# Accurate estimation of solar PV power plant capacity factors in Uruguay through detailed quality control and satellite gap filling

Sandino Rehermann Cabrera
*Instituto Tecnológico Regional Centro-Sur (ITR)*
*Universidad Tecnológica (UTEC)*
Durazno, Uruguay
sandino.rehermann@estudiantes.utec.edu.uy

Vívian Teixeira-Branco
*Instituto Tecnológico Regional Centro-Sur (ITR)*
*Universidad Tecnológica (UTEC)*
Durazno, Uruguay
vivian.teixeira@utec.edu.uy

João Victor Furtado Frazão de Medeiros
*Centro de Energias Renováveis (CER)*
*Universidade Federal de Pernambuco (UFPE)*
Recife, Brasil
joao.furtado@ufpe.br

Rodrigo Alonso-Suárez
*Laboratorio de Energía Solar (LES)*
*Universidad de la República (UdelaR)*
Salto, Uruguay
r.alonso.suarez@gmail.com

***Abstract*—This study implements a methodology to produce accurate, gap-free time series of solar irradiance and PV generation data for a large photovoltaic (PV) power plant in Uruguay. Addressing the challenge of data gaps in solar energy monitoring, the work employs quality control techniques, satellite-based solar irradiance estimation, solar radiation and PV models, and gap-filling procedures to generate a comprehensive five-year dataset at 10-minute intervals. The resulting dataset allows detailed analysis of solar resource availability, PV production and capacity factors. The study finds an average capacity factor of $22.4\%$ over the five-year period, with monthly variations ranging from $14.1\%$ to $28.1\%$. This work provides the first precise assessment of PV plant capacity factors in Uruguay, providing valuable insights for grid management and future solar energy investments.**

***Index Terms*—PV, capacity factors, data quality control, solar satellite estimates, gap filling**

## I. Introduction

The accurate estimation of solar photovoltaic (PV) power generation and capacity factors is a critical aspect for the optimization of investment strategies in the renewable energy sector. The capacity factor determines the efficacy of PV systems by relating the actual generation to the maximum possible generation when the systems operate continuously at rated capacity. Information on capacity factors is needed to operate, adapt or plan the electricity infrastructure [1].

In the solar energy sector, the presence of data gaps due to communication failures or discarded data from quality control procedures poses a significant challenge to the generation of complete and reliable time series data. Such datasets are essential for detailed analysis of PV plant operational indicators over time and can be used to improve solar PV generation assessment tools. Missing data can lead to incorrect assessments and misguided decisions, highlighting the need to implement robust quality control and gap-filling techniques. Various methods, including time series interpolation, statistical filling and machine learning techniques, can be used to estimate missing values [2], [3]; other techniques use satellite data to fill the gaps in the time series [4]. Of these options, satellite gap-filling is the only one that uses the actual cloud cover over the site, resulting in more accurate estimates of solar irradiance and PV generation to replace missing data.

This work implements a methodology to generate a multi-year gap-free solar irradiance and PV generation time series for one of the largest PV power plants in Uruguay. The techniques used are based on solar irradiance estimation from satellite imagery, data site adaptation, PV generation models and interpolation techniques. The starting point of the analysis is the data recorded at the PV plant, which is quality controlled, filtered and gap filled. The uncertainty introduced by the gap-filling models is quantified and compared with other studies. The generated time series allow comparison with reference studies in Uruguay [5]–[9] and enable the first accurate assessment of PV generation and associated capacity factors in Uruguay. Both analyses provide key data for grid decisions and the promotion of new PV investments.

## II. Data

### A. PV plant measured data

The environmental and operational data of the large-scale PV plants installed in Uruguay are public and available on the ADME[1] website. The PV plant known as "La Jacinta", located in the northwest of Uruguay (latitude $-31.43°$S and longitude $-57.91°$W), is considered for this study as it is one of the largest PV plants in the country. It is a fixed panel PV plant with a DC capacity of 65 MWp and a nominal capacity of 50 MW. The variables used in this study are the Global Horizontal Irradiance (GHI), the Global Tilted Irradiance (GTI) and the PV Power Production (PPV) on a 10-minute time scale. The

[1]Uruguay's Electricity Market Administrator, https://adme.com.uy/.

period from 2018 to 2022 (five full years) was chosen due to the availability of 10-minute satellite imagery, the same time base as the data recorded at the PV plant.

### B. Solar satellite data

Geostationary satellite imagery can be used to accurately estimate hourly and sub-hourly solar irradiance [10]–[12]. The 10-minute GHI at the site was estimated using the satellite CIM-McClear model [13], which employs the McClear clear sky model [14], a regionally adjusted cloud index parameterization, and a locally adjusted satellite ground reflectance model. The McClear model estimates ($\text{GHI}_{\text{csk}}$) were downloaded from the SoDa platform (http://www.soda-pro.com) and the satellite imagery (GOES-16 visible channel #2) was provided by the Solar Energy Laboratory (LES, http://les.edu.uy/). The satellite all-sky GHI estimates were produced for this work. The McClear estimates have also been used to calculate the clear sky index, $k_c = \text{GHI}/\text{GHI}_{\text{csk}}$, to be used for the final interpolation step of missing data.

## III. Methodology

This work involves a detailed analysis of the GHI, GTI and PPV time series. Quality control techniques are used for each variable and for groups of them, as they have relationships that should be respected to some extent.

The three variables were first visually inspected to identify clearly anomalous data. GHI measurements that did not pass the Baseline Surface Radiation Network (BSRN) filters [15] or had large anomalous deviations from solar satellite estimates were removed. A performance assessment of the satellite data was conducted to determine the maximum allowable deviation and compared with other studies using the same technique in the region. PPV samples with power generation limited by operational constraints imposed by the system operator were removed, as they affect the relationship between PV power output and solar irradiance magnitudes. In this sense, the work aims to estimate PV production and capacity factors in Uruguay without operational restrictions (losses are then attributed to O&M interventions or inefficiencies in grid operation). Additional filters have been applied in the GHI vs. GTI and GTI vs. PPV spaces to detect clear outliers from the general relationship between variables.

The number and length of GHI data gaps after filtering are identified and filled with GHI site-adapted satellite estimates. The availability of satellite estimates determines how many GHI data gaps remain to be filled by interpolation. In such cases, a linear interpolation based on the clear sky index is used. The previous gap filling (satellite data and interpolation) allows the 10-minute GHI time series to be completed within the five-year period. Very little interpolation was required (less than $0.2\%$ of the data), and at most five consecutive 10-minute samples had to be interpolated, with only two of these events. GTI gaps are filled with GTI data estimated from GHI using the transposition model known as HDKR [16]. Transposition models estimate the GTI with knowledge of the direct and diffuse components of solar radiation. These components were estimated from the GHI time series using the univariate empirical separation model of Ruiz-Arias et al. [17], with coefficients adapted to northwestern Uruguay [18]. Finally, the PV power gaps were filled with PPV data estimated from GTI and solar elevation ($\alpha_s$) and azimuth ($\gamma_s$) angles (in deg.), using the multivariate linear regression model shown in Eq. (1) trained with the non-filled data. As the GHI dataset is complete after gap filling, the GTI and PPV datasets are also complete after this procedure.

$$\begin{aligned}\text{PPV}_{\text{est}} = \quad & a_1\ \text{GTI} + a_2\ \text{GTI}^2 + a_3\ \text{GTI}^3 + a_4\ \text{GTI}^4 \\ & \ldots + b_1\ \alpha_s + b_2\ \alpha_s^2 + c_1\ \gamma_s + c_2\ \gamma_s^2\end{aligned} \tag{1}$$

The completed data sets are used to perform a detailed analysis of solar resource availability, PV production and PV capacity factor. Annual and monthly values are obtained. The capacity factor (CF) is the ratio of the annual/monthly generation ($E_t$) divided by the maximum generation of the system in continuous operation [19]. The latter is calculated as the contractual nominal power ($P_{\text{nom}}$) multiplied by the time considered ($t$, month or year), as shown in the bottom part of Eq. (2). This metric provides an important insight into the actual performance of the system compared to its maximum possible performance.

$$\text{CF} = \frac{E_t}{P_{\text{nom}} \times t} \tag{2}$$

The performance of the models used to calculate the GTI and PPV from the GHI values and used for gap filling is also evaluated. The Mean Bias Deviation (MBD) and the Root Mean Square Deviation (RMSD) metrics are used to quantify the models performance. Their relative values are expressed as a percentage of the average of the measurements.

## IV. Results

### A. Models performance assessment

Gaps are filled using GHI satellite estimates (CIM-McClear model), the separation (RA1) and transposition (HDKR) models and the PPV regression model. GHI can only be filled using GHI satellite data and interpolation. The filtered GHI data represent $8.8\%$ of the total daylight samples, of which about $8.6\%$ are filled with satellite data. Measured GTI data can be filled with GTI estimated from measured or estimated GHI, and PPV can be filled with PPV obtained from measured GTI or estimated GTI from GHI. In all cases, where applicable, the measurements are preferred as the starting point for gap filling. The filtered GTI and PPV data represent $3.3\%$ and $11.0\%$ of the daylight samples respectively. Therefore, the models used to estimate the GHI, GTI and PPV need to be evaluated.

Table I shows the performance indicators of the satellite GHI and the RA1+HDRK GTI estimation. The first row shows the GHI assessment of the CIM-McClear model without any site adaptation post-processing (CIM-wa). It can be seen that this model slightly overestimates the GHI, with an rMBD of $4.9\%$. Therefore, a site adaptation procedure [20] was applied, using the linear regression coefficients to correct for the bias in the satellite estimates. The metrics for the site corrected satellite GHI estimates

(CIM-sa) are shown in the second row of Table I. As expected for this type of site adjustment procedure, the rMBD decreases from $+4.9\%$ to $+0.1\%$[2], significantly reducing the bias of the satellite estimates. For the relative rRMSD, the fitted model shows a slight reduction from $16.0\%$ to $15.2\%$. The values found are similar to other studies in the region, where values of $15.6\%$ and $17.1\%$ were reported in Refs. [21] and [22], respectively. Fig. 1 shows the scatter plots of the original and site-adapted GHI satellite estimates compared to the ground measurements. The left plot shows that the original satellite model overestimates, especially for low GHI values. The right plot shows the site-adapted satellite data, where the regression fit is practically the same as the $x = y$ line.

The last row of Table I shows the performance of the RA1+HDKR GTI estimation model. A low bias of $+0.5\%$ and an rRMSD of $5.5\%$ are found. This is comparable to other studies carried out in Uruguay using the same model, which report an rRMSD of around $4.1\%$ [9]. It should be noted that Ref. [9] used reference measurements from a dedicated solar radiation measurement network with class A sensors, calibrated every two years and regularly maintained. Table II shows the performance indicators of the PV estimation model. It is observed that the model is unbiased with a relatively low rRMSD of $6.8\%$. Compared to other studies, Ref. [23] found an rRMSD of $11.4\%$ for daily power estimation using TRNSYS and the isotropic tilted plane transposition model [24]. Ref. [25] used different physical approaches to estimate PV power in the Brazilian semi-arid region and found rRMSD values ranging from $12.5\%$ to $16.9\%$.

TABLE I
PERFORMANCE METRICS FOR GHI SOLAR SATELLITE ESTIMATES AND GTI ESTIMATION FROM GHI AT A 10-MINUTE TIME SCALE.

| | **MBD** W/m$^2$ | **rMBD** % | **RMSD** W/m$^2$ | **rRMSD** % | **ave. ref.** W/m$^2$ |
|---|---|---|---|---|---|
| GHI CIM-wa | $+21.4$ | $+4.9$ | 69.0 | 16.0 | 431.6 |
| GHI CIM-sa | $+0.5$ | $+0.1$ | 65.4 | 15.2 | 431.6 |
| GTI RA1+HDKR | $+2.3$ | $+0.5$ | 24.8 | 5.5 | 451.5 |

TABLE II
PERFORMANCE INDICATORS FOR PPV ESTIMATES AT 10-MINUTE SCALE. THE REFERENCE PPV VALUE FOR NORMALIZATION IS 23.0 MW.

| | **MBD** MW | **rMBD** % | **RMSD** MW | **rRMSD** % |
|---|---|---|---|---|
| PPV regression model | $\simeq 0.0$ | $\simeq 0.0$ | 1.5 | 6.8 |

Fig. 2 shows the scatter plots for GTI (left) and PPV (right) estimation. The spread in these scatter plots is smaller than in Fig. 1, which is consistent with the lower rRMSD of the PPV and GTI estimates compared to the GHI satellite estimate. A slightly higher scatter is observed for PPV than for GTI, which can be explained by PV model limitations with module temperature or soiling effects. The PPV scatter plot is slightly asymmetric, with more samples above the red line compensated by a few larger outliers below the red line. This is confirmed by Fig. 3, which shows the error histogram of the PPV estimate together with a Gaussian distribution (in red) as a reference. More frequencies are observed for positive errors and a slightly larger tail is observed for negative errors.

[2]This low residual rMBD value is due to a few negative site adapted GHI samples that need to be restored to their original value for data consistency.

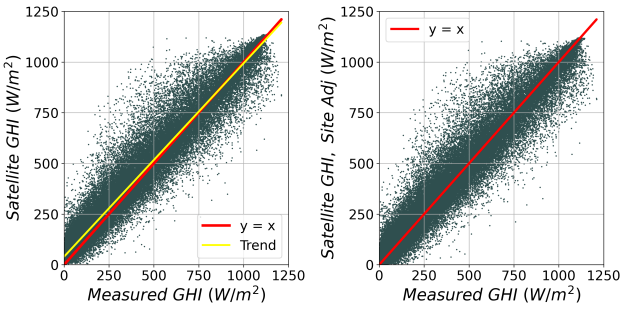


Fig. 1. CIM-McClear GHI scatter plot for original (left) and site-adapted (right) satellite estimates. A regression line (yellow) is added to the left plot.

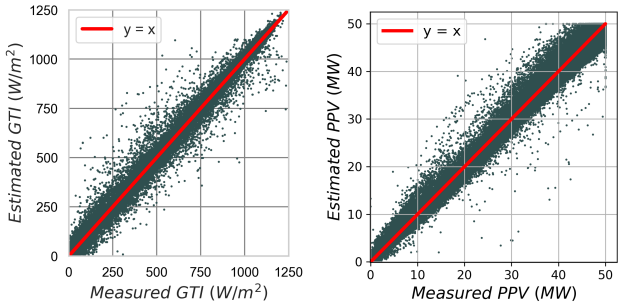


Fig. 2. Scatter plots between the GTI satellite estimate and the PPV estimate versus their corresponding measured values.

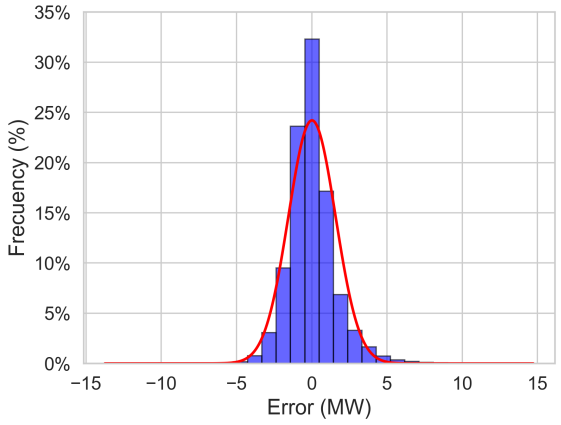


Fig. 3. Error distribution of the PPV, with a normal distribution (in red)

This performance assessment shows that the models used to fill the GHI, GTI and PPV gaps are adequate. In fact, these choices are all among the lowest uncertainty options for this process. In particular, it should be noted that they introduce low or negligible bias, which makes

them particularly suitable for annual and monthly analysis (for which the averages mainly filter uncertainty, not bias).

### *B. Solar resource and PV production characterization*

The monthly averages per year, calculated from the complete 10-minute data series (without gaps) of GHI, GTI and PPV, are shown in Fig. 4. The average of the whole data period (2018–2022) is shown with a black line, and the reference values for each variable are shown with a blue line. In the case of the GHI, the reference values come from the Uruguayan solar map (MSU[3], [5]). For the GTI, the available information is from the Uruguayan Typical Meteorological Year (AMTU, [7]). The PPV is compared with the data provided by the national electricity grid operator, UTE [4]. These studies serve as a reference for the results obtained in this work.

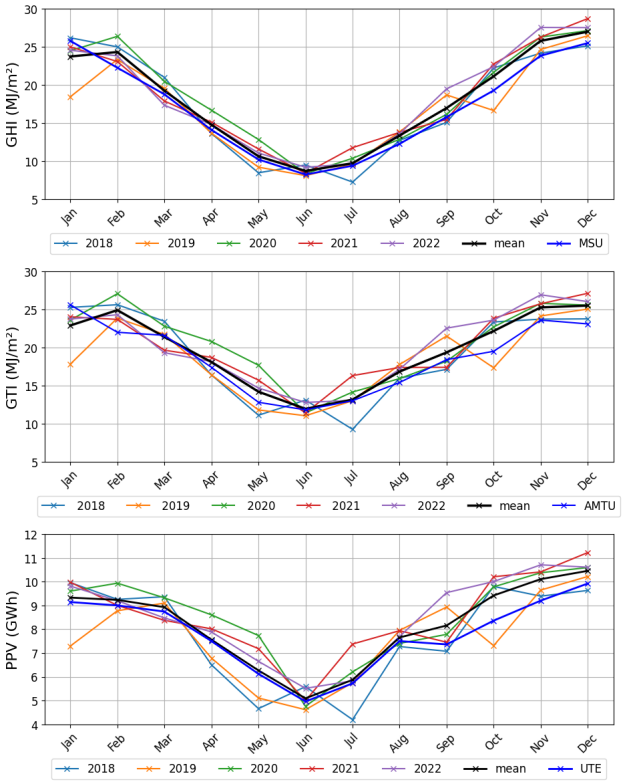


Fig. 4. Monthly evolution of GHI, GTI and PPV values.

Fig. 4 shows the expected seasonal pattern in the region, with higher values for each variable in summer and lower values in winter. This pattern is more pronounced for the GHI. The monthly means are clearly related as the variations are similar. Higher inter-annual monthly variability is observed in January, July and October over this five-year period. In particular, the year 2019 shows the lowest values to a considerable extent in January and October. For the GHI, the monthly averages derived from this work closely match the MSU estimates. The absolute differences range from 0.3 MJ/m$^2$ in July to 2.1 MJ/m$^2$ in January, with a MAD of 1.1 MJ/m$^2$ over the whole period. A similar pattern is observed between the obtained GTI series and the AMTU. The maximum difference compared to the AMTU monthly averages is 2.7 MJ/m$^2$ in January, while the minimum difference is about 0.1 MJ/m$^2$ in July. The PPV seasonal profile is compared to the UTE reference, which is the sum of reported monthly production. The estimated average monthly production varies between 5.1 GWh in July and 10.5 GWh in December. This PPV is closely aligned with the UTE reports, whose values range between 5.0 GWh and 9.9 GWh for the same months. For the whole period, the electricity generation estimated in this work amounts to 490 GWh, while UTE reports about 468 GWh, which is 4.8% less. This difference is mainly due to operational restrictions (both from O&M and curtailed power), so it is expected that the estimated values consistently exceed the actual production in each month.

The monthly capacity factors calculated for each year and the average for the whole period are shown in Fig. 5. The annual relative anomalies are shown in Fig. 6 (the difference between the annual average for each year and their mean value, expressed as a percentage of the mean value). As can be seen in Fig. 5, the CF is also affected by the seasonal variation and follows the same trends as the other magnitudes. The average monthly capacity factors vary between 14.1% (in June) and 28.1% (in December). In December 2021 the capacity factor exceeded 30%. The capacity factor for the five-year period is 22.4%. The anomalies range from +3.9% to −6.7%. The results in the Fig. 6 give a first indication that, the PV plant produced about 6% less energy during 2018-2019, while produced about 4% more energy over the 2020-2022 period.

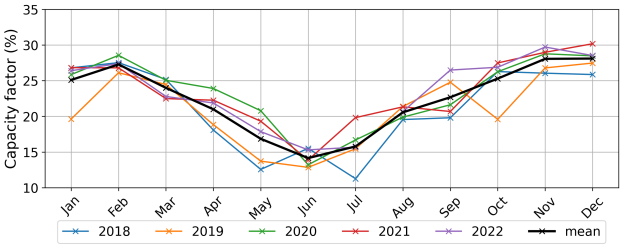


Fig. 5. Monthly evolution of the solar PV capacity factors.

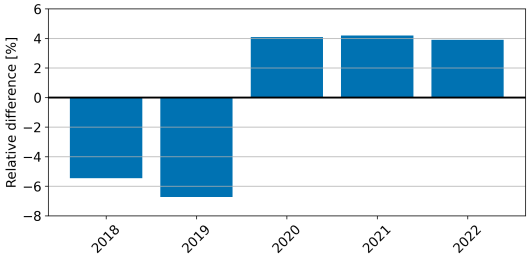


Fig. 6. Percentage anomalies of generation year by year

In 2017, Ref. [8] (Uruguay's solar map) estimated a capacity factor of 17.2% for a PV plant in the same location, but using DC capacity instead of nominal contractual power as used in this work. In [6], similar values were found for the Asahi PV plant in Salto from 2010

[3]http://les.edu.uy/online/msuv2/

[4]https://www.ute.com.uy/institucional/ute/utei/composicion-energetica

to 2013. The 4-year average CF calculated by the authors was 17.6%. Performing the same calculation as in the two previous works, but with the data from this work, the CF obtained is 17.4%. Although the similarity is remarkable, Uruguay's solar map is based on 17 years of satellite estimates, while this study averages only 5 years. A longer history of PV plant operation would be required for a conclusive comparison.

## V. Conclusions

This study developed a 5-year, gapless, high temporal resolution solar irradiance and PV generation time series for the 'La Jacinta' power plant in northwestern Uruguay. The performance of the models used was quantified and compared with other studies in the literature as well as with real data provided by the grid operator. Accurate annual and monthly PV production and capacity factors were estimated from this dataset, leading to an accurate assessment of PV plant operation over the period. This study is the first detailed assessment of solar capacity factors in the region through careful quality control and low uncertainty gap filling. Future work aims to evaluate other PV generation estimation models and calculate other plant performance indicators. In addition, a similar study is planned for single axis tracker plants. This will allow the comparison of production performance for different types of PV plant technologies in Uruguay.

## References


[1] N. Bolson, P. Prieto, and T. Patzek, "Capacity factors for electrical power generation from renewable and nonrenewable sources," *Proceedings of the National Academy of Sciences*, vol. 119, no. 52, 2022.

[2] H. Demirhan and Z. Renwick, "Missing value imputation for short to mid-term horizontal solar irradiance data," *Applied Energy*, vol. 225, pp. 998–1012, 2018.

[3] N. B. Mohamad, A.-C. Lai, and B.-H. Lim, "A case study in the tropical region to evaluate univariate imputation methods for solar irradiance data with different weather types," *Sustainable Energy Technologies and Assessments*, vol. 50, p. 101764, 2022.

[4] M. Schwandt, K. Chhatbar, R. Meyer, K. Fross, I. Mitra, R. Vashistha, G. Giridhar, S. Gomathinayagam, and A. Kumar, "Development and test of gap filling procedures for solar radiation data of the Indian SRRA measurement network," *Energy Procedia*, vol. 57, pp. 1100–1109, 2014. 2013 ISES Solar World Congress.

[5] R. Alonso-Suárez, G. Abal, R. Siri, and P. Muse, "Satellite-derived solar irradiation map for Uruguay," *Energy Procedia*, vol. 57, pp. 1237–1246, 2014.

[6] D. Oroño, R. Alonso-Suárez, G. Crapuchetti, G. Hermida, and M. Puppo, "Simulation of PV power plant's output in Uruguay," *Proceedings of the 4th International Workshop on Integration of Solar Power into Power Systems*, pp. 1–9, 2014.

[7] R. Alonso-Suárez, G. Abal, M. Bidegain, and P. Modernell, "Año Meteorológico Típico para Aplicaciones de Energía Solar - AMTUes: series horarias típicas para 5 sitios del Uruguay," tech. rep., Facultad de Ingeniería, UdelaR, Uruguay, 2016.

[8] R. Alonso-Suárez, *Estimación del recurso solar en Uruguay mediante imágenes satelitales*. PhD thesis, Universidad de la República del Uruguay, 2017.

[9] I. Piccioli, "Modelado de la irradiancia solar sobre superficies inclinadas," Master's thesis, Universidad de la República del Uruguay, 2022.

[10] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, and F. Vignola, "A new operational model for satellite-derived irradiances: description and validation," *Solar Energy*, vol. 73, no. 5, pp. 307–317, 2002.

[11] C. Rigollier, M. Lefevre, and L. Wald, "The method Heliosat-2 for deriving shortwave solar radiation from satellite images," *Solar Energy*, vol. 77, no. 2, pp. 159–169, 2004.

[12] R. Alonso-Suárez, G. Abal, R. Siri, and P. Musé, "Brightness-dependent Tarpley model for global solar radiation estimation using GOES satellite images: application to Uruguay," *Solar Energy*, vol. 86, pp. 3205–3215, Nov 2012.

[13] A. Laguarda, G. Giacosa, R. Alonso-Suárez, and G. Abal, "Performance of the site-adapted CAMS database and locally adjusted cloud index models for estimating global solar horizontal irradiation over the Pampa Húmeda," *Solar Energy*, vol. 199, pp. 295–307, 2020.

[14] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, A. Benedetti, J. W. Kaiser, and J. J. Morcrette, "McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions," *Atmospheric Measurement Techniques*, vol. 6, pp. 2403–2418, 2013.

[15] L. McArthur, "Baseline Surface Radiation Network operations manual," Tech. Rep. WCRP-121/ WMO TD-No. 1274, World Climate Research Programme – WMO, April 2005.

[16] D. Reindl, W. Beckman, and J. Duffie, "Evaluation of hourly tilted surface radiation models," *Solar Energy*, vol. 45, no. 1, pp. 9–17, 1990.

[17] J. Ruiz-Arias, H. Alsamamra, J. Tovar-Pescador, and D. Pozo-Vázquez, "Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions," *Energy Conversion and Management*, vol. 51, no. 5, pp. 881–893, 2010.

[18] G. Abal, D. Aicardi, R. Alonso-Suárez, and A. Laguarda, "Performance of empirical models for diffuse fraction in Uruguay," *Solar Energy*, vol. 141, pp. 166–181, 2017.

[19] IEC61724-1, "Photovoltaic system performance - guidelines for measurement, data exchange, and analysis," *International Electrotechnical Commission*, 2021.

[20] J. Polo, S. Wilbert, J. Ruiz-Arias, R. Meyer, C. Gueymard, M. Súri, L. Martín, T. Mieslinger, P. Blanc, I. Grant, J. Boland, P. Ineichen, J. Remund, R. Escobar, A. Troccoli, M. Sengupta, K. Nielsen, D. Renne, N. Geuder, and T. Cebecauer, "Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets," *Solar Energy*, vol. 132, pp. 25–37, 2016.

[21] A. Laguarda, P. Iturbide, X. Orsi, M. Denegri, S. Luza, L. Burgos, V. Stern, and R. Alonso-Suárez, "Validación de modelos satelitales Heliosat-4 y CIM-ESRA para la estimación de irradiancia solar en la Pampa Húmeda," *Energías Renovables y Medio Ambiente*, vol. 48, pp. 1–9, 2021.

[22] P. Iturbide, X. Orsi, M. Denegri, S. Fioretti1, S. Luza, V. Stern, R. Alonso-Suárez, and F. Ronchetti, "Modelos de machine learning para estimar la radiación solar horizontal en la Pampa Húmeda con información satelital multiescala," *Avances en Energías Renovables y Medio Ambiente*, 2023.

[23] J. D. Mondol, Y. G. Yohanis, and B. Norton, "Solar radiation modelling for the simulation of photovoltaic systems," *Renewable Energy*, vol. 33, no. 5, pp. 1109–1120, 2008.

[24] B. Liu and R. Jordan, "Daily insolation on surfaces tilted towards equator," *ASHRAE J.;(United States)*, vol. 10, 1961.

[25] J. Medeiros, "Avaliação de modelos utilizados na estimativa da geração de energia de uma usina fotovoltaica localizada no semiárido brasileiro," Master's thesis, Universidade Federal de Pernambuco, 2023.