Forecasting Solar Radiation Using a Long-Term Memory Neural Network for Lima - Peru

A Romero* and A Guevara*

Faculty of Sciences, National University of Engineering (UNI), Lima, Peru

* Email: and.romero@uni.pe

Abstract. Weather patterns have been of great interest for a long time. Especially today, where clean energy is a must. In this research an approach to predict solar radiation using deep learning algorithms is presented. To do this, we use data obtained from NASA POWER VIEW for 11 years in the city of Lima, we process all the data and we use a recurrent neural network (RNN) architecture, mainly we use the Long-Short Term Memory (LSTM). This will allow our RNN to learn the patterns and can give us an accurate prediction, likewise we will validate these results by regression metrics. The results show that studying solar radiation should not be limited to the time series in days, but rather use time frequencies, such as weeks and months. According to these results, our RNN architecture gives us optimal performance, adapting to all trends and approaching the random values of solar irradiation.

1. Introduction

Nowadays, most of the resources available to generate energy are limited and will be depleted due to the demand for them. Additionally, these resources entail problems of water and soil contamination as a result of their use, these resources are called fossil fuels. Their short and long-term impacts negatively influence society and the ecosystem. [11]

According to the GLOBAL STATUS REPORT ON RENEWABLES 2021, energy demand has been increasing between 2009 and 2019. Renewable energies have only covered a quarter of this increase. In addition, a majority growth in the use of solar PV can be observed between 2008 and 2018 [18]. This happened, because in recent years the costs of producing wind and solar energy have been significantly reduced. [12]

Solar energy is currently one of the main energy sources with the greatest development in the future. It is a renewable resource, causing a lower environmental impact compared to other types of energy sources [1]. Currently, technologies based on the collection of solar energy are being used to generate electricity [10]. Technologies such as solar PV, thermal collectors, and others [13]. These approaches have already been proven and are widely practiced throughout the world as renewable alternatives to conventional non-hydroelectric technologies.[10]. These mechanisms, energy collectors, depend on the meteorological conditions and the limits of the solar PV [3]. The amount of solar irradiance that reaches a photovoltaic cell is the main reason in the output power performance. The solar irradiance in a photovoltaic cell maintains dependence with geographical location, the weather and panel orientation with respect to the sun as well as the sky. [2]

Taking knowledge and predicting weather patterns have been of great interest for many years, an example of this is the work entitled "Deterministic nonperiodic flow" by Lorenz (1963) [14], where weather modeling is carried out assuming that it is chaotic and reducing it to a system of nonlinear coupled partial equations. The solutions of these equations can give a weather approximation, but they are very sensitive to errors in measurements by electronic instruments. [15]
This research aims to predict the climate, specifically solar irradiance, studied from deterministic data, that is, a time series. In this sense, a time series can be analyzed using a neural network architecture, specifically a neural network that can store data for long periods of time and use it to create dependencies with the most recent values.

A RNN architecture was implemented for time series in order to forecast daily, weekly and monthly solar radiation in Lima by collecting satellite data and validating using regression metrics.

The prediction of the movement that solar irradiation would have has various applications in solar systems, such as the management of resources, the monitoring of control systems and their maintenance. Therefore, it is of vital importance to promote the economic viability and efficiency of solar-based systems.

2. Methodology

In this section, we will review the methodology used to study solar irradiance prediction. We propose a scheme, as shown in Figure 1. We begin with the data collection from our study site. Subsequently, we carry out a traditional data treatment using Python. Then, we incorporate the selected and processed data into our recurring neural network model. Finally, we validate the results with the metrics.

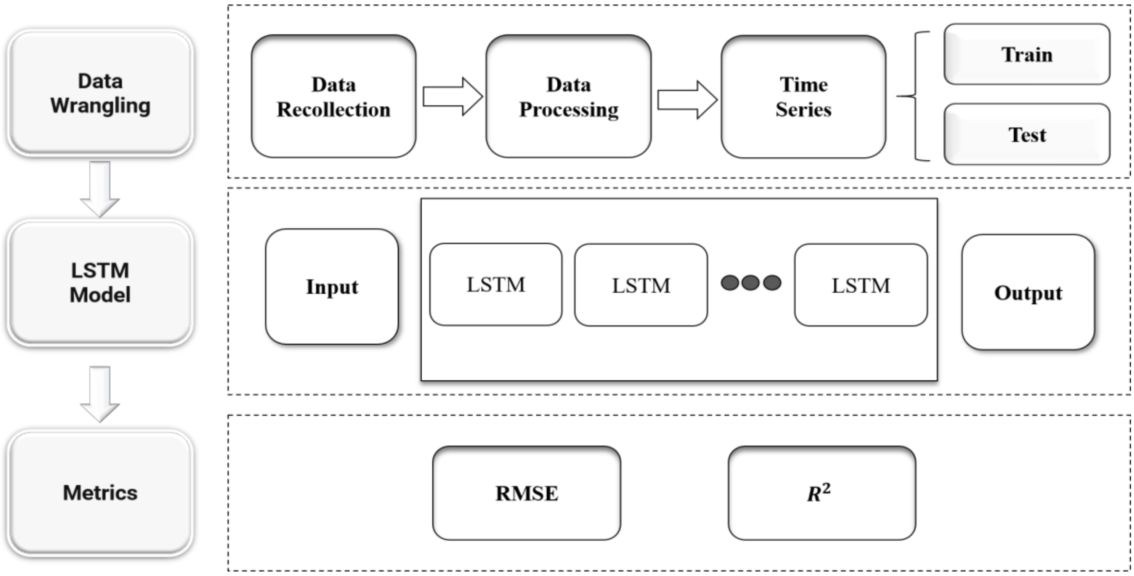

Figure 1. Procedure of the applied deep learning methodology.

2.1. Data Recollection

The data obtained from solar irradiance were based on satellite data. Due to the limited availability of open data, it was decided to use data from NASA satellites through Power View. We obtained data on solar irradiance in a daily period of about 11 years (from January 2010 to February 2021), these data consist of incidents of solar radiation on a horizontal surface (kWhr/m^2/day).

To use POWER VIEW, we must specify the coordinates of our study place, as shown in Figure 2, for our study we are located in the city of Lima-Peru, specifically in the Faculty of Sciences of the National University of Engineering, with latitude of -12.0170 and longitude of -77.0506. POWER

VIEW [4], will provide us with various types of climatological data, only data from incidents of solar radiation on a horizontal surface were collected.

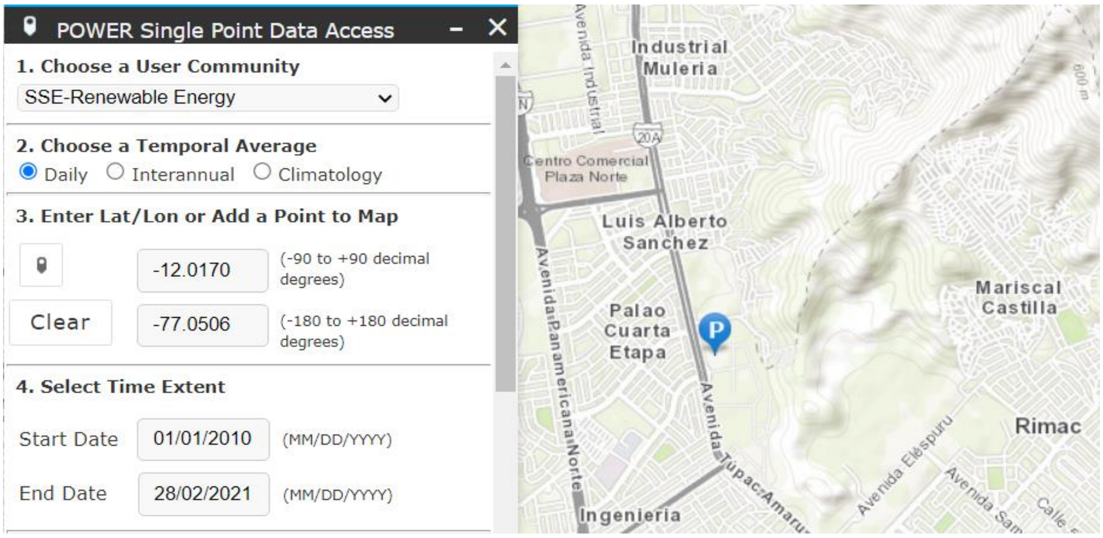

Figure 2. Solar irradiance extraction using Longitude and Latitude for the study area. These data was obtained from the NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science/Applied Science Program.[4]

2.2. Time Series Analysis

Data is the main source of research, a particular approach is to study it to find patterns, correlations and dependencies between variables. In this work we use Python in the processing of solar irradiance data. Analyzing trends, averages and plotting the results.

It began by cleaning the data and plotting the solar irradiance over time, as shown in Figures 3, 4 and 5. By viewing Figure 3, we can observe the maximums and minimums that it presents in various years, these being related to the seasons. weather of the year. Likewise, a histogram showing the frequency of solar irradiance values per day (Figure 4). Where it is shown that there are two peaks, which indicate the probability of a maximum and a minimum in a cyclical way.

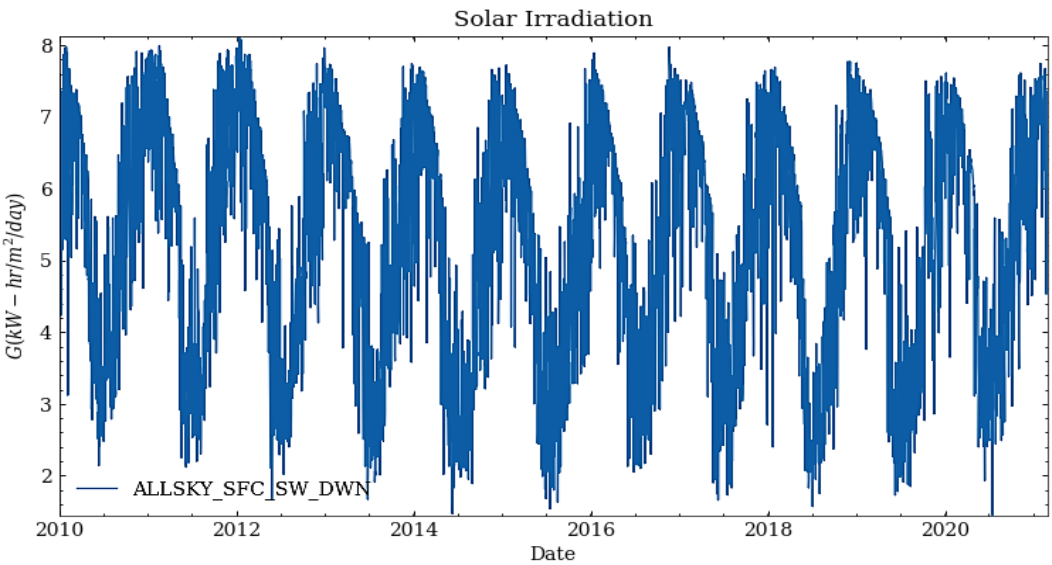

Figure 3. Graph of the time series that forms the solar irradiance in our study area. Data obtained from the NASA Langley Research Center (LaRC) POWER Project.[4]

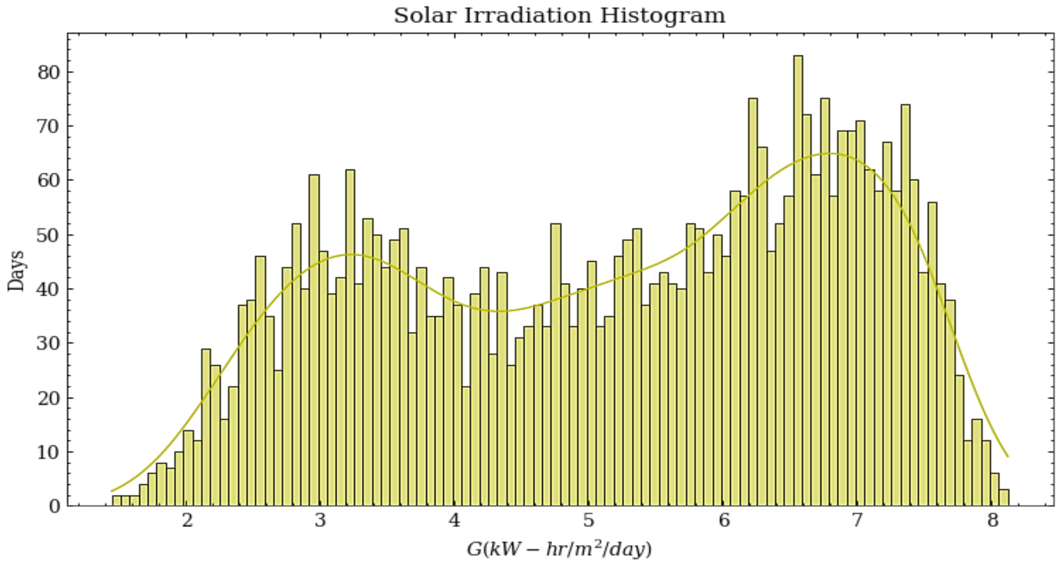

Figure 4. Histogram of solar irradiance. These data was obtained from the NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science/Applied Science Program.[4]

Knowing how the solar irradiance changes over time is a support to identify the trend that it will take in the coming years. Figure 5 shows the annual average, a decrease was observed in the years 2014 and 2015.

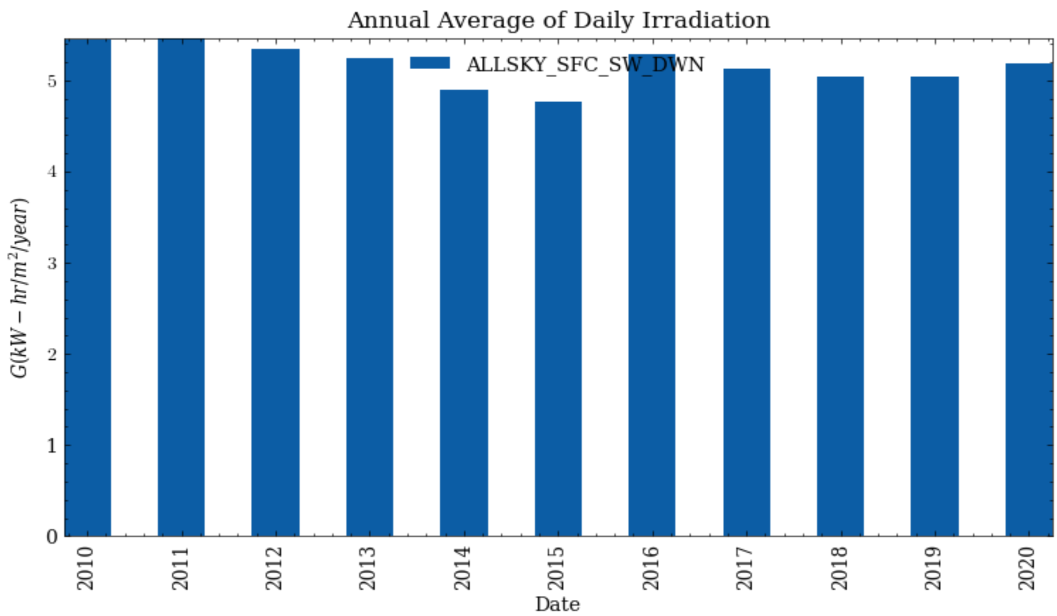

Figure 5. Annual average of solar irradiation. These data was obtained from the NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science/Applied Science Program.[4]

2.2.1. Hodrick-Prescott Filter. This filter allows us to get a smooth curve about the data provided. Likewise, it gives us a trend that is usually close to what is expected.
Let y_t a time series for $t = 1,2,3,\ldots,T$. If $s_t$is the series trend, then the measure of cyclical fluctuations is given by: $c_t = y_t - s_t$

The authors propose that the following equation is minimized by time series trend component.[6]

$$\sum_{t=1}^{T}(y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1}((s_{t+1} - s_t) - (s_t - s_{t-1}))^2$$

Where λ is the penalty parameter and controls the smoothness of the time series.

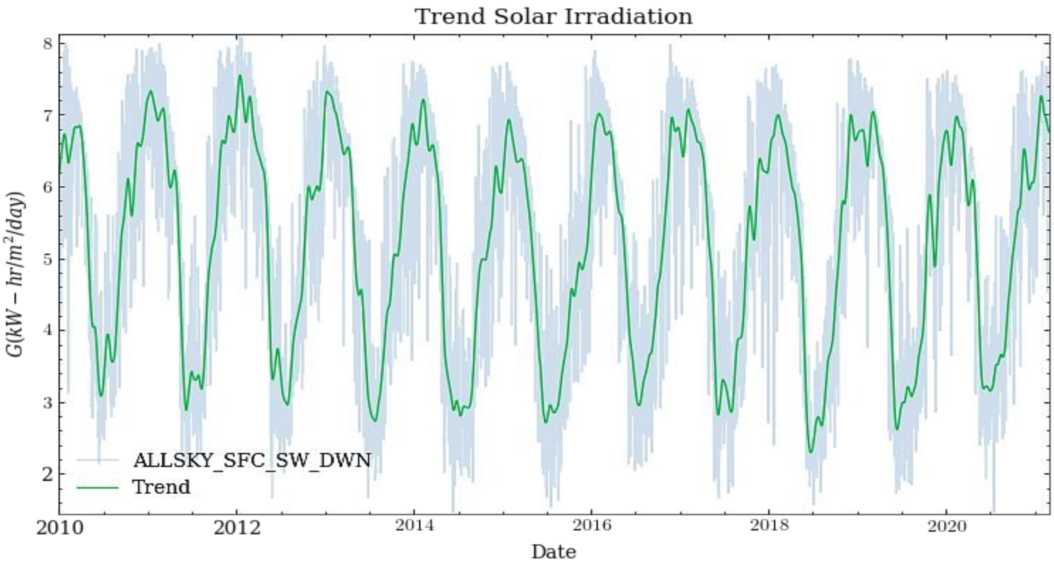

Figure 6. Extraction of the solar irradiance trend using the Hodrick-Prescott Filter.

2.2.2. Rolling Window: Estimating the parameters in a moving window with a fixed size through the time series helps us to evaluate the constancy of the parameters of a model. If the estimated parameters are constant, the difference in the moving windows should be close. If at any point the parameters change, continuous estimates should capture this instability. [7]

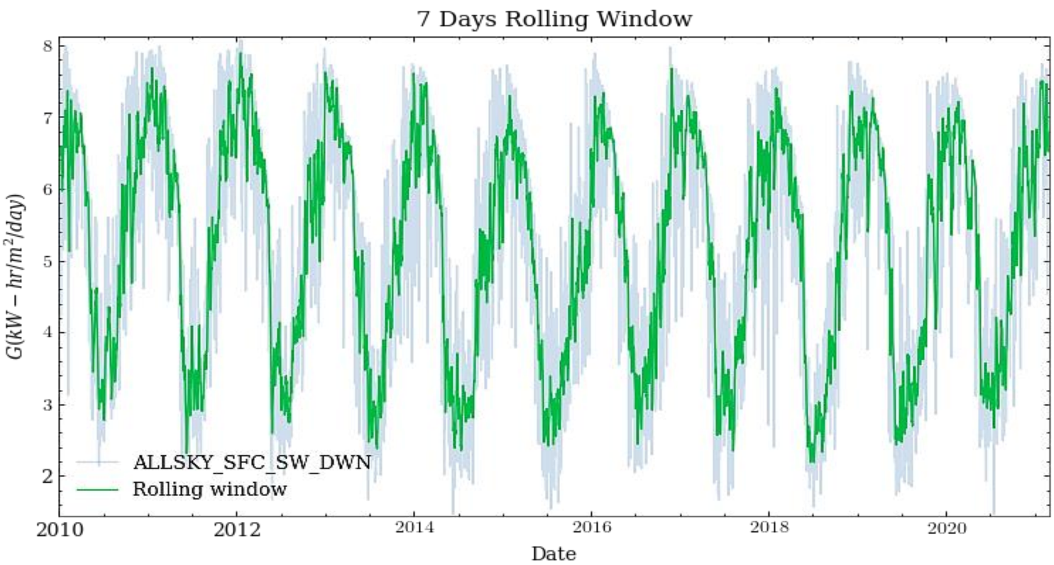

Figure 7. Graph of a rolling window with parameters of 7 days.

An interesting section is the decomposition of time series. Decomposing the time series in trend, into predictable changes that occur over a period of time (seasonality) and the residual (difference between trend and seasonality), provides useful information to better understand the analysis of the collected data.
Figures 8,9 and 10 are decompositions into days, months, and years for solar radiation time series. In our case we use an additive model.

$$x_t = trend + seasonality + residual$$

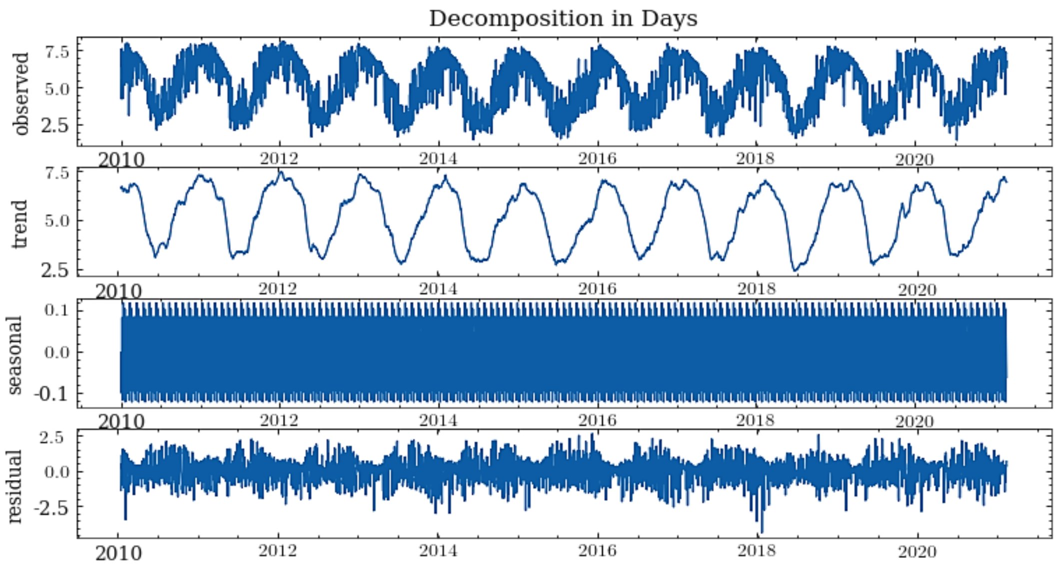

Figure 8. Decomposition into base elements of the time series in days.

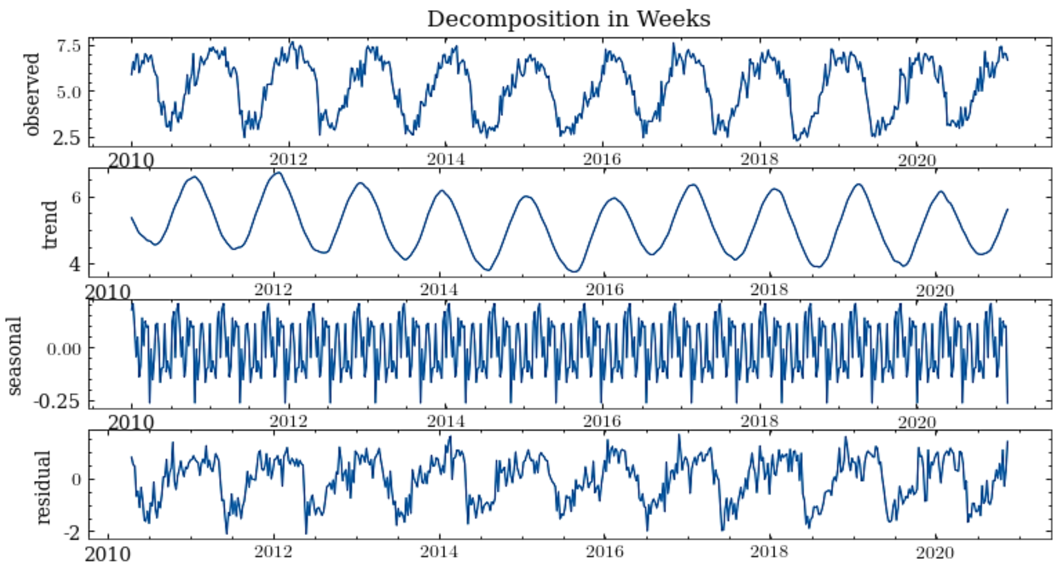

Figure 9. Decomposition into base elements of the time series in weeks.

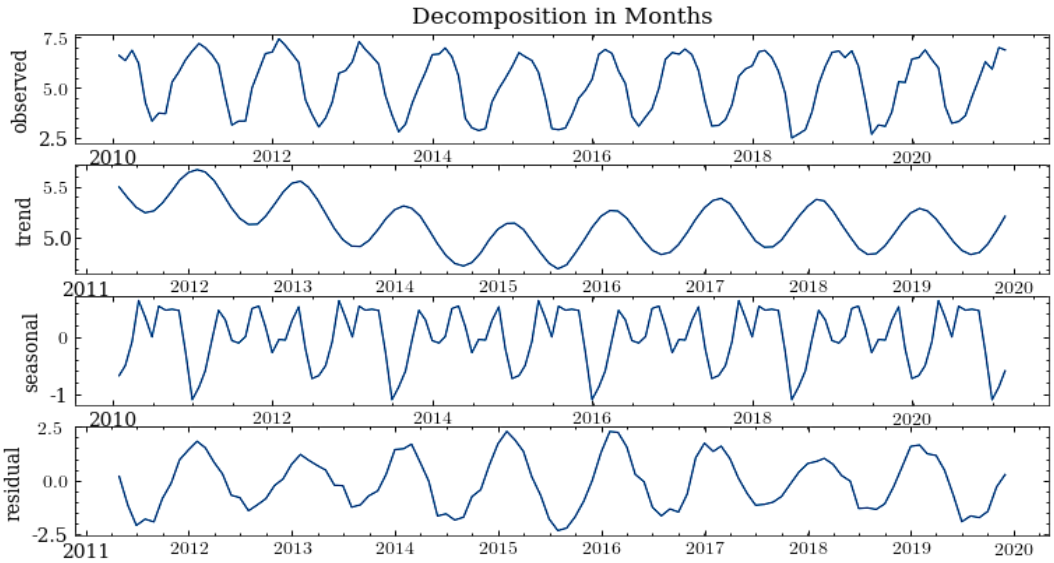

Figure 10. Decomposition into base elements of the time series in months.

2.3. Train/Test Split

Once the data has been analysed, through its trend, its frequency and seasonality. Next, the data is separated into training data and testing data. This is common and necessary in methodologies based on machine learning. Thus, we take 80% of the data already processed to train and the remaining 20% to validate our training.

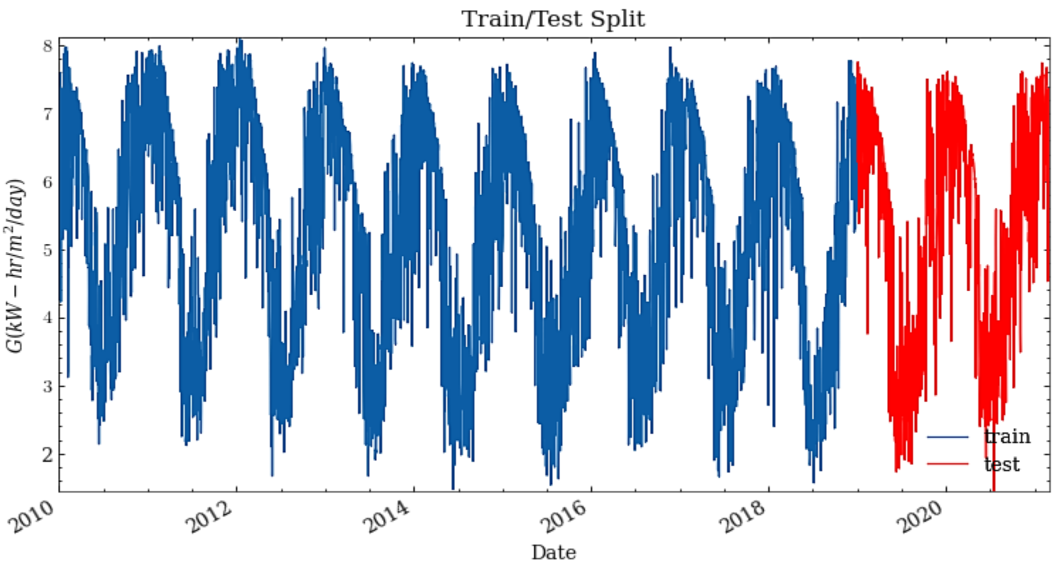

Figure 12. Data split into training and validation data for our LSTM model.

2.4. LSTM Architecture

The Long short-term memory (LSTM) architecture is a kind of recurrent neural network (RNN) that, unlike its predecessor, allows to improve long-term dependencies because it maintains the connection between what was learned in initial layers. [8]

In this long-term memory, terms can be updated with new adhesion in the training process in each of its layers, which gives us versatility for the following processes. As we can see in the image, each network consists of four layers.

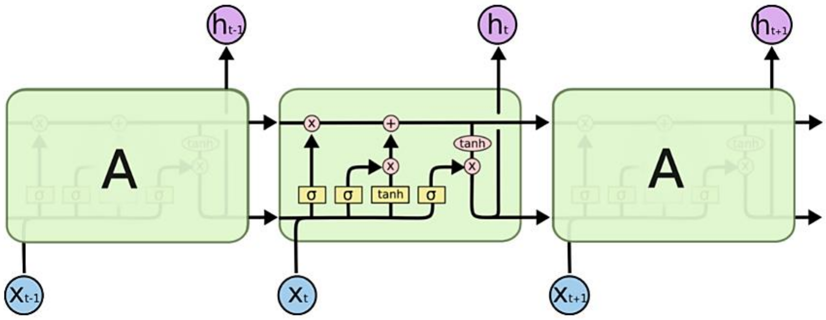

Figure 12. LSTM Architecture [5]

In the first, we can allow what information we are going to keep or eliminate from a previous process. h_{t-1} ,x_t values are from a previous process and W_f, b_f values are trained.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

In the second layer we have two processes in one we decide that we are going to save in the cell and in the other which would be the candidates to be updated

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\hat{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

For the third layer is when we decide to add a new state update of this cell.

$$C_t = f_t \times C_{t-1} + i_t \times \hat{C}_t$$

As a final part of the process, we have to generate our output and status values that will be used by the next cell.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \times tanh(C_t)$$

2.5. Regression Metrics

2.5.1. Root Mean Square Error (RMSE)

It is a measure that allows us to determine the difference between the predicted values for our model. [9]

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}}$$

Where y_t are the predicted values.

2.5.2. Coefficient of determination (R^2)

With this measure, we can determine the quality of our predictions when compared to the actual values. The closer R^2 is to one, the better predicted our model will be.[15]

$$R^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2}$$

For which $\sigma_{XY}$is the covariance for (X,Y), σ_X is the variance of the variable X and σ_Y is the variance of the variable Y.

3. Results & Discussion

In this section, the results obtained by predicting solar irradiance through the time series are presented and discussed.

The data collected, through the NASA satellite in 11 years, were processed and applied in the specific LSTM model. The performance of the LSTM can be seen in Figure 13, where at a glance we can see how our prediction maintains trends. These trends were previously analyzed, and now we corroborate the seasonal and cyclical form of the time series. Our metrics indicate that the performance in the coefficient of determination $R^2 = 74\%$, as shown in Table 1. Contrasting with the results obtained by Banalxmi et al. [18] for a single city, they show correlation metrics of up to 72% in models such as LSTM, GRU, CNN; our LSTM derived model has a higher correlation.

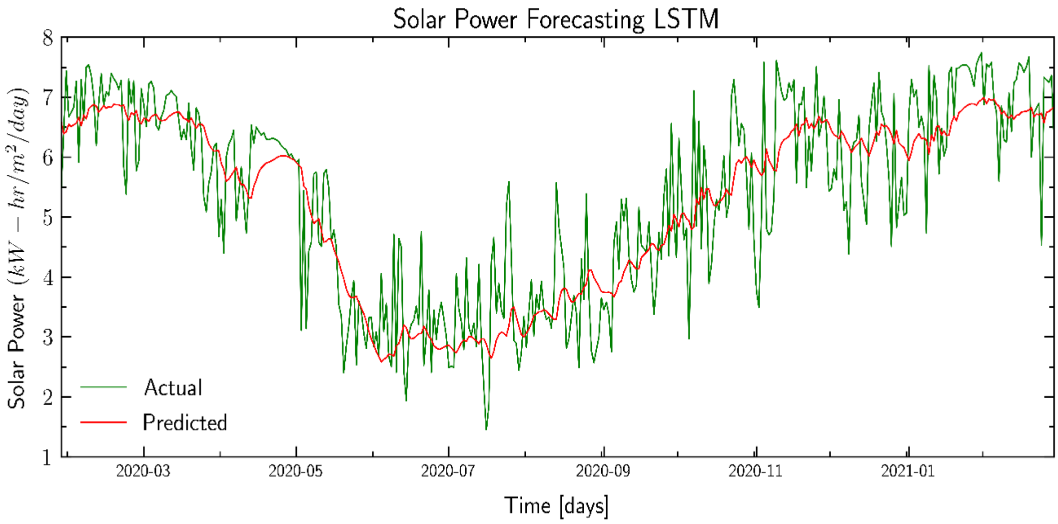

Figure 13. Comparison between the prediction of the solar irradiance in 360 days and its validation value.

Analyzing a time series can be taken to different time frequencies, therefore, we previously analyzed and now we predict the values of solar irradiance for weeks and for months.

Having an abundance of data is essential for deep learning, as shown in Figure 14A, it has been predicted with acceptable efficiency and this correlates with Figure 14B, where convergence is much faster compared to the other frequencies of weather. The more we shorten our frequency, the smoother the curve; as we can see in Figures 14C, 14E. Visually the prediction improves, but convergence takes much longer compared to the daily model, as the figures 14D and 14F shows.

A numerical value is shown for each time frequency; thus, we have a comparative table with the results of the metrics for each prediction. As can be seen, the best result was obtained with a frequency in months, where the R^2 and the RMSE has a direct relationship in this optimization. The results in weeks and days have a moderate efficiency due to the randomness in their time series, being more complicated to predict the next step with optimal accuracy. In Figure 15, we can see the productions with an arrangement of the standard deviation, where we have our best adjusted data for the monthly time series. Other algorithms, such as those studied in Lunche Wang et al. [19] have a higher efficiency when evaluated in daily frequencies. The proposed algorithm can equalize them when used in other frequencies such as the monthly. The smoother the curve, the better the prediction.

Table 1. Comparison between time frequencies.

LSTM	R^2(%)	RMSE
Days	74	0.811
Weeks	86	0.512
Months	92	0.422

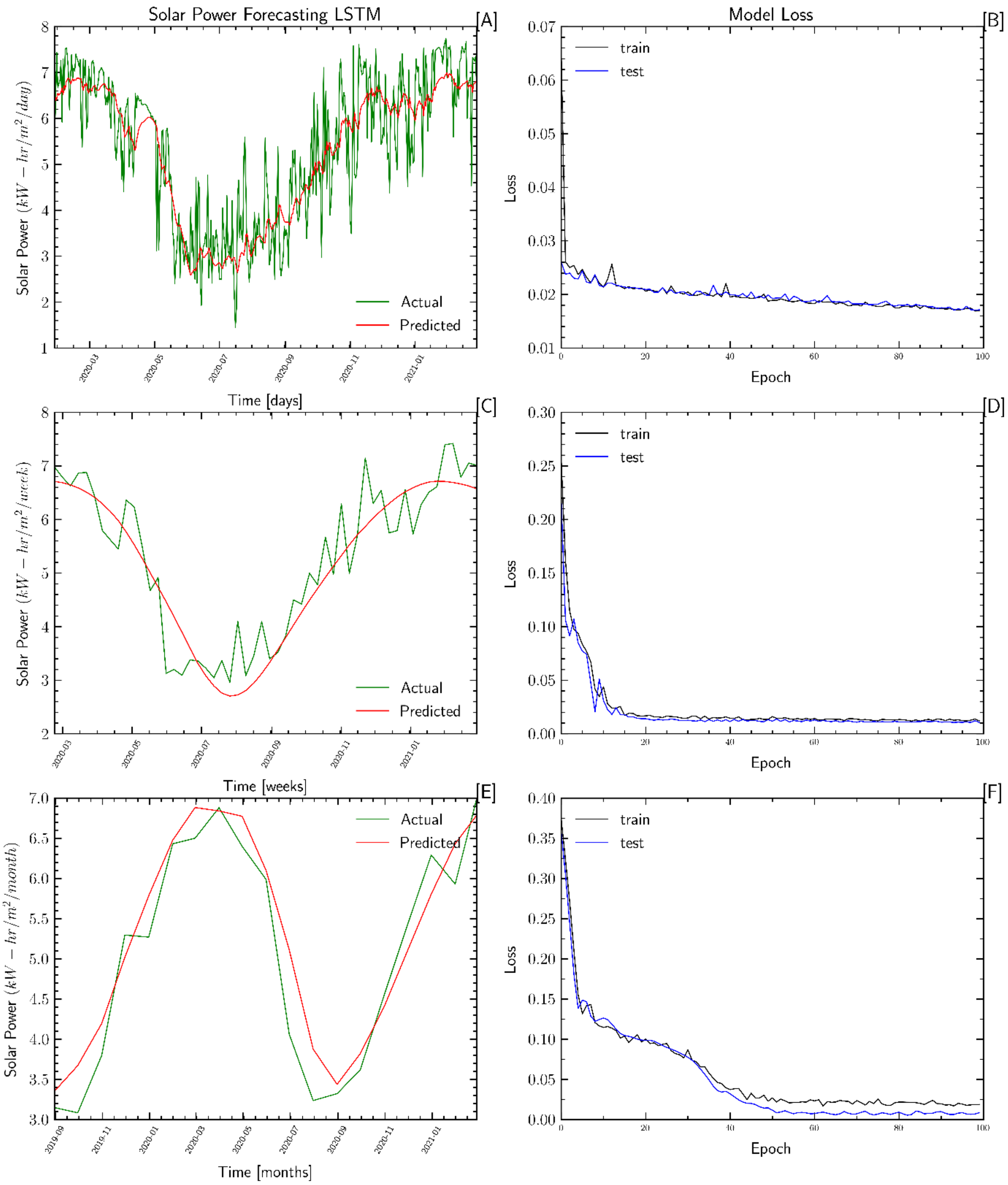

Figure 14. Analysis and prediction of solar irradiance in time frequencies. A) daily prediction, B) cost function in training in days, C) weekly prediction, D) cost function in training in weeks, E) monthly prediction, F) cost function in training in months

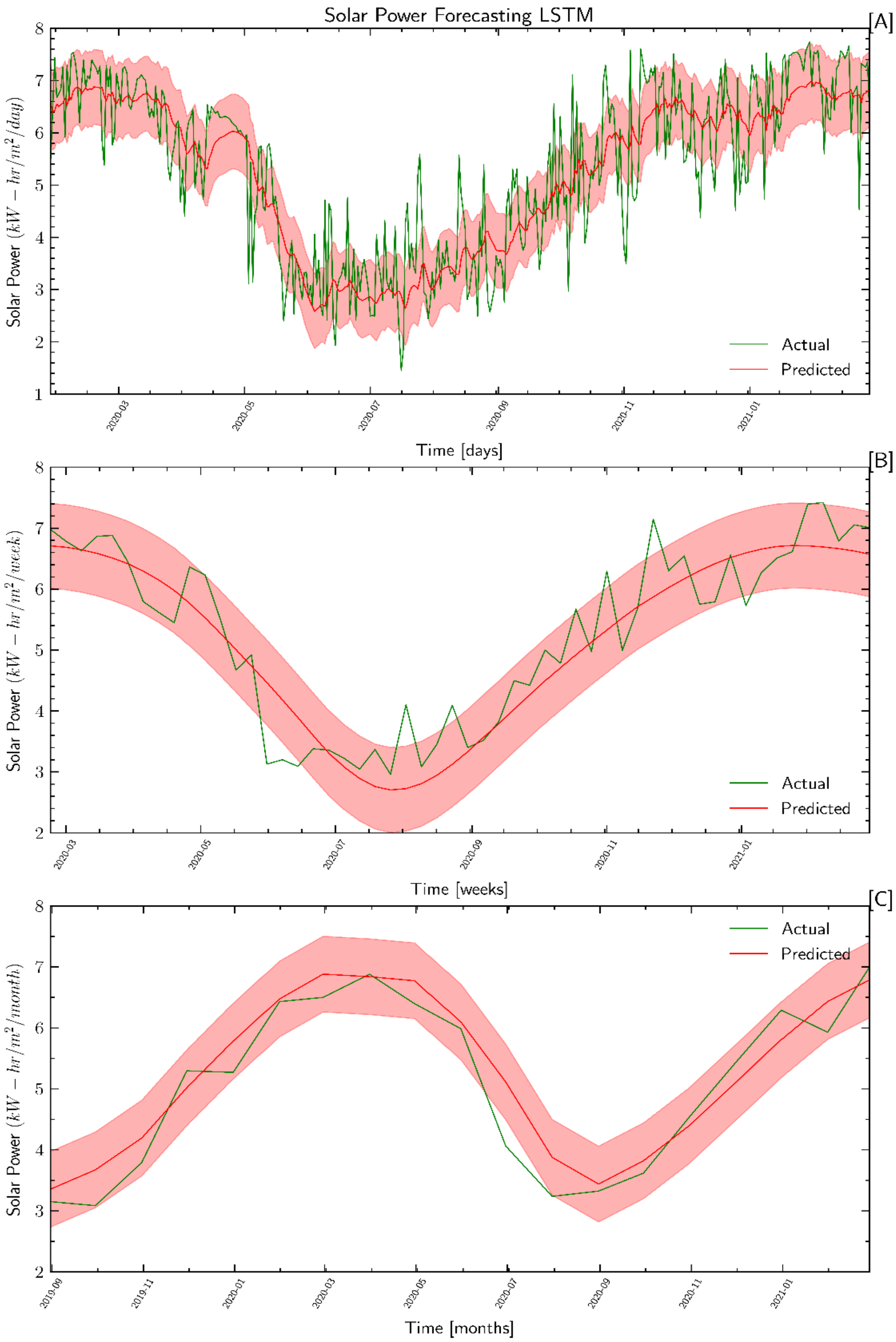

Figure 15. Predictions of solar irradiance with its added standard deviation. A) daily prediction with standard deviation, B) weekly prediction with standard deviation, C) monthly prediction with standard deviation.

4. Conclusions

We have proposed and validated a model for the study of solar radiation over time. The LSTM Recurrent Neural Network has been used, which is an ideal model to study the climate. Despite not having abundant data, it was enough to be able to train and obtain a good result in the regression. The determination coefficient (R^2) being 74% for the daily model, 85.5% for the weekly model and 92% for the month model. Resulting in that the monthly model is the one with the highest efficiency.

To improve this work, it is the possibility of correlating with data obtained by a pyranometer located in Lima. These values obtained after passing through the atmosphere will allow us to obtain a "correction factor" with the predicted values with the Recurrent Neural Network. Which would give us greater reliability to the predicted values and therefore greater precision.

References

[1] Foster, R., Ghassemi, M., and Cota, A. 2009 Solar energy: renewable energy and the environment. *CRC press*.

[2] Lubitz, W. D. 2011 Effect of manual tilt adjustments on incident irradiance on fixed and tracking solar panels. *Applied energy*, **88**(5), 1710-1719.

[3] Su, Y., Chan, L. C., Shu, L., and Tsui, K. L. 2012 Real-time prediction models for output power and efficiency of grid-connected solar photovoltaic systems. *Applied energy*, **93**, 319-326.

[4] NASA 2020 POWER data access viewer.

[5] Olah, C. 2015, Agosto 27 Understanding LSTM Networks. *Colah's blog*. Retrieved 2021, from https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[6] Muñoz, E., and Kikut, A. 1994 El filtro de Hodrick y Prescott: una técnica para la extracción de la tendencia de una serie. *Banco C.*

[7] Zivot, E., and Wang, J. 2007 Modeling financial time series with S-Plus®. *Springer Science & Business Media.* **19**

[8] Hochreiter, S., and Schmidhuber, J. 1997 Long short-term memory. *Neural computation*, **9**(8), 1735-1780.

[9] Chai, T., and Draxler, R. R. 2014 Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development*, **7**(3), 1247-1250.

[10] Kabir, E., Kumar, P., Kumar, S., Adelodun, A. A., and Kim, K. H. 2018 Solar energy: Potential and future prospects. *Renewable and Sustainable Energy Reviews*, **82**, 894-900.

[11] Zoghi, M., Ehsani, A. H., Sadat, M., javad Amiri, M., and Karimi, S. 2017 Optimization solar site selection by fuzzy logic model and weighted linear combination method in arid and semi-arid region: A case study Isfahan-IRAN. *Renewable and Sustainable Energy Reviews*, **68**, 986-996.

[12] Renewables 2021 Global Status Report from https://www.ren21.net/reports/global-status-report/ [Accessed June 2021]

[13] Sukhatme, S. P., and Nayak, J. K. 2017 Solar energy. *McGraw-Hill Education*.

[14] Lorenz, E., 1963a: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20** 130–141

[15] Danforth C. and Lewiston. 2001 Why the Weather is Unpredictable, An Experimental and Theoretical Study of the Lorenz Equations from https://cdanfort.w3.uvm.edu/research/danforth-bates-thesis.pdf [Accessed June 2021]

[16] Barrett, J. P. 1974 The coefficient of determination—some limitations. *The American Statistician*, **28**(1), 19-20.

[17] Ranalder, L., Busch, H., Hansen, T., Brommer, M., Couture, T., Gibb, D., ... and Sverrisson, F. 2021 Renewables in Cities 2021 Global Status Report. *REN21 Secretariat: Paris, France*.

[18] Brahma, B., and Wadhvani, R. 2020. Solar irradiance forecasting based on deep learning methodologies and multi-site data. *Symmetry*, **12**(11), 1830.

[19] Wang, L., Kisi, O., Zounemat-Kermani, M., Salazar, G. A., Zhu, Z., and Gong, W. (2016). Solar radiation prediction using different techniques: model evaluation and comparison. *Renewable and*

Sustainable Energy Reviews, **61**, 384-397.