

UCUENCA

Universidad de Cuenca

Facultad de Ingeniería

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Modelamiento de patrones diarios de producción solar fotovoltaica orientada a la mejora de la eficiencia y precisión en las proyecciones energéticas: un enfoque estadístico


Trabajo de titulación previo a la obtención del título de Ingeniero en Electrónica y Telecomunicaciones

Autor:

Micaela Cristina Tobar Palacios

Director:

Danny Vinicio Ochoa Correa

ORCID:  0000-0001-5633-1480

Cuenca, Ecuador

2025-09-23

Resumen

Este trabajo de titulación desarrolla un modelo estadístico para caracterizar y predecir la producción de energía solar fotovoltaica (FV) considerando sus variaciones anuales y diarias a partir de los datos de tres grupos de paneles solares a lo largo de un año. Para modelar la variabilidad anual, se emplearon modelos autorregresivos de media móvil (ARMA) aplicados a datos de series temporales sin tendencia, permitiendo identificar patrones de dependencia temporal en la generación de energía. A nivel diario, el trabajo se enfocó en ajustar distribuciones beta a percentiles de producción fotovoltaica y al agrupamiento de perfiles diarios, lo que proporcionó una caracterización refinada de los patrones de generación horaria. Además, a partir de los modelos ARMA obtenidos, se generaron series sintéticas con propiedades estadísticas equivalentes a las series originales, facilitando la simulación y el análisis predictivo. Los métodos implementados en este estudio proporcionan una herramienta robusta para mejorar la precisión en las proyecciones de generación solar y optimizar la gestión de sistemas fotovoltaicos, contribuyendo así a la planificación y operación eficiente de la energía renovable.

Palabras clave del autor: producción fotovoltaica, modelo ARMA, percentiles, análisis estadístico



El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Cuenca ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por la propiedad intelectual y los derechos de autor.

Repositorio Institucional: <https://dspace.ucuenca.edu.ec/>

Abstract

This research develops a statistical model to characterize and predict photovoltaic (PV) solar energy production, considering its annual and daily variations based on data from three groups of solar panels over one year. To model annual variability, autoregressive moving average (ARMA) models were applied to detrended time series data, allowing the identification of temporal dependency patterns in energy generation. At the daily level, the work focused on fitting beta distributions to percentiles of photovoltaic production, and on clustering the daily profiles, which provided a refined characterization of hourly generation patterns. Furthermore, based on the obtained ARMA models, synthetic time series with statistical properties equivalent to the original series were generated, facilitating simulation and predictive analysis. The methods implemented in this study provide a robust tool to improve the accuracy of solar generation projections and optimize the management of photovoltaic systems, thereby contributing to the efficient planning and operation of renewable energy.

Author Keywords: photovoltaic production, ARMA model, percentiles, statistical analysis



The content of this work corresponds to the right of expression of the authors and does not compromise the institutional thinking of the University of Cuenca, nor does it release its responsibility before third parties. The authors assume responsibility for the intellectual property and copyrights.

Institutional Repository: <https://dspace.ucuenca.edu.ec/>

Índice de contenido

Introducción	10
1.1 Antecedentes y contexto	10
1.2 Objetivos	12
1.3 Panorama metodológico	13
1.4 Estructura del Trabajo de Titulación	14
Revisión de la literatura	15
Metodología	28
3.1 Recolección de datos	30
3.2 Preprocesamiento de datos	30
3.3 Análisis de variación anual	31
3.3.1 Analema	31
3.3.2 Transformación Box-Cox	33
3.3.3 Procesamiento para modelado ARMA	33
3.3.4 Generación de series sintéticas	36
3.4 Análisis de variación diaria	37
3.4.1 Cálculo de percentiles y ajuste de la función beta	37
3.5 Perfiles diarios	38
3.6 Factor de Planta	40
3.7 Conclusiones del capítulo	41
Resultados y Discusión	42
4.1 Descripción general	42
4.2 Ajuste estacional (Anual)	43
4.2.1 Modelo Lineal	43
4.2.2 Estandarización de los residuos	46
4.2.3 Modelo ARMA	47
4.2.4 Evaluación del Modelo	49
4.2.5 Generación de series simuladas	51
4.3 Variación Diaria	53
4.4 Perfiles diarios	58
4.5 Factor de Planta	60

UCUENCA

5

4.6 Conclusiones del capítulo..... 61

Conclusiones **63**

Referencias..... **66**

Índice de figuras

Fig. 1 Diagrama de flujo de la metodología	29
Fig. 2 Distribución de paneles	30
Fig. 3 Variación anual de amanecer y atardecer	32
Fig. 4 Diagrama de flujo de modelamiento ARMA.....	34
Fig. 5 Ejemplos de comportamiento de ACF y PACF en modelo AR y MA	36
Fig. 6 Acumulado diario	39
Fig. 7 Acumulados Estandarizados.....	40
Fig. 8 Valores diarios con modelo lineal	45
Fig. 9 Desviación estándar de residuales	47
Fig. 10 Funciones de autocorrelación y autocorrelación Parcial.....	48
Fig. 11 Evaluación de residuales modelo ARMA (a) SFV1. (b) SFV2. (c) SFV3.....	50
Fig. 12 Series Simuladas.....	52
Fig. 13 Ajuste de percentiles de variación horaria	54
Fig. 14 Percentiles de SFV1	55
Fig. 15 Percentiles de SFV2.....	56
Fig. 16 Percentiles de SFV3.....	57
Fig. 17 Dendrograma de agrupamiento según perfil diario	58
Fig. 18 Curvas de acumulados diarios promedio por grupo	59
Fig. 19 Perfiles de producción diaria promedio por grupo	60

Índice de tablas

Tabla I Estudios sobre rendimiento de plantas FV	22
Tabla II Estudios sobre predicción solar	24
Tabla III Estudios sobre métodos de imputación de datos	27
Tabla IV Horas de Salida y Puesta del Sol	32
Tabla V Comportamiento de ACF y PACF en modelo ARMA.....	35
Tabla VI Valores para Ecuación de Modelo Lineal	44
Tabla VII Modelo Lineal de la Desviación Estándar	46

Agradecimientos

Quisiera expresar mi más sincero agradecimiento a todas las personas que contribuyeron al desarrollo de este trabajo de titulación. En primer lugar, agradezco al Ing. Danny Ochoa Correa, cuyo asesoramiento y guía fueron fundamentales para la realización de este trabajo. Su orientación me permitió avanzar con claridad y precisión en cada etapa; su sólida visión investigativa me proporcionó un horizonte real y actualizado en el campo de las energías renovables, factores sin los cuales no hubiese llegado a este punto.

Asimismo, agradezco al director de carrera, el Ing. Alcides Araujo, por su respaldo y confianza en este proyecto. Su apoyo fue invaluable para la consecución de este trabajo.

Un reconocimiento especial a mi padre, quien desempeñó un papel clave en este estudio. Su apoyo, conocimientos y constante motivación fueron esenciales para llevar a cabo esta investigación. La visión y conocimientos que me transmitió fueron determinantes para dar forma a este estudio. Su capacidad para analizar problemas desde una perspectiva crítica y su enfoque metódico me ayudaron a definir la dirección de la investigación, interpretar los datos de manera rigurosa y superar los desafíos analíticos que surgieron en el proceso. Sin su apoyo intelectual y motivación constante, este trabajo no habría sido posible.

No puedo dejar de mencionar a mi familia, cuyo apoyo incondicional me acompañó en todo momento. A mi madre, mi hermana y mi prometido, gracias por su paciencia, comprensión y aliento a lo largo de este proceso. Finalmente, un agradecimiento especial a mi gata Freya, cuya compañía fue un recordatorio constante de felicidad en las horas de trabajo.

A todos ustedes, mi más profundo agradecimiento.

Lista de abreviaturas

ACF	Función de Autocorrelación (por sus siglas en inglés)
AIC	Criterio de Información de Akaike (por sus siglas en inglés)
ANFIS	Sistema de Inferencia Neurodifusa Adaptativa (por sus siglas en inglés)
ANN	Redes Neuronales Artificiales (por sus siglas en inglés)
AR	Autoregresivo (por sus siglas en inglés)
ARIMA	Autoregresivo Integrado de Medias móviles (por sus siglas en inglés)
ARMA	Autoregresivo de Medias móviles (por sus siglas en inglés)
BIC	Criterio de Información Bayesiano (por sus siglas en inglés)
FV	Fotovoltaica
GAN	Red Generativa Antagónica (por sus siglas en inglés)
GEI	Gases de Efecto Invernadero
GSR	Radiación Solar Global diaria
GW	Gigawatts
kNN	k vecinos cercanos (por sus siglas en inglés)
kW	Kilowatts
kWh	Kilowatts hora
MA	Medias móviles (por sus siglas en inglés)
NaN	No es un número (por sus siglas en inglés)
NN	Redes Neuronales (por sus siglas en inglés)
NOAA	Oficina Nacional de Administración Oceánica y Atmosférica (por sus siglas en inglés)
NWP	Predicción Numérica del Clima (por sus siglas en inglés)
PACF	Función de Autocorrelación parcial (por sus siglas en inglés)
PCA	Análisis de Componentes Principales (por sus siglas en inglés)
R	Lenguaje de Programación R
SARIMA	Autorregresivo Integrado de Medias móviles Estacional (por sus siglas en inglés)
SCADA	Control de Supervisión y Adquisición de Datos (por sus siglas en inglés)
SFV	Solar Fotovoltaica
SMCS	Simulación Secuencial de Monte Carlo (por sus siglas en inglés)
WGS84	Sistema Geodésico Mundial de 1984 (por sus siglas en inglés)

Introducción

1.1 Antecedentes y contexto

En las dos últimas décadas, ha aumentado la conciencia sobre la afección del clima global por los efectos ambientales negativos derivados del uso de la energía proveniente de combustibles de origen fósil para las actividades humanas, debido a las emisiones de gases de efecto invernadero (GEI) [1]. Esta situación ha impulsado una creciente conciencia sobre la importancia de la conservación ambiental y la mitigación de impactos que pueden lograrse con un modelo energético más sostenible [2]. La conciencia social sobre los problemas ambientales y sus posibles soluciones tendrá un efecto directo en la forma de afrontar los retos energéticos del futuro. Esto significará, sin duda, una reducción del consumo de energía proveniente del uso de combustibles fósiles; así como un aumento de la producción de energías renovables o generadas a partir de recursos naturales inagotables. Este nuevo modelo energético debe basarse en fuentes de energía sostenibles desde el punto de vista medioambiental y disponibles en cantidades suficientes que garanticen su disponibilidad a largo plazo [3]. Las energías renovables, especialmente la solar fotovoltaica, han adquirido un papel fundamental en este escenario, gracias a su capacidad para generar electricidad de manera limpia y sostenible, sin emisiones de gases de efecto invernadero ni contaminantes atmosféricos durante su fase de operación [4]. La energía solar ha sido considerada como un recurso ilimitado y gratuito de energía: además, es barata, abundante, naturalmente libre y no contaminante [5].

La energía solar fotovoltaica (FV) ha experimentado un crecimiento significativo en los últimos años, convirtiéndose en una opción cada vez más atractiva tanto a nivel técnico como económico [5], [6].

A nivel global, diversos países han apostado por el desarrollo de la energía solar fotovoltaica como parte de sus estrategias de transición energética hacia un futuro más sostenible [7]. Esta tendencia se ha visto reflejada en la implementación de políticas de incentivo, subsidios y programas de financiamiento destinados a promover la adopción de esta tecnología [8]. Los países en desarrollo están invirtiendo cada vez más en energía renovable, lo que ha llevado a un aumento de la generación de energía fotovoltaica en la última década; esperándose que los sistemas fotovoltaicos

se extiendan a nivel mundial [9]. Asimismo, la inversión en investigación y desarrollo ha contribuido a mejorar la eficiencia de los paneles solares en los procesos de captura, conversión y distribución, y reducir los costos de instalación, aumentando así su eficiencia [10] y competitividad en el mercado energético [11] en comparación con otras fuentes de energía alternativas como la eólica, la de biomasa y la hidroeléctrica a pequeña escala.

Entre otras tecnologías renovables, la conversión directa de la radiación solar en electricidad mediante tecnología FV ha mostrado una de las capacidades de crecimiento más notables a nivel mundial: en 2024, las instalaciones solares mundiales alcanzaron casi 600 GW, un impresionante aumento del 33 % con respecto al año anterior. La energía solar representó el 81 % de toda la nueva capacidad de energía renovable añadida a nivel mundial [12]. La energía solar se utiliza directa o indirectamente en un número sustancial de aplicaciones, como la generación de electricidad, el calentamiento/refrigeración de aire y agua, etc. [1]. Se espera que las energías renovables emergentes, la solar y la eólica, desempeñen un papel importante a la hora de abastecer al menos entre el 5 y el 10% de la demanda total de energía eléctrica a nivel mundial [2]. Para mantener el impulso global se requerirá abordar diversos desafíos específicos de cada región. Los mercados más avanzados deben centrarse en mejorar la flexibilidad de la red, adaptar los marcos políticos a las necesidades de las energías renovables variables, priorizar el almacenamiento en baterías y agilizar los procesos de permisos y conexión a la red. Para los países menos adelantados, la clave reside en reducir las brechas de inversión. En todas las regiones, desarrollar una fuerza laboral cualificada y establecer objetivos ambiciosos en materia de energía solar y almacenamiento, son tareas esenciales [12].

El entendimiento del fenómeno de incidencia de la radiación solar en la superficie terrestre es de fundamental importancia para comprender los procesos físicos que tienen lugar en el sistema tierra-atmósfera [13] y para la evaluación del rendimiento de las tecnologías solares establecidas, así como la estimación de la viabilidad de las instalaciones en el futuro [1]. Los pronósticos optimizados de la irradiancia son necesarios para predecir el posible rendimiento de una planta de generación eléctrica solar FV. Este tipo de estudios se basa en los patrones de irradiancia, con la visión de clasificar los días en función de la transición de nubes y la variabilidad de la irradiancia directa, puesto que la respuesta del sistema depende del tipo de día [14]. El análisis de estos patrones permite garantizar la eficiencia y el manejo de las centrales de generación eléctrica alimentadas por energía solar maximizando su rendimiento [6], [9], [14].

La previsión de la producción es una tarea fundamental para facilitar la penetración a gran escala de la generación fotovoltaica en la red eléctrica; resultando en menores costos y restricciones en la producción. La precisión de la previsión está limitada tanto por la naturaleza cambiante del clima como por la incertidumbre sobre la respuesta física de los sistemas fotovoltaicos a determinadas condiciones. La alta variabilidad de

la irradiancia solar proviene principalmente del movimiento de las nubes y su bloqueo estocástico de la luz solar [15].

La ausencia de mediciones durante la producción y el almacenamiento de energía es un problema común en la industria fotovoltaica. La producción puede estimarse generalmente por características similares de los inversores y/o por señales externas correlacionadas. La generación fotovoltaica depende de condiciones meteorológicas [9] por tanto, es necesario desarrollar relaciones empíricas como sustituto de la medición de la radiación solar [13] formulando modelos de previsión y estimación de estos datos [2]. Para ello, es habitual encontrar modelos autorregresivos, métodos no paramétricos o técnicas de agrupamiento [16]. El análisis estadístico de los registros históricos de producción fotovoltaica permite identificar tendencias y patrones temporales, facilitando así el desarrollo de modelos predictivos precisos [17]. Estos modelos mejoran las proyecciones de producción futura, optimizando la planificación del mantenimiento, la gestión de la carga de la red eléctrica y la integración con otras fuentes de energía.

En el caso específico de Ecuador, un país caracterizado por su abundante recurso solar, la energía solar fotovoltaica representa una oportunidad única para impulsar el desarrollo sostenible y reducir la huella ambiental del sector energético [18]. Pese a su enorme potencial, aún existen desafíos en términos de infraestructura, regulación y financiamiento que deben ser abordados para aprovechar plenamente el potencial de esta tecnología. Sin embargo, el creciente interés de diferentes sectores sociales y productivos, la voluntad política y marcos regulatorios favorables en el país sugieren un camino prometedor hacia la integración de la energía solar fotovoltaica en la matriz energética nacional [19].

Idealmente, la información sobre la radiación solar de cada ubicación específica debería adquirirse desde estaciones de medición que utilicen instrumentos de alta precisión. Sin embargo, mediciones fiables de radiación solar global son relativamente escasas en muchos lugares del mundo debido a la baja densidad de estaciones de registro; y al costo y mantenimiento de los dispositivos [1].

En un contexto de crisis energética en el Ecuador y una fuerte dependencia de fuentes como la hidroelectricidad, la diversificación de la matriz energética se vuelve crucial para garantizar la seguridad y la estabilidad del suministro eléctrico [20].

1.2 Objetivos

Con estos antecedentes, la pregunta central de esta investigación es hasta qué punto se puede modelar la captura de energía de los paneles solares según las variaciones horarias y estacionales, permitiendo proyecciones confiables y la simulación de series. Se postula que una caracterización detallada de la captura de energía revelará patrones analizables y proyectables. Específicamente, se explorará la modelación de series temporales con variaciones diarias y estacionales en la captura de energía

mediante un análisis exhaustivo de los datos de producción de energía de diferentes grupos de paneles FV, incluyendo la identificación de patrones diarios mediante percentiles, así como la modelación de la producción fotovoltaica anual.

Lo antes expuesto se especifica en los siguientes objetivos:

Objetivo General

Plantear un modelo predictivo de producción fotovoltaica que integre variaciones temporales a partir de datos históricos orientado a mejorar la precisión de las proyecciones de producción solar fotovoltaica.

Objetivos específicos

1. Caracterizar la producción fotovoltaica y su variabilidad temporal en función de los datos históricos.
2. Identificar patrones diarios o agrupaciones en la generación fotovoltaica, basados en las variaciones temporales, que influyen en la eficiencia operativa de los paneles.
3. Implementar un modelo de captura de energía solar que considere la distribución por grupos de paneles.
4. Generar un modelo para simulación de serie de tiempo tipo ARMA para la cantidad de energía capturada por grupos de paneles.
5. Utilizar técnicas de modelamiento estadístico para estimar valores en casos de pérdidas de datos, permitiendo así completar y validar los conjuntos de datos empleados en el análisis.

1.3 Panorama metodológico

Los registros reales de medición se obtendrán del sistema SCADA de la granja solar fotovoltaica del laboratorio de Micro-Red de la Facultad de Ingeniería de la Universidad de Cuenca, luego de un preprocesamiento de los datos alojados en las memorias de los servidores de los que dispone el laboratorio. Con estos registros reales de producción fotovoltaica, la validación de los datos disponibles es un primer paso indispensable debido a posibles datos faltantes o erróneos, lo que plantea un desafío para obtener un modelo preciso si no se aborda adecuadamente [21]. En primer lugar, se implementará una validación de los registros de medición asociados a cada agrupación de paneles de la granja solar, a través del diagrama de diferencial acumulado de masas. Para el análisis anual, se tomarán valores de radiación total diaria para cada grupo. Posteriormente, se realizará la remoción de estacionalidad por medio de un ajuste sinusoidal. La declinación causada por la pendiente del eje de rotación de la Tierra con respecto a la eclíptica alrededor del Sol y la rotación de la Tierra implican cambios periódicos adicionales de la radiación solar (analema). Estos procesos evalúan el componente periódico anual. El componente periódico diario se evalúa mediante la rotación de la Tierra. El estado de la atmósfera implica cambios tanto estocásticos como periódicos. La turbidez de la atmósfera y la cobertura de

nubes tienen, además un fuerte componente estocástico. Las sumas mensuales de radiación global se utilizan ampliamente para varios propósitos. Es una buena herramienta para calcular los almacenamientos estacionales o analizar la eficiencia (estacional, anual) de las instalaciones solares. Las sumas diarias de radiación son los valores promedio utilizados con más frecuencia. La varianza de esta variable debe considerarse en el diseño de almacenamiento de corta duración utilizado en sistemas fotovoltaicos [2].

Luego de esto, se estimarán los parámetros de un ajuste autorregresivo de medias móviles (ARMA), que permita generar series estocásticas de la producción fotovoltaica. Por otro lado, para el análisis diario se emplearán curvas de percentiles que resultarán en la comprensión estadística de la radiación a lo largo del día. Al estar la información limitada a un año, se emplearán criterios estadísticos como la raíz del error cuadrático medio (RMSE) para la evaluación del modelo. Para estos análisis se hará uso de herramientas informáticas tales como Excel y el lenguaje de programación R, dentro del IDE RStudio; estas últimas, herramientas de acceso abierto.

1.4 Estructura del Trabajo de Titulación

Este trabajo de titulación está estructurado en cinco capítulos, cada uno de los cuales se basa en el anterior para abordar el análisis y modelado de los datos de radiación solar y producción fotovoltaica. El capítulo 2 presenta una revisión de la literatura donde se explora la investigación existente sobre el modelado de la radiación solar, el rendimiento del sistema fotovoltaico y las técnicas estadísticas para el análisis de series de tiempo y perfiles diarios. Destaca los nichos de investigación en la literatura actual y posiciona esta investigación dentro del contexto académico más amplio. El capítulo de metodología detalla el enfoque analítico empleado en este estudio. Describe el proceso de recopilación de datos, las técnicas de eliminación de tendencias de series de tiempo, el modelado ARMA para el análisis de residuos y el ajuste de funciones beta a las curvas diarias de producción FV. Además, explica cómo se optimizaron y consolidaron los parámetros de las funciones beta. El capítulo 4 presenta los resultados de los análisis estadísticos y los esfuerzos de modelado. Incluye los modelos ARMA derivados, las series temporales sintetizadas y las funciones beta ajustadas. El capítulo también analiza las implicaciones de los hallazgos, comparando el desempeño de los modelos y su aplicabilidad en escenarios del mundo real. Por último, el capítulo final resume los hallazgos clave y su importancia. Analiza las limitaciones del estudio y ofrece recomendaciones para futuras investigaciones y aplicaciones prácticas en la optimización de sistemas solares fotovoltaicos y la planificación energética.

Revisión de la literatura

El análisis del estado del arte permite situar esta investigación en el contexto de los estudios sobre energía solar fotovoltaica, estableciendo su fundamentación teórica y metodológica. Se examinan los principales enfoques para el modelado y análisis de la producción fotovoltaica, con énfasis en metodologías estadísticas y técnicas de pronóstico aplicadas a series temporales.

En esta sección, se analizan estudios previos que han abordado la caracterización de la variabilidad diaria y estacional de la generación FV, así como la integración de modelos estadísticos específicos para distintas ubicaciones. Además, se discuten los desafíos en la aplicación de estos modelos a datos reales, destacando la necesidad de enfoques más robustos que mejoren la precisión de las proyecciones y optimicen la gestión de la energía solar. Esta revisión proporciona el marco conceptual y metodológico sobre el cual se desarrolla el presente estudio.

Un recurso principal en el campo de la energía solar es el libro "Modeling Solar Radiation at the Earth's Surface: Recent Advances," editado por Viorel Badescu [22]. Este libro es un compendio de las últimas metodologías y desarrollos en el modelado de la radiación solar y sus aplicaciones, y sirve como un primer acercamiento al análisis cuantitativo de la radiación. Basándose en ese libro como una información preliminar, se puede aplicar estos conocimientos hacia enfoques más específicos de sistemas fotovoltaicos que utilizan la radiación solar como recurso. Específicamente, el capítulo 11 [23] detalla cómo modelar series temporales para la radiación solar utilizando el proceso ARMA a través de la metodología Box-Jenkins. Este capítulo se enfoca en los factores individuales que contribuyen a los datos globales en la serie temporal. Lo cual permite la comprensión de los mecanismos físicos subyacentes y mejora la capacidad de analizar y predecir el comportamiento futuro basado en la serie temporal.

De la misma manera los autores Tyagi et al. [11] combinan avances recientes en investigación sobre energía solar, cubriendo temas como tecnología celular, sistemas térmicos y aplicaciones. Este libro se centra en el creciente papel de la energía fotovoltaica y colectores solares térmicos, respaldados por iniciativas gubernamentales y avances de la industria. Se describen temas como las futuras hojas de ruta energéticas, las innovaciones tecnológicas en energía y los sistemas solares térmicos. Además, se destacan otras aplicaciones industriales y domésticas de la energía solar como la refrigeración, la desalinización, la gasificación y la cocina.

Un desafío importante en el análisis del desempeño de los sistemas fotovoltaicos radica en la variabilidad y complejidad de los factores que influyen, como las condiciones ambientales, la degradación a nivel de componentes y las ineficiencias a nivel de sistema. Por ejemplo, la variabilidad temporal y espacial de la irradiación solar o las fluctuaciones de temperatura pueden complicar la identificación del comportamiento real del sistema frente al ruido ambiental.

Con esto en mente, el objetivo de Mohammadi et al. [1] es desarrollar un método preciso y eficiente para predecir la radiación solar global diaria (GSR) mediante la identificación de los parámetros de entrada más relevantes utilizando un sistema de inferencia neurodifusa adaptativa (ANFIS). De esta forma se busca optimizar la precisión de la predicción y minimizar la complejidad del modelo, teniendo en cuenta la variabilidad de las condiciones climáticas y las características de la radiación solar en diferentes ubicaciones. Los resultados obtenidos enfatizan la importancia de la selección apropiada de los parámetros de entrada para predecir la GSR diaria. Una predicción adecuada, simple y precisa de este tipo es rentable para diseñar y evaluar adecuadamente el rendimiento de los sistemas de energía solar, lo que posteriormente conduce a beneficios técnicos y económicos.

De manera similar, los mecanismos de degradación, como la degradación inducida por potencial o la suciedad, pueden ocurrir a diferentes velocidades, lo que complica aún más las evaluaciones de desempeño. Estos factores introducen incertidumbres que pueden ocultar las tendencias de desempeño o conducir a diagnósticos incorrectos, por lo que es ideal reducir estas degradaciones.

Con la intención de supervisar y monitorear el funcionamiento de sistemas fotovoltaicos, los autores Vergura et al. [24] examinan un conjunto de medidas energéticas para obtener información offline sobre la planta fotovoltaica, permitiendo identificar las áreas donde se producen fallos o malfuncionamientos. También se resalta la capacidad de realizar un seguimiento en tiempo real de la planta fotovoltaica utilizando datos muestrales. El uso de estadística descriptiva e inferencial del algoritmo propuesto es una herramienta eficaz para la detección de fallos y la monitorización remota de plantas fotovoltaicas.

Varios estudios buscan analizar el rendimiento de paneles solares FV en zonas específicas para evaluar la productividad de esas zonas. El estudio de Pashiardis et al. [25] ofrece un análisis estadístico valioso y una comparación de la radiación solar en dos sitios de Chipre que representan climas costeros y del interior. Al analizar los componentes de la radiación, los índices de claridad y los factores de turbidez, los autores caracterizan los distintos regímenes de radiación y destacan la mayor intensidad de radiación global en el sitio costero en comparación con el sitio del interior. Estos conocimientos son fundamentales para comprender el potencial de energía solar específico del sitio.

En el artículo de Bocca et al. [26] se describe una metodología para generar modelos de estimación de electricidad fotovoltaica para instalaciones en grandes regiones donde sólo hay disponibles unos pocos datos dispersos o estaciones de medición. El objetivo del método de modelado propuesto es generar un estimador preciso de la energía fotovoltaica de instalaciones potenciales utilizando sólo datos de unas pocas ubicaciones de referencia dispersas, infiriendo así el potencial estimado en una gran área de interés. De la misma manera, Ekström et al. [27] presentó una metodología basada en Monte Carlo para el análisis de la generación de energía fotovoltaica en lugares donde no se miden las fuentes de generación. La metodología presentada se puede utilizar para analizar la generación solar agregada en escenarios futuros con áreas geográficas de diferentes tamaños.

La radiación solar es el parámetro más importante en el diseño y estudio de dispositivos de conversión de energía solar. En el estudio de Oturanc et al. [28], se utilizaron métodos estadísticos para analizar los datos de radiación en Turquía, país que recibe abundante energía solar debido a su ubicación. También se desarrolló un modelo no lineal entre la radiación global diaria promedio mensual y la temperatura ambiente. Para el mismo país, se presenta el artículo de Saglam [2] donde el autor analiza el rendimiento de un panel FV conectado a una carga resistiva permanente. Se registraron y compararon continuamente la salida eléctrica y los parámetros meteorológicos como temperatura, radiación solar, y humedad. Se desarrolló una ecuación matemática utilizando el ajuste de curvas para explorar la relación entre el rendimiento fotovoltaico y las condiciones climáticas, destacando el impacto de los factores ambientales en la producción de energía solar.

El autor Almarshoud [29] explora propiedades relativas de algunos sitios en Arabia Saudita en relación con la producción de energía solar. La investigación se realizó para tres posibles modos de seguimiento solar: ángulo de inclinación fijo, modos de seguimiento de 1 eje y de 2 ejes. Las características de rendimiento se han determinado utilizando tres indicadores: factor de rendimiento, factor de capacidad y relación de rendimiento.

Los autores Amusan y Otokunefor [5] analizan la producción FV en Nigeria e informan sobre el efecto de la nubosidad ligera y densa en el rendimiento de salida del módulo solar. Concluyen que la nubosidad tiene un efecto significativo en el rendimiento de salida del módulo solar. Resaltan que una región que experimenta nubosidad alta perderá una cantidad alta de rendimiento del módulo solar. Por lo que, caracterizar el recurso solar es crucial para dimensionar la efectividad de sistemas de energía que integren esta tecnología de manera confiable.

Laevens et al. [17] se presenta un problema de estimación clásico para calcular la energía generada por sistemas de energía solar fotovoltaica, a nivel diario, anual, regional y nacional. En el caso de los Países Bajos, esta investigación demuestra que es necesaria una medida de la generación de energía solar basada en la irradiación.

Los resultados de este método pueden ser de gran utilidad para la formulación de políticas a nivel regional, donde a menudo se realizan esfuerzos para estimular iniciativas de energía renovable.

El objetivo de Arabali et al. [30] es desarrollar un marco estocástico para el dimensionamiento óptimo y el análisis de confiabilidad de los sistemas de energía híbridos que integran fuentes de energía renovable y almacenamiento de energía. El marco aborda las incertidumbres en la energía eólica, la energía fotovoltaica y la demanda de carga utilizando modelos de promedio móvil autorregresivo (ARMA). Combina un método de optimización de búsqueda de patrones con simulación secuencial de Monte Carlo (SMCS) para minimizar los costos del sistema y, al mismo tiempo, cumplir con los requisitos de confiabilidad.

El trabajo de Martínez y Forero [10] se ha realizado para conocer la relación entre la eficiencia de dos tipos de paneles fotovoltaicos, monocristalinos y policristalinos, y la variación climática en la ciudad de Bogotá-Colombia. Con esta información recolectada se realizó un estudio con todas las variables correlacionadas a través de la herramienta estadística análisis de componentes principales (PCA) donde se vio el impacto de cada variable climática sobre la potencia generada en ambos paneles y por tanto su impacto sobre la eficiencia. Las variables con correlación positiva más alta con la potencia generada por los paneles fotovoltaicos fueron la temperatura y la irradiancia, mientras que la variable que tuvo mayor correlación negativa fue la humedad.

En el artículo de Agoua et al. [31] los autores proponen un modelo estadístico espacio-temporal para mejorar la previsión a corto plazo de la producción fotovoltaica. Por otro lado, Fliess et al. [32] destacan la volatilidad de los datos de producción fotovoltaica y resaltan la estacionalidad. Buscan generar bandas de predicción mediante procesos de descomposición aditiva de la serie de tiempo, afirmando que la implementación es posible sin técnicas arduas de aprendizaje automático y sin datos históricos extensos.

Los modelos de media móvil autorregresiva (ARMA) se han empleado ampliamente en el análisis de series temporales debido a su capacidad para capturar dependencias temporales y comportamiento estocástico. En el contexto de la producción de energía solar fotovoltaica, los modelos ARMA son particularmente útiles para modelar y pronosticar la producción de energía al tener en cuenta patrones en datos históricos y la aleatoriedad inherente. Estos modelos proporcionan una base sólida para comprender la dinámica de la producción de energía y se pueden ampliar o combinar con otros enfoques para mejorar la precisión predictiva.

La metodología de modelado ARMA para la producción FV se propone en "A Guide to Solar Power Forecasting using ARMA Models" [33] y es utilizada por Mbaye et al. [34] reportan usar un modelo ARMA para hacer forecasting en períodos cortos de tiempo. Duran y Minchala [35] aplican modelos ARMA para la predicción de producción

fotovoltaica en comparación con otros modelados de mayor recurso computacional. Al igual que Martin et al. [36] cuyo trabajo presenta una comparación de modelos estadísticos basados en series de tiempo aplicados para predecir valores semidiarios de irradiancia solar global.

David et al. [37] se plantean como objetivo evaluar el rendimiento de una combinación comúnmente utilizada de dos modelos lineales: ARMA y GARCH en econometría para proporcionar pronósticos probabilísticos de la irradiancia solar. Una metodología similar es usada por Carrillo et al. [15] donde se basan en el procesamiento de señales gráficas para la reconstrucción de señales y la predicción con un modelo lineal, espacio-temporal autorregresivo (ST-AR). Los autores proponen un marco robusto basado en gráficos que explota la correlación entre diferentes sistemas para pronosticar la producción fotovoltaica local en función de mediciones imperfectas.

La realización de pronósticos precisos a nivel nacional es crucial para optimizar la gestión energética, fundamentar el desarrollo de políticas y promover la sostenibilidad ambiental. El estudio de Cabello-López et al. [38] tiene como objetivo abordar los desafíos asociados con la variabilidad significativa en la producción de energía renovable y su impacto en la estabilidad de la red mejorando la precisión de los métodos de pronóstico existentes. Asimismo, el artículo de Das et al. [39] realizó una revisión sistemática y exhaustiva de la literatura sobre modelos y técnicas de pronóstico directo de energía fotovoltaica.

Varios estudios de pronósticos resaltan la correlación entre la producción FV y factores meteorológicos específicos de la zona. Huang et al. [40] estudian dos métodos de predicción para sistemas fotovoltaicos: con un modelo físico y un modelo estadístico. Establecen un modelo físico basado en la construcción de sistemas fotovoltaicos y un modelo estadístico de redes neuronales basado en datos históricos, cuyas entradas son datos como irradiancia solar, temperatura del aire, nubes, humedad y posición del sol. Concluyen que la mejora futura de los métodos de predicción energética se basa principalmente en la mejora de la previsión meteorológica.

Kardakos et al. [41] exploran dos métodos de pronóstico para plantas de energía fotovoltaica conectadas a la red: un modelo SARIMA que incorpora pronósticos de radiación solar basados en NWP y un modelo ANN con múltiples entradas. Al comparar estos métodos con modelos SARIMA puros y de persistencia más simples, los autores demuestran que la incorporación de datos de radiación solar externa conduce a valores de error más bajos y una mejor precisión de pronóstico del día siguiente. Sin embargo, el estudio también señala que los modelos más simples funcionan adecuadamente durante el verano en condiciones climáticas estables.

La incorporación de datos de radiación solar en los modelos de pronóstico, como se demuestra en estos estudios, se alinea con los objetivos de esta investigación para mejorar las predicciones de generación de energía fotovoltaica. Al validar la

superioridad de los modelos basados en la radiación, este trabajo proporciona un punto de referencia para evaluar enfoques alternativos de pronóstico en diversas condiciones climáticas.

Por último, un problema común al trabajar con datos de producción FV tomada de sistemas reales es la imperfección de los datos. En la mayoría de los casos, estas imperfecciones se presentan como datos faltantes y se abre paso a una gran cantidad de estudios que buscan lidiar con este problema y encontrar una solución adecuada.

Demirhan et al. [42] se centran en la estimación de valores faltantes en series de irradiancia solar por minuto, hora, día y semana utilizando una amplia variedad de métodos de imputación. Se identificó a la interpolación, el promedio móvil ponderado y los filtros de Kalman como métodos de imputación prometedores para series de irradiancia solar. El método adecuado debe considerar la proporción y distribución de los valores faltantes junto con la frecuencia de la serie de interés.

Por otro lado, los autores Kim et al. [9] concluyen que la imputación de datos faltantes más adecuada para su aplicación a la previsión fotovoltaica es el método de *k-nearest neighbors* (KNN). Al igual que los autores Sundararajan y Sarwat [43] que concluyen que KNN y random forest tienen las menores diferencias y ayudan a las empresas de servicios públicos a realizar predicciones más precisas de la generación para la planificación de la distribución.

Iwueze et al. [44] proponen nuevos métodos para estimar valores faltantes en datos de series temporales. Al comparar con métodos existentes, bajo los supuestos establecidos, las estimaciones del nuevo método basado en la descomposición completa de una serie son las mejores en términos de medidas de precisión.

De igual forma, Zhang et al. [45] proponen un nuevo método solarGAN para la imputación de datos solares multivariados, en el que se realizan las modificaciones necesarias en la entrada de la red generativa antagónica (GAN) para abordar de manera efectiva los datos de series de tiempo. Los estudios de caso sobre un conjunto de datos públicos muestran que el solarGAN propuesto supera a varios métodos de imputación de datos basados en GAN y aprendizaje automático con al menos un 23,9% de reducción del error cuadrático medio.

Como síntesis del presente capítulo, se han organizado tres tablas que resumen los estudios más relevantes revisados. La primera tabla (Tabla I) recoge estudios dedicados a la evaluación y caracterización del rendimiento de sistemas solares en distintas regiones y condiciones. La segunda tabla (Tabla II) agrupa trabajos centrados en la predicción de la generación fotovoltaica mediante distintos enfoques metodológicos ya sea mediante estadística clásica métodos nuevos como redes neuronales. Finalmente, la tercera tabla (Tabla III) presenta investigaciones orientadas a la imputación de datos faltantes en series temporales solares, una problemática

común en contextos de monitoreo y predicción. Estas tablas permiten visualizar de forma estructurada los objetivos, metodologías, métricas empleadas y resultados cuantitativos principales de cada estudio, facilitando la comparación entre enfoques y sus respectivos aportes.

Tabla I
Estudios sobre rendimiento de plantas FV

	Autor	Título	Objetivo	Metodología utilizada	Métricas empleadas	Resultados Principales
[1]	A. Arabali et al.	“Stochastic Performance Assessment and Sizing for a Hybrid Power System of Solar/Wind/Energy Storage	Optimizar el dimensionamiento y evaluar la confiabilidad de un sistema híbrido de energía renovable en condiciones de incertidumbre	Modelo estocástico ARMA y método de optimización basado en búsqueda de patrones junto con una simulación secuencial de Monte Carlo	Costo total del sistema (NPC), confiabilidad del suministro	Los costos se encuentran en el rango de 8.07 a 8.30 M\$ para para valores de energía esperada no suministrada (EENS) entre 0.09 y 0.1
[3]	A. F. Almarshoud	“Performance of solar resources in Saudi Arabia	Evaluar el potencial solar y la eficiencia de producción FV en diversas regiones de Arabia Saudita	Análisis estadístico de datos históricos de radiación solar en 40 estaciones	Promedios mensuales, factor de capacidad, tasa de conversión	Factor de capacidad promedio: 18–22%; producción media anual: ~1700 kWh/kWp
[6]	B. P. M. Laevens et al.	“An observational method for determining daily and regional photovoltaic solar energy statistics	Estimar la producción FV diaria y regional mediante observaciones empíricas	Fusión de datos de producción e irradiancia modelada con muestreo Monte Carlo	Rendimiento energético (kWh/kWp)	rendimientos energéticos anuales específicos en el rango de 877-946 kWh/kWp y 838-899 kWh/kWp para 2016 y 2017 respectivamente
[9]	G. Oturanc et al.	“Statistical Analysis of Solar Radiation Data	Analizar estadísticamente datos de radiación solar y estimar irradiancia con modelos empíricos	Ajuste de modelo de Angström modificado, regresión no lineal H-temperatura	Índice de claridad (KT), horas de sol, promedios mensuales, coeficientes de regresión	Horas de sol entre 9.4–14.6 h/día; KT entre 0.54–0.60; el modelo no lineal predice H con buena precisión
[13]	J. Amusan and E. Otokunefor	“The Effect of Cloud on the Output Performance of a Solar Module	Comparar el efecto de la nubosidad ligera (escasa) y densa (gruesa) en el rendimiento de salida del módulo solar	Comparación experimental del rendimiento del módulo solar en condiciones despejadas y nubladas utilizando mediciones en intervalos de 15min	Potencia y % de pérdidas	Pérdida observada del 23,80 % debido a una ligera nubosidad y del 66,75 % debido a nubosidad densa

[15]	J. Ekström et al.	“A statistical approach for hourly photovoltaic power generation modeling with generation locations without measured data	Clasificar días basado en forma de curvas de irradiancia diaria.	Morfología matemática y técnicas clásicas de agrupación	FANOVA, anchos de silueta medios	Valor p de la prueba FANOVA para los cuatro grupos < 0.05
[22]	R. Martinez and E. Forero	“Estimation of energy efficiency in solar photovoltaic panels considering environmental variables	Determinar la relación entre la eficiencia de paneles FV monocristalinos y policristalinos, y la variación climática.	Recopilación de datos experimentales de paneles y monitoreo meteorológico análisis estadístico mediante PCA.	Principal Components Analysis (PCA)	La eficiencia promedio para el panel monocristalino fue del 9,45% mientras que en el panel policristalino fue del 8,48%.
[23]	S. Pashiardis et al.	“Statistical analysis for the characterization of solar energy utilization and inter-comparison of solar radiation at two sites in Cyprus	Caracterización e intercomparación de los dos sitios en Chipre con respecto a las intensidades de radiación global, de haz y difusa.	Análisis estadístico y comparación entre sitios (promedios, desviaciones, correlaciones)	Media, coeficiente de variación, asimetría y curtosis	Sitios caracterizados por tasa de irradiación diaria relativamente alta. Frecuencia alta de días despejados. Larnaca presenta tasas de radiación global superior a Anthalassa
[25]	S. Vergura et al.	“Descriptive and Inferential Statistics for Supervising and Monitoring the Operation of PV Plants	Diagnosticar fallas de inversor en plantas FV mediante análisis estadístico en tiempo real	Supervisión offline y monitoreo online basado en estadística descriptiva e inferencial	Media, mediana, desviación estándar, outliers, intervalos de confianza	Anomalías fuera del 95% de confianza; corriente atípica previa a falla; validación exitosa del sistema

Tabla II
Estudios sobre predicción solar

	Autor	Título	Objetivo	Metodología utilizada	Métricas empleadas	Resultados Principales
[2]	A. Bocca et al.	“Estimating photovoltaic energy potential from a minimal set of randomly sampled data	Estimar generación FV en zonas con pocos datos de medición	Modelación de estimadores de producción FV	Mean Absolute Error (MAE)	Electricidad FV generada MAE no excede 6%, y es menor al 4% en el 90% de los casos. El valor p obtenido es de 0.26 para nivel de significación de 5%.
[4]	A. Mbaye	“ARMA model for short-term forecasting of solar potential: application to a horizontal 24ermal24 of Dakar site.	Generar un modelo para predicción de producción FV a corto plazo	Modelación ARMA	AIC, MAE, RMSE	Se obtuvo un RMSE =0.629 y un coeficiente de determinación, $R^2 = 0.963$. El MAE=0.528 y MBE=0.012 (Mean Bias Error).
[8]	E. Kardakos et al.	Application of Time Series and Artificial Neural Network Models in Short-term Forecasting of PV Power Generation	Predecir la generación FV de plantas conectadas mediante modelos ARIMA y redes neuronales	Seasonal ARIMA y ANN con 7 entradas y 1 capa oculta	RMSE Normalizado	Los modelos ANN (NRMSE=11.26) y el modelo SARIMA modificado (NRMSE=11.12) son superiores en comparación con el modelo de persistencia.
[16]	J. F. Duran and L. I. Minchala	“A Comparative Study on Time Series Prediction of Photovoltaic-Power Production Through Classic Statistical Techniques and Short-Term Memory Networks	Comparar modelos estadísticos clásicos (ARIMA) y redes neuronales LSTM para predicción FV	SARIMA vs LSTM para series temporales fotovoltaicas	BIC, RMSE, NRMSE y tiempo de computación	NRMSE < 1% para todos los modelos, para SARIMA: sube RSME para predicciones a largo plazo mientras que para LSTM se mantiene constante

[17]	K. Mohammadi et al.	"Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure	Identificar los parámetros más relevantes para la predicción de la radiación solar global diaria.	Selección de variables de entrada de un Sistema de Inferencia Neurodifusa Adaptativa (ANFIS) para predicción adecuada	MABE, RMSE y coeficiente de correlación R	Considerar las combinaciones más relevantes de 2 parámetros de entrada es la opción más adecuada para que todas las ciudades logren mayor precisión y menor complejidad en las predicciones.
[18]	L. Martín et al.	"Prediction of global solar irradiance 25ermal25s time series 25ermal25s: Application to solar 25ermal power plants energy production planning	Predicción de irradiancia global para planificación de plantas solares térmicas	Comparación de métodos AR, NN, y lógica difusa con índice de claridad y componente perdido como entradas	rRMSE	La mejora del modelo final frente a la persistencia presenta RMSE= 25,58% para el último horizonte temporal de predicción.
[19]	M. David et al.	"Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models	Pronóstico probabilístico de irradiancia usando modelos ARMA-GARCH	Modelos ARMA-GARCH para capturar heterocedasticidad	RSME, CRPSS, KSI	El modelo es sobreconfiado cuando la probabilidad del pronóstico es inferior al 60%-70% y poco confiado cuando está por encima de este umbral.
[20]	M. Fliess et al.	"Prediction bands for solar energy: New short-term time series forecasting techniques	Proponer bandas de predicción para energía solar a corto plazo	Técnicas de estimación nuevas y procedimientos de extrapolación clásicos con énfasis en estacionalidad	Mean Interval Length (MIL), Prediction Interval Coverage Probability (PICP)	Para la irradiancia, el error absoluto medio (EMA) entre los pronósticos y los valores reales es del 26,6% frente al 35,44%.
[21]	R. E. Carrillo et al.	"High-Resolution PV Forecasting from Imperfect Data: A Graph-Based Solution	Desarrollar un método robusto para la predicción de energía FV en múltiples emplazamientos	Procesamiento de señales gráficas y un modelo autorregresivo espacio-temporal (ST-AR) aplicado a datos FV.	NRMSE	NRMSE = 13% para datos reales y NRMSE = 9% para datos sintéticos

[26]	T. Cabello-López et al.	“Forecasting solar energy production in Spain: A comparison of univariate and multivariate models at the national level	Comparar modelos univariados y multivariados para pronóstico solar en España	Modelos ARIMA univariados y VAR multivariados	MAE y tiempo de ejecución	Mejora del 47,58% y 41% de la MAE media para los enfoques multivariados y univariados, respectivamente
[31]	X. G. Agoua et al.	“Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production	Desarrollar un modelo de predicción de energía FV de corta duración (0,-6 horas)	Modelo estocástico autoregresivo	RMSE	Se logran mejoras de RMSE entre el 15% y 20%. Para evaluación espacio-temporal se logra reducción de 28%
[32]	Y. Huang et al.	“Comparative study of power forecasting methods for PV stations	Comparar métodos de predicción para estaciones FV	análisis de modelos físicos y estadísticos a tres diferentes rangos de tiempo	nRMSE	nRMSE de todo el año = 12.5% para el modelo físico y 10.5% para el modelo estadístico NN.

Tabla III
Estudios sobre métodos de imputación de datos

	Autor	Título	Objetivo	Metodología utilizada	Métricas empleadas	Resultados Principales
[5]	A. Sundararajan and A. I. Sarwat	“Evaluation of Missing Data Imputation Methods for an Enhanced Distributed PV Generation Prediction	Comparación de diferentes métodos de imputación de valores perdidos	kNN, random forest y random imputation	Métricas de error y magnitud de efectos	kNN presenta RMSE=8.09, Random forest RSME=29 y Random RMSE=212
[10]	H. Demirhan and Z. Renwick	“Missing value imputation for short to mid-term horizontal solar irradiance data	Evaluar métodos de imputación para datos de irradiancia solar en diferentes rangos de tiempo	Comparación entre 36 métodos de imputación basado en datos reales y completos de irradiación solar	rMAE, rRMSE, MASE	Buen rendimiento de MA para tasas de datos faltantes entre 5%, 25%, y 50%. Filtros de Kalman también funcionan bien con el 50% de datos faltantes
[12]	I. S. Iwueze et al.	“Comparison of Methods of Estimating Missing Values in Time Series	Comparar métodos para estimar valores faltantes en series temporales	Las medidas de precisión se calculan a partir de las desviaciones de estimaciones vs. valores reales.	MAE, MAPE y RMSE	Decomposing Without the Missing Value (DWMV) produjo la mejor precisión entre los otros métodos: MAE=7.6, MAPE=4.3 y RMSE=10.5
[27]	T. Kim et al.	“Analysis and Impact Evaluation of Missing Data Imputation in Day-ahead PV Generation Forecasting	Evaluar el impacto de diferentes métodos de imputación en la precisión de la predicción de energía FV.	Comparación de cuatro métodos de imputación a datos meteorológicos en un modelo de predicción de regresión de vectores de soporte (SVR).	RMSE, MRE, RMSD, MRD	Cuando se aplica el método de imputación a los datos de entrenamiento, el RMSD máximo es 5 kW y el MRD < 2%. kNN produce el error más pequeño.
[30]	W. Zhang et al.	“SolarGAN: Multivariate Solar Data Imputation Using Generative Adversarial Network	Imputación multivariante usando redes generativas adversarias (GAN)	SolarGAN, GAN multivariantes para imputación de datos solares	Error cuadrático medio (MSE)	MSE reducido en un 25% respecto a métodos clásicos de imputación

Metodología

Este capítulo describe los métodos y procedimientos usados para analizar los datos de la serie de tiempo de producción fotovoltaica para cada grupo de paneles solares instalados en el laboratorio de Micro-Red del campus Balzay de la Universidad de Cuenca. El objetivo de este análisis es modelar el comportamiento de la producción fotovoltaica a lo largo del tiempo y generar una serie de tiempo simulada que refleje los componentes estadísticos de la serie de tiempo original. Para lograr este análisis, la metodología incluye pasos clave como: preprocesamiento de los datos, desestacionalización, modelado ARMA de residuales y la generación sintética de una serie de tiempo. Además, se busca caracterizar estadísticamente la producción FV diaria con curvas de percentiles. Las curvas de percentiles describen la variabilidad de los datos a lo largo del año mostrando patrones típicos y el rango de niveles de producción a lo largo de los días. A estas curvas se les ajustaron funciones que describan su comportamiento con la intención de obtener formas parametrizadas para el cálculo de cualquier valor de percentil. Cada uno de estos pasos se describe en detalle para garantizar la transparencia y reproducibilidad del análisis.

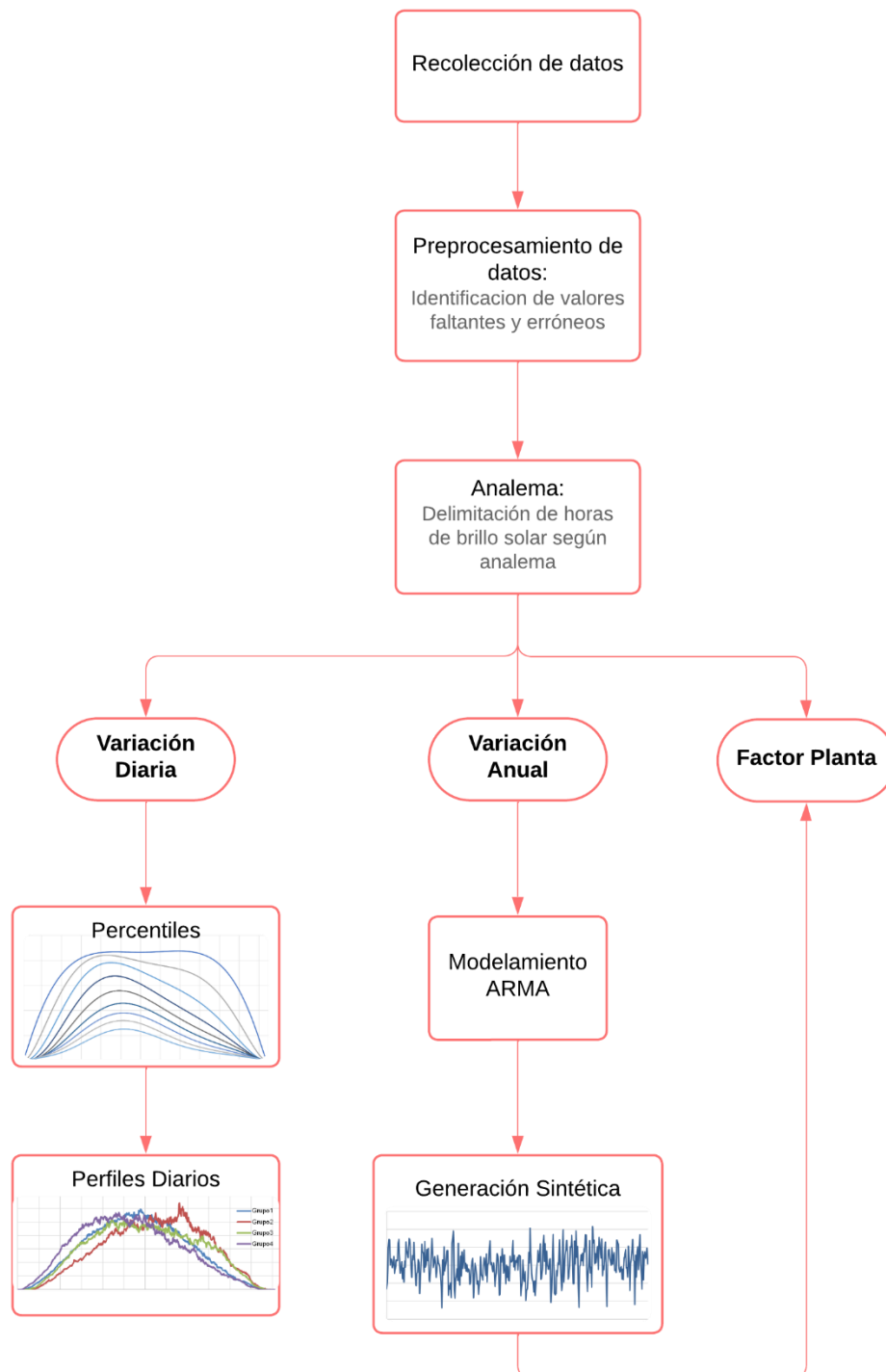
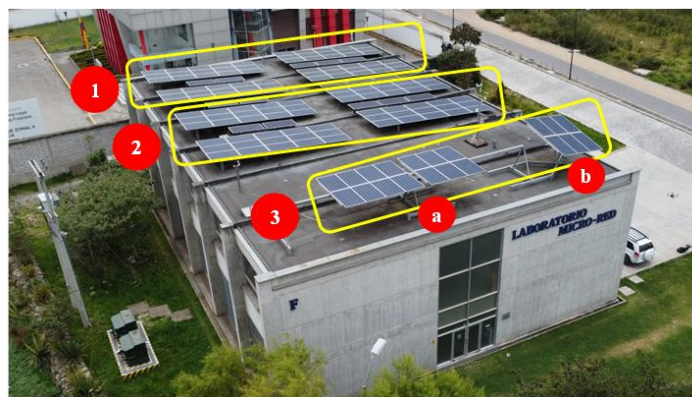


Fig. 1 Diagrama de flujo de la metodología
Fuente: Elaboración propia.

3.1 Recolección de datos

Se obtienen los datos históricos de producción de electricidad de la granja solar del edificio del laboratorio de Micro-Red registrada durante todo el año 2022. Los datos representan la potencia en kW generada por cada uno de los arreglos fotovoltaicos disponibles. La frecuencia de muestreo o resolución temporal de los datos es de un minuto, por lo que se tienen 1440 datos diarios por cada día del año.

Existen 3 grupos de datos, correspondientes a los grupos de paneles solares: SFV1 (Fijos monocristalinos), SFV2 (Fijos policristalinos) y SFV3 (Móviles policristalinos), cada uno con una potencia nominal de 15kW, 15kW y 5kW respectivamente, cuya ubicación se presenta en la Fig. 2.



1. SFV1 – Fijo: 60 Paneles silicio monocristalino (15kW)
2. SFV2 – Fijo: 60 Paneles silicio policristalino (15kW)
3. SFV3 – Móvil : 20 Paneles silicio policristalino (5kW)
movimiento en a) 1 eje; b) 2 ejes.

Fig. 2 Distribución de paneles

Fuente: Archivo del laboratorio de Micro-red, Ing. Danny Ochoa.

Asimismo, se obtienen los datos históricos de la estación meteorológica que se encuentra en el mismo lugar. Estos datos fueron usados para encontrar valores numéricos que pudieran resultar erróneos al comparar el comportamiento de los datos de radiación solar y producción FV, que son las variables con mayor correlación [26].

3.2 Preprocesamiento de datos

Los datos obtenidos del SCADA tienen que ser transformados para ser leídos correctamente por los programas informáticos Excel y R. La transformación implica verificar el formato de las fechas para evitar errores en la lectura de día vs mes. Los datos se leen en el formato mm/dd/aaaa. Para la lectura en R se necesita que el documento en donde se encuentran los datos no exista espacios.

Aparte del formato de la data original, se encuentran y se reportan los valores de error, celdas vacías o valores *Not a Number* (NaN). Para facilitar la interpretación de los datos por el software, los valores de error (NaN) son eliminados de la data, es decir se dejan las celdas vacías.

Como un primer paso para el tratamiento de los datos se pretende eliminar estos valores erróneos o no existentes. Para esto se examinó los datos buscando vacíos o valores negativos durante el día que indican algún error. Esto permitió eliminar días que presentan anomalías como una falta de datos, días en los que se corta la producción FV o en los que no se registran valores sino hasta algunas horas después de la salida del sol.

Un paso que facilitó el tratamiento de los datos fue transformar el vector a una tabla. Para esto, se procedió a separar la variable de tiempo en los valores de hora y del día para así tener en una columna las fechas y en otra los minutos. Con esto se realizó la transformación del vector a una tabla: *pivot table* con los datos del día en las columnas y los datos de la hora en las filas. Esto también permitió la obtención de valores totales diarios.

Los valores diarios que resultan del acumulado de potencia representan un indicador de producción FV. Debido a que la frecuencia de muestreo de potencia es de minuto a minuto, en términos de energía (3.1), se tiene que la producción FV en kWh es la potencia [kW] por 1/60 h.

$$\text{Energía} = \text{potencia} \times \text{tiempo} \quad (3.1)$$

De tal forma que, la producción FV diaria, será el acumulado diario de potencia dividido para 60.

Debido a que los datos fueron tomados cada minuto independientemente de las horas de presencia solar o no, se consideró que no es necesario procesar los datos de las horas de la noche. Esto implica un análisis centrado únicamente en el comportamiento de la producción FV debido al recurso solar. Se obtiene así una reducción a casi la mitad de los datos que se requieren procesar evitando correlaciones triviales de valores registrados en la ausencia de sol.

3.3 Análisis de variación anual

3.3.1 Analema

La inclinación de eje terrestre con respecto a la eclíptica, la rotación de la tierra respecto a su eje y su traslación alrededor del sol ocasiona diferentes patrones de radiación en su superficie, de acuerdo con su ubicación geográfica, que se capturan en la analema. Este permite establecer con exactitud las horas de radiación solar. En principio, para optimizar la data, se podría descartar el tiempo en que no se encuentra

luz solar, acortando la información desde y hasta las horas aproximadas de salida y puesta del sol. Sin embargo, para obtener mejor precisión, mediante la analema se puede obtener la hora exacta de salida y puesta del sol en cada día del año. Para esto se hizo uso de una hoja de cálculo proveniente de la NOAA¹ que describe el patrón solar según las coordenadas de latitud y longitud [46]. Ingresando las coordenadas² del lugar de emplazamiento de los paneles y seleccionando las fechas para las cuales se requieren los datos, en la hoja electrónica se realizan automáticamente los cálculos que proporcionan las horas exactas en el día de salida y puesta del sol. Debido a que los paneles solares se encuentran muy cercanos a la línea ecuatorial, la variación entre estas horas es muy pequeña pero no despreciable, y permite obtener gráficos con mejor resolución e indicación de los rangos de horas reales de presencia y ausencia de luz solar. Estos datos se evidencian, en la Tabla IV, donde se presenta una descripción general de las horas de salida y puesta del sol con promedios mínimos máximos y variación máxima.

Tabla IV
Horas de Salida y Puesta del Sol

	Promedio	Mínimo	Máximo	Delta
Salida de sol	6:12:45	5:52:54	6:24:18	0:31:24
Puesta de sol	18:19:33	18:06:02	18:36:55	0:30:52

Fuente: Elaboración propia.

Con datos diarios de presencia de luz solar se pueden eliminar los datos tomados por el sistema en horas de la noche. Esto facilita el procesamiento al reducir significativamente el volumen de datos. La variación en la salida y puesta del sol se puede apreciar en la

Fig. 3 donde se evidencia la curva azul para la salida del sol y la curva naranja para la puesta de sol.

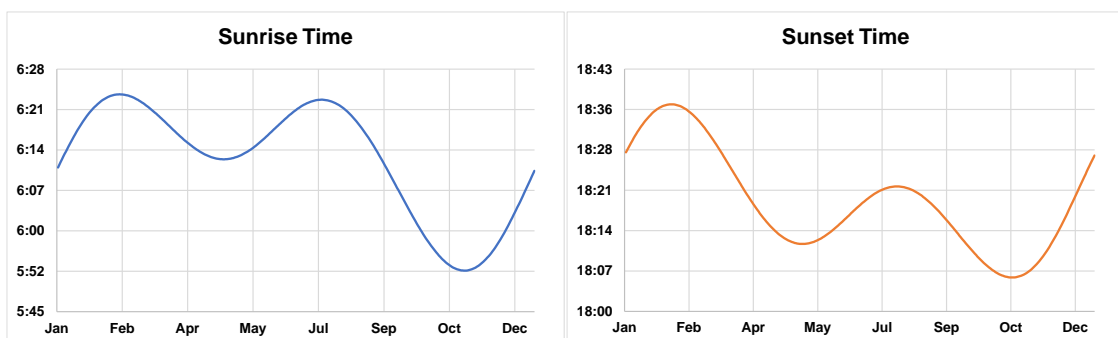


Fig. 3 Variación anual de amanecer y atardecer

Fuente: Elaboración propia.

¹ NOAA: National Oceanic and Atmospheric Administration

² Coordenadas WGS84 de ubicación de los paneles: -2.89198553771969, -79.0385313693698

3.3.2 Transformación Box-Cox

Una herramienta que se puede considerar necesaria en el procesamiento de los datos es la transformación Box-Cox, que permite normalizar los residuos de las series de tiempo [45]. Para esta transformación se usa la ecuación (3.2):

$$y^T = \frac{y^\lambda - 1}{\lambda} \quad (3.2)$$

Donde y^T es la variable transformada y λ es el parámetro que controla la transformación. λ se estima a partir de los datos buscando corregir sesgos en la distribución de errores. Muchos programas estadísticos modernos (incluyendo R) incorporan potentes rutinas de Box-Cox, y es relativamente sencillo utilizar un script o función para examinar automáticamente un amplio rango de valores λ y determinar rápidamente la transformación óptima [45]. Al implementar esta función en R para los tres grupos de paneles, se busca que la transformación normalice los residuos y equalice la varianza de la data.

3.3.3 Procesamiento para modelado ARMA

El procedimiento de Box-Cox puede revelar la necesidad o no de hacer una transformación de la variable. En el caso que la transformación sea significativa, ésta se la aplica en conjunción con el modelo lineal seleccionado. El modelo lineal se empleó para remover la estacionalidad; y en caso necesario se aplicaría una estandarización de la varianza para procurar un proceso estacionario en sentido amplio. Los residuos de este proceso se ajustaron a un modelo ARMA, con parámetros óptimos identificados a través de la función de autocorrelación (ACF), de la función de autocorrelación parcial (PACF) y criterios de selección. Se verificaron los residuos para confirmar que se asemejaran al ruido blanco. Finalmente, el modelo ARMA se probó simulando series temporales sintéticas a través de un proceso inverso. Todo esto se describe en el diagrama de flujo de la Fig. 4 inspirado en el diagrama de [48].

Para el modelamiento ARMA de los datos se procede primero a encontrar un modelo lineal de la señal. Esto permite desestacionalizar o hacer un *detrending*, que permita descomponer la serie de tiempo.

El modelo lineal que se busca encontrar es de característica periódica y cíclica, es decir, tiene un período de un año. Para esto se utilizan funciones senoidales con período de 365 días y corresponden a la función presentada en la ecuación (3.3):

$$ML = A * \cos(2\pi fX) + B * \text{sen}(2\pi fX) \quad (3.3)$$

Los valores para cada una de las incógnitas se obtienen al correr la función de modelo lineal en R.

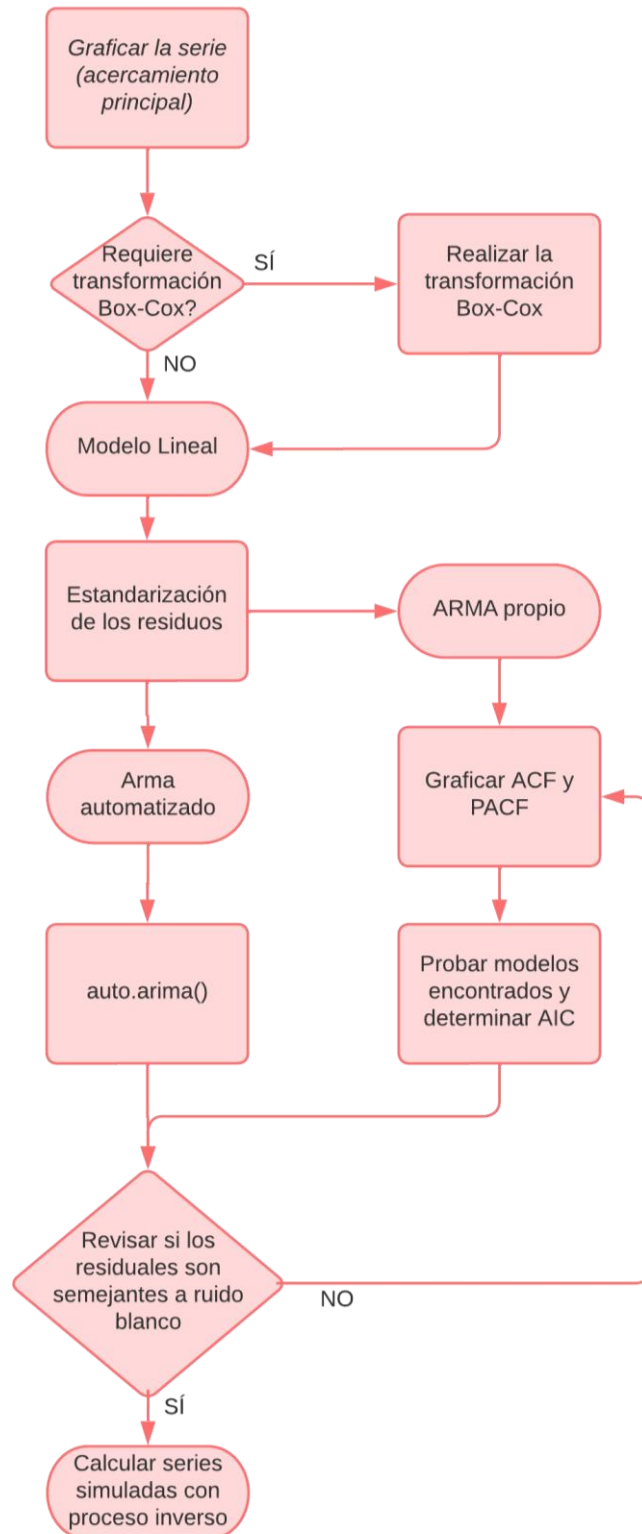


Fig. 4 Diagrama de flujo de modelamiento ARMA
Fuente: Elaboración propia, adaptado de [48].

Se obtienen resultados diferentes para cada grupo de paneles que se reportan en la sección de resultados. En el caso del grupo de paneles de silicio monocristalino (SFV1) fue necesario incorporar además una sinusoidal que tenga un período de medio año, mientras que en los otros dos grupos se mantuvo el período único cíclico anual. En cualquier caso, con este modelo lineal se puede encontrar un componente de la señal que este desestacionalizado.

Para el caso en que, del análisis de residuales, se encuentre que las varianzas no son constantes durante el año es necesario realizar una estandarización de la varianza al dividir los datos para un valor de desviación estándar obtenida dentro de una ventana de 7 días. Para encontrar este valor, se realizó un procedimiento similar al anterior donde se encuentra un modelo lineal, en este caso de los residuos.

Este modelo se usa para determinar la función aproximada de la desviación estándar y se obtienen los puntos para los cuales se van a dividir los datos. De esta manera queda estandarizada la variable y se puede pasar a la siguiente etapa donde se busca calcular el modelo ARMA de los residuos.

El modelo ARMA se consigue al implementar la función en R para obtener los valores ideales de ϕ 's y θ 's. Sin embargo, también se puede hacer un análisis de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para comparar el grado del modelo (AR y MA).

Las funciones de autocorrelación están relacionadas con el modelo ARMA como se explica en la Tabla V tomada de [49] y [50]. El comportamiento de ACF y PACF disminuye gradualmente con el número de rezagos o *lags*, o disminuye bruscamente después de un *lag* determinado. Estos comportamientos se ejemplifican en la Fig. 5 tomados de modelos generados con el propósito de resaltar las características descritas.

Tabla V
Comportamiento de ACF y PACF en modelo ARMA

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Fuente: Elaboración propia adaptado de [49], [50]

Los gráficos ayudan a seleccionar los parámetros p y q del modelo ARMA o verificar que el modelo escogido automáticamente por el programa sea el correcto.

Finalmente, para comprobar que el modelo ARMA obtenido es adecuado, se verifica que los valores residuales se asemejen a ruido blanco. Una vez que se comprueba que es ruido blanco se procede a realizar series sintéticas usando el proceso inverso.

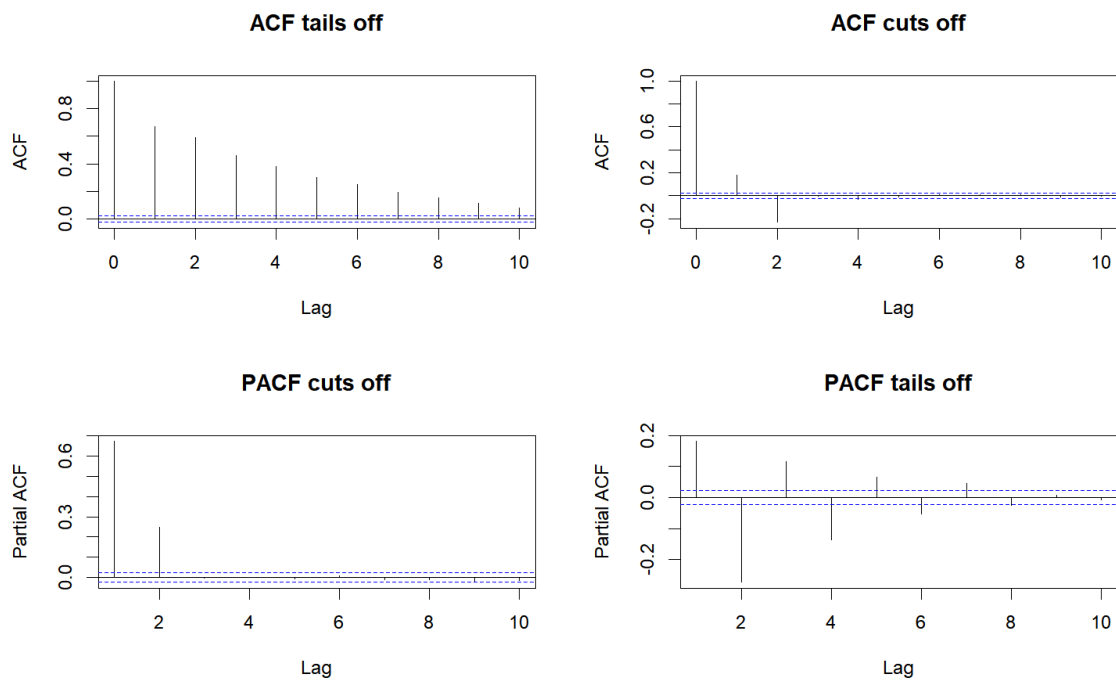


Fig. 5 Ejemplos de comportamiento de ACF y PACF en modelo AR y MA
Fuente: Elaboración propia.

3.3.4 Generación de series sintéticas

Después de identificar los modelos ARMA para cada grupo de paneles solares, se generan series de tiempo sintéticas para simular la producción fotovoltaica anual con características similares a la data original. Ese enfoque permite crear datos adicionales que reflejen los patrones originales de la serie de tiempo, incluyendo cualquier autocorrelación inherente, varianza y dependencias temporales.

La generación de series sintéticas se desarrolló invirtiendo el proceso de modelamiento ARMA, usando los parámetros estimados para cada grupo de paneles. Para cada modelo se generó datos aleatorios similares a ruido blanco que simulen los residuos observados en la serie de tiempo original. A continuación, se presenta un resumen de los pasos a seguir:

1. Generación de ruido blanco: Se generó una serie que servirá como entrada para cada modelo ARMA. El ruido blanco es una serie estocástica normalmente distribuida, con media cero y desviación estándar igual a la obtenida en los residuos originales. Este ruido asegura que la serie sintética presente el mismo comportamiento estocástico.
2. Aplicación del modelo ARMA: Al ruido blanco generado se aplica los parámetros del modelo ARMA para cada grupo de paneles. Esto genera una autocorrelación entre los datos que refleja la estructura temporal de los residuos ajustados al modelo ARMA.

3. Desestandarización de la varianza: Una vez obtenida la serie ARMA, se reintroduce la variabilidad de la desviación estándar. Esto inyecta diferencias en la varianza de la serie generada, que se presentan en la serie de tiempo original, ya que se descubrió que esta no era uniforme.
4. Reintegración de la tendencia: La serie generada se combina con el modelo lineal que describe la tendencia de los datos para restaurar los patrones temporales de largo plazo. Asegurando así que la serie sintética presente las mismas variaciones estacionales de la data original.

Para confirmar la precisión de la serie sintética, se compararon métricas estadísticas clave, como la media, la varianza y la estructura de autocorrelación, entre la serie temporal sintética y la original. Al validar estas propiedades, nos aseguramos de que los datos sintéticos conservaran las características esenciales de la producción de energía solar observada y pudieran usarse de manera confiable en análisis posteriores.

3.4 Análisis de variación diaria

Para el análisis de la variación diaria de la producción fotovoltaica se construyen patrones diarios calculando varios percentiles para cada día. Mostrando así una caracterización de patrones de producción FV diarios.

Una herramienta útil para la caracterización estadística del comportamiento de la producción FV diaria para una determinada hora, es la representación de percentiles. Los percentiles demuestran la probabilidad de excedencia de un valor de radiación a una hora cualquiera. Con este análisis se pretende obtener una clara probabilidad de excedencia de producción FV a una hora específica dentro del intervalo de radiación dado por la analema.

Inicialmente, se tomaron los datos acumulados de 5 minutos para todos los días con un proceso de ajuste para estandarizar la longitud diaria según la analema. Esto resulta en uniformidad para la longitud de días, con relación a la posición del sol en el cielo, a diferencia del horario regular.

3.4.1 Cálculo de percentiles y ajuste de la función beta

Mediante el uso de MS Excel, se calcularon los percentiles para representar la distribución de la producción FV a lo largo del día. Esto se hizo automáticamente con funciones existentes de Excel. Luego, para cada curva de percentil se ajustó una distribución beta, que es comúnmente usada para modelar variables limitadas dentro de un rango [50].

La función de densidad beta está dada por (3.4):

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \text{ for } x \in [0,1] \quad (3.4)$$

Para encontrar el mejor ajuste se buscó minimizar la suma de diferencia de cuadrados entre la función beta ajustada y la curva original de percentil. Cada curva de función beta tiene 3 parámetros α , β y k que se ajustan, buscando obtener la mejor coincidencia entre los percentiles observados y la función beta.

Este proceso de ajuste resultó en 3 parámetros para cada valor de percentil, lo que resultó en más de 60 parámetros en total para describir todas las curvas encontradas. Con la idea de optimizar el ajuste reduciendo la cantidad de parámetros necesarios, se busca modelar el comportamiento de cada variable. Tomando en cuenta la variación de estos parámetros se logró consolidar estos a un grupo de parámetros más manejable.

El comportamiento observado de cada una de las variables (α , β , k) evidencia una relación funcional que puede ser estimada para describirlas en forma general. Al encontrar funciones en términos de la hora del día y el percentil en análisis, se reduce el número de parámetros necesarios para caracterizar las curvas de percentiles. De tal forma que se puede obtener la curva para cualquier percentil.

3.5 Perfiles diarios

Para analizar los patrones en los datos de radiación solar, se implementó un enfoque de agrupamiento utilizando los perfiles de producción FV en horas de luz solar diarias. El agrupamiento permite identificar perfiles diarios similares entre si formando un conjunto que tenga características diferentes a otros grupos. Al agrupar los datos se busca identificar los diferentes perfiles diarios que se presentan en los datos independientemente de la cantidad de producción total diaria.

Para cada día, se registraron los datos puros de potencia tomando en cuenta los límites de presencia de luz solar según la analema como se presentó en la Tabla IV anteriormente. Esto permite delimitar los datos para no considerar los valores en la noche y agrupar los perfiles de captura sin sesgos introducidos por similitudes en horas donde no hay producción. Una vez que se tuvo los datos con los minutos delimitados, se procedió a realizar un acumulado de los valores para obtener las curvas diarias que servirán para la clasificación de los datos. Estas curvas de acumulados presentan la progresión de la radiación solar a lo largo del día. Sin embargo, debido a que cada día se obtienen diferentes totales de producción, los acumulados también presentan estas diferencias en la cantidad máxima que alcanzan. La Fig. 6 ejemplifica esto, al presentar una muestra de las curvas acumuladas donde se evidencia que, si bien todas las curvas comienzan en 0, alcanzan diferentes valores máximos diarios.

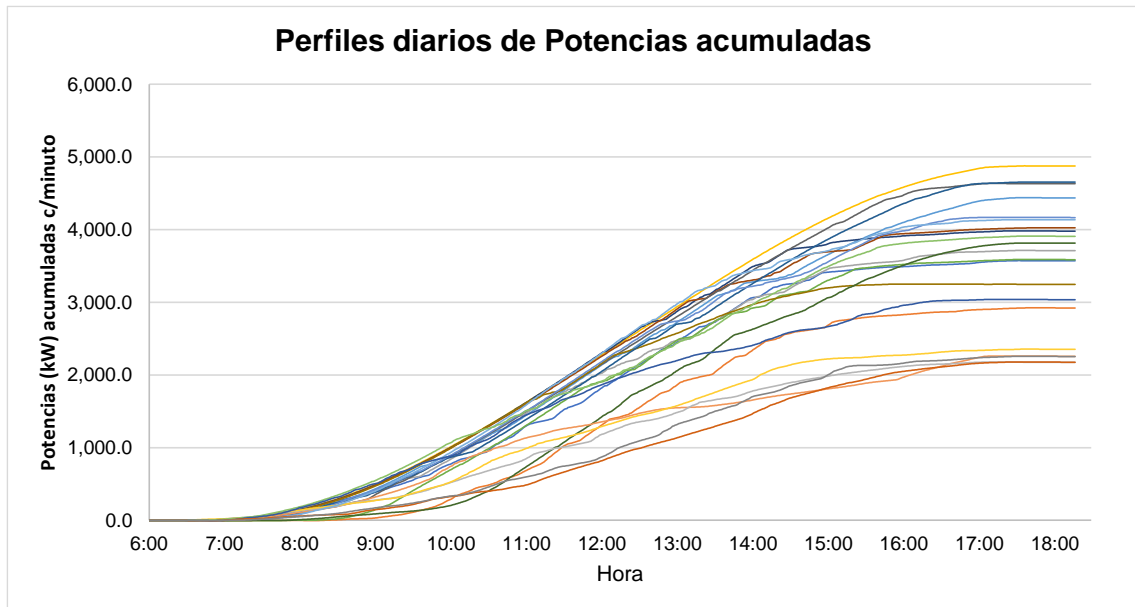


Fig. 6 Acumulado diario
Fuente: Elaboración propia.

Esto puede complicar el análisis de agrupaciones ya que se agruparían los días por cantidad total de energía capturada y no por su perfil diario. Considerando que la agrupación se desea hacer por tipo de perfil diario, se debe tomar un paso de estandarización y así obtener las todas curvas en el rango (0,1) y poder clasificarlas según su comportamiento gráfico y no la producción alcanzada.

Por lo tanto, se busca dividir los valores acumulados para el total, o el valor final, logrando así una estandarización de las curvas. Este proceso de normalización eliminó las diferencias en los valores absolutos y permitió que el enfoque se desplazara a la forma de las curvas acumuladas diarias. Un ejemplo de la estandarización se presenta en la Fig. 7

A partir de este punto ya se puede proceder con el agrupamiento. Las curvas diarias estandarizadas se analizaron utilizando procesos de agrupamiento en R. El objetivo era identificar grupos que representaran distintos perfiles diarios de luz solar, que pudieran corresponder a variaciones en las condiciones climáticas, las estaciones u otros factores.

Una vez obtenidos los grupos de datos, se pueden dividir cada día en sus grupos respectivos para analizar las características conjuntas que los definen. El análisis de las características de cada grupo se lo puede hacer gráficamente con el objetivo de visualizar los diferentes patrones diarios. Para esto se calculó la curva acumulada promedio para cada grupo, lo que proporcionó una forma representativa de cada uno. Posteriormente, estas curvas acumuladas promedio se desagregaron para derivar los

perfiles diarios promedio para cada grupo. Este proceso proporcionó una comprensión detallada de los patrones típicos de luz solar asociados con cada grupo.

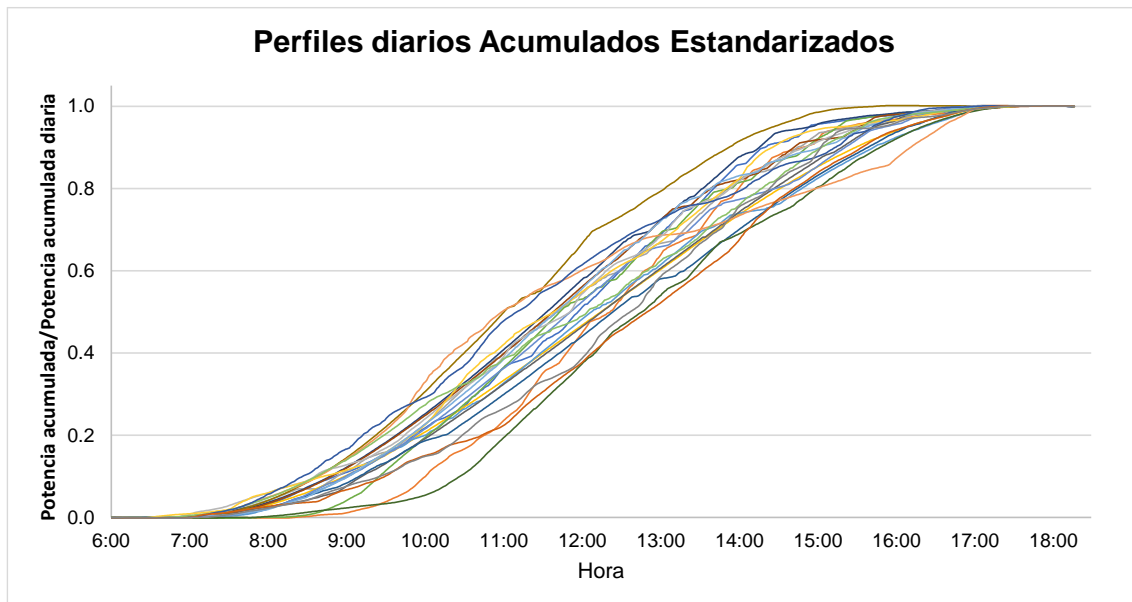


Fig. 7 Acumulados Estandarizados
Fuente: Elaboración propia.

Esta metodología permitió la identificación de patrones característicos en la acumulación diaria de radiación solar, lo que proporcionó información sobre la variabilidad y la previsibilidad de la disponibilidad de energía solar.

3.6 Factor de Planta

El factor de capacidad o factor de planta es una métrica clave para evaluar el rendimiento de los paneles solares, ya que representa la relación entre la producción de energía real en un período específico y la producción máxima posible si los paneles operaran a plena capacidad de forma continua. En este estudio, se calcula el factor de capacidad de cada grupo de paneles solares para cuantificar su eficiencia y comparar su rendimiento a lo largo del tiempo.

Para calcular el factor de planta de cada grupo de paneles solares, se utiliza la siguiente fórmula (3.5):

$$FP = \frac{\text{Producción Total de Energía}}{\text{Capacidad Instalada} \times \text{Período de Tiempo}} \quad (3.5)$$

donde:

- **Producción Total de Energía** se refiere a la energía real producida por los paneles durante el período de estudio.

- **Capacidad Instalada** es la potencia máxima nominal de cada grupo de paneles.

Este cálculo permite normalizar el rendimiento, teniendo en cuenta las diferencias de capacidad entre los grupos de paneles. Al analizar el factor de capacidad en múltiples períodos de tiempo, este estudio evalúa las variaciones en el rendimiento debido a factores como las condiciones climáticas, los cambios estacionales y la posible degradación de los paneles.

Para cada serie temporal, se calculó el factor de capacidad diario. Estos valores del factor de capacidad ofrecen información sobre la consistencia y confiabilidad de la producción de energía solar de cada grupo de paneles, complementando el modelado ARMA y la generación de datos sintéticos discutidos en secciones anteriores.

3.7 Conclusiones del capítulo

En este capítulo, se ha descrito la metodología utilizada para analizar las series temporales de producción de energía solar de tres grupos de paneles solares. Siguiendo un proceso sistemático de preprocesamiento de datos, identificación de modelos ARMA y generación de series temporales sintéticas, este estudio establece una base sólida para comprender los patrones subyacentes en la producción de energía solar. Además, al presentar los percentiles diarios se permite describir de forma consolidada el comportamiento diario de la producción FV. Los pasos seguidos garantizan la reproducibilidad del análisis y la fiabilidad de los datos sintéticos generados, que se utilizarán en los capítulos posteriores para análisis y validaciones adicionales.

Resultados y Discusión

Este capítulo presenta los resultados del análisis de la producción fotovoltaica para cada grupo de paneles solares. El análisis parte de la transformación de la serie de tiempo para remover tendencias, seguido del modelado ARMA de los residuales. El objetivo principal del análisis estacional es detectar patrones temporales subyacentes en los datos y evaluar los modelos a través del diagnóstico de residuales. Mientras que el análisis diario se centra en describir estadísticamente los patrones diarios de producción FV. Los resultados discutidos en este capítulo incluyen las características estadísticas de los modelos ajustados, la validación de los modelos mediante la generación de datos sintéticos y el ajuste de curvas de percentiles diarios.

4.1 Descripción general

Para el análisis de los datos, se buscó calcular valores de potencia acumulados que permitan visualizar patrones anuales y diarios. Los datos son tomados cada minuto y representan valores de potencia. La potencia nominal de los paneles es de 15kW para los dos primeros grupos de paneles y de 5kW para el tercer grupo. Con esto en consideración, las sumas presentan datos de potencia acumulada, es decir, no es un acumulado de energía, sino una sumatoria de potencia. En caso de que se requiera los valores de energía correspondientes es conveniente proceder con lo indicado en la ecuación (3.1).

A la potencia acumulada diaria, se le debe dividir para el número de minutos considerados. En este caso se tomaron 737 minutos diarios (de luz solar), por lo que se utilizan las siguientes ecuaciones donde P_a es la potencia acumulada en kW, P_r es la potencia promedio para la duración de la radiación y se calcula encontrando el promedio de la potencia en ese rango de tiempo con la ecuación (4.1):

$$P_r = \frac{P_a}{737} \quad (4.1)$$

Para encontrar la energía promedio diaria en kWh, se multiplica la potencia promedio por la cantidad de tiempo en horas. La cantidad de horas es la cantidad de minutos considerados para 60 minutos como se muestra en la ecuación (4.2):

$$E_{prom} = P_r * \frac{737}{60} = P_r * 12.28 \quad (4.2)$$

Para el cálculo del factor planta, se debe tomar en cuenta las 24 horas de trabajo a la potencia nominal de 15kW, la energía posible de producción o Energía nominal diaria se considera como la potencia nominal del panel multiplicado por las horas de trabajo como se muestra en la ecuación (4.3). Debido a que se considera un día entero, se tomaran las 24 horas de trabajo.

$$E_{nominal} = P_{nominal} * tiempo = 15 * 24 = 360 \quad (4.3)$$

Por último, el cálculo del factor planta se realiza mediante la ecuación (4.4)

$$FP = \frac{P_r * 12.28}{360} \quad (4.4)$$

4.2 Ajuste estacional (Anual)

El primer aspecto de la modelación de la serie de tiempo consistió en un *detrending* estacional a través de un modelo lineal. Para esto, se empleó conjuntamente la transformación Box-Cox, que resultó significativa sólo para el grupo SFV3, ya que para los demás grupos el valor de lambda es muy cercano a 1, indicando que no es necesario realizar la transformación. En el caso de SFV3 el valor de lambda es de 0.505 por lo que se procede a realizar la transformación usando la ecuación (3.2) mostrada anteriormente.

Al aplicar el procedimiento Box Jenkins para modelar la serie anual mediante un modelo ARMA, (función *auto.arima* en R) se obtuvo residuos semejantes a ruido blanco. Además, se logró simular una serie de tiempo con el mismo comportamiento estadístico que las series de tiempo originales.

4.2.1 Modelo Lineal

El objetivo de este ajuste lineal es encontrar una función periódica, de un año, que ajuste la variación estacional de la serie. Con esto se busca quitar el componente periódico y así desestacionalizar la serie.

Las ecuaciones encontradas para cada grupo de paneles se presentan a continuación, con la variable X correspondiendo al número de día del año como se muestra en la ecuación (4.5):

$$Producción = Intercepto + A * \cos\left(\frac{2\pi}{365} * X\right) + B * \sen\left(\frac{2\pi}{365} * X\right) + C * \cos\left(\frac{2\pi}{182.5} * X\right) \quad (4.5)$$

Tabla VI
Valores para Ecuación de Modelo Lineal

Grupo	Ecuación de modelo lineal
SFV1	$ml = 3210.07 + 219.9 \cos(2\pi/365 * X) - 267.08 \operatorname{sen}(2\pi/365 * X) - 254.03 \cos(2\pi/182.5 * X)$
SFV2	$ml = 3062.4 + 300.11 \cos(2\pi/365 * X) - 157.28 \operatorname{sen}(2\pi/365 * X)$
SFV3	$ml = 1161.5 + 212.32 \cos(2\pi/365 * X) - 129.59 \operatorname{sen}(2\pi/365 * X)$

Fuente: Elaboración propia.

Las curvas obtenidas para el modelo lineal se pueden observar en los gráficos presentados en la Fig. 8. Donde, se tienen los valores acumulados diarios en color negro, y en color rojo se dibuja el modelo lineal que representa el componente estacional de la serie.

Las curvas presentan cierto grado de similitud en su tendencia, con el mínimo valor encontrado entre mayo y junio para todos los grupos y el máximo en diciembre para el grupo de paneles SFV2 y SFV3, mientras que el máximo para SFV1 es en el mes de octubre.

Esta diferencia se puede dar por el término agregado para el primer grupo con frecuencia de medio año. Sin embargo, todos los grupos coinciden con el mínimo entre mayo y junio. Esto podría estar relacionado con la cantidad de lluvias en ese mes que a su vez resulta en menor radiación por la alta presencia de nubes.

Aparte, una vez obtenido el modelo lineal que describe la estacionalidad de los datos se puede calcular nuevamente un promedio, obtenido del modelo lineal, que refleje la tendencia de los datos y no se vea afectado por los datos perdidos o faltantes. El cálculo del promedio para el modelo lineal es simple, puesto que se toma el valor del intercepto. Este valor será el promedio siempre que se considere un período completo, y debido a que se buscó encontrar un modelo para el período anual completo, se puede considerar a este valor como un promedio efectivo. Usando esta técnica se encontró que el promedio general de los datos es menor considerando el modelo lineal.

El primer acercamiento a los datos resulta en un promedio $\bar{\mu}$, este promedio se ve afectado por los datos faltantes. Mientras que, una vez obtenido el modelo lineal, se puede calcular el promedio de este, para encontrar un valor de promedio $\hat{\mu}$ menos influenciado por la ausencia de datos. Los datos perdidos influyen variables globales como promedio total, factor planta y demás, pero, debido a la estacionalidad que se presenta, si la ausencia de datos ocurre en temporadas altas, el promedio presentaría un sesgo negativo. Por lo tanto, al tener los valores de las series simuladas, se asegura que las variables globales calculadas no presenten esos sesgos.

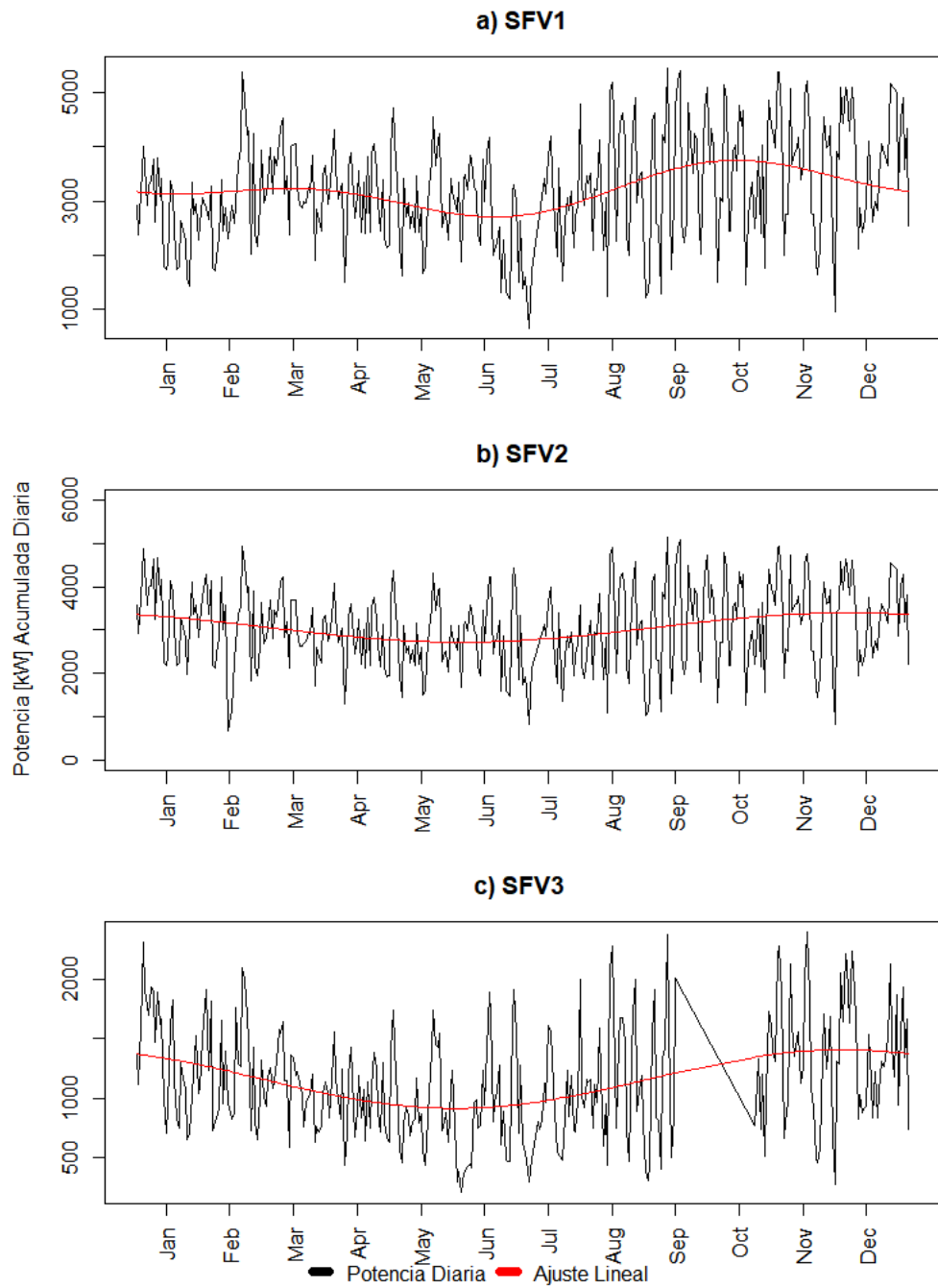


Fig. 8 Valores diarios con modelo lineal
Fuente: Elaboración propia.

4.2.2 Estandarización de los residuos

Una vez desestacionalizada la serie, se continuó con el análisis de los residuos para el modelado adicional. En este punto, se observó que existe una diferencia en la varianza a lo largo del año. Es decir, la varianza observada de los residuos es menor en la primera mitad del año en comparación con la segunda mitad del año. Esta variabilidad en la desviación estándar puede afectar el resultado del modelo ARMA, ya que estos modelos asumen una varianza constante a lo largo del tiempo. Por lo tanto, se pretende estandarizar esta varianza para optimizar el modelo ARMA.

Para esto se realizó un proceso similar al anterior donde se buscó estandarizar la varianza según el método descrito en la sección anterior. El modelo lineal que describe la desviación estándar para una ventana de cada grupo de residuos y permite realizar la estandarización se presenta en la Tabla VII:

Tabla VII
Modelo Lineal de la Desviación Estándar

Grupos	Ecuación de modelo lineal de residuos
SFV1	$ml = 834.61 - 45.43 \cos(2\pi/365 * X) - 242.51 \sen(2\pi/365 * X)$
SFV2	$ml = 828.68 - 19.97 \cos(2\pi/365 * X) - 182.54 \sen(2\pi/365 * X)$
SFV3	$ml = 11.966 - 0.861 \cos(2\pi/365 * X) - 3.4152 \sen(2\pi/365 * X)$

Fuente: Elaboración propia.

Una vez obtenidas estas curvas, se puede dividir cada valor residual para la desviación estándar estimada con el modelo lineal en ese punto, como se explica en la ecuación (4.6), donde $e(t)$ respresenta los residuales originales y $\sigma(t)$ es la desviacion estandar en funcion del tiempo.

$$e_{estandarizados}(t) = \frac{e(t)}{\sigma(t)} \tag{4.6}$$

Lo que se busca con eso es que los datos sean estacionarios en sentido amplio para aplicar el modelo ARMA. El resultado del modelo lineal para la estandarización se presenta en la Fig. 9, donde se evidencian los valores de la desviación estándar a lo largo del año, mostrando mayor variabilidad en la segunda mitad del año. La figura también presenta el modelo lineal ajustado que describe esta variación de forma cíclica/periódica.

Con estos métodos se logró obtener una serie de tiempo estacionaria de residuos, que podrán ser modelados de una mejor manera por el modelo ARMA.

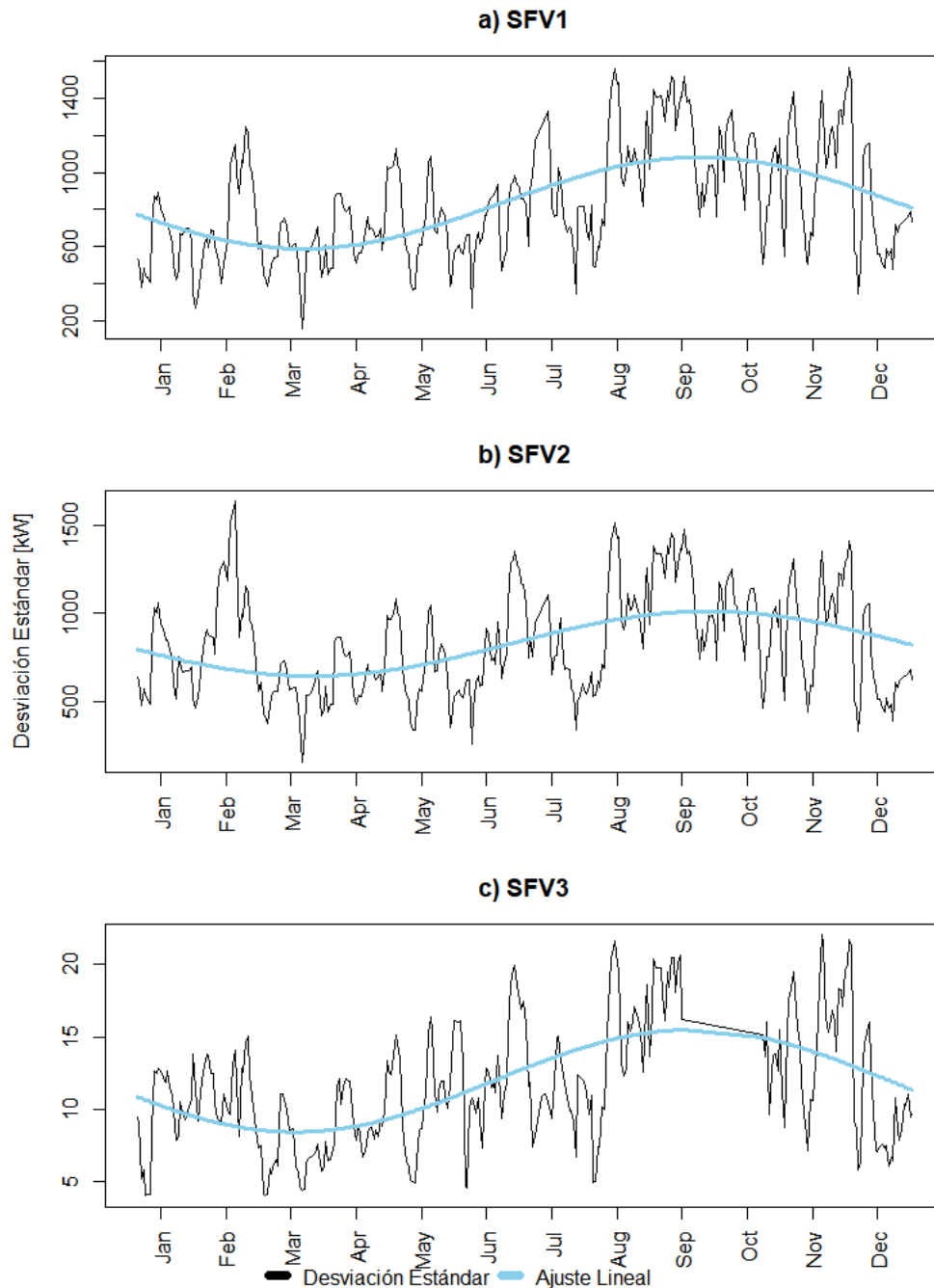


Fig. 9 Desviación estándar de residuales
Fuente: Elaboración propia.

4.2.3 Modelo ARMA

Para identificar los modelos ARMA(p,q) apropiados para las series temporales residuales de los tres grupos de paneles solares, se analizaron la función de

autocorrelación (ACF) y la función de autocorrelación parcial (PACF). Estos gráficos ayudan a determinar el orden de los componentes autorregresivos (AR) y de media móvil (MA), ya que revelan retrasos significativos en la estructura de autocorrelación de los datos representando los valores para el grado del modelo ARMA adecuado en cada caso. Las autocorrelaciones significativas se indican mediante valores que exceden los intervalos de confianza, marcados como líneas discontinuas en los gráficos.

En la Fig. 10 se presentan los gráficos de PACF y ACF para los tres grupos de paneles. Las funciones alcanzan rápidamente los límites de significancia a pesar de su tipo de comportamiento, lo que indica que se van a obtener modelos de orden bajo, pero no resulta fácil determinar exactamente el valor de p y q . Por ejemplo, en ninguno de los casos se tienen un comportamiento claro de PACF mientras que, los tres casos de ACF presentan un comportamiento de disminución gradual. Aun así, debido a que los pulsos caen por debajo de los límites tempranamente, no se tiene una definición clara del tipo de modelo ARMA adecuado. Por lo que, se buscaría ajustar varios modelos buscando encontrar el más adecuado al comparar criterios de información. Cada modelo se puede evaluar con el resultado de la función *arima()* en R, teniendo en cuenta que esta ajusta el modelo ARMA de orden especificado a los datos, para luego comparar los resultados de los diferentes modelos, buscando el mejor. Esto también se puede hacer de forma automática con el uso de funciones en R para automatizar la selección de los parámetros óptimos del modelo ARMA. La función *auto.arima()* evalúa varias combinaciones de órdenes AR y MA en función de criterios de información como el Criterio de Información de Akaike (AIC) o el Criterio de Información Bayesiano (BIC), lo que garantiza que se seleccione el modelo que mejor se ajuste a cada grupo.

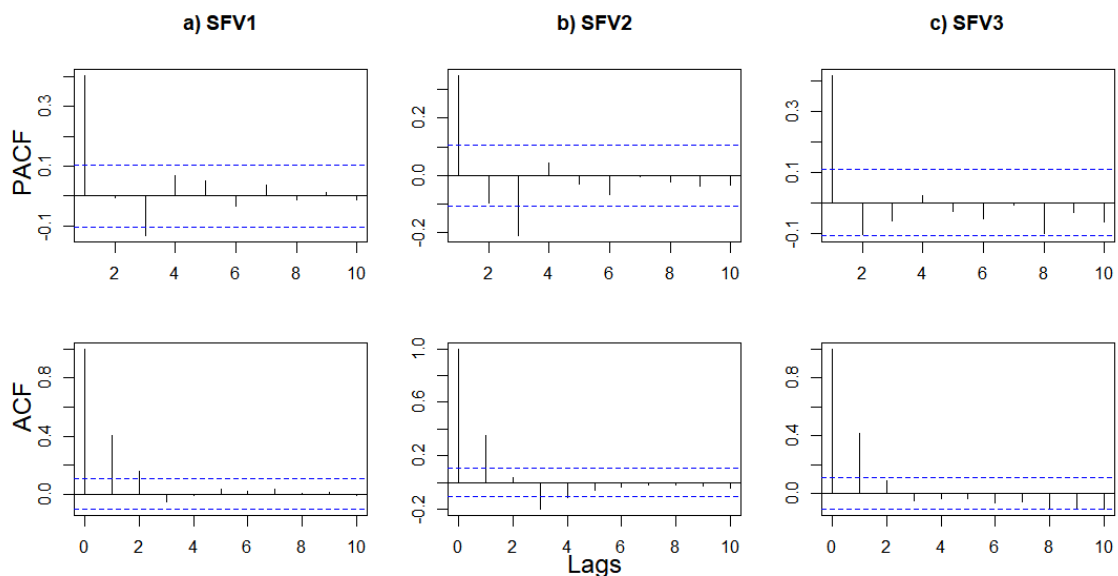


Fig. 10 Funciones de autocorrelación y autocorrelación Parcial
Fuente: Elaboración propia.

Los residuos del primer grupo de paneles solares se modelaron mejor utilizando un proceso ARMA(0,2) resultando en un proceso solamente de media móvil. Los coeficientes para los términos MA se estimaron en 0.4248 y 0.2571 respectivamente. Este modelo sugiere que el valor residual actual está influenciado principalmente por los dos términos de error anteriores, con una influencia decreciente para el segundo rezago. La varianza del error, σ^2 , es de 1.063, lo cual se debe tener en cuenta al momento de generar las series simuladas.

Para el segundo grupo de paneles solares, se identificó un modelo ARMA(2,2) con media distinta de cero como el que mejor se ajusta. Este modelo incorpora dos términos autorregresivos y dos términos de media móvil con promedio 0.0005. Los coeficientes para los términos AR son -0.2356 y -0.3067, mientras que los términos MA tuvieron coeficientes de 0.6071 y 0.5538. Esto indica que los residuos están influenciados tanto por sus valores pasados como por los términos de error pasados, y la interacción entre estos componentes es más compleja que en el primer grupo. La varianza del error, σ^2 tiene un valor estimado de 0.9859. En este ajuste, el valor de la media y del primer coeficiente de la parte autorregresiva (AR1), resultan no ser significativos debido a que estos valores son menores al doble de su error estándar (s.e.) calculado.

El tercer grupo de paneles solares se ajustó a un proceso ARMA(1,0), que indica únicamente un proceso autorregresivo de un solo término. El coeficiente AR se estimó en 0.4195. Esto sugiere que los residuos están determinados principalmente por su valor pasado inmediato, sin influencia significativa de los términos de error pasados. La varianza residual estimada es de 1.037, lo que indica un modelo más simple con una variabilidad inexplicada relativamente baja.

Estos modelos fueron elegidos en función del valor mínimo de AIC y su capacidad para capturar adecuadamente la estructura de autocorrelación observada en los residuos. Las diferencias en las estructuras del modelo ARMA entre los tres grupos de paneles solares reflejan patrones temporales distintos en sus series temporales residuales. El modelo ARMA(0,2) del primer grupo enfatiza las dependencias de error a corto plazo, el modelo ARMA(2,2) del segundo grupo captura un equilibrio entre los residuos rezagados y los errores, y el modelo ARMA(1,0) del tercer grupo sugiere una relación más simple donde los residuos dependen predominantemente de su valor pasado más reciente.

4.2.4 Evaluación del Modelo

El objetivo del modelado era asegurar que los residuales presenten características de ruido blanco, confirmando que los patrones sistemáticos de la serie de tiempo se capturaron adecuadamente. La evaluación del modelo se presenta con los gráficos de la Fig. 11.

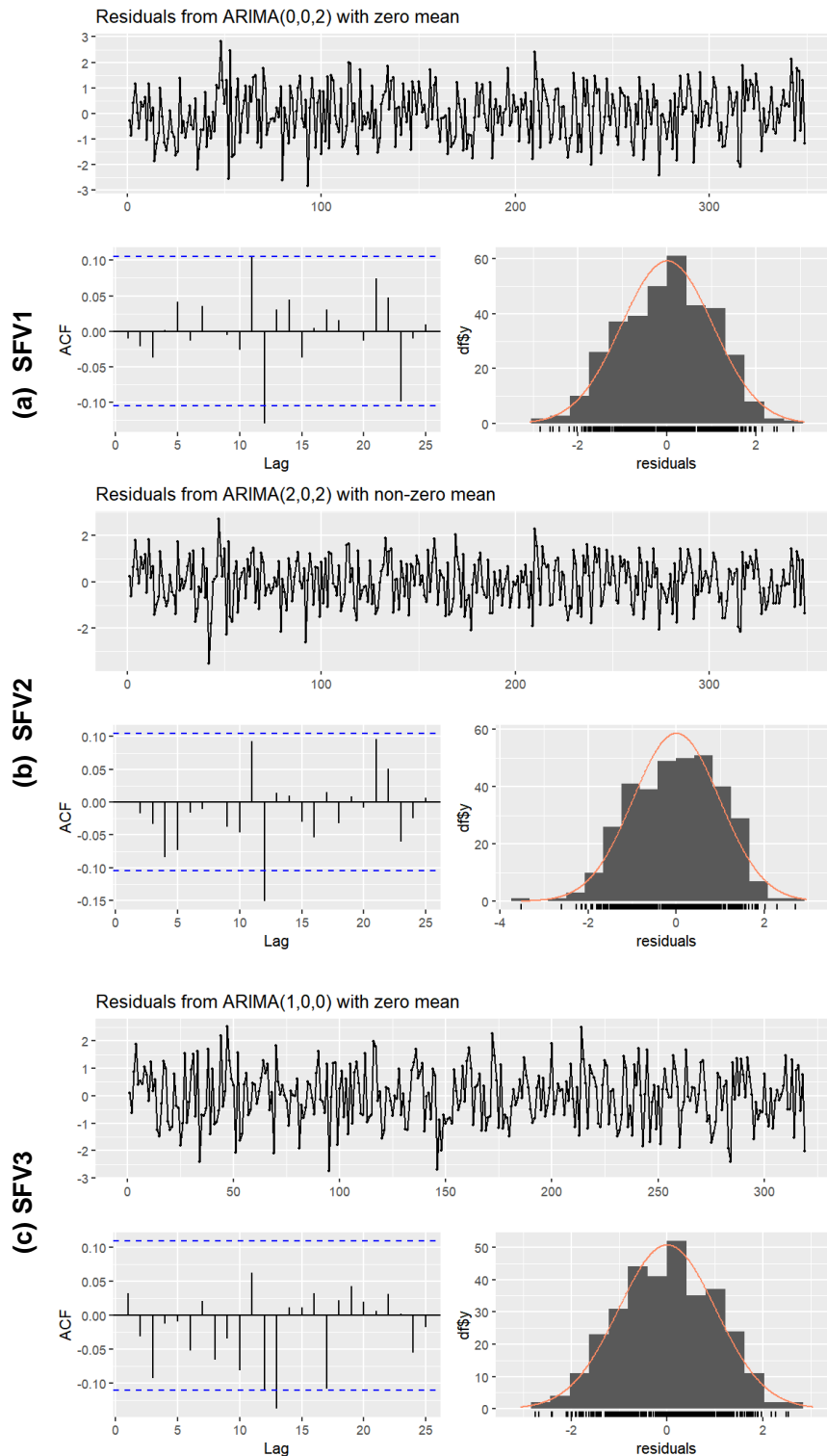


Fig. 11 Evaluación de residuos modelo ARMA (a) SFV1. (b) SFV2. (c) SFV3.
Fuente: Elaboración propia.

Para verificar el ajuste del modelo se grafica los residuos y se comparan con ruido blanco. Si los modelos ARMA fueron especificados apropiadamente con un correcto

modelamiento, los residuales deberían presentar una naturaleza aleatoria y ausencia de autocorrelación. La Fig. 11 muestra la evaluación de los residuales de los modelos ARMA. La característica principal es verificar si se aproximan a ruido blanco, lo que implica que no presenta correlación y tienen varianza constante a lo largo del tiempo.

Los residuales presentan características consistentes con ruido blanco, incluyendo la falta de patrones discernibles y fluctuaciones no correlacionadas. Esto confirma la eficacia del modelo ARMA en capturar las variaciones sistemáticas de la serie de tiempo. Los residuales de los modelados ARMA para cada grupo de paneles se asemejan a ruido blanco, por lo que se comprueba la validez de los modelos como representantes eficaces de las series de tiempo.

Los intervalos de predicción de los modelos ARMA se basan en la suposición de que los residuos no están correlacionados y se distribuyen normalmente. Si alguna de estas suposiciones no se cumple, los intervalos de predicción pueden ser incorrectos. Por este motivo, siempre se deben trazar la ACF y el histograma de los residuos para comprobar las suposiciones antes de generar intervalos de predicción [48].

4.2.5 Generación de series simuladas

La generación de series simuladas se logró siguiendo el proceso inverso al empleado para el modelamiento ARMA, comenzando por la generación de ruido blanco. Las series que se van a generar presentan las mismas características estadísticas, temporales y estacionales de las series de tiempo originales, ya que son creadas y adaptadas con los coeficientes encontrados en cada sección. Las series temporales simuladas permiten evaluar la capacidad de los modelos de replicar la dinámica temporal y la variabilidad de la producción de energía solar observada.

Las series simuladas pueden ser generadas para cualquier intervalo de tiempo o período necesario. En este caso, en la Fig. 12 se presenta la serie generada para el período de un año, comparable con las series originales. En esta figura se puede apreciar que los datos simulados conservaron los patrones temporales clave presentes en las series originales.

La comparación visual entre las series observadas y simuladas (Fig. 8 y Fig. 12) revela que los datos sintéticos capturan las variaciones estacionales características de la producción FV. Esto significa que la generación de series sintéticas ha modelado eficazmente la data original, ya que es capaz de producir nuevos datos que siguen las mismas tendencias y tienen la misma estacionalidad que el original.

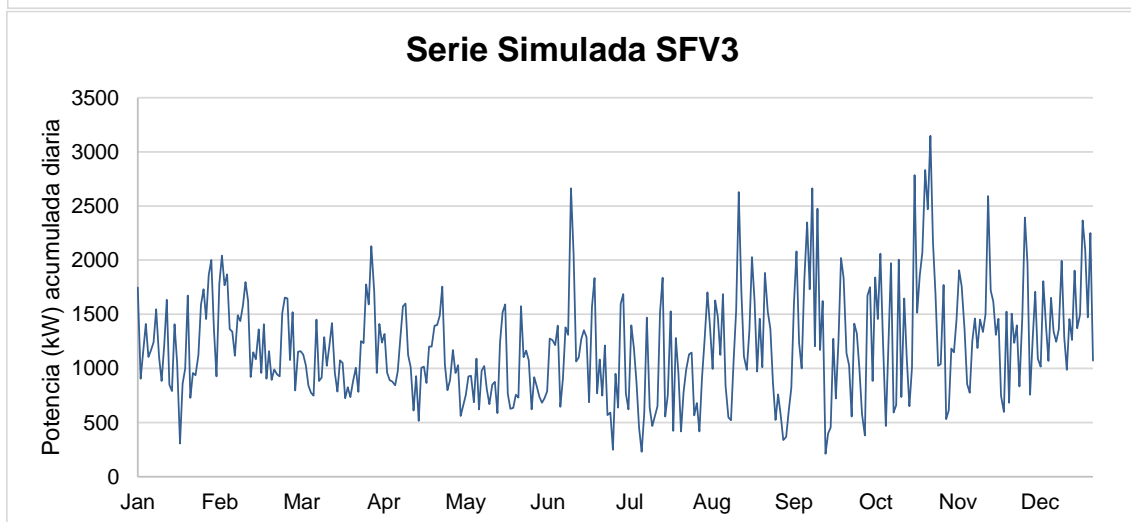
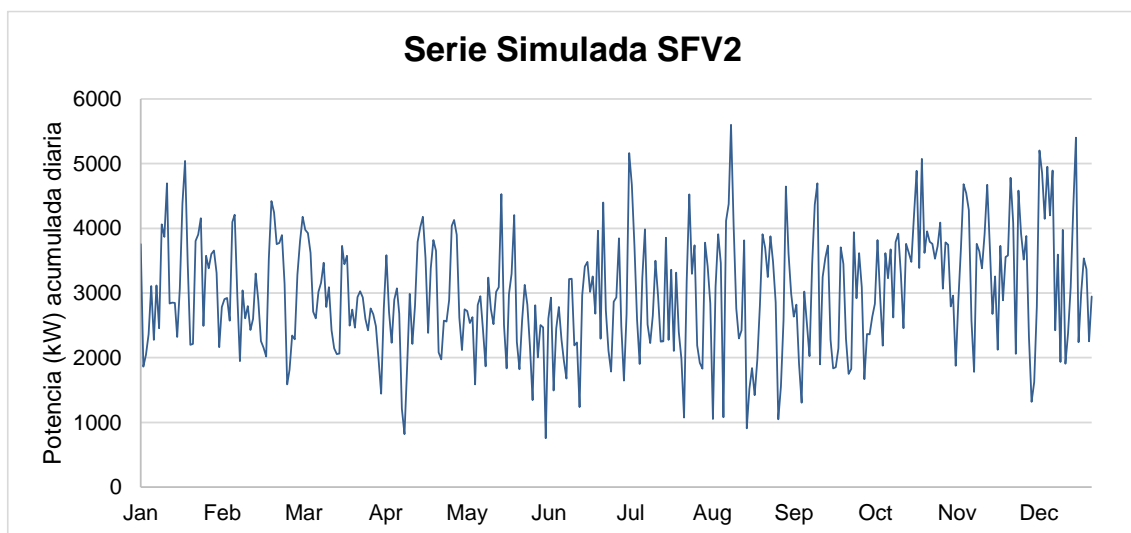
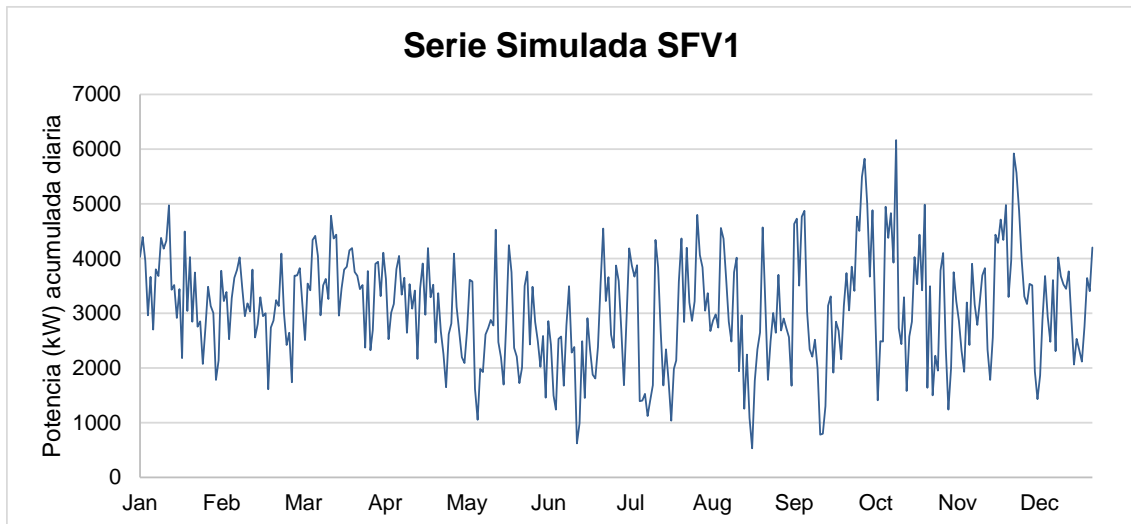


Fig. 12 Series Simuladas
Fuente: Elaboración propia.

Estos resultados destacan el papel fundamental de la serie ARMA simulada para vincular los datos observados con el modelado predictivo, mejorando en última instancia la comprensión y la gestión de los sistemas de energía FV. La utilidad de estas series simuladas se extiende más allá de la validación; también permiten el análisis de escenarios generando perfiles de producción en condiciones variables.

Las series sintéticas y su capacidad para capturar la naturaleza aleatoria de los datos incluyendo sus patrones intrínsecos sirven de base para un campo amplio de aplicaciones. Con base a estas series generadas se pueden crear métodos de imputación de datos con intervalos de confianza simulados, basados en el comportamiento de los datos observados. Además, estas series pueden servir como base para algoritmos de predicción que requieren de un conjunto extenso de datos para entrenamiento, garantizando que las series presentan la información relevante de los datos originales.

La estructura del modelo ARMA es probabilística y puede generar resultados tanto normales como extraordinarios. Al ser estos transformados con el modelo lineal, proveen los resultados finales de potencia acumulada que, en algunas simulaciones, puede superar el valor de potencia nominal máxima del panel o ser menores a cero. Con el fin de restringir estos casos, se descarta la simulación que presente valores fuera de los límites y se genera una nueva serie que cumpla con estas restricciones, o se truncan los valores para que se encuentren dentro de estos.

4.3 Variación Diaria

Para analizar la variabilidad y la distribución de la radiación solar diaria a lo largo del año, se calcularon los percentiles diarios en base a los datos colectados. Estas curvas proporcionan un resumen completo de la distribución de la producción solar fotovoltaica a lo largo del día en diferentes percentiles. Los percentiles proporcionan información de valores típicos, máximos/altos y mínimos/bajos de las condiciones de radiación solar durante el año.

Los percentiles se calcularon utilizando Excel para datos de radiación solar diaria obtenidos en un año. El análisis incluyó percentiles comunes como el 10, el 25, el 50 (mediana), el 75 y el 90, que representan niveles de radiación solar bajos, típicos y altos, respectivamente. Una hipótesis con los percentiles obtenida de la literatura es que el percentil 90 es relacionado con el índice de día despejado. Es decir que el percentil refleja los niveles de radiación en el 10% superior de los días más soleados, por lo que es un indicador de días de alta producción o pico en términos de producción de energía solar.

Para suavizar las curvas de percentiles y reducir la cantidad de parámetros necesarios para describirlas, se ajustó una función beta a los percentiles calculados. Esto permite una representación más eficiente y permite la generación de curvas de percentiles para cualquier valor dado. Los percentiles se calcularon utilizando datos diarios de

potencia acumulada cada 5 minutos y las curvas resultantes se ajustaron a funciones beta estimando los parámetros α , β y k . Estos parámetros se optimizaron para minimizar el error entre las curvas observadas y ajustadas. Un ejemplo de este ajuste se presenta en la Fig. 13 tomando las curvas de los percentiles 75, 50 y 25, del grupo de paneles policristalinos SFV2.

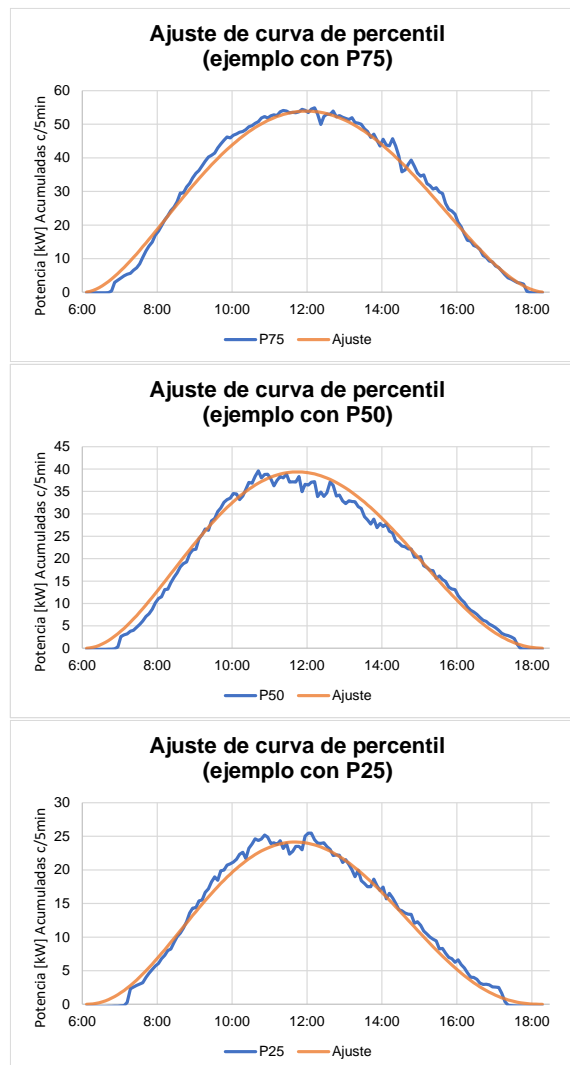


Fig. 13 Ajuste de percentiles de variación horaria
Fuente: Elaboración propia.

Las siguientes figuras presentan las curvas de percentiles diarios (superior) y sus correspondientes curvas beta ajustadas (inferior) para cada grupo de paneles. Las curvas observadas revelan la variabilidad real en la producción, mientras que las curvas ajustadas demuestran cómo una función beta puede aproximar estos patrones.

La Fig. 14 muestra los percentiles para el grupo de paneles monocristalinos y las curvas ajustadas. El ajuste de las funciones beta para los percentiles del SFV1 se logró con dos funciones logarítmicas para α y β y con una beta para k .

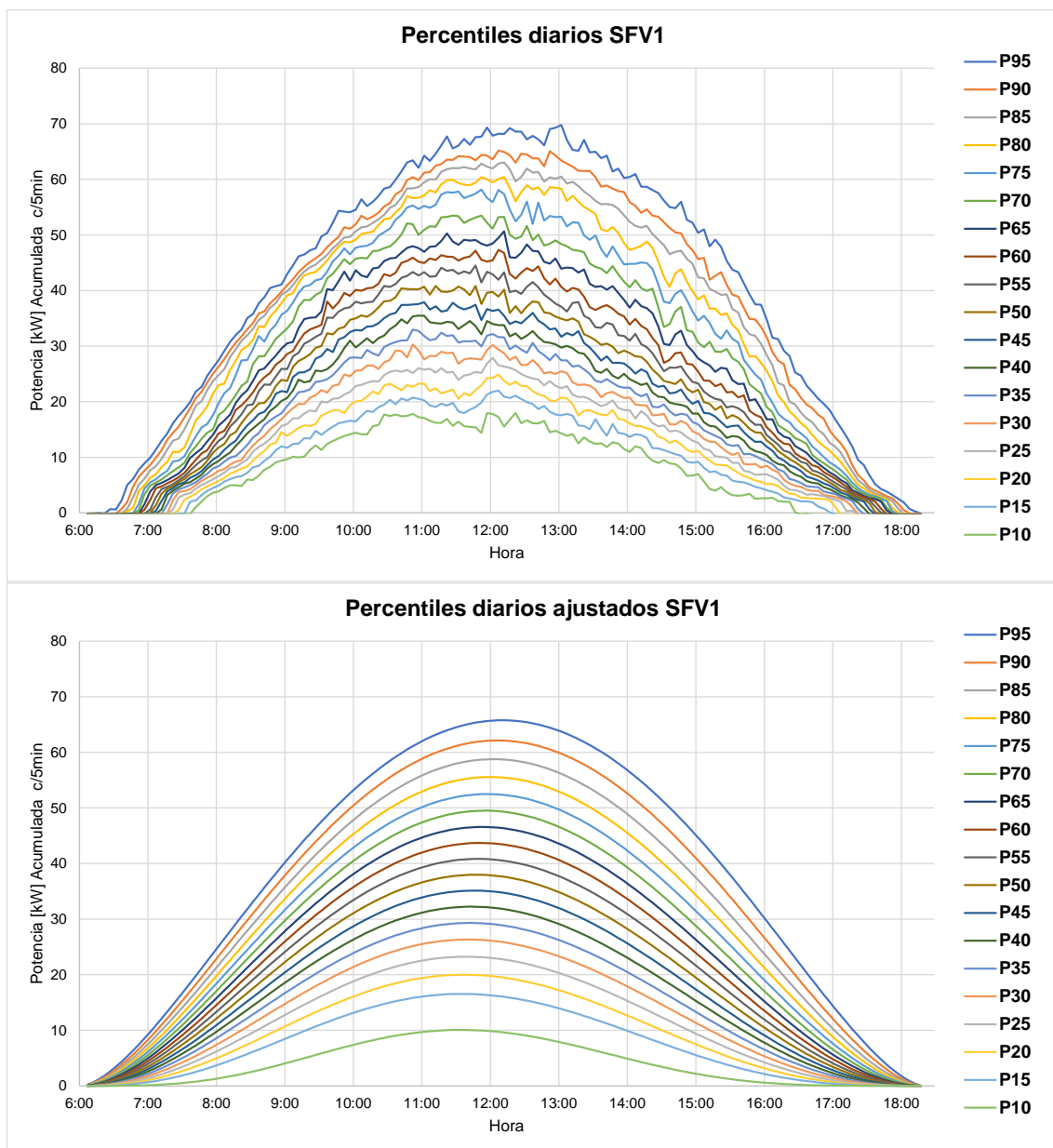


Fig. 14 Percentiles de SFV1
Fuente: Elaboración propia.

Las curvas presentan una distribución aproximadamente uniforme y asimetría ligeramente positiva para los percentiles bajos que se normaliza en los percentiles más altos. Esto indica una leve variación en la producción FV diaria, donde usualmente en las mañanas se captura más radiación solar.

Para el caso de los paneles fijos policristalinos, se presenta la Fig. 15 con sus curvas reales y ajustadas

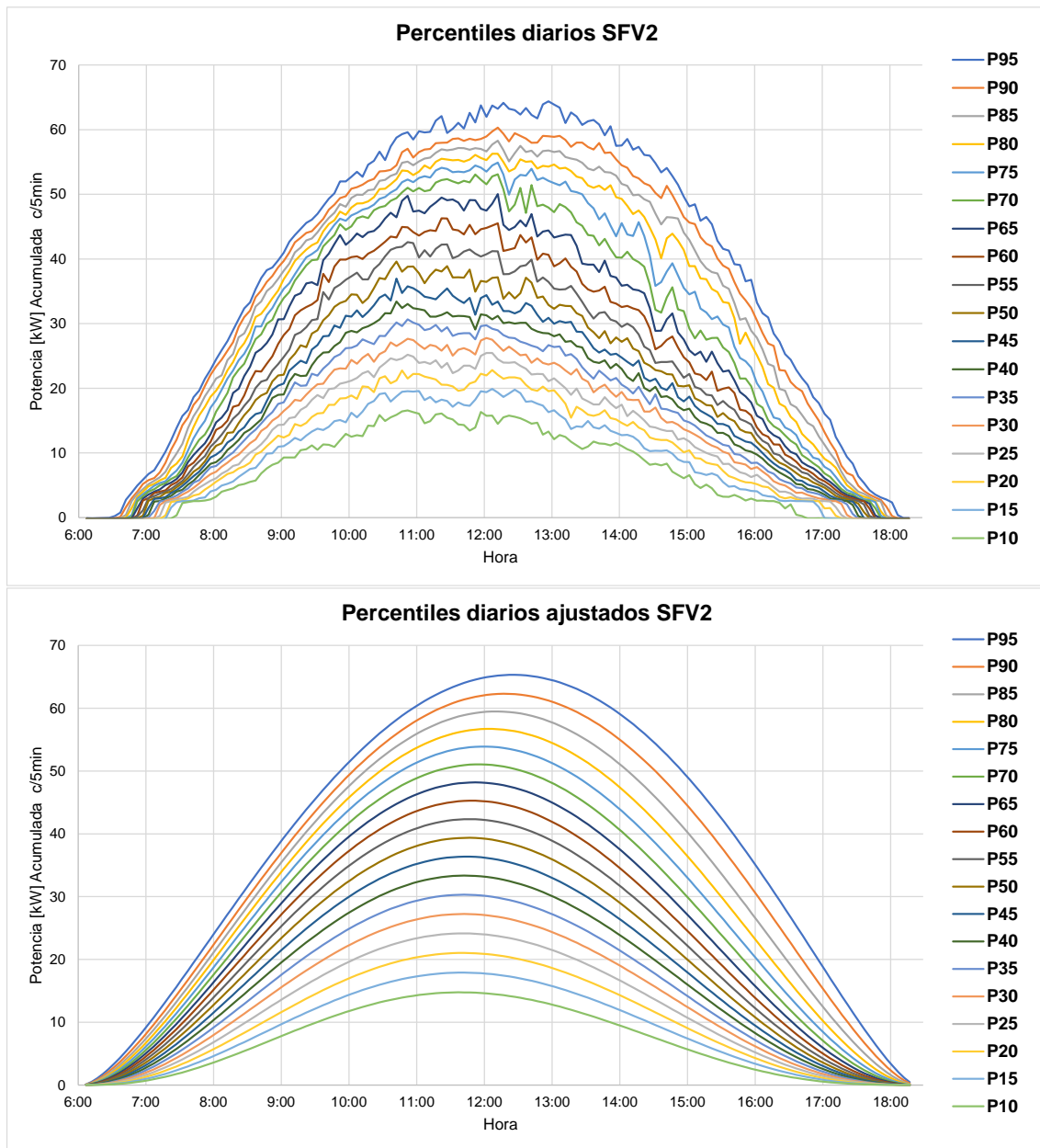


Fig. 15 Percentiles de SFV2
Fuente: Elaboración propia.

El ajuste de las funciones beta para los percentiles del SFV2 se logró con funciones doble beta para α y β y con una beta para k . Al igual que el caso anterior, las curvas presentan cierto grado de asimetría en valores bajos de percentiles que se normalizan con los valores más altos indicando el mismo fenómeno. Sin embargo, en este caso la asimetría parece ligeramente más evidente.

Finalmente, se tiene el caso de los paneles móviles que se presenta en la Fig. 16. Las curvas que se ajustaron en ese caso son 5 parámetros debido a su comportamiento diferente. En este caso para el ajuste se utilizaron funciones doble beta que son las resultantes de una suma de dos betas normales.

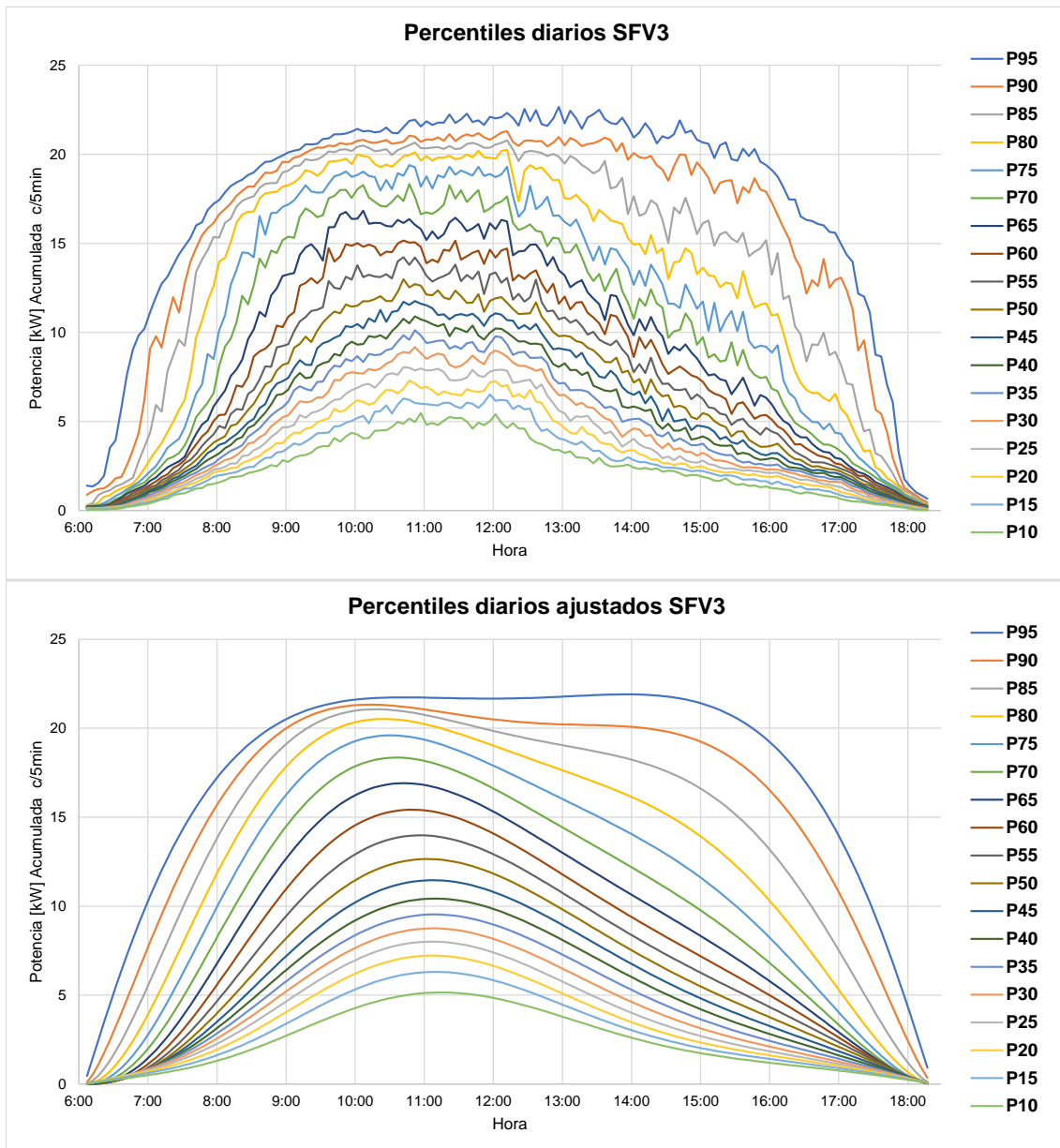


Fig. 16 Percentiles de SFV3
Fuente: Elaboración propia.

Estos gráficos resaltan características y comportamientos clave en los datos con una representación convincente de los fenómenos observados. La asimetría está claramente visible y muestra que la abrumante mayoría de la producción se concentra en la primera mitad del día. Otra característica por notar en esta figura es el

aplanamiento de la curva en los percentiles más altos que sólo resalta el límite de la capacidad de producción de los paneles (5kWh). Sin embargo, gracias a estas curvas se deduce que esto sólo ocurre en el 15% de los casos para las horas de la mañana y únicamente en el 5% de los casos en la tarde.

4.4 Perfiles diarios

El algoritmo de agrupamiento se seleccionó para agrupar los días con formas de curva similares, lo que revela patrones subyacentes en los datos. Usando la función *hclust()* en R, con el método de *Ward* se logró el agrupamiento requerido. Los resultados de la agrupación jerárquica se visualizan claramente en el dendrograma presentado en la Fig. 17.

El dendrograma ilustra las relaciones jerárquicas entre los perfiles de horas de luz solar diaria en función de la forma de sus curvas acumuladas, y por ende, de sus perfiles diarios. Cada rama representa un grupo de días con patrones similares, y la altura de las ramas indica el grado de disimilitud entre los grupos. A su vez, las ramas que se juntan a menor altura indican un mayor grado de similitud o semejanza entre los días.

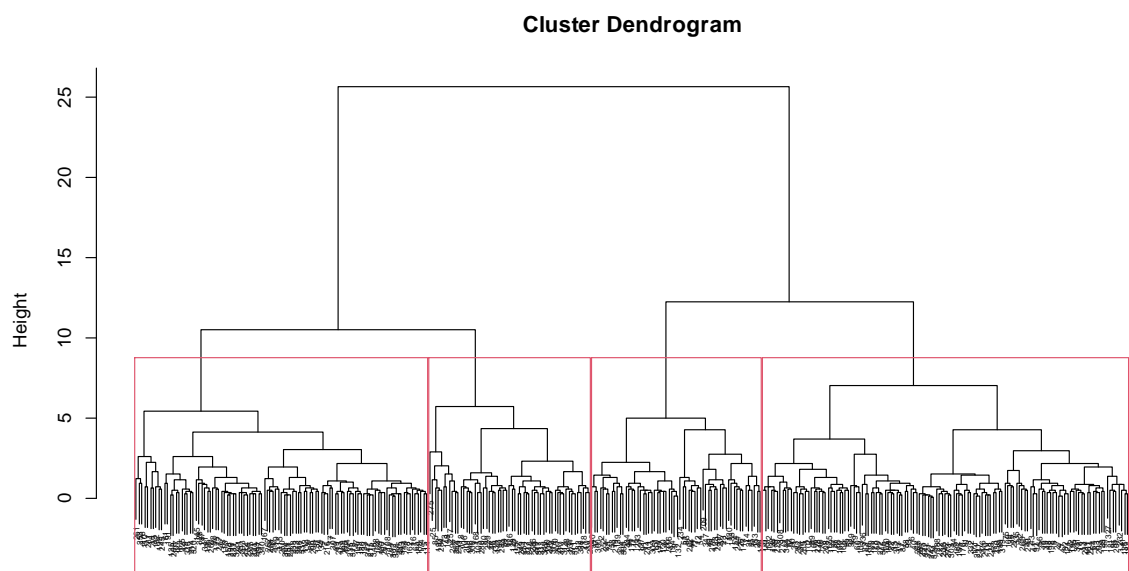


Fig. 17 Dendrograma de agrupamiento según perfil diario
Fuente: Elaboración propia.

El dendrograma se utilizó para determinar el número óptimo de grupos o perfiles diarios. Mediante el análisis de la distancia vertical en la que se fusionan las ramas, se busca una separación adecuada de los posibles grupos. Con base en la estructura observada y la agrupación natural de los datos, se identificaron 4 grupos,

correspondientes a las divisiones en rojo del gráfico. Cada grupo correspondiente a un patrón distinto en las curvas de luz solar acumulada.

Como se puede observar en el gráfico de la Fig. 17, los grupos no son homogéneos y presentan cantidades diferentes de días para cada uno. El primer grupo contiene 122 días, el segundo grupo es el más pequeño y contiene sólo 57 días, mientras que el tercero y cuarto contienen 94 y 76 días respectivamente.

Estos grupos reflejan la variabilidad en los perfiles de luz solar diaria, potencialmente vinculada a cambios estacionales, cobertura de nubes u otros factores meteorológicos. La característica intrínseca de cada grupo se puede determinar mediante el perfil que presenta cada uno. Para determinar estos perfiles diarios, se encontró una curva de potencia representativa al calcular un promedio de cada grupo. Estos promedios se pueden identificar en la Fig. 18, donde se visualizan los diferentes patrones de captura.

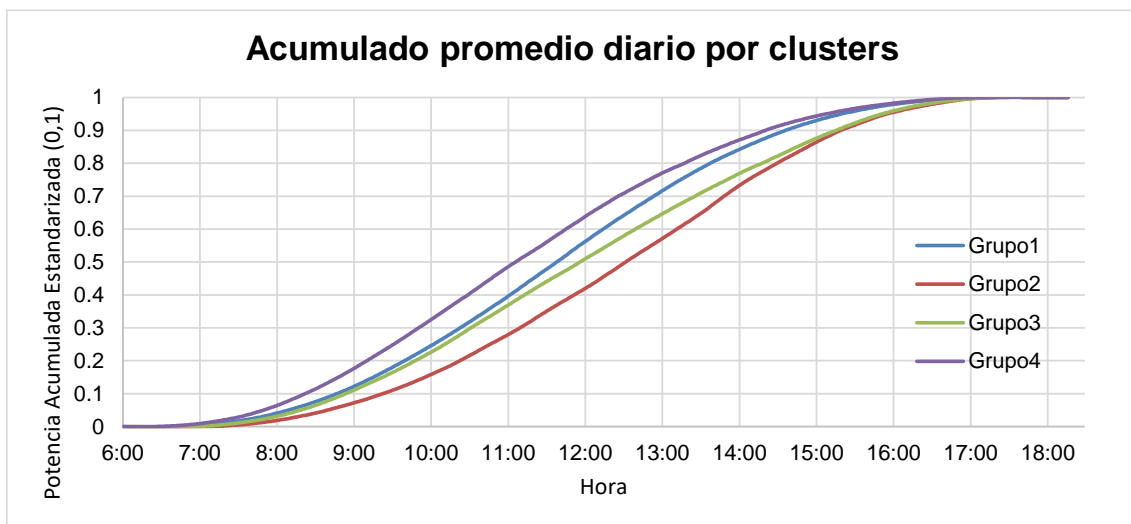


Fig. 18 Curvas de acumulados diarios promedio por grupo
Fuente: Elaboración propia.

Basándose en el gráfico de la Fig. 18 se pueden identificar los diferentes patrones de potencia para cada grupo, donde una pendiente mayor indica mayor captura de radiación solar en esa hora. Las curvas, al tener diferentes pendientes a lo largo del día implican que cada grupo presenta diferente nivel de producción a lo largo del día, implicando que se han separado los días según su perfil de potencia correctamente. Por ejemplo, la curva del grupo 2 parece tener la menor cantidad de producción en la tarde, mientras que la curva del grupo 4 tiene la mayor producción en la mañana.

Sin embargo, para un gráfico más intuitivo, se puede desagregar la serie con el objetivo de facilitar la visualización de los perfiles diarios. Esto se presenta en la Fig. 19, donde se visualizan las curvas promedio correspondientes a los perfiles diarios de cada grupo.

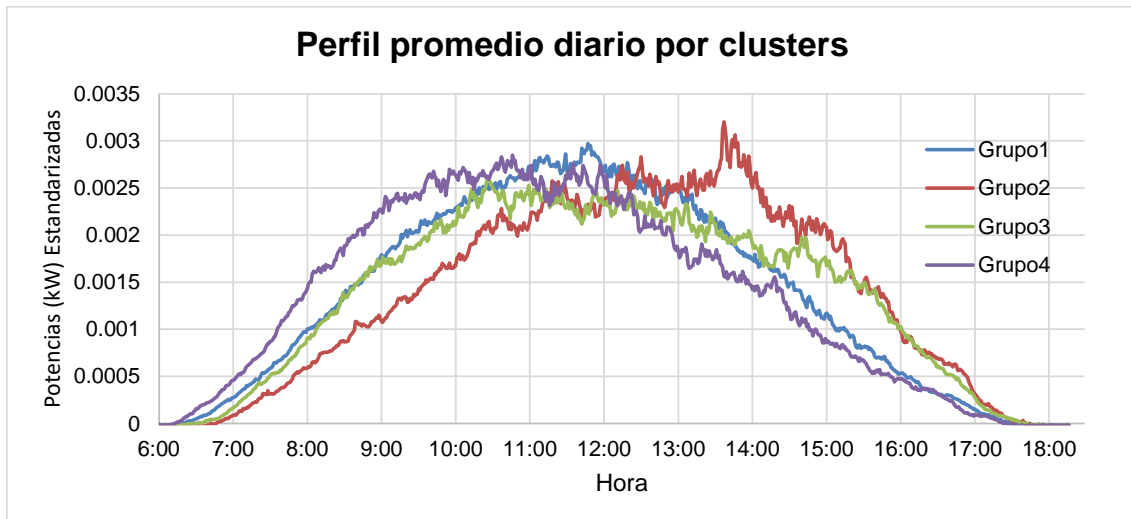


Fig. 19 Perfiles de producción diaria promedio por grupo
Fuente: Elaboración propia.

En la Fig. 19 se puede apreciar de manera más clara lo dicho anteriormente donde la curva del grupo 2 presenta baja producción en la mañana y alta producción en la tarde. Mientras que el grupo 4 presenta lo opuesto, mayor producción en la mañana y baja producción en la tarde. El caso del grupo 1 y grupo 3, presentan producción similar en la mañana, pero a medio día existe más producción por parte del grupo 1, con caída en la tarde, mientras que el grupo 3 se presenta de manera más uniforme sin picos de producción, pero a su vez sin caídas.

La forma promedio de cada uno de los grupos con mayor variación y caída de producción en las tardes, sugiere que el principal factor que influye en este comportamiento es la presencia de nubes y lluvia que se da mayormente en las tardes. Esto, por supuesto, sin dejar de lado factores como la orientación de los paneles fijos y el seguimiento solar de los paneles móviles.

4.5 Factor de Planta

El factor de planta calculado (4.4) para el sistema fotovoltaico durante el año fue del 14.9% para el grupo de paneles SFV1, en el caso del segundo grupo, SFV2, el factor planta es de 14.19% y por último para el grupo SFV3 se tiene un factor de planta de 15.97%. Este último generalmente presenta un valor más alto por la propiedad de seguimiento al sol que mejora su rendimiento.

De la misma manera se obtuvo el factor planta de las series simuladas anteriormente. Los parámetros ARMA derivados de los datos observados se utilizaron para generar 100 realizaciones independientes de la producción anual de energía solar para cada grupo de paneles solares. El factor de planta para cada grupo de paneles solares se

determinó promediando los valores obtenidos en las 100 simulaciones. Esto proporcionó una medida sólida del valor esperado, a diferencia de un valor único que pudo representar una simulación inusualmente alta o baja. El valor encontrado para el factor de planta de las series simuladas para SFV1 es de 14.86%. Para el caso de las series simuladas SFV2, el valor es de 14.2%. Por último, el valor encontrado para las series simuladas de SFV3, es de 16.07%.

Los valores encontrados para el factor planta de las series guardan mucha relación con los valores de las series originales. Corroborando nuevamente la correcta aplicación del modelo ARMA y como este reprodujo eficazmente las características clave de la producción de energía solar. Las discrepancias observadas entre los factores de capacidad de la serie simulada y los datos originales podrían atribuirse al alcance limitado del conjunto de datos observado, que abarca solamente un año. Un conjunto de datos de varios años podría producir un factor de capacidad promedio más representativo, que posiblemente se alinearía más estrechamente con los valores simulados. Esto resalta la necesidad de una recopilación de datos a más largo plazo para refinar la validación del modelo y mejorar la precisión de las predicciones.

4.6 Conclusiones del capítulo

Estos resultados son fundamentales para comprender la radiación solar esperada a lo largo del año y pueden aportar al diseño de sistemas solares fotovoltaicos y a la planificación energética al identificar períodos de alta y baja disponibilidad de energía solar.

La caracterización de los datos a través de modelos ARMA permite comprender el comportamiento de la captura de energía a lo largo del año. Además, con la simulación de las series se abre paso a nuevas investigaciones de monitoreo, predicción y optimización de la planta, basados en el comportamiento estadístico de los datos.

Las curvas beta ajustadas reducen la complejidad de la descripción de los perfiles percentiles diarios, lo que permite generar curvas suaves para cualquier percentil deseado. Este enfoque es particularmente útil para fines de previsión y simulación. La comparación de las curvas observadas y ajustadas indica una fuerte concordancia, como lo demuestran las desviaciones mínimas en todos los percentiles y horas del día. Esto sugiere que las funciones beta capturan eficazmente la variabilidad en la producción de energía fotovoltaica.

La agrupación de los días según su forma o perfil diario indican los diferentes comportamientos de la producción diaria. Las curvas promedio resaltan la variabilidad de la captura diaria específica de la ubicación de los paneles. Esta información revela cómo varía la producción a lo largo de las estaciones al ayudar a diferenciar entre días despejados, nublados y parcialmente nublados según la hora del día, lo que proporciona información sobre cómo el clima afecta la producción. Por otro lado,

también permite identificar problemas con los equipos, problemas de sombreado o anomalías en los datos por las desviaciones de los perfiles esperados.

Conclusiones

Esta investigación proporcionó un análisis exhaustivo de los datos de producción fotovoltaica recopilados durante un año de tres grupos de paneles solares en el laboratorio de Micro-red del campus Balzay de la Universidad de Cuenca. Los objetivos abarcaron la caracterización cuantitativa de la captura de energía solar, la identificación de patrones o grupos diarios y el modelado de la captura de energía para la estimación de la producción. A continuación, se detallan las principales conclusiones del estudio.

El análisis cuantificó con éxito la energía solar capturada por cada grupo de paneles, ofreciendo información detallada sobre su rendimiento. La aplicación de herramientas estadísticas y técnicas de modelado reveló variaciones en la producción debido a cambios estacionales, características específicas del panel y condiciones ambientales. Considerando este comportamiento se modeló efectivamente cada grupo de datos abriendo paso a la optimización de recursos y predicción futura.

Al analizar los perfiles y percentiles diarios, se identificaron patrones distintos en la captura de energía solar. Estos patrones se modelaron utilizando funciones beta, que describieron de manera efectiva la distribución de la producción diaria a lo largo del año. El enfoque de reducción de parámetros agilizó aún más el proceso de caracterización, reduciendo la complejidad sin comprometer la precisión.

El modelado ARMA de los tres grupos de paneles resultó efectivo para demostrar el comportamiento individual de cada grupo de paneles que al realizar una comparación entre ellos se presentan hallazgos interesantes. Los grupos de paneles presentan modelados ARMA diversos, con modelos AR o MA puros cuando se espera que tengan cierto grado de similitud. Esto puede indicar cuestiones interesantes a considerar de los datos tomando en cuenta las diferencias preexistentes entre los paneles (tecnología y funcionamiento).

Por un lado, las variaciones en las propiedades físicas de los paneles podrían generar patrones distintos en la producción de energía. Por otro, incluso pequeñas diferencias en la orientación, el ángulo de inclinación o la alineación con respecto al sol pueden afectar la captura de energía y dar como resultado diferentes características de las series temporales. Si los modelos ARMA difieren significativamente, esto sugiere que las series temporales de producción no son totalmente intercambiables. Esto podría complicar los métodos de imputación que se basan en correlaciones sólidas entre los paneles, lo que pone de relieve la importancia de tener en cuenta estas dinámicas específicas de los paneles y su comportamiento conjunto.

No obstante, la implementación de modelos ARMA en datos de series temporales estandarizados y sin tendencia proporcionó herramientas sólidas para estimar la producción solar. Las series temporales sintéticas generadas a partir de estos modelos demostraron una alta fidelidad a los datos observados, capturando tanto las variaciones a corto plazo como las tendencias a largo plazo. Esto abre paso a múltiples estudios y aplicaciones que se apoyen en las series sintéticas para realizar optimizaciones, imputación de datos y predicciones con intervalos de confianza fiables.

La capacidad de los modelos ARMA para generar series sintéticas realistas subraya su utilidad para simular la producción de energía solar en condiciones similares. Esto tiene implicaciones significativas para predecir la producción de energía solar y desarrollar estrategias basadas en datos para optimizar los sistemas de energía. El trabajo futuro podría centrarse en la incorporación de otras variables, como los datos meteorológicos, para mejorar la precisión del modelo.

En general esta investigación contribuye al campo de la energía renovable al ofrecer herramientas prácticas para analizar y predecir la producción de energía solar. Los hallazgos tienen implicaciones para optimizar el rendimiento de los paneles, mejorar la precisión de los pronósticos y gestionar las brechas de datos en proyectos de monitoreo solar futuros.

Una limitación importante del análisis es la dependencia de un solo año de datos observados para el desarrollo del modelo. La producción de energía solar puede variar significativamente de un año a otro debido a factores como los patrones climáticos, las anomalías estacionales y los cambios en el rendimiento del sistema. Con sólo un año de datos, el factor de planta observado puede no representar completamente el rendimiento promedio a largo plazo de los paneles solares. Por el contrario, los modelos ARMA generan simulaciones basadas en los patrones estadísticos derivados de los datos disponibles, que pueden suavizar las anomalías a corto plazo y aproximarse a una tendencia más general.

El proceso de agrupamiento de los perfiles diarios de producción fotovoltaica se implementó exitosamente, lo que permitió la identificación de patrones de comportamiento distintivos en la generación de energía solar. Al analizar las curvas acumuladas y estandarizarlas, se formaron grupos basados en la forma y variabilidad de los perfiles de producción diaria, no en la cantidad de producción generada. Cada grupo destaca un comportamiento único diario en la respuesta a factores ambientales como la nubosidad o los cambios estacionales. Esta clasificación demuestra la eficacia del agrupamiento para capturar y caracterizar la diversidad de la producción FV diaria, lo que proporciona información valiosa para el modelado energético y la optimización del sistema.

Los perfiles diarios ayudan a refinar los modelos predictivos para la producción FV a corto y largo plazo. Esto abre paso a un análisis de clasificación, donde cada día subsiguiente puede ser clasificado en uno de los grupos según su perfil de producción. A su vez esta información se puede utilizar para simular escenarios futuros, lo que ayuda en la planificación y el análisis de riesgos. Los modelos personalizados para cada clúster mejoran la precisión de las estimaciones de producción, guiando las decisiones sobre el

tamaño del sistema, la capacidad de la batería o la integración con otras fuentes de energía renovable. Por otro lado, la agrupación de perfiles diarios proporciona un marco sólido para imputar datos de producción faltantes según el promedio del grupo o el perfil coincidente más cercano.

Referencias

- [1] K. Mohammadi, S. Shamsirband, A. Kamsin, P. C. Lai, and Z. Mansor, "Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure," *Renew. Sustain. Energy Rev.*, vol. 63, pp. 423–434, Sep. 2016, doi: 10.1016/j.rser.2016.05.065.
- [2] S. Saglam, "Meteorological parameters effects on solar energy power generation," *WSEAS Trans. Circuits Syst.*, vol. 9, Oct. 2010.
- [3] J. Ordóñez, E. Jadraque, J. Alegre, and G. Martínez, "Analysis of the photovoltaic solar energy capacity of residential rooftops in Andalusia (Spain)," *Renew. Sustain. Energy Rev.*, vol. 14, no. 7, pp. 2122–2130, Sep. 2010, doi: 10.1016/j.rser.2010.01.001.
- [4] P. G. V. Sampaio and M. O. A. González, "Photovoltaic solar energy: Conceptual framework," *Renew. Sustain. Energy Rev.*, vol. 74, pp. 590–601, Jul. 2017, doi: 10.1016/j.rser.2017.02.081.
- [5] J. Amusan and E. Otokunefor, "The Effect of Cloud on the Output Performance of a Solar Module," Feb. 2023.
- [6] S. Shedd, B. M. Hodge, A. Florita, and K. Orwig, "Statistical Characterization of Solar Photovoltaic Power Variability at Small Timescales: Preprint".
- [7] Y. Kassem, H. Çamur, and S. M. A. Alhuoti, "Solar Energy Technology for Northern Cyprus: Assessment, Statistical Analysis, and Feasibility Study," *Energies*, vol. 13, no. 4, p. 940, Feb. 2020, doi: 10.3390/en13040940.
- [8] M. Del P. Pablo-Romero and Department of Economic Analysis-Chair of Energy and Environmental Economics-Facultad de CC.Económicas y Empresariales, Ramon y Cajal 1. Seville, 41018, Spain, "Solar Energy: Incentives to Promote PV in EU27," *AIMS Energy*, vol. 1, no. 1, pp. 28–47, 2013, doi: 10.3934/energy.2013.1.28.
- [9] T. Kim, W. Ko, and J. Kim, "Analysis and Impact Evaluation of Missing Data Imputation in Day-ahead PV Generation Forecasting," *Appl. Sci.*, vol. 9, no. 1, Art. no. 1, Jan. 2019, doi: 10.3390/app9010204.
- [10] R. Martinez and E. Forero, "Estimation of energy efficiency in solar photovoltaic panels considering environmental variables," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 437, p. 012008, Oct. 2018, doi: 10.1088/1757-899X/437/1/012008.
- [11] H. Tyagi, A. K. Agarwal, P. R. Chakraborty, and S. Powar, Eds., *Advances in Solar Energy Research*. in Energy, Environment, and Sustainability. Singapore: Springer Singapore, 2019. doi: 10.1007/978-981-13-3302-6.
- [12] "Global Market Outlook for Solar Power 2025-2029 - SolarPower Europe." Accessed: Jun. 04, 2025. [Online]. Available: <https://www.solarpowereurope.org/insights/outlooks/global-market-outlook-for-solar-power-2025-2029/detail>
- [13] G. M. Singh and S. S. Bhatti, "Statistical comparison of global and diffuse solar radiation correlations," *Energy Convers. Manag.*, vol. 30, no. 2, pp. 155–161, Jan. 1990, doi: 10.1016/0196-8904(90)90027-V.
- [14] M. Gastón-Romeo, T. Leon, F. Mallor, and L. Ramírez-Santigosa, "A Morphological Clustering Method for daily solar radiation curves," *Sol. Energy*, vol. 85, no. 9, pp. 1824–1836, Sep. 2011, doi: 10.1016/j.solener.2011.04.023.

- [15] R. E. Carrillo, M. Leblanc, B. Schubnel, R. Langou, C. Topfel, and P.-J. Alet, "High-Resolution PV Forecasting from Imperfect Data: A Graph-Based Solution," *Energies*, vol. 13, no. 21, Art. no. 21, Jan. 2020, doi: 10.3390/en13215763.
- [16] I. de-Paz-Centeno, M. T. García-Ordás, Ó. García-Olalla, and H. Alaiz-Moretón, "Imputation of missing measurements in PV production data within constrained environments," *Expert Syst. Appl.*, vol. 217, p. 119510, May 2023, doi: 10.1016/j.eswa.2023.119510.
- [17] B. P. M. Laevens, O. ten Bosch, F. P. Pijpers, and W. G. J. H. M. van Sark, "An observational method for determining daily and regional photovoltaic solar energy statistics," *Sol. Energy*, vol. 228, pp. 12–26, Nov. 2021, doi: 10.1016/j.solener.2021.08.077.
- [18] J. P. Muñoz-Vizhñay, M. V. Rojas-Moncayo, and C. R. Barreto-Calle, "Incentivo a la generación distribuida en el Ecuador," *Ingenius*, no. 19, pp. 60–68, Jan. 2018, doi: 10.17163/ings.n19.2018.06.
- [19] G. S. Inca Yajamín, D. F. Cabrera Carrión, D. F. Villalta Gualán, R. C. Bautista Zurita, and H. D. Cabrera Carrion, "Evaluación de la actualidad de los sistemas fotovoltaicos en Ecuador: avances, desafíos y perspectivas," *Cienc. Lat. Rev. Científica Multidiscip.*, vol. 7, no. 3, pp. 9493–9509, Jul. 2023, doi: 10.37811/cl_rcm.v7i3.6835.
- [20] M. Farghali *et al.*, "Strategies to save energy in the context of the energy crisis: a review," *Environ. Chem. Lett.*, vol. 21, no. 4, pp. 2003–2039, Aug. 2023, doi: 10.1007/s10311-023-01591-5.
- [21] G. E. A. P. A. Batista and M. C. Monard, "An analysis of four missing data treatment methods for supervised learning," *Appl. Artif. Intell.*, vol. 17, no. 5–6, pp. 519–533, May 2003, doi: 10.1080/713827181.
- [22] V. Bădescu, *Modeling solar radiation at the earth's surface: recent advances*. Heidelberg: Springer, 2008.
- [23] J. Boland, "Time Series Modelling of Solar Radiation," in *Modeling Solar Radiation at the Earth's Surface: Recent Advances*, V. Bădescu, Ed., Berlin, Heidelberg: Springer, 2008, pp. 283–312. doi: 10.1007/978-3-540-77455-6_11.
- [24] S. Vergura, G. Acciani, V. Amoroso, G. E. Patrono, and F. Vacca, "Descriptive and Inferential Statistics for Supervising and Monitoring the Operation of PV Plants," *IEEE Trans. Ind. Electron.*, vol. 56, no. 11, pp. 4456–4464, Nov. 2009, doi: 10.1109/TIE.2008.927404.
- [25] S. Pashiardis, S. A. Kalogirou, and A. Pelengaris, "Statistical analysis for the characterization of solar energy utilization and inter-comparison of solar radiation at two sites in Cyprus," *Appl. Energy*, vol. 190, pp. 1138–1158, Mar. 2017, doi: 10.1016/j.apenergy.2017.01.018.
- [26] A. Bocca, L. Bottaccioli, E. Chiavazzo, M. Fasano, A. Macii, and P. Asinari, "Estimating photovoltaic energy potential from a minimal set of randomly sampled data," *Renew. Energy*, vol. 97, pp. 457–467, Nov. 2016, doi: 10.1016/j.renene.2016.06.001.
- [27] J. Ekström, M. Koivisto, J. Millar, I. Mellin, and M. Lehtonen, "A statistical approach for hourly photovoltaic power generation modeling with generation locations without measured data," *Sol. Energy*, vol. 132, pp. 173–187, Jul. 2016, doi: 10.1016/j.solener.2016.02.055.
- [28] G. Oturanc, A. Hepbasli, and A. Genc, "Statistical Analysis of Solar Radiation Data," *Energy Sources*, vol. 25, no. 11, pp. 1089–1097, Nov. 2003, doi: 10.1080/00908310390233531.
- [29] A. F. Almarshoud, "Performance of solar resources in Saudi Arabia," *Renew. Sustain. Energy Rev.*, vol. 66, pp. 694–701, Dec. 2016, doi: 10.1016/j.rser.2016.08.040.
- [30] A. Arabali, M. Ghofrani, M. Etezadi-Amoli, and M. S. Fadali, "Stochastic Performance Assessment and Sizing for a Hybrid Power System of Solar/Wind/Energy Storage," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 363–371, Apr. 2014, doi: 10.1109/TSTE.2013.2288083.

- [31] X. G. Agoua, R. Girard, and G. Kariniotakis, "Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 538–546, Apr. 2018, doi: 10.1109/TSTE.2017.2747765.
- [32] M. Fliess, C. Join, and C. Voyant, "Prediction bands for solar energy: New short-term time series forecasting techniques," *Sol. Energy*, vol. 166, pp. 519–528, May 2018, doi: 10.1016/j.solener.2018.03.049.
- [33] B. Singh and D. Pozo, "A Guide to Solar Power Forecasting using ARMA Models," in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, Bucharest, Romania: IEEE, Sep. 2019, pp. 1–4. doi: 10.1109/ISGTEurope.2019.8905430.
- [34] A. Mbaye, "ARMA model for short-term forecasting of solar potential: application to a horizontal surface of Dakar site.," *OAJ Mater. Devices*, vol. 4, no. 1, Art. no. 1, May 2019, Accessed: May 20, 2024. [Online]. Available: <https://caip.co-ac.com/index.php/materialsanddevices/article/view/75>
- [35] J. F. Duran and L. I. Minchala, "A Comparative Study on Time Series Prediction of Photovoltaic-Power Production Through Classic Statistical Techniques and Short-Term Memory Networks," in *2023 9th International Conference on Control, Decision and Information Technologies (CoDIT)*, Rome, Italy: IEEE, Jul. 2023, pp. 1512–1517. doi: 10.1109/CoDIT58514.2023.10284303.
- [36] L. Martín, L. F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, and M. Cony, "Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning," *Sol. Energy*, vol. 84, no. 10, pp. 1772–1781, Oct. 2010, doi: 10.1016/j.solener.2010.07.002.
- [37] M. David, F. Ramahatana, P. J. Trombe, and P. Lauret, "Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models," *Sol. Energy*, vol. 133, pp. 55–72, Aug. 2016, doi: 10.1016/j.solener.2016.03.064.
- [38] T. Cabello-López, M. Carranza-García, J. C. Riquelme, and J. García-Gutiérrez, "Forecasting solar energy production in Spain: A comparison of univariate and multivariate models at the national level," *Appl. Energy*, vol. 350, p. 121645, Nov. 2023, doi: 10.1016/j.apenergy.2023.121645.
- [39] U. Das *et al.*, "Forecasting of Photovoltaic Power Generation and Model Optimization," *Renew. Sustain. Energy Rev.*, Jan. 2017.
- [40] Y. Huang, J. Lu, C. Liu, X. Xu, W. Wang, and X. Zhou, "Comparative study of power forecasting methods for PV stations," in *2010 International Conference on Power System Technology*, Oct. 2010, pp. 1–6. doi: 10.1109/POWERCON.2010.5666688.
- [41] E. Kardakos, M. Alexiadis, S. Vagropoulos, C. Simoglou, P. Biskas, and A. Bakirtzis, *Application of Time Series and Artificial Neural Network Models in Short-term Forecasting of PV Power Generation*. 2013. doi: 10.1109/UPEC.2013.6714975.
- [42] H. Demirhan and Z. Renwick, "Missing value imputation for short to mid-term horizontal solar irradiance data," *Appl. Energy*, vol. 225, pp. 998–1012, Sep. 2018, doi: 10.1016/j.apenergy.2018.05.054.
- [43] A. Sundarajan and A. I. Sarwat, "Evaluation of Missing Data Imputation Methods for an Enhanced Distributed PV Generation Prediction," in *Proceedings of the Future Technologies Conference (FTC) 2019*, K. Arai, R. Bhatia, and S. Kapoor, Eds., in *Advances in Intelligent Systems and Computing*. Cham: Springer International Publishing, 2020, pp. 590–609. doi: 10.1007/978-3-030-32520-6_43.
- [44] I. S. Iwueze, E. C. Nwogu, V. U. Nlebedim, U. I. Nwosu, and U. E. Chinyem, "Comparison of Methods of Estimating Missing Values in Time Series," *Open J. Stat.*, vol. 8, no. 2, Art. no. 2, Mar. 2018, doi: 10.4236/ojs.2018.82025.
- [45] W. Zhang, Y. Luo, Y. Zhang, and D. Srinivasan, "SolarGAN: Multivariate Solar Data Imputation Using Generative Adversarial Network," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 743–746, Jan. 2021, doi: 10.1109/TSTE.2020.3004751.
- [46] N. US Department of Commerce, "ESRL Global Monitoring Laboratory - Global Radiation and Aerosols." Accessed: Aug. 02, 2024. [Online]. Available: <https://gml.noaa.gov/grad/solcalc/calcdetails.html>

- [47] J. Osborne, "Improving your data transformations: Applying the Box-Cox transformation," *Pract. Assess. Res. Eval.*, vol. 15, no. 1, Art. no. 1, Jan. 2010, doi: 10.7275/qbpc-gk17.
- [48] 8.7 ARIMA modelling in R | *Forecasting: Principles and Practice (2nd ed)*. Accessed: Jul. 26, 2024. [Online]. Available: <https://otexts.com/fpp2/arma-r.html>
- [49] A. Aue, *Time Series Analysis*. LibreTexts, 2021.
- [50] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*, Fourth edition. in Wiley online library. Hoboken, N.J: J. Wiley & Sons, 2008. doi: 10.1002/9781118619193.
- [51] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*, 5th ed. Hoboken, NJ: Wiley, 2011.