



## Determining Suitable Probability Distribution Models for Annual Precipitation Data (A Case Study of Mazandaran and Golestan Provinces)

Mahdavi, Mohammad

Faculty of Natural Resources, University of Tehran, Karaj, Iran

Khaled Osati (Corresponding author)

Watershed Management, University of Tehran and member of Young Researchers Club of Kurdistan, Karaj, Iran

Tel: 98-935-862-6454 E-mail: khaled.ausati@gmail.com

Sayed Ali Naghi Sadeghi

Combating desertification Engineering, University of Tehran, Karaj, Iran

Bakhtiar Karimi

Drainage and Irrigation Engineering

University of Tehran and member of Young Researchers Club of Kermanshah, Karaj, Iran

Jalil Mobaraki

Watershed Management, University of Tehran, Karaj, Iran

### Abstract

Statistical distributions can be used for data development in shortage data situations, as in many parts of Iran station. The aims of this study are to select the best frequency distribution to estimate average annual precipitation and assess the effects of data length on the selection of suitable distribution. Therefore 65 stations data of Mazandaran and Golestan provinces were analyzed. Relative residual mean square (RMS) was used to determine the best fitted distribution to any annual series and precipitation was estimated for different return periods. Relative frequency of first classes of fitted distributions showed that normal and Pearson distributions fitness decreased and Gumbel distribution had more fitness with data series by increasing statistical period length. The best-fitted distribution is Pearson with 15-year data; log Pearson for 20, 25 and 30-year periods. Based on Moment method and total given scores, two-parameter normal distribution has the best fitness in all statistical periods.

**Keywords:** Annual precipitation, Mazandaran, Golestan, Frequency distribution, Relative Residual Mean Square

### 1. Introduction

Precipitation is a key component of the hydrological cycle and one of the most important parameters for various natural and socio-economic systems: Water resources management, agriculture and forestry, tourism, flood protection, to name just a few (Schmidli, 2005). The study of consequences of global climate change on these systems requires scenarios of future precipitation change as input to hydrologic cycle. Hydrological and meteorological data show no random behavior. Then they can be analyzing by some statistical methods based on frequency analyses of precipitation and flood data. Therefore, statistical distributions can be employed for the studies such as the design of water structure, the management of water resource and watershed, and the determination of effective factors about hydrologic cycle. However, it is necessary to determine the best-fitted distribution to studied data. The primary aim of frequency analysis is to relate the magnitude of extreme events to their frequency of occurrence using the probability of distributions (Chow et al., 1988). The aim of this study is to determine suitable probability of distribution models for annual precipitation data in study area. Seven well-known probable distributing models including two-parameter standard normal, two-parameter log normal, three-parameter log normal, two-parameter gamma, Pearson type III, log Pearson type III and Gumbel with moment and maximum likelihood parameters which are tested to determine the best fitted distributions as well as precipitation in different return periods.

In Iran and other countries, many hydrologist and other experts have analyzed precipitation data. Mashayekhi (1972) considered Iran precipitation as normal distribution. Khalili (1973, 1976) studied 90-year Tehran precipitation data and 10-year central Alburz precipitation data and indicated credibility of gamma distribution. After investigating probability density function of monthly and yearly precipitation data of oldest station of Iran, Haghghatjou (2002) found that log Pearson type III is the best-fitted distribution to monthly data and there are no suitable distributions for yearly precipitation. After investigating of probable suitable distributions for meaning, maximizing and minimizing discharges in Mazandaran hydrological stations, Gholami (2000) concluded that more stations Gumbel distributions (L-moment method) and three-parameter log normal (moment method) had the most fitness for maximizing and minimizing annual discharges, as well as Gumbel (L-moment) for meaning and minimizing annual discharges. Bedeustani (1999), performed a study in the east Azarbayjan, indicated that although there was no suitable distribution for discharging estimation in short term, three-parameter log normal distribution and then log Pearson type III distribution showed more fitness with data series by increasing statistical period length. In addition, for maximizing precipitation in short and long terms, three-parameter log normal and three-parameter log normal along with log Pearson type III distributions were suitable, respectively. Markovich (1965) used minimum Square method for flowing assessment and concluded that gamma distribution had the best fitness among other distributions. Keshtkar (2001, 2006) compared moment and L-moment Methods to determine the probability of distributing parameters and suitable distributin for annual discharge series. 20 and 17 hydrometric station was chosen for meaning, maximizing and minimizing annual discharges and maximizing peak discharges respectively. In the central plateau watershed, the best fitted distribution for different annual discharges was studied. Results showed the best fitted distributions: for minimum discharges: Pearson distribution type III ( L moment method); for Medium annual discharges: Pearson type III and Log Pearson type III distributions (L moment method); for Maximum annual discharges: Pearson type III (L moment method), Log Pearson type III and two-parameter Log normal (moment method) distributions; for Maximum annual momentous discharges: Log Pearson type III (moment), Pearson type III, three-parameter Log normal and two-parameter Log normal (L moment method) distributions. Dinpashoh (2004) studied the regionalization of Iran's precipitation climate by using Data from 77 synoptic stations in Iran. After selecting twelve variables from the 57 candidate variables, using Procrustes Analysis, the selected variables used to regionalize Iran's precipitation climate by factor analysis and clustering techniques. The H and Z-statistics, which were based on the L-moment technique, were used to test the homogeneity of each region and selected the distribution which best-fitted annual precipitation records- in that region. The countrywide area was divided into six regions with homogeneous and one with heterogeneous precipitation climates. They demonstrated Southern coastal of Caspian Sea as B homogeneous region i.e. the best-fitted distribution for annual precipitation records in that region of Generalized Logistic.

Tase (1982) assessed 50-year precipitation data of 82 meteorological stations in Japan and considered normal distribution for these data. Reich (1983) indicated monthly precipitation as a function of Gumbel distribution based on 15-year data. Based on results of a research conducting in the Seyhan river basin in Turkey (Topaloglu, 2002), Gumbel, log-Logistic, Pearson type III, log-Pearson type III and log-normal-3 distributions were applied to the series of annual instantaneous flood peaks and annual peak daily precipitation for 13 flow gauging and 55 precipitation gauging stations. According to the evaluations of chi-squared tests, Gumbel (the moments methods) for both hydrological and precipitation stations were founded to be the best models. Based on the K-S test, log-normal-3 (the moments methods) and log Pearson type III (the moments methods) models were determined to be the best for hydrological and precipitation stations, respectively. Dahamsheh et al (2007) described the structural characteristics and temporal and spatial variation of annual precipitation data in Jordan, together with possible projections for the future and using the probability distribution. Based upon their results, annual precipitation in Jordan was consistent in time with evidence of randomness. Finally they identified three distinct regions as 3-parameter lognormal distribution in the east, gamma distribution in the south and log-Pearson Type III distribution in the remainder of the country. Based on the chi-square goodness of fit test, Campbell (1981) found that the log Pearson type III distribution was the best distribution for culvert design in small-forested watersheds. Whereas Gumbel distribution with L-Moment method has been found to be a suitable distribution for peak flows by Gingras and Adamowski(1992), Pilon, and Adamowski (1992), Wallis(1988).

## **2. Method and Materials**

The basin of Caspian Sea Rivers is located in the north slope of central Mountains and stretched along shore from Sefidrood delta to Bandar gaz. It is limited to Sefidrood watershed from the west, Central Basin from the south, Semnan and Khorasan basin from southern east and the Caspian Sea from the North. General patterns of atmospheric circulations over the Caspian Sea are controlled by the Eurasian High and Low Pressure Cells (Iceland Low, Indian Low, Azores High, Siberian High and Central Asian High; Lahijani, 2007). The Iranian coast has a sub-tropical climate and a rainfall pattern shows a strong gradient from west (around 1900mm in Anzali) to east (around 196mm in Hassan Gholi). Distribution of rainy weather is uniform throughout the year on the coast. The of average relative humidity is mostly above 75%. Mean summer temperature is around 26 1C, with fewer than 30 snowy and freezing days during the year (Lahijani, 2007).

In this study, precipitation-gauging data from stations of Golestan and Mazandaran provinces has used to investigate fitness of statistical distributions (figure 1).

Totally, a 30-year (1972-2001) common period data of 65 stations have been selected based on statistical period length and closeness to present time. After analyzing the outliers by U.S. Water Resources Council method (1981), the missing data has been completed by normal ratio method (Mahdavi, 2002) with attention to available number of years. Furthermore the homogeneity of data was tested by Double Mass Curve and Runs tests (Mahdavi, 2002).

Using HYFA (Hydrological Frequency Analysis) program, the best-fitted statistical distribution is determined for each station in periods of 15, 20, 25 and 30-year. In HYFA application, goodness of fit of relative residual mean square and chi-square test has been used and the parameters of the distributions were estimated by the methods of moments and maximum likelihood methods.

In this study, normal, two-parameter log-normal, three-parameter log-normal, Gumbel, two parameter gamma, Pearson type III and log Pearson type III distributions have been used (more information about distribution calculations are available in Applied Hydrology book, Mahdavi,2002 and Betül S thesis). After selecting the best-fitted distribution, it is possible to estimate mean annual precipitation for return periods of 2, 5, 10, 20, 25, 50, 100, 500 and 1000 year. By analyzing relative residual mean square and chi-square test tables (in HYFA output) in different time series (15, 20, 25 and 30-year), the best-fitted distributions ranked. Then scores 7-1 has been given to any distribution respectively, based on priority of them. Finally, the best-fitted probability of distribution selected by relative frequency of first classes (by analyzing the best-fitted distribution in any station) and total given scores for each statistical distribution.

### 3. Results

In HYFA software, using deviation table and volumes of relative residual mean square, the best statistic distribution, that has the least deviation, has been selected and introduced as the best-fitted distribution to data. Then in this distribution, by attending to determine average and occurrence probability (or return period) relating curve can be drawn. By this curve, volumes of considered variety with certain return period can be determined. HYFA output presents some of the statistical characteristics such as mean, standard deviation, variance, skewness, etc. Some statistical characteristics of 30-year data were tabulated in Table 1 and location of the studied stations depicted in fig. 1.

Obtained scores of distributions indicated that in all stations, there is no apparent predominance for any distribution (confirms with results of Haghighatjou study). In all stations, relative frequency of first Score of the best fitness in four assessed period changes as figures 2 & 3.

Based on the sum of given Scores to each distribution, in Moment method, two-parameter Log-normal distribution for four statistical periods was the most suitable distribution. As well as in Maximum likelihood method, by increasing the period length, Pearson, gamma, three-parameter log-normal and log Pearson type III have the best fitness with scores of 216, 210, 225 and 219 respectively (table 2 & 3).

Relative frequency of first classes and separated investigation of two methods used to estimating distributions parameters introduce Gumbel and normal (moment method), gamma and Pearson distributions (Maximum likelihood method) for 15-year period, log Pearson, Pearson, gamma and Gumbel distributions (Maximum likelihood method) for 20-year period, log Pearson, Pearson and gamma distributions (Maximum likelihood method) for 25-year period and log Pearson and gamma distributions (Maximum likelihood method) for 30-year period as the best fitted models.

According to the relative frequency of the best fitted distribution and comparative investigation of two methods used to estimating the parameters of distributions, the best fitted distributions are Pearson and normal ones (moment method) for periods of 15-year, log Pearson, gamma and normal distributions (moment method) for 20-year, log Pearson, Pearson and normal distributions (moment method) for 25-year and log Pearson, gamma and normal distributions (moment method) for 30-year.

### 4. Discussion and Conclusion

Based on obtained Scores of distributions in all stations, there is no apparent predominance for any distribution (confirms with results of Haghighatjou study). Therefore, it is more important to select the frequency distribution to estimate the value of annual precipitation data based on the best-fitted distribution for any region.

The changes in statistical period length in studied stations changes the best fitted distribution for 88 percent of stations and only 8 stations have no changes in the best fitted distribution (Table 4) that indicated relatively high variability of annual precipitation and insufficient data. It indicated the importance of appropriate distributing of station and the need to use the more data for researches.

Based on total given scores, two-parameter normal distribution has the best fitness in all statistical periods. Totally, it is noticed that by increasing statistical period length, normal and Pearson distributions are decreased and Gumbel distribution has more fitness with data series, based on relative frequency of the best-fitted distributions. Superiority of Pearson distribution in 15-year periods changed as log Pearson for 20, 25 and 30-year periods that maybe a result of

data short period or occurrence of extreme events (i.e. drought and wet year). Period length also effects skewness, standard deviation and distribution parameters. Therefore completing the gages and regular data collection has a significant role in hydrological data analysis. Having a long-term data also can be effective in appropriating interpretation of extreme hydrological events (i.e. drought and wet years) occurrence.

## 5. Acknowledgements

We are very grateful of experts in Ministry of Energy (especially Mr. Babae) for helpful attempts in completing the data. We thanks to the attempts of employee of university of Tehran, Natural Resource College.

## References

- Bedeustani, H. (1999). Assessment of the best frequency distribution to prediction the flood and maximum precipitation with specific occurrence probability in East Azerbaijan. MSc., University of Tarbiat Modares(noor).
- Betül S. (2003). Role of rainfall pattern and basin physiography on the probability distribution on floods. PhD thesis, Philosophy in Civil Engineering Hydraulics & Hydrology & Water Resources Program, Dokuz Eylül University.
- Campbell, A. (1981). Flood frequency analysis of small forested watersheds for culvert design. M.Sc. thesis.
- Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied Hydrology*. McGraw-Hill, New York, NY.
- Dahamsheh, A. & Aksoy, H. (2007). Structural characteristics of annual precipitation data in Jordan. *Theoretical and Applied Climatology*, 88(3-4), 201-212, DOI 10.1007/s00704-006-0247-3, Printed in The Netherlands.
- Dinpashoh, Y. (2004). Selection of variables 5 for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *J. Hydrology.*, 297, 109–123,
- Gholami, A. (2000). Assessment of the best statistical distribution to minimum, mean and max discharge data using the L-Moment method (Mazandaran). MSc., University of Tarbiat Modares(noor).
- Gingras, D. and Adamowski, K. (1992). Coupling of nonparametric frequency and L-moment analysis for mixed distribution identification. *Water Resour. Bull.* 28 2, pp. 263–272.
- Haghighatjou, P. (2002). Probability distribution functions as applied to monthly and annual precipitation of old station in Iran. *J.Agro.Sci. Nature. Resour*, Vol.9 (3).
- Keshtkar, A.R. (2001). Assessment and fitting the best statistical distribution to, mean max and min discharge data (central plateau). MSc. Thesis, University of Tehran.
- Keshtkar A. R. (2006). A study of the appropriate probability distributions for annual flow series, using L-moment method in arid and semi-arid regions. *Iranian J. Natural Res.*, Vol. 59, No. 1, 2006.
- Khalili, A. (1973). Precipitation pattern of central Alburz. Arch. Meteo. Geophys, *Bioclimatology Series B* 21(2-3): 215-232.
- Khalili, A. (1976). Precipitation climatology of the Tehran area. *Meteo. magazine*, Vol. 105:293-306.
- Lahijani, H.A.K. (2007). Evidence for late Holocene highstands in Central Guilan–East Mazandaran, *South Caspian coast, Iran*. Quaternary International (2007), doi:10.1016/j. quaint.2007.10.005
- Mahdavi, M. (2002). *Applied hydrology*, vol. 1-2, University of Tehran publication.
- Markovich, R.D. (1965). Probability of best fit to distributions of annual precipitation and run off. Hydro.Paper.no.8, Colorado State University, Fort Collins, Co.
- Mashayekhi, T. (1972). Analysis of annual Precipitation data of Nivar Meteorology Stations. *j. of meteorology of total Country*.
- Moosavi, H. (1987). Estimation of maximum daily discharge in Fars and adjacent without data Watershed. M.Sc., University of Shiraz, Irrigation department, pages 315.
- Pilon, P.J. and K. Adamowski. (1992). The value of regional information to flood frequency analysis using the method of L-moments. *Can. J. Civ. Eng.*, 19, 137-147.
- Reich, T. (1983). Frequency distribution of precipitation amounts. Gerlands Beitrage Zur Geophysik, Leipzig.92(5):377-390.
- Schmidli J., C. M. Goodess, C. Frei, M. R. Haylock, Y. Hundecha, J. Ribalaygua, T. Schmith. (2005). Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps. *Journal of Geophysical Research*, DOI:10.1029/.
- Tase, N. (1982). Regional occurrences of wet and dry years in Japan. Instit. Of Geoscience, Tsukuba University, *Annual Report*, no.8:37-40.

Topaloglu, F. (2002). Determining suitable probability distribution models for flow and precipitation series of the Seyhan River Basin. *Turk j. Agric for* 26(2002) 187-194 © T.BÜTAK.

Wallis, J. R. (1988). Catastrophes, computing and containment: living in our restless habitat. *Speculation in Science and Technology*, 11(4). 295-315.

Table 1. The statistical characteristic of given station data and location of the station.

Station	Annual precipitation Average (mm)	St. deviation	Skewness	Coefficient of variation
Inchehbroon	206.60	51.37	0.01	0.25
Ghaleh Jigh	351.75	87.45	0.50	0.25
Ghazanghayeh	264.09	87.10	-0.21	0.33
Maraveh Tappeh	347.29	80.41	1.03	0.23
Torshakly	214.27	68.45	0.51	0.32
Hootan	273.92	71.41	0.36	0.26
Tangrah	679.86	120.81	0.77	0.18
Tamr	503.54	124.26	0.78	0.25
Galikesh	759.27	193.21	1.38	0.25
Gonbad Kavooos ,	413.46	94.88	0.75	0.23
Lazooreh	864.96	184.59	1.21	0.21
Bahlakeh Dashli	427.32	97.08	0.78	0.23
Pas poshteh	960.18	214.79	0.74	0.22
Nodeh	825.63	172.06	-0.12	0.21
Araz Kooseh	449.67	81.73	0.80	0.18
Ghazaghly	359.89	78.46	0.06	0.22
Sad Gorgan	344.43	87.40	0.69	0.25
Tirtash	565.24	118.65	0.63	0.21
Salian Tappeh	332.22	70.67	0.60	0.21
Shilat Tazeh Abad	596.75	348.50	1.69	0.58
Taghi Abad	592.53	106.72	0.81	0.18
Nahar Khoran	750.51	192.57	0.61	0.26
Jafa Kandeh	490.43	136.23	1.48	0.28
Til Abad	253.30	70.79	0.24	0.28
Siah Ab	490.60	98.82	0.52	0.20
baghoo	548.95	115.32	0.33	0.21
Fazel Abad	678.58	95.95	-0.16	0.14
GHafar hajee	443.22	91.79	0.33	0.21
Barkola	475.30	96.02	-0.26	0.20
Abloo	711.07	203.24	1.55	0.29
Afrachal	834.09	384.61	1.17	0.46
Darabkola	751.01	171.47	0.41	0.23

Continued: Table 1.

Station	Annual Precipitation Average (mm)	St. deviation	Skewness	Coefficient of variation
Solaiman Tangeh	584.34	161.99	1.72	0.28
Rig Cheshmeh	750.03	146.97	0.67	0.20
Kord Khail	672.07	155.00	0.60	0.23
Shirgah	964.99	218.47	-0.33	0.23
Kiakola	670.99	184.67	0.07	0.28
Arab Khail	770.1	126.38	-0.09	0.16
Ghoran Talar	1074.26	244.84	0.60	0.23
Babol	656.8	148.34	0.53	0.23
Mian Dasht	798.06	165.56	0.46	0.21
Larim	732.47	149.83	0.55	0.20
Naharestagh	400.29	106.41	0.01	0.27
Panjab	202	65.9	0.65	0.33
Razan , Noor	320.98	110.56	0.5	0.34
Karehsang	946.23	236.14	0.92	0.25
Sorkh Rood	1137.26	307.48	0.63	0.27
Mahmudabad	1003.48	157.97	-0.37	0.16
Chamestan	781.28	164.62	0.43	0.21
Alam Kola	1023.76	185.79	0.79	0.18
Khair Rood Kenar	1395.31	368.22	0.91	0.26
Sardab Rood	1157.92	245.74	-0.33	0.21
Kelar Abad	1512.09	263.71	0.35	0.17
Abbas Abad	1552.77	185.6	0.32	0.12
Haratbar	1369.54	323.26	0.5	0.24
Pole Shir Rood ,	1245.06	318.25	0.82	0.26
Gelibin	1048.66	316.41	-0.53	0.30
Kangsar	1075.95	323.45	0.3	0.30
Rostam Rood	1088.12	226.25	-0.02	0.21
Alamdeh	881.84	268.77	0.6	0.30
Tooskatook	1114.29	214.43	0.05	0.19
Sefid Chah	406.16	77.09	0.75	0.19
Nozar Abad	669.65	151.08	0.26	0.23
Talarom	539.53	88.93	0.29	0.16
Paeen Zarandin	763.96	199.97	0.04	0.26

Table 2. Total given Scores for each statistical distribution fitted to annual precipitation data in different time series 15, 20, 25 and 30-year based on Moment method.

Gumbel	Log Pearson Type III	Pearson Type III	Two Parameter gamma	Three-Parameter Log-normal	Two-Parameter Log-normal	normal	Distribution Type / Statistical Period(year)
228	374	354	189	324	175	176	15
206	377	355	196	316	176	194	20
187	382	355	197	331	152	216	25
198	377	338	182	350	162	213	30

Table 3. Total given scores for each statistical distribution fitted to annual precipitation data in different time series 15, 20, 25 and 30-year based on Maximum likelihood method.

Gumbel	Log Pearson Type III	Pearson Type III	Two Parameter gamma	Three-Parameter Log-normal	Two-Parameter Log-normal	normal	Distribution Type / Statistical Period(year)
353	228	216	229	254	283	257	15
330	236	282	210	241	259	262	20
327	257	256	229	225	243	283	25
336	219	271	229	228	244	293	30

Table 4. The best fitted statistical distributions to stations without change in different time series.

station	The best fitted distribution
Hootan	normal distribution (each one of two methods)
Nodeh	Pearson distribution (moment method)
Araz Kooseh	Three Parameter Log normal distribution (maximum likelihood method)
Fazel Abad	Log Pearson type 3 distribution (moment method)
Barkola	Pearson distribution (moment method)
Shirgah	Log Pearson type 3 distribution (moment method)
Naharestagh	Log Pearson type 3 distribution (moment method)
Sardab Rood	Pearson distribution (moment method)

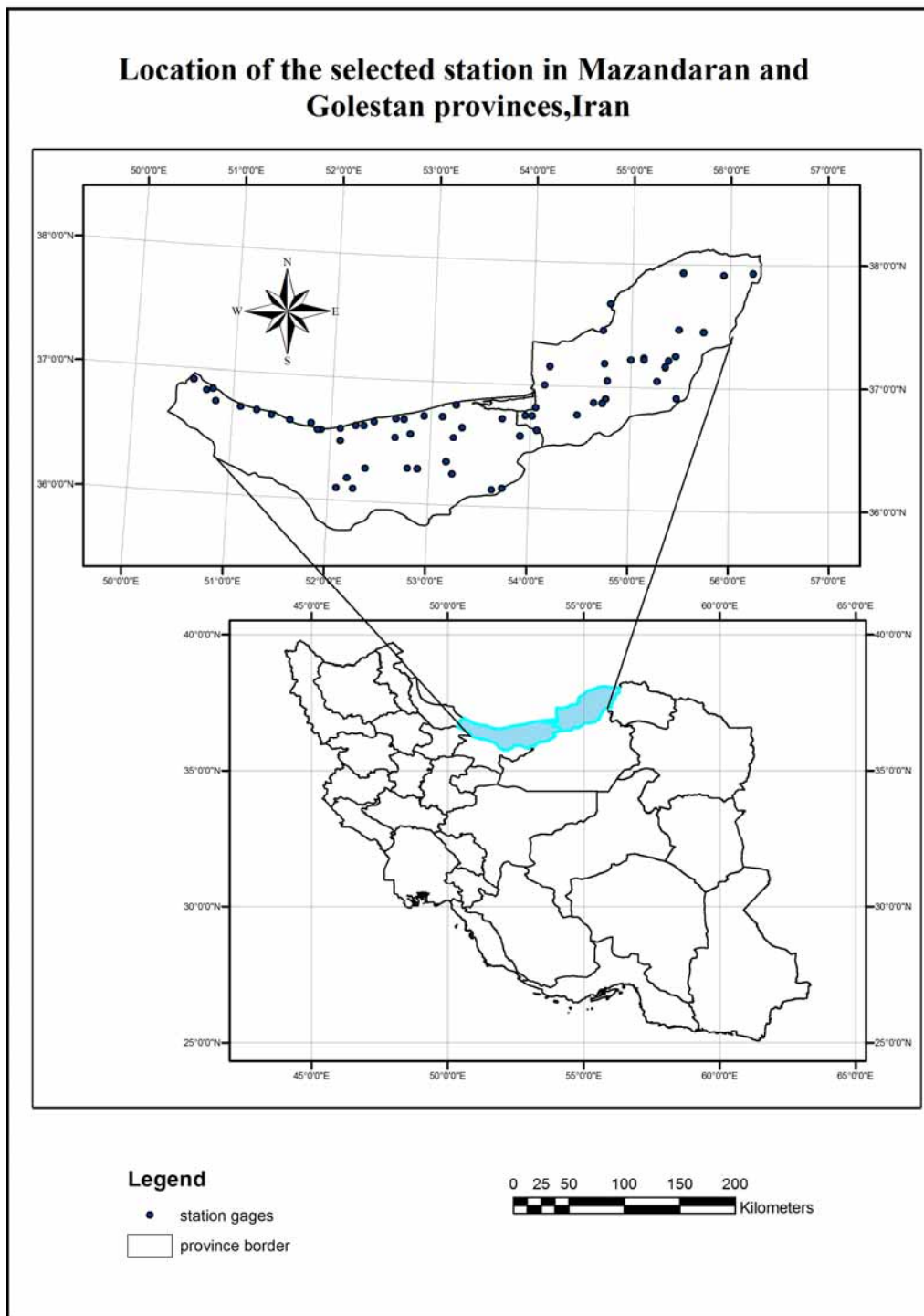
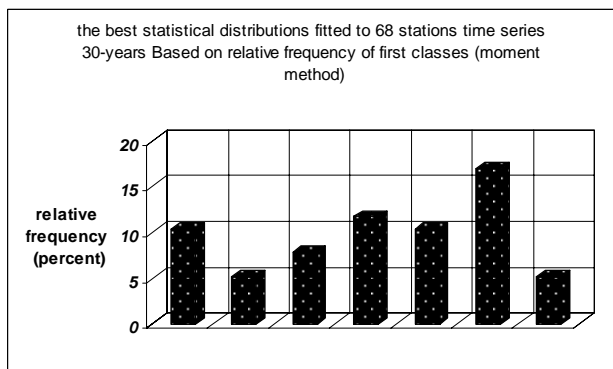
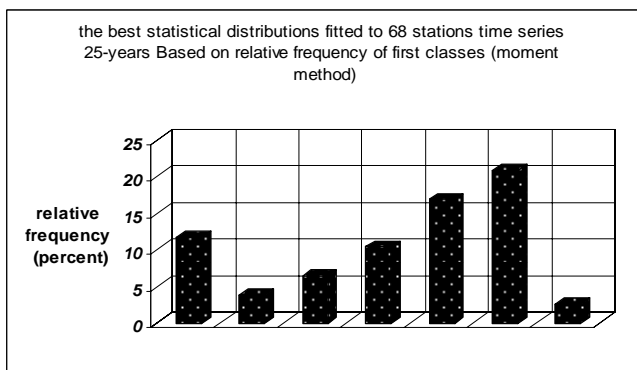
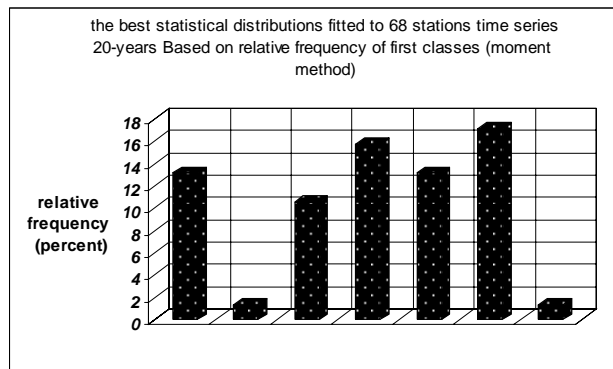
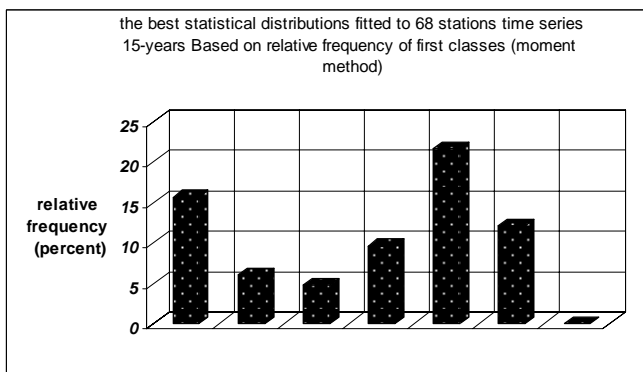


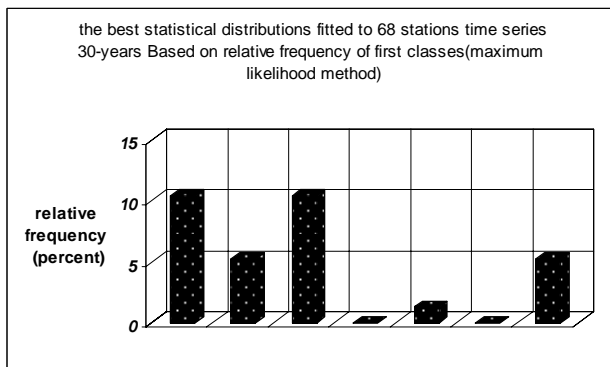
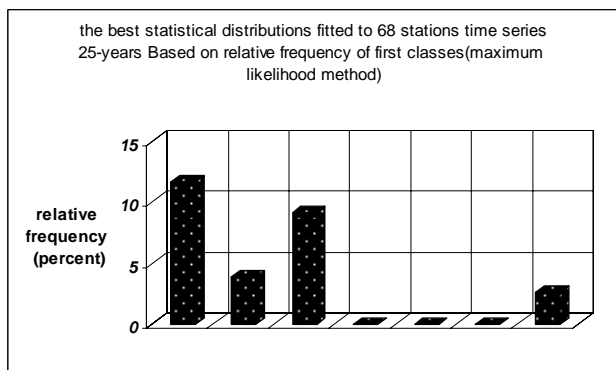
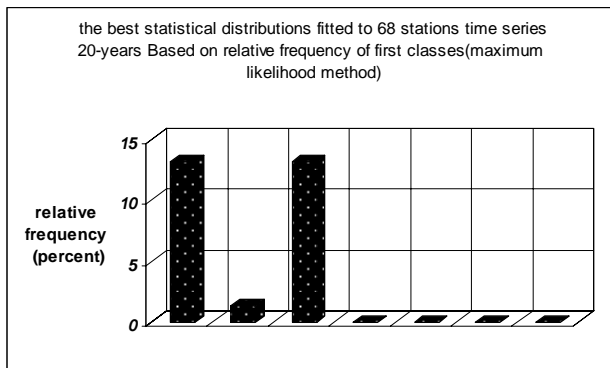
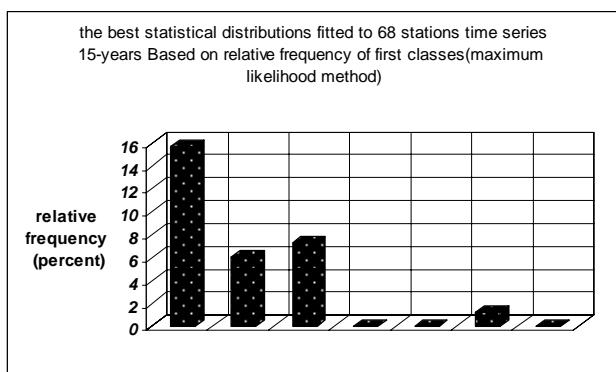
Figure 1. Location of the selected station in Mazandaran and Golestan provinces





Note: Distribution from left to right: normal, Two-Parameter Log-normal, Three-Parameter Log-normal, Two parameter gamma, Pearson Type III and Log Pearson Type III , Gumbel distributions

Figure 2. The best statistical distributions fitted to 65 stations in different time series 15, 20, 25 and 30-year based on relative frequency of first classes (moment method).



Note: Distribution from left to right: normal, Two-Parameter Log-normal, Three-Parameter Log-normal, Two parameter gamma, Pearson Type III and Log Pearson Type III, Gumbel distributions

Figure 3. The best statistical distributions fitted to 65 stations in different time series 15, 20, 25 and 30-year based on relative frequency of first classes (maximum likelihood method).