ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ

Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ

СОВЕТ МОЛОДЫХ УЧЕНЫХ

Студенттер мен жас ғалымдардың «ҒЫЛЫМ ЖӘНЕ БІЛІМ - 2014» атты IX халықаралық ғылыми конференциясы

IX Международная научная конференция студентов и молодых ученых «НАУКА И ОБРАЗОВАНИЕ - 2014»

The IX International Scientific Conference for students and young scholars «SCIENCE AND EDUCATION-2014»

2014 жыл 11 сәуір

11 апреля 2014 года

April 11, 2014

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ

Студенттер мен жас ғалымдардың
«Ғылым және білім - 2014»
атты IX Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ

СБОРНИК МАТЕРИАЛОВ
IX Международной научной конференции
студентов и молодых ученых
«Наука и образование - 2014»

PROCEEDINGS
of the IX International Scientific Conference
for students and young scholars
«Science and education - 2014»

2014 жыл 11 сәуір

Астана

УДК 001(063)
ББК 72
Ғ 96

Ғ 96

«Ғылым және білім – 2014» атты студенттер мен жас ғалымдардың IX Халықаралық ғылыми конференциясы = IX Международная научная конференция студентов и молодых ученых «Наука и образование - 2014» = The IX International Scientific Conference for students and young scholars «Science and education - 2014». – Астана: http://www.enu.kz/ru/nauka/nauka-i-obrazovanie/, 2014. – 5831 стр. (қазақша, орысша, ағылшынша).

ISBN 978-9965-31-610-4

Жинаққа студенттердің, магистранттардың, докторанттардың және жас ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті мәселелері бойынша баяндамалары енгізілген.

The proceedings are the papers of students, undergraduates, doctoral students and young researchers on topical issues of natural and technical sciences and humanities.

В сборник вошли доклады студентов, магистрантов, докторантов и молодых ученых по актуальным вопросам естественно-технических и гуманитарных наук.

УДК 001(063)
ББК 72

ISBN 978-9965-31-610-4

ВЫВОДЫ

В данной работе построена модель простейшей нейронной сети с помощью языка функционального программирования Erlang. Размер кода составил 100 строк с комментариями, сама модель после создания заняла менее 10 КБ в ОЗУ. Благодаря использованию в архитектуре Erlang модели акторов и легковесности процессов виртуальной машины BEAM, дальнейшее усложнение топологии нейронной сети не вызовет труда. Этому будут посвящены наши дальнейшие исследования.

Список использованных источников

1. Томпсон С. Программирование в Erlang / С. Томпсон, Ф. Чезарини // М.:ДМК Пресс, 2012. - 488с.
2. Gene I. Sher Handbook of neuroevolution through Erlang / Gene I. Sher // NY.:Springer, 2013. - 835 p.
3. Armstrong J. Concurrent Programming in ERLANG / J. Armstrong, R.Virding, C. Wikstrom, M. Williams // NJ.:Englewood Cliffs, 1996. – 205 p.
4. Логика мышления. [Электронный ресурс]. Режим доступа: http://www.aboutbrain.ru/

UDC004.6:004.42

ANALYSIS OF ENVIRONMENTAL AND ENGINEERING DATA USING MATLAB PROGRAMMING TOOL

Rakhmetov Almaz Muratovich

almaz.rakhmetov@gmail.com

Postgraduate at Eurasian National University, Astana

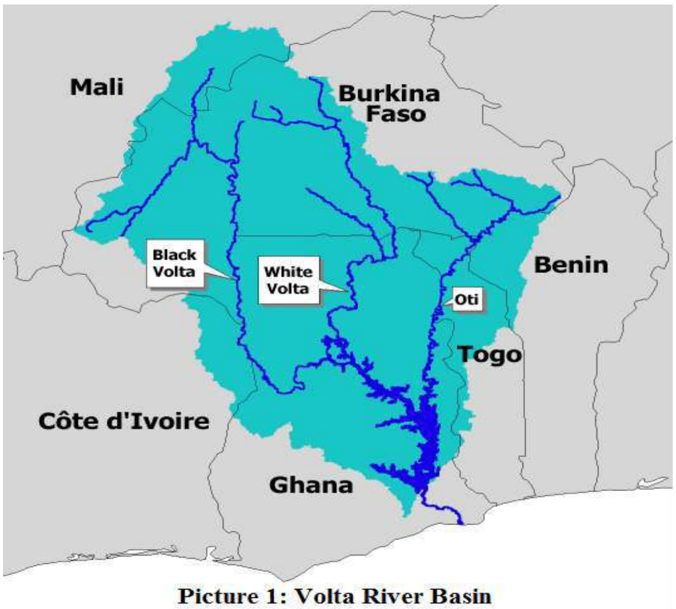
Supervisor – Sharipbayev A.A

Abstract - Rainfall and flow statistical datasets can provide decision makers with surface-water information needed for activities such as water quality regulation, infrastructure design, and water supply planning and management [3]. Therefore, it is important topic for environmental scientists. Scientists use different kinds of statistical methods to process data. It is hard to manually analyze and visualize the data they observed, therefore, they need to know different kinds of computing tools and programming tools for their studies. This paper is based on developing MATLAB functions in order to analyze and visualize Volta River hydrological data.

1. Introduction

Volta is one of the great rivers of Africa which is located in Volta Basin on the Western Africa. Volta Basin lies within latitudes $5^o.30$ N and 14^o 30 N and longitudes $2^o.00$ E and $5^o.30$ W and covers estimated 400 000 km^2. The basin lies mainly in Burkina Faso (43%) and Ghana (42%) with minor parts (15%) in Togo, Benin, Cote d'Ivoire and Mali (picture 1).

Volta River plays very important role for those countries. Volta River provided fish for citizens in basin proximity, and saved human lives when millions of people were dying due to the famine caused by a drought of 1970s. Moreover, a lot of dams to generate hydro-energy were built in the Volta River. For example, Akosombo dam



Picture 1: Volta River Basin

generates 80% of the power produced in the country. Therefore, environmental scientists always are making studies about Volta River Basin area. This paper also contributes to the Volta River study, particularly, the rainfall and stream flow amounts of the river.

The paper is organized as follows: Section 2 puts light on the data to be used for the analyses. The analyses of monthly statistics and annual basis are discussed in section 3 and 4 accordingly. Whereas sections 3 and 4 only analyzes rainfall and stream flow separately, in section 5 analyzes made by considering the correlation coefficient and line of regression between two data sets. Section 6 mentions conclusion and the future work.

2. Datasets

The two type of data used in the paper are the historical stream flow data and rainfall data. The stream flow data includes 48 years, from 1954 to 2001, which is provided by station Bui. Rainfall data includes 52 years, from 1954 to 2005, which is provided by Volta River Authority, Engineering Department. Stream flow data given in $x10^6 m^3/s$ (millions meter cubed per second), and rainfall data given in mm (millimeter). Data is stored in two separate Excel files. Data representation is based on month period. Since the rainfall data only contains data from 1954 to 2005 and the stream flow records contain data up to 2001, only year 1954–2001 will be included for comparison and relationship analyses from both datasets. In the following paragraphs ***monthlyRain*** is used as two dimensional array of rainfall data, and ***monthlyFlow*** is used as stream flow data.

3. Calculating Monthly Statistics.

MonthlyStatistics (code 1) function is used to calculate main statistical parameters of each month. These statistical parameters include mean, median, maximum, and standard deviation. MonthlyStatistics helps environmentalists to determine certain characteristics of certain months such as which one is wetter or dryer than others.

MonthlyStatistics function has only one input parameter called ***X*** which is two dimensional array where rows denote years and columns denote months. In this function simple MATLAB built in functions are used to compute mean, median, maximum, and standard deviation of every month. At the last line, *uitable* object is created to write all the statistics on spreadsheet.

Code 1: MonthlyStatistics.m:

```
function MonthlyStatistics(X);
…
Table(1,:) = mean(X,1);
Table(2,:) = median(X,1);
Table(3,:) = max(X,[],1);
Table(4,:) = std(X,1,1);
…
table = uitable('data',Table,'ColumnNames',columnnames)
```

Below given monthly statistics of rainfall (table 1) and stream flow data (table 2) While Table 1 clearly shows that months from April to October are wetter than other months, Table 2 shows that months through June to November have more stream flow amount than other months.

mm	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Mean	5.9	23.9	78.4	112.3	161.7	163.4	118	103.3	199.1	147.5	18.7	7.1
Median	0	17.7	65.3	110.6	151.1	159	104.4	94	195.4	135.9	9.1	0
Max	76.3	93.9	239.9	254	407.5	305.8	326.2	318	308.7	395.9	112.5	37.8
St Dev	13.8	24.8	49.4	44.6	71.9	51.2	74.7	59.8	64.9	72.4	23.7	10.8

Table 1: Statistical Parameters for Rainfall Data

$x10^6 m^3/s$	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Mean	29.1	13.8	8.3	9.3	29.3	83.9	187.1	418	837.7	595.9	175.8	66.3
Median	21.2	10	6.1	7.7	23.8	66.1	161.4	408.2	837.4	512.3	140.7	56.6

Max	87.6	63.3	53.5	30	138	343	619.1	1537	2417	1537	490.6	172.3
St Dev	22.5	12.4	9.1	6.2	22.6	66.2	127.3	237.9	413.5	324.3	111.2	40.4

Table 2: Statistical Parameters for Stream Flow Data

4. Periodical Mean Graph.

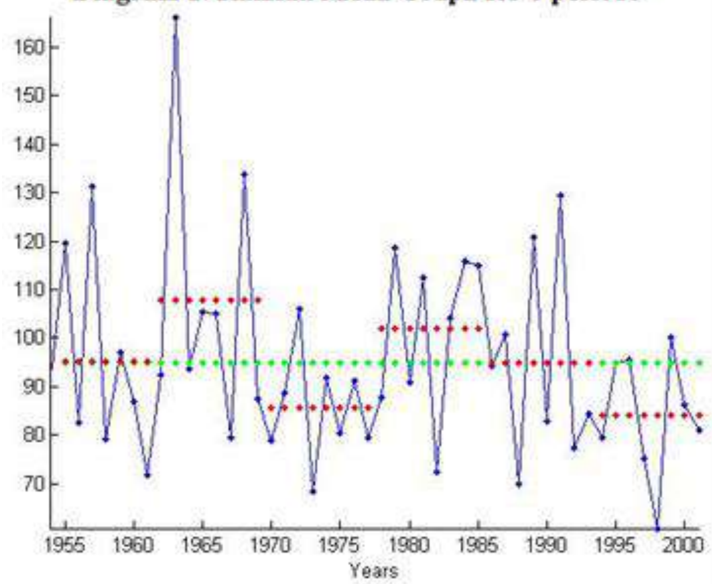
Whereas MonthlyStatistics calculates means on monthly basis, in the following section analyses were carried out on annual basis and the annual mean were obtained as average of monthly data for each calendar year. Graphical data presentation is more effective than tables of summary statistics. Therefore, DrawMeanGraph (Code 2) function plots a graph which shows how means of data are changing over a period of times.

DrawMeanGraph function has three input parameters such as ***X***, ***Y***, and ***period***. ***X*** and ***Y*** are column vectors and ***period*** is the number of equally divided population parts. ***X*** is vector of years, and ***Y*** is the data to plot. In Line 4 "Axis" function is used to limit the x and y axes of the graph to be plotted. "Hold on" is used to enable add plots to the one graph. To plot the graph MATLAB build in function "plot" is used. In Line 6, two sub graph parameters are passed to plot function. First one is to plot just yearly mean graph, and second one is to plot the mean of the whole years. Each sub graph consists X axis, Y axis, and string of characters. Lines from 8 to 15 contains algorithm which will plot the graphs of each period of years. Previous is the variable responsible for the last year of the previous period of years; while current is the variable responsible for the last year of the current period of years. ***XX*** and ***YY*** are x-axis and y-axis of the current sub graph.

Code 2: DrawMeanGraph.m:

```
1: function drawMeanGraph(X, Y, period)
…
4:axis([min(X), max(X), min(Y), max(Y)]);
5: hold on;
6: plot(X, Y, '.-', X(1:n), mean(Y), 'g.-');
7: previous = 0;
8: for i=1:1:period,
9:    hold on;
10:    current = round (n*(i/period));
11:    XX = X(previous+1:current);
12:    YY = mean(Y(previous+1:current));
13:plot(XX, YY, 'r.-');
14:    previous = current;
15: end
```

Diagram 1: Rainfall Mean Graph for 6 periods



For Volta River Basin case, DrawMeanGraph is used to plot the mean graph of rainfall data. In order to obtain rainfall mean graph, DrawMeanGraph function should be passed with years and ***monthlyRain*** data. For six periods of years function would be called as: ***drawMeanGraph(1954:1:2001, mean(monthlyRain, 2), 6).*** The result of this function is shown in Diagram 1.

Diagram 1 has blue, green, and blue dots. Blue dots denote rainfall amount in certain years. Green dotted line is the average of the whole rainfall data. Red dotted lines are averages of means of each period of years. Whole average helps to identify which periods of years where more dry or wet than other periods of years. Figure 2 shows that in Volta River Basin has been a clear change in time series in rainfall since 1970. This change is occurred due to the population expansion and increased land use practices after 1970s. This observation had earlier been made by Gyau-Boakye and Tumbulto [2]. However, according to the graph on Figure 2, rainfall amount significantly increased after 1980s, and average is going downwards year by year. Moreover, the rainfall amount considerably fallen down in the period from 1992 till 2000, which is almost equal to the 1970s severe drought period.

5.1 Least Square Line and Correlation Coefficient.

Functions described above only analyzes rainfall and stream flow separately. They do not show any relationship between two data sets. The following function called LeastSquareLine (Code 4), plots loess line that traces the center of the scatter plot cloud. Loess line helps to detect if the two variables displayed in the figure are correlated or independent to each other [5]. To find out loess line's equation least square line method is used. Shaefer and Theodore [4] used Least Square Line method with following equation:

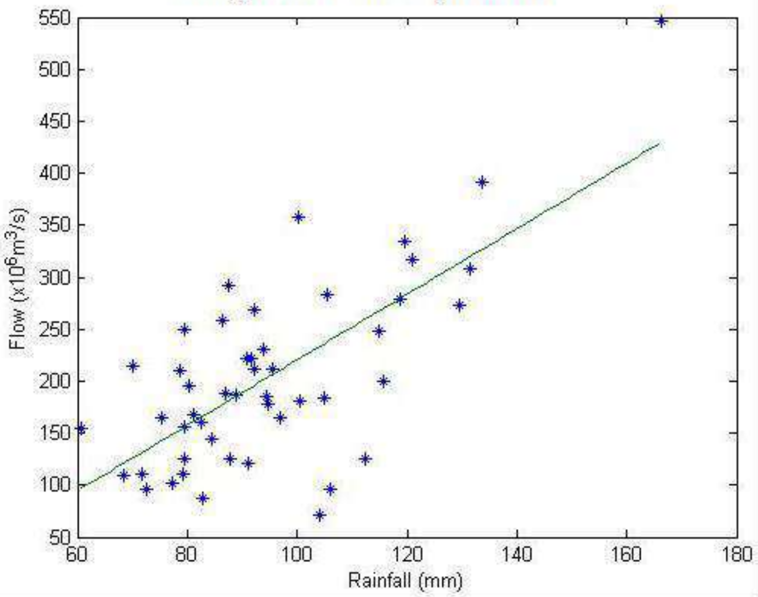
$$Y = a_0 + a_1X, \quad \text{where} \quad a_0 = \overline{Y} - a_1\overline{X} \quad \text{and} \quad a_1 = \frac{n\sum_{i=1}^{n} X_iY_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}$$

In order to use above equation, first thing to implement is finding five sums by MATLAB build in "sum" function. In the Line 3 and 4, sum(V) function is used to compute $\sum_{i=1}^{n} V_i$, and in line 5 and 6 sum(V.*E) is used to compute $\sum_{i=1}^{n} V_iE_i$. Then in line 8 and 9, a1 and a0 are calculated by using before computed "sums" and mean function. At the end of the LeastSquareLine, plot function is used to plot the regression of Y on X, and then correlation coefficient with significance level is computed.

Code 3: LeastSquareLine.m

```
1: function r = LeastSquareLine(X, Y)
2: n = size(X, 1);
3: sumX = sum(X);
4:sumY = sum(Y);
5: sumXY = sum(X. * Y);
6: sumXX = sum(X. * X);
7: sumX2 = sumX * sumX;
8: a1 = ((n * sumXY) - (sumX * sumY))
/((n * sumXX) - sumX2);
9: a0 = mean(Y) - (a1 * mean(X));
10: x = min(X):1:max(X);
11: y = a0 + (a1 * x);
12: plot(X, Y, '*', x, y)
13:xlabel('Rainfall')
14:ylabel('Flow')
15: [r, p] = corrcoef(X, Y)
```

Diagram 2: Least Square Line



For Volta River Basin case, LeastSquareLine discerns the relationship between rainfall and flow records. Therefore, two data should be passed as means on annual basis: ***LeastSquareLine(mean(monthlyRain, 2), mean(monthlyFlow, 2)).*** The result of this function call is shown in Diagram 2.

Diagram 2 has blue asterisks and a line. Every asterisk means that how much m^3/s flow amount is at the given amount of rainfall. Line is the result of the least squares line method equation. The line in the graph is tending upwards. It means that amount of stream flow is tend to increase when rainfall amount is increased.

5.2 Correlation Change.

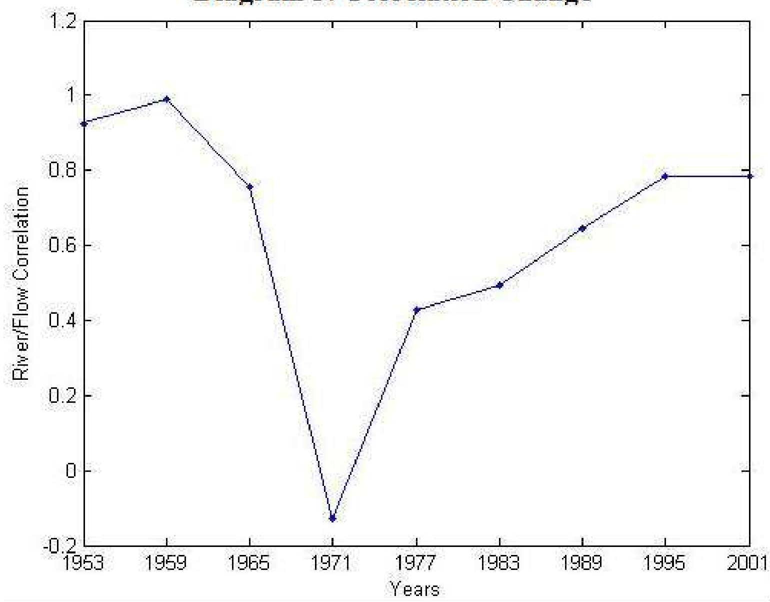
Following function shows how relationship between rainfall and stream flow is changing over years. Function is called DrawCorrelationChange (Code 5). The function plots a graph of correlation coefficient change in the given number of years.

This function has three input parameters. First two are column vectors ***X*** and ***Y***, and third one is a variable ***years***. Whereas, ***X*** is rainfall data and ***Y*** is stream flow data, ***years*** is the number of years in one period. First, number of years in the dataset is computed in line 2. Then from line 3 to line 12 small algorithm is written to compute correlation coefficient of each period. Algorithm started with "for" statement to iterate through each period of years. Variable ***ll*** denotes first year of the certain period, and ***rr*** denotes last year. Those two variables help to get the subset of data (certain period of years) from rainfall and stream flow data. Then MATLAB build in "corrcoef" function is used to calculate correlation coefficient of the selected period. The algorithm is repeated again with minor changes in lines 14 to 22, so that in case of last period should contain less years than user's given parameter the function would not give error message. At the last lines of the DrawCorrelationChange sets the label names of horizontal and vertical axis of plot.

Code 5: DrawMeanGraph.m:

```
1:function DrawCorrelationChange(X, Y, years)
2: n = size(X, 1);
3: for i=1:years:n,
4:    ll = i;
5:if (i + years - 1) <= n
rr = (i+years-1);
6:        elserr = n;
7:end
8:    meanX = mean(X(ll:rr, :), 2);
9:    meanY = mean(Y(ll:rr, :), 2);
10:   R = corrcoef(meanX, meanY);
11:r(ceil(i/years)) = R(1, 2);
12: end
13: k = ceil(i/years);
14: if i < n
…
22: end
23:plot([1:k],r,'.-');
...
```

Diagram 3: Correlation Change



The use of the DrawCorrelationChange for analyzes depends on how scientist wants to analyze the data. If scientist wants to observe changes occurring in small amount of years, then he or she passes small number as the ***years*** parameter. If he or she wants to analyze changes occurring in large periods of years, then he or she passes big number. To analyze Volta River Basin, appropriate call of DrawCorrelationChange is: ***DrawCorrelationChange(monthlyRain, monthlyFlow, 6).*** The result of this function call shown in Diagram 3.

Diagram 4 has eight periods, each contain six years. In the first three two periods the stream flow was highly dependent on the rainfall amount, because correlation coefficient was between 0.8 and 1. However, between years 1965 and 1971 relationship between rainfall and stream flow demonstrates abrupt fall to the negative correlation. Negative correlation coefficient means that there was not any relationship between rainfall and stream flow. Again, it is because of the change in the time series around 1970. At that period there was a severe drought in that area. The causes of the drought are loss of natural vegetation, poor water management, and global sea surface temperature changes, according to Aiguo Dai et al [1]. That is why rainfall is significantly decreased, and did little affect on stream flow. However, correlation coefficient is rising upward after drought period, which means that rainfall had some effects on stream flow in not dry areas.

6. Conclusion

Extracting an useful information from raw data and making analysis by means of contemporary programming tools is very useful for making decisions on different kinds of problems by observing the dynamic of progress or regression in particular situation as this paper concerns the environmental issue of Volta River Basin area. From the analyses given in this paper it must be predictable the periods of drought and wetness, and prepare specialists to prevent ecological problems of the area.

Four different approaches were made to analyze the data. First approach, monthly statistics analysis, points out that Volta River stream flow increases in summer and fall and decreases in other two seasons. It is mainly because of the low rainfall amount in colder seasons.

Second approach, periodical mean graph analysis, shows that rainfall amount is significantly decreased in 1970s, mainly because of bad water management. Therefore, countries that are neighboring with Volta River should use water and dams effectively. According to the plotted graph the rainfall average was decreasing starting from 1991. Therefore, 1970’s drought might be repeated in recent times if people are not going to effectively use water resources.

In the third approach, Least Square Line method was used to analyze the relationship between rainfall and stream flow data. Analysis showed that stream flow significantly depends on rainfall. That is why due to the decrease in rainfall amount, stream flow amount is decreases too in cold seasons. However, the relationship between two data is lower in the drought seasons.

The last approach analyzes the changes of the coherence between rainfall and stream flow over periods of years. The result shows that at the drought periods two data is less coherent than at the wetter periods.

In the future those kinds of data analyses with programming tools can be used to analyze the disastrous situation in Aral Sea and give useful information and solutions for environmental scientists and government to prevent the disappearance of the sea by effectively using water resources and dumbs in certain periods of time.

Literature

1. Aiguo Dai, Peter j. Lamb, Kevin E. Trenberth, Mike Hulme, Philip D. Jones, Pingoing Xie. (2004) *The Recent Sahel Drought is Real. International Journal of Climatology.*
2. Gyau-Boakye P. & Tumbulto J.W. (1999). The Volta Lake and Declining Rainfall and

   Streamflows in the Volta River Basin.*Water Research Institute, Accra, Ghana.*
3. Risley J.C., Stonewall A., Haluska T.L. (2009) Estimating flow-duration and low-flow frequency statistics for unregulated streams in Oregon. 1 pg.
4. Shaefer S.J., Theodore L. (2007) *Probability and Statistic for Environmental Science.* Boca Raton, Florida: CRC press. 287-304 pp.
5. Song S. Qian (2010).Environmental and Ecological Statistics with R. Duke University, Durham, North Carolina, USA. 25-47 pp.