

Identification of Precipitation Types by Cluster Analysis Method (Case Study: Zahedan. Iran)

H. Nazaripour

PhD Student in Climatology, University of Sistan & Baluchestan.Iran.

M. Khosravi.*

Associate Professor of Climatology, University of Sistan & Baluchestan.Iran.

Abstract

In this paper daily Precipitation Data's from 1.1.1966 to 31.12.2005 examined for identification precipitation types. First, days of precipitation extracted. Then seven precipitation and four temporal parameters was defined for them and finally, matrix of 616*11 for 901 precipitation days were created. A hierarchical cluster analysis by Ward method applied on this matrix and five different precipitation types were identified. According to this classification, Zahedan precipitation types are: 1-medium, severe and frequent occurrence precipitation type 2-Very low, very short term, more frequent occurrence and very quiet type 3-Low, long term and quiet type 4-High, low occurrence and long durability type 5-Very high, low durability and very severe type. These five types, particularly displayed the Zahedan region precipitation characteristics.

Key words: precipitation, cluster analysis, precipitation type, Zahedan.

Introduction

The cluster analysis, one of the statistical methods, is used in reducing data and finding real groups. The purpose of cluster analysis application is creating homogenous groups from different people (Greshten Garbeh, 1999: 148). This analysis is going to be done by two methods: the first one is hierarchical method and the second one unhierarchical. In hierarchical method, the numbers of groups in every level is different and each person belongs to a special group and there is no possibility of their displacement (Bacher, 1996: 424). In unhierarchical method, each person does not belong to any special group but the degree of their

reference to each one of these groups is determined (Stat Soft, 1994: 323).

Cluster analysis is mostly implemented by climatologists for various purposes relevant to climates of different geographical scales. Kaufman and Rousseuw have studied the introductions of categorizing data by application of cluster analysis (Kaufman & Rousseuw, 1990:168). Domroes identified three principal components and five precipitation regimes in Iran by the application of principal component analysis and cluster analysis on monthly precipitation of 71 stations in Iran (Domroes and et al, 1998: 152). Jackson has analyzed the comparison of results obtained from the application of different

methods of cluster analysis in climatic categorizing of pluviometry stations in the district of Hare (Jackson and et al, 1998: 9914). Romero have analyzed the categorization of atmospheric circulation patterns leading to heavy precipitation in Spain (Romero and et al, 1999: 775). Singh has analyzed the principal components of high, medium and low precipitation years with the help of cluster analysis (Sing, 1999: 642). Littmann has categorized the 500 hpa level by the implementation of cluster analysis of pressure related data and Geopotential height and has studied the relation of obtained synoptic types and Mediterranean basin precipitation (Littmann, 2000: 168). Baldwin has applied the cluster analysis method for categorizing of precipitation areas (Baldwin and et al, 2002: 3). Alijani has classified the thermal areas by the application of cluster analysis method (Alijani, 1993: 87). Heidari and Alijani have introduced the first climatic classification in Iran by performing multivariable analysis (Heidari and Alijani, 1999: 63). Torabi have classified Iran into five climatic areas (Caspian, cold mountainous, hot semi-dry, hot dry and hot dry coastal) by using of cluster analysis method (Torabi and et al, 2001: 44). Masoodian (2003: 80) has identified three factors by applying factor analysis on monthly precipitation of 120 stations in Iran which is the same with the spatial scope identified by Domroes (Domroes and et al, 1998: 152). Atayi has zoned the precipitation areas in Iran with the help of statistical methods (Principal component analysis, factor analysis and cluster analysis) and has studied the comparison

of obtained results for each one. He classified precipitation areas in Iran into seven precipitation zones by the implementation of cluster analysis (Atayi, 2004). Gerami Motlagh has zoned the climatic areas in Bushehr province by cluster analysis and has identified six climatic zones (Gerami Motlagh, 2004). Masoodian has identified twelve different precipitation regimes in Iran, which clarify more details in comparison with five precipitation regimes of Domroes (1998: 152), by the way of cluster analysis of monthly precipitation percentage in Iran (Masoodian, 2005: 47). Masoodian and Atayi have identified five precipitation areas with performing cluster analysis of monthly precipitation of near to half of the century in Iran (Masoodian and Atayi, 2005: 1). Mohammadi and Masoodian have analyzed the synoptic types of Sanandaj station with the help of cluster analysis and have related them to circulatory patterns of 500 hpa level (Mohammadi and Masoodian, 2007: 39). Nazaripour has studied the synoptic of heavy precipitation in Bushehr province and has typified the precipitation in Bushehr province with performing of cluster analysis (Nazaripour, 2007). Nazaripour and Khoshal have conducted some investigation on synoptic types of Naein station's climate (Nazaripour, 1386: 114). They also have studied the synoptic types of Khor biabanak Station and with the help of cluster analysis, identified three distinguished types for it (Nazaripour and Khoshal, 2007: 27). Nazaripour and Khosravi have investigated the important effects of patterns leading to extreme precipitation for reducing the flood

damage in Bushehr province and with performing of cluster analysis, have classified the characteristics of precipitation which lead to floods (Nazaripour and Khosravi, 2009: 112).

Disregarding practical benefits, the identification of climatic synoptic types in Iran, which includes various climates, based on either one variable or multi variables is theoretically influential and it makes the comprehension of reasons for existing different climates and their effective factors much better. The identification of precipitation types can make the characteristics of precipitation in a station and district more tangible and it clarifies the reason of fluctuations in precipitation as well. Consideration of the characteristics of each precipitation type can help us with understanding of other environmental phenomena.

Data and Methodology

The data used in this paper is the information relevant to daily precipitation of Zahedan station from January 1, 1966 up to December 2005. First of all, the characteristics of precipitation including the range of annual change (figure 1), monthly precipitation (Figure 2) and seasonal distribution of precipitation (Figure 3) and some statistical characteristics of precipitation (average, standard deviation, coefficient of variability and variance) (Table 1) and rainfall totals (Figure 4) have been considered for getting a general view on Zahedan precipitation. Then the precipitation days in Zahedan (more than 0.1 mm) have been extracted. The frequency of precipitation days during the study is 901 days were calculated.

Frequency of rainy days for month (Figure 5) and days (Figure 6) of the year are plotted. Based on these, the matrix of 901×1 for precipitation days reckoned. Then, the parameters of the starting year and month of precipitation, ending of it, the monthly frequency, peak monthly precipitation, the sum of monthly precipitation, the daily durability of precipitation, peak daily precipitation, the sum of daily precipitation and daily precipitation intensity for precipitation days are considered (Table 2). According to these parameters, the initial matrix changed into 616×11 . It is important to know that in Table 2, the last column namely the precipitation types is the output of cluster analysis and it is not any components of initial matrix. The variables of year, month, the starting and ending day are out of the process of analysis. Finally, these set-aside variables are considered for final analysis. From now on, this matrix is set as the principal one in precipitation types of Zahedan. Because the data includes different units so standardization after the process of analysis is important to unify the weight of each variable for separating the precipitation types. Here, because the purpose is unifying the weight of variables, the following process is to be considered (Equation 1).

1: Standardizing Method:

$$STND_{ij} = \frac{Data_{ij} - Min_j}{Max_j - Min_j}$$

$STND_{ij}$ the standardized value of variable j in the day of i ; the value of variable j in the day of i ; Min_j the

minimum value of variable j ; Max_j the maximum value of variable j . In standardization, the value of each data is subtracted from its mean and divided by its standard deviation. The mean of standardized matrix is zero and its SD is one.

So the matrix ($std611*7$) is taken as the base for determining the Euclidean Distance. Because there was no idea for the numbers of groups before categorization, performing of cluster analysis for identification of groups would seem practical. Accordingly, all the variables would be compared with each other one by one in order to determine the degree of their similarity and then they all would be clustered based on their degree of similarities. Therefore, there are two outstanding steps in cluster analysis: the first one is determining the degree of similarities between each person and the second one is the kind of merging persons based on their degree of similarities. Based on the way chosen for accounting the level of similarities and the kind of merging, a cluster analysis can be performed by different methods. There are different ways for determining the degree of similarities. In climatology, when the scale of variables and domains like the present study's data are different, for determining the degree of dissimilarity the Euclidean Distance is used like the following (Equation 2).

2-1: Euclidean Distance Computation:

$$d_{rs}^2 = (X_r - X_s)(X_r - X_s)$$

2-2: Euclidean Distance Computation for standardized

$$\text{matrix: } d_{rs}^2 = (X_r - X_s)D^{-1}(X_r - X_s)'$$

Consequently, after determining the value of Euclidean Distance on the standardized matrix $S = std611*7$, the matrix of distance D comes out. After determining the degree of similarity, the method for showing the merging items, which consist of the up most similarity, can be applied. There are different introduced methods for merging which include: full, average, weight, central, and medium and Ward amalgamation (Equation 3).

3- Linkage Computation by Ward

$$\text{Method: } d(r,s) = \frac{n_r n_s d_{rs}^2}{(n_r + n_s)}$$

Here d_{rs}^2 is the distance between the group r and s which is determined by Ward method. Because in this case the value of dispersion within groups reaches to lowest point and the homogeneity of obtained groups reaches to highest point. In Ward method, a member places in a cluster in which the new within-cluster matrix has the lowest value.

Accordingly, the cluster diagram for precipitation data is drawn based on the mentioned methods (Figure 7) and that being the case, five precipitation types, as the main types for Zahedan, are obtained (Figure 8). The horizontal diagram shows the numbers of types and the vertical one shows the variety of heights for types compared with each other. In cluster analysis, the meeting points determine the numbers of synoptic types or map patterns (the level of cluster numbers). Like the level of determining the number of

Table 1 continued

.....
2005	4	30	30	5	5	7	1	0.1	0.1	0.1	2
2005	5	1	3	6	4.3	10	3	4.3	9	3	3
2005	5	24	24	6	4.3	10	1	0.2	0.2	0.2	2
2005	5	27	27	6	4.3	10	1	0.7	0.7	0.7	2
2005	5	31	31	6	4.3	10	1	0.2	0.2	0.2	2
2005	6	1	1	2	0.6	1.1	1	0.6	0.6	0.6	2
2005	6	11	11	2	0.6	1.1	1	0.5	0.5	0.5	2
2005	8	31	31	1	1	0.1	1	0.1	0.1	0.1	2

Table 2: Statically characteristics of Zahedan Precipitation(1966-2005)

	Average	Standard Deviation	Cv(%)	Variance
Monthly Precipitation	6.5	6.6	100	44
Yearly Precipitation	79.5	40.5	51	1645

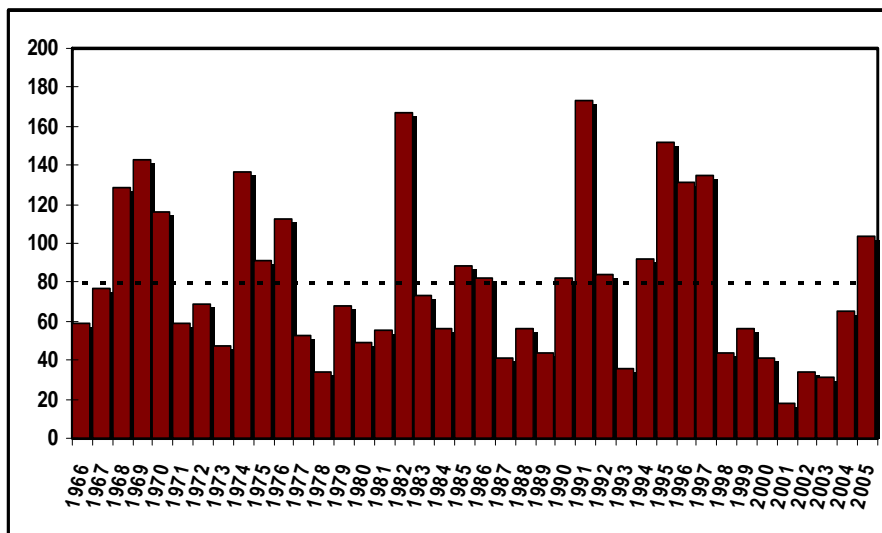


Figure 1: Annual variations of Zahedan Precipitation (1966-2005)

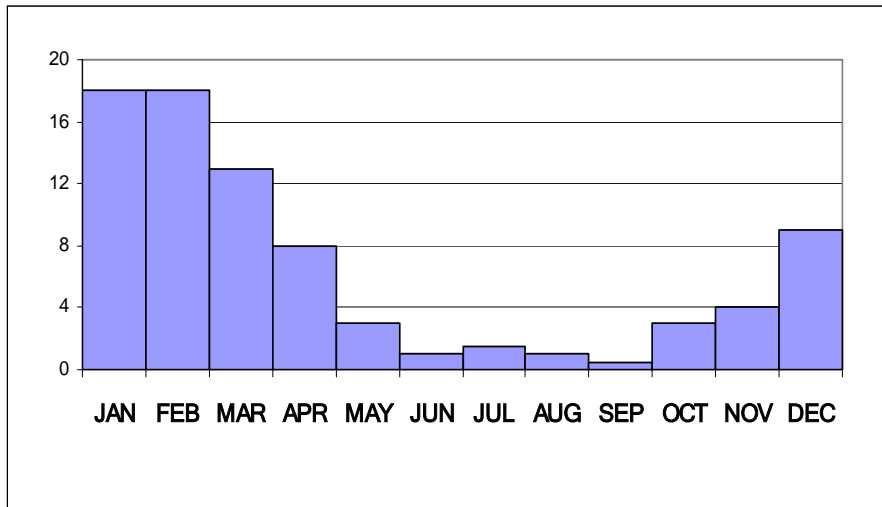


Figure 2: Monthly variations of Zahedan Precipitation (1966-2005)

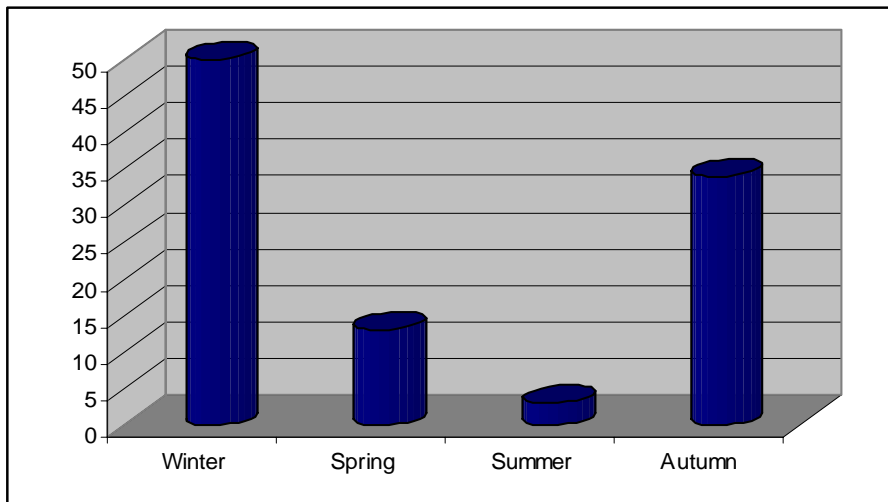


Figure 3: Seasonal fluctuations of Zahedan Precipitation (1966-2005)

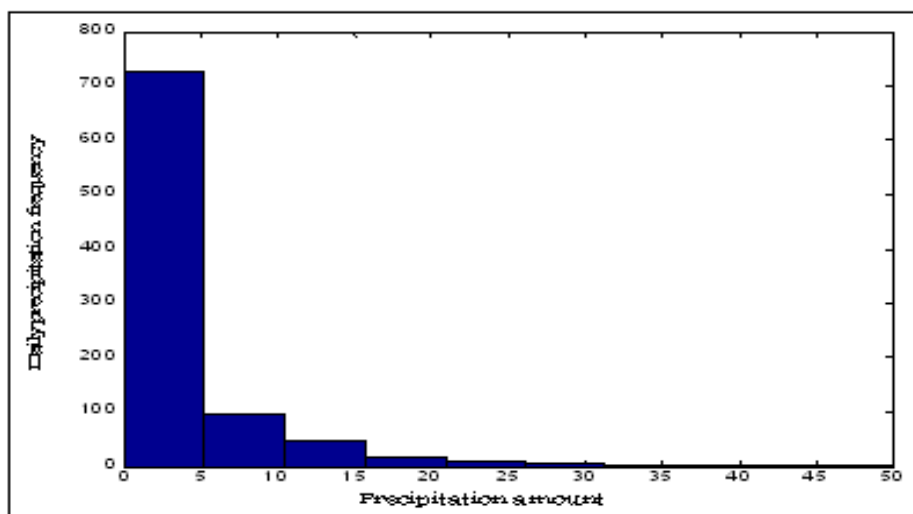


Figure 4: Frequency diagram of precipitation days than the amount of precipitation (1966-2005)

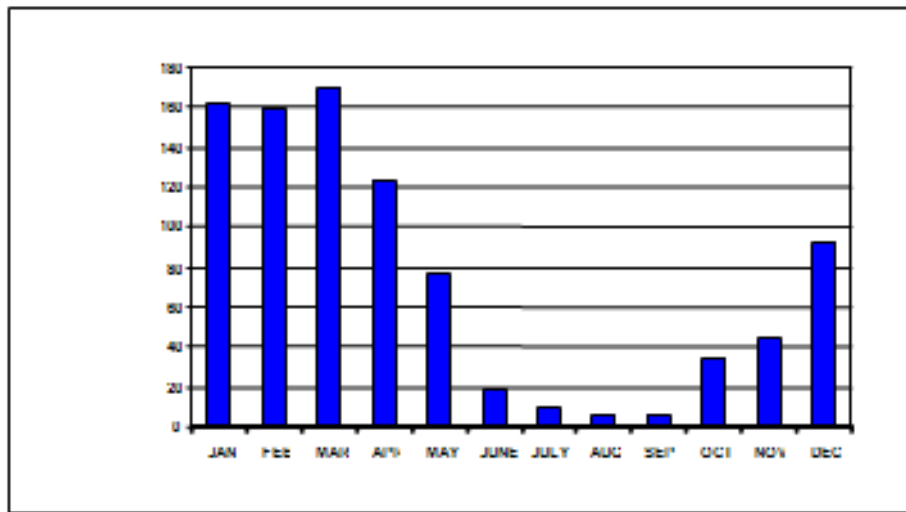


Figure 5: Monthly distribution of precipitation days in Zahedan (1966-2005)

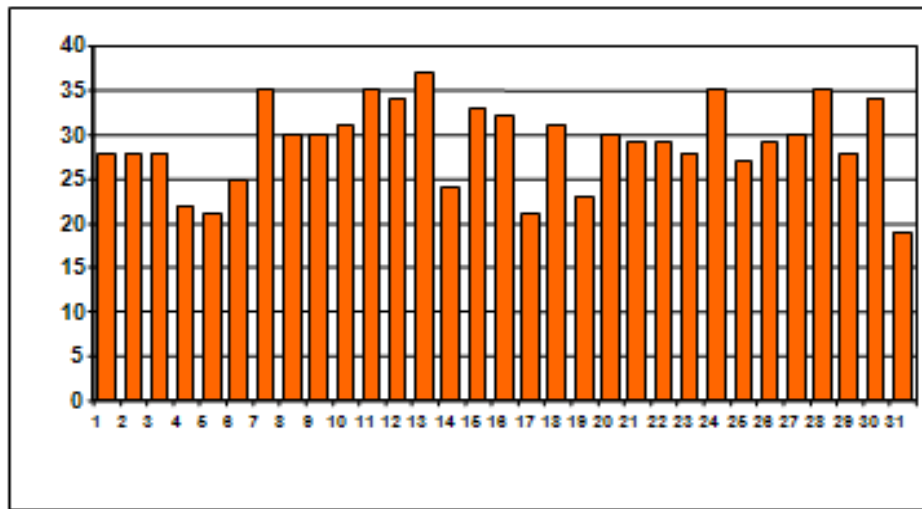


Figure 6: Daily distribution of precipitation days in Zahedan (1966-2005)

Discussion and conclusion

Performing a cluster analysis on the standard (std 616×7) and the merging of days based on Ward method showed that Zahedan has almost five distinguished precipitation types (Figure 7 & 8).

The precipitation types of Zahedan 1-Medium, severe and frequent occurrence type

Considering the peak monthly, daily precipitation type and its intensity, the number one precipitation type places at the

second stage after the type number 5 and considering the sum of monthly and daily precipitation after types number 4 and 5, it locates at the third stage (Table 3) and considering the daily precipitation durability after types 3 and 4, it stands at the third stage (Table 4). Therefore, it can be claimed that this precipitation type has medium durability in precipitation and is lower than 5 days, mostly 1 to 2 days and it is considered as severe precipitation in classification of precipitation intensity. This type is known as medium

precipitation type in amount. From the occurrence point of view, this type locates at the third stage after types 2 and 3. Like types 2 and 3, this type has high occurrence and it is more likely to occur (Table 5). Considering the monthly frequency, it has occurred in all the months of a year except June and September. But it's the most frequent occurrence is in the winter. Again from the high occurrence point of view, winter, fall, spring and summer are placed in this order (Table 6).

2- Very low, very short term, more frequent occurrence and very quiet type

The number two precipitation type is located in the last stage with respect to peak monthly and daily precipitation, the sum of monthly and daily precipitation as well as precipitation intensity (Table 3) and also in the matter of daily precipitation durability; it locates in the last stage (Table 4). Consequently, it is fair to say this precipitation type has very low durability as if it only occurs in one-day durability, besides it possesses the most one-day durability frequency. In classification of precipitation intensity, it is known as very quiet precipitation. With respect to occurrence, this type is the highest, i.e. it is more likely to occur (Table 5). Regarding the monthly frequency, it is visible in all the months of a year and considering the frequency of seasonal occurrence, winter, spring, fall and summer stands in this order (Table 6).

3- Low, high durability and low intensity type

Regarding the peak and sum of monthly and daily precipitation as well as precipitation intensity, the precipitation type number three stands before type number 2 in the fourth stage (Table 3). From the viewpoint of daily precipitation durability, it places in the second stage after type 4 and includes the 2-3-4 and 5-day durability. However, the frequency of 2 and 3-day durability is more and it does not have one-day durability as well (Table 4). So it can be concluded that this precipitation type has high durability in precipitation characteristics and is known as quiet in classification of precipitation intensity. This type is popular as low precipitation type in amount. Regarding occurrence, after type 2 this type has dedicated the highest occurrence to itself, i.e. it is more likely to occur (Table 5). With reference to monthly frequency, it occurs in all the months except August. Seasons of winter, spring, fall and summer are the orders of seasonal occurrence frequency (Table 6).

4- High, low occurrence and very long durability type with winter precipitation regime

With regard to peak monthly and daily precipitation type and intensity, the No.4 precipitation type stands after types number 1 and 5 in the third stage, and respecting the sum of monthly and daily precipitation, it is in the second stage after type 5 (Table 3) and regarding the daily precipitation durability it places in the first stage (Table 4). So we can claim that this type has a very long durability and just

possesses two frequencies of 6 and 7-day durability, it is the only type which includes durability more than 5 days and its frequency of durability occurrence is very low. In the class of precipitation intensity, it is medium and it is high in amount. From the occurrence viewpoint during the study, it had been the lowest and occurred just 4 times. It is not more likely to occur and does not follow any special regularity (Table 5). With respect to monthly frequency, it was seen in February and March and it does not have any winter seasonal occurrence as well (Table 6). It can be concluded that precipitation with very long durability is more likely to occur just in winter.

5- Very high, low durability and very severe type

Considering the peak and sum of monthly and daily precipitation, as well as precipitation intensity, type number 5 stands in the first stage (Table 1) and in regard to daily precipitation durability, it is the fourth (Table 4). So this precipitation type has low durability and only has two frequencies of 3 and 4-day durability. In the classes of precipitation intensity and amount, it is severe and very high. With reference to occurrence during the study, it has the lowest occurrence after type number 4, as if it occurred only 17 times. It is not more likely to occur and does not follow any special regularity (Table 5). It only occurs in winter and fall, i.e. it has the winter-fall seasonal occurrence (Table 6). So it can be concluded that very high and severe precipitation is more likely to occur only in fall and winter.

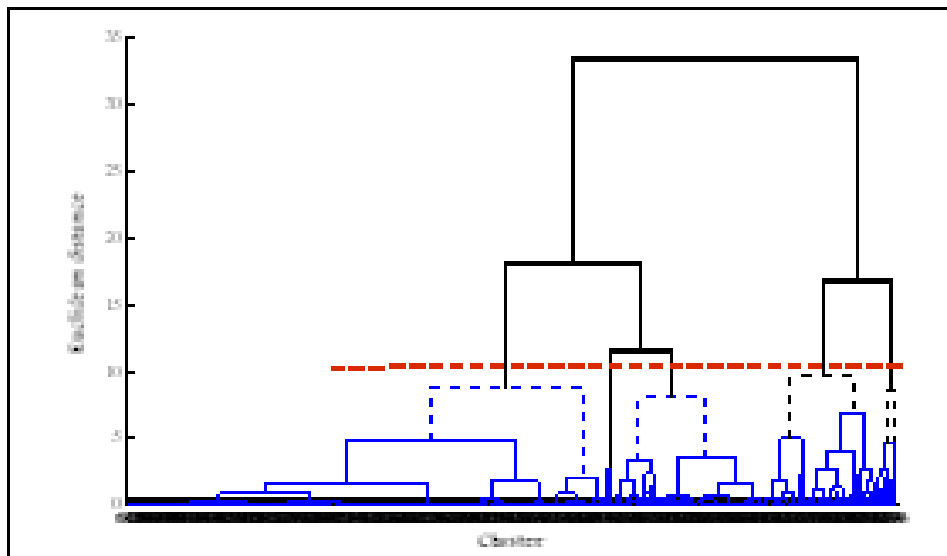


Figure 7: Total dendrogram of Zahedan precipitation Type zahedan (1966-2005)

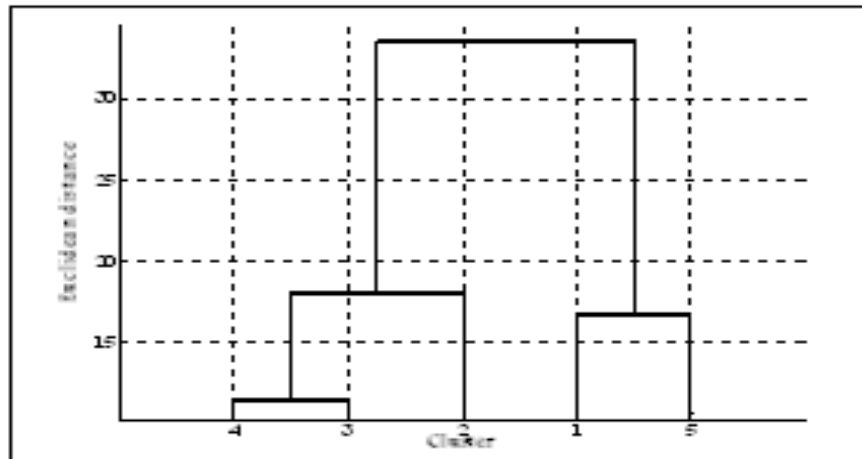


Figure 8: Sub dendrogram of Zahedan precipitation Type zahedan (1966-2005)

Table 3: Precipitation Types characteristics of Zahedan (1966-2005)

	Type number	Monthly Precipitation maximum	Monthly Precipitation Sum	Daily Precipitation maximum	Daily Precipitation Sum	Intensity Precipitation
Mean	1	14.9	31.4	11.9	14.27	8.6
Max	1	52	85.3	24.1	28.4	23
Mean	2	6.49	12.23	1.5	1.5	1.5
Max	2	52	85.3	7	7	7
Mean	3	7	15	3.1	4.5	1.8
Max	3	30	72.8	12	16	5.5
Mean	4	10.45	39.3	10.4	25.3	4
Max	4	14	54.9	14	37.1	6.18
Mean	5	27.5	46.8	27.5	37.5	16.1
Max	5	52	85.3	52	67.8	42

Table 4: Durability properties of precipitation types of Zahedan (1966-2005)

day	7 day	6 day	5 day	4 day	3 day	2 day	1 day	Type number
Frequency	0	0	1	3	11	41	33	1
	0	0	0	0	0	0	۳۸۴	2
	0	0	0	۵	32	86	0	3
	2	2	0	0	0	0	0	4
	0	0	0	3	3	10	1	5

Table 5: Yearly Frequency of precipitation days type zahedan (1966-2005)

Year	Type1	Type2	Type2	Type2	Type2
1966	2	9	1	1	0
1967	1	13	3	0	1
1968	4	12	4	0	1
1969	2	11	3	0	2
1970	3	3	2	1	1
1971	2	9	3	0	0
1972	1	8	7	0	0
1973	0	5	1	0	1
1974	3	11	2	0	2
1975	1	9	3	0	1
1976	6	15	3	1	0
1977	2	11	3	0	0
1978	0	9	2	0	0
1979	2	10	5	0	0
1980	1	9	2	0	0
1981	2	5	3	0	0
1982	6	17	3	0	0
1983	2	13	7	0	0
1984	2	6	2	0	0
1985	2	1	2	0	1
1986	5	11	3	0	0
1987	3	8	2	0	0
1988	1	9	2	0	0
1989	0	9	4	0	1
1990	4	17	2	0	0
1991	5	9	1	0	1
1992	1	11	8	0	1

Table 5 continued

1993	1	9	1	0	0
1994	2	14	3	0	1
1995	3	14	6	0	1
1996	5	15	1	0	0
1997	4	20	5	0	0
1998	2	9	2	0	0
1999	2	5	6	0	0
2000	2	4	1	0	0
2001	0	6	1	0	0
2002	2	6	2	0	0
2003	0	8	3	0	0
2004	0	3	2	0	0
2005	2	11	6	1	1

Table 6: Monthly Frequency of precipitation days type zahedan (1966-2005)

Month	Type 1	Type2	Type 3	Type 4	Type 5
JAN	20	57	22	0	6
FEB	22	53	18	2	6
MAR	14	78	22	2	1
APR	8	54	23	0	0
MAY	4	43	11	0	0
JUN	0	15	2	0	0
JULY	3	2	1	0	0
AUG	2	3	0	0	0
SEP	0	3	1	0	0
OCT	3	14	4	0	1
NOV	4	27	2	0	1
DEC	9	34	16	0	2

Conclusion

The results obtained from the cluster analysis of precipitation characteristics in Zahedan in almost half of the last century show that:

- Zahedan has almost five distinguished precipitation types
- The highest and lowest precipitation is for types five and two respectively
- The most and the least severe precipitation goes back to types five and two respectively
- The longest term durability is for type four and the shortest term durability is type two's
- The lowest number of precipitation seasons for all the types is 1 season (type 4) and the most number is 4 seasons
- Very high and severe precipitation occurs in fall and winter
- Precipitation with long term durability occurs only in winter
- The highest frequency of durability is related to one-day durability and lowest one is for 6 and 7-day durability
- Type number 4 has had the fewest annual occurrences and is not likely to occur anymore. But types 2 and 3 have had the highest annual occurrences and are more likely to occur more.

Reference

-Alijani B (2002) Determination of the Rain Producing Weather types in Tehran Using Vorticity Computation Method, Iranian Journal of Research in Geography, 63-64:114-132.

-Alijani B(1993) Thermal Area of Azarbayjan by cluster analysis technical, the Humanities science Journal of Teachers Training University, 2-3:87-104.

-Atayi H (2004) Regionalization of Precipitation in Iran PhD Thesis of Climatology, Esfahan , Department Geography, Esfahan University.

-Bacher, J., 1996: Cluster analyse. Munchen: Oldenbourg, 424 pp. Book.

-Baldwin M E and Lakshmiarahan S(2002) Rainfall classification using histogram analysis: An example of data mining in meteorology. Technical Report, School of Computer Science, University of Oklahoma, Norman, OK. 342.

-Domroes M and Kaviani M and Schaefer D(1998): An analysis of regional and intra-annual precipitation variability over Iran using multivariate statistical methods, Theory. Appl. Climatol., 61(3-4), 151-159.

-Gerstengarbe F.W., P.C. Werner, and K. Fraedrich(1999) Applying Non-Hierarchical Cluster Analysis Algorithms to Climate Classification: Some Problems and their Solution, Theor. Appl. Climatol. 64, 143-150.

-Ghrami motlagh A(2004) Climatic Regionalization of bushehr province, M.A. Thesis of Climatology, Esfahan , Department Geography, Esfahan University.

-Heidari H and Alijani B (1999) Climatic Classification of Iran by multi variable method, Iranian Journal of Research in Geography, 37:57-74.

- Jackson, I.J., Weinand, H, (1995): Classification of tropical rainfall stations: a comparison of clustering techniques. *Int. J. Climatol.* 15, 985–994.
- Kaufman, L., Rousseeuw, P.J (1990): *Finding Groups in Data: An Introduction to Cluster Analysis.* Wiley, New York, 344 pp.
- Littmann, T. (2000): An empirical classification of weather types in the Mediterranean Basin and their interrelation with rainfall, *Theor. Appl. Climatol.* 66, pp 161-171.
- Masoodian S.A and Ataei H, (2005) Regionalization of Precipitation seasonal of Iran Using Cluster Analysis, the Humanities science Journal of Isfahan University, 1:1-13.
- Masoodian S.A, (2003), Geographic distribution of precipitation in Iran bay rotated PCA, *Journal of Geography and Development*, 1:79-88.
- Masoodian S.A, (2005) Regionalization of Precipitation Regimes of Iran Using Cluster Analysis, *Journal of Research in Geography*, 52:47-61.
- Masoodian S. A, (2005) The Relationship between Geopotential Height 500 hpa Circulation Patterns and Precipitation over Mond Basin, *Journal of Geography and Development*, 5:161-183
- Mohammadi B and Masoodian S.A, (2007) The Relationship between synoptic climatic types of Sanandaj Station with circulation patterns of 500 hpa, *Iranian Journal of Geography and Development*, 9:39-56.
- Nazaripour H and Khoshhal J (2007) Identification Synoptic Type of Khor Biabanak Station (1985-2004), *The Scientific and Technical Journal of IRIMO (Nivar)*, 66-67:27-38.
- Nazaripour H and Khoshhal J, (2007) Identification Synoptic Type of Naein Station (1992-2004), *The Scientific and Technical Journal (Atlas)*, Vol:4. No: 7-114-126.
- Nazaripour H and Khosravi M (2009) Identification role of Synoptic patterns Super Heavy Precipitations In decrease flood destruction In Bushehr Province, *Articles collections of National Conference for effects reduction on weather and climatic hazards, Ardabil. Iran.* pp 112-130.
- Nazaripour H, (2007) Synoptic Analysis heavy precipitation of Boshehr province, M.A. Thesis of Climatology, Esfahan, Department Geography, Isfahan University.
- Romero, R., Sumner, G., Ramis, C., Genoves, A (1999): A classification of the atmospheric circulation patterns producing significant daily rainfall in the Spanish Mediterranean area. *Int. J. Climatol.* 19, 765–785.
- Singh, C.V. (1999): Principal components of monsoon rainfall in normal, flood and drought years over India. *Int. J. Climatol.* 19, 639–652.
- StatSoft, (1994): *Statistica. Vol. III, Statistics II.* Tulsa: Stat-Soft Technical Support.
- Torabi S and Jahanbakh S and Alijani B and Shafiei Kh (2001) *Climatic*

Classification of Iran by multi variable method, Iranian Journal of Research in Geography,39:38-56.

-Yarnal B (1993) Synoptic Climatology in Environmental Analysis, London, a

Primer Belhaven Press. Translated by Masoodian, S.A, (2006), Esfahan University Press.97-140.