Research Plan

Project code: EP001

Project title: Prediction of Solar Power Potential in Singapore using Numerical Weather Prediction Models and Machine Learning

a) Rationale:

When it comes to energy generation, solar power remains the most promising renewable energy source in Singapore. With an average solar irradiance of 1,580 kWh/(m^2*year) and about 50% more solar radiation than temperate countries, solar photovoltaic (PV) generation has the greatest potential for wider deployment in Singapore. [1] Moreover, a report by National Renewable Energy Lab [2] has indicated that from 2010 to 2017 there has been approximately a 70% reduction in the total PV system hardware cost. Combining great potential and decreasing cost, there is a strong demand to deploy solar power widely in Singapore. Precise solar radiation prediction will be the key to successful integration of solar power.

Currently, Weather Research and Forecast (WRF) model [3] has been widely used for the weather forecast in the study [4] [5]. Additionally, the model has been proved to be successful in power output forecasting in the context of Smart Grids and Renewable Energy on [6]. However, the WRF model has its limitation of long running time [7] [8] and it took a few weeks in total to generate the whole dataset of solar radiation data at multiple different locations across one year used in this study.

Due to the high cost of time in applying the WRF model, there is a strong demand to develop alternative ways to make the necessary predictions. [9] Machine Learning [10] that research different algorithms, which can learn information from datasets, can be a useful tool to make predictions with much shorter time taken [11]. Essentially, various algorithms can accept a given set of training examples and derive a general trend over time. [5] There are different machine learning algorithms, such as Support Vector Machine (SVM), Random Forest, Time Series and so on. [12] [13] [14] [15] [16] [17] [18] [19] The focus of this study is to learn about the algorithms and apply them to this problem. With the best model evaluated by different metrics like MAE, MSE, and RMSE, we can have an alternative way to make accurate predictions of solar radiation with less time. [20]

b) Research Question(s), Hypothesis(es), Engineering goal(s) and Expected Outcome(s)

The **Research Question** is that by building machine learning models in replace of WRF, can we find an alternative way to predict solar radiation in Singapore with less time taken?

The **hypothesis** is that through computational experiments we can build one or more machine learning models that can make solar potential predictions with faster speed.

The **engineering goal** is to develop a machine learning model to interpolate the solar radiation potential of multiple certain locations across Singapore based on the data generated by the WRF model for the past one year. This model can then be used to make predictions of the future solar potential at a certain time of interest at any given location.

The **expected outcome** is to build a machine learning model that can accurately predict the solar radiation data based on location and time with the smallest error with the much shorter time taken compared with the WRF model.

c) Procedures, Risk and Safety, Data Analysis

Procedures

The original data is the solar radiation data at multiple different locations across the country generated by the WRF model for the period 1 Jan 2018, 8 am - 1 Jan 2019, 8 am. It consists of four columns and processed to the format that python can accept.

In order to have a general idea about solar radiation potential in Singapore, graphs of Direct Solar Radiation (SWDIR) and Diffuse Solar Radiation (SWDIF) values at various locations between different time intervals (a day, a week and a month namely) will be plotted. As the predictions are in terms of CSR, defined as

$$\text{CSR} = \frac{\text{SWDIF}}{\text{SWDIF} + \text{SWDIR}} \qquad (1)$$

with its range from 0 to 1 by definition, the NaN values (not a number) of CSR with both SWDIR and SWDIF zero should be removed or converted to a usable format.

After preprocessing the data, we can start to work on building the model. Different inputs can help machine learning models to derive different features. In order to derive main features of the dataset, including seasonality, spatial correlation and so on [21], the training features are not only restricted with the time and corresponding CSR but also include the location features and the correlation of CSR at different places. A few models based on classification or regression (Support Vector Machine, Random Forest, Time series model and so on) [12] [13] [14] [15] [16] [17] [18] [19] will then be built and able to accurately predict the CSR for any time during the year.

The models built are going to be tested with the test sets separated ahead and predictions made will be compared against the actual result in the test set. By using different metrics including the mean average error (MAE), mean square error (MSE) and root mean squared error (RMSE), the accuracy of each model can be quantified and compared against one another. The model with the best performance will be chosen.

Risk and Safety

As this project is purely computational, there are very few safety issues involved.

However, there are risks present as the accuracy of the data decreases. In order to obtain a general trend from the dataset, details and small variance are intentionally ignored. Moreover, machine learning models usually have the restriction of computing power. As a result, lower accuracy of machine learning models is caused compared with the WRF model. [22]

Data Analysis

In order to assess how accurate a model is, a baseline model, which creates a trend of constant straight line regardless input values, is created and compared against all the other models built to know the improvement of the model in its accuracy. To quantify the difference in comparison, different metrics are employed including Mean Average Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE). MAE can be used to measure the overall variation with the same weight for all values, while MSE will give higher weight to larger variations from the actual result. RMSE, with the same rationale behind as MSE, can be interpreted as the standard deviation from the actual trend and can be easily understood for having the same units as the corresponding variable. By combining the result of four metrics, the accuracy of a model can be completely reflected.

d) Bibliography

1. https://www.ema.gov.sg/solar_photovoltaic_systems.aspx (accessed on 20 Nov 2018) ↵

2. Ran Fu, David Feldman, Robert Margolis, Mike Woodhouse, and Kristen Ardani. U.S. solar photovoltaic system cost benchmark: Q1 2017. Sep. ↵

3. Numerical Weather Prediction (Weather Models) https://www.weather.gov/media/ajk/brochures/NumericalWeatherPrediction.pdf (accessed on 15 Nov 2018) ↵

4. T.M.Giannaros, V.Kotroni, and K.Lagouvardos. Predicting lightning activity in Greece with the weather research and forecasting (wrf) model. Atmospheric Research, vol. 156. 1 - 13, 2015. ↵

5. Terren-Serrano, Guillermo. "Machine Learning Approach to Forecast Global Solar Radiation Time Series."(2016).[http://digitalrepository.unm.edu/ece_etds/249] (Stephan Rasp and Sebastian Lerch. Neural networks for post-processing ensemble weather forecasts. arXiv: 1805.09091v1 [star.ML] 23 May 2018) ↵↵

6. D.Carvalho, A. Rocha, M. Gomez-Gesterira, and C. S. Santos. Sensitivity of the WRD model wind simulation and wind energy production estimates to planetary boundary layer parameterizations for onshore and offshore areas in the Iberian peninsula. Applied Energy, vol. 135, pp. 234-246, 2014. ↵

7. A.M. Guerrero-Higueras, E.Garcia-Ortega, J.Lorenzana and V.Matellan. Schedule WRF model executions in parallel computing environments using Python. ↵

8. Tolstykh, Mikhail & Frolov, Alexander. (2005). Some Current Problems in Numerical Weather Prediction. Izvestiya Atmospheric and Oceanic Physics. 41. 285-295. ↵

9. U. Divya, Chitra Pasupathi. A machine learning approach to predict solar radiation for solar energy based devices. By International Journal of Computer Science & Engineering Technology (IJCSET) ↵

10. http://fortune.com/2016/09/14/data-machine-learning-solar/ (accessed on 10 Nov 2018) ↵

11. Haupt, Sue & Kosovic, Branko. (2015). Big Data and Machine Learning for Applied Weather Forecasts Forecasting Solar Power for Utility Operations. 10.1109/SSCI.2015.79. ↵

12. Cyril Voyant, Gilles Notton, Soteris Kalogirou, Marie-Laure Nivet, Christophe Paoli, Fabrice Motte, Alexis Fouilloy. Machine Learning methods for solar radiation forecasting: a review. Horizon 2020 project (H2020-LCE-2014-3 - LCE-08/2014 - 646529) TILOS "Technology Innovation for the Local Scale, Optimum Integration of Battery Energy Storage". ↵↵

13. Philippe Lauret, Cyril Voyant, Ted Soubdhan, Mathieu David, Philippe Poggi. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. Solar Energy, Elsevier, 2015, pp.00. ↵↵

14. Seckin Karasu, Aytac Altan, Zehra Sarac, Rifat Hacioglu. Prediction of solar radiation based on machine learning methods. Bülent Ecevit University (BAP Project No: 2012-17-15-01 and 2014-75737790-01) and International Engineering Research Symposium-UMAS 2017(Duzce University) ↵↵

15. Sotiris Vardoulakis, Bernard E.A. Fisher, Koulis Pericleous, Norbert Gonzalez-Flesca. Modelling air quality in street canyons: a review. Atmospheric Environment, Elsevier, 2003, 37 (2), pp.155-182. ↵↵

16. Wei-Zhen Lu, Wen-Jian Wang. Potential assessment of the "Support Vector Machine" method in forecasting ambient air pollutant trend ↵↵

17. Sanyam Gupta, Infumath K, Govind Singhal. Weather Prediction Using Normal Equation Method and Linear regression Techniques. (IJCSIT)↵↵

18. Stephan Rasp and Sebastian Lerch. Neural networks for post-processing ensemble weather forecasts. arXiv: 1805.09091v1 [star.ML] 23 May 2018↵↵

19. Min-Kyu Baek and Duehee Lee. Spatial and Temporal Day-Ahead Total Daily Solar Irradiation Forecasting: Ensemble Forecasting Based on the Empirical Biasing↵↵

20. Ricardo Aler, Ricardo Martin, Jose M. Valls, and Ines M. Galvan. A Study of Machine Learning Techniques for Daily Solar Energy Forecasting using Numerical Weather Models.↵

21. Xie-Kang Wang, Wei-Zhen Lu. Seasonal Variant of air pollution index: Hong Kong case study↵

22. Aoife M.Foleyabd, Paul G.Leahyab, AntoninoMarvugliac, Eamon J.McKeoghabCurrent methods and advances in forecasting of wind power generation ↵