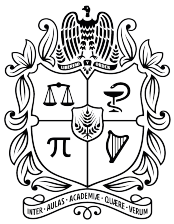


UNIVERSIDAD
NACIONAL
DE COLOMBIA

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:
Meteorologìa energética, energìa 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

Acknowledgments

Gobernación de Nariño through a scholar credit managed by the Fundación Centro de Estudios Básicos y Aplicados - CEIBA supported this thesis. Alianza del Pacífico funded the research visit to Chile University. Furthermore, the Energy Center of the University of Chile supported the fieldwork in Huatacondo town. The meteorological information was obtained in the project analysis of non-conventional renewable energy resources complementarity in Colombia

Finally, I want to thanks to research group GIPEM for supporting me during the development of my work, especially to my advisor Belizza Ruiz.

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 worldwide available resource that could drive electrification processes in regions with low socio-economic indexes. Therefore, to know solar irradiance behavior and data is increasingly a mandatory activity. However, some interesting sites, generally socio-economic outcast places, do not rely on solar irradiance data, and if information exists, it is not complete. Therefore, researchers use some techniques to estimate this energy resource with information from other meteorological variables as temperature. 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 empirical temperature-based models in tropical and mountainous environments. Moreover, this work proposes a new empirical technique that models solar irradiance in some areas better than the three techniques mentioned. Statistical error comparison allows us to choose the best model for each location and the data imputation model. Hargreaves and Samani's model presented better results in the Pacific zone, and the proposed model showed better results in the Andean and Amazon zones. Another significant result is the linear relationship between the new empirical model constants and the altitude 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 Feed Forward Network, Multiple Layer Feed Forward Network, 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 environments.

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.

Resumen

Artículo [I] Incluir a las comunidades en proyectos socio-técnicos require abordar aspectos sociales, energéticos, ambientales, políticos y culturales. Dirigir un proyecto energético con un enfoque socio-técnico require el análisis en conjunto del proyecto y la comunidad impactada. En este sentido, este trabajo se enfoca en formular una metodología que facilite la priorización de proyectos y la participación de la comunidad. Para evaluar a la comunidad se adpatan los índices de desarrollo humano y los índices de los objetivos de desarrollo sustentable a la información disponible para Nariño. El valor presente neto es la herramienta usada para la evaluación del proyecto.El proceso de análisis jerárquico permite evaluar la comunidad y el proyecto conjuntamente y establecer objetivos de priorización. Por otra parte, la metodología de co-costrucción es la base de la directriz propuesta para trabajo con la comunidad. Esta investigación encontró que existe una relación entre los proyectos que buscan mejorar la calidad de vida y la educación en Nariño.

Artículo [II] La irradiancia solar es un recurso ampliamente disponible en el planeta, 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 para estimar la irradiancia solar. Una de esas técnicas son los 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 interpolacióne espacial en un ambiente tropical y montañoso. La información usada es del departamento de Nariño-Colombia que tiene un área de 33.268 km^2. El modelo propuesto muestra un mejor comportamiento en sitios localizado a más de 2.500 msmnl, 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 en-

ergía solar. Existen varias técnicas para pronosticar la irradiancia solar global, como las númericas y las estadísticas. En este contexto, esta investigación compara cuatro enfoques de pronóstico estadístico: Promedio móvil integrado autorregresivo, red neuronal de una capa, red neuronal de multiples capas y memoria a corto y plazo, 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 jerarquico, 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.

Content

List of Figures

List of Tables

List of Symbols

Symbols with Latin alphabet

Symbol	Definition
a	Empirical coefficient
A	The semivariance matrix between the ith and jth point
b	Empirical coefficient
c	Empirical coefficient
d_i	Distance between the x_i and x_0
D	Julian day
H	Global horizontal irradiance at ground level
H_0	Daily extraterrestrial solar irradiance
I_0	Hourly extraterrestrial solar irradiance
I_{cs}	Hourly clear-sky global solar irradiance
I_{mt}	Global solar irradiance in the time t
I_{sc}	Solar constant
K_t	Clearness Index
n	The number of sampled points
m	Air mass
p	Exponent parameter
r	The search window radius
T_h	Hourly measured temperature
T_{max}	Maximum daily temperature
T_{mean}	Mean daily temperature
T_{min}	Minimum daily temperature

Symbol	Definition
T_R	Ratio between the daily minimum and maximum temperature
ΔT	Difference between the daily maximum and minimum temperature
$u(x)$	Non-stationary mean
x_0	Target point
x_i	Not sampled point
z	Measured point
$z(x_i + h)$	The value at an h distance from x_i

Symbols with Greek alphabet

Symbol	Definition
ϕ	Latitude
δ	Solar declination
ω_s	Hour angle at sunset
β	Solar altitude
β_k	Unknown coefficients estimated from the data
$\epsilon(x)$	The spatially correlated random residual function
λ_i	The weight of the sampled the point
τ	Atmospheric transmittance
μ	Stationary mean
ψ	The Lagrange multiplier

Abbreviations

Abbreviation	Term
AHP	Analytic Hierarchy Process
ANN	Artificial Neural Network
ARMA	Auto Regressive Moving Average
ARIMA	Auto Regressive Integrated Moving Average
AWS	Automatic Weather stations
CEPAL	Comission for Latin America and the Caribbean
CWS	Conventional Weather Stations
DANE	Departamento Administrativo Nacional de Estadística
IDEAM	Instituto de Hidrología, Meteorología y Estudios Ambientales
IDW	Inverse Distance Weighting
IPSE	Instituo de Planificación y Soluciones Energéticas para las Zonas no Interconectadas
PERS	Programa de Energización Rural Sostenible
UPME	Unidad de Planeación Minero Energética

1 Introduction

Renewable energies are a locally available resource with environmental and social benefits, which allow community integration in energy projects (Dawoud et al., 2019). The proliferation of solar-powered systems has prompted studies to estimate the potential of the solar resource. Solar irradiance information is essential in several fields, such as electricity generation, weather forecasting, agriculture, and ecology (Moreno et al., 2011). Notably, in renewable energy applications, the well-designed power plant relies on the availability of solar irradiance potential information. The lack or low quality of information is a barrier to analyzing solar irradiance potential (Janjai et al., 2005; J. Li & Heap, 2014). This issue could arise in the inadequate planning of solar energy projects, negatively impacting electricity generation's growth based on this renewable source.

Encouraging electricity generation with solar energy requires understanding the resource. However, in developing countries, the solar irradiance data are not widely available because of the scarcity of weather stations that measure this variable and the equipment calibration and maintenance requirements. In some cases, although the databases have an open-access, the process of getting information might be extended (Bakirci, 2009). Therefore, in the last years, the solar irradiance studies have used other relevant variables like temperature, sunshine, relative humidity, among others, to estimate this variable.

The solar researchers have developed models and methods to estimate the daily or monthly solar insolation, from empirical models to artificial intelligence. The empirical models are usually based on astronomical, geometrical, physical, and meteorological factors, highlighting the last (Besharat et al., 2013). The empirical model implementation depends on the availability and consistency of the data (Akinoglu, 2008a). In this sense, Almorox et al. recommended that the empirical model's data should be simultaneous and reliable in the study site (Almorox et al., 2011). Regardless of the solar insolation estimation method used, the data collected at ground level offer information from a point source.As a result, it needs to estimate the insolation values in the not sampled points to obtain continuous data over the study area (J. Li & Heap, 2011). The standardization of the information from isolated measurements is possible with the regional interpolation techniques by constructing a regional model (Moreno et al., 2011). Overall, there are two categories for the interpolation techniques: non-geostatistical and geostatistical. However, all methods share the same general estimation formula (J. Li & Heap, 2008).

Furthermore, solar energy for generating electricity drives the studies in forecasting irradiance (Reikard, 2009). Solar irradiance forecasting allows estimating the electricity generation in the long- medium- and short-term. This information is crucial to maintain the balance between the electricity demand and

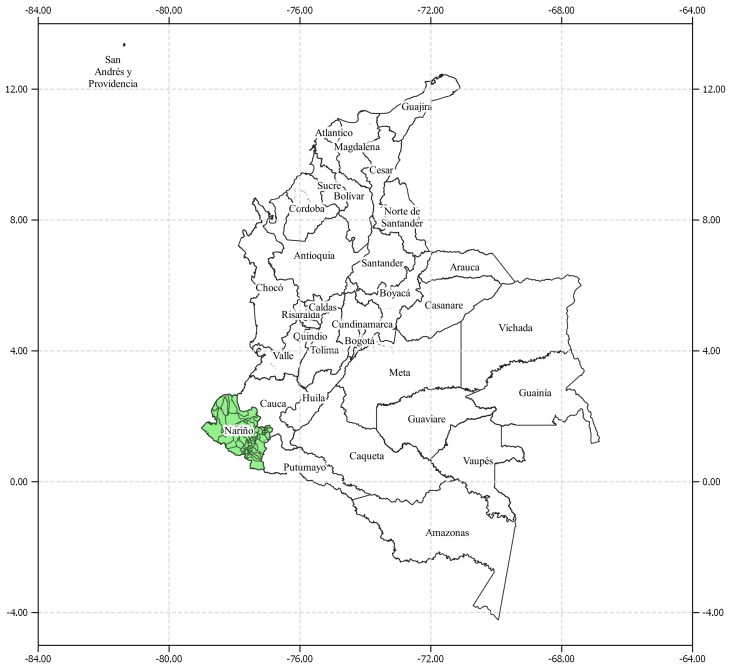
supply (Dannecker, 2015) and minimizing costs associated with the start and shutdown of conventional power plants (Badosa et al., 2017). These models facilitate the implementation of PV systems, both on-grid and off-grid. Data collected in extended periods are used to understand the behavior and predict future solar radiation values in a specific location (Suehrcke, 2000). Time series analysis examines the recorded data over time to develop a mathematical model to describe a variable (Shumway & Stoffer, 2011). There are several approaches to forecasting solar irradiance: persistence methods, autoregressive models, and soft computing techniques (Demirhan & Renwick, 2018; Diagne et al., 2013; Inman et al., 2013). The artificial neural network, fuzzy logic, and hybrids are robust for modeling the physical processes' stochastic nature, like the solar irradiance, because of their capacity to compensate systematic errors and problematic learnable deviation (Paulescu et al., 2013). The selection of a forecasting method depends on the desired timescale, i.e., intra-hour, intra-day, medium, and long term. Statistical approaches usually perform well for short-term and soft computing techniques for long-term analysis (Coimbra et al., 2013; Demirhan & Renwick, 2018)

The use of solar energy with photovoltaic (PV) technology has grown in recent years. It has taken decades for solar energy is feasible economically in developing countries. Today, the associated costs to this energy resource are competitive, mainly in rural areas where electricity from conventional resources could be more expensive. Also, solar energy projects allow including the community in their planning and execution; as a consequence, improving the project sustainability over time. When the population is an active part of the energy projects, their development eases the interaction between the communities, the funding entities, and the project's developers. Consequently, the joint analysis of both projects and communities addresses energy projects from a socio-technical view. Besides, sustainable energy projects correlate with the populations' socio-economic development, including population growth, urbanization, and industrialization.

Creating a value chain around the solar resource for the department of Nariño from an energy approach is the focus of this work. This department is located in the south of Colombia, as Figure , and it is characterized by its diverse population, its three environmental regions: Pacific, Andean, and Amazon, and the need to improve the quality of life of the population that has been affected by the armed conflict. In terms of electricity supply, the department is divided into two zones, the first receives electricity from the National Interconnected System (SIN), and the second is called the Non-Interconnected Zone (ZNI), which obtains electricity from local plants that use fossil fuels. The Colombian Congress enacted Law 1715 to promote electricity generation with non-conventional renewable energy sources in 2014 (El Congreso de Colombia, 2014). This law promotes distributed generation and self-generation of electricity. Therefore, considering the regulatory framework of Colombia and the characteristics of Nariño, this work proposes a prioritization methodology for productive projects and a community guideline to joint work between the company and the community. Furthermore, this research assesses three empirical temperature-based models: Hargreaves and Samani (HS), Bristow and Campbell (BC), and Okundamiya and Nzeako (ON), to estimate solar insolation in a tropical and mountainous environment. Finally, this study applies four state-of-the-art prediction models for global solar irradiance and insolation forecasting with one day ahead in tropical and mountainous environments with incomplete datasets: i) Autoregressive Moving Average (ARIMA), ii) Single Layer Feed Forward Network (SL-FFN),

iii) Multi-Layer Feed Forward Network (ML-FFN), and iv) Long Short-Term Memory (LSTM).

Figure 1-1. Nariño location in Colombia



For the first part, we analyzed several community evaluation components, such as the Human Development Index (HDI) and Sustainable Development Goals Index (SDGI), and adapted those to Nariño's characteristics and the available information in the national or regional public databases. The Net Present Value (NPV) was the measurement used to project evaluation because this is a commonly used measurement for cost-benefit evaluation in PV power projects. The AHP was the method for prioritizing projects and communities. The method processed six prioritization objectives to present different perspectives for making decisions. Additionally, this part of the work proposes the community's participation guidelines to integrate its priorities, concerns, and interests with the stakeholders in developing the project. In this part of the work, the authors found a relationship between the objectives that sought to improve education and the population's quality of life.

To estimate solar insolation potential, we analyzed three empirical models based on temperature. The objective of these models was, on the one hand, to fill the solar insolation data gaps in the data series, and on the other hand, to increase the number of points sampled with estimated solar insolation data to improve the quality of solar energy potential maps. Additionally, we propose a new empirical relationship based on the logistic regression model to estimate the solar insolation with the daily temperature difference. The proposed model presents a linear relationship between the empirical constants and the site's altitude. The linear relationship allows calculating the empirical constants of the Conventional Weather Stations (CWS) using their altitude. As a result, it was possible to increase the density of sampled points of solar insolation in the area of Nariño. This result was relevant since the performance of spatial interpolation techniques depends, among other things, on the density of sampled points and their clustering.

We compare four spatial interpolation techniques in the global solar insolation mapping part: Inverse Distance Weighted (IDW) and simple, ordinary, and universal Kriging to obtain spatially continuous solar insolation information in the studied region. The techniques used information collected in eight AWS, and the solar insolation estimated with temperature data, through empirical models, in sixteen locations with CWS. Cross-validation showed that the ordinary Kriging was the best among the analyzed

methods because it presented less bias throughout the year due to its robustness.

In the last part of this research, and mainly to encourage the use of solar energy in the on-grid areas, we implement four forecast models on an ahead-day horizon with an hourly and daily timestamp. In this part, the missing data imputation combines a linear interpolation of the subsequent values with the average of past values measured at the same hour of the imputed data. The results show that neural network-based techniques have better performance than ARIMA. The LSTM model performs better in the Pacific area on AWS with less amount measured data. In comparison, SLP achieves the best performance on AWS with more input data.

This thesis is organized as follows: Chapter 2 corresponds to the prioritization process of energy projects and the community participation guide. Chapter 3 shows the results of the three empirical temperature-based models and the proposed model based on logistic regression, which supported the imputation of daily solar insolation data. Chapter 4 presents the daily solar insolation estimation process using the empirical models of Hargreaves and Samani and the proposed model, and the application of four spatial interpolation techniques that result in twelve daily solar insolation potential maps for each month. Chapter 5 shows the results of applying four solar irradiance forecasting models on an ahead-day in an hourly range. Finally, Chapter 6 collects all the conclusions of the research work.

2 A multi-criteria methodology for prioritization of social projects and community participation: Nariño study case[1]

Abstract

The joint analysis of projects and communities allows for the possibility of addressing energy projects from a socio-technical view. Including communities in the development of projects requires a better understanding of their social, economic, environmental, and cultural affairs. Consequently, this work focuses on a methodological formulation to prioritize energy projects considering community participation. The community evaluation considered the Human Development Index and Sustainable Development Goals and adjusted them according to the information available about Nariño, and the project evaluation uses the Net Present Value. The Analytic Hierarchy Process evaluates communities and projects jointly and establishes prioritization objectives. Moreover, the co-construction methodology is the basis to formulate guidelines for community work. This research found a close relationship between the projects for improving life quality and education in Nariño.

Keywords

Energy projects; Analytic Hierarchy Process; Rural electrification; Human Development Index; Sustainable Development Goal Index

2.1 Introduction

Including the community in the energy project's planning and execution, such as rural electrification or productive projects, improves the initiatives' sustainability over time. When the population is an active part of the energy projects, their development eases the interaction between the communities, the funding entities, and the project's developers. Consequently, the joint analysis of both projects

[1] This chapter is derived in part from an article published in International Journal of Sustainable Energy, 10 Feb 2021, available online: https://doi.org/10.1080/14786451.2021.1883612

and communities addresses energy projects from a socio-technical view. Besides, sustainable energy projects correlate with the populations' socio-economic development, including population growth, urbanization, and industrialization. Renewable energies are locally available resources with environmental and social benefits which allows for the community integration in energy projects (Dawoud et al., 2019). The last aspect is relevant because the traditional approach for rural electrification has been weak in terms of involving the communities in the development of successful projects. When there is a gap between the community and the project development, it presents difficulties with sustainability over time (Palma-Behnke et al., 2019). The rural electrification projects are essential because electricity is a significant factor in the economy and improves the populations' quality of life; therefore, their sustainability over time is also relevant and should be achieved (Balbás Egea & Eguren Egiguren, 2019). Despite this, the International Energy Agency declared that about 1,1 billion people did not have access to electricity in 2017, and about 84 % live in rural areas (International Energy Agency, 2017).

The lack of rural electrification initiatives is faulty, especially in regions with a high population, as in the case of Nariño's. There, the rural population makes up 53 % of the total population. The main advantages of rural electrification involve better lighting systems for dwellings, educational and health centers, better working conditions for domestic activities, reduction of rural migration to cities because of the creation of jobs through the improvement of operations related to the agricultural sector's use of technology, the environment preservation through using renewable sources for electricity generation, among others (Camblong et al., 2009). Accordingly, Nariño's government identified three proposals in its development plan 2016-2019: Territorial peace and social equity, growth and green innovation, and regional integration (Gobernación de Nariño, 2016a). Likewise, it established seven strategic areas to achieve the proposals; each had an action program covering one or more objectives. One of these programs is rural electrification to encourage other fields to develop in the state of Nariño.

The traditional approach to supplying electricity to rural or isolated populations have been through grid enlargement, either at transmission or distribution level or with stand-alone generators that use fossil fuels (Ubilla et al., 2014). The first approach usually requires a high capital investment making it economically unfeasible when there are low levels of demand for electricity. The second approach, although requiring low investment, results in high operational costs. Hence, the electricity cost is high in Nariño's off-grid zones due to the fossil fuel transportation costs from the closest harbors to the communities (Universidad de Nariño et al., 2014). Another element to highlight related to using fossil fuels for generating electricity is the greenhouse gas emissions that negatively impact the environment. As a result, there are programs and initiatives to reduce their consumption and introduce renewable energies for electricity generation (Opoku et al., 2020). Additionally, photovoltaic technology (PV) is a suitable option considering the populations' terrain, resource availability, and technology (Figueirêdo Neto & Rossi, 2019).

The electricity generation with renewable resources at the local level allows for the community's inclusion into the project and transforms a technical project in a socio-technical project. Therefore, the long-term objective of the Institute for Planning and Promotion of Energy Solutions for Non-Interconnected

Areas[2] (IPSE) would be to include the communities in the electricity generation projects to sustain them over time (Rodriguez, 2011). Accordingly, IPSE should improve the projects' implementation, adapting it to a communities' needs. It would reduce the gap between communities and project development (Superintendencia de Servicios Públicos Domiciliarios SSPD, 2019). The previous statement requires evaluating several aspects and formulating strategies to ease the work between the stakeholders, the community, and the funding entity (Marinakis et al., 2017).Consequently, this research focuses on formulating a two-part methodology to make decisions, determine budgets, and formulate public policy in the energy sector that increases energy project's sustainability over time. The first part consists of prioritizing the project development's order using the Analytical Hierarchy Process (AHP) methodology that processed six components from the communities and project evaluation using the Net Present Value (NVP). The second part shows a working guideline to include the community in energy projects.

This work proposes a prioritization methodology for productive projects and a community guideline for joint work. For the first part, the authors analyzed several community evaluation components, such as the Human Development Index (HDI) and Sustainable Development Goals (SDG) and adapted those to Nariño's characteristics and the available information in the regional and national public databases. That information is crucial to select projects with the best solutions in social, environmental, and economic fields (Zore et al., 2018). The selected projects came from the Sustainable Rural Energization Program[3] (PERS), and the selecting criterion was the energy resource, in this case, the solar irradiance. The Net Present Value (NPV) was the measurement used for the project evaluation because this is a commonly used measurement for cost-benefit evaluation in PV power projects (Yang et al., 2018). The AHP was the method for prioritizing projects and communities. The method processed six prioritization objectives to present different perspectives for making decisions. The second part of this methodology focuses on the community's participation guidelines to integrate its priorities, concerns, and interests with the stakeholders in developing the project (Aslani, 2014). The proposed criteria for the community evaluation is a reference framework that could facilitate the project evaluation in other fields. Likewise, the community participation guideline is a process that invites joint work between the community, the funding entity, and the project's developers to sustain the energy project over time. That guideline includes a process that can be modified according to the project and the community needs.

The authors found a direct correlation when the objective is to improve the education or life quality conditions. Therefore, public policies mixing these objectives could enhance the results. It is necessary to analyze more productive projects and to collect more information from the communities. It would improve the characterization and include the cultural traditions in the project development. The guideline proposed by the authors would enhance the probability of accomplishing successful projects and including the community as an active part of the project development. The article contains the following sections; Section 2.2 presents the methodology used in this work; Section 2.3 describes the case study; Section 2.4 presents the results obtained in the state of Nariño, and finally, Section 2.5 presents the conclusions.

[2] IPSE by its Spanish name
[3] PERS by its Spanish name

2.2 Methodology

This research work follows a methodology synthesized in two parts, as Figure **2-1** shows. TThe first part consists of a rural projects' prioritization process that involves both a community and projects' assessment. The second part consists of general guidelines based on the results from the first part of the methodology for installing energy projects in Nariño.

The rural projects'prioritization process required information from PERS, Statistical National Administrative Department[4], and Planning Unit Energy Mining[5], among other institutions. These entities acted directly with the community, collecting primary information through surveys to construct databases and draw up plans and reports. Therefore, the information used in this study reflects the living conditions of the communities in the state of Nariño. Taking the databases into account, the authors considered five dimensions: education, health, security, quality of life, and economy. These aspects let understand sustainability beyond the traditional dimensions since including indicators that assess the quality of life, cultural and institutional relationships, among others (Feleki et al., 2018). The NPV was the tool to evaluate the projects; besides, it constituted another dimension in the information math processing (Ubilla et al., 2014). Dimensions linked to community and the NPV evaluating projects constitute the six dimensions analyzed under the multi-criteria approach that tries them as criteria in the evaluation process. Multi-criteria Decision Making (MCDM) tools allow the evaluation and prioritization of complex problems with high uncertainty, conflicting objectives, multi-interests, among others (Vaidya & Kumar, 2006). Also, the MCDM facilitates the classifying and arranging a finite number of decision alternatives based on attributes or decision criteria, which describe different characteristics (Mary & Suganya, 2016). In the social projects, MCDM lets distribute the budgets efficiently; therefore, the coordination and the development of a project follow a previously established development guideline. When there is a public investor involved, the project evaluation results could be the basis for a program, even a policy (J. Pacheco & Contreras, 2008). The most appropriate and frequently used MCDM in the energy planning field is the AHP, followed by PROMETHEE and ELECTRE (Abdullah & Najib, 2016; Wang et al., 2009).

The AHP has two prime advantages: the quantification and weighting of criteria regarding a given objective and different options for achieving such an objective, and the mathematical processing more understandable than other methods(P. H. Dos Santos et al., 2019). Additionally, AHP transform empirical comparison into numerical values. AHP orders the evaluation process elements. In the first step, AHP states the objective, in the second, it establishes the criteria and sub-criteria, and in the third, it presents alternatives that would be selected projects from PERS. After crossing the objective, criteria, and alternatives, the AHP gives a list of prioritized projects (Saaty, 1990, 2008). ELECTRE III is the most common method from the ELCTREE family used in energy planning, the objective of this method is to choose the alternative that accomplish with most of the criteria. However, the method sometimes present difficult to find the best alternative; therefore, becomes necessary to use a more detailed method. PROMETHEE II is the method commonly used in the energy planning field from the PROMETHEE family. A main dif-

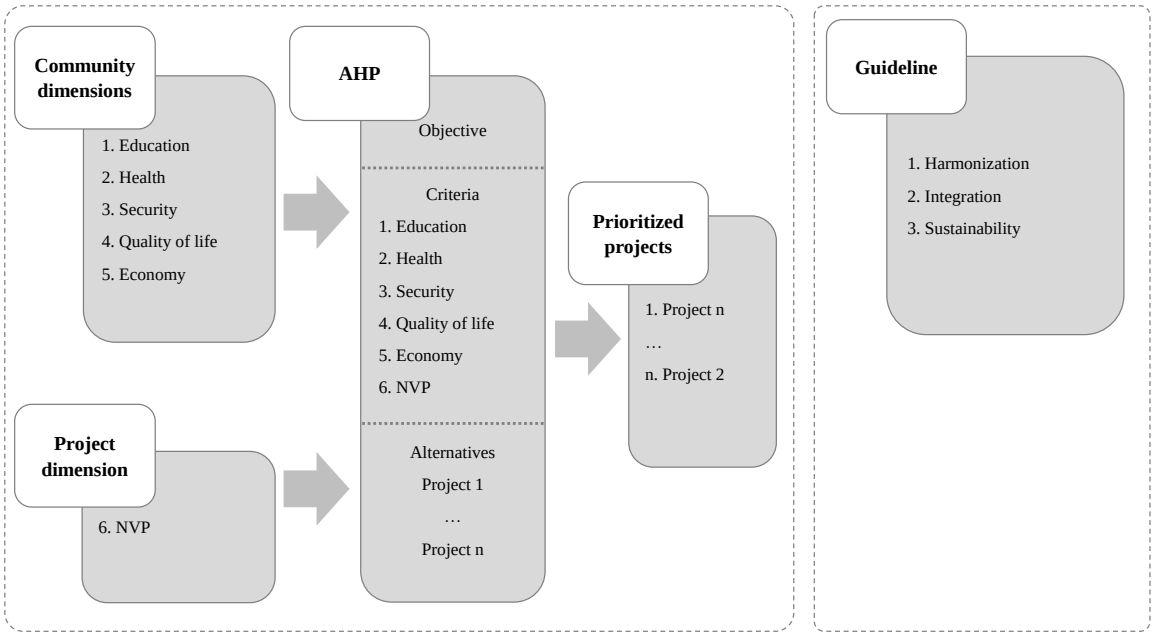
[4] DANE by their Spanish name
[5] UPME by their Spanish name

ference between PROMETHEE II and ELECRTE III is the calculation procedure. For the decision makers the first one method is easily to understand than the second one method. These types of methods, named outranking methods, are usually used to categorize the alternatives in acceptable or unacceptable, but not as methods for the alternative selection (Løken, 2007). Of the three MCDM mentioned above, the AHP has the best properties, that is why the authors chose it (Mardani et al., 2018).

The methodology's second part consists of drawing up a guideline that eases the interaction between the community and the local government or another type of entity promoting socio-energy projects. The guideline's purpose is to insert the community in the project development to create a management system according to the community's tradition and culture and, in turn, to arouse the sustainability of the project. It is convenient to highlight that the starting point to draw up the guidelines was the co-construction methodology because this seeks to empower the community through the decision-making processes to achieve sustainable solutions over time (Montedonico et al., 2018). Nevertheless, the authors modified it, including as one of the requirements, the direct participation of community members in the work team that execute the specific project.

The guideline must cover the community features. Accordingly, Nariño's case is relevant because this territory has diverse population groups as African Colombians (42 collective territories corresponding to 18,80 % of the total population), indigenous (55 indigenous reservation corresponding to 10,79 % of the total population), and half-bred (70,41 % of the total population), changing terrain as valleys, enormous mountains, and forest regions, and an economy that needs dynamism. Those aspects demonstrate that this guideline would have a place in different environments (Ministerio de Cultura, 2020).

Figure 2-1. Methodology scheme.



2.2.1 Community evaluation

At the community level, several indexes exist to explain the environmental, social, and economic aspects that justify political decision-making without a universally accepted collection of indexes for evaluating such populations (Feleki et al., 2018).Although some governmental policies follow the profit maximization principle, some indexes are relevant because of their broader approach, such as System of Environmental-Economic Accounting (SEEA), Adjusted Net Savings (ANS), Living Planet Index (LPI), Genuine Progress Indicator (GPI), Human Development Index (HDI), Sustainable Development Goal Index (SDGI), among others (Arbeláez-Arias, 2006). Even though the indicators mentioned above could offer a whole context, the State of Nariño's available public information was insufficient to compute them. Consequently, the authors only used indexes from both HDI and SDGI that are applicable to the available information (Bertelsmann Stiftung & Sustainable Development Solutions Network (SDSN), 2016; United Nations, 2018).

Considering the situation in the state of Nariño, the authors analyzed five dimensions for the community evaluation: education, health, security, quality of life, and economy. Each dimensions has assosiated with one or more components or social characteristics, as shown in Table **2-1**.

Table 2-1. Dimensions of community evaluation.

Dimension*	Components	Description
Education	Literacy	The older population share than 15 years that read and write.
	Kindergarten Net Coverage	It correspond to the ratio between population aged five years old who attend school and the total population in this age group.
	Primary Net Coverage	It correspond to the ratio of the population aged between six and ten years old who attend school and the total population in this age group.
	Secondary Net Coverage	It correspond to the ratio of the population aged between eleven and fourteen years old who attend school and the total population in this age group.
	Middle Net Coverage	It correspond to the ratio of the population aged between fifteen and sixteen years old who attend school and the total population in this age group.
	National Test Score	Score reached by 11th grade students in the national test Saber 11.
Health	Social Security Coverage	Percentage of the population affiliated to any health insurance.
	Infant Mortality Rate	Child deaths under one year old per 1.000 children born alive.
Security	Armed Conflict Incidence	It measures the armed actions against the State forces, homicides, kidnapping, victims of anti-personnel mines, forced displacement, and area with illicit crops.

Table 2-1 continued from previous page

Dimension*	Components	Description
Quality of life	Housing deficit	Quantitative and qualitative housing deficit. The first one estimates the houses'number that should be built for the total families. The second one refers to the houses' quality, overcrowding, and public services access.
	Unsatisfied Basis Needs	It correspond to overcrowded households and inappropriate conditions, inadequate public services, high economic dependence, school-age children without school attendance.
Economy	Economic Importance	Value added by the municipality with standard prices
	Incidence of multidimensional poverty in rural zones	Illiteracy, low educational level, barriers to early childhood service access, child labor, informal work, long-term unemployment, access to the health system, barriers for access to care given a need, access to freshwater and aqueducts.

*These dimensions correspond to prioritization objectives.

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

All components are comparable because they are quantifiable; however, it is necessary to realize a normalization process because each one has different measurement units. In this work, the normalization procedure is the min-max one, as shown in Equation 2-1,

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

where $x^{'}$ is the index value, x is the value without normalized, $max(x)$is the maximum original values, $min(x)$ is the minimum original values.

A number between 0 and 1 represents each component. When the dimension has more than one component, the dimension value is the average of normalized components, as shown in Equation 2-2. Although some components appear in various dimensions, all of them have the same weight. The authors decided to give the same weight for components because the DANE does not offer information thereon,

$$D_i = \frac{1}{n}\sum_{j=1}^{n} x_j^{'} \tag{2-2}$$

where D_i is the dimension i, $x_j^{'}$ is the indicator j in the dimension i, n is the amount of index into the dimension i.

2.2.2 Project evaluation

The academic community utilizes several methods to perform economic analysis with specific characteristics in their implementation. However, the principle is the same in those methods: "the capital budgeting approach for calculating the economic return of a project a sequence of discounted cash flow" (Žižlavský, 2014). In this research, the NPV evaluates the projects because this is an investment appraisal analysis tool used in the energy sector projects when the costs are the primary variable (Konstantin & Konstantin, 2018). The NPV is a popular and sophisticated economic evaluation technique, which is the sum of the present values of inflows and outflows over a period and the discount rate that introduce a risk rate (Gaspars-Wieloch, 2019). In the project evaluation, the authors considered prioritizing the projects with the lower NPV,

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

where R_a is the net cash inflow-outflows during a period a, i is the discount rate, n is the number of evaluated periods, following the economic maximization approach, and implementing the maximum number of projects. This tool is useful when the economic benefits are equal in more than one project or not measurable. Social projects involve different variables related to welfare that are not measurable from economic benefits (Ubilla et al., 2014). For example, the security feeling by public lighting growth or diseases reduction by freshwater access, among other aspects,

2.2.3 Dimension weighting

The authors use the AHP methodology for weighing the dimensions. The following steps adjusted the methodology to the specific case (Saaty, 1990, 2008):

- Define the problem and the ultimate objective. In this case, there is no expert panel to determine the importance of one criterion over another. Therefore, the procedure starts with formulating six prioritization objectives, named O1 to O6 (see Table 6), evaluated subsequently. Namely, when analysis befalls on the one objective, it takes more relevance than the others that obtain the same assessment. This procedure covers all the prioritization objectives one by one.

- Organize decisions hierarchically starting from the objective, followed by the middle levels composed by the criteria, sub-criteria, and alternatives. In this case, the criteria are the dimensions established for the community and project evaluation, and there are no sub-criteria.

- Build pairwise matrices and compare each objective with the criterion of the lower immediately levels. In this case, there are six pairwise matrices; each one corresponds to a prioritization objective. The matrix's upper-level elements correspondent to the objectives and the matrix's fist column elements correspond to the criteria. A comparison between them, keeping into account

the Saaty scale values and recommendations presented in Table **2-2**, results in a square matrix with ones on the diagonal and prioritization values in the matrix's other positions. This matrix must follow the reciprocity, homogeneity, and consistency principles explained below.

Table 2-2. Saaty scale.

Value	Definition	Description
1	Same importance	Two activities contribute equally to achieve the objective
3	Moderate importance	One activity contributes slightly more to achieve the objective over another
5	Strong importance	One activity contributes strongly more to achieve the objective over another
7	Very strong and demonstrate the importance	One activity contributes much more to achieve the objective over another
9	Extreme importance	Absolute evidence favors an activity over another
2, 4, 6, 8	Intermediate values	When needed an agreement between the parties
Reciprocal	$a_{ij=1} = \frac{1}{a_{ij}}$	Hypothesis of the method

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

- Use obtained prioritization results from the pairwise matrices to weigh them with the level immediately below. The pairwise matrix's eigenvector gives other the priorities level; each element of this vector corresponds to each criterion's weight to evaluate the alternatives. In other words, eigenvector's each element evaluates the criteria that evaluated the alternatives. Finally, the alternatives'prioritization order results of a sum of exponential terms where each base is the dimension value, and the exponent is an eigenvector's element.

The expression $n(n-1)/2$ gives the amount of comparisons, where n correspond to the criteria, as shown in Equation 2-4,

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

with:

$$1 \leq i, j \leq n \tag{2-5}$$

where a_{ij}: pairwise comparison between dimensions i and j, n: number of analyzed dimensions. In this case, as there are six criteria, the number of comparison is fifteen.

The matrix must accomplish with the reciprocity, homogeneity, and consistency properties, that means (Aznar & Guijarro, 2012; J. Pacheco & Contreras, 2008):

- Reciprocity: $a_{ij} = x$ then $a_{ji} = 1/x$ with $1/9 \leq x \leq 9$
- Homogeneity: i and j have the same significance then $a_{ij} = a_{ji} = 1$
- Consistency: If a_{ij} represents the alternative's importance i over alternative j, and a_{jk} represents the alternative's importance j over alternative k, ideally $a_{jk} * a_{kj} = a_{ij} \; \forall \; 1 \leq i, j, k \leq n$; however, as there are subjective evaluations in the pairwise comparisons, the process is iterative until the Consistency Ratio (CR) of A is smaller than 0,1.

  To compute the CR, we follow the next process:

$$\lambda_{max} = V * B \tag{2-6}$$

  where λ_{max} maximum eigenvalue of the pairwise comparison matrix A, V priority vector or eigenvector obtained from matrix A, B row vector containing the sum of column elements of matrix A.

  The last result is useful to compute the Consistency Index (CI) that represents the estimation error's variance of a_{ij} with Equation (2-7):

$$CI = \frac{\lambda_{max} - n}{n - 1}. \tag{2-7}$$

  Finally, the CR is the ratio between CI and RI, as Equation 2-8 shows

$$CR = \frac{CI}{RI} \tag{2-8}$$

  where RI is the random index given by Saaty.

The last step consists of computing the final prioritization value linking the community, project, and weighting results by Equation 2-9. Due to different municipalities having the same kind of project, the community dimensions'assessment is the average of all communities involved in the project,

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

2.2.4 Guideline to work with the community

The guideline objective is to stimulate the community's active participation in the project to develop it without opposition, arouse the population's belongingness with technical labors, and maintain the project on time. The guideline formulation took elements coming from the co-construction methodology and public entities'information of the state of Nariño. The co-construction methodology seeks to empower the community through its role in the decision-making processes to generate a sustainable solution over time (Montedonico et al., 2018). It is important to remark that the population's

distinctiveness demands to adjust the guideline. Consequently, this proposal constitutes a framework guideline for community interaction with three steps: harmonization, integration, and sustainability. The first one seeks to define a common purpose in the interdisciplinary team and characterize the community. The second one deals with the integration of the community, stakeholders, and work team. The last attempts to establish a local management scheme to maintain the project over time.

Unlike the co-construction methodology, this guideline covers the work team's formation and training to create the idea of achieving a successful project before starting the work community. This work team must be interdisciplinary, including professionals from several studies fields. Additionally, the guideline highlights the importance of involving local companies in the project execution, reducing the operation and maintenance costs, and facilitating interaction with the community. Furthermore, this guideline encourages creating a local regulation board to formulate initiatives that overcome barriers and ease the project development. Likewise, this local board considers community cultural traditions to operate as a regulatory entity.

2.3 Case study description

The state of Nariño is in the southwestern region of Colombia. The Nariño's population face power struggles with the troubles derived from an armed conflict that started more than five decades ago. Among the conflict's consequences are the forced displacement that spreads misery belts around the capital cities, land dispossession for planting coca leaf, opium poppy, oil palm, and illegal mining extraction, among other damaging aspects (Ávila et al., 2014). Those most affected population is the rural one, causing a significant impact on the state because 53 % of the Nariño's population lives in rural areas. According to the DANE's data, in 2018, Nariño's rural population had an Unsatisfied Basic Need Rate of 26,61 %, while in the urban areas, this rate was 16,20 % (Departamento Administrativo Nacional de Estadísitica - DANE, 2018d). Regarding the public services for the same year, 13,5 % of the families did not have electricity service, and 29,2 % and 53,1 % of these families did not have any water system or sewage networks, respectively. Even, in San Juan de Pasto, Nariño's capital city, there is only a partial network of natural gas covering 22,15 % of the total population, the rest of the population use LPG cylinders inside houses endangering the families (Ministerio de Minas y Energía, 2018). The state of Nariño's infrastructure level has been classified as medium-high by the CEPAL. This categorization accounted the access

Table 2-3. Classification of localities by hour of electricity supply.

Hours	Localities
0	14
1-6	438
7-12	144
13-18	1
19-23	0
24	6
Without information	36
Total	609

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

to public services, ICTs, roads, airports, and harbors. The same institution gave a medium-low score in competitiveness compared with other Colombian states to the state of Nariño (Ramírez J. et al., 2017).

In 2018, there were 516.398 dwellings, from which 480.134 had electricity service, and 36.264 houses did not. From all electricity consumers, 88,4 % were on-grid users, and 11,6 % were off-grid users. From 11,6 % off-grid users, 8,4 % lived in urban areas and 91,6 % were rural users (Unidad de Planeación Minero Energética, 2019). Urban and rural off-grid users obtain the electricity service from the municipality's power plants that use fossil fuels like diesel. Regarding the service indicators, 71,9 % of the locations without telemetry have electricity from 1 to 6 hours per day, while in 5,9 % remaining does not have information about electricity supply (see Table **2-3**) (Arbeláez Pérez, 2019). Eight localities with telemetry report an average of 7 hours per day of electricity supply (Mossos, 2019). Moreover, the average electricity price is about 0,33[6] US$ kWh (Superintendencia de Servicios Públicos Domiciliarios SSPD, 2019).

2.4 Results obtained

2.4.1 Project prioritization

Figure **2-2** shows the hierarchy scheme for the decision making about the alternatives in the state of Nariño. It is convenient remember that the dimensions correspond to the criteria that evaluate six prioritization objectives. The criteria are education, health, quality of life, security, and economy for the community's assessment, and financial for the project's evaluation.

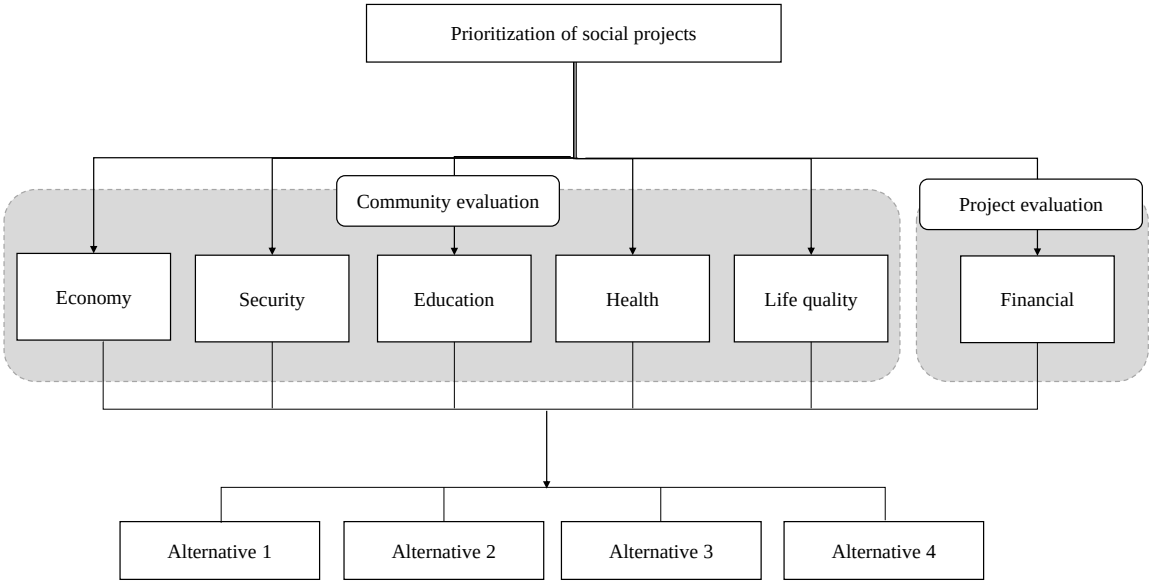
The solar resource as input in the electricity generation was the criterion for choosing alternatives because this research is part of a larger project that seeks to promote solar energy projects; therefore, it is essential to offer elements to create a value chain around the solar energy in Nariño. The following were the selected projects from PERS'alternatives:

- School electrification: electrification of 256 schools located in 11 Pacific zone municipalities using solar PV technology. The objective of the project is to put into operation the computation room and in the future, an ethernet network in each school. To achieve this, the project seeks to install solar power plants with a capacity between 1,5 to 2 kW (Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE, 2014b).

- Rural electrification: electrification of *El Sande* rural community using solar kits. The community is composed of 364 inhabitants who live in 54 dwellings. As the information did not let us know if the community had a distribution network; therefore, it proposed using a 2 kW-solar kit with a battery system offering a three days autonomy because this is an off-grid area (Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE, 2014c).

[6] 1 COP $ is equal to 0,00027 US$

Figure 2-2. Scheme for prioritization of social projects.



- Solar public lighting: the project proposes to install 213 solar lighting systems every 18 meters along the streets of the municipality of *San Lorenzo*. Each 40 W-solar lighting system equates to 70 W-conventional light system installed. Moreover, each one has a three days autonomy through an energy storage system (Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE, 2014d).

- Water pumping system by solar energy: water pumping systems using solar energy for irrigating crops and providing water for the population in the municipality of *Taminango*. The pump moves 16 m3/day of water to cover the need for irrigation and for human being's consumption. The number of systems considered corresponds to the number of properties between 1 and 5 hectares reported in the institutional databases (2.304) (Universidad de Nariño; Unidad de Planeación Minero Energética; USAID; IPSE, 2014a)

The methodology section's procedure led the authors to normalize each component of Table **2-1** values using Equation 2-1 and calculate the dimensions'indexes using Equation 2-2. The computing deemed information from all 64 municipalities. Table **2-4**'s two first rows indicate the best and worst condition values for each dimension. For instance, the best condition for the education, health, and economy dimensions occurs when the result is one (Oficina de planeación educativa, 2018; DANE & Banco de la República de Colombia, 2016; Departamento Administrativo Nacional de Estadísitica - DANE, 2017, 2018b,c,e,a,d, 2016, 2005; Instituto Departamental de Salud de Nariño, 2018).

Table **2-4** shows the results from the community evaluation in the municipalities linked to the projects. *San Lorenzo* has a better indicator (0,49) than *Roberto Payán* (0,09) in the education dimension. Concerning health, *Tumaco* presents better conditions (0,69) than *Santa Cruz* (0,13). However, *Tumaco* has

the worst security conditions in the state (1,00). In the life quality dimension, the population of *La Tola* presents the worst situation (0,95). *Tumaco* has the highest score (0,53) in economic importance because it is a harbor, and *Taminango* shows the worst position (0,19). These results show that the municipalities located in the state's Pacific zone are the most affected in the evaluated dimensions.

Table 2-4. Municipalities evaluation results.

	Education	Health	Security	Quality of life	Economy
Better condition	1	1	0	0	1
Worst condition	0	0	1	1	0
Municipality	Education	Health	Security	Quality of life	Economy
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
Best indicator					
Worst indicator					

The NPV was the chosen criterion for the projects' evaluation. In this step, the authors made an economic analysis of each project, considering all the investment costs and a discount rate of 10 % evaluated in six years. Table **2-5** shows the normalized results of the financial evaluation of the projects. The rural electrification and solar public lighting projects have the same normalized value, although the NPV is slightly different. It is valuable to point out that the water pumping project using solar energy is the most expensive because it would cover 2.304 beneficiaries, corresponding to the total rural properties.

Table 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

The following step was to compute the weights for each dimension with the AHP method In this case, since there is no expert panel to determine the importance of one dimension over another, six prioritization objectives are presented, named from O1 to O6. Each objective emphasizes one dimension over

the others and Table **2-6** shows the results of the weights.

Table 2-6. Weighting by prioritization objective.

	O1	O2	O3	O4	O5	O6
Education	0,083	0,083	0,083	0,083	0,083	0,583
Health	0,083	0,083	0,083	0,083	0,583	0,083
Security	0,083	0,083	0,083	0,583	0,083	0,083
Quality of life	0,083	0,083	0,583	0,083	0,083	0,083
Economic	0,083	0,583	0,083	0,083	0,083	0,083
Financial	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

Table **2-7** presents the results of the projects' prioritization by objectives. The schools'electrification is the main alternative from the prioritization objectives related to the financial (O1), life quality (O3), security (O4), and education (O6) objectives. The results are coherent because the municipalities involved in this project showed the lowest indicators in the dimensions considered. If the objective is to improve life quality, security, and education, the projects crossing these dimensions are the school's electrification ones. Moreover, if the available budget determines the decision, the school's electrification is also a good option because the project's normalized NPV value is the third option, which means that the project is not the most expensive one.

The objectives related to the economic (O2) and health (O5) issues prioritize the electrification project for the indigenous reservation. These prioritization objectives highlight the importance of the objective population's economic and health dimensions over the other dimensions. The collected information shows that *Santacruz* community has the worst health condition between the communities evolved in the evaluated alternatives (see Table **2-4**). In the economic dimension, the indicators show that the *San Lorenzo* and *Santacruz* communities are in the same conditions; however, *Santacruz* had worse conditions than *San Lorenzo* in security, health, and education dimensions. It let us conclude that the electrification improves the economic and health conditions of the population. Results are consistent because electricity powers working tools, electric devices like refrigerators to preserve meals, and technological entertainment devices.

For all prioritization objectives, the last project in importance is the water pumping using solar energy. As the community evaluated concerning this project has better indicators than other communities, the results were coherent except in the economic dimension. The public lighting project using solar energy is the third priority since the evaluated community for such a project has better conditions than some other evaluated communities in the education and security dimensions.

Table 2-7. Projects prioritization.

Project	O1	O2	O3	O4	O5	O6
Solar pumping	4,66	4,56	4,38	4,09	4,44	4,40
Rural electrification	5,74	5,63	5,57	5,18	5,67	5,62
Solar public lighting	5,58	5,47	5,32	4,97	5,34	5,32
School electrification	5,76	5,59	5,66	5,54	5,55	5,69

The first in importance are in dark gray

The school electrification projects are a priority in four proposed prioritization objectives. The authors did a more detailed analysis of this project to understand why this project is a priority in the state. The school electrification project was disaggregated in the project by the municipality, and each project was evaluated with the same weighting of Table **2-6**.

The project's benefits quantification method followed the suggestions from the Manual of Benefits Evaluation and Quantification by the National Planning Department (Dirección de Inversiones y Finanzas Públicas, 2006). In the manual suggests the following assumptions considered in this research:

- The school electrification increases both access to information and education quality.
- Reduction of school dropouts and a higher probability of reaching middle education level by students.
- Increase of the future incomes of at least one legal minimum wage.

Under the last considerations, the project's benefits include at least the annual income for who finishes the middle education level and recover about ten laptops per school that were disused because there was no electricity.

Table **2-8** shows the project's assessment results stemmed from the NPV and Internal Return Rate (IRR) for each municipality. The results agreed with the scale economies, since the largest project had the highest IRR, while the smallest project had even a negative IRR. Therefore, the results presented the break-even point with the execution of at least ten schools onwards.

Table **2-8** shows the project's assessment results stemmed from the NPV and Internal Return Rate (IRR) for each municipality. These results agreed with the scale economies, since the largest project had the highest IRR, while the smallest project had even a negative IRR. Likewise, the results identified the break-even point with the execution of at least ten schools onwards. The same Table also exhibits the prioritization objectives results for each municipality of which there is a noticeable result. The school electrification project's analysis from all prioritization objectives identifies to the municipality of Tumaco because of offering the lowest results respect to other municipalities. The last is concordant with the municipality's socioeconomic issues that put it in a better position than the others in the Pacific zone; consequently, this project had the highest IRR, as Table **2-8** shows.

Although in the O1, the most relevant factor is the financial one, the project execution order is not equal to the NPV standardization order. It occurs because the project evaluation includes other decision variables as health, education, security, and others. When the economy is the leading dimension, as O2 shows, the school electrification project in *Cumbitara* achieves the first place because this municipality has one of the lowest economies among the evaluated communities. Therefore, the execution of this project would improve the economic conditions in this area. In O4, when the security dimension leads the analysis of school electrification, the municipality of *Barbacoas* is a priority because it has the second worse condition in this dimension after the municipality of *Tumaco*. As *Tumaco* has better conditions in the other dimensions than *Barbacoas*, this took the first position for O4.

Table 2-8. Financial evaluation of the school electrification projects.

Project	Number of schools	Normalized NPV	IRR	O1	O2	O3	O4	O5	O6
Tumaco	53	0,00	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	9,60	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	1,00	-3,43	5,743	5,500	5,695	5,198	5,597	5,655

Table **2-9**'s cells contain a number meaning the same ranking in the executing order for both prioritization objectives. This correlation matrix is a useful tool for decision-makers who should determine the merging of objectives. Another characteristic is that the matrix's diagonal corresponds to the evaluated municipalities. The most outstanding result is the correlation between O3 (life quality) and O6 (education); this means that both objectives have the same ranking in executing the alternatives. In other words, to improve education to increase life quality and vice-versa.

Table 2-9. Correlation matrix in the school electrification projects by dimensions

	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

Furthermore, when education improves, the population's economy also progresses, as confirmed by the correlation between O2 and O6. This result also confirms the assumption to compute the IRR where investments in education increase the population's incomes. Accordingly, detailed research of these correlations would support the sectoral initiative formulation for the education and economy sectors, impacting the life quality and boosting social development. For future research works, it could be interesting to insert more dimensions about communities and projects to contribute to the formulation of broader socio-energy policies.

2.4.2 Guideline for community work

Taking as a basis the co-construction methodology, Nariño's population information derived from public entities, and field research in the Huatacondo village, the authors propose a three-stage guideline

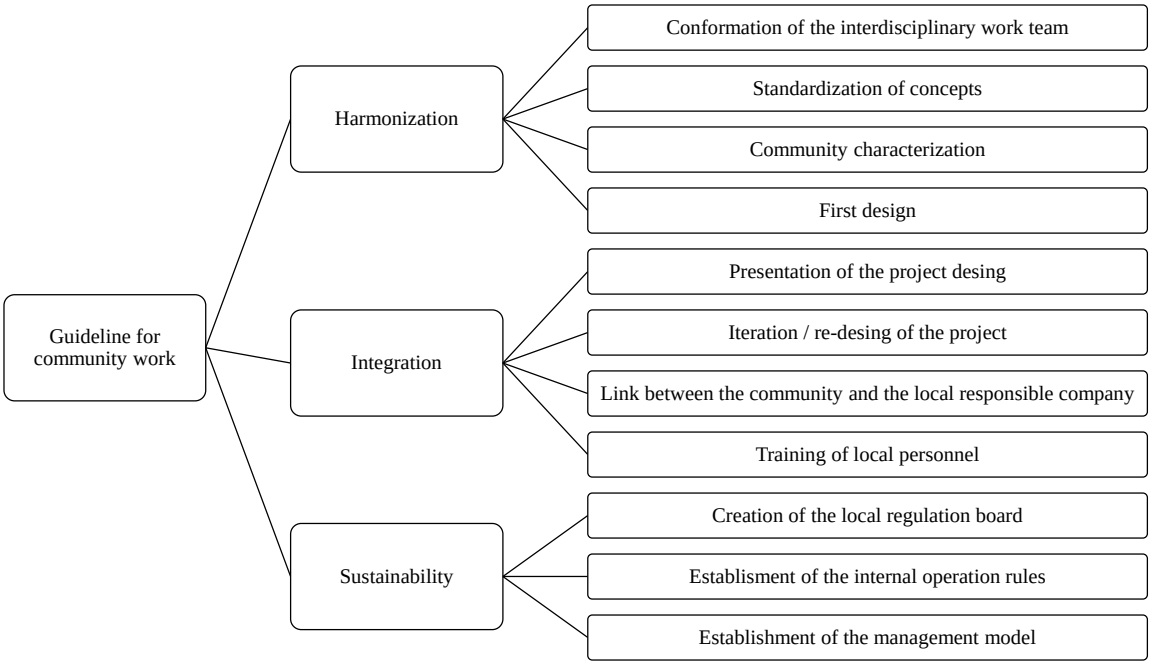
for promoting the joint work between developers or funding entities and communities. Figure **2-3** synthesizes the guideline for impelling solar PV technology projects for the electrification.

The first step in the harmonization stage focuses on forming the interdisciplinary working group to define the energy project's objective. The second step searches to standardize concepts and terms to talk the same language among the working group's members. This step is essential because the group's members may have different ideas about successful, sustainable, and pertinent projects, and positive and negative impacts stemmed from them. Consequently, before designing a solution or interacting with the community, it is crucial to clarify both the objective and the project's expected impacts. The joint work strengthens the project because the population would know its purpose and impacts; likewise, it would state their concerns. The third step consists of characterizing the community from the point of view social, economic, cultural, among others. This characterization stands out to the community's stakeholders and recognizes the community's needs to offer suitable solutions. The first stage's final step consists of drawing up the project's first design considering a technical, economic, and social view. Using primary or secondary information related to the study place is vital.

The integration stage begins with the presentation of the project's design to the community. This socialization is essential because it arouses a dialog around the project's crucial aspects that could be troubling. Those spaces should advertise the changes that eventually would re-design the project, including opinions and proposals from people. The aim is to achieve a consensus; consequently, the process is repetitive; namely, it occurs several times until achieving such consensus. Although it is not necessary to hold technical debates, the community must feel that it is a fundamental part of the decision-making process to formulate the project. The third step searches the executing company. Ideally, the company should be local because it eases the project execution, the maintenance operations, and the relationship with the community. However, when there are no local companies to develop the project, it is suggested to involve a local supervisory body that may be constituted by academy members, as a university research institution, local population, among others, to supervise impartially activities developed by the executing company. The last step trains local staff to perform primary operations and maintenance activities. This guideline intends to minimize the foreign staff dependence to carry out routine tasks and reduce operation and maintenance costs.

The sustainability stage is the last. It seeks to maintain the project over time. Therefore, communication among the funding entity, the executing company, and the community is constant. The first step consists of creating a local regulation board; conveniently, this entity must be legitimate for the community and act according to its customs and culture. In the second step, the national level's regulatory entities would help the local regulatory board establish concordant rules, rights, and obligations with the national regulation and laws. This step is imperative because several troubles would be solved transparently. The final step creates a management model for long-term operation in which the company and the community have responsibilities because they would part of the organizational structure.

Figure 2-3. Guideline for community work



2.5 Discussion

This paper proposed a methodology for the selection of projects and its implementation in a community. Additionally, the prioritization methodology evaluates relevant areas and the cost of implementing solar-energy-based. The criteria chose for community evaluation consider the SDGI and the HDI and the NPV results were the criteria for project evaluation. Also, the prioritization presents a combination of those indexes, which were the criteria used in the AHP methodology to give an execution order. Furthermore, the guideline for community work emphasized the integration of the community in the design and operation of the implemented solution, providing the project with a higher probability of success. In this case, the authors considered the characteristics of Nariño and identified relevant areas for the community evaluation by municipalities. The dimensions chosen were education, health, quality of life, economy, and security. However, it is convenient to include more dimensions to understand in a better way the conditions of the population and include cultural aspects of the communities.

For the project evaluation, the authors consider the NPV, giving more importance to those with lower costs. However, this lone indicator does not allow for the recognition of social benefits that may arise from the project. Hence, in future work, the use of additional quantifiers that measure the economic benefit that arose from social projects could motivate the attention of potential investors.

The alternatives evaluated with this methodology were four projects, along with the communities involved. The results show that in four of six prioritization objectives, the electrification of schools is the priority, and in two objectives, the priority was the electrification of an indigenous reservation. Given

the above results, the authors decided to carry out a more detailed analysis of the project of electrification of the schools. The analysis showed that if the IRR is a decision variable, it is better to implement larger projects; in this case, more than ten. On the other hand, the results showed a correlation between education and the quality of life because O3 and O6 have six projects in the same order of execution; therefore, the accomplishment of education projects improves the quality of life and vice-versa. In this sense, it would be convenient to undertake more studies that confirm this finding and thus unify efforts in the public policies in these sectors to strengthen the results

The researchers expected that, with the implementation of this prioritization methodology, it would be possible to identify the optimal execution order of social projects according to a given objective. Additionally, with the guideline, the objective is to make the community an active part of the solution, to improve the sustainability of the project and avoid or reduce the dependence on an external entity, and thus empower the community and promote their social, economic, education and cultural development, among other aspects. It is also essential to highlight the primary role of the local regulation board in the success of the project. Therefore, future research should analyze aspects such as regulation, election mechanisms of the board members, among others, deeply.

3 Assessing empirical models for estimate global solar irradiance using air temperature in tropical and mountainous environment. Part I: imputation[1]

Abstract

Solar irradiance is an available resource that could support electrification in regions with low socio-economic indices. Therefore, it is increasingly necessary to understand solar irradiance behavior and data. However, some places, especially those with low socio-economically impoverished population, do not have solar irradiance data, and if such information exist, it is likely incomplete. Therefore, researchers estimate this energy resource using information from other meteorological variables, such as temperature. Nevertheless, a broad analysis of these techniques in tropical and mountainous environments has yet to be conducted. As such, current research analyzes the performance of three well-known empirical temperature-based models in tropical and mountainous environments. Moreover, this work proposes a new empirical relationship that can models solar irradiance in some areas better than the other three models. Statistical error comparison permints the selection of the best model for each location and the best data imputation model. Hargreaves and Samani's model had better results in the Pacific zone, while the proposed model demostrates better results in the Andean and Amazon zones. Another significant result was the linear relationship between the new empirical model constants and the altitude above 2.500 MASL.

Keywords

Solar irradiance, Data imputation, Bristow and Campbell, Hargreaves and Samani, temperature-based empirical models.

[1] This article had the participation of Belizza Ruiz

3.1 Introduction

It has taken decades before solar energy becomes economically feasible in developing countries. Today, the associated costs of harnessing this energy resource are competitive in certain situations such as rural and isolated electrification. Isolated and rural areas for which electricity could be supplied by expanding electricity transport systems over areas with low-population density discourage investments because such as expansion would require an electricity demand sufficient to turn a profit. Despite these limitations, the State must ensure access to electricity services for its entire population, which could be achieved via non-conventional energy sources, such as solar energy. Before doing so, however, it is essential to understand the behaviors of such sources. In developing countries, solar irradiance data are often not available because of the scarcity of weather stations that measure this variable, and the insufficient or incomplete calibration and maintenance of metering equipment. In some cases, even when open-access databases are available, the actual acquisition of such information might be difficult (Bakirci, 2009). Therefore, in recent years, solar irradiance studies have increasingly relied on empirical methods that incorporate relevant variables, such as temperature, sunshine, and relative humidity, to estimate solar irradiance.

Researchers have developed models and methods, from empirical models to artificial intelligence, to estimate solar insolation for different time frames. The simplicity, acceptance, adaptability, and low computational cost are the advantages of the empirical models that led us to analyze them. Most empirical models in this regard are based on astronomical, geometrical, physical, and, especially, meteorological factors (Besharat et al., 2013). Meterological factors include temperature, sunshine duration, and cloudiness, the latter of which is particularly important for understanding solar irradiance behavior. Of these factors, sunshine duration and temperature are typically recorded in relevant databases (Benson et al., 1984), and are consequently implemented most often to estimate solar insolation (Bakirci, 2009). As the implementation of an effective empirical model depends on data availability and consistency (Akinoglu, 2008b), some researchers have recommended revising data simultaneity and reliability at study sites (Almorox et al., 2011). Unlike temperature data, which are recorded by weather stations, sunshine duration data are often not available. Therefore, temperature-based empirical models are a more convenient option for estimating global solar radiation (Fan et al., 2018).

In 1982, Hargreaves and Samani presented the first temperature-based model for estimating solar insolation based on daily temperature differences (Hargreaves & Samani, 1982). This was followed, in 1984, by a temperature-based empirical model, developed by Bristow and Campbell, that focused on difference between the daily maximum and minimum temperatures (Bristow & Campbell, 1984).In 20011, Chen et al. evaluated the performance of the temperature-based models in China, they found that support vector machine models using maximum and minimum temperatures as input and polynomial kernel function outperform empirical models (Chen et al., 2011). In 2014, H. Li et al. presented a temperature-based model for China that built upon the Hargreaves and Samani model (H. Li et al., 2014). Quansah et al. evaluated temperature- and sunshine-based empirical models in Ghana (Quansah et al., 2014), and C. M. Dos Santos et al. assessed nine temperature-based models for Brazil (C. M. Dos Santos et

al., 2014). In 2017, Rivero et al. validated the Hargreaves and Samani model for Mexico (Rivero et al., 2017), and Jamil & Akhtar compared empirical models for India's subtropical and humid environments (Jamil & Akhtar, 2017). Although several studies analyze the empirical models' behavior in different places globally, there is no the same amount of research analyzing the efficacy of temperature-based models for tropical and mountainous environments.

Toward this end, the purpose of this research was to assess three empirical temperature-based models-Hargreaves and Samani (HS), Bristow and Campbell (BC), and Okundamiya and Nzeako (ON)-with respect to their capacity to estimate solar insolation in a tropical and mountainous environment. To achieve this goal, statistical validation determines and compare the performance of the three models. Aditionally, the authors propose a new method based on logistic regression model, to estimate solar insolation according to daily temperature difference. The data used as input for these empirical models came from IDEAM's[2] Automatic Weather Stations (AWS) located in the state of Nariño, which has three environmental zones: Pacific, Andean, and Amazonian. This allowed for the evaluation of the models in different weather and physiographic conditions. The database was randomly divided into two parts: the first for calibrating the models, and the second for validating the models statistically. Before using the empirical models, the data passed a quality control procedure to improve the results' reliability. In this regard, R-CRAN was the software used to carry out the computational process needed for the research (R Core Team, 2020).

This work has six sections organized as follows: Section 3.2.1 presents a description of the case study. Section 3.2.2 Data quality control for both solar irradiance and temperature data, Section 3.2.4 Empirical temperature-based models, Section 3.2.5 contains the proposed model, Section 3.2.6 Statistical validation of empirical models. Section 3.3 shows the results and discussion. Finally, Section 3.4 presents the conclusions.

3.2 Materials and methods

3.2.1 Site and dataset

The use of solar energy in production processes requires precise knowledge of solar irradiance behavior in specific locations, including weather patterns, in addition to other factors, such as orographic characteristics and socio-economic population data pertinent to the potential impacts of solar energy. For instance, Nariño's location ($00°31'08''N$ and $02°41'08''N$ latitude and $76°51'19''W$ and $79°01'34''W$ longitude) favors the harnessing of solar energy, as the region receives considerable and consistent solar irradiance year round, with little variation compared to other latitudes. This is notable because, once installed, solar power systems would not need trackers to reach the maximum power point, thereby reducing initial investment and maintenance costs and facilitating the easier installation of additional

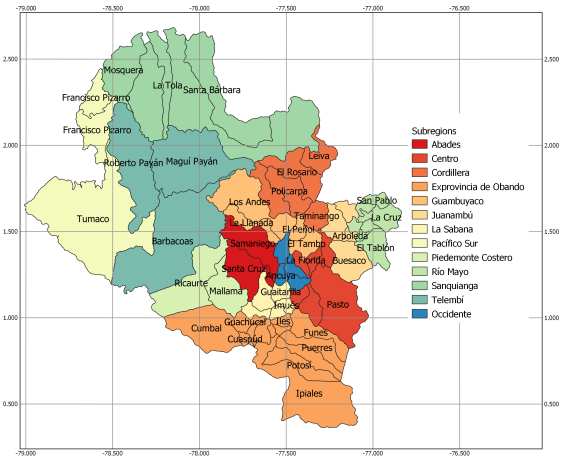
[2] Institute of Hydrology, Meteorology and Environmental Studies, IDEAM by iys Spanish name

solar power systems.

Despite having low solar irradiance variability, the State of Nariño is located in an Intertropical Convergence Zone given its close proximity to the Equator, which in turn influences weather behavior, causing unimodal or bimodal rainy seasons and increased cloud cover due to high humidity in low altitudes. The latter occurs because the region's mountain ranges retain the moist air masses coming from the Pacific Ocean. More specifically, when incoming air masses collide with the western mountain range, the Pacific foothills become more humid and tend to maintain this humidity (Instituto Geográfico Agustín Codazzi - IGAC, 2014). Hence, it is appropriate to divide the weather and orographic conditions of Nariño into three environmental regions: the Pacific, the Andean, and the Amazon (Martínez, 2018).

The Pacific region covers 52 % of the total state territory and is costituted by *Telembí*, *Pacífico Sur*, *Sanquianga*, and *Piedemonte Costero* administrative subregions (see Figure **3-1**). This region includes the *Mira-Mataje* binational watershed, and a mangrove forest located in the Sanquianga natural reserve. Besides, it has two climatic zones: the Pacific flatlands and the Pacific foothills. The first one has humidity higher than 80 %, temperatures higher than $26°C$, and the precipitation range is between 3.000 and 5.000 mm/year. The second one presents a high humidity, temperature between $18°C$ to $24°C$, and precipitations between 4.000 and 6.000 mm/year. Nonetheless, there is a high precipitation zone located between *Junín* and *Barbacoas* localities with a 9.000 mm/year level (Gobernación de Nariño, 2016a; Instituto Geográfico Agustín Codazzi - IGAC, 2014).

Figure 3-1. Administrative subregions of the Nariño state.



The second zone is the Andean region, constituted by *La Sabana, Los Abades, Occidente, Cordillera, Centro, Juanambú, Río Mayo*, and *Guambuyaco* administrative subregions (see Figure **3-1**). In this region, the Andean mountain range is divided into the western and central mountain range, covering approximately 40 % of the area of the State (CORPONARIÑO, 2001; Instituto Geográfico Agustín Codazzi - IGAC, 2014). In the western mountain range are the *Chiles, Cumbal,* and *Azufral* volcanos. In the central mountain range are the *Galeras* and *Doña Juana* volcanos, and also, the *Túquerres-Ipiales* high flatlands, as well as the *Atriz Valley*. This region shows a bimodal precipitation behavior, between 800 and 2.200 mm/year, with peaks in April-May and October-November. In the *Río Mayo, Juanambú*, and *Guaitara* watershed, there is a temperature between 16 to 24 $°C$ and precipitation of 1.000 to 1.800 mm/year. In the north zone towards *Patía*, the annual precipitation is less than 1.000 mm/year, with a temperature above 24 $°C$ (Gobernación de Nariño, 2016a).

The third region is the Nariño Amazon watershed, which takes 8 % of the territory composed by the *Exprovincia de Obando* administrative subregion. About 16,4 % of the state population lives in this area. This region has two climate zones: the mountainous and the flatland zone. The first one, located between

the *Patía* and *Putumayo* rivers, present temperatures between 6 $°C$ to 11 $°C$, and receives precipitation about 2.000 mm/year. The second one has tropical weather influenced by the cloudy jungle, and present precipitation between 500 to 1.500 mm/year (Gobernación de Nariño, 2016a). It is convenient to highlight that the mining activity is a significative gainful activity; in fact, about 98,38 % of the State's minerals come from this region (Instituto Geográfico Agustín Codazzi - IGAC, 2014).

In the State of Nariño, eight AWS measure solar irradiance (see Table **3-1**). Figure **3-2** shows the location of automatic and conventional weather stations in the State. . In the Pacific region, there are three AWS located in the State of Nariño and one in the neighboring State of Cauca. The AWS's altitude range is between 16 and 512 MASL. In the Andean region, there are five AWS with altitudes between 1.005 and 3.120 MASL. Finally, in the Amazon region, there is one AWS at 3.577 MASL. The maximum altitude difference between all AWS is 3.561 MASL. Such difference allows evaluating the solar irradiance behavior in a wide range of altitudes and three diverse environmental regions.

Figure 3-2. AWS location.

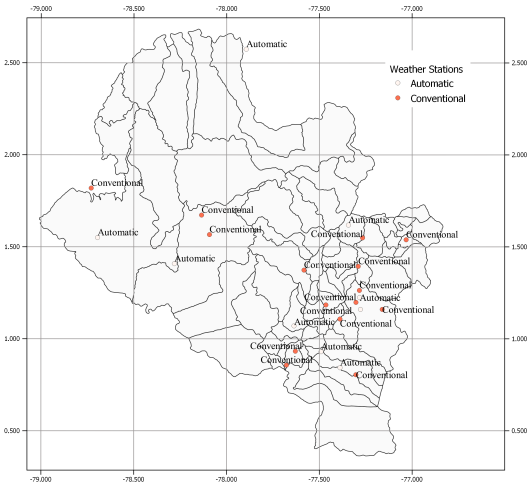


Table 3-1. Automatic Weather Stations.

Name	Latitude [°]	Longitude [°]	Altitude [MASL]	Period	Region
Biotopo	1,41	-78,28	512	2005-2017	Pacific
Altaquer	1,56	-79,09	1.1010	2013-2014	Pacific
Granja el Mira	1,55	-78,69	16	2016-2017	Pacific
Cerro-Páramo	0,84	-77,39	3.577	2005-2017	Amazon
La Josefina	0,93	-77,48	2.449	2005-2017	Andean
Viento Libre	1,62	-77,34	1.005	2005-2017	Andean
Universidad de Nariño	1,23	-77,28	2.626	2005-2017	Andean
Botana	1,16	-77,27	2.820	2005-2017	Andean
El Paraiso	1,07	-77,63	3.120	2005-2017	Andean

3.2.2 Solar irradiance data quality control

The data quality control is a procedure intended to improve the reliability of the time series data. The procedure includes the analysis of database structure, the comparison with fixed and flexible limits, and time consistency. These steps have previously yielded reliable results in studies aimed at understanding weather data quality control in Spain (Estévez et al., 2011). The quality control procedure followed in the current research relied on guidelines presented in regulation UNE500540, which outlines successive

analytical procedures of quality control for different weather variables (AENOR, 2004).

1. The first step consists of checking the database structure. In our case, the data exhibited the following structure: AWS code, weather variable code, date, hour, and value. This step only maintains the values with the described structure (Rivero et al., 2017).

2. The second step, the fixed-range step, UNE500540 suggests using the most restrictive condition between the measurement device limit and the physical phenomenon limit. The physical limit is the maximum extraterrestrial solar irradiance of each location computed with Equation 3-1:

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

$$sin\beta = cos\phi cos\delta cos\omega_s + sin\phi sin\delta \tag{3-2}$$

where D is the Julian day, I_{sc} is the solar constant $(1.367\,[W/m^2])$ representing the energy from the Sun per unit area of the surface perpendicular to the irradiance propagation direction (Şen, 2008), I_0 is the maximum extreterrestrial solar irradiance, β is the solar altitude, ϕ is the latitude of the site, δ is the solar declination, and ω_s is the hour angle. In this case of study, AWS have pyranometers Kipp & Zonen CMP11 with an upper operation limit of $4.000W/m^2$ (Kipp & Zonen, 2000). The most restrictive condition is the last, the calculated global solar irradiance in the time (AENOR, 2004). Consequently, $I_0 \geq I_{mt}$, where I_{mt} is the measured global solar irradiance at the time t.

3. The third step, the flexible-range step, UNE500540 suggests comparing the time series with the maximum and minimum values of a time series validated. In this case, there is no time series previously validated; therefore, we used, as an alternative, the restriction proposed for Estévez et al. with a modification. Namely, instead of using I_0, they use the following range $0,03I_{cst} \leq I_{mt} \leq I_{cst}$, where I_{mt} is the measured global solar irradiance at time t, and I_{cst} is the clear-sky global solar irradiance at time t

$$I_{cs} = I_0\tau. \tag{3-3}$$

Considering that the clear-sky global solar irradiance I_{cs} is equal to τ times I_0 and τ is the atmospheric transmittance. The first step is to use the Kreith & Kreider model to estimate the atmospheric transmittance and then calculate I_{cs} with Equation 3-4 (Şen, 2008)

$$\tau = 0,56\left(e^{-0,65m} + e^{-0,095m}\right) \tag{3-4}$$

with:

$$m = \frac{1}{sin\beta} \tag{3-5}$$

where m is the air mass.

4. The fourth step, the time-consistency step, consists of analyzing the changing hour by hour of the global solar irradiance. This analysis follows the restriction $\left|I_{cs_t} - I_{cs_{t-1}}\right| > \left|I_{m_t} - I_{m_{t-1}}\right|$. It is a useful test for detecting data storage and connection troubles in the datalogger. A high sampling frequency, for instance, every ten minutes, would increase the effectiveness of the test.

Aside from the tests mentioned above, it is also necessary to consider the zero offset, which results from thermal imbalances in the pyranometer. This phenomenon occurs because the sensor does not absorb any measurable irradiance values in the pyranometer's spectral range, resulting in erroneous values (Kipp & Zonen, 2000). It is fundamental to offset this unbalance because neglecting it would entail underestimated solar irradiance records between 0,7 % to 4,3 %. Despite their importance, several environmental factors influence the measurement process, making it difficult to correct it for all measurement instruments in all locations and environments (Serrano et al., 2015). Although various approaches to correct the zero-offset adapted to specific environmental conditions and instruments exist, the current research did not follow these approaches.

3.2.3 Temperature data quality control

The validation of temperature data follows the Estévez et al.; Rivero et al. recommendations. The quality control of temperature data comprises five validation steps: structure, range, step, consistency, and persistence.

1. The first step consists of verifying the database structure, likewise that the solar irradiance validation.

2. The second step, the range-test step, involves two evaluation methods: the sensor range method and a data validation criteria defined by other researchers method. In this case, the sensor is a ROTRONIC HYGROCLIP 2 RTD PT100 with a range between $-40°C$ and $60°C$. Rivero et al. recommend a temperature interval of between $-50°C$ and $70°C$, while Estévez et al. suggested a range of $-30°C$ to $50°C$. We followed the Estévez et al. recommendation because it is inside the temperature instrument range.

3. The third step, the step test, requires the fulfillment of the following requirements: $|T_h - T_{h-1}| < 4$; $|T_h - T_{h-2}| < 7$; $|T_h - T_{h-3}| < 9$; $|T_h - T_{h-6}| < 15$; and $|T_h - T_{h-12}| < 25$, where T_h is the temperature at hour h. Although this evaluation does not consider other climatology aspects that can affect temperature variation, like humidity, wind speed, cloudiness, and precipitation, we applied these restrictions. Another fundamental requirement is that the daily step restriction fulfills the following condition $T_{max} - T_{min} < 30°C$ (Herrera-Grimaldi et al., 2019), where T_{max} is the maximum daily temperature, and T_{min} is the minimum daily temperature.

4. The fourth step, internal-consistency test, assesses the accomplishment of the following conditions: $T_{max} > T_{mean} > T_{min}$; $T_{max}\,(D) > T_{min}\,(D-1)$; $T_{min}\,(D) \leq T_{max}\,(D-1)$, where

T_{mean} is the mean daily temperature.

5. The fifth step, persistence test, verifies the measurement variability; therefore, the data must accomplish the following conditions: $T_{max}\,(D) \neq T_{max}\,(D-1) \neq T_{max}\,(D-2), T_{min}\,(D) \neq T_{min}\,(D-1) \neq T_{min}\,(D-2)$.

The analysis for the step test, internal consistency test, and persistency test followed the Estévez et al. quality control process because it offered promising results in data analysis in Spain (Estévez et al., 2011; Herrera-Grimaldi et al., 2019).

3.2.4 Empirical temperature-based models

Although the air temperature is a standard variable in the weather stations, it was no frequent to estimate solar insolation using this variable. The temperature started to be relevant to estimate solar insolations when agricultural studies modeled solar insolation to analyze the crop production rates. Consequently, researchers from other knowledge fields as solar one paid attention to the maximum, minimum, and mean temperature values to model their processes (Paulescu, 2008). The most traditional models using the maximum and minimum temperatures are the Hargreaves and Samani, and Bristow and Campbell models (Besharat et al., 2013; Bristow & Campbell, 1984). Both models are the basis of innovative approaches adapted to the location conditions.

Before presenting the empirical models, it is necessary to introduce the daily extraterrestrial solar insolation on a horizontal surface H_0 math expression because this is part of most empirical models:

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

The solar declination δ and the sunset hour angle ω_s are calculated with Equation 3-7 and Equation 3-8, respectively:

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

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

Hargreaves and Samani model

Hargreaves and Samani (HS) proposed a linear relationship between the temperature difference's square root and the fraction between the extraterrestrial and terrestrial solar insolation for different periods in 1982,

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

where H is the global solar insolation on a horizontal surface, and a is the empirical coefficient. This model did not consider the effects of cloudiness, humidity, latitude, elevation, and topography, among others, in the specific location for which the model is used (Samani, 2000; Allen, 1997) Allen stated that the HS's model has better behavior in a monthly time-frame than daily because the variables follow a mean trend, resulting in a consistent relationship between $T_{max} - T_{min}$ and H/H_0 (Allen, 1997).

The empirical coefficient a represents the change rate from the maximum and minimum temperatures difference with the ratio between the extraterrestrial and terrestrial solar insolation. Initially, HS proposed an empirical coefficient calibrated with an eight-year time series from Central Valley in the Davis County (California), see Table 2 (Besharat et al., 2013; Hargreaves & Samani, 1982). HS presented a relevant conclusion about weather elements affecting this coefficient. They found that relative humidity above 54 % affects the empirical coefficient, and the maximum and minimum temperatures difference is smaller when it is nearly this percentage. As shown in Table **3-2**, humid climates have minor values than a tenth while the other regions have higher values than a tenth (Hargreaves & Samani, 1982). A thorough process uses empirical coefficients according to the region under study. Accordingly, Rivero et al. pointed out that using a fixed empirical coefficient for a large area with different regions may lead to significant errors because topography influences the temperature, advective environment, and vegetation (Rivero et al., 2017).

Table 3-2. HS empirical coefficients.

a	Region type
0,16	Arid and semi-arid
0,17	Interior regions
0,19	Coastal regions
0,10-0,09	Humid climates

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

Bristow and Campbell model

Bristow and Campbell (BC) presented a model based on two assumptions. The first one is the linear relationship between the absorbed (net) and incoming solar insolation, and the second assumption consists of despising the heat flux coming soil for a daily period because on average it is near zero. Sensible and latent heat are components of the absorbed insolation, and the sensible heat produce diurnal temperatures higher than during the nighttime (Bristow & Campbell, 1984). The last is understandable because solar irradiance heat more the air masses due to the short-wave radiation. While during the night, there is less long-wave emission from the Earth to the atmosphere reducing the temperature (Meza F. & Varas E., 2000). Furthermore, under ideal conditions, the temperature tend towards the dew-point during night, and it is minimum just before sunrise, resulting in a significant difference between daily maximum and minimum air temperature. This phenomenon allows modeling solar insolation as a function of temperature difference (Bristow & Campbell, 1984). Under these assumptions, the authors presented their model,

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

where a, b, and c are empirical coefficients, a represents the maximum H/H_0 in the study area, b and c determine how soon the maximum H/H_0 is achieved with ΔT increases. Empirical coefficients rep-

resent the regional characteristics from arid to humid environments (Bristow & Campbell, 1984). The range in daily temperature extremes (ΔT) was calculated as:

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

Temperature difference depends on the daily maximum T_{max}, and minimum T_{min} for a day D given in $(^{\circ}C)$. BC concluded that the minimum temperatures average of two consecutive days reduces the hot or chilly air masses' effects, avoiding overestimated and underestimated solar insolation values. Accordingly, Dos Santos et al. stated that the advection is not a common phenomenon in tropics zones; therefore, the temperature change ΔT estimation in these regions would satisfy the Equation 3-12. Those authors also highlighted that this equation is better for sites with high altitudes, as in the study case (C. M. Dos Santos et al., 2014)

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

Another weather effect affecting the solar insolation estimation is the rain. The rainy effect decreases setting $\Delta T\left(D\right)$ equal to 0,75 times the measured. If $\Delta T\left(D-1\right)$ is less than $\Delta T\left(D-2\right)$ about $(2^{\circ}C)$, the first is multiplied by 0,75 (Bristow & Campbell, 1984). However, when the rainy period is long, the relation between the solar insolation and ΔT reaches an equilibrium. Therefore, it does not require adjustments (Goodin et al., 1999).

Models implemented in a tropical environment

The knowledge of solar irradiance behavior in tropical zones is a mandatory subject for this research. The cases analyzed belong to Africa (Nigeria Abuja, Benin City, Katsina, Lagos, Nsukka, and Yola), Brazil (Água Branca, Pao de Azucar, Santana do Ipanema, Palmeira dos Índios, Arapicara, Maceió, Corcuripe, Sao Jose da Laje), and Mexico. Because Nariño is in a tropical area, the results from those three cases constitute relevant inputs for our research.

Okundamiya & Nzeako (ON) proposed a linear model represented with Equation 3-13 (Okundamiya & Nzeako, 2011),

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

where T_R is the monthly average of the ratio between the daily minimum and maximum temperature T_{min}/T_{max}, and a,b,c are empirical coefficients. The model's statistics validation showed that the coefficient of determination between 0,809 and 0,952.

Nwokolo & Ogbulezie reviewed empirical models implemented in West Africa to compute the global solar irradiance. They found that the soft computing models have better accuracy than the empirical

models since they can be easier to adapt to several weather conditions because they allow more inputs, as models or variables, to strengthen their reliability (Nwokolo & Ogbulezie, 2017).

Dos Santos et al. studied the performance of the ten temperature-based models in Northeastern Brazil. They found that the models did not show significant changes after adjusting them concerning the rainy periods' effects. Dos Santos and colleagues also concluded that the HS model had a better performance in hinterlands and interior lands, and the BC model showed the best performance in humid and coastal zones (C. M. Dos Santos et al., 2014).

Rivero et al. compared the values of the original HS model's empirical coefficients with new values stemmed from that model calibrated for Mexico with local data. A noticeable aspect of that research consists of Mexico's climate zones classification using the Köppen-Geiger system. Furthermore, the authors did another classification based on the clearness index, as shown in Table **3-3**. to support the AWS data. It is useful to remember that the clearness index is the ratio between the terrestrial solar insolation and the extraterrestrial solar insolation. A relevant conclusion showed that regardless of solar irradiance peaks during the day, it is possible to obtain similar clearness index values. Another conclusion noted that fixed empirical coefficients lead to significant error, especially in zones with a temperature difference below 15 $^{\circ}C$. The authors proposed Equation 3-14 to overcome the identified troubles derived from fixed coefficients (Rivero et al., 2017),

Table 3-3. Daily clearness index classification ranges.

K_T range	Day type
$0,00 < K_T \leq 0,20$	Cloudy/Overcast
$0,20 < K_T \leq 0,60$	Partially cloudy
$0,60 < K_T \leq 0,75$	Sunny
$0,75 < K_T \leq 1,00$	Very sunny

Source: (Rivero et al., 2017)

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

where a_{HS} is the calculated empirical value used in the original HS equation, a_1 and a_2 are empirical coefficients of the Rivero et al. model.

3.2.5 Proposed empirical model

The proposed model originated from observing the scatter plot between the clearness index and the daily temperature difference in each AWS. The authors concluded that the logistic model would offer useful results. This model has successfully studied human growth, animal and biological processes, energy use patterns, and atmospheric applications (Agami Reddy, 2011; Moon & Kim, 2020). The logistic regression commonly describes the relationship between binary variables and a predictor (Kleinbaum & Klein, 2010).Although the relationship between extraterrestrial and terrestrial solar insolation does not offer binary results, it varies into a range minor than one showing similar behavior. The use of this technique "arises in estimating relationships in which the dependent variable is continuous, but is limited in range" (Manning, 1996). Manning's statement describes the proposed model's relationship between the dependent and independent variables because the relationship of solar insolations is continuous uniquely for the range defined by the minimum and maximum temperatures. The logistic regression

does not assume that neither variables nor predictors have a normal distribution; this affirmation was another criterion to choose this approach (Harrenll, 2015).

The logistic regression model is part of the Generalized Linear Model (GLM), which describes the relationship between the mean of the dependent and independent variables with a more complex relationship than $y = a + bx$. In logistic regression, which is a statistical method, the coefficients have a similar meaning than in linear regression; namely a is the log-odds of success at $x = 0$, while b is the change in the log-odds of success corresponding to a one-unit increase in x (Casella & Berger, 2002). This regression also assumes a linear relationship between the dependent and independent variables' log-odds (Moon & Kim, 2020).

$$y = \frac{1}{1 + e^{-z}} \tag{3-15}$$

The logistic model comes from the logistic function; see Equation 3-15. y is the predicted variable, z is a linear sum $a + \sum b_i x_i$, where x_i is the i independent variable x, and a and b are constants (see Equation 3-16) (Kleinbaum & Klein, 2010).

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

The temperature change ΔT is the predictor variable. Therefore, the proposed model is the following:

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

3.2.6 Statistical validation

Statistical validation is a mandatory step that allows comparing the predictor model with the real measures to determine its suitability. There are five main validation techniques: subjective assessment, dispersion indicators, overall performance indicators, distribution similitude indicators, and visual indicators (Gueymard, 2014; Mayer & Butler, 1993). The first one consists of evaluating the model by experts. However, these techniques are open to misinterpretation since it is a subjective test with personal bias (Mayer & Butler, 1993). The second one is appropriate when the data have the same time framework, location, and treatment, among other characteristics. This validation technique measures the difference between the modeled and real value. Table **3-4** shows some dispersion measured, expressed in absolute units or percentages (Mayer & Butler, 1993).

Table 3-4. Statistical error measurements

Measurement	Definition	Formulae[2]
Mean percentage error (MPE)	To indicate that is a better model when its value is close to zero, and the ratio of the standard deviation of the measured and computed value is near to one.	$\frac{100}{n}\sum_{i=1}^{n}(p_i - o_i)/o_i$

Table 3-4 continued from previous page

Measurement	Definition	Formulae[2]
Mean absolute error (MAE)	It is the average vertical distance between each predicted and observed point.	$\frac{1}{n}\sum_{i=1}^{n}\left\|p_i - o_i\right\|$
Root mean square error (RMSE)	It provides a measure of the error size and is sensitive to outlier values because this measure gives much weight on large errors.	$\left[\frac{1}{n}\sum_{i=1}^{n}\left(p_i - o_i\right)^2\right]^{1/2}$
Mean bias error (MBE)	This measure provides information on the long-term performance of the model, when the model has a systematic error that presents overestimated or underestimated predictors. Low values of MBE are desirable, though it should be noted that an overestimated data set will cancel another underestimated data set.	$\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)$
Standard Deviation of the residual (%)	It shows the difference between the standard deviation of the predicted and observed dataset.	$\left(\frac{100}{O_m}\right)\frac{1}{n}\left[\sum_{i=1}^{n} n\left(p_i - o_i\right)^2 - \left[\sum_{i=1}^{n}\left(p_i - o_i\right)\right]^2\right]^{1/2}$
Uncertainty at 95 % (U_{95})	It measures certainty confidence; a lower value is expected.	$1,96\left(SD^2 + RMSE^2\right)^{1/2}$

[2] p_i is the predicted value,o_i is the observed value, O_m is the mean observed value, n is the amount of data

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

The third technique, although similar to the second, evaluates different fields of study to solar energy. This indicator gives information about the model efficiency using greater values that mean better performance. The fourth technique compares the model cumulative frequency distribution with the reference values; this research did not implement this technique (Gueymard, 2014). The last technique shows the simulated and real data features using plots. It is a conventional informative method; nevertheless, the data presentation format could lead to misinterpretations. As a result, a desirable alternative is the observed vs. predicted scatter plots with the linear fit, indicating the perfect adjust. It presents "goodness of fitting as a vertical deviation from the perfect line," indicating some biases (Mayer & Butler, 1993). A suitable visualization method is the Taylor diagram since it combines three dispersion indicators: Pearson correlation, RMSE, and standard deviation (Gueymard, 2014).

3.3 Results and discussion

The results have four subsections. The first one presents the global solar irradiance quality control results. The second one shows the results of the temperature data validation procedure. The third one contains the HS, BC, NO, and the proposed models' empirical coefficients. The last subsection comprises the imputation results and the daily solar insolation for the AWS studied.

3.3.1 Global solar irradiance quality control

A previously step to the quality control procedure is the data adjustment stemmed from the calibration constant. In this case, there is no calibration constant for the *Biotopo* AWS affecting the time series quality; likewise, this AWS had the smallest recordings amount among all AWS. The total data were 47.612 for the period studied, corresponding to 34,50 % approximately. The AWS registered information during 57,59 % of the measured period 2005-2017. Table **3-5** shows the results of the global solar irradiance validation procedure. The first step evaluating the database structure presents 5.843 recordings on average with the incorrect structure. The AWS with more recordings with the incorrect data structure is *Viento Libre*; this step confirmed that 9.715 data did not have the database structure, corresponding to 12,55 % of the total data. The AWS with fewer recordings with the incorrect data structure is *Biotopo*; this step confirmed that 1.055 data did not have the database structure, corresponding to 2,22 % of the total. The fixed range validation discards 36,71 % of data on average. *Biotopo* lost the most data amount corresponding to 47,62 % of the total, and the *Universidad de Nariño* lost the fewest data amount about 22,09 % of the total. The flexible range test results present that there are information losses of 39,11 %. *Biotopo* is the AWS with the lowest losses of about 15,14 %, and *Guapi* has the highest losses of about 54,22 %. Although the last test is not mandatory, it is useful to point out that 44,06 % of the recordings did not overcome this level. Considering only the mandatory steps, about 35,27 % of the data overcome these validation steps. The *Universidad de Nariño*'s AWS had the most data amount approving the validation process, and the *Guapi*'s AWS had the fewest data amount approving the validation process.

Table 3-5. Solar irradiance validation results

AWS Name	AWS Code	Number of Data	Data Step 1	Data Step 2	Data Step 3	Data Step 4
Biotopo	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
Average		79.493	73.651	47.510	28.208	15.519

Table **3-6** presents the amount of days classified by the number of recordings between 6:00 and 18:00. The AWS has mainly days with 10 and 11 measurements per day. It is worth noting that on average just 1,26 % days had complete information (13 measures per day), with a maximum of 4,25 % in *Cerro Páramo* and a minimum of 0,28 % in *Biotopo*. Consequently, the model implementation only considered the days with at least six recordings during the daytime period to avoid sub estimating the resource-approximately 95,81 % of the recordings that overcome the mandatory validation levels allowed daily aggregation. Results confirm the importance of improving the measurement instruments' maintenance and calibration procedures to increase the useful information amount and, in turn, raise the reliability.

Table 3-6. Number of days classified by irradiance measures

	Number of measures per day													Total of
AWS Name	1	2	3	4	5	6	7	8	9	10	11	12	13	days
Biotopo	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

We considered the Rivero et al. classification and modified the item about partially cloudy days, as shown in Table **3-7**. Table **3-7** shows that on average there is about 64,70 % of partially high cloudiness days in the analyzed AWS. As a result, there are losses between 20 % and 40 % of the extraterrestrial solar irradiance when hitting the ground level. It is a relevant result because clouds affect global solar irradiance, and it is a variable that is not easy to model. Consequently, the resource estimation in this tropical and mountainous environment is more complicated than in other environments

Table 3-7. Days classification with the clearness index

K_t	$0,00 < K_t \leq 0,20$	$0,20 < K_t \leq 0,40$	$0,40 < K_t \leq 0,60$	$0,60 < K_t \leq 0,75$	$0,75 < K_t \leq 1,00$	
	Number of days					
AWS Name	Cloudy	Partially high cloudiness	Partially low cloudiness	Sunny	Very sunny	Amount of days*
Biotopo	760	**1.188**	67	0	0	2.015
Viento Libre	105	**1.458**	1.179	0	0	2.745
Universidad de Nariño	243	**2.611**	846	1	0	3.701
Cerro Páramo	**1.380**	1.232	137	1	0	2.793
La Josefina	88	**991**	271	1	0	1.351
Botana	451	**2.704**	711	2	0	3.868
El Paraiso	167	**1.859**	543	1	0	2.570
Guapi	130	**746**	125	0	0	1.001

*Days with $\Delta T, T_R, T_{max}, T_{min}, K_t$ complete information

3.3.2 Temperature data quality control

Table **3-8** shows the results of the temperature validation procedure. In the first step, there were information losses of 8,64 % on average. *Cerro Páramo* is the most critical AWS with 16,95 % of missing data. In the fixed range test, *Guapi* had data losses of about 11,62 %. In the step test, about 35,07 % of the data did not pass the validation requirement. Taking as a base the starting values, 57,08 % of the data overcome the quality control procedure.

Table 3-8. Temperature validation results

AWS Name	AWS Code	Number of Data	Data Step 1	Data Step 2	Data Step 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 Paraiso	52055220	91.699	85.713	84.792	54.371
Guapi	53045040	84.371	81.014	71.598	49.039

Table **3-9** shows the number of days by daytime temperature measurements. Results represent that about 57,42 % of the days had 11 or 12 measured per day (88,46 % of information). Finally, from the total data that overcome the hourly validation test, only 68,03 % was suitable for feeding the models because just this percentage of days have equal or more than 6 measures per day.

Table 3-9. Amount of days by daily temperature measurements

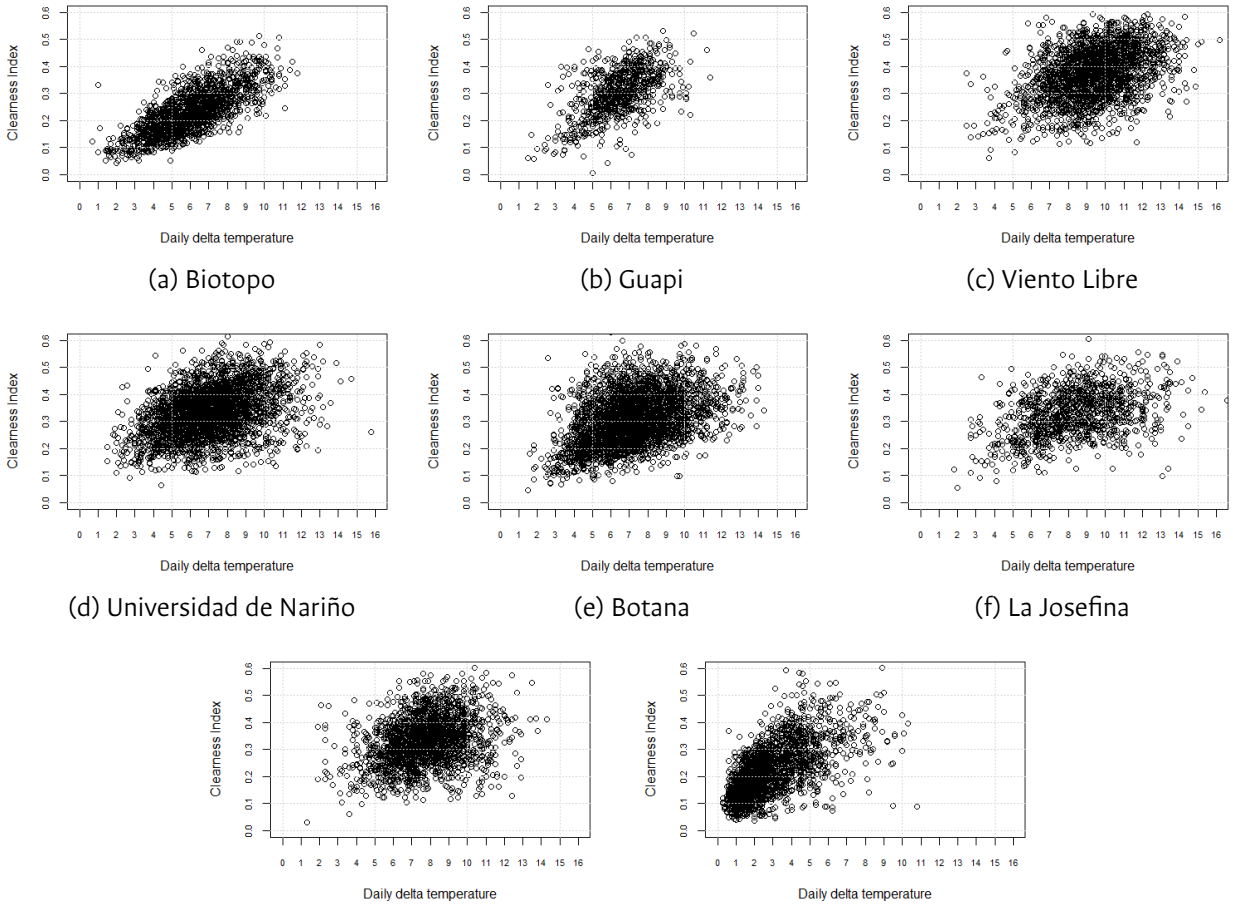
	Number of measures per day												Total of
AWS Name	1	2	3	4	5	6	7	8	9	10	11	12	days
Biotopo	33	9	17	13	26	46	56	85	130	257	605	846	2.123
Viento Libre	79	68	91	91	94	120	124	192	264	500	873	696	3.192
Universidad de Nariño	74	26	24	31	47	71	113	198	309	599	1.170	1.379	4.041
Cerro Páramo	101	49	58	47	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 Paraiso	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

3.3.3 Empirical models calibration and validation

Figure **3-3** shows the relationship between the daily clearness index and daily delta of temperature for all studied AWS. Figure **3-3** (a) and (b) show the relationship for the AWS located in the Pacific zone. In these cases, the daily delta of temperature is between 1 to 10 $C°$, and the clearness index is between 0,1 to 0,5, this means that from all total global solar irradiance in a clear sky condition between 10 % to 50 % reaches the ground level. Figure **3-3** (c) to (g) show the relationship for AWS located in the Andean zone. *Universidad de Nariño* and *Botana* have similar ranges of daily delta of temperature and clearness index. *Viento Libre* has more data in the upper level of the daily delta of temperature and clearness index. *Cerro Páramo* has more concentration of data in lower values of clearness index and daily delta of temperature, this indicates high cloudiness and short daily difference of temperature.

Table **3-10** shows the empirical coefficients of the HS, BC, ON and the proposed model. The BC model's a and b empirical constants showed a growing trend with the altitude, while the c empirical constant showed the opposite behavior. The HS model's empirical coefficients did not present a significant variation. The ON model's empirical coefficient for the Bitopo AWS was the unique negative value. It is necessary more studies and data to understand physically this result. In general terms, the ON model's c coefficients presented a standard deviation of 0,00030.

Figure 3-3. Delta of Temperature-Clearness Index



(a) Biotopo (b) Guapi (c) Viento Libre

(d) Universidad de Nariño (e) Botana (f) La Josefina

(g) Paraiso (h) Cerro Páramo

Table 3-10. Empirical coefficients

	BC			HS
AWS	a	b	c	a
Biotopo	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

Paraiso	0,5850	0,3183	0,4818	0,1199	
Guapi	0,4471	0,1156	1,2478	0,1208	
	ON			Proposed model	
AWS	a	b	c	a	b
Biotopo	-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

Table **3-11** presents the results of seven statistical validation measurements for each AWS. Considering the RMSE, SD, MAE, U_{95} and MAPE results, the proposed model has a better performance than the other ones. BC's model had better results for MAE and MPE. The proposed model showed better results in AWS located at altitudes above 2.500 MASL. In contrast, the HS's model showed better performance at altitudes below 2.500 MASL. However, if the objective is to use a unique model to estimate solar irradiance from temperature data in the State, the proposed model will be the best.

The RMSE results showed that the lowest value occurred in the Guapi AWS $878,75 Wh/(m^2 day)$, and the highest value in the *Cerro Páramo* AWS $1.209,76 Wh/(m^2 day)$; on average, the RMSE was $1.046,69 Wh/(m^2 day)$. ON's model presented the biggest RMSE with $1.058,77 Wh/(m^2 day)$, followed by the BC's model with $1.052,96 Wh/(m^2 day)$, HS's model with $1.046,66 Wh/(m^2 day)$, and the proposed model with $1.028,37 Wh/(m^2 day)$.

The Standard Deviation (SD) of the residual shows that *Cerro Páramo* presented more scattered values than the other AWS. comparing the four models and analyzing the average SD for, the best option is the proposed model. The MBE shows that BC's model made an underestimation of 77,79 $Wh/(m^2 day)$ in *Cerro Páramo*. All models overestimate the resource in *Viento Libre* by 167,77 $Wh/(m^2 day)$, 163,13 $Wh/(m^2 day)$, 162,63 $Wh/(m^2 day)$, and 160,23 $Wh/(m^2 day)$, with the ON, HS, BC and Proposed model, respectively.

The MAE results shows that the proposed model had the best performance with an average error of 821,41 $Wh/(m^2 day)$. U_{95} results presented that, on average, the following result 2.076,49 $Wh/(m^2 day)$, 2.065,11 $Wh/(m^2 day)$, 2.034,13 $Wh/(m^2 day)$, 2.016,86 $Wh/(m^2 day)$ for ON, BC, HS, and proposed model, respectively. Confirming that the proposed model is better than the other ones.

The MPE results presented the BC's model as the best one with 14,45 % on average, and the HS's model as the worst with 15,31 % on average. MAPE showed that proposed model as the best option, with 35,12 % on average, followed by HS and ON models. MAPE results were consistent with the other statistical

results.

Table 3-11. Summary of the empirical model's results

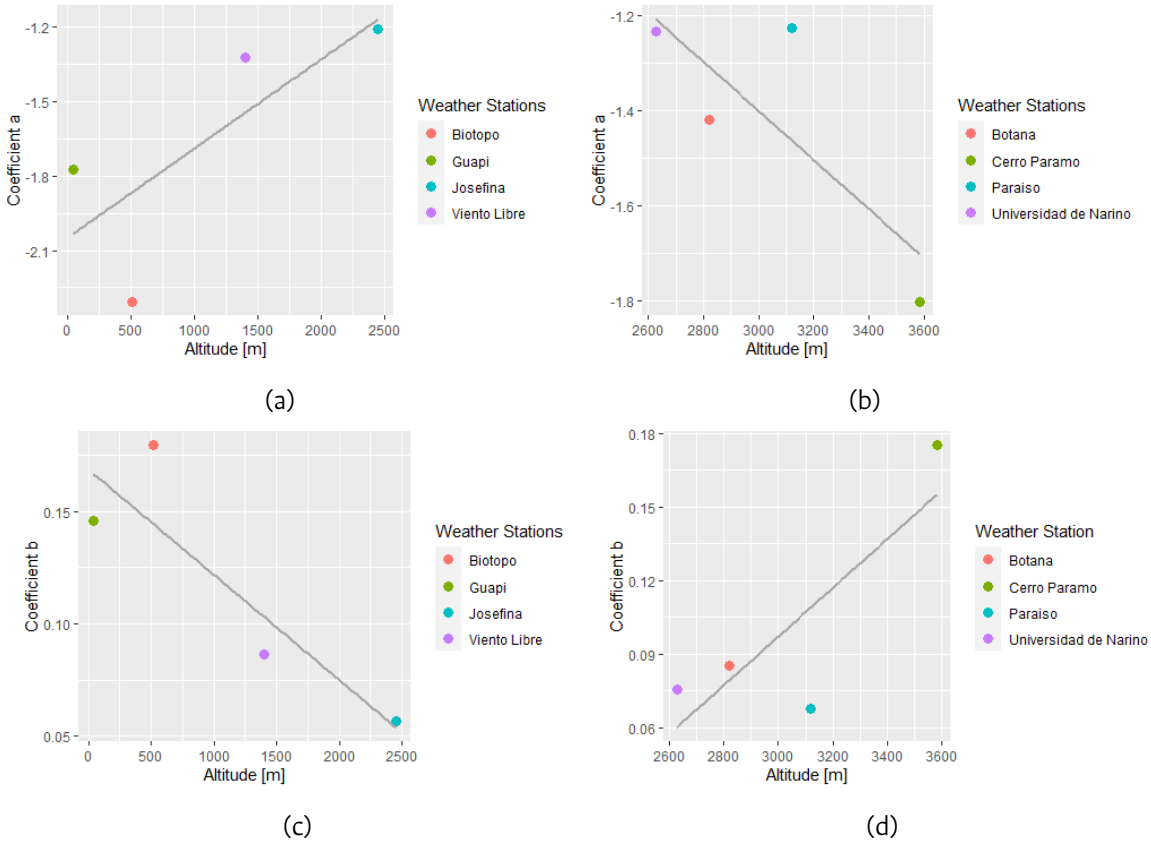
	RMSE $[Wh/m^2day]$				SD %			
AWS	BC	HS	ON	Proposed	BC	HS	ON	Proposed
Biotopo	1.152,62	**993,64**	1.155,07	1.113,48	49,90	**43,11**	50,08	48,29
Viento Libre	1.086,47	1.080,72	1.110,62	**1.077,35**	29,21	29,05	29,86	**28,97**
Cerro Páramo	1.194,57	1.209,76	1.196,56	**1.152,72**	56,04	56,77	56,15	**54,08**
Universidad de Nariño	1.032,45	1.083,73	1.032,24	**1.019,14**	31,89	33,55	31,88	**31,47**
Botana	1.052,42	1.070,23	1.077,17	**1.042,68**	34,46	35,05	35,23	**34,14**
Josefina	1.009,00	1.066,18	999,28	**984,75**	31,29	33,06	30,99	**30,53**
Paraiso	930,82	990,32	938,02	**921,32**	27,82	29,58	28,04	**27,54**
Guapi	965,40	**878,75**	961,21	915,53	31,75	**28,91**	31,58	30,12
	MBE $[Wh/m^2day]$				MAE $[Wh/m^2day]$			
AWS	BC	HS	ON	Proposed	BC	HS	ON	Proposed
Biotopo	-77,79	**-2,01**	-45,24	-37,29	916,08	**800,52**	917,96	885,10
Viento Libre	162,63	163,13	167,77	**160,23**	863,76	**861,64**	883,10	862,86
Cerro Páramo	37,90	**21,29**	27,22	33,37	940,46	946,29	929,90	**887,34**
Universidad de Nariño	90,20	**62,04**	93,68	93,72	845,12	884,48	841,37	**833,30**
Botana	48,24	**42,30**	71,63	47,13	866,59	881,19	888,56	**860,18**
Josefina	**5,28**	-20,58	16,92	22,18	769,70	830,06	767,58	**760,43**
Paraiso	-23,17	-42,52	-15,89	**-14,98**	757,02	806,64	760,64	**748,67**
Guapi	-36,98	**-16,38**	-53,45	-27,50	753,77	**696,35**	773,11	733,40
	U_{95} $[Wh/m^2day]$				MPE %			
AWS	BC	HS	ON	Proposed	BC	HS	ON	Proposed
Biotopo	2.261,26	**1.949,37**	2.266,06	2.184,48	**16,22%**	19,52%	17,93%	18,33%
Viento Libre	2.130,26	2.118,99	2.177,61	**2.112,38**	15,34%	15,32%	15,39%	**15,18%**
Cerro Páramo	2.343,94	2.373,74	2.347,83	**2.261,82**	28,77%	27,90%	**27,64%**	28,69%
Universidad de Nariño	2.024,57	2.125,13	2.024,17	**1.998,46**	13,73%	**12,78%**	13,82%	13,80%
Botana	2.063,85	2.098,79	2.112,40	**2.044,75**	14,22%	**11,32%**	15,08%	14,11%
Josefina	1.978,59	1.941,90	1.959,53	**1.931,04**	**12,24%**	20,51%	12,52%	12,70%
Paraiso	1.825,23	1.941,90	1.839,34	**1.806,59**	8,24%	**7,70%**	8,59%	8,51%
Guapi	1.893,21	**1.723,28**	1.885,00	1.795,41	6,82%	7,46%	**6,11%**	7,12%
	MAPE %							
AWS	BC	HS	ON	Proposed				
Biotopo	49,13%	**44,53%**	49,71%	48,13%				
Viento Libre	30,09%	29,99%	30,46%	**29,94%**				
Cerro Páramo	56,68%	56,62%	55,00%	**53,67%**				
Universidad de Nariño	31,44%	32,47%	31,26%	**30,97%**				

Botana	34,42%	**32,58%**	35,35%	34,07%
Josefina	30,83%	37,42%	30,74%	**30,56%**
Paraiso	26,75%	28,29%	26,91%	**26,47%**
Guapi	28,15%	**26,14%**	28,65%	27,14%

Lowest values are in bold

The proposed model's empirical coefficients showed a relationship with altitude. Figure **3-4** presents the empirical coefficients' lineal adjustments for two altitude ranges covering the AWS's locations. On the left side, figures represent the relationship between empirical coefficients and the altitudes above 2.500 MASL, and on the right side, figures show the remaining AWS for higher altitudes. Statistics analysis indicates that R^2 for the a coefficient is 0,5995 and 0,5262 in the first and second cases, respectively. The R^2 for the b coefficient is 0,8152 and 0,6069 in the first and second cases, respectively. This result is remarkable since the solar irradiance and temperature change with the altitude, the first due to the rise of beam irradiance caused by the reduction of scattered and absorbing molecules and aerosol (Blumthaler, 2012).

Figure 3-4. a and b empirical model-altitude relation in the proposed model



3.3.4 Imputation of daily solar insolation data

According to the AWS location, the imputation process to estimate the daily solar insolation data followed the best empirical model based on the research results. Table **3-12** shows the imputed data for each AWS. *La Josefina* was the AWS with the most imputed values; it was necessary to fill 2.870 missing data. *Botana* was the AWS with less imputed values, filling 572 missing data. An average among all AWS allows concluding that about four years had missing data; therefore, it was necessary to fill them employing temperature data. Time series before and after the daily solar insolation imputation process for each AWS allow us present Figure **3-5**.

Table 3-12. Number of imputed days by AWS

AWS Name	Days
Biotopo	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

Figure **3-6** shows the monthly daily average solar insolation for all analyzed AWS. The AWS located in the Pacific zone presented a similar behavior; namely, they registered a peak in August and September, and the lower level in November. *Viento Libre* is the only AWS that recorded values close to 4.000 $Wh/(m^2day)$ in August. The *Botana* and *Universidad de Nariño* on the Andean region showed a peak between October and November. *Cerro Páramo* and *Viento Libre* exhibited certain energy complementarity because the lower level in the first one is to compensate with the higher level of the second one.

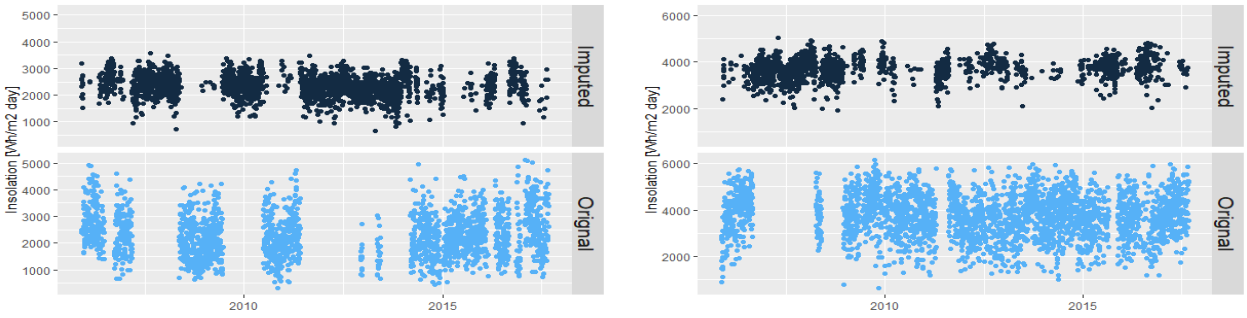
3.4 Conclusions

The global solar irradiance data's validation levels have a strong influence on the empirical variables results. It is evident when the results present that 60,89 % overcome the mandatory validation steps. From validated value, 95,81 %, which corresponds to the number of days with at least six recordings, constituted the empirical model's calibration information. However, this percentage represented, on average, 33,90 % of the total information recorded in the AWS. Besides, the days with complete information just reached 1,26 %. This result indicates that the time series' quality is no optimum; therefore, it is necessary to improve and increase the maintenance and calibration procedures.

In the State of Nariño, the AWS's performance is a determining factor, considering the predominance of partially high cloudiness that represents 64,7 % of the days. In other words, in Nariño, there is a high cloud interaction that difficult the solar insolation estimation; therefore, it is fundamental to increase the reliable measurement systems. Consequently, it is essential to regularly establish a plan to do these procedures and follow high-quality and widely accepted standards. It was also notorious, the need for installing more AWS to increase the sampling points.

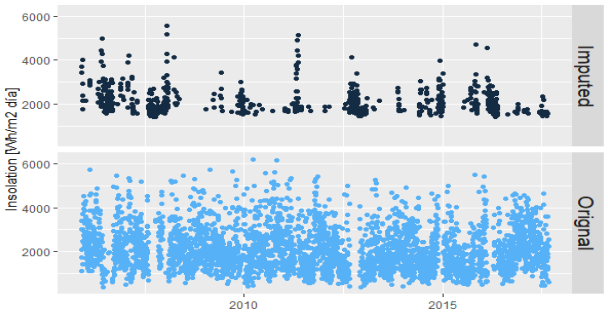
Regarding the temperature measures, only 92,78 % was useful for the empirical calibration and imputation from the total data that overcame the hourly validation steps. Besides, the number of temperature

Figure 3-5. Data imputation

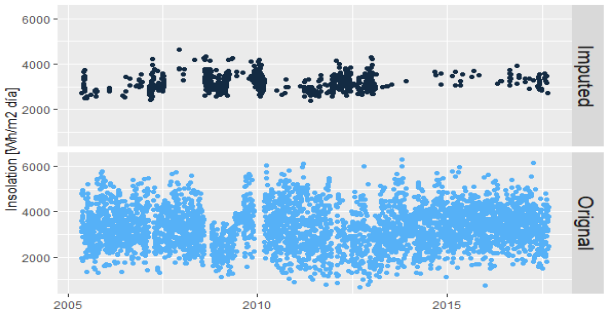


(a) Biotopo

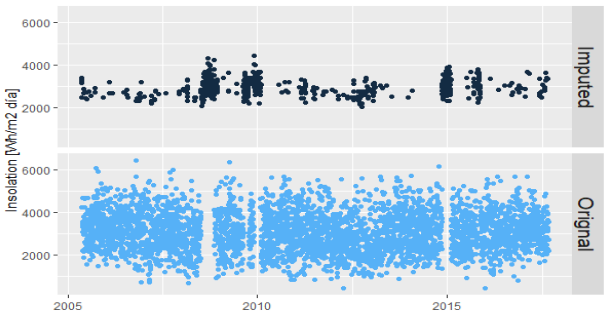
(b) Viento Libre



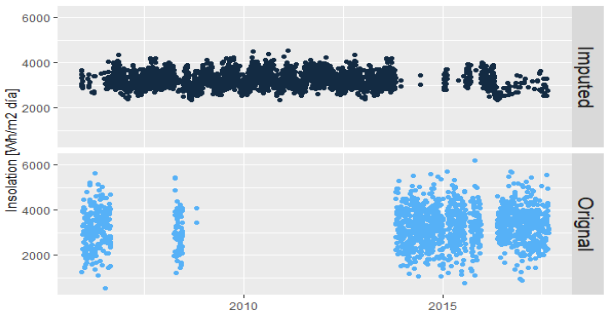
(c) Cerro Páramo



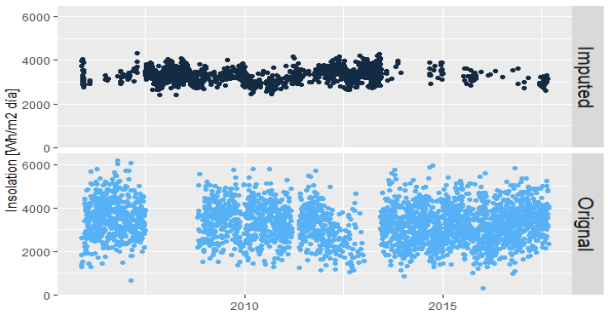
(d) Universidad de Nariño



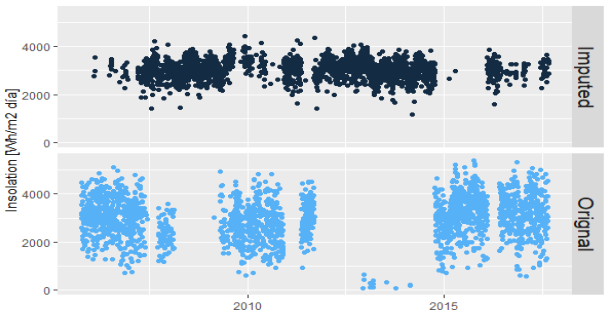
(e) Botana



(f) La Josefina

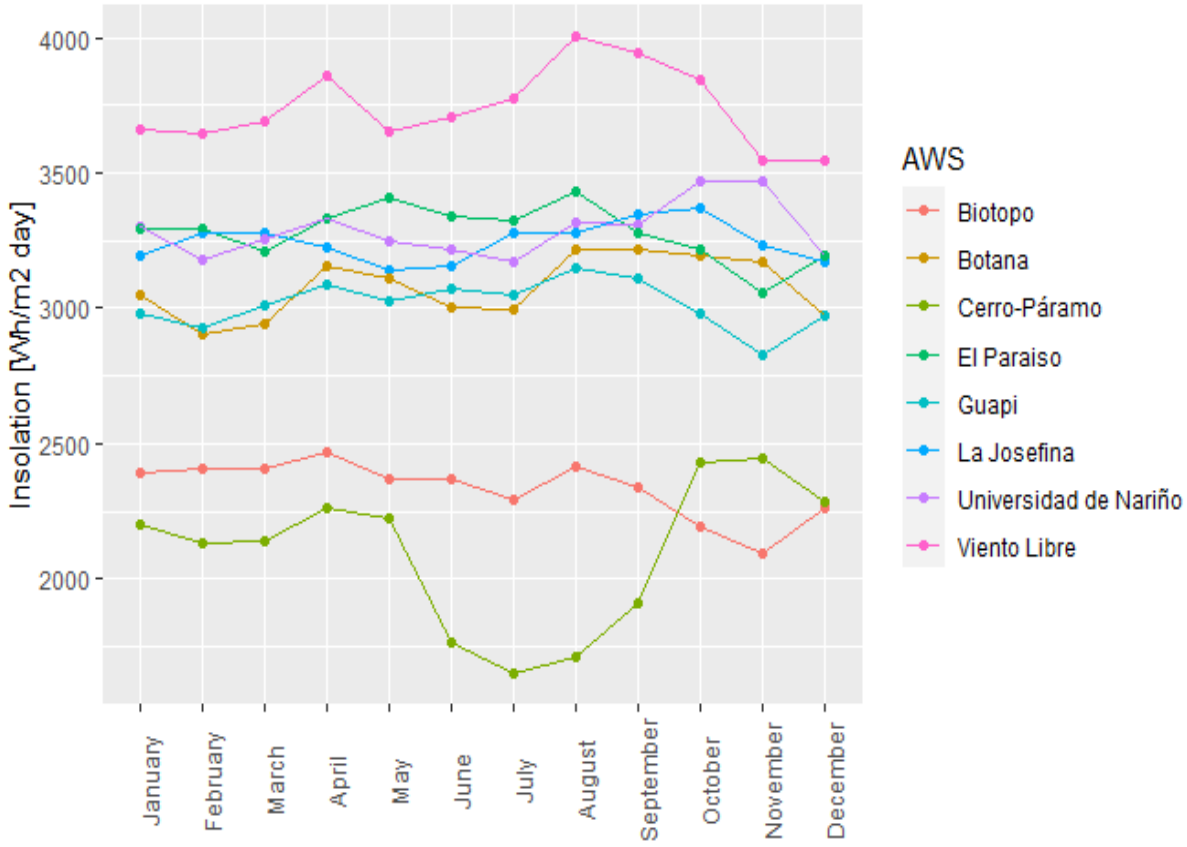


(g) Paraiso



(h) Guapi

Figure 3-6. Monthly daily insolation



daytime measurements presented peaks in eleven and twelve daily values; this means that most of the days used for modeling and filling the database by the imputation process had 88,46 % of the total information on average.

The proposed model showed a linear relationship between the empirical coefficients against the altitude. R^2 showed a better adjustment between the empirical constants and the altitude in sites above 2.500 MASL than sites below this altitude. This result is consistent with the temperature in tropical zones and the global solar irradiance in high altitudes.

Results from RMSE, SD, MAE, U_{95} and MAPE statistics tests pointed out that the proposed model had better performance in five of the eight evaluated cases. These cases are in the Andean and Amazon zones, with altitudes above 2.500 MASL. The proposed model was useful to input the information in the Andean and Amazon AWS. In the Pacific zones' AWS, the Hargreaves and Samani's model was the best option, followed by this research's proposed model. The proposed model had a stable performance in this tropical and mountainous environment. However, it is necessary to analyze more information coming from other places with the same characteristics. This prime requirement needs to increase AWS's number and the time series' quality in the tropical and mountainous environments.

4 Assessing empirical models for estimate global solar irradiance using air temperature in tropical and mountainous environment. Part II[1]

Abstract

Solar energy potential maps are enablers for making decisions about installing photovoltaic solar energy systems. However, the lack of in situ solar irradiance information recorded is a barrier to support decision-making in this field of study. Therefore, this research increases the solar irradiance sample points using temperature-based empirical methods such a Hargreaves and Samani and Logistic models, and use the leave-one-out cross-validation to assess the performance of four spatial interpolation techniques in tropical and mountainous environments that cover an area of 33.268 km^2. The analysis allowed concluding that the Hargreaves and Samani model is better for Pacific zone, while Logistic is better for Andean and Amazon zones. Also, ordinary Kriging method was the best interpolation technique because presented the lower bias.

Keywords

Temperature-based models, Hargreaves and Samani, spatial interpolation techniques, solar radiation mapping.

4.1 Introduction

Solar irradiance information is essential in fields of study as electricity generation, weather forecasting, agricultural production, and ecological behavior (Moreno et al., 2011). This research approaches the solar electricity generation subject that requires high-quality information to design power plants optimally. The lack of data or low-quality information is a barrier to know the solar potential and de-

[1] This article had the participation of Belizza Ruiz

sign power plants. This complicated situation is deeper in isolated areas with nonexistent or deficient road infrastructure like mountainous regions and tropical environments (Janjai et al., 2005; J. Li & Heap, 2014). Despite the above, governmental entities in charge of offering electricity service solutions in isolated areas promote electrification projects using solar irradiance. Stakeholders from public and private sectors face a lack of information affecting the successful development of projects; therefore, they use solar energy potential maps to make decisions and avoid inadequate planning.

Academic literature and governmental reports present solar irradiance atlas for most global territory, showing solar potential indicative values. A share of this information uses satellite data to draw maps while others use *in situ* data. Ground-level data offers full information about the environment where the weather variable records exist; however, the Automatic Weather Stations (AWS) amount is no suitable in all cases. Therefore, it is necessary to estimate the solar irradiance values for places without solar irradiance sensors to obtain information about the not sampled points, and obtain spatially continuous data over the study area (J. Li & Heap, 2011). TInterpolation techniques allow designing useful regional models to standardize recordings from isolated areas (Moreno et al., 2011). Results quality depends on spatial density and clustering of the sample points, surface type, among other aspects (J. Li & Heap, 2011). Overall, there are two categories for the interpolation techniques: non-geostatistical and geostatistical. However, all methods share the same general estimation formula (J. Li & Heap, 2008).The non-geostatistical techniques estimate no-sampled points values with adjusted mathematical functions according to sampled points information. The geostatistical techniques establish a spatial correlation between data (Martín & Dominguez, 2019). The sampled points number is decisive to define the information quality because the solar irradiance value is affected not only by cloudiness, aerosols, relative humidity, among other factors, but also by the physiography and soil type of the study site (Jeffrey et al., 2001; Şen, 2008).

This research assesses a combination of temperature-based empirical models and spatial interpolation techniques in a tropical and mountainous environment to elaborate monthly solar insolation maps. Accordingly, this research uses in situ measurements of global solar irradiance collected in eight AWS located in the State of Nariño in the Colombian southwest. The sampled points increased by implementing two temperature-based empirical models to estimate global solar irradiance in sixteen locations with Conventional Weather Stations (CWS). Likewise, this research used spatial interpolation techniques such as Inverse Distance Weight (IDW) and Simple, Ordinary, and Universal Kriging to obtain continuous solar insolation information in the studied region. Moreover, the leave-one-out cross-validation is the process implemented to select the best spatial interpolation technique. R-CRAN is the software, and gstat is the package used to realize all calculations.

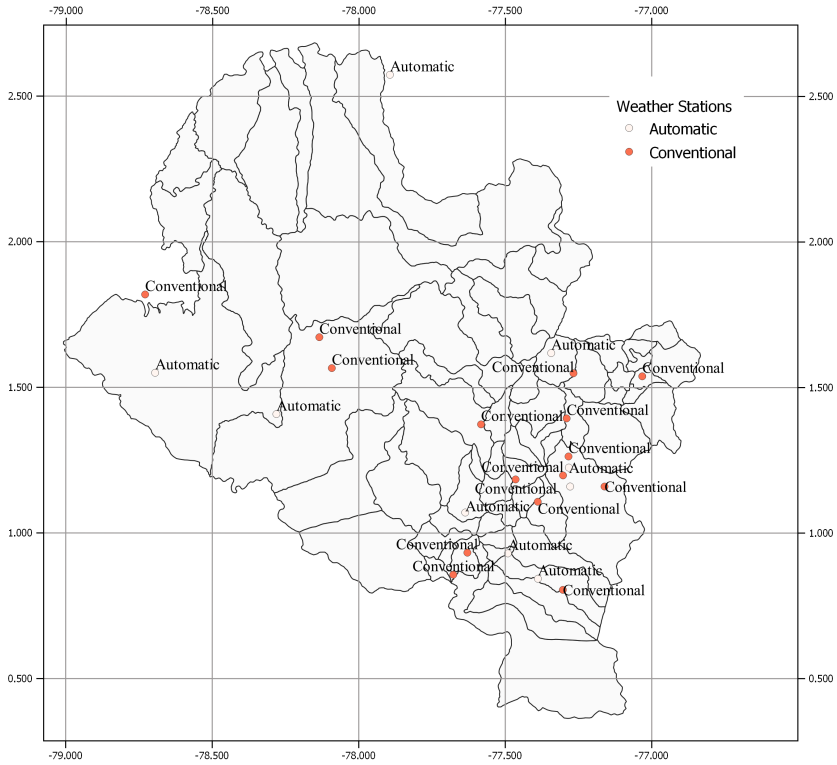
The results show that the Hargreaves and Samani empirical model performs better in the Pacific zone. In the Andean zone, the linear relationship between the empirical constants and the altitude, derived from the logistic model, reduces the overestimation. Although this method could have adjustments stemmed from future research works, it has presented pleasing results (Hoyos-Gómez, Ruiz, 2020). Therefore, it estimates information for the Andean region. The Ordinary Kriging interpolation technique was the best among the techniques analyzed because it presented less bias during all year because its robustness.

This article had the following sections: Section 4.2 shows the description of the data. Section 4.3 presents the theoretical framework explaining the spatial interpolation techniques used in this work. Section 4.4 contains the methodology followed in this research. Section 4.5 shows the results. Finally, Section 4.6 presents the conclusions.

4.2 Description of the data

Table **4-1** and **4-2** list the AWS and CWS classified by region, respectively, and Figure **4-1** shows their location in the State of Nariño. There are 5 AWS and 11 CWS in the Andean region, 3 AWS and 4 CWS in the Pacific region, and 1 AWS and a 1 CWS in the Amazon region. AWS altitude ranges between 16 and 3.577 MASL, and CWS altitude range between 1 to 3.141 MASL. The AWS measures global solar irradiance and temperature, while the CWS measures temperature uniquely. The AWS density in the State of Nariño is about $4.158km^2$ per station approximately; adding the CWS, the density improves, passing to $1.386km^2$ per station.

Figure 4-1. Location of weather stations



The global solar irradiance and temperature recorded in the AWS were submitted to a quality control procedure described in (Hoyos-Gómez and Ruiz, 2020). The temperature information of the CWS was obteined already validated. Table **4-3** and Table **4-4** present the number of days classified by the number of data items recorded. Regarding global solar irradiance the days mainly have 10 or 11 data items recorded, while in the temperature case mainly the day have 11 to 12 data items recorded. It is important remark that just 1,26 % of the total measured days record all the day time values of global solar irradiance.

Table 4-1. Automatic Weather Stations

Name	Latitude [°]	Longitude[°]	Altitude $MASL$	Period	Region
Biotopo	1,41	-78,28	512	2005-2017	Pacific
Altaquer	1,56	-79,09	1.1010	2013-2014	Pacific
Granja el Mira	1,55	-78,69	16	2016-2017	Pacific
Cerro-Páramo	0,84	-77,39	3.577	2005-2017	Amazon
La Josefina	0,93	-77,48	2.449	2005-2017	Andean
Viento Libre	1,62	-77,34	1.005	2005-2017	Andean
Universidad de Nariño	1,23	-77,28	2.626	2005-2017	Andean
Botana	1,16	-77,27	2.820	2005-2017	Andean
El Paraiso	1,07	-77,63	3.120	2005-2017	Andean

Table 4-2. Conventional Weather Stations

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

4.3 Theoretical framework

The growing solar technology use has increased the interest in information about solar potential point by point on territories. That knowledge entails obtaining continuous spatial data of solar irradiance that arise from the spatial interpolation processes. The spatial interpolation methods estimate solar irradiance values in non-sampled points from measured data using spatial variability information (J. Li & Heap, 2008). The regionalized variable theory, the semivariogram method, and the cumulative semivariogram are the approaches used to model the spatial variability using solar irradiance and distance (Şen, 2008).

Table 4-3. Amount of days classified by daily irradiance measures

	Number of measures per day													Total of
AWS Name	1	2	3	4	5	6	7	8	9	10	11	12	13	days
Biotopo	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 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

Table 4-4. Amount of days by daily temperature measures

	Number of days												Total of
AWS Name	1	2	3	4	5	6	7	8	9	10	11	12	days
Biotopo	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 Paraiso	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

The sampling density and spatial distribution of the points with recorded data, cluster sampling, surface type, data variance, among other things, influence the spatial interpolation methods performance (J. Li & Heap, 2011). There are mainly two interpolation categories: non-geostatistical and geostatistical. In the first category, the methods partially model the spatial autocorrelation through mathematical functions; some of these methods are Natural Neighbor (NaN), Inverse Distance Weight (IDW), Triangular Irregular Network (TIN), Regression models, among others. In the second category, the methods simulate the spatial data autocorrelation and evaluate the uncertainty of the results to carry out the interpolation processes such as Kriging that which is the most commonly used method (Bhattacharjee et al., 2019; Martín & Dominguez, 2019; Sankar et al., 2018).

All methods share the same general estimation represented by Equation 4-1:

$$\widehat{z}\left(x_0\right) = \sum_{i=1}^{n} \lambda_i z\left(x_i\right) \tag{4-1}$$

where $\widehat{z}$ is the estimated value in the x_0 point, z is the measured value in x_i point, λ_i is the weight of the sampled point, and n is the number of sampled points used for the estimation (J. Li & Heap, 2008). This

expression arose from considering that samples from close weather stations have a higher probability of presenting similar values and behaviors than samples taken in distant weather stations with each other (J. Li & Heap, 2008). The main objective of any method is to determine these weights as better as possible (Şen, 2008). The following subsection presents an explanation of IDW and Kriging techniques.

4.3.1 Inverse Distance Weighted (IDW)

IDW estimates the value of non-sampled points through the linear combination of the nearby sampled points. The inverse of the Euclidean distance allows computing the weight of a sampled point over a non-sampled point, as Equation 4-2 indicated,

$$\lambda_i = \frac{1/d_i^p}{\sum_{i=1}^{n} 1/d_i^p} \tag{4-2}$$

where d_i is the distance between x_i and x_0, p is the exponent parameter, and n is the number of sampled points used for the estimation (J. Li & Heap, 2008). This expression shows that the farthest sampled point has the lowest contribution to the calculation (Martín & Dominguez, 2019).The main factor influencing the method accuracy is the exponent selected arbitrarily as the neighborhood size. When the p parameter value increases, the weight value diminishes. Moreover, if the p parameter value changes, the type of interpolation also changes. The IDW behaves as moving average interpolation when $p = 0$, linear interpolation with $p = 1$, and weighted moving average when p is different to 1. The most common exponent is $p = 2$ called as Inverse Distance Squared (IDS) (J. Li & Heap, 2011; Introduction to Spatial Analysis, 2009). For IDS λ_i is calculated with the following expression $\lambda_i = \left(\frac{r-d_i}{d_i}\right)^2$, where r is the radius of the search window that can be fixed or variable (Bhattacharjee et al., 2019).

4.3.2 Kriging

Kriging is a generalized least-square regression algorithm with an exceptional unbiased prediction (Jamaly & Kleissl, 2017). The accuracy of the Kriging depends on the variogram to obtain a minimum-variance predictor. The variogram is estimated using the semivariance, which considers both the distance and the variation between the estimated and measured points (Introduction to Spatial Analysis, 2009). Spatial correlation analyzes the data change according to the semivariance, which is the difference between the measured and estimated point variances as Equation 4-3 presents (M. A. Oliver & Webster, 1990),

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

where $\widehat{\gamma}\left(h\right)$ is the semivariance, which is the half of the variance between the estimation point and the nearby measured points, n is the number of pairs of sample points separated by distance h, $z\left(x_i\right)$ is the

sample value in the location x_i, and $z\left(x_i+h\right)$ is the value at a h distance from x_i. At short distances, low h values, small semivariance values, when h increase the semivariance also increase.

Equation 4-3 allows calculate the experimental variogram that consist of semivariances at a finite set of lags. The reliability of the experimental variogram is related to the quality of the data and their density (A. M. Oliver & Webster, 2015). The next step is to fit a function to the experimental variogram known as the theoretical variogram that describe the main characteristics of the sampled points. The slope of the theoretical variogram quantifies the spatial autocorrelation between the sampling points at lag intervals (Martín & Dominguez, 2019). There are two ways to model the theoretical variogram unbounded and bounded. In the first case, the variogram increase indefinitely at lag increase. In the second case, the variogram follows a second-order stationary process, it reaches an upper limit that is the sill variance (Webster & Oliver, 2007). Additionally, the variogram has a range that determine the limit of the spatial correlation where the autocorrelation becomes zero. Additionally, when there is a zero distance between the estimation and measured point, the semivariogram is zero. However, if there are measurement or scale errors in the variable or both, the semivariogram presents a nugget effect. In other words, the semivariogram has a high value in short distances (Viera Díaz, 2002).

The variogram in the base of the Kriging methods implementation. Equation 4-4 shows the Kriging general estimator, where μ is a stationary mean, λ_i is the Kriging weight,$\mu\left(x_0\right)$ is the mean of samples considered in the influence area, n is the number of sampled points (J. Li & Heap, 2008). It follows the structure of the general estimation showed in Equation 4-1,

$$\widehat{z}\left(x_0\right)-\mu=\sum_{i=1}^{n}\lambda_i\left[z\left(x_i\right)-\mu\left(x_0\right)\right] \tag{4-4}$$

The weights arise from a system of linear equations that accomplish with two conditions. The first one is the following restriction $E\left\{\widehat{z}\left(x_0\right)-z\left(x_i\right)\right\}=0$ to obtain unbiased weights. The second one is that the weights must minimize the variance $\sigma\left(x_i\right)=Var\left\{\widehat{z}\left(x_0\right)-z\left(x_i\right)\right\}=0$ (Moreno et al., 2011). Although there are several Kriging methods, there is only one objective: to estimate continuous values for the no-sampled points (Dai et al., 2003).

Simple Kriging (SK)

SK has the most straightforward mathematical formulation among the Kriging methods. This model assumes that the mean and covariance are previously known and constant in all locations (Martín & Dominguez, 2019). Consequently, when the mean and covariance are different from the assumed starting values, the method does not show optimal results, limiting its applicability (Olea, 1999). Equation 4-5 shows the SK estimator model,

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

where μ is a known stationary mean for the whole interpolated surface. In SK $1 - \sum_{i=1}^{n} \lambda_i$ is not necessarily equal to 0; however, the second term of 4-5 ensure that the prediction are unbiased (J. Li & Heap, 2008; A. M. Oliver & Webster, 2015). SK works with covariances, C, more than semivariances; therefore, the weights are calculated as follows:

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

Ordinary Kriging (OK)

OK is the most general and widely used Kriging method that assumes that the variation is random and spatially depend, and a constant unknown mean and a variance that depends only on lag distance, Equation 4-7 shows the estimation model of the OK

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

This method follows the next restriction $[1 - \sum_{i=1}^{n} \lambda_i] = 0$ to accomplish with the non-bias condition (J. Li & Heap, 2008). OK aims to minimize the errors' variance and uses lineal combinations to estimate the data set weights reducing the bias (Kiš, 2016), as Equation 4-8 and 4-9 show

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

$\widehat{\gamma}\,(x_i - x_0)$ is the semivariance between the ith and jth point. The theoretical variogram must guarantee that the variance is not negative (A. M. Oliver & Webster, 2015; Webster & Oliver, 2007)

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

Following that $\sum_{i=1}^{n} \lambda_i = 1, \psi(x_0)$ is the Lagrange multiplier used to minimize the variance. In a matrix form the weights are estimated as follow:

$$\lambda = A^{-1} b. \tag{4-10}$$

Universal Kriging (UK)

UK is a method used for spatial process with a trend or drift and a not stationary mean; therefore, this method incorporates the mean as a local trend in function of the coordinated (X, Y). The variation in $z(x)$ have a systematic component, which could be decomposed into a linear combination of deterministic functions, additionally to random component expressed by $z(x) = u(x) + \varepsilon(x)$. $u(x)$ replace

the mean μ, is a deterministic component that varies smoothly, $\varepsilon(x)$ is a spatially correlated random residual function (Bhattacharjee et al., 2019; J. Li & Heap, 2008; Olea, 1999; Kiš, 2016; A. M. Oliver & Webster, 2015; Webster & Oliver, 2007).

As in this case there is a not stationary process, the experimental variogram does not estimated the variogram of the random residual, $\varepsilon(x)$, as in the two above presented methods. Consequently, the variogram is estimated from $\varepsilon(x) = z(x) - u(x)$. To estimate the variogram without bias, it is need separate $u(x)$ from $\varepsilon(x)$; then $u(x)$ can be expressed as follows:

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

Where β_k, with $k = 0, 1, \ldots, K$ are unknown coefficients estimated from the data, and $f_k\left(x\right)$ are known deterministic functions of the spatial coordinates that describes the drift. For a linear trend there are three functions $f_0 = 1$, $f_1 = x_1$, and $f_2 = x_2$. For a quadratic trend it is need add three functions $f_3 = x_1^2$, $f_4 = x_1 x_2$ and $f_5 = x_2^2$. For example, with a linear trend the equation is expressed as follow $z(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon(x)$. Finally, with the functions to estimate $u(x)$, and $\varepsilon(x)$, it is possible predict the $\widehat{z}$ value at any x_0 point

$$\widehat{z}\left(x_0\right) = \sum_{i=1}^{n} \lambda_i f_k\left(x_i\right) \tag{4-12}$$

the expectation is

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

the estimator is unbiased if satisfied $\sum_{i=1}^{n} \lambda_i f_k(x_i) = f_x(x_0)$For all $\quad k = 0, 1, \ldots, K$

4.3.3 Cross-Validation

The cross-validation assesses and compares the regional interpolation techniques. The first step consists of defining a base case in which all sampled points allow estimating the no-sampled points. Afterward, calculations repeat to find the value of a removed sampled point and so forth until calculating such values as sampled points (Sankar et al., 2018). The best interpolation technique stemmed from error analysis that is the average of the comparisons between results from each calculation with removed sampled points and the base case sampled points (Berrar, 2018; Şen, 2008).

4.4 Methodology

Research development covers three stages. The first stage consists of estimating solar insolation using empirical models. The second part arises from implementing the spatial interpolation techniques using

the R-CRAN and the gstat package. The third part comprises the cross-validation to select the best option that allows mapping the global solar irradiance.

The Hargreaves and Samani model is the empirical method for estimating solar insolation in the Pacific zone. In the Andean and Amazon zones, the linear relationship between the logistic model empirical coefficients and the AWS's altitude is the technique proposed to estimate the empirical coefficients associated with the CWS. Coefficients are valuable because they belong to the temperature-based empirical model that estimates the global solar insolation for each CWS.

The second stage consists of mapping the global solar insolation using the spatial interpolation techniques explained in 4.3.1 and 4.3.2.At the beginning of the mapping process, the authors of this research constructed a regularly spaced grid with a resolution of 100 meters. Then, the Kriging techniques adjust a function to the experimental variogram represented by a scatter plot between the semivariance and the distance. The experimental variogram values allow determining the starting values with which compute the theoretical variogram. In this research, the starting values fulfilled the next recommendations: the third part of the experimental variogram maximum distances constitute the range, the mean of the first three values of the experimental variogram forms the nugget, and the mean of the last five values of the experimental variogram establishes the partial sill (Pebesma, 2016). The gstat package has twenty options to determine the theoretical variogram function (Pebesma & Graeler, 2020). The authors fitted all the available functions and chose the best option for each case.

The last stage conducted the cross-validation to determine the better interpolation technique and obtain the monthly global solar insolation maps. Every interpolation technique had an associated error analysis that consisted of calculating Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Standard Deviation (SD), and Mean Percentage Error (MPE). The techniques with fewer errors indicated the best option for each month. It is convenient to highlight that the statistical errors arose from the average of all results obtained by each iteration during the cross-validation process.

4.5 Results and discussion

The Hargreaves and Samani and logistic models allowed estimating the daily solar insolation using air temperature data to increase the number of sampled points. The Hargreaves and Samani model allowed calculating the empirical constant with AWS data located in the Pacific region; consequently, such an empirical constant served to estimate solar insolation using air temperature data of the CWS locations.

It is convenient to remind that there are three AWS in the Pacific zone; however, only one has a broad time framework and coinciding with the CWS time frameworks. Consequently, weather information from the *Biotopo* AWS allowed estimating the CWS solar irradiance behavior. Although Quansah et al. suggest a maximum distance of 20 km to transpose weather information from a point to another, the

authors of this research could not use that criterion because the time frameworks of the nearest weather stations did not coincide. Nevertheless, it is convenient to show the distances among the weather stations. There is a distance of 30,18 km between *Granja el Mira* AWS and *CCCP del Pacífico CWS*, 67,54 km between *Biotopo* AWS and *CCCP del Pacífico* CWS, 63,71 km between *Granja el Mira* AWS and *Barbacoas* CWS, and 68,79 km between *Biotopo* AWS and *Barbacoas* CWS.

The linear relationship between the AWS altitude and the empirical constants of the logistic model allowed computing the empirical constants of the CWS located in the Andean and Amazon zones. It is a relevant find because the use of the linear relationship make ease the estimation of global solar insolation using air temperature in the Andean and Amazon zones, that in this case does not have a considerable amount of weather stations

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

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

where h is the altitude of the weather station. The values of the empirical coefficients are replaced in Equation 4-16.

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

where H is the daily average global solar insolation, H_0 is the average daily extraterrestrial solar insolation, a and b are the empirical coefficients, ΔT is the difference between the maximum and minimum daily temperature. Table **4-5** presents the results of the empirical coefficients obtained for the CWS using Equation 4-14 and Equation 4-15. All values of the a empirical coefficient are negative; independently, if the CWS is located both above or below 2.500 MASL, in this case the maximum value is in *Obonuco*, the minimum is in *Bombona*, with a standard deviation of 0,1009. In the b empirical coefficient case, all values are positive, the maximum value is in *Común el automática*, and the minimum value is in *Monopamba* with a standard deviation of 0,0133.

The performance calculations used the four AWS data, two of the Pacific zone and two of the Andean zone, that had recordings of one year on average. Recorded data passed a quality control process; subsequently, a comparison between results from empirical models and data in situ arose to compute the statistical errors presented in Table **4-6**. The bolded results are the best ones.

For *Altaquer* and *Granja el Mira* AWS located in the Pacific zone, the logistic model had a better performance in RMSE, SD, MAE,MBE and U_{95} than the Hargreaves and Samani model that presented better results in MPE, and MAPE uniquely. In the MBE, the Hargreaves and Samani model underestimated the resource in 442,78 $[Wh/m^2day]$ for *Granja el Mira* and 333,11 $[Wh/m^2day]$ for Altaquer, while the Logistic model overestimate it in 145,14 $[Wh/m^2day]$ and 76,19 $[Wh/m^2day]$ in *Altaquer* and *Granja el Mira*, respectively. Even though the logistical model did not have the best behavior in all statistical error

Table 4-5. Estimated empirical coefficients for Andean and Amazon zone.

CWS	Altitude $[MASL]$	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

measurements, it is suitable to estimate the solar resource with air temperature data in places below 2.500 MASL.

For *Ospina Perez* and *Sandona* AWS located in the Andean zone, the logistic model showed lower errors in all statistical measurements. Regarding the MBE results, the two empirical models overestimated the solar resource. However, the logistical model reduced the overestimation by 181,09 $[Wh/m^2day]$ (51,12 %) and 225,15 $[Wh/m^2day]$ (18,55 %) for *Ospina Perez* and *Sandona*, respectively. Finally, the authors concluded that the logistical model presented better results than the Hargreaves and Samani model in altitudes above 2.500 MASL.

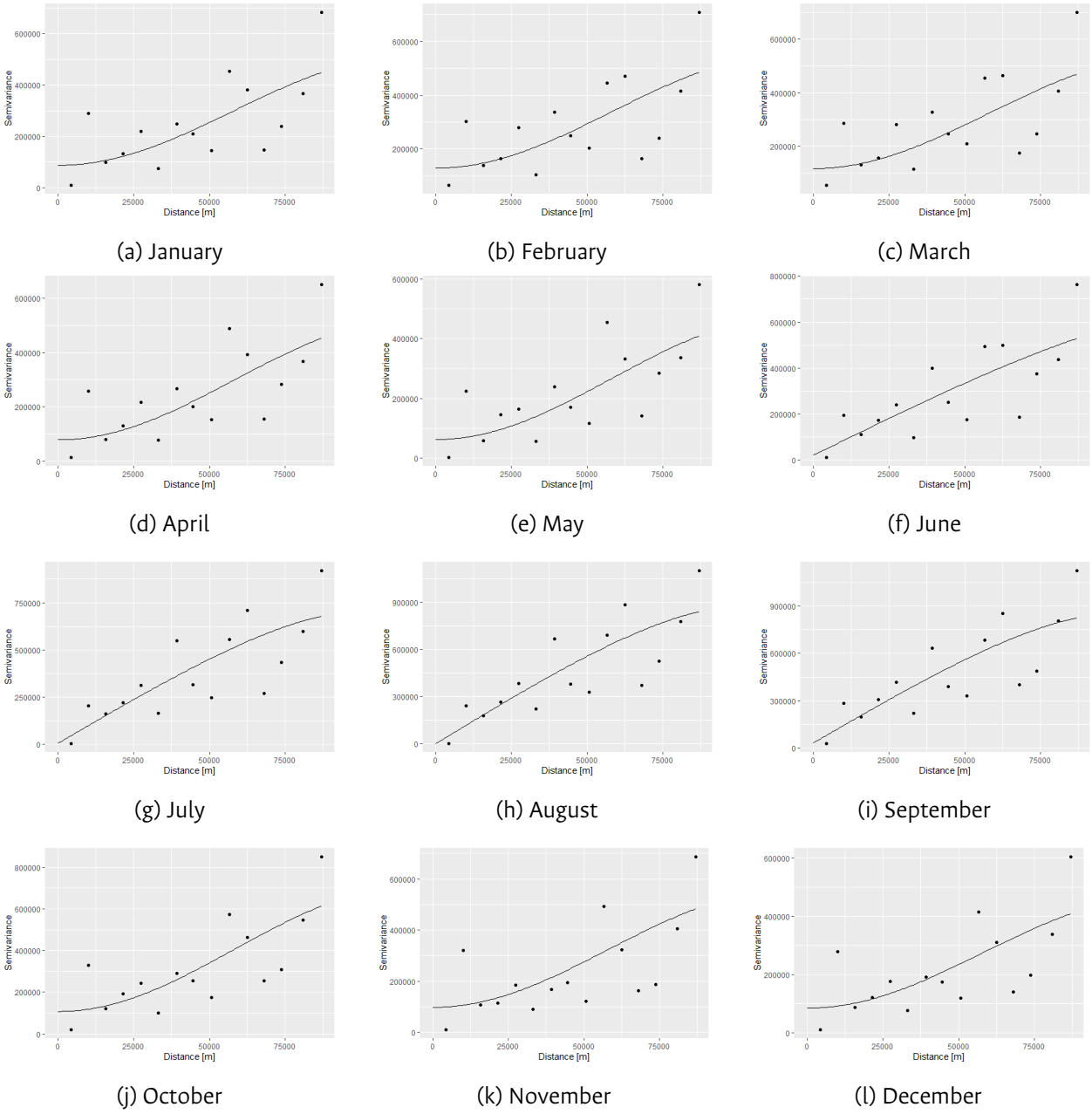
Table 4-6. Comparison of the empirical models.

AWS Name	N	RMSE $[Wh/m^2day]$	SD %	MAE $[Wh/m^2day]$	MBE $[Wh/m^2day]$	MPE %	MAPE %	U_{95} $[Wh/m^2day]$
Altaquer - HS	390	761,09	24,49	596,88	-333,11	**-4,23 %**	**22,30 %**	1.492,51
Altaquer – Logistic	390	**670,59**	**23,43**	**529,73**	**145,14**	14,14 %	24,24 %	**1.315,17**
Granja el Mira - HS	305	854,87	**24,81**	684,98	-442,78	**3,47 %**	**33,75 %**	1.676,26
Granja el Mira - Logistic	305	**781,91**	25,99	**596,66**	**76,19**	20,57 %	34,81 %	**1.533,40**
Ospina Perez - HS	231	987,70	27,44	797,83	355,08	16,80 %	27,29 %	1.936,64
Ospina Perez - Logistic	231	**934,93**	**27,35**	**736,86**	**173,99**	**11,09 %**	**24,43 %**	**1.833,25**
Sandona - HS	221	1.375,84	22,83	1.249,26	1.213,27	49,93 %	50,70 %	2.697,03
Sandona - Logistic	221	**1.179,92**	**22,69**	**1.034,94**	**988,12**	**41,30 %**	**42,32 %**	**2.313,07**

The base case used the IDW with $p = 2$ as exponent because IDS is the most used IDW methods. Implementing the Kriging methods entails determining the theoretical variograms represented by Figure **4-2** for each month. We implemented the best variogram model available in the gstat package for each month. The wave was the option selected from gstat package to model January to June and October to December, and Circular was the option for modelling May to September

The experimental variogram of May and December shows the lowest semivariance; their values were

Figure 4-2. Experimental and Theoretical variogram.



(a) January (b) February (c) March

(d) April (e) May (f) June

(g) July (h) August (i) September

(j) October (k) November (l) December

352.019 and 322.911, respectively, see Figure **4-2**(e) and Figure **4-2**(l). Therefore, it expects the lowest statistical errors for those months. In August and September, the semivariance difference had the highest values; their values were 847.928 and 801.113, respectively, see Figure **4-2**(h) and Figure **4-2**(i).Consequently, these months could show the highest statistical errors.

Table **4-7** shows the statistical errors stemmed from using the IDW and Kriging methods. The cross-validation leave-one-out was the technique implemented to compute the errors. Concerning the RMSE, there was an improvement of 11,71 %, 11,94 %, and 2,07 % for OK, SK and UK, respectively, regarding

to IDW results. Figure **4-3** confirms the RMSE results for every interpolation technique. Analyzing the average results, the SK and OK preset the more promising result. Also, the RMSE results show that all interpolation techniques had a similar performance throughout the year, see Figure **4-3**. During September, the RMSE reached the highest value 636,785 $Wh/(m^2day)$ with IDW, because the outlier values influence the results, as the semivariance value showed. May presents the lowest RMSE value, 410,781 $Wh/(m^2day)$ using SK. There is an reduction of the error of 14,63 % and 14,62 % with OK and SK against the IDW results in August.

Regarding the MAE results, the OK and SK techniques improved by 3,98 % and 5,10 %, respectively, and the UK technique reduced by 6,78 %, as Figure **4-4** presents. September shows the highest value obtained with the UK technique 455,055 $Wh/(m^2day)$, and May had the lowest value 296,33 $Wh/(m^2day)$. SK and OK show similar results during all year.

Concerning MBE, all Kriging methods had better results than IDW by 1,01, 0,063, and 0,022 times with OK, SK, and UK, respectively, see Figure **4-5**. The MBE results show that OK is the method with less bias; also, this is aslo the only technique with underestimation. It is crucial in the power plant design because the resource determines the power plant's size to generate the needed electricity.

The SD results shows that the Kriging methods reduce the dispersion by 11,40 %, 11,97 %, and 2,10 % with OK, SK and UK, respectively, see Figure **4-6**. The SD had similar results to RMSE because high dispersion could lead to outlier values. From July to September, there is more dispersion. Therefore, during these months, the weather in the territory is more heterogeneous than in the other months. Finally, the MPE results showed an improvement of OK, SK and UK by 46,58 %, 16,81 % and 20,98 %, respectively, see Figure **4-7**. The MPE results show that OK had better results during all year.

Table 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

Figure 4-3. RMSE results

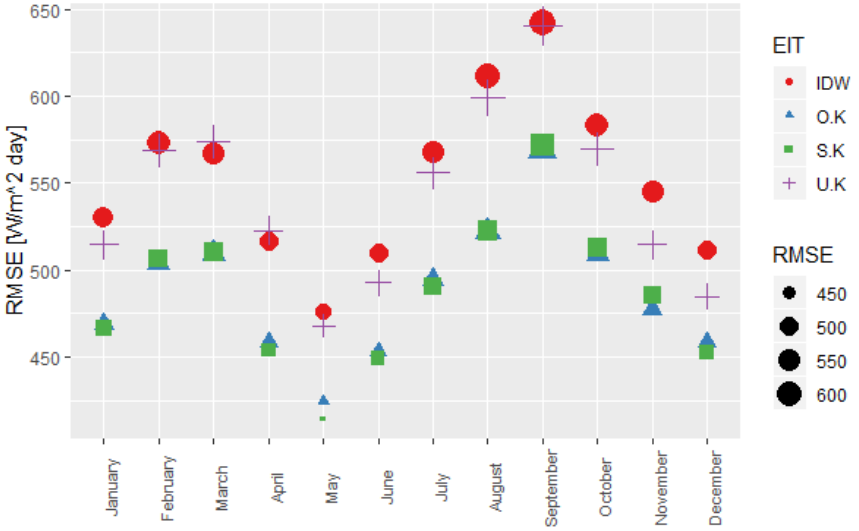


Figure 4-4. MAE results

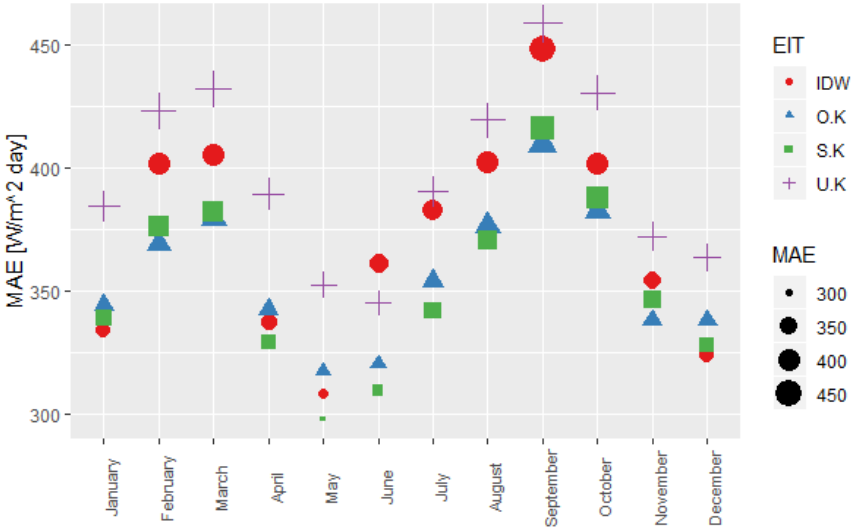


Figure 4-5. MBE results

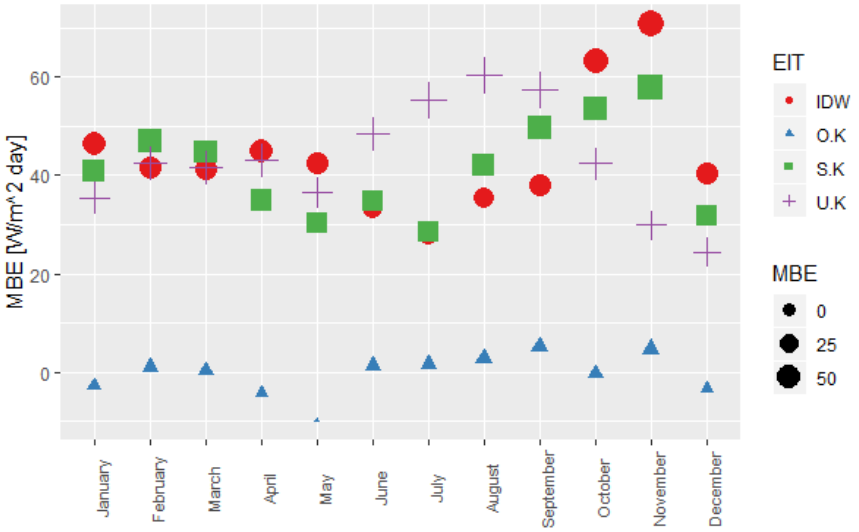
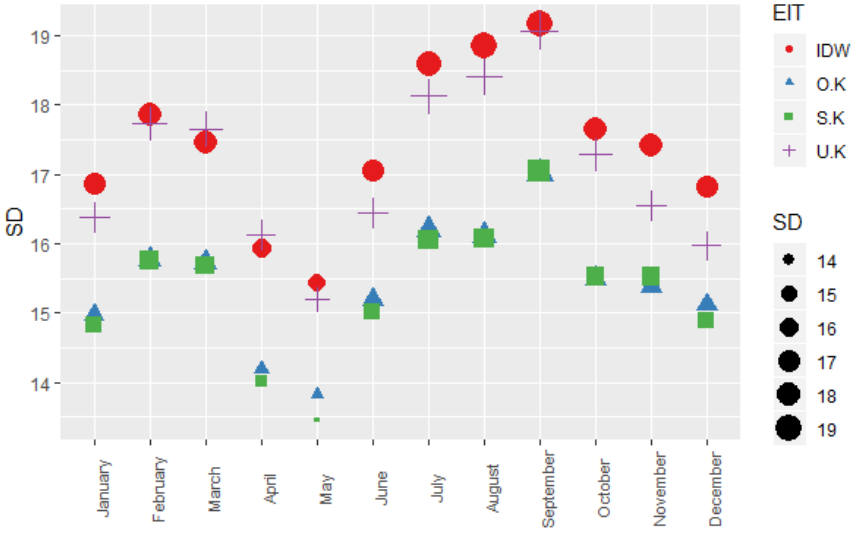
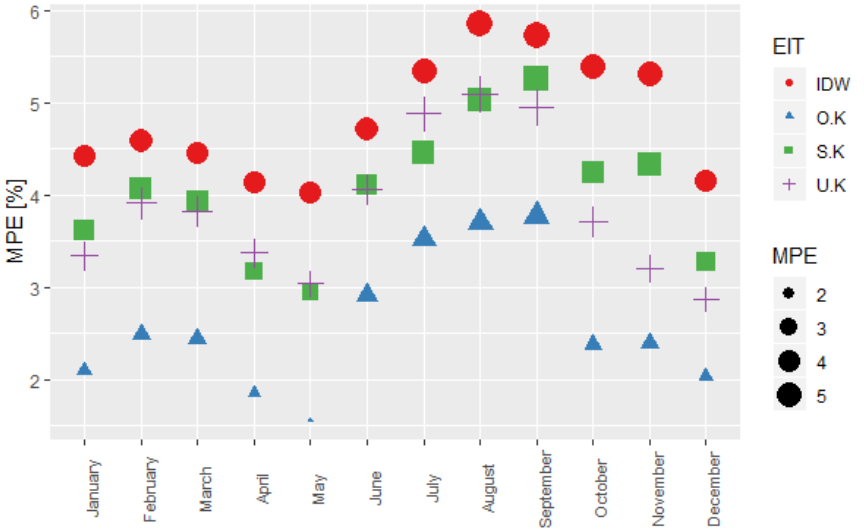


Figure 4-6. SD results



In general, Kriging methods showed better results than IDW. Withing the Kriging techniques SK had better results in RMSE (11,94 %), MAE (5,10 %), and SD (11,97 %), and OK had better results in MBE (1,01 times), and MPE (46,58 %). Based on results, OK was the method selected to map global solar insolation in the analyzed tropical and mountainous areas due to that it is the technique with less bias and less percentage error. Figure **4-8** to Figure **4-9** shows the monthly daily average global solar insolation from January until December. Comparing the obtained results in this research against the Solar Atlas from the IDEAM, the researchers agree that the Andean zone has significant potential during all year. However, there is a difference between the estimated potential in these two works. Overall, the IDEAM results present a higher potential than the estimated potential in this research. For example, in the

Figure 4-7. MPE results



Obonuco CWS, the IDEAM present potential of 3.677,5 $[Wh/m^2day]$, and this research estimate the potential of 3.298,6 $[Wh/m^2day]$ in January. In *Apto. San Luis*, IDEAM estimate a potential of 3.902,6 $[Wh/m^2day]$, and this research estimate the potential of 2.194,1 $[Wh/m^2day]$ in March. Also, IDEAM used two CWS (*Obonuco* and *Apto. San Luis*) and five AWS (*La Josefina, Botana, Cerro Páramo, Viento Libre*, and *Paraiso*) to estimate the potential in *Nariño*, this research increase the number of AWS and CWS used. Therefore, the density, which is an important factor in the interpolation techniques, increase with this research.

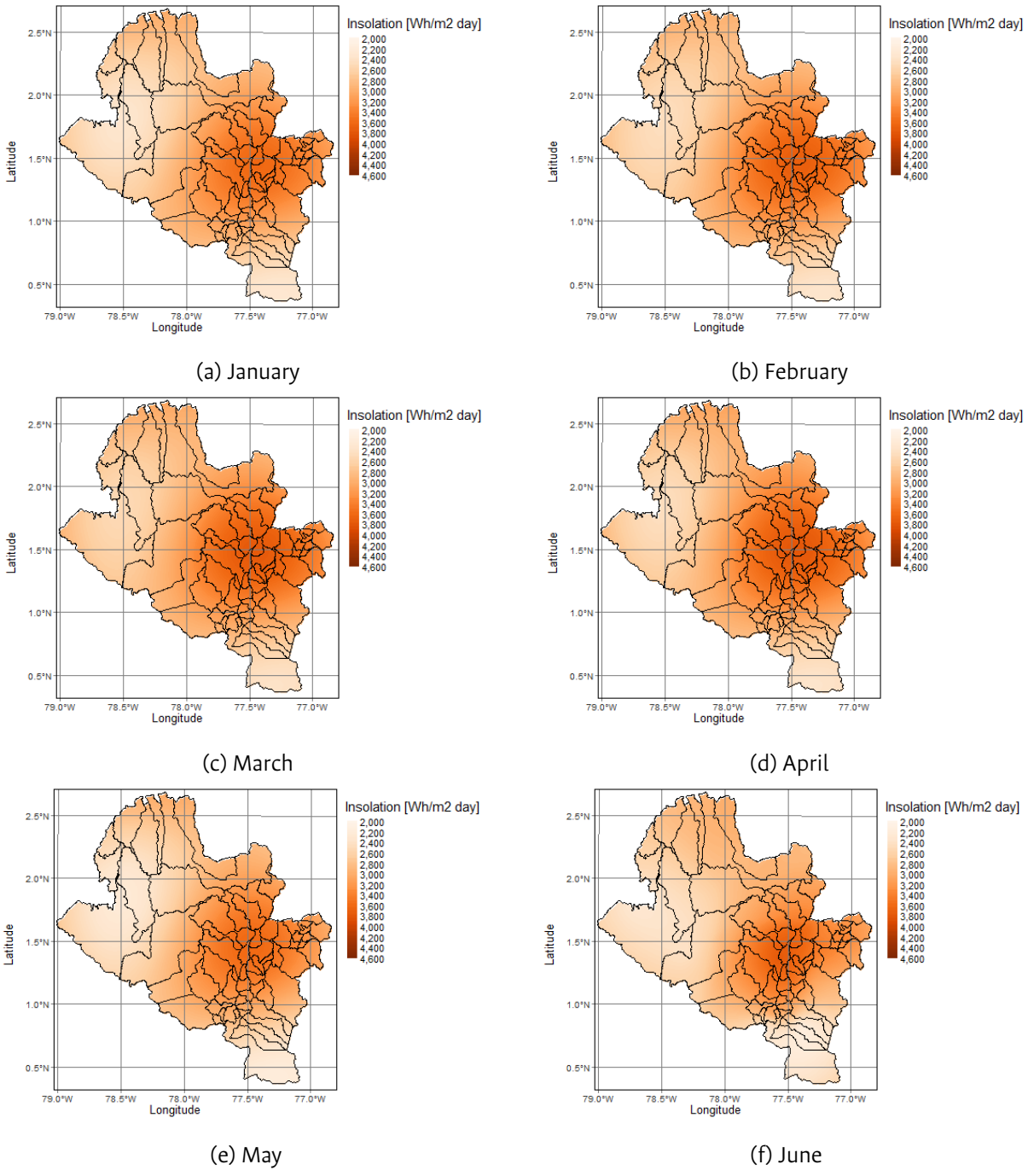
From the maps, it is possible to conclude that a typical solar power plant with a 1 kW capacity, evaluated in the lower and better case, located in the Andean zone could generate between 73,15 kWh and 81,24 kWh in December and October, respectively. The same power plant located in the Pacific zone could generate between 52,44 kWh and 61,13 kWh in November and March, respectively, and in the Amazon zone, the electricity generation is between 51,31 kWh and 62,42 kWh in June and October, respectively.

4.6 Conclusions

Solar insolation estimation allowed confirming that the relationship between AWS altitude and logistic model empirical coefficients had better performance in the Andean than the Pacific zones. The Hargreaves and Samani model had better results in the Pacific zone.

The authors estimated solar insolation with the HS model in the Pacific zone, and results presented an underestimated resource by 442,78 $[Wh/(m^2)day]$ and 333,11 $[Wh/(m^2)day]$ for *Granja el Mira* and *Altaquer*, respectively. It is essential to increase the AWS number on the territory to improve the results.

Figure 4-8. First semester solar insolation



(a) January

(b) February

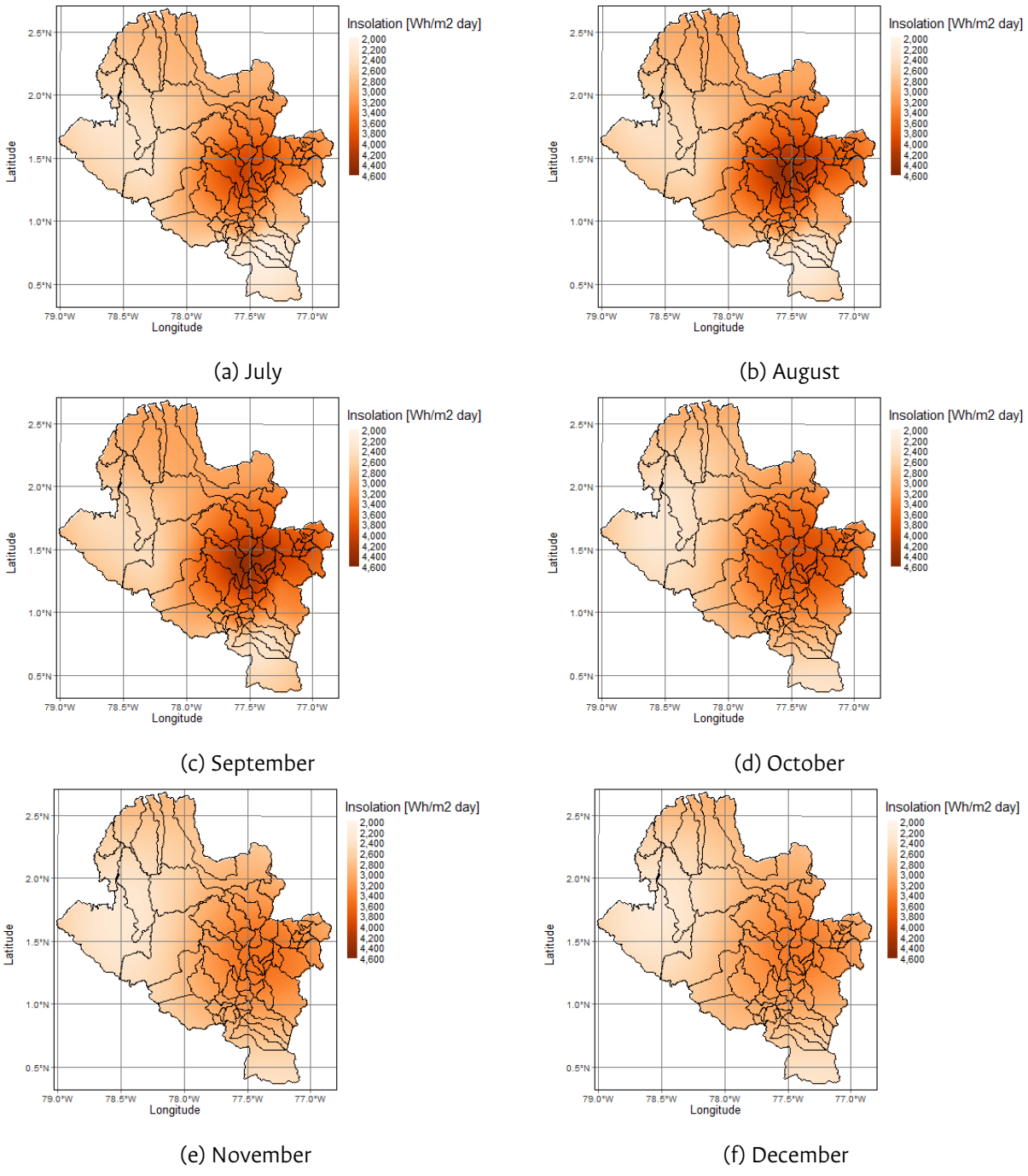
(c) March

(d) April

(e) May

(f) June

Figure 4-9. Second semester solar insolation



(a) July

(b) August

(c) September

(d) October

(e) November

(f) December

The authors estimated the solar insolation in the Andean zone with the logistic model proposed at the beginning of this work. The comparison between measured and estimated values showed an overestimation in *Sandona* and *Ospina Perez*. However, regarding the HS model, the estimation is reduced 51,81 % and 18,55 % for *Ospina Perez* and *Sandona*, respectively. In global terms, the logistical model had better performance than HS model in altitudes above 2.500 MASL.

Comparing between IDW and the Kriging methods, OK presented an improvement in the statistical errors. The following figures confirms that conclusion: RMSE 11,71 %, MAE 3,98 %, MBE 1,01 times, SD 11,40 % and MPE 46,58 %. Although SK showed better results in MAE, SD, and RMSE than OK, the bias with OK is lower. Therefore, OK was the technique selected for mapping the solar insolation potential. UK was the technique that presented lower improvements against the IDW, even in MAE, which reduced the 6,78 % of the accuracy.

From the maps, it is possible to conclude that the obtained estimation is lower than the presented in the IDEAM solar radiation maps over the whole Nariño territory. However, it is also possible to determine the existence of a zone with a high potential for electricity generation with solar PV power plants.

5 Short-term forecasting of global solar irradiance and insolation in tropical environments with incomplete data[1]

Abstract

Accurate mechanisms for forecasting solar irradiance and insolation 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 compares four forecasting approaches Autoregressive Integrated Moving Average, Single Layer Feed Forward Network (SL-FNN), Multiple Layer Feed Forward Network (ML-FNN), and Long Short-Term Memory (LSTM) 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 environments.

5.1 Introduction

The proliferation of solar power generation systems has promoted the interest in solar irradiance and insolation forecasting models. Accurate systems allow estimating the electricity generation in the long-medium- and short-term. This information is crucial to maintain the balance between energy demand and supply (Dannecker, 2015), as well as minimize costs associated with start and shutdown of conventional power plants (Badosa et al., 2017). The solar irradiance is the amount of solar energy on an specific area during a specific time interval, in this case solar irradiance refers to the solar energy in a hourly interval $[Wh/m^2]$. Solar insolation measures the cumulative solar energy on a surface, in this case solar insolation is the solar energy in a daily interval $[W/m^2day]$ (Sandia National Laboratories, 2021)

Time series analysis is the process of examining recorded data over time in order to develop a mathematical model (Shumway & Stoffer, 2011). These models facilitate the implementation of photo-voltaic

[1] This article had the participation of Belizza Ruiz and Francisco Ruiz

(PV) systems, both on-grid and off-grid. Data recorded in extended periods of time is used to understand the behavior and predict future solar irradiance values in a specific location (Suehrcke, 2000). Furthermore, the study of solar irradiance becomes an important subject due to local or global laws seeking to reduce greenhouse gas (GHG) emissions.

The solar irradiance is a time-dependent phenomenon composed of a deterministic and stochastic part (Boland, 2008). Mathematical models allow predicting the exact future value of the deterministic part. On the other hand, the stochastic part outputs a future value between two limits with a confidence interval (Box et al., 2016). The forecasting accuracy relies on the stochastic component ability to model the solar radiation changes induced by the clouds (Inman et al., 2013). There are several approaches to forecast the solar radiation such as persistence methods that assume that the value at time $t + 1$ is equal to value at time t (Diagne et al., 2013), autoregressive models, e.g., autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA), which allow modeling stationary and non-stationary variations and describing complex nonlinear atmospheric phenomena (Inman et al., 2013), and soft computing techniques, e.g., support vector machine (SVM), artificial neural network (ANN), and fuzzy and genetic algorithms (GA) (Demirhan & Renwick, 2018). The ANN, fuzzy logic, and hybrids are robust for modeling the physical processes' stochastic nature, like the solar irradiance because of their capacity to compensate systematic errors and problematic learnable deviation (Paulescu et al., 2013).

The selection of a forecasting method depends on the desired timescale, e.g., intra-hour (15 min to 2 h), intra-day (1h to 6h) and day ahead (1 day to 3 days) (Diagne et al., 2013). Statistical approaches usually perform well for short-term forecasting, such as ARIMA and ANN. For long-term analysis, soft computing techniques are frequently preferred (Coimbra et al., 2013; Demirhan & Renwick, 2018). The autoregressive models describe the characteristics and behavior of the time series using an autoregression process (Antonanzas et al., 2016; Dannecker, 2015). ARIMA is an extension of the ARMA model, which models non-stationary time series. The motivation for using these forecasting models is their robustness to random errors and outliers (Diagne et al., 2013; Sobri et al., 2018). An ANN is a statistical model that establishes a relationship between the input and output data during a training process through layers formed by interconnected nodes of inputs, outputs, hidden layers, and activation functions. It has become one of the most popular solar power forecasting technique (Mazorra-Aguiar & Díaz, 2018; Antonanzas et al., 2016). Long Short-Term Memory (LSTM) network is an advanced Recurrent Neural Network (RNN), which has been recently used in the renewable energy field (Chandola et al., 2020). LSTM learns the dependence between successive data (Ghimire, Deo, Downs, & Raj, 2019). Some studies suggest that LSTM outperforms other state-of-the-art models in forecasting day-ahead solar irradiance (Husein & Chung, 2019).

The implementation of these forecasting techniques requires historical data sets. The datasets used in this study were collected by the Colombian Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM) in 12 Automatic Weather Stations (AWS) located in the Nariño - Colombia. This region is situated in the tropic between the $0°21'54''$ and $2°41'10''$ latitudes and $-79°0'43''$ to $-76°50'13''$ longitude. The altitude of the AWS ranges from 16 to 3.577 MASL. This location has three geographical

subregions, called the Pacific, Andean and Amazonia zones. As usual in solar irradiance measurements, these datasets have a considerable amount of missing data. To overcome this problem, Layanun et al. proposed a data-missing imputation technique for a seasonal ARIMA-based forecasting method in Bangkok in Thailand that averages the data classified by weather type determined with temperature and humidity information (Layanun et al., 2017). Likewise, Rodríguez-Rivero et al. applied an average smoothing technique to fill the time series's missing data to short-term forecasting. Ogunsola & Song used three approaches to solar irradiance data imputation Singular Spectral Analysis (SSA), Statistically Adjusted Solar Radiation (SASR), and Temperature-based Approach (TBA). They recommend using different imputation techniques based on the gap length (Ogunsola & Song, 2014). In our experiments, we implement an imputation technique that replaces the missing data with a value estimated from the existing data (Demirhan & Renwick, 2018; Moritz & Bartz-Beielstein, 2017).

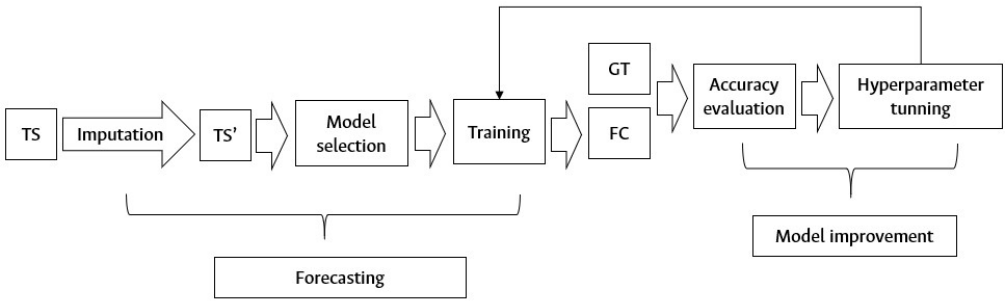
This study applies four state-of-the-art prediction models for global solar irradiance and insolation one day-ahead forecasting in tropical and mountainous environments with incomplete data: i) ARIMA, ii) Single Layer Feedforward Neural Network (SL-FNN), iii) Multi-Layer Feedforward Neural Network (ML-FNN), iv) and Long Short-Term Memory (LSTM). In the irradiance case, the models forecast the hourly irradiance values one day ahead. For the insolation, which in this case is the sum of the irradiance hourly values, the models forecast the insolation one day ahead. The missing data imputation combines a linear interpolation of the subsequent values with the average of past values measured at the same hour of the imputed data for hourly data, and a TBA for daily data. To remove the deterministic effect, it is the clearness index used instead of the global solar irradiance values directly. The results show that the NN-based techniques outperform ARIMA. LSTM model has a better performance in AWS with less amount of data in the Pacific zone. SL-FNN achieve the best behavior in the AWS when more data is available.

5.2 Materials and Methods

This section contains the description of four forecasting models (the autoregressive statistical method ARIMA, and three models based on neural network architectures: SL-FNN, ML-FNN, and LSTM).

Figure **5-1** describes our forecasting methodology. As a preprocessing step, we perform "quality control" on the input data by removing anomalous data. After that, we fill the time series gaps (missing data). In the model selection step, we consider four options: ARIMA, SL-FNN, ML-FNN and LSTM. In the training step, the model's parameters are computed in such a way that fit the training data. To evaluate the model, we use the parameter values updated until the t-th day to predict the irradiance or insolation in the $t+1$-th day. We use the performance measures to tune the hyperparameters.

Figure 5-1. This diagram shows the stages for training the forecasting models and hyperparameter tunning where TS, TS', FC, and GT stand for input time series, imputed time series, forecasting (model output), and ground-truth (value provided by the AWS), respectively.



5.2.1 ARIMA

ARIMA consists of three components: i) autoregression (AR), ii) integration (I), and iii) moving average (MA). AR and MA deal with the stochastic elements, and I renders the time series stationarity. ARIMA is denoted as ARIMA (p, d, q), where p and q are the AR and MA order, respectively, and d is the number of derivatives applied to the time series (e.g., $d = 0$ means that the time series is already stationary, and $d = 2$ means that two derivatives are needed to make the data stationary) (Agami Reddy, 2011). ARIMA (p, d, q) is defined by:

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

or

$$\left(1 - \sum_{i=1}^{p} \Phi_i B^i\right)(1-B)^d x_t = \left(1 + \sum_{i=1}^{q} \theta_i B^q\right) u_t \tag{5-2}$$

where B is the backward shift operator that may be treat as a complex number, $\Delta = 1 - B$ is the backward difference, Φ_p and θ_q are polynomials of order p and q respectively. $(1-B)^d x_t$ is the responsible to transform the non-stationary time series in a stationary one. After applying this transformation, it is possible to use any forecasting strategy for stationary data (Montgomery et al., 2008; Shumway & Stoffer, 2011).

AR describes the past behavior of the time series and series residual at the actual time as a weighted linear combination of values of a dataset of a stochastic process x_t (Dannecker, 2015) and a white noise u_t as follows

$$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_p x_{t-p} + u_t = \sum_{i=1}^{p} \Phi_i x_{t-i} + u_t. \tag{5-3}$$

Using the backshift operator the equation 5-3 is expressed as follows:

$$u_t = \left(1 - \sum_{i=1}^{p} \Phi_i B^i\right) x_t. \tag{5-4}$$

MA describes the time-series' random perturbations by a weighted linear combination of previous values of a white noise error. Thus, the time series is represented as a set of uncorrelated and normal-distributed random variables, as follows:

$$x_t = \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} = \left(1 - \sum_{i=1}^{p} \theta_i B^q\right) u_t. \tag{5-5}$$

5.2.2 Feedforward Neural Networks (FNN)

An FNN is an ensemble of units, known as neurons, connected by synaptic joints each one with a weight coefficient (Blaga et al., 2019). An FNN can be divided into three parts: input, hidden layers, and output. The first part receives the input data. The hidden layers connect the input and output. The output layer outcomes the computed values (Premalatha & Valan Arasu, 2016). The FNN training requires an iterative backpropagation procedure that learns an input-output mapping. This process has four steps: i) forward propagation of the training pattern input, ii) error calculation by a loss function that compares estimated and reference values, iii) backpropagation of the error to recompute each weight Δw_{ij} from the output to the hidden layer, and iv) weight's w_{ij} updating: $w_{ij}^{new} = w_{ij}^{old} + \lambda \Delta w_{ij}$, where λ is known as learning rate (Blaga et al., 2019).

During forward propagation, the inputs $\mathbf{x}^{i-1}$ are multiplied by the weights w, the individual results are summed-up, and a bias b is added to the results as an offset value as follows

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

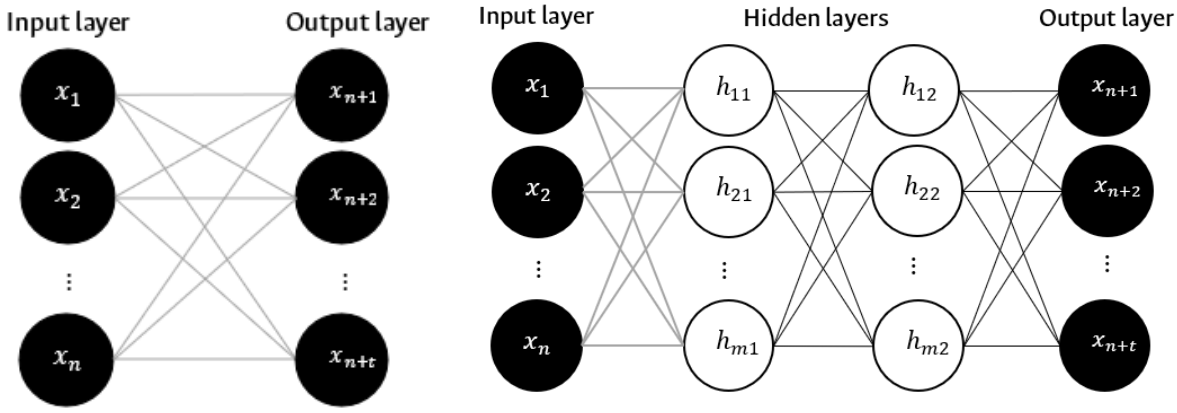
where i is the current layer (if $i = 1$, $\mathbf{x}^0$ is the input), n is the n-th neuron in the current layer, k is the k-th neuron in the previous layer, and N_{i-1} is the number of neurons in the previous layer. This result is passed through a so-called activation function (Ghimire, Deo, Raj, & Mi, 2019) such that

$$\mathbf{x}^{(i)} = f(\mathbf{z}^{(i)})$$

where $\mathbf{x}^i$ is the current layer's output, and $f(\cdot)$ is typically used to bound the signal or induce non-linear interactions. In this way, the signal passes all the layers until reaching the output. The weights are initialized with random values. Both weights and bias are iteratively updated until a stopping criterion is fulfilled (Munawar & Wang, 2020).

In this study, we consider two FNN architectures: a Single Layer FNN (SL-FNN) that links directly input and outputs without activation functions, and a multi-layer FNN (ML-FNN) with two hidden layers and ReLU activation functions between them. Note that our SL-NN and ML-NN architectures learn a linear and non-linear mapping, respectively. Figure **5-2** illustrates both of them.

Figure 5-2. Architecture FNN



(a) FNN with a single hidden layer

(b) FNN with multiple hidden layers

5.2.3 LSTM

In addition to the FNN models described above, we consider a recurrent neural network (RNN) architecture. RNN is an efficient tool to deal with temporal patterns due to the capacity of remember previous data (Chandola et al., 2020). However, training a RNN is usually difficult due to the vanishing gradient problem (Husein & Chung, 2019). Given an input $x_{t-1} = x_1, x_2, x_n$, the output x_t in a RNN is given by:

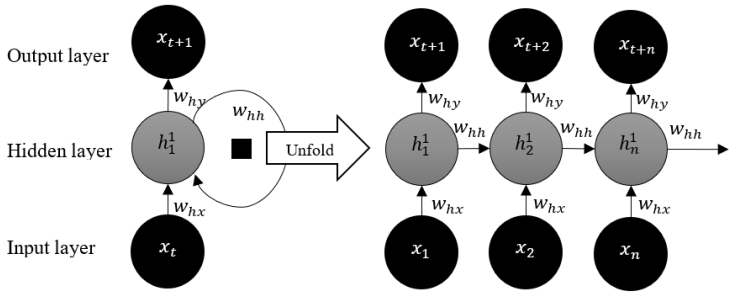
$$\begin{aligned} x_t &= w_{hx_t} h_t + b_{x_t} \\ h_t &= H(w_{hx_{t-1}} x_{t-1} + w_{hh} h_{t-1} + b_h) \end{aligned} \tag{5-6}$$

where $w_{hx_{t-1}}$, w_{hh}, w_{hx_t} are input-hidden, hidden-hidden and hidden-output weight matrices, b_h, and b_{x_t} are hidden and output bias vectors, respectively. H term is the hidden layer activation function.

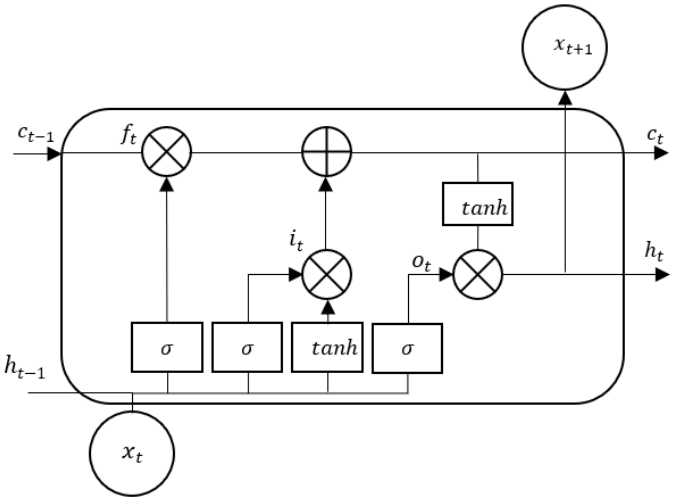
LSTM is an advanced RNN that resolves the gradient problem including an explicit memory to the network. LSTM has an input, forget gate, output gate, and a cell unit that serves as memory for a defined time interval (see Figure **5-4**a). The gates control the flow of information that enter and leave the cell unit, see Figure **5-4**b (Husein & Chung, 2019). The forget gate f_t determine the influence of the previous state on the current state. The input gate i_t receives the new information to update the cell state. The output gate o_t provide the information based on the cell state. The sigmoid function σ adjust the output values of these gates to a value between 0 to 1, that are interpreted as a probability (Kwon et al., 2020).

The explicit memory of the LSTM makes this technique appropriate for a short-term, near real-time forecast model (Ghimire, Deo, Downs, & Raj, 2019). The LSTM unit has three operation states: 1) Input gate activate: the cell memory accumulates new information. 2) Forget gate activate: the cell erases the accumulated information. 3) Output gate activate: the final output propagates to the ultimate state.

Figure 5-3. RNN



(a) RNN Structure



(b) LSTM architecture

In the LSTM forward propagation the forget gate computes

$$f_t = \sigma(w_{xf}x_t + w_{hf}h_{t-1} + b_f) \tag{5-7}$$

where h_{t-1} represents the last previous state, and $\sigma(\cdot)$ is the logistic sigmoid function. The LSTM cell internal status is updated with a conditional self-loop weight f_t as follows:

$$c_t = f_t c_{t-1} + i_t \tanh(w_{xc}x_t + w_{hc}h_{t-1} + b_c). \tag{5-8}$$

The cell unit c_t is a linear self-loop controlled by the forget gate unit f_t that determines the forward contribution of c_{t-1}. Then, the external input unit i_t is calculated similarly to the forget gate with its own parameters:

$$i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + b_i). \tag{5-9}$$

The output h_t of the LSTM cell can also be shut off, via the output gate o_t, which also uses a sigmoid unit for gating:

$$h_t = o_t \tanh(c_t) \tag{5-10}$$

where $o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + b_o)$, and i, f, o and c are the input, forget, output gate and cell, respectively. Note that i, f, o and c have the same size as the hidden vector h.

5.2.4 Quality control data

To reduce the risk of biases induced by inaccurate data, we apply a quality control process that consist of three stages: i) Data completeness checking: samples that lack essential information (station code, variable code, date and time, and data value) are removed from the data set. ii) anomaly detection based on fixed limits: to avoid over estimate the resource, we discard the samples that exceed the maximum hourly extraterrestrial solar irradiance value (I_0). Therefore, it is required that $I_0 \geq I_{mt}$ where I_{mt} is the measured global solar irradiance at time t, and I_0 is

$$I_{sc}\left[1 + 0,033cos\left(360\frac{D-3}{365}\right)\right] * sin\beta \tag{5-11}$$

where D is the Julian day, I_{sc} is the solar constant $(1.367\,[W/m^2])$ representing the energy from the Sun per unit area of the surface perpendicular to the irradiance propagation direction (Şen, 2008) and $sin\beta = cos\phi cos\delta cos\omega_s + sin\phi sin\delta$. iii) anomaly detection based on hourly limits: to avoid underestimate the resource, we use as lower bound the 3% of the clear-sky global solar irradiance (I_{cst}), where $I_{cst} = I_0\tau$. We consider the samples under the lower bound as anomalous values and remove them from the dataset.

To estimate the atmospheric transmittance τ and I_{cst}, we implement the Kreith & Kreider model (Şen, 2008):

$$\tau = 0,56\left(e^{-0,65/\sin\beta} + e^{-0,095/\sin\beta}.\right) \tag{5-12}$$

The procedure above follows the UNE500540 regulation (AENOR, 2004), and the recommendations in (Estévez et al., 2011), which applied quality control to solar irradiance data in Spain.

5.2.5 Data imputation

Data imputation is the process of completing the missing data with reasonable values. In this case, the authors use the clearness index K_t, which is the ratio between the solar irradiance at ground level and the clear sky global solar irradiance, computed with the Kreith & Kreider model, instead of dealing with the solar irradiance data directly.

In our experiments, for hourly data we filled out the missing data on the first day with a value of 1. Afterward, the technique considers three cases: if the missing data is at i) 6, ii) 18, or iii) between 7 to 17 hours. In the first case, the missing data is the average between the value at the same hour of the day before with the clearness index value of the 7 hours of the current day. On the second case, the value is the result of the average between the data at 17 hours of the current day and the value at 18 hours of the day before. In the third case, the imputed value corresponds to the average between the value at the same hour the day before and the value of the immediately previous and following hours. If the immediately next value is missed, it is not considered in the calculation.

For daily data imputation, we used empirical temperature-based models such as Hargreaves and Samani (Hargreaves & Samani, 1982),

$$\frac{H}{H_0} = a\left(T_{max} - T_{min}\right)^{0,5} \tag{5-13}$$

where H is the daily global solar insolation, H_0 is the daily extraterrestrial global solar insolation, a is an empirical coefficient computed from each dataset, T_{max} is the maximum daily temperature, and T_{min} is the daily minimum temperature, and a new approach based on logistic regression,

$$\frac{H}{H_0} = \frac{1}{1 + e^{-(a+b(T_{max}-T_{min}))}} \tag{5-14}$$

where a and b are empirical coefficient computed from each dataset. These models estimate the solar insolation from the difference between the daily maximum and minimum temperature.

5.2.6 Model performance criteria

We use statistical validation to measure the forecasting performance. Table **5-1** shows the computed errors. For MAE and RMSE, the lower the better, and for MBE, the closer to zero the better (Abreu et al., 2018). The performance evaluation only considers the measure data, since the imputed data is used exclusively for training. Therefore, the amount of days for the error estimation is different from each AWS's total amount of data.

Table 5-1. Statistical errors

Measurement	Definition	Formula*
Mean absolute error (MAE)	It is the average vertical distance between each predicted and observed point. This measure quantifies the error with more emphasis on the mean and less on individual extreme events.	$\frac{1}{n}\sum_{i=1}^{n}\lvert p_i - o_i\rvert$
Root mean square error (RMSE)	It provides a measure of the error size and is sensitive to outlier values because this measure gives much weight on large errors. It captures variability rather than the overall trend.	$\left[\frac{1}{n}\sum_{i=1}^{n}\left(p_i - o_i\right)^2\right]^{1/2}$
Mean bias error (MBE)	This measure provides information on the long-term performance of the model, when the model has a systematic error that presents overestimated or underestimated predictors. Low values of MBE are desirable, though it should be noted that an overestimated data set will cancel another underestimated data set.	$\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)$

* p_i is the predicted value,o_i is the observed value, n is the amount of data

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 Experimental set-up

5.3.1 Location and dataset

In this study, we use irradiance data from twelve AWS (see Table **5-2**) located in Nariño, Colombia, as shown in Figure **5-4**. This region is located in the Intertropical Convergence Zone, where the Andean mountain range splits into two mountain ranges. This zone is formed by three sub-regions, the Pacific, Andean, and Amazon. The altitude of all the AWS ranges from 42 to 3.577 MASL. These geographical characteristics allow assessing the studied forecasting techniques on different physio-graphic and environmental conditions.

Figure 5-4. Location of weather stations

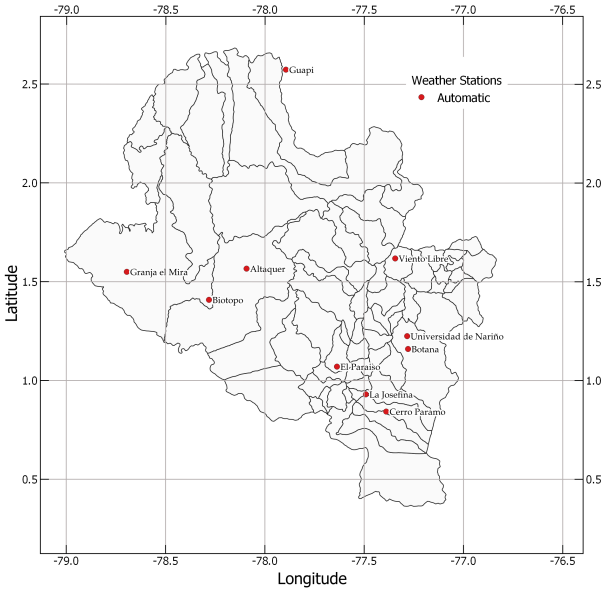
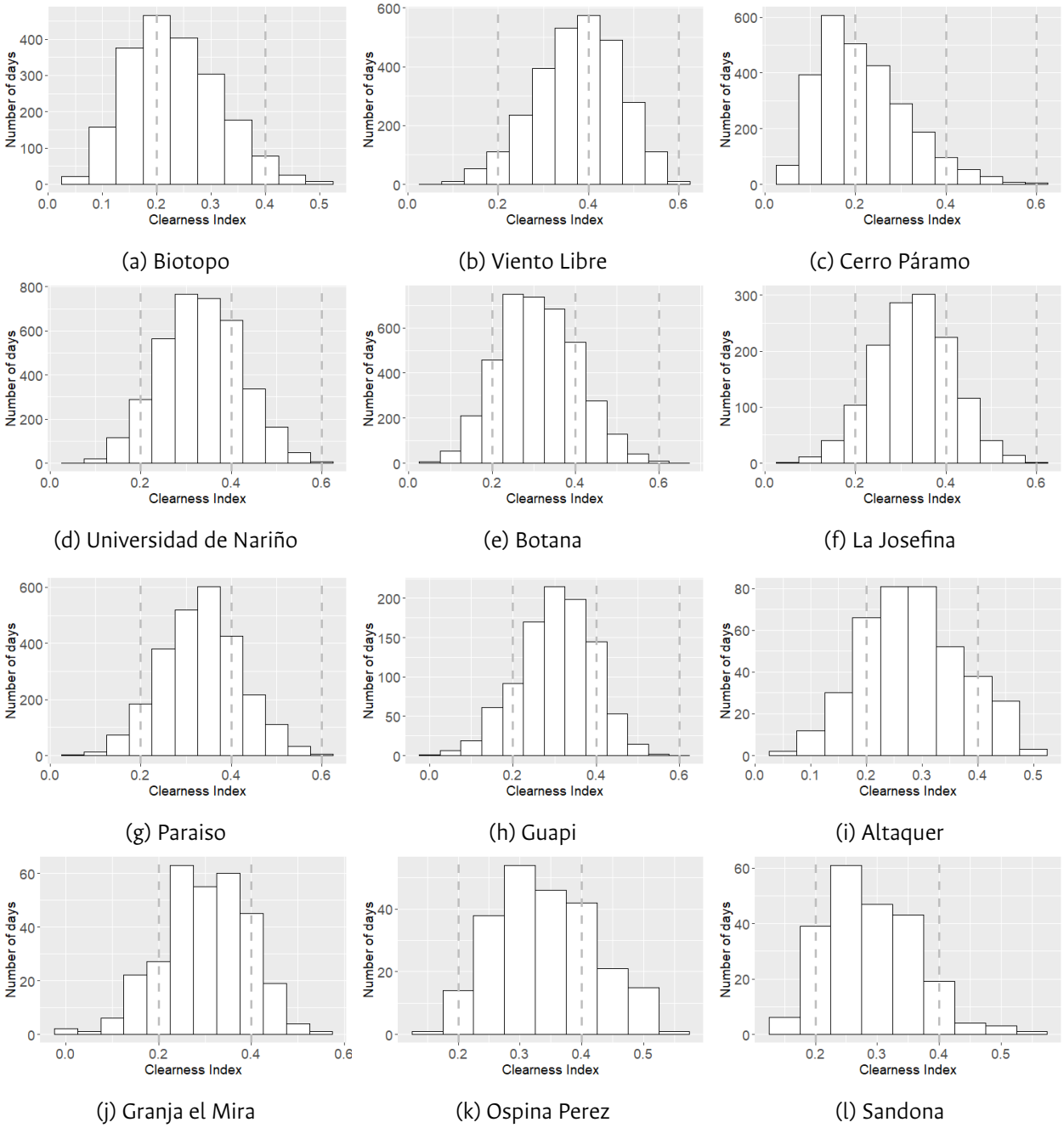


Table 5-2. Automatic Weather Stations

Name	Latitude	Longitude	Altitude	Period	Region
Biotopo	1,41	-78,28	512	2005-2017	Pacific
Altaquer	1,56	-79,09	1.1010	2013-2014	Pacific
Granja el Mira	1,55	-78,69	16	2016-2017	Pacific
Guapi	2,57	-77,89	42	2005-2017	Pacific
Cerro-Páramo	0,84	-77,39	3.577	2005-2017	Amazon
La Josefina	0,93	-77,48	2.449	2005-2017	Andean
Viento Libre	1,62	-77,34	1.005	2005-2017	Andean
Universidad de Nariño	1,23	-77,28	2.626	2005-2017	Andean
Botana	1,16	-77,27	2.820	2005-2017	Andean
El Paraiso	1,07	-77,63	3.120	2005-2017	Andean
Sandona	1,30	-77,46	1.838	2013-2017	Andean
Ospina Perez	1,25	-77,48	1.619	2013-2017	Andean

Figure **5-5** shows the global solar radiation data grouped by the clearness index K_t, based on the following categories (Rivero et al., 2017): cloudy days $0,0 < K_t \leq 0,2$, partially high cloudiness $0,2 < K_t \leq 0,4$, partially low cloudiness $0,4 < K_t \leq 0,6$, sunny $0,6 < K_t \leq 0,75$, very sunny $0,75 < K_t \leq 1,0$. Note that *Biotopo* and *Cerro Páramo* have mostly cloudy days, and the others have mostly partially high cloudy days.

Figure 5-5. Days classification with clearness index



(a) Biotopo (b) Viento Libre (c) Cerro Páramo

(d) Universidad de Nariño (e) Botana (f) La Josefina

(g) Paraiso (h) Guapi (i) Altaquer

(j) Granja el Mira (k) Ospina Perez (l) Sandona

5.3.2 Data pre-processing

Forecasting methods aim to model the global solar irradiance's stochastic component. Therefore, in the pre-processing stage, we remove the time series' trend to reduce the deterministic effect. To this end, we use the clearness index K_t (Benali et al., 2019; Diagne et al., 2013). The K_t ranges between 0 to 1 and indicates the amount of solar irradiance in a clear sky that reaches the Earth surface. Therefore,

$K_t = 0$ is obtained on an completely overcast day, while $K_t = 1$ is achieved on a very sunny day.

5.3.3 Forecasting models' architectures

We use the forecasting models for the one-day ahead prediction of solar irradiance and insolation. We grouped the neural network architectures into SL-FNN, ML-FNN, and LSTM, and design two architectures per group (for irradiance and insolation forecasting in each case). Each architecture outputs a vector $\in R^{13}$ and R, respectively. In SL-FNN, the input directly feeds the output, so, the size of the input is 130 for solar irradiance forecasting (13 values per day), and 10 for insolation (1 value per day), and the ouput is 13 and 1, respectively. Our ML-FNN consists of two hidden layers with 130, and 10 neuron per hidden layer for irradiance and insolation, respectively, and ReLU activation functions between them. In LSTM, we use a a hidden layer (memory) with the same size as the input (130 or 10 accordingly) and a fully connected layer between this layer and the output. Note that there is not activation function that bounds the output in any of the neural network models.

5.3.4 Training

For training ARIMA, we use a sliding window of 10 days. We move the window one day at a time and fit the model each shift. For tuning the hyperparamters p, d, and q (see (5-1)), we carry out an exhaustive search in $p \in \{1, 2, 3\}$, $d \in \{0, 1, 2\}$, and $q \in \{1, 2, 3\}$, and choose the best (p, d, q) each step according to the Akaike information criterion (AIC). Likewise, the neural network models use the 10-day sliding window as input and forecast the next day data. For tuning the learning rate, we randomly choose an AWS (Biotopo), and select the one that outputs the lowest RMSE in $\lambda \in \{1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. We set $\lambda = 10^{-2}$ then. For updating the network parameters (weights and biases), we use batches of 10 consecutive windows paired with their corresponding next-day data. As loss function, we use the mean square error (MSE) that compares the next day prediction and ground-truth. We do not introduce additional regularization terms to the loss function such as dropouts. We compute the forecasting performance progressively each step before updating the networks' parameters disregarding imputed outputs. In our experiments, we use the built-in ARIMA algorithm of the Python module "statsmodel 0.11.1," and the deep-learning framework "Pytorch 1.4.0" for implementing the neural network models.

5.4 Results and discussion

This section contains the results and analyses of the application of imputation methods and forecasting models on irradiance and insolation data.

5.4.1 Imputation results

For the irradiance data analyses, our time frame of interest is from 6:00 to 18:00. Therefore, the collected data of a day is complete when the dataset contains 13 data values in this interval (one per hour). On average (for all the 12 AWS), $69,26\%$ of training data was imputed. Table **5-3** contains the imputation details (total number of measured and missing values on the left, and number of missing data per time series quarter on the right). For instance, in *Biotopo*, there are 10.971 missing values in the first quarter (data collected between 2005/11/26 and 2008/11/04).

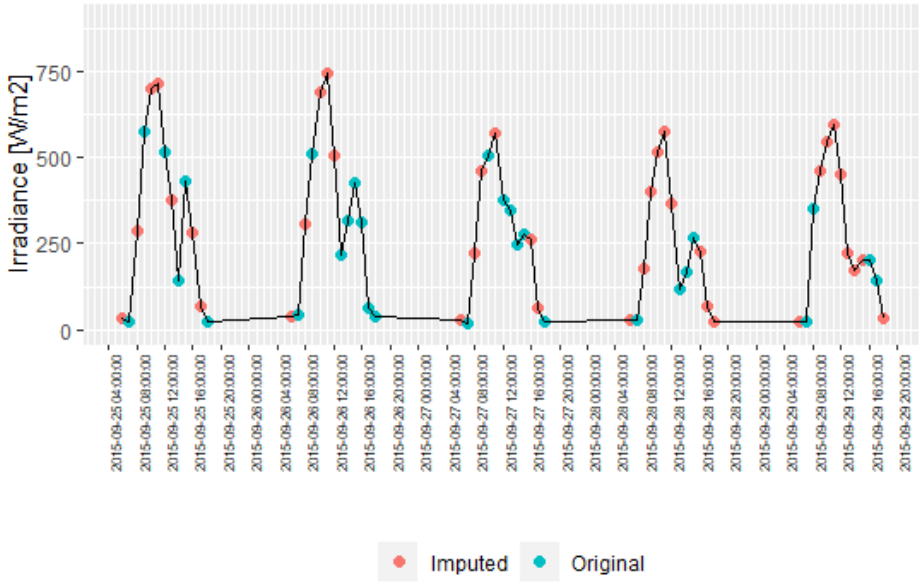
Table 5-3. Details of the hourly data used in the forecasting experiments.

		Time series details				Missing values per quarter			
AWS Name	Time frame	# of measured values	# of missing values	% of missing values	Total	1^{sd}quarter # of data	2^{nd}quarter # of data	3^{th}quarter # of data	4^{th}quarter # of data
Biotopo	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
Guapi	2005/10/09-2017/08/30	13.247	43.225	$76,5\%$	56.472	10.347	10.565	13.040	9.273
Cerro Páramo	2005/11/25-2017/08/31	22.072	33.802	$60,5\%$	55.835	9.204	7.959	8.209	8.430
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
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
Paraiso	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

La Josefina is the AWS with most missing data. For training the forecasting models with this station, 84,2 % data needs to be estimated. Most of its missing values are in the second quarter of the time series with 13.880 missing values $(29,5\%)$. The longest gap with consecutive missing values is 12.999, approximately 2,7 years, and the one-size gap (one missed value between two known values) occurs 2.242 times. During 2010 there is not any register, therefore, for training the models, all values during that period were refilled. Figure **5-6** shows the imputation process in a five-day time frame. As it can be seen, the imputed values follow the trend of the actual values. The statistical error shows that, in this case, the imputation process underestimates the resource by $-15,53[Wh/m^2]$ (see Table **5-4**).

In the *Biotopo* time series, the largest percentage of missing data corresponds to the third quarter $(29,6\%)$, between the position 27.933 to 41.898. The longest gap with missing values is 7.252, and the one-size gap occurs 2.560 times. In *Guapi*, there are 13.040 $(30,2\%)$ missing values during the third quarter of the time series, the longest gap size with missing values is 3.913, and the one-size gap occurs 2.582 times. In the second quarter of the *Ospina Perez* time series, there are 1.392 missing values $(33,7\%)$, the longest gap size with missing values is 1.813, and the one-size gap occurs 564 times. *Sandona* does not register any data during the second quarter, reaching 1.391 missing values $(33,8\%)$ then, the longest gap size with missing values is 1.808, and the one-size gap occurs 456 times. *Viento Libre* lacks 12.450 values $(31,3\%)$ during the first quarter of the time series, the longest gap size is 7.325, and the one-size gap occurs 4.944 times. *Paraiso* lacks 11.259 values $(29,4\%)$ in the first quarter of the

Figure 5-6. Data imputation *La Josefina* in 2015



time series, the longest gap size values is 3.856, and the one-size gap occurs 4.475 times. *Granja el Mira* has 1.044 missing values $(30, 1\%)$ in the third quarter, the longest gap size is 621, and the one-size gap occurs 513 times. In *Universidad de Nariño*, the most critical case is the second quarter that lacks 9.655 values $(26, 1\%)$, the longest gap size is 914, and the one-size gap occurs 4.612 times. *Botana* has 9.666 missing values $(27, 3\%)$ in the second quarter, the longest gap size is 1.661, and the one-size gap occurs 5.791 times. *Cerro Páramo* has 9.204 missing values $(27, 2\%)$ during the first quarter, the longest gap size is 1.150, and the one-size gap occurs 3.151 times. In *Altaquer*, the most critical section is the second quarter with 922 missing values $(29, 7\%)$, the longest gap size is 445, and the one-size gap occurs 632 times.

Viento Libre, Cerro Páramo and *Paraiso* lack the largest number of items in the first quarter of the time series, with an average of $(29, 3\%)$. In the *Altaquer, Universidad de Nariño, La Josefina, Botana, Ospina Perez* and *Sandona*, the second quarter of the time series presents the largest percentage of missing data $(30, 0\%$ on average). In *Sandona*, there is not any registered data during that time frame. In the third quarter of the time series, *Biotopo, Granja el Mira* and *Guapi* lack $(30, 0\%)$ of the data on average, being the quarter with the largest percentage of missing values.

Table **5-4** shows the error estimates of the imputation process in the irradiance data. According to MAE, *Cerro Páramo* exhibits less extreme events. Consequently, the imputed values follow the mean trend. The RMSE shows that *Sandona* and *Granja el Mira* have more variability and outlier values. The average RMSE is 118,15 $[Wh/m^2]$. In the AWS located in Pacific, the RMSE is 115,84 $[Wh/m^2]$, 127,63 $[Wh/m^2]$ in the Andean zone, and 103,69 $[Wh/m^2]$ in the Amazon zone. The MBE shows that in all AWS the resource was underestimated by $12, 83[Wh/m^2]$ on average, with a minimum in *Ospina Perez* of $2, 78[Wh/m^2]$ and a maximum in *Guapi* of $16, 79[Wh/m^2]$. Likewise, the MAE is 81,07 $[Wh/m^2]$ on average. In the Pacific zone, MAE is 76,75 $[Wh/m^2]$, 85,92 $[Wh/m^2]$ in the Andean zone, and 64,45

$[Wh/m^2]$ in the Amazon zone.

Table 5-4. Statistical errors in the imputation process for solar irradiance forecasting

Regions	AWS Names	MAE $[W/m^2]$	RMSE$[W/m^2]$	MBE$[W/m^2]$
Pacific	Biotopo	69,46	104,16	-14,72
	Altaquer	68,24	107,37	-15,82
	Granja el Mira	92,22	134,83	-16,18
	Guapi	77,09	117,00	-16,79
Amazon	Cerro Páramo	64,45	103,69	-15,07
Andean	Viento Libre	80,41	118,04	-11,28
	Universidad de Nariño	85,41	125,66	-9,81
	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
	Sandona	86,34	136,88	-11,94
	Average	81,07	118,15	-12,82

Table **5-5** shows the errors of the imputation of insolation data. The RSME shows that the Logistic model outperform Hargreaves and Samani model in the Andean and Amazon zones, and in the Pacific zone in the AWS with lower measured data such as *Altaquer* and *Granja el Mira*. In the Andean zone, The average RMSE of the logistic model is $1.022,86[Wh/m^2day]$, and *Sandona* and *Viento Libre* show more variability than the other AWS located in this zone. The average MAE of the logistic model is $833,89[Wh/m^2day]$. In the Pacific zone, the average RMSE with logistic model is $870,38[Wh/m^2day]$ and with Hargreaves and Samani model is $933,50[Wh/m^2day]$. The average MAE of the logistic model is $686,22[Wh/m^2day]$ and with the Hargreaves and Samani model is $737,93[Wh/m^2day]$. In this zone, the logistic model outperform Hargreaves and Samani model in the AWS with lower amount of data (*Altaquer* and *Granja el Mira*). The MBE does not follow a patter as the RMSE and MAE; therefore, this error should be analyzed for each particular case.

5.4.2 Irradiance forecasting

After applying the data imputation process, we train and test the four forecasting models: ARIMA, SL-FNN, ML-FNN, and LSTM. In total, we train 96 forecasting models (8 models for each AWS) in an hourly and daily timestamp. In order to remove the deterministic part of the time series, we use the clearness index instead of solar irradiance or insolation data directly. To quantify the forecasting models' performance, we calculate the MAE, RMSE, and MBE error measures for each AWS.

During 2007, 2008, and 2010, rainfall increased in Colombia because of La Niña. This weather pattern was stronger in 2010 from June to December, with a peak in October and November that registered an Oceanic Niño Index (ONI) of -1,7. Likewise, in 2009, 2015 and 2016, temperature increased because of El Niño. The ONI reached an intensity of 2,6 in December 2015, being the highest value registered by

Table 5-5. Statistical errors in the imputation process for insolation forecasting. Lowest values are in bold for each empirical model

		RMSE $[Wh/m^2day]$		MBE $[Wh/m^2day]$		MAE $[Wh/m^2day]$	
Regions	AWS Names	HS	Logistic	HS	Logistic	HS	Logistic
Pacific	Biotopo	**993,64**	1.113,48	**-2,01**	-37,29	**800,52**	885,10
	Altaquer	761,09	**670,59**	-333,11	**145,14**	596,88	**529,73**
	Granja el Mira	854,87	**781,91**	-442,78	**76,19**	684,98	**596,66**
	Guapi	**878,75**	915,53	**-16,38**	-27,50	**696,35**	733,40
Amazon	Cerro Páramo	1.209,76	**1.152,72**	**21,29**	33,37	946,29	**887,34**
Andean	Viento Libre	1.080,72	**1.077,35**	163,13	**160,23**	**861,64**	862,86
	Universidad de Nariño	1.083,73	**1.019,14**	**62,04**	93,72	884,48	**833,30**
	Botana	1.070,23	**1.042,68**	**42,30**	47,13	881,19	**860,18**
	Josefina	1.066,18	**984,75**	**-20,58**	22,18	830,06	**760,43**
	Paraiso	990,32	**921,32**	-42,52	**-14,98**	806,64	**748,67**
	Ospina perez	987,70	**934,93**	355,08	**173,99**	797,83	**736,86**
	Sandona	1.375,84	**1.179,92**	1.213,27	**988,12**	1.249,26	**1.034,94**

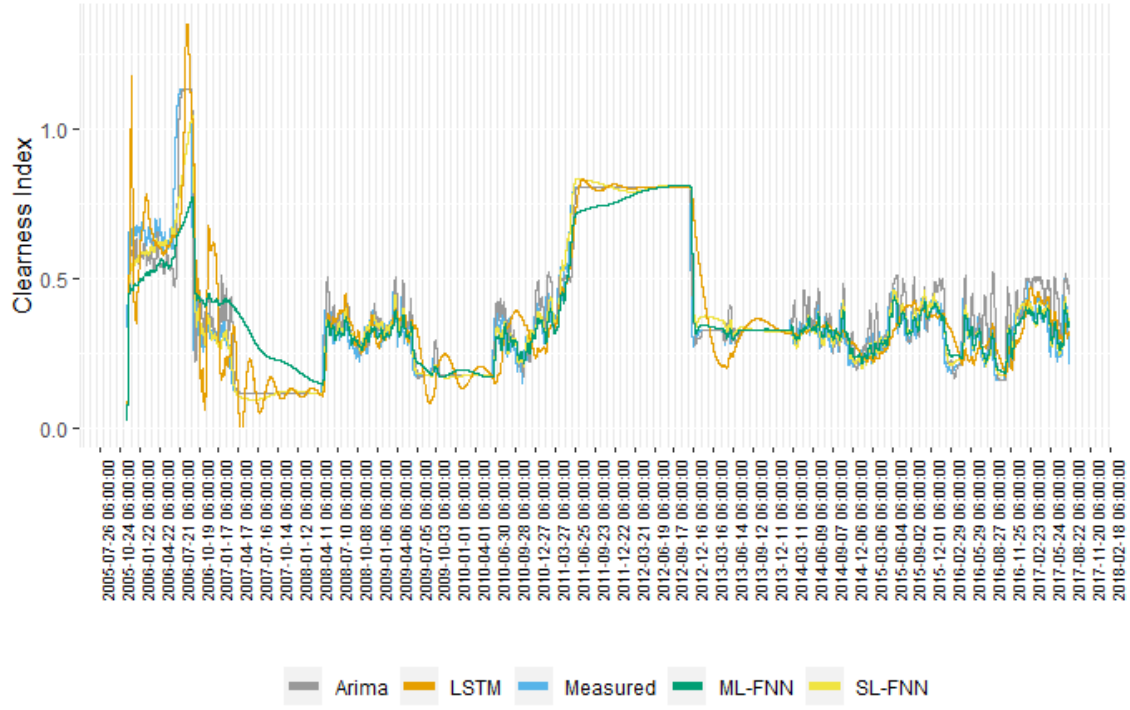
the National Oceanic and Atmospheric Administration (NOAA) since 1950. These natural phenomenons affect the performance of the forecasting techniques due to changes in the cloudiness (National Oceanic and Atmospheric Administration, 2020).

Figure **5-7** contains the hourly measured and predicted data (10-day average) in *Biotopo*. Figure **5-7 (a)** shows the time series in the complete interval of acquisition. In this experiment, ARIMA, SL-FNN and ML-FNN adapt better to fast changes than LSTM, which exhibits a delayed response on these cases. To better understand the models' behavior, Figures **5-7 (b)** and **(c)** show the eight-month starting and ending time intervals. ARIMA outperforms the others in the first part of the time series prediction. We attribute this observation to the large number of parameters in neural networks, which require more training data than ARIMA. On the other hand, Fig. **5-7 (c)** shows that the neural network-based models output a more accurate prediction when observed more data.

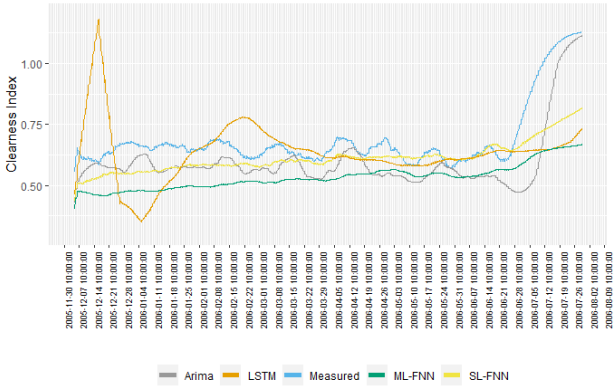
Figure **5-8** shows *Altaquer* results. Figure **5-8 (a)** shows the time series in the complete interval of acquisition. Figure **5-8 (a)** shows that ML-FNN and LSTM do not have enough data inputs to adjust their weights. Figure **5-8 (b)** shows that ARIMA outperform the neural network-based models at the beginning; however, SL-FNN outperform ARIMA when there is six month of measures as input data on average. Figure **5-8 (c)** confirms that SL-FNN model outperform the others when trained with enough data.

Table **5-6** contains the errors for the complete time series and errors by quarters of the time series. In *Biotopo*, the MAE of the complete time series show that ARIMA is the best model, followed by SL-FNN, LSTM, and ML-FNN. We observe that ARIMA, SL-FNN, and LSTM follow a mean trend better than ML-FNN, and that using these models instead of ML-FNN improves the results in 25,54 %, 21,11 %, and 6,62 % respectively. Analyzing the MAE result by quarters, ARIMA is the best model in the first quarter, as Figure **5-7 (b)** shows. However, SL-FNN and LSTM outperform ARIMA in the next two quarters, and in

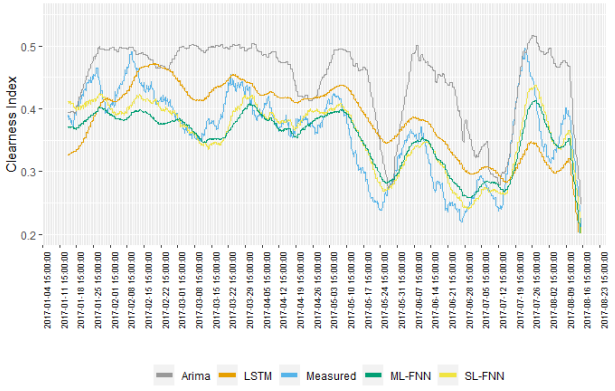
Figure 5-7. Irradiance forecasting results in *Biotopo*



(a) Complete



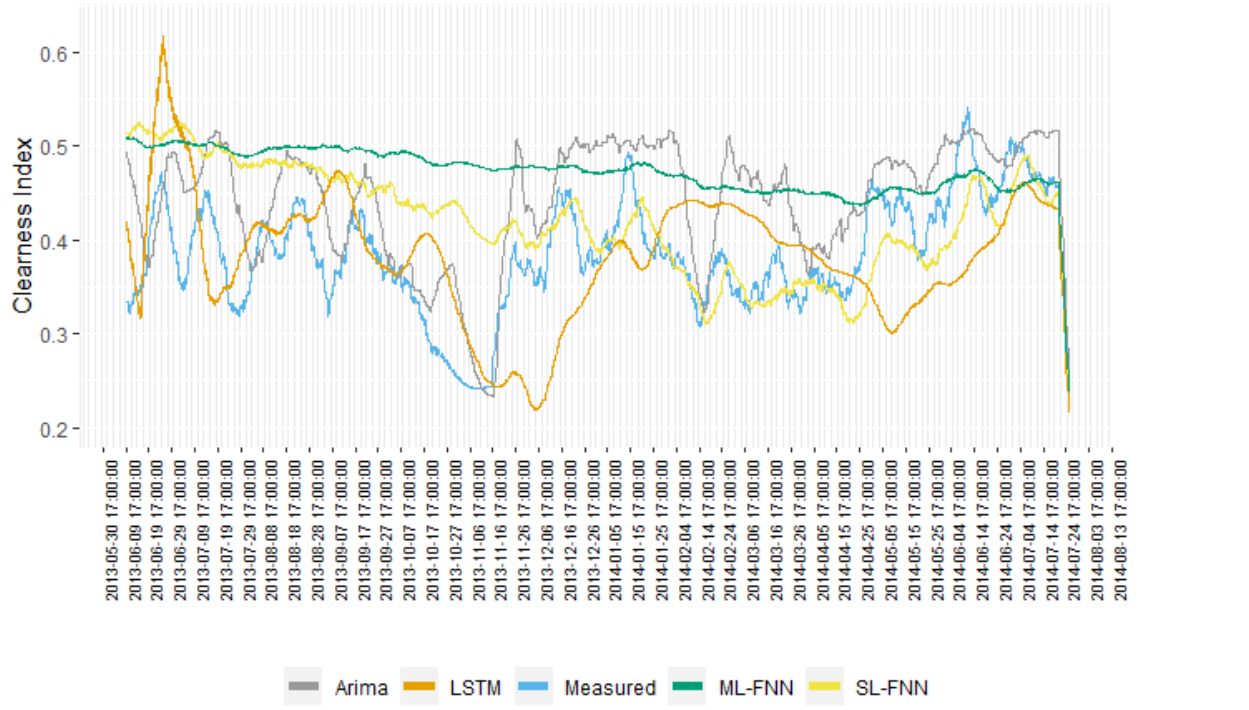
(b) Starting



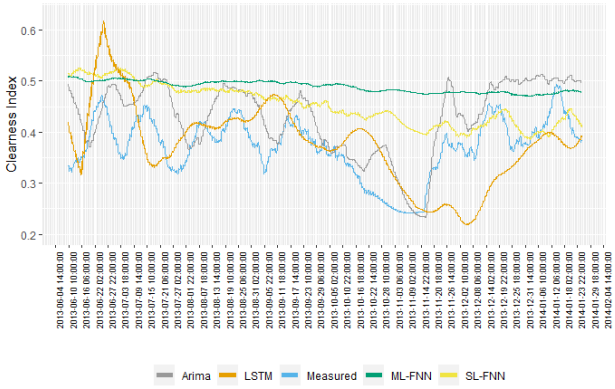
(c) Ending

the final quarter, ARIMA has the most significant error, as shown in Figure **5-7 (c)**. The RMSE shows that in the complete time series analysis SL-FNN is the best model followed by LSTM, ML-FNN and ARIMA. Considering that RMSE is sensitive to outliers, these results show that LSTM is the model that forecasts less atypical values. In the first quarter, ARIMA is the best model, in comparison with ML-FNN, LSTM and SL-FNN there is an error reduction of 41,73 %, 32,43 % and 12,68 % respectively. Nevertheless, in the final quarter, ML-FNN, SL-FNN and LSTM outperform ARIMA by 49,35 %, 47,42 %, and 45,05 %. Finally, regarding the model bias measured with MBE, LSTM is the least biased followed by ML-FNN, SL-FNN and ARIMA.

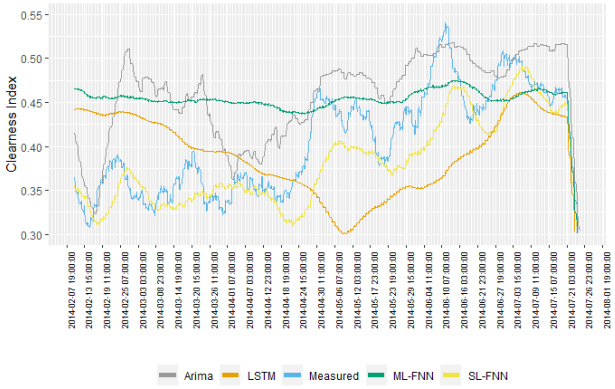
Figure 5-8. Irradiance forecasting results in *Altaquer*



(a) Complete



(b) Starting



(c) Ending

In *Altaquer*, MAE shows that LSTM is the best model, followed by ARIMA, SL-FNN, and ML-FNN. In the first, second, and third quarters, ARIMA is the best model since the neural network-based models do not have enough information to adjust their parameters. However, in the second and third quarters, LSTM reduces the MAE difference with respect to ARIMA, and in the final quarter, LSTM outperforms ARIMA by 13,28 %. RMSE and MAE behave similarly for this AWS. ARIMA is the best model in the first three quarters outperforming LSTM by 14,92 %, 5,8 % and 0,92 %, SL-FNN by 13,63 %, 46,45 % and 38,62 %, and ML-FNN by 21,87 %, 39,62 %, 30,03 %. LSTM and SL-FNN outperform ARIMA in the final quarter by 10,83 %, and 8,87 % respectively. As a result, LSTM is the best model in the complete time series, followed by ARIMA, SL-FNN and ML-FNN. MBE shows that LSTM is the model with less bias

outperforming SL-FNN, ARIMA, and ML-FNN by 40,27 %, 64,11 % and 98,07 %, respectively.

In *Granja el Mira*, MAE and RSME show that ARIMA is the best model for the complete time series. However, ARIMA is also the most biased model. Analyzing the models by quarters, ARIMA exhibits the best MAE and RMSE in the first three quarters. In the final quarter, SL-FNN is the best model. Regarding the MBE, ARIMA exhibits less bias in the first three quarters, and SL-FNN in the final quarter. However, the MBE in the complete time series shows ML-FNN as the least biased model. MBE is a measure that in long-term could compensate overestimation and underestimation periods, as shown in Table **5-1**.

In *Viento Libre*, the AWS located in the Andean zone with less cloudiness (see Figure **5-5**), SL-FNN has the lowest MAE and RMSE, followed by ARIMA, ML-FNN, and LSTM in the complete time series. In the first three-quarters, ARIMA is the best model, and in the final quarter, this model has the highest error. In the first quarter, ARIMA and SL-FNN have similar RMSE values, and in the second and third quarters, ARIMA outperforms the others. Nevertheless, in the final quarter, the neural network-based models outperform ARIMA. ARIMA is less biased during the first three quarters. However, LSTM outperforms ARIMA by 99,49 % in the last quarter.

In *Universidad de Nariño*, MAE shows that the neural network-based models outperform ARIMA in the complete time series. However, in the analysis by quarters, ARIMA outperforms neural-network-based models during the first three quarters. Regarding the RMSE, in the complete time series analysis, the neural network-based models outperform ARIMA. Nevertheless, the difference between ARIMA and LSTM is approximately 0,81 %. In the first three quarters, ARIMA has a lower error than the neural network-based models. The MBE shows LSTM as the least biased model in the complete time series. Besides, LSTM has the lowest bias in the first and fourth quarters.

Cerro Páramo is the only AWS located in the Amazonia zone. Also, it is the AWS with more cloudiness and the highest altitude, 3.577 MASL. In the complete time series analysis, MAE and RSME show that SL-FNN is the best model, followed by ML-FNN, LSTM, and ARIMA. The neural network-based models outperform ARIMA in the first quarter. This result is opposite to the ones obtained with all the AWS analyzed above, where ARIMA outperforms the neural network-based models during the first three quarters. The cloud cover amount of *Cerro Páramo* could explain these results. SL-FNN is the least biased model, followed by LSTM, ML-FNN, and ARIMA.

La Josefina is a critical case regarding the percentage of missing data (almost 80 % of the time series). MAE shows that SL-FNN better fits the mean trend of the measured data in the complete time series, followed by LSTM, ARIMA, and ML-FNN. ARIMA is the model with the highest RMSE. By using SL-FNN, LSTM and ML-FNN, RMSE reduces by 90,19 %, 88,29 % and 87,82 % respectively. MBE shows that LSTM is less biased in the complete time series. In the first quarter, ARIMA is the best model regarding MAE, and SL-FNN is the best model in the second, third, and fourth quarters. The RMSE shows that SL-FNN has the lowest error in all four quarters.

In *Botana*, the neural network-based models exhibit lower MAE and RMSE than ARIMA in the complete time series analysis. The MBE shows that ML-FNN is the most biased model, and LSTM, SL-FNN, and

ARIMA reduce the bias by 96,83 %, 77,55 %, and 77,25 %, respectively. MAE shows that ARIMA is the best model in the first three quarters, and SL-FNN is the best in the fourth quarter. ARIMA, SL-FNN, and ML-FNN have a similar RMSE with a difference of 0,06 % on average in the first quarter. In the following two quarters, ARIMA is the best model. SL-FNN is the best model in the fourth quarter. MBE shows that ML-FNN is more biased than LSTM, SL-FNN and ARIMA by 96,83 %, 77,55 % and 77,25 % respectively.

In *Paraiso* SL-FNN has the lowest MAE and RMSE values in the complete time series analysis. ARIMA exhibits the lowest MAE and RMSE error in the first three quarters. In the last quarter, SL-FNN is the best model, followed by ML-FNN, LSTM, and ARIMA. MBE presents that LSTM has the lowest bias, followed by SL-FNN, ML-FNN, and ARIMA.

In *Ospina Perez*, MAE and RSME show that ARIMA is the best model, and ML-FNN is the worst model in the complete time series. In the first quarter, SL-FNN has the lowest MAE. From the second to the fourth quarter, LSTM has the lowest MAE. The RMSE present that ARIMA is the best model in the first quarter and LSTM is the best model from the second to the fourth quarter. Additionally, LSTM is the less biased model followed by SL-FNN, ARIMA, and ML-FNN in the complete time series. In *Sandona*, MAE and RMSE show that ARIMA is the best in the complete time series and quarters analysis. The MBE indicates that LSTM is the less biased model, followed by ARIMA. Also, the percentage difference between these models is 0,8 %.

In *Guapi*, SL-FNN exhibits the lowest error in the complete time series regarding MAE and RMSE followed by ARIMA, ML-FNN, and LSTM. In the first and fourth quarters, SL-FNN shows the lowest MAE and RMSE error.ARIMA is the best option in the second and third quarters. Regarding the bias, SL-FNN shows the lowest value, followed by LSTM, ML-FNN, and ARIMA.

Universidad de Nariño and *Botana* are located closer to each other, as illustrated in Figure **5-4**, and the altitude difference is 194 MASL. However, *Botana* present more cloudiness, see Figure **5-5**. Therefore, this condition could affect the performance of the forecasting models. In *Universidad de Nariño* the MPE is 23,53 %, 19,52 %, 17,77 % and 12,84 % for ARIMA, ML-FNN, LSTM and SL-FNN respectively. In *Botana*, MPE is 41,88 %, 29,61 %, 26,34 %, and 23,21 % for ARIMA, ML-FNN, LSTM and SL-FNN respectively. These results show that cloudy conditions increase the statistical errors. ARIMA present the biggest increase with 18,35 %, followed by SL-FNN (10,37 %), ML-FNN (10,09 %), and LSTM (8,57 %).

MAE shows that ARIMA and SL-FNN follow the mean trend in the Pacific zone and have lowest errors. Additionally, MAE exhibits that ML-FNN has the lowest performance in *Biotopo* and *Altaquer*. The RMSE shows that ARIMA and SL-FNN have less variability than ML-FNN and LSTM. MBE shows that LSTM forecasts the future values with less bias. However, the MPE in the Pacific Zone shows that in *Biotopo* and Guapi, which have more than one year of measurements, LSTM is the best model, and in *Altaquer* and *Granja el Mira*, with one year of measurements on average, SL-FNN is the best option. SL-FNN is the best model in *Cerro Páramo* because it has the lowest variability, as shown by its MAE and RMSE, and has less bias. MPE shows that SL-FNN has an error of 27,29 %, LSTM of 34,96 %, ML-FNN of 39,65 %, and ARIMA of 96,96 %. Therefore, in cloudy environments, the neural network-based models outperform ARIMA, as shown by the statistical error measurements. In the AWS with more than one year of measurements

located in the Andean zone, SL-FNN is the model with less MAE error, followed by ML-FNN, LSTM, and ARIMA in the AWS located above 2.000 MASL.

In *Viento Libre*, with an altitude below 2.000 MASL, SL-FNN is the best model, followed by ARIMA, ML-FNN, and LSTM. The RMSE shows that SL-FNN and ML-FNN predict future values with less variability. Additionally, RSME shows that ARIMA has a lower performance in AWS located above 2.000 MASL. Regarding the bias, LSTM is the best model in all AWS located in the Andean zone. In the AWS with one year of measurements located in the Andean zone, ARIMA outperforms the neural network-based models in MAE and RMSE. In this case, ML-FNN has the most significant error. The short length of the time series could explain this result. ML-FNN might suffer overfitting since it exhibits poor performance when there is not enough training data.

In the Pacific zone there is not a clear model that has the best performance in all AWS. In *Altaquer*, which has 54,7 % of missing data and prevalence of partially high cloudiness, LSTM outperform the other models. SL-FNN is the best model in *Guapi*; this AWS has 76,5 % of missing data and mainly partially high cloudiness. In *Granja el Mira* and *Biotopo* there is not a model that has the best performance in all error measures. Considering RMSE and MAE, ARIMA is the best model for *Granja el Mira*. *Biotopo*, which contains 77,1 % of missing data and prevalent cloudy days, considering all errors, SL-FNN is the best model.

In the Andean zone, RMSE and MAE show that SL-FNN is the best option for all AWS with more than one year of measures, independently of the missing data percentage and cloud cover amount, while ARIMA is the best model in AWS with one year of measures on average. LSTM is the model with less bias in all AWS.For *Cerro Páramo*, the only AWS located in the Amazon zone with mainly cloudy days, SL-FNN outperforms the other model in all errors.

AWS	Model	Complete			MAE $[K_t]$				RMSE$[K_t]$				MBE$[K_t]$			
		MAE $[K_t]$	MBE $[K_t]$	RMSE $[K_t]$	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Biotopo	ARIMA	**0.093**	0.014	0.193	**0.392**	0.566	0.829	0.151	**0.325**	0.411	0.496	0.306	-0.025	-0.558	-0.812	0.044
	SL-FNN	0.098	-0.003	**0.153**	0.523	0.527	0.828	0.119	0.396	0.402	0.522	0.161	-0.105	-0.509	-0.793	0.002
	LSTM	0.116	**0.002**	0.179	0.659	**0.520**	**0.746**	0.129	0.506	**0.393**	**0.473**	0.168	**0.016**	**-0.494**	**-0.695**	**0.001**
	ML-FNN	0.125	-0.002	0.184	0.858	0.944	1.351	**0.116**	0.558	0.693	0.820	**0.155**	0.041	-0.898	-1.259	-0.001
Altaquer	ARIMA	0.182	0.058	0.228	**0.733**	**0.338**	**0.590**	0.199	**0.449**	**0.284**	**0.392**	0.240	0.236	**0.072**	**0.057**	0.042
	SL-FNN	0.186	0.037	0.232	0.860	0.684	0.978	0.177	0.527	0.530	0.639	0.219	0.413	0.557	0.723	-0.033
	LSTM	**0.175**	**-0.002**	**0.223**	0.844	0.411	0.624	**0.172**	0.534	0.314	0.398	**0.214**	**0.175**	0.265	0.330	-0.063
	ML-FNN	0.216	0.088	0.259	0.978	0.670	0.957	0.200	0.574	0.493	0.584	0.241	0.433	0.383	0.384	**0.016**
Granja el Mira	ARIMA	**0.169**	-0.013	**0.219**	**0.603**	**0.135**	**0.177**	0.184	**0.412**	**0.107**	**0.119**	0.227	**-0.048**	**-0.126**	**-0.156**	-0.029
	SL-FNN	0.177	0.011	0.228	0.924	0.926	1.267	**0.153**	0.572	0.699	0.797	**0.194**	0.205	0.922	1.258	**-0.010**
	LSTM	0.268	-0.007	0.351	1.343	0.686	1.027	0.265	0.858	0.495	0.608	0.371	-0.277	0.672	0.998	0.082
	ML-FNN	0.217	**0.003**	0.272	1.033	0.442	0.790	0.189	0.638	0.397	0.555	0.234	0.181	0.439	0.786	-0.049
Guapi	ARIMA	0.130	0.010	0.205	**0.490**	**0.255**	**0.354**	0.177	0.357	**0.205**	**0.244**	0.251	0.088	**0.162**	**0.168**	0.014
	SL-FNN	**0.106**	**0.000**	**0.157**	0.479	0.488	0.647	**0.140**	**0.331**	0.386	0.434	**0.193**	**-0.004**	-0.393	-0.456	-0.004
	LSTM	0.197	0.002	0.336	0.550	0.320	0.581	0.273	0.386	0.280	0.392	0.379	0.016	-0.265	-0.471	0.072
	ML-FNN	0.152	-0.003	0.215	0.752	0.413	0.731	0.156	0.480	0.399	0.556	0.209	0.131	-0.193	-0.482	**0.003**
Viento Libre	ARIMA	0.145	-0.014	0.206	**0.182**	**0.147**	**0.231**	0.176	**0.205**	**0.149**	**0.201**	0.226	**0.004**	**0.093**	**0.125**	-0.024
	SL-FNN	**0.117**	-0.004	**0.164**	0.273	0.284	0.485	**0.124**	0.243	0.252	0.347	**0.164**	-0.026	-0.183	-0.283	-0.002
	LSTM	0.156	**0.000**	0.247	0.513	0.424	0.633	0.152	0.652	0.329	0.412	0.196	0.041	-0.344	-0.474	**0.000**
	ML-FNN	0.153	-0.007	0.203	0.647	0.758	0.976	0.138	0.465	0.600	0.655	0.176	-0.126	-0.614	-0.689	-0.008
Universidad de Nariño	ARIMA	0.173	-0.015	0.228	**0.484**	**0.125**	**0.199**	0.195	**0.319**	**0.117**	**0.158**	0.241	-0.035	**-0.084**	**-0.116**	-0.014
	SL-FNN	**0.144**	-0.008	**0.194**	0.511	0.613	0.977	0.136	0.334	0.469	0.614	**0.178**	-0.045	-0.591	-0.934	-0.004
	LSTM	0.169	**0.000**	0.226	0.560	0.495	0.758	0.147	0.389	0.359	0.452	0.188	**-0.022**	-0.484	-0.735	**0.000**
	ML-FNN	0.158	0.000	0.209	0.519	0.311	0.521	**0.141**	0.343	0.280	0.382	0.183	-0.126	-0.220	-0.339	-0.006
Josefina	ARIMA	0.091	-0.016	1.294	**0.299**	0.160	0.218	0.163	0.293	0.134	0.159	0.218	0.007	-0.138	-0.173	-0.019
	SL-FNN	**0.074**	0.003	**0.127**	0.365	**0.083**	**0.152**	**0.124**	**0.286**	**0.099**	**0.140**	**0.168**	0.048	**-0.006**	**-0.026**	**0.002**
	LSTM	0.089	**0.000**	0.151	0.370	0.259	0.337	0.156	0.302	0.210	0.237	0.205	**-0.005**	-0.181	-0.181	-0.003

		Complete			MAE $[K_t]$				RMSE$[K_t]$				MBE$[K_t]$			
	ML-FNN	0.103	0.018	0.158	0.640	0.274	0.441	0.134	0.412	0.220	0.292	0.174	0.239	-0.008	0.091	-0.004
Botana	ARIMA	0.178	-0.003	0.228	**0.681**	**0.125**	**0.210**	0.183	0.433	**0.112**	**0.154**	0.234	**0.000**	**-0.015**	**0.009**	-0.007
	SL-FNN	**0.149**	-0.003	**0.194**	0.682	0.455	0.707	**0.136**	0.433	0.344	0.443	**0.178**	-0.017	-0.032	-0.268	-0.002
	LSTM	0.162	**0.000**	0.211	0.715	0.268	0.461	0.153	0.479	0.226	0.311	0.194	-0.009	-0.254	-0.434	**0.000**
	ML-FNN	0.153	0.012	0.196	0.684	0.179	0.251	0.142	**0.433**	0.150	0.183	0.181	0.160	-0.141	-0.173	-0.001
Paraiso	ARIMA	0.156	-0.022	1.578	**0.286**	**0.114**	**0.207**	0.189	**0.245**	**0.106**	**0.148**	0.238	-0.037	**0.014**	**0.050**	-0.010
	SL-FNN	**0.123**	-0.004	**0.172**	0.318	0.375	0.534	**0.141**	0.254	0.282	0.335	**0.184**	-0.030	-0.344	-0.472	**-0.003**
	LSTM	0.138	**-0.001**	0.192	0.329	0.426	0.626	0.167	0.261	0.311	0.377	0.216	**-0.014**	-0.400	-0.574	-0.004
	ML-FNN	0.137	-0.009	0.186	0.419	0.317	0.550	0.148	0.321	0.302	0.418	0.190	-0.176	-0.116	-0.317	-0.008
Ospina Perez	ARIMA	**0.123**	-0.033	**0.193**	**0.841**	0.353	0.530	0.180	**0.519**	0.261	0.325	0.240	-0.223	-0.162	-0.148	-0.045
	SL-FNN	0.163	-0.026	0.219	0.839	0.288	0.504	0.171	0.524	0.294	0.411	0.225	-0.317	0.046	0.163	0.062
	LSTM	0.169	**-0.015**	0.236	0.924	**0.200**	**0.381**	**0.166**	0.588	**0.207**	**0.292**	**0.217**	**-0.022**	**0.002**	**0.021**	-0.070
	ML-FNN	0.213	-0.129	0.267	0.842	0.418	0.585	0.175	0.536	0.315	0.367	0.225	-0.441	-0.314	-0.379	**-0.041**
Sandona	ARIMA	**0.132**	-0.015	**0.199**	**0.817**	**0.303**	**0.536**	**0.158**	**0.503**	**0.275**	**0.382**	**0.205**	-0.128	-0.088	-0.108	**0.005**
	SL-FNN	0.284	-0.159	0.394	0.875	0.317	0.595	0.162	0.545	0.312	0.443	0.210	-0.168	0.028	0.098	-0.027
	LSTM	0.259	**-0.013**	0.328	1.124	0.308	0.572	0.213	0.711	0.295	0.414	0.255	**-0.006**	**0.011**	0.069	0.064
	ML-FNN	0.328	-0.150	0.410	0.894	0.619	0.879	0.198	0.556	0.484	0.581	0.244	-0.087	0.434	**0.507**	0.062
Cerro Páramo	ARIMA	0.210	0.074	0.349	0.822	**0.129**	**0.188**	0.213	1.026	**0.103**	**0.128**	0.296	0.254	**-0.111**	**-0.151**	0.069
	SL-FNN	**0.136**	**0.001**	**0.186**	**0.614**	0.494	0.774	**0.131**	**0.409**	0.388	0.506	**0.185**	**0.009**	-0.485	-0.758	-0.005
	LSTM	0.177	0.002	0.277	0.677	0.362	0.560	0.260	0.454	0.266	0.338	0.430	0.018	-0.349	-0.536	**0.001**
	ML-FNN	0.156	0.003	0.214	0.756	0.512	0.801	0.162	0.469	0.391	0.507	0.249	0.149	-0.496	-0.770	-0.024

Table 5-6. Irradiance statistical errors of the one day-ahead forecasting process

5.4.3 Insolation forecasting

In this study, we also apply the forecasting models for the one-ahead day prediction of daily solar insolation. The daily solar insolation is the sum of hourly irradiances measured between 6:00 and 18:00 hours. Table **5-7** shows the errors for a daily time stamp.

In *Biotopo*, MAE shows that LSTM is the best option for daily global solar irradiance forecasting, followed by ML-FNN, ARIMA, and SL-FNN. It indicates that LSTM fits the time series' mean trend better than ML-FNN, ARIMA, and SL-FNN. As a result, using LSTM reduces the MAE error by 6,41 %, 17,73 %, and 20,23 % in comparison with ML-FNN, ARIMA and SL-FNN. Analyzing the MAE by quarters, LSTM outperforms the other models in the first and fourth quarters, and ARIMA is the best model in the second and third quarters. The RMSE shows that LSTM is the model with the lowest variability, followed by ML-FNN, ARIMA, and SL-FNN. When using LSTM, ML-FNN or ARIMA instead of SL-FNN, the RMSE decreases by 23,33 %, 16,41 % and 9,07 %, correspondingly. Furthermore, LSTM presents less bias than the other models and reduces the bias by 89,98 %, 44,88 %, and 33,47 % in comparison with SL-FNN, ARIMA, and ML-FNN.

In *Altaquer*, with one year of measurements on average, LSTM outperforms the other models. MAE shows that LSTM and ARIMA are the best options. LSTM improves the results by 58,43 %, 48,43 % and 6,75 % in comparison with SL-FNN, ML-FNN and ARIMA. Also, the analysis by quarters shows that LSTM has the best behavior. RSME shows LSTM as the best option, followed by ARIMA, ML-FNN, and SL-FNN. LSTM reduces the RMSE in 60,01 %, 53,85 %, and 8,18 % in comparison with SL-FNN, ML-FNN and ARIMA, respectively. LSTM is the less biased model followed by ARIMA, ML-FNN, and SL-FNN. MAPE shows that LSTM has an error of 32,63 %, ARIMA of 38,47 %, ML-FNN of 65,95 % and SL-FNN of 72,53 %. The large proportion of missing data and short length of the time series might explain the large errors obtained in this case.

LSTM is the best model in *Granja el Mira* considering MAE, RMSE, and MBE. MAE shows that LSTM improves the performance by 27,71 %, 13,55 %, and 11,60 % in comparison with SL-FNN, ARIMA and ML-FNN. RMSE shows that LSTM has less variability than ML-FNN, ARIMA, and SL-FNN by 12 %, 17,75 % and 30,61 %. The analysis by quarters shows that MAE and RMSE have similar results. LSTM is the less biased model followed by ARIMA, ML-FNN, and SL-FNN. MAPE shows that LSTM has an error of 26,16 %, ML-FNN of 29,10 %, ARIMA of 30,75 % and SL-FNN of 35,71 %.

In *Viento Libre*, the statistical errors show ML-FNN as the model with less MAE, RMSE, and MBE errors, followed by LSTM, SL-FNN, and ARIMA. MAE shows that ML-FNN and LSTM have a similar error with a difference of 0,16 %. Considering RMSE, ARIMA model exhibits the largest errors, and using ML-FNN, LSTM or SL-FNN reduce the error by 19,13 %, 18,94 % or 2,52 % respectively. MBE shows that the models with less bias are ML-FNN and LSTM. These models reduce the bias by 77,95 % and 99,11 % compared to SL-FNN and ARIMA. Therefore, ML-FNN and LSTM reduce the bias in comparison with SL-FNN and ARIMA. MAPE shows that on average LSTM has an error of 19,11 %, ML-FNN of 19,25 %, SL-FNN of 21,95 % and ARIMA of 23,11 %.

In *Universidad de Nariño*, ML-FNN and SL-FNN outperform the other models. MAE shows that ML-FNN has the lowest error in both, the complete time series, and quarters analysis, followed by SL-FNN. RSME shows that using ML-FNN instead of SL-FNN, ARIMA and LSTM reduces the outliers by 1,9 %, 18,11 %, and 25,4 %, respectively. MBE presents the neural network-based models as the models with less bias. MAPE shows that ARIMA has an error of 28,21 %, LSTM of 24,89 %, SL-FNN of 23,64 % and ML-FNN of 23,49 %.

In *Cerro Páramo* the MAE, RMSE and MBE shows SL-FNN as the best option for daily solar insolation. Using SL-FNN instead ML-FNN, ARIMA and LSTM improve MAE in 5,84 %, 10,55 % and 10,82 %. In the analysis by quarters, SL-FNN is the best option, followed by ARIMA in the three first quarters and by ML-FNN in the last quarter. RSME shows that the SL-FNN model reduces the variability 6,26 %, 9,96 %, and 16,69 % in comparison with ML-FNN, LSTM, and ARIMA models respectively. MBE result shows that ARIMA has more bias than ML-FNN, LSTM, and SL-FNN, increasing the error on average by 65,56 %, 88,80 %, and 92,50 %, respectively. MAPE shows that SL-FNN is the best model with an error of 38,03 % followed by ML-FNN with 39,52 %, ARIMA with 40,90 % and LSTM with 42,75 %.

In *La Josefina*, the MAE, RMSE, and MBE show that LSTM is the best models followed by ML-FNN, ARIMA, and SL-FNN. Using LSTM instead ML-FNN, ARIMA and SL-FNN reduce the MAE in 5,54 %, 17,62 % and 23,34 % respectively. Considering RMSE, LSTM, ML-FNN and ARIMA improve the results by 33,42 %, 31,43 %, and 17,30 % in comparison with SL-FNN. LSTM has the lower bias, followed by ML-FNN, ARIMA, and SL-FNN. MAPE shows that LSTM with an error of 14,39 %, ML-FNN of 15,42 %, ARIMA of 17,52 % and SL-FNN of 19,06 %. MAPE values are lower than 20 %; however, the amount of missing and imputed data of this AWS is considerable, which could affect the error measurements.

In *Botana*, the MAE and RMSE show that ML-FNN is the best model. ML-FNN reduces the MAE error, in comparison with SL-FNN, LSTM and ARIMA, in 2,02 %, 3,54 % and 19,28 % respectively. RMSE shows that ML-FNN has less variability than SL-FNN, LSTM and ARIMA reduce the error by 1,70 %, 6,78 % and 19,80 % respectively. In *Botana*, LSTM is less biased than SL-FNN, ML-FNN and ARIMA. MAPE shows that on average ML-FNN has an error of 27,23 %, SL-FNN of 27,65 %, LSTM of 28,21 % and ARIMA of 33,52 %. The neural network-based models outperform the ARIMA model. Furthermore, in *Universidad de Nariño* and *Botana*, which are AWS located close to each other, the MAE and MBE have similar values with the four forecasting models. However, RMSE shows that in cloudy environments, the neural network-based models' variability is lower than ARIMA. Additionally, MAPE shows that the neural network-based models have an error lower than 20 % on average.

In *Paraiso*, the neural network-based models outperform ARIMA. MAE shows that LSTM is the best model, followed by ML-FNN, SL-FNN, and ARIMA. Also, LSTM reduces the MAE by 3,20 %, 4 %, and 19,59 % in comparison with ML-FNN, SL-FNN and ARIMA. Considering RMSE, the LSTM model is the best, followed by ML-FNN, SL-FNN, and ARIMA. Also, LSTM improves the result by 2,22 % SL-FNN 4,63 % and 19,88 % in comparison with ML-FNN, SL-FNN and ARIMA models. Additionally, LSTM is the model with the lowest bias, followed by SL-FNN, ML-FNN, and ARIMA. Also, the neural network-based model reduces the bias on average by 82,52 % in comparison with ARIMA. MAPE shows that on average the error of LSTM is 19,84 %, ML-FNN of 20,71 %, SL-FNN of 20,86 & and ARIMA of 24,63 %. In this

case, LSTM is the only model with average error below 20 %.

In *Ospina Perez*, ML-FNN has the largest error due to the low amount of training data. Also, this AWS has a high amount of imputed values resulting in a low amount of data for statistical errors measures calculation. Considering the MAE and RMSE, SL-FNN is the best forecasting model, followed by ARIMA and LSTM. The improvement obtained with using SL-FNN, ARIMA, or LSMT instead of ML-FNN model is 59,53 % in MAE and 55,28 % in RSME on average. Regarding the MBE, the model with less bias is SL-FNN followed by LSTM. SL-FNN reduces the bias by 2,66 %, 8,83 % and 98,85 % in comparison with LSTM, ARIMA and ML-FNN. The MAPE shows that SL-FNN has the lowest error with 13,95 %, followed by ARIMA with 16,30 %, LSTM with 20,92 % and ML-FNN with 43,14 %.

In *Sandona*, ML-FNN has the largest MAE, RMSE, and MBE values. This AWS has one year of measurements on average, as *Ospina Perez*. Therefore, in the Andean zone, ML-FNN needs as input more than one year of measurements to describe the variability of this zone's global solar insolation. MAE shows that LSTM is the best option, followed by SL-FNN, ARIMA, and ML-FNN. The RMSE presents the LSTM model as the best forecasting option reducing the error by 4,61 % in comparison with SL-FNN, 12,47 % in comparison with ARIMA, and 33,02 % in comparison with ML-FNN. Regarding the bias, the SL-FNN and LSTM have the lowest bias. MAPE shows that LSTM has an average error of 19,38 %, SL-FNN of 21,52 %, ARIMA of 23,58 % and ML-FNN of 30,07 %.

In *Guapi*, the MAE, RSME, and MBE show that the neural network-based models outperform ARIMA. Contrasting these models with the ARIMA model, the MAE error is reduced by 19,73 %, 15,07 % and 14,12 % with LSTM, SL-FNN and ML-FNN respectively. RMSE shows that LSTM is the best model, followed by SL-FNN, ML-FNN, and ARIMA. LSTM reduces the RMSE value by 1,15 %, 2,09 % and 20,85 % in comparison with SL-FNN, ML-FNN and ARIMA respectively. The MBE results shows that LSTM is the less biased model, followed by SL-FNN, ML-FNN, and ARIMA. Therefore, the statistical errors show that the neural network-based models outperform ARIMA.

AWS	Model	Complete			MAE$[K_t]$				RMSE $[K_t]$				MBE $[K_t]$			
		MAE$[K_t]$	MBE$[K_t]$	RMSE $[K_t]$	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Biotopo	ARIMA	0.061	0.004	0.081	0.246	**0.109**	**0.156**	0.072	0.163	**0.085**	**0.103**	0.093	0.023	**-0.050**	**-0.037**	0.006
	SL-FNN	0.062	0.008	0.089	0.349	0.943	1.394	0.062	0.260	0.672	0.814	0.080	0.099	0.942	1.391	0.001
	LSTM	**0.050**	**0.001**	**0.067**	**0.205**	0.669	0.985	**0.058**	**0.144**	0.490	0.597	**0.074**	**0.015**	0.656	0.959	0.000
	ML-FNN	0.054	0.003	0.075	0.266	0.726	1.096	0.061	0.195	0.519	0.643	0.078	0.044	0.722	1.088	**0.000**
Altaquer	ARIMA	0.087	0.007	0.112	0.293	0.230	0.343	0.100	0.192	0.179	0.224	0.129	**0.023**	-0.082	**-0.034**	0.008
	SL-FNN	0.180	0.088	0.234	1.524	0.909	1.435	0.112	0.804	0.660	0.848	0.133	1.524	0.909	1.433	-0.065
	LSTM	**0.075**	**0.002**	**0.093**	**0.308**	**0.192**	**0.313**	0.093	**0.189**	**0.157**	**0.210**	0.112	0.031	**0.016**	0.107	**-0.003**
	ML-FNN	0.162	0.076	0.219	1.432	0.942	1.491	**0.087**	0.773	0.686	0.886	**0.105**	1.428	0.942	1.491	0.035
Granja el Mira	ARIMA	0.079	0.007	0.104	0.297	**0.079**	**0.162**	0.082	0.189	**0.087**	**0.127**	0.106	**0.005**	**0.027**	**0.067**	0.005
	SL-FNN	0.092	-0.018	0.119	0.542	0.612	0.909	0.070	0.330	0.450	0.552	0.088	-0.441	-0.612	-0.905	-0.007
	LSTM	**0.066**	**0.002**	**0.082**	**0.282**	0.195	0.313	0.065	**0.167**	0.151	0.199	0.081	0.043	0.184	0.287	**0.002**
	ML-FNN	0.077	0.010	0.097	0.380	0.375	0.591	**0.061**	0.239	0.278	0.359	**0.077**	0.182	0.375	0.585	0.019
Viento Libre	ARIMA	0.075	0.013	0.100	0.233	0.803	0.848	0.080	0.160	0.759	0.761	0.105	0.037	-0.774	-0.790	0.013
	SL-FNN	0.071	0.010	0.097	0.325	1.162	1.625	**0.064**	0.247	0.846	0.971	**0.081**	0.131	1.162	1.623	0.002
	LSTM	0.062	0.000	0.081	0.187	**0.127**	**0.178**	0.067	0.125	**0.104**	**0.126**	0.084	-0.004	**0.064**	**0.052**	**0.000**
	ML-FNN	**0.062**	**0.000**	**0.081**	**0.185**	0.376	0.516	0.068	**0.124**	0.283	0.323	0.086	**0.004**	0.366	0.497	0.002
Universidad de Nariño	ARIMA	0.080	0.012	0.105	0.326	0.337	0.493	0.083	0.209	0.253	0.309	0.104	0.053	0.242	0.303	0.011
	SL-FNN	0.068	-0.001	0.087	0.310	0.234	0.329	0.063	0.194	0.176	0.206	0.079	-0.010	-0.175	-0.212	**0.000**
	LSTM	0.072	**0.000**	0.114	0.366	0.462	0.738	0.065	0.349	0.403	0.539	0.080	**0.002**	0.343	0.503	0.000
	ML-FNN	**0.068**	0.000	**0.085**	**0.273**	**0.128**	**0.231**	**0.062**	**0.170**	**0.125**	**0.175**	**0.078**	0.002	**0.066**	**0.107**	0.000
Cerro Páramo	ARIMA	0.074	0.009	0.105	0.297	0.417	0.509	0.069	0.203	0.340	0.356	0.094	0.034	0.319	**0.315**	0.008
	SL-FNN	**0.066**	**-0.001**	**0.087**	**0.278**	**0.200**	**0.343**	**0.059**	**0.186**	**0.176**	**0.242**	**0.079**	-0.014	**-0.192**	-0.328	0.000
	LSTM	0.074	0.001	0.098	0.312	0.824	1.186	0.066	0.210	0.608	0.725	0.087	**0.013**	0.807	1.152	**0.000**
	ML-FNN	0.071	0.003	0.094	0.343	0.761	1.142	0.061	0.230	0.548	0.675	0.080	0.024	0.755	1.130	0.002
Josefina	ARIMA	0.052	0.005	0.074	0.200	**0.133**	**0.191**	0.076	0.142	**0.110**	**0.135**	0.101	0.015	**0.082**	**0.088**	0.011
	SL-FNN	0.056	0.011	0.090	0.352	0.949	1.425	0.062	0.289	0.677	0.832	0.081	0.150	0.948	1.423	0.002
	LSTM	**0.043**	**0.000**	**0.060**	**0.168**	0.444	0.652	**0.060**	**0.120**	0.324	0.394	0.078	**0.010**	0.432	0.628	**0.000**

		Complete			MAE$[K_t]$				RMSE $[K_t]$				MBE $[K_t]$			
	ML-FNN	0.046	0.001	0.062	0.177	0.263	0.385	0.060	0.125	0.195	0.236	**0.078**	0.012	0.258	0.375	0.000
Botana	ARIMA	0.089	0.012	0.114	0.364	**0.212**	**0.341**	0.085	0.229	**0.180**	**0.240**	0.108	0.052	**0.152**	**0.220**	0.012
	SL-FNN	0.074	-0.001	0.093	0.334	0.259	0.380	0.065	0.207	0.206	0.256	0.082	-0.014	-0.255	-0.371	0.000
	LSTM	0.075	**0.000**	0.099	0.350	0.966	1.415	0.066	0.241	0.695	0.835	0.084	**-0.002**	-0.866	-1.215	**0.000**
	ML-FNN	**0.072**	0.002	**0.091**	**0.328**	0.471	0.718	**0.065**	**0.205**	0.343	0.431	**0.082**	0.027	0.468	0.713	0.001
Paraiso	ARIMA	0.072	0.009	0.096	0.222	**0.322**	**0.382**	0.084	0.165	**0.273**	**0.282**	0.106	0.032	**-0.232**	**-0.199**	0.012
	SL-FNN	0.061	0.002	0.081	0.237	0.525	0.808	**0.067**	0.175	0.380	0.480	0.085	0.026	0.522	0.803	-0.001
	LSTM	**0.058**	**0.001**	**0.077**	**0.195**	0.562	0.830	0.067	**0.145**	0.410	0.501	**0.084**	**0.012**	0.547	0.801	-0.001
	ML-FNN	0.060	0.003	0.079	**0.231**	0.414	0.625	0.069	0.162	0.296	0.367	0.087	0.044	0.412	0.621	**0.000**
Ospina Perez	ARIMA	0.051	0.006	0.085	0.334	0.255	0.417	0.081	0.222	0.222	0.301	0.107	0.037	-0.131	**-0.162**	0.008
	SL-FNN	**0.045**	**-0.001**	**0.069**	**0.315**	**0.147**	**0.254**	**0.063**	**0.198**	**0.128**	**0.176**	**0.080**	-0.024	**0.106**	0.178	-0.009
	LSTM	0.067	-0.002	0.099	0.500	0.336	0.505	0.087	0.302	0.264	0.334	0.108	**-0.024**	-0.299	-0.420	**-0.003**
	ML-FNN	0.134	0.065	0.188	1.269	0.945	1.414	0.067	0.688	0.682	0.837	0.086	1.269	0.945	1.413	0.034
Sandona	ARIMA	0.062	0.007	0.084	0.305	0.577	0.653	0.073	0.206	0.505	0.511	0.096	0.024	-0.495	-0.488	0.012
	SLP	0.058	-0.004	0.075	**0.264**	**0.203**	**0.255**	0.056	**0.168**	**0.167**	**0.180**	0.078	**-0.023**	**-0.124**	**-0.095**	-0.020
	LSTM	**0.052**	**0.005**	**0.070**	0.302	0.239	0.409	**0.053**	0.196	0.198	0.271	0.073	0.085	0.232	0.392	**-0.005**
	ML-FNN	0.079	0.021	0.105	0.556	0.373	0.612	0.055	0.337	0.282	0.374	**0.071**	0.473	0.373	0.609	-0.007
Guapi	ARIMA	0.062	0.005	0.086	0.261	0.157	0.246	0.079	0.173	0.130	0.170	0.104	0.022	**-0.023**	**0.023**	0.005
	SL-FNN	0.053	0.000	0.069	**0.224**	**0.118**	**0.183**	0.059	**0.145**	**0.094**	**0.122**	0.077	**0.002**	-0.069	-0.085	**0.000**
	LSTM	**0.050**	**0.000**	**0.068**	0.229	0.230	0.326	**0.058**	0.151	0.174	0.207	**0.076**	0.004	-0.197	-0.259	0.000
	ML-FNN	0.054	0.001	0.070	0.231	0.379	0.580	0.059	0.151	0.278	0.350	0.077	0.006	0.371	0.566	0.001

Table 5-7. Insolation statistical errors of the one day-ahead forecasting process

The AWS used to compare the forecasting models are located in an Intertropical Convergence Zone, which implies there is high cloud cover amount and precipitation. This situation could affect the models' accuracy because it increases the variability of the global solar irradiance and the forecasting error. Additionally, the significant amount of missing data might also affect the forecasting accuracy.

In the Pacific zone, *Biotopo* and *Altaquer* (located in the Pacific foothill) face high humidity and a rainy environment. Consequently, MPE and MAPE show that the forecasting models have a large error in the irradiance and insolation. In this case, only LSTM has an error close to 30 %. In the other AWS of the Pacific zone, SL-FNN is the best option in the irradiance forecasting for the AWS with more than one year of measurements. LSTM is the best option for all the AWS in a insolation forecasting. As future work, changes in the memory of LSTM could be introduced to analyze the capability of this to follow the clearness index variability of the global solar irradiance.

In the Andean zone, SL-FNN is the best model considering MAE, RMSE, and MPE in five of the seven AWS in the irradiance forecasting. However, LSTM is the less biased model in all cases. In the insolation forecasting, the neural network-based models outperform ARIMA. In *Cerro Páramo* that is located in the Amazon zone and is the most cloudy AWS, SL-FNN is the best option in the irradiance and insolation forecasting. The neural network-based models outperform ARIMA in the irradiance forecasting. In conclusion, SL-FNN is the best forecasting model for a cloudy environment for global solar irradiance.

5.5 Conclusions

In this work, we implemented four forecasting models: ARIMA, SL-FNN, ML-FNN, and LSTM for forecasting global solar irradiance with one-day ahead horizons in an hourly timestamp and global solar insolation with one-day ahead horizons in an daily timestamp. One of the challenges we tackled was the handling of missing data for which we implemented an imputation process. The training process required the imputation of more than 50 % of the time series values. *Altaquer* that provided the most complete time series, required the imputation of 54,7 % of the data, and *La Josefina* was the most critical case and required the imputation of 84,2 % of the data. The statistical errors of the solar irradiance imputation showed that the resource was underestimated. The RMSE and MAE result showed that the Andean zone the error is bigger than in the Pacific zone. For the solar irradiance imputation, we use the empirical temperature-based models such as Hargreaves and Samani and Logistic, the first one for the Pacific zone and the second one for the Andean and Amazon zones. To reduce the statistical errors, it would be worth exploring alternative imputation techniques for solar irradiance and insolation data.

In the hourly forecasting, in almost all cases, the SL-FNN, ML-FNN and LSTM outperformed ARIMA. In Pacific zone, LSTM was the best model when training with the shortest time series (*Altaquer*), and SLP when training with longer ones (*Biotopo* and *Guapi*). Also, the neural network-based models showed a better performance in cloudy conditions in this zone (*Biotopo* and *Altaquer*). This result indicates that the solar irradiance variability induced by the cloud motion might be better modeled by neural network-

based models.

In the Andean zone, SL-FNN presents better results in the MAE and RMSE error measurement when the time series is longer i.e. in *Viento Libre*, *Botana*, *La Josefina*, *Paraiso*, and *Universidad de Nariño* cases. When the time series is shorter (*Ospina Perez* and *Sandona*), ARIMA is the model with less RMSE and MAE error. Therefore, for the Andean zone one year of measurements on average was not enough for training the neural network-based models. As a result, in such cases ARIMA is a better option. The MBE error measurement shows LSTM as the model with less bias in all AWS of this zone. In the Amazonian zone, the *Cerro Páramo* AWS, SL-FNN is the best model in all error measurements. Overall, the SL-FNN model outperforms the other models in global solar irradiance in a one-day ahead horizon with an hourly frequency. As a conclusion, SL-FNN is the best option for sites with high altitude and cloudiness.

The statistical error measurements of the global solar insolation forecasting in a daily timestamp show that LSTM is the best option in the Pacific zone regardless the time series length. In the Andean zone, the neural network-based model exhibit lower statistical errors. However, the model performance is dependent on the amount of training input data. For example in the shorter time series LSTM and SL-FNN are the better models, while ML-FNN is the model with lower performance. The results presented in this study do not exhibit a clear pattern that indicates that there is one forecasting technique that outperforms the others overall in the Andean zone. As observed in other studies (Dannecker, 2015), we found that the performance of each forecasting model depends on the specific task.

6 Thesis Conclusions

This thesis proposed a methodology for the selection of projects and its implementation in a community. Additionally, the prioritization methodology evaluates relevant areas and the cost of implementing solar-energy-based. The criteria chose for community evaluation consider the SDGI and the HDI and the NPV results were the criteria for project evaluation. Also, the prioritization presents a combination of those indexes, which were the criteria used in the AHP methodology to give an execution order. Furthermore, the guideline for community work emphasized the integration of the community in the design and operation of the implemented solution, providing the project with a higher probability of success. In this case, the authors considered the characteristics of Nariño and identified relevant areas for the community evaluation by municipalities. The dimensions chosen were education, health, quality of life, economy, and security.

The alternatives evaluated with this methodology were four projects, along with the communities involved. The results show that in four of six prioritization objectives, the electrification of schools is the priority, and in two objectives, the priority was the electrification of an indigenous reservation. Given the above results, the authors decided to carry out a more detailed analysis of the project of electrification of the schools. The analysis showed that if the IRR is a decision variable, it is better to implement larger projects; in this case, more than ten. On the other hand, the results showed a correlation between education and the quality of life because O3 and O6 have six projects in the same order of execution; therefore, the accomplishment of education projects improves the quality of life and vice-versa. In this sense, it would be convenient to undertake more studies that confirm this finding and thus unify efforts in the public policies in these sectors to strengthen the results

The researchers expected that, with the implementation of this prioritization methodology, it would be possible to identify the optimal execution order of social projects according to a given objective. Additionally, with the guideline, the objective is to make the community an active part of the solution, to improve the sustainability of the project and avoid or reduce the dependence on an external entity, and thus empower the community and promote their social, economic, education and cultural development, among other aspects. It is also essential to highlight the primary role of the local regulation board in the success of the project. Therefore, future research should analyze aspects such as regulation, election mechanisms of the board members, among others, deeply.

The validation levels of global solar irradiance data have a strong influence on the results of the empirical variables. Considering all the recorded information, just 60,89 % overcome the mandatory validation steps proposed. From that value, 95,81 %, which corresponds to the number of days with at least six

values measured, was the information used in the empirical models calibration. However, this percentage represent on average 33,90 % of the total information recorded in the AWS. Besides, the days with complete information just reached 1,26 %. This result indicates the quality of the time series and the necessity to improve and increase the maintenance and calibration procedures.

In the Nariño state, the performance of the AWS is a determining factor, considering the predominance of partially high cloudiness days that represent 64,7 % of the days. In other words, in Nariño, there is a high cloud interaction that difficult the solar insolation estimation; therefore, the need for reliable measures increases.

Regarding the temperature measures, from the amount of data that overcome the hourly validation steps, a 92,78 % were taken for the empirical calibration and imputation. Besides, the number of temperature daytime measurements present peaks in eleven and twelve daily values; this means that most of the days taken for modelling and filling the database by the imputation process, have the 88,46 % of the total information on average.

The proposed model showed a linear relationship between the empirical coefficients against the altitude of the study site. The empirical coefficients were classified between those above and below 2.500 MASL The lineal adjust for cases below 2.500 MSSL shows an R^2 of 0.5995 and 0.5262 for a and b respectively. The lineal adjustment for cases above 2.500 MASL presents an R^2 of 0,8182 and 0,6069 for a and b respectively. This result is consistent with the behavior of the temperature in tropical zones and the global solar irradiance in high altitudes.

When the RMSE, SD, MAE, U_{95} and MAPE were considered, the proposed model had better performance in five out of the eight evaluated cases. These cases are in the Andean and Amazon zone, with altitudes above 2.500 MASL As a result, the proposed model is used to input the information in the Andean and Amazon AWS. For the AWS located in the Pacific zones, the HS model was the best option, followed by the proposed model in this research.

In the Pacific zone, HS is the applied model to estimate solar insolation. However, the statistical error results show that the resource is underestimated by 442,78 $[Wh/m^2day]$ and 333,11 $[Wh/m^2day]$ for Granja el Mira and Altaquer. Consequently, to improve this zone's results, it is needed to increase the weather station network's density.

The proposed model was the model used to estimate the solar insolation in the Andean zone. The comparison between measured and estimated values showed an overestimation. However, regarding the HS model, the estimation is reduced 51,81 % and 18,55 % for Ospina Perez and Sandona, respectively. The proposed model has better performance than HS model in altitudes above 2.500 MASL.

Comparing IDW to the Kriging techniques, OK presented an improvement in the statistical errors as follow: RMSE 11,71 %, MAE 3,98 %, MBE 100,92 %, SD 11,40 % and MPE 46,58 %. Although SK showed better results in MAE, SD, and RMSE than OK, the bias with OK is lower. Therefore, OK was the technique selected for mapping the solar insolation potential. UK was the technique that presented lower

improvements against the IDW, even in MAE, which reduced the 6,78 % of the accuracy.

From the maps, it is possible to conclude that the obtained estimation is lower than the presented in the IDEAM solar radiation maps over the whole Nariño territory. It is also possible to determine the existence of a zone with a high potential for electricity generation with solar PV power plants.

The errors measurements of the forecasting models, present that the neural network-based models outperform ARIMA model. In the AWS which have longer time series, the SLP was the best model followed by MLP. In the AWS with shorter the LSTM was the best model. Also, in the Pacific zone in cloudy environments the LSTM model outperforms the others. It is a valuable result due to the variability in the solar irradiance induced by the cloud motion is better described with this models. Therefore, in cloudy environment, the LSTM is the recommended model to forecasting global solar irradiance. In the Andean zone, the neural network-based forecasting models present better results in the MAE and RMSE error measurement when the time series is longer i.e. in *Viento Libre*, *Botana*, *La Josefina*, and *Paraiso* cases. The MBE error measurement present the SLP as the model with less bias. Unlike to the Pacific zone, in this zone the LSTM was the model with more bias, regardless the cloudy level. In the *Ospina Perez* and *Sandona*, which are shorter time series, SLP is better in cloudy environments. In the Amazonian zone, the *Cerro Páramo* AWS shows that SLP is the best model in all error measurements. Moreover, the time series plot allows conclude that ARIMA model predict the future values following a mean trend, more than describing the variability of the time series. In overall, the SLP model outperforms the other model in global solar irradiance in a one-day ahead horizon with a hourly frequency.

The global solar insolation forecasting model's statistical error measurements in a daily timestamp show that in the AWS located in the Pacific zone with more than one year of measures, the mean error is similar with all the forecasting models. However, the ARIMA model has a bigger variance. While in the AWS located in the Pacific zone with one year of measures on average, the SLP model presents the biggest errors due to the low training input data. Overall in the AWS located in the Andean zone with more than one year of measurements, the neural network-based models have better performance than the ARIMA model. However, in the AWS with one year of measures on average located in the Andean zone, the MLP model presents the biggest errors. Therefore, in the Andean zone, the neural network-based model presents lower statistical errors. However, the model performance is dependent on the amount of training input data.

Recommendations

It is convenient to include more dimensions to understand in a better way the conditions of the population and include cultural aspects of the communities. For the project evaluation, the authors consider the NPV, giving more importance to those with lower costs. However, this lone indicator does not allow for the recognition of social benefits that may arise from the project. Hence, in future work, the use of additional quantifiers that measure the economic benefit that arose from social projects could motivate

the attention of potential investors.

Due to the quality control results, it is important to establish a plan to do calibration and maintenance procedures regularly and to follow high-quality and widely accepted standards. It was also notorious, the need for installing more AWS to increase the sampling points.

The proposed empirical model had a good performance in this tropical and mountainous environment. However, it is necessary to analyze more information coming from other places with the same characteristics. To achieve this, the prime requirement is to increase the number of AWS and the quality of the time series in the tropical and mountainous environments.

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. (2008a). 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

Akinoglu, B. (2008b). 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. Mougeot, 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. ⊠., & 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 montly 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 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 minimim 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). Micro-grids 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=XuSTldnWP5MKAP8J3YUHlBbfFCM

CORPONARIÑO. (2001). *Plan De Gestion 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 Estadísitica - 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 Estadísitica - DANE. (2009). *Metodología Déficit de Vivienda* (Tech. Rep.). Bogotá: Departamento Adminstrativo Nacional de Estadística.

Departamento Administrativo Nacional de Estadísitica - 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 Estadísitica - 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 Estadísitica - 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 Estadísitica - 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 Estadísitica - 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 Estadísitica - 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 Estadísitica - 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 Estadísitica - 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

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*. Bogotà: 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/WEO2017SpecialReport_EnergyAccessOutlook.pdf%0Ahttp://dx.doi.org/10.1016/j.enpol.2016. doi: 10.1787/9789264285569-en

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 Latioamericano y del Caribe de Planificación Económica y Social - ILPES. Retrieved from 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 di ff use 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

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/ColombiaCultural/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, ☒. 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

National Oceanic and Atmospheric Administration. (2020). *Climate Prediction Center*. Retrieved from https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php

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

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 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

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

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

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

Rodriguez, 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. Servces Sciences*, *1*(1), 83. Retrieved from http://www.rafikulislam.com/uploads/resourses/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)

Sandia National Laboratories. (2021). *Irradiance & Insolation*. Retrieved from https://pvpmc.sandia.gov/modeling-steps/1-weather-design-inputs/irradiance-and-insolation-2/

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=9tv0taI8l6YC 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

Ubilla, K., Jiménez-Estévez, G. A., Hernádez, 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/CoberturadelSistemaIntercontecadoNacional/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

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). *Geoestadìsitica Aplicada*. Instituto de Geofìsica UNAM; Instituto de Geofìsica 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, ⊠., 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