

## COMP 6730 Major Assignment Report: U6631257

1. I decided to store the dataset as a pandas dataframe object as it automatically converts all column data into series and elements to float or int depending on the values. I stored date as string object for future use. It was very suitable for me to use it in future rather than converting it to other data types. Code within function **check\_data\_validity(data)** ensures that each column does not contain missing values or any inconsistent data. If there is any inconsistent data, imputation is done using mean of that column or appropriate method and whole dataframe is returned.

Code within function **most\_average\_rainfall(data)**, the difference of every month's rainfall and average rainfall is calculated and if this is less than the minimal difference, min\_difference is updated and the process continues till last month. After this, the index of month with the minimum difference is sent to **index\_to\_name\_month(month\_index)** function which returns the name of the month. This name along with the year is returned as the month with most average rainfall.

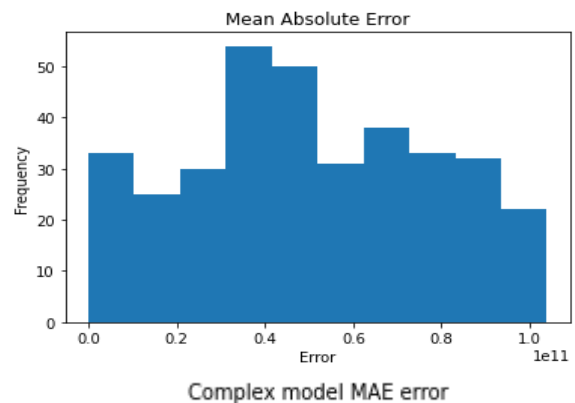
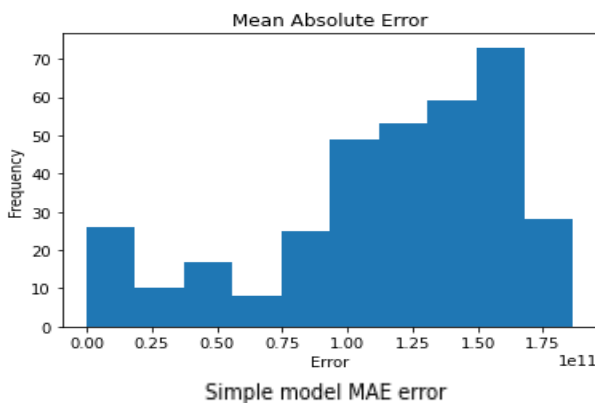
Code within **hottest\_month(data)** keeps track of 2 lists for sum of max\_temperatures of each month and number of times a value has been added to sum of these months each as a value in the list. Each month's max\_temperature is added to its respective element in the sum list. Temperature average of each month is calculated using this sum and count for each month resulting in average max\_temperature for all 12 months. Then a maximum of this is calculated and its index is used to return the hottest month name. This works for any order of months of a year and also if no months is given.

2. As far as my models are concerned, Complex model is best compared to Simple model. The mean absolute error generated by models in predicting volumes to actual volumes are:

Simple model : 114630065672.65038 litres

Complex model : 50499523338.908485 litres

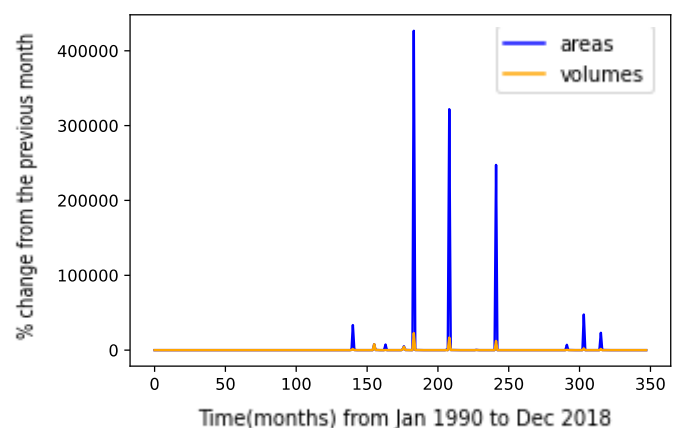
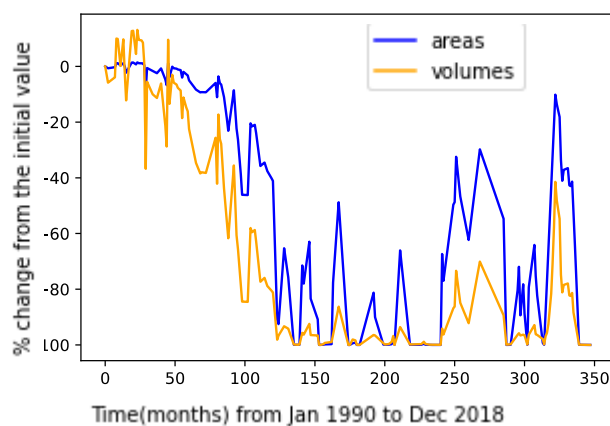
Complex model is better than simple model as its error is only 44% of simple model's error indicating it's error is almost 0.55 times lower than error generated by simple model. As MAE is negatively oriented, the model with lower value(complex) is preferred.



3. For solving Question 4, I assumed the maximum of area as catchment area, volume[0], area[0] as initial volume, surface area respectively. Evaporation rate is constant for simple model. I calculated total volume at  $i^{\text{th}}$  month is equal to volume in previous month + volume induced through rainfall – volume evaporated. Volume induced through rainfall is rainfall in that month times catchment area whereas evaporated volume is surface area of previous month times evaporation rate. Evaporation rate changes every month in complex model depending on all given factors as mentioned in problem. They are somewhat realistic but not fully. Catchment area would also include surrounding areas of the lake in realistic sense. The model would be improved if surface area of this month is used instead of surface area of previous month but it is not possible in this model. I would use the average of this month's and previous month's surface area to calculate this month's volume for a more improved model and also increase the catchment area if possible. I would also prefer the changing evaporation rate to the constant one.

#### 4. Plot for area vs volume :

There are 2 plots. The first one describes % change in volume and areas of lake compared to initial value whereas the other describes % change in volume and areas of lake compared to its



immediate previous values over the months from 1990 to 2018. Areas , volumes lists are calculated as % change as required for both plots and graphs are plotted as above. These images are vector images, so you can zoom in to have a better view at the plots. In first plot, the points where areas show increase compared to volumes is point where area increases more for less increase in volume. There are points where volume increases more for less increase in area at the start of the graph.

#### References:

Medium. 2018. *Mean Absolute Error (MAE) ~ Sample Calculation*. [online] Available at: <https://medium.com/@ewuramaminka/mean-absolute-error-mae-sample-calculation-6eed6743838a> [Accessed 28 May 2020].