Solar Power Generation Forecast

Noah Howren

CS458

Introduction

As there are more and more environmental restrictions in place to protect the planet, it is essential that we understand the strengths and drawbacks of different renewable resources in various climate conditions that exist across the planet. Solar power is already an important renewable energy source today and going into the future the best way to understand the scalability and optimization of this resource is to be able to reasonably understand the interactions with certain climate conditions and the power output of a given solar panel grid. This in mind, considering technology has now provided anyone with decent programming abilities to take advantage of data analysis capabilities of certain programming language libraries to take data, analyze it, and forecast results going forward. With the proper tuning of aspects that make up these libraries, one can certainly use the capabilities to reasonably forecast solar power generation.

In this project I was tasked with taking data of 3 different solar power plants power production hour by hour and the corresponding climate data recorded and creating an accurate forecasting solution using a decided upon model. To do this I created a mental roadmap for myself with a series of goals to solve the various problems that go into optimizing this forecasting. I knew I was going to have to start by importing the data into time series models divided by which plant the data was recorded from. Also as part of this data initialization portion of the project, I went about splitting the data into its proper test and training sets.

After initializing this data, I would have to determine which library would be optimal for analyzing this time series data and creating accurate forecasts. Following this, I knew it would be essential to further optimize this library and the parameters in place to achieve the best results possible. After receiving the results, it is important to analyze the data, through outputting its graphs and metrics like MAE and RMSE to ensure the successfulness of the forecasting.

Background and Related Work

There are a few journal articles that pertain to problems similar to that of which I'm tasked with solving for this project. The first being *Short-term forecasting of temperature driven electricity load using time series and neural network model* by Nengbao Liu, Vahan Babushkin, and Afshin Afshari. This article in particular was studying the optimal model (between SARIMAX and ANN) for a week ahead forecasting system for a temperature driven electricity load. The second article I took note of was *Fuzzy modeling to forecast an electric load time series* by Cesar Machado Pereira, Nival Nunes de Almeida, and Maria L.F. Velloso. This article took a look at the Sarimax model in comparison to a Fuzzy modeling system and its utility in forecasting the electrical load required in the metropolitan region of Bahia State.

These articles had different findings that impacted my process in deciding which model would work best for my case as well as helping with the process of tackling the problem in the end. The first important finding from these articles was from the first which found in comparison to a ANN approach, a SARIMAX was a decent bit more accurate for forecasting a temperature driven electricity load both in the case of predicting a week at a time, and a day at a time. Furthermore, the second article found that SARIMAX was once again more apt in forecasting an electrical load when compared to a Fuzzy inference model. The conclusion I can draw from this, is that SARIMAX is a quite good linear regression approach to handle problems of electricity production and consumption. I attribute this to the seasonality functionality that this model has. This seasonality functionality definitely has a positive effect on solving these types of issues because as the climate conditions change through the seasons, there is a measurable impact on renewable energy production as well as electricity consumption.

Methods

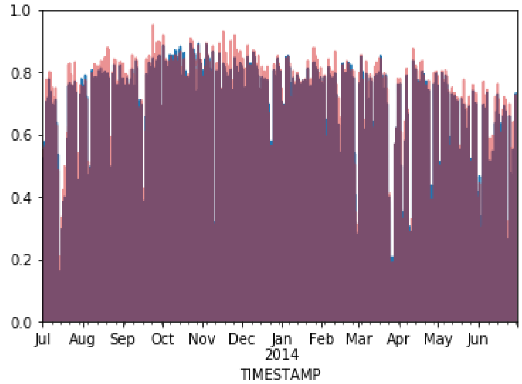
I began my work on this problem by implementing a pandas time series import of the spreadsheet containing the full dataset. When I had a list of all the data points as a whole, I parsed it down into which plant the data had come from using exact indices found from the indexing of the file. From there, I once again parse these individual plant datasets to their individual training and testing sets using indices calculated from the indices in the original file. In order to ensure I have the proper points in each dataset, I output the head of each dataset.

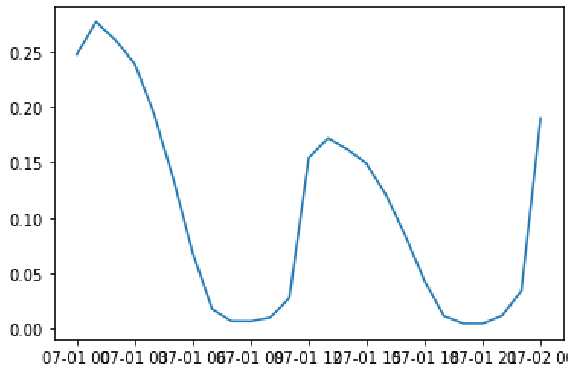
With the findings from previous similar works in mind, I figured that the SARIMAX linear regression model definitely would be a great choice in tackling this project due to the aforementioned seasonality functionality. For implementing this model I had to do seasonal decomposition for each of the plants training dataset. Following that, I implemented a seasonal iteration functionality for each plant to make the modeling go over more optimally.

Finally, I did some work to ensure that I was doing each process correctly. First, I output graphs for a full test dataset length and for a single day forecast on the end of the dataset. Following that I did calculations of both the mean absolute error and root mean square error for each plant individually, and of the dataset as a whole.

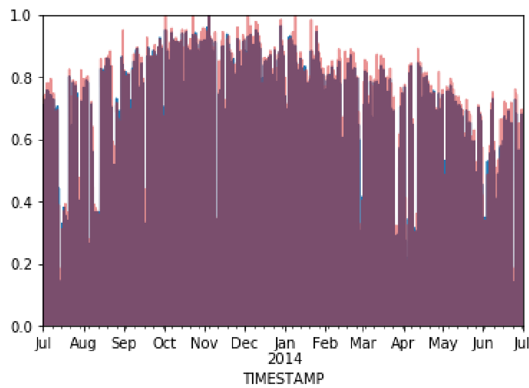
Evaluation of Results

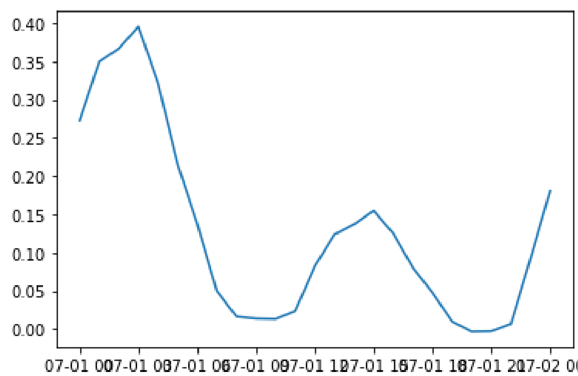
<u>PLANT 1</u>



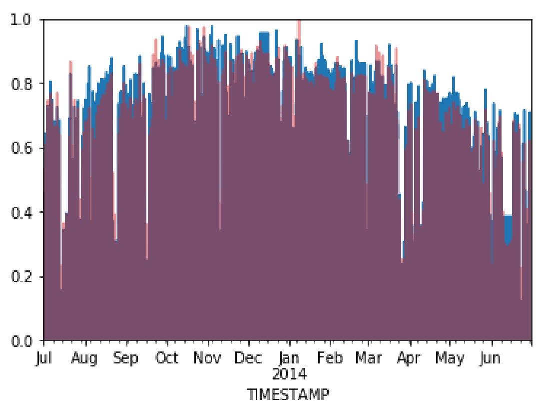


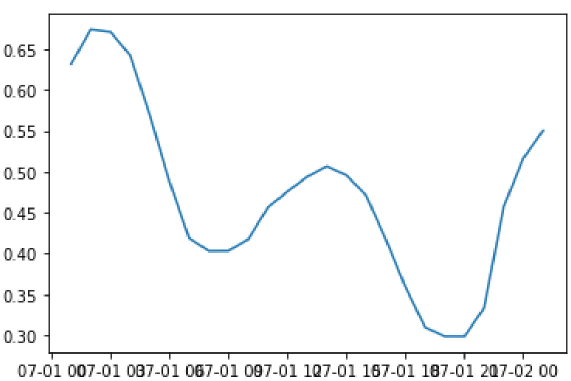
<u>PLANT 2</u>





<u>PLANT 3</u>





The purple area on each graph on the right represents the area where predicted and actual overlap. If the purple stops and there is pink for a certain datapoint, that pink area is an overprediction. If the purple stops and there is blue above it for a certain datapoint, that blue is the actual prediction, showing there is underprediction for that datapoint.

The way I achieved these results is from running my python notebook file (SPGProject--NoahHowren.ipynb) running on an Ubuntu virtual machine with the notebook file open in Visual Studio Code.

ZONEID	***PLANT 1***	***PLANT 2***	***PLANT 3***	***OVERALL***
MAE	0.0551	0.0634	0.0578	0.0588
RMSE	0.0844	0.0922	0.0864	0.0876

From these results, one can draw the conclusion that plant 1 was most successful in forecasting with this model. Furthermore we see that plant 2 is least successful in forecasting with this model, but by no means, extremely inaccurate in its forecasting.

Conclusion

I learned a lot about the practical applications of linear regression modeling in this project. It was very interesting to create a program with the functionality to do as accurate forecasting as I have for this dataset. I can conclude from the results in this project that the SARIMAX modeling system is quite successful in this use case. I look forward to applying the knowledge and experience I acquired in doing this project to future work.

References

Liu, N., Babushkin, V., & Afshari, A. (2014). Short-term forecasting of temperature driven electricity load using time series and neural network model. *Journal of Clean Energy Technologies*, *2*(4), 327-331. https://www.researchgate.net/profile/Afshin-Afshari/publication/263167726_Short-Term_Forecasting_of_Temperature_Driven_Electricity_Load_Using_Time_Series_and_Neural_Network_Model/links/5729a0e908ae2efbfdb7f84c/Short-Term-Forecasting-of-Temperature-Driven-Electricity-Load-Using-Time-Series-and-Neural-Network-Model.pdf

Pereira, C. M., de Almeida, N. N., & Velloso, M. L. (2015). Fuzzy modeling to forecast an electric load time series. *Procedia Computer Science*, *55*, 395-404. https://www.sciencedirect.com/science/article/pii/S1877050915015641