LAKE MY BREADTH AWAY

Ethan Arsht, Jaskirat Kaur, Setu Loomba, Grey Xu

MAIN CONTENT

01

INTRODUCTION

Water bodies, especially lakes, play a critical role in the global hydrological cycle. The rate of evaporation and loss of water bodies can significantly influence local and regional climates, water availability, and ecosystem health. Over the years, understanding lake evaporation dynamics has become crucial due to the increasing impacts of climate change on freshwater resources.

Attributes:

The processed GLEV dataset comprises the following attributes:

1. **Lake Open Areas:** Monthly data capturing the surface areas of 1.42 million lakes. This attribute is vital for understanding the spatial extent of lakes and how it varies over time, which can be influenced by factors such as seasonal changes, human activities, and climate patterns.
2. **Evaporation Rates:** Monthly evaporation rates from these lakes, offering insights into the speed at which water is lost to the atmosphere. By analyzing this attribute, one can gauge the factors affecting evaporation rates, such as temperature, humidity, and wind speed.
3. **Weather Data:** maximum daily temperature, minimum daily temperature, precipitation, snow depth from the national center of environmental information (NCEI). This have a series of the weather stations and we select a specific station as a case study for understanding the relationship between weather and water evaporation rate.

THOUGHT MAPS

EVAPORATION RATE

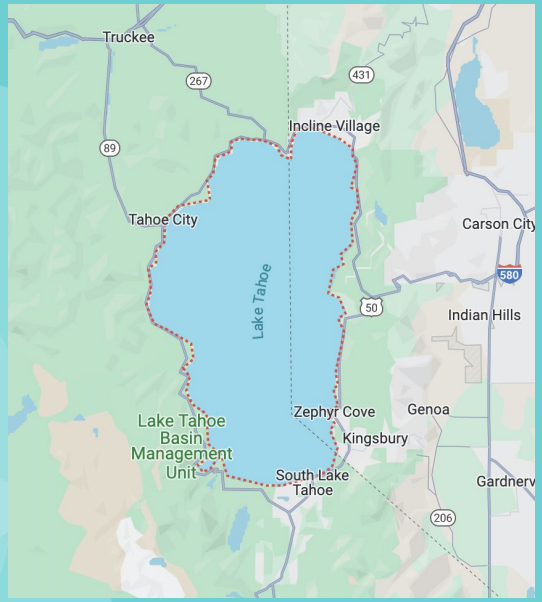
WEATHER DATA

WATER BODY LOSS

02

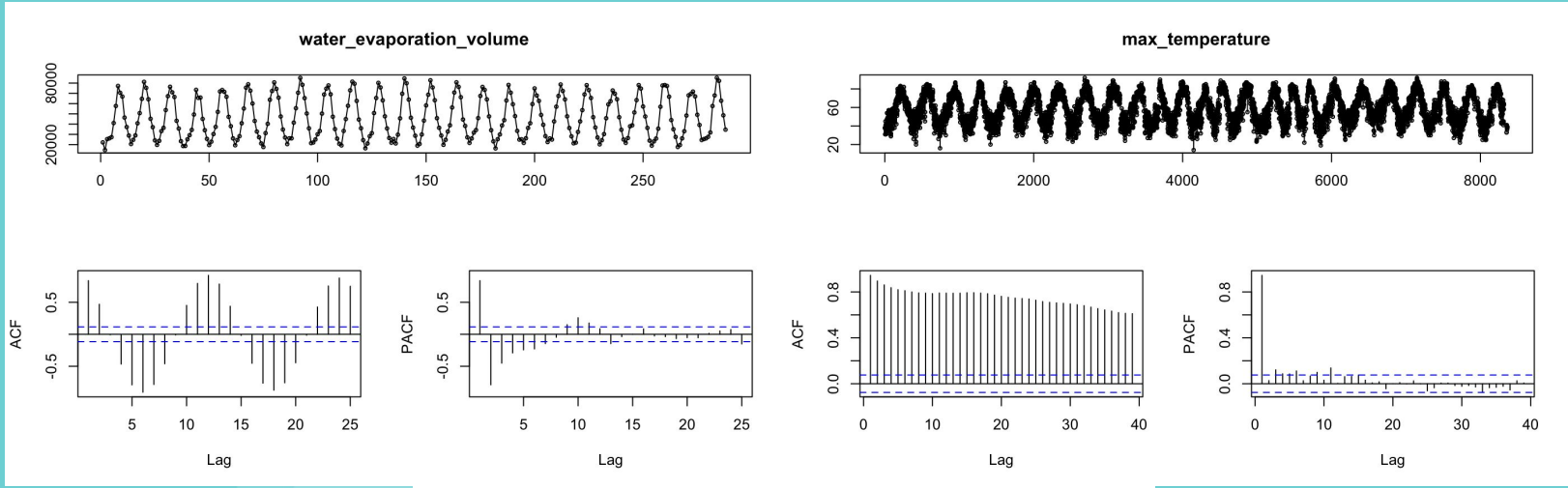
WATER EVAPORATION AND WEATHER

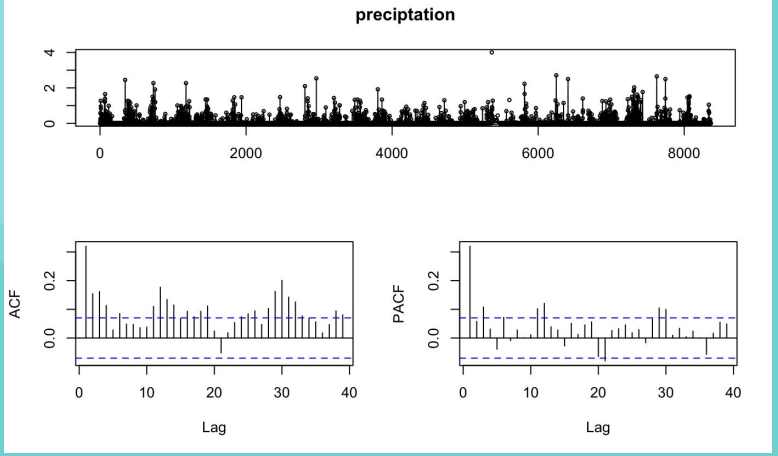
CASE STUDY: LAKE TAHOE



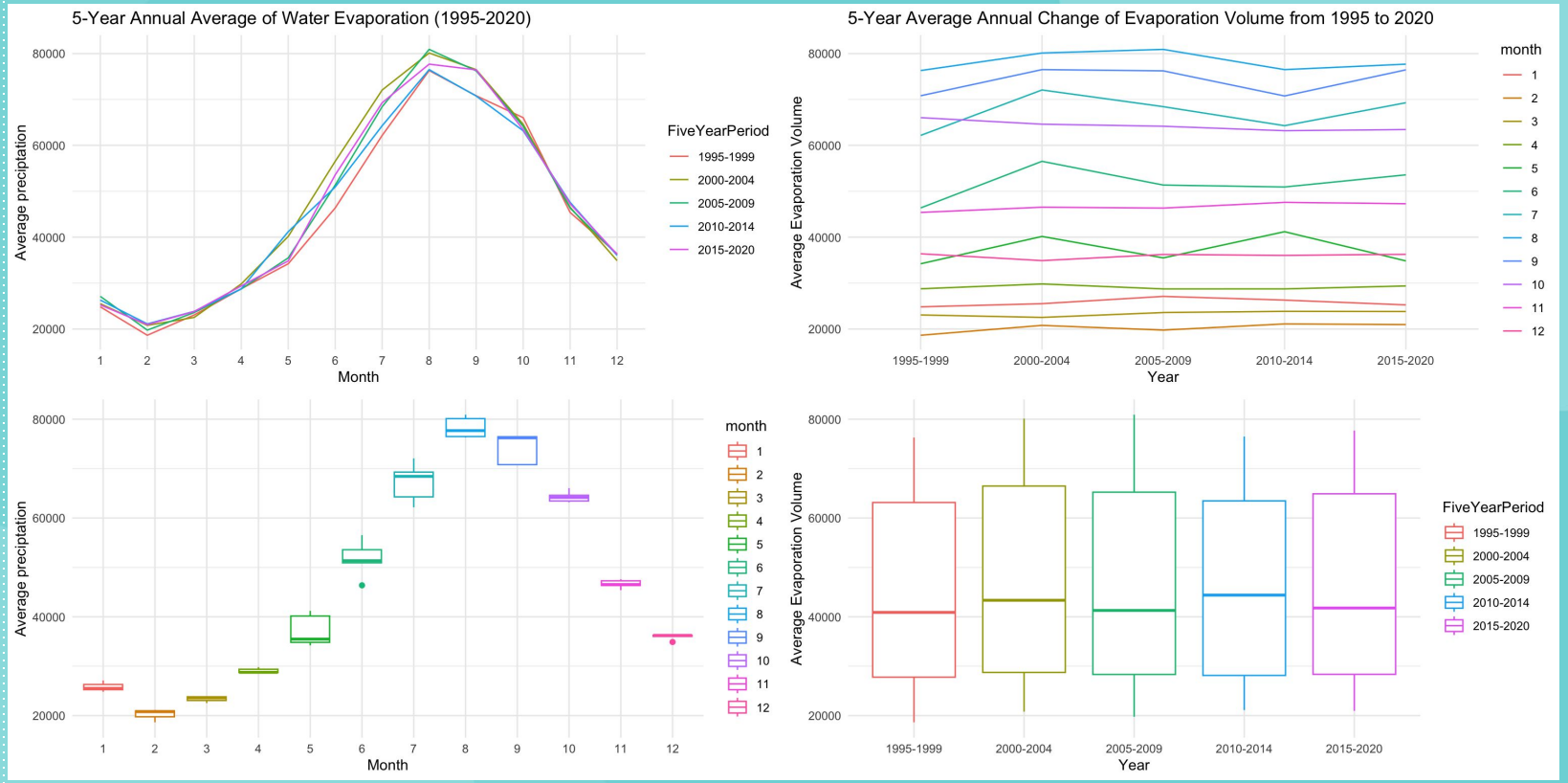
Lake Tahoe, a large freshwater lake in the Sierra Nevada mountains of the United States, straddling the border between California and Nevada, is an ideal subject for a case study given its fluctuating weather and temperature condition around the season.

TIME SERIES PLOTS

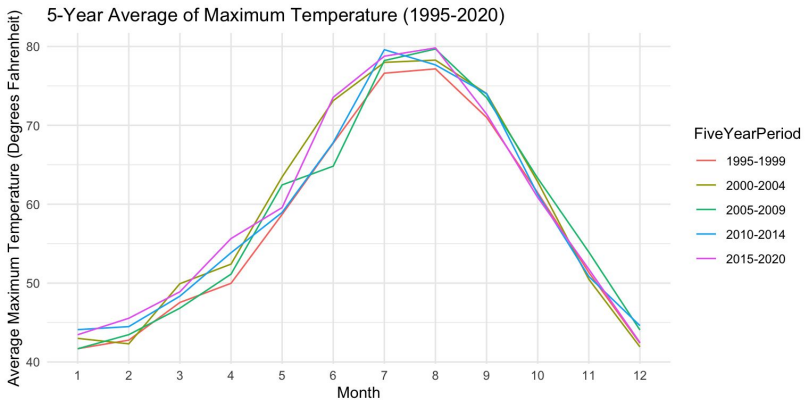


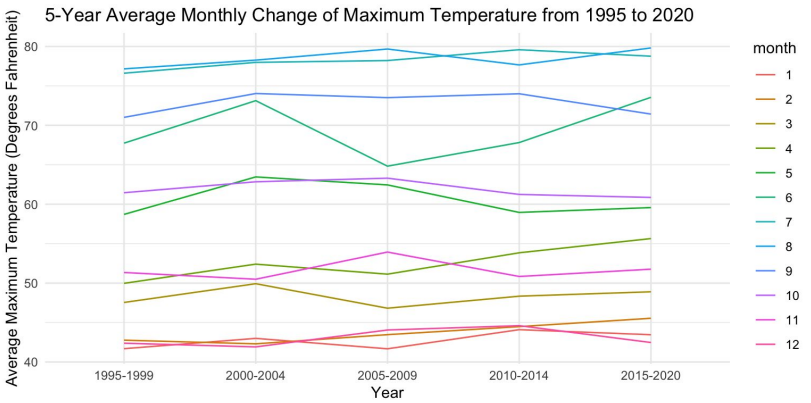


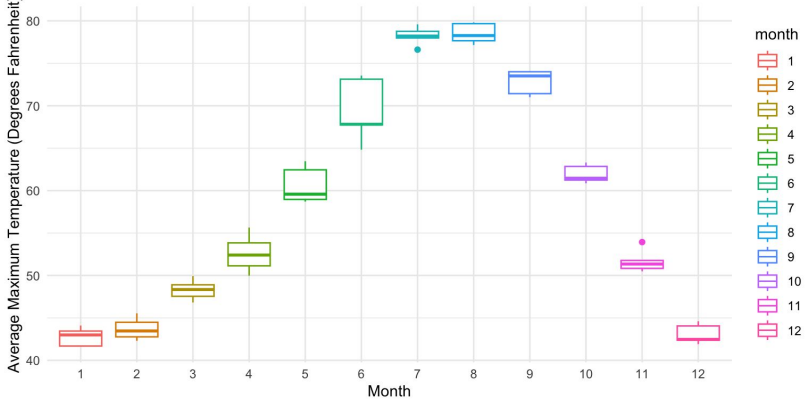
WATER EVAPORATION

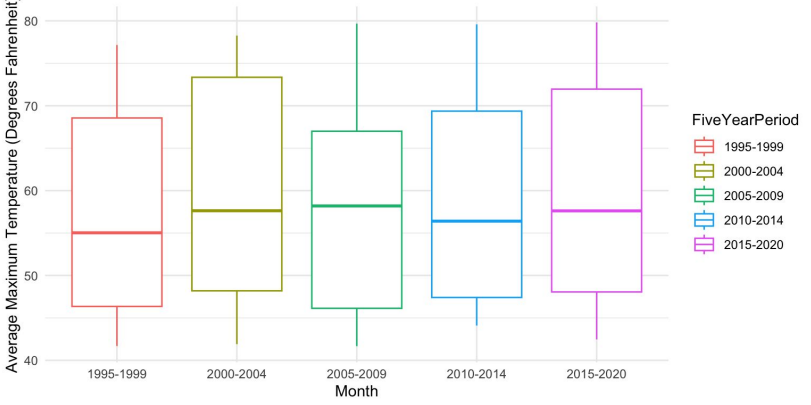


MAXIMUM TEMPERATURE

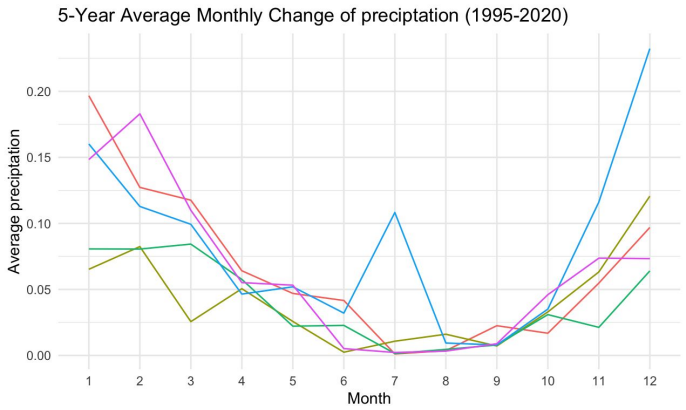


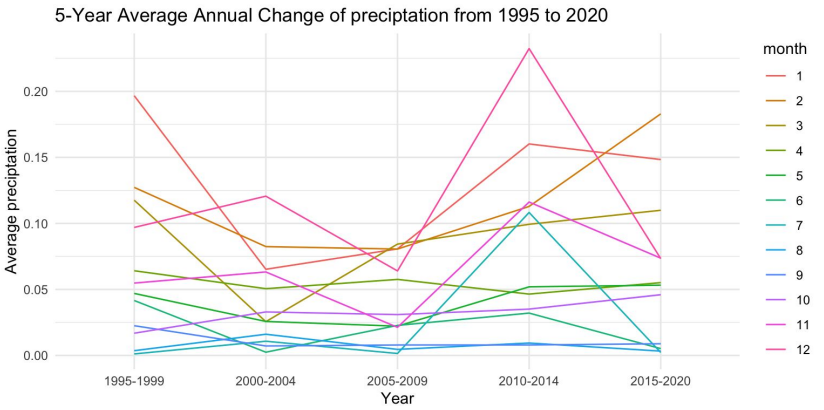


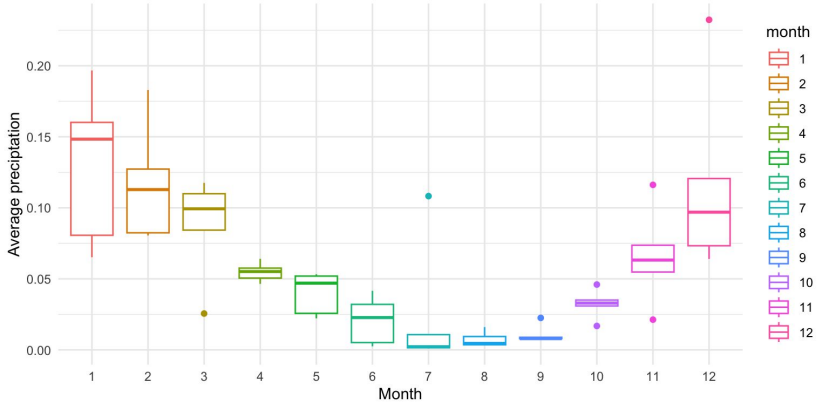


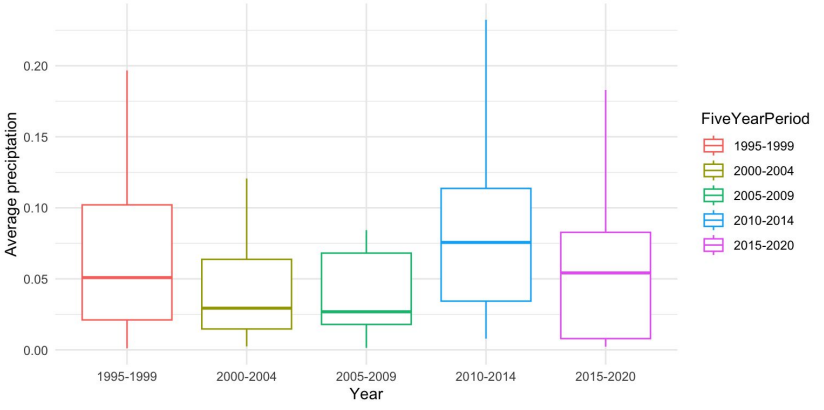


PRECIPITATION







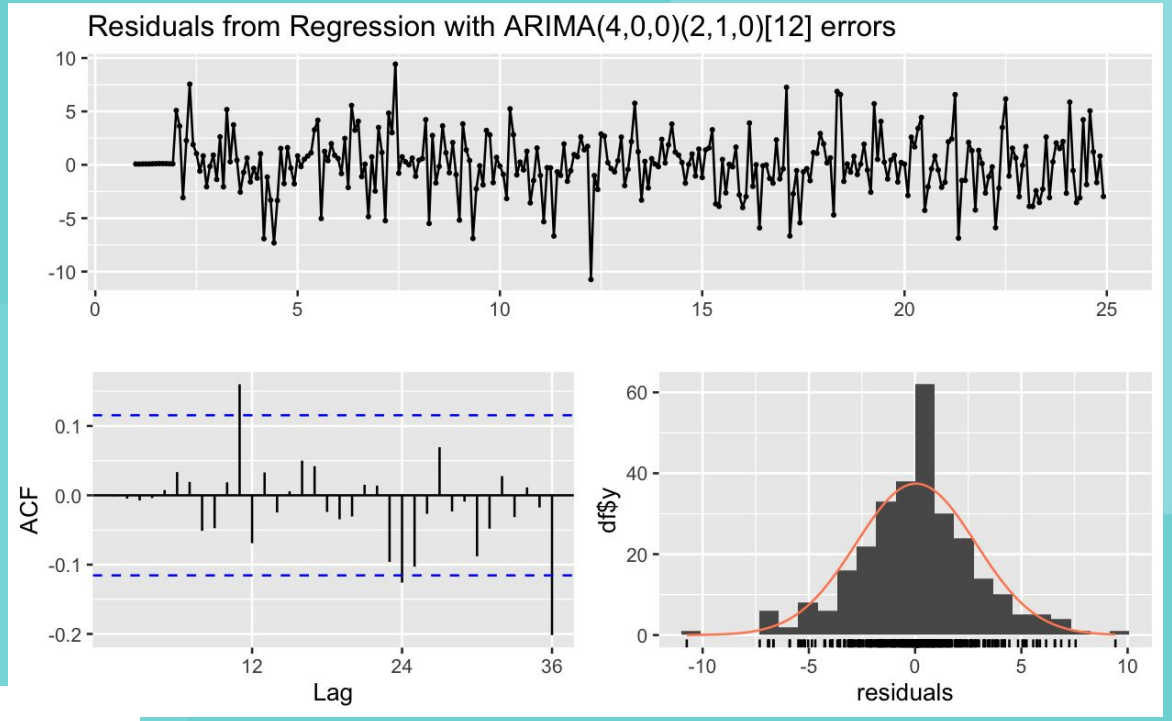


- **Data Characteristics:**
  a. **Seasonality:**
     - Strong seasonality observed in water evaporation rate and temperature.
     - Precipitation also shows seasonality but with opposite peaks compared to temperature and evaporation .
  b. **Trend:**
     - No clear trend in evaporation rate and precipitation.
     - Steady increase in average daily maximum temperature noted over time, with some fluctuations in the past decade.
  c. **Multiplicative/Additive:**
     - The seasonality in the data appears to be additive.
- **Correlation Analysis:**
  a. Temperature and water evaporation volume: Correlation coefficient of 0.846055.
  b. Precipitation and evaporation volume: Correlation coefficient of -0.3956236.
- **Implication for Modeling:**
  a. Consider using ARIMAX to incorporate these features (temperature and precipitation) as regressors in our models.

COMPARING MODELS

	AICc	MAE	RMSE
Seasonal Naive		3925.979	5306.727
ARIMA(4,0,1)(2,1,0)[12]	1387.18	2925.805	3896.681
ARIMAX with max temperature as a regressor (4,0,1)(2,1,0)[12]	1391.84	2880.762	3920.87
ARIMAX with precipitation as a regressor (4,0,1)(2,1,0)[12]	1395.1	2898.258	3895.713
ARIMAX with max temperature and precipitation as regressors (4,0,0)(2,1,0)[12]	1393.23	2864.487	3909.507

BEST MODEL:




```
        Ljung-Box test

data:  Residuals from Regression with
ARIMA(4,0,0)(2,1,0)[12] errors
Q* = 21.967, df = 18, p-value = 0.2334

Model df: 6.   Total lags used: 24
```

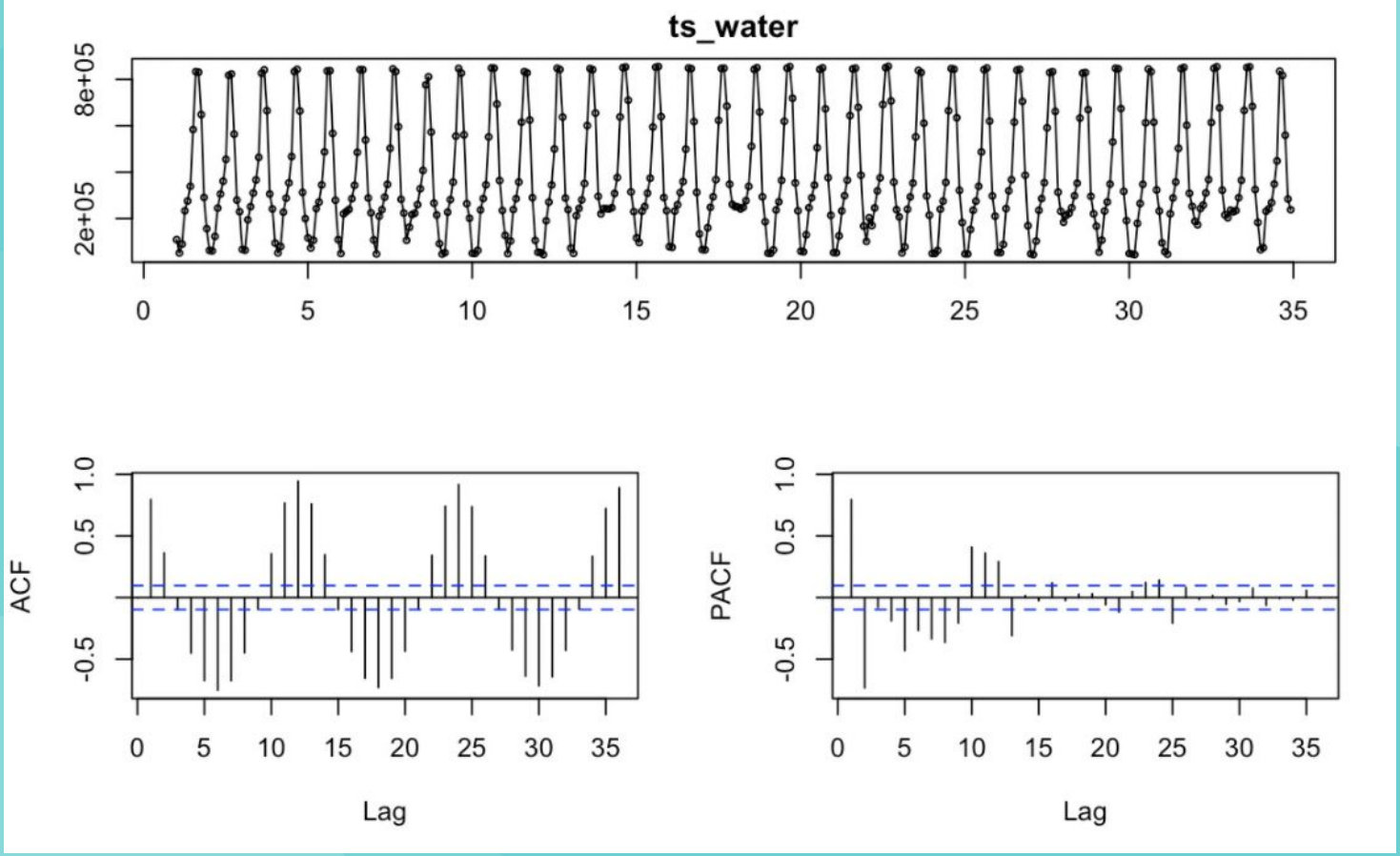
03

WATER BODY LOSS AND EVAPORATION

DATA INPUT

- **Objective:** Utilize predicted water evaporation rates from weather data to analyze changes in water bodies.
- **Data Adjustment:** Time series inputs modified to include a broader range of lake data across the U.S., moving beyond a single-lake focus.
- **Methodology:** Employ a random sample of 100 lakes, calculating average size and evaporation rates to identify general lake body patterns.
- **Limitation:** Each station's weather data must be meticulously downloaded and aligned with its respective lake, followed by a comprehensive data cleaning process. We aim to refine this aspect of our research in future studies.
- **Current Approach:** Given the strong correlation between water evaporation rates and weather variables like temperature, evaporation rate is used as the primary regressor.

OPEN WATER AREA DATA



DATA CHARACTERISTICS

Trend Analysis: The data does not exhibit a strong or consistent pattern.

Seasonality: The presence of a 12-month cycle suggests that seasonal differencing is necessary to achieve data stabilization.

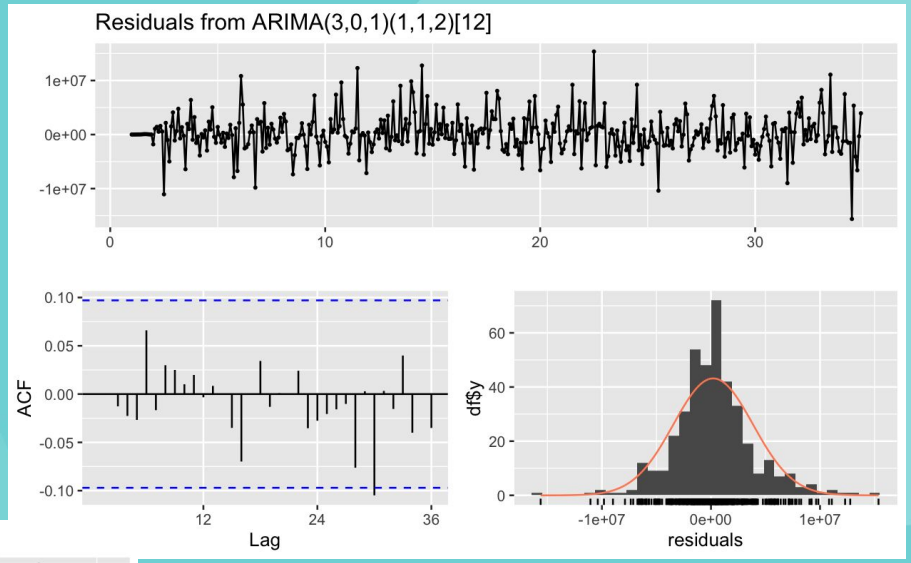
Autocorrelation Function (ACF): The autocorrelation does not show a decaying pattern across all lags, indicating non-stationarity.

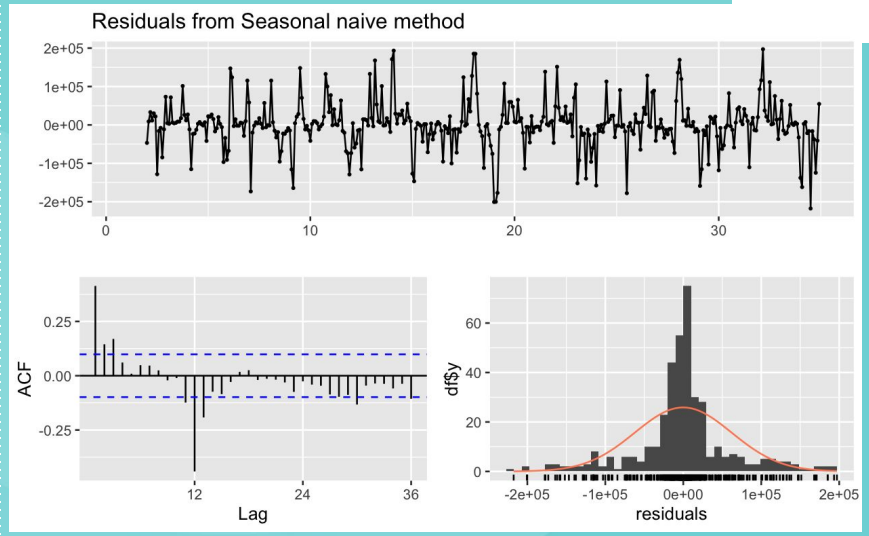
Partial Autocorrelation Function (PACF): There is evidence of a non-seasonal component with a notable drop at lag 3. Additionally, a seasonal decay pattern is observed with a significant drop at lag 12, suggesting a possible seasonal order of p = 1.

Correlation Analysis: There is a strong correlation (0.8039305) between the water area and evaporation volume.

Implication for Modeling: Given these findings, it would be appropriate to consider using ARIMAX or VAR models. Including the evaporation volume as a regressor in the ARIMA model for water area could enhance the predictive accuracy.

BASELINE MODELS: ARIMA VS SNAIVE

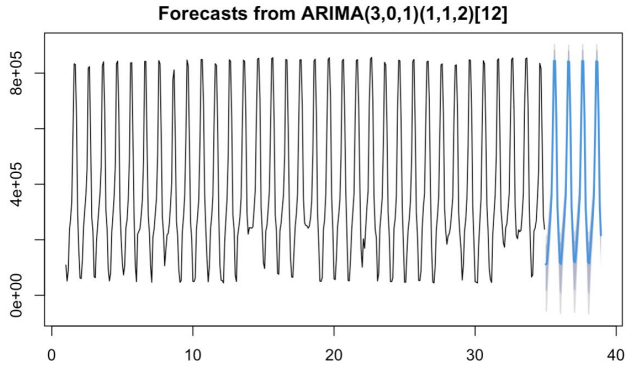




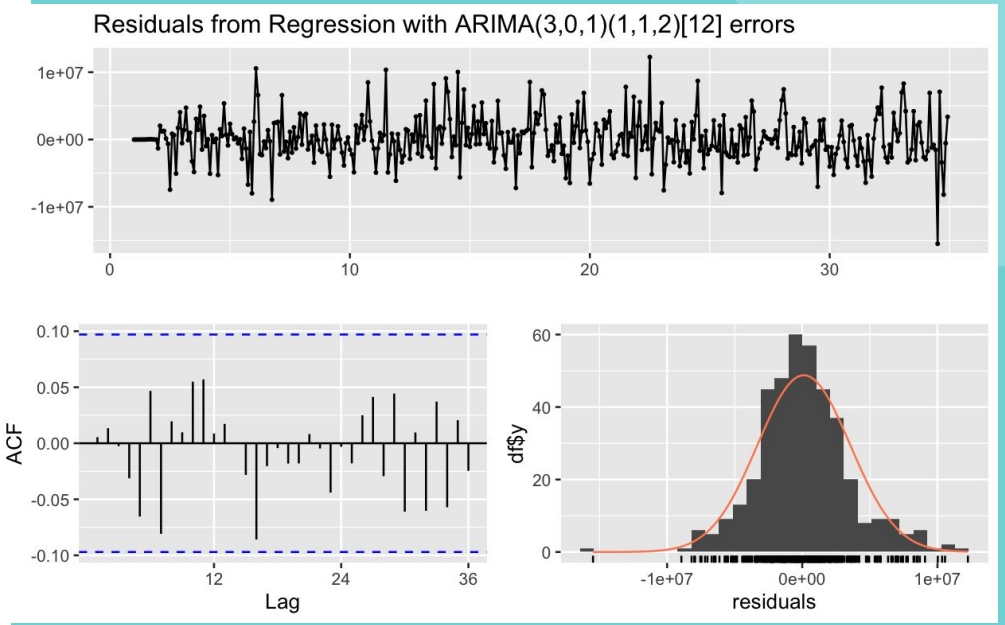
MODEL PERFORMANCE

	AICc	MAE	RMSE
Seasonal Naive		39583.94	62043.26
ARIMA(3,0,1)(1,1,2)[12]	13136.25	29324.31	40756.01
ARIMAX with max evaporation volume as a regressor (3,0,1)(1,1,2)[12]	13089.51	28866.66	39496.1
ETS(A,N,A)	14851.65	43207.6	32174.17

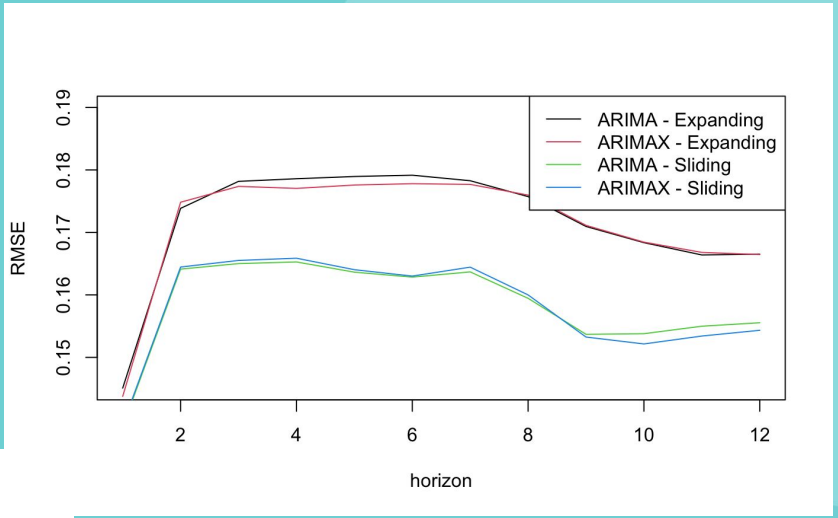
BEST MODEL: ARIMAX

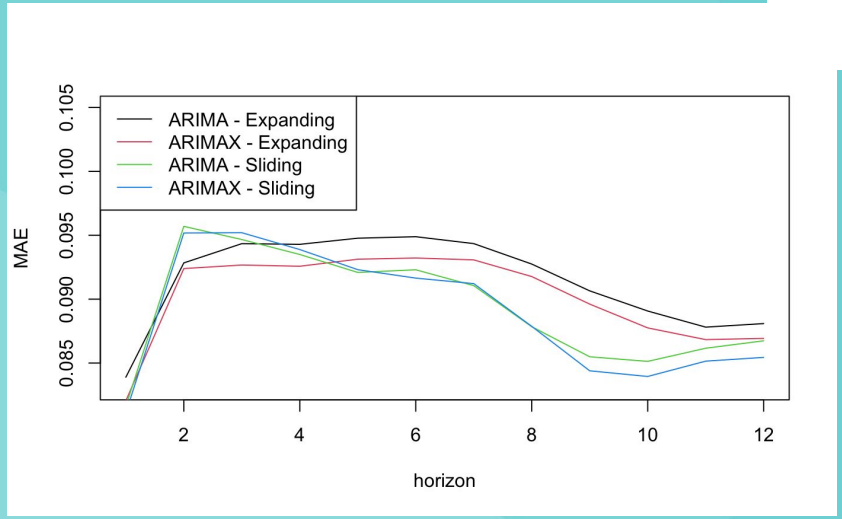



```
        Ljung-Box test

data:  Residuals from Regression with
ARIMA(3,0,1)(1,1,2)[12] errors
Q* = 13.736, df = 17, p-value = 0.6857

Model df: 7.   Total lags used: 24
```



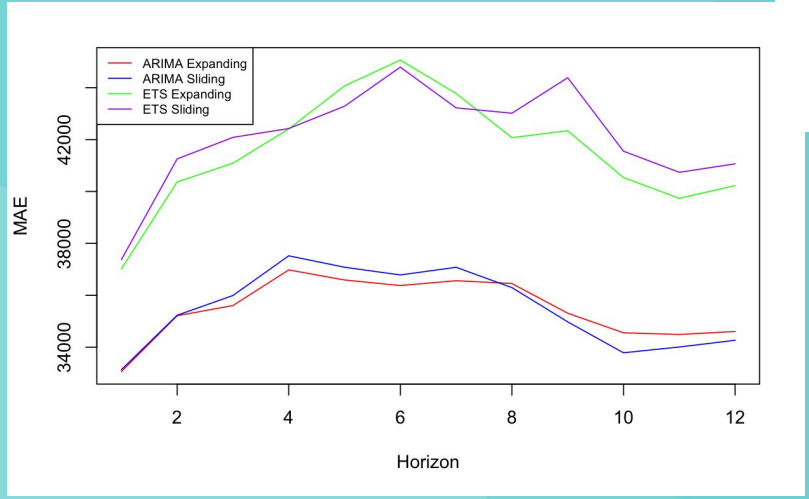
CROSS VALIDATION RESULTS

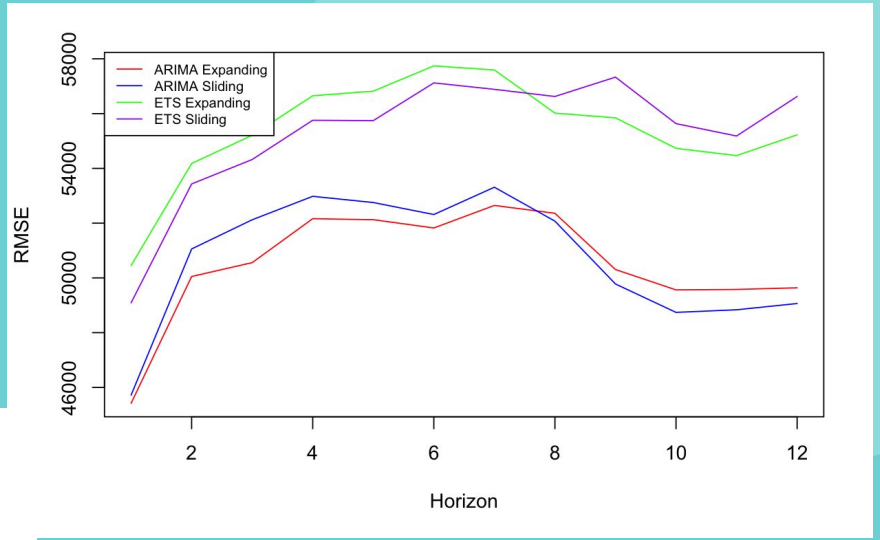




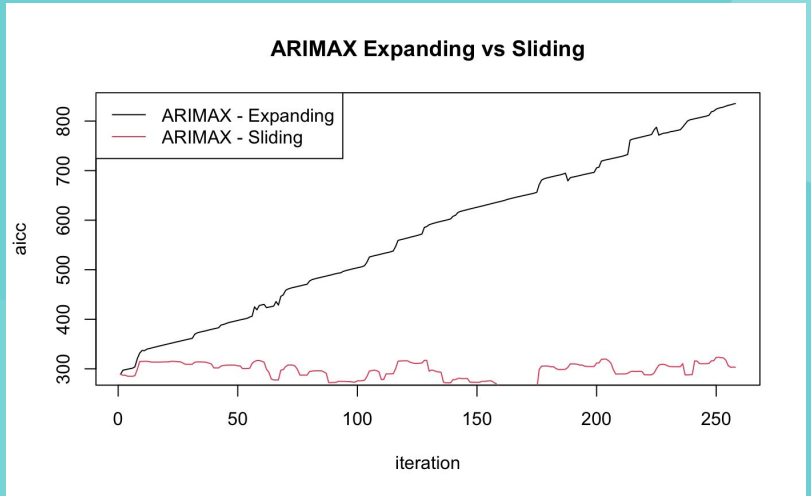
Side note: model can't be estimated using original scale over certain time windows, we have applied a logarithmic transformation (log base 10) to the water body data for this particular cross validation.

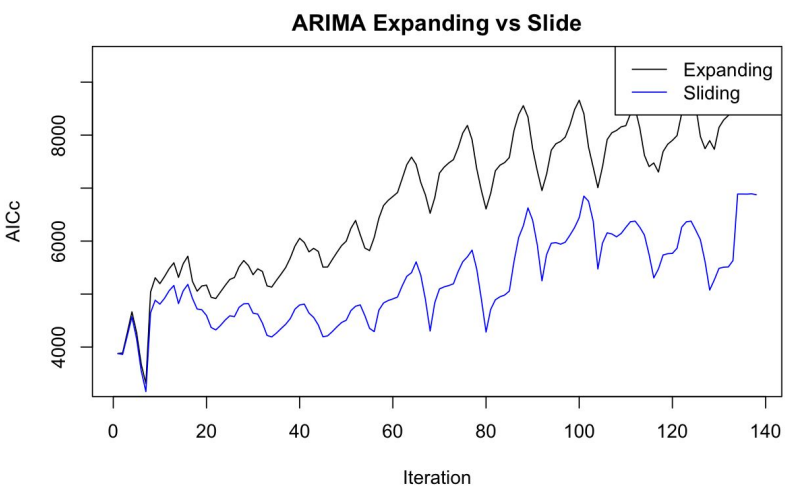
CROSS VALIDATION RESULTS

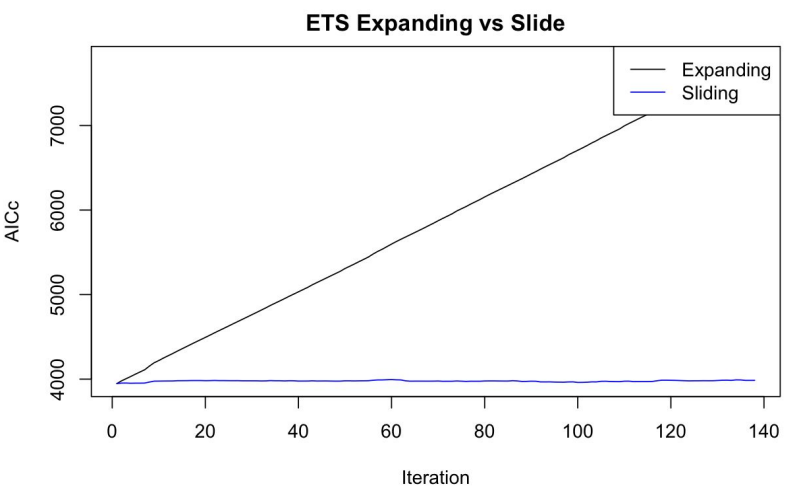




CROSS VALIDATION RESULTS



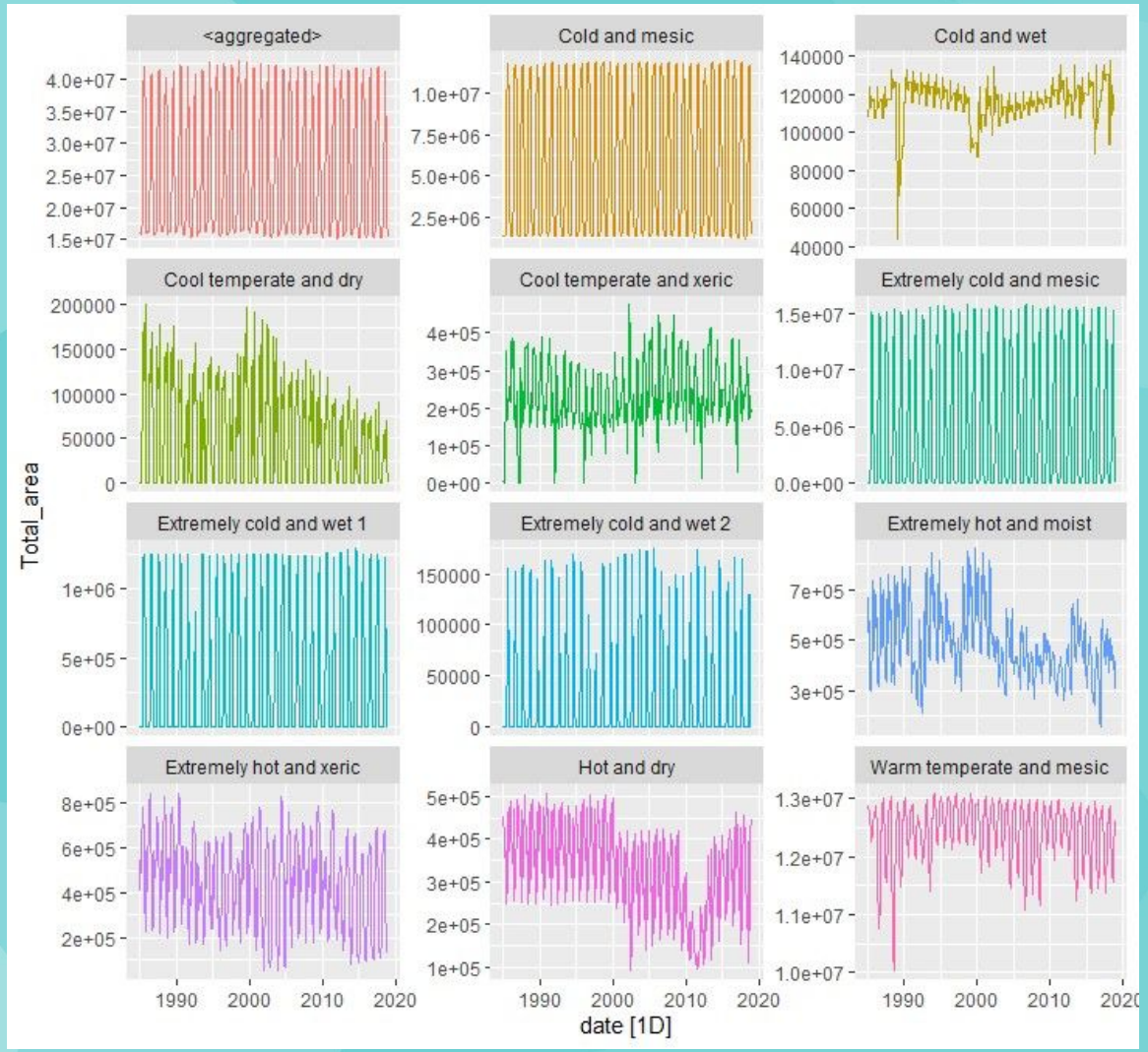




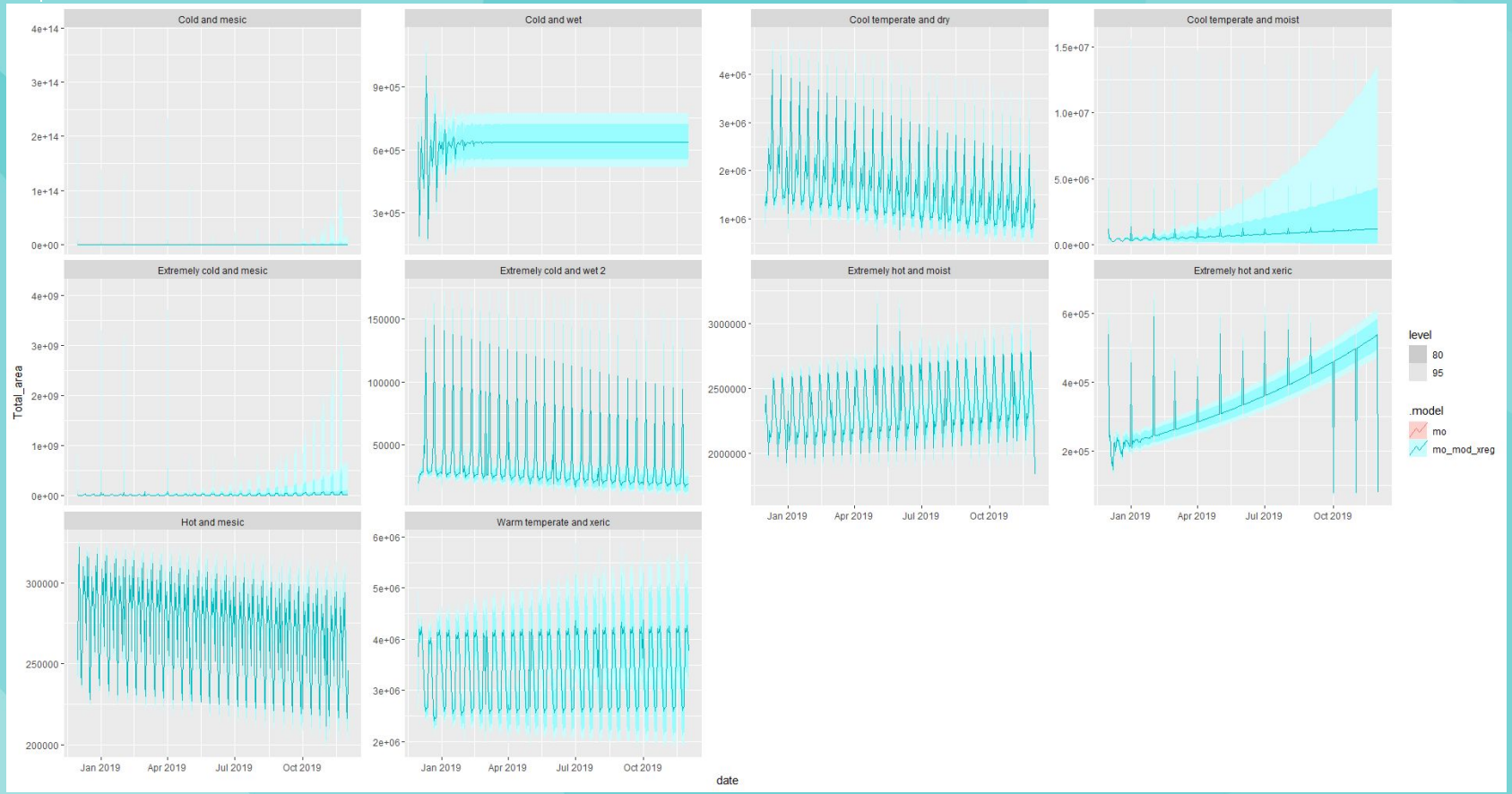
04

LAKE CLASSIFICATION

LAKE HIERARCHY



HIERARCHICAL FORECASTS



05

WEB APPLICATION ON PREDICTION

https://ethanapps.shinyapps.io/lake_predictor/