IE360
TERM PROJECT
FORECASTING SOLAR ENERGY PRODUCTION

Ertuğrul UYAR – 2018402063
Gizem GÖKTEN – 2017402015
Selin GÜL – 2018402216

CONTENT:

1- INTRODUCTION

1.1 Problem Description

In this report, the aim is to analyze the given data in a statistical perspective with the help of R. Then to implement regression and time series analysis for forecasting solar energy production. and examine the regression model in terms of goodness of fit and residual standard error whereas time series model in terms of the error metrics like AICc and BIC.
During the 'Competition Phase' we are expected to forecast hourly solar energy production of KIVANC 2 GES (Güneş Enerjisi Santrali). The given data includes weather measurements for nine locations nearby the solar power plant and it's in hourly form from February 1st 2021 to the last day of updated data. For each day, we are provided production data that consists of hourly production values and weather data that consists of hourly values at the given coordinates (latitude and longitude) with some variables. The data contains 4 weather variables; DSWRF, REL_HUMIDITY, CLOUD_LOW_LAYER, TEMP and each will be explained briefly;

TEMP : Temperature at the provided location. We may expect to see seasonality in the temperature data. And although it seems that there can be positive correlation between temperature and solar power production, it's important to mention that high temperatures can decrease the efficiency of panels.
REL_HUMIDITY : Relative humidity at the provided location which "indicates a present state of absolute humidity relative to a maximum humidity given the same temperature."[1]
DSWRF : The downward shortwave radiation flux. Forecasting of DSWRF is used for "building energy usage modeling and optimization"[2].
CLOUD_LOW_LAYER : This is total cloud cover data (in terms of percentage) for low-level types of clouds.

1.2 Summary of the Proposed Approach

We began this report with constructing a descriptive analysis, which will be covered below. After examining the data in a broad perspective, we checked for the stationary of the data and tried to discover attributes that make it nonstationary. Then we see that data has trend and seasonality patterns so lag1 value is determined to be added to the model for representing the trend and lag24 value is determined to be added for representing daily seasonality, these steps aimed to make data stationary. Then ggpairs plot was created to analyze the correlations between regressors and the regression model was constructed with these knowledge.

Basically, the same steps were implemented for the ARIMA and ARIMAX models; examined data was taken into consideration, correlation between regressors were analyzed and the models were constructed in line with their statistically reasonable processes, which will be explained in later parts.

1.3 Descriptive Analysis of the Given Data

Plotting the hourly solar energy production:

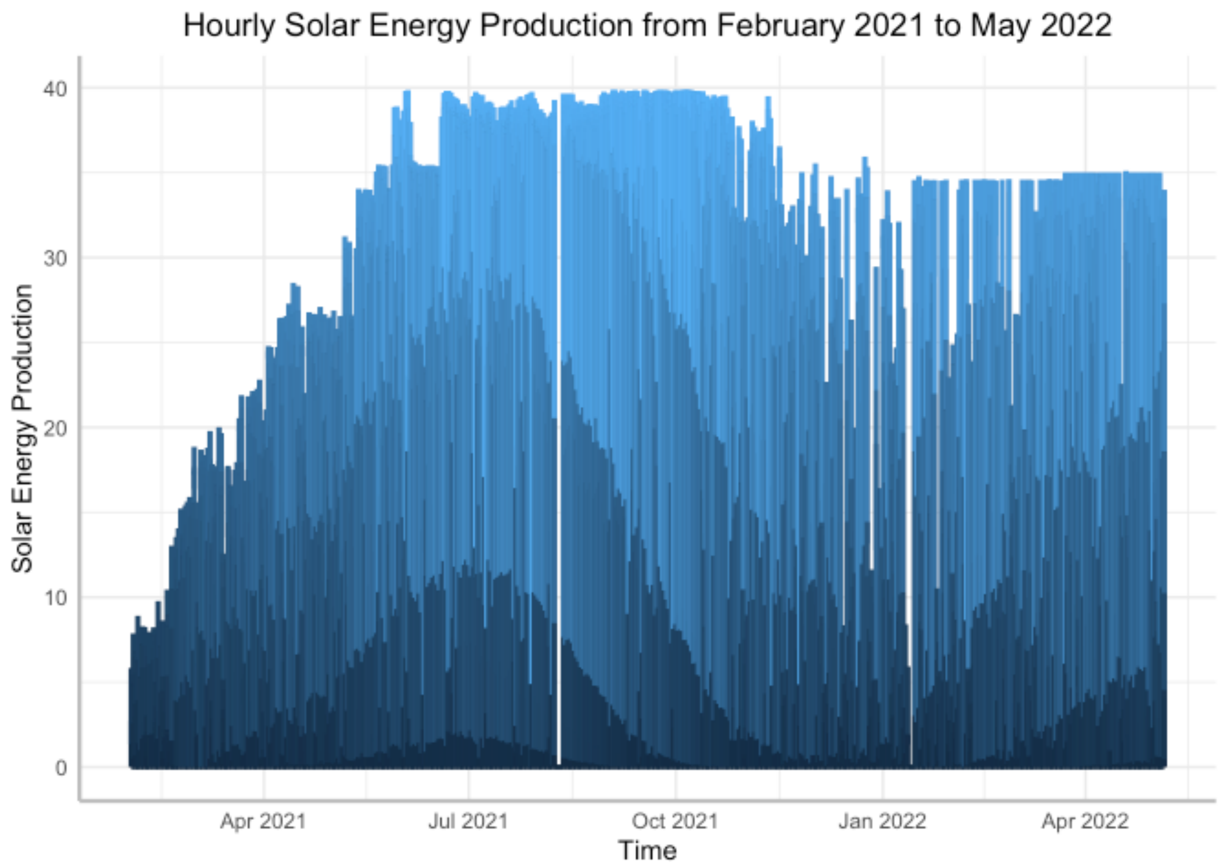


Figure 1: Hourly Solar Energy Production

This plot is not very easy to interpret and cannot give insights at first sight. Therefore, rather than examining hourly series to find patterns, we are to evalute the daily time series of the solar production.

Aggregating the hourly data into daily data and plotting the daily time series of solar energy production:

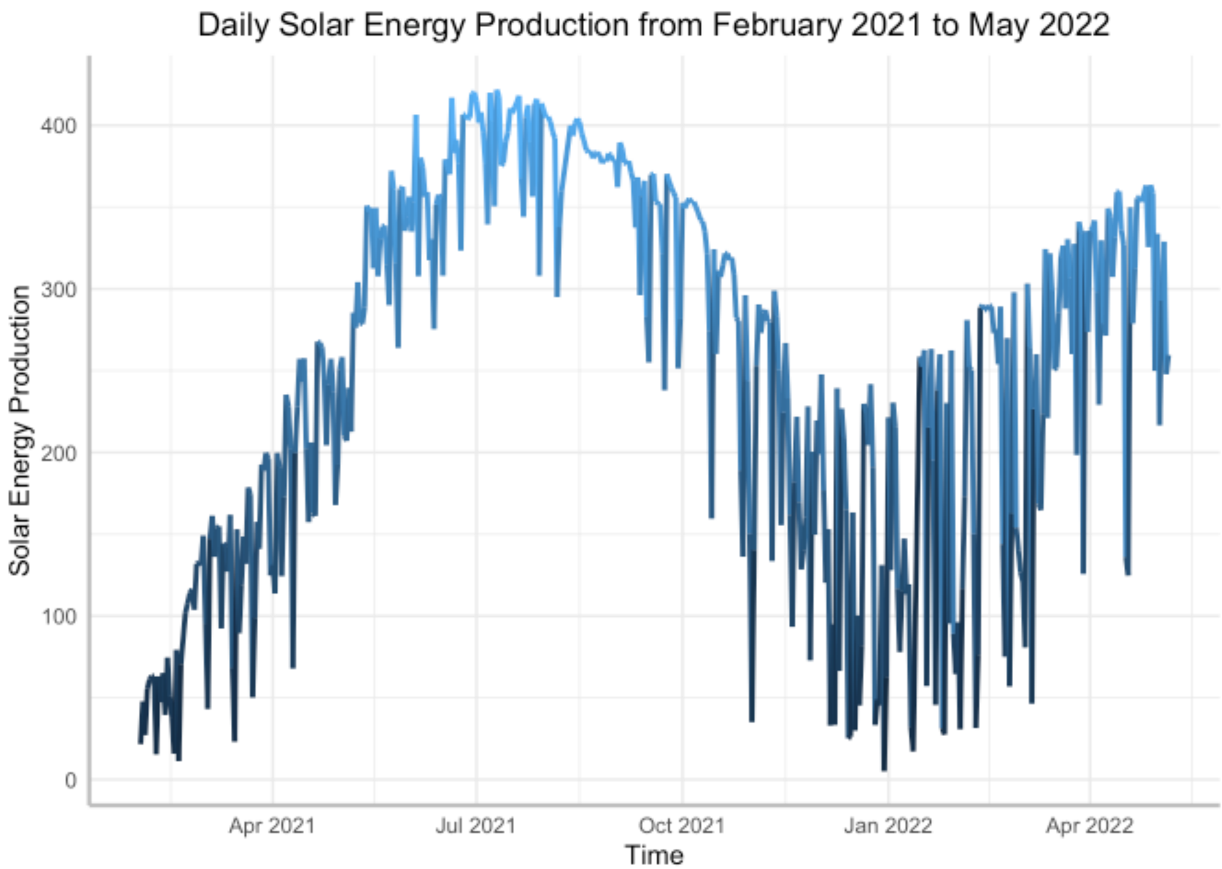


Figure 2: Daily Solar Energy Production

We can deduce from this plot that the solar energy production increases in summer and decreases in winter. We cannot conclude that the seasonality is stable across all years since we do not have enough data to reach this claim. Another observation is that the variance in winter months are higher than during the summer, which is somehow reasonable since in summer, the amount of sunshine is more consistent than it is in winter.

Aggregating hourly times series data into weekly data and plotting the weekly solar energy production to check if we catch a more interpretible pattern:

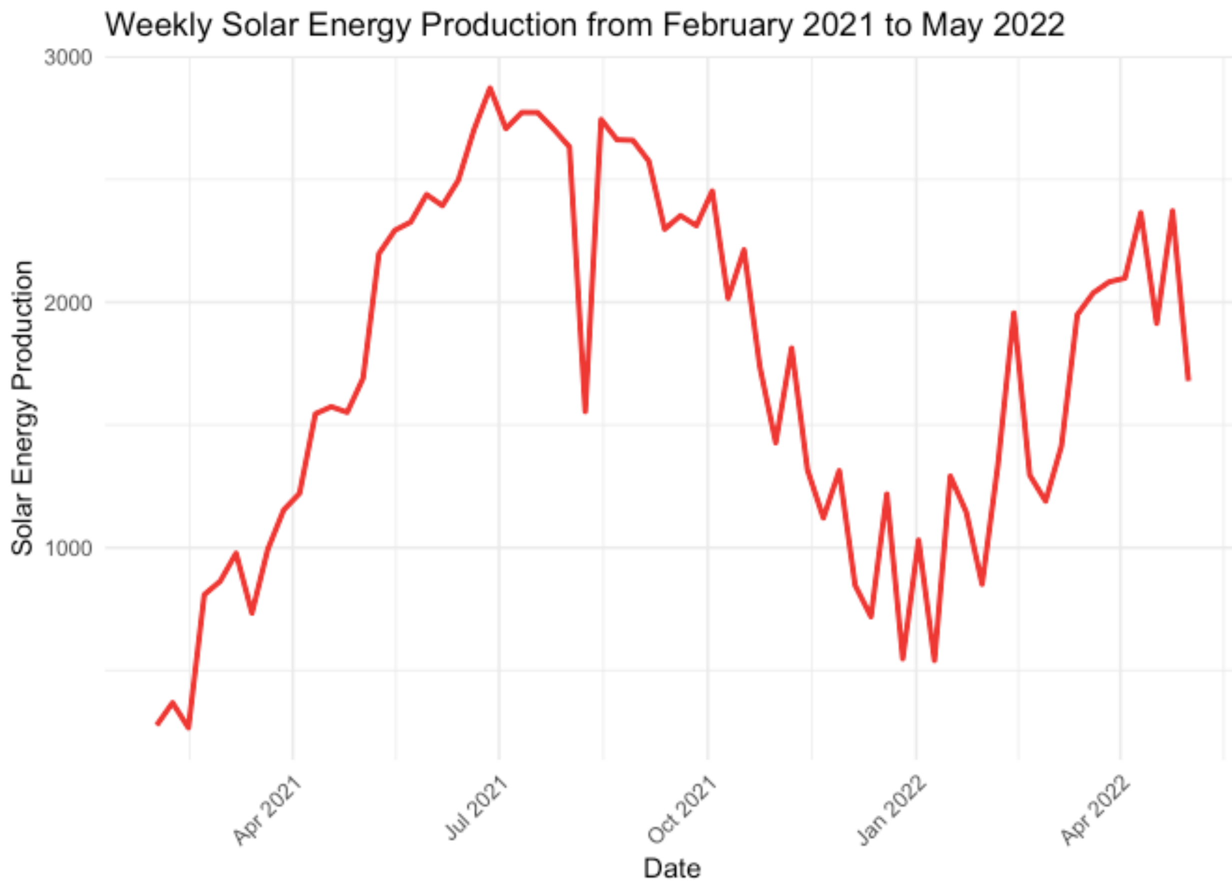


Figure 3: Weekly Solar Energy Production

We can see the observations that we deduce from daily data more clear in this weekly time series. As we are interested to predict in may and june, we can have a first impression of getting higher production rates as the time passes.

Autocorrelation function to check if there is an existing pattern to be extracted:

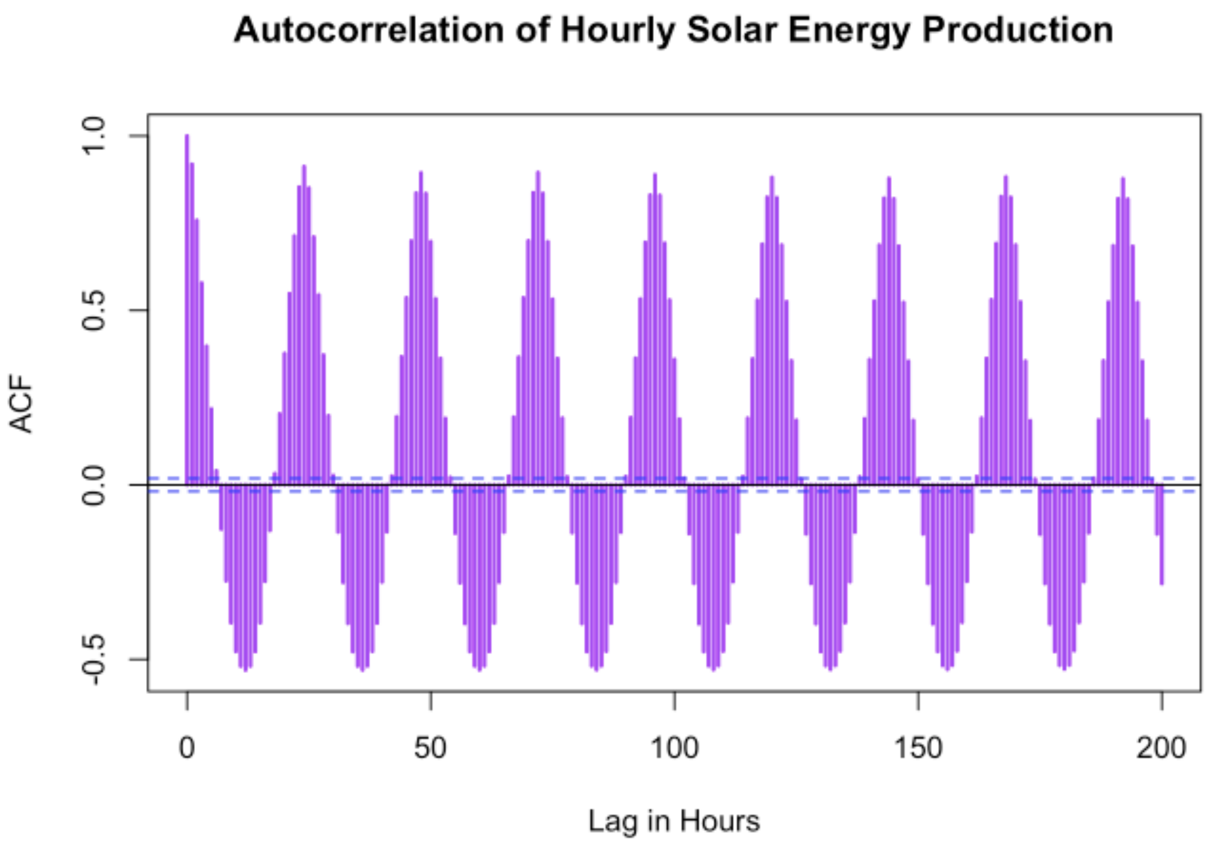


Figure 4: Autocorrelation of Hourly Solar Energy Production

We can see from the autocorelation function that the solar energy production data shows a seasonal pattern. Since it is the hourly data and the data shows a pattern that repeats itself every 24 hours, it can obviously be said that there is a daily seasonality, that is, the pattern repeats itself in a daily manner, consecutive hours shows similar behaviour and same hours of the days reveal a similar production value.

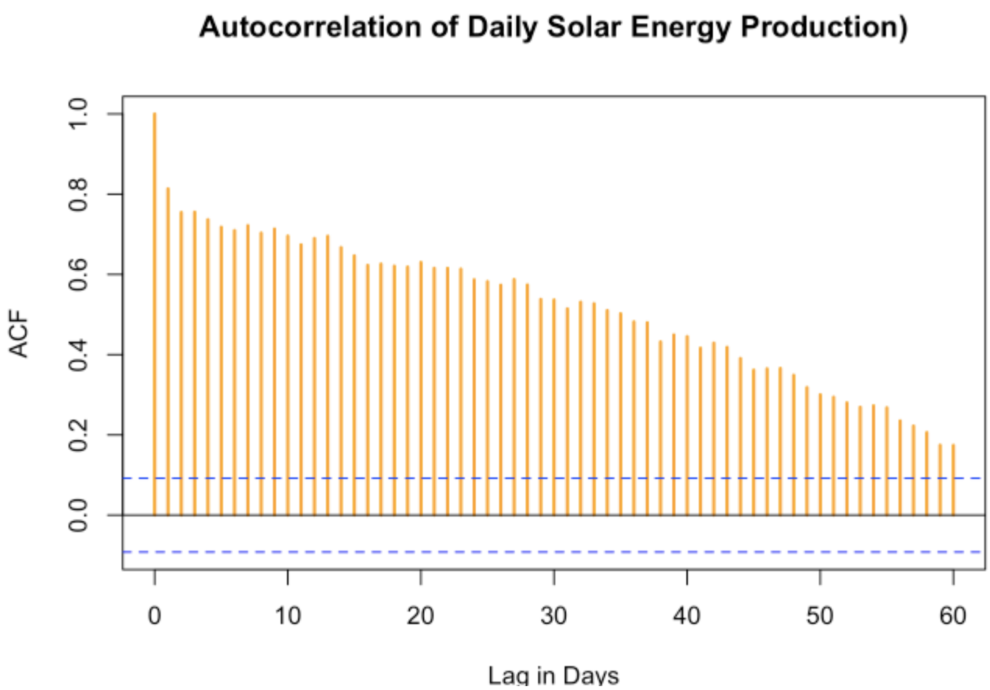


Figure 5: Autocorrelation of Daily Solar Energy Production

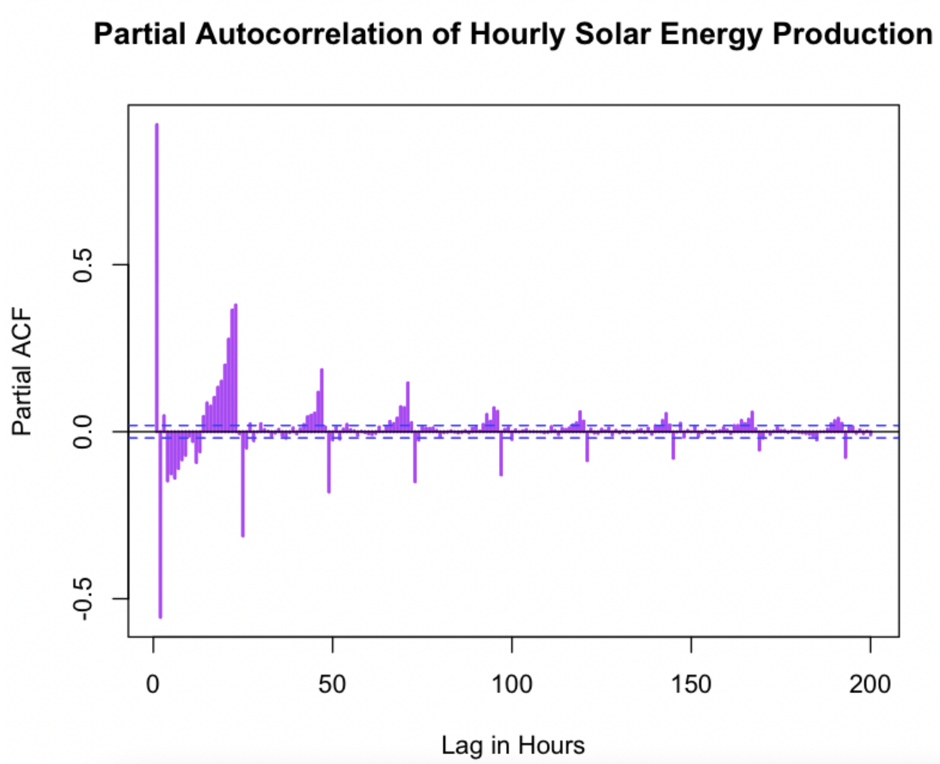


Figure 6: Partial Autocorrelation of Hourly Solar Energy Production

The partial autocorrelation function reveals that there are a lot of significant spikes but the most important ones are Lag1 and Lag24. We can suggest from Lag1 that there is a trend component in this time series and we can deduce from Lag24, 48, 72… that there is a daily seasonality, as we deduced from previous plots.

Partial autocorrelation function of daily time series data of solar energy production:

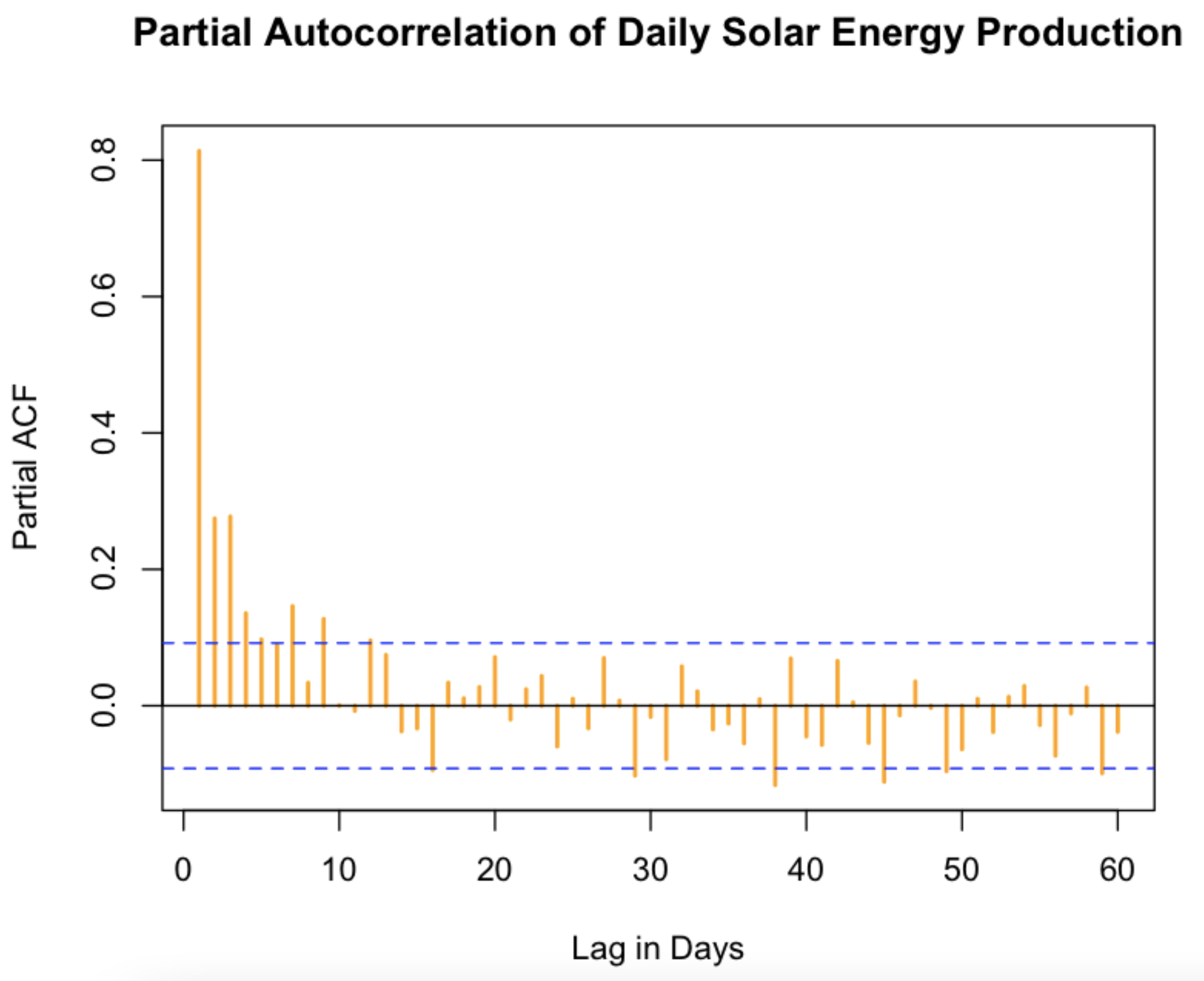


Figure 7: Partial Autocorrelation of Daily Solar Energy Production

This plot reveals nothing new that can contribute to our model selection process. The most significant spike is at lag1 although there are some spikes at lag2 an lag3, which may be helpful, therefore important to keep in mind during the model building process.

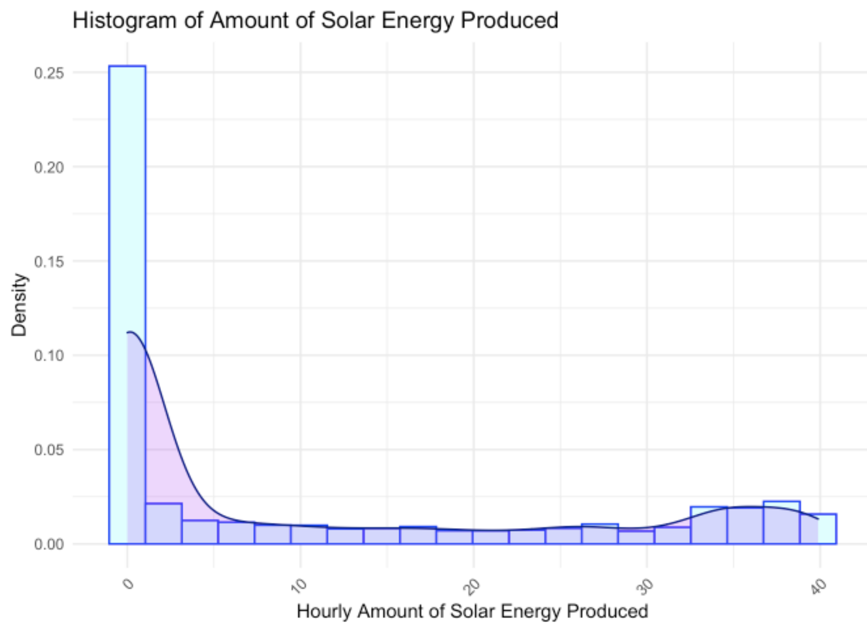


Figure 8: Histogram of Solar Energy Production by hours

This histogram revels that, after exluding zeros from the interpretation since there are a lot of dark hours for every day, the rest approximately follows a uniform distribution, which means violating the normality assumption.

In order to check the stationarity and transformation requirement of the data, the roll-mean and roll-var plots are given:

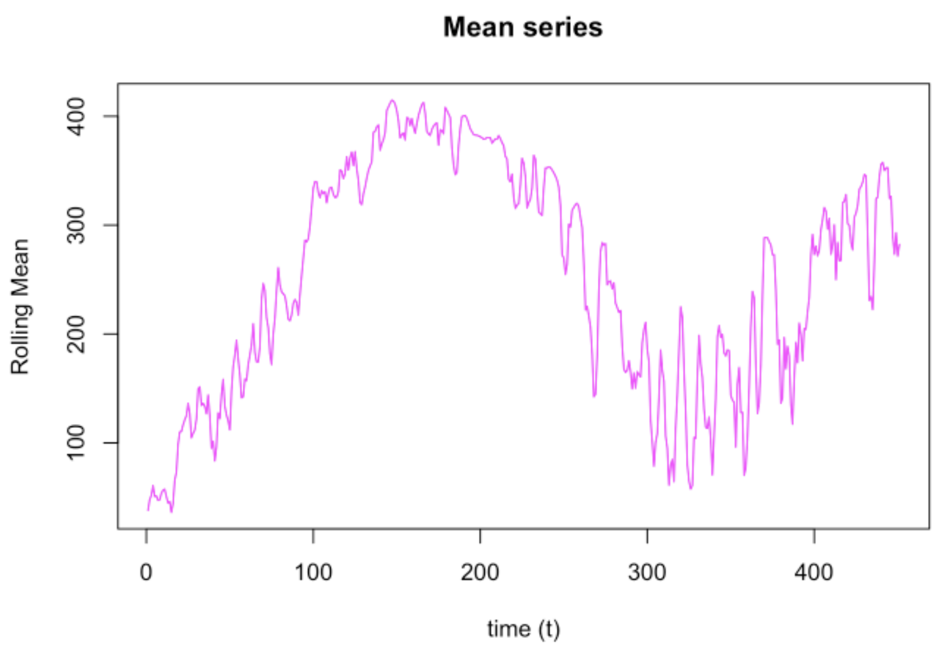


Figure 9: Mean Series of Solar Energy Production

We can see that there is a seasonal pattern in the data, which means that the data is not stationary.

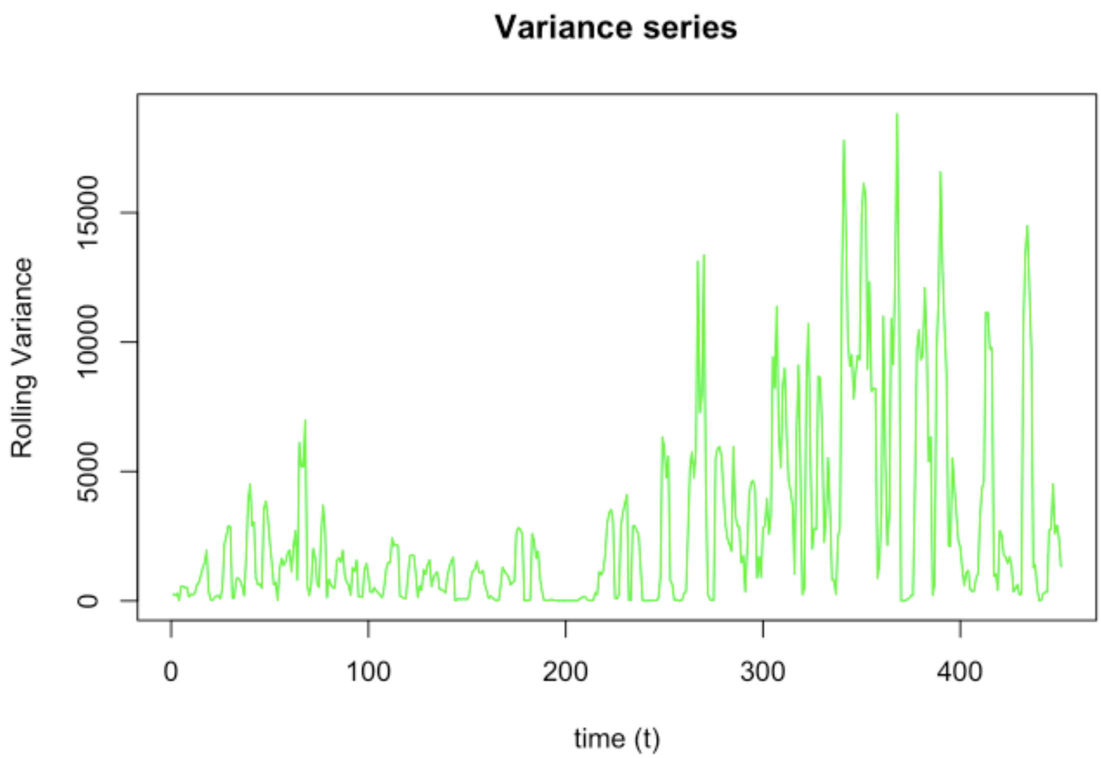


Figure 10: Variance Series of Solar Energy Production

From variance series, we can see that the variance does not show any pattern, meaning that the data idoes not need to be transformed. In order to eliminate non-stationarity from our data, we can difference the series and move on from differenced data to build some of our models.

In order to check the correlations beetween production data and the mean values of the four variables, ggpairs is used:

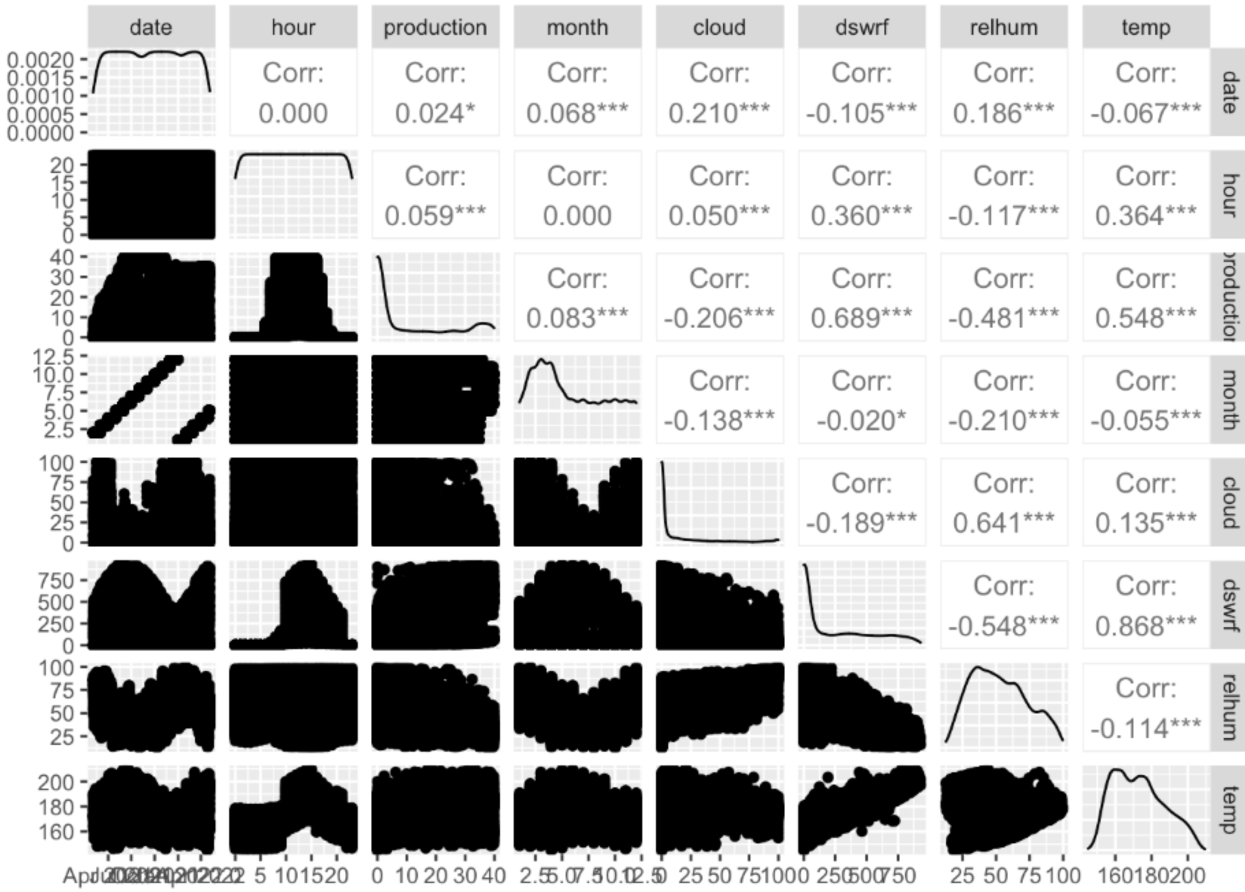


Figure 11: Correlation graph for production and other variables

We can see that there are significant correlations between the production value and the temp, dswrf, and relhum variables. These are to be used during the model building and their significances are to be checked again.

In order to further examine the relationship between production amounts and the variables, four variables' lattitude and longtitude based amounts and production amounts are plotted using ggpairs:

For CLOUD_LOW_LAYER:

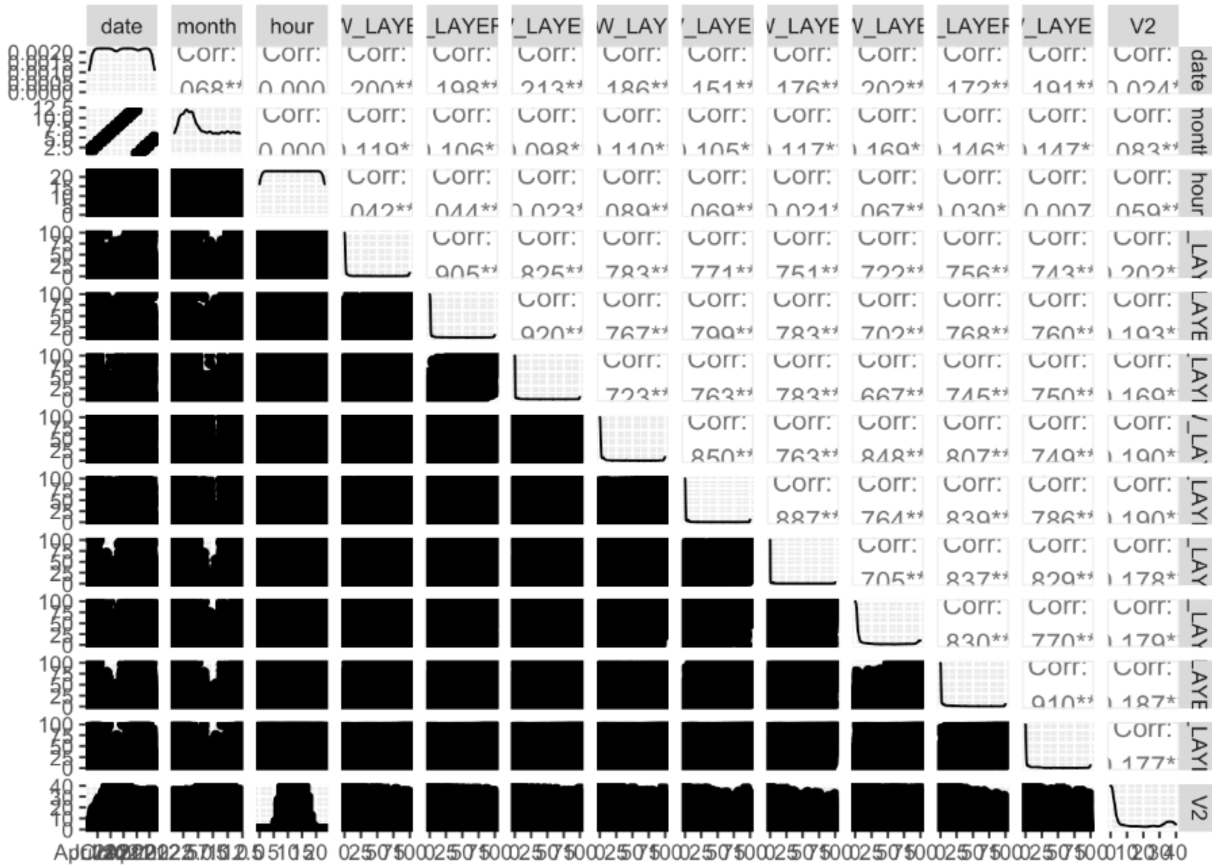


Figure 12: Correlation graph for production and CLOUD_LOW_LAYER_X_Y

We can see that the most correlated value of this variable with production is when the lattitude is 36.75 and the longtitude is 33.25.

For DSWRF:

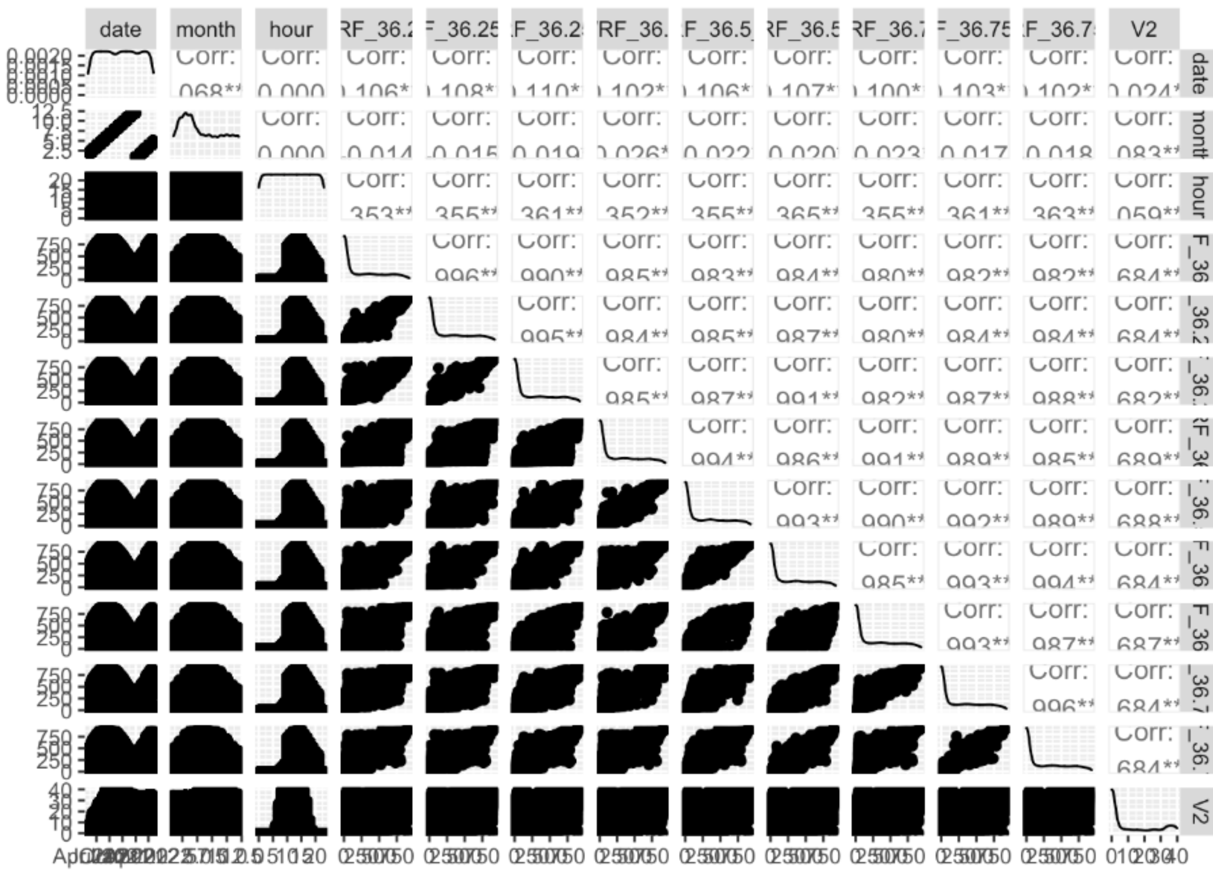


Figure 13: Correlation graph for production and DSWRF_X_Y

We can see that the most correlated value of this variable with production is when the lattitude is 36.25. and the longtitude is 33.

For REL_HUMIDITY:

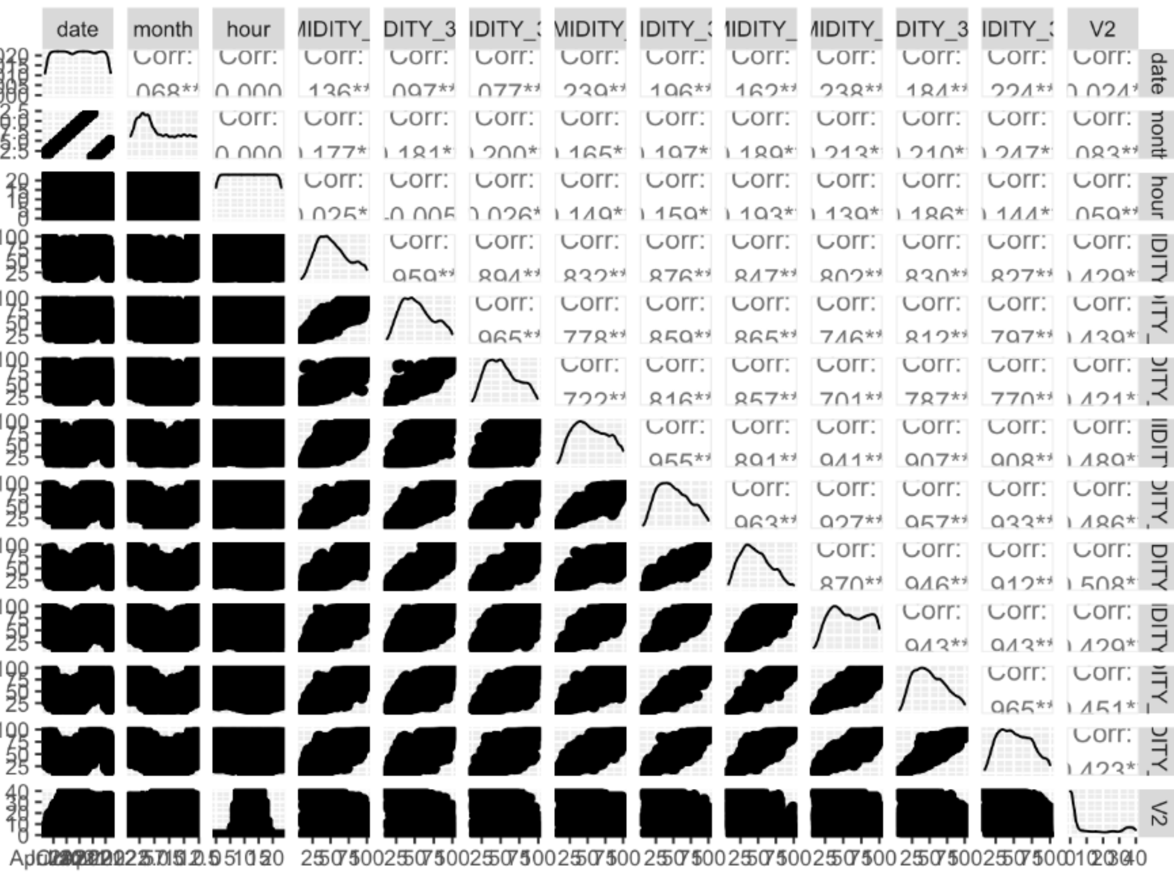


Figure 14: Correlation graph for production and REL_HUMIDITY_X_Y

We can see that the most correlated value of this variable with production is when the lattitude is 36.50 and the longtitude is 33.50.

For TEMP:

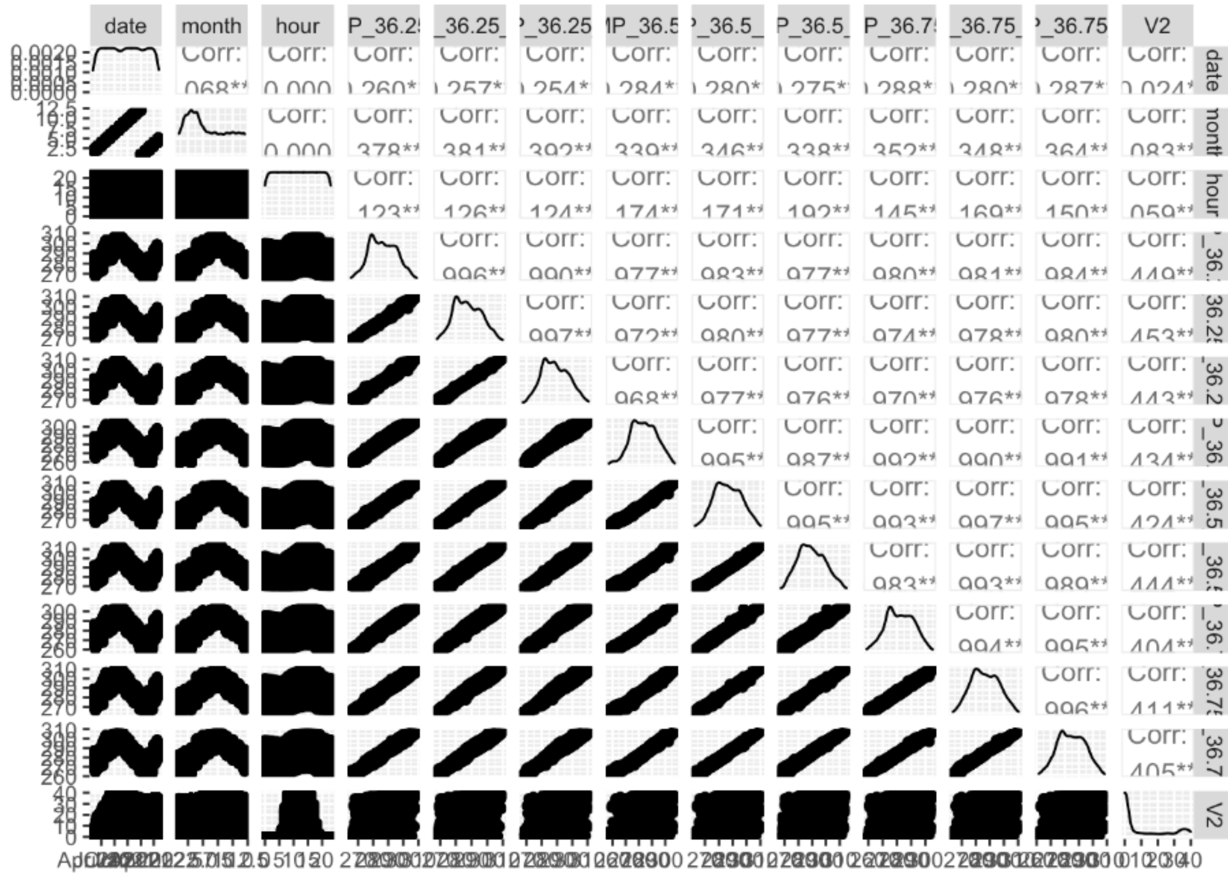


Figure 15: Correlation graph for production and TEMP_X_Y

We can see that the most correlated value of this variable with production are when the lattitude is 36.25 and the longtitude is 33.25 and the lattitude is 36.25 and the longtitude is 33.50.

These variables are to be checked to see if they are reasonable to use in our models.

In order to check the stationarity of the data, we can additionally excercise the KPSS test:

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 5 lags.
##
## Value of test-statistic is: 1.2836
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Figure 15: KPSS Unit Root Test for Stationarity

From the test results, we can see that the test-statistic is reasonably high and therefore we reject the null hypothesis of having a stationary data and will take necessary actions.

2- RELATED LITERATURE

Forecasting is a very significant process to determine the features of a process, service and facility starting from their design processes. Also it should be continued as long as they serve in order to improve their processes and use the resources effectively. In this manner, forecasting solar power production has a powerful impact on determining the distribution of energy. Besides this, since it's a sustainable energy production, optimizing its processes would be a game-changer movement both for the production and consumption of the energy.

In that sense, several articles are reviewed to gain a comprehensive background. And even though we are given the weather variables not to make detailed searches about them but to follow a data-driven approach to understand them, the importance of forecasting the downward shortwave radiation flux impresses us. Estimation of surface downward shortwave radiation over China from AVHRR data based on four machine learning methods paper fascinated us as emphasizing the importance of forecasting DSWRF values accurately and it was very inspiring the way they perform forecasts.

Other than that, we found various articles about forecasting solar energy production, Assessment of forecasting techniques for solar power production with no exogenous inputs[4], Online short-term solar power forecasting[5] and 2D-interval forecasts for solar power production[6] are some examples of work of art. Away from seeing their methodologies while working on similar projects, they are very helpful to gain a perception about the influential and challenging parts of the study.

3- APPROACH

Our general approach is to build regresion model, ARIMA model and ARIMA with regressors model.Even if we decided to move forward with ARIMAX, these three approaches are explained.

First: we divided the data into test and training sets to evaluate the performances of the models.

For the regression model:

We used the variables that seem to be significant during the analysis:

```
##
## Call:
## lm(formula = V2 ~ month + hour + relhum + dswrf + CLOUD_LOW_LAYER_36.75_33.25 +
##     REL_HUMIDITY_36.5_33.5 + TEMP_36.25_33.25 + TEMP_36.25_33.5 +
##     TEMP_36.5_33.5, data = alldatatr[month == 5 | month == 6 &
##     hour >= 5 & hour <= 20])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.5394  -7.0732  -0.9734   5.9928  29.7252
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -2.739e+02  3.098e+01  -8.841  < 2e-16 ***
## month                       6.692e+00  6.890e-01   9.712  < 2e-16 ***
## hour                       -6.300e-01  5.990e-02 -10.517  < 2e-16 ***
## relhum                      1.546e-01  8.574e-02   1.803  0.07168 .
## dswrf                       2.615e-02  1.501e-03  17.427  < 2e-16 ***
## CLOUD_LOW_LAYER_36.75_33.25 -1.790e-01  3.590e-02  -4.987 7.06e-07 ***
## REL_HUMIDITY_36.5_33.5      -7.512e-02  8.555e-02  -0.878  0.38010
## TEMP_36.25_33.25            1.655e+00  5.074e-01   3.262  0.00114 **
## TEMP_36.25_33.5             2.061e-01  4.638e-01   0.444  0.65687
## TEMP_36.5_33.5             -1.003e+00  1.939e-01  -5.173 2.72e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.693 on 1154 degrees of freedom
## Multiple R-squared:  0.5938, Adjusted R-squared:  0.5906
## F-statistic: 187.4 on 9 and 1154 DF,  p-value: < 2.2e-16
```

Figure 16: reg1 model

Based on the stars that show the significance levels, we exlude relhum, REL_HUMIDITY and TEMP variables.

```
##
## Call:
## lm(formula = V2 ~ month + hour + dswrf + CLOUD_LOW_LAYER_36.75_33.25 +
##     TEMP_36.25_33.25 + TEMP_36.5_33.5, data = alldatatr[month ==
##     5 | month == 6 & hour >= 5 & hour <= 20])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -27.200  -7.251  -1.107   6.091  30.073
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -2.384e+02  2.592e+01  -9.196  < 2e-16 ***
## month                        7.297e+00  6.461e-01  11.294  < 2e-16 ***
## hour                        -6.300e-01  5.814e-02 -10.836  < 2e-16 ***
## dswrf                        2.554e-02  1.404e-03  18.192  < 2e-16 ***
## CLOUD_LOW_LAYER_36.75_33.25 -1.599e-01  3.512e-02  -4.553 5.86e-06 ***
## TEMP_36.25_33.25             1.676e+00  1.850e-01   9.060  < 2e-16 ***
## TEMP_36.5_33.5              -9.372e-01  1.917e-01  -4.890 1.15e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.709 on 1157 degrees of freedom
## Multiple R-squared:  0.5913, Adjusted R-squared:  0.5892
## F-statistic:   279 on 6 and 1157 DF,  p-value: < 2.2e-16
```

Figure 17: reg2 model

Adjusted R-squared value decresed, therefore, we did not exclude some variables that we decided not to move on with.

We saw from the descriptive analysis that there are some significant spikes and a seasonality. Therefore, we added corresponding varibales:

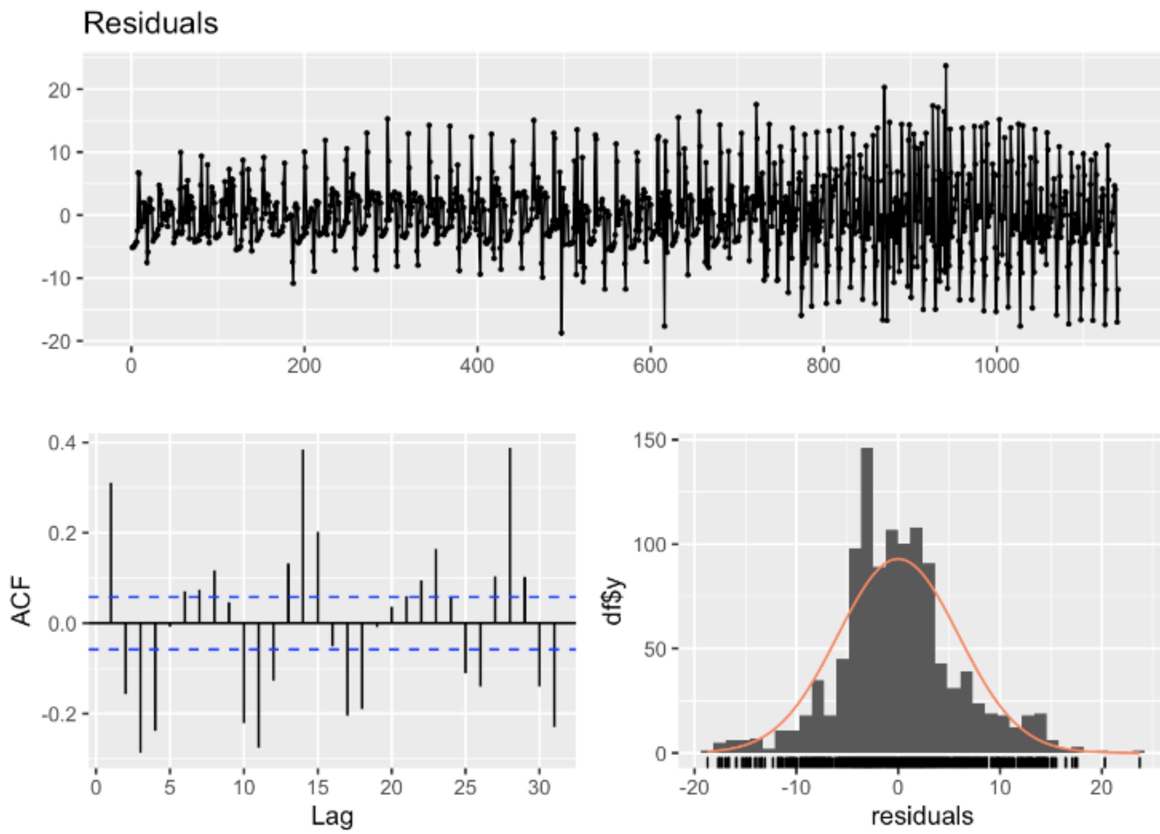
```
##
## Call:
## lm(formula = V2 ~ month + hour + seasonality + Lag1 + Lag24 +
##     relhum + dswrf + CLOUD_LOW_LAYER_36.75_33.25 + REL_HUMIDITY_36.5_33.5 +
##     TEMP_36.25_33.25 + TEMP_36.25_33.5 + TEMP_36.5_33.5, data = alldatatr[month ==
##     5 | month == 6 & hour >= 5 & hour <= 20])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.3142  -3.5386  -0.4856   2.9927  23.1212
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -36.309165  21.260036  -1.708   0.0879 .
## month                        0.830880   0.690554   1.203   0.2291
## hour                        -0.227531   0.038614  -5.893 5.02e-09 ***
## seasonality                 -0.000022   0.001098  -0.020   0.9840
## Lag1                         0.927302   0.024701  37.542  < 2e-16 ***
## Lag24                        0.107203   0.015021   7.137 1.70e-12 ***
## relhum                       0.080753   0.053371   1.513   0.1305
## dswrf                       -0.002539   0.001162  -2.186   0.0290 *
## CLOUD_LOW_LAYER_36.75_33.25 -0.089320   0.022222  -4.019 6.22e-05 ***
## REL_HUMIDITY_36.5_33.5      -0.012396   0.053007  -0.234   0.8151
## TEMP_36.25_33.25             1.371169   0.326307   4.202 2.85e-05 ***
## TEMP_36.25_33.5             -0.657964   0.303375  -2.169   0.0303 *
## TEMP_36.5_33.5              -0.592873   0.121537  -4.878 1.22e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.954 on 1127 degrees of freedom
##   (24 observations deleted due to missingness)
## Multiple R-squared:  0.8478, Adjusted R-squared:  0.8461
## F-statistic:   523 on 12 and 1127 DF,  p-value: < 2.2e-16
```

Figure 18: reg3 model

After adding those terms, the adjusted R-squared value significantly increased. However, some terms seem to be insignificant. We exclude those terms and found our final regression model as reg4 even if there occured a decrese after removing these variables in order not to cause multicollinearity.

```
##
## Call:
## lm(formula = V2 ~ month + hour + Lag1 + Lag24 + relhum + dswrf +
##     CLOUD_LOW_LAYER_36.75_33.25 + TEMP_36.25_33.25 + TEMP_36.5_33.5,
##     data = alldatatr[month == 5 | month == 6 & hour >= 5 & hour <=
##         20])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.7201  -3.4219  -0.4599   3.0240  23.7223
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -38.292038  19.790719  -1.935  0.05326 .
## month                        0.671099   0.443199   1.514  0.13025
## hour                        -0.220421   0.037463  -5.884 5.28e-09 ***
## Lag1                         0.926009   0.024601  37.641  < 2e-16 ***
## Lag24                        0.104145   0.014913   6.984 4.90e-12 ***
## relhum                       0.064456   0.022854   2.820  0.00488 **
## dswrf                       -0.002763   0.001143  -2.418  0.01578 *
## CLOUD_LOW_LAYER_36.75_33.25 -0.088129   0.022117  -3.985 7.19e-05 ***
## TEMP_36.25_33.25             0.719103   0.127514   5.639 2.16e-08 ***
## TEMP_36.5_33.5              -0.590301   0.120032  -4.918 1.00e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.96 on 1130 degrees of freedom
##   (24 observations deleted due to missingness)
## Multiple R-squared:  0.8471, Adjusted R-squared:  0.8458
## F-statistic: 695.4 on 9 and 1130 DF,  p-value: < 2.2e-16
```

Figure 19: reg4 model




```
##
##  Breusch-Godfrey test for serial correlation of order up to 13
##
## data:  Residuals
## LM test = 392.21, df = 13, p-value < 2.2e-16
```

Figure 20: Breusch-Godfrey test

This model seems to have normal distribution even if it has undesired peaks in ACF plot. Unfortunately we were not able to remove this unwanted situation.

After finding the model, we apply the model to the test data set and get predictions.

```
##       121       122       123       124       125       126       127
##  4.0762123 3.9720182 3.9569161 3.7387362 3.7324903 3.3694011 3.4752905
##       128       129       130       131       132       133       134
##  8.2483890 26.8981402 36.5094390 28.8080528 29.8128867 28.8643234 29.7210571
##       135       136       137       138       139       140       141
## 24.2826656 15.5557947 6.6687880 1.6452443 1.4010987 0.3956536 -2.6618483
##       142       143       144
## -2.4336004 -1.0762267 -4.3505321
```

Figure 21: Results of Regression Model

For the ARIMA model:

We built our model to predict hourly production amount and in order to do so, we built model for every hour for which there is a production, which means we excluded 0 production hours: from 20 to 05 and perform forecasting for each hours between and including 06-19. Here is the example for hour==12 and we get the result for each hour by changing the hour variable by hand.

First, we differenced the data as we found out that the data is not stationary.

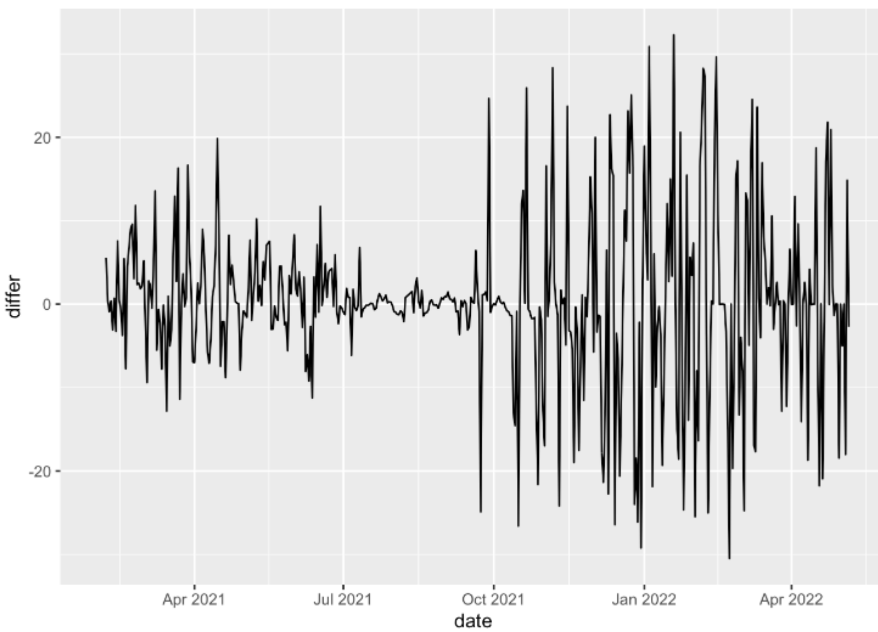


Figure 21: Differenced data

After that, we applied KPSS test to see if there needs to be further differencing

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 5 lags.
##
## Value of test-statistic is: 0.1076
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Figure 22: KPSS test for differenced data

Since the test-statistic is reasonably small, there is no need to difference the data further.

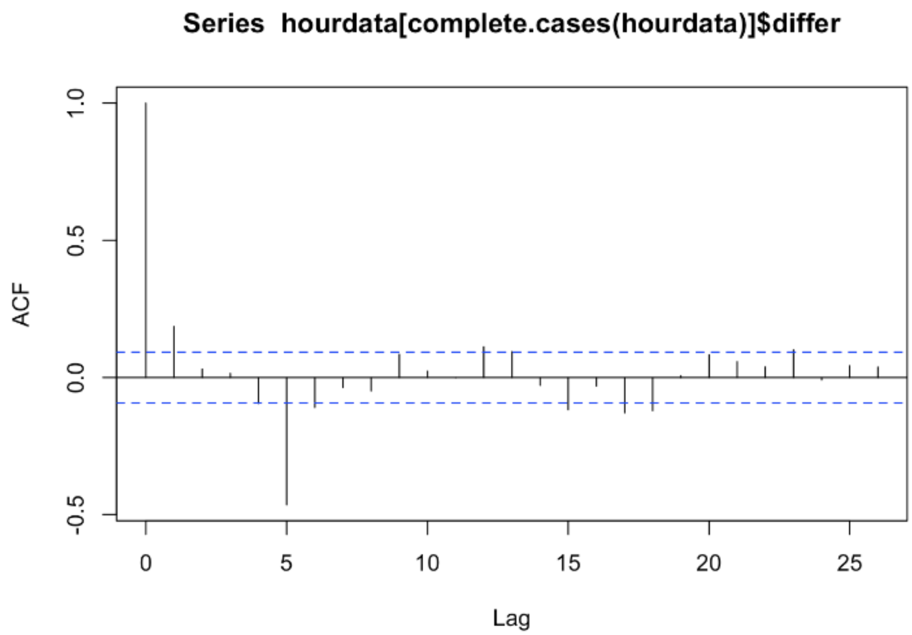


Figure 23: ACF for differenced one hour data

Looking at the PCF plot, we see that there are two significant spikes at lag 1 and 5.

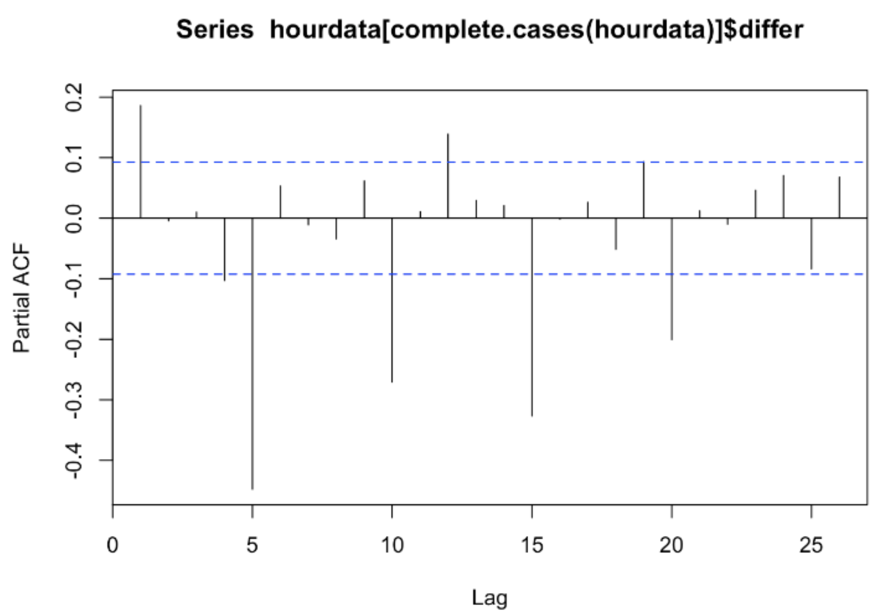


Figure 24: PACF for differenced one hour data

From the PACF plot, we can see that there are significant spikes at lag 1, 5,10,15..

In order to find the appropriate ARIMA model, we used auto.arima function:

```
## Series: hourdata$differ
## ARIMA(5,0,0) with zero mean
##
## Coefficients:
##          ar1     ar2     ar3      ar4      ar5
##       0.1417  0.0078  0.0250  -0.0137  -0.4491
## s.e.  0.0421  0.0427  0.0429   0.0429   0.0422
##
## sigma^2 = 74.72:  log likelihood = -1603.64
## AIC=3219.27   AICc=3219.46   BIC=3243.92
```

Figure 25: Proposed Model by auto.arima

It gave us the ARIMA(5,0,0) model to use.

After having the model, we used it to perform forecasting and get the results:

mo... <dbl>	V2 <dbl>	cloud <dbl>	relhum <dbl>	temp <dbl>	dswrf <dbl>	week <date>	differ <dbl>	predicted_differ <dbl>	forecastval <dbl>
4	35.00	0.0000000	14.61111	191.3153	842.8044	2022-04-24	0.00	0.00000	35.00000
4	35.00	0.0000000	24.22222	195.6204	826.4444	2022-04-24	0.00	0.00000	35.00000
4	16.56	0.0000000	23.66667	194.6047	787.3867	2022-04-24	-18.44	-18.44000	16.56000
5	35.00	10.6666667	47.05556	199.4425	777.2978	2022-05-01	0.00	0.00000	35.00000
5	28.66	0.2888889	44.98889	202.3432	780.8067	2022-05-01	-5.01	-5.01000	28.66000
5	35.00	2.3666667	41.78889	202.1769	803.7267	2022-05-01	0.00	0.00000	35.00000
5	16.98	2.8777778	38.98889	201.0118	802.3533	2022-05-01	-18.02	-18.02000	16.98000
5	31.40	69.2666667	53.86667	194.6252	502.3511	2022-05-01	14.84	14.84000	31.40000
5	32.28	9.1333333	39.06667	197.0454	818.7711	2022-05-01	-2.72	-2.72000	32.28000
2	NA	71.1111111	78.55556	189.5000	243.3333	2021-01-31	NA	1.52863	30.18863

Table 1: Resulting table of ARIMA(5,0,0) and the forecasted value

ARIMAX

We added CLOUD_LOW_LAYER variable to our arima model, which decreased the AICc value significantly.

```
## Series: complete_dat$differ
## Regression with ARIMA(5,0,0) errors
##
## Coefficients:
##          ar1      ar2     ar3      ar4      ar5     xreg
##      -0.0664  -0.0041  0.1085  -0.0358  -0.4881  -0.1002
## s.e.  0.1105   0.1110  0.1149   0.1156   0.1151   0.0409
##
## sigma^2 = 86.71:  log likelihood = -242.14
## AIC=498.29   AICc=500.19   BIC=513.72
```

Figure 26: Model features of ARIMAX

And the result:

relhum <dbl>	temp <dbl>	dswrf <dbl>	week <date>	differ <dbl>	cloud_diff <dbl>	res <dbl>	predicted_differ <dbl>	forecastval <dbl>
24.22222	195.6204	826.4444	2022-04-24	0.00	0.0	0.2275646	0.000000	35.00000
23.66667	194.6047	787.3867	2022-04-24	-18.44	0.0	-8.9811587	-18.440000	16.56000
47.05556	199.4425	777.2978	2022-05-01	0.00	6.6	3.6054484	0.000000	35.00000
44.98889	202.3432	780.8067	2022-05-01	-5.01	0.0	-5.4639231	-5.010000	28.66000
41.78889	202.1769	803.7267	2022-05-01	0.00	0.0	1.1716525	0.000000	35.00000
38.98889	201.0118	802.3533	2022-05-01	-18.02	0.0	-18.2340540	-18.020000	16.98000
53.86667	194.6252	502.3511	2022-05-01	14.84	100.0	9.2382138	14.840000	31.40000
39.06667	197.0454	818.7711	2022-05-01	-2.72	-6.5	-4.7517460	-2.720000	32.28000
78.55556	189.5000	243.3333	2021-01-31	NA	NA	NA	1.263756	29.92376
54.48889	176.9901	321.1111	2021-01-31	NA	NA	NA	1.263756	36.26376

Table 1: Resulting table of ARIMA(5,0,0) with regressor and the forecasted value

4- RESULT

Hourly production amounts are mostly reasonable and with some exceptions, our model gave an average error rate comparing to the other contestants. Since we did not prepare a model evaluation by wMAPE and other metrics such as fBias, as we are to mention in future work, we could not provide a formal comparison of our models and corresponding results.

5- CONCLUSION and FUTURE WORK

In conclusion, we built three models and used ARIMAX model to predict future values. Beforehand, we perform descriptive analysis to understand the data and which actions were needed. Then, we built our model in such a way that gave us the prediction for each our by changing hour value by hand.

Forecasting solar energy production based on weather was a challenging task involving dedicated research, evaluation and reevaluation, and high analytical skills.

When it comes to what we could do better: Rather than forecasting every hour one by one, we could come up with more efficient and effective solution, such as dividing time slots. We could do a more comprehensive descriptive analysis and work more on to transform the data into desired form containing no stationarity, with normal distribution and no significant spikes at ACF and PACF values. In addition, we could do more work on how to evaluate our models based on wMAPE and other metrics. Even if we tried to make our model better during the submission phase, we did not change our model thoroughly. For example, rather than using CLOUD_LOW_LAYER variable in our ARIMAX model, we could use some other significant variables.

6- References

[1] "Humidity", Wikipedia, 21 May 2022, https://en.wikipedia.org/wiki/Humidity

[2] "DATA PRODUCTS: DOWNWARD SHORTWAVE RADIATION (SURFACE)", Goes-R, https://www.goes-r.gov/products/baseline-DSR.html#:~:text=The%20downward%20shortwave%20radiation%20

[3] Wei, Yu, et al. "Estimation of Surface Downward Shortwave Radiation over China from AVHRR Data Based on Four Machine Learning Methods." Solar Energy, vol. 177, Jan. 2019, https://reader.elsevier.com/reader/sd/pii/S0038092X18311083?token=07EA2B09AA0A9D84214E5AAC6B8FC7C6B80C91C4DCB5FABCA519E6076121AB46BD8CA54B93AECECBCFD35D8A85C8E1E7&originRegion=eu-west-1&originCreation=20220607180421

[4] Pedro, Hugo T. C., and Carlos F. M. Coimbra. "Assessment of Forecasting Techniques for Solar Power Production with No Exogenous Inputs - ScienceDirect." Assessment of Forecasting Techniques for Solar Power Production with No Exogenous Inputs - ScienceDirect, May 2012, https://reader.elsevier.com/reader/sd/pii/S0038092X12001429?token=F0FA289ECDF21EDD79361F914870B626D0BC43AB95951319CA0FB4AFA5D6BF6932FFA4272D47A9A104EFB1174A0FE578&originRegion=eu-west-1&originCreation=20220607181136

[5] Bacher, Peder, et al. "Online Short-Term Solar Power Forecasting - ScienceDirect." Online Short-Term Solar Power Forecasting - ScienceDirect, July 2009, https://reader.elsevier.com/reader/sd/pii/S0038092X09001364?token=7C6308AD219E39717D4E4C4F1FF2BDBD9C40EA5DA8D85E5F8E701C01C78392866DFF9E1CAB072A46B6393460C58867EF&originRegion=eu-west-1&originCreation=20220607181618

[6] Rana, Mashud, et al. "2D-Interval Forecasts for Solar Power Production - ScienceDirect." 2D-Interval Forecasts for Solar Power Production - ScienceDirect, September 2015, https://reader.elsevier.com/reader/sd/pii/S0038092X15004545?token=29759299FAAE7AC95377D3A95351FA6BA97CBAAF0A18007389776C16EEC0E7EF22F529BE6FDF3FB9500DB340978BADA3&originRegion=eu-west-1&originCreation=20220607181854