Assignment 2: Watersheds and Areal Estimates of Precipitation

Due: Wednesday, 6 February 2019

Background - Catchment water balance

The water balance of a catchment can be expressed as:

$$\Delta S = P - E - Q \tag{1}$$

where ΔS is the change in water storage within the catchment over a specified time interval, P is the precipitation input, E is the evapotranspiration loss and Q is the streamflow exiting the catchment. The validity of Eq. [1] requires that there are no net exchanges of groundwater into or out of the catchment, and no artificial imports or exports of water (e.g., for domestic water supply).

Over a period of one or more years, changes in catchment storage (soil moisture, groundwater) will often be negligible compared to the amounts of streamflow, precipitation and evapotranspiration. This assumption may not be valid in glacierized catchments, where glacial recession may produce a significant decrease in catchment storage. If ΔS can be neglected, then the water balance can be expressed as

$$Q = P - E. \tag{2}$$

Assignment Part 1: Quality Control of a Given Gauge (10 marks)

Measurements obtained from rain gauges are prone to error. Uncertainty may arise from: a) gauge malfunction; b) shielding by overhanging vegetation (causing under- or over catch); c) gauge relocation; or d) wind effects. It is especially important to estimate the reliability of a given rain gauge if it is to be used for climate-change detection and analysis studies. This is because changes due to climate (e.g. changes in the mean annual precipitation) may be close to those errors introduced by a-d.

The dataset 'ltppt.txt' consists of annual precipitation series for four stations (Fort St. John, Prince George, Quesnel, Williams Lake) in BC for the period 1936-2001. Your task is to assess the reliability of the Williams Lake record. In addition to the possibility of an artificial trend introduced by the station being relocated, there are missing values in the data. Replace the missing values by developing a linear relation (i.e. $y = mx + b$) between the Williams Lake record and another station.

Once the missing values have been replaced, evaluate the long-term behavior of measurements from the gauge using the double mass method (see Chapter 4 in Dingman 2002).

The Instructions

- Download the 'ltppt.txt' file from the course website and put into a directory called A2. **Change the start location of your R shortcut to reflect the new working directory.** Once you have R started we will open the text file and assign the contents to a dataframe called 'ltppt':

```
> ltppt=read.table('ltppt.txt', header=T)
```

- Examine the degree of correlation between all stations using the cor() function:

```
> cor(ltppt, use='pairwise.complete.obs')
```

  Because your file contains missing values, the extra argument tells R to just use the cases where all data exist. You could also type **help(cor)** to see how this function works in general.[1]

- Once you have figured out the station most highly correlated with the Williams Lake record, produce a linear model that will be used to estimate the missing values of the Williams Lake record. To produce a linear model in R, use the 'lm' function

```
> pptmod = lm(y~x)
```

  where y (dependent variable) is predicted by x (independent). Once you have made your linear model object use the summary function to obtain the slope and intercept.

- Replace missing value (NA's) with estimated value based on the linear regression model. First, type:

```
> ltppt
```

  and note the index values for the years with missing precipitation data at Williams Lake. Then to predict any value for Williams Lake with your precipitation model (linear regression) and the Quesnel data (for example), you could now use the following:

[1]The Pearson correlation coefficient is a standardized method of expressing the covariation between two variables x and y. It is expressed mathematically as:

$$r = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{(N-1)\cdot\sigma_x\cdot\sigma_y} \tag{3}$$

where σ_x is the standard deviation of variable x. Values for r vary between -1 and 1. Squaring r gives the proportion of explained variance expressed by the relation (i.e., the coefficient of determination, r^2).

```
> ltppt[i,5]=ltppt[i,4]*m+b
```

where i denotes the index value for the missing data, and m and b are the slope and intercept values from the linear regression model. Repeat this process for all missing values, adjusting the index values as needed. Then type again:

```
> ltppt
```

to ensure all missing precipitation values at the Williams Lake station are now replaced by values from the linear regression model.

- Construct a double mass curve comparing the average yearly precipitation values of Fort St. John (FSJ), Prince George (PG), and Quesnel (Q) against those of Williams Lake (WL). First, calculate average yearly precipitation from FSJ, PG, and Q, use the `apply()` function:

```
> regional.av = apply(ltppt[,2:4], 1, mean)  ##calculates row-wise means
```

Use the `cumsum()` function to construct the double mass curve plot:

```
> plot(cumsum(ltppt[,5]), cumsum(regional.av))
```

- Determine the year when the Williams Lake record started to deviate from the others (break in slope). To see the position of the individual points on your scatterplot, use

```
> plot.ts (cumsum(ltppt$williams),cumsum(regional.av) )
```

- Estimate the correction factor (K) that is required to adjust the Williams Lake record[2] by building two new linear models for the data before and after the slope change. To simplify this process, consider defining the following variables:

```
> regavcm = cumsum(regional.av)
> wlcm = cumsum(ltppt[,5])
```

Then one can build the linear regression models before and after the breakpoint using:

[2] The correction factor K is

$$K = \frac{m_1}{m_2}$$

where m_1, m_2 are the slopes of the lines before and after the change respectively.

```
> early=lm(regavcm[i1:i2]~wlcm[i1:i2])
> late=lm(regavcm[i3:i4]~wlcm[i3:i4])
```

where the index values for the early period range from i1 to i2 and for the late period, from i3 to i4. To obtain the details of the linear regression models, then type:

```
> summary(early)
> summary(late)
```

The Writeup

1. Provide a table of the correlations between stations (two decimal places only please). (2 marks)
2. Which station has the highest correlation? (2 marks)
3. What is the regression equation for your linear model of annual precipitation at Williams Lake? Is the model significant? What is the R^2 value? (2 marks)
4. Provide a properly annotated double-mass curve. How does the Williams Lake record compare to the regional average? What year does the Williams Lake record begin to deviate from the regional average? (2 marks)
5. What are the two linear models before and after the change (provide 2 regression equations and R^2 values), and what is the value of K? (2 marks)

Assignment Part 2: Estimating Missing Gauge Measurements (15 marks)

One of the challenges in the application of the catchment water balance method is estimating the spatially averaged precipitation depth over the catchment area. Several approaches are available, each having advantages and disadvantages.

In this exercise, you will examine the spatial pattern of precipitation in a steep, rugged catchment, particularly in relation to its variation with elevation. You will then apply the catchment water balance approach to estimate the annual water yield, using several methods for spatially averaging the precipitation measurements.

Listed in Table 1 are eight tipping bucket rain gauges, their coordinates (UTM), mean annual precipitation (MAP) and storm precipitation measurements (mm). These data can be found in the file 'gauges.txt' on the course website for Assignment 2. You need to establish basin-average precipitation for the Cheakamus Lake Catchment (southwestern British Columbia) for this rainstorm that occurred on 31 August 1991.

Gauge	Easting (m)	Northing (m)	Elevation (m)	Mean Annual P (mm a^{-1})	31 August 1991 (mm)
A	502497	5533664	1830	1550	60
B	504636	5540321	890	1200	45
C	509369	5531979	1360	1320	51
D	512145	5537627	2000	1745	80
E	514418	5534120	1900	1600	NA
F	516970	5539288	1500	1400	55
G	520792	5539583	1350	1200	48
H	519791	5534932	2400	2200	112

Table 1: Rain Gauge Network for Cheakamus Lake

Instructions

1. Download the 'gauges.txt' file from the course website and put into a directory called A2. **If you have not done so already, change the start location of your R shortcut to reflect the new working directory.** Once you have R started we will open the text file and assign the contents to a dataframe called 'gauges':

   ```
   > gauges=read.table('gauges.txt', header=T)
   ```

   and then verify that the dataframe has loaded the data properly.

2. Next, you need to estimate a missing value for rain gauge (E) on 31 August 1991 using the following five approaches:

   - Simple Average
   - Normal Ratio method: [3]

   ```
   > ratios  = gauges$MAP[5]/gauges$MAP #normal ratios
   > WP = ratios*gauges$Aug31 #weighted storm precip totals
   > NR.hat = sum(WP,na.rm=T)/7
   ```

   - Inverse Distance (use $w_i \sim \frac{1}{d_g}$ where w_i is the ith weight, d_g is the distance (km) between gauges, and the scaling exponent $b = 1$).[4] To calculate the distance between gauges, use the `dist()` command:

   ```
   > xy = data.frame(gauges$Easting, gauges$Northing)
   > d = dist(xy,upper=TRUE)
   > d.stnE = c(d[4], d[10],d[15], d[19], NA, d[23],d[24],d[25])
   ```

   Then to calculate the weights d_g^{-1}, type:

   ```
   > invd = 1/d.stnE
   > D = sum(invd,na.rm=T)
   > WP = invd*gauges$Aug31
   > ID.hat = sum(WP,na.rm=T)/D
   ```

   - Inverse Distance Squared (here use $w_i \sim \frac{1}{d_g^2}$ where w_i is the ith weight, d_g is the distance (km) between gauges, and the scaling exponent $b = 2$).
   - A linear model of storm precipitation (P) versus elevation (Z) to estimate the storm total at the missing station

3. Use the mean of your estimates from 1-5 to estimate the precipitation total for gauge E for the rainstorm on 31 August 1991.

[3] The normal ratio uses the equation:

$$\hat{p_o} = \frac{1}{G} \cdot \sum_{g=1}^{G} \frac{P_o}{P_g} \cdot p_g \tag{4}$$

where $\hat{p_o}$ is the estimated value, using G gauges, P_o, P_g are the annual averages from the missing and non-missing gauges respectively, and p_g is the storm total from gauge (g).

[4] Inverse distance estimates $\hat{p_o}$ from:

$$\hat{p_o} = \frac{1}{D} \cdot \sum_{g=1}^{G} d_g^{-b} \cdot p_g \tag{5}$$

where d_g is the distance between missing and available (g) gauge, and b is a scaling exponent (i.e. higher b, quicker drop off with distance). D is the sum of the weighting coefficients d_g^{-b}.

The Writeup

1. Provide a table showing the weights assigned to each station for the ratio method and the two inverse distance weighting methods. (3 marks)
2. Provide a summary table with the five estimates of the missing storm P value, and their mean. (6 marks)
3. Prepare an annotated plot showing the relation between elevation and storm precipitation, and plot the linear model using the `abline(lmod)` command. (3 marks)
4. Comment on the relation between Z and P using the output from the model summary. (3 marks)

Assignment Part 3: Areal Estimation of Precipitation (15 marks)

Now that you have estimated the missing gauge data at station E, derive estimates of total precipitation for the 216 km^2 Cheakamus Lake watershed for this storm (31 August 1991) using:

- Simple Arithmetic Mean
- Distance Weighted Mean to centroid of the watershed
- Isohyetal Method

Mapping the watershed

- To map the basin and station locations in R, type the following commands from your R prompt (remember, not the commented parts):

```
## Load the shapefiles library - this can be downloaded from:
## https://cran.r-project.org/web/packages/shapefiles/index.html
## if not already installed on your computer.
## Then under the "Packages" tab, click on "Install package(s) from local zip files".
> library(shapefiles)
## Read in shapefile for the Cheakamus Lake watershed
> chk=read.shapefile('cheak_watershed')
## Extract just the x,y coordinates from the shapefile and overwrite shape object
> chk=chk$shp$shp[[1]]$points
## Plot the points to define the watershed boundary
## type = 'l' will plot the data as a series of lines and 'col' = color of line or
## point
> plot(chk, type='l', col='red', xlab='Easting (m)', ylab='Northing (m)')
## Add the gauges to the plot
> points(gauges$Easting, gauges$Northing, col='blue', pch=2)
## Estimate/plot watershed centroid using the basin's mean coordinates
> centx=mean(chk[,1])  # Average easting
> centy=mean(chk[,2])  # Average northing
> points(centx, centy, col='black', pch=4)
```

Save this plot for your writeup.

Simple Arithmetic Mean

Compute the simple arithmetic mean for gauges A-H after substituting the missing data for station E from the mean of all methods in question 2 of Assignment Part II. Substitute the missing value (MV) for station E in your dataframe as follows:

```
> gauges$Aug31[5] = MV
```

This provides your first estimate of basin average precipitation for 31 August 1991.

Distance Weighted Mean

The first step here is to create a new table with the information for gauges A-H and adding one more (gauge I) with the coordinates for the watershed centroid, while keeping all of its other information as "NA". You can also insert the missing precipitation data for station E in this new table. Save this new table as 'gauges2.txt' and then load it as follows:

```
> ngauges=read.table('gauges2.txt', header=T)
> ngauges
```

Then follow the same approach for the inverse distance (use $w_i \sim \frac{1}{d_g}$ where w_i is the ith weight, d_g is the distance (km) between the gauges and the watershed centroid, and the scaling exponent $b = 1$) used in Assignment Part II.

To calculate the distance between the gauges and watershed centroid, use the `dist()` command:

```
> xy = data.frame(ngauges$Easting, ngauges$Northing)
> d = dist(xy,upper=TRUE)
> d.stnE = c(d[8], d[15],d[21], d[26], d[30], d[33], d[35], d[36], NA)
```

Then to calculate the weights d_g^{-1}, type:

```
> invd = 1/d.stnE
> D = sum(invd,na.rm=T)
> WP = invd*ngauges$Aug31
> ID.hat = sum(WP,na.rm=T)/D
```

This provides you the second estimate of areally-averaged precipitation for the Cheakamus Lake catchment.

Isohyetal Method

Mountainous terrain causes precipitation to vary greatly in space. Often these patterns may reflect orographic forcing or other localized phenomena. In these cases it may be more meaningful to derived weighted precipitation based on the isohyetal method. This method multiplies the averaged isohyetal value with the area enclosed by the two isohyets ($P_{(i-1)}$ and $P_{(i+1)}$ and the watershed divide to estimate the fractional component of total basin precipitation for that polygon. These estimates are summed for the entire watershed and divided by the catchment area to derive a weighted estimate of total precipitation.

Isohyet (mm)	Fractional Area
< 40	0.01
$40 - 60$	0.40
$60 - 80$	0.35
$80 - 100$	0.15
100	0.09

Table 2: Fractional areas for several isohyets observed on 31 August 1991.

From the contours of isohyets plotted on the watershed map, one can establish their fractional areas for each (see Table 2).

Use the equation:

$$\hat{P} = \sum_{i=1}^{N} w_i p_i \tag{6}$$

where w_i is the fractional area of each isohyet (i) and p_i is the mean value of the range of precipitation for a given isohyet range. Then from the information in Table 2, calculate the isohyet-weighted total storm precipitation (in mm).

The Writeup

1. Submit the watershed map that identifies the location of the precipitation gauges and watershed centroid. (3 marks)
2. Show your calculations for the three different estimates of basin total storm precipitation (in mm), and provide a summary table. (6 marks)
3. How does the isohyetal method of estimating the precipitation field compare to the other estimates? (3 marks)
4. How do your estimates compare to the observed runoff from the catchment for the 31 August event (104 mm) during a period of 24 hours? Speculate on why this might be. (3 marks)