Solar Panel Time Series Analysis Report

Caleb Dayley and Josh Christensen

Abstract

In this study, we try to use data from a single solar panel over 3 years to find what effects time has on the power generation of solar panels in generating electricity and how effectively we can predict that power generation in the future. We approach this problem from a time series perspective because we would expect the power generation of the solar panel to be highly correlated with itself in days just before and after each measurement. We also find that correlation between measurements taken around a year apart is also significant. We fit an AR-1 model with time and a 3 level fourier curve as covariates in order to account for the correlation between days and years. We find that time does have a negative impact on the effectiveness of solar panels indicating that solar panels do tend to wear over time.

Introduction

The prevalence of solar power has grown over the years. Solar power provides a more environmentally friendly option in comparison to fossil fuels. However, the amount of energy produced by a solar panel can vary for multiple reasons including weather, panel angle, direction of the sun, cloud cover, panel cleanliness, etc. In this particular study, we want to know the effect that time has on the power generation of solar panels to see how fast solar panels degrade. We have a dataset that has the daily amount of kWh (kilowatt hours) produced by a single solar panel each day for 3 years. In this study, we try to answer the following questions:

- How do solar panels degrade over time?
- How many years till the panels have lost 50
- Can we obtain projections of power for the next year?

Exploratory Data Analysis

By looking at a plot of power generated by our solar panel over time, we can see there are some obvious patterns. First of all, we can see that our panel produces more energy on average during the summer months. Second, we can see that the average power output does appear to decrease each year. We can evaluate whether these patterns are significant by fitting an appropriate model to the data.

Intuition tells us that we would expect the amount of power generated each day to be highly correlated with the previous day because of the expectation that the weather, cloud cover, direction of the sun, and other variables of a given day tend to be similar to the days just before it. An autocorrelation plot of our data confirms this thought by showing our measured values are highly correlated with one another. If we don't account for this correlation, the estimated accuracy of our model's expected effect of time on power produced will most likely be incorrect.

Model Exploration

Different types of models were considered for this analysis. However, with a data set that only includes time as an explanatory variable, models such as tree ensemble models or neural networks are not viable options. We also considered a simple linear regression model using time as a covariate but because of the non-linear relationship between power production and time and the dependence of measurements on one another, this model would not adequately fit our data. In order to fit our model, we decided to try fitting

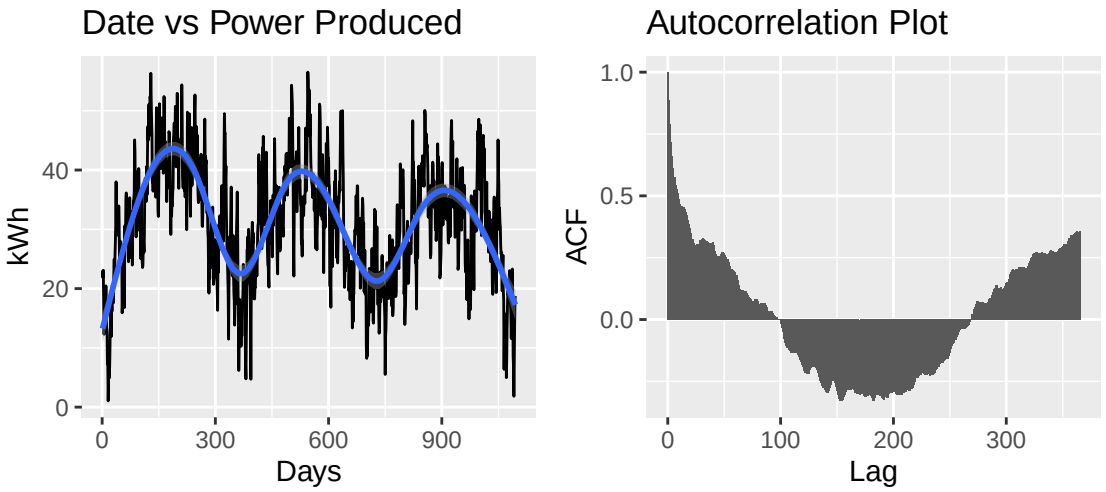

Figure 1: EDA plots

different time series models in order to capture the autocorrelation as well as the non-linear relationship in our data set.

We considered three different time series models for our analysis. Our first model was a ARIMA(2,0,1) with time as a covariate. The second was an AR(1) model with time and a one level fourier curve as covariates while the third was an AR(1) model with time and a three level fourier curve as covariates. We used root prediction mean square error (RPMSE) in order to compare the predictive capabilities of each model to each other as well as to the standard deviation of our time series. Below is a table summarizing the cross-validated results.

Table 1: Model comparison

Model	RPMSE
Mean	10.126
ARIMA(2,0,1)	12.323
AR(1) w/ 1 level fourier	7.763
AR(1) w/ 3 level fourier	7.582

We found that our AR(1) model with a three level fourier curve predicted out of sample better than any other model and also predicted better than using the mean of all our observations. Based on this finding, we will continue our analysis using this model.

Model Specification:

$$y_t = X_t\beta + \epsilon_t$$
$$\epsilon_t = \phi\epsilon_{t-1} + \omega_t$$
$$\omega \sim N(0, \sigma^2)$$

- Our β vector holds the intercept and all the average estimated effects of time, our sine curves, and our cosine curves holding everything else constant. The intercept represents our estimation of the solar power output on the first day of observed data begins.
- Our X_t matrix contains a 1 to include the intercept as well as the time, cosine terms, and sine terms from our fourier curves such that each β term matches with its covariate of any given t observation.
- ϵ represents our correlated error term which follows an AR(1) model, this indicates that each error is updated based on the correlated error of the previous measurement

- ω represents our uncorrelated errors which we assume follow a normal distribution with 0 as a mean and σ^2 as the variance.

With our chosen model, we are assuming that our uncorrelated ω errors are independent and follow a normal distribution with equal variance. We also assume a linear framework for our model. We will visit the validity of these assumptions later on in the study.

Model Justification and Performance Evaluation

Variable selection was straightforward in this analysis due to the low number of available covariates. We included time and a fourier approximation of the data in our mean process. We included the time variable to have an estimate of the average effect over time after accounting for temporal correlation and seasonality. We included the six fourier terms (third order approximation) to account for seasonality in our data. We made our approximation third order because it performed better in-sample and out-of-sample compared to the first and second order approximations.

We included one autoregressive term in our correlation function. We included this variable and none others because this was the correlation structure that resulted in the lowest AIC. The positive ϕ indicates that each observation has positive correlation with the previous observation.

The assumptions for our model include that our ω errors are independent and identically normally distributed with mean 0 and variance σ^2, and model linearity (not necessarily linear relationships). We check normality of the data by decorrelating residuals and checking for normality. We check constant variance and linearity by plotting decorrelated residuals against fitted values and checking that vertical spread is consistent across the horizontal axis and there are no remaining curved trends. We also check that we have eliminated correlation in our residuals by displaying an acf plot that shows we have captured the majority of the correlation. All plots mentioned are included in Figure 2.

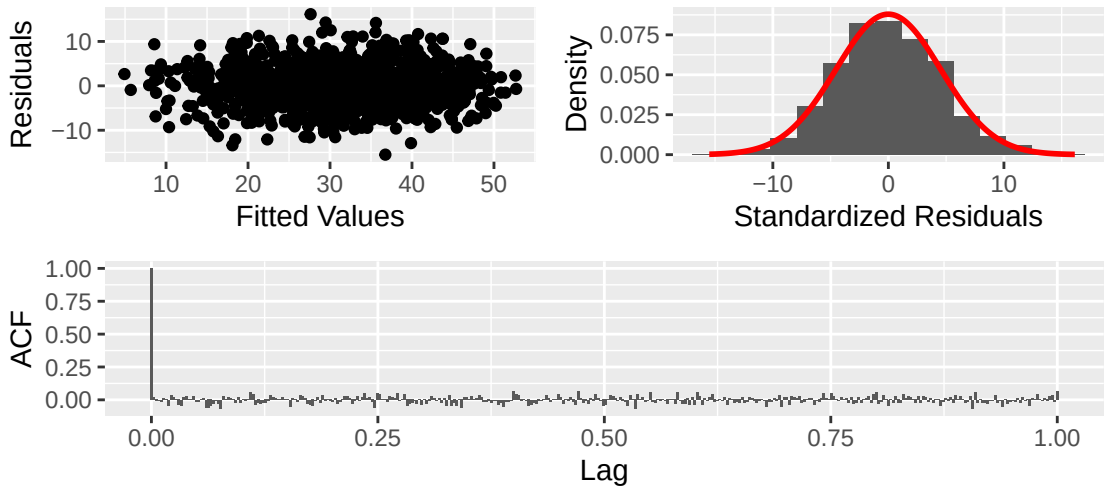

Figure 2: Evaluation of assumptions

To evaluate model fit we measured in-sample root mean square error (RMSE) and pseudo-R^2. Our RMSE of 4.54 compared to the standard deviation of the data, 10.13, indicates that our model fits the data quite well. This is confirmed by our pseudo-R^2 of 0.799.

One of the primary goals of this analysis was to be able to predict future output from our solar panels. We performed cross-validation on our analysis by taking training sets that were a year and a half in size and predicting the coming year. We did this 180 times by shifting both training and test sets by one day each time and then averaged our RPMSE, bias, and out-of-sample pseudo-R^2 across the 180 iterations. The

results are summarized in Table 2. The RPMSE indicates how far off our predictions are on average from the truth. Our bias indicates whether we consistently under or overestimate and by how much. Pseudo-R^2 gives a rough approximation of the percentage of variability we are capturing. All metrics indicate fairly robust prediction for a time-series forecast this far into the future.

Table 2: Prediction Evaluation

	Estimates
RPMSE	7.642
Bias	-0.674
Pseudo-R^2	0.351

Results

Now that we have decided on our model we can use the parameter estimates from our model to address our questions of interest. The parameter estimates from our model are summarized in Table 3 below.

Table 3: Parameter Estimates with Intervals

	Estimate	Lower	Upper
ϕ	0.777	0.739	0.814
β_0	33.982	31.512	36.452
Time (days)	-0.005	-0.009	-0.001
Fourier $\sin_1$	-0.264	-2.020	1.493
Fourier $\cos_1$	-9.356	-11.042	-7.670
Fourier $\sin_2$	-1.330	-3.030	0.369
Fourier $\cos_2$	-2.397	-4.072	-0.723
Fourier $\sin_3$	0.047	-1.626	1.719
Fourier $\cos_3$	-1.026	-2.682	0.630

Our model also estimated σ^2 to be 20.74. In the context of our problem, ϕ represents the parameter for the AR(1) correlation model that we used to capture temporal correlation. The β_0 parameter represents the intercept for all observations. The time coefficient represents how much kWh generation degrades on average per day after accounting for the other variables in our model. The Fourier coefficients represent how much the estimate for kWh changes per day according to each Fourier component.

We now address the questions posed at the beginning of this analysis. We first recognize that our time coefficient indicates how our solar panels are degrading over time after accounting for seasonal trends and temporal correlation. Our coefficient value of -0.005 indicates that our solar panel generates 0.005 fewer kWh per day on average for every day elapsed or about 1.818 fewer kWh per day on average for every year elapsed.

We answer the question of how long it takes a solar panel to lose 50% of its power generating ability by taking β_0 and seeing how long it takes the coefficient for time to reduce that number by 50%. β_0 is the starting point after accounting for all other trends we've modeled including seasonality, degradation and correlation. Using this method we estimate that it will take 9.34 years on average for the solar panel to lose 50% of its power generating ability.

The final goal of our analysis was to forecast a year of power generation into the future. Figure 3 displays predictions for the next year along with 95% prediction bands.

Conclusions

The main goals of this study were to learn about solar panel degradation, to determine how long it takes a panel to lose 50% of its power generating ability and to project power generation a year into the

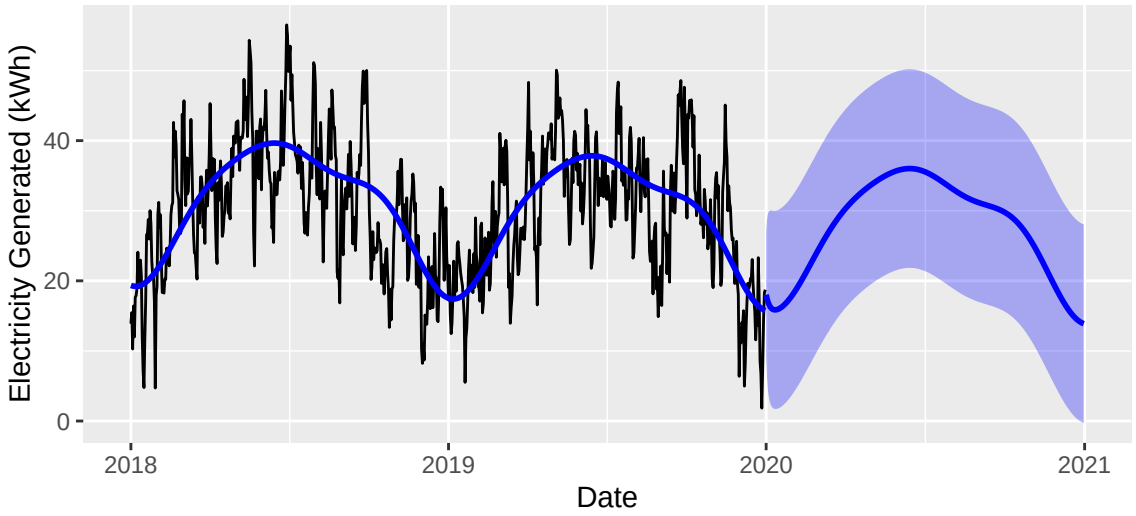

Figure 3: Power projections for the coming year

future. Each of these goals was specifically addressed in the results section and we have justified why it is appropriate to use our model to answer these questions.

One shortcoming of this study is that we did not explore other ways to account for seasonal trends outside of a Fourier approximation and traditional seasonal autoregressive or moving average components. While the Fourier terms performed quite well, there may be better ways to account for the seasonal trend. Another shortcoming is related to the data. As always, more data is better, but that is particularly obvious when we are only dealing with a single solar panel for a single household. Generalizing decisions from this single panel is not advisable, and it would be much preferred to obtain data from other solar panels and perform a longitudinal analysis.

Going forward, we suggest that new data be requested on other panels as a first measure. If new data cannot be obtained we would also suggest looking into other methods of dealing with the seasonal trend in our time series. Another option for next steps would be to collect data on other covariates that may affect energy generation such as weather, panel angle, panel cleanliness, etc.

Teamwork

We both worked on every aspect of the analysis. We both evaluated different models. We both worked on the write up with Caleb taking charge of the first two sections and Josh leading on the last three. We both created graphics for the final paper. We both reviewed all sections of the final paper.