ANALYSIS SOME GAUGING STATIONS IN IRAQ BY USING DISCORDANCY MEASURE AND HETEROGENEITY MEASURE +

تحليل بعض محطات قياس التصاريف في العراق باستخدام مقياس التباعد ومقياس التجانس

Ahmed Sami Nasser*

Abstract:

This research deals with the analysis of a group of gauging stations which consists of eleven gauging stations that are used to measure the discharge of Tigris branches within Iraq. They were supposed as a local region which has the same frequency distribution. The discordancy measure was applied in order to determine the deviated station within the region and the heterogeneity measure was applied in order to conclude whether the suggested region is homogeneous or not.

If the discordancy measure of each station is compared to the critical value, it will be concluded that there is no deviated station. However, ALTON COBRY station (no. 7) can be shifted to another region so that the whole region can be considered homogenous as the heterogeneity measure is (0.064) which is located within the accepted homogeneous limitations.

It was also concluded that there is a certain effect of the number of data for each station on the discordancy measure and sub-sequentially on the heterogeneity decision. Therefore, it was found that the lower the number of data for one station, the higher the number of deviated stations and sub-sequentially the higher the heterogeneity.

As a benefit of this research, the analysis which was applied to conclude the heterogeneity measure for the region, so can be found an adequate frequency distribution for a homogenous region and can be used to find a relationship between flood or drought discharges within return period especially for river branches that have no gauging station.

المستخلص:

هذه البحث يتناول تحليل مجموعة تتألف من احد عشر محطة قياس التصاريف لروافد نهر دجلة في العراق،وتم افتراضها منطقة(محلية) واحدة تملك نفس التوزيع الاحصائي ، باستخدام مقياس التباعد وذلك لتحديد المحطة الشاذة في المنطقة ومقياس التجانس ،لاستنتاج فيما لو كانت المنطقة المقترحة متجانسة ام لا .

اذا قارنا مقياس التباعد لكل محطة في المنطقة مع القيمة الحرجة سوف نستنتج انه لا توجد محطة شاذة. لكن محطة "التون كوبري" رقم (٧) يمكن ان تزحف الى منطقة اخرى.استنتج بان المنطقة متجانسة لان قيمة مقياس التجانس للمنطقة(٠,٠٦٤) وهي تقع ضمن حدود قبول التجانس.

كذلك استنتج ان هناك تأثير لعدد البيانات لكل محطة على مقياس التباعد وعلى قرار التجانس حيث وجد انه عندما تقل عدد البيانات لمحطة واحدة سيزداد عدد المحطات الشاذة وكذلك يزداد عدم التجانس.

+ **Received on 20/4/2010 , Accepted on 4/5/2011 .**

* **Assist lecture / Institute of Aumara technology**

الفائدة من هذا البحث لاستنتاج مقياس عدم التجانس للمنطقة ، وبالتالي لاستباط التوزيع الاحـــصائي الملائـــم للمنطقة المتجانسة وعندها يمكن ايجاد علاقة بين التصريف الفيضاني او تصريف الشحة مع فترة العودة وخـــصوصاً للروافد التي لا تملك محطة قياس للتصريف.

Introduction:

Extreme environmental events such as floods cause a lot of damage to life and property of human society also droughts cause damage to agriculture, so the analysis of frequency distributions is important.

[1] in this , it was selected type of distributions for specify site, and it was estimated the upper and lower limits for observed data according to selected type of distribution, so can find unusual points for data in site by graphical method ,and then can find the suitable type of distribution according to the number of unusual points , this procedure should be repeated for every sites.

A "region" is a group of sites each of which is assumed to have data drawn from the same frequency distribution as coated as by [2].

This study, it was focused on two critical points as important part in frequency analysis for max. Or min. discharge for some sites , the first one was the assignment of some sites to region and specify the unusual sites in a region based on L moment [3] , and the second was assessment of regional homogeneity based on L moment [4].

L moments:

The L-moment of x is defined as the functions of probability-weighted moments. Probability weighted moments are calculated from the ranked observations X_j an estimator of Probability weighted moments are defined as [2]:

$$b_r = n^{-1}\sum_{j=1}^{n} X_j \frac{(j-1)(j-2).......(j-1)}{(n-1)(n-2)........(n-i)} \quad ....1$$

n: number of data for each site
i: ith site
j: jth rank data inter ith site
for any distribution, the first four L-moments can be calculated from:

$$\left.\begin{aligned} l_1 &= b_0 \\ l_2 &= 2b_1 - b_0 \\ l_3 &= 6b_2 - 6b_1 + b_0 \\ l_4 &= 20b_3 - 30b_3 + 12b_1 - b_0 \end{aligned}\right\} \quad ....2$$

The L-mean, l_1 , is ameasure of central tendency and the L-standard deviation, l_2 , is a measure of dispersion , hence, the dimensionless L-moment ratios are defined by [3]:

$$\left.\begin{aligned} t &= l_1 / l_2 \; (L - Coefficient\ of\ variation, L - C_V) \\ t_3 &= l_3 / l_2 \; (L - Skewness) \\ t_4 &= l_4 / l_2 \; (L - Kurtosis) \end{aligned}\right\} \quad ....3$$

Discordancy measure:

In this test, L-moment ratio (L-coefficient of variation (L- C_v), L-skewness, and L-kurtosis) of a site is used to describe that site in three –dimensional space [2]. A group of homogeneous sites may form a cluster of such points. A point, which is far from the center of the cluster, can be discarded. Discordancy measurement of site can be calculated by D_i [2], which is :

$$D_i = \frac{N}{3(N-1)}(u_i - \bar{u})^T S^{-1}(u_i - \bar{u}) \quad \text{.......4}$$

Where:
N=number of sites

$u_i = \begin{bmatrix} t^i & t_3^i & t_4^i \end{bmatrix}^T$ vector of L- moment ratios for site containing the t, t_3 and t_4 values.

Mean of vector u_i : $\bar{u} = N^{-1}\sum_{i=1}^{N} u_i$ 5

S : covariance matrix of u_i is define as:

$$\left.\begin{aligned} S &= N^{-1}\sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T && \mathit{for} \quad N < 30 \\ S &= (N-1)^{-1}\sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T && \mathit{for} \quad N \geq 30 \end{aligned}\right\} \quad \text{.......6}$$

The site I is declared to be discordant, if D_i is greater than the critical value, of the discordancy statistic given in a tabular from by [5] As shown in table (1)

Table (1) critical values for the discordancy statistics D_i

N	α /N	INV. F	Di
5	0.02	1350.50466	1.333
6	0.016667	59.16588856	1.648
7	0.014286	22.96752761	1.917
8	0.0125	14.7659418	2.140
9	0.011111	11.48242917	2.329
10	0.01	9.779538241	2.491
11	0.009091	8.757888276	2.632
12	0.008333	8.08558785	2.757
13	0.007692	7.614226793	2.869
14	0.007143	7.268227632	2.971
>=15	0.006667	7.005312663	3.000

Hosking and Wallis [2] initially suggested the criterion $D_i \geq 3$, but this is not satisfactory for small regions. We can see that D_i satisfies the algebraic bound[4] $D_i \leq (N-1)/3$ .for a test with significance level α (take α =0.1) an approximate critical value of $\max_i D_i$ is (N-1)Z/(N-4+3Z),where Z is the upper 100α /N percentage point of an F distribution with 3 and N-4 degrees of freedom[4] .

Heterogeneity measure:-

The idea underlying [2] heterogeneity statistics is to measure the sample variability of the L-moment ratios and compare in homogeneous region.

Suppose that N sites of observations of the same variable in different measurment sites are available, and that one wishes to verify if they can be grouped form a statistically homogeneous region:

Let $X_{i,j}$ be the jth observation in the ith sites , sorted in ascending order

$(X_{i,1} \leq X_{i,2} \leq \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \leq X_{i,ni})$ where i=1,2,…………N

After that calculate the L- coefficient of variation (L-C_v), coefficient L-skewness, and coefficient of L-kurtosis for each site from eq. (3).

The regional averaged L-C_v , L- skewness ,and L-kurtosis coefficients are

$$t^R = \frac{\sum_{i=1}^{N} n_i t^i}{\sum_{i=1}^{N} n_i} \qquad t_3^R = \frac{\sum_{i=1}^{N} n_i t_3{}^i}{\sum_{i=1}^{N} n_i} \qquad t_4{}^R = \frac{\sum_{i=1}^{N} n_i t_4{}^i}{\sum_{i=1}^{N} n_i} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 7$$

And the weighted standard deviation

$$V = \left\{ \sum_{i=1}^{N} n_i (t^i - t^R)^2 / \sum_{i=1}^{N} n_i \right\}^{\frac{1}{2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 8$$

An heterogeneity measure, which is called θ_{HW_1}

$$\theta_{HW_1} = \frac{V - \mu_V}{\sigma_V} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 9$$

θ_{HW_1} can be approximated by normal distribution with zero mean and unit variance [1]. Following [4], the region under analysis can therefore be regarded as "acceptably homogenous" if θ_{HW_1} <1 ,"possible heterogeneous" if 1< θ_{HW_1} <2 and "definitely heterogonous" if $\theta_{HW_1} \geq 2$.

Case analysis:

In this study, it is take max. annual discharge for eleven gauging stations in Iraq from 1960 to 2007 for" MANKUBA" station, "BALKEIAN" station,"GENDEIN" station ,"ASKI-KALK" station," ZAKHO" station ,up stream" DOKAN "dam station," ALTON-KOBRY" station," KARKUK" station, up stream "DERBENDEKHAN "dam station, "DALYA" up stream HMREEN dam station, and "ENGANA" station [6] and[7]. The data are shown in fig.(1a) and the location of them are illustrated in fig(1b).The number of data available for every sites and the mean for every them are illustrated in table(2).

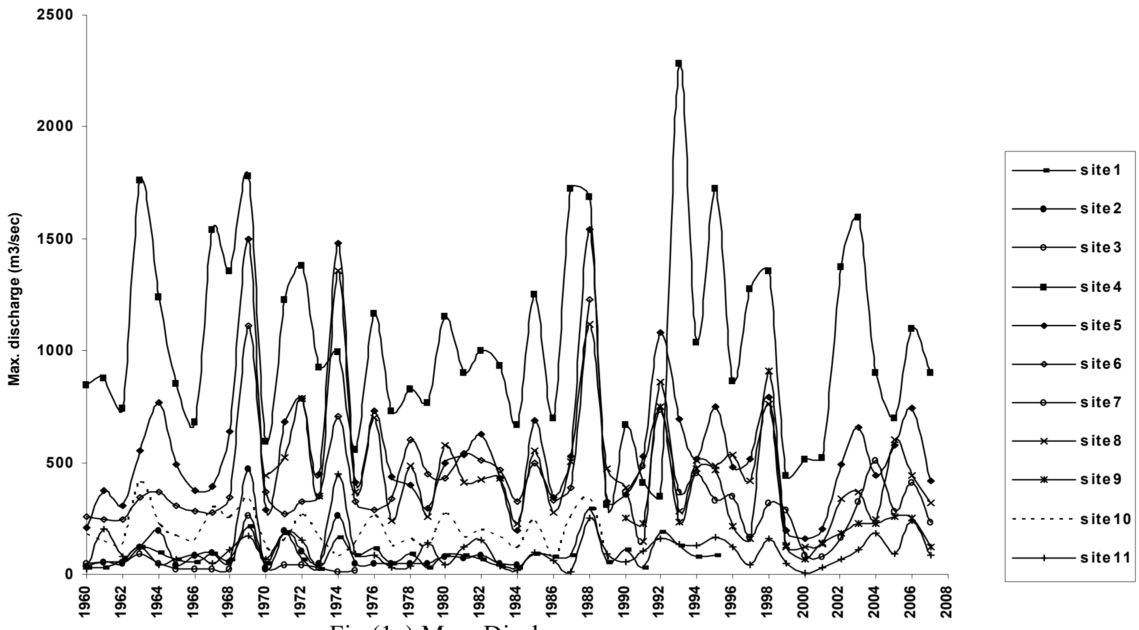


Fig.(1a) Max. Discharge vs. year.

Fig.(1a) Max. Discharge vs. year.

.

Fig.(1b) location of gauging stations

Table (2) no. of data for each sites and the magnitude of mean

site	n	Mean
1	36	90.083
2	25	97.160
3	16	53.063
4	48	1026.021
5	30	559.625
6	48	431.379
7	29	331.222
8	28	463.526
9	38	293.889
10	18	192.567
11	48	105.708

Results and discussion:

1- It is written program by using BASIC language to compute the (D_i) and (θ_{HW1}).

It is found in the first step that t, t3, and t_4 for every gauging stations .these results have been calculated by applying the program, Moreover Di has also calculated for eleven gauging stations, these results are stated in table (3) . The group average L-moment values of Lcv, L skewness and L Kurtosis are t^R=0.3120, t_3^R=0.2903 , t_4^R=0.2410.

Fig.(2) states that the point that marked by ( ) and its coordinates the group average L-moment have the values of Lcv and L skewness .Gauging station no.7 have a great discordance, Therefore this point far away from the group average L-moment values. It is also see that the gauging station no. 1 is near the group average L-moment values .From table (3) can be seen that this point have min. discordance.

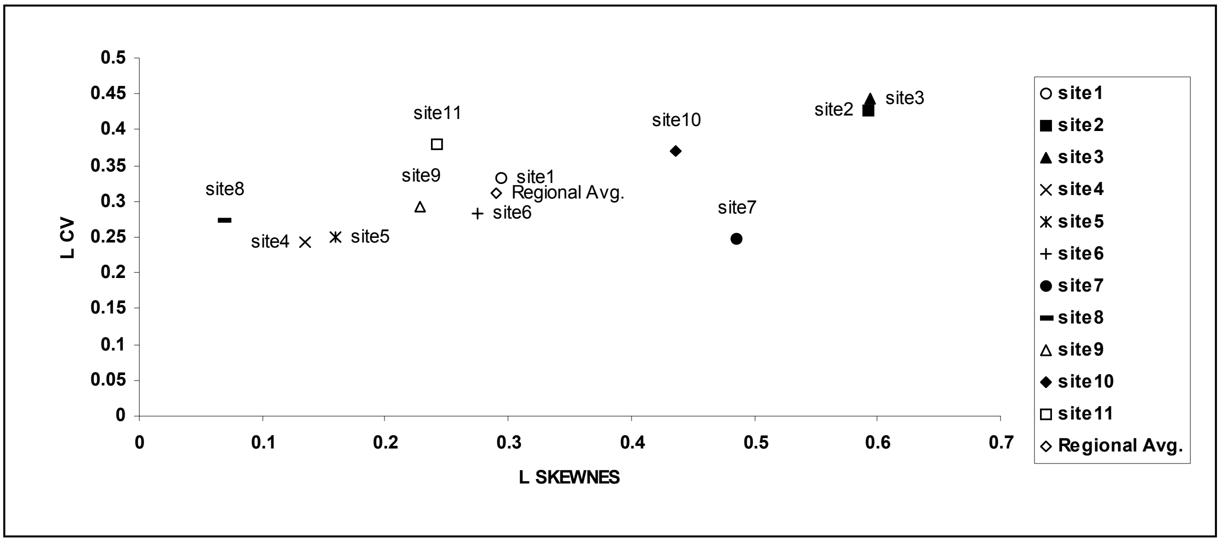
There is no evidence of gross errors in the data, when we comparing di values with critical value D_i which state in table (1). It could be concluded that there is no site is cordance whereas gauging station no.7 can be shifted to another region because this point has largest d_i in a region. This can be concluded when we compare t,t3, and t_4 for this site with the group average L-moment values of Lcv , L skewness and L Kurtosis which was very discordance.

The second step, It was tested the region for heterogeneity the value θ_{HW_1} =0.064 Therefore, the region is an acceptably homogenous.

Table (3) Results of t, t_3, t_4, and D_i for each site

i	t	t3	t4	d
1	0.3337	0.2949	0.233	0.1448
2	0.4276	0.5925	0.3921	1.2779
3	0.4444	0.5938	0.5272	1.8862
4	0.2418	0.1351	0.1072	0.7141
5	0.2497	0.1593	0.1275	0.5598
6	0.2822	0.2747	0.2475	0.1396

7	0.2468	0.4848	0.3369	2.3121
8	0.2732	0.0686	0.2198	1.9631
9	0.2917	0.228	0.2492	0.2624
10	0.3704	0.4361	0.3518	0.1891
11	0.379	0.2414	0.1731	1.551



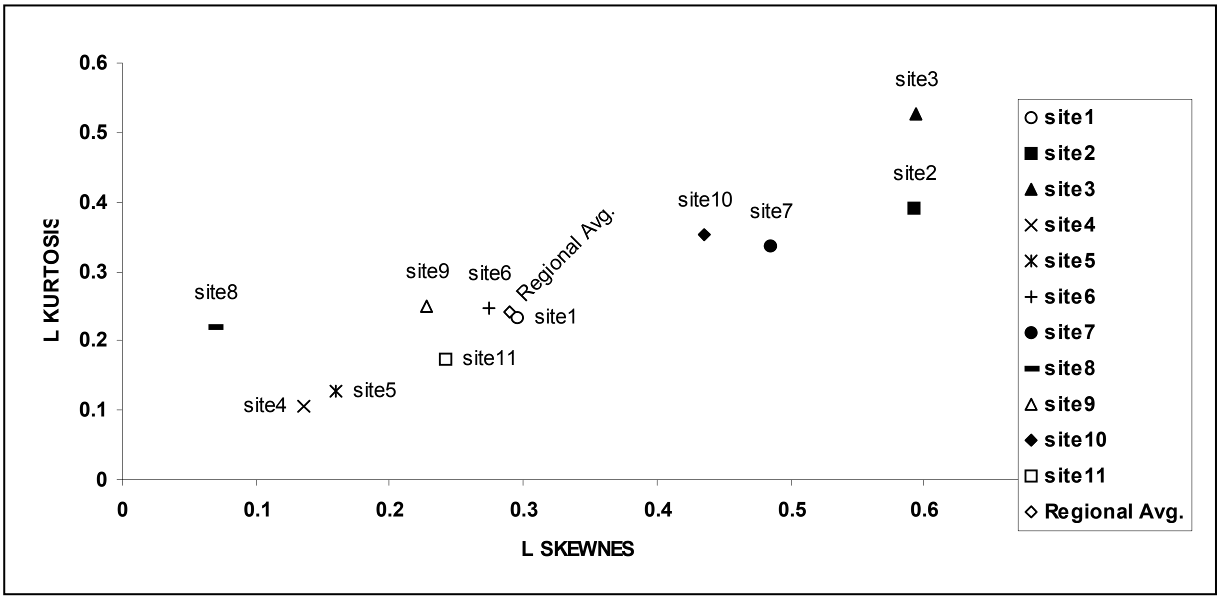


Fig.(2) L moment ratio for five gauging stations in Iraq.

2-For the purpose of study the effect of data number for every sites on decision heterogeneity. It was assumed that data number for every sites are constant except the site no.1, The number of data was taken 36,30,20,10, and 5.

It is found in table (4) that the decrease (n) for site no.1 this will increase the number of sites discordance .

If compare site no.1 that have n=5 with site no.1 that have n=36 ,it is concluded that there are seven sites that have increased(d) moreover the gross error increase if we decrease the data number and the analysis will be more accurate if we take long data for every sites .

This is very clear and it is illustrated in table (5) which shows that the in formations are stated as follow: L CV and L skewness are increased for site no.1 when the n=5 .Table (6) is also indicated that (θ_{HW1})will increase if decrease data number .

Table(4) Results of (d) for each site with variable (n) for site 1

n	36	30	20	10	5
site	d				
1	0.1448	0.0349	0.1263	0.8438	2.8320
2	1.2779	0.6267	0.2522	0.0287	0.0880
3	1.8862	1.2075	2.0540	1.1958	0.6375
4	0.7141	1.5429	1.5253	2.2097	2.2029
5	0.5598	1.7023	1.7275	1.3667	0.6459
6	0.1396	0.0640	0.1915	0.1131	0.1975
7	2.3121	2.7551	0.9031	0.5319	0.6580
8	1.9631	1.8839	1.7187	2.9949	2.7020
9	0.2624	0.2323	1.1851	0.4366	0.4720
10	0.1891	0.1896	0.5141	0.9758	0.4070
11	1.5510	0.7609	0.8023	0.3031	0.1573

Table(5) Results of Lcv,L skewness and L kurtosis for group with variable (n) for site1

n	36	30	20	10	5
L CV	0.3120	0.3534	0.3849	0.3872	0.3877
L SKEWNESS	0.2903	0.2876	0.3050	0.3109	0.3410
L KURTOSIS	0.2410	0.2275	0.1979	0.2060	0.2182

Table (6)values of heterogeneity measure for group with variable (n) for site1

site 1	
n	θ_{HW1}
36	0.064
30	0.062
20	0.082
10	0.103
5	0.126

Conclusions:

1-The gauging stations are regard it has data drawn from the same frequency distribution and the region is homogeneous. Therefore it can be found the suitable type of distribution and then estimate the relationship between flood or drought discharge with return period for this region.

2-Analysis any region should be taken long data to get more accurate.

References

1- Kite, G.W., *Frequency and risk analysis in hydrology*, water resources publications, Colorado, 1977.

2- Hosking, J.R.M., and Wallis,J.R. *"Some statistics useful regional frequency analysis " water resources research*, Vol.29, No.2,pp271-281, 1993.

3- Hosking, J.R.M. *"L-moments: analysis and estimation of distributions using linear combinations of order statistics" J.R.Stat.Soc.*, Ser.B, 52,pp105-124, 1990.

4- Viglione,A.,Laio,F.,and Claps,P. *" A comparison of homogeneity test for regional frequency analysis" water resources research*. Vol.43, W03428, 2007.

5- Hosking, J.R.M., and Wallis, J.R., *Regional frequency analysis: An approach based on L –moment, Cambridge University press*, cambrige, UK. , 1997.

6- Ministry of Irrigation, *Discharges for Selected Gauging Stations in Iraq (1975-1990)*, .D.G. Maintenance and Operation of Irrigation Projects, Ministry of Irrigation, Baghdad, 1990.

7- Ministry of Irrigation, *Discharges for Selected Gauging Stations in Iraq (1990-2007)* compacted disk, .D.G. Maintenance and Operation of Irrigation Projects, Ministry of Irrigation, Baghdad, 2010.