IRRADIANCE PREDICTION OVER TIME SERIES FOR THE USE OF PHOTOVOLTAIC ENERGY

Mónica Yolanda Moreno Revelo

momorenor@unal.edu.co

Doctorado en ingeniería-linea automática
Universidad Nacional de Colombia

September 19 2024

UNIVERSIDAD NACIONAL DE COLOMBIA

Contents

UNIVERSIDAD NACIONAL DE COLOMBIA

Motivation

- The use of fossil fuel-based energy worldwide is around 80%. [1]
- Approximately 52% of the Colombian territory is not *electrically interconnected*. [2].
- Call 933 Minciencias.
- Increase the accuracy of prediction.

Problem Statement

Given $m \in N$, $I_k \in R^N$ where N is the amount of samples in the time serie, the relationship between I_k and its past values $I_{k-1}, I_{k-2}..., I_{k-m}$ can be described by

$$I_k = \phi_1 f(I_k - 1) + \phi_2 g(I_k - 2) + ... + \phi_m f(I_k - m) \quad (1)$$

There are several ways to calculate the coefficients ϕ and the functions involving irradiance values [3], [4].

Problem statement

Based on the previous description, the following problems can be identified:

1. **Non-stationarity and non-linearity due to climate variability prevents models from capturing the complex relation among the data. Moreover, it difficult to represent complex dependencies over time, including long-term progressions and seasonal patterns** [3], [4], [5].
2. Most methods are focused on specific locations and cannot be generalized to other locations [6].

UNIVERSIDAD NACIONAL DE COLOMBIA

Problem statement

1. Data augmentation due to missing data can lead to problems related to the failure to extract intrinsic properties and task dependence[7],[4].
2. There is uncertainty regarding the selection of input variables for the forecasters, as well as uncertainty about the number of samples chosen for the look-back and look-forward periods [8], [4].
3. Hyper parameters tuning [9].

Classification according to the prediction method

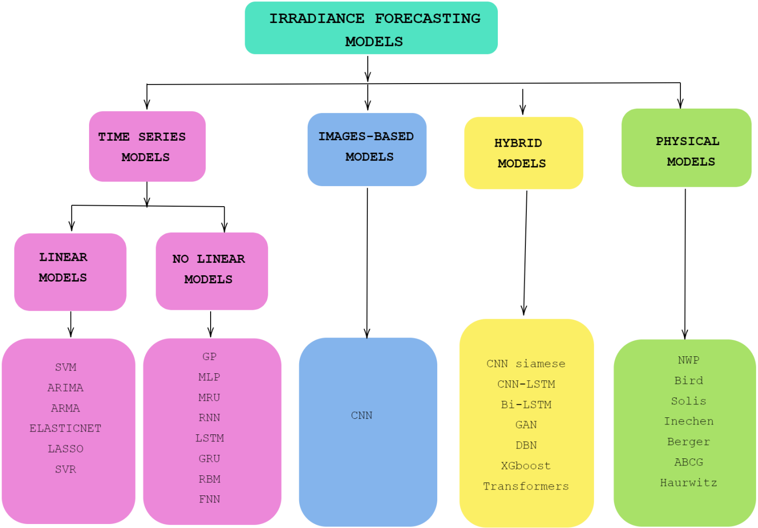

Figure: State of the art in irradiance prediction

DATASETS

The following datasets are used to make experiments

Dataset	Variable	Max Value	Min Value
Solcast	Global Solar irradiance (w/m^2)	1134	0
Ideam	Global solar irradiance (w/m^2)	1103	0
Fronius	Global solar irradiance (w/m^2)	1773.5	0
Jenna	Temperature (℃)	37.3	-22.6

Table: Datasets Information

UNIVERSIDAD NACIONAL DE COLOMBIA

Solcast

Solcast operates a global cloud tracking and prediction system that uses near real-time satellite imagery from 11 weather satellites and weather data from 7 numerical weather prediction (NWP) models [10].

Geographic Coverage	Temporal Coverage
Global (except for the ocean and polar region) Udenar is used in the research	Since January 2007 January 2014-April 2024 is used in the research
Time Recording	**Variables**
5 minutes 60 minutes is used in the research	Global Solar Irradiance (w/m^2) Precipitation (mm/h) Temperature (℃), etc

Table: Solcast Information

Ideam

IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales) plays a crucial role in monitoring and prediction weather patterns in order to manage water resources, and conducting environmental assessments in Colombia [11].

Udenar 52045080		
Temporal Coverage	**Time recording**	**Variables**
Since 01/01/16 until 31/07/21	One hour	Irradiation (whr/m^2)
Since 01/11/16 until 18/04/24		Precipitation (mm) Temperature (℃) Min temperature (℃) Max temperature (℃) Humididty (%)

Table: Ideam Information

Fronius

Fronius is an Austrian company that specializes in solar energy, welding, and battery charging technology. This technology is currently being used at Udenar to measure data on the energy produced in the area [8].

Geographic Coverage	Temporal Coverage
Udenar	January 2023-June 2023
Time Recording	**Variables**
5 minutes 60 minutes is used in the research	Voltage (V) PV production (Kw.hr/day), etc

Table: Fronius Information

UNIVERSIDAD NACIONAL DE COLOMBIA

Jenna

Jena Climate is weather time series dataset recorded at the Weather Station of the Max Planck Institute for Biogeochemistry in Jena, Germany [12].

Geographic Coverage	Temporal Coverage
Jenna, Germany	January 2009-December 2016
Time Recording	**Variables**
10 minutes 60 minutes is used in the research	14 (temperature, *pressure*, humidity, etc)

Table: Jenna Information

PREDICTORS

As a preliminar analysis the following predictors are used:

- Long Short Term Memory (LSTM)
- Gated Recurrent Unit (GRU)
- Extreme Gradient Boosting (XGB)

The previous methods are used in a time-based cross-validation scheme to evaluate them across all samples.

LSTM

LSTM is a improved version of recurrent neural network (RNN) since the network can remember both short term and long term values [9].An LSTM unit is composed of cells, each with an input gate, output gate, and forget gate.

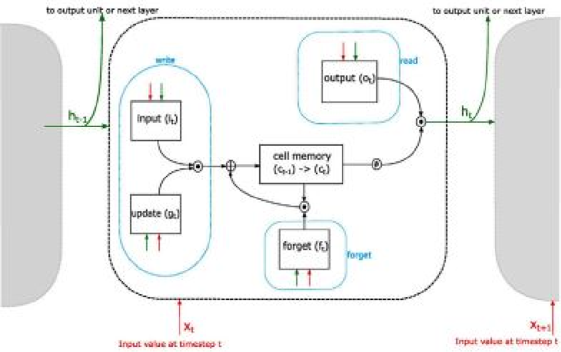

Figure: LSTM General architecture. Taken from [13])

GRU

GRU is another type of RNN designed to address problems related to vanishing and exploding gradients. Unlike LSTM, GRU is simpler to apply and compute [4].

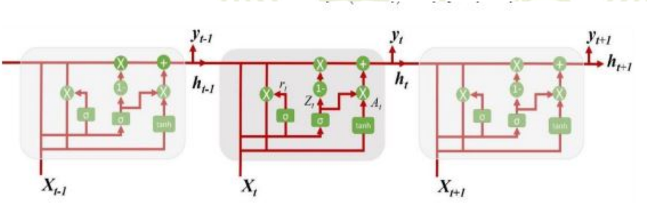

Figure: GRU General architecture. Taken from [4]

XGB

An XGBoost method is an ensemble of several models that work together to fit the residuals of each preceding model. Therefore, XG-Boost can easily learn interactions among features [14].

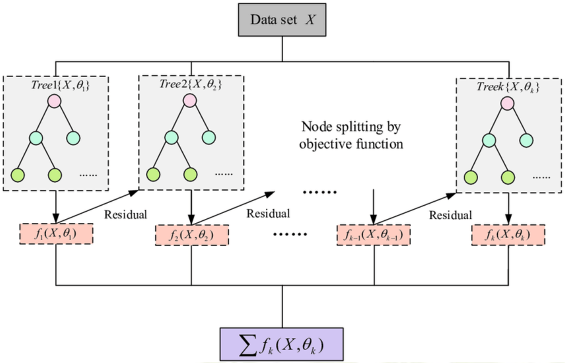

Figure: XGB General architecture. Taken from [15]

Architectures

Based on the research carried out by [16], the next architectures are evaluated over the datasets.

	LSTM and GRU
Inputs	None (ghi or temperature)
Outputs	1(ghi or temperature)
Number of layers	3 (input, hidden with 20 neurons, output)
Epochs	30
Learning algorithm	Adam (learning rate=0.07)
	XGB
Inputs	None (ghi or temperature)
Outputs	1(ghi or temperature)
Booster	gbtree
n estimators	100
max depth	4

Table: Architectures used

METRICS: Auto Correlation

Auto correlation measures the similarity between observations of a time series at different time lags. This metric is used to determine whether the series are independent and to establish the number of samples required for certain tests in the experimental section.

$$k(t, t-k) = \frac{Cov(X_t, X_{t-k})}{\sqrt{Var(X_t)Var(X_{t-k})}} \tag{2}$$

where $Cov(.)$ represents the covariance and $var(.)$ the variance.

UNIVERSIDAD NACIONAL DE COLOMBIA

METRICS: Performance metrics

The next METRICS are used to evaluate the accuracy of the predictors.

metric	Abrev	Equation	Range
Mean Bias Error [6]	MBE	$\frac{1}{N}\sum_{i=1}^{N}(p_m - p_p)$	$[\infty, \infty]$
Root Mean Square Error [17]	RMSE	$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_m - p_p)^2}$	$[0, \infty]$
Determination coefficient [7]	R^2	$1 - \frac{\sum_{i=1}^{N}(p_m - p_p)^2}{\sum_{i=1}^{N}(p_m - \bar{p}_m)^2}$	$[0, 1]$

Table: METRICS used

UNIVERSIDAD NACIONAL DE COLOMBIA

EXPERIMENTS

To apply the previously described prediction methods, first a preprocessing step is carried out, followed by four test set.

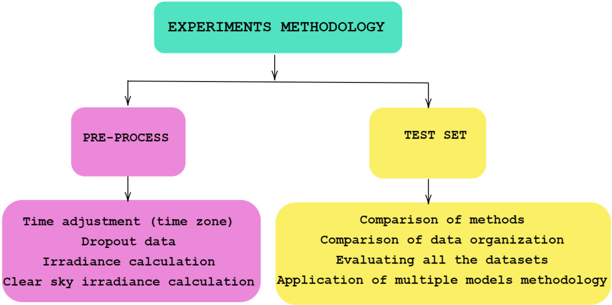

Figure: Methodology of experiments

Pre process

1. **Time adjustment** is applied to Solcast data to align it with the Colombian time zone.
2. **Dropout night data** is done since there is no irradiance at this time.
3. Since Fronius and Ideam record PV production and irradiation respectively, there is a need to **calculate irradiance** first.
4. **Clear sky irradiance** is also calculated to make future experiments.

Pre process

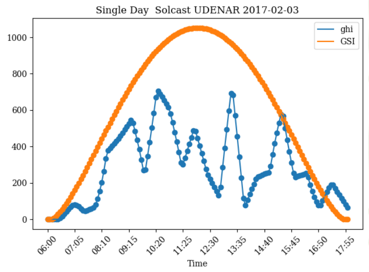

Figure: Measured and theoretical GHI from Solcast

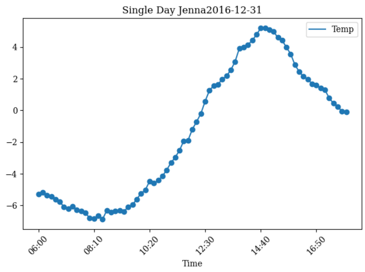

Figure: Measured temperature from Jenna

Figure: Parameters after the pre process

Pre process

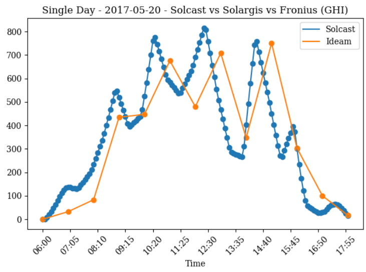

Figure: Measured GHI from Solcast and Ideam

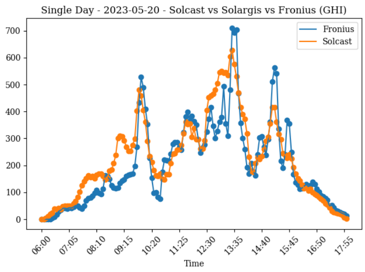

Figure: Measured GHI from Solcast and Fronius

Figure: Parameters after the pre-process

Tests methodology

Firstly, a **Normalization** between 0 and 1 is applied to all datasets and subsequently four test set are carried out.

	Test set 1
Objective	To select the best predictor among those compared
predictor	LSTM, GRU, XGB
Input samples	One
Output samples	One
Dataset	Solcast
Tests done	Three (1A, 1B, 1C)

Table: Test set 1 description

Tests methodology

Since XGB was the best predictor, the next test sets are developed with this one.

	Test set 2
Objective	To select the data organization to feed the predictors
predictor	XGB
Input samples	Thirteen/Seven/One hundred thirty
Output samples	One/One/Thirteen
Dataset	Solcast
Tests done	Three (2A, 2B, 2C)

Table: Test set 2 description

UNIVERSIDAD NACIONAL DE COLOMBIA

Tests methodology

Since using thirteen samples to predict the next one proved to be the most effective technique for feeding the predictors; the subsequent tests are conducted using this methodology.

	Test set 3
Objective	To evaluate the method on different datasets
predictor	XGB
Input samples	Thirteen
Output samples	One
Dataset	Ideam, Fronius, Jenna
Tests done	Three (3A, 3B, 3C)

Table: Test set 3 description

Tests methodology

Finally, hour predictions are made using a different model.

Test set 4	
Objective	To evaluate an hourly model methodology
predictor	XGB
Input Samples	Thirteen
Output Samples	One
Dataset	Solcast, Jenna
Tests done	Two (4A, 4B)

Table: Test set 4 description

UNIVERSIDAD NACIONAL DE COLOMBIA

RESULTS: Test set 1

Single Input Single Output - Solcast dataset						
predictor	Training			Testing		
	MBE (w/m^2)	RMSE (w/m^2)	R2	MBE (w/m^2)	RMSE (w/m^2)	R2
LSTM	21.17	179.62	0.55	23.21	186.53	0.55
GRU	23.08	182.52	0.54	24.75	189.53	0.53
XGB	**0**	**174.52**	**0.57**	**2.51**	**183.11**	**0.56**

Table: Test set 1 results

RESULTS: Test set 1

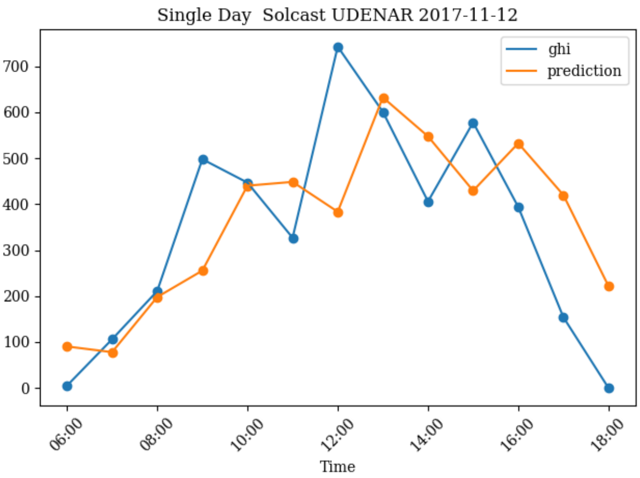

Figure: Measured GHI vs predicted GHI using XGB

UNIVERSIDAD NACIONAL DE COLOMBIA

RESULTS:Test set 2

The number of samples used in the first and third tests is determined based on the recommendation given by [11], while the number of samples in the second test is determined by applying correlation METRICS, obtaining the following results.

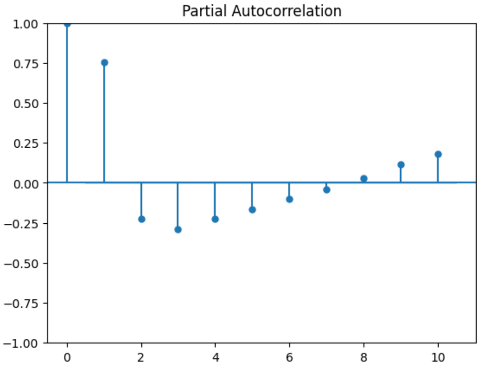

Figure: Solcast autocorrelation

RESULTS: Test set 2

XGB - Solcast dataset						
Input Samples/ Output Samples	Training			Testing		
	MBE (w/m^2)	RMSE (w/m^2)	R2	MBE (w/m^2)	RMSE (w/m^2)	R2
13/1	**0.02**	**130.46**	**0.76**	2.73	**140.38**	**0.74**
7/1	0.04	138.31	0.73	**-1.48**	147.97	0.72
130/13	0	151.31	0.68	3.77	168.64	0.63

Table: Test set 2 results

UNIVERSIDAD NACIONAL DE COLOMBIA

RESULTS: Test set 2

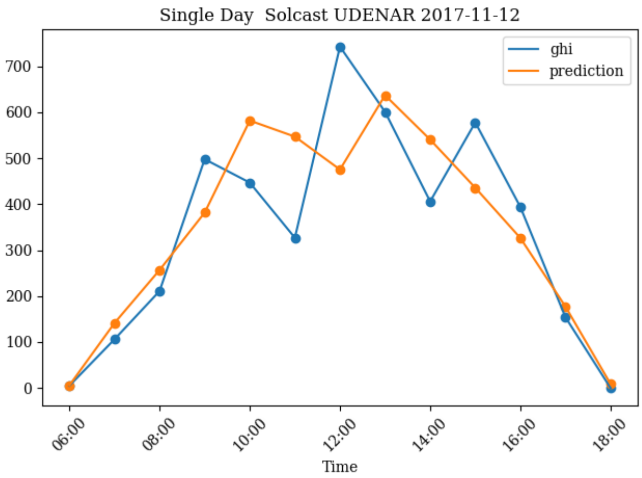

Figure: Measured GHI vs predicted GHI using 13 samples to predict the next one

RESULTS: Test set 3

Multiple Input Single Output-XGB						
Dataset	**Training**			**Testing**		
	MBE (w/m^2)	**RMSE (w/m^2)**	**R2**	**MBE (w/m^2)**	**RMSE (w/m^2)**	**R2**
Ideam	**0**	128.49	0.78	-2.38	**139.66**	**0.72**
Fronius	0	**100.72**	**0.84**	**-2.23**	151.87	0.59
Jenna	0	1.37	0.98	0.031	1.48	0.97

Table: Test set 3 results

RESULTS: Test set 3

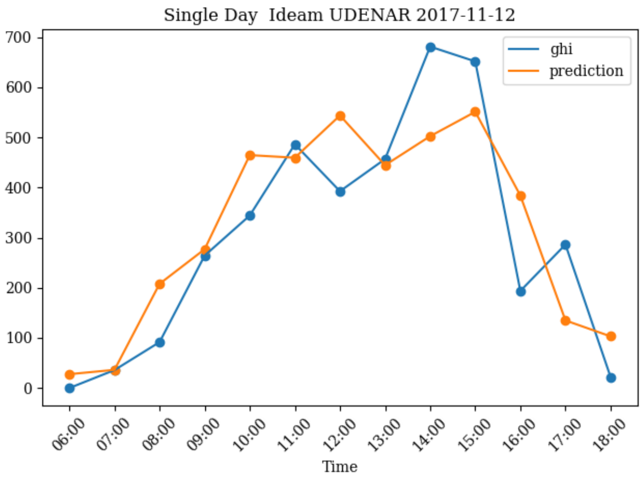

Figure: Measured GHI vs predicted GHI on IDEAM

UNIVERSIDAD NACIONAL DE COLOMBIA

RESULTS: Test set 4

Multiple Input Single Output-XGB-Model per hour						
Dataset	Training			Testing		
	MBE	RMSE	R2	MBE	RMSE	R2
Solcast (w/m^2)	-0.03	78.63	0.73	1.81	121.69	0.4
Jenna (℃)	0	1.73	0.95	0.22	2.56	0.86

Table: Test set 4 results

UNIVERSIDAD NACIONAL DE COLOMBIA

RESULTS: Test set 4

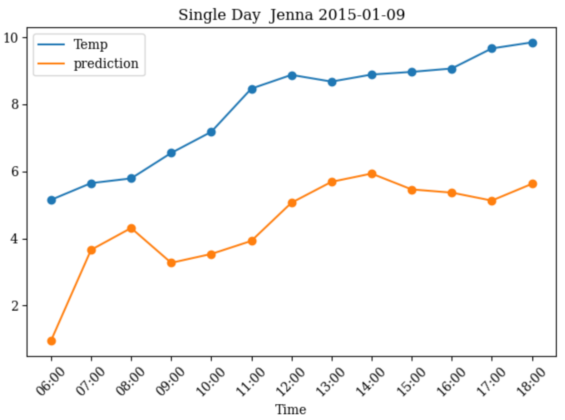

Figure: Measured Temperature vs predicted Temperature on Jenna

Final discussion Solcast

Tests summary - Solcast - XGB						
Test set	Training			Testing		
	MBE (w/m2)	RMSE (w/m2)	R2	MBE (w/m2)	RMSE (w/m2)	R2
1C	**0**	174.52	0.57	2.51	183.11	0.56
2A	0.02	130.46	**0.76**	2.73	140.38	**0.74**
4A	-0.03	**78.63**	0.73	**1.81**	**121.69**	0.4

Table: Solcast Results Summary

Final discussion Solcast

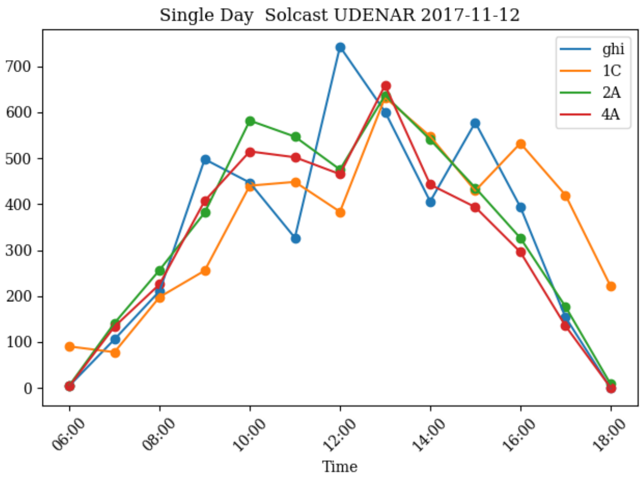

Figure: Measured GHI vrs predicted GHI with different methods

Final discussion Jenna

Tests summary - Jenna - XGB						
Test set	Training			Testing		
	MBE (w/m2)	RMSE (w/m2)	R2	MBE (w/m2)	RMSE (w/m2)	R2
3C	**0**	1.37	0.98	0.03	1.48	0.97
4B	0	1.73	**0.955**	0.22	2.56	**0.86**
[18]	-	**0.80**	-	**-**	**0.87**	-

Table: Jenna Results Summary

UNIVERSIDAD NACIONAL DE COLOMBIA

Final discussion Jenna

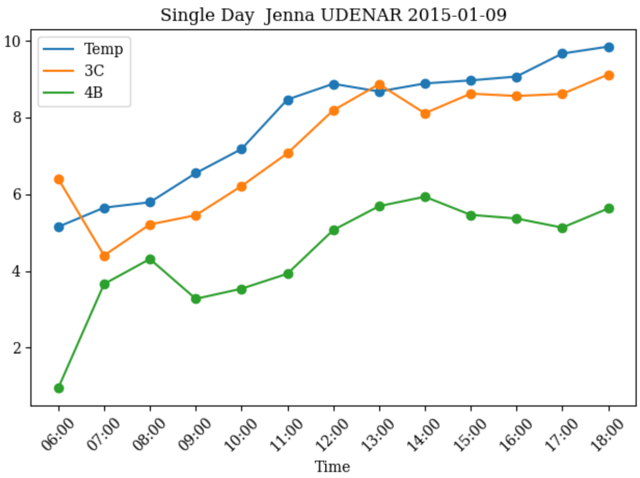

Figure: Measured Temperature vrs predicted Temperature with different methods

Final discussion

The most important results are the following:

- As shown in test set 2, using several samples to predict the next one significantly improves the results if compared to using only one previous sample.
- According to test set 4 it is concluded that it is better to use a different model for each hour instead of using only one model to predict the GHI the next thirteen hours. Moreover, predicting all the values (multiple output) in a day decreases the accuracy of the prediction as seen in test 2C.
- Comparing Test 2A and Test 2B, it is concluded that if a window approach with a different analysis is carried out, it improves the accuracy.

Future work

As further research the next tasks are proposed:

- To evaluate other methods such as: Gaussian process, hybrid methods, transformers.
- To find a relationship between Solcast and Ideam Data to take advantage of the two kind of data.

References I

[1] Noor Bariah Mohamad, An-Chow Lai, and Boon-Han Lim. “A case study in the tropical region to evaluate univariate imputation methods for solar irradiance data with different weather types”. In: **Sustainable Energy Technologies and Assessments** 50 (2022), p. 101764.

[2] Gabriel Narvaez et al. “The impact of climate change on photovoltaic power potential in Southwestern Colombia”. In: **Heliyon** 8.10 (2022).

[3] Jwaone Gaboitaolelwe et al. “Machine learning based solar photovoltaic power forecasting: a review and comparison”. In: **IEEE Access** 11 (2023), pp. 40820–40845.

[4] Abbas Mohammed Assaf et al. “A review on neural network based models for short term solar irradiance forecasting”. In: **Applied Sciences** 13.14 (2023), p. 8332.

UNIVERSIDAD NACIONAL DE COLOMBIA

References II

[5] Meenu Ajith and Manel Martínez-Ramón. “Deep learning algorithms for very short term solar irradiance forecasting: A survey”. In: **Renewable and Sustainable Energy Reviews** 182 (2023), p. 113362.

[6] Laith Abualigah et al. “Wind, solar, and photovoltaic renewable energy systems with and without energy storage optimization: A survey of advanced machine learning and deep learning techniques”. In: **Energies** 15.2 (2022), p. 578.

[7] Qingsong Wen et al. “Time series data augmentation for deep learning: A survey”. In: **arXiv preprint arXiv:2002.12478** (2020).

[8] Gentiana Alija, Armend Ymeri, and Nexhmi Krasniqi. “Study aspects for the solar-photovoltaic system with installation Capacity of 60.3 kwp”. In: (2021).

UNIVERSIDAD NACIONAL DE COLOMBIA

References III

[9] Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019". In: **Applied soft computing** 90 (2020), p. 106181.

[10] Abdulrahman Almarshoud et al. "Validation of satellite-derived solar irradiance datasets: a case study in Saudi Arabia". In: **Future Sustainability** 2.2 (2024), pp. 1–7.

[11] Laura S Hoyos-Gómez, Jose F Ruiz-Muñoz, and Belizza J Ruiz-Mendoza. "Short-term forecasting of global solar irradiance in tropical environments with incomplete data". In: **Applied Energy** 307 (2022), p. 118192.

[12] K Venkatachalam et al. "DWFH: An improved data-driven deep weather forecasting hybrid model using Transductive Long Short Term Memory (T-LSTM)". In: **Expert systems with applications** 213 (2023), p. 119270.

UNIVERSIDAD NACIONAL DE COLOMBIA

References IV

[13] Pratima Kumari and Durga Toshniwal. “Deep learning models for solar irradiance forecasting: A comprehensive review”. In: **Journal of Cleaner Production** 318 (2021), p. 128566.

[14] Hanjin Zhang et al. “A novel hybrid transformer-based framework for solar irradiance forecasting under incomplete data scenarios”. In: **IEEE Transactions on Industrial Informatics** (2024).

[15] Rui Guo et al. “Degradation state recognition of piston pump based on ICEEMDAN and XGBoost”. In: **Applied Sciences** 10 (Sept. 2020), p. 6593. DOI: 10.3390/app10186593.

[16] Garazi Etxegarai et al. “An analysis of different deep learning neural networks for intra-hour solar irradiation forecasting to compute solar photovoltaic generators’ energy production”. In: **Energy for Sustainable Development** 68 (2022), pp. 1–17.

UNIVERSIDAD NACIONAL DE COLOMBIA

References V

[17] Bryan Lim and Stefan Zohren. "Time-series forecasting with deep learning: a survey". In: **Philosophical Transactions of the Royal Society A** 379.2194 (2021), p. 20200209.

[18] Hemant Yadav and Amit Thakkar. "NOA-LSTM: An efficient LSTM cell architecture for time series forecasting". In: **Expert Systems with Applications** 238 (2024), p. 122333.

Thanks

UNIVERSIDAD NACIONAL DE COLOMBIA