Solar Irradiance Intraday Forecasting using LSTM Networks, ground measurements and Satellite Information

Ximena Orsi Millán[1], Paula Iturbide[1], Valeria Stern[1], Dario Licata Caruso[1], Anahí Lanson[1] and Rodrigo Alonso Suarez[2]

[1] Grupo de Estudios de la Radiación Solar (GERSolar), Instituto de Ecología y Desarrollo Sustentable (INEDES). Univ. Nacional de Luján, Buenos Aires, (Argentina).

[2] Laboratorio de Energía Solar, Dpto. de Física del CENUR Litoral Norte, Udelar, (Uruguay).

Abstract

Accurate forecasting of solar irradiance at photovoltaic (PV) sites enhances grid reliability and supports key operational tasks such as reserve planning and energy trading. This study presents a hybrid, dual-input recurrent neural network for intraday forecasting of global horizontal irradiance (GHI) in the Argentine Pampas incorporating past ground measurements and satellite images. The model combines sequential information from recent observations with exogenous predictors representing spatial and geometric characteristics. Sequential inputs include the clear-sky index (k_c), reflectance indices derived from visible-channel satellite imagery at multiple spatial scales, and meteorological variables. The exogenous branch encodes static features such as station coordinates, station identifiers, and solar zenith angle. The network predicts the hourly k_c for lead times from 1 to 6 hours, which is then converted to GHI using clear-sky estimates. Results show that integrating satellite-derived reflectance and meteorological predictors substantially improves forecast accuracy compared with models based solely on ground-based measurements, with the largest gains observed at longer horizons and for reflectance indices aggregated over wider spatial areas.

Keywords: GHI Forecast, LSTM Networks, Satellite Indices, Meteorological Information

1. Introduction

A short-term forecasting framework for global horizontal irradiance (GHI) was developed for the Argentine Pampas using a hybrid, dual-input recurrent neural architecture. The model combines sequential information from recent observations with exogenous predictors that encode spatial and geometric context. Sequential inputs include the clear-sky index (k_c), reflectance indices derived from visible-channel geostationary satellite imagery, and local meteorological variables, while the exogenous branch incorporates static features such as latitude, longitude, station identifiers, and the solar zenith angle and its first time derivative. In this work, the term *exogenous* refers to non-sequential inputs that are not processed by the recurrent layer but provide complementary spatial or geometric information to improve the forecast. Satellite reflectance indices provide high-frequency (10 min time rate), multiscale characterization of cloud cover and its temporal evolution, enabling the network to capture both short-term variability and broader spatial patterns that affect surface irradiance.

As a methodological approach, the clear-sky index (k_c) was employed as the target variable, defined as the ratio between the measured GHI and the theoretical clear-sky GHI value. This choice allows the geometric variability associated with the apparent motion of the Sun to be eliminated, facilitates learning by using a normalized and dimensionless variable, and improves comparability between different stations and atmospheric conditions. The predicted GHI is then reconstructed from the estimated k_c by multiplying it by the corresponding clear-sky value. Model performance is evaluated directly on GHI for practical reasons and to ensure comparability with other studies reported in the literature.

Over the past two decades, solar radiation forecasting has evolved from classical statistical approaches to machine learning techniques and, more recently, to deep architectures capable of capturing the nonlinear and highly dynamic nature of the solar resource. Early autoregressive models, widely used for intraday forecasting horizons, proved useful under stable-sky conditions, but their predictive ability rapidly deteriorates in the presence of variable cloudiness or abrupt transitions (Aguiar et al. 2015, Marchesoni & Alonso-Suárez, 2020).

In this context, the incorporation of satellite and modeled meteorological information made it possible to improve the representation of atmospheric states. In particular, the combination of ground observations, numerical weather prediction outputs, and variables derived from satellite imagery within probabilistic neural schemes produced substantial improvements: reductions of up to 20% in root mean square error were reported compared to purely statistical models, especially for 1–3-hour horizons, where cloud evolution is the dominant factor.

Complementary approaches have also been developed combining evolutionary strategies and machine learning techniques to optimize solar irradiance prediction. In particular, evolutionary artificial neural networks, which integrate genetic algorithms to simultaneously adjust both the architecture and the synaptic weights of the network, have shown superior performance compared to conventional backpropagation-trained networks. This approach led to a reduction in root mean square error and improved model stability under daily and seasonal atmospheric variability (Guijo-Rubio et al., 2021). In parallel, ensemble machine learning models incorporating spatiotemporal parameters have proven highly effective for very short-term forecasting, exploiting information from nearby meteorological stations and the local dynamics of cloud cover. These models, based on feedforward neural networks optimized with spatial and temporal information, achieved 10-minute irradiance forecasts with an RMSE of 50.8 W/m² and accuracy improvements above 11% relative to persistence, demonstrating their potential for real-time photovoltaic power management (Rodríguez et al., 2021).

The development of machine learning and deep learning techniques marked a paradigm shift in irradiance estimation, as it enabled replacing empirical or physically based relationships with representations learned directly from data. Recent reviews show that deep neural networks consistently outperform linear, support vector, or random forest models, particularly for short horizons and under variable atmospheric conditions (Assaf et al., 2023; Chen et al., 2024). Hybrid models that combine convolutional and LSTM networks (CNN-LSTM) are particularly effective, leveraging the former's ability to extract spatial patterns (such as cloud structure and displacement) and the latter's ability to model temporal dependencies. In several studies, these architectures achieved normalized errors (nRMSE) below 12% and coefficients of determination above 0.9, values highly competitive with physical or persistence methods.

The use of satellite information has proven to be a turning point in model accuracy. The availability of databases such as SARAH-2.1 or imagery from the COMS-MI and EUMETSAT satellites has made it possible to generate models entirely based on satellite data, capable of covering extensive geographical areas without requiring local calibration using ground measurements. Some recent works propose fully data-driven spatiotemporal models that, from sequences of visible-channel reflectance or effective albedo, predict the evolution of cloudiness (Marchesoni-Acland et al., 2023) and irradiance several steps ahead (Nielsen et al., 2021). These architectures, generally based on U-Nets or ConvLSTM encoder–decoder structures, learn directly from patterns of cloud movement and formation, and have demonstrated superior performance to optical-flow or reanalysis-based models. In evaluations over the European continent, root mean square errors were reduced by more than 15% compared to ERA5 and TV-L1 (Nielsen et al., 2021), while maintaining spatial coherence between predicted and observed fields, even under strong convective cloud conditions.

In other regions, such as South Asia, deep LSTM networks have been shown to capture with high fidelity the daily variability of irradiance and its relationship with local meteorological variables. Models integrating temperature, humidity, cloud cover, and wind speed achieved coefficients of determination above 0.9 and mean absolute errors on the order of 40 W/m² for horizons up to 24 hours, highlighting their robustness to seasonality and changing climatic regimes (Jayasankar et al. 2024).

The application of these techniques in South America has followed a similar trend. In Uruguay and southern Brazil, recurrent and LSTM networks trained with clear-sky index series demonstrated that normalization through kc not only facilitates model learning but also improves its ability to generalize across different sites. The inclusion of satellite-derived cloudiness variables reduced the normalized error (nRMSE) to values between 10% and 13%, significantly outperforming purely autoregressive approaches (Coleto-Rola and Alonso-Suárez 2020). These results confirmed that the combined use of ground and satellite information is key to improving the accuracy of hourly forecasts.

More recently, the integration of neural approaches with structured learning algorithms has revealed a new path to improve irradiance prediction. Hybrid models combining deep networks with tree-based methods (such as Random Forest or Gradient Boosting) take advantage of both the explanatory capacity of interpretable models and the representational power of neural networks (Thaker et al.2024). This type of architecture, applied to GHI forecasting from EUMETSAT satellite image pixels and meteorological variables, achieved normalized errors close to 10% for horizons of 1 to 6 hours, outperforming persistence and NWP simulations in accuracy.

The results of these studies consistently indicate that improving short-term prediction fundamentally depends on the model's ability to capture cloud transitions and their impact on surface irradiance. The combination of high-frequency satellite data with ground-based radiation measurements, together with neural architectures capable of representing temporal and spatial dependencies, has significantly reduced forecasting uncertainty. In this context, the present work adopts and adapts these methodological advances to develop an intraday GHI

forecasting system for the Argentine Pampas region, based on LSTM networks trained with sequences of ground data, multiscale satellite reflectance, and modeled meteorological variables. The proposal aims to provide empirical evidence on the applicability of these techniques in mid-latitude regions of the Southern Hemisphere, contributing to the development of operational tools for photovoltaic generation management and the integration of solar resources into power grids with high renewable penetration.

2. Data and methods

2.1. Data

The database spans the 2019–2021 period and includes global horizontal irradiance (GHI) measurements from three Argentine stations: Luján, Concepción del Uruguay, and Paraná. Measurements were obtained with Class A pyranometers compliant with the ISO 9060:2018 standard, originally recorded at 1-minute resolution and later integrated into 10-minute averages to match the time rate of GOES-16 satellite imagery. Table 1 lists the geographic coordinates of the three stations.

Table 1: Geographic coordinates of the measurement stations.

Location	Latitude	Longitude
Luján	-34.56	-59.06
Paraná	-31.85	-60.54
Concepción del Uruguay	-32.49	-58.35

Strict data-quality control was applied following the Baseline Surface Radiation Network (BSRN) guidelines (McArthur, 2005), complemented by visual inspection to remove outliers and anomalous sequences. Nighttime periods were excluded by retaining only data satisfying $\cos(\theta_z) > 0.1$

Reflectance indices were derived from GOES-16 Band 2 (red channel), corresponding to the visible planetary reflectance, while clear-sky irradiance and modeled meteorological variables were obtained from the National Renewable Energy Laboratory (NREL) database. The meteorological variables originate from the MERRA-2 reanalysis (Gelaro et al., 2017), which provides a physically consistent atmospheric dataset by assimilating satellite and ground-based observations.

The data were structured into temporal sequences of consecutive 10-minute samples. This sequential organization of predictors serves as the input to the proposed recurrent neural network described in the next subsection. The dataset covers three full years, used respectively for training (2019), validation (2020), and testing (2021). Table 2 summarizes the number of valid temporal sequences available for each station, year, and forecast horizon. As can be seen, Concepción del Uruguay consistently exhibits fewer valid samples due to data availability.

Table 2: Number of valid temporal sequences by station, year, and forecast horizon.

Horizon	Luján			Paraná			Concepción del Uruguay		
	2019	2020	2021	2019	2020	2021	2019	2020	2021
1h	2233	2073	2530	2264	2349	2104	2017	1762	1348
2h	1932	1819	2205	1942	2055	1813	1728	1505	1159
3h	1638	1560	1890	1658	1762	1543	1448	1244	938
4h	1351	1307	1588	1380	1467	1280	1179	983	745
5h	1084	1047	1280	1108	1175	1022	929	749	567
6h	805	798	977	843	873	762	696	513	398

2.2. Proposed model

We design a hybrid, dual-input recurrent neural network to forecast hourly global horizontal irradiance (GHI) for lead times from 1 to 6 hours.

The architecture consists of two main branches:

- a sequential branch that ingests short windows of recent observations to capture short-term dynamics, and
- an exogenous branch that encodes spatial and geometric predictors.

Figure 1 illustrates the flow through both branches and the fusion stage. For each forecast, the sequential branch receives six consecutive 10-minute samples per variable (i.e., one hour). Its output is then concatenated with the exogenous feature vector and passed through a fully connected layer, after which a linear output neuron maps the result to the hourly clear-sky index (k_c). The GHI forecast at each prediction horizon is computed as

$$\widehat{\mathrm{GHI}}(t + \Delta t) = k_c \,.\, \mathrm{GHI}_{cs}(t + \Delta t) \qquad \text{(eq . 1)}$$

where $\mathrm{GHI}_{cs}(t + \Delta t)$ is the clear-sky irradiance corresponding to the target hour $t + \Delta t$ at each station and Δt is the forecast lead time (1–6 h) corresponding to each prediction horizon

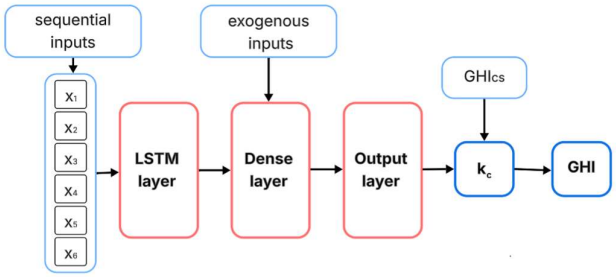

Figure 1: Hybrid dual-input architecture of the proposed model. The sequential branch processes a 1-hour window (six 10-minute samples) through an LSTM, while the exogenous branch provides spatial and geometric features. After concatenation, a Dense layer and a linear output neuron produce the hourly *kc* forecast; multiplying by clear-sky GHI yields the GHI estimate.

To assemble the input sequences for every forecast horizon, six consecutive 10-minute samples of all input variables are taken (representing one hour of data), starting daily at 06:30 local time and repeating throughout the day until the last hour allowed by the selected horizon. An operational delay of 30 minutes is applied between the last input and the beginning of the forecast window. Figure 2 shows the temporal arrangement of the input and forecast intervals. All time labels correspond to the start of each interval (10 minutes for sequential inputs and 1 hour for forecasts).

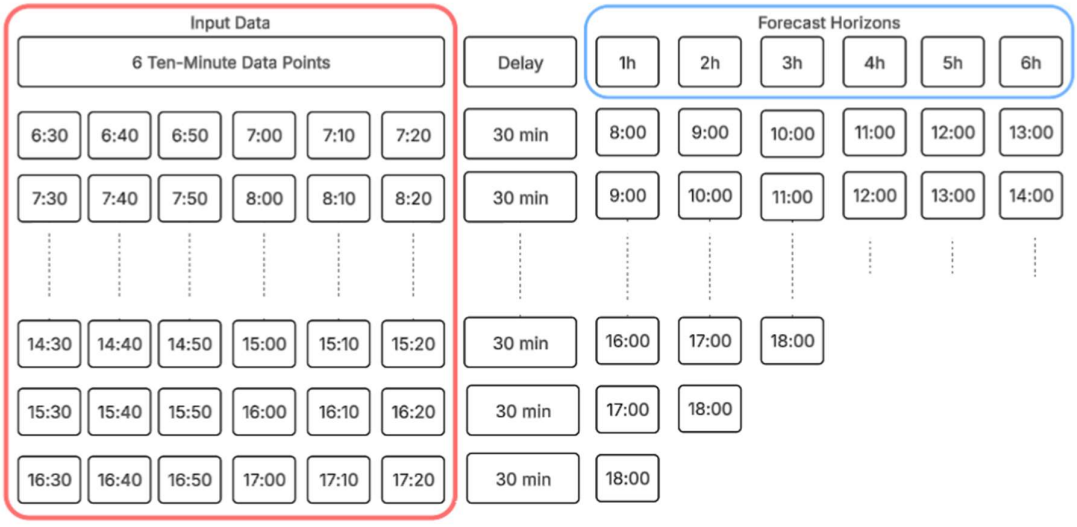

Figure 2: Temporal arrangement of input and forecast windows. Each forecast uses six 10-minute inputs, followed by a 30-minute operational delay and hourly forecast windows for horizons from 1 h to 6 h. All time labels correspond to the start of each interval.

The sequential input is processed by a Long Short-Term Memory (LSTM) layer. LSTM networks are a type of recurrent neural architecture designed to capture temporal dependencies in sequential data. Through gated mechanisms, they selectively retain or discard information over different time spans, effectively mitigating the vanishing-gradient problem present in conventional recurrent networks. This makes them particularly suitable for time-dependent forecasting tasks such as solar irradiance prediction, where the current state depends on several preceding observations.

The LSTM layer comprises 80 units with tanh activation. Its output is concatenated with a vector of exogenous features including the solar zenith angle and its first time derivative, latitude, longitude, and a categorical station identifier represented through one-hot encoding. The fused representation is passed through a fully connected (Dense) layer with 80 units and ReLU activation, followed by dropout regularization. Finally, a linear output neuron predicts the hourly clear-sky index (kc) at the target horizon, which is then multiplied by the corresponding clear-sky GHI to yield the irradiance estimate

Model training uses the RMSprop optimizer (learning rate=0.001) with mean squared error (MSE) as the loss and mean absolute error (MAE) as a monitoring metric. We apply early stopping (patience = 20 epochs, restoring the best weights) and ReduceLROnPlateau to adapt the learning rate when the validation loss stagnates. Each model is trained for up to 300 epochs with a batch size of 40, and the final configuration corresponds to the minimum validation loss. Implementation uses the Functional API of TensorFlow/Keras (v2.18).

Seven configurations of sequential input variables were evaluated, combining the clear-sky index (k_c), reflectance-based indices derived from satellite images at increasing spatial aggregations, and meteorological predictors. The performance of each configuration is analyzed in Section 3.

2.3 Error metrics and persistence model

To evaluate the performance of the forecasting models for each forecast horizon, several error metrics were computed. The most used are the root mean square error (RMSE, eq. 2), the mean absolute error (MAE, eq. 3), and the mean bias error (MBE, eq. 4). In addition, their normalized versions, denoted by the subscript n (e.g., $RMSE_n$, MAE_n, MBE_n), express the errors as a percentage of the mean observed value. These normalized metrics enable a fairer comparison of model performance across different sites or irradiance scales, eliminating the influence of absolute magnitude. The definition of these three metrics is as follows:

$$\mathrm{RMS}E_{\Delta t} = \sqrt{\frac{1}{N}\sum_t (Y^m(t+\Delta t) - Y^0(t+\Delta t))^2} \qquad \text{(eq. 2)}$$

$$\mathrm{MAE}_{\Delta t} = \frac{1}{N}\sum_t |Y^m(t+\Delta t) - Y^0(t+\Delta t)| \qquad \text{(eq. 3)}$$

$$\mathrm{MBE}_{\Delta t} = \frac{1}{N}\sum_t Y^m(t+\Delta t) - Y^0(t+\Delta t) \qquad \text{(eq. 4)}$$

where Y^0represents the observed irradiance, Y^m the predicted value, Δt the forecast horizon, and N the total number of samples.

The performance of each forecasting model is compared with a persistence model to justify the added complexity of the proposed approach.

In the persistence model, the hourly clear-sky index k_c corresponding to the input hour is assumed to remain constant throughout the forecast horizon. The predicted global horizontal irradiance ($\widehat{\mathrm{GHI}}_{\mathrm{per}}$) for a future time $t + \Delta t$ is obtained by multiplying this persisted value of k_c by the clear-sky irradiance ($\mathrm{GHI}_{\mathrm{cs}}$) for the same forecast time, as expressed in eq. 5:

$$\widehat{GHI}_{per}(\mathrm{t} + \Delta\mathrm{t}) = k_c(\mathrm{t}) \,.\, \mathrm{GHI}_{\mathrm{cs}}(\mathrm{t} + \Delta\mathrm{t}) \qquad \text{(eq. 5)}$$

The comparison is expressed through the Forecasting Skill (FS, eq. 6), defined as:

$$FS = 1 - \frac{RMSE_{mod}}{RMSE_{per}} \qquad \text{(eq. 6)}$$

where $\mathrm{RMSE}_{\mathrm{mod}}$ corresponds to the forecasting model and $\mathrm{RMSE}_{\mathrm{per}}$ to the persistence model. The FS value indicates the relative improvement of the model over persistence, according to the following interpretation:

· FS < 0 : the model performs worse than the persistence model,
· FS = 0 : the model performs equal to the persistence model,
· FS > 0 : the model performs better than the persistence model, and
· FS = 1 : the model achieves perfect forecasting performance

3. Numerical results and discussion

In this section, we analyze the intraday GHI forecasts obtained from the proposed hybrid neural network. The model was evaluated under seven distinct input configurations (Config. 1–7), which differ in the composition of the sequential input variables and are summarized in Table 3. These configurations were designed to assess the impact of different combinations of predictors on forecast performance. The sequential inputs include the clear-sky index (k_c), satellite-derived reflectance indices (R_1, R_2, R_3) with varying spatial aggregation, and a set of meteorological predictors (M) comprising relative humidity, wind speed, precipitation, air temperature, and dew point temperature. Additionally, the performance of the best configuration is further examined across different clear-sky index regimes. The reflectance indices (R_1, R_2, R_3) correspond to increasing spatial coverages of approximately 12x12km^2, 30x30km^2, and 90x90km^2, respectively, representing progressively coarser aggregations of the satellite reflectance fields.

Table 3: Sequential-input configurations evaluated. R1, R2, R3 indicate increasing spatial coverage of reflectance índices; M denotes meteorological predictors.

Configuration	Sequential input
1	k_c
2	$k_c + R_1$
3	$k_c + R_1 + M$
4	$k_c + R_2$
5	$k_c + R_2 + M$
6	$k_c + R_3$
7	$k_c + R_3 + M$

Tables 4–6 report RMSE (Wh m^{-2}) for the seven input configurations across the six forecast horizons at Luján, Paraná, and Concepción del Uruguay.

Table 4: RMSE for Luján for different forecast horizons (best results in bold) and inputs configurations.

Configuration	1 h	2 h	3 h	4 h	5 h	6 h
1) k_c	96.8	122.4	136.8	143.9	145.1	139.2
2) $k_c + R_1$	84.4	114.3	127.9	135.9	138.1	133.9
3) $k_c + R_1 + M$	82.2	110.3	120.5	127.3	126.7	124.2
4) $k_c + R_2$	79.0	111.7	124.5	134.2	136.6	133.4
5) $k_c + R_2 + M$	77.9	108.2	119.4	125.7	131.0	123.8
6) $k_c + R_3$	**70.0**	100.0	116.6	126.7	131.6	129.2
7) $k_c + R_3 + M$	70.3	**97.7**	**113.9**	**121.3**	**122.5**	**123.3**

Table 5: RMSE for Paraná for different forecast horizons (best results in bold) and inputs configurations.

Configuration	1 h	2 h	3 h	4 h	5 h	6 h
1) k_c	94.0	122.4	139.2	152.4	149.7	141.4
2) $k_c + R_1$	83.0	112.9	130.9	143.5	144.1	134.7
3) $k_c + R_1 + M$	82.2	109.3	126.0	136.6	135.6	130.0
4) $k_c + R_2$	79.1	110.7	127.7	141.8	142.8	134.3
5) $k_c + R_2 + M$	79.2	107.0	124.7	136.9	135.2	130.3
6) $k_c + R_3$	71.5	101.8	122.2	135.1	138.6	131.9
7) $k_c + R_3 + M$	**70.8**	**100.0**	**118.5**	**131.9**	**132.2**	**129.8**

Table 6: RMSE for Concepción del Uruguay for different forecast horizons (best results in bold) and inputs configurations.

Configuration	1 h	2 h	3 h	4 h	5 h	6 h
1) k_c	102.4	135.3	161.9	176.9	183.2	180.2
2) $k_c + R_1$	91.2	126.3	154.5	169.3	179.3	178.8
3) $k_c + R_1 + M$	88.5	120.4	143.2	157.4	158.0	158.5
4) $k_c + R_2$	86.8	124.8	153.5	171.4	178.9	178.6
5) $k_c + R_2 + M$	85.2	120.4	141.9	155.0	159.2	**157.3**
6) $k_c + R_3$	**78.2**	116.1	142.9	161.3	173.7	173.5
7) $k_c + R_3 + M$	78.8	**113.5**	**136.8**	**150.6**	**153.0**	158.2

.
RMSE generally increases with forecast time, as expected for intraday GHI forecasts. Some performance drops are observed at the 6-h horizon in specific configurations. These are attributable to changes in the evaluation sample at longer horizons (fewer cases and a different distribution of clear-sky hours and conditions), rather than to actual performance improvements. This interpretation is consistent with the increase observed in normalized metrics (e.g., nRMSE%; see Fig. 3), except for the Luján station in the $k_c + R_2 + M$ configuration.

The inclusion of meteorological features further reduces the RMSE in almost all environments, with the most evident benefits at longer horizons (≥3 h). This is consistent with the slower temporal variability of meteorological variables. Some configurations at the 1- and 2-hour horizons show slight increases in RMSE, but from the 3-hour horizon onward, the *"+M "* variants predominate over the configurations without *M*.

At all stations, the best-performing configuration (highlighted in bold in the tables) typically corresponds to the combination using the largest spatial coverage (R_3) together with meteorological predictors (M), i.e., $k_c + R_3 + M$. The only exceptions are the 1-hour forecasts at Luján and Concepción, where the best results are obtained with R_3 alone—consistent with the expectation that meteorological variables contribute less at shorter lead times—and the 6-hour forecast at Concepción, where the lowest RMSE is achieved with R_2.

Figure 3 presents the normalized error metrics expressed as percentages — nMAE (%), nRMSE (%), the coefficient of determination (R^2), and the Forecast Skill index relative to the persistence model.
The configuration that uses only the clear-sky index as sequential input (Model 1, k_c) consistently yields the weakest performance. Even so, it attains a positive—albeit modest—Forecast Skill, particularly at the 1-h horizon and most notably at Concepción del Uruguay. When reflectance indices are added as sequential inputs (Configurations 2–7), all metrics improve: nMAE and nRMSE decrease, while R^2 and Forecast Skill increase. On average, combinations that include reflectance indices with larger spatial coverage (R_3) outperform those based on smaller coverage (R_1 and R_2).

The inclusion of meteorological variables (M) has a limited and not fully systematic effect at the shortest horizons (1–2 h). From the 3-h horizon onward, however, models that incorporate M exhibit a clearer and more consistent improvement across stations.

Concepción del Uruguay shows the lowest forecast skill overall: it is very poor at the 1-h horizon and increases only mildly with lead time, ending around values comparable to the first-horizon skill at Luján and Paraná. For Concepción, both nMAE(%) and nRMSE(%) rise more steeply from the 3-h horizon onward, and R^2 decays faster than at the other sites—except in the configurations that include meteorological inputs, where the degradation is partially mitigated. A plausible explanation is the smaller number of valid samples available at Concepción (cf. Table 2), which may translate into higher variance and reduced model generalization.

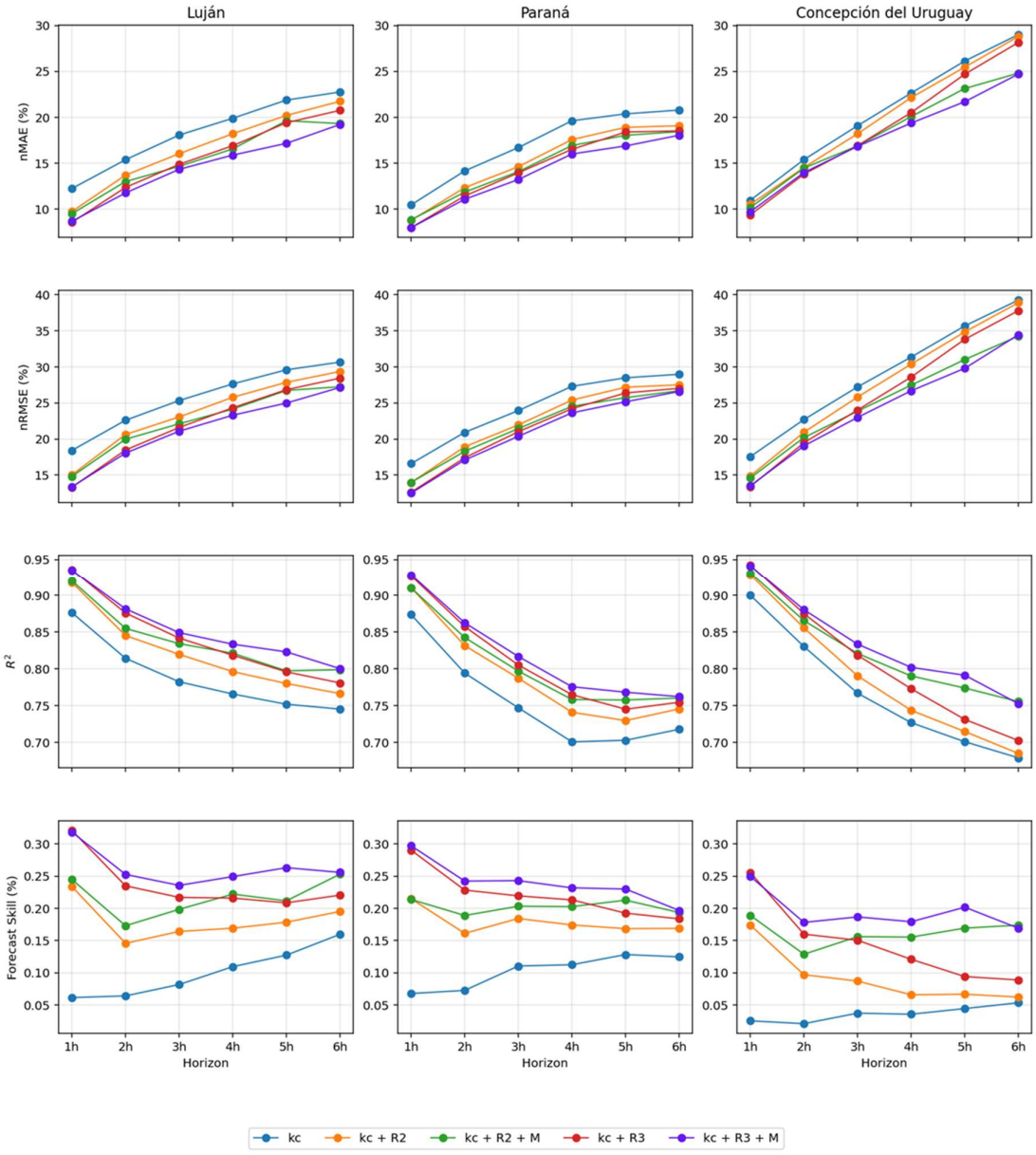

Figure 3: Performance metrics for the hybrid LSTM models at the three stations (Luján, Paraná, and Concepción del Uruguay) as a function of forecast horizon (1–6 h). The panels show, from top to bottom, the normalized mean absolute error (nMAE %), normalized root mean square error (nRMSE %), coefficient of determination (R^2), and forecast skill (FS %) relative to the persistence model. Results are presented for the five configurations of the sequential variables. Adding satellite-derived reflectance indices and meteorological predictors leads to a consistent reduction in relative errors and an improvement in R^2 and forecast skill across stations and horizons.

Finally, Tables 7–9 present the absolute and relative metrics for all three stations under Configuration 7. The performance metrics for Luján and Paraná are quite similar, with Luján exhibiting only a slightly better Forecast Skill. Conversely, the Concepción del Uruguay station shows a comparatively lower performance than the other two. This difference is potentially attributable to the smaller volume of available data for this specific location

Table 7: Hourly forecast performance at the Lujan station: MBE, MAE, RMSE (Wh/m² and %) and Forecast Skill (FS) by lead time (Δt).

Δt (h)	MBE		MAE		RMSE		FS
	(Wh/m^2)	(%)	(Wh/m^2)	(%)	(Wh/m^2)	(%)	(%)
1	-4.3	-0.8	45.5	8.7	70.3	13.4	31.9
2	-12.9	-2.4	63.6	11.8	97.7	18.0	25.3
3	-12.2	-2.3	77.4	14.3	113.9	21.1	23.6
4	-12.0	-2.3	82.6	15.9	121.3	23.3	24.9
5	-6.9	-1.4	84.1	17.2	122.5	25.0	26.3
6	-15.6	-3.4	87.3	19.2	123.3	27.1	25.6

Table 8: Hourly forecast performance at the Paraná station: MBE, MAE, RMSE (Wh/m² and %) and Forecast Skill (FS) by lead time (Δt).

Δt (h)	MBE		MAE		RMSE		FS
	(Wh/m^2)	(%)	(Wh/m^2)	(%)	(Wh/m^2)	(%)	(%)
1	-5.7	-1.0	45.1	8.0	70.8	12.5	29.7
2	-14.3	-2.4	64.7	11.1	100.0	17.1	24.2
3	-5.7	-1.0	76.8	13.2	118.5	20.4	24.3
4	-16.8	-3.0	89.3	16.0	131.9	23.6	23.2
5	-17.1	-3.3	88.6	16.9	132.2	25.2	23.1
6	-32.2	-6.6	88.0	18.0	129.8	26.6	19.7

Table 9: Hourly forecast performance at the Concepción del Uruguay station: MBE, MAE, RMSE (Wh/m² and %) and Forecast Skill (FS) by lead time (Δt).

Δt (h)	MBE		MAE		RMSE		FS
	(Wh/m^2)	(%)	(Wh/m^2)	(%)	(Wh/m^2)	(%)	(%)
1	-23.0	-3.9	56.6	9.7	78.8	13.5	25.0
2	-36.3	-6.1	83.2	14.0	113.5	19.1	17.8
3	-34.6	-5.8	100.1	16.8	136.8	23.0	18.7
4	-30.8	-5.5	109.3	19.4	150.6	26.7	17.9
5	-18.7	-3.6	111.4	21.7	153.0	29.8	20.2
6	-13.9	-3.0	113.4	24.7	158.2	34.5	16.9

4.Conclusions

We introduced a hybrid, dual-input recurrent model for intraday forecasting of global horizontal irradiance (GHI) in the Argentine Pampas. The architecture fuses an LSTM layer—fed with short windows of recent observations—with an exogenous branch that represents site-specific and geometric predictors (latitude, longitude, station identifier) and solar geometry (solar zenith angle and its first time derivative). The sequential inputs included the clear-sky index (k_c), satellite-derived reflectance indices at multiple spatial aggregations (R_1, R_2, R_3), and meteorological variables. The network predicts hourly k_c for lead times from 1 to 6 h, which is then mapped to GHI using the corresponding clear-sky GHI for the target hour.

Augmenting k_c sequences with reflectance information improves forecast accuracy across stations, with larger spatial aggregations (R_3) generally yielding the best performance. The inclusion of meteorological predictors further reduces errors in most settings, with the clearest benefits emerging from 3 h onward, consistent with the slower temporal variability of these variables.

4. References

Aguiar, L. M., Pereira, B., David, M., Díaz, F., Lauret, P., 2015. Use of satellite data to improve solar radiation forecasting with Bayesian artificial neural networks. Solar Energy, 122, 1309 – 1324.

Assaf, A.M., Haron, H., Abdull Hamed, H.N., Ghaleb, F.A., Qasem, S.N., Albarrak, A.M., 2023. A review on neural network based models for short term solar irradiance forecasting. Applied Sciences, 13(14), 8332.

Chen, S., Li, C., Stull, R., Li, M., 2024. Improved satellite-based intra-day solar forecasting with a chain of deep learning models. Energy Conversion and Management, 313, 118598.

Coleto-Rola, C., Alonso-Suárez, R., 2020. Pronóstico de la irradiación solar horaria en Uruguay y sur de Brasil utilizando redes neuronales recurrentes. Anais Congresso Brasileiro de Energia Solar - CBENS.
Jayasankar, K.C., Anandhakumar, G., Kalaimurugan, A., 2024.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., … & Zhao, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of climate*, *30*(14), 5419-5454.

Guijo-Rubio, D., Durán-Rosal, A. M., Gutiérrez, P. A., Gómez-Orellana, A. M., Casanova-Mateo, C., Sanz-Justo, J., ... & Hervás-Martínez, C. (2020). Evolutionary artificial neural networks for accurate solar radiation prediction. Energy, 210, 118374.

Jayasankar, K. C., Anandhakumar, G., Kalaimurugan, A. (2024). Prediction of solar radiation using deep LSTM-based machine learning algorithms. Journal of Environmental Nanotechnology, 13(3), 1–8.

McArthur, L. (2005). Baseline Surface Radiation Network Operations Manual. Technical Report WCRP-121/WMO TD-No. 1274 World Climate Research Programme – WMO.

Marchesoni, F., Alonso-Suárez, R., 2020. Intra-day solar irradiation forecast using RLS filters and satellite images. Renewable Energy. 161. 1140-1154. 10.1016/j.renene.2020.07.101.

Marchesoni-Acland, F., Herrera, A., Mozo, F., Camiruaga, I., Castro, A., Alonso-Suárez, R. (2023). Deep learning methods for intra-day cloudiness prediction using geostationary satellite images in a solar forecasting framework. Solar Energy, 262, 111820.

Nielsen, H., Iosifidis, A., Karstoft, H.,2021. IrradianceNet: Spatiotemporal deep learning model for satellite-derived solar irradiance short-term forecasting. Solar Energy, 228, 659-669.

Rodríguez, F., Martín, F., Fontán, L., & Galarza, A. (2021). Ensemble of machine learning and spatiotemporal parameters to forecast very short-term solar irradiation to compute photovoltaic generators' output power. Energy, 229, 120647.

Thaker, J.; Höller, R.; Kapasi, M., 2024. Short-Term Solar Irradiance Prediction with a Hybrid Ensemble Model Using EUMETSAT Satellite Images. Energies, 17, 329