The entropy theory as a tool for modelling and decision-making in environmental and water resources

Vijay P Singh
Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA 70808-6405, USA

Abstract

Since the development of the entropy theory in the late 1940s and of the principle of maximum entropy (POME) in the late 1950s, there has been a proliferation of applications of the entropy theory in a wide spectrum of areas, including environmental and water resources. The real impetus to entropy-based modelling in water resources was provided in 1970s. A great variety of entropy-based applications in environmental and water resources have since been reported, and new applications continue to unfold. Most of these applications have, however, been in the realm of modelling and a relatively few applications have been reported on decision-making. This paper revisits the entropy theory and discusses its usefulness in environmental and water resources, and is concluded with comments on its implications in developing countries.

Introduction

Environmental and water resource systems are inherently spatial and complex, and our understanding of these systems is less than complete. Many of the systems are either fully stochastic, or part-stochastic and part-deterministic. Their stochastic nature can be attributed to randomness in one or more of the following components that constitute them: system structure (geometry); system dynamics; forcing functions (sources and sinks); and initial and boundary conditions. As a result, a stochastic description of these systems is needed, and the entropy theory enables development of such a description.

Engineering decisions concerning environmental and water resource systems are frequently made with less than adequate information. Such decisions may often be based on experience, professional judgment, thumb rules, crude analyses, safety factors, or probabilistic methods. Usually, decision-making under uncertainty tends to be relatively conservative. Quite often, sufficient data are not available to describe the random behavior of such systems. Although probabilistic methods allow for a more explicit and quantitative accounting of uncertainty, their major difficulty occurs due to the availability of limited or incomplete data. Small sample sizes and limited information render estimation of probability distributions of system variables with conventional methods quite difficult. This problem can be alleviated by use of the entropy theory which enables determination of the least-biased probability distributions with limited knowledge and data. Where the shortage of data is widely rampant as is normally the case in developing countries, the entropy theory is particularly appealing. The objective of this paper is to revisit the entropy theory and underscore its usefulness for both modelling and decision-making in environmental and water resources.

Entropy theory

The entropy theory is comprised of three main parts: Shannon entropy, principle of maximum entropy, and principle of minimum cross entropy. Before discussing these parts, it will be instructive to briefly discuss the meaning of entropy.

Meaning of entropy

The zeroth law of thermodynamics is related to the concept of temperature T, the first law of thermodynamics is related to the concept of internal energy U, and the second law of thermodynamics is related to the thermodynamic variable, called entropy, S, which is defined for a system as:

$$dS = \frac{dQ}{T}, \oint dS = 0 \quad (1)$$

where:
- dS is the change in entropy
- dQ is the change in heat
- T is the temperature.

$\oint$ indicates that the integral is evaluated for a complete traversal of the system response cycle. In Eq. (1), temperature is a state variable.

Heat is disordered energy. Energy can exist without disorder. The general principle is that energy becomes heat as soon as it is disordered. Conversely, disorder can exist without energy, and disorder becomes heat as soon as it is energised. Thus, to specify heat two numbers are needed: one to measure the quantity of heat, and the other to measure the quantity of disorder. The quantity of heat energy is measured in terms of calories and the quantity of disorder is measured in terms of entropy.

If there is a connection between disorder and entropy, then disorder, like entropy, must increase in natural processes. This is indeed the case, i.e. there is a tendency for a natural process to proceed toward a state of greater disorder. To illustrate, consider the confluence of two rivers having different sediment concentrations C_1 and C_2. Downstream of the confluence, the downstream reach attains an intermediate concentration C. The river system has been more disordered in this natural process because we have lost our ability to classify sediment concentration. The statement that discharge in the river corresponds to concentration C is weaker than the statement that discharge in River A corresponds to sediment concentration C_1 and discharge in River B corresponds to sediment

☎(001)- (225)-388 6697; fax (001)- (225)-388 8652;
e-mail: cesing@unix1.sncc.lsu.edu
Received 9 February 1999; accepted in revised form 6 August 1999.

concentration C_2 .

In statistical mechanics, disorder has a precise meaning and its connection with entropy is expressed as:

$$S = k \ln w = -k \ln p \qquad (2)$$

where:

k is the Boltzmann constant

S is the entropy of the system

w, called the disorder parameter, is the probability that the system will exist in the state it is in, relative to the possible states it could be in.

Eq. (2) defines the Boltzmann entropy and connects a thermodynamic or macroscopic quantity, the entropy, with a statistical or microscopic quantity, the probability. This also helps place the second law of thermodynamics on a statistical basis. Thus, entropy has been employed in thermodynamics as a measure of the degree of ignorance about the true state of a system. Algebraically, it is proportional to the logarithm of the probability of the state the system in. In a hydraulic system, if there were no energy loss the system would be orderly and organised. It is the energy loss and its causes that make the system disorderly and chaotic. Thus, entropy can be interpreted as a measure of the amount of chaos within a system. In hydraulics, entropy is a measure of the amount of irrecoverable flow energy which is expended by the hydraulic system to overcome friction. The system converts a portion of its mechanical energy to heat energy which then is dissipated to the external environment. Thus, the process equation in hydraulics expressing energy (or head) loss originates indeed in the entropy concept.

The direction of disorder in which natural processes occur (toward higher entropy) is determined by the laws of probability (toward a more probable state). The equilibrium state is the state of maximum entropy thermodynamically and the most probable state statistically. It must, however, be noted that fluctuations may occur about an equilibrium distribution (for example, the Brownian motion). This means that entropy increase in every spontaneous process is not a certainty. Indeed in certain processes it may decrease. Over a sufficiently long time horizon, however, even the most improbable states might occur. For example, the water in a pond suddenly freezes in spring or extreme winds, such as tornadoes, strike an area. Although such occurrences are possible, the probability of their happening, when computed, is incredibly small. Thus, the second law of thermodynamics shows the most probable course of events, not the only possible ones.

Entropy is an extensive property like mass, energy, volume, momentum, charge, or number of atoms of chemical species, but unlike these quantities, it does not obey a conservation law. Since entropy of a system is an extensive property, the total entropy of the system equals the sum of entropies of individual parts:

$$S = \sum_{i=1}^{m} S_i = \sum_{i=1}^{m} k \ln p_i + cons \qquad (3)$$

where:

S_i is the entropy of the i th subsystem,

p_i is the probability of being in the i th state,

cons is a constant, and

k is constant.

The most probable distribution of energy in a system is the one that corresponds to the maximum entropy of the system:

$$S = \sum_{i=1}^{m} k \ln p_i = \max imum \qquad (4)$$

This occurs under the condition of dynamic equilibrium. During evolution toward a stationary state, the rate of entropy production per unit mass should be minimum, compatible with external constraints. This is the Prigogin principle:

$$\frac{dS}{dt} = a \min imum \qquad (5)$$

In thermodynamics, entropy is decomposed into two parts: entropy exchanged between the system and its surroundings; and entropy produced in the system itself:

$$dS = dS_e + dS_i \qquad (6)$$

According to the second law of thermodynamics, the entropy of a closed and isolated system always tends to increase.

Shannon entropy

Eq. (2) can be extended to quantify disorder of a system by including all probable states. This led Shannon (1948) to develop the entropy theory for expression of information or uncertainty. To understand the informational aspect of entropy we consider a set consisting of n events. We view uncertainty as a situation where we do not know which event among n events will occur. Thus, uncertainty is about which one of those events actually occurs. Based on one's knowledge about the events, the uncertainty can be more or less. For example, the total number of events is a piece of information and the number of those events with non-zero probability is another piece of information. The probability distribution of the events, if known, provides a certain amount of information. Shannon (1948) defined a quantitative measure of uncertainty associated with a probability distribution or the information content of the distribution in terms of entropy, called Shannon entropy or informational entropy. The uncertainty can be quantified with entropy taking into account all different kinds of available information. Thus, entropy is a measure of the amount of uncertainty represented by the probability distribution and is a measure of the amount of chaos or of the lack of information about a system. If complete information is available, entropy = 0. Otherwise, it is greater than zero. The Shannon entropy is the weighted Boltzmann entropy.

Consider, for example, a drainage basin schematised by a channel network in which channels are idealised as single lines and links are network segments, as shown in Figure 1. A new link is formed by the junction of no more than two links and sources are the points further upstream in the network. The magnitude of a link is the number of sources upstream draining into that link. The magnitude n of the channel network is that of the outlet link and equals the number of first order streams. Let each link be associated with an average elevation and can have one of the m average elevations, e_i , i = 1, 2, 3, ..., m. Let p_i be the probability of a link having elevation e_i . The Shannon entropy S of the drainage basin system can be expressed as:

$$S = -\sum_{i=1}^{m} p_i \ln p_i \qquad (7)$$

For a continuous variable, the Shannon entropy can be written as

$$S = -\int_0^{\infty} f(x) \ln f(x) dx \qquad (8)$$

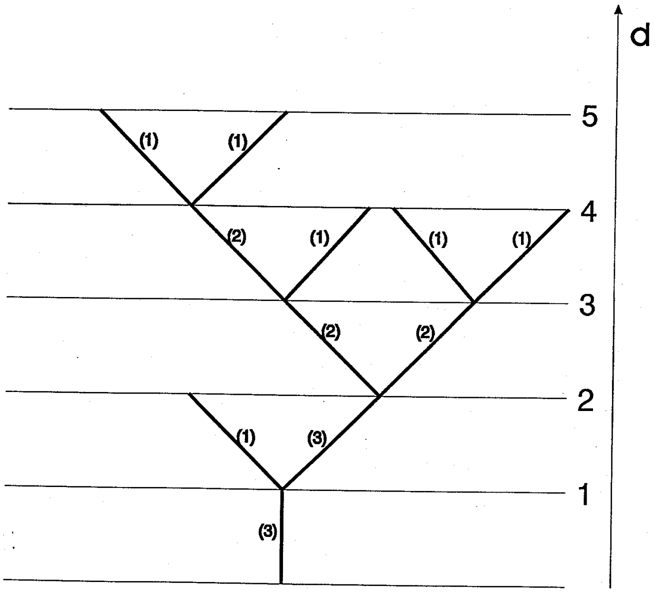

Figure 1

Definition of a simplified drainage network formed by links, external nodes (sources), and internal nodes. The Horton order of each link is reported in parentheses. In this figure the network magnitude n (the number of sources) is 6, the topological diameter D (maximum topological distance d from the outlet) is 5, and the Horton order is 3.

Principle of maximum entropy

Jaynes (1957) formulated the principle of maximum entropy (POME) a full account of which is presented in a treatise by Levine and Tribus (1978) and Tribus (1968). According to POME, when making inferences based on incomplete information, the probability distribution to be drawn must have the maximum entropy permitted by the available information expressed in the form of constraints. According to the Shannon entropy as an information measure, the POME-based distribution is favored over those with less entropy among those which satisfy the given constraints. Thus, entropy defines a kind of measure on the space of probability distributions. Intuitively, distributions of higher entropy represent more disorder, are smoother, are more probable, are less predictable, or assume less. The POME-based distribution is maximally noncommittal with regard to missing information and does not require invocation of ergodic hypotheses.

Mathematically, we maximise the Shannon entropy:

$$\text{Maximise} \quad S = -\int_0^\infty f(x) \ln f(x)\, dx \tag{9}$$

subject to the available information (or knowledge). This knowledge is expressed in terms of constraints, such as:

$$\int_0^\infty g_i(x)\, f(x) dx = E[g_i(x)], i = 0,1,2,..... \tag{10}$$

Solution of Eq. (9), subject to Eq. (10), can be obtained by using the method of Lagrange multipliers. This yields

$$f(x) = \exp[-\lambda_0 - \sum_{i=1}^{m} \lambda_i g_i(x)] \tag{11}$$

where:

λ_i, i = 0, 1, 2, ..., m, are Lagrange multipliers, which can be expressed in terms of constraints given by Eq. (10).

Principle of minimum cross entropy

According to Laplace's principle of insufficient reason, all outcomes of an experiment should be considered equally likely unless there is information to the contrary. For example, taking the previous example of a drainage basin, the probability of a link having an elevation is 1/m, if there is no information available on elevations of links.

Suppose we guess a probability distribution for a random variable X as Q = {q_1, q_2, q_3, ..., q_n }based on intuition or theory. This constitutes the prior information in terms of a prior distribution. To verify our guess, we take a set of observations X = {x_1, x_2, x_3, ..., x_n} and compute moments based on these observations. To derive the distribution P = {p_1, p_2, p_3,..............., p_n} of X, we take all the given information and make the distribution as near to our intuition and experience as possible. Thus, the principle of minimum cross entropy (POMCE) is expressed, when the cross entropy, D(P,Q), is minimised, as:

$$D(P,Q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \tag{12}$$

On the basis of intuition, experience or theory, a random variable may have an *a priori* probability distribution. Then, the Shannon entropy is maximum when the probability distribution of the random variable is that one which is as close to the *a priori* distribution as possible. This is referred to as the principle of minimum cross entropy which minimises the Bayesian entropy (Kullback and Leibler, 1951). This is equivalent to maximising the Shannon entropy. Here minimising D(P,Q) is equivalent to maximising the Shannon entropy.

Entropy theory as a decision tool in environmental and water resources

Although the entropy theory has been applied in recent years to a great variety of problems in environmental and water resources, its potential as a decision-making tool has not been fully exploited. What follows is a discussion highlighting this potential. Fundamental to the concepts presented below is the need of probability distributions which can be derived using the entropy theory.

Information of data

One frequently encounters a situation in which to exercise freedom of choice, evaluate uncertainty or measure information gain or loss. The freedom of choice, uncertainty, disorder, information content, or information gain or loss has been variously measured by relative entropy, redundancy, and conditional and joint entropies employing conditional and joint probabilities. As an example, in the analysis of empirical data, the variance has often been interpreted as a measure of uncertainty and as revealing gain or loss in information. However, entropy is another measure of dispersion - an alternative to variance. This suggests that it is possible to determine the variance whenever it is possible to determine the entropy measures, but the reverse is not necessarily true. However, variance is not the appropriate measure if the sample size is small. The advantage of the entropy measure is that it is derived from a theoretical basis and not a personal preference, and that it takes into consideration small values with appropriately small weights.

Since entropy is a measure of uncertainty or chaos, and variance σ^2 is a measure of variability, the connection between them is of interest. In general, an explicit relation between entropy and variance does not exist. However, for certain distributions, S can be expressed as a function of σ^2. If two distributions have common variance, an entropy-based measure of affinity or closeness between the distributions, A(.,.), can be defined (Mukherjee and Ratnaparkhi, 1986). The affinity between two distributions is defined by the absolute difference between entropies of the two distributions, which can be shown to be the expectation of the likelihood ratio:

$$A(f_1, f_2) = \left| S_1 - S_2 \right| \tag{13}$$

where f_1 and f_2 are the two probability density functions. Eq. (13) can be cast as:

$$A(f_1, f_2) = \left| E[\log(f_2 / f_1)] \right| \tag{14}$$

which is the expectation of the likelihood ratio. This measure differs from Kullback's minimum distance information criterion. Likewise, a similarity function is defined as one minus the quotient of the affinity between any two distributions and the maximum value of affinity between the distributions:

$$S(f_i, f_j) = 1 - \frac{A(f_i, f_j)}{\max[A(f_i, f_j)]}, i \neq j \tag{15}$$

where S_1 and S_2 are entropies of the two distributions. Thus, affinity (distance) is a monotonically decreasing function of similarity. The similarity factor can be used to cluster or group models.

To measure correlation or dependence between any two variables, an informational coefficient of correlation r_0 is defined as a function of transinformation, T_0:

$$r_o = [1 - \exp(-2T_o)]^{0.5}, 0 \leq r_0 \leq 1 \tag{16}$$

$$T_0 = \iint p(x,y) \ln \frac{p(x,y)}{p(x)\,p(y)} dx\,dy \tag{17}$$

where:

p(x) and q(y) are marginal probability density functions of x and y, respectively

p(x, y) is the joint probability density function of (x, y).

The transinformation expresses the upper limit of common information between two variables and represents the level of dependence (or association) between the variables. This is also referred to as mutual information. It represents the upper limit of transferable information between the variables, and its measure is given by r_0, which is a better measure of correlation than is the ordinary correlation coefficient, r. The ordinary correlation coefficient r measures the amount of information transferred between variables under specified assumptions, such as linearity and normality. An inference similar to that of the ordinary correlation coefficient, r, can be drawn by defining the amount (in per cent) of transferred information by the ratio T/T_0, where T can be computed in terms of ordinary r.

Criteria for model selection

Usually there are more models than one needs and a choice has to be made as to which model to choose. Akaike (1973) formulated a criterion, called Akaike information criterion (AIC), for selecting the best model from amongst several models. The information criterion AIC provides a method of model identification and can be expressed as minus twice the logarithm of the maximum likelihood plus twice the number of parameters used to find the best-fit model:

$$AIC = -2\log(\text{maximized likelihood}) + 2k \tag{18}$$

where k is the number of parameters in the distribution. In case there are several models, the model giving the minimum value of AIC should be selected. When the maximum likelihood is identical for two models, the model with the smaller number of parameters should be selected, for that will lead to smaller AIC and comply with the principle of parsimony.

Hypothesis testing

Another important application of the entropy theory is in testing of hypotheses (Tribus, 1969). With use of Bayes' theorem in logarithmic form, an evidence function is defined for comparing two hypotheses. The evidence in favor of a hypothesis over its competitor is the difference between the respective entropies of the competition and the hypothesis under test. Defining surprisal as the negative of the logarithm of the probability, the mean surprisal for a set of observations is expressed as:

$$\eta_k = -\ln p(x_k) \tag{19}$$

where x_k is the observation in the set of observations. The mean surprisal for the set of m observations is expressed as:

$$\overline{\eta_k} = \frac{1}{m}\sum_{i=1}^{m} \eta_k = -\frac{1}{m}\sum_{i=1}^{m} \ln p(x_k) \tag{20}$$

Therefore, the evidence function, EV, for two hypotheses is obtained as the difference between the two values of the mean surprisal multiplied by the number of observations:

$$EV = m(\overline{\eta_1} - \overline{\eta_1}) \tag{21}$$

Risk assessment

In common language, risk is the possibility of loss or injury and the degree of probability of such loss. Rational decision-making requires a clear and quantitative way of expressing risk. In general, risk cannot be avoided and a choice has to be made between risks. There are different types of risk, such as business risk, social risk, economic risk, safety risk, investment risk, occupational risk, etc. To put risk in proper perspective, it is useful to clarify the distinction between risk, uncertainty, and hazard.

The notion of risk involves both uncertainty and some kind of loss or damage. Uncertainty reflects the variability of our state of knowledge or state of confidence in a prior evaluation. Thus, risk is the sum of uncertainty plus damage:

$$\text{risk} = \text{uncertainty} + \text{damage} \tag{22}$$

Hazard is commonly defined as a source of danger and involves a scenario identification (e.g. failure of a dam) and the consequent measure of that scenario or the measure of the damage. Risk encompasses the likelihood of conversion of that source into the actual delivery of loss, injury, or some form of damage. Thus, risk is the ratio of hazard to safeguards:

$$risk = \frac{hazard}{safeguards} \qquad (23)$$

By increasing safeguards, risk can be reduced but it is never zero. Since awareness of risk reduces risk, awareness is a part of safeguards. Qualitatively, risk is subjective and is relative to the observer. Risk involves the probability of scenario and its consequence resulting from happening of the scenario. Thus, one can say that risk is probability and consequence. Risk is related to entropy, for the latter is employed to derive the probability distribution.

Safety evaluation

Safety has two components: reliability and probabilistic risk assessment (PRA). In reliability, a safety margin or probability of failure is defined when maximum external loads are specified (i.e., design-basis loads such as design discharge). There can be three domains of safety: safe domain (no failure); potentially unsafe domain (failure is possible); and unsafe (failure) domain (a failure is certain). The safe and unsafe domains are separated by a limit state surface. A system will fail if the failure indicator reaches the limit state surface. If we consider a probability density function of failure at a value of the failure indicator, then the cumulative probability of the failure indicator defines fragility.

In PRA, the probability of failure from loads exceeding design basis loads is considered. Through introduction of a hazard function, the probability density function of external loads is specified. Consider a hydraulic system with a set of random parameters. The system can fail in many ways and every failure mode can be described with a corresponding failure indicator involving a number of different failure modes which is a function of hazard parameters and the random parameters. For example, in case of an earth dam, depending upon the failure mode, a failure indicator could be erosion at the bottom, reservoir water level, water leakage, displacement, etc. Hazard parameters could be extreme rainfall, reservoir level, peak discharge, depth of water at the dam top, etc. Structural random parameters could be strengths of materials, degree of riprap, degree of packing, internal friction, etc. Entropy factors in safety evaluation through obtaining probability distributions.

Reliability analysis

Reliability of a system can be defined as the probability that the system will perform its intended function for at least a specified period of time under specified environmental conditions. Different measures of reliability are applied to different systems, depending upon their objective. Indeed, the use of a particular system determines the kind of reliability measure that is most meaningful and most useful. As an example, the reliability measure of a dam is the probability of its survival during its expected life span. On the other hand, the reliability measure associated with hydroelectric power plant components is the failure rate, since the failure of a plant is of primary concern. Furthermore, at different times during the operating life a system may be required to have a different probability of successfully performing its required function under specified conditions. The term "failure" means that the system is not capable of performing its required function. We only consider the case where the system is either capable of performing its functions or not and exclude the case involving varying degrees of capability.

If the reliability is defined as the probability of success, that is, the system will perform its intended function for at least a defined period of time, then the reliability function can be computed directly from the knowledge of the failure time distribution. If the system is resurrected through repair and maintenance then the mean failure time is known as the mean (operating) time between failures. The mean time to failure is the expected time during which the system will perform successfully, also expressed as the expected life.

The rate at which failures occur in a time interval is the failure rate and is defined by the probability that a failure per unit time occurs in the interval, provided that a failure has not occurred prior to the beginning of the interval. The hazard rate (or hazard function) is defined by the limit of failure as the length of the time interval approaches zero. This implies the instantaneous failure rate. The importance of entropy in reliability engineering lies in its ability to incorporate prior knowledge in the estimated reliability.

Entropy theory as a modelling tool in environmental and water resources

A historical perspective on entropy applications in environmental and water resources is given in Singh and Fiorentino (1992) and Singh (1997). Harmancioglu and Singh (1998) discussed the use of entropy in water resources. In what follows is a discussion of entropy-based applications.

Derivation of probability distributions

Often needed are frequency or probability distributions that satisfy the given information. The entropy theory is ideally suited to that end. Indeed POME has been employed to derive a variety of distributions some of which have found wide applications in environmental and water resources. Many of these distributions have been summarised in Singh et al. (1986), Singh and Fiorentino (1992) and Singh (1998). When there are no constraints, then POME yields a uniform distribution. As more constraints are introduced, the distribution becomes more peaked and possibly skewed. In this way, the entropy reduces from a maximum for the uniform distribution to zero when the situation is entirely deterministic.

Consider, as an example, the case of deriving the instantaneous unit hydrograph (IUH) for a watershed. The IUH can be considered as a probability distribution of the time of travel (t). Thus, the time of travel is construed as the random variable. It is assumed that rainfall occurs uniformly over the entire watershed and that abstractions, such as surface detention, depression storage, antecedent condition, infiltration, etc., are satisfied by the watershed first before occurrence of surface runoff. Thereafter, the rainfall would be the rainfall excess which leads to surface runoff. This assumption is essential for consideration of the IUH.

When rainfall excess occurs over the space of the watershed, the rainfall excess starts travelling toward the outlet. Since the watershed surface has virtually infinite number of points at which the travel is initiated, the number of travel paths is infinite. Once a water mass starts to travel, its travel path is pretty much determined first by the slope and other characteristics of the overland plane and then by the channel network. Thus, it is not unrealistic to surmise that the number of values of the travel times is infinite. The water droplets falling at the watershed outlet will take virtually no time to reach the outlet, i.e., the travel time is zero. The water droplets falling at the most remote portion will take the longest path and consequently the longest time, equal to the time of concentration, T_c. Thus, t varies from zero to T_c . However, for simplicity of analysis, it is assumed that t varies from zero to infinity.

For deriving the IUH, we invoke our elementary understanding of watershed hydrology and express it in quantitative terms called constraints. Since IUH is a probability distribution, it must satisfy

$$\int_0^\infty f(t)\,dt = 1 \tag{24}$$

Since the time of travel has a range from 0 to T_c, we may wish to express an average value of t as

$$\int_0^\infty t\, f(t)\,dt = \bar{t} = E[t] = k \tag{25}$$

It may be recalled that the average travel time is the same as the lag time of the watershed, which has been found to have a strong correlation with watershed characteristics, such as area, slope, etc. (Singh, 1988).

To minimise the impact of extreme values on the value of k, we may wish to express the average of the logarithmically transformed values of travel times:

$$\int_0^\infty \ln t\, f(t)\,dt = \overline{\ln t} = E[\ln t] \tag{26}$$

If Eqs. (25) and (26) constitute the information we have on the travel time, then we can invoke POME to derive the probability density function of t, or IUH. To that end, we maximise the Shannon entropy Eq. (8), subject to Eqs. (24) to (26). This is an optimisation problem and can be solved using the method of Lagrange multipliers. To that end, introducing the Lagrange multipliers (λ_0, λ_1, λ_2), the function, L(f), to be maximised becomes:

$$L(f) = -\int_0^\infty f(t)\ln f(t)\,dt - \lambda_0\left[\int_0^\infty f(t)\,dt - 1\right] - \lambda_1\left[\int_0^\infty t\, f(t)\,dt - k\right] - \lambda_1\left[\int_0^\infty \ln t\, f(t)\,dt - \overline{\ln t}\right] \tag{27}$$

Taking partial derivatives of L with respect to f and setting the resulting equation to zero, the following results:

$$\partial L(f) = -\int_0^\infty [\ln f(t) + 1 + (\lambda_0 - 1) + \lambda_1 t + \lambda_2 \ln t]\,\partial f(t)\,dt = 0 \tag{28}$$

Eq. (28) yields:

$$f(t) = \exp[-\lambda_0 - \lambda_1 t - \lambda_2 \ln t] \tag{29}$$

Eq. (29) specifies the entropy-based probability density function. The Lagrange multipliers (λ_0, λ_1, λ_2) are to be expressed in terms of Eqs. (24) to (26). To that end, the zeroth Lagrange multiplier, also called the potential function, is obtained by substituting Eq. (29) in Eq. (24):

$$\lambda_0 = \ln \int_0^\infty \exp[-\lambda_1 t - \lambda_2 \ln t] \tag{30}$$

Solution of Eq. (30) yields

$$\exp(\lambda_0) = \frac{1}{\lambda_1^{\,1-\lambda_2}}\Gamma(1-\lambda_2) \tag{31}$$

The zeroth Lagrange multiplier is also expressed as:

$$\lambda_0 = (\lambda_2 - 1)\ln \lambda_1 + \ln[\Gamma(1-\lambda_2)] \tag{32}$$

Differentiating Eqs. (30) and (32) with respect to λ_1 and λ_2 and equating each derivative, respectively, to the constraints in Eqs. (25) and (26) and solving, we obtain:

$$\frac{\partial \lambda_0}{\partial \lambda_1} = -E[t] = \frac{\lambda_2 - 1}{\lambda_1} \tag{33}$$

$$\frac{\partial \lambda_0}{\partial \lambda_2} = \ln \lambda_1 + \frac{\partial}{\partial \lambda_2}\Gamma(1-\lambda_2) = -E[\ln t] \tag{34}$$

Eqs. (33) and (34) yield:

$$E[t] = \frac{1-\lambda_2}{\lambda_1} \tag{35}$$

$$E[\ln t] = -\ln \lambda_1 + \Psi(1-\lambda_2) \tag{36}$$

where $\Psi(b)$ = d [ln $\Gamma(b)$]/db is the digamma function, with b = 1- λ_2 . Substituting Eq. (35) and (36) in Eq. (29), one obtains:

$$f(t) = \frac{1}{a\,\Gamma(b)}\left(\frac{t}{a}\right)^{b-1}\exp\left[-\frac{t}{a}\right] \tag{37}$$

where λ_1 = 1/a and b = 1- λ_2. Eq. (34) is the familiar 2-parameter gamma distribution. It is least biased and consistent with the data expressed in terms of the constraints.

If, for example, one also has an idea about the time taken by the antecedent condition to be satisfied. Let that time be t_0 . This can be thought of as the time to ponding. Then the constraints can be written as:

$$\int_{t_0}^\infty \ln(t - t_0)\, f(t)\,dt = E[\ln(t - t_0)] \tag{38}$$

$$\int_{t_0}^\infty (t - t_0)\, f(t)\,dt = E[t] = k \tag{39}$$

Following the same procedure, the resulting IUH equation would be:

$$f(t) = \frac{1}{a\,\Gamma(b)}\left(\frac{t-t_0}{a}\right)^{b-1}\exp\left[-\frac{t-t_0}{a}\right] \tag{40}$$

which is the Pearson type 3 distribution.

Parameter estimation

It is desirable to estimate parameters of a probability distribution in terms of the given constraints (Jaworski, 1987). The entropy theory accomplishes precisely that. Singh (1998) has described POME-based estimation for a number of probability distributions used in environmental and water resources. He has also discussed a comparison of the POME method with the methods of moments, maximum likelihood estimation, and some others. The comparison shows that the POME method is either comparable to or better than other methods.

As an example, consider the case of the gamma distribution used in hydrology. As an IUH model, it is frequently written as:

$$h(t) = \frac{1}{k\,\Gamma(n)}\left(\frac{t}{k}\right)^{n-1}\exp\left[-\frac{t}{k}\right] \tag{41}$$

where k is the reservoir lag time and n is the number of linear reservoirs. Eq. (41) is was proposed by Nash (1958) and is a response of a cascade of equal linear reservoir subjected to an instantaneous burst of rainfall excess having a unit volume. This equation has two parameters k and n, which need to be estimated. Singh and Rajagopal (1986) developed an entropy-based parameter estimation, called parameter space expansion method. For Eq. (41), this method works as follows. First, we determine the constraints on which Eq. (41) is based. This is obtained by substituting Eq. (41) in Eq. (24):

$$S = -\int_0^\infty f(t) \ln f(t)\, dt \tag{42}$$

$$= -\int_0^\infty \frac{1}{k\,\Gamma(n)} \left(\frac{t}{k}\right)^{n-1} \exp\left[-\frac{t}{k}\right] \ln\left\{ \frac{1}{k\,\Gamma(n)} \left(\frac{t}{k}\right)^{n-1} \exp\left[-\frac{t}{k}\right]\right\} dt$$

The constraints are:

$$\int_0^\infty t\, f(t)\, dt = \bar{t} \tag{43}$$

$$\int_0^\infty \ln(t)\, f(t)\, dt = E[\ln t] = \overline{\ln t} \tag{44}$$

The Pome-based formulation of the distribution is given by:

$$f(t) = \exp[-\lambda_0 - \lambda_1 t - \lambda_2 \ln t] \tag{45}$$

which is the same as Eq. (29). Therefore,

$$E[t] = k$$
$$E[\ln t] = \ln k + \Psi(n), \Psi(n) = d[\ln \Gamma(n)]/dn \tag{46}$$

Singh and Singh (1985) applied Eqs. (45) and (46) to a number of rainfall-runoff events from a large experimental rainfall-runoff facility. The rainfall intensity was uniform in both time and space. Parameters n and k were estimated by POME and also by the method of moments (MOM), method of cumulants (MOC), maximum likelihood estimation (MLE), and method of least squares (MOLS). Using the parameter values estimated by these methods, the unit hydrographs were generated and convoluted with appropriate rainfall hyetographs to generate runoff hydrographs. For a sample events, observed and computed runoff hydrographs are shown in Fig. 2. For this event, POME, MOM and MOLS were comparable. This was also true for other events. Thus, Pome offers a competitive method of parameter estimation.

Maximum entropy-spectral analysis for flow forecasting

Burg (1968) defined entropy in terms of power spectrum W(f), where f is frequency. For a stationary stochastic process x(t), the Burg entropy, H(f), is expressed as:

$$H(f) = \int_{-w}^{+w} \ln[W(f)]\, df \tag{47}$$

where w is the frequency band. The spectrum W(f) can be written in terms of the Fourier series as:

$$W(f) = \frac{1}{2w} \sum_{n=-\infty}^{\infty} \rho(n) \exp[-i 2\pi n f \Delta t] \tag{48}$$

where Δt is the sampling time interval, $i = (-1)^{0.5}$, and $\rho(n)$ is the autocorrelation defined as:

$$\rho(n) = \int_{-w}^{+w} W(f) \exp[i 2\pi f n \Delta t]\, df \quad , -m \le n \le m \tag{49}$$

where m is lag. It has been shown (Krstanovic and Singh, 1993a, b) that maximisation of Eq. (47) is equivalent to maximisation of:

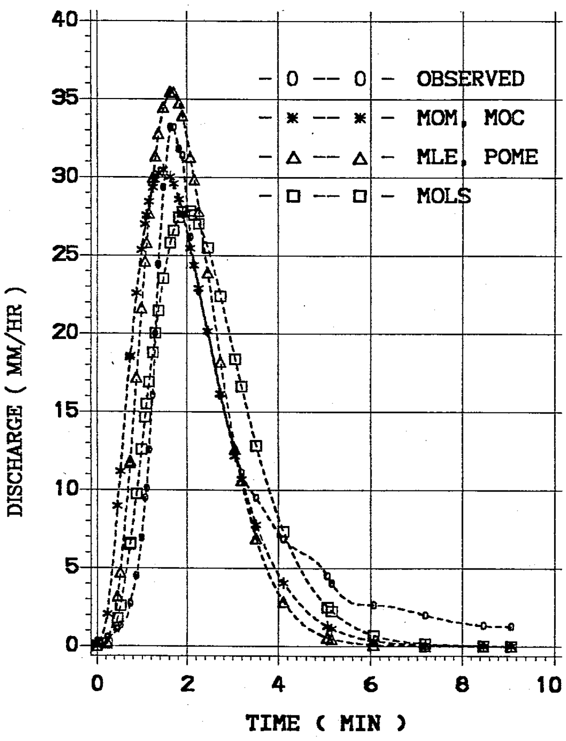

Figure 2
A comparison of observed and computed runoff hydrographs for event 3. The methods of parameter estimation are MOM, MOC, MLE, POME, and MOLS

$$H(f) = \frac{1}{2} \ln(2w) + \frac{1}{4w} \int_{-w}^{+w} \ln[W(f)]\, df \tag{50}$$

subject to constraints defined by Eq. (49). Substitution of Eq. (48) in Eq. (50) and maximisation leads to MESA.

The maximum entropy spectral analysis (MESA), introduced by Burg (1975), has several advantages over conventional spectral analysis methods. It has short and smooth spectra with high-degree resolutions. The statistical characteristics which are used in stochastic model identification can also be estimated using MESA, thus permitting integration of spectral analysis and computations related to stochastic model development. Jaynes (1982) has shown that MESA and other methods of spectral analysis, such as Schuster, Blackman-Tukey, maximum likelihood, Bayesian, and autoregressive (AR, ARMA or ARIMA) models are not in conflict, and that AR models are a special case of MESA. Krstanovic and Singh (1991a,b) employed MESA for long-term streamflow forecasting. Krstanovic and Singh (1993a,b) extended the MESA method to develop a real-time flood forecasting model. Figure 3 shows streamflow regeneration for River Orinoco for the case of forward forecastinmg (Singh and Krstanovic, 1991b). This figure clearly shows a close agreement between observed and MESA-computed streamflows.

Basin geomorphology

Entropy plays a fundamental role in characterisation of landscape. Using the entropy theory for morphological analysis of river basin networks, Fiorentino et al. (1993) found the connection between entropy and the mean basin elevation as:

$$y = -a \ln b + a S \tag{51}$$

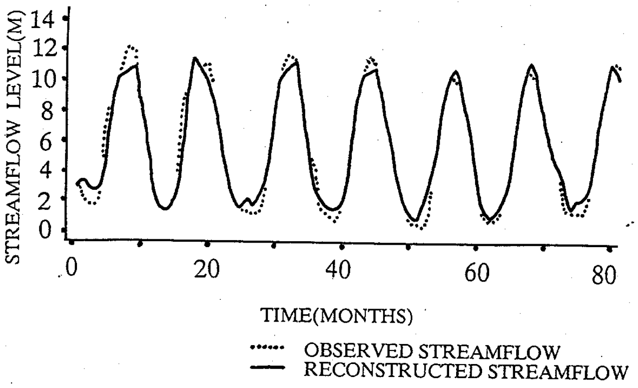

Figure 3
Streamflow regeneration for River Orinoco (forward forecasting)

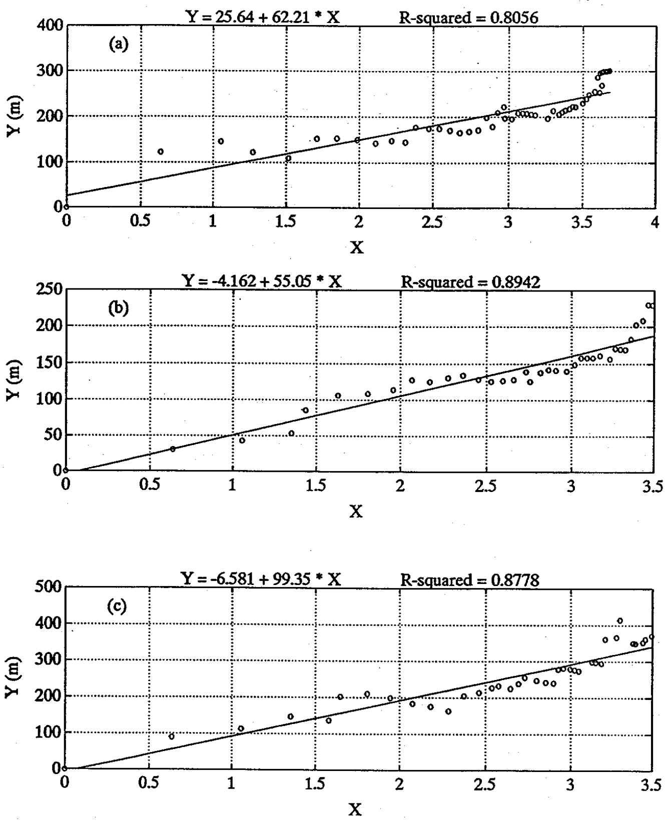

Figure 4
Average elevation versus Shannon entropy of subnetworks whose outlets lie on the main channel and respective least squares lines, for (a) Arcidiaconata, (b) Lapilloso, and (c) Vulgano basins.

where:

y is the mean basin elevation,
S is entropy, and
a and b are parameters having physical significance.

Similarly, the relation between the fall in elevation from the source to the outlet of the main channel and the entropy of its drainage basin was found to be linear:

$$F = \alpha + \beta S \qquad (52)$$

where F is the fall in the elevation from the source to the outlet of the main channel, and α and β are parameters. In a similar vein, the relation between the elevation of a node and the logarithm of its distance from the source was found to be:

$$y_\delta = y_0 - F\,\frac{\ln \delta}{\ln D} \qquad (53)$$

where:

y_0 is the elevation of the source of the channel
y_δ is the elevation of the downstream node at a distance from the source
D is the topological diameter
F is the fall in elevation from the source to the outlet.

When a basin was ordered following the Horton-Strahler ordering scheme, a linear relation was found between the drainage entropy and the basin order:

$$S = \Omega \ln R_L - \ln(R_L - 1) \qquad (54)$$

where R_L is the stream length ratio, and S is the order of the basin. This relation can be characterised as a measure of the basin network complexity. The basin entropy was also found to be linearly related to the logarithm of the magnitude of the basin network:

$$S = (\Omega - 1)\ln R_L + \ln\left[\frac{R_L}{R_L - 1}\right] \qquad (55)$$

This relation led to a non-linear relation between the network diameter and magnitude where the exponent was found to be related to the fractal dimension of the drainage network:

$$S = \ln\left(\frac{R_L}{R_L - 1} n^{1/c}\right) \qquad (56)$$

where:

c = 1.75, and
n is the magnitude of the drainage basin.

Fiorentino et al. (1993) verified Eqs. (51) to (56) on a number of basins in Italy and close agreements between entropy-based derivations and observed geomorphologic variables. Figure 4 shows a relation between the average elevation and Shannon entropy for three subbasins whose outlets lie on the main channel. In the figure the x-axis denotes entropy and the y-axis the average elevation..

Design of hydrologic networks

The purpose of measuring networks is to gather information in terms of data. Fundamental to evaluation of these networks is the ability to determine if the networks are gathering the needed information optimally. The entropy theory is a natural tool to make that determination. Krstanovic and Singh (1992a,b) employed the theory for space and time evaluation of rainfall networks in Louisiana. The decision whether to keep or to eliminate a rain-gauge was based entirely on reduction or gain of information at that gauge. In this manner the best combination of rain-gauges was suggested. Yang and Burn (1994) employed a measure of information flow, called directional information transfer index (DIT), between gauging stations in the network. The DIT is defined as:

$$DIT = \frac{T}{S} = \frac{S - S_{lost}}{S} = 1 - \frac{S_{lost}}{S} \qquad (57)$$

where:

S is the entropy (or information content) of the station X
T is the transinformation (or mutual information) between station X and Y
S_{lost} is the information lost (or the amount of information transmitted).

The value of DIT varies from 0 where no information is transmitted and the stations are independent to one where no information is lost and the stations are fully dependent. Between two stations of one pair, the station with higher DIT value should be retained because of its greater capability of inferring information at the other side.

Note that DIT is not symmetrical, for DIT (X,Y) = T/S(X) for station X is, in general, not equal to DIT (Y, X) = T/S(Y) for station Y. DIT (X, Y) describes the fractional information inferred by station X about station Y, whereas DIT (Y, X) is the fractional information inferred by station Y about station X. Between two stations of one pair, the station with higher DIT value should be retained because of its greater capability of inferring information at other stations. The concept of DIT can be extended to regionalisation of networks.

Reliability of water distribution systems

Entropy-based measures have been developed for evaluation of reliability and redundancy of water distribution networks. These measures accurately reflect changes in the network reliability. However, the redundancy of a network also depends on the ability of the network to respond to the failure of one of its links. Awumah et al. (1990) applied these measures to evaluate reliability/redundancy of a range of network layouts and showed that the entropy-based redundancy measure was a good indicator of the relative performance implications of different levels of redundancy.

The network redundancy (or reliability) measure of a water distribution system is expressed as:

$$R = \sum_{j=1}^{N} \left(\frac{Q_j}{Q_0} R_j \right) - \sum_{j=1}^{N} \left(\frac{Q_j}{Q_0} \ln \frac{Q_j}{Q_0} \right) \quad (58)$$

where:
R is the network redundancy
Q_0 is the sum of flows in all links in the network
Q_j is total flow into node j
N is the number of nodes.

R_j is expressed as:

$$R_j = -\sum \frac{q_{ij}}{Q_j} \ln \frac{q_{ij}}{Q_j} \quad (59)$$

where:
q_{ij} is the flow in the link from node i to node j
Q_j is the set of nodes on the upstream ends of links incident upon node j.

The right side of Eq. (58) is the sum of two terms. The first terms expresses an algebraic sum of weighted entropy measure at each of the constituent nodes, and the second term defines the redundancy among the N nodes. In a similar manner, Awumah et al. (1990) extended the analysis to take account of the ability of a network to responds to failure of one of its links.

Subsurface hydrology

In groundwater engineering, it is often true that few measurements of aquifer and flow parameters, such as hydraulic conductivity, are available, and there is a large degree of uncertainty in the measured values of fundamental flow parameters. Woodbury and Ulrych (1993) used the principle of minimum relative entropy (POMRE) to determine these parameters. Barbe et al. (1994) applied POME to derive a probability distribution for piezometric head in one-dimensional steady groundwater flow in confined and unconfined aquifers, subject to the total probability law and the conservation of mass. For example, the cumulative probability distribution (cdf) of piezometric head h in a confined aquifer was found to be:

$$F(h) = \frac{h - h_0}{h_u - h_0} \quad (60)$$

where:
h_0 is the piezometric head at the lower end of the aquifer at x = 0
h_u is the head at the upper end at x = L
L is the length of the aquifer.

The probability density function of h, f(h), is:

$$f(h) = \frac{1}{h_u - h_0} \quad (61)$$

which is a uniform distribution. This is based on the assumption that nothing is known about the distribution of head and therefore there is no constraint.

From a few measurements of transmissivity (T) based on pumping tests and of piezometric head, Bos (1990) employed POME and Bayes' theorem to derive the probability distribution of transmissivity.

Application in hydraulics

Yang (1994) showed that the fundamental theories in hydrodynamics and hydraulics can be derived from variational approaches based on maximisation of entropy, minimisation of energy, or minimisation of energy dissipation rate. Chiu and Murray (1992) applied POME to determine the probability distribution of velocity in non-uniform open channel flow:

$$\frac{u}{u_m} = \frac{1}{m} \ln \left[1 + (e^m - 1) \frac{y - y_0}{y_m - y_0} \right] \quad (62)$$

where:
u is the velocity that monotonically increases with y
m is an entropy parameter
u_m is the maximum velocity in a channel cross-section
y_m is the maximum value of y
y_0 is the minimum value of y.

The entropy-based velocity distribution fits experimental data very well and is of great practical value in hydraulic modelling.

Water quality assessment

Environmental pollution can be perceived as a result of discharge of material and heat into the environment (water, air, and/or soil) through human activity of production and consumption. When a compound is added to pure water, the compound will dissolve and diffuse throughout water. The dissolution and diffusion imply an increase in the entropy of the solution (by virtue of its definition and the second law of thermodynamics) and an increase in the degree of pollution. This suggests that an increase in entropy implies water pollution. Water is extensively used in cooling, washing, disposal of waste material, and dissipation of waste heat. Water pollution can then be viewed as water initially containing a low value of entropy being eventually discharged with high value of entropy which, in turn, increases the entropy of the environment. Thus, entropy can serve as a comprehensive index for assessment of pollution control. To extend the argument further, the diversity of species of organisms in water or the diversity index (DI) is related to the degree of pollution. In general, the number of species decreases as the degree of pollution increases. The DI can be calculated from the entropy theory as:

$$DI = -\sum_{i=1}^{n} p_i \ln p_i \qquad (63)$$

where p_i is the number of organisms of species i divided by the total number of organisms present in the water. Thus, DI can be used to evaluate water quality of a water body in time and space, as well as to compare water quality of different water bodies. Because entropy increases with pollution, energy is required to abate pollution, remove pollutants from water and purify it and in turn decrease entropy of the polluted water. This suggests that the efficiency of water treatment systems can be expressed by the entropy production and thus their thermodynamical efficiency can be evaluated.

Design of water quality networks

The entropy theory, when applied to water quality monitoring network design, yields promising results, especially in the selection of technical design features, such as monitoring sites, time frequencies, variables to be sampled, and sampling duration. Furthermore, it permits a quantitative assessment of efficiency and benefit/cost parameters. Harmancioglu and Singh (1998) reviewed the advantages as well as the limitations of the entropy method as applied to the design of water quality monitoring networks. Given an observed change in water quality levels at a downstream location, the entropy-based formulation predicts the probabilities of each possible water quality level at each of the upstream stations.

Optimisation

Optimisation is widely needed in modelling as well as decision making in environmental and water resources. The entropy theory is a potentially powerful tool in constrained optimisation as shown by Templeman and Xingsi (1987). This is a promising new area of research.

Implications for developing countries

One of the main problems plaguing environmental and water resources development in developing countries is the lack of data or lack of sufficient data. Frequently, either the data are missing or it is incomplete, or it is not of good quality or the record is not of sufficient length. As a result, more often than not, it is the data that dictate the type of model to be used and not the availability of modelling technology. Many conventional models are not applicable when their data needs are not met. Furthermore, subjective information such as professional experience, judgment, and thumb or empirical rules have played a significant role in hydrologic practice in many developing countries. Conventional models do not have the capability to accommodate such subjective information, although such information may be of good quality or high value. The potential for application of the entropy theory is enormous in developing countries, for it maximises the use of information contained in data, however little it may be, and it permits use of subjective information. Thus, in the face of limited data the entropy theory results in a reliable solution of the problem at hand. Furthermore, it offers an objective avenue for drawing inferences as to the model results. In addition, the entropy-based modelling is efficient, requiring relatively little computational effort and is versatile in its applicability across many disciplines.

Concluding remarks

The entropy theory is versatile, robust, and efficient. It permits determination of the least-biased probability distribution of a random variable, subject to the available information. Furthermore, it suggests if the available information is adequate or not and if not, then additional information should be sought. In this way it brings the model, the modeller, and the decision-maker closer. As an objective measure of information or uncertainty, the entropy theory allows to communicate with nature as illustrated by its application to design of data acquisition systems, design of environmental and hydrologic networks, and assessment of reliability of these systems or networks. In a similar vein, it helps better understand physics or science of natural systems such as landscape evolution, geomorphology, and hydrodynamics. A wide variety of seemingly disparate or dissimilar problems can be meaningfully solved with use of entropy.

References

AKAIKE H (1973) Information theory and an extension of the maximum likelihood principle. In: BN Petrov and F Csaki (ed.) *Proc. 2nd Int. Symp. on Information Theory*, Publishing House of the Hungarian Acad. of Sci., Budapest, Hungary, .

AWUMAH K, GOULTER IC and BHATT SK (1990) Assessment of reliability in water distribution networks using entropy-based measures. *Stochastic Hydrol. and Hydraulics* **4** 309-320.

BARBE DE, CRUISE JF and SINGH VP (1994) Derivation of a distribution for the piezometric head in groundwater flow using stochastic and statistical methods. In: Hipel KW (ed.) *Hydrology and Environmental Engineering*, Vol. **2**, Kluwer Academic Publishers, Dordrecht, The Netherlands.

BOS R (1990) Aquifer Parameter Identification by the Maximum Entropy and Bayes' Theorem. Unpublished Report, Technion-Israel Institute of Technology, Haifa, Israel.

BURG JP (1968) A new technique for time series data. Paper presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, The Netherlands. 12-13 August.

BURG JP (1975) Maximum Entropy Spectral Analysis. Unpublished Ph.D. Thesis, Stanford Univ., Palo Alto, California.123 pp.

CHIU CL and MURRAY DW (1992)Variation of velocity distribution along nonuniform open channel flow. J. of Hydraulic Eng. **118** (1) 989-1001.

FIORENTINO M, CLAPS P and SINGH VP (1993) An entropy-based morphological analysis of river basin networks. Water Resour. Res. **29**(4) 1215-1224.

HARMANCIOGLU NB and SINGH VP (19??) Entropy in environmental and water resources. In: *Encyclopedia of Hydrology and Water Resources*, Kluwer Academic Publishers, Boston, Massachusetts. 225-241.

JAWORSKI W (1987) Higher-order moments and the maximum entropy inference: The thermodynamic limit approach. *J. of Phys. A: Math. and Gen*. **20** 915-926.

JAYNES ET (1957) *Information Theory and Statistical Mechanics, I. Physical Review*. **106** 620-630.

JAYNES ET (1982) On the rationale of maximum entropy methods. *Proc. of the IEEE* **70** 939-952.

KRSTANOVIC PF and SINGH VP (1991a) A univariate model for long-term streamflow forecasting: 1. Development. *Stochastic Hydrol. and Hydraulics* **5** 173-188.

KRSTANOVIC PF and SINGH VP (1991b) A univariate model for long-term streamflow forecasting: 2. Application. *Stochastic Hydrol. and Hydraulics* **5** 189-205.
KRSTANOVIC PF and SINGH VP (1992a) Evaluation of rainfall networks using entropy: 1. Theoretical development. *Water Resour. Manage.* **6** 279-293.
KRSTANOVIC PF and SINGH VP (1992b) Evaluation of rainfall networks using entropy: II. Application. *Water Resour. Manage*. **6** 295-314.
KRSTANOVIC PF and SINGH VP (1993a) A real-time flood forecasting model based on maximum entropy spectral analysis: 1. Development. *Water Resour. Manage.* **7** 109-129.
KRSTANOVIC PF and SINGH VP (1993b) A real-time flood forecasting model based on maximum entropy spectral analysis: 2. Application. *Water Resour. Manage*. **7** 131-151.
KULLBACK S and LEIBLER RA (1951) On information and sufficiency. *Annals of Math. Stat*. **22** 79-86.
LEVINE RD and TRIBUS M (eds.) (1978) *The Maximum Entropy Formalism*. The MIT Press, Cambridge, Massachusetts. 498 pp.
MUKHERJEE D and RATNAPARKHI MV (1986) On the functional relationship between entropy and variance with related applications. *Communications on Statistical Theory and Methods* **15** (1) 291-311.
NASH JE (1958) Determining runoff from rainfall. *Proc. of the Inst. of Civil Eng*.**10** 163-184.
SHANNON CE (1948) A mathematical theory of communications, I and II. *Bell System Techn. J*. **27** 379-443.
SINGH VP (1988) *Hydrologic Systems, Vol. 1: Rainfall-Runoff Modelling.* Prentice Hall, Englewood Cliffs, New Jersey.
SINGH VP (1997) The use of entropy in hydrology and water resources. *Hydrol. Processes* **11** 587-626.
SINGH VP (1998) *Entropy-Based Parameter Estimation in Hydrology*. Kluwer Academic Publishers, Boston, Massachusetts, 1998.
SINGH VP and FIORENTINO M (ed.) (1992) *Entropy and Energy Dissipation in Water Resources*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
SINGH VP and SINGH K (1985) Derivation of the gamma distribution by using the principle of principle of maximum entropy (POME). Water Resour. Bull. **21** (6) 941-952.
SINGH VP and RAJAGOPAL AK (1986) Some recent advances in application of the principle of maximum entropy (POME) in hydrology. IAHS Publication No. **164** 353-364.
SINGH VP, RAJAGOPAL AK and SINGH K (1986) Derivation of some frequency distributions using the principle of maximum entropy (POME). Advances in Water Resour. **9** (2) 91-106.
TEMPLEMAN AB and XINGSI L (1987) A maximum entropy approach to constrained nonlinear programming. *Eng. Optimization* **12** 191-205.
TRIBUS M (1969) *Rational Description: Decision and Designs*. Pergamon Press, New York.
WOODBURY AD and ULRYCH TJ (1993) Minimum relative entropy: Forward probabilistic modelling. *Water Resour. Res*. **29** 2847-2860.
YANG CT (1993) Variational theories in hydrodynamics and hydraulics. *J. of Hydraulic Eng*. **120** (6) 737-756.
YANG Y and BURN D H (1994) An entropy approach to data collection network design. *J. of Hydrol*. **157** 307-324.