**CEG8501 Coursework**

**Introduction**

Congratulations on your new position as field hydrologist for Hobbiton Environmental Monitoring (HEM) Ltd. There has been a recent incident at the Buckleberry ferry port on the River Brandywine and the ferry port has been destroyed. The East Farthing Council have decided that rather than replacing the ferry port, they would like to investigate the option of replacing the ferry service with a bridge. We at HEM have been commissioned to evaluate the potential for construction at this location.

You should prepare a written report of no more than 20 pages, including figures and tables. Title and contents pages are not included in the 20 page limit – everything else to be considered for assessment is to be included in the 20 page limit. You may submit a neatly presented Jupyter notebook instead if you prefer. Present the report in sections as organised in this document. It is suggested that you carry out the statistical analysis in Python, but you are free to use any suitable package (Excel, Minitab, SPSS, R *etc*.). Preliminary data checking or manipulation is desirable.

*Assessment Criteria*

* Correct execution of analyses and appropriate choice of methods, distributions etc.; (40%)
* Demonstration of understanding and limitations of methods, justification of choices, and interpretation of results; (50%)
* Clear, legible presentation of results and interpretation (10%)

**Deadline: 4.00pm Friday 11th November**

Warning: before carrying out the statistical analysis in this assignment you should carry out “exploratory data analysis” and quality checking to see if there are:

* any missing data
* any erroneous data (e.g. negative, or large values which are clearly incorrect)
* any samples which are too small or not matching in size if they are to be compared against other samples

**Task 1**

You need to install a meteorological monitoring site in the catchment. There is space for a 18.6 m2 pen, and we need at least one AWS, TBR and borehole each onsite. Fences should be 1.3m high to prevent cows from entering. We have a budget of £3000. Cost up the materials required for the monitoring site for approval by East Farthing Council.

Fencing = 0.2p per cm2

AWS = £572.30

TBR = £321.00

Borehole = £630.00

**Task 2**

After visiting the proposed site on a windy day, you become concerned that a nearby apple tree will damage your equipment if apples are blown into the area. You’re trying to impress your new boss and so you do a quick back of the envelope calculation to check if the apples will blow onsite. You look up some typical wind speeds in the area. This is what you sketch out:

* Apple start height = 3.2m
* Turbulence around the tree creates an initial velocity of 24m/s at an angle of 22 degrees above horizontal.
* Proposed site is 19m away.

Will the apple fall onsite?

**Task 3**

After installing a level logger on the River Brandywine, your next bit of field work is to get some data to create your stage-discharge relationship. You decide to calculate the current river flow by conducting some salt gauging. You have collected the data and recorded it in file salt\_gauging.csv. What is the flow?

*Info:*

* *Timestep=2s*
* *Calibration factor = 3.91x10-3 cm/µS*
* *5kg salt used in 35l water*
* *Background conductivity = 0.041 µS/cm*

*Procedure:*

1. *Calculate the electrical conductivity above background conductivity for each timestep by subtracting the background conductivity*
2. *Find the point in the record where the conductivity begins to increase above the background level*
3. *Calculate the area under the curve from this point, noting the timestep of 2 seconds*
4. *Multiply the area under the curve by the calibration factor to convert toseconds.*
5. *Calculate the volume of additional water added to the river (tracer volume i.e. the volume of water that was mixed with salt plus 0.32 multiplied by number of kg salt added (1kg salt occupies 0.32 L when dissolved in water))*
6. *Calculate the flow (l/s) (by dividing the volume of additional water by the calibrated area under the curve)*
7. *Calculate the tracer volume flow (l/s) (the tracer volume divided by the duration of the test)*
8. *Correct the calculated river flow for the tracer flow*

**Task 4**

You have installed 3 boreholes and recorded the following water table level readings:

|  |  |  |  |
| --- | --- | --- | --- |
| x | y | Ground Elevation (masl) | Water table (mbgl) |
| 380270 | 499200 | 345 | 11 |
| 380330 | 499400 | 256 | 14 |
| 380450 | 498900 | 223 | 15 |

You are planning on installing another borehole at 380290, 499100, (ground elevation 340 masl). What would you expect the water table level to be approximately?

**After a few months working on the Buckleberry ferry project, you are promoted to the position of data scientist. In this new role you are asked to analyse existing rainfall, flow and temperature records for the area to understand extremes, dependencies and trends.**

**Task 5**

**Data**

* Daily rainfall records for the Brandywine catchment are given (1961-2015) in **daily\_rainfall\_brandywine.csv**
* Annual rainfall records for the Brandywine and neighbouring Greyflood catchment in **annual\_rainfall\_brandywine\_greyflood.csv**

**Task (5a)** Consider a number of probability distributions to use to represent the observed data, bearing in mind the shape (i.e. skewness, tails) of the observed (empirical) distribution in each case. Candidate distributions to try are: normal, 3-parameter lognormal, gamma, exponential.

Assess the fit graphically, and choose the most suitable in each case. Provide a plot of the frequency distribution (pdf) as well as the parameters of the distribution.

**Task (5b)** Using the most suitable distribution in each case, plot the cumulative distribution function (cdf) and then use the cdf to estimate the 10th, 50th, 90th and 99th percentile values of rainfall.

**Task (5c)** For the annual rainfall series for Brandywine and Greyflood (separately), present the overall mean with a 95% confidence interval using a suitable method.

**Task (5d)** For the two annual rainfall series at Brandywine and Greyflood, carry out a test to estimate the difference in the means and assess whether they are significantly different. For the Brandywine, split the record into two equal halves and test these 2 samples to see if they are significantly different (i.e. if there has been a change in the mean rainfall at this location over the period of the observations).

**Task 6**

**Data**

* Standard annual average rainfall (SAAR) estimates for 24 gauges: in **SAAR\_elevation.csv** together with location (easting and northing);
* Maps of rain gauge network and elevation (in appendix)

**Task 6a** Use the data given in sheet *SAAR* to derive a simple linear regression relation between rainfall and elevation. Present a scatterplot of the regression relation, and calculate the standard error. Comment on the accuracy and reliability of the relation, with reference to the requirements for linear regression to be valid.

**Task 6b** Use the data given in sheet *SAAR* to derive a multiple regression relation between rainfall and [*elevation, easting, northing*]. Present the regression relation, and the standard error. Comment on any improvement on the simple regression above, taking into account the *p-values* of the coefficients to decide if the extra terms are meaningful (i.e. p <0.05). Referring to maps of rainfall and topography in the area, do the signs of the coefficients for easting and northing have any physical meaning?

**Task 6c** For a new ungauged location at (E,N) (380000,500000) and elevation of 400 m, use your best regression relation from 2b to estimate the SAAR. **Task 7**

**Data**

* Daily rainfall records for Brandywine (1961-2015): in **daily\_rainfall\_brandywine.csv**
* Daily flow for the Brandywine at Buckleberry in **Flows.csv**

**Task 7a** Resample the daily rainfall data to monthly and annual time series. Use the daily, monthly and annual rainfall data to generate a correlogram for suitable lags. Comment on the results and explain the differences as the duration increases from daily to annual. Now do the same for the monthly flow data and comment on the results.

**Task 7b**

Resample the daily flow data to a monthly timeseries. Now de-seasonalise the flow data. Use  where  is observed flow for month t, is the mean for calendar month and is the standard deviation for calendar month . Some formulae to help with this are given in the spreadsheet.

Generate the correlogram for these new data and compare with the results using the ‘raw’ observed data: explain the difference.

**Appendix (for task 6)**





