Data Analysis & Inference

Acea Smart Water Analytics Time Series Analysis

Authors' Name	Student ID
Islam Anwar	201900337
Aliyu Bello	201900104

FACULTY OF SCIENCE & INNOVATION,
THE UNIVERSITIES OF CANADA

May 23, 2026

Abstract

—

Time series analysis is a vital aspect of data analysis exploration. The analysis of variables through time is important to identify underlying impacts or shocks that can impact observation values. Time series analysis is performed on the Acea Smart Water Analytics data set, which observes the effect of features such as rainfall, temperature and depth to groundwater through time. The data is preprocessed through stationarization, transformations, differencing. Auto-correlation analysis is performed prior to modeling under auto-ARIMA models and Prophet models. Results show that auto-ARIMA best represents and forecasts the data set.

Time Series Analysis — Auto-correlation Analysis — auto-ARIMA — Forecasting

Contents

1 Introduction

Time series analysis is a specific method of analyzing a sequence of observations collected across a certain interval of time. Univariate time series are a very common form of time series where there is only one dependent variable affected by several independent features. Multivariate time series focuses on multiple dependent variables each affected by their respective independent variables.

Our problem of choice is performing a time series analysis for the Acea Smart Water data set. The problem presented by this data set is to estimate the influence of features on the water availability of several water bodies (lakes, rivers, aquifers). By forecasting accurate values of water availability per day, models should be able to effectively capture volumes of water in each waterbody present.

The Acea Smart Water Analytics is comprised of 9 independent data sets, each representing a certain water body with differing features that impact each respective water body. We will be considering the petrignano waterbody for this analysis. This waterbody is of type aquifer, which is an underground layer of permeable rocks such as sand or gravel. The aquifer waterbody is influenced by the following features: rainfall, humidity, subsoil, temperatures, and drainage volumes.

This paper covers various statistical methods and techniques that will be employed to perform sufficient time series analysis of our data. Statistical techniques such as data visualization of our dependent variable through time. Data preprocessing to ensure our data is clean and ready to be used for modeling and further exploration. Furthermore, feature engineering will be deployed to identify the most significant features affecting water availability, as well as exploratory data analysis using visuals and specific tests for stationarity and other properties of time series data. Finally, we will perform modeling using a healthy variety of time series models to interpret and identify which time series model is most suitable for our data.

2 Statistical Methods & Analysis

This section covers an array of statistical methods and analyses performed on our time series data set. Analyses including data visualization, preprocessing procedure, feature engineering for identifying the most impactful features affecting our time series, as well as exploratory data analysis, and finally the modelling process.

2.1 Data Visualization

Our first aspect of analysis is data visualization. Fig. (1) displays the time series in a data frame format including the feature columns as well as the target variable. Our independent variables are as follows: rainfall, temperature, drainage volume (volume of water taken from drinking water treatment plant), and river hydrometry (groundwater level in m). Our dependent variable is the depth to groundwater, which is the groundwater level from the ground floor in metres.

	date	rainfall	depth_to_groundwater	temperature	drainage_volume	river_hydrometry
0	2009-01-01 00:00:00	0.000000	-31.140000	5.200000	-24530.688000	2.400000
1	2009-01-02 00:00:00	0.000000	-31.110000	2.300000	-28785.888000	2.500000
2	2009-01-03 00:00:00	0.000000	-31.070000	4.400000	-25766.208000	2.400000
3	2009-01-04 00:00:00	0.000000	-31.050000	0.800000	-27919.296000	2.400000
4	2009-01-05 00:00:00	0.000000	-31.010000	-1.900000	-29854.656000	2.300000

Figure 1: Data Overview

We plotted the time series plot of each column including our target variable to visualize if there was any trend or pattern through time. Figs. (2-6) represents the time series plot for all of our columns of the dataset. Fig. (2) shows that there is a series of peaks and troughs throughout the period from 2009 to 2020 with numerous number of high amounts of rainfall. This could potentially indicate seasonality but an independent test will provide us with a more definitive result.

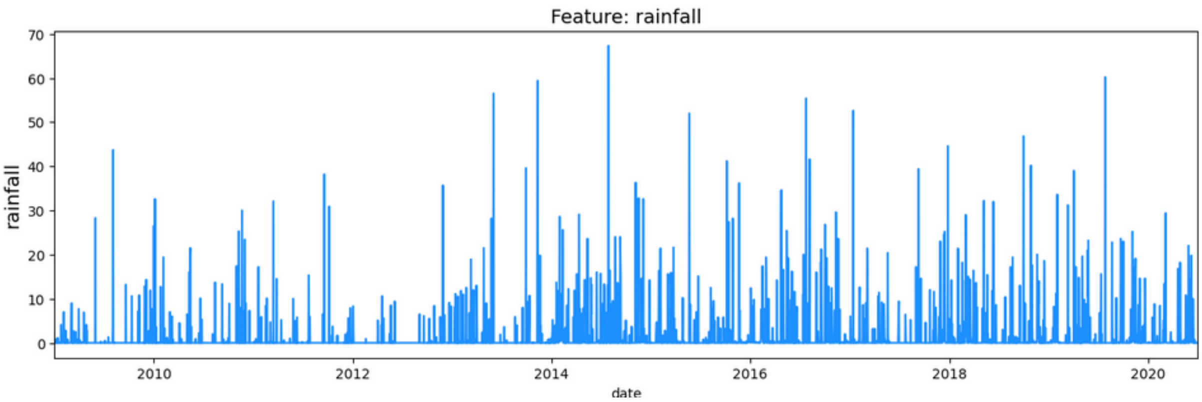

Figure 2: Rainfall TS Plot

Regarding fig. (3), the depth to groundwater plots a fluctuating trend in its values over time. Furthermore, the y-axis values it is encompassing indicate a non-stationarity in the mean and variance, which is something that may need to be remedied later on.

Concerning Fig. (4), we instantly notice a consistent set of peaks and troughs throughout several years. There is a high probability of seasonality for our temperature variable, which is to be expected. Regarding stationarity, the center of the data points is not centered around 0. This indicates a non-stationarity in the mean, however, this can be remedied. But the plot shows stationarity in the variance, which is promising.

Fig. (5) seems to show a much more interesting plot, with a somewhat consistent trend through time. However, there may be some potential outliers around 2011 and 2012 with some more near 2019. However, without these outliers, there may be stationarity in the

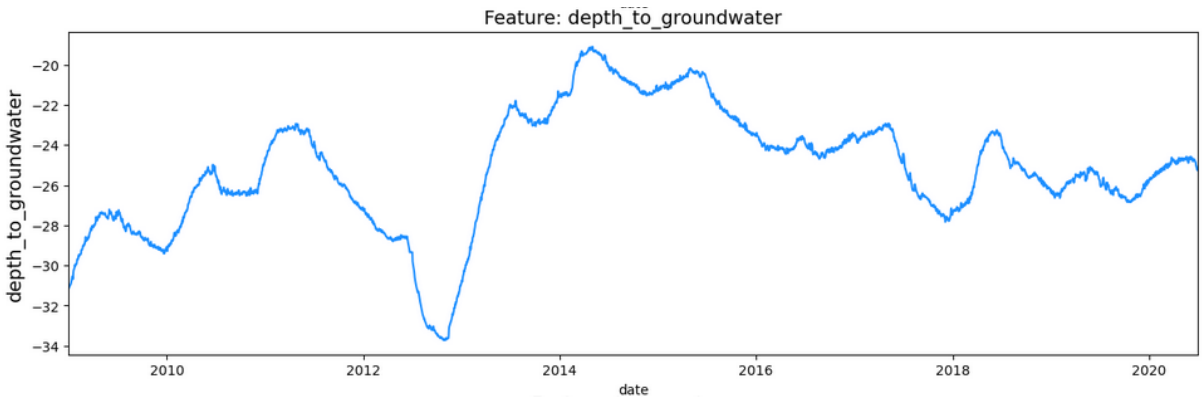

Figure 3: Depth to Groundwater TS Plot

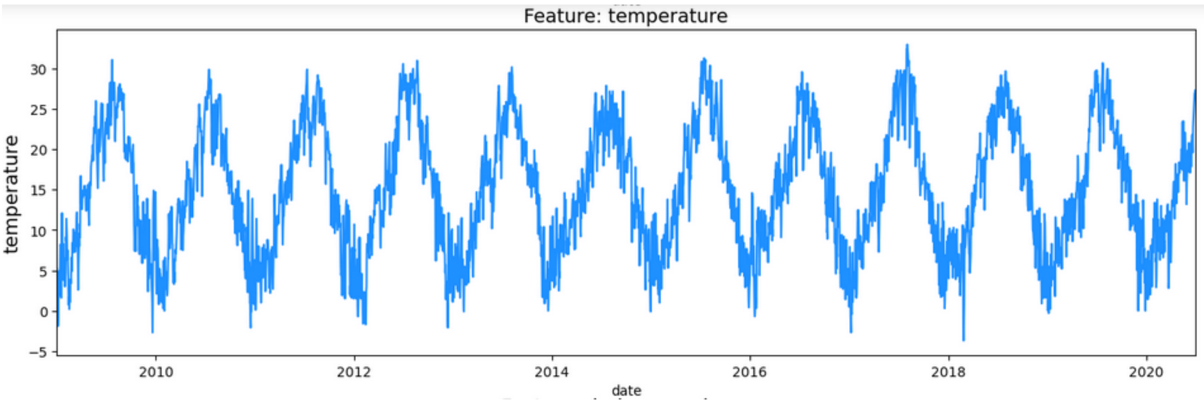

Figure 4: Temperature TS Plot

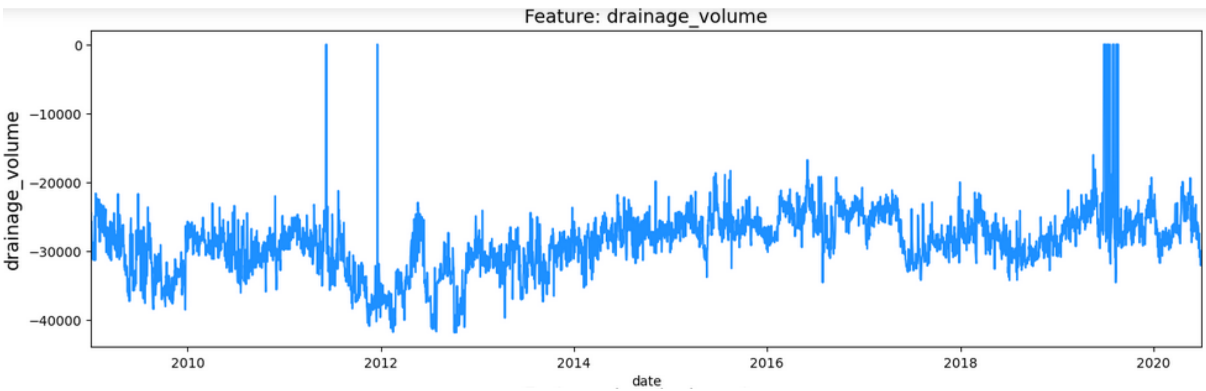

Figure 5: Drainage Volume TS Plot

variance, unlike in the mean, which can be mitigated using remedial techniques covered later.

Fig. (6) shares similar characteristics to the previously mentioned figure. We have a somewhat consistent set of fluctuations through time with a few potential outliers. We may have stationarity in the variance however, the data points are not centered around the mean.

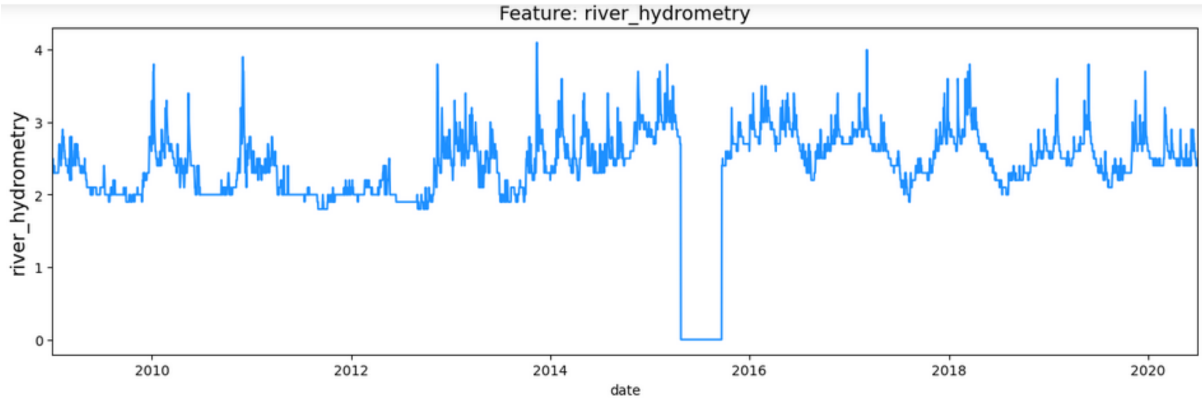

Figure 6: Hydrometry TS Plot

2.2 Data Preprocessing

2.2.1 Chronological Order and Equidistant Time Intervals

Our first step in the data preprocessing aspect of this project is to ensure our date column is in chronological order and that we have constant time intervals between observations. We achieved the chronological order of our dataset by sorting the data frame by timestamps. Regarding the equidistant timestamps, we performed a check between each pair of consecutive observations to see if the difference was constant throughout the entire dataset as shown in Fig. (7). We ended up having a total of 4198 timestamps or observations in our data.

	date	delta
0	2009-01-01	NaT
1	2009-01-02	1 days
2	2009-01-03	1 days
3	2009-01-04	1 days
4	2009-01-05	1 days

Figure 7: Time Series Observation Time Interval

2.2.2 Missing Values

Secondly, we checked for missing values in our dataset and found some columns with missing values. This is also supported by fig. (9, 10, *another*).

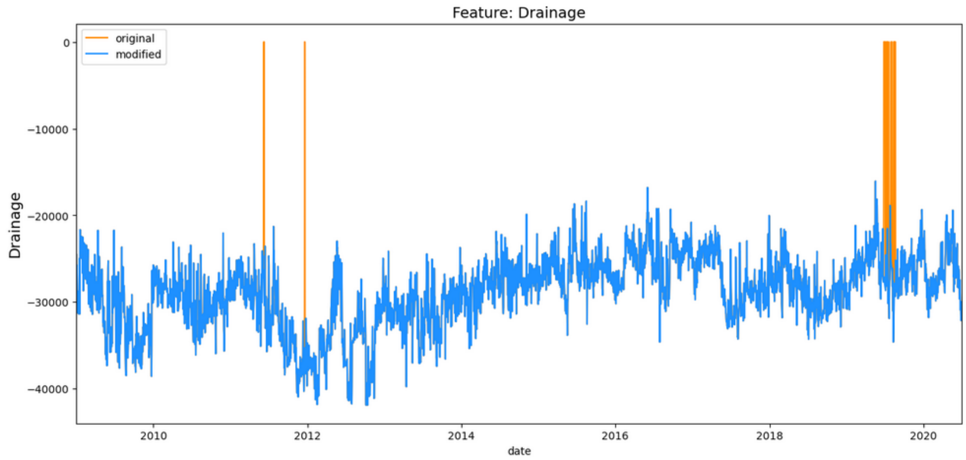

Figure 8: Missing Values in Drainage Volume Feature

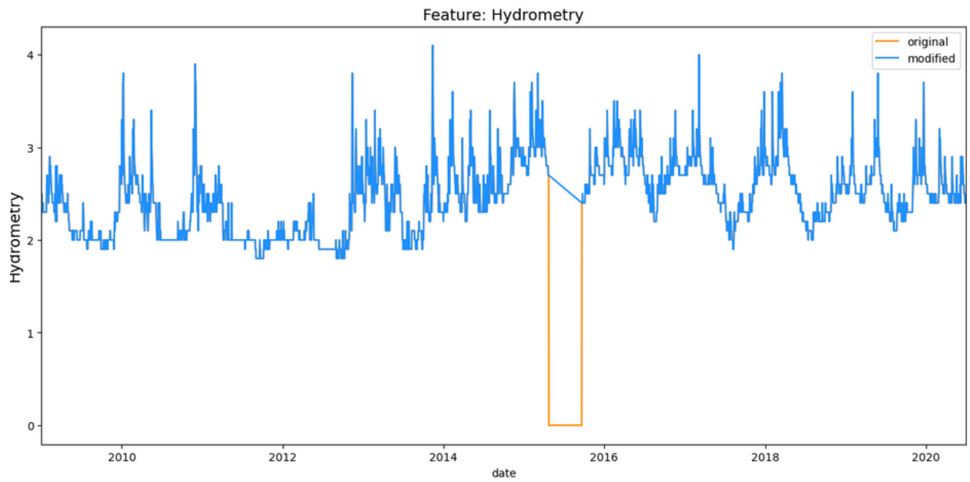

Figure 9: Missing Values in River Hydrometry Feature

There are various methods to handle missing values in our data set. Such methods include filling empty values with extremely high values to classify that observation as an outlier. Other methods are also available such as filling with the mean value of the observations, as well forward filling, and also interpolating using nearby observations in time. As shown in fig. (*4 from app*), we tried to plot a feature through time with every single remedial action and it seems that our best strategy to handle missing values is by interpolation as it produces the closest modified set of observations to the original data.

2.2.3 Smoothing / Resampling

Smoothing or resampling is another vital stage of preprocessing to help us understand the data more thoroughly. We have opted to resample our data. There are two kinds of resampling: upsampling, where we increase the frequency of samples by shortening the time interval, and downsampling, where we decrease the frequency of samples by lengthening the time interval of the data. We decided to choose to downsample to ensure we are dealing with a discrete time series problem rather than a continuous problem. We are downsampling to weekly intervals to help with analyzing the data.

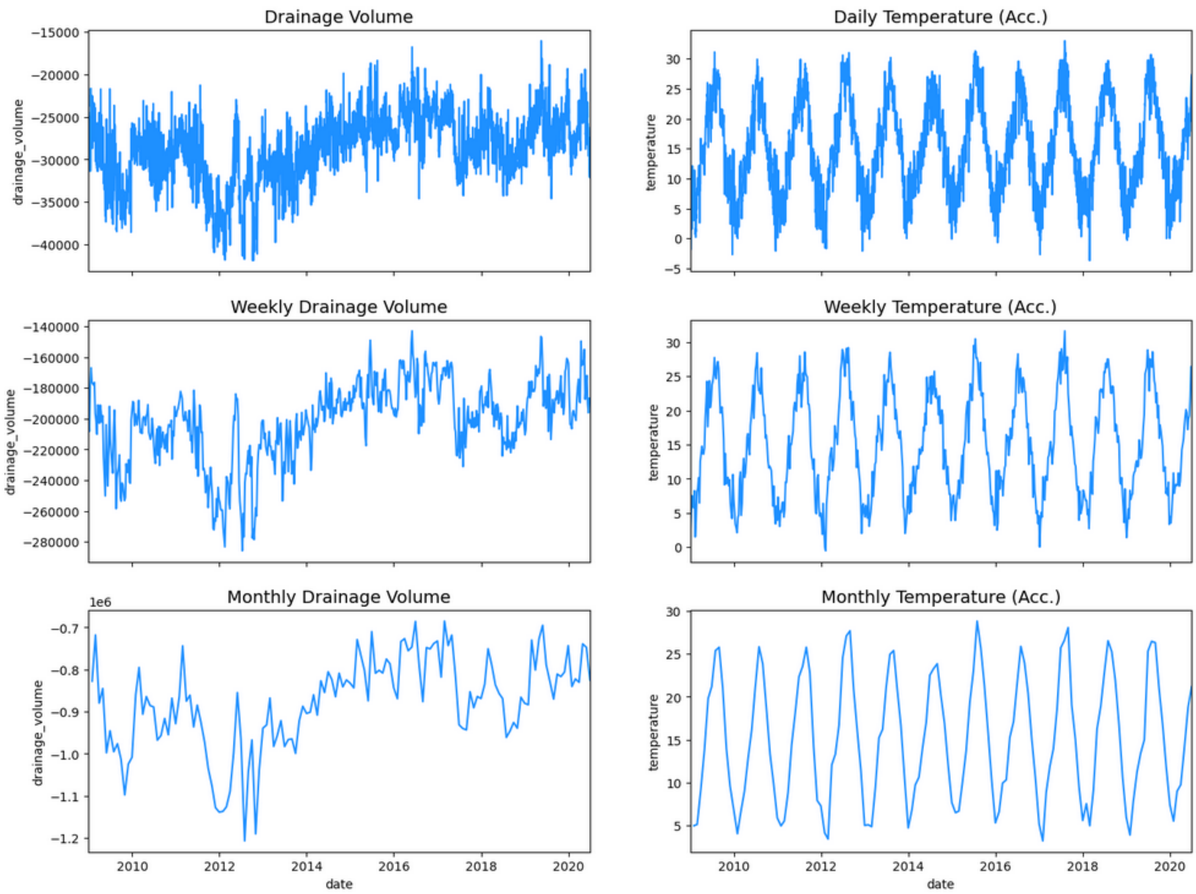

Figure 10: Data After Downsampling

2.2.4 Stationarity

Stationarity describes the time series that has a constant mean, constant variance, and constant covariance. We will check stationarity for our time series via three different approaches:

1. Plotting time series to check for trends or seasonality
2. Series partitioning to compare the mean and variance for each partition of the data
3. Formal test using Augmented Dickey-Fuller (ADF) Test

Regarding time series plots, we can see from fig. (11) that there is non-stationarity in the mean and standard deviation. However, this can be mitigated.

Concerning a formal test, we will use a unit root test such as the ADF to check for non-stationarity. A unit test is a characteristic of a time series that makes it non-stationary.

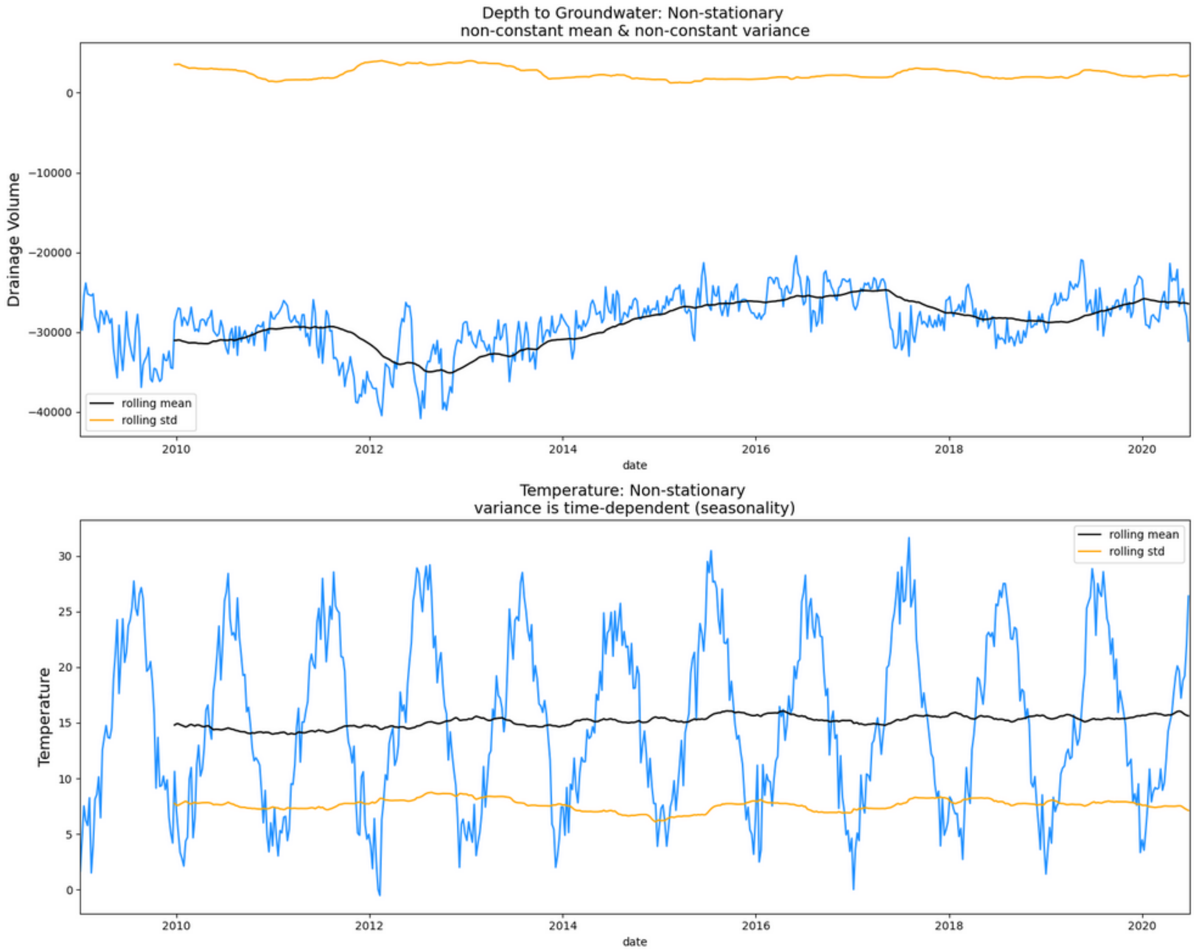

Figure 11: Non-Stationarity in Data

The ADF tests whether a time series has a unit root as its null hypothesis and the reverse as its alternative hypothesis. There are two ways for the null hypothesis to be rejected as shown in equations (1) and (2). One case is when the p-value of a feature is less than or equal to α, which is defaulted to 0.05. Similarly, if the ADF statistic is less than the critical value, then the null hypothesis is also rejected. This indicates no unit root and thus stationarity. Fig. (*from app*) illustrates the output of the Augmented Dickey-Fuller test and the critical values for different values of α.

$$p \leq \alpha \tag{1}$$

$$ADF Statistic < Critical Value \tag{2}$$

We also drew graphs for each variable through time along with their respective ADF statistic and p-values. These are shown in Fig. (12). We can see that some variables will reject the null hypothesis at a value of 0.01 for α. However, we need them all to be stationary at a certain value. Therefore, α set to 0.05 is an appropriate value for continuing our analysis.

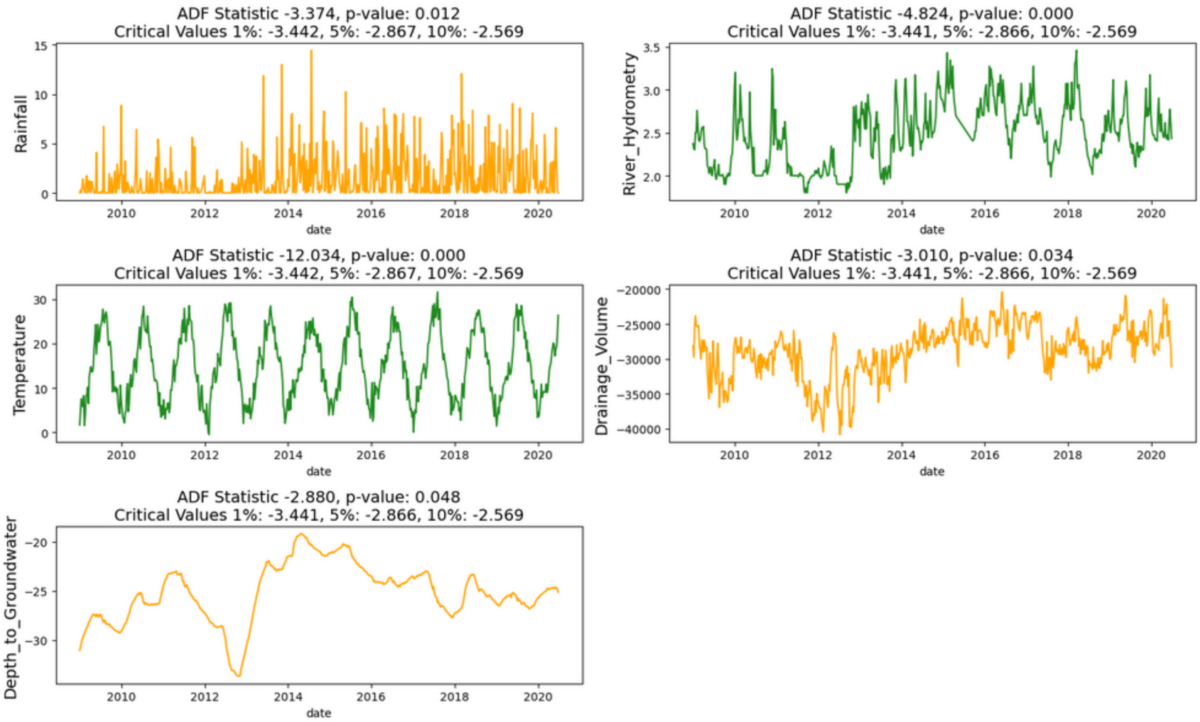

Figure 12: ADF Statistic Graphs

2.2.5 Transformations

To run time series models such as ARIMA, they require that our data is stationary. Therefore, we will perform a transformation to stabilize the non-constant variance in our series. We will perform a log transformation since we have no negative values in our time series. Fig. (13) shows our data after the transformation. It resembles a more normal distribution and the variance has been slightly mitigated.

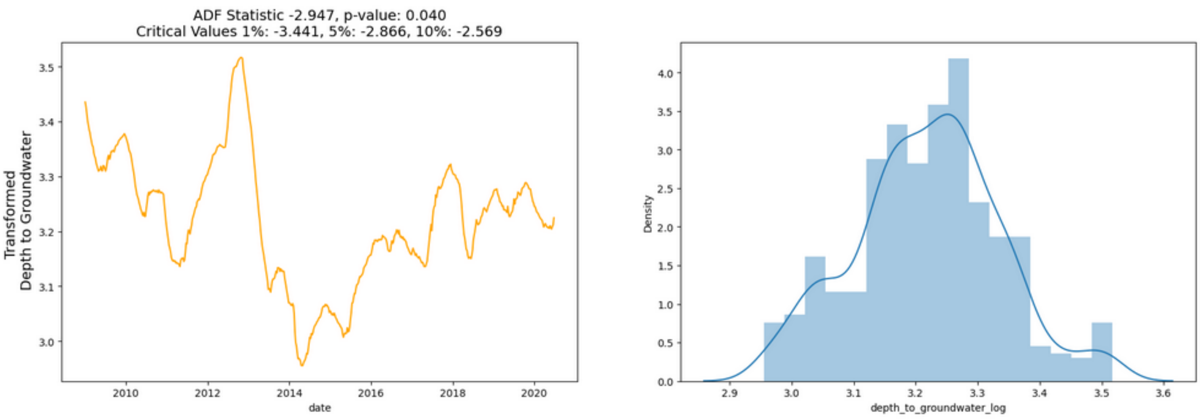

Figure 13: Log Transformation Results

2.2.6 Differencing

Concerning differencing, we will take only the first-order difference as too many will result in some values of 0 in our data and loss of information. Fig. (14) shows the target variable after transforming and differencing the data. It is now stationary in the mean as well as the variance with some notable outliers present.

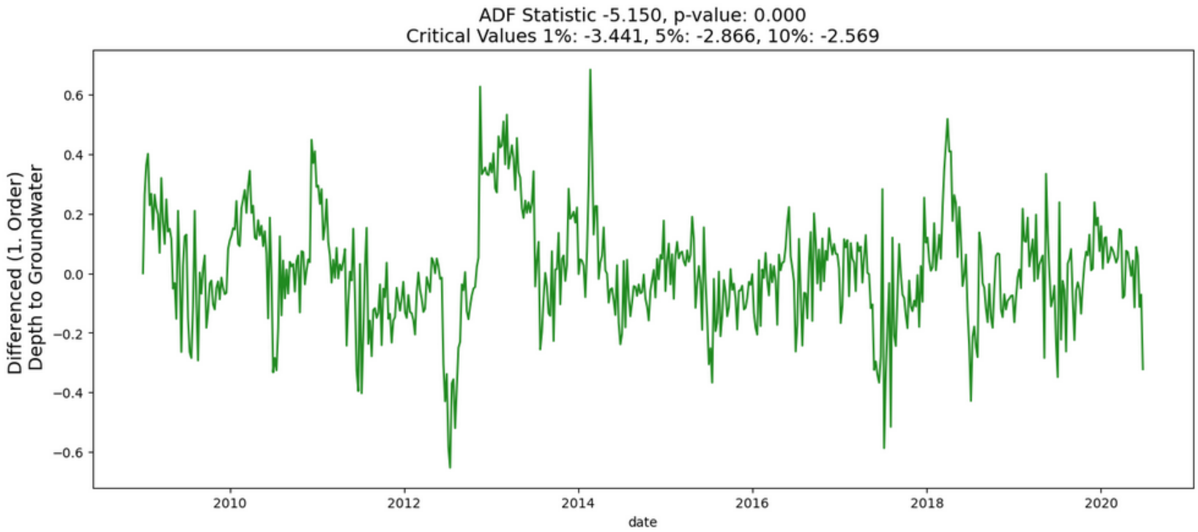

Figure 14: First Order Difference

2.3 Feature Engineering

Regarding feature engineering, we wanted to produce effective features and some necessary actions before our modeling phase of the project.

2.3.1 Encoding Cyclical Features

A key step we performed on our data set was to encode our time features as cyclical features. This means forming a monthly cycle for every single year as shown in Fig. (15). A graph with incremental steps is illustrated showing the cycles throughout the period of our data.

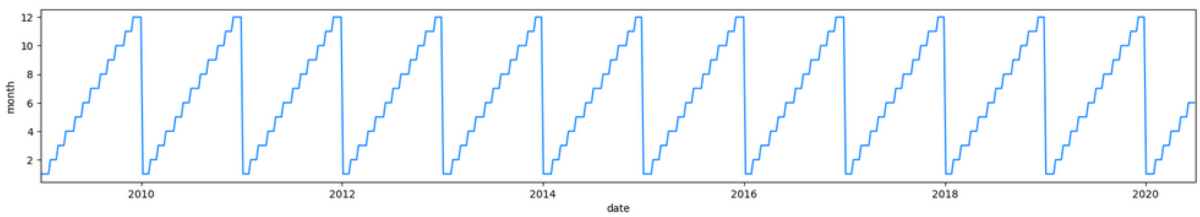

Figure 15: Cyclic Time Features

Furthermore, to improve our analysis of the data later on we tried to form through fig. (15) a transformation into a sinusoidal function through time such that later on we can observe graphs such as the auto-correlation and partial auto-correlation function graphs respectively, and find the cut-off points. Fig. (16) illustrates this in terms of taking the sine and cosine of the month variable.

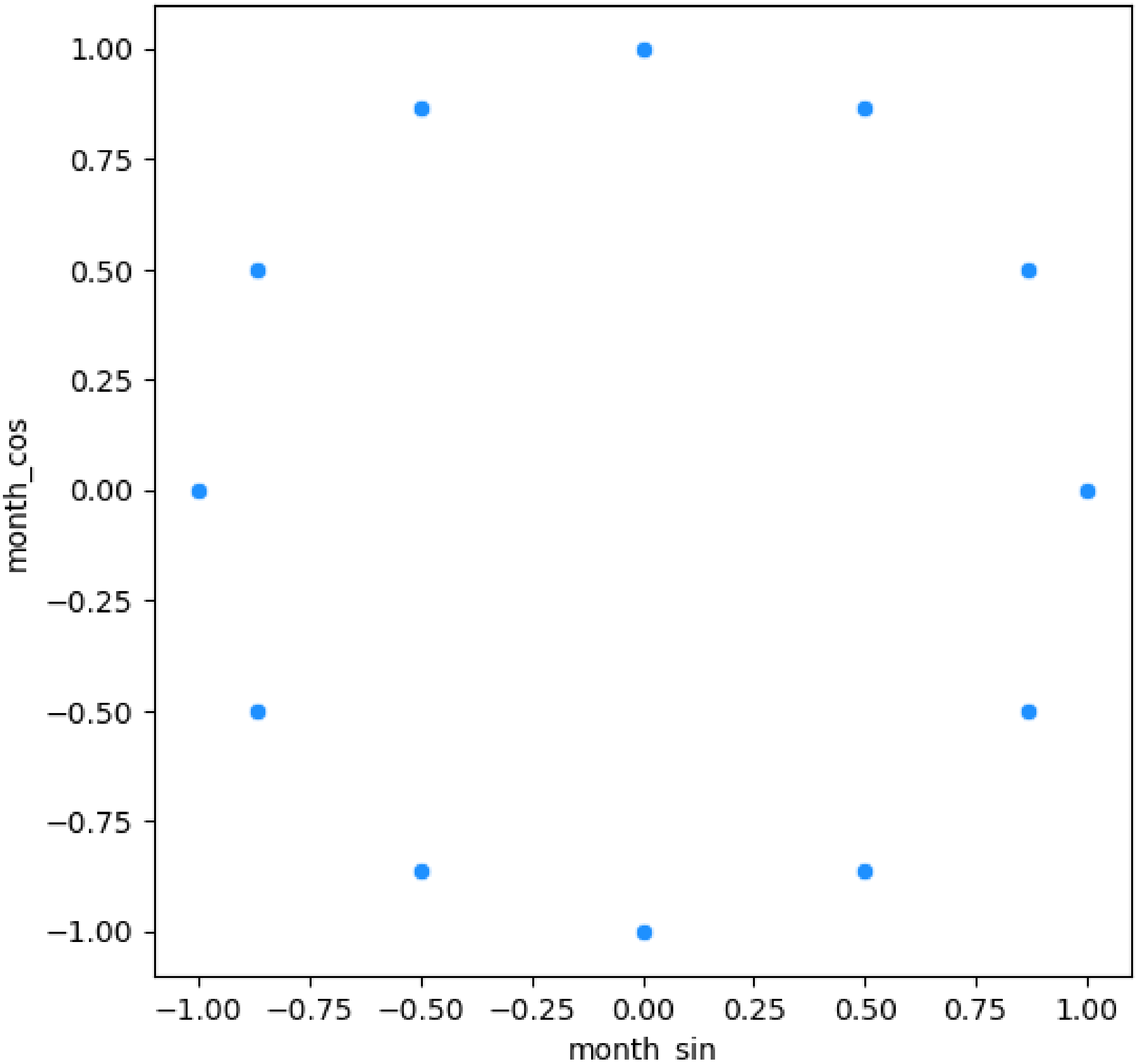

Figure 16: Sinusoidal Time Features

2.3.2 Time Series Decomposition

A time series is comprised of a combination of the following components:

1. Trend - the increasing or decreasing value in the series
2. Seasonality - the repeating short-term cycles in the series
3. Level - the average value in the series
4. Noise - the random variation in the series

Decomposing a time series helps us to generalize the series and understand its details more thoroughly during analysis. In our case, we will use a seasonal decomposition function that uses a moving average as its underlying method. We are assuming an additive model with a period frequency of 52 since we have 52 weeks per year.

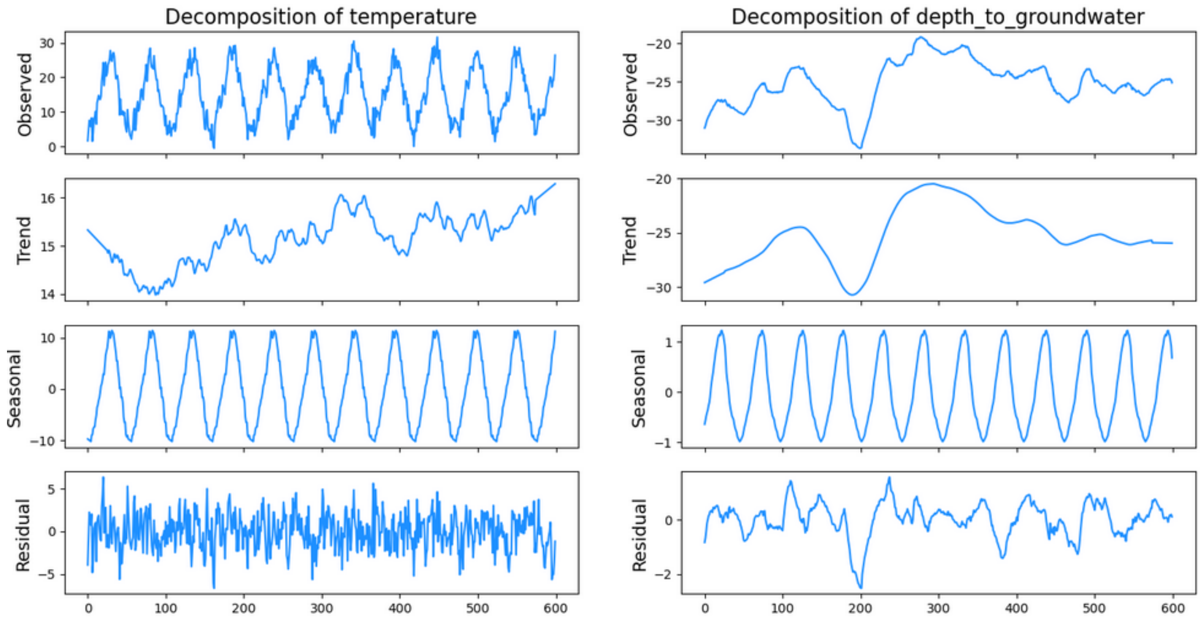

Figure 17: Seasonal Decomposition of 2 Variables

2.3.3 Extracting Lags

For computing the lags, fig. (*in app*) shows the functions used to get them. Using the shift function we were able to compute the lags of our data.

2.4 Exploratory Data Analysis

This section focuses on extracting some knowledge from our data. One of the things we computed was the seasonal indices of our data features as well as the target variable. Fig. (*in app*) illustrates the seasonal indices of features through a monthly time interval. As we can see, there is seasonality in a lot of the features we have.

Furthermore, we computed and visualized the correlation matrix between all our core features as well as the correlation matrix of the lagged features as shown in fig. (*in app*). We can see that the lagged features are more highly correlated than the core features from our original data set.

2.4.1 Autocorrelation Analysis

After stationarizing the time series by differencing as mentioned previously, we now use an autocorrelation plot and partial autocorrelation plot to determine whether an autoregressive model (AR) or moving average (MA) is needed to fix any autocorrelation still present in the series. Fig. (18) shows the autocorrelation plot of our dependent variable, and as we can see it alternates for each lag, so perhaps a value of MA(1). Similarly, with the partial autocorrelation plot, we can conclude that lags are decreasing rapidly and cut off at lag 3 approximately.

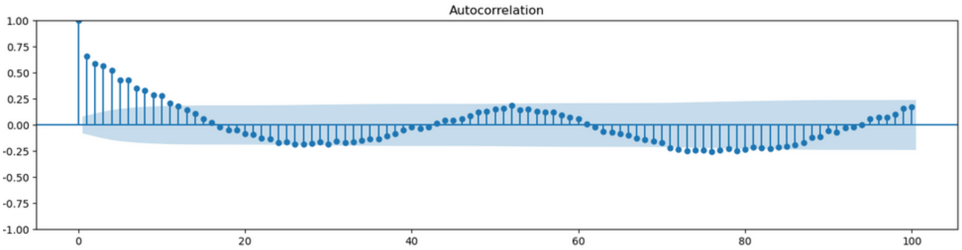

Figure 18: ACF Plot

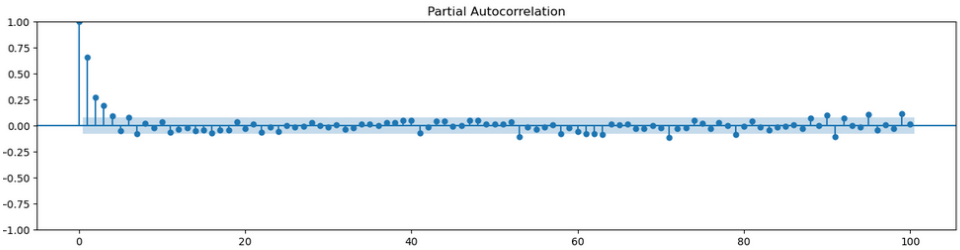

Figure 19: PACF Plot

2.5 Modeling

For the modeling stage of this project, we performed a univariate time series on each of our variables including depth to groundwater, rainfall, temperature, and so on through time. We chose 2 models for our univariate analysis: Prophet, and Auto-ARIMA. Prophet is an open-source library for performing univariate time series forecasting and it does this using an additive time series forecasting model. Implementation details support all effects of time series including trend, seasonality, and cyclability.

For Prophet, we computed the root mean squared error as our evaluation metrics while for Auto-ARIMA, we attempted to show a summary of each variable by minimizing the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). By using different values of p, q, and d in our ARIMA models, we could decide which values are most appropriate to minimize the AIC and BIC.

3 Results and Discussion

3.1 Results

Regarding the results of our modeling stage, we started with the depth to groundwater variable, plotting the rolling window to see which training splits are most appropriate.

After performing our training test split of 85 percent to the training and 15 percent to the testing, we trained our Prophet model and used it for forecasting. fig. (31) illustrates the forecast data as well as the confidence interval of the variable. Furthermore, we trained the Auto-ARIMA model, and as shown in Fig. (32), we tested different values of p and q for both the moving average and auto-regressive parts of our ARIMA model as well as the integration of the ARIMA model.

The best model was ARIMA(1,1,1) based on a stepwise search to minimize the AIC evaluation metric. SARIMAX results in Fig. (32) also display the BIC and Hannan-Quinn Information Criterion (HQIC) as two other evaluation metrics for model selection. These values support our decision that auto-ARIMA would outperform the Prophet model due to lower values in their respective information criterion.

We also plotted the diagnostic plots of the ARIMA model for the variable depth to groundwater. As fig. (33) shows, our standardized residuals are based around 0. The Q-Q plot and histogram show univariate normality with some notable outliers present toward the end of the graphs.

We performed the same procedures for the following variables:

- Rainfall
- Temperature
- River Hydrometry
- Drainage Volume

3.2 Conclusion

In conclusion, each variable data was trained using a variety of time series models and each reported back with unique insights. Features such as depth to groundwater, river hydrometry, and drainage volume show promising results for forecasting with low AIC, BIC, and HQIC. They all present normal distributions post-modeling with a lack of outliers.

The temperature feature displays a clear seasonality effect both in the analysis stage as well as during the training process. Forecasting is also seasonal depending on the time of the year. However, the rainfall feature lags as a lone variable. Plots for the prophet forecast are unclear and scattered. AIC, BIC, and HQIC values are also significantly larger than the other features. This could be due to the nature of the rainfall feature alone, it might be dependent on other feature interactions to interpret or analyze it better.

Areas of improvement to be made in this project is to perform multivariate analysis on this list of features to understand the various interactions involved as they simultaneously pass through time. An appropriate measure for this would be Vector Auto Regression (VAR) for multivariate analysis.

4 Appendix

4.1 Annotated Computer Output (Tables & Figures)

```
df = df.drop('delta', axis=1)
df.isna().sum()
```

```
date                    0
rainfall                0
depth_to_groundwater   27
temperature             0
drainage_volume         1
river_hydrometry        0
dtype: int64
```

Figure 20: Missing Values across Data Set

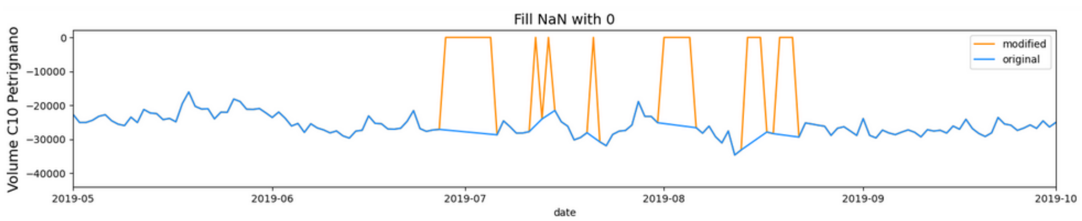

Figure 21: Remedy Missing Values by Outlier

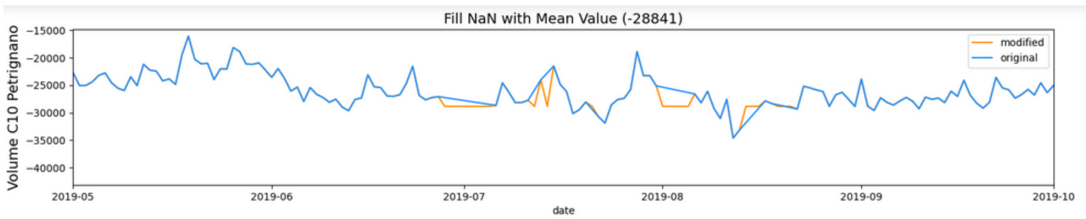

Figure 22: Remedy Missing Values by Mean Value

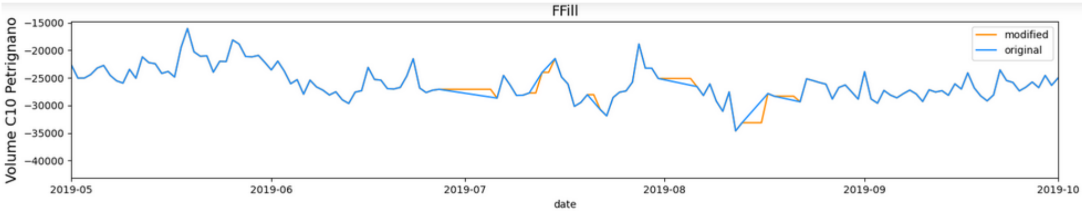

Figure 23: Remedy Missing Values by Forward Fill

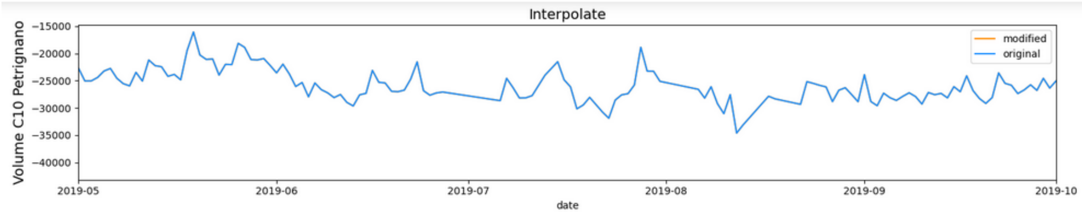

Figure 24: Remedy Missing Values by Interpolation

```
from statsmodels.tsa.stattools import adfuller

result = adfuller(df['depth_to_groundwater'].values)
result
```

```
(-2.880201649316664,
 0.04769919092020842б,
 7,
 592,
 {'1%': -3.441444394224128,
  '5%': -2.866434537627б454,
  '10%': -2.5693766637372l7},
 -734.3154255877616)
```

Figure 25: Augmented Dickey-Fuller Test

```
weeks_in_month = 4

for column in core_columns:
    df[f'{column}_seasonal_shift_b_2m'] = df[f'{column}_seasonal'].shift(-2 * weeks_in_month)
    df[f'{column}_seasonal_shift_b_1m'] = df[f'{column}_seasonal'].shift(-1 * weeks_in_month)
    df[f'{column}_seasonal_shift_1m'] = df[f'{column}_seasonal'].shift(1 * weeks_in_month)
    df[f'{column}_seasonal_shift_2m'] = df[f'{column}_seasonal'].shift(2 * weeks_in_month)
    df[f'{column}_seasonal_shift_3m'] = df[f'{column}_seasonal'].shift(3 * weeks_in_month)
```

Figure 26: Computing Lags

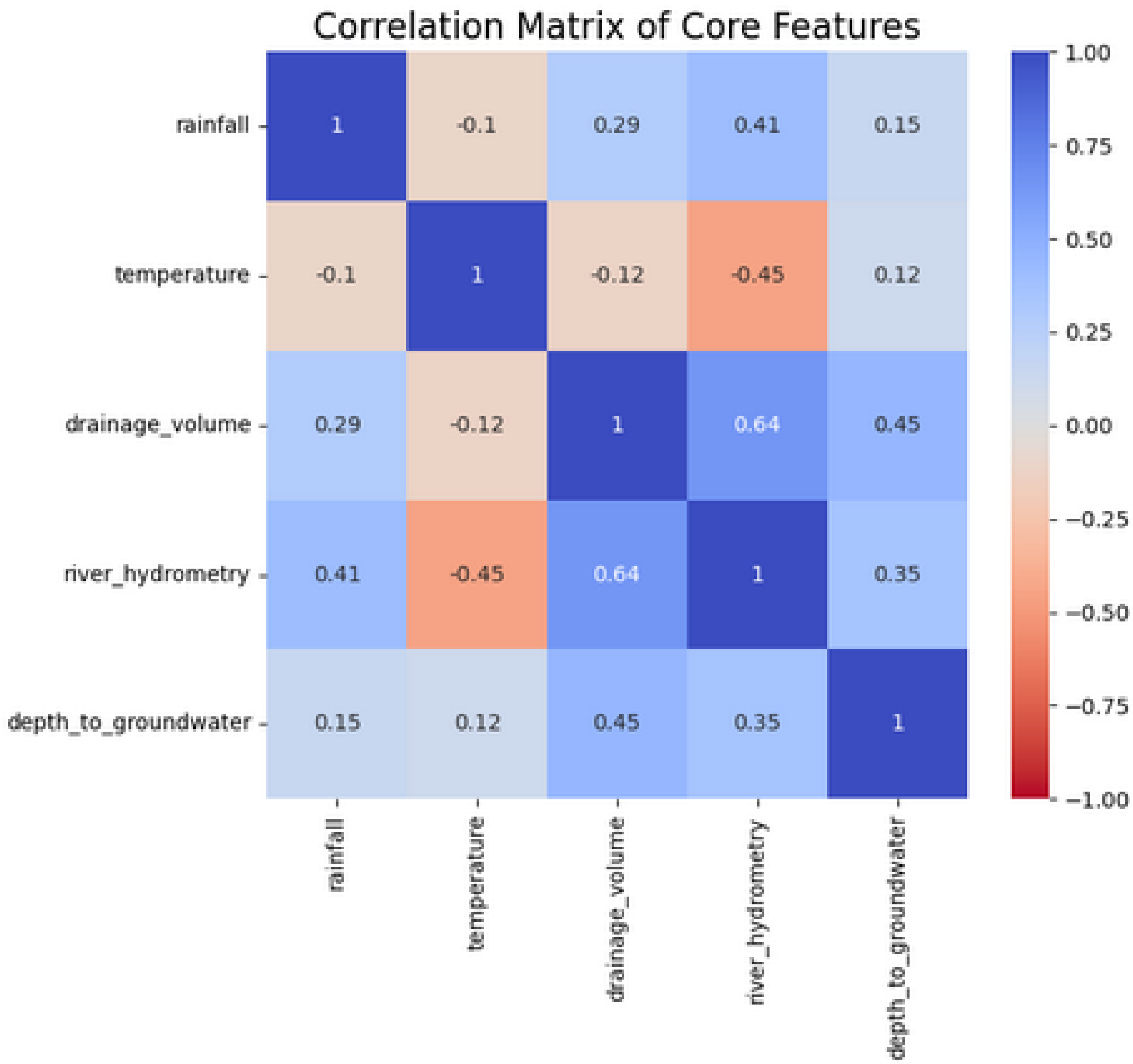

Figure 27: Correlation Matrix of Original Features

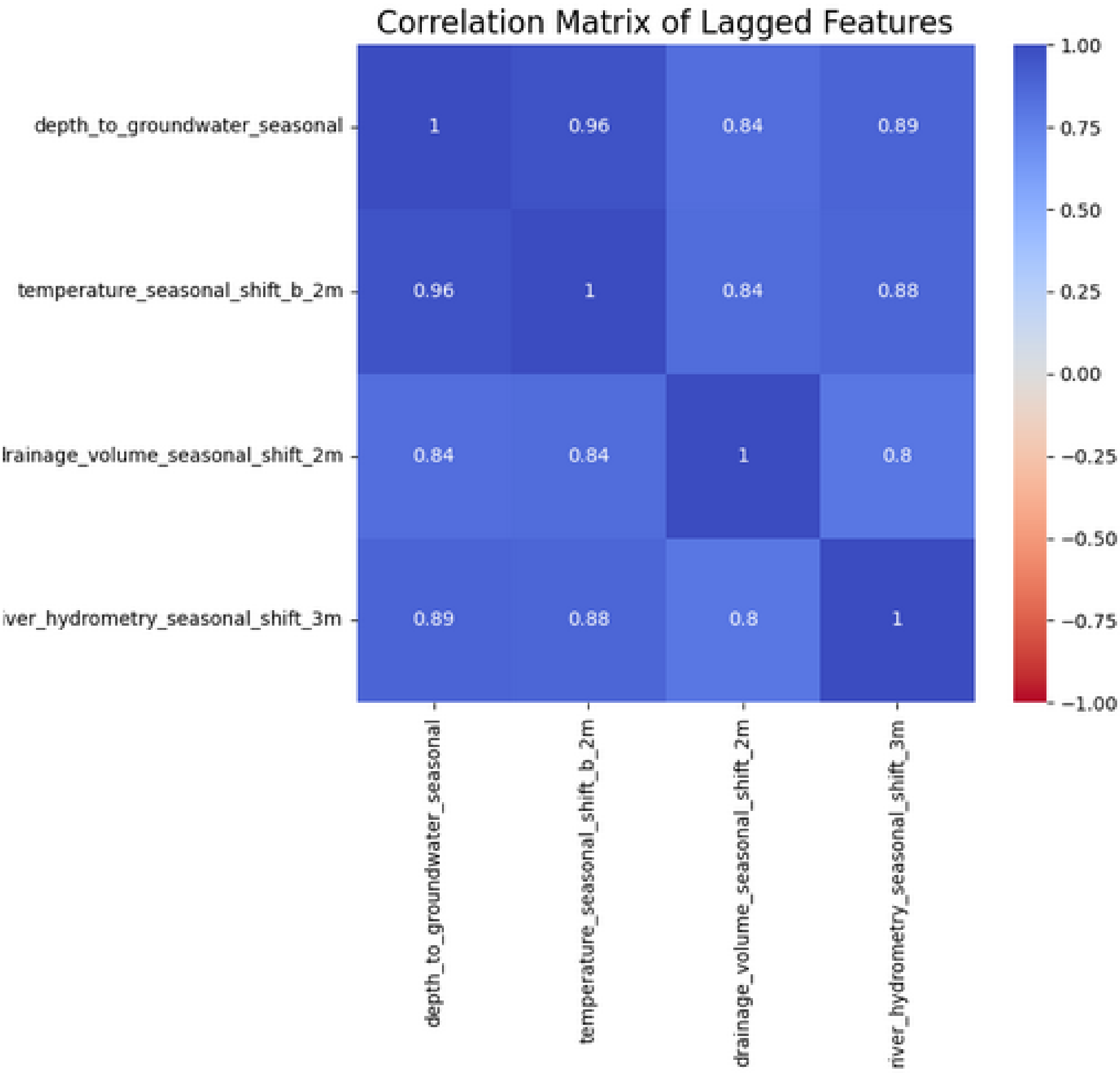

Figure 28: Correlation Matrix of Lagged Features

Figure 29: Seasonal Components of Variables

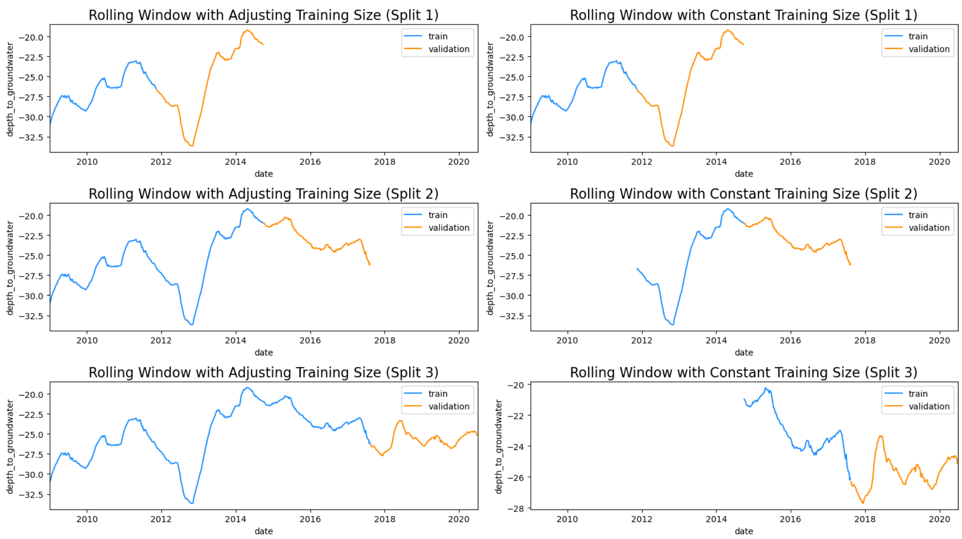

Figure 30: Rolling Window Depth to Groundwater

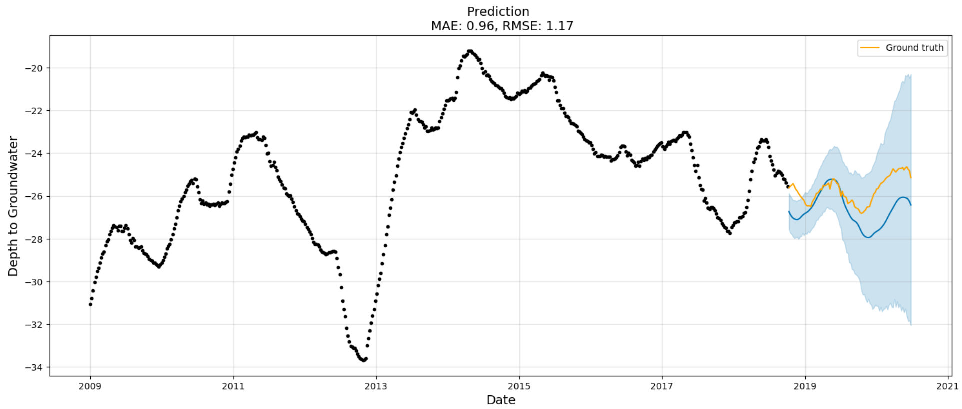

Figure 31: Prophet Forecast Depth to Groundwater

```
Performing stepwise search to minimize aic
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=-631.136, Time=0.25 sec
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=-242.692, Time=0.05 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=-574.047, Time=0.15 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=-427.347, Time=0.08 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=-243.054, Time=0.02 sec
 ARIMA(2,1,1)(0,0,0)[0] intercept   : AIC=-629.209, Time=0.41 sec
 ARIMA(1,1,2)(0,0,0)[0] intercept   : AIC=-629.237, Time=0.37 sec
 ARIMA(0,1,2)(0,0,0)[0] intercept   : AIC=-492.779, Time=0.16 sec
 ARIMA(2,1,0)(0,0,0)[0] intercept   : AIC=-611.065, Time=0.14 sec
 ARIMA(2,1,2)(0,0,0)[0] intercept   : AIC=-628.351, Time=0.54 sec
 ARIMA(1,1,1)(0,0,0)[0]             : AIC=-632.995, Time=0.09 sec
 ARIMA(0,1,1)(0,0,0)[0]             : AIC=-428.258, Time=0.05 sec
 ARIMA(1,1,0)(0,0,0)[0]             : AIC=-575.735, Time=0.04 sec
 ARIMA(2,1,1)(0,0,0)[0]             : AIC=-631.069, Time=0.19 sec
 ARIMA(1,1,2)(0,0,0)[0]             : AIC=-631.097, Time=0.14 sec
 ARIMA(0,1,2)(0,0,0)[0]             : AIC=-494.001, Time=0.07 sec
 ARIMA(2,1,0)(0,0,0)[0]             : AIC=-612.866, Time=0.06 sec
 ARIMA(2,1,2)(0,0,0)[0]             : AIC=-630.210, Time=0.29 sec

Best model:  ARIMA(1,1,1)(0,0,0)[0]
Total fit time: 3.106 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  510
Model:               SARIMAX(1, 1, 1)   Log Likelihood                 319.497
Date:                Sat, 08 Apr 2023   AIC                           -632.995
Time:                        08:42:23   BIC                           -620.297
Sample:                             0   HQIC                          -628.016
                                - 510
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.9195      0.021     43.765      0.000       0.878       0.961
ma.L1         -0.4885      0.037    -13.357      0.000      -0.560      -0.417
sigma2         0.0167      0.001     24.809      0.000       0.015       0.018
===================================================================================
Ljung-Box (L1) (Q):                   0.02   Jarque-Bera (JB):               185.01
Prob(Q):                              0.90   Prob(JB):                         0.00
Heteroskedasticity (H):               1.17   Skew:                             0.22
Prob(H) (two-sided):                  0.32   Kurtosis:                         5.92
===================================================================================
```

Figure 32: Auto-ARIMA Summary Depth to Groundwater

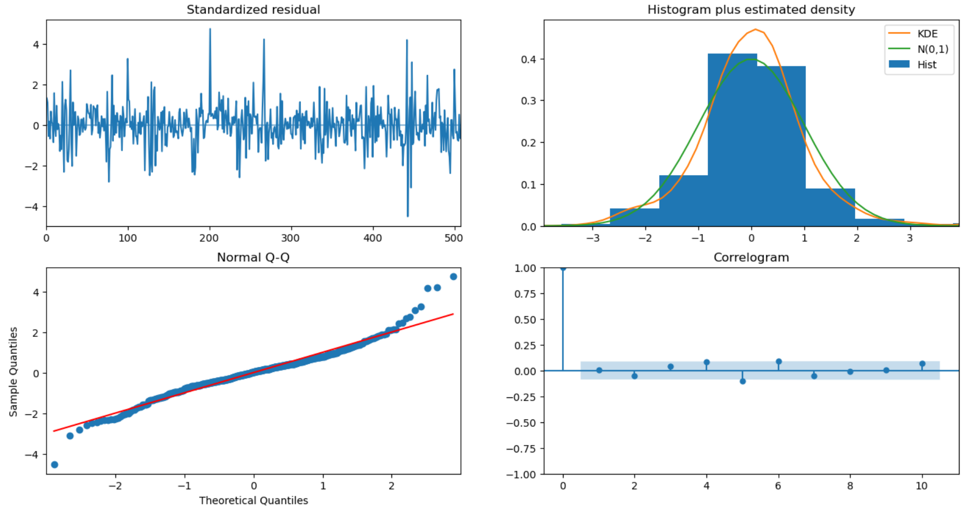

Figure 33: Model Diagnostics Depth to Groundwater

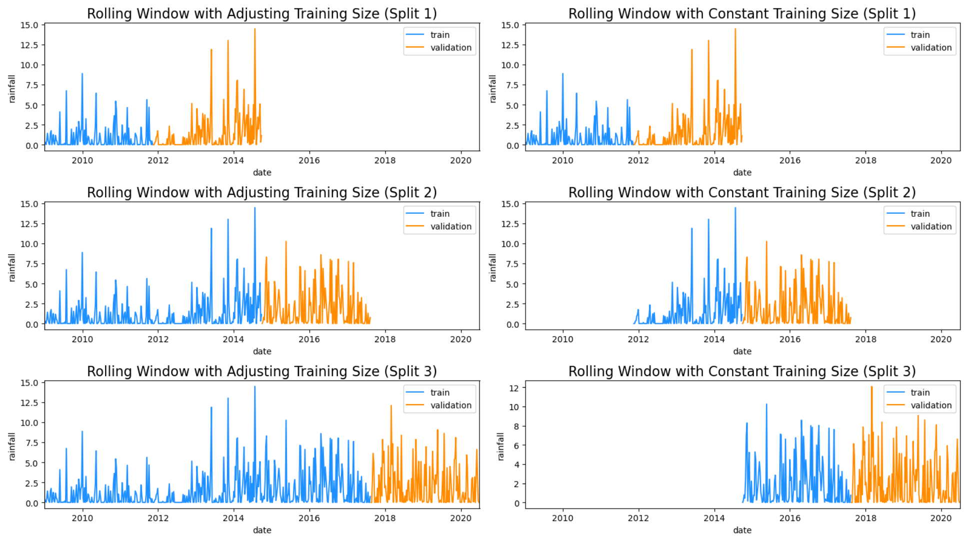

Figure 34: Rolling Window Rainfall

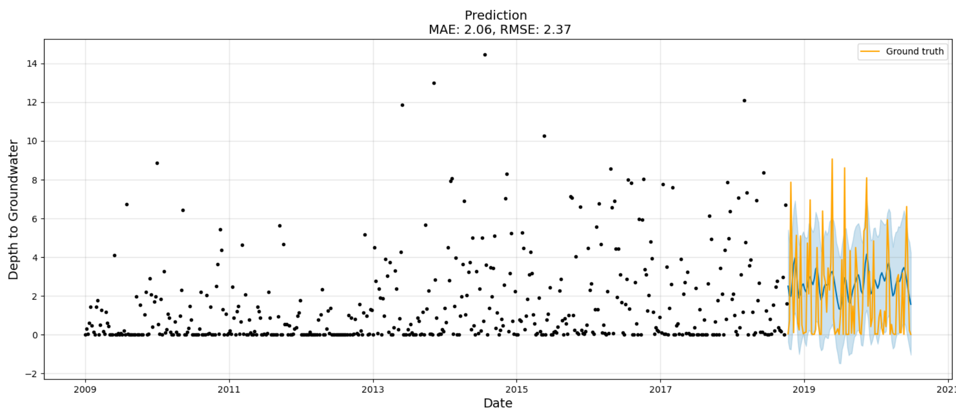

Figure 35: Prophet Forecast Rainfall

```
Performing stepwise search to minimize aic
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=2268.699, Time=0.10 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=2470.451, Time=0.01 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=2379.860, Time=0.02 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=2410.902, Time=0.03 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=2267.354, Time=0.15 sec
 ARIMA(2,0,0)(0,0,0)[0]             : AIC=2351.710, Time=0.03 sec
 ARIMA(3,0,1)(0,0,0)[0]             : AIC=2269.217, Time=0.22 sec
 ARIMA(2,0,2)(0,0,0)[0]             : AIC=2269.795, Time=0.39 sec
 ARIMA(1,0,2)(0,0,0)[0]             : AIC=2267.489, Time=0.15 sec
 ARIMA(3,0,0)(0,0,0)[0]             : AIC=2337.835, Time=0.04 sec
 ARIMA(3,0,2)(0,0,0)[0]             : AIC=2270.935, Time=0.42 sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=2266.280, Time=0.41 sec
 ARIMA(1,0,1)(0,0,0)[0] intercept   : AIC=2267.116, Time=0.33 sec
 ARIMA(2,0,0)(0,0,0)[0] intercept   : AIC=2280.473, Time=0.07 sec
 ARIMA(3,0,1)(0,0,0)[0] intercept   : AIC=2268.210, Time=0.54 sec
 ARIMA(2,0,2)(0,0,0)[0] intercept   : AIC=inf, Time=0.49 sec
 ARIMA(1,0,0)(0,0,0)[0] intercept   : AIC=2281.346, Time=0.04 sec
 ARIMA(1,0,2)(0,0,0)[0] intercept   : AIC=2266.370, Time=0.42 sec
 ARIMA(3,0,0)(0,0,0)[0] intercept   : AIC=2281.261, Time=0.09 sec
 ARIMA(3,0,2)(0,0,0)[0] intercept   : AIC=inf, Time=0.59 sec

Best model:  ARIMA(2,0,1)(0,0,0)[0] intercept
Total fit time: 4.532 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  510
Model:               SARIMAX(2, 0, 1)   Log Likelihood               -1128.140
Date:                Sat, 08 Apr 2023   AIC                           2266.280
Time:                        08:53:41   BIC                           2287.453
Sample:                             0   HQIC                          2274.581
                                - 510
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0131      0.017      0.755      0.450      -0.021       0.047
ar.L1          1.0712      0.049     21.799      0.000       0.975       1.168
ar.L2         -0.0802      0.045     -1.795      0.073      -0.168       0.007
ma.L1         -0.9605      0.023    -42.229      0.000      -1.005      -0.916
sigma2         4.8708      0.240     20.303      0.000       4.401       5.341
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              1145.67
Prob(Q):                              0.97   Prob(JB):                         0.00
Heteroskedasticity (H):               3.17   Skew:                             2.15
Prob(H) (two-sided):                  0.00   Kurtosis:                         8.95
===================================================================================
```

Figure 36: Auto-ARIMA Summary Rainfall

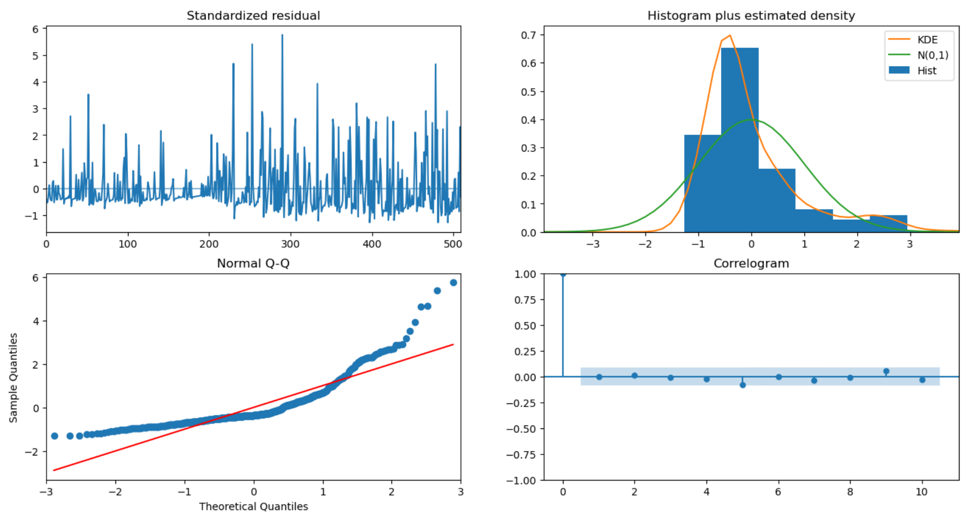

Figure 37: Model Diagnostics Rainfall

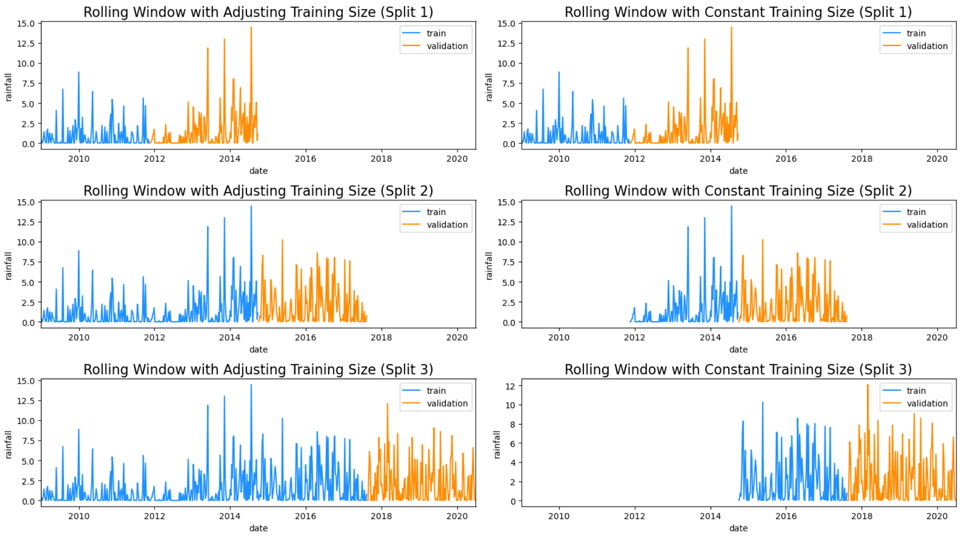

Figure 38: Rolling Window Temperature

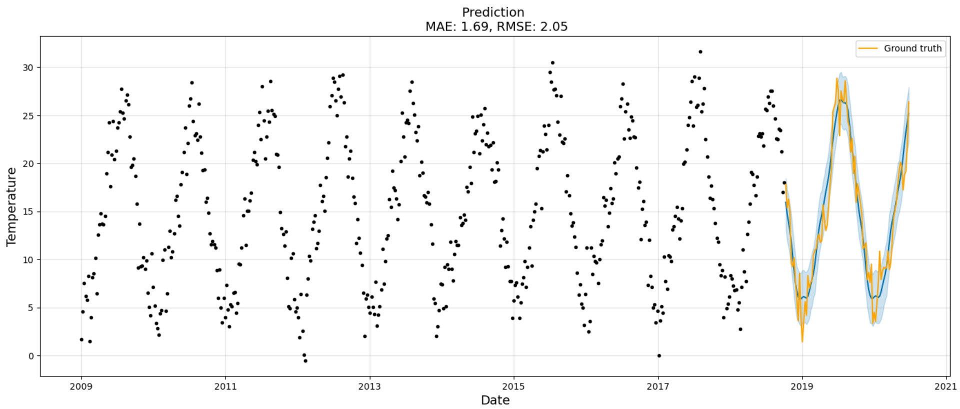

Figure 39: Prophet Forecast Temperature

```
Performing stepwise search to minimize aic
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=2445.058, Time=0.07 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=4338.136, Time=0.01 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=2459.639, Time=0.03 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=3751.909, Time=0.04 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=2446.919, Time=0.12 sec
 ARIMA(1,0,2)(0,0,0)[0]             : AIC=2446.737, Time=0.09 sec
 ARIMA(0,0,2)(0,0,0)[0]             : AIC=3381.498, Time=0.10 sec
 ARIMA(2,0,0)(0,0,0)[0]             : AIC=inf, Time=0.05 sec
 ARIMA(2,0,2)(0,0,0)[0]             : AIC=2418.034, Time=0.17 sec
 ARIMA(3,0,2)(0,0,0)[0]             : AIC=2403.451, Time=0.25 sec
 ARIMA(3,0,1)(0,0,0)[0]             : AIC=2447.087, Time=0.14 sec
 ARIMA(3,0,3)(0,0,0)[0]             : AIC=2447.728, Time=0.50 sec
 ARIMA(2,0,3)(0,0,0)[0]             : AIC=2405.827, Time=0.36 sec
 ARIMA(3,0,2)(0,0,0)[0] intercept   : AIC=2290.514, Time=0.65 sec
 ARIMA(2,0,2)(0,0,0)[0] intercept   : AIC=2304.786, Time=0.51 sec
 ARIMA(3,0,1)(0,0,0)[0] intercept   : AIC=2439.902, Time=0.26 sec
 ARIMA(3,0,3)(0,0,0)[0] intercept   : AIC=inf, Time=nan sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=2439.398, Time=0.16 sec
 ARIMA(2,0,3)(0,0,0)[0] intercept   : AIC=2275.802, Time=0.67 sec
 ARIMA(1,0,3)(0,0,0)[0] intercept   : AIC=2431.689, Time=0.18 sec
 ARIMA(1,0,2)(0,0,0)[0] intercept   : AIC=2439.046, Time=0.14 sec

Best model:  ARIMA(2,0,3)(0,0,0)[0] intercept
Total fit time: 5.132 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  510
Model:               SARIMAX(2, 0, 3)   Log Likelihood               -1130.901
Date:                Sat, 08 Apr 2023   AIC                           2275.802
Time:                        09:08:38   BIC                           2305.443
Sample:                             0   HQIC                          2287.423
                                - 510
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.2173      0.005     41.693      0.000       0.207       0.227
ar.L1          1.9815      0.002    995.616      0.000       1.978       1.985
ar.L2         -0.9960      0.002   -506.723      0.000      -1.000      -0.992
ma.L1         -1.5421      0.046    -33.413      0.000      -1.633      -1.452
ma.L2          0.2878      0.085      3.388      0.001       0.121       0.454
ma.L3          0.2706      0.046      5.860      0.000       0.180       0.361
sigma2         4.8959      0.305     16.069      0.000       4.299       5.493
===================================================================================
Ljung-Box (L1) (Q):                   0.07   Jarque-Bera (JB):                 1.12
Prob(Q):                              0.79   Prob(JB):                         0.57
Heteroskedasticity (H):               0.88   Skew:                            -0.11
Prob(H) (two-sided):                  0.40   Kurtosis:                         3.09
```

Figure 40: Auto-ARIMA Summary Temperature

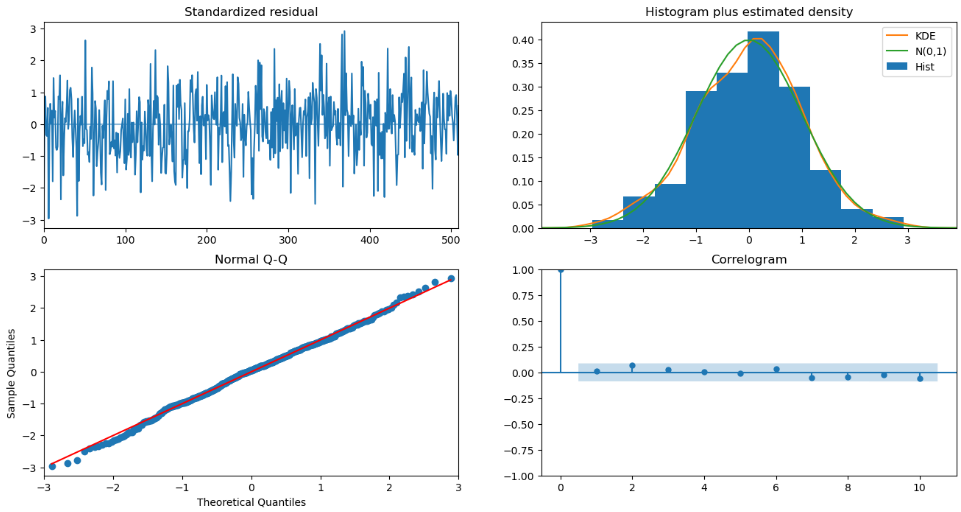

Figure 41: Model Diagnostics Temperature

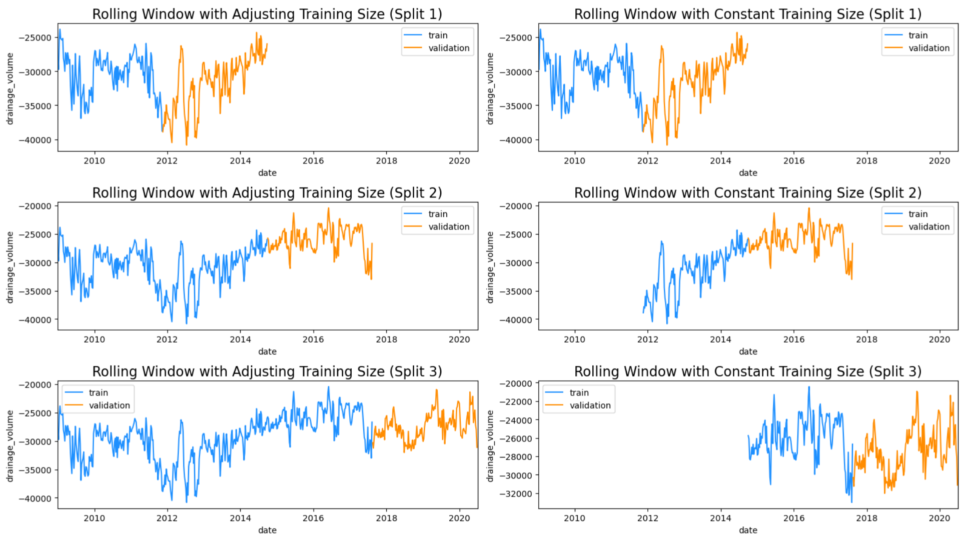

Figure 42: Rolling Window Drainage Volume

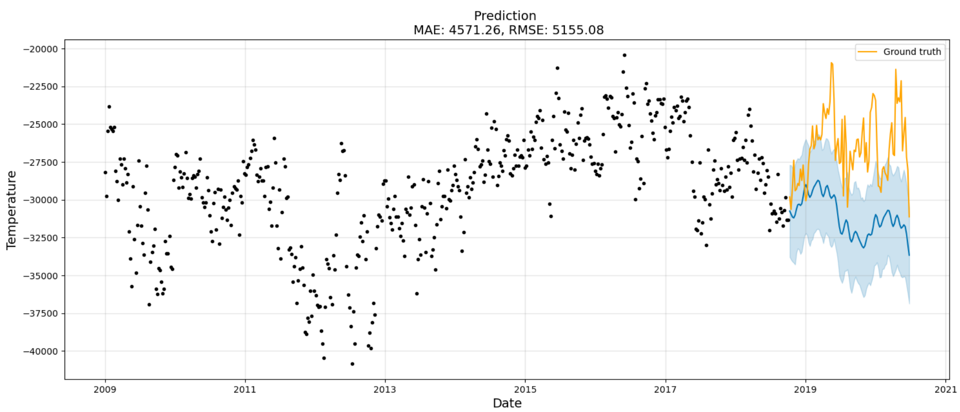

Figure 43: Prophet Forecast Drainage Volume

```
Performing stepwise search to minimize aic
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=9111.879, Time=0.03 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=11946.700, Time=0.01 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=inf, Time=0.01 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=inf, Time=0.08 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=9090.176, Time=0.23 sec
 ARIMA(2,0,0)(0,0,0)[0]             : AIC=inf, Time=0.02 sec
 ARIMA(3,0,1)(0,0,0)[0]             : AIC=9091.099, Time=0.23 sec
 ARIMA(2,0,2)(0,0,0)[0]             : AIC=9090.681, Time=0.26 sec
 ARIMA(1,0,2)(0,0,0)[0]             : AIC=9101.985, Time=0.04 sec
 ARIMA(3,0,0)(0,0,0)[0]             : AIC=inf, Time=0.03 sec
 ARIMA(3,0,2)(0,0,0)[0]             : AIC=inf, Time=0.57 sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=9092.787, Time=0.12 sec

Best model:  ARIMA(2,0,1)(0,0,0)[0]
Total fit time: 1.654 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  510
Model:               SARIMAX(2, 0, 1)   Log Likelihood               -4541.088
Date:                Sat, 08 Apr 2023   AIC                           9090.176
Time:                        09:11:25   BIC                           9107.114
Sample:                             0   HQIC                          9096.817
                                - 510
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.6631      0.052     31.741      0.000       1.560       1.766
ar.L2         -0.6632      0.052    -12.661      0.000      -0.766      -0.561
ma.L1         -0.9053      0.031    -29.619      0.000      -0.965      -0.845
sigma2      3.235e+06   1.78e-09   1.82e+15      0.000    3.23e+06    3.23e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.30   Jarque-Bera (JB):                29.89
Prob(Q):                              0.58   Prob(JB):                         0.00
Heteroskedasticity (H):               0.71   Skew:                            -0.17
Prob(H) (two-sided):                  0.03   Kurtosis:                         4.14
===================================================================================
```

Figure 44: Auto-ARIMA Summary Drainage Volume

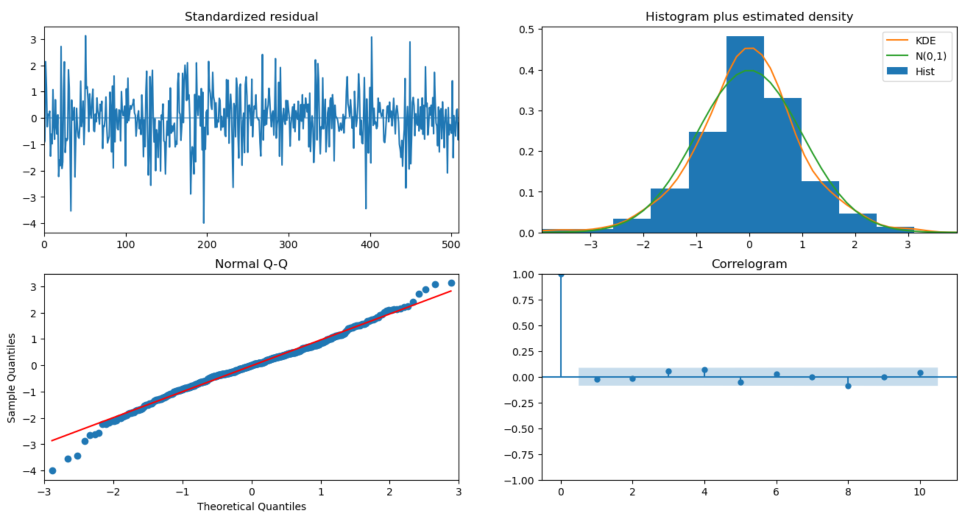

Figure 45: Model Diagnostics Drainage Volume

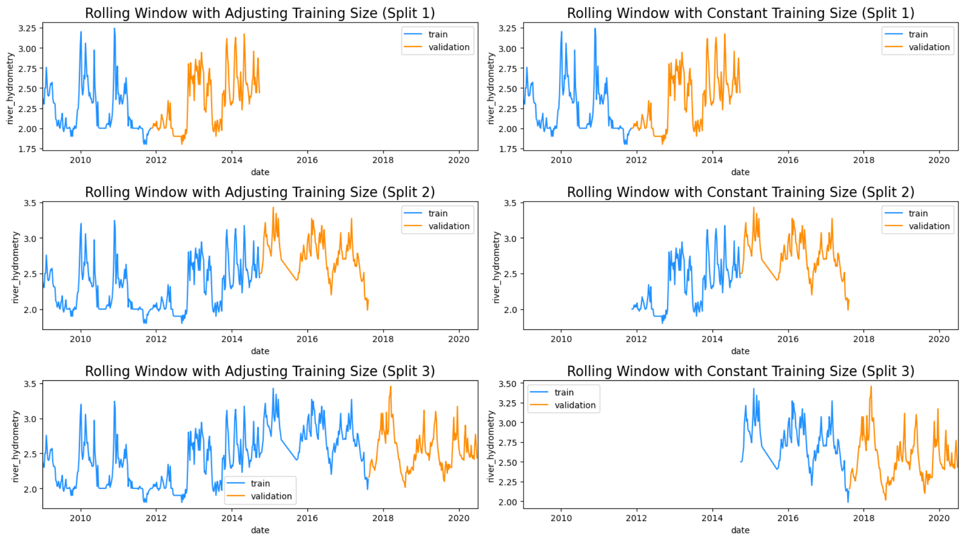

Figure 46: Rolling Window River Hydrometry

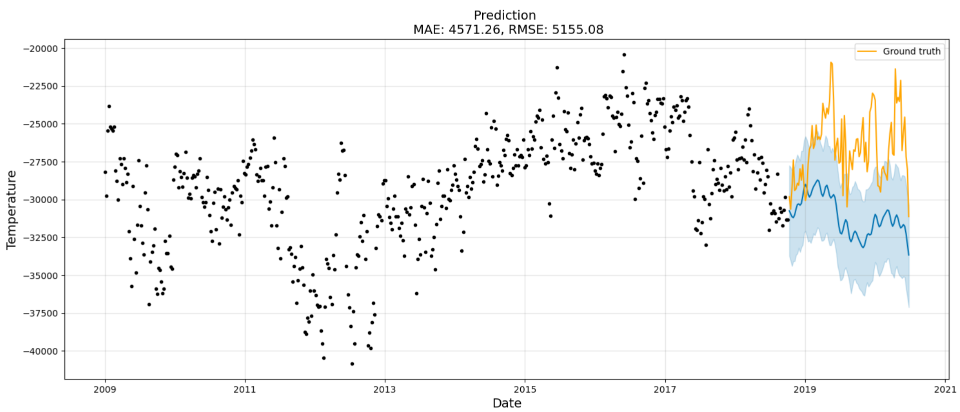

Figure 47: Prophet Forecast River Hydrometry

```
Performing stepwise search to minimize aic
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=9111.879, Time=0.03 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=11946.700, Time=0.01 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=inf, Time=0.01 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=inf, Time=0.08 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=9090.176, Time=0.23 sec
 ARIMA(2,0,0)(0,0,0)[0]             : AIC=inf, Time=0.02 sec
 ARIMA(3,0,1)(0,0,0)[0]             : AIC=9091.099, Time=0.23 sec
 ARIMA(2,0,2)(0,0,0)[0]             : AIC=9090.681, Time=0.26 sec
 ARIMA(1,0,2)(0,0,0)[0]             : AIC=9101.985, Time=0.04 sec
 ARIMA(3,0,0)(0,0,0)[0]             : AIC=inf, Time=0.03 sec
 ARIMA(3,0,2)(0,0,0)[0]             : AIC=inf, Time=0.57 sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=9092.787, Time=0.12 sec

Best model:  ARIMA(2,0,1)(0,0,0)[0]
Total fit time: 1.654 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  510
Model:               SARIMAX(2, 0, 1)   Log Likelihood               -4541.088
Date:                Sat, 08 Apr 2023   AIC                           9090.176
Time:                        09:11:25   BIC                           9107.114
Sample:                             0   HQIC                          9096.817
                                - 510
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.6631      0.052     31.741      0.000       1.560       1.766
ar.L2         -0.6632      0.052    -12.661      0.000      -0.766      -0.561
ma.L1         -0.9053      0.031    -29.619      0.000      -0.965      -0.845
sigma2      3.235e+06   1.78e-09   1.82e+15      0.000    3.23e+06    3.23e+06
===================================================================================
Ljung-Box (L1) (Q):                   0.30   Jarque-Bera (JB):                29.89
Prob(Q):                              0.58   Prob(JB):                         0.00
Heteroskedasticity (H):               0.71   Skew:                            -0.17
Prob(H) (two-sided):                  0.03   Kurtosis:                         4.14
===================================================================================
```

Figure 48: Auto-ARIMA Summary River Hydrometry

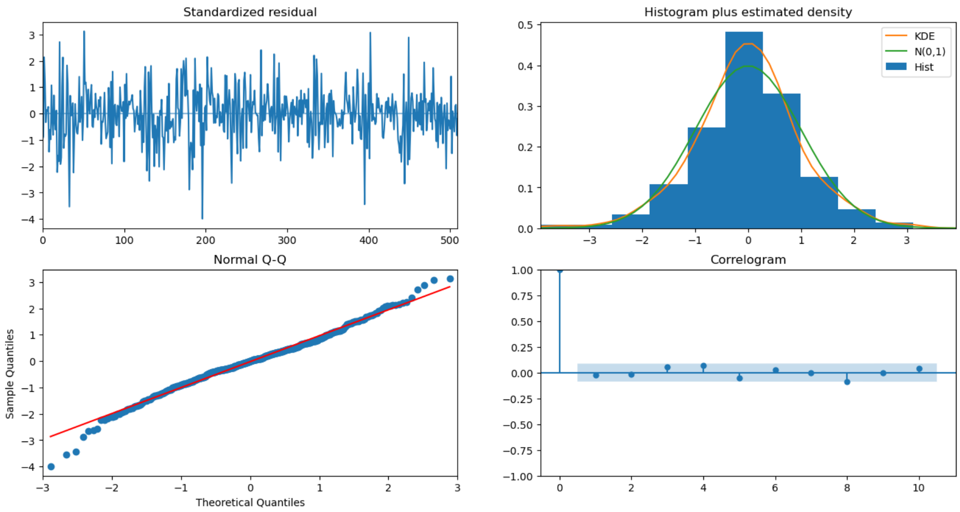

Figure 49: Model Diagnostics River Hydrometry

References