# Forecast of River Flows To Evaluate Feasibility of Run-of-River Hydro Projects for SinoHydro

*Group B4*


| | |
|---|---|
| Aditya Pisupati | 61711022 |
| Tej Vohra | 61710736 |
| Naveen Badampudi | 61710591 |
| Bharadwaj Mundle | 61710928 |
| Anoop Thomas | 61710865 |
| Nishidh Biyani | 61710712 |

# Executive Summary

## Description of Business Problem

Run-of-river (ROR) hydro projects harness the energy in flowing water to generate electricity. There is total dependency on the flow rate measured in cusecs ($m^3s^{-1}$). As there is an absence of a dam and reservoir, the power generation is susceptible to the effects of daily and seasonal flow infrequencies. SinoHydro is a global Hydro Project construction company, looking for construction projects in the next quarter. It is looking to bid for six ROR hydro projects and wants us to forecast the river flows for next three years for the following rivers:

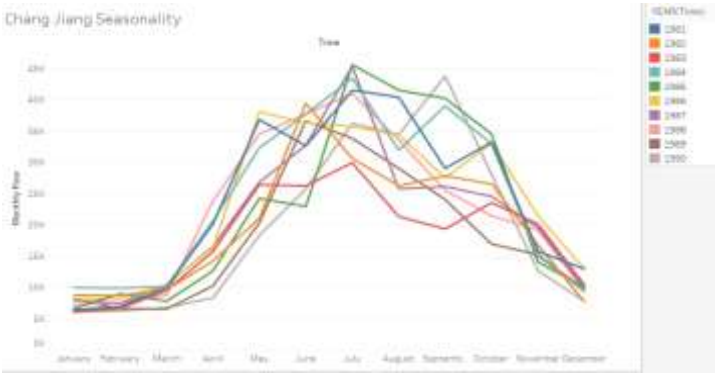
- **Madison River, West Yellowstone**
- **Chang Jiang, Han Kou**
- **Pigeon River, Ontario**
- **Funder River, Logager**
- **Piper's Hole, Newfoundland**
- **Mitta Mitta River, Tallandoon**

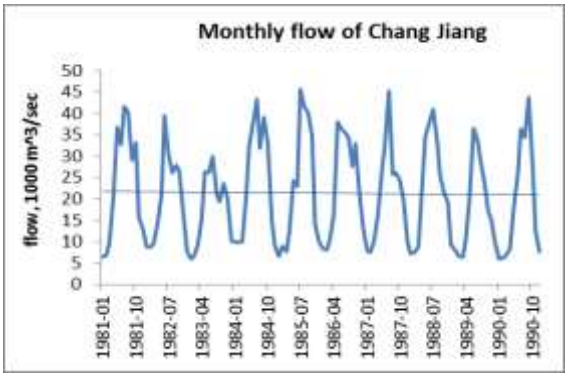
The data is available for the years 1981 – 1990 and consists of the mean monthly flow rate. Flow rate predictions for three years will also be used to estimate expected power generation and revenue. Over prediction in such a scenario is expensive and shall be penalised as we do not wish the company to take up projects that are unprofitable.

## Initial Analysis

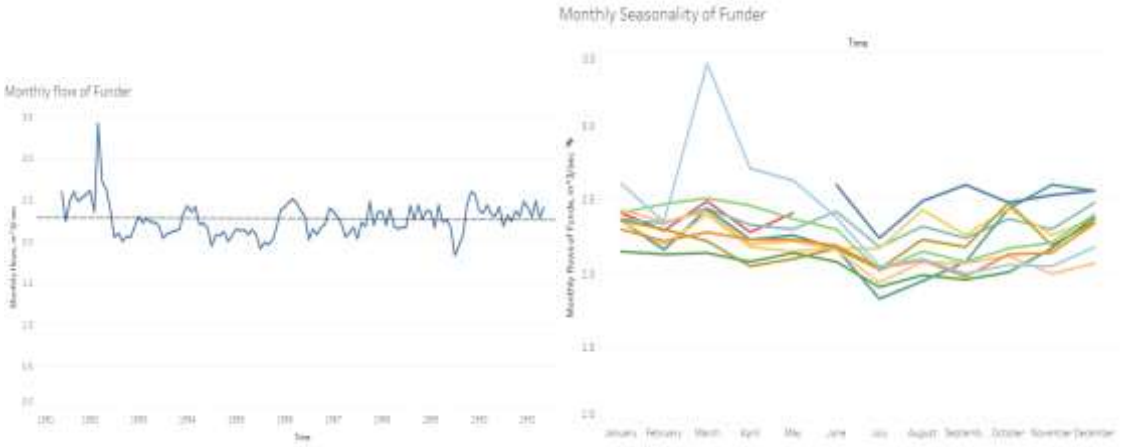
We plotted the monthly flows of each river against time to understand the trend and the seasonality (shown for two rivers below) which would help us in choosing an appropriate model for forecast. These would also give an insight into the external factors if any at play.

**Chang Jiang:** We observe no trend in the average monthly flows and a seasonality of 12 as shown in the plots below. The peaks are observed from the months of June to august which coincide with the rainy season.




**Funder:** We observe no trend in the average monthly flows and a seasonality of 12 can be observed even though the change is slight since the values are small as shown in plots below




### Method Selection

The viability of methods was evaluated based on the time series components available in the series. Following our initial analysis, we concluded that the data has no trend and has only level, noise and seasonality. For a metric like river flow this conclusion is very intuitive because barring external factors the flow rate is not expected to grow year on year. Hence the models highlighted in green have been finalised to try on the river flow data.

**Viability of methods for Forecasting River Flows**

| | |
|---|---|
| **Previous Month Naive** | High monthly seasonality; Poor results |
| **12 month Naive** | Good approach |
| **Moving Averages** | Constant prediction; does not work for seasonality |
| **Exponential Smoothing** | Constant prediction; does not work for seasonality |
| **Double Exponential** | Does not work for seasonality |
| **Holts Winters** | Good approach ; incorporates seasonality and trend |
| **Linear Regression** | Log of flow rate has been taken to take care of negative prediction; dummy variables capture seasonality |
| **Ensemble** | Good approach |
| **ARIMA** | Works only for short term forecast |

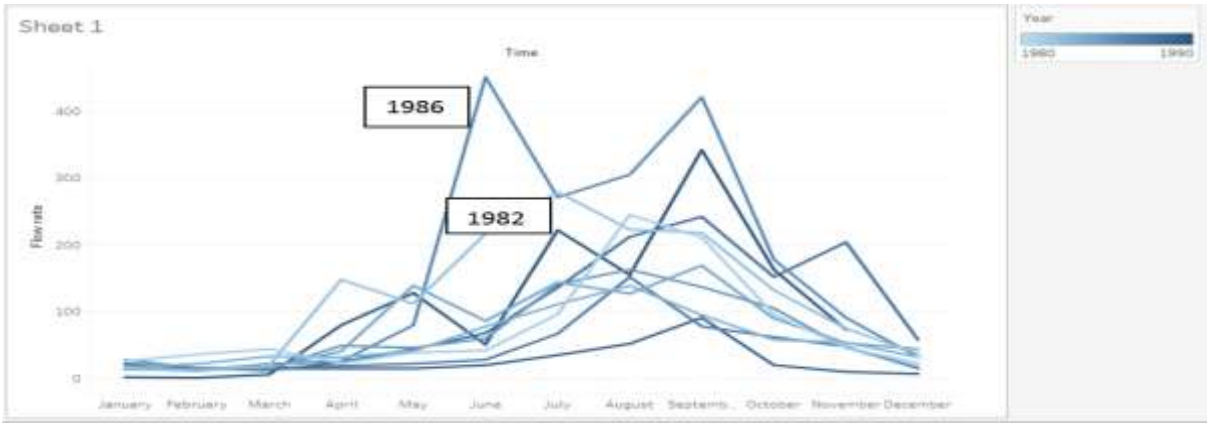
## Technical Summary

### *Data Partition*

For our analysis for all the 6 rivers, we used monthly data spanning from 1980 to 1990, 120 data points. To be able to predict the flow rate for the next 3 years, we split the data into validation, 84 data points and training, 36 data points.

### *Extreme values*

We noticed outliers in our data set for certain rivers, for example the Piper's Hole and the Mitta Mitta. We dealt with extreme values by replacing the outlier value with a value one season ahead based on seasonality of 12. The Seasonality graph of Mitta Mitta River, with outliers from the years 1982 and 1986 is shown below.

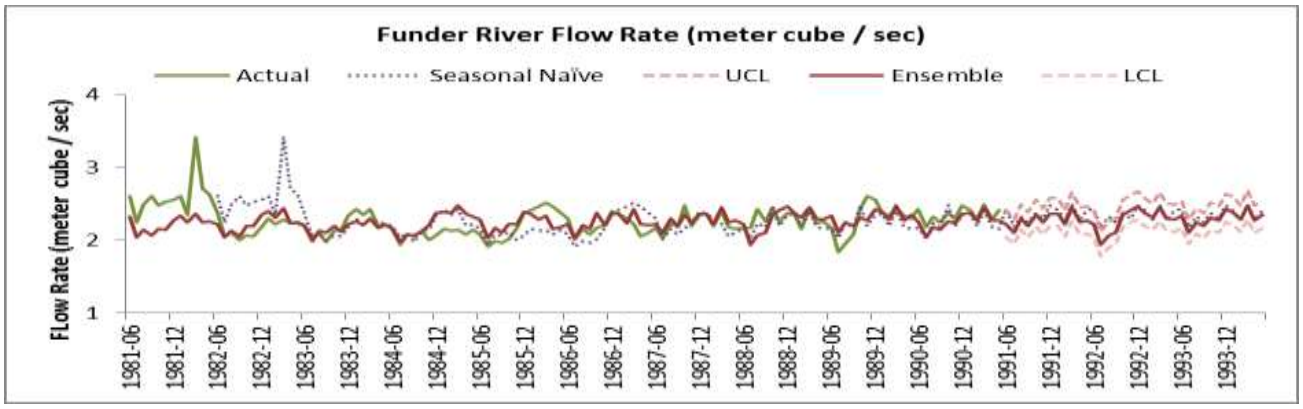


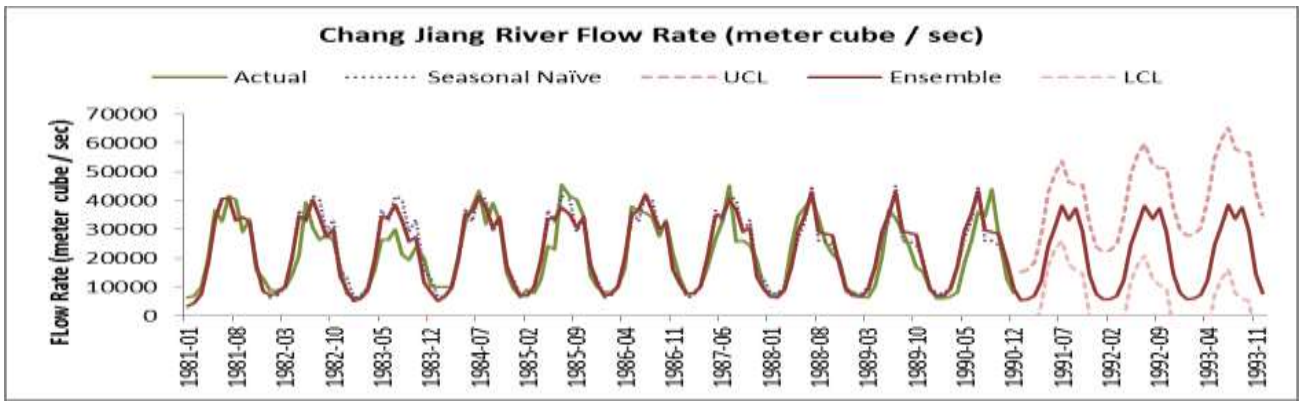

## Method Performance & Metrics

The following methods were tried on the six data sets and were evaluated visually and using the metrics of RMSE, MAPE and Average Error. It is important to notice that the scale of river flows of the rivers varies widely. With most rivers having flow rates in ’00s, Chang Jiang has a flow rate in ’00,000. Hence to avoid the scale effect in model performance, MAPE is preferred to RMSE. Moreover, Average error and residual plots give insight into over prediction and under prediction.

## Evaluation

The primary evaluation of the models was done visually and then MAPE and RMSE metrics were used to finalise the model. The following charts showcase the performance of seasonal Naïve and the best model for three of the rivers which showed excellent result



As we can see above, the ensemble shows the best forecast against the validation period. Forecasts also include prediction intervals to make decision making to eliminate projects.

**MAPE in % terms for the six rivers**

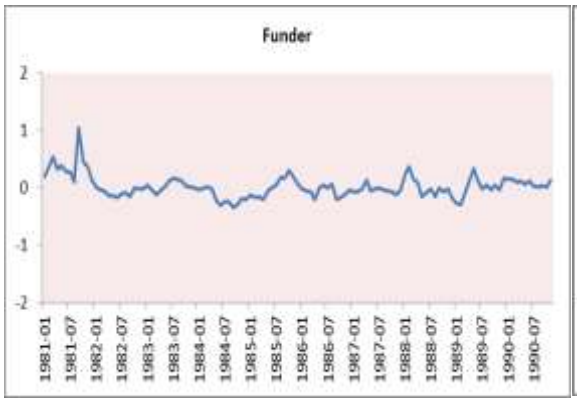
| | 12 month Naive | | Holts Winter | | Linear Regression | | Ensemble | |
|---|---|---|---|---|---|---|---|---|
| | Training | Validation | Training | Validation | Training | Validation | Training | Validation |
| Funder, Logger | 8.8 | 4.9 | 5 | 6.3 | 5.7 | 5 | NA | **4.6** |
| Madison, Yellowstone | 10.6 | 17.5 | 5.03 | 16.81 | 6.3 | **15.6** | NA | **15.7** |
| Mitta, Victoria | 31.3 | 126 | 32.4 | 130 | 27.5 | 81.2 | NA | 104 |
| Chang Jiang, Han Kou | 19.6 | 25.7 | 20.4 | 26.6 | 14.8 | 27.9 | 16.5 | **24.2** |
| Pigeon, Ontario | 76.7 | 106.8 | 42.5 | 46.7 | 58.5 | 84.7 | NA | 56.8 |
| Pipers Hole, Newfoundland | 35.2 | 43.2 | 65.3 | 60.8 | 21 | **26** | NA | NA |

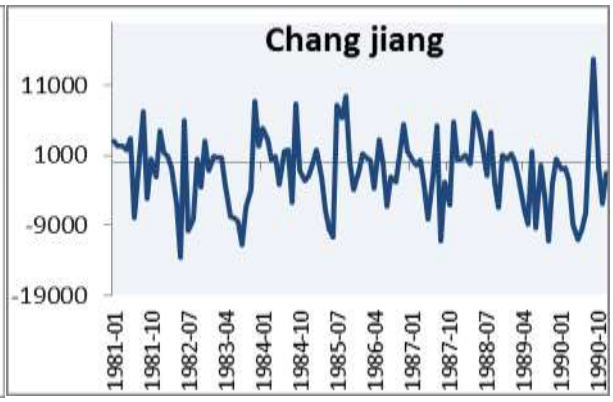
**RMSE for all the models**

| | 12 month Naive | | Holts Winter | | Linear Regression | | Ensemble | |
|---|---|---|---|---|---|---|---|---|
| | Training | Validation | Training | Validation | Training | Validation | Training | Validation |
| Funder, Logger | 0.27 | 0.14 | 0.16 | 0.18 | 0.18 | 0.15 | NA | **0.14** |
| Madison, Yellowstone | 3.12 | 3.41 | 1.47 | 3.3 | 1.77 | 3.14 | NA | 3.15 |
| Mitta, Victoria | 39.25 | 59.63 | 27.72 | 52.15 | 25.10 | 50.04 | NA | 72.76 |
| Chang Jiang, Han Kou | 6703 | 6030 | 5105 | 6185 | 4640 | 5745 | 5615 | 5721 |
| Pigeon, Ontario | 13.29 | 14.61 | 10.71 | 11.36 | 8.19 | 11.41 | NA | 10.48 |
| Pipers Hole, Newfoundland | 8.9 | 9.4 | 18.3 | 16.1 | 3.1 | 4.8 | 10.2 | 13.7 |

## Over prediction

Over prediction, a very important outcome to be prevented can be checked by looking at the residual plots and Average error table mentioned below.



**Average Error to understand the over-prediction**

| | 12 month Naive | | Holts Winter | | Linear Regression | | Ensemble | | |
|---|---|---|---|---|---|---|---|---|---|
| | Training | Validation | Training | Validation | Training | Validation | Training | Validation | |
| Funder, Logger | -0.05 | 0.07 | -0 | 0.03 | 0.01 | 0.07 | NA | 0.07 | Acceptable |
| Madison, Yellowstone | -0.1 | -2 | -0.01 | -1.7 | 4.22 | -1.29 | NA | -1.52 | Acceptable Over-Prediction |
| Mitta, Victoria | NA | 22.41 | 0.33 | 14.1 | 0 | 0.45 | NA | 26.3 | Under Prediction |
| Chang Jiang, Han Kou | -1463 | -1023 | -309 | -2362 | 539 | -2280 | -886 | -1702 | Over Prediction |
| Pigeon River, Ontario | 1.07 | -4.72 | -0.36 | 3.81 | 0 | -3.84 | NA | -0.02 | Acceptable |
| Pipers Hole, Newfoundland | 0.8 | 3.2 | -0.84 | 5.32 | 1 | 1.35 | NA | NA | Acceptable |

## Conclusion and Future Scope

- The river flow rate is not just a function of time, but is also affected by several external variables as time series components. The model can be improved by considering external factors such as the temperature, humidity and rainfall. However, the models shall be deployed for the perusal of the client, upon whose request the model may be rebuilt when future data becomes available.
- The models with good performance on MAPE and the one that have substantial under prediction are to be used for decision making.
- The Control limits and prediction intervals shall perform as better indicators than the point estimates.

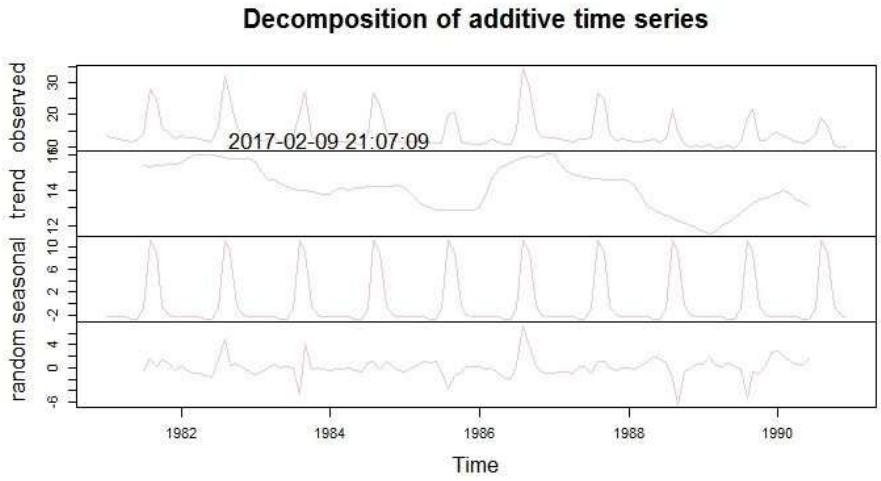
## Recommendation

- Based on our current analysis we recommend that Sino-Hydro bids for the project at Funder river and the Pipers' Hole river
- The following sites have forecasts in acceptable limits, but we leave it to the discretion of the clients to go forward with the projects:
    1. Maddison River
    2. Pigeon River
- Chang Jiang and Mitta Mitta do not have forecasts which seem promising and hence project bidding at these two river sites should be avoided with the current state of information

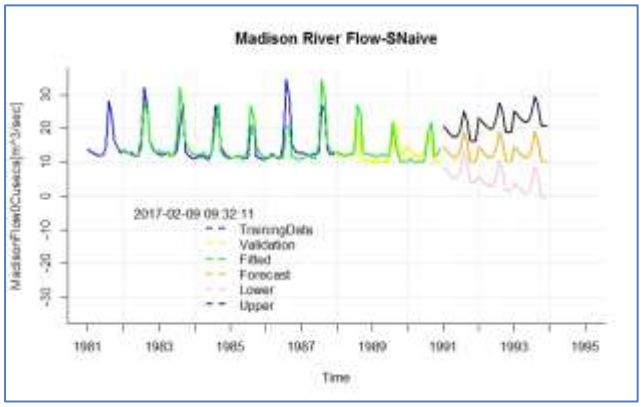
# Appendix

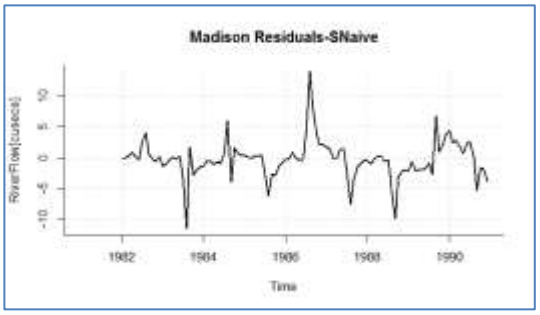
## Maddison River Plots

1. Components of the original time series

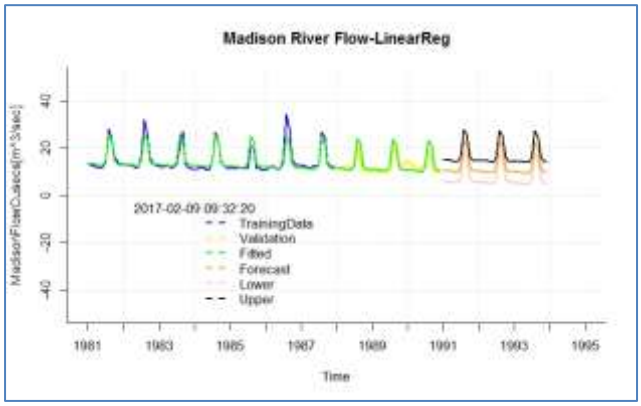


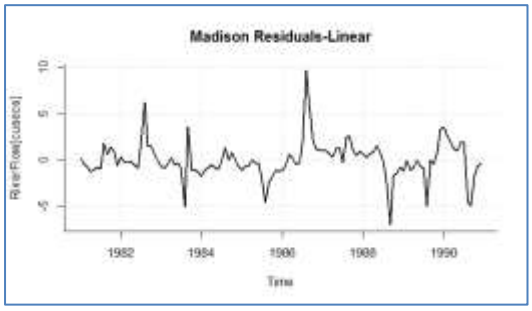

2. Naïve forecasts



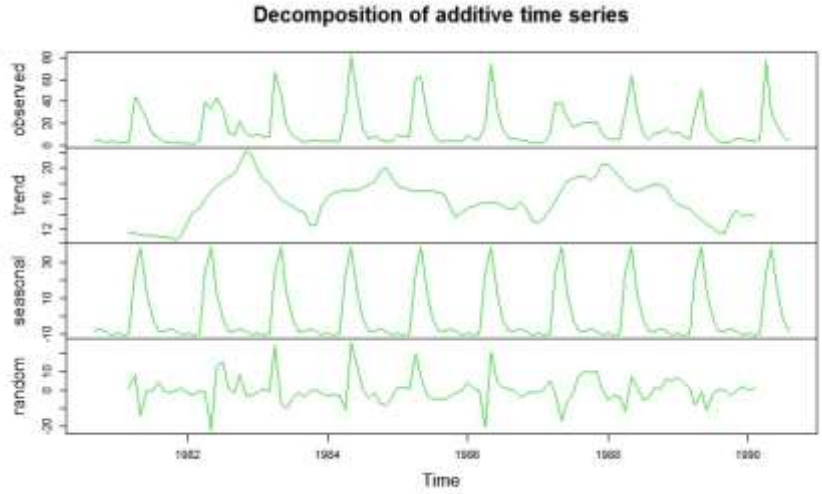

3. Best Model - Linear Regression (According to MAPE)



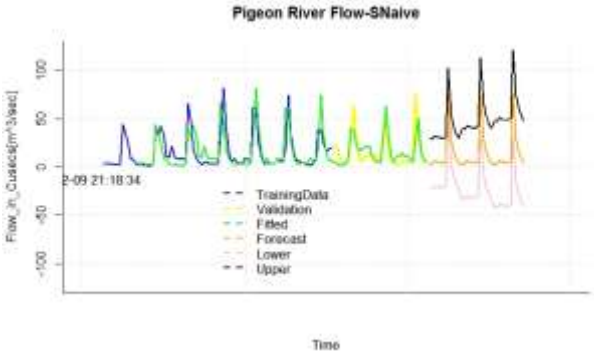



## Pigeon Plots

1. Components of the original time series

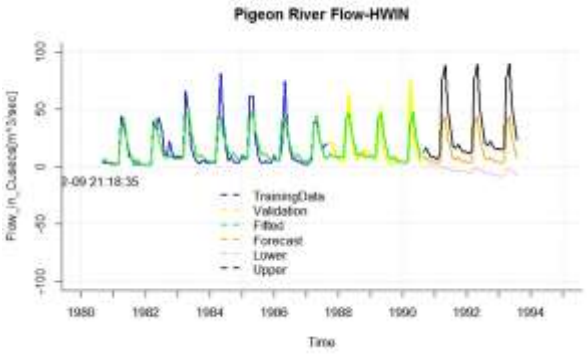


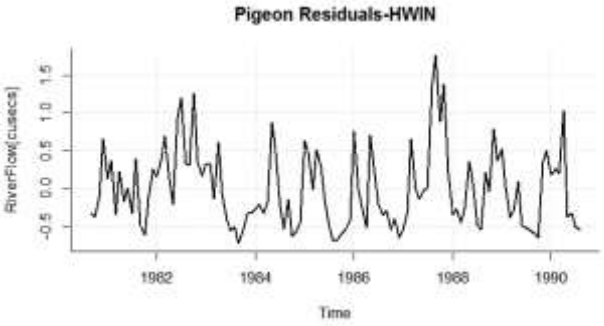

2. Naïve Forecast





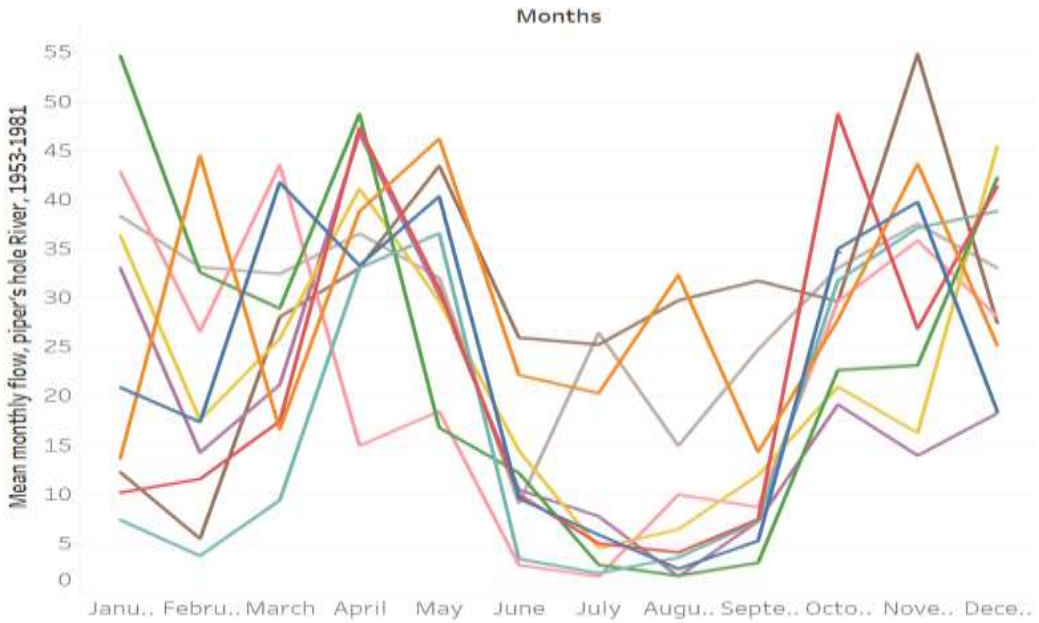
3. Best Model – Holt-Winter Forecast (According to MAPE)



## Pier's Hole Plots

1. Components of the original time series




2. Forecast