An analysis of satellite-based Machine Learning models to estimate global solar irradiance at a horizontal plane

Paula Iturbide[1(✉)] [0009-0005-5534-5182] and Rodrigo Alonso-Suarez[2] [0000-0002-5216-5232] and Franco Ronchetti[3,4] [0000-0003-3173-1327]

[1] Grupo de Estudios de la Radiación Solar (GERSolar), Instituto de Ecología y Desarrollo Sustentable (INEDES). Univ. Nacional de Luján, CP 6700, Buenos Aires, Argentina.
[2] Laboratorio de Energía Solar, Facultad de Ingeniería, Universidad de la República, Uruguay.
[3] Instituto de Investigación en Informática LIDI, Universidad Nacional de La Plata, Buenos Aires, Argentina.
[4] Comisión de Investigaciones Científicas de la Pcia. de Buenos Aires (CICPBA), Buenos Aires, Argentina.
paula.iturbide@sas.com

Abstract. Accurate solar resource information is a fundamental requirement for solar energy ventures. The lack of precision in solar radiation data can significantly affect the success of the projects. Argentina has solar radiation ground measurement networks. The information obtained through this method is limited due to its spatial sparsity, since it is only possible to measure with appropriate quality in some sites across the territory. To overcome this limitation, it is common to generate models capable of estimating solar radiation through satellite images, which provide spatial resolution. This work develops and validates an empirical model for this purpose based on Machine Learning (ML), demonstrating that it is a useful and accurate tool to be considered. This allows ventures that make use of this type of energy to have greater certainty in the availability of the resource, and therefore in the decision-making process. Variables obtained from images of the geostationary meteorological satellite GOES-16, McClear clear-sky model estimates, and geometrically calculated information are used as input to the algorithms. The results of the ML models are compared with estimates from pre-existing models for the region that incorporate physical modelings, such as Heliosat-4 and CIM-ESRA. The evaluation shows a higher performance of the ML methods when multi-scale satellite information is used as input. The incorporation of multi-scale satellite data is not yet implemented in solar radiation physical modeling, which is an advantage of ML modeling.

Keywords: First Keyword, Second Keyword, Third Keyword.

1 Introduction

There are different ways to estimate solar radiation from satellites. One type of model, called physical models, attempts to solve the radiative transfer equations of solar radiation through the atmosphere. To do this, they use the information on atmospheric components that influence the absorption and scattering of solar radiation. Their performance depends on the quality with which these variables are known, which are operationally estimated by satellite through inverse problems or by atmospheric models. Another type of model, called statistical or empirical, does not attempt to model the physical attenuation phenomena through the atmosphere but adjusts the estimates to ground measurements using a set of coefficients or parameters. These models are therefore simpler to calculate but may have extrapolation problems to other sites where local conditions can be different. An intermediate category is that of hybrid models (semi-empirical), where a physical basis is assumed and parameters with physical meaning are adjusted to ground data.

Among the satellite models available for the Pampa Humeda region with a physical basis, some that are particularly relevant for this work are the CIM-ESRA [9] and the Heliosat-4 [19]. The first is a semi-empirical model based on information from the visible channel of the GOES-16 satellite (located at latitude 75º W), operated by NOAA, while the second is a physical model based on information from the European Meteosat Second Generation (MSG) satellite, located at the Greenwich meridian.

The CIM-ESRA model replicates the strategy used by other satellite models such as SUNY [16][18] or Heliosat-2 [21]. In these models, global horizontal irradiance (GHI) is estimated as the irradiation that would occur under clear sky conditions modulated by an empirical factor based on a cloudiness index (CIM is due to Cloud Index Model) obtained from satellite images. That is,

$$I = I_{cs} F(C) \tag{1}$$

where I is the estimated solar irradiance, Ics is the irradiance that would occur under clear sky conditions (estimated by a clear sky model), and F is a cloud attenuation factor adjusted locally (or regionally) that depends on a cloudiness index (C). CIM-ESRA uses the ESRA clear sky model [21], whose only input variable is the Linke turbidity for an air mass equal to two, TL [12]. This variable represents the state of the cloud-free atmosphere using a single parameter, which models the aerosols and water vapor effects. The CIM-ESRA model uses seasonal average cycles of TL for its operation [10].

The operational model of the Copernicus Atmosphere Monitoring Service (CAMS) is Heliosat-4 [19]. This model consists of two combined physical sub-models: McClear [11], which estimates irradiance under clear sky conditions, and McCloud, which adds the effect of cloudiness. Both McClear and McCloud are based on outputs from the

radiative transfer model LibRadTran [4], which is computationally expensive. McClear uses inputs such as aerosols properties, the precipitable water column, and ozone obtained from the CAMS reanalysis database (atmospheric modeling), as well as daily estimates of land surface albedo obtained from the MODIS satellite instrument (a low-earth orbit meteorological satellite). McCloud estimates the attenuation of solar radiation in the presence of clouds with MSG images, and the use of an adapted APOLLO/SEV methodology [7]. The Heliosat-4 estimates are available as a free internet service for the geographic coverage of the MSG satellite and are widely used (on the other hand, the McClear model is available with global coverage).

Currently, machine learning techniques such as artificial neural networks (ANN), k-nearest neighbors (kNN), support vector regression (SVR), extreme learning machines (ELM), and tree-based ensembles are also used to estimate solar radiation [5]. The input variables used in some works include the month, latitude, longitude, and altitude, as well as meteorological variables such as vapor pressure, land surface temperature, day length, temperature, humidity, precipitation, cloudiness, wind speed, evaporation, and the global solar radiation is the output of the models [20]. The data is divided into training and validation sets, the models are trained with the training set and validated with the remaining data.

Other works use an empirical clear sky model combined with variables such as temperature, humidity, and pressure obtained from available meteorological databases on land as input to a neural network [6]. A second work in this same line uses values from meteorological stations similar to those mentioned above [22], proposing different models (different combinations of input variables) all based on neural networks. A third work [15] compares the use of neural networks for different sky conditions where, according to the authors, they find poor performance. All these works make use of ground measurements. Works using satellite information under the ML framework are scarce for solar radiation [5][23]. It is interesting to note that the use of ground-based measurements limits the estimation tools significantly, both in terms of data availability and extrapolation capacity to sites where there are no ground measurements. Satellite information does not have these limitations, as it is accessible and available, even operationally, and is capable of observing all points in large continental regions with spatial resolution suitable for solar energy projects (the pixel width is typical of 1 km, enough to model a large-scale solar photovoltaic plant). Therefore, satellite-based methods have a much higher potential.

2 Materials and methods

2.1 Ground-based measurements

The GERSolar R&D group manages a ground-based solar radiation measurement network that covers the Buenos Aires Province. The global horizontal irradiance (GHI) measurements are recorded in all stations with a 1-minute resolution, being the most common solar radiation data. In this work, a 10-minute time scale is used, which is the same as the GOES-16 satellite image rate. Therefore, the minute resolution ground data are averaged to obtain their 10-minute value, following the criterion that each average should have at least 2/3 of the corresponding minute data. The quality of measurements is a prerequisite for any meaningful use. The 10-minute GHI data series were subjected to a quality control algorithm consisting of four successive filters (Table 1) and a visual inspection of the series to eliminate shading periods or other anomalous data. The selected filters are standard in the field [14]. The objective of the quality procedure is to select a typical data set, where missing, anomalous, or error-affected data, and measurements affected by isolated or short-lived phenomena (such as over-irradiance) are excluded [2].

Table 1. Filters applied to ground-based measurements

Filter	Criterion	Description
1	$\alpha_s > 7^{\circ}$	Minimun solar height
2	$-2W/m^2 < I_h < I_0.1.2.cos\theta_z^{1.2} + 50W/m^2$	BSRN levels (Long y Shi, 2008)
3	$0W/m^2 < I_h < I_h^{ESRA}(TL = 1.8)$	Elevations of a clear sky model
4	$ktp < 0.89$	Bound Perez clarity index

The first filter sets a minimum solar altitude of 7º and excludes samples taken during sunrise and sunset. This filter aims to eliminate ground data with higher relative error due to the measuring equipment, which presents known directional errors at high incidence angles. The second filter uses a widely used criterion that establishes a range of values for extremely rare cases. This criterion is based on the solar zenith angle [13] and is recommended by the worldwide Baseline Solar Radiation Network (BSRN). The third filter imposes an upper limitation on the measured values, which correspond to clear sky conditions. This filter uses the ESRA model with very low atmospheric turbidity (TL = 1.8), which ensures an upper limit for GHI [1]. Finally, the fourth filter imposes a maximum value on the modified clearness index by Perez et al. [17]. The modified clearness index is defined as a regular clearness index (the GHI normalized by its corresponding top-of-the-atmosphere value) but without its dependence on solar

altitude. In summary, each filter has its own criteria and is used to exclude inaccurate or unreliable data. In this section, the results for the Luján station operated by GERSolar along with INTA for the years 2019-2021 are shown, and its precise location is indicated in Table 2.

Table 2. Location of the measurement station and its measurement equipment. The total number N (filtered) corresponds to the reliable 10-minute measurements in the 2019-2021 period.

Estation	Latitude (Degrees)	Longitude (Degrees)	Equipment	N Totals (leaked data)
Luján, ARG	-34.558	-59.062	CMP11 & CMP21	62,592

The solarimetric station in Luján has data loggers from Campbell Scientific and uses two Kipp & Zonen CMP11 and CMP21 pyranometers as measurement equipment. These devices are Class A according to the equipment classification standard for solar radiation measurement (ISO 9060:2018). The data series was assembled by selecting the higher quality equipment during any concurrent measurement period, and one of the data series was chosen when the other device failed or the measurement had errors. Fig. 1 shows the distribution of the data points for the period considered in this article, after the series underwent the quality control. The white gaps correspond to missing data.

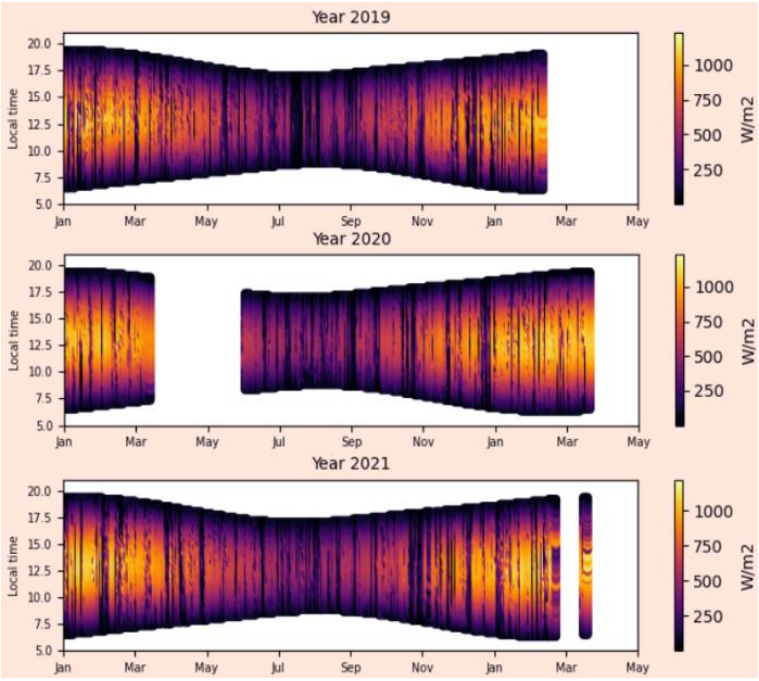

Fig. 1. Plot of ground data after subjecting it to quality control algorithms.

2.2 Satellite images

The visible channel images (C02 channel, centered at 0.64 μm) of the GOES-16 meteorological geostationary satellite were used. This satellite is part of the geostationary satellite network for Earth observation that covers the entire globe. This particular satellite is operated by the National Oceanic and Atmospheric Administration (NOAA) of the United States and has been generating images for the entire American continent since 2018 with a time interval between 10 and 15 minutes. The satellite is located over the Earth's equator at -75°W longitude in geostationary orbit. Its spatial resolution varies along the image, being 500 m at its nadir. Over the Pampa Húmeda region, the pixel size varies between 1 and 3 km. The visible channel is suitable for estimating solar radiation because diurnal cloudiness is recognizable and quantifiable. Clouds are typically more reflective than the background (the Earth's surface), and therefore distinguishable.

The two classical variables calculated from visible channel satellite images are the reflectance factor (FR) and the planetary reflectance (RP). The latter is also known as Earth albedo and is denoted as ρ. The FR is a normalization of the radiance measured by the satellite from each pixel with respect to the maximum it is capable of measuring (i.e., the solar radiation incident on the top of the atmosphere normalized by the spectral response of the satellite radiometer). It is therefore in the interval [0, 1] and contains, in addition to daylight cloudiness information, spatial information about the variable illumination of the Sun over the Earth's surface. The quantity RP contains the additional normalization to remove this spatial dependence and is effectively the reflectivity of the Earth, in its strict physical sense.

2.3 Data set configuration

The following input variables were used for the machine learning algorithms:

- Cos(z) - Cosine of the solar zenith angle. It is the angle between the local zenith (the local vertical) and the Sun's direction. This quantity varies instantaneously according to the relative movement between Earth and the Sun and also depends on the position on Earth's surface (latitude and longitude of the site).

- McClear estimates - The McClear clear sky model is widely used in solar radiation simulation. Its main advantages include that the estimates are freely available to download, and that it is one of the most accurate models [24]. The model was developed by Lefèvre et al. [11] and uses atmospheric information such as pressure, temperature, humidity, and aerosol optical thickness. The model employs a physics simulation algorithm based on radiative transfer cal-

culations and atmospheric modeling to estimate direct, diffuse, and global solar radiation under clear sky conditions for a specific location. These estimates were downloaded here from the CAMS web portal.

- Multiscale satellite information (F01-F18 and R01-R18) - These are spatial averages of the Reflectance Factor and Earth Albedo, respectively, in square latitude-longitude cells of incremental size. The numbers from 01 to 18 indicate the increasing cell size in which the satellite information is averaged, meaning cell sides from 0.01° to 0.90° respectively. The relationship between the cell side and numbering is not linear. For smaller cell sizes, the spacing between sides has a higher resolution, and vice versa for larger cell sizes.

2.4 Definition of training and validation sets

The aim of the ML algorithms is to learn how to estimate GHI from satellite data by adjusting their parameters to ground measurements. The annual seasonality of GHI is characteristic. Hence, it is an adequate choice to select the training and testing data sets on an annual basis. A data set corresponding to two of the three available years was used for training and the remaining year for validation. This was done for all possible year permutations. Taking the validation set in this way avoids three possible biases. The first one is related to the random distribution of data in the training and validation sets as consecutive time samples can be very similar to each other. The second one is related to the particularities of each year as one year could be sufficiently different from another. The third one is related to incomplete data as there is a two-month gap in 2020 that was unrecoverable, and for the years 2019 and 2021, the month of December is missing.

2.5 Supervised learning models

Three ML models were considered: a linear multiple regression, which allows for a quick white box estimation; a Feed Forward Neural Network that uses a 100 neurons hidden layer, ReLU activation function, and Adam optimizer; and a Random Forest with 30 estimators. The latter two are non-linear models.

3 Results and discussion

3.1 Performance reference

GHI estimations from the CIM-ESRA and Heliosat-4 satellite models were downloaded for the Luján station and the period 2019-2021 from the websites

http://les.edu.uy/online/stack-loc/ (CIM-ESRA, LES web portal) and https://www.soda-pro.com/web-services (Heliosat-4, CAMS web portal). The CIM-ESRA estimations are available on a 10-minute scale for different Latin American stations, including Luján, Argentina. The CAMS estimations are not available at the 10-minute temporal resolution used in this article, so one-minute data were downloaded and then integrated into the 10-minute time scale. The data from these models were used as a performance reference for the developed statistical ML models against available databases for the region.

The performance metrics used in this work are the MBE, RMSE, MAE, and R^2 (Mean Bias Error, Root Mean Square Error, Mean Absolute Error, and R-squared) respectively. For the first three, their corresponding relative values are also reported, which are named as MBEn, RMSEn, and MAEn, respectively. These metrics are expressed as a percentage of the measurement average value, which is 420 W/m^2 for this station. MBE is the systematic bias of the models, and RMSE and MAE measure the error dispersion with different weighting laws, being the former more sensitive to outliers. The evaluation results of these models can be seen in Table 3.

Table 3. Performance metrics of the Heliosat-4 (CAMS) and CIM-ESRA models. MBE, RMSE, and MAE are measured in W/m2, their relative versions in %, and R^2 is dimensionless.

	MBEn	**RMSE**	**RMSEn**	**MAE**	**MAEn**	**R^2**
CAMS	-0.99	93.38	22.22	54.94	13.07	0.894
CIM-ESRA	1.63	74.79	17.07	49.34	11.26	0.928

It is possible to notice that the estimates from the CIM-ESRA model fit more accurately to the region compared to those obtained by the Heliosat-4 model. The superior performance of CIM-ESRA is due to two reasons: (a) it uses GOES-16 information instead of MSG, which has better-viewing angles for the Pampa Húmeda region, and (b) it is a semi-empirical model whose adjustable parameters were determined for the region (previously) by using information from 10 sites in the 2010-2017 period [8]. The performance reference is therefore given by the CIM-ESRA model, which also uses information from the same satellite as the input data used for the machine learning algorithms in this work.

3.2 Implementation of Machine Learning Models

The variables detailed in section 2.3 were used as input for the implemented models. Different combinations of input variables were tested for the implemented models, and it was found that for the black box algorithms, it was sufficient to have information

from the satellite image variables. Tables 4 and 5 show the results. The metrics are presented for each validation year, with the fit to the other two remaining years of the 2019-2021 period. The last column shows the average performance over the 3 validation years, which is used here as a comparison value with the CIM-ESRA reference in Table 3.

Table 4. Results of ML models using all input variables.

	Linear Regression				Neural Networks 100 Hidden ReLU-Adam				Random Forest n=30			
Metric	**2019**	**2020**	**2021**	**Mean**	**2019**	**2020**	**2021**	**Mean**	**2019**	**2020**	**2021**	**Mean**
MBE	0.08	-1.32	1.10	**-0.05**	-0.59	-1.45	-2.52	**-1.52**	1.26	-1.99	1.73	**0.33**
MBEn	0.02	-0.28	0.25	**0.00**	-0.14	-0.31	-0.58	**-0.34**	0.30	-0.43	0.40	**0.09**
RMSE	70.49	69.05	70.79	**70.11**	68.84	67.19	68.61	**68.21**	71.88	69.14	70.44	**70.49**
RMSEn	16.86	14.79	16.37	**16.01**	16.47	14.40	15.87	**15.58**	17.19	14.81	16.29	**16.10**
MAE	45.08	40.88	44.98	**43.65**	40.74	37.59	39.81	**39.38**	42.29	39.16	42.21	**41.22**
MAEn	10.78	8.76	10.40	**9.98**	9.74	8.05	9.21	**9.00**	10.12	8.39	9.76	**9.42**
R^2	0.94	0.95	0.94	**0.94**	0.94	0.95	0.95	**0.95**	0.94	0.95	0.94	**0.94**

Table 5. Results of the ML models using only satellite variables (excluding the clear sky model and the geometric variable Cos(z)).

	Linear Regression				Neural Networks 100 Hidden ReLU-Adam				Random Forest n=30			
Metric	**2019**	**2020**	**2021**	**Mean**	**2019**	**2020**	**2021**	**Mean**	**2019**	**2020**	**2021**	**Mean**
MBE	-0.95	18.50	-9.97	**2.53**	-5.33	-0.26	0.40	**-1.73**	0.45	0.82	2.35	**1.21**
MBEn	-0.23	3.96	-2.31	**0.48**	-1.28	-0.06	0.09	**-0.41**	0.11	0.18	0.54	**0.28**
RMSE	176.82	211.23	179.33	**189.13**	70.13	67.73	69.64	**69.16**	72.08	69.16	71.28	**70.84**
RMSEn	42.29	45.26	41.48	**43.01**	16.77	14.51	16.11	**15.80**	17.24	14.82	16.49	**16.18**
MAE	135.59	150.77	141.22	**142.53**	44.20	41.07	42.67	**42.65**	44.21	41.40	44.51	**43.37**
MAEn	32.43	32.31	32.66	**32.47**	10.57	8.80	9.87	**9.75**	10.58	8.87	10.29	**9.91**
R^2	0.62	0.54	0.63	**0.60**	0.94	0.95	0.94	**0.95**	0.94	0.95	0.94	**0.94**

When using all variables, the statistical techniques exhibit similar performance, although it is possible to notice that the neural network's performance is slightly better. If we look at, for example, RMSEn, the neural network's performance is 0.4% higher than the others on average. It also shows a 1.5% RMSEn gain compared to the CIM-ESRA reference. When comparing the linear regression with and without the clear sky estimates and the Cos(z) variable, a marked downgrade in performance is noticed, unlike what happens in the case of nonlinear models (only a loss of 0.2% in RMSEn is observed). This suggests that for these types of techniques, the satellite variables would suffice to estimate solar radiation (F01-F18 and R01-R18), but this is not the case for the simple linear regression algorithm. In that case, linear regression loses the necessary temporal reference to estimate a magnitude with a marked geometric daily and annual

component like the GHI. However, machine learning algorithms do not lose it. The geometric information is implicit in the relationship between FR and RP, and ML algorithms are capable of learning from it. Fig. 2 shows the predictions of the neural network that only uses satellite information for two days with cloudy sky conditions corresponding to January (top) and July (bottom). Fig. 3 shows the estimation of the same model but for two clear days during the year.

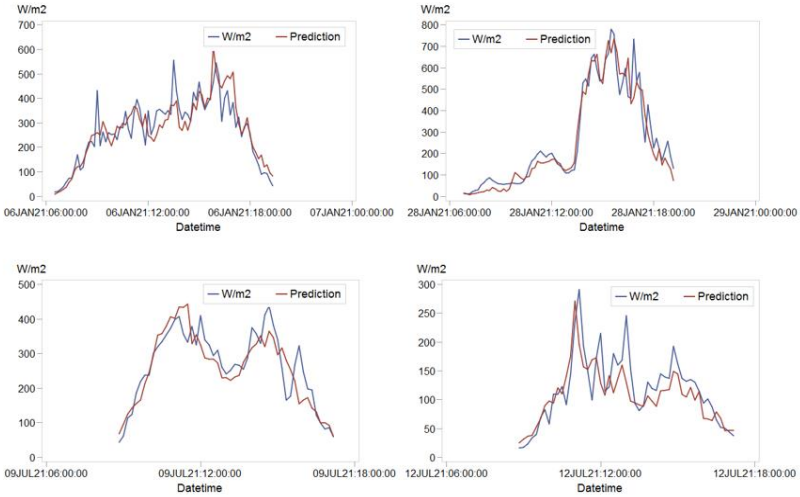

Fig. 2. Comparative plots between ground measurements and the neural network that uses only satellite variables for two days in January 2021 (top) and two days in July (bottom) under cloudy sky conditions.

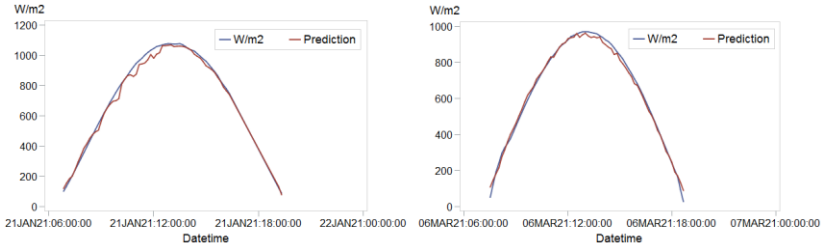

Fig. 3. Comparative plots between ground measurements and the neural network that uses only satellite variables for two days with clear sky conditions.

It can be appreciated in the previous plots that the model is capable of capturing, in general, the presence of clouds and their effect on GHI with a good level of success. This is reflected in the metrics of Table 5. In the case of clear sky days, it can be observed that the predicted curve is not completely smooth. This could be due to the non-use of the clear sky model or the calculated geometric information of Cos(z), and could be indicating the approximate temporal positioning of the algorithm derived from the joint use of FR and RP.

4 Conclusions and future work

In this study, it was observed that the neural network algorithm had the best performance, followed by the random forest method, even after eliminating input variables such as the clear sky model and cosine of the solar zenith angle. In contrast, a simple linear regression showed poor performance when these input variables were removed, as it could not reconstruct the necessary temporal reference to estimate GHI. In terms of comparison with the CIM-ESRA reference model, better performance was observed for the empirical ML models. The comparative metrics between CIM-ESRA and the neural network are all favorable to the latter, with MAEn of 11.3% vs 9.7%, RMSEn of 17.1% vs 15.8%, and MBEn of +1.6% vs -0.4%. However, it will be important in future work to analyze the behavior of the empirical model in other locations in the Pampa Húmeda region, i.e., to analyze its spatial extrapolation using a testing set in different measuring sites. For future work, it would also be appropriate to consider other relevant satellite variables, such as the satellite cloud index [3] and information from the satellite's infrared channels. These parameters can provide additional information about the cloud cover or atmospheric transmissivity, which could improve the accuracy of ML solar radiation estimates. Therefore, it would be valuable to evaluate the inclusion of these variables in the model and measure their impact on its performance.

References

1. Abal, G., Aicardi, D., Alonso-Suárez, R., and Laguarda, A. (2017). Performance of empirical models for diffuse fraction in uruguay. Solar Energy, 141:166–181.
2. Aristegui, R.; Iturbide, P.; Stern, V.; Lell, J.; Righini, R. (2019). Variabilidad de corto plazo y valores extremos de la irradiancia solar en la Pampa Húmeda Argentina. Avances en Energías Renovables y Medio Ambiente (AVERMA), vol. 23, pág. 19-30.
3. Cano, D., Monget, J., Albuisson, M., Guillard, H., Regas, N., and Wald, L. (1986). A method for the determination of global solar radiation from meteorological satellite data. Solar Energy, 37:31–39.
4. Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling, T., and Bugliaro, L. (2016). The Libradtran software package for radiative transfer calculations (version 2.0.1). Geoscientific model development, 9:1647–1672.

5. Huang, G., Li, Z., Li, X., Liang, S., Yang, K., Wang, D., Zhang, Y. (2019). Estimating surface solar irradiance from satellites: Past, present, and future perspectives. Remote Sensing of Environment, 233, https://doi.org/10.1016/j.rse.2019.111371.
6. Jiménez, V. A., Will, A., & Rodríguez, S. (2017). Estimación de radiación solar horaria utilizando modelos empíricos y redes neuronales artificiales. Ciencia y tecnología, (17), 29-45.
7. Kriebel, K., Gesell, G., Kastner, M., and Mannstein, H. (2003). The cloud analysis tool APOLLO: Improvements and validations. International Journal of Remote Sensing, 24(12):2389–2408.
8. Laguarda, A., Iturbide, P., Orsi, X., Denegri, M. J., Luza, S., Burgos, B. L., Alonso-Suárez, R. (2021). Validación de modelos satelitales Heliosat-4 y CIM-ESRA para la estimación de irradiancia solar en la Pampa Húmeda. Energías Renovables y Medio Ambiente, 48, 1-9.
9. Laguarda, A., Giacosa, G., Alonso-Suárez, R., and Abal, G. (2020). Performance of the site-adapted CAMS database and locally adjusted cloud index models for estimating global solar horizontal irradiation over the Pampa Húmeda region. Solar Energy, 199:295–307.
10. Laguarda, A., and Abal, G. (2017). Clear-sky broadband irradiance: first model assessment in Uruguay. Proceedings of the ISES Solar World Congress, ISBN 978-39-81465-97-6, pp. 1360-1371, doi:10.18086/swc.2017.21.05.
11. Lefevre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., ... and Morcrette, J. J. (2013). McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmospheric Measurement Techniques, 6(9), 2403-2418.
12. Linke, F. (1922). Transmissions-koeffizient und trübungsfaktor. Meteorological Magazine Beiträge Zur Physik der Atmosphäre Beitr, 10:91–103.
13. Long, C. N., & Shi, Y. (2008). An automated quality assessment and control algorithm for surface radiation measurements. The Open Atmospheric Science Journal, 2(1).
14. McArthur, L. (2005). Baseline Surface Radiation Network (BSRN) Operations Manual. Td-no. 1274, wrcp/wmo, World Meteorological Organization (WMO, www.wmo.org).
15. Olivera, L., Atia, J., Amet, L., Osio, J., Morales, M., & Cappelletti, M. (2020). Uso de redes neuronales artificiales para la estimación de la radiación solar horaria bajo diferentes condiciones de cielo. Avances en Energías Renovables y Medio Ambiente-AVERMA, 24, 232-243.
16. Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., F., and Vignola (2002). A new operational model for satellite-derived irradiances: description and validation. Solar Energy, 73:307–317.
17. Perez, R., Ineichen, P., Seals, R., & Zelenka, A. (1990). Making full use of the clearness index for parameterizing hourly insolation conditions. Solar Energy, 45(2), 111-114.
18. Perez, R., Schlemmer, J., Hemker, K., Kivalov, S., Kankiewicz, A., and Gueymard, C. (2015). Satellite-to-irradiance modeling - a new version of the SUNY model. En 42nd Photovoltaic Specialist Conference (PVSC), pp. 1–7.
19. Qu, Z., Oumbe, A., Blanc, P., Espinar, B., Gesell, G., Gschwind, B., Klüser, L., Lefevre, M., Saboret, L., Schroedter-Homscheidt, M., and Wald, L. (2017). Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method. Meteorologische Zeitschrift, 26(1):33–57.

20. Raichijk, C. (2008). Estimación de la irradiación solar global en Argentina mediante el uso de redes neuronales. Energías Renovables y Medio Ambiente (ISSN 0328-932X). Vol. 22, pp. 1 - 6.
21. Rigollier, C., Lefevre, M., and Wald, L. (2004). The method Heliosat-2 for deriving shortwave solar radiation from satellite images. Solar Energy, 77(2):159–169.
22. Sayago, S., Bocco, M., Ovando, G., & Willington, E. A. (2011). Radiación solar horaria: modelos de estimación a partir de variables meteorológicas básicas. Avances en Energías Renovables y Medio Ambiente, 15.
23. Verbois, H., Saint-Drenan, Y.-M., Becquet, V., Gschwind, B., Blanc, P. (2023). Retrieval of surface solar irradiance from satellite using machine learning: pitfalls and perspectives, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-243.
24. Yang, D. (2020). Choice of clear-sky model in solar forecasting. Journal of Renewable and Sustainable Energy 12, 026101, https://doi.org/10.1063/5.0003495.