2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10020615

Solar Irradiance Prediction Using Transformer-based Machine Learning Models

Ayda Demir[1], Luis Felipe Gutiérrez[2], Akbar Siami Namin[2], Stephen Bayne[1]
[1]Department of Electrical and Computer Engineering, [2]Department of Computer Science
[1,2]Texas Tech University
{*ayddemir, luis.gutierrez-espinoza, akbar.namin, stephen.bayne*}*@ttu.edu*

***Abstract*—This paper presents a study of irradiance prediction using a transformer-based machine learning model for the photovoltaic (PV) renewable energy system. We explore the forecast of irradiance using ten years of data from the Texas Mesonet Data Archive at the Reese Center in Lubbock, Texas. After training our transformer model with 90% of the data, we show that it can fit the irradiance trend successfully, while the testing phase with the remaining 10% of the dataset indicates that the transformer can predict irradiance trends that align with the observed values. Additionally, we borrow the concept of rolling LSTMs to generate a rolling transformer in order to extrapolate values of irradiance even when the observed values are not available. Our extrapolation results show that the transformer can extrapolate irradiance values accurately in the short-term, but it is less precise in the long-term. To remedy this, we aim to explore more thoroughly the hyperparameter configuration of our model in order to move towards our goal of including machine learning methods in the control of PV systems.**

***Index Terms*—Forecasting, solar panels, renewable energy, deep learning, machine learning, transformer, time series.**

I. Introduction

The adoption of renewable energy sources stands as one of the main measures to reduce the carbon print caused by fossil fuels. Moreover, renewable energy sources are also leveraged around the world to complement an energy grid that is already powered by heavily pollutant sources; this is critical when the energy requirements in a given area rise or the access to traditional energy sources gets restricted. For instance, there is a growing concern in the European Union because of the 2021/2022 gas crisis and the threat to its decarbonization goals set to be achieved by 2050 [1]. Therefore, government authorities and private organizations consider energy from renewable sources as an important step to remedy this matter. This study relates to one particular renewable source: solar energy.

Solar cells are electronic devices that use sunlight energy, or *irradiance*, to convert it into electrical energy by the photovoltaic effect, which is a phenomenon that allows the generation of voltage and current exposing a material to sunlight. Solar cells are the electrical building blocks of traditionally used solar panels. These panels come together to create a solar string. As seen in the Figure 1, many solar strings are brought together in highly-productive energy environments to form a solar array [2].

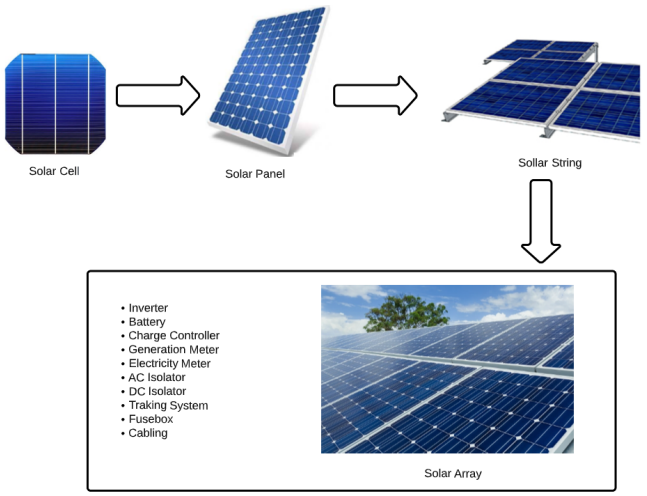


Fig. 1: Diagram from solar cell to solar array.

The most critical parameters to generate energy from solar panels are temperature and irradiance. Equation 1 presents how the solar power correlates with temperature and irradiance, where P_{PV} is the solar panel power, $P_{PV_{STC}}$ is the solar panel power in standard test conditions, G_T is the irradiance, and T_j is the temperature [3]. Figure 2 shows and end-to-end diagram of the energy flow from generation to user delivery. The process starts with the solar panel generating energy from irradiance. Next, after the energy generated from the solar panel is transmitted to the DC-DC converter, it is sent to the inverter before it reaches the final user or is dispatched to the main grid. Note that one evident drawback of solar energy is that it cannot be generated outside daylight hours. To solve this issue, solar energy can be dispatched to a battery system after the inverter step with the purpose of storing it for future use.

$$P_{PV} = P_{PV_{STC}} * \frac{G_T}{1000} * [(1 - \mu) * (T_j - 25)] \quad (1)$$

As mentioned previously, irradiance is a critical factor in the performance of solar panels. Figure 3a shows an I-V curve, i.e., values of the output of voltage (x-axis) and short circuit current (y-axis) in the solar panel for different values

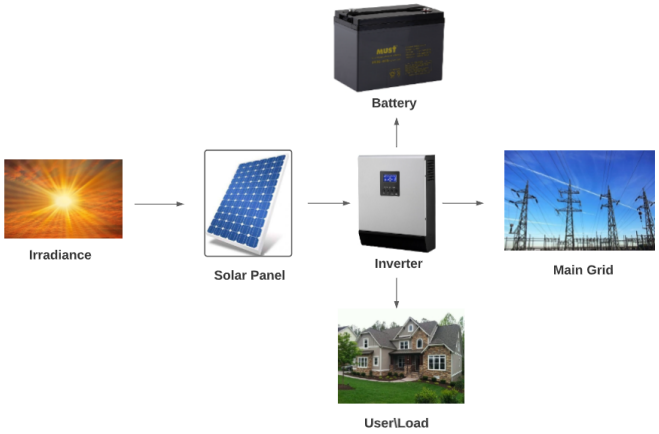
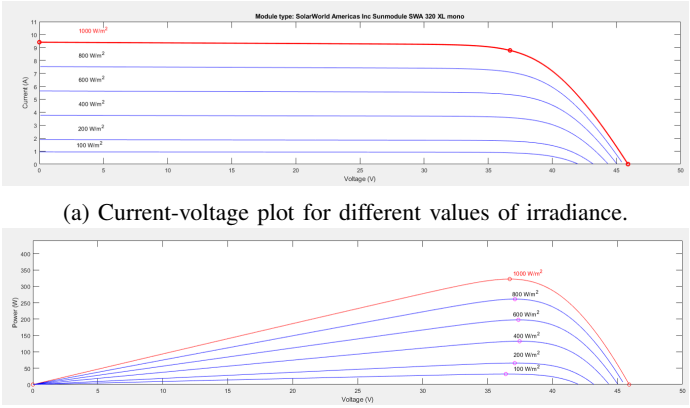


Fig. 2: Energy flow from generation to delivery.



(a) Current-voltage plot for different values of irradiance.

(b) Power-voltage plot for different values of irradiance.

Fig. 3: PV module power-voltage, current-voltage at different irradiance levels.

of irradiance. On the other hand, Figure 3b depicts the values of voltage (x-axis) and power (y-axis) vary according values of irradiance. One commonality between Figures 3a and 3b is that both plots place the highest output of current and power when irradiance is at peak levels and, thus, the solar panel is working at maximum capacity.

Due to the essential nature of irradiance in the energy generation of solar panels, several researchers have put their efforts in the prediction of irradiance for different purposes related to renewable energy generation. For instance, Yu et al. [4] employed the Long Short Term Memory (LSTM) cell, an ubiquitous machine learning model in time series analysis, to perform irradiance forecasting during complicated weather conditions. In addition to LSTMs, Jalali et al. [5] incorporated convolutional neural networks (CNNs) to create a convolutional LSTM model in order to predict future values of irradiance.

Even though LSTMs and CNNs have demonstrated to achieve good performance in tasks related to time series analysis and predictions, a novel machine learning method, the *Transformer*, has gained momentum in time series forecasting after positioning as a state-of-the-art model in other research fields such as natural language processing and computer vision; related to our field, transformers have also been used for energy load forecasting [6], showing that the method is promising in the renewable energy sector.

In this study, we leverage the transformer machine learning model for irradiance forecasting. We use ten years of irradiance data retrieved from the West Texas Mesonet Data Archive at the Reese Center in Lubbock, Texas. As other researchers in this field, we aim to take advantage of the predicted irradiance values to anticipate the power generated since, as shown previously in Equation 1, irradiance is a relevant term for power generation. Most importantly, knowing beforehand irradiance values will allow us to anticipate control decisions so the solar system can be prepared for the forecast conditions. We expect this to be a step forward towards the inclusion of state-of-the-art machine learning methods in the control of solar systems and microgrids.

It is worth noting that this research study is contextualized in a larger and more comprehensive project that focuses on incorporating machine learning approaches into the control of microgrids. We have already explored weather-related forecasting using LSTMs for temperature in this direction [7].

The key contributions of this paper are:

1) We implement and train a transformer model specifically designed to forecast irradiance values using historical irradiance data.
2) We explore how a rolling transformer can be used to perform an unsupervised prediction in a short (i.e., 48 hours or 2 days) and long-term (i.e., 384 hours or 16 days) period.
3) We show that transformers can be suitable for taking the first step towards the prediction of produced energy in solar panels through the prediction of irradiance. Hence, transformers are promising for automating the control of solar panel system.

The rest of this paper is organized as follows: Section II summarizes a few studies relevant to our research; Section III provides a description of the transformer model used in this paper; Section IV presents the dataset that we utilized in the experiments; Section V shows the results that we obtained after carrying out our experiments; finally, Section VI concludes our study and sheds light on future research directions.

II. Related Work

Forecasting of weather-related data is a common topic in machine learning literature: some of the weather features predicted are temperature [7], wind speed [8], humidity [9], and irradiance [10], which is the one relevant to our study. Irradiance estimations are made using different methods. For instance, Qing & Niu [10] investigated the forecasting of irradiance using the LSTM machine learning model for a specific area, which is the island of Santiago, Cape Verde. In the aforementioned study, LSTM results were compared

with other methods, such as linear least square regression and multilayered feedforward neural networks using backpropagation algorithm (BPNN), revealing that better results are obtained using LSTMs. Deep learning-based approaches to time series analysis such as LSTM [11] and Temporal Convolutional Networks (TCN) [12] have already demonstrated their effectiveness in predicting time series data.

In the same directions, researchers used ensemble models, such as boosted trees, bagged trees, random forest, and generalized random forest, to predict solar irradiance [13]. In [13], Villegas-Mier et al. utilize an optimized random forest and adaboost models to predict irradiance from different locations in the city of Queretaro, Mexico, reaching a 95% when comparing to the baseline models.

In this study, we use a state-of-the-art model from natural language processing and computer vision: the transformer. Transformers have a more complex structure when compared to more traditional deep learning models such as CNNs and LSTMs, and they achieve better results in tasks where convolutional and recurrent models were top-performers [14]. Additionally, the transformer model allows for a much faster training stage since it can be parallellized without breaking the sequence order, a feature that is impossible in sequential models such as the LSTM. Our work relies heavily on the transformer configuration presented by Wu et al. [15], whose purpose is to predict the occurrence of influenza-like illness. More details about transformers are presented in Section III.

III. The Transformer-based Model

This section briefly presents the technical background and architecture of the transformer model that we used in our study.

A. Transformers for Time Series Forecasting

Transformers are state-of-the-art models in fields such as Natural Language Processing (NLP) and Computer Vision (CV), consistently outperforming traditional deep learning architectures like LSTMs and CNNs [14]. The breakthrough and capabilities of transformers are being expanded to the domain of time series analysis and forecasting. A notable research work is presented by Wu et al. [15], where the authors develop a transformer-based framework to forecast the amount of influenza-like illness. In this study, we adapt the introduced framework from [15] to carry out irradiance forecasting. Next, we briefly present the model utilized in our work.

1) Transformer Architecture: Figure 4 shows the architecture of the transformer model used in this study. As in [15], we construct the model using the encoder and decoder components presented in the seminal work on transformers by Vaswani et al. [16], where the initial architectures of the encoder, decoder, and the attention mechanism were presented. A similar encoder-decoder architecture is also adapted for clustering time series data [17].

a) Positional encoding layers: One crucial advantage of transformers over traditional recurrent models is that they can undergo parallelized training. This contrasts with recurrent models, which rely on sequential training procedures to learn the underlying pattern in the data. Nonetheless, despite the parallelization capabilities of their training process, transformers still are provided with information regarding the order of the sequences; this is done using "*positional encoding layers*". Positional encoding layers inject information about the local or global position of the samples in a sequence. This allows the model to preserve the order of the sequence while training non-sequentially. Positional encoding layers leverage trigonometric functions and calculate each entry of the positional encoding matrix, P, as follows [16]:

$$P(k, 2i) = \sin \frac{k}{n^{2i/d}}$$
$$P(k, 2i+1) = \cos \frac{k}{n^{2i/d}}$$

where k is the index of the sample in the sequence, d is the dimension of the embedding output, i is the column index, and n is a scalar set to $10,000$ by the authors in [16].

b) Input layers: Both the encoder and decoder components get their input using an input layer. Each input layer is a simple linear feed-forward component.

c) Encoder component: The encoder component consists of a fixed number of sub-layers called *encoder layers*. Figure 5a depicts the internal structure of the encoder layer. The first operation in the encoder layer is *self-attention*, a vital mechanism in the transformer model. For the sake of space, we omit the definition of self-attention, this can be consulted in [16]. Next, an aggregation and normalization layer is applied, alongside a feed-forward linear layer. Finally, another aggregation and normalization is applied. The output of the encoder layers is then fed to the decoder layers, as seen in Figure 4.

d) Decoder component: The decoder component has an internal structure similar to that of the encoder component. In addition to the self-attention and aggregation/normalization layers, the decoder contains an *encoder-decoder attention* mechanism.

2) Hyperparameters selection: As in the case of previous state-of-the-art models, such as LSTMs and CNNs, hyperparameters selection is of uttermost importance for transformer architectures. Despite the fast training process of transformers, this can be a very resource-consuming step. Because of this, we borrow part of the selection of hyperparameters from [15], which already demonstrated to be suitable in the task of time series forecasting. The hyperparameters configuration of our model is presented in Table III.

3) Construction of the Training dataset: In addition to the transformer architecture, Figure 4 shows an example of a single training data point used in our experiment. Each sample in our training set consists of a sequence with a window size of 48 consecutive observations (i.e., $(x_1, x_2, ..., x_{47}, x_{48})$) that is fed to the encoder component, and a sequence of 48 observations (i.e., $(x_{48}, x_{49}, ..., x_{94}, x_{95})$) that is fed into the decoder. The encoder's and decoder's sequences are used

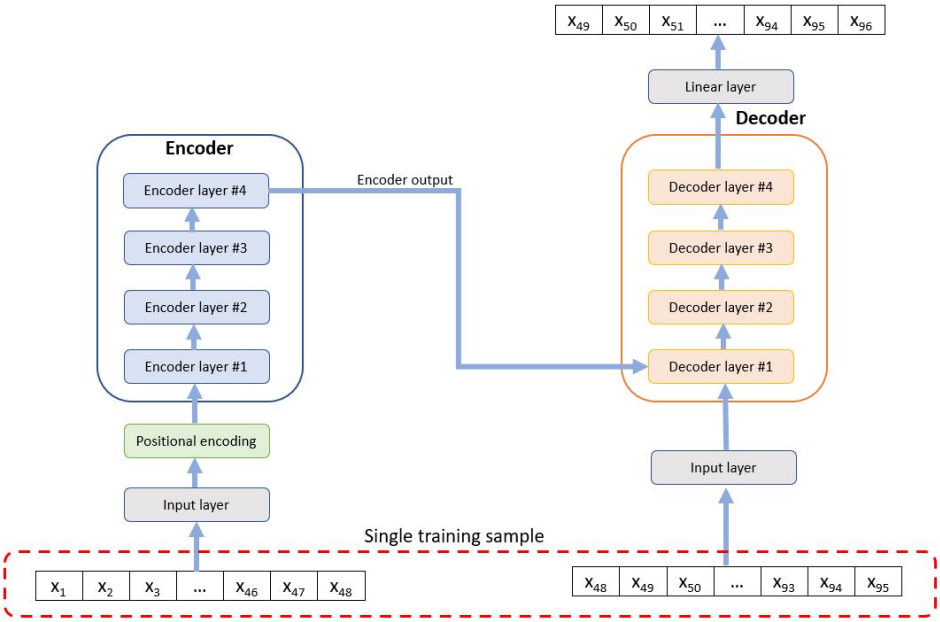
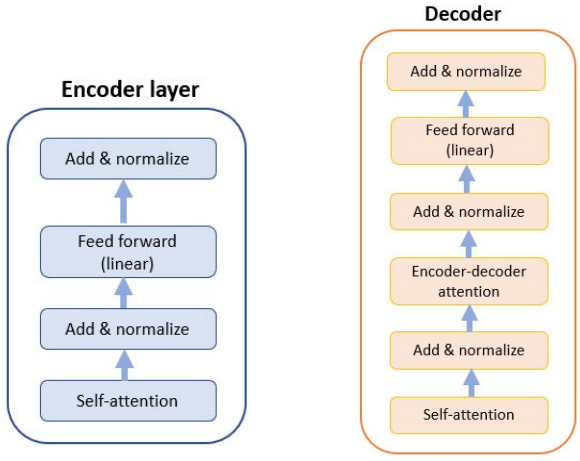


Fig. 4: Transformer architecture.



(a) Internal architecture of an encoder layer.

(b) Internal architecture of a decoder layer.

Fig. 5: Internal architecture of encoder and decoder layers.

to predict the decoder's output (i.e., $(x_{49}, x_{50}, ..., x_{95}, x_{96})$), which is the final prediction. Following the work in [16], we mask the decoder's sequence in order to prevent the leakage of future information into the current target prediction. We detail our dataset in Section IV.

4) Optimizer and loss function: We trained our transformer model using the Adam optimizer [18] using a fixed learning rate of $2 \cdot 10^{-4}$ for 10 epochs, alongside the Mean Squared Error loss function, defined as

TABLE I: Selection of hyperparameters for our transformer model.

Hyperparameter	Value
# encoder layers	4
# decoder layers	4
Embedding dimension	512
# hidden units (encoder input layer)	2,048
# hidden units (decoder input layer)	2,048
# attention heads (encoder)	8
# attention heads (decoder)	8
Dropout prob. (regularization)	0.2

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(observed(i) - predicted(i))^2,$$

where n is the total number of samples.

IV. Dataset

The dataset used in our study was obtained from the West Texas Mesonet Data Archive at the Reese Center in Lubbock, Texas. Specifically, we utilized values of irradiance taken from the period between January 1st 2012 and March 11th 2022. These values were obtained using sensors that captured irradiance values in W/m^2 at different intervals, ranging from every three to fifteen minutes. Originally, the dataset consists of over 1.9 million observations taken in roughly 10 years. Since the samples are taken at different time intervals, we aggregate and average the data hourly; this reduces the dataset to 89,279 samples over the same time span. Table II shows the main descriptive statistics of the dataset.

TABLE II: Descriptive statistics of the dataset.

Statistic	Value
Count	89,279
Mean	226.09
Std	311.66
Min.	0.00
Max.	1,124.90

TABLE III: Summary of sizes of the data splits.

Split	# samples
Training set	80,784
Testing set	8,495

A. Training/Testing Split

We applied a hold-out split to validate our model with 90% of consecutive data reserved for the training set, and the remaining 10% was used to construct the testing set. In addition, we transformed the data using standardization in order to avoid negative effects caused by the scale of the data.

V. Results

We present our results in a 3-fold manner:

1) We show how the model succeeds in fitting the irradiance data.
2) We use our held-out testing set to test our model on unseen data; we show that the predicted trends agree with the observed trends.
3) We apply the concept of *rolling* LSTMs to the transformer model to generate predictions in an unsupervised fashion, i.e., without having the observed data for irradiance.

A. Model Fitting

We trained the model presented in Section III using the hyperparameter configuration shown in Table III for 10 epochs. Figure 6 shows an elbow graph of training loss values (MSE) per number of epochs using the standardized values of irradiance. The elbow graph shows that the training process is stable, as there are no significant changes in the loss values while increasing the number of epochs. We contrast this finding with an elbow graph generated after training the model using unnormalized data in Figure 7, which shows that training the transformer model with unnormalized data can cause the training to become unstable, as can be in the loss raise at epoch 6 in Figure 7. Furthermore, we observe that the loss values drop rapidly after only one epoch in the case of normalized data, and it plateaus starting at two epochs onwards. This suggests that the transformer model is capable of learning the underlying pattern in the signal using a relatively low number of epochs, which is also the case for unnormalized data in Figure 7, although the same result is achieved at a higher number of epochs.

Figure 8a shows a single sample from our training set (last sample) alongside the prediction of our model done in training time where the x-axis is the time; whereas, the y-axis is the

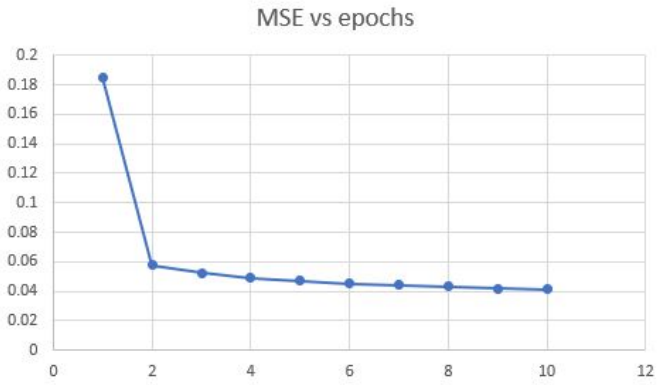


Fig. 6: Training loss values (MSE) per number of epochs (normalized data).

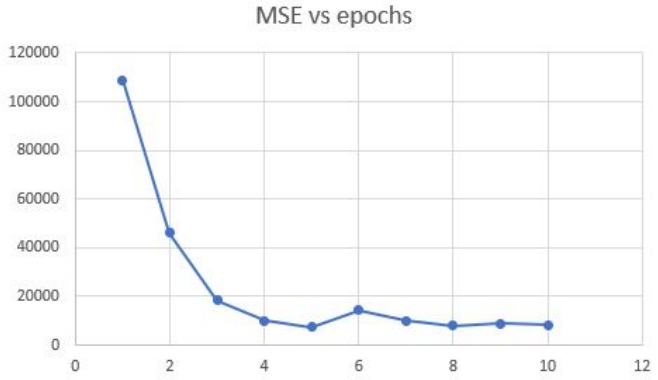


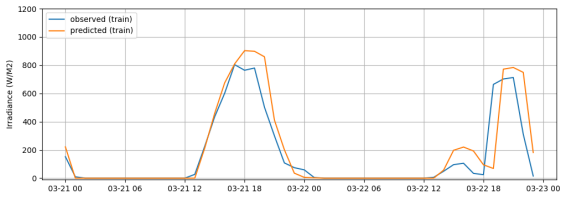
Fig. 7: Training loss values (MSE) per number of epochs (unnormalized data).

amount of irradiance. Recall a single sample consists of 48 consecutive observations of irradiance (see Figure 4). Note that the model is able to learn the expected periodic peaks of irradiance throughout 48 hours, including the abrupt decrease during March 22nd 2021, probably due to a temporary overcast.

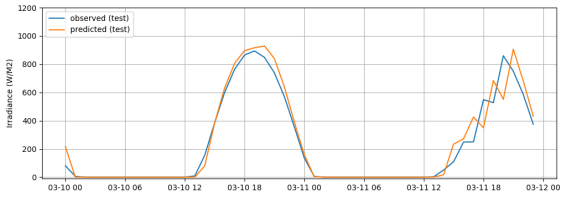
B. Model Evaluation at Testing Stage

Next, we evaluated our transformer model on the testing dataset. Recall we reserved the last 10% of our dataset for testing purposes. As we did with a single training sample, Figure 8b shows an example of a data point from the testing set alongside the prediction from our model. One notable finding from Figure 8b is that the model is able to predict the ascending irradiance wavelets present during the daylight hours of March 11th 2022.

As for the complete testing set, displaying it entirely is not feasible since it is approximately one year of data, and the fluctuations of irradiance would not be visible in a resolution of one day. To remedy this, we chose to display 16 days (384 hours/steps) worth of irradiance data; these observations correspond to the period between February 22nd 2022 01:00 and March 10th 2022 00:00. Figure 9a shows the observed values of irradiance and the predictions of our model for the



(a) One sample from the training set, alongside our transformer model's prediction for irradiance (March 21st 2021 00:00 - March 22nd 2021 23:00, errors are scaled back to the original magnitudes).



(b) One sample from the testing set, alongside our transformer model's prediction for irradiance (March 10th 2022 00:00 - March 11th 2022 23:00, errors are scaled back to the original magnitudes).

Fig. 8: Examples of single observations from our training and testing sets.

TABLE IV: Descriptive statistics of errors at testing and extrapolation time.

	Original scale				Normalized			
	Testing		Extrapolation		Testing		Extrapolation	
	MSE	RMSE	MSE	RMSE	MSE	RMSE	MSE	RMSE
Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Max	189,830	435	824,285	907	608	24.6	2,646	51.4
Avg.	4,035	63	43,303	208	12.2	3.4	138	11.7
Std.	6,524	80	77,431	278	20.2	4.4	247	15.7
Avg. (only daylight)	7,107	84	75,242	274	22	4.6	240	15.4
Std. (only daylight)	9,149	95	62,872	250	28	5.2	201	14.1

aforementioned 16 days. After the inspection of the observed and predicted trends, we noticed that the model could also predict the correct patterns of irradiance throughout the days, including some of their irregularities such as the ones seen during March 2nd 2022 or March 6th 2022.

Nevertheless, we also identify a drawback that also arises in Figure 8a and Figure 8b: even though there exists a pattern in the predictions that resembles the original pattern in the training data (i.e., daylight and night hours), the model almost consistently overestimates the values of irradiance. This could represent a critical problem when integrating the model to a real-life scenario of control. This is supported by Table IV, which shows some descriptive statistics of errors during testing stage for both normalized and in the original scale.

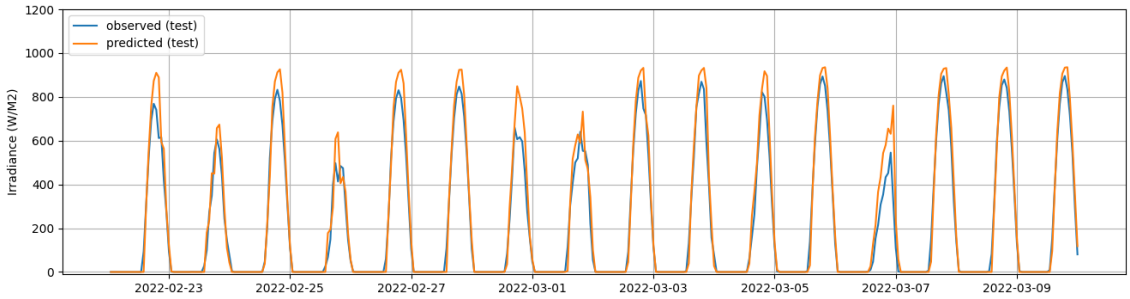
In order to avoid a misleading error reading, we calculate the average error with and without considering non-daylight hours, this prevents artificially low average errors as the model easily predicts non-daylight hours. Note that the average testing error during daylight is 84 W/m^2 (RMSE), whereas the maximum error found in the predictions during the testing stage is 435 W/m^2.

C. Exploring Irradiance Extrapolation with a Rolling Transformer

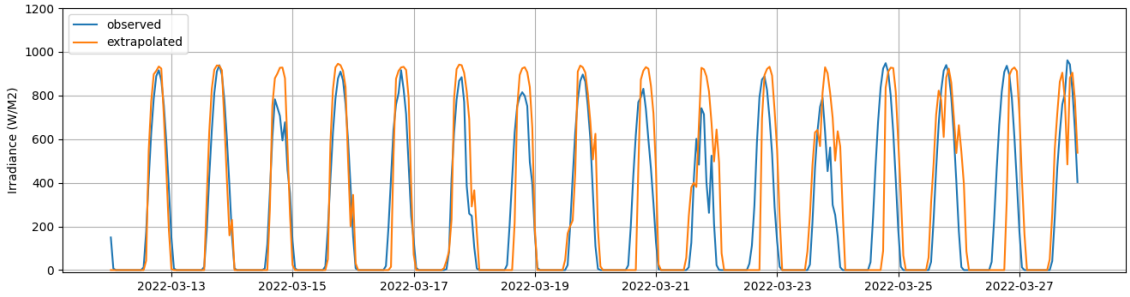
Rolling LSTMs [11] rely on predicted outputs to perform prediction of future values. We leverage the concept of rolling to perform extrapolation (i.e., prediction of future values of irradiance without having the actual observed values) of irradiance values, but using our transformer model instead of an LSTM. The core idea remains unchanged: we feed each predicted sequence of 48 observations at step t as a query sequence to the transformer to predict the sequence at step $t + 1$. The purpose of this is to perform prediction of future irradiance values even when the actual observed data is not available. In addition, this offers insights into how the model would behave if it predicts in a completely unsupervised manner. We withheld 16 days worth of data from the testing set to conduct our analysis regarding extrapolation.

Figure 9b shows the extrapolated values of irradiance in this part of the experiment. Even though the rolling transformer does not need actual observed values to make predictions in an ongoing basis, we plot the observed and predicted irradiance values in order to assess the model performance. According to Figure 9b, the model performs well during the first 48 hours, it succeeds in predicting irregular irradiance peak patterns such as the one during March 22nd 2022 or March 24th 2022. Nevertheless, it is also noticeable that the model overestimates values of irradiance, an effect we also found during the training and testing phases; this can be confirmed quantitatively in Table IV, where the average error is 244 W/m^2 in the original scale.

Note that the maximum error in the extrapolation step is 907 W/m^2, which is a particularly large error. Moreover, Figure 9b depicts another interesting phenomenon during the extrapolation phase: the frequency of the extrapolated signal seems to be increasing the longer the model is used to predict irradiance values in an unsupervised manner. This causes the extrapolated values to start offsetting with respect to the actual observed values, i.e., the two signals start reducing their overlap, which at the same time causes the massive errors like the 907 W/m^2 (original scale) seen in Table IV. We confirm this finding after plotting the trend of a moving average (window size = 24 hours) of MSE errors for the extrapolated irradiance values, this plot is shown in Figure 10. Figure 10 shows that the average daily MSE error of the extrapolated data has an oscillating and generally ascending pattern that gets steeper the more extended the extrapolated period is. This has direct implications in the usability of our current transformer model in a real-life scenario for long-term irradiance prediction, which would be less favorable than short-term predictions, as seen in Figure 10. However, this does not mean that the transformer models are not suitable for long-term irradiance prediction, but that this specific architecture and hyperparameters configuration needs to be improved to achieve better results. This is one of our main research directions to improve the results presented in this study.



(a) Last 16 days of testing data, alongside our model's predictions. The values are observations and predictions between February 22nd 2022 01:00 and March 10th 2022 00:00.



(b) 16 days of extrapolated data between March 12th 2022 and March 27th 2022, alongside the observed values of irradiance in that period.

Fig. 9: Testing stage and extrapolation of irradiance values.

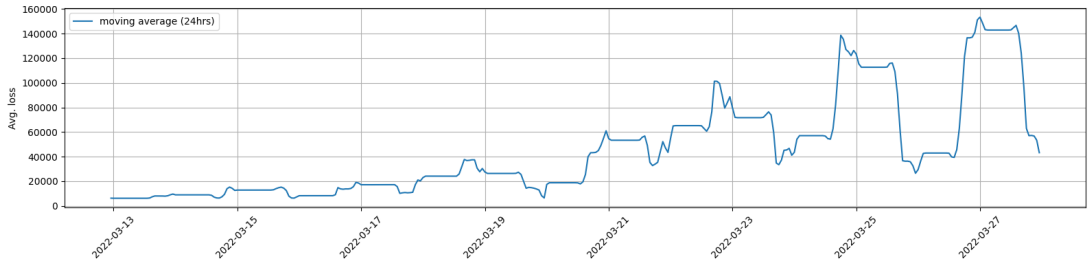


Fig. 10: Trend of a moving average (window size = 24 hours) for MSE errors of extrapolated irradiance values.

VI. Conclusion & Future Work

Weather prediction is essential in many different areas because it affects most of human endeavors, including renewable energy production and management. One of the most important renewable energy sources is solar energy through the use of solar panels. We require weather information on irradiance and temperature to calculate power from photovoltaic panels effectively.

In this study, we performed radiation forecasting for the solar panel using the transformer method. We used data containing ten years of information from the West Texas Mesonet Data Archive between 2012 and 2022. Since the values of irradiance are sampled in different intervals, we average the data hourly in order to get a fixed number of 24 observations per day.

After training the model, we explore different results regarding the training, testing, and extrapolation steps. The results show that the transformer model is able to fit the data and follow the daily irradiance pattern. In the testing prediction and extrapolation stages, we show the forecast for 16 days of data (384 hours). Our results suggests that the model can perform well at the testing stage as the predicted and observed irradiance trends hold a significant overlap, which yields to low errors. As for the extrapolation stage, we notice that the short-term predictions are more accurate than the long-term ones, as both signals (i.e., observed and extrapolated) start to offset the longer the prediction horizon is. Because of this, we aim to explore more thoroughly a different configuration of hyperparameters and architecture that allow us to improve extrapolated values of irradiance.

Finally, on the machine learning side, future work includes

irradiance prediction with multiple features, such as temperature, humidity, pressure, dew point, etc. On the electrical side, we aim to predict solar panel energy use predicted temperature (our previous work) and irradiance. We want to build an energy management system with predicted power that includes machine learning at its core. This smart system will be able to make fast decisions in critical situations. Finally, when the solar system includes a battery, we aim to predict its life cycle and control the charge and discharge processes.

ACKNOWLEDGEMENT

We would like to thank Dr. Argenis Bilbao for providing access to the dataset. This work is supported in part by the National Science Foundation (NSF) under Grant No: 1821560.

REFERENCES

[1] M. Mišík, "The eu needs to improve its external energy security," *Energy Policy*, vol. 165, p. 112930, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421522001550

[2] S. S. Mohammed, D. Devaraj, and T. I. Ahamed, "Modeling, simulation and analysis of photovoltaic modules under partially shaded conditions," *Indian Journal of Science and Technology*, vol. 9, no. 16, pp. 1–8, 2016.

[3] A. Hajiah, T. Khatib, K. Sopian, and M. Sebzali, "Performance of grid-connected photovoltaic system in two sites in kuwait," *International Journal of Photoenergy*, vol. 2012, 2012.

[4] Y. Yu, J. Cao, and J. Zhu, "An lstm short-term solar irradiance forecasting under complicated weather conditions," *IEEE Access*, vol. 7, pp. 145 651–145 666, 2019.

[5] S. M. J. Jalali, S. Ahmadian, A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "Automated deep cnn-lstm architecture design for solar irradiance forecasting," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 54–65, 2021.

[6] A. L'Heureux, K. Grolinger, and M. A. Capretz, "Transformer-based model for electrical load forecasting," *Energies*, vol. 15, no. 14, p. 4993, 2022.

[7] A. Demir, L. F. Gutiérrez, S. Bayne, and A. Bilbao, "Temperature prediction in microgrids using lstms: A case study," in *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2022, pp. 1237–1242.

[8] H. Liu, X. Mi, and Y. Li, "Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, lstm network and elm," *Energy Conversion and Management*, vol. 159, pp. 54–64, 2018.

[9] M. I. Hutapea, Y. Y. Pratiwi, I. M. Sarkis, I. K. Jaya, and M. Sinambela, "Prediction of relative humidity based on long short-term memory network," in *AIP Conference Proceedings*, vol. 2221, no. 1. AIP Publishing LLC, 2020, p. 060003.

[10] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by lstm," *Energy*, vol. 148, pp. 461–468, 2018.

[11] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparison of arima and lstm in forecasting time series," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. IEEE, 2018, pp. 1394–1401.

[12] S. Gopali, F. Abri, S. Siami-Namini, and A. S. Namin, "A comparison of tcn and lstm models in detecting anomalies in time series data," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 2415–2420.

[13] C. G. Villegas-Mier, J. Rodriguez-Resendiz, J. M. Álvarez-Alvarado, H. Jiménez-Hernández, and Á. Odry, "Optimized random forest for solar radiation prediction using sunshine hours," *Micromachines*, vol. 13, no. 9, p. 1406, 2022.

[14] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, "Transformers in time series: A survey," *arXiv preprint arXiv:2202.07125*, 2022.

[15] N. Wu, B. Green, X. Ben, and S. O'Banion, "Deep transformer models for time series forecasting: The influenza prevalence case," *arXiv preprint arXiv:2001.08317*, 2020.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[17] N. Tavakoli, S. Siami-Namini, M. A. Khanghah, F. M. Soltani, and A. S. Namin, "An autoencoder-based deep learning approach for clustering time series data," *SN Applied Sciences*, vol. 2, pp. 1–25, 2020.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.