

Project Assignment

Code explanation

Function name: `most_average_rainfall(data)`

This part of code is to find the month and year had a rainfall closest to average. I start with find the mean value of the rainfall using `numpy.mean` method. Considering there could be multiple dates have rainfall close to the mean value. The values with the minimal difference with the mean value are identified first and then located the dates of these values.

Function name: `hottest_month(data)`

This part of code used to find the hottest month throughout the data. The hottest month is the month with the highest mean temperature. So, for each month I calculated the mean `max_temperature` across the sample period, and find the month has the highest mean `max_temperature`. For easily tracing back which month has the highest mean `max_temperature`. I created a list contains 12 months and locate each month data to calculate the mean `max_temperature` value, and then collect these values into a list corresponding to the index of the month in month list.

Statistics about Lake George

The largest area covered by the lake	144269375.0 Km ²
The average volume of the lake	56640293251.61218 Litres
Month and year had a rainfall closest to average	2000, 08 and 2005, 08
The hottest month/mean temperature	01 / 26.726551724137927

Evaluation of simple model and complex model

	MAE	RMSE
Simple model	68257521489.03296	83949051674.77284
Complex model	14048808998.391115	20370353412.01747

Both the MAE and RMSE measure the magnitude of the difference between predicted volumes and the actual volume. The MAE and RMSE are closer to zero means a smaller difference between predicted volumes and the actual volume. The result shows the complex model is better than the simple model as there is smaller difference between modelled volume and actual volume.

Assumptions of models

At the Part (a) simple model, the first assumption is the evaporation is constant each month. The second assumption is the catchment area and the surface area are both equal to the maximum surface area. But the assumptions are not realistic, according to the Wikipedia and the Geological sites of NSW, the catchment area of Lake George is small, has a catchment area of about 932km². This partially explained when assumed the catchment area is the maximum area, the modelled volume is significant higher than the actual volume (Fig.3). Based on the given data, I modified model by assumed the catchment area and surface area are both equal to the mean value of area, but considering the volume was really high sometimes at compare with the other times was completely dry. The mean replace catchment area and surface area is not a reasonable way to improved model (Fig.5).

Appendix

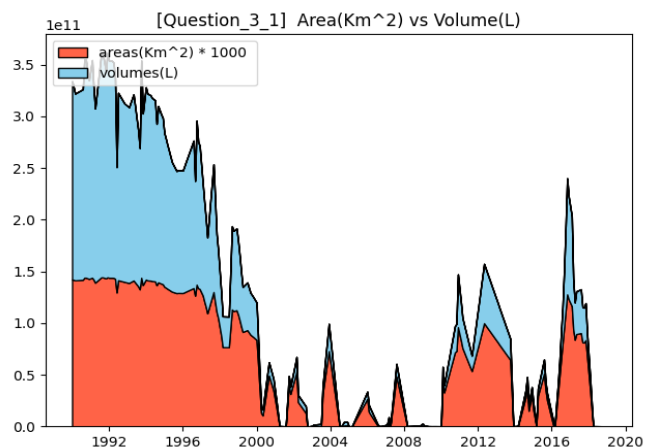


Fig.1 Changes in area and water volume

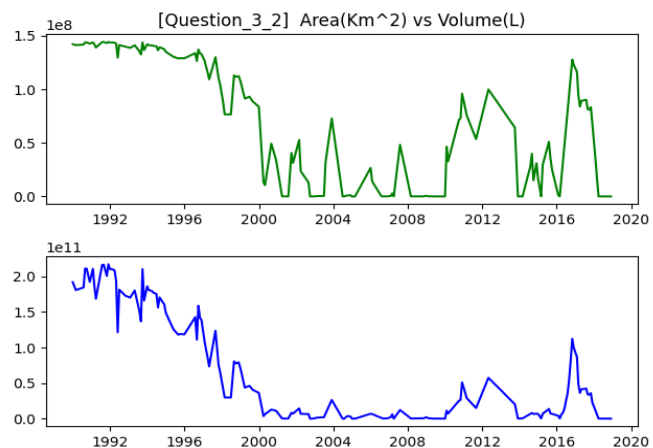


Fig.2 Changes in area and water volume

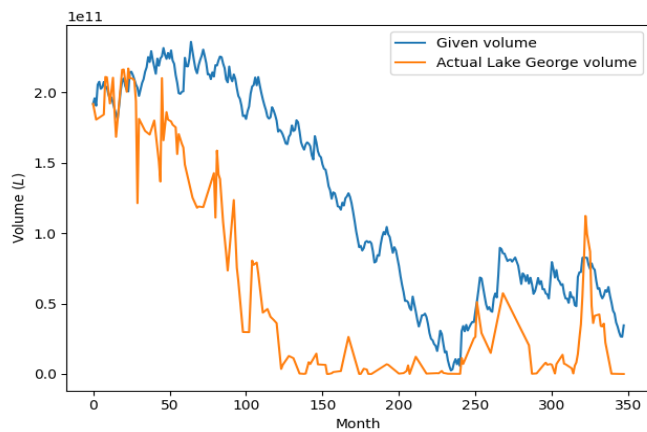


Fig.3 Simple model

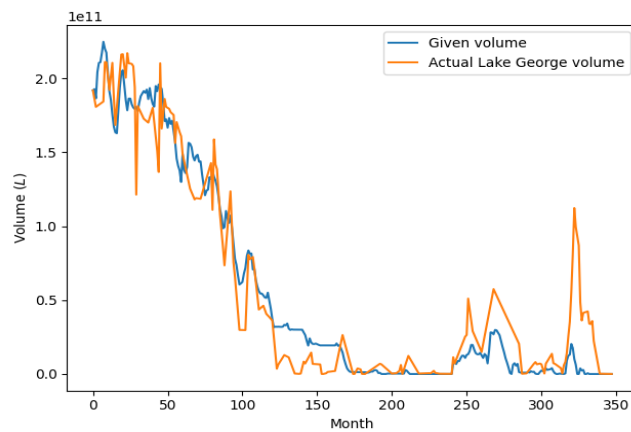


Fig.4 Complex model

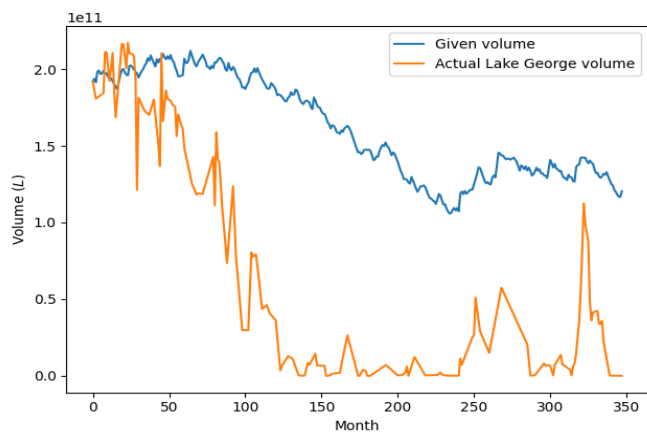


Fig.5 Simple model (catchment and surface area are equal to mean area)