Time Series Application on Lake Erie Monthly Water Levels (1921-1970)

Line Shi Traky Deng Wendy Xu

Introduction

This poster provides a graphical representation for the Lake Erie time series project, in which models were developed to conclude and predict the behaviors of the water levels. The study data includes 600 monthly water levels of Lake Erie, from January 1921 to December 1970.

As the fourth-largest lake by surface area in the five Great Lakes in North America, Lake Erie exhibits prominent seasonal fluctuation in its water levels due to its small and shallow features. Understanding and having the ability to predict the behavior of the lake water levels are of interest in this study and in research as the water levels carry economic and environmental meanings for the surrounding households and businesses.

Project Breakdown

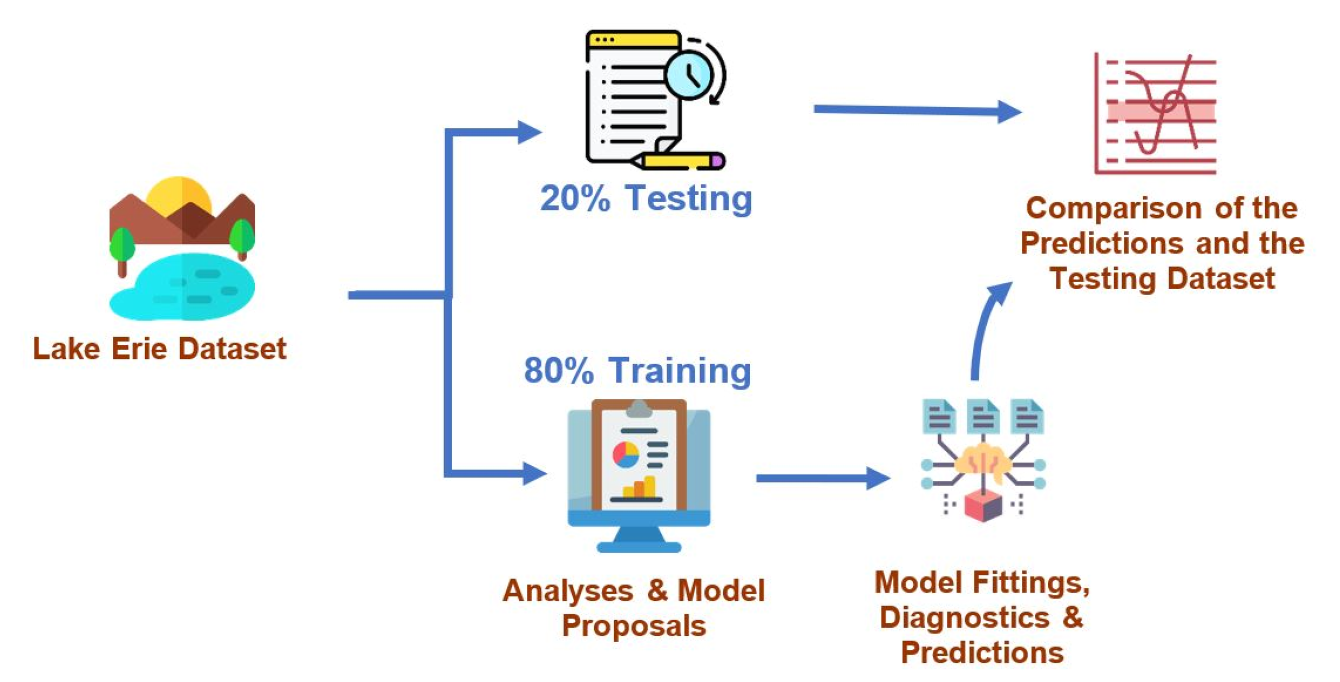


Figure 1: Flow Chart of the Project

Initial Observations

The following features were identified based on the original series:

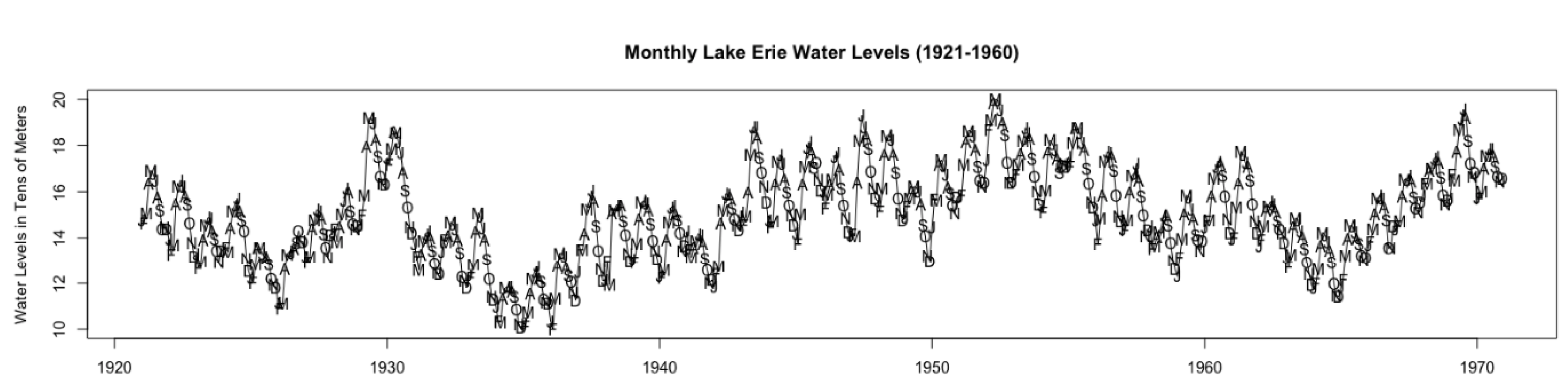


Figure 2: Monthly Water Level of Lake Erie 1921-1970

- **Non-constant mean** overtime
- Relatively **constant variance**
- Strong **seasonality** - higher water levels in summer months and lower water levels in winter months

Data Splitting

Based on the procedures proposed, we split the dataset into two subsets:

- **80%** for training the models (480 observations)
- **20%** for testing the models' predictions (120 observations)

Analyses & Model Proposals

Based on the plot of the original dataset (Figure 2), we performed a first degree differencing.

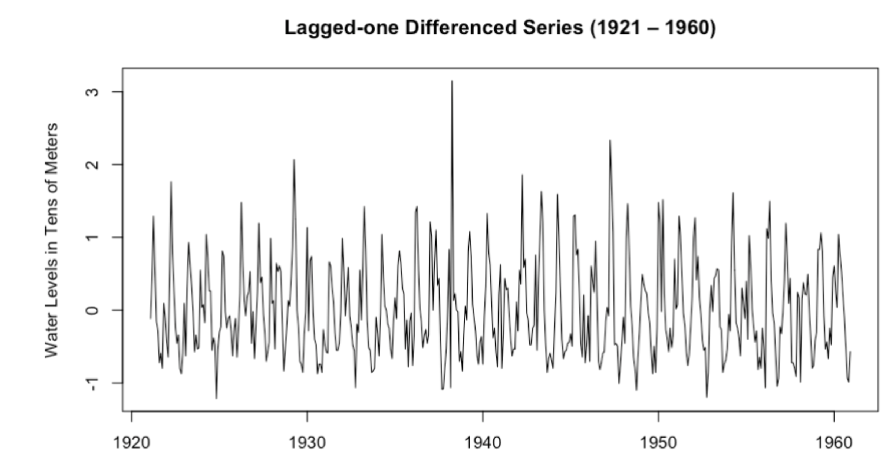


Figure 3: Plot of Lagged-One Differenced Series

According to Figure 3, the Dickey-Fuller test, and the periodogram of the transformed data, we conclude the data:

- has a **constant mean** around 0 and a **constant variance**
- is **stationary** (Dickey-Fuller test with $p < 0.01$)
- has a **seasonality** of 12

We first fit the differenced series to a simple **deterministic trend model** with a seasonal component of 12. After fitting the model, we examine the ACF and PACF plots (Figure 4) to answer of question whether the model describe the behavior of the data well:

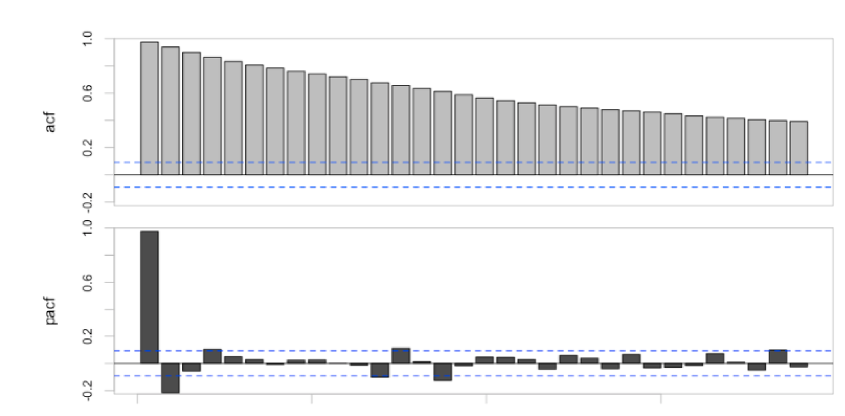


Figure 4: ACF and PACF Plot of Residuals

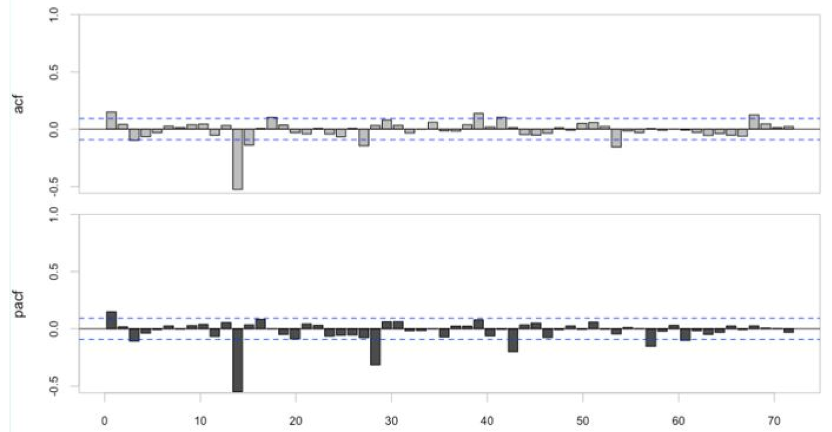


Figure 5: ACF and PACF Plot of Lagged-One Seasonal Differenced Dataset

While the seasonality seems to be modelled, it is apparent that there are significant non-seasonal autocorrelations at many lags. Recognizing the limitation of the capability of a deterministic trend model, we opt for a more flexible seasonal model, a **seasonal ARIMA** (a.k.a. **SARIMA**) models, in which we can model both seasonal and non-seasonal behaviors multiplicatively.

To identify the SARIMA orders, we take a seasonal difference on top of the once differenced series recoginizing that the twice seasonal differencing over-difference the series. The ACF and PCAF plots can be seen in figure 5, from which we propose the following six models:

Model	AIC
ARIMA $(1,1,0)\times(1,1,1)_{12}$	522.16
ARIMA $(1,1,0)\times(0,1,1)_{12}$	511.74
ARIMA $(1,1,1)\times(1,1,1)_{12}$	553.87
ARIMA $(1,1,1)\times(0,1,1)_{12}$	553.48
ARIMA $(3,1,3)\times(1,1,1)_{12}$	550.16
ARIMA $(3,1,3)\times(0,1,1)_{12}$	550.56

Models with an AR order of 1 in their seasonal components generally perform worse than models without an AR order of 1 in their seasonal components. The MA(1) in the non-seasonal component also does not seem to be improving the model fit. We narrow our selection down to two models: $ARIMA(1,1,0)\times(0,1,1)_{12}$ and $ARIMA(3,1,3)\times(1,1,1)_{12}$.

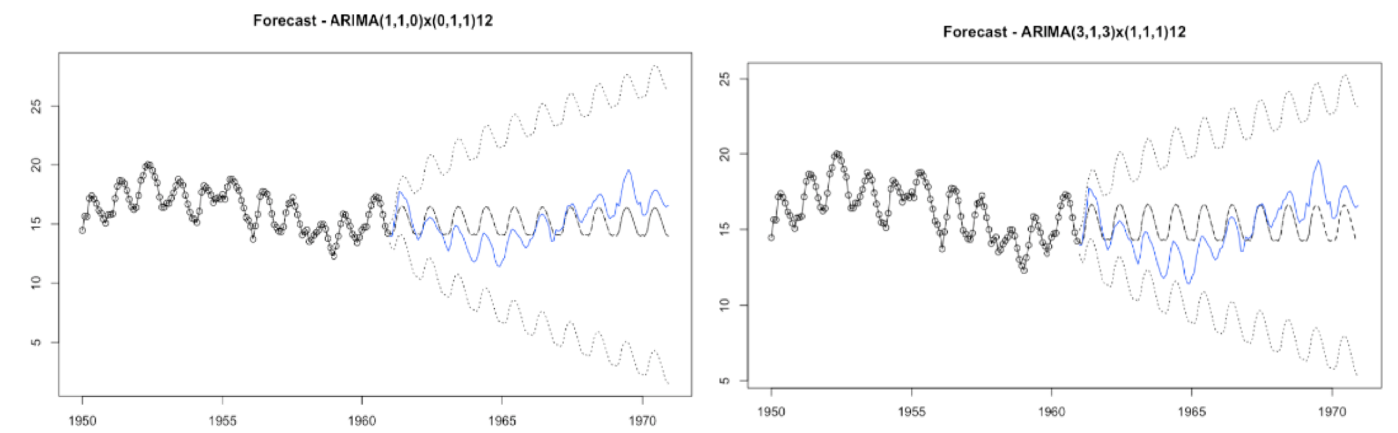
Model Selections

The principle of parsimony was respected during the model selection, which motivates us to fit and predict using a simpler model with a slightly worse AIC, and a more complex model with the smallest AIC amongst all the proposed.

Model	Equations
ARIMA$(1,1,0)\times(0,1,1)_{12}$	$\underbrace{(1-0.2006B)}_{\text{non-seasonal AR(1)}}\underbrace{(1-B)}_{\text{non-seasonal difference(1)}}\underbrace{(1-B^{12})}_{\text{seasonal difference(1)}}y_t=\underbrace{(1-B^{12})}_{\text{seasonal MA(1)}}e_t$
ARIMA$(3,1,3)\times(1,1,1)_{12}$	$\underbrace{(1-0.7124B)(1-0.0655B)(1+0.1842B)}_{\text{non-seasonal AR(3)}}\underbrace{(1-B)}_{\text{non-seasonal difference(1)}}\underbrace{(1+0.0743B^{12})}_{\text{seasonal AR(1)}}\underbrace{(1-B^{12})}_{\text{seasonal difference(1)}}y_t = \underbrace{(1-0.5217)(1-0.1059B)(1+0.0172B)}_{\text{non-seasonal MA(3)}}\underbrace{(1-B^{12})}_{\text{seasonal MA(1)}}e_t$

Predictions

Figure 6 illustrate the 10-year predictions generated by the two selected models.



$\text{ARIMA}(1,1,0)\times(0,1,1)_{12}$ $\text{ARIMA}(3,1,3)\times(1,1,1)_{12}$

Figure 6:Predictions vs. Testing dataset

The predictions of both models successfully mimic the approximate seasonal trend in the original series but have a less fluctuated mean compared with the original series.The ARIMA $(1,1,0)\times(0,1,1)_{12}$ model generates slightly lower monthly forecast than the ARIMA $(3,1,3)\times(1,1,1)_{12}$ model and a wider 95% prediction intervals.

Conclusions

In this study, we applied time series modelling techniques, including model proposal and residual diagnosis, to the Lake Erie dataset to understand and predict the behaviours of the monthly water levels (in tens of meters). The two models we selected based on autocorrelation and partial autocorrelation plot produce similar predictions; however, it is difficult to determine which model is more suited for our data. We ended our analyses on the note that the application of seasonal ARIMA models is capable of capturing and simulating certain behaviours of the data, but cannot be overarching with the practice of parsimony and the consideration of the variance-bias tradeoff.

References

Changnon, S. A. (2004). Temporal Behavior of Levels of the Great Lakes and Climate Variability.*Journal of Great Lakes Research, 30*(1), 184–200.

Cryer, J. D., Chan, K.-sik. (2011). *Time series analysis with applications in R.* New York: Springer.

The Editors of Encyclopaedia Britannica. (2019, February 20). Lake Erie. Retrieved from https://www.britannica.com/place/Lake-Erie