Solar Power

Max Smith and Abe Durrant

10/25/2021

Abstract:

Here we look at the daily power generation in kWh of a single solar panel over 3 years. We investigate the decrease in time of power production of solar panels by looking at this data. We fit the data to an ARIMA model and determined that there is a decrease of about 1.971 on average from year to year in the production of power from the solar panels. We also calculated that it would take just over 8 years for our solar panel to reach 50% power generating capacity if the trend continues. Finally, we predicted the power generation for the next 365 days for this solar panel and determined that our model predicts power fairly well. It gives us more information than just looking at the mean.

Introduction

Solar panels are increasing in popularity as an environmentally friendly alternative to fossil fuels and a way to cut down on ones electricity bill. Understanding how solar panels degrade over time is important in understanding their sustainability and effectiveness. The amount of solar energy generated depends on several variables such as direction the panel faces, amount of cloud cover and other factors. However, we are looking at how a single solar panel degrades over time by looking at the amount of energy it produces per day over a period of 3 years. We have been tasked with looking into how the panels degrade over time, how many years til the panels have lost 50% of their power generating capabilities on average, and a forecast for daily power generation over the following year.

Exploratory Data Analysis

As we began looking at the data, we plotted it with a linear model. This is shown below in figure one. Initially we see that a negative trend is shown and the solar panel does appear to be slowly losing its power production capabilities. We also see that there is a strong seasonal component showing that over a year, we have times where we can expect to see more or less power generation. Since the data is collected over time it is correlated with one another. The power generation from one day is highly correlated with the data previous to it. This is shown in Figure 2 below. This plot shows the autocorrelation from one day to the next day and the other days following. We see that power generation is highly correlated with the day after and that continues to taper. We must choose a model that can account for this such as a time series. If we don't do this we will miss a lot of the predictive power and add uncertainty due to this this correlation to answer the questions at hand.

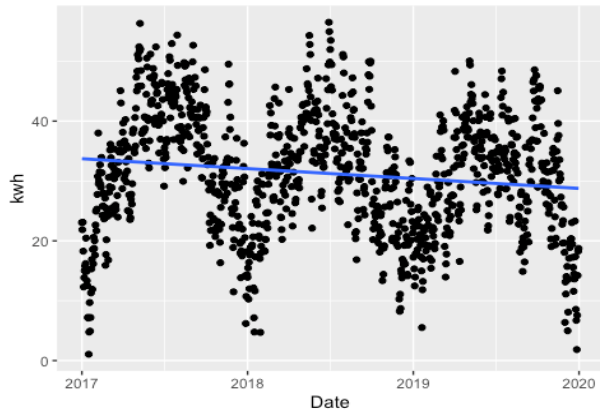


Figure 1

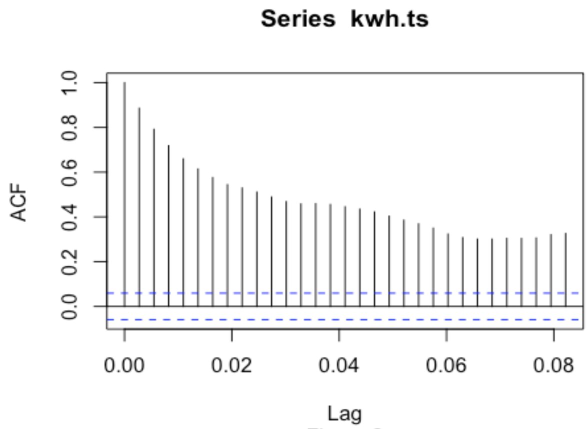


Figure 2

Modeling and Methods

To answer our questions of interest, we explored our data and determined that in order to best answer the questions and fit the data we needed to use a time series model that accounted for the correlation. First we fit an ARIMA model to the data. We included the yearly seasonality that we found. Below is an ARIMA model to explain what is happening.

$$Model: Y_t = Y_{t-365} + \delta + \phi(Y_{t-1} - Y_{t-366}) + w_t,$$

$$Where: w_t \sim N(0, \sigma^2)$$

Y_t and Y_{t-365} represent the power generated in the current period and the period one year previously.

δ is the drift which is the change from the previous year. The parameter, ϕ, expresses what parts of the values (Y_{t-1} and Y_{t-1}) from the last periods are relevant in estimating the current one. Finally, we have w_t which is the error term which is normally distributed with mean 0 and variance σ^2.

Strengths of this model are higher predictive power and less uncertainty if we do indeed have correlated data. When the data is correlated and we use a standard regression model we will see unbiased estimates but our predictions and intervals will be wrong. Weaknesses of this model are that it is computationally heavy. While it works on a dataset of this size it does take a long time. On a larger dataset it would be very computationally expensive. Another weakness is that it can over fit the data easily.

To use this model we have certain assumptions we have to meet. First we need to see normality in our errors. When we plot our errors we want to see them looking approximately normal. We also need a constant variance in our data. This means that our variance isn't changing throughout our data. Next, we assume that our data is linear. Due to the way that we set up our model we do assume linearity from year to year and we need to check that assumption. Finally, we are assuming that our data is not independent. We are also assuming that we are accounting for all dependence through our model.

Model Justification and Performance Evaluation

As our data comes with the single explanatory variable of Date, we are only using that in our time series model. We fit a Sarima model with parameters p = 1, D = 1, and S = 365. These parameters were chosen through the use of the "auto.arima" function in R. This function returns the best model according to an AIC value. Our value of p represents that our model is an AR model order 1, while D represents the differencing based on the seasonal trend (set at S = 365)

Our model runs under several assumptions which we can quickly check. The first is linearity. We are assuming our data is a linear combination of previous data points. We have seen no trends that lead us to believe this is violated such as an exponential decrease in kWh over years. We are also assuming our data is multivariate normal which we check by looking at the histogram of residuals which is shown in Figure 3. There is no cause for concern looking at this plot and the residuals appear to be normally distributed. We are also assuming constant variance which we can check by looking at a plot of the standardized residuals vs the fitted values shown in Figure 4. Here we don't see any changing variance across the data so we can continue with this assumption. Finally, while we are assuming dependence, we assume that our model accounts for this and that our residuals are no longer correlated. A look at the ACF plot of Residuals shows us that we have indeed accounted for the correlation across most of our data. There is a small spike at 365 days, so our differencing in our model isn't fully accounting for the correlation across years, however it has reduced it to be very small.

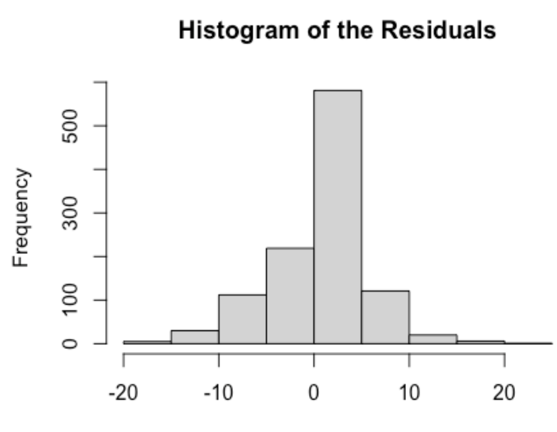


Figure 3

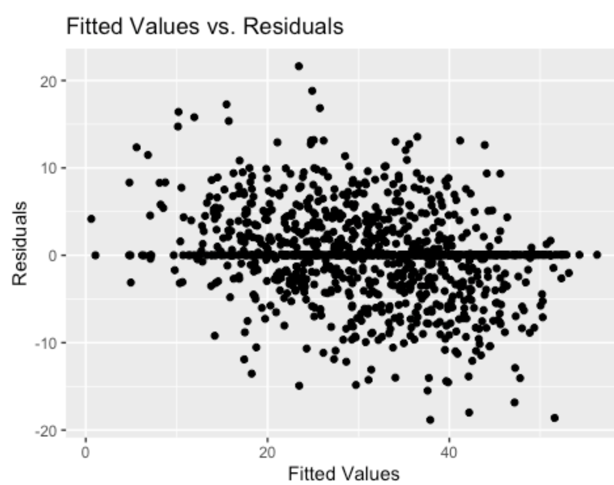
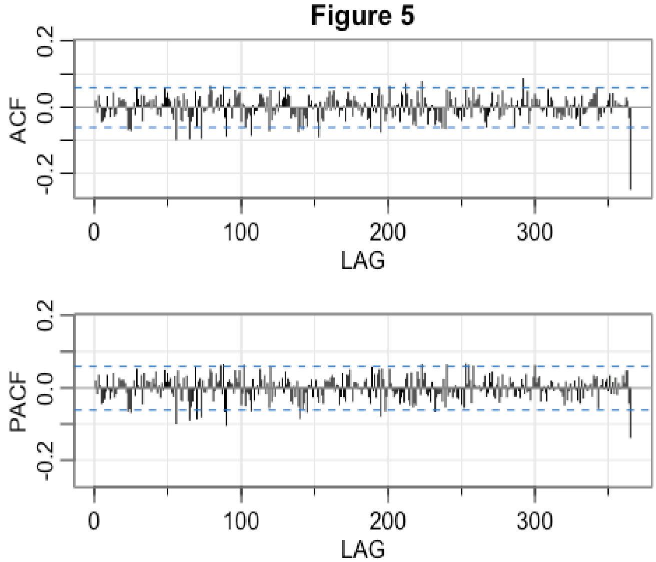


Figure 4

Figure 5



Our model fits the data well overall. The RMSE of our model on the data it was trained on is 5.013 and our MAE is 3.29. This means that we are off by about 3.29 on average from our fitted values vs the actual data points. We also calculated a Pseudo R-squared value of 0.77. This is a way to estimate how well our model does at explaining the variation in kWh. 0.77 means that our model is doing a good job overall at explaining the variation in kWh.

We ran a cross-validation study where we took different subsets at the end of our data and used the rest of the data to predict them. We predicted 12 different subsets of the data starting with a full year in our test set and then 30 days less in each iteration. From this cross-validation study, we returned a RPMSE of 9.1 and a Bias of -0.467. This means our predictions were off by approximately 9.1 on average across all the test sets. While this is lower than the standard deviation of kWh, it is much higher than the RMSE from the full model. This is likely due to the fact that without 3 full years of data it is hard to account for the difference over years. Our bias of -0.467 means that our predictions were slightly high on average.

Results

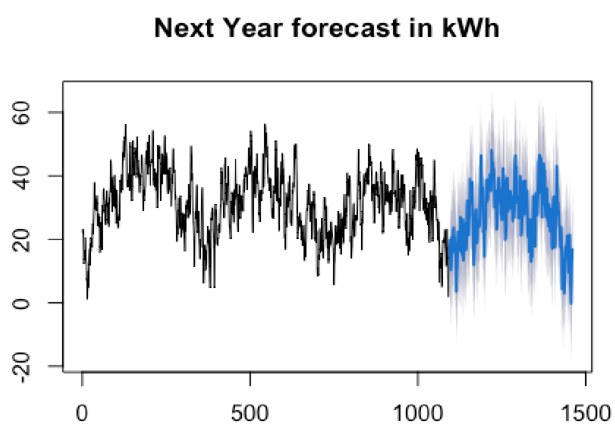
Our models estimated parameters were δ and ϕ. δ is the drift which was estimated to be -0.0054, with a standard error of 0.0025, which is the change of power production by day from the previous year. The parameter ϕ expresses what parts of the values in kWh from the last periods are relevant in estimating the current one. This value was estimated to be 0.7544 with standard error 0.0242.

Finding how the panels degrade over time can be difficult since the amount of solar energy is different across the seasons of the year. To account for this, we decided to look at a few summary statistics of each year in terms of solar energy in order to compare the energy over time (year to year). We looked at the mean and median as well as the first and third quartiles. Below is a table including these statistics for the 3 years of observed data. We can see that there is a downward trend of about 2 kilowatt hours per year across each of these metrics. Our predictions for 2021 also show this with a predicted mean of 27.40, median of 27.77, 25th quantile of 20.29, and 75th quantile of 34.64. Additionally, we can look at the drift in our model summary which is calculated to be -0.0054. This means that our model is accounting for a decrease in power generation of 0.0054 each day of the year. This tells us that our model is predicting that the yearly decrease in kWh generated is about 1.971 kWh.

year	mean	max	min	median	90th quantile	75th quantile	25th quantile	10th quantile
2017	33.27699	56.32	1.06	33.94	46.252	41.03	25.58	18.632
2018	31.07370	56.51	4.71	31.34	43.278	37.67	24.08	18.812
2019	29.37419	50.05	1.84	29.74	41.302	36.61	22.26	17.292

If current trends continue then we estimate it will be approximately 8.4 years until the panels have lost 50% power. This is assuming a yearly decrease on average of 1.971. However, this estimate is only based on the current trend and would not account for any sort of exponential change over years (hard to identify with only 3 years of data). We do not know if after a certain period of time whether or not the panels will start degrading faster, or perhaps level out at a certain level.

Our projections for the next 365 days can be seen in the following graph along with the estimated errors. Summary statistics are included in the table discussed above.



Overall, these solar panels do appear to be degrading steadily over time. There is a seasonal trend each year and so we primarily looked at yearly data in order to see how the panels are degrading. Our model has given predictions for the next year that show the same trend upward in the summer months and an overall drop in energy level for the year. After 8.4 years we predict that the panels will have lost about 50% of their power.

Conclusion

We met the goals of the study as we were able to answer the original analysis questions that we had. We found an approximation for the yearly decrease in kWh, and we were able to project future solar panel performance. Our model also shows the seasonal trend there is with these particular solar panels.

Our model did have shortcomings in its estimated predictive power and also may not account for possible changes in how much kWh is decreasing over time. While it was predicting better than the mean, it still had large errors at times. The next steps in analyzing the energy power of these solar panels could include collecting data over a longer period of time (to detect any long-term trends) as well as looking at other variables that may affect their performance (temperatures, amount of sunlight, etc.). This could come from getting data from more than a single set of solar panels. Panels may degrade differently in different conditions and our model can’t account for that.

Teamwork

Max did abstract, Introduction, and Model and Methods Abe did Model Justification and Evaluation, Results, and Conclusion We both looked over each others work and did EDA