At Home Solar Forecasting: Long Short-Term Memory Residential Solar Power Forecasting with Personal Weather Station Measurements

Daniel Shannon
djshanno@syr.edu
Syracuse University
College of Engineering and Computer Science
June 13th, 2024

https://github.com/radioxeth/photovoltaic-weather-station

Abstract
We combine Personal Weather Station (PWS) and Residential Solar Power (RSP) data to forecast solar power generation using next-day local weather forecasts. Our approach is to train a Long Short-Term Memory (LTSM) neural network using 15-minute aggregated PWS and RSP data collected from March 2020 through May 2024. We then apply the model to a National Weather Service (NWS) forecast to predict the next day's solar production. To reduce the dimensionality of the data, we will explore the data to determine independent features. The independent features of the data will train the model based on a 24 hour sliding window. A sample of data is 24 hours of 15 minute interval data, which results in 96 data points per sample. The dataset contains 4 years of 15-minute interval solar and weather data.

Through LSTM and proper data modeling, we are able to generate RSP forecasts that closely resemble the recorded power profile for a 24 hour period. The best error achieved is an MSE of 448 MW. This is achieved using 12 input features with 8 hidden layers and 256 nodes per layer, and a sigmoid activation function after 90 epochs of training a LSTM model. One issue is dealing with winter precipitation - snow can cover solar panels for several days after a precipitation event. This could be addressed through a more complex model or through additional features, such as snowfall and UV intensity.

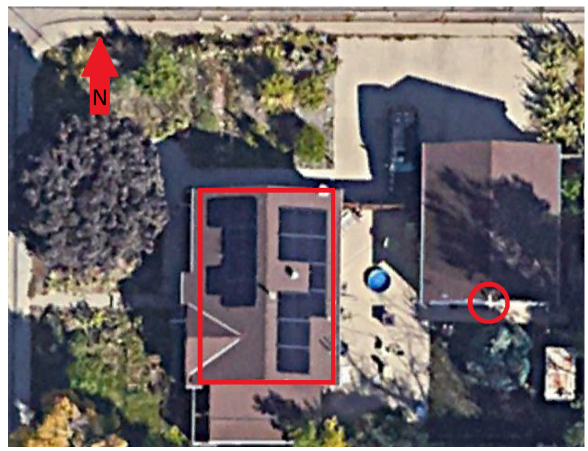

Figure 1. Residential solar panel array (box) and personal weather station (circle) subject location in Denver, CO.

Introduction

Three years of historical 5-minute PWS and 15-minute RSP data collected from a single family home in Denver, Colorado (**Figure 1)** are used to forecast next-day solar power output. The collected data is examined and the independent features are extracted to match the 2^d=n rule for determining the number of features compared to the number of available samples. In addition to manual feature reduction, Principal Component

Analysis is used in an attempt to improve the model's performance.

The features are passed as a 24 hour sliding window to a LSTM Recurrent Neural Network (RNN). The performance is measured by comparing the cumulative predicted power to the recorded power generation for each day. These cumulative values are then used to calculate the Mean Square Error for the collection of days. The model is trained on historical PWS and collected RSP data. The 24 forecast provided by NOAA is used to model the next 24 hour power generation cycle.

Previous Work

There are over 250,000 personal weather stations globally ("PWS Network Overview") and over 2.3 million residential solar installations in the US as of 2022 (Barbose et al.). As more individuals install RSP, there may arise the desire to forecast the performance of panels for the purpose of utility bill prediction and emergency preparedness in cases where the panels are connected to an on-site battery. Additionally, utilities need to understand the load of power generation relative to the capacity of the available infrastructure.

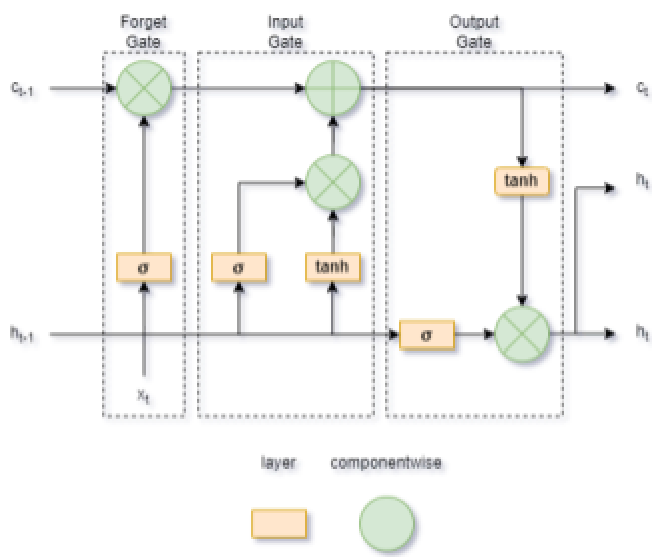

Figure 2. State flow through the memory gates of a long short-term memory RNN model. *c(t)* is the cell state and h(t) is the unit output. (Qin, Chuan, et al.)

Feature Reduction

The PWS used in this experiment generates over 30 data features. A number of them are variations of the base data types collected, including precipitation, temperature, humidity, pressure, windspeed, and wind direction. There are derivatives of these such as dewpoint, wind chill, and heat index, as well as pressure trend, maximum, minimum, and average values over a period of time.

De Guia, et al. have shown that Principal Component Analysis (PCA) can be used with data normalization to train LSTM to predict solar irradiance values. This is similar to photovoltaic power generation as they are directly related. Variance can be measured for each principal component to reduce the variance in each component.

Forecasting Methods

Along with increasing residential solar panel installations, there is a growing interest in forecasting residential solar power generation. Classifiers such as random forests and decision tree models have been used to forecast solar and weather data. Random forests perform best out of statistical based methods when compared to linear regression, decision trees, and support vector machines, per Carrera et al.

Numerous studies show the usefulness of LSTM and RNN models in building weather and solar power forecasting models (Bhagyanidhi et al.) (de Guia et al.) (Hossain and Mahmood) (Razavi et al.). LSTM models are variances of RNNs that focus on the vanishing gradient problem (Qin et al.). RNNs are designed to recognize patterns in sequences of data, such as time series data. Like other neural networks, RNNs, like ANNs (Artificial Neural Networks), have an input layer, a number of hidden layers, and an output layer. Each layer contains one or more nodes and each node of each layer is connected to each node of the next layer by weights. The weights are used to determine the relevant features in mapping the classification to the training data. Weights, along with the activation function, allow the algorithm to minimize the error between the predicted and actual output values.

LSTM models are an extension of RNNs in that they retain information from previous inputs. The failure of RNNs with regards to time series data is that they process each input one sample at time, so they lose the information from the previous sample. LSTM fixes this by including gates and cell state. This allows the neural network to maintain information from previous training inputs, thus allowing for a better assessment of time-series based classification. One downside to LSTM models is that addressing the vanishing gradient can negatively impact runtime of training as more information is retained.

Performance Evaluation

Regression metrics can be used to measure the performance of time series LSTM models. Mean square error (MSE) (**1**), mean absolute error (MAE) (**2**), explained variance score (EVS) (**3**), and regression score functions (R2) (**4**) have been used in previous studies to evaluate the performance of LSTMs (Bhagyanidhi et al.).

$$MSE = \frac{1}{n}\Sigma(y - \widehat{y})^2 \quad \mathbf{1}$$

$$MAE = \frac{1}{n}\Sigma|y - \widehat{y}| \quad \mathbf{2}$$

$$EVS = 1 - \frac{var(y-\widehat{y})}{var(y)} \quad \mathbf{3}$$

$$R^2 = 1 - \frac{SS_{sum}}{SS_{total}} \quad \mathbf{4}$$

Experiment Design

Design Overview

The overall experiment design consists of data collection, sampling, normalization, outlier analysis, dimension reduction, regression, training, and performance evaluation (**Figure 3**). Data collection is completed through RESTful APIs. Sampling, normalization, and outlier analysis level the sample to 15-minute intervals and reduce variance in the values. We explore the performance of difference feature extraction methods including manual and PCA to train LSTM neural networks for using weather based forecasts to predict solar power generation.

Data Collection

PWS and RSP datasets are collected through their available Application Programming Interfaces (API). The PWS collects temperature, humidity, wind, humidity and pressure data and retains the data in 5-minute aggregated intervals. These parameters are available through the API aggregated as MAX, MIN, and AVG values. UV intensity and solar irradiance are also available through the PWS API. Because the PWS used does not collect these data, they are returned as non values. The RSP data consists of power and energy, aggregated into 15-minute intervals.

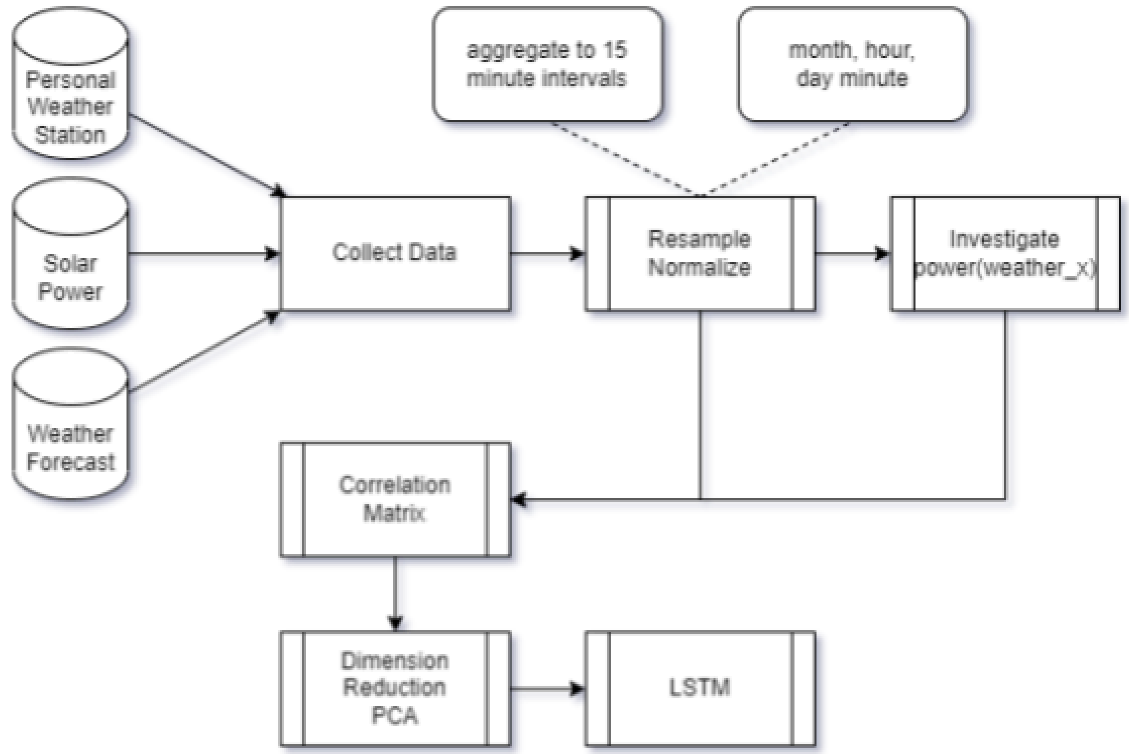

Figure 3. Experiment design and data flow from data collection to model generation.

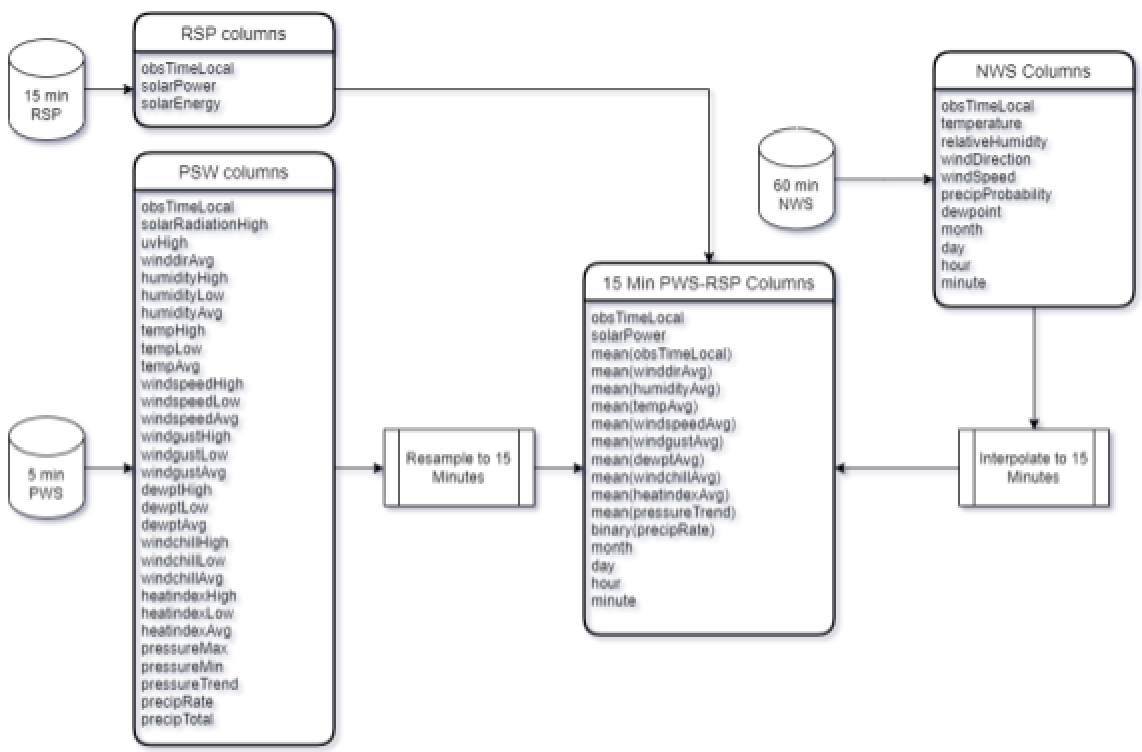

Figure 4. Data resampling and feature reduction flow for the collected RSP, PWS, and NSW forecast data.

Upon collecting the PWS and RSP data, the PWS data is resampled from 5-minute intervals to 15-minute intervals using their respective aggregation functions - max, min, and mean. The end result of scraping and resampling the data is 31 dimensions, including local observation time as seen in **Figure 4**. After extracting and leveling the 15-minute data, we separate the data into 24 hour samples - resulting in 96 rows per sample.

The data is normalized using the standard scalar. Each non-solar or time feature is standardized by removing the mean and scaling to the unit variance.

Dimension Reduction…

The 31 available dimensions can be reduced either by manually selecting the dimensions or through Principal Component Analysis (PCA). With 4 years of 15-minute data, we have ~140,160 rows. Partitioning the samples into 24 hour sections over 4 years yields 1,460 samples. To approximate the number of dimensions needed to correctly fit the model to the data, the guidance of **(5)** can be applied to determine the optimal number of features to use.

$$2^d = n \tag{5}$$

d is the number of dimensions and n is the number of samples. 10 dimensions is a good starting point to satisfy (**5**) with 1,460 samples.

Principal Component Analysis

Examining the variance in **Figure 5.** indicates that 8-10 components might be the optimal number of components to use for model training and testing. Although PCA does not perform as well as manual selection, the component variance results agree with the number of dimensions to use.

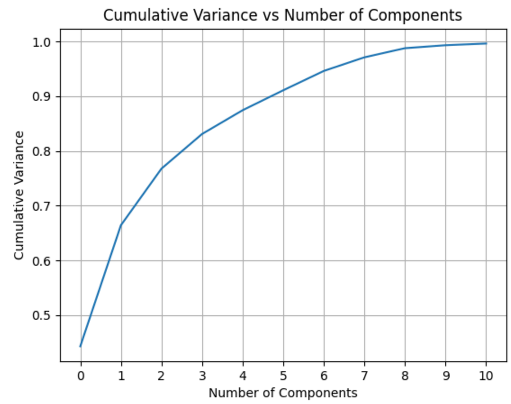

Figure 5. The PCA variance curve begins to elbow and flatten at 8-10 components.

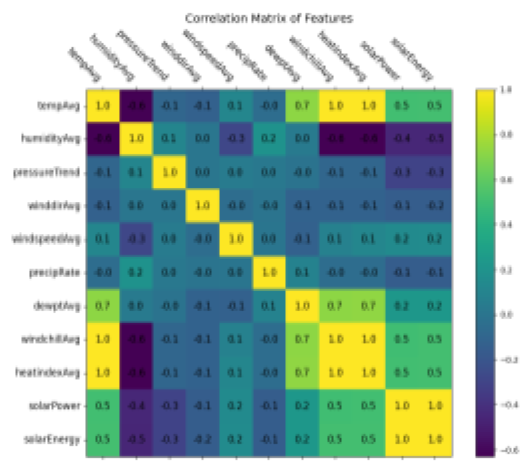

Figure 6. Correlation plot of the average values gathered from the PWS.

Manual Dimension Reduction

A correlation plot can be used to guide which features have the least amount of correlation. Features with high correlation can be discarded. In **Figure 6**, it can be seen that windchillAvg and heatindexAvg are highly correlated and can be dropped.

Four time features, six measured PWS features, one calculated PWS feature, and one measured RSP feature can be used as the final set of features to use in model training and inference:

- Time Features:
    - **Month**
    - **Day**
    - **Hour**
    - **Minute**
- PWS Measured
    - **Temperature**
    - **Humidity**
    - **Pressure**
    - **Wind speed**
    - **Wind direction**
    - **Precipitation**
- PWS Calculated
    - **Dew point**
- RSP Measured
    - **Solar power**

LSTM Training
In order to increase the number of samples for training, we implement a 24 hour sliding window with a 1 hour delta.

A sliding window can be used in training to increase the number of available samples. If n is the number of samples, w is the window size, and Δt is the window shift, then (**5**) can be applied to increase the number of training samples.

$$n_{final} = (n - 1)\frac{w}{\Delta t} + 1 \quad \mathbf{5}$$

Given n=1,095 days, w=24hrs, and Δt=1hr, a sliding window increases the number of samples available from ~1,095 to ~26,257 for 3 years of data. Since each sample has 96 rows, this increases the number of rows from 105,120 to 2,520,672.

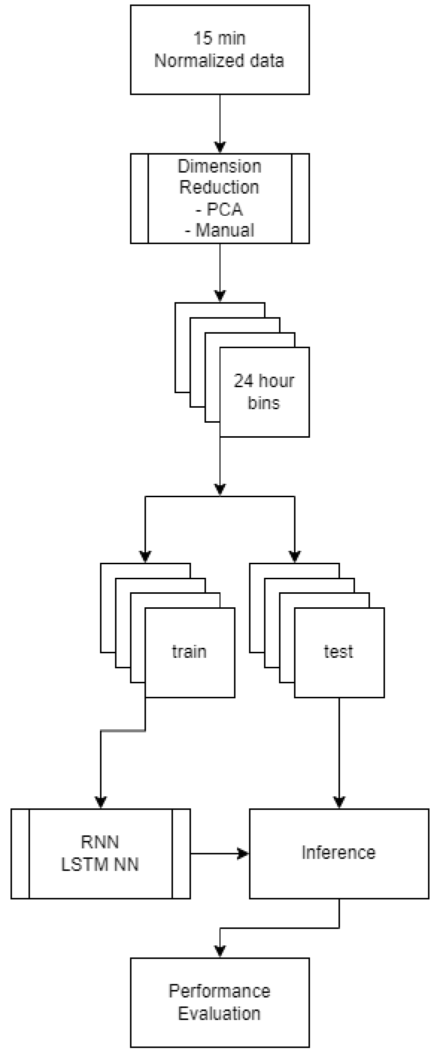

Figure 7. Model training design.

Several different model parameters were tuned during training. This includes the number of layers, nodes in each layer, activation function, epochs of training, and loss criteria.

Previous work showed that 4 hiddens layers with 128 nodes performed best for rainfall forecasting (Bhagyanidhi et al.). From here, models tracked the cumulative forecasted and recorded power while altering the model parameters. A loss threshold is implemented to prevent overfitting on trained data.

Performance Evaluation
To evaluate the performance of the model the collected samples are split 75/25 in order to capture full years of data. This split arises naturally as we can use the first 3 years of data as training data and the last 1 year as testing data. The cumulative forecasted and recorded power values from each testing sample is used to determine the MSE, MAE, and R2 values.

Results.
The model itself has seen the best results with a sigmoid activation function, with 8 layers of 256 nodes and minimizing the errors on the Mean Square Error criterion with a window loss threshold of 0.1%. When combined with 11 features that are low in correlation, we are able to achieve an r2 for the cumulative power of 0.85, an MSE of 448 MW/d, and an MAE of 15 kW/d during inference testing (**Table 1**). Applying the NWS weather forecast yields an r2 of -0.43, an MSE of 702 MW/d and an MAE of 22 kW/d (**Table 2**).

The MSE for individual days ranges from 10^2 to 10^6 W/d and the r2 ranges from -10^5 to 0.997. The model handles some situations very well, but fails on others. The largest errors are due to days when the model forecasts some power generation, but the actual power generation is 0. This is possibly due to snow remaining on the solar panels for several days after large snow events. Given the location, Denver, CO, it is possible to snow several inches or feet and then be sunny and warm for a few days while the snow remains on the panels.

Inference Results (393 samples)		
Metric	**Cumulative Production**	**Daily Mean**
MSE	448.35 MW/d	0.35 MW
MAE	15.36 kW/d	0.27 kW
r2	0.85	-18,370.12

Table 1. Test inference results for metrics of every samples' cumulative production and the mean of each day's regression metric computed individually.

Forecast Results (9 samples)		
Metric	**Cumulative Production**	**Daily Mean**
MSE	702.42 MW/d	0.86 MW
MAE	22.21 kW/d	0.47 kW
r2	-0.43	.75

Table 2. NWS forecast results for metrics of every samples' cumulative production and the mean of each day's regression metric computed individually.

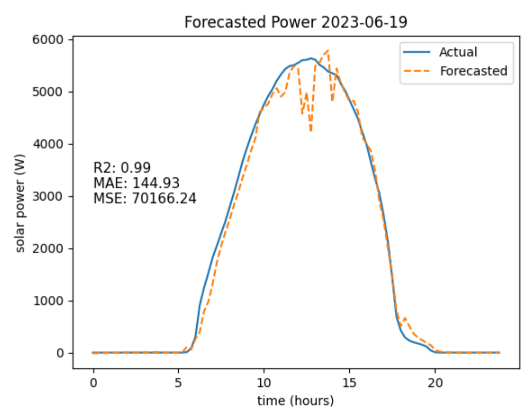

Figure 8. An accurate, simple forecast with a high r2 value and relatively low MAE.

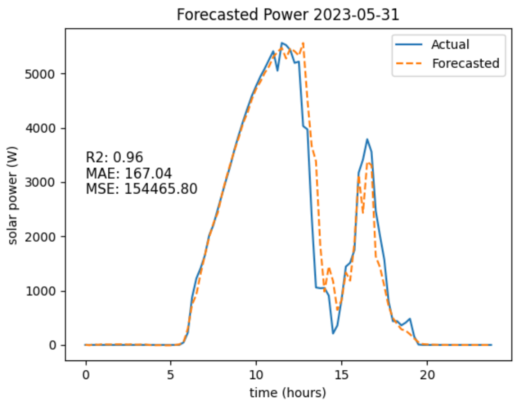

Figure 9. A relatively accurate, complex forecast with a high r2 value.

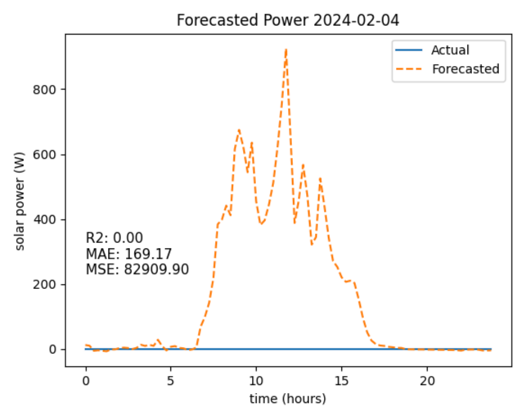

Figure 10. An inaccurate forecast, likely due to snow on the panels with a low r2 value.

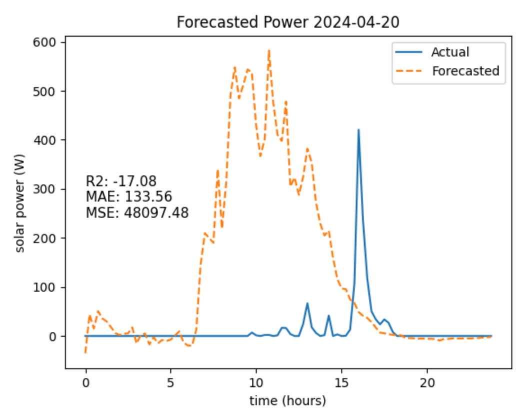

Figure 11. An impossible forecast with a negative r2 value. Negative power generation is not possible.

After training and testing, the model is then applied to NWS weather forecasts. The model performs worse on the forecasted weather than the recorded weather as seen in **Figure 12**. Despite this challenge, sometimes the weather is predictable and stable, which results in good power production forecasts as seen in **Figure 13.** This is likely because the weather forecasts are forecasts and, additionally, do not have as many features as the PWS.

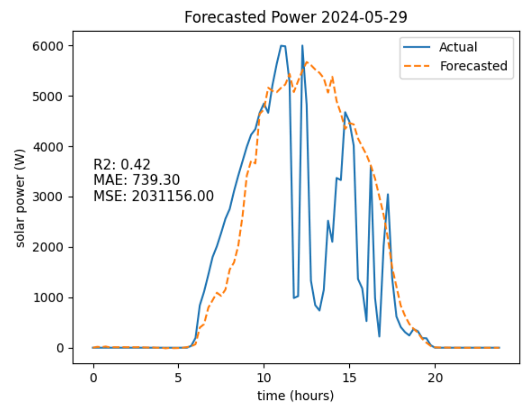

Figure 12. A poor solar power forecast using the NWS linearly interpolated weather forecast for inference.

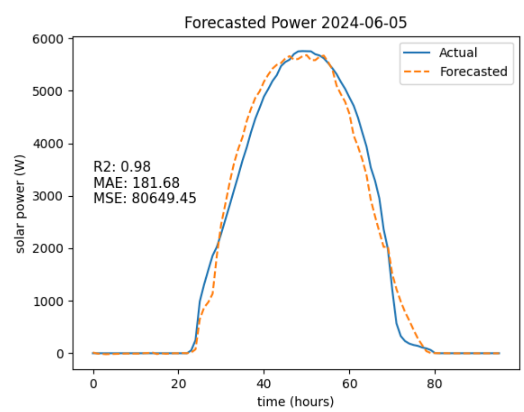

Figure 13. A great solar power forecast using the NWS linearly interpolated weather forecast for instance. The weather and solar production were very predictable on June 5th, 2024.

Setting a loss threshold using a rolling window during training improves the performance of the model. The best improvements are seen when the loss threshold is 0.001. This improves not only the error of the model, but also the time spent training the model. Since the loss levels out at around 100 epochs, training to 300 or 500 epochs over-fits the model (**Figure 14**).

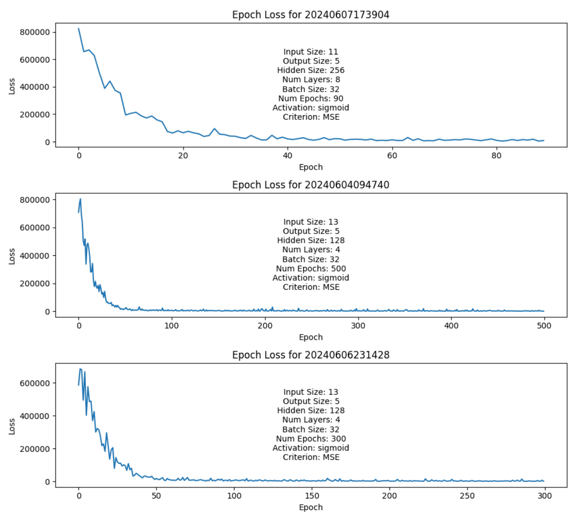

Figure 14. Training loss by epoch for the top three performing models. Notice that the top performing model only trained to 90 epochs with a loss threshold of 0.001.

Future Work

Work can be done to improve the features used and the normalization of those features in both the PWS and NWS forecast data sets. It could prove useful to derive a feature that indicates the probability of snow on the solar panels. Also, this experiment did not perform a thorough outlier analysis, which would be helpful in smoothing out the erratic nature of power and weather. In addition to outlier analysis, it may help to run the data through smoothers, such as a regression algorithm to have more predictable behavior in both weather and solar forecasts.

Additional work can be done to tune the model. Only a handful of hidden layers, layer size, and activation functions were used in this project. This LSTM uses a linear module state (Shannon lstm.py#22). An attempt was made to try a polynomial or sinusoidal module state to reflect the cyclic nature of the data, but nothing substantial came out of the effort. It would be worth exploring different ways to take advantage of the data's cyclic nature within the model itself.

In essence, this project is a starting point for proper solar power forecasting as there are many avenues of data features, forecasts, and models to explore.

Conclusion

Long short-term memory models are a suitable choice to handle the intricacies of weather and solar forecasting. The feature extraction process affects the performance of the model more so than tuning the model - at least at this stage of forecasting. More work can be done to improve the features used in the model, as well as the model itself. Power forecasting is a promising field with much research being put into it. This could prove useful to homeowners and utilities with residential solar panels providing them energy.

Works Cited

Barbose, Galen, et al. “Residential Solar-Adopter Income and Demographic Trends: 2022 Update.” *UC Open Access Publications*, 2022.

Bhagyanidhi, A., et al. “An Approach for Rainfall Prediction Using Long Short Term Memory Neural Network.” *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, 2020, https://ieeexplore-ieee-org.libezproxy2.syr.edu/document/9250809.

Carrera, Berny, et al. “PVHybNet: a hybrid framework for predicting photovoltaic power generation using both weather forecast and observation data.” *IET Renewable Power Generation*, vol. 14, no. 22, 2020, pp. 219-2201, https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-rpg.2018.6174.

de Guia, J., et al. “Using Stacked Long Short Term Memory with Principal Component Analysis for Short Term Prediction of Solar Irradiance based on Weather Patterns.” *2020 IEEE REGION 10 CONFERENCE (TENCON)*, vol. 10.1109/TENCON50793.2020.9293719, 2020, https://ieeexplore.ieee.org/abstract/document/9293719.

Hossain, M., and H. Mahmood. “Short-Term Photovoltaic Power Forecasting Using an LSTM Neural Network and Synthetic Weather Forecast.” *IEEE Access*, vol. 8, 2020. *https://ieeexplore.ieee.org/abstract/document/9200614*, 10.1109/ACCESS.2020.3024901.

Kim, Seul-Gi, et al. “A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning.” *MDPI Open Access Journals*, vol. 11, no. 5, 2019, https://www.mdpi.com/2071-1050/11/5/1501.

“PWS Network Overview.” *Weather Underground*, https://www.wunderground.com/pws/overview. Accessed 6 May 2024.

Qin, Chuan, et al. “Long short-term memory with activation on gradient.” *ScienceDirect*, vol. 164, 2023, https://www-sciencedirect-com.libezproxy2.syr.edu/science/article/pii/S0893608023002125.

Razavi, S., et al. “From Load to Net Energy Forecasting: Short-Term Residential Forecasting for the Blend of Load and PV Behind the Meter.” *IEEE Access*, vol. 8, 2021, https://ieeexplore.ieee.org/abstract/document/9292948.

Shannon, Daniel. “photovoltaic-weather-station.” *GitHub*, https://github.com/radioxeth/photovoltaic-weather-station/tree/main.