Characterizing Coastal Stratocumulus Cloud Impacts On Solar Irradiance For Short-Term Forecasting

Gabriela Pallauta Pérez[1,2] and Mónica Zamora Zapata[1,2]

[1] Departamento de Ingeniería Mecánica, Universidad de Chile, Santiago (Chile)

[2] Solar Energy Research Center, SERC (Chile)

Abstract

Short-term solar forecasting is essential for optimizing photovoltaic systems. In coastal regions, the variability caused by morning stratocumulus clouds challenges precise forecasts. We use meteorological data from Antofagasta, Chile, to detect important variables for solar irradiance: correlations highlight a strong influence of wind speed, net radiation, and relative humidity on solar irradiance, and Convergent Cross Mapping identifies a causal influence of dry bulb and dew point temperatures and atmospheric pressure. These variables and their importance are incorporated into a Long Short-Term Memory model with attention to predict time series of solar irradiance, demonstrating an improved performance over the model without attention, achieving lower MSE, MAE, RMSE, and MAPE values, as well as reduced variability across cross-validation folds, indicating both higher accuracy and more consistent results. SHapley Additive exPlanations analysis confirms a greater variable importance for the attention model.

Keywords: Solar forecasting, Coastal stratocumulus, Solar variability, Machine Learning.

1. Introduction

Accurate short-term solar energy forecasting is essential for optimizing the operation of on-grid and off-grid photovoltaic (PV) systems. Effective forecasting can improve grid performance, facilitate efficient energy management and storage, and improve equipment reliability. Machine learning models have proven to be an effective tool for predicting global solar irradiance with great accuracy, addressing the challenges posed by dynamic cloud cover (Voyant, 2017). Moreover, predictive models using machine learning techniques have significantly improved over traditional models, especially under unstable meteorological conditions, such as those observed in the French islands of Corsica, Guadeloupe, and Réunion (Lauret, 2015).

Solar irradiance forecasting becomes particularly complicated by the fluctuating dynamics of coastal stratocumulus (Sc), which can cause abrupt variations in solar radiation, especially during the morning hours. The diurnal cycle of coastal Sc clouds is closely tied to daily solar insolation. These clouds typically form overnight, thin at sunrise, and gradually fragment into broken cloud fields until they dissipate. The thinning and fragmentation of these clouds lead to energy deficits and strong ramps and variability that can disrupt inverter or grid operations. These variations can occur over timescales ranging from minutes to hours (Zamora Zapata, 2020). Recent observational studies have revealed characteristic patterns in the dissipation of Sc clouds (Luccini, 2021), developing methodologies to detect their fragmentation through solar irradiance measurements.

Understanding the complexity of the weather system linked to this type of cloud provides guidance for selecting meteorological variables whose measurement may involve high operational costs. Although more accurate models are often difficult to interpret, causal analysis tools such as Convergent Cross Mapping (CCM) can overcome the limitations of more classical correlation approaches. CCM explores relationships between time series of a pair of variables: to determine the causal relevance of one variable over another, a reconstruction of nonlinear state spaces is performed, evaluating the ability to reconstruct one variable from the patterns within the other variable. As more information is given, a converging trend is expected if there is causality. This exercise allows the identification of causal directionality in high-dimensional systems with nonlinear interactions, noise, or apparently weak correlations (Sugihara et al., 2012; Tsonis, 2017). These advantages motivated its use in our previous study on the interaction between sea breeze and solar irradiance (Pallauta-Perez et al., 2025).

10.18086/swc.2025.04.11 Available at http://proceedings.ises.org

The most popular models nowadays are machine learning models, which, in turn, make predictions by assigning an importance value to each feature. This importance can also be determined in a more informed way, or what we call a model with attention. This is the case of SHAP (SHapley Additive exPlanations) a game-theoretic approach designed to explain the output of any machine learning model in relation to the variables used for a prediction (Lundberg & Lee, 2017).

In this study, we focus on Antofagasta, a coastal city in a desert climate. Antofagasta experiences frequent cloudy days between May and November, with cloud cover reaching up to 50% in August and September, primarily due to coastal Sc clouds (Zamora Aguirre, 2010). These conditions make Antofagasta an ideal natural laboratory for studying the impact of Sc clouds on solar irradiance. For this site, the causality between sea breeze dynamics and global horizontal irradiance (GHI) in coastal Sc dissipation has been previously explored (Pallauta Pérez, 2025); here, we aim to build upon this work by examining both correlation and causal relationships with additional atmospheric variables, thereby complementing traditional correlation analysis with a more robust causal inference approach. This information is then used to enhance forecasting performance by feeding a Long Short-Term Memory (LSTM) model with an attention mechanism. Finally, the novel model behavior is validated through SHAP analysis, resulting in improved short-term solar irradiance forecasting accuracy.

2. **Data and methods**

2.1 Data description

Surface observations were gathered at Cerro Moreno Airport in Antofagasta, Chile (23.4° S, 70.4° W, 135 m above sea level), covering 773 days between December 2013 and February 2016 (Muñoz, 2025). The dataset includes a 5-minute resolution time series of surface temperature (T), relative humidity (RH), downwelling shortwave irradiance or global horizontal irradiance (GHI), dew point temperature (Td), atmospheric pressure (p), wind speed magnitude (w), and wind direction (α). These high-frequency measurements enable the detailed study of boundary layer conditions and the diurnal evolution of coastal stratocumulus clouds.

To incorporate temporal patterns into the model, two cyclic features were added to capture the periodic nature of the day and hour. The day of the year and decimal hour were computed and then transformed into sine and cosine signals to represent their cyclicity continuously, thus avoiding discontinuities at the edges. These features allow the model to recognize seasonal and hourly patterns, improving its predictive capacity by linking meteorological data to its temporal context. Additionally, a standard normalization was applied to all variables by subtracting the mean and dividing by the standard deviation.

Then, a dataset organized into temporal windows was prepared: for each full day (288 five-minute points), the model takes 2.5 hours of historical data (09:30–12:00) as features and issues a forecast with a 2.5-hour horizon (12:00–14:30), ensuring that each window contains exactly 30 points.

2.2 Methodological workflow

Following the steps outlined in Figure 1, we analyze meteorological data to study the cloud fragmentation and dissipation process by analyzing GHI as a proxy for cloudiness. The dissipation process is divided into three stages: starting from a cloudy state at sunrise, the cloud layer first thins uniformly, leading to a progressive increase in GHI toward clear-sky values (Pallauta- Pérez, 2025). Subsequently, the cloud layer transitions into a broken cloud field, introducing intense GHI variability. Finally, full cloud dissipation occurs, and GHI reaches the clear-sky curve.

To better understand the dynamics of cloud dissipation and its impact on solar variability, we employ CCM, a nonlinear state space reconstruction methodology. CCM is designed to identify causal networks and determine the directionality of interactions between variables, distinguishing causality from mere correlation in time series data originating from complex systems (Tsonis, 2017). Its ability to detect causality even in noisy environments and between seemingly uncorrelated variables makes it particularly valuable for this type of study (Sugihara, 2012).

Figure 1:. Overview of the methodological workflow, from data preprocessing and causality analysis to model training and interpretability assessment using SHAP.

2.3 Correlation and Causality

We study the relationship among the variables of interest by analyzing intercorrelations and exploring causality using CCM. Specifically, we examine how atmospheric variables influence solar irradiance, focusing on their effect on GHI time series.

We first determine the optimal parameters for CCM, the time lag (τ) using mutual information, and the embedding dimension (E). The analysis focuses on $\tau = 1$ (original sampling), which allows the assessment of direct causal relationships. This choice provides three main advantages: it maintains consistency with the original sampling scale, enabling straightforward physical interpretation; it captures immediate interactions between variables without temporal mediation; and it facilitates comparison across different variable pairs by using a unified temporal framework.

The nonlinearity of the variables was verified by examining the decay in prediction accuracy of the Simplex method over forecasting horizons ranging from 1 to 10 units, where each unit corresponds to the sampling interval (Sugihara, 2012). To preserve the original patterns and seasonal variability of the variables, the data were not normalized. The robustness of the CCM results was confirmed through the convergence of Spearman's correlation coefficient and its statistical significance ($p < 0.05$), ensuring the reliability of the inferred relationships.

The embedding dimension (E) was determined using the *EmbedDimension* function from the pyEDM library, ensuring adequate reconstruction of the underlying system dynamics. A total of 773 days of data, including both cloud dissipation and non-dissipation periods, recorded at 5-minute intervals, were analyzed. The embedding dimension (E) was systematically evaluated across various library sizes, with the most stable and representative value selected for each variable.

2.4 LSTM Architecture

As shown in Figure 2, the architecture begins with an LSTM layer containing 14 units, which returns the full sequence, producing an output for each of the 30 input time steps, corresponding to a 2.5-hour forecast. This layer includes regularization on both the input and recurrent weights, helping to prevent overfitting by penalizing large or unnecessary parameters.

Next, a *BatchNormalization* layer is applied, which stabilizes and accelerates training by rescaling activations. This is followed by a 20 % Dropout, which randomly disconnects a portion of the neurons during training to enhance model robustness.

The second LSTM block contains 128 units and also returns sequences. This layer slightly reduces the model dimensionality and is regularized more softly than the previous layer. The normalization and Dropout strategy is repeated, this time with a 30 % Dropout rate. The third and final LSTM layer has 64 units and does not return sequences, meaning that it condenses all temporal information into a single vector. This vector is passed through a new *BatchNormalization* and another 20 % Dropout.

The model is trained using a set of input data and corresponding target values for up to 250 epochs with batches of 100 samples. Two additional regularization techniques are applied: *EarlyStopping*, which halts training if the validation loss does not improve for 15 consecutive epochs and restores the best-performing weights, and *ReduceLROnPlateau*, which halves the learning rate if the validation loss does not improve over five epochs. Training is conducted without data shuffling to preserve the temporal structure inherent in the time series.

The code implements an LSTM model for predictive time series analysis, employing a methodology that integrates temporal cross-validation to ensure rigorous and unbiased performance assessment. The TimeSeriesSplit approach ensures that, in each iteration, the training set chronologically precedes the test set, preventing the model from inadvertently "seeing the future" during training. This procedure mitigates overly optimistic performance estimates and preserves temporal consistency throughout the validation process. In this study, the dataset is divided into five sequential, non-overlapping folds, where each test segment follows its corresponding training segment in time, guaranteeing a faithful temporal progression across all splits.

During training, 10% of the training data is allocated for internal validation, allowing the *EarlyStopping* mechanism to monitor performance and terminate training once overfitting is detected. Model performance is evaluated using a comprehensive set of metrics, MSE, MAE, RMSE, R^2, and MAPE, which collectively capture different dimensions of predictive accuracy and error variability, providing a robust assessment of the model's generalization capability over approximately 156 days per fold, covering 773 consecutive days of surface observations spanning December 2013 to February 2016.

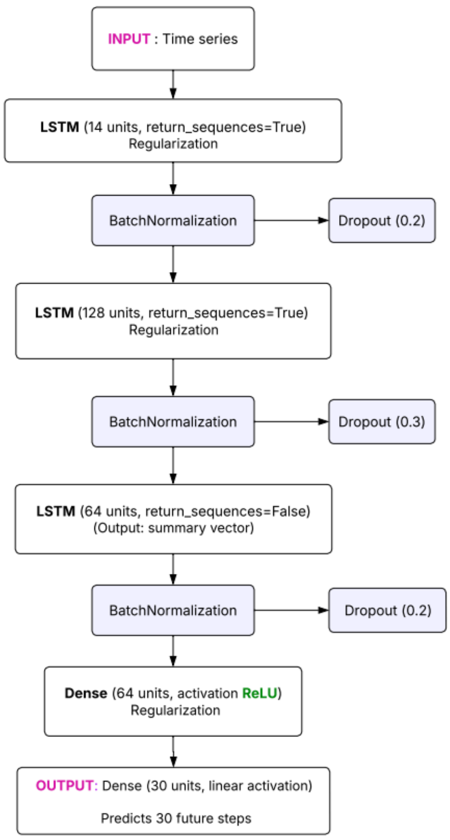

Figure 2: Schematic representation of the LSTM architecture used for solar irradiance forecasting, showing the sequence of LSTM, normalization, and dropout layers, along with the dense output layer that generates the 30-step prediction horizon.

2.4 Performance evaluation

In the evaluation of LSTM models for time series or forecasting problems, it is essential to use appropriate metrics to quantify the model's performance. In this application, the following metrics were used:

- Mean Squared Error (MSE): The MSE measures the average of the squared differences between the predicted and actual values.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad \text{(eq. 1)}$$

where n is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

- Root Mean Squared Error (RMSE): The RMSE is the square root of the MSE.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad \text{(eq. 2)}$$

- Mean Absolute Error (MAE): The MAE measures the average of the absolute differences between the predicted and actual values.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad \text{(eq. 3)}$$

- Coefficient of Determination (R^2): R^2quantifies the proportion of the variance in the dependent variable that is explained by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad \text{(eq. 4)}$$

where $\bar{y}_i$ is the mean of the actual values.

2.5 Attention mechanism

The attention mechanism is a technique introduced in deep learning that allows the model to selectively focus on different parts of the input sequence when generating an output. It helps address the limitation of fixed-size internal memory by dynamically weighting the importance of different parts of the input for each output step.

The architecture used in this study is based on the Transformer model (Vaswani et al., 2017) and its multi-head self-attention mechanism. However, several adaptations are incorporated to modify the original architecture, designed for natural language processing, for application in multivariate time series forecasting. Among these modifications is a causal mask to preserve the autoregressive nature of the prediction and a learnable feature-weighting mechanism. These extensions transform the original language-oriented architecture into a specialized model for time series forecasting with built-in interpretability. In contrast, the baseline model without attention processes the input sequence uniformly through recurrent layers, without explicitly distinguishing the relative importance of past time steps or individual features.

To compute these weights, a normalized vector is required to indicate the initial weight values, which are derived from the causality coefficients obtained through Convergent Cross Mapping (CCM). This means that the attention mechanism is guided by the causal influence of each variable, as quantified by CCM. Then, a weight vector is created, defined as a trainable tensor whose dimension matches the number of features. Variables related to temporal information, such as day of year and hour of day, are assigned initial equal weights of 0.7.

First, learnable weights are applied using a sigmoid function to keep them within the range [0, 1], highlighting the most relevant input features. Next, a temporal attention matrix is applied to ensure causality, preventing attention from being assigned to future time steps. Attention weights are computed by applying a softmax transformation to the masked logits, which represent the raw, unnormalized outputs of the model before activation.

The model then combines both effects by weighting the features with the learnable sigmoid weights and multiplying the weighted inputs by the temporal attention matrix, thereby capturing historical influence over time with varying importance across features. Finally, the method outputs a tensor of identical shape to the input, since the layer maintains both temporal resolution and feature dimensionality. The transformation occurs through two consecutive operations: feature weighting, followed by causal temporal attention.

The training procedure for this model follows the same methodology described in Section 2.4, using identical hyperparameters, optimization strategy, and regularization techniques (*EarlyStopping* and *ReduceLROnPlateau*). Training is likewise performed without data shuffling to maintain the temporal order of the input sequences, ensuring consistent and comparable evaluation across models.

2.3 SHapley Additive exPlanations (SHAP)

SHAP is a technique that quantifies the contribution of each variable to the outcome of a machine learning model. It provides insights into how the model can be improved and helps interpret the underlying phenomenon.

In LSTM models, SHAP enables the interpretation of how each time step and feature influences the result, highlighting key patterns and interactions across different time windows. The signed SHAP values indicate whether the contribution of a variable is positive or negative. However, while SHAP clarifies correlations within the model, it does not imply causality. Predictive models assume that historical patterns will persist, but this does not guarantee that such relationships will hold in the future (Lundberg & Lee, 2017).

The methodology calculates marginal differences between predictions when each feature is included or excluded, averaging across all possible feature combinations. This process ensures desirable mathematical properties such as fairness and consistency, meaning that if a feature's impact increases, its SHAP value cannot decrease.

In practice, the code computes SHAP values to explain how input variables influence the predictions of an LSTM model with multiple outputs. It selects 250 samples from the training set as the background (baseline reference) and 3 samples from the test set to be explained. Because *KernelExplainer* requires flattened data, the three-dimensional inputs are converted into two-dimensional vectors.

Finally, for each of the model's five outputs, a SHAP explainer is trained to estimate the individual contribution of every variable to each prediction. The results are then organized, labeled by output and sample, and combined into a single dataset for interpretation.

3. **Results**

3.1 Correlation and Causality

A correlation analysis identified the most significant relationships between atmospheric variables and GHI, highlighting those with statistically relevant coefficients. In this case, values above 0.5 were observed for variables such as wind speed, net radiation, and relative humidity, suggesting a direct influence of these parameters on GHI variations. However, to complement this traditional evaluation based on Spearman correlations, as shown in Figure 3, the CCM technique was applied. The results show that temperature, despite exhibiting a relatively low correlation coefficient of 0.43, presents a causal relationship with GHI close to 0.8, indicating a stronger influence than suggested by correlation analysis. Likewise, dew point temperature, with an almost null correlation of 0.02, revealed a significant causal influence from GHI, with a unidirectional effect reaching 0.7. This suggests that this variable may play a more relevant role in GHI reconstruction than what is captured by traditional linear methods. Similarly, atmospheric pressure, with a weak negative correlation of –0.13, reaches a causal strength close to 0.7.

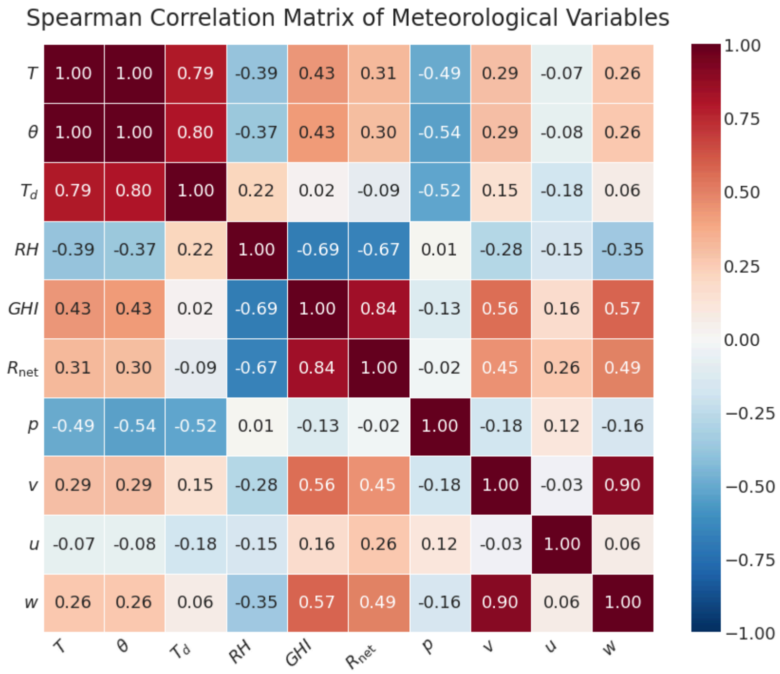

Figure 3: Correlation map for the most relevant variables in the dataset

These results highlight the value of CCM for the analysis of atmospheric time series, as it allows the identification of variables that may have a physically significant effect on cloud dynamics, even when no apparent linear relationship with irradiance exists. In this context, it is confirmed that relying solely on linear correlations could lead to the underestimation of key factors driving solar variability.

Table 1 presents an analysis of the characteristic times (τ) of different meteorological variables, where τ represents the optimal lag based on mutual information between a variable and its time-shifted version. High τ values indicate that a variable changes slowly, showing strong self-correlation over time, while low τ values suggest a faster response to external perturbations.

Table 1: CCM properties for each variable in relation to GHI.

Variables	Embedding dimension (E)	Characteristic time (τ)	Asymptotic correlation coefficient for causality (ρ)
RH	2	8	0.9
R_{net}	2	10	0.95
p	4	10	0.7
T	2	16	0.8
θ	2	16	0.8
T_d	3	21	0.7
u	4	5	0.7
v	4	11	0.5
w	4	10	0.7

It is important to note that the results from Table 1 do not necessarily imply a purely causal mechanism in a strict sense, but rather reflect the degree of dynamic coupling between variables. In CCM, the reported values indicate how well GHI can be reconstructed from each variable, that is, the extent to which each atmospheric parameter contributes to or shares information with irradiance dynamics. High causality coefficients, such as those for RH and Rnet, suggest that these variables are strongly coupled with GHI, likely acting as primary drivers or co-evolving components within the same physical system. Therefore, the CCM analysis captures not only causal influences but also mutual dependencies, offering a more comprehensive view of the underlying interactions that govern solar irradiance variability.

3.2 Forecasting performance

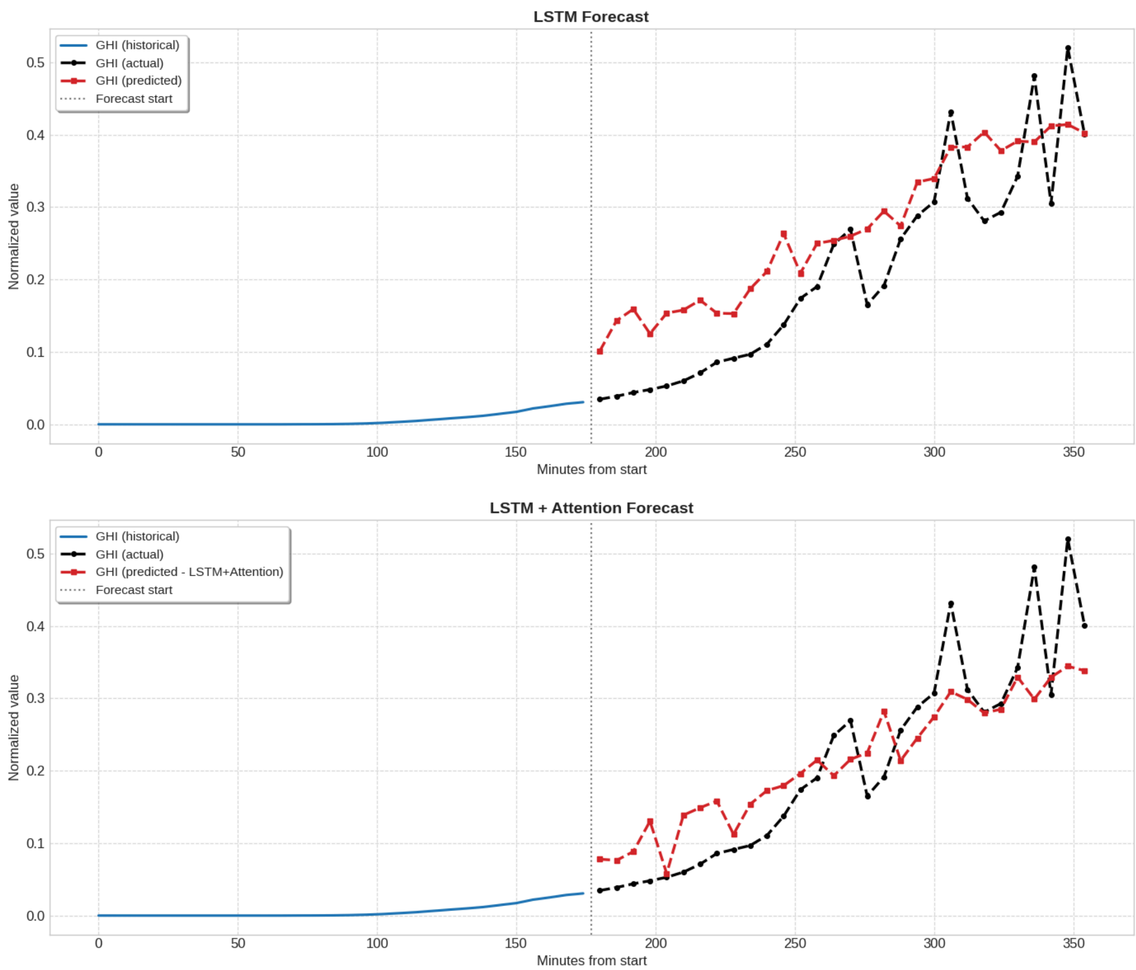

Figure 4: Comparison between the standard LSTM model (top) and the LSTM with the Attention mechanism (bottom) for a sample forecast window. Both plots display the historical GHI values (blue), the actual measured GHI during the forecast period (black dashed line), and the predicted GHI (red dashed line). The vertical gray dotted line indicates the transition point between the historical input window and the forecast horizon.

When comparing the forecasts produced by both models in Figure 4, the version incorporating the attention mechanism performs slightly better across nearly all cross-validation results. The MSE (0.0135 vs. 0.0138), MAE (0.0916 vs. 0.0922), and RMSE (0.1149 vs. 0.1156) are marginally lower, indicating an improved ability to minimize both absolute and squared errors. Similarly, the MAPE decreases from 0.5981 to 0.5828, reflecting enhanced relative prediction accuracy. Both models were trained for a comparable number of epochs (118.8 and 116.4 on average); however, the attention-based model exhibited lower standard deviations across folds, suggesting more stable and consistent performance throughout the temporal cross-validation process.

3.3 Assessing variable influences

By performing the SHAP analysis, we are able to assess the most influential variables in each forecasting model. In the LSTM model without attention, thermal variables, such as temperature and potential temperature, emerge as the most influential, while wind components and wind speed contribute significantly at specific times. Variables like relative humidity exhibit transient spikes in importance, whereas global radiation and atmospheric pressure maintain more stable but relatively minor effects.

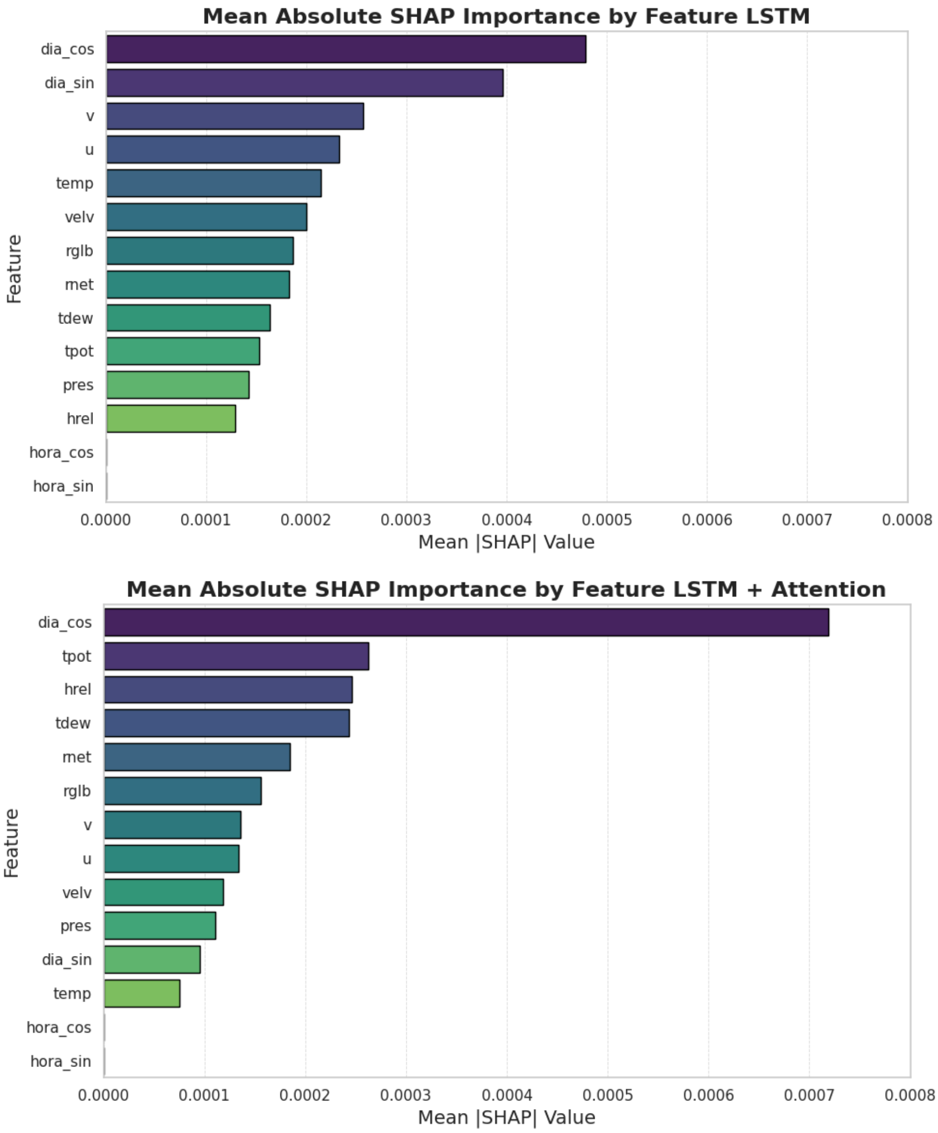

Figure 5: Mean absolute SHAP value for each input variable. The figure illustrates the average contribution of each feature to the model's predictions, obtained by aggregating SHAP values across all time steps and cross-validation folds. Higher bars indicate greater influence on the model's output. Features are ranked in descending order of importance, highlighting those most relevant to the forecasting task.

As shown in Figure 5, the attention-based model displays a differentiated behavior: variable importance is generally higher, reflecting the mechanism's ability to better identify and prioritize the most relevant features. Net radiation gains increased relevance, indicating improved capture of key physical relationships. Among the wind-related variables, mean wind speed stands out as the most influential, surpassing the zonal and meridional components. The hourly variable retains low importance, consistent with the model without attention. Overall, the enhanced discrimination in variable importance observed in the attention-based model suggests a deeper understanding of underlying relationships and more stable learning, ultimately leading to improved forecasting accuracy.

4. Conclusions

In conclusion, the incorporation of the attention mechanism into the LSTM model resulted in measurable improvements in short-term solar forecasting. The attention-based model achieved lower MSE, MAE, RMSE, and MAPE values, indicating enhanced accuracy in both absolute and relative error metrics. Although both models have room for improvement, the attention model showed a smaller deficit, reflecting better explanatory capacity. Additionally, the attention mechanism contributed to more consistent performance across cross-validation folds, as evidenced by reduced variability in the metrics.

Thermal variables dominate in the LSTM model without attention, while wind components and humidity show variable influence. The attention-based model highlights net radiation and mean wind speed as most important, demonstrating improved feature prioritization, more stable learning, and enhanced forecasting accuracy.

Future research will evaluate different forecasting configurations by modifying the input information and forecasting horizon. Additionally, we should continue exploring the implementation of machine learning with attention models for solar irradiance forecasting, evaluating their performance in terms of predictive capability and interpretability of the involved variables across different time horizons, and validating them in multiple locations under diverse local cloudiness conditions.

Additionally, attention mechanisms should explore the temporal scales at which these causal relationships can influence irradiance behaviour, considering the bidirectional nature of interactions and the potential for distinct temporal dynamics across system variables.

5. Acknowledgments

This research was funded by ANID Fondap Solar Energy Research Center (SERC-Chile) project number 1523A0006.

6. References

Lauret, P., Voyant, C., Soubdhan, T., David, M., Poggi, P., 2015. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. Solar Energy 112, 446–457. https://doi.org/10.1016/j.solener.2014.12.014.

Luccini, E., Rivas, E., 2021. Kinematics and dynamics of the stratocumulus average diurnal dissipation process at Atacama Desert. Atmospheric Research 255, 105523. https://doi.org/10.1016/j.atmosres.2021.105523.

Muñoz, R. Surface and Ceilometer Data-Cerro Moreno, Antofagasta, 2025. Available online: https://zenodo.org/records/14903408 (accessed on 20 October 2025).

Pallauta Pérez, G., Muñoz, R.C., Arrué, P., Zamora Zapata, M., 2025. The role of sea breeze in the coastal stratocumulus dissipation transition in Antofagasta and San Diego. Atmosphere 16, 437. https://doi.org/10.3390/atmos16040437.

Sugihara, G., May, R., Ye, H., Hsieh, C.h., Deyle, E., Fogarty, M., Munch, S., 2012. Detecting causality in complex ecosystems. Science 338, 496–500.

Tsonis, A., 2017. Convergent cross mapping: Theory and an example. In: Advances in Nonlinear Geosciences. Springer, Cham, pp. 587–600.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., Fouilloy, A., 2017. Machine learning methods for solar radiation forecasting: A review. Renewable Energy 105, 569–582. https://doi.org/10.1016/j.renene.2016.12.095.

Wood, R., 2012. Stratocumulus clouds. Monthly Weather Review 140, 2373–2423.

Zamora Aguirre, R., 2010. Caracterización observacional de la capa límite marina en Antofagasta. Universidad de Chile. Disponible en: https://repositorio.uchile.cl/handle/2250/102299.

Zamora Zapata, M., Norris, J., Kleissl, J., 2020. Coastal stratocumulus dissipation dependence on initial conditions and boundary forcings in a mixed-layer model. Journal of the Atmospheric Sciences 77, 2717–2741. https://doi.org/10.1175/JAS-D-19-0254.1.

Zelinka, M.D., Randall, D.A., Webb, M.J., Klein, S.A., 2017. Clearing clouds of uncertainty. Nature Climate Change 7, 674–678.Van der Geer, J., Hanraads, J.A.J., Lupton, R.A., 2000. The art of writing a scientific article. Journal of Scientific Communication. 163, 51-59