BOĞAZİÇİ UNIVERSITY

IE360: STATISTICAL FORECASTING AND TIME SERIES

“Solar Power Forecasting”
Project Report
Spring 2024

Prepared By

Ahmet ÇELİKER- 2020402126
Ahmet Emir ŞENER – 2021402240
Emir TAMER –2021402249

Table of Content

1. INTRODUCTION

The goal of this project is to predict the hourly solar power output of Edikli GES (Güneş Enerjisi Santrali) for the following day, utilizing time series data that spans from January 1, 2022, to the most recent available data, which is up to two days before the forecast date. Managing total production, transporting energy, and adhering to legal regulations are challenges faced by these facilities. To improve production planning, the facility seeks to predict the probable hourly energy production for the next day, measured in megawatt-hours (MWh). The hourly energy production from the panels is influenced by multiple factors. Direct sunlight exposure is crucial for maximizing panel capacity. Additionally, the condition of the panels, including dust coverage, electronic and material integrity, as well as legal, financial, and logistical constraints, can significantly impact production at any given time. Since direct sunlight is essential, utilizing weather forecasts for prediction is logical. A comprehensive weather dataset is available, including hourly measurements of cloud coverage, temperature, shortwave radiation flux, solar radiation, and snow.

The report will begin by creating an aggregated daily linear regression model to serve as a baseline comparison. Following this, we will develop time series linear regression models incorporating various regressors. After constructing these models, we will compare their performance based on their Weighted Mean Absolute Percentage Errors (WMAPEs). The model with the best performance will then be used to make the final prediction. The dataset includes hourly production values along with corresponding weather conditions for those hours.

It's crucial to factor in the operational constraints of the Edikli GES when addressing the hourly solar energy production issue. While the facility has a capacity of 15 MW, licensing restrictions limit production to 10 MW, imposing a tangible upper limit on our modeling efforts and adding complexity to real-world problem. Additionally, fluctuations in production exceeding 10 MW within a defined time frame are attributed to government-led expansions in electricity generation, presenting another significant challenge for our model.

To forecast hourly electricity production amounts of Edikli GES, we first tried to identify the results of base models, then tried to create 4 clusters and evaluate the summed productions with summed regressors. Due to the hardness of disaggregation of summed forecast data, finally we resulted our final model with the hourly production data where we used month feature interaction terms.

For this project, both R and Python codes have been used for visualization, forecasting, and modeling purposes with the following libraries:

Python:

```
import pandas as pd
import numpy as np
from datetime import datetime
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf
import statsmodels.api as sm
from scipy.stats import boxcox
import seaborn as sns
from scipy.stats import norm
```

R:

```
require(base), require(data.table), require(datasets), require(dplyr), require(forecast)
require(GGally), require(ggplot2), require(ggstats), require(graphics),
require(grDevices), require(lubridate), require(methods), require(readr), require(stats),
require(utils), library(readr), library(ggplot2), library(dplyr), library(ggcorrplot),
library(readr), library(tidyr), library(Rcpp), library(zoo), require(rpart), require(rattle),
library(openxlsx)
```

2. Preparing the Data

a. Arranging the Data: There are two data sets to be used in the project where one of them is containing the daily production amount data and the other one contains the the regressors: related cloud cover(**TCDC_low.cloud.layer, TCDC_middle.cloud.layer, TCDC_high.cloud.layer, TCDC_entire.atmosphere)**, downward shortwave radiation (**DSWRF_surface**), solar radiation (**USWRF_top_of_atmosphere, USWRF_surface, DLWRF_surface**), snow (**CSNOW_surface)** and temperature (**TMP_surface)** data for 25 grid points. We started by turning the long format regressor data to wide format. Then, we merged the regressor data set with the production data set. Regressors are calculated the inverse distance weights between a data point and a set of grid points and then normalizes these weights.

b. Filling the Missing Data: To prevent the models from encountering NAs due to missing production data, we filled each NA position by taking the average of the production data from the previous and next hour. We used a for loop starting from the first date, so if there are two consecutive days with NAs, they will both be filled using the same method.

c. Adding Dates and Simple Regressors: We added month and hour information columns. During the model building process, new datasets were created using columns such as lag columns and interaction term columns, categorical variable columns which are necessary for the model.

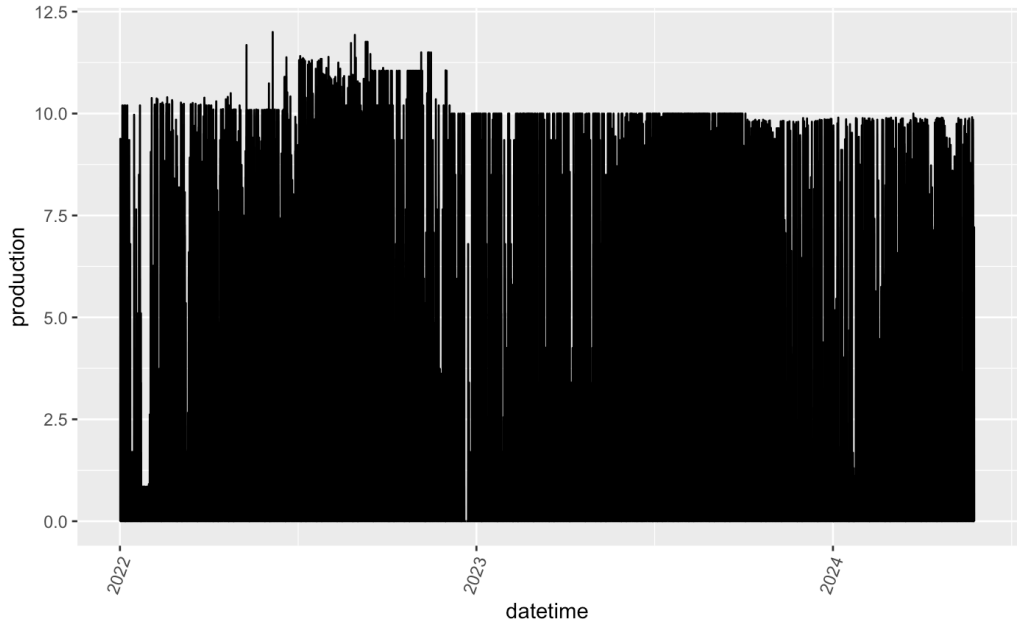
d. Deleting Hours with No Production: Between 8 p.m. and 5 a.m., there is no significant production, so we deleted the rows within these hours. By default, these hours are assigned “0” at the forecast phase.

e. Adding Inverse Distance Weights of Regressors and Capacity Regressors: Because there is regressor data for 25 different locations and considering that points closer to the center have a greater influence on production, a value was obtained for each regressor using the inverse distance method, where the weight of each regressor is determined based on its proximity to the center. We added average of these locations for each regressor as a new regressors.

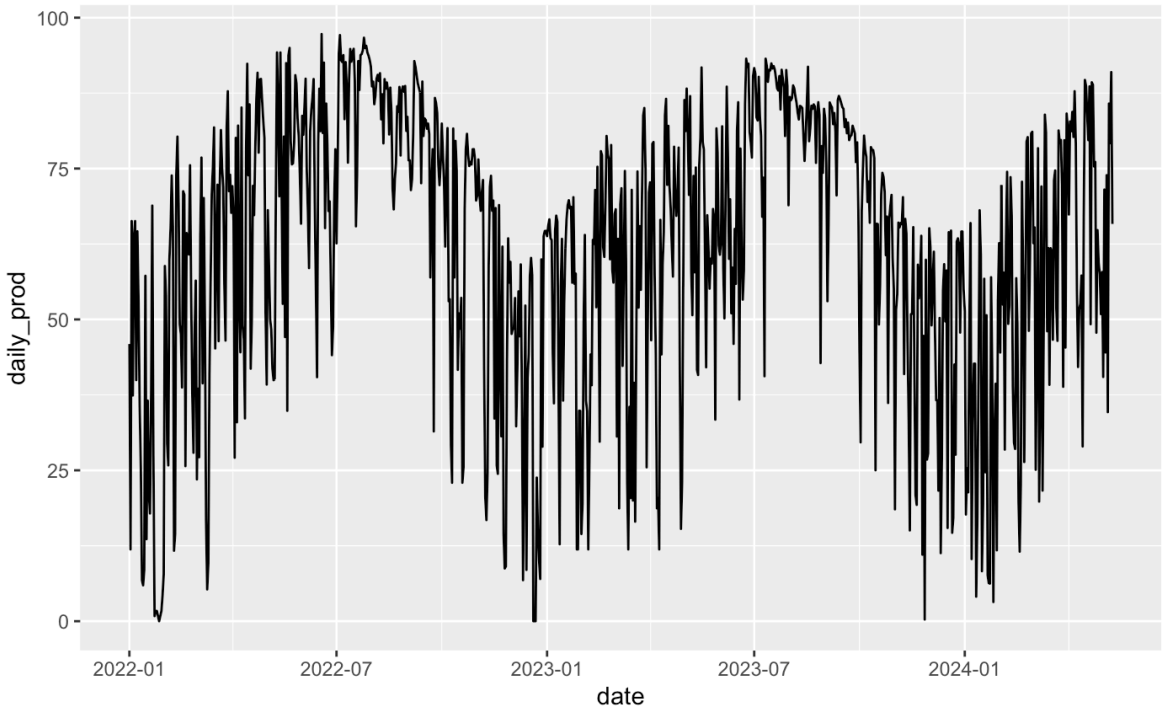
```r
calculate_weights <- function(data_point, grid_points) {
  weights <- vector()
  for (i in 1:length(grid_points)) {
    distance <- sum((grid_points[[i]] - data_point)^2)
    inv_distance <- 1 / distance
    weights <- c(weights, inv_distance)
  }
  total_weight <- sum(weights)
  weights <- weights / total_weight
  return(weights)
}

data_point <- c(38.29, 34.97)
grid_points <- list(
  c(37.75, 34.5), c(37.75, 34.75), c(37.75, 35), c(37.75, 35.25), c(37.75, 35.5),
  c(38, 34.5), c(38, 34.75), c(38, 35), c(38, 35.25), c(38, 35.5),
  c(38.25, 34.5), c(38.25, 34.75), c(38.25, 35), c(38.25, 35.25), c(38.25, 35.5),
  c(38.5, 34.5), c(38.5, 34.75), c(38.5, 35), c(38.5, 35.25), c(38.5, 35.5),
  c(38.75, 34.5), c(38.75, 34.75), c(38.75, 35), c(38.75, 35.25), c(38.75, 35.5)
)
```

```r
weights <- calculate_weights(data_point, grid_points)

weather_info_weighted <- data.frame(matrix(0, nrow = nrow(weather_wide[, -(1:2)]), ncol = 10))
colnames(weather_info_weighted) <- paste0("weighted_group_", 1:10)

for (k in 1:10) {
  weighted_sum <- numeric(nrow(weather_wide[, -(1:2)]))
  start_col <- (k - 1) * 25 + 1
  end_col <- start_col + 24

  for (j in start_col:end_col) {
    col_name <- colnames(weather_wide[, -(1:2)])[j]
    weighted_sum <- weighted_sum + weather_wide[, -(1:2)][[col_name]] * weights[(j - start_col + 1)]
  }

  weather_info_weighted[, k] <- weighted_sum
}

colnames(weather_info_weighted) <- c("dswrf_surface",
"tdc_low.cloud.layer","tdc_middle.cloud.layer","tdc_high.cloud.layer","tdc_entire.atmosphere","uswrf_top_of_atmosphere",
                                     "csnow_surface","dlwrf_surface","uswrf_surface","tmp_surface")

weather_info_weighted <- cbind(datetime_info, weather_info_weighted)
tail(weather_info_weighted, 100)
```

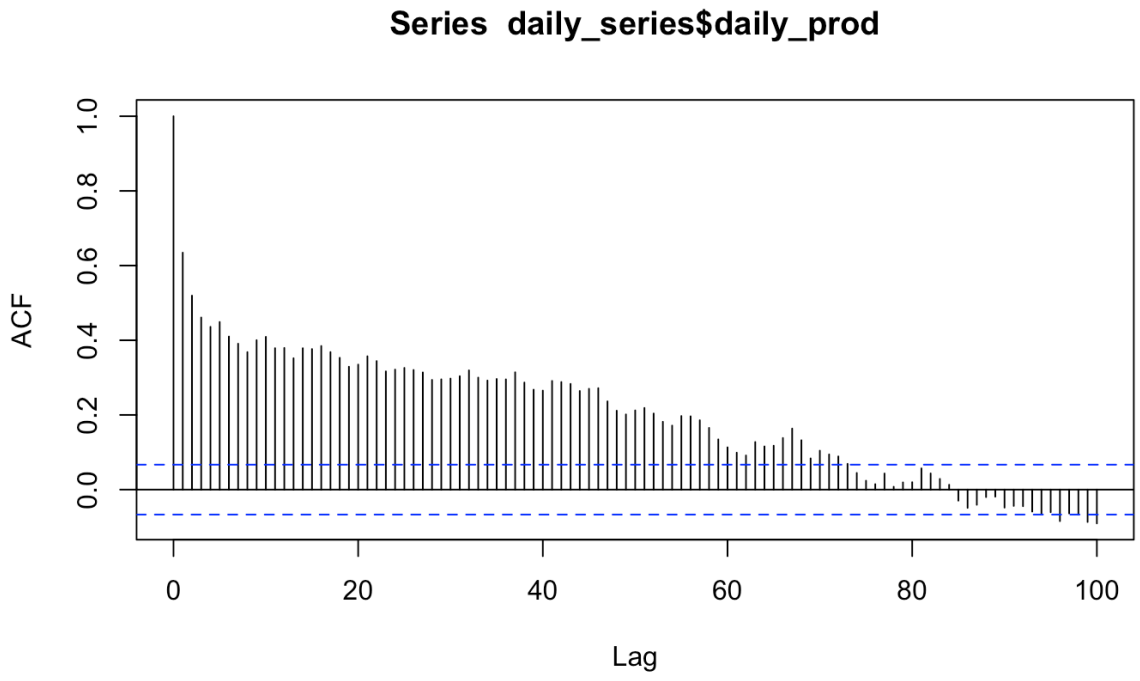
The merging process aligns the weather data and production datasets based on their corresponding dates.



The above figure shows the levels of energy production of the facility. One can easily detect the maximum production thresholds of 10 and 12.5 in different time intervals. As can also be seen, there is generally a production that increases from the beginning of the year and decreases towards the end of the year. The reasons for these decreases are seasonal. There is a seasonality that shows a decrease in the winter months and an increase in the summer months. To make a clear interpretation, hourly data was aggregated and converted into daily production data.

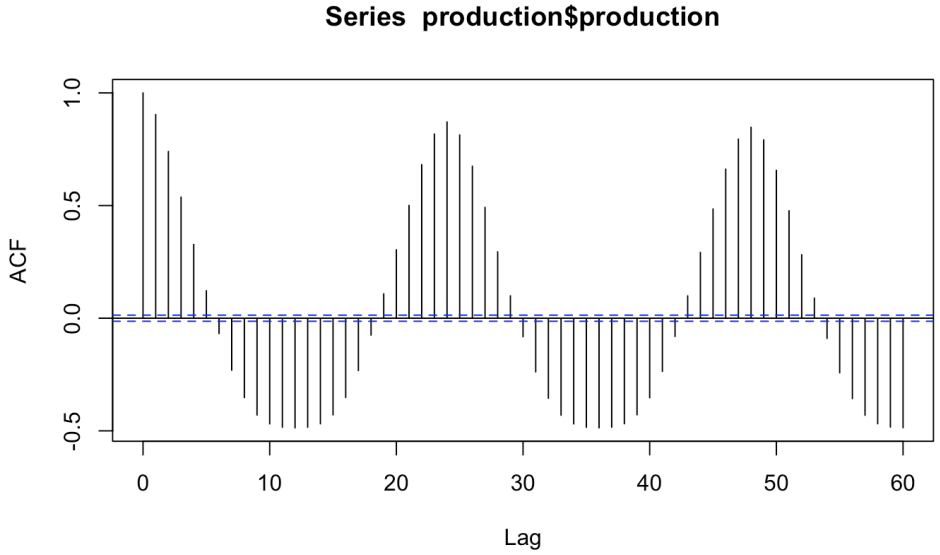


As can be understood from this graph, there is a typical seasonality. There is also a seasonality with a similar pattern in the predictor variables. The ACF plot shows us the seasonality that we observe more accurately.

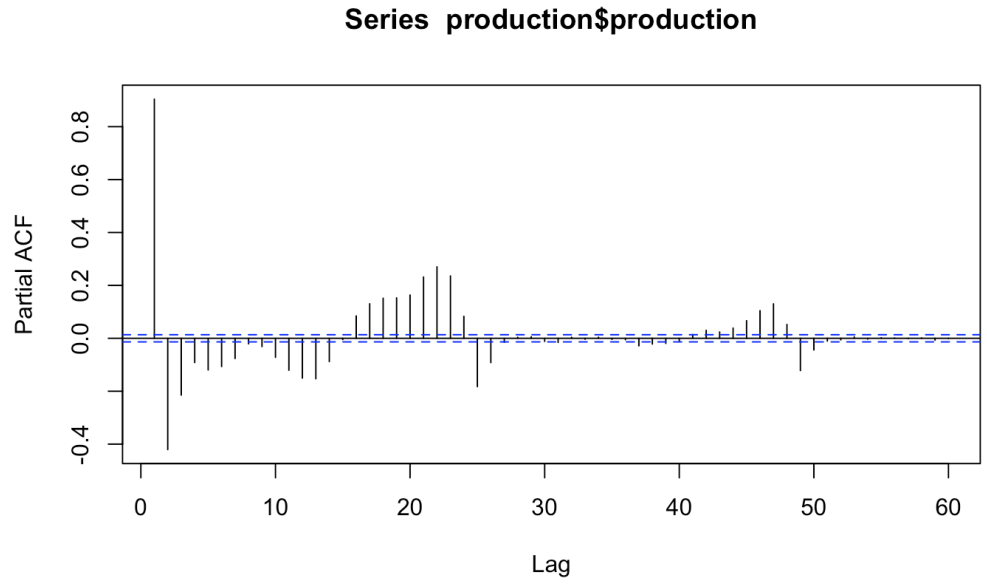


An **ACF plot with a comb-like shape** can be an indicator of the presence of **trend and seasonality**. This suggests that there is a pattern in the time series data and that this pattern tends to repeat over time.

The ACF plot of hourly production is shown below. This behavior indicates that the time series exhibits seasonality at the positive and negative peaks of the sinusoidal ACF, which occur at lag 12.

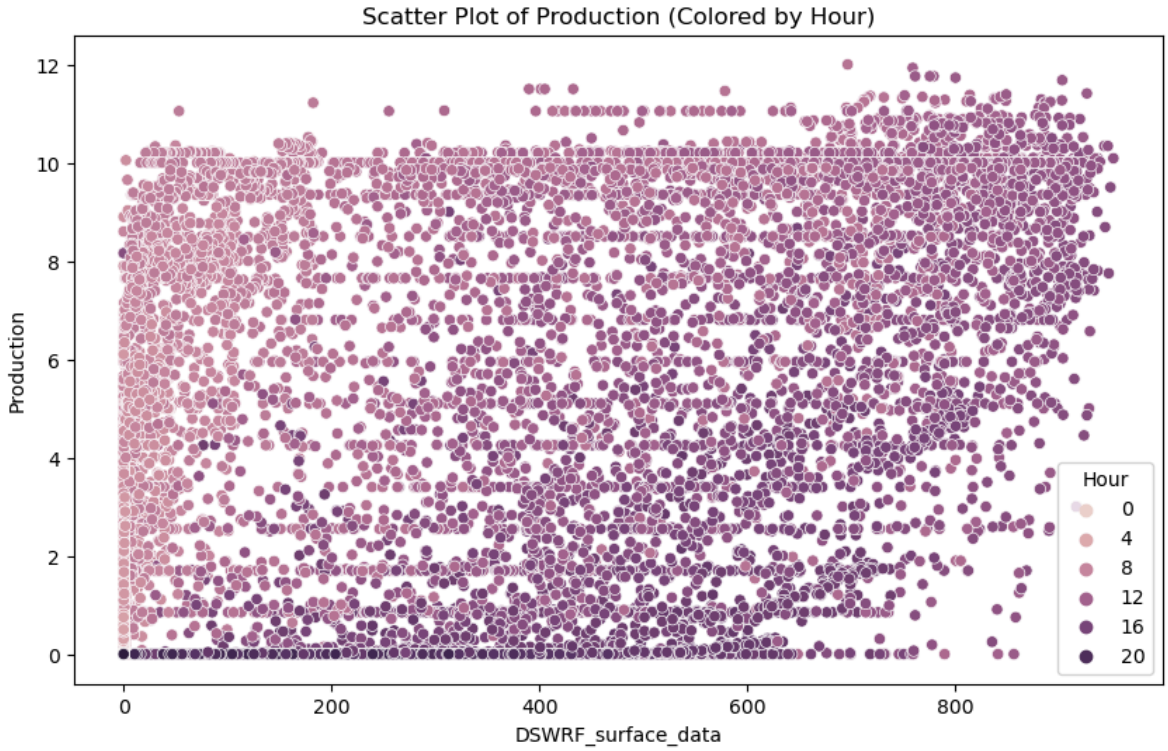


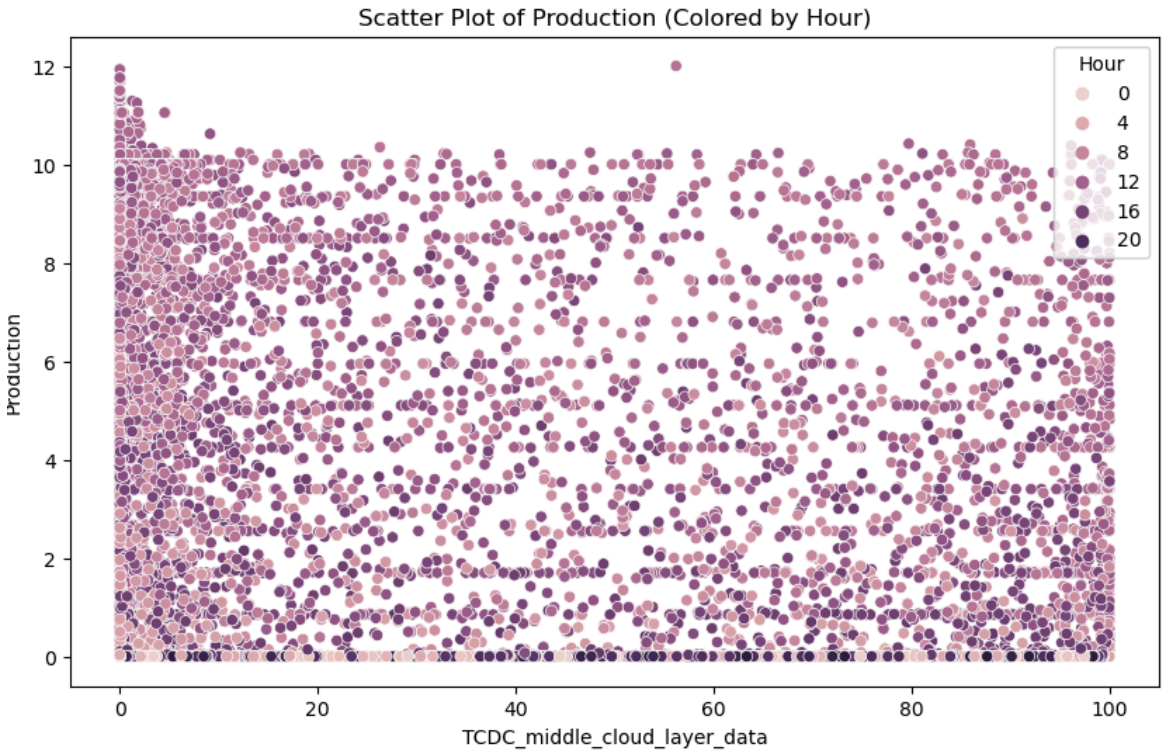
The Partial Autocorrelation Function (PACF) plot of electricity production shown here provides insights into the time series data by illustrating the partial correlation between observations at different lags.

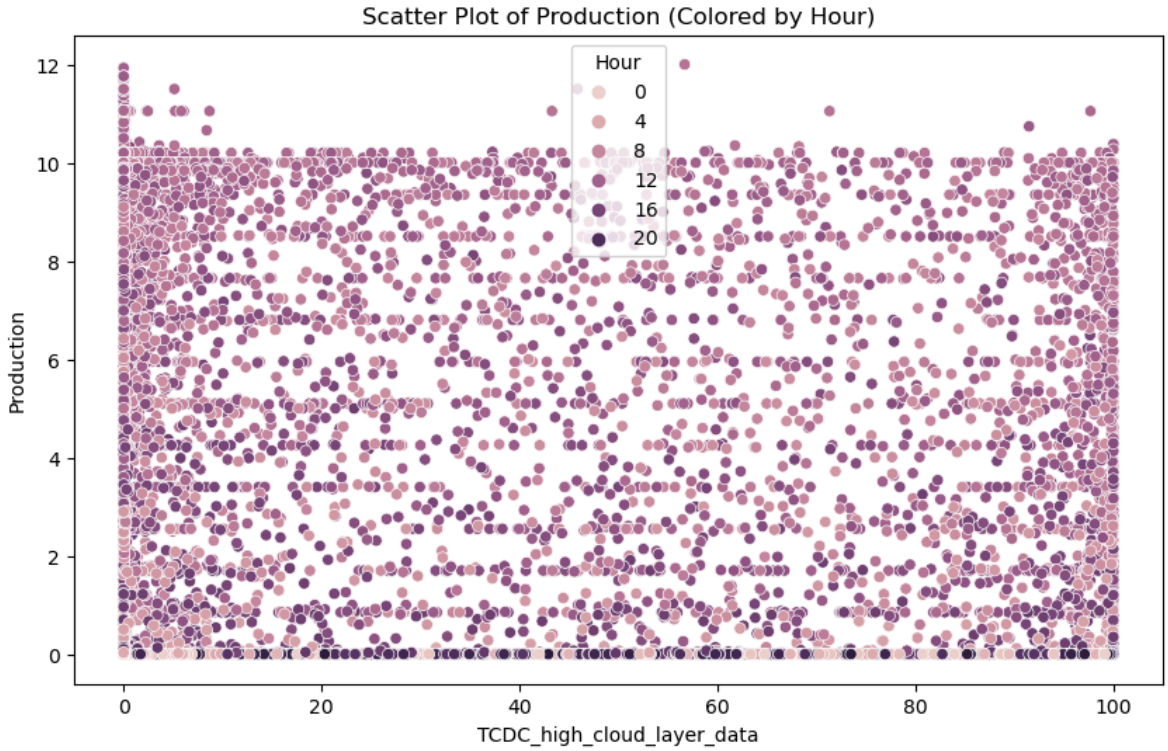


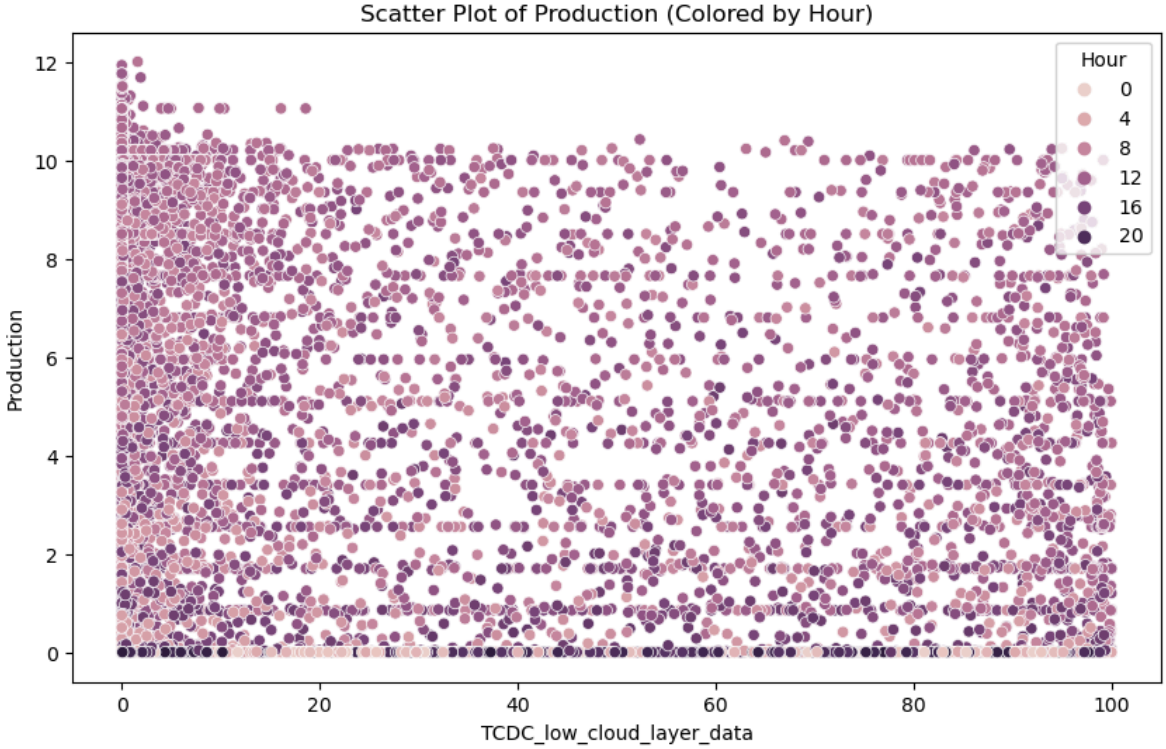
There is a significant positive spike at lag 1, indicating that the current value of electricity production is highly correlated with its immediate previous value. This suggests strong short-term autocorrelation. The sinusoidal pattern in the PACF plot, with alternating positive and negative spikes, reinforces the presence of seasonality in the data. The magnitude of the spikes decreases as the lag increases, indicating that while there is seasonality, the influence of past values diminishes over time.

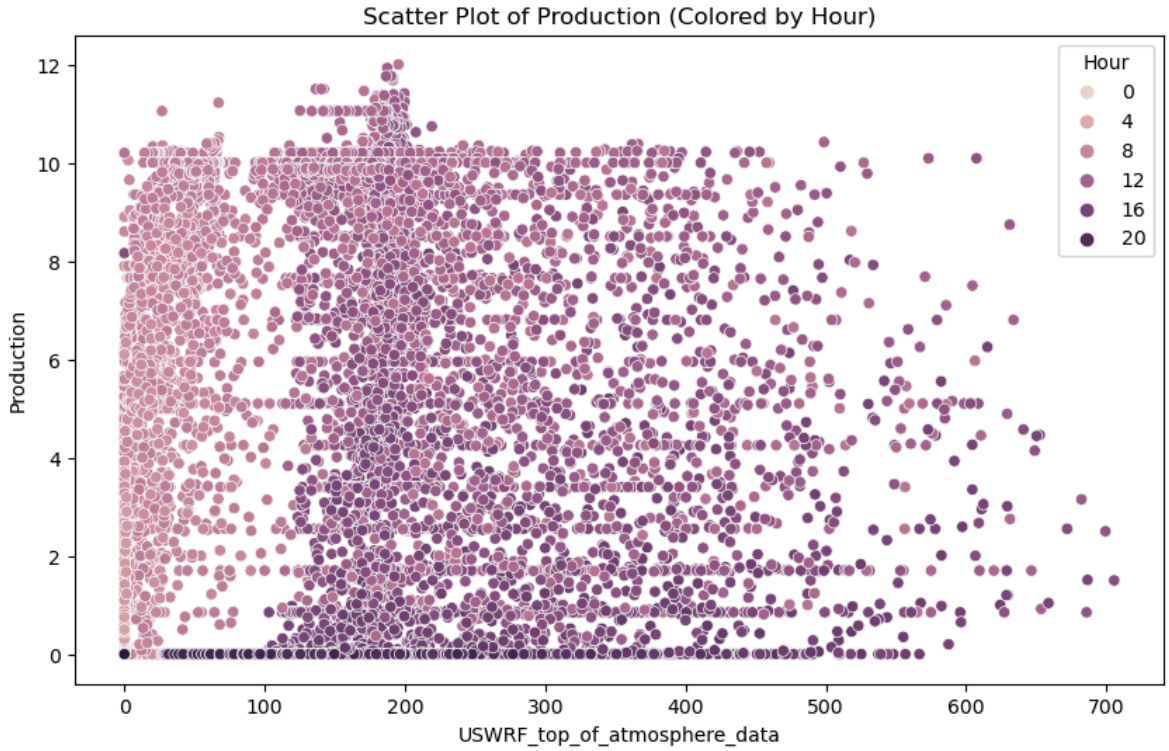
To obtain the relationships between production amounts vs the regressors, all the regressors have been scattered with production amounts by hueing the production data by hours as bellow. From the hued scatterplots, it can be interpreted that for different hours, the effects of regressors on production are different and different relationships between regressors and production amount could be observed.

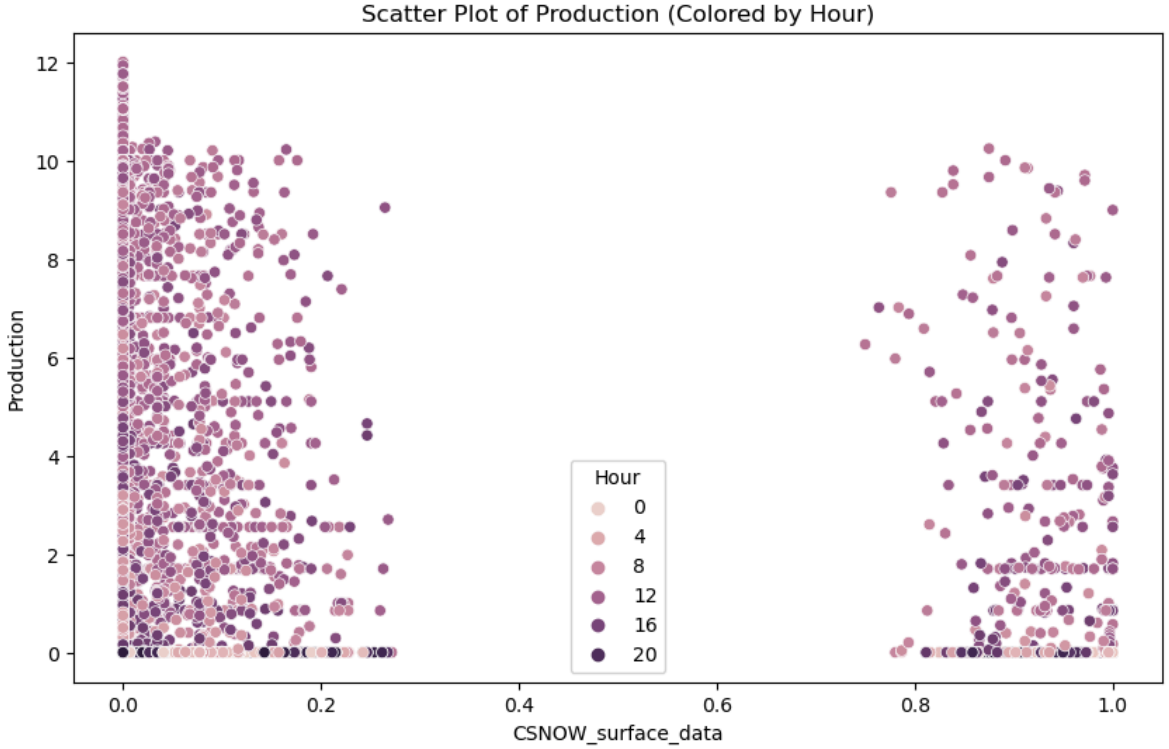




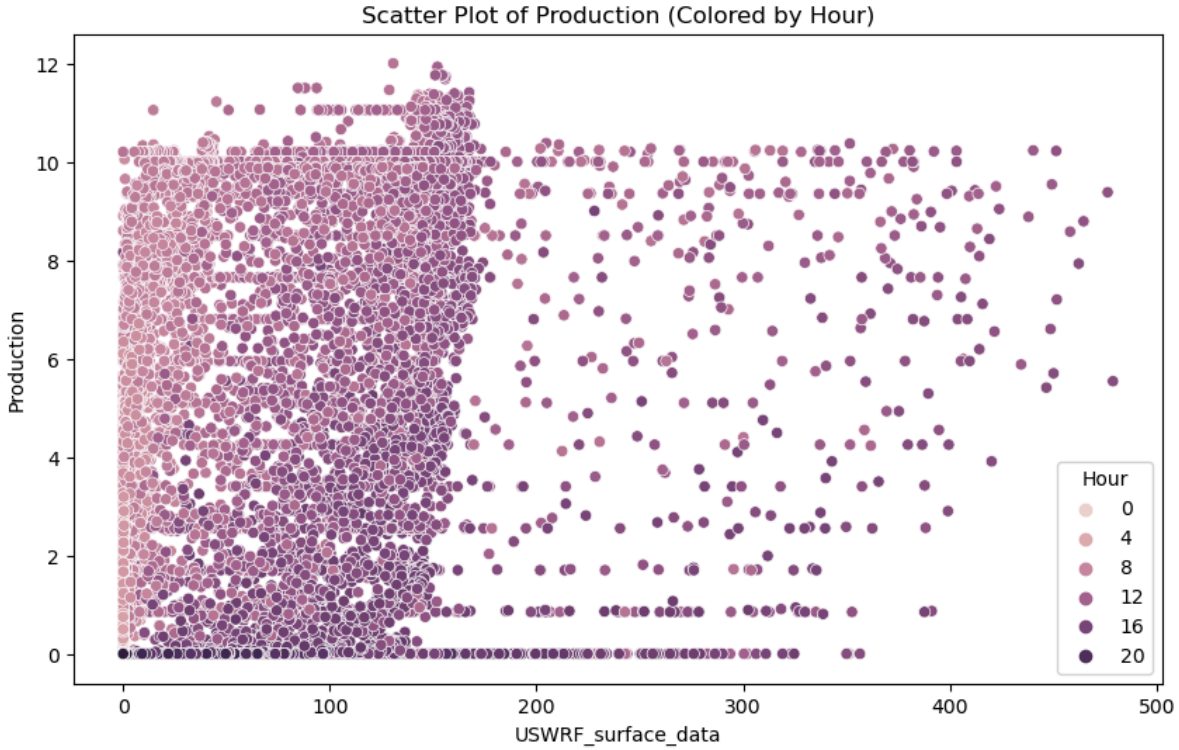

Scatter Plot of Production (Colored by Hour)
Hour
0
4
8
12
16
20
Production
TCDC_high_cloud_layer_data
0
2
4
6
8
10
12
0
20
40
60
80
100

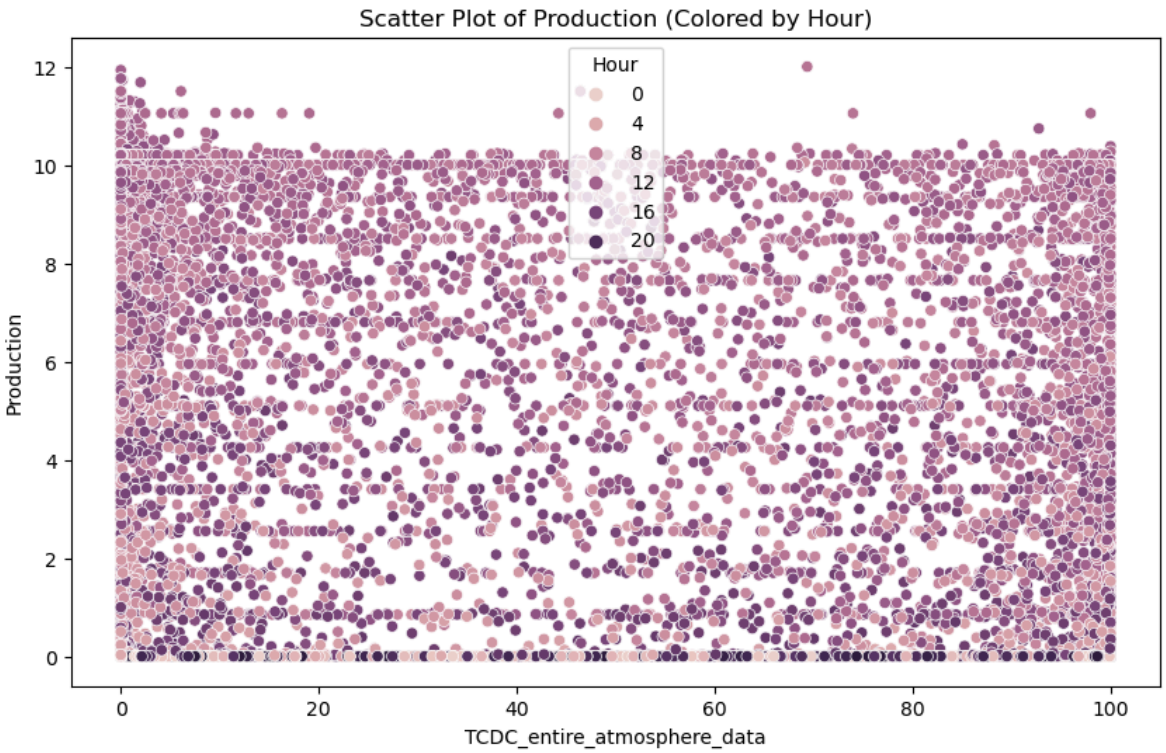

Scatter Plot of Production (Colored by Hour)
Hour
0
4
8
12
16
20
Production
TCDC_low_cloud_layer_data
0
2
4
6
8
10
12
0
20
40
60
80
100

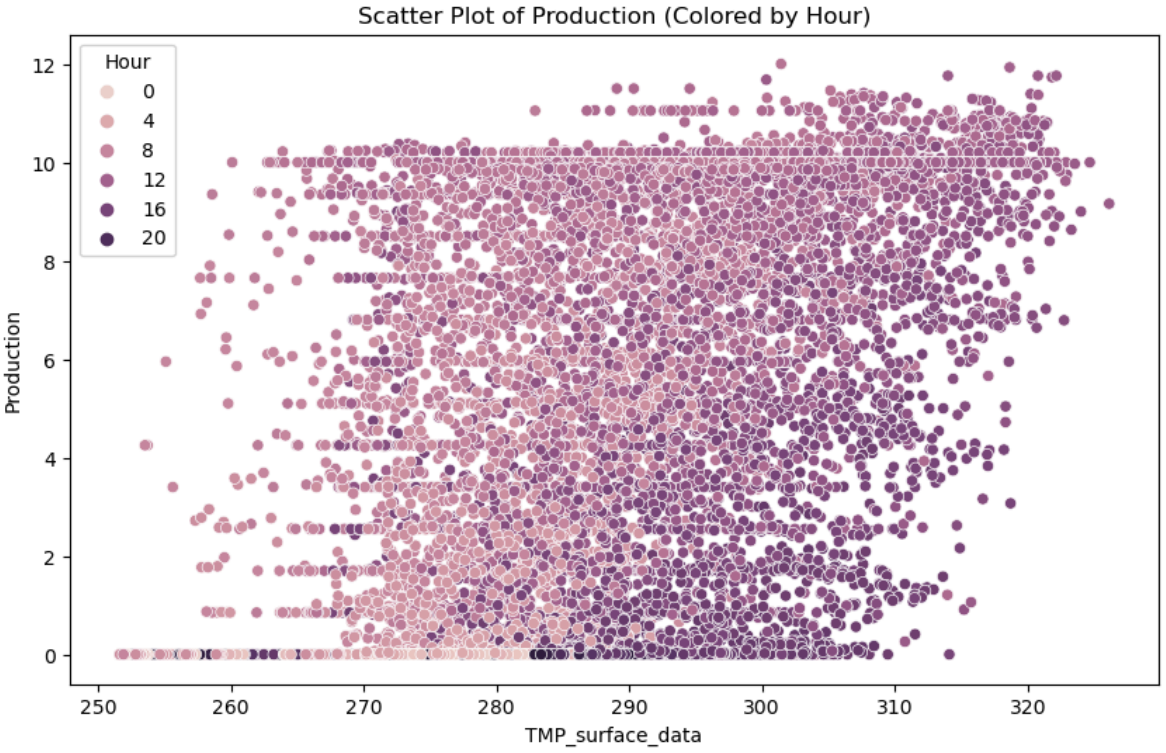

Scatter Plot of Production (Colored by Hour)
Hour
0
4
8
12
16
20
Production
USWRF_top_of_atmosphere_data
0
2
4
6
8
10
12
0
100
200
300
400
500
600
700


Scatter Plot of Production (Colored by Hour)
Hour
0
4
8
12
16
20
Production
CSNOW_surface_data
0
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0


Scatter Plot of Production (Colored by Hour)
Production
DLWRF_surface_data
Hour
0
4
8
12
16
20
0
2
4
6
8
10
12
0
100
200
300
400


Scatter Plot of Production (Colored by Hour)
Production
USWRF_surface_data
Hour
0
4
8
12
16
20
0
2
4
6
8
10
12
0
100
200
300
400
500


Scatter Plot of Production (Colored by Hour)
Hour
0
4
8
12
16
20
Production
TCDC_entire_atmosphere_data
0
20
40
60
80
100
0
2
4
6
8
10
12


Scatter Plot of Production (Colored by Hour)
Hour
0
4
8
12
16
20
Production
TMP_surface_data
250
260
270
280
290
300
310
320
0
2
4
6
8
10
12

3. Related Literature

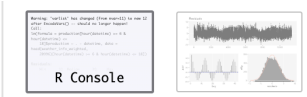
Time series analysis models (such as ARIMA and SARIMA) perform relatively good for data with seasonal patterns. However, if the output is heavily affected by the externalities, time series analysis approaches are not enough themselves and regression-based approaches are required. Solar energy production data does not have seasonalities similar to a daily gross sales data of a retail company, which has strong weekly and monthly seasonalities. Solar data has somewhat hourly seasonality. It is heavily affected by external variables, such as cloud low layer and DSWRF. For these reasons, we decided to apply linear regression for forecasting to catch effects of externalities.

Besides, it is important to understand the regressors to be able to create models and make correct interpretations. **TCDC_low.cloud.layer, TCDC_middle.cloud.layer, TCDC_high.cloud.layer, TCDC_entire.atmosphere** time series are regressors for indicating the cloud cover amounts for different type of clouds. There are few radiation time series data which are used as regressors, downward shortwave radiation **DSWRF_surface**, solar radiation for hourly mean of upward solar radiation flux at the nominal top of atmosphere **USWRF_top_of_atmosphere,** solar radiation for hourly mean of upward solar radiation flux at the surface **USWRF_surface,** and **DLWRF_surface.** Besides, categorical **CSNOW_surface** variable is used, which takes value 1 when there is snow and 0 otherwise. Temperature **TMP_surface** data is used in Kelvin.

4. Modelling Approaches

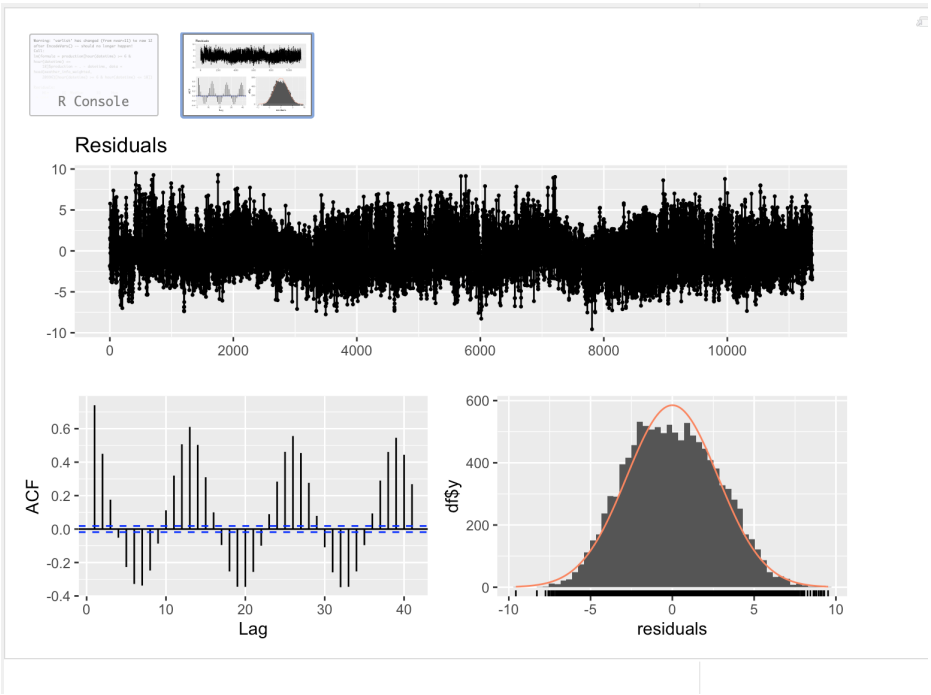
4.1.1.Linear Regression with All Weather Related Regressors

Since data has a seasonal behavior (hourly-monthly-seasonally), using weather data with time-related regressors can be good approach. Models built step by step (by adding regressors) and their performances measured based on adjusted R squared and residual plots. These models are built to forecast 06.00-18.00 because there is no significant production before 06.00 and after 18.00, the forecasts for these hours are zero. Therefore, there is no need to build a model to predict those hours.



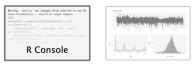

```
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -1.054e+02  1.633e+00 -64.582  < 2e-16 ***
dswrf_surface           -7.380e-03  2.625e-04 -28.111  < 2e-16 ***
tdc_low.cloud.layer      2.979e-02  1.564e-03  19.044  < 2e-16 ***
tdc_middle.cloud.layer   6.759e-03  1.199e-03   5.639 1.75e-08 ***
tdc_high.cloud.layer     4.044e-03  1.305e-03   3.098  0.00195 **
tdc_entire.atmosphere   -8.571e-04  1.594e-03  -0.538  0.59086
uswrf_top_of_atmosphere  2.763e-03  3.570e-04   7.739 1.09e-14 ***
csnow_surface           -7.117e-02  1.568e-01  -0.454  0.64983
dlwrf_surface           -8.915e-02  1.457e-03 -61.185  < 2e-16 ***
uswrf_surface            8.736e-03  7.794e-04  11.208  < 2e-16 ***
tmp_surface              4.718e-01  6.896e-03  68.412  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.795 on 11364 degrees of freedom
Multiple R-squared:  0.4742,    Adjusted R-squared:  0.4737
F-statistic:  1025 on 10 and 11364 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 14

data:  Residuals
LM test = 7763, df = 14, p-value < 2.2e-16
```

Residual plots of the only with all regressors:



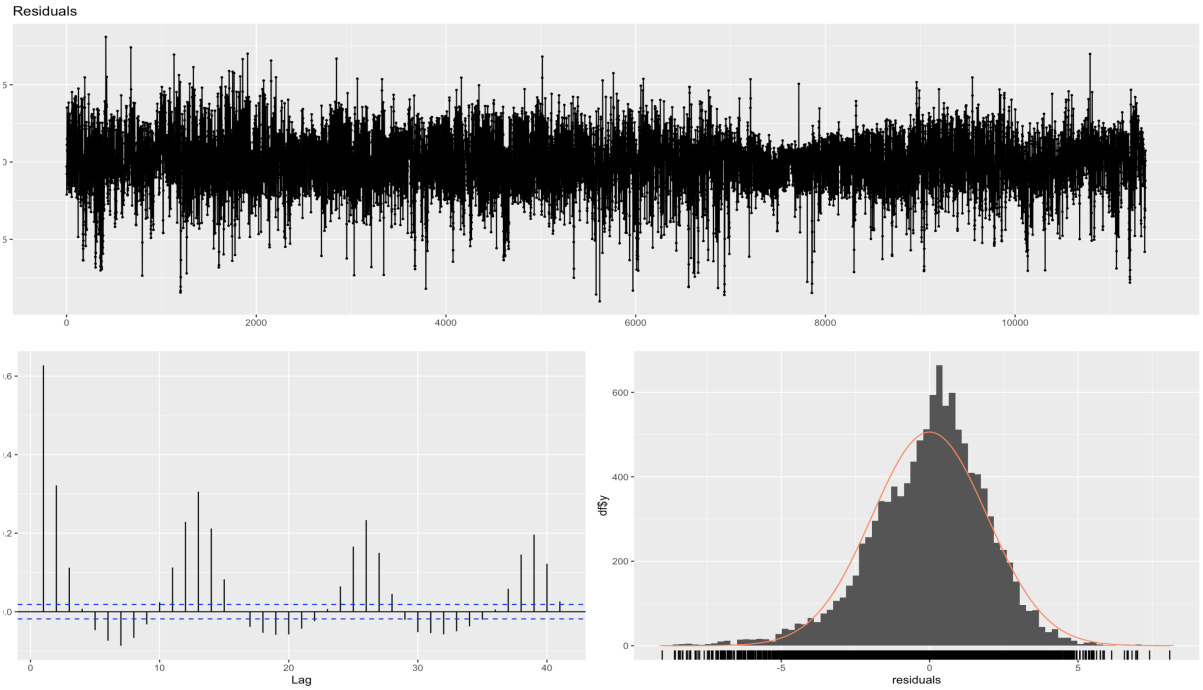
Although all the regressors are significant and are correlated with the production, these regressors are not enough to explain the model just by themselves. Residuals are highly correlated due to seasonal behavior of the data which has already been observed in introduction part. Therefore, month and hour information is added to obtain a better model. Besides, in our python code, we have obtained scatterplots for residuals vs predictors and residuals vs predicted values.

4.1.2.Model with All Weather Related Regressors + Time Related Regressors (hour and month)




```
Warning: 'varlist' has changed (from nvar=14) to new 16 after EncodeVars() -- should no longer happen!
Call:
lm(formula = production[hour(datetime) >= 6 & hour(datetime) <=
    18]$production ~ . - datetime - mon - hour + as.factor(hour) +
    as.factor(mon), data = head(model1_month_hour_data, 20996)[hour(datetime) >=
    6 & hour(datetime) <= 18])

Residuals:
    Min      1Q  Median      3Q     Max
-9.0089 -1.1848  0.1848  1.2695  8.0911

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)            -3.309e+01  1.815e+00 -18.229  < 2e-16 ***
dswrf_surface          -1.995e-03  2.838e-04  -7.029 2.19e-12 ***
tdc_low.cloud.layer    -3.126e-03  1.162e-03  -2.691 0.007139 **
tdc_middle.cloud.layer -1.306e-03  8.563e-04  -1.525 0.127283
tdc_high.cloud.layer   -2.967e-03  9.311e-04  -3.187 0.001443 **
tdc_entire.atmosphere  -3.835e-03  1.140e-03  -3.364 0.000769 ***
uswrf_top_of_atmosphere  1.788e-03  4.387e-04   4.075 4.63e-05 ***
csnow_surface          -7.493e-01  1.110e-01  -6.748 1.57e-11 ***
dlwrf_surface          -2.613e-02  1.392e-03 -18.768  < 2e-16 ***
uswrf_surface           5.852e-03  6.384e-04   9.167  < 2e-16 ***
tmp_surface             1.681e-01  7.100e-03  23.678  < 2e-16 ***
as.factor(hour)11      -3.768e-01  9.594e-02  -3.928 8.63e-05 ***
as.factor(hour)12      -7.518e-01  1.013e-01  -7.423 1.23e-13 ***
as.factor(hour)13      -1.239e+00  1.053e-01 -11.767  < 2e-16 ***
as.factor(hour)14      -2.169e+00  1.061e-01 -20.437  < 2e-16 ***
as.factor(hour)15      -3.678e+00  1.045e-01 -35.191  < 2e-16 ***
as.factor(hour)16      -5.282e+00  9.522e-02 -55.472  < 2e-16 ***
as.factor(hour)17      -6.347e+00  9.579e-02 -66.258  < 2e-16 ***
as.factor(hour)18      -6.619e+00  1.049e-01 -63.102  < 2e-16 ***
as.factor(hour)6       -5.282e+00  1.824e-01 -28.955  < 2e-16 ***
as.factor(hour)7       -3.348e+00  1.754e-01 -19.085  < 2e-16 ***
as.factor(hour)8       -1.209e+00  1.628e-01  -7.428 1.18e-13 ***
as.factor(hour)9       -2.056e-02  1.483e-01  -0.139 0.889734
as.factor(mon)Aug      -3.547e-01  1.247e-01  -2.844 0.004459 **
as.factor(mon)Dec      -5.884e-01  1.265e-01  -4.653 3.31e-06 ***
as.factor(mon)Feb       2.715e-01  1.096e-01   2.477 0.013272 *
as.factor(mon)Jan      -7.064e-01  1.210e-01  -5.839 5.41e-09 ***
as.factor(mon)Jul       3.343e-01  1.077e-01   3.102 0.001924 **
as.factor(mon)Jun       2.445e-01  1.019e-01   2.400 0.016396 *
as.factor(mon)Mar       3.755e-01  9.095e-02   4.129 3.68e-05 ***
as.factor(mon)May       1.855e-01  8.412e-02   2.205 0.027497 *
as.factor(mon)Nov      -6.881e-01  1.137e-01  -6.050 1.49e-09 ***
as.factor(mon)Oct      -3.365e-01  1.004e-01  -3.351 0.000807 ***
as.factor(mon)Sep      -1.882e-01  1.067e-01  -1.764 0.077705 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.945 on 11341 degrees of freedom
Multiple R-squared:  0.7459,    Adjusted R-squared:  0.7452
```

Residual plots of the averaged data+time related regressors (hour and month) model:

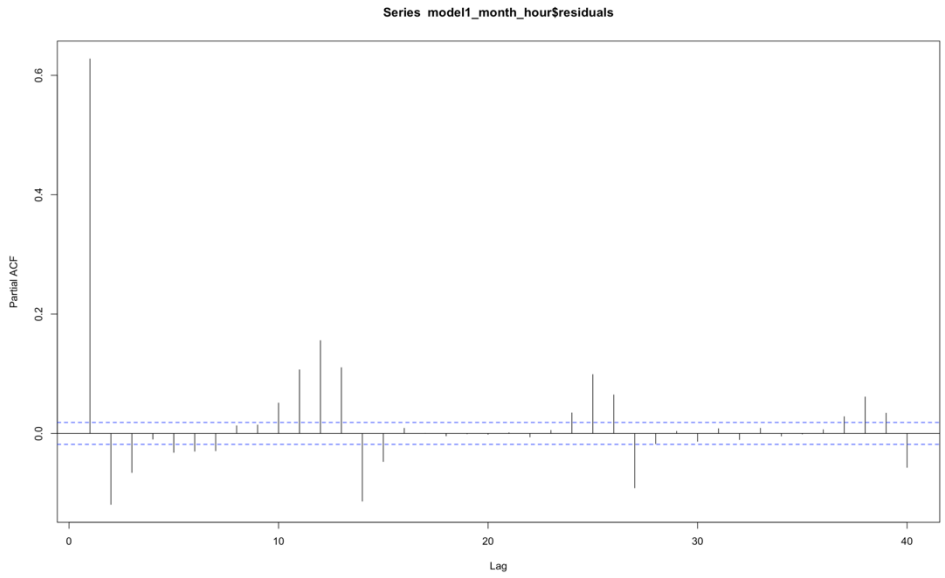


After incorporating hour and month information into the model, the adjusted R-squared value significantly increased, indicating an improved fit of the model to the data. Additionally, the autocorrelation of the residuals has decreased substantially, though it remains above acceptable limits. To further mitigate this issue, the inclusion of lagged variables is being considered.

Before proceeding with the addition of lagged variables, it is crucial to perform a Partial Autocorrelation Function (PACF) analysis on the residuals. This step is necessary because, while the Autocorrelation Function (ACF) plot shows sinusoidal behavior and high correlation across all lags, it does not provide sufficient information about the independent effects of individual lagged variables. The PACF analysis, on the other hand, will help identify the specific lags that have a significant partial correlation with the residuals, thereby indicating which lagged variables should be included in the model.

By conducting a PACF analysis, we can determine the precise lagged variables that independently influence the residuals. This insight will enable us to refine the model further, aiming to reduce residual autocorrelation to acceptable levels and enhance the overall predictive accuracy. The PACF analysis will guide the selection of appropriate lagged variables, ensuring that the model captures all relevant temporal dependencies and improves its performance in predicting electricity production.

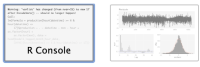
PACF plot of the residuals:



The two highest correlations are observed at lag 1 and lag 12. Since this model focuses on the time period from 06:00 to 18:00, the correlation at lag 12 corresponds to the correlation at lag 24 in the original data. The correlations at lag 1 and lag 24 were noted before the modeling process, and their origins were explained. However, post-

modeling, it is observed that these correlations persist, indicating that the regressors in this model cannot fully account for this relationship. Lag 1 and lag 24 are deemed inappropriate for prediction in the model since they would use data from two days prior (d-2). Therefore, lag 48 has been added to the model instead and an autoregressive model is constructed.

4.1.3. Model with All Weather Related Regressors + Time Related Regressors (hour and month) + Lagged Variables



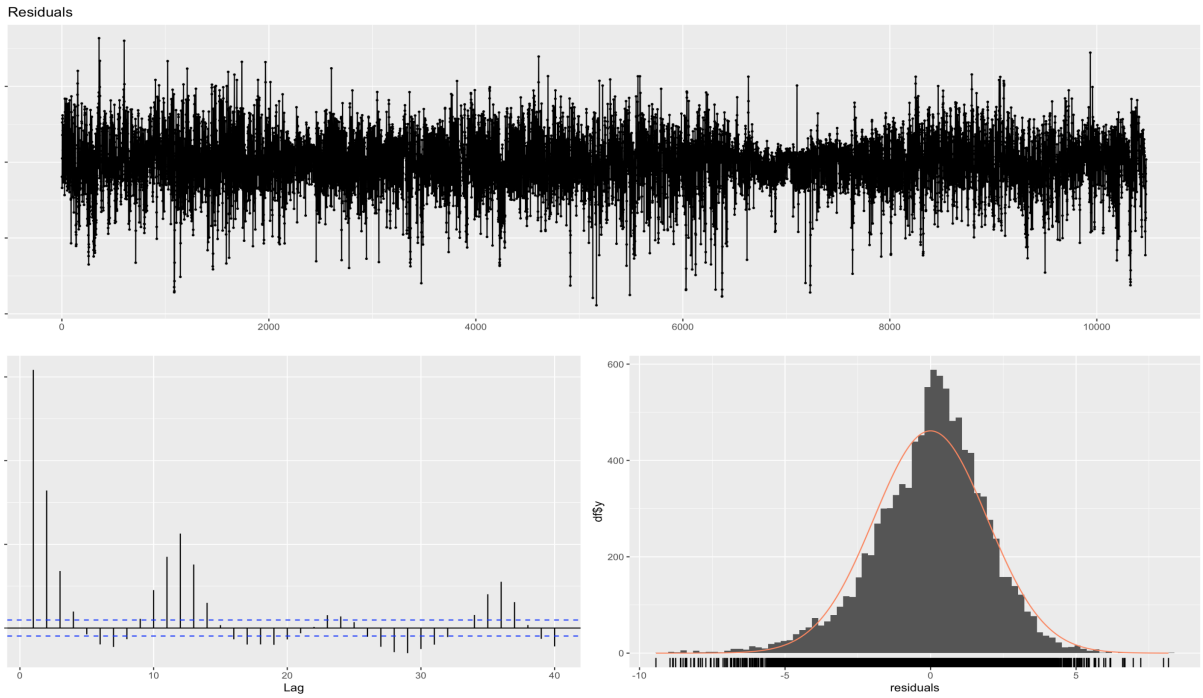

```
Warning: 'varlist' has changed (from nvar=15) to new 17 after EncodeVars() -- should no longer happen!
Call:
lm(formula = production[hour(datetime) >= 6 & hour(datetime) <=
    17]$production ~ . - datetime - mon - hour + as.factor(hour) +
    as.factor(mon), data = head(model1_lagged_month_hour_data,
    20996)[hour(datetime) >= 6 & hour(datetime) <= 17])

Residuals:
    Min      1Q  Median      3Q     Max
-9.4361 -1.1264  0.1487  1.2448  8.1746

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -2.614e+01  1.903e+00 -13.738  < 2e-16 ***
dswrf_surface           -1.050e-03  2.908e-04  -3.610 0.000308 ***
tdc_low.cloud.layer     -4.387e-03  1.236e-03  -3.550 0.000387 ***
tdc_middle.cloud.layer  -2.108e-03  8.988e-04  -2.345 0.019038 *
tdc_high.cloud.layer    -2.795e-03  9.909e-04  -2.821 0.004796 **
tdc_entire.atmosphere   -4.800e-03  1.201e-03  -3.997 6.46e-05 ***
uswrf_top_of_atmosphere  2.131e-03  4.604e-04   4.628 3.74e-06 ***
csnow_surface           -8.713e-01  1.160e-01  -7.511 6.33e-14 ***
dlwrf_surface           -2.409e-02  1.456e-03 -16.549  < 2e-16 ***
uswrf_surface            4.732e-03  6.519e-04   7.259 4.17e-13 ***
tmp_surface              1.364e-01  7.458e-03  18.290  < 2e-16 ***
lag_48                   1.729e-01  8.495e-03  20.351  < 2e-16 ***
as.factor(hour)11       -3.476e-01  9.649e-02  -3.602 0.000317 ***
as.factor(hour)12       -6.852e-01  1.024e-01  -6.695 2.27e-11 ***
as.factor(hour)13       -1.104e+00  1.069e-01 -10.324  < 2e-16 ***
as.factor(hour)14       -1.885e+00  1.089e-01 -17.304  < 2e-16 ***
as.factor(hour)15       -3.122e+00  1.104e-01 -28.286  < 2e-16 ***
as.factor(hour)16       -4.318e+00  1.076e-01 -40.148  < 2e-16 ***
as.factor(hour)17       -5.155e+00  1.130e-01 -45.636  < 2e-16 ***
as.factor(hour)6        -4.036e+00  1.947e-01 -20.728  < 2e-16 ***
as.factor(hour)7        -2.436e+00  1.839e-01 -13.248  < 2e-16 ***
as.factor(hour)8        -6.813e-01  1.682e-01  -4.051 5.13e-05 ***
as.factor(hour)9         2.769e-01  1.522e-01   1.819 0.068915 .
as.factor(mon)Aug       -2.567e-01  1.302e-01  -1.972 0.048653 *
as.factor(mon)Dec       -4.661e-01  1.306e-01  -3.570 0.000358 ***
as.factor(mon)Feb        2.278e-01  1.142e-01   1.996 0.046010 *
as.factor(mon)Jan       -6.702e-01  1.254e-01  -5.342 9.36e-08 ***
as.factor(mon)Jul        2.830e-01  1.120e-01   2.527 0.011526 *
as.factor(mon)Jun        2.122e-01  1.063e-01   1.996 0.045980 *
as.factor(mon)Mar        3.695e-01  9.488e-02   3.894 9.90e-05 ***
as.factor(mon)May        1.581e-01  8.764e-02   1.804 0.071327 .
as.factor(mon)Nov       -5.851e-01  1.170e-01  -4.999 5.85e-07 ***
as.factor(mon)Oct       -2.754e-01  1.036e-01  -2.657 0.007890 **
as.factor(mon)Sep       -1.441e-01  1.111e-01  -1.296 0.194864
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.948 on 10442 degrees of freedom
  (24 observations deleted due to missingness)
```

```
Residual standard error: 1.948 on 10442 degrees of freedom
  (24 observations deleted due to missingness)
Multiple R-squared:  0.7313,    Adjusted R-squared:  0.7304
F-statistic: 861.2 on 33 and 10442 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 37

data:  Residuals
LM test = 4414.4, df = 37, p-value < 2.2e-16
```

The purpose of adding lagged variables to the model, which already includes all weather-related regressors and time-related regressors (hour and month), is to address any remaining autocorrelation in the residuals and improve the overall predictive accuracy of the model. The lag 48 variable is highly significant ($p < 0.001$) with a positive coefficient of 0.1729,

indicating that the value of electricity production 48 hours ago has a significant positive impact on the current value.

Residual plots of Model with averaged data + time related regressors (hour and month) + lagged variable:



Although the adjusted R-squared value remained the same after the addition of lagged variables, the correlation between residuals decreased. However, there is still significant autocorrelation present in the residuals. Additionally, the residuals do not exhibit constant variance, as evidenced by a notable increase in variance after lag 30.
In summary, while the inclusion of lagged variables has improved the model by reducing some degree of autocorrelation, it has not fully resolved all the issues. The residuals still show significant autocorrelation, indicating that the model is not completely capturing the time-dependent structure of the data. Moreover, the presence of heteroscedasticity (non-constant variance of residuals) suggests that the model assumptions are not entirely met, potentially leading to inefficiency in the estimates and unreliable predictions.

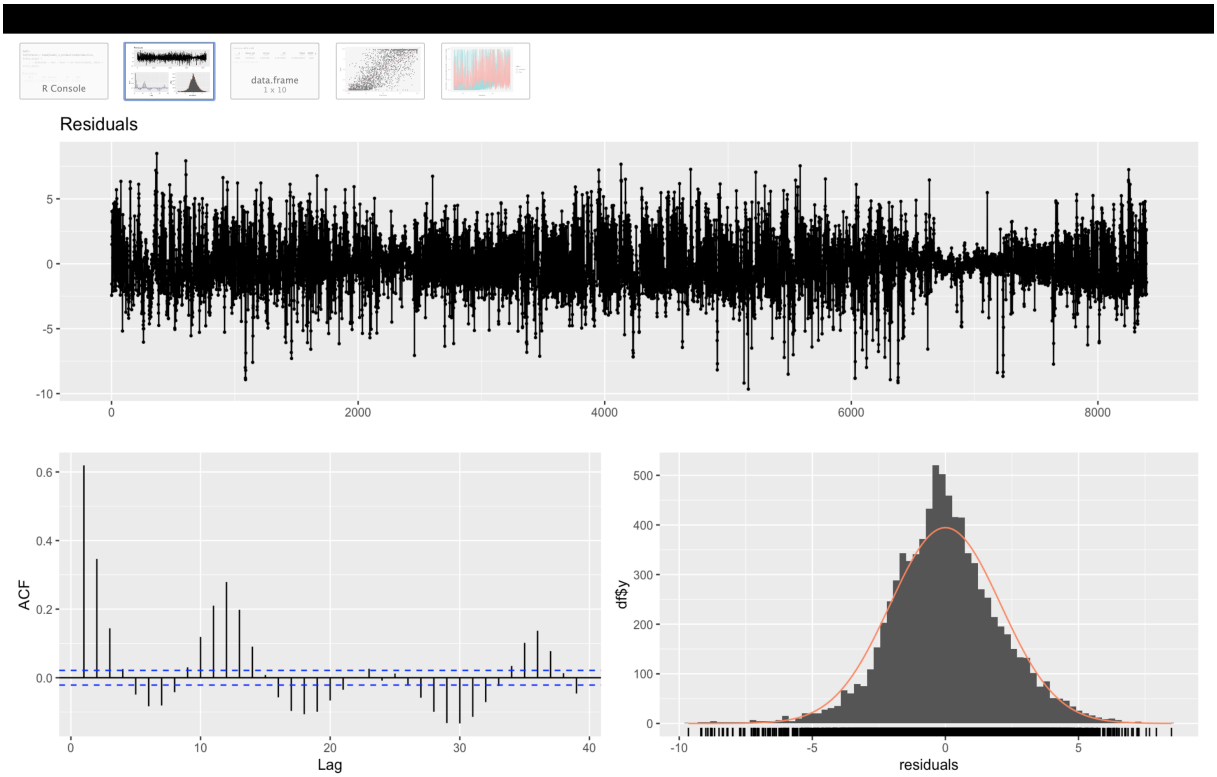
4.1.4. Model with All Weather Related Regressors + Time Related Regressors (hour and month) + Lagged Variables + Hourly Interaction Term for DSWRF

Creating an interaction term for each hour involves multiplying the dswrf_surface (direct shortwave radiation flux at the surface) variable by the corresponding hour of the day. This approach allows the model to capture the variation of solar radiation throughout the day more accurately.

The primary motivation behind adding this interaction term is the recognition that the impact of solar radiation on electricity production varies depending on the time of day. By incorporating the hourly interaction terms, the model can better account for the changes in solar radiation's effect at different times, leading to a more precise and nuanced representation of the relationships between the independent variables.
The interaction term dswrf_surface * hour allows the model to reflect these hourly changes more accurately, recognizing that the influence of solar radiation on electricity production is not constant throughout the day.

```
Call:
lm(formula = head(model_x_production$production, train_size) ~
    . - datetime - mon - hour + as.factor(mon), data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-9.6546 -1.2960 -0.0721  1.2126  8.4922

Coefficients: (1 not defined because of singularities)
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               -3.372e+01  2.580e+00 -13.073  < 2e-16 ***
dswrf_surface             -5.978e-03  4.657e-04 -12.837  < 2e-16 ***
tdc_low.cloud.layer        2.058e-03  1.558e-03   1.320 0.186727
tdc_middle.cloud.layer    -2.181e-03  1.129e-03  -1.932 0.053419 .
tdc_high.cloud.layer      -1.020e-03  1.224e-03  -0.834 0.404483
tdc_entire.atmosphere     -3.775e-03  1.471e-03  -2.566 0.010317 *
uswrf_top_of_atmosphere    1.650e-03  3.274e-04   5.039 4.77e-07 ***
csnow_surface             -6.672e-01  1.437e-01  -4.644 3.47e-06 ***
dlwrf_surface             -3.131e-02  1.941e-03 -16.133  < 2e-16 ***
uswrf_surface              3.195e-03  8.251e-04   3.872 0.000109 ***
tmp_surface                1.585e-01  1.071e-02  14.796  < 2e-16 ***
lag_48                     3.654e-01  8.981e-03  40.688  < 2e-16 ***
hour_6_interaction_dswrf  -3.302e-01  5.115e-02  -6.456 1.13e-10 ***
hour_7_interaction_dswrf   6.954e-03  6.121e-03   1.136 0.255950
hour_8_interaction_dswrf   1.822e-02  2.439e-03   7.469 8.88e-14 ***
hour_9_interaction_dswrf   1.618e-02  1.455e-03  11.116  < 2e-16 ***
hour_10_interaction_dswrf  7.450e-03  2.715e-04  27.445  < 2e-16 ***
hour_11_interaction_dswrf  6.582e-03  2.597e-04  25.348  < 2e-16 ***
hour_12_interaction_dswrf  6.047e-03  2.522e-04  23.976  < 2e-16 ***
hour_13_interaction_dswrf  5.623e-03  2.452e-04  22.931  < 2e-16 ***
hour_14_interaction_dswrf  4.888e-03  2.372e-04  20.602  < 2e-16 ***
hour_15_interaction_dswrf  3.820e-03  2.299e-04  16.620  < 2e-16 ***
hour_16_interaction_dswrf  2.045e-03  2.382e-04   8.586  < 2e-16 ***
hour_17_interaction_dswrf         NA         NA      NA       NA
as.factor(mon)Aug         -1.701e-01  1.617e-01  -1.052 0.292821
as.factor(mon)Dec         -1.576e-01  1.408e-01  -1.119 0.263012
as.factor(mon)Feb          8.741e-01  1.353e-01   6.459 1.11e-10 ***
as.factor(mon)Jan         -1.899e-01  1.293e-01  -1.468 0.142015
as.factor(mon)Jul          3.087e-01  1.344e-01   2.298 0.021611 *
as.factor(mon)Jun          4.288e-01  1.370e-01   3.130 0.001752 **
as.factor(mon)Mar          7.080e-01  1.211e-01   5.846 5.22e-09 ***
as.factor(mon)May          2.049e-01  1.164e-01   1.761 0.078334 .
as.factor(mon)Nov         -1.765e-01  1.222e-01  -1.445 0.148560
as.factor(mon)Oct         -8.913e-02  1.266e-01  -0.704 0.481512
as.factor(mon)Sep         -1.053e-01  1.427e-01  -0.738 0.460607
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.086 on 8361 degrees of freedom
  (24 observations deleted due to missingness)
Multiple R-squared:  0.6979,    Adjusted R-squared:  0.6967
```

Residual plots of Model with averaged data + time related regressors (hour and month) + lagged variable:

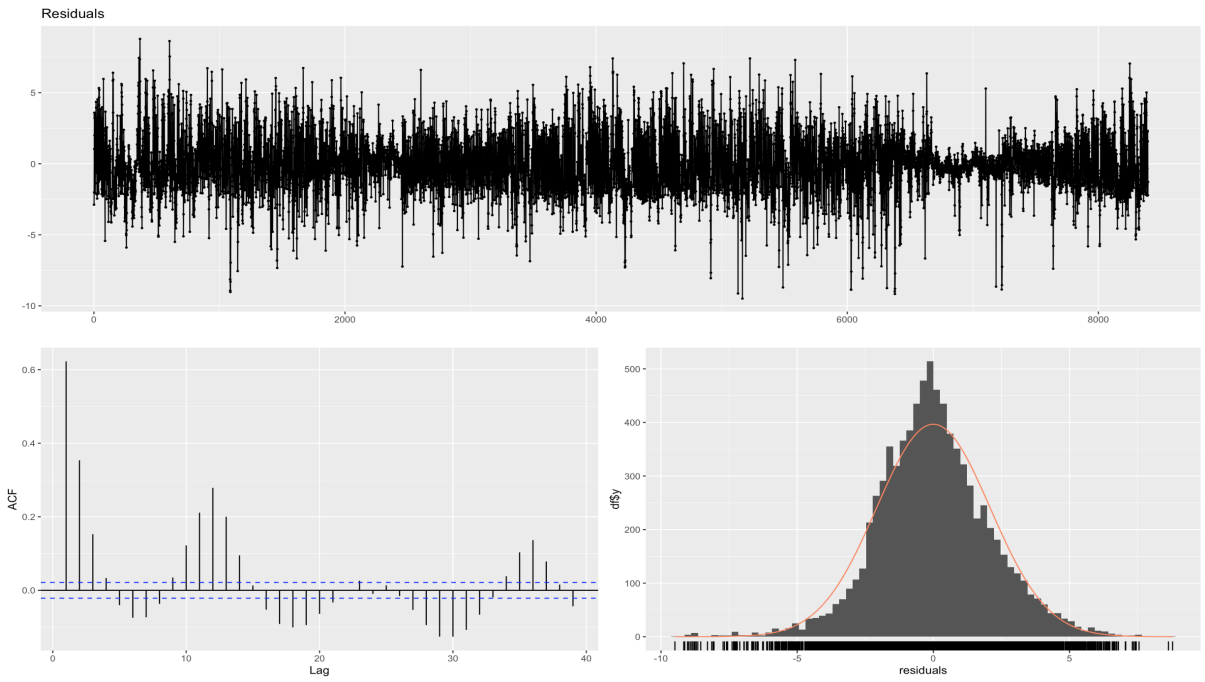


Adding hourly DSWRF interaction terms didn't improve the model. Additionally, due to monthly seasonality, the DSWRF value is directly influenced by the month and hour of the day. To capture this interaction, a monthly DSWRF interaction term was added to the model. The logic behind this change lies in enhancing the model's ability to capture the variability of solar radiation on both daily and monthly scales. Solar radiation levels fluctuate based on the time of day and the month of the year. For instance, during summer months, there might be higher solar irradiance compared to winter months.

4.1.5. Model with All Weather Related Regressors + Time Related Regressors (hour and month) + Lagged Variables + Hourly Interaction Term for DSWRF + Monthly Interaction Term for DSWRF

```
Call:
lm(formula = head(model_x1_production$production, train_size) ~
    . - datetime - mon - hour, data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-9.4882 -1.3072 -0.0748  1.2209  8.7837

Coefficients: (2 not defined because of singularities)
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -3.186e+01  2.448e+00 -13.019  < 2e-16 ***
dswrf_surface               -7.140e-03  6.697e-04 -10.662  < 2e-16 ***
tdc_low.cloud.layer          3.227e-03  1.564e-03   2.063 0.039139 *
tdc_middle.cloud.layer      -1.736e-03  1.123e-03  -1.545 0.122292
tdc_high.cloud.layer        -7.495e-04  1.205e-03  -0.622 0.533834
tdc_entire.atmosphere       -2.721e-03  1.466e-03  -1.856 0.063503 .
uswrf_top_of_atmosphere      2.135e-03  3.425e-04   6.234 4.76e-10 ***
csnow_surface               -6.428e-01  1.433e-01  -4.487 7.33e-06 ***
dlwrf_surface               -3.442e-02  1.959e-03 -17.572  < 2e-16 ***
uswrf_surface                1.840e-03  8.473e-04   2.172 0.029915 *
tmp_surface                  1.548e-01  1.037e-02  14.929  < 2e-16 ***
lag_48                       3.660e-01  8.953e-03  40.883  < 2e-16 ***
hour_6_interaction_dswrf    -2.387e-01  4.373e-02  -5.459 4.92e-08 ***
hour_7_interaction_dswrf     1.594e-02  5.069e-03   3.146 0.001664 **
hour_8_interaction_dswrf     2.093e-02  2.070e-03  10.110  < 2e-16 ***
hour_9_interaction_dswrf     1.741e-02  1.286e-03  13.536  < 2e-16 ***
hour_10_interaction_dswrf    7.399e-03  2.719e-04  27.211  < 2e-16 ***
hour_11_interaction_dswrf    6.532e-03  2.611e-04  25.014  < 2e-16 ***
hour_12_interaction_dswrf    5.996e-03  2.540e-04  23.611  < 2e-16 ***
hour_13_interaction_dswrf    5.570e-03  2.468e-04  22.570  < 2e-16 ***
hour_14_interaction_dswrf    4.831e-03  2.383e-04  20.274  < 2e-16 ***
hour_15_interaction_dswrf    3.758e-03  2.299e-04  16.343  < 2e-16 ***
hour_16_interaction_dswrf    2.020e-03  2.387e-04   8.464  < 2e-16 ***
hour_17_interaction_dswrf           NA         NA      NA       NA
month_Jan_interaction_dswrf -8.371e-05  5.617e-04  -0.149 0.881537
month_Feb_interaction_dswrf  2.994e-03  5.552e-04   5.392 7.17e-08 ***
month_Mar_interaction_dswrf  2.390e-03  5.158e-04   4.634 3.64e-06 ***
month_Apr_interaction_dswrf  1.342e-03  4.998e-04   2.685 0.007269 **
month_May_interaction_dswrf  1.775e-03  5.052e-04   3.514 0.000443 ***
month_Jun_interaction_dswrf  2.111e-03  5.172e-04   4.082 4.51e-05 ***
month_Jul_interaction_dswrf  2.204e-03  5.066e-04   4.351 1.37e-05 ***
month_Aug_interaction_dswrf  1.856e-03  5.142e-04   3.610 0.000308 ***
month_Sep_interaction_dswrf  1.385e-03  5.083e-04   2.724 0.006456 **
month_Oct_interaction_dswrf  7.726e-04  5.099e-04   1.515 0.129803
month_Nov_interaction_dswrf  3.509e-04  5.382e-04   0.652 0.514422
month_Dec_interaction_dswrf         NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.089 on 8361 degrees of freedom
  (24 observations deleted due to missingness)
```

```
Residual standard error: 2.089 on 8361 degrees of freedom
  (24 observations deleted due to missingness)
Multiple R-squared:  0.697,     Adjusted R-squared:  0.6958
F-statistic: 582.7 on 33 and 8361 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 39

data:  Residuals
LM test = 3858.6, df = 39, p-value < 2.2e-16
```



Neither the adjusted R square nor the ACF plot showed any improvement in this model. Therefore, we decided to abandon further development in the aggregate model and focus on working with hourly models. This approach allows us to capture trends and patterns at specific hours in more detail.

Till this point, we have reported different models to understand the effects of different approaches. Now we will be proceeding with our detailly evaluated and compared models.

4.2. Clustered Models

One of the other approaches that we considered through our project is to define 4 different clusters and work on them separately. Our first cluster was consisting of hours between 19:00 to 03:00, where the production values always be zero, so we defined zero vector for them. For our second cluster, with hours 04:00, 05:00, and 18:00; we applied simple moving average by determining the window lengths for all the three hours separately. Our third cluster was

consisting of the sum of the data between 06:00-08:00 and 15:00-17:00 hours. We have defined third cluster with this hour since there can be observed considerable amount of production, but the total amount of the production never reaches 10, which is the upper bound. We have worked on the summed daily data for cluster three and tried to disaggregate it to hours. Our last cluster was consisting of the summed daily data of hours from 09:00 to 14:00. It can be observed that between these hours, the threshold production level of 10 is considerably reached, therefore we have defined this cluster separate then the third cluster.

The provided code below is to determine the best window lengths for our cluster 2, simple moving average:

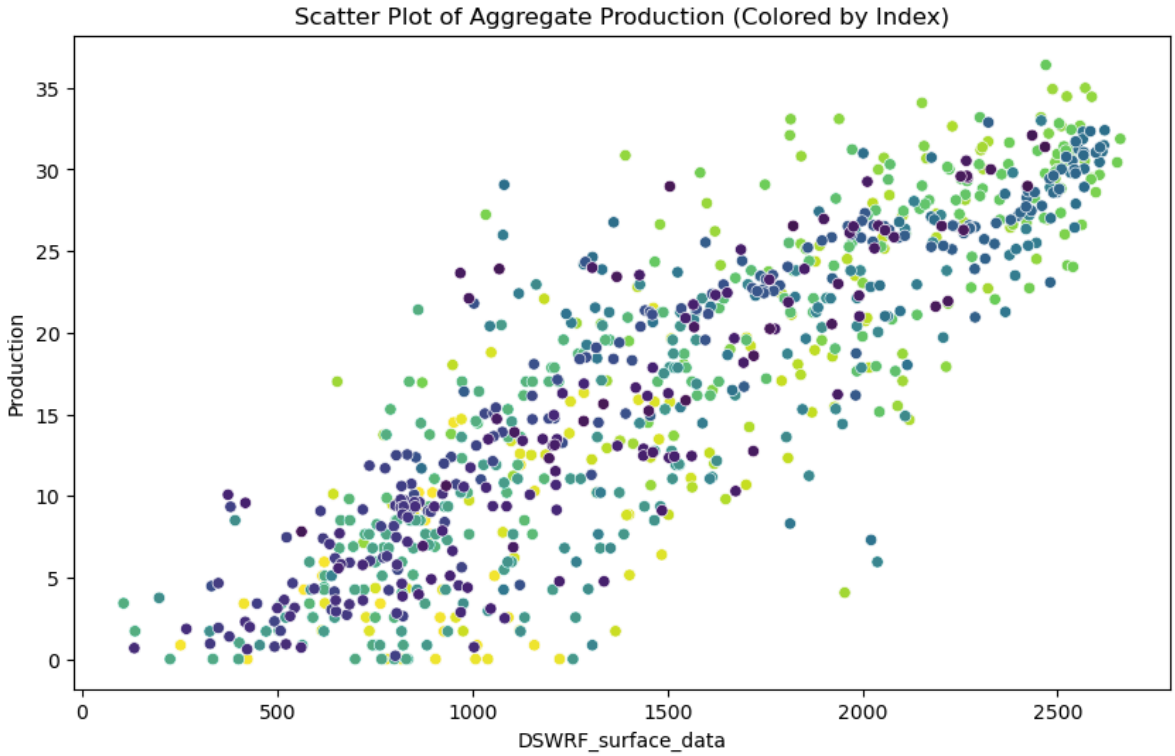
```{r}
production_sma_hour4 <- production %>%
  filter(hour(datetime) == 4)
production_sma_hour4$production_lag2 <- lag(production_sma_hour4$production, n = 2)

production_sma_hour5 <- production %>%
  filter(hour(datetime) == 5)
production_sma_hour5$production_lag2 <- lag(production_sma_hour5$production, n = 2)

production_sma_hour18 <- production %>%
  filter(hour(datetime) == 18)
production_sma_hour18$production_lag2 <- lag(production_sma_hour18$production, n = 2)

moving_average <- function(x, n) {
  sapply(seq_along(x), function(i) {
    if (i < n) {
      NA
    } else {
      mean(x[(i-n+1):i])
    }
  })
}

wmape <- function(x, n, f) {
  ma <- moving_average(x, n)
  mean(abs(f - ma) / mean(f, na.rm = TRUE), na.rm = TRUE)
}

best_window_length <- function(x, f, max_window_size = 10) {
  n_values <- seq_len(max_window_size)
  wmape_values <- sapply(n_values, function(n) wmape(x, n, f))
  print(wmape_values)
  which.min(wmape_values)
```

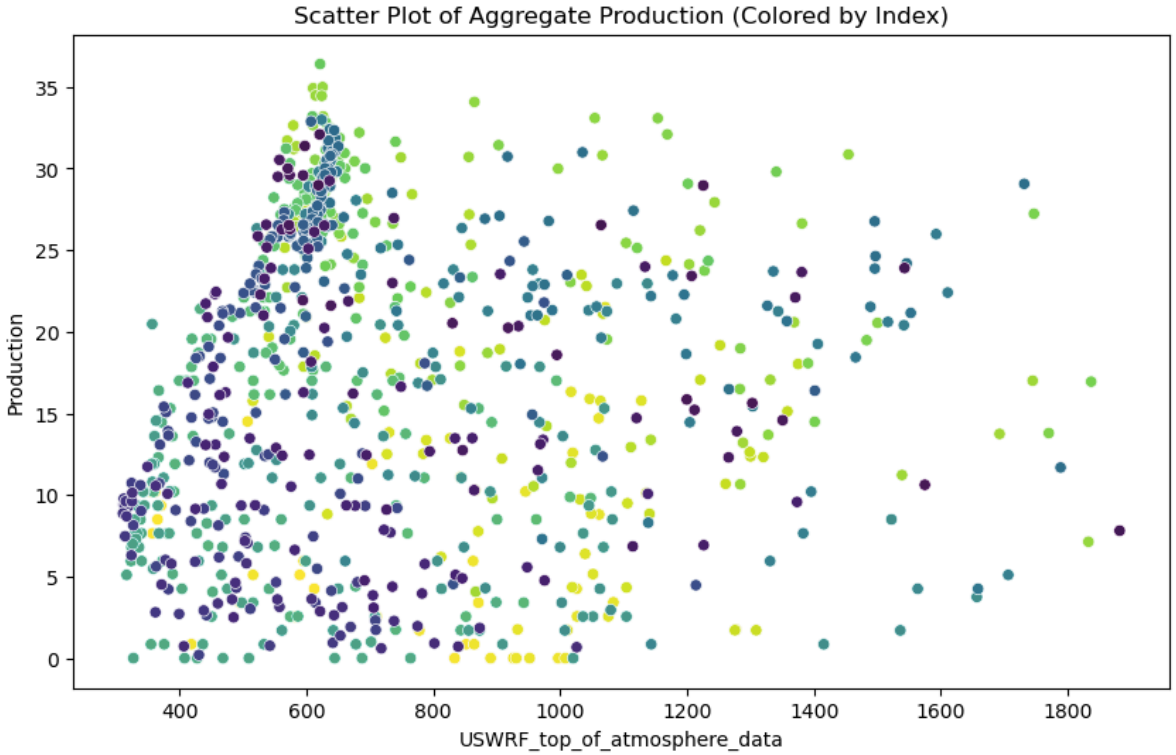
```
}

best_window_length(production_sma_hour4$production,
production_sma_hour4$production_lag2)
best_window_length(production_sma_hour5$production,
production_sma_hour5$production_lag2)
best_window_length(production_sma_hour18$production,
production_sma_hour18$production_lag2)
```

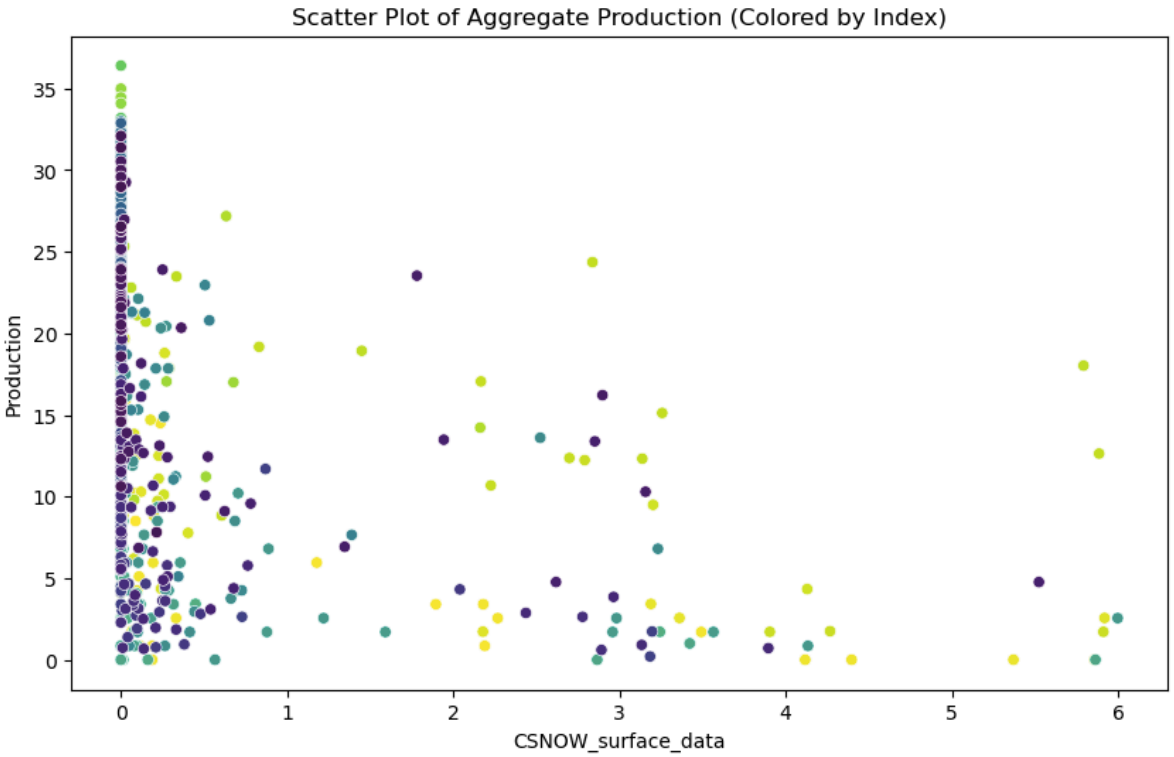
Once we determine the window lengths, we have forecasted simple moving average values with the following code:

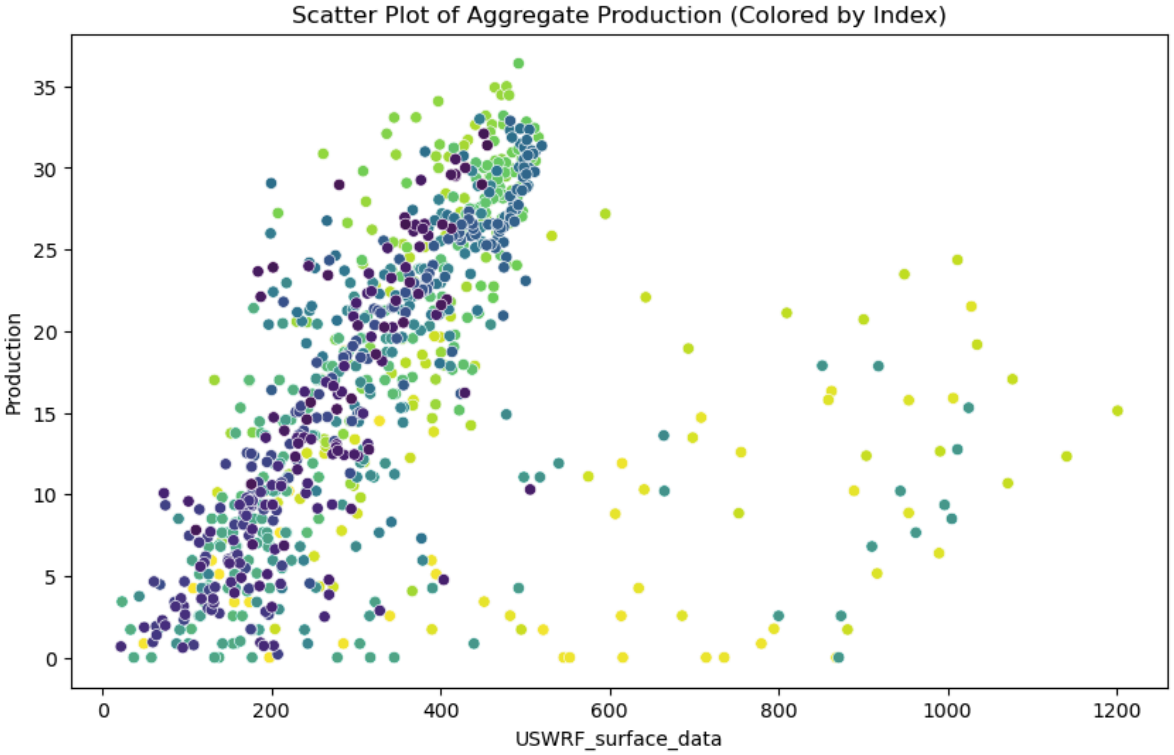
```{r}
moving_average <- function(x, n) {
  sapply(seq_along(x), function(i) {
    if (i < n) {
      NA
    } else {
      mean(x[(i-n+1):i])
    }
  })
}
production_sma_hour4$sma_forecasts <-
moving_average(production_sma_hour4$production_lag2, 3)
production_sma_hour5$sma_forecasts <-
moving_average(production_sma_hour5$production_lag2, 3)
production_sma_hour18$sma_forecasts <-
moving_average(production_sma_hour18$production_lag2, 3)
production_sma_hour4 <- production_sma_hour4 %>%
  filter(as.Date(datetime) >= as.Date('2024-02-01'))
production_sma_hour5 <- production_sma_hour5 %>%
  filter(as.Date(datetime) >= as.Date('2024-02-01'))
production_sma_hour18 <- production_sma_hour18 %>%
  filter(as.Date(datetime) >= as.Date('2024-02-01'))
```

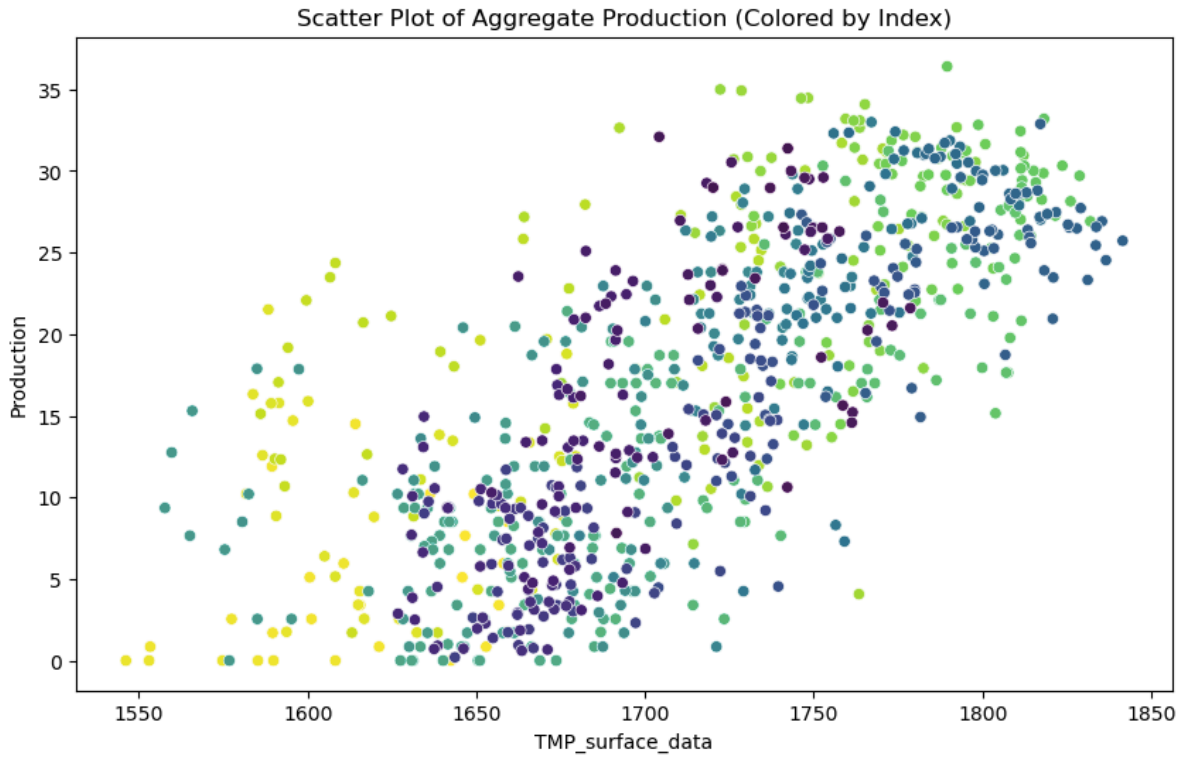
For cluster 3, the production vs regressors scatter plots where we hued according to the dates are as given below. We also tried to sum all cloud layer data in this section since we obtained better R-squared results compared to the individual values. It can be interpreted that between some production amounts, and regressors, high autocorrelation and linear relationships could be observed which means that using summed production and summed regressors data would improve our model results.

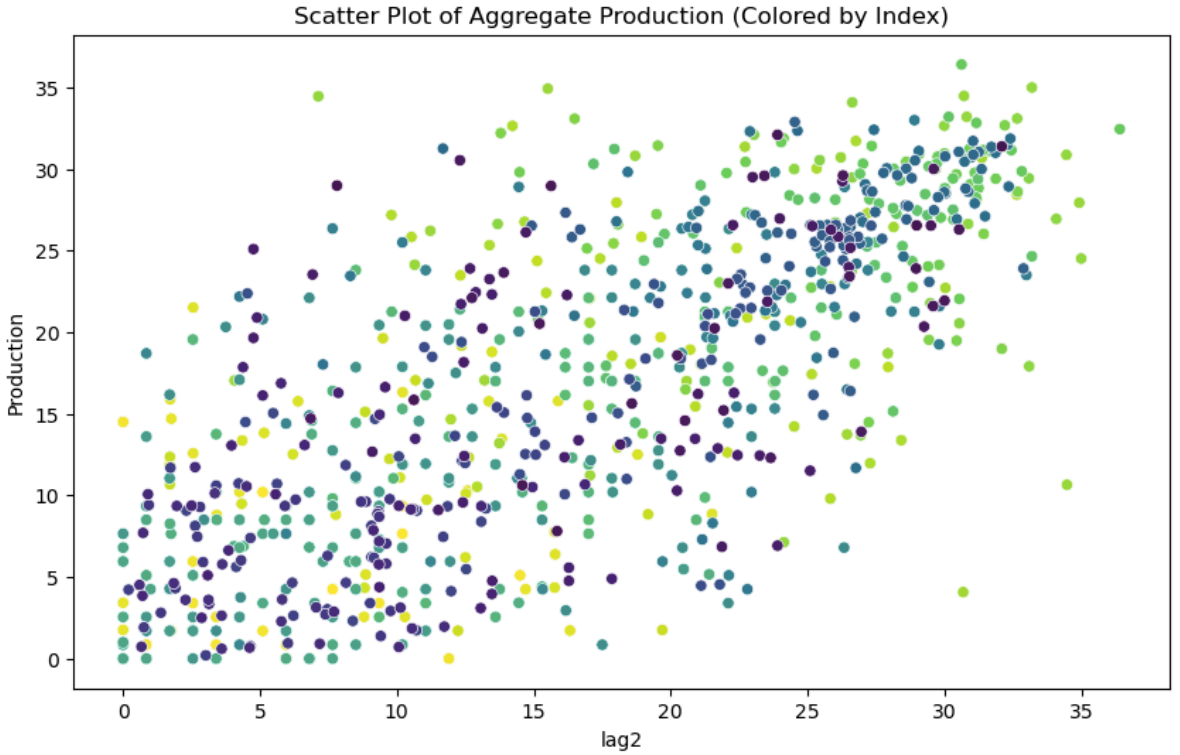

Scatter Plot of Aggregate Production (Colored by Index)
Production
DSWRF_surface_data
0
5
10
15
20
25
30
35
0
500
1000
1500
2000
2500

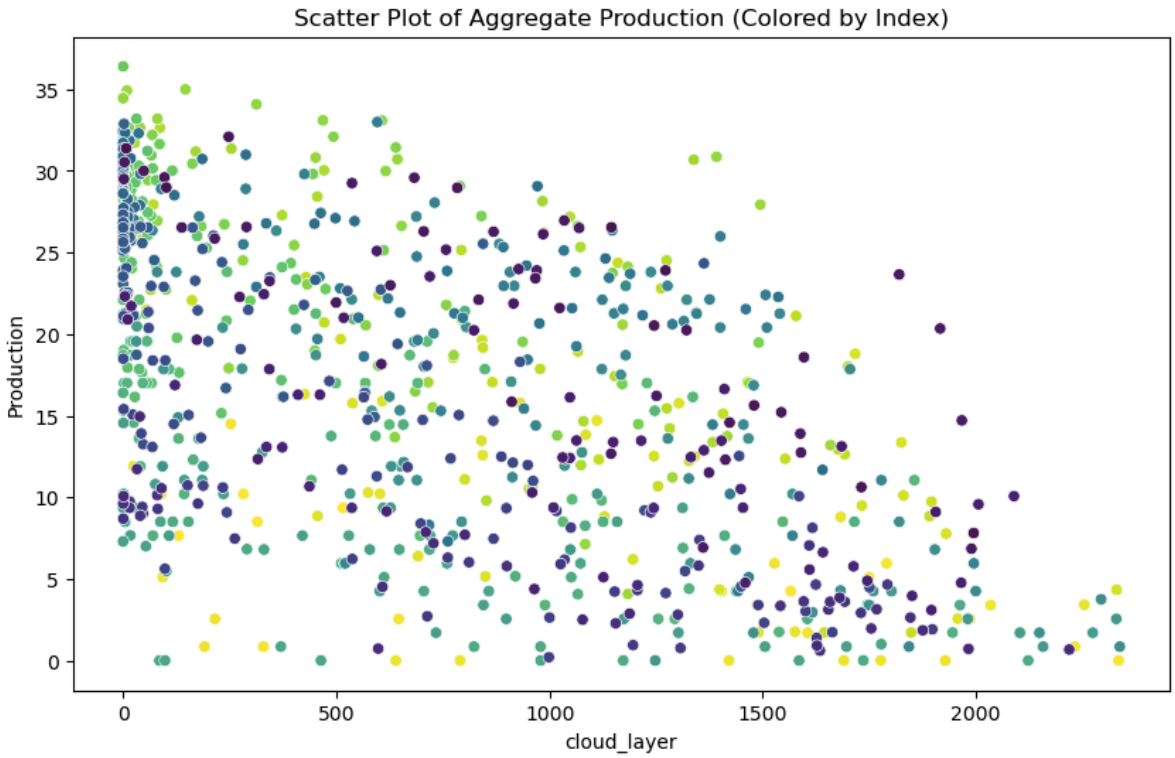

Scatter Plot of Aggregate Production (Colored by Index)
Production
USWRF_top_of_atmosphere_data
0
5
10
15
20
25
30
35
400
600
800
1000
1200
1400
1600
1800


Scatter Plot of Aggregate Production (Colored by Index)
Production
CSNOW_surface_data
0
5
10
15
20
25
30
35
0
1
2
3
4
5
6


Scatter Plot of Aggregate Production (Colored by Index)
Production
USWRF_surface_data
0
5
10
15
20
25
30
35
0
200
400
600
800
1000
1200

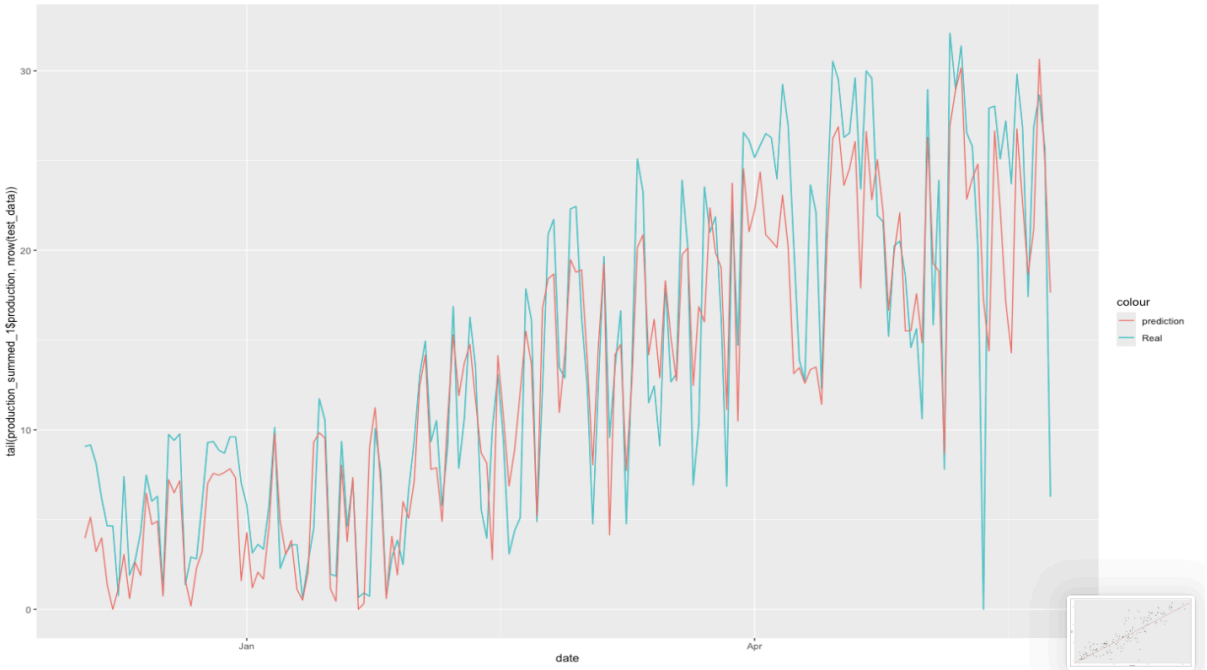

Scatter Plot of Aggregate Production (Colored by Index)
Production
TMP_surface_data
0
5
10
15
20
25
30
35
1550
1600
1650
1700
1750
1800
1850

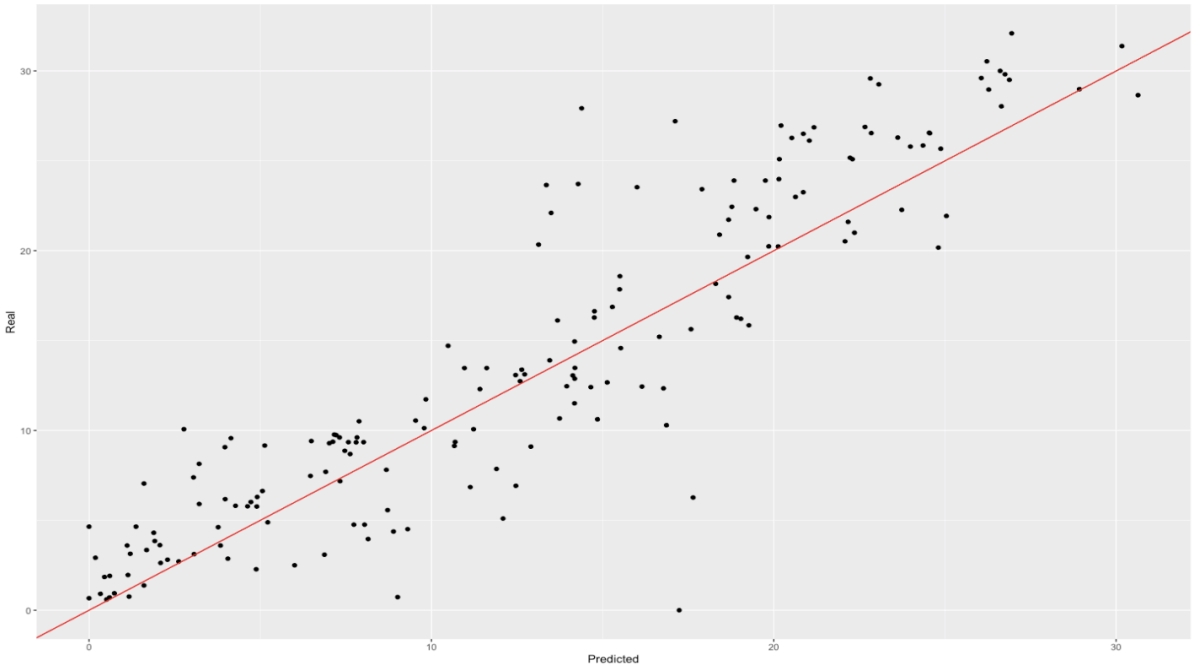

Scatter Plot of Aggregate Production (Colored by Index)
Production
lag2
0
5
10
15
20
25
30
35

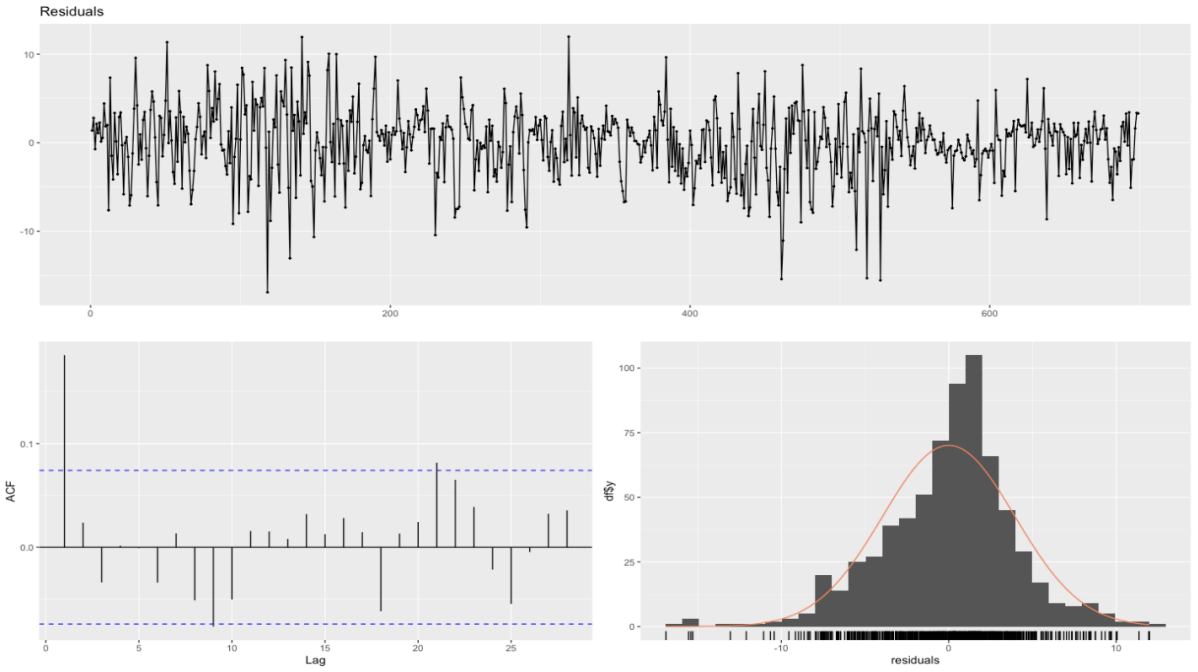


For summed cluster 3 model, one of our regression models where we used monthly dummy variables multiplied with dswrf_surface values and the results are as given below. It can be observed that the distribution of residuals does not show any reliable problem and the real vs predicted data shows close enough to linear trend. Besides, the model evaluation values such as WMAPE, RMSE, MAD, and Adjusted R-squared are 0.2104, 3.8799, 2.899, and 0.8211 respectively, which the model evaluation metrics are better compared to the previous models.

```
Call:
lm(formula = head(production_summed_1$production, train_size) ~
    . - lag_48 - date - mon - dswrf_surface, data = train_data)

Residuals:
     Min       1Q   Median       3Q      Max
-16.8911  -2.1751   0.3723   2.3088  11.9802

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                -79.371621  30.589354  -2.595 0.009671 **
tdc_low.cloud.layer         -0.004008   0.002331  -1.719 0.086049 .
tdc_middle.cloud.layer      -0.002156   0.001733  -1.244 0.213906
tdc_high.cloud.layer        -0.002508   0.001775  -1.412 0.158286
tdc_entire.atmosphere       -0.004746   0.002194  -2.164 0.030836 *
uswrf_top_of_atmosphere      0.005360   0.001168   4.588 5.34e-06 ***
csnow_surface               -0.016316   0.213774  -0.076 0.939185
dlwrf_surface               -0.007040   0.003348  -2.103 0.035867 *
uswrf_surface               -0.003177   0.002558  -1.242 0.214704
tmp_surface                  0.056576   0.021254   2.662 0.007954 **
month_Jan_interaction_dswrf  0.004353   0.002325   1.872 0.061579 .
month_Feb_interaction_dswrf  0.007377   0.001861   3.963 8.18e-05 ***
month_Mar_interaction_dswrf  0.008242   0.001568   5.255 1.98e-07 ***
month_Apr_interaction_dswrf  0.007561   0.001409   5.367 1.10e-07 ***
month_May_interaction_dswrf  0.008377   0.001371   6.108 1.70e-09 ***
month_Jun_interaction_dswrf  0.007813   0.001397   5.593 3.24e-08 ***
month_Jul_interaction_dswrf  0.007655   0.001501   5.098 4.45e-07 ***
month_Aug_interaction_dswrf  0.007126   0.001712   4.162 3.55e-05 ***
month_Sep_interaction_dswrf  0.007537   0.001821   4.138 3.94e-05 ***
month_Oct_interaction_dswrf  0.007910   0.002111   3.748 0.000194 ***
month_Nov_interaction_dswrf  0.006327   0.002585   2.447 0.014640 *
month_Dec_interaction_dswrf  0.002539   0.002903   0.875 0.382152
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.023 on 677 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.8265,    Adjusted R-squared:  0.8211
F-statistic: 153.6 on 21 and 677 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 25

data:  Residuals
LM test = 47.961, df = 25, p-value = 0.00377
```


tail(production_summed_1$production, nrow(test_data))
30
20
10
0
Jan
Apr
date
colour
prediction
Real

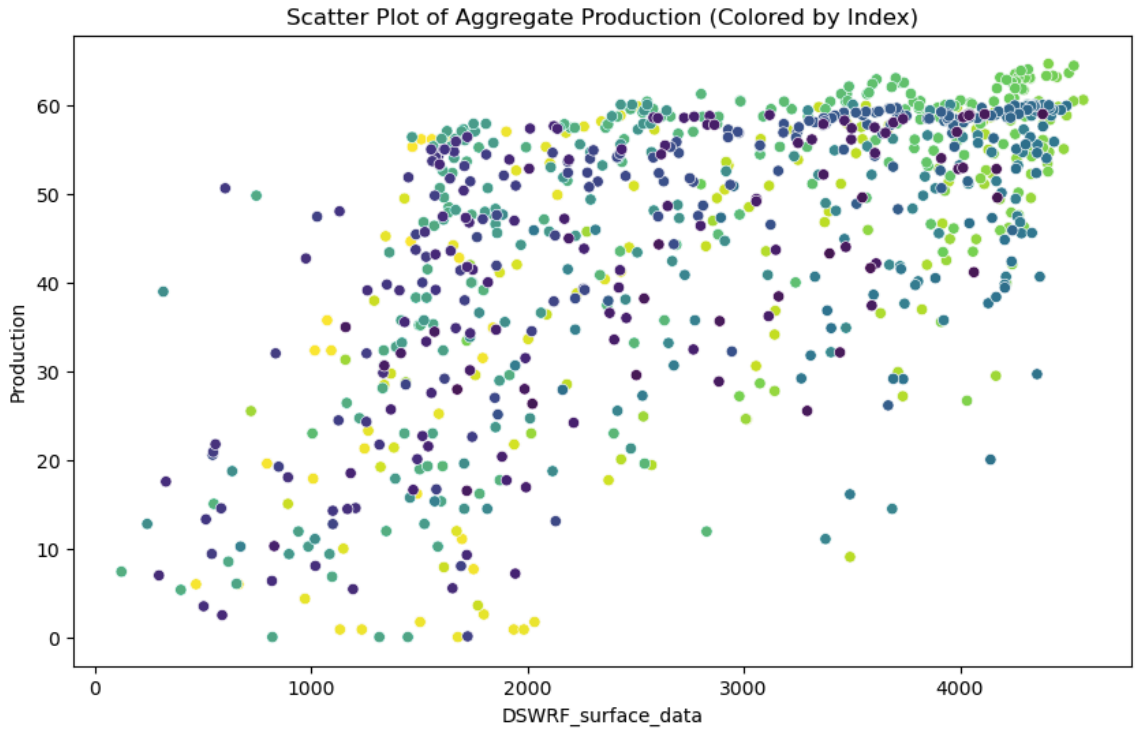

Real
30
20
10
0
0
10
20
30
Predicted

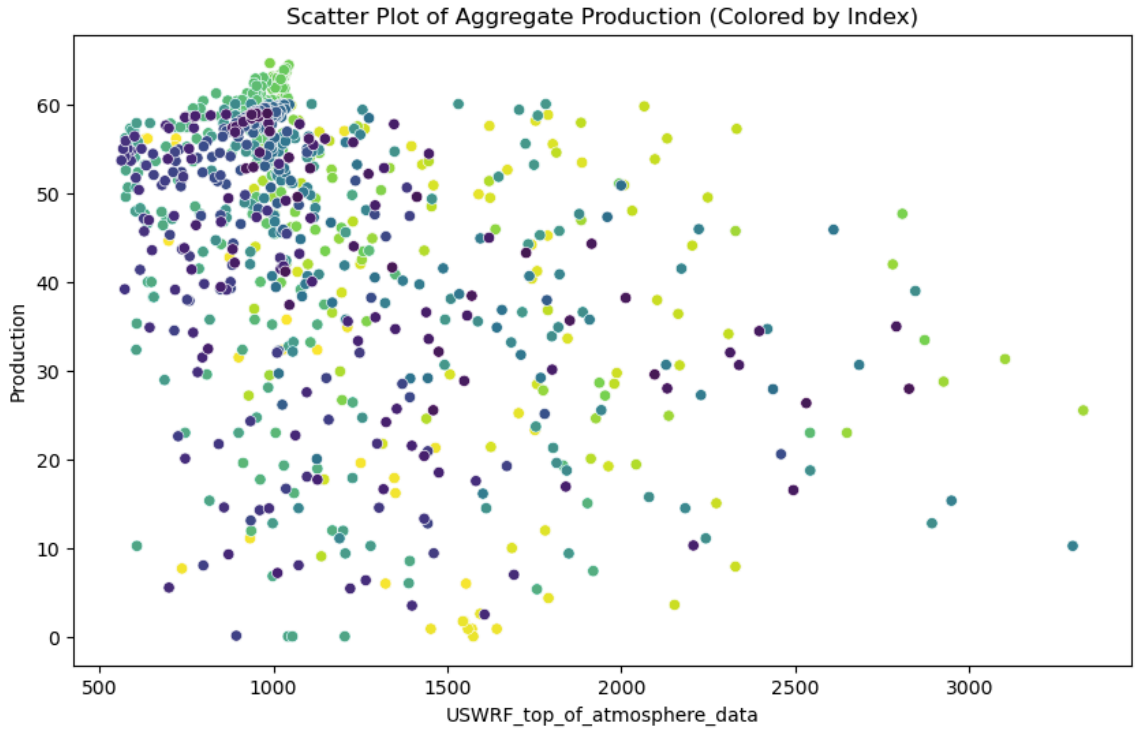


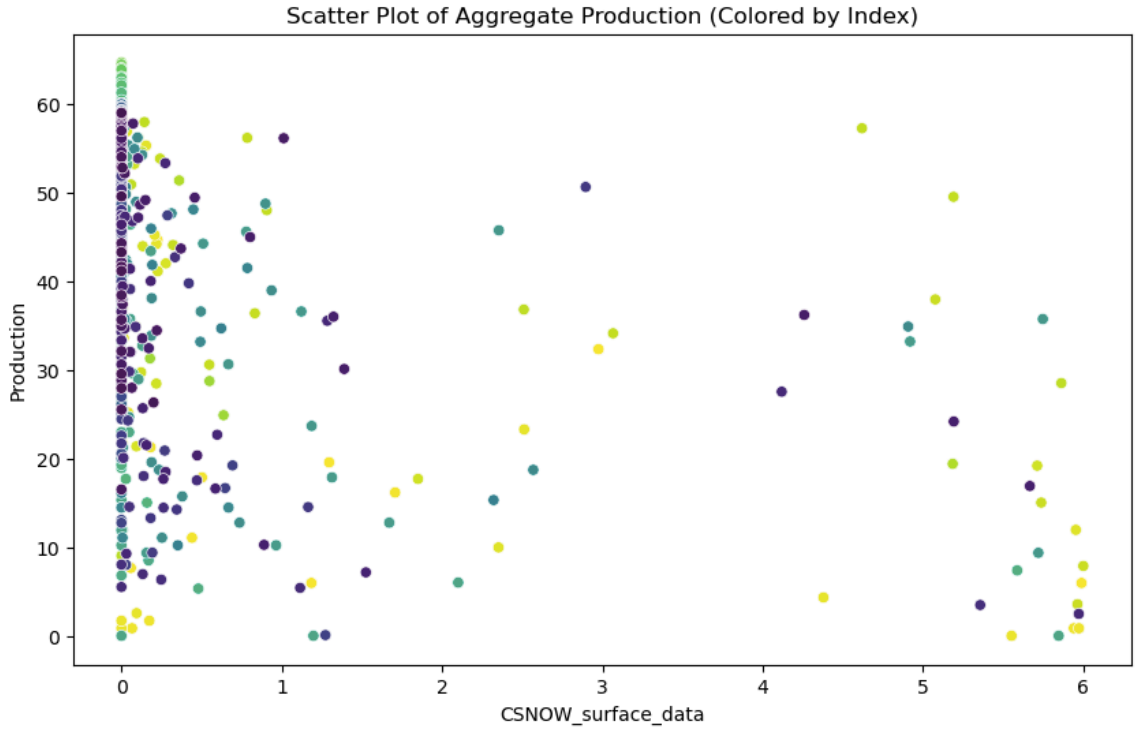
n	mean_val	sd_val	CV	FBias	MAPE	RMSE	MAD	MADP	WMAPE
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
174	13.77845	9.035031	0.6557364	0.07267855	Inf	3.8799	2.899512	0.2104382	0.2104382

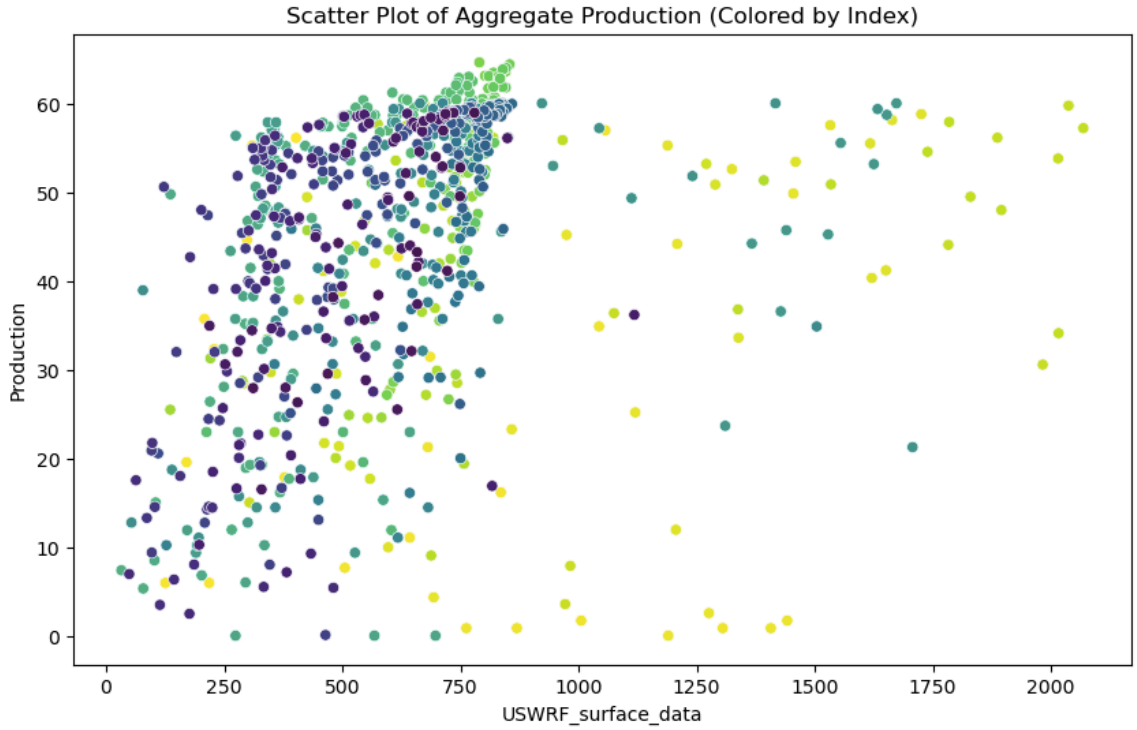
1 row

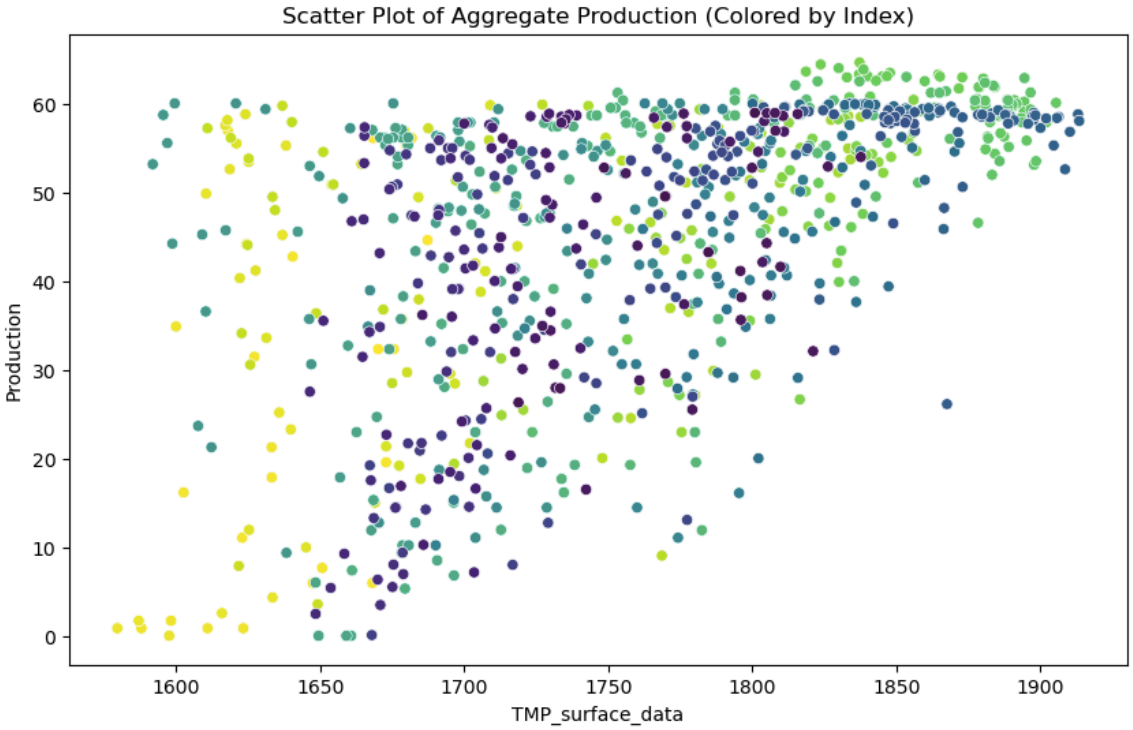
For cluster 4, the scatter plots for production amounts vs regressors are as following where we hued scatters with respect to their indices:

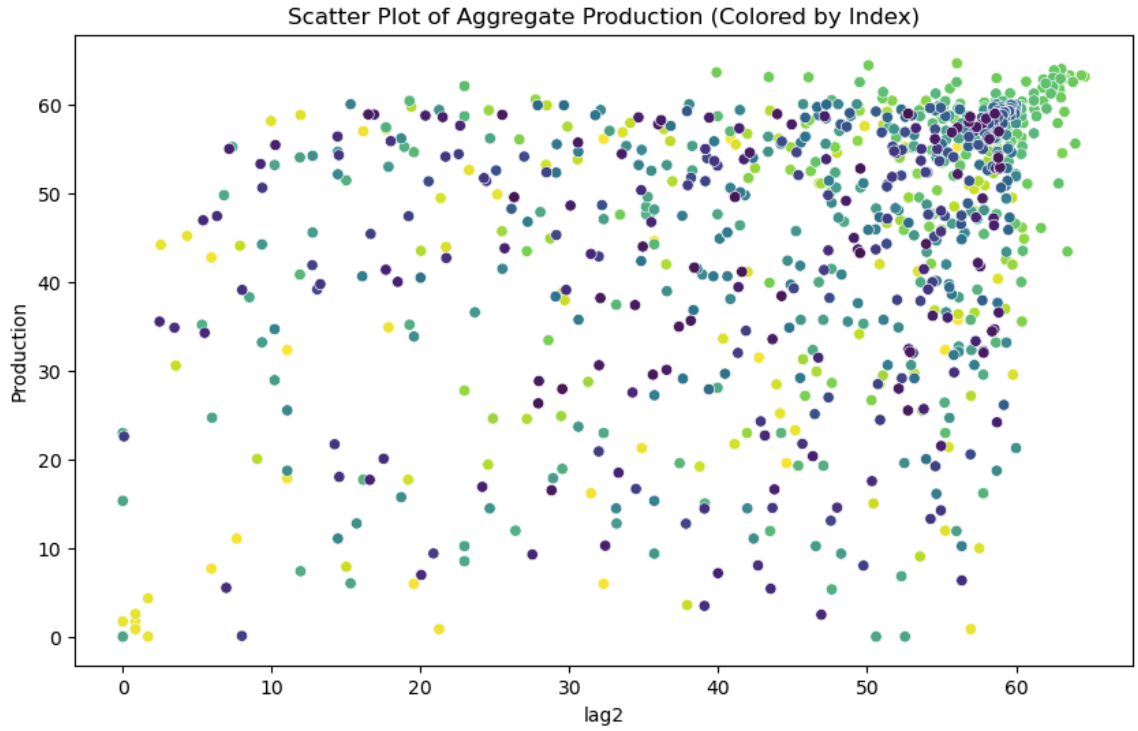


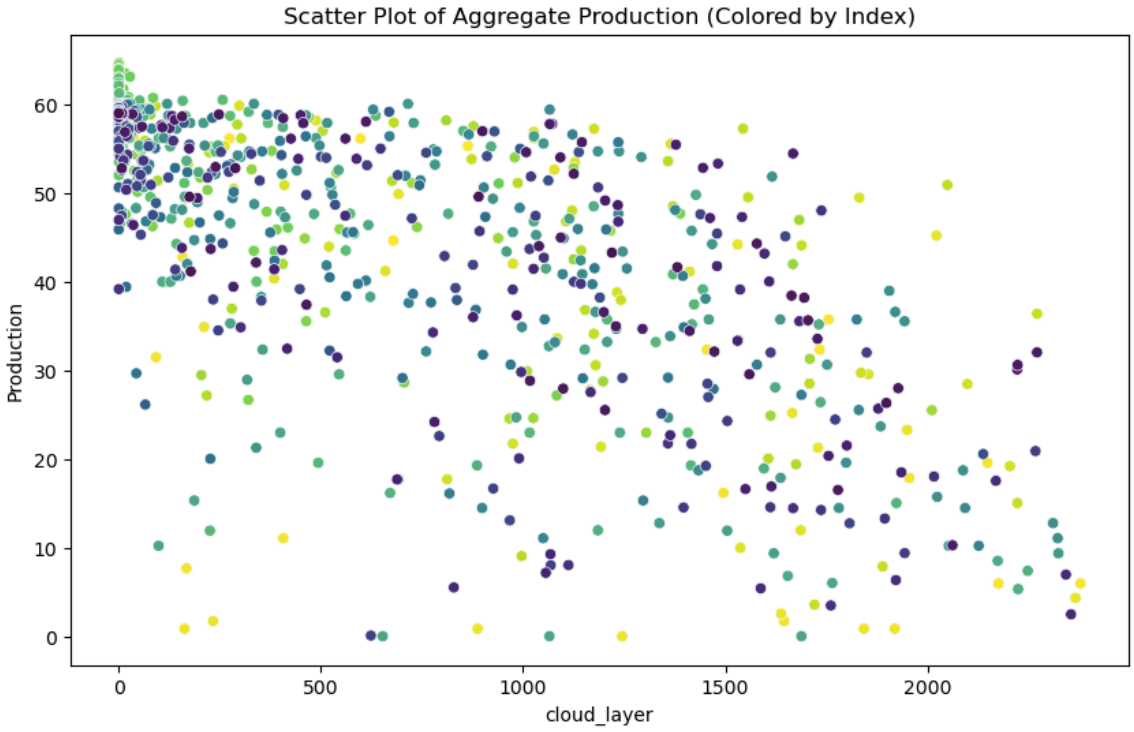

Scatter Plot of Aggregate Production (Colored by Index)
Production
USWRF_top_of_atmosphere_data
0
10
20
30
40
50
60
500
1000
1500
2000
2500
3000


Scatter Plot of Aggregate Production (Colored by Index)
Production
CSNOW_surface_data
0
10
20
30
40
50
60
0
1
2
3
4
5
6


Scatter Plot of Aggregate Production (Colored by Index)
Production
USWRF_surface_data
0
10
20
30
40
50
60
0
250
500
750
1000
1250
1500
1750
2000

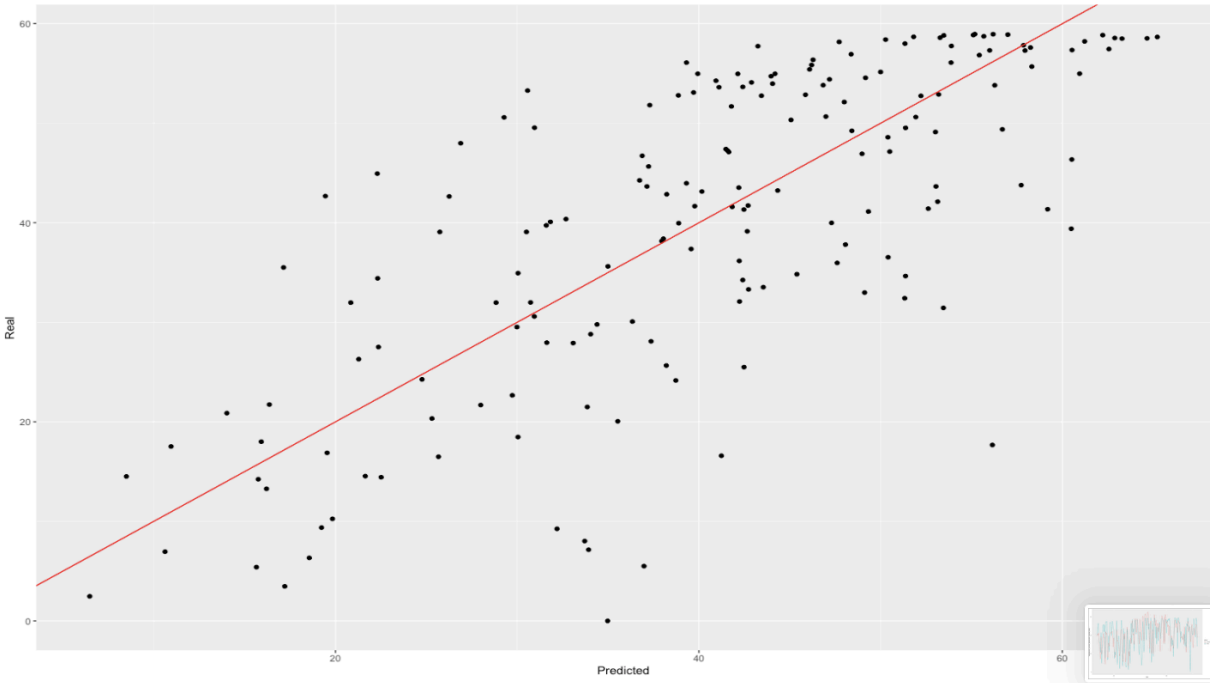

Scatter Plot of Aggregate Production (Colored by Index)
Production
TMP_surface_data
0
10
20
30
40
50
60
1600
1650
1700
1750
1800
1850
1900

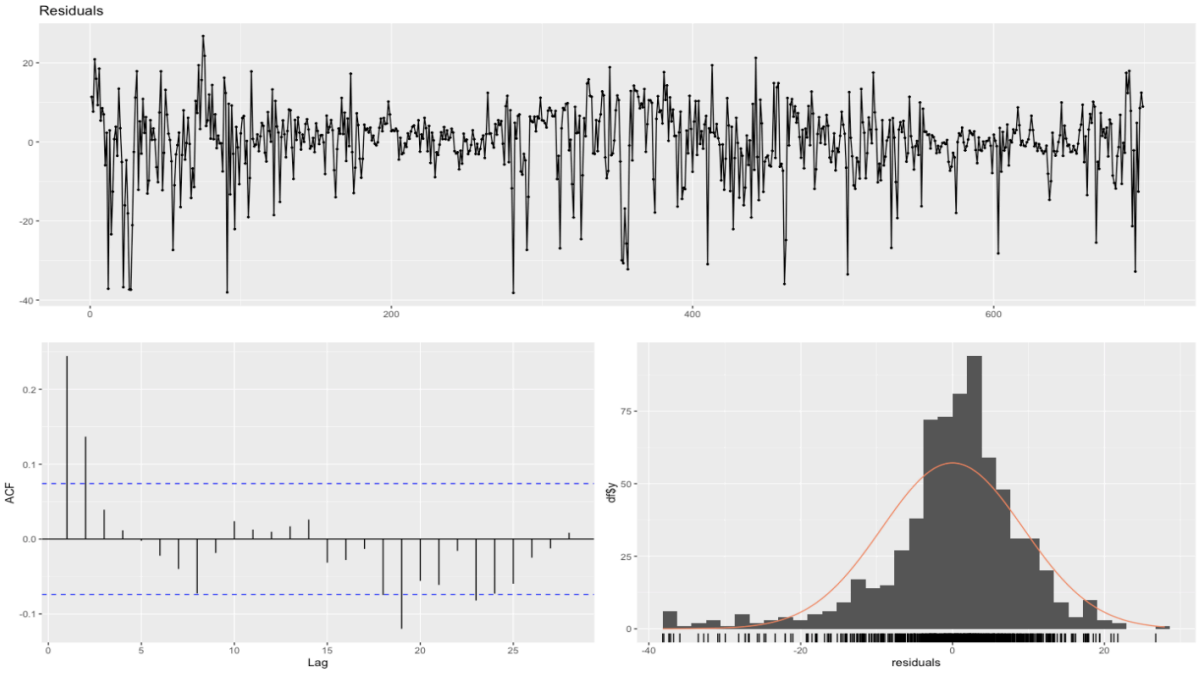


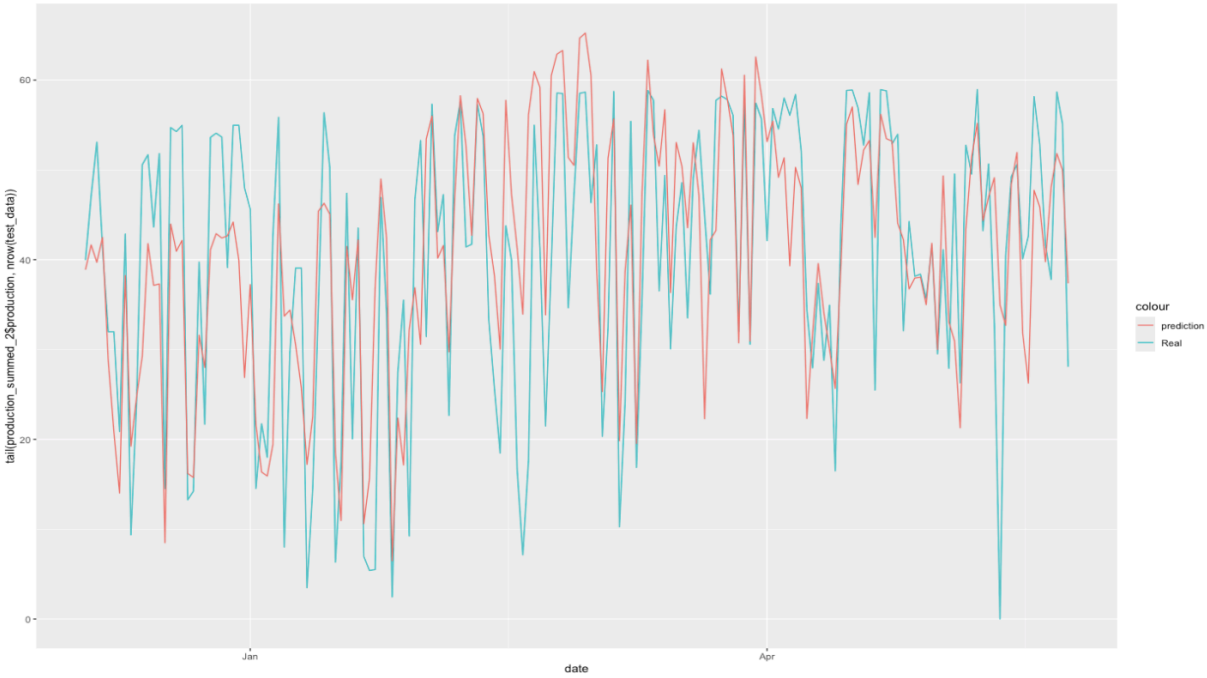


In previous models, we obtained highest coefficient for dswrf_surface and through our research about the regressors, we learned that the most important regressor is dswrf_surface. For cluster 4, it can be observed that there exists a kink like shape in our most important regressor, dswrf_surface. Therefore, we will be trying to develop more than one model, where in one of the models the dswrf_surface data has been used as piecewise non-linear regressor. The results of our first model are as given below:

```
Call:
lm(formula = head(production_summed_2$production, train_size) ~
    . - lag_48 - date - mon - dswrf_surface, data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-38.178  -3.312   1.010   5.287  26.765

Coefficients: (1 not defined because of singularities)
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -2.106e+02  4.646e+01  -4.534 6.85e-06 ***
tdc_low.cloud.layer         -8.659e-03  5.494e-03  -1.576 0.115450
tdc_middle.cloud.layer      -1.181e-02  3.559e-03  -3.319 0.000951 ***
tdc_high.cloud.layer         1.388e-03  4.253e-03   0.326 0.744319
tdc_entire.atmosphere       -1.092e-02  4.830e-03  -2.260 0.024118 *
uswrf_top_of_atmosphere      4.606e-03  2.132e-03   2.161 0.031068 *
csnow_surface               -1.076e+00  4.367e-01  -2.464 0.013998 *
dlwrf_surface               -2.704e-02  5.639e-03  -4.794 2.01e-06 ***
uswrf_surface                1.520e-04  2.482e-03   0.061 0.951187
tmp_surface                  1.651e-01  3.051e-02   5.409 8.79e-08 ***
dswrf_surface_log            1.351e+00  2.975e+00   0.454 0.649867
month_Jan_interaction_dswrf  2.107e-01  2.973e-01   0.708 0.478891
month_Feb_interaction_dswrf  1.688e+00  3.567e-01   4.732 2.71e-06 ***
month_Mar_interaction_dswrf  1.030e+00  4.225e-01   2.438 0.015006 *
month_Apr_interaction_dswrf -2.283e-01  4.882e-01  -0.468 0.640277
month_May_interaction_dswrf -3.647e-01  5.246e-01  -0.695 0.487115
month_Jun_interaction_dswrf -3.006e-01  5.407e-01  -0.556 0.578371
month_Jul_interaction_dswrf  7.796e-02  5.720e-01   0.136 0.891639
month_Aug_interaction_dswrf -5.072e-01  6.247e-01  -0.812 0.417073
month_Sep_interaction_dswrf  9.905e-03  5.404e-01   0.018 0.985382
month_Oct_interaction_dswrf  9.118e-02  4.166e-01   0.219 0.826821
month_Nov_interaction_dswrf  2.316e-02  3.134e-01   0.074 0.941110
month_Dec_interaction_dswrf         NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.464 on 677 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.6429,    Adjusted R-squared:  0.6318
F-statistic: 58.03 on 21 and 677 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 72.273, df = 26, p-value = 3.079e-06
```







Description: df [1 × 10]

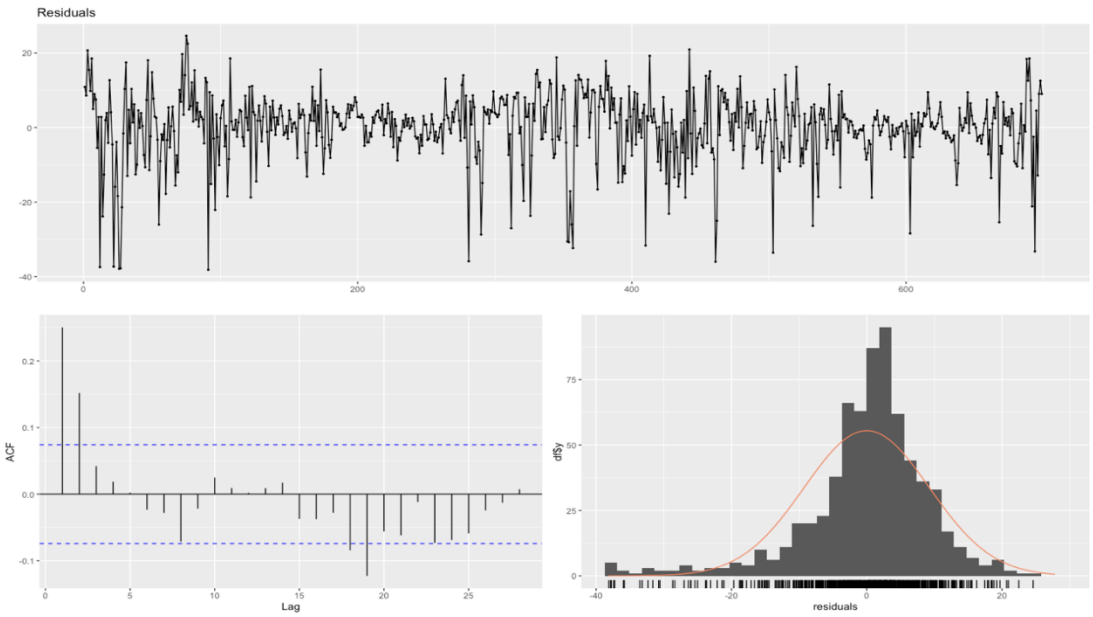
n	mean_val	sd_val	CV	FBias	MAPE	RMSE	MAD	MADP	WMAPE
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
174	40.09483	15.74612	0.392722	-0.01421695	Inf	11.0537	8.57806	0.2139443	0.2139443

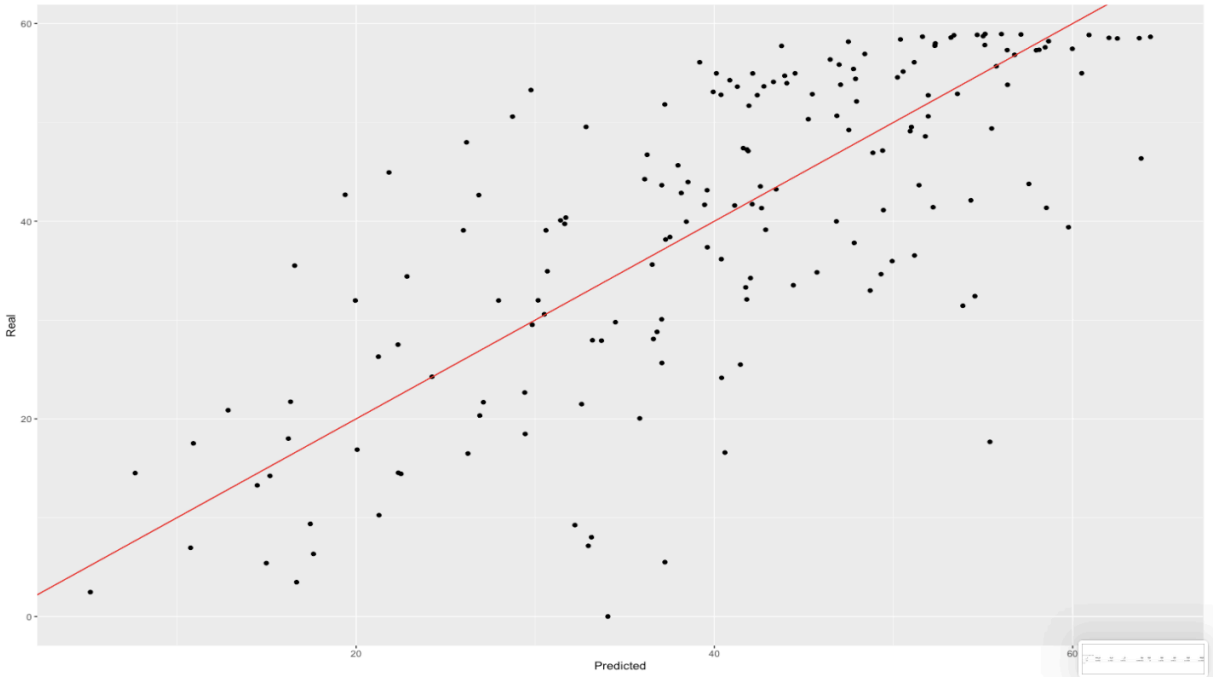
1 row

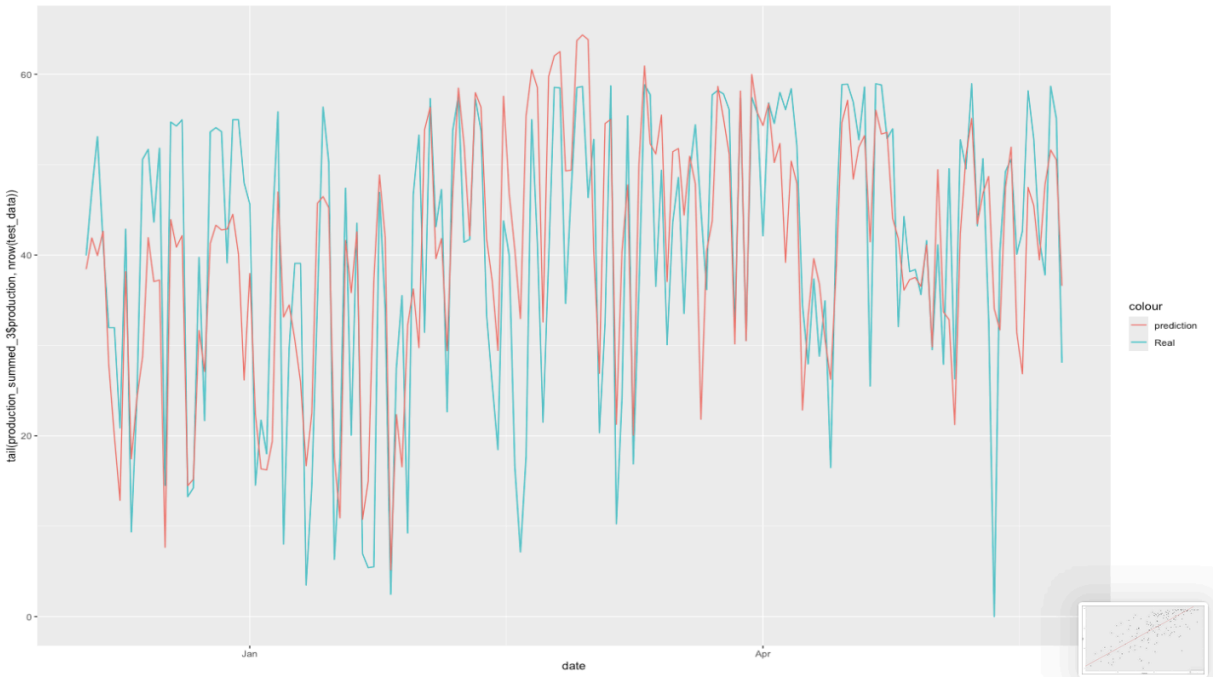
For our second summed model for cluster 4, we have tried different dswrf_surface values and by comparing the results, decided that the best dswrf_surface value to separate our model into two pieces is 2800, where we obtained better model evaluation metric results. Besides, we have tried to apply different piecewise nonlinear properties for other regressors however we were not able to obtain better results. The output for the piecewise nonlinear model is as given

below:

```
Call:
lm(formula = head(production_summed_3$production, train_size) ~
    . - lag_48 - dswrf_surface - date, data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-38.154  -3.315   1.194   4.978  24.611

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)               -2.333e+02  4.968e+01  -4.697 3.20e-06 ***
tdc_low.cloud.layer       -6.669e-03  5.390e-03  -1.237  0.21637
tdc_middle.cloud.layer    -1.128e-02  3.541e-03  -3.184  0.00152 **
tdc_high.cloud.layer       1.737e-03  3.970e-03   0.438  0.66186
tdc_entire.atmosphere     -1.092e-02  4.440e-03  -2.459  0.01420 *
uswrf_top_of_atmosphere    4.640e-04  3.006e-03   0.154  0.87739
csnow_surface             -1.111e+00  4.344e-01  -2.559  0.01072 *
dlwrf_surface             -2.968e-02  5.797e-03  -5.120 3.98e-07 ***
uswrf_surface              6.355e-03  3.850e-03   1.651  0.09922 .
tmp_surface                1.926e-01  3.367e-02   5.719 1.61e-08 ***
monAug                    -4.121e+00  2.978e+00  -1.384  0.16688
monDec                    -6.104e+00  5.608e+00  -1.088  0.27677
monFeb                     7.792e+00  4.224e+00   1.845  0.06550 .
monJan                    -4.018e+00  5.278e+00  -0.761  0.44675
monJul                     2.732e+00  2.277e+00   1.200  0.23052
monJun                     6.481e-01  2.285e+00   0.284  0.77678
monMar                     6.809e+00  2.566e+00   2.653  0.00815 **
monMay                     3.414e-02  1.997e+00   0.017  0.98637
monNov                    -5.163e+00  4.498e+00  -1.148  0.25140
monOct                    -2.958e+00  3.279e+00  -0.902  0.36735
monSep                    -6.460e-01  2.825e+00  -0.229  0.81922
dswrf_surface_higherpiece -3.429e-03  2.842e-03  -1.206  0.22815
dswrf_surface_lowerpiece  -2.215e-03  3.066e-03  -0.723  0.47014
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.428 on 676 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.6461,    Adjusted R-squared:  0.6345
F-statistic: 56.09 on 22 and 676 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 75.765, df = 26, p-value = 9.273e-07
```



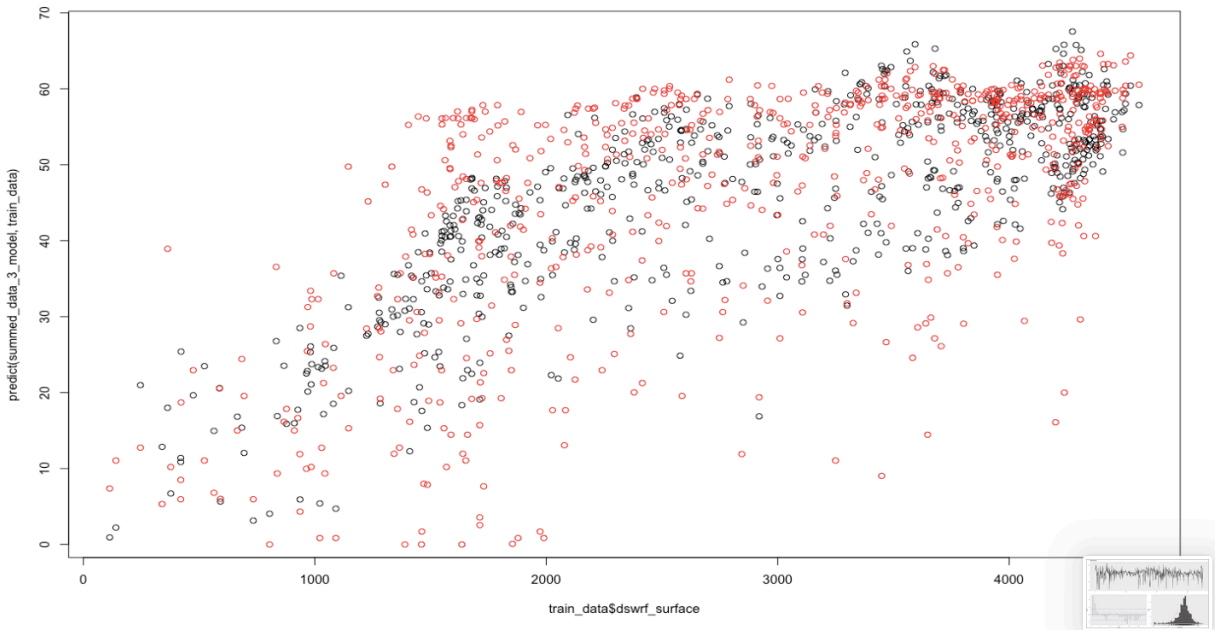




Description: df [1 × 10]

n	mean_val	sd_val	CV	FBias	MAPE	RMSE	MAD	MADP	WMAPE
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
174	40.09483	15.74612	0.392722	-0.009874612	Inf	11.00766	8.496111	0.2119004	0.2119004

1 row



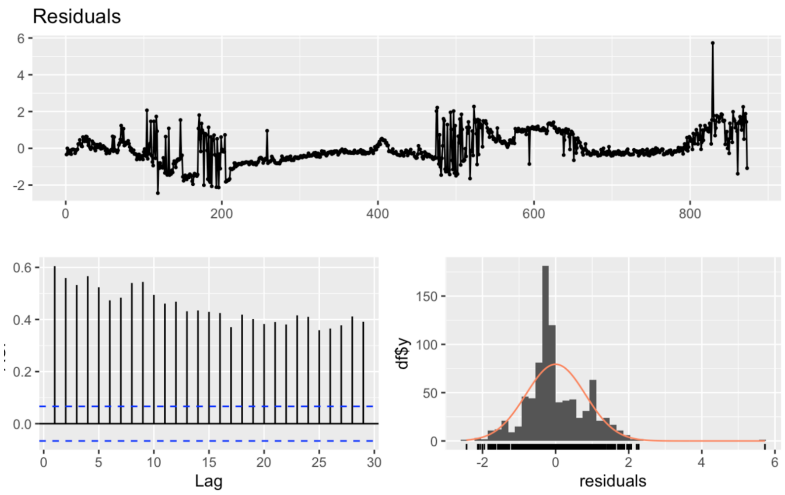
For fourth cluster, our linear model yielded WMAPE, RMSE, and MAD values 0.21394, 11.0537, and 8.57806 respectively. The piecewise nonlinear model yielded WMAPE, RMSE, MAD, and Adjusted R-squared values 0.2119, 11.0076, and 8.496 respectively. When we compare two models for cluster 4, it can be observed that in all three model evaluation metric results, small improvements could be observed.

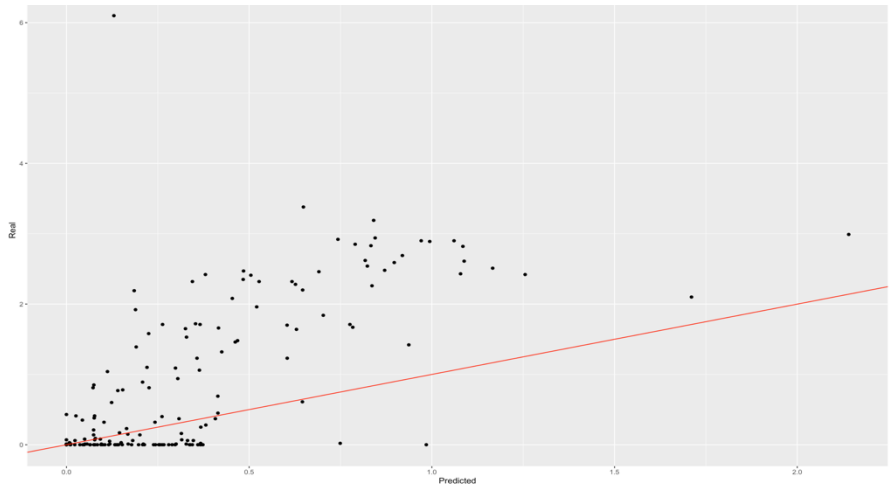
4.3. Hourly Models

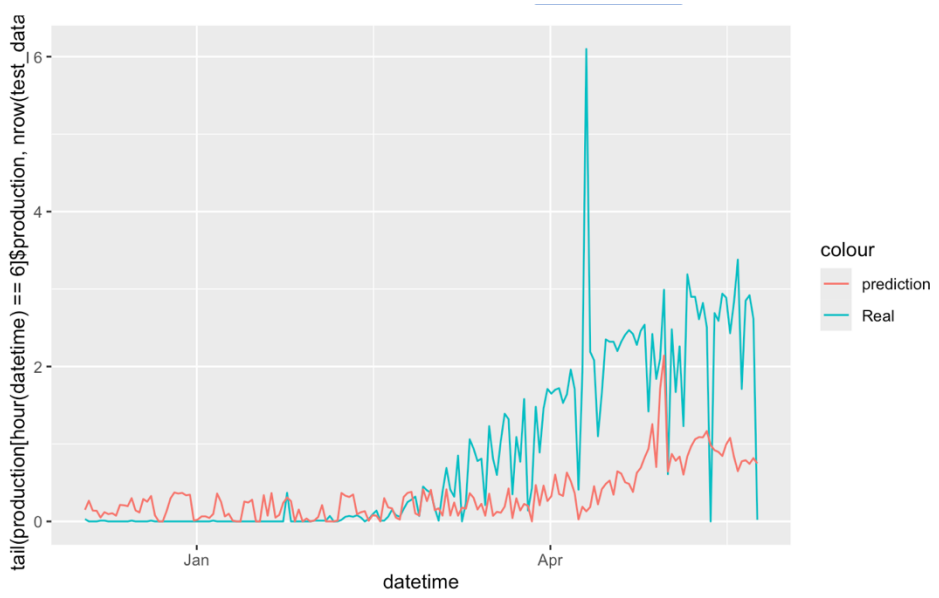
In the final model phase, we treated each hour independently and built a model for each hour. We deleted hours between 6 a.m. and 9 p.m. since they return 0 for most of the data so for the remaining 12 hours of the day, we built 12 different models. Hourly production vs days scatter plots can be reached from our python code. For hourly models, we have tried to use dswrf_surface data as given. However, using logged dswrf_surface data yielded better results for all hours. Besides, we have tried to transform all other regressors however we did not observe any significant improvement. On the other hand, we used regression trees on residuals to identify if any significant regressor would have different impact on the results however, we were unable to reach any significant conclusions from regression trees.

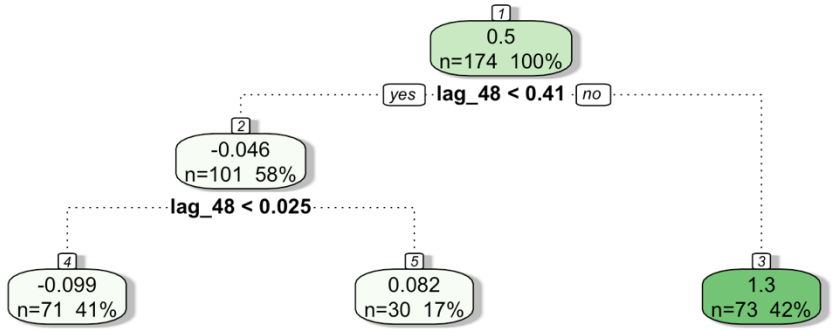
4.3.1 Hour 6

```
Call:
lm(formula = head(production[hour(datetime) == 6]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1090 -0.4162 -0.1191  0.4252  2.4169

Coefficients: (7 not defined because of singularities)
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                -9.4097699  4.0024155  -2.351   0.0190 *
tdc_low.cloud.layer        -0.0011679  0.0021869  -0.534   0.5935
tdc_middle.cloud.layer     -0.0002469  0.0014699  -0.168   0.8667
tdc_high.cloud.layer       -0.0006608  0.0015910  -0.415   0.6780
tdc_entire.atmosphere      -0.0024285  0.0019733  -1.231   0.2189
uswrf_top_of_atmosphere     0.0767413  0.0525313   1.461   0.1445
csnow_surface               0.2155673  0.1880158   1.147   0.2520
dlwrf_surface              -0.0013862  0.0031552  -0.439   0.6606
uswrf_surface              -0.5264654  1.0720461  -0.491   0.6235
tmp_surface                 0.0371057  0.0172413   2.152   0.0317 *
dswrf_log                   1.7191691  0.3872420   4.440 1.05e-05 ***
month_Jan_interaction_dswrf        NA         NA      NA       NA
month_Feb_interaction_dswrf        NA         NA      NA       NA
month_Mar_interaction_dswrf        NA         NA      NA       NA
month_Apr_interaction_dswrf 2.6971933  1.2964909   2.080   0.0379 *
month_May_interaction_dswrf -0.6527368  0.3634247  -1.796   0.0729 .
month_Jun_interaction_dswrf -0.3740572  0.3552645  -1.053   0.2928
month_Jul_interaction_dswrf -0.3072498  0.3900362  -0.788   0.4311
month_Aug_interaction_dswrf -0.6135402  0.7026601  -0.873   0.3829
month_Sep_interaction_dswrf        NA         NA      NA       NA
month_Oct_interaction_dswrf        NA         NA      NA       NA
month_Nov_interaction_dswrf        NA         NA      NA       NA
month_Dec_interaction_dswrf        NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7812 on 683 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.3952,    Adjusted R-squared:  0.3819
F-statistic: 29.76 on 15 and 683 DF,  p-value: < 2.2e-16
```





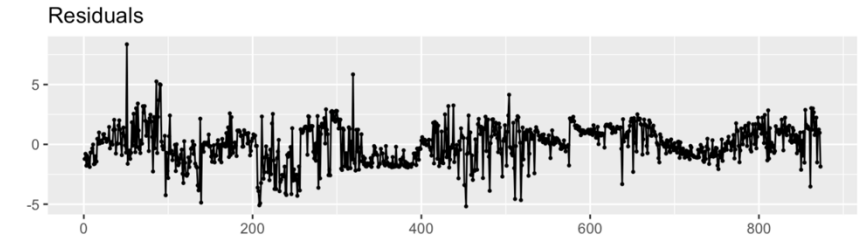


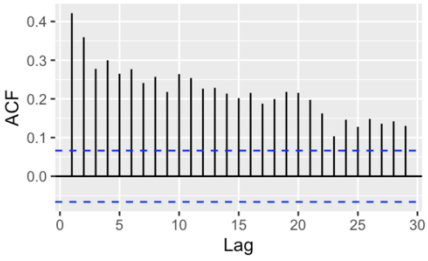


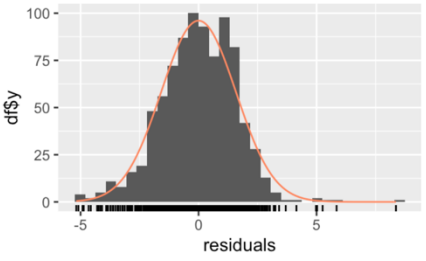
Rattle 2024-Jun-04 16:34:07 ahmetceliker

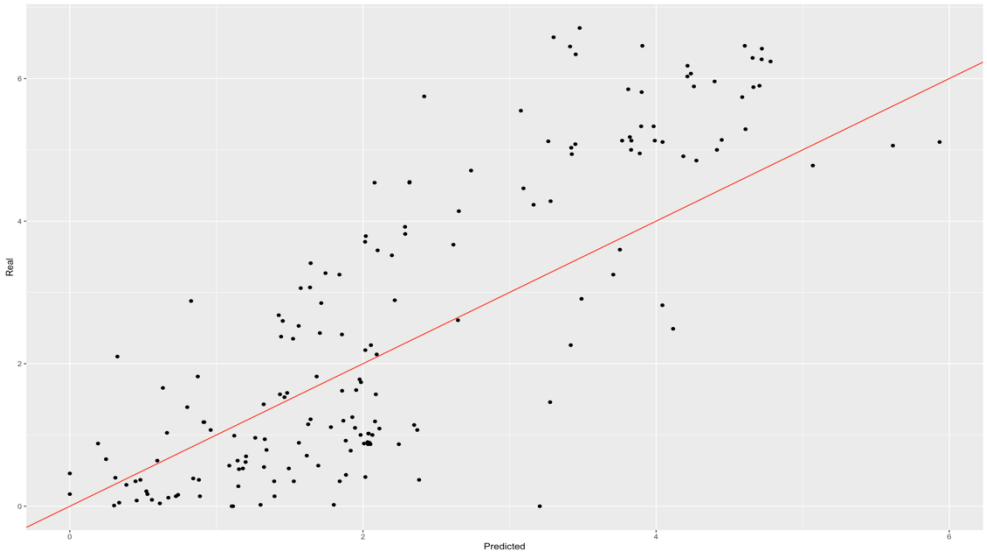
4.3.2. Hour 7

```
Call:
lm(formula = head(production[hour(datetime) == 7]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2395 -1.1360  0.0651  1.1501  8.3242

Coefficients: (4 not defined because of singularities)
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                -35.835715   9.213614  -3.889 0.000110 ***
tdc_low.cloud.layer         -0.008218   0.004887  -1.682 0.093095 .
tdc_middle.cloud.layer      -0.001465   0.003265  -0.449 0.653676
tdc_high.cloud.layer        -0.003338   0.003551  -0.940 0.347479
tdc_entire.atmosphere       -0.001529   0.004515  -0.339 0.735044
uswrf_top_of_atmosphere      0.005895   0.020024   0.294 0.768535
csnow_surface                0.518138   0.396636   1.306 0.191882
dlwrf_surface               -0.015481   0.007275  -2.128 0.033687 *
uswrf_surface                0.226628   0.287244   0.789 0.430402
tmp_surface                  0.152683   0.039599   3.856 0.000126 ***
dswrf_log                    0.702174   0.264327   2.656 0.008082 **
month_Jan_interaction_dswrf         NA         NA      NA       NA
month_Feb_interaction_dswrf         NA         NA      NA       NA
month_Mar_interaction_dswrf  0.930837   0.293045   3.176 0.001558 **
month_Apr_interaction_dswrf -0.106669   0.092386  -1.155 0.248658
month_May_interaction_dswrf -0.103522   0.083928  -1.233 0.217828
month_Jun_interaction_dswrf -0.096292   0.084375  -1.141 0.254173
month_Jul_interaction_dswrf -0.102094   0.092347  -1.106 0.269313
month_Aug_interaction_dswrf -0.213845   0.099947  -2.140 0.032744 *
month_Sep_interaction_dswrf -0.135397   0.120937  -1.120 0.263296
month_Oct_interaction_dswrf  1.733745   0.562972   3.080 0.002156 **
month_Nov_interaction_dswrf         NA         NA      NA       NA
month_Dec_interaction_dswrf         NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.686 on 680 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.4914,    Adjusted R-squared:  0.4779
F-statistic:  36.5 on 18 and 680 DF,  p-value: < 2.2e-16
```

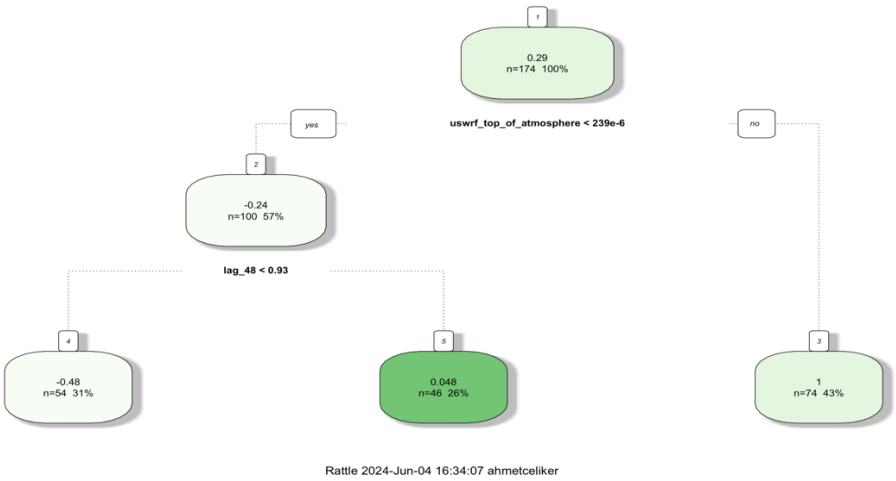

Residuals
5
0
-5
0
200
400
600
800


ACF
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
Lag


df$y
100
75
50
25
0
-5
0
5
residuals

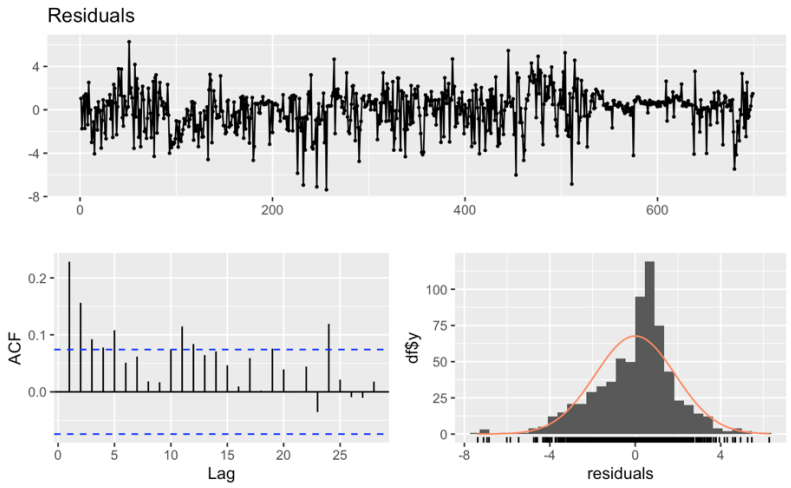

Real
6
4
2
0
0
2
4
6
Predicted

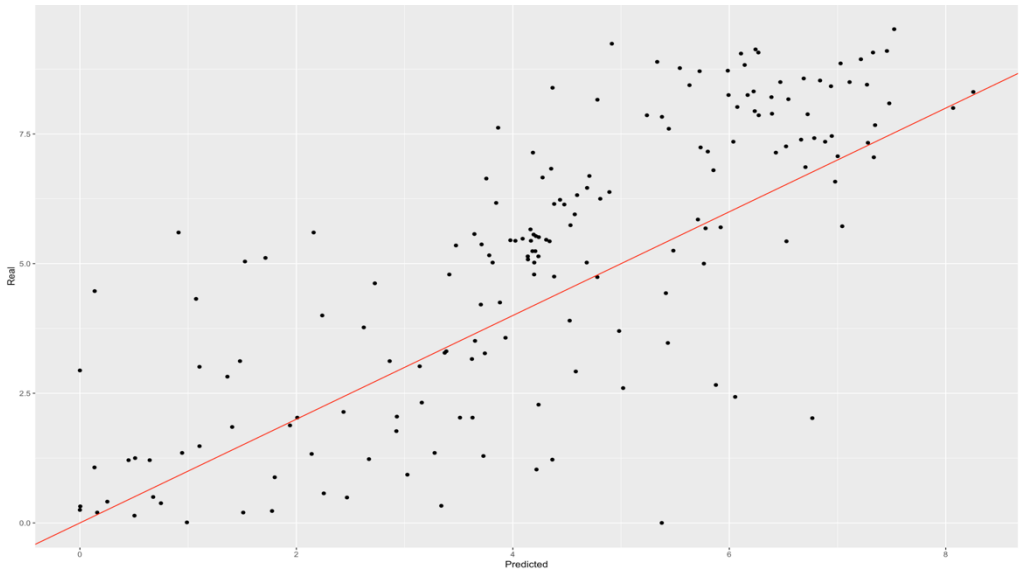

tail(production[hour(datetime) == 7]$production, nrow(test_data))
6
4
2
0
Jan
Apr
datetime
colour
prediction
Real

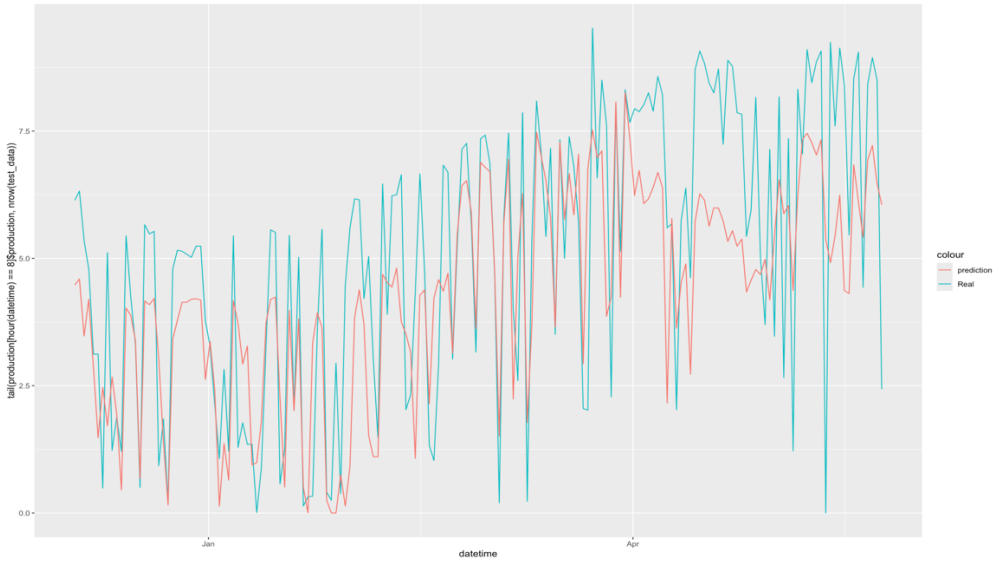


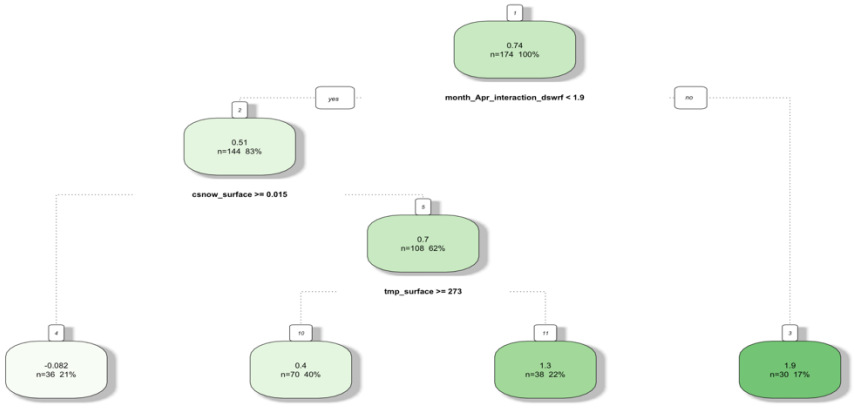
4.3.3.Hour 8

```
Call:
lm(formula = head(production[hour(datetime) == 8]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-7.3755 -1.0036  0.3246  1.0678  6.2749

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  9.591e+00  1.045e+01   0.918 0.359138
tdc_low.cloud.layer         -1.966e-02  5.635e-03  -3.489 0.000517 ***
tdc_middle.cloud.layer      -9.973e-03  3.785e-03  -2.634 0.008621 **
tdc_high.cloud.layer         6.710e-03  4.203e-03   1.596 0.110886
tdc_entire.atmosphere       -1.323e-02  5.178e-03  -2.555 0.010827 *
uswrf_top_of_atmosphere     -2.166e-03  9.151e-03  -0.237 0.812950
csnow_surface               -5.701e-01  4.663e-01  -1.223 0.221895
dlwrf_surface                3.756e-04  7.982e-03   0.047 0.962483
uswrf_surface                1.264e-02  8.427e-02   0.150 0.880839
tmp_surface                 -2.019e-02  4.443e-02  -0.454 0.649645
dswrf_log                    1.346e+00  2.416e-01   5.571 3.66e-08 ***
month_Jan_interaction_dswrf -1.594e+01  6.493e+00  -2.454 0.014369 *
month_Feb_interaction_dswrf  5.388e-02  1.506e-01   0.358 0.720563
month_Mar_interaction_dswrf -1.486e-02  4.305e-02  -0.345 0.730061
month_Apr_interaction_dswrf -5.806e-02  2.536e-02  -2.290 0.022348 *
month_May_interaction_dswrf -3.002e-02  2.283e-02  -1.315 0.188921
month_Jun_interaction_dswrf -2.779e-02  2.312e-02  -1.202 0.229866
month_Jul_interaction_dswrf -2.214e-02  2.543e-02  -0.871 0.384311
month_Aug_interaction_dswrf -3.915e-02  2.880e-02  -1.359 0.174492
month_Sep_interaction_dswrf -3.895e-02  3.224e-02  -1.208 0.227526
month_Oct_interaction_dswrf -1.210e-03  4.299e-02  -0.028 0.977560
month_Nov_interaction_dswrf  1.566e-02  1.025e-01   0.153 0.878604
month_Dec_interaction_dswrf  9.459e-01  1.492e+00   0.634 0.526447
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.905 on 676 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.5484,    Adjusted R-squared:  0.5337
F-statistic: 37.32 on 22 and 676 DF,  p-value: < 2.2e-16
```





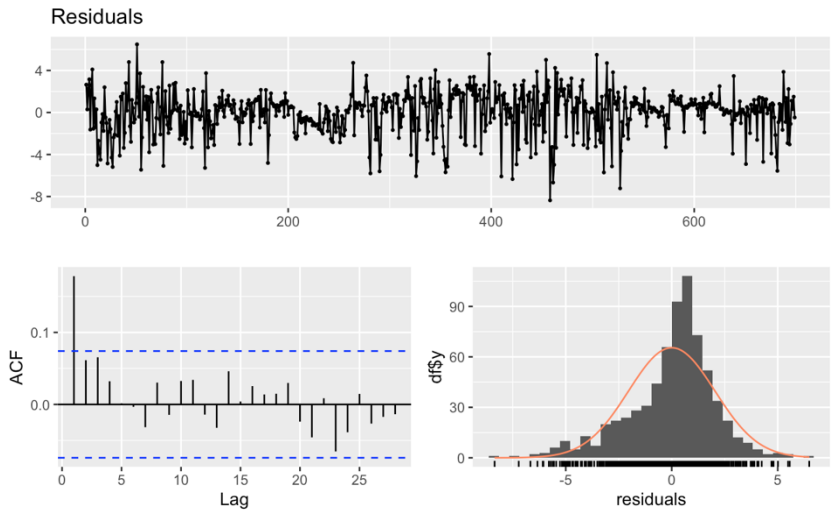


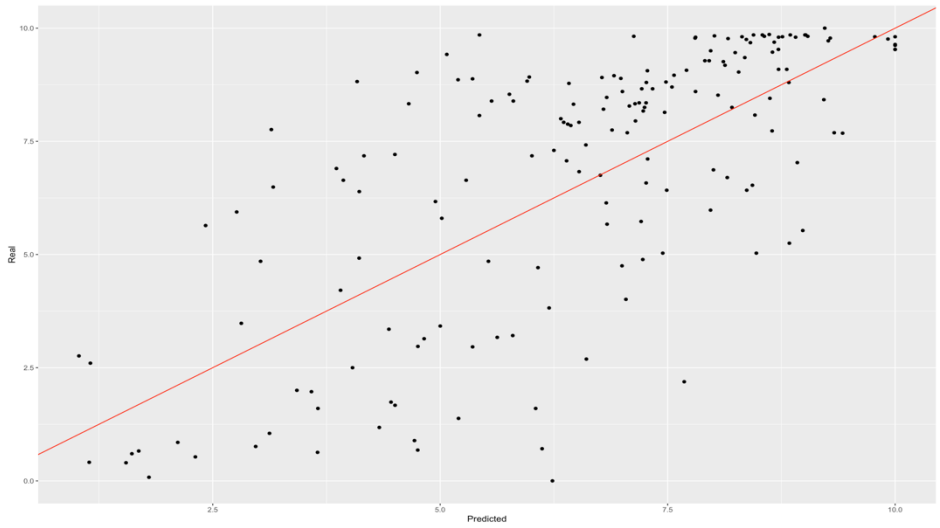


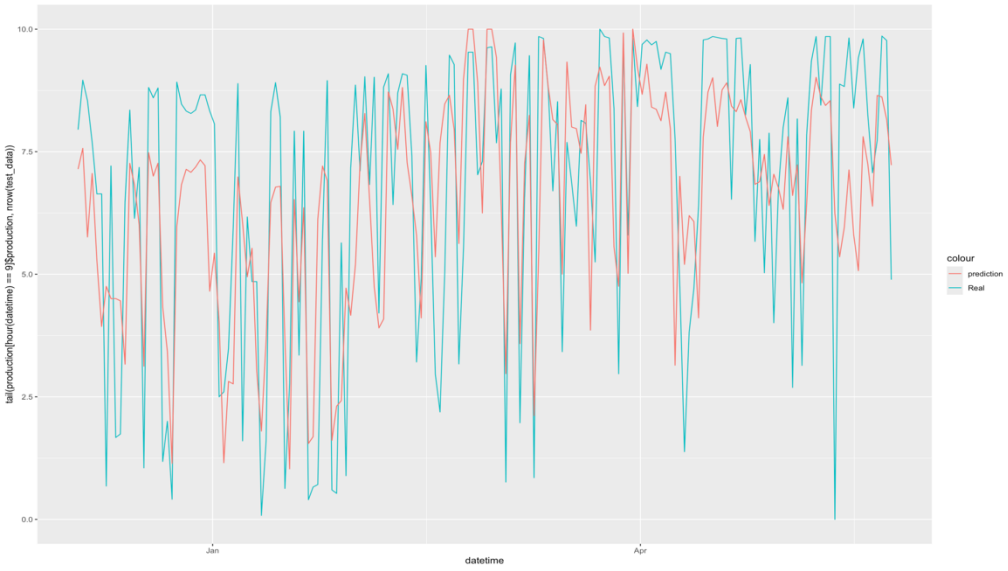
Rattle 2024-Jun-04 16:34:09 ahmetceliker

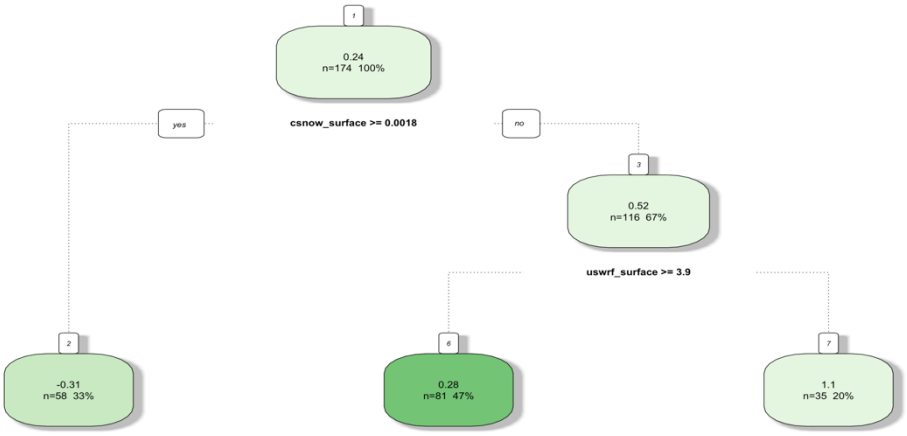
4.3.4.Hour 9

```
Call:
lm(formula = head(production[hour(datetime) == 9]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-8.3576 -0.9500  0.3239  1.2032  6.4859

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -3.449e+01  1.122e+01  -3.074 0.002197 **
tdc_low.cloud.layer           -1.314e-02  6.579e-03  -1.997 0.046195 *
tdc_middle.cloud.layer        -1.018e-02  4.313e-03  -2.359 0.018596 *
tdc_high.cloud.layer           4.658e-03  5.098e-03   0.914 0.361129
tdc_entire.atmosphere         -1.271e-02  6.086e-03  -2.088 0.037146 *
uswrf_top_of_atmosphere        6.186e-04  6.752e-03   0.092 0.927025
csnow_surface                 -7.801e-01  4.966e-01  -1.571 0.116680
dlwrf_surface                 -1.963e-02  8.349e-03  -2.351 0.019021 *
uswrf_surface                  7.047e-02  3.276e-02   2.151 0.031820 *
tmp_surface                    1.620e-01  4.703e-02   3.445 0.000606 ***
dswrf_log                      8.837e-01  4.570e-01   1.934 0.053540 .
month_Jan_interaction_dswrf -9.248e-02  6.031e-02  -1.533 0.125622
month_Feb_interaction_dswrf  1.319e-02  3.103e-02   0.425 0.670999
month_Mar_interaction_dswrf -2.528e-02  1.970e-02  -1.283 0.199915
month_Apr_interaction_dswrf -4.098e-02  1.363e-02  -3.006 0.002746 **
month_May_interaction_dswrf -3.727e-02  1.179e-02  -3.161 0.001642 **
month_Jun_interaction_dswrf -3.608e-02  1.187e-02  -3.039 0.002466 **
month_Jul_interaction_dswrf -3.656e-02  1.310e-02  -2.791 0.005397 **
month_Aug_interaction_dswrf -5.269e-02  1.575e-02  -3.345 0.000868 ***
month_Sep_interaction_dswrf -4.592e-02  1.754e-02  -2.618 0.009034 **
month_Oct_interaction_dswrf -3.918e-02  2.072e-02  -1.891 0.059102 .
month_Nov_interaction_dswrf -5.144e-02  2.884e-02  -1.784 0.074915 .
month_Dec_interaction_dswrf -3.390e-02  5.712e-02  -0.593 0.553111
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.074 on 676 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.5068,    Adjusted R-squared:  0.4907
F-statistic: 31.57 on 22 and 676 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 39.806, df = 26, p-value = 0.04076
```





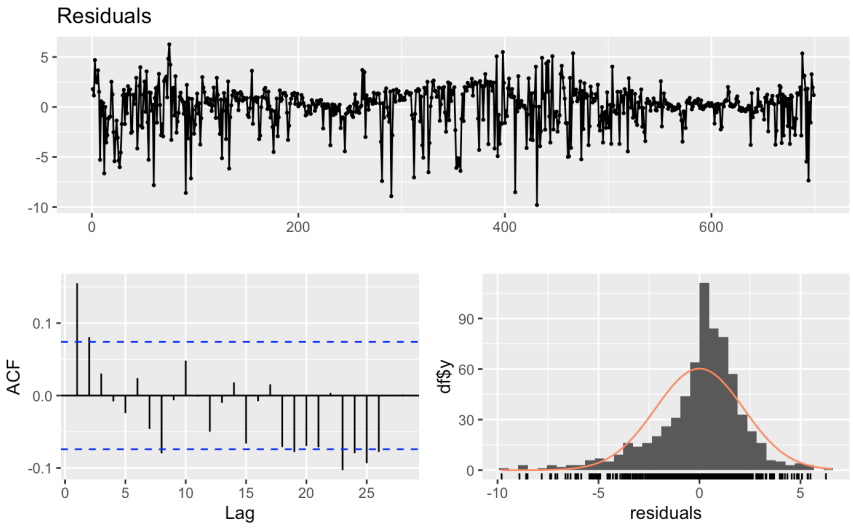


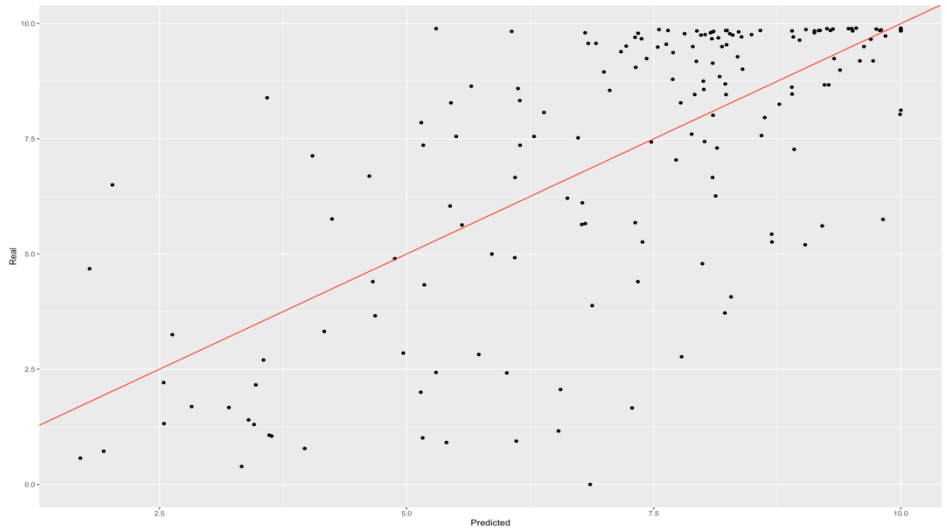


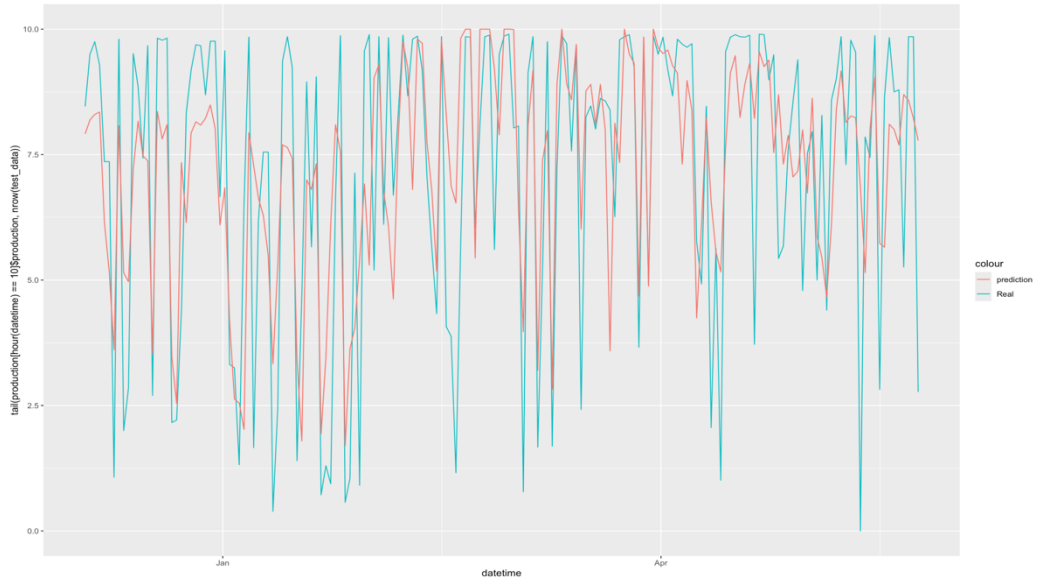
Rattle 2024-Jun-04 16:34:11 ahmetceliker

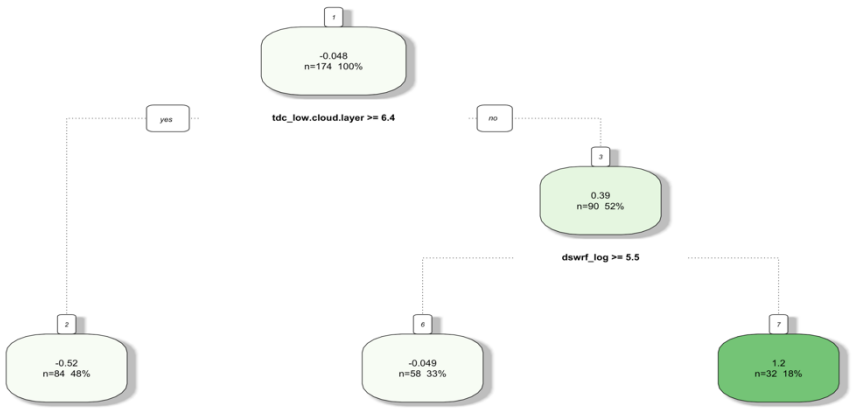
4.3.5 Hour 10

```
Call:
lm(formula = head(production[hour(datetime) == 10]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-9.7982 -0.8323  0.3121  1.3035  6.2577

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -46.212902  12.588416  -3.671 0.000261 ***
tdc_low.cloud.layer          -0.012873   0.006713  -1.917 0.055602 .
tdc_middle.cloud.layer       -0.008575   0.004057  -2.114 0.034910 *
tdc_high.cloud.layer         -0.001064   0.004852  -0.219 0.826550
tdc_entire.atmosphere        -0.005063   0.005722  -0.885 0.376522
uswrf_top_of_atmosphere       0.002554   0.002236   1.142 0.253659
csnow_surface                -0.882749   0.585602  -1.507 0.132170
dlwrf_surface                -0.029746   0.008652  -3.438 0.000622 ***
uswrf_surface                 0.003660   0.004773   0.767 0.443466
tmp_surface                   0.208966   0.051268   4.076 5.13e-05 ***
dswrf_log                     0.739963   0.634839   1.166 0.244190
month_Jan_interaction_dswrf  -0.009176   0.006812  -1.347 0.178412
month_Feb_interaction_dswrf  -0.001189   0.005411  -0.220 0.826132
month_Mar_interaction_dswrf  -0.004040   0.004268  -0.947 0.344087
month_Apr_interaction_dswrf  -0.006144   0.003651  -1.683 0.092889 .
month_May_interaction_dswrf  -0.006335   0.003409  -1.858 0.063591 .
month_Jun_interaction_dswrf  -0.005903   0.003424  -1.724 0.085192 .
month_Jul_interaction_dswrf  -0.005579   0.003649  -1.529 0.126750
month_Aug_interaction_dswrf  -0.007453   0.004117  -1.810 0.070675 .
month_Sep_interaction_dswrf  -0.006541   0.004339  -1.508 0.132122
month_Oct_interaction_dswrf  -0.006072   0.004714  -1.288 0.198217
month_Nov_interaction_dswrf  -0.007373   0.005600  -1.317 0.188432
month_Dec_interaction_dswrf  -0.006032   0.007049  -0.856 0.392499
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.22 on 676 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.4456,    Adjusted R-squared:  0.4276
F-statistic:  24.7 on 22 and 676 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 68.891, df = 26, p-value = 9.61e-06
```





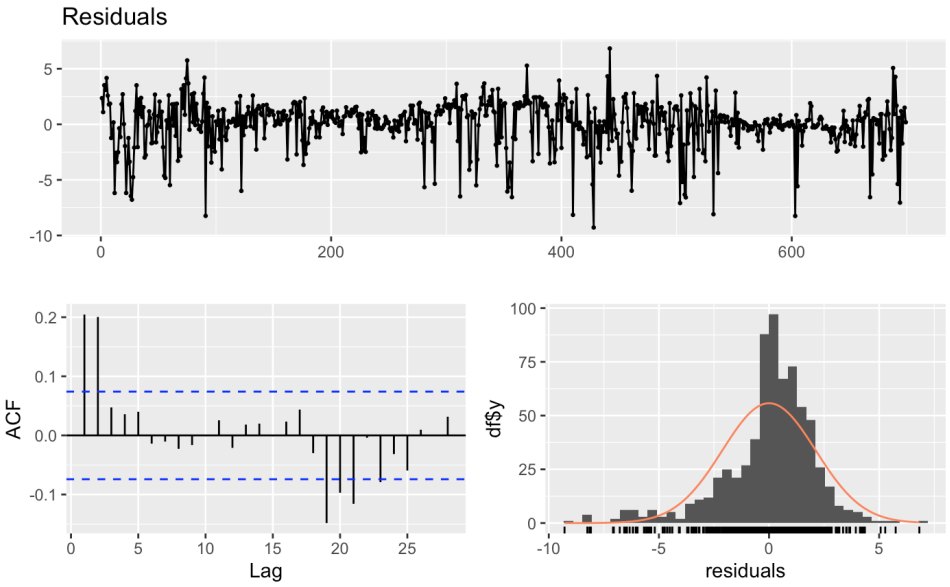


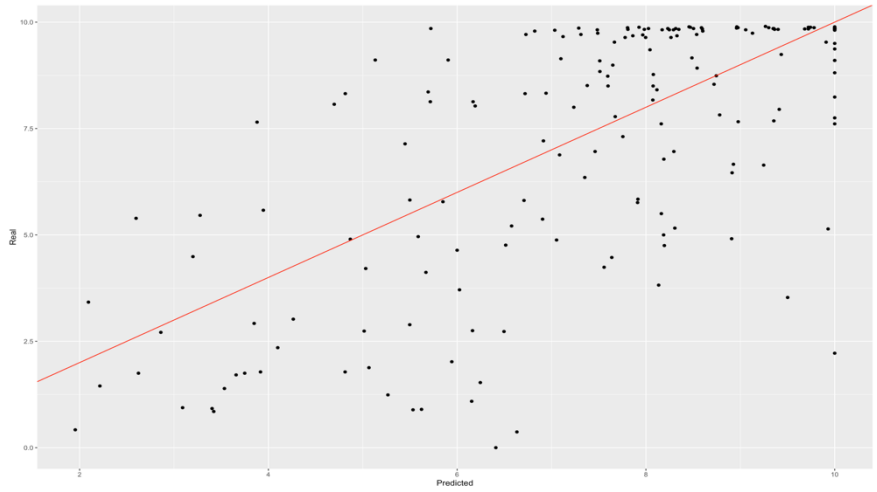


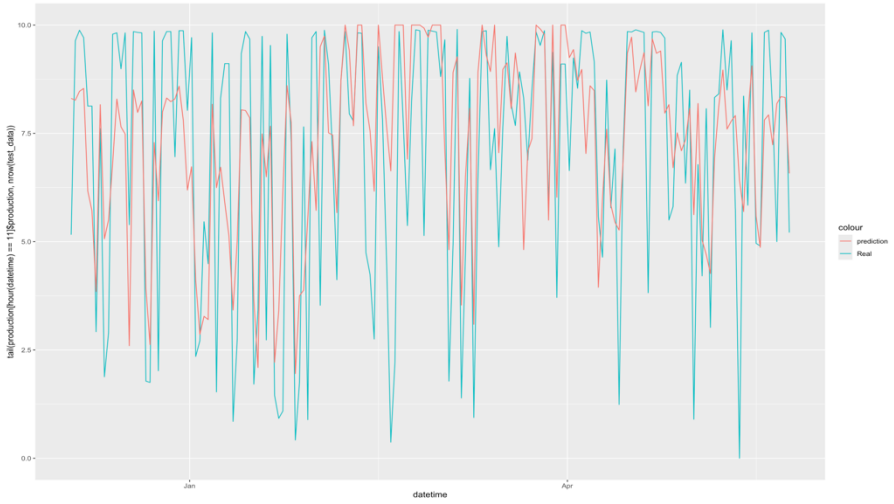
Rattle 2024-Jun-04 16:34:13 ahmetceliker

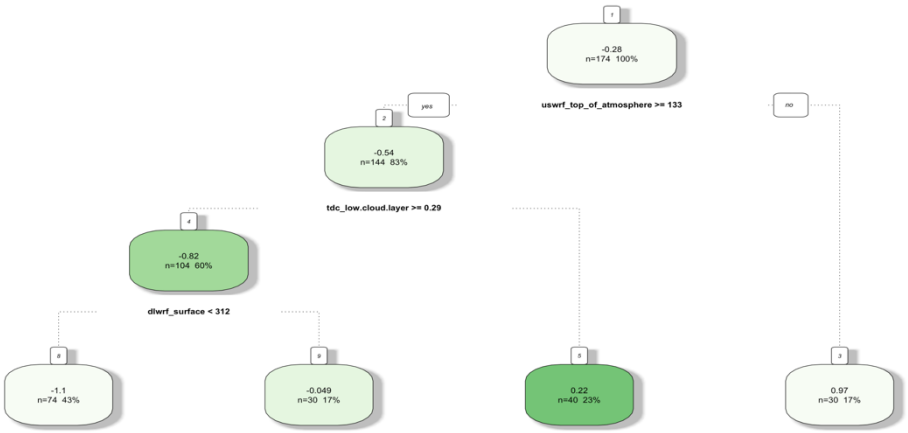
4.3.6.Hour 11

```
Call:
lm(formula = head(production[hour(datetime) == 11]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-9.2837 -0.6641  0.2532  1.2438  6.8233

Coefficients: (1 not defined because of singularities)
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                -40.944020   9.673859  -4.232 2.63e-05 ***
tdc_low.cloud.layer         -0.009527   0.006688  -1.424 0.154784
tdc_middle.cloud.layer      -0.009554   0.004142  -2.307 0.021370 *
tdc_high.cloud.layer         0.004745   0.004950   0.958 0.338171
tdc_entire.atmosphere       -0.012593   0.005721  -2.201 0.028060 *
uswrf_top_of_atmosphere      0.006985   0.002447   2.855 0.004435 **
csnow_surface               -0.852783   0.539034  -1.582 0.114104
dlwrf_surface               -0.026060   0.007247  -3.596 0.000346 ***
uswrf_surface               -0.002287   0.003066  -0.746 0.456057
tmp_surface                  0.179925   0.038636   4.657 3.86e-06 ***
dswrf_log                    0.711833   0.635563   1.120 0.263109
month_Jan_interaction_dswrf -0.082883   0.088864  -0.933 0.351309
month_Feb_interaction_dswrf  0.247510   0.104365   2.372 0.017991 *
month_Mar_interaction_dswrf -0.009964   0.121796  -0.082 0.934825
month_Apr_interaction_dswrf -0.300938   0.137625  -2.187 0.029110 *
month_May_interaction_dswrf -0.423505   0.146802  -2.885 0.004040 **
month_Jun_interaction_dswrf -0.350901   0.151832  -2.311 0.021126 *
month_Jul_interaction_dswrf -0.270467   0.159348  -1.697 0.090092 .
month_Aug_interaction_dswrf -0.395813   0.174685  -2.266 0.023775 *
month_Sep_interaction_dswrf -0.263504   0.153051  -1.722 0.085585 .
month_Oct_interaction_dswrf -0.165008   0.120796  -1.366 0.172392
month_Nov_interaction_dswrf -0.117020   0.093409  -1.253 0.210721
month_Dec_interaction_dswrf         NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.153 on 677 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.4656,    Adjusted R-squared:  0.449
F-statistic: 28.09 on 21 and 677 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 82.696, df = 26, p-value = 8.003e-08
```





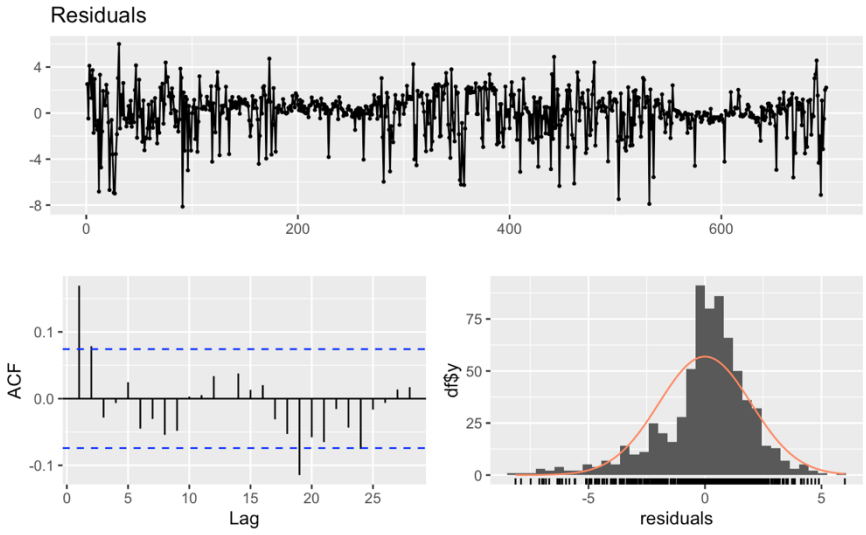


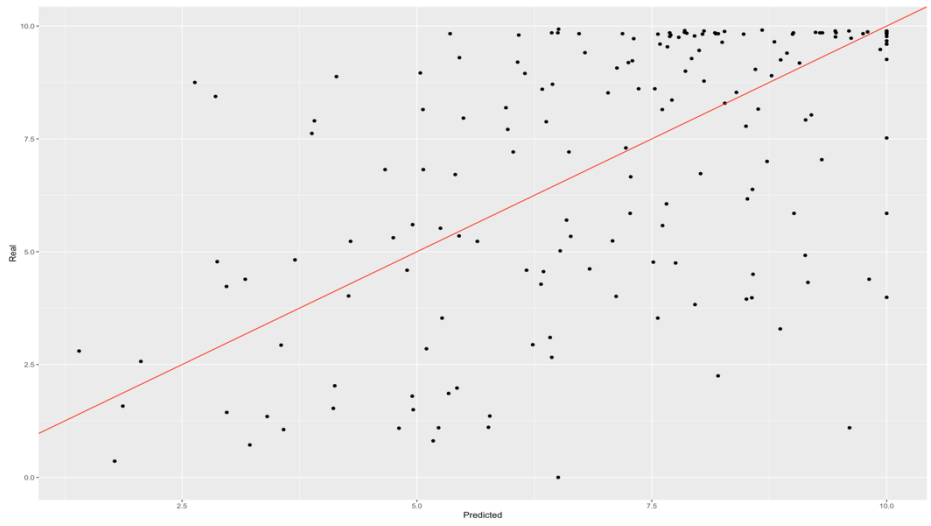


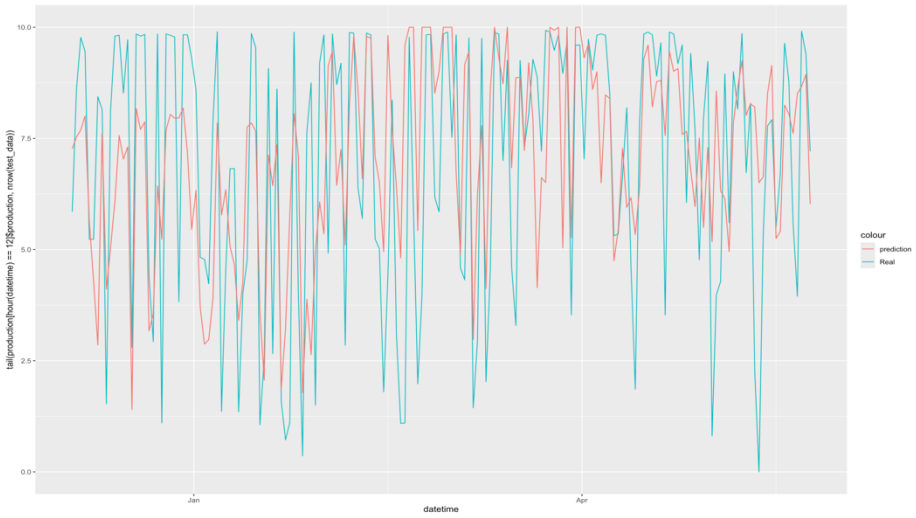
Rattle 2024-Jun-04 16:34:15 ahmetceliker

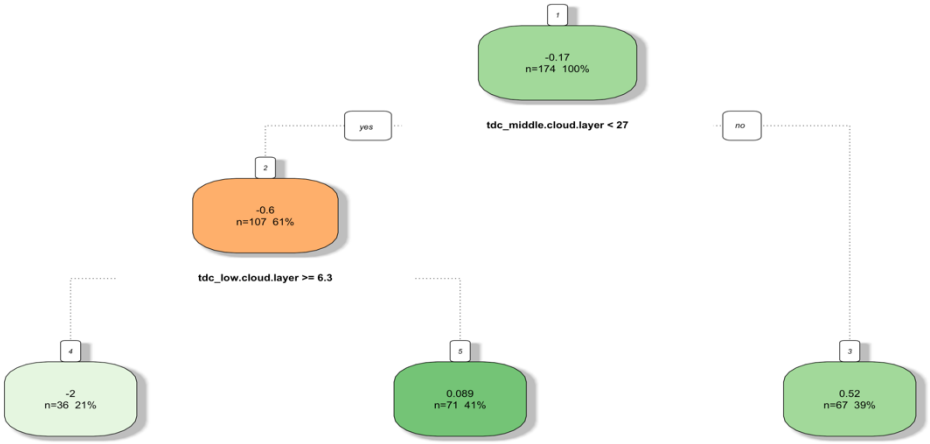
4.3.7.Hour 12

```
Call:
lm(formula = head(production[hour(datetime) == 12]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-8.1326 -0.6536  0.2610  1.1457  6.0026

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                -34.531806   9.626029  -3.587 0.000358 ***
tdc_low.cloud.layer         -0.011156   0.006537  -1.707 0.088364 .
tdc_middle.cloud.layer      -0.012986   0.004201  -3.091 0.002074 **
tdc_high.cloud.layer         0.004472   0.004900   0.913 0.361730
tdc_entire.atmosphere       -0.014579   0.005620  -2.594 0.009687 **
uswrf_top_of_atmosphere      0.004729   0.002159   2.190 0.028848 *
csnow_surface               -0.723164   0.479292  -1.509 0.131812
dlwrf_surface               -0.024782   0.006720  -3.688 0.000244 ***
uswrf_surface                0.001807   0.003432   0.526 0.598803
tmp_surface                  0.152566   0.037148   4.107  4.5e-05 ***
dswrf_log                    1.046461   0.700104   1.495 0.135454
month_Jan_interaction_dswrf -0.006323   0.004968  -1.273 0.203544
month_Feb_interaction_dswrf -0.001406   0.004149  -0.339 0.734694
month_Mar_interaction_dswrf -0.002441   0.003437  -0.710 0.477728
month_Apr_interaction_dswrf -0.004379   0.003030  -1.445 0.148926
month_May_interaction_dswrf -0.004354   0.002863  -1.521 0.128813
month_Jun_interaction_dswrf -0.004515   0.002862  -1.577 0.115184
month_Jul_interaction_dswrf -0.003612   0.003009  -1.200 0.230439
month_Aug_interaction_dswrf -0.004231   0.003314  -1.277 0.202193
month_Sep_interaction_dswrf -0.004172   0.003496  -1.193 0.233135
month_Oct_interaction_dswrf -0.004458   0.003821  -1.167 0.243665
month_Nov_interaction_dswrf -0.005185   0.004454  -1.164 0.244726
month_Dec_interaction_dswrf -0.005396   0.005212  -1.035 0.300949
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.01 on 676 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.5139,    Adjusted R-squared:  0.4981
F-statistic: 32.49 on 22 and 676 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 47.621, df = 26, p-value = 0.005974
```





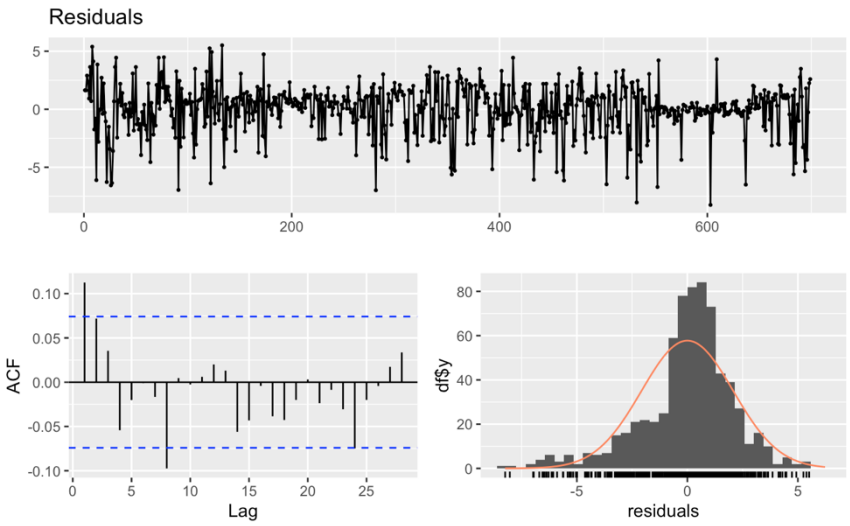


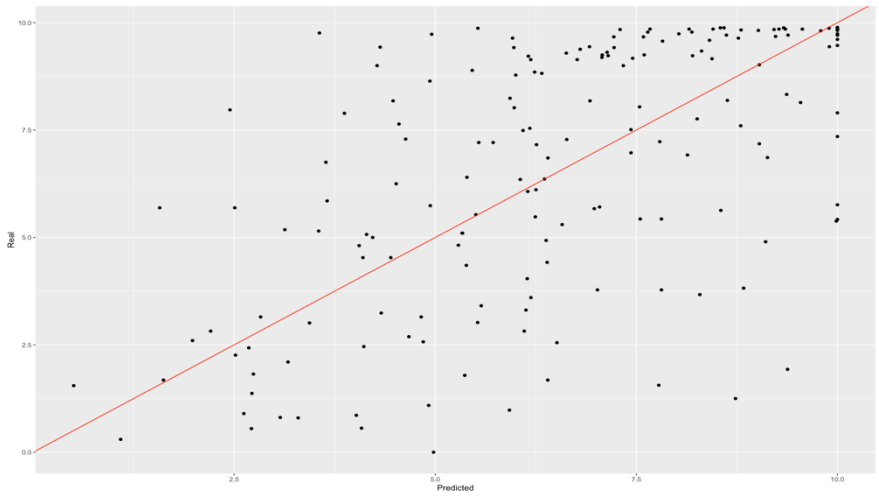


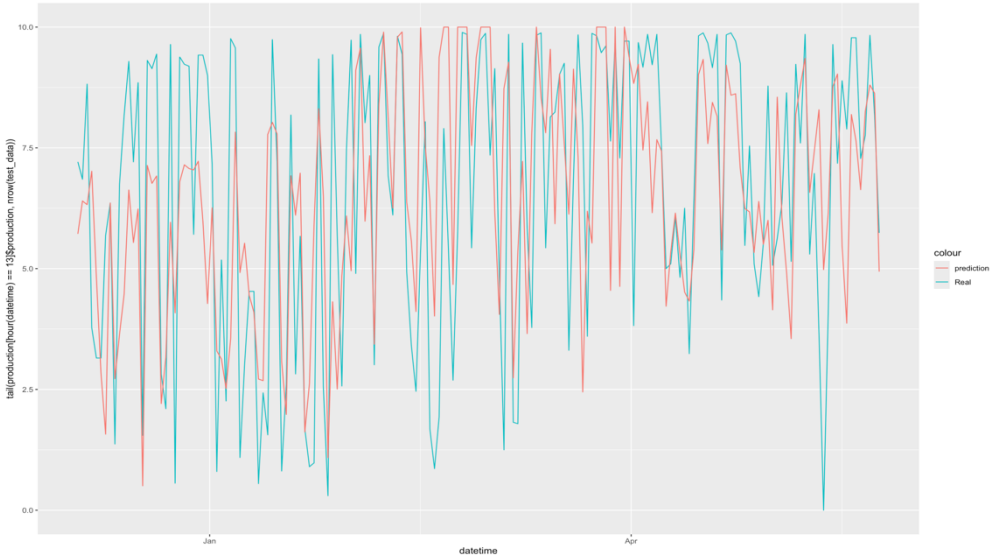
Rattle 2024-Jun-04 16:34:16 ahmetceliker

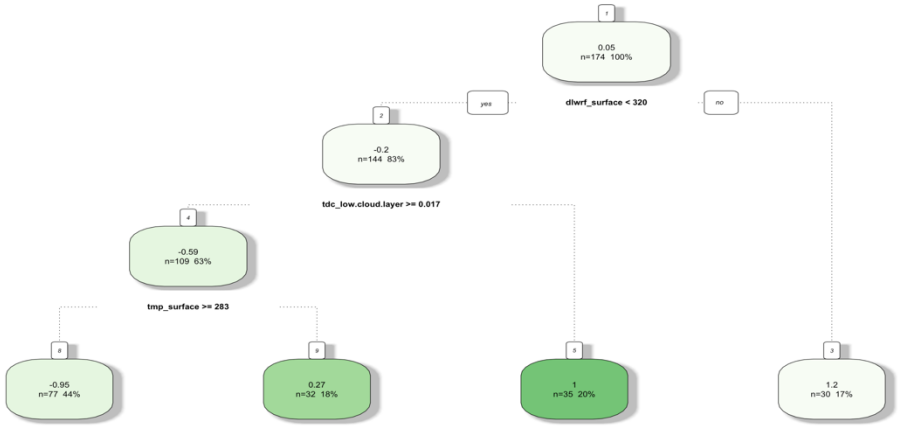
4.3.8.Hour 13

```
Call:
lm(formula = head(production[hour(datetime) == 13]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-8.2426 -0.7705  0.2458  1.1719  5.5106

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -2.752e+01  9.681e+00  -2.843 0.004604 **
tdc_low.cloud.layer         -1.193e-02  7.181e-03  -1.662 0.097031 .
tdc_middle.cloud.layer      -1.011e-02  4.586e-03  -2.205 0.027818 *
tdc_high.cloud.layer         2.476e-04  5.417e-03   0.046 0.963564
tdc_entire.atmosphere       -1.561e-02  6.196e-03  -2.519 0.011988 *
uswrf_top_of_atmosphere      3.680e-03  2.378e-03   1.548 0.122157
csnow_surface               -8.029e-01  5.112e-01  -1.571 0.116708
dlwrf_surface               -2.464e-02  6.797e-03  -3.625 0.000311 ***
uswrf_surface                1.489e-03  3.472e-03   0.429 0.668098
tmp_surface                  1.389e-01  3.688e-02   3.765 0.000181 ***
dswrf_log                    1.909e-01  7.693e-01   0.248 0.804050
month_Jan_interaction_dswrf -7.518e-04  5.032e-03  -0.149 0.881275
month_Feb_interaction_dswrf  3.218e-03  4.249e-03   0.757 0.449085
month_Mar_interaction_dswrf  1.431e-03  3.551e-03   0.403 0.687118
month_Apr_interaction_dswrf -1.061e-03  3.159e-03  -0.336 0.737021
month_May_interaction_dswrf -6.335e-04  2.994e-03  -0.212 0.832473
month_Jun_interaction_dswrf -6.692e-04  2.996e-03  -0.223 0.823347
month_Jul_interaction_dswrf -4.758e-06  3.135e-03  -0.002 0.998789
month_Aug_interaction_dswrf -7.228e-04  3.425e-03  -0.211 0.832907
month_Sep_interaction_dswrf -9.889e-05  3.612e-03  -0.027 0.978164
month_Oct_interaction_dswrf -8.253e-04  3.971e-03  -0.208 0.835406
month_Nov_interaction_dswrf -1.920e-03  4.622e-03  -0.415 0.677998
month_Dec_interaction_dswrf -2.468e-03  5.318e-03  -0.464 0.642749
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.109 on 676 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.5413,    Adjusted R-squared:  0.5263
F-statistic: 36.26 on 22 and 676 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 34.772, df = 26, p-value = 0.1167
```


Residuals
5
0
-5
0
200
400
600
ACF
0.10
0.05
0.00
-0.05
-0.10
0
5
10
15
20
25
Lag
df$y
80
60
40
20
0
-5
0
5
residuals


Real
Predicted
10.0
7.5
5.0
2.5
0.0

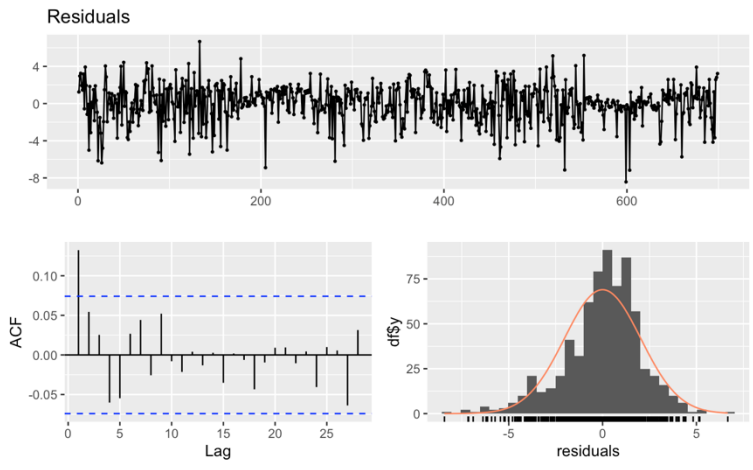

tail(production[hour(datetime) == 13]$production, nrow(test_data))
datetime
Jan
Apr
colour
prediction
Real

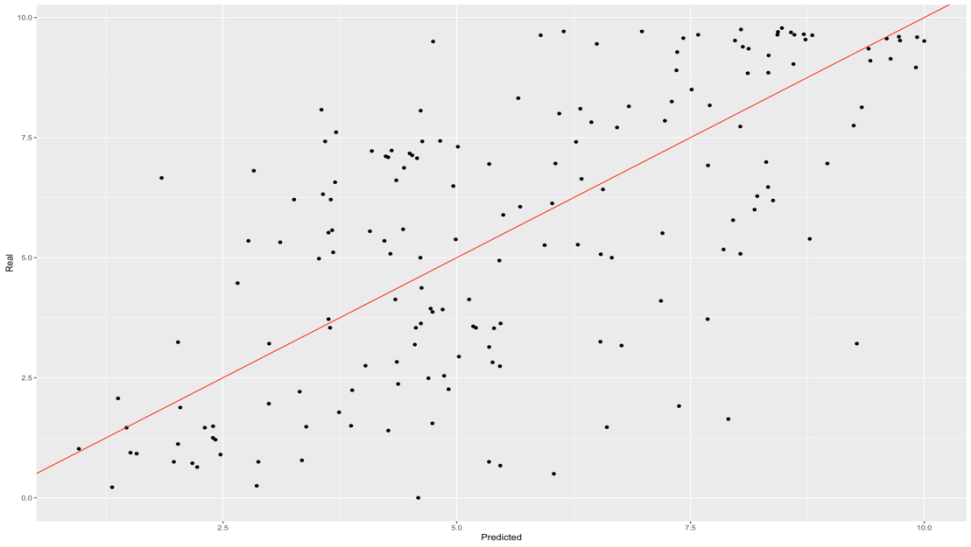


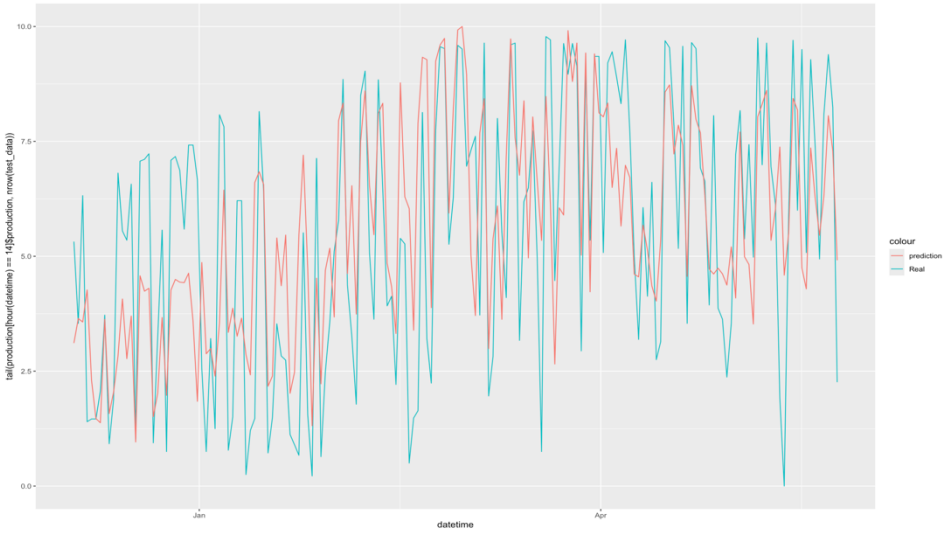
Rattle 2024-Jun-04 16:34:18 ahmetceliker

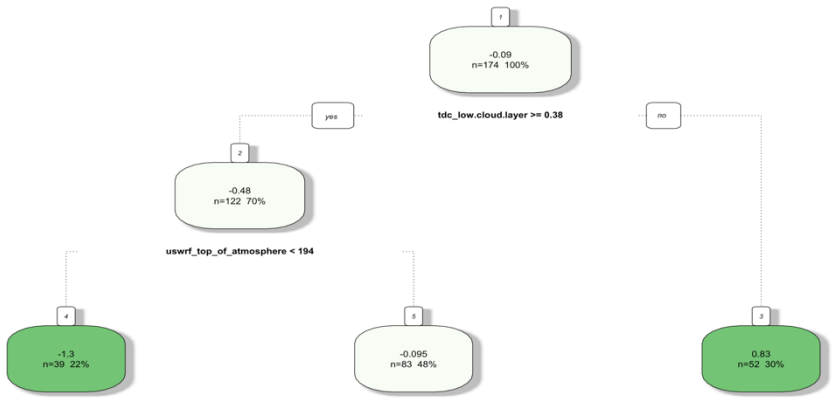
4.3.10.Hour 14

```
Call:
lm(formula = head(production[hour(datetime) == 14]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-8.4281 -0.9694  0.2062  1.3106  6.6717

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -16.924357   9.785496  -1.730 0.084170 .
tdc_low.cloud.layer           0.002153   0.007223   0.298 0.765784
tdc_middle.cloud.layer       -0.010348   0.004650  -2.225 0.026396 *
tdc_high.cloud.layer          0.002919   0.005416   0.539 0.590126
tdc_entire.atmosphere        -0.019414   0.006279  -3.092 0.002070 **
uswrf_top_of_atmosphere       0.004250   0.002392   1.777 0.076030 .
csnow_surface                -0.742957   0.514053  -1.445 0.148839
dlwrf_surface                -0.023355   0.006638  -3.518 0.000463 ***
uswrf_surface                -0.002197   0.003320  -0.662 0.508364
tmp_surface                   0.102432   0.035883   2.855 0.004441 **
dswrf_log                    -0.329335   0.795658  -0.414 0.679067
month_Jan_interaction_dswrf   0.003063   0.004917   0.623 0.533527
month_Feb_interaction_dswrf   0.006594   0.004195   1.572 0.116438
month_Mar_interaction_dswrf   0.005229   0.003552   1.472 0.141506
month_Apr_interaction_dswrf   0.003069   0.003181   0.965 0.334931
month_May_interaction_dswrf   0.003094   0.003025   1.023 0.306854
month_Jun_interaction_dswrf   0.002982   0.003029   0.984 0.325318
month_Jul_interaction_dswrf   0.004160   0.003160   1.317 0.188407
month_Aug_interaction_dswrf   0.003851   0.003433   1.122 0.262313
month_Sep_interaction_dswrf   0.003953   0.003618   1.093 0.274965
month_Oct_interaction_dswrf   0.002328   0.003986   0.584 0.559352
month_Nov_interaction_dswrf   0.001551   0.004621   0.336 0.737305
month_Dec_interaction_dswrf  -0.001711   0.005229  -0.327 0.743554
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.067 on 676 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:   0.54,     Adjusted R-squared:  0.5251
F-statistic: 36.08 on 22 and 676 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 28.826, df = 26, p-value = 0.319
```





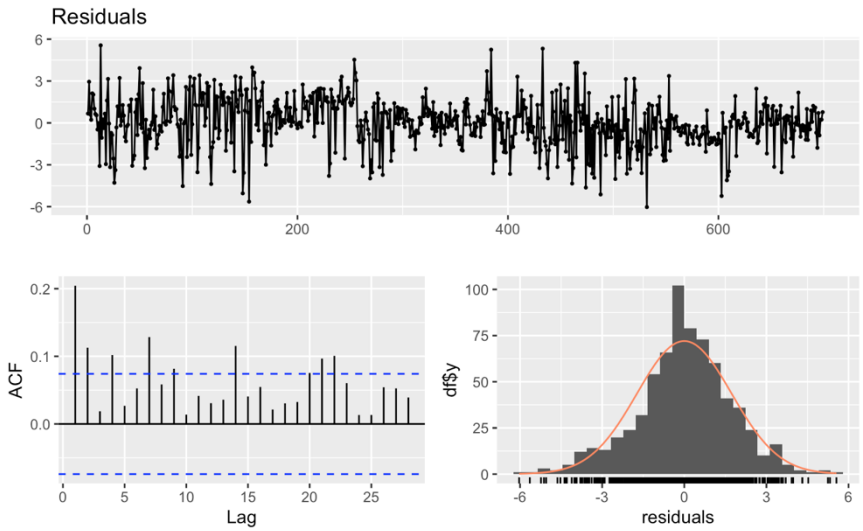


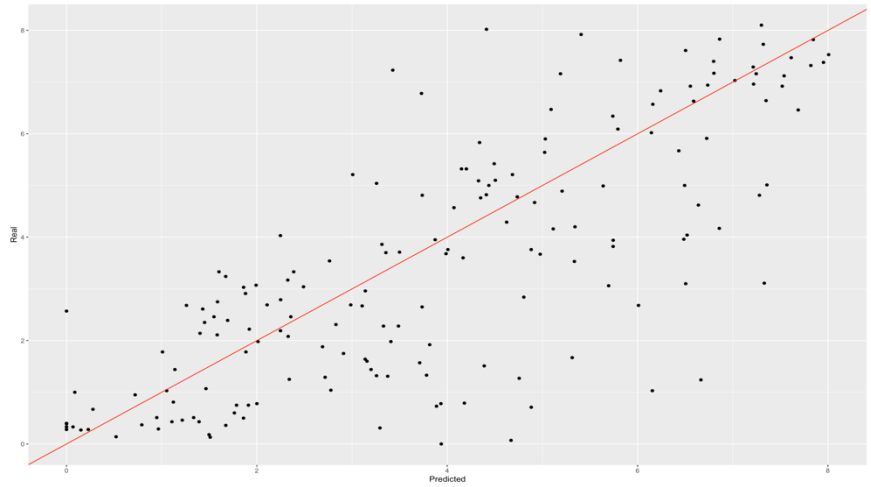


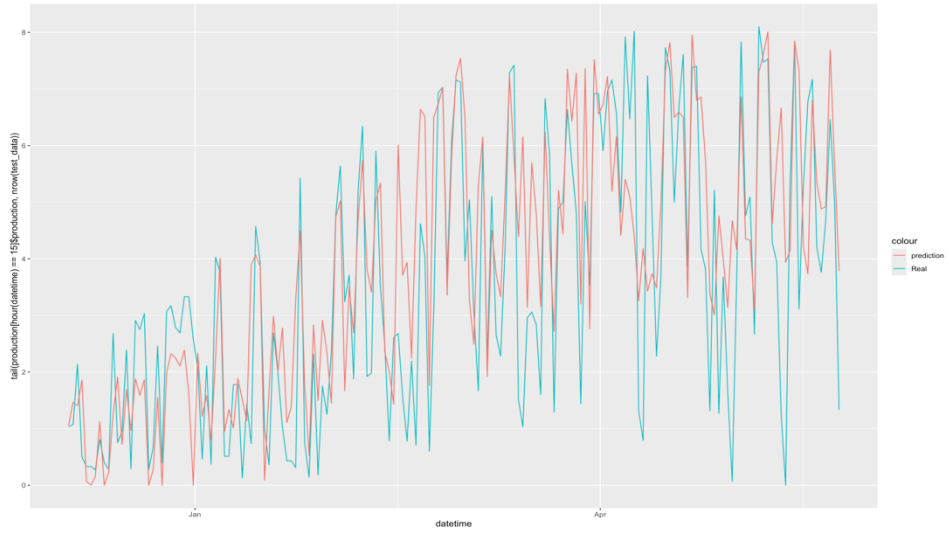
Rattle 2024-Jun-04 16:34:20 ahmetceliker

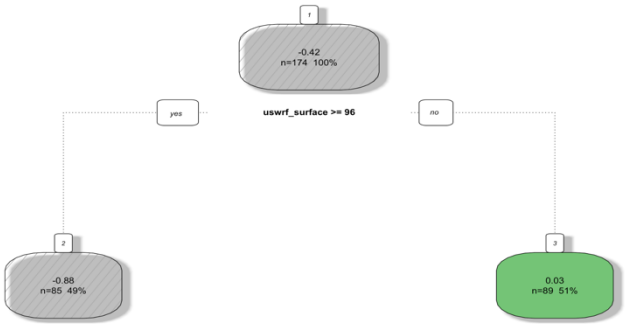
4.3.11.Hour 15

```
Call:
lm(formula = head(production[hour(datetime) == 15]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.0310 -0.9501  0.0047  1.0550  5.5568

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                -3.787e+01  8.432e+00  -4.491 8.35e-06 ***
tdc_low.cloud.layer         7.943e-03  6.073e-03   1.308 0.191306
tdc_middle.cloud.layer     -9.948e-03  3.999e-03  -2.488 0.013098 *
tdc_high.cloud.layer       -6.324e-04  4.650e-03  -0.136 0.891875
tdc_entire.atmosphere      -1.530e-02  5.508e-03  -2.777 0.005631 **
uswrf_top_of_atmosphere     7.498e-03  2.022e-03   3.708 0.000226 ***
csnow_surface               2.175e-01  4.326e-01   0.503 0.615242
dlwrf_surface              -2.748e-02  5.611e-03  -4.898 1.21e-06 ***
uswrf_surface              -6.651e-04  2.824e-03  -0.236 0.813862
tmp_surface                 1.808e-01  3.084e-02   5.861 7.20e-09 ***
dswrf_log                  -9.728e-01  6.919e-01  -1.406 0.160198
month_Jan_interaction_dswrf  4.727e-03  4.185e-03   1.129 0.259106
month_Feb_interaction_dswrf  6.521e-03  3.582e-03   1.820 0.069157 .
month_Mar_interaction_dswrf  5.628e-03  3.056e-03   1.842 0.065948 .
month_Apr_interaction_dswrf  4.324e-03  2.751e-03   1.572 0.116443
month_May_interaction_dswrf  4.971e-03  2.621e-03   1.897 0.058248 .
month_Jun_interaction_dswrf  4.248e-03  2.625e-03   1.618 0.106109
month_Jul_interaction_dswrf  4.076e-03  2.743e-03   1.486 0.137787
month_Aug_interaction_dswrf  3.670e-03  2.970e-03   1.236 0.217009
month_Sep_interaction_dswrf  3.334e-03  3.132e-03   1.065 0.287433
month_Oct_interaction_dswrf  1.776e-03  3.458e-03   0.513 0.607778
month_Nov_interaction_dswrf  1.230e-03  4.011e-03   0.307 0.759217
month_Dec_interaction_dswrf  4.537e-04  4.496e-03   0.101 0.919648
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.753 on 676 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.636,     Adjusted R-squared:  0.6241
F-statistic: 53.68 on 22 and 676 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 68.513, df = 26, p-value = 1.089e-05
```





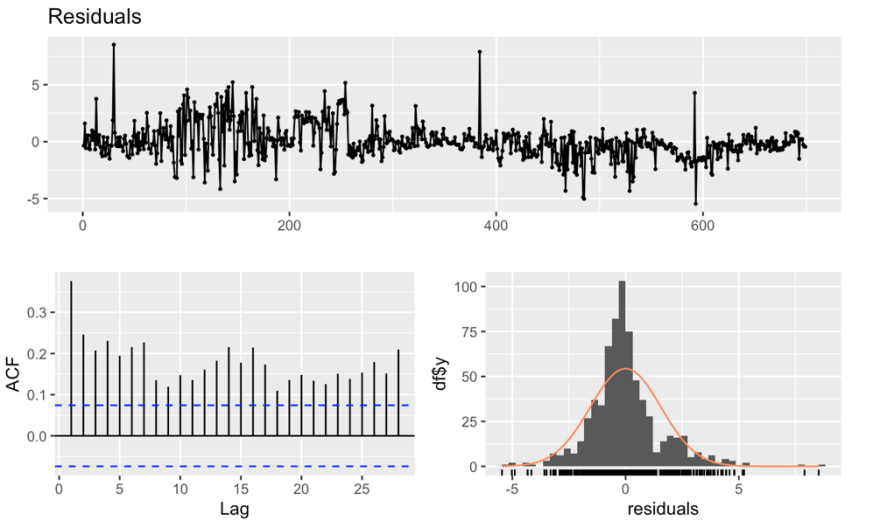


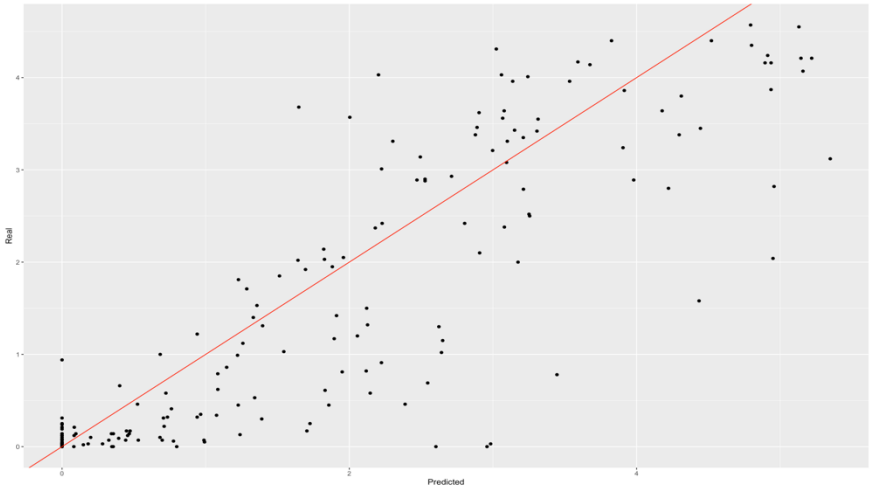


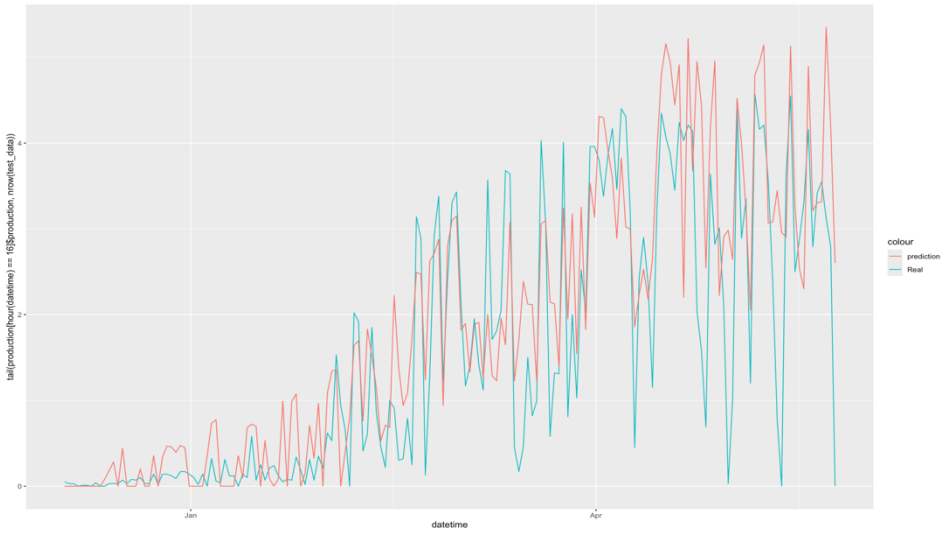
Rattle 2024-Jun-04 16:34:22 ahmetceliker

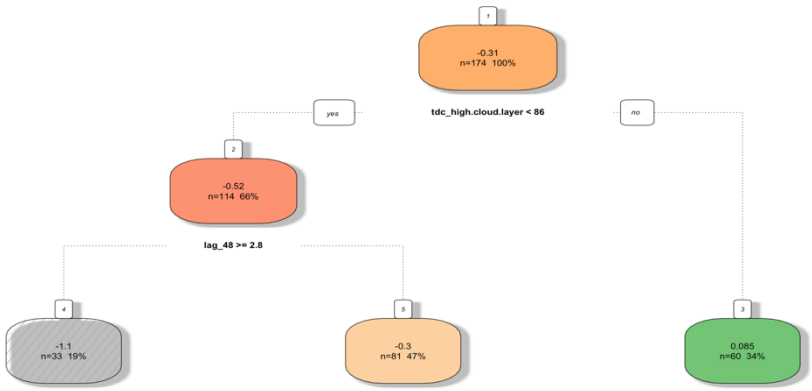
4.3.12.Hour 16

```
Call:
lm(formula = head(production[hour(datetime) == 16]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-5.4558 -0.7650 -0.1653  0.5916  8.5220

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                -1.812e+01  8.013e+00  -2.261 0.024087 *
tdc_low.cloud.layer         1.304e-03  3.580e-03   0.364 0.715869
tdc_middle.cloud.layer     -2.756e-03  2.656e-03  -1.037 0.299902
tdc_high.cloud.layer       -7.119e-03  2.622e-03  -2.715 0.006801 **
tdc_entire.atmosphere      -1.976e-03  3.676e-03  -0.538 0.591078
uswrf_top_of_atmosphere     6.349e-03  1.387e-03   4.578  5.6e-06 ***
csnow_surface              -3.087e-02  4.100e-01  -0.075 0.940003
dlwrf_surface              -6.902e-03  5.855e-03  -1.179 0.238847
uswrf_surface              -5.582e-04  2.763e-03  -0.202 0.839963
tmp_surface                 8.474e-02  3.250e-02   2.607 0.009326 **
dswrf_log                  -1.119e+00  4.167e-01  -2.686 0.007405 **
month_Jan_interaction_dswrf  8.341e-03  3.767e-03   2.214 0.027159 *
month_Feb_interaction_dswrf  8.470e-03  3.039e-03   2.787 0.005470 **
month_Mar_interaction_dswrf  7.647e-03  2.579e-03   2.966 0.003128 **
month_Apr_interaction_dswrf  8.481e-03  2.346e-03   3.615 0.000322 ***
month_May_interaction_dswrf  8.401e-03  2.248e-03   3.737 0.000202 ***
month_Jun_interaction_dswrf  7.437e-03  2.235e-03   3.327 0.000925 ***
month_Jul_interaction_dswrf  7.294e-03  2.289e-03   3.187 0.001504 **
month_Aug_interaction_dswrf  7.859e-03  2.541e-03   3.092 0.002068 **
month_Sep_interaction_dswrf  6.911e-03  2.798e-03   2.470 0.013767 *
month_Oct_interaction_dswrf  5.308e-03  3.384e-03   1.568 0.117276
month_Nov_interaction_dswrf  6.362e-03  4.183e-03   1.521 0.128807
month_Dec_interaction_dswrf  7.427e-03  4.617e-03   1.609 0.108191
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.584 on 676 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.5848,    Adjusted R-squared:  0.5713
F-statistic: 43.28 on 22 and 676 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 163.9, df = 26, p-value < 2.2e-16
```



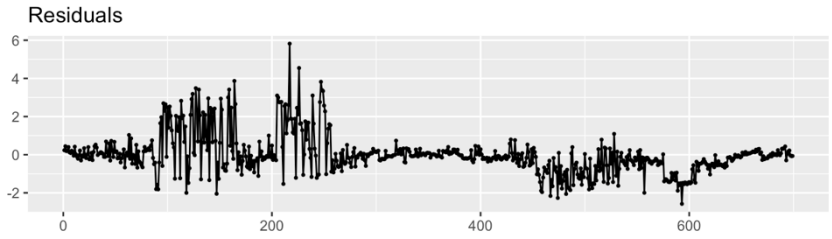


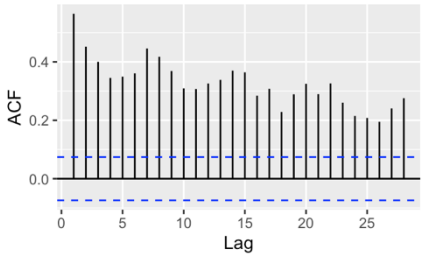


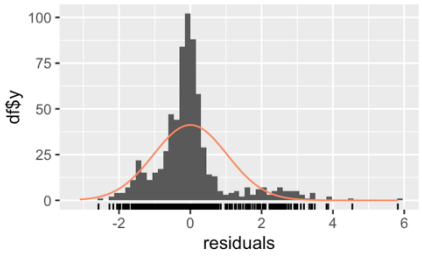


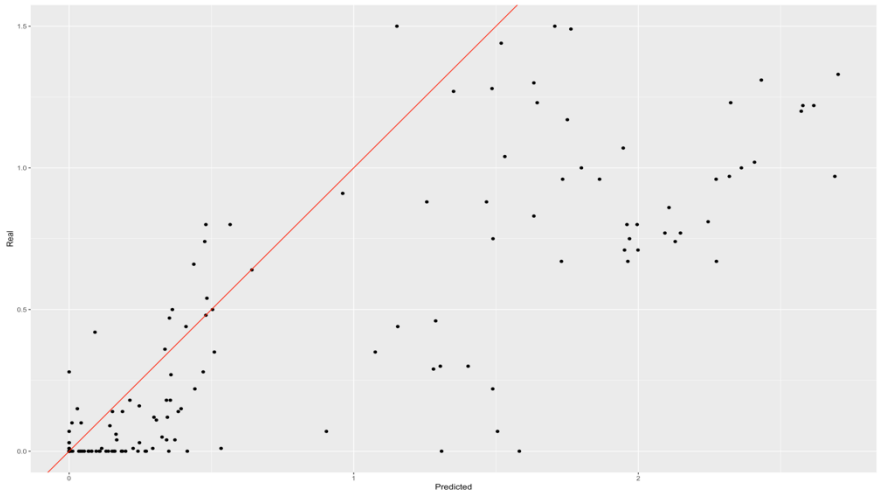
4.3.13.Hour 17

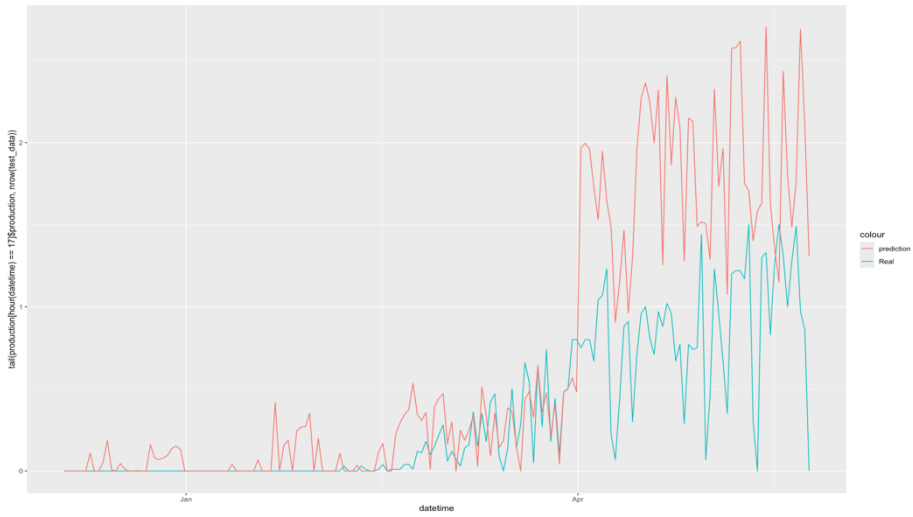
```
Call:
lm(formula = head(production[hour(datetime) == 17]$production,
    train_size) ~ . - lag_48 - datetime - mon - dswrf_surface,
    data = train_data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.5751 -0.4752 -0.0923  0.2022  5.8207

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               -6.5395332  5.5812516  -1.172 0.241732
tdc_low.cloud.layer       -0.0007274  0.0025292  -0.288 0.773734
tdc_middle.cloud.layer    -0.0009060  0.0018527  -0.489 0.624981
tdc_high.cloud.layer      -0.0047939  0.0018290  -2.621 0.008963 **
tdc_entire.atmosphere      0.0027834  0.0025589   1.088 0.277098
uswrf_top_of_atmosphere    0.0039462  0.0010015   3.940 8.99e-05 ***
csnow_surface              0.1925206  0.2759017   0.698 0.485550
dlwrf_surface             -0.0050477  0.0040676  -1.241 0.215053
uswrf_surface             -0.0017927  0.0021828  -0.821 0.411775
tmp_surface                0.0348597  0.0229318   1.520 0.128942
dswrf_log                 -0.6395808  0.2817935  -2.270 0.023541 *
month_Jan_interaction_dswrf  0.0053728  0.0028036   1.916 0.055741 .
month_Feb_interaction_dswrf  0.0040006  0.0022181   1.804 0.071742 .
month_Mar_interaction_dswrf  0.0035040  0.0018584   1.885 0.059792 .
month_Apr_interaction_dswrf  0.0059132  0.0016622   3.557 0.000401 ***
month_May_interaction_dswrf  0.0060449  0.0015921   3.797 0.000160 ***
month_Jun_interaction_dswrf  0.0050331  0.0015848   3.176 0.001561 **
month_Jul_interaction_dswrf  0.0047150  0.0016169   2.916 0.003662 **
month_Aug_interaction_dswrf  0.0059309  0.0018006   3.294 0.001040 **
month_Sep_interaction_dswrf  0.0042428  0.0020049   2.116 0.034695 *
month_Oct_interaction_dswrf  0.0037368  0.0025095   1.489 0.136938
month_Nov_interaction_dswrf  0.0057703  0.0032456   1.778 0.075873 .
month_Dec_interaction_dswrf  0.0070696  0.0036594   1.932 0.053785 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.05 on 676 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.4705,    Adjusted R-squared:  0.4533
F-statistic: 27.31 on 22 and 676 DF,  p-value: < 2.2e-16


        Breusch-Godfrey test for serial correlation of order up to 26

data:  Residuals
LM test = 323.34, df = 26, p-value < 2.2e-16
```

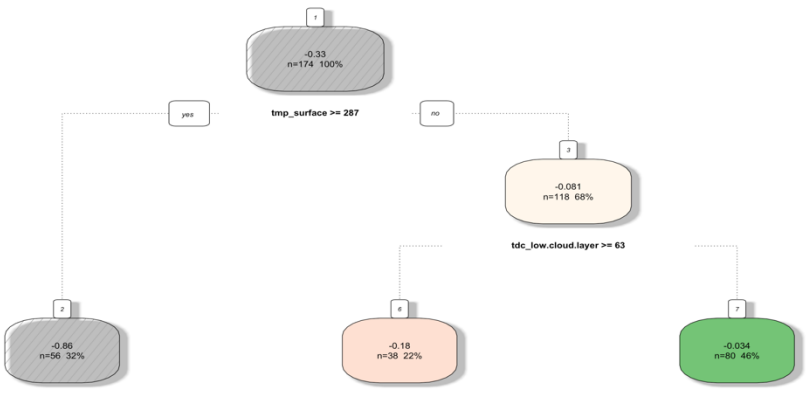












Rattle 2024-Jun-04 16:34:26 ahmetceliker

5. Baseline Method

At day t, we are trying to forecast the hourly production of dat t+1, with most recent production data being t-1. Therefore, we have evaluated the baseline method, which is the naïve forecast over the test period. The following code is used for evaluating the baseline method over the test period, and the model evaluation results are as following. It can be observed that none of our model performs worse than the baseline method when we make test-train split for our data.

```{r}
naïve_production <- production

naïve_production$production_lag2 <- lag(naïve_production$production, n = 48)

naïve_production <- naïve_production %>% filter(datetime >= as.Date('2024-02-01'))

naïve_production <- naïve_production %>% filter(datetime <= as.Date('2024-05-15'))

print(accu(naïve_production$production, naïve_production$production_lag2))
```

n <int>	mean_val <dbl>	sd_val <dbl>	CV <dbl>	FBias <dbl>	MAPE <dbl>	RMSE <dbl>	MAD <dbl>	MADP <dbl>	WMAPE <dbl>
2497	2.529127	3.468145	1.371282	0.005019611	NaN	2.178313	1.09163	0.4316232	0.4316232

6. Model Comparison

After we evaluate all the model accuracy results seperately, we have tried to combine our findings and evaluate the aggregated results. We have evaluated 4 different models in this step. It should be taken into account that since we were trying to forecast the hourly production of day t+1 at day t, by using the most recent known production amount t-1, we have tried to use the 48 hour lagged production amounts and appended them as the regressors. However, since the EPİAS was not able provide the t-1 production amounts correctly, we tried to use 72 hour lagged production amounts. However using 72 hour lagged production amounts did not improve our model so we did not use any lagged values as regressors in any of our final models.

6.1 Model with disaggregated clusters according to past data with sma for cluster 2

We have used 4 different clusters, one of them consisting of zero vectors, second one consisting of sma results and the third and fourth ones consisting of summed production amounts and summed regressors. To disaggregate the summed forecast data, we have divided hourly production to summed daily production amounts for each cluster and found the averages in train horizon. For cluster three, the hourly production ratios to summed cluster ratios for hours 06:00, 07:00, 08:00, 15:00, 16:00, 17:00 are 0.028560312, 0.150408073, 0.363920275, 0.300924708, 0.120982008, 0.035204624 respectively. For cluster four, the hourly production ratios to summed cluster ratios for hours 09:00, 10:00,

11:00, 12:00, 13:00, 14:00 are 0.163615062, 0.174684177, 0.176403689, 0.178151998, 0.166458302, 0.140686771 respectively. Besides of using only production data for finding ratios, we also tried some other combinations using the features. For instance we tried to multiply production ratios with dswrf_surface ratios and tmp_surface ratios and take the cubed root of the values in order to find different weights however, none of our trials yielded better results. By multiplying summed forecasts with this ratio for each cluster, we have formed our complete forecast model, and the test-train splitted model evaluation results are as given below:

```
test_horizon_forecasts <- data.frame(
  date = as.Date(production_sma_hour4$datetime),
  hour_0 = zero_vector,
  hour_1 = zero_vector,
  hour_2 = zero_vector,
  hour_3 = zero_vector,
  hour_4 = production_sma_hour4$sma_forecasts,
  hour_5 = production_sma_hour5$sma_forecasts,
  hour_6 = results_cluster_3$V1,
  hour_7 = results_cluster_3$V2,
  hour_8 = results_cluster_3$V3,
  hour_9 = results_cluster_4$V1,
  hour_10 = results_cluster_4$V2,
  hour_11 = results_cluster_4$V3,
  hour_12 = results_cluster_4$V4,
  hour_13 = results_cluster_4$V5,
  hour_14 = results_cluster_4$V6,
  hour_15 = results_cluster_3$V4,
  hour_16 = results_cluster_3$V5,
  hour_17 = results_cluster_3$V6,
  hour_18 = production_sma_hour18$sma_forecasts,
  hour_19 = zero_vector,
  hour_20 = zero_vector,
  hour_21 = zero_vector,
  hour_22 = zero_vector,
  hour_23 = zero_vector
)
```

n <int>	mean_val <dbl>	sd_val <dbl>	CV <dbl>	FBias <dbl>	MAPE <dbl>	RMSE <dbl>	MAD <dbl>	MADP <dbl>	WMAPE <dbl>
2736	2.568615	3.481532	1.355412	-0.005831485	NaN	1.26608	0.6693058	0.2605707	0.2605707

From the model evaluation results provided above, it can be observed that our WMAPE result have been found as 0.2605. It can be concluded that, since our WMAPE results for cluster 3 and cluster 4 by themselves were lower, this disaggregation process was not sufficient for converting summed data to hourly data. One of the main reasons for this is the cyclic behavior in production amounts. Since the lengths of the sunshine times are not constant throughout the year and there exists non-constant cyclic weather behaviors, using constant ratios for test horizon did not yield expected results. Thus, we tried to find and use

different ratios for each month however, since the production amounts have been differing due to the weather conditions and instead of observing significant weather conditions, cyclic behaviors where the lengths of the frequencies are not constant were being observed, even we tried different ratios, we did not get any significant improvement. Thus, we have moved to the following model.

6.2 Model with disaggregated clusters according to hourly production forecast data with sma for cluster 2

We indicated that using clustered and summed productions and regressors were yielding better forecast results in model evaluation part. To disaggregate the summed production amounts, we used the hourly forecast models, divided each hourly production forecast to summed production forecasts for each cluster. We tried this for both cluster 3 and cluster 4 but obtained better results when it is applied to only cluster 4. Here is the following code and the model evaluation results:

```
test_horizon_forecasts <- data.frame(
  date = as.Date(production_sma_hour4$datetime),
  hour_0 = zero_vector,
  hour_1 = zero_vector,
  hour_2 = zero_vector,
  hour_3 = zero_vector,
  hour_4 = production_sma_hour4$sma_forecasts,
  hour_5 = production_sma_hour5$sma_forecasts,
  hour_6 = results_cluster_3$V1,
  hour_7 = results_cluster_3$V2,
  hour_8 = results_cluster_3$V3,
  hour_9 = forecasts_c4 * aggregate_c4$hour_9_normalized,
  hour_10 = forecasts_c4 * aggregate_c4$hour_10_normalized,
  hour_11 = forecasts_c4 * aggregate_c4$hour_11_normalized,
  hour_12 = forecasts_c4 * aggregate_c4$hour_12_normalized,
  hour_13 = forecasts_c4 * aggregate_c4$hour_13_normalized,
  hour_14 = forecasts_c4 * aggregate_c4$hour_14_normalized,
  hour_15 = results_cluster_3$V4,
  hour_16 = results_cluster_3$V5,
  hour_17 = results_cluster_3$V6,
  hour_18 = production_sma_hour18$sma_forecasts,
  hour_19 = zero_vector,
  hour_20 = zero_vector,
  hour_21 = zero_vector,
  hour_22 = zero_vector,
  hour_23 = zero_vector
)
```

n	mean_val	sd_val	CV	FBias	MAPE	RMSE	MAD	MADP	WMAPE
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2736	2.568615	3.481532	1.355412	-0.005831486	NaN	1.268448	0.666033	0.2592966	0.2592966

Compared to the previous model, slight improvement could be observed in this version of the model however it can not be concluded that this disaggregation method also works sufficiently well enough. Since the most of the problem was arising from disagregation of the production forecast, we decided to try combined versions of summed clusters and hourly productions.

6.3 Model with disaggregated clusters and hourly production forecasts with sma for cluster 2

We have tried to use hourly production forecasts combined with clustered summed forecasts and evaluate the results. We have tried different combinations of hourly production forecast and disaggregated forecast and the best model with model evaluation results are as given below.

```
test_horizon_forecasts <- data.frame(
  date = as.Date(production_sma_hour4$datetime),
  hour_0 = zero_vector,
  hour_1 = zero_vector,
  hour_2 = zero_vector,
  hour_3 = zero_vector,
  hour_4 = production_sma_hour4$sma_forecasts,
  hour_5 = production_sma_hour5$sma_forecasts,
  hour_6 = results_cluster_3$V1,
  hour_7 = results_cluster_3$V2,
  hour_8 = results_cluster_3$V3,
  hour_9 = results_hour_9,
  hour_10 = results_hour_10,
  hour_11 = results_hour_11,
  hour_12 = results_hour_12,
  hour_13 = results_hour_13,
  hour_14 = results_hour_14,
  hour_15 = results_cluster_3$V4,
  hour_16 = results_cluster_3$V5,
  hour_17 = results_cluster_3$V6,
  hour_18 = production_sma_hour18$sma_forecasts,
  hour_19 = zero_vector,
  hour_20 = zero_vector,
  hour_21 = zero_vector,
  hour_22 = zero_vector,
  hour_23 = zero_vector
)
```

n	mean_val	sd_val	CV	FBias	MAPE	RMSE	MAD	MADP	WMAPE
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2736	2.568615	3.481532	1.355412	-0.002520383	NaN	1.2646	0.6659809	0.2592763	0.2592763

It can be observed that we were unable to reach less WMAPE value by this method compared to the other two methods.

6.4 Model with hourly production forecasts with sma for cluster 2

Best results for hourly production have been found by taking the logarithm of dswrf_surface data and appending monthly interaction regressors multiplied with logged dswrf_surface data as we have discussed above. The code for combining the hourly productions and the sma for cluster 2 with the model evaluation results are as given below:

```
test_horizon_forecasts <- data.frame(
  date = as.Date(production_sma_hour4$datetime),
  hour_0 = zero_vector,
  hour_1 = zero_vector,
  hour_2 = zero_vector,
  hour_3 = zero_vector,
  hour_4 = production_sma_hour4$sma_forecasts,
  hour_5 = production_sma_hour5$sma_forecasts,
  hour_6 = results_hour_6,
  hour_7 = results_hour_7,
  hour_8 = results_hour_8,
  hour_9 = results_hour_9,
  hour_10 = results_hour_10,
  hour_11 = results_hour_11,
  hour_12 = results_hour_12,
  hour_13 = results_hour_13,
  hour_14 = results_hour_14,
  hour_15 = results_hour_15,
  hour_16 = results_hour_16,
  hour_17 = results_hour_17,
  hour_18 = production_sma_hour18$sma_forecasts,
  hour_19 = zero_vector,
  hour_20 = zero_vector,
  hour_21 = zero_vector,
  hour_22 = zero_vector,
  hour_23 = zero_vector
)
```

n	mean_val	sd_val	CV	FBias	MAPE	RMSE	MAD	MADP	WMAPE
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2736	2.568615	3.481532	1.355412	0.002900262	NaN	1.235857	0.6504011	0.2532108	0.2532108

It can be observed that for hourly model, model evaluation results yielded less WMAPE value compared to the other three models. Thus, our final model can be stated as hourly production forecast model with logged dswrf_surface data multiplied with monthly dummy variables.

7. Submission Phase, Conclusion and Future Works

At the beginning of the submission phase, we have been using the forecast results of 6.1 model. However, we have observed that the disaggregation process of the summed production forecasts did not work well due to the need of change in ratios caused by weather cycles and the daily weather fluctuations causing to hourly production to fluctuate. We have observed that in some days, where high amounts of production are expected, our model 6.1 was underpredicting the production amounts where there should be high amounts of production. We have interpreted the result that these hourly underestimations have been occurring since our model was not able to explain the low amounts of productions at peak times. Therefore, as future work, the cause of low amount of production should be understood and appended to the model. Besides, we have tried to find new disaggregation method to be able to catch-up with real life production by considering the hourly weather fluctuations. Therefore, we have multiplied summed production data with ratios that we obtained using hourly production forecasts, model 6.2. We have made some of our submissions by using model 6.2 and some submissions with different combinations of all models however, after trying different combinations of clustered and hourly production, we have decided to use our last model, 6.4 which is the hourly production model as our final model. Besides, for some days, we have uploaded hourly production forecasts by using additional regressor, 48 hour lagged production data, however; the production data provided by EPİAS was not always correct for day t-1 and it has been getting updated one day later. Therefore, we have stopped using lags as predictors. Our final Model, 6.4 yields 0.2532 WMAPE value for test period, which is better compared to baseline method WMAPE value 0.4316. Trying to understand unexpected drops in production amounts would be important to considered for future work. Besides, working on the weather condition cycles and trying to establish a model where the model is trained in a similar weather conditioned time horizons might also improve the model since the effects of weather conditions depends also on hour, date, and sunshining time of the day. Both of our Python and R code can be reached from the provided links. The following results are the forecasted production data for our final model 6.4.

The model can be enhanced by exploring additional strategies and modeling techniques. Below are some future directions that may lead to improvements:

- In the exploration of alternative methods for hourly modeling strategies, ARIMA (AutoRegressive Integrated Moving Average) and SARIMA (Seasonal AutoRegressive Integrated Moving Average) models should be considered. These

time series models are well-suited for capturing temporal dependencies and seasonality in hourly data, potentially offering improved accuracy for certain hours.

- Adding features related to maximum capacity and interactions with average regressors could enhance model performance. Future work should include creating new features such as max capacity and (max capacity) * (average of each regressor) for each hour, integrating these features into existing models, and analyzing their significance and contribution.
- Furthermore, incorporating other dummy variables or nonlinear relations could help mitigate underprediction issues, improving the model's accuracy.
- To enhance the model, the effects of variables on real-life electricity production can be analyzed, and variable values can be utilized in various combinations and formulas.

8. Appendicies

date	hour_4	hour_5	hour_6	hour_7	hour_8	hour_9	hour_10	hour_11	hour_12	hour_13	hour_14	hour_15	hour_16	hour_17
<date>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2024-02-01	0.000000000	0.00000000	0.16823611	1.5370016	4.021870	7.208184	8.635785	8.674019	8.275941	8.239401	7.353114	4.529163	1.3310241	0.000000000
2024-02-02	0.000000000	0.00000000	0.23304824	1.8394037	4.638437	8.259311	9.071214	9.029159	8.625059	8.712699	7.780081	4.837861	1.3238793	0.000000000
2024-02-03	0.000000000	0.00000000	0.14683412	1.3881338	3.839541	6.472395	6.809715	6.786220	6.114199	5.676330	4.257273	1.525303	0.0754381	0.000000000
2024-02-04	0.000000000	0.00000000	0.16799868	0.7826381	1.793952	4.509587	5.590103	6.320257	6.502834	6.468359	5.704358	3.086372	0.5329338	0.000000000
2024-02-05	0.000000000	0.00000000	0.04139180	0.2434697	1.231149	3.665580	4.313401	4.699145	4.343264	2.676588	2.926958	2.407716	0.6880112	0.071171939
2024-02-06	0.000000000	0.00000000	0.05417423	0.2773216	1.144830	4.017930	7.175218	7.771831	7.326082	7.430889	6.777460	4.249062	1.5149124	0.000000000
2024-02-07	0.000000000	0.00000000	0.35960764	2.0284683	4.915249	8.614980	9.320170	9.333971	8.724429	8.870604	7.850270	5.277395	1.6013600	0.000000000
2024-02-08	0.000000000	0.00000000	0.39088969	2.0648232	4.770925	8.267300	8.758110	8.445727	7.606035	7.291680	5.875909	3.492459	0.8317232	0.000000000
2024-02-09	0.000000000	0.00000000	0.37200113	1.9482617	4.660010	7.507657	6.752599	6.899191	6.193838	5.914995	5.010997	3.042904	1.6936776	0.000000000
2024-02-10	0.000000000	0.00000000	0.33586607	1.9477377	5.030878	8.681652	9.367223	9.376785	8.716115	8.767299	7.342825	4.570831	1.4551363	0.000000000
2024-02-11	0.000000000	0.00000000	0.27690122	1.5661751	3.947934	7.101291	9.270709	9.337944	8.671500	8.884838	7.534704	4.793668	1.1659498	0.000000000
2024-02-12	0.000000000	0.00000000	0.32783362	1.7470927	3.760078	6.525782	7.518393	7.369108	6.543258	6.018887	4.455626	2.105905	0.5926984	0.000000000
2024-02-13	0.000000000	0.00000000	0.19703463	1.0587629	3.281535	5.693529	6.602127	6.735934	6.002175	5.217259	3.926430	1.764983	0.7446738	0.000000000
2024-02-14	0.000000000	0.00000000	0.16112438	0.3394143	1.220153	3.867902	4.898268	5.202399	4.525593	3.678015	2.774685	1.215803	0.7222474	0.100908260
2024-02-15	0.000000000	0.00000000	0.26927842	1.4926582	4.431597	7.825949	9.295312	9.412070	8.702053	8.935017	7.955154	5.455617	2.0497813	0.099548349
2024-02-16	0.000000000	0.00000000	0.24216825	1.4966343	4.508838	7.270862	7.913699	7.812649	7.009600	6.861336	5.575897	3.151933	1.3205097	0.000000000
2024-02-17	0.000000000	0.00000000	0.30403088	0.7729584	2.352465	4.885545	6.073821	6.459218	5.530863	5.519449	5.180771	3.328611	0.8189091	0.000000000
2024-02-18	0.000000000	0.00000000	0.35274642	1.8432670	4.551715	7.513638	5.670093	5.114772	3.819731	3.516633	2.825545	1.875164	0.9699854	0.210030359
2024-02-19	0.000000000	0.00000000	0.30690717	1.6578521	4.844318	8.232550	9.334601	9.381722	8.509673	8.307787	7.009206	4.322627	1.5677141	0.265423675
2024-02-20	0.000000000	0.00000000	0.37730837	1.7432285	4.704494	8.336448	9.881504	10.000000	9.155001	9.354144	8.393159	5.991707	2.3087518	0.288105178
2024-02-21	0.000000000	0.00000000	0.22269231	1.4778620	4.904539	7.670187	9.722891	9.839309	9.048221	9.393302	8.379903	5.889298	2.2632723	0.314111792
2024-02-22	0.000000000	0.00000000	0.25105407	1.0654297	3.317516	5.220935	4.872281	5.787779	4.948369	4.207548	3.314818	1.377775	1.1402896	0.515295917
2024-02-23	0.000000000	0.00000000	0.35101941	1.8565466	5.764608	8.673248	9.725479	9.883172	9.091084	9.406207	8.303560	5.812823	2.4166289	0.267285510
2024-02-24	0.000000000	0.00000000	0.39060735	2.1540823	6.712764	9.993330	10.000000	10.000000	9.287157	9.684364	8.656857	6.072883	2.5001650	0.216836766
2024-02-25	0.000000000	0.00000000	0.38392371	2.0313822	6.770631	9.938193	10.000000	10.000000	9.387270	9.850301	8.799376	6.326148	2.6507548	0.257529051
2024-02-26	0.000000000	0.00000000	0.28212956	1.7619217	6.109806	8.582819	8.606350	8.884516	7.582747	6.959373	5.374076	2.998315	0.9547538	0.029718275
2024-02-27	0.000000000	0.00000000	0.23846497	1.1252522	3.876229	5.945063	7.580277	8.718234	8.053431	8.453371	7.313528	5.045609	2.5606890	0.293964790
2024-02-28	0.000000000	0.00000000	0.44999215	2.2447509	7.209476	10.000000	10.000000	10.000000	9.576999	10.000000	8.939744	6.499585	2.8497712	0.325705027
2024-02-29	0.000000000	0.00000000	0.37983276	2.0181552	7.096024	10.000000	10.000000	10.000000	9.663956	10.000000	9.088571	6.753851	2.9007398	0.354326827
2024-03-01	0.000000000	0.00000000	0.43641691	2.1644004	6.818881	9.418689	9.829098	9.809113	9.824335	9.816300	8.771223	6.066690	1.8913846	0.220474490
2024-03-02	0.000000000	0.00000000	0.31001017	1.6457168	4.864964	6.490040	6.731016	6.932820	6.720495	6.412108	5.353757	2.982829	1.7975428	0.354582279
2024-03-03	0.000000000	0.00000000	0.31107519	0.6377361	1.547133	2.804441	3.386752	3.929968	4.806711	4.061725	3.678717	2.308972	1.2630359	0.004606404
2024-03-04	0.000000000	0.00000000	0.32740913	1.6502860	5.774004	7.603998	7.638764	8.354554	8.550455	8.180655	7.408616	4.753329	1.7337355	0.259605382
2024-03-05	0.000000000	0.00000000	0.45308290	2.2461531	7.045986	9.251895	8.892588	8.744548	8.826761	8.706367	8.189459	5.698868	1.9067824	0.247915376
2024-03-06	0.000000000	0.00000000	0.20346736	0.6219362	2.302246	3.592650	2.950370	2.845405	2.684032	2.960395	2.825821	1.921140	1.2656116	0.240923770
2024-03-07	0.000000000	0.00000000	0.38943007	1.5133438	5.087792	7.249260	6.702368	5.477928	5.282233	4.855099	4.993937	4.086285	1.9178037	0.370414870
2024-03-08	0.000000000	0.00000000	0.27008797	1.7149723	6.302809	8.180374	7.995222	7.701910	7.585537	7.107599	6.005908	3.386644	1.3782986	0.121928002
2024-03-09	0.000000000	0.00000000	0.13743583	0.4217537	1.755575	2.419004	2.704707	2.522568	3.731623	3.747493	3.633883	3.196656	1.1026818	0.467767102
2024-03-10	0.000000000	0.00000000	0.16270278	0.7464043	3.581070	5.527488	8.848253	8.572354	8.118099	7.334425	6.941976	4.680798	1.8643069	0.354132254
2024-03-11	0.000000000	0.00000000	0.43159061	2.3123316	7.625336	9.739974	10.000000	10.000000	9.938015	10.000000	9.499923	6.740427	1.7158388	0.183855175
2024-03-12	0.000000000	0.00000000	0.38456760	2.0270448	7.070259	8.790521	8.613048	8.776487	8.678036	8.043559	7.283294	5.318773	2.9235832	0.313500040

date	hour_4	hour_5	hour_6	hour_7	hour_8	hour_9	hour_10	hour_11	hour_12	hour_13	hour_14	hour_15	hour_16	hour_17
<date>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2024-04-23	0.000000000	0.32666667	0.77749809	4.5825523	6.143004	8.250060	9.288022	9.432377	9.194344	8.822574	7.820721	6.623619	4.6057113	1.777820999
2024-04-24	0.000000000	0.33666667	0.99241388	4.9237914	6.183467	7.992766	7.882289	8.112168	7.879929	7.446290	6.488317	5.467145	4.0991008	1.707849888
2024-04-25	0.000000000	0.41333333	1.02691825	4.1613174	5.172902	7.020163	8.490450	8.209628	7.921360	6.739045	4.956397	3.332300	2.4544894	1.055811677
2024-04-26	0.000000000	0.47000000	1.20519538	3.9956266	5.355912	7.000017	7.390943	6.971349	7.196633	6.761255	4.994747	2.957149	3.8820410	1.710836817
2024-04-27	0.000000000	0.50333333	1.57719358	4.4266139	5.646414	7.485523	7.823052	7.627747	6.377489	5.887523	4.933009	4.599790	4.5872699	1.676974759
2024-04-28	0.000000000	0.43000000	0.93244501	3.6206643	5.182053	6.485069	7.295380	7.314546	7.809725	6.840618	4.854292	3.856832	2.1535721	1.194740317
2024-04-29	0.000000000	0.48666667	2.10356756	4.1128975	5.718817	7.137158	7.305606	7.495847	5.840398	6.090301	4.623832	3.160978	2.6961148	1.284895695
2024-04-30	0.000000000	0.47000000	2.50085936	4.3374813	5.155105	6.966030	7.856308	8.106948	7.519996	6.412126	5.355335	4.462132	2.7013650	1.223374221
2024-05-01	0.000000000	0.60000000	0.88266446	3.5801195	5.563902	6.466688	7.094859	5.826369	5.308187	4.474874	4.281378	4.002152	2.4334318	1.110914631
2024-05-02	0.000000000	0.57666667	1.12254765	4.5466336	6.723835	8.067494	8.643341	8.191173	8.235690	8.415848	7.782491	6.629914	4.2577683	2.005548790
2024-05-03	0.000000000	0.52333333	1.11076075	3.8415958	5.917015	6.781219	5.735702	4.967913	6.006749	6.176991	5.141036	4.416541	3.7621768	1.520427787
2024-05-04	0.000000000	0.55000000	1.16172213	4.1538967	6.216748	7.422773	5.619157	4.691714	6.068869	5.172921	4.993215	4.249378	2.9919738	1.700141158
2024-05-05	0.000000000	0.56000000	0.90416047	2.6628941	4.428818	4.831838	4.431643	4.047834	4.919514	3.849816	3.467459	3.270161	2.0741512	0.996119503
2024-05-06	0.003333333	0.60666667	1.20046070	3.8280439	6.313219	6.739939	6.036912	6.846273	7.534494	8.019442	8.027753	7.104605	4.6231803	2.289357822
2024-05-07	0.003333333	0.52666667	1.30130930	4.5345150	7.455252	8.536468	8.480486	8.065775	8.287658	8.634496	8.345281	7.406325	4.7160112	2.264016592
2024-05-08	0.006666667	0.58666667	1.39967813	5.0412166	7.694839	9.164208	9.187619	8.881151	8.875466	9.210123	8.679190	7.768294	4.8744955	2.268787484
2024-05-09	0.010000000	0.63000000	1.49365605	4.9659293	7.541933	8.790269	7.975369	7.527058	7.680717	6.766089	5.581427	4.614794	2.9888623	1.564213827
2024-05-10	0.020000000	0.63000000	1.48482091	4.9822151	7.292200	8.608596	8.100718	7.667348	7.887980	7.437917	6.106601	5.500483	2.9478183	1.492753375
2024-05-11	0.026666667	0.61000000	1.58095066	5.0100891	7.491260	8.694159	8.326022	7.957583	7.915569	8.186021	7.475132	6.421569	3.3356471	1.228742143
2024-05-12	0.036666667	0.62666667	1.38252402	3.5266831	5.485741	6.347854	6.542123	6.188834	6.198947	4.931753	4.525823	3.859105	2.7857222	1.382666483
2024-05-13	0.036666667	0.67333333	1.31743082	2.7243186	4.991091	5.257191	4.736292	5.327074	6.386287	6.049099	5.390721	4.046679	2.7426309	1.477895551
2024-05-14	0.026666667	0.42333333	1.14497981	2.9172340	5.342204	6.158456	8.081218	7.842422	8.161196	8.608745	8.453805	7.615949	4.9084690	2.380090687
2024-05-15	0.030000000	0.46666667	1.22182301	3.5781039	6.425863	7.288496	8.934462	8.900868	8.772546	8.923429	8.260891	7.045263	3.1512582	1.437729023

date	hour_4	hour_5	hour_6	hour_7	hour_8	hour_9	hour_10	hour_11	hour_12	hour_13	hour_14	hour_15	hour_16	hour_17
<date>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
2024-03-13	0.000000000	0.00000000	0.38577769	1.9216776	6.561408	8.114941	8.365532	8.477251	8.160054	7.486715	6.604419	4.083960	1.2822217	0.188718714
2024-03-14	0.000000000	0.00000000	0.36249891	1.7124910	5.971899	8.081295	9.507467	9.710067	9.416346	9.045726	8.230159	5.586090	1.6216927	0.216066227
2024-03-15	0.000000000	0.00000000	0.26851816	0.9329434	3.695396	4.806778	5.905818	6.629190	6.656884	5.896111	4.816261	2.819995	2.2443943	0.388024488
2024-03-16	0.000000000	0.00000000	0.42697760	2.3029971	7.410310	9.279657	8.367444	8.136370	7.963622	8.590130	7.782856	5.049856	1.9667280	0.378281860
2024-03-17	0.000000000	0.00000000	0.34208728	1.7003003	5.904029	7.814708	8.500677	8.573794	8.260916	7.126809	6.264360	4.320272	2.0325019	0.178174768
2024-03-18	0.000000000	0.00000000	0.34548929	2.1536797	6.678818	7.874800	7.820808	7.530036	6.975954	5.993081	5.125504	2.828570	1.2106572	0.000000000
2024-03-19	0.000000000	0.00000000	0.28061806	2.0777882	6.021983	7.513264	8.746459	8.881490	8.660879	8.614323	8.174862	5.666353	2.9400307	0.428631587
2024-03-20	0.000000000	0.00000000	0.37998294	2.7202433	7.168349	8.424333	8.062905	8.152688	7.785628	7.312065	6.095645	3.922279	3.0047245	0.481034180
2024-03-21	0.000000000	0.01000000	0.49288544	1.8863138	3.145378	4.235344	3.391848	4.682534	4.257838	2.776919	2.797653	2.697177	2.0625137	0.335600617
2024-03-22	0.000000000	0.02666667	0.11838430	2.4095209	6.764109	8.819234	7.990959	6.479640	6.058152	5.945246	5.814015	4.689745	1.9992904	0.661318374
2024-03-23	0.000000000	0.02666667	0.39527812	3.4981310	7.712486	9.193773	6.591718	6.617842	6.045748	5.094661	5.481397	3.944345	1.3391446	0.347796118
2024-03-24	0.000000000	0.01666667	0.27728756	3.4484643	7.215560	8.830739	9.931651	10.000000	10.000000	10.000000	9.609823	6.776460	3.1316753	0.478107991
2024-03-25	0.000000000	0.01666667	0.34521750	3.7711155	7.369053	9.045694	9.349612	9.479783	9.395047	9.515262	8.505528	5.830672	1.9815733	0.289658516
2024-03-26	0.000000000	0.02666667	0.19302729	1.7708840	3.832188	5.648420	9.424676	9.734403	9.747571	9.934042	9.554466	6.845313	3.1163776	0.474483855
2024-03-27	0.000000000	0.04000000	0.09957748	1.4329364	4.144917	4.540905	4.350604	4.979072	4.974009	4.628824	4.839178	2.909157	1.5077802	0.064957919
2024-03-28	0.000000000	0.02333333	0.49129914	5.0025634	8.229942	9.773368	9.637748	9.942770	9.583707	9.709718	9.149772	6.654086	3.0892581	0.437067525
2024-03-29	0.000000000	0.01333333	0.45310634	2.7069708	4.311575	4.878694	4.555692	5.446446	5.188245	4.696607	3.962190	2.513068	1.6741187	0.483484957
2024-03-30	0.000000000	0.02666667	0.47376178	5.8256697	8.463723	10.000000	10.000000	10.000000	9.978375	9.797887	9.095510	6.769619	3.3504199	0.513448974
2024-03-31	0.000000000	0.04333333	0.47388918	5.6237861	7.517988	9.034645	9.207474	9.911250	9.505651	8.957339	7.813864	5.819767	3.0301483	0.461681101
2024-04-01	0.000000000	0.08000000	0.55823992	3.6040645	6.701481	8.655444	9.418448	9.266591	9.288638	8.884489	8.145037	6.621814	4.0505103	1.544058186
2024-04-02	0.000000000	0.08333333	0.66274042	4.0783798	7.271996	9.325734	9.481643	9.435682	9.577612	9.285924	8.457390	7.089603	4.0637404	1.571387635
2024-04-03	0.000000000	0.10000000	0.63187468	3.8137458	6.543869	8.354592	9.050992	8.659125	8.554965	7.533965	6.531458	5.075071	3.7452532	1.554708383
2024-04-04	0.000000000	0.10333333	0.60187076	3.8462393	6.604874	8.344582	8.955070	8.943908	8.973979	8.515809	7.408152	6.011349	3.4542809	1.391462574
2024-04-05	0.000000000	0.11666667	0.73531593	4.2809721	6.876109	8.268009	7.406287	7.144015	6.754868	6.500288	5.823324	4.317974	2.8229863	1.256032984
2024-04-06	0.000000000	0.12333333	0.60453588	4.2218706	7.204275	8.805076	9.028441	8.598611	8.476577	7.699669	6.936732	5.323736	3.6995991	1.588784954
2024-04-07	0.000000000	0.13666667	0.50679664	3.7922443	6.806652	8.060200	8.249448	8.404037	8.365464	7.478355	6.668995	5.025465	2.9946339	1.387122094
2024-04-08	0.000000000	0.14000000	0.16998303	0.8121663	2.239928	3.115746	4.535711	4.133612	4.898223	4.622331	4.795470	4.426218	2.9016227	1.234987370
2024-04-09	0.000000000	0.15000000	0.37753900	3.4717000	6.163881	7.102277	6.365969	6.189132	5.717351	5.730983	4.660537	3.334687	1.8520876	0.741209340
2024-04-10	0.000000000	0.12000000	0.37275635	1.4585852	3.745818	5.028311	8.122718	7.578523	7.302707	6.209768	5.675579	4.075875	2.1415823	0.906572718
2024-04-11	0.000000000	0.16333333	0.40715249	3.1045373	5.015260	6.262070	6.327308	5.838576	5.928077	5.399883	5.127034	3.389920	2.3401685	1.200743435
2024-04-12	0.000000000	0.62333333	0.56263814	3.3144651	5.190285	6.336143	5.490173	5.430884	6.021426	4.694427	4.422881	3.755854	2.0954259	0.746535818
2024-04-13	0.000000000	0.66000000	0.40687117	1.8611808	3.035950	4.358250	5.059397	5.254176	5.570701	4.880207	4.315696	3.467932	2.4590332	1.131170239
2024-04-14	0.000000000	0.64666667	0.55691901	4.2313308	6.344897	7.876310	7.365271	6.802356	6.344867	5.428086	5.257182	4.867054	3.6886763	1.587627354
2024-04-15	0.000000000	0.19333333	0.57315741	4.6395644	6.989570	8.806514	9.191123	9.378035	9.346394	9.087042	8.618825	7.218387	4.5339668	1.823046555
2024-04-16	0.000000000	0.26000000	0.58334774	4.7891156	6.944733	9.061185	9.446766	9.715993	9.628483	9.425745	8.811295	7.653913	4.8240733	1.865196542
2024-04-17	0.000000000	0.27000000	0.61292855	4.2851801	6.263647	8.007579	8.200296	8.449303	8.354312	7.802189	7.273599	6.285976	4.6244782	1.782535476
2024-04-18	0.000000000	0.29000000	0.76062472	4.9975944	6.796333	8.823871	8.920515	9.062761	8.877379	8.558591	7.940422	6.433268	4.1359851	1.633813530
2024-04-19	0.000000000	0.36333333	0.68637937	4.9718363	6.818337	8.951668	9.285930	9.391637	8.853718	8.275006	7.457765	6.310636	4.6166247	1.845761865
2024-04-20	0.000000000	0.41000000	0.69871178	4.8539893	6.546956	8.469885	8.054222	8.146444	7.698525	5.795950	4.793901	3.402562	2.1668417	1.040792905
2024-04-21	0.000000000	0.46333333	0.60739555	4.6537607	6.187012	8.361324	9.536720	9.686875	9.505663	9.309665	8.781279	7.783034	4.9071772	1.930022658
2024-04-22	0.000000000	0.35000000	0.59160479	4.7204305	6.388602	8.564851	9.134487	9.346892	9.072680	8.707502	8.038369	6.626978	3.5424384	1.538226318