

UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# Forecasting global solar radiation in Nariño

**Laura Sofía Hoyos Gómez**

Universidad Nacional de Colombia  
Facultad de Ingeniería y Arquitectura  
Departamento de eléctrica, electrónica y computación  
Manizales, Colombia  
2021

# **Forecasting global solar radiation in Nariño**

**Laura Sofía Hoyos Gómez**

Trabajo de grado presentado como requisito parcial para optar al título de:  
**Doctora en ingeniería**

Advisor:  
(Ph.D., Profesora) Belizza Janet Ruiz Mendoza

Línea de Investigación:  
Meteorología energética, energía solar  
Grupo de Investigación en Potencia, Energía y Mercados - GIPEM

Universidad Nacional de Colombia  
Facultad de Ingeniería y Arquitectura  
Departamento de eléctrica, electrónica y computación  
Manizales, Colombia  
2021

*To my family*

## **Agradecimientos**

La Gobernación de Nariño a través de un crédito-beca administrado por la Fundación Centro de Estudios Básicos y Aplicados - Ceiba financió esta tesis doctoral. La visita de investigación a la Universidad de Chile fue financiada con una beca de la Alianza del Pacífico. Además, el trabajo de campo realizado en la localidad de Huatacondo fue financiado por el Centro de Energía de la Universidad de Chile. La información meteorológica usada en esta investigación fue obtenida en el marco del proyecto complementariedad de fuentes de energía renovable no convencionales en Colombia.

Finalmente, agradezco al Grupo de Investigación en Potencia, Energía y Mercados - GIPEM por el apoyo brindado para el desarrollo de mi trabajo, especialmente a mi directora la profesora Belizza Janet Ruiz Mendoza.

## Resumen

**Artículo [I]** El análisis conjunto de proyectos y comunidades permite abordar proyectos energéticos desde un enfoque sociotécnico. Incluirá las comunidades en el desarrollo de proyectos requiere una mejor comprensión de sus Asuntos Sociales, económicos, ambientales y culturales. En consecuencia, este trabajo se centra en Formular una metodología para priorizar proyectos energéticos, considerando la participación de la comunidad. En la evaluación de las comunidades se consideró los Índices de Desarrollo Humano y los Índices de Objetivos de Desarrollo Sustentable y ajustó estos de acuerdo con la información disponible para Nariño, la evaluación de los proyectos se hizo con el Valor Presente Neto. Para la evaluación conjunta de las comunidades y los proyectos se utilizó la metodología de análisis de proceso jerárquico y se estableció seis objetivos de priorización. Además, la metodología de co-construcción fue el soporte para formular lineamientos para trabajo comunitario. Esta investigación encontró una relación cercana entre los proyectos que contribuyen a mejorar la calidad de vida y la educación en Nariño.

**Artículo [II]** La irradiancia solar es un recurso ampliamente disponible que podría contribuir al proceso de electrificación en lugares con bajos índices socio económicos. No obstante, en algunos lugares, la información de este recurso no está disponible o tiene baja calidad. Para superar este problema algunos investigadores han desarrollado técnicas como modelos empíricos basados en temperatura para estimar el recurso. Sin embargo, no hay un amplio análisis del comportamiento de esas técnicas en ambientes tropicales y montañosos. Por lo tanto, esta investigación analiza el comportamiento de tres modelos empíricos basados en temperatura y un modelo propuesto bajo estas condiciones ambientales. Los errores estadísticos calculados permiten elegir el mejor modelo para cada punto evaluado. Con este modelo se hace la imputación de datos con el fin de incrementar la calidad de las bases de datos analizadas. El modelo propuesto se ajusta mejor a la zona Andina y amazónica, mientras el modelo de Hargreaves y Samani tiene mejores resultados en la zona Pacífica. Además, el modelo propuesto presenta una relación lineal entre las constantes empíricas y la altitud de las estaciones meteorológicas localizadas por encima de los 2.500 msnm.

**Artículo [III]** Los mapas que muestran el potencial de la energía solar facilitan el uso del recurso solar. Sin embargo, la falta de información de irradiancia solar son una barrera para elaborar este tipo de herramientas. Este investigación propone estimar la irradiancia global solar con datos de temperatura usando el modelo empírico de Hargreaves y Samaani y uno propuesto, para incrementar el número de puntos muestreados. Además, se implementa la técnica de validación cruzada conocida como dejar uno por fuera para evaluar el rendimiento de cuatro técnicas de interpolación espacial en un ambiente tropical y montañoso. La información usada es del departamento de Nariño-Colombia que tiene un área de  $33.268 \text{ km}^2$ . El modelo propuesto muestra un mejor comportamiento en sitios localizado a más de 2.500 msnm, ubicados en la zona Andina y Amazonica. Además, Kriging ordinario es la mejor técnica

de interpolación espacial.

**Artículo [IV]** Los modelos de pronóstico de irradiancia solar impulsan las aplicaciones que usan energía solar. Existen varias técnicas para pronosticar la irradiancia solar global, como las numéricas y las estadísticas. En este contexto, esta investigación compara cuatro enfoques de pronóstico estadístico: Promedio móvil integrado autorregresivo (ARIMA), Perceptrón de una sola capa (SLP), Perceptrón de múltiples capas (MLP) y memoria a corto plazo (LSTM), en un horizonte de un día por delante, utilizando conjuntos de datos incompletos medidos en un entorno tropical y montañoso. Los resultados muestran que los modelos basados en redes neuronales superan al modelo ARIMA. Además, LSTM tiene un mejor rendimiento con un número reducido de datos de entrada y en entornos de nubosidad.

**Palabras clave:** Participación comunitaria, Electrificación rural, Proceso analítico jerárquico, Proyectos energéticos, Índice de desarrollo humano, Índice de metas de desarrollo sostenible, modelos basados en temperatura, iomputación d datos, Hargreaves y Samani, Técnicas de interpolación espacial, Mapeo de radiación solar.

# Abstract

**Article [I]** Introducing the community to technical projects requires a deal with the social, energy and environmental policies as well as the cultural field. To address an energy project from a socio-technical view requires the joint analysis of both the project and the community. This work focuses on the formulation of a methodology to ease the prioritization of projects and community participation. To evaluate the community, the Human Development Index and Sustainable Development Goal Index are adjusted to the context and available information of Nariño. The Net Present Value is used for the project evaluation. The Analytic Hierarchy Process allows for the evaluation of the community and project jointly and establishing prioritization objectives. Moreover, the co-construction methodology is the basis to formulate guidelines to work with the community. This research found that there is a relationship between the projects that seek to improve the quality of the life and education in Nariño.

**Article [II]** Solar irradiance is a widely available resource that could drive electrification processes in regions with low socio-economic indexes. However, in some places, solar irradiance resource information is not available or has low quality. As a result, some techniques allow us to estimate this resource from other meteorological variables' information. Nevertheless, there is not a broad analysis of these techniques in tropical and mountainous environments. Therefore, this research analyzes the performance of three well known, and one proposed empirical temperature-based models in a tropical and mountainous environment. Statistical error comparison allows us to choose the best model for each location and the model for the data imputation. Hargreaves and Samani model present better results in the Pacific zone. In the Andean and Amazon zone, the proposed model fit better. Also, the proposed model's empirical constants show a linear relationship with the altitude in AWS above 2.500 MASL.

**Article [III]** The solar energy potential maps are an enabler for solar energy use. However, the lack of solar irradiance information is a barrier to elaborating on this type of decision tool. This research proposed the estimation of solar irradiance using air temperature data to increase the sampled points with the Hargreaves and Samani and a proposed empirical model. Also, the leave-one-out cross-validation is the technique used to assess the performance of four spatial interpolation techniques in a tropical and mountainous environment. The information came from Nariño state in Colombian that covers an area of  $33.268 km^2$ . The proposed empirical model shows better performance in sites with an altitude above 2.500 MASL, located in the Andean and Amazon zone. Further, Ordinary Kriging was the interpolation technique with the best behavior.

**Article[IV]** Accurate mechanisms for forecasting solar irradiance boost solar energy applications. There are several techniques to forecast global solar irradiance, such as numerical weather prediction and statistical techniques. In this context, this research compare four forecasting approaches Autoregressive Integrated Moving Average, Single Layer Perceptron, Multiple Layer Perceptron, and Long Short-Term Memory in a one-day ahead horizon using incomplete datasets measured in a tropical and mountainous environment. The results show that the neural network-based models outperform the ARIMA model. Furthermore, LSTM has better performance with a low number of input data and in cloudiness environ-

ments.

**Keywords:** Community participation, Rural electrification, Analytic Hierarchy Process, Multicriteria Approach, Energy projects, Human Development Index, Sustainable Development Goal Index, Temperature based models, data imputation, Hargreaves and Samani, Spatial interpolation techniques, solar radiation mapping.

# Contenido

<b>Agradecimientos</b>	<b>iv</b>
<b>Resumen</b>	<b>v</b>
<b>Lista de símbolos</b>	<b>xvi</b>
<b>1 Introducción</b>	<b>1</b>
<b>2 Metodología multicriterio para priorización de proyectos sociales y participación Comunitaria: caso de estudio Nariño</b>	<b>5</b>
2.1 Introducción . . . . .	5
2.2 Metodología . . . . .	8
2.2.1 Evaluación comunitaria . . . . .	10
2.2.2 Evaluación del proyecto . . . . .	12
2.2.3 Ponderación de dimensiones . . . . .	13
2.2.4 Directriz para trabajo comunitario . . . . .	15
2.3 Descripción del caso de estudio . . . . .	16
2.4 Resultados obtenidos . . . . .	17
2.4.1 Priorización de proyectos . . . . .	17
2.4.2 Directriz para el trabajo comunitario . . . . .	23
2.5 Conclusiones . . . . .	24
<b>3 Evaluación de modelos empíricos para estimar irradiancia solar global usando datos de temperatura en un ambiente tropical y montañoso. Parte I: imputación</b>	<b>26</b>
3.1 Introducción . . . . .	27
3.2 Materiales y métodos . . . . .	28
3.2.1 Descripción de los datos . . . . .	28
3.2.2 Control de calidad de los datos de irradiancia solar . . . . .	31
3.2.3 Control de calidad de datos de temperatura . . . . .	32
3.2.4 Modelos empíricos basados en temperatura . . . . .	33
3.2.5 Modelo empírico propuesto . . . . .	36
3.2.6 Validación estadística . . . . .	37
3.3 Resultados y discusiones . . . . .	39
3.3.1 Control de calidad de datos irradiancia solar global . . . . .	39

3.3.2	Control de calidad de datos de temperatura . . . . .	41
3.3.3	Calibración y validación de modelos empíricos . . . . .	41
3.3.4	Imputación de datos de insolación solar diaria . . . . .	45
3.4	Conclusiones . . . . .	46
<b>4</b>	<b>Evaluación de modelos empíricos para estimar irradiancia solar global usando datos de temperatura en un ambiente tropical y montañoso: Parte II: mapeo</b>	<b>50</b>
4.1	Introducción . . . . .	50
4.2	Descripción de los datos . . . . .	52
4.3	Marco teórico . . . . .	54
4.3.1	Pondedor inverso por distancia (IDW) . . . . .	55
4.3.2	Kriging . . . . .	56
4.3.3	Validación cruzada . . . . .	59
4.4	Metodología . . . . .	59
4.5	Resultados y discusión . . . . .	60
4.6	Conclusiones . . . . .	68
<b>5</b>	<b>Pronóstico de corto plazo de irradiancia solar global en ambientes tropicales</b>	<b>74</b>
5.1	Introducción . . . . .	74
5.2	Materiales y métodos . . . . .	76
5.2.1	ARIMA . . . . .	77
5.2.2	ANN . . . . .	78
5.2.3	LSTM . . . . .	78
5.2.4	Control de calidad de datos . . . . .	81
5.2.5	Imputación de datos . . . . .	82
5.2.6	Criterios de comportamiento del modelo . . . . .	82
5.3	Configuración experimental . . . . .	83
5.3.1	Localización y conjunto de datos . . . . .	83
5.3.2	Pre-procesamiento de los datos . . . . .	83
5.3.3	Diseño del modelo . . . . .	84
5.4	Resultados y discusión . . . . .	86
5.4.1	Resultados de la imputación de datos horario . . . . .	86
5.4.2	Resultados del pronóstico . . . . .	89
5.5	Conclusiones . . . . .	109
<b>6</b>	<b>Conclusiones de la tesis</b>	<b>118</b>
6.1	Recomendaciones . . . . .	120

# **Lista de Figuras**

<b>2-1</b>	Esquema de la metodología . . . . .	10
<b>2-2</b>	Esquema de priorización de proyectos . . . . .	17
<b>2-3</b>	Guía para el trabajo comunitario . . . . .	23
<b>3-1</b>	Subregiones administrativas . . . . .	29
<b>3-2</b>	Localización de las AWS . . . . .	30
<b>3-3</b>	Temperatura contra índice de claridad . . . . .	45
<b>3-4</b>	Relación entre los coeficientes empíricos a y b y la altitud del sitio . . . . .	46
<b>3-5</b>	Imputación de datos . . . . .	47
<b>3-7</b>	Insolación solar global diaria por mes . . . . .	48
<b>4-1</b>	Localización de las estaciones meteorológicas . . . . .	53
<b>4-2</b>	Variograma experimental y teórico . . . . .	63
<b>4-4</b>	RMSE . . . . .	65
<b>4-5</b>	MAE . . . . .	66
<b>4-6</b>	MBE . . . . .	66
<b>4-7</b>	SD . . . . .	67
<b>4-8</b>	MPE . . . . .	67
<b>4-9</b>	Enero . . . . .	68
<b>4-10</b>	Febrero . . . . .	68
<b>4-11</b>	Marzo . . . . .	69
<b>4-12</b>	Abril . . . . .	69
<b>4-13</b>	Mayo . . . . .	70
<b>4-14</b>	Junio . . . . .	70
<b>4-15</b>	Julio . . . . .	71
<b>4-16</b>	Agosto . . . . .	71
<b>4-17</b>	Septiembre . . . . .	72
<b>4-18</b>	Octubre . . . . .	72
<b>4-19</b>	Noviembre . . . . .	73
<b>4-20</b>	Diciembre . . . . .	73
<b>5-1</b>	Estructura de una ANN . . . . .	79
<b>5-2</b>	. . . . .	80

<b>5-3</b>	Localización de las estaciones meteorológicas . . . . .	83
<b>5-4</b>	Clasificación de los días por el índice de claridad . . . . .	85
<b>5-5</b>	Esquema del algoritmo . . . . .	86
<b>5-6</b>	Distribución de los NAs en la La Josefina . . . . .	87
<b>5-7</b>	Imputación de datos La Josefina en 2015 . . . . .	87
<b>5-8</b>	Resultados del pronóstico horario en Biotopo . . . . .	90
<b>5-9</b>	Resultados del pronóstico horario en Guapi . . . . .	91
<b>5-10</b>	Resultados del pronóstico horario en Altaquer . . . . .	92
<b>5-11</b>	Resultados del pronóstico horario en Granja el Mira . . . . .	93
<b>5-12</b>	Resultados del pronóstico horario en Viento Libre . . . . .	94
<b>5-13</b>	Resultados del pronóstico horario en Universidad de Nariño . . . . .	95
<b>5-14</b>	Resultados del pronóstico horario en Botana . . . . .	96
<b>5-15</b>	Resultados del pronóstico horario en La Josefina . . . . .	97
<b>5-16</b>	Hourly forecasting results in Paraiso . . . . .	98
<b>5-17</b>	Resultados del pronóstico horario en Ospina Perez . . . . .	99
<b>5-18</b>	Resultados del pronóstico horario en Sandona . . . . .	100
<b>5-19</b>	Resultados del pronóstico horario en Cerro Páramo . . . . .	101
<b>5-20</b>	Resultados estadísticos . . . . .	101
<b>5-21</b>	Resultados estadísticos . . . . .	102
<b>5-22</b>	Resultados estadísticos en Viento Libre . . . . .	102
<b>5-23</b>	Resultados estadísticos . . . . .	102
<b>5-24</b>	Resultados estadísticos . . . . .	103
<b>5-25</b>	Resultados estadísticos . . . . .	103
<b>5-26</b>	Resultados estadísticos en Cerro Páramo . . . . .	103
<b>5-27</b>	Resultados del pronóstico diario en Biotopo . . . . .	104
<b>5-28</b>	Resultados del pronóstico diario en Guapi . . . . .	105
<b>5-29</b>	Resultados del pronóstico diario en Altaquer . . . . .	106
<b>5-30</b>	Resultados del pronóstico diario en Granja el Mira . . . . .	107
<b>5-31</b>	Resultados del pronóstico diario en Viento Libre . . . . .	108
<b>5-32</b>	Resultados del pronóstico diario en Universidad de Nariño . . . . .	109
<b>5-33</b>	Resultados del pronóstico diario en Botana . . . . .	110
<b>5-34</b>	Resultados del pronóstico diario en La Josefina . . . . .	111
<b>5-35</b>	Resultados del pronóstico diario en Paraiso . . . . .	112
<b>5-36</b>	Resultados del pronóstico diario en Ospina Perez . . . . .	113
<b>5-37</b>	Resultados del pronóstico diario en Sandona . . . . .	114
<b>5-38</b>	Resultados del pronóstico diario en Cerro Páramo . . . . .	115
<b>5-39</b>	Error estadístico diario . . . . .	115
<b>5-40</b>	Error estadístico diario . . . . .	116
<b>5-41</b>	Error estadístico diario en Viento Libre . . . . .	116
<b>5-42</b>	Error estadístico diario . . . . .	116
<b>5-43</b>	Error estadístico diario . . . . .	117

<b>5-44</b> Errores estadísticos diarios . . . . .	117
<b>5-45</b> Error estadístico diario en Cerro Páramo . . . . .	117

# **Lista de Tablas**

<b>2-1</b>	Dimensiones para la evaluación de la comunidad . . . . .	11
<b>2-2</b>	Escala de Saaty . . . . .	14
<b>2-3</b>	Clasificación de localidades por horas de suministro eléctrico . . . . .	16
<b>2-4</b>	Resultados de la evaluación de la comunidad . . . . .	19
<b>2-5</b>	Normalized NPV . . . . .	20
<b>2-6</b>	Ponderación por objetivos de priorización . . . . .	20
<b>2-7</b>	Priorización de proyectos . . . . .	21
<b>2-8</b>	Evaluación financiera del proyecto de electrificación escolar . . . . .	22
<b>2-9</b>	Matriz de correlación en los proyectos . . . . .	22
<b>3-1</b>	Estaciones meteorológicas automáticas . . . . .	30
<b>3-2</b>	HS coeficientes empíricos . . . . .	34
<b>3-3</b>	Clasificación por índice de claridad . . . . .	36
<b>3-4</b>	Medidas de errores estadísticos . . . . .	38
<b>3-5</b>	Resultados de validación de datos de irradiancia global . . . . .	40
<b>3-6</b>	Cantidad de días con datos entre las 6 y 18 horas . . . . .	40
<b>3-7</b>	Clasificación por índice de claridad . . . . .	41
<b>3-8</b>	Resultados de validación de temperatura . . . . .	41
<b>3-9</b>	Cantidad de días por mediciones diarias de temperatura . . . . .	42
<b>3-10</b>	Coeficientes empíricos . . . . .	42
<b>3-11</b>	Resumen de los resultados de los modelos empíricos . . . . .	43
<b>3-12</b>	Imputación por AWS . . . . .	46
<b>4-1</b>	Estaciones meteorológicas automáticas . . . . .	52
<b>4-2</b>	Estaciones meteorológicas convencionales . . . . .	53
<b>4-3</b>	Número de días clasificados por cantidad de mediciones diarias de irradiancia global solar . . . . .	54
<b>4-4</b>	Número de días clasificados por cantidad de mediciones diarias de temperatura . . . .	54
<b>4-5</b>	Coeficientes empíricos estimados para la zona Andina y Amazónica . . . . .	62
<b>4-6</b>	Comparación de los modelos empíricos . . . . .	62
<b>4-7</b>	Spatial interpolation methods results . . . . .	64
<b>5-1</b>	Statistical errors . . . . .	82

<b>5-2</b>	Estaciones meteorológicas automáticas . . . . .	84
<b>5-3</b>	Resultados de imputación . . . . .	88
<b>5-4</b>	Errores estadísticos del proceso de imputación . . . . .	89

# Listado de símbolos

## Símbolos con letras del alfabeto latino

Símbolo	Definición
a	Coeficiente empírico
A	Semivarianza de la matriz entre el $i^{th}$ y el $j^{th}$ punto
b	Coeficiente empírico
c	Coeficiente empírico
$d_i$	Distancia entre el $x_i$ y $x_0$
D	Día Juliano
H	Irradiancia global solar horizontal a nivel del suelo
$H_0$	Irradiancia solar extraterrestre
$I_0$	Irradiancia solar extraterrestre horaria
$I_{cs}$	Irradiancia solar global en cielo despejado horaria
$I_{mt}$	Irradiancia solar global en el tiempo $t$
$I_{sc}$	constante solar
$K_t$	Índice de claridad
n	Número de puntos muestreados
m	Masa de aire
p	Exponente
r	El radio de la ventana de búsqueda
$T_h$	Temperatura horaria medida
$T_{max}$	Temperatura máxima diaria
$T_{mean}$	Temperatura media diaria
$T_{min}$	Temperatura mínima diaria

<b>Símbolo</b>	<b>Definición</b>
$T_R$	Relación entre la temperatura diaria mínima y máxima
$\Delta T$	Diferencia entre la de temperatura diaria mínima y máxima
$u(x)$	Media no estacionaria
$x_0$	Punto objetivo
$x_i$	Punto no muestreado
$z$	Punto muestreado
$z(x_i + h)$	Valor a una distancia $h$ desde $x_i$

## Símbolos con letras Griegas

<b>Símbolo</b>	<b>Definición</b>
$\phi$	Latitud
$\delta$	Declinación solar
$\omega_s$	Ángulo horario al atardecer
$\beta$	Altitud solar
$\tau$	Transmitancia atmosférica

## Abreviaciones

<b>Abreviación</b>	<b>Termino</b>
AHP	Proceso Analítico Jerárquico
ANN	Red neuronal artificial
ARMA	Promedio móvil autoregresivo
ARIMA	Promedio móvil autoregresivo integrativo

<b>Abreviación</b>	<b>Termino</b>
AWS	Estación meteorológica automática
CEPAL	Comisión Económica para América Latina y el Caribe
CWS	Estación meteorológica convencional
DANE	Departamento Administrativo Nacional de Estadística
IDEAM	Instituto de Hidrología, Meteorología y Estudios Ambientales
IDW	Ponderado inverso por distancia
IPSE	Instituto de Planificación y Soluciones Energéticas para las Zonas no Interconectadas
PERS	Programa de Energización Rural Sostenible
UPME	Unidad de Planeación Minero Energética

# 1 Introducción

Las energías renovables son un recurso disponible localmente con beneficios ambientales y sociales, que permiten la integración comunitaria en proyectos energéticos (Dawoud et al., 2019). La proliferación de sistemas de generación de electricidad con energía solar ha impulsado estudios para estimar el potencial del recurso solar, pronosticar el recurso en diferentes horizontes de tiempo, e incluir a la comunidad dentro de estos proyectos. Incluir a la comunidad en la planificación y ejecución de proyectos energéticos, como electrificación rural o proyectos productivos, mejora la sostenibilidad de las iniciativas en el tiempo. Cuando la población es parte activa de los proyectos energéticos, su desarrollo facilita la interacción entre las comunidades, las entidades financieras y los desarrolladores del proyecto. En consecuencia, el análisis conjunto de proyectos y comunidades aborda los proyectos energéticos desde un punto de vista socio-técnico. Además, los proyectos de energía sostenible se correlacionan con el desarrollo socioeconómico de las poblaciones, debido al crecimiento de la población, la urbanización y la industrialización.

Dentro de las energías renovables, el aprovechamiento de la energía solar con tecnología fotovoltaica ha crecido en las últimas décadas. En la actualidad, los costos asociados a esta tecnología son competitivos, principalmente en las zonas rurales donde la electricidad generada con recursos convencionales podría resultar más costosa. Fomentar la generación de electricidad con energía solar requiere conocer el recurso. Sin embargo, en los países en desarrollo, los datos de irradiancia solar no están ampliamente disponibles, debido a la escasez de estaciones meteorológicas que midan esta variable y los requisitos de calibración y mantenimiento de los equipos. En algunos casos, aunque las bases de datos tienen un acceso abierto, el proceso de obtención de información puede ser extenso (Bakirci, 2009). Por ello, en los últimos años, los estudios de irradiancia solar han utilizado otras variables relevantes como temperatura, brillo solar, humedad relativa, entre otras, para estimar esta variable.

Los investigadores solares han desarrollado modelos y métodos para estimar la insolación solar diaria o mensual, con modelos empíricos o inteligencia artificial. Los modelos empíricos suelen estar basados en factores astronómicos, geométricos, físicos y meteorológicos, destacando el último (Besharat et al., 2013). La implementación del modelo empírico depende de la disponibilidad y consistencia de los datos (Akinoglu, 2008). En este sentido, Almorox et al. recomendó que los datos del modelo empírico sean simultáneos y confiables en el sitio de estudio (Almorox et al., 2011). Sin importar el método de estimación de insolación solar que se use, los datos recopilados a nivel del suelo ofrecen información de una fuente puntual. Por lo tanto, es necesario estimar los valores de insolación solar en los puntos no muestreados, construyendo datos espacialmente continuos sobre el área de estudio (J. Li & Heap,

2011). La estandarización de la información de las mediciones aisladas es posible con las técnicas de interpolación regional, a través de la construcción de un modelo regional (Moreno et al., 2011). En general, hay dos categorías para las técnicas de interpolación: no geoestadísticas y geoestadísticas. Sin embargo, todos los métodos comparten la misma fórmula de estimación general (J. Li & Heap, 2008).

Además de investigaciones para determinar el potencial de insolación solar en áreas, también se ha impulsado los estudios de pronóstico de irradiancia solar (Reikard, 2009). El pronóstico de la irradiancia solar permite estimar la generación eléctrica a largo, mediano y corto plazo. Esta información es crucial para mantener el equilibrio entre la demanda y la oferta de energía eléctrica (Dannecker, 2015) y minimizar los costos asociados con el arranque y parada de las centrales eléctricas convencionales (Badosa et al., 2017). El pronóstico consiste en el análisis de series de tiempo, que es el proceso de examinar los datos registrados a lo largo del tiempo para desarrollar un modelo matemático que describa las características de los datos (Shumway & Stoffer, 2011). Estos modelos facilitan la implementación de sistemas fotovoltaicos (PV), tanto dentro como fuera de la red eléctrica. Las series de tiempo registradas en períodos prolongados se utilizan para comprender el comportamiento y predecir los valores futuros de irradiancia solar en una ubicación específica (Suehrcke, 2000).

Existen varios enfoques para pronosticar la irradiancia solar: métodos de persistencia, modelos autorregresivos y técnicas de computación suave (Demirhan & Renwick, 2018; Diagne et al., 2013; Inman et al., 2013). Las redes neuronales (ANN), la lógica difusa y los enfoques híbridos son robustos para modelar la naturaleza estocástica de los procesos físicos, como la irradiancia solar, debido a su capacidad para compensar errores sistemáticos y desviaciones (Paulescu et al., 2013). La selección de un método de pronóstico depende de la escala de tiempo deseada, es decir, intrahorario, intradiario, mediano y largo plazo. Los enfoques estadísticos generalmente funcionan bien para pronósticos a corto plazo, como ARIMA y ANN. Para el análisis a largo plazo, se prefieren con frecuencia las técnicas de computación suave (Coimbra et al., 2013; Demirhan & Renwick, 2018).

Este trabajo está enfocado en crear una cadena de valor al rededor del recurso solar para el departamento de Nariño desde un enfoque energético. Este departamento está localizado al sur de Colombia, se caracteriza por su población diversa, sus tres regiones ambientales: Pacífica, Andina y Amazónica, y las necesidad de mejorar la calidad de vida de la población que ha sido afectada por el conflicto armado. En términos de suministro de electricidad, el departamento se divide en dos zonas, la primera recibe electricidad del Sistema Interconectado Nacional (SIN), y la segunda se denomina como Zona No Interconectada (ZNI) que obtiene la electricidad de plantas locales que usan combustibles fósiles. En el 2014 el Congreso de Colombia promulgó la Ley 1715 para fomentar la generación de electricidad con fuentes de energía renovable no convencional (El Congreso de Colombia, 2014). Esa ley promueve la generación distribuida y la autogeneración de electricidad. Por lo tanto, considerando el marco normativo de Colombia y las características de Nariño, este trabajo propone una metodología de priorización de proyectos productivos y una guía comunitaria para el trabajo conjunto entre empresa y comunidad. Además, esta investigación evalúa tres modelos empíricos basados en la temperatura: Hargreaves y Samani (HS), Bristow y Campbell (BC) y Okundamiya y Nzeako (ON), para estimar la insolación solar en

---

un ambiente tropical y montañoso. Por último, este estudio aplica cuatro modelos de predicción de última generación para el pronóstico de la irradiancia solar global en ambientes tropicales y montañosos con datos incompletos: i) ARIMA, ii) Perceptrón de capa única (SLP), iii) Perceptrón de múltiples capas (MLP) ) y iv) LSTM.

Para la primera parte, los autores analizaron varios componentes de la evaluación comunitaria, como el Índice de Desarrollo Humano y los Objetivos de Desarrollo Sostenible, y los adaptaron a las características de Nariño y a la información disponible en las bases de datos públicas nacionales o regionales. El valor actual neto (VAN) fue la medida utilizada para la evaluación de proyectos, porque es una medida de uso común para la evaluación de costos y beneficios en proyectos de energía fotovoltaica. El Analytic Herachy Process (AHP) fue el método para priorizar proyectos y comunidades. El método procesó seis objetivos de priorización para presentar diferentes perspectivas para la toma de decisiones. Adicionalmente, es esta parte el trabajo propone los lineamientos de participación de la comunidad para integrar sus prioridades, inquietudes e intereses con los actores en el desarrollo del proyecto energético con el fin de incrementar la probabilidad de éxito de estos. En esta parte del trabajo de encontró una relación entre los objetivos que buscaban mejorar la educación y la calidad de vida de la población.

Para la estimación del potencial de insolación solar, los autores analizaron tres modelos empíricos basados en temperatura. El objetivo de estos modelos fue, por una parte, llenar los vacíos de los datos de insolación solar que tenían las series de datos, y por otra parte incrementar el número de puntos muestreados con datos estimados de insolación solar, con el fin de mejorar la calidad de los mapas de potencial de energía solar. Además de los modelos empíricos analizados, los autores proponen una nueva relación basada en el modelo de regresión logística para estimar la insolación solar con la diferencia diaria de temperatura. Como resultado, el modelo propuesto presenta una relación lineal entre las constantes empíricas y la altitud del sitio. Esta relación lineal permitió calcular las constantes empíricas que facilitaron la estimación de la insolación solar en puntos donde solo se tenía datos de temperatura. Finalmente, con los resultados de los modelos empíricos se logró incrementar la densidad de puntos muestreados de insolación solar en el área del departamento. Este resultado fue relevante ya que el desempeño de las técnicas de interpolación espacial está ligado a la densidad de puntos muestreados y a la agrupación de estos.

Con el objetivo de mostrar el potencial de la insolación solar promedio mensual que tiene el departamento de Nariño, y al mismo tiempo incentivar la construcción de plantas de generación de electricidad que usa este recurso, principalmente en las ZNI. Esta investigación utiliza mediciones *in situ* de la insolación solar global recolectadas en ocho AWS ubicadas en el departamento de Nariño, y la insolación solar estimada, con datos de temperatura, a través de modelos empíricos en diecisésis ubicaciones con estaciones meteorológicas convencionales (CWS). Con esa información, los autores utilizaron técnicas de interpolación espacial como el Ponderado Inverso por Distancia (IDW) y Kriging simple, ordinario y universal para obtener información de insolación solar espacialmente continua en la región estudiada. La validación cruzada mostró que el método de interpolación Kriging ordinario fue el mejor entre los métodos analizados, porque presentó el menor sesgo durante todo el año debido a su robustez.

En la última parte de esta investigación, y con el fin principalmente de incentivar el uso de la energía solar en las zonas que pertenecen al SIN, se implementan cuatro modelos de pronóstico en un horizonte de un día de adelanto en un rango horario y diario. En esta parte, la imputación de datos faltantes horarios combina una interpolación lineal de los valores subsiguientes con el promedio de valores pasados medidos a la misma hora de los datos imputados. Mientras que para los datos en un rango diario se usan los modelos empíricos basados en temperatura. Los resultados muestran que las técnicas de basadas en redes neuronales tienen un mejor rendimiento que ARIMA. El modelo LSTM tiene un mejor rendimiento en la zona del Pacífico en las AWS con menos cantidad de datos medidos. Mientras que SLP logra el mejor comportamiento en las AWS con más datos de entrada.

Esta tesis está organizada de la siguiente manera, el Capítulo 2 corresponde al proceso de priorización de proyectos energéticos y la guía de participación comunitaria. El Capítulo 3 muestra los resultados de la implementación de tres modelos empíricos basados en temperatura, y el modelo propuesto basado en la regresión logística, que fueron el soporte para la imputación de datos de insolación solar diaria. El Capítulo 4 presenta el proceso de estimación de insolación solar diaria usando los modelos empíricos de Hargreaves y Samani el propuesto, y la aplicación de cuatro técnicas de interpolación espacial que dan como resultado doce mapas de potencial de insolación solar diaria para cada mes. El capítulo 5 muestra los resultado de aplicar cuatro modelos de pronóstico de irradiancia solar en un horizonte de un día de adelanto en un rango horario. Finalmente, el Capítulo 6 recoge todas las conclusiones del trabajo de investigación.

## **2 Metodología multicriterio para priorización de proyectos sociales y participación Comunitaria: caso de estudio Nariño case<sup>1</sup>**

### **Resumen**

El análisis conjunto de proyectos y comunidades permite abordar proyectos energéticos desde un enfoque sociotécnico. Incluir a las comunidades en el desarrollo de proyectos requiere una mejor comprensión de sus aspectos Sociales, económicos, ambientales y culturales. En consecuencia, este trabajo se centra en formular una metodología para priorizar proyectos energéticos, considerando la participación de la comunidad. En la evaluación de las comunidades se consideró los Índices de Desarrollo Humano y los Índices de Objetivos de Desarrollo Sustentable y ajustó estos de acuerdo con la información disponible para Nariño, la evaluación de los proyectos se hizo con el Valor Presente Neto. Para la evaluación conjunta de las comunidades y los proyectos se utilizó la metodología de análisis de proceso jerárquico y se estableció seis objetivos de priorización. Además, la metodología de co-construcción fue el soporte para formular lineamientos para trabajo comunitario. Esta investigación encontró una relación cercana entre los proyectos que contribuyen a mejorar la calidad de vida y la educación en Nariño.

**Palabras clave** Proyecto energético; Proceso Analítico Jerárquico; Electrificación Rural; Índices de Desarrollo Humano; Índice de Desarrollo Sustentable.

### **2.1 Introducción**

Incluir a la comunidad en la planificación y ejecución de proyectos energéticos, tales como electrificación rural o proyectos productivos, mejora la sustentabilidad de estas iniciativas. Cuando la población es una parte activa de los proyectos energéticos su desarrollo facilita la interacción entre las comunidades,

---

<sup>1</sup>Este artículo tuvo el apoyo Belizza Ruiz and Patricio Mendoza

las entidades de financiación y los desarrolladores del proyecto. Consecuentemente, el análisis conjunto de los proyectos y las comunidades aborda los proyectos energéticos desde un enfoque sociotécnico. Además, los proyectos energéticos sustentables están correlacionados con el desarrollo socioeconómico de la población, debido al crecimiento de la población, urbanización e industrialización. Las energías renovables son un recurso localmente disponible con beneficios ambientales y sociales que permiten integrar la comunidad dentro de los proyectos energéticos (Dawoud et al., 2019). El último aspecto es relevante porque el enfoque tradicional para la electrificación rural ha sido débil en términos de involucrar las comunidades en el desarrollo de los proyectos. Cuando hay una brecha entre la comunidad y el proyecto, éste presenta dificultades para sostenerse en el tiempo (Palma-Behnke et al., 2019). Los proyectos de electrificación rural son esenciales porque la electricidad es un factor significativo en la economía y mejora la calidad de vida de la población. Por lo tanto, su sustentabilidad sobre el tiempo es también relevante y debería ser un objetivo principal en los proyectos energéticos (Balbás Egea & Egiguren, 2019). Sin embargo, la Agencia Internacional de Energía declaró que cerca de 1.100 millones de personas no tenían acceso a electricidad en 2017, y cerca del 84 % vivían en zonas rurales (International Energy Agency, 2017).

La falta de iniciativas de electrificación rural es un limitante para el desarrollo, especialmente en regiones con una alta población en estas zonas, como es el caso de Nariño, donde la población rural alcanza el 53 % del total de la población. Las principales ventajas de la electrificación rural son mejoras en los sistemas de iluminación para hogares, mejoras en centros de educación y salud, mejores condiciones de trabajo para actividades domésticas, reducción de la migración rural a ciudades debido a la creación de trabajos a través de la tecnificación del sector agrícola, preservación del ambiente a través del uso de fuentes de energía renovable para la generación de electricidad, entre otros (Camblong et al., 2009). En este sentido, el Gobierno de Nariño identificó tres propósitos en su plan de desarrollo 2016-2019 paz territorial y equidad social, crecimiento e innovación verde, e integración regional (Gobernación de Nariño, 2016a). Asimismo, éste estableció siete áreas estratégicas para alcanzar esos propósitos, cada uno con un programa de acción cubriendo uno o más objetivos. Uno de esos programas es la electrificación rural para impulsar el desarrollo en otros campos.

El enfoque tradicional para suministrar electricidad a áreas rurales o aisladas ha sido a través de la extensión del sistema de transmisión o distribución, o con el uso de unidades de generación que usan combustibles fósiles (Ubilla et al., 2014). El primer enfoque usualmente requiere alta inversión, haciendo este económicamente inviable cuando hay bajos niveles de demanda de electricidad. El segundo enfoque, aunque requiere baja inversión, resulta en altos costos operacionales; por lo tanto, la tarifa de electricidad es elevada, como ocurre en las zonas no interconectadas de Nariño, debido a los costos de transporte de los combustibles fósiles desde el puerto marítimo más cercano a las comunidades (Universidad de Nariño et al., 2014). Otro elemento para resaltar relacionado al uso de combustibles fósiles para la generación de electricidad es la emisión de gases de efecto invernadero que impactan negativamente el ambiente. Como resultado existen programas e iniciativas para reducir su consumo e introducir energías renovables para la generación de electricidad (Opoku et al., 2020). Adicionalmente la tecnología fotovoltaica es una opción viable considerando las condiciones de terreno, la disponibili-

dad del recurso y la madurez de la tecnología (Figueirêdo Neto & Rossi, 2019).

La generación de electricidad con fuentes renovables disponibles a nivel local permite la inclusión de la comunidad dentro de los proyectos, y transforma estos en proyectos sociotécnicos. Por lo tanto, el objetivo a largo plazo del Instituto de Planificación y Promoción de Soluciones Energéticas para Zonas no Interconectadas (IPSE) será la inclusión de las comunidades en los proyectos de generación de electricidad para mantener estos en el tiempo (Rodriguez, 2011). En consecuencia, el IPSE deberá mejorar la implementación de proyectos, adaptando estos a las necesidades de las comunidades. Además, reducir la brecha entre las comunidades y el desarrollo de los proyectos (Superintendencia de Servicios Públicos Domiciliarios SSPD, 2019). Lograr este objetivo requiere evaluar diversos aspectos de la comunidad y formular estrategias para facilitar el trabajo entre los interesados, la comunidad y la entidad financiadora (Marinakis et al., 2017). Para aportar en este aspecto, esta investigación se enfoca en formular una metodología de dos partes para la toma de decisiones, determinar presupuesto, y formular políticas públicas en el sector energético que incrementen la sustentabilidad de los proyectos energéticos. La primera parte consiste en la priorización de proyectos usando la metodología multicriterio proceso analítico jerárquico (AHP) compuesta de seis dimensiones, cinco para la evaluación de la comunidad, y una para la evaluación del proyecto, que fue el Valor Presente Neto (VPN). La segunda parte, muestra una guía para trabajo con la comunidad con el fin de incluir ésta como una parte activa de los proyectos energéticos.

Este trabajo propone una metodología de priorización para proyectos productivos y una guía de trabajo comunitario. Para la primera parte, los autores analizaron varios componentes para evaluar la comunidad, tales como los índices de desarrollo humano (HDI) y los índices de los Objetivos de Desarrollo sostenible (SDGI), y adaptaron estos a las características de Nariño y a la información disponible en las bases de datos públicas regionales y nacionales. Dicha información es crucial para la selección de proyectos con la mejor solución en aspectos sociales, ambientales y económicos (Zore et al., 2018). Los proyectos seleccionados provienen del Programa de Energización Rural Sostenible (PERS), y el criterio de selección fue el recurso energético, en este caso, la irradiancia solar. Para la evaluación de proyectos se utilizó el VPN porque esta es una medida comúnmente usada para la evaluación de costo-beneficio en los proyectos fotovoltaicos (Yang et al., 2018). La AHP fue el método para la priorización de proyectos y comunidades. El método procesó seis objetivos de priorización, con el fin de presentar diferentes perspectivas para los tomadores de decisiones. La segunda parte, propone una guía para la participación de la comunidad, con el fin de exponer sus prioridades, preocupaciones e intereses con los desarrolladores del proyecto (Aslani, 2014). Los criterios propuestos para la evaluación de la comunidad son un marco de referencia que pueden facilitar la evaluación de proyectos en otros campos. Asimismo, la guía de la participación comunitaria es un proceso que invita al trabajo conjunto entre la comunidad, la entidad financiadora y los desarrolladores del proyecto para mantener el proyecto energético en el largo plazo. Esta guía incluye un proceso que se puede modificar de acuerdo con el proyecto y las necesidades de la comunidad.

Los autores encontraron una correlación directa cuando el objetivo es mejorar las condiciones de cali-

dad de vida o educación. Por lo tanto, las políticas públicas que junten estos objetivos podrían fortalecer los resultados finales. Sin embargo, es necesario analizar más proyectos productivos y recolectar más información de las comunidades con el fin de confirmar estos resultados. Además, para mejorar la caracterización e incluir las tradiciones culturales en el desarrollo de los proyectos. La guía propuesta por los autores incrementaría la probabilidad de llevar a cabo proyectos exitosos e incluir la comunidad como una parte activa del proyecto. El artículo tiene las siguientes secciones; la Sección 2.2 presenta la metodología usada en este trabajo; la sección 2.3 describe el caso de estudio; la sección 2.4 presenta los resultados obtenidos para el departamento de Nariño, y finalmente, la sección 2.5 presenta las conclusiones.

## 2.2 Metodología

Este trabajo de investigación sigue una metodología sintetizada en dos partes, como muestra la Figura 2-1. La primera parte, consiste en un proceso de priorización de proyectos rurales que implica tanto una evaluación comunitaria como de proyectos. La segunda parte, consiste en pautas generales basadas en los resultados de la primera parte de la metodología para la instalación de proyectos energéticos en Nariño.

El proceso de priorización de proyectos rurales tomó información de PERS, el Departamento Administrativo Nacional de Estadística (DANE), la Unidad de Planeación Minero Energética (UPME), entre otras instituciones. Estas entidades trabajaron directamente con la comunidad recolectando información primaria a través de encuestas, para construir una base de datos y esbozar planes y reportes. Por lo tanto, la información usada en este estudio refleja las condiciones de vida de las comunidades en Nariño. Teniendo en cuenta las bases de datos, los autores consideraron cinco dimensiones: educación, salud, seguridad, calidad de vida, y economía. Estos aspectos permiten entender la sustentabilidad más allá de las dimensiones estándares al incluir indicadores que evalúan la calidad de vida y las relaciones culturales e institucionales (Feleki et al., 2018). El VPN fue la herramienta para evaluar proyectos, y ésta se consideró como otra dimensión. Las dimensiones para evaluar a la comunidad y el VPN constituyeron las seis dimensiones analizadas bajo el enfoque multicriterio, que las utilizó como criterios en el proceso de evaluación.

Las herramientas de toma de decisiones multicriterio (MCDM) permiten la evaluación y priorización de problemas complejos con alta incertidumbre, objetivos conflictivos, múltiples intereses, entre otros (Vaidya & Kumar, 2006). Adicionalmente, las MCDM facilitan la clasificación y organización de un número finito de alternativas de decisión basado en atributos o criterios, las cuales describen diferentes características (Mary & Suganya, 2016). En los proyectos sociales, las MCDM permiten distribuir el presupuesto eficientemente; por lo tanto, la coordinación y el desarrollo de un proyecto sigue un lineamiento previamente establecido. Cuando un inversionista público está involucrado, la evaluación de proyectos podría resultar en la base de un programa e inclusive una política (J. Pacheco & Contr-

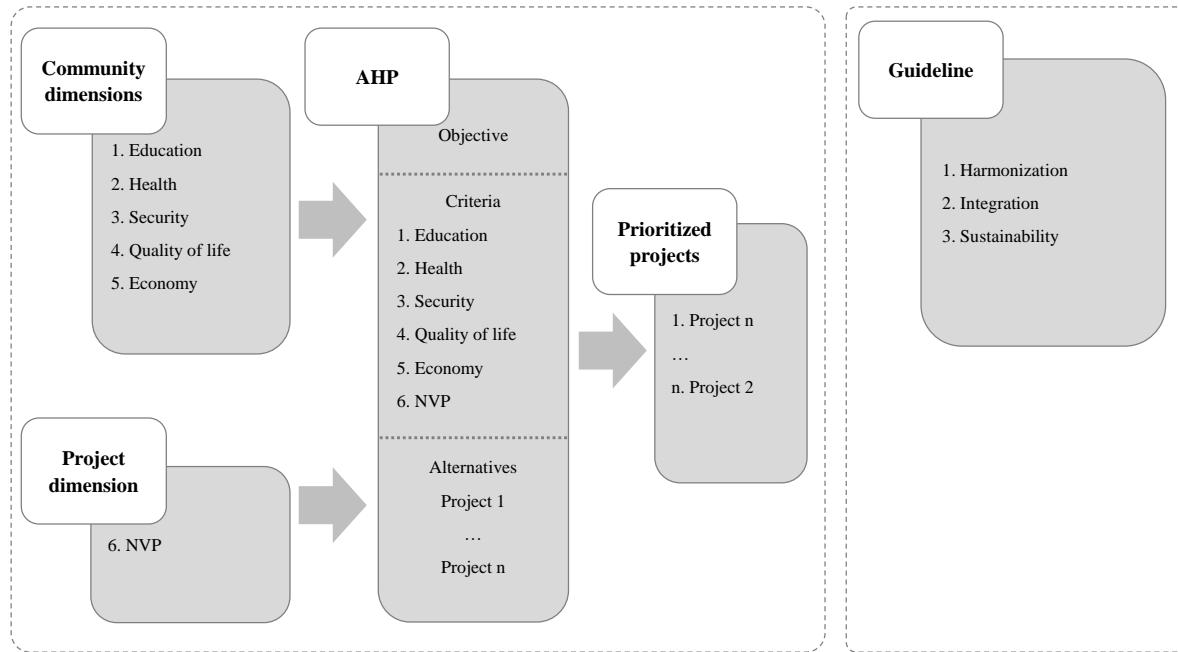
eras, 2008). El más apropiado y frecuentemente usado MCDM en la planificación energética es el AHP, seguido de PROMETHEE y ELECTREE (Abdullah & Najib, 2016; Wang et al., 2009).

El AHP tiene dos principales ventajas: la cuantificación y ponderación de criterios dado un objetivo y diferentes opciones para alcanzar éste, y el proceso matemático es más entendible que otros métodos (P. H. Dos Santos et al., 2019). Adicionalmente, el AHP transforma comparaciones empíricas en valores numéricos. En la parte superior de la jerarquía se ubica el objetivo; en el segundo se establecen los criterios y sub criterios con los que se evaluarán las alternativas que están en la base de la jerarquía (Saaty, 1990, 2008). ELECTRE III es el método más común de la familia ELECTRE utilizado en la planificación energética. El objetivo de este método es elegir la alternativa más compatible con los criterios. Sin embargo, a veces el método presenta desafíos para encontrar la mejor alternativa; por lo tanto, se hace necesario utilizar un método más sencillo. PROMETHEE II es el método comúnmente utilizado en el campo de la planificación energética de la familia PROMETHEE. La principal diferencia entre PROMETHEE II y ELECRTE III es el proceso de cálculo. Para los responsables de la toma de decisiones, el primer método es más fácil de entender que el segundo método. Estos tipos de métodos, denominados métodos de clasificación, se utilizan normalmente para clasificar las alternativas como aceptables o inaceptables, pero no como métodos para La selección de alternativas (Løken, 2007). De los tres MCDM mencionados anteriormente, el AHP tiene las mejores propiedades y es por eso que fue elegido por los autores (Mardani et al., 2018).

La segunda parte de la metodología consiste en una directriz que facilite la interacción entre la comunidad y el gobierno local u otro tipo de entidad que promueva proyectos socio-energéticos. El objetivo de la directriz es involucrar a la comunidad en el desarrollo del proyecto para crear un sistema de gestión de acuerdo con la tradición y la cultura de la comunidad y, a su vez, aumentar las posibilidades de éxito del proyecto. Es conveniente destacar que el punto de partida para elaborar las directrices fue la metodología de co-construcción porque ésta busca empoderar a la comunidad a través de los procesos de toma de decisiones (Montedonico et al., 2018). Sin embargo, los autores la modificaron incluyendo como uno de los requisitos, la participación directa de los miembros de la comunidad en el equipo de trabajo que ejecutan el proyecto específico.

Un factor para destacar en la directriz es que ésta debe adaptarse a las características de la comunidad. En consecuencia, el caso de Nariño es relevante porque este territorio cuenta con diversos grupos de población, tales como afrocolombianos (42 territorios colectivos correspondientes al 18,80 % de la población total), los indígenas (55 reservas indígenas correspondientes al 10,79 % de la población total) y los grupos mixtos (70,41 % de la población total). Además, la diversidad del terreno que tiene valles, enormes montañas, y bosques, también debe ser considerada. Estos aspectos demuestran que esta directriz tendría aplicación en diferentes ambientes (Ministerio de Cultura, 2020).

**Figura 2-1.** Esquema de la metodología



## 2.2.1 Evaluación comunitaria

A nivel comunitario, existen varios índices para explicar los aspectos ambientales, sociales y económicos que justifican la toma de decisiones políticas. Sin embargo, no hay una colección universalmente aceptada de índices para evaluar dichas poblaciones (Feleki et al., 2018). Aunque algunas políticas gubernamentales siguen el principio de maximización de beneficios, otros índices también son relevantes debido a su enfoque más amplio, como el sistema de contabilidad ambiental-económica (SEEA), ahorro neto ajustado (ANS), índice planeta vivo (LPI), indicador de progreso genuino (GPI), índice de desarrollo humano (HDI), índice de objetivos de desarrollo sostenible (SDGI), entre otros (Arbeláez-Arias, 2006). A pesar de que los indicadores mencionados anteriormente podrían ofrecer un contexto completo, la información pública disponible del departamento de Nariño era insuficiente para calcularlos. En consecuencia, los autores sólo utilizaron índices tanto de HDI como de SDGI que se aplican a la información disponible (Bertelsmann Stiftung & Sustainable Development Solutions Network (SDSN), 2016; United Nations, 2018).

Teniendo en cuenta todos los factores en el departamento de Nariño, los autores analizaron cinco dimensiones para la evaluación comunitaria: educación, salud, seguridad, calidad de vida y economía. Cada dimensión tiene asociado uno o más componentes o características sociales, como se muestra en la Tabla 2-1.

**Tabla 2-1.** Dimensiones para la evaluación de la comunidad

Dimensión*	Componente	Descripción
Educación	Alfabetismo	La población mayor (15+ años) que puede leer y escribir.
	Cobertura de preescolar	Correspondiente a la proporción entre la población cinco años que asiste a la escuela y la población total en este grupo de edad.
	Cobertura neta de escuela primaria	Corresponde a la relación entre la población entre seis y diez años que asiste a la escuela y la población total en este grupo de edad.
	Cobertura neta de escuela secundaria	Corresponde a la proporción entre la población de once y catorce años que asiste a la escuela y la población total en este grupo de edad.
	Cobertura de neta de educación media	Corresponde a la relación entre la población de quince y diecisésis años que asisten a la escuela y la población total en este grupo de edad.
Salud	Puntuación nacional estandarizada	Corresponde a la puntuación de los estudiantes de 11º grado en la prueba nacional llamada Saber 11.
	Cobertura del seguro social	Corresponde al porcentaje de la población afiliada a cualquier seguro médico.
Seguridad	Tasa de mortalidad infantil	Corresponde a las muertes de menores de un año por cada 1.000 niños nacidos vivos.
	Incidencia del conflicto armado	Mide acciones armadas contra las fuerzas del Estado, homicidios, secuestros, víctimas de minas antipersonas, desplazamiento forzado y cultivos ilícitos.
Calidad de vida	Déficit de vivienda	Déficit cuantitativo y cualitativo de vivienda. La primera estima el número de casas necesarias para el total de familias. El segundo se refiere a la calidad, el hacinamiento y el acceso a los servicios públicos de las casas.
	Necesidades básicas insatisfechas	Corresponde a hogares superpoblados y en condiciones inaceptables, servicios públicos inadecuados, alta dependencia económica, niños en edad escolar que no asisten a la escuela.
Economía	Importancia económica	Valor añadido por el municipio con precios estándar.
	Incidencia de la pobreza multidimensional en las zonas rurales	El analfabetismo, el bajo nivel educativo, las barreras al acceso a los servicios de la primera infancia, el trabajo infantil, el trabajo informal, el desempleo de larga duración, el acceso al sistema de salud, las barreras de acceso a la atención dada una necesidad, el acceso al agua dulce y los acueductos.

\*Las dimensiones corresponden a los criterios usados en la metodología AHP, y se organizan como objetivos de priorización

Fuente:(Departamento Administrativo Nacional de Estadística - DANE, 2019, 2009; Gobernación de Nariño, 2016b)

Todos los componentes de la Tabla 2-1 son comparables; sin embargo, es necesario completar un proceso de normalización porque tienen diferentes unidades de medida. Para este trabajo, el procedimiento de normalización es el de mínimo-máximo, como se muestra en la Ecuación 2-1.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2-1)$$

Donde  $x'$ : es el valor del componente,  $x$ : es el valor original sin normalización,  $\max(x)$ : es el valor original máximo,  $\min(x)$ : es el valor original mínimo.

Un número entre 0 y 1 representa cada componente. Cuando la dimensión tiene más de un componente, el valor de la dimensión es el promedio de los componentes normalizados, como se muestra en la Ecuación 2-2. Aunque algunos componentes aparecen en varias dimensiones, todos ellos tienen el mismo peso. Los autores decidieron dar el mismo peso para los componentes porque el DANE no ofrece información al respecto.

$$D_i = \frac{1}{n} \sum_{j=1}^n x'_j \quad (2-2)$$

Donde  $D_i$  es la dimensión  $i$ ,  $x'_j$  es el indicador  $j$  en la dimensión  $i$ ,  $n$  es la número de componentes de la dimensión  $i$ .

## 2.2.2 Evaluación del proyecto

La comunidad académica utiliza varios métodos para realizar análisis económicos. Sin embargo, el principio es el mismo en estos métodos: "el enfoque de presupuesto de capital para calcular el rendimiento económico de un proyecto es una secuencia de flujo de caja descontado" (Žižlavský, 2014). En esta investigación, el VPN evalúa los proyectos porque se trata de una herramienta de análisis de evaluación de inversiones utilizada en los proyectos del sector energético cuando los costes son la variable principal (Konstantin & Konstantin, 2018). El VPN es una técnica de evaluación económica popular y sofisticada, que es la suma de los valores actuales de entradas y salidas durante un período y la tasa de descuento que introducen una tasa de riesgo (Gaspars-Wieloch, 2019). En la evaluación del proyecto, los autores consideraron dar prioridad a los proyectos con el VPN más bajo, siguiendo el enfoque de maximización económica. Esta herramienta es útil cuando los beneficios económicos son iguales en más de un proyecto o no medibles (Ubilla et al., 2014). Los proyectos sociales implican diferentes variables relacionadas con el bienestar que no son medibles a partir de beneficios económicos. Por ejemplo, la sensación de seguridad por el alumbrado público, o la reducción de la enfermedad por el acceso a agua dulce, entre otros aspectos.

$$VPN = \sum_{a=1}^n \frac{R_a}{(1+i)^a} \quad (2-3)$$

Donde  $R_a$  es la entrada neta de efectivo de las salidas durante un período  $a$ ,  $i$  es la tasa de descuento,  $n$  es el número de períodos evaluados.

### 2.2.3 Ponderación de dimensiones

Los autores utilizan la metodología AHP para ponderar las dimensiones. Los siguientes pasos ajustaron la metodología al caso específico (Saaty, 1990, 2008):

- Definir el problema y el objetivo final. En este caso, no existe un panel de expertos que determine la importancia de un criterio sobre otro. Por lo tanto, el procedimiento comienza con la formulación de seis objetivos de priorización, denominados O1 a O6 (véase la Tabla 2-6), evaluados posteriormente. A saber, cuando se está analizando un objetivo, éste tiene más relevancia sobre los otros que mantienen la misma relevancia en el proceso de evaluación. Este procedimiento abarca todos los objetivos de priorización uno por uno.
- Organizar las decisiones jerárquicamente a partir del objetivo, seguido de los niveles medios compuestos por los criterios, subcriterios y alternativas. En este caso, los criterios son las dimensiones establecidas para la evaluación de la comunidad y del proyecto, y no hay subcriterios.
- Crear matrices de comparación por pares y comparar los criterios entre ellos. En este caso, hay seis matrices; cada una corresponde a un objetivo de priorización. Los elementos de nivel superior de la matriz corresponden a los objetivos, y los elementos de evaluación corresponden a los criterios. Una comparación entre ellos, manteniendo los valores de escala Saaty y las recomendaciones presentadas en la Tabla 2-2, da como resultado una matriz cuadrada con una diagonal de unos y los valores de las comparaciones en las otras posiciones de la matriz. Esta matriz debe seguir los principios de reciprocidad, homogeneidad y consistencia que se explican a continuación.
- Usar los resultados de priorización obtenidos de las matrices por pares para ponderar los niveles inmediatamente subsiguientes. El vector propio de la matriz por pares da el nivel de prioridades; cada elemento de este vector corresponde al peso de cada criterio para evaluar las alternativas. Por último, el orden de priorización de las alternativas resulta de la suma de términos exponentiales donde cada base es el valor de la dimensión y el exponente es un elemento de vector propio.

La expresión  $n(n-1)/2$  proporciona el número de comparaciones, donde  $n$  corresponde a los criterios, como se muestra en la Ecuación 2-4. En este caso, como hay seis criterios, el número de comparaciones

**Tabla 2-2.** Escala de Saaty

Valor	Definición	Descripción
1	Misma importancia	Dos actividades contribuyen igualmente a alcanzar el objetivo
3	Importancia moderada	Una actividad contribuye un poco más a alcanzar el objetivo sobre otra
5	Importancia fuerte	Una actividad contribuye activamente más a alcanzar el objetivo sobre otra
7	Importancia muy fuerte y demostrada	Una actividad contribuye mucho más a alcanzar el objetivo sobre otra
9	Importancia extrema	La evidencia absoluta favorece una actividad sobre otra
2, 4, 6, 8	Valores intermedios	Cuando se necesita un acuerdo entre las partes
Reciprocidad	$a_{ij}=1/a_{ji}$	La hipótesis del método

Fuente: (J. Pacheco & Contreras, 2008, p. 51)

es quince.

$$A = [a_{ij}] = \begin{pmatrix} a_1/a_1 & a_1/a_2 & \dots & a_1/a_n \\ a_2/a_1 & a_2/a_2 & \dots & a_2/a_n \\ \vdots & \vdots & & \vdots \\ a_n/a_1 & a_n/a_2 & \dots & a_n/a_n \end{pmatrix} \quad (2-4)$$

con:

$$1 \leq i, j \leq n \quad (2-5)$$

Donde  $a_{ij}$  corresponde a la comparación por pares entre criterios  $i$  y  $j$ ,  $n$  corresponde al número de criterios analizados.

La matriz debe cumplir con las propiedades de reciprocidad, homogeneidad y consistencia (Aznar & Guijarro, 2012; J. Pacheco & Contreras, 2008):

- Reciprocidad:  $a_{ij} = x$  entonces  $a_{ji} = 1/x$  con  $1/9 \leq x \leq 9$
- Homogeneidad:  $i$  y  $j$  tienen la misma importancia entonces  $a_{ij} = a_{ji} = 1$
- Consistencia: si  $a_{ij}$  representa la importancia del criterio  $i$  sobre  $j$ , y  $a_{jk}$  representa la importancia del criterio  $j$  sobre  $k$ , idealmente  $a_{jk} * a_{kj} = a_{ij} \forall 1 \leq i, j, k \leq n$ ; sin embargo, como hay evaluaciones subjetivas en la matriz de comparaciones, el proceso es iterativo hasta lograr una Consistency Ratio (CR) de  $A$  menor a 0,1.

Para calcular CR, se usa el siguiente proceso:

$$\lambda_{max} = V * B \quad (2-6)$$

Donde  $\lambda_{max}$  es el valor propio máximo de la matriz de comparación por pares  $A$ ,  $V$  es el vector o eigenvector de prioridades obtenido de la matriz  $A$ ,  $B$  es el vector de fila que contiene la suma de los elementos de columna de la matriz  $A$ .

El último resultado es útil para calcular el *Consistency Index* (CI) que representa la varianza del error de estimación de  $a_{ij}$ :

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (2-7)$$

Por último, CR es igual a la relación entre CI y RI, como muestra la Ecuación 2-8

$$CR = \frac{CI}{RI} \quad (2-8)$$

Donde  $CI$  es el índice de consistencia,  $RI$  es el índice aleatorio dado por Saaty.

El último paso consiste en calcular el valor de priorización final que vincula los resultados de la comunidad, el proyecto y la ponderación usando la Ecuación 2-9. Debido a que un proyecto tiene más de una comunidad asociada, la evaluación de las dimensiones de la comunidad es el promedio de todas las comunidades involucradas en el proyecto.

$$I = \sum_{i,j=1}^{n,m} D_i^{V_j} \quad (2-9)$$

## 2.2.4 Directriz para trabajo comunitario

El objetivo es estimular la participación activa de la comunidad en el proyecto para desarrollarlo sin oposición. Esto se puede lograr permitiendo a la población sentirse parte del proyecto, donde pueden ser empleados en trabajos técnicos básicos y en el proceso de toma de decisiones. La formulación de la directriz tomó elementos procedentes de la metodología de co-construcción y la información de las entidades públicas de Nariño. La metodología de co-construcción busca empoderar a la comunidad a través de su papel en los procesos de toma de decisiones para generar una solución sostenible a lo largo del tiempo (Montedonico et al., 2018). Es importante señalar que el carácter distintivo de la población exige que se ajuste la directriz a sus condiciones. En consecuencia, esta propuesta constituye una guía marco para la interacción de la comunidad con tres pasos: armonización, integración y sostenibilidad. El primero busca definir un propósito común en el equipo interdisciplinario y caracterizar a la comunidad. El segundo se ocupa de la integración de la comunidad, las partes interesadas y el equipo de trabajo, y el último intenta establecer un esquema de gestión local para mantener el proyecto a lo largo del tiempo.

A diferencia de la metodología de co-construcción, esta directriz se asegura de que todos los que forman parte del equipo de trabajo tengan el mismo objetivo y estén de acuerdo con lo que hay que hacer en la comunidad. Este equipo de trabajo debe ser interdisciplinario, incluyendo profesionales de varios campos de estudio. Además, la directriz destaca la importancia de involucrar a las empresas locales en la ejecución del proyecto, reduciendo los costos de operación y mantenimiento, y facilitando la interacción con la comunidad. Además, esta directriz fomenta la creación de un consejo de regulación local para formular iniciativas que superen las barreras y faciliten el desarrollo del proyecto. Del mismo modo, esta junta local considerará las tradiciones culturales comunitarias y funcionará como una entidad reguladora.

## 2.3 Descripción del caso de estudio

El departamento de Nariño se encuentra en la región suroeste de Colombia. La población de Nariño tiene problemas derivados de un conflicto armado que comenzó hace más de cinco décadas. Entre las consecuencias del conflicto se encuentran los desplazamientos forzados, despojo de tierras para plantaciones de hoja de coca, amapola y palma de aceite y la extracción de minerales de forma ilegal entre otros aspectos (Ávila et al., 2014). La población más afectada ha sido la de las zonas rurales, provocando un impacto significativo en el departamento, ya que el 53 % de la población total del vive en estas zonas. Según los datos del DANE, en 2018, la población rural de Nariño tenía una taza de necesidades básicas insatisfechas del 26,61 %, mientras que las personas de las zonas urbanas tenía una taza de 16,20 % (Departamento Administrativo Nacional de Estadística - DANE, 2018d). Respecto a los servicios públicos, en el mismo año el 13,5 % de las familias no tenían acceso a servicio de electricidad, el 29,2 % no tenía acceso a acueducto y el 53,1 % no tenía alcantarillado. Incluso en San Juan de Pasto, la capital de Nariño, hay una red de gas natural que solo cubre el 22,15 % de la población, el resto de la población utiliza cilindros con Gas Licuado de Petróleo (GLP) incrementando el riesgo de accidentes en los hogares (Ministerio de Minas y Energía, 2018). En cuanto a infraestructura, el departamento de Nariño ha sido clasificado como medio-alto por la CEPAL. Esta categorización consideró el acceso a los servicios públicos, las tecnologías de la información y las comunicaciones, las carreteras, los aeropuertos y los puertos. La misma institución dio un puntaje medio-bajo en competitividad al departamento en comparación con otros departamentos en Colombia (Ramírez J. et al., 2017).

**Tabla 2-3.** Clasificación de localidades por horas de suministro eléctrico

Horas	Localidades
0	14
1-6	438
7-12	144
13-18	1
19-23	0
24	6
Sin información	36
Total	609

Fuente: (Arbeláez Pérez, 2019; Mossos, 2019)

En 2018, había 516.398 viviendas, de las cuales 480.134 tenían electricidad, y 36.264 no tenían este

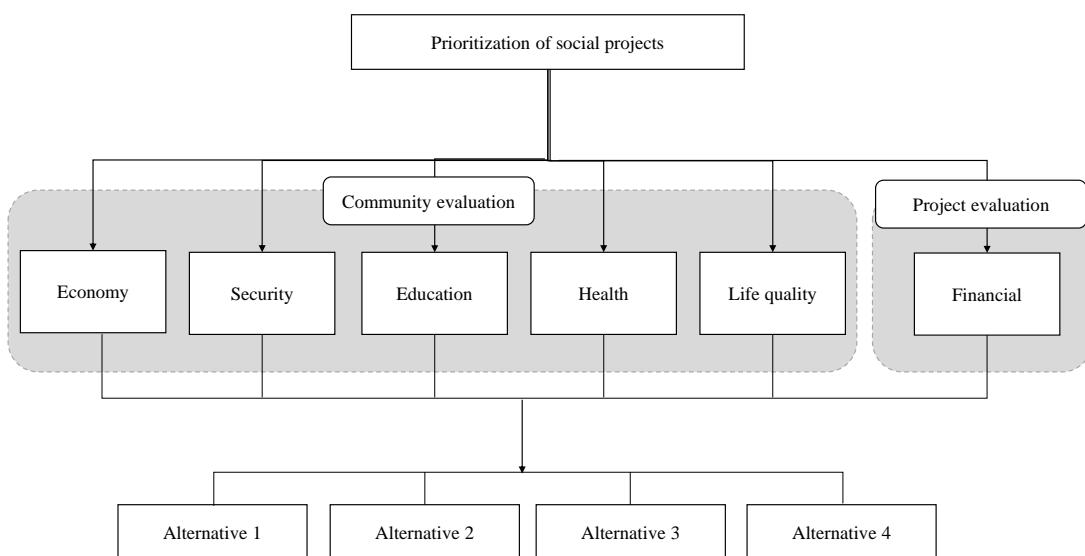
servicio. De todos los consumidores de electricidad, el 88,33 % eran usuarios Conectados al SIN, y el 11,66 % eran usuarios no interconectados. Del 11,66 % de usuarios fuera de la red, el 8,35 % vivía en zonas urbanas y el 91,64 % eran usuarios rurales (Unidad de Planeación Minero Energética, 2019). Los usuarios de las zonas no interconectadas obtiene el suministro de electricidad a través del uso de centrales eléctricas municipales que utilizan combustibles fósiles como el diésel. En cuanto a los indicadores de servicio, el 71,49 % de las localidades sin telemetría tienen electricidad de 1 a 6 horas al día, mientras que el 5,91 % restante no dispone de información sobre el suministro de electricidad (ver Tabla 2-3) (Arbeláez Pérez, 2019). Las localidades con telemetría, en promedio tienen siete horas diarias de suministro de electricidad (Mossos, 2019). Además, el precio promedio de la electricidad es de unos 0,33<sup>2</sup> US\$ kWh (Superintendencia de Servicios Públicos Domiciliarios SSPD, 2019).

## 2.4 Resultados obtenidos

### 2.4.1 Priorización de proyectos

La Figura 2-2 muestra el esquema jerárquico para la toma de decisiones sobre las alternativas en el departamento de Nariño. Es importante recordar que las dimensiones corresponden a los criterios que evalúan seis objetivos de priorización. Los criterios son: educación, salud, calidad de vida, seguridad, y economía, para la evaluación de la comunidad y finanzas para la evaluación del proyecto.

**Figura 2-2.** Esquema de priorización de proyectos



<sup>2</sup>1 COP \$ es equivalente a 0,00027 US\$

En esta investigación se seleccionaron los proyectos cuyo insumo fuera la energía solar. Esta investigación es parte de un proyecto más grande que busca promover proyectos con energía solar; por lo tanto, es esencial ofrecer elementos para crear una cadena de valor a su alrededor. Los siguientes fueron los proyectos seleccionados entre las alternativas presentadas en PERS:

- Electrificación escolar: electrificación de 256 escuelas ubicadas en 11 municipios de la zona del Pacífico, utilizando tecnología solar fotovoltaica. El objetivo del proyecto es poner en funcionamiento la sala de cómputo y, en el futuro, una red Ethernet en cada escuela. Para ello, el proyecto busca instalar plantas de energía solar con una capacidad de entre 1,5 y 2 kW (Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE, 2014b).
- Electrificación rural: electrificación de la comunidad rural de El Sande utilizando kits solares. La comunidad está compuesta por 364 habitantes que viven en 54 viviendas. Dentro de la información disponible no fue posible determinar si la comunidad tenía una red de distribución. Como consecuencia, se propuso utilizar un kit solar de 2 kW con un sistema de baterías que ofreciera una autonomía de tres días, porque se trata de una zona no interconectada (Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE, 2014c).
- Alumbrado solar: el proyecto propone instalar 213 sistemas de iluminación solar cada 18 metros a lo largo de las calles del municipio de San Lorenzo. Cada sistema de iluminación solar de 40 W equivale a un sistema de luz convencional de 70 W. Además, cada uno tiene una autonomía de tres días a través de un sistema de almacenamiento de energía (Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE, 2014d).
- Sistemas de bombeo de agua por energía solar: sistemas de bombeo de agua que utilizan energía solar para irrigar cultivos y proporcionar agua a la población en el municipio de Taminango. La bomba mueve  $16\text{ m}^3/\text{da}$  de agua para cubrir las necesidades de riego y el consumo por parte del ser humano. El número de sistemas considerados corresponde al número de propiedades de entre 1 y 5 hectáreas reportados en las bases de datos institucionales (2.304) (Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE, 2014a)

El procedimiento de la sección de metodología llevó a los autores a normalizar cada componente de los valores de la Tabla 2-1 utilizando la Ecuación 2-1 y calcular los índices de las dimensiones utilizando la Ecuación 2-2. Para este cálculo se consideró la información de los 64 municipios. Las dos primeras filas de la Tabla 2-4 indican los mejores y peores valores de condición para cada dimensión. Por ejemplo, la mejor condición para las dimensiones de educación, salud y economía se produce cuando el resultado es uno (Oficina de planeación educativa, 2018; DANE & Banco de la República de Colombia, 2016; Departamento Administrativo Nacional de Estadística - DANE, 2017, 2018b,c,e,a,d, 2016, 2005; Instituto Departamental de Salud de Nariño, 2018).

La Tabla 2-4 muestra los resultados de la evaluación comunitaria en los municipios vinculados a los proyectos. San Lorenzo tiene un mejor indicador (0,49) que Roberto Payán (0,09) en la dimensión ed-

ucativa. En cuanto a la salud, Tumaco presenta mejores condiciones (0,69) que Santa Cruz (0,13). Sin embargo, Tumaco tiene las peores condiciones de seguridad en el departamento (1,00). En la dimensión de la calidad de vida, la población de La Tola presenta la peor situación (0,95). Tumaco tiene la puntuación más alta (0,53) en importancia económica, porque es una ciudad portuaria, y Taminango muestra la peor posición (0,19). Estos resultados muestran que los municipios ubicados en la zona del Pacífico son los más afectados en las dimensiones evaluadas.

**Tabla 2-4.** Resultados de la evaluación de la comunidad

	Educación	Salud	Seguridad	Calidad de vida	Economía
Mejor condición	1	1	0	0	1
Peor condición	0	0	1	1	0
Municipio	Educación	Salud	Seguridad	Calidad de vida	Economía
Barbacoas	0,28	0,68	0,97	0,76	0,52
Cumbitara	0,30	0,38	0,67	0,74	0,25
El Charco	0,15	0,52	0,49	0,89	0,38
La Tola	0,19	0,49	0,22	0,95	0,36
Mosquera	0,17	0,28	0,12	0,91	0,44
Olaya Herrera	0,17	0,49	0,55	0,77	0,44
Francisco Pizarro	0,24	0,34	0,17	0,82	0,34
Policarpa	0,39	0,31	0,87	0,39	0,30
Roberto Payán	0,09	0,53	0,86	0,80	0,48
Tumaco	0,26	0,69	1,00	0,62	0,53
Santa Bárbara	0,19	0,68	0,31	0,34	0,44
Santacruz	0,23	0,13	0,11	0,67	0,22
San Lorenzo	0,49	0,44	0,04	0,53	0,22
Taminango	0,47	0,40	0,10	0,50	0,19
Mejor indicador					
Peor indicador					

El VPN fue el criterio elegido para la evaluación de los proyectos. En este paso, los autores realizaron un análisis económico de cada proyecto, teniendo en cuenta todos los costes de inversión y una tasa de descuento del 10 %, evaluada a lo largo de seis años. La Tabla 2-5 muestra los resultados normalizados de la evaluación financiera de los proyectos. Los proyectos de electrificación rural y alumbrado público solar tienen el mismo valor normalizado, aunque el VPN es ligeramente diferente. Es imperativo señalar que el proyecto de bombeo de agua utilizando energía solar es el más caro porque cubriría a 2.304 beneficiarios, correspondientes al total de propiedades rurales.

El siguiente paso fue calcular las ponderaciones para cada dimensión con el método AHP. En este caso, dado que no existe un panel de expertos para determinar la importancia de una dimensión sobre otra, se

**Tabla 2-5.** Normalized NPV

Alternatives	Normalized score
School electrification	0,91
Rural electrification	1,00
Solar public lighting	1,00
Solar water pumping	0,00

presentan seis objetivos de priorización, denominados O1 a O6. Cada objetivo enfatiza una dimensión sobre las demás y la Tabla 2-6 muestra los resultados de los ponderados.

La Tabla 2-7 presenta los resultados de la priorización de los proyectos por objetivos. La electrificación de las escuelas es la principal alternativa de los objetivos de priorización relacionados con los objetivos financieros (O1), calidad de vida (O3), seguridad (O4) y educación (O6). Los resultados son coherentes porque los municipios involucrados en este proyecto mostraron los indicadores más bajos en las dimensiones consideradas. Si el objetivo es mejorar la calidad de vida, la seguridad y la educación, los proyectos más importantes para lograr estos objetivos son los de electrificación de las escuelas. Además, si el presupuesto disponible determina la decisión, la electrificación de las escuelas también es una buena opción porque el valor normalizado del VPN del proyecto lo coloca en tercer puesto, lo que significa que el proyecto no es el más costoso.

**Tabla 2-6.** Ponderación por objetivos de priorización

	O1	O2	O3	O4	O5	O6
Educación	0,083	0,083	0,083	0,083	0,083	0,583
Salud	0,083	0,083	0,083	0,083	0,583	0,083
Seguridad	0,083	0,083	0,083	0,583	0,083	0,083
Calidad de vida	0,083	0,083	0,583	0,083	0,083	0,083
Economía	0,083	0,583	0,083	0,083	0,083	0,083
Financiero	0,583	0,083	0,083	0,083	0,083	0,083
Total	1,000	1,000	1,000	1,000	1,000	1,000

Los objetivos relacionados con los objetivos economía (O2) y de salud (O5) priorizan el proyecto de electrificación para la reserva indígena. Estos objetivos de priorización ponen el énfasis de la importancia en las dimensiones económica y salud de la población sobre las demás dimensiones. La información recopilada muestra que la comunidad de Santacruz tiene las peores condiciones de salud entre las comunidades involucradas en las alternativas evaluadas (ver Tabla 2-4). En la dimensión económica, los indicadores muestran que las comunidades de San Lorenzo y Santacruz están en las mismas condiciones; sin embargo, Santacruz tiene peores condiciones que San Lorenzo respecto a las dimensiones de seguridad, salud y educación. Lo anterior, nos permite concluir que la electrificación mejora las condiciones económicas y de salud de la población. Los resultados son consistentes porque la electricidad facilita el uso de herramientas de trabajo, dispositivos eléctricos como refrigeradores para preservar alimentos y dispositivos de entretenimiento, entre otros.

Para todos los objetivos de priorización, el proyecto de menor importancia es el bombeo de agua con energía solar. Como la comunidad evaluada con este proyecto tiene mejores indicadores que otras comunidades, excepto en la dimensión económica, los resultados fueron coherentes. El proyecto de

alumbrado público que utiliza energía solar es la tercera prioridad ya que la comunidad evaluada para este proyecto tiene mejores condiciones que algunas otras comunidades evaluadas en las dimensiones de educación y seguridad.

Los proyectos de electrificación escolar son una prioridad en cuatro objetivos de priorización propuestos. Los autores hicieron un análisis más detallado de este proyecto para entender por qué este proyecto es una prioridad en el departamento. El proyecto de electrificación de la escuela fue desagregado en proyectos por municipio, y cada proyecto fue evaluado con la misma ponderación de la Tabla 2-6.

**Tabla 2-7.** Priorización de proyectos

Project	O1	O2	O3	O4	O5	O6
Bombeo de agua	4,66	4,56	4,38	4,09	4,44	4,40
Electrificación rural	5,74	5,63	5,57	5,18	5,67	5,62
Alumbrado público	5,58	5,47	5,32	4,97	5,34	5,32
Electrificación escuelas	5,76	5,59	5,66	5,54	5,55	5,69

El método de cuantificación de beneficios del proyecto siguió el Manual de Evaluación de Beneficios y Sugerencias de Cuantificación del Departamento Nacional de Planificación (Dirección de Inversiones y Finanzas Públicas, 2006). El manual sugiere los siguientes supuestos considerados en esta investigación:

- La electrificación de la escuela aumenta tanto el acceso a la información como la calidad de la educación.
- Reducción de la deserción escolar y una mayor probabilidad de alcanzar el nivel de educación media para los estudiantes.
- Incrementa la posibilidad de que un estudiante que termine la educación media en el futuro reciba un salario mínimo legal mensual.

De acuerdo con las tres consideraciones anteriores, los beneficios del proyecto incluían los ingresos anuales para aquellos con un diploma de escuela secundaria y el rescate de unas diez computadoras portátiles por escuela que están en desuso porque no hay electricidad.

La Tabla 2-8 muestra los resultados de la evaluación del proyecto derivados del VPN y la Tasa de Retorno Interno (IRR) para cada municipio. Los resultados están de acuerdo con las economías de escala, ya que el proyecto más grande tuvo la IRR más alta, mientras que el proyecto más pequeño tuvo una IRR negativa. Por lo tanto, los resultados presentaron el punto de equilibrio con la implementación de al menos diez escuelas.

La misma Tabla 2-8 también muestra los resultados de los objetivos de priorización. Bajo los objetivos analizados, el proyecto de electrificación en el municipio de Tumaco tiene el resultado más bajo respecto a los demás municipios. Esto último es concordante con los aspectos económicos del municipio que lo colocan en una mejor posición que los demás de la zona del Pacífico.

En el objetivo O1, el factor más relevante es el financiero. Sin embargo, el orden de ejecución de los

proyectos no es igual a la orden de estandarización VPN. Es porque la evaluación del proyecto incluye otras variables de decisión: salud, educación, seguridad, entre otras. Cuando la economía es la dimensión principal, como muestra objetivo O2, el proyecto de electrificación escolar en Cumbitara tiene el primer lugar, porque este municipio tiene una de las economías más bajas entre las comunidades evaluadas. Por lo tanto, la ejecución de este proyecto mejoraría las condiciones económicas. En el objetivo O4, cuando la dimensión de seguridad lidera el análisis, la electrificación escolar en el municipio de Barbacoas es una prioridad, porque tiene la segunda condición en esta dimensión después del municipio de Tumaco. Como Tumaco tiene mejores condiciones en las otras dimensiones que Barbacoas, éste tomó la primera posición para la dimensión O4.

**Tabla 2-8.** Evaluación financiera del proyecto de electrificación escolar

Proyecto	Numero de escuelas	NPV izado	Normal- izado	IRR	O1	O2	O3	O4	O5	O6
Tumaco	53	0		16,77	4,782	4,487	4,577	4,782	4,377	4,648
Roberto Payán	45	0,26		16,09	5,303	5,473	5,634	5,670	5,443	5,695
Olaya Herrera	40	0,36		15,11	5,363	5,494	5,612	5,487	5,463	5,644
El Charco	24	0,69		12,19	5,626	5,587	5,732	5,504	5,503	5,716
Cumbitara	19	0,79		9,6	5,720	5,701	5,696	5,653	5,628	5,671
La Tola	17	0,83		8,11	5,664	5,560	5,728	5,290	5,485	5,653
Santa Bárbara	15	0,86		6,43	5,585	5,416	5,276	5,252	5,261	5,555
Francisco Pizarro	14	0,88		5,47	5,685	5,561	5,651	5,234	5,565	5,620
Barbacoas	11	0,94		1,85	5,764	5,503	5,668	5,779	5,397	5,647
Policarpa	10	0,96		0,32	5,790	5,652	5,461	5,741	5,642	5,599
Mosquera	8	1		-3,43	5,743	5,500	5,695	5,198	5,597	5,655

Las celdas de la Tabla 2-9 contienen un número que muestra la misma clasificación en el orden de ejecución para ambos objetivos de priorización. Esta matriz de correlación es una herramienta útil para los responsables de la toma de decisiones que deben determinar la fusión de objetivos. Otra característica es que la diagonal de la matriz corresponde a los municipios evaluados. El resultado más destacado es la correlación entre los objetivos O3 (calidad de vida) y O6 (educación) porque ambos objetivos tienen seis proyectos en el mismo ranking en la ejecución de las alternativas.

En otras palabras, mejorar la educación significa mejorar la calidad de vida y viceversa.

Además, cuando la educación mejora, la economía de la población también progresará, como lo confirma la correlación entre los objetivos O2 y O6. Este resultado también confirma la suposición de calcular

**Tabla 2-9.** Matriz de correlación en los proyectos

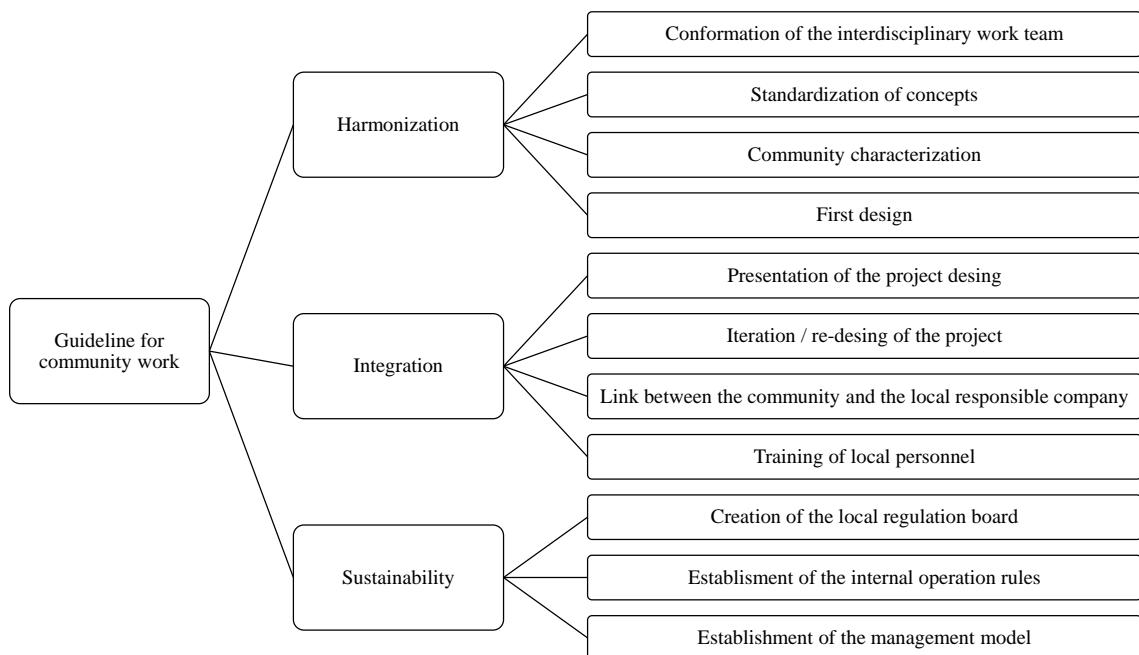
	O1	O2	O3	O4	O5	O6
O1	11	1	1	3	4	1
O2	1	11	3	2	3	4
O3	1	3	11	1	2	6
O4	3	2	1	11	2	1
O5	4	3	2	2	11	2
O6	1	4	6	1	2	11

el IRR donde las inversiones en educación aumentan los ingresos de la población. En consecuencia, una investigación detallada de estas correlaciones apoyaría la iniciativa sectorial para los sectores de la educación y la economía, impactando la calidad de vida e impulsando el desarrollo social. En futuras investigaciones, podría ser interesante insertar más dimensiones sobre las comunidades y los proyectos para contribuir a la formulación de políticas socio-energéticas más amplias

## 2.4.2 Directriz para el trabajo comunitario

Basándose en la metodología de co-construcción, la información poblacional de Nariño registrada por entidades públicas, e investigaciones de campo en la localidad de Huatacondo. Los autores proponen una directriz en tres etapas para promover el trabajo conjunto entre las entidades de financiación y las comunidades. La Figura 2-3 sintetiza la directriz para fomentar proyectos de tecnología solar fotovoltaica para la electrificación.

**Figura 2-3.** Guía para el trabajo comunitario



El primer paso en la etapa de armonización se centra en la formación del grupo de trabajo interdisciplinario para definir el objetivo del proyecto energético. El segundo paso busca estandarizar conceptos y términos para utilizar el mismo lenguaje entre los miembros del grupo de trabajo. Este paso es esencial porque los miembros del grupo pueden tener diferentes ideas sobre proyectos exitosos, sostenibles y pertinentes, e impactos positivos y negativos derivados de ellos. En consecuencia, antes de diseñar una solución o interactuar con la comunidad, es crucial aclarar tanto el objetivo como los impactos esperados del proyecto. El trabajo conjunto fortalece el proyecto porque la población conocería su propósito e impactos; de la misma manera, ellos deben exponer sus preocupaciones. El tercer paso

consiste en caracterizar a la comunidad desde el punto de vista social, económico y cultural. Con el fin de reconocer las necesidades de la comunidad y ofrecer soluciones adecuadas. El último paso de la primera etapa consiste en la elaboración del primer diseño del proyecto, teniendo en cuenta una visión técnica, económica y social. En este paso se usa información primaria o secundaria del sitio donde se implementará el proyecto.

La etapa de integración comienza con la presentación del diseño del proyecto a la comunidad. Esta socialización es esencial porque estimula un diálogo en torno a los aspectos cruciales del proyecto que podrían ser preocupantes. Esos espacios deben propiciar los cambios que eventualmente rediseñarían el proyecto, incluyendo opiniones y propuestas de la comunidad. Aunque no es necesario profundizar en debates técnicos, la comunidad debe sentir que es una parte fundamental del proceso de toma de decisiones en el proyecto. El tercer paso implica la búsqueda de la empresa a cargo. Idealmente, la empresa debe ser local porque facilita la ejecución del proyecto, las labores de operación y mantenimiento y la relación con la comunidad. Sin embargo, cuando no hay empresas locales para desarrollar el proyecto, se sugiere involucrar a un organismo de supervisión local que pueda ser constituido por miembros académicos, por ejemplo, un equipo de investigación universitaria, población local, entre otros, para supervisar imparcialmente las actividades desarrolladas por la empresa elegida. El último paso consiste en capacitar al personal local para realizar operaciones básicas y actividades de mantenimiento. Esta directriz tiene la intención de minimizar la dependencia del personal externo para llevar a cabo tareas rutinarias y reducir los costos de operación y mantenimiento.

La etapa de sostenibilidad es la última y busca mantener el proyecto en el tiempo. Por lo tanto, la comunicación entre la entidad de financiación, la empresa elegida y la comunidad debe ser constante. El primer paso consiste en crear una junta reguladora local; esta entidad debe ser legítima para la comunidad y actuar de acuerdo con sus costumbres y cultura. En el segundo paso, las entidades reguladoras a nivel nacional ayudarían a la junta reguladora local a establecer normas, derechos y obligaciones concordantes con la reglamentación y las leyes nacionales. El último paso crea un modelo de gestión para la operación a largo plazo en el que la empresa y la comunidad tienen responsabilidades, porque formarían parte de la estructura organizativa.

## 2.5 Conclusiones

Este documento propuso una metodología para seleccionar e implementar proyectos en las comunidades de Nariño. La metodología de priorización de la investigación evalúa las áreas relevantes de las comunidades y los costos de implementación de proyectos basados en energía solar. Los criterios para evaluar las comunidades consideraron el SDGI y el IDH, y el criterio para evaluar los proyectos consideró el VPN. Es importante destacar que la metodología de priorización presentó una combinación de aquellos componentes que alimentaron la metodología AHP para tener un orden de clasificación. Además, la directriz de la comunidad hizo hincapié en su integración en las actividades de diseño y operación del

proyecto, aumentando sus posibilidades de éxito. En el caso de estudio, los autores consideraron las características de Nariño e identificaron áreas relevantes para evaluar comunidades. Las dimensiones más cercanas a los aspectos socioeconómicos de las comunidades eran la educación, la salud, la calidad de vida, la economía y la seguridad. Sin embargo, es imperativo incluir más dimensiones para comprender mejor las condiciones de vida de la población e incluir los aspectos culturales de las comunidades.

Los autores consideraron el VPN para evaluar proyectos, y el criterio de selección resaltó los proyectos mostrando los costos más bajos. Sin embargo, este indicador no reconoce de forma exclusiva los beneficios sociales derivados de los proyectos. Por lo tanto, es recomendable utilizar cuantificadores adicionales que midan los beneficios económicos de los proyectos sociales, motivando la atención de los posibles inversionistas para futuras obras.

Las alternativas evaluadas con esta metodología fueron cuatro proyectos. Los resultados mostraron que, en cuatro de los seis objetivos de priorización, la electrificación de la escuela es la prioridad, y en dos objetivos, la prioridad era la electrificación de una reserva indígena. Dados los resultados anteriores, los autores decidieron realizar un análisis más detallado del proyecto de electrificación de la escuela. El análisis mostró que, si el IRR es una variable de decisión, la ejecución debe tener al menos diez proyectos. Por otro lado, los resultados mostraron una correlación entre la educación y la calidad de vida porque los objetivos O3 y O6 tienen seis proyectos en el mismo orden de ejecución; por lo tanto, la realización de proyectos educativos mejora la calidad de vida y viceversa. Este es un hallazgo relevante para mostrar a los responsables de la toma de decisiones, ya que podrían realizar más estudios que confirmen esta constatación y unifiquen los esfuerzos en las políticas públicas del sector energético para fortalecer los resultados.

La metodología sugerida por los autores identifica el orden de ejecución óptimo de los proyectos sociales de acuerdo con los objetivos de priorización. Además, la directriz involucra a la comunidad como parte activa de la solución, mejorando la sostenibilidad del proyecto a lo largo del tiempo y reduciendo la dependencia externa del personal. Por lo tanto, empodera a la comunidad y promueve su desarrollo social, económico, educativo y cultural, entre otros. También es esencial destacar el papel principal del consejo regulador local para lograr su éxito. Por lo tanto, la investigación futura debe analizar aspectos como la regulación, los miembros de la junta, y los mecanismos que estos eligen.

# **3 Evaluación de modelos empíricos para estimar irradiancia solar global usando datos de temperatura en un ambiente tropical y montañoso. Parte I: imputación<sup>1</sup>**

## **Resumen**

La irradiancia solar es un recurso ampliamente disponible que podría contribuir al proceso de electrificación en lugares con bajos índices socio económicos. No obstante, en algunos lugares la información de este recurso no está disponible o tiene baja calidad. Para superar este problema algunos investigadores han desarrollado técnicas como modelos empíricos basados en temperatura para estimar el recurso. Sin embargo, no hay un amplio análisis del comportamiento de esas técnicas en ambientes tropicales y montañosos. Por lo tanto, esta investigación analiza el comportamiento de tres modelos empíricos basados en temperatura y un modelo propuesto bajo estas condiciones ambientales. Los errores estadísticos calculados permiten elegir el mejor modelo para cada punto evaluado. Con este modelo se hace la imputación de datos con el fin de incrementar la calidad de las bases de datos analizadas. El modelo propuesto se ajusta mejor a la zona Andina y amazónica, mientras el modelo de Hargreaves y Samani tiene mejores resultados en la zona Pacífica. Además, el modelo propuesto presenta una relación lineal entre las constantes empíricas y la altitud de las estaciones meteorológicas localizadas por encima de los 2.500 msnm.

**Palabras clave** Irradiancia solar, Imputación de datos, Bristow y Campbell, Hargreaves y Samani, modelos empíricos basados en temperarura

---

<sup>1</sup>Este artículo tuvo la participación de Belizza Ruiz

### 3.1 Introducción

Pasaron algunas décadas antes de que la energía solar fuera viable económicamente en los países en desarrollo. Actualmente, los costos asociados al aprovechamiento de este recurso energético son competitivos bajo características específicas. Tradicionalmente, el suministro de electricidad para zonas rurales o aisladas se ha hecho a través de la extensión de los sistemas de transmisión o distribución. Sin embargo, la baja densidad poblacional hace que ésta sea una opción poco rentable. En consecuencia, las regiones aisladas no cumplen los requisitos de rentabilidad. Aunque las zonas aisladas tienen baja densidad de población, el Estado tiene el deber de ofrecer servicios de electricidad para toda su población. Por lo tanto, el deber del Estado podría estar satisfecho con la promoción de energías no convencionales como la energía solar. Sin embargo, antes de fomentar la generación de electricidad con energía solar, es fundamental entender el comportamiento de los recursos. En los países en desarrollo, los datos de irradiacia solar no están disponibles debido a la escasez de estaciones meteorológicas que miden esta variable y a los requisitos de calibración y mantenimiento de los equipos de medición. En algunos casos, aunque las bases de datos tienen acceso abierto, el proceso para obtener información puede ser largo (Bakirci, 2009). Por lo tanto, en los últimos años, los estudios de irradiancia solar han utilizado métodos como los empíricos que utilizan variables como la temperatura, el brillo solar, la humedad relativa, entre otros, para estimar la irradiancia solar.

Los investigadores han desarrollado modelos y métodos para estimar la insolación solar, en diferentes escalas de tiempo, desde modelos empíricos hasta inteligencia artificial. La simplicidad, aceptación, adaptabilidad y bajo costo computacional son ventajas de los modelos empíricos. Los modelos empíricos se basan generalmente en factores astronómicos, geométricos, físicos y meteorológicos, destacando los últimos (Besharat et al., 2013). Algunos de los factores meteorológicos son la nubosidad, la temperatura y el brillo solar que describen el estado del cielo; sin embargo, la nubosidad es el factor más determinante para conocer el comportamiento de insolación solar. De las tres variables mencionadas anteriormente, las variables más comúnmente medidas son el brillo solar y la temperatura (Benson et al., 1984); en consecuencia, estas variables son las más usadas para estimar la insolación solar (Bakirci, 2009). La implementación del modelo empírico depende de la disponibilidad y coherencia de los datos (Akinoglu, 2008). Almoroz, Hontoria & Benito recomiendan que los datos usados tengan simultaneidad y fiabilidad (Almoroz et al., 2011). Los datos de brillo solar no se miden con la misma frecuencia que los datos de temperatura en las estaciones meteorológicas. Por lo tanto, los modelos empíricos basados en la temperatura son una opción conveniente para estimar la insolación solar global (Fan et al., 2018).

Hargreaves y Samani presentaron en 1982 el primer modelo basado en la diferencia de temperatura diaria para estimar la insolación solar (Hargreaves & Samani, 1982). Bristow y Campbell propusieron un modelo empírico basado en la temperatura en 1984 (Bristow & Campbell, 1984). En 2011, Cheng y otros evaluaron el rendimiento de los modelos basados en la temperatura para China (Chen et al., 2011). En 2014, Li et al. presentó un modelo basado en la temperatura sustentado en el modelo Hargreaves y Samani para China (H. Li et al., 2014). Quansah et al. evaluaron modelos empíricos basados en la temperatura y brillo solar en Ghana (Quansah et al., 2014). Dos Santos et al. evaluaron nueve modelos

basados en la temperatura para Brasil (C. M. Dos Santos et al., 2014). En 2017, Rivero y otros validaron el modelo Hargreaves y Samani para México (Rivero et al., 2017). Jamil y Akhtar compararon modelos empíricos para los ambientes subtropicales y húmedos de la India (Jamil & Akhtar, 2017). Aunque varios estudios analizan el comportamiento de los modelos empíricos en diferentes lugares a nivel mundial, no hay la misma cantidad de investigaciones que analicen el comportamiento de los modelos basados en la temperatura en entornos tropicales y montañosos.

El propósito de esta investigación es evaluar tres modelos empíricos basados en la temperatura: Hargreaves y Samani (HS), Bristow y Campbell (BC), y Okundamiya y Nzeako (ON) para estimar la insolación solar en un entorno tropical y montañoso. La validación estadística define el mejor modelo porque evalúa su rendimiento y lo compara con los demás. Además, los autores proponen un nuevo método basado en el modelo de regresión logística para estimar la insolación solar con la diferencia de temperatura diaria. La información para los métodos empíricos provino de las Estaciones Meteorológicas Automáticas (AWS) del Instituto Hidrológico, Meteorológico y de Estudios Ambientales-IDEAM ubicadas en el departamento de Nariño<sup>2</sup>, que cuenta con tres zonas ambientales: Pacífico, Andino y Amazonia. Estos permiten evaluar los modelos en diferentes condiciones climáticas y fisiográficas. La base de datos se divide en dos partes aleatoriamente: la primera parte para la calibración del modelo y la segunda parte para validarla estadísticamente. Antes de utilizar los modelos empíricos, los datos pasan un procedimiento de control de calidad para mejorar la fiabilidad de los resultados. R-CRAN es el software utilizado para llevar a cabo el proceso computacional en esta investigación.

El artículo está organizado en las siguientes seis secciones: La Sección 3.2 describe los métodos y materiales. La subsección 3.2.1 describe las características del conjunto de datos. Las subsecciones 3.2.2 y 3.2.3 describen los procedimientos de control de calidad aplicados tanto a la irradiación solar como a los datos de temperatura. La subsección 3.2.4 presenta los modelos empíricos basados en la temperatura, mientras que la subsección 3.2.5 presenta el modelo propuesto por los autores. La subsección 3.2.6 evalúa la validación estadística de modelos empíricos. La sección 3.3 presenta los resultados y la discusión. Finalmente, la Sección 3.4 presenta las conclusiones.

## 3.2 Materiales y métodos

### 3.2.1 Descripción de los datos

El uso de la energía solar en los procesos de producción requiere un conocimiento preciso del comportamiento de irradiancia solar, incluyendo los estudios meteorológicos y orográficos de la ubicación y el análisis de los impactos socio-económicos en la población. La ubicación de Nariño favorece el aprovechamiento de la energía solar. Las coordenadas geográficas del departamento están entre una

---

<sup>2</sup>Para la administración del Estado, el territorio se divide en departamento, distritos, municipios, y territorios indígenas

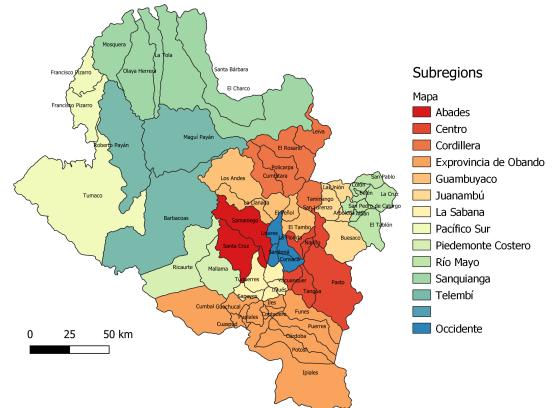
latitud  $00^{\circ}31'08''N$  y  $02^{\circ}41'08''N$  y una longitud  $76^{\circ}51'19''W$  y  $79^{\circ}01'34''W$ . En esta posición geográfica recibe una cantidad considerable de irradiancia solar durante todo el año, con una baja variación en comparación con otras latitudes. Lo anterior es un asunto notable porque los sistemas de energía solar no necesitarían rastreadores para alcanzar el punto de potencia máximo. Esta característica física destaca la reducción de los costes iniciales de inversión y mantenimiento y facilita la instalación de sistemas solares.

A pesar de la baja variabilidad de la irradiación solar, Nariño se encuentra en una Zona de Convergencia Intertropical dada su cercanía al Ecuador, lo que influye en el comportamiento meteorológico, provocando estaciones lluviosas unimodales o bimodales y una alta nubosidad debido a la alta humedad relativa en altitudes bajas. La última condición se incrementa porque las cordilleras conservan las masas de aire húmedo procedentes del Océano Pacífico. Consecuentemente, cuando las masas de aire chocan con la cordillera occidental, la estribación del Pacífico se vuelve húmeda y mantiene dicha humedad (Instituto Geográfico Agustín Codazzi - IGAC, 2014). Por lo tanto, el clima y la condición orográfica dividen a Nariño en tres regiones ambientales: el Pacífico, el Andino y el Amazonas (Martínez, 2018).

La región del Pacífico cubre el 52 % del territorio total y está constituida por las subregiones administrativas Telembí, Pacífico Sur, Sanquianga y Piedemonte Costero (ver la Figura 3-1). Esta región incluye la cuenca hidrográfica binacional Mira-Mataje y un bosque de manglares ubicado en la reserva natural de Sanquianga. Además, tiene dos zonas climáticas: las llanuras del Pacífico y las estribaciones del Pacífico. El primero tiene una humedad relativa superior al 80 %, temperaturas superiores a  $26^{\circ}C$ , y precipitaciones entre 3.000 y 5.000 mm/año. El segundo presenta una alta humedad, temperatura entre  $18^{\circ}C$  a  $24^{\circ}C$ , y precipitaciones entre 4.000 y 6.000 mm/año. Además, hay una zona situada entre las localidades de Junín y Barbacoas con precipitaciones de unos 9.000 mm/año (Gobernación de Nariño, 2016a; Instituto Geográfico Agustín Codazzi - IGAC, 2014).

La segunda zona es la región Andina, constituida por las subregiones La Sabana, Los Abades, Occidente, Cordillera, Centro, Juanambú, Río Mayo y Guambuyaco (ver Figura 3-1). En esta región, la cordillera andina se divide en las cordilleras occidental y central, cubriendo aproximadamente el 40 % del territorio del departamento (CORPONARIÑO, 2001; Instituto Geográfico Agustín Codazzi - IGAC, 2014). En la cordillera occidental se encuentran los volcanes Chiles, Cumbal y Azufral. En la cordillera central se encuentran los volcanes Galeras y Doña Juana, las llanuras altas de Túquerres-Ipiales y el Valle de Atriz. Esta región muestra un comportamiento de precipitación bimodal, entre 800 y 2.200 mm/año, con picos en abril-mayo y octubre-noviembre. En la cuenca hidrográfica de Río Mayo, Juanambú y Guáitara, hay temperaturas entre 16 y  $24^{\circ}C$  y precipitaciones entre 1.000 y 1.800 mm/año. En la zona norte

**Figura 3-1.** Subregiones administrativas



hacia Patía, la precipitación anual es inferior a 1.000 mm/año, con temperaturas superiores a 24 °C (Gobernación de Nariño, 2016a).

La tercera región es la cuenca hidrográfica amazónica, que cubre el 8 % del territorio de Nariño y está constituida por la subregión Exprovincia de Obando. Alrededor del 16,4 % de la población del departamento vive en esta área. Esta región tiene dos zonas climáticas: las zonas montañosas y las planicies. La zona montañosa se encuentra entre los ríos Patía y Putumayo, presenta temperaturas entre 6 °C y 11 °C, y recibe precipitaciones de unos 2.000 mm/año. La segunda zona tiene un clima tropical influenciado por la selva nublada y presenta precipitaciones entre 500 y 1.500 mm/año (Gobernación de Nariño, 2016a). Es conveniente destacar que la actividad minera para extraer minerales preciosos es una actividad importante y lucrativa; de hecho, alrededor del 98,38 % de los minerales extraídos en el departamento provienen de esta región (Instituto Geográfico Agustín Codazzi - IGAC, 2014).

En Nariño hay ocho AWS que miden la irradiancia solar (ver Tabla 3-1), La Figura 3-2 muestra la ubicación de las estaciones meteorológicas automáticas y convencionales en el departamento. En la región del Pacífico hay tres AWS ubicadas en Nariño y una en el departamento vecino del Cauca. El rango de altitud de AWS está entre 16 y 512 m.s.n.m. En la región Andina, hay cinco AWS con altitudes entre 1.005 y 3.120 m.s.n.m. Por último, en la región amazónica, hay una AWS a 3.577 m.s.n.m. La diferencia de altitud máxima entre todas las AWS es de 3.561 m.s.n.m. Esta diferencia permite evaluar el comportamiento de la irradiancia solar en una amplia gama de altitudes y en tres regiones ambientales diversas.

**Figura 3-2.** Localización de las AWS



**Tabla 3-1.** Estaciones meteorológicas automáticas

Nombre	Latitud	Longitud	Altitud	Periodo	Región
Biotopo	1,41	-78,28	512	2005-2017	Pacífico
Altaquer	1,56	-79,09	1.1010	2013-2014	Pacífico
Granja el Mira	1,55	-78,69	16	2016-2017	Pacífico
Cerro-Páramo	0,84	-77,39	3.577	2005-2017	Amazonia
La Josefina	0,93	-77,48	2.449	2005-2017	Andina
Viento Libre	1,62	-77,34	1.005	2005-2017	Andina
Universidad de Nariño	1,23	-77,28	2.626	2005-2017	Andina
Botana	1,16	-77,27	2.820	2005-2017	Andina
El Paraíso	1,07	-77,63	3.120	2005-2017	Andina

### 3.2.2 Control de calidad de los datos de irradiancia solar

El control de calidad de datos es un procedimiento que busca mejorar la fiabilidad de las series temporales. En este caso, incluye un análisis de la estructura, la comparación con los límites fijos y flexibles, y la consistencia del tiempo. Estos pasos ofrecieron resultados fiables en estudios de control de calidad de los datos meteorológicos en España (Estévez et al., 2011). El procedimiento de control de calidad seguido en esta investigación tomó como pauta el reglamento UNE500540 que propone sucesivos procedimientos analíticos de control de calidad para diferentes variables meteorológicas (AENOR, 2004).

El primer paso consiste en comprobar la estructura de la base de datos. En nuestro caso, los datos tienen la siguiente estructura: código de la AWS, código de la variable meteorológica, fecha, hora y valor. Este paso solo mantiene los valores con la estructura descrita (Rivero et al., 2017). Para el de rango fijo, UNE500540 sugiere utilizar la condición más restrictiva entre el límite del dispositivo de medición y el límite del fenómeno físico (AENOR, 2004). Las AWS de Nariño tienen piranómetros Kipp & Zonen CMP11 con un límite de operación superior de  $4.000W/m^2$  (Kipp & Zonen, 2000). El límite físico es la irradiancia solar extraterrestre máxima de cada ubicación calculada con la Ecuación 3-1. Por lo tanto, la condición más restricción es la última; en consecuencia,  $I_0 \geq I_{mt}$ , donde  $I_0$  es la máxima irradiancia solar extraterrestre, y  $I_{mt}$  es la irradiancia global solar medida en el tiempo  $t$ .

$$I_0 = I_{sc} \left[ 1 + 0,033 \cos \left( 360 \frac{D-3}{365} \right) \right] * \sin \beta \quad (3-1)$$

$$\sin \beta = \cos \phi \cos \delta \cos \omega_s + \sin \phi \sin \delta \quad (3-2)$$

Para la prueba de rango flexible, UNE500540 sugiere comparar la serie temporal con los valores máximo y mínimo de una serie temporal validada. En este caso, no hay ninguna serie temporal validada previamente; por lo tanto, los autores utilizaron, como alternativa, la restricción propuesta por Estévez et al. con una modificación. Es decir, en lugar de utilizar  $I_0$ , utilizan el siguiente rango definido por la irradiancia solar global en cielo despejado  $0,03I_{cst} \leq I_{mt} \leq I_{cst}$  (Estévez et al., 2011). Teniendo en cuenta que la irradiancia solar global en cielo despejado es igual  $\tau$  veces  $I_0$ . El primer paso es estimar  $\tau$ , que es la transmitancia atmosférica, usando el modelo de Kreith & Kreider, y posteriormente calcular la irradiancia global en cielo despejado con la Ecuación 3-3.

$$\tau = 0,56 (e^{-0,65m} + e^{-0,095m}) \quad (3-3)$$

Donde:

$$m = \frac{1}{\sin \beta} \quad (3-4)$$

$$I_{cs} = I_0 \tau \quad (3-5)$$

El cuarto paso consiste en analizar el cambio hora a hora de la irradiancia solar global. Este análisis, conocido como consistencia del tiempo, sigue la restricción  $|I_{cst} - I_{cst-1}| > |I_{mt} - I_{mt-1}|$ . Es una prueba útil para detectar problemas de almacenamiento de datos y conexión en el registrador de datos. Una alta frecuencia de muestreo, por ejemplo, cada diez minutos, aumentaría la eficacia de la prueba.

Además de las pruebas mencionadas anteriormente, es necesario considerar la corrección por cero, que resulta de desequilibrios térmicos en el piranómetro. Este fenómeno se produce porque el sensor no absorbe irradiancia medible en el rangopectral del piranómetro, lo que resulta en valores erróneos (Kipp & Zonen, 2000). No compensar este desequilibrio implica una subestimación del recurso entre el 0,7 % y el 4,3 %. A pesar de su importancia, varios factores ambientales influyen en el proceso de medición, lo que dificulta establecer una corrección válida para todos los instrumentos de medición en todos los lugares y entornos (Serrano et al., 2015). Aunque existen varios enfoques para corregir la compensación por cero adaptada a condiciones e instrumentos ambientales específicos, esta investigación no siguió ningún proceso.

### 3.2.3 Control de calidad de datos de temperatura

La validación de los datos de temperatura siguió las recomendaciones de Estévez et al.; Rivero et al. El control de calidad de datos de temperatura tiene cinco pasos de validación: estructura, rango, paso, coherencia y persistencia. El primer paso consiste en verificar la estructura de la base de datos, de la misma manera que la validación de irradiancia solar. La prueba de rango tiene dos maneras de evaluarla. El rango instrumental determina la primera manera. En este caso, el sensor es un ROTRONIC HYGROCLIP 2 RTD PT100 con un rango  $-40^{\circ}\text{C}$  y  $60^{\circ}\text{C}$ . Los criterios de los investigadores definen la segunda manera de validar los datos. En consecuencia, Rivero et al. recomiendan un intervalo de temperatura entre  $-50^{\circ}\text{C}$  y  $70^{\circ}\text{C}$ , mientras Estévez et al. sugiere un rango de  $-30^{\circ}\text{C}$  a  $50^{\circ}\text{C}$ . En esta investigación se tomó la recomendación de Estévez et al. porque está dentro del rango de medición del instrumento de temperatura. Del mismo modo, el análisis para los siguientes niveles (prueba de paso, prueba de consistencia interna y prueba de persistencia) siguió el proceso de control de calidad de Estévez et al. porque ofrecía resultados prometedores en el análisis de datos en España (Estévez et al., 2011; Herrera-Grimaldi et al., 2019). La prueba de paso requiere cumplir los siguientes requisitos:  $|T_h - T_{h-1}| < 4$ ;  $|T_h - T_{h-2}| < 7$ ;  $|T_h - T_{h-3}| < 9$ ;  $|T_h - T_{h-6}| < 15$ ; y  $|T_h - T_{h-12}| < 25$ . Aunque esta evaluación no tiene en cuenta otros aspectos climatológicos como la humedad relativa, la velocidad del viento, la nubosidad y la precipitación, los autores aplicaron esas restricciones. Otra restricción es de rango diario  $T_{max} - T_{min} < 30^{\circ}\text{C}$  (Herrera-Grimaldi et al., 2019).

La consistencia interna evalúa el cumplimiento de las siguientes restricciones:  $T_{max} > T_{mean} > T_{min}$ ;  $T_{max}(d) > T_{min}(d-1)$ ;  $T_{min}(d) \leq T_{max}(d-1)$ . Por último, la prueba de persistencia verifica la variabilidad de la medición; por lo tanto, los datos deben cumplir las condiciones  $T_{max}(d) \neq T_{max}(d-1) \neq T_{max}(d-2)$ ,  $T_{min}(d) \neq T_{min}(d-1) \neq T_{min}(d-2)$  (Estévez et al., 2011).

### 3.2.4 Modelos empíricos basados en temperatura

Aunque la temperatura del aire es una variable comúnmente medida en las estaciones meteorológicas, ésta no era frecuente usada para estimar la insolación solar. La temperatura comenzó a ser relevante para estimar la insolación solar cuando se modeló ésta con datos de temperatura en estudios agrícolas. En consecuencia, los investigadores de otros campos del conocimiento, como la energía solar, prestaron atención a los valores de temperatura máxima, mínima y media en sus estudios (Paulescu, 2008). Los modelos más tradicionales que utilizan las temperaturas máximas y mínimas son los modelos de Hargreaves y Samani, y Bristow y Campbell (Besharat et al., 2013; Bristow & Campbell, 1984). Ambos modelos son la base de nuevos enfoques adaptados a las condiciones de ubicación.

Antes de presentar los modelos empíricos, es necesario introducir la expresión matemática para calcular la insolación solar extraterrestre diaria sobre una superficie horizontal, porque es parte de la mayoría de los modelos empíricos.

$$H_0 \left( \frac{Wh}{m^2 da} \right) = \frac{24}{\pi} I_{sc} \left( 1 + 0,033 \cos \left( \frac{360D}{365} \right) \right) * \left( \cos \phi \cos \delta \cos \omega_s + \frac{\pi}{180} \omega_s \sin \phi \sin \delta \right) \quad (3-6)$$

Las Ecuaciones 3-7 y 3-8 permiten calcular la declinación solar  $\delta$  y el ángulo horario  $\omega_s$  respectivamente:

$$\delta = 23,45 \sin \left[ \frac{360(D + 284)}{365} \right] \quad (3-7)$$

$$\omega_s = \cos^{-1} [-\tan(\delta) \tan(\phi)] \quad (3-8)$$

#### Modelo de Hargreaves y Samani

Hargreaves y Samani (HS) propusieron una relación lineal entre la raíz cuadrada de la diferencia de temperatura y la fracción entre la insolación solar extraterrestre y terrestre para diferentes períodos de tiempo en 1982. Este modelo no tuvo en cuenta los efectos de la nubosidad, humedad relativa, latitud, elevación y topografía, entre otros, en la ubicación específica para la que se utiliza el modelo (Samani, 2000; Allen, 1997).

$$\frac{H}{H_0} = a (T_{max} - T_{min})^{0,5} \quad (3-9)$$

Allen declaró que el modelo de HS tiene un mejor comportamiento en un período de tiempo mensual que en el diario porque las variables siguen una tendencia media, lo que resulta en una relación consistente entre  $T_{max} - T_{min}$  y  $H/H_0$  (Allen, 1997).

El coeficiente empírico representa la tasa de cambio de la diferencia de temperaturas máximas y mínimas con la relación entre la insolación solar extraterrestre y terrestre. Inicialmente, el modelo HS propuso un coeficiente empírico calibrado con una serie temporal de ocho años de Central Valley en el Condado de Davis (California), véase la Tabla 3-2 (Besharat et al., 2013; Hargreaves & Samani, 1982). HS presentó una conclusión relevante sobre los elementos meteorológicos que afectan a este coeficiente. Encontraron que la humedad relativa por encima del 54 % afecta al coeficiente empírico, ya que la diferencia de temperaturas máximas y mínimas es menor en estos ambientes. Como se muestra en la Tabla 3-2, los climas húmedos tienen valores menores que 0,1 mientras que las otras regiones tienen valores más altos que 0,1 (Hargreaves & Samani, 1982). Un proceso minucioso utiliza coeficientes empíricos según la región en estudio. En consecuencia, Rivero et al. señalaron que el uso de un coeficiente empírico fijo para una gran área con diferentes regiones puede dar lugar a errores significativos porque la topografía, la advección y la vegetación influyen en la temperatura (Rivero et al., 2017).

**Tabla 3-2.** HS coeficientes empíricos

a	Tipo de región
0,16	Árido y semi-árido
0,17	Regiones interiores
0,19	Regiones costeras
0,10-0,09	Climas húmedos

Fuente: (Besharat et al., 2013; Hargreaves & Samani, 1982)

### Modelo de Bristow y Campbell

Bristow y Campbell (BC) presentaron un modelo basado en dos supuestos. La primera es la relación lineal entre la insolación solar absorbida y entrante, y la segunda suposición consiste en despreciar el flujo de calor que viene del suelo durante un período diario porque está cerca de cero. Sin embargo, es conveniente mencionar que el calor sensible producido por las temperaturas diurnas es más alto que durante la noche (Bristow & Campbell, 1984). Lo último es comprensible porque la irradiancia solar calienta más las masas de aire debido a la radiación de onda corta. Mientras que durante la noche, hay menos emisión de onda larga de la Tierra a la atmósfera reduciendo la temperatura (Meza F. & Varas E., 2000). Además, en condiciones ideales, la temperatura es mínima justo antes del amanecer, lo que resulta en una diferencia significativa entre la temperatura máxima y la temperatura mínima del aire diaria. Este fenómeno permite modelar la insolación solar en función de la diferencia de temperatura (Bristow & Campbell, 1984). Bajo estas suposiciones, los autores presentaron su modelo.

$$\frac{H}{H^0} = a [1 - e^{(-b\Delta T^c)}] \quad (3-10)$$

$a$ ,  $b$ , y  $c$  son coeficientes empíricos,  $a$ , representa la relación máxima entre la insolación solar extraterrestre y terrestre,  $b$  y  $c$  que tan rápido el máximo valor de  $H/H_0$  es alcanzado con cambio de  $\Delta T$ . Los coeficientes empíricos representan las características regionales de ambientes áridos a húmedos (Bristow & Campbell, 1984).

$$\Delta T(D) = T_{max}(D) - \frac{T_{min}(D) + T_{min}(D+1)}{2} \quad (3-11)$$

El cambio de temperatura depende de las temperaturas máximas  $T_{max}$  y mínimas  $T_{min}$  diarias durante un día dado  $D$ . Bristow & Campbell concluyeron que el promedio de temperaturas mínimas de dos días consecutivos reduce los efectos de las masas de aire caliente o frío, evitando valores de insolación solar sobreestimados y subestimados. Además, C. M. Dos Santos et al. afirmaron que la advección no es un fenómeno común en las zonas de los trópicos; por lo tanto, la estimación del cambio de temperatura  $\Delta T$  en estas regiones se hace con la Ecuación 3-12. Esos autores también destacaron que esta ecuación es mejor para sitios con grandes altitudes, como en el caso de estudio (C. M. Dos Santos et al., 2014).

$$\Delta T(D) = T_{max}(D) - T_{min}(D) \quad (3-12)$$

Para reducir el efecto de la lluvia es necesario ajustar el rango de la temperatura diaria, ajustando  $\Delta T(J)$  igual a 0,75 veces el estimado  $\Delta T(D)$ . Además, si  $\Delta T(D-1)$  es inferior a ( $2^{\circ}C$ ) respecto a  $\Delta T(D-2)$ , el primer factor es multiplicado por 0,75 (Bristow & Campbell, 1984). Sin embargo, cuando los períodos lluviosos son largos, la relación entre la insolación solar y  $\Delta T$  alcanza un equilibrio. Por lo tanto, no requiere ajustes (Goodin et al., 1999).

### **Modelos implementados en un ambiente tropical**

El conocimiento del comportamiento de modelos de estimación de la radiación solar en zonas tropicales es un tema obligatorio para esta investigación. Los casos analizados pertenecen a África (Nigeria Abuja, Benin City, Katsina, Lagos, Nsukka y Yola), Brasil (Agua Branca, Pao de Azucar, Santana do Ipanema, Palmeira dos Índios, Arapicara, Maceió, Coruripe, São José da Laje) y México. Debido a que Nariño se encuentra en una zona tropical, los resultados de estos tres casos constituyen insumos relevantes para nuestra investigación.

Okundamiya & Nzeako propusieron el modelo lineal representado con la Ecuación 3-13 (Okundamiya & Nzeako, 2011).  $T_R$  es el promedio mensual de la relación entre las temperaturas mínimas y máximas diarias,  $a$ ,  $b$  y  $c$  son coeficientes empíricos. La validación estadística del modelo mostró un coeficiente de determinación entre 0.809 y 0,952.

$$\frac{H}{H_0} = a + bT_R + cT_{max} \quad (3-13)$$

Nwokolo y Ogbulezie analizaron modelos empíricos implementados en África Occidental para calcular la radiación solar global. Encontraron que los modelos de computación blanda tienen mejor precisión

que los modelos empíricos, ya que pueden ser más fáciles de adaptar a varias condiciones climáticas porque permiten más entradas, como modelos o variables, para reforzar su fiabilidad (Nwokolo & Ogbulezie, 2017).

C. M. Dos Santos et al. estudiaron el desempeño de los diez modelos basados en la temperatura en el noreste de Brasil. Encontraron que los modelos no mostraban cambios significativos después de ajustarlos con respecto a los efectos de los períodos lluviosos. Dos Santos y sus colegas también concluyeron que el modelo HS tenía un mejor rendimiento en tierras interiores, y el modelo BC mostró el mejor rendimiento en zonas húmedas y costeras (C. M. Dos Santos et al., 2014).

Rivero et al. compararon los valores de los coeficientes empíricos del modelo original de HS con nuevos valores derivados de ese modelo calibrado para México con datos locales. Un aspecto notable de esa investigación consiste en la clasificación de las zonas climáticas de México utilizando el sistema Koppen-Geiger. Además, los autores realizaron otra clasificación basada en el índice de claridad, como se muestra en la Tabla 3-3 para describir los datos de las AWS. Es útil recordar que el índice de claridad es la relación entre la insolación solar terrestre y extraterrestre. Una conclusión relevante mostró que independientemente de los picos de irradiancia solar durante el día, es posible obtener valores de índice de claridad similares. Otra conclusión señaló que los coeficientes empíricos fijos conducen a un error significativo, especialmente en zonas con una diferencia de temperatura inferior a 15 °C. Los autores propusieron la Ecuación 3-14 para superar los problemas identificados derivados de coeficientes fijos (Rivero et al., 2017).

**Tabla 3-3.** Clasificación por índice de claridad

$K_T$ Rango	Tipo de día
$0,00 < K_T \leq 0,20$	Nublado
$0,20 < K_T \leq 0,60$	Parcialmente nublado
$0,60 < K_T \leq 0,75$	Soleado
$0,75 < K_T \leq 1,00$	Muy soleado

Fuente: (Rivero et al., 2017)

$$a_{HS} = a_1 + a_2 (\Delta T) + a_3 (\Delta T)^2 \quad (3-14)$$

### 3.2.5 Modelo empírico propuesto

El modelo propuesto se originó a partir de la observación de la gráfica de dispersión entre el índice de claridad y la diferencia de temperatura diaria en cada AWS. Los autores concluyeron que el modelo logístico ofrecería resultados útiles. Este modelo ha estudiado con éxito el crecimiento humano, los procesos animales y biológicos, los patrones de uso de energía y las aplicaciones atmosféricas (Agami Reddy, 2011; Moon & Kim, 2020). La regresión logística suele describir la relación entre las variables binarias y un predictor (Harrenll, 2015). Aunque la relación entre la insolación solar extraterrestre y terrestre no ofrece resultados binarios, varía en un rango entre 1 y 0 (Kleinbaum & Klein, 2010). El uso de esta técnica “surge en la estimación de relaciones en las que la variable dependiente es continua, pero está limitada en rango” (Manning, 1996). La declaración de Manning describe la relación del modelo propuesto entre las variables dependientes e independientes, porque la relación de las insolaciones solares es continua

únicamente para el rango definido por las temperaturas mínimas y máximas. La regresión logística no supone que ni las variables ni los predictores tengan una distribución normal; esta afirmación fue otro criterio para elegir este enfoque.

El modelo de regresión logística forma parte del Modelo Lineal Generalizado (GLM), que describe la relación entre la media de las variables dependientes e independientes con una relación más compleja que  $y = a + bx$ . La regresión logística, que es un método estadístico, los coeficientes tienen un significado similar al de la regresión lineal; a saber  $a$  es la probabilidad de éxito en  $x = 0$ , mientras  $b$  es el cambio en la probabilidad de éxito correspondiente a un aumento de una unidad en  $x$  (Casella & Berger, 2002). Esta regresión también supone una relación lineal entre las probabilidades de las variables dependientes e independientes (Moon & Kim, 2020).

$$y = \frac{1}{1 + e^{-z}} \quad (3-15)$$

La función logística provino del modelo logístico representado por la Ecuación 3-15.  $y$  es la variable predicha,  $z$  es una suma lineal  $a + \sum b_i x_i$ , donde  $x_i$  es la variable independiente, y  $a$  y  $b$  son constantes (Kleinbaum & Klein, 2010).

$$y = \frac{1}{1 + e^{-(a+bx)}} \quad (3-16)$$

El cambio de temperatura es la variable predictora. Por lo tanto, el modelo propuesto es el siguiente.

$$\frac{H}{H_0} = \frac{1}{1 + e^{-(a+b\Delta T)}} \quad (3-17)$$

A pesar de tener una variable dependiente no binaria, los resultados del modelo propuesto mostraron un rendimiento estable.

### 3.2.6 Validación estadística

La validación estadística es un paso obligatorio que permite comparar el modelo predictor con las medidas reales para determinar su idoneidad. Existen cinco técnicas principales de validación: evaluación subjetiva, indicadores de dispersión, indicadores generales de rendimiento, indicadores de similitud de distribución e indicadores visuales (Gueymard, 2014; Mayer & Butler, 1993). El primero consiste en evaluar el modelo por expertos. Sin embargo, estas técnicas están abiertas a la interpretación errónea ya que es una prueba subjetiva con sesgo personal (Mayer & Butler, 1993). El segundo es apropiado cuando los datos tienen el mismo marco de tiempo de trabajo, ubicación y tratamiento, entre otras características. Esta técnica de validación mide la diferencia entre el valor modelado y el valor real. La Tabla

**3-4** muestra algunas medidas de dispersión, expresada en unidades absolutas o porcentajes (Mayer & Butler, 1993).

**Tabla 3-4.** Medidas de errores estadísticos

Medida	Definición	Formula <sup>2</sup>
Error medio porcentual (MPE)	Los valores cercanos a cero indican un mejor modelo y que la relación de la desviación estándar del valor medido y calculado está cerca a uno.	$\frac{1}{n} \sum_{i=1}^n (p_i - o_i) / o_i$
Error medio absoluto (MAE)	Es la distancia vertical media entre cada punto predicho y observado.	$\frac{1}{n} \sum_{i=1}^n  p_i - o_i $
Error medio cuadrático (RMSE)	Proporciona una medida del tamaño del error y es sensible a los valores atípicos porque esta medida da mucho peso a errores grandes.	$\left[ \frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2 \right]^{1/2}$
Error medio de desviación (MBE)	Esta medida proporciona información sobre el rendimiento a largo plazo del modelo cuando el modelo tiene un error sistemático que presenta predictores sobreestimados o subestimados. Los valores bajos de MBE son deseables, aunque debe tenerse en cuenta que un conjunto de datos sobreestimado cancelará otro conjunto de datos subestimado.	$\frac{1}{n} \sum_{i=1}^n p_i - o_i$
Desviación estándar (SD)	Muestra la diferencia entre la desviación estándar de los conjuntos de datos predichos y observados.	$\left( \frac{100}{O_m} \right) \frac{1}{n} \left[ \sum_{i=1}^n n (p_i - o_i)^2 - [\sum_{i=1}^n (p_i - o_i)]^2 \right]^{1/2}$
Incertidumbre al 95 % ( $U_{95}$ )	Mide la confianza de la certeza; un valor más bajo es mejor.	$1,96 (SD^2 + RMSE^2)^{1/2}$

<sup>2</sup>  $p_i$  valor predicho,  $o_i$  valor observado,  $O_m$  promedio del valor observado,  $n$  cantidad de datos

Fuente: (Almorox et al., 2011; C. M. Dos Santos et al., 2014; Gueymard, 2014; J. Li & Heap, 2011; Mayer & Butler, 1993)

La tercera categoría, aunque similar a la segunda, evalúa diferentes campos de estudio a la energía solar. Este indicador proporciona información sobre la eficiencia del modelo utilizando. La cuarta categoría compara la distribución de frecuencia acumulativa del modelo con los valores de referencia; esta investigación no implementó esta categoría (Gueymard, 2014). La última categoría muestra las operaciones de datos simuladas y reales mediante gráficas. Es un método informativo convencional; sin embargo, el formato de presentación de datos podría dar lugar a interpretaciones erróneas. Como resultado, una alternativa deseable son las gráficas de dispersión observadas frente a las predichas con el ajuste lineal, lo que indica el ajuste perfecto. Presenta “la bondad del ajuste como una desviación vertical de la línea perfecta,” lo que indica algunos sesgos (Mayer & Butler, 1993). Un método de visualización adecuado

es el diagrama de Taylor, ya que combina tres indicadores de dispersión: correlación de Pearson, RMSE y desviación estándar (Gueymard, 2014).

## 3.3 Resultados y discusiones

Los resultados tienen cuatro subsecciones. El primero presenta los resultados de control de calidad de irradiancia solar global. El segundo muestra los resultados del procedimiento de validación de datos de temperatura. El tercero contiene los coeficientes empíricos HS, BC, ON los del modelo propuesto. La última subsección comprende los resultados de la imputación y la insolación solar diaria para las AWS estudiadas.

### 3.3.1 Control de calidad de datos irradiancia solar global

Un paso previo al procedimiento de control de calidad es el ajuste de datos, aplicando la constante de calibración. En este caso, no hay ninguna constante de calibración para Biotoxo, afectando la calidad de las series temporales; esta AWS tenía la cantidad de datos almacenados más pequeña entre todas las AWS. Los datos totales fueron de 47.612 para el período estudiado, correspondientes al 34,50 % aproximadamente de los datos que deberían existir. En promedio la información registrada por AWS durante el período de medición 2005-2017 fue del 57,59 %.

La Tabla 3-5 muestra los resultados del procedimiento de validación de la irradiancia solar global. El primer paso para evaluar, la estructura de la base de datos, presenta 5.843 datos almacenados en promedio con la estructura incorrecta. La AWS con más datos con la estructura incorrecta es Viento Libre con 9.715 datos erróneos, correspondiente al 12,55 % de los datos totales. La AWS con menos grabaciones con la estructura de datos incorrecta es Biotoxo; este paso confirmó que 1.055 datos no tenían la estructura correcta, correspondiente al 2,22 % del total de datos. La validación de rango fijo descartó en promedio 36,71 % datos. Biotoxo perdió la mayor cantidad de datos correspondiente al 47,62 % del total, y Universidad de Nariño perdió el menor número de datos de alrededor del 22,09 % del total. Los resultados de las pruebas de rango flexible presentan que hay pérdidas de información del 39,11 % en promedio. Biotoxo es la AWS con las pérdidas más bajas, alrededor del 15,14 %, y Guapi tiene las pérdidas más altas, alrededor del 54,22 %. Aunque la última prueba no es obligatoria, es útil señalar que el 44,06 % de los datos no superaron este nivel. Teniendo en cuenta únicamente los pasos obligatorios, alrededor del 35,27 % de los datos superan estos pasos de validación. La AWS de la Universidad de Nariño tiene la mayor cantidad de datos que aprobó el proceso de validación, y Guapi tiene la menor cantidad de datos que aprobó el proceso de validación.

La Tabla 3-6 presenta el número de días con datos entre las 6:00 y las 18:00. Cabe señalar que sólo el 1,26 % de días disponía de información completa, con un máximo del 4,25 % en Cerro Páramo y un

**Tabla 3-5.** Resultados de validación de datos de irradiancia global

Nombre	Código	Datos	Paso 1	Paso 2	Paso 3	Paso 4
Biotoxo	51025060	47.612	46.557	24.385	20.699	12.883
Viento Libre	52035040	77.424	67.709	40.777	26.835	12.311
Universidad de Nariño	52045080	98.452	93.338	72.715	37.481	21.033
Cerro Páramo	52055150	90.440	81.940	57.407	36.661	25.709
La Josefina	52055170	55.909	54.041	29.966	14.183	7.427
Botana	52055210	98.928	90.777	51.416	38.327	20.847
El Paraiso	52055220	88.408	82.135	54.371	29.033	14.394
Guapi	53045040	78.773	72.708	49.039	22.452	12.747

mínimo del 0,28 % en Biotoxo. En consecuencia, la implementación del modelo solo consideró los días con al menos seis datos durante el período diurno para evitar subestimar del recurso, aproximadamente el 95,81 % de los datos que superaron los niveles de validación obligatorios fueron usados para la agregación diaria.

Los resultados confirman la importancia de mejorar los procedimientos de mantenimiento y calibración de los instrumentos de medición para aumentar la cantidad de información útil y, a su vez, aumentar la fiabilidad de los resultados.

**Tabla 3-6.** Cantidad de días con datos entre las 6 y 18 horas

Nombre	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Biotoxo	13	16	15	19	30	49	104	190	350	528	622	207	6	2.149
Viento Libre	43	50	77	86	121	175	313	444	702	654	390	119	11	3.185
Universidad de Nariño	27	42	56	73	97	173	340	403	730	1.014	778	308	63	4.104
Cerro Páramo	34	39	42	45	59	83	145	283	483	952	1.080	362	160	3.767
La Josefina	109	71	22	24	25	57	78	195	293	386	274	134	6	1.674
Botana	16	20	26	43	82	142	246	469	798	1.066	839	336	14	4.097
El Paraiso	55	83	110	137	158	205	285	364	614	787	514	149	13	3.474
Guapi	185	193	180	153	111	101	139	239	335	447	549	182	75	2.889

Los autores consideraron la clasificación de Rivero et al. y modificaron el elemento sobre días parcialmente nublados, como se muestra en la Tabla 3-7. En el conjunto de datos analizado hay principalmente días parcialmente nublados (64,70 %), como muestra la Tabla 3-7. En otras palabras, aproximadamente del 20 % al 40 % de la irradiancia solar extraterrestre llega al nivel del suelo. Este es un resultado interesante debido a que las nubes son un factor crucial que influye en la irradiancia solar global y es una variable que no es fácil de modelar. Por tanto, la estimación del recurso en este ambiente tropical y montañoso es más compleja que en otros ambientes.

**Tabla 3-7.** Clasificación por índice de claridad

$K_t$	$0,00 < K_t \leq 0,20$	$0,20 < K_t \leq 0,40$	$0,40 < K_t \leq 0,60$	$0,60 < K_t \leq 0,75$	$0,75 < K_t \leq 1,00$	Cantidad de días*
Nombre	Nublado	Nubosidad parcialmente alta	Nubosidad parcialmente baja	Soleado	Muy soleado	
Biotopo	760	<b>1.188</b>	67	0	0	2.015
Viento Libre	105	<b>1.458</b>	1.179	0	0	2.745
Universidad de Nariño	243	<b>2.611</b>	846	1	0	3.701
Cerro Páramo	<b>1.380</b>	1.232	137	1	0	2.793
La Josefina	88	<b>991</b>	271	1	0	1.351
Botana	451	<b>2.704</b>	711	2	0	3.868
El Paraíso	167	<b>1.859</b>	543	1	0	2.570
Guapi	130	<b>746</b>	125	0	0	1.001

\*Días con información completa de  $\Delta T$ ,  $T_R$ ,  $T_{max}$ ,  $T_{min}$ ,  $K_t$

### 3.3.2 Control de calidad de datos de temperatura

La Tabla 3-8 muestra los resultados del procedimiento de validación de datos de temperatura. En el primer paso, hay una pérdida en promedio del 8,64 % de los datos. Cerro Páramo es la AWS más crítica con una pérdida del 16,95 %. En el siguiente paso, la prueba de rango fijo, Guapi tuvo una pérdida de datos del 11,62 %. En la prueba de paso, en promedio, el 35,07 % de los datos no pasan el requisito de validación. Tomando como base los valores iniciales, el 57,08 % de los datos superan el procedimiento de control de calidad.

**Tabla 3-8.** Resultados de validación de temperatura

Nombre	Código	Datos	Paso 1	Paso 2	Paso 3
Biotopo	51025060	52.848	47.268	46.436	24.385
Viento Libre	52035040	77.424	67.969	67.962	40.777
Universidad de Nariño	52045080	100.740	94.880	92.886	72.715
Cerro Páramo	52055150	91.850	76.280	69.410	57.407
La Josefina	52055170	55.728	52.867	52.707	29.966
Botana	52055210	98.952	91.077	91.047	51.416
El Paraíso	52055220	91.699	85.713	84.792	54.371
Guapi	53045040	84.371	81.014	71.598	49.039

La Tabla 3-9 muestra el número de días según las mediciones de temperatura durante el día. Eso significa que aproximadamente el 57,42 % de los días tiene el 88,46 % de la información. Finalmente, de la cantidad de datos que superan la prueba de validación horaria, el 68,03 % de la información es apta para ser utilizada en los modelos empíricos.

### 3.3.3 Calibración y validación de modelos empíricos

La Figura 3-3 muestra la relación entre el índice de claridad diaria y el delta diario de temperatura para las AWS estudiadas. La Tabla 3-10 muestra los coeficientes empíricos de HS, BC, ON y el modelo

**Tabla 3-9.** Cantidad de días por mediciones diarias de temperatura

Nombre	1	2	3	4	5	6	7	8	9	10	11	12	Total
Biotoxo	33	9	17	13	26	46	56	85	130	257	605	846	2.123
Viento Libre	79	68	91	91	94	120	124	192	264	500	873	696	3.192
Universidad de Nariño	74	26	24	31	47	71	113	198	309	599	1.170	1.379	4.041
Cerro Páramo	101	49	58	47	67	86	133	190	348	590	754	701	3.124
La Josefina	61	21	36	46	50	78	109	130	197	405	573	558	2.264
Botana	38	11	27	46	62	102	147	253	386	679	1.114	1.221	4.086
El Paraíso	74	42	65	67	92	102	155	236	373	541	915	909	3.571
Guapi	20	8	10	21	18	21	45	79	168	405	919	1.349	3.063

propuesto. Las constantes empíricas  $a$  y  $b$  del modelo BC muestran una tendencia creciente con la altitud de las AWS, mientras que la constante empírica  $c$  muestra el comportamiento opuesto. Los coeficientes empíricos del modelo HS no presentan variación significativa, la diferencia entre el valor máximo y mínimo es de 0,037. El coeficiente empírico  $a$  del modelo ON en Bitopo es el único valor negativo. Los coeficientes  $c$  del modelo ON presentan una desviación estándar baja de 0,00030.

La Tabla 3-11 presenta los resultados de siete mediciones de validación estadística para cada AWS. Considerando los resultados de RMSE, SD, MAE,  $U_{95}$  y MAPE, el modelo propuesto tiene un mejor desempeño que los demás. El modelo de BC tuvo mejores resultados en MBE y MPE. El modelo propuesto muestra mejores resultados en las AWS ubicadas en altitudes superiores a los 2.500 msnm. Mientras que el modelo de HS muestra un mejor desempeño en las AWS ubicadas en altitudes por debajo de los 2.500 msnm. Sin embargo, si el objetivo es utilizar un modelo único para estimar la irradiancia solar global a partir de los datos de temperatura en Nariño, el modelo propuesto será el mejor.

**Tabla 3-10.** Coeficientes empíricos

	Bristow & Campbell			Hargreaves & Samani
AWS	$a$	$b$	$c$	$a$
Biotoxo	0,5075	0,0735	1,1908	0,0970
Viento Libre	0,5942	0,1499	0,8655	0,1248
Cerro Páramo	0,5922	0,2595	0,6153	0,1340
Universidad de Nariño	0,4893	0,3282	0,6568	0,1263
Botana	0,6288	0,1964	0,6415	0,1169
Josefina	0,6039	0,2563	0,5350	0,1138
Paraíso	0,5850	0,3183	0,4818	0,1199
Guapi	0,4471	0,1156	1,2478	0,1208
	Okundamiya & Nzeako			Modelo propuesto
AWS	$a$	$b$	$c$	$a$
Biotoxo	-0,1838	-0,3871	0,0264	-2,3058
				0,1786

Viento Libre	0,0578	-0,2666	0,0168	-1,3499	0,0912
Cerro Páramo	0,1084	-0,1572	0,0257	-1,7914	0,1706
Universidad de Nariño	0,2679	-0,2416	0,0112	-1,2211	0,0747
Botana	0,0621	-0,1547	0,0191	-1,4489	0,0898
Josefina	0,1770	-0,1589	0,0118	-1,2299	0,0608
Paraiso	0,2617	-0,1775	0,0100	-1,1667	0,0607
Guapi	0,0717	-0,5202	0,0228	-1,8043	0,1495

En los resultados de RMSE, el valor más bajo es  $878,75 \text{ Wh}/(\text{m}^2\text{da})$  en Guapi, y el valor más alto es  $1.209,76 \text{ Wh}/(\text{m}^2\text{da})$  en Cerro Páramo, en promedio el RMSE es  $1.046,69 \text{ Wh}/(\text{m}^2\text{da})$ . El modelo ON presenta el RMSE más grande con  $1.058,77 \text{ Wh}/(\text{m}^2\text{da})$ , seguido por el modelo BC con  $1.052,96 \text{ Wh}/(\text{m}^2\text{da})$ , el modelo HS con  $1.046,66 \text{ Wh}/(\text{m}^2\text{da})$ , y el modelo propuesto con  $1.028,37 \text{ Wh}/(\text{m}^2\text{da})$ .

La desviación estándar del residual muestra que Cerro Páramo presentó valores más dispersos que las otras AWS. El promedio de la desviación estándar presenta al modelo propuesto como la mejor opción, con un valor de 35,64. Los resultados de MBE muestran que el modelo de BC subestima el recurso en  $77,79 \text{ Wh}/(\text{m}^2\text{da})$  en Cerro Páramo. Mientras que todos los modelos sobre estiman el recurso, por ejemplo en Viento Libre por  $167,77 \text{ Wh}/(\text{m}^2\text{da})$ ,  $163,13 \text{ Wh}/(\text{m}^2\text{da})$ ,  $162,63 \text{ Wh}/(\text{m}^2\text{da})$  y  $160,23 \text{ Wh}/(\text{m}^2\text{da})$ , con el modelo ON, HS, BC y propuesto , respectivamente.

Los resultados del MAE describen al modelo propuesto como el de mejor desempeño con un error promedio de aproximadamente  $821,41 \text{ Wh}/(\text{m}^2\text{da})$ , mientras que el modelo ON alcanza un error promedio de aproximadamente  $845,27 \text{ Wh}/(\text{m}^2\text{da})$ .  $U_{95}$  en promedio tiene los siguientes resultados  $2.076,49 \text{ Wh}/(\text{m}^2\text{da})$ ,  $2.065,11 \text{ Wh}/(\text{m}^2\text{da})$ ,  $2.034,13 \text{ Wh}/(\text{m}^2\text{da})$ ,  $2.016,86 \text{ Wh}/(\text{m}^2\text{da})$  para ON, BC, HS y el modelo propuesto respectivamente. Esto confirma que el modelo propuesto es mejor que los demás.

MPE presentó al modelo de BC como el mejor con un 14,45 % de error en promedio y al modelo de HS como el peor con un 15,31 % en promedio. MAPE mostró el modelo propuesto como la mejor opción con un 35,12 % de error en promedio. Los resultados de MAPE son consistentes con los resultados de otras mediciones estadísticas.

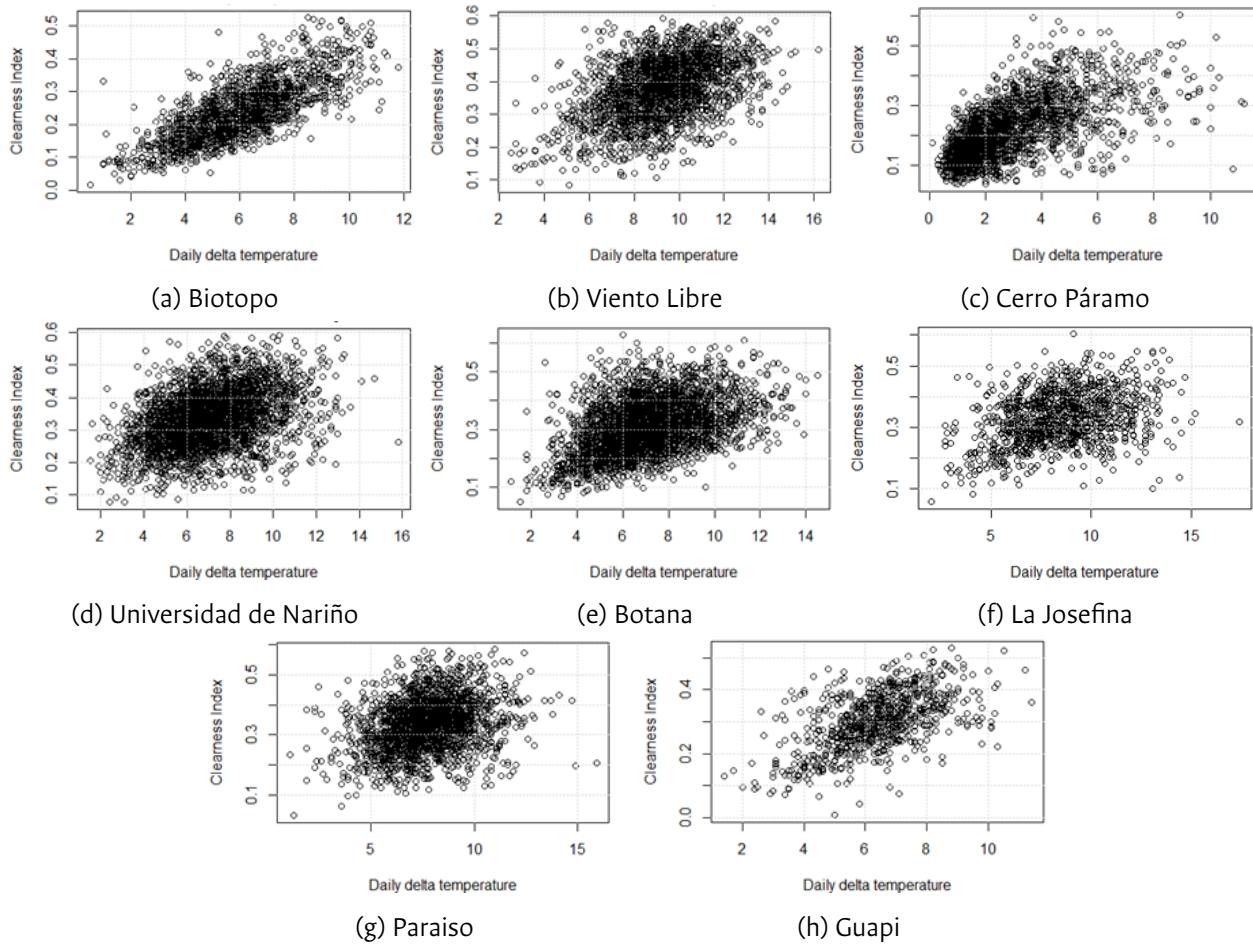
**Tabla 3-11.** Resumen de los resultados de los modelos empíricos

AWS	RMSE				SD			
	BC	HS	ON	Propuesto	BC	HS	ON	Propuesto
Biotoxo	1.152,62	<b>993,64</b>	1.155,07	1.113,48	49,90	<b>43,11</b>	50,08	48,29
Viento Libre	1.086,47	1.080,72	1.110,62	<b>1.077,35</b>	29,21	29,05	29,86	<b>28,97</b>
Cerro Páramo	1.194,57	1.209,76	1.196,56	<b>1.152,72</b>	56,04	56,77	56,15	<b>54,08</b>
Universidad de Nariño	1.032,45	1.083,73	1.032,24	<b>1.019,14</b>	31,89	33,55	31,88	<b>31,47</b>
Botana	1.052,42	1.070,23	1.077,17	<b>1.042,68</b>	34,46	35,05	35,23	<b>34,14</b>

Josefina	1.009,00	1.066,18	999,28	<b>984,75</b>	31,29	33,06	30,99	<b>30,53</b>
Paraiso	930,82	990,32	938,02	<b>921,32</b>	27,82	29,58	28,04	<b>27,54</b>
Guapi	965,40	<b>878,75</b>	961,21	915,53	31,75	<b>28,91</b>	31,58	30,12
MBE								
AWS	BC	HS	ON	Propuesto	BC	HS	ON	Propuesto
Biotoxo	-77,79	<b>-2,01</b>	-45,24	-37,29	916,08	<b>800,52</b>	917,96	885,10
Viento Libre	162,63	163,13	167,77	<b>160,23</b>	863,76	<b>861,64</b>	883,10	862,86
Cerro Páramo	37,90	<b>21,29</b>	27,22	33,37	940,46	946,29	929,90	<b>887,34</b>
Universidad de Nariño	90,20	<b>62,04</b>	93,68	93,72	845,12	884,48	841,37	<b>833,30</b>
Botana	48,24	<b>42,30</b>	71,63	47,13	866,59	881,19	888,56	<b>860,18</b>
Josefina	<b>5,28</b>	-20,58	16,92	22,18	769,70	830,06	767,58	<b>760,43</b>
Paraiso	-23,17	-42,52	-15,89	<b>-14,98</b>	757,02	806,64	760,64	<b>748,67</b>
Guapi	-36,98	<b>-16,38</b>	-53,45	-27,50	753,77	<b>696,35</b>	773,11	733,40
$U_{95}$								
AWS	BC	HS	ON	Proposed	BC	HS	ON	Proposed
Biotoxo	2.261,26	<b>1.949,37</b>	2.266,06	2.184,48	<b>16,22%</b>	19,52%	17,93%	18,33%
Viento Libre	2.130,26	2.118,99	2.177,61	<b>2.112,38</b>	15,34%	15,32%	15,39%	<b>15,18%</b>
Cerro Páramo	2.343,94	2.373,74	2.347,83	<b>2.261,82</b>	28,77%	27,90%	<b>27,64%</b>	28,69%
Universidad de Nariño	2.024,57	2.125,13	2.024,17	<b>1.998,46</b>	13,73%	<b>12,78%</b>	13,82%	13,80%
Botana	2.063,85	2.098,79	2.112,40	<b>2.044,75</b>	14,22%	<b>11,32%</b>	15,08%	14,11%
Josefina	1.978,59	1.941,90	1.959,53	<b>1.931,04</b>	<b>12,24%</b>	20,51%	12,52%	12,70%
Paraiso	1.825,23	1.941,90	1.839,34	<b>1.806,59</b>	8,24%	<b>7,70%</b>	8,59%	8,51%
Guapi	1.893,21	<b>1.723,28</b>	1.885,00	1.795,41	6,82%	7,46%	<b>6,11%</b>	7,12%
MAPE								
AWS	BC	HS	ON	Proposed				
Biotoxo	49,13%	<b>44,53%</b>	49,71%	48,13%				
Viento Libre	30,09%	29,99%	30,46%	<b>29,94%</b>				
Cerro Páramo	56,68%	56,62%	55,00%	<b>53,67%</b>				
Universidad de Nariño	31,44%	32,47%	31,26%	<b>30,97%</b>				
Botana	34,42%	<b>32,58%</b>	35,35%	34,07%				
Josefina	30,83%	37,42%	30,74%	<b>30,56%</b>				
Paraiso	26,75%	28,29%	26,91%	<b>26,47%</b>				
Guapi	28,15%	<b>26,14%</b>	28,65%	27,14%				

Mejores valores son resaltados

En el modelo propuesto, los coeficientes empíricos siguen una tendencia cuando estos valores se grafican contra la altitud del sitio. La Figura 3-4 muestra los ajustes lineales de los coeficientes empíricos

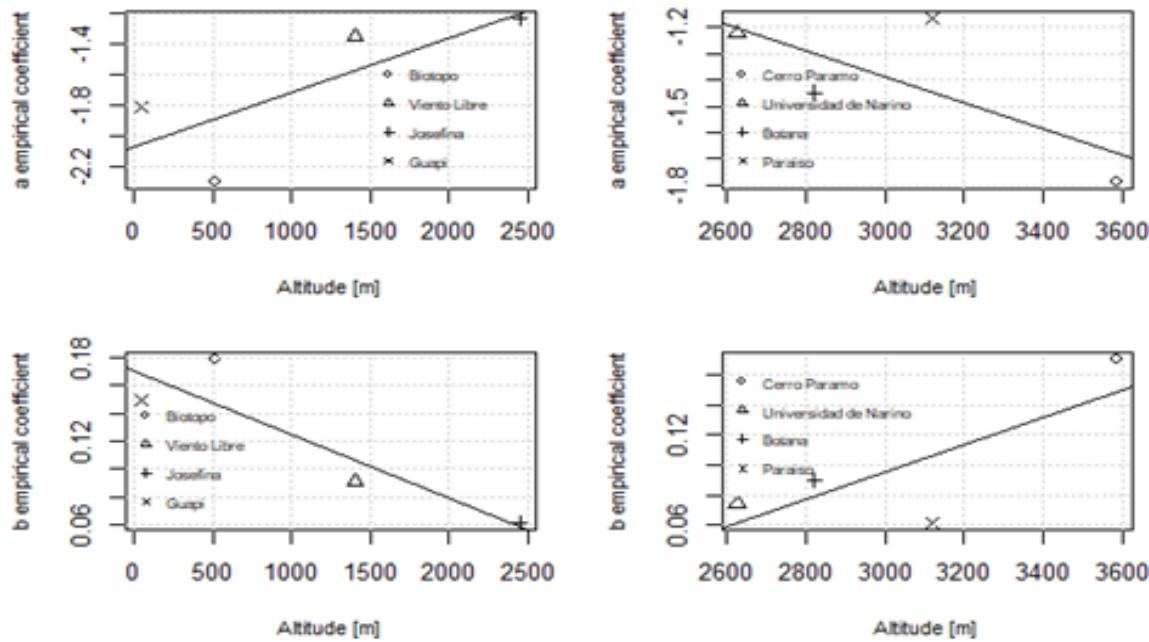
**Figura 3-3.** Temperatura contra índice de claridad

con la altitud, clasificada por la altitud de las AWS. En el lado izquierdo se encuentran los valores de las AWS con altitud por encima de los 2.500 msnm, y en el lado derecho las AWS restantes. El  $R^2$  de  $a$  es 0,5995 y 0,5262 en el primer y segundo caso, respectivamente. El  $R^2$  de  $b$  es 0,8152 y 0,6069 en el primer y segundo caso, respectivamente. Este es un resultado notable ya que en las zonas tropicales la temperatura cambia con la altitud, también la irradiancia solar aumenta con el crecimiento de la altitud, principalmente debido al aumento de la irradiancia directa, causado por la reducción de las moléculas de dispersión y absorción, así como del aerosol. (Blumthaler, 2012).

### 3.3.4 Imputación de datos de insolación solar diaria

Para el proceso de imputación de los datos de insolación solar diaria, los autores utilizaron en cada AWS el mejor modelo empírico. La Tabla 3-12 muestra la cantidad de datos imputados para cada AWS. La Josefina fue la AWS con la mayor cantidad de datos imputados con 2.870. Botana fue la AWS que menos requirió la imputación con 572. Haciendo un promedio entre todas las AWS, se puede concluir

**Figura 3-4.** Relación entre los coeficientes empíricos a y b y la altitud del sitio



que aproximadamente cuatro años fueron imputados mediante datos de temperatura. La Figura 3-5 muestra la serie de tiempo antes y después de la imputación diaria para cada AWS.

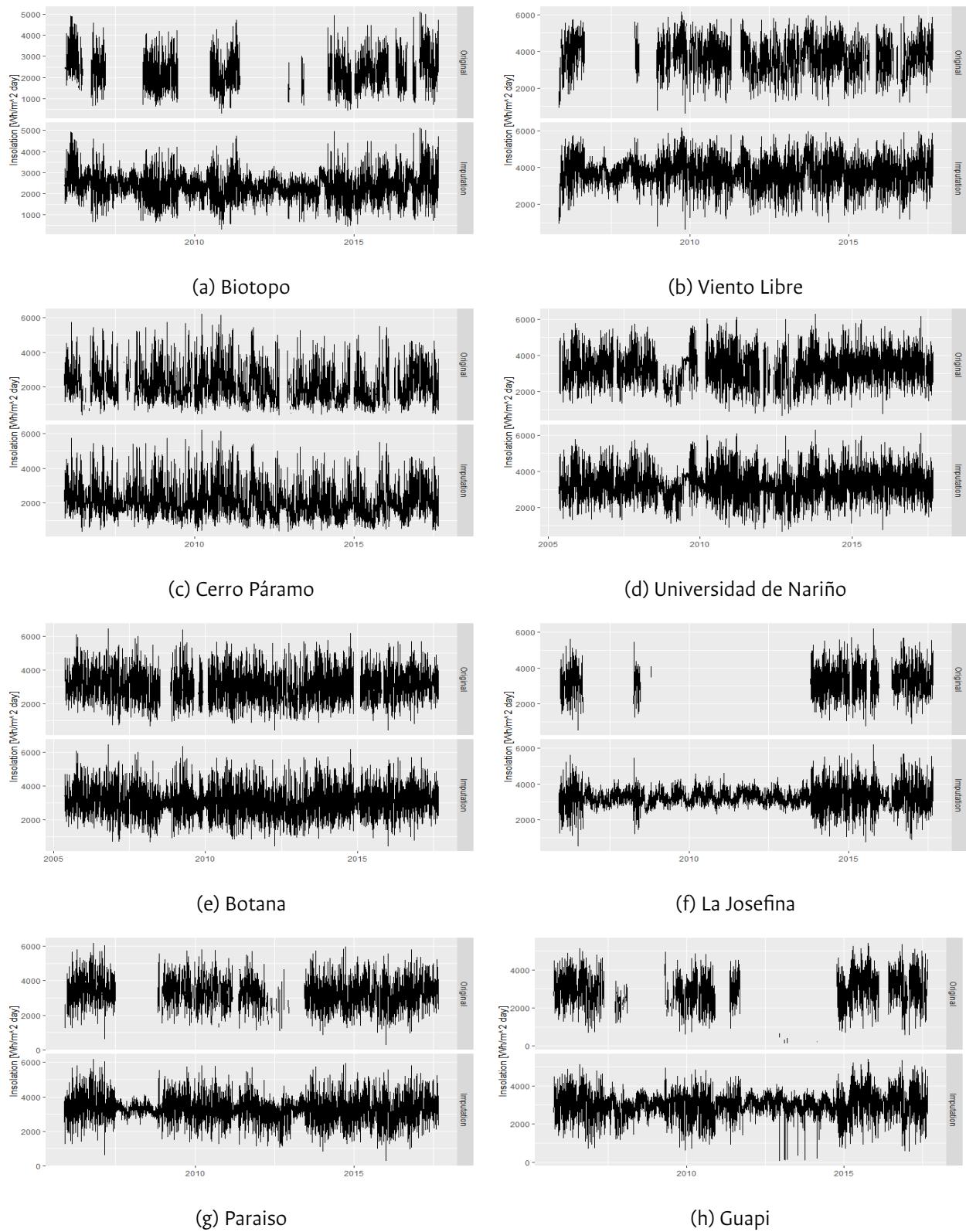
La Figura 3-7 muestra la insolación solar promedio diaria mensual para todas las AWS analizadas. Las AWS ubicadas en la zona del Pacífico presentaron un comportamiento similar con un pico en agosto y septiembre, y el nivel más bajo en noviembre. Viento Libre fue la única AWS que registró valores cercanos a 4.000 Wh/ ( $m^2 da$ ) en agosto. Las AWS de la zona Andina, Botana y Universidad de Nariño mostraron un pico entre octubre y noviembre. Además, entre Cerro Páramo y Viento Libre hay cierta complementariedad, porque el nivel más bajo en la primera se compensa con el nivel más alto de la segunda.

**Tabla 3-12.** Imputación por AWS

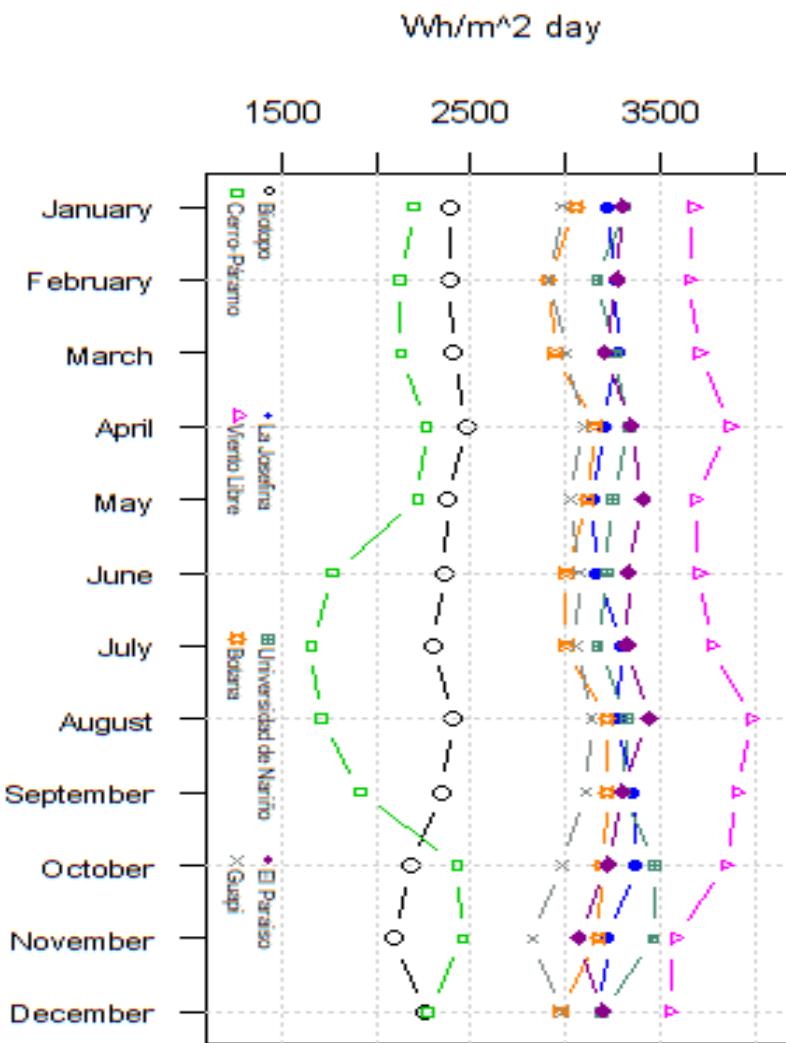
AWS	Imputation
Biotoxo	2.241
Viento Libre	1.502
Cerro Páramo	749
Universidad de Nariño	686
Botana	585
La Josefina	2.872
Paraiso	1.369
Guapi	2.277

### 3.4 Conclusiones

Los niveles de validación de los datos de irradiancia solar global tienen una fuerte influencia en los resultados de las variables empíricas. Considerando toda la información registrada, solo un 60,89 % supera los pasos de validación obligatorios propuestos. De ese valor, 95,81 %, que corresponde al número de días con al menos seis valores medidos, fue la información uti-

**Figura 3-5.** Imputación de datos

**Figura 3-7.** Insolación solar global diaria por mes



lizada en la calibración de los modelos empíricos. Sin embargo, este porcentaje representa en promedio el 33,90 % de la información total registrada en las AWS. Además, los días con información completa apenas llegan al 1,26 %. Este resultado indica la calidad de la serie temporal y la necesidad de mejorar y aumentar los procedimientos de mantenimiento y calibración.

En el departamento de Nariño, el desempeño de las AWS es un factor determinante, considerando el predominio de días parcialmente nublados que representan el 64,7 % de los días. En otras palabras, en Nariño existe una alta interacción de nubes que dificulta la estimación de la insolación solar; por tanto, aumenta la necesidad de medidas fiables. En consecuencia, es importante establecer un plan para realizar estos procedimientos con regularidad y seguir estándares de alta calidad y ampliamente aceptados. También fue notoria la necesidad de instalar más AWS para aumentar los puntos de muestreo.

En cuanto a las medidas de temperatura, de la cantidad de datos que superan los pasos de validación horaria, se tomó un 92,78 % para la calibración e imputación empírica. Además, el número de mediciones de temperatura diurnas presenta picos en once y doce valores diarios; esto significa que la mayoría de los días que se usan para modelar y llenar la base de datos con el proceso de imputación, tienen el 88,46 % de la información total en promedio.

El modelo propuesto mostró una relación lineal entre los coeficientes empíricos y la altitud del sitio de estudio. Los coeficientes empíricos se clasificaron entre aquellos por encima y por debajo de 2.500 msnm. El ajuste lineal para casos por debajo de 2.500 msnm muestra un  $R^2$  de 0,5995 y 0,5262 para  $a$  y  $b$  respectivamente. El ajuste lineal para casos por encima de 2.500 msnm presenta un  $R^2$  de 0,8182 y 0,6069 para  $a$  y  $b$  respectivamente. Este resultado es consistente con el comportamiento de la temperatura en zonas tropicales y la irradiancia solar global en grandes altitudes.

Cuando se consideraron RMSE, SD, MAE,  $U_{95}$  y MAPE, el modelo propuesto tuvo un mejor desempeño en cinco de los ocho casos evaluados. Estos casos se encuentran en la zona Andina y Amazónica, con altitudes superiores a los 2.500 msnm. Como resultado, el modelo propuesto se utiliza para imputación en las AWS de la zona Andina y Amazónica. Para la AWS ubicada en la zona del Pacífico, el modelo HS fue la mejor opción, seguido por el modelo propuesto en esta investigación. El modelo propuesto tuvo un buen desempeño en este ambiente tropical y montañoso. Sin embargo, es necesario analizar más información proveniente de otros lugares con las mismas características. Para lograr esto, el requisito principal es aumentar el número de AWS y la calidad de las series de tiempo en los entornos tropicales y montañosos.

# **4 Evaluación de modelos empíricos para estimar irradiancia solar global usando datos de temperatura en un ambiente tropical y montañoso: Parte II: mapeo<sup>1</sup>**

## **Resumen**

Los mapas de potencial de energía solar son elementos facilitadores para la toma de decisiones sobre la instalación de sistemas de energía solar fotovoltaica. Sin embargo, la falta de información registrada *in situ* sobre irradiancia solar es una barrera para apoyar la toma de decisiones en este campo de estudio. Por lo tanto, esta investigación combina la estimación de irradiancia solar utilizando métodos empíricos basados en la temperatura como Hargreaves y Samani y el modelo logístico, para incrementar los puntos muestrados, con técnicas de interpolación espacial. Además, utiliza la validación cruzada de dejar uno fuera para evaluar el rendimiento de cuatro técnicas de interpolación espacial en entornos tropicales y montañosos que cubren un área de  $33.268 \text{ km}^2$ . El análisis permitió concluir que el modelo de Hargreaves y Samani es mejor para la zona Pacífica, mientras que el Logístico es mejor para las zonas Andina y Amazónica. Además, el método kriging ordinario fue la mejor técnica de interpolación porque presentó el menor sesgo.

**Palabras clave** Modelos basados en temperatura, Hargreaves y Samani, técnicas de interpolación espacial, mapeo de irradiancia solar.

## **4.1 Introducción**

La información sobre irradiancia solar es fundamental en campos de estudio como la generación de electricidad, el pronóstico del clima, la producción agrícola y la ecología (Moreno et al., 2011). Esta

---

<sup>1</sup>Este artículo tuvo la participación de Belizza Ruiz

investigación aborda el tema de la generación de electricidad con energía solar, que requiere información de alta calidad para diseñar plantas de potencia de manera óptima. La falta de datos o información de baja calidad es una barrera para conocer el potencial solar y diseñar centrales eléctricas. Esta situación es más marcada en áreas aisladas con infraestructura vial inexistente o deficiente, como regiones montañosas y ambientes tropicales (Janjai et al., 2005; J. Li & Heap, 2014). A pesar de lo anterior, las entidades gubernamentales encargadas de ofrecer soluciones de servicio eléctrico en áreas aisladas, promueven proyectos de electrificación mediante irradiación solar. Las partes interesadas de los sectores público y privado enfrentan una falta de información que afecta el desarrollo exitoso de los proyectos; por lo tanto, utilizar mapas de potencial de energía solar para tomar decisiones y podría evitar una planificación inadecuada.

La literatura académica y los informes gubernamentales presentan un atlas de irradiancia solar para la mayor parte del territorio global, mostrando valores indicativos del potencial solar. Una parte de esta información utiliza datos satelitales para elaborar mapas, mientras que otros utilizan datos *in situ*. Los datos a nivel del suelo ofrecen información completa sobre el entorno donde existen los registros de variables meteorológicas; sin embargo, la cantidad de estaciones meteorológicas automáticas (AWS) no es adecuada en todos los casos. Por lo tanto, es necesario estimar los valores de irradiancia solar para lugares sin sensores de irradiancia solar, para obtener información sobre los puntos no muestreados y obtener datos espacialmente continuos sobre el área de estudio (J. Li & Heap, 2011). Las técnicas de interpolación permiten diseñar modelos regionales útiles para estandarizar muestras de puntos aislados (Moreno et al., 2011). La calidad de los resultados depende de la densidad espacial y la agrupación de los puntos muestreados, el tipo de superficie, entre otros aspectos (J. Li & Heap, 2011). En general, hay dos categorías para las técnicas de interpolación: no geoestadísticas y geoestadísticas. Sin embargo, todos los métodos comparten la misma fórmula de estimación general (J. Li & Heap, 2008). Las técnicas no geoestadísticas estiman valores de puntos sin muestreo con funciones matemáticas ajustadas de acuerdo con la información de puntos muestreados. Las técnicas geoestadísticas establecen una correlación espacial entre los datos (Martín & Dominguez, 2019). El número de puntos muestreados es decisivo para definir la calidad de la información porque el valor de la irradiancia solar se ve afectado no solo por nubosidad, aerosoles, humedad relativa, entre otros factores, sino también por la fisiografía y tipo de suelo del sitio de estudio (Jeffrey et al., 2001; Sen, 2008).

Esta investigación evalúa una combinación de modelos empíricos basados en la temperatura y técnicas de interpolación espacial en un ambiente tropical y montañoso para elaborar mapas mensuales de insolación solar. En consecuencia, esta investigación utiliza mediciones *in situ* de la irradiancia solar global recolectadas en ocho AWS ubicadas en el departamento de Nariño en el suroeste de Colombia. Los puntos muestreados aumentaron mediante la implementación de dos modelos empíricos basados en temperatura para estimar la irradiancia solar global en diecisésis ubicaciones con estaciones meteorológicas convencionales (CWS). Asimismo, se utilizó técnicas de interpolación espacial como el Ponderado Inverso por Distancia (IDW) y Kriging simple, ordinario y universal para obtener información de insolación solar espacialmente continua en la región estudiada. Además, la validación cruzada de dejar uno fuera es el proceso implementado para seleccionar la mejor técnica de interpolación espacial. R-CRAN

es el software y gstat es el paquete utilizado para realizar todos los cálculos.

Los resultados muestran que el modelo empírico de Hargreaves y Samani funciona mejor en la zona del Pacífico. En la zona Andina, la relación lineal entre las constantes empíricas y la altitud, derivada del modelo logístico, reduce la sobreestimación. Si bien este método podría tener ajustes derivados de futuras investigaciones, ha presentado resultados satisfactorios (Hoyos-Gómez, Ruiz, 2020). Por lo tanto, estima información para la región andina. El método de interpolación Kriging ordinario fue la mejor entre los métodos analizadas porque presentó menos sesgo durante todo el año debido a su robustez.

Este artículo tiene las siguientes secciones: La sección 4.2 muestra la descripción de los datos. La sección 4.3 presenta el marco teórico que explica las técnicas de interpolación espacial utilizadas en este trabajo. La sección 4.4 contiene la metodología seguida en esta investigación. La sección 4.5 muestra los resultados y la discusión. Finalmente, la Sección 4.6 presenta las conclusiones.

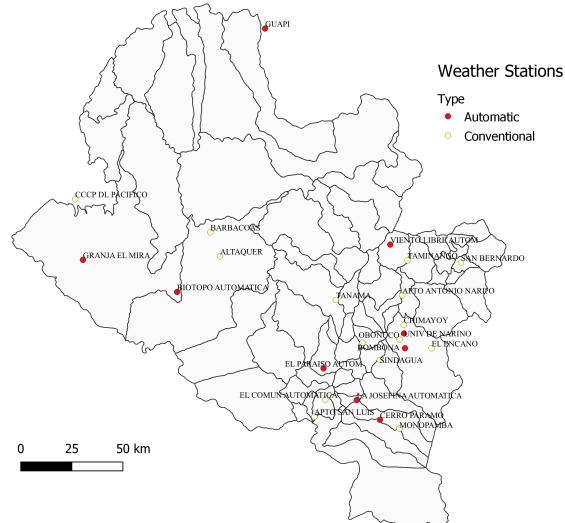
## 4.2 Descripción de los datos

La Tabla 4-1 y 4-2 enumeran las AWS y CWS clasificadas por región, respectivamente, y la Figura 4-1 muestra su ubicación en el departamento de Nariño. Hay 5 AWS y 11 CWS en la región Andina, 3 AWS y 4 CWS en la región del Pacífico, y 1 AWS y 1 CWS en la región amazónica. La altitud AWS varía entre 16 y 3.577 msnm, y la altitud CWS varía entre 1 y 3.141 msnm. La AWS mide la irradiancia solar global y la temperatura, entre otras variables, mientras que el CWS mide solo temperatura. La densidad de las AWS en el departamento de Nariño es de aproximadamente  $4.158 \text{ km}^2$  por estación aproximadamente; agregando las CWS, la densidad mejora, pasando a  $1.386 \text{ km}^2$  por estación.

**Tabla 4-1.** Estaciones meteorológicas automáticas

Nombre	Latitud	Longitud	Altitud	Periodo	Región
Biotopo	1,41	-78,28	512	2005-2017	Pacífico
Altaquer	1,56	-79,09	1.1010	2013-2014	Pacífico
Granja el Mira	1,55	-78,69	16	2016-2017	Pacífico
Cerro-Páramo	0,84	-77,39	3.577	2005-2017	Amazonia
La Josefina	0,93	-77,48	2.449	2005-2017	Andino
Viento Libre	1,62	-77,34	1.005	2005-2017	Andino
Universidad de Nariño	1,23	-77,28	2.626	2005-2017	Andino
Botana	1,16	-77,27	2.820	2005-2017	Andino
El Paraíso	1,07	-77,63	3.120	2005-2017	Andino

La irradiancia solar global y la temperatura registradas en las AWS fueron sometidas a un procedimiento de control de calidad descrito en (Hoyos-Gómez y Ruiz, 2020). La información de temperatura de las

**Figura 4-1.** Localización de las estaciones meteorológicas**Tabla 4-2.** Estaciones meteorológicas convencionales

Nombre	Latitud	Longitud	Altitud	Periodo	Región
CCCP del Pacífico	1,82	-78,73	1	2005-2017	Pacífico
Altaquer	1,56	-79,09	1.1010	2005-20134	Pacífico
Granja el Mira	1,55	-78,69	16	2005-2017	Pacífico
Obonuco	1,19	-77,30	2.710	2005-2015	Andean
Apto. Antonio Nariño	1,39	-77,29	1.796	2005-2017	Andino
San Bernardo	1,53	-77,03	2.190	2005-2017	Andino
Taminango	1,55	-77,27	1.875	2005-2017	Andino
Común el antomática	0,93	-77,63	3.141	2007-2017	Andino
Apto. San Luis	0,86	-77,67	2.961	2005-2017	Andino
Bombona	1,18	-77,46	1.493	2005-2017	Andino
Tanama	1,37	-77,58	1.500	2005-2017	Andino
Sindagua	1,11	-77,39	2.800	2005-2017	Andino
Barbacoas	1,67	-78,13	32	2005-2012	Pacífico
Monopamba	0,99	-77,15	2.719	2006-2016	Amazonia
El Encano	1,15	-77,16	2.830	2005-2017	Andino
Chimayoy	1,26	-77,28	2.745	2005-2014	Andino

CWS fue obtenida ya validada. La Tabla 4-3 y la Tabla 4-4 presentan el número de días clasificados por el número de datos registrados. Con respecto a la irradiancia solar global, los días tienen principalmente 10 u 11 datos registrados, mientras que en el caso de la temperatura principalmente el día tiene de 11 a 12 datos registrados. Es importante señalar que solo el 1,26 % del total de días medidos registra todos los valores diurnos de la irradiancia solar global.

**Tabla 4-3.** Número de días clasificados por cantidad de mediciones diarias de irradiancia global solar

Nombre	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Biotoxo	13	16	15	19	30	49	104	190	350	528	622	207	6	2.149
Viento Libre	43	50	77	86	121	175	313	444	702	654	390	119	11	3.185
Universidad de Nariño	27	42	56	73	97	173	340	403	730	1.014	778	308	63	4.104
Cerro Páramo	34	39	42	45	59	83	145	283	483	952	1.080	362	160	3.767
La Josefina	109	71	22	24	25	57	78	195	293	386	274	134	6	1.74
Botana	16	20	26	43	82	142	246	469	798	1.066	839	336	14	4.097
El Paraíso	55	83	110	137	158	205	285	364	614	787	514	149	13	3.474
Guapi	185	193	180	153	111	101	139	239	335	447	549	182	75	2.889

**Tabla 4-4.** Número de días clasificados por cantidad de mediciones diarias de temperatura

Nombre	1	2	3	4	5	6	7	8	9	10	11	12	Total
Biotoxo	33	9	17	13	26	46	56	85	130	257	605	846	2.123
Viento Libre	79	68	91	91	94	120	124	192	264	500	873	696	3.192
Universidad de Nariño	74	26	24	31	47	71	113	198	309	599	1.170	1.379	4.041
Cerro Páramo	101	49	58	47	67	86	133	190	348	590	754	701	3.124
La Josefina	61	21	36	46	50	78	109	130	197	405	573	558	2.264
Botana	38	11	27	46	62	102	147	253	386	679	1.114	1.221	4.086
El Paraíso	74	42	65	67	92	102	155	236	373	541	915	909	3.571
Guapi	20	8	10	21	18	21	45	79	168	405	919	1.349	3.063

### 4.3 Marco teórico

El creciente uso de la tecnología solar fotovoltaica ha aumentado el interés en la información sobre el potencial solar en los territorios. Ese conocimiento conlleva a la obtención de datos espaciales continuos de irradiancia solar que surgen de los procesos de interpolación espacial. Los métodos de interpolación espacial estiman los valores de irradiancia solar en puntos no muestrados a partir de datos medidos utilizando la información de variabilidad espacial (J. Li & Heap, 2008). La teoría de la variable regionalizada, el método del semivariograma y el semivariograma acumulativo son los enfoques utilizados para modelar la variabilidad espacial utilizando la irradiancia solar y la distancia (Şen, 2008).

La densidad y la distribución espacial de los puntos muestrados, la agrupación de los puntos muestrados, el tipo de superficie, la varianza de los datos, entre otras cosas, influyen en el rendimiento de

los métodos de interpolación espacial (J. Li & Heap, 2011). Hay tres categorías de interpolación: no geoestadísticas, geoestadísticas y una combinación de estas dos. En la primera categoría, los métodos modelan parcialmente la autocorrelación espacial a través de funciones matemáticas; algunos de estos métodos son Vecino natural (NaN), Ponderado inverso por distancia (IDW), Red irregular triangular (TIN), Modelos de regresión, entre otros. En la segunda categoría, los métodos simulan la autocorrelación de los datos espaciales y evalúan la incertidumbre de los resultados para realizar los procesos de interpolación como Kriging que es el método más utilizado (Bhattacharjee et al., 2019; Martín & Dominguez, 2019; Sankar et al., 2018).

Todos los métodos comparten el mismo estimador general, presentado en la Ecuación 4-1, basado en la idea de que las muestras de estaciones meteorológicas cercanas tienen una mayor probabilidad de ser similares que las muestras tomadas en estaciones meteorológicas remotas (J. Li & Heap, 2008).

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (4-1)$$

Donde  $\hat{z}$  es el valor estimado en el punto  $x_0$ ,  $z$  es el valor medido en el punto  $x_i$ ,  $\lambda_i$  es el peso del punto muestreado y  $n$  es el número de puntos muestreados utilizados para la estimación (J. Li & Heap, 2008). El principal objetivo de cualquier método es determinar estos pesos lo mejor posible (Şen, 2008). La siguiente subsección presenta una breve explicación de las técnicas IDW y Kriging.

### 4.3.1 Ponderado inverso por distancia (IDW)

IDW estima el valor de los puntos no muestreados a partir de una combinación lineal de los puntos muestreados cercanos. El inverso de la distancia euclíadiana es la función que se usa para calcular el peso de cada punto muestreado sobre un punto no muestreado, ver Ecuación 4-2. Es decir, el punto muestreado más lejano tiene la contribución más baja al valor calculado del punto no muestreado (Martín & Dominguez, 2019).

$$\lambda_i = \frac{1/d_i^P}{\sum_{i=1}^n 1/d_i^P} \quad (4-2)$$

Donde  $d_i$  es la distancia entre  $x_i$  y  $x_0$ ,  $p$  es el parámetro del exponente y  $n$  es el número de los puntos muestreados utilizados para la estimación (J. Li & Heap, 2008). El factor principal que influye en la precisión del método es el valor del parámetro del exponente. La selección de este parámetro y el tamaño de la vecindad son arbitrarios, pero cuando el valor del parámetro  $p$  aumenta, la contribución del valor más lejano disminuye. Además, si el valor del parámetro  $p$  cambia, el tipo de interpolación también cambia, el IDW es una interpolación de promedio móvil cuando  $p = 0$ , interpolación lineal con  $p = 1$  y ponderada media móvil cuando  $p$  es diferente a 1. El exponente más común es  $p = 2$  llamado Distancia inversa al cuadrado (IDS) (J. Li & Heap, 2011; Introduction to Spatial Analysis, 2009). Para IDS

$\lambda_i$  se calcula con la siguiente expresión  $\lambda_i = \left(\frac{r-d_i}{d_i}\right)^2$ , donde  $r$  es el radio de la ventana de búsqueda (Bhattacharjee et al., 2019).

### 4.3.2 Kriging

Kriging es un algoritmo de regresión de mínimos cuadrados generalizado con una predicción insesgada (Jamaly & Kleissl, 2017). La precisión de Kriging depende del variograma para obtener un predictor de varianza mínima. El variograma se estima utilizando la semivarianza, que considera tanto la distancia como la variación entre los puntos estimados y medidos (Introduction to Spatial Analysis, 2009). La correlación espacial analiza el cambio de datos de acuerdo con la semivarianza, que es la diferencia entre las varianzas de los puntos medidos y estimados como la Ecuación 4-3 presenta (M. A. Oliver & Webster, 1990).

$$\hat{\gamma}(x_i, x_0) = \hat{\gamma}(h) = \frac{1}{2n} \sum_{i=1}^n (z(x_i) - z(x_i + h))^2 \quad (4-3)$$

Donde  $\hat{\gamma}(h)$  es la semivariancia, que es la mitad de la varianza entre el punto estimado y los puntos medidos,  $n$  es el número de pares de puntos de muestra separados por una distancia  $h$ ,  $z(x_i)$  es el valor de muestra en la ubicación  $x_i$ , y  $z(x_i + h)$  es el valor a una distancia  $h$  de  $x_i$ . A distancias cortas, valores bajos de  $h$ , valores pequeños de semivarianza, cuando  $h$  aumentan, la semivarianza también aumenta.

La ecuación 4-3 permite calcular el variograma experimental que consta de semivarianzas en un conjunto finito de intervalos. La fiabilidad del variograma experimental está relacionada con la calidad de los datos y su densidad (A. M. Oliver & Webster, 2015). El siguiente paso es ajustar una función al variograma experimental conocido como variograma teórico, que describe las principales características de los valores muestrados. La pendiente del variograma teórico cuantifica la autocorrelación espacial entre los puntos de muestreo a determinadas distancias (Martín & Dominguez, 2019). Hay dos formas de modelar el variograma ilimitado y limitado. En el primer caso, el variograma aumenta indefinidamente al aumentar la distancia. En el segundo caso, el variograma sigue un proceso estacionario de segundo orden, es decir alcanza un límite superior que es la varianza del umbral (Webster & Oliver, 2007). Además, el variograma teórico tiene un rango que determina el límite de la correlación espacial donde la autocorrelación se vuelve cero. Además, cuando hay una distancia cero entre la estimación y el punto medido, el semivariograma es cero. Sin embargo, si hay errores de medición o escala en la variable o en ambos, el semivariograma presenta un efecto de nugget. En otras palabras, el semivariograma tiene un valor alto en distancias cortas (Viera Díaz, 2002).

El variograma es la base de la implementación de los métodos de Kriging. La Ecuación 4-4 muestra el estimador general de Kriging, donde  $\mu$  es una media estacionaria,  $\lambda_i$  es el peso de Kriging,  $\mu(x_0)$  es la

media de las muestras consideradas en el área de influencia,  $n$  es el número de puntos muestreados (J. Li & Heap, 2008). Siguiendo la estructura de la estimación general mostrada en la Ecuación 4-1.

$$\hat{z}(x_0) - \mu = \sum_{i=1}^n \lambda_i [z(x_i) - \mu(x_0)] \quad (4-4)$$

Los pesos surgen de un sistema de ecuaciones lineales que cumplen con dos condiciones. La primera es la siguiente restricción  $E\{\hat{z}(x_0) - z(x_i)\} = 0$  para obtener pesos no sesgados. El segundo es que los pesos deben minimizar la varianza  $\sigma(x_i) = Var\{\hat{z}(x_0) - z(x_i)\} = 0$  (Moreno et al., 2011). Aunque existen varios métodos de Kriging, solo hay un objetivo que es estimar valores continuos para los puntos sin muestreo (Dai et al., 2003).

### Kriging Simple (SK)

SK tiene la formulación matemática más sencilla entre los métodos de Kriging. Este modelo asume que la media y la covarianza son conocidas a priori y constantes en todas las ubicaciones (Martín & Dominguez, 2019). Como resultado, cuando la media y covarianza en el sitio estimado es diferente de los valores iniciales asumidos, el método no muestra resultados óptimos, lo que lo convierte en un modelo de aplicabilidad limitada (Olea, 1999). La Ecuación 4-5 muestra el modelo de estimación de SK.

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) + \left[ 1 - \sum_{i=1}^n \lambda_i \right] \mu \quad (4-5)$$

Donde  $\mu$  es una media estacionaria conocida para toda la superficie interpolada. En SK  $1 - \sum_{i=1}^n \lambda_i$  no es necesariamente igual a 0; sin embargo, el segundo término de 4-5 asegura que la predicción no tenga sesgo (J. Li & Heap, 2008; A. M. Oliver & Webster, 2015). SK trabaja con covarianzas,  $C$ , más que con las semivarianzas; por lo tanto, los ponderados se calculan de la siguiente manera:

$$\sum_{i=1}^n \lambda_i C(x_i, x_j) = C(x_0, x_j) \quad \text{For all } j = 1, 2, \dots, n \quad (4-6)$$

### Kriging Ordinario (OK)

OK es el método de Kriging más general y ampliamente utilizado, éste asume que la variación es aleatoria y dependiente espacialmente, y una media desconocida que es constante y una varianza que depende

solo de la distancia a los puntos muestrados, la Ecuación 4-7 muestra el modelo de estimación de OK

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) + \left[ 1 - \sum_{i=1}^n \lambda_i \right] \mu(x_0) \quad (4-7)$$

Este método tiene la siguiente restricción  $[1 - \sum_{i=1}^n \lambda_i] = 0$  para cumplir con la condición de no sesgo (J. Li & Heap, 2008). OK tiene como objetivo minimizar la varianza de los errores y utiliza combinaciones lineales para estimar los ponderados del conjunto de datos reduciendo el sesgo (Kiš, 2016), como la Ecuación 4-8 y 4-9 muestran

$$Var[\hat{z}(x_0)] = E[\{\hat{z}(x_0) - z(x_0)\}^2] = 2 \sum_{i=1}^n \lambda_i \hat{\gamma}(x_i - x_0) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \hat{\gamma}(x_i - x_j) \quad (4-8)$$

$\hat{\gamma}(x_i - x_0)$  es la semivariancia entre el i-ésimo y el j-ésimo punto. El variograma teórico debe garantizar que la varianza no sea negativa A. M. Oliver & Webster (2015); Webster & Oliver (2007)

$$\sum_{i=1}^n \lambda_i \hat{\gamma}(x_i - x_0) + \psi(x_0) = \hat{\gamma}(x_i - x_j) \quad (4-9)$$

Cumpliendo con  $\sum_{i=1}^n \lambda_i = 1$ ,  $\psi(x_0)$  es el multiplicador de Lagrange utilizado para minimizar la varianza. En forma de matriz, los ponderados se estiman de la siguiente manera:

$$\lambda = A^{-1}b \quad (4-10)$$

### Kriging Universal (UK)

UK es un método utilizado para procesos espaciales con una tendencia y una media no estacionaria; por lo tanto, este método incorpora la media como una tendencia local en función de las coordenadas ( $X, Y$ ). La variación en  $z(x)$  tiene un componente sistemático, que podría descomponerse en una combinación lineal de funciones deterministas; además, tiene un componente aleatorio expresado por  $z(x) = u(x) + \varepsilon(x)$ . Donde  $u(x)$  reemplaza la media  $\mu$ ,  $\varepsilon(x)$  es una función residual aleatoria correlacionada espacialmente (Bhattacharjee et al., 2019; J. Li & Heap, 2008; Olea, 1999; Kiš, 2016; A. M. Oliver & Webster, 2015; Webster & Oliver, 2007).

Como en este caso hay un proceso no estacionario, el variograma experimental no estima el variograma del residuo aleatorio,  $\varepsilon(x)$ , como en los dos métodos presentados anteriormente. En consecuencia, el variograma se estima a partir de  $\varepsilon(x) = z(x) - u(x)$ . Para estimar el variograma sin sesgo, es necesario separar  $u(x)$  de  $\varepsilon(x)$ ; entonces  $u(x)$  se puede expresar de la siguiente manera:

$$u(x) = \sum_{k=0}^K \beta_k f_k(x) \quad (4-11)$$

Donde  $\beta_k$ , con  $k = 0, 1, \dots, K$  son coeficientes desconocidos estimados a partir de los datos, y  $f_k(x)$  son funciones deterministas conocidas de las coordenadas espaciales que describen la tendencia. Para una tendencia lineal, hay tres funciones  $f_0 = 1$ ,  $f_1 = x_1$  y  $f_2 = x_2$ . Para una tendencia cuadrática es necesario agregar tres funciones  $f_3 = x_1^2$ ,  $f_4 = x_1 x_2$  y  $f_5 = x_2^2$ . Por ejemplo, con una tendencia lineal, la ecuación se expresa como sigue  $z(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon(x)$ . Finalmente, con las funciones para estimar  $u(x)$  y  $\varepsilon(x)$ , es posible predecir el valor  $\hat{z}$  en cualquier punto  $x_0$

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i f_k(x_i) \quad (4-12)$$

La esperanza es

$$E[\hat{z}(x_0)] = \sum_{k=0}^K \sum_{i=1}^n \beta_k \lambda_i f_k(x_i) \quad (4-13)$$

El estimador es insesgado si satisface  $\sum_{i=1}^n \lambda_i f_k(x_i) = f_k(x_0)$  para todos  $k = 0, 1, \dots, K$

### 4.3.3 Validación cruzada

La validación cruzada evalúa y compara las técnicas de interpolación espacial. El primer paso consiste en definir un caso base en el que todos los puntos muestreados permitan estimar los puntos no muestreados. Posteriormente, los cálculos se repiten para encontrar el valor de un punto muestreado eliminado y así sucesivamente hasta calcular valores como puntos muestreados (Sankar et al., 2018). La mejor técnica de interpolación se determina a partir del análisis de los errores de las comparaciones entre los resultados de cada cálculo con puntos muestreados eliminados y los puntos muestreados del caso base (Berrar, 2018; Sen, 2008).

## 4.4 Metodología

El desarrollo de la investigación abarca tres etapas. La primera etapa consiste en estimar la insolación solar mediante modelos empíricos. La segunda parte es la implementación de las técnicas de interpolación espacial utilizando el paquete gstat de R-CRAN. La tercera parte comprende la validación cruzada para seleccionar la mejor opción que permita mapear la irradiancia solar global.

El modelo de Hargreaves y Samani es usado para estimar la insolación solar en la zona del Pacífico. En las zonas Andina y Amazónica, la relación lineal entre los coeficientes empíricos del modelo logístico

y la altitud de las AWS es la técnica propuesta para estimar los coeficientes empíricos asociados a las CWS. Los coeficientes son valiosos porque pertenecen al modelo empírico basado en la temperatura que estima la insolación solar global para cada CWS.

La segunda etapa consiste en mapear la insolación solar global utilizando las técnicas de interpolación espacial explicadas en 4.3.1 y 4.3.2. Al comienzo del proceso de mapeo, los autores de esta investigación construyeron un cuadrícula con una resolución de 100 metros. Luego, las técnicas de Kriging ajustan una función al variograma experimental representado por un diagrama de dispersión entre la semivarianza y la distancia. Los valores del variograma experimental permiten determinar los valores iniciales con los que se calcula el variograma teórico. En esta investigación, los valores iniciales cumplieron con las siguientes recomendaciones: la tercera parte de las distancias máximas del variograma experimental constituyen el rango, la media de los tres primeros valores del variograma experimental determinan la varianza del *nugget* y la media de los cinco últimos valores del variograma experimental establecen el umbral parcial (Pebesma, 2016). El paquete gstat tiene veinte opciones para determinar la función del variograma teórica (Pebesma & Graeler, 2020). Los autores ajustaron todas las funciones disponibles y eligieron la mejor opción para cada caso.

La última etapa fue la validación cruzada para determinar la mejor técnica de interpolación y obtener los mapas mensuales de insolación solar global. Cada técnica de interpolación tenía un análisis de error asociado que consistía en calcular el error cuadrático medio (RMSE), el error absoluto medio (MAE), el error de sesgo medio (MBE), la desviación estándar (SD) y el error porcentual medio (MPE). Las técnicas con menos errores indicaron la mejor opción para cada mes. Es conveniente resaltar que los errores estadísticos surgieron del promedio de todos los resultados obtenidos por cada iteración durante el proceso de validación cruzada.

## 4.5 Resultados y discusión

Los modelos logísticos y de Hargreaves y Samani permitieron estimar la insolación solar diaria utilizando datos de temperatura del aire para aumentar el número de puntos muestrados. El modelo de Hargreaves y Samani permitió calcular la constante empírica con datos de AWS ubicadas en la región del Pacífico; en consecuencia, dicha constante empírica sirvió para estimar la insolación solar utilizando datos de temperatura del aire de las ubicaciones de CWS.

Es conveniente recordar que existen tres AWS en la zona del Pacífico; sin embargo, solo una tiene consistencia temporal con las CWS de la zona. En consecuencia, la información meteorológica de Biotopo permitió estimar el comportamiento de la irradiancia solar de las CWS. Aunque Quansah et al. sugieren una distancia máxima de 20 km para trasponer la información meteorológica de un punto a otro, los autores de esta investigación no pudieron utilizar ese criterio, porque los marcos temporales de las estaciones meteorológicas más cercanas no coincidían. Sin embargo, es conveniente mostrar las dis-

tancias entre las estaciones meteorológicas. Hay una distancia de 30,18 km entre Granja el Mira AWS y CCCP del Pacífico, 67,54 km entre Biotopo AWS y CCCP del Pacífico, 63,71 km entre Granja el Mira y Barbacoas, y 68,79 km entre Biotopo y Barbacoas.

La relación lineal entre la altitud de las AWS y las constantes empíricas del modelo logístico permitió calcular las constantes empíricas de las CWS ubicados en las zonas Andina y Amazónica. Es un hallazgo relevante porque el uso de la relación lineal facilita la estimación de la insolación solar global utilizando la temperatura del aire en las zonas andina y amazónica, que en este caso no cuenta con una cantidad considerable de estaciones meteorológicas.

$$a = \begin{cases} -2,0565 + 0,000357h & h \leq 2.500 \\ 0,0910 - 0,0004932h & h > 2.500 \end{cases} \quad (4-14)$$

$$b = \begin{cases} 0,1703 - 4,566e^{-5}h & h \leq 2.500 \\ -0,8 + 9,182e^{-5}h & h > 2.500 \end{cases} \quad (4-15)$$

Donde  $h$  es la altitud de la estación meteorológica. Los valores de los coeficientes empíricos se reemplazan en la Ecuación 4-16.

$$\frac{H}{H_0} = \frac{1}{1 + e^{-(a+b\Delta T)}} \quad (4-16)$$

Donde  $H$  es la insolación solar global promedio diaria,  $H_0$  es la insolación solar extraterrestre promedio diaria,  $a$  y  $b$  son los coeficientes empíricos,  $\Delta T$  es la diferencia entre la temperatura diaria máxima y mínima. La Tabla 4-5 presenta los resultados de los coeficientes empíricos obtenidos para las CWS usando la Ecuación 4-14 y la Ecuación 4-15. Todos los valores del coeficiente empírico  $a$  son negativos; independientemente, si la CWS se ubica tanto por encima como por debajo de 2.500 msnm, en este caso el valor máximo está en Obonuco, el mínimo está en Bombona, con una desviación estándar de 0,1009. En el caso del coeficiente empírico  $b$ , todos los valores son positivos, el valor máximo está en Común el automático y el valor mínimo está en Monopamba con una desviación estándar de 0,0133.

Para determinar el desempeño del modelo logístico, su usaron los datos de cuatro AWS, dos de la zona del Pacífico y dos de la zona andina, que tuvieron registros de un año en promedio y que no se usaron inicialmente para la determinación de la relación lineal. Los datos registrados pasaron un proceso de control de calidad; posteriormente, surgió una comparación entre los resultados de los modelos empíricos y los datos *in situ* para calcular los errores estadísticos presentados en la Tabla 4-6. Los resultados en negrita son los mejores.

Para Altaquer y Granja el Mira ubicados en la zona del Pacífico, el modelo logístico tuvo un mejor desempeño en RMSE, SD, MAE, MBE y  $U_{95}$  que el modelo de Hargreaves y Samani, que presentó mejores resultados en MPE y MAPE. En el MBE, el modelo de Hargreaves y Samani subestimó el recurso en 442,78 [Wh/m<sup>2</sup>da] para Granja el Mira y 333,11 [Wh/m<sup>2</sup>da] para Altaquer, mientras que el modelo

**Tabla 4-5.** Coeficientes empíricos estimados para la zona Andina y Amazónica

CWS	Altitud	a	b
Obonuco	2.710	-1,245394	0,06885508
Apto. Antonio Nariño	1.796	-1,424296	0,08828996
San Bernardo	2.190	-1,283602	0,07029916
Común el automática	3.141	-1,457943	0,10843070
Apto. San Luis	2.961	-1,369175	0,09190260
Bombona	1.493	-1,532495	0,10212553
Tanama	1.500	-1,529995	0,10180589
Sindagua	2.800	-1,289778	0,07711913
Monopamba	2.719	-1,249832	0,06968149
El Encano	2.830	-1,304572	0,07987382
Chimayoy	2.745	-1,262654	0,07206888
Taminango	1.875	-1,396086	0,08468267

logístico lo sobreestima en 145,14 [ $Wh/m^2da$ ] y 76,19 [ $Wh/m^2da$ ] en Altaquer y Granja el Mira, respectivamente. Si bien el modelo logístico no tuvo el mejor comportamiento en todas las mediciones de error estadístico, es adecuado estimar el recurso solar con datos de temperatura del aire en lugares por debajo de los 2.500 msnm.

Para Ospina Perez y Sandona ubicados en la zona andina, el modelo logístico mostró menores errores en todas las mediciones estadísticas. En cuanto a los resultados de MBE, los dos modelos empíricos sobreestimaron el recurso solar. Sin embargo, el modelo logístico redujo la sobreestimación en 181,09 [ $Wh/m^2da$ ] (51,12 %) y 225,15 [ $Wh/m^2da$ ] (18,55 %) para Ospina Perez y Sandona, respectivamente. Finalmente, los autores concluyeron que el modelo logístico presentó mejores resultados que el modelo de Hargreaves y Samani en altitudes superiores a 2.500 msnm.

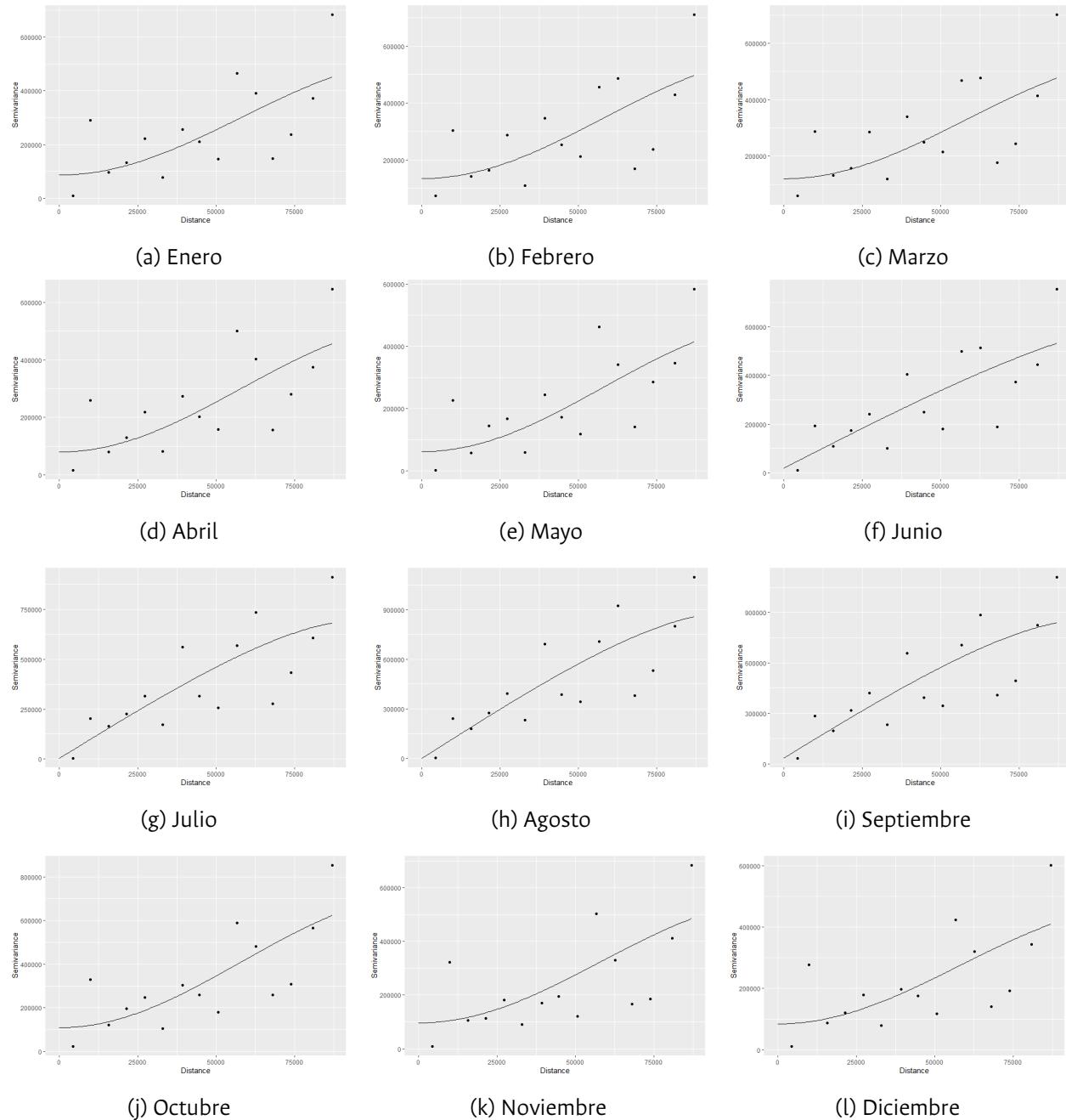
**Tabla 4-6.** Comparación de los modelos empíricos

AWS	N	RMSE	SD	MAE	MBE	MPE	MAPE	$U_{95}$
Altaquer (HS)	390	761,09	24,49	596,88	-333,11	<b>-4,23 %</b>	<b>22,30 %</b>	1.492,51
Altaquer - proposed	390	<b>670,59</b>	<b>23,43</b>	<b>529,73</b>	<b>145,14</b>	14,14 %	24,24 %	<b>1.315,17</b>
Granja el Mira (HS)	305	854,87	<b>24,81</b>	684,98	-442,78	<b>3,47 %</b>	<b>33,75 %</b>	1.676,26
Granja el Mira - proposed	305	<b>781,91</b>	25,99	<b>596,66</b>	<b>76,19</b>	20,57 %	34,81 %	<b>1.533,40</b>
Ospina Perez (HS)	231	987,70	27,44	797,83	355,08	16,80 %	27,29 %	1.936,64
Ospina Perez - proposed	231	<b>934,93</b>	<b>27,35</b>	<b>736,86</b>	<b>173,99</b>	<b>11,09 %</b>	<b>24,43 %</b>	<b>1.833,25</b>
Sandona (H&S)	221	1.375,84	22,83	1.249,26	1.213,27	49,93 %	50,70 %	2.697,03
Sandona - proposed	221	<b>1.179,92</b>	<b>22,69</b>	<b>1.034,94</b>	<b>988,12</b>	<b>41,30 %</b>	<b>42,32 %</b>	<b>2.313,07</b>

El caso base de las técnicas de interpolación espacial utilizó el IDW con  $p = 2$  como exponente, porque IDW es el método IDW más utilizado. Implementar los métodos de Kriging implica determinar los variogramas teóricos representados por la Figura 4-2 para cada mes. Los autores implementaron el mejor

modelo de variograma disponible en el paquete gstat para cada mes. Wave fue la opción seleccionada del paquete gstat para modelar de enero a junio y de octubre a diciembre, y Circular fue la opción para modelar de mayo a septiembre.

**Figura 4-2.** Variograma experimental y teórico



El variograma experimental de mayo y diciembre muestra la semivarianza más baja; sus valores fueron 352.019 y 322.911, respectivamente, ver Figura 4-2e y Figura 4-2l. Por lo tanto, se espera los errores estadísticos más bajos para esos meses. En agosto y septiembre, la diferencia de semivarianza tuvo los

valores más altos; sus valores fueron 847.928 y 801.113, respectivamente, ver Figura 4-2h y Figura 4-2i. En consecuencia, estos meses podrían mostrar los errores estadísticos más altos.

La Tabla 4-7 muestra los errores estadísticos derivados del uso de los métodos IDW y Kriging. La validación cruzada de dejar uno fuera fue la técnica implementada para calcular los errores. Respecto a RMSE, hubo una mejora del 11,71 %, 11,94 % y 2,07 % para OK, SK y UK, respectivamente, con respecto a los resultados de IDW. La Figura 4-4 confirma los resultados de RMSE para cada técnica de interpolación. Analizando los resultados promedio, SK y OK establecen el resultado más prometedor. Además, los resultados de RMSE muestran que todas las técnicas de interpolación tuvieron un rendimiento similar durante todo el año, consulte la Figura 4-4. Durante septiembre, el RMSE alcanzó el valor más alto 636,785 Wh/(m<sup>2</sup>da) con IDW, debido a que los valores atípicos influyen en los resultados, como lo mostró el valor de semivarianza. Mayo presenta el valor RMSE más bajo, 410,781 Wh/(m<sup>2</sup>da) usando SK. Hay una reducción del error del 14,63 % y del 14,62 % con OK y SK frente a los resultados de IDW en agosto.

En cuanto a los resultados de MAE, las técnicas OK y SK mejoraron en un 3,98 % y 5,10 %, respectivamente, y la técnica de UK redujo el error un 6,78 %, como presenta la Figura 4-5. Septiembre muestra el valor más alto con la técnica UK 455,055 Wh/(m<sup>2</sup>da), y mayo tuvo el valor más bajo 296,33 Wh/(m<sup>2</sup>da). SK y OK muestran resultados similares durante todo el año.

Con respecto a MBE, todos los métodos de Kriging tuvieron mejores resultados que IDW en 1,01, 0,063 y 0,022 veces con OK, SK y UK, respectivamente, ver Figura 4-6. Los resultados de MBE muestran que OK es el método con menos sesgo; Además, ésta es también la única técnica con subestimación. Esta información es crucial en el diseño de la planta de energía porque el recurso determina el tamaño de la planta para generar la electricidad necesaria para cubrir la demanda.

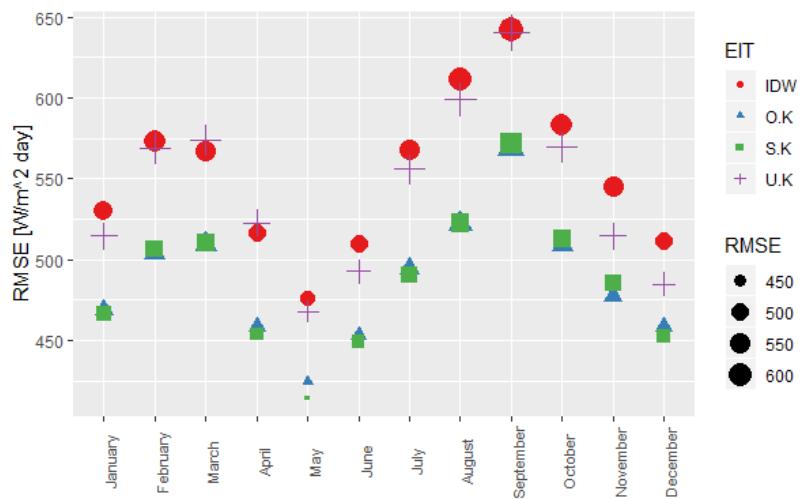
Los resultados de SD muestran que los métodos de Kriging reducen la dispersión en un 11,40 %, 11,97 % y 2,10 % con OK, SK y UK, respectivamente, ver Figura 4-7. La SD tuvo resultados similares a RMSE porque una alta dispersión podría conducir a valores atípicos. De julio a septiembre, hay más dispersión. Por tanto, durante estos meses, el clima en el territorio es más heterogéneo que en el resto de meses. Finalmente, los resultados de MPE mostraron una mejora en OK, SK y UK de un 46,58 %, 16,81 % y 20,98 %, respectivamente, ver Figura 4-8. Los resultados de MPE muestran que OK tuvo mejores resultados durante todo el año.

**Tabla 4-7.** Spatial interpolation methods results

Month	RMSE	MAE	MBE	MPE	SD	EIT	RMSE	MAE	MBE	MPE	SD	EIT
January	525,342	330,598	45,869	4,35	16,729	IDW	465,330	342.332	-2.860	2.05	14.875	O.K
February	567,242	397,357	41,182	4,50	17,712	IDW	500,239	366.794	0.906	2.44	15.661	O.K
March	561,394	400,581	41,040	4,39	17,307	IDW	505,381	376.748	0.199	2.40	15.622	O.K
April	511,992	333,541	44,521	4,07	15,819	IDW	455,076	341.022	-4.360	1.82	14.113	O.K
May	471,550	306,39	42,209	3,97	15,319	IDW	421,250	315.984	-9.763	1.52	13.736	O.K

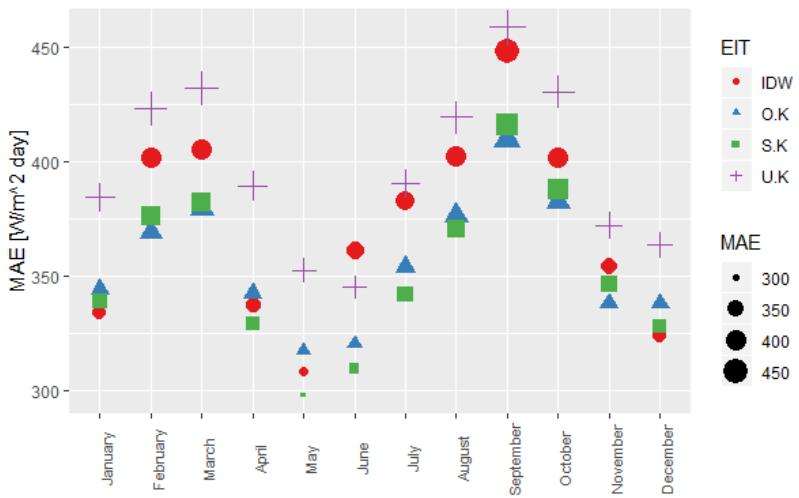
June	506,003	360,751	33,047	4,65	16,952	IDW	450,789	319.848	1.264	2.87	15.134	O.K
July	564,144	381,971	27,776	5,29	18,486	IDW	491,535	352.424	1.542	3.48	16.127	O.K
August	607,336	399,639	34,836	5,78	18,753	IDW	518,485	374.029	2.713	3.65	16.036	O.K
September	636,785	443,480	37,219	5,64	19,045	IDW	564,540	407.237	4.903	3.71	16.913	O.K
October	576,885	396,533	62,608	5,31	17,496	IDW	504,945	380.054	-0.427	2.33	15.405	O.K
November	539,667	352,341	70,571	5,25	17,276	IDW	472,927	336.586	4.637	2.35	15.270	O.K
December	506,778	321,878	40,389	4,10	16,692	IDW	454,732	336.019	-3.541	1.99	15.025	O.K
January	462,785	336.920	40.141	3.55	14.738	S.K	509,716	381.682	34.512	3.26	16.257	U.K
February	502,179	373.044	46.191	4,00	15.655	S.K	563,363	419.714	41.599	3.82	17.589	U.K
March	505,895	378.639	44.025	3.87	15.578	S.K	568,021	427.894	40.631	3.75	17.513	U.K
April	450,275	327.556	34.090	3.10	13.925	S.K	517,755	386.887	42.070	3.29	16.005	U.K
May	410,781	296.330	29.787	2.89	13.363	S.K	462,906	349.814	35.557	2.97	15.054	U.K
June	446,634	309.531	34.535	4.06	14.950	S.K	487,127	343.213	47.358	3.98	16.277	U.K
July	487,870	340.342	28.128	4.41	15.980	S.K	550,103	387.010	54.272	4.80	17.960	U.K
August	518,563	368.247	41.566	4.97	15.987	S.K	592,957	416.214	59.400	5.01	18.247	U.K
September	567,399	413.525	48.972	5.19	16.936	S.K	633,792	455.055	56.646	4.86	18.912	U.K
October	508,207	384.880	52.490	4.16	15.422	S.K	564,131	427.383	41.536	3.63	17.164	U.K
November	480,752	344.644	57.366	4.26	15.412	S.K	508,971	369.595	28.939	3.12	16.408	U.K
December	448,857	325.603	31.080	3.21	14.796	S.K	479,935	360.711	23.534	2.79	15.839	U.K

Figura 4-4. RMSE

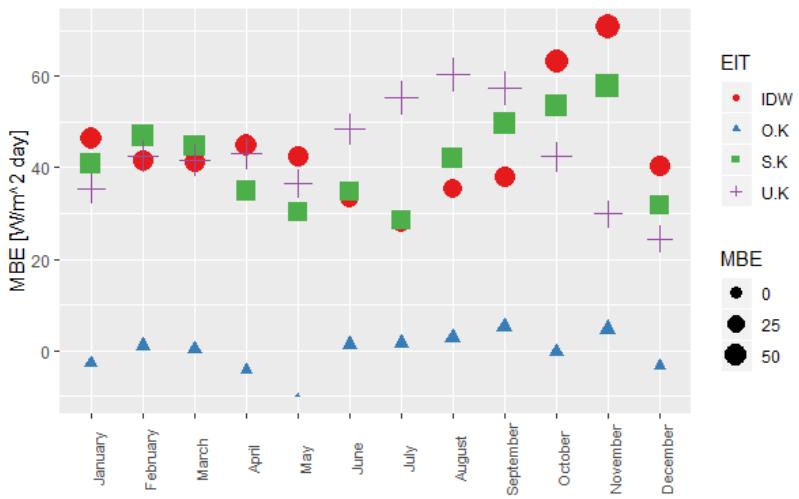


En general, los métodos de Kriging mostraron mejores resultados que el de IDW. Con los métodos de Kriging, SK tuvo mejores resultados en RMSE (11,94 %), MAE (5,10 %) y SD (11,97 %), y OK tuvo mejores resultados en MBE (1,01 veces) y MPE (46,58 %). Con base en los resultados, OK fue el método seleccionado para mapear la insolación solar global en las áreas tropicales y montañosas analizadas, debido a que es la técnica con menos sesgo y menor porcentaje de error. Desde la Figura 4-9 hasta la Figura 4-20

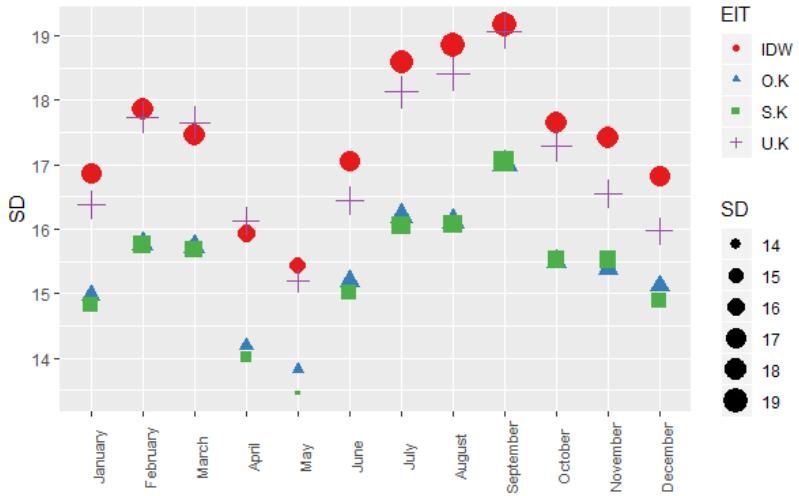
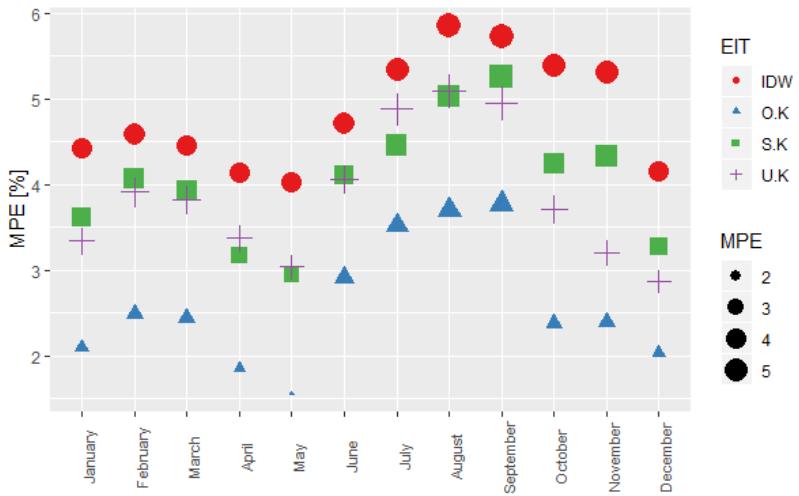
**Figura 4-5. MAE**



**Figura 4-6. MBE**



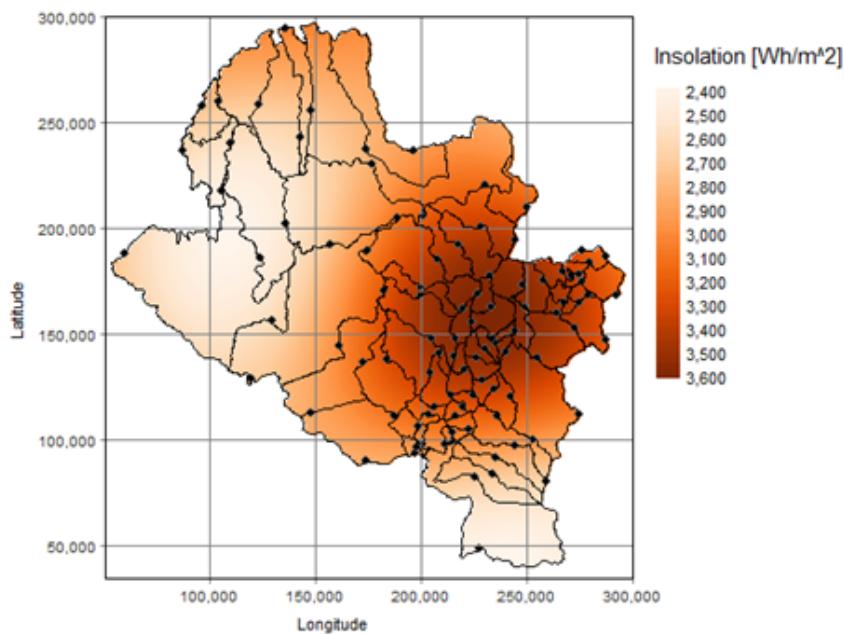
muestran la insolación solar global promedio diaria mensual desde enero hasta diciembre. Al comparar los resultados obtenidos en esta investigación con el Atlas Solar del IDEAM, los investigadores coinciden en que la zona andina tiene un potencial significativo durante todo el año. Sin embargo, existe una diferencia entre el potencial estimado en estas dos investigaciones. En general, los resultados de IDEAM presentan un potencial más alto que el potencial estimado en esta investigación. Por ejemplo, en Obonuco el potencial estimado por el IDEAM es de 3.677,5 [Wh/m<sup>2</sup>da], y esta investigación estima el potencial en 3.298,6 [Wh/m<sup>2</sup>da] en enero. En Apto. San Luis, el IDEAM estiman un potencial de 3.902,6 [Wh/m<sup>2</sup>das], y esta investigación estima el potencial en 2.194,1 [Wh/m<sup>2</sup>das] en marzo. Además, el IDEAM usó dos CWS (Obonuco y Apto. San Luis) y cinco AWS (La Josefina, Botana, Cerro Páramo, Viento Libre y Paraiso) para estimar el potencial en Nariño, esta investigación aumenta la can-

**Figura 4-7. SD****Figura 4-8. MPE**

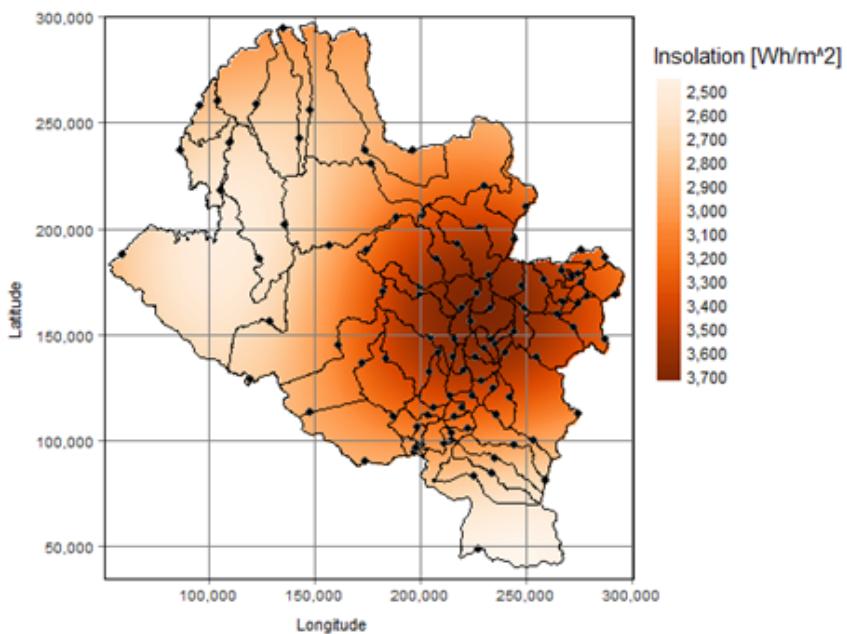
tidad de AWS y CWS utilizados. Por tanto, la densidad, que es un factor importante en las técnicas de interpolación, aumenta con esta investigación.

De los mapas se puede concluir que una planta de energía solar típica con una capacidad de 1 kW, evaluada en el peor y mejor caso, ubicada en la zona andina podría generar entre 73,15 kWh y 81,24 kWh en diciembre y octubre, respectivamente. La misma central ubicada en la zona Pacífico podría generar entre 52,44 kWh y 61,13 kWh en noviembre y marzo, respectivamente, y en la zona amazónica, la generación de electricidad está entre 51,31 kWh y 62,42 kWh en junio y octubre, respectivamente.

**Figura 4-9.** Enero

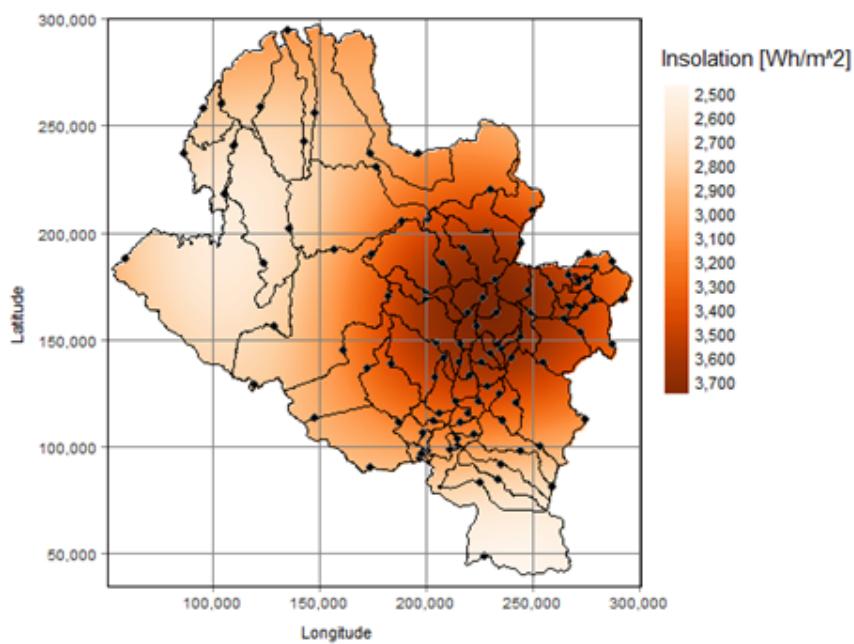
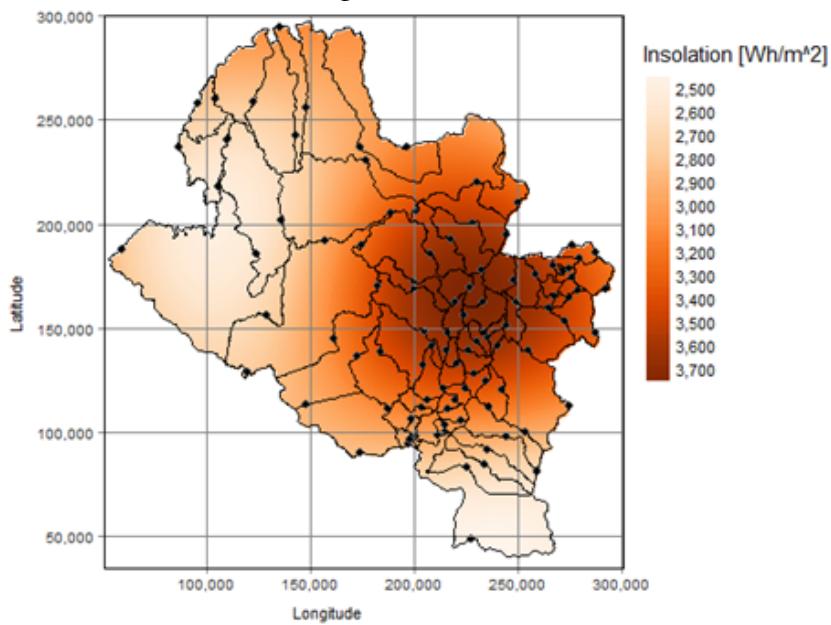


**Figura 4-10.** Febrero



## 4.6 Conclusiones

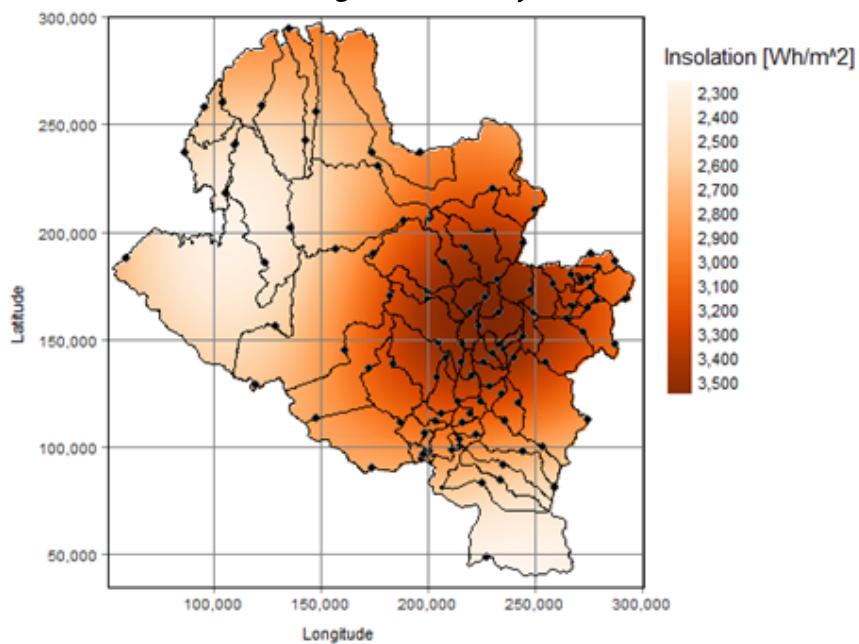
La estimación de la insolación solar permitió confirmar que la relación entre la altitud AWS y los coeficientes empíricos del modelo logístico tuvo un mejor desempeño en la zona andina que en la del

**Figura 4-11.** Marzo**Figura 4-12.** Abril

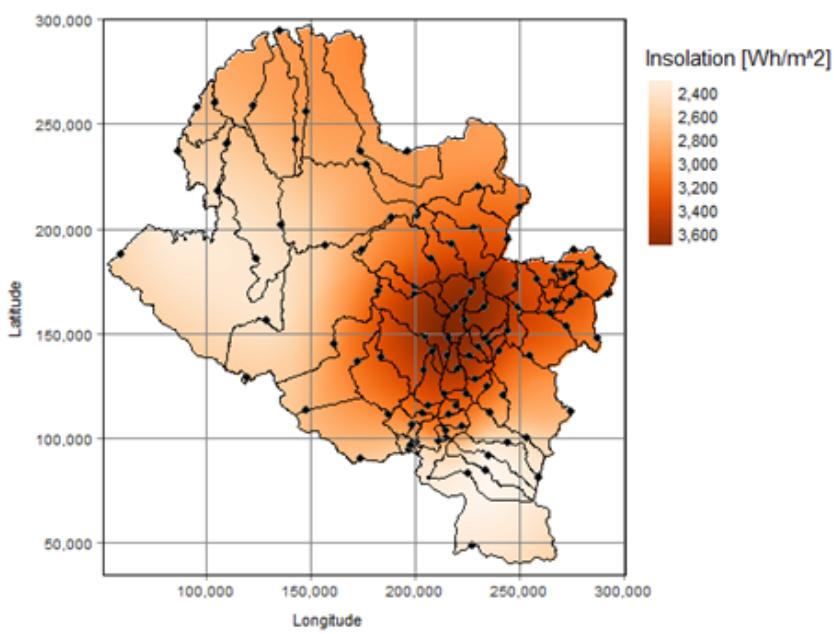
Pacífico. El modelo de Hargreaves y Samani tuvo mejores resultados en la zona del Pacífico.

Los autores estimaron la insolación solar con el modelo Hargreaves y Samani en la zona del Pacífico, y los resultados presentaron un recurso subestimado por 442,78 [Wh/m<sup>2</sup>da] y 333,11 [Wh/m<sup>2</sup>da] para Granja el Mira y Altaquer, respectivamente. Es fundamental aumentar el número de AWS en el territorio

**Figura 4-13.** Mayo

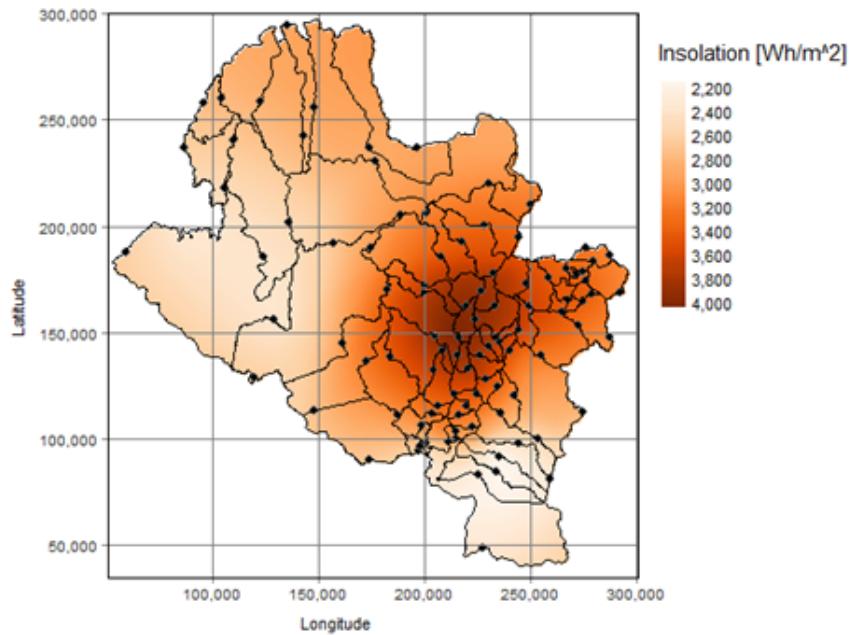
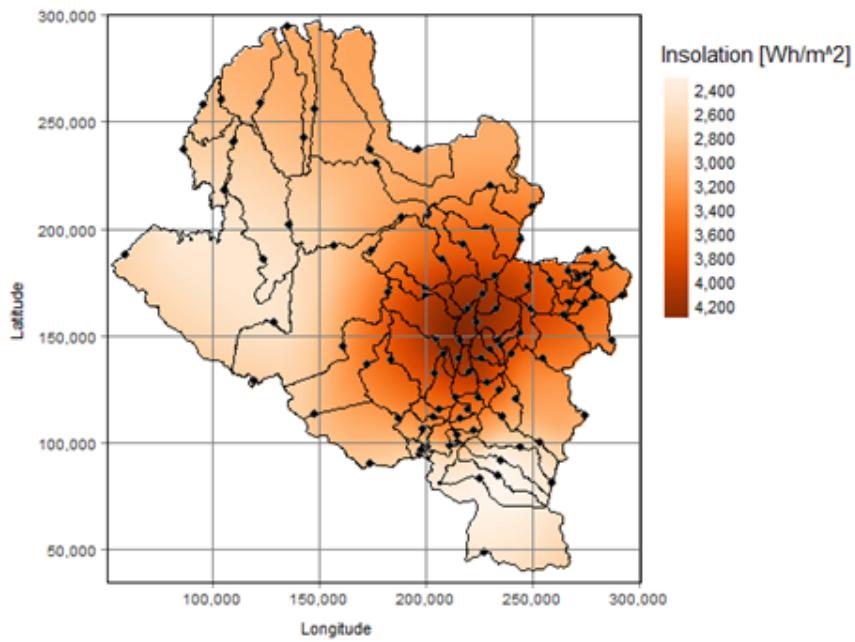


**Figura 4-14.** Junio



para mejorar los resultados.

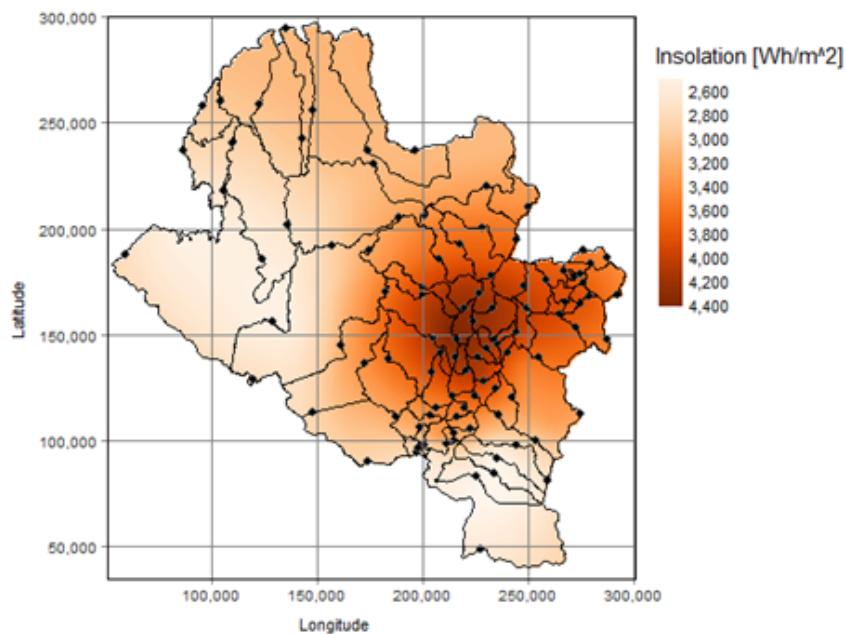
Los autores estimaron la insolación solar en la zona andina con el modelo logístico propuesto en la Parte I de este trabajo. La comparación entre los valores medidos y estimados mostró una sobreestimación en Sandona y Ospina Perez. Sin embargo, con respecto al modelo HS, la sobreestimación se reduce 51,81

**Figura 4-15. Julio****Figura 4-16. Agosto**

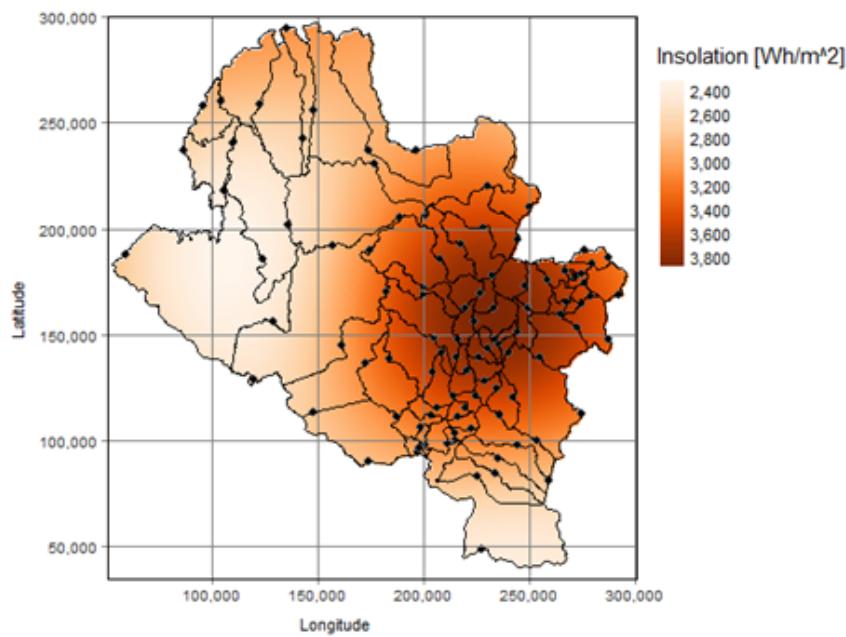
% y 18,55 % para Ospina Perez y Sandona, respectivamente. En términos globales, el modelo logístico tuvo un mejor desempeño que el modelo HS en altitudes superiores a los 2.500 msnm.

Al comparar los métodos IDW y Kriging, OK presentó una mejora en los errores estadísticos. Las siguientes cifras confirman esa conclusión: RMSE 11,71 %, MAE 3,98 %, MBE 1,01 veces, SD 11,40 % y MPE

**Figura 4-17.** Septiembre

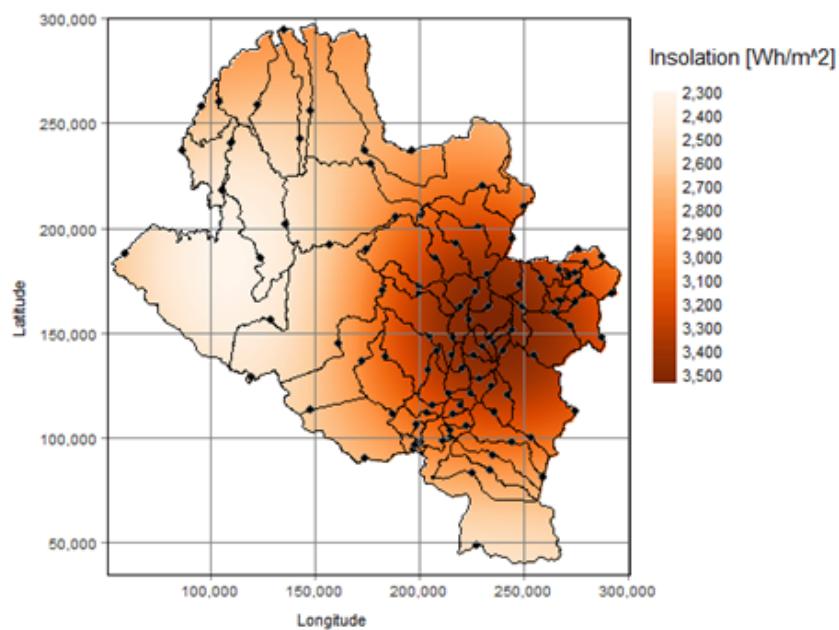
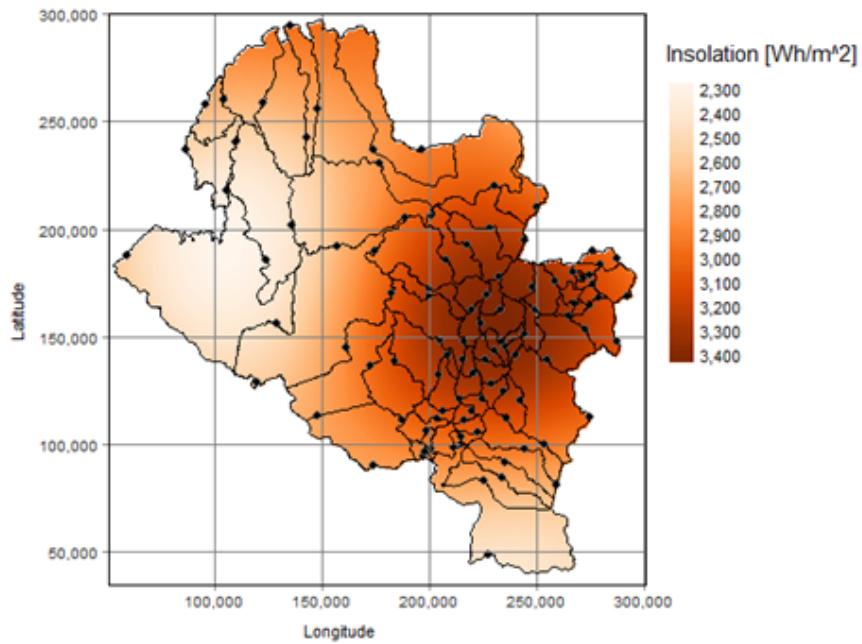


**Figura 4-18.** Octubre



46,58 %. Aunque SK mostró mejores resultados en MAE, SD y RMSE que OK, el sesgo con OK es menor. Por lo tanto, OK fue la técnica seleccionada para mapear el potencial de insolación solar. UK fue la técnica que presentó menores mejoras frente al IDW, incluso en MAE, que redujo el 6,78 % de la precisión.

De los mapas se puede concluir que la estimación obtenida es menor a la presentada en los mapas de

**Figura 4-19.** Noviembre**Figura 4-20.** Diciembre

radiación solar del IDEAM en todo el territorio de Nariño. Sin embargo, también es posible determinar la existencia de una zona con un alto potencial de generación de electricidad con plantas de energía solar fotovoltaica.

# 5 Pronóstico de corto plazo de irradiancia solar global en ambientes tropicales<sup>1</sup>

## Resumen

La previsión de datos de irradiancia solar es un tema esencial para impulsar las aplicaciones de energía solar. Existen varias técnicas para pronosticar la irradiancia solar global, como la predicción numérica del clima y las técnicas estadísticas. En este contexto, el objetivo de este artículo es implementar cuatro técnicas de pronóstico ARIMA, SLP, MLP y LSTM, en un horizonte de día anticipado utilizando conjuntos de datos incompletos medidos en un ambiente tropical y montañoso. La técnica de imputación implementada llenó en promedio xxxx datos de missinda con un error en MAE de  $81,07\text{ Wh/m}^2$ , RMSE de  $118,15\text{ Wh/m}^2$  y MBE de  $-12,83\text{ Wh/m}^2$ . Además, los resultados muestran que los modelos basados en redes neuronales superan al modelo ARIMA. Además, LSTM tiene un mejor rendimiento con un número reducido de datos de entrada y en entornos nublados.

## 5.1 Introducción

La proliferación de sistemas de generación de energía solar ha impulsado el interés por pronosticar la irradiancia solar. Los sistemas precisos de predicción de la irradiancia solar permiten estimar la generación de electricidad a largo, medio y corto plazo. Esta información es crucial para mantener el equilibrio entre la oferta y la demanda de energía (Dannecker, 2015) y minimizar los costos asociados con el inicio y cierre de las centrales eléctricas convencionales (Badosa et al., 2017).

El análisis de series de tiempo es el proceso de examinar los datos registrados a lo largo del tiempo para desarrollar un modelo matemático (Shumway & Stoffer, 2011). Estos modelos facilitan la implementación de sistemas fotovoltaicos (FV), tanto dentro como fuera de la red. Las series de tiempo registradas en períodos prolongados se utilizan para comprender el comportamiento y predecir los valores futuros de irradiancia solar en una ubicación específica (Suehrcke, 2000). Además, el estudio de la irradiancia solar se convierte en un tema importante debido a las leyes locales o globales que buscan

---

<sup>1</sup>Este artículo tuvo la participación de Belizza Ruiz y Francisco Ruiz

introducir este tipo de energía en la matriz energética para reducir las emisiones de gases de efecto invernadero (GEI).

La irradiancia solar es un fenómeno dependiente del tiempo compuesto por una parte determinista y estocástica (Boland, 2008). Los modelos matemáticos permiten predecir el valor futuro exacto de la parte determinista. Por otro lado, la parte estocástica genera un valor futuro entre dos límites con un intervalo de confianza (Box et al., 2016). La precisión del pronóstico se basa en la capacidad del componente estocástico para modelar los cambios de irradiancia solar inducidos por las nubes (Inman et al., 2013). Hay varios enfoques para pronosticar la irradiancia solar: los métodos de persistencia son un modelo de pronóstico simple donde el valor en el tiempo  $t + 1$  es igual al valor en el tiempo  $t$  (Diagne et al., 2013), modelos autorregresivos como el autorregresivo de promedio móvil (ARMA), autorregresiva integrado de promedio móvil (ARIMA), que permite modelar variaciones estacionarias y no estacionarias y describir fenómenos atmosféricos no lineales complejos (Inman et al., 2013)-y técnicas de computación blanda, como máquina de vectores de soporte (SVM), red neuronal artificial (ANN) y algoritmos difusos y genéticos (GA) (Demirhan & Renwick, 2018). La ANN, la lógica difusa y los híbridos son robustos para modelar la naturaleza estocástica de los procesos físicos, como la irradiancia solar, debido a su capacidad para compensar errores sistemáticos y desviaciones aprendibles problemáticas (Paulescu et al., 2013).

La selección de un método de pronóstico depende de la escala de tiempo deseada, es decir, intra-horario, intra-diario, mediano y largo plazo. Los enfoques estadísticos generalmente funcionan bien para pronósticos a corto plazo, como ARIMA y ANN. Para el análisis a largo plazo, se prefieren con frecuencia las técnicas de computación blanda (Coimbra et al., 2013; Demirhan & Renwick, 2018). Los modelos autorregresivos describen las características y el comportamiento de la serie temporal mediante un proceso de autorregresión y representan el 14 % de los modelos utilizados para el pronóstico de la energía solar (Antonanzas et al., 2016; Dannecker, 2015). ARIMA es una extensión del modelo ARMA, que permite modelar series de tiempo no estacionarias. La solidez a errores aleatorios y valores atípicos fomenta el uso de estos modelos de pronóstico (Diagne et al., 2013; Sobri et al., 2018). La ANN es un modelo estadístico que establece una relación entre los datos de entrada y salida durante un proceso de entrenamiento a través de capas formadas por nodos interconectados de entradas, salidas, capas ocultas y funciones de activación. Esta se ha convertido en la técnica de pronóstico de energía solar más popular (Mazorra-Aguiar & Díaz, 2018; Antonanzas et al., 2016). Long Short-Term Memory (LSTM) es una red neuronal recurrente (RNN) avanzada, que se ha utilizado recientemente en el campo de las energías renovables (Chandola et al., 2020) .LSTM aprende la dependencia entre datos sucesivos (Ghimire, Deo, Downs, & Raj, 2019). Algunos estudios sugieren que LSTM supera a otros modelos de vanguardia en el pronóstico de la irradiancia solar diaria (Husein & Chung, 2019).

La implementación de estas técnicas de pronóstico requiere conjuntos de datos históricos de radiación solar. En esta investigación, los autores utilizan los datos recopilados por el Instituto Colombiano de Hidrología, Meteorología y Estudios Ambientales (IDEAM) en 12 Estaciones Meteorológicas Automáticas (AWS) ubicadas en Nariño - Colombia. Esta región está situada en el trópico entre las latitudes

0°21'54" y 2°41'10" y longitudes –79°0'43" a –76°50'13", con altitudes que varían entre los 16 a 3.577 msnm. El área de estudio tiene tres subáreas, denominadas zonas del Pacífico, Andina y Amazonia. Como es habitual, estos conjuntos de datos tienen una cantidad considerable de datos faltantes debido a la pérdida de conexión entre los equipos de medición y almacenamiento, falta de calibración de las estaciones que la dejan fuera de operación, entre otras. Para superar este problema, Layanun et al. propuso una técnica de imputación de datos para un método de pronóstico estacional basado en ARIMA en Bangkok, Tailandia, que promedia los datos clasificados por tipo de clima determinados con información de temperatura y humedad relativa (Layanun et al., 2017). Asimismo, Rodríguez-Rivero et al. aplicó una técnica de suavizado de promedios para llenar los datos faltantes de la serie de tiempo para pronóstico a corto plazo (Rodríguez-Rivero et al., 2017). Ogunsola & Song utilizó tres enfoques para la imputación de datos de irradiancia solar, análisis espectral singular (SSA), irradiancia solar ajustada estadísticamente (SASR) y enfoque basado en la temperatura (TBA). Además, en este estudio se recomendó utilizar diferentes técnicas de imputación basadas en la longitud de los datos faltantes (Ogunsola & Song, 2014). Nuestros experimentos implementaron una técnica de imputación que reemplaza los datos faltantes con un valor estimado a partir de los datos existentes en un rango horario (Demirhan & Renwick, 2018; Moritz & Bartz-Beielstein, 2017). Este estudio aplica cuatro modelos de predicción de vanguardia para el pronóstico de la irradiancia solar global en entornos tropicales y montañosos con datos incompletos: i) ARIMA, ii) Perceptrón de capa única (SLP), iii) Perceptrón de múltiples capas (MLP), iv) LSTM. La imputación de datos faltantes combina una interpolación lineal de los valores subsiguientes con el promedio de valores pasados medidos a la misma hora de los datos imputados. Para eliminar el efecto determinista, se utiliza el índice de claridad en lugar de los valores de irradiancia solar global directamente. Los resultados muestran que las técnicas basadas en redes neuronales tienen un mejor rendimiento que ARIMA. El modelo LSTM tiene un mejor rendimiento en AWS con menos cantidad de datos en la zona del Pacífico. SLP logra el mejor comportamiento en AWS con muchos más datos de entrada.

Este artículo tiene cinco secciones, la Sección 5.2 presenta las técnicas de pronóstico. La sección 5.3 contiene la descripción de los datos utilizados. La sección 5.4 presenta los resultados obtenidos. Finalmente, la Sección 5.5 muestra las conclusiones.

## 5.2 Materiales y métodos

En esta sección, los autores describen los cuatro modelos de predicción utilizados para pronosticar la irradiancia solar global: el método estadístico autorregresivo ARIMA y tres modelos basados en arquitecturas de redes neuronales: SLP, MLP y LSTM.

### 5.2.1 ARIMA

Este modelo se divide en tres componentes: i) autorregresión (AR), ii) integración (I) y iii) media móvil (MA). AR y MA se ocupan de los elementos estocásticos, y I es el responsable de transformar la serie de tiempo en una estacionaria. ARIMA se denota como ARIMA  $(p, d, q)$ ,  $p$  es el orden de AR,  $q$  es el orden de MA y  $d$  derivadas aplicadas a la serie de tiempo, tal que  $d = 0$  significa que la serie de tiempo ya es estacionaria, y  $d = 2$  significa que las dos derivadas son necesarias para que los datos sean estacionarios (Agami Reddy, 2011). Por tanto, ARIMA  $(p, d, q)$  es:

$$\Phi_p(B) \Delta^d x_t = \theta_q(B) u_t, \quad u_t \sim WN(0, \sigma^2) \quad (5-1)$$

o

$$x_t + \Phi_1 x_{t-1} + \cdots + \Phi_p x_{t-p} = u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q} \quad (5-2)$$

$$\Phi(B) x_t = \left(1 - \sum_{i=1}^p \Phi_i B^i\right) (1 - B)^d x_t = \left(1 + \sum_{i=1}^q \theta_i B^q\right) u_t \quad (5-3)$$

Donde  $B$  es el operador de desplazamiento hacia atrás,  $\Delta = 1 - B$  es la diferencia hacia atrás,  $\Phi_p$  y  $\theta_p$  son polinomios de orden  $p$  y  $q$  respectivamente.  $(1 - B)^d x_t$  es el encargado de transformar la serie temporal no estacionaria en una estacionaria. Después de aplicar esta transformación, es posible utilizar cualquier estrategia de pronóstico para datos estacionarios (Montgomery et al., 2008).

AR describe el comportamiento pasado de la serie temporal y la serie residual en el tiempo real como una combinación lineal ponderada de valores de un conjunto de datos de un proceso estocástico  $x_t$  (Dannecker, 2015) y un ruido blanco  $u_t$  como sigue

$$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_p x_{t-p} + u_t = \sum_{i=1}^p \Phi_i x_{t-i} + u_t \quad (5-4)$$

Usando el operador de retroceso  $B^j x_t$  la Ecuación 5-4 se expresa de la siguiente manera:

$$u_t = \left(1 - \sum_{i=1}^p \Phi_i B^i\right) x_t \quad (5-5)$$

MA describe las perturbaciones aleatorias de la serie temporal mediante una combinación lineal ponderada de valores previos de un error de ruido blanco. Por lo tanto, la serie de tiempo se representa como un conjunto de variables aleatorias no correlacionadas y con distribución normal, de la siguiente manera:

$$x_t = \theta_1 x_{t-1} + \theta_2 x_{t-2} + \cdots + \theta_q x_{t-q} + u_t = \left(1 - \sum_{i=1}^q \theta_i B^q\right) u_t \quad (5-6)$$

### 5.2.2 ANN

Una ANN es un conjunto de unidades conocidas como neuronas, conectadas por articulaciones sinápticas, cada una con un coeficiente de peso (Blaga et al., 2019). Una ANN tiene tres capas, entrada, oculta y salida. La primera recibe los datos de entrada. La capa oculta conecta las capas de entrada y salida utilizando las neuronas. La capa de salida genera la información calculada (Premalatha & Valan Arasu, 2016).

El entrenamiento de la ANN requiere un procedimiento de retropropagación iterativo que aprende un mapeo de entrada-salida. Este proceso tiene cuatro pasos: i) propagación hacia adelante de la entrada del patrón de entrenamiento, ii) cálculo del error mediante una función de pérdida que compara los valores estimados y de referencia, iii) propagación hacia atrás del error para volver a calcular cada peso  $\Delta w_{ij}$  desde la salida a la capa oculta, y iv) el peso  $w_{ij}$  se actualiza de la siguiente manera  $w_{ij}^{nuevo} = w_{ij}^{antiguo} + \lambda \Delta w_{ij}$ , donde  $\lambda$  se conoce como tasa de aprendizaje (Blaga et al., 2019).

Durante la propagación hacia adelante, las entradas se multiplican inicialmente por los pesos y se suman los resultados individuales. Posteriormente, se agrega un sesgo a los resultados como un valor de compensación que regula la señal. El tercer paso es la activación de la función de transferencia para modular las sumas (las funciones logística y gaussiana son funciones de transferencia comunes). De esta forma, la señal pasa por todas las capas hasta llegar a la salida. Los pesos se inicializan con valores aleatorios. Tanto las ponderaciones como el sesgo se actualizan iterativamente hasta que se cumple un criterio de detención (Munawar & Wang, 2020). Un perceptrón multicapa (MLP) es una ANN con múltiples capas ocultas. Un MLP con  $N_I + 1$  capas se expresa de la siguiente manera

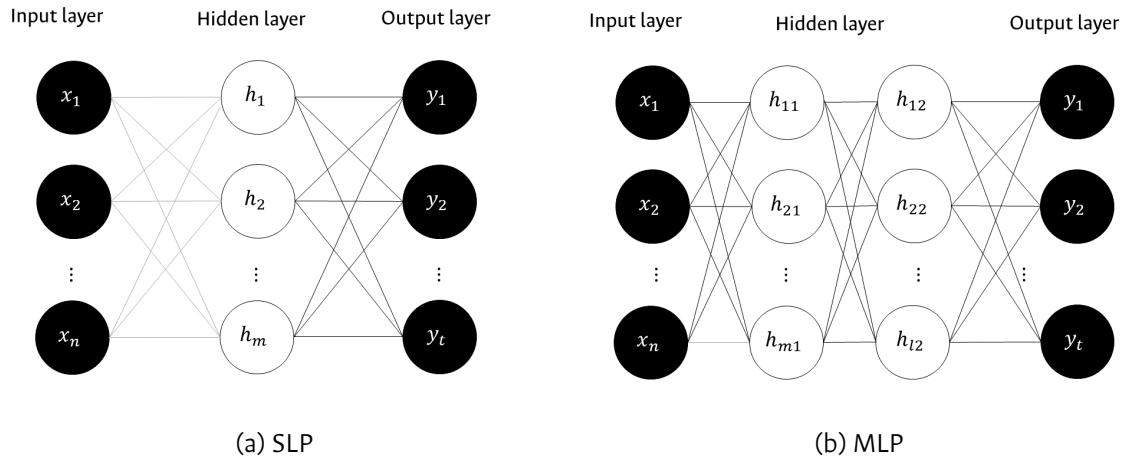
$$\mathbf{x}^{(i)} = f(\mathbf{z}^{(i)}) \quad (5-7)$$

$$z_n^{(i)} = \sum_{k=1}^{N_{i-1}} w_{nk}^{(i)} x_k^{i-1} + b_n^{(i)} \quad (5-8)$$

donde  $\mathbf{x}^{(0)}$  es la entrada (serie de tiempo del día actual),  $\mathbf{x}^{(N_I)}$  es la irradiancia solar prevista,  $\mathbf{x}^{(i)}$  por  $0 < i < N_I$  son capas ocultas,  $w_{nk}^{(i)}$  es el peso que conecta la neurona  $k$ -ésima en la capa anterior con el  $n$ -ésimo en la capa actual,  $b_n^{(i)}$  es un sesgo neuronal, y  $f(\cdot)$  es una función de activación (Ghimire, Deo, Raj, & Mi, 2019). La Figura 5-1 ilustra ambos tipos de arquitecturas, SLP y MLP.

### 5.2.3 LSTM

RNN es una herramienta eficiente para resolver problemas con patrones complejos debido a la capacidad de recordar información importante de la entrada que influye en el resultado final (Chandola et al., 2020). Sin embargo, entrenar a una RNN es difícil debido a que el gradiente de retropropagación crece o se reduce en cada paso de tiempo (Husein & Chung, 2019). Dada una entrada  $x_t = x_1, x_2, x_n$ , la salida  $y_t$  en un RNN viene dada por:

**Figura 5-1.** Estructura de una ANN

$$y_t = w_{hy}h_t + b_y \quad (5-9)$$

$$h_t = H(w_{hx}x_t + w_{hh}h_{t-1} + b_h)$$

donde  $w_{hx}$ ,  $w_{hh}$ ,  $w_{hy}$  son matrices de ponderaciones de entrada-oculta, oculta-oculta y oculta-salida,  $b_h$  y  $b_y$  son vectores de desviación de salida y ocultos, respectivamente.  $H$  es la función de activación de capa oculta.

LSTM es una RNN avanzado que resuelve el problema del gradiente al incluir una memoria explícita para la red. LSTM tiene una puerta de entrada, olvido y salida, y una unidad de celda que sirve como memoria durante un intervalo de tiempo definido (ver Figura 5-2a). Las puertas controlan el flujo de información que entra y sale de la unidad de celda, vea la Figura 5-2b (Husein & Chung, 2019). La puerta de olvido  $f_t$  determina la influencia del estado anterior en el estado actual. La puerta de entrada  $i_t$  recibe la nueva información para actualizar el estado de la celda. La puerta de salida  $o_t$  proporciona la información basada en el estado de la celda. La función sigmoidea  $\sigma$  ajusta los valores de salida de estas puertas a un valor entre 0 y 1, que se interpreta como una probabilidad (Kwon et al., 2020).

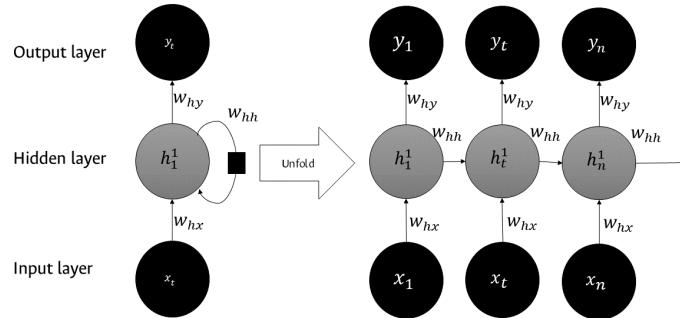
La memoria explícita del LSTM hace que esta técnica sea ideal para un modelo de pronóstico a corto plazo, casi en tiempo real (Ghimire, Deo, Downs, & Raj, 2019).

El LSTM tiene tres estados de funcionamiento:

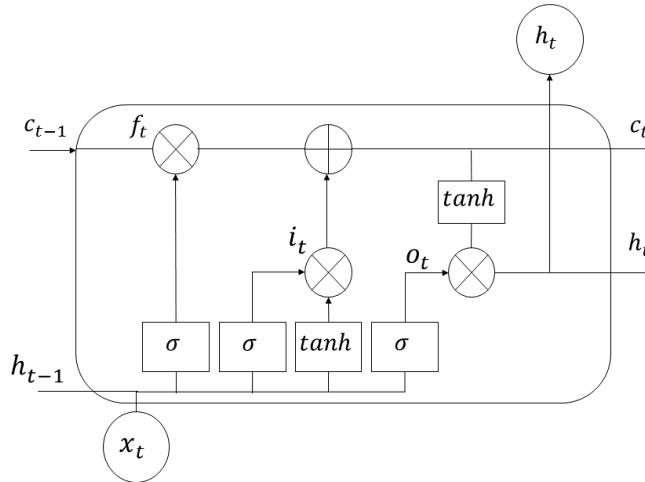
1. Activación de la puerta de entrada: la memoria de la celda acumula nueva información
2. Olvidar puerta activada: la celda borra la información acumulada
3. Activación de la puerta de salida: la salida final se propaga al estado final

Desde la Ecuación 5-10 a Ecuación 5-13 se muestran las ecuaciones de propagación hacia adelante de

Figura 5-2



(a) RNN Structure



(b) LSTM architecture

LSTM

$$f_t = \sigma(w_{xf}x_t + w_{hf}h_{t-1} + b_f) \quad (5-10)$$

Donde  $h_{t-1}$  representa el último estado anterior,  $\sigma(\dots)$  es la función sigmoidal logística.

El estado interno de la celda LSTM se actualiza con el ponderado del bucle propio condicional  $f_t$  de la siguiente manera:

$$c_t = f_t c_{t-1} + i_t \tanh(w_{xc}x_t + w_{hc}h_{t-1} + b_c). \quad (5-11)$$

La unidad de celda  $c_t$  es un bucle automático lineal que está controlado por una unidad de puerta de olvido  $f_t$ , que determina la contribución directa de  $c_{t-1}$  a un valor entre 0 y 1.

La unidad de entrada externa  $i_t$  se calcula de manera similar a la puerta de olvido, pero con sus propios parámetros:

$$i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + b_i). \quad (5-12)$$

La salida  $h_t$  de la celda LSTM también se puede apagar, a través de la puerta de salida  $o_t$ , que también usa una unidad sigmoidea para la activación:

$$h_t = o_t \tanh(c_t) \quad (5-13)$$

$$o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + b_o) \quad (5-14)$$

donde  $i$ ,  $f$ ,  $o$  y  $c$  son la entrada, el olvido, la puerta de salida y la celda, todas tienen el mismo tamaño que el vector oculto  $h$ , y  $w$  es la matriz de ponderados.

#### 5.2.4 Control de calidad de datos

Para reducir el riesgo de sesgos inducidos por datos inexactos, los autores aplican un proceso de control de calidad que consta de tres etapas: i) Verificación de la integridad de los datos: muestras que carecen de información esencial (código de estación, código de variable, fecha y hora, y valor de los datos) se eliminan del conjunto de datos. ii) detección de anomalías basada en límites fijos: para evitar sobreestimar el recurso, descartamos las muestras que superan el valor máximo horario de radiación solar extraterrestre ( $I_0$ ). Por lo tanto, se requiere que  $I_0 \geq I_{mt}$  donde  $I_{mt}$  es la irradiancia solar global medida en el momento  $t$ , y  $I_0$  es

$$I_0 = I_{sc} \left[ 1 + 0,033 \cos \left( 360 \frac{D-3}{365} \right) \right] * \sin \beta \quad (5-15)$$

Donde  $D$  es el día juliano,  $I_{sc}$  es la constante solar ( $1.367 [W/m^2]$ ) que representa la energía desde el Sol de la superficie perpendicular a la dirección de propagación de la irradiancia por unidad de área (Şen, 2008) y  $\sin \beta = \cos \phi \cos \delta \cos \omega_s + \sin \phi \sin \delta$ . iii ) detección de anomalías basada en límites horarios: para evitar subestimar el recurso, utilizamos como límite inferior el 3 % de la irradiancia solar global en cielo despejado ( $I_{cst}$ ), donde  $I_{cst} = I_0 \tau$ . Consideramos las muestras por debajo del límite inferior como valores anómalos y las eliminamos del conjunto de datos.

Para estimar la transmitancia atmosférica  $\tau$  y  $I_{cst}$ , los autores implementan el modelo de Kreith y Kreider (Şen, 2008):

$$I_{cs} = I_0 \tau \quad (5-16)$$

El procedimiento anterior sigue la normativa UNE500540 (AENOR, 2004), y las recomendaciones de (Estévez et al., 2011), que aplicaba el control de calidad a los datos de irradiancia solar en España.

### 5.2.5 Imputación de datos

La imputación de datos es el proceso de completar los datos faltantes con valores razonables. En este caso, los autores utilizan el índice de claridad, que es la relación entre la irradiancia solar a nivel del suelo y la radiación solar extraterrestre, en lugar de tratar directamente los datos de irradiancia solar.

En nuestros experimentos, los autores completaron los datos faltantes el primer día con un valor de 1. Luego, la técnica considera tres casos: si los datos faltantes están en i) 6, ii) 18, o iii) entre 7 y 17 horas. En el primer caso, el dato faltante es el promedio entre el valor en la misma hora del día anterior con el valor del índice de claridad de las 7 horas del día actual. En el segundo caso, el valor es el resultado del promedio entre los datos a las 17 horas del día actual y el valor a las 18 horas del día anterior. En el tercer caso, el valor imputado corresponde a la media entre el valor a la misma hora del día anterior y el valor de las horas inmediatamente anterior y siguiente. Si el dato inmediatamente siguiente no está disponible, éste no se considera en el cálculo.

### 5.2.6 Criterios de comportamiento del modelo

Los autores utilizan la validación estadística para medir el rendimiento de las técnicas de pronóstico. La Tabla 5-1 muestra los errores calculados. Cuanto menor sea el MAE y el RMSE, mejor, y cuanto más cerca de cero se muestre el MBE, mejor (Abreu et al., 2018). La evaluación del desempeño solo considera los datos de la medida, ya que los datos imputados se utilizan exclusivamente para entrenamiento. Por lo tanto, la cantidad de días para la estimación del error es diferente de la cantidad total de datos de cada AWS.

**Tabla 5-1.** Statistical errors

Medida	Definición	Fórmula*
Error medio absoluto (MAE)	Es la distancia vertical entre los valores predichos y medidos. Esta medida cuantifica el error con más énfasis sobre la media y menos sobre los eventos extremos individuales	$\frac{1}{n} \sum_{i=1}^n  p_i - o_i $
Error medio cuadrático (RMSE)	Esta medida proporciona el tamaño del error; sin embargo, es sensible a valores atípicos porque da más peso a los grandes errores. Además, ésta captura la variabilidad más que la tendencia general de la predicción	$\left[ \frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2 \right]^{1/2}$
Error medio de desviación (MBE)	Esta medida proporciona información sobre el rendimiento a largo plazo del modelo, ya que cuando el modelo tiene un error sistemático que presenta predicciones sobreestimadas o subestimadas éstas se pueden compensar en el corto plazo	$\frac{1}{n} \sum_{i=1}^n p_i - o_i$

**Tabla 5-1 continuación de la página anterior**

Medida	Definición	Fórmula*
* $p_i$ es el valor predicho, $o_i$ es el valor observado, $n$ es la cantidad de datos		

Source: (Almorox et al., 2011; C. M. Dos Santos et al., 2014; Gueymard, 2014; J. Li & Heap, 2011; Mayer & Butler, 1993; Blaga et al., 2019)

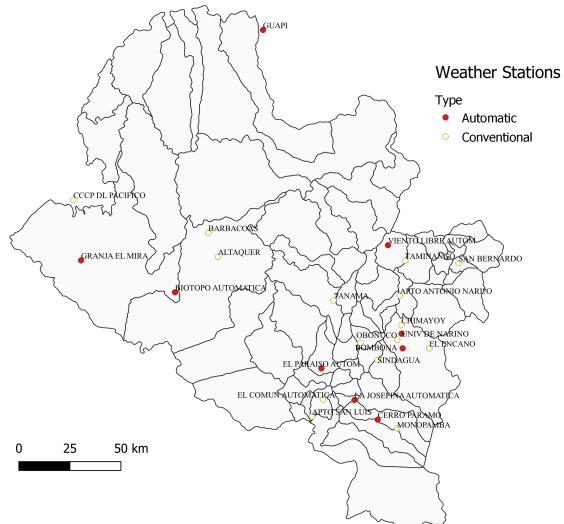
## 5.3 Configuración experimental

### 5.3.1 Localización y conjunto de datos

En este estudio, los autores utilizan datos de irradiancia solar global de doce AWS (ver Tabla 5-2) ubicados en Nariño, Colombia, como se muestra en la Figura 5-3. Esta región está ubicada en la Zona de Convergencia Intertropical y es donde la cordillera de los Andes se divide en dos cordilleras. Esta zona está formada por tres subregiones, Pacífica, Andina y Amazónica. La altitud de todas las AWS varía de 42 a 3.577 msnm. Estas características geográficas permiten evaluar el desempeño de las técnicas de predicción estudiadas en diferentes condiciones fisiográficas y ambientales.

La Figura 5-4 muestra los histogramas del índice de claridad  $k_t$ , basado en las siguientes categorías (Rivero et al., 2017): días nublados  $0, 0 < k_t \leq 0, 2$ , nubosidad parcialmente alta  $0, 2 < k_t \leq 0, 4$ , nubosidad parcialmente baja  $0, 4 < k_t \leq 0, 6$ , soleado  $0, 6 < k_t \leq 0, 75$ , muy soleado  $0, 75 < k_t \leq 1, 0$ . Las líneas verticales discontinuas de cada figura separan las categorías. Biotopo y Cerro Páramo tienen días mayormente nublados, y las otras tienen principalmente días con nubosidad parcialmente alta.

**Figura 5-3.** Localización de las estaciones meteorológicas



### 5.3.2 Pre-procesamiento de los datos

Los métodos de pronóstico tienen como objetivo modelar el componente estocástico de la irradiancia solar global. Por lo tanto, en la etapa de pre-procesamiento, los autores eliminan la tendencia de la serie temporal para reducir el efecto determinista. Con este fin, los autores utilizan el índice de cielo

**Tabla 5-2.** Estaciones meteorológicas automáticas

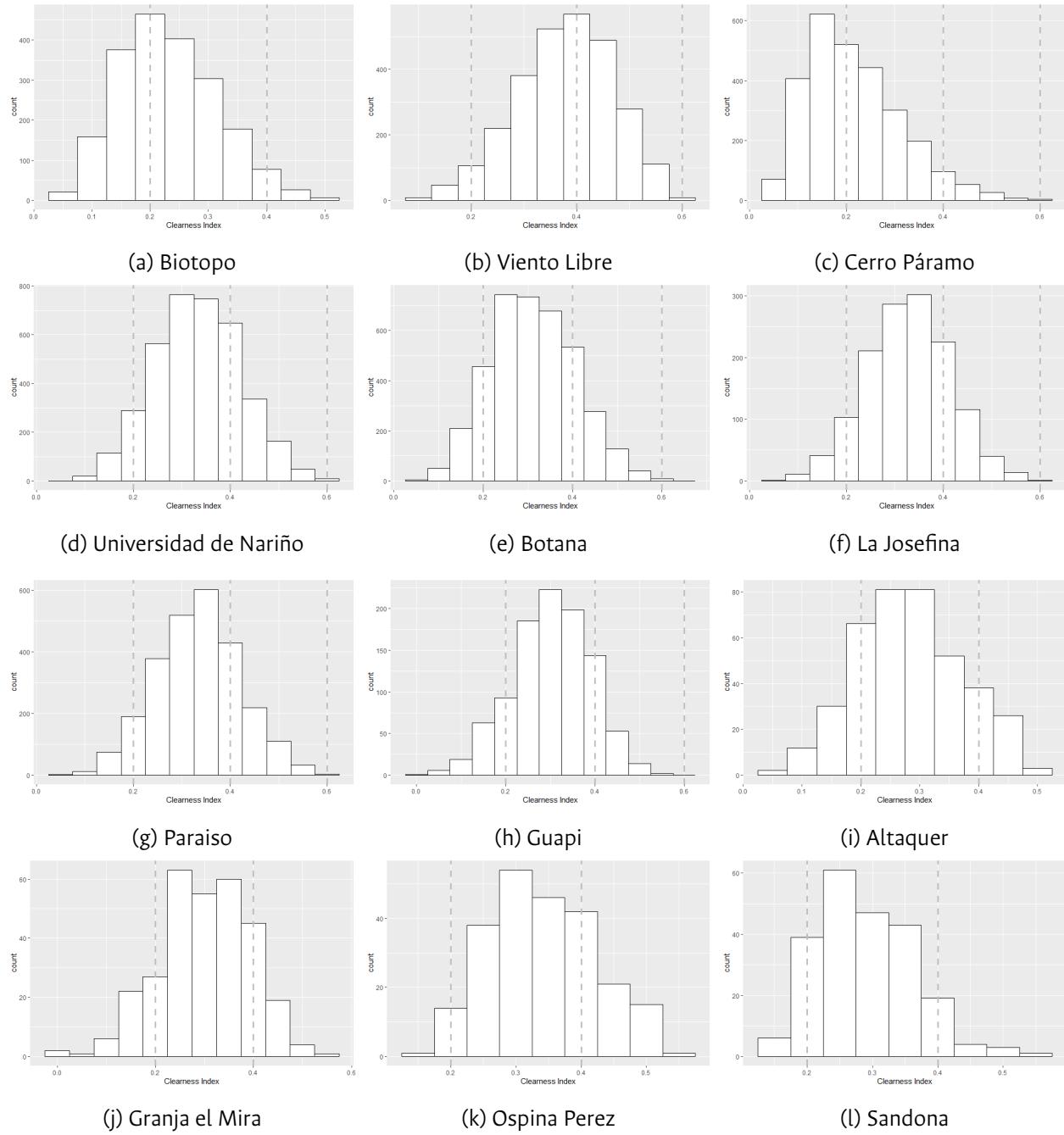
Nombre	Latitud	Longitud	Altitud	Periodo	Región
Biotoxo	1,41	-78,28	512	2005-2017	Pacífico
Altaquer	1,56	-79,09	1.1010	2013-2014	Pacífico
Granja el Mira	1,55	-78,69	16	2016-2017	Pacífico
Cerro-Páramo	0,84	-77,39	3.577	2005-2017	Amazonas
La Josefina	0,93	-77,48	2.449	2005-2017	Andina
Viento Libre	1,62	-77,34	1.005	2005-2017	Andina
Universidad de Nariño	1,23	-77,28	2.626	2005-2017	Andina
Botana	1,16	-77,27	2.820	2005-2017	Andina
El Paraíso	1,07	-77,63	3.120	2005-2017	Andina
Sandoná	1,30	-77,46	1.838	2013-2017	Andina
Ospina Perez	1,25	-77,48	1.619	2013-2017	Andina
Guapi	2,57	-77,89	42	2005-2017	Pacífico

despejado  $k_{tcs}$ , calculado como la irradiancia solar global dividida por la irradiancia solar global de cielo despejado (Benali et al., 2019; Diagne et al., 2013). El  $k_{tcs}$  varía entre 0 y 1 e indica la cantidad de irradiancia solar global en un cielo despejado que llega a la superficie de la Tierra. Por lo tanto,  $k_{tcs} = 0$  se obtiene en un día completamente nublado, mientras que  $k_{tcs} = 1$  se obtiene en un día muy soleado.

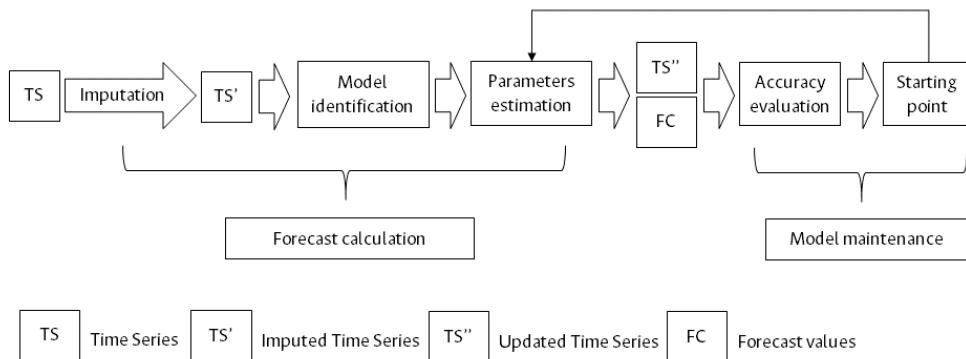
### 5.3.3 Diseño del modelo

Nuestro algoritmo (Fig. 5-5) utiliza una ventana de diez días y predice el undécimo día. Incorpora información histórica tomando los parámetros resultantes de la ventana de búsqueda anterior como punto de partida de los parámetros de la siguiente búsqueda.

En esta investigación, la serie de tiempo sin procesar se somete a un proceso de control de calidad, para aumentar la confiabilidad de los resultados. Una vez finalizado el proceso de control de calidad, las series temporales obtenidas inician la etapa de pronóstico. Esta etapa comienza con el proceso de imputación, que consiste en llenar los vacíos con valores consistentes, en el caso del pronóstico con datos horarios. La serie de tiempo imputada es la entrada para los modelos de pronóstico. El siguiente paso en la etapa de cálculo de pronóstico es la identificación del modelo que es la selección del modelo entre ARIMA, SLP, MLP y LSTM. Una vez definido el modelo, el siguiente paso es la estimación de los parámetros, es decir, el cálculo de las ponderaciones y la determinación del orden polinómico de las ecuaciones de ARIMA. La salida del algoritmo son los valores de pronóstico y una serie de tiempo actualizada que contiene tanto los valores medidos o imputados como los valores pronosticados. La serie de tiempo actualizada es la entrada para la etapa de mantenimiento del modelo, donde la diferencia

**Figura 5-4.** Clasificación de los días por el índice de claridad

entre el valor medido y pronosticado determina la precisión del modelo de pronóstico. El error medido es la base para ajustar el orden de los pesos y polinomios de los modelos de pronóstico para la siguiente ventana de búsqueda. Además, los valores de estas ventanas de búsqueda se toman como punto de partida de la siguiente ventana de búsqueda.

**Figura 5-5.** Esquema del algoritmo

## 5.4 Resultados y discusión

Esta sección contiene dos subsecciones. La primera presenta los resultados de imputación para un rango horario. El segundo muestra el resultado del pronóstico de la irradiancia solar utilizando ARIMA, SLP, MLP y LSTM para un horizonte horario y diario.

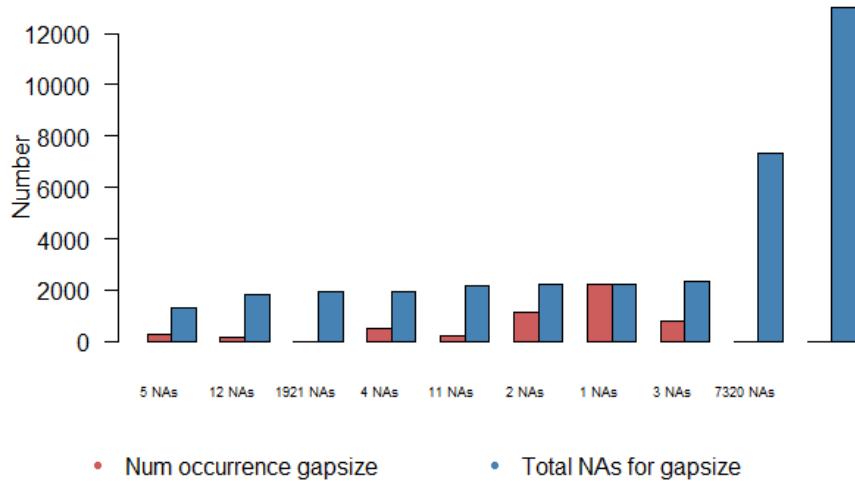
### 5.4.1 Resultados de la imputación de datos horario

El rango horario de interés es de 6:00 a 18:00. Por lo tanto, un día está completo cuando el conjunto de datos contiene todos sus 13 valores de datos (uno por hora). En promedio, para las 12 AWS se imputaron 69,26% de datos de entrenamiento. La Tabla 5-3 contiene los detalles de imputación. En la primera parte, contiene el número de valores medidos y perdidos en todo el conjunto de datos. En la segunda parte, contiene el número de datos faltantes para cada cuarta parte del conjunto de datos completo, por ejemplo, en Biotopo, los datos que se midieron entre 2005/11/26 y 2008/11/04 tienen 10,971 valores faltantes.

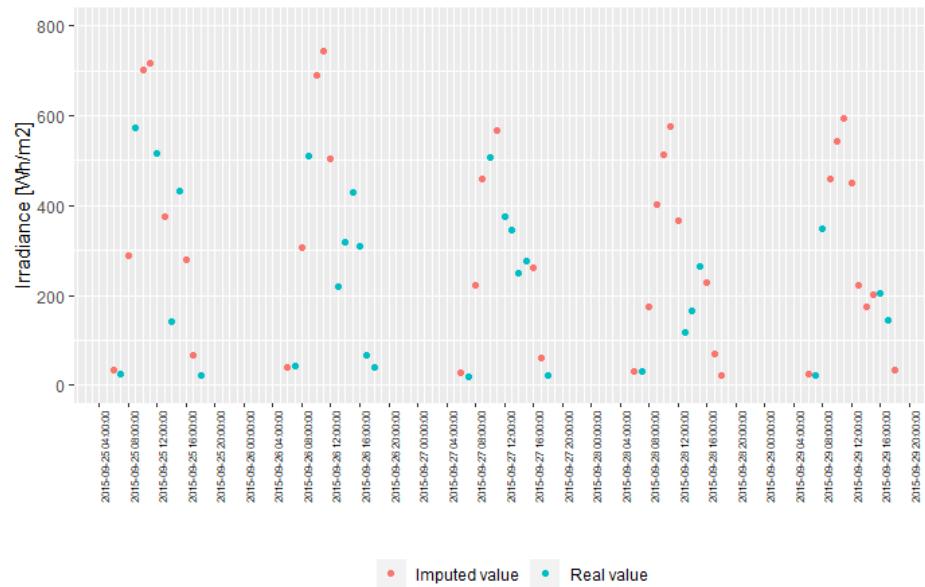
La Josefina es la AWS con más datos perdidos. Requería estimar el 84,2% de los datos. La mayoría de sus valores perdidos se encuentran en la segunda parte de la serie temporal, con 13.880 valores perdidos (99,4%). En la Figura 5-6, las barras rojas indican el número de veces que aparece un espacio faltante de determinado número (secuencia de valores perdidos) y las barras azules muestran el total de valores perdidos con el tamaño del espacio correspondiente. La brecha más larga con valores perdidos consecutivos es 12,999, aproximadamente 2,7 años, y la brecha de un dato faltante (un valor perdido entre dos valores conocidos) ocurre 2.242 veces. Durante 2010 no existe ningún registro, por lo que para el entrenamiento de los modelos se llenaron todos los valores durante ese período. La Figura 5-7 muestra los resultados de imputación en un marco de tiempo de cinco días. Como puede verse, los valores imputados siguen la tendencia de los valores reales. Los resultados del error estadístico

muestran que, en este caso, el proceso de imputación subestima el recurso en  $-15,53 [Wh/m^2]$  (ver Tabla 5-4).

**Figura 5-6.** Distribución de los NAs en la La Josefina



**Figura 5-7.** Imputación de datos La Josefina en 2015



En Biotopo el mayor porcentaje de datos faltantes se ubica en la tercera parte de la serie temporal (91,3%), entre la posición 27.933 a 41.898. La brecha más larga con valores perdidos es 7.252. La brecha con un dato perdido ocurre 2.560 veces. En Guapi hay 13.040 (92,4%) valores perdidos durante la tercera parte de la serie de tiempo. El tamaño de espacio más largo con valores perdidos es 3.913. La brecha con un dato perdido ocurre 2.582 veces. En la segunda parte de la serie temporal de Ospina Perez hay 1.392 valores perdidos (99,8%). El tamaño de espacio más largo con valores perdidos es 1.813. La brecha con un dato perdido ocurre 564 veces. Sandona no registra datos durante la segunda parte

de la serie de tiempo alcanzando 1.391 valores perdidos (100%). No obstante, el tamaño de espacio más largo con valores perdidos es 1,808. La brecha con un dato perdido ocurre 456 veces. Viento Libre durante la primera parte de la serie temporal perdió 12.450 valores (88, 9%). El tamaño de espacio más largo con valores perdidos es 7.325. La brecha con un dato perdido ocurre 4.944 veces. Paraíso perdió 11.259 valores (80, 6%) en la primera parte de la serie temporal. El tamaño de espacio más largo con valores perdidos es 3.856. La brecha con un dato perdido ocurre 4.475 veces. Granja el Mira tiene 1.044 valores perdidos (80, 5%) en la tercera parte de la serie temporal. El tamaño de espacio más largo con valores perdidos es 621. El espacio con un dato perdido ocurre 513 veces. En Universidad de Nariño el caso más crítico ocurre en la segunda parte de la serie de tiempo, ya que carece de 9.655 valores ((66, 1%) del total de datos). El tamaño de espacio más largo con valores perdidos es 914. El espacio con un dato perdido ocurre 4.612 veces. Botana en la segunda parte de la serie temporal perdió 9.666 valores (66, 2%). El tamaño de espacio más largo con valores perdidos es 1.661. La brecha con un dato perdido ocurre 5.791 veces. Cerro Páramo tiene 9.204 valores perdidos (65, 9%) durante el primer trimestre de la serie temporal. El tamaño de espacio más largo con valores perdidos es 1.150. La brecha con un dato perdido ocurre 3.151 veces. En Altaquer la parte más crítica es la segunda parte de la serie temporal con 922 valores perdidos (64, 9%). El tamaño de espacio más largo con valores perdidos es 445. El espacio con un dato perdido ocurre 632 veces.

Viento Libre, Cerro Páramo y Paraíso tiene la mayor cantidad de datos perdidos en la primera parte de la serie temporal, con un promedio de (78, 46%). En Altaquer, Universidad de Nariño, La Josefina, Botana, Ospina Pérez y Sandona, la segunda parte de la serie temporal presenta el mayor porcentaje de datos faltantes 82, 74% en promedio. En el caso de Sandona no hay información registrada durante la segunda parte del período de tiempo. En la tercera parte de la serie temporal, Bitopo, Granja el Mira y Guapi tienen en promedio (88, 06%) de datos faltantes, siendo la mayor pérdida de información.

**Tabla 5-3.** Resultados de imputación

Estación	Marco de tiempo	Resultados de imputación				Distribución de los valores perdidos			
		# de mediciones	# de datos perdidos	Porcentaje de valores perdidos	Total	1 <sup>st</sup> cuarto	2 <sup>nd</sup> cuarto	3 <sup>rd</sup> cuarto	4 <sup>th</sup> cuarto
Bitopo	2005/11/26-2017/08/31	12.811	43.050	77, 1%	55.861	10.971	10.710	12.755	8.614
Altaquer	2013/05/31-2014/08/10	2.576	3.105	54, 7%	5.681	698	922	723	762
Granja el Mira	2016/07/29-2017/08/31	1.723	3.464	66, 8%	5.187	819	805	1.044	796
Viento Libre	2005/11/13-2017/08/31	16.288	39.742	70, 9%	56.030	12.450	9.485	8.494	9.313
Universidad de Nariño	2005/05/12-2017/08/31	21.375	37.060	63, 4%	58.435	9.467	9.655	9.300	8.638
Cerro Páramo	2005/11/25-2017/08/31	22.072	33.802	60, 5%	55.835	9.204	7.959	85.209	8.430
La Josefina	2005/11/28-2017/08/31	8.815	47.020	84, 2%	55.835	12.146	13.880	11.576	9.418
Botana	2005/05/12-2017/08/31	23.036	35.399	60, 6%	58.435	8.577	9.666	8.242	8.914
Paraíso	2005/11/23-2017/08/31	17.551	38.349	68, 6%	55.900	11.259	9.540	9.530	8.020
Ospina Perez	2016/06/18-2017/08/17	1.449	4.128	74, 0%	5.577	747	1.392	1.162	827
Sandona	2016/06/16-2017/08/17	1.454	4.110	73, 9%	5.564	713	1.391	1.162	844
Guapi	2005/10/09-2017/08/30	13.247	43.225	76, 5%	56.472	10.347	10.565	13.040	9.273

La Tabla 5-4 muestra el resultado de los errores estadísticos del proceso de imputación. Los resultados del MAE muestran que Cerro Páramo presenta eventos menos extremos; en consecuencia, los valores imputados siguen una tendencia media. Los resultados de RMSE presentan a Sandona y Granja el Mira como las AWS con más variabilidad y valores atípicos. El MBE muestra que en todas las AWS el recurso

se subestimó en promedio por 12, 83[Wh/m<sup>2</sup>], con un máximo en Ospina Perez de 2, 78[Wh/m<sup>2</sup>] y un mínimo en Guapi de 16, 79[Wh/m<sup>2</sup>]. En promedio, el MAE presenta un error de 81, 07[Wh/m<sup>2</sup>], en la zona Pacífico MAE es 76, 75[Wh/m<sup>2</sup>], en la zona andina es 85, 92[Wh/m<sup>2</sup>], y en la zona Amazónica es 64, 45[Wh/m<sup>2</sup>]. El resultado de RMSE muestra en promedio un error de 118, 15[Wh/m<sup>2</sup>], en las AWS ubicadas en el Pacífico el error es 115, 84[Wh/m<sup>2</sup>], en la zona andina 127, 63[Wh/m<sup>2</sup>], y en la zona Amazónica 103, 69[Wh/m<sup>2</sup>].

**Tabla 5-4.** Errores estadísticos del proceso de imputación

Estación	MAE [Wh/m <sup>2</sup> ]	RMSE[Wh/m <sup>2</sup> ]	MBE[Wh/m <sup>2</sup> ]
Biotopo	69,46	104,16	-14,72
Altaquer	68,24	107,37	-15,82
Granja el Mira	92,22	134,83	-16,18
Viento Libre	80,41	118,04	-11,28
Universidad de Nariño	85,41	125,66	-9,81
Cerro Páramo	64,45	103,69	-15,07
La Josefina	83,06	126,03	-15,53
Botana	89,84	131,31	-13,83
Paraiso	86,46	123,44	-10,2
Ospina Perez	89,95	132,07	-2,78
Sandonia	86,34	136,88	-11,94
Guapi	77,09	117,00	-16,79
Average	81,07	118,15	-12,82

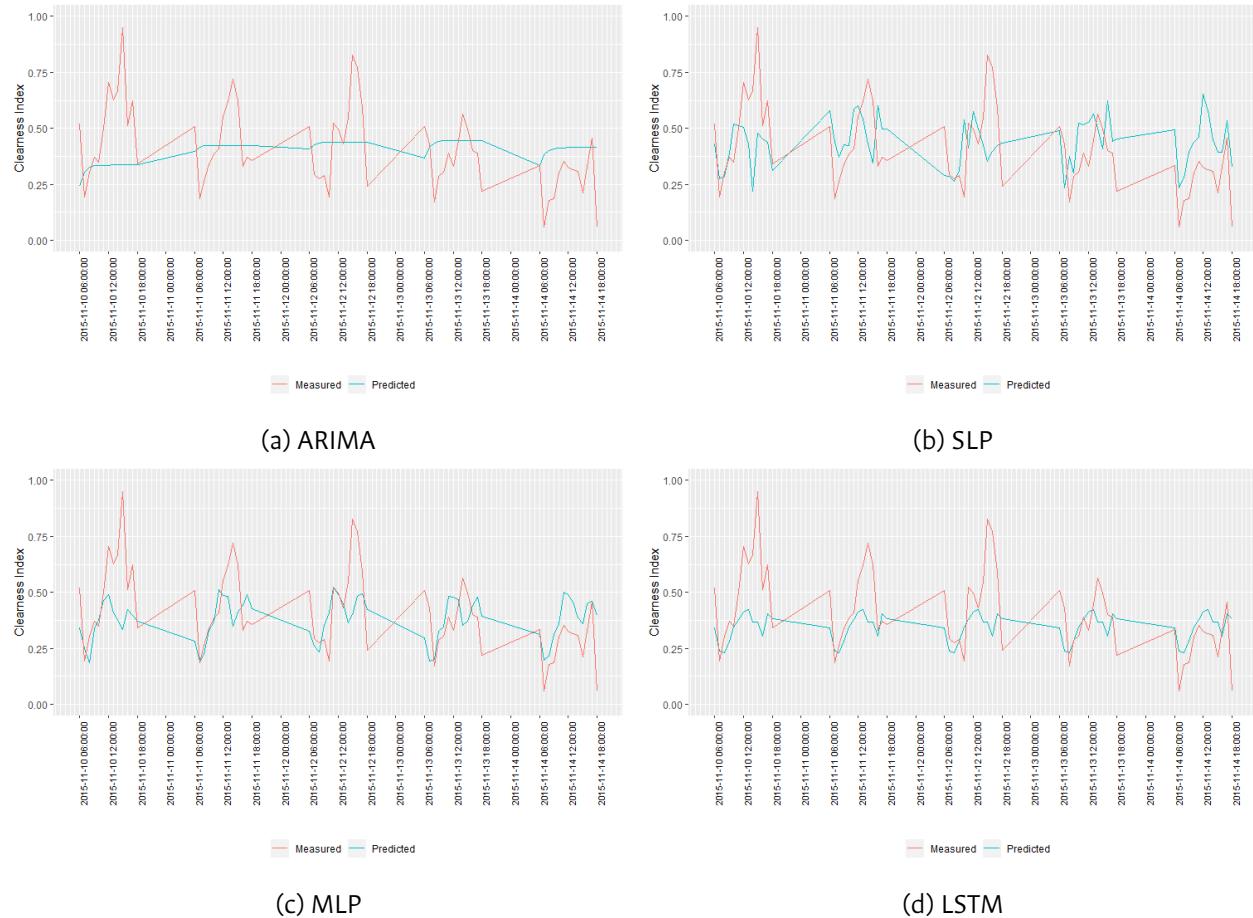
#### 5.4.2 Resultados del pronóstico

Después de aplicar el proceso de imputación de datos, los autores entrena y prueban los cuatro modelos de pronóstico descritos en la Sección 5.2: ARIMA, SLP, MLP y LSTM. En total, entrena 48 modelos de pronóstico (4 modelos para cada AWS). Es conveniente recordar que durante 2007, 2008 y 2010, las precipitaciones aumentaron en Colombia a causa de La Niña. Este patrón climático fue más fuerte en 2010 de junio a diciembre, con un pico en octubre y noviembre que registró un Índice Oceánico del Niño (ONI) de -1,7. Asimismo, en 2009, 2015 y 2016, la temperatura aumentó debido a El Niño. El ONI alcanzó una intensidad de 2,6 en diciembre de 2015, siendo el valor más alto registrado por la Administración Nacional Oceánica y Atmosférica (NOAA) desde 1950. Estos fenómenos naturales afectan el desempeño de las técnicas de pronóstico debido a cambios en la nubosidad.

La Figura 5-8 muestra los resultados obtenidos con los modelos de pronóstico implementados en Biotopo en un horizonte horario durante cinco días. SLP es el modelo que mejor describe la variabilidad del índice de claridad en esta AWS. En este caso, el modelo ARIMA sigue una tendencia media de la serie temporal sin seguir la variabilidad.

La Figura 5-9 presenta que los modelos basados en redes neuronales superan al modelo ARIMA en Guapi AWS. Como resultado, SLP, MLP y LSTM describen la variabilidad de la irradiancia solar mejor que ARIMA. En Biotopo y Guapi ubicados en la zona del Pacífico, los mejores modelos son los mode-

**Figura 5-8.** Resultados del pronóstico horario en Biotopo

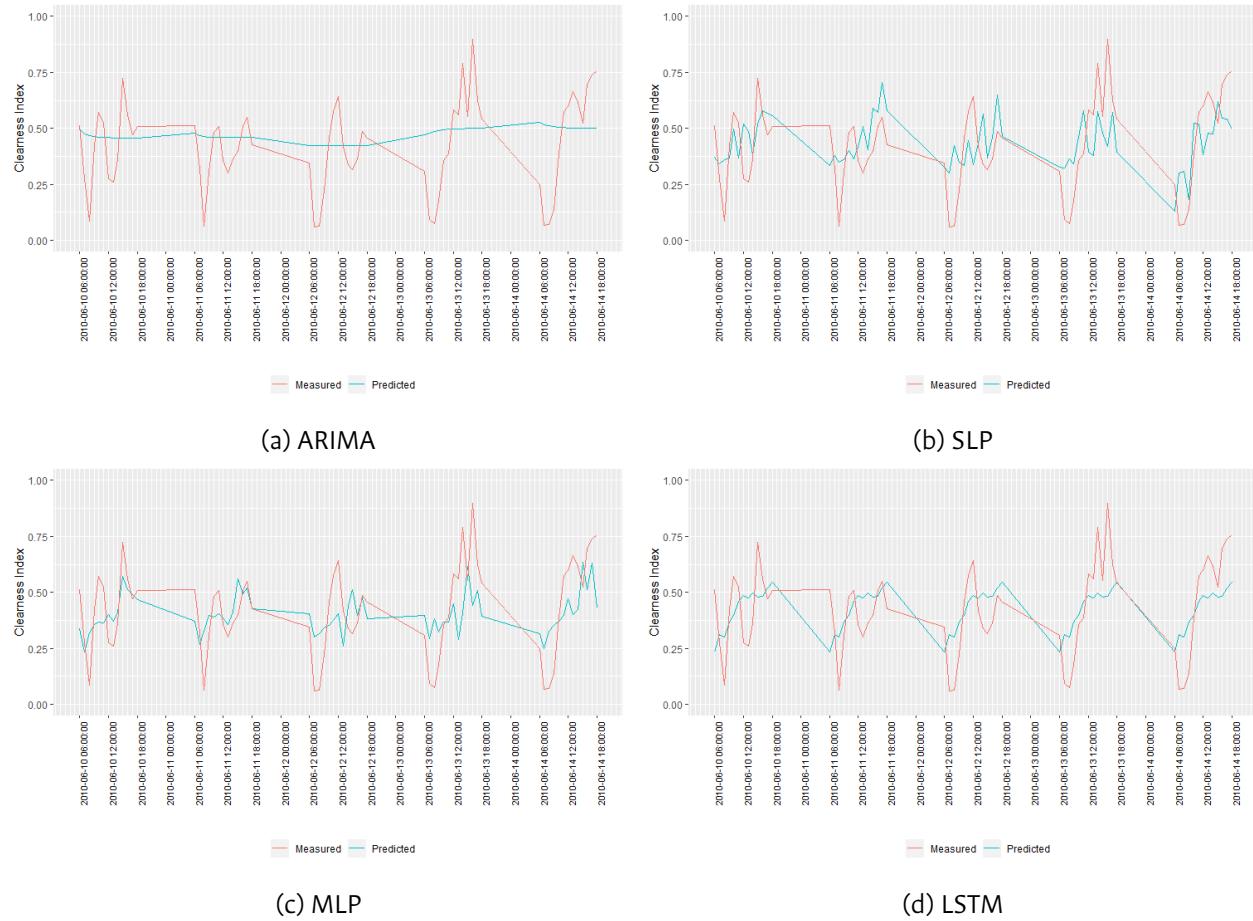


los basados en redes neuronales para un horizonte de pronóstico por hora. SLP y MLP presentan más variabilidad, y LSTM suaviza la variabilidad de series de tiempo más que otros modelos de este tipo. En ambas AWS el modelo ARIMA sigue una tendencia media más que la variabilidad de la serie temporal. Este resultado muestra que ARIMA no tiene suficiente robustez para seguir la variabilidad de la irradiancia solar global en períodos cortos de tiempo.

En Altaquer los modelos basados en redes neuronales presentan mejores resultados que el modelo ARIMA. Además, es posible observar que todos los modelos de pronóstico sobreestiman el recurso. El modelo MLP sobreestima el recurso más que los otros modelos basados en redes neuronales, consulte la Figura 5-10.

En Granja el Mira, ARIMA y LSTM tienen un comportamiento similar, ambos modelos siguen una tendencia media más que la variabilidad de la serie temporal. En contraste, SLP y MLP describen mejor la variabilidad de la irradiancia solar global, como muestra la Figura 5-11. Conviene señalar que en Granja el Mira se imputan el 66,8 % de los datos. Por tanto, un gran porcentaje de datos imputados en una

**Figura 5-9.** Resultados del pronóstico horario en Guapi

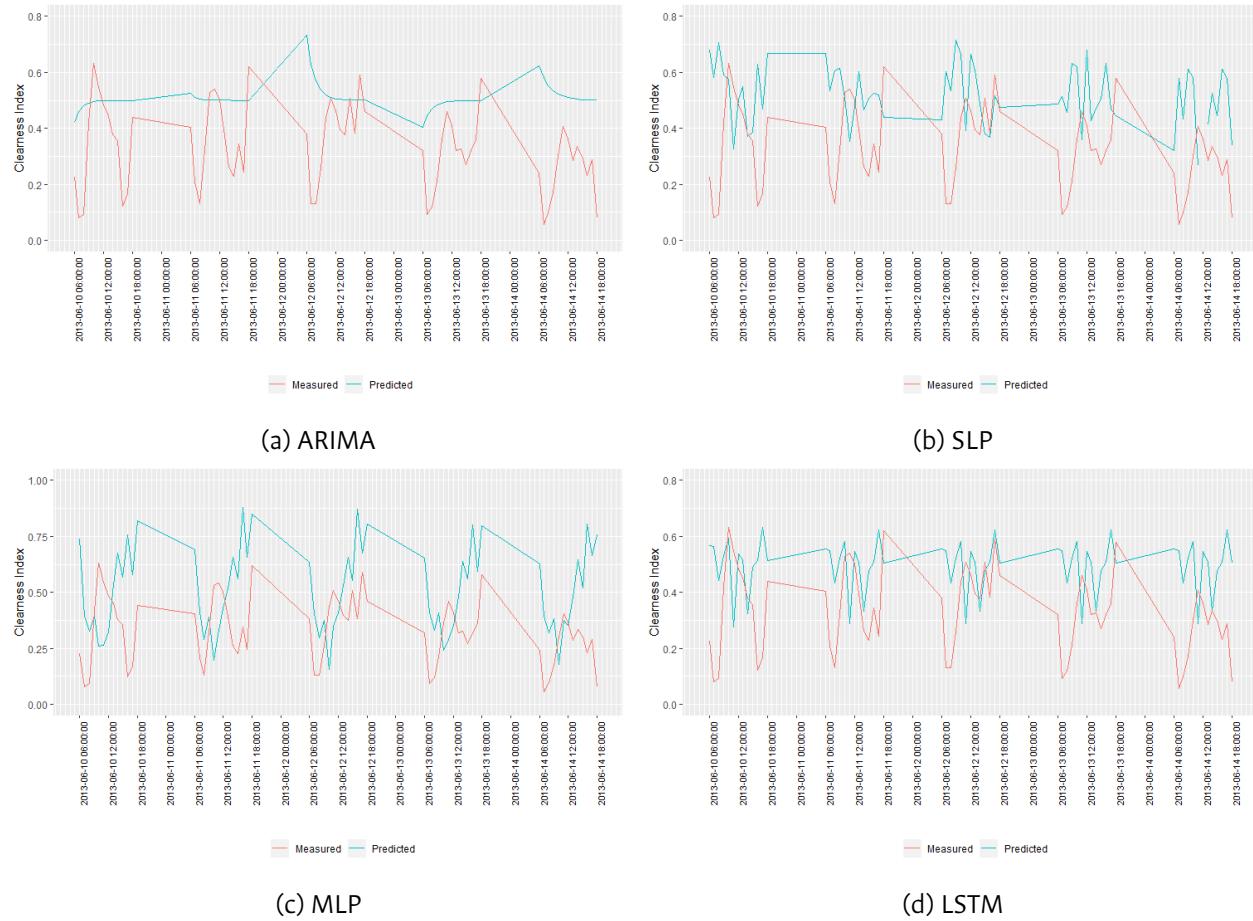


serie de tiempo corta, como en este caso, podría afectar el desempeño de los modelos de pronóstico. Además, la Figura 5-10 y 5-11 muestran que el modelo ARIMA predice valores futuros siguiendo una tendencia media debido a la alta variabilidad de la serie de tiempo en una marca de tiempo por hora.

En Viento Libre, la AWS ubicada en la zona andina con menos nubosidad, la Figura 5-12 muestra que los modelos basados en redes neuronales superan al modelo ARIMA. A pesar de la reducción de la perturbación por la nubosidad, el modelo ARIMA no sigue la variabilidad de la irradiancia solar global con la misma fiabilidad que los otros modelos.

Universidad de Nariño y Botana están ubicados cerca uno del otro; sin embargo, Botana presenta más nubosidad que Universidad de Nariño. Por lo tanto, esta condición podría afectar el sesgo de los modelos de pronóstico. En Universidad de Nariño los modelos de pronóstico con mayor variabilidad son SLP y MLP. Por el contrario, ARIMA y LSTM siguen una tendencia media. En el caso de LSTM, incluso si sigue una tendencia media, describe los cambios principales en la irradiancia solar global mejor que ARIMA, como se muestra en la Figura 5-13.

**Figura 5-10.** Resultados del pronóstico horario en Altaquer



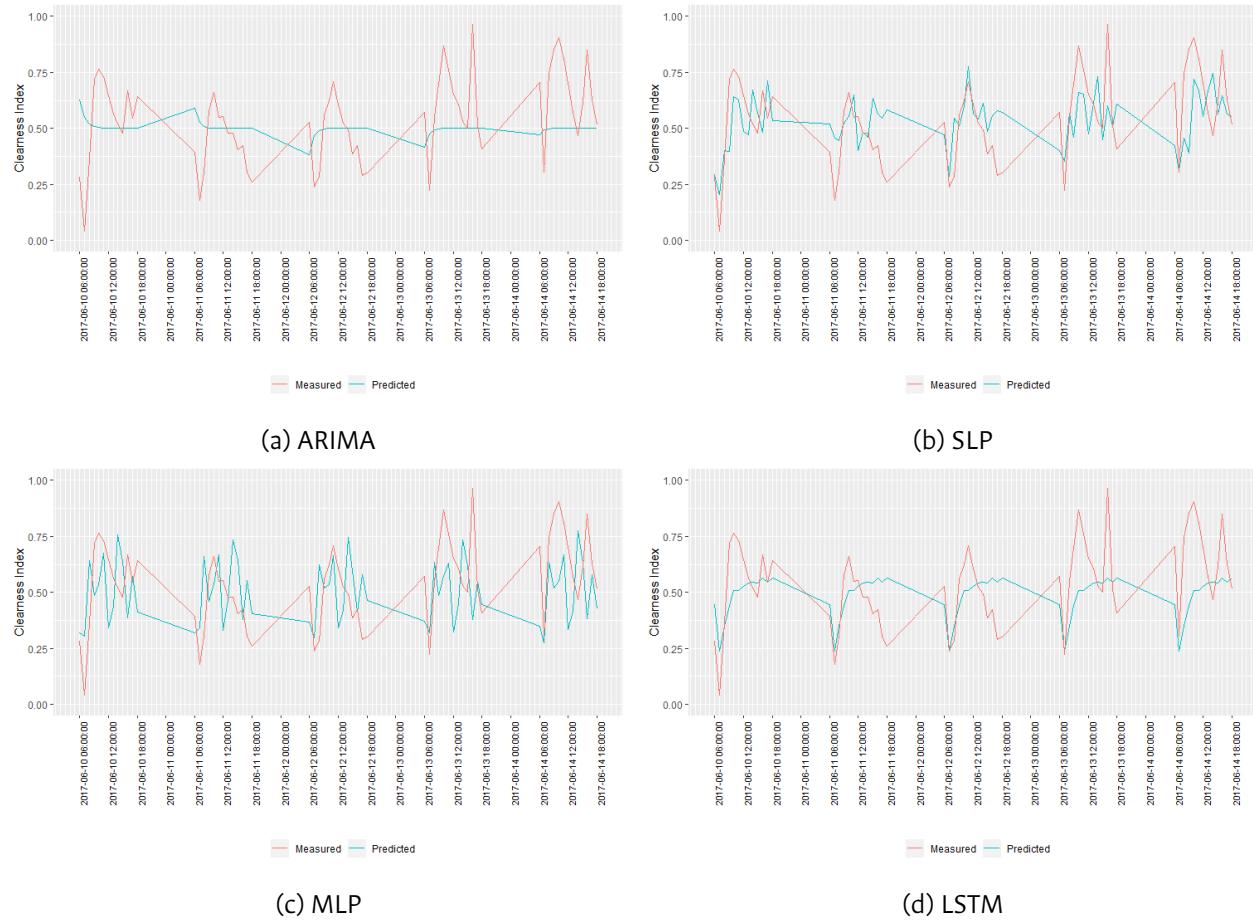
La Figura 5-14 muestra que en Botana todos los modelos de pronóstico subestiman el recurso. Además, es posible observar que SLP es el mejor modelo de pronóstico seguido por MLP.

La Josefina es un caso crítico por el porcentaje de datos faltantes, casi el 80 % de la serie temporal. En este caso, MLP y SLP siguen la variabilidad de la serie temporal. ARIMA y LSTM describen la tendencia media de la serie temporal.

La Figura 5-16 presenta los resultados del pronóstico en Paraiso. En este caso, LSTM y SLP describen la variabilidad de la serie temporal. Además, al igual que en los demás AWS ubicados en la zona andina, el modelo ARIMA describe una tendencia media. Los resultados de los AWS ubicados en la zona andina, los cuales tienen una serie de tiempo con más de un año de mediciones, presentan a los modelos basados en redes neuronales como la mejor opción de pronóstico. Asimismo, es posible observar que el modelo ARIMA en todos los casos describe los valores futuros como una tendencia media de la serie de tiempo en un horizonte horario.

En Ospina Perez el modelo MLP es la mejor técnica de pronóstico debido a que sigue la variabilidad de

**Figura 5-11.** Resultados del pronóstico horario en Granja el Mira

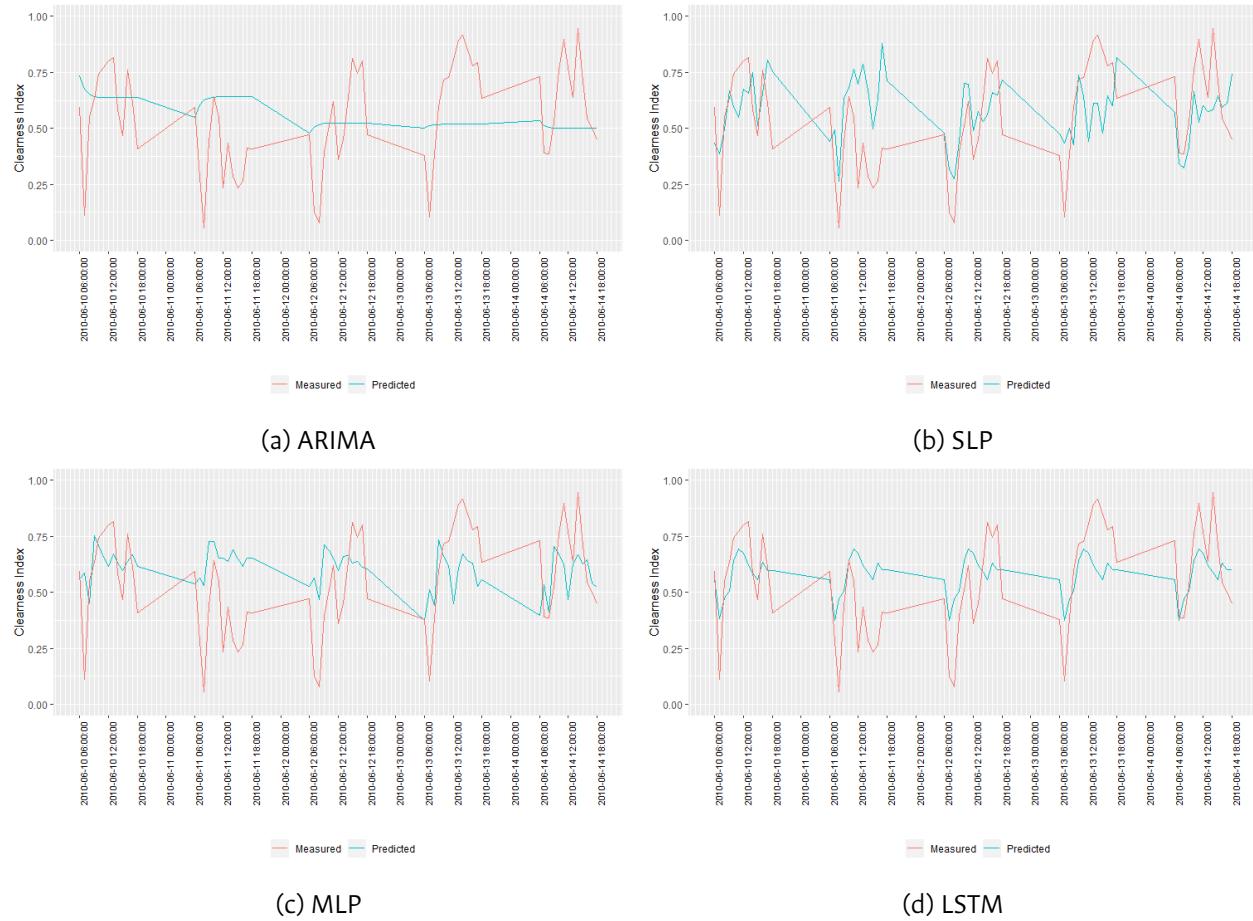


la irradiancia solar global con menos sesgo. En esta AWS, como en las otras AWS ubicadas en la zona andina, el modelo ARIMA describe los valores futuros con una tendencia media, como muestra la Figura 5-17.

Sandona es una AWS ubicado en la zona andina con un año de mediciones en promedio. En este caso, SLP y MLP tienen el mejor desempeño en pronosticar valores futuros. LSTM sigue la variabilidad de la serie temporal; sin embargo, sobreestima el recurso. ARIMA, como en los demás casos, pronostica los valores futuros siguiendo una tendencia media. Los resultados de Ospina Perez y Sandona son consistentes con los resultados obtenidos en el otro AWS ubicado en la zona andina con más mediciones, los modelos basados en redes neuronales superan al modelo ARIMA. Además, el modelo ARIMA tiene dificultades para describir la variabilidad de la serie temporal; por lo tanto, este modelo predice valores futuros como la tendencia media.

Cerro Páramo es la única AWS ubicado en la zona Amazónica, además, es el AWS con más nubosidad. La Figura 5-19 muestra que los modelos basados en redes neuronales superan al modelo ARIMA. SLP

**Figura 5-12.** Resultados del pronóstico horario en Viento Libre

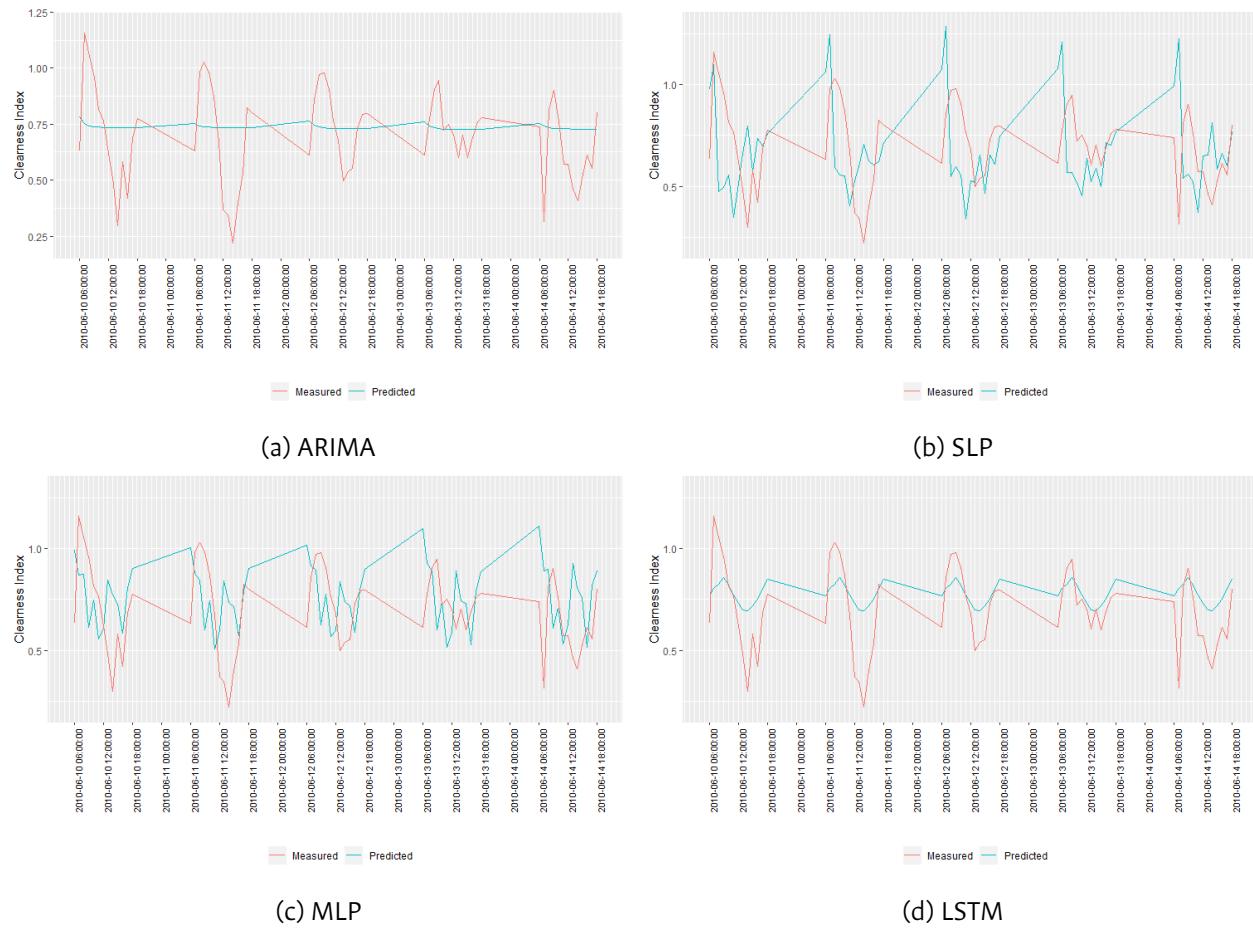


suaviza las variaciones de las series de tiempo más en comparación con los otros modelos basados en redes neuronales. En este caso, ARIMA modela la tendencia media de la serie temporal; por tanto, los valores futuros predichos con este modelo siguen una tendencia media más que la variabilidad de la serie temporal.

Para cuantificar el desempeño del modelo de pronóstico en un horizonte horario, los autores calculan las medidas de error MAE, RMSE y MBE para cada AWS. En Biotopo las medidas de error MAE, RMSE y MBE muestran que SLP es el mejor modelo, seguido del modelo MLP, consulte la Figura 5-20a. Contrastando los resultados del error estadístico de todos los modelos, utilizando SLP, hay una mejora del 11,53 %, 9,8 % y 78,81 % en MAE, RMSE y MBE en comparación con LSTM. Usando MLP en lugar de LSTM, el error se reduce en 7,34 %, 6,35 % y 35,52 % en MAE, RMSE y MBE. Los resultados de RMSE y MAE muestran que ARIMA tuvo un 4,74 % y un 4,51 % menos de errores que LSTM; sin embargo, LSTM presenta menos sesgo que ARIMA, reduciendo el error en un 16,98 %.

La Figura 5-20b muestra que considerando las tres medidas de error, el mejor modelo en Guapi es

**Figura 5-13.** Resultados del pronóstico horario en Universidad de Nariño

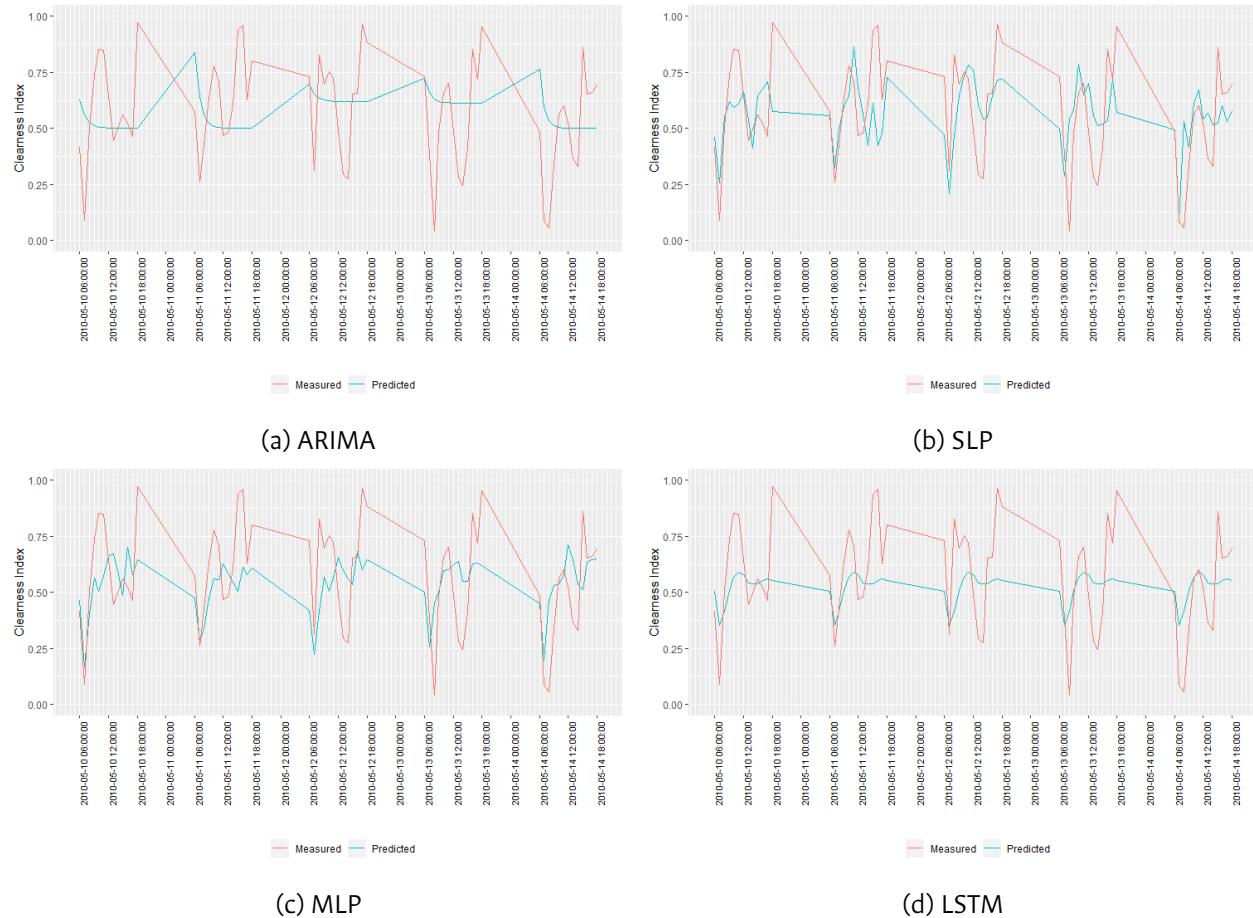


SLP, seguido de MLP y ARIMA. En este caso, LTSM tiene un rendimiento inferior al resto en todas las mediciones de error, aumentando el error en un promedio de 9,26 %, 8,44 % y 48,2 % en MAE, RMSE y MBE, respectivamente.

Los modelos basados en redes neuronales superan a ARIMA en entornos nublados en la zona del Pacífico. En Altaquer, LSTM es el mejor modelo considerando las medidas de error RMSE, MAE y MBE, vea la Figura 5-21a. MAE y RMSE muestran que LSTM es el mejor modelo en Granja el Mira, y que ARIMA tiene menos sesgo, vea la Figura 5-21b.

Los resultados muestran que en la zona del Pacífico (Biotopo, Altaquer, Granja el Mira y Guapi), LSTM supera a los demás cuando el conjunto de entrenamiento es más pequeño, y SLP supera a los demás cuando el conjunto de entrenamiento es mayor. MLP puede sufrir un sobreajuste ya que presenta un rendimiento deficiente cuando no hay suficientes datos de entrenamiento. Los resultados de MBE en Altaquer y Granja el Mira muestran que los modelos basados en redes neuronales presentan menos sesgo en condiciones nubladas. En consecuencia, describen un mejor comportamiento de influencia de

**Figura 5-14.** Resultados del pronóstico horario en Botana

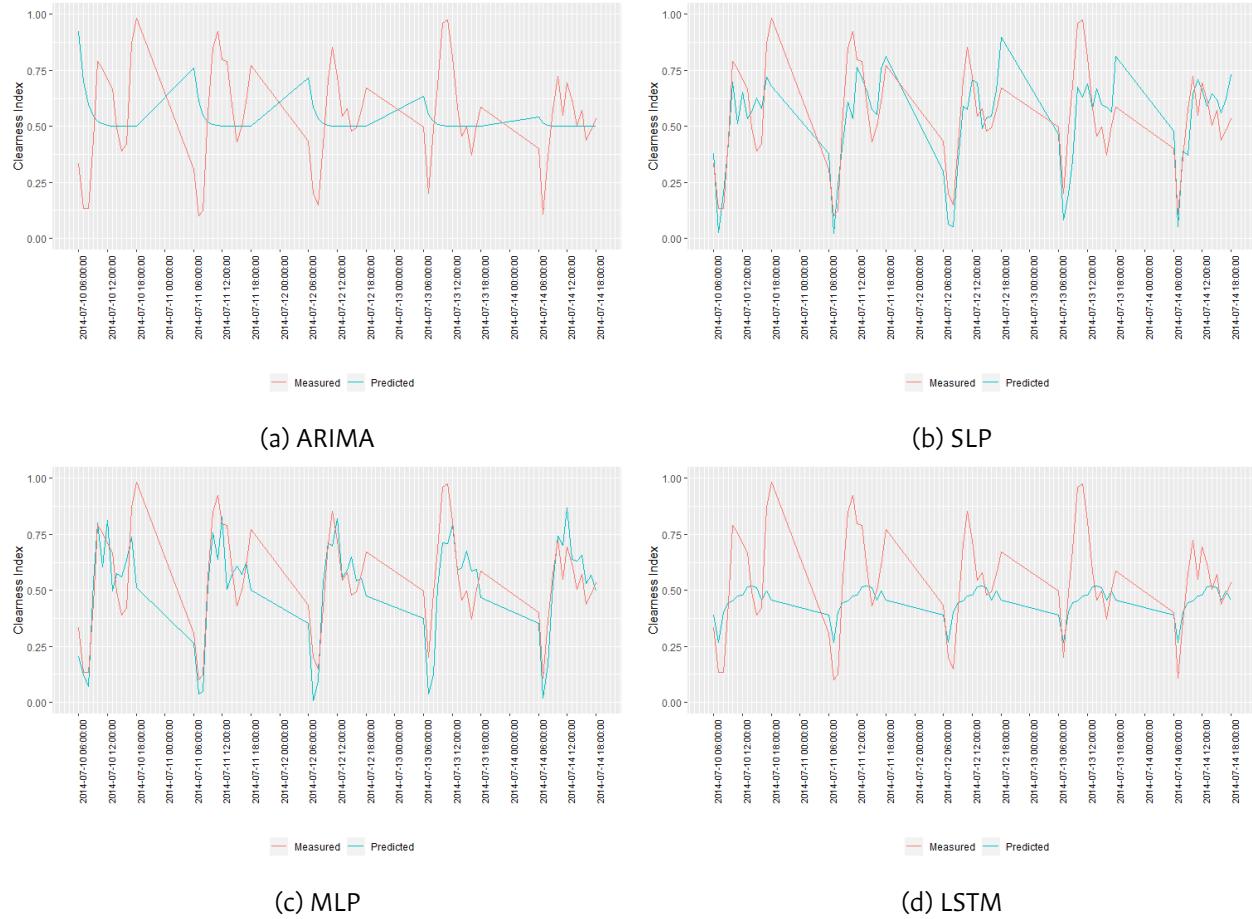


las nubes. Por lo tanto, LSTM fue la mejor opción para pronosticar la irradiancia solar global para los conjuntos de datos más pequeños. El modelo SLP fue la mejor opción para el pronóstico de un día de anticipación, seguido por el modelo MLP en conjuntos de datos más grandes.

Viento Libre es el AWS con menos nubosidad ubicado en la zona andina, ver la Figura 5-4. Ésta es la única AWS con una altitud por debajo de 1.100 msnm en esta zona, consulte la Tabla 5-2. En cuanto a RMSE y MAE, los modelos basados en redes neuronales superan al modelo ARIMA en un 23,43 % y un 23,30 %. El resultado de MBE expone que todos los modelos de pronóstico subestiman el recurso, y el modelo LSTM tiene más varianza que los otros, seguido por ARIMA, vea la Figura 5-22.

La Figura 5-23 muestra el resultado de los errores estadísticos para Universidad de Nariño y Botana AWS. Estos AWS están ubicados más cerca uno del otro, como se ilustra en la Figura 5-3, y la diferencia de altitud es de 194 msnm. Botana tiene más días nublados que Universidad de Nariño, ver Figura 5-4. En ambos casos, los modelos basados en redes neuronales (SPL en particular) superan a ARIMA. El MBE muestra que LSTM tiene menos sesgo en entornos nublados.

**Figura 5-15.** Resultados del pronóstico horario en La Josefina

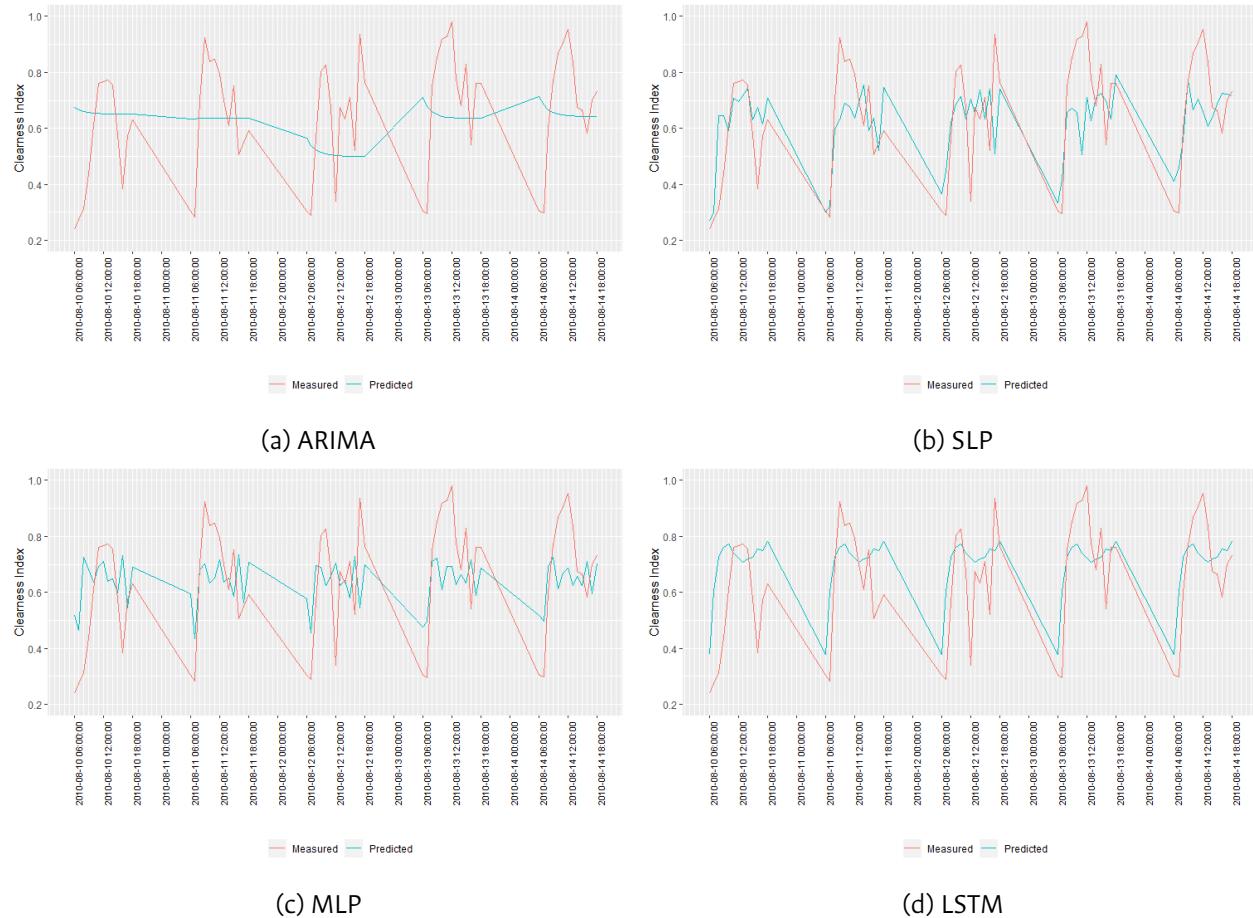


La Figura 5-24 presenta los resultados de errores estadísticos para La Josefina y Paraíso AWS. En ambos casos, los modelos basados en redes neuronales presentan errores más pequeños.

La Figura 5-25a muestra los resultados de errores estadísticos de Ospina Perez. En los casos de RMSE y MAE, el modelo LSTM es mejor que los otros modelos. Los resultados de MBE muestran que MLP tiene menos sesgo. La Figura 5-25b contiene el resultado para Sandona AWS, en este caso, ARIMA y SLP son la mejor opción de pronóstico. En general, para las AWS ubicado en la zona andina, los modelos basados en redes neuronales son los mejores modelos de pronóstico. En las AWS con más datos, el SLP supera a ARIMA y los otros modelos basados en redes neuronales.

En Cerro Páramo la única AWS ubicada en la zona de la Amazonia, los modelos basados en redes neuronales exhiben un mejor desempeño que ARIMA, ver Figura 5-26. Esta AWS tiene días mayormente nublados, vea la Figura 5-4, y es la AWS con la mayor altitud, 3.577 msnm. Los resultados presentan los modelos basados en redes neuronales como la mejor opción para el pronóstico de la radiación solar global en entornos nublados.

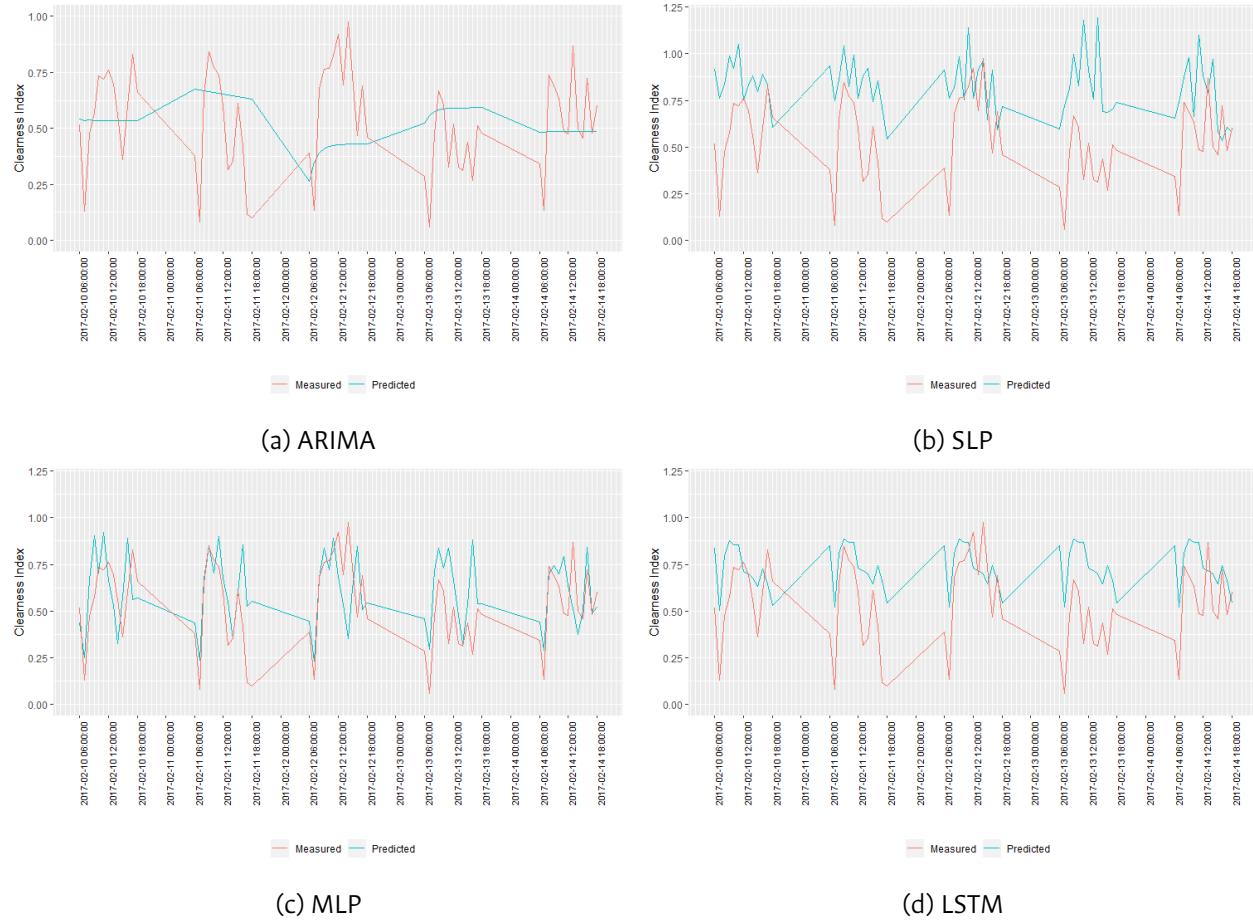
**Figura 5-16.** Hourly forecasting results in Paraiso



Los modelos ARIMA, SLP, MLP y LSTM se implementaron para pronosticar la insolación solar diaria con un día por delante. La Figura 5-27 presenta el resultado de los modelos ARIMA, SLP, MLP y LSTM en un horizonte diario para Bitopo. ARIMA sigue la variabilidad del índice de claridad. Sin embargo, es posible observar que cuando hay picos, el modelo de pronóstico los describe con dos días de retraso en promedio. El SLP representa la variabilidad del índice de claridad mejor que MLP debido a que la información no es suficiente para ajustar todos los parámetros de MLP. En este caso, LSTM predice valores futuros como un promedio de los valores medidos. Por tanto, este modelo no describe la variabilidad del índice de claridad.

En la AWS Guapi, ubicada en la zona del Pacífico, el modelo ARIMA describe la variabilidad del índice de claridad mejor que los modelos SLP, MLP y LSTM, consulte la Figura 5-28. En este caso, como en Biotopo AWS, el modelo LSTM describe los valores futuros como un promedio de las mediciones. Por lo tanto, los modelos ARIMA y basados en redes neuronales, SLP y MLP siguen mejor la variabilidad de la insolación solar.

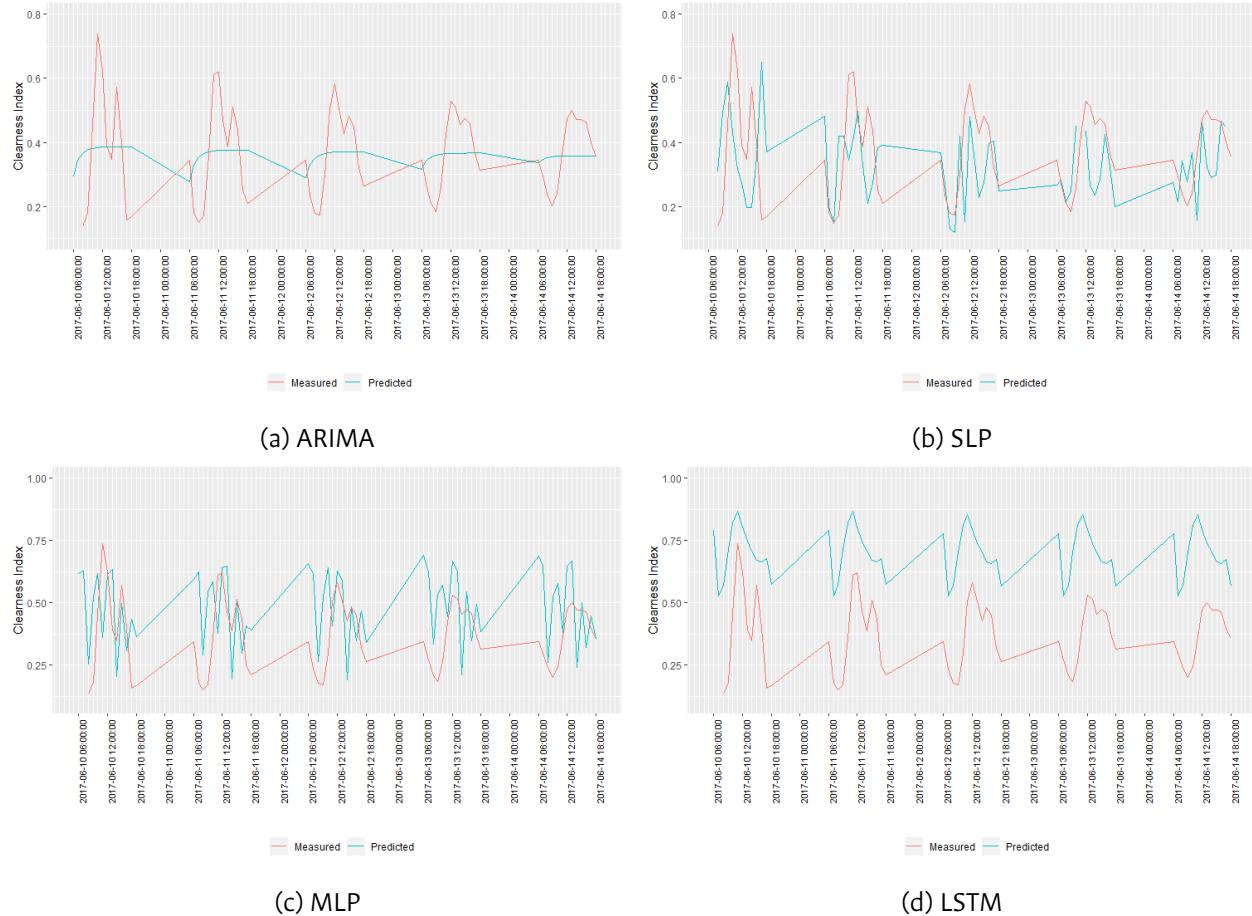
**Figura 5-17.** Resultados del pronóstico horario en Ospina Perez



En Altaquer, una AWS con un año de mediciones en promedio, el modelo ARIMA supera a los modelos basados en redes neuronales. En este caso, los modelos basados en redes neuronales no tienen suficiente información de entrada para entrenar los pesos utilizados en estos modelos. La Figura 5-29 muestra que los modelos SLP y MLP sobreestiman el recurso durante el primer trimestre del tiempo de pronóstico. Una vez que los modelos SLP y MLP ajustan los pesos, los modelos siguen la variabilidad del índice de claridad. En el caso LSTM, es posible observar que este modelo no es tan sensible a la variabilidad del índice de claridad como los modelos SLP y MLP. Por lo tanto, LSTM predice los valores siguiendo una media de valores pasados; sin embargo, este modelo describe las principales alteraciones del índice de claridad. Como resultado, en este caso, ARIMA es el modelo más sensible, mientras que LSTM es el modelo más conservador.

En el caso de Granja el Mira, la Figura 5-30 muestra que el modelo ARIMA supera a los modelos basados en redes neuronales. Como en el caso de Altaquer, los modelos SLP y MLP tienen baja precisión mientras que el proceso de aprendizaje ajusta los pesos para estimar los valores futuros. El modelo LSTM presenta baja variación; por lo tanto, los valores futuros son una media de los valores pasados. Como

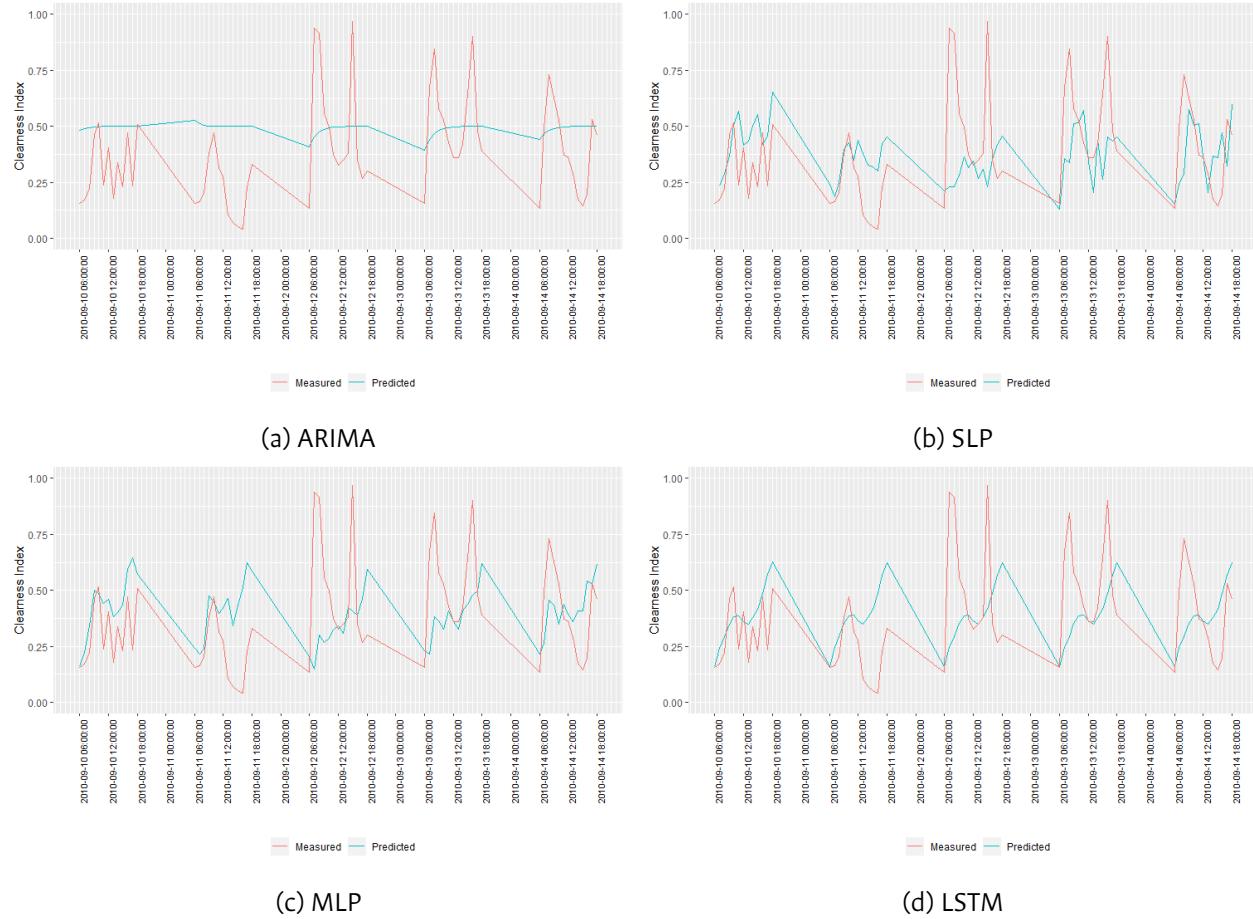
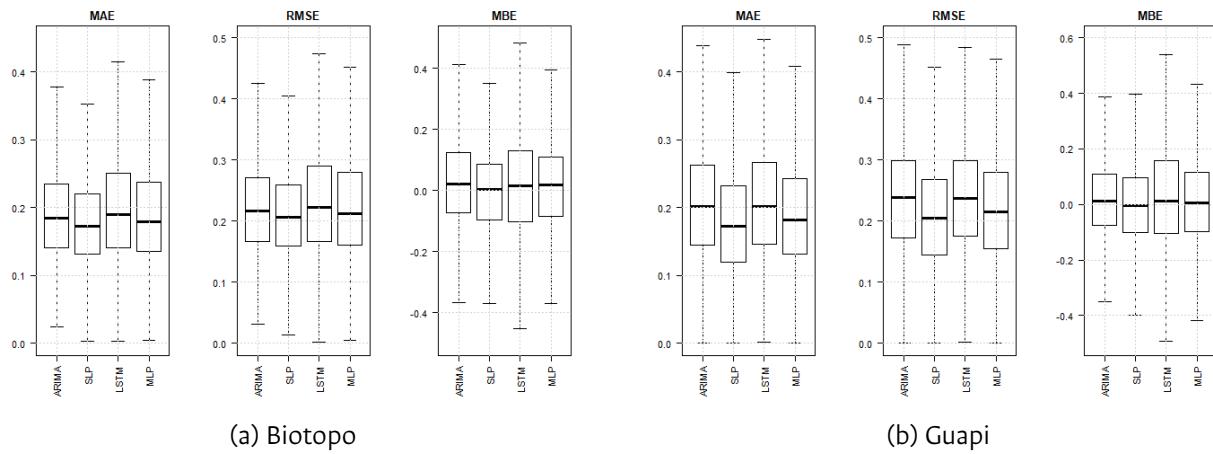
**Figura 5-18.** Resultados del pronóstico horario en Sandona



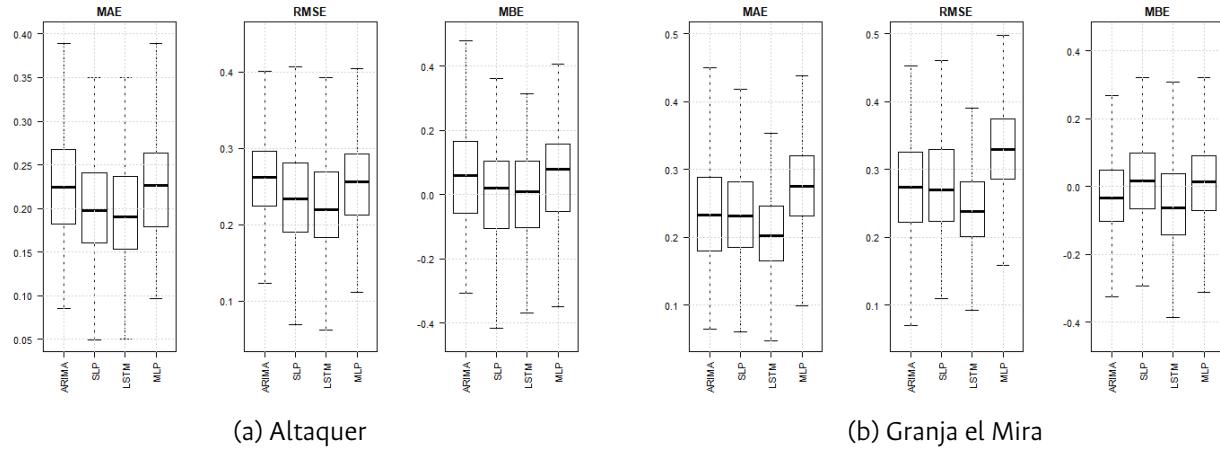
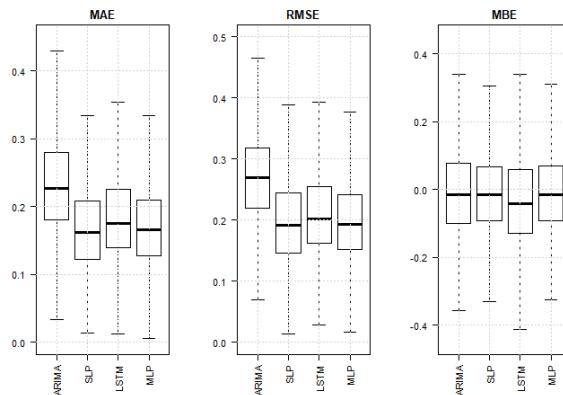
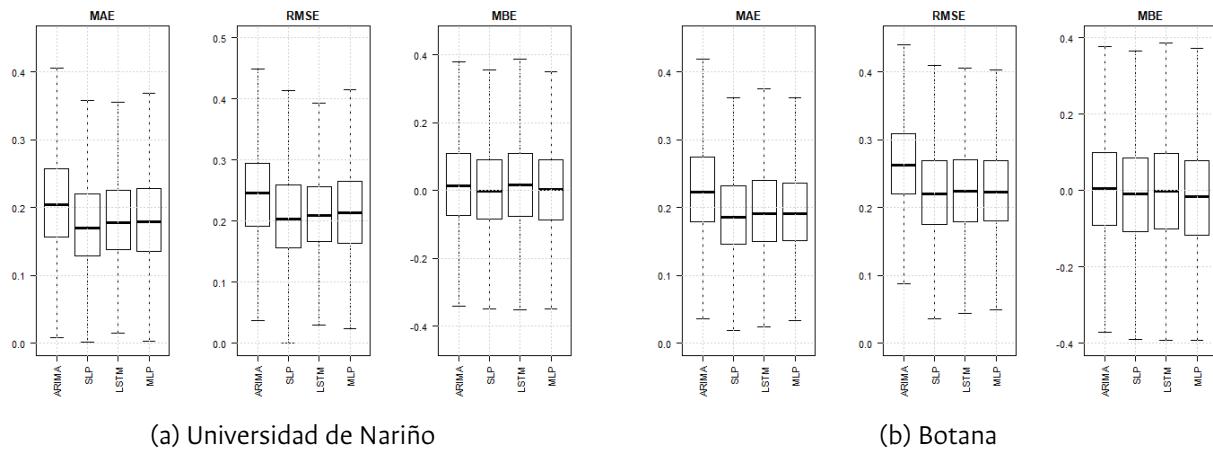
resultado, el ARIMA es el modelo con más robustez para seguir la variabilidad del índice de claridad, y el LSTM es el modelo conservador. Los gráficos de pronóstico de insolación diaria muestran que en las AWS ubicadas en la zona del Pacífico, el modelo estadístico clásico ARIMA es la mejor opción para el pronóstico de insolación solar en una marca de tiempo diaria.

En Viento Libre se puede observar que el modelo ARIMA describe la variabilidad del índice de claridad; sin embargo, hay un retraso de aproximadamente dos días entre los picos de los valores medidos y pronosticados, como muestra la Figura 5-31. Por tanto, los resultados podrían dar lugar a errores en la toma de decisiones. Los modelos basados en redes neuronales tienen baja variabilidad, es más notable en los resultados del modelo LSTM.

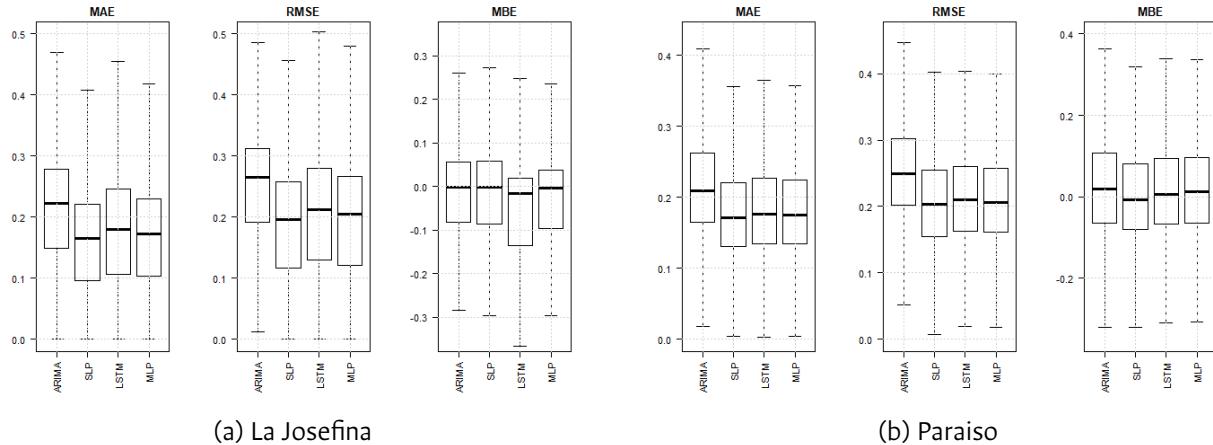
Universidad de Nariño y Botana AWS están ubicados uno cerca del otro. En ambos casos, el modelo ARIMA es la mejor opción para el pronóstico de la insolación solar con un día de anticipación, vea la Figura 5-32 y 5-33. Sin embargo, en el caso de Universidad de Nariño, ARIMA tuvo mejores resultados que en el caso de Botana, debido al nivel de nubosidad de la segunda área. Además, los modelos SLP,

**Figura 5-19.** Resultados del pronóstico horario en Cerro Páramo**Figura 5-20.** Resultados estadísticos

MLP y LSTM pronostican los valores futuros basados en la media de los valores pasados más que en la variación del índice de claridad. Por lo tanto, los modelos basados en redes neuronales son los modelos

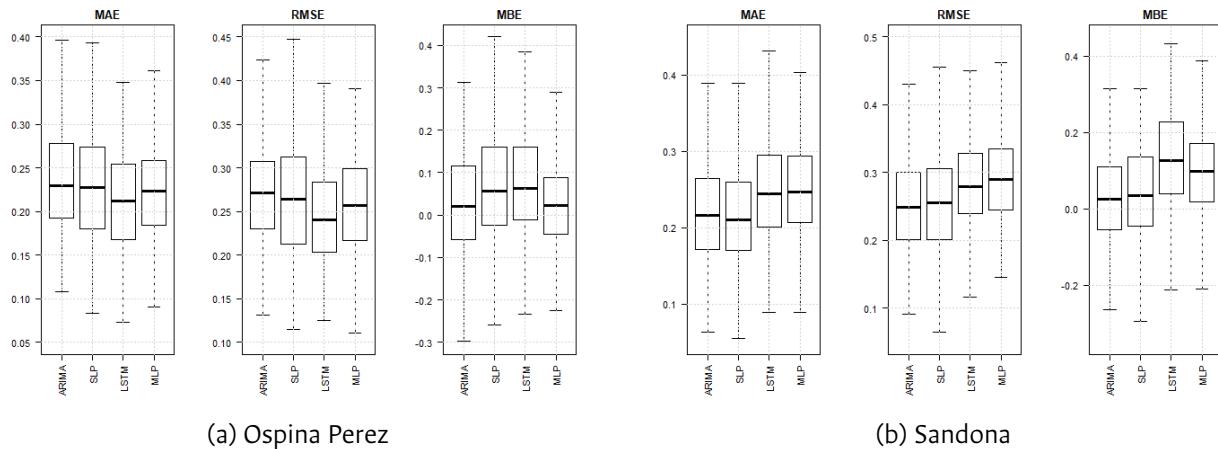
**Figura 5-21.** Resultados estadísticos**Figura 5-22.** Resultados estadísticos en Viento Libre**Figura 5-23.** Resultados estadísticos

conservadores.

**Figura 5-24.** Resultados estadísticos

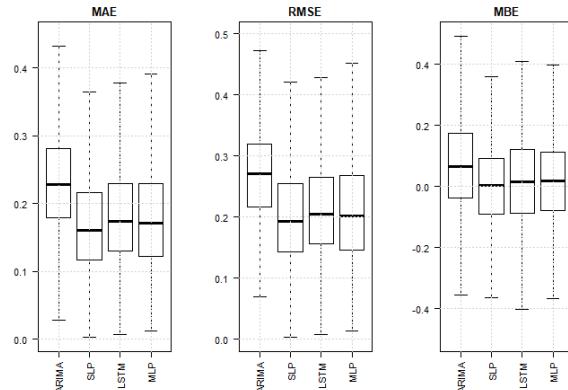
(a) La Josefina

(b) Paraiso

**Figura 5-25.** Resultados estadísticos

(a) Ospina Perez

(b) Sandona

**Figura 5-26.** Resultados estadísticos en Cerro Páramo

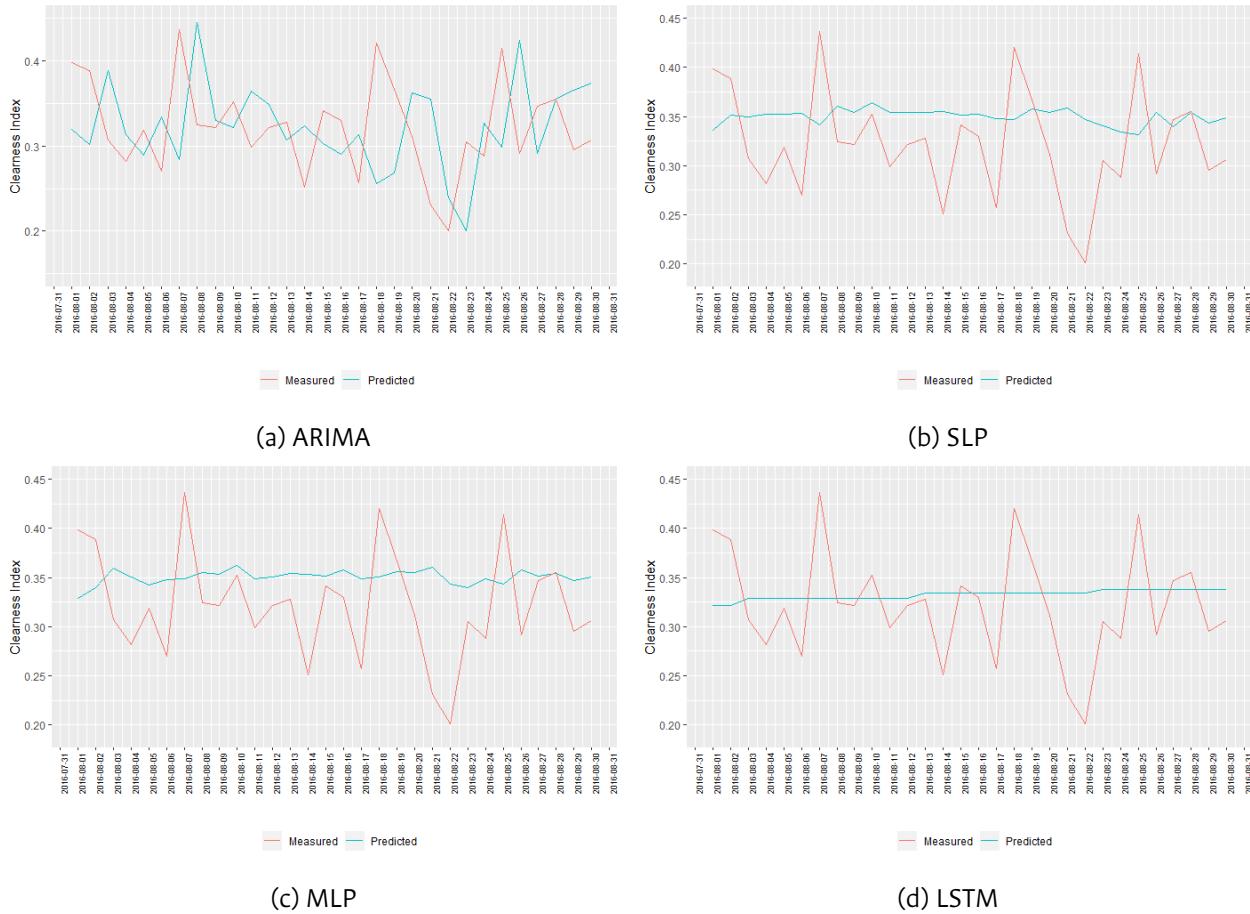
En La Josefina los resultados muestran que el modelo ARIMA tiene suficiente información de entrada para predecir la variabilidad del índice de claridad. Sin embargo, como en otros AWS, el modelo ARIMA

**Figura 5-27.** Resultados del pronóstico diario en Biotopo

pronostica los valores con un retraso de dos días. En los modelos basados en redes neuronales, SLP constituye la mejor opción. El LSTM es el modelo con baja variabilidad, consulte la Figura 5-34.

Paraiso presenta resultados similares a las demás AWS ubicados en la zona andina. ARIMA es el modelo con más robustez para describir la variación del índice de claridad. Sin embargo, los picos medidos y pronosticados tienen un retraso. Con respecto a los modelos basados en redes neuronales, el modelo SLP describe la variabilidad del índice de claridad mejor que MLP y LSTM, consulte la Figura 5-35.

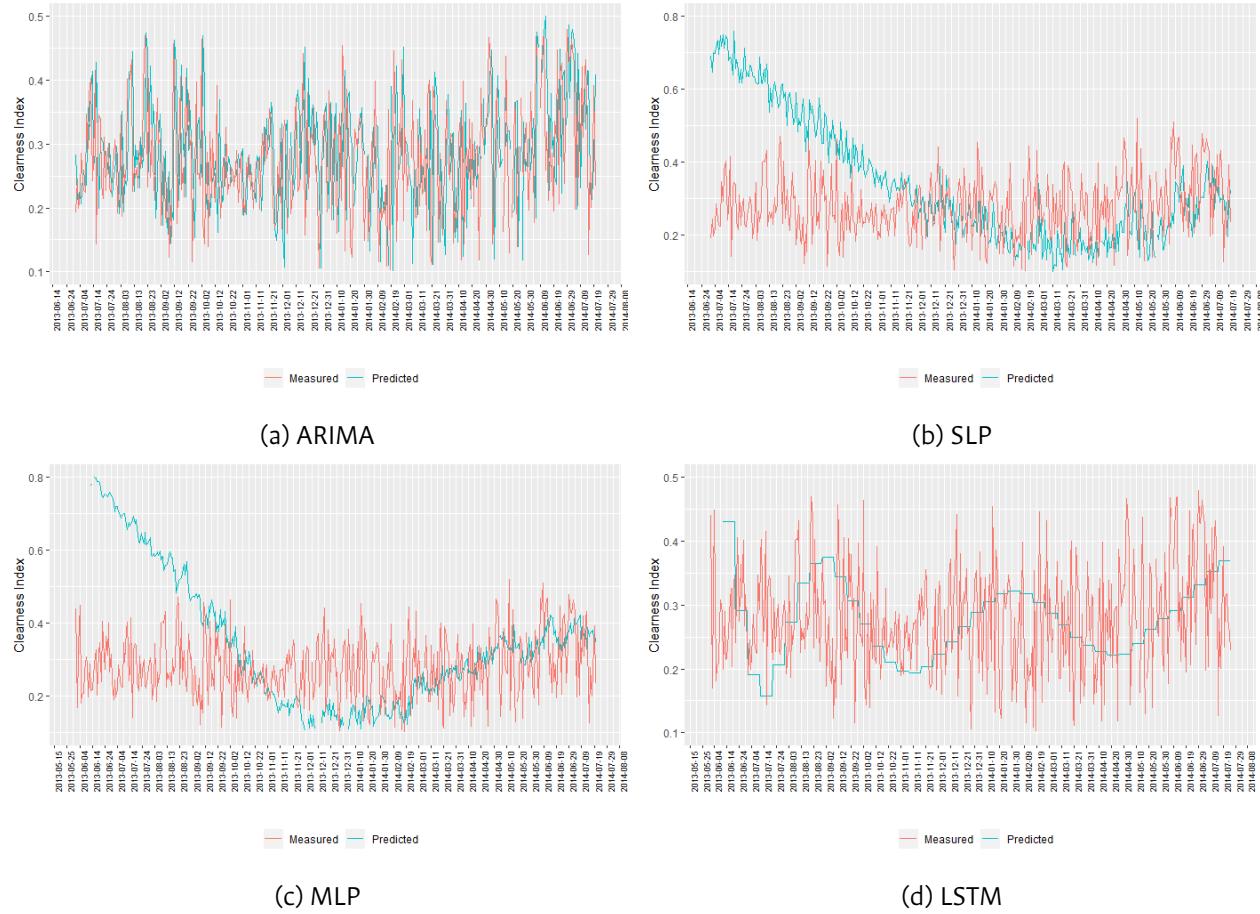
Ospina Perez y Sandona son AWS ubicadas en la zona andina. Estos AWS tienen un año de mediciones en promedio. Por lo tanto, la Figura 5-36 y la Figura 5-37 muestran la serie de tiempo completa. Ospina Perez tiene más datos imputados que Sandona. En ambos casos, el modelo ARIMA sigue mejor la variabilidad del índice de claridad. En los modelos basados en redes neuronales, SLP ajusta los pesos más rápidamente que MLP. Como resultado, SLP describe la serie de tiempo mejor que MLP. En el caso de LSTM, el modelo al inicio describe los principales cambios, y al final de la serie temporal, el modelo describe los valores futuros con la media de los valores pasados, como ocurrió en otros los otros AWS ubicados en la zona andina.

**Figura 5-28.** Resultados del pronóstico diario en Guapi

La Figura 5-38 presenta el pronóstico de la insolación solar global en un horizonte diario usando el modelo ARIMA, SLP, MLP y LSTM. Cerro Páramo es el único AWS analizado ubicado en la zona amazónica. En este caso, el modelo ARIMA responde a los cambios del índice de claridad. Sin embargo, como ocurre en los casos Andino y Pacífico, existe un retraso entre los valores medidos y pronosticados. Además, de los modelos basados en redes neuronales, SLP es la mejor opción seguida de los modelos MLP y LSTM.

El RMSE, MBE y MAPE fueron los errores estadísticos calculados para determinar la precisión de los modelos de pronóstico, ARIMA, SLP, MLP y LSTM. La Figura 5-39 presenta los resultados de los errores estadísticos para Biotopo y Guapi. En el caso de Biotopo, los resultados de RMSE muestran que MLP es el modelo con el error medio más bajo, seguido de LSTM y MLP. Además, el modelo LSTM presenta menos varianza que los otros modelos. MLP es el modelo con menos sesgo y varianza seguido por LSTM. El MAPE muestra que MLP es el modelo con menos error medio y LSTM el modelo con menos varianza. En el caso de Guapi la RSME presenta a los modelos basados en redes neuronales con menos error. SLP es el mejor modelo, seguido de MLP, LSTM y ARIMA. Los resultados de MBE muestran que

**Figura 5-29.** Resultados del pronóstico diario en Altaquer

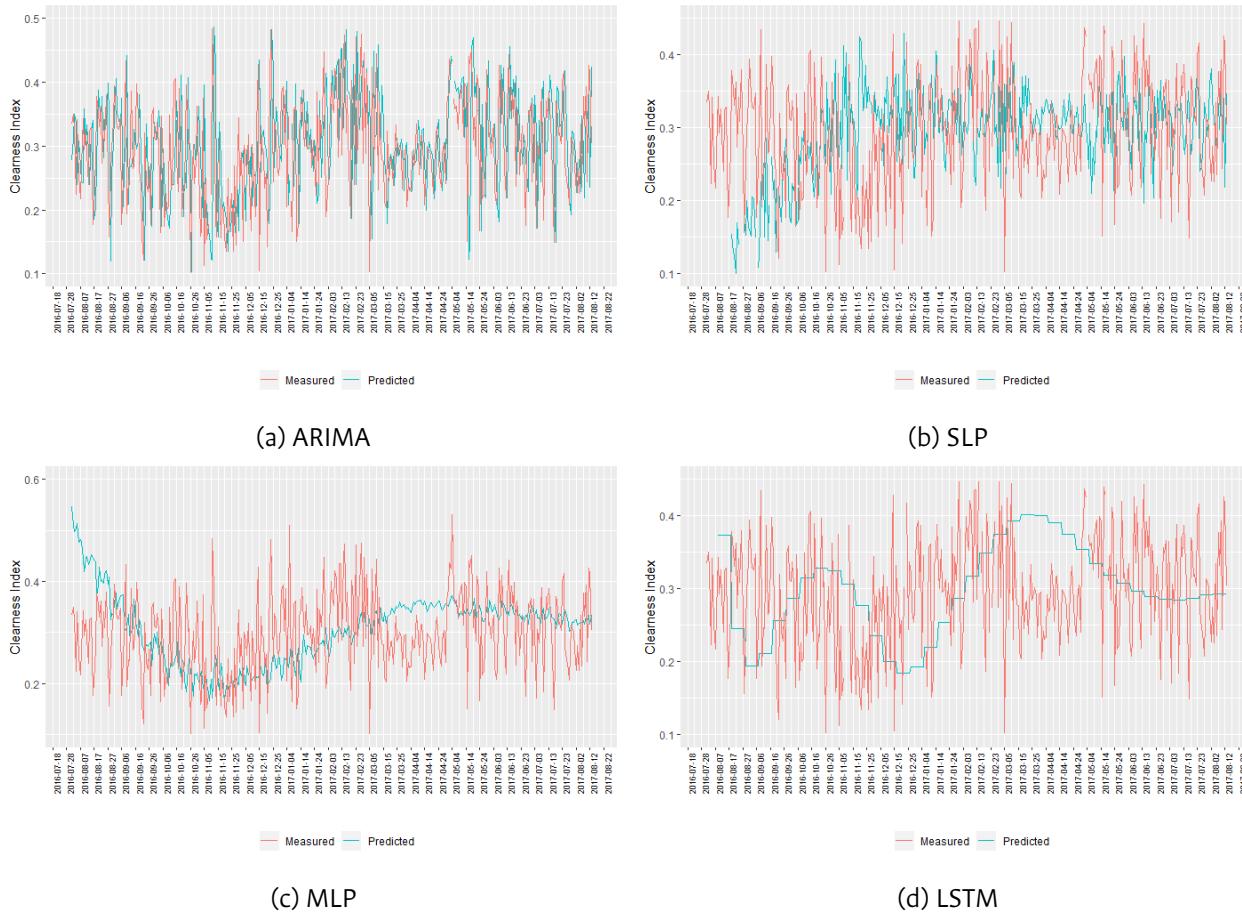


SLP y MLP tienen un sesgo similar. Además, el SLP y MLP tienen errores similares siendo los más bajos, seguidos de LSTM y ARIMA. Por lo tanto, los resultados del error estadístico presentan que los modelos basados en redes neuronales superan al modelo ARIMA en las AWS ubicadas en el Pacífico.

En Altaquer los errores estadísticos muestran que ARIMA y LSTM son las mejores opciones. RSME presenta LSTM como la mejor opción, seguida de ARIMA. En el caso de Altaquer, SLP tiene los valores de RMSE más altos. El resultado MBE de los modelos MLP y SLP tiene una mayor varianza debido a la sobreestimación en los valores pronosticados causada por los pesos de estos modelos al comienzo del período de pronóstico. El MAPE muestra a los modelos ARIMA y LSTM como las mejores opciones, mientras que SLP y MLP tienen un error medio y una varianza mayores. Por tanto, en este caso, los modelos ARIMA y LSTM son las mejores opciones. En el caso de Granja el Mira la diferencia de errores es menor que en el caso de Altaquer. Además, en este caso, RMSE, MBE y MAPE muestran que el modelo MLP es la mejor opción en esta AWS, seguido de LSTM, ARIMA y SLP, consulte la Figura 5-40.

En Viento Libre los errores estadísticos describen a MLP como el modelo con menor valor en RMSE y MBE, seguido de LSTM. Los resultados de MAPE muestran que los modelos basados en redes neuronales

**Figura 5-30.** Resultados del pronóstico diario en Granja el Mira



tienen un porcentaje de error medio menor. Sin embargo, ARIMA tiene menos varianza que los otros modelos, debido a la robustez de este modelo para describir la variación del índice de claridad, ver Figura 5-41.

En Universidad de Nariño y Botana las medidas de error estadístico presentan que los errores medios tienen valores similares con los cuatro modelos de pronóstico. Sin embargo, la varianza de los modelos basados en redes neuronales es menor que la varianza del modelo ARIMA, consulte la Figura 5-42. Una razón por la que este resultado podría ser el retraso entre los valores pronosticados y medidos observados en los gráficos de series de tiempo.

La Figura 5-43a muestra que los modelos basados en redes neuronales superan al modelo ARIMA en Paraiso. En el caso de La Josefina, los resultados de RMSE, MBE y MAPE presentan que LSTM y MLP son los mejores modelos.

La Figura 5-44a presenta las medidas de errores estadísticos de Ospina Perez. Es posible observar que en este caso, el modelo MLP tiene el mayor error debido a la baja cantidad de datos de entrada

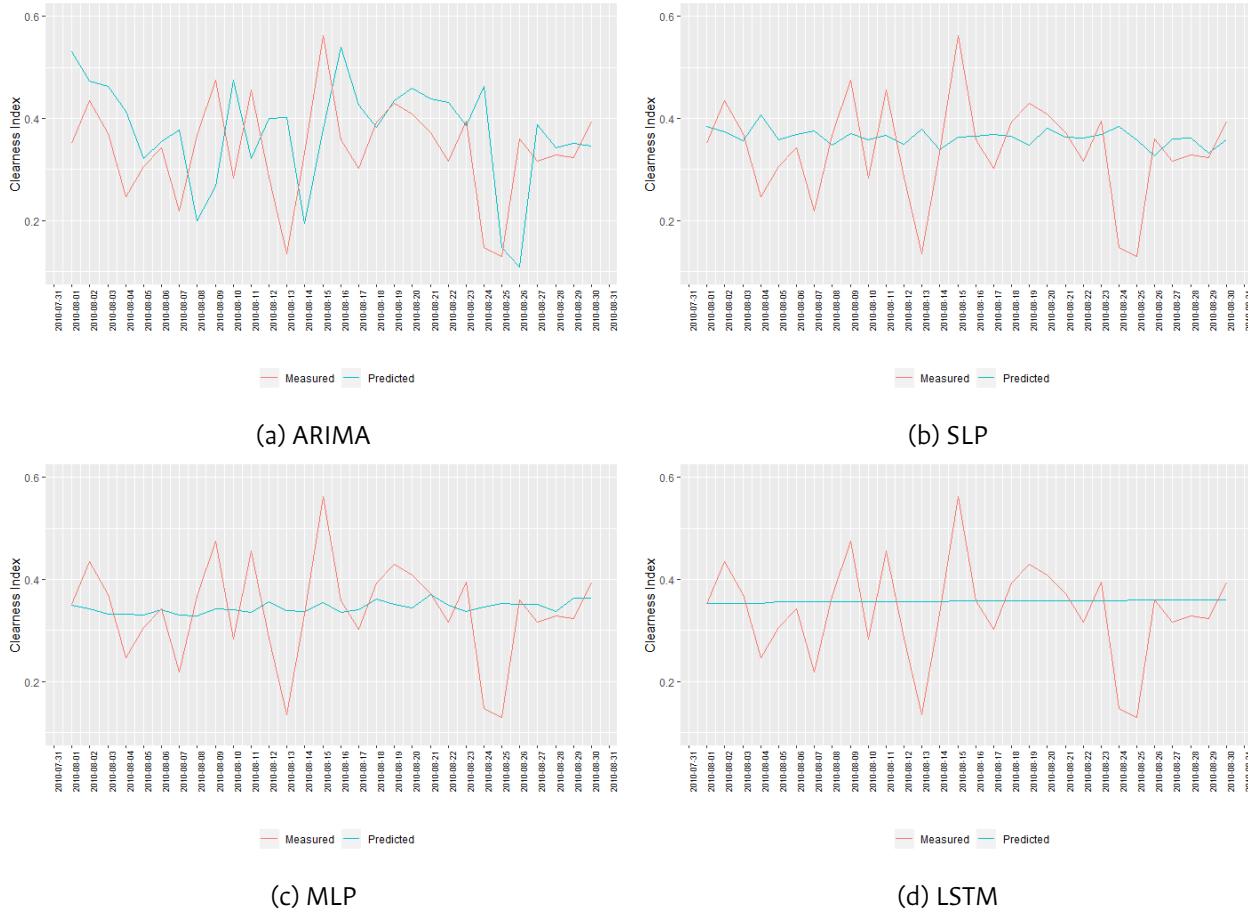
**Figura 5-31.** Resultados del pronóstico diario en Viento Libre



para entrenar el modelo y ajustar el peso de éste. Además, esta AWS tiene una gran cantidad de valores imputados que resultan en una baja cantidad de datos para el cálculo de medidas de errores estadísticos. En el caso de Sandona, MLP y LSTM presentan los errores medios más grandes en RMSE, MBE y MAPE. En este caso, SLP es la mejor opción porque el error medio y la varianza son menores.

En Cerro Páramo el resultado de las mediciones de errores estadísticos presenta a SLP como la mejor opción para el pronóstico de la insolación solar diaria debido a su baja varianza en RSME, MBE y MAPE. Además, el resultado de las mediciones de error estadístico presenta a LSTM como el modelo con el mayor error medio y varianza en RMSE y MBE. Sin embargo, el resultado de MAPE muestra que el modelo ARIMA tiene los mayores errores porcentuales medios.

**Figura 5-32.** Resultados del pronóstico diario en Universidad de Nariño

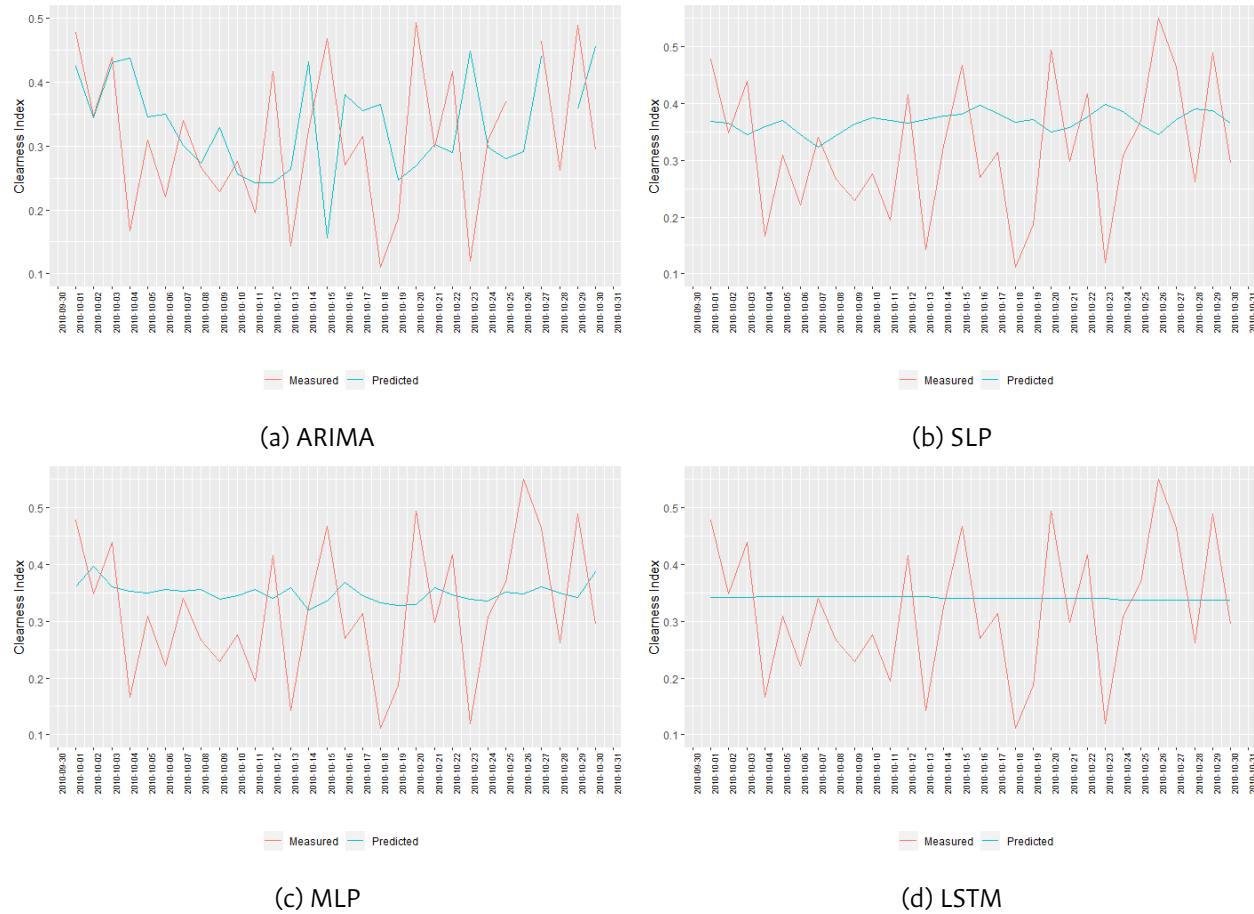


## 5.5 Conclusions

En este trabajo, los autores implementaron cuatro modelos de pronóstico: ARIMA, SLP, MLP y LSTM para pronosticar la irradiancia solar global con horizontes de un día por delante en una marca de tiempo por hora e insolación solar global con horizontes de un día por delante en una marca de tiempo diaria. La información utilizada provino de doce AWS ubicados en el departamento de Nariño en Colombia. Uno de los desafíos fue lidiar con los datos faltantes para implementar los modelos de pronóstico.

El proceso de imputación en todos los casos imputó más del 50 % de los valores de la serie temporal, el mejor caso fue Granja el Mira con 54,1 %, y el peor caso fue La Josefina con 84,2 %. El resultado anterior mostró que se completó la mayor parte de los datos de la serie temporal. Los resultados del error estadístico muestran que en promedio los datos imputados subestiman el recurso en todos los territorios. Los resultados de RMSE y MAE presentan que en la zona andina el error es mayor que en la zona del Pacífico. Para reducir los errores estadísticos es conveniente realizar más estudios analizando otras técnicas de imputación.

**Figura 5-33.** Resultados del pronóstico diario en Botana



Las medidas de error de los modelos de pronóstico, presentan que los modelos basados en redes neuronales superan al modelo ARIMA en el horizonte horario. En el AWS que tiene series de tiempo más largas, el SLP fue el mejor modelo seguido por MLP. En el AWS con más corto, el LSTM fue el mejor modelo. Además, en la zona del Pacífico en ambientes nublados, el modelo LSTM supera a los demás. Es un resultado valioso de la variabilidad en la irradiancia solar inducida por el movimiento de las nubes que se describe mejor con este modelo. Por lo tanto, en un ambiente nublado, el LSTM es el modelo recomendado para pronosticar la irradiancia solar global.

En la zona andina, los modelos de pronóstico basados en redes neuronales presentan mejores resultados en la medición de errores MAE y RMSE cuando la serie temporal es más larga, es decir, en Viento Libre, Botana, La Josefina, y casos Paraiso. La medición del error MBE presenta al SLP como el modelo con menos sesgo. A diferencia de la zona del Pacífico, en esta zona el LSTM fue el modelo con más sesgo, independientemente del nivel de nubosidad. En Ospina Perez y Sandona, que son series de tiempo más cortas, SLP es mejor en ambientes nublados. En la zona amazónica, la AWS Cerro Páramo muestra que SLP es el mejor modelo en todas las medidas de error. Además, la gráfica de series de tiempo per-

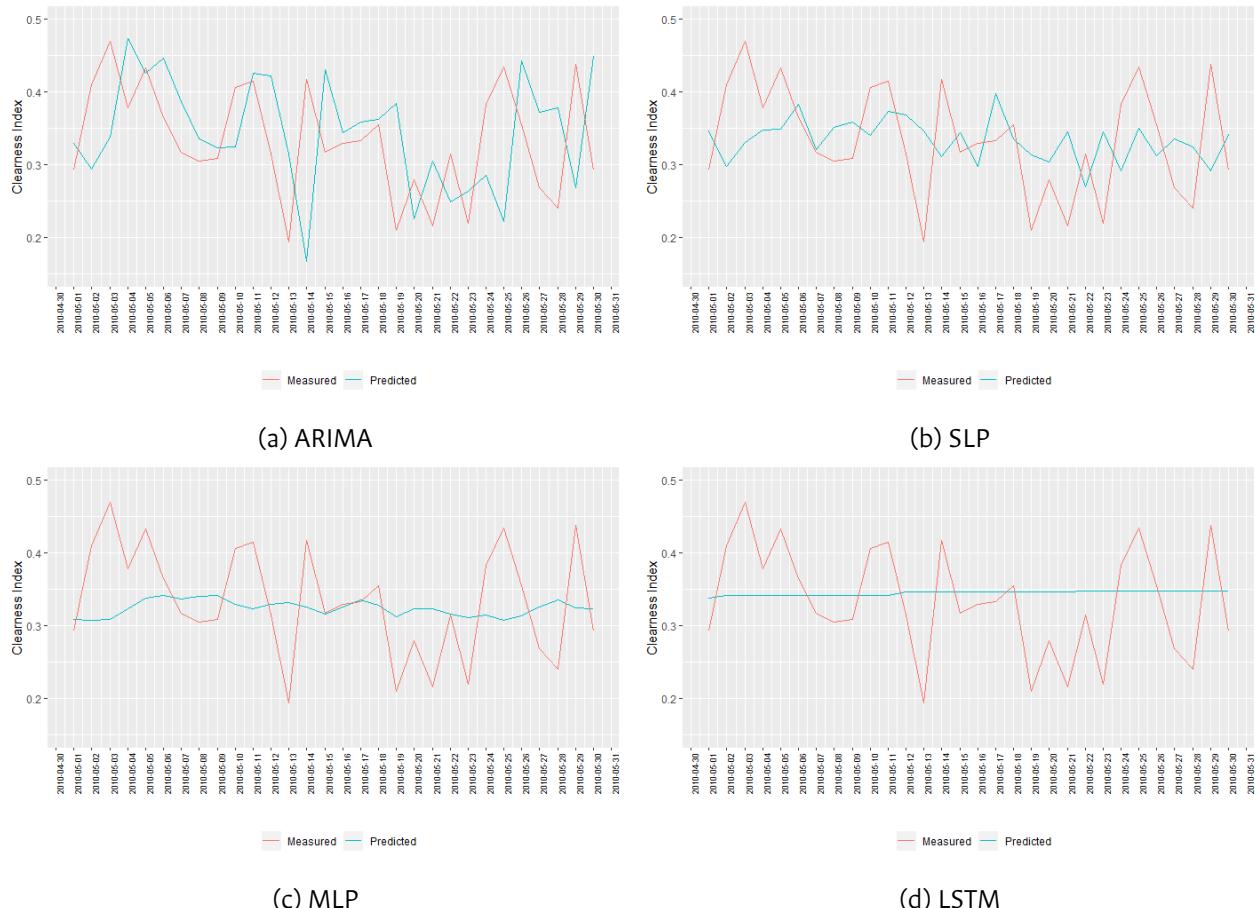
**Figura 5-34.** Resultados del pronóstico diario en La Josefina

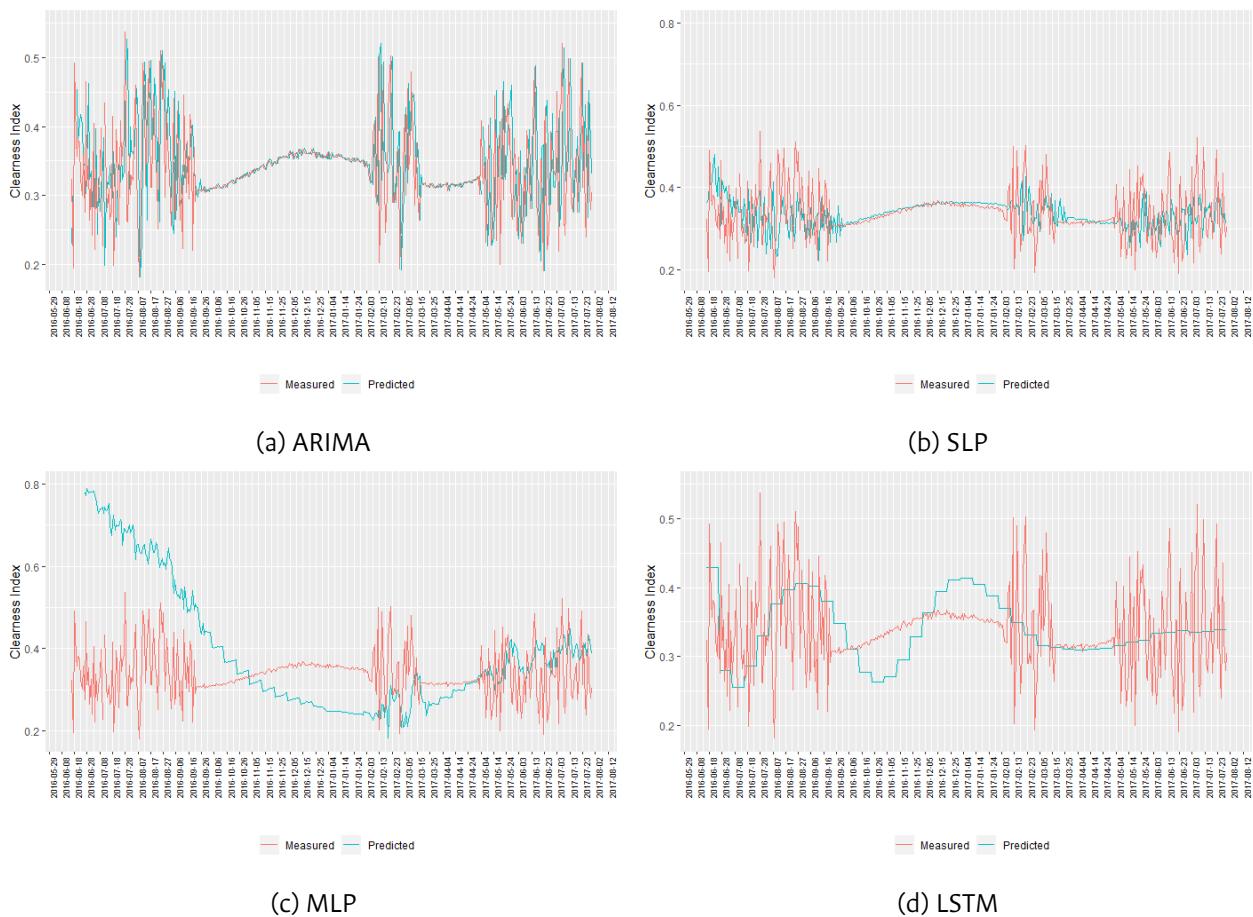


mite concluir que el modelo ARIMA predice los valores futuros siguiendo una tendencia media, más que describir la variabilidad de la serie de tiempo. En general, el modelo SLP supera al otro modelo en irradiancia solar global en un horizonte de un día por delante con una frecuencia horaria.

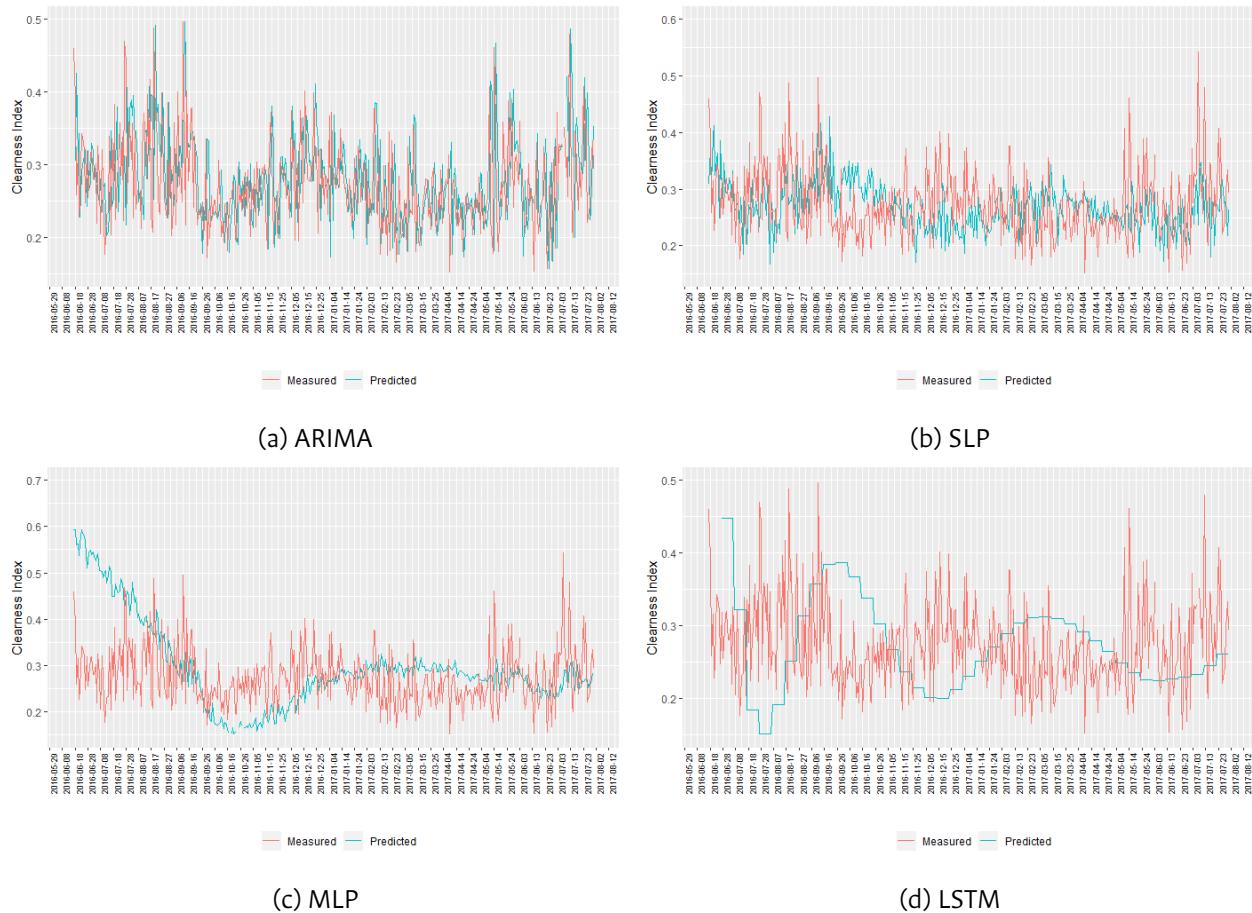
Las mediciones de error estadístico del modelo de pronóstico de insolación solar global en una marca de tiempo diaria muestran que en el AWS ubicado en la zona del Pacífico con más de un año de medidas, el error medio es similar con todos los modelos de pronóstico. Sin embargo, el modelo ARIMA tiene una mayor variación. Mientras que en la AWS ubicada en la zona del Pacífico con un año de medidas en promedio, el modelo SLP presenta los mayores errores debido a los bajos datos de entrada de entrenamiento. En general, en el AWS ubicado en la zona andina con más de un año de mediciones, los modelos basados en redes neuronales tienen un mejor desempeño que el modelo ARIMA. Sin embargo, en el AWS con un año de medidas en promedio ubicadas en la zona andina, el modelo MLP presenta los mayores errores. Por tanto, en la zona andina, el modelo basado en redes neuronales presenta menores errores estadísticos. Sin embargo, el rendimiento del modelo depende de la cantidad de datos de entrada de entrenamiento.

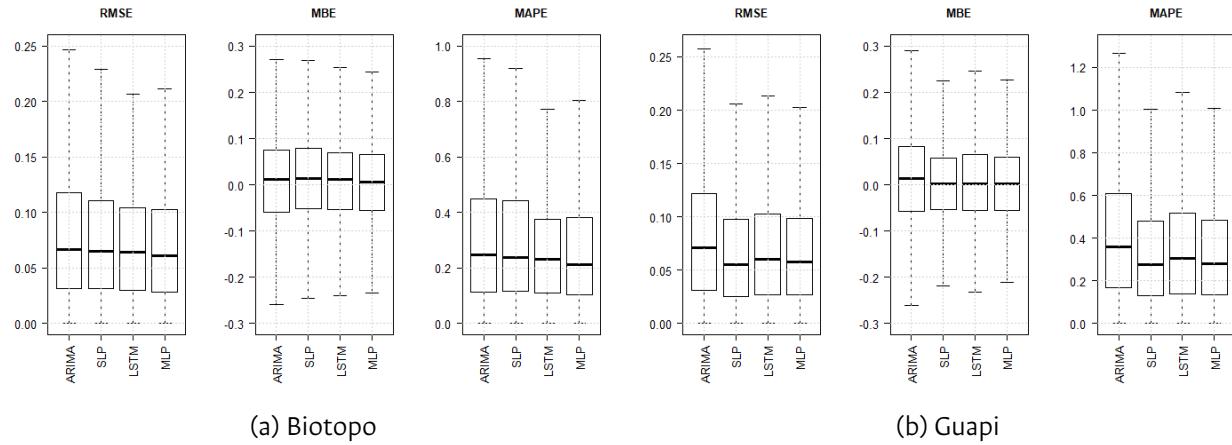
**Figura 5-35.** Resultados del pronóstico diario en Paraiso

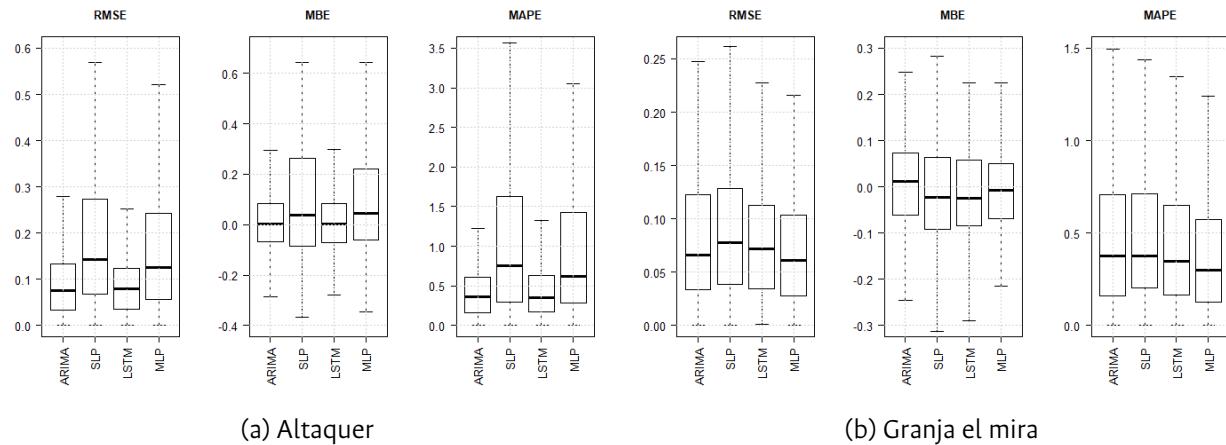
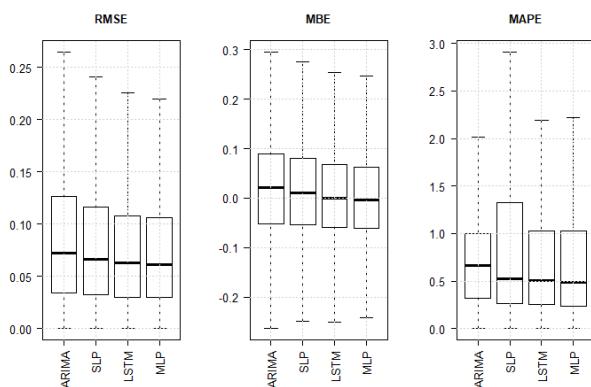
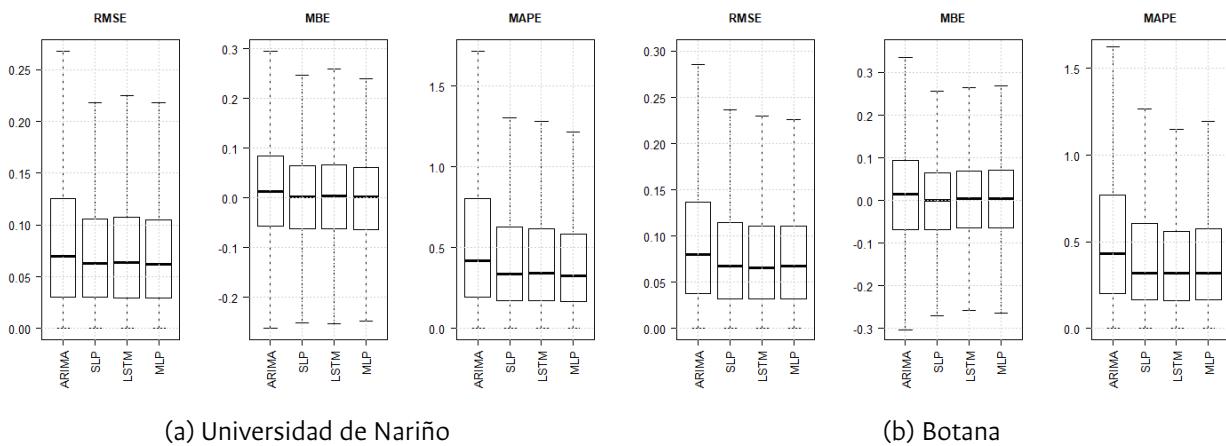


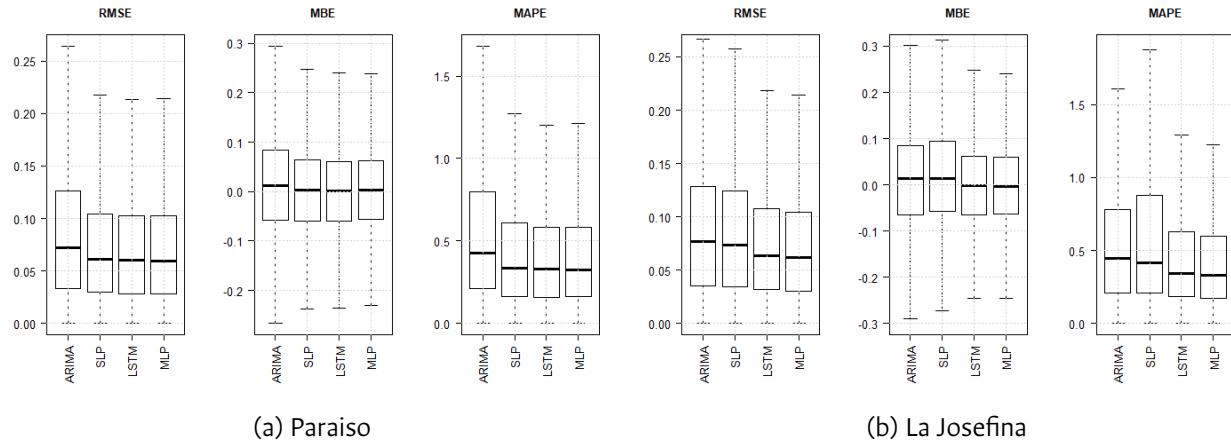
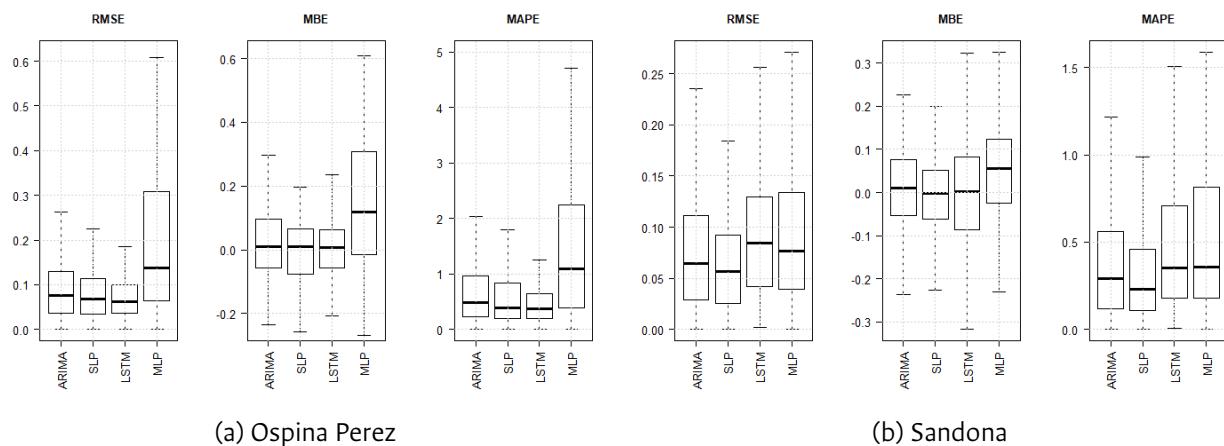
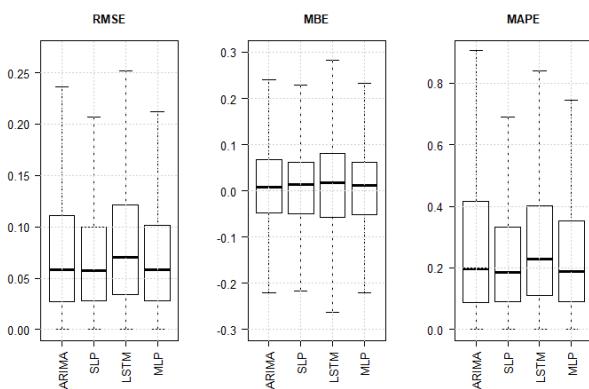
**Figura 5-36.** Resultados del pronóstico diario en Ospina Perez

**Figura 5-37.** Resultados del pronóstico diario en Sandona



**Figura 5-38.** Resultados del pronóstico diario en Cerro Páramo**Figura 5-39.** Error estadístico diario

**Figura 5-40.** Error estadístico diario**Figura 5-41.** Error estadístico diario en Viento Libre**Figura 5-42.** Error estadístico diario

**Figura 5-43.** Error estadístico diario**Figura 5-44.** Errores estadísticos diarios**Figura 5-45.** Error estadístico diario en Cerro Páramo

## 6 Conclusiones de la tesis

Esta tesis propuso una metodología para la selección de proyectos y su implementación en una comunidad. Además, la metodología de priorización evalúa áreas relevantes y el costo de implementación de proyectos energéticos basados en energía solar. Los criterios elegidos para la evaluación comunitaria consideran el SDGI y el IDH y los resultados del VPN fueron los criterios para la evaluación del proyecto. Asimismo, la priorización presenta una combinación de esos índices, que fueron los criterios utilizados en la metodología AHP para dar un orden de ejecución. Además, la guía de trabajo comunitario enfatizó la integración de la comunidad en el diseño y operación de la solución implementada, otorgando al proyecto una mayor probabilidad de éxito. En este caso, los autores consideraron las características de Nariño e identificaron áreas relevantes para la evaluación comunitaria por parte de los municipios. Las dimensiones elegidas fueron educación, salud, calidad de vida, economía y seguridad.

Las alternativas evaluadas con esta metodología fueron cuatro proyectos, junto con las comunidades involucradas. Los resultados muestran que en cuatro de seis objetivos de priorización, la electrificación de escuelas es la prioridad, y en dos objetivos, la prioridad fue la electrificación de un resguardo indígena. Dados los resultados anteriores, los autores decidieron realizar un análisis más detallado del proyecto de electrificación de las escuelas. El análisis mostró que si la TIR es una variable de decisión, es mejor implementar proyectos más grandes; en este caso, más de diez. Por otro lado, los resultados mostraron una correlación entre educación y calidad de vida porque O3 y O6 tienen seis proyectos en el mismo orden de ejecución; por tanto, la realización de proyectos educativos mejora la calidad de vida y viceversa. En este sentido, sería conveniente realizar más estudios que confirmen este hallazgo y así unificar esfuerzos en las políticas públicas de estos sectores para fortalecer los resultados.

Los investigadores esperan que, con la implementación de esta metodología de priorización, sea posible identificar el orden óptimo de ejecución de los proyectos sociales de acuerdo con un objetivo determinado. Adicionalmente, con la directriz, el objetivo es hacer de la comunidad parte activa de la solución, mejorar la sostenibilidad del proyecto y evitar o reducir la dependencia de una entidad externa, y así empoderar a la comunidad y promover su desarrollo social, económico, educación y cultural, entre otros aspectos. También es esencial resaltar el papel principal de la junta reguladora local en el éxito del proyecto. Por tanto, las investigaciones futuras deben analizar en profundidad aspectos como la regulación, los mecanismos de elección de los miembros del consejo, entre otros.

Los niveles de validación de los datos de irradiancia solar global tienen una fuerte influencia en los resultados de las variables empíricas. Considerando toda la información registrada, solo un 60,89 %

superá los pasos de validación obligatorios propuestos. De ese valor, 95,81 %, que corresponde al número de días con al menos seis valores medidos, fue la información utilizada en la calibración de los modelos empíricos. Sin embargo, este porcentaje representa en promedio el 33,90 % de la información total registrada en el AWS. Además, los días con información completa apenas llegan al 1,26 %. Este resultado indica la calidad de la serie temporal y la necesidad de mejorar y aumentar los procedimientos de mantenimiento y calibración.

En el departamento de Nariño, el desempeño de la AWS es un factor determinante, considerando el predominio de días parcialmente nublados que representan el 64,7 % de los días. En otras palabras, en Nariño existe una alta interacción de nubes que dificulta la estimación de la insolación solar; por tanto, aumenta la necesidad de medidas fiables.

En cuanto a las medidas de temperatura, de la cantidad de datos que superan los pasos de validación horaria, se tomó el 92,78 % para la calibración e imputación empírica. Además, el número de mediciones de temperatura diurnas presenta picos en once y doce valores diarios; esto significa que la mayoría de los días que se tardan en modelar y llenar la base de datos por el proceso de imputación, tienen el 88,46 % de la información total en promedio.

El modelo propuesto mostró una relación lineal entre los coeficientes empíricos contra la altitud del sitio de estudio. Los coeficientes empíricos se clasificaron entre aquellos por encima y por debajo de 2.500 MSSL. El ajuste lineal para casos por debajo de 2.500 MSSL muestra un  $R^2$  de 0.5995 y 0.5262 para a y b respectivamente. El ajuste lineal para casos por encima de 2.500 msnm presenta un  $R^2$  de 0,8182 y 0,6069 para a y b respectivamente. Este resultado es consistente con el comportamiento de la temperatura en zonas tropicales y la irradiancia solar global en grandes altitudes.

Cuando se consideraron RMSE, SD, MAE,  $U_{95}$  y MAPE, el modelo propuesto tuvo un mejor desempeño en cinco de los ocho casos evaluados. Estos casos se encuentran en la zona Andina y Amazónica, con altitudes superiores a los 2.500 msnm. Como resultado, el modelo propuesto ha utilizado como entrada la información en la AWS Andina y Amazónica. Para el AWS ubicado en las zonas del Pacífico, el modelo HS fue la mejor opción, seguido del modelo propuesto en esta investigación.

En la zona del Pacífico, HS es el modelo aplicado para estimar la insolación solar. Sin embargo, los resultados del error estadístico muestran que el recurso está subestimado en 442,78 [ $Wh/m^2da$ ] y 333,11 [ $Wh/m^2da$ ] para Granja el Mira y Altaquer. En consecuencia, para mejorar los resultados de esta zona, es necesario aumentar la densidad de la red de estaciones meteorológicas.

El modelo propuesto fue utilizado para estimar la insolación solar en la zona andina. La comparación entre los valores medidos y estimados mostró una sobreestimación. Sin embargo, respecto al modelo HS, la estimación se reduce en un 51,81 % y un 18,55 % para Ospina Pérez y Sandona, respectivamente. El modelo propuesto tiene un mejor desempeño que el modelo HS en altitudes superiores a 2.500 msnm.

Al comparar IDW con las técnicas de Kriging, OK presentó una mejora en los errores estadísticos de la

siguiente manera: RMSE 11,71 %, MAE 3,98 %, MBE 100,92 %, SD 11,40 % y MPE 46,58 %. Aunque SK mostró mejores resultados en MAE, SD y RMSE que OK, el sesgo con OK es menor. Por lo tanto, OK fue la técnica seleccionada para mapear el potencial de insolación solar. UK fue la técnica que presentó menores mejoras frente al IDW, incluso en MAE, que redujo el 6,78 % de la precisión.

De los mapas se puede concluir que la estimación obtenida es menor a la presentada en los mapas de radiación solar del IDEAM en todo el territorio de Nariño. También es posible determinar la existencia de una zona con un alto potencial de generación eléctrica con plantas de energía solar fotovoltaica.

Las mediciones de errores de los modelos de pronóstico, presentan que los modelos basados en redes neuronales superan al modelo ARIMA. En las AWS que tienen series de tiempo más largas, el SLP fue el mejor modelo seguido por MLP. En las AWS con series de tiempo más cortas el LSTM fue el mejor modelo. Además, en la zona del Pacífico en entornos nublados, el modelo LSTM supera a los demás. Es un resultado valioso debido a que la variabilidad en la irradiancia solar inducida por el movimiento de las nubes se describe mejor con estos modelos. Por lo tanto, en un ambiente nublado, el LSTM es el modelo recomendado para pronosticar la irradiancia solar global.

En la zona andina, basados en la medición de errores MAE y RMSE los modelos de pronóstico basados en redes neuronales presentan mejores resultados cuando la serie temporal es más larga, es decir, en Viento Libre, Botana, La Josefina, y casos Paraiso. MBE presenta al modelo SLP como el modelo con menos sesgo. A diferencia de la zona del Pacífico, en esta zona LSTM fue el modelo con más sesgo, independientemente del nivel de nubosidad. En Ospina Perez y Sandona, que son series de tiempo más cortas, SLP es mejor en ambientes nublados. En la zona amazónica, la AWS Cerro Páramo muestra que SLP es el mejor modelo en todas las medidas de error. Además, el gráfico de series de tiempo permite concluir que el modelo ARIMA predice los valores futuros siguiendo una tendencia media, más que describiendo la variabilidad de la serie de tiempo. En general, el modelo SLP supera al otro modelo en irradiancia solar global en un horizonte de un día por delante con una frecuencia horaria.

## 6.1 Recomendaciones

Es conveniente incluir más dimensiones para comprender de mejor manera las condiciones de la población e incluir aspectos culturales de las comunidades. Para la evaluación del proyecto, los autores consideran el VPN, dando más importancia a los de menor costo. Sin embargo, este único indicador no permite el reconocimiento de los beneficios sociales que puedan derivarse del proyecto. De ahí que, en trabajos futuros, el uso de cuantificadores adicionales que midan el beneficio económico que surgió de los proyectos sociales podría motivar la atención de potenciales inversores.

Debido a los resultados del control de calidad, es importante establecer un plan para realizar procedimientos de calibración y mantenimiento de las estaciones meteorológicas con regularidad y seguir estándares de alta calidad y ampliamente aceptados. También fue notoria la necesidad de instalar más

AWS para aumentar los puntos de muestreo.

El modelo empírico propuesto tuvo un buen desempeño en este ambiente tropical y montañoso. Sin embargo, es necesario analizar más información proveniente de otros lugares con las mismas características. Para lograr esto, el requisito principal es aumentar el número de AWS y la calidad de las series de tiempo en los entornos tropicales y montañosos.

# References

- Abdullah, L., & Najib, L. (2016). Sustainable energy planning decision using the intuitionistic fuzzy analytic hierarchy process: choosing energy technology in Malaysia. *International Journal of Sustainable Energy*, 35(4), 360–377. Retrieved from <https://doi.org/10.1080/14786451.2014.907292> doi: 10.1080/14786451.2014.907292
- Abreu, E. F., Canhoto, P., Prior, V., & Melicio, R. (2018). Solar resource assessment through long-term statistical analysis and typical data generation with different time resolutions using GHI measurements. *Renewable Energy*, 127, 398–411. Retrieved from <https://doi.org/10.1016/j.renene.2018.04.068> doi: 10.1016/j.renene.2018.04.068
- AENOR. (2004). *Redes de estaciones meteorológicas automáticas: directrices para la validación de registros meteorológicos procedentes de redes de estaciones automáticas. Validación en tiempo real.*
- Agami Reddy, T. (2011). *Applied Data Analysis and Modelling for Energy Engineers and Scientists*. Springer London. doi: 10.1007/978-1-4419-9613-8
- Akinoglu, B. (2008). Recent Advances in the Relations between Bright Sunshine Hours and Solar Irradiation. In *Modeling solar radiation at the earth's surface* (pp. 115–143). Springer. doi: doi.org/10.1007/978-3-540-77455-6\}5
- Allen, R. G. (1997). Self-Calibrating Method for Estimating Solar Radiation From Air Temperature. *Journal of Hydrologic Engineering*, 2(250), 56–67.
- Almorox, J., Hontoria, C., & Benito, M. (2011). Models for obtaining daily global solar radiation with measured air temperature data in Madrid (Spain). *Applied Energy*. doi: 10.1016/j.apenergy.2010.11.003
- Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de Pison, F. J., & Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. *Solar Energy*, 136, 78–111. Retrieved from <http://dx.doi.org/10.1016/j.solener.2016.06.069> doi: 10.1016/j.solener.2016.06.069
- Arbeláez-Arias, F.-A. (2006). *Desarrollo sostenible y sus indicadores* (Tech. Rep.). Cali: Centro de Investigaciones y Documentación Socioeconómica. Retrieved from <http://bibliotecavirtual.clacso.org.ar/Colombia/cidse-univalle/20121116025351/Doc93>.

- Arbeláez Pérez, O. A. (2019). *Informe mensual de localidades sin telemetría de las ZNI* (Tech. Rep.). Centro Nacional de Monitoreo.
- Aslani, A. (2014). Private sector investment in renewable energy utilisation: Strategic analysis of stakeholder perspectives in developing countries. *International Journal of Sustainable Energy*, 33(1), 112–124. doi: 10.1080/14786451.2012.751916
- Ávila, A. F., Escobar, E., & Torres Tobar, C. (2014). DEPARTAMENTO DE NARIÑO (Tech. Rep.). Fundación Paz y Reconciliación; Redprodepaz.
- Aznar, J., & Guijarro, F. (2012). *Nuevos métodos de valoración: modelos multicriterio*. Retrieved from <http://medcontent.metapress.com/index/A65RM03P4874243N.pdf>
- Badosa, J., Gobert, E., Grangereau, M., & Kim, D. (2017). Day-Ahead Probabilistic Forecast of Solar Irradiance: A Stochastic Differential Equation Approach. In P. Drobinski, M. Mogeot, D. Picard, R. Plougonven, & P. Tankov (Eds.), *Mathematics & statistics* (p. 22). Paris: Springer. Retrieved from <http://www.springer.com/series/10533>
- Bakirci, K. (2009). *Models of solar radiation with hours of bright sunshine: A review*. doi: 10.1016/j.rser.2009.07.011
- Balbás Egea, J. I., & Eguren Egiguren, J. A. (2019). Bases for a sustainable energy model. Case study: Basque autonomous community. *International Journal of Sustainable Energy*, 38(9), 884–903. doi: 10.1080/14786451.2019.1609474
- Benali, L., Notton, G., Fouilloy, A., Voyant, C., & Dizene, R. (2019). Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renewable Energy*, 132, 871–884. doi: 10.1016/j.renene.2018.08.044
- Benson, R. B., Paris, M. V., Sherry, J. E., & Justus, C. G. (1984). Estimation of daily and monthly direct, diffuse and global solar radiation from sunshine duration measurements. *Solar Energy*, 32(4), 523–535. doi: doi.org/10.1016/0038-092X(84)90267-6
- Berrar, D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1-3, 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X
- Bertelsmann Stiftung, & Sustainable Development Solutions Network (SDSN). (2016). *Índice y paneles de los ODS. Informe global* (Tech. Rep.). Retrieved from <http://sdgindex.org/assets/files/SDG-Index-ES-02.pdf>
- Besharat, F., Dehghan, A. A., & Faghih, A. R. (2013). *Empirical models for estimating global solar radiation: A review and case study*. doi: 10.1016/j.rser.2012.12.043
- Bhattacharjee, S., Ghosh, S. K., & Chen, J. (2019). *Semantic Kriging for Spatio-temporal Prediction* (Vol. 839). Springer. doi: doi.org/10.1007/978-981-13-8664-0

- Blaga, R., Sabadus, A., Stefu, N., Dughir, C., Paulescu, M., & Badescu, V. (2019). A current perspective on the accuracy of incoming solar energy forecasting. *Progress in Energy and Combustion Science*, 70, 119–144. Retrieved from <https://doi.org/10.1016/j.pecs.2018.10.003> doi: 10.1016/j.pecs.2018.10.003
- Blumthaler, M. (2012). Solar Radiation of the High Alps. In C. Lütz (Ed.), *Plants in alpine regions cell physiology of adaptation and survival strategies* (pp. 11–20). Springer Wien New York. doi: 10.1007/978-3-7091-0136-0
- Boland, J. (2008). Time Series Modelling of Solar Radiation. In V. Badescu (Ed.), *Modeling solar radiation at the earth surface* (pp. 283–311). Retrieved from [http://link.springer.com/10.1007/978-1-4471-4649-0\\_5](http://link.springer.com/10.1007/978-1-4471-4649-0_5) doi: 10.1007/978-1-4471-4649-0\}5
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis : forecasting and control*.
- Bristow, K. L., & Campbell, G. S. (1984). On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology*, 31(2), 159–166.
- Camblong, H., Sarr, J., Niang, A. T., Curea, O., Alzola, J. A., Sylla, E. H., & Santos, M. (2009). Microgrids project, Part 1: Analysis of rural electrification with high content of renewable energy sources in Senegal. *Renewable Energy*, 34(10), 2141–2150. Retrieved from <http://dx.doi.org/10.1016/j.renene.2009.01.015> doi: 10.1016/j.renene.2009.01.015
- Casella, G., & Berger, R. L. (2002). *Statistical Inference* (Second ed.). Thomson.
- Chandola, D., Gupta, H., Tikkiwal, V. A., & Bohra, M. K. (2020). Multi-step ahead forecasting of global solar radiation for arid zones using deep learning. *Procedia Computer Science*, 167(Iccids 2019), 626–635. Retrieved from <https://doi.org/10.1016/j.procs.2020.03.329> doi: 10.1016/j.procs.2020.03.329
- Chen, J.-l., Liu, H.-b., Wu, W., & Xie, D.-t. (2011). Estimation of monthly solar radiation from measured temperatures using support vector machines - A case study. *Renewable Energy*, 36(1), 413–420. Retrieved from <http://dx.doi.org/10.1016/j.renene.2010.06.024> doi: 10.1016/j.renene.2010.06.024
- Coimbra, C. F., Kleissl, J., & Marquez, R. (2013). Overview of Solar-Forecasting Methods and a Metric for Accuracy Evaluation. In J. Kleissl (Ed.), *Solar energy forecasting and resource assessment* (First ed., chap. Chapter 8). Elsevier. Retrieved from [https://books.google.com.co/books?hl=en&lr=&id=94KI0\\_SPwW8C&oi=fnd&pg=PP1&dq=Solar+energy+forecasting+and+resource+assessment&ots=HcVnQHR7Mt&sig=XuSTldnWP5MKAP8J3YUH1BbfFCM](https://books.google.com.co/books?hl=en&lr=&id=94KI0_SPwW8C&oi=fnd&pg=PP1&dq=Solar+energy+forecasting+and+resource+assessment&ots=HcVnQHR7Mt&sig=XuSTldnWP5MKAP8J3YUH1BbfFCM)
- CORPONARIÑO. (2001). *Plan De Gestión Ambiental Regional 2002 - 2012* (Tech. Rep.). San Juan de Pasto: Corponariño. Retrieved from <http://corponarino.gov.co/expedientes/pgar20022012/pgar2002-2012.pdf>

- Dai, K. Y., Liu, G. R., Lim, K. M., & Gu, Y. T. (2003). Comparison between the radial point interpolation and the Kriging interpolation used in meshfree methods. , 32, 60–70. doi: 10.1007/s00466-003-0462-z
- DANE, & Banco de la República de Colombia. (2016). *Coyuntura económica regional*.
- Dannecker, L. (2015). *Energy Time Series Forecasting*. Springer Vieweg. doi: 10.1007/978-3-658-11039-0
- Dawoud, F., Jbour, A., Al-salaymeh, A., Qoaider, L., & Fink, T. (2019). Innovative solutions for Renewable Energy and Energy Efficiency in Jordan. , 20(4), 201–216.
- Demirhan, H., & Renwick, Z. (2018). Missing value imputation for short to mid-term horizontal solar irradiance data. *Applied Energy*, 225(March), 998–1012. Retrieved from <https://doi.org/10.1016/j.apenergy.2018.05.054> doi: 10.1016/j.apenergy.2018.05.054
- Departamento Administrativo Nacional de Estadística - DANE. (2005). *Déficit de vivienda*. Retrieved from <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/deficit-de-vivienda>
- Departamento Administrativo Nacional de Estadística - DANE. (2009). *Metodología Déficit de Vivienda* (Tech. Rep.). Bogotá: Departamento Administrativo Nacional de Estadística.
- Departamento Administrativo Nacional de Estadística - DANE. (2016). *Valor Agregado según ramas de actividad económica y PIB (Clasificación Cuentas Nacionales) Serie 2000 - 2014p, Base 2005\**. Retrieved from <http://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-trimestrales#pib-por-rama-de-actividad>
- Departamento Administrativo Nacional de Estadística - DANE. (2017). *Indicador de importancia económica municipal*. Retrieved from <https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-departamentales/indicador-de-importancia-economica-municipal>
- Departamento Administrativo Nacional de Estadística - DANE. (2018a). *Encuesta de la calidad de vida (ECV)*. Retrieved from <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/calidad-de-vida-ecv>
- Departamento Administrativo Nacional de Estadística - DANE. (2018b). *Estadísticas vitales nacimientos y defunciones*. Retrieved from <https://www.dane.gov.co/index.php/estadisticas-por-tema/salud/nacimientos-y-defunciones>
- Departamento Administrativo Nacional de Estadística - DANE. (2018c). *Fuerza laboral y educación*. Retrieved from <https://www.dane.gov.co/index.php/estadisticas-por-tema/educacion/fuerza-laboral-y-educacion>
- Departamento Administrativo Nacional de Estadística - DANE. (2018d). *Necesidades Básicas Insatisfechas*. Retrieved from <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/necesidades-basicas-insatisfechas-nbi>

- Departamento Administrativo Nacional de Estadística - DANE. (2018e). *Pobreza y desigualdad*. Retrieved from <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-y-desigualdad>
- Departamento Administrativo Nacional de Estadística - DANE. (2019). *Necesidades básicas insatisfechas (NBI)*. Retrieved from <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/necesidades-basicas-insatisfechas-nbi>
- Diagne, M., Mathieu, D., Lauret, P., Boland, J., & Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, 27, 65–76. Retrieved from <http://dx.doi.org/10.1016/j.rser.2013.06.042> doi: 10.1016/j.rser.2013.06.042
- Dirección de Inversiones y Finanzas Públicas. (2006). *Manual de valoración y cuantificación de beneficios* (Tech. Rep.). Departamento Nacional de Planeación.
- Dos Santos, C. M., De Souza, J. L., Ferreira Junior, R. A., Tiba, C., de Melo, R. O., Lyra, G. B., ... Lemes, M. A. M. (2014). On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. *Energy*. doi: 10.1016/j.energy.2014.04.116
- Dos Santos, P. H., Neves, S. M., Sant'Anna, D. O., Oliveira, C. H. d., & Carvalho, H. D. (2019). The analytic hierarchy process supporting decision making for sustainable development: An overview of applications. *Journal of Cleaner Production*, 212, 119–138. Retrieved from <https://doi.org/10.1016/j.jclepro.2018.11.270> doi: 10.1016/j.jclepro.2018.11.270
- El Congreso de Colombia. (2014). POR MEDIO DE LA CUAL SE REGULA LA INTEGRACIÓN DE LAS ENERGÍAS RENOVABLES NO CONVENCIONALES AL SISTEMA ENERGÉTICO (No. May). Retrieved from [http://www.upme.gov.co/Normatividad/Nacional/2014/LEY\\_1715\\_2014.pdf](http://www.upme.gov.co/Normatividad/Nacional/2014/LEY_1715_2014.pdf)
- Estévez, J., Gavilán, P., & Giráldez, J. V. (2011). Guidelines on validation procedures for meteorological data from automatic weather stations. *Journal of Hydrology*, 402(1-2), 144–154. doi: 10.1016/j.jhydrol.2011.02.031
- Fan, J., Chen, B., Wu, L., Zhang, F., Lu, X., & Xiang, Y. (2018). Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions. *Energy*, 144, 903–914. Retrieved from <https://doi.org/10.1016/j.energy.2017.12.091> doi: 10.1016/j.energy.2017.12.091
- Feleki, E., Vlachokostas, C., & Moussiopoulos, N. (2018). Characterisation of sustainability in urban areas: An analysis of assessment tools with emphasis on European cities. *Sustainable Cities and Society*, 43(July), 563–577. Retrieved from <https://doi.org/10.1016/j.scs.2018.08.025> doi: 10.1016/j.scs.2018.08.025
- Figueirêdo Neto, G. S., & Rossi, L. A. (2019). Photovoltaic energy in the enhancement of indigenous education in the Brazilian Amazon. *Energy Policy*, 132(May), 216–222. Retrieved from <https://doi.org/10.1016/j.enpol.2019.05.037> doi: 10.1016/j.enpol.2019.05.037

- Gaspars-Wieloch, H. (2019). Project Net Present Value estimation under uncertainty. *Central European Journal of Operations Research*, 27(1), 179–197. doi: 10.1007/s10100-017-0500-0
- Ghimire, S., Deo, R. C., Downs, N. J., & Raj, N. (2019). Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia. *Journal of Cleaner Production*, 216, 288–310. Retrieved from <https://doi.org/10.1016/j.jclepro.2019.01.158> doi: 10.1016/j.jclepro.2019.01.158
- Ghimire, S., Deo, R. C., Raj, N., & Mi, J. (2019). Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy*, 253(April), 113541. Retrieved from <https://doi.org/10.1016/j.apenergy.2019.113541> doi: 10.1016/j.apenergy.2019.113541
- Gobernación de Nariño. (2016a). Plan participativo de Desarrollo Departamental. *Plan de Desarrollo Departamental de Nariño*, 255. doi: 10.1017/CBO9781107415324.004
- Gobernación de Nariño. (2016b). *Plan participativo de Desarrollo Departamental* (Tech. Rep.). Gobernación de Nariño.
- Goodin, D. G., Hutchinson, J. M. S., Vanderlip, R. L., Knapp, M. C., & Goodin, D. G. (1999). Estimating Solar Irradiance for Crop Modeling Using Daily Air Temperature Data. *AGROCLIMATOLOGY*, 91, 845–851.
- Gueymard, C. A. (2014). A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. *Renewable and Sustainable Energy Reviews*, 39, 1024–1034. Retrieved from <http://dx.doi.org/10.1016/j.rser.2014.07.117> doi: 10.1016/j.rser.2014.07.117
- Hargreaves, G. H., & Samani, Z. A. (1982). Estimating Potential Evapotranspiration. *Journal of the Irrigation and Drainage Division*, 108(IR3), 225–230.
- Harrenll, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis* (Vol. 13) (No. 5). Springer. doi: 10.1007/978-3-319-1925-7
- Herrera-Grimaldi, P., García-Marín, A. P., & Estévez, J. (2019). Multifractal analysis of diurnal temperature range over Southern Spain using validated datasets. *Chaos*, 29(6). doi: 10.1063/1.5089810
- Husein, M., & Chung, I. Y. (2019). Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach. *Energies*, 12(10). doi: 10.3390/en12101856
- Inman, R. H., Pedro, H. T., & Coimbra, C. F. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6), 535–576. Retrieved from <http://dx.doi.org/10.1016/j.pecs.2013.06.002> doi: 10.1016/j.pecs.2013.06.002
- Instituto Departamental de Salud de Nariño. (2018). *Informe de gestión programa de vigilancia de calidad del agua año 2017* (Tech. Rep.). Pasto.

- Instituto Geográfico Agustín Codazzi - IGAC. (2014). *Nariño características geográficas*. Bogotá: Imprenta Nacional de Colombia.
- International Energy Agency. (2017). International Energy Agency - Energy Access Outlook 2017: From poverty to prosperity. *Energy Procedia*, 94(March), 144. Retrieved from [http://www.iea.org/publications/freepublications/publication/WE02017SpecialReport\\_EnergyAccessOutlook.pdf](http://www.iea.org/publications/freepublications/publication/WE02017SpecialReport_EnergyAccessOutlook.pdf) doi: [10.1016/j.enpol.2016.10.1787/9789264285569-en](https://doi.org/10.1016/j.enpol.2016.10.1787)
- Introduction to Spatial Analysis. (2009). *Introduction to Spatial Analysis*.
- J. Pacheco, & Contreras, E. (2008). *Manual metodológico de evaluación multicriterio para programas y proyectos*. Santiago de Chile: Instituto Latinoamericano y del Caribe de Planificación Económica y Social - ILPES. Retrieved from [http://www.fundacionpobreza.cl/biblioteca-temas.php?id\\_tema=14](http://www.fundacionpobreza.cl/biblioteca-temas.php?id_tema=14)
- Jamaly, M., & Kleissl, J. (2017). Spatiotemporal interpolation and forecast of irradiance data using Kriging. *Solar Energy*, 158(February), 407–423. Retrieved from <http://dx.doi.org/10.1016/j.solener.2017.09.057> doi: 10.1016/j.solener.2017.09.057
- Jamil, B., & Akhtar, N. (2017). Comparison of empirical models to estimate monthly mean diffuse solar radiation from measured data : Case study for humid-subtropical climatic region of India. *Renewable and Sustainable Energy Reviews*, 77(February), 1326–1342. Retrieved from <http://dx.doi.org/10.1016/j.rser.2017.02.057> doi: 10.1016/j.rser.2017.02.057
- Janjai, S., Laksanaboonsong, J., Nunez, M., & Thongsathitya, A. (2005). Development of a method for generating operational solar radiation maps from satellite data for a tropical environment. *Solar Energy*, 78, 739–751. doi: 10.1016/j.solener.2004.09.009
- Jeffrey, S. J., Carter, J. O., Moodie, K. B., & Beswick, A. R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling and Software*, 16(4), 309–330. doi: 10.1016/S1364-8152(01)00008-1
- Kipp & Zonen. (2000). Instruction Manual Pyranometer/ Albedometer CM11 e CM14.
- Kiš, I. M. (2016). Comparison of Ordinary and Universal Kriging interpolation techniques on a depth variable (a case of linear spatial trend), case study of the Šandrovac Field. *The Mining-Geology-Petroleum Engineering Bulletin*, 31(2), 41–58. doi: 10.17794/rgn.2016.2.4
- Kleinbaum, D. G., & Klein, M. (2010). *Logistic Regression: a self-learning text* (No. 3). Springer. doi: 10.1007/978-1-4419-1742-3
- Konstantin, P., & Konstantin, M. (2018). *Power and Energy Systems Engineering Economics*. Gewerbestrasse: Springer. doi: <https://doi.org/10.1007/978-3-319-72383-9>

- Kwon, B. S., Park, R. J., & Song, K. B. (2020). Short-Term Load Forecasting Based on Deep Neural Networks Using LSTM Layer. *Journal of Electrical Engineering and Technology*, 15(4), 1501–1509. Retrieved from <https://doi.org/10.1007/s42835-020-00424-7> doi: 10.1007/s42835-020-00424-7
- Layanun, V., Suksamosorn, S., & Songsiri, J. (2017). Missing-data Imputation for Solar Irradiance Forecasting in Thailand. In *Sice annual conference* (pp. 1234–1239). Kanazawa.
- Li, H., Cao, F., Wang, X., & Ma, W. (2014). A Temperature-Based Model for Estimating Monthly Average Daily Global Solar Radiation in China. *The Scientific World Journal*, 2014. doi: doi.org/10.1155/2014/128754
- Li, J., & Heap, A. D. (2008). A Review of Spatial Interpolation Methods for Environmental Scientists. *Australian Geological Survey Organisation*, 68(2008/23), 154. doi: [http://www.ga.gov.au/image{\\\_}cache/GA12526.pdf](http://www.ga.gov.au/image{\_}cache/GA12526.pdf)
- Li, J., & Heap, A. D. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, 6(3-4), 228–241. Retrieved from <http://dx.doi.org/10.1016/j.ecoinf.2010.12.003> doi: 10.1016/j.ecoinf.2010.12.003
- Li, J., & Heap, A. D. (2014). Environmental Modelling & Software Spatial interpolation methods applied in the environmental sciences : A review. *Environmental Modelling and Software*, 53, 173–189. Retrieved from <http://dx.doi.org/10.1016/j.envsoft.2013.12.008> doi: 10.1016/j.envsoft.2013.12.008
- Løken, E. (2007). Use of multicriteria decision analysis methods for energy planning problems. *Renewable and Sustainable Energy Reviews*, 11(7), 1584–1595. doi: 10.1016/j.rser.2005.11.005
- Manning, R. L. (1996). Logit regressions with continuous dependent variables measured with error. *Applied Economics Letters*, 3(3), 183–184. doi: 10.1080/135048596356636
- Mardani, A., Jusoh, A., Halicka, K., Ejdys, J., Magruk, A., & Ungku, U. N. (2018). Determining the utility in management by using multi-criteria decision support tools: a review. *Economic Research-Ekonomska Istrazivanja*, 31(1), 1666–1716. Retrieved from <https://doi.org/10.1080/1331677X.2018.1488600> doi: 10.1080/1331677X.2018.1488600
- Marinakis, V., Papadopoulou, A. G., & Psarras, J. (2017). Local communities towards a sustainable energy future: needs and priorities. *International Journal of Sustainable Energy*, 36(3), 296–312. Retrieved from <https://doi.org/10.1080/14786451.2015.1018264> doi: 10.1080/14786451.2015.1018264
- Martín, A. M., & Dominguez, J. (2019). Solar Radiation Interpolation. In J. Polo, L. Martín-Pomares, & A. Sanfilipo (Eds.), *Solar resources mapping* (pp. 301–311). Springer. doi: 10.1007/978-3-319-97484-2{\\_}12

- Martínez, A. G. (2018). Nariño: Departamento de Nariño Colombia - Informacion detallada Nariño Colombia. Retrieved from <https://www.todacolombia.com/departamentos-de-colombia/narino.html>
- Mary, S. A. S. A., & Suganya, G. (2016). Multi-Criteria Decision Making Using ELECTRE. *Circuits and Systems*, 07(06), 1008–1020. doi: 10.4236/cs.2016.76085
- Mayer, D. G., & Butler, D. G. (1993). Statistical validation. *Ecological Modelling*, 68(1-2), 21–32. doi: 10.1016/0304-3800(93)90105-2
- Mazorra-Aguiar, L., & Díaz, F. (2018). Solar Radiation Forecasting with Statistical Models. In R. Perez (Ed.), *Wind field and solar radiation characterization and forecasting*. (pp. 171–198). Springer. doi: 10.1007/978-3-319-76876-2{\\_}6
- Meza F., & Varas E. (2000). Estimation of mean monthly solar global radiation as a function of temperature. *Agricultural and Forest Meteorology* 100 (2000) 231–241. , 100, 231–241.
- Ministerio de Cultura. (2020). Sistema Nacional de Información Cultural. Retrieved from <http://www.sinic.gov.co/SINIC/ColCultural/ColCulturalBusca.aspx?AREID=3&COLTEM=216&IdDep=52&SECID=8>
- Ministerio de Minas y Energía. (2018). Hidrocarburos - Ministerio de Minas y Energía. Retrieved from <https://www.minminas.gov.co/cobertura-nacional1>
- Montedonico, M., Herrera-Neira, F., Marconi, A., Urquiza, A., & Palma-Behnke, R. (2018). Co-construction of energy solutions: Lessons learned from experiences in Chile. *Energy Research and Social Science*, 45(July), 173–183. Retrieved from <https://doi.org/10.1016/j.erss.2018.08.004> doi: 10.1016/j.erss.2018.08.004
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. New Jersey: John Wiley.
- Moon, S. H., & Kim, Y. H. (2020). An improved forecast of precipitation type using correlation-based feature selection and multinomial logistic regression. *Atmospheric Research*, 240(February), 104928. Retrieved from <https://doi.org/10.1016/j.atmosres.2020.104928> doi: 10.1016/j.atmosres.2020.104928
- Moreno, A., Gilabert, M. A., & Martínez, B. (2011). Mapping daily global solar irradiation over Spain: A comparative study of selected approaches. *Solar Energy*, 85(9), 2072–2084. doi: 10.1016/j.solener.2011.05.017
- Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, 9(1), 207–218. Retrieved from <https://cran.r-project.org/web/packages/imputeTS/vignettes/imputeTS-Time-Series-Missing-Value-Imputation-in-R.pdf>
- Mossos, M. A. (2019). *Informe mensual de telemetría* (Tech. Rep.). Centro Nacional de Monitoreo.

- Munawar, U., & Wang, Z. (2020). A Framework of Using Machine Learning Approaches for Short-Term Solar Power Forecasting. *Journal of Electrical Engineering and Technology*, 15(2), 561–569. Retrieved from <https://doi.org/10.1007/s42835-020-00346-4> doi: 10.1007/s42835-020-00346-4
- Nwokolo, S. C., & Ogbulezie, J. C. (2017, 5). A quantitative review and classification of empirical models for predicting global solar radiation in West Africa. *Beni-Suef University Journal of Basic and Applied Sciences*. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S2314853517300458> doi: 10.1016/j.bjbas.2017.05.001
- Oficina de planeación educativa. (2018). Secretaría Departamental De Nariño. , 169. Retrieved from [http://www.sednarino.gov.co/SEDNARINO12/phocadownload/2018/Descargas/2\\_BOLETINESTADISTICO2017.pdf](http://www.sednarino.gov.co/SEDNARINO12/phocadownload/2018/Descargas/2_BOLETINESTADISTICO2017.pdf)
- Ogunsola, O. T., & Song, L. (2014). Restoration of long-term missing gaps in solar radiation. *Energy and Buildings*, 82, 580–591. Retrieved from <http://dx.doi.org/10.1016/j.enbuild.2014.07.088> doi: 10.1016/j.enbuild.2014.07.088
- Okundamiya, M. S., & Nzeako, A. N. (2011, 5). Empirical Model for Estimating Global Solar Radiation on Horizontal Surfaces for Selected Cities in the Six Geopolitical Zones in Nigeria. *Journal of Control Science and Engineering*, 2011, 1–7. Retrieved from <http://www.hindawi.com/journals/jcse/2011/356405/> doi: 10.1155/2011/356405
- Olea, R. A. (1999). Simple Kriging. In *Geostatistics for engineers and earth scientists* (pp. 7–30). Boston, MA: Springer US. Retrieved from [http://link.springer.com/10.1007/978-1-4615-5001-3\\_2](http://link.springer.com/10.1007/978-1-4615-5001-3_2) doi: 10.1007/978-1-4615-5001-3{\\_}2
- Oliver, A. M., & Webster, R. (2015). *Basic Steps in Geostatistics: The Variogram and Kriging*. Springer. doi: 10.1007/978-3-319-15865-5ISSN
- Oliver, M. A., & Webster, R. (1990). Kriging : a method of interpolation for geographical information systems. *International journal of geographical information systems*, 4(3), 313–332.
- Opoku, R., Adjei, E. A., Ahadzie, D. K., & Agyarko, K. A. (2020). Energy efficiency, solar energy and cost saving opportunities in public tertiary institutions in developing countries: The case of KNUST, Ghana. *Alexandria Engineering Journal*, 59(1), 417–428. Retrieved from <https://doi.org/10.1016/j.aej.2020.01.011> doi: 10.1016/j.aej.2020.01.011
- Palma-Behnke, R., Jiménez-Estévez, G., Sáez, D., Montedonico, M., Mendoza-Araya, P., Hernández, R., & Muñoz, C. (2019). Lowering electricity access barriers by means of participative processes applied to microgrid solutions : The Chilean case. *Proceedings of the IEEE*, 1–15. doi: 10.1109/JPROC.2019.2922342
- Paulescu, M. (2008). Solar Irradiation via Air Temperature Data. In V. Badescu (Ed.), *Modeling solar radiation at the earth surface* (pp. 175–193). Springer.

- Paulescu, M., Paulescu, E., Gravila, P., & Badescu, V. (2013). *Weather Modeling and Forecasting of PV Systems Operation*. London: Springer London. Retrieved from <http://link.springer.com/10.1007/978-1-4471-4649-0> doi: 10.1007/978-1-4471-4649-0
- Pebesma, E. (2016). *Fitting variogram models in gstat*. Retrieved from <https://www.r-spatial.org/r/2016/02/14/gstat-variogram-fitting.html>
- Pebesma, E., & Graeler, B. (2020). *Package 'gstat' Title Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation* (Tech. Rep.). Retrieved from <https://github.com/r-spatial/gstat/issues/>
- Premalatha, N., & Valan Arasu, A. (2016). Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms. *Journal of Applied Research and Technology*, 14(3), 206–214. Retrieved from <http://dx.doi.org/10.1016/j.jart.2016.05.001> doi: 10.1016/j.jart.2016.05.001
- Quansah, E., Amekudzi, L. K., Preko, K., Aryee, J., Boakye, O. R., Boli, D., & Salifu, M. R. (2014, 1). Empirical Models for Estimating Global Solar Radiation over the Ashanti Region of Ghana. *Journal of Solar Energy*, 2014, 1–6. Retrieved from <http://www.hindawi.com/journals/jse/2014/897970/> doi: 10.1155/2014/897970
- Ramírez J., J. C., de Aguas P., J. M., & De Aguas, M. (2017). *Escalafón de la competitividad de los departamentos de Colombia 2017* (Tech. Rep.). Bogotá. Retrieved from [https://repositorio.cepal.org/bitstream/handle/11362/43156/1/S1800010\\_es.pdf](https://repositorio.cepal.org/bitstream/handle/11362/43156/1/S1800010_es.pdf)
- Reikard, G. (2009). Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83(3), 342–349. Retrieved from <http://dx.doi.org/10.1016/j.solener.2008.08.007> doi: 10.1016/j.solener.2008.08.007
- Rivero, M., Orozco, S., Sellschopp, F. S., & Loera-Palomo, R. (2017). A new methodology to extend the validity of the Hargreaves-Samani model to estimate global solar radiation in different climates: Case study Mexico. *Renewable Energy*. doi: 10.1016/j.renene.2017.08.003
- Rodríguez, H. (2011). *Observatorio de energías renovables en América Latina y el Caribe: Colombia*. OLADE; ONUDI.
- Rodríguez-Rivero, C., Pucheta, J., Laboret, S., Sauchelli, V., & Patiño, D. (2017). Short-Term Series Forecasting By Complete and Incomplete Datasets. *Journal of Artificial Intelligence and Soft Computing Research*, 7(1), 5–16.
- Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, 48(1), 9–26. doi: 10.1016/0377-2217(90)90057-I
- Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *Int. J. Services Sciences*, 1(1), 83. Retrieved from <http://www.rafikulislam.com/uploads/resources/197245512559a37aadea6d.pdf> doi: 10.1504/IJSSCI.2008.017590

- Samani, Z. (2000). Estimating Solar Radiation and Evapotranspiration Using Minimum Climatological Data. *Journal of Irrigation and Drainage Engineering*, 126, 265–267. doi: 10.1061/(ASCE)0733-9437(2000)126:4(265)
- Sankar, G., Kumar, P., & Maiti, R. (2018). Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon ( SOC ). *Journal of the Saudi Society of Agricultural Sciences*, 17(2), 114–126. Retrieved from <https://doi.org/10.1016/j.jssas.2016.02.001> doi: 10.1016/j.jssas.2016.02.001
- Şen, Z. (2008). *Solar Energy Fundamentals and Modeling Techniques*. Springer. doi: 10.1007/978-1-84800-134-3
- Serrano, A., Sanchez, G., & Cancillo, M. L. (2015). Correcting daytime thermal offset in unventilated pyranometers. *Journal of Atmospheric and Oceanic Technology*, 32(11), 2088–2099. doi: 10.1175/JTECH-D-15-0058.1
- Shumway, R. H., & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications With R Examples* (Vol. 102; G. Casella, S. Fienberg, & I. Olkin, Eds.). Springer. Retrieved from <http://books.google.com/books?id=9tv0taI816YC> doi: 10.1007/978-1-4419-7865-3
- Sobri, S., Koohi-Kamali, S., & Rahim, N. A. (2018, 1). Solar photovoltaic generation forecasting methods: A review. *Energy Conversion and Management*, 156, 459–497. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0196890417310622> doi: 10.1016/J.ENCONMAN.2017.11.019
- Suehrcke, H. (2000). On the relationship between duration of sunshine and solar radiation on the Earth's surface: Angström's equation revisited. *Solar Energy*, 68(5), 417–425.
- Superintendencia de Servicios Públicos Domiciliarios SSPD. (2019). *Zonas No Interconectadas – ZNI: Diagnóstico de la Prestación del Servicio de Energía Eléctrica 2019*. Retrieved from [https://www.superservicios.gov.co/sites/default/archivos/Publicaciones/Publicaciones/2019/Nov/diagnostico\\_de\\_la\\_prestacion\\_del\\_servicio\\_zni\\_-\\_07-11-2019-lo\\_1.pdf](https://www.superservicios.gov.co/sites/default/archivos/Publicaciones/Publicaciones/2019/Nov/diagnostico_de_la_prestacion_del_servicio_zni_-_07-11-2019-lo_1.pdf)
- Ubilla, K., Jiménez-Estévez, G. A., Hernández, R., Reyes-Chamorro, L., Irigoyen, C. H., Severino, B., & Palma-Behnke, R. (2014). Smart microgrids as a solution for rural electrification: Ensuring long-term sustainability through cadastre and business models. *IEEE Transactions on Sustainable Energy*, 5(4), 1310–1318. doi: 10.1109/TSTE.2014.2315651
- Unidad de Planeación Minero Energética. (2019). *Índice de Cobertura de Energía Eléctrica - ICEE 2018* (Tech. Rep.). Bogotá: Unidad de Planeación Minero Energético. Retrieved from <http://www.siel.gov.co/Inicio/CoberturadelSistemaIntercontadoNacional/ConsultasEstadisticas/tabid/81/Default.aspx>

- United Nations. (2018). *Calculating the human development indices-graphical presentation Inequality-adjusted Human Development Index (IHDI) Knowledge Human Development Index (HDI) Long and healthy life A decent standard of living Human Development Index (HDI) Knowledge Long and* (Tech. Rep.). Retrieved from [http://hdr.undp.org/sites/default/files/hdr2018\\_technical\\_notes.pdf](http://hdr.undp.org/sites/default/files/hdr2018_technical_notes.pdf)
- Universidad de Nariño, Unidad de Planeación Minero Energética-UPME, Usaid, U. S. A. f. I. D., & Ipse, I. D. P. Y. P. D. S. E. P. L. Z. N. I. (2014). Diagnóstico energético y social del departamento de Nariño. , 127.
- Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE. (2014a). *Bombeo De Agua Para Riego Utilizando Energía Solar* (Tech. Rep.). San Juan de Pasto. Retrieved from <http://www1.upme.gov.co/sgic/sites/default/files/BombeoSolarTaminango.pdf>
- Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE. (2014b). *Diseño De Red Inalámbrica Rural Para Acceso a Internet En Las Instituciones Educativas Pertenecientes a Las Comunidades Negras De Las Subregiones De Sanquianga* , (Tech. Rep.).
- Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE. (2014c). *Energía Solar Fotovoltaico Como Estrategia Alternativa y Sostenible de Energización en el Municipio de Santacruz* (Tech. Rep.). San Juan de Pasto.
- Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE. (2014d). *Estudio Para La Implementación De Un Sistema De Alumbrado Fotovoltaico En El Municipio* (Tech. Rep.). San Juan de Pasto.
- Vaidya, O. S., & Kumar, S. (2006). Analytic hierarchy process: An overview of applications. *European Journal of Operational Research*, 169(1), 1–29. doi: 10.1016/j.ejor.2004.04.028
- Viera Díaz, M. A. (2002). *Geoestadística Aplicada*. Instituto de Geofísica UNAM; Instituto de Geofísica y Astronomía CITMA.
- Wang, J. J., Jing, Y. Y., Zhang, C. F., & Zhao, J. H. (2009). Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews*, 13(9), 2263–2278. doi: 10.1016/j.rser.2009.06.021
- Webster, R., & Oliver, A. M. (2007). *Geoestatistics for Environmental Scientists* (Second ed., Vol. 14; S. Senn, M. Scott, & V. Barnett, Eds.). John Wiley. doi: 10.1097/00005344-198900149-00008
- Yang, S., Zhu, X., & Guo, W. (2018). Cost-Benefit Analysis for the Concentrated Solar Power in China. *Journal of Electrical and Computer Engineering*, 2018. doi: 10.1155/2018/4063691
- Žižlavský, O. (2014). Net Present Value Approach: Method for Economic Assessment of Innovation Projects. *Procedia - Social and Behavioral Sciences*, 156(November 2014), 506–512. doi: 10.1016/j.sbspro.2014.11.230

- Zore, Đ., Čuček, L., Širovnik, D., Novak Pintarič, Z., & Kravanja, Z. (2018). Maximizing the sustainability net present value of renewable energy supply networks. *Chemical Engineering Research and Design*, 131, 245–265. doi: 10.1016/j.cherd.2018.01.035