applied sciences

MDPI

Article

A New Empirical Approach for Estimating Solar Insolation Using Air Temperature in Tropical and Mountainous Environments

Laura Sofía Hoyos-Gomez [1,2,*] and Belizza Janet Ruiz-Mendoza [1]

[1] Grupo de Investigación en Potencia, Energía y Mercados (GIPEM), Departamento de Eléctrica, Electrónica y Computación, Facultad de Ingeniería y Arquitectura, Campus La Nubia, Edificio S, Universidad Nacional de Colombia-Sede Manizales, Salón S 203, km 7 vía al Aeropuerto, Manizales 055440, Colombia; bjruizm@unal.edu.co

[2] Electrónica y Biomédi-ca-FIMEB, Facultad de Ingeniería Mecánica, Universidad Antonio Nariño-Sede Bogotá, Bogotá 111511, Colombia

* Correspondence: lshoyosg@unal.edu.co

Abstract: Solar irradiance is an available resource that could support electrification in regions that are low on socio-economic indices. Therefore, it is increasingly important to understand the behavior of solar irradiance. and data on solar irradiance. Some locations, especially those with a low socio-economic population, do not have measured solar irradiance data, and if such information exists, it is not complete. There are different approaches for estimating solar irradiance, from learning models to empirical models. The latter has the advantage of low computational costs, allowing its wide use. Researchers estimate solar energy resources using information from other meteorological variables, such as temperature. However, there is no broad analysis of these techniques in tropical and mountainous environments. Therefore, in order to address this gap, our research analyzes the performance of three well-known empirical temperature-based models—Hargreaves and Samani, Bristol and Campbell, and Okundamiya and Nzeako—and proposes a new one for tropical and mountainous environments. The new empirical technique models daily solar irradiance in some areas better than the other three models. Statistical error comparison allows us to select the best model for each location and determines the data imputation model. Hargreaves and Samani's model had better results in the Pacific zone with an average RMSE of 936, 195 $\left[\mathrm{Wh/m^2\ day}\right]$, SD of 36, 01%, MAE of 748, 435 $\left[\mathrm{Wh/m^2\ day}\right]$, and U_{95} of 1.836, 325 $\left[\mathrm{Wh/m^2\ day}\right]$. The new proposed model showed better results in the Andean and Amazon zones with an average RMSE of 1.032, 99 $\left[\mathrm{Wh/m^2\ day}\right]$, SD of 34, 455 $\left[\mathrm{Wh/m^2\ day}\right]$, MAE of 825, 46 $\left[\mathrm{Wh/m^2\ day}\right]$, and U_{95} of 2.025, 84 $\left[\mathrm{Wh/m^2\ day}\right]$. Another result was the linear relationship between the new empirical model constants and the altitude of 2500 MASL (mean above sea level).

Keywords: temperature-based models; data imputation; tropical environment; logistic regression; Hargreaves and Samani; Bristow and Campbell

Citation: Hoyos-Gomez, L.S.; Ruiz-Mendoza, B.J. A New Empirical Approach for Estimating Solar Insolation Using Air Temperature in Tropical and Mountainous Environments. *Appl. Sci.* **2021**, *11*, 11491. https://doi.org/10.3390/app112311491

Academic Editors: Harry D. Kambezidis, Jürgen Reichardt and Basil Psiloglou

Received: 12 July 2021
Accepted: 10 November 2021
Published: 3 December 2021

1. Introduction

Solar energy is a promising renewable energy source for supplying the growing energy demand. It has taken decades for solar energy to become economically feasible in developing countries. Today, the associated cost of harnessing this energy resource is competitive in certain situations, such as rural and isolated electrification. In isolated and rural areas, low population density discourages investments for expanding electricity transport systems due to uncertainty in the profits. The state could face such limitations in ensuring access to electricity services for rural and isolated populations through non-conventional energy sources, such as solar energy. In developing countries, solar irradiance data are often unavailable because of the scarcity of weather stations that measure this variable and the insufficient or incomplete calibration and maintenance of metering equipment [1].

In some cases, even when open-access databases are available, information gathering is extensive [2]. Therefore, in recent years, studies analyzing the correlation and methods for estimating solar insolation have increased. Empirical models incorporate relevant variables, such as temperature, sunshine, and relative humidity, to estimate solar insolation, as academic literature shows. The choice of the weather variable to estimate solar irradiance depends on available information, which is temperature in this case. We selected temperature because variable data were available in the national database, and AWS measured temperature and solar irradiance simultaneously in the analyzed locations.

1.1. Background

From empirical models to artificial intelligence, researchers develop models to estimate solar insolation for different time frames. Simplicity, acceptance, adaptability, and low computational cost are the advantages of empirical models [3]. Empirical models are usually based on astronomical, geometrical, physical, and meteorological factors [4]. The meteorological factors include the cloudiness, temperature, and sunshine that describe the condition of the sky. Of these, cloudiness is the most crucial factor for determining solar irradiance behavior. The databases record sunshine duration and temperature [5]; consequently, these variables are most widely used to estimate solar insolation [2]. The implementation of the empirical model depends on data availability and consistency [6]. Given this, the authors of [7] recommend revising the study site's data simultaneity and reliability. In this case, all the weather stations recorded temperature data. Therefore, temperature-based empirical models are a convenient option for estimating global solar insolation [8].

In 1982, Hargreaves and Samani presented the first temperature-based model for estimating solar insolation based on daily temperature differences [9]. In 1984, Bristow and Campbell proposed a temperature-based empirical model that inputs the difference between daily maximum and minimum temperatures [10]. In 2011, Cheng 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 functions outperform empirical models [11]. In 2014, Li et al. presented a temperature-based model for China based on the Hargreaves and Samani model [12]. Quansah et al. evaluated temperature-based and sunshine-based empirical models in Ghana [13], and Dos Santos et al. assessed nine temperature-based models for Brazil [14]. In 2017, Rivero et al. validated Hargreaves and Samani's model for Mexico [15], while Jamil and Akhtar compared empirical models for India's subtropical and humid environments [16]. Although several studies analyzed the empirical models' behavior in different places globally, there is no similar amount of research analyzing the efficacy of temperature-based models for tropical and mountainous environments.

1.2. Contribution

To this end, the purpose of this research study was to assess three empirical temperature-based models—Hargreaves and Samani (HS), Bristow and Campbell (BC), and Okundamiya and Nzeako (ON)—for their capacity to estimate solar insolation in a tropical and mountainous environment. Statistical validation determines and compares the performance of the three models. Additionally, we proposed a new method based on the logistic regression model in order to estimate daily solar insolation according to daily temperature differences. This new approach offers an option to estimate daily solar insolations based on the temperature difference that could improve the estimation of this resource in sites with different latitudes with low computational costs. The IDEAM's (Institute of Hydrology, Meteorology and Environmental Studies, abbreviated *IDEAM* according to its Spanish translation) Automatic Weather Stations (AWSs) located in the State of Nariño, Colombia, measured the information used as input. The three environmental zones of the area, Pacific, Andean, and Amazonian, allowed evaluating 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 reliability of the results. R-CRAN was the software used to carry out the computational process needed for the research.

1.3. Paper Structure

The article is organized into the following sections: Section 2 describes the materials and methods used; Section 2.1 describes the characteristics of the dataset; Sections 2.2 and 2.3 outline the quality control procedures applied to both solar irradiance and temperature data; Section 2.4 presents the empirical temperature-based models; Section 2.3.4 introduces the author's proposed model; Section 2.4 assesses the statistical validation of empirical models; Section 3 presents the results and discussion; and finally, Section 4 presents the conclusions.

2. Materials and Methods

2.1. Site and Dataset

The weather stations located in Nariño, Colombia (00°31′08″ N and 02°41′08″ N latitude; 76°51′19″ W and 79°01′34″ W longitude), measured solar irradiance and temperatures used as input for the empirical models. The proximity to the Equator and the Intertropical Convergence Zone influences the weather of this place, causing unimodal or bimodal rainy seasons and increasing cloud cover due to high humidity in low altitudes. Additionally, the western mountain range is a barrier for the moist air masses from the Pacific Ocean; thus, the Pacific foothills become more humid and tend to maintain this humidity [17]. Hence, it is appropriate to divide the weather and orographic conditions of Nariño into three environmental regions: the Pacific, the Andean, and the Amazon [18].

The Pacific zone covers 52% of the total territory of Nariño, and this zone includes the *Mira-Mataje* binational watershed and a mangrove forest located in the *Sanquianga* natural reserve. Additionally, the region has two climatic zones: the Pacific flatlands and the Pacific foothills. The Pacific flatlands have a humidity level higher than 80%, an average temperature higher than 26 °C, and annual precipitation of between 3000 and 5000 mm/year. The Pacific foothills have high humidity, an average temperature between 18 °C and 24 °C, and annual precipitation between 4000 and 6000 mm/year. However, between the *Junín* and *Barbacoas* localities, the annual precipitation is approximately 9000 mm/year [17–19]. The Andean zone constitutes 40% of the total area of Nariño, and in this zone, the Andes split into two ranges: western and central [17,20]. The western mountain range exhibits bimodal precipitation behavior between 800 and 2200 mm/year, with peaks in April–May and October–November. In the central mountain range, the temperatures vary between 16 °C to 24 °C, and precipitation averages between 1000 to 1800 mm/year. In the north zone toward *Patia*, annual precipitation is less than 1000 mm/year, with temperatures typically above 24 °C [19]. The Amazon zone covers 8% of the Nariño territory and has two climate zones: mountainous and flatland. The mountainous zone is located between the *Patía* and *Putumayo* rivers with temperatures averaging between 6 °C and 11 °C and annual precipitation of about 2000 mm/year. The flatland zone has tropical weather influenced by the cloudy jungle and precipitation between 500 to 1500 mm/year [19].

There are 9 AWSs measuring solar irradiance, 8 located in Nariño and 1 positioned in the neighboring state of Cauca, and 16 Conventional Weather Stations (CWS) measuring temperature (see Tables 1 and 2), located as shown in Figure 1. In the Pacific region, there are three AWSs. The altitude of these AWS ranges between 16 and 512 MASL. In the Andean region, there are five AWSs located at altitudes between 1005 and 3120 MASL. Finally, in the Amazon region, there is one AWS at an altitude of 3577 MASL. The maximum altitude difference between all AWSs is 3561 MASL. This difference allows evaluating solar irradiance behavior in a wide range of altitudes in the three diverse environmental regions. We reviewed a study on the same zone that used Landsat 7 (1999–2015), MODIS (2005–2015) and VAISALA 3TIER (2000–2015) satellite data; however, when comparing the satellite and ground data, we found a significant difference between them. The satellite data resulted in a maximum solar irradiance potential of approximately 245 W/m^2 day with MODIS and

Landsat 7 database [21]. In contrast, the ground measurements show a maximum potential of approximately 3800 W/m^2 day. Furthermore, the results of the national Solar Atlas are closer to our results than satellite ones.

Table 1. Automatic weather stations in Nariño.

Name	Latitude	Longitude	Altitude	Period	Region
Biotopo	1.41	−78.28	512	2005–2017	Pacific
Altaquer *	1.56	−78.09	101	2013–2014	Pacific
Granja el Mira	1.55	−78.70	16	2016–2017	Pacific
Cerro Páramo	0.84	−77.39	3577	2005–2017	Amazon
La Josefina	0.93	−77.48	2449	2005–2017	Andean
Viento Libre	1.62	−77.34	1005	2005–2017	Andean
Universidad de Nariño	1.23	−77.28	2626	2005–2017	Andean
Botana	1.16	−77.27	2820	2005–2017	Andean
El Paraiso	1.07	−77.63	3120	2005–2017	Andean
Guapi **	2.57	−77.89	42	2005–2017	Pacific

* This AWS was not used in this study because the measurement period is short for this study. ** This AWS is in the neighboring state of Cauca.

Table 2. Conventional weather stations in Nariño.

Name	Latitude	Longitude	Altitude	Period	Region
CCCP del Pacífico	1.82	−78.73	1	2005–2017	Pacific
Altaquer	1.56	−78.09	1010	2005–2013	Pacific
Granja el Mira	1.55	−78.69	16	2005–2017	Pacific
Obonuco	1.19	−77.30	2710	2005–2015	Andean
Apto. Antonio Nariño	1.39	−77.29	1796	2005–2017	Andean
San Bernardo	1.53	−77.03	2190	2005–2017	Andean
Taminango	1.55	−77.27	1875	2005–2017	Andean
Común el automática	0.93	−77.63	3141	2007–2017	Andean
Apto. San Luis	0.86	−77.67	2961	2005–2017	Andean
Bombona	1.18	−77.46	1493	2005–2017	Andean
Tanama	1.37	−77.58	1500	2005–2017	Andean
Sindagua	1.11	−77.39	2800	2005–2017	Andean
Barbacoas	1.67	−78.13	32	2005–2012	Pacific
Monopamba	0.99	−77.15	2719	2006–2016	Amazon
El Encano	1.15	−77.16	2830	2005–2017	Andean
Chimayoy	1.26	−77.28	2745	2005–2014	Andean

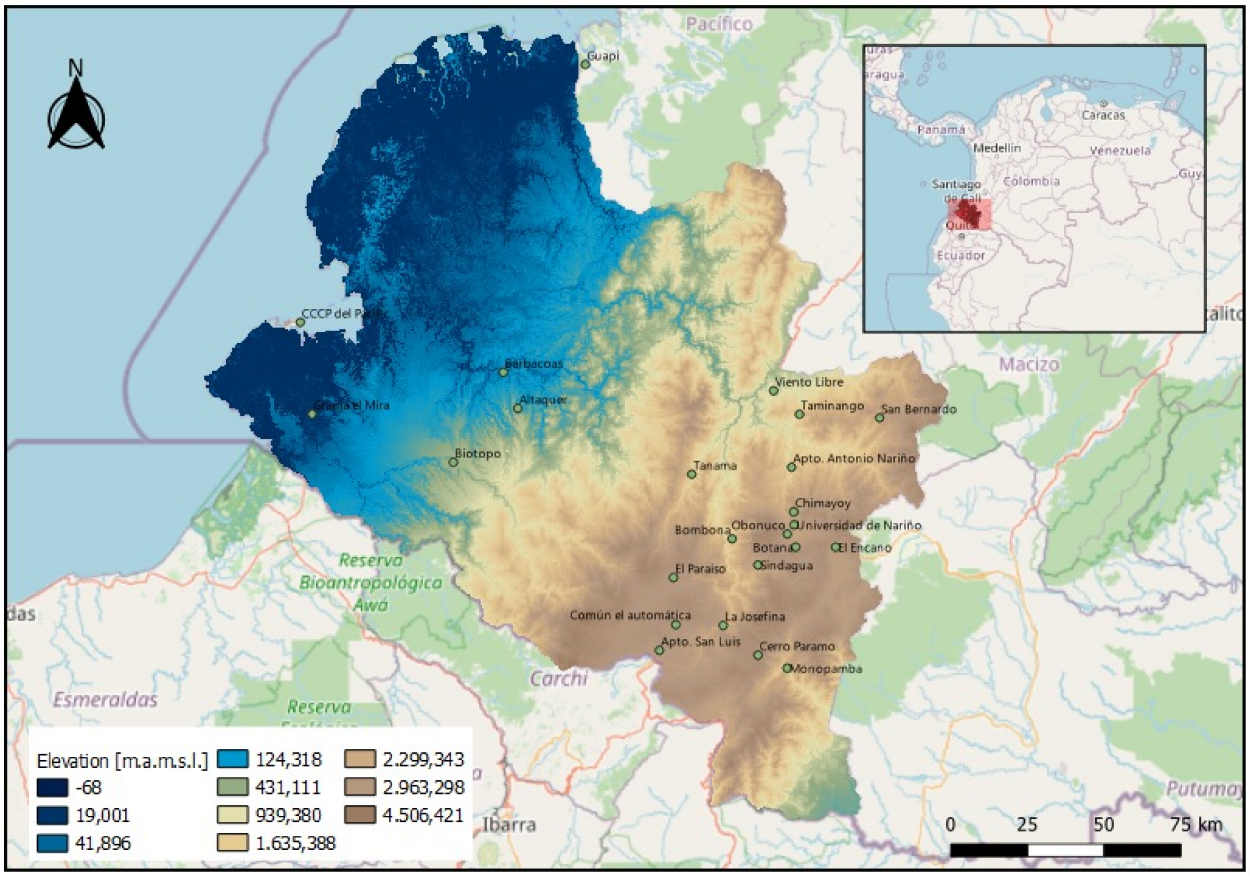

Figure 1. Automatic and conventional weather stations in Nariño. Source: made by the authors with information from OpenStreetMap.

2.2. Data Quality Control

Data quality control is a procedure intended to improve the reliability of time series data. The procedure includes the analysis of database structure, the comparison with fixed and flexible limits, and time consistency. The application of these steps offered reliable results in studies of weather data quality control in Spain [22]. The quality control procedure followed in the current research study relied on guidelines presented in regulation UNE500540, which outlines successive analytical procedures of quality control for different weather variables [23]. The first step consists of checking the database structure, the second step compares the data with the maximum extraterrestrial solar irradiance, the third step compares the time series with the maximum and minimum clear-sky global solar irradiance limits, and the fourth step consists of analyzing changes in global solar irradiance hour by hour.

2.2.1. Solar Irradiance Data Quality Control

The first step consists of checking 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 [15]. For the fixed-range step, UNE500540 suggests using the most restrictive condition between the measurement device limit and the physical phenomenon limit. Nariño's AWS have Kipp & Zonen CMP11 pyranometers with an upper operation limit of 4000 W/m^2 [24]. The physical limit is the

maximum extraterrestrial solar irradiance I_0 of each location computed with the following equation [25]:

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

where I_{sc} is the solar constant, D is the Julian day, β is the solar altitude, and δ is the declination calculated as follows:

$$\delta = 23.45 \times sin\left[\frac{360 \times (D + 284)}{365}\right],$$

where φ is the latitude, and ω_s is the hour angle calculated as follows.

$$\omega_s = \cos^{-1}[-\tan(\delta)\tan(\varphi)].$$

Therefore, the most restrictive condition is the extraterrestrial solar irradiance [22]. Consequently, $I_0 \geq I_{mt}$, where I_{mt} is the measured global solar irradiance by a pyranometer at time t.

For the flexible range test, UNE500540 suggests comparing the time series with the maximum and minimum values of a validated time series. In this case, there was no time series previously validated; therefore, we used, as an alternative, the restriction proposed for Estévez et al. with one modification. Namely, instead of using I_0, we used the following range defined by the clear-sky global solar irradiance: $(0.03 * I_{\mathrm{cs}} \leq I_{mt} \leq I_{\mathrm{cs}})$ [21]. The clear-sky global solar irradiance is equal to τ times I_0, $I_{cs} = I_0 * \tau$, and τ is the atmospheric transmittance estimated with Kreith and Kreider's model as follows [25]:

$$\tau = 0.56 \times \left(e^{-0,65 \times m} + e^{-0,095 \times m}\right)$$

where the air mass m is defined as follows.

$$m = \frac{1}{\sin(\beta)}$$

The fourth step consists of analyzing changes in global solar irradiance hour by hour. This analysis, known as time consistency, follows the restriction $|I_{cst} - I_{cst-1}| > |I_{mt} - I_{mt-1}|$. Time consistency is a useful means 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.

In addition to 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 [24]. It is essential to offset this imbalance because neglecting it would entail underestimated solar irradiance records between 0.7% and 4.3%. Several environmental factors influence the measurement process, so it difficult to correct all measurement instruments in all locations and environments [26]. Although various approaches to correct the zero-offset adapted to specific environmental conditions and instruments exist, the current research did not follow such approaches.

2.2.2. Temperature Data Quality Control

We carried out five temperature validation procedures, structure, range, step, consistency, and persistence, based on the recommendations of Estévez et al. and Rivero et al. The first step consists of verifying the database structure and that of the solar irradiance validation. The range test involves two evaluation methods: the instrumental range method determined by the ROTRONIC HYGROCLIP 2 RTD PT100 with a temperature range between −40 °C and 60 °C and the use of data validation criteria defined by other

researchers. Rivero et al. recommended a temperature interval between −50 °C and 70 °C, while Estévez et al. suggested a range of −30 °C to 50 °C. We followed Estévez et al.'s recommendation because it is within the temperature range of the sensor. Likewise, the analysis of the step test, internal consistency test, and persistency test also followed Estévez et al.'s quality control process because it offered promising data analysis results in Spain [22,27].

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 air temperature measured at time h. Although this evaluation does not consider other climatology aspects that can affect temperature variation, such as humidity, wind speed, cloudiness, and precipitation, the authors of this study applied these restrictions. Another fundamental requirement is that the daily step restriction fulfills the following condition: $T_{max} - T_{min} < 30$ °C, where T_{max} and T_{min} are the maximum and minimum daily temperatures [26]. Internal consistency assesses the accomplishment of the following conditions: $T_{max} > T_{mean} > T_{min}$; $T_{max}(d) > T_{min}(d-1)$; and $T_{min}(d) \leq T_{max}(d-1)$. Finally, the persistence test verifies measurement variability; therefore, the data must accomplish the conditions $T_{max}(d) \neq T_{max}(d-1) \neq T_{max}(d-2)$ and $T_{min}(d) \neq T_{min}(d-1) \neq T_{min}(d-2)$, where d is the day number [22].

2.3. Empirical Temperature-Based Models

Although air temperature is a standard variable measured in weather stations, it has been infrequently used to estimate solar insolation. Temperature started to be relevant in estimating solar insolation when agricultural studies modeled solar insolation to analyze crop production rates. Consequently, researchers from other fields increasingly focus on the maximum, minimum, and mean temperature values in order to model solar insolation [28]. The most traditional models incorporating maximum and minimum temperatures are the Hargreaves and Samani and Bristow and Campbell models [4,10]. Both models are the basis of innovative approaches adapted to specific location conditions [7,12,27,29–32].

2.3.1. Hargreaves and Samani's Model

In 1982, Hargreaves and Samani (HS) proposed a linear relationship between the square root of the temperature difference and the fraction between the extraterrestrial and terrestrial global solar insolation for different periods, as follows:

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

where a is the empirical coefficient, H is the global solar insolation on a horizontal surface, and H_0 is the daily extraterrestrial solar insolation calculated as follows.

$$H_0\left(\frac{Wh}{m^2 day}\right) = \frac{24}{\pi} \times I_{sc} \times \left(1 + 0,033 * cos\left(\frac{360 * D}{365}\right)\right) \times (cos(\varphi) \times cos(\delta) \times sin(\omega_s) + \frac{\pi}{180} \times \omega_s \times sin(\varphi) \times sin(\delta)).$$

This model, however, did not consider the effects of cloudiness, humidity, latitude, elevation, or topography, among others, in the specific location for which the model is used [33,34]. Allen stated that the HS model demonstrates better behavior in a monthly time frame than in a daily time frame because the variables follow a mean trend, resulting in a consistent relationship between $T_{max} - T_{min}$ and H/H_0 [35].

The empirical coefficient represents the rate of change from the maximum and minimum temperature difference with the ratio between the extraterrestrial and terrestrial solar insolation. Initially, the HS model proposed an empirical coefficient calibrated with an eight-year time series from Central Valley in Davis County, California (see Table 3) [4,9]. They found that a relative humidity above 54% affects the empirical coefficient, and that the maximum and minimum temperature difference is smaller when it is nearly this percentage. As shown in Table 3, in humid regions, the empirical coefficient has a value less than 0.10 while other regions have values higher than 0.16 [9]. 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 result in significant errors because topography influences temperature, the advective environment, and vegetation [15].

Table 3. Hargreaves and Samani's 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: [4,14].

2.3.2. Bristow and Campbell's Model

Bristow and Campbell (BC) presented a model based on two assumptions. The first assumption is that of a linear relationship existing between the absorbed and incoming solar insolation, while the second assumption is that of disregarding the heat flux coming from the soil for a daily period because it is close to zero. However, it is convenient to mention that the sensible heat produced by diurnal temperatures is higher than that produced during the nighttime [10]. This is understandable because solar irradiance heats air masses more due to short-wave radiation, whereas at night, there is less long-wave emission from the Earth to the atmosphere, reducing the temperature [29]. Furthermore, under ideal conditions, the temperature is minimum just before sunrise, resulting in a significant difference between daily maximum and minimum air temperature. This phenomenon permits the modeling of solar insolation as a function of temperature difference [10]. Under these assumptions, the authors presented their model:

$$\frac{H}{H_0} = a\left[1 - e^{(-b\Delta T^c)}\right],$$

where a, b, and c are empirical coefficients. a represents the maximum ratio between the extraterrestrial and terrestrial global solar insolation H/H_0; b and c determine how soon the model achieves such a maximum by temperature increases ΔT. Empirical coefficients represent the regional characteristics from arid to humid environments [10].

Temperature difference depends on the daily maximum T_{max} and minimum T_{min} temperatures for a day D given in °C. Bristow and Campbell concluded that the minimum temperature averages of two consecutive days reduces the effect from hot or cold air masses, in turn, avoiding overestimated and underestimated solar insolation values. Accordingly, Dos Santos et al. stated that advection is not a common phenomenon in tropical zones; therefore, the temperature change ΔT estimation in these regions is $\Delta T(D) = T_{max}(D) - T_{min}(D)$. Dos Santos et al. also highlighted that this equation is better for sites with high altitudes, as in the study case [14].

Another weather effect modifying the solar insolation estimation is rain. The effect of rain decreases, setting $\Delta T(J)$ equal to 0.75 times the estimated $\Delta T(D)$. If $\Delta T(D-1)$ is less than $\Delta T(\mathrm{D}-2)$ by about 2 ºC, the first factor is multiplied by 0.75 [10]. However, when the rainy period is long, the relation between solar insolation and ΔT reaches an equilibrium. Therefore, it does not require adjustments [36].

2.3.3. Models Implemented in Tropical Environments

Knowledge of solar irradiance behavior in tropical zones is a mandatory requirement for this research study. The analyzed cases are from 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, and Sao Jose da Laje), and Mexico. Due to the fact that Nariño is in a tropical area, the results from these three cases constitute relevant inputs for the current research.

Okundamiya and Nzeako (ON) proposed a linear model as follows [37]:

$$\frac{H}{H_0} = a + bT_R + cT_{max},$$

where T_R is the monthly average of the ratio between the daily minimum and maximum temperatures (T_{min}/T_{max}), and a, b and c are empirical coefficients. The statistical validation of the model yielded a coefficient of determination between 0.809 and 0.952.

Nwokolo and Ogbulezie reviewed empirical models implemented in West Africa in order to compute global solar irradiance. They found that the soft computing models have better accuracy than the empirical models since they can be more easily adapted to several weather conditions. This is because soft computing models allow more inputs, as models or variables, to strengthen their reliability [38].

Dos Santos et al. studied the performance of 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 et al. also concluded that the HS model demonstrates better performance in the hinterlands and interior regions, whereas the BC model showed the best performance in humid and coastal zones [14].

Rivero et al. compared the values of the original empirical coefficients of the HS with new values generated from the model calibrated for Mexico with local data. A notable element of that research was the classification of Mexico's climate zones using the Köppen–Geiger system. Furthermore, the authors developed another classification based on the clearness index, as shown in Table 4, to support the AWS data. It is useful to remember that the clearness index is the ratio between terrestrial solar insolation and extraterrestrial solar insolation. A relevant conclusion in this respect was that regardless of solar irradiance peaks during the day, it is possible to obtain similar clearness index values. Another notable conclusion was that fixed empirical coefficients result in significant error, especially in zones with a temperature difference below 15 °C. The authors proposed the following equation to overcome the identified issue derived from fixed coefficients [15]:

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

where a_{HS} is the Hargreaves and Samani equation coefficient.

Table 4. 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: [15].

2.3.4. Proposed Empirical Model

The proposed model originated from observing the scatter plot between the clearness index and the daily temperature difference in each AWS. We concluded that the logistic model would offer useful results, because this model has successfully studied human growth, animal and biological processes, energy use patterns, and atmospheric applications [39,40]. The logistic regression commonly describes the relationship between binary variables and a predictor [41]. Although the relationship between extraterrestrial and terrestrial solar insolation does not offer binary results, it varies in a range between 0 and 1. The use of this technique *"arises in estimating relationships in which the dependent variable is continuous*, but is limited in range" [42]. Manning's statement describes the relationship between the dependent and independent variables in the proposed model because solar insolation relationships are continuous, and the temperature range is defined by the minimum and maximum temperatures. Logistic regression does not assume that

neither variables nor predictors have a normal distribution; this affirmation was another criterion to choose this approach [43].

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 to 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 [44]. This regression also assumes a linear relationship between the dependent and independent variables' log-odds [40].

The logistic function came from the logistic model and is represented as follows:

$$y = \frac{1}{1 + e^{-z}},$$

where y is the predicted variable, and z is a linear sum $a + \sum b_i x_i$ where x_i is the independent variable, and a and b are constants [41].

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

In this case, the temperature change ΔT is the predictor variable. Therefore, the proposed model is defined as follows.

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

2.4. Statistical Validation

Statistical validation is a mandatory step that allows the comparison of the predictor model with real measures to determine the suitability of the model. There are five main validation techniques: subjective assessment, dispersion indicators, overall performance indicators, distribution similitude indicators, and visual indicators [45,46]. We implement the second technique, because it is appropriate when the data have the same time frame, location, and treatment, among other characteristics. This validation technique measures the difference between the modeled and real value. Table 5 shows some measured dispersion indicators, expressed in absolute units or percentages [45].

Table 5. Error measurements.

Measurement	**Definition**	**Formula (p_i is The Predicted Value, o_i Is the Observed Value, P_m Is the Mean Predicted Value, O_m Is the Mean Observed Value, and n Is the Amount of Data)**
Mean of percent error (MPE)	Values close to zero indicate a better model and suggest that the ratio of the standard deviation of the measured and computed value is near one.	$\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)/o_i$
Mean absolute error (MAE)	The average vertical distance between each predicted and observed point.	$\frac{1}{n}\sum_{i=1}^{n}\lvert p_i - o_i\rvert$
Root mean square error (RMSE)	This provides a measure of the error size but is sensitive to outlier values because the measure gives more weight to large errors.	$\left[\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2\right]^{1/2}$

Table 5. *Cont.*

Measurement	Definition	Formula (p_i is The Predicted Value, o_i Is the Observed Value, P_m Is the Mean Predicted Value, O_m Is the Mean Observed Value, and n Is the Amount of Data)
Mean bias error (MBE)	This measure provides information on the model's long-term performance when the model includes a systematic error that presents overestimated or underestimated predictors. Low MBE values are desirable, although it should be noted that an overestimated dataset will cancel out another underestimated dataset.	$\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)$
Standard Deviation of the residual (SD)	This measure shows the difference between the standard deviation of the predicted and observed datasets.	$\left(\frac{100}{O_m}\right)\frac{1}{n}\left[\sum_{i=1}^{n} n(p_i - o_i)^2 - \left[\sum_{i=1}^{n}(p_i - o_i)\right]^2\right]^{\frac{1}{2}}$
Uncertainty at 95% (U_{95})	This is a measure of certainty confidence; a lower value is expected.	$1.96\left(SD^2 + RMSE^2\right)^{1/2}$

Source: [7,14,45–47].

3. Results and Discussion

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

Figure 2 shows the scheme followed for implementing the models. We obtained the raw data of AWS and conducted a quality control process for data cleaning. Subsequently, we selected the days with at least six measured per day, aggregated the hourly irradiance to estimate the insolation, and chose the maximum and minimum daily temperature. We randomly selected 80% of the data of AWS for empirical model calibration and 20% for statistical validation.

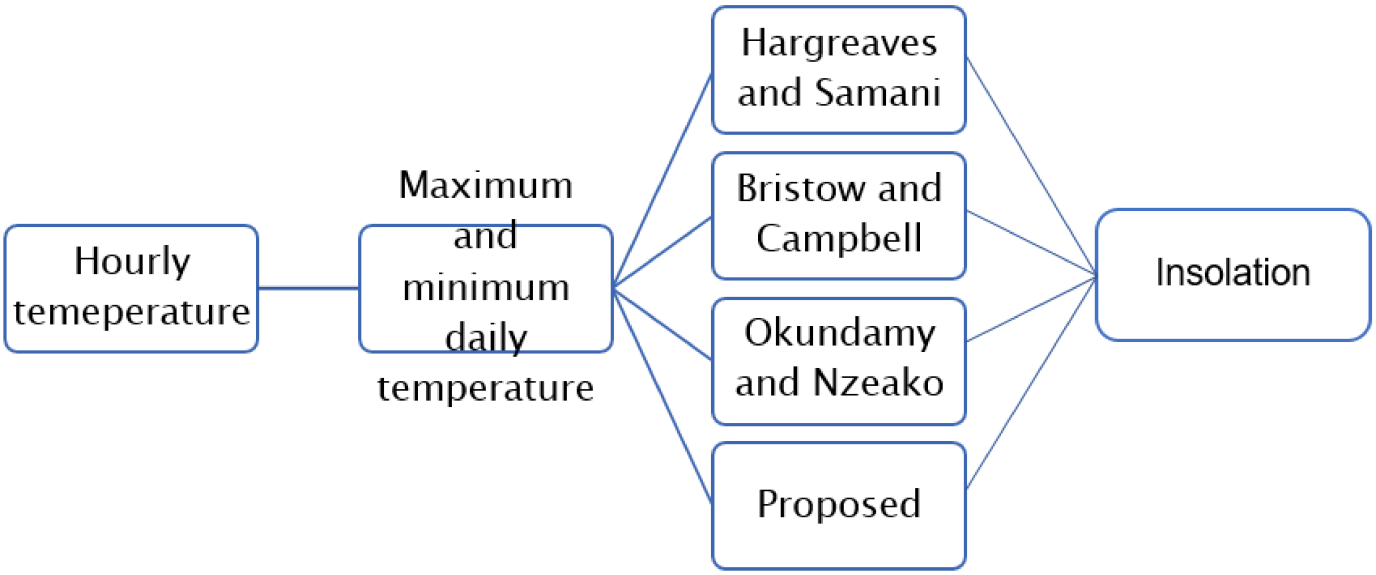

Figure 2. Scheme for empirical model implementation.

3.1. Quality Control: Global Solar Irradiance and Temperature

A step undertaken prior to the quality control procedure is data adjustment stemming from the calibration constant. In this case, there was no calibration constant for the *Biotopo* AWS affecting the quality of the time series; likewise, this AWS also had the smallest number of recordings among all AWS. The total amount of data (47,612) for the period studied corresponded to 34.50% of the total data that should be recorded. The AWS registered information during 57.59% of the measured period 2005–2017.

Table 6 shows the results of the global solar irradiance validation procedure. The first step, which evaluated the database structure, presents 5843 recordings on average with the incorrect structure. The AWS with the most recordings that had an incorrect data structure was *Viento Libre*; this step confirmed that 9715 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 was *Biotopo*; this step confirmed that 1055 data did have the database structure, corresponding to 2.22% of the total. The fixed range validation discarded 36.71% of data on average. *Biotopo* lost the most data, corresponding to 47.62% of the total, whereas *Universidad de Nariño* lost the fewest data, amounting to about 22.09% of the total. The flexible range test results reveal information losses of 39.11%. The AWS with the lowest number of losses was *Biotopo*, about 15.14%, while *Guapi* had the highest number of losses, about 5422%. Although the last test was not mandatory, it is useful to point out that 44.06% of the recordings did not overcome the time consistency level. Considering only the mandatory steps, about 35.27% of the data overcame these validation steps. *Universidad de Nariño's* AWS had the most data approving the validation process, whereas *Guapi's* AWS had the fewest data approving the validation process.

Table 6. Solar irradiance validation results. Each step represents the amount of data that overcome the validation level.

Name	Code	Data	Step 1	Step 2	Step 3	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	7427
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	15,919

The first row of Table 7 represents the recorded data number between 06:00 and 18:00. The subsequent rows show the day numbers with the recorded data amount indicated in the heading—for instance, *Biotopo* has 13 days with only one record. The columns highlighted in gray represent that each day has between 10 and 11 recordings per day. It is worth noting that only 1.26% of days had complete information, 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 underestimating the resource—approximately 95.81% of the recordings that overcame the mandatory validation levels accomplished this requirement.

Table 7. Classification of days according to irradiance recording per day. The rows contain the number of days classified by the amount of measured data between the 6:00 and 18:00 per day.

Name \ Data Number per Day *	1	2	3	4	5	6	7	8	9	10	11	12	13	Total of Days
Biotopo	13 **	16	15	19	30	49	104	190	350	528	**622**	207	6	2149
Viento Libre	43	50	77	86	121	175	313	444	702	**654**	390	119	11	3185
Universidad de Nariño	27	42	56	73	97	173	340	403	730	**1014**	778	308	63	4104
Cerro Páramo	34	39	42	45	59	83	145	283	483	952	**1080**	362	160	3767
La Josefina	109	71	22	24	25	57	78	195	293	**386**	274	134	6	1674
Botana	16	20	26	43	82	142	246	469	798	**1066**	839	336	14	4097
El Paraiso	55	83	110	137	158	205	285	364	614	**787**	514	149	13	3474
Guapi	185	193	180	153	111	101	139	239	335	447	**549**	182	75	2889

* This number represents the amount of data per day. ** This number represents the number of days in all time series with one measure per day.

The results confirm the importance of improving the maintenance and calibration procedures of the measurement instruments in order to increase the amount of useful information, in turn increasing reliability.

We considered the Rivero et al.'s classification and modified the item concerning partially cloudy days, as Table 8 shows. Table 8 shows that, in the analyzed AWS, on average, 64.70% of days were classified as partially high cloudiness. This corresponds to losses between 20% and 40% of extraterrestrial solar irradiance when reaching ground level. This result is important not only because clouds affect global solar irradiance but also because, as a variable, it is not easy to model. Consequently, the resource estimation in this tropical and mountainous environment is more complicated than in other environments.

Table 8. Days' classification according to cloudiness.

k_t	$0.00<k_t\leq0.20$	$0.20<k_t\leq0.40$	$0.40<k_t\leq0.60$	$0.60<k_t\leq0.75$	$0.75<k_t\leq1.00$	
Name	**Cloudy**	**Partially High Cloudiness**	**Partially Low Cloudiness**	**Sunny**	**Very Sunny**	**Number of Days ***
Biotopo	760	**1188**	67	0	0	2015
Viento Libre	105	**1458**	1179	0	0	2745
Universidad de Nariño	243	**2611**	846	1	0	3701
Cerro Páramo	**1380**	1232	137	1	0	2793
La Josefina	88	**991**	271	1	0	1351
Botana	451	**2704**	711	2	0	3868
El Paraiso	167	**1859**	543	1	0	2570
Guapi	130	**746**	125	0	0	1001
Average	16.99%	**64.70%**	18.28%	0.03%	0.00%	100%

* Days with ΔT, T_R, T_{max}, T_{min}, and k_t complete information.

Table 9 shows the results of the temperature validation procedure. In the first step, missing information was 8.64% on average. *Cerro Páramo* AWS missed the most information amount, with 16.95% missing data. In the fixed range test, *Guapi* had data missing of about 11.62%. In the step test, about 35.07% of the data did not pass the validation requirement. Considering the starting values as the base, 57.08% of the data overcame the quality control procedure.

Table 9. Temperature validation results. Each step presents the amount of data that overcome the validation level.

Name	Code	Data	Step 1	Step 2	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 10 shows, as Table 7, the record numbers by day. The results indicate that about 57.42% of the days had 88.46% of useful information. Finally, from the total amount of data that overcame the hourly validation test, only 68.03% of data were suitable for feeding the models.

Table 10. Classification of days according to irradiance recording per day. The rows contain the number of days classified by the amount of data measured between 6:00 and 18:00.

Name \ Data Number per Day *	1	2	3	4	5	6	7	8	9	10	11	12	Total of Days
Biotopo	33 **	9	17	13	26	46	56	85	130	257	605	**846**	2123
Viento Libre	79	68	91	91	94	120	124	192	264	500	**873**	696	3192
Universidad de Nariño	74	26	24	31	47	71	113	198	309	599	1170	**1379**	4041
Cerro Páramo	101	49	58	47	67	86	133	190	348	590	**754**	701	3124
La Josefina	61	21	36	46	50	78	109	130	197	405	**573**	558	2264
Botana	38	11	27	46	62	102	147	253	386	679	1114	**1221**	4086
El Paraiso	74	42	65	67	92	102	155	236	373	541	**915**	909	3571
Guapi	20	8	10	21	18	21	45	79	168	405	919	**1349**	3063

* This number represents the amount of data per day. ** This number represents the number of days in all time series with one measure per day.

3.2. Model Development and Performance

3.2.1. Model Development

By observing the scatter plot from Figures 3–10, we concluded that the logistic model would offer helpful results because logistic regression commonly describes the relationship between binary variables and a predictor [40]. Although the relationship between extraterrestrial and terrestrial solar insolation does not offer binary results, it varies in a range between 0 and 1. Furthermore, logistic regression does not assume that neither variables nor predictors have a normal distribution [42].

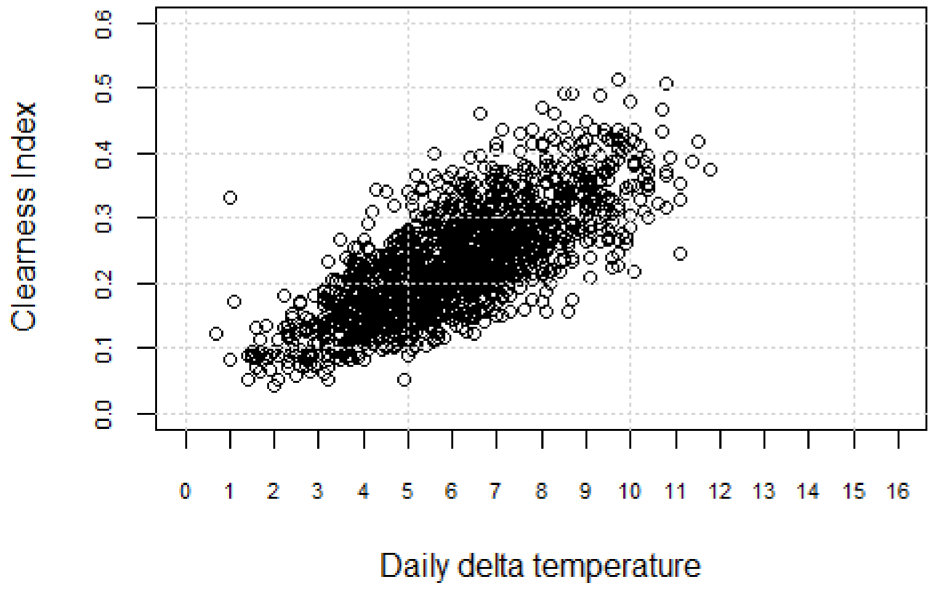

Figure 3. Daily delta of temperature–clearness index for Biotopo.

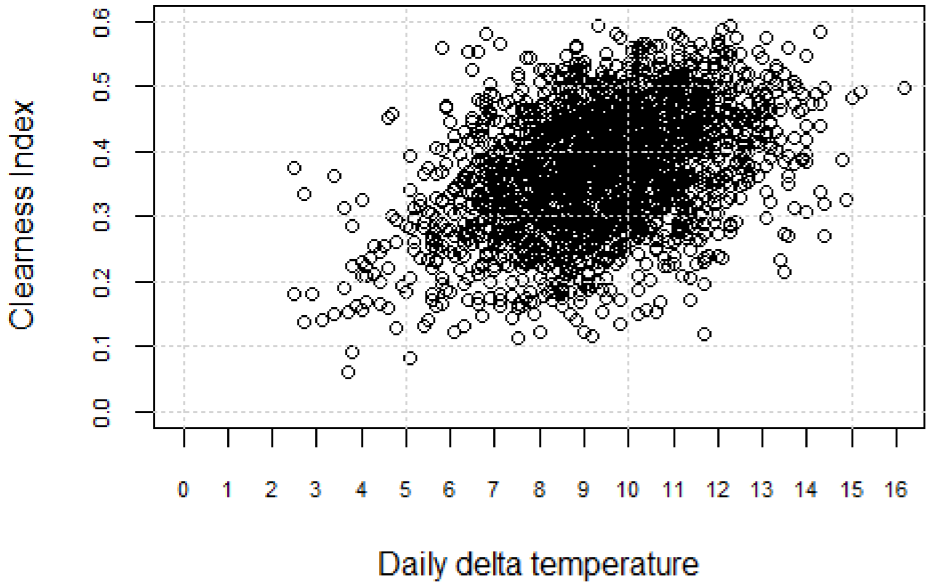

Figure 4. Daily delta of temperature–clearness index for Viento Libre.

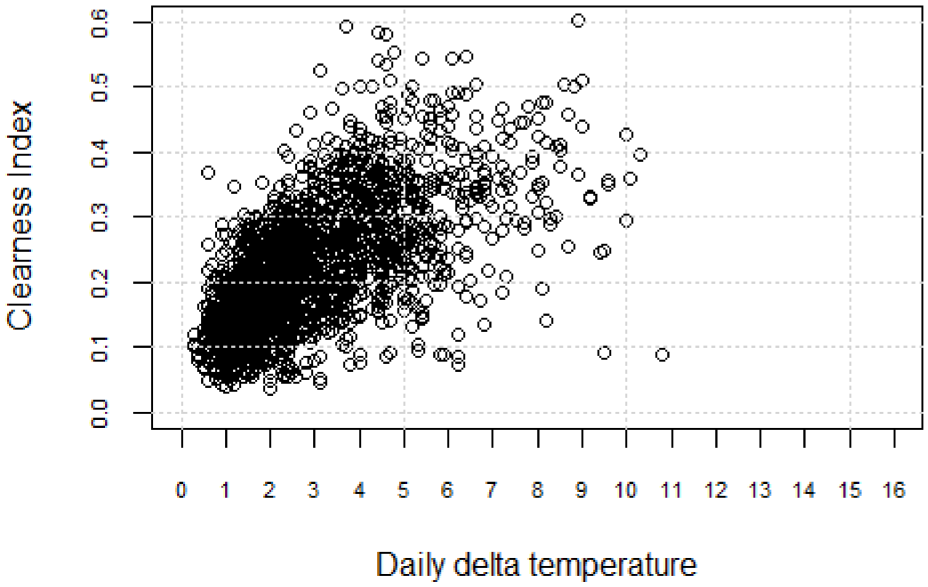

Figure 5. Daily delta of temperature–clearness index for Cerro Páramo.

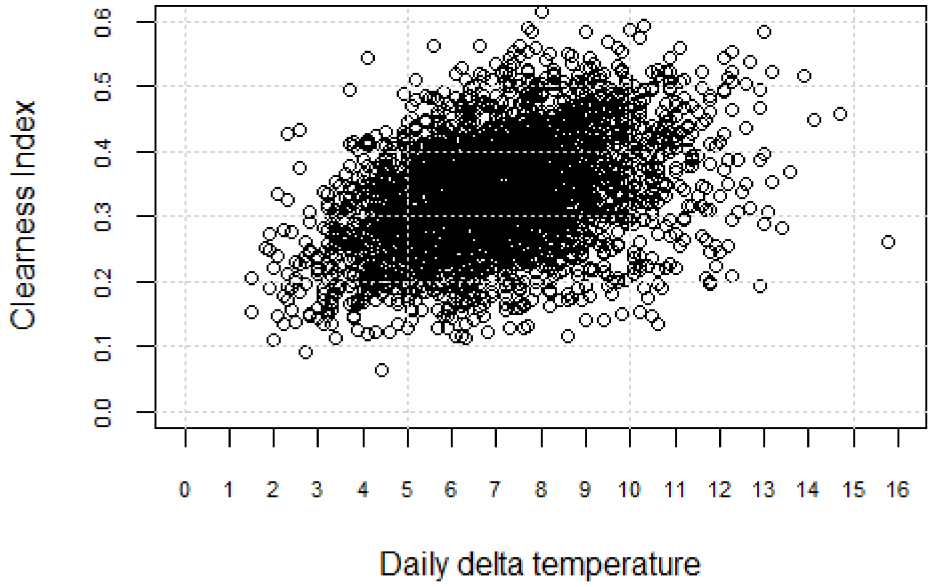

Figure 6. Daily delta of temperature–clearness index for Universidad de Nariño.

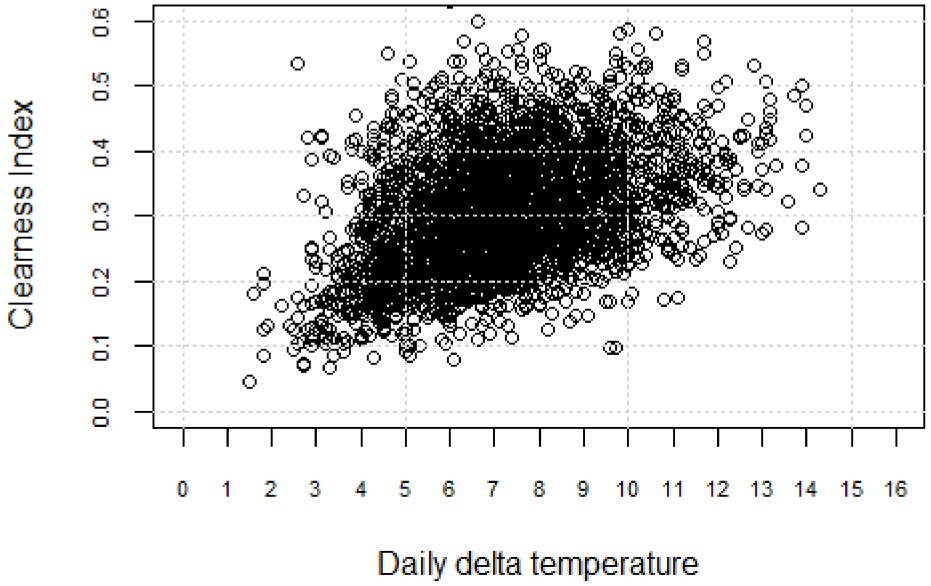

Figure 7. Daily delta of temperature–clearness index for Botana.

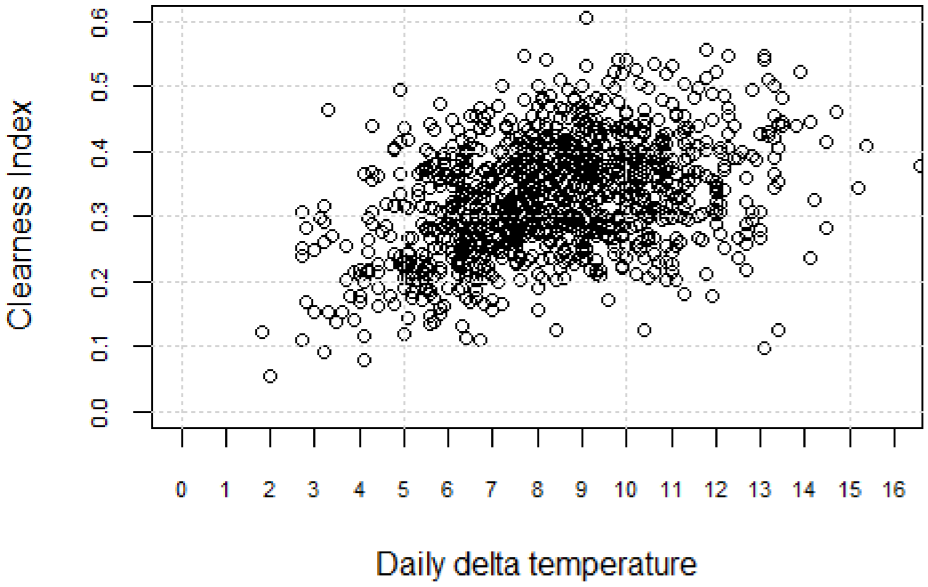

Figure 8. Daily delta of temperature–clearness index for La Josefina.

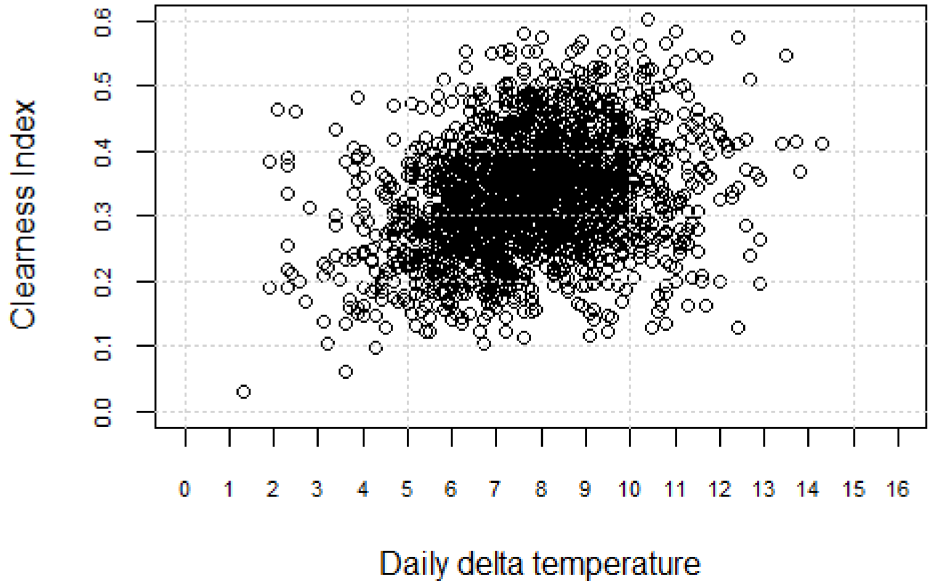

Figure 9. Daily delta of temperature–clearness index for Paraiso.

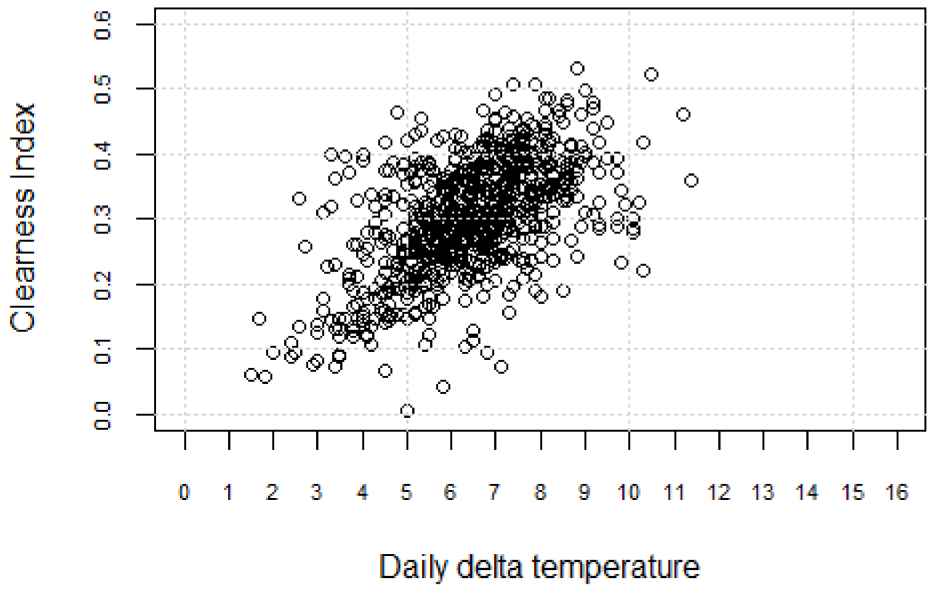

Figure 10. Daily delta of temperature–clearness index for *Guapi*.

3.2.2. Performance

Figures 3–10 show the relationship between the daily clearness index and daily delta of temperature for the studied AWS. Figures 3 and 10 show that there is a moderate linear relationship between the daily clearness index and daily delta of temperature in the Pacific zone. Figure 5 shows that, in *Cerro Páramo*, the clearness index and the delta of temperature possess a moderate linear relationship. In the Andean zone, the clearness index and delta of temperature exhibited a weak linear relationship.

Table 11 shows the empirical coefficients of the HS, BC, and ON models and those of the proposed model. The *a* and *b* empirical constants of the BC model demonstrated a growing trend corresponding to altitude, while the *c* empirical constant showed the opposite behavior. The empirical coefficients of the HS model did not present a significant variation among AWS. This means that the AWSs have similar humid conditions, considering the original values given by [9]. The empirical coefficient of the ON model for the *Bitopo* AWS revealed the unique negative value. More studies and data are necessary to understand this result. However, there were two influencing factors: *Biotopo* AWS did not have calibration factor, and the data number was the lowest.

Table 11. Empirical coefficients.

	Bristow and Campbell		
AWS	*a*	*b*	*c*
Biotopo	0.5075	0.0735	1.1908
Viento Libre	0.5942	0.1499	0.8655
Cerro Páramo	0.5922	0.2595	0.6153
Universidad de Nariño	0.4893	0.3282	0.6568
Botana	0.6288	0.1964	0.6415
Josefina	0.6039	0.2563	0.5350
Paraiso	0.5850	0.3183	0.4818
Guapi	0.4471	0.1156	1.2478

Table 11. *Cont.*

Hargreaves and Samani			
AWS	*a*		
Biotopo	0.0970		
Viento Libre	0.1248		
Cerro Páramo	0.1340		
Universidad de Nariño	0.1263		
Botana	0.1169		
Josefina	0.1138		
Paraiso	0.1199		
Guapi	0.1208		
Okundamiya and Nzeako			
AWS	*a*	*b*	*c*
Biotopo	−0.1838	−0.3871	0.0264
Viento Libre	0.0578	−0.2666	0.0168
Cerro Páramo	0.1084	−0.1572	0.0257
Universidad de Nariño	0.2679	−0.2416	0.0112
Botana	0.0621	−0.1547	0.0191
Josefina	0.1770	−0.1589	0.0118
Paraiso	0.2617	−0.1775	0.0100
Guapi	0.0717	−0.5202	0.0228
Proposed model			
AWS	*a*	*b*	
Biotopo	−2.3058	0.1786	
Viento Libre	−1.3499	0.0912	
Cerro Páramo	−1.7914	0.1706	
Universidad de Nariño	−1.2211	0.0747	
Botana	−1.4489	0.0898	
Josefina	−1.2299	0.0608	
Paraiso	−1.1667	0.0607	
Guapi	−1.8043	0.1495	

Table 12 presents the results of seven statistical validation measurements for each AWS. Considering $RMSE$, SD, MAE, U_{95}, and $MAPE$, the proposed model had better performance than the other models. That said, the BC model had better results for MAE and MPE. The proposed model showed better results in AWS located at altitudes above 2500 MASL. In contrast, the HS model showed better performance at altitudes below 2500 MASL. However, if the objective is to use a unique model to estimate solar irradiance from temperature data in the state, the proposed model remains the best.

Table 12. Summary of the empirical model's statistics results. The best result for each AWS is in bold.

$RMSE\ [Wh/m^2\ day]$				
AWS	BC	HS	ON	Proposed
Biotopo	1.152,62	**993,64**	1.155,07	1.113,48
Viento Libre	1.086,47	1.080,72	1.110,62	**1.077,35**
Cerro Páramo	1.194,57	1.209,76	1.196,56	**1.152,72**
Universidad de Nariño	1.032,45	1.083,73	1.032,24	**1.019,14**
Botana	1.052,42	1.070,23	1.077,17	**1.042,68**
Josefina	1.009,00	1.066,18	999,28	**984,75**
Paraiso	930,82	990,32	938,02	**921,32**
Guapi	965,40	**878,75**	961,21	915,53
SD				
AWS	BC	HS	ON	Proposed

Table 12. *Cont.*

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

RMSE showed that the lowest value occurred in the Guapi AWS ($878,75 Wh/m^2 day$), while the highest value was in the Cerro Páramo AWS ($1.209,76 Wh/m^2 day$); on average, RMSE was $1.046,69 Wh/m^2 day$. The ON model presented the highest RMSE, with

1.058, 77Wh/m^{2}day; followed by the BC model, with 1.052, 96Wh/m^{2}day; the HS model, with 1.046, 66Wh/m^{2}day; and the proposed model, with 1.028, 37Wh/m^{2}day.

The standard deviation showed that *Cerro Páramo* presented more scattered values than the other AWSs. By comparing the four models and analyzing the average standard deviation, the best option is the proposed model. The *MBE* showed that the BC model made an underestimation of 77, 79Wh/m^{2}day in *Cerro Páramo*. All models—ON, HS, BS, and the proposed model—overestimated the solar resource in *Viento Libre* by 167, 77Wh/m^{2}day, 163, 13Wh/m^{2}day, 162, 63Wh/m^{2}day, and 160, 23Wh/m^{2}day, respectively.

The MAE showed that the proposed model had the best performance, with an average error of 821, 41Wh/m^{2}day, while the ON model reached an average error of 8.45, 27Wh/m^{2}day. U_{95} yielded, on average, the following results: 2.076, 49Wh/m^{2}day 2.065, 11Wh/m^{2}day, 2.034, 13Wh/m^{2}day, and 2.016, 86Wh/m^{2}day for ON, BC, HS, and proposed models, respectively, thereby confirming that the proposed model performed better than the other models.

MPE presented the BC model as the best-performing model, with 14.45% on average, whereas the HS model was the worst-performing, with 15.31% on average. MAPE revealed the proposed model as the best option, with 35.12% followed by the HS and ON models. The MAPE results were consistent with the other statistical results.

The proposed model's empirical coefficients showed a relationship with altitude. Figure 11 presents the linear adjustment of the empirical coefficients for two altitude ranges covering the location of AWS. On the left side, the figures represent the relationship between empirical coefficients and the altitudes below 2500 MASL; on the right side, the figures show the remaining AWS. Statistical analysis indicated that R^2 for the *a* coefficient was 0.5262 and 0.5995 in the first and second cases, respectively. R^2 for the *b* coefficient was 0.6069 and 0.8152 in the first and second cases, respectively. This result is remarkable since solar irradiance and temperature change with respect to altitude. These changes occur because of higher altitude and fewer molecules and aerosols scattering and absorbing solar irradiance [47,48].

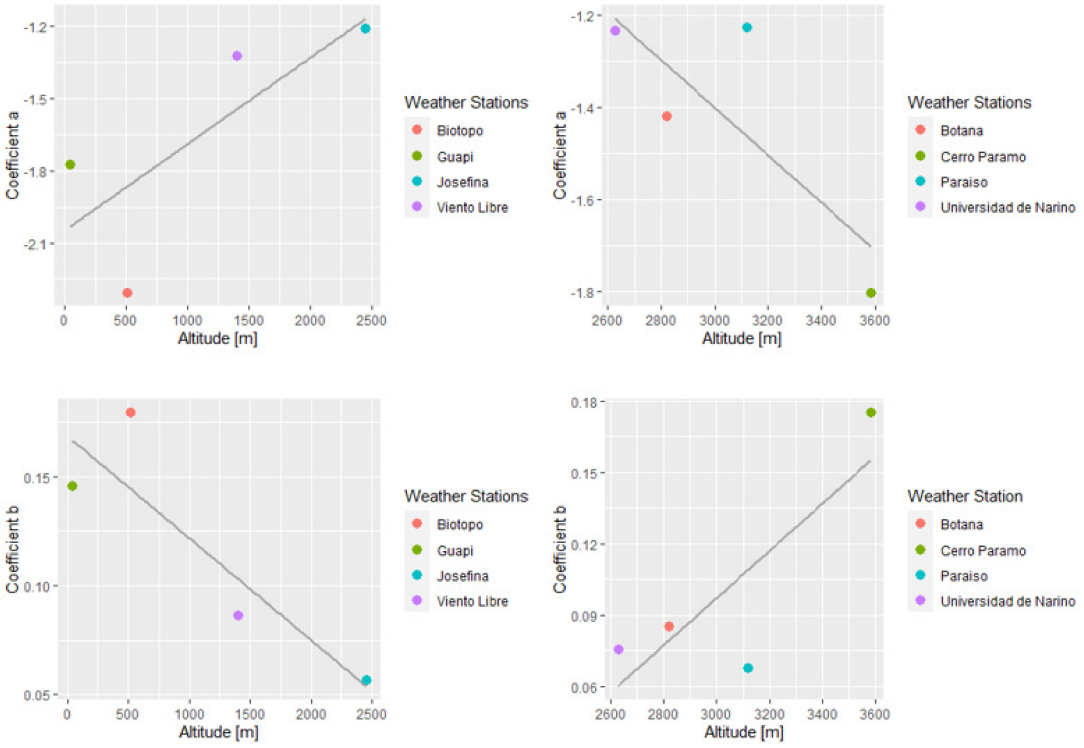

Figure 11. Coefficients a and b empirical model–altitude relation in the proposed model.

3.3. Imputation of Daily Solar Insolation Data

According to the AWS location, the imputation process used to estimate the daily solar insolation data followed the best empirical model based on the research results. Table 13 shows the number of imputed data for each AWS. *La Josefina* had the most imputed values; it was necessary to fill in 2870 missing data. *Botana* had the lowest imputed values, filling 572 missing data. On average, the empirical models allowed imputing about four years of missing data; therefore, it was necessary to fill these empty gaps by using temperature data. The time series before and after the daily solar insolation imputation process for each AWS is presented in Figures 12–19.

Table 13. Imputation by AWS.

AWS	Imputation
Biotopo	2241
Viento Libre	1502
Cerro Páramo	749
Universidad de Nariño	686
Botana	585
La Josefina	2872
Paraiso	1369
Guapi	2277

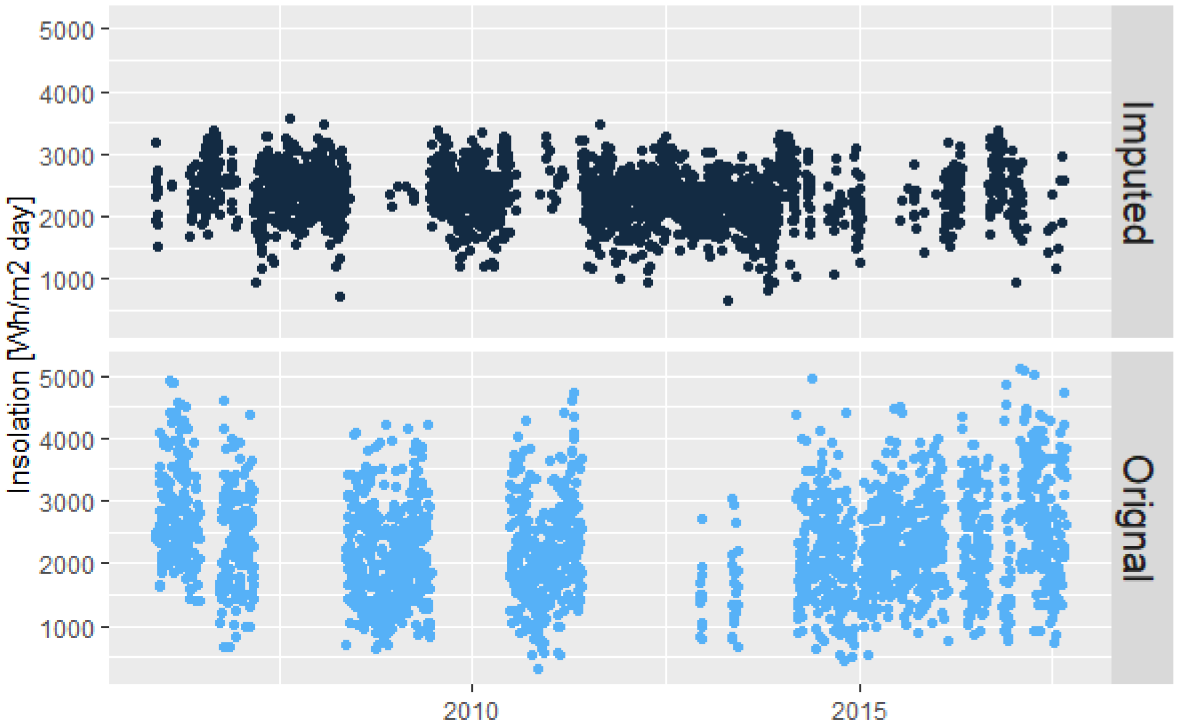

Figure 12. Biotopo imputation. The dark blue points are the estimated values using air temperature information, and the light blue points are the measured values.

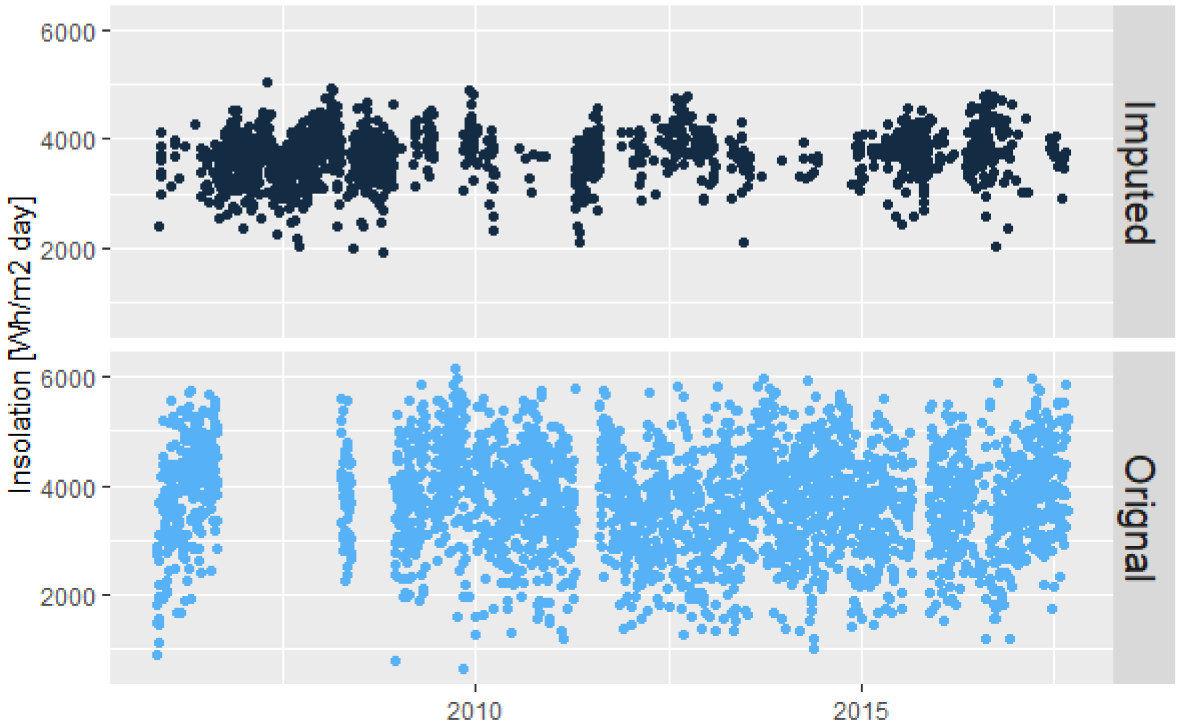

Figure 13. Viento Libre imputation.

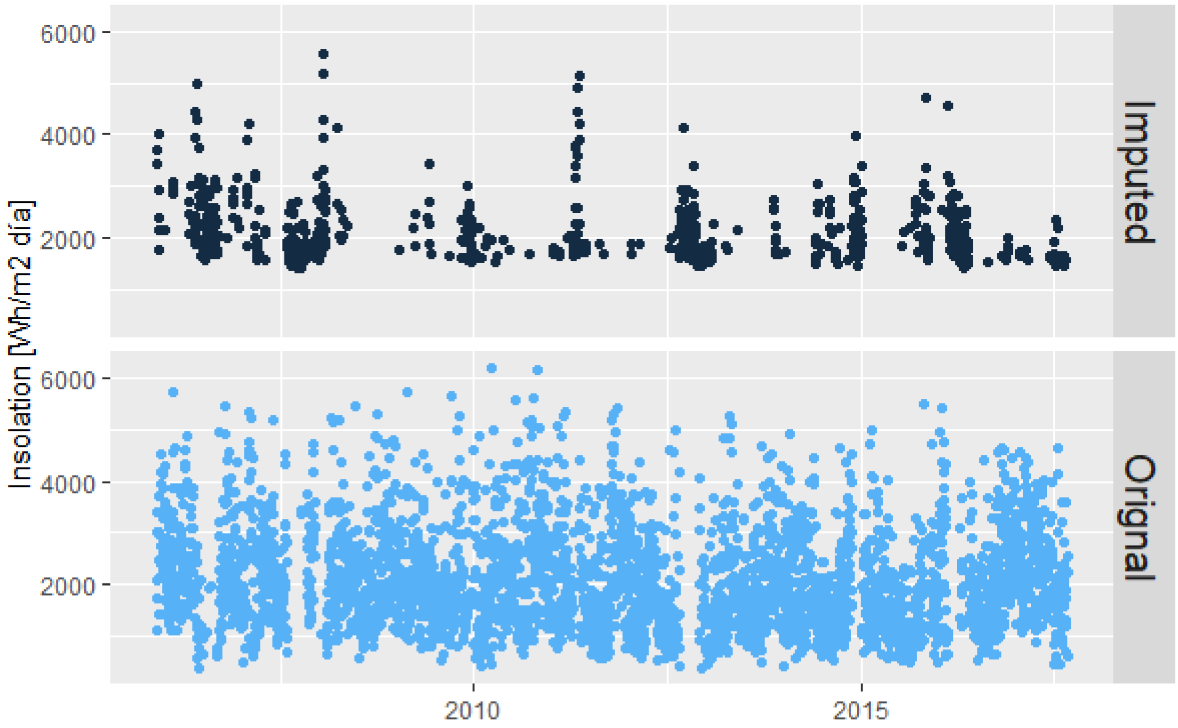

Figure 14. Cerro Páramo imputation.

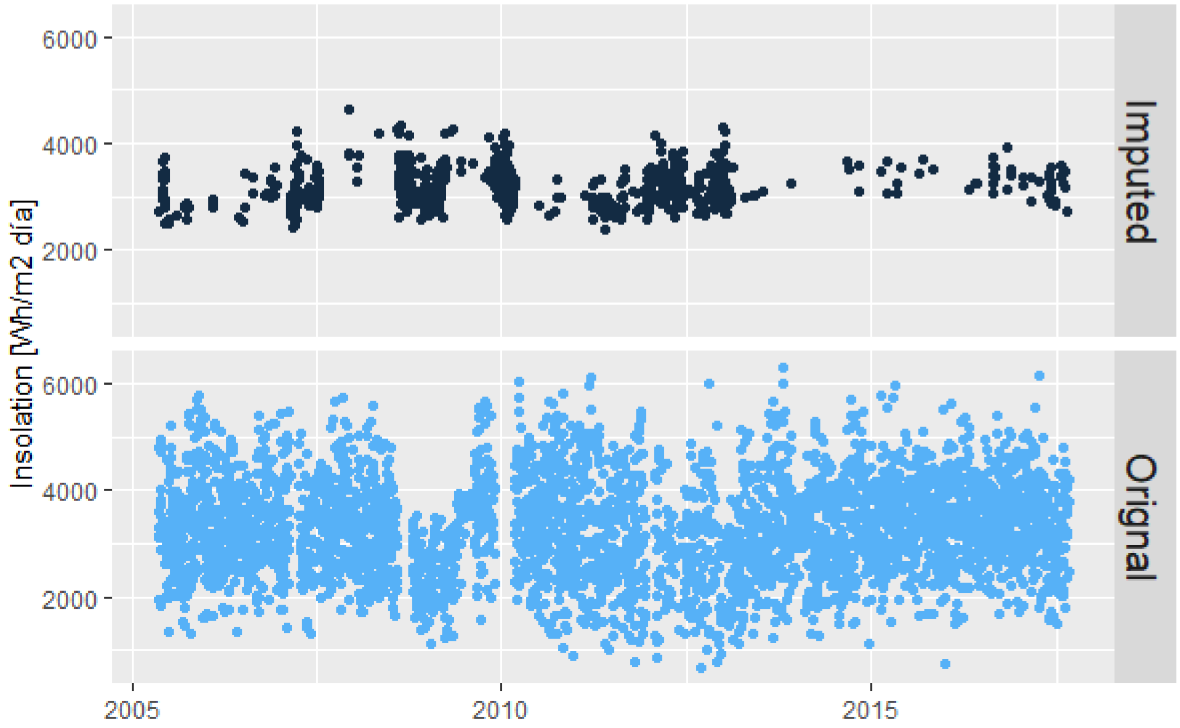

Figure 15. Universidad de Nariño imputation.

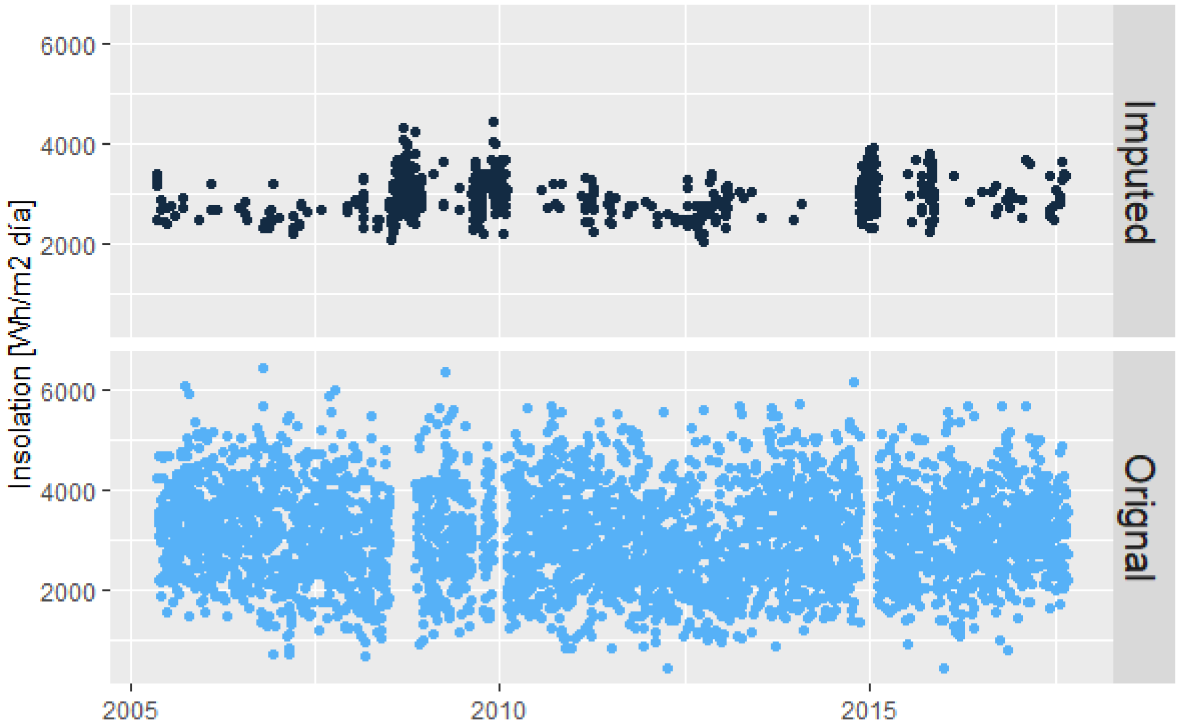

Figure 16. Botana imputation.

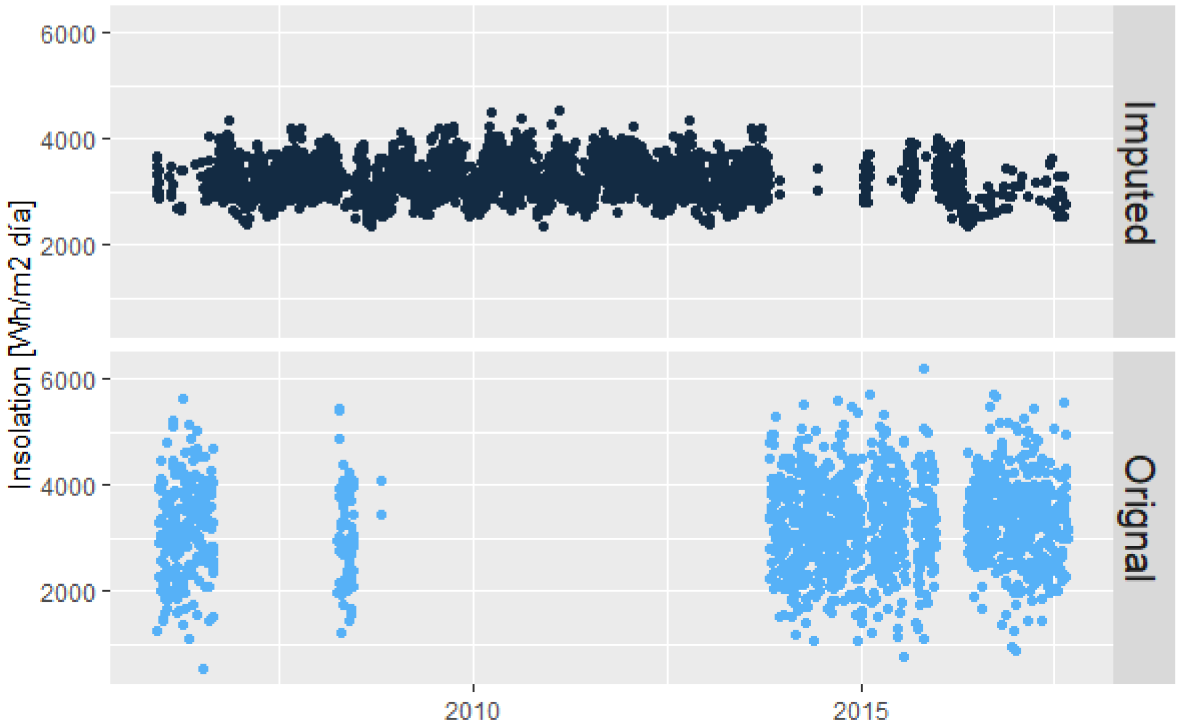

Figure 17. La Josefina imputation.

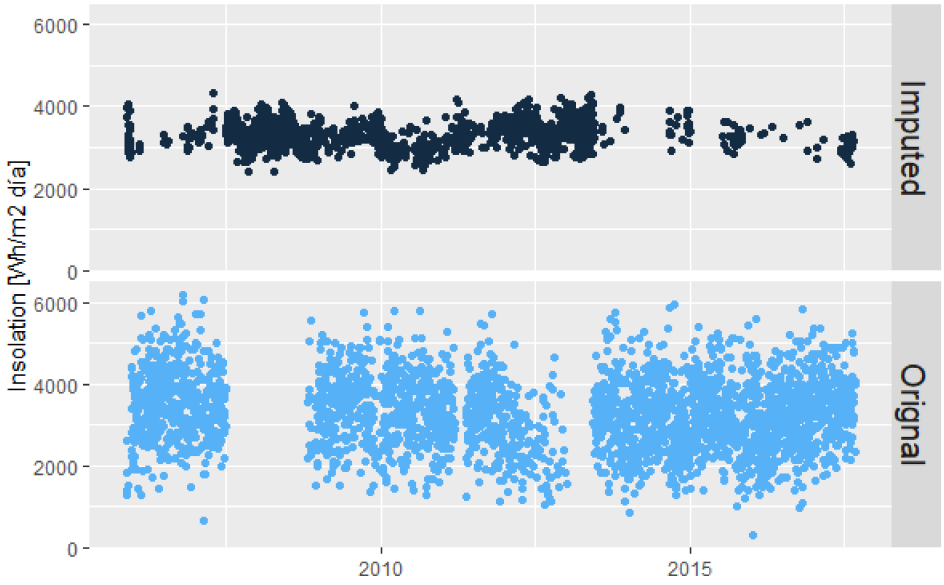

Figure 18. Paraiso imputation.

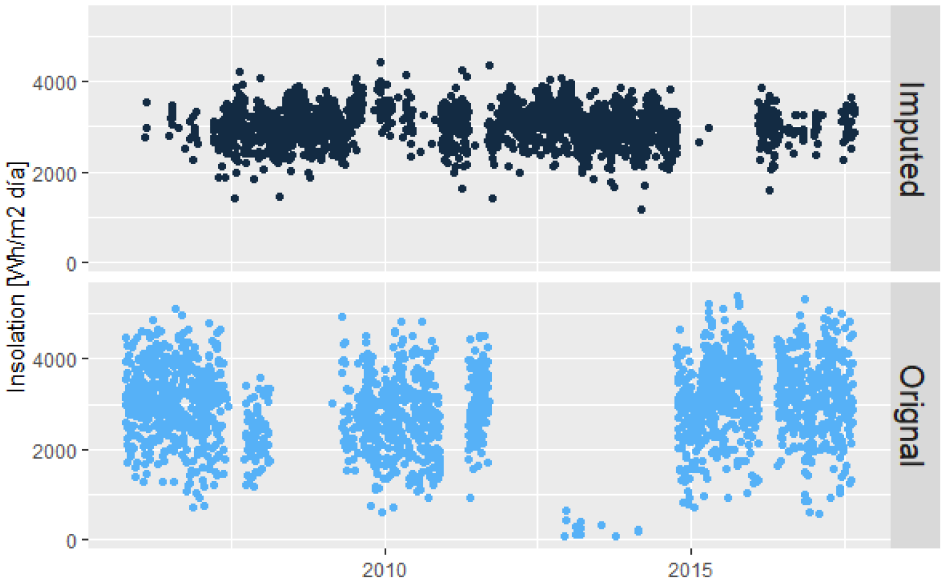

Figure 19. Guapi imputation.

Figure 20 shows the monthly daily average solar insolation for all analyzed AWSs. The AWS located in the Pacific zone presented similar behaviors; namely, they registered a peak in August and September and a lower level in November. *Viento Libre* was the only AWS that recorded values close to 4000 Wh/m^2 day in August. The *Botana* and *Universidad de Nariño* AWSs in the Andean region showed a peak between October and November. The *Cerro Páramo* and *Viento Libre* AWSs exhibited certain energy complementarity because the lower level in the first AWS compensates the higher level of the second AWS.

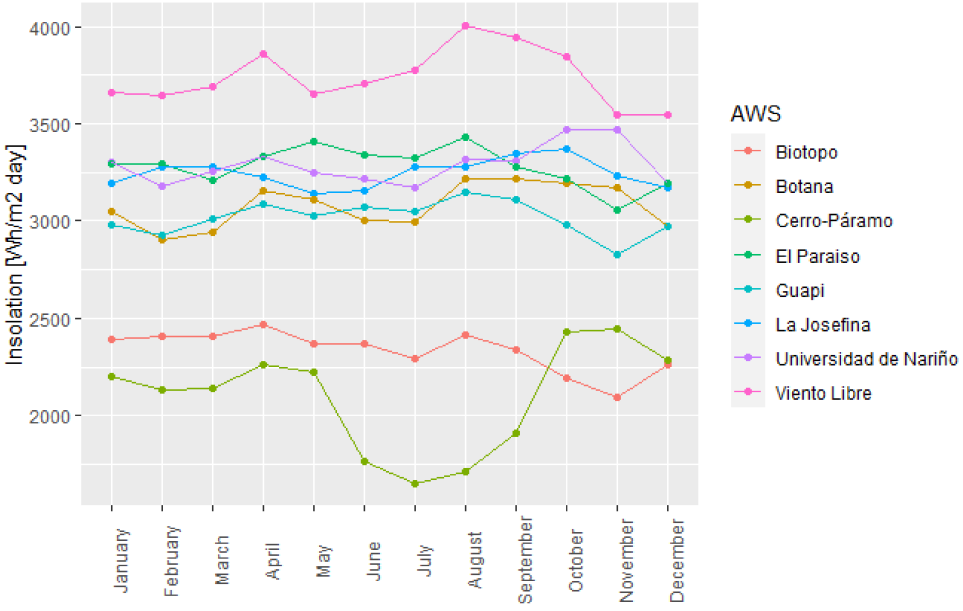

Figure 20. Daily monthly insolation.

4. Conclusions

The validation levels of the global solar irradiance data strongly influenced the results of the empirical variables. The best evidence was that the results showed that 60.89% of the

data overcame the mandatory validation steps. From the dataset overcoming the quality control process, 95.81% corresponded to the days with at least six recordings, and those days constituted the empirical model's calibration information. However, this percentage represented, on average, 33.90% of the total information recorded in the AWS. In addition to this, only 1.26% of the days had complete information. This result indicates that the time series quality is not optimum; therefore, it is necessary to improve and increase maintenance and calibration procedures.

In the state of Nariño, the performance of the AWS is a determining factor when considering the predominance of partial high cloudiness, representing 64.7% of the total measured days. Accordingly, in Nariño, high cloudiness complicates the estimation of solar insolation; therefore, it is fundamental to improve the reliability of measurement systems. Consequently, it is also essential to regularly establish a plan to carry out these procedures at a high-quality level and according to widely accepted standards. Installing more AWSs to increase the sampling points is also recommended.

Regarding temperature measures, only 92.78% of the measures were useful for the empirical calibration and imputation from the total data that overcame the hourly validation steps. Additionally, the days with more records had between 11 and 12 data inputs per day; 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 and altitude. R^2 showed better adjustments between the empirical constants and the altitude in sites above 2500 MASL than at sites below this altitude. This result is consistent with the temperature in tropical zones where there is high humidity at low altitudes and global solar irradiance in high altitudes.

The results from $RMSE$, SD, MAE, U_{95}, and $MAPE$ statistical tests demonstrated that the proposed model exhibited better performance in five of the eight evaluated cases (*Viento Libre*, *Cerro Páramo*, *Universidad de Nariño*, *Botana*, *Josefina*, and *Paraiso*), all of which were in the Andean and Amazon zones at altitudes above 2500 MASL (see Table 12). The proposed model was useful for imputing information in the Andean and Amazon AWSs, with a mean value in RMSE of $1.022,86\ \mathrm{W/m^2 day}$ for the Andean zone and $1.152,72\ \mathrm{W/m^2 day}$ in the Amazon zone. In the Pacific zones' AWSs, Hargreaves and Samani's model was the best option, followed by the model proposed in this study.

The proposed model showed stable performance in the tropical and mountainous environment of Nariño. However, it is necessary to analyze more information coming from other locations with the same characteristics. Most importantly, the number of AWSs and the quality of time series data in tropical and mountainous environments must be increased.

Author Contributions: Conceptualization, L.S.H.-G. and B.J.R.-M.; methodology, L.S.H.-G. and B.J.R.-M.; software, L.S.H.-G.; validation, L.S.H.-G. and B.J.R.-M.; formal analysis, L.S.H.-G. and B.J.R.-M.; investigation, L.S.H.-G. and B.J.R.-M.; resources, L.S.H.-G. and B.J.R.-M.; data curation, L.S.H.-G.; writing—original draft preparation, L.S.H.-G.; writing—review and editing, L.S.H.-G. and B.J.R.-M.; visualization, L.S.H.-G.; supervision, B.J.R.-M.; project administration, B.J.R.-M.; funding acquisition, L.S.H.-G. and B.J.R.-M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors appreciate the support of *Fundación Ceiba*, which funded the doctoral studies of Laura Sofia Hoyos-Gómez, and El Fondo Nacional para el financiamiento de la Ciencia, la Tecnología y la Innovación Fondo Francisco José de Caldas, and the National University of Colombia agreement FP44842-260-2017.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge El Fondo Nacional para el financiamiento de la Ciencia, la Tecnología y la Innovación Fondo Francisco José de Caldas, and the National University of Colombia under the project titled "complementariedad de fuentes no convencionales de energía para la generación de electricidad", and Fundación Ceiba. We also thank engineer Carlos Vargas-Arguelles for his assistance.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

a, b, c	Empirical coefficients
H	Global horizontal insolation
H_0	Daily extraterrestrial solar irradiance
I_0	Hourly extraterrestrial solar irradiance
I_{cs}	Hourly clear-sky global solar irradiance
I_{cst}	Hourly clear-sky global solar irradiance at time t
I_{mt}	Global solar irradiance at time t
I_{sc}	Solar constant
K_t	Clearness index
T_h	Hourly measured temperature
T_{max}	Maximum daily temperature
T_{mean}	Mean daily temperature
T_{min}	Minimum daily temperature
T_R	Ratio between the daily minimum and maximum temperature
ΔT	Difference between the daily maximum and minimum temperature
φ	Latitude
δ	Solar declination
ω_s	Sunset hour angle
D	Julian day
β	Solar altitude
τ	Atmospheric transmittance

References

1. Hassan, G.E.; Youssef, M.E.; Mohamed, Z.E.; Ali, M.A.; Hanafy, A.A. New Temperature-based Models for Predicting Global Solar Radiation. *Appl. Energy* **2016**, *179*, 437–450. [CrossRef]
2. Bakirci, K. Models of solar radiation with hours of bright sunshine: A review. *Renew. Sustain. Energy Rev.* **2009**, *13*, 2580–2588. [CrossRef]
3. Jahani, B.; Dinpashoh, Y.; Raisi, A. Evaluation and development of empirical models for estimating daily solar radiation. *Renew. Sustain. Energy Rev.* **2017**, *73*, 878–891. [CrossRef]
4. Besharat, F.; Dehghan, A.A.; Faghih, A.R. Empirical models for estimating global solar radiation: A review and case study. *Renew. Sustain. Energy Rev.* **2013**, *21*, 798–821. [CrossRef]
5. Benson, R.B.; Paris, M.V.; Sherry, J.E.; Justus, C.G. Estimation of daily and montly direct, diffuse and global solar radiation from sunshine duration measurements. *Sol. Energy* **1984**, *32*, 523–535. [CrossRef]
6. Akinoglu, B. Recent Advances in the Relations between Bright Sunshine Hours and Solar Irradiation. In *Modeling Solar Radiation at the Earth's Surface*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 115–143. [CrossRef]
7. Almorox, J.; Hontoria, C.; Benito, M. Models for obtaining daily global solar radiation with measured air temperature data in Madrid (Spain). *Appl. Energy* **2011**, *88*, 1703–1709. [CrossRef]
8. Fan, J.; Chen, B.; Wu, L.; Zhang, F.; Lu, X.; Xiang, Y. Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions. *Energy* **2018**, *144*, 903–914. [CrossRef]
9. Hargreaves, G.H.; Samani, Z.A. Estimating Potential Evapotranspiration. *J. Irrig. Drain. Div.* **1982**, *108*, 225–230. [CrossRef]
10. Bristow, K.L.; Campbell, G.S. On the relationship between incoming solar radiation and daily maximum and minimim temperature. *Agric. For. Meteorol.* **1984**, *31*, 159–166. [CrossRef]
11. Chen, J.; Liu, H.; Wu, W.; Xie, D. Estimation of monthly solar radiation from measured temperatures using support vector machines—A case study. *Renew. Energy* **2011**, *36*, 413–420. [CrossRef]
12. Li, H.; Cao, F.; Wang, X.; Ma, W. A Temperature-Based Model for Estimating Monthly Average Daily Global Solar Radiation in China. *Sci. World J.* **2014**, *2014*, 128754. [CrossRef]
13. Quansah, E.; Amekudzi, L.K.; Preko, K.; Aryee, J.; Boakye, O.R.; Boli, D.; Salifu, M.R. Empirical Models for Estimating Global Solar Radiation over the Ashanti Region of Ghana. *J. Sol. Energy* **2014**, *2014*, 897970. [CrossRef]

14. Dos Santos, C.M.; De Souza, J.L.; Ferreira, R.A., Jr.; Tiba, C.; de Melo, R.O.; Lyra, G.B.; Teodoro, I.; Lyra, G.B.; Lemes, M.A.M. On modeling global solar irradiation using air temperature for Alagoas State, Northeastern Brazil. *Energy* **2014**, *71*, 388–398. [CrossRef]
15. Rivero, M.; Orozco, S.; Sellschopp, F.S.; Loera-Palomo, R. A new methodology to extend the validity of the Hargreaves-Samani model to estimate global solar radiation in different climates: Case study Mexico. *Renew. Energy* **2017**, *114*, 1340–1352. [CrossRef]
16. Jamil, B.; Akhtar, N. 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. *Renew. Sustain. Energy Rev.* **2017**, *77*, 1326–1342. [CrossRef]
17. Instituto Geográfico Agustín Codazzi—IGAC. *Nariño Características Geográficas*; Imprenta Nacional de Colombia: Bogotà, Colombia, 2014.
18. Martínez, A.G. Nariño: Departamento de Nariño Colombia—Informacion Detallada Nariño Colombia 2018. Available online: https://www.todacolombia.com/departamentos-de-colombia/narino.html (accessed on 2 October 2018).
19. Gobernación de Nariño. Plan participativo de Desarrollo Departamental. *Plan Desarro. Dep. Nariño* **2016**, *255*. [CrossRef]
20. Corponariño. Plan de Gestion Ambiental Regional 2002–2012; San Juan de Pasto, Colombia. 2001. Available online: https://www.google.com.hk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwj3msf58Jv0AhVyk1YBHWhcB_0QFnoECAYQAQ&url=http%3A%2F%2Fcorponarino.gov.co%2Fexpedientes%2Fpgar20022012%2Fpgar2002-2012.pdf&usg=AOvVaw0oNP5AtXgL_y7Cn00TvEic (accessed on 1 November 2021).
21. Estévez, J.; Gavilán, P.; Giráldez, J.V. Guidelines on validation procedures for meteorological data from automatic weather stations. *J. Hydrol.* **2011**, *402*, 144–154. [CrossRef]
22. AENOR. Redes de Estaciones Meteorológicas Automáticas: Directrices Para la Validación de Registros Meteorológicos Procedentes de Redes de Estaciones Automáticas. Validación en Tiempo Real. 2004. Available online: https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0031912 (accessed on 1 November 2021).
23. Kipp, Z. Instruction Manual Pyranometer/Albedometer CM11 e CM14 2000. Available online: https://www.google.com.hk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi-7ovK8pv0AhVFp1YBHYzaBMkQFnoECAUQAQ&url=https%3A%2F%2Fs.campbellsci.com%2Fdocuments%2Fus%2Fmanuals%2Fkippzonen_manual_cmp-series.pdf&usg=AOvVaw0V6zEJQuWa8A1q7zAXlzJW (accessed on 1 November 2021).
24. Şen, Z. *Solar Energy Fundamentals and Modeling Techniques*; Springer: Berlin/Heidelberg, Germany, 2008. [CrossRef]
25. Serrano, A.; Sanchez, G.; Cancillo, M.L. Correcting daytime thermal offset in unventilated pyranometers. *J. Atmos. Ocean. Technol.* **2015**, *32*, 2088–2099. [CrossRef]
26. Herrera-Grimaldi, P.; García-Marín, A.P.; Estévez, J. Multifractal analysis of diurnal temperature range over Southern Spain using validated datasets. *Chaos* **2019**, *29*, 063105. [CrossRef] [PubMed]
27. Paulescu, M. Solar Irradiation via Air Temperature Data. In *Modeling Solar Radiation at the Earth Surface*; Badescu, V., Ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 175–193.
28. Meza, F.; Varas, E. Estimation of mean monthly solar global radiation as a function of temperature. *Agric. For. Meteorol.* **2000**, *100*, 231–241. [CrossRef]
29. Moreno, A.; Gilabert, M.A.; Martínez, B. Mapping daily global solar irradiation over Spain: A comparative study of selected approaches. *Sol. Energy* **2011**, *85*, 2072–2084. [CrossRef]
30. Ul Rehman Tahir, Z.; Hafeez, S.; Asim, M.; Amjad, M.; Farooq, M.; Azhar, M.; Amjad, G.M. Estimation of daily diffuse solar radiation from clearness index, sunshine duration and meteorological parameters for different climatic conditions. *Sustain. Energy Technol. Assess.* **2021**, *47*, 101544. [CrossRef]
31. Mujabar, S.; Chintaginjala Venkateswara, R. Empirical models for estimating the global solar radiation of Jubail Industrial City, the Kingdom of Saudi Arabia. *SN Appl. Sci.* **2021**, *3*, 95. [CrossRef]
32. Samani, Z. Estimating Solar Radiation and Evapotranspiration Using Minimum Climatological Data. *J. Irrig. Drain. Eng.* **2000**, *126*, 265–267. [CrossRef]
33. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration-Guidelines for Computing Crop Water Requirements—FAO Irrigation and Drainage Paper 56*; FAO: Rome, Italy, 1998.
34. Allen, R.G. Self-Calibrating Method for Estimating Solar Radiation From Air Temperature. *J. Hydrol. Eng.* **1997**, *2*, 56–67. [CrossRef]
35. Goodin, D.G.; Hutchinson, J.M.S.; Vanderlip, R.L.; Knapp, M.C.; Goodin, D.G. Estimating Solar Irradiance for Crop Modeling Using Daily Air Temperature Data. *Agroclimatology* **1999**, *91*, 845–851. [CrossRef]
36. Okundamiya, M.S.; Nzeako, A.N. Empirical Model for Estimating Global Solar Radiation on Horizontal Surfaces for Selected Cities in the Six Geopolitical Zones in Nigeria. *J. Control. Sci. Eng.* **2011**, *2011*, 805–812. [CrossRef]
37. Nwokolo, S.C.; Ogbulezie, J.C. A quantitative review and classification of empirical models for predicting global solar radiation in West Africa. *Beni-Suef Univ. J. Basic Appl. Sci.* **2018**, *7*, 367–396. [CrossRef]
38. Agami Reddy, T. *Applied Data Analysis and Modelling for Energy Engineers and Scientists*; Springer: London, UK, 2011. [CrossRef]
39. Moon, S.H.; Kim, Y.H. An improved forecast of precipitation type using correlation-based feature selection and multinomial logistic regression. *Atmos. Res.* **2020**, *240*, 104928. [CrossRef]
40. Kleinbaum, D.G.; Klein, M. *Logistic Regression: A Self-Learning Text*; Springer: Berlin/Heidelberg, Germany, 2010. [CrossRef]
41. Manning, R.L. Logit regressions with continuous dependent variables measured with error. *Appl. Econ. Lett.* **1996**, *3*, 183–184. [CrossRef]

42. Harrenll, F.E. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 13. [CrossRef]
43. Casella, G.; Berger, R.L. *Statistical Inference*, 2nd ed.; Thomson: Toronto, ON, Canada, 2002.
44. Mayer, D.G.; Butler, D.G. Statistical validation. *Ecol. Model.* **1993**, *68*, 21–32. [CrossRef]
45. Gueymard, C.A. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. *Renew. Sustain. Energy Rev.* **2014**, *39*, 1024–1034. [CrossRef]
46. Li, J.; Heap, A.D. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecol. Inform.* **2011**, *6*, 228–241. [CrossRef]
47. Blumthaler, M. Solar Radiation of the High Alps. In *Plants in Alpine Regions Cell Physiology of Adaptation and Survival Strategies*; Lütz, C., Ed.; Springer: New York, NY, USA, 2012; pp. 11–20. [CrossRef]
48. Cabrera, O.; Champutiz, B.; Calderon, A.; Pantoja, A. Landsat and MODIS satellite image processing for solar irradiance estimation in the department of Narino-Colombia. In Proceedings of the 2016 21st Symposium on Signal Processing, Images and Artificial Vision, STSIVA 2016, Bucaramanga, Colombia, 31 August–2 September 2016; pp. 1–6. [CrossRef]