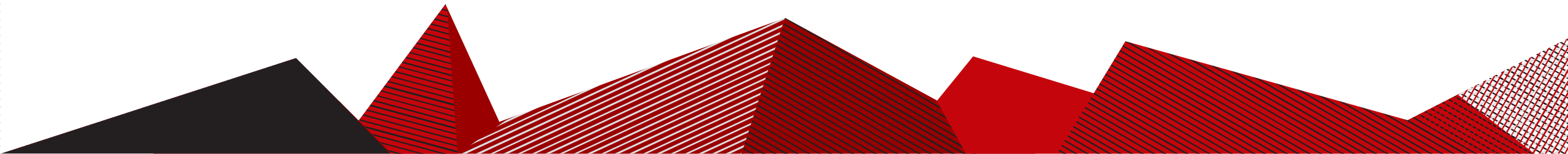
Global Historical Precipitation Time Series Analysis

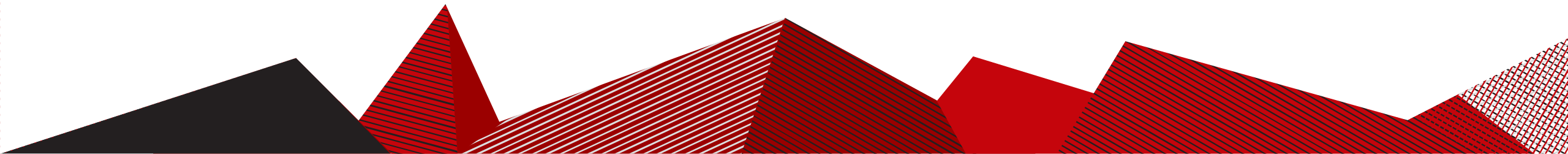
Yuhan Meng, Xueqian Zhang, Chenghui Li, Jitian Zhao

Outline：

- Data Description
- Data Preprocessing
- CHTC data processing
- Result

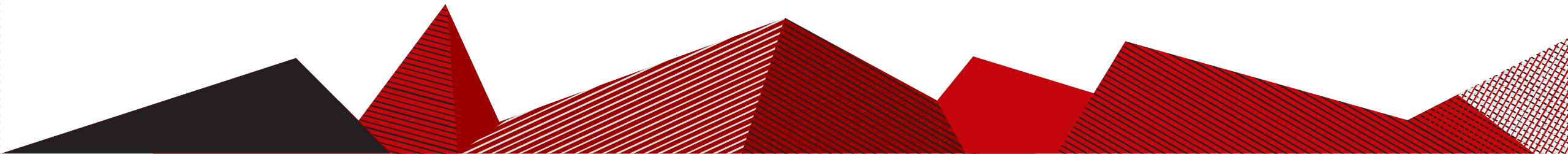


Data Description

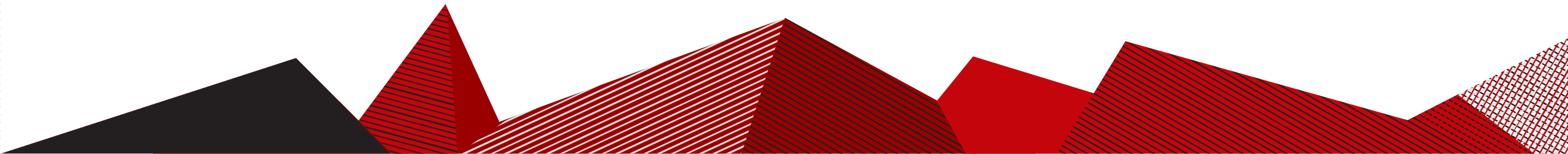


Data description

- **URL**: https://www.ncei.noaa.gov/data/global-historical-climatology-network-daily/
- **Data Context**: The Global Historical Climatology Network–daily dataset is a set of daily climate summaries from thousands of weather stations around the world.
- **Data Format**: Each file represents climate data collected over years at one station. There are a lot of missing values in our data. Some station might lack **precipitation** data and use snowfall instead, some only have data in a short period of time.



Data Preprocessing



Data processing

DATE	NAME	PRCP	PRCP_ATTRIBUTES	SNOW	SNOW_ATTRIBUTES
2006-07-28	FREETOWN 1.0 NE, NY US	0	,,N		
2006-07-29	FREETOWN 1.0 NE, NY US	25	,,N		
2006-07-30	FREETOWN 1.0 NE, NY US	0	,,N		
2006-07-31	FREETOWN 1.0 NE, NY US	0	,,N		
2007-11-10	FREETOWN 1.0 NE, NY US			20	,,N
2007-11-16	FREETOWN 1.0 NE, NY US			15	,,N
2007-11-18	FREETOWN 1.0 NE, NY US			3	,,N
2007-11-19	FREETOWN 1.0 NE, NY US			10	,,N
2007-11-23	FREETOWN 1.0 NE, NY US			18	,,N
2007-11-24	FREETOWN 1.0 NE, NY US			3	,,N
2007-11-26	FREETOWN 1.0 NE, NY US			20	,,N
2007-11-28	FREETOWN 1.0 NE, NY US			5	,,N
2007-11-30	FREETOWN 1.0 NE, NY US			3	,,N

Data processing

- **Problems**:

❖Missing value, the date of rainfall is not continuous and even lose data of whole months.

❖When a day doesn’t rain the record is 0.

❖Snowfall records instead of rainfall records.

- **Solutions:**

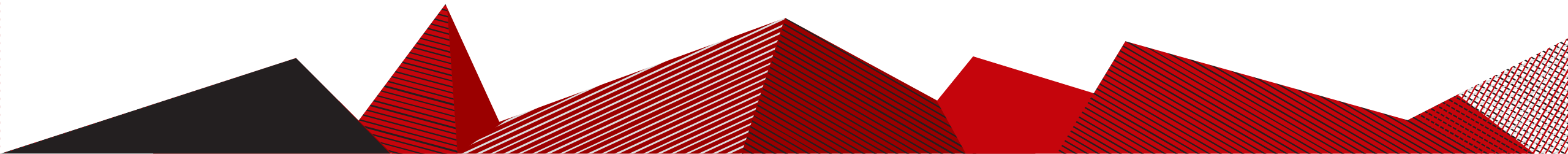
❖Use the mean of season to constrain it.

❖Choose file that every seasons has data.

❖Consider data longer than 7 years.

❖Convert snowfall to rainfall by multiplying $\frac{1}{15}$.

CHTC data processing



CHTC data processing

Location Selection

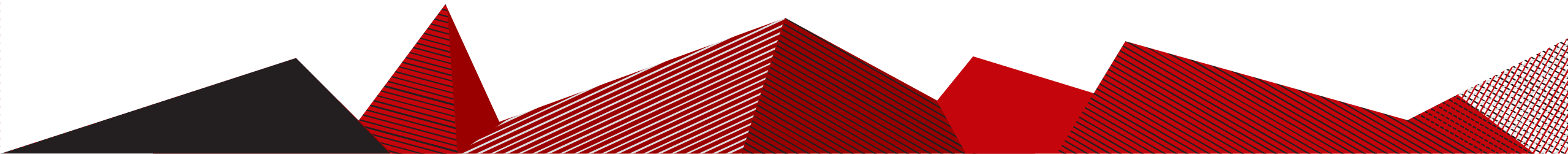
Write a **match.sh** to choose only stations from CA, CH, UK and US these four large countries.

spliting and parallel processin

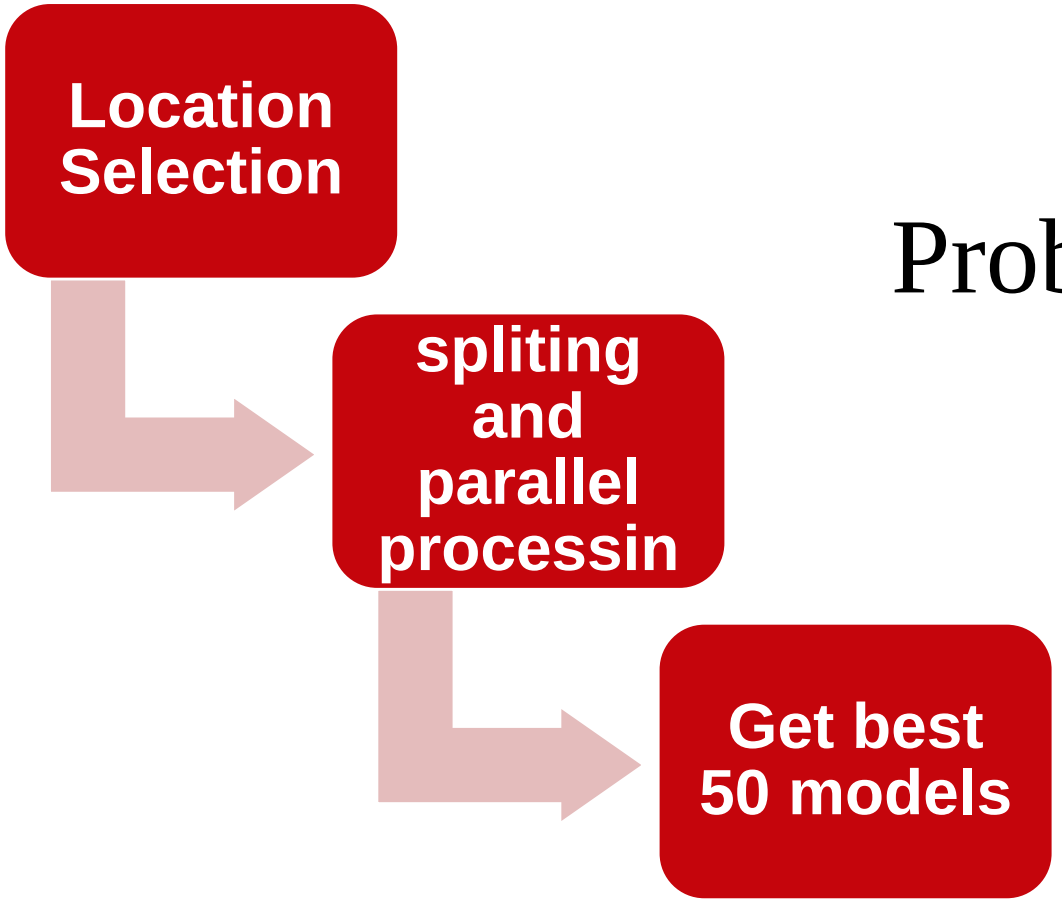
Fit ARIMA model for each time series and compute AIC.

Get best 50 models

Merge the output AIC together and sort them from smallest to largest and extract only the first 50.



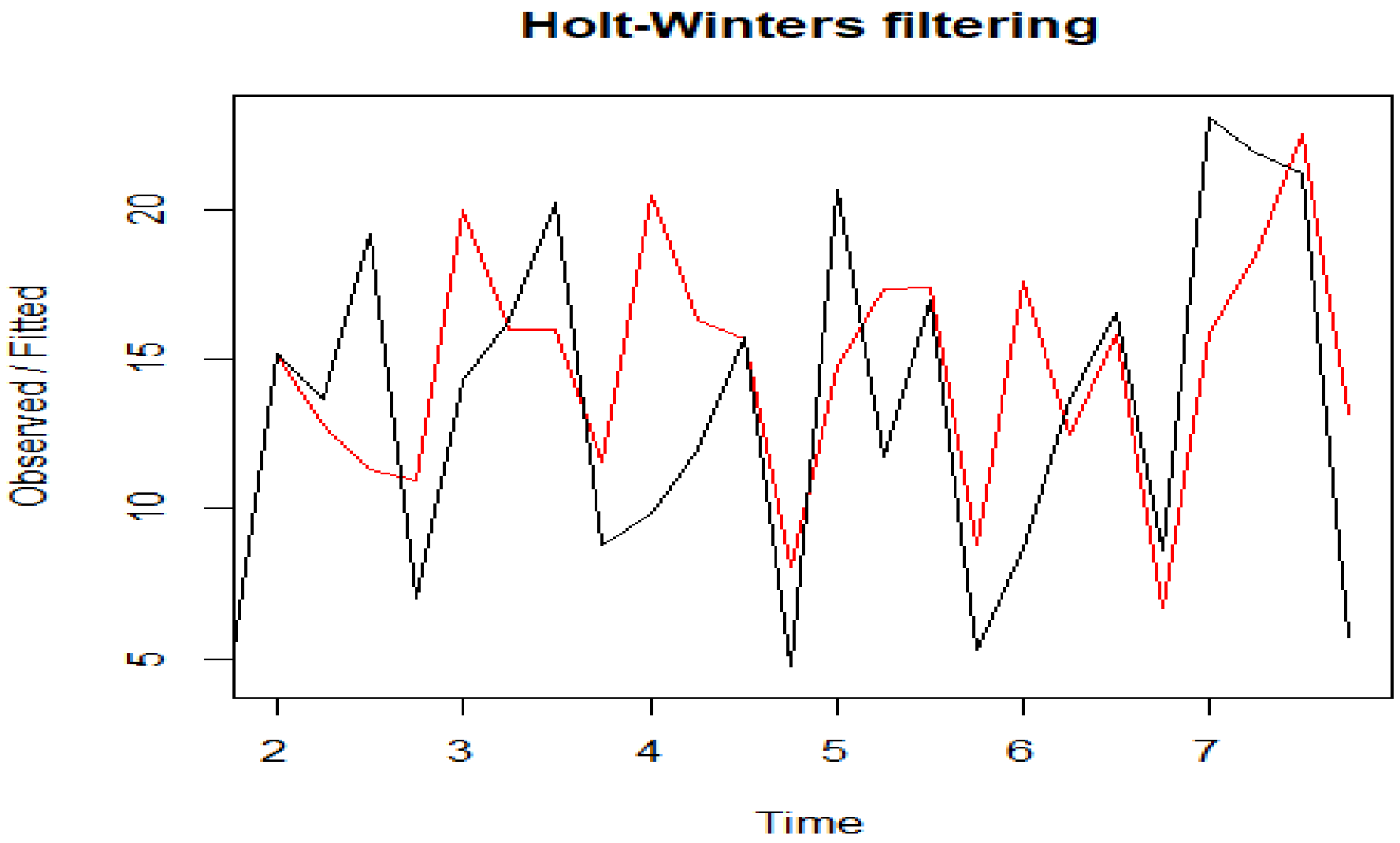
CHTC data processing



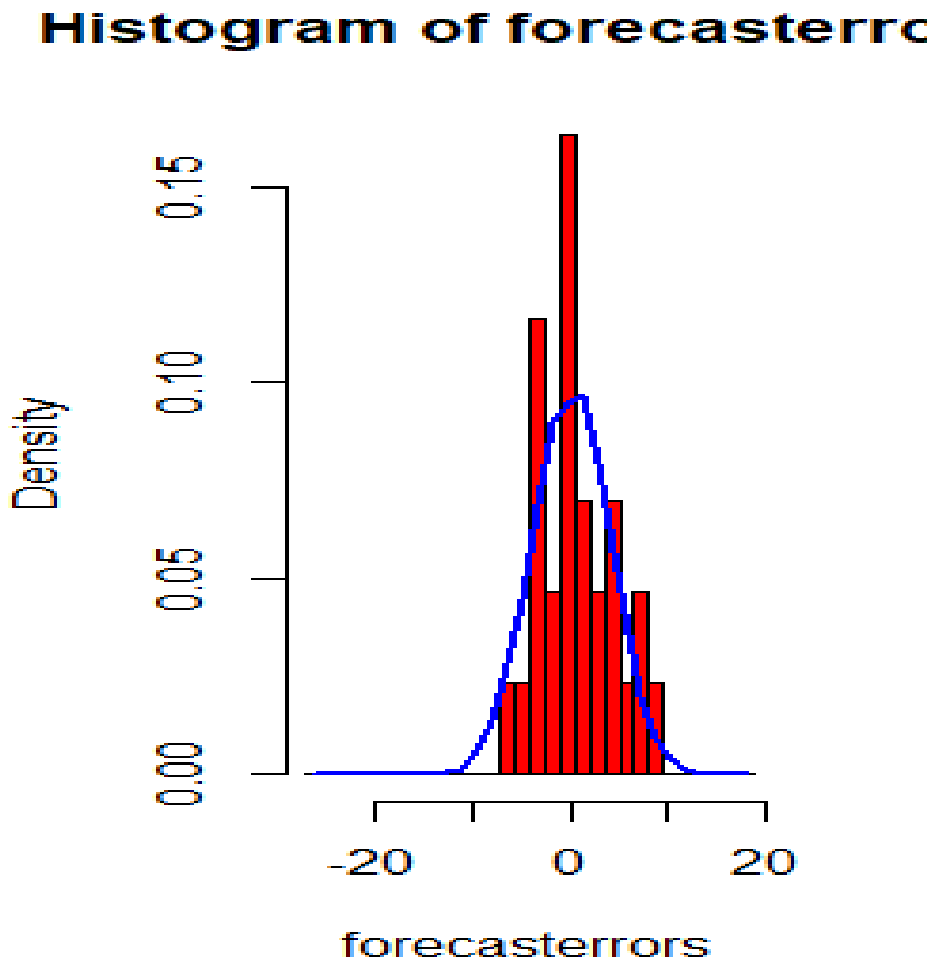
Problems in the processing!

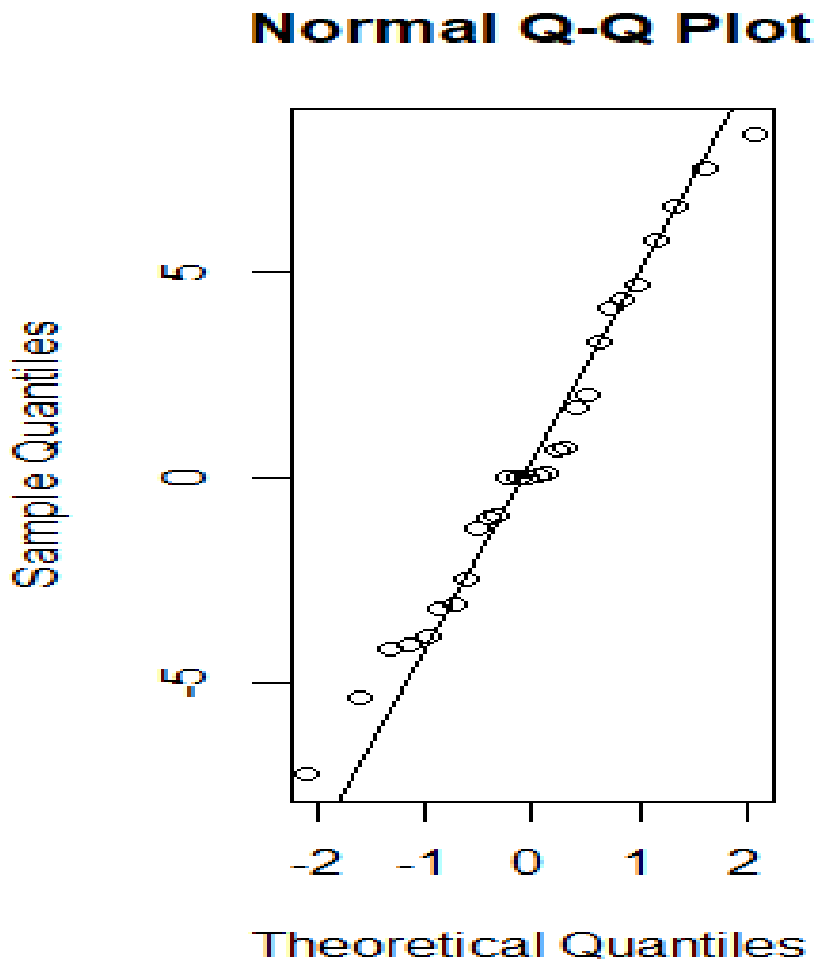
Result

Filtering


Holt-Winters filtering
Observed / Fitted
5
10
15
20
2
3
4
5
6
7
Time

Residual Plots





Filtering