Suraj Shrestha

Standard error estimation of regional regression method used in Nepal for hydrological data analysis

Regression analysis of historical hydrological data

Helsinki Metropolia University of Applied Sciences

Bachelors Degree

Environmental Engineering

Thesis

13.01.2017

Author(s) Title Number of Pages Date	Suraj Shrestha Standard error estimation of regional regression method used in Nepal for hydrological data analysis 37 pages + 29 appendices 13 01 2017
Degree	Bachelor's Degree
Degree Programme	Environmental Engineering
Specialisation option	Waste and water treatment
Instructor(s)	Heerman waiba Project Manager at Garjang Upatyaka Hydropower Ltd Veli-Matti Taavitsainen, Principal Lecturer, Helsinki Metropolia UAS

Hydrological studies are made for various water related projects; hydropower is one of the sectors that require hydrological studies. There are number of ungauged sites in Nepal where there is no hydrological station. The regional regression methods for hydrological studies were developed by Department of Hydrology and Meteorology jointly with Water and Energy Commission Secretariat with limited information of hydrology. The equation has been used in past for the hydrological studies of ungauged sites.

In this thesis, regression models were developed using the historical hydrological data available in *Hydrological estimation in Nepal* report and the standard error was estimated for each model. The statistical analysis was done using R software. R is a free software environment for statistical computing and graphics.

The topographical factor of Nepal also supports the fact that these models are not reliable for all the river conditions. There are rivers that are above 3000 m and elevation only cannot be a single factor to influence the hydrology. The results of the thesis suggest there must be specification for the use of these formulas in future and establishment of more hydrological station, updating daily data can be done for the better hydrological studies

Keywords	regression model, prediction interval, standard error, relative standard error ,ungauged sites

Acknowledgement

I would like to thank Helsinki Metropolia University of Applied Science for giving me an opportunity to study and my teacher's for supporting me during my period in university. I take this opportunity to thank my supervisor Veli-Matti Taavitsainen for guiding me throughout the thesis writing process and my work supervisor Heeraman Waiba at Garjung Hydropower for academic support.

I deeply thank to Sudip Bhattarai from the Department of Hydrology and Meteorology for providing the important documents for thesis. My special thanks reserved for our senior Lecturer Kaj Lindedahl and my friends Bijay Karki, Rupesh Sharma, Navagantuk Sharma and Sunaina Shrestha for their immense support.

The thesis was challenging for me; I am solely responsible for any kind of mistake in thesis. I dedicate the thesis to all of my friends and family who supported me.

Helsinki 13.01.2017
Suraj Shrestha

Contents

List of Figures

List of Tables

List of Abbreviations

DHM	Department of Hydrology and Meteorology
Const	Constant
Coef	Coefficient
Avg Elv	Average Elevation
PIann	Annual Precipitation
A<3k	Area below 3000 meter
A<5k	Area below 5000 meter
MULTR	Multiple Regression Programs
WESC	Water and Energy Commission Secretariat
FORTRAN	Formula Translation
GEV	Generalized Extreme Value
RSE	Relative Standard Error
SE	Standard Error
m	meter
m^3/s	Cubic meter per second
km^2	Square Kilometer

1 Introduction

There are numbers of ungauged catchment areas in Nepal that require hydrological studies for hydropower projects for various purposes such as flood frequency analysis, low flow analysis and flow duration of a river. In 1962, the hydrological survey was started by the Government of Nepal at; a section under the Department of Electricity currently the section is known as the Department of Hydrology and Meteorology (DHM) (Environment, 2006) that comes under the Ministry of Environment Government of Nepal. To conduct the hydrological analysis in Nepal for various water projects, DHM, jointly with the Water and Energy Commission Secretariat (WESC) in 1980, developed the regional models to study the different hydrological characters such as mean monthly flow, flood for the ungauged sites.

These models were published in 1990 titled *Methodologies for Estimating Hydrologic Characteristics of Ungauged Locations in Nepal*. The regional regression model mentioned in the main report (Sharma & Adhikari, 2004) used the data that were available up to 1985. This statistical analysis was done by using FORTRAN based multiple regression programs (MULTR). (Sharma & Adhikari, 2004). The data used for these models showed substantial variation for all the parameters for example the range for the average elevation was from 911 m to 4863m and 989mm to 3741mm for the annual precipitation.

Monsoon dominates the hydrology of Nepal; during the summer monsoon from June to September, Nepal receives around 80 percent of its precipitation and rest of year it is extra monsoon period or northeast monsoon period. Monsoon Wetness index is the average rainfall from June 15 to September 15. There is a topographic influence most of the rivers also receive input from snow that melts during summer that increase the flow of water in most of the rivers in Nepal.

The thesis has three objectives. The first objective was to understand the regional method used in Nepal, for hydrological analysis. The second objective was to know whether the historical data would give the same coefficient values used in regional method. The third objective was to calculate the standard error of the regional method to know the accuracy of such methods.

2 Background

2.1 Regional regression method

The regional regression equations were developed using historical data, and the equations were used for hydrological studies of the ungauged sites. The equation and data were obtained from a report *Hydrological estimation in Nepal* (Sharma & Adhikari, 2004).Since the constants and the coefficients of these models were used to determine the hydrological character of the ungauged sites, there is always higher probability of error in the predicted values. Understanding such error can be beneficial for any kind of the water project such as hydropower and irrigation. The standard error estimation of such models is a measure of accuracy of predictions made by such models. These are the four regional methods developed using the historical data collected from 51 hydrological station of Nepal.

1. Long term mean monthly flow
2. Flow duration
3. Low flow analysis
4. Flood analysis

2.2 Hydrological terminology

Hydrological terminology related to this thesis is explained below:

- Catchment area or drainage area is defined as the entire area of river basin where the surface runoff drains to the river, through rain, melting of snow or any other activates.
- Gauging station are the sites where the hydrometric measurement of the water level or surface elevation and volumetric discharge or flow is recorded. The data are recorded in daily basis, for the variables like extreme flow and low flow.
- Ungauged sites are sites where hydrometric measurements such as stream flow are not recorded.
- The parameters required for the ungauged sites are obtained from the gauged station in the similar region. It is expected to have similar characteristics such

that the hydrological response should be similar. Few factor such as spatial proximity of another gauging station, similarity in mean elevation, area and slope of catchment are also factor for selection of donor-gauged station.

- Annual Precipitation is a total rainfall in a year and in Nepal 80% of the rainfall is during the monsoon period.
- Elevation is height above or below a fixed reference point. The range for the average elevation in context of Nepal is from 911 m to 4863 m from the sea level.
- Catchment area (km^2) under 3000 meter and 5000 meters are other two parameter used for the analysis.
- Annual exceedance probability P is the probability that a specified magnitude will be equalled or exceeded in any given year. (Risley, et al., 2013)
- Return period or Recurrence interval the recurrence interval or return period T is the average interval in years between successive occurrences of annual exceedance probability P are reciprocal to each other. (Ralph & Wesley, 2001):

$$T = \frac{1}{P} \ and \ P = \frac{1}{T}$$

- A two-year flood is a flood that has an annual exceedance probability of 0.5 or 50%
- A hundred-year flood is a flood that has an annual exceedance probability of 1%.

2.3 Long-term mean monthly flow

The long-term mean monthly flow is the average flow of a month that gives information on the variability of the flow and the quantity of the water available. (Sharma & Adhikari, 2004). To calculate the long-term mean monthly flow, two methods are selected drainage area ratio method and regional regression method developed by DHM. (BPC Hydroconsult;Practical Action, 2002)

2.4 Drainage-area ratio method

Eq.1 is the drainage area ratio method based on the assumption that the stream flow of ungauged station can be estimated by multiplying the ratio of the catchment area of ungauged station by the stream flow for the nearby stream flow-gauging station This method is also known as the catchment area ratio method (CAR) (Emerson, et al., 2005):

$$\tilde{Y}ij = \left(\frac{A_y}{A_x}\right) X_{IJ}, \quad (1)$$

where, $\tilde{Y}ij$ is the estimated stream flow, in meter cube per second (m^3/s) for month *i* and year *j* for the ungagged station. A_y = the drainage area or catchment area, in square kilometer (km^2), for the ungauged station. A_x = the drainage area or catchment area, in square kilometer (km^2), for the stream flow gauging station and, X*ij* = the stream flow, in meter cube per second (m^3/s) for month *i* and year *j* for the stream flow gaging station.

2.5 Regional regression method for the long term mean monthly flow

The regional regression formula for long-term flow was developed for the mean monthly flow of the ungauged station in Nepal. The equation uses for the long –term mean monthly different constant value for each month and a coefficient for average annual precipitation, average elevation and a basin area below 3000 meter and 5000 m. These values were created calculating the regression of historical data of mean monthly flow of each month and average annual precipitation, average elevation and basin area below the 3000 meter and 5000 m from the 51 hydrological stations. There are two different equations; Eq.2 is for all the nine months expect for March, April, and May and the Eq. 3 is for the March, April and May.

$$Q_{flow} = e^{a+b\ln(x)+c\ln(y)+d\ln(z)}, \quad (2)$$

where Q mean is the monthly flow and subscript $flow$ is the mean monthy flow for all the months beside March, April and May, a is the constant for a particular month, b is the coefficient for the average elevation x, c is the coefficient for annual precipitation *y*, d is the coefficient for (catchment area below 3000 m) *z*.

For regression analysis, Eq.2 is inserted into the following linear formula.

$$\ln Q = a + \ln bx + \ln cy + \ln dz \qquad (2b)$$

Where Q_{flow} stands for the response variable and x, y, z is the independent variable. The Eq.3 is for the March, April and May.

$$Q_{flow} = [a + f\sqrt{e}]^2, \qquad (3)$$

where, Q is mean monthly flow and the subscript $flow$ is mean monthly flow for March, April and May, a is the constant for month, f is the coefficient of catchment area below 5000 m *e.*

For regression analysis, Eq.3 is inserted into following linear form.

$$\sqrt{Q_{flow}} \; = a + f\sqrt{e} \qquad (3b)$$

Where $\sqrt{Q_{flow}}$ stands for the response variable and $\sqrt{x}$ stands for the independent variable.

2.6 Flow duration and Flow Duration Curve

Flow duration data is the mean flow value measured over a specified time interval that has been exceeded various percentages of the specified time interval. For example, 80% exceedance probability represents that the high flow has been exceeded 80 % of all days of the flow record. (Risley, et al., 2013)

A flow duration curve is a plot of flow duration data where the discharged flow is plotted against percentage of time. It shows the percentage of time where the flow in a stream is likely to equal or exceed the value of particular interest. The area under the flow duration curve gives the average daily flow, and the median daily flow is the 50% value. Characterization of the ability of the basin (or the reservoir) to provide flows of various magnitude in relation to the amount of time. (H.M, 2005/2006)

Similarly, the shape of the curve provides further information about the basin in relation to the seasonal variations. The shape of the curve in the high-flow region indicates the type of flood regime the basin is likely to have, whereas, the shape of the low-flow region characterizes the ability of the basin to sustain low flows during dry seasons. Steep curve is the result of flood caused by the rain. However, the snowmelt flood will yield much flatter curve (Oregon State University, 2005). These are the most important factors, which directly affect the performance of any hydropower plant.

Eq. 4 is regional regression equation for the flow of all the exceedance probability beside 0% and 100%, whereas Eq.5 and Eq.6 are for the 0% and 100 % exceedance probability (Sharma & Adhikari, 2004);

$$Q_{\%} = e^{a+b\ln(x)+c\ln(y)+d\ln(z)}, \tag{4}$$

where Q is the flow and subscript is $\%$ is the probability of exceedance for all the flow excluding 0% and 100%, a is the constant for a particular probability of exceedance, b is the coefficient for the average elevation x, c is the coefficient for annual precipitation y, d is the coefficient for (catchment area below 3000 m) z.

For analysis regression, Eq.4 is inserted into the following linear formula.

$$\ln Q = a + \ln bx + \ln cy + \ln dz, \tag{4b}$$

where $Q_{\%}$ stands for the response variable and x, y, z are the independent variables.

The independent variable of the flow for 0% probability of exceedance are average elevation and area under 3000 m whereas the independent variable for 100% probability of exceedance are annual precipitation and area under 5000 m. The equation 5 is for the 100% probability of exceedance.

$$Q_{100\%} = \left[a + b\sqrt{y} + c\sqrt{z}\right]^2, \tag{5}$$

Q is the flow and subscript is 100% is the probability of exceedance, a is the constant of flow for 100% probability of exceedance, b is the coefficient for the average elevation y, c is the coefficient for the area under 3000 m z.

For regression analysis, Eq.5 inserted into the following formula

$$\sqrt{Q_{100\%}} = a + b\sqrt{y} + c\sqrt{z} \ , \qquad (5b)$$

where $\sqrt{Q_{100\%}}$ stands for the response variable and the independent variables are $\sqrt{y}$ and$\sqrt{z}$

The equation Eq.6 is for 0% probability of exceedance.

$$Q_{0\%} = \left[d + e\sqrt{g} + f\sqrt{h}\right]^2 \ , \qquad (6)$$

where Q is the flow and subscript is 0% is the probability of exceedance ,d is the constant for 0% probability of exceedance, e is the coefficient for the average elevation g, f is the coefficient for the area under 3000 m h.

For regression analysis, Eq.6 is inserted into the following linear formula.

$$\sqrt{Q_{0\%}} = d + e\sqrt{g} + f\sqrt{h} \ , \qquad (6b)$$

where $\sqrt{Q_{0\%}}$ stands for the response variable and the independent variables are $\sqrt{g}$ and$\sqrt{h}$

2.7 Regional regression method for low flow analysis

Low flow analysis is to determine allowable water transfers and withdrawals to determine a minimum downstream release. Eq.7 is a regional method for assessing low-flow characteristics for 1 day, 7 days , 30 days and for a monthly duration, e.g. 1-day low flow is the lowest value obtained from each year’s daily streamflow and the seven-day low flow is the minimum value obtained from the consecutive seven-day average.

Estimation of low flow status on a river is important for designing a single-purpose or multi-purpose water resources project considering extreme condition regarding the availability of adequate water supply. The information of low flow is needed to determine maximum power that a run of river hydropower can generate during the dry season.

$$Q_{d.y} = \left[a + b\sqrt{e}\right]^2, \qquad (7)$$

where Q is the flood and subscript d and y are the day and return period, a is the constant is the constant for the particular average of a day of a period, b is a coefficient for same average of a day and a return period , e is the area below the 5000 m elevation of a catchment area.

For regression analysis Eq.7is put into the following linear formula.

$$\sqrt{Q_{d.y}} \ = a + b\sqrt{e}, \tag{7b}$$

where $\sqrt{Q_{d.y}}$ stands for the response variable and $\sqrt{e}$ stands for the independent variable

2.8 Flood Frequency analysis

Flood is the extreme event that does not occur frequently but it does occur, it has its effects. The tendency of flood also depends on the geographical characteristics of the area, such as elevation in context of Nepal. There are various methods for the flood frequency analysis, such as Gumbel method, Log Pearson Type III distribution and normal distribution that uses stream flow as its variable, but the formula developed by Nepal for uses basin area under certain elevation.

Nepal developed its own regional regression flood frequency analysis formula also known as (WECS /DHM) method from long-term flow data collected from the 51 hydrological stations of the Nepal. The length of the record varied from 11 to 34 years, and the stations having records for less than 10 years were excluded. The constant and coefficient in the formula are for basin area under 3000 m of the any hydrological station of Nepal. There are two different equations for flood flow Eq. 8 and Eq.9. Eq.8 is for the 2-years and 100-year flood that has different constant and power value. Eq.9 is for the rest of the return periods. Eq.9 uses the standard normal variate denoted by *s* and the flood flows obtained from equations Eq.8 for 2-year and 100-year flood.

$$Q_T = b_0 \, a^{b_1} \quad , \tag{8}$$

where Q is the flood flow and subscript T = 2 and T = 100 are the 2 year and 100 year return period, a is the area under 3000 meter, b_0 is the constant for 2 year and 100 year flood , b_1 the power for the 2 year and 100 year flood.

For regression analysis, Eq.8 is inserted into the linear formula.

$$\ln Q = \ln(b_0) + b_1 \ln(a) \quad (8b)$$

Where Q stands for the response variable and a is the independent variable. For other years than two or 100-year flood, Q_T is calculated by Eq. 9 below

$$Q_T = \exp(lnQ_2 + s\sigma_l) \quad , \quad (9)$$

where

$$\sigma_l = \frac{1}{2.326} \ln(\frac{Q_{100}}{Q_2})$$

Q is the flood flow and T is the return period for rest of the return period, a is the area under 3000 meter, T is the return period.

Table 1 shows standard normal variates for various return periods.

Table 1 Values for the standard normal variate for different return period

Return Period (T) (in years)	Standard Normal Variate(s)
2	0
5	0.842
10	1.282
20	1.645
50	2.054
100	2.326
200	2.576
500	2.878
1000	3.090
5000	3.540

10000	3.719

3 Methodology

3.1 Linear regression

Regression is the study of the dependence. The simple and multiple regression models were developed using historical data. The simple linear regression model is for modeling the linear relationship between dependent variable *y* and the independent variable *x.* The simple regression model is written as following form

$$y = \beta_0 + \beta_1 x + \mathcal{E},$$

Where *y* is the dependent variable, β_0 is *y* intercept, β_1 is gradient or the slope of regression line, x is the independent variable, and Ɛ is the random error. It is usually assumed that error Ɛ is normally distributed with E(Ɛ) = 0 and a constant variance Var(Ɛ)=$\boldsymbol{\sigma}^2$.

The second type of regression is the multiple linear regression with one dependent variable and more than one independent variable. The multiple linear regression assumes that the response variables is a linear function of the model parameters.
The general form of the multiple linear regression model as follows (Yan ,, et al., 2009);

$$y = \beta_0 + \beta_1 x + \cdots \beta_1 x_p + \mathcal{E},$$

Here y is the dependent variable, $\beta_0, \beta_1, \beta_2 \beta_p$ are the regression coefficients,$x_1, x_2 x_3 x_n$ are the independent variables in the model. It is usually assumed that error Ɛ is normally distributed with E(Ɛ) = 0 and a constant variance Var(Ɛ)=$\boldsymbol{\sigma}^2$. (Yan ,, et al., 2009);

3.2 Error analysis

3.2.1 Prediction interval

The prediction interval is an interval that contains future results with a given probability. The prediction interval applies to individual observation and they measure how much uncertainty is associated with a single estimated value ŷ (Jarman, 2013).

3.2.2 Rules of thumb of normal probabilities

In many cases, it is accurate enough to use approximate normal probabilities given by the following rules of thumb: If X~N (μ, σ²) then

1. $P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 68\%$
2. $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95\%$
3. $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 99.7\%$

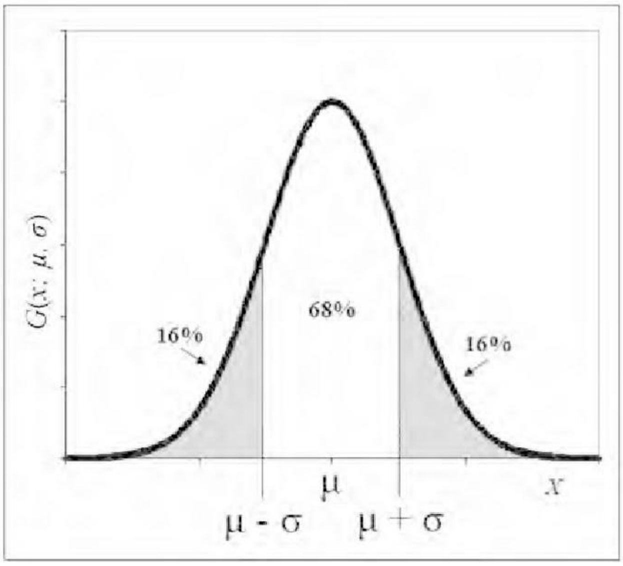

Figure 1. 68% normal distribution **(Roterman-konieczna, 2009)**

In Figure 1, the three-sigma rule presented graphically-the range specified with values μ- σ and μ+σ represents the range for normally distributed measurements whose probability of meeting equals 68%. The hydrological data are normally distributed data and 68 % is the normal probability for standard error.

3.2.3 Standard error of regression line (SE)

The standard error estimates the variation of observed y values around the regression line, and this value can be used to put a margin of error or prediction interval around a y value. (Jarman, 2013).

3.2.4 Relative standard error (RSE)

Relative standard error is a measure of sampling error, which is obtained by expressing the standard error as a percentage of the estimate.

$$RSE(estiamate) = 100 \times \frac{SE(estimate)}{estimate}$$

The relative standard errors of less than 25% are sufficiently reliable, and relative standard errors between 25% and 50% should be used in caution. Estimates with relative standard errors greater than 50% are considered too unreliable for general use. (Austrailan Health Minsters' Council, 2014)

However, it was not possible to visualize the prediction interval lines for multiple regression models, but upper, lower and fitted bounds were obtained by ‘predict’ function of R. To calculate the standard error, the values from the fitted bound were subtracted from the upper bound, whereas to get the relative standard error, the subtracted values were divided by fitted values and multiplied by 100, to get the results as percentage.

3.2.5 Linear regression and error analysis terminology

- Standard Uncertainty represents standard error that the 68% of the time the true value of the measured quantity falls with the stated uncertainty (Bell, 1999)
- Uncertainty is a quantification of the doubt about the measurement result (Bell, 1999)
- Error is the difference between the measured value and the true value of the things being measured (Bell, 1999)
- Absolute error is the amount of physical error in a measurement period.
- The adjusted R squared is goodness of fit measure and when it is close to 1 that indicates all the variability of response data is around its mean
- Dependent variable is a variable whose value is depended upon the other variable
- Independent variable is a variable whose value is not depended upon other variables
- P- Value is the level of marginal significance within a statistical hypothesis test representing the probability of the occurrence of a given event.

- Significance level (α) is the probability of rejecting the null hypothesis when it is true. The most common values are 0.05, 0.01or 0.001
- Null Hypothesis (***H_0***) is type of hypothesis used in statics that purpose that no statistical significance exists in set of given observation.
  H_0 in this study is response variable or dependent dose not depends upon independent variable

4 Data

According to (Sharma & Adhikari, 2004) hydrological data of the major river basins in Nepal are available for more than 40 years till 1995 from the 51 regular stream -gauging stations. The following hydrological data are presented in Appendix A to I from page 1 to 7;

1. Mean monthly flow from all 51 hydrological station of Nepal
2. Average elevation of all the 51 hydrological stations
3. Annual precipitation
4. Catchment area under 3000 and 5000 meter
5. Flow duration
6. Low flow data in different return year period such as 2year flood in 1 day
7. Constant and coefficient values for all the methods

The daily flow data for the station 439.9 of Midhim khola was obtained from DHM. The data had many missing values. To fulfill this requirement data were predicted and generated using the observed data to fill the gap. Historical data is not sufficient to predict the risk caused by extreme event such as flooding and drought (Yadav, 2002).

Moving mean cannot be an option for the data generation due to the significant periodicity in the streamflow in Nepal. Months in the monsoon season are serially correlated with each other or with those of the previous year. Therefore, the significant deterministic component is other than the mean. Most of the data in hydrology are serially correlated where the data have the serial dependence.

To predict the missing values in the data, partial least square model was used from the r package the package name is ‘pls’. The model was validated by cross validation and the best predicting model is selected to predict the data (Harlad & Tormod, 1989)

5 Results

5.1 Result of the drainage area ratio method

The hydrograph is a plot of the variation of discharge of water with respect to time. Figure 1 is a hydrograph that shows the comparison between the predicted mean monthly flows data of the four rivers with respect to Midhim khola (Midhim River). These Rivers were selected as the possible donor drainage areas using Eq.1. Khudi River was selected as a donor catchment because of the similar area size and location as both in lamjung districts. Mean monthly flow data for all the four rivers is available in Appendix J in page 7.

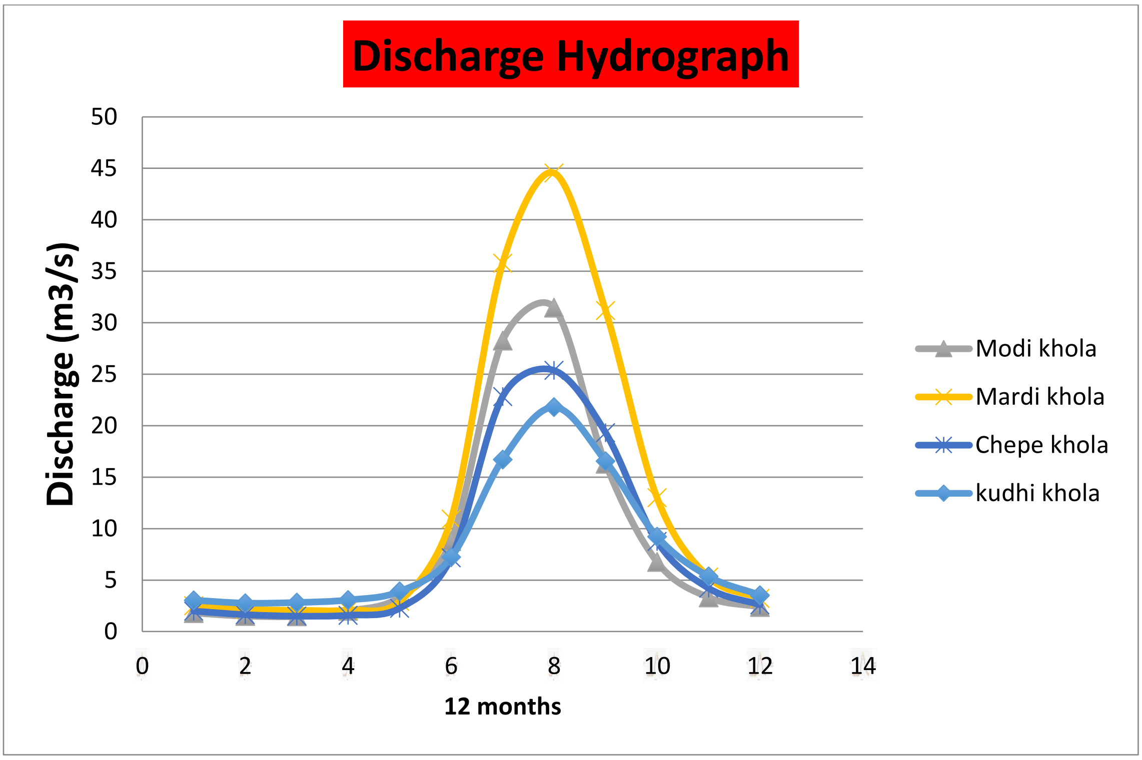

Figure 2: Hydrograph of mean monthly flow of four rivers

5.1.1 Comparison between new and actual data

The historical data for Khudi River is only available from 1983 to 1995. The data misses some of its value for year 1987, 1988 and 1992, therefore partial least square (pls) were conducted to provide the missing values. The result for the new and actual data is in Appendix K in page 8 and the R script for a cross validation and a selected model is given in Appendix R from page 10 to 14.

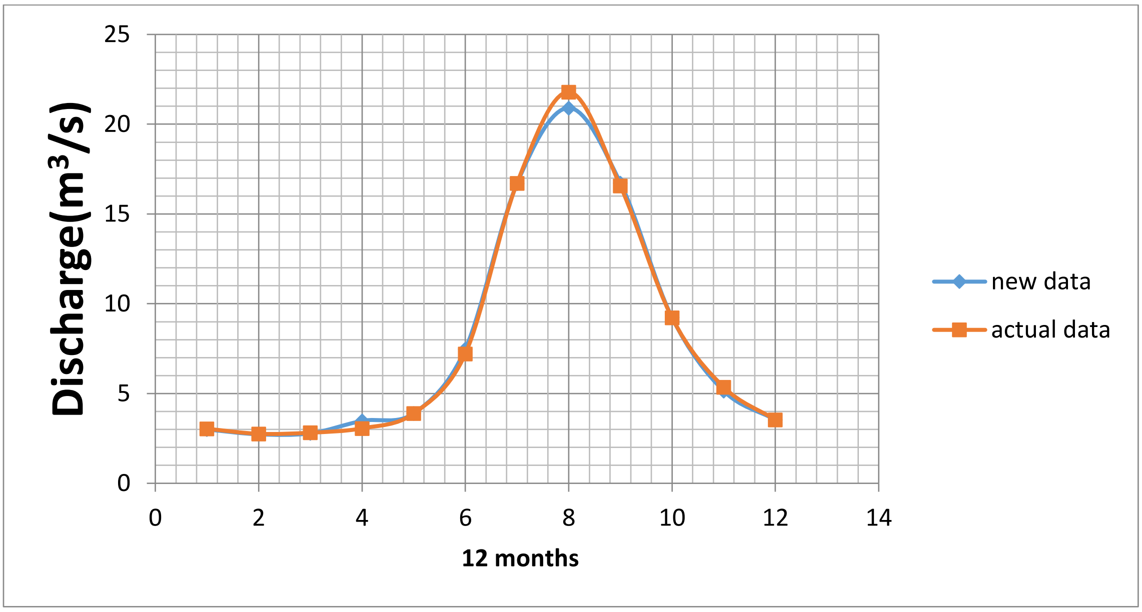

Figure 3: Comparison between the discharge hydrograph of actual and new data

Both the lines in Figure 2 are of Midhim River. The difference between the red line and blue line is that the red line was obtained with actual data of Khudi River, whereas the blue line was obtained by filling the missing values in the data of Khudi River. The data is presented in Appendix K in page 8.

5.1.2 Comparison between predicted mean monthly flow data from two methods

The hydrograph in Figure 3 shows the difference between the two lines. The blue indicates the data obtained using the regional regression equations Eq.2 and Eq.3 whereas the red line indicates the data obtained using Eq.1. The data obtained from both the method are given in Appendix M in page 8 and 9.

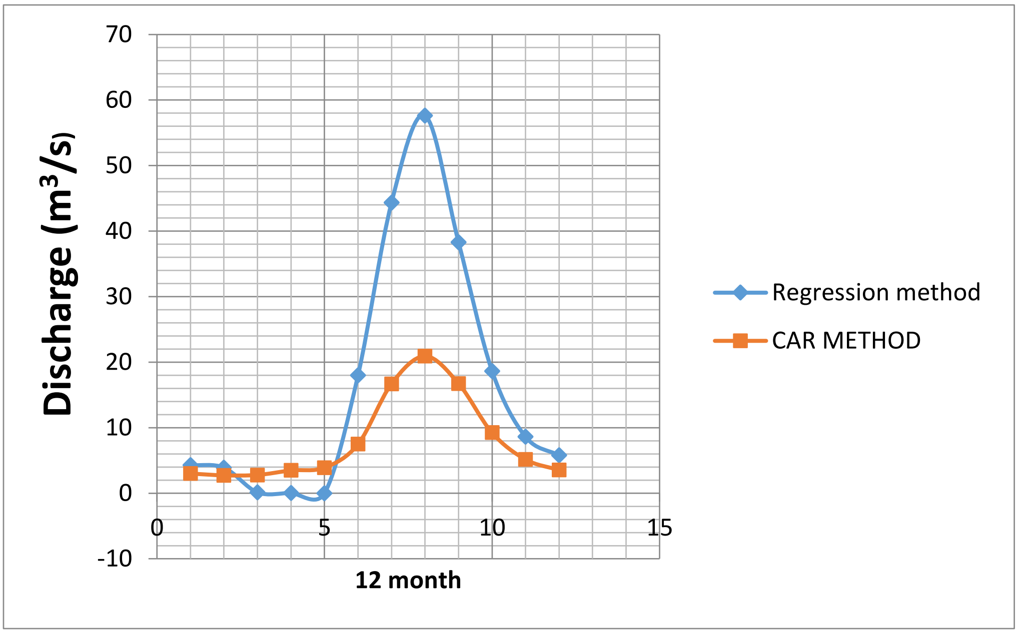

Figure 4: Hydrograph comparison between two methods

5.2 Results of regional regression method for mean monthly flow

There are two different models for 12 months. The first model is a multiple linear regression model using Eq.2 in linearized form (Eq.2b) that uses log transformation for 9 months excluding March, April and May. The second one is a simple linear regression model using Eq.3 in linearized form (Eq.3b) that uses square root as transformation for March, April and May.

5.2.1 Results of a models

The January model is taken as an example for the mean monthly flow to show what kind of results were obtained and how they are interpreted. The R script and the results for rest of 8 months are presented in Appendix S from page 14 to 20.

The model uses Eq.2 in linearized form (Eq.2b) and it is called Janmodel. The dependent variable in the model is Jan that is the mean monthly flow for January from all the 51 stations, and the average elevation (AE), catchment area under 3000 m (CA3) and annual precipitation (AP) are independent variables

```
lm(formula = log(Jan) ~ log(AE) + log(AP) + log(CA3), data = hydro)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.90087 -0.14450 -0.00959  0.11742  0.99233 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -16.77791    1.75506  -9.560 1.34e-12 ***
log(AE)       1.36253    0.10393  13.110  < 2e-16 ***
log(AP)       0.47771    0.20384   2.344   0.0234 *  
log(CA3)      0.81762    0.03335  24.513  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3492 on 47 degrees of freedom
Multiple R-squared:  0.9559,	Adjusted R-squared:  0.9531 
F-statistic: 339.6 on 3 and 47 DF,  p-value: < 2.2e-16
```

The adjusted R- squared is 95%, which indicates good fit. The model p-value is less tha n ($2.2e^{-16}$), and all individual coefficient p-values are below the level of significance, and thus the response variable depends significantly on the independent variable. (cf. p. 15)

The final Janmodel using Eq.2 is

$$Q_{January} = e^{-16.77+1.33\ln(\mathrm{AE})+0.47771\ln(\mathrm{AP})+0.1762\ln(\mathrm{CA3})}.$$

Figure 5 show the plot of actual versus predicted flows. The actual values are the response data whereas predicted values were generated in R.

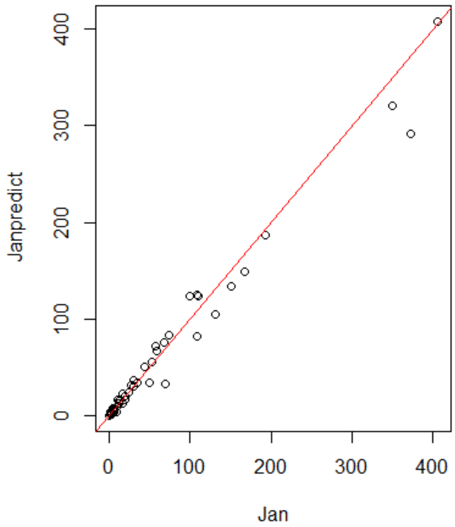

Figure 5: Actual versus predicted stream flow values for January

The plot shows that the model is accurate and there is a strong correlation between the predicted and actual values.

Table 1 shows results for Janmodel, and it gives the prediction errors for the values in the range of the data, in a table all the combination of independent variables in 3 levels are given. The first three columns gives the independent variables, catchment area below 3000 m (CA3), average elevation (AE) and annual precipitation (AP). The 'Janfit' column gives the fitted value or predicted flow value whereas Janlwr and Janupr are the, lower and upper prediction limit of the flow. The prediction level of 68% is used in concordance of standard uncertainty

Table 2 Prediction errors result for Janmodel

	CA3 (m^2)	AE (m)	AP (mm)	Janfit (m^3/s)	Janlwr (m^3/s)	Janupr (m^3/s)	Janerror*100 (%)
1	11	911	989	0.1067	0.0694	0.1641	53.8127
2	9745.5	911	989	27.418	18.585	40.450	47.5294
3	19480	911	989	48.302	32.643	71.473	47.9711
4	11	2887	989	0.5137	0.3342	0.7896	53.7014

5	9745.5	2887	989	132.00	90.755	191.98	45.4434
6	19480	2887	989	232.54	159.61	338.79	45.6930
7	11	4863	989	1.0453	0.6729	1.6239	55.3435
8	9745.5	4863	989	268.61	183.56	393.07	46.3348
9	19480	4863	989	473.21	323.03	693.20	46.4913
10	11	911	2365	0.1618	0.1100	0.2381	47.1494
11	9745.5	911	2365	41.583	28.150	61.426	47.7190
12	19480	911	2365	73.255	49.207	109.06	48.8721
13	11	2887	2365	0.7791	0.5329	1.1391	46.2066
14	9745.5	2887	2365	200.19	138.25	289.88	44.8002
15	19480	2887	2365	352.67	241.91	514.15	45.7884
16	11	4863	2365	1.5854	1.0741	2.3401	47.6015
17	9745.5	4863	2365	407.38	280.33	592.01	45.3209
18	19480	4863	2365	717.67	490.84	1049.3	46.2123
19	11	911	3741	0.2015	0.1358	0.2990	48.3943
20	9745.5	911	3741	51.767	33.935	78.969	52.5478
21	19480	911	3741	91.197	59.216	140.45	54.0056
22	11	2887	3741	0.9699	0.6596	1.4261	47.0336
23	9745.5	2887	3741	249.22	166.87	372.21	49.3516
24	19480	2887	3741	439.05	291.41	661.47	50.6607
25	11	4863	3741	1.9737	1.3316	2.9255	48.2239
26	9745.5	4863	3741	507.15	338.87	758.99	49.6580
27	19480	4863	3741	893.44	592.17	1348.0	50.8764
mean				193.11	130.53	285.77	48.6201

The lower flow limit is within the range of 0.069 to 592.17 m^3/s, and the average is 130.53 while the upper flow limit is 0.164 to 1347.99 m^3/s and, the average is 285.77 m^3/s. The average of predicted flow value is 193.11 m^3/s, and, the average of the relative standard error for all the 27 experiments is 49%., which is below 50% thus the model moderately reliable and can be used in caution.

The March model is shown here as an example. The model uses Eq.3 in the linearized form (Eq.3), and it is called Marchmodel. Mar is the flow of the river from all the stations in March and it is a dependent variable while the area under 5000 meter is an independent variable.

```
Call:
lm(formula = sqrt(Mar) ~ sqrt(CA5), data = hydro)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.3949 -0.4312  0.1315  0.7107  1.7705 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.383586   0.204175   1.879   0.0662 .  
sqrt(CA5)   0.091134   0.002689  33.887   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9449 on 49 degrees of freedom
Multiple R-squared:  0.9591,	Adjusted R-squared:  0.9582 
F-statistic:  1148 on 1 and 49 DF,  p-value: < 2.2e-16
```

The adjusted R square value is 95%, which indicates good fit. The p value of the model is ($2.2e^{-16}$) and the individual coefficient p value is below the level of significance, and th us the response variable depends significantly on the independent variable. (cf. p. 15)

The final Marchmodel using Eq.3 is

$$Q_{March} = [0.383586 + 0.091\sqrt{\text{CA5}}\,]^2$$

Figure 6 is a plot of the actual value in March versus predicted flow value. It shows the strong correlation, and the model fits the data

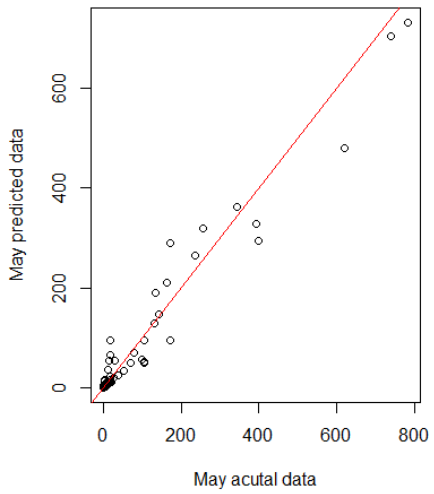

Figure 6: Actual vs. predicted flow of March

Figure 7 is a plot of streamflow versus catchment area under 5000m of March, the two red lines are the prediction interval of 68 %, and the blue line is the fitted value or the predicted flow.

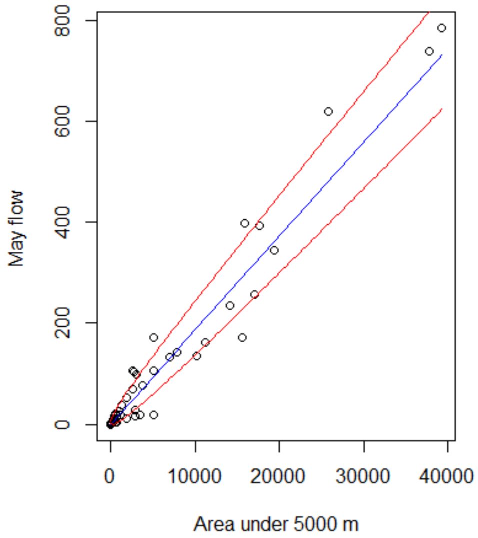

Figure 7: Prediction interval plot of the March model

The minimum, maximum, mean predicted flow values and relative standard error for the mean monthly flows of all months are presented in Appendix N in page 9. The first three columns are for the predicted flow values in m^3/s and the remaining three are for minimum relative standard error, maximum relative standard error and mean relative standard error. The predicted flow has been below 1 m^3/s for all the months beside the month for June, July and August.

March, April and May show the huge range between the minimum and maximum values of relative standard error. The reason for the high relative standard error is extremely low predicted flow and the small absolute error. The result for the March can be taken as an example where the maximum predicted flow value is $340m^3/s$ that gives 11% relative standard error, whereas the minimum predicted flow is $0.5m^3/s$ that gives 418% relative standard error. It can be concluded that the flow must be above $350m^3/s$ to have small relative standard error and the model is sufficiently reliable. March, April and May uses Eq.3, and the rest of the months uses Eq.2 that have the small range between the minimum and the maximum relative standard error. The minimum relative standard errors March, April and May are below 16% which shows the model is sufficiently reliable. The minimum relative standard error for other months excluding June and July are around 50%, thus the model is moderately reliable and can be used with caution. However, the model for the June and July can be used to understand the order of magnitude of flow during these months.

5.3 Results for the regional regression equation of exceedance probability

5.3.1 Result of the 0%, 5% and 100% probability of exceedance

There are three different equations, the probability of exceedance 5%, 20%, 40%, 60%, 80% and 95% uses Eq.4 whereas Eq.5 and Eq.6 is for 100% and 0% probability of exceedance. The R script for the probability of exceedance is given in Appendix T from page 20 to 23 and results is given in Appendix P in page 10.

The result of 5% probability of exceedance is shown here as an example. The discharge for 5% probability of exceedance means there is 5% chance that the flow will occur in a year.

The model uses the Eq.4 in a linearized form (Eq.4b), and it is called Flow5model. It is a multiple linear model where an average elevation (AE), catchment area under 3000 m (CA3) and annual precipitation (AP) are independent variables and Flow5 is the flow for 5% probability of exceedance for all the stations is a dependent variable.

```
lm(formula = log(Flow5) ~ log(AE) + log(AP) + log(CA3))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.59580 -0.19568  0.03429  0.15113  0.85768 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -14.08009    1.66823  -8.440 5.66e-11 ***
log(AE)       1.10683    0.09879  11.204 7.21e-15 ***
log(AP)      0.67401    0.19375   3.479   0.0011 ** 
log(CA3)      0.87579    0.03170  27.624  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3319 on 47 degrees of freedom
Multiple R-squared:  0.9604,      Adjusted R-squared:  0.9579 
F-statistic: 380.4 on 3 and 47 DF,  p-value: < 2.2e-16
```

The adjusted R^2 is 95% which indicates good fit. The p value is less than ($2.2e^{-16}$) and a lso all individual coefficients p-values are below the level of significance that rejects the null hypothesis that means the response variable depends significantly on the independ ent variable. (cf. p. 15)

The final model using Eq.4 is

$$Q_{5\%} = e^{-14.08+1.10\ln(\mathrm{AE})+0.674\ln(\mathrm{AP})+0.8757\ln(\mathrm{CA3})}$$

Table 3 gives the result for the Flow5model it shows the prediction errors for the values in the range of the data, in a table all the combination of independent variables in 3 levels are given. The first three columns gives the independent variables, catchment area below 3000 m (CA3), average elevation (AE) and annual precipitation (AP). The Flow5fit column gives the predicted flow value whereas Flow5upr and Flow5lwr give the upper and lower prediction limit. The column Flow5error*100 is gives the relative standard error of the 27 experiments. The prediction level of 68% is used in concordance of standard uncertainty.

Table 3 Result of prediction error of Flow5model

	AE m	CA3 km^2	AP (mm)	Flow5fit (m^3/s)	Flow5upr (m^3/s)	Flow5lwr (m^3/s)	Flow5error*100 (%)
1	911	11	989	1.235	1.8593	0.8201	50.571
2	2887	11	989	4.426	6.6601	2.9417	50.467
3	4863	11	989	7.883	11.982	5.1863	51.995
4	911	9745.5	989	470.87	681.43	325.37	44.718
5	2887	9745.5	989	1687.9	2409.8	1182.2	42.773
6	4863	9745.5	989	3006	4316.8	2093.3	43.604
7	911	19480	989	863.62	1253.4	595.07	45.130
8	2887	19480	989	3095.7	4427.1	2164.8	43.005
9	4863	19480	989	5513.3	7925.4	3835.4	43.750
10	911	11	2365	2.222	3.2082	1.5394	44.364
11	2887	11	2365	7.966	11.430	5.5519	43.485
12	4863	11	2365	14.187	20.541	9.7987	44.786
13	911	9745.5	2365	847.43	1227.9	584.86	44.895
14	2887	9745.5	2365	3038	4318.8	2136.6	42.172

15	4863	9745.5	2365	5410	7717.8	3792.3	42.658
16	911	19480	2365	1554	2268.8	1064.8	45.970
17	2887	19480	2365	5571	7972.4	3893.5	43.094
18	4863	19480	2365	9922	14238	6915.1	43.490
19	911	11	3741	3.0271	4.4052	2.0802	45.525
20	2887	11	3741	10.851	15.653	7.5222	44.256
21	4863	11	3741	19.325	28.092	13.294	45.366
22	911	9745.5	3741	1154.3	1724.5	772.69	49.394
23	2887	9745.5	3741	4137.9	6058.5	2826.1	46.417
24	4863	9745.5	3741	7369.3	10811	5023.3	46.702
25	911	19480	3741	2117.2	3191.7	1404.4	50.750
26	2887	19480	3741	7589.3	11205	5140.5	47.636
27	4863	19480	3741	13516	19982	9142.5	47.837
mean				2849.4	4141.96	1960.74	45.73372

The 68% prediction interval of the flow for 5% probability of exceedance from the 27 experiments lies within the range of 1.859256 to 19981.78 m3/s in the upper bound and 0.82000 to 9142.517 m3/s in the lower bound. The average of the fitted value or the predicted flow value is 2849.4 m^3/s and the average of relative standard error from the 27 different for the flow is 45.7 %, which is below 50% so the model is moderately reliable.

The result of the 100% probability of exceedance is given below. The model uses Eq.5 in linearized form (Eq.5b) and it is called Flow100model. The model uses different inde pendent variable then Flow5model and Flow0model .The annual precipitation (AP) and catchment area below 5000 m (CA5) are independent variables, and Flow100 is the flo w for 100% probability of exceedance for all the stations is a dependent variable.

```
Call:
lm(formula = sqrt(Flow100) ~ sqrt(AP) + sqrt(CA5))

Residuals:
    Min      1Q  Median      3Q     Max
-3.5413 -0.5160  0.0600  0.5433  1.9946

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.133405   1.264956  -1.687   0.0982 .
sqrt(AP)     0.047038   0.026669   1.764   0.0841 .
sqrt(CA5)    0.076521   0.003288  23.271   <2e-16 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.073 on 48 degrees of freedom
Multiple R-squared:  0.9255,      Adjusted R-squared:  0.9224
F-statistic: 298.3 on 2 and 48 DF,  p-value: < 2.2e-16
```

The adjusted R squared is 92% which indicates good fit and the p-value for the model i s (2.2e-[16]) and also all coefficient p – values are below the level of significance, and thu s the response variable depends significantly on the independent variable. (cf. p. 15)

The final Flow100model using Eq.5

$$Q_{100} = [-2.133 + 0.048\sqrt{\mathrm{AP}} + 0.76\sqrt{CA5}]^2$$

Table 4 gives the prediction error in the range of the data for the Flow100model.In a table all the combination of independent variables in 3 levels are given. The first two columns gives the independent variables, catchment area below 5000 m (CA5), and annual precipitation (AP). Flow100fit gives the predicted flow value whereas Flow100upr and Flow100lwr column gives the upper and lower prediction limit. The prediction limit here is 68% in concordance of standard uncertainty. The Flow100error*100 gives column gives the relative standard error of all nine experiments.

Table 4 Prediction errors result for the Flow100model

	AP mm	CA5 km^2	Flow100fit (m^3/s)	Flow100upr (m^3/s)	Flow100lwr (m^3/s)	Flow100error * 100(%)
1	989	17	0.1147	0.6940	2.2810	505.19
2	2365	17	0.2205	2.4706	0.4002	1020.2
3	3741	17	1.1217	4.9655	0.0121	342.66
4	989	19645	101.43	125.72	79.738	23.954
5	2365	19645	118.36	144.46	94.860	22.050
6	3741	19645	131.54	161.78	104.42	22.994
7	989	39273	210.55	246.24	177.66	16.950
8	2365	39273	234.66	273.11	199.13	16.385
9	3741	39273	253.07	297.17	212.51	17.427

These result shows that the range of the relative standard error is 17% to 24% excluding the relative standard error of very low predicted flow value. This model is sufficiently reliable if the predicted flow value is above 100m^3/s When the predicted flow value is below 100m^3/s than the relative standard error is high though the absolute error is moderate than the model is not good model.

The result of the 0% probability of exceedance is given below. The model uses Eq.6 in linearized form (Eq.6b) and it is called Flow0model. The model uses different independent variables then the previous two models Flow5model and Flow100model. Flow0 is the flow for the 0% probability of exceedance that is the dependent variable and the average Elevation (AE) and catchment area below 3000 m (CA3) are the independent variables.

```
lm(formula = sqrt(Flow0) ~ sqrt(AE) + sqrt(CA3))

Residuals:
   Min     1Q Median     3Q    Max 
-9.698 -3.694 -1.589  2.652 25.696 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -12.80420    4.31501  -2.967  0.00467 ** 
sqrt(AE)      0.36616    0.08415   4.351 7.05e-05 ***
sqrt(CA3)     0.52910    0.02938  18.010  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.744 on 48 degrees of freedom
Multiple R-squared:  0.8908,	Adjusted R-squared:  0.8863 
F-statistic: 195.8 on 2 and 48 DF,  p-value: < 2.2e-16
```

The adjusted $R^2$is 86 % which indicates moderate good fit. The p-value for the model is (2.2e-^{16}) and all the individual coefficient p- values are below the level of significance th at rejects t-he null hypothesis, and that means the response variable depends significan tly on the independent variable. (cf. p. 15)

The relative standard error of 0% probability of exceedance is given in Appendix N in p age 9. The relative standard error value is between 18% to 31% excluding the very hig h relative standard error. The reason for the very higher relative standard error is the ve ry low predicted flow values. The model is reliable only if the predicted flow value is abo ve 2000 m^3/s

The final Flow0model using Eq.6

$$Q_{0\%} = [-12.80 + 0.036\sqrt{\mathrm{AE}} + 0.52\sqrt{CA3}]^2$$

5.3.2 Results for the other probability of exceedance

The minimum, maximum and mean value for the predicted flow value and relative standard error for all the specified probability of exceedance is given in Appendix O in page 9. The first three columns gives the minimum, maximum and mean value for the predicted flow value and the rest of the three columns gives the minimum, maximum and mean value for the relative standard error. The appendix O shows that the minimum relative standard error for other probability of exceedance beside 0%, 5% and 100% are above 50% that concludes models are not reliable for predicting the values for ungauged station. The R script and the results for all the specified probability of exceedance is given in Appendix S from page 18 to 23.

5.4 Results for the flow duration curve

Figure 7 is the flow duration curve is plotted data calculated using regional method for an ungauged site that has constant and coefficient values for calculating the flows for specified probability of exceedance. The independent variable used in the equations depends upon the exceedance of probability. The Eq.4 Eq.5 and Eq.6 were used to get these values. The values for the independent variable and the result are given in Appendix L in page 8.

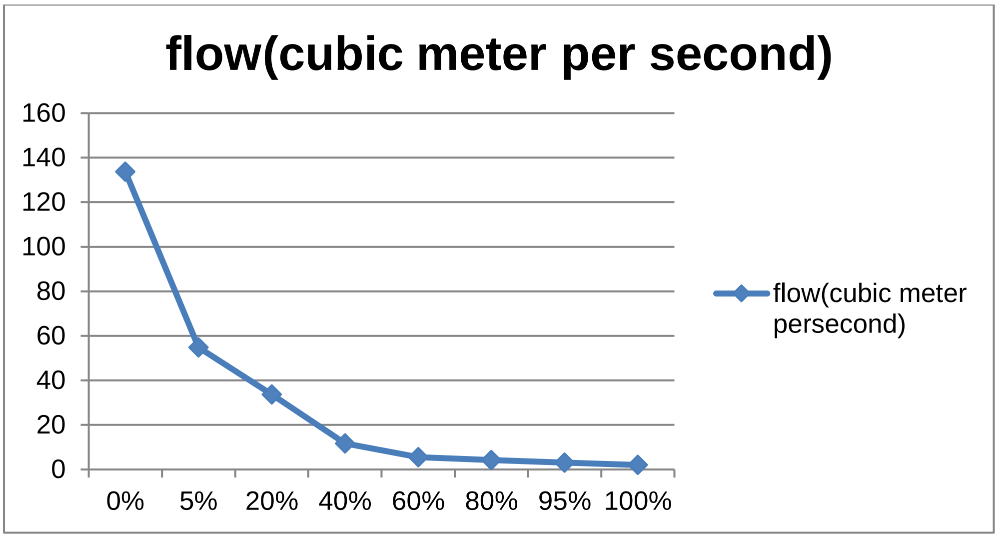

Figure 8 Flow duration curve using regional regression method

In the other hand the Figure 8 is the flow duration curve developed using the flow data of the mean monthly flow data of Midhim Khola (Midhim River) the flow duration curve can be plotted using the 'fdc' function of 'hydroSTM' Package. (Bigiarini, 2015)The mean monthly flow data is given in the Appendix K in page 8.

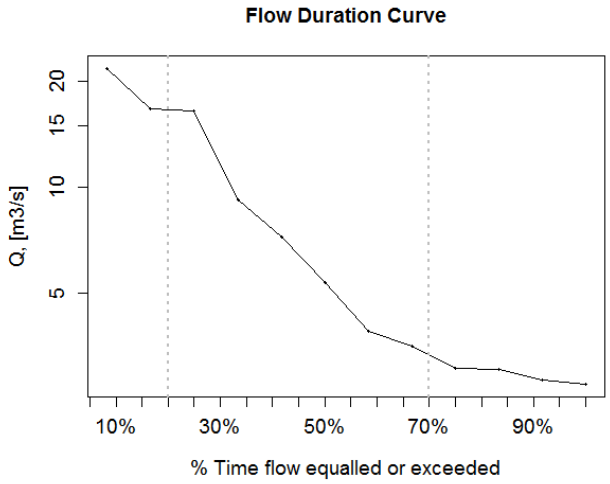

Figure 9 Flow duration curve of using data of drainage area method

5.5 Results for the regional method for low flow

5.5.1 The results are for the 2 year 1 day low flow

The result of the 2-year 1-day low flow is shown as an example. The model uses the Eq.7 in linearized form (Eq.7b) and it is called dayone2yrmodel.The dependent variable is the low flow that is 1-day mean minimum flow that occurs on average once in a 2 years .The catchment area below 5000 meter is only the independent variable in regional regression of low flow analysis.

```
Call:
lm(formula = sqrt(dayone2yr) ~ sqrt(CA5), data = daka)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.8092 -0.4454  0.1050  0.7659  1.9593 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.214351   0.252525   0.849      0.4    
sqrt(CA5)   0.081496   0.003326  24.501   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.169 on 49 degrees of freedom
Multiple R-squared:  0.9245,	Adjusted R-squared:  0.923 
F-statistic: 600.3 on 1 and 49 DF,  p-value: < 2.2e-16
```

The adjusted R square is 92%, which indicates model fits the data. The p- value is (2.2 $e^{-16,}$) and all individual coefficients p- values are below the level of significance, so null h ypothesis is rejected that means the response variable is significantly depends on the in dependent variable (cf .p. 15).

The final dayone2yrmodel using Eq. 7 is

$$Q_{1.2} = [0.214 + 0.081\sqrt{\text{CA5}}]^2$$

Figure 10 is the plot for the actual versus predicted values and it shows that the strong correlation between the actual and the model is an accurate model.

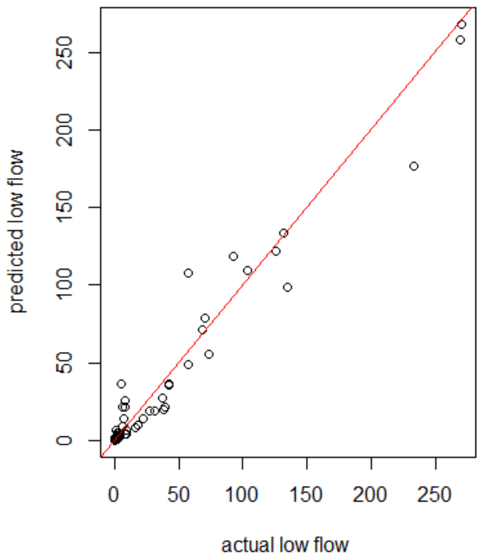

Figure 10: Actual versus predicted flow value for 2 year 1 day low flow

The green lines in Figure 11 the upper and lower bound of 68 % prediction interval whereas the blue is fitted line. The figure shows that the predicted flow data are close to regression line.

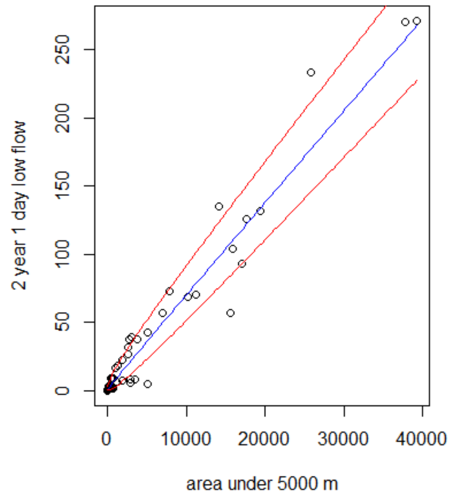

Figure 11:2 year 1 day low flow versus catchment area below 5000 m

The result for prediction error of the low flow is given in Appendix Q in page 10. The table has the minimum, maximum and mean flow values in m^3/s unit for predicted flow. The prediction limit is 68% in concordance with standard uncertainty. The table also contains minimum, maximum and the mean of relative standard error. The minimum relative standard error beside 10 years 30 days flow for all the low flow is below 18% whereas the maximum relative standard error is around 1600 %. The huge relative standard error is reasonable due to the extremely small predicted flow values which are around 0.3 m^3/s and the small absolute error, if these huge maximum relative standard error are excluded than all the models excluding 10 years 30-day model have very low relative standard error. For example, the maximum predicted flow value for 2 years 1-day low flow model is 340 m^3/s that gives the minimum relative standard error value of 16 %, it concludes that the predicted flow must be above 340 m^3/s so the model is sufficiently reliable. The results shows that all the models beside 10 year 30days model is sufficiently reliable The R script for the regression analysis is given in Appendix U in page 23 to 28.

5.6 Flood analysis results

The regression model for the 2-year flood model and 100-year flood model uses Eq.8 in the linearized form Eq.8b and it is called flood2model and flood100model respectively. The flood2model uses the 2-year flood data and flood100model uses the 100 year flood data obtained from Generalized Extreme Value (GEV) distribution as the dependent variable or response variable and the catchment area under 3000 m as the independent variable for both the model . The *Hydrological estimation in Nepal* (Sharma & Adhikari, 2004) mentions that the long term data from the 51 hydrological station were fitted using consolidated frequency analysis package (Pilon & Harvey, 1993) in Generalized Extreme Value (GEV) distribution to obtain the 2-year and 100-year flood. The R script for the regression analysis of flood data is given Appendix V in page 29

The result is of 2-year flood is given below

```
Call:
lm(formula = log(yr2) ~ log(CA3), data = flood)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.95111 -0.34213 -0.01421  0.28281  1.34967 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.75825    0.34455   2.201   0.0327 *  
log(CA3)     0.86570    0.04888  17.711   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5411 on 47 degrees of freedom
Multiple R-squared:  0.8697,	Adjusted R-squared:  0.8669 
F-statistic: 313.7 on 1 and 47 DF,  p-value: < 2.2e-16
```

The adjusted R squared value is 86%, which indicates moderately good fit. The p- value is ($2.2e^{-16,}$) and all individual coefficients p- values are below the level of significance, and thus the response variable depends significantly on the independent variable. (cf. p . 15)

The result of the 100-year flood is given below

```
Call:
lm(formula = log(yr100) ~ log(CA3), data = flood)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.5031 -0.4281 -0.1025  0.3595  2.0504 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.62912    0.46571   7.793 5.20e-10 ***
log(CA3)     0.64712    0.06607   9.795 6.23e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 0.7314 on 47 degrees of freedom
Multiple R-squared:  0.6712,   Adjusted R-squared:  0.6642
F-statistic: 95.94 on 1 and 47 DF,  p-value: 6.226e-13
```

The adjusted R squared value is 66 %, which indicates it is not a good fit. The p- value i s ($2.2e^{-16,}$) and all individual coefficients p- values are below the level of significance, an d thus the response variable depends significantly on the independent variable. (cf. p. 1 5)

.

The flood2model and flood100model uses Eq.8 in linearized form (Eq.8b) and the coefficients of the Eq.8b b_0 and b_1 are obtained from the exponential of the intercept and slop of the log and they are 2.13 and 0.86 for the flood2model, and they are close to coefficients of the original values 2.29 and 0.86 for 2-year flood analysis. However, the coefficients b_0 and b_1 obtained from the flood100model are 37.67 and 0.64 that are not close enough compare to the original value 20.7 and 0.72. The average relative standard error for flood2model is 75% and the average relative standard error for flood100model is 113% in concordance of standard uncertainty. The relative standard errors for both the models are above 50% so the models are not a good model

There are few possible reasons for such a huge difference coefficients values for flood100model. The first reason can be the stations that were selected for the flood analysis. According to the *Hydrological estimation in Nepal* report (Sharma & Adhikari, 2004) the long term data from the 51 hydrological stations excluding the stations that have data less than 10 years of rivers in Nepal were used to compute the flood frequency for each station. The period of hydrological record table in report (Sharma & Adhikari, 2004) contains the information on the stations used and the length of the record for each station. However, the period of hydrological record table is missing the station number 439.7 that was used in all the hydrological analysis for mean monthly flow, flow duration and low flow analysis. The same table incudes the new station 602.5 that was used for any hydrological analysis. This shows that there is some mistake in regarding the recording of the data for the stations.

The period of hydrological record table in the report (Sharma & Adhikari, 2004) shows that the station number 627.5 is only the station that has data less than 10 years whereas the report (Sharma & Adhikari, 2004)mentions that only 49 stations out of 51 stations are included for the flood frequency analysis.

The report (Sharma & Adhikari, 2004) also mentions that the consolidated frequency analysis package (Pilon & Harvey, 1993) was used to fit the long-term data from the 51 hydrological stations in the GEV distribution. The consolidated frequency analysis package has used the maximum flow values of the river for each station. The maximum flow values of the rivers are not available in the report (Sharma & Adhikari, 2004) so, it is not possible to compute the flood data for 2 year and 100 -year return period. It would have been possible to compute the flood values if the maximum flow values were available and compare it with exiting flood data that are fitted in the GEV distribution that are available in report (Sharma & Adhikari, 2004).

6 Conclusion

Hydrological studies are very important for hydropower. There are a number of ungauged sites in Nepal, and the only option for hydrological studies for ungauged sites is to select a similar donor catchment area. The results hydrological study conducted for this thesis show, that the quality of data is poor and results of the regional regression methods show that the models for low flow analysis are only sufficiently reliable in concordance of the standard uncertainty.

The regional regression method uses coefficients developed from historical data for average elevation, area below 3000 and 5000 meters and annual precipitation. The mean monthly flow values obtained from regional regression method using Eq.2 and Eq.3 are double the values given by drainage area ratio method. The drainage area ratio area method using Eq.1 uses area of the donor and the ungauged catchment area and stream flow values of the donor catchment area as its parameters. The regional method and drainage area ratio method uses different parameters so each method gives systematically different results.

The regional regression model is a local model developed using the local data so regional regression model is only for the ungauged sites in Nepal. The relative standard error is small for the March, April and May using Eq.3 and they are sufficiently reliable, whereas the relative standard error of the model is below 50% for the other months using Eq.2 and they are moderately reliable beside June and July. Though the relative standard error of models is moderate, the user must take into consideration regarding the failure of the model and must be aware regarding the consequences that may occur while using this model.

The relative standard errors for models using Eq.4 for flow duration are above 50% that concludes the model is not a good model to predict the flow data. The user must find other alternative to calculate the flow duration. However, the results of the model can be used to get the information order of magnitude of the flow .The model used for the low flow analysis gave the small relative standard error below 25% that concludes model is sufficiently reliable for predicting low flow values. The relative standard error of the model using Eq.8 for 2-year flood and 100-year flood are above 75% and 100% respectively. These models are not the good model to predict the flood data yet the results from these models can be used to get information regarding the order of magnitude of the flood. There are other formulas for the flood analysis such as gumble distribution and, log normal distribution that are reliable than the regional regression method.

The data from 51 hydrological stations shows that the range of elevation, precipitation and catchment areas is huge. In addition, there are a number of ungauged sites in different parts of the country thus using these formulas is not a wise option for hydrological studies. These methods are not much reliable in high Himalayan region because the hydrology of the Himalayan region is affected by various other factors rather than elevation such as snow and glaciers. The elevation of the Terai region is very low, and the hydrology is affected by the interaction between surface water and ground water thus the result will be different. The flood type is different between theTerai, region (landform with low elevation) and the mountain (landform with high elevation). The flood in the Terai region is static; during the monsoon, most of the part of the Terai region is affected by flood. There should is a need for further study regarding the effect such floods and ground water on hydrology.

There are many rivers above 3000 m; permanent snow and melting of snow will affect the hydrology of these rivers. The study of snow and ice hydrology will help to generate good quality data. Each year data must be updated for all the hydrological stations.

References

Austrailan Health Minsters' Council, 2014. *Aboriginal and Torres Strait islaner Health Performance Framework 2014 Report,* s.l.: Department of the Prime Minister and Cabinate.

Bell, S., 1999. *A beginner's guide to uncertainty of measurement.* [Online]
Available at:
https://www.wmo.int/pages/prog/gcos/documents/gruanmanuals/UK_NPL/mgpg11.pdf
[Accessed 2016].

Bigiarini, M. Z., 2015. *CRAN.* [Online]
Available at: https://cran.r-project.org/web/packages/hydroTSM/hydroTSM.pdf
[Accessed 2016].

BPC Hydroconsult;Practical Action, 2002. *Civil works guidelines for micro hydropower in Nepal,* s.l.: Practical Action, Nepal.

Emerson, D. G., Vecchia, A. V. & Da, A. L., 2005. *Evaluation of Drainage-Area Ratio Method Used to Estimate Streamflow for the Red River of the North Basin, North Dakota and Minnesota,* Virginia: U.S. Geological Survey.

Environment, G. o. N. M. o. P. &., 2006. *Department Of Hydrology and Meteorology.* [Online]
Available at: http://www.dhm.gov.np/
[Accessed 19 November 2016].

H.M, R., 2005/2006. *Hydrology Principles Analysis Design.* s.l.:NEW AGE INTERNATIONAL PUBLISHERS .

Harlad , M. & Tormod, N., 1989. *Multivariate Calibration.* Chichester: Wiley.

Jarman, K. H., 2013. *The Art of Data Analysis: How to Answer Almost Any Question Using Basic Statistics.* 2nd ed. s.l.:Wiely.

Oregon State University, 2005. *Flow duration analysis.* [Online]
Available at: http://streamflow.engr.oregonstate.edu/analysis/flow/
[Accessed 21 November 2016].

Pilon, P. J. & Harvey, D. K., 1993. *CONSOLIDATED FREQUENCY ANALYSIS(CFA),* OTTAWA,ONTARIO: s.n.

Ralph, W. A. & Wesley, J. P., 2001. *Water resource engineering.* 1st ed. s.l.:Pearson.
Risley, J., Stonewall, A. & Haluska, T., 2013. *Scientific Investigations Report 2008-5126,* s.l.: U.S.GEOLOGICAL SURVEY.

Roterman-konieczna, i., 2009. *Statistics by prescription.* s.l.:Jagiellonian University Press.
Sharma, K. P. & Adhikari, N. R., 2004. *Hydrological estimation in Nepal,* Kathmandu: Deparment of Meterology and Hydrology.

Yadav, S. K., 2002. *Hydrological analysis for bheri-babi hydropower project Nepal.* [Online]
Available at: http://www.ub.uib.no/elpub/Norad/2002/ntnu/thesis01.pdf
[Accessed December 2016].

Yan ,, Xin, S. & Xiao Gang, 2009. *Linear regression analysis:theory and computing.* 1st ed. s.l.:World Scientfic.

Appendices

Appendix A Mean monthly flow m^3/s for 51 hydrological stations in Nepal

stn .no	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
120	20.5	18.9	19	21.3	27.4	50	101	129	98.4	51.1	31.5	24.2
170	1.41	1.36	1.54	1.13	1.17	11.6	22.8	33.7	17.7	6.07	2.63	1.79
240	132	117	134	204	398	740	1190	1390	929	419	235	166
250	151	135	148	215	394	734	1390	1710	1130	488	263	186
260	74.8	70	77.1	91.6	132	281	726	987	674	252	125	89.3
270	99.3	85.7	82.1	103	162	351	973	1330	987	377	183	123
280	350	311	329	435	738	1450	3180	4260	2910	1240	624	429
286	4.71	4.07	3.41	2.78	3.42	6.93	27.4	45.4	38.5	18.3	7.71	5.06
290	17.6	14.7	12	10	15.5	53.7	224	225	233	84.5	32.7	21.2
330	16.6	13.9	12	10.1	10.6	33.1	123	220	176	72.6	29.9	20.4
339.5	6.09	5.14	4.67	3.74	4.24	18.6	67.6	90.3	73.4	30.3	12.2	7.4
350	27.5	22.3	18.6	14.9	17.9	61.4	213	347	295	103	50.1	33.9
360	30.4	25.3	20.7	15.9	18	111	348	475	395	155	62.1	36.5
406.5	10.7	8.64	8.82	11.9	20.9	52.2	147	165	96.2	42.9	20.6	14.6
410	59.4	50.1	50.4	66.8	106	291	766	875	609	257	126	80.7
415	4.75	3.86	3.18	2.97	6.32	33.4	100	94.3	67.3	26.8	9.89	6.15
420	111	90.5	81.2	89.8	135	367	1164	1426	1035	457	227	150
428	3.38	2.89	2.78	2.8	3.94	14.7	48.2	60	42	17.5	7.13	4.35
430	12.9	11.4	11.3	13	19.1	49.8	130	147	103	54.5	25.3	16.9
439.3	3.89	3.53	3.61	3.92	4.98	9.24	21.4	27.9	21.3	12.7	7.88	4.85
439.7	52.4	45.5	45.7	55.8	105	254	520	665	424	169	95.3	65.6
440	5.69	4.66	4.31	4.56	6.14	20	61.5	72.3	57	26	12.5	7.75
445	35.7	31.2	36.2	60.6	106	223	394	422	310	160	84.2	51.1
446.8	2.48	2.03	1.81	1.96	2.85	12.1	35.8	43.8	30.7	12	5.45	3.35
447	50.1	44.3	46.1	57.4	98.8	250	526	594	397	174	95	64.5
448	9.38	7.11	5.42	5.65	9.2	34.7	99	130	94.4	42.1	21.7	13.1
450	372	302	282	362	620	1717	3929	4604	3332	1628	828	504
460	9.33	7.9	6.62	6.61	7.77	21.2	59.9	74.9	64.9	30	15.5	11
465	5.99	5.09	4.69	5.92	6.29	15.2	49.6	63.5	56.4	19.5	10.3	7.28
470	1.91	1.57	1.43	1.39	1.89	5.69	21.6	30	24.7	9.62	4.19	2.61
505	0.33	0.27	0.25	0.25	0.31	0.92	2.74	3.88	3.14	1.38	0.66	0.44
539	20.4	15.6	12.5	12.6	28.3	110	415	416	351	121	43.9	26.3
600.1	69.7	71.5	87.8	105	171	437	716	780	566	238	116	82.9
602	6.09	5.14	4.76	6.75	16	31	53	56.4	50.9	24.6	12.6	8.18
604.5	108	108	127	158	257	618	1080	1140	869	398	199	135
606	167	156	197	229	345	718	1370	1710	1230	604	348	242
610	24.2	21.4	20.6	25	37.5	88.3	183	258	163	77.7	41.7	28.9
620	12.3	10.6	10.1	11.3	15.9	48.9	129	160	114	49.2	23.9	16
627.5	3.36	2.97	2.77	3.01	4.71	10	22.5	29.2	22.4	12.1	7.23	4.32
630	57.9	48.9	46.4	53	78.1	216	605	762	504	220	116	75.2
640	1.31	1.13	1.02	0.92	0.98	1.81	4.57	5.94	4.93	3.4	2.03	1.58
647	30.5	25.5	24.3	29.9	53	167	427	461	307	130	62.6	40.5
650	6.14	5.25	4.76	5.26	8.66	39.6	97.7	98.8	54.4	24.5	12.2	8.19
652	109	91.7	84.1	92.7	144	436	1210	1570	1020	462	225	144
660	14.6	12.1	11.1	12.5	17.5	50.5	137	158	111	55	28.6	19.6
670	44.1	37.1	35.9	41.6	69.6	237	544	590	409	192	90.3	57.6
680	193	164	150	162	235	790	2380	2760	1950	886	384	248
690	67.6	55.3	53.6	78.7	172	444	822	888	667	325	151	91.7
695	406	348	353	440	784	1890	3890	4390	3410	1630	867	555
728	6.31	5.54	5.55	6.63	10.9	25.9	52.5	49.8	47.4	20.6	10.2	6.88
795	11.6	9.51	8.59	10.2	18.5	61.6	165	161	114	51.7	24.3	16.1

Appendix B Data of area, average elevation, annual precipitation and area below 3000 and 5000 m

Area	Averege Elevation	Annual precipitation	Catchment area under 3000 m	Catchment area under 5000 m
1175	3073	1883	598	1095
190	1831	1508	190	190
21438	4122	989	4080	15830
23229	3917	1033	5811	17620
7386	2532	1641	5185	6963
13309	3080	1409	6551	11258
45857	3331	1289	19480	37775
811	1452	1341	811	811
2924	964	1497	2924	2924
1924	1799	1660	1868	1924
662	1656	1572	501	662
3527	1571	1643	3464	3527
5072	1236	1639	5009	5072
642	3064	3175	364	542
7109	3812	1324	2273	5093
418	1197	2685	418	418
12234	2821	1631	7026	10221
124	2382	3741	93	124
590	2786	3565	355	538
133	2586	2048	94	133
3937	3812	2607	1230	2745
309	1905	1791	258	309
3968	4245	1687	723	2655
154	2163	1849	121	154
4643	4358	1702	693	3010
630	1725	2196	552	630
32099	3041	1878	16165	25802
471	970	2012	471	471
426	1174	1909	426	426
169	911	1942	169	169
17	2060	2174	17	17
2922	1058	1749	2922	2922
25447	4863	2366	445	15597
409	1586	2122	373	410
27241	4734	2299	1457	17136
29532	4453	1947	3695	19427
2388	4586	3028	295	1342
594	3346	2658	254	513
113	4166	2219	11	101
4904	3417	2389	2122	3743
69	1849	1562	50	69
2948	4183	1975	732	1864
330	2785	1892	222	326
10141	3235	2003	5011	7892
921	2957	1592	576	812
3650	3810	1625	1423	2597
17593	3010	1779	9884	14182
5948	2883	1668	3664	5048
53689	3733	1785	17859	39273
404	1664	2519	400	408
1178	1242	2321	1016	1182

Appendix C Low flow Frequency m^3/s for 1 day, 7 days and 30 days

		1 day			7 day			30 day	
stn .no	2yr	10yr	20yr	2yr	10yr	20yr	2yr	10yr	20yr
120	16.73	13.46	12.61	17	13.5	12.6	17.5	13.8	13
170	0.52	0.31	0.27	0.572	0.327	0.283	0.703	0.433	0.395
240	104	90.2	87.5	108	93.3	90.4	113	98.3	95.2
250	126	104	97.5	125	103	96.4	132	108	101
260	57.2	38.6	32.8	59.5	43.1	38.7	63.6	46.5	42.4
270	70.7	56.3	51.6	73.7	59.1	54.3	77.1	62.2	57.5
280	270	219	201	276	225	208	290	234	219
286	1.2	0.82	0.75	1.34	0.93	0.86	1.79	1.29	1.23
290	6.07	3.96	3.68	6.47	4.17	3.82	7.55	4.88	4.38
330	7.27	4.83	4.1	7.61	4.96	4.23	8.72	5.99	5.31
339.5	1.81	0.87	0.69	2.09	1.01	0.808	3.06	1.88	1.61
350	8.27	5.08	4.5	9.23	6.08	5.54	13	8.95	7.9
360	4.88	1.66	1.24	6.02	2.2	1.61	10.1	4.83	3.66
406.5	8.13	6.78	6.54	8.23	7.02	6.81	8.46	7.34	7.18
410	42.8	30.9	27.2	44.1	35.5	33.6	46.7	38.3	36.5
415	1.81	0.95	0.65	1.98	1.08	0.774	2.43	1.41	1.02
420	69	53.2	50	71.2	54.8	51.3	75.4	59.8	56.9
428	1.93	1.34	1.24	2.01	1.37	1.25	2.24	1.55	1.42
430	9.28	7.32	6.99	9.58	7.68	7.37	10.2	8.14	7.77
439.3	2.96	2.11	1.79	3.03	2.26	1.96	3.28	2.39	2.08
439.7	37.9	31	29.5	38.9	31.9	30.5	40.6	34.1	33
440	3.3	2.33	2.12	3.45	2.47	2.26	3.84	2.7	2.54
445	27.1	22.9	22.1	28	23.6	22.7	30	24.9	23.8
446.8	1.27	0.52	0.36	1.29	0.573	0.427	1.5	0.789	0.65
447	39.6	35	34.1	40.8	35.5	34.4	42.7	37.2	35.8
448	3.32	1.92	1.66	3.65	2.12	1.79	4.54	251	1.98
450	233	180	168	248	191	177	267	208	192
460	4.36	2.91	2.61	4.73	3.22	2.92	5.37	3.79	3.5
465	2.97	1.88	1.75	3.12	2.06	1.93	3.63	2.71	2.61
470	0.8	0.56	0.51	0.89	0.6	0.54	1.04	0.763	0.723
505	0.15	0.07	0.05	0.165	0.078	0.0595	0.181	0.096	0.083
539	7.97	5.68	5.37	9.36	6.43	5.86	10.4	7.59	7.28
600.1	57.25	53	52.5	62.8	57.1	56.2	69.1	59.2	55.9
602	3.54	2.2	1.84	3.77	2.39	2	4.14	2.69	2.31
604.5	92.9	75.1	71.6	96.6	77.6	73.9	102	81.9	77.8
606	132	104	98.2	136	111	107	148	111	104
610	18.2	12	10.4	18.8	12.3	10.6	19.7	12.7	11
620	9.28	7.13	6.62	9.47	7.31	6.78	9.77	7.58	7.08
627.5	2.3	1.6	1.32	2.38	1.69	1.4	2.68	1.93	1.58
630	37.8	30.8	30	39.4	32.1	31.2	42.2	34.9	33.9
640	0.49	0.213	0.17	0.567	0.266	0.216	0.707	0.428	0.385
647	22.2	19.5	18.7	22.9	20	19.2	23.8	21	20.1
650	4.02	3.05	2.88	4.18	3.24	3.09	4.4	3.48	3.35
652	73.2	57.5	53	75	60.3	56.4	79.4	64.7	60.6
660	9.45	7.21	6.58	9.95	7.68	7.02	10.8	8.34	7.55
670	31.8	22.9	20.8	32.8	24.1	22.2	34.6	25.4	23.4
680	135	109	102	138	112	106	144	120	114
690	42.3	29.8	27.6	44	31.2	28.9	47.2	33.8	31.3
695	271	190	167	285	212	193	307	231	210
728	3.07	2.38	2.29	3.334	2.53	2.4	3.73	2.95	2.84
795	6.49	5.06	4.85	6.88	5.28	5.01	7.4	5.68	5.41

Appendix D Flow duration m^3/s for the specified probability of exceedance

stn .no	0%	5%	20%	40%	60%	80%	95%	100%
120	227	141	86.7	37.8	37.8	20.2	16.1	12.1
170	73.4	28	11.4	3.15	3.15	1.14	0.699	0.35
240	1915	1411	958	403	403	136	111	90.7
250	2316	1737	1042	434	434	156	127	92.6
260	1308	1020	570	180	180	75	60	36
270	1775	1391	818	245	245	94.1	77.7	45
280	5698	4352	2720	979	979	354	272	204
286	87.4	63	22.4	6.72	4.48	2.94	1.96	0.84
290	678	365	138	38.2	20.3	11.4	7.31	4.06
330	453	255	106	30.4	16.4	11.6	8.51	6.08
339.5	173	103	56.7	13	6.48	5.13	2.97	2.16
350	536	380	190	55	29	18	12	8
360	939	536	282	70.5	31	19.7	11.3	2.82
406.5	241	191	93.1	29.5	16	10.3	7.76	6.2
410	1110	893	586	181	83.7	58.3	44.6	33.5
415	218	111	59.8	15	5.68	4.19	2.69	0.299
420	2290	1503	884	265	133	92.8	75.1	53
428	121	59.5	33.3	9.63	4.2	2.98	1.93	1.23
430	188	158	104	34.7	16.3	11.9	9.41	7.92
439.3	52.8	28.4	20.5	8.93	4.94	3.78	3.05	2
439.7	853	639	391	155	65.9	49.4	43.3	37.1
440	132	71.9	48.1	13.9	6.73	4.64	3.48	2.55
445	541	445.2	302	135	66.8	39.8	30.2	22.3
446.8	81.4	43.9	28.4	7.1	3.23	1.94	1.29	0.516
447	906	580	360	148	68	50	42	36
448	190	142	86.7	25.2	11.4	7.09	4.33	1.58
450	6849	4638	3092	1237	510	340	263	186
460	168	82.6	45.2	16.8	10.1	6.97	5.16	2.84
465	289	75.6	27.9	10.5	6.77	4.58	3.18	2.39
470	950	34.9	15.2	4.66	2.24	1.43	0.986	0.627
505	5.9	4.18	2.58	0.75	0.406	0.234	0.16	0.0738
589	1018	473	284	66.2	27	16.2	10.8	6.75
600.1	1124	792	566	198	108	79.24	67.92	53.77
602	85.6	66.7	42.6	20.7	8.97	5.75	4.37	3.45
604.5	1797	1212	866	346	165	117	95.3	77.9
606	2356	1781	1250	500	269	194	144	113
610	325	237	155	61.4	31.1	22.1	17.2	7.36
620	511	165	100	30.1	15.5	11.5	9.02	6.51
627.5	67.4	27	19.2	8.84	4.37	2.91	2.29	1.56
630	633	742	464	151	76.6	51	41.8	32.5
640	11.6	7.75	3.75	2.03	1.3	0.975	0.725	0.45
647	632	485	294	95.6	41.2	27.9	23.5	19.1
650	112	110	54.9	16.8	7.93	5.49	3.97	2.44
652	4408	1584	932	303	135	97.9	74.6	60.6
660	259	172	113	37.6	18.3	12.9	10.2	7.52
670	830	601	407	146	58.2	40.7	29.1	21.3
680	3810	2831	1973	515	232	172	146	120
690	1380	954	620	254	105	66.8	44.5	28.6
695	7378	4452	3180	1352	572	398	318	239
728	81.1	67.2	33.6	14.7	7.56	5.67	4.2	3.36
795	376	184	93.4	29.7	14.6	9.72	7.02	4.86

Appendix E Two year and hundred year flood data

stn .no	yr2	yr100
120	222	525
170	144	1050
240	2260	3860
250	3190	8870
260	2900	10100
270	2680	7120
280	8860	20300
286	272	1330
290	2080	10400
330	617	1390
339.5	318	2520
350	1430	17000
360	2700	10100
406.5	642	3010
410	1880	3650
415	509	1500
420	4500	9200
428	134	676
430	290	7170
439.3	60.9	179
440	263	799
445	715	2900
448.8	146	681
447	1020	3720
448	512	1670
450	9360	15100
460	535	1260
465	479	4080
470	309	1080
505	9.77	197
589	3610	37100
600.1	1350	1580
602	239	1000
604.5	2670	6110
606	3820	6110
610	451	3420
620	610	2350
630	1840	6200
640	27.5	261
647	1030	1750
650	528	9660
652	3350	11700
660	369	743
670	1520	3620
680	6020	13200
690	2520	6720
695	7270	21200
728	170	5780
795	3300	8340

Appendix F Khudi River data used as a donor catchment area

Year	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Annual year
1983	2.68	2.58	2.62	2.63	3.03	3.52	14.1	21.2	20.8	14.9	7.71	4.72	8.37
1984	3.99	3.37	3.31	3.38	4.01	10.8	22.3	19.6	16.8	8.29	4.95	3.94	8.73
1985	3.72	3.47	3.46	3.44	3.62	4.6	17.1	15.8	17.4	13.4	7.72	5.56	8.27
1986	4.88	4.59	4.66	4.71	4.8	7.76	19.4	20.8	20	7.64	4.87	3.51	8.96
1987	2.55	2.03	2.01	3.41	4.06	5.66	21.6	52.5	NA	NA	NA	NA	NA
1988	5.25	5.03	5.05	5.28	5.45	NA	NA	NA	NA	10.4	5.33	4.23	NA
1989	4.1	3.57	3.66	4.26	7.24	12.7	29.2	33.8	25.6	11.9	6.26	3.95	12.2
1990	3.08	2.74	3.25	3.71	4.82	9.28	24.1	27.8	22.4	17.4	9.69	5.31	11.1
1991	4.33	3.84	3.71	3.98	5.34	11.6	18.3	26.9	25.6	9.6	4.16	3.72	10.1
1992	NA	NA	NA	2.85	3	5.77	13.6	25.6	20.6	15.7	8.85	6.32	NA
1993	5.14	4.65	4.54	4.92	5.8	10.1	20.3	30.7	20.9	9.39	5.12	3.97	10.5
1994	3.18	3.06	3.2	3.63	3.78	7.19	24.1	28.5	21.4	10.5	6.02	4.44	9.92
1995	3.76	3.46	3.82	4.7	9.87	21.9	33.1	31.6	21.4	11.9	11.7	4.75	13.5
Average:	3.89	3.53	3.61	3.92	4.98	9.24	21.4	27.9	21.2	11.8	6.86	4.53	10.2

Appendix G Most plausible relationships for the average annual hydrograph

Month	Constant	Coef.of Avg Elv	Coef.of Ann Plann	Coef.ofof A<3k	Coefof.of A<5k
Jan	-16.77	1.36	0.470	0.82	
Feb	-17.200	1.42	0.456	0.814	
Mar	0.384				0.091
Apr	0.18				0.104
May	0.0001				0.136
Jun	-19.5	1.61	0.709	0.872	
Jul	-16.3	1.26	0.759	0.884	
Aug	-14.7	1.24	0.622	0.871	
Sep	-13.7	1.09	0.594	0.872	
Oct	-15.3	1.21	0.600	0.846	
Nov	-16.7	1.36	0.543	0.826	
Dec	-17	1.39	0.504	0.822	

Appendix H Most plausible relationship for different flow duration

Month	constant	Coef.of.Avg Elv	Coeff.of Ann Ptn	Const of A<3k	Const of A<5k
0%	-12.8	0.366		0.59	
5%	-13.6	1.108	0.607	0.874	
20%	-17	1.359	0.716	0.883	
40%	-19	1.554	0.656	0.859	
60%	-18.3	1.535	0.513	0.832	
80%	-19.4	1.589	0.559	0.834	

95%	-21.2	1.732	0.598	0.842	
100%	-2.18		0.048		0.07

Appendix I Most plausible relationship for low flow analysis

Return Period	Day	Constant Cd.T	Coef Fd.Td	Std	r2
	1	0.2144	0.0815	0.0033	0.925
	7	0.2362	0.083	0.0033	0.929
2	30	0.3026	0.0854	0.0031	0.938
	Monthly	0.3397	0.086	0.003	0.94
10	1	0.0859	0.0729	0.0032	0.915
	7	0.092	0.0748	0.0031	0.921
	30	0.1807	0.0766	0.003	0.93
	Monthly	0.2138	0.0777	0.0031	0.94
20	1	0.0698	0.0703	0.0031	0.912
	7	0.0662	0.0726	0.0031	0.918
	30	0.1609	0.0742	0.003	0.927
	Monthly	0.1945	0.0754	0.0031	0.929

Appendix J Table for mean monthly flow for 4 rivers

Month	Khudi river	Mardi river	Chepe river	Modi river
January	3.0384	2.5096	1.9820	1.7918
February	2.7573	2.1458	1.6378	1.4996
March	2.8197	2.0641	1.4940	1.4706
April	3.0619	2.0790	1.5776	2.0017
May	3.8898	2.9254	2.2628	3.3415
June	7.2173	10.915	7.1526	8.7815
July	16.715	35.788	22.828	28.249
August	21.792	44.550	25.369	31.478
September	16.559	31.185	19.319	16.304
October	9.2169	12.994	8.7570	6.7637
November	5.3583	5.2940	4.2114	3.3092
December	3.5383	3.2299	2.5903	2.3891
Annual mean	7.9971	12.973	8.2651	8.9483

Appendix K New and actual data of Midhim River

Months	New data (m^3/s)	Raw data (m^3/s)
Jan.	2.9970	3.0384
Feb.	2.7272	2.7573
Mar.	2.7837	2.8197
Apr.	3.4765	3.0619
May	3.8946	3.8898
Jun.	7.4829	7.2173
Jul.	16.697	16.715
Aug.	20.908	21.792
Sep.	16.737	16.559
Oct.	9.2331	9.2169
Nov.	5.1454	5.3583
Dec.	3.5353	3.5383
Average	7.9682	7.9971

Appendix L Midhim River variable values and flow values from regional method

Catchment area below 3000 m (km)	85.79	Probability of Exceedance	Flow(m^3/s)
Catchment area below 5000 m (km)	103.69	0%	133.73
Annual Precipitation (mm)	3610	5%	54.799
Average Elevation (m)	2666.5	20%	33.670
		40%	11.664
		60%	5.5571
		80%	4.1652
		95%	3.0341
		100%	2.0073

Appendix M Predicted data for Midhim River using two different methods

Month	Regional Regression method	Drainage area ratio method
Jan	4.2992	3.0384
Feb	3.8975	2.7573
Mar	0.1266	2.8197
Apr	0.0367	3.0619
May	0.0000	3.8898
Jun	17.988	7.2173
Jul	44.339	16.715
Aug	57.627	21.792
Sep	38.316	16.559

Oct	18.650	9.2169
Nov	8.6116	5.3583
Dec	5.7689	3.5383
mean	16.638	7.9971

Appendix N Predicted flow and relative standard error for mean monthly flow

Month	Min fitted (m^3/s)	Max fitted (m^3/s)	Mean fitted (m^3/s)	Min relative standard error (%)	Max relative standard error (%)	Mean relative standard error (%)
Jan	0.1067	893.44	193.11	44.800	55.343	48.620
Feb	0.0897	772.77	166.29	46.826	57.933	50.848
Mar	0.5766	340.18	172.52	11.491	418.37	24.196
Apr	0.3717	431.60	217.14	13.761	882.17	33.466
May	0.3159	731.32	365.81	15.913	1895.8	48.095
Jun	0.4359	3065.9	794.96	69.237	75.973	72.792
Jul	1.5183	6345.3	1788.7	60.798	66.564	63.843
Aug	1.9202	7526.8	2138.2	49.286	53.789	51.665
Sept	1.6189	5053.0	1502.9	49.606	54.144	52.004
Oct	0.7149	2226.5	643.68	47.443	51.751	49.720
Nov	0.3218	1146.4	319.23	48.164	52.549	50.481
Dec	0.2077	766.40	212.08	46.373	50.568	48.590

Appendix O Prediction error result of the model Flow0model

	AE (m)	CA3 (km^2)	Flow0fit(m3/s)	Flow0upr(m3/s)	Flow0lwr(m3/s)	Relative standard error
1	911	11	0.00	50.377	50.311846	NA
2	2887	11	74.384	242.38	2.8252	225.84
3	4863	11	209.81	468.59	53.621	123.34
4	911	9745.5	2548.2	3344.5	1860.0	31.249
5	2887	9745.5	3493.0	4375.4	2710.0	25.260
6	4863	9745.5	4220.1	5200.4	3342.0	23.231
7	911	19480	5197.6	6379.4	4136.7	22.739
8	2887	19480	6515.1	7768.3	5372.1	19.235
9	4863	19480	7495.5	8846.5	6256.4	18.024

Appendix P Predicted flow value and relative standard error of probability of exceedance

Exceed-ance Prob-ability	Min fit (m3/s)	Max fit (m^3/s)	Mean fit(m3/s)	Min relative standard error (%)	Max relative standard error (%)
0%	0.00001	7495.5	3306.0	18.02 4	NA
5%	1.2348	13516	2849.5	42.172	51.995
20%	0.5127	10066	1899.5	50.224	62.291
40%	0.1685	3402.1	636.27	53.767	66.856
60%	0.1355	1215.4	295.80	53.951	67.094
80%	0.0671	1004.7	197.62	50.657	62.848
95%	0.0409	855.82	160.72	56.351	70.198
100%	0.1147	253.07	116.78	16.385	1020.2

Appendix Q Prediction errors results for low flow

Low flow	Min fitted (m3/s)	Max fitted (m3/s)	Mean fitted (m3/s)	Min relative standard error (%)	Max relative error (%)	Mean relative error (%)
2 year 1 day	0.303	267.8	135.12	16.189	910.36	38.066
10 year 1 day	0.149	211.04	105.98	17.398	1462	46.271
20 year 1 day	0.129	195.8	98.26	17.866	1616.4	48.771
2 year 7 day	0.334	278.19	140.46	15.609	824.31	36.068
10 year 7 day	0.16	222.64	111.83	16.691	1352.3	43.812
20 year 7 day	0.133	208.63	104.68	17.063	1538.2	46.472
2 year 30 day	0.428	296.53	150.07	14.394	640.33	31.820
10 year 30 day	0.957	228.54	117.78	34.257	993.39	70.256
20 year 30 day	0.217	220.73	111.2	14.395	640.33	34.576

Appendix R R script and result for partial least square

Year 1988

```
DF      <- read.table('AVQ4393b.TXT',header=TRUE)
i88     <- c(1:4,7:9,11:13)
j88     <- c(7:9,13) # June,july,august and december
S       <- DF[i88,-c(j88,1,14)]
U       <- DF[i88,j88]
print(cbind(S,U))
SU <- data.frame(U=I(as.matrix(U)),S=I(as.matrix(S)))
Model1<- plsr(U~S, data=SU,validation='CV')
print(summary(Model1))
Upred <- predict(Model1,newdata=as.matrix(DF[6,-c(j88,1,14)]))[,,1:5]
print(Upred)

Model 1
Data:        X dimension: 10 8
```

```
             Y dimension: 10 4
Fit method: kernelpls
Number of components considered: 8

VALIDATION: RMSEP
Cross-validated using 10 leave-one-out segments.

Response: Jun.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV           5.405    5.304    4.944    3.202    3.045    2.922    5.788    6.439    3.258
adjCV        5.405    5.203    4.677    3.122    2.952    2.807    5.539    6.143    3.092

Response: Jul.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV           6.018    5.860    5.716    5.210    9.888    12.04    7.774    7.815    29.72
adjCV        6.018    5.761    5.522    5.086    9.477    11.50    7.394    7.444    28.21

Response: Aug.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV           6.271    4.495    4.309    4.607    6.637    8.053    8.678    8.317    38.12
adjCV        6.271    4.325    4.175    4.511    6.419    7.756    8.287    7.924    36.18

Response: Dec.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV          0.7245   0.8100   0.8072   0.3830   0.4174   0.4442   0.8157   0.7728    7.162
adjCV       0.7245   0.8115   0.8672   0.3783   0.4076   0.4331   0.7840   0.7445    6.797

TRAINING: % variance explained
      1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
X     32.5532    55.89    96.98    99.30    99.78    99.99    99.99   100.00
Jun.  60.4680    88.32    88.38    92.11    94.02    94.04    97.26    99.97
Jul.  52.3353    67.15    69.94    75.49    84.87    96.84    97.04    97.98
Aug.  70.9443    71.51    74.46    74.72    80.66    91.41    95.79    96.94
Dec.   0.7746     4.90    85.91    87.26    87.26    87.27    89.08    91.91
NULL
>
> Upred <- predict(Model1,newdata=as.matrix(DF[6,-c(j88,1,14)]))[,,1:5]
> print(Upred)
       1 comps   2 comps   3 comps   4 comps   5 comps
Jun.  1.245511 10.645385 10.637774  9.450059 13.664840
Jul. 13.189789 20.822138 20.762892 19.148700 29.535230
Aug. 14.737103 13.184826 13.248387 12.883368 21.495446
Dec.  4.518981  4.034049  3.995578  4.091587  4.073357
```

Year 1992

```
DF      <- read.table('AVQ4393b.TXT',header=TRUE)
i92     <- c(1:4,7:9,11:13)
j92     <- 2:4 # Jan to mar
A      <- DF[i92,-c(j92,1,14)]
B      <- DF[i92,j92]
print(cbind(A,B))
AB <- data.frame(B=I(as.matrix(B)),A=I(as.matrix(A)))
Model2<- plsr(B~A, data=AB,validation='CV')
print(summary(Model2))
Bpred <- predict(Model2,newdata=as.matrix(DF[10,-c(j92,1,14)]))[,,1:2]
print(Bpred)
```

```
Model2
Data:         X dimension: 10 9
              Y dimension: 10 3
Fit method: kernelpls
Number of components considered: 8

VALIDATION: RMSEP
Cross-validated using 10 leave-one-out segments.

Response: Jan.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV           0.821   0.9188   0.7177   0.8224    1.203    1.222    1.227    1.169    1.113
adjCV        0.821   0.8970   0.7026   0.7789    1.152    1.171    1.174    1.109    1.056

Response: Feb.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV           0.725   0.8193   0.6489   0.7711    1.073    1.052   0.8594   0.8653   0.8230
adjCV        0.725   0.8039   0.6361   0.7194    1.023    1.001   0.8189   0.8225   0.7808

Response: Mar.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV          0.6491   0.7068   0.6483   0.7734    1.083    1.077   0.8823   0.9735   0.9500
adjCV       0.6491   0.6894   0.6375   0.7260    1.032    1.026   0.8432   0.9313   0.9017

TRAINING: % variance explained
      1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
X       44.95    79.48    81.78    92.86    98.02    99.63    99.89    99.95
Jan.    38.21    66.87    78.62    81.16    85.71    91.14    99.86   100.00
Feb.    31.34    62.38    80.59    84.31    91.79    96.92    99.07    99.96
Mar.    35.37    47.04    67.52    75.41    84.79    91.71    93.70    99.78
NULL
>
> Bpred <- predict(Model2,newdata=as.matrix(DF[10,-c(j92,1,14)]))[,,1:2]
> print(Bpred)
      1 comps  2 comps
Jan. 3.224791 3.083950
Feb. 3.004258 2.874845
Mar. 3.120058 3.049026
```

Year 1987

```
DF      <- read.table('AVQ4393b.TXT',header=TRUE)
i87     <- c(1:4,7:9,11:13)
j87     <- 10:13 # September to December
X       <- DF[i87,-c(j87,1,14)]
Y       <- DF[i87,j87]
print(cbind(X,Y))
XY <- data.frame(Y=I(as.matrix(Y)),X=I(as.matrix(X)))
Model <- plsr(Y~X, data=XY,validation='CV')
print(summary(Model))
Ypred <- predict(Model,newdata=as.matrix(DF[5,-c(j87,1,14)]))[,,1:6]
print(Ypred)

Model

Data:         X dimension: 10 8
              Y dimension: 10 4
Fit method: kernelpls
```

```
Number of components considered: 8

VALIDATION: RMSEP
Cross-validated using 10 leave-one-out segments.

Response: Sep.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV           3.057    2.959    2.521    2.981    3.607    4.611    6.249    11.09    39.98
adjCV        3.057    2.914    2.457    2.918    3.499    4.447    5.995    10.57    37.96

Response: Oct.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV           3.231    3.441    4.610    4.162    3.350    3.031    4.782    3.272    3.346
adjCV        3.231    3.414    4.544    4.067    3.274    2.950    4.599    3.124    3.176

Response: Nov.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV           2.531    2.942    4.039    3.719    3.476    2.509    2.095    4.244    13.00
adjCV        2.531    2.905    3.932    3.618    3.368    2.430    2.018    4.046    12.34

Response: Dec.
       (Intercept)  1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
CV          0.7245   0.8289    1.141    1.004    1.039   1.0321    1.120    2.023    2.146
adjCV       0.7245   0.8194    1.118    0.978    1.010   0.9999    1.082    1.937    2.037

TRAINING: % variance explained
       1 comps  2 comps  3 comps  4 comps  5 comps  6 comps  7 comps  8 comps
X     79.59335   93.187    97.85    99.55    99.95   100.00   100.00   100.00
Sep.  33.15686   64.154    64.57    68.88    69.34    70.11    71.02    73.69
Oct.   0.04151    2.535    37.71    61.28    74.35    78.66    97.35    99.84
Nov.  13.15050   26.013    42.01    61.45    80.55    92.30    93.76    95.94
Dec.   2.77538    6.022    37.02    48.27    60.48    66.94    71.64    98.65
NULL
>
> Ypred <- predict(Model,newdata=as.matrix(DF[5,-c(j87,1,14)]))[,,1:6]
> print(Ypred)
        1 comps   2 comps   3 comps   4 comps   5 comps   6 comps
Sep. 24.781632 35.269462 34.981368 35.998053 35.628107 35.631329
Oct. 11.359193 14.502813 17.287631 19.801144 17.720314 17.712255
Nov.  8.671275  3.079570  4.550558  6.338256  4.368271  4.357846
Dec.  4.143504  3.339169  3.925335  4.314683  3.863716  3.861502
>
```

Appendix S R Script and results for mean monthly flow

```
#   Regression of Long term mean monthy Equation
hydro<-read.csv("hydro.csv"  header = TRUE)
AE<-(hydro$AE)# Averege elevation in (m)
AP<-(hydro$AP)# Annual Precipitation (mm)
CA3<-(hydro$CA3)# Area (m^2)under 3000 m elevation (m)
CA5<-(hydro$CA5)# Area(m^2) under 5000 m elevation (m)
Area3<- seq(min(CA3),max(CA3),length.out=3)
AVerElve<- seq(min(AE),max(AE),length.out=3)
Appp<- seq(min(AP),max(AP),length.out=3)

#MJANUARY
Jan<- (hydro$Jan)
Janmodel<-lm( log(Jan)~ log(AE)+log(AP)+log(CA3),data= hydro)
Janpredict<-exp(predict(Janmodel))
```

```
print(summary(Janmodel))
plot(Jan,Janpredict)
abline(0,1,col='red')
Jandesign <-  data.frame(expand.grid(CA3=Area3,AE=AVerElve,AP=Appp))
Janpred<-exp(predict(Janmodel,newdata=Jandesign,
                     interval = 'prediction',level=0.68))
Janfit<-Janpred[,'fit']
Janlwr<-Janpred[,'lwr']
Janupr<-Janpred[,'upr']
Janerror<-(Janupr-Janfit)/Janfit
print(cbind(Jandesign,Janfit,Janlwr,Janupr,Janerror*100))
print(cbind(min(Janfit),max(Janfit),mean(Jan-
fit),min(Janlwr),max(Janlwr),mean(Janlwr),min(Janupr),max(Janupr),mean(Janupr),min(Janer-
ror*100),max(Janerror*100),mean(Janerror*100)))

#FEBURARY

Feb<- (hydro$Feb)
Febmodel<-lm( log(Feb)~log(AE)+ log(AP)+log(CA3),data= hydro)
Febpredict<-exp(predict(Febmodel))
print(summary(Febmodel))
plot(Feb,Febpredict)
abline(0,1)
Febdesign= data.frame(expand.grid(CA3=Area3,AE=AVerElve, AP=Appp))
Febpred<-exp(predict(Febmodel,newdata=Febdesign, interval='prediction',level=0.68))
Febfit<-Febpred[,'fit']
Feblwr<-Febpred[,'lwr']
Febupr<-Febpred[,'upr']
Feberror<-(Febupr-Febfit)/Febfit
print(cbind(Febdesign,Febfit,Feblwr,Febupr,Feberror*100))
print(cbind(min(Febfit),max(Febfit),mean(Feb-
fit),min(Feblwr),max(Feblwr),mean(Feblwr),min(Febupr),max(Febupr),mean(Febupr),min(Feber-
ror*100),max(Feberror*100),mean(Feberror*100)))

# JUNE
Jun<- (hydro$Jun)
Junemodel<-lm( log(Jun)~log(AE)+ log(AE)+log(CA3),data= hydro)
Junepredict<-exp(predict(Junemodel))
print(summary(Junemodel))
plot(Jun,Junepredict)
abline(0,1)
Junedesign= data.frame( expand.grid( CA3=Area3,AE=AVerElve,AP=Appp))
Junepred<-exp(predict(Junemodel,newdata=Junedesign, interval='prediction',level=0.68))
Junfit<-Junepred[,'fit']
Junlwr<-Junepred[,'lwr']
Junupr<-Junepred[,'upr']
Junerror<-(Junupr-Junfit)/Junfit
print(cbind(Junedesign,Junfit,Junlwr,Junupr,Junerror*100,Junerror95*100))
print(cbind(min(Junfit),max(Junfit),mean(Junfit),min(Junlwr),max(Junlwr),mean(Jun-
lwr),min(Junupr),max(Junupr),mean(Junupr),min(Junerror*100),max(Junerror*100),mean(Junerror*10
0)))

#JULY
Jul<- (hydro$Jul)
Julymodel<-lm( log(Jul)~log(AE)+ log(AE)+log(CA3),data= hydro)
Julypredict<-exp(predict(Julymodel))
print(summary(Julymodel))
plot(Jul,Julypredict)
abline(0,1)
Julydesign= data.frame( expand.grid( CA3=Area3,AE=AVerElve,AP=Appp))
Julypred<-exp(predict(Julymodel,newdata=Julydesign, interval='prediction',level=0.68))
Julfit<-Julypred[,'fit']
```

```
Jullwr<-Julypred[,'lwr']
Julupr<-Julypred[,'upr']
Julerror<-(Julupr-Julfit)/Julfit
print(cbind(Julydesign,Julfit,Jullwr,Julupr,Julerror*100))
print(cbind(min(Julfit),max(Julfit),mean(Jul-
fit),min(Jullwr),max(Jullwr),mean(Jullwr),min(Julupr),max(Julupr),mean(Julupr),min(Juler-
ror*100),max(Julerror*100),mean(Julerror*100)))

# August
Aug<- (hydro$Aug)
Augustmodel<-lm( log(Aug)~log(AE)+ log(AE)+log(CA3),data= hydro)
Augustpredict<-exp(predict(Augustmodel))
print(summary(Augustmodel))
plot(Aug,Augustpredict)
abline(0,1)
Augustdesign= data.frame( expand.grid( CA3=Area3,AE=AVerElve,AE=Appp))
Augustpred<-exp(predict(Augustmodel,newdata=Augustdesign, interval='prediction',level=0.68))
print(cbind(Augustdesign,Augfit,Auglwr,Augupr,Augerror*100))
print(cbind(min(Augfit),max(Augfit),mean(Augfit),min(Auglwr),max(Auglwr),mean(Auglwr),min(Au-
gupr),max(Augupr),mean(Augupr),min(Augerror*100),max(Augerror*100),mean(Augerror*100)))

#September
Sep<- (hydro$Sep)
Septembermodel<-lm( log(Sep)~log(AE)+ log(AE)+log(CA3),data= hydro)
Septemberpredict<-exp(predict(Septembermodel))
print(summary(Septembermodel))
plot(Sep,Septemberpredict)
abline(0,1)
Septemberdesign= data.frame( expand.grid( CA3=Area3,AE=AVerElve,AP=Appp))
Septemberpred<-exp(predict(Septembermodel,newdata=Septemberdesign,          interval='predic-
tion',level=0.68))
Sepfit<-Septemberpred[,'fit']
Seplwr<-Septemberpred[,'lwr']
Sepupr<-Septemberpred[,'upr']
Seperror<-(Sepupr-Sepfit)/Sepfit
print(cbind(Septemberdesign,Sepfit,Seplwr,Sepupr,Seperror*100,Seperror95*100))
print(cbind(min(Octfit),max(Octfit),mean(Octfit),min(Octlwr),max(Octlwr),mean(Octlwr),min(Oc-
tupr),max(Octupr),mean(Octupr),min(Octerror*100),max(Octerror*100),mean(Octerror*100)))

#October
Oct<- (hydro$Oct)
Octobermodel<-lm( log(Oct)~log(AE)+ log(AE)+log(CA3),data= hydro)
Octoberpredict<-exp(predict(Octobermodel))
print(summary(Octobermodel))
plot(Oct,Octoberpredict)
abline(0,1)
Octoberdesign= data.frame( expand.grid( CA3=Area3,AE=AVerElve,AP=Appp))
Octoberpred<-exp(predict(Octobermodel,newdata=Octoberdesign,                interval='predic-
tion',level=0.68))
print(summary(Octoberpred))
Octfit<-Octoberpred[,'fit']
Octlwr<-Octoberpred[,'lwr']
Octupr<-Octoberpred[,'upr']
Octerror<-(Octupr-Octfit)/Octfit
print(cbind(Octoberdesign,Octfit,Octlwr,Octupr,Octerror*100,Octerror95*100))
print(cbind(min(Octfit),max(Octfit),mean(Octfit),min(Octlwr),max(Octlwr),mean(Octlwr),min(Oc-
tupr),max(Octupr),mean(Octupr),min(Octerror*100),max(Octerror*100),mean(Octerror*100)))

# November
Nov<- (hydro$Nov)
Novembermodel<-lm( log(Nov)~log(AE)+ log(AE)+log(CA3),data= hydro)
Novemberpredict<-exp(predict(Novembermodel))
```

```
print(summary(Novembermodel))
plot(Nov,Novemberpredict)
abline(0,1)
Novemberdesign= data.frame( expand.grid( CA3=Area3,AE=AVerElve,AP=Appp))
Novemberpred<-exp(predict(Novembermodel,newdata=Novemberdesign,              interval='predic-
tion',level=0.68))
Novfit<-Novemberpred[,'fit']
Novlwr<-Novemberpred[,'lwr']
Novupr<-Novemberpred[,'upr']
Noverror<-(Novupr-Novfit)/Novfit
print(cbind(Novemberdesign,Novfit,Novlwr,Novupr,Noverror*100,Noverror95*100))
print(cbind(min(Novfit),max(Novfit),mean(Novfit),min(Novlwr),max(Novlwr),mean(Novlwr),min(Novu
pr),max(Novupr),mean(Novupr),min(Noverror*100),max(Noverror*100),mean(Noverror*100)))

#December
Dec<- (hydro$Dec)
Decembermodel<-lm( log(Dec)~log(AE)+ log(AE)+log(CA3),data= hydro)
Decemberpredict<-exp(predict(Decembermodel))
print(summary(Decembermodel))
plot(Dec,Decemberpredict)
abline(0,1)
Decemberdesign= data.frame( expand.grid( CA3=Area3,AE=AVerElve,AP=Appp))
Decemberpred<-exp(predict(Decembermodel,newdata=Decemberdesign,              interval='predic-
tion',level=0.68))
Decfit<-Decemberpred[,'fit']
Declwr<-Decemberpred[,'lwr']
Decupr<-Decemberpred[,'upr']
Decerror<-(Decupr-Decfit)/Decfit
print(cbind(Decemberdesign,Decfit,Declwr,Decupr,Decerror*100))
print(cbind(min(Decfit),max(Decfit),mean(Decfit),min(Declwr),max(Declwr),mean(Declwr),min(De-
cupr),max(Decupr),mean(Decupr),min(Decerror*100),max(Decerror*100),mean(Decerror*100)))

# These are 3 month that uses onl one independent variables
# March
Mar<- (hydro$Mar)
Marchmodel<-lm(sqrt(Mar)~sqrt(CA5), data= hydro)
Marchpredict<-predict(Marchmodel)^2
print(summary(Marchmodel))
plot(Mar,Marchpredict)
abline(0,1)
Elve<- seq(min(hydro$CA5),max(hydro$CA5),length.out=101)
plot(hydro$CA5,Mar)
Marpred <- predict(Marchmodel,newdata=data.frame(CA5=Elve),interval='prediction',level=0.68)^2
lines(Elve,Marpred[,'fit'],col='blue')
lines(Elve,Marpred[,'upr'],col='red')
lines(Elve,Marpred[,'lwr'],col='red')
Marchfit<-Marpred[,'fit']
Marchupr<- Marpred[,'upr']
Marchlwr<-Marpred[,'lwr']
Marcherror<-((Marchupr-Marchfit)/Marchfit)
mean(Marcherror*100)
print(cbind(min(Marchfit),max(Marchfit),mean(Marchfit),min(Marchlwr),max(Marchlwr),mean(March-
lwr),min(Marchupr),max(Marchupr),mean(Marchupr),min(Marcherror*100),max(Marcher-
ror*100),mean(Marcherror*100)))

# April

Apr<- (hydro$Apr)
Aprilmodel<-lm(sqrt(Apr)~sqrt(CA5), data= hydro)
Aprilpredict<-predict(Aprilmodel)^2
print(summary(Aprilmodel))
plot(Apr,Aprilpredict)
```

```
abline(0,1)
Elve<- seq(min(hydro$CA5),max(hydro$CA5),length.out=101)
plot(hydro$CA5,Apr)
Aprpred <- predict(Aprilmodel,newdata=data.frame(CA5=Elve),interval='prediction',level=0.68)^2
lines(Elve,Aprpred[,'fit'],col='blue')
lines(Elve,Aprpred[,'upr'],col='red')
lines(Elve,Aprpred[,'lwr'],col='red')
Aprilfit<-Aprpred[,'fit']
Aprilupr<- Aprpred[,'upr']
Aprillwr<-Aprpred[,'lwr']
Aprilerror<-((Aprilupr-Aprilfit)/Aprilfit)
mean(Aprilerror)
print(cbind(min(Aprilfit),max(Aprilfit),mean(April-
fit),min(Aprillwr),max(Aprillwr),mean(Aprillwr),min(Aprilupr),max(Aprilupr),mean(Aprilupr),min
(Aprilerror*100),max(Aprilerror*100),mean(Aprilerror*100)))
# May#

May<- (hydro$May)
maymodel<-lm(sqrt(May)~sqrt(CA5), data= hydro)
Maypredict<-predict(maymodel)^2
print(summary(maymodel))
plot(May,Maypredict)
abline(0,1)
Elve<- seq(min(hydro$CA5),max(hydro$CA5),length.out=101)
plot(hydro$CA5,May)
Maypred <- predict(maymodel,newdata=data.frame(CA5=Elve),interval = 'prediction',level=0.68)^2
lines(Elve,Maypred[,'fit'],col='blue')
lines(Elve,Maypred[,'lwr'],col='red')
lines(Elve,Maypred[,'upr'],col='red')
Mayfit<-Maypred[,'fit']
Mayupr<- Maypred[,'upr']
Maylwr<-Maypred[,'lwr']
Mayerror<-((Mayupr-Mayfit)/Mayfit)
mean(Mayerror*100)
print(cbind(min(Mayfit),max(Mayfit),mean(May-
fit),min(Maylwr),max(Maylwr),mean(Maylwr),min(Mayupr),max(Mayupr),mean(Mayupr),min(May-
error*100),max(Mayerror*100),mean(Mayerror*100)))
```

```
> print(summary(Febmodel))

Call:
lm(formula = log(Feb) ~ log(AE) + log(AP) + log(CA3), data = hydro)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.77941 -0.17815 -0.02229  0.14349  1.01562 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -17.20264    1.82093  -9.447 1.94e-12 ***
log(AE)       1.42311    0.10783  13.197  < 2e-16 ***
log(AP)       0.45624    0.21149   2.157   0.0361 *  
log(CA3)      0.81175    0.03461  23.457  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3623 on 47 degrees of freedom
Multiple R-squared:  0.9532,	Adjusted R-squared:  0.9502 
F-statistic: 319.3 on 3 and 47 DF,  p-value: < 2.2e-16

> print(summary(Marchmodel))

Call:
lm(formula = sqrt(Mar) ~ sqrt(CA5), data = hydro)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.3949 -0.4312  0.1315  0.7107  1.7705 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.383586   0.204175   1.879   0.0662 .  
sqrt(CA5)   0.091134   0.002689  33.887   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 0.9449 on 49 degrees of freedom
Multiple R-squared:  0.9591,     Adjusted R-squared:  0.9582
F-statistic:  1148 on 1 and 49 DF,  p-value: < 2.2e-16

> print(summary(Aprilmodel))

Call:
lm(formula = sqrt(Apr) ~ sqrt(CA5), data = hydro)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5945 -0.5554  0.2360  0.7708  2.2489

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.181220   0.273997   0.661    0.511
sqrt(CA5)   0.103917   0.003609  28.794   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.268 on 49 degrees of freedom
Multiple R-squared:  0.9442,     Adjusted R-squared:  0.9431
F-statistic: 829.1 on 1 and 49 DF,  p-value: < 2.2e-16

> print(summary(maymodel))

Call:
lm(formula = sqrt(May) ~ sqrt(CA5), data = hydro)

Residuals:
    Min      1Q  Median      3Q     Max
-5.4754 -0.6222  0.1026  1.0412  3.4198

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0005779  0.4104434  -0.001    0.999
sqrt(CA5)    0.1364632  0.0054063  25.242   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.9 on 49 degrees of freedom
Multiple R-squared:  0.9286,     Adjusted R-squared:  0.9271
F-statistic: 637.1 on 1 and 49 DF,  p-value: < 2.2e-16

> print(summary(Junemodel))

Call:
lm(formula = log(Jun) ~ log(AE) + log(AE) + log(CA3), data = hydro)

Residuals:
     Min       1Q   Median       3Q      Max
-1.50610 -0.21816  0.02398  0.25071  1.12923

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -14.04456    1.16745  -12.03 4.26e-16 ***
log(AE)       1.65258    0.15058   10.97 1.10e-14 ***
log(CA3)      0.81433    0.04319   18.86  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5085 on 48 degrees of freedom
Multiple R-squared:  0.9206,     Adjusted R-squared:  0.9172
F-statistic: 278.1 on 2 and 48 DF,  p-value: < 2.2e-16

> print(summary(Julymodel))

Call:
lm(formula = log(Jul) ~ log(AE) + log(AE) + log(CA3), data = hydro)

Residuals:
     Min       1Q   Median       3Q      Max
-1.25189 -0.19749  0.01318  0.21885  0.92533

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.45827    1.05394  -9.923 3.26e-13 ***
log(AE)       1.30667    0.13594   9.612 9.09e-13 ***
log(CA3)      0.82219    0.03899  21.088  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.459 on 48 degrees of freedom
Multiple R-squared:  0.9279,     Adjusted R-squared:  0.9249
F-statistic: 308.8 on 2 and 48 DF,  p-value: < 2.2e-16

> print(summary(Augustmodel))

Call:
lm(formula = log(Aug) ~ log(AE) + log(AE) + log(CA3), data = hydro)

Residuals:
     Min       1Q   Median       3Q      Max
-1.01849 -0.20516  0.03212  0.18072  0.82868

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.04812    0.88910  -11.30 3.97e-15 ***
log(AE)       1.28201    0.11468   11.18 5.81e-15 ***
```

```
log(CA3)      0.81915    0.03289   24.91  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3873 on 48 degrees of freedom
Multiple R-squared:  0.9469,     Adjusted R-squared:  0.9447
F-statistic:   428 on 2 and 48 DF,  p-value: < 2.2e-16

> print(summary(Septembermodel))

Call:
lm(formula = log(Sep) ~ log(AE) + log(AE) + log(CA3), data = hydro)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.93033 -0.20394 -0.00593  0.19085  0.92388 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -9.16259    0.89386 -10.251 1.12e-13 ***
log(AE)      1.12540    0.11529   9.762 5.55e-13 ***
log(CA3)     0.82375    0.03307  24.912  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.3893 on 48 degrees of freedom
Multiple R-squared:  0.9443,     Adjusted R-squared:  0.942
F-statistic: 407.1 on 2 and 48 DF,  p-value: < 2.2e-16

> print(summary(Octobermodel))

Call:
lm(formula = log(Oct) ~ log(AE) + log(AE) + log(CA3), data = hydro)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.99976 -0.23364  0.01614  0.19531  0.94245 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -10.70402    0.86155  -12.42  < 2e-16 ***
log(AE)       1.24086    0.11112   11.17 6.05e-15 ***
log(CA3)      0.79760    0.03187   25.03  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3752 on 48 degrees of freedom
Multiple R-squared:  0.9473,     Adjusted R-squared:  0.9451
F-statistic: 431.1 on 2 and 48 DF,  p-value: < 2.2e-16

> print(summary(Novembermodel))

Call:
lm(formula = log(Nov) ~ log(AE) + log(AE) + log(CA3), data = hydro)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.09803 -0.21680  0.01152  0.18903  0.99779 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -12.48735    0.87237  -14.31   <2e-16 ***
log(AE)       1.39089    0.11252   12.36   <2e-16 ***
log(CA3)      0.78199    0.03227   24.23   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.38 on 48 degrees of freedom
Multiple R-squared:  0.947,      Adjusted R-squared:  0.9448
F-statistic: 428.7 on 2 and 48 DF,  p-value: < 2.2e-16

> print(summary(Decembermodel))

Call:
lm(formula = log(Dec) ~ log(AE) + log(AE) + log(CA3), data = hydro)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.0600 -0.1805 -0.0342  0.1711  0.9464 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -13.09480    0.84538  -15.49   <2e-16 ***
log(AE)       1.41615    0.10904   12.99   <2e-16 ***
log(CA3)      0.78102    0.03127   24.97   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3682 on 48 degrees of freedom
Multiple R-squared:  0.9504,     Adjusted R-squared:  0.9483
F-statistic: 459.7 on 2 and 48 DF,  p-value: < 2.2e-16
```

Appendix T R script and results for probability of exceedance

```
prob<-read.csv("prob.csv", header = TRUE )
AE<-(prob$AE)
AP<-(prob$AP)
CA3<-(prob$CA3)
CA5<-(prob$CA5)
Area3<- seq(min(CA3),max(CA3),length.out=3)
AVerElve<- seq(min(AE),max(AE),length.out=3)
Appp<- seq(min(AP),max(AP),length.out=3)
Area5<-seq(min(CA5),max(CA5),length.out=3)

#0%
Flow0 <-(prob$X0)
Flow0model <-lm(sqrt(Flow0)~sqrt(AE)+sqrt(CA3))
Flow0predict<-(predict(Flow0model))^2
print(summary(Flow0model))
plot(Flow0,Flow0predict)
abline(0,1)
Flow0design<-data.frame(expand.grid( AE=AVerElve,CA3=Area3))
Flow0pred<-predict(Flow0model,newdata=Flow0design,interval='prediction', level=0.68)^2
Flow0fit<-Flow0pred[,'fit']
Flow0lwr<-Flow0pred[,'lwr']
Flow0upr<-Flow0pred[,'upr']
Flow0error<-(Flow0upr-Flow0fit)/Flow0fit
print(cbind(Flow0design,Flow0fit,Flow0upr,Flow0lwr,Flow0error*100))
print(cbind(min(Flow0fit),max(Flow0fit),mean(Flow0fit),min(Flow0lwr),max(Flow0lwr),mean(Flow0l
wr),min(Flow0upr),max(Flow0upr),mean(Flow0upr),min(Flow0error*100),max(Flow0er-
ror*100),mean(Flow0error*100)))

#5% exceedence probability
Flow5 <-(prob$X5)
Flow5model <-lm(log(Flow5)~log(AE)+log(AP)+log(CA3))
Flow5predict<-exp(predict(Flow5model))
print(summary(Flow5model))
abline(0,1)
Flow5design<-data.frame(expand.grid( AE=AVerElve,CA3=Area3,AP=Appp))
Flow5pred<-exp(predict(Flow5model,newdata=Flow5design,interval='prediction', level=0.68))
Flow5fit<-Flow5pred[,'fit']
Flow5lwr<-Flow5pred[,'lwr']
Flow5upr<-Flow5pred[,'upr']
Flow5error<-(Flow5upr-Flow5fit)/Flow5fit
print(cbind(Flow5design,Flow5fit,Flow5upr,Flow5lwr,Flow5error*100))
print(cbind(min(Flow5fit),max(Flow5fit),mean(Flow5fit),min(Flow5lwr),max(Flow5lwr),mean(Flow5l
wr),min(Flow5upr),max(Flow5upr),mean(Flow5upr),min(Flow5error*100),max(Flow5er-
ror*100),mean(Flow5error*100)))

#20% exceedence probability
Flow20 <-(prob$X20)
Flow20model <-lm(log(Flow20)~log(AE)+log(AP)+log(CA3))
Flow20predict<-exp(predict(Flow5model))
print(summary(Flow20model))
Flow20design<-data.frame(expand.grid( AE=AVerElve,CA3=Area3,AP=Appp))
Flow20pred<-exp(predict(Flow20model,newdata=Flow20design,interval='prediction', level=0.68))
Flow20fit<-Flow20pred[,'fit']
Flow20lwr<-Flow20pred[,'lwr']
Flow20upr<-Flow20pred[,'upr']
Flow20error<-(Flow20upr-Flow20fit)/Flow20fit
print(cbind(Flow20design,Flow20fit,Flow20upr,Flow20lwr,Flow20error*100))
```

```
#40% exceedence probability

Flow40 <-(prob$X40)
Flow40model <-lm(log(Flow40)~log(AP)+log(AE)+log(CA3))
Flow40predict<-exp(predict(Flow40model))
print(summary(Flow40model))
Flow40design<-data.frame(expand.grid( AE=AVerElve,CA3=Area3,AP=Appp))
Flow40pred<-exp(predict(Flow40model,newdata=Flow40design,interval='prediction', level=0.68))
Flow40fit<-Flow40pred[,'fit']
Flow40lwr<-Flow40pred[,'lwr']
Flow40upr<-Flow40pred[,'upr']
Flow40error<-(Flow40upr-Flow40fit)/Flow40fit
print(cbind(Flow40design,Flow40fit,Flow40upr,Flow40lwr,Flow40error*100))

#60% exceedence probability
Flow60 <-(prob$X60)
Flow60model <-lm(log(Flow60)~log(AP)+log(AE)+log(CA3))
Flow60predict<-exp(predict(Flow60model))
print(summary(Flow60model))
Flow60design<-data.frame(expand.grid( AE=AVerElve,CA3=Area3,AP=Appp))
Flow60pred<-exp(predict(Flow60model,newdata=Flow60design,interval='prediction', level=0.68))
Flow60fit<-Flow60pred[,'fit']
Flow60lwr<-Flow60pred[,'lwr']
Flow60upr<-Flow60pred[,'upr']
Flow60error<-(Flow60upr-Flow60fit)/Flow60fit
print(cbind(Flow60design,Flow60fit,Flow60upr,Flow60lwr,Flow60error*100))

#80% exceedence probability
Flow80 <-(prob$X80)
Flow80model <-lm(log(Flow80)~log(AE)+log(AP)+log(CA3))
Flow80predict<-exp(predict(Flow80model))
print(summary(Flow80model))
Flow80design<-data.frame(expand.grid( AE=AVerElve,CA3=Area3,AP=Appp))
Flow80pred<-exp(predict(Flow80model,newdata=Flow80design,interval='prediction', level=0.68))
Flow80pred95<-exp(predict(Flow80model,newdata=Flow80design,interval='prediction'))
Flow80fit<-Flow80pred[,'fit']
Flow80lwr<-Flow80pred[,'lwr']
Flow80upr<-Flow80pred[,'upr']
Flow80error<-(Flow80upr-Flow80fit)/Flow80fit
print(cbind(Flow80design,Flow80fit,Flow80upr,Flow80lwr,Flow80error*100))

#95% exceedence probability
Flow95 <-(prob$X95)
Flow95model <-lm(log(Flow95)~log(AE)+log(AP)+log(CA3))
Flow95predict<-exp(predict(Flow95model))
print(summary(Flow95model))
Flow95design<-data.frame(expand.grid( AE=AVerElve,CA3=Area3,AP=Appp))
Flow95pred<-exp(predict(Flow95model,newdata=Flow95design,interval='prediction', level=0.68))
Flow95fit<-Flow95pred[,'fit']
Flow95lwr<-Flow95pred[,'lwr']
Flow95upr<-Flow95pred[,'upr']
Flow95error<-(Flow95upr-Flow95fit)/Flow95fit
print(cbind(Flow95design,Flow95fit,Flow95upr,Flow95lwr,Flow95error*100))

# 100% exceedence probability
Flow100 <-(prob$X1)
Flow100model <-lm(sqrt(Flow100)~sqrt(AP)+sqrt(CA5))
Flow100predict<-(predict(Flow100model))^2
print(summary(Flow100model))
plot(Flow100,Flow100predict)
abline(0,1)
Flow100design<-data.frame(expand.grid( AP=Appp,CA5=Area5))
```

```
Flow100pred<-predict(Flow100model,newdata=Flow100design,interval='prediction', level=0.68)^2
Flow100fit<-Flow100pred[,'fit']
Flow100lwr<-Flow100pred[,'lwr']
Flow100upr<-Flow100pred[,'upr']
Flow100error<-(Flow100upr-Flow100fit)/Flow100fit
print(cbind(Flow100design,Flow100fit,Flow100upr,Flow100lwr,Flow100error*100))
> print(summary(Flow20model))

Call:
lm(formula = log(Flow20) ~ log(AE) + log(AP) + log(CA3))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.88056 -0.19808  0.01086  0.19564  0.94107 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -17.29567    1.92939  -8.964 9.65e-12 ***
log(AE)       1.35981    0.11426  11.901 8.71e-16 ***
log(AP)       0.76074    0.22408   3.395   0.0014 ** 
log(CA3)      0.88183    0.03667  24.049  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3839 on 47 degrees of freedom
Multiple R-squared:  0.9517,	Adjusted R-squared:  0.9486 
F-statistic: 308.5 on 3 and 47 DF,  p-value: < 2.2e-16

> print(summary(Flow40model))

Call:
lm(formula = log(Flow40) ~ log(AP) + log(AE) + log(CA3))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.92926 -0.20872 -0.02339  0.25455  1.04800 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -19.10212    2.03992  -9.364 2.55e-12 ***
log(AP)       0.67576    0.23692   2.852  0.00643 ** 
log(AE)       1.55646    0.12080  12.884  < 2e-16 ***
log(CA3)      0.85666    0.03877  22.097  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4059 on 47 degrees of freedom
Multiple R-squared:  0.9476,	Adjusted R-squared:  0.9443 
F-statistic: 283.4 on 3 and 47 DF,  p-value: < 2.2e-16

> print(summary(Flow60model))

Call:
lm(formula = log(Flow60) ~ log(AP) + log(AE) + log(CA3))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.87894 -0.30928 -0.05351  0.26153  0.99351 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -15.80498    2.04561  -7.726 6.54e-10 ***
log(AP)       0.11399    0.23758   0.480    0.634    
log(AE)       1.61698    0.12114  13.348  < 2e-16 ***
log(CA3)      0.83453    0.03888  21.466  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.407 on 47 degrees of freedom
Multiple R-squared:  0.9492,	Adjusted R-squared:  0.9459 
F-statistic: 292.5 on 3 and 47 DF,  p-value: < 2.2e-16

> print(summary(Flow80model))

Call:
lm(formula = log(Flow80) ~ log(AE) + log(AP) + log(CA3))

Residuals:
     Min       1Q   Median       3Q      Max 
-0.87799 -0.20113 -0.02628  0.19961  1.03029 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -19.35979    1.94304  -9.964  3.6e-13 ***
log(AE)       1.59185    0.11507  13.834  < 2e-16 ***
log(AP)       0.55367    0.22567   2.453   0.0179 *  
log(CA3)      0.83053    0.03693  22.491  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 0.3866 on 47 degrees of freedom
Multiple R-squared:  0.9512,	Adjusted R-squared:  0.948 
F-statistic: 305.1 on 3 and 47 DF,  p-value: < 2.2e-16

> print(summary(Flow95model))

Call:
lm(formula = log(Flow95) ~ log(AE) + log(AP) + log(CA3))

Residuals:
    Min      1Q  Median      3Q     Max 
-1.0065 -0.2262 -0.0262  0.2128  1.1665 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -21.05848    2.11894  -9.938 3.91e-13 ***
log(AE)       1.73560    0.12548  13.831  < 2e-16 ***
log(AP)       0.58393    0.24610   2.373   0.0218 *  
log(CA3)      0.83751    0.04027  20.798  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4216 on 47 degrees of freedom
Multiple R-squared:  0.9456,	Adjusted R-squared:  0.9422 
F-statistic: 272.6 on 3 and 47 DF,  p-value: < 2.2e-16
```

Appendix U R Script and results for low flow analysis

```
# Regression low flow equation
daka<-read.csv('daka.csv', header = TRUE)
AE<-(daka$AE)
AP<-(daka$AP)
CA3<-(daka$CA3)
CA5<-(daka$CA5)
# 2 yr 1 day
dayone2yr<-(daka$dayone2yr)
dayone2yrmodel<-lm (sqrt(dayone2yr)~sqrt(CA5),data= daka)
dayone2yrpredict<-predict(dayone2yrmodel)^2
print(summary(dayone2yrmodel))
plot(dayone2yr,dayone2yrpredict,xlab='actual low flow ',ylab='predicted low flow')
abline(0,1,col='red')
Elve<- seq(min(daka$CA5),max(daka$CA5),length.out=101)
plot(daka$CA5,dayone2yr,xlab = 'area under 5000 m',ylab='2 year 1 day low flow')
dayone2yrpred<-  predict(dayone2yrmodel,newdata=data.frame(CA5=Elve),interval  =  'prediction'
,level=0.68)^2
lines(Elve,dayone2yrpred[,'fit'],col='blue')
lines(Elve,dayone2yrpred[,'lwr'],col='red')
lines(Elve,dayone2yrpred[,'upr'],col='red')
dayone2yrfit<-(dayone2yrpred[,'fit'])
dayone2yrlwr<-(dayone2yrpred[,'lwr'])
dayone2yrupr<-(dayone2yrpred[,'upr'])
dayone2yrerror<-(dayone2yrupr-dayone2yrfit)/dayone2yrfit
dayone2yrerror*100
print(cbind(dayone2yrfit,dayone2yrupr,dayone2yrlwr,dayone2yrerror*100))

#  1 day 10 yr

dayone10yr<-(daka$dayone10yr)
dayone10yrmodel<-lm (sqrt(dayone10yr)~sqrt(CA5),data= daka)
dayone10yrpredict<-predict(dayone10yrmodel)^2
print(summary(dayone10yrmodel))
abline(0,1,col='red')
Elve<- seq(min(daka$CA5),max(daka$CA5),length.out=101)
plot(daka$CA5,dayone10yr, xlab = 'Area under 5000m ', ylab='10 year 1 day low flow')
dayone10yrpred<-   predict(dayone10yrmodel,newdata=data.frame(CA5=Elve),interval  =  'predic-
tion',level=0.68)^2
lines(Elve,dayone10yrpred[,'fit'],col='blue')
lines(Elve,dayone10yrpred[,'lwr'],col='red')
```

```
lines(Elve,dayone10yrpred[,'upr'],col='red')
dayone10yrfit<-dayone10yrpred[,'fit']
dayone10yrlwr<-dayone10yrpred[,'lwr']
dayone10yrupr<-dayone10yrpred[,'upr']
dayone10yrerror<-(dayone10yrupr-dayone10yrfit)/dayone10yrfit
print(cbind(dayone10yrfit,dayone10yrlwr,dayone10yrupr,dayone10yrerror*100))
# 1 day 10 year
print(cbind(min(dayone10yrfit),max(dayone10yrfit),mean(dayone10yrfit)))

# day 1 20 yrdays
dayone20yr<-(daka$dayone20yr)
dayone20yrmodel<-lm (sqrt(dayone20yr)~sqrt(CA5),data= daka)
dayone20yrpredict<-predict(dayone20yrmodel)^2
print(summary(dayone20yrmodel))
abline(0,1,col='red')
Elve<- seq(min(daka$CA5),max(daka$CA5),length.out=101)
plot(daka$CA5,dayone20yr)
dayone20yrpred<-   predict(dayone20yrmodel,newdata=data.frame(CA5=Elve),interval   =   'predic-
tion',level=0.68)^2
lines(Elve,dayone20yrpred[,'fit'],col='blue')
lines(Elve,dayone20yrpred[,'lwr'],col='red')
lines(Elve,dayone20yrpred[,'upr'],col='red')
dayone20yrfit<-dayone20yrpred[,'fit']
dayone20yrlwr<-dayone20yrpred[,'lwr']
dayone20yrupr<-dayone20yrpred[,'upr']
dayone20yrerror<-(dayone20yrupr-dayone20yrfit)/dayone20yrfit
print(cbind(dayone20yrfit,dayone20yrlwr,dayone20yrupr,dayone20yrerror*100))
print(cbind(min(dayone20yrfit),max(dayone20yrfit),mean(dayone20yrfit)))
#  7 days 2 year
daysev2yr<-(daka$daysev2yr)
daysev2yrmodel<-lm (sqrt(daysev2yr)~sqrt(CA5),data= daka)
daysev2yrpredict<-predict(daysev2yrmodel)^2
print(summary(daysev2yrmodel))

abline(0,1,col='red')
Elve<- seq(min(daka$CA5),max(daka$CA5),length.out=101)
plot(daka$CA5,daysev2yr)
daysev2yrpred<-    predict(daysev2yrmodel,newdata=data.frame(CA5=Elve),interval    =    'predic-
tion',level=0.68)^2
lines(Elve,daysev2yrpred[,'fit'],col='blue')
lines(Elve,daysev2yrpred[,'lwr'],col='red')
lines(Elve,daysev2yrpred[,'upr'],col='red')
daysev2yrfit<-daysev2yrpred[,'fit']
daysev2yrlwr<-daysev2yrpred[,'lwr']
daysev2yrupr<-daysev2yrpred[,'upr']
daysev2yrerror<-(daysev2yrupr-daysev2yrfit)/daysev2yrfit
print(cbind(daysev2yrfit,daysev2yrlwr,daysev2yrupr,daysev2yrerror*100))
print(cbind(min(daysev2yrfit),max(daysev2yrfit),mean(daysev2yrfit)))

# 10 year 7 days
daysev10yr<-(daka$daysev10yr)
daysev10yrmodel<-lm (sqrt(daysev10yr)~sqrt(CA5),data= daka)
daysev10yrpredict<-predict(daysev10yrmodel)^2
print(summary(daysev10yrmodel))
abline(0,1,col='red')
Elve<- seq(min(daka$CA5),max(daka$CA5),length.out=101)
plot(daka$CA5,daysev10yr)
daysev10yrpred<-   predict(daysev10yrmodel,newdata=data.frame(CA5=Elve),interval   =   'predic-
tion',level=0.68)^2
lines(Elve,daysev10yrpred[,'fit'],col='blue')
lines(Elve,daysev10yrpred[,'lwr'],col='red')
lines(Elve,daysev10yrpred[,'upr'],col='red')
```

```
daysev10yrfit<-daysev10yrpred[,'fit']
daysev10yrlwr<-daysev10yrpred[,'lwr']
daysev10yrupr<-daysev10yrpred[,'upr']
daysev10yrerror<-(daysev10yrupr-daysev10yrfit)/daysev10yrfit
print(cbind(daysev10yrfit,daysev10yrlwr,daysev10yrupr,daysev10yrerror*100))
print(cbind(min(daysev10yrfit),max(daysev10yrfit),mean(daysev10yrfit)))
# 20yr 7 days
daysev20yr<-(daka$daysev20yr)
daysev20yrmodel<-lm (sqrt(daysev20yr)~sqrt(CA5),data= daka)
daysev20yrpredict<-predict(daysev20yrmodel)^2
print(summary(daysev20yrmodel))
abline(0,1,col='red')
Elve<- seq(min(daka$CA5),max(daka$CA5),length.out=101)
plot(daka$CA5,daysev20yr)
daysev20yr   <-   predict(daysev20yrmodel,newdata=data.frame(CA5=Elve),interval   =   'predic-
tion',level=0.68)^2
lines(Elve,daysev20yr[,'fit'],col='blue')
lines(Elve,daysev20yr[,'lwr'],col='red')
lines(Elve,daysev20yr[,'upr'],col='red')
daysev20yrfit<-daysev20yr[,'fit']
daysev20yrlwr<-daysev20yr[,'lwr']
daysev20yrupr<-daysev20yr[,'upr']
daysev20yrerror<-(daysev20yrupr-daysev20yrfit)/daysev20yrfit
print(cbind(daysev20yrfit,daysev20yrlwr,daysev20yrupr,daysev20yrerror*100))
print(cbind(min(daysev20yrfit),max(daysev20yrfit),mean(daysev20yrfit)))
# 2 yr 30 days
daythir2yr<-(daka$daythir2yr)
daythir2yrmodel<-lm (sqrt(daythir2yr)~sqrt(CA5),data= daka)
daythir2yrpredict<-predict(daythir2yrmodel)^2
print(summary(daythir2yrmodel))
abline(0,1,col='red')
Elve<- seq(min(daka$CA5),max(daka$CA5),length.out=101)
plot(daka$CA5,daythir2yr)
daythir2yr   <-   predict(daythir2yrmodel,newdata=data.frame(CA5=Elve),interval   =   'predic-
tion',level=0.68)^2
lines(Elve,daythir2yr[,'fit'],col='blue')
lines(Elve,daythir2yr[,'lwr'],col='red')
lines(Elve,daythir2yr[,'upr'],col='red')
daythir2yrfit<-daythir2yr[,'fit']
daythir2yrlwr<-daythir2yr[,'lwr']
daythir2yrupr<-daythir2yr[,'upr']
daythir2yrerror<-(daythir2yrupr-daythir2yrfit)/daythir2yrfit
print(cbind(daythir2yrfit,daythir2yrlwr,daythir2yrupr,daythir2yrerror*100))
print(cbind(min(daythir2yrfit),max(daythir2yrfit),mean(daythir2yrfit)))
# 10 yr 30 days
daythir10yr<-(daka$daythir10yr)
daythir10yrmodel<-lm (sqrt(daythir10yr)~sqrt(CA5),data= daka)
daythir10yrpredict<-predict(daythir10yrmodel)^2
print(summary(daythir10yrmodel))
abline(0,1,col='red')
Elve<- seq(min(daka$CA5),max(daka$CA5),length.out=101)
plot(daka$CA5,daythir10yr)
daythir10yr   <-   predict(daythir10yrmodel,newdata=data.frame(CA5=Elve),interval   =   'predic-
tion',level=0.68)^2
lines(Elve,daythir10yr[,'fit'],col='blue')
lines(Elve,daythir10yr[,'lwr'],col='red')
lines(Elve,daythir10yr[,'upr'],col='red')
daythir10yrfit<-daythir10yr[,'fit']
daythir10yrlwr<-daythir10yr[,'lwr']
daythir10yrupr<-daythir10yr[,'upr']
daythir10yrerror<-(daythir10yrupr-daythir10yrfit)/daythir10yrfit
print(cbind(daythir10yrfit,daythir10yrlwr,daythir10yrupr,daythir10yrerror*100))
```

```
print(cbind(min(daythir10yrfit),max(daythir10yrfit),mean(daythir10yrfit)))
# # 20 yr 30 days
daythir20yr<-(daka$daythir20yr)
daythir20yrmodel<-lm (sqrt(daythir20yr)~sqrt(CA5),data= daka)
daythir20yrpredict<-predict(daythir20yrmodel)^2
print(summary(daythir20yrmodel))
abline(0,1,col='red')
Elve<- seq(min(daka$CA5),max(daka$CA5),length.out=101)
plot(daka$CA5,daythir20yr)
daythir20yr   <-   predict(daythir20yrmodel,newdata=data.frame(CA5=Elve),interval   =   'predic-
tion',level=0.68)^2
lines(Elve,daythir20yr[,'fit'],col='blue')
lines(Elve,daythir20yr[,'lwr'],col='red')
lines(Elve,daythir20yr[,'upr'],col='red')
daythir20yrfit<-daythir20yr[,'fit']
daythir20yrlwr<-daythir20yr[,'lwr']
daythir20yrupr<-daythir20yr[,'upr']
daythir20yrerror<-(daythir20yrupr-daythir20yrfit)/daythir20yrfit
print(cbind(daythir20yrfit,daythir20yrlwr,daythir20yrupr,daythir20yrerror*100))
```

```
> print(summary(dayone2yrmodel))

Call:
lm(formula = sqrt(dayone2yr) ~ sqrt(CA5), data = daka)

Residuals:
    Min      1Q  Median      3Q     Max
-3.8092 -0.4454  0.1050  0.7659  1.9594

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.214380   0.252527   0.849      0.4
sqrt(CA5)   0.081495   0.003326  24.501   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.169 on 49 degrees of freedom
Multiple R-squared:  0.9245,   Adjusted R-squared:  0.923
F-statistic: 600.3 on 1 and 49 DF,  p-value: < 2.2e-16

> print(summary(dayone10yrmodel))

Call:
lm(formula = sqrt(dayone10yr) ~ sqrt(CA5), data = daka)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9874 -0.4357  0.0461  0.7649  1.8321

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.085987   0.240263   0.358    0.722
sqrt(CA5)   0.072872   0.003165  23.026   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.112 on 49 degrees of freedom
Multiple R-squared:  0.9154,   Adjusted R-squared:  0.9137
F-statistic: 530.2 on 1 and 49 DF,  p-value: < 2.2e-16

> print(summary(dayone20yrmodel))

Call:
lm(formula = sqrt(dayone20yr) ~ sqrt(CA5), data = daka)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9593 -0.4294  0.0475  0.7173  1.9157

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.068988   0.237409   0.291    0.773
sqrt(CA5)   0.070262   0.003127  22.468   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.099 on 49 degrees of freedom
Multiple R-squared:  0.9115,   Adjusted R-squared:  0.9097
F-statistic: 504.8 on 1 and 49 DF,  p-value: < 2.2e-16

> print(summary(daysev2yrmodel))

Call:
```

```
lm(formula = sqrt(daysev2yr) ~ sqrt(CA5), data = daka)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6917 -0.4316  0.1380  0.7405  2.1842 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.236193   0.248476   0.951    0.346    
sqrt(CA5)   0.082971   0.003273  25.351   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.15 on 49 degrees of freedom
Multiple R-squared:  0.9292,	Adjusted R-squared:  0.9277 
F-statistic: 642.7 on 1 and 49 DF,  p-value: < 2.2e-16

> print(summary(daysev10yrmodel))

Call:
lm(formula = sqrt(daysev10yr) ~ sqrt(CA5), data = daka)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.9380 -0.3640  0.1240  0.7446  1.7608 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.092035   0.237120   0.388      0.7    
sqrt(CA5)   0.074829   0.003123  23.958   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.097 on 49 degrees of freedom
Multiple R-squared:  0.9213,	Adjusted R-squared:  0.9197 
F-statistic:   574 on 1 and 49 DF,  p-value: < 2.2e-16

> print(summary(daysev20yrmodel))

Call:
lm(formula = sqrt(daysev20yr) ~ sqrt(CA5), data = daka)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.9643 -0.4648  0.1217  0.7035  1.8185 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.066189   0.234465   0.282    0.779    
sqrt(CA5)   0.072551   0.003088  23.492   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.085 on 49 degrees of freedom
Multiple R-squared:  0.9185,	Adjusted R-squared:  0.9168 
F-statistic: 551.9 on 1 and 49 DF,  p-value: < 2.2e-16

> print(summary(daythir2yrmodel))

Call:
lm(formula = sqrt(daythir2yr) ~ sqrt(CA5), data = daka)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.2042 -0.4511  0.1350  0.6973  2.3250 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.302594   0.237229   1.276    0.208    
sqrt(CA5)   0.085367   0.003125  27.320   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.098 on 49 degrees of freedom
Multiple R-squared:  0.9384,	Adjusted R-squared:  0.9371 
F-statistic: 746.4 on 1 and 49 DF,  p-value: < 2.2e-16

> print(summary(daythir10yrmodel))

Call:
lm(formula = sqrt(daythir10yr) ~ sqrt(CA5), data = daka)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6693 -0.7226 -0.1279  0.4413 13.3364 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.677729   0.475166   1.426     0.16    
sqrt(CA5)   0.072865   0.006259  11.642 1.02e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 2.199 on 49 degrees of freedom
Multiple R-squared:  0.7345,  Adjusted R-squared:  0.729
F-statistic: 135.5 on 1 and 49 DF,  p-value: 1.022e-15

> print(summary(daythir20yrmodel))

Call:
lm(formula = sqrt(daythir20yr) ~ sqrt(CA5), data = daka)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5289 -0.4152  0.1008  0.6905  1.7837

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.160570   0.226375   0.709    0.481
sqrt(CA5)   0.074159   0.002982  24.871   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.048 on 49 degrees of freedom
Multiple R-squared:  0.9266,  Adjusted R-squared:  0.9251
F-statistic: 618.5 on 1 and 49 DF,  p-value: < 2.2e-16
```

Appendix V R script for flood analysis

```
# Regression of  region flood analysis equation
flood<-read.csv('newflood.csv',header=TRUE)
#extract.var(flood)
CA3<- flood$CA3
yr2<- flood$yr2
yr100<- flood$yr100
stn.no <-flood$stn..no

# CA3 is the catchment area belwo3 3000 m
# yr2 is 2 year flood
# yr100 is 100 year flood
flood2model<-lm (log(yr2)~log(CA3),data=flood)
flood100model<-lm (log(yr100)~log(CA3),data=flood)
print(summary(flood2model))
print(summary(flood100model))
print(a2 <- exp(coefficients(flood2model)[1]))
print(b2 <- coefficients(flood2model)[2])
print(a100 <- exp(coefficients(flood100model)[1]))
print(b100 <- coefficients(flood100model)[2])
print(Table <- rbind(c(a2,b2),c(a100,b100)))
#=2.29*Z2^0.86

lve<- seq(min(flood$CA3),max(flood$CA3),length.out=101)
plot(flood$CA3,yr2,xlab = 'area under 3000m', ylab = ' 2 year flood', log='x')
flood2lo<-exp(predict(flood2model,interval='prediction',
                      newdata=data.frame(CA3=lve),level=0.68))
lines(lve,flood2lo[,'fit'],col='blue')
lines(lve,flood2lo[,'lwr'],col='red')
lines(lve,flood2lo[,'upr'],col='red')
flood2lofit<-(flood2lo[,'fit'])
flood2lolwr<-(flood2lo[,'lwr'])
flood2loupr<-(flood2lo[,'upr'])
flood2loerror<-(flood2loupr-flood2lofit)/flood2lofit
print(cbind(flood2lofit,flood2loupr,flood2lolwr,flood2loerror*100))

#100 yr flood
lve<- seq(min(flood$CA3),max(flood$CA3),length.out=101)
plot(flood$CA3,yr100,xlab = 'area under 3000m', ylab = ' 100 year flood', log='x')
flood100lo<-exp(predict(flood100model,interval='prediction',
                        newdata=data.frame(CA3=lve),level=0.68))
lines(lve,flood100lo[,'fit'],col='blue')
lines(lve,flood100lo[,'lwr'],col='red')
lines(lve,flood100lo[,'upr'],col='red')
flood100lofit<-(flood100lo[,'fit'])
flood100lolwr<-(flood100lo[,'lwr'])
flood100loupr<-(flood100lo[,'upr'])
flood100loerror<-(flood100loupr-flood100lofit)/flood100lofit
print(cbind(flood100lofit,flood100loupr,flood100lolwr,flood100loerror*100))
```