RESEARCH

Predicting Solar Power Supply Generation in the Philippines

Sandro Luis R Silva*† and Jac Lin T Yu†

Abstract

Due to the increasing focus towards sustainable global efforts against climate change, renewable energy generation, more specifically solar power generation, has experienced a significant growth in the past few years. However, solar power production by nature is still considered highly volatile and intermittent, due to its high dependency on various factors such as irridiation, cloud cover, temperature, and other weather parameters. As such, the inherent volatile nature of solar poses significant challenges in accurate forecasting techniques not only to solar power plant operators but also to power grid operators. This paper provides a deep learning model approach in forecasting solar power production with multiple weather variables. In the paper, four sites across the Philippines were considered, and four variations of the LSTM model were implemented. For two power plants, the best model had a single layer (32-nodes) LSTM, while the other two, utilized a two layer LSTM model with 64 and 32 nodes. The modified Mean Absolute Error across all four sites reduced by at least 60% when compared to industry standard.

Keywords: Long-Short Term Memory; LSTM; Time Series; Power; Solar Power; Deep Learning

*Correspondence: ssilva@aim.edu
Asian Institute of Management, Paseo de Roxas, Manila, Philippines
Full list of author information is available at the end of the article
†Equal contributor

1 Introduction

With the increasing devastating effects of climate change, governments and industries have long been developing policies to mitigate its impact. In the COP21 or the 2015 Paris Climate Conference, all participating countries have agreed to reduce carbon emission in order to contain global warming to below 2 °C. In addition, the United Nation's 2030 Sustainable Development Goals, which aims to address various societal issues, include provisions to produce affordable and clean energy (SDG 7) and provide climate action (SDG 13). Thus, the global energy industry has veered its focus away from traditional fuel sources to renewable and sustainable solutions, such as hydroelectric, solar and wind.

1.1 Statement of the Problem

The shift towards renewable energy poses new threats and challenges to existing power grids across the world due to the inherent variability and intermittency of solar power. The problem then is "Using neural networks, can solar power generation be predicted better than what industry currently uses given weather data?"

1.2 Significance

Accurate forecasting of solar power can help solar power plant operators reduce the risk of unnecessary market penalties. In addition, forecasting accurate solar power generation may provide power grid operators the ability to balance and schedule the distribution of generated power, for not only renewable power operators, but also conventional (and rigid) power plants, such as coal and natural gas. The overall

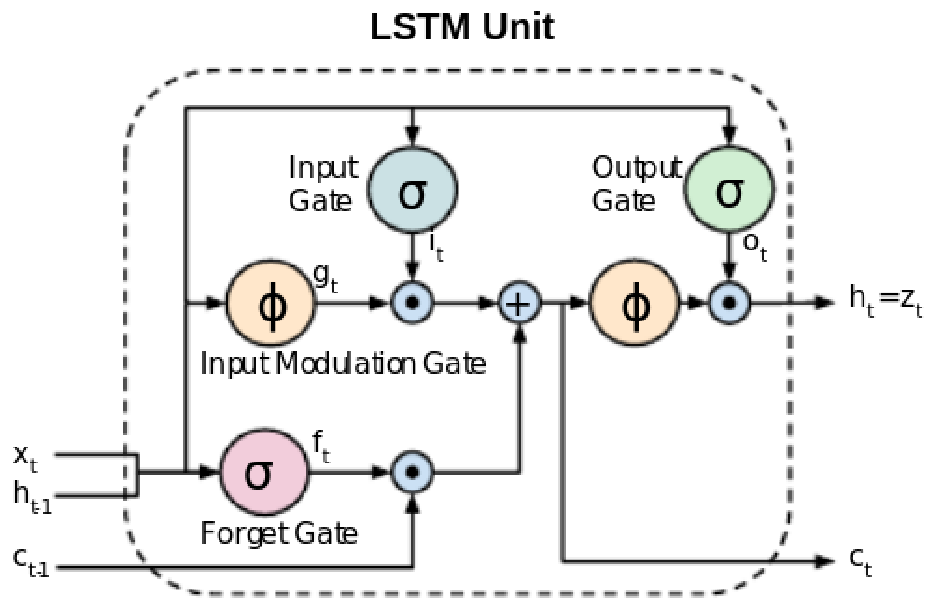

Figure 1 Structure of an LSTM cell

significance of this study is to ensure a more secure and stable power grid even with high solar power penetration.

1.3 Scope and Limitation

The study only considers four solar power plants located in North Luzon, Greater Metro Manila, Visayas. No plants in the South Luzon and Mindanao regions were considered. Only 2 years and 8 months worth of weather data was included in the study. This data, however, does not include solar irridiance, an important indicator of solar power generation, due to the lack of availability.

2 Proposed Framework

2.1 Long-Short Term Memory

Long Short Term Memory or LSTM is a deep learning method proposed in 1997 to address the vanishing or exploding gradient problem encountered in Recurrent Neural Networks when dealing with time series forecasting.(1) It does this by integrating a memory cell that allows the learning of complex and long-term temporal dynamics of the data.(2) This memory cell determines which unimportant data should be forgotten and which important feature should be remembered. Figure 1 shows the gates mechanism that allows the interaction of the memory cell with its environment.(3) The input gate (i_t) seen in equation 1 and input modulation gate (g_t) seen in equation 2, where $\mathbf{W}$ are the weights matrices and $\mathbf{b}$ are the bias terms, allows the updating of the cell.

$$\mathbf{i}_t = \sigma(\mathbf{W}_{xi}\mathbf{x}^t + \mathbf{W}_{hi}\mathbf{h}^{(t-1)} + \mathbf{b}_i) \tag{1}$$

$$\mathbf{g}_t = tanh(\mathbf{W}_{xg}\mathbf{x}^t + \mathbf{W}_{hg}\mathbf{h}^{(t-1)} + \mathbf{b}_g) \tag{2}$$

Equation 3 shows the computation of the forget gate (f_t) which determines if the information is suppressed or allowed to pass through and equation 4 shows the computation of the output gate (o_t) which decides how the hidden states are updated.

$$\mathbf{f}_t = \sigma(\mathbf{W}_{xf}\mathbf{x}^t + \mathbf{W}_{hf}\mathbf{h}^{(t-1)} + \mathbf{b}_f) \tag{3}$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{xo}\mathbf{x}^t + \mathbf{W}_{ho}\mathbf{h}^{(t-1)} + \mathbf{b}_o) \tag{4}$$

At the current time step t, the cell is updated by equation 5 and the values of the hidden units are updated by equation 6.

$$\mathbf{C}^t = (\mathbf{C}^{(t-1)} \odot \mathbf{f}_t) \oplus (\mathbf{i}_t \odot \mathbf{g}_t) \tag{5}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot tanh(\mathbf{C}_t) \tag{6}$$

2.2 Evaluation Metric

Traditionally, researchers utilized classical statistical indicators used in solar power forecasting, specifically the Mean Absolute Error (MAE) (4)(5).

Mean Absolute Error (MAE): MAE refers to the average distance between the measured values and forecasting model. This analysis is appropriate to evaluate uniform prediction errors.

$$MAE = \frac{1}{N}\sum_{n=1}^{N} |P_{pred} - P_{act}| \tag{7}$$

where P_{pred} defines the predicted energy generation by the model and P_{act} defines the actual energy generation

However, the MAE fails to consider power grid elements and requirements. It is therefore important to utilize performance metrics that asses power generation forecasting by considering various forecast time-scales, and capacity restrictions. In addition,Sobri et al. indicated that depending on the use case, metrics assessments can be classified into four types i.e., statistical, ramp characterization, uncertainty quantification, and economic metrics. (5) As such, Almeida et al., implemented a modified MAE, defined as cvMAE, which evaluates a model by that penalize the hourly or daily energy error. The cvMAE measures the goodness of the predictions for applications requiring hourly predictions during a period of a day. (6)

$$cvMAE = \frac{MAE}{P_{mean}} \tag{8}$$

where P_{mean} defines the mean of measured power in a given time-horizon.

In this study, a modified version of the MAE and $cvMAE$ will be utilized. The proposed metric takes into consideration not only classical statistical error measurement observed in the MAE, but also energy industry related considerations, such as day-ahead projected generation and rated capacity of each specific plant.

$$mMAE = \frac{\frac{1}{N}\sum_{n=1}^{N} |P_{pred} - P_{act}|}{P_{max}} = \frac{MAE}{P_{max}} \tag{9}$$

where P_{max} indicates a solar power plant's peak installed capacity, or the intended full-load sustained output of the power plant.

3 Methodology

3.1 Dataset

Energy generation, since December 26, 2016 until August 25, 2019, were collected from the Philippines' Wholesale Electricity Spot Market (WESM). To reduce the scope of the study, only four solar power plants located across Luzon and Visayas were selected. These plants, along with the location and rated dependable capacity, are seen in Table 1.

Corresponding weather data, in each of these four locations, were collected separately. The weather data contained 10 features: Temperature (in degrees Fahrenheit), Humidity (in percentage), Dew Point (in degrees Fahrenheit), Pressure (in inches), Precipitation (in inches), Cumulative Precipitation (in inches), Wind Speed (in mph), Wind Gust (in mph), Wind Direction (direction of wind), and Weather Condition.

Granularity of the data is in 1-hr intervals, which results in a total of 23160 observations per plant. The experiment implemented a 80-10-10 training, validation and test split.

3.2 Models

3.2.1 Industry Benchmark Model

For the industry benchmark, P_{pred} is defined as each solar power plant's day-ahead hourly projection submission to the market operator, referred to as the Ex-Ante or RTD in the Philippine energy trading industry, and P_{act} is defined as the actual scheduled dispatch by the solar plant considering the market operator's dispatching schedule, referred to as the Ex-Post or RTX in the Philippine energy trading industry. To contextualize, the market operator utilizes each solar power plant's day-ahead submission in the scheduling, prioritization, and dispatching of energy production. Since solar power plant's are considered non-scheduled generators, or must-run units, these energy producers are prioritized by the market operators and are allowed to generate at their full capacity.

3.2.2 Proposed Models

Four proposed models were utilized in this study all focusing on the LSTM neural network. Networks 1 and 2 considered a single LSTM layer with nodes, 64 and 32 respectively, while networks 3 and 4 considered a stacked two layer LSTM network with nodes 64-32 and 32-64 respectively. An additional feed-forward hidden layer with an rectified linear unit activation function was utilized at the end of all four networks, where the number of hidden nodes corresponded to the number of hidden nodes of the last LSTM layer. Network 1 and 4 utilized 32 nodes, while networks 2 and 3 utilized 64 nodes. All four networks considered the RMSProp optimizer, dropouts and recurrent dropouts of 0.2, batch size of 128, and an epoch setting of 50.

4 Results

Table 2 shows the results of for each location. The stacked 2 layer LSTM with 32 and 64 nodes is the best performing model for the Clark Solar Power Plant and the Valenzuela Solar Power Plant, whereas the single layer LSTM model with 32 nodes best fits the First Toledo Solar Power Plant and the Subic Solar Power Plant.

For the Clark Solar Power Plant, the initial industry benchmark error was computed at 12.01%, and the 2 layer LSTM with 32 and 64 nodes model error is at 2.86%, which translates to a 76.19% error reduction.

For the First Toledo Solar Power Plant, a 15.34% industry benchmark error was computed, and the single layer LSTM with 32 nodes model produced only a 4.30% model error, which translates to a 71.97% error reduction.

For the Subic Solar Power Plant, the initial industry benchmark error was computed at 3.64%, and the single layer LSTM with 32 nodes model error is at 1.36%, which translates to a 62.64% error reduction.

For the Valenzuela Solar Power Plant, a 12.48% industry benchmark error was computed, and the two layer LSTM with 32 and 64 nodes model produced only a 2.58% model error, which translates to a 71.33% error reduction.

5 Conclusion and Recommendation

Improvements brought upon by the utilization of deep learning models can reduce costs for both suppliers and the national grid operators, and increase the efficiency of power generation.

Further studies can be conducted using different neural network architectures, varying numbers of nodes, layers, and optimizers of the LSTM layer, adding dense layers with different nodes. Including data that are main drivers of solar power generation, such as quality and angle of solar panels and the hourly irridiance, can drastically improve results. A study that produces one model for all solar power plants may also be considered. Location and descriptors of the power plant can be included which is fed to a CNN network together with an LSTM model will consider both spatio-temporal features.

Declaration

Availability of data and materials

The datasets generated during and/or analysed during the current study are available in the Wholesale Electricity Spot Market (http://www.wesm.ph/) and Weather Underground (https://www.wunderground.com/) websites.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Author's contributions
SS performed pre-processing of the data for the development of the models. SS and JY performed model selection and evaluation. Both were major contributors in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements
Not applicable

References

1. Zhang J, Chi Y, Xiao L. Solar Power Generation Forecast Based on LSTM. In: 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS); 2018. p. 869–872.
2. Qing X, Niu Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy. 2018 apr;148:461–468.
3. Rassem A, El-Beltagy M, Saleh M. Cross-Country Skiing Gears Classification using Deep Learning. 2017 06;.
4. Wang K, Qi X, Liu H. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. Applied Energy. 2019;251(C):1–1. Available from: https://ideas.repec.org/a/eee/appene/v251y2019ic30.html.
5. Sobri S, Koohi-Kamali S, Abd Rahim N. Solar photovoltaic generation forecasting methods: A review. Energy Conversion and Management. 2018 01;156:459–497.
6. Almeida MP, Perpiñán O, Narvarte L. PV power forecast using a nonparametric PV model; 2015. .

Tables

Table 1 Power Plant Description

Power Plant	Location	Installed Capacity
Clark Solar Power Plant	Mabalacat, Pampanga	22 MW
First Toledo Solar Power Plant	Toledo, Cebu	60 MW
Subic Solar Power Plant	Olongapo, Zambales	100 MW
Valenzuela Solar Power Plant	Valenzuela, NCR	8.5 MW

Table 2 Summary of Benchmark vs Model Performances

Power Plant	Best Model	Industry Benchmark	Model Performance	Improvement
Clark Solar Power Plant	LSTM+LSTM with 32 & 64 nodes	2.86%	12.01%	76.19%
First Toledo Solar Power Plant	LSTM with 32 nodes	4.30%	15.34%	71.97%
Subic Solar Power Plant	LSTM with 32 nodes	1.36%	3.64%	62.64%
Valenzuela Solar Power Plant	LSTM+LSTM with 32 & 64 nodes	2.58%	12.48%	79.33%