

UNIVERSIDAD
NACIONAL
DE COLOMBIA

ANÁLISIS DEL RECURSO SOLAR MEDIANTE MODELOS ESTADÍSTICOS DE PREDICCIÓN DE CORTO PLAZO EN LA SABANA DE BOGOTÁ

Omar Andrés Villarreal Mesa

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería Mecánica
Bogotá, Colombia

2020

Análisis del recurso solar mediante modelos estadísticos de predicción de corto plazo en la sabana de Bogotá

Omar Andrés Villarreal Mesa

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título
de:

Magister en Ingeniería Mecánica

Director (a):

Ph.D. Javier Rosero García

Línea de Investigación:

Energías renovables

Grupo de Investigación:

Electrical Machine and Drives EM&D

Universidad Nacional de Colombia

Facultad de ingeniería, Departamento de Ingeniería Mecánica

Bogotá, Colombia

2020

Agradecimientos

Especial gratitud con la Red de Monitoreo de Calidad del Aire de Bogotá de la Secretaría de Ambiente Distrital cuyo personal suministró los datos meteorológicos que fueron el insumo primordial a partir del cual se pudo desarrollar el presente trabajo. Por su presta disposición, su diligente colaboración y su profesionalismo.

Resumen

En las últimas décadas debido al cambio climático, se han incrementado los esfuerzos para la reducción del uso de combustibles fósiles y las emisiones de gases de efecto invernadero. Las fuentes renovables de energía han incrementado su participación en la matriz de generación de energía eléctrica, impulsadas por una reducción sensible de los costos de instalación y operación gracias a los nuevos desarrollos tecnológicos. No obstante, estas fuentes presentan grandes retos para su integración en las redes eléctricas, particularmente por ser consideradas como fuentes no gestionables por su naturaleza variable. En este trabajo se implementa una metodología de predicción para generar pronósticos a corto plazo (predicción para el día siguiente) de la irradiación solar promedio horaria en superficie. El método de predicción se construye y valida con información de estaciones meteorológicas de la Red de Monitoreo de Calidad del Aire de la Secretaría de Ambiente de Bogotá. La precisión del modelo fue evaluada a través del error medio cuadrático (RMSE) y el error medio absoluto (MAE).

Palabras clave: Irradiación solar, predicción de la oferta, series de tiempo, redes neuronales, energía solar fotovoltaica

Abstract

In recent years due to the emergence of climate change problem efforts have been increased to reduce fossil fuel consumption and greenhouse gasses emissions. Renewable energy sources have increased their participation in the electric power generation matrix, driven by a significant reduction in installation and operating costs through to new technological developments. However, these sources have great challenges for its integration into electricity grids, particularly since they are considered as non-manageable sources because of their variable nature. In this work a prediction methodology is implemented to generate short-term forecasts (day ahead predictions) of hourly average solar irradiation at surface. The forecasting method is built and validated with information from meteorological station of the Red de Monitoreo de Calidad del Aire de Bogotá. The precision of the model was evaluated through the root mean square error (RMSE) and mean absolute error (MAE).

Keywords: solar irradiation, forecasting methods, time series, neural networks, solar photovoltaic energy

Contenido

	Pág.
1. Introducción	1
2. Justificación	3
3. Objetivos.....	5
3.1 Objetivo general	5
3.2 Ejemplos de presentación y citación de figuras	5
4. Marco teórico.....	7
4.1 Sistema eléctrico colombiano.....	7
4.2 Energía solar.....	8
4.3 Integración de las energías renovables en Colombia	10
4.4 Radiación solar	12
4.5 Modelos de predicción de la irradiación solar	13
4.5.1 Modelos físicos	14
4.5.1.1 Métodos físicos basados en imágenes satelitales	14
4.5.1.2 Métodos físicos basados en imágenes tierra-cielo.....	15
4.5.1.3 Modelos numéricos de predicción meteorológica	16
4.5.2 Modelos estadísticos	18
4.5.2.1 Modelos autorregresivos.....	19
4.5.2.2 Modelos de promedios móviles.....	20
4.5.2.3 Modelos autorregresivos de promedios móviles	20
4.5.2.4 Modelos ARIMA.....	21
4.5.2.5 Modelos SARIMA	22
4.5.2.6 Modelo de redes neuronales ANN	23
5. Estado del arte.....	25
5.1 Modelos de predicción a corto plazo de la irradiación solar	25
5.2 Métodos de validación de predicciones	29
5.2.1 Métricas de precisión	29
5.2.2 Evaluación del desempeño de predicciones	31
6. Metodología	33
6.1 Pre-procesamiento de datos	34
6.2 Selección del modelo	34
6.3 Ajuste y generación de predicciones	36
6.4 Validación del modelo	37

7. Análisis de resultados	39
7.1 Pre-procesamiento de datos	39
7.2 Selección del modelo	43
7.3 Análisis de la irradiación	45
7.3.1 Comportamiento de la irradiación solar horaria promedio.....	45
7.3.2 Análisis de la variabilidad de la irradiación solar	46
7.3.3 Variación entre emplazamientos de estudio	49
7.4 Modelo de series de tiempo	51
7.4.1 Identificación del modelo	52
7.4.1.1 Prueba de Dick and Fulley	52
7.4.1.2 Análisis de la función de autocorrelación y de autocorrelación parcial....	53
7.4.1.3 Análisis de la función de autocorrelación y de autocorrelación parcial....	56
7.4.2 Obtención de parámetros	57
7.4.3 Diagnóstico del modelo	57
7.4.3.1 Comportamiento de los residuales	58
7.4.3.2 Correlación de los residuales	58
7.4.3.3 Gráfica cuantil-cuantil.....	59
7.4.3.4 Prueba de Ljung-Box.....	60
7.5 Modelo de redes neuronales.....	60
7.5.1 Selección de los parámetros de rezagos del modelo.....	61
7.5.2 Selección del número de nodos del modelo	62
7.6 Generación de predicciones	64
7.7 Validación de predicciones	67
8. Conclusiones	71
8.1 Conclusiones	71
8.2 Recomendaciones	72
9. Bibliografía	83

Lista de figuras

	Pág.
Figura 1. Esquema del Sistema Eléctrico Nacional [10].....	8
Figura 2. Evolución de costos de las energías renovables [16]	10
Figura 3. Capacidad de generación eléctrica del SIN a diciembre de 2014 [7]	11
Figura 4. Modelo WRF regional de temperatura para Estados Unidos	18
Figura 5. Modelo biológico y matemático de redes neuronales [47].....	24
Figura 6. Diagrama de flujo metodología	33
Figura 7. Selección del modelo de predicción	44
Figura 8. Comportamiento de la irradiación solar promedio.....	46
Figura 9. Diagrama de cajas irradiación CDAR 2015.....	47
Figura 10. Irradiación global promedio horaria mes de septiembre de 2015.....	49
Figura 11. Irradiación global promedio horaria mes de abril de 2016.....	50
Figura 12. Prueba de Dick and Fulley (CADR) [65]	53
Figura 13. Función de autocorrelación – CDAR.....	54
Figura 14. Función de autocorrelación parcial – CDAR	54
Figura 15. Función de autocorrelación diferenciada cada 13 rezagos – CDAR	55
Figura 16. Función de autocorrelación parcial diferenciada cada 13 rezagos – CDAR ...	55
Figura 17. Residuales del modelo SARIMA (1,0,1,1,1,2) ₁₃ – CDAR	58
Figura 18. Autocorrelación de los residuales – CDAR	59
Figura 19. Gráfica cuantil-cuantil de los residuales – CDAR.....	59
Figura 20. Descomposición de la serie de radiación global promedio horaria.....	61
Figura 21. Visualización de la ARNN.....	64
Figura 22. Pronóstico 8-15 de octubre de 2016 - CDAR.....	65
Figura 23. Pronóstico 18-24 de mayo de 2016 - CDAR	66
Figura 24. Pronóstico 9-15 de octubre de 2016 – San Cristóbal.....	79
Figura 25. Pronóstico 18-24 de mayo de 2016 – San Cristóbal.....	79
Figura 26. Pronóstico 9-15 de octubre de 2016 – Guaymaral.....	80
Figura 27. Pronóstico 9-15 de octubre de 2016 – Guaymaral.....	80
Figura 28. Pronóstico 22-28 de diciembre de 2015 – Tunal.....	81
Figura 29. Pronóstico 22-28 de diciembre de 2015 – Tunal.....	81
Figura 30. Pronóstico 22-28 de diciembre de 2015 – Kennedy.....	82

Lista de tablas

	Pág.
Tabla 1. Matriz de selección de modelo	35
Tabla 2. Matriz de calificación	36
Tabla 3. Ubicación estaciones RMCAB [62].....	40
Tabla 4. Características estaciones RMCAB [62].....	40
Tabla 5. Disponibilidad de la estación meteorológica	41
Tabla 6. Disponibilidad de estaciones RMCAB	43
Tabla 7. Matriz de evaluación de modelos de predicción	44
Tabla 8. AIC - CDAR.....	56
Tabla 9. Parámetros del modelo SARIMA (1,0,1,1,1,2) ₁₃ - CDAR	57
Tabla 10. Ljung-Box test – CDAR	60
Tabla 11. Parámetros de rezago autorregresivo para el modelo ARRN	62
Tabla 12. Número de nodos en la capa escondida para el modelo ARNN	63
Tabla 13. Periodos de pronóstico.....	66
Tabla 14. Desempeño de los modelos de pronóstico temporada seca.....	67
Tabla 15. Desempeño de los modelos de pronóstico temporada de lluvias	68
Tabla 16. Test de Dick and Fulley (San Cristóbal)	75
Tabla 17. Test de Dick and Fulley (Guaymaral)	75
Tabla 18. Test de Dick and Fulley (Kennedy).....	76
Tabla 19. Test de Dick and Fulley (Tunal).....	76

1.Introducción

En la actualidad hay una preocupación creciente entre la comunidad académica y las autoridades gubernamentales con respecto al calentamiento global progresivo que se ha registrado durante las últimas décadas. Las emisiones de gases de efecto invernadero (CO₂, CH₄, NO₂, etc.) de origen antrópico se han incrementado exponencialmente desde la era industrial, impulsados principalmente por el crecimiento económico y poblacional [1]. Esto ha elevado las concentraciones atmosféricas de estos gases a niveles sin precedentes en la historia reciente del planeta. Es altamente probable que el calentamiento global observado desde mediados del siglo XX se deba a los efectos del aumento de la concentración de estos gases, los cuales absorben la radiación infrarroja de forma más intensa que los demás constituyentes de la atmósfera e irradian esta energía de vuelta a la superficie terrestre [2].

La oferta energética mundial está dominada por la explotación de combustibles fósiles (carbón, gas natural y petróleo) [3]. Sin embargo, está demostrado que el ingente uso de estos combustibles es una de las grandes fuentes de emisión de gases de efecto invernadero. Este hecho, sumado a la naturaleza agotable de las reservas de combustibles fósiles ha creado una presión sobre el mercado energético, el cual está migrando hacia tecnologías de aprovechamiento de fuentes limpias, principalmente la energía solar fotovoltaica y la energía eólica, cuyos costos de instalación ya están en el rango de la generación a partir de combustibles fósiles [4].

Entre los efectos del cambio climático se cuentan: incremento de los niveles del mar, aumento en la temperatura promedio de la superficie terrestre, disminución de la extensión de las áreas de nieves y hielos, aumento de los eventos meteorológicos extremos y la acidificación oceánica [5]. Aunque Colombia no contribuye significativamente a la emisión de gases de efecto invernadero, el país si es altamente vulnerable a los efectos del cambio

climático, debido a su actual dependencia del recurso hídrico para la generación energía eléctrica [6].

En Colombia, la matriz energética está constituida principalmente por las centrales hidroeléctricas (75%), otra porción significativa está representada por las centrales térmicas de gas natural y carbón (19%), y las energías renovables tienen tan solo una participación marginal en el mercado (<1%) [7]. A pesar de que la generación hidroeléctrica es considerada tradicionalmente como energía limpia por su baja huella de carbono, ésta también conlleva graves impactos ambientales y sociales como alteración de los regímenes hídricos, transporte de sedimentos, desplazamiento de comunidades, desmejora de la seguridad alimentaria de la población afectada aguas abajo y procesos de erosión asociados con la construcción de la represa entre otros [8].

La integración de las energías renovables en el mercado energético colombiano presenta grandes retos. Particularmente la energía solar y la energía eólica dependen de fuentes que no son controlables ni almacenables, lo que las cataloga como una fuente de energía no gestionable. Su comportamiento es altamente volátil al depender de variables meteorológicas, lo que dificulta la predicción del recurso y afecta la confiabilidad en el suministro de energía. Finalmente, estos recursos están restringidos a ciertas ubicaciones geográficas y pueden requerir de infraestructura para la interconexión [9].

En este trabajo se presenta un estudio de predicción a corto plazo de la irradiación solar. Se realizó una revisión bibliográfica de modelos de predicción de recurso solar con énfasis en los modelos de predicción para el día siguiente. Se realizaron pronósticos en un horizonte de siete días usando un modelo de series de tiempo estacional SARIMA y un modelo de redes neuronales autorregresivo. Se utilizaron datos meteorológicos obtenidos de la Red de Monitoreo de Calidad del Aire de Bogotá. Los modelos ajustados fueron probados para cinco emplazamientos distintos en la sabana de Bogotá.

2. Justificación

El mercado energético a nivel mundial se encuentra inmerso en una “transición energética”, marcada por cambios en los patrones de uso de la energía que afectan los recursos, los usuarios, la infraestructura y los servicios. Una de las principales características de esta transición es la migración hacia fuentes de energías renovables principalmente solar y eólica, disminuyendo la generación a partir de combustibles fósiles.

La principal ventaja de las energías renovables es que son recursos en la práctica inagotables y que pueden ser aprovechados de manera sostenible, los procesos de generación entregan energía limpia y libre de contaminantes, por lo que no contribuyen al efecto invernadero y el calentamiento global. No obstante, las energías renovables tienen grandes dificultades para asegurar un suministro confiable de energía debido a que las fuentes primarias dependen de fenómenos climáticos, meteorológicos e hidrológicos los cuales son por naturaleza variables, incrementan la incertidumbre de generar la energía esperada lo que disminuye la confiabilidad de la red.

En Colombia, el mercado energético mayorista es operado por la empresa XM, quien tiene a cargo las funciones de coordinador del Centro Nacional de Despacho – CND y operador del mercado intra-diario. El precio de la electricidad en este mercado se establece de común acuerdo con base en la oferta y la demanda y sin intervención del estado. Para las plantas generadoras es muy importante conocer con antelación su capacidad de generación, ya que en caso de no poder cumplir con la oferta planteada el generador será sancionado asignándole una disponibilidad de cero para las próximas 24 horas. Esto último es particularmente importante para las plantas de generación a partir de fuentes no renovables, las cuales al depender de recursos por naturaleza variables necesitan desarrollar modelos de predicción de su oferta con la mayor precisión posible.

La predicción del recurso solar en sus diferentes escalas de tiempo es crucial para el diseño y operación de sistemas de generación fotovoltaica, incluyendo los sistemas de almacenamiento de energía, los sistemas de control de carga residenciales e industriales, y para el desarrollo y planificación del mercado energético y su marco regulatorio. Esto permite a los operadores de la red optimizar la distribución de energía, identificar necesidades de balance con otras fuentes disponibles, planear las actividades de mantenimiento en los sitios de producción y tomar las medidas necesarias en caso de eventos extremos.

3. Objetivos

3.1 Objetivo general

- Implementar una metodología de predicción que permita generar pronósticos de la irradiación solar promedio horaria y aplicarla en la sabana de Bogotá.

3.2 Ejemplos de presentación y citación de figuras

- Evaluar modelos de predicción a corto plazo aplicados en pronósticos de la irradiación solar para el día siguiente.
- Establecer un modelo de predicción a corto plazo del recurso solar para ser implementado en la sabana de Bogotá.
- Validar la idoneidad del modelo para representar el comportamiento de la irradiación solar con los datos disponibles de estaciones meteorológicas.
- Evaluar el desempeño del modelo para generar pronósticos de la irradiación solar en diferentes emplazamientos en la sabana de Bogotá.

4. Marco teórico

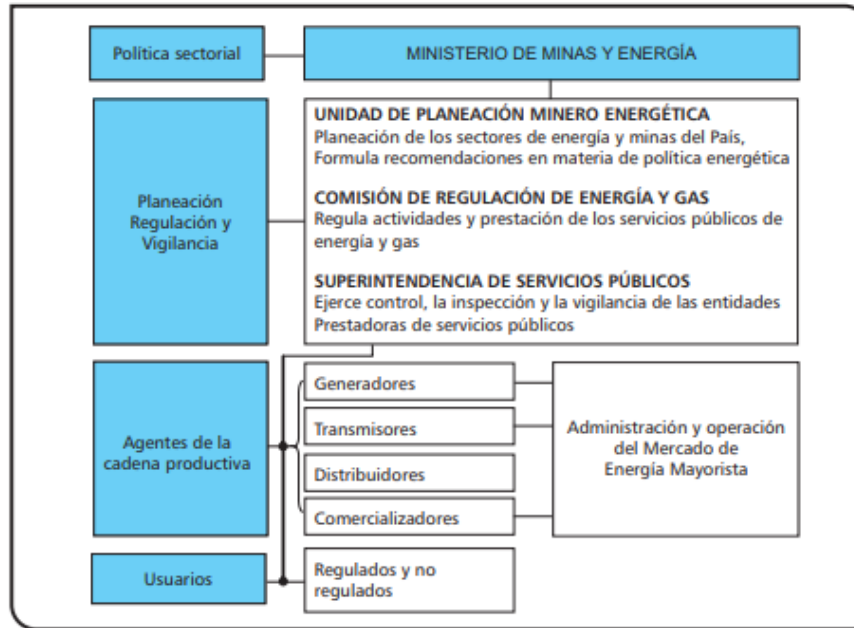
4.1 Sistema eléctrico colombiano

En Colombia existe una clara separación de roles entre el estado y los inversionistas en el mercado energético. El estado tiene como función fijar las políticas públicas energéticas, regular el mercado, ejercer la vigilancia y control de los actores y llevar a cabo la planeación estratégica para la expansión. En tanto los agentes privados son los encargados de la administración e inversión en el mercado. En la Figura 1 se muestra un esquema del sistema eléctrico nacional [10].

En la Figura 1 se muestra un esquema del Sistema Eléctrico Nacional. El marco regulatorio colombiano diferencia claramente las actividades que desarrollan los agentes para la prestación del servicio de electricidad en generación, transmisión, distribución y comercialización de energía eléctrica [11]. En la generación y comercialización tienen un tipo de mercado de libre competencia. En tanto en la transmisión y distribución se tiene una regulación que reconoce el carácter de monopolio natural, buscando en todo caso condiciones mínimas de competencia en donde sea posible [12].

Las empresas comercializadoras adquieren la energía eléctrica en una bolsa de energía o mercado energético mayorista que busca garantizar un modelo de libre oferta y demanda sin la intervención del estado. Los grandes generadores presentan sus ofertas a la bolsa con una anticipación de 24 horas para la operación del día siguiente, los agentes comercializadores están en la libertad de elegir cual oferta tomar observando sus contratos de distribución suscritos. Como contraparte los grandes consumidores y comercializadores están en libertad de celebrar contratos bilaterales con las empresas generadoras, los cuales deben ser registrados en un sistema de administración de intercambios comerciales [12].

Figura 1. Esquema del Sistema Eléctrico Nacional [10]



El estricto cumplimiento de los compromisos por parte de las generadoras garantiza el equilibrio entre la oferta y la demanda y la estabilidad del mercado eléctrico, la empresa XM tiene a cargo el control de la operación integrada de los recursos de generación, interconexión y transmisión; cuando las empresas operadoras no cumplen con su oferta energética durante el día, como sanción se las declara no disponibles para el día siguiente [13]. Esto hace que sea necesario para los operadores conocer con anticipación su potencial de generación, y es más sensible para la generación a partir de energías renovables que tienen naturaleza variable y una muy alta incertidumbre para su predicción.

4.2 Energía solar

La energía solar es la energía que se obtiene de la radiación electromagnética procedente del sol, la energía solar es la fuente principal a partir de la cual se desarrollan directa o indirectamente todos los procesos físicos y biológicos en la tierra [14]. Se estima que la energía recibida del sol en la superficie terrestre es de alrededor de 1.08×10^{14} kW; si tan solo el 0,1% de esta energía se convirtiera con una eficiencia del 10%, sería suficiente para generar cuatro veces la capacidad instalada de generación mundial, que fue de 3,000 GW en 2013 [15].

Los principales inconvenientes para la explotación a gran escala de la energía solar, y en general de las energías renovables radican en que estos recursos presentan grandes variaciones de hora a hora, así como estacionalmente, lo que las clasifica como energías no gestionables. Además, debido a la disponibilidad restringida de dichos recursos en ciertas ubicaciones geográficas, la generación debe ser llevada a cabo en el sitio donde se encuentra el recurso, lo que puede representar una barrera grande en términos de su interconexión al mercado energético.

La energía solar fotovoltaica es la conversión directa de la radiación solar en electricidad. Esto se logra a través de una celda fotovoltaica, las celdas son dispositivos hechos de dos tipos de materiales semiconductores: uno de carga positiva (p) y uno de carga negativa (n). Cuando la celda es expuesta a la irradiación solar, los fotones de la luz “arrancan” un electrón debido al efecto fotovoltaico, la diferencia de potencial causada en la celda genera una corriente DC que puede ser aprovechada [16].

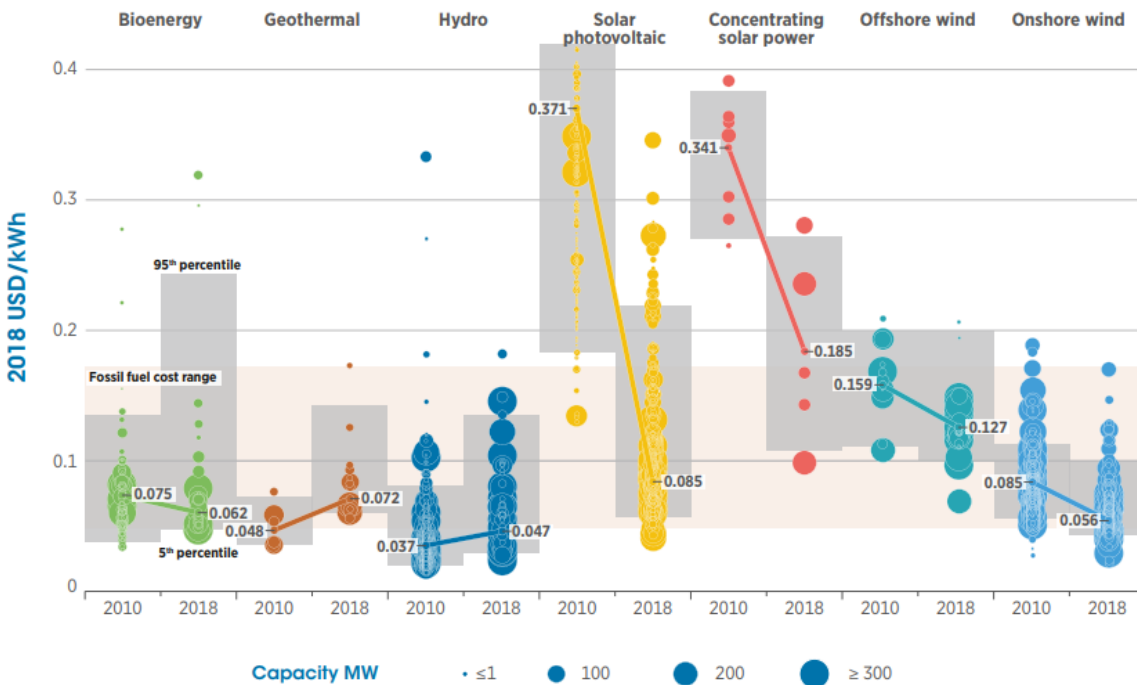
Las celdas individuales tienen un voltaje de salida limitado a aproximadamente 0,5 V y corrientes de salida limitadas a alrededor de 7 A, por lo que es necesario combinarlas en arreglos en serie y/o en paralelo para obtener mayores voltajes y corrientes. Un módulo fotovoltaico, más conocido como panel solar consta de un arreglo de 36 celdas conectadas en serie para producir un voltaje máximo de salida de 17 V, con una corriente de 7 A [17]. La eficiencia del módulo siempre será menor que la eficiencia de las celdas debido a las pérdidas de energía.

Por su sencillez, modularidad y confiabilidad los sistemas de generación fotovoltaica presentan un gran campo de aplicación, desde pequeños sistemas de generación como fuente de electricidad en el lugar de consumo hasta instalaciones de gran capacidad conectadas a la red de distribución [18].

El costo de las primeras celdas de silicio desarrolladas en 1958 era exorbitante para ser usado en aplicaciones industriales (\$1000 USD/W), la primera aplicación práctica de las celdas fue para las misiones espaciales, donde el precio no era un obstáculo mayor y el sol era la única fuente de energía disponible [18]. La crisis energética de los años 70's,

dirigió la atención al desarrollo de fuentes alternativas de energía e impulsó el fomento de la investigación en materiales semiconductores para paneles solares. Pero es durante los últimos 10 años que se ha experimentado una mejora en los costos que nos ha llevado al despegue final de la energía solar, pasando de un costo de instalación de 8.5 USD/W en 2009 a 2.99 USD/W en 2019 [16].

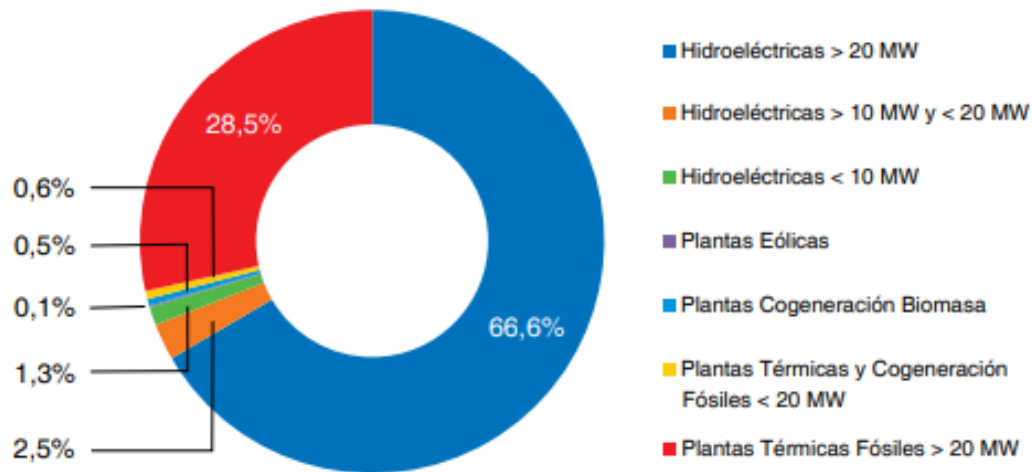
Figura 2. Evolución de costos de las energías renovables [16]



4.3 Integración de las energías renovables en Colombia

La matriz de energía eléctrica de Colombia puede considerarse como limpia o con una baja huella de carbono. Alrededor de un 70% de la capacidad instalada de generación proviene de plantas de generación hidroeléctricas, las plantas termoeléctricas que generan a partir de carbón y gas natural representan un 28,5%, en tanto las energías renovables tienen una participación marginal en la matriz (ver Figura 3.) [7]. La predominancia de la energía hidroeléctrica se explica por las ventajas geográficas con que cuenta el país dada su orografía predominantemente montañosa.

Figura 3. Capacidad de generación eléctrica del SIN a diciembre de 2014 [7]



Colombia tiene un importante potencial de recursos naturales para generación que aseguran la autosuficiencia del país por lo menos en el corto plazo, incluso durante los últimos lustros el país se ha convertido en un exportador de energía en la región [19]. Además, el hecho de que el sistema eléctrico sea relativamente bajo en emisiones de gases de efecto invernadero ha hecho que la matriz energética se haya mantenido sin variaciones significativas. No obstante, a causa de la alta dependencia del recurso hídrico, el país es vulnerable a fenómenos de sequía como el fenómeno del niño, los cuales se proyecta que se intensificarán como consecuencia del cambio climático [20].

El estado colombiano ha venido realizando esfuerzos para diversificar la matriz energética mediante la incorporación de energías renovables. En particular la Ley 1715 de 2015 entrega incentivos tributarios como deducciones del impuesto de renta, exención de aranceles, reducción del IVA y depreciación acelerada de los activos relacionados para aquellas empresas que inviertan en proyectos de generación a partir de fuentes de energía no convencionales como energía eólica, solar fotovoltaica, biomasa, geotérmica, mareomotriz y pequeñas centrales hidroeléctricas [21].

El principal reto para la integración de las energías renovables en el mercado eléctrico colombiano radica en su naturaleza variable al depender del comportamiento de variables meteorológicas y la dificultad de predicción del recurso primario [9]. Esto impacta en

particular a las empresas generadoras, ya que por su diseño el mercado energético colombiano requiere declarar con antelación de un día el potencial disponible de generación. Cuando esta oferta no se cumple el administrador del mercado puede sancionar a la empresa penalizando su despacho del día siguiente [22].

4.4 Radiación solar

Se denomina radiación solar al espectro electromagnético emitido por el sol, este tiene una longitud de onda característica que va de los 200 a 4000 nm, comprendiendo principalmente los espectros infrarrojo, visible y ultravioleta. Mientras la radiación solar se propaga libremente en el vacío, la irradiación solar es una medida de la energía que alcanza la superficie de la atmósfera; esta energía se ve atenuada de acuerdo con el cuadrado de la distancia [23].

La irradiación solar se expresa como una densidad de flujo energético y sus unidades son de vatios por metro cuadrado (W/m^2). La irradiación solar en la superficie de la atmósfera se toma típicamente como $1366 W/m^2$, pero la naturaleza elíptica de la órbita terrestre hace que este valor varíe entre $1321 W/m^2$ en el apogeo (alrededor del 4 de Julio) hasta $1415 W/m^2$ en el perigeo (alrededor del 3 de enero) [24].

La energía producida por un sistema de generación fotovoltaico depende directamente de la radiación incidente sobre el sistema. La radiación solar que alcanza la superficie terrestre es el resultado de múltiples y complejas interacciones que se dan en la atmósfera. Tres procesos meteorológicos modifican principalmente la radiación solar que pasa a través de la atmósfera con dirección a la superficie terrestre: dispersión, absorción y reflexión. Estos procesos se originan en la interacción de la radiación con los gases y partículas suspendidas que se encuentran en la atmósfera terrestre [25].

A la radiación que alcanza la superficie de la tierra sin ser modificada se le llama radiación solar directa. En cambio, a la radiación que alcanza la superficie después de ser dispersada se le llama radiación solar difusa. No toda la radiación solar directa y difusa está disponible en superficie, del mismo modo que en la atmósfera, la radiación solar recibida puede ser redirigida debido a la reflexión espacial, a este fenómeno se le conoce

como albedo. El albedo o reflectividad en la superficie terrestre depende del tipo de material y puede variar del 5-80%. La radiación solar que incide sobre un objeto es una combinación de radiación solar directa y difusa, a esta combinación se le llama radiación solar global [26].

Colombia por su posición geográfica en una zona tórrida cercana al ecuador es un país privilegiado, no obstante, el país también se encuentra en la región andina por lo que su clima es variado lo que afecta la disponibilidad de la radiación solar. En el país, el estudio más extenso de la radiación solar es el atlas de la radiación solar del IDEAM [27]. En el atlas se observa que la mejor región del país para la generación es la zona de la costa Caribe, debido al clima predominante seco de la zona. No obstante, otras zonas del país como los llanos orientales tienen un buen potencial en las épocas secas del año.

4.5 Modelos de predicción de la irradiación solar

Predecir la irradiación solar disponible es el primer paso para modelar el desempeño de un sistema de conversión de energía solar y asegurar una integración estable con la red eléctrica [28]. Debido a la alta complejidad de la dinámica atmosférica, y la interacción con la radiación en la atmósfera, no es posible predecir la irradiación solar disponible de forma analítica. Los modelos de predicción toman como base mediciones meteorológicas o imágenes satelitales.

Dependiendo de la aplicación y de la escala de tiempo, se selecciona la metodología de predicción adecuada. Para predicción de muy corto plazo los modelos que usan mediciones directas in situ son adecuados. Los pronósticos intra hora con una alta resolución espacial y temporal se basan en imágenes cielo-tierra tomadas a través de sistemas automatizados. Para una escala desde 30 minutos hasta 6 horas las predicciones de movimiento de nubes con base en imágenes satelitales muestran buen rendimiento. Los modelos estadísticos son adecuados para horizontes de predicción de una hora hasta varios días adelante. Finalmente, para la integración a la red se requiere de predicciones con uno o varios días de antelación, estos pronósticos se basan en modelos numéricos de predicción del clima (NWP) [29].

En general, los modelos de predicción se clasifican en dos categorías: modelos físicos y modelos estadísticos.

4.5.1 Modelos físicos

Los modelos físicos se dividen en: modelos numéricos de predicción meteorológica (NWP) que se alimentan de observaciones obtenidas a nivel del suelo o con ayuda de instrumentos flotantes; otro tipo de modelos de predicción físicos se basan en imágenes de las nubes tomadas a través de satélites, o imágenes tierra-cielo tomadas por sistemas de control automáticos al nivel del suelo [29].

La predicción a corto plazo (0-3 horas) implica el uso de mallas espaciales y temporales muy finas para detectar fluctuaciones de alta frecuencia como la sombra de las nubes y la influencia del contenido de aerosoles en la atenuación de la radiación difusa. Para estos pronósticos, el factor principal de variabilidad en la irradiación es la cobertura de la nubosidad.

En cambio, los pronósticos a más largo plazo (un horizonte de horas a días) requieren recolectar datos suficientes para una inicialización precisa del modelo y una representación realística de las nubes. Para la predicción se precisa de observaciones mucho más largas (varios años) para compilar estadísticas robustas que puedan resolver la media y la desviación teniendo en cuenta variaciones estacionales. Mientras que los modelos de predicción numérica han experimentado avances considerables en las últimas décadas, la caracterización de las nubes y su compleja influencia sobre el estado del clima sigue siendo uno de los eslabones más débiles en la predicción del recurso solar [24].

4.5.1.1 Métodos físicos basados en imágenes satelitales

Los sistemas de observación satelital son una herramienta integral para aplicar en la predicción del recurso solar, con una utilidad que abarca todo el espectro de las necesidades de los usuarios desde la evaluación del recurso para pronósticos unas horas adelante hasta escalas que permiten modelar la generación en un horizonte de varios días después.

Las predicciones con base en imágenes satelitales son una implementación híbrida de modelos físicos y aproximaciones estocásticas. La base de los modelos físicos para el pronóstico de la irradiación solar depende totalmente de la interacción entre los componentes atmosféricos como los gases y los aerosoles con la radiación solar. Estas interacciones se simulan a través de modelos de transferencia de la radiación, en tanto las técnicas estadísticas se definen sobre la base de regresiones entre la irradiación solar medida a nivel del suelo y medidas simultaneas tomadas por instrumentos satelitales. Los parámetros de las ecuaciones de regresión incluyen el ángulo del cenit, el índice de cobertura de nubes y la transmisividad de la atmósfera [30].

Las imágenes de satélite gracias a su alta resolución temporal y espacial también son una herramienta útil para detectar el movimiento de las nubes y predecir los cambios meteorológicos en el muy corto plazo (30 minutos a 6 horas) debido a la nubosidad. Las predicciones se basan en la aplicación de modelos de campos de vectores de movimiento de nubes (CMVs). El procedimiento para derivar los CMVs consiste en identificar dos imágenes sucesivas y calcular el vector de desplazamiento entre dos zonas a partir de la diferencia cuadrática media en cada pixel. Los vectores de campo obtenidos se utilizan para predecir la posición de las nubes en el siguiente instante [31].

4.5.1.2 Métodos físicos basados en imágenes tierra-cielo

La predicción del recurso solar en escalas temporales y espaciales muy pequeñas (0-100m y 0-30 min) es un trabajo que presenta grandes desafíos, y el cual no puede lograrse de forma precisa a través de imágenes satelitales, ya que éstas tienen problemas técnicos como fallas en la frecuencia de captura de imágenes, errores de navegación y efectos de paralaje, que pueden resultar en una incorrecta georreferenciación de las nubes [32].

Los modelos de predicción basados en imágenes cielo-tierra utilizan fotografías tomadas por sistemas automáticos, conocidos como Total Sky Imager (TSI). Básicamente, un sistema TSI es una cámara fotográfica colocada al nivel del suelo, la cual captura imágenes del cielo con una cobertura de 180° con una frecuencia constante. Los sistemas TSI se pueden utilizar tanto para la predicción de la radiación normal directa como de la radiación total [33].

Las proyecciones de las condiciones de radiación solar observadas se basan en el historial de mediciones inmediatamente anterior. Las imágenes tomadas se procesan para detectar la localización y estructura de las nubes. La posición e influencia de las nubes en el futuro se deducen de su movimiento, la imagen de la nube se propaga hacia adelante en el tiempo a través de algoritmos de movimiento de vectores de forma análoga a los modelos satelitales. Los sistemas TSI son útiles para la predicción de la radiación en un horizonte de hasta 15 minutos. En términos del error medio cuadrático normalizado (RMSE) el error de las predicciones varía entre 18% - 24% para horizontes de predicción desde 30 segundos hasta 15 minutos. Sin embargo, debido a su naturaleza física y su potencial de desarrollo, los métodos TSI han sido adoptados rápidamente por muchos grupos de investigación y plantas de generación durante los últimos años, no solo para la predicción de la irradiación solar sino para aplicaciones meteorológicas en general [34].

El horizonte de pronóstico de los sistemas TSI está limitado a alrededor de 30 minutos debido a su rango limitado de visión. Una solución para extender su alcance es ubicar un arreglo de puntos de medición para obtener más información. Sin embargo, el costo asociado a cada TSI (\$2,000 USD) y la naturaleza dinámica de los campos locales de nubes pueden limitar la correlación de las imágenes [29].

4.5.1.3 Modelos numéricos de predicción meteorológica

Los modelos numéricos de predicción meteorológica se basan puramente en la física atmosférica. Tratan de como las observaciones del clima se pueden usar para elaborar pronósticos de su estado futuro a través de modelos físico-matemáticos. Sorprendentemente, el primer modelo numérico meteorológico fue planteado en 1922 por L. F. Richardson, no obstante, no fue hasta el advenimiento de la computación en los años 1960s que se lograron tiempos de cálculo menores que los tiempos de predicción [35].

Manipular los ingentes volúmenes de datos de entrada y ejecutar los complejos cálculos necesarios para los modelos numéricos de predicción meteorológica modernos requiere de poderosas computadoras. La naturaleza compleja de las ecuaciones diferenciales que gobiernan la física atmosférica, y el comportamiento caótico de las variables

meteorológicas, hace que los pronósticos a partir de este método no sean confiables en un horizonte mayor a 14 días [36].

Los modelos NWP primero recogen datos de las condiciones iniciales de la atmósfera a través de satélites, observaciones en tierra y mediciones de variables meteorológicas a diferentes alturas con globos diseñados para este fin. Después, esta información tiene que pasar por un proceso llamado asimilación de datos, este es un proceso complejo que busca reducir los errores debido al ruido introducido porque los instrumentos miden simultáneamente diferentes cantidades en volúmenes de control distintos. El siguiente paso es el planteamiento de las ecuaciones que gobiernan la dinámica atmosférica: segunda ley de Newton para fluidos, ecuaciones termodinámicas y ecuaciones de transferencia por radiación se integran y solucionan para obtener la tasa de cambio de las propiedades de la atmósfera, con esta tasa de cambio se proyecta un estado futuro de la atmósfera y se repite el proceso hasta el tiempo de predicción. Finalmente, a los resultados se les aplica un post procesamiento estadístico de ensayo de prueba y error, para comparar las salidas con las observaciones, determinar una correlación estadística y corregir el error [37].

Los modelos NWP se clasifican en dos tipos de modelos: modelos globales y modelos regionales. En los modelos globales se lleva a cabo una simulación global o mundial del comportamiento de la atmósfera. Mientras que los modelos regionales se simulan para una escala más pequeña (continental o de país), los modelos regionales generalmente se valen de los modelos globales para establecer las condiciones de frontera, por lo que se puede decir que los modelos regionales propagan los errores de los modelos globales. Los modelos NWP más utilizados son Global Forecast System (GFS), Climate Forecast System, North American Mesoscale (NAM), European Centre for Medium-Range Weather Forecasts (ECMWF) y Weather Research Forecasting (WRF) [38].

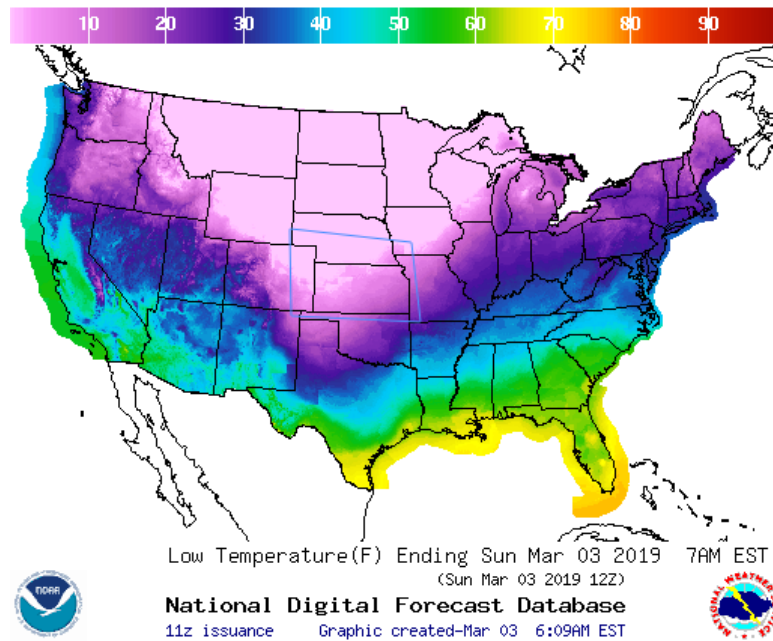


Figura 4. Modelo WRF regional de temperatura para Estados Unidos

La radiación solar es tan solo una de más de una decena de parámetros que manejan los modelos NWP, en la Figura 5 se muestra la salida de un modelo WRF norteamericano para la temperatura. Por tanto, los modelos NWP no desarrollan al detalle las leyes físicas de transferencia de calor para la radiación solar. En general, los modelos NWP tienen un buen desempeño para predicciones de la radiación solar con un horizonte desde un día hasta varios días. Son una herramienta eficaz para varias aplicaciones como para planear el despacho de plantas solares a medio plazo, pero su aplicación principal radica en la predicción del clima para meteorología.

4.5.2 Modelos estadísticos

Los modelos estadísticos para la predicción del recurso solar se basan en operaciones con series de tiempo. Una serie de tiempo se define como un conjunto de datos con un orden cronológico dado con respecto a un punto de referencia. Las aplicaciones de las series de tiempo abarcan un amplio abanico de áreas como econometría, análisis demográfico, procesamiento de señales, simulación de procesos biológicos, control y aseguramiento de calidad, análisis de inventarios y prospectiva de la oferta y demanda eléctrica [39].

Los modelos estadísticos presentan una serie de ventajas con respecto a los modelos físicos: no es necesario conocer las ecuaciones que gobiernan la física atmosférica, son muy simples, ya que al no tratar con los procesos atmosféricos requieren de mucho menos recursos computacionales que los modelos físicos, tienen un bajo tiempo de desarrollo y son fáciles de implementar. Por contra, al tratarse de modelos de tipo caja negra, donde el sistema se estudia desde el punto de vista de sus entradas y salidas, ignorando los procesos al interior, la interpretación de los resultados puede ser complicada, además, los modelos estadísticos requieren de bases de datos robustas ya que es necesario entrenar los modelos con todas las posibles situaciones que puedan presentarse [40].

Los modelos estadísticos se pueden dividir en dos categorías: métodos de regresión y métodos de inteligencia artificial. Algunos ejemplos de métodos de regresión son modelos autorregresivos (AR), modelos de promedio móvil (MA) y modelos autorregresivos de promedio móvil integrados (ARIMA). Las técnicas de regresión presentan limitaciones debido a la naturaleza caótica del comportamiento de variables como el movimiento de las nubes y las mediciones meteorológicas. Así, las técnicas de inteligencia artificial como las redes neuronales artificiales (ANN) y algoritmos genéticos también han tenido éxito reciente en su implementación para la predicción del recurso solar [41].

4.5.2.1 Modelos autorregresivos

Los modelos estadísticos autorregresivos (AR) son una de las técnicas más básicas aplicadas para modelar series de tiempo. Siendo X_t una variable aleatoria indexada en el tiempo, en un modelo autorregresivo los valores futuros de la variable X_t se calculan a partir de una combinación lineal de los valores previos de la misma serie de datos, sumado a un término de ruido blanco o error (Z_t). De la forma más general, un modelo de autorregresivo se puede expresar según la ecuación (1) [42]:

$$X_t = Z_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} \quad (1)$$

Donde ϕ_1, \dots, ϕ_p son los parámetros del modelo y dan el peso que tiene cada uno de los valores previos en el cálculo del valor actual de la serie de tiempo, el subíndice p indica el orden del modelo autorregresivo.

4.5.2.2 Modelos de promedios móviles

Otro tipo de modelos aplicado en el ajuste de series de tiempo son los modelos de promedios móviles (MA). En un modelo de promedios móviles se estima el valor de una variable aleatoria a partir de una combinación lineal de los términos de ruido asociados a las variables anteriores, un ejemplo de aplicación práctica puede ser la determinación del precio de venta de un producto en el mercado, el precio puede verse afectado por anuncios de descuentos o noticias producidas en los días anteriores. En el modelo de promedios móviles la variable X_t se calcula según la ecuación (2):

$$X_t = \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (2)$$

Donde nuevamente $\theta_1, \dots, \theta_q$ representan los parámetros del modelo y dan el peso que tiene cada término para el cálculo del valor actual en la serie de tiempo y el subíndice q indica el orden del modelo de promedios móviles.

4.5.2.3 Modelos autorregresivos de promedios móviles

Muchos de los procesos que se pueden describir a partir de series de tiempo son representadas mejor cuando se utiliza una combinación de términos autorregresivos y de promedios móviles para el ajuste. Estos modelos son conocidos como modelos ARMA, los modelos ARMA están formados por dos partes, una parte de modelo autorregresivo (AR) y otra parte de promedios móviles (MA). En general, un modelo ARMA se puede expresar según la ecuación (3):

$$X_t = Z_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (3)$$

Los parámetros ϕ_p y θ_q representan aquí también los coeficientes del modelo, mientras que los subíndices p y q denotan el orden de cada parte del modelo.

4.5.2.4 Modelos ARIMA

Los modelos de predicción presentados antes se aplican para series de tiempo estacionarias. Una serie de tiempo estacionaria se define como aquella en que la media y la varianza de los datos son constantes en el tiempo. En estos casos es posible transformar una serie de tiempo para que sea estacionaria removiendo las tendencias presentes, para esto se aplican operaciones de diferencias sucesivas a la serie de tiempo [43].

Para discutir los modelos ARIMA, primero debemos introducir el operador de retraso, la ecuación (4) define el operador de retraso:

$$BX_t = X_{t-1} \quad (4)$$

A partir del operador de retraso, el operador de diferencias se calcula con la ecuación (5):

$$\nabla^d X_t = (1 - B)^d X_t \quad (5)$$

Los modelos ARIMA incluyen un componente autorregresivo (AR), un componente de promedios móviles (MA) y un componente de integrado (I) en el que se transforma la serie en estacionaria después de d operaciones de diferencias sucesivas. La ecuación general del modelo ARIMA (p, d, q) se expresa según la ecuación (6):

$$Y_t = Z_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (6)$$

Donde:

$$Y_t = \nabla^d X_t \quad (7)$$

De forma más concisa, la ecuación (6) puede se puede reescribir en la ecuación (8):

$$\phi(B)Y_t = \theta(B) Z_t \quad (8)$$

Donde:

$$\phi(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (9)$$

$$\theta(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_2 B^2 \quad (10)$$

Donde d es orden del modelo de diferencias sucesivas, generalmente un orden de 1 o 2 es suficiente para eliminar las tendencias de las series de datos y no se recomienda diferenciar por encima de esos valores, p es el orden del modelo autorregresivo y q es el orden del modelo de promedios móviles. La decisión sobre el orden de los modelos autorregresivos y de promedios móviles se hace con base en procesos iterativos de minimización, en la selección del modelo también es importante considerar el principio de la parsimonia, que indica que siempre se debe preferir el modelo más sencillo posible [44].

4.5.2.5 Modelos SARIMA

Los datos de una serie de tiempo pueden exhibir una característica llamada estacionalidad, esto se observa cuando los valores de la serie X_t presentan un componente periódico además de las correlaciones con los valores más recientes; es decir que las observaciones se ven afectadas de forma similar cada s observaciones o rezagos, donde s se define como la estacionalidad de la serie. Entre los ejemplos de una serie estacional están las ganancias trimestrales obtenidas por una empresa, las ventas por mes de un almacén o fenómenos que se vean afectados por el clima como la contaminación atmosférica. La inclusión de un componente estacional en un modelo mejora notablemente el desempeño del modelo.

En estos casos para eliminar la estacionalidad se procede con la diferenciación de la serie cada s rezagos, cuando la serie presenta además de la estacionalidad una tendencia, la serie debe transformarse primero en una serie estacionaria a través de la diferenciación introducida antes. De forma general, un modelo SARIMA $(p, d, q, P, D, Q)_s$ se puede expresar según la ecuación (11):

$$\Phi_P(B^s)\phi_p(B)(1 - B^s)^D(1 - B)^d X_t = \Theta_Q(B^s)\theta_q(B)Z_t \quad (11)$$

Donde:

$$\phi(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (12)$$

$$\Theta_Q(B^S) = 1 + \theta_1 B^S + \theta_2 B^{2S} + \dots + \theta_Q B^{QS} \quad (13)$$

$$\theta(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (14)$$

$$\Phi_Q(B^S) = 1 - \Phi_1 B^S + \Phi_2 B^{2S} + \dots + \Phi_P B^{PS} \quad (15)$$

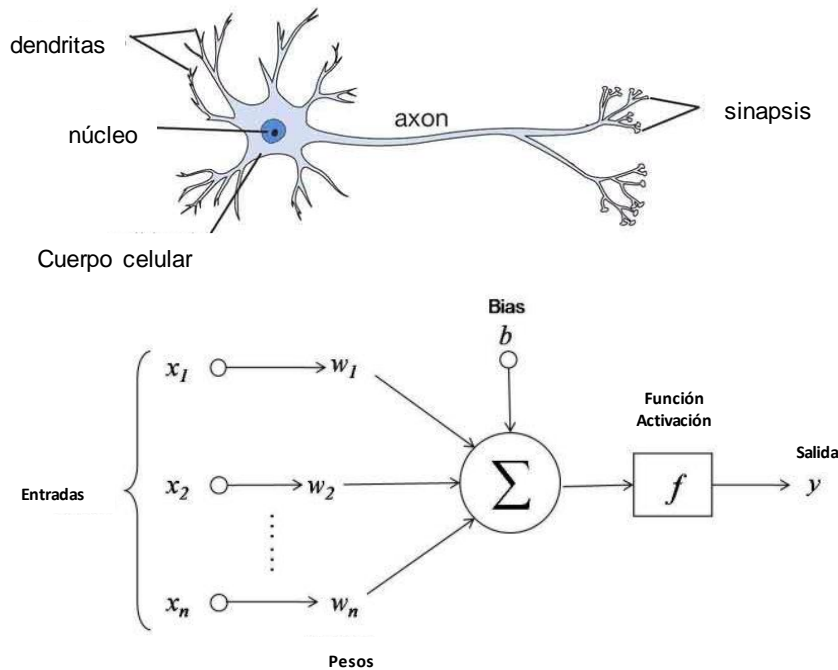
Los parámetros p , d y q son los mismos del modelo ARIMA, en tanto P representa los términos autorregresivos estacionales, D el orden de diferenciación estacional, al igual que en los modelos ARIMA no se recomienda que este término tenga un orden muy alto, finalmente Q es el orden de los términos de promedios móviles estacionales.

4.5.2.6 Modelo de redes neuronales ANN

Las técnicas de aprendizaje de máquina o inteligencia artificial (machine learning) permiten utilizar los ordenadores y otros dispositivos con capacidad computacional para que aprendan a extraer los patrones y relaciones que haya en un conjunto de datos por sí solos. Estos patrones se pueden usar posteriormente para predecir comportamientos y como apoyo en la toma de decisiones. Las técnicas de inteligencia artificial pueden ayudar a mejorar los modelos de pronóstico de la irradiación solar, ya que permiten manejar los componentes no lineales que presentan los datos [45].

Dentro de las técnicas de aprendizaje de maquina se encuentran los modelos de redes neuronales artificiales, estos modelos establecen una analogía con la estructura del sistema nervioso, el cual está compuesto de cientos de miles de neuronas que se comunican entre ellas por impulsos eléctricos a través de un proceso llamado sinapsis; las neuronas en general tienen una estructura compuesta por un cuerpo celular o soma del que se desprende un denso árbol de ramificaciones (árbol dendrítico) y del cual parte una fibra tubular denominada axón, que también se ramifica en su extremo final para conectarse con otras neuronas (ver Figura 7) [46]. Desde un punto de vista funcional, las neuronas constituyen procesadores de información simples, poseen un canal de entrada de información: las dendritas, un órgano de cómputo: el soma, y un canal de salida: el axón (ver Figura 5) [47].

Figura 5. Modelo biológico y matemático de redes neuronales [47]



La neurona artificial es la unidad fundamental de un modelo de redes neuronales artificiales, se trata de un elemento simple de cálculo que a partir de un vector de entrada ($x_i(t)$) procedente del exterior o de otras neuronas vecinas, emite una única salida que se transmite idéntica a otras neuronas posteriormente. El peso sináptico (w_i) define la intensidad de la interacción entre la neurona y la entrada (ver Figura 5), cuando se tiene una sinapsis fuerte la información que se transmite contribuye en buena medida a la definición del nuevo estado o alteración que se produce en la neurona receptora.

La red neuronal se somete a un proceso de aprendizaje, en éste las neuronas artificiales aprenden a través del ajuste de los pesos sinápticos que caracterizan a las conexiones. Los pesos son adaptados de acuerdo a la información extraída de los patrones de entrenamiento nuevos que se van presentando. Se distinguen tres tipos de aprendizaje: aprendizaje supervisado en el que se proporciona a la red las entradas y salidas correctas, aprendizaje no supervisado en el que no se suministran los patrones de salida, sino solo los de entrada y se deja a la red clasificarlos en función de características comunes que encuentre entre ellos y aprendizaje híbrido en el que solo se le dice a la red si la respuesta acierta o falla [47]

5. Estado del arte

5.1 Modelos de predicción a corto plazo de la irradiación solar

En la literatura se encuentran numerosos trabajos acerca de predicción de la radiación solar a corto plazo. Las técnicas de predicción se basan principalmente en modelos ARIMA, modelos de redes neuronales artificiales, técnicas híbridas que mezclan modelos ARIMA y redes neuronales, modelos basados en imágenes cielo-tierra y modelos numéricos de predicción meteorológica NWP.

Mohammed H. Alsharif, M. Younes y Jeong Kim [48] plantean un modelo integrado auto-regresivo de promedios móviles estacional (SARIMA) para predecir la radiación solar promedio diaria y promedio mensual en Seúl, Corea del Sur, el modelo fue alimentado con los datos de la administración meteorológica de Corea durante los años 1981-2017. Los investigadores encontraron que un modelo ARIMA (1,1,2) es el que mejor se ajusta al comportamiento de la radiación promedio diaria, mientras que un modelo SARIMA (4,1,1) con doce intervalos para los componentes autorregresivos y de promedios móviles es el más apropiado para el promedio mensual. Según los resultados, la radiación solar mensual media varía de 176-377 Wh/m². El error cuadrático medio (RMSE) fue de 33.18 Wh/m² para la radiación mensual, mientras que el RMSE fue de 104.26 Wh/m² para la radiación diaria.

V. Prema y K. Uma [49] desarrollan una serie de modelos de predicción a corto plazo para la irradiancia solar en Karnataka, India. Los pronósticos se realizan con un día de anticipación utilizando diferentes modelos de series de tiempo, incluyendo modelos de promedios móviles, de suavizado exponencial y de descomposición multiplicativa. Los modelos se alimentaron con los datos de la estación meteorológica de Karnataka en el

periodo comprendido entre el 01/01/2012 y el 30/05/2013, los primeros nueve meses de ese periodo se utilizaron para el ajuste de los modelos, mientras que los siguientes nueve meses se usaron para la validación. Los investigadores encontraron que el modelo de descomposición multiplicativa fue el que mejor funcionó para la predicción, mostrando un error porcentual medio absoluto (MAPE) de alrededor de 9.28%, también encontraron que en general los modelos probados ofrecen un desempeño aceptable para predicciones en un horizonte de hasta dos semanas.

Una metodología de predicción para la demanda de energía bajo una red inteligente en el campus de Bogotá de la Universidad Nacional de Colombia es aplicada por J. Rosero, A. García y O. Duarte [50]. Los autores realizan una caracterización de los patrones de consumo de la red con base en los días de la semana y los periodos académicos, seleccionando 5 tipos de curvas de carga diferentes. Posteriormente realizan el ajuste de un modelo de series de tiempo para las curvas de carga, los autores encontraron que un modelo ARIMA (1,2,1) se ajusta al comportamiento de la demanda de energía, las mejores predicciones las obtuvieron para los días no laborales con un RMSE promedio de 0.58%. Una comparación del desempeño de modelos de predicción de series de tiempo estacionales SARIMA y modelos de redes neuronales utilizadas para el pronóstico de la generación fotovoltaica de diferentes emplazamientos a lo largo de Grecia es presentado por E. Kardakos, M. Alexiadis, S y A. G. Bakirtzis [51]. Los autores encontraron que el modelo que mejor ajustaba los datos en términos del error cuadrático medio era un modelo SARIMA (3,1,2) x (3,1,2)₂₄ con un factor de corrección exógeno según el pronóstico del tiempo (11.12%). Los modelos de redes neuronales mostraron un RMSE similar (11.42% y 11.26%) en tanto el modelo SARIMA sin el factor de corrección dio un RMSE promedio anual de 12.89%.

Sujan Ghimire, C. Ravinesh y Nathan J.Downs [52] aplican un modelo de aprendizaje automático basado en redes neuronales para la predicción de la irradiancia solar en cinco localidades de Australia, los resultados fueron comparados con otras técnicas de inteligencia artificial así como modelos de series de tiempo. El entrenamiento y evaluación de la red neuronal se realizó utilizando datos meteorológicos desde 1979 hasta 2015 para los cinco emplazamientos seleccionados, la comparación de los resultados muestra que el modelo de redes neuronales dio como resultado un error medio cuadrático (RMSE) más

bajo (1.7–2.3 MJ m⁻²/día en relación con 2.1–5.9 MJ m⁻²/día), un RMSE relativo menor (9.07–12.47% vs. 10.98–29.15%), y un error medio absoluto (MAE) más bajo (7.97–11.74 vs 9.27–33.96).

Otro método de predicción utilizando redes neuronales orientado al pronóstico del índice de claridad, el cual es un parámetro que puede ser usado para el cálculo de la radiación global y difusa, fue presentado por T. Khatibab, A. Mohamed y K. Sopian [53]. El modelo planteado se basa en el modelo de percepción multicapa y tiene cuatro entradas y una salida. Las entradas son latitud, longitud, el número de día y la relación de la duración del día, la salida es el índice de claridad. Los autores utilizaron los datos de 28 estaciones meteorológicas para la investigación, 23 estaciones para entrenar la red, mientras que las otras 5 estaciones se usaron para probar la red. Además, los datos de irradiancia solar de las estaciones fueron utilizados para calcular la radiación global difusa con una ecuación propuesta por los autores para Malasia. El MAPE, el error de sesgo medio (MBE) y el RMSE para la irradiancia solar global pronosticada es del 5,92%, 1,46% y 7,96. El MAPE en la estimación de la irradiancia solar difusa es del 9,8%.

El modelo ANN + Wavelet, es un modelo de redes neuronales que utiliza una función ondícula o transformada de wavelet como función de activación. A. Yadav y L. Behera [54] utilizan un modelo ANN + Wavelet para predecir la radiación solar con una frecuencia horaria en Rajasthan, India, el modelo toma como entradas parámetros meteorológicos como temperatura, humedad, velocidad y dirección del viento, presión atmosférica, radiación normal directa (DNI), radiación horizontal difusa (DHI) y radiación global horizontal (GHI). Se alimentó el modelo con datos del periodo 1999-2013. Los resultados obtenidos muestran un RMSE de 14.96% cuando se usa la transformada wavelet y 18.29% sin la transformada wavelet, en tanto el MAE fue de 9.62% con la transformada wavelet y de 13.79% sin la transformada wavelet. Finalmente, el coeficiente de correlación R² fue de 79.12 para con la transformada wavelet y 67.12 sin la transformada wavelet.

Otra variante de los modelos de redes neuronales es conocida como modelo de redes adaptativas basadas en sistemas de inferencia difusa (ANFIS), este tipo de modelo combina los principios de las redes neuronales artificiales y la lógica difusa de operaciones tipo if-then para modelar los procesos de razonamiento. H. Yadav, Y. Pal y M. Tripathi [55] presentan una investigación en la que aplican un modelo ANFIS para la predicción en un

horizonte de 24 horas de la generación fotovoltaica en India. El modelo propuesto por los investigadores tiene cinco capas y tres operaciones de lógica difusa inmersas. El entrenamiento del modelo se realizó con los datos de la generación promedio del mercado indio por un año. Los resultados muestran que el modelo propuesto exhibe un MAPE de 4.1%, 3.1%, 9.3% y 3.9% para el invierno, verano, temporada de lluvias y otoño respectivamente. El desempeño del modelo propuesto se comparó con los de un modelo ARIMA y un modelo de redes neuronales clásico (ANN), los investigadores encontraron que el modelo ANFIS propuesto mejora los pronósticos de los modelos de referencia.

Otra técnica de inteligencia artificial que se ha aplicado para la predicción del recurso solar a corto plazo son las máquinas de vectores de soporte, estas son un conjunto de algoritmos de aprendizaje supervisado para la solución de problemas de clasificación y regresión. En su trabajo M. Z. Hassan, K. M. Ali y Jashnil Kumar [56] realizan un estudio de la radiación solar en el aeropuerto de Nadi en Fiji, Islas Fiyi para la radiación diaria durante el año 2015. Los investigadores prueban técnicas de regresión lineal como mínimos cuadrados y algoritmos de vectores de soporte usando diferentes funciones Kernel para generar un modelo de predicción. El modelo propuesto por los investigadores tiene una mejora en el RMSE de 32% en comparación con el modelo de referencia.

Un trabajo realizado para estimar la radiación solar incidente a nivel de superficie en distintas zonas de Argentina a partir de modelos estadísticos usando como entrada imágenes satelitales fue elaborado por R. Righini [57]. El autor utiliza fotografías tomadas por un satélite GOES 8 para calcular la irradiancia global promedio en cinco localidades de Argentina a través del modelo de Tarpey durante los periodos de mayo de 2000 y enero de 2001. Las constantes de ajuste del modelo fueron estimadas usando datos pinanométricos de tierra. Righini realizó pronósticos con un error medio cuadrático porcentual del 13% para el mes de enero y del 24% para el mes de mayo, el autor reporta un error medio cuadrático mucho más bajo para los pronósticos diarios.

R. Perez [58] propone un modelo para la predicción de la irradiancia solar horaria utilizando los datos del índice de nubosidad disponibles en la base de datos de pronósticos digitales nacionales del servicio meteorológico de Estados Unidos. En su trabajo presenta una evaluación del rendimiento del modelo frente a las irradiancias medidas en tierra y a través

de satélites en Albania, Nueva York. Perez divide sus predicciones en horizontes temporales de 0-4 horas, 4-8 horas, 8-26 horas y 26-76 horas, en sus resultados Perez reporta que la precisión de las predicciones disminuye al aumentar el horizonte de predicción, aunque el error medio cuadrático (RMSE) permanece estable en un rango de 30-40% después de validar con datos meteorológicos.

Por otra parte, Caldas M. y Suarez A. [59] presentan una metodología para predecir la irradiancia solar promedio en un horizonte de uno a diez minutos con base en una combinación de imágenes satelitales y mediciones de irradiancia solar en tiempo real en un lugar de Uruguay. La validación del modelo se realiza con una muestra de 5238 puntos de tiempo de pronóstico en un periodo de seis meses, la incertidumbre del pronóstico fue evaluada por separado para días claros, nublados y parcialmente nublados, con un error medio cuadrático (RMSE) de 11.4% para horizontes de pronóstico de 6 a 10 minutos en condiciones parcialmente nubladas, para condiciones despejadas el modelo alcanza un error máximo de 28.2% a un horizonte de 4 minutos hasta un mínimo de 16.3% para un horizonte de predicción de diez minutos en adelante.

5.2 Métodos de validación de predicciones

5.2.1 Métricas de precisión

En la literatura se encuentran diferentes métricas para cuantificar la precisión de los pronósticos de radiación solar y generación fotovoltaica. La selección del estadístico más adecuado depende de la necesidad, por ejemplo, los operadores del sistema necesitan métricas que castiguen efectivamente los costos de los errores en el pronóstico, mientras que para los investigadores son de más utilidad los indicadores de rendimiento de los diferentes modelos o de un solo modelo bajo diferentes condiciones meteorológicas [29]. Seleccionar adecuadamente el conjunto de datos de validación sobre el que se evaluarán las predicciones es crucial. El conjunto de datos de prueba debe excluir todos los datos que se usaron para entrenar los modelos, para que la evaluación se realice con datos independientes.

Uno de los estadísticos más utilizados es el error cuadrático medio (RMSE). El RMSE es ampliamente utilizado en problemas estadísticos de regresiones y por la industria energética para la evaluación del desempeño de predicciones del recurso solar, el RMSE está dado por la ecuación (16):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N}} \quad (16)$$

Donde Y_i es el valor observado de la variable, \hat{Y}_i es el valor predicho y N es el número de datos. El RMSE proporciona una buena estimación del error, ya que el término cuadrático da un mayor peso a los errores de mayor magnitud, no obstante, esto puede ser un obstáculo cuando se tienen errores extremos que puedan verse penalizados excesivamente. La principal desventaja del RMSE es que es una medida de error absoluto, por lo que no permite comparaciones cuando se trata de predicciones de diferentes órdenes de magnitud.

Otra métrica ampliamente utilizada en estadística es el error medio absoluto (MAE). El error medio absoluto se define según la ecuación (17):

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i - \hat{Y}_i| \quad (17)$$

El error medio absoluto al igual que el error medio cuadrático también es una métrica de medición del error global, pero a diferencia del RMSE no penaliza excesivamente los errores extremos de predicción.

Finalmente, la desviación media de parcialidad o sesgo (MBD) es un estadístico que permite calcular el sesgo de las predicciones, el MBD se define según la ecuación (18):

$$MBD = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i) \quad (18)$$

Este estadístico es particularmente útil para su aplicación en predicciones de generación eléctrica, ya que conocer el sesgo general del pronóstico permite a los operadores de sistemas de energía asignar recursos disponibles para compensar los errores de previsión en el proceso de despacho.

Las métricas introducidas arriba pueden ser normalizadas para permitir comparar el desempeño de predicciones en zonas con grandes diferencias meteorológicas, o en el caso de pronósticos de generación para comparar sistemas de diferentes tamaños, expresando su valor con respecto a un valor de referencia, típicamente se utiliza para esto el valor medio del conjunto de datos observados [40].

$$RMSE(\%) = \left(\frac{RMSE}{\bar{y}} \right) \cdot 100 \quad (19)$$

$$MAE(\%) = \left(\frac{MAE}{\bar{y}} \right) \cdot 100 \quad (20)$$

$$MBD(\%) = \left(\frac{MBD}{\bar{y}} \right) \cdot 100 \quad (21)$$

Donde \bar{y} es una estimación del valor medio del conjunto de datos de la predicción

$$\bar{y} = \sum_{i=1}^N \frac{Y_i}{N} \quad (22)$$

5.2.2 Evaluación del desempeño de predicciones

Dado que la precisión de los pronósticos depende en gran medida de la ubicación y el horizonte de predicción utilizado para la evaluación, así como de otros factores, es difícil evaluar la calidad de un pronóstico a partir de métricas de precisión solamente a través de las métricas de precisión. Se puede obtener más información comparando las precisiones de diferentes pronósticos con un conjunto común de datos de prueba.

Algunos métodos de predicción simples pueden servir como puntos de referencia para evaluar los pronósticos. Una referencia muy común para evaluar los pronósticos es el modelo de persistencia, que básicamente asume que las cosas permanecen igual, lo que en el caso de la radiación solar conlleva proyectar los valores pasados del índice de radiación en el futuro. Otra referencia comúnmente usada son pronósticos basados en condiciones climáticas normales a través de modelos auto-regresivos simples [60]. Cualquiera sea la referencia seleccionada, el índice de desempeño se define como:

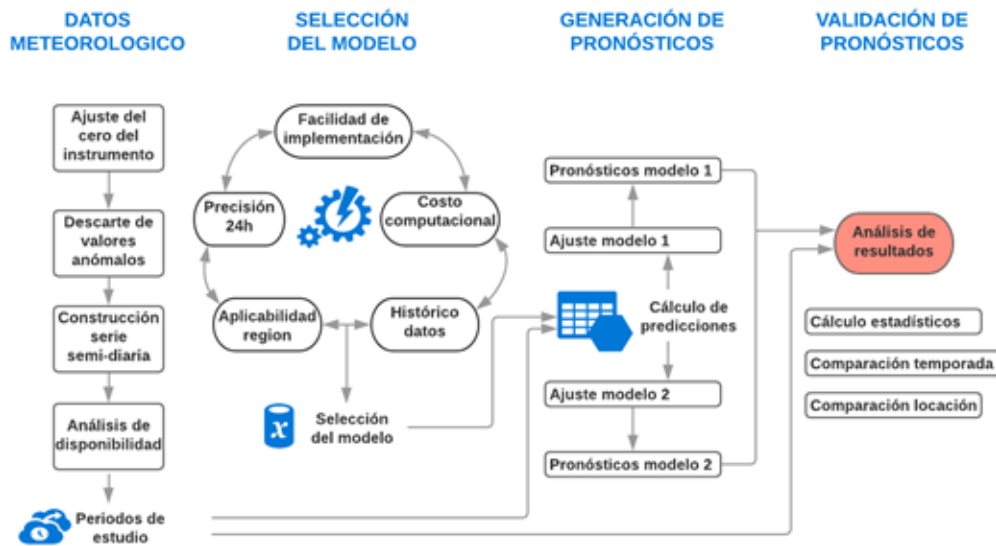
$$SS = \frac{metric_{reference} - metric_{forecast}}{metric_{reference} - metric_{perfect\ forecast}} \quad (13)$$

La evaluación de desempeño también se puede utilizar para identificar las condiciones en que los pronósticos funcionan relativamente bien o mal. Por ejemplo, los modelos numéricos de predicción meteorológica (NWP) exhiben un índice de desempeño menor con respecto al modelo de persistencia para horizontes de tiempo muy cortos (2-3 horas), indicando que los modelos NWP tienen muy poca habilidad para realizar predicciones en ese horizonte con respecto a horizontes más amplios.

6. Metodología

La metodología se compone de cuatro pasos principales: pre-procesamiento de datos meteorológicos, selección del modelo de predicción, ajuste del modelo y generación de predicciones y validación de predicciones. En la Figura 8 se muestra un esquema de la metodología.

Figura 6. Diagrama de flujo metodología



En este trabajo se utilizaron datos disponibles de la Red de Monitoreo de Calidad del Aire de Bogotá (RMCAB) de la Secretaría de Ambiente del Distrito. La red de calidad del aire cuenta con trece estaciones de monitoreo operativas distribuidas a lo largo de la ciudad, sus estaciones además de registrar concentraciones de contaminantes típicos (PM, SOX, NOX, CO, O3) para propósitos ambientales, también registran variables meteorológicas como precipitación, velocidad y dirección predominante del viento, temperatura, humedad relativa, presión atmosférica y radiación solar [61].

Los datos obtenidos por cada una de las estaciones se reciben en una estación central donde se someten a un proceso de validación final. Se usaron los datos de radiación para las estaciones que tengan disponibles registros de radiación solar promedio horaria.

6.1 Pre-procesamiento de datos

Para asegurar que el instrumento de medición está bien calibrado, se aplica un proceso consistente en verificar que durante la noche (entre las 06:00 p.m. y las 5:50 a.m.), cuando la lectura del instrumento que ser de cero, se registre efectivamente este valor; en el caso en que hubiera un registro por encima del cero, dicho valor se sustrae de la medición del día siguiente para asegurar la calibración correcta.

Además, se deben descartar lecturas extremas que superen los límites físicos para la radiación, para esto se comparan las mediciones registradas con el cálculo de la irradiación solar extra-terrestre máxima para ese periodo del año, en el caso en que las lecturas superen ese valor, se procede a descartar los registros.

Los registros obtenidos fueron filtrados para tener series de datos únicamente en el periodo de tiempo en que los instrumentos registran valores de radiación, es decir que solamente estamos interesados en las mediciones durante el día. Bogotá está ubicada 4° al norte del ecuador terrestre, por lo que la duración del día se puede aproximar como constante y de alrededor de doce horas a lo largo del año, por esto se descartaron todos los registros de obtenidos en el periodo entre las 07:00 p.m. y las 05:00 a.m.

Finalmente, es de vital importancia el análisis de la disponibilidad de los datos radiométricos, para asegurar que estos tienen una continuidad temporal y son adecuados para ajustar un modelo de predicción.

6.2 Selección del modelo

Como se expuso antes, los modelos de predicción para la irradiación solar se pueden dividir en modelos físicos y modelos estadísticos. Los modelos físicos se basan en la resolución a través de métodos numéricos de las ecuaciones que gobiernan la dinámica

atmosférica, apoyados por mediciones meteorológicas, imágenes satelitales y/o fotografías cielo-tierra, requieren de información precisa y extensa sobre las condiciones iniciales de la atmósfera. Proveen soluciones en escalas espaciales amplias por lo que son utilizados para el desarrollo de mapas de radiación solar. Los modelos físicos entregan soluciones en una escala global o meso-escala, por lo que se debe aplicar un post-procesado para considerar las condiciones locales del emplazamiento.

Por otra parte, los modelos estadísticos son modelos de tipo caja negra, no requieren de un conocimiento detallado de la dinámica atmosférica y son fácilmente implementables por su sencillez, no necesitan grandes recursos computacionales para su desarrollo en comparación con los modelos físicos. La principal desventaja de estos es que están restringidos al emplazamiento en que se desarrolla la implementación, requieren de datos específicos de la locación y son muy sensibles al tamaño y la calidad de los datos de entrada, ya que deben ser entrenados con todas las posibles situaciones que se puedan presentar para tener un desempeño adecuado.

Para este trabajo se realizó una revisión bibliográfica de modelos aplicados para predicción del recurso solar en un horizonte de corto plazo (predicción del tipo del día siguiente), los modelos analizados fueron evaluados y se eligió el más apropiado para ser implementado en la sabana de Bogotá a través de una matriz de selección de 5 criterios, los criterios de selección y la matriz de evaluación se muestran en las Tablas 1 y 2.

Tabla 1. Matriz de selección de modelo

Criterio	Modelos				
	I	II	III	IV	V
Facilidad de implementación					
Uso de recursos computacionales					
Histórico de datos					
Aplicabilidad en distintas regiones					
Precisión para pronósticos del día siguiente					

Tabla 2. Matriz de calificación

	Valoración	Efecto
Muy alto	5	Tiene un alto impacto positivo para la implementación
Alto	4	Afecta positivamente la implementación
Medio	2	No afecta la implementación
Bajo	1	Afecta negativamente la implementación

Se valoraron los siguientes criterios para la selección del modelo: facilidad de implementación, consumo de recursos computacionales para la simulación, volumen de datos requeridos para el entrenamiento del modelo, posibilidad de implementación en otros emplazamientos y precisión reportada en la literatura para pronósticos del día siguiente. A cada uno de estos cinco criterios se les asignó una valoración entre 0 y 5 dependiendo de que tanto afecta positivamente la implementación del modelo, siendo cinco la mayor afectación positiva para la aplicación del modelo.

6.3 Ajuste y generación de predicciones

Una vez seleccionada la metodología de predicción, se procede con el entrenamiento y ajuste del modelo. Como se expuso antes, la cantidad y calidad de los datos de entrada es vital para el desempeño adecuado del modelo de pronóstico. Se analizará la disponibilidad de registros de cada una de las estaciones meteorológicas de las que se cuente con datos, para seleccionar un periodo de registros con una muy buena continuidad. En la literatura para pronósticos de irradiación promedio horaria se recomienda que se tengan datos ininterrumpidos por un periodo de por lo menos tres años, pero preferiblemente de más de once años para poder considerar los efectos de la variación de la irradiación solar en el largo plazo debido a los ciclos solares.

Con base en el análisis de la disponibilidad de los datos de las estaciones meteorológicas, se decidirá si los datos son adecuados para implementar un modelo de pronóstico en cada emplazamiento y se seleccionará el periodo de datos más adecuado para el ajuste del modelo. Se generarán predicciones de la irradiación solar promedio horaria por un periodo de por lo menos catorce días para la evaluación de desempeño del modelo en las diferentes locaciones seleccionadas.

6.4 Validación del modelo

A partir de las predicciones de la irradiación solar obtenidas en el paso anterior, se realizó la validación del modelo seleccionado para comprobar su precisión e idoneidad para representar el comportamiento de la radiación solar semi-diaria en cada uno de los emplazamientos seleccionados.

Las predicciones de los modelos fueron evaluadas tomando como referencia las mediciones de radiación de las estaciones meteorológicas. Para esta comparación se emplearon los estadísticos de: error medio cuadrático (RMSE), error medio absoluto (MAE) y desviación media de sesgo (MBD). Así mismo se determinó el coeficiente de correlación R^2 para decidir si el modelo provee un ajuste adecuado de los datos.

El periodo de registros utilizado para la evaluación de desempeño del modelo también fue seleccionado tomando como base el análisis de la disponibilidad de mediciones de las estaciones meteorológicas llevado a cabo en el paso anterior. Se generaron predicciones por siete días y se tomaron dos periodos diferentes para cada emplazamiento para identificar el efecto del clima sobre las predicciones.

7. Análisis de resultados

A continuación, se presentan los resultados del trabajo, primero se muestra el pre-procesamiento realizado a los datos de la Red de Monitoreo de Calidad del Aire de Bogotá, se recibieron datos de cinco estaciones meteorológicas que cuentan con registros de irradiación promedio horaria, posteriormente se expone la selección del modelo, se seleccionaron dos modelos de tipo estadístico para el pronóstico de la irradiación, después se analiza el comportamiento de la irradiación promedio horaria a lo largo del año a partir de diagramas de cajas, se presenta el proceso de ajuste para cada uno de los dos modelos seleccionados, las predicciones generadas y finalmente se analizan los resultados en términos del error cuadrático medio RMSE y el error medio absoluto MAE, se establecen comparaciones entre los emplazamientos y en diferentes periodos en el mismo emplazamiento.

7.1 Pre-procesamiento de datos

Se obtuvieron datos correspondientes a la irradiación global en superficie provenientes de 5 estaciones meteorológicas de la Red de Monitoreo de Calidad del Aire de Bogotá (RMCAB) que cuentan con registros disponibles de promedios horarios, los datos fueron suministrados por la secretaría de ambiente distrital. En las Tablas 2 y 3 se muestra la ubicación geográfica y las características principales de cada una de las estaciones meteorológicas de las que se utilizó información.

Tabla 3. Ubicación estaciones RMCAB [62]

Ubicación estaciones meteorológicas						
Estación	Latitud	Longitud	Altitud	Localidad	Dirección	Tipo de zona
Guaymaral	4° 47' 1,5" N	74° 2' 39,1" W	2580 m	Suba	Autopista Norte #205-59	Suburbana
Parque Simón Bolívar	4° 39' 30,5" N	74° 5' 2,3" W	2577 m	Barrios Unidos	Calle 63 # 59A-06	Urbana
Kennedy	4° 37' 30,2" N	74° 9' 40,8" W	2580 m	Kennedy	Carrera 80 # 40-55 sur	Urbana
Tunal	4° 34' 34,4" N	74° 7' 51,4" W	2589 m	Tunjuelito	Carrera 24 # 49-86 sur	Urbana
San Cristóbal	4° 34' 21,2" N	74° 5' 1,7" W	2688 m	San Cristóbal	Carrera 2 este # 12 - 78 sur	Urbana

Tabla 4. Características estaciones RMCAB [62]

Estación	Altura del suelo	Altura punto de muestra al suelo	Altura viento al suelo
Guaymaral	0 m	4 m	10 m
P. Simón Bolívar	0 m	4,6 m	10 m
Kennedy	3 m	7 m	10 m
Tunal	0 m	3 m	10 m
San Cristóbal	0 m	3 m	10 m

Se obtuvieron datos para el periodo correspondiente entre el 01 de enero de 2010 y el 30 de septiembre de 2019. Los datos suministrados por la Red de Monitoreo de Calidad del Aire fueron sometidos a un pre-procesamiento para descartar valores anómalos y para construir las series de tiempo semi-diaras que serán utilizadas para el entrenamiento y evaluación del modelo.

Es de vital importancia el análisis de la disponibilidad de los datos radiométricos, para asegurar que estos tienen una continuidad temporal y son adecuados para ajustar un modelo de predicción.

Para analizar la disponibilidad de las estaciones, los datos obtenidos fueron ordenados por años y por meses y se cuantificaron los días en que la estación meteorológica estuvo disponible, en aquellos casos en los que no se contó con registros de promedios horarios las doce horas completas se procedió a descartar la serie de datos. El resumen de la disponibilidad de las estaciones para cada uno de los años en que se tiene registros se muestra en la Tabla 5.

Tabla 5. Disponibilidad de la estación meteorológica

Estación	Año	E	F	M	A	M	J	J	A	S	O	N	D	Disponib
Centro de Alto Rendimiento Deportivo	2010	31	28	28	26	28	26	31	29	22	29	30	31	92,9%
	2011	31	28	31	30	31	30	23	28	30	30	27	31	95,9%
	2012	28	29	31	29	31	30	31	31	30	31	25	31	97,5%
	2013	31	28	31	30	31	30	31	30	30	28	30	31	98,9%
	2014	31	28	30	30	31	30	27	30	19	31	30	30	95,1%
	2015	31	28	31	30	31	30	31	31	29	31	30	31	99,7%
	2016	31	29	31	30	31	30	31	31	30	31	30	31	100,0%
	2017	31	25	10	30	31	30	31	31	30	31	30	12	88,2%
	2018	26	28	31	30	20	30	31	31	30	31	30	31	95,6%
2019	31	28	31	30	31	30	31	31	30	-	-	-	100,0%	
San Cristóbal	2012	29	29	28	30	31	30	27	26	28	30	13	28	89,9%
	2013	28	28	18	13	28	30	31	19	0	10	28	18	68,8%
	2014	31	28	29	30	31	30	31	7	19	30	30	31	89,6%
	2015	31	28	31	30	31	30	31	31	30	31	30	31	100,0%
	2016	29	29	31	30	31	30	31	31	30	31	30	31	99,5%
	2017	31	28	31	30	31	30	31	31	30	31	30	31	100,0%
	2018	31	28	31	30	31	30	31	31	30	30	30	31	99,7%
	2019	31	28	31	30	31	30	31	29	30	-	-	-	99,3%
Guaymaral	2010	31	25	31	28	29	29	29	20	16	24	0	0	71,8%
	2011	28	28	25	30	31	30	17	30	30	28	29	31	92,3%
	2012	31	29	28	29	31	20	31	31	30	31	30	29	95,6%
	2013	29	28	31	30	31	30	31	25	30	31	30	31	97,8%
	2014	31	28	31	30	31	29	30	30	30	31	29	31	98,9%
	2015	31	28	31	30	31	30	31	30	30	31	30	31	99,7%
	2016	31	29	31	30	31	30	31	31	30	24	30	25	96,4%
	2017	31	28	31	30	31	30	31	31	30	31	30	25	98,4%
	2018	31	28	31	30	29	30	31	31	30	31	23	31	97,5%
	2019	31	28	31	30	31	30	31	31	30	-	-	-	100,0%
Kennedy	2010	31	28	31	29	30	24	31	30	26	31	30	29	95,9%
	2011	30	27	31	30	31	29	18	0	18	31	29	30	83,3%
	2012	31	29	30	30	29	30	31	21	30	31	30	31	96,4%
	2013	31	28	31	30	31	30	31	31	30	31	30	31	100,0%

Estación	Año	E	F	M	A	M	J	J	A	S	O	N	D	Disponib
	2014	29	28	31	30	31	29	31	31	30	31	30	31	99,2%
	2015	31	28	31	30	31	29	31	31	30	31	30	31	99,7%
	2016	28	17	15	30	27	30	31	31	30	31	30	28	89,6%
	2017	28	28	29	30	31	30	31	31	30	31	30	31	98,6%
	2018	31	28	31	30	31	30	31	31	30	31	30	31	100,0%
	2019	31	27	31	30	31	30	31	31	30	-	-	-	99,6%
Tunal	2010	18	28	20	30	28	30	28	27	30	31	24	30	88,8%
	2011	31	28	31	30	31	30	31	29	30	31	30	31	99,5%
	2012	31	29	24	27	31	28	31	25	30	31	30	31	95,1%
	2013	31	28	23	30	31	28	31	31	30	31	30	27	96,2%
	2014	29	28	31	30	31	30	31	31	30	31	30	31	99,5%
	2015	31	28	31	30	31	30	31	31	30	30	30	31	99,7%
	2016	31	29	31	30	31	30	31	31	30	31	30	31	100,0%
	2017	30	28	31	30	31	30	31	29	29	31	30	31	98,9%
	2018	31	28	31	30	31	30	31	31	30	31	30	31	100,0%
	2019	31	28	31	30	31	30	31	31	30	-	-	-	100,0%

De acuerdo con lo expuesto antes, en la literatura se recomienda la selección de periodos de estudio con una disponibilidad de registros mayor al 99%; por esto se utilizó un código de colores para facilitar la selección del periodo, representando el color verde cuando el mes tiene una disponibilidad mayor al 96,6% y los datos pueden ser adecuados si hay una continuidad temporal en los registros.

En la Tabla 5 se observa que las estaciones meteorológicas analizadas tuvieron cortes operativos ocasionales, algunas veces por periodos de tiempo largos, pero en general hay una continuidad aceptable de las mediciones y se pueden elegir periodos de estudio de más de dos años para cada estación con una disponibilidad mayor al 99%. Para la selección del periodo de estudio de cada una de las estaciones se tuvo en cuenta el criterio de que estos periodos coincidieran en lo posible para poder comparar el análisis de los registros y los pronósticos generados para cada una de las estaciones entre estos.

Los periodos de estudio seleccionados se muestran en la Tabla 6.

Tabla 6. Disponibilidad de estaciones RMCAB

Estación	Periodo	Disponibil
CDAR	01 de octubre de 2014 - 31 de enero de 2017	99,86%
San Cristóbal	01 de noviembre de 2014 - 31 de julio de 2019	99,78%
Guaymaral	01 de septiembre de 2013 - 15 de octubre de 2016	99,54%
Kennedy	01 de septiembre de 2012 - 31 de diciembre de 2015	99,76%
Tunal	01 de febrero de 2014 - 31 de julio de 2019	99,91%

Para la evaluación del pronóstico se tuvo cuidado de seleccionar periodos en los que se tenga la mayor cantidad de estaciones meteorológicas disponibles para poder hacer comparaciones entre emplazamientos. Además, se generaron predicciones para dos periodos del año para estudiar el efecto del clima sobre la precisión del pronóstico: se seleccionaron los siguientes periodos de predicción: 9–15 de octubre de 2016 y 18–24 de mayo de 2016 para las estaciones CDAR, San Cristóbal, Guaymaral y Tunal y 23–29 diciembre de 2015 para las estaciones Guaymaral y Kennedy.

7.2 Selección del modelo

Para la selección del método de predicción se evaluaron los siguientes modelos: modelo de series de tiempo aplicado por Mohammed H. Alsharif, M. Younes en Seúl, Corea del Sur, modelo de redes neuronales (ANN) de Sujun Ghimier y C. Ravinesh aplicado en diferentes localidades de Australia, modelo de máquinas de soporte de vectores (SVM) de M. Hassan, K. Ali y Jashnil Kumar utilizado en el aeropuerto de Islas Fiji, modelo numérico meteorológico de Pérez para el estado de Nueva York y metodología Total Sky Imager (TSI) utilizado por Caldas y Suarez en Uruguay.

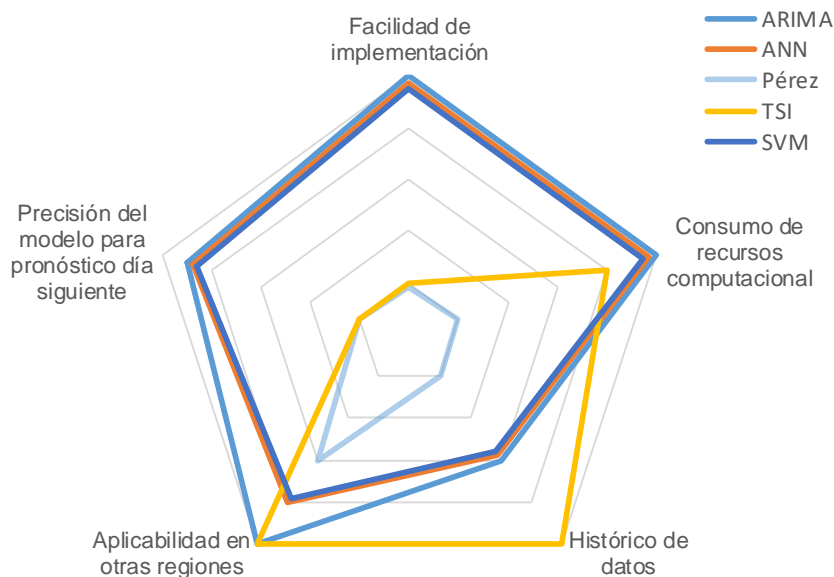
En la Tabla 7 se muestra la matriz de selección del modelo, junto con los resultados obtenidos después de la evaluación, para la selección del modelo se asignó a cada uno de los cinco criterios un valor entre 1 y 5, siendo 5 el que más afectaba positivamente la implementación del modelo y 1 que lo hacía más difícil. En la Figura 7 se muestra un gráfico de radar para la selección del modelo.

Tabla 7. Matriz de evaluación de modelos de predicción

Criterio	Modelos				
	ARIMA	ANN	Pérez	TSI	SVM
Facilidad de implementación	5	5	0,9	1	5
Consumo de recursos computacional	5	5	1	4	5
Histórico de datos	3	3	1	5	3
Aplicabilidad en otras regiones	5	4	3	5	4
Precisión del modelo para pronóstico día siguiente	4,5	4,5	1	1	4,5
Promedio	4,5	4,3	1,38	3,75	4,3

El modelo meteorológico de Pérez tiene varios inconvenientes para su implementación: requiere de una vasta cantidad de datos de variables atmosféricas tomados en diferentes puntos geográficos para poder desarrollar la solución, utiliza grandes cantidades de recursos computacionales para su implementación y no muestra buenos resultados para predicciones del tipo de pronósticos para el día siguiente.

Figura 7. Selección del modelo de predicción



La metodología de imágenes cielo tierra (TSI) tiene dos grandes ventajas: por una parte, no requiere de registros históricos de datos, y además la metodología se puede aplicar en otras latitudes; sin embargo, para su implementación se precisa de equipos de fotografía y adquisición de datos con los que la universidad no cuenta en estos momentos, y aunque muestra buenos resultados en el muy corto plazo (15 minutos a una hora), no permite generar pronósticos para el día siguiente.

Dentro de los modelos analizados se examinaron tres modelos estadísticos: uno de series de tiempo (ARIMA) y dos técnicas de inteligencia artificial (redes neuronales y vectores de soporte). Estos modelos tienen prácticamente las mismas fortalezas y debilidades: requieren de un volumen de datos de entrada relativamente pequeño en comparación con los modelos meteorológicos, en la literatura se reporta una muy buena precisión para pronósticos para el día siguiente en los emplazamientos seleccionados, tienen un bajo uso de recursos computacionales y son fácilmente implementables. La única diferencia remarcable está en la sensibilidad de los modelos de inteligencia artificial al cálculo de los parámetros. Se decide aplicar un modelo de series de tiempo y un modelo de redes neuronales.

7.3 Análisis de la irradiación

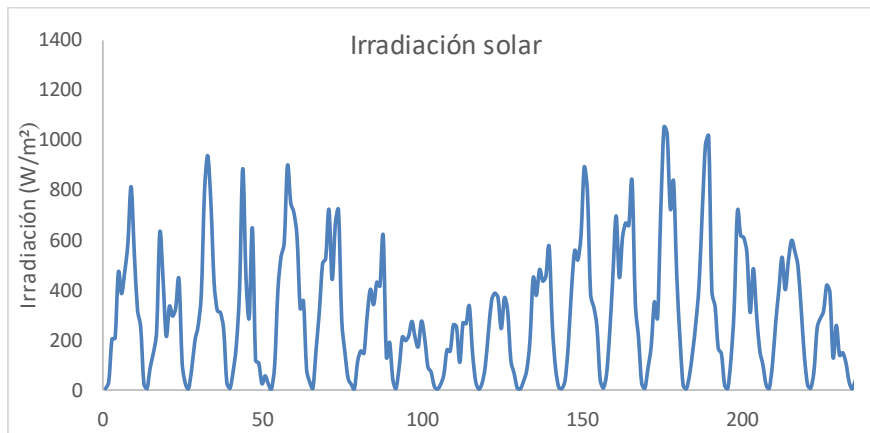
7.3.1 Comportamiento de la irradiación solar horaria promedio

La irradiación solar en superficie se ve afectada por factores geográficos y meteorológicos, principalmente es influenciada por la época del año y la latitud, ya que dependiendo de la posición geográfica del emplazamiento la duración del día varía a lo largo del año; no es este el caso de estudio ya que los emplazamientos seleccionados están todos cerca al ecuador, por lo que la duración del día se puede aproximar como constante a lo largo del año. Otro factor que influencia significativamente la irradiación en superficie es la elevación del sol, que depende también de la latitud y de la época del año, por esto las regiones polares a pesar de recibir más luz en verano reciben un nivel de radiación bajo o moderado.

Además, factores meteorológicos como la nubosidad hacen que la radiación solar disminuya, la cantidad de radiación atenuada por la nubosidad depende del tipo de nube,

de su desarrollo y de su posición, por lo que esta variable es difícil de incorporar en modelos empíricos. Otros parámetros como la altitud, la contaminación atmosférica y el albedo también pueden tener una influencia significativa, no obstante, los emplazamientos seleccionados se encuentran aproximadamente bajo las mismas condiciones, por lo que estos otros parámetros no muestran una influencia significativa. En la Figura 8 se muestra el comportamiento de la irradiación global promedio horaria para tres semanas del mes de octubre de 2014.

Figura 8. Comportamiento de la irradiación solar promedio



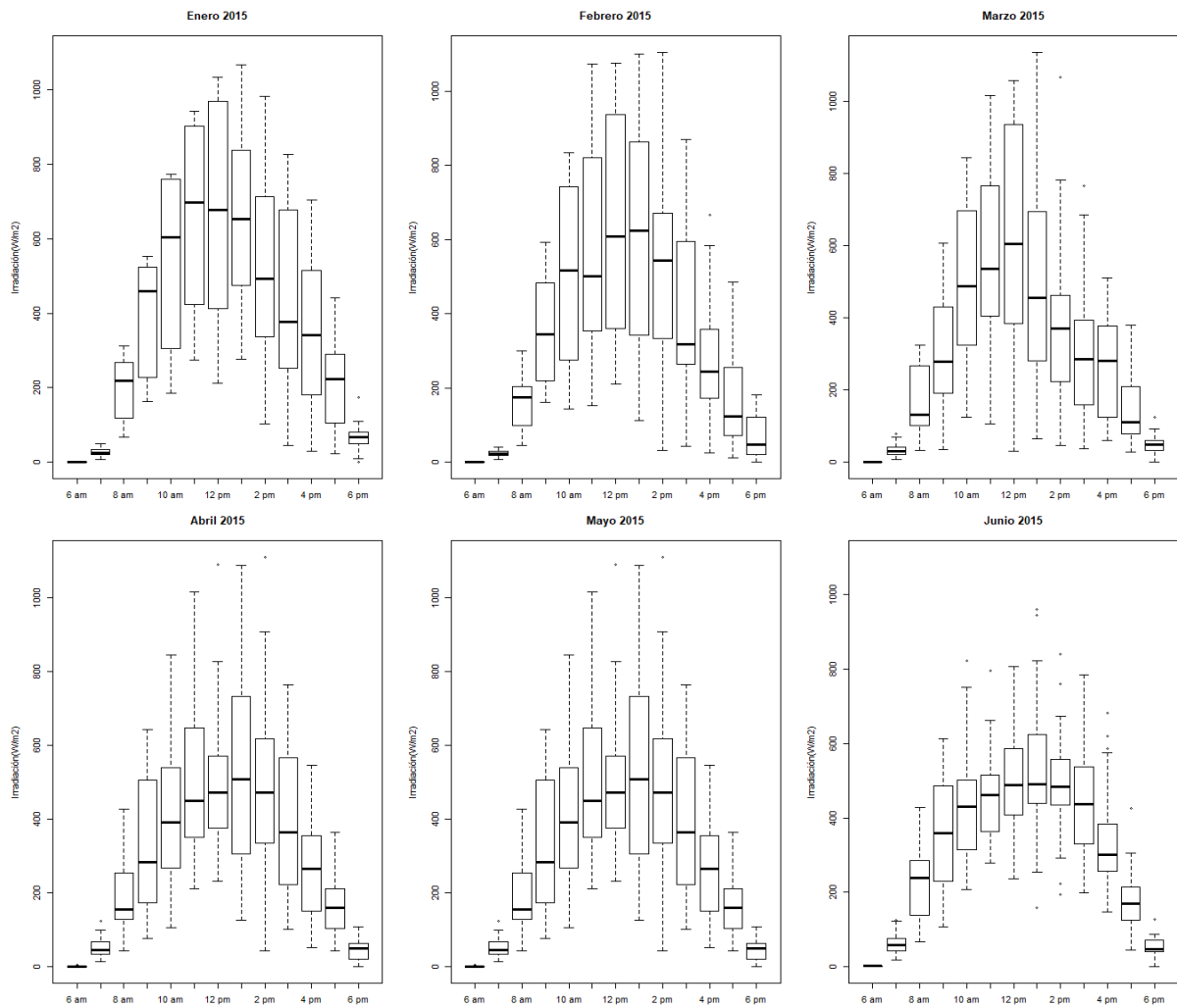
En la gráfica anterior se observa que la irradiación global en superficie tiene un comportamiento cíclico y similar a uno parabólico, alcanzando su valor máximo alrededor del mediodía, cuando la elevación del sol es máxima; sin embargo, la presencia de nubes o precipitación afectan la irradiación solar global disminuyendo su valor promedio significativamente, además, las rachas de nubosidad también pueden causar oscilaciones en la irradiación como se puede apreciar para algunos días del mes de octubre.

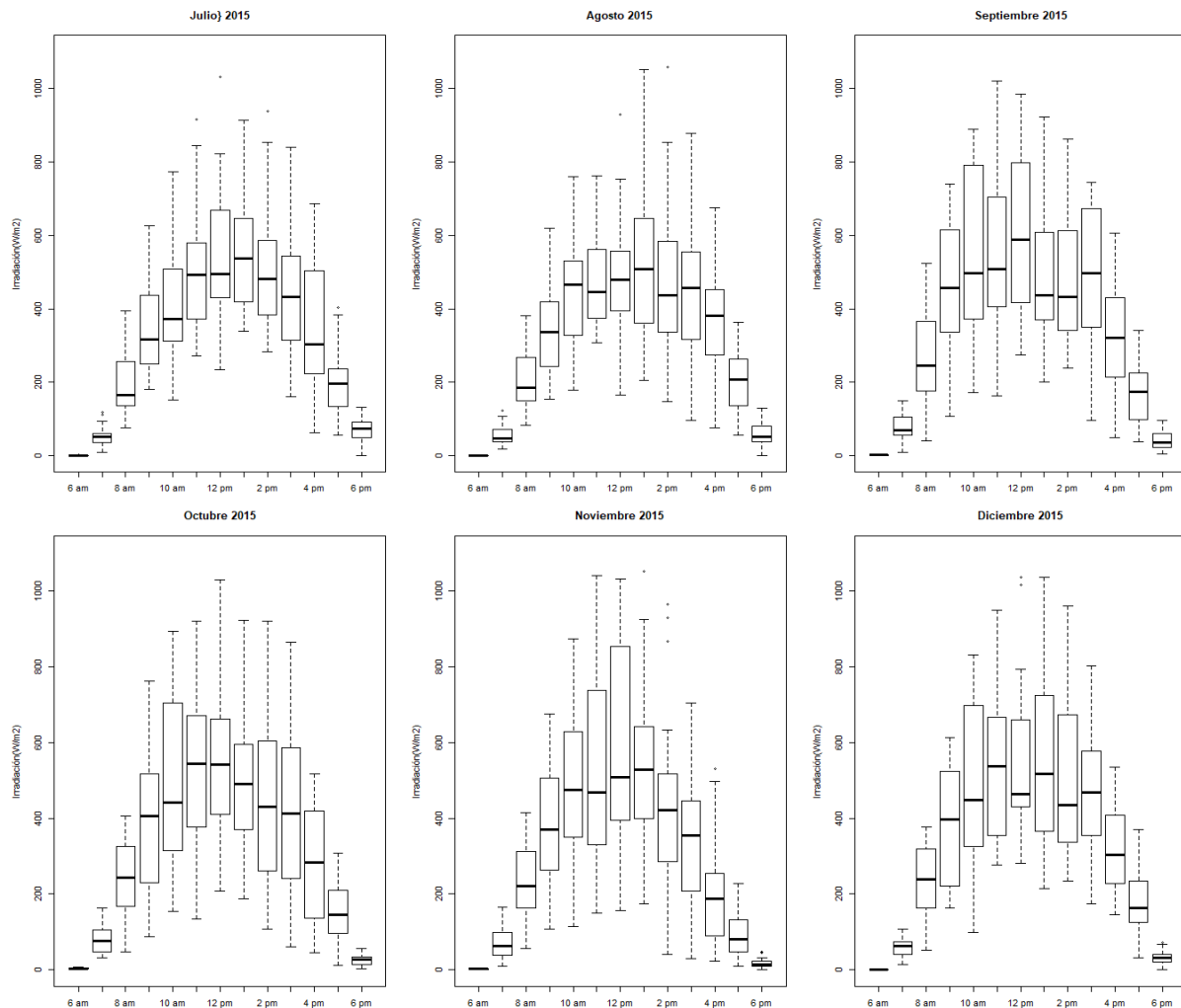
7.3.2 Análisis de la variabilidad de la irradiación solar

Una forma más adecuada para analizar el comportamiento de la irradiación global en superficie durante los diferentes meses del periodo de estudio es a través de diagramas de Box-Whisker. Los diagramas de Box-Whisker o diagramas de cajas son gráficos de tipo histograma que permiten visualizar conjuntamente la media y varianza de un conjunto de datos [63]. Para representar un diagrama de cajas se dibuja un rectángulo desde el cuartil

1 (Q1) hasta el cuartil 3 (Q3), la media se dibuja como una línea inscrita dentro del rectángulo, en tanto se traza una línea perpendicular discontinua que parte de los cuartiles Q1 y Q3 y cuya longitud corresponde a la distancia intercuartílica, en el caso de que se tengan valores por fuera de esta línea, éstos se representan como puntos. Los diagramas de cajas para los datos de irradiación en superficie de la estación CDAR a lo largo del año 2015 se muestran en la Figura 9.

Figura 9. Diagrama de cajas irradiación CDAR 2015



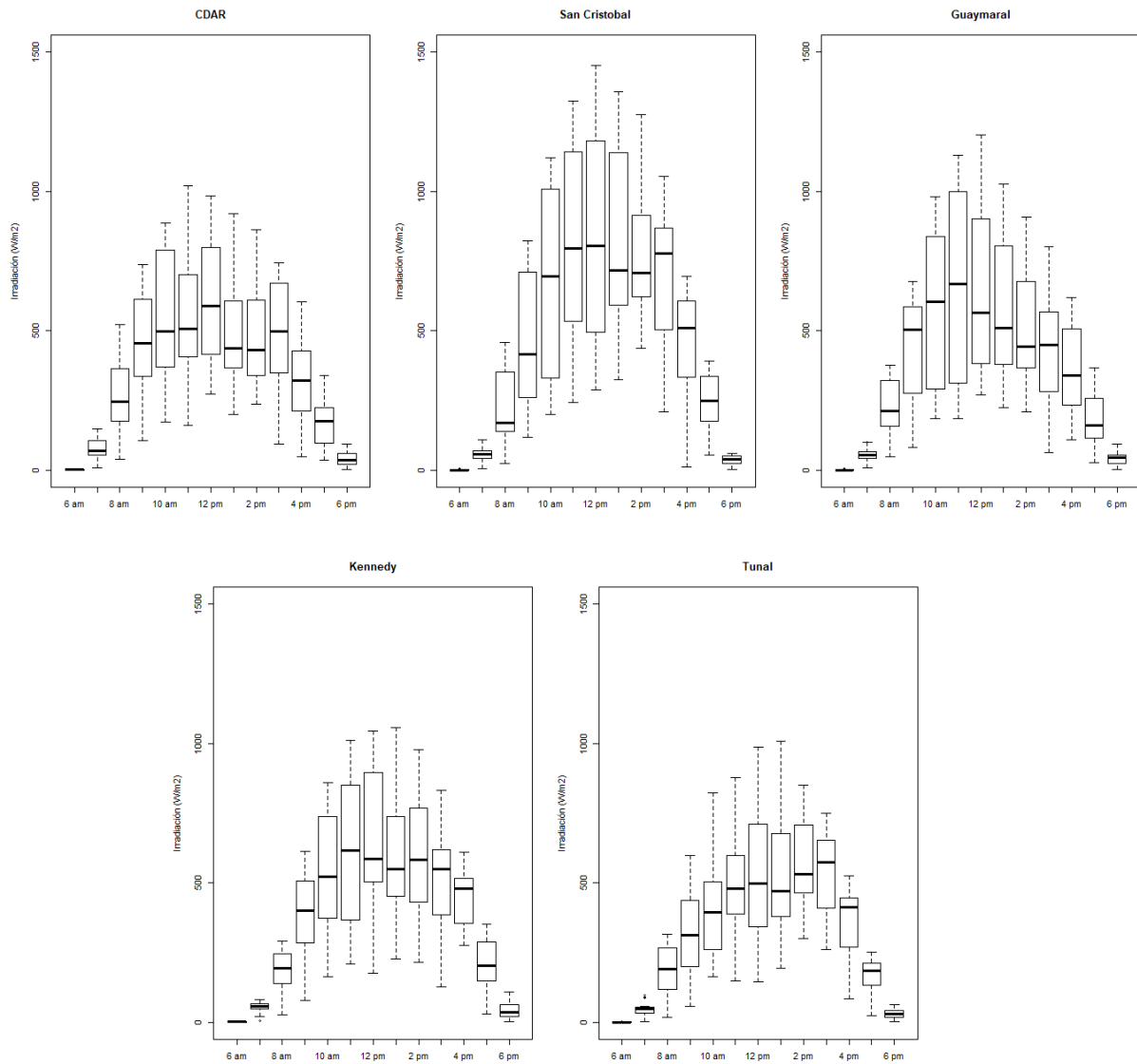


En la Figura 9 se observa que la variabilidad de la irradiación es significativa a lo largo del año, los meses donde hay una menor irradiación promedio son abril, mayo y octubre, por el contrario, durante los meses de enero, febrero y septiembre la irradiación promedio alcanza los valores más altos. Finalmente, la varianza de los datos es más alta durante los meses predominante soleados (enero, febrero y septiembre), no obstante, los meses en los que se está en transición entre temporada de lluvias y temporada seca también se presenta una variación significativa de la irradiación global promedio.

7.3.3 Variación entre emplazamientos de estudio

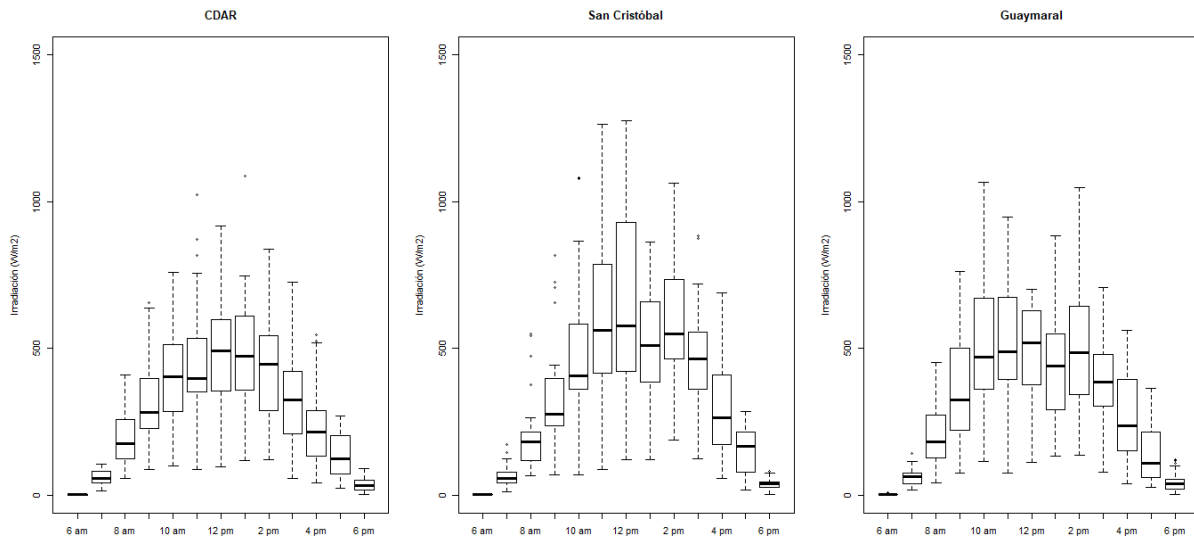
La variación de la irradiación global en superficie entre los emplazamientos de estudio seleccionados también se analizó a través de diagramas de cajas. A continuación, se muestran los diagramas de cajas para la irradiación global solar promedio horaria durante los meses de septiembre de 2015 y abril de 2016 para las cinco estaciones de las que se cuenta con datos de irradiación disponibles: Centro de Alto Rendimiento Deportivo (CDAR), San Cristóbal, Tunal, Kennedy y Guaymaral.

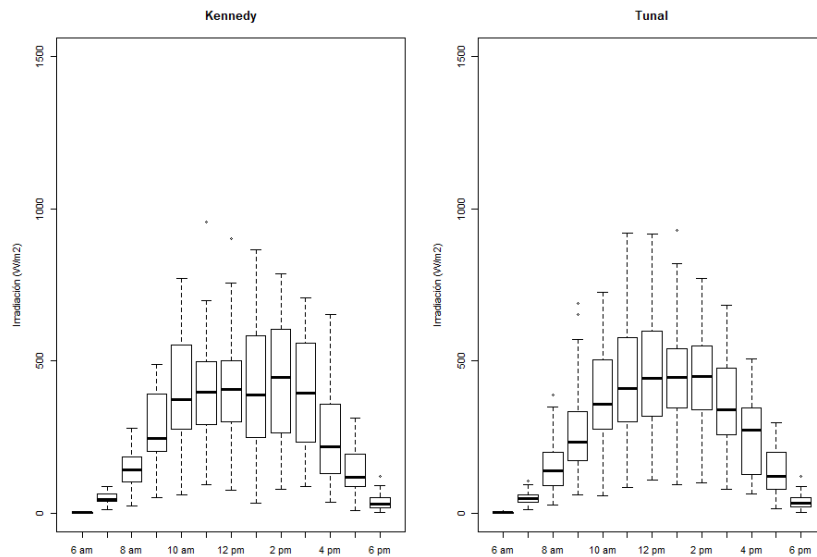
Figura 10. Irradiación global promedio horaria mes de septiembre de 2015



Los meses seleccionados representan el comportamiento típico de la irradiación global en superficie durante la temporada seca (septiembre) y la temporada de lluvias (abril). En la Figura 10 se observa que, para el mes de septiembre, la estación que presenta los registros más altos de irradiación es la de San Cristóbal, esto puede atribuirse a que por su posición geográfica esta estación se encuentra a una mayor altitud que las demás (ver Tabla 3) por lo que la irradiación global es mayor al depender esta de la altura, al contrario, las estaciones que presentan menores registros de irradiación son las de Tunal y CDAR. Por otra parte, al analizar el comportamiento de la irradiación global en superficie para la temporada de lluvias (ver Figura 11, abril 2016) se observa que, bajo condiciones de alta nubosidad, las diferencias en la irradiación promedio horaria son considerablemente menores entre los cinco emplazamientos estudiados. Esto puede explicarse porque bajo condiciones de alta nubosidad la radiación solar difusa es la que más contribuye a la irradiación global, y esta se ve afectada principalmente por la nubosidad, la cual es muy alta en todos los cinco emplazamientos durante la temporada de lluvias.

Figura 11. Irradiación global promedio horaria mes de abril de 2016





Finalmente, se debe anotar que tanto durante la temporada de lluvias como durante la temporada seca se observa que la variación de la irradiación en superficie es mucho más alta en la estación meteorológica de San Cristóbal comparada con las demás estaciones estudiadas; de hecho, para el mes de mayo de 2016 se observa una cantidad considerable de registros fuera del rango de $2/3Q1$ y $2/3Q4$. Esto puede atribuirse a que la estación se encuentra muy cerca a los cerros orientales, en esta se presentan microclimas dominados por cambios bruscos en la nubosidad debido a la alta humedad relativa y velocidad del viento, se pueden presentar condiciones soleadas y lluviosas durante el mismo día.

7.4 Modelo de series de tiempo

Para el ajuste del modelo de pronóstico se utilizó la metodología de Box-Jenkins [64]. La metodología de Box-Jenkins comprende los procesos aplicados para identificar, ajustar y evaluar un modelo de tipo ARMA, ARIMA o SARIMA apropiado para describir una serie temporal. Esta metodología se compone de tres pasos principales:

1. Identificación del modelo: esta fase consiste en la identificación del tipo de proceso estocástico al que corresponden los datos para encontrar los valores adecuados de los parámetros p , d y q del modelo; esto se hace a partir del análisis de los gráficos o correlogramas de las funciones auto-correlación (ACF) y auto-correlación parcial (PACF). Además, en esta etapa se realiza una prueba de estacionariedad a la serie para identificar si cumple con los supuestos de estacionariedad. A partir

del análisis de los correlogramas también se identifica si la serie posee algún tipo de estacionalidad que deba ser incluida en el modelo.

2. Estimación de parámetros: en esta etapa se estiman los coeficientes de los términos autorregresivos, de medias móviles y de rezago (p , d y q) del modelo seleccionado, la estimación se realiza a través de los algoritmos de cálculo, usando algún paquete estadístico, los algoritmos más utilizados son los de verosimilitud o de mínimos cuadrados no lineales.
3. Comprobación del modelo: en esta fase se realizan algunas pruebas al modelo para corroborar que éste se ajusta a las especificaciones de un proceso univariado estacionario. Se efectúan pruebas para garantizar que el modelo estimado cumple con las condiciones de estacionariedad e invertibilidad y que existe significancia estadística en los rezagos incorporados, también se analizan los residuales del modelo para verificar que estos son un proceso totalmente aleatorio y por tanto son independientes y tienen una media y varianza constante. Si la estimación no es adecuada se debe regresar al paso uno y buscar un modelo mejor.

7.4.1 Identificación del modelo

7.4.1.1 Prueba de Dick and Fulley

El primer paso para el ajuste de un modelo de series de tiempo es la realización del test de Dick and Fulley o prueba de estacionariedad para identificar si la serie se comporta como una serie estacionaria, es decir que la media y la varianza son constantes en el tiempo. En caso de que la serie no sea estacionaria, la variable se diferencia d veces hasta que la serie obtenida sea de tipo estacionario; el número de operaciones de diferencias d representa el orden de integración del modelo ARIMA.

Los resultados mostrados de aquí en adelante se obtuvieron utilizando los softwares estadísticos E-VIEWS y R-studio, el cual es ampliamente usado para el ajuste de modelos de tipo de series de tiempo. En la Figura 12 se muestran los resultados del test de Dick and Fulley aplicado a la serie de tiempo de irradiación global de la estación del centro de alto rendimiento deportivo (CADR) durante el periodo de estudio seleccionado.

Figura 12. Prueba de Dick and Fulley (CADR) [65]

Null Hypothesis: W_M_ has a unit root
 Exogenous: Constant
 Lag Length: 37 (Automatic - based on SIC, maxlag=37)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-10.09633	0.0000
Test critical values:		
1% level	-3.430867	
5% level	-2.861653	
10% level	-2.566872	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(W_M_)
 Method: Least Squares
 Date: 10/25/19 Time: 10:48
 Sample (adjusted): 10/03/2014 18:00 9/30/2016 18:00
 Included observations: 9439 after adjustments

Para el test de Dick and Fulley la hipótesis nula H_0 es que la serie no se comporta como una serie estacionaria, en tanto la hipótesis alterna H_1 es que la serie se comporta como una serie estacionaria. En este caso, los resultados de la prueba arrojan un valor p de 0.0000, por tanto, se puede rechazar la hipótesis nula y se asume que la serie se comporta como una serie estacionaria.

Los resultados de las pruebas de la prueba de Dick and Fulley para los otros cuatro emplazamientos estudiados se muestran en el Anexo A.

7.4.1.2 Análisis de la función de autocorrelación y de autocorrelación parcial

El siguiente paso es el análisis de la autocorrelación y la autocorrelación parcial mediante el estudio de sus correlogramas. La función de autocorrelación define como el valor actual de un proceso estocástico (y_t) es influenciado por los valores pasados de la serie (y_{t-k}); el correlograma es un gráfico que muestra la autocorrelación en función del rezago (k). Por otra parte, la función de autocorrelación parcial mide la correlación entre dos observaciones separadas por un tiempo k después de descontar el efecto de los demás términos de desfase vecinos ($y_{t-1}, y_{t-2}, \dots, y_{t-k}$) [66]. Las funciones de autocorrelación y autocorrelación parcial se muestran en las Figuras 13 y 14.

Figura 13. Función de autocorrelación – CDAR

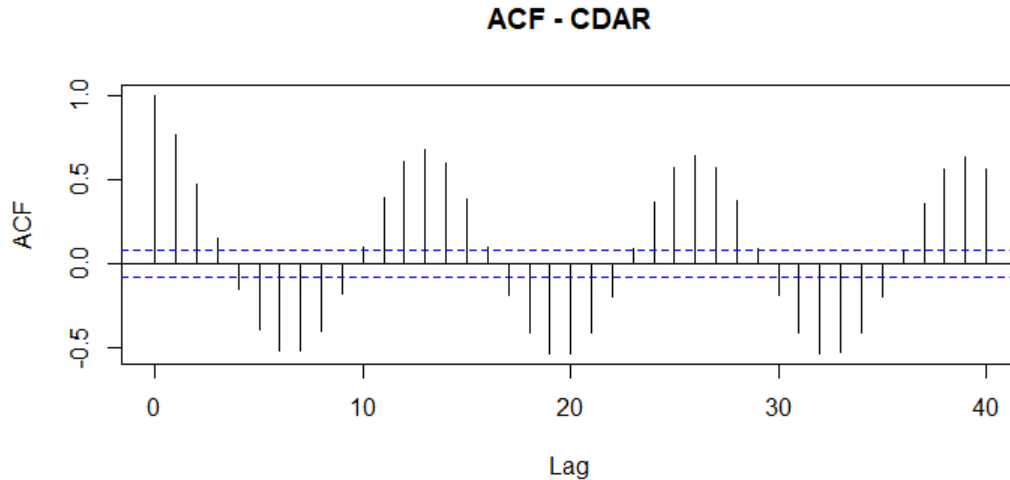
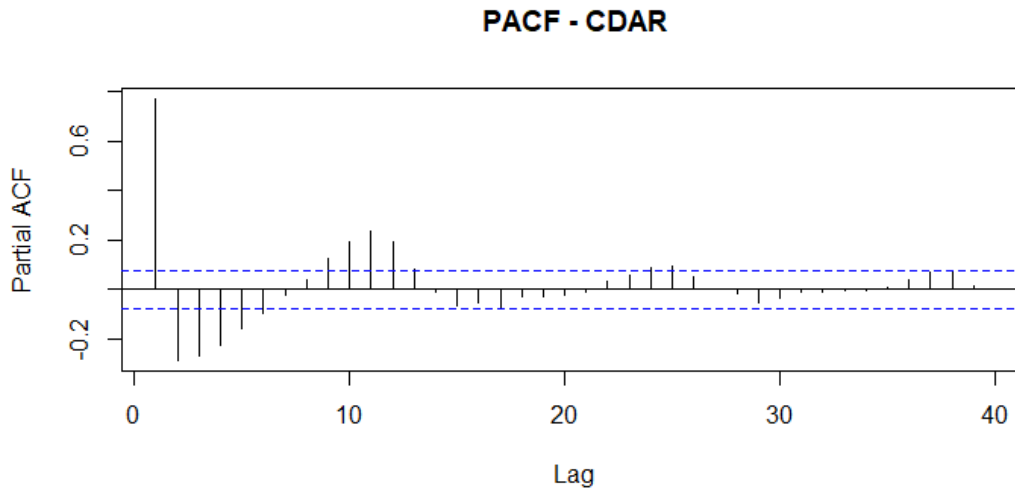


Figura 14. Función de autocorrelación parcial – CDAR



El comportamiento observado para la función de autocorrelación muestra un patrón de oscilación que se repite cada 13 observaciones (se tienen 13 registros de irradiación por día). Esto nos indica la presencia de estacionalidad, por tanto, se procede a diferenciar la serie cada 13 rezagos. A continuación, se muestran los correlogramas para la autocorrelación y la autocorrelación parcial para la serie después de diferenciarla cada 13 rezagos en las Figuras 15 y 16.

Figura 15. Función de autocorrelación diferenciada cada 13 rezagos – CDAR
ACF - CDAR diferenciada

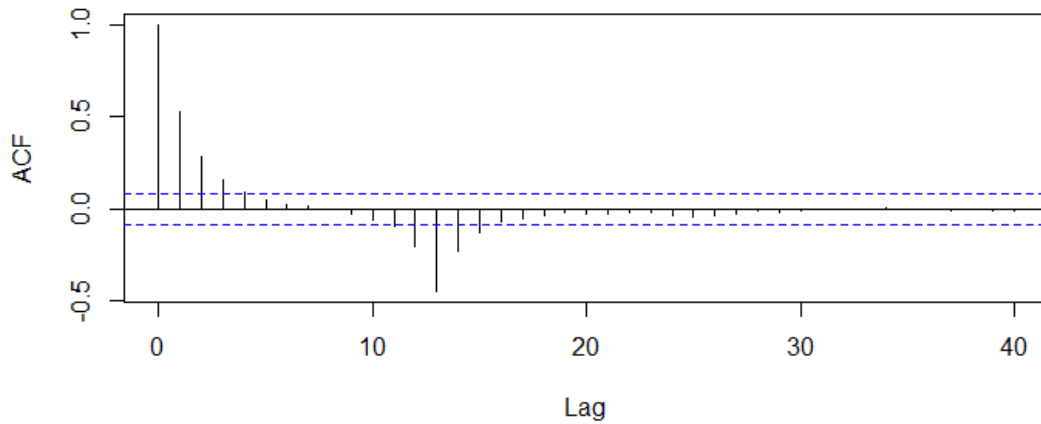
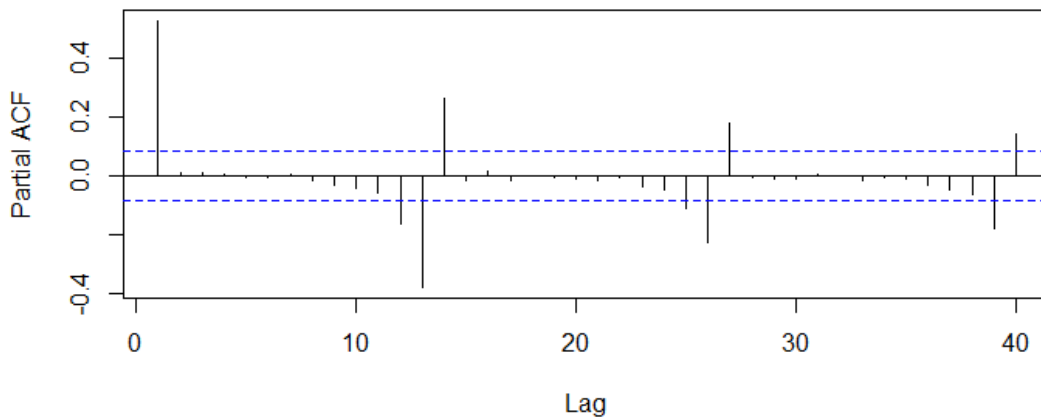


Figura 16. Función de autocorrelación parcial diferenciada cada 13 rezagos – CDAR
PACF - CDAR diferenciada



No es sencillo determinar los parámetros de un modelo SARIMA solo a través del análisis de los correlogramas de autocorrelación y autocorrelación parcial de la serie; con el análisis de los gráficos se pueden seleccionar posibles valores para los términos, pero es necesario ensayar varios modelos a través de un proceso iterativo para seleccionar el que mejor ajuste los datos. Para la selección de cualquier modelo de series de tiempo se debe tener en cuenta el principio de parsimonia, que consiste en elegir el modelo más simple posible; para un modelo SARIMA es bueno restringirse a que la suma de los parámetros cumpla que [67]:

$$p + d + q + P + D + Q \leq 6 \quad (9)$$

En este caso, en la función de autocorrelación (Figura 18) podemos observar dos picos en los rezagos 1 y 2 que nos indican que el término de promedios móviles q podría ser 1 o 2, además se observa un pico en el rezago 13 que indica un término de promedios móviles estacional Q de 1. Con respecto a la autocorrelación parcial se observa un pico en el rezago 1 que se puede atribuir a un término autorregresivo p de 1, se observa también picos significativos en el rezago 13, lo que sugiere que el término autorregresivo estacional puede ser 1 o 2.

7.4.1.3 Análisis de la función de autocorrelación y de autocorrelación parcial

Se ensayaron diferentes modelos SARIMA (p, d, q, P, D, Q, s) con los siguientes valores para los órdenes del modelo: $p=(1,2)$, $d=0$, $q = (1, 2)$, $P=(1,2)$, $D=1$, $Q=(1,2)$, $s=13$. Los modelos se compararon evaluando el estadístico Akaike Information Criteria (AIC). El AIC da una medida del ajuste relativo de un modelo estadístico para un conjunto de datos. En la selección de un modelo a través del AIC, se elige el que tenga el menor valor de este. Otro criterio de decisión para la selección de un modelo estadístico es la suma cuadrática de errores (SSE), como uno de los supuestos del modelo es que el error es aleatorio, el SSE del modelo deberá ser lo más pequeño posible.

Tabla 8. AIC - CDAR

p	d	q	P	D	Q	s	AIC	SSE
1	0	1	1	1	1	13	119086,5	160748230
1	0	1	1	1	2	13	119077,3	160529555
1	0	1	2	1	1	13	119080,9	160607070
1	0	1	2	1	2	13	119085,5	160657902
1	0	2	1	1	1	13	119088,3	160744309
1	0	2	1	1	2	13	119078,9	160516760
1	0	2	2	1	1	13	119082,6	160602579
2	0	1	1	1	1	13	119088,7	160751452
2	0	1	1	1	2	13	119079,2	160519126
2	0	1	2	1	1	13	119083,1	160609562
2	0	1	2	1	2	13	119088	160664803
2	0	2	1	1	1	13	119090,1	160741476

A partir del análisis del AIC se selecciona un modelo SARIMA (1,0,1,1,1,2)₁₃ por ser el que presenta el menor valor del AIC. Con respecto al SSE, el modelo que presenta menor SSE es el (1,0,2,1,1,2)₁₃, sin embargo, nos decidimos por el modelo (1,0,1,1,1,2)₁₃ porque para la selección del modelo pesa más el AIC.

7.4.2 Obtención de parámetros

Una vez se han seleccionado los órdenes del modelo a ajustar, se procede con el cálculo de los coeficientes del modelo. Este cálculo se hace a través de rutinas estadísticas que pueden incluir diferentes funciones objetivo. Una de las más comunes es la de mínimos cuadrados ordinarios. El algoritmo busca el vector de parámetros que mejor ajuste el valor de la función objetivo a través de un proceso iterativo.

En la Tabla 10 se muestran los coeficientes obtenidos para el modelo SARIMA (1,0,1,1,1,2)₁₃.

Tabla 9. Parámetros del modelo SARIMA (1,0,1,1,1,2)₁₃ - CDAR

	Estimate	SE	t.value	p.value
ar1	0.5985	0.0141	425.963	0.0000
ma1	-0.0064	0.0175	-0.3637	0.7161
sar1	0.5698	0.1165	48.923	0.0000
sma1	-1.4781	0.1219	-121.295	0.0000
sma2	0.4867	0.1187	41.011	0.0000
constant	0.0038	0.0056	0.6868	0.4922

Por lo tanto, el modelo se puede escribir como:

$$(1 + \theta_1 B^S)(1 + \theta_1 B)(1 - B^S)^D Y_t = (1 - \Phi_1 B^S + \Phi_2 B^{2S})(1 - \phi_1 B) Z_t \quad (12)$$

7.4.3 Diagnóstico del modelo

El último paso de la metodología de Box-Jenkins es el análisis de los residuales del modelo para garantizar que estos son generados por un proceso puramente aleatorio. Una de las pruebas realizadas durante el diagnóstico del modelo es el test de Ljung-Box, esta prueba tiene por objetivo demostrar que no queda ninguna correlación remanente en los

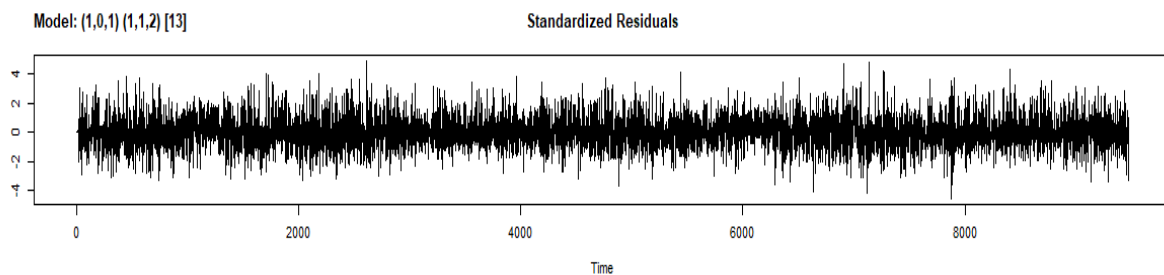
residuales, el estadístico de prueba utilizado es el valor p ; un valor p por encima del nivel de significancia nos indica que se puede rechazar la hipótesis nula y por tanto no hay autocorrelación en los residuales en este caso.

Además, se analiza el correlograma de la autocorrelación de los residuales para confirmar que no quede ninguna correlación remanente y también se grafica la serie de los residuales para revisar que estos poseen un comportamiento completamente aleatorio.

7.4.3.1 Comportamiento de los residuales

La serie de tiempo de los residuales se muestra en la Figura 17.

Figura 17. Residuales del modelo SARIMA $(1,0,1,1,1,2)_{13}$ – CDAR

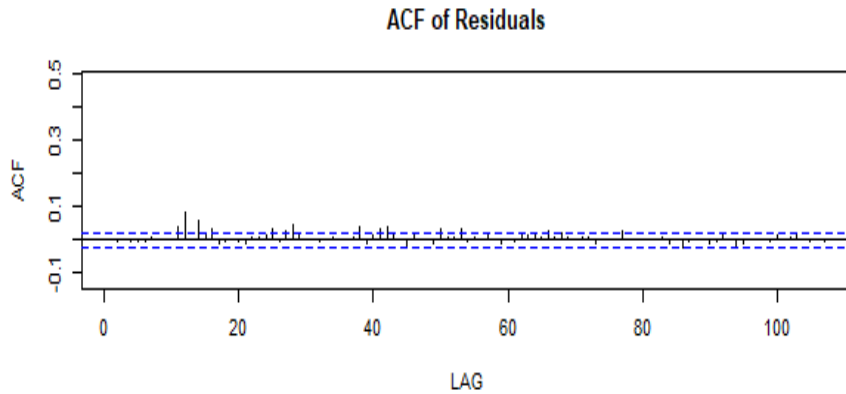


En la Figura 17 se observa que los residuales tienen un comportamiento aleatorio, la media y la varianza también parecen constantes en el tiempo, por lo que se puede afirmar que los residuales obedecen a ruido blanco.

7.4.3.2 Correlación de los residuales

A continuación, se analiza la función de autocorrelación de los residuales para garantizar que no queda ninguna correlación remanente en los rezagos. En la Figura 18 se muestra el correlograma de la autocorrelación de los residuales.

Figura 18. Autocorrelación de los residuales – CDAR

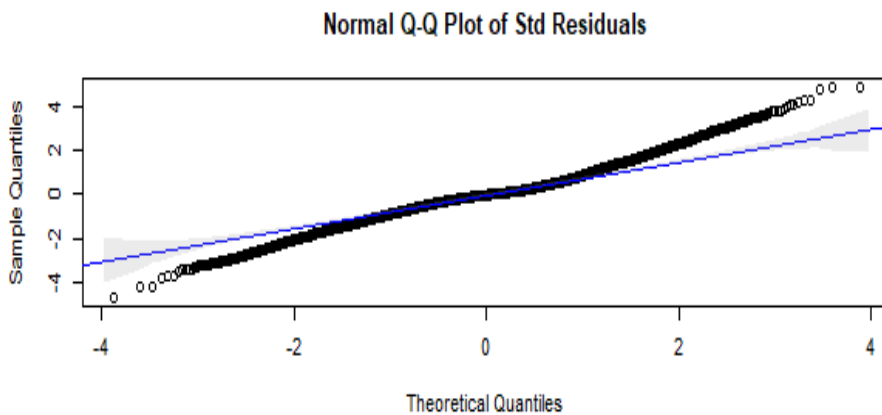


En la Figura 18, se observa un pico pequeño alrededor del rezago 13, no obstante, podemos ignorarlo y asumir que se trata de un fenómeno aleatorio. Por tanto, se concluye que no queda ninguna correlación remanente en la serie.

7.4.3.3 Gráfica cuantil-cuantil

Otra manera de representar los residuales es a través de una gráfica cuantil-cuantil, este es un método gráfico utilizado para comparar dos distribuciones de probabilidad graficando sus cuantiles juntos. En este caso, se grafican los cuantiles de los residuales del modelo junto con los de una distribución normal, para verificar que estos tienen un comportamiento normal se revisa que se ajusten a la recta mostrada en el gráfico.

Figura 19. Gráfica cuantil-cuantil de los residuales – CDAR



En la Figura 19 observamos que los datos se ajustan bien a una distribución normal, a pesar de que en los extremos se desvía un poco, en torno a los primeros cuartiles teóricos el ajuste es muy bueno.

7.4.3.4 Prueba de Ljung-Box

Finalmente, otra manera de comprobar que no existe ninguna correlación en los residuales es a través del test de Ljung-Box. En la prueba de Ljung-Box se prueba si un grupo de correlaciones para una serie de tiempo es diferente de cero. En lugar de probar la aleatoriedad para cada rezago por separado, esta prueba una aleatoriedad global con base en un número de rezagos dado.

Los resultados del test de Ljung-Box se muestran en la Tabla 10. Un valor sugerido para el número de rezagos en el cálculo del estadístico es el logaritmo en base 10 del número total de datos, en este caso ese número de rezagos es de 4.

Tabla 10. Ljung-Box test – CDAR

p-value	#rezagos
0.9808	4

En la prueba de Ljung-Box la hipótesis nula es que no existe correlación en los residuales de la serie. En este caso, los resultados del test indican que no se puede rechazar la hipótesis nula, ya que el valor p está muy por encima del nivel de significancia. Por tanto, no hay correlación remanente en los residuales.

7.5 Modelo de redes neuronales

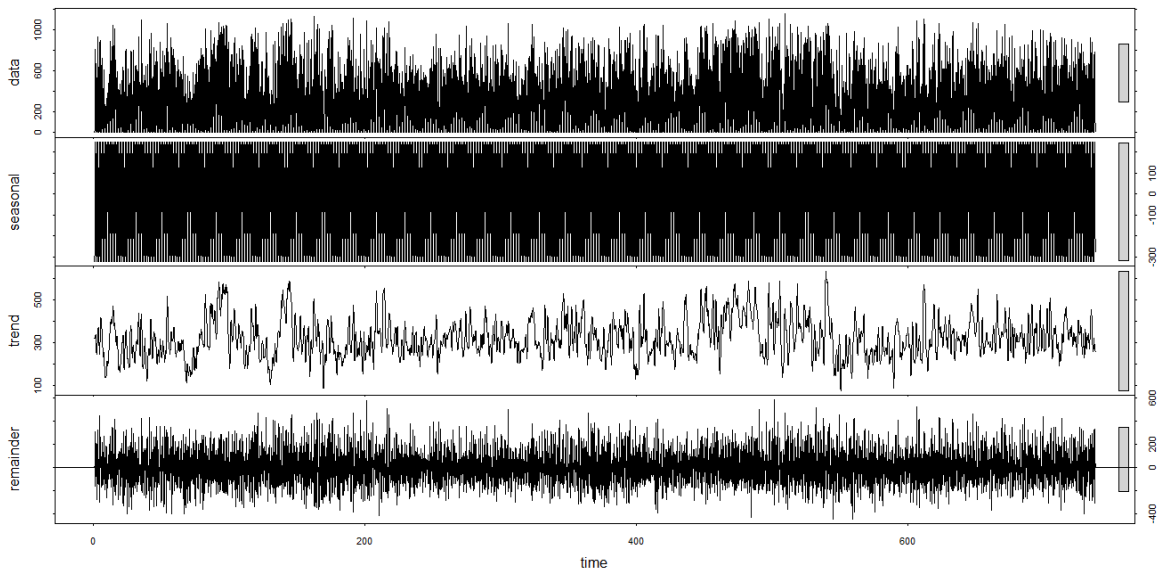
El modelo de redes neuronales se ajustó mediante el software estadístico R-studio. Se utilizó la función nnetar que ajusta un modelo no lineal autorregresivo (ARNN) con una función de activación logística, la función nnetar es comúnmente aplicado en el pronóstico de series de tiempo univariadas. La función programa una red neuronal pre-alimentada de una sola capa escondida, que tiene como entradas más importantes: p que indica los rezagos del modelo auto-regresivo no estacional, P que da los rezagos del modelo autorregresivo estacional, y k que representa el número de neuronas o nodos en la capa

escondida. El modelo ARNN(p,P,k)s es análogo al modelo SARIMA((p,0,0) (P,0,0)s) solo que a diferencia del modelo SARIMA, este incorpora términos no lineales en la correlación. La función nnetar prueba un número de redes idénticas, inicializando los pesos de manera aleatoria para seleccionar el modelo que mejor ajuste los datos.

7.5.1 Selección de los parámetros de rezagos del modelo

Para la selección del orden del modelo autorregresivo no estacional, cuando se trata de una serie de tiempo estacional primero se debe descomponer la serie [68]. Esto se hace a través de la función stl, la función stl descompone una serie de tiempo en tres partes: tendencia, estacionalidad y variación aleatoria. En la Figura 20 se muestra la descomposición de la serie de radiación.

Figura 20. Descomposición de la serie de radiación global promedio horaria



En la Figura 20 se muestra una fuerte tendencia estacional en la serie de tiempo, la cual ya fue analizada para el modelo SARIMA. Para descomponer la serie se requiere conocer con anticipación el periodo de la estacionalidad. En este caso, conocemos del análisis para el modelo de series de tiempo que la estacionalidad es de 13, por tanto, este argumento se ingresa a la serie de tiempo de irradiación, antes de realizar la descomposición.

Una vez se tiene la serie descompuesta, se procede con el cálculo del AIC de distintos modelos autorregresivos que ajusten la serie estacional descompuesta. Se selecciona el

orden del modelo que arroje el menor AIC. En la Tabla 11 se muestran algunos valores del AIC calculado para distintos órdenes de un modelo AR que ajuste la serie estacional descompuesta.

Tabla 11. Parámetros de rezago autorregresivo para el modelo ARRN

p	d	q	AIC	SSE
1	0	0	120612.0	164373584
2	0	0	120613.9	164371379
3	0	0	120615.8	164370389
4	0	0	120616.8	164362528
5	0	0	120615.4	163573682
...
12	0	0	120384.3	160139454
13	0	0	120370.9	159882892
14	0	0	120372.5	159875889
...
25	0	0	120355.6	159225473
26	0	0	120351.7	159128291
27	0	0	120351.7	159093071
...
38	0	0	120347.6	158653082
39	0	0	120347.2	158615440
40	0	0	120348.9	158615440

De acuerdo con los AIC para diferentes modelos autorregresivos para la serie estacional descompuesta, se observa que en general el estadístico de ajuste es mínimo cuando el número de rezagos es un múltiplo de 13 (13, 26, ...); es importante anotar que órdenes superiores de 39 y múltiplos de 13 para el modelo AR arrojan un menor AIC, sin embargo, el costo computacional de estos es muy alto en comparación con la mejora obtenida, por lo que se sugiere ensayar términos de 13 y 26 para la selección del modelo. El orden del modelo auto-regresivo estacional se seleccionó como 1 de manera análoga al modelo de series de tiempo expuesto antes.

7.5.2 Selección del número de nodos del modelo

El último parámetro para la selección del modelo es el número de nodos en la capa escondida de la red. Nuevamente se realiza un proceso iterativo ensayando distintos

números de nodos desde 1 hasta 20. Además, el proceso se realizó para los tres órdenes del modelo auto-regresivo estacional seleccionados arriba. Como el ajuste computacional del modelo se debe a un proceso puramente aleatorio, para la evaluación de los modelos no es posible usar estadísticos de ajuste como se hizo antes; por tanto, la evaluación se realizará con respecto a la precisión de los modelos ensayados para el pronóstico de un conjunto de datos de prueba. Se usó el periodo entre el 9 y el 15 de octubre de 2016 de los datos de la estación Centro de Alto Rendimiento Deportivo – CDAR. Estos mismos datos también fueron usados para las predicciones del modelo SARIMA que se exponen en la sección 8.6. En la Tabla 12 se muestra la evaluación del ajuste de los modelos en términos del error cuadrático medio RMSE.

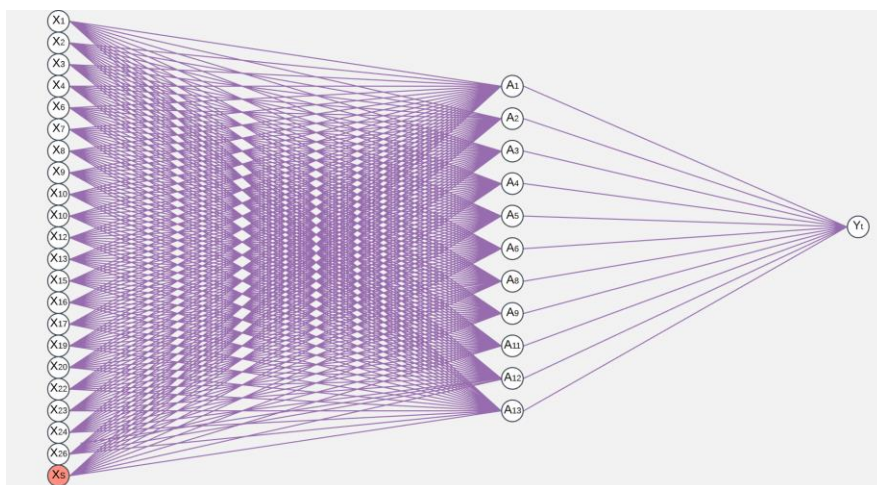
Tabla 12. Número de nodos en la capa escondida para el modelo ARNN

p	P	k	RMSE	p	P	k	RMSE
13	1	1	137,75	26	1	1	120,62
13	1	2	134,83	26	1	2	107,53
13	1	3	135,59	26	1	3	81,63
13	1	4	166,30	26	1	4	67,82
13	1	5	88,01	26	1	5	49,65
13	1	6	79,32	26	1	6	57,22
13	1	7	56,30	26	1	7	56,27
13	1	8	45,04	26	1	8	51,94
13	1	9	47,84	26	1	9	52,14
13	1	11	48,75	26	1	11	48,94
13	1	12	55,01	26	1	12	52,87
13	1	13	53,28	26	1	13	38,09
13	1	14	59,42	26	1	14	42,94
13	1	15	41,69	26	1	15	48,55
13	1	16	63,26	26	1	16	47,00
13	1	17	64,50	26	1	17	42,39
13	1	18	59,48	26	1	18	52,36
13	1	19	52,58	26	1	19	51,87
13	1	20	51,06	26	1	20	39,50

En total se analizaron 40 modelos con diferentes parámetros autorregresivo no estacional y nodos en la capa escondida. Del análisis del RMSE para el pronóstico del 9 al 15 de noviembre de 2016, se concluye que el modelo que mejor ajusta los datos es un modelo con 26 rezagos para el término autorregresivo no estacional, 1 rezago para el término

autorregresivo estacional, 13 nodos en la capa escondida y un periodo de 13 rezagos, es decir un modelo ARNN (26,1,17)13. Un esquema del modelo autorregresivo de redes neuronales implementado se muestra en la Figura 21.

Figura 21. Visualización de la ARNN



7.6 Generación de predicciones

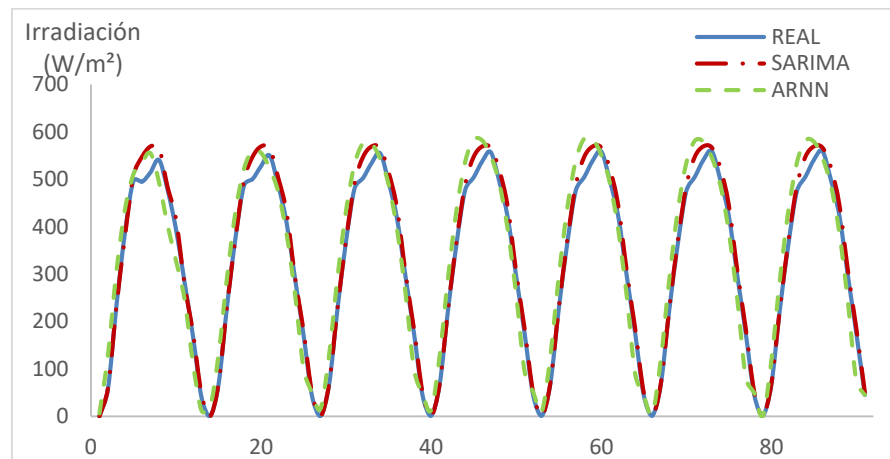
Una vez ajustados por completo los modelos de predicción, se procedió a hacer un pronóstico por un periodo de 7 días para cada uno de los emplazamientos de estudio. Los mismos modelos de series de tiempo y redes neuronales ajustados para los registros del Centro de Alto Rendimiento Deportivo – CDAR, fueron utilizados para generar las predicciones en los otros cuatro emplazamientos. Los modelos fueron re-entrenados con los datos de cada estación.

Las predicciones se realizaron según los periodos de pronóstico obtenidos del análisis de la disponibilidad de las estaciones meteorológicas descrito en la sección 8.2. En la Figura 23 se muestra el pronóstico para la segunda semana del mes de octubre de 2016 en la estación del Centro de Alto Rendimiento Deportivo - CDAR.

Este periodo se seleccionó porque los primeros días del mes de octubre hacen parte de la segunda temporada seca del año (agosto a octubre). En la Figura 22 se observa que bajo condiciones principalmente soleadas. El pronóstico generado se ajusta muy bien al

comportamiento observado de la irradiación global promedio horaria. Mostrando un error medio cuadrático porcentual (RMSE) de tan solo 19,5 W/m² (6.1%) para el modelo de series de tiempo y 50,4 W/m² (15,8%) para el modelo autorregresivo de redes neuronales, así mismo, el error medio absoluto (MAE) fue de 6,8 W/m² y 44,4 W/m² respectivamente (4.3% y 13.4%).

Figura 22. Pronóstico 8-15 de octubre de 2016 - CDAR

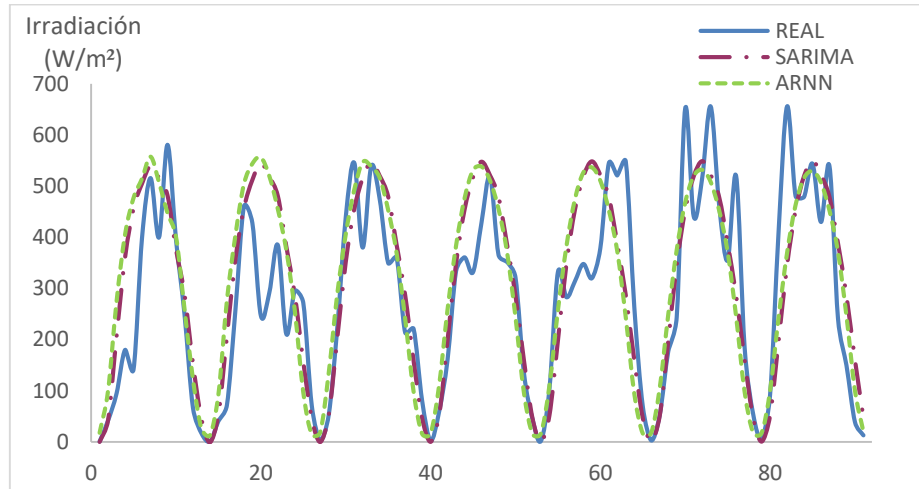


Por otra parte, cuando se presentan nubosidad y precipitaciones, la irradiación en superficie se ve disminuida abruptamente como consecuencia de la atenuación de la irradiación producida; como ya se mostró antes en el análisis de los registros, estas variaciones se presentan casi aleatoriamente ocasionando grandes diferencias en las curvas de radiación diaria. En la Figura 23 se observa el pronóstico para siete días durante la temporada de lluvias (18-24 de mayo de 2016).

En la Figura 24 se observa que los modelos ensayados sobre-estiman significativamente la irradiación solar global promedio horaria cuando hay presencia de nubosidad. En este caso, el RMSE fue de 112,4 W/m² (39.6%) para el modelo de series de tiempo y de 121,1 W/m² (46.5%) para el modelo autorregresivo de redes neuronales, el MAE fue de 81,8 W/m² y 103,7 W/m² respectivamente (28,7% y 32,4%). Recordemos que los modelos presentados en este trabajo son de tipo estadístico y univariados, basan su funcionamiento en la extrapolación de los valores futuros a partir de la serie de datos anteriores, por lo que les es imposible ajustarse a estas variaciones abruptas en la irradiación debido a otros factores externos. Otros modelos estadísticos de tipo inteligencia artificial han intentado

incorporar correlaciones con otras variables meteorológicas, pero su horizonte de pronóstico es de tan solo unas horas adelante.

Figura 23. Pronóstico 18-24 de mayo de 2016 - CDAR



Para las demás estaciones meteorológicas se definieron periodos de pronóstico teniendo cuidado de que estos coincidieran en la mayor cantidad posible de estaciones, para poder así comparar diferencias presentadas entre emplazamientos. Los periodos de pronóstico definidos para cada estación se muestran en la Tabla 13.

Tabla 13. Periodos de pronóstico

Estación	Periodo pronóstico
CDAR	18/5/2016-24/05/2016 8/10/2016-14/10/2016
San Cristóbal	18/5/2016-24/05/2016 8/10/2016-14/10/2016
Guaymaral	18/5/2016-24/05/2016 8/10/2016-14/10/2016
Tunal	18/5/2016-24/05/2016 22/12/2015-28/12/2015
Kennedy	22/12/2015-28/12/2015

Las gráficas de los pronósticos generados para las estaciones meteorológicas de San Cristóbal, Guaymaral, Tunal y Kennedy se muestran en los Anexos B, C, D y E.

7.7 Validación de predicciones

La validación de las predicciones se realizó a través de los estadísticos del error cuadrático medio (RMSE) y el error medio absoluto (MAE). El resumen de las predicciones generadas para las estaciones de Centro de Alto Rendimiento Deportivo, Guaymaral, San Cristóbal y Kennedy se muestran en las Tablas 14 y 15.

Tabla 14. Desempeño de los modelos de pronóstico temporada seca

MODELO DE SERIES DE TIEMPO					
Estación	Periodo	RMSE	RMSE%	MAE	MAE%
CDAR	09/10-16/10 2016	19,5 W/m ²	6,1%	13,8 W/m ²	4,3%
San Cristóbal	09/10-16/10 2016	185,0 W/m ²	47,4%	140,2 W/m ²	35,9%
Guaymaral	09/10-16/10 2016	98,0 W/m ²	31,0%	51,4 W/m ²	16,3%
Tunal	23/12-29/12 2016	103,2 W/m ²	25,1%	76,7 W/m ²	18,7%
Kennedy	23/12-29/12 2016	114,0 W/m ²	25,7%	86 W/m ²	19,4%
MODELO DE REDES NEURONALES AUTORREGRESIVO					
Estación	Periodo	RMSE	RMSE%	MAE	MAE%
CDAR	09/10-16/10 2016	49,8 W/m ²	16,0%	51,2 W/m ²	14,1%
San Cristóbal	09/10-16/10 2016	187,0 W/m ²	47,9%	135,0 W/m ²	34,6%
Guaymaral	09/10-16/10 2016	99,1 W/m ²	31,3%	52,6 W/m ²	16,6%
Tunal	23/12-29/12 2016	123,6 W/m ²	30,1%	91,4 W/m ²	22,2%
Kennedy	23/12-29/12 2016	141,4 W/m ²	31,9%	108 W/m ²	24,4%

En la Tabla 14 se observa que en general, los dos modelos de pronóstico implementados tienen un desempeño aceptable para condiciones predominantemente soleadas, se observa un ajuste muy bueno en términos del error medio absoluto porcentual (MAPE) para la estación del Centro de Alto Rendimiento Deportivo, el ajuste también es aceptable en las estaciones de Guaymaral, Kennedy y Tunal. No obstante, es más malo para la estación de San Cristóbal durante el mismo periodo. Esto puede deberse a las condiciones del emplazamiento con frecuentes precipitaciones y nubosidades bajas que afectan la radiación solar; como la estación de San Cristóbal está ubicada en los cerros orientales, la alta humedad relativa proveniente de los bosques unida a las altas velocidades del viento

puede causar fenómenos de micro-climas, esto ya fue discutido en el análisis de la variabilidad de la irradiación solar de las estaciones expuesto en la sección 8.3.3.

Por otra parte, los pronósticos para los cinco emplazamientos seleccionados en temporada de lluvias se muestran en la Tabla 15. En la Tabla 16 se observa que bajo condiciones de alta nubosidad y/o precipitaciones la precisión en el pronóstico de la irradiación global promedio horaria se ve afectada sensiblemente en comparación con el desempeño en tiempo seco. Esto debido a que durante la temporada de lluvias se presentan variaciones abruptas en la irradiación a lo largo del día. Estas variaciones se suceden de forma prácticamente aleatoria, por lo que los cambios son imposibles de capturar en el horizonte de pronóstico seleccionado de un día adelante, ya que se producen en escalas de tiempo de horas y son generados por fenómenos de clima tropical, en el cual rachas de vientos o cambios de presión atmosférica pueden cambiar el patrón del clima.

Tabla 15. Desempeño de los modelos de pronóstico temporada de lluvias

MODELO DE SERIES DE TIEMPO					
Estación	Periodo	RMSE	RMSE%	MAE	MAE%
CDAR	18/05-24/05 2016	112,4 W/m ²	39,6%	81,6 W/m ²	28,7%
San Cristóbal	18/05-24/05 2016	161,0 W/m ²	43,8%	117, W/m ²	31,9%
Guaymaral	18/05-24/05 2016	150,0 W/m ²	41,0%	102 W/m ²	28,0%
Tunal	18/05-24/05 2016	114,4 W/m ²	38,4%	82,7 W/m ²	28,5%
Kennedy	-	-	-	-	-
MODELO DE REDES NEURONALES AUTORREGRESIVO					
Estación	Periodo	RMSE	RMSE%	MAE	MAE%
CDAR	18/05-24/05 2016	121,0 W/m ²	42,7%	92 W/m ²	32,4%
San Cristóbal	18/05-24/05 2016	167,3 W/m ²	45,5%	119,8W/m ²	32,6%
Guaymaral	18/05-24/05 2016	157,6 W/m ²	43,2%	109 W/m ²	29,9%
Tunal	18/05-24/05 2016	117,2 W/m ²	38,4%	117,2 W/m ²	30,4%
Kennedy	-	-	-	-	-

A pesar de las dificultades expuestas, en conjunto, los modelos ensayados producen resultados aceptables para el pronóstico en los cinco emplazamientos con un error medio porcentual (MAPE) de 18,7% y 22,3% para los modelos SARIMA y ARNN respectivamente

en temporada seca, el promedio mejora hasta 14,4% y 19,0% si excluimos los datos de la estación San Cristóbal. Mientras que el MAPE fue de 29,3% y 31,3% para los mismos modelos en temporada de lluvias.

Finalmente, destaca el hecho que el modelo de redes neuronales implementado no consigue mejorar el desempeño del modelo de series de tiempo, a pesar de que la precisión es bastante cercana para algunos emplazamientos. Esto puede explicarse porque se implementó un modelo que a pesar de incluir componentes no lineales en la correlación, es meramente auto-regresivo, al no incluir términos de promedios móviles en comparación con el modelo de series de tiempo, es menos adecuado para manejar los componentes aleatorios de la serie de tiempo. Otro factor que podría influenciar es el hecho de que la selección del modelo se realizó para los datos del Centro de Alto Rendimiento Deportivo y como las técnicas de inteligencia artificial son altamente sensibles a la selección de los parámetros, el modelo implementado en el CDAR puede no ser el más apropiado para las demás estaciones.

8. Conclusiones

8.1 Conclusiones

- Se realizó una revisión bibliográfica de modelos de predicción de recurso solar con énfasis en los modelos de predicción a corto plazo. Se encontró que, para un horizonte de pronóstico del día siguiente, los modelos de predicción más adecuados son los modelos estadísticos de series de tiempo y las técnicas de inteligencia artificial.
- La latitud de los emplazamientos estudiados permite aproximar la duración del día como constante a lo largo del año. Por lo que los registros obtenidos de irradiación solar horaria evidencian un comportamiento estacional de las series de tiempo. Las transformaciones aplicadas a los datos han permitido obtener una serie temporal estacionaria luego de aplicar una diferenciación cada 13 rezagos.
- La irradiación solar global promedio horaria en superficie puede predecirse aceptablemente en un horizonte de pronóstico de 24 horas a través de un modelo de series de tiempo SARIMA (1,0,1,1,2)₁₃ y un modelo de redes neuronales autorregresivo con una capa escondida con 13 neuronas en la capa escondida y 26 nodos de entrada para componentes autorregresivos y 1 nodo autorregresivo estacional.
- El error en el pronóstico evaluado a través del error medio absoluto porcentual (MAE) fue de 14% en promedio en todos los emplazamientos para predicciones durante tiempo seco (excluyendo la estación San Cristóbal) y de 29% para predicciones en temporada de lluvias con el modelo de series de tiempo SARIMA, para el modelo de redes neuronales autorregresivo ARNN se obtuvo un desempeño similar de 19% para temporada seca (nuevamente excluyendo la estación de San Cristóbal) y de 31,3% para temporada de lluvias.

- Las diferencias evidenciadas para las predicciones entre el modelo de series de tiempo y el modelo de redes neuronales autorregresivo se pueden explicar porque el modelo de redes neuronales a diferencia del modelo de series de tiempo no incorporó términos de promedios móviles. Los modelos puramente autorregresivos tienen problemas manejar los componentes aleatorios de la serie. Las técnicas de inteligencia artificial también son muy sensibles a la selección de los parámetros y en la selección del modelo solo se tuvo en cuenta los registros de la estación del Centro de Alto Rendimiento Deportivo.
- Se encontraron diferencias significativas entre emplazamientos y épocas del año de la predicción, estas se atribuyen al clima tropical predominante en la zona de la sabana de Bogotá, en el que se presentan cambios abruptos en la nubosidad debido a caídas de temperatura y presencia de lluvias, estos cambios son imposibles de manejar con los modelos y el horizonte de pronóstico seleccionados.
- La calidad de los datos de irradiación solar y la continuidad de los registros es fundamental para obtener buenos resultados de predicción. Esta afirmación se basa en el hecho de que previamente se obtuvieron datos meteorológicos de otras entidades: Instituto de Hidrología y Estudios Ambientales – IDEAM, Corporación Autónoma Regional de Cundinamarca – CAR y datos propios del grupo de máquinas y motores eléctricos (EM&D) de la Universidad Nacional de Colombia, no obstante, la baja disponibilidad de los datos impidió la selección de un modelo que ofreciera un buen coeficiente de ajuste.

8.2 Recomendaciones

- Se encontró que la precisión del pronóstico es altamente influenciada por factores externos como la nubosidad, la cual domina el clima en la sabana de Bogotá, observando grandes diferencias en los pronósticos entre la temporada seca y la temporada de lluvia. Se recomienda probar una metodología de predicción directamente sobre la nubosidad, como por ejemplo el uso de técnicas de imágenes cielo-tierra o metodologías híbridas de imágenes satelitales.

- Otra alternativa podría ser el ajuste de una red neuronal multivariable con operaciones de lógica difusa que permita generar pronósticos en un horizonte de más corto plazo, en los que se correlacione la irradiación global en superficie no solo con los valores anteriores de esta, sino que permita establecer relaciones con cambios de tendencia en las otras variables meteorológicas: i.e. el descenso abrupto de la temperatura está directamente relacionado con la presencia de altas nubosidades y mayor posibilidad de precipitación que afectan la irradiación en superficie.

A. Prueba de raíz unitaria

Tabla 16. Test de Dick and Fulley (San Cristóbal)

Augmented Dickey-Fuller Unit Root Test on IRRADIACION_W_M_		
Null Hypothesis: IRRADIACION_W_M_ has a unit root		
Exogenous: Constant		
Lag Length: 41 (Automatic - based on SIC, maxlag=46)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-17.72983	0.0000
Test critical values:	1% level	-3.430470
	5% level	-2.861477
	10% level	-2.566777
*MacKinnon (1996) one-sided p-values.		
Augmented Dickey-Fuller Test Equation		
Dependent Variable: D(IRRADIACION_W_M_)		
Method: Least Squares		
Date: 10/26/19 Time: 16:36		
Sample (adjusted): 9/04/2014 09:00 6/30/2019 18:00		
Included observations: 22474 after adjustments		

Tabla 17. Test de Dick and Fulley (Guaymaral)

Augmented Dickey-Fuller Unit Root Test on IRRADIACION_W_M_		
Null Hypothesis: IRRADIACION_W_M_ has a unit root		
Exogenous: Constant		
Lag Length: 11 (Automatic - based on SIC, maxlag=16)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.711813	0.0001
Test critical values:	1% level	-3.447441
	5% level	-2.868968
	10% level	-2.570794
*MacKinnon (1996) one-sided p-values.		
Augmented Dickey-Fuller Test Equation		
Dependent Variable: D(IRRADIACION_W_M_)		
Method: Least Squares		
Date: 10/26/19 Time: 16:39		
Sample (adjusted): 9/01/2016 18:00 9/30/2016 18:00		
Included observations: 378 after adjustments		

Tabla 18. Test de Dick and Fulley (Kennedy)

Augmented Dickey-Fuller Unit Root Test on W_M_		
Null Hypothesis: W_M_ has a unit root		
Exogenous: Constant		
Lag Length: 41 (Automatic - based on SIC, maxlag=42)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-14.34808	0.0000
Test critical values:		
1% level	-3.430606	
5% level	-2.861537	
10% level	-2.566810	
*Mackinnon (1996) one-sided p-values.		
Augmented Dickey-Fuller Test Equation		
Dependent Variable: D(W_M_)		
Method: Least Squares		
Date: 10/26/19 Time: 16:35		
Sample (adjusted): 9/04/2012 09:00 11/30/2015 18:00		
Included observations: 15259 after adjustments		

Tabla 19. Test de Dick and Fulley (Tunal)

Augmented Dickey-Fuller Unit Root Test on IRRADIACION_W_M_		
Null Hypothesis: IRRADIACION_W_M_ has a unit root		
Exogenous: Constant		
Lag Length: 38 (Automatic - based on SIC, maxlag=42)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-13.47362	0.0000
Test critical values:		
1% level	-3.430583	
5% level	-2.861527	
10% level	-2.566804	
*Mackinnon (1996) one-sided p-values.		
Augmented Dickey-Fuller Test Equation		
Dependent Variable: D(IRRADIACION_W_M_)		
Method: Least Squares		
Date: 10/26/19 Time: 16:43		
Sample (adjusted): 2/04/2014 06:00 6/30/2017 18:00		
Included observations: 16133 after adjustments		

B. Pronósticos San Cristóbal, Guaymaral, Tunal y Kennedy

Figura 24. Pronóstico 9-15 de octubre de 2016 – San Cristóbal

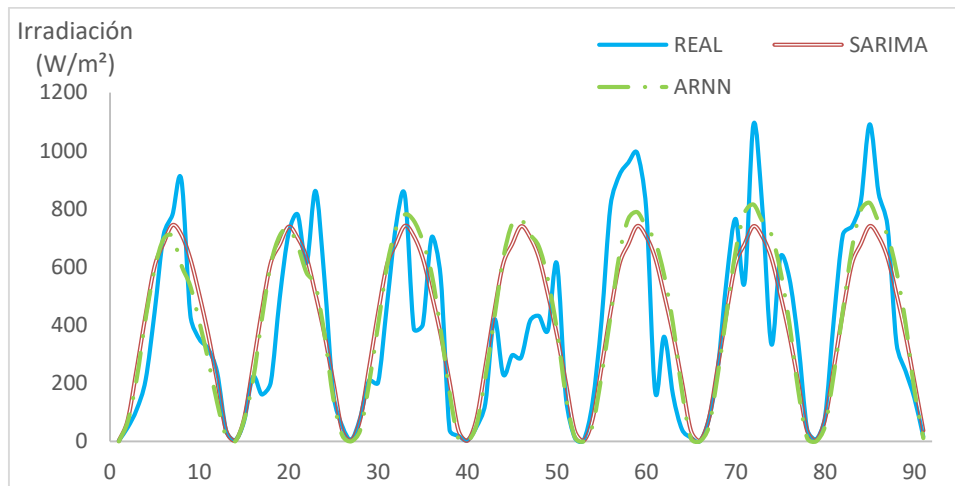


Figura 25. Pronóstico 18-24 de mayo de 2016 – San Cristóbal

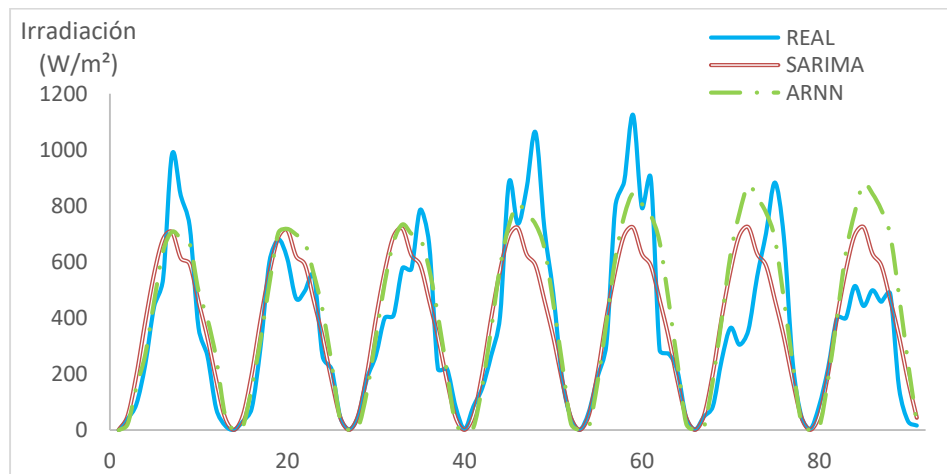


Figura 26. Pronóstico 9-15 de octubre de 2016 – Guaymaral

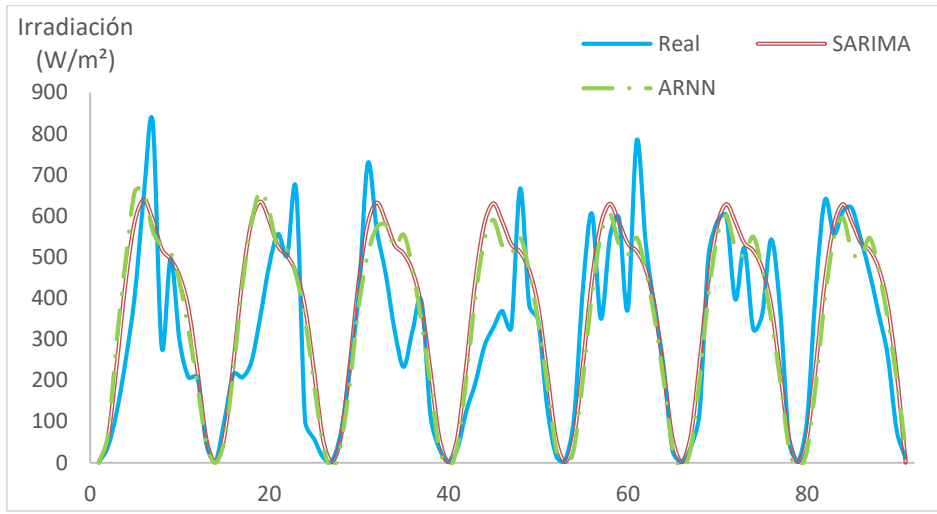


Figura 27. Pronóstico 9-15 de octubre de 2016 – Guaymaral

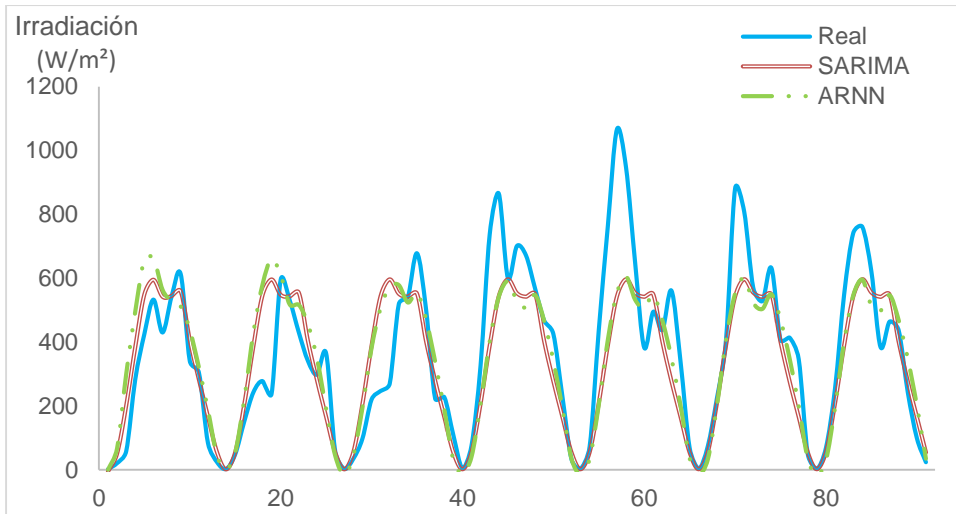


Figura 28. Pronóstico 22-28 de diciembre de 2015 – Tunal

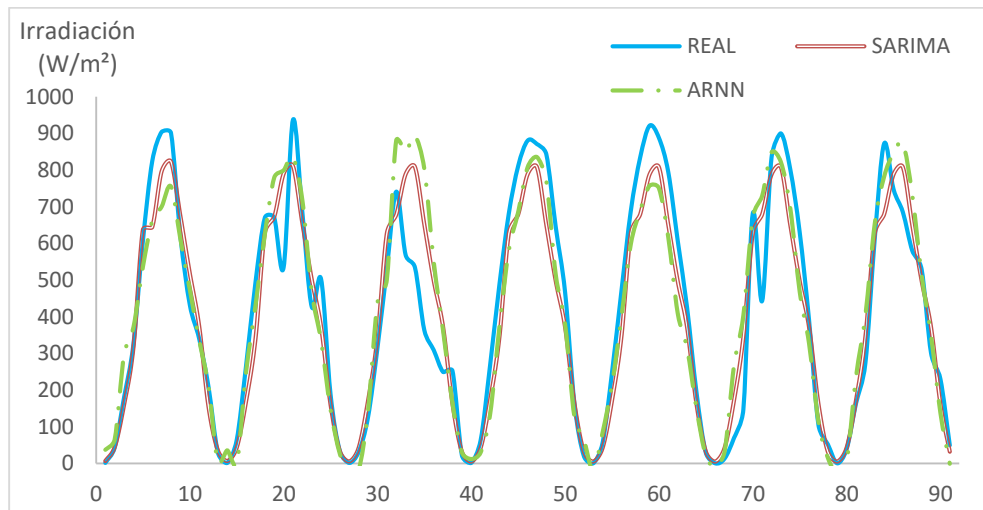


Figura 29. Pronóstico 22-28 de diciembre de 2015 – Tunal

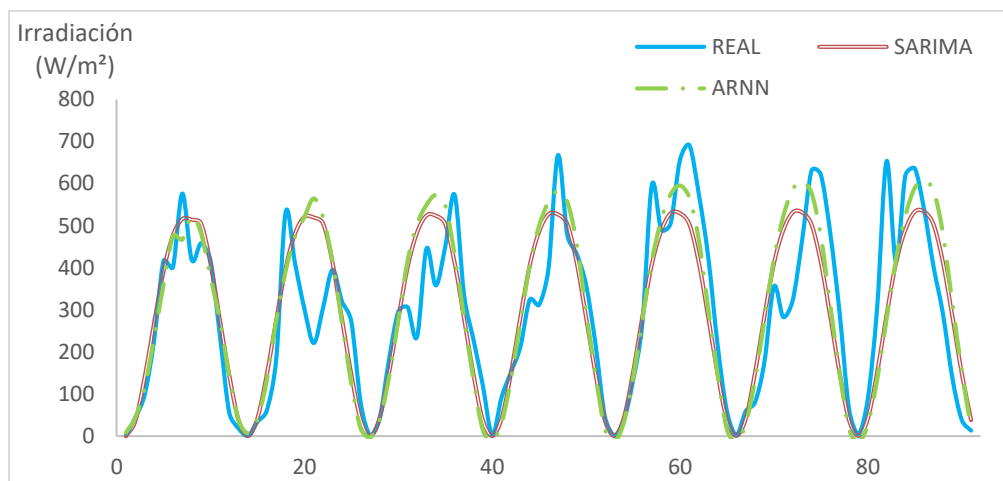
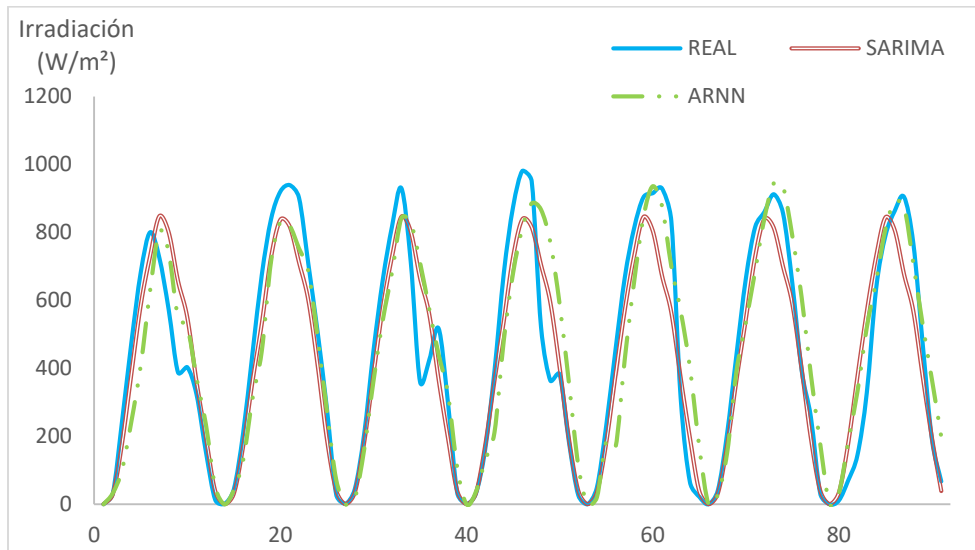


Figura 30. Pronóstico 22-28 de diciembre de 2015 – Kennedy



9. Bibliografía

- [1] World Wildlife Foundation, «The effects of climate change,» WWF, 2018. [En línea]. Available: <https://www.wwf.org.uk/impactsofclimatechange>. [Último acceso: 2018 28 03].
- [2] F. H. Cocks, Energy demand and climate change, issues and resolutions, Hong Kong: Wiley-VCH, 2009.
- [3] BP, «Statistical Review of World Energy 2017,» Workbook, London, 2017.
- [4] H. R. Murcia, «Desarrollo de la energía solar en Colombia y sus perspectivas,» *Revista de Ingeniería, Universidad de los Andes*, p. 9, 2008.
- [5] I. P. o. C. C. IPCC Special Reports Emission Scenarios, WMO UNEP, 2000.
- [6] Ministerio de Ambiente y Desarrollo Sostenible, «El ABC de los compromisos de Colombia para la COP21,» *WWF-Colombia*, p. 17, 2015.
- [7] E. García , A. Corredor , L. Calderón y M. Gomez, Análisis Costo Beneficio de Energías Renovables no Convencionales en Colombia, Bogotá: Fedesarrollo, 2013.
- [8] L. Gomez, Análisis de la viabilidad técnica y financiera de la generación de energía eléctrica a partir del recurso eólico en el municipio de Pereira, Medellín: Universidad EAFIT, 2015.
- [9] UPME, Unidad de Planeación Minero Energética, Integración de las energías renovables no convencionales en Colombia, Bogotá: Ministerio de Minas y Energía, 2015.
- [10] UPME, Plan de Expansión Eléctrica 2006-2020, Bogotá: Unidad de Planeación Minero Energética, 2006.

- [11] C. d. I. R. d. Colombia, Ley 143 de 1994, Bogotá: Congreso de la República de Colombia, 1994.
- [12] D. C. Corzo, ANÁLISIS DEL SECTOR DE ENERGÍA ELÉCTRICA COLOMBIANO Y REGIONAL, Bogotá: UNIVERSIDAD COLEGIO MAYOR DE NUESTRA SEÑORA DEL ROSARIO, 2013.
- [13] Comisión de Regulación de Energía y Gas, Resolución 140 de 2009, Bogotá: Ministerio de Minas y Energía, 2009.
- [14] B. Sorensen, Renewable Energy, London: Elsevier, 2004.
- [15] World Energy Council, World Energy Resources: Solar, New York: World Energy Council, 2013.
- [16] A. Goetzberger y V. U. Hoffmann, Photovoltaic solar energy generation, Springer, 2005.
- [17] S. Kalogirou, Solar energy engineering : processes and system, Burlington: Elsevier Inc, 2009.
- [18] F. Kreith y Y. Goswami, Handbook of Energy and Renewable Energy, CRC Press, 2007.
- [19] Unidad de Planeación Minero Energética - UPME, FORMULACIÓN DE UN PLAN DE DESARROLLO PARA LAS FUENTES NO CONVENCIONALES DE ENERGÍA EN COLOMBIA (PDFNCE), Bogotá: UPME, 2010.
- [20] S. Cortés y A. Arango Londoño, «Energías renovables en Colombia: una aproximación desde la economía,» *Revista Ciencias Estratégicas*, vol. 25, nº 38, pp. 375-390, 2017.
- [21] S. Hoyos, C. J. Franco y I. Dyner, «Integración de las fuentes no convencionales de energía renovable al mercado eléctrico y su impacto sobre el precio,» *Ingeniería y Ciencia*, vol. 13, nº 26, pp. 115-146, 2017.
- [22] O. Duarte y J. García Rendón, «Estimación del precio marginal del sistema eléctrico colombiano: una mirada desde la organización industrial,» *Ecos de economía*, vol. 19, nº 41, 2015.
- [23] O. Perpiñan Lamigueiro, Energía solar fotovoltaica, Creative Commons, 2018.

- [24] J. Kleissl, *Solar Energy Forecasting and Resource Assessment*, Waltham, USA: Elsevier, 2013.
- [25] J. Insunza, *Meteorología Descriptiva*, Concepción, Chile: Universidad de Concepción, Departamento de Física de la Atmósfera y del Océano, 2018.
- [26] M. Pidwirny, *Fundamentals of Physical Geography*, Okanagan: British Columbia Okanagan, 2014.
- [27] IDEAM, *Atlas de la Radiación Solar de Colombia*, Bogotá: Instituto de Hidrología, Meteorología y Estudios Ambientales, 2005.
- [28] D. Masa Bote, *Contribución a la integración de sistemas fotovoltaicos a la red eléctrica: recurso solar y predicción de generación*, Madrid: Escuela Técnica Superior de Ingenieros de Telecomunicación - Universidad Politécnica de Madrid, 2014.
- [29] D. I. Chaturvedi y Isha, «Solar Power Forecasting: A Review,» *International Journal of Computer Applications*, vol. 145, nº 6, pp. 28-50, 2016.
- [30] J. Du, M. Qilong, Z. Penglin, G. Jinhui, Y. Jun y Y. Bangshen, «Shor Term Solar Irradiance Forecasts Using Sky Images and Radiative Transfer Models,» *Energies*, vol. 11, nº 1107, 2018.
- [31] H. Seung Yang, K. Yeol Bae, H.-S. Park y D. Keun Sung, «Solar Power Prediction Based on Satellite Images and Support Vector Machine,» *IEEE Transactions on Sustainable Energy*, vol. 7, nº 3, pp. 1255-1263, 2016.
- [32] G. Kariniotakis, *Renewable Energy Forecasting*, Duxford, United Kingdom: Woodhead Publishing, 2017.
- [33] F. Lopes, H. Silva, R. Salgado, A. Cavaco, P. Canhoto y M. Collares-Pereira, «Short-term forecasts of GHI and DNI for solar energy systems operation: assessment of the ECMWF integrated forecasting system in southern Portugal,» *Solar Energy*, vol. 170, pp. 14-30, 2018, August.
- [34] H. Yang, B. Kurtz, A. Nguyen, C. W. Chow, M. Ghonima y J. Kleissl, «Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego,» *Energy Procedia*, vol. 49, pp. 2340-2350, 2014.
- [35] T. Tomkins Warner, *Numerical Weather and Climate Prediction*, Cambridge, England: Cambridge University Press, 2011.

- [36] J. Coiffier, *Fundamentals of Numerical Weather Prediction*, Cambridge, United Kingdom: Cambridge University Press, 2011.
- [37] T. Tomkins Warner, *Numerical Weather and Climate Prediction*, Cambridge, United Kingdom: Cambridge University Press, 2011.
- [38] N. O. a. A. Administratrion, «National Centers for Environmental Information,» NOAA, 2019. [En línea]. Available: <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/numerical-weather-prediction>. [Último acceso: 10th February 2019].
- [39] T. Schreiber, «Interdisciplinary application of nonlinear time series methods,» *Physics Reports*, vol. 308, pp. 1-64, 1999.
- [40] L. M. Pomares, *Análisis y predicción de series temporales de irradiancia solar global mediante modelos estadísticos*, Madrid: Universidad Complutense de Madrid, 2012.
- [41] R. Inman, H. T. C. y C. F. Coimbra, «Solar forecasting methods for renewable energy integration,» *Progress in Energy and Combustion Science*, vol. 39, pp. 535-576, 2013.
- [42] R. Perez, *Wind Field and Solar Radiation Characterization and Forecasting*, New York: Springer International Publishing, 2018.
- [43] M. P. Gonzalez Casimiro, *Análisis de series temporales - Modelos ARIMA*, Bilbao, España: Universidad del País Vasco, 2009.
- [44] R. R. Shumway y D. Stoffer, *Time Series Analysis and its Applications with R Examples*, 2017: SpringerLink, 1999.
- [45] A. R. Valdez Alvarado, *Introducción al machine learning*, La Paz, Bolivia: Facultad de Ciencias Puras y Naturales - Universidad Mayor de San Andrés, 2018.
- [46] P. Isasi Vifiuela y I. Galván León, *Redes de Neuronas Artificiales: Un Enfoque Práctico*, Madrid, España: Pearson Prentice Hall, 2004.
- [47] B. Martín del Brio y A. Sanz Molina, *Redes Neuronales y Sistemas Difusos*, Madrid, España: RA-MA, 2006.
- [48] M. Alsharif, M. Younes y J. Kim, «Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation: The Case of Study of Seoul, Soth Korea,» *Symmetry*, vol. 11, nº 240, 2019.

- [49] V. Prema y K. Uma Rao, «Development of statistical time series models for solar power prediction,» *Renewable Energy*, vol. 83, pp. 100-109, 2014.
- [50] J. Rosero, Á. Zambrano y Ó. Duarte, «Implementation of an Energy Demand Forecasting Model under a Smart Grids Environment,» de *2018 IEEE PES Transmission & Distribution Conference and Exhibition - Latin America (T&D-LA)*, Lima, Perú, 2018.
- [51] E. G. Kardakos, M. C. Alexiadis, C. K. Vagropoulos, P. N. Simoglu, A. G. Biskas y A. G. Bakiris, «Application of Time Series and Artificial Neural Network Models in Short Term Forecasting of PV Generation,» *48th International Universities' Power Engineering Conference (UPEC)*, vol. Dublin, pp. 1-6, 2013.
- [52] S. Ghimire, C. D. Ravinesh, N. Downs y N. Raj, «Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia,» *Journal of Cleaner Production*, vol. 216, pp. 288-310, 2019.
- [53] T. Khatib, M. Azad, S. K. y M. M., «Solar Energy Prediction for Malaysia Using Artificial Neural Networks,» *International Journal of Photoenergy*, vol. Tomo 2012, 2012.
- [54] A. Yadav Pratap y L. Behera, «Solar Radiation forecasting using neural networks and Wavelet Transform,» *IFAC Proceedings Volumes*, vol. 47, nº 1, pp. 890-896, 2014.
- [55] H. K. Yadav, Y. Pal y M. M. Tripathi, «Short-Term PV Power Forecasting Using Adaptive Neuro-Fuzzy Inference System,» de *IEEE 8th Power India International Conference (PIICON)*, India, 2018.
- [56] M. Z. Hassan, K. M. E. Ali, A. Shawkat Ali y J. Kumar, «Forecasting Day-ahead Solar Radiation Using Machine Learning Approach,» *4th Asia-Pacific World Congress on Computer Science and Engineering*, pp. 252-258, 2017.
- [57] R. Righini, «EMPLEO DE IMÁGENES SATELITALES GOES 8 EN EL CÁLCULO DE LA IRRADIACIÓN SOLAR GLOBAL EN DISTINTAS ZONAS DE ARGENTINA,» *ASADES*, vol. 8, nº 2, pp. 11.01-11.06, 2004.
- [58] R. Perez, K. Moore, S. Wilcox, D. René y A. Zelenka, «Forecasting solar radiation – Preliminary evaluation of an approach based upon the national forecast database,» *Solar Energy*, vol. 81, nº 6, pp. 809-812, 2007.

- [59] M. Caldas y R. Alonso-Suárez , «Very short-term solar irradiance forecast using all-sky imaging and real-time irradiance measurements,» *Renewable Energy*, 2019.
- [60] A. H. Murphy, «Skill scores based on the mean square error and their relationships to the correlation coefficient,» *Monthly weather review*, vol. 116, nº 12, p. 2417–2424, 1988.
- [61] Red de Monitoreo de Calidad del Aire de Bogotá, «Secretaría distrital de Ambiente,» 2013. [En línea]. Available: <http://ambientebogota.gov.co/red-de-calidad-del-aire>. [Último acceso: 29 Septiembre 2019].
- [62] Red de Monitoreo de Calidad del Aire de Bogotá, «Estacions RMCAB,» Secretaría de Ambiente Distrital, 2013. [En línea]. Available: <http://ambientebogota.gov.co/estaciones-rmcab#>. [Último acceso: 2 Octubre 2019].
- [63] W. Alpha, «Box-and-Whisker Plot,» 2019. [En línea]. Available: <http://mathworld.wolfram.com/Box-and-WhiskerPlot.html>. [Último acceso: 4 octubre 2019].
- [64] G. E. P. Box y G. Jenkins, *Time Series Analysis: Forecasting and Control*, San Francisco CA: Holden-Day, 1976.
- [65] Q. M. S. (QMS), «Dick and Fulley Test,» *EViews* 11, 1994.
- [66] Minitab, «Interpret the partial autocorrelation function (PACF),» 29 10 2019. [En línea]. Available: <https://support.minitab.com/en-us/minitab/19/help-and-how-to/modeling-statistics/time-series/how-to/partial-autocorrelation/interpret-the-results/partial-autocorrelation-function-pacf/>. [Último acceso: 29 10 2019].
- [67] R. Hyndman y G. Athanasopoulos, *Forecasting: Principles and Practice*, Sidney: OTexts, 2018.
- [68] Aggarwal y C. Charu, *Neural Networks and Deep Learning*, Springer, 2018.