Evaluation and Selection of Streamflow Network Stations Using Entropy Methods

Nermin SARLAK
İnönü University, Department of Civil Engineering, Malatya-TURKEY
e-mail: nsarlak@inonu.edu.tr
A. Ünal SORMAN
Middle East Technical University, Department of Civil Engineering, Ankara-TURKEY

Abstract

The entropy methods provide reliable results in evaluating the performance of existing stream-gauging networks. Two entropy methods used to design network problems for selecting the priority stations in the multivariate case are presented in this study. One of these entropy methods is based on the combination of stations with the least transinformation, which was developed with normal and log-normal distributions. The other method, based on ranking stations, is applied for normal, log-normal, and gamma distributions. The aim of this study is to investigate the effect of distribution types on these entropy methods. For this reason, these entropy methods were applied under different distributions for the annual observations of 5 runoff stations in the Kızılırmak Basin. It was found that the stations selected to continue observations on the existing stream gauging network were not the same for different distribution types. This indicates that the distribution type for streamflow data is very important for these methods.

Key words: Entropy method, Normal distribution, Log-normal distribution, Gamma distribution.

Introduction

Sampling data in hydrology is essentially a way of communicating with the natural system, which is uncertain prior to the making of any observation. Each collected sample represents a signal from the natural system. Redundant information does not help to reduce the uncertainty further; it only increases the costs of obtaining data. On the basis of this analogy, a methodology based on the entropy concept of information theory has been developed for the evaluation of hydrological data networks (Özkul, 1996). Entropy is a measure of the degree of uncertainty of random hydrological processes. Since the reduction of uncertainty by means of making observations is equal to the amount of information gained, the entropy criterion indirectly measures the information content of a given series of data. Once the statistical structure of a process is known, its entropy can be computed and expressed in specific units (Harmancıoğlu, 1981). Amorocho and Espildora (1973) considered that the entropy concept, as defined by Shannon, provided satisfactory results in the comparison of various mathematical models developed for the same hydrological process, and in the selection of the most appropriate model. The concept of a hydrologic network as a communication channel, which is designed for transmitting hydrologic information, was introduced by Caselton and Husain (1980). Harmancıoğlu and Yevjevich (1987) used the entropy method on monthly observed data of an extremely polluted river basin. Entropy-based measures were used in this study to evaluate the goodness of information transfer by regression. The results of this study revealed that the association between most of the water quality variables was insignificant. Husain (1989) presented a simple methodology, using the entropy concept, to estimate regional hydrologic uncer-

tainty and information at both gauged and ungauged grids in a basin. Yang and Burn (1994) developed a new methodology for data collection network design. The approach employed a measure of the information flow between gauging stations in the network, which is referred to as directional information transfer. Non-parametric estimation was used to approximate the multivariate probability density functions, which were required in the entropy calculations. The directional information transfer is found useful in a network study for measuring the association between gauging stations. Harmancıoğlu et al. (2003) discussed an entropy-based approach for the assessment of combined spatial/temporal frequencies of monitoring networks. The results were demonstrated with water quality data observed along the Mississippi River in Louisiana, USA. The authors emphasized that the entropy method used was the best method for utilizing different techniques in combination, and to investigate network features from different perspectives before a final decision was made for network assessment and redesign. A new concept of entropy was developed for normal and log-normal distributions by Markus et al. (2003). The entropy approach was applied to information theory for evaluating stations through their information transmission to and from other stations. The aim of this study is to examine the results of 2 entropy methods, with different distribution assumptions, on a network of stream gauging stations located in the Kızılırmak Basin in order to rank the stations according to their level of importance. Stations' rankings based on the entropy methods for different distributions were obtained and the effects of the distribution types are shown.

Entropy Methods

There are 2 methods to design network problems in multivariate cases. The first method, Method 1, is proposed by Harmancıoğlu (1981). The objective of this method is to minimize transinformation by choosing an appropriate number of monitoring stations, using a stochastic approach, in spatial orientation in order to design network stations. The combination of stations with the least transinformation reflects the variability of the quality along the river without producing redundant information. Such an approach foresees the monitoring of a variable at points where there is the most variation or uncertainty. Accordingly, existing sampling sites can be

sorted in the order of decreasing uncertainty or decreasing informativeness. Thus, the first station is the one where the highest uncertainty about the variable occurs. The following stations serve to reduce this uncertainty further so that the last station offers the least amount of information. The other entropy approach (Method 2) is applied to information theory to evaluate stations through their information transmission to and from other stations (Markus et al., 2003). The following procedures for both Method 1 and Method 2 are applied to select the best combination of stations for the multivariate case.

Method 1:

The stochastic dependence between 2 processes causes their marginal entropies and the total entropy to decrease. The same is true for more than 2 variables that are stochastically dependent upon each other.

For the multivariate case, the total entropy of M stochastically independent variables X_m (m=1, ... ,M) is:

$$H(X_1, X_2, ..., X_M) = \sum_{m=1}^{M} H(X_m) \qquad (1)$$

If significant stochastic dependence occurs between the variables, the total entropy has to be expressed in terms of conditional entropies added to the marginal entropy of one of the variables (Özkul, 1996):

$$H(X_1, X_2, ..., X_M) = H(X_1) + \sum_{m=2}^{M} H(X_m|X_1, ..., X_{m-1}) \qquad (2)$$

As it is mentioned, entropy is a function of the probability distribution of a process. Therefore, the multivariate joint and conditional probability distribution functions of M variables should be determined to compute the related entropies:

$$H(X_1, X_2, ..., X_m) = -\int_{-\infty}^{\infty} .. \int_{-\infty}^{\infty} f(x_1, ..., x_m) \cdot$$
$$log f(x_1, ..., x_m) dx_1 dx_2...dx_m \qquad (3)$$

$$H(X_m|X_1,...,X_{m-1}) = -\int\limits_{-\infty}^{\infty}..\int\limits_{-\infty}^{\infty}f(x_1,...,x_m)\cdot$$
$$logf(x_m|x_1,...,x_{m-1})dx_1dx_2...dx_m$$

$$(4)$$

where f(x_1,..,x_m) is the multivariate probability density function of related distribution.

The next step in the computation of total, marginal, or conditional entropies is to determine the type of probability distribution function that best fits the analyzed process. Harmancıoğlu (1981) proposed the multivariate normal or log-normal probability distribution functions because of their simplicity in mathematical computations. If a multivariate normal distribution is assumed, the joint entropy of X is obtained using Eq. (5) (Harmancıoğlu, 1981):

$$H(X) = (M/2)\ell n2\pi + (1/2)\ell n|C| + M/2 \qquad (5)$$

where M is the number of variables and $|C|$ is the determinant of the covariance matrix C. Eq. (5) provides a single value for the entropy of M variables and the unit of entropy is napier, since logarithms are taken to the base e. If logarithms of observed values are used, the same procedure can be applied for log-normal distribution.

In the above formula, the covariance matrix C involves the cross covariances, C_{ij} of M different variables:

$$C = \begin{bmatrix} C_{11} & C_{12} & . & . & . & C_{1M} \\ C_{21} & C_{22} & . & . & . & C_{2M} \\ . & . & & . & . & . \\ C_{M1} & C_{M2} & . & . & . & C_{MM} \end{bmatrix}_{(M*M)} \qquad (6)$$

For a single variable, the covariance matrix includes the autocovariances as a measure of the serial dependence within the process. When both the serial and cross covariances are considered, the matrix includes both the auto and cross covariances (Harmancıoğlu, 1981; Özkul, 1996):

$$C = \begin{bmatrix} C_{11}(0) & ... & C_{11}(K) & ... & C_{1M}(0) & ... & C_{1M}(-K) \\ C_{11}(K) & ... & C_{11}(0) & ... & C_{1M}(K) & ... & C_{1M}(0) \\ ... & ... & ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... & ... & ... \\ C_{M1}(0) & ... & C_{M1}(-K) & ... & C_{MM}(0) & ... & C_{MM}(K) \\ C_{M1}(K) & ... & C_{M1}(0) & ... & C_{MM}(K) & ... & C_{MM}(0) \end{bmatrix}_{[M*(K+1)][M*(K+1)]} \qquad (7)$$

The covariance matrix in Eq. (7) is a nonsingular, positive-definite, and symmetrical matrix. Eq. (5) can also be used in the calculation of conditional entropies as the difference between joint entropies for 3 variables:

$$H(X|Y,Z) = H(X,Y,Z) - H(Y,Z) \qquad (8)$$

Consequently, in Method 1, existing stations in a basin are listed in the order of priority. The benefits for each combination of sampling sites are measured in terms of the least transinformation or the highest conditional entropy produced by that combination. Therefore, addition or elimination of new stations lead to a decrease or increase in transinformation and conditional entropies (Özkul, 1996).

Method 2:

Marcus et al. (2003) defined a new concept of entropy, which is a fractional reduction of the entropy of X by R(X, Y):

$$R(X,Y) = \frac{T(X,Y)}{H(X)} \qquad (9)$$

which also can be viewed as a reduction of the uncertainty of X if Y is known, or information received by X from Y. H(X) is the marginal entropy of a single variable X. T(X,Y) is the information transferred from X to Y, the transinformation.

$$H(X) = \int\limits_{-\infty}^{\infty} f(x) \log \frac{1}{f(x)} dx \qquad (10)$$

where f(x) is the probability density function of related distribution.

$$T(X,Y) = H(X) + H(Y) - \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x,y) \log \frac{1}{f(x,y)} dxdy. \qquad (11)$$

Similarly, the information sent (transmitted) from X to Y is defined as:

$$S(X,Y) = \frac{T(X,Y)}{H(Y)}. \qquad (12)$$

Equations (9) and (12), which describe the relationship between 2 variables, X and Y, are adapted to the network of stream gauges (Markus et al., 2003). Using these equations, information received and sent to station m is defined as:

$$R(m) = R(X(m), \hat{X}(m)) \qquad (13)$$

$$S(m) = S(X(m), \hat{X}(m)) \qquad (14)$$

where X(m) represents the data at site m. The quantity, $\hat{X}(m)$at station m, is obtained by multiple linear regression as:

$$\hat{X}(m) = a(m) + \sum_{j=1}^{M-1} Y_j(m).b_j(m) \qquad (15)$$

where Y_j(m) is a matrix of data from all other stations, a(m) and b(m) are parameters of the multiple regression between site m and all other sites, and M is the number of stations. As the relationships between data at different sites are found to be linear or close to linear, this assumption of linearity is deemed appropriate.

In this method, the concept of entropy is used to determine the stations with the highest amounts of S(m) and R(m). If R(m) is large relative to other stations, it indicates that the station denoted as m

receives a lot of information. On the other hand, stations sending more information, having larger S(m), are considered more valuable and should remain active. Finally, the net information transfer, N(m), is defined as the difference between S(m) and R(m):

$$N(m) = S(m) - R(m) \qquad (16)$$

Stations with positive N(m) are considered more valuable in regional analyses. If the number of stations in the network is to be reduced, such a station is more likely to be retained in the network than a station with a negative N(m) (Markus et al., 2003).

Method 2 can be applied for normal, log-normal, and gamma distributions. The marginal and joint entropy terms are calculated. The computation of these terms for normal and log-normal distributions is straightforward. Appropriate probability density functions of the distributions are then incorporated into the related equations. In the case of gamma distribution, the parameters, σ and λ, can be calculated using the derived method of moment estimators:

i. The expected value of X: $\mu_1(X) = \sigma\lambda$ (17a)

ii. The variance of X: $\mu_2(X) = \sigma^2\lambda$ (17b)

The marginal entropy of the gamma probability density function was defined by Husain (1989) as:

$$H(X) = -(\lambda - 1)\psi(\lambda) + \Gamma(\lambda+1)/\Gamma(\lambda) + \ell n(\sigma\Gamma(\lambda)) \qquad (18)$$

where $\psi(\lambda)$ is the digamma function:

$$\psi(\lambda) = \partial/\partial\lambda (\ell n\Gamma(\lambda)). \qquad (19)$$

Due to limitations in the derivation of bivariate gamma distribution functions and complexities in their mathematical computations, application of bivariate gamma distribution functions is very limited. However, the bivariate gamma distribution, as proposed by Husain (1989), can be transformed to normalized variates, z and w. The information transmission relationship is defined by Eqs. (20a) and (20b):

$$1 \Big/ \sqrt{2\pi} \int\limits_{-\infty}^{z} e^{-0.5t^2} dt = \int\limits_{0}^{X} f(t; \sigma_x, \lambda_x) dt \qquad (20a)$$

$$1 \Big/ \sqrt{2\pi} \int\limits_{-\infty}^{w} e^{-0.5t^2} dt = \int\limits_{0}^{Y} f(t; \sigma_y, \lambda_y) dt \qquad (20b)$$

In the above expressions, X and Y are variables with univariate gamma distributions with parameters (σ_x, λ_x) and (σ_y, λ_y), respectively. z and w are normalized variates of X and Y, respectively, with a mean of zero and standard deviation of unity. If ρ_{zw} is the correlation coefficient between z and w, then the information transmitted by variable Y about X, i.e. T(X,Y), or by variable X about Y, i.e. T(Y, X) is simplified as (Husain, 1989):

$$T(X, Y) = T(Y, X) = -\frac{1}{2} \ell n \left(1 - \rho_{zw}^2\right). \qquad (21)$$

Case study of Kızılırmak River Streamflow Gauging Stations in Turkey

Five streamflow gauging stations located along the Kızılırmak River were selected as the data set used in this study of the application of 2 entropy methods, since it has more water monitoring stations and a

longer history of record keeping than any other river in Turkey. As can be seen in Figure 1, station EIE 1501 is located before the Hirfanlı, Kesikköprü, and Kapulukaya Dams. The other station, EIE 1541 is located on the tributary (Delice) of the main river, which is the most important stream of the Kızılırmak River. The remaining stations, EIE 1503, EIE 1528, and EIE 1536, are located downstream of existing dams. Since these 3 streamflow stations have been affected by existing dams, records from these stations were converted to natural streamflow characteristics. The streamflow gauging station (i.e. EIE 1501) located upstream of these dams was used as the reference to obtain the natural values (unaffected form) of the downstream streamflow characteristics at these 3 stations. For this reason, seasonal correlations of the streamflow values, which have been observed before the construction of the dams, were obtained both for the affected and the unaffected gauging stations. Using these seasonal correlation coefficients, streamflow values of the affected gauging stations were adjusted by regression equation in order to obtain the natural records. The inflow and outflow data of the dam reservoirs were collected for processing from The State Hydraulic Works (DSI)
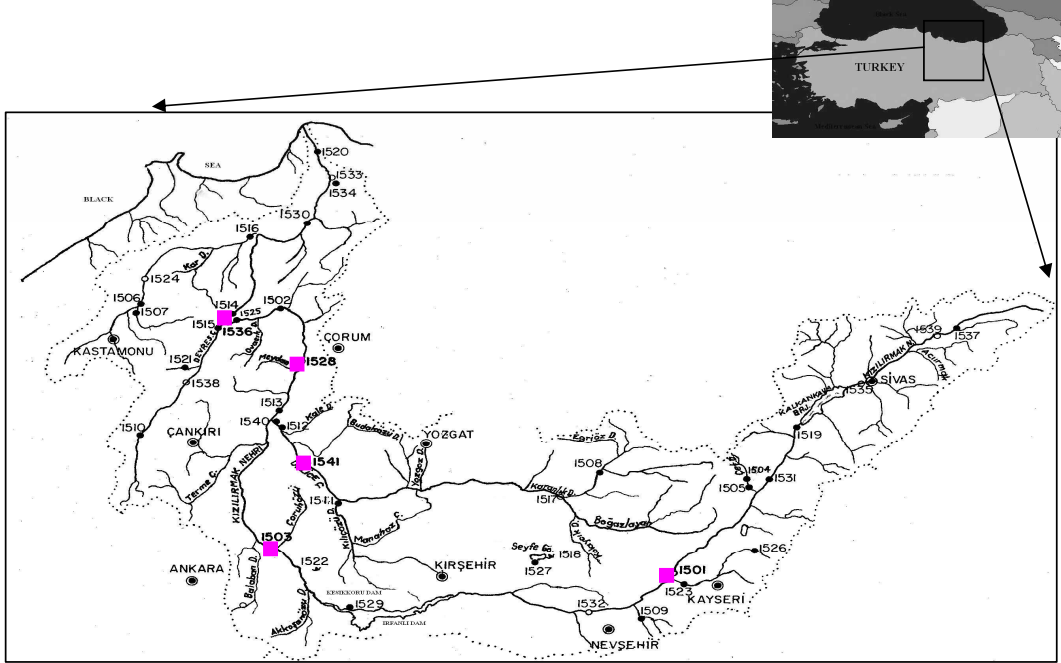

Figure 1. Kızılırmak Basin and location of the streamflow gauging stations.

95

(Sarlak, 2005). They were used to obtain the change in storage of the dams' reservoirs ($\pm\Delta S$) so that the corrected streamflow values could be checked. After obtaining $\pm\Delta S$, these values were compared with the differences of affected and natural streamflows in order to find out whether there was a correlation between them, and significant correlations (0.80) were found. Therefore, it was assumed that the converted values of the natural cases were acceptable.

Application of Method 1 (Normal and Log-normal Distributions)

The methodology described for Method 1 was applied to the annual streamflow values of 5 stream gauging stations in the Kızılırmak River, according to the normal and log-normal distributions. Although the period of observation varied for each station, a common period of 41 years, between 1955 and 1995, was considered for all stations. The procedure of Method 1 for selecting the best combination of stations, based on the minimum transinformation principle of normal distribution, is summarized below. Using the logarithm of the data set, the same procedure was also followed for the log-normal distribution case.

i. Since 5 stream gauging stations were considered in this study, the data set for each station was represented by X_m, where m (m=1, ... ,M) represents the station number (M=5).

ii. The marginal entropy H (X_m) of the variable for each station was computed, first by using Eq. (5), where M was replaced by 1. As can be seen from Table 1, the marginal entropy value of the EIE 1536 streamflow gauging station was greater than the marginal entropy value of the other stations. Because of this, it was selected as the first priority station, X_1, to continue

making observations. In other words, the highest uncertainty occurred concerning the variables in this location.

iii. Next, the selected station (EIE 1536) was coupled with every other station in the network to select the pair that led to the least transinformation. The EIE 1541 station, which fulfilled this condition, was marked as the second priority location, X_2. A pair of stations was selected that had the highest joint entropy and the least transinformation. Accordingly, these stations produced the highest amount of information when they were run together.

iv. As a third step, the conditional entropies and transinformations of the (X_1, X_2) pair were computed with every other station in the network to select the third station with the least transinformation.

v. The same procedure was continued by considering successive combinations of 4 and 5 stations and selecting the combination that produced the least transinformation and redundant information.

vi. The stations were ranked according to their priority orders. For example, the highest rank (r=5) represents the first priority order, which means that this station was necessary for this network to remove uncertainty.

The stations selected using Method 1, with their priority orders for normal and log-normal distributions, are presented in Tables 1 and 2, respectively.

When each streamflow series was assumed to be normally distributed, EIE 1536, which was the station located furthest downstream on the main river, was selected as the highest priority station using Method 1. Moreover, Method 1 selected the station farthest upstream on the tributary as the station

Table 1. Selection of sampling stations for normal distribution.

Station no.	Station Added	Marginal Entropy	Joint Entropy	Conditional Entropy	Transinfor mation	Rank (r)
	(M)	(napierian)	(napierian)	(napierian)	(napierian)	-
1536	1	8.74	8.74	-	-	5
1541	2	5.88	12.82	6.94	1.795	4
1501	3	7.56	18.06	10.50	2.319	3
1528	4	8.47	22.89	14.42	3.630	2
1503	5	7.97	25.83	17.85	5.039	1

with the second highest priority. Accordingly, stations EIE 1536 and EIE 1541 constituted the pair with the least amount of redundant information. The third location was station EIE 1501, which was located farthest upstream on the main river. As can be seen in Table 1, the joint entropy and transinformation values increased, contributing the other stations to the network under consideration. The percentages of redundant information varied with the addition of each new station to the combination. Decisions can be made about which stations in the network should be discontinued according to the amount of transinformation, which was determined beforehand.

The results of Method 1 under log-normal distribution for this network, which was similar to normal distribution computations, are presented in Table 2. It is observed from Table 2 that stations EIE 1541 and EIE 1501 were selected as the first and the second priority stations. As expected, the last station in the list was EIE 1528, since it produced information that was highly redundant with that of station EIE 1536, due to its location. Transinformation values also increased under log-normal distribu-

tion. However, the amount of transinformation values showed differences under normal and log-normal distributions.

This emphasized that the selection of an appropriate distribution type is very important before beginning any analysis.

Application of Method 2 (Normal and Log-normal distributions)

Method 2 was also applied to the Kızılırmak Basin. Equations (9) and (13) were used to compute the total information received by station m; R (m). Equations (12) and (14) were then used to compute the total information sent by station m; S (m). Finally, Eq. (16) was used to compute the total net information transfer for station m; N (m). As previously mentioned, if logarithms of observed values are used, the same procedure can be employed for log-normal distribution. The information transfer parameters S (m), R (m), and N (m), as well as the station ranks based on these parameters are shown in Table 3 (normal distribution) and in Table 4 (log-normal distribution).

Table 2. Selection of sampling stations for log-normal distribution.

Station no.	Station Added	Marginal Entropy	Joint Entropy	Conditional Entropy	Transinformation	Rank (r)
	(M)	(napierian)	(napierian)	(napierian)	(napierian)	-
1541	1	-2.08	-2.08	-	-	5
1501	2	-2.58	-5.79	-3.22	1.14	4
1536	3	-3.06	-10.40	-7.34	1.54	3
1503	4	-2.88	-17.12	-14.24	3.84	2
1528	5	-3.08	-30.10	-27.03	9.91	1

Table 3. Station ranking according to information transmitted, S(m), information received, R(m), and net information, N(m), for normal distribution.

Station no.	Information transfer			Rank (r)		
	Send	Received	Net	Send	Received	Net
	S(m)	R(m)	N(m)	S(m)	R(m)	N(m)
	(napierian)	(napierian)	(napierian)	-	-	-
1501	0.9229	0.9228	0.0002	2	2	2
1503	0.9251	0.9238	0.0013	3	3	5
1541	0.9181	0.9178	0.0004	1	1	4
1528	1.2347	1.2350	-0.0003	5	5	1
1536	1.0896	1.0894	0.0002	4	4	3

Table 4. Station ranking according to information transmitted, S(m), information received, R(m), and net information, N(m), for log- normal distribution.

Station no.	Information transfer			Rank (r)		
	Send	Received	Net	Send	Received	Net
	S(m)	R(m)	N(m)	S(m)	R(m)	N(m)
	(napierian)	(napierian)	(napierian)	-	-	-
1501	-96.05	-96.91	0.139	5	5	4
1503	-358.21	-358.57	-0.358	3	3	2
1541	-118.05	-117.50	0.551	4	4	5
1528	-2320.40	-2320.92	-0.524	1	1	1
1536	-1892.09	-1892.38	-0.287	2	2	3

The stations having the lowest rank (r = 1 or r = 2) were less important in the information transfer process. It was not necessary to continue observations at these stations. On the other hand, the stations having higher ranks (r = 4 or r = 5), according to this method, should be retained in the network.

While station EIE1536 was selected as the first priority station by Method 1 under normal distribution assumption, the same station was selected by Method 2 as the station with the third highest priority for remaining in the streamflow gauging network, according to net information. Similarly, station EIE 1503, which was chosen as the station with the least priority by Method 1 with normal distribution, was selected as the most important station by Method 2 with normal distribution. Nonetheless, it was observed that while the results of the 2 methods were varied under normal distribution assumption, the results for the methods were the same under log-normal distribution.

Application of Method 2 (Gamma distributions)

Method 2 for gamma distribution was applied using Eq. (18) for marginal entropy and Eq. (21) for transinformation. The rest of the procedure for Method 2 for gamma distribution was exactly the same as it was for the other distributions. Finally, the information transfer parameters S (m), R (m), and N (m) for gamma distribution are shown in Table 5. According to Method 2 results obtained from gamma distribution, station EIE 1503 was the most important station to remain in this network. Moreover, station EIE 1501, which was the station farthest upstream, was selected as the least important station for the information transfer process.

As mentioned earlier, stations' ranking varied for both methods under different distributions. However, with log-normal distribution, the stations' rankings were similar. Nevertheless, it was observed that this ordering under log-normal distribution was quite reasonable if compared with the results of the other distributions. Since the stations farthest upstream and located in different tributaries (EIE 1501 and EIE 1541) had the highest uncertainty, these stations were selected as the second and the third priority stations for this network, as would be expected.

Although the aim of this study is to demonstrate the affect of the distribution type on each entropy method, the underlying distributions of data sets were checked by goodness-of-fit tests, based on statistical theory. Among them, the Kolmogorov-Smirnov and Shapiro-Wilk tests stand out. The results of these tests are summarized in Table 6. According to these tests for each series, the distribution types were obtained as log-normal. Therefore, this can explain the similarity of station rankings determined by the 2 entropy methods with log-normal distribution.

Summary and Conclusions

In this study, 2 entropy methods were applied to a network of gauging stations located in the Kızılırmak Basin. The aim of the study is to rank the stations according to the priority for their selection to remain in the network, which is dependent upon their information contribution. The combination of stations in Method 1 was selected with the minimum transinformation. The first station selected in this method was the station having the highest priority, which means that this station must be retained in the network. The stations were also ranked by Method 2. Lower ranked stations indicate less importance on a network and these are the stations that could be

Table 5. Station ranking according to information transmitted, S(m), information received, R(m), and net information, N(m), for gamma distribution.

Station no.	Information transfer			Rank (r)		
	Send	Received	Net	Send	Received	Net
	S(m)	R(m)	N(m)	S(m)	R(m)	N(m)
	(napierian)	(napierian)	(napierian)	-	-	-
1501	0.6349	0.6350	-0.000121	1	1	1
1503	0.6903	0.6902	0.000098	2	2	5
1541	0.7200	0.7201	-0.000096	3	3	2
1528	1.0479	1.0479	0.000006	5	5	3
1536	0.9935	0.9935	0.000025	4	4	4

Table 6. The results of Shapiro-Wilk and Kolmogorov-Smirnov tests.

Station no.	Shapiro-Wilk test		Kolmogorov-Smirnov tests	
	R^2	R^2, critical	d values	d, critical values $\alpha = 0.20$
1501	0.98	≈ 1.0	0.065263	< 0.167
1503	0.98	≈ 1.0	0.108561	< 0.167
1541	0.97	≈ 1.0	0.081364	< 0.167
1528	0.98	≈ 1.0	0.110004	< 0.167
1536	0.98	≈ 1.0	0.10914	< 0.167

discontinued. Higher ranks indicate the stations that should be retained in the network. For the multivariate case of Method 1, the multivariate density function must be applied, whereas the bivariate density function of distribution is adequate for Method 2. Although the mathematical definition of entropy is easily developed for skewed distributions in bivariate cases, the computational procedure becomes much more difficult when multivariate distributions are considered. Because of this, Method 1 is applied with normal and log-normal distributions, while Method 2 is applied for normal, log-normal, and gamma distributions.

In order to demonstrate the effect of the distribution type on each entropy method, station rankings with Method 1 and Method 2, for different distributions, were obtained. It was observed that the priority level of each station on the existing stream gauging network changed with different distribution types, as shown in Figure 2.

As a result, in the determination of appropriate distribution for a streamflow series, it is important to rely on the results, which can be obtained from entropy methods. For example, it is obvious that planning a stream gauging network system under normal distribution for data sets having right-skewed distribution may not be reliable.

From this point of view, the selection of appropriate distributions for variables is a crucial part of any issue in which an optimum network system is to be planned with entropy methods. For that reason, the authors emphasize that the distribution types for data series should be determined properly before applying the 2 entropy methods. Otherwise, the results obtained may be invalid.

The best fitting probability distribution for each station was determined as log-normal by 2 goodness-of-fit tests, the Shapiro-Wilk and Kolmogorov-Smirnov tests. According to the Shapiro-Wilk test, the squared correlation values for log-normal distribution for each station were calculated and are shown in Table 6. These values were quite close to one. Therefore, the null hypothesis, in which the distributions of the population were log-normal, is accepted for each series. The Kolmogorov-Smirnov test is a distance test; the maximum distance is obtained between the theoretical and empirical distributions. These values and critical values for each series are
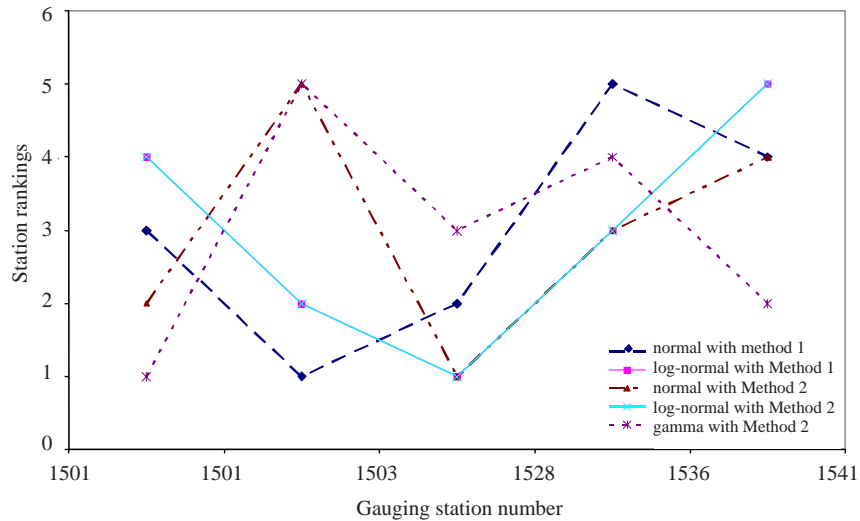
Figure 2. Station rankings based on 2 different entropy methods for different distributions using annual discharges.

presented in Table 6. Since these values were less than the corresponding table values (critical values), the null hypothesis that the distributions of the population were log-normal distribution is accepted for each series, as well.

Therefore, the ranking for log-normal distribution can be adopted for the Kızılırmak network in this study.

Acknowledgment

References

Amorocho, J. and Espildora, B., "Entropy in the Assessment of Uncertainty of Hydrologic Systems and Models", Water Resources Research, 9, 1511-1522, 1973.

Caselton, W.F. and Husain T., "Hydrologic Networks: Information Transinformation", Journal of the Water Resources Planning and Management Division, ASCE, 106, WR2, July, 503-529, 1980.

Harmancıoğlu N., "Measuring the Information Content of Hydrological Process by the Entropy Concept", Ege University, The Journal of Civil Engineering Faculty, The Special Issue of 100th Birth Year of Ataturk. 13-40, Izmir, 1981.

Harmancıoğlu, N.B. and Yevjevich, V., "Transfer of Hydrologic Information among River Points", Journal of Hydrology, 91, 103-118, 1987.

Harmancıoğlu, N.B., Fistikoglu, O. and Özkul, S., "Integrated Technologies for Environmental Monitoring and Information Production", Kluwer Academic Publishers, 496pp, 2003.

Husain, T., "Hydrologic Uncertainty Measure and Network Design", Water Resources Bulletin, 25, 527-534, 1989.

Markus M., Knapp H.V. and Tasker G.D., "Entropy and Generalized Least Square Methods in Assessment of the Regional Value of Streamgages", Journal of Hydrology 283, 107-121, 2003.

Sarlak N., "Evaluation and Modeling of Streamflow Data: Entropy Method, Autoregressive Models with Asymmetric Innavations and Artificial Neural Networks", PhD thesis, Metu, Ankara, Turkey (2005).

Özkul S., "Space / Time Design of Water Quality Monitoring Networks by the Entropy Method", PhD thesis, Dokuz Eylül University, Faculty of Civil Engineering, İzmir, 169 pages, (1996).

Yang, Y. and Burn, D.H., "An Entropy Approach to Data Collection Network Design." Journal of Hydrology, 157, 307-324, 1994.