entropy

MDPI

Article

An Extension to the Revised Approach in the Assessment of Informational Entropy

Turkay Baran, Nilgun B. Harmancioglu *, Cem Polat Cetinkaya and Filiz Barbaros

Faculty of Engineering, Civil Engineering Department, Dokuz Eylul University, Tinaztepe Campus, Buca, 35160 Izmir, Turkey; turkay.baran@deu.edu.tr (T.B.); cem.cetinkaya@deu.edu.tr (C.P.C.); filiz.barbaros@deu.edu.tr (F.B.)

* Correspondence: nilgun.harmancioglu@deu.edu.tr; Tel.: +90-542-413-9300

Received: 29 September 2017; Accepted: 20 November 2017; Published: 29 November 2017

Abstract: This study attempts to extend the prevailing definition of informational entropy, where entropy relates to the amount of reduction of uncertainty or, indirectly, to the amount of information gained through measurements of a random variable. The approach adopted herein describes informational entropy not as an absolute measure of information, but as a measure of the variation of information. This makes it possible to obtain a single value for informational entropy, instead of several values that vary with the selection of the discretizing interval, when discrete probabilities of hydrological events are estimated through relative class frequencies and discretizing intervals. Furthermore, the present work introduces confidence limits for the informational entropy function, which facilitates a comparison between the uncertainties of various hydrological processes with different scales of magnitude and different probability structures. The work addresses hydrologists and environmental engineers more than it does mathematicians and statisticians. In particular, it is intended to help solve information-related problems in hydrological monitoring design and assessment. This paper first considers the selection of probability distributions of best fit to hydrological data, using generated synthetic time series. Next, it attempts to assess hydrometric monitoring duration in a netwrok, this time using observed runoff data series. In both applications, it focuses, basically, on the theoretical background for the extended definition of informational entropy. The methodology is shown to give valid results in each case.

Keywords: uncertainty; information; informational entropy; variation of information; continuous probability distribution functions; confidence intervals

1. Introduction

The concept of entropy has its origins in classical thermodynamics and is commonly known as "thermodynamic entropy" in relation to the second law of thermodynamics. Such a non-probabilistic definition of entropy has been used widely in physical sciences, including hydrology and water resources. Typical examples on the use of "thermodynamic entropy" in water resources involve problems associated with river morphology and river hydraulics [1,2].

Boltzmann's definition of entropy as a measure of disorder in a system was given in probabilistic terms and constituted the basis for statistical thermodynamics [3–5]. Later, Shannon [6] followed up on Boltzmann's definition, claiming that the entropy concept could be used to measure disorder in systems other than thermodynamic ones. Shannon's entropy is what is known as "informational entropy", which measures uncertainty (or, indirectly, information) about random processes. As uncertainty and information are the two most significant yet the least clarified problems in hydrology and water resources, researchers were intrigued by the concept of informational entropy. Thus, it has found a large number of diverse applications in water resources engineering.

Within a general context, the entropy principle is used to assess uncertainties in hydrological variables, models, model parameters, and water-resources systems. In particular, versatile uses of the concept range from specific problems, such as the derivation of frequency distributions and parameter estimation, to broader cases such as hydrometric data network design. The most distinctive feature of entropy in these applications is that it provides a measure of uncertainty or information in quantitative terms [7–19].

On the other hand, researchers have also noted some mathematical difficulties encountered in the computation of various informational entropy measures. The major problem is the controversy associated with the mathematical definition of entropy for continuous probability distribution functions. In this case, the lack of a precise definition of informational entropy leads to further mathematical difficulties and, thus, hinders the applicability the concept in hydrology. This problem needs to be resolved so that the informational entropy concept can be set on an objective and reliable theoretical basis and thereby achieve widespread use in the solution of water-resources problems based on information and/or uncertainty.

Some researchers [20,21] attempted to revise the prevailing definition of informational entropy, where entropy relates to the amount of reduction of uncertainty, or indirectly to the amount of information gained through measurements of a random variable. The study presented extends on the revised definition of Jaynes [20] and Guiasu [21] to describe informational entropy, not as an absolute measure of information, but as a measure of the variation of information. The mathematical formulation developed herein does not depend on the use of discretizing intervals when discrete probabilities of hydrological events are estimated through relative class frequencies and discretizing intervals. This makes it possible to obtain a single value for the variation of information instead of several values that vary with the selection of the discretizing interval. Furthermore, the extended definition introduces confidence limits for the entropy function, which facilitates a comparison between the uncertainties of various hydrological processes with different scales of magnitude and different probability structures.

It must be noted that the present work is intended for hydrologists and environmental engineers more than for mathematicians and statisticians. In particular, entropy measures have been used to help solve information-related problems in hydrological monitoring design and assessment. These problems are manifold, ranging from the assessment of sampling frequencies (both temporal and spatial) and station discontiuance to statistical analyses of observed data. For the latter, this paper considers the selection of probability distributions of best fit to hydrological data. Hence, the informational entropy concept is used here only in the temporal domain. To test another feature of entropy measures, the present work also attempts to assess hydrometric monitoring duration in a gauging network, this time using observed runoff data series. In both applications, the paper focuses, basically, on the theoretical background for the extended definition of informational entropy, and the results are shown to give valid results.

2. Mathematical Difficulties Associated with Informational Entropy Measures

Entropy is a measure of the degree of uncertainty of random hydrological processes. It is also a quantitative measure of information contained in a series of data since the reduction of uncertainty equals the same amount of gain in information [7,22]. Within the scope of Mathematical Communication Theory, later known as Information Theory, Shannon [6] and later Jaynes [23] defined informational entropy as the expectation of information or, conversely, as a measure of uncertainty. If S is a system of events, $E_1, E_2, \ldots, E_n$, and $p(E_k) = p_k$ the probability of the k-th event recurring, then the entropy of the system is:

$$H(S) = -\sum_{k=1}^{n} p_k \ln p_k \quad (1)$$

With,

$$\sum_{k=1}^{n} p_k = 1$$

Shannon's entropy as given in Equation (1) is originally formulated for discrete variables and always assumes positive values. Shannon extended this expression to the continuous case by simply replacing the summation with an integral equation as:

$$H(X) = -\int_{-\infty}^{+\infty} f(x) \cdot \ln f(x) \cdot dx \tag{2}$$

With,

$$\int_{-\infty}^{+\infty} f(x) \cdot dx = 1$$

For the random variable $X \in (-\infty, +\infty)$, and where $H(X)$ is denoted as the marginal entropy of X, i.e., the entropy of a univariate process. Equation (2) is not mathematically justified, as it is not valid under the assumptions initially made in defining entropy for the discrete case. What researchers proposed for solving this problem has been to approximate the discrete probabilities p_k by $f(x)\Delta x$, where $f(x)$ is the relative class frequency and Δx, the size of class intervals. Under these conditions, the selection of Δx becomes a crucial problem, such that each specified class interval size gives a different reference level of zero uncertainty with respect to which the computed entropies are measured. In this case, various entropy measures become relative to the discretizing interval Δx and change in value as Δx changes. The unfavorable result here is that the uncertainty of a random process may assume different values at different selected values of Δx for the same variable and the same probability distribution function. In certain cases, the entropy of a random variable even becomes negative [16,17,22,24–27], a situation which contradicts Shannon's definition of entropy as the selection of particular Δx values produces entropy measures varying within the interval $(-\infty, +\infty)$. On the contrary, the theoretical background for the random variable X, $H(X)$ defines the condition:

$$0 \leq H(X) \leq \ln N \tag{3}$$

where N is the number of events X assumes. The condition above indicates that the entropy function has upper (In N) and lower (0 when X is deterministic) bounds, assuming positive values in between [6,8,10–13,16,17,22,24–28]. The discrepancies encountered in practical applications of the concept essentially result from the above errors in the definition of entropy for continuous variables.

Another significant problem is the selection of the probability distribution function to be used in the definition of entropy, as in Equation (2). The current expression for continuous entropy produces different values when different distribution functions are assumed for the same variable. In this case, there is the need for a proper selection of the distribution function which best fits the process analyzed. One may consider here a valid criterion in the form of confidence limits to assess the suitability of the selected distribution function for entropy computations.

Further problems are encountered when the objective is to compare the uncertainties of two or more random variables with widely varying means and thus with different scales of magnitude. For instance, if entropy values are computed, using the same discretizing interval Δx, for two variables with means of 100 units and 1 unit, respectively, the results become incomparable due to the improper selection of the reference level of zero uncertainty for each variable. Such a problem again stems from the inclusion of the discretizing interval Δx in the definition of entropy for continuous variables. Comparison of uncertainties of different variables is an important aspect of entropy-based hydrometric network design procedures, where the aforementioned problem leads to subjective evaluations of information provided by the network [7,19].

It follows from the above discussion that the main difficulty associated with the applicability of the informational entropy concept in hydrology is the lack of a precise definition for the case of the continuous variables. It is intended in this study to resolve this problem by extending the revised approach proposed by Guiasu [21] so that the informational entropy can be set on an objective and reliable theoretical basis in order to discard subjective assessments of information conveyed by hydrological data or of the uncertainty of hydrological processes.

3. The Revised Definition of Informational Entropy for Continuous Variables

To solve the difficulties associated with the informational entropy measure in the continuous case, some researchers have proposed the use of a function $m(x)$ such that the marginal entropy of a continuous variable X is expressed as:

$$H(X) = -\int_{-\infty}^{+\infty} f(x) \cdot ln\left[\frac{f(x)}{m(x)}\right] \cdot dx \tag{4}$$

"where $m(x)$ is an 'invariant measure' function, proportional to the limiting density of discrete points" [20]. The approach seemed to be statistically justified; however, it still remained uncertain what the $m(x)$ function might represent in reality. Jaynes [20] also discussed that it could be an a priori probability distribution function, but there were then controversies over the choice of a priori distribution such that the problem was unresolved [8].

In another study, Guiasu [21] referred to Shannon's definition of the informational entropy for the continuous case. He considered that the entropy $\{H_S\}$ for the continuous variable X within an interval $[a, b]$ is:

$$H_S = -\int f(x) \cdot \ln f(x) \cdot dx \tag{5}$$

When the random variable assumes a uniform probability distribution function as:

$$f(x) = \frac{1}{(b-a)} \; x \in [a, b] \tag{6}$$

Then the informational entropy H_S for the continuous case within this interval can be expressed as:

$$H_S = \ln(b-a) \tag{7}$$

If the interval $[a, b]$ is discretized into N equal intervals, the variable follows a discrete uniform distribution and its entropy $\{H_N\}$ can be expressed as:

$$x \in [a, b] \tag{8}$$

When N goes to infinity, H_N will also approach infinity. In this case, Guiasu [21] claims that, although H_S and H_N are similarly defined, H_S will not approach H_N when $N \to \infty$. Accordingly, Guiasu [21] proposed an expression similar to that of Jaynes [20] for informational entropy in the continuous case as:

$$H(X/X^*) = -\int f(x) \cdot \ln\left[\frac{f(x)}{m(x)}\right] \cdot dx \tag{9}$$

which he called as the variation of information. In Equation (9), X^* represents a priori information (i.e., information available before making observations on the variable X) and X is the a posteriori information (i.e., information obtained by making observations). Similarly, $m(x)$ is the a priori and $f(x)$ the a posteriori probability density function for the random variable X.

In previous studies by the authors [8,10–13], informational entropy has been defined as the variation of information, which indirectly equals the amount of uncertainty reduced by making observations. To develop such a definition, two measures of probability, p and q with (p and $q \in$ K),

are considered in the probability space (Ω, K). Here, q represents a priori probabilities (i.e., probabilities prior to making observations). When a process is defined in such a probability space, the information conveyed when the process assumes a finite value A $\{A \in \mathrm{K}\}$ in the same probability space is:

$$I = -\ln\left(\frac{p(A)}{q(A)}\right) \tag{10}$$

The process defined in Ω can assume one of the finite and discrete events $(A_1, \ldots, A_n) \in \mathrm{K}$; thus, the entropy expression for any value A_n can be written as:

$$H(p/q) = -\ln\left(\frac{p(A_n)}{q(A_n)}\right)(n = 1, \ldots, N) \tag{11}$$

The total information content of the probability space (Ω, K) can be defined as the expected value of the information content of its elementary events:

$$H(p/q) = -\sum p(A_n) \cdot \ln\left(\frac{p(A_n)}{q(A_n)}\right) \tag{12}$$

Similarly, the entropy $H(X/X^*)$ of a random process X defined in the same probability space can be defined as:

$$H(X/X^*) = -\sum p(x_n) \cdot \ln\left(\frac{p(x_n)}{q(x_n)}\right) \tag{13}$$

where, $H(X/X^*)$ is in the form of conditional entropy, i.e., the entropy of X conditioned on X^*. Here, the condition is represented by an a priori probability distribution function, which can be described as the reference level against which the variation of information in the process can be measured.

Let us assume that the a priori $\{q(x)\}$ and a posteriori $\{p(x)\}$ probability distribution functions of the random variable X are known. If the ranges of possible values of the continuous variable X are divided into N discrete and infinitesimally small intervals of width Δx, the entropy expression for this continuous case can be given as:

$$H(X/X^*) = -\int p(x) \cdot \ln\left(\frac{p(x)}{q(x)}\right) \cdot dx \tag{14}$$

The above expression describes the variation of information (or, indirectly, the uncertainty reduced by making observations) to replace the absolute measure of information content given in Equation (2). This definition is essentially in conformity with those given by Jaynes [20] and Guiasu [21] for continuous variables. When the same infinitesimally small class interval Δx is used for the a priori and a posteriori distribution functions, the term Δx drops out in the mathematical expression of marginal entropy in the continuous case. Thus, this approach eliminates the problems pertaining to the use of Δx discretizing class intervals involved in the previous definitions of informational entropy [8,10–13].

At this point, the most important issue is the selection of a priori distribution. In case the process X is not observed at all, no information is available about it so that it is completely uncertain. In probability terms, this implies the selection of the uniform distribution. In other words, when no information exists about the variable X, the alternative events it may assume may be represented by equal probabilities or simply by the uniform probability distribution function.

If the a priori $\{q(x)\}$ is assumed to be uniform, and a posteriori $\{p(x)\}$ distribution of X is assumed to be normal, the informational entropy $H(X/X^*)$ can be expressed as:

$$H(X/X^*) = \ln\sqrt{2\pi} + \ln\sigma + \frac{1}{2} - \ln(b - a) \tag{15}$$

By integrating Equation (14). The first three terms in this equation represent the marginal entropy of X and the last term stands for the maximum entropy. Accordingly, the variation of information can be expressed simply as:

$$H(X/X^*) = H(X) - H_{\max} \tag{16}$$

If the a posteriori distribution of X is assumed to be lognormal, the informational entropy $H(X/X^*)$ becomes:

$$H(X/X^*) = \ln\sqrt{2\pi} + \ln\sigma_y + \mu_y + \frac{1}{2} - \ln(b-a) \tag{17}$$

with and μ_y and σ_y being the mean and standard deviation of $y = \ln x$.

If the a posteriori distribution of X is assumed to be 2-parameter gamma distribution with parameters α and β,

$$f_{(x)} = \frac{1}{\beta^{\alpha}\,\Gamma_{(\alpha)}}\, x^{\alpha-1}\, e^{\frac{-x}{\beta}}\; x \geq 0 \tag{18}$$

The informational entropy $H(X/X^*)$ becomes:

$$H(X/X^*) = \ln[\beta\cdot\Gamma(\alpha)] + \mu_x/\beta - (\alpha-1)\cdot\Phi(\alpha) - \ln(b-a) \tag{19}$$

where, μ_x is the mean of the series, α the shape parameter, and β the scale parameter.

In the above, entropy as the variation of information measures the amount of uncertainty reduced by making observations when the a posteriori distribution is estimated.

The maximum amount of information gained about the process X defined within the interval $[a, b]$ is $H_{\max}$. Thus, the expression in Equation (16) will assume negative values. However, since $H(X/X^*)$ describes entropy as the variation of information, it is possible to consider the absolute value of this measure.

When the a posteriori probability distribution function is correctly estimated, the information gained about the random variable will increase as the number of observations increases. Thus, when this number goes to infinity, the entropy $H(X/X^*)$ will approach zero. In practice, it is not possible to obtain an infinite number of observations; rather, the availability of sufficient data is important. By using the entropy measure $H(X/X^*)$, it possible to evaluate the fitness of a particular distribution function to the random variable and to assess whether the available data convey sufficient information about the process.

4. Mathematical Interpretation of the Revised Definition of Informational Entropy

4.1. The Distance between Two Continuous Distribution Functions as Defined by the Euclidian Metric

The approach used to obtain Equation (16) is essentially a means of measuring the distance between the points in probability space, described by the a priori $\{q(x)\}$ and a posteriori $\{p(x)\}$ distribution functions. The distance between these two functions can be determined by different measures like the metric concept, which enables one to see whether the two functions coincide.

According to the Euclidian metric, the distance between $p(x)$ and $q(x)$ functions defined in the same probability space (Ω, K) is:

$$I = \int [p(x) - q(x)]^2 dx \tag{20}$$

If $p(x)$ is the standard normal, and $q(x)$, the standard uniform distribution function, one obtains:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\left\{\sqrt{1 + \frac{3\sqrt{\pi}}{x}\left[1 - \frac{4}{3}F(x)\right]}\right\} \tag{21}$$

By integrating Equation (20) to obtain the difference function $\Phi(x)$. The $F(x)$ function in Equation (21) represents the cumulative probabilities for the standard normal distribution. When the

above difference function is equal to zero, $p(x)$ and $q(x)$, which are described as two points in the (Ω, K) probability space, will coincide at the same point. When the difference function assumes a minimum value, this will indicate a point of transition between $p(x)$ and $q(x)$, where the two functions can be expressed in terms of each other. The same point also refers to a minimum number of observations required to produce information about the process X. When the difference function is described as in Figure 1, the presence of such a minimum value can be observed. The difference function $\Phi(x)$ decreases until $x = x_0$, where it passes through a minimum value. At point x_0, the two functions $p(x)$ and $q(x)$ approach each other until the distance between them is approximated by a constant C. After this point, when x approaches infinity, the difference function gradually increases; and finally, the difference between $p(x)$ and $q(x)$ approaches zero at infinity. One may define x_0 as the point where the two probability functions can be used interchangeably with an optimum number of observations.

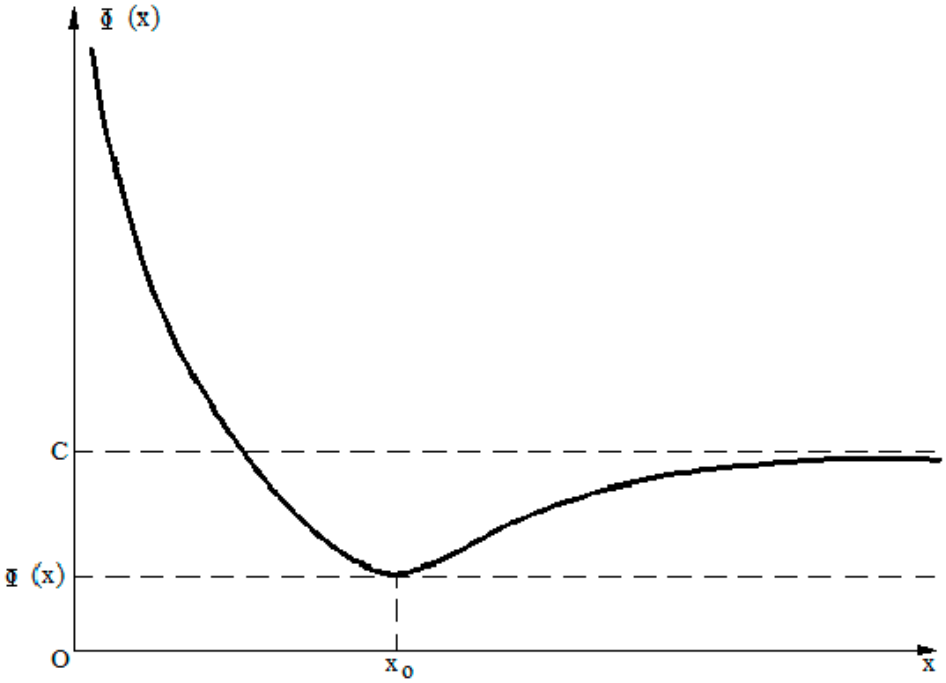

Figure 1. The difference function as defined by the Euclidian metric.

The purpose of observing the random variable X is to obtain a realistic estimate of its population parameters and to achieve reliable information about the process to allow for correct decisions in planning. On the other hand, each observation entails a cost factor; therefore, planners are interested in delineating how long the variable X has to be observed. Equation (16) is significant from this point of view. By defining the variation of information as the reduction of uncertainty via sampling, the point where no more increase or change in variation of information is obtained actually specifies the time point when sampling can be stopped. This is a significant issue which may be employed in considerations of gauging station discontinuance.

4.2. The Distance between Two Continuous Distribution Functions as Defined by Max-Norm

The max-norm can also be used to measure the distance between two functions defined in the probability space and to assess whether these two functions approach each other. According to the max-norm, the distance between two functions $p(x)$ and $q(x)$ is defined as:

$$\Delta(p, m) = \sup_{-\infty<x<+\infty} |p(x) - q(x)| \tag{22}$$

When $p(x)$ is used to represent the standardized normal and $q(x)$, the standardized uniform distribution functions, the difference function $\{h(x)\}$ will be:

$$h(x) = p(x) - q(x) \tag{23}$$

It may be observed in Figure 2 that, the critical points of the difference function are at h_0, h_1, and h_2 so that the difference between the two functions $\{\Delta(p,q)\}$ can be expressed as:

$$\Delta(p,q) = \max\{h_0, h_1, h_2\} \tag{24}$$

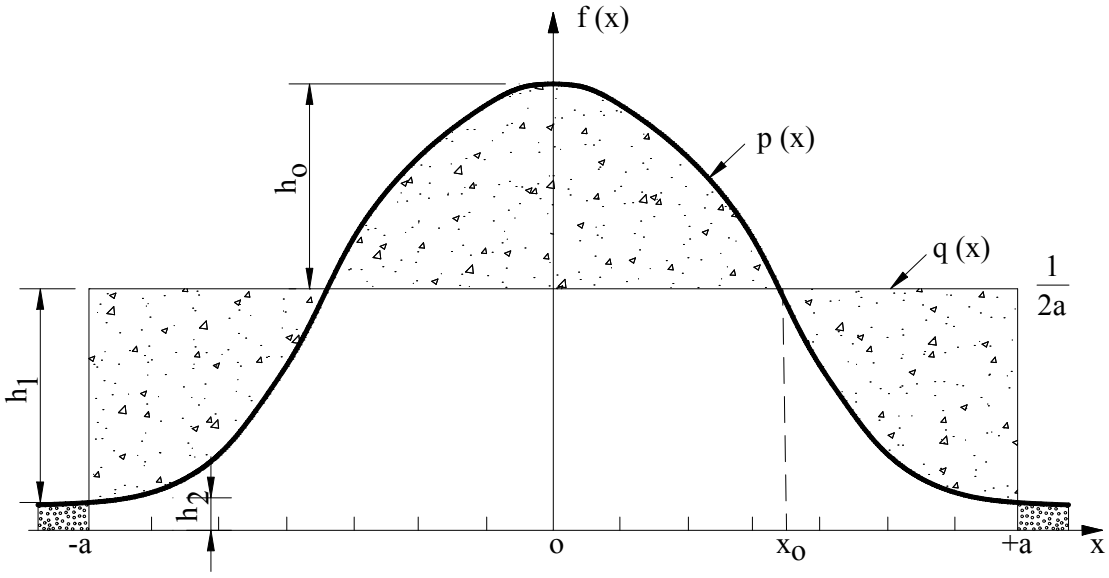

Figure 2. Critical values of the difference function as defined by the max-norm.

Based on the half-range value "a" in Figure 2, the critical points h_0, h_1, and h_2 can be obtained as:

$$h_0 = \frac{1}{\sqrt{2\pi}}\frac{1}{2a} \tag{25}$$

$$h_1 = \frac{1}{\sqrt{2\pi}}\frac{1}{2a}e^{(-a^2/2)} \tag{26}$$

$$h_2 = \frac{1}{\sqrt{2\pi}}e^{(-a^2/2)} \tag{27}$$

The problem here is then to find the distance between the two functions as the half-range value "a" which minimizes $\Delta(p,q)$ of Equation (24). The critical half-range value "a" that satisfies this supremum is:

$$a = \frac{3}{4}\sqrt{2\pi} \tag{28}$$

At the above critical half-range value "a", which is obtained by the max-norm, it is possible to use the two functions $p(x)$ and $q(x)$ interchangeably with an optimum number of observations.

When two points represented by the a posteriori and a priori distribution functions, $p(x)$ and $q(x)$, respectively, in the same probability space approach each other, this indicates, in information terms, an information increase about the random process analyzed. The case when the two points coincide represents total information availability about the process. Likewise, when $H(X/X^*)$ of Equation (16) approaches zero in absolute terms, this indicates a gain of total information about the process X defined

within the interval [a, b]. One obtains sufficient information about the process when the variation information, as described by the Euclidian metric, approaches a constant value.

4.3. Asymptotic Properties of Shannon's Entropy

Vapnik [29] analyzed and provided proofs for some asymptotic properties of Shannon's entropy of the set of events on the sample size N. He used these properties to prove the necessary and sufficient conditions of uniform convergence of the frequencies of events to their probabilities.

In the work of Vapnik [23], it is shown that the sequence:

$$\frac{H(S)}{N},\ N = 1,\ 2,\ \ldots, \tag{29}$$

has a limit c, when N goes to infinity. The lemma:

$$\lim_{N\to\infty} \frac{H(S)}{N} = c\ \ 0 \le c \le 1, \tag{30}$$

was proved by Vapnik [29] and was claimed to "repeat the proof of the analogous lemma in information theory for Shannon's entropy". Vapnik [29] also proved that, for any N, the sequence of Equation (29) is an upper bound for limit of Equation (30).

Vapnik [29] proved the above lemmas for Shannon's entropy, based on the discrete case of Equation (1). However, they are also valid for the continuous case as described by the Euclidian metric. Thus, it is possible to restate, using Vapnik's proofs, that the upper bound $H_{\max}$ of Shannon's entropy will be reached as the number of observations increases to approach the range of the population ($N\to\infty$) and that the variation of information of Equation (16) approaches a constant value "c".

In the next section, the derivation of the constant "c" is demonstrated for the case when the a priori distribution function is assumed to be uniform and the a posteriori function to be normal. These assumptions comply with the limits ($0 \le c \le 1$) defined for the discrete case as in Equation (30).

5. Further Development of the Revised Definition of the Variation of Information

If the observed range [a, b] of the variable X is considered also as the population value of the range, R, of the variable, the maximum information content of the variable may be described as:

$$H_{\max} = \ln R \tag{31}$$

With;

$$R = b - a\ a < x < b \tag{32}$$

When the a posteriori distribution of the variable is assumed to be normal, the marginal entropy of X becomes:

$$H(X) = \ln\sqrt{2\pi} + \ln\sigma + 1/2 \tag{33}$$

If the variable is actually normally distributed and if a sufficient number of observations are obtained, the entropy of Equation (16) will approach a value which can be considered to be within an acceptable region. This is the case where one may infer that sufficient information has been gained about the process.

When sufficient information is made available about X, it will be possible to make the best estimates for the mean (μ), variance (σ), and the range (R) of X. For this purpose, the variable has to be analyzed as an open series in the historic order. According to the approach used, the information gained about the process will continuously increase as the number of observations increase. Similarly, $H_{\max}$ and $H(X)$ will also increase, while $H(X/X^*)$ will decrease. When the critical point is reached, where the variable can be described by its population parameters, $H_{\max}$ will approach a constant value; $H(X)$ will also get closer to this value with $H(X/X^*)$ approaching a constant value of "c" as in Figure 3.

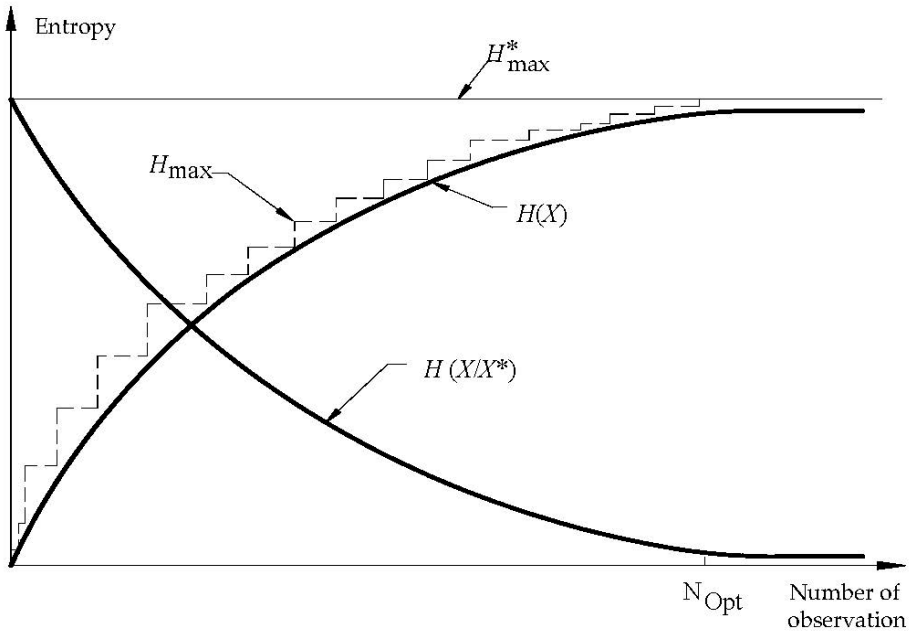

Figure 3. Maximum entropy $\{H_{max}\}$, marginal entropy $\{H(X)\}$ and entropy as the variation of information $\{H(X/X^*)\}$ versus the number of observations.

Determination on Confidence Limits for Entropy Defined by Variation of Information

The confidence limits (*acceptable region*) of entropy can be determined by using the a posteriori probability distribution functions. If the normal $\{N(0,1)\}$ probability density function is selected, the maximum entropy for the standard normal variable z is;

$$H_{max}(z) = \ln R_z, \tag{34}$$

with the range of z being,

$$R_z = 2a \tag{35}$$

Here, the value a describes the half-range of the variable. Then, the maximum entropy for variation x with N (μ,σ) is;

$$H_{max}(x) = \ln(R_z\sigma) \tag{36}$$

If the critical half-range value is foreseen as:

$$a = 4\sigma, \tag{37}$$

then the area under the normal curve may be approximated to be 1.

For the half-range value, replacing the appropriate values in Equation (16), one obtains the acceptable entropy value for the normal probability density function as:

$$H(X/X^*)_{cr} = 0.6605, \tag{38}$$

using natural logarithms. When the entropy $H(X/X^*)$ of the variable which is assumed to be normal remains below the above value, one may decide that the normal probability density function is acceptable and that a sufficient amount of information has been collected about the process.

If the a posteriori distribution function is selected as lognormal LN(μ_y, σ_y), the variation of information for the variable x can be determined as:

$$H(X/X^*) = \ln\left[2Sinh(a\sigma_y)\right] - \ln\sigma_y - 1.4189 \tag{39}$$

Here, since lognormal values will be positive, one may consider $0 \leq x \leq \infty$. Then the acceptable value of $H(X/X^*)$ for the lognormal distribution function will be;

$$H(X/X^*) = a\sigma_y - \ln \sigma_y - 1.4189 \tag{40}$$

According to Equation (40), no single constant value exists to describe the confidence limit for lognormal distribution. Even if the critical half-range is determined, the confidence limits will vary according to the variance of the variable. However, if the variance of x is known, the confidence limits can be computed.

6. Application

6.1. Application to Synthetic Series to Test the Fit of Probability-Distribution Functions

It is often difficult in practice to find long series of complete hydrological data. Thus, it is preferred here to test the above methodology on synthetically generated data for the purposes of evaluating the fit of different probability distribution functions. For this purpose, normal $\{N(\mu, \sigma)\}$ and lognormal $\{LN(\mu_y, \sigma_y)\}$ distributed time series are produced, using uniformly distributed series derived by the Monte Carlo method. Ten-year time series are obtained with normal $\{N(\mu, \sigma)\}$ and lognormal $\{LN(\mu_y, \sigma_y)\}$ distributions, respectively. Each series covered a period of (10×365) days with cumulative data for each year as $(i \times 365$; where $i = 1, ..., N)$.

To test the methodology, $N(8, 10)$ distributed 3650 synthetic data are divided into subgroups with 365 data in each. First, maximum informational entropy (H_{max}) is determined, using Equation (31) and the whole time series. Assuming that the a posteriori distribution is normal, marginal entropies ($H(X)$) and, finally, the informational entropy values ($H(X/X^*)$ are computed for the normal distribution using Equation (15). Consecutive values of these entropy measures are computed first for 365 generated data, next for 2×365 data, and for the last year 10×365 data. The confidence limits for the case of a posteriori normal distribution is determined by Equation (38). Figure 4 shows the results of this application. If a lognormal posteriori distribution is assumed for this series, which is actually normally distributed, this assumption is rejected on the basis of the computed confidence limits for normal distribution. Otherwise, the assumption is accepted. In Figure 4, the $H(X/X^*)$ values fall below the confidence level determined for normal distribution so that the assumption of a posteriori lognormal distribution is rejected.

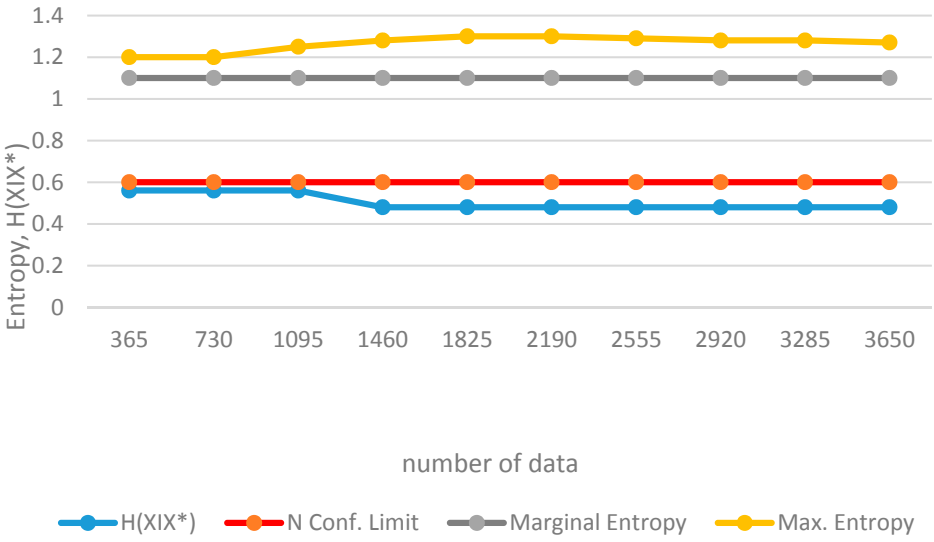

Figure 4. Normal distributed synthetic series, by the assumption of a posteriori normal distributed probability function (where; N Conf. Limit is the confidence limit for normal distribution).

If the same application is repeated by using the confidence limit for lognormal distribution, as in Figure 5, the assumption of a posterior lognormal distribution is rejected as the $H(X/X^*)$ values stay above the confidence level determined for lognormal distribution.

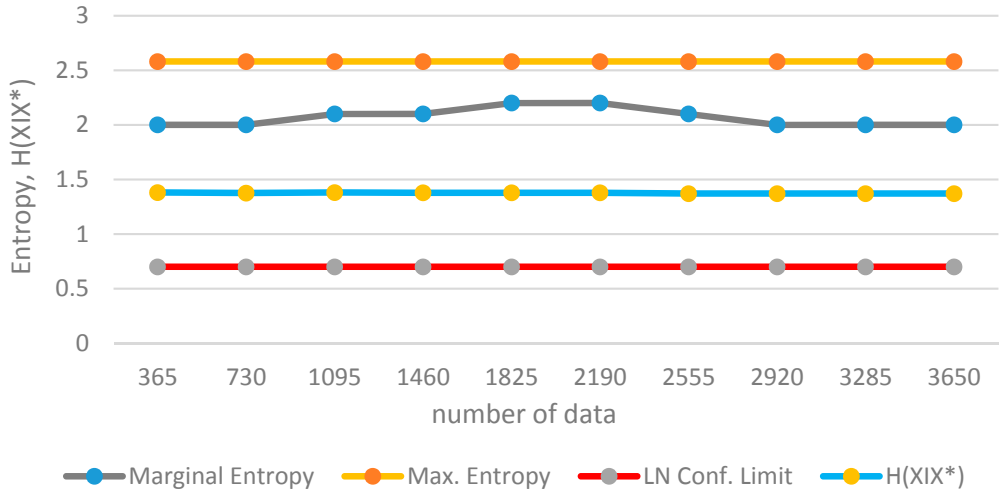

Figure 5. Normal distributed synthetic series, by the assumption of a posteriori lognormal distributed probability function (where; LN Conf. Limit is the limit of confidence for lognormal distribution).

Similar exercises may be run by generating lognormal distributed synthetic series and assuming the posteriori distribution first as lognormal (Figure 6) and then as normal distribution (Figure 7).

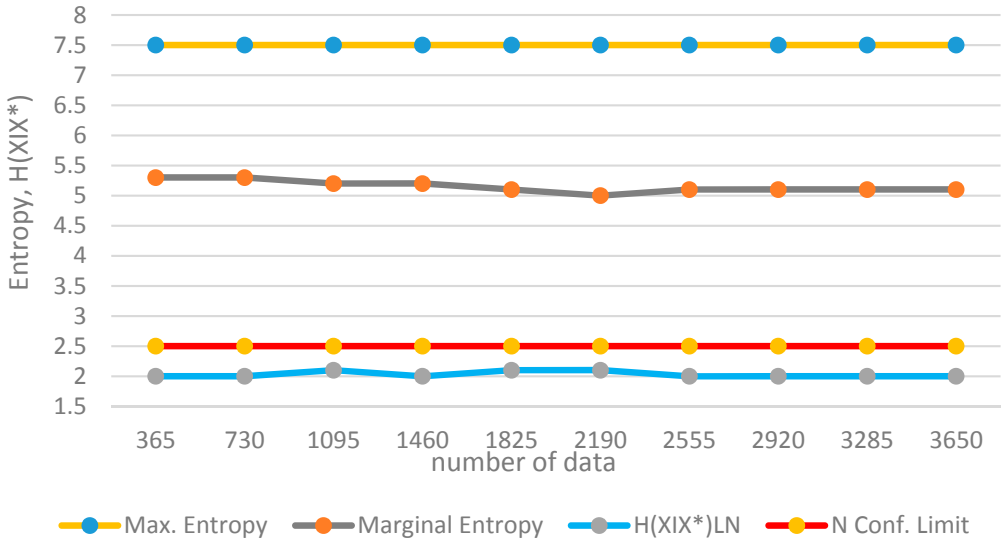

Figure 6. Lognormal distributed synthetic series, by the assumption of a posteriori lognormal distributed probability function (where; N Conf. Limit is the confidence limit for normal distribution).

The above exercises show that comparisons between assumptions of a posteriori normal and lognormal distributions on the basis of entropy-based confidence limits for each distribution give valid results by checking how the variation of information values behave with respect to the confidence limits.

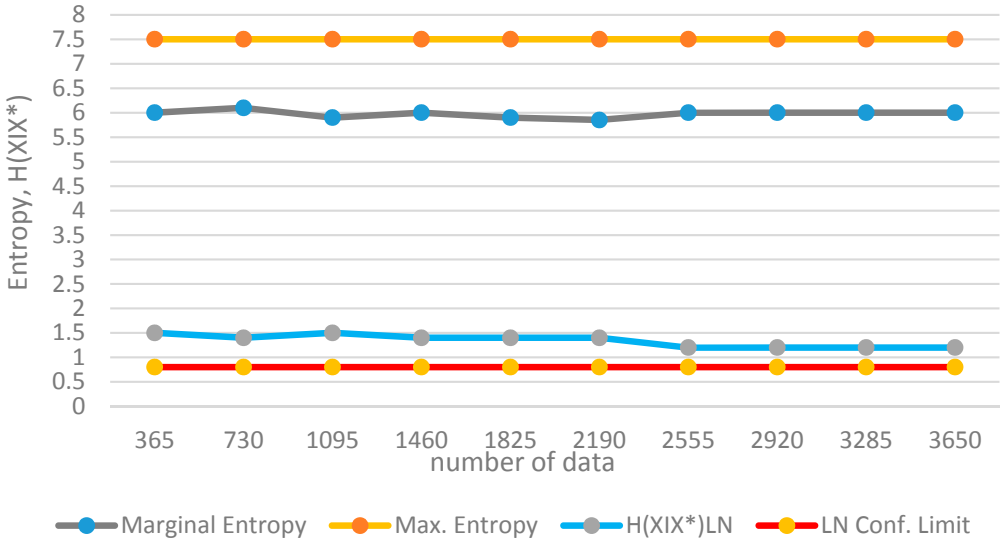

Figure 7. Lognormal distributed synthetic series, by the assumption of a posteriori normal distributed probability function (where; LN Conf. Limit is the limit of confidence for lognormal distribution).

6.2. Application to Runoff Data for Assessment of Sampling Duration

An important question regarding hydrometric data-monitoring networks is how long the observations should be continued. Considering the "data rich, information poor" data networks of our times, researchers and decisionmakers have wondered whether monitoring could be discontinued at certain sites, as data observation is a cost and labor-consuming activity [30,31]. To date, none of the approaches proposed for the problem of station discontinuance have found universal acceptance. Entropy measures as described in this work may as well be employed when a monitoring activity reaches an optimal point in time after which any new data does not produce new information. This feature of entropy measures is shown in Figure 3, where the marginal entropy of the process $H(X)$ approaches the total uncertainty Hmax as the number of observations (N) increase. Finally, a point is reached where $H_{\max}$ and $H(X)$ coincide after a certain number of observations, which can be defined as N_{opt}. After this point on, observed data do not produce new information, and thus monitoring can be discontinued. Certainly, the probability distribution of best fit to observed series must be selected first to evaluate this condition. This is an important feature of entropy measures as they can be used to infer about station discontinuance, based also on the selection of the appropriate distribution functions.

To test the above aspect of the entropy concept, observed runoff data at two monitoring stations (Kuskayasi and Akcil) in the Ceyhan river basin in Turkey are employed (Figure 8). The Ceyhan basin has been subject to several investigations and projects for the development of water schemes; thus, it is intended here to evaluate the monitoring activities in the basin in terms of entropy measures. Although there are other gauging stations along the river, their data are not homogeneous due to already-built hydraulic structures. Kuskayasi and Akcil are the two stations where natural flows are observed, although their common observation periods cover only 8 years.

The observations at Kuskayasi were discontinued after 1980 and Akcil after 1989. Thus, for the purposes of this application their common period between 1973 and 1980 is selected. Daily data for the observation period of 8 years are used, where the mean daily runoff at Kuskayasi is 10.8 m^3/s and that at Akcil is 27.18 m^3/s. The standard deviations are 11.77 m^3/s and 22.48 m^3/s, respectively.

Next, the fits of normal and lognormal distributions are tested at both stations again with the entropy concept. This analysis is followed by the computation of marginal entropies ($H(X)$, $H_{\max}$ and the variation of information $H(X)/H^*$) for these two distribution functions. The computations are carried out in a successive manner, using the first year's 365 data, the second year's 720, and so on until the total number of 2920 data are reached. Certainly, $H_{\max}$ changes with the total of data

observed from the beginning of the observation period, assuming a ladder-like increase as in Figure 3, where H^* is used to represent H_{max} for the total observation period of 2920 daily data.

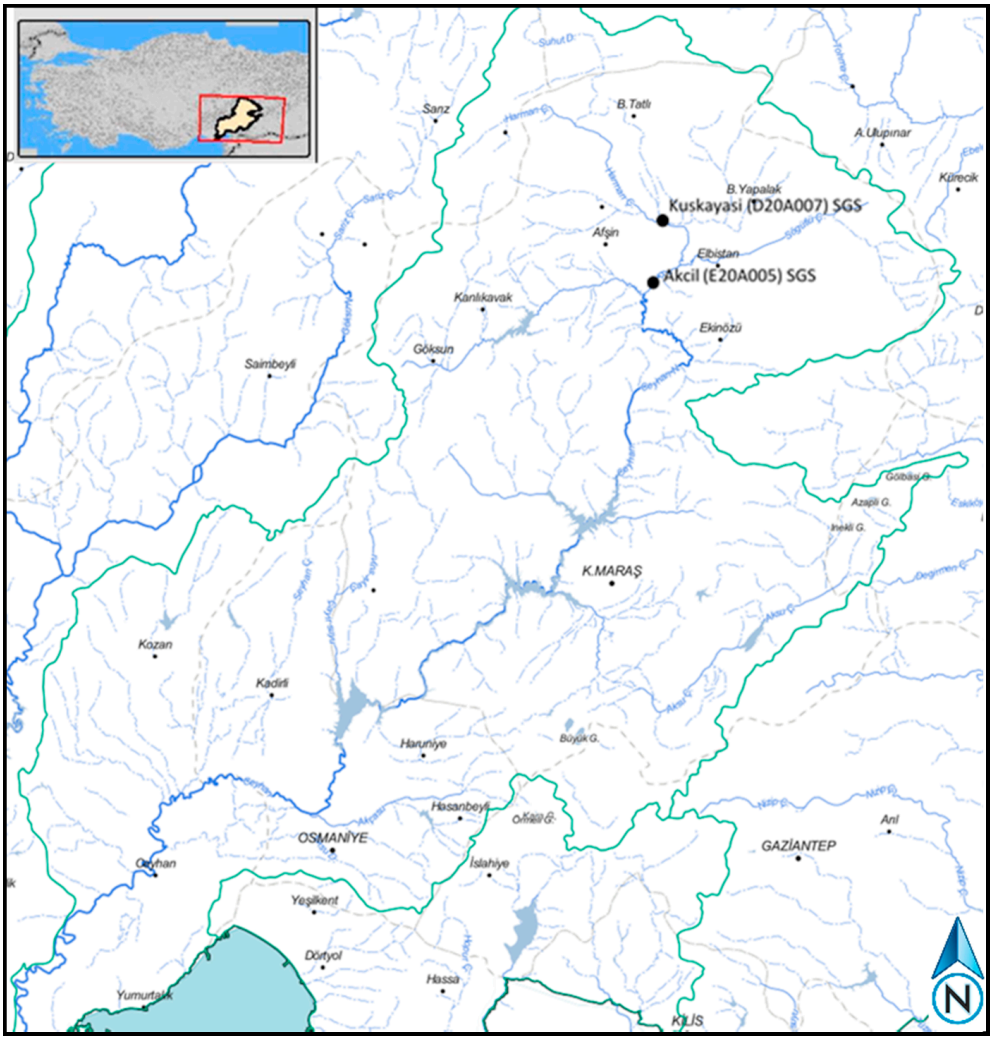

Figure 8. Ceyhan river basin in the south of Turkey and selected monitoring sites (Kuskayasi and Akcil).

Figures 9 and 10 show figures similar to Figure 3 under the assumption of normal and lognormal distributions fit to daily data for 8 years. Although both distributions seem to be sufficient, normal distribution shows more distinctively how $H(X)$ approaches the total entropy H^*. It may seem unusual for an upstream station with daily observations to reflect a normal distribution; yet this is physically due to karstic contributions to runoff, which stabilize the flows.

Whether the normal or lognormal distributions are selected, it can be observed in Figures 9 and 10 that 2920 observations are not sufficient to reach H^*. Although $H(X)$ approaches H^*, the optimal number of observations is not yet reached with only 8 years of observations.

Results for the downstream Akcil station are shown in Figures 11 and 12. Here, again, normal distribution appears to give a better fit to observed data. As can be observed especially in Figure 11, $H(X)$ closely approaches H^* for 8 years of data. If observations could be continued after 8 years of 2920 data, most probably the optimum number of observations would be reached.

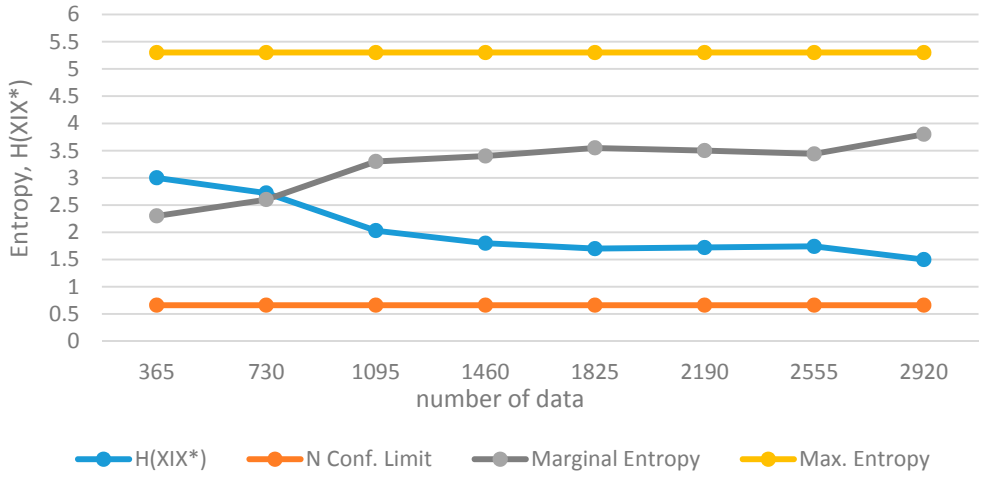

Figure 9. Kuskayasi (1973–1980), by the assumption of a posteriori normal distribution function (where; N Conf. Limit is the limit for normal distribution).

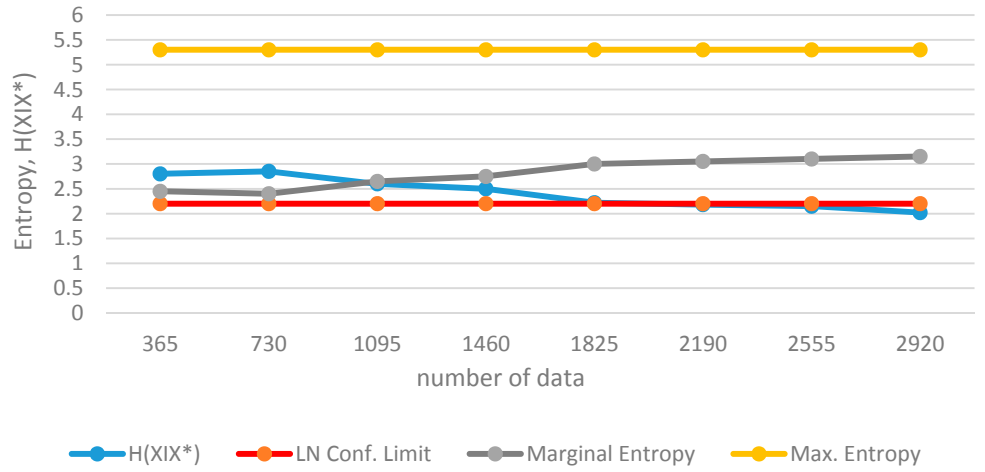

Figure 10. Kuskayasi (1973–1980), by the assumption of a posteriori lognormal distribution function (where; LN Conf. Limit is the limit of confidence for lognormal distribution).

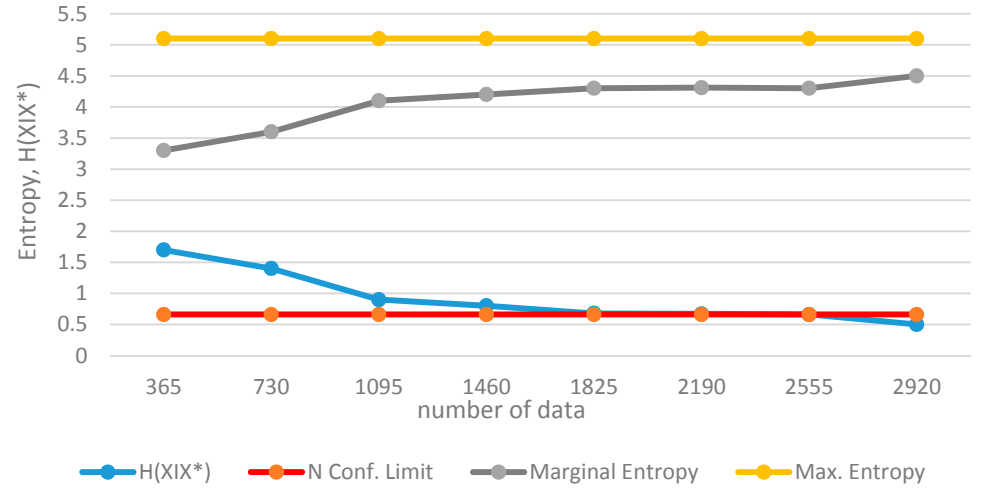

Figure 11. Akcil (1973–1980), by the assumption of a posteriori normal distribution function (where; N Conf. Limit is the limit of confidence for normal distribution).

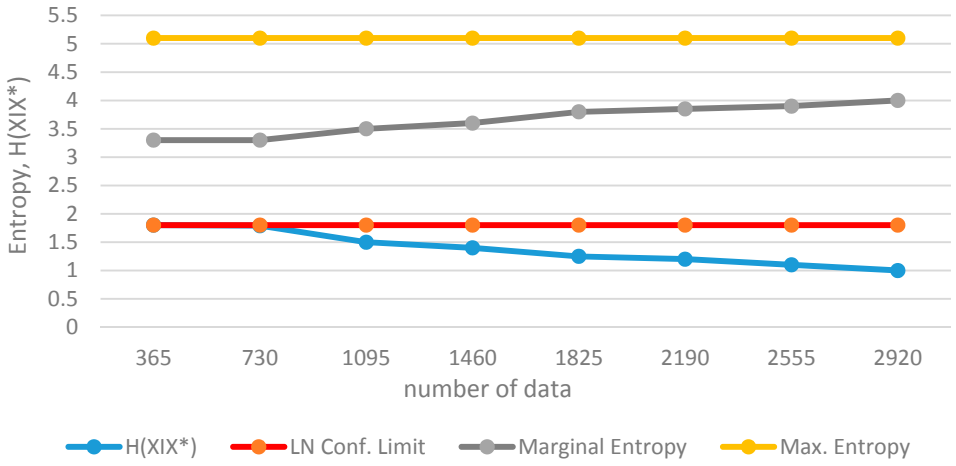

Figure 12. Akcil (1973–1980), by the assumption of a posteriori lognormal distribution function (where; LN Conf. Limit is the limit of confidence for lognormal distribution).

It is concluded on the basis of results obtained through the above application that, if sufficiently long observed time series are available, the entropy principle can be effectively used to infer on an important feature of hydrometric monitoring, i.e., sampling duration or station discontinuance.

7. Conclusions

The extension to the revised definition of informational entropy developed in this paper resolves further major mathematical difficulties associated with the assessment of uncertainty, and indirectly of information, contained in random variables. The description of informational entropy, not as an absolute measure of information but as a measure of the variation of information, has the following advantages:

- It eliminates the controversy associated with the mathematical definition of entropy for continuous probability distribution functions. This makes it possible to obtain a single value for the variation of information instead of several entropy values that vary with the selection of the discretizing interval when, in the former definitions of entropy for continuous distribution functions, discrete probabilities of hydrological events are estimated through relative class frequencies and discretizing intervals.
- The extension to the revised definition introduces confidence limits for the entropy function, which facilitates a comparison between the uncertainties of various hydrological processes with different scales of magnitude and different probability structures.
- Following from the above two advantages, it is further possible through the use of the concept of the variation of information to:

    - determine the contribution of each observation to information conveyed by data;
    - determine the probability distribution function which best fits the variable;
    - make decisions on station discontinuance.

The present work focuses basically on the theoretical background for the extended definition of informational entropy. The methodology is then tested via applications to synthetically generated data and observed runoff data and is shown to give valid results. For real-case observed data, long duration series with sufficient length and quality are needed. Currently, studies are being continued by the authors on long series of runoff, precipitation and temperature data.

It follows from the above discussions that the use of the concept of variation of information and of confidence limits makes it possible to:

- *determine the contribution of each observation to information conveyed by data*;
- *calculate the cost factors per information gained*;
- *determine the probability distribution function which best fits the variable*;
- *select the model which best describes the behavior of a random process*;
- *compare the uncertainties of variables with different probability density functions*;
- *make decisions on station discontinuance*.

The above points are different problems to be solved by the concept of entropy, and further extensions of the methodology are required to address each of them.

Acknowledgments: We gratefully acknowledge the support received from the authors' EU Horizon2020 Project entitled FATIMA (FArming Tools for external nutrient Inputs and water Management, Grant No. 633945) for providing the required funds to cover the costs towards publishing in open access.

Author Contributions: Turkay Baran and Nilgun B. Harmancioglu conceived and designed the experiments; Turkay Baran and Filiz Barbaros performed the experiments; Cem P. Cetinkaya and Filiz Barbaros analyzed the data; Turkay Baran, Nilgun B. Harmancioglu and Cem P. Cetinkaya contributed reagents/materials/analysis tools; Nilgun B. Harmancioglu wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Singh, V.P.; Fiorentino, M. (Eds.) A Historical Perspective of Entropy Applications in Water Resources. In *Entropy and Energy Dissipation in Water Resources*; Water Science and Technology Library; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1992; Volume 9, pp. 155–173.
2. Fiorentino, M.; Claps, P.; Singh, V.P. An Entropy-Based Morphological Analysis of River Basin Networks. *Water Resour. Res.* **1993**, *29*, 1215–1224. [CrossRef]
3. Wehrl, A. General Properties of Entropy. *Rev. Mod. Phys.* **1978**, *50*, 221–260. [CrossRef]
4. Templeman, A.B. Entropy and Civil Engineering Optimization. In *Optimization and Artificial Intelligence in Civil and Structural Engineering*; NATO ASI Series (Series E: Applied Sciences); Topping, B.H.V., Ed.; NATO: Washington, DC, USA, 1989; Volume 221, pp. 87–105, ISBN 978-94-017-2490-6.
5. Schrader, R. On a Quantum Version of Shannon's Conditional Entropy. *Fortschr. Phys.* **2000**, *48*, 747–762. [CrossRef]
6. Shannon, C.E. A Mathematical Theory of Information. In *The Mathematical Theory of Information*; The University of Illinois Press: Urbana, IL, USA, 1948; Volume 27, pp. 170–180.
7. Harmancioglu, N.; Alpaslan, N. Water Quality Monitoring Network Design: A Problem of Multi-Objective Decision Making. *JAWRA* **1992**, *28*, 179–192. [CrossRef]
8. Harmancioglu, N.; Singh, V.P.; Alpaslan, N. Versatile Uses of The Entropy Concept in Water Resources. In *Entropy and Energy Dissipation in Water Resources*; Singh, V.P., Fiorentino, M., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1992; Volume 9, pp. 91–117, ISBN 978-94-011-2430-0.
9. Harmancioglu, N.; Alpaslan, N.; Singh, V.P. Assessment of Entropy Principle as Applied to Water Quality Monitoring Network Design. In *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*; Water Science and Technology Library; Hipel, K.W., McLeod, A.I., Panu, U.S., Singh, V.P., Eds.; Springer: Dordrecht, The Netherlands, 1994; Volume 3, pp. 135–148, ISBN 978-94-017-3083-9.
10. Harmancioglu, N.B.; Yevjevich, V.; Obeysekera, J.T.B. Measures of information transfer between variables. In Proceedings of the Fourth International Hydrology Symposium—Multivariate Analysis of Hydrologic Processes, Fort Collins, CO, USA, 15–17 July 1985; Colorado State University: Fort Collins, CO, USA, 1986; pp. 481–499.
11. Harmancioglu, N.B.; Yevjevich, V. Transfer of hydrologic information among river points. *J. Hydrol.* **1987**, *91*, 103–118. [CrossRef]
12. Harmancioglu, N.; Cetinkaya, C.P.; Geerders, P. Transfer of Information among Water Quality Monitoring Sites: Assessment by an Optimization Method. In Proceedings of the EnviroInfo Conference 2004, 18th International Conference Informatics for Environmental Protection, Geneva, Switzerland, 21–23 October 2004; pp. 40–51.

13. Baran, T.; Bacanli, Ü.G. An Entropy Approach for Diagnostic Checking in Time Series Analysis. *SA Water* **2007**, *33*, 487–496.
14. Singh, V.P. The Use of Entropy in Hydrology and Water Resources. *Hydrol. Process.* **1997**, *11*, 587–626. [CrossRef]
15. Singh, V.P. The Entropy Theory as a Decision Making Tool in Environmental and Water Resources. In *Entropy Measures, Maximum Entropy Principle and Emerging Applications. Studies in Fuzziness and Soft Computing*; Karmeshu, Ed.; Springer: Berlin, Germany, 2003; Volume 119, pp. 261–297, ISBN 978-3-540-36212-8.
16. Harmancioglu, N.; Singh, V.P. Entropy in Environmental and Water Resources. In *Encyclopedia of Hydrology and Water Resources*; Herschy, R.W., Fairbridge, R.W., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1998; Volume 5, pp. 225–241, ISBN 978-1-4020-4497-7.
17. Harmancioglu, N.; Singh, V.P. Data Accuracy and Data Validation. In *Encyclopedia of Life Support Systems (EOLSS)*; Knowledge for Sustainable Development, Theme 11 on Environmental and Ecological Sciences and Resources, Chapter 11.5 on Environmental Systems; Sydow, A., Ed.; UNESCO Publishing-Eolss Publishers: Oxford, UK, 2002; Volume 2, pp. 781–798, ISBN 0 9542989-0-X.
18. Harmancioglu, N.B.; Ozkul, S.D. Entropy-based Design Considerations for Water Quality Monitoring Networks. In *Technologies for Environmental Monitoring and Information Production*; Nato Science Series (Series IV: Earth and Environmental Sciences); Harmancioglu, N.B., Ozkul, S.D., Fistikoglu, O., Geerders, P., Eds.; Springer: Dordrecht, The Netherlands, 2003; Volume 23, pp. 119–138, ISBN 978-94-010-0231-8.
19. Ozkul, S.; Harmancioglu, N.B.; Singh, V.P. Entropy-Based Assessment of Water Quality Monitoring Networks. *J. Hydrol. Eng.* **2000**, *5*, 90–100. [CrossRef]
20. Jaynes, E.T. *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*; Rosenkrantz, R.D., Ed.; Springer: Dordrecht, The Netherlands, 1983; ISBN 978-94-009-6581-2.
21. Guiasu, S. *Information Theory with Applications*; Mc Graw-Hill: New York, NY, USA, 1977; 439p, ISBN 978-0070251090.
22. Harmancioglu, N.B. Measuring the Information Content of Hydrological Processes by the Entropy Concept. *J. Civ. Eng. Fac. Ege Univ.* **1981**, 13–88.
23. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [CrossRef]
24. Harmancioglu, N. Entropy concept as used in determination of optimal sampling intervals. In Proceedings of the Hydrosoft 1984, International Conference on Hydraulic Engineering Software. Interaction of Computational and Experimental Methods, Portorož, Yugoslavia, 10–14 September 1984; Brebbia, C.A., Maksimovic, C., Radojkovic, M., Eds.; Editions du Tricorne: Geneva, Switzerland; pp. 99–110.
25. Harmancioglu, N.B. An Entropy-based approach to station discontinuance. In *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*; Time Series Analysis and Forecasting; Hipel, K.W., McLeod, I., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1994; Volume 3, pp. 163–176.
26. Harmancioglu, N.B.; Alpaslan, N. *Basic Approaches to Design of Water Quality Monitoring Networks*; Water Science and Technology; Elsevier: Amsterdam, The Netherlands, 1994; Volume 30, pp. 49–56.
27. Harmancioglu, N.B.; Cetinkaya, C.P.; Barbaros, F. Environmental Data, Information and Indicators for Natural Resources Management. In *Practical Environmental Statistics and Data Analysis*; Rong, Y., Ed.; ILM Publications: Buchanan, NY, USA, 2011; Chapter 1, pp. 1–66.
28. Schultze, E. Einführung in die Mathematischen Grundlagen der Informationstheorie. In *Lecture Notes in Operations Research and Mathematical Economics*; Springer: Berlin, Germany, 1969; 116p, ISBN 978-3-642-86515-2.
29. Vapnik, V.N. *Statistical Learning Theory*; Wiley Interscience: New York, NY, USA, 1998; 736p, ISBN 978-0-47-03003-4.
30. Harmancioglu, N.B.; Singh, V.P.; Alpaslan, N. *Environmental Data Management*; Water Science and Technology Library; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1998; 298p, ISBN 0792348575.
31. Harmancioglu, N.B.; Fistikoglu, O.; Ozkul, S.D.; Singh, V.P.; Alpaslan, N. *Water Quality Monitoring Network Design*; Water Science and Technology Library; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1999; 290p, ISBN 978-94-015-9155-3.