ISSN No:-2456-2165

Correlation of Different Weather Parameters using R Statistical Analysis Software for Swat River

Muhammad Irfan
MS Student at Iqra National University
Peshawar Pakistan
Dr. M. Mehboob Alam
Engr. HOD at City University
Peshawar Pakistan

Usama Habib Bangash
MS Student at Iqra National University
Peshawar Pakistan
Fawad Khan
Lecturer at Civil Engineering department
Iqra national University Peshawar

Abstract:- This thesis also review the correlation between discharge (cusec),rainfall (mm), gauge (ft.), daily minimum temperature (℃) and daily maximum temperature (℃) by using R statistical analysis software. The correlation matrix for the above multiple parameters also made for Swat region using Pearson, spearmen and Kendell correlation methods in R. This study also explain the correlation of rainfall vs discharge and gauge vs discharge for 2010 flood by using R. Correlations and graphical view of different statistical data produced different set of functions in R, these functions may be used for another river of the same features.

Keywords:- *Flood frequency analysis, Gumbel and extreme value theory type-1, Pearson, spearman and Kendell methods for correlation, Swat river annual peak discharge.*

I. INTRODUCTION

A. Brief introduction of Swat

Swat valley is present in the north of one of the province of Pakistan is Khyber Pakhtunkhwa at 35 degree latitude north and 72 degree and 30 degree longitude east. Swat total area is 5337 sq.km, there are two tehsils one is Matta and other is Swat which comprises of 683 sq.km area and tehsil Swat having 4654 sq.km. Out of total 5337 sq.km area 497969 acres area cover by forest. Swat majority area is comprises of mountains and located in the foothills of mountainous range of Hindukush. The Hindukush range spread from north towards south direction with different gradients in the Swat region. These gradient start from 600m in the south and reach to 6000m above the mean sea level towards north. During summer, the temperature is moderate and short in lower Swat while in upper Swat the temperature is cool and pleasant.

The month which is high temperature is June with average temperature are max 33C° and min 16C° respectively. The month which is coldest is January with average temperature are max 11C° and min -2C° respectively. Usually in Swat region winter season is for long term and starts from November ends at March; during this duration, snowfall and rainfall occur. In district Swat, average annual precipitation ranges from 1000mm to 12000mm.

B. Available Data Format

The data is available comprises of Gauge (ft.), discharge (cubic feet per second), daily maximum temperature (centigrade), daily minimum temperature (centigrade) and daily precipitation (mm). The data is available for the last 10 years which is from 2005 to 2015. The sample of the data is in table 1.

Date	Year	PPT(mm)	Daily Min Tem	Daily Max Tem	Discharge	Gauge
1/1/2006	2006	17	7	8	2444.8	2
1/2/2006	2006	21	7	11	2829.1	2
1/3/2006	2006	8	4	13	2829.1	2
1/4/2006	2006	0	3	14	2646	2
1/5/2006	2006	0	2	12	2829.1	2
1/6/2006	2006	0	2	18	2646	2
1/7/2006	2006	0	0	20	2646	2
1/8/2006	2006	0	9	20	2444.8	2
1/9/2006	2006	0	6	16	2444.8	2
1/10/2006	2006	0	6	16	2444.8	2
1/11/2006	2006	0	7	23	2444.8	2

Table 1. Format of the available data

C. Dealing with missing values

During collection of data, there must be some missing values due to human errors or some other technical problems or worst condition of weather where it may not be possible to take reading. Due to these reasons, many NAs values may incorporate in dataset. While using R the code may not run for the specific function when there is missing values (usually the missing values set by default -1000 in dataset). To remove the missing values from the dataset first import the data into R, find the summary of dataset for finding the number of missing values.

D. Time Series of different weather parameters

A time series is a sequence of observation of data points measured over a time interval. The observations ordered in time, as successive observation may be dependent. Time series plotted via line charts or scatter plots where time, is in X-axis as an independent variable on which we have low or no control and the data points plotted on Y-axis.

ISSN No:-2456-2165

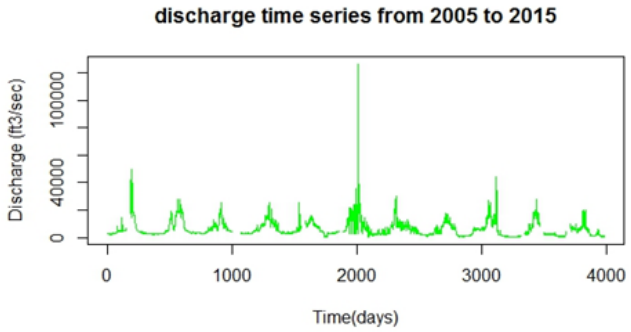

Fig 1:- Daily Discharge of Swat River for the last ten years from 2005 to 2015 using R statistical analysis software.

> plot.ts(first$`discharge (cuses)`, main= "discharge time series from 2005 to 2015", xlab="Time(days)", ylab="Discharge (ft3/sec)" ,col= "green")

E. Daily Precipitation

Usually rainfall measured in (mm). Rainfall event measured using different instruments like standard rain gauge and recording rain gauge, which is weighing and tipping buckets. Optical rain gauge also used for measuring rainfall. Daily Rainfall data for Swat region acquired from metrological department Lahore. In winter seasons, Western depression brings some amount of rainfall to northern and western areas of Pakistan in mid-December to March. Early summer is generally arid, rainfall varies between 25mm to 130mm. thunderstorm brings some rainfall to northern areas like Swat, Dir and Peshawar etc. from April to June. Late summer (monsoon) July to September heavy rainfall occur, while dry days and very little rainfall occurs in post monsoon from July to September. The following graph shows daily rainfall data at Swat region. The daily data initially converted into CSV format and imported into R statistical analysis program for further processing. Using the following code the graphical view of rainfall at upper Swat region.

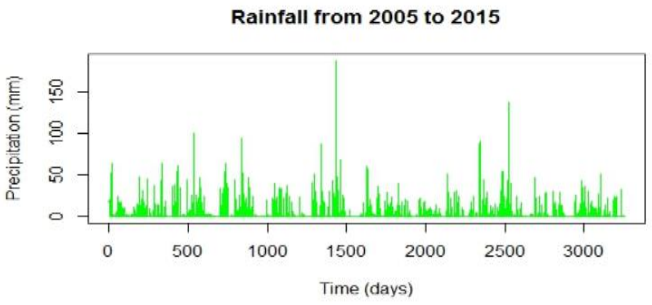

Fig 2:- Daily Rainfall using R software for the last 10 year from 2005 to 2015 of Swat region

> plot.ts(forboxplot$`ppt(mm)`, main="Rainfall from 2005 to 2015", xlab="Time (days)", ylab="Precipitation (mm)", col="green")

F. Daily Maximum Temperature

Temperature is the degree of hotness and coldness. It is average of hourly temperature. There are two type of device used to measure temperature one of which is Recording-type called thermograph with bimetallic strip or a metal filled with alcohol or mercury is used and temperature recorded on chart at every instant. Second type is non-recording, thermometer used for non-recording type. Mercury thermometer used for measuring high temperature while for low temperature alcoholic thermometer is used. There are other devices used to measure the temperature i.e. electrical resistance thermometer, Thermo couples, Gas bulb thermometer and many mores. For Swat region maximum daily temperature and minimum daily temperature, data in centigrade is acquired form the metrological department Lahore. By using the following code in R statistical analysis program for daily maximum temperature, we get the graphical view as below. The graph shows that some region in Swat where temperature goes beyond 40℃ each year.

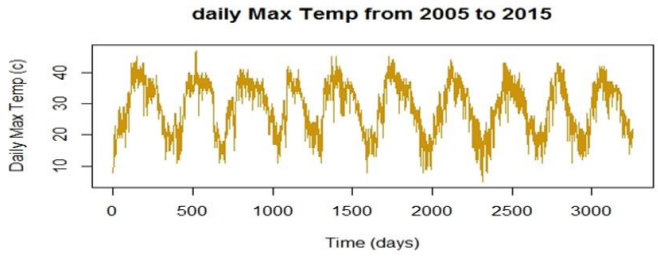

Fig 3:- Daily maximum temperature for last ten year from 2005 to 2015.

>plot.ts(forboxplot$`Daily Max Tem`, main="daily Max Temp from 2005 to 2015", xl ab="Time (days)", ylab="Daily Max Temp (c)", col="#CD950C")

G. Daily Minimum Temperature

In Swat region, there is sever clod in the months from Dec to Feb in which the temperature goes beyond 0℃ and a lot of snowfall occur in the upper region of Swat. Using the following code in R get the graphical view of minimum temperature at Swat region. The graph shows that there are some region in Swat where the temperature reaches to 0℃.

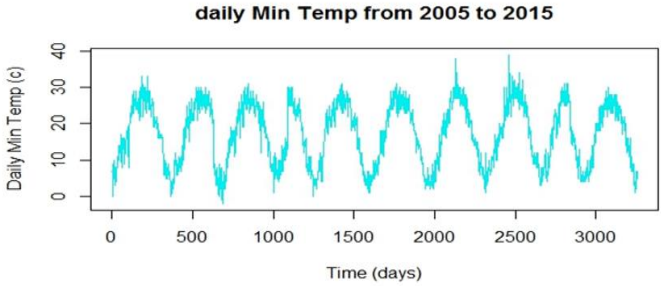

Fig 4:- Daily minimum temperature for the last ten years from 2005 to 2015

>plot.ts(forboxplot$`Daily Min Tem`, main="daily Min Temp from 2005 to 2015", xlab="Time (days)", ylab="Daily Min Temp (c)", col="#00EEEE")

ISSN No:-2456-2165

H. Stage Frequency

Stage is height of water level in a stream above some arbitrary zero datum at some specified station. This stage used to measure the discharge in open channels. It is difficult to make a direct, continuous measurement of the rate of flow in a stream but relatively simple to obtain a continuous record of stage. The primary field data gathered at most stream flow measurement stations that are "River Stage". Satisfactory only if there is an adequate correction between stage and discharge that is Rating Curve. For measuring stage value there are two types one is the recording type and second is non-recording type. The non-recording type comprises of vertical staff gauge, inclined staff gauge, section staff gauge, wire-weight gauge. The vertical staff gauge used for narrow sections. Inclined staff gauges graduated in such a manner that it directly gives the vertical stage. Section staff gauge used where there is no structure available, which is accessible to all stages. While Wire-Weight gauge is the number of revolutions of the drum counted whose circumference known (30 cm). The recording type gauges gives continuous record of stage at every instant of time. A float is used to known about stilling well level of water. The fig 3.5 shows the graphical view of different stages at their corresponding stage value. The graphs drawn using the following code in R statistical analysis program. The horizontal line shows that the discharge is same at that specific stage for many days.

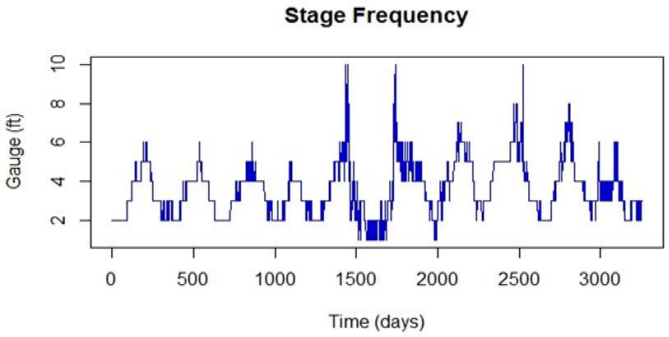

Fig 5:- Daily stage reading of Swat River for the last ten years from 2005 to 2015.

>plot.ts(forboxplot$guage, main="Stage Frequency", xlab="Time (days)", ylab="Gauge (ft)", col="#0000CD")

I. Problem Statement

Keeping in view the importance of correlation between different weather parameters using Pearson, Spearman and Kendall for daily Precipitation in mm, daily gauge readings in feet, daily maximum temperature in centigrade, and daily temperature in centigrade and daily discharge of Swat River for the last 10 years from 2005 to 2015. To build these relation R statistical analysis software with different unique codes for different graphs. These codes will be used for other rivers of the same features. These relation values will give the idea of how they depend each other.

J. Objective

Correlation matrix of different weather parameters using different method which is Pearson correlation matrix, Spearman correlation matrix and Kendall correlation matrix with the help of R statistical analysis software for ten year of daily discharge, daily maximum temperature, daily minimum temperature, daily gauge reading and daily precipitation.

II. METHODOLOGY AND DATA ASSUMPTIONS

A. Discharge and Gauge relationship

Discharge (or surface runoff) mainly horizontal water flow occurring at the surface in rivers and streams. It does not include the groundwater flow (or subsurface flow). Since discharge is only defined in streams or rivers. Many processes can cause surface runoff. Subsurface flow is also consider runoff (surface discharge) only in that case when the subsurface flow behave like stream flow under the ground and reaches to the surface water at some outlet. Surface runoff can also be depend on the topography, soil characteristic example is hydraulic conductivity which depends strongly on the soil wetness, hydraulic conductivity means that precipitated water infiltrate vertically into the ground. But when the intensity of precipitation is more causes not to percolate water into the soil, but all the water appears as runoff on the surface. In case of natural soil when excess of infiltration occur soil become fully saturated. The onward precipitation lead to excess runoff. Different methods to measures the discharge of river includes tracer method (instantaneous method), velocity area method with current meter, velocity area method using the dipping bar, bucket method, float method, manning equation and many more. In hydrology, graph between discharge and stage is called rating curve. The stage, which is the height of water in the given river usually noted by installing different gauge stations at suitable locations. While discharge in the given river measured across the stream channel with the help of flow meter or another most accurate method. Different methods to measure discharge of stream flow defend over the range of stream flow height. Rating curve is the graph plotted between discharge on x-axis and stage of the given river on y-axis in Cartesian coordinate system. The rating curve develop by following two steps. The first step is to establish relationship between stage and discharge by measuring stage and discharge by using different methods. When this relationship is not change with time called permanent control, when change with time is called shifting control. This change is due to erosion or deposition of sediments, which causes to change the stage measurement. The relationship establish in the first step used to calculate theoretical discharge just by measuring stage of the river. For plotting, such graphs first need to download GGPLOT package and install in R statistical analysis software. By using the following code in R

ISSN No:-2456-2165

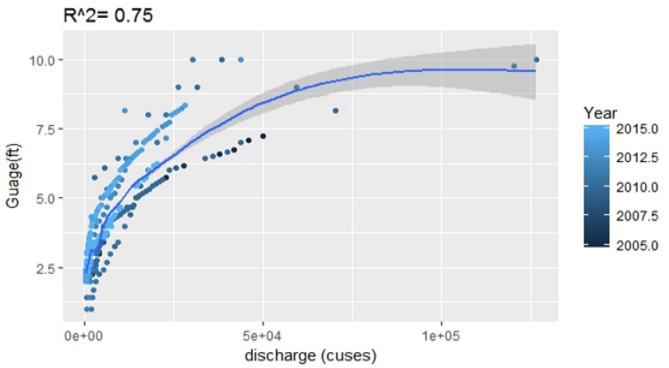

Fig 6:- Relation of daily discharge to gauge of Swat River for the last ten year.

> ggplot(first)+geom_point(aes(x=`discharge (cuses)`, y= `Guage(ft)`, colour= Year))+geom_smooth(aes(x=`discharge (cuses)`, y= `Guage(ft)`, colour= Year)) + labs(title="R^2= 0.75")

B. Daily Rainfall and daily discharge relationship

Stream flow of Swat River is changing continuously in each season from day to day even from hour to hour. The main factor of the stream flow is the precipitation on the upper Swat watershed causes runoff. Precipitation Results River to rise up to certain stage and the rise depend on the quantity and intensity of rainfall. The rise of river is also occur when there is rainfall occur far up region in the watershed, which remember that much runoff causes to rise up the river drains by the outflow point. Larger river means larger watershed surface area means that size of the river depend on the size of the surface area of watershed and smaller size river means smaller watershed surface area. Likewise different river size have different behavior to flood and rainfall. Large Size River have the slower rate of rise and fall while smaller size river have the greater rate of rise and fall. Smaller size watershed causes a flood 100 times as much water to flow in base flow each minute which cause to rise and fall the river in minute and hours, while large river take days to rise and fall but when rise causes flood for many days with greater destruction. In R, software using the following code get the graphical relationship between daily rainfall and their corresponding stream flow.

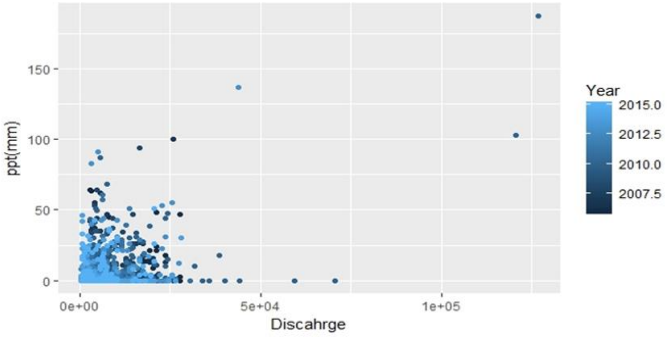

Fig 7:- Daily discharge and daily precipitation graphical relationship for 10 year from 2005 to 2015

ggplot(forboxplot)+aes(x=Discahrge, y= `ppt(mm)`, main="Discharge vs precipitation",colour= Year)+geom_point()

C. Discharge and Gauge relationship each year

Stream gauges continuously measured stage. This continuous record stage translated to river discharge by applying the stage-discharge relation (also called rating). Stage-discharge relations also developed for stream gauges by physically measuring the flow of the river with a mechanical current meter at a wide range stages. For developing a relation of stage-discharge, curve of Swat River by using 10 year data from 2005 to 2015. The data comprises of daily discharge data corresponding to their respective stage reading. As per available data, there are more values of discharge recorded at single stage value. In addition, single discharge reading there are different stage reading. The stage-discharge relation graph shows discharge of the river Swat in cubic feet per second when the stage is in feet. In the below graphs there are two different type of circle one is dark circle and the other is light circle. Dark circle represent concurrent measurement of more values of discharge at the same stage and the light circle represent the concurrent measurement of stage and discharge for one reading. In each graph per year shows, that there are few circles which are light and are at maximum gauge. These maximum gauge values connect with the maximum discharge of the respective year. That maximum value of discharge shows the maximum flood occur in that year. For example in 2010flood the discharge value is (126692 cubic feet per second) which is shown in 2010 stage-discharge graph on light circle. The below graphs shows the stage-discharge relationship each year. For drawing, the stage-discharge graphs use the following syntax in R-project software.

par(mfrow=c(2,2)) > plot(`discharge (cuses)`[Year=="2008"], `guage(ft)`[Year=="2008"], xlab="discharge(cuses)", ylab="gauge(ft)", main="Plot for Year 2008")

ISSN No:-2456-2165

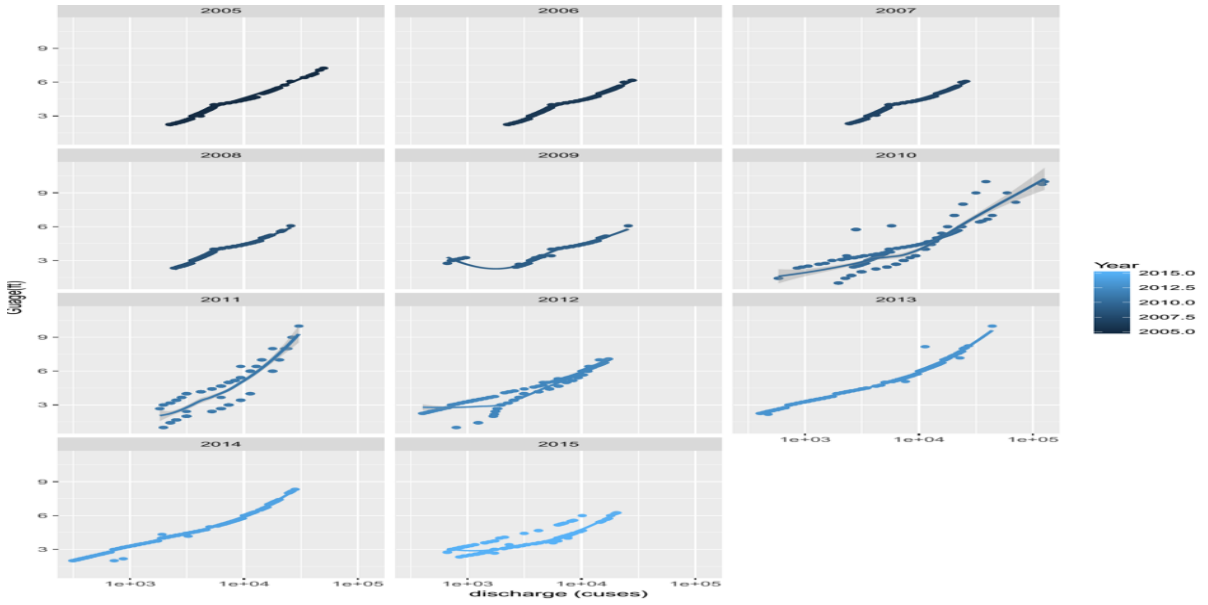

Fig 8:- Each year relation of daily discharge and daily gauge reading

III. CORRELATION OF DIFFERENT WEATHER PARAMETERS

A. Pearson correlation method

Correlation tool used to measure strength relationship between two variables. The correlation values between +1 and -1 indicate coefficient varies perfect relationship. When the correlation value near to 0 means that relationship becomes weaker. The different types of correlation are Pearson correlation, Kendall rank correlation, Spearman correlation and Point-Biserial correlation. Pearson correlation coefficient measure strength between different variable. A formula used to produce this coefficient value, which is between +1 and -1. Negative value shows that the relationship between variables is negatively correlated, positive coefficient value means that the relationship between variable is positive correlated.

$$r = \frac{[n\sum(ab) - (\sum a)(\sum b)]}{\sqrt{[(n\sum a^2) - (\sum a)^2][(n\sum b^2) - (\sum b)^2]}}$$

Where n= number of couples of variables.
∑ab= Sum of the product of couple variable.
∑a= Sum of a variables.
∑b= Sum of b variable.

Pearson Correlation degree		
From	To	Remarks
Near to ±1		Perfect Correlation
±0.5	±1	High degree relation
±0.30	±0.49	Moderate degree relation
±0.9 and below		Low relation
0		No relation

Table 2. Pearson Correlation degrees

B. Spearman Correlation Method

Spearman correlation method is a statistical measure to find the strength relationship between paired data. Formula used for Spearman correlation method as.

$$r = \frac{1 - 6\sum d^2}{n(n^2 - 1)}$$

Where d= ranks difference.
d^2 = squared ranks difference.
n= Number of observations.
Spearman correlation coefficient is in between -1 and +1 denoted by "r". Correlation value describes the strength using the following guide.

Spearman Correlation Ranges		
From	To	Remarks
0	0.19	Very Weak
0.2	0.39	Weak
0.4	0.59	Moderate
0.6	0.79	Strong
0.8	1	Very Strong

Table 3. Spearman Correlation Ranges

C. Kendall Correlation Method

Another method is Kendall correlation which is not a strong considerable test to find the strong suit between two dependent variables. Let assume two sample x and y, such that each trial quantity is n and we see that total number of sets formed from x*y is(n(n-1))/2. The following formula is used

to determine the correlation based on Kendall correlation value.

$$r = \frac{n_c - n_d}{0.5n(n-1)}$$

Where n_c = number when order in the similar manner.
And n_d = number when order in the unlike manner.

Correlation values when in-between 0.1 and 0.29 denote a small relationship, when in-between 0.30 and 0.49 shows an intermediate relationship and 0.5 and above denote a large relationship.

Correlation coefficient range is same for Pearson and Spearman lies between -1 to +1. Pearson coefficient +1 indicate that two dependent variable increasing with a constant rate, which occurs in the form of straight line. In this case the Spearman correlation coefficient is similarly +1 represent in the form of straight line. The Pearson correlation factor is positive however value is less than +1, while the Spearman coefficient quiet equals to +1 while one variable increases with the increase of another variable but the amount of increase is not consistent. For example Pearson = +0.8 but Spearman = +1.

When the data is random or the relationship is non-existent, the correlation coefficient is nearly zero. The relation between two variables is a perfect line with a decreasing slop for both correlations. For correlation coefficient using R function is cor( ) to find a correlation matrix of different weather parameters. The code used in R software is.

>cor (x, method = c (“pearson”, “kendall”, “spearman”))

Where “x” is data set or numeric data. “Method” means that one of the correlation methods used to find the correlation coefficient. Already known method for correlation coefficient is Pearson method, which is the linear relationship between two dependent variable. Where Kendall and Spearman methods are not parametric correlation methods.

```
cor ( x, method= “pearson”, used= “ complete.obs”)
res<- cor (disacharge_data)    round (res,2)
```

Using the given code in R statistical analysis correlation value between daily discharge (cusec) and daily precipitation (mm) is 0.3 while using Pearson correlation method, which shows moderate degree relation. 0.15 while using Spearman correlation method shows very week relation and 0.11 in Kendall correlation method shows small relationship. Relation between stage (ft.) and discharge (cusec) is 0.72 using Pearson method shows high degree relation. Which is 0.75 in Spearman correlation method shows strong relation and 0.6 in Kendall correlation method shows large association. Relationship between rainfall and daily minimum temperature is 0.02, 0.04 and 0.03 is Pearson, Spearman and Kendall correlation shows low degree, week and small relation respectively.

D. Graphical relation of different weather parameters

On the basis of available data of weather parameters which is daily precipitation data in (mm), day-to-day lowest temperature data in (℃), day-to-day extreme temperature data in (℃), day-to-day river discharge data in (cusec) and gauge reading in (ft.). All the parameters are combine for the same date using Microsoft Access. These parameters are plot against the daily discharge data. The first plot shows the relationship between the daily discharge (cusec) and daily precipitation (mm). The graph shows that more frequent precipitation (mm) occur in Swat region is in between (0 to 50mm) which causes frequent discharge from (0 to 25000 cusec). Few numbers of precipitation (mm) occur in the last 10 years are in between (50 to 100mm) which causes discharge in between (25000 to 40000 cusec). There are only two to three values of precipitation occur in the last 10 years specifically occur in year 2010 the rainfall was 164mm, which causes the maximum discharge not happened in the previous years (60000 to 170000 cusec).The second plot is between daily discharge (cusec) and daily maximum temperature (℃), and daily minimum temperature (℃). The plot shows that at the lower temperature (0 to 20℃) the discharge is also lower (0 to 5000 cusec). As the temperature rises (20 to 35℃), the discharge value also increases reaches to 170000 cusec. This maximum discharge occur due to maximum precipitation occur (164mm). Due to rise in temperature, causes melt the glaciers present in the upper swat region, which occurs in the form of discharge also; causes increase the stage in river Swat. This discharge is added to the discharge occur due to precipitation in swat region in this range of temperature.

- par(mfrow = c(2,2))
- plot(Combine_data$ Discharge, Combine_data$ ppt(mm)).
- plot(Combine_data$ Discharge, Combine_data$ daily Min temp)).
- plot(Combine_data$ Discharge, Combine_data$ daily Max temp).
- plot(Combine_data$ Discharge, Combine_data$ gauge).

Using the above code the graphs for combine data is shown as.

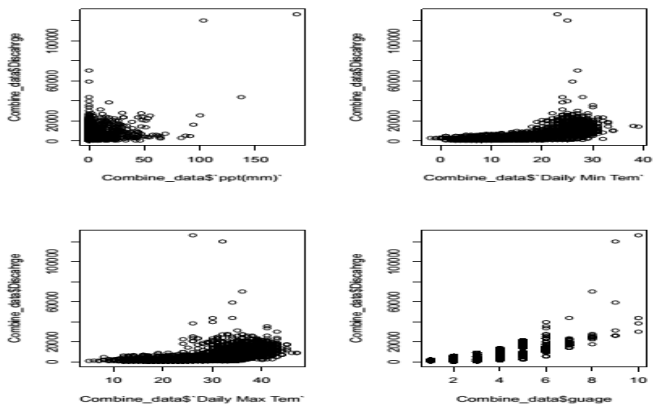

Fig 9:- Graphical relationship of different weather parameters

ISSN No:-2456-2165

IV. CONCLUSION AND RECOMMENDATION

A. Pearson Correlation Matrix

Using the following codes in R statistical analysis software the matrix obtained for Pearson method

```
cor ( x, method= "pearson", used= " complete.obs")
res<- cor (disacharge_data)    round (res,2)
```

Correlation Matrix using Pearson Method					
	Discharge (cusec)	Precipitation (mm)	Daily Max Temp (C)	Daily Min Temp(C)	Gauge (ft.)
Discharge(cusec)	1	0.3000483	0.5055827	0.575378	0.7202447
Precipitation(mm)	0.3000483	1	-0.09434546	0.02201245	0.1574657
Daily Max Temp©	0.5055827	-0.09434546	1	0.8919306	0.577406
Daily Min Temp©	0.575378	0.02201245	0.8919306	1	0.6882277
Gauge (ft.)	0.7202447	0.1574657	0.577406	0.577406	1

Table 4. Correlation matrix using R statistical analysis software for Pearson correlation

B. Spearman Correlation Matrix

Using the following code in R statistical analysis software obtained the matrix for Spearman method.

```
cor ( x, method= "spearman", used= " complete.obs")
res<- cor (disacharge_data)    round (res,2)
```

Correlation Matrix using Spearman Method					
	Discharge (cusec)	Precipitation (mm)	Daily Max Temp©	Daily Min Temp(C)	Gauge (ft.)
Discharge(cusec)	1	0.1492861	0.7105167	0.750419	0.75279
Precipitation(mm)	0.1492861	1	-0.1074895	0.04798654	0.129611
Daily Max Temp©	0.7105167	-0.1074895	1	0.8887582	0.619502
Daily Min Temp©	0.750419	0.04798654	0.8887582	1	0.72374
Gauge (ft.)	0.7527898	0.1296107	0.6195019	0.7237397	1

Table 5. Correlation matrix using R statistical analysis software for Spearman correlation

C. Kendall Correlation Matrix

Correlation Matrix using Kendall Method					
	Discharge (cusec)	Precipitation (mm)	Daily Max Temp©	Daily Min Temp(C)	Gauge (ft.)
Discharge(cusec)	1	0.1155024	0.5129532	0.5549135	0.600355
Precipitation(mm)	0.1155024	1	-0.08458666	0.03685426	0.109771
Daily Max Temp©	0.5129532	-0.08458666	1	0.7191484	0.476143
Daily Min Temp©	0.5549135	0.03685426	0.7191484	1	0.577407
Gauge (ft.)	0.6003549	0.1097713	0.4761428	0.5774072	1

Table 6. Correlation matrix using R statistical analysis software for Kendall correlation

In Pearson Correlation method the relation value of 0.72 for Swat River shows that there is a perfect correlation between daily gauge readings and daily discharge value. Means that most of the discharge flows within the streams. The extra water from flood due to precipitation overflows through the banks which cause to reduce the correlation value. The relation value 0.11 between precipitation and daily discharge shows that there is week relation between precipitation and discharge means that all the water in the Swat River is not due to the precipitation but this is a perennial River which flows throughout the year due to the melting of glaciers and from other ground water reservoirs. The different correlation value for the same parameters like gauge vs discharge is 0.72 in Pearson method, 0.75 in the Spearman correlation and 0.6 in the Kendall correlation method which is Perfect correlation, Strong and large relationship according to the above used methods.

There is a set of syntax for graphical view and for construction of different matrices which will be used to construct similar graphs and matrices of any type of parameters.

ACKNOWLEDGMENT

I am very thankful my thesis supervisor Engr.Dr.M.Mahboob Alam (specialization in Water resources engineering and management) was the chairman of Civil Department IQRA National University Peshawar. Professor Mahboob Alam was always guide me whenever I face trouble situation or any difficulty about my research. During the scheduled time. He always allowed this paper to be my own work, and put me in the right path whenever I diverted.

REFERENCES

1. Arora, H. & Singh, V. P., 1989, "A comparative evaluation of the estimators of the log Pearson (LP) typeIII distribution", Hydrol. 105, 19-37.
2. Pearson type three distribution", J. Hydrol. 110, 239-257.
3. Hydrology Department WAPDA, Lahore, 2005.
4. Irrigation & Power Department, Peshawar, 2002.
5. WAPDA. 2002, "Feasibility Study on the Development of Munda Dam Multipurpose Project in Islamic Republic of Pakistan", Japan International Cooperation Agency (JICA), Final Report, Vol 3