Assessment of Data Sources and Solar Irradiation Satellite Estimation Models for Rural Villages in the Colombian Amazon

Luis E. Ordoñez, Víctor A. Bucheli, and Eduardo F. Caicedo

Abstract— **Despite global efforts to adopt renewable energy, many remote regions still lack reliable electrical services. Addressing this requires a thorough analysis of solar resource data to identify viable solutions for these underserved areas. We evaluate the error in monthly solar radiation data from a satellite image-based Random Forest (satellite RF) model by using data from IDEAM meteorological stations and NASA sources. By rigorously comparing these datasets, we aim to assess the reliability of solar radiation estimation sources in the Amazon region. The results help establish confidence in various data sources, essential for utilizing estimated solar energy data in renewable energy research. We compared the data using the Relative Root Mean Squared Error (Relative RMSE) on a monthly scale. On the one hand, the relative RMSE between NASA and IDEAM ranges from 6.86% to 20.93%. On the other hand, the error between satellite RF model and IDEAM fluctuates between 6.56% and 12.33%. Similarly, the error between satellite RF model and NASA ranges from 4.80% to 15.27%. The findings indicate that the error in NASA data is higher compared to the error in satellite RF model data when benchmarked against IDEAM. Despite the limited number of meteorological stations and a maximum error of 20.93% between the two predictive data sources compared to ground-based observed data, we consider it reliable to use estimated solar radiation data for developing effective renewable energy solutions in remote locations.**

Link to graphical and video abstracts, and to code: https://latamt.ieeer9.org/index.php/transactions/article/view/9174

Index Terms—**Amazon region, Meteorological stations, Photovoltaic Systems, Renewable Energy, Rural villages, Solar Radiation.**

I. Introduction

The growing global emphasis on renewable energy is driven by the need to reduce greenhouse gas emissions and mitigate climate change, prompting organizations worldwide to promote sustainable development [1]. According to the United Nations [2], energy is a critical factor in this effort, with fossil fuel combustion for electricity and heat generation being a major contributor to greenhouse gases.

Despite significant advancements in renewable energy technologies, many remote areas in Colombia, particularly the Non-Interconnected Zones (ZNI), continue to experience limited and unreliable electricity access, as reported in the 2021 Sectoral Report by the Superintendence of Public Utility Services [3]. This lack of access negatively impacts the quality of life, education, productivity, and access to information and communication technologies (ICT) in these regions, as highlighted by Bustos González et al. [4]. In the Amazon region, 13 service providers operated in ZNIs across 31 municipalities, supplying electricity to 396 localities and 105,150 households in 2019. However, the electricity coverage reached only 57.13%, largely due to the region's vast size and challenging geographical conditions.

Accurate monitoring of solar radiation is essential for assessing renewable energy potential in these underserved regions. However, surface-based solar radiation monitoring, particularly using pyranometers, faces significant challenges. While pyranometers are highly accurate in measuring irradiation, they are costly, require regular maintenance, and are sparsely deployed in remote areas like the Putumayo department, where IDEAM's weather station network is limited, and few stations include solar radiation sensors [5]. This scarcity of ground-based data poses significant obstacles to reliable solar energy assessments in the region [6].

To overcome these limitations, researchers have increasingly turned to predictive models based on satellite data, which provide broader spatial coverage but present their own challenges. Satellite data can be influenced by atmospheric conditions such as cloud cover, and models may require calibration to local conditions to improve their accuracy. Despite these difficulties, satellite-based models offer a valuable alternative for estimating solar radiation in regions where surface-based monitoring is inadequate.

This research introduces a novel approach by evaluating solar radiation data in the Colombian Amazon by comparing satellite RF models and NASA's solar radiation estimates with ground-based measurements from IDEAM. The main objective is to rigorously assess the accuracy and reliability of each predictive model individually, given the region's limited availability of ground-based data. Specifically, the study compares these models to assess their practical applicability in solar energy planning for underserved areas. By offering a detailed comparison of multiple predictive models, the research provides crucial insights into their performance in

The associate editor coordinating the review of this manuscript and approving it for publication was Gabriel Pinto (*Corresponding author: Luis Eduardo Ordoñez Palacios*).

L. E. O. Palacios, V. A. B Guerrero, and E. F. C. Bravo are with Universidad del Valle, Cali, Colombia (e-mails: luis.ordonez.palacios@correounivalle.edu.co, victor.bucheli@correounivalle.edu.co, and eduardo.caicedo@correounivalle.edu.co).

real-world applications, helping to inform renewable energy strategies and improve energy access in remote regions.

Our findings provide essential guidance for addressing energy access challenges in areas with sparse data and limited infrastructure, contributing to the broader goal of sustainable energy development in Colombia's Amazon region and other similar underserved areas worldwide. Moreover, this research helps bridge the gap between theoretical predictive models and the practical demands of solar energy planning in regions with constrained data collection capabilities.

This document is structured into sections that include related works, materials and methods, results, discussion, and conclusions.

II. Related Works

This research evaluates predictive sources of solar radiation data compared to ground-based observed data. It also compares data from predictive sources of solar energy data in locations where meteorological stations are not available. In this section, previous works addressing evaluations applied to various energy sources are described.

In this context, the research by Martins et al. [7] validates solar and wind resource assessment models as part of the international SWERA project, aimed at providing a database to promote the integration of renewable energy into the energy matrix of several countries. They describe the methodology used to produce solar maps, based on the analysis of seasonal and annual averages of daily solar irradiation during the period from 1995 to 2002. The comparison between the estimated and measured values of daily global solar irradiation, it is observed that the estimates provided by the BRASIL-SR model align well with the actual data, with a mean bias error (MBE) of 0.2% and a root mean square error (RMSE) of 11.4% of the average global solar irradiation.

Escobar et al. [8] evaluate Chile's solar energy resources, incorporating measurements from a network of ground stations, satellite estimates from the Chile-SR model, and simulations assessing the potential for solar thermal, photovoltaic, and concentrated solar energy. Their comparison of satellite data with ground station data from 2010 to 2014 shows an agreement with a relative RMSE of 8.9%.

Sarazola et al. [9] evaluated the performance of six solar irradiation estimation models in Uruguay, using quality-controlled ground measurements over an extended period. The LCIM model, adapted for the region and based on GOES-East satellite imagery, demonstrates the highest accuracy and spatial consistency, with a root mean squared deviation of 6% and a mean bias of less than 1%.

Ferrari et al. [10] evaluate tools for urban energy planning. They select 17 tools designed to assess energy services, sources, and technologies at the urban or district level. Among them, 6 user-friendly tools were identified, such as energyPRO, HOMER, iHOGA, EnergyPLAN, SIREN, and WebOpt. These applications offer energy calculations and are feasible for widespread use. They provide detailed information on the general overview, functionalities, structure, user interface, required input data, and results generated by these tools.

Riveros-Rosas et al. [11] implement the Heliosat 2 model in Mexico using images from the GOES 13 satellite. The results showed an average annual relative error between the modeled and measured data. When comparing modeled results with meteorological service stations, they observe an average annual relative RMSE of 12% was observed at 4 stations, and an RMSE below 20%, with minimum values between 6% and 8%, at 5 stations.

This work represents an innovative and significant contribution to assessing solar resources in the Amazon region, an area of vital ecological importance with unique challenges in terms of energy access. The research's novelty lies in both its methodological approach and the specific findings for the Amazon. This study addresses a significant gap in knowledge about the potential of solar energy in this region, offering valuable information for sustainable energy planning in one of the most important and challenging ecosystems on the planet.

III. Methodology

This section formulates the questions that drove the research and identifies the sources of information used. We provide a detailed description of the data sets used and the methods we use to compare and evaluate solar resources in remote villages in the Amazon region. In addition, we describe the metrics to evaluate the results, and we present the architecture of the model.

A. Research Questions

The researchers have utilized mathematical, statistical, and predictive methods to analyze the behavior of solar radiation in remote areas. Colombia faces challenges in accurately assessing solar radiation due to a limited number of meteorological observation stations. We address this constraint through a comparative analysis approach. We evaluate the discrepancies between two distinct sources of solar radiation data and the actual measurements recorded by available weather monitoring stations. Additionally, we examine the differences between these two predictive data sources.

This study raises several questions: Q1: How accurate are the estimations from data sources in regions with similar climatic conditions? Q2: Which data source represents the lowest error in solar radiation estimations? The answers to these questions are found in the results and conclusions of this document.

B. Protocol Data

We reliy on three sources of solar energy data: The Colombian Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM), NASA, and the satellite RF algorithm from the work of Ordoñez Palacios et al. [12]. We detail the data observed from the ground in the IDEAM database in [5] in Table I. Additionally, Table II shows three locations in the Putumayo department where IDEAM lacks meteorological observation data.

TABLE I
IDEAM'S SOLAR RADIATION DATA

Station	Municipality	Departament	Latitude	Longitude	Altitude
Florencia	Florencia	Caquetá	1.7330	-75.645027	600
Acueducto Mocoa	Mocoa	Putumayo	1.15733333	-76.651833	650
El Pepino	Mocoa	Putumayo	1.08288888	-76.667111	760

TABLE II
LOCATIONS WITH SATELLITE RF AND NASA DATA

Location	Municipality	Departament	Latitude	Longitude
San José del Guineo	Villagarzón	Putumayo	0.9451371	-76.634792
Santa Rosa de Juanambú reservation	Puerto Caicedo	Putumayo	0.7295021	-76.591350
Alto Lorenzo reservation	Puerto Asís	Putumayo	0.3651866	-76.550081

We use solar energy data from IDEAM covering the period between the years 2011 and 2018. Meanwhile, we obtained monthly averages of solar radiation from NASA based on data spanning a 22-year period between June 1983 and June 2005. We obtained the NASA data through the HOMER Pro energy resource optimization tool [13]. Furthermore, the research by Ordoñez Palacios et al. [12] utilizes satellite images from the years 2012, 2013, and 2014 for hourly solar radiation estimation.

Fig. 1 presents a geographical map of the study sites in the Colombian Amazon region. We selected these sites due to their location in rural areas with limited access to meteorological data, which is highly relevant for studying the feasibility of solar energy generation projects.

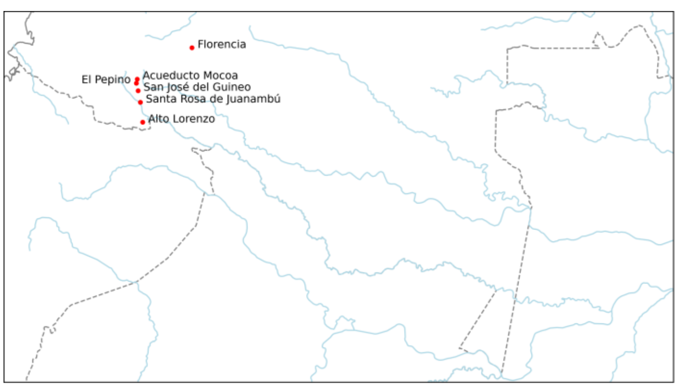

Fig. 1. Study sites in the Colombian Amazon region.

Table III displays the monthly averages of solar radiation obtained from the three sources of information in different locations in the Amazon region.

TABLE III
MONTHLY AVERAGES OF SOLAR ENERGY

Location	Florencia (Florencia - Caquetá)			Acueducto Mocoa (Mocoa - Putumayo)			El Pepino (Mocoa - Putumayo)		
Month/Source	**RF**	**NASA**	**IDEAM**	**RF**	**NASA**	**IDEAM**	**RF**	**NASA**	**IDEAM**
January	3.86	4.24	3.95	3.88	4.38	3.92	4.05	4.38	3.83
February	3.79	4.02	3.85	4.14	4.33	3.50	3.99	4.33	4.02
March	3.76	3.78	3.66	3.63	4.30	3.24	3.68	4.30	4.14
April	3.76	3.78	3.57	3.78	4.27	3.77	3.90	4.27	3.91
May	3.57	3.79	3.91	3.53	4.26	3.79	3.52	4.26	3.82
June	3.44	3.57	3.62	3.31	4.40	3.51	3.37	4.40	4.43
July	3.54	3.55	3.46	3.61	4.38	3.51	3.60	4.38	3.94
August	3.88	3.73	3.70	3.88	4.45	3.49	4.07	4.45	3.95
September	4.14	4.18	4.01	4.23	4.52	3.75	4.26	4.52	3.68
October	4.34	4.30	3.67	4.56	4.50	3.88	4.62	4.50	4.51
November	3.88	4.21	3.81	4.25	4.36	3.70	4.16	4.36	4.02
December	4.22	4.18	4.06	4.28	4.28	3.54	4.04	4.28	3.85
Annual average	**3.85**	**3.94**	**3.77**	**3.93**	**4.37**	**3.63**	**3.94**	**4.37**	**4.01**

We found that the IDEAM database lacks ground-based measurement stations for the locations mentioned in Table II. However, Table IV shows the monthly averages estimated by NASA, obtained through the Homer tool; the monthly averages generated by the satellite RF algorithm developed by Ordoñez Palacios et al. [12]; and the data for the period between 2011 and 2018, provided by the Power | Data Access Viewer project [14], which includes solar and meteorological datasets from NASA's research.

TABLE IV
NASA, POWER AND SATELLITE RF MONTHLY AVERAGES

Location	San José del Guineo (Villagarzón)			Santa Rosa de Juanambú reservation (Puerto Caicedo)			Alto Lorenzo reservation (Puerto Asís)		
Month/Source	**RF**	**NASA**	**POWER**	**RF**	**NASA**	**POWER**	**RF**	**NASA**	**POWER**
January	4.00	4.38	4.67	3.76	3.94	4.67	3.95	3.94	4.67
February	3.60	4.33	4.21	3.71	3.68	4.21	3.75	3.68	4.21
March	3.85	4.30	3.96	3.62	3.50	3.96	3.62	3.50	3.96
April	3.83	4.27	4.13	3.81	3.63	4.13	3.70	3.63	4.13
May	3.55	4.26	3.89	3.50	3.58	3.89	3.47	3.58	3.89
June	3.49	4.40	3.71	3.64	3.38	3.71	3.50	3.38	3.71
July	3.45	4.38	3.67	3.50	3.40	3.67	3.42	3.40	3.67
August	3.91	4.45	4.37	3.90	3.63	4.37	3.90	3.63	4.37
September	4.18	4.52	4.91	4.36	4.09	4.91	4.20	4.09	4.91
October	4.58	4.50	4.77	4.62	4.29	4.77	4.58	4.29	4.77
November	3.88	4.36	4.68	4.28	4.11	4.68	4.05	4.11	4.68
December	4.12	4.28	4.65	4.53	3.93	4.65	4.12	3.93	4.65
Annual average	**3.87**	**4.37**	**4.30**	**3.94**	**3.76**	**4.30**	**3.85**	**3.76**	**4.30**

We calculated the monthly averages of solar radiation from the IDEAM and Power project data sources using data we collected between 2011 and 2018, depending on availability.

However, we find that the monthly data from the satellite RF model and NASA/Homer model correspond to different time periods. While this discrepancy may introduce uncertainty when comparing the different sources, we could not select a consistent period to calculate the monthly averages in these cases.

C. Proposed Model

Currently, various sources of solar energy data complement the information provided by meteorological observation stations. However, it is important to assess the level of confidence in the data estimated by mathematical, statistical, or predictive models. Therefore, we find it relevant to compare the data that these models provide with real data that climate monitoring stations obtain. We used the Relative Root Mean Squared Error (Relative RMSE) as an evaluation metric to calculate the error between the estimated data and the real data by normalizing the RMSE with respect to the actual values. We use it to gain a clearer understanding of the error in relation to the magnitude of the observed values. The result of the relative RMSE is expressed as a percentage, which makes it easier to interpret the error in relation to the actual observed values [7] [8] [9] [11].

We checked the IDEAM database and chose the Amazon region. We aimed to obtain ground-based observed solar radiation data. Then, we compared these data with data provided by satellite RF and NASA. The results allowed us to validate the reliability of the data estimated by sources that predict solar radiation data. We also found it important to analyze the error in the estimated solar energy data across different regions located at various altitudes above sea level. Finally, we focused the study on remote locations in the Putumayo department where the number of IDEAM meteorological stations is scarce.

We averaged the IDEAM and satellite RF data monthly, considering solar radiation values between 6 am and 6 pm. Solar radiation values outside this range tend to be zero. The HOMER Pro tool allows us to import monthly averages from the internet based on geographical coordinates. Additionally, we generated area diagrams that plot with colors the area under a line to indicate the differences between the data provided by the two data sources.

We note that IDEAM's network of meteorological observation stations is limited, and we cannot access real data for the areas listed in Table II. Consequently, we focused the analysis solely on comparing the discrepancies between the satellite RF algorithm's estimations and NASA's data sources.

We performed the calculations on the datasets and constructed the graphs using the Python programming language and the Jupyter Notebook development interface. We have placed the notebooks, and the data used in this work in a GitHub repository [15].

D. Model Architecture

Fig. 2 presents the data flow from the information sources to the evaluation of monthly solar radiation averages. We can use this information to determine which predictive data source has the lowest error compared to the IDEAM data source.

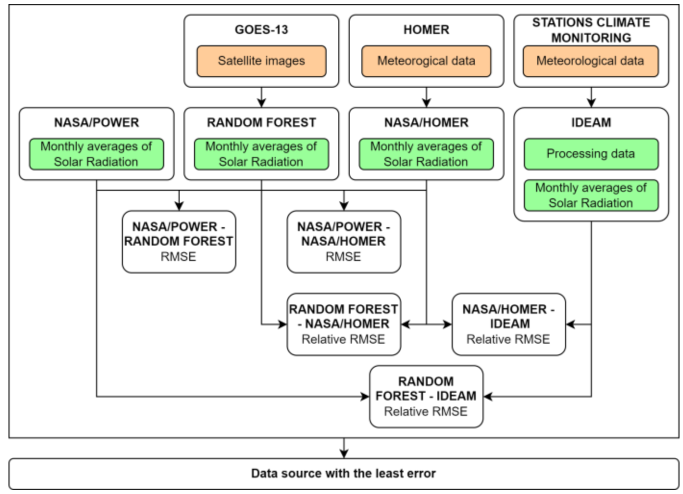

Fig. 2. Solar Radiation Data Evaluation Workflow.

IV. Results

Table V shows the evaluation of two solar energy data sources, using the relative RMSE. We considered three sources of solar radiation data in this study. We made the comparisons considering the Colombian Amazon region and selected three different locations in the departments of Caquetá and Putumayo. We aimed to understand the error in estimated solar radiation data compared to observed data from meteorological stations in places with similar climatic conditions.

TABLE V
SATELLITE RF, NASA AND IDEAM DATA RESULTS

Location	Source 1	Source 2	RMSE	Relative RMSE (%)
Florencia (Caquetá)	RF	NASA	0.181	4.80
	RF	IDEAM	0.248	6.56
	NASA	IDEAM	0.259	6.86
Acueducto Mocoa (Putumayo)	RF	NASA	0.555	15.27
	RF	IDEAM	0.448	12.33
	NASA	IDEAM	0.760	20.93
El Pepino (Mocoa - Putumayo)	RF	NASA	0.523	13.05
	RF	IDEAM	0.409	10.22
	NASA	IDEAM	0.429	10.70

Figs. 3 and 4 visually present the area diagrams that explain how the monthly averages of solar energy behave in the Amazon region. Fig. 3 shows that the relative RMSE error between the data estimated by NASA and the satellite RF model is 4.8%, indicating that the data are similar. In contrast, the satellite RF and NASA data show errors of 6.56% and 6.86%, respectively, compared to IDEAM.

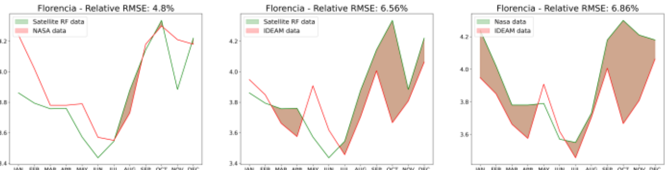

Fig. 3. Results of the comparison between Satellite RF and NASA data, Satellite RF and IDEAM data, and NASA and IDEAM data at the Florencia location.

Fig. 4 demonstrates that the relative RMSE between the data estimated by NASA and satellite RF is 15.27% and 13.05% for Acueducto Mocoa and El Pepino respectively. Likewise, the error of satellite RF and NASA data compared to real data from IDEAM is 12.33% and 20.93% for Acueducto. For El Pepino, the error is 10.22% and 10.70%.

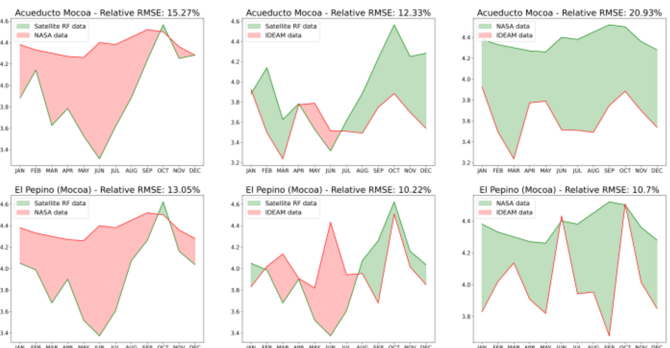

Fig. 4. Results of the comparison between Satellite RF and NASA data, Satellite RF and IDEAM data, and NASA and IDEAM data at the locations of Mocoa and El Pepino.

The Amazon region has a humid equatorial climate with high temperatures, high relative humidity and abundant rainfall [16]. This applies to departments in the Amazon region, such as Caquetá and Putumayo. The relative RMSE between the two predictive sources of solar energy data, NASA and satellite RF, is 4.8%, 15.27% and 13.05% in Florencia, the Acueducto Mocoa, and El Pepino in Putumayo. Considering the low error, we can infer that the estimations from data sources in places with similar climatic conditions are quite accurate. In fact, the error compared to the ground-observed data source (IDEAM) is below 15.27%, except for the 20.93% error between IDEAM and NASA data for Acueducto Mocoa.

Table VI presents the results obtained from the evaluation of solar energy data between NASA and satellite RF. We took data from places located in the municipalities of Villagarzón, Puerto Caicedo, and Puerto Asís in the department of Putumayo. IDEAM does not have solar energy monitoring stations in these locations.

Fig. 5 shows that the RMSE between the data estimated by NASA and satellite RF is 0.574, 0.260 and 0.147 for Villagarzón, Puerto Caicedo, and Puerto Asís, respectively. This highlights the importance of predictive data sources that complement the limited number of weather stations. It also supports the theory that predictive data sources provide accurate data in places with similar weather conditions. The 0.574 error does not greatly exceed the 0.555 value we obtained from comparing NASA and satellite RF data at the Acueducto Mocoa meteorological station. We could not calculate the relative RMSE in this case due to the lack of real data from meteorological stations in the described locations.

TABLE VI
SATELLITE RF, NASA AND POWER DATA RESULTS

Location	Source 1	Source 2	RMSE
	RF	NASA	0.574
San José del Guineo (Villagarzón)	RF	POWER	0.193
	NASA	POWER	0.660
	RF	NASA	0.260
Santa Rosa de Juanambú reservation (Puerto Caicedo)	RF	POWER	0.260
	NASA	POWER	0.000
	RF	NASA	0.147
Alto Lorenzo reservation (Puerto Asís)	RF	POWER	0.147
	NASA	POWER	0.000

On the other hand, we found that data obtained through the Homer tool and provided by NASA are identical for Puerto Caicedo and Puerto Asís to the data from the Power project, also by NASA. This makes sense because both datasets come from research conducted by the same institution, though the error is greater compared to Villagarzón, where the RMSE is 0.660. Additionally, the graph shows that the data from the Power project resembles those obtained through the satellite RF algorithm, with RMSE values of 0.192, 0.260, and 0.147, respectively.

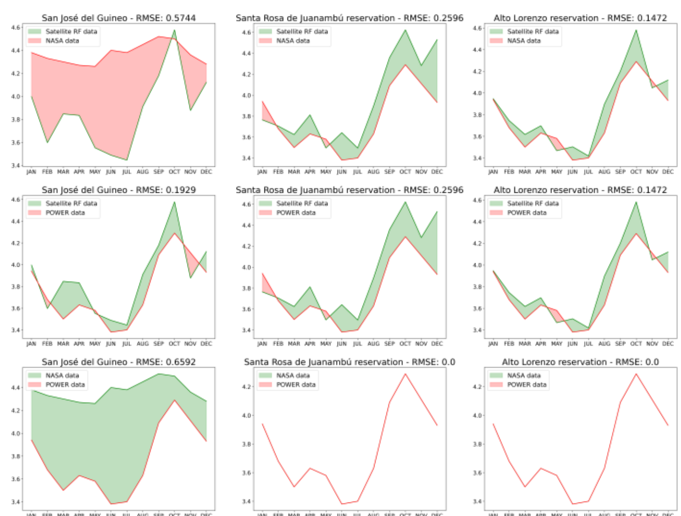

Fig 5. Results of the comparison between Satellite RF and NASA data, Satellite RF and POWER data, and NASA and POWER data at the locations of San José del Guineo, Santa Rosa de Juanambú and Alto Lorenzo.

IDEAM data shows that Putumayo has only two solar radiation measurement stations in Mocoa, confirming the lack of sufficient ground-based solar energy observation instruments. Therefore, we consider it crucial to use other predictive data sources to make the most of the solar resource.

Figs. 3, 4, and 5 compare the performance of two predictive data sources using area diagrams that show how solar radiation estimates vary over time. These visualizations highlight discrepancies between the data, with the areas representing the magnitude of differences in monthly estimations. The diagrams reflect the relative RMSE values, offering a clearer view of the error magnitude between predictive models and ground-truth data. For instance, larger areas indicate higher error margins, which matches the higher relative RMSE values. These figures visually correlate with the numerical results, making it easier to understand the comparative accuracy of each data source in this study.

V. Discussion

Although various entities monitor climate behavior, insufficient measurement station coverage fails to cover entire territories. Consequently, many researchers have developed predictive models to complement existing meteorological station data. In this context, we need to assess the accuracy of these models by comparing them with real ground observations.

This research extends the work of Ordoñez Palacios et al. [12], who evaluated the performance of seven solar radiation predictive models using four metrics. The satellite RF algorithm, which performed best based on satellite imagery, achieved an R^2 of 0.82 and an RMSE of 107.05 W/m^2 on an hourly basis. In this study, we compare satellite RF estimates with another predictive source, NASA, and data from IDEAM meteorological stations.

This study aims to understand solar radiation behavior in remote areas of Colombia and assess the accuracy of predictive data sources to facilitate solar energy generation projects, thereby addressing the issue of electricity access. This, in turn, promotes community development and reduces poverty, which is why we chose the Amazon region. We found that IDEAM lacks sufficient meteorological stations to assess the region's energy potential. Therefore, we used data from satellite RF, NASA (via the Homer tool), and the Power project to compare monthly solar energy averages in three vulnerable Amazon locations, as shown in Table VI.

The metrics used reveal significant discrepancies among the relative RMSE values of the three data sources. We observe the highest relative error between NASA estimates and IDEAM data, reaching 20.93%, which highlights the limitations of satellite data granularity in remote regions. In contrast, satellite RF shows a maximum error of 12.33% compared to IDEAM, indicating a better ability to adjust values to local conditions, possibly due to its focus on additional variables.

When comparing these results with previous studies in regions with different climates, such as the work of Escobar et al. [8] in Chile and Sarazola et al. [9] in Uruguay, we note that model accuracy varies significantly depending on the geographical context. While we achieved a relative RMSE of 8.9% on an hourly basis in Chile and 6% in Uruguay at an hourly scale, the higher values obtained in the Colombian Amazon are reasonable, given the challenge of working in areas with limited meteorological station networks and extreme climatic variations.

Although this research centers on the Colombian Amazon, we can generalize the insights gained to other regions with limited meteorological station coverage or similar climatic conditions. The observed discrepancies in predictive data sources, such as the differences between satellite RF and NASA, highlight the need for region-specific evaluations of predictive models. In areas with sparse ground-based observations, such as many parts of Latin America, Africa, or Southeast Asia, we can apply the insights from this research to improve the accuracy and reliability of solar radiation estimations. This, in turn, can facilitate the development of renewable energy projects in remote or underserved communities. Therefore, expanding predictive model validation to these areas could improve energy access and sustainability efforts on a broader scale.

The results of this study highlight the importance of continuing to evaluate and improve predictive models in hard-to-reach regions. The satellite RF model performs better compared to NASA, suggesting its potential for future applications in designing microgrids and solar energy projects in the Amazon. However, we need to continue refining methodologies to reduce discrepancies and improve predictive data reliability.

VI. Conclusions

We compare the results from satellite RF and NASA data sources with ground-observed data from IDEAM. We also evaluate the error between the two predictive data sources. Table V shows that the relative RMSE between NASA's monthly data and IDEAM's data ranges from 6.86% to 20.93%. The error between satellite RF and IDEAM fluctuates from 6.56% to 12.33%. Similarly, the error between satellite RF and NASA ranges from 4.80% to 15.27%.

The results show that the highest error between NASA and IDEAM data is 20.93%, while between satellite RF and IDEAM it is 12.33%. The lowest errors were 6.86% and 6.56%, respectively. For the comparison between satellite RF and NASA estimations, the maximum error is 15.27%, indicating that satellite RF estimates lower solar radiation values than NASA. NASA's error is higher compared to IDEAM, suggesting that satellite RF is more accurate. Therefore, satellite RF is better suited for estimating solar potential and designing microgrids in the Amazon region.

This study highlights the practical relevance of using predictive data sources, particularly in regions with limited meteorological stations, such as the Colombian Amazon. The relatively low error rates, especially the 12.33% observed between satellite RF and IDEAM, suggest that we can reliably use predictive models for solar energy planning in remote areas. This is essential for designing microgrids and solar power systems that promote sustainable development and improve energy access in isolated communities. Additionally, the observed discrepancies, such as the 20.93% error between NASA and IDEAM, emphasize the need to refine these models to reduce uncertainty and ensure more precise estimations in areas without ground stations.

Given the small number of meteorological stations and the maximum error of 20.93% between the predictive data sources (satellite RF and NASA) compared to IDEAM, we find it reliable to use estimated solar radiation data in areas without measuring stations.

When comparing data from areas without IDEAM stations, we find that NASA's data through the Homer tool are identical to those from the Power Project in two of the three locations in the Colombian Amazon. We also observe that the data granularity from both sources is not optimal, as the Power Project data are identical across the three selected areas, despite their location in different municipalities.

For future work, we recommend incorporating additional sources of predictive data to consolidate this research and enhance confidence in using estimated data for photovoltaic system studies. We suggest employing neural network models, such as generative adversarial networks and transformers, which are effective in processing sequences and time series. Additionally, we advise implementing transfer learning techniques, using pre-trained models on similar datasets, to improve estimation accuracy in future studies.

REFERENCES

[1] Naciones Unidas, «Apoyar el desarrollo sostenible y la acción climática», United Nations. Accedido: 28 de junio de 2023. [En línea]. Disponible en: https://www.un.org/es/our-work/support-sustainable-development-and-climate-action

[2] Naciones Unidas, «Energías renovables: energías para un futuro más seguro», United Nations. Accedido: 28 de junio de 2023. [En línea]. Disponible en: https://www.un.org/es/climatechange/raising-ambition/renewable-energy

[3] Superservicios, «Informe Sectorial de la Prestación del Servicio de Energía Eléctrica 2021». 2021. [En línea]. Disponible en: https://www.superservicios.gov.co/sites/default/files/inline-files/informe_sectorial_zni_2021%20%281%29.pdf

[4] J. F. Bustos González, A. L. Sepúlveda, y K. T. Aponte, «Zonas no interconectadas eléctricamente en Colombia: problemas y perspectiva». Accedido: 28 de junio de 2023. [En línea]. Disponible en: https://fce.unal.edu.co/centro-editorial/docs/econografos-escuela-economia/65-zonas-no-interconectadas-electricamente-en-colombia-problemas-y-perspectiva

[5] IDEAM, «Consulta y Descarga de Datos Hidrometeorológicos». Accedido: 9 de junio de 2021. [En línea]. Disponible en: http://dhime.ideam.gov.co/atencionciudadano/

[6] V. L. L. Jiménez, «Propuesta Metodológica para el Rediseño de una Red Meteorológica en un Sector de la Región Andina Colombiana», Publicaciones E Investig., vol. 8, pp. 55-76, mar. 2014, doi: 10.22490/25394088.1281.

[7] F. R. Martins, E. B. Pereira, y S. L. Abreu, «Satellite-derived solar resource maps for Brazil under SWERA project», Sol. Energy, vol. 81, n.o 4, pp. 517-528, abr. 2007, doi: 10.1016/j.solener.2006.07.009.

[8] R. A. Escobar et al., «Estimating the potential for solar energy utilization in Chile by satellite-derived data and ground station measurements», Sol. Energy, vol. 121, pp. 139-151, nov. 2015, doi: 10.1016/j.solener.2015.08.034.

[9] I. Sarazola, A. Laguarda, J. C. Ceballos, y R. Alonso-Suárez, «Benchmarking of modeled solar irradiation data in Uruguay at a daily time scale», IEEE Lat. Am. Trans., vol. 21, n.o 9, Art. n.o 9, ago. 2023.

[10] S. Ferrari, F. Zagarella, P. Caputo, y M. Bonomolo, «Assessment of tools for urban energy planning», Energy, vol. 176, pp. 544-551, jun. 2019, doi: 10.1016/j.energy.2019.04.054.

[11] D. Riveros-Rosas, María. E. Sánchez-Diaz, V. M. Velasco-Herrera, A. E. González-Cabrera, M. Valdés-Barrón, y L. F. Zarzalejo, «Implementation of the heliosat 2 model in Mexico from GOES 13 satellités images», AIP Conf. Proc., vol. 2445, n.o 1, p. 150004, may 2022, doi: 10.1063/5.0085648.

[12] L. E. Ordoñez Palacios, V. Bucheli Guerrero, y H. Ordoñez, «Machine Learning for Solar Resource Assessment Using Satellite Images», Energies, vol. 15, n.o 11, Art. n.o 11, ene. 2022, doi: 10.3390/en15113985.

[13] HOMER, «HOMER Pro - Microgrid Software for Designing Optimized Hybrid Microgrids». Accedido: 9 de junio de 2021. [En línea]. Disponible en: https://www.homerenergy.com/products/pro/index.html

[14] NASA, «POWER | Data Access Viewer». Accedido: 21 de marzo de 2023. [En línea]. Disponible en: https://power.larc.nasa.gov/data-access-viewer/

[15] L. E. Ordóñez Palacios, «Evaluación de datos del recurso solar». Accedido: 29 de junio de 2023. [En línea]. Disponible en: https://github.com/luise-phd/evalDatosRecursoSolar

[16] DatosMundial, «Clima: Amazonía en Colombia», DatosMundial.com. Accedido: 21 de junio de 2023. [En línea]. Disponible en: https://www.datosmundial.com/america/colombia/clima-amazonia.php

Luis Ordoñez Palacios received the title of Engineer in Systems Engineering from the Corporación Universitaria Autonóma del Cauca in 2005. He received his Specialization in Project Management in 2017 at the Universidad del Tolima and received his doctorate in Engineering in 2024 at the Universidad del Valle, Cali, Colombia. He is a Professor at the Faculty of Engineering at the Escuela de Ingenieria de Sistemas y Computación of the Univalle Research Group on Artificial Intelligence. His current research interests are in the field of Artificial Intelligence and renewable energy.

Víctor Bucheli Guerrero received his degree in Systems Engineering, followed by a master's degree in systems Engineering and Computation, and a Ph.D. in Engineering (2003-2013). He is currently a full professor at the School of Computer Systems at Universidad del Valle in Cali, Colombia. His research interests include artificial intelligence, network science, and knowledge management. His research involves developing and applying complex network and machine learning models. He has authored or co-authored over 30 publications in international journals and conferences and holds a senior research rank awarded by MinCiencias. He also leads the Artificial Intelligence Group at Universidad del Valle.

Eduardo Caicedo Bravo received the degree of Engineer in Electrical Engineering from the Universidad del Valle in 1984. He received his M.Sc. in 1993 and his PhD in industrial computer in 1996 at the Universidad Politécnica de Madrid. He is a Professor of the Electrical and Electronic Engineering School at the Perception and Intelligent Systems Group, Universidad del Valle. His current research interests are in the field of computer intelligence and smart grids.