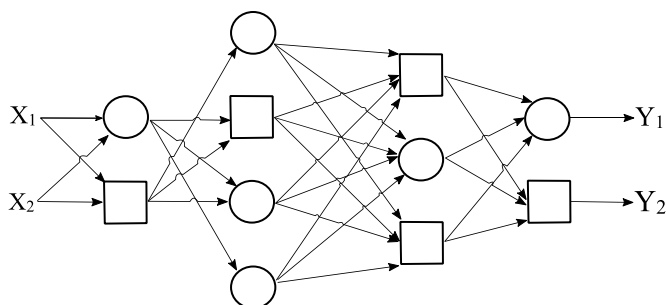


Figura 9. Topología de una red adaptable con nodos fijos y adaptables. Tomado de: [28]

Para reflejar diferentes capacidades adaptativas, ambos nodos circulares y cuadrados se utilizan en la representación de una red adaptativa. Un nodo cuadrado (nodo adaptativo) tiene parámetros, mientras que un nodo circular (nodo fijo) no tiene ninguno. El conjunto de parámetros en una red adaptativa es la unión de todos los parámetros de cada nodo adaptativo [31]. Para lograr un mapeo de entrada-salida deseado, estos parámetros se actualizan de acuerdo con los datos de entrenamiento dados y un procedimiento de aprendizaje basado en gradiente, aunque se pueden aplicar varias técnicas híbridas para aumentar su rendimiento y velocidad de aprendizaje.

De acuerdo con la topología de los sistemas de inferencia difusa, estos sistemas están clasificados en tres tipos, mencionados anteriormente. Donde los dos primeros tipos utilizan el modelo de trabajo de Mamdani, el cual utiliza una técnica de aprendizaje supervisado (aprendizaje de retro propagación) para aprender los parámetros de las funciones de membresía. El sistema de inferencia difusa tipo 3 utiliza un modelo de trabajo de Takagi-Sugeno, estos hacen uso de una mezcla de retro propagación para aprender las funciones de membresía y la estimación de mínimos cuadrados medios para determinar los coeficientes de las combinaciones lineales en las conclusiones de la regla [32].

Por lo tanto, ANFIS cae en el tercer tipo de sistema de inferencia difusa, donde en su núcleo, la técnica divide la representación del conocimiento previo en subconjuntos para reducir el

espacio de búsqueda y utiliza el algoritmo de propagación inversa para ajustar los parámetros difusos. El sistema resultante es una red neuronal adaptativa funcionalmente equivalente a un sistema de inferencia Takagi-Sugeno de primer orden, donde la relación entrada-salida es lineal, y se muestra en la Figura 10.

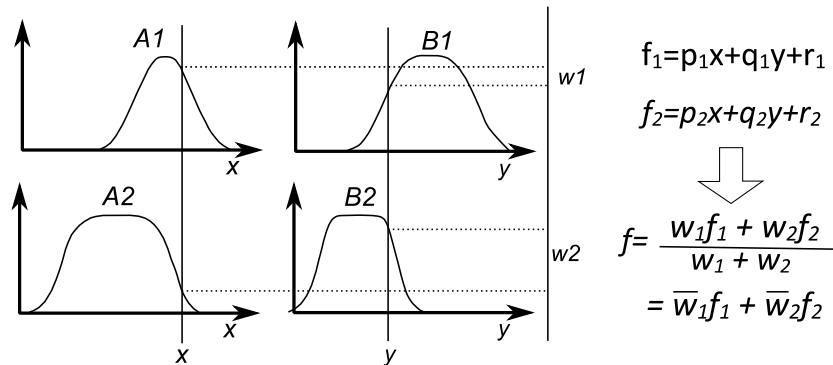


Figura 10. Razonamiento difuso tipo 3. Adaptada de: [31]

En un sistema Sugeno de primer orden, un conjunto de reglas típico con dos reglas borrosas si-entonces se puede expresar como:

$$\text{Regla 1: Si } x \text{ es } A_1 \text{ y } y \text{ es } B_1, \text{ entonces } f_1 = p_1x + q_1y + r_1$$

$$\text{Regla 2: Si } x \text{ es } A_2 \text{ y } y \text{ es } B_2, \text{ entonces } f_2 = p_2x + q_2y + r_2$$

Donde x y y son las entradas nítidas al nodo i , A_i y B_i son los conjuntos difusos en el antecedente, f_i es la salida dentro de la región difusa especificada por la regla difusa; y p_i , q_i y r_i son los parámetros de diseño que se determinan durante el proceso de entrenamiento.

Seguido a esto, la estructura ANFIS se compone de cinco capas: capa difusa, capa de producto, capa normalizada, capa de difuminación y capa de salida total. La estructura ANFIS para dos entradas, una salida y dos reglas se muestra en la Figura 11.

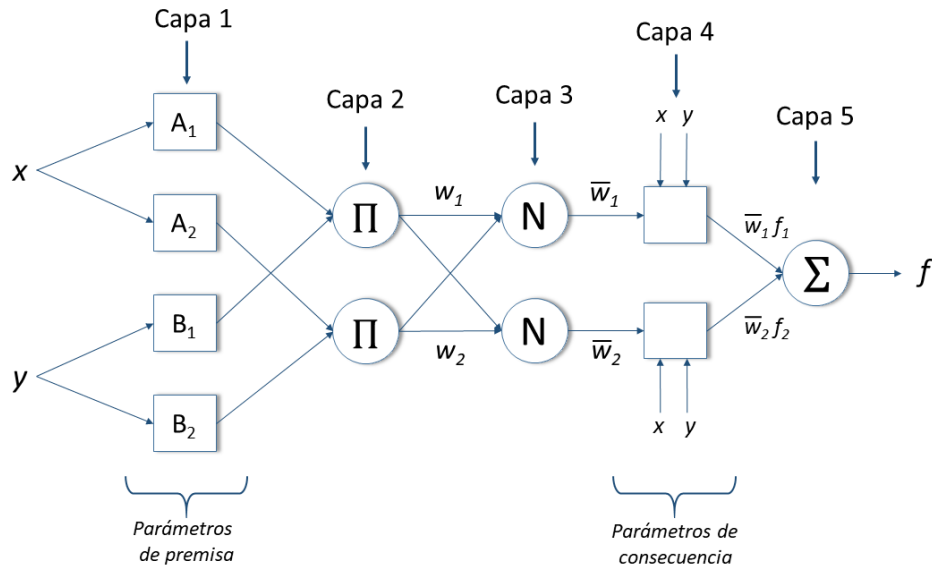


Figura 11. Estructura básica de una red ANFIS. Adaptada de:[33]

Cada capa realiza una tarea en particular en el sistema de inferencia difusa, definidas por [31] y serán explicadas a continuación.

Capa 1 (capa difusa): La primera capa es la capa difusa que convierte las entradas en un conjunto difuso por medio de funciones de membresía. Cada nodo i en esta capa es un nodo cuadrado con una función de nodo:

$$O_i^1 = \mu_{A_i}(x)$$

Donde x es la entrada al nodo i , y A , es la etiqueta lingüística (pequeña, grande, etc.) asociada con esta función de nodo. En otras palabras, O_i es la función de pertenencia de A , y especifica el grado en que la entrada x satisface el cuantificador A_i . Por lo general, $\mu_{A_i}(x)$ se elige con forma de campana con un rango de 0 a 1, aunque también puede ser una función de membresía triangular o trapezoidal como se muestra en la Figura 12.

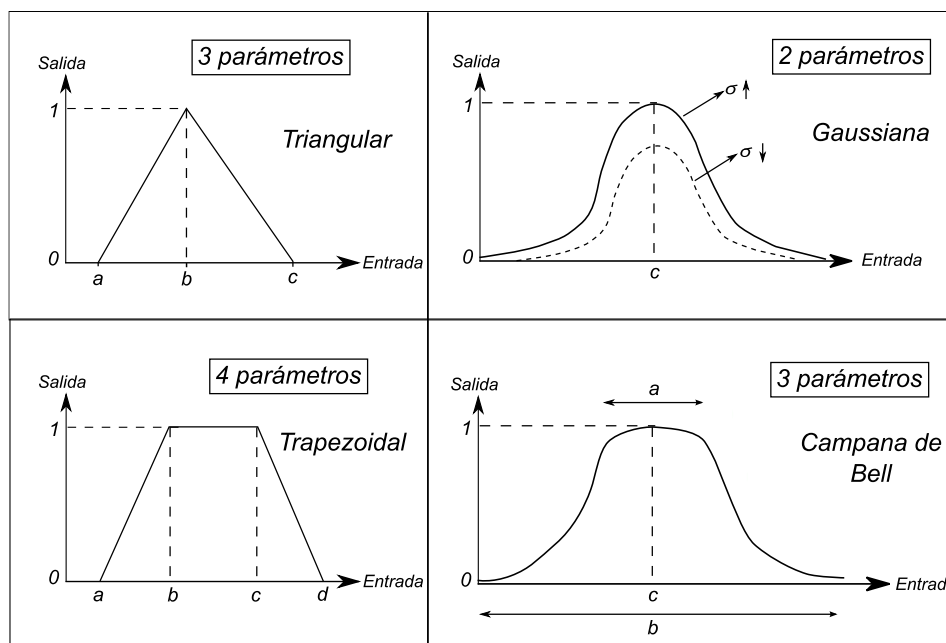


Figura 12. Tipología de funciones de membresía

Los parámetros en esta capa se denominan parámetros de premisa mostrados en la Tabla 2.

Tabla 2.

Modelo matemático de las funciones de membresía

Nombre de las FMs	Ecuación
FM Triangular	$u_{Ai}(x) = \max \left\{ \min \left(\frac{x-a}{b-a}, \frac{c-x}{c-b} \right), 0 \right\}$
FM Trapezoidal	$u_{Ai}(x) = \max \left\{ \min \left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c} \right), 0 \right\}$
FM Campana de Gauss	$u_{Ai}(x) = e^{-\frac{(x-c)^2}{2\sigma^2}}$
FM Campana de Bell	$u_{Ai}(x) = \frac{1}{1 + \left \frac{x-c}{a} \right ^2}$

Capa 2 (capa de producto): Esta capa tiene los nodos circulares representados con Π en la Figura 11. La función del nodo es aplicar el operador de regla (AND/OR) entre las señales entrantes con el objeto de servir como una señal de salida, de la siguiente manera:

$$\omega_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \quad i = 1, 2.$$

Donde ω_i representa la fuerza de disparo de cada regla.

Capa 3 (capa de normalización): los nodos en esta capa se denotan con la letra N, su función es calcular la relación entre la fuerza de disparo de cada regla en específico con respecto a la suma de todas las fuerzas de disparo de todas las reglas, así:

$$\bar{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2}, \quad i = 1, 2.$$

Donde $\bar{\omega}_i$ es denominado como fuerza de disparo normalizada.

Capa 4 (capa de difuminación): Los nodos en esta capa se denotan con un cuadrado y calculan la salida ponderada de cada función lineal de la siguiente manera:

$$O_i^4 = \bar{\omega}_i f_i = \bar{\omega}_i (p_i x + q_i y + r_i)$$

donde $\bar{\omega}_i$ es la salida de la capa de normalización, y (p_i, q_i, r_i) son los coeficientes de una combinación lineal en el sistema de inferencia de Sugeno. Los parámetros en esta capa se denominarán parámetros consecuentes.

Capa 5 (capa de salida): esta capa se constituye por un único nodo de forma circular etiquetado por Σ , su función es calcular la salida general como la suma de todas las entradas entrantes, es decir:

$$O_i^5 = \text{salida general} = \sum_1 \bar{\omega}_i f_i = \frac{\sum \bar{\omega}_i f_i}{\sum \omega_i}$$

Por lo tanto, se construye una red adaptativa que es funcionalmente equivalente a un sistema de inferencia difusa de tipo 3, y su razonamiento difuso se presentó anteriormente en la Figura 10.

ANFIS utiliza un algoritmo de aprendizaje híbrido para la estimación de la premisa y los parámetros consiguientes. El algoritmo de aprendizaje híbrido estima los parámetros

consecuentes en un paso hacia adelante y los parámetros de la premisa en un paso hacia atrás. En la fase de avance, la información se propaga hacia adelante hasta la capa 4, donde los parámetros subsiguientes se optimizan mediante un algoritmo de regresión de mínimos cuadrados. En la fase hacia atrás, las señales de error se propagan hacia atrás y los parámetros de la premisa se actualizan mediante un algoritmo de gradiente descendiente (GD). Este error medido generalmente se define por la suma de la diferencia al cuadrado entre los valores medidos y modelados y se minimiza a un valor deseado [31]. Luego, el resultado general final se puede reescribir como:

$$f_{salida} = (\overline{\omega_1} \cdot x)p_1 + (\overline{\omega_2} \cdot y)q_1y + (\overline{\omega_1})r_1 + (\overline{\omega_1} \cdot x)p_2 + (\overline{\omega_2} \cdot y)q_2y + (\overline{\omega_1})r_2$$

Donde x y y son los parámetros de entrada del modelo, ω_1 y ω_2 son las fuerzas de disparo normalizadas de las reglas difusas y p_i , q_i , r_i son los parámetros consecuentes.

3. Metodología

Para el desarrollo del proyecto de investigación, se mostrará un listado de actividades para evidenciar cómo fue alcanzado cada uno de los objetivos específicos propuestos anteriormente, explicando con detalle cada actividad que se realizó para su cumplimiento.

3.1 Adquisición de datos de radiación solar

Se identificaron tres bases de datos con información histórica de la radiación solar global en la ciudad de Cúcuta. Seguido a esto, se buscaron entidades que tuvieran un registro histórico de mediciones de variables meteorológicas en especial la radiación solar global, de las cuales se seleccionó una base de datos nacional (IDEAM) y dos internacionales (NASA, PVGIS).

El IDEAM cuenta con estaciones meteorológicas con sensores para medir directamente la radiación solar. Para obtener información de esta entidad, se realizó una solicitud formal por diferentes medios (correo electrónico, aplicativo web, llamada telefónica, etc.) para obtener datos históricos de variables climáticas disponibles en las estaciones meteorológicas de esta entidad en la ciudad de Cúcuta [34]. Las bases de datos internacionales (NASA [35], PVGIS [36]) obtienen su información por medio de satélites y para obtener información de estas entidades se debe ingresar a su página oficial y descargar las variables deseadas en un tiempo determinado. Para esta investigación se definió un rango de tiempo de 10 años, donde fueron seleccionados los datos comprendidos desde el 2005 hasta el 2015.

Al obtener la información de las tres bases de datos se verificó si se encuentra la cantidad de mediciones en el periodo comprendido; la base de datos del IDEAM presenta brechas de días y meses de datos no medidos, además algunos datos no cumplen con el rango de medición en el que se debe encontrar la variable; debido a esto, por medio del software Matlab se realizó un proceso de depuración de los datos con el objetivo de clasificar aquellos valores cuyas medidas

estén dentro del rango correspondiente. La base de datos de la NASA también presenta ausencia de mediciones y valores atípicos que fueron depurados; por último, la base de datos del PVGIS no cuenta con todos sus datos, aunque todas sus mediciones están dentro del rango.

La base de datos del IDEAM y del PVGIS proporcionan la información de la radiación solar en unidades de W/m^2 por cada hora; sin embargo, la información suministrada por la NASA contiene la sumatoria de la radiación solar en W/m^2 por día.

Teniendo en cuenta que la base de datos suministrada por el IDEAM es una institución del gobierno colombiano con mediciones en sitio, se debe tener su información como prioridad en este estudio; pero debido a que la información no se encuentra completa y en algunos casos los datos se salen del rango, no brinda la suficiente veracidad para este trabajo.

Por lo tanto, el propósito de este primer análisis es identificar qué tan precisa es la información suministrada por las bases de datos de la NASA y PVGIS con respecto a los datos del IDEAM. Para ello con los datos ya depurados se sincronizan las tres bases de datos con el objetivo de que cada una quede con la misma cantidad de información. Los datos ya procesados se someten a un análisis de similitud, evaluado bajo diferentes errores y coeficientes estadísticos por medio de la herramienta computacional Matlab.

Una vez identificada la base de datos que tenga mejores resultados con respecto a la del IDEAM, se utilizan los datos completos de esta base de datos para realizar el análisis estadístico, que será evidenciado más adelante en este documento.

3.2 Instrumentación para medición y registro de datos

Para la medición de la radiación solar global en una superficie horizontal, se utilizó un piranómetro y un datalogger para la adquisición de los datos en la Universidad Francisco de Paula Santander. Estos dispositivos de medición y registro de datos pertenecen al Grupo de

Investigación de Microelectrónica Aplicada y Control (GIDMAC) del departamento de Electricidad y Electrónica. Para hacer correcto uso de estos equipos se solicitó un permiso por medio de una carta al director del grupo de investigación. Para poder realizar la instalación del piranómetro se procedió a pedir un permiso a la vicerrectoría administrativa para acceder a las instalaciones del edificio de laboratorios de electrónica, donde como respuesta nos redirigieron a la dependencia de servicios generales y ellos fueron la entidad encargada de realizar la instalación.

Los dispositivos fueron instalados en el techo del edificio de laboratorios de ingeniería electrónica, más exactamente en el área del techo que comprende la oficina del grupo de investigación (GIDMAC) mostrado en la Figura 13. Se seleccionó esta área de manera estratégica cerca del grupo de investigación con el fin de poder hacer la toma de mediciones dentro de la oficina y evitar mayor gasto en cables de conexión además de no tener que hacer traslado de equipos.



Figura 13. Grupo de investigación de microelectrónica aplicada y control (GIDMAC)

Después de realizar un estudio de campo con la herramienta Solar Pathfinder que está disponible en el grupo de investigación (GIDMAC), se identifica que el área seleccionada no

presenta sombras que puedan alterar las mediciones tomadas en el transcurso del año garantizando que las mediciones de radiación solar global no están siendo interrumpidas por algún objeto (edificio, árboles, cableado, etc.).

En la Figura 14 se muestra un diagrama de la instrumentación requerida para la toma de mediciones, lo cual inicia por medio de un sensor de radiación solar global denominado piranómetro, su modelo es CMP6 con estándar ISO 9060 de la empresa Kipp & Zonen, el cual se encarga de medir la variable y enviar la información en una diferencia de potencial, esta diferencia de potencial es medida por un datalogger de modelo CR300 de la marca Campbell Scientific, el cual funciona como un transductor para este caso y se encarga de enviar los datos a una computadora por medio del software especializado LoggerNet.



Figura 14. Diagrama representativo de la instrumentación requerida

La empresa Campbell Scientific brinda un software licenciado para el uso adecuado del datalogger, este software denominado LoggerNet versión 4.5 [37] es un software de soporte que permite programar, comunicar y descargar datos del datalogger al PC. Para este caso se procedió a realizar el proceso de comunicación entre el datalogger y el PC por medio del software, la Figura 15 enseña el menú principal del software.

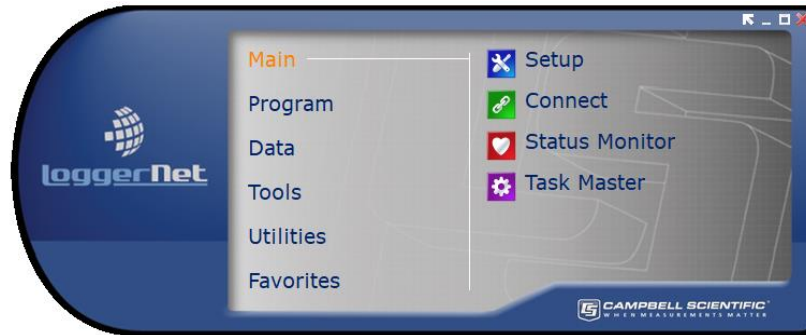


Figura 15. Menú principal del software LoggerNet

Por medio del seguimiento de una serie de pasos por medio del conector micro USB se logra la comunicación entre ambos dispositivos. Seguido a esto, se realizó el proceso de configuración del datalogger para el reconocimiento del piranómetro con el fin de poder realizar la toma de mediciones de radiación solar. Se programó el datalogger de la siguiente manera:

1. Se seleccionó el datalogger con el cual se haría la adquisición de datos (modelo CR300 Series).
2. Se realizó el proceso de selección del sensor así: Sensors > Meterologicals > Solar Radiation > CMP3/CMP6/CMP11 Pyranometer. Se escoge esta ruta debido a que el sensor que se va a utilizar es el piranómetro modelo CMP6 mencionado anteriormente.
3. Se definió el intervalo de tiempo en el que el datalogger realizará las mediciones, donde serán tomadas en un intervalo de 5 minutos.
4. Se define el tipo de salida que se desea de las mediciones tomadas, esto se debe a que el datalogger toma 30 mediciones cada 10 segundos durante los 5 minutos. El tipo de salida se definió como el promedio de las 30 mediciones.
5. Se realizaron las conexiones indicadas por el software mostrado en la Figura 16 y se envía el programa al datalogger para hacer inicio de la toma de mediciones.

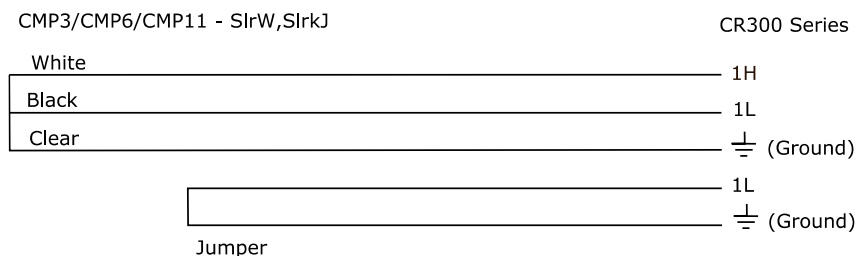


Figura 16. Diagrama de conexiones del piranómetro y datalogger

Después de que el piranómetro es instalado y el datalogger está correctamente programado se inicia el proceso de toma de mediciones de radiación solar global en las instalaciones de la Universidad Francisco de Paula Santander.

3.3 Análisis estadístico de la radiación solar

Una vez seleccionada la base de datos que presente mejores resultados en el análisis de similitud, se procede a realizar análisis estadísticos con el fin de evaluar la variabilidad de la radiación solar y encontrar los factores causantes de dicha variación durante un periodo de determinado. El análisis de similitud se hace con el fin de poder escoger una base de datos que sea precisa con respecto a las mediciones tomadas por el IDEAM, debido a que para realizar un análisis estadístico se requiere la mayor cantidad de muestras posibles de una población para garantizar resultados óptimos y confiables.

En la investigación, la variable puesta en estudio es la radiación solar en la ciudad de Cúcuta, Norte de Santander, Colombia; con latitud $N7^{\circ}53'38.08''$ y longitud $O72^{\circ}30'28.15''$, evaluada en el periodo de 10 años comprendidos desde el año 2005 hasta el año 2015.

Se realiza el análisis de la varianza de un solo factor (ANOVA); su estudio se basa en analizar la varianza de un parámetro sometiéndolo a una prueba de hipótesis y analizando cómo se comporta la variable dentro y entre los tratamientos. Dichos tratamientos para este caso son los años de estudio evaluados; la idea de realizar este modelo de estadística inferencial es verificar si

existe o no una variabilidad estadísticamente significativa de la radiación solar global durante el periodo en estudio.

Estos análisis fueron realizados mediante el software estadístico R Studio Versión 1.1.463 – © 2009-2018 RStudio, Inc. Por último, para realizar un análisis detallado de la información y complementar los resultados obtenidos por el modelo ANOVA, se desarrolla el test de Bonferroni, con el comando “*pairwise.t.test(SOLRAD, YEAR, p.adj=“bonf”)*”, el software ejecuta múltiples pruebas *t* entre todos los posibles pares de tratamientos, con el propósito de buscar la diferencia más significativa que exista entre los años de estudio.

Por medio de estos estudios se logra identificar la variabilidad existente en la radiación solar global. Además, se realizaron diagramas y gráficos estadísticos que describen el comportamiento de la radiación solar global anual y se definen algunas causas por las cuales este parámetro presenta variabilidad en la zona geográfica seleccionada.

3.4 Creación de la estructura ANFIS

Para la construcción de un sistema adaptativo de inferencia neuro difusa (ANFIS), se utilizó la base de datos del IDEAM que, a pesar de no tener su información completa, presentaba un número importante de datos que podía ser usado para el modelamiento de la red de inteligencia artificial. En esta sección, los datos del IDEAM son prioridad debido al uso de equipos de medición en sitio que pueden presentar una mayor fiabilidad de la información suministrada. Después de solicitar formalmente la información, se obtiene la base de datos con variables climáticas en la ciudad de Cúcuta de la estación meteorológica satelital de la Universidad Francisco de Paula Santander. Las variables brindadas por el IDEAM son la humedad relativa, temperatura ambiente, brillo solar, velocidad del viento y radiación solar global. Cada conjunto de datos viene definido con el día, mes y año en el que se tomó la medición; adicionalmente, el

IDEAM toma sus mediciones de manera horaria lo que significa que la base de datos también suministra la hora en que fue tomada la medición.

Después de analizar y procesar los datos, se eliminaron registros donde hacían falta por lo menos una de las 5 variables meteorológicas; también fueron eliminadas las variables que se salían de su rango de operación, quedando al final con un total de 23 220 registros. A continuación, se presentan los pasos implementados para la creación de la red de inteligencia artificial.

3.4.1 Normalización. El proceso de normalización es un método común utilizado en el preprocesamiento de datos, y consiste en transformar los datos de entrada de un rango amplio (original) a un rango más pequeño [38]. El rango de normalización puede variar ya que debe especificarse de acuerdo con la función de membresía que se utilizará en la red; en este estudio, el rango de operación con el que se va a trabajar será de 0 a 1. Se restringen estos datos con el fin de mantener todas las variables con un mismo rango de operación, minimizar el error de regresión, mejorar la precisión y mantener la correlación entre el conjunto de datos. Su implementación está definida por la siguiente fórmula [39]:

$$X_{norm} = \frac{X_{actual} - X_{min}}{X_{max} - X_{min}}$$

Donde X_{norm} representa a los datos normalizados, X_{actual} es la entrada original, X_{max} y X_{min} son los valores máximos y mínimos de los datos de entrada, respectivamente. Ya con las variables normalizadas se procede a construir la red.

3.4.2 Creación de las redes ANFIS con diferentes FMs. Se implementó una serie de estructuras ANFIS, considerando los tres parámetros de diseño para este tipo de redes: el número de funciones de membresía, su tipo y la estrategia de aprendizaje a aplicar. El software Fuzzy Logic Toolbox TM de Matlab fue la herramienta utilizada para la implementación de ANFIS, ya

que proporciona funciones de línea de comandos y una aplicación para el entrenamiento de sistemas de inferencia difusa de tipo Sugeno que utilizan datos específicos para el aprendizaje de entrada / salida.

En la Figura 17 se muestra la estructura de la red ANFIS de tipo Sugeno con sus respectivas entradas y salidas. Las variables de entrada seleccionadas para esta investigación son las 4 variables meteorológicas que brinda el IDEAM en su base de datos. Adicionalmente, se selecciona como otra variable de entrada la hora en la que fue tomada la medición, esto con el fin de evaluar si la integración del parámetro puede llegar a optimizar los resultados de la red de estimación.

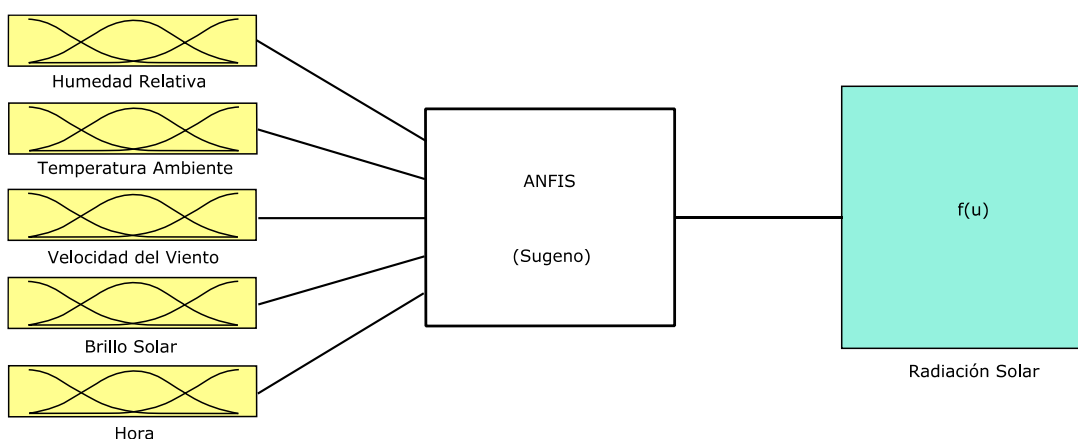
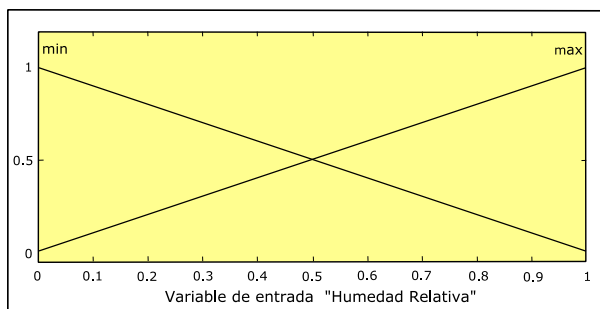
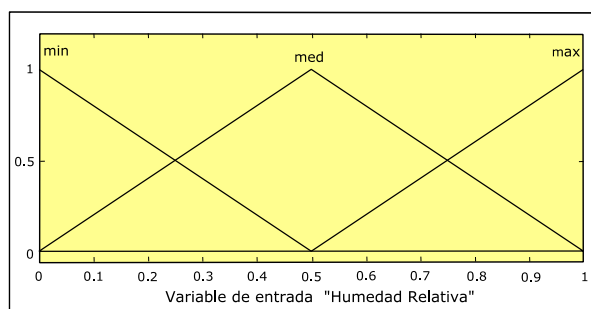


Figura 17. Caracterización de la red ANFIS

El número de funciones de membresía (FMs) en evaluación fueron 2 y 3, y los parámetros iniciales para el mapeo de las FMs fueron en un rango de [0,1]. Las figuras 18, 19, 20, 21 y 22 muestran gráficos de los diferentes tipos de funciones de membresía que fueron aplicados para la evaluación de la red; el objetivo es identificar cuál combinación arroja el error mínimo en el entrenamiento de la red.

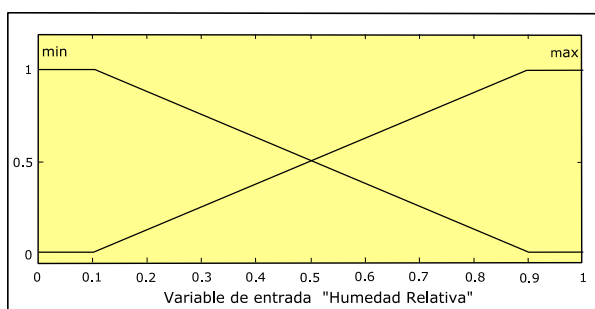


a.

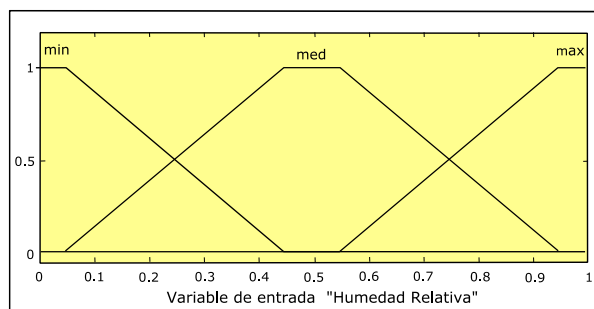


b.

Figura 18. Representación gráfica de la función de membresía en Matlab tipo: Triangular

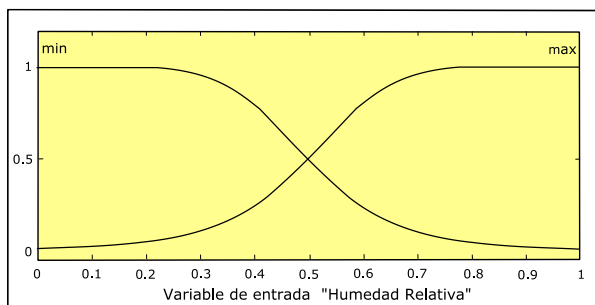


a.

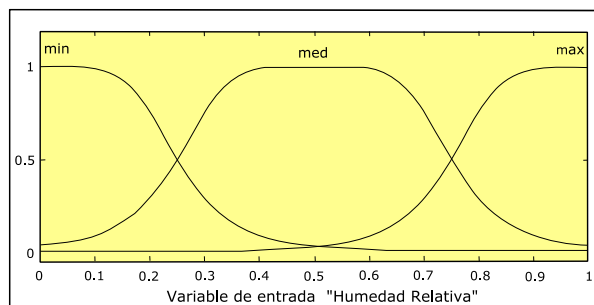


b.

Figura 19. Representación gráfica de la función de membresía en Matlab tipo: Trapezoidal



a.



b.

Figura 20. Representación gráfica de la función de membresía en Matlab tipo: Campana de Bell

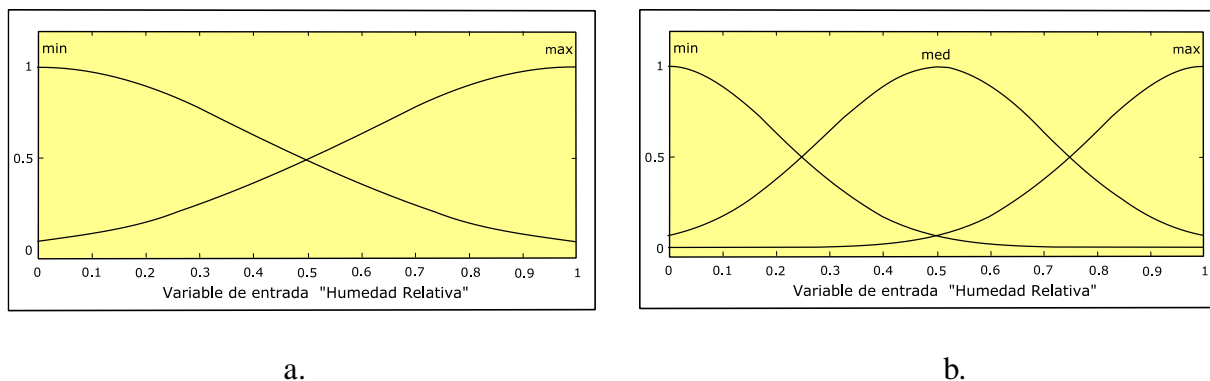


Figura 21. Representación gráfica de la función de membresía en Matlab tipo: Campana de Gauss

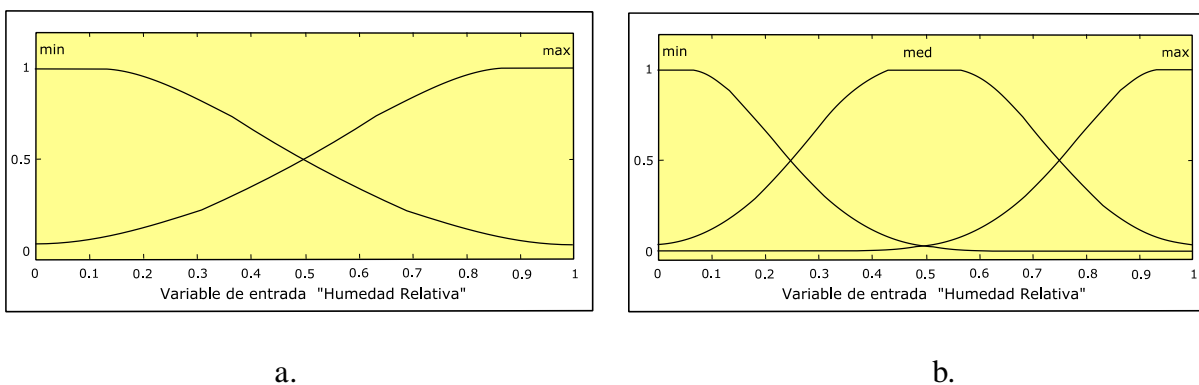


Figura 22. Representación gráfica de la función de membresía en Matlab tipo: Campana de Gauss 2

3.4.3 Entrenamiento de la red. Después de la asignación de las funciones de membresía con sus correspondientes parámetros de premisa se procede al entrenamiento de la red. Para ello se deben pre-establecer las reglas if-then (si – entonces) de acuerdo a las posibles combinaciones de los datos de entrada, asociando cada variable como una combinación estratégica para la predicción de la radiación solar global; esto se hace con el objetivo de que este proceso de inicialización mejore la actualización de los pesos en el entrenamiento del modelo ANFIS. El número de reglas posibles que regularán la red neuro difusa se define con la siguiente ecuación:

$$\#Reglas = FM^n$$

Donde FM es el número de funciones de membresía y n es el número de variables de entrada. Para dos y tres funciones de membresía, se obtuvo un total de 32 y 243 reglas. Con las reglas estructuradas se procede a entrenar la red; cabe mencionar que el número de reglas debe ser igual a las funciones de membresía de la salida. Del total de los datos, el 80% (18 576 de los 23 220 datos para cada variable) se utilizaron para el entrenamiento de la red.

En el entrenamiento, se implementaron dos estrategias de aprendizaje: una basada en el enfoque de propagación hacia atrás y otra en un algoritmo híbrido derivado (el cual combina el método gradiente descendiente con el estimador de mínimos cuadrados), que ofrece la Toolbox de Matlab. Se definió un número total de épocas con el fin de monitorizar el comportamiento del error RMSE en el transcurso de las épocas y definir cuando el error deja de disminuir significativamente. Con todos los parámetros ajustados se realizó el entrenamiento supervisado de la red.

3.4.4 Comparación y selección de la red. Al obtener los resultados del entrenamiento de cada configuración de la red, se realizó un proceso de comparación y selección. Los parámetros que se tuvieron en cuenta fueron los siguientes: tiempo de entrenamiento, número de épocas y valor de RMSE.

Los tres parámetros mencionados anteriormente fueron tomados al finalizar la etapa de entrenamiento. Luego se hace el proceso de evaluación y selección con el fin de identificar cual configuración está arrojando mejores valores de error en un menor tiempo y con menor cantidad de épocas transcurridas. Aquella configuración de red que cumpla de manera óptima estos tres indicadores, es seleccionada para realizar el proceso de verificación de los datos, cuya función de membresía y método de aprendizaje se denominará como el más apropiado para la solución del problema planteado.

3.4.5 Ejecución y validación de la red. Al finalizar el entrenamiento y comparación de las diferentes topologías de la red ANFIS, se definió cuál red es la que se ajusta más para la estimación de la radiación solar global. Con la red seleccionada se realizó el proceso de validación de la red con el 20 % de los datos faltantes, es decir, con 4 644 de los 23 220 registros pertenecientes a la base de datos. Este proceso se realiza para analizar el comportamiento de la red ante información totalmente desconocida para ella, es decir, que no recibe ningún dato de retroalimentación para evaluar si su salida es correcta o no. En este paso solo se usan las variables de entrada y se obtiene la correspondiente salida, que es comparada de forma externa con la variable real de la radiación solar.

Con el 20 % de los datos se procede a extraer los datos de salida pertenecientes a la red seleccionada por medio del comando “*output=evalfis(inputs,nombre_de_la_red)*” donde *output* es la salida que se desea obtener, *inputs* son las entradas del 20 % de los registros y *nombre_de_la_red* es la red previamente seleccionada que presentó el mejor desempeño en la fase de entrenamiento.

La red arroja los resultados en un rango de 0 a 1, debido al proceso de normalización de las variables de entrada explicado anteriormente. Por ello, se procede a convertir la salida de la red en el rango de operación de la radiación solar global.

3.5 Evaluación del desempeño de la red.

Para comparar los resultados de la red ANFIS con otros modelos en la literatura, se calcularon diferentes indicadores de error: RMSE, MAE, MAPE, MBE y R². Las siguientes ecuaciones muestran las expresiones matemáticas para cada indicador, respectivamente.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_{medido} - X_{estimado})^2} \quad (\text{Wh/m}^2 \text{ ó MJ/m}^2)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |X_{medido} - X_{estimado}| \quad (\text{Wh/m}^2 \text{ ó MJ/m}^2)$$

$$MAPE = \left(\frac{1}{N} \sum_{i=1}^N \left| \frac{X_{medido} - X_{estimado}}{X_{medido}} \right| \right) \times 100 \quad (\%)$$

$$MBE = \frac{1}{N} \sum_{i=1}^N (X_{medido} - X_{estimado}) \quad (\text{Wh/m}^2 \text{ ó MJ/m}^2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (X_{medido} - X_{estimado})^2}{\sum_{i=1}^N (X_{estimado})^2} \quad \text{Adimensional}$$

El MAE se define como una cantidad que se utiliza para medir qué tan cerca están los valores pronosticados con los valores medidos. El MBE es el error de pronóstico promedio que representa el error sistemático de un modelo, este error contrarresta los resultados negativos con los positivos. El RMSE indica el nivel de dispersión que produce el modelo ANFIS. Un RMSE más bajo indica que el modelo ANFIS desarrollado está teniendo una buena precisión de predicción [40]. R^2 de manera similar al indicador R , muestra la correlación entre los datos esperados y estimados. Los valores más grandes de R^2 indican una mayor coincidencia de tendencias en los datos medidos por los resultados del modelo [30], [41]. Finalmente, el indicador MAPE representa el porcentaje promedio de la diferencia entre los valores predichos y reales [42].

4. Resultados

En esta sección se presentan los resultados obtenidos en el proyecto, representados en gráficos y diagramas estadísticos.

4.1 Adquisición de datos de radiación solar

Los datos de radiación solar fueron seleccionados en un periodo de 10 años; al obtener la información de las tres bases de datos, éstas se comparan con el fin de verificar la cantidad de información que contiene cada una de ellas. Los resultados se muestran en la Figura 23 donde se comparan las tres fuentes de información según la cantidad de días por año, resaltando que los 365 días del año equivalen al 100%.

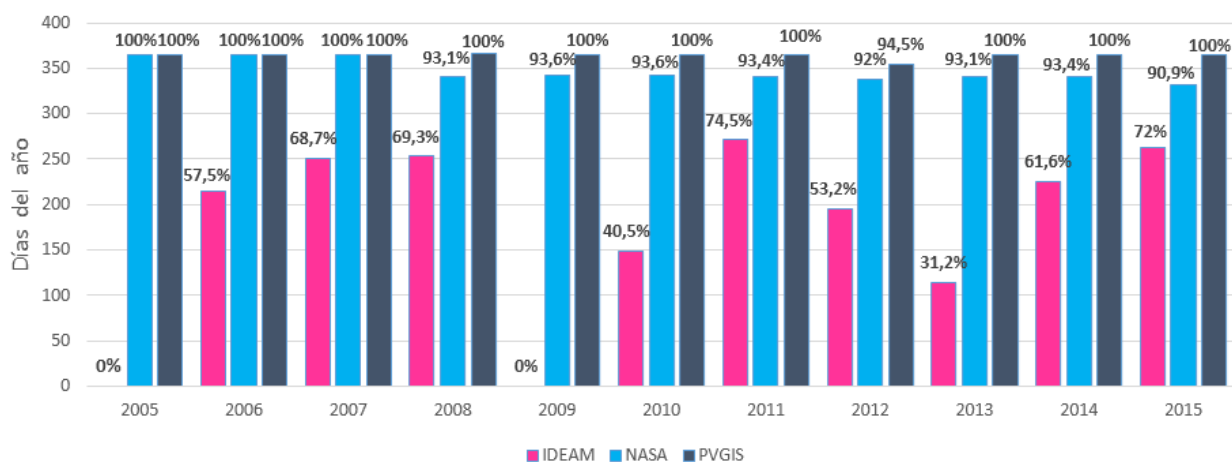


Figura 23. Conteo de datos anuales

Como se observa en la Figura 23, la cantidad de información suministrada por la base de datos del IDEAM es escasa en comparación con las bases de datos de la NASA y PVGIS. En los años 2005 y 2009 se presenta una ausencia total de las mediciones; por lo tanto, utilizar estos datos afectaría negativamente la precisión del análisis estadístico debido a que existe una menor cantidad de muestras para el estudio.

La base de datos de la NASA y la de PVGIS, en comparación con la del IDEAM, suministran información casi completa de la variable solar; un breve ejemplo de esto se muestra en la

Figura 24, haciendo notorio la ausencia de datos del IDEAM en el mes de enero pero que en su mayoría están presentes en las otras bases de datos, para el mismo mes durante todos los años hasta el 2015.

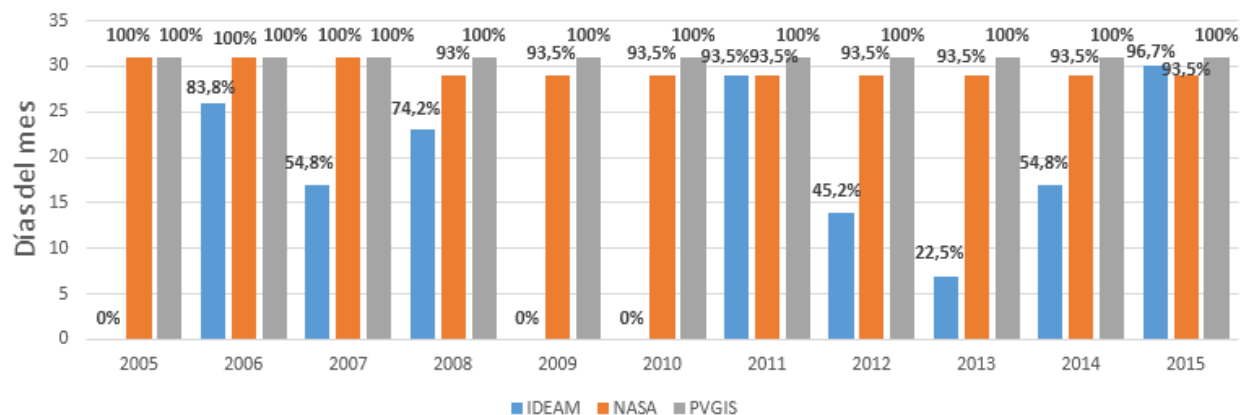


Figura 24. Conteo de datos del mes de enero

Teniendo en cuenta la información desplegada en las figuras 23 y 24, el propósito del análisis es identificar qué tan precisos son los datos suministrados por las bases de datos de la NASA y PVGIS con respecto a los datos del IDEAM, con el fin de aprovechar una mayor cantidad de datos. El análisis de similitud entre las bases de datos consiste en hallar diferentes coeficientes y errores estadísticos para validar la información. En la Tabla 3, se dan a conocer los resultados obtenidos de este análisis.

Tabla 3.

Análisis de similitud de las bases de datos

Error estadístico	IDEAM-NASA	IDEAM-PVGIS
Coef. Correlación	0.6431	0.8469
Coef. R^2	0.4486	0.7299
RMSE	1425.3 W/m^2	770.64 W/m^2
MAPE	23.12 %	12.63 %
MBE	13.45 %	1.31 %

El coeficiente de correlación mostrado en la Tabla 3, permite conocer el grado de asociación lineal entre dos variables cuantitativas demostrando que entre IDEAM-NASA existe un coeficiente de correlación moderado y entre el IDEAM-PVGIS hay una fuerte relación lineal. El coeficiente R^2 indica la calidad del modelo (PVGIS, NASA) para replicar los resultados de la referencia, que en este caso es el IDEAM [24].

Los resultados demuestran que la base de datos de PVGIS arroja un RMSE de 770.64 W/m^2 , siendo un valor inferior al de la NASA con un error de 1425.3 W/m^2 , demostrando que el PVGIS presenta un menor nivel de dispersión de los datos con respecto a los del IDEAM. El Mean Absolute Percentage Error (MAPE) mide la precisión del modelo a nivel porcentual, basando su análisis en el porcentaje promedio de la diferencia entre los valores reales (IDEAM) y pronosticados (NASA y PVGIS) [42], teniendo un error superior la NASA con un 23.12 %. El Mean Bias Error (MBE) es el error de pronóstico promedio que representa el error sistemático de un modelo; este error contrarresta los resultados negativos con los positivos, demostrando que la base de datos de PVGIS es más precisa con un MBE del 1.31 % con respecto a la NASA (13.45 %).

Los resultados presentados en la Tabla 3 indican que la base de datos de PVGIS es la que más se acerca a los valores del IDEAM con una calidad del modelo R^2 de aproximadamente 0.73. En comparación con la base de datos de la NASA, PVGIS arroja mejores resultados tanto en los coeficientes de correlación como en los errores estadísticos; estos resultados son los criterios para seleccionar a PVGIS como la base de datos para hacer los análisis de estadística inferencial detallados en las siguientes subsecciones.

4.2 Instrumentación para la medición y registro de datos

Una vez fue realizada la instalación del piranómetro y configurado el datalogger, se inicia el proceso de la toma de mediciones de la radiación solar global. El día 6 de diciembre del 2019 se inició el proceso de las mediciones tomando registros cada 5 minutos y siendo almacenados en la memoria interna del datalogger para eventualmente ir retirando su información. En la Figura 25 y Figura 26 se presentan un par de gráficas que define el comportamiento de la radiación solar en un día soleado y un día nublado.

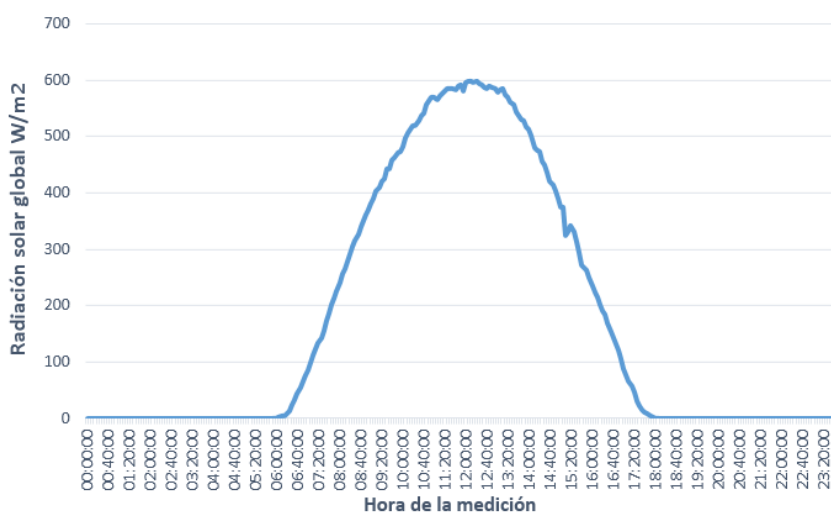


Figura 25. Datos medidos por el piranómetro a nivel diario para un día soleado

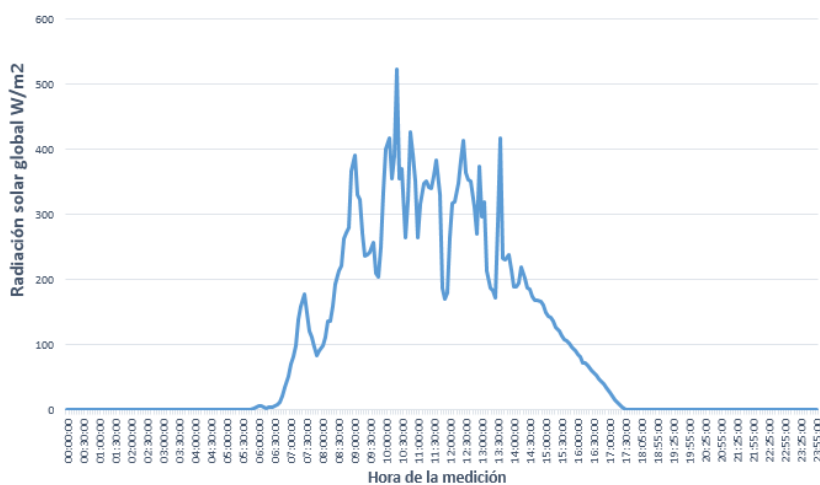


Figura 26. Datos medidos por el piranómetro a nivel diario para un día nublado

En el transcurso de un día en la ciudad de Cúcuta, la radiación solar tiene un crecimiento de 6 am a 12 pm y decrece entre las 12 pm y 6 pm; en un día soleado en la mayoría de los casos la radiación solar global no posee cambios abruptos, con un índice bajo de presencia de nubes como se muestra en la Figura 25. Sin embargo, para un día con presencia de las nubes la radiación solar puede presentar cambios abruptos en su potencia, a medida que una nube interrumpa la radiación solar directa que capta el piranómetro en conjunto con los demás tipos de radiación, mostrado en la Figura 26. Cabe mencionar que las fechas de medición para las figuras son: 5 de marzo del 2020 para la Figura 25 y 7 de diciembre del 2019 para la Figura 26.



Figura 27. Resumen de la irradiación solar en el mes de febrero

En la Figura 27 se ilustra un resumen del comportamiento de la irradiación solar en el mes de febrero. Con los datos tomados a una frecuencia de 5 minutos por el datalogger, se realizan las operaciones matemáticas pertinentes con el fin de obtener valores de irradiancia en los meses donde se encuentran los datos completos tomados de inicio a fin, esto debido a que se tomaron 3 meses de mediciones, de los cuales enero y febrero son los meses que poseen todos los días de mediciones. En la Tabla 4 se presenta un resumen de la irradiancia solar en estos dos meses; los registros fueron tomados desde el piranómetro instalado en los laboratorios de electrónica.

Tabla 4.

Resumen de irradiación solar para los meses de enero y febrero del 2020, dados en Wh/m².

Año	Día	Enero	Febrero
2020	1	2331.8	4093.4
2020	2	3528.2	3453.5
2020	3	3746.4	3672.0
2020	4	3845.7	3949.1
2020	5	2243.7	3981.0
2020	6	3512.3	4313.9
2020	7	3583.8	4166.5
2020	8	3589.1	3601.7
2020	9	3664.5	3426.2
2020	10	4079.9	4119.5
2020	11	1394.0	4124.5
2020	12	3672.7	3764.1
2020	13	1829.7	3024.9
2020	14	2821.9	1806.5
2020	15	3501.8	3750.2
2020	16	3519.0	3983.7
2020	17	2765.9	2600.1
2020	18	2466.3	4408.0
2020	19	2828.2	4132.9
2020	20	2975.5	2736.7
2020	21	2549.1	2579.1
2020	22	1260.6	2854.0
2020	23	3322.3	1800.6
2020	24	1404.0	2746.9
2020	25	2163.3	3012.7
2020	26	1096.2	3677.0
2020	27	3305.7	3216.1
2020	28	2973.2	2941.3
2020	29	3409.3	3147.2
2020	30	3209.8	-
2020	31	3227.4	-

De la Tabla 4 se evidencian días de muy baja radiación solar, debido a que estos meses presentaron días de alta nubosidad y precipitación. El promedio en horas solares pico (HSP) para

el mes de enero fue de 2.9 HSP y para el mes de febrero de 3.4 HSP, donde se denomina una hora solar pico como aquella que posee 1 kWh/m².

4.3 Análisis estadístico de la radiación solar

Conocer la variabilidad de la radiación solar global en una localidad determinada validará si la zona seleccionada es apta para la generación de energía renovable. Por tal motivo, el modelo estadístico ANOVA es seleccionado para evaluar si existe variabilidad significativa de la radiación solar global en la ciudad de Cúcuta, con el fin de establecer si existen patrones de comportamiento para proyecciones futuras de la radiación o si se deben agrupar una gran cantidad de datos históricos para evaluar el potencial solar que se pueda obtener en los años siguientes. En la Tabla 5 se dan a conocer los resultados del análisis de la varianza ANOVA para todos los años en general.

Tabla 5.

Análisis de la varianza de un solo factor

ANOVA	Df	Sum of Sq	Mean Sq	F Value	P
Factor (year)	9	7.198x10 ⁷	7.997x10 ³	5.128	6.28x10 ⁻⁷
Residuals	3641	5.579x10 ⁹	1.559x10 ³	-	-

Para realizar el análisis se adoptaron los datos de la última década disponible en la base de datos PVGIS, donde los datos del 2012 no se tienen en cuenta debido a que no se encuentran completos y podrían indicar resultados incorrectos.

En este orden de ideas, un valor p de 6.28×10^{-7} , significa que la radiación solar global ha variado en los últimos 10 años. Un valor F mayor que la unidad y un valor p mucho menor que 0.05, demuestra que al menos una de las medias poblacionales es diferente de las otras, procediendo a rechazar la hipótesis nula.

Teniendo en cuenta los resultados del análisis ANOVA se define que existe una importante variabilidad de la radiación solar en la última década y se procede a evaluar de forma más específica la evolución año a año de la radiación solar en la región bajo análisis. Lo anterior, por medio de un diagrama de cajas y bigotes ya que es una gráfica que describe características importantes del comportamiento de la muestra, tanto en términos de dispersión como de asimetría. Es un diagrama que suministra información sobre los valores máximos y mínimos, los cuartiles, la mediana y sobre la existencia de valores atípicos y la simetría de la distribución; en adición a esto, el diagrama indica el promedio de la radiación a través de rombos rojos.

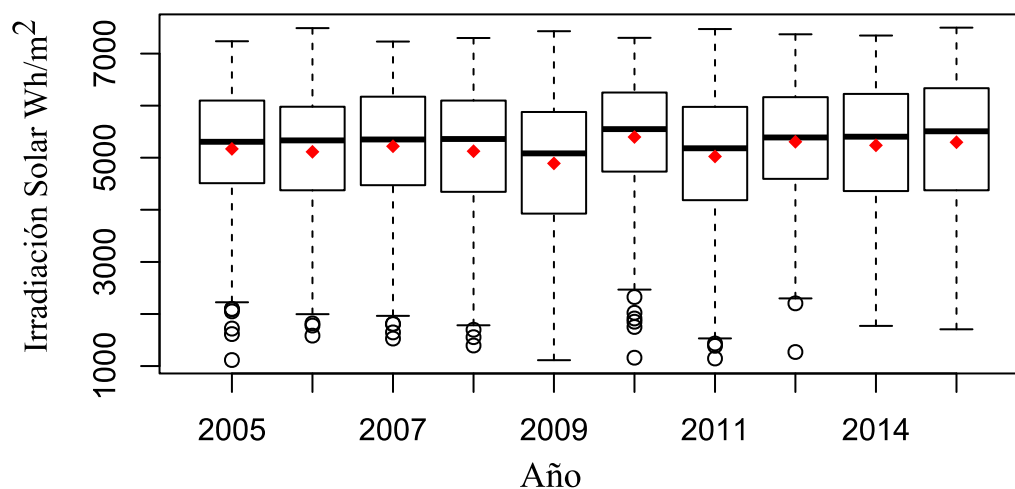


Figura 28. Resumen de la última década de radiación solar global en Cúcuta

En la Figura 28 se puede apreciar el resumen de los valores de radiación solar global para los años en estudio, donde la variable se mantiene aproximadamente constante a excepción de tres años. El valor promedio mínimo corresponde al año 2009, y el valor promedio máximo al año 2010; el rango intercuartílico en donde están contenidos el 50 % de los datos, se mantiene aproximadamente constante para los 10 años en estudio; el rombo rojo indica a donde apuntan todos los datos de cada año como media aritmética, la cual no está muy alejada de la mediana.

Desde otra perspectiva, la Figura 29 presenta el resumen del promedio anual de la radiación solar global mostrada en horas pico; el gráfico muestra con mayor claridad que los años 2009 y 2010 son los que poseen el menor y mayor valor en promedio de horas solares pico, respectivamente, y que los demás años no poseen una variación abrupta de la radiación.

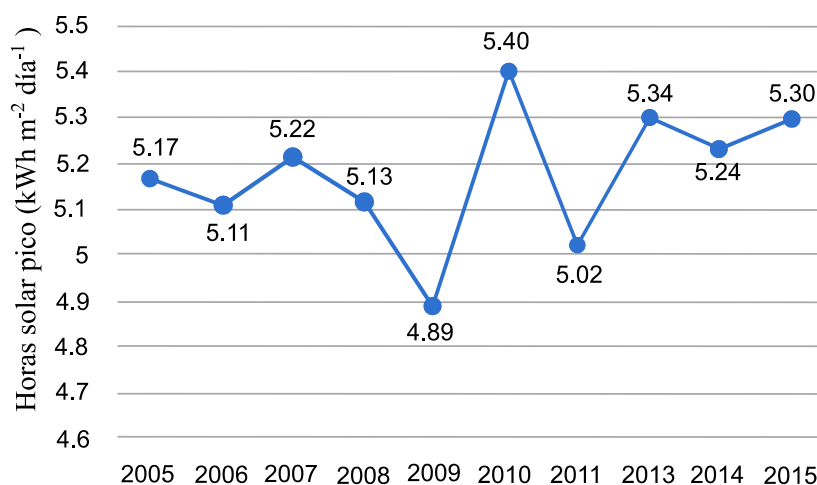


Figura 29. Resumen de la radiación solar en horas pico.

A pesar de que las figuras 28 y 29 presentan los fuertes kWh m⁻² día⁻¹ cambios en términos de radiación solar que se pueden presentar entre años para la ciudad de Cúcuta (soportando las primeras conclusiones del análisis ANOVA), se realiza el test de Bonferroni con el fin de validar cuáles años son los que presentan un valor p estadísticamente significativo. Teniendo en cuenta lo anterior se procede a realizar el test de Bonferroni, el cual está basado en la ecuación

$$\alpha' = \frac{\alpha}{m}$$

Donde el alfa prima (α') es el cociente entre el alfa nominal preestablecido como 0.05 y el número de pares posibles entre los tratamientos (m) [24]. Teniendo el nuevo valor de alfa, se realiza una prueba de hipótesis para identificar si existen valores de probabilidad menores que el alfa prima y proceder a rechazar la hipótesis nula (H_0).

La Tabla 6 muestra los resultados del test de Bonferroni, desarrollado por medio del software estadístico R Studio, que se encarga de ajustar el alfa prima para visualizar los resultados como si se estuviera evaluando sobre un alfa de 0.05.

Tabla 6.

Test de Bonferroni.

T. Bonferroni	2005	2006	2007	2008	2009	2010	2011	2013	2014	2015
2006	1.000	x	x	x	x	x	x	x	x	x
2007	1.000	1.000	x	x	x	x	x	x	x	x
2008	1.000	1.000	1.000	x	x	x	x	x	x	x
2009	0.117	0.713	0.017	0.500	x	x	x	x	x	x
2010	0.609	0.097	1.000	0.146	2×10^{-6}	x	x	x	x	x
2011	1.000	1.000	1.000	1.000	1.000	0.002	x	x	x	x
2013	1.000	1.000	1.000	1.000	0.36×10^{-3}	1.000	0.106	x	x	x
2014	1.000	1.000	1.000	1.000	0.008	1.000	0.918	1.000	1.000	x
2015	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

En la Tabla 6, las casillas sombreadas son los valores de probabilidad de mayor significancia, es decir, valores significativamente menores que $\alpha = 0.05$, notando que la mayoría de las casillas sombreadas tienen relación con el año 2009. El valor más pequeño y de mayor significancia está situado entre el 2009 y 2010 con un valor p de 2×10^{-6} ; este valor indica que la mayor diferencia se encuentra en este par de años lo que permite concluir que en el año 2009 se presenta la mediana más baja y en el 2010 la mediana más alta, así como lo demuestra la Figura 28.

De acuerdo con estos resultados surgen las siguientes preguntas ¿a qué se debe la variación de la radiación solar en los años 2009 y 2010? ¿Cuáles son los meses en los que se presenta la mayor y menor presencia de radiación solar?

En los años 2009, 2010 y 2011 ocurrieron fenómenos naturales conocidos como fenómeno de El Niño y fenómeno de La Niña, también conocido como fenómeno del niño oscilación sur (ENOS), los cuales afectaron la gran mayoría del territorio colombiano. Esto justifica que el valor promedio de radiación solar más bajo se encuentra en el año 2009 debido a que en los

primeros tres meses del año el fenómeno de La Niña estuvo presente en la región Andina (región de Colombia donde está localizada la ciudad de Cúcuta), como es evidenciado en la Tabla 7.

Tabla 7.

Porcentaje de precipitación en la región Andina, Colombia. Tomado de [43]

2009	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
Por debajo de lo normal (<90%)	1	2	1	76	74	25	78	39	97	67	61	69
Normal (90 - 100%)	5	5	8	21	19	24	14	30	2	23	25	17
Por encima de lo normal (>110%)	93	93	91	4	7	51	7	30	1	9	13	13

En la tabla se clasifican los meses significativos con anomalías de precipitación “por encima de lo normal” o “por debajo de lo normal”. El criterio para definir un mes significativo es que el área afectada fuese igual o superior al 50 % del total de territorio de la región natural respectiva. Así se representan en la tabla con color rosa los casos significativos “por encima de lo normal”, y en color azul los casos “por debajo de lo normal”.

Los primeros tres meses del año 2009 indican una presencia de precipitación mayor al 90 %, esto ocasionó un decaimiento en el valor promedio de la radiación solar global, generando valores por debajo de los 3000 Wh/m² por día, estos valores tan bajos son la razón por la cual el año 2009 presenta la mediana más baja de radiación solar global en comparación con los demás años.

Por otra parte, el fenómeno de El Niño se acentuó en los primeros 4 meses del 2010, esto ocasionó una mayor ausencia de la precipitación, generando para el territorio en estudio, un aumento de las horas solar pico, donde se encontraron valores máximos de horas solares pico por encima de los 6000 Wh/m² por día, como se muestra en la Figura 30.

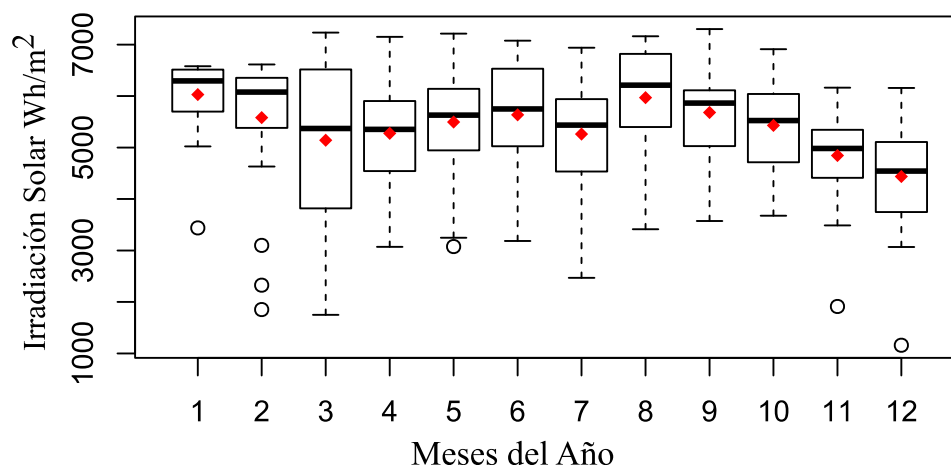


Figura 30. Evaluación de la variabilidad mensual de radiación solar global en el año 2010

En el segundo semestre del año se logra visualizar un decremento de la mediana y el promedio debido a que el fenómeno de La Niña vuelve a entrar en rigor afectando también los valores promedios de radiación solar global del año 2011, el cual fue el segundo año con menor valor de dicha variable, mostrado en la Figura 29.

Adicionalmente, durante la última década disponible en la base de datos del PVGIS se logra hacer un estudio detallado de la información enfocándose en los meses; cada mes es sometido al análisis de varianza de un solo factor (ANOVA), donde se determina el valor p y el valor F para así identificar en qué meses la hipótesis nula fue rechazada y verificar cuáles meses del año presentan mayor dispersión en sus datos; estos resultados se muestran en la Tabla 8.

Tabla 8.

Análisis de la varianza de todos los meses del año

ANOVA	Ene	Feb	Mar	Abr	May	Jun
Valor P	1.53×10^{-7}	4.43×10^{-8}	1.38×10^{-12}	0.004	0.828	0.009
Valor F	5.866	6.288	9.390	2.749	0.562	2.491

a)

ANOVA	Jul	Ago	Sep	Oct	Nov	Dic
Valor P	0.219	0.228	0.046	0.019	0.102	0.001
Valor F	2.202	1.316	1.939	2.242	1.647	3.149

b)

Basado en los resultados de la Tabla 8, los meses que tienen mayor dispersión en sus valores según el test de ANOVA son enero, febrero y marzo; teniendo en cuenta que el mes que posee una mayor dispersión en sus medidas es el mes de marzo, indicado en la Figura 30 con un mayor rango intercuartílico y comprendido desde aproximadamente 2000 Wh/m² y por encima de 7000 Wh/m². En las figuras 31 y 32 se muestra claramente cómo fue el comportamiento de la radiación solar global a nivel mensual en la última década.

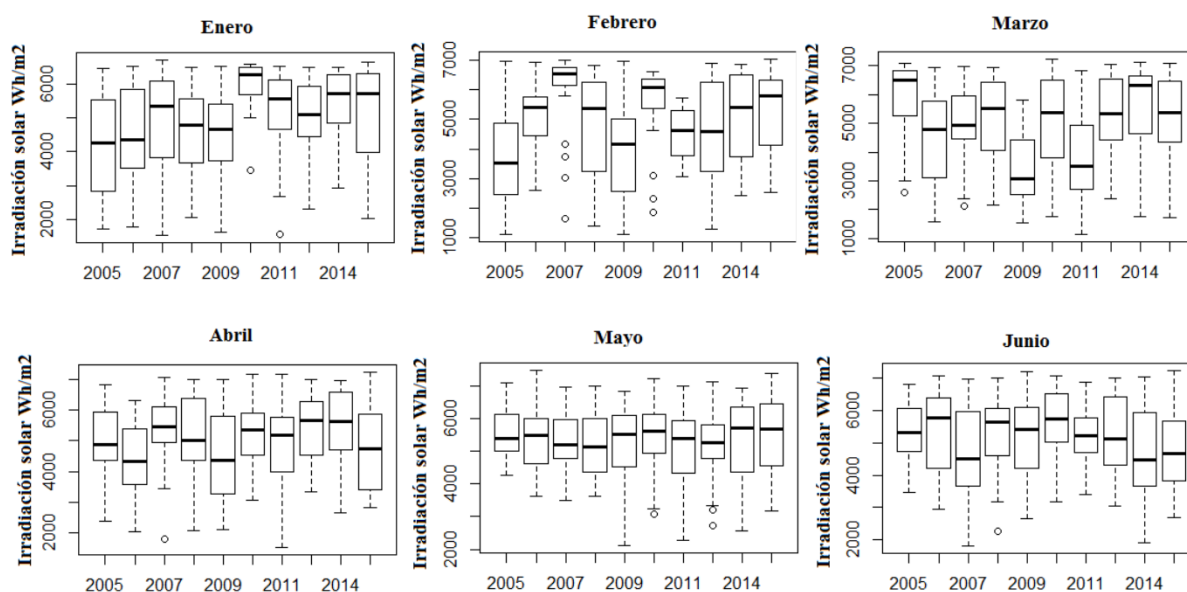


Figura 31. Resumen de la radiación solar en el primer semestre del año de la última década

La Figura 31 muestra un resumen estadístico de la radiación solar global para los primeros seis meses, para la elaboración de este diagrama de caja y bigotes se recopiló y agrupó la información mensualmente de los registros comprendidos en la última década puesta en estudio. Asimismo, se realizó el mismo procedimiento para los últimos seis meses mostrado en la Figura 32.

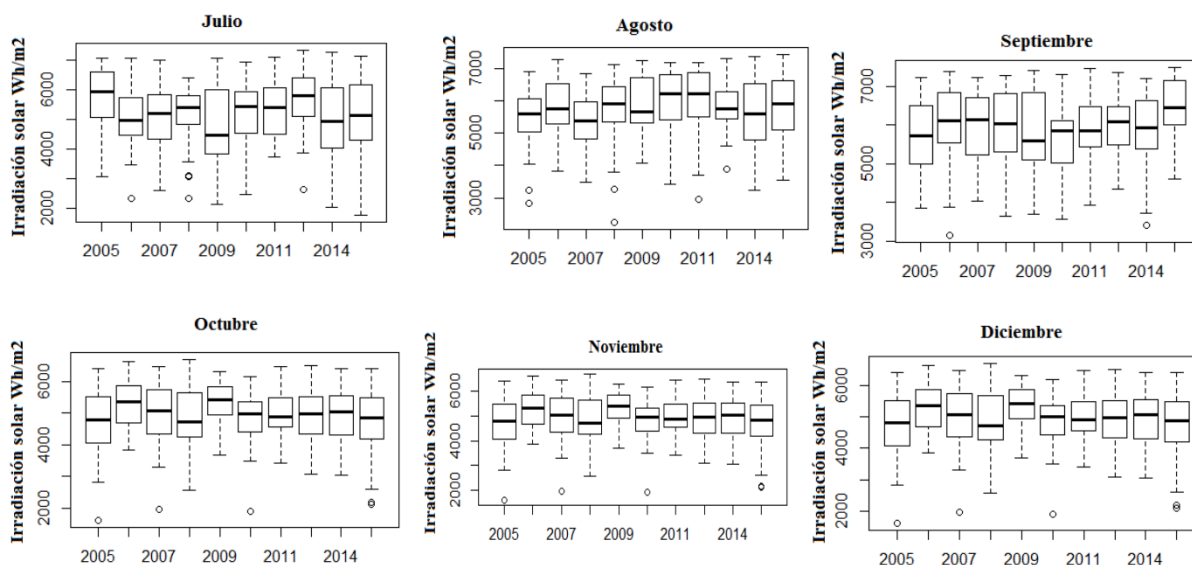


Figura 32. Resumen de la radiación solar en el segundo semestre del año de la última década

En consecuencia, se procede a revisar cuáles son los meses con los valores promedios máximos y mínimos de radiación solar global. La Figura 33 enseña una breve descripción de los niveles de radiación solar global promedio en cada mes, arrojando resultados en Wh/m^2 por día.

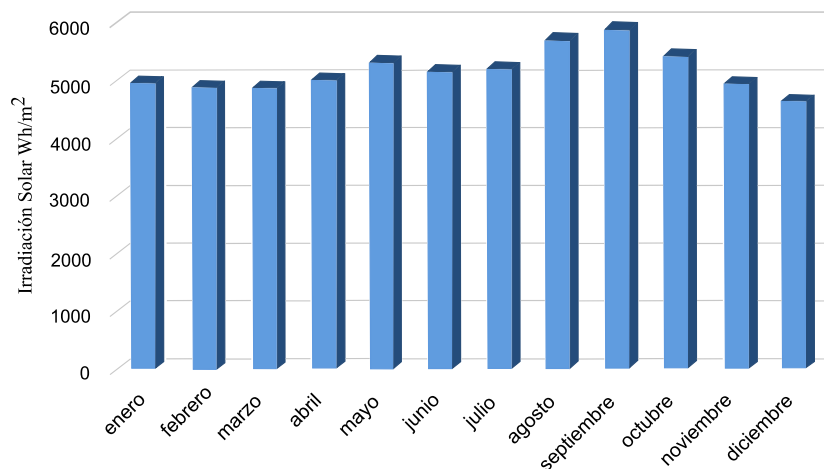


Figura 33. Resumen mensual de la radiación solar global

De la figura se puede identificar al mes de septiembre como el mes con mayor registro de insolación con un promedio de 5916.95 Wh/m^2 por día, y a diciembre como el mes con menor registro de insolación con un promedio de 4652.63 Wh/m^2 por día.

4.4 Creación de la estructura ANFIS

Con el 80 % de los datos se realizó el proceso de entrenamiento para cada topología de la red ANFIS. En la Figura 34 se muestra un ejemplo del proceso de entrenamiento con algoritmo de aprendizaje de propagación hacia atrás para una red ANFIS de dos funciones de membresía.

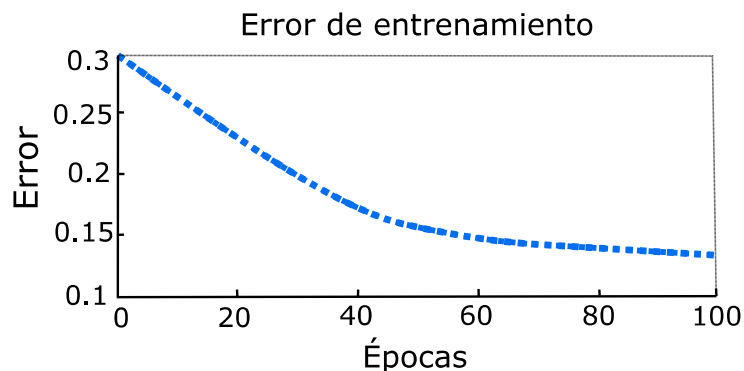


Figura 34. Entrenamiento de una red ANFIS con 2 FMs

De la gráfica se puede observar que a medida que transcurren las 100 épocas, el error RMSE tiende a disminuirse significativamente; de la época 1 a la 50 se logra ver un notorio decrecimiento del error y a medida que transcurren las épocas después de la época 50 el RMSE sigue disminuyendo, pero en pequeñas proporciones hasta que llega a un punto de convergencia.

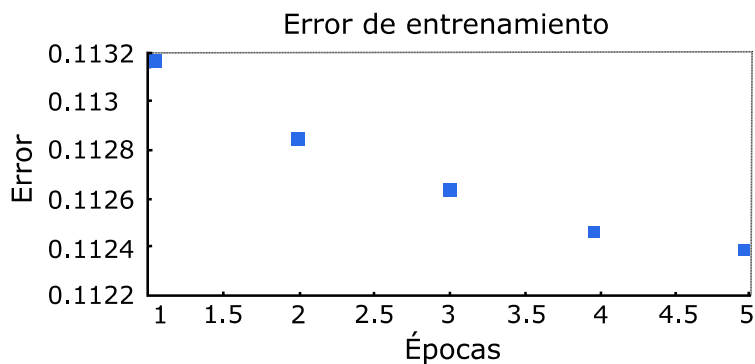


Figura 35. Entrenamiento de una red ANFIS con 3 FMs

La Figura 35 muestra el entrenamiento de una red ANFIS de 3 funciones de membresía con algoritmo de aprendizaje híbrido. De la figura se observa que sigue el mismo comportamiento de decrecimiento de la Figura 34, sin embargo, el modelo de entrenamiento con tres funciones de

membresía hace que el tiempo de transición de una época a otra sea más prolongado; por otra parte, desde la etapa 1 el error es menor a diferencia del entrenamiento con 2 FMs.

Después de realizar las pruebas pertinentes de entrenamiento para todas las topologías de la red, se procedió comparárlas de acuerdo con su desempeño. En las Tablas 9 y 10 se muestra un resumen de los resultados obtenidos del entrenamiento de la red con 2 y 3 FMs, además representan el comportamiento del error RMSE de acuerdo con cada tipo de FMs y algoritmo de aprendizaje.

Tabla 9.

Entrenamiento ANFIS con 2 funciones de membresía

Parámetro	ANFIS 1	ANFIS 2	ANFIS 3	ANFIS 4	ANFIS 5	ANFIS 6	ANFIS 7	ANFIS 8	ANFIS 9	ANFIS 10
Tipo FM	Triangle	Triangle	Gaussian	Gaussian	Gaussian2	Gaussian 2	Gbell	Gbell	Trapezoidal	Trapezoidal
Algoritmo Aprendizaje	Hybrid	Back Propag	Hybrid	Back Propag	Hybrid	Back Propag	Hybrid	Back Propag	Hybrid	Back Propag
Nº Épocas	66	100	25	150	52	131	50	78	30	116
Tiempo	18min	5min	6min	7min	11min	7min	14min	6min	9min	7min
RMSE	0.1142	0.1339	0.1132	0.1332	0.1128	0.1334	0.1129	0.1340	0.1141	0.1272

Según los resultados obtenidos por la Tabla 9 y la Tabla 10, se realiza el proceso de evaluación y selección de la red. Teniendo en cuenta el tiempo de entrenamiento respecto al número de épocas, las redes que cuentan con 2 FMs llegan a su punto mínimo de error en un menor tiempo; sin embargo, las redes con tres FMs presentaron un mejor rendimiento en el error con un número menor de épocas, aunque con un tiempo considerablemente mayor.

Tabla 10.

Entrenamiento ANFIS con 3 funciones de membresía

Parámetro	ANFIS 1	ANFIS 2	ANFIS 3	ANFIS 4	ANFIS 5	ANFIS 6	ANFIS 7	ANFIS 8	ANFIS 9	ANFIS 10
Tipo FM	Triangle	Triangle	Gaussian	Gaussian	Gaussian2	Gaussian2	Gbell	Gbell	Trapezoidal	Trapezoidal
Algoritmo Aprendizaje	Hybrid	Back Propag	Hybrid	Back Propag	Hybrid	Back Propag	Hybrid	Back Propag	Hybrid	Back Propag
N° Épocas	5	5	5	5	5	5	5	5	5	5
Tiempo	2h 0min	42min	1h 57min	26min	1h 54min	40min	2h 2min	21min	2h 3min	41min
RMSE	0.1117	0.3117	0.1106	0.3473	0.1106	0.3005	0.1111	0.3006	0.1117	0.3035

Considerando los resultados, se puede observar que el método de optimización híbrido genera mejores resultados para las redes de 2 y 3 FMs. Por lo tanto, se tienen en cuenta las redes ANFIS que trabajen bajo el algoritmo de aprendizaje híbrido y se escoge aquella que presente un RMSE mínimo. Para este caso la red de inteligencia artificial que cumple con todos los parámetros y presenta un RMSE menor es la que se encuentra sombreada en la Tabla 10, cuyos parámetros son: tres funciones de membresía, modelo de optimización híbrido, 5 épocas, tiempo de ejecución 1 hora con 57 minutos y un RMSE de 0.1106. Este modelo presenta el valor mínimo de RMSE de todas las topologías y es escogido como la configuración esencial para el modelo de estimación de radiación solar.

Después de la fase de entrenamiento se presentan nuevos parámetros en las funciones de membresía, con los cuales la red ANFIS se ajusta con el objeto de llegar a su mejor punto de optimización, estos parámetros son llamados parámetros de consecuencia como es mencionado en el marco teórico y se muestran en la Tabla 11. Cabe mencionar que los parámetros de premisa preestablecidos para la estructura ANFIS con tres funciones de membresía tipo Campana de Gauss son los siguientes: Min (0.2123, 0), Med (0.2123, 0.5) y Max (0.2123, 0).

Tabla 11.

Parámetros de consecuencia de la red ANFIS

Variable de entrada	Mínimo		Medio		Máximo	
Humedad Relativa	0.2096	0.0034	0.2087	0.5073	0.2386	0.9862
Temperatura Ambiente	0.2348	0.0128	0.2076	0.494	0.209	0.9958
Velocidad del Viento	0.2002	-0.0013	0.1978	0.5081	0.2244	0.9933
Brillo Solar	0.2143	0.001	0.2126	0.502	0.213	1.002
Hora	0.2043	-0.0039	0.2086	0.4969	0.2166	0.997

Asimismo, en la Figura 36 se muestra una representación gráfica de la modificación de los parámetros de premisa en las funciones de membresía. Se presenta como ejemplo la variable de entrada humedad relativa.

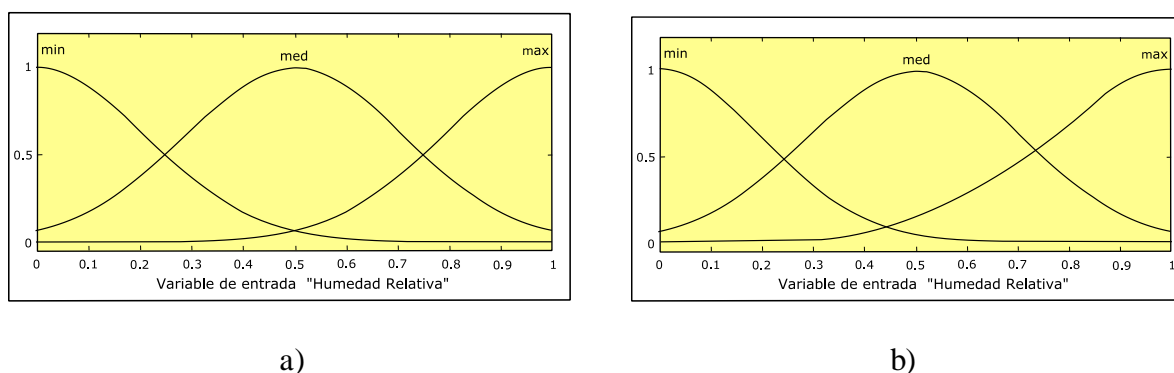


Figura 36. Modificación de la función de membresía: a) Parámetros de premisa b) Parámetros de consecuencia

Una vez seleccionada la red ANFIS definida como ANFIS 3 en la Tabla 10, se realiza un proceso de verificación de los datos del cual se escogieron aleatoriamente 84 registros de las variables de entrada y salida con el fin de evaluar el desempeño de la red. Los resultados se aprecian en la Figura 37. De la figura se observa que las cruces azules representan la salida de la

red de estimación y los asteriscos rojos representan los datos reales, esto indica la existencia de una alta correlación entre los datos experimentales y el modelo de estimación; indicando que la red ANFIS seleccionada está respondiendo satisfactoriamente a la solución del problema.

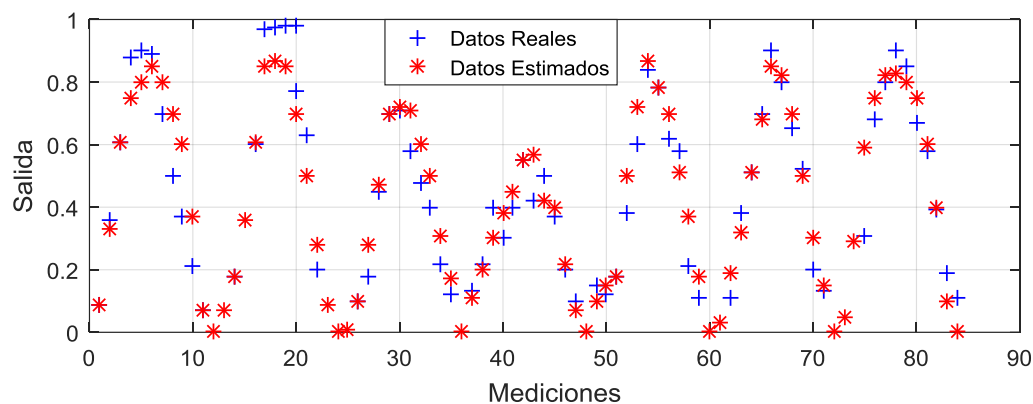


Figura 37. Comparación entre el desempeño del sistema y el conjunto de datos.

Por último, en la Figura 38 se enseña la correlación existente entre algunas de las variables de entrada y la salida de la red, específicamente, como ejemplo se muestra la humedad relativa y el brillo solar como entradas evaluadas del modelo. Además, se evidencia la influencia que presentan las variables en el comportamiento de la radiación solar.

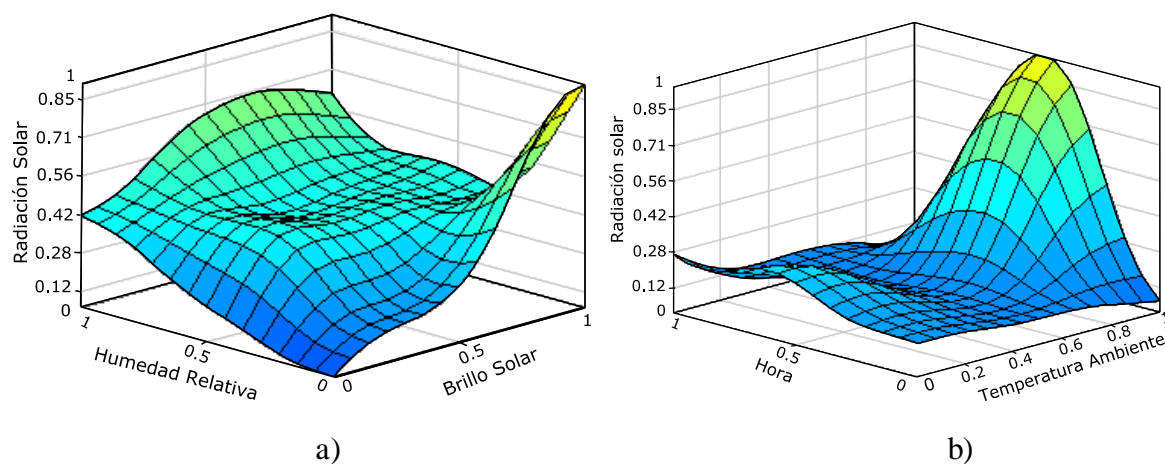


Figura 38. Relación entre los datos de entrada y salida del análisis cualitativo: a) Humedad relativa vs Brillo Solar b) Hora vs Temperatura ambiente.

A partir de las gráficas, se puede identificar que la temperatura y el brillo presentan una relación directamente proporcional con la variable de salida del modelo. Por otro lado, la humedad no presenta una influencia tan significativa como la del brillo solar, indicando que la humedad tiene un efecto moderado sobre la precisión de la salida del modelo.

4.5 Evaluación del desempeño de la red

El rendimiento del modelo utilizado en este estudio se calculó por medio de errores y coeficientes estadísticos. La finalidad de la evaluación es medir el desempeño de la red de estimación con valores de entrada desconocidos por el modelo de estimación. En la Tabla 12 se muestran los resultados obtenidos de la evaluación de la red; estos resultados se obtuvieron aplicando las ecuaciones mencionadas en la metodología por medio del software Matlab, donde la variable medida se obtiene del 20 % de los registros de la base de datos destinados para la validación, y la variable estimada es la salida de la red ANFIS.

Tabla 12.

Evaluación del desempeño de la red ANFIS

Indicador de error	Resultado
MAPE (%)	27.8
MAE (Wh/m ² - MJ/m ²)	97.93 - 0.352
RMSE (Wh/m ² - MJ/m ²)	142.23 - 0.512
MBE (Wh/m ² - MJ/m ²)	22.14 - 0.079
R ²	0.9115

Aunque los datos originales proporcionados por el IDEAM expresan la irradiación en Wh/m², la Tabla 12 también muestra los datos en MJ/m² para fines de comparación, ya que esta unidad

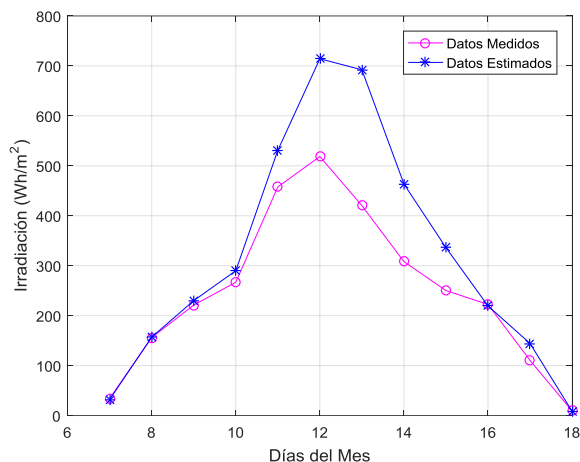
también es común en la literatura. Según el Sistema Internacional de Unidades, la equivalencia entre estas unidades es $1 \text{ Wh} = 3.6 \times 10^{-3} \text{ Mega Joules}$.

Según [44], en aplicaciones prácticas, $\text{MAPE} \leq 10\%$ significa alta precisión, $10\% \leq \text{MAPE} \leq 20\%$ significa buena predicción, $20\% \leq \text{MAPE} \leq 50\%$ significa predicción razonable y $\text{MAPE} \geq 50\%$ significa predicción inexacta. Teniendo en cuenta lo anterior, el modelo podría realizar estimaciones con una predicción razonable.

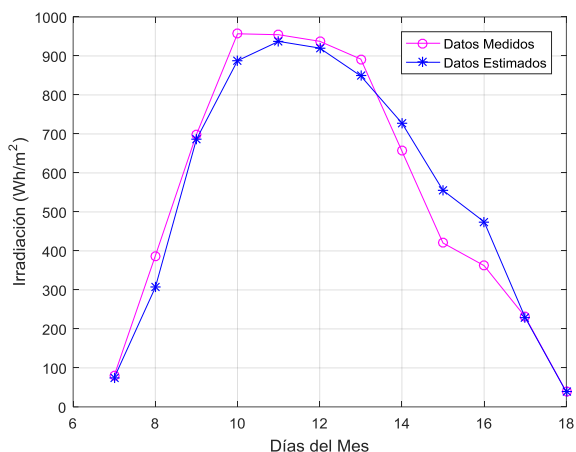
En general, el coeficiente de determinación (R^2) representa el porcentaje de datos más cercano a la línea de mejor ajuste del conjunto de datos general [40]. Los resultados indican claramente que el 91 % de los valores de radiación solar global pronosticados están muy cerca de los valores medidos según el modelo ANFIS elegido.

Los valores más pequeños de RMSE y MAE implican una aproximación más cercana de los valores medidos por los modelos. Los valores R^2 más grandes indican una coincidencia más fuerte de tendencias en los datos medidos por los resultados del modelo. El RMSE y el MAE son métricas básicas ampliamente utilizadas para evaluar el desempeño de los modelos de predicción [30]. Según lo mencionado anteriormente, con un $\text{RMSE} = 142.23 \text{ Wh/m}^2$, $\text{MAE} = 97.93 \text{ Wh/m}^2$ y $R^2 = 0.9115$, el modelo de inteligencia artificial estaría cumpliendo con los parámetros para considerarse como un buen modelo para replicar.

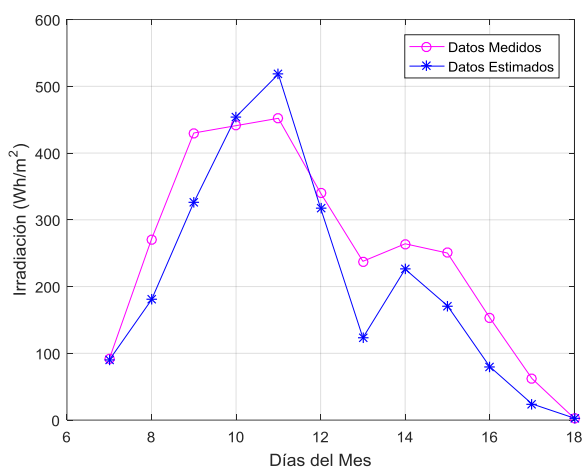
Se genera una comparación de los datos medidos y estimados en la Figura 39 para las ciudades consideradas con cuatro días aleatorios en los meses de mayo, agosto, noviembre y diciembre de los últimos años proporcionados por el IDEAM. Como no hay estaciones en Colombia, estos meses se seleccionaron tratando de incluir diferentes escenarios durante todo el año.



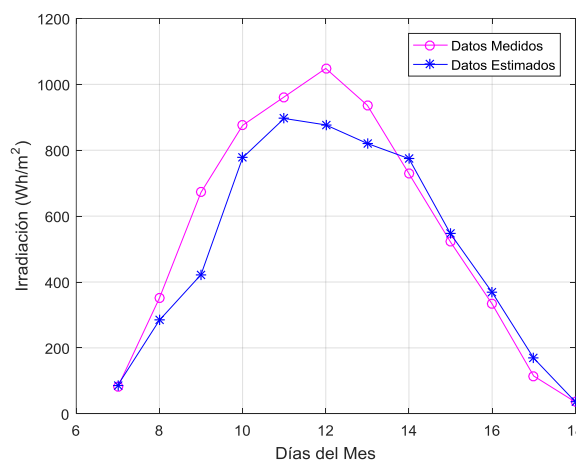
a)



b)



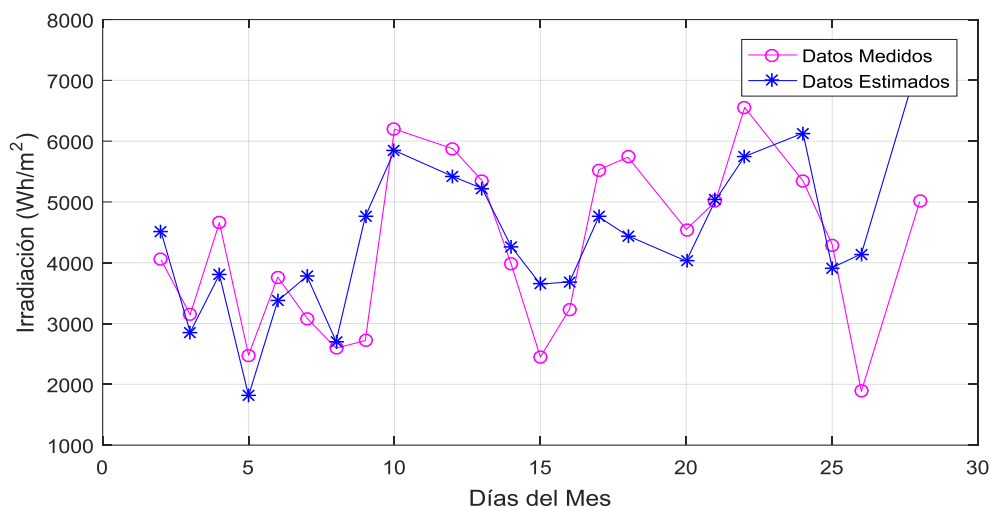
c)



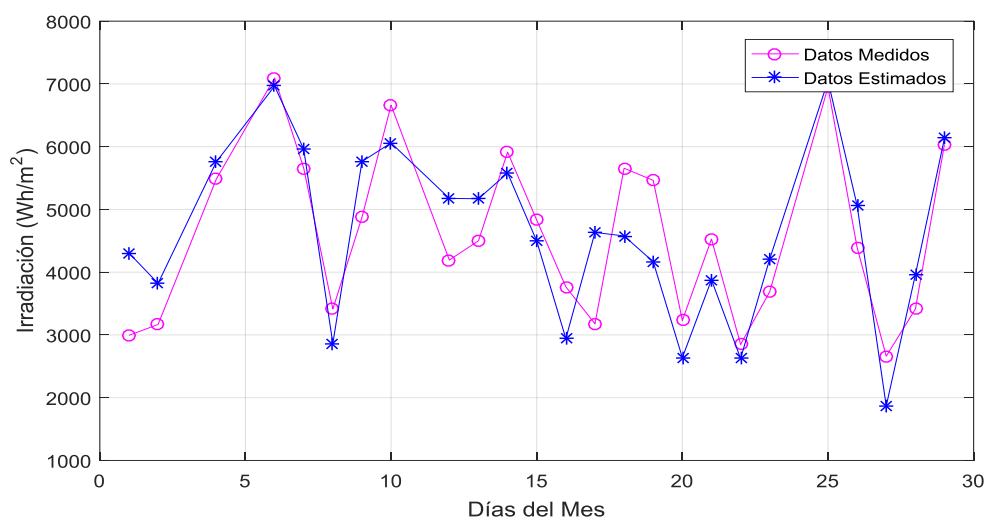
d)

Figura 39. Datos Medidos vs. Estimados para el modelo de estimación por hora en la ciudad de Cúcuta: a) mayo 17, 2015. b) agosto 22, 2014. c) noviembre 16, 2014. d) junio 10, 2015.

Una comparación similar de la Figura 39 para la estimación por hora se realiza en la Figura 40, para comparar los resultados de manera diaria, se utiliza un par de meses con los datos proporcionados por el IDEAM. Se escogió el año 2015 para la extracción de los datos debido a que es el año que presenta la cantidad mínima de datos faltantes con respecto a los otros años.



a)



b)

Figura 40. Datos Medidos vs. Estimados para el modelo de estimación por día en la ciudad de Cúcuta: a) abril, 2015. b) junio, 2015.

Por otro lado, los resultados en términos de los indicadores de error de algunos trabajos en la literatura para los diferentes rangos de estimación se presentan en la Tabla 13, y se destacan en azul como indicador de rendimiento los estudios de investigación donde el error es igual o superior a los presentados en el trabajo actual.

Tabla 13.

Características de modelos ANFIS de la literatura

Referencia	Ciudad - país	Variables de entrada	Indicadores de error
[30]	Yucatán, México	Tmax, Tmin, radiación solar extraterrestre, días lluviosos	R ² =0.645 RMSE(MJ/m ²)=2.801 MAE(MJ/m ²)=2.153
[16]	Argel, Argelia	Temperatura ambiente, brillo solar, radiación solar total diaria	R ² =0.98
[45]	Jodhpur, India	Radiación solar global, ángulo incidente, ángulo cenital, insolación extraterrestre	R ² =0.978 RMSE(MJ/m ²)= 0.026 MBE(MJ/m ²)=4.47x10 ⁻⁴
[46]	India	Radiación solar directa, radiación solar difusa, temperatura ambiente	RMSE(MJ/m ²)=0.071 MAE(MJ/m ²)=0.056 MAPE(%)=5.130
[47]	Isfahan, Iran	Tmax, brillo solar, brillo solar máximo	R=0.989 RMSE(MJ/m ²)=2.353
[48]	Argelia	Latitud, longitud, altitud	RMSE(MJ/m ²)=0.0215 MAPE(%)=2.2
[49]	Tamilnadu, India	Radiación solar global, latitud, longitud	R ² =0.98 RMSE(MJ/m ²)=0.0078 MBE(MJ/m ²)=1.031
[15]	Iseyin, Nigeria	Tmax, Tmin, brillo solar	R ² =0.6567 RMSE(MJ/m ²)=1.758
[50]	Kerman, Irán	Brillo solar, radiación solar global, radiación solar extraterrestre	RMSE(MJ/m ²)=1.3052 MBE(MJ/m ²)=1.017 MAPE(%)=23.052

Cabe mencionar que usar los indicadores de error solo para seleccionar qué modelo es mejor que otro podría ser irresponsable considerando que cada modelo está entrenado con condiciones y objetivos muy diferentes. En consecuencia, la comparación en la Tabla 13 es solo una descripción general destinada a mostrar el comportamiento de modelos similares.

5. Conclusiones

Con la presente investigación se puede inferir que por medio de la aplicación del test de Bonferroni y el modelo ANOVA, la radiación solar global en la ciudad de Cúcuta, Norte de Santander, en promedio, permanece constante y no posee mucha variabilidad alrededor de los años. Dicha variable empieza a tener fluctuaciones por encima y por debajo del promedio, dependiendo si el territorio está siendo sometido a algún fenómeno natural (El Niño y La Niña); es decir, la radiación solar en la ciudad de Cúcuta permanece aproximadamente constante si y solo si no ocurren fenómenos naturales de gran magnitud.

Los meses que poseen un mayor porcentaje de radiación solar global son agosto y septiembre, según los estudios realizados en la última década; en promedio poseen al día más de 6 horas pico, y el mes que posee un menor porcentaje de radiación solar es diciembre, en promedio se registra un valor diario menor a las 5 horas solares pico.

La red adaptativa de inferencia neuro difusa (ANFIS) que fue propuesta en esta investigación, utiliza cinco variables meteorológicas de entrada (humedad relativa, temperatura ambiente, velocidad del viento, brillo solar y hora de la medición) para estimar la radiación solar global en la ciudad de Cúcuta. Lo anterior como una herramienta adicional de información de la variable en mención, teniendo las condiciones actuales de las bases de datos de la región. El modelo fue evaluado y responde con error estadístico $RMSE = 142.23 \text{ Wh/m}^2$ y un coeficiente de determinación $R^2 = 0.9115$. Estos valores definen al modelo como un modelo de estimación de buena precisión.

El uso de modelos de inteligencia artificial para la estimación de la radiación solar en sectores del territorio colombiano donde no se posee una amplia cantidad de estaciones meteorológicas en funcionamiento para la adquisición de registros de este parámetro, permite que a partir de otras

variables de fácil acceso se puedan determinar los valores de radiación solar global. Esto representa una ventaja importante para el despliegue de este tipo de modelos en la región, porque su aplicación garantiza un alto nivel de precisión y confiabilidad.

Recomendaciones

Con el modelo propuesto se pueden completar las brechas de valores no medidos de radiación solar en la base de datos del IDEAM. Asimismo, se puede actualizar la base de datos para los siguientes años después del 2015.

Con los datos medidos del piranómetro instalado se podría hacer una adaptación para que dichos datos puedan ser observados desde una página web en tiempo real.

Para realizar una evaluación más exhaustiva del modelo de estimación, este se podría poner a prueba con una implementación física que mida en tiempo real todas las variables del estudio, por medio de sensores que puedan medir las variables de entrada (humedad relativa, temperatura ambiente, brillo solar, velocidad del viento, hora) y simultáneamente la variable de salida. Esto con el fin de tener una mayor precisión de la medición de los datos para poder identificar si la red de inteligencia artificial está trabajando de la manera más precisa en sus estimaciones.

Al modelo propuesto se le pueden añadir otras variables de entrada como (latitud, longitud y altitud) con el cual pueda ser entrenado con datos de otra ciudad o región y validarlo con datos reales de radiación solar. Esto con el fin de poder crear un modelo que se pueda ajustar a cualquier localidad independientemente de las condiciones climáticas.

Referencias

- [1] Instituto de Hidrología Meteorología y Estudios Ambientales de Colombia [IDEAM], “Atlas de Radiación Solar, Ultravioleta y Ozono de Colombia,” 2016.
- [2] T. Markvart, *Solar Electricity*, 2nd ed. Chichester, 2000.
- [3] L. A. Mejía Calderón, J. Mosquera Ramírez, and J. S. Betancourth Rojas, “Síntesis dimensional de un mecanismo esférico para la orientación de paneles solares,” *Entre Cienc. e Ing.*, vol. 11, no. 22, p. 43, 2017, doi: 10.31908/19098367.3548.
- [4] M. D. M. y E. UPME, IDEAM, Ambiente, “Atlas de Radiación Solar de Colombia Atlas de Radiación Solar de Colombia,” p. 175, 2005.
- [5] H. O. Benavides, “Información técnica sobre Gases de Efecto Invernadero y el cambio climático,” *Ideam*, pp. 1–102, 2007, doi: IDEAM–METEO/008-2007.
- [6] Unidad de Planeación Minero Energética [UPME], “Integración de las energías renovables no convencionales en Colombia,” 2015.
- [7] CELSIA, “Energía solar en Colombia, panorama en cifras,” *Energía solar en Colombia: así es el panorama en cifras*. [Online]. Available: <https://blog.celsia.com/energia-solar-en-colombia-panorama-en-cifras/>. [Accessed: 29-Nov-2019].
- [8] XM S.A., “Informes Anuales XM.” [Online]. Available: <http://informesanuales.xm.com.co/2015/SitePages/operacion/2-6-Capacidad-efectiva-neta.aspx>. [Accessed: 02-Dec-2019].
- [9] J. Gómez Ramírez, J. D. Murcia Murcia, and I. Cabeza Rojas, “La energía solar fotovoltaica en Colombia: potenciales, antecedentes y perspectivas,” *Univ. St. Tomás*, p. 19, 2017.
- [10] UPME, *Integración de las energías renovables no convencionales en Colombia*. 2015.

- [11] G. Capizzi, C. Napoli, and F. Bonanno, “Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 23, no. 11, pp. 1805–1815, 2012, doi: 10.1109/TNNLS.2012.2216546.
- [12] Instituto de Hidrología Meteorología y Estudios Ambientales de Colombia [IDEAM], “Solicitud de Información,” 2016.
- [13] A. Khosravi, R. O. Nunes, M. E. H. Assad, and L. Machado, “Comparison of artificial intelligence methods in estimation of daily global solar radiation,” *J. Clean. Prod.*, vol. 194, pp. 342–358, 2018, doi: 10.1016/j.jclepro.2018.05.147.
- [14] W. Contreras, M. G. Galban, and S. B. Sepúlveda, “Análisis estadístico de la radiación solar en la ciudad de Cúcuta,” *Entre Cienc. e Ing.*, vol. 12, no. 23, p. 16, 2018, doi: 10.31908/19098367.3698.
- [15] L. Olatomiwa, S. Mekhilef, S. Shamshirband, and D. Petković, “Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria,” *Renew. Sustain. Energy Rev.*, vol. 51, pp. 1784–1791, 2015, doi: 10.1016/j.rser.2015.05.068.
- [16] A. Mellit, A. H. Arab, N. Khorissi, and H. Salhi, “An ANFIS-based forecasting for solar radiation data from sunshine duration and ambient temperature,” *2007 IEEE Power Eng. Soc. Gen. Meet. PES*, 2007, doi: 10.1109/PES.2007.386131.
- [17] C. Montoya Rasero, “La energía solar fotovoltaica en el CTE,” *Electra*, no. 137, pp. 94–100, 2006.
- [18] A. de empresas de energías Renovables, “¿Qué es la energía fotovoltaica?” [Online]. Available: <https://www.appa.es/appa-fotovoltaica/que-es-la-energia-fotovoltaica/>. [Accessed: 03-Dec-2019].

- [19] Agencia estatal de meteorología de España, “La radiación solar.”
- [20] U. B. De Castro *et al.*, “Calor y temperatura,” p. 1, 2019.
- [21] O. Perpiñan Lamigueiro, “E S Fotovoltaica,” p. 194, 2012.
- [22] J. Barañaño, “Central de información y discusión de energía en Chile,” 2010. [Online]. Available: <http://www.centralenergia.cl/2010/10/12/potencial-solar-de-chile/>. [Accessed: 16-Jan-2020].
- [23] K. and Zonen, “CMP6-Piranometro.” [Online]. Available: <http://kippzonen.es/Product/214/CMP6-Piranometro#.XjH7i2gzBIU>.
- [24] W. Navidi, *Estadística para ingenieros y científicos*, vol. 66. 2012.
- [25] M. S. Hossain, N. A. Rahim, M. M. Aman, and J. Selvaraj, “Application of ANOVA method to study solar energy for hydrogen production,” *Int. J. Hydrogen Energy*, vol. 44, no. 29, pp. 14571–14579, 2019, doi: 10.1016/j.ijhydene.2019.04.028.
- [26] A. H. Assi, M. H. Al Shamisi, and H. A. N. Hejase, “MATLAB tool for predicting the global solar radiation in UAE,” in *International Conference on Renewable Energies for Developing Countries, REDEC 2012*, 2012, pp. 1–8, doi: 10.1109/REDEC.2012.6416698.
- [27] K. Bakker, K. Whan, W. Knap, and M. Schmeits, “Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation,” *Sol. Energy*, vol. 191, no. September, pp. 138–150, 2019, doi: 10.1016/j.solener.2019.08.044.
- [28] P. Ponce Cruz, *Inteligencia artificial con aplicaciones a la ingeniería*. Ciudad de México, 2010.
- [29] A. A. Serna, E. M. Acevedo, and E. M. Serna, “Principios de la Inteligencia Artificial en las Ciencias Computacionales,” *Actas Ing.*, vol. 3, pp. 354–361, 2017.
- [30] V. H. Quej, J. Almorox, J. A. Arnaldo, and L. Saito, “ANFIS, SVM and ANN soft-

- computing techniques to estimate daily global solar radiation in a warm sub-humid environment,” *J. Atmos. Solar-Terrestrial Phys.*, vol. 155, no. September 2016, pp. 62–70, 2017, doi: 10.1016/j.jastp.2017.02.002.
- [31] J. R. Jang, “ANFIS : Adaptive-Network-Based Fuzzy Inference System,” doi: 10.1109/21.256541.
- [32] A. Abraham, “Adaptation of Fuzzy Inference System Using Neural Learning,” vol. 83, pp. 53–83, 2005, doi: 10.1007/11339366_3.
- [33] A. M. Abdulshahed, A. P. Longstaff, and S. Fletcher, “The application of ANFIS prediction models for thermal error compensation on CNC machine tools,” *Appl. Soft Comput. J.*, vol. 27, pp. 158–168, 2015, doi: 10.1016/j.asoc.2014.11.012.
- [34] IDEAM, “Instituto de Hidrología, Meteorología y Estudios Ambientales.” [Online]. Available: <http://www.ideam.gov.co/solicitud-de-informacion>. [Accessed: 22-Jan-2020].
- [35] NASA, “NASA Prediction of Worldwide Energy Resources.” [Online]. Available: <https://power.larc.nasa.gov/>. [Accessed: 22-Jan-2020].
- [36] PVGIS, “Photovoltaic Geographical Information System.” [Online]. Available: <https://ec.europa.eu/jrc/en/pvgis>. [Accessed: 22-Jan-2020].
- [37] “loggernet,” *Download LoggerNet 4.5*. [Online]. Available: <https://www.campbellsci.es/loggernet>.
- [38] H. B. Yıldırım, Ö. Çelik, A. Teke, and B. Barutçu, “Estimating daily Global solar radiation with graphical user interface in Eastern Mediterranean region of Turkey,” *Renew. Sustain. Energy Rev.*, vol. 82, no. xxxx, pp. 1528–1537, 2018, doi: 10.1016/j.rser.2017.06.030.
- [39] U. K. Das *et al.*, “Forecasting of photovoltaic power generation and model optimization:

- A review,” *Renew. Sustain. Energy Rev.*, vol. 81, no. April 2017, pp. 912–928, 2018, doi: 10.1016/j.rser.2017.08.017.
- [40] N. Premalatha and A. Valan Arasu, “Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms,” *J. Appl. Res. Technol.*, vol. 14, no. 3, pp. 206–214, 2016, doi: 10.1016/j.jart.2016.05.001.
- [41] A. Qazi, H. Fayaz, A. Wadi, R. G. Raj, N. A. Rahim, and W. A. Khan, “The artificial neural network for solar radiation prediction and designing solar systems: A systematic literature review,” *J. Clean. Prod.*, vol. 104, pp. 1–12, 2015, doi: 10.1016/j.jclepro.2015.04.041.
- [42] K. Chiteka and C. C. Enweremadu, “Prediction of global horizontal solar irradiance in Zimbabwe using artificial neural networks,” *J. Clean. Prod.*, vol. 135, pp. 701–711, 2016, doi: 10.1016/j.jclepro.2016.06.128.
- [43] G. Hurtado Moreno and O. C. González, “Evaluación De La Afectación Territorial De Los Fenómenos El Niño/La Niña Y Análisis De La Confiabilidad De La Predicción Climática Basada En La Presencia De Un Evento,” *Ideam*, p. 38, 2011.
- [44] Ö. Çelik, A. Teke, and H. B. Yildirim, “The optimized artificial neural network model with Levenberg-Marquardt algorithm for global solar radiation estimation in Eastern Mediterranean Region of Turkey,” *J. Clean. Prod.*, vol. 116, pp. 1–12, 2016, doi: 10.1016/j.jclepro.2015.12.082.
- [45] D. Jain and M. Lalwani, “Prediction of irradiation: A comparative study of ANFIS,” *Proc. IEEE Int. Conf. Circuit, Power Comput. Technol. ICCPCT 2017*, 2017, doi: 10.1109/ICCPCT.2017.8074227.
- [46] H. K. Yadav, Y. Pal, and M. M. Tripathi, “Short-Term PV Power Forecasting Using

- Adaptive Neuro-Fuzzy Inference System,” *8th IEEE Power India Int. Conf. PIICON 2018*, pp. 1–6, 2018, doi: 10.1109/POWERI.2018.8704445.
- [47] K. Mohammadi, S. Shamshirband, A. Kamsin, P. C. Lai, and Z. Mansor, “Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure,” *Renew. Sustain. Energy Rev.*, vol. 63, pp. 423–434, 2016, doi: 10.1016/j.rser.2016.05.065.
- [48] A. Mellit, S. A. Kalogirou, S. Shaari, H. Salhi, and A. Hadj Arab, “Methodology for predicting sequences of mean monthly clearness index and daily solar radiation data in remote areas: Application for sizing a stand-alone PV system,” *Renew. Energy*, vol. 33, no. 7, pp. 1570–1590, 2008, doi: 10.1016/j.renene.2007.08.006.
- [49] T. R. Sumithira and A. Nirmal Kumar, “Prediction of monthly global solar radiation using adaptive neuro fuzzy inference system (ANFIS) technique over the state of Tamilnadu (India): A comparative study,” *Appl. Sol. Energy (English Transl. Geliotekhnika)*, vol. 48, no. 2, pp. 140–145, 2012, doi: 10.3103/S0003701X1202020X.
- [50] K. Mohammadi, S. Shamshirband, D. Petković, and H. Khorasanizadeh, “Determining the most important variables for diffuse solar radiation prediction using adaptive neuro-fuzzy methodology; Case study: City of Kerman, Iran,” *Renew. Sustain. Energy Rev.*, vol. 53, pp. 1570–1579, 2016, doi: 10.1016/j.rser.2015.09.028.