HYDROLOGICAL PROCESSES, VOL. 11, 587–626 (1997)

THE USE OF ENTROPY IN HYDROLOGY AND WATER RESOURCES

V. P. SINGH

Water Resources Program, Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA 70803-6405, USA

ABSTRACT

Since the development of the entropy theory by Shannon in the late 1940s and of the principle of maximum entropy (POME) by Jaynes in the late 1950s there has been a proliferation of applications of entropy in a wide spectrum of areas, including hydrological and environmental sciences. The real impetus to entropy-based hydrological modelling was provided by Amorocho and Espildora in 1972. A great variety of entropy applications in hydrology and water resources have since been reported, and new applications continue to unfold. This paper reviews the recent contributions on entropy applications in hydrology and water resources, discusses the usefulness and versatility of the entropy concept, and reflects on the strengths and limitations of this concept. The paper concludes with comments on its implications in developing countries. © 1997 by John Wiley & Sons, Ltd.

Hydrological Processes, vol. 11, 587–626 (1997).

(No. of Figures: 0 No. of Tables: 0 No. of Refs: 159)

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: (1) system structure (geometry), (2) system dynamics, (3) forcing functions (sources and sinks), and (4) 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 are frequently made with less than adequate information. Such decisions concerning the environmental and water resource systems often may be based on experience, professional judgment, thumb rules, crude analyses, safety factors or probabilistic methods. Quite often, sufficient data are not available to describe the random behaviour of such systems. Although probabilistic methods allow for a more explicit and quantitative accounting of uncertainty, their major difficulty occurs owing to 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. This characteristic of entropy is particularly appealing in developing countries where shortage of data is rampant. The entropy theory provides a natural way for determining risk associated with an environmental or a water resources system, and can serve as a basis for risk and reliability analysis.

In recent years, the entropy theory has been applied to a great variety of problems in hydrology and water resources. Singh and Rajagopal (1987) discussed advances in application of the principle of maximum entropy (POME) in hydrology. Rajagopal *et al.* (1987) presented new perspectives for potential applications of entropy in water resources research. Singh (1989) reported on hydrological modelling using entropy. A historical perspective on entropy applications in water resources was presented by Singh and Fiorentino (1992). Harmancioglu *et al.* (1992) discussed the use of entropy in water resources. Alpaslan *et al.* (1992) discussed the role of entropy, and Harmancioglu *et al.* (1992) its application in design and evaluation of

CCC 0885–6087/97/060587–40$17·50

Received 23 March 1996
Accepted 7 June 1996

water quality monitoring networks. These surveys have discussed the state of the art of entropy-based modelling and applications in environmental and water resources, and no attempt will be made to repeat what is already discussed in these surveys. The objective of this study is to provide a survey of the entropy concept and its more recently reported applications in hydrology and water resources. Thus, an attempt will be made to update and supplement the discussion contained in the above surveys.

CONCEPT OF ENTROPY

Clausius coined the word 'entropy' from the Greek meaning transformation. Thus, entropy originated in physics and occupies an exceptional position among physical quantities. It does not appear in the fundamental equations of motion. Its nature is, rather, a statistical or probabilistic one, for it can be interpreted as a measure of the amount of chaos within a quantum mechanical mixed state. It is an extensive property like mass, energy, volume, momentum, charge, number of atoms of chemical species, but, unlike these quantities, it does not obey a conservation law.

In physical sciences entropy relates macroscopic and microscopic aspects of nature and determines the behaviour of macroscopic systems in equilibrium (or close to equilibrium). Entropy is not an observable; that means there does not exist an operator with the property that its expectation value in some state would be its entropy (Wehri, 1978). It is, rather, a function of a state. For example, if the state is described by the density matrix, its entropy is given by the van Neumann formula.

Entropy is viewed in three different but related contexts and is hence typified by three forms as discussed in what follows.

Thermodynamic entropy

The concept of entropy is defined in thermodynamics as a state variable for systems in thermal equilibrium. The definition of experimental entropy H_E in the form given in 1850 by Clausius can be written as

$$H_E = \int_{\text{reversible path}} \frac{dQ}{T} \tag{1a}$$

where T is the absolute temperature and Q is the amount of heat entering the system. In this manner, entropy is defined as a function of the macroscopic variables such as pressure and temperature, and its numerical value can be measured experimentally (up to a constant). This constant is provided by the third law of thermodynamics, according to which H_E vanishes at the absolute zero of temperature (Gull, 1991). Thus, the classical definition of thermodynamic entropy is

$$H = \int \frac{dE}{T} \text{ or } dHT = dE \tag{1b}$$

where H is entropy, E is energy per unit mass and T is absolute temperature.

Entropy is an extensive property of a system and the total entropy is equal to the sum of entropies of individual parts. For a system composed of m subsystems or states it can be expressed as (Prigogin, 1967):

$$H = \sum_{i=1}^{m} H_i = \sum_{i=1}^{m} k \ln p_i + \text{constant} \tag{2}$$

where H_i is entropy of the ith subsystem, p_i is probability of being in the ith state and k is a constant. Lewis and Randall (1961) state that the most probable distribution of energy in a system is such that the entropy of the whole system is a maximum.

$$H = \sum_{i=1}^{m} k \ln p_i = \text{a maximum} \tag{3a}$$

This occurs under the condition of dynamic equilibrium. During evolution towards a stationary state the rate of entropy production per unit mass should, as shown by Prigogin (1967), be a minimum compatible with external constraints. This is referred to as the Prigogin principle, expressed as

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \text{a minimum} \tag{3b}$$

This principle is of great value in explaining fluvial processes and river behaviour. The principle of minimum entropy is a by-product of Onsager's variational principle (Lavenda, 1978). In generalized thermodynamics, entropy is decomposed into two parts:

$$\mathrm{d}H = \mathrm{d}H_\mathrm{e} + \mathrm{d}H_\mathrm{i} \tag{4}$$

where $\mathrm{d}H_\mathrm{e}$ is the entropy exchanged between the system and its surroundings and $\mathrm{d}H_\mathrm{i}$ is the entropy produced in the system. Equation (4) expresses the entropy balance. These two types of entropies can be identified with use of the Gibbs relation.

Statistical–mechanical entropy

In 1866 Boltzmann formulated the kinetic theory of gases, giving birth to statistical thermodynamics. He studied a gas of N particles each in a six-dimensional phase space of position and momentum, and studied how collisions led to an equilibrium distribution. He divided the phase space (position–momentum space) available to the molecules into discrete cells. He then defined entropy, called Boltzmann entropy, H_B, in terms of an H function as

$$H_\mathrm{B} = -k_\mathrm{B}H \tag{5a}$$

where

$$H = \int \mathrm{d}^3x\, \mathrm{d}^3p\rho \log \rho \tag{5b}$$

where $\rho(x, p, t)$ is the distribution of particles, k_B is Boltzmann's constant, p is the phase volume and x is the space coordinate. The notation $\mathrm{d}^3x\, \mathrm{d}^3p$ stands for phase–volume element in the space of one particle. Lebowitz (1993) has presented a comprehensive historical account of Boltzmann's entropy.

Nearly 35 years later Gibbs, in 1901, formulated statistical mechanics and developed what is popularly known as the Gibbs algorithm using the canonical ensemble. The ensemble is defined by the N-particle distribution function, or Liouville function $\rho(x_1, p_1; x_2, p_2; \ldots, x_N, p_N; t)$ which gives the probability density in the full phase space of the system. The term phase stands for the collection of all coordinates and momenta that is the Gibbs ensemble. The Gibbs entropy H_G is defined in terms of the joint probability distribution of the N particles as

$$H_\mathrm{G} = -k_\mathrm{B} \int \mathrm{d}\tau \rho_N \log \rho_N \tag{6}$$

where ρ_N is the probability density function (pdf) for the N-particle system, and $\mathrm{d}\tau$ represents phase–volume elements in the full phase space. Although Gibbs did not explain how or why he took the logarithm of the probability of a system linearly proportional to the energy and number of particles, thereby giving rise to his statistical analogues, he did state that the expected value of $\log \rho$ was a minimum. This then suggests that the formalisms of Gibbs and of Jaynes (1957a,b) are essentially the same.

With a little reinterpretation of the Boltzmann distribution function, the Boltzmann entropy can be expressed as

$$H_\mathrm{B} = -k_\mathrm{B}N \int \mathrm{d}\tau_1 \rho_1 \log \rho_1 \tag{7}$$

where ρ_1 is the single-particle probability density, and $d\tau_1$ stands for the phase-volume element in the space of one element. Thus, Boltzmann established a connection between the variable of state, 'entropy', derived from phenomenological considerations, and the amount of chaos (or disorder) of a system:

$$H_B = k_B \ln W \tag{8}$$

where H_B is the Boltzmann entropy of the system, W is the number of macroscopically indistinguishable microscopic configurations or the number of ways to realize a given macrostate and k_B is a positive constant. Here, chaos means the number of microstates that have the same macroscopic properties. This equation was generalized by Gibbs. The analogy of Gibbs' formula was introduced by von Neumann in quantum mechanics:

$$H_B(\rho) = -k_B T_r \rho \ln \rho \tag{9}$$

where ρ is the density matrix characterizing the system state, k_B is the Boltzmann constant, and T_r stands for integration over all phase space. In a system with W different microstates, each of them occurring with the same probability, the density matrix of this system is $\rho = (1/W)p$, p being a W-dimensional projection.

In the Gibbs definition, there is always one system with N particles in it, not N systems each with one particle. Jaynes (1965) has shown that the Gibbs entropy, when maximized (i.e. for the canonical ensemble), can be identified with the thermodynamic entropy defined by Clausius. In other words, the Gibbs entropy of the canonical ensemble is numerically equal to the experimental entropy defined by Clausius. More generally, H_G is defined for all probability distributions, not just the canonical ensemble, therefore,

$$H_G \leqslant H_E \tag{10}$$

with equality if and only if the distribution ρ_N is canonical.

The expression for the change of the Boltzmann entropy shows that it ignores both the internal energy and the effect of the interparticle forces on the pressure. This is because it is defined in terms of the single-particle distribution. For the case of a perfect gas only, the Boltzmann entropy equals the Clausius experimental entropy and also the maximized Gibbs entropy. In general, the Boltzmann entropy is not equal to the experimental entropy.

For an isolated system the experimental entropy can also increase, that is,

$$\Delta H_E \geqslant 0 \tag{11}$$

with equality only if changes are reversible. The Gibbs entropy is actually a constant of the motion and the constancy of the Gibbs theoretical entropy is needed to prove the second law of thermodynamics.

The Gibbs algorithm yields uncertainty of our predictions, not the observed temporal fluctuation. Let a quantity of interest be f. We use the Gibbs algorithm to set up an equilibrium ensemble. The ensemble average of f is $\langle f \rangle$ and its variance $(\Delta f)^2 = \langle (f - \langle f \rangle)^2 \rangle$. Then Δf represents our degree of ignorance or uncertainty about the quantity f expressed by the ensemble, not the temporal fluctuation of f. To show that, we define a long-term time average as

$$\bar{f} = \frac{1}{T} \int_0^T f(t)\, dt \tag{12}$$

and a long-term variance as

$$(\delta f)^2 = \frac{1}{T} \int_0^T (f(t) - \bar{f})^2\, dt = \bar{f}^2 - (\bar{f})^2 \tag{13}$$

From ensemble averages, we can show $\langle f \rangle = \langle \bar{f} \rangle$. However,

$$\langle (\delta f)^2 \rangle = (\Delta f)^2 + (\Delta \bar{f})^2 \tag{14}$$

The second term is not necessarily zero. If a time average is computed over a short time interval, then the observed variation in f can be less than the Δf of the equilibrium ensemble. However, the long-term variation of f can actually be greater than Δf, depending upon a particular property of the probability density function (pdf) of the ensemble.

Information — theoretical entropy

To understand the informational aspect of entropy we consider a set consisting of n events: $\{E_1, E_2, \ldots, E_n\}$. We view uncertainty as a situation where we do not know which one of 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 n is a piece of information and the number of those events with non-zero probability is another piece of information. The probability distribution on $\{E_1, E_2, \ldots, E_n\}$, if known, provides a certain amount of information that might reduce the amount of uncertainty. The uncertainty can be quantified with entropy taking into account all different kinds of available information. This then suggests that the entropy measure is conditional, the conditions being in the form of information about the possible outcomes of the experiment.

We consider a message consisting of words or digits. The message can be represented as a sequence of binary digits. If the length of a 'word' is n, then n digits are needed to characterize it. The set E_n of all words of length n contains 2^n elements. Therefore, the amount of information needed to characterize one element of it is $\log_2$ of the number of elements of $E_n = \log_2 N$, with $N = 2^n$. This means that the amount of information needed to characterize an element of any set of power N (not necessarily of the form 2^n) is $\log_2 N$. Let E be a union of $E_1 \cup \ldots \cup E_K$ of pairwise disjointed sets, N_i = number of elements of E_i. Let $p_i = N_i/N$, $N = \Sigma N_i$. If it is known that an element of E belongs to some E_i, then the amount of additional information needed to completely determine it is $\log_2 N_i$. Thus, the average amount of information needed to determine an element, provided its affiliation to a particular E_i is known, is

$$\Sigma (N_i/N) \log_2 N_i = \Sigma p_i \log_2 (N p_i) = \Sigma p_i \log_2 p_i + \log_2 N \tag{15a}$$

This shows that the information equal to $\log_2 N$ is needed if it is not known as to which E_i a given element belongs to. Therefore, the corresponding lack of information is

$$H = -\Sigma p_i \log_2 p_i \tag{15b}$$

which is the Shannon entropy (1948a, b). Thus, H expresses a measure of the amount of uncertainty represented by the probability distribution. By interpreting E as a set of N measurements and p_i as the probability of finding the system in the ith microstate, the Shannon entropy equals the statistical mechanical entropy. Tribus (1961a, b) demonstrated that all of the laws of classical thermodynamics and, in particular, the concept of heat and temperature could be defined from the Shannon entropy using the Jaynes principle of maximum entropy (POME).

Thus entropy is a measure of the amount of chaos, or of the lack of information about a system. If complete information is available, i.e. if there is a pure state, entropy = 0. Otherwise it is greater than zero, and the bigger the value the more microstates and the smaller the statistical weight. $H(\rho) \geqslant \ln(1/p_1)$, p_1 being the biggest eigenvalue (= operator norm) of ρ. The entropy is a measure of our ignorance about a system, described by a density matrix, or, in a classical sense, by a probability distribution.

In communication theory, entropy is a measure of information. It implies freedom of choice to select between alternative messages. This concept is gainfully exploited by Pierce (1978) in his examination of the relation between information theory and search theory. The amount of search effort plays a critical role in this relation. For large values of this effort the optimal search policy produces the greatest information gain.

There is, however, a lack of consensus as to the meaningfulness of entropy. The implication that entropy is a measure of the meaningfulness of a particular decomposition of a set of values is erroneous. Shannon and Weaver (1949) assumed that all communication errors carried an equal penalty. Marschak (1971) argues that this assumption may be the cause of misunderstandings about the meaning of the entropy measure. Given the equal penalty assumption, entropy does measure the information value of a system. However, a measure based on the probability distribution cannot determine the information value of a system when unequal penalties are assigned to the various possible errors. According to Marschak, the value of an information system, or informativeness, can decrease even with an increase in information transmitted. Entropy does not measure all of the relevant aspects of information even in the sense of uncertainty as to the message transmitted. This weakens the argument for using entropy as a measure of information loss resulting from aggregation of records.

Fama (1965) and Theil and Leehders (1965), amongst others, used entropy to evaluate the predictive ability of forecasters, using information inaccuracy defined as $\Sigma\, q_i \log(q_i/p_i)$, where q_i is the fraction of values for which outcome i occurs and p_i is the fraction for which outcome i is forecast. Information inaccuracy equals the reduction in uncertainty from knowing the outcome as opposed to the forecast. In this case, all forecast errors may not carry an equal penalty as in the case of security prices. Then, the measure fails to distinguish between inaccuracy resulting from an overly optimistic forecast and that from an overly pessimistic forecast.

A similar criticism applies to the use of entropy as a measure of decentralization in an organization (Murphy and Hasenjaeger, 1973). However, interpretation of entropy in the distribution of tasks and decisions within an organization as a measure of disorder or decentralization in the decision making process is of great value. The decomposition property of the entropy measure permits insights into the nature of entropy decentralization in a hierarchical system. This decision making ability may be attributed to entropy within the informal decision units, entropy within a supremal unit or entropy between and at the various decision making levels themselves.

Principle of maximum entropy

Jaynes (1957a, b; 1982) formulated the principle of maximum entropy (POME), a full account of which is presented in a treatise by Levine and Tribus (1978), Kapur (1983a,b), Kapur and Kesavan (1992) and in Rosenkrantz (1983). 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 Shannon's interpretation of entropy as an information measure, the POME-based distribution is favoured over those with less entropy among those that satisfy experimentally 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 non-committal with regard to missing information and does not require invocation of ergodic hypotheses. The POME formalism is simpler and applies equally well to reversible and irreversible processes.

POME is a variational principle for assignment of probabilities under certain types of constraints, called testable information. Such constraints correspond to the probability distribution directly. A constraint is some piece of information that results in modifying the probability distribution. A physical constraint can be defined as any physical influence that exerts a systematic tendency on the outcome of an experiment. For example, for a discrete pdf, $\{p_i\}$, the ensemble average $\langle x\rangle \equiv \Sigma x_i p_i$ of a quantity x constitutes testable information. According to POME, the probabilities should be assigned by maximizing the entropy

$$H = -\Sigma\, p_i \log \frac{p_i}{m} \tag{16a}$$

under the constraint $\Sigma_i p_i = 1$ and $\langle x\rangle = x_0$, where $\{m_i\}$ is a suitable measure over the space of possibilities (hypothesis space).

Indeed, most frequency distributions produced in real experiments are maximum entropy distributions. When a POME-based distribution departs statistically significantly from an experimental one, it provides conclusive evidence of the existence of new constraints that were not taken into account in the calculation. This points to the further strength of POME, that it provides the most efficient procedure by which, if unknown constraints exist, they can be discovered. In this manner, POME brings out the physics by pointing to the existence of other constraints than those already used.

Evans (1978) proposed an approach for determining constraints in the maximum entropy formalism with use of the concept of essergy. The term essergy is interpreted to mean the essential aspect or essence of energy, or the energy in the form essential for the production of power, and is defined as

$$E_i = K \ln \frac{P_i}{P_{i0}} \tag{16b}$$

and therefore the expected value of essergy is given by

$$\langle E \rangle = K \sum_{i=1}^{n} P_i \ln \frac{P_i}{P_{i0}} \tag{16c}$$

where K is a unit conversion constant (with units of energy in most thermodynamic applications), P_i is the probability of outcome i conditional upon the knowledge of the event and the reference level of the background knowledge ('0') and P_{i0} is the probability of outcome i based upon reference level of background knowledge. Evans (1978) postulated that the actual probabilities $\{P_i\}$ are the ones for which the essergy $\langle E \rangle / K$ is a minimum subject to the constraints that characterize the system and the reference level. This postulate permits determining probabilities $\{P_i\}$ in a system where the constraints are non-linear in $\{P_i\}$ where POME does not apply.

Evans (1969) has discussed how to choose m. If there is no additional information other than normalization, POME assigns equal probabilities to all possible events, in accordance with Bernoulli's principle of insufficient reason [or principle of indifference (Keyes, 1921)]. Shore and Johnson (1980), Gull and Skilling (1984) and Gull (1991), among others, argue that POME is the only variational principle for the assignment of probability distributions, once the hypothesis space has been defined. Tikochinsky *et al.* (1984a,b) reached the same conclusion. Jaynes (1982) has justified it in numerous other ways. Kesavan and Kapur (1989) generalized POME.

Tikochinsky *et al.* (1984b) discussed three independent but complementary ways for deriving the form of the distribution based on two assumptions: (1) the experiment can be repeated a finite number of times, and (2) the theoretical expectation values of the relevant observables are estimated from their measured sampled averages. All three led to a unique distribution which was identical with the one derived by use of POME. Evans (1969) and Snickers and Weibull (1977), amongst others, renamed POME as the principle of minimum information (POMI).

At a philosophical level, building models for systems is ontology; the study of dynamics. Our knowledge about the state of the system is epistemology. Thus, statistical mechanics works at the level of inference. As hydrologists, it is our task to choose an appropriate space of possibilities, enumerate the possible states of the system and investigate its dynamics. As an example, the Gibbs ensemble represents the probability that our N-particle system is in a particular microstate. We make inferences about the state of a system, given incomplete information. We know the values of the macroscopic variables, but there are many microstates compatible with this macrostate. We set up a probability distribution (of the ensemble) using POME and the available constraints.

Principle of minimum cross-entropy

Kullback and Leibler (1951) formulated the principle of minimum cross entropy (POMCE) which can be expressed as follows. Suppose we guess a probability distribution for a random variable x as

$Q = \{q_1, q_2, \ldots, q_n\}$ based on intuition, experience or theory. This constitutes the prior information in the form of a prior distribution. To verify our guess we take some observations $x = (x_1, x_2 \ldots, x_n)$ and compute some moments of the distribution. To derive the distribution $P = \{p_1, p_2, \ldots, p_n\}$ of x we take all the given information and make the distribution as near to our intuition and experience as possible. Thus, POMCE is expressed as where the cross-entropy D is minimized. If no *a priori* distribution is available and if, according to Laplace's principle of insufficient reason, Q is chosen to be a uniform distribution U then

$$D(P, Q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \tag{17a}$$

$$D(P, U) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{1/n} = \ln n - \left(- \sum_{i=1}^{n} p_i \ln p_i \right) \tag{17b}$$

Hence, minimizing $D(P, U)$ is equivalent to maximizing the Shannon entropy. Because D is a convex function, its local minimum is its global minimum. Thus, a posterior distribution P is obtained by combining a prior Q with the specified constraints. The distribution P minimizes the cross (or relative) entropy with respect to Q, defined by equation (17a), where the entropy of Q is defined as Equation (16a). Cross-entropy minimization results asymptotically from Bayes' theorem. Its detailed treatment is found in Smith and Grandy (1995).

Shore and Johnson (1980) provide an axiomatic derivation of POMCE. van Campenhout and Cover (1981) show that the conditional distribution of a given random variable x is the (normalized) product of the POME distribution and the initial distribution. If the initial distribution is uniform, then the conditional distribution is the POME distribution. Kesavan and Kapur (1989) generalized POMCE. Campbell (1970) established an equivalence between POMCE and 'Gauss' principle. POMCE is related to Bayes' theorem, since the former yields a canonical distribution while the latter yields the Darwin–Fowler method used for justifying the canonical distribution.

POMCE can be employed for hypothesis testing as explained by Jaynes (1978). Let us consider a null hypothesis H_0 corresponding to a given distribution $Q = \{q_i\}$ obtained experimentally. The fitted distribution is found to be $P = \{p_i\}$. The objective is to test the null hypothesis if the fitted distribution is sufficiently close to the observed distribution. A widely used test is the χ^2 test defined as

$$\chi^2 = N \sum_{i=1}^{N} \frac{(q_i - p_i)^2}{q_i} \tag{18a}$$

when the value of χ^2 exceeds a certain critical value at a chosen significance level, the null hypothesis is rejected. The critical value is independent of N. Indeed, a preferred method of hypothesis testing is the use of $D(P, Q)$ of Equation (17a):

$$D(P, Q) = N \sum_{i=1}^{N} p_i \ln(p_i/q_i) \tag{18b}$$

which can be justified by Bayes' theorem. For small deviations $[|\, q_i - p_i \,| = 0(N^{-1/2})]$, one can show that

$$D(P, Q) = \frac{1}{2} \chi^2 \tag{18c}$$

The rejection of the null hypothesis is evidence of a new physical constraint and the nature of the deviation gives a clue as to the nature of the new constraint.

Algorithmic randomness and entropy

In the definitions of the Boltzmann, Gibbs and Shannon entropies the concept of ensemble is invoked. Thus, a probabilistic formalism is employed in which the entropy of a definite, completely known microstate

is always zero, and where the membership of this microstate in an ensemble defines the set of probabilities (the density function) used to calculate the entropy of the system. Both Boltzmann and Gibbs entropies selectively discard part of the information about the physical system which justifies introduction of the probabilistic description. The Shannon entropy is a measure of missing information or ignorance. Zurek (1989) proposed a new definition of entropy which does not use this ensemble strategy and which can be applied to the individual microstates of the system. This entropy is based on the definition of randomness. Definition of entropy as algorithmic information content is based on the intuitive notion of what a random number or random configuration is.

Random means difficult to describe or to reproduce. Algorithmic information content, algorithmic randomness, algorithmic entropy or sometimes algorithmic complexity, $K(s)$, of a binary string s is defined as the length, in the number of digits, of the shortest programme for a universal computer that can produce the state in question:

$$K(s) \simeq |\, s \,| \tag{19}$$

When s is interpreted as a binary representation of an integer, Equation (19) implies $K(s) \simeq \log_2(s)$. This measure of randomness bypasses the concept of probability and can be used to define the entropy of a completely specified microscopic state of an individual dynamical system. Algorithmic randomness is given by the shortest programme that generates the description of the state in question as an output.

Zurek (1989) suggested a new definition of physical entropy including both of its complementary contributions — randomness and ignorance. He defined physical entropy, S_d, as the sum of the missing information and of the length of the most concise record expressing the information already at hand:

$$S_d = H_d + K(d) \tag{20}$$

where H_d is the information about the actual microstate which is still missing in spite of the availability of the relevant data d as measured by the Shannon entropy, and $K(d)$ is the algorithmic randomness given by the size of the most concise description of the already available relevant data. This definition is appealing in hydrology but has not been applied to date.

General representation of entropy

Several attempts have been made to provide, possibly, more general measures of uncertainty along three main lines (Forte, 1984): (1) representation of entropy in terms of the probabilities and some other parameters that contain the Shannon entropy as a special case such as Renyi's entropy of order α, entropies of order, α, β, etc. (Aczel and Daroczy, 1975); (2) entropies that depend only on the possible outcomes of the experiment, eventually through their information content without probabilities (Forte, 1969, 1972); and (3) entropies that depend explicitly on the outcome and/or functions of the outcome (random variables, random vectors) (Forte and Cictleo, 1979).

Consider a set of events $\{E_1, E_2, \ldots, E_n\}$. If a function of probabilities of the events provides the amount of uncertainty then the probabilities are presumably given. The uncertainty concerns which one of the events occurs, knowing the probabilities $\{p_1, p_2, \ldots, p_n\}$ where p_i is associated with E_i. Then one can reason that information is carried not only by the numbers p_i but also by the fact that each p_i is associated with a specific event E_i. Thus, the corresponding measure of uncertainty can be represented by a functional I_n $\{E_1, E_2, \ldots, E_n; p_1, p_2, \ldots, p_n\}$ of the events and their probabilities giving rise to a general form of entropy (Aczel, 1984):

$$I_n = -a \sum_{i=1}^{n} p_i \ln p_i + \sum_{i=1}^{n} p_i g(E_i) \tag{21}$$

where a is an arbitrary real constant, and g is an arbitrary real-valued function of the events. The function I_n is not necessarily symmetric in the probabilities but is in the couples (E_i, p_i), $i = 1, 2, \ldots, n$. It measures the amount of uncertainty associated with the random variable, ignoring its range.

If the probabilities are given but their distribution on $\{E_1, E_2, \ldots, E_n\}$ is not, then the corresponding entropy is symmetric and is the Shannon entropy. On the other hand, if the only information on the events $\{E_1, E_2, \ldots, E_n\}$ is the total number of those with non-zero probability, then the corresponding entropy does not depend on $\{p_1, p_2, \ldots, p_n\}$, but should contain the information, i.e. the cardinality m of the set of non-zero probabilities. Harley's entropy ($\log m$) is one of these entropies. In the event, no information is given at all about the probabilities or any other quantity associated with the events, then the corresponding entropy depends only on the partition, i.e. $I_n = I_n\{E_1, E_2, \ldots, E_n\}$. These entropies are those without probabilities. Aczel (1984) surveyed generalized information measures.

Non-probabilistic entropy

De Luca and Termini (1972) defined a non-probabilistic entropy in the setting of fuzzy sets entropy. This entropy is a global measure of the indefiniteness connected with the situation under consideration and can be regarded as a measure of a quantity of information not necessarily related to random experiments. This is useful in situations where indefiniteness arises more from a sort of intrinsic ambiguity rather than from a statistical variation. Thus, entropy is taken to provide a measure of the degree of fuzziness of the situation. It can also be regarded as an average intrinsic information which is received when making a decision. For a fuzzy set f, the degree of fuzziness, called entropy, $d(f)$, can be expressed as

$$d(f) = H(f) + H(\bar{f}) \tag{22}$$

where $\bar{f}$, the complement of f (not the algebraic complement), is defined point by point as

$$\bar{f}(x) = 1 - f(x) \tag{23}$$

The quantity $d(f)$ depends only on the values assumed by f and satisfies the following properties: (1) $d(f) = 0$ if, and only if, f takes on the values of 0 or 1; (2) $d(f)$ is maximum if, and only if, $f = (1/2)$; (3) $d(f) \geqslant d(f^*)$ where f^* is any fuzzy set such that $f^*(x) \geqslant f(x)$ if $f(x) \geqslant (1/2)$ and $f^*(x) \leqslant f(x)$ if $f(x) \leqslant (1/2)$. From Equation (22), one notes that $d(f) = \mathrm{d}(\bar{f})$.

The function $H(f)$ is similar to the Shannon entropy but is conceptually quite different:

$$H(f) = -K \sum_{i=1}^{N} f(x_i) \ln f(x_i) \tag{24}$$

where N is the number of elements of the set I, and K is a positive constant. Thus, $d(f)$ can be written using the Shannon entropy function $S(x)$

$$S(x) = -x \ln x - (1 - x)\ln(1 - x)$$

as

$$d(f) = K \sum_{i=1}^{N} S[f(x_i)] \tag{25}$$

$d(f)$ satisfies the first two properties.

In a random experiment, the total entropy can be expressed as

$$H_t = H(p_1, p_2, \ldots, p_N) + m(f, p_1, p_2, \ldots, p_N) \tag{26}$$

which describes the total uncertainty that we have in making a provision about the element of I which will occur as a result of the random experiment and in taking a decision about the value 1 or 0 which has to be attached to the element itself. If $m = 0$, which occurs in the absence of fuzziness, then H_t reduces to the Shannon entropy. If $H\{p_1, p_2, \ldots, p_n\} = 0$, which means there is no random experiment and only a fixed element, say, x_i will occur, then $H_t = S(f(x_i))$.

When a problem is not well defined, the Shannon entropy is not a sufficient measure, and another entropy is sought that is a measure of imprecision. When added to the information entropy, a total measure, called combined entropy, of both uncertainty and imprecision is the result. The fuzzy entropy $d[f(x_1), f(x_2), \ldots, f(x_n)]$, where $x = (x_1, x_2, \ldots, x_n)$ in an imprecise manner with fuzzy support $f(x_i)$, defines the average imprecision given by Equation (25). Imprecision occurs when it is not clear where x can be recognized as a member of $\{x_i\}$ or not. Complete ambiguity is recognized by $f(x_i) = (1/2)$, where $S(f(x_i)) = S_i$ is a maximum. $S(f(x_i)) = 0$ if, and only if, $f = 0$ or $f = 1$. A vagueness that leans towards recognition of x as a member of $\{x_i\}$ has a value of $(1/2) < f(x_i) \leqslant 1$, and S_i is decreased.

In a fuzzy situation, it is seldom clear if x is not a member of $\{x_i\}$. When x occurs it is not clear as to which x_i $(i = 1, 2, \ldots, n)$ it should be assigned. This fuzziness is recognized by the various $f(x_i)$, $0 \leqslant f(x_i) \leqslant 1$. The average measure of this fuzziness or imprecision is given by Equation (25).

Thus, in a practical problem, two types of information exist. The first is associated with the countable results of experiments and can be determined using the Shannon entropy. This is the objective information. The second is associated with professional wisdom, professional judgment, understanding and experience, and can be seldom expressed functionally. This type of information is termed subjective. A measure of this subjective impression is given by the fuzzy entropy. Thus, the combined entropy $F(p,f) =$ total entropy H_t is

$$F(p,f) = -\sum_{i=1}^{n} p_i \ln p_i + D\sum_{i=1}^{n} p_i S_i[f(x_i)] \tag{27}$$

in which D is a multiplier and needs to be determined. $F(p,f)$ satisfies a set of desiderata: (1) F is continuous in p_i and $f(x_i)$; (2) $F(p,f) = F(p) + F(f\,|\,p)$ where $F(f\,|\,p) =$ conditional entropy given p; (3) if $p_i = 1$, then $F(p) = 0$, and $F(p,f) = S_i(f(x_i))$; (4) if all $f(x_i) = 0$ or 1, then $F(p,f) = H$.

As a result of the combined entropy, the probability distribution needs to be revised (Brown, 1980). Let this distribution be $\bar{p}_i$ in place of p_i. Then the construction of $\bar{p}_i$ is such that

$$-\sum_{i=1}^{n} \bar{p}_i \ln \bar{p}_i = F(p,f) \tag{28}$$

and the criterion for multiplier, D, could be that of all $f(x_i) = (1/2)$ then $\bar{p}_i = (1/n)$. Then

$$D = -1{\cdot}4427\left(\ln \frac{1}{n} + H\right) \tag{29}$$

and

$$F(p,f) = -\sum_{i=1}^{n} p_i \ln p_i - 1{\cdot}4427\left(\ln \frac{1}{n} - \sum_{i=1}^{n} p_i \ln p_i\right)\sum_{i=1}^{n} p_i S_i \tag{30a}$$

Determination of D in this manner leads to the maximum combined entropy (including uncertainty and imprecision) when all $f(x_i) = (1/2)$ and all $p_i = (1/n)$. For $H < F < F_{\max}$ one or more $f(x_i) > (1/2)$ and the distribution of p_i is more peaked. At $F = H$, all $f(x_i) = 1$ and $p_i = p_i$. Therefore, the criterion for D limits $f(x_i)$ to $(1/2) \leqslant f(x_i) \leqslant 1$ and p_i to lie inclusively between $(1/n)$ and p_i.

In principle, p_i may be estimated from Equations (28) and (30a). The acceptable p_i lies between p_i and (1/2). The probability of x lying between θ_1 and θ_2 $(\theta_2 > \theta_1)$ is obtained by

$$\bar{p}(\theta_1 \geqslant x \geqslant \theta_2) = \bar{p}(x \geqslant \theta_1) - \bar{p}(x \geqslant \theta_2) \tag{30b}$$

where θ_1 and θ_2 are any arbitrary values.

Properties of entropies

For a discrete stochastic process generating signals s_i ($i = 1, 2, \ldots, n$) at random with probabilities p_i ($i = 1, 2, \ldots, n$) the entropy measure

$$H = -\sum_{i=1}^{n} p_i \log p_i \tag{31a}$$

is simply another measure of dispersion, which can be related to the moments of the probability function. Its virtues stem from its decomposition and interpretation as a measure of information, freedom of choice, disorder or uncertainty (Horowitz and Horowitz, 1976).

The most important properties of the entropy measure are: (1) any tendency toward equalization of the p_i for a given n increases H_i; (2) when p_i are all equal, the greater the value of n is the greater the value of H; (3) $0 \leqslant H \leqslant \log_2 n$; (4) the measure is defined for conditional probabilities, including posterior probabilities determined by Bayes' theorem; and (5) entropy can be decomposed into a sum of 'between' groups and 'within' groups entropies.

If s_i itself sets into motion a stochastic process that generates other 'signals', t_{ij}, with probabilities q_{ij} ($j = 1, 2, \ldots, J_i$), then the entropy of the latter process is defined by

$$H_i = -\sum_{j=1}^{J_i} q_{ij} \log_2 q_{ij} \tag{31b}$$

The entropy of the total process can be written as

$$H = -\Sigma p_i \log p_i + \Sigma p_i H_i = H_B + H_w \tag{31c}$$

where H_B is the entropy 'between' groups, H_i is the entropy 'within' the ith group, and H_w is the expected 'within' group entropy. Therefore, we can explain the overall entropy in terms of the entropy within and between variously grouped components.

ENTROPY AND DATA ANALYSIS

The relative entropy $H/\log n$, redundancy $\log n - H$ and conditional and joint entropies employing conditional and joint probabilities have been variously interpreted as measuring freedom of choice, uncertainty and disorder, information content or information gain or loss. In the analysis of empirical data, entropy is another measure of dispersion, an alternative to the variance σ^2, which has also been interpreted as a measure of uncertainty and as revealing gain or loss in information. 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. The argument for favouring an entropy measure over the alternatives must be based on either the desire to satisfy the theoretical properties discussed above or on the grounds that the data are amenable to the calculation of the entropy measures.

Hart (1971) compared entropy to the classical statistical measures of dispersion. He argues that when the sample size is sufficiently large to justify the use of a statistical distribution it is preferable to use the parameters of the underlying distribution. Entropy and the entropy-based measures can be obtained directly from these parameters. He reasons that variance is the appropriate measure of concentration because: (1) other measures are governed by the variance; (2) the variance can be decomposed into variances between and within groups or sets of data; and (3) classical statistical tests can be used to test for statistically significant changes in the variance. 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, that it lends itself to decomposition and that it takes into consideration small values with appropriately small

weights. Philippatos and Wilson (1972) suggested the use of entropy rather than the variance as a measure of risk of a security portfolio whose components yielded stochastic returns.

Relation between entropy and variance

Mukherjee and Ratnaparkhi (1986) investigated the functional relationship between entropy H and variance σ^2 for some well-known distributions, including normal and generalized normal, gamma and generalized gamma, exponential, uniform, Weibull, extreme-value type 1, Poisson, logistic, Pareto and several others. Since entropy is a measure of uncertainty or chaos, and variance is a measure of variability, the connection between them is of interest to investigate the stochastic nature of a phenomenon.

In general, an explicit relation between H and σ^2 does not exist. However, for certain distributions which can be appropriately reparameterized, H can be expressed as a function of σ^2. Under the restriction that the variance is known or fixed, this functional relationship can be compared for different distributions. The entropies of the aforementioned distributions are convex functions of variance, and the normal distribution possesses the highest entropy followed by logistic, Laplace and extreme-value distributions.

Mukherjee and Ratnaparkhi (1986) provided an entropy-based measure of affinity or closeness between different distributions under the assumption of distributions having common variance. The affinity between two distributions with density functions f_1 and f_2, and entropies of H_1 and H_2, was defined as

$$A(f_1, f_2) = |H_1 - H_2| \tag{32a}$$

This can also cast as

$$A(f_1, f_2) = |-E[\log(f_1/f_2)]| = |E[\log(f_2/f_1)]| \tag{32b}$$

which is the expectation of the likelihood ratio. This measure differs from Kullback's minimum distance information criterion. They also defined a similarity function as

$$S(f_i, f_j) = 1 - \frac{A(f_i, f_j)}{\max[A(f_i, f_j)]} \tag{32c}$$

for any i and j, and $i \neq j$. Thus, affinity (distance) is a monotonically decreasing function of similarity. The similarity factor can be used to cluster or group models. Grouping of distributions suggests that entropy implies a concentration of density towards the location followed by long tails.

Entropy and correlation

Linfoot (1957) defined the informational coefficient of correlation r_0, $0 \leqslant r_0 \leqslant 1$, as

$$r_0 = [1 - \exp(-2T_0)]^{0.5} \tag{33}$$

where T_0 is the logarithmic index of correlation or transinformation defined as

$$T_0 = \sum_{ij} (p_{ij} \log p_{ij} - p_{ij} \log p_i q_i) \tag{34a}$$

in which p_i is the probability of $x = x_i \{i = 1, 2, \ldots, n)$; q_i is the probability of $y = y_i$; p_{ij} is the joint probability of $x = x_i$ and $y = y_j$; $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$;

$$p_i = \sum_j p_{ij} \tag{34b}$$

and

$$q_j = \sum_i p_{ij} \tag{34c}$$

for discrete variables, and

$$T_0 = \int\int [\, p(x, y)\log\, p(x, y) - p(x, y)\log[\, p(x)q(y)]]\, \mathrm{d}x\, \mathrm{d}y \tag{34d}$$

for continuous variables where $p(x)$ and $q(y)$ are the marginal probability density distributions of x and y, respectively. Equation (34d) can be cast as

$$T_0 = \int\int p(x, y)\log\left(\frac{p(x, y)}{p(x)q(y)}\right) \mathrm{d}x\, \mathrm{d}y \tag{34e}$$

T_0, the upper limit of common information between two processes, represents the level of dependence (or association) between the variables x and y. If x and y are statistically independent then $T_0 = 0$ and $r_0 = 0$. For complete dependence between x and y, $r_0 = 1$. T_0 represents the upper limit of transferable information between x and y, and its measure is given by r_0. The ordinary correlation coefficient r measures the amount of information transferred between variables under specified assumptions such as linearity and normality. By comparing with r_0, the goodness of information transfer by regression and that of the underlying assumptions can be inferred because of its interpretation in terms of the quantity of information. r_0 is a better measure of correlation than r. Also, r_0 has the advantage in that it is an informational measure of correlation which is a generalization of the ordinary correlation coefficient of a normal distribution. An inference similar to that of r can be drawn by defining the amount (in percent) of transferred information by the ratio T/T_0, where T can be computed in terms of ordinary r for an assumed type of regression equation.

As an example, consider a bivariate probability distribution (Whittaker and Robinson, 1944) given as

$$p(x, y) = \frac{1}{2\pi}(ab - h^2)\exp\left[-\frac{1}{2}(ax^2 + by^2 + 2hxy)\right] \tag{35}$$

where a, b and h are parameters; $a > 0$, $ab - h^2 > 0$. This type of distribution is of great value in modelling a variety of hydrological processes, such as rainfall. The classical correlation coefficient r for this case is found to be:

$$r = -\frac{h}{(ab)^{0.5}} \tag{36a}$$

The informational measures T_0 is found to be

$$T_0 = \frac{1}{2}\ln\frac{ab}{ab - h^2} \tag{36b}$$

Equation (36a) and (36b) show that

$$r = [1 - \exp(-2T_0)]^{0.5} \tag{36c}$$

which is similar to Equation (33). Indeed, the same result is obtained when $p(x, y)$ is given in a more general form. An important advantage of informational measures r_0 and T_0 is that they are invariant under a change

of parameterization, and r_0 reduces to r when $p(x, y)$ is a normal distribution. Harmancioglu *et al.* (1992a,b,c) employed T_0 and r_0 for investigating stream gauging and groundwater recharge.

Horibe (1985) defined a correlation coefficient ρ between two random variables x and y in terms of entropy. If the correlation distance between x and y is denoted by $d(x, y)$ then the correlation coefficient follows:

$$\rho(x, y) = 1 - d(x, y) \tag{37}$$

where

$$d(x, y) = H(x \mid y)/H(x), \quad \text{if} \quad H(x) \geqslant H(y) \tag{38a}$$

and

$$d(x, y) = H(y \mid x)/H(y), \quad \text{if} \quad H(x) \leqslant H(y) \tag{38b}$$

where $H(\cdot)$ is the marginal entropy (uncertainty) of $(\cdot)$, $H(x \mid y)$ is the conditional entropy which measures the uncertainty of x remaining after knowing y and $H(y \mid x)$ is the conditional entropy of y given x.

Equation (38) is an entropic form of ρ and admits an intuitive interpretation. If $\rho(x, y) = a$, and $H(x) \geqslant H(y)$, then $(1 - a)H(x) = H(x \mid y)$. This means that 100 percent of the information (uncertainty) $H(x)$ that x has, is gained or transmitted (respectively reduced) by knowing y. In a case involving several random variables, d or ρ gives a natural or information topology.

Kvalseth (1987) extended the work of Horibe (1985). Recalling the definition of transformation $T(x, y)$,

$$T(x, y) = H(x) - H(x \mid y) \textit{ or } T(y, x) = H(y) - H(y \mid x) \tag{39}$$

the informational association measure ρ restricted to the interval [0, 1] was defined as

$$\rho(x, y) = \frac{T(x, y)}{D} \tag{40}$$

where D is some approximate norming quantity with three alternative expressions:

$$D_1 = \min[H(x), H(y)] \tag{41}$$

$$D_2 = [H(x) + H(y)]/2 \tag{42}$$

$$D_3 = \max[H(x), H(y)] \tag{43}$$

For each of the three D_i values, $T/D_i = 0$ $(i = 1, 2, 3)$ when x and y are independent and $T/D_i \leqslant 1$, since

$$T \leqslant D_1 \leqslant D_2 \leqslant D_3 \tag{44}$$

which follows from Equation (39) and the fact that $H(x \mid y) \geqslant 0$ and $H(y \mid x) \geqslant 0$. Kvalseth recommended D_2 for use. He also derived confidence intervals and performed hypothesis testing for the population ρ.

Rajski (1961) defined a function by means of entropy which is a distance in the set of discrete probability distributions. This distance measures the dependence between the transmitted and the received signals. This functional

$$d(x, y) = 1 - \frac{T(x, y)}{H(x, y)}, \quad H(x, y) \neq 0 \tag{45}$$

is a distance in the set x. $H(x, y)$ is the joint entropy of (x, y), and $T(x, y)$ is the full information. If $x = y$, then $d(x, y) = 0$, $H(x) = H(y) = H(x, y)$ and $T(x, y) = H(x) + H(y) - H(x, y)$.

CRITERIA FOR MODEL SELECTION

Akaike (1972) formulated a criterion, called Akaike information criterion (AIC), for selecting the best model from amongst several models. This criterion is derived from an analysis of the maximum likelihood estimate

(MLE) and yields an estimate of a measure of fit of the model. It also provides a method of model identification (Akaike, 1974, 1985). The information criterion, AIC, can be expressed as

$$AIC = -2 \log(\text{maximized likelihood}) + 2k \tag{46}$$

where k is the number of parameters θ used to find the best fit model, and the maximum likelihood value is obtained by maximizing the log-likelihood function for the model $f(x, \theta)$ with respect to the parameters θ. When there are several models specifying $f(x, \theta)$, 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. Turkman (1985) applied AIC to make a choice of extremal (Gumbel, Frechet and Weibull) models and analysed its effectiveness in making the choice.

Tribus (1969) and Tribus *et al.* (1964) employed entropy in hypothesis testing. With use of Bayes' theorem in logarithmic form, an evidence function was defined for comparing two hypotheses. The evidence in favour of a hypothesis over its competitor is the difference between the respective entropies of the competition and the hypothesis under test. Defining surprisal, η_k, as the negative of the logarithm of the probability, the surprisal function was expressed as

$$\eta_k = -\ln p(x_k) \tag{47}$$

where x_k is the kth observation in the set of observations. The mean surprisal η in the set of m observations was given by

$$\bar{\eta} = \frac{1}{m} \sum_{k=1}^{m} \eta_k \tag{48}$$

Therefore, the evidence function Q for two hypotheses 1 and 2 is obtained as

$$Q = \sum_{k=1}^{m} \eta_{k,1} - \sum_{k=1}^{m} \eta_{k,2} = m(\bar{\eta}_1 - \bar{\eta}_2) \tag{49}$$

Entropy appears through defining the probability distribution of x required in Equation (47).

RISK

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. Kaplan and Garrick (1981) presented an excellent treatment of risk. To put risk in proper perspective, it is useful to clarify the distinction between risk, uncertainty and hazard, as well as between probability and frequency and the relativity of risk.

Probability is a numerical measure of a state of knowledge, a degree of belief, a state of confidence — a number used to communicate a state of mind. It is thus subjective, changeable and not measurable. Frequency, on the other hand, corresponds to the outcome of an experiment involving repeated trials and is, thus, a hard (well-defined, objective) measurable number. Frequency is used to calibrate the probability scale. Furthermore, the study of frequency-type information is the subject of statistics, and probability the science of handling the lack of data. The outcome of an experiment can be predicted in the form of a probability curve $p(\phi)$ against frequency ϕ as a way of expressing our state of knowledge; $p(\phi)$ is the probability of frequency curve.

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. Symbolically,

$$\text{risk} = \text{uncertainty} + \text{damage} \tag{50}$$

Hazard (H) is commonly defined to be a source of danger and can be defined as a set of doublets: $H = \{(s_i, x_i)\}$, where s_i is a scenario identification (e.g. failure of a dam) and x_i is the consequence measure of that scenario or the measure of the damage. Randomness reflects the variability of the measurable results. Risk encompasses the likelihood of conversion of that source into actual delivery of loss, injury or some form of damage. Symbolically,

$$\text{risk} = \frac{\text{hazard}}{\text{safeguards}} \tag{51}$$

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. Kaplan and Garrick (1981) presented a first-level quantitative definition and then a second-level definition generalizing the former. For the first-level definition, risk R is defined as a set of triplets:

$$R = [(s_i, p_i, x_i)], \quad i = 1, 2, \ldots, n \tag{52}$$

where p_i is the probability of scenario s_i, and x_i is the consequence resulting from happening of s_i. Thus, one can say that risk is probability and consequence. This leads to the arising of a risk curve encompassing multiple scenarios and consequences. Indeed, a family of curves may be needed to communicate the idea of risk. In order to be able to include all scenarios not otherwise included in the earlier definition one may modify Equation (52) as

$$R = [(s_i, p_i, x_i)], \quad i = 1, 2, \ldots, n + 1 \tag{53}$$

Thus, this includes all the scenarios that have been thought of and also an allowance for those not thought of.

In a frequency format, R can be expressed as a set of triplets:

$$R = [(s_i, \phi_i, x_i)], \quad i = 1, 2, \ldots, n + 1 \tag{54}$$

where ϕ is the frequency of occurrence with the cumulative frequency

$$\Phi_i = \sum_{x_j \geqslant x_i} \phi_i \tag{55}$$

where the sum is over all scenarios having damage equal to or greater than x_i.

However, the degree of uncertainty depends upon our total state of knowledge as of right now; and upon all the evidence, data and experience with similar courses of action in the past. There is uncertainty about the curve $\Phi(x)$ which can be expressed by embedding this curve in a space of curves and constructing a probability distribution over this space. This gives rise to a family of curves $\Phi_p(x)$, with the parameter being the cumulative probability, which is the pictorial form of level 2 definition of risk. Thus, we have a set of triplets in the form:

$$R = [(s_i, p_i(\phi_i), \mathrm{x}_i)] \tag{56a}$$

which includes uncertainty in frequency. In a similar vein, if there is uncertainty in the damage then

$$R = [(s_i, p_i(\phi_i), \xi_i(x_i))] \tag{56b}$$

or more generally,

$$R = [(s_i, p_i(\phi_i, x_i))] \tag{56c}$$

using a joint distribution on ϕ_i and x_i. The level 2 definition of risk is absolute and objective and depends upon the evidence at hand.

To illustrate the concept of risk discussed above, we consider the scenario of a flooding event in an urban watershed m and wish to know the frequency ϕ_m of this event. Three types of information are available: (1) the background knowledge of drainage; (2) the past experience of flooding; and (3) experience of flooding in hydrologically similar watersheds. In our specific urban watershed we may have had k_m occurrences in T_m years. Similarly, this type of data would consist of a set of doublets: (k_i, T_i), $i = 1, 2, \ldots j$, giving the experience of flooding in urban watersheds deemed similar to watershed m. The three types of available information can be combined into a probability curve, $p(\phi_m \mid E)$ expressing our state of knowledge about ϕ_m by Bayes' theorem

$$p(\phi_m \mid E) = p(\phi_m)\left[\frac{P(E \mid \phi_m)}{P(E)}\right] \tag{57}$$

where $p(\phi_m \mid E)$ is the posterior probability to be assigned to ϕ_m after having the evidence E, $p(\phi_m)$ is the prior probability assigned to ϕ_m before learning the evidence E, $p(E \mid \phi_m)$ is the likelihood or the conditional probability that evidence E would be observed if the true frequency were actually ϕ_m and $p(E)$ is the prior probability of the evidence E.

The information types (1) and (3) are expressed in the prior $p(\phi_m)$. The watershed specific information type (2) constitutes the evidence and factors in the calculation through the likelihood function:

$$p(E \mid \phi_m) = \left[\frac{(\phi_m T_m)^{k_m}}{k_m!}\right]\exp[-\phi_m T_m] \tag{58a}$$

The prior $p(E)$ is then the sum, or integral, of the numerator in Equation (57):

$$p(E) = \int_0^\infty p(\phi_m)p(E \mid \phi_m)\, \mathrm{d}\phi_m \tag{58b}$$

This ensures that the normalization of $p(\phi_m \mid E)$ is correct. Entropy is invoked in deriving the prior distributions.

SAFETY

Safety has two components: reliability and probabilistic risk assessment (PRA) (Goodman, 1985). 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). The safe and unsafe domains are separated by a limit state surface. Here, uncertainty in failure criteria is taken into consideration.

In PRA, 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. Here, safe and unsafe domains are separated by a potentially unsafe domain which is tantamount to introducing fragility.

The concept of fragility is now introduced. Consider a hydraulic system with random parameters $y_1, y_2, \ldots, y_m$. The system can fail in many ways and every failure mode can be described with a corresponding failure indicator $g_i (i = 1, 2, \ldots, N$, N is the number of different failure modes) which is a

function of hazard parameters $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_m$. Thus,

$$g_i = g_i(x_1, x_2, \ldots, x_n;\ y_1, y_2, \ldots, y_m) \tag{59a}$$

For an earth dam, depending upon 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.

In the case of a one-mode failure system, failure will occur if the failure indicator reaches the limit state surface:

$$g(x_1, x_2, \ldots, x_n;\ y_1, y_2, \ldots, y_m) = 0 \tag{59b}$$

This surface distinguishes the safe domain ($g \leqslant 0$) from the unsafe domain ($g > 0$) in the space of random variables $x_1, x_2, \ldots, x_n; y_1, y_2, \ldots, y_m$. Implied herein is that a failure always occurs at the same value of the failure indicator g.

The random parameter variability has two levels: technical and physical. The technical variability can be detected by regular technical means, but the physical variability is beyond these means. Non-measurable differences in random parameters can lead to measurable variability in the failure indicator value corresponding to system failure. This is termed intrinsic uncertainty in the value of the failure indicator criterion. There is also extrinsic uncertainty in the criterion value owing to missing random parameters or neglect of variability of some parameters.

The space of random parameters can be divided into three domains:

(1) Safe domain, D_s (no failure)

$$g(x_1, x_2, \ldots, x_n;\ y_1, y_2, \ldots, y_m) \leqslant g_{\min} \tag{60a}$$

(2) Potentially unsafe domain, D_p (failure is possible)

$$g_{\min} \leqslant g(x_1, x_2, \ldots, x_n;\ y_1, y_2, \ldots, y_m) < g_{\max} \tag{60b}$$

(3) Unsafe (failure) domain, D_f (a failure is certain)

$$g(x_1, x_2, \ldots, x_n;\ y_1, y_2, \ldots, y_m) \geqslant g_{\max} \tag{60c}$$

Consider a probability density function of failure at a value of the failure indicator g as $f(g)$. A cumulative probability $F(g)$ then defines fragility as

$$F(g) = \int_{g_{\min}}^{g} f(g_*)\, \mathrm{d}g_* \tag{61a}$$

Therefore, fragility is a conditional probability of the failure at a failure indicator value g_* not exceeding g:

$$F(g) = P[g_* \leqslant g] \tag{61b}$$

The total probability of failure per year P_F can be expressed as

$$P_F = \int h(x_1, x_2, \ldots, x_n;\ y_1, y_2, \ldots, y_m) F\,[g(x_1, x_2, \ldots, x_n;\ y_1, y_2, \ldots, y_m)]\, \mathrm{d}x_1\, \mathrm{d}x_2 \ldots \mathrm{d}x_n\, \mathrm{d}y_1\, \mathrm{d}y_2 \ldots, \mathrm{d}y_m \tag{61c}$$

where $h(\ldots)$ is a joint distribution function of hazard and structural parameters. Or

$$P_F = \int_{-\infty}^{\infty} h(g)F(g)\,\mathrm{d}g \tag{62a}$$

Alternatively,

$$P_F = \int_{-\infty}^{\infty} H(g)f(g)\,\mathrm{d}g \tag{62b}$$

where $H(g)$ is a hazard function, meaning the probability of excedance of the failure indicator g:

$$H(g) = \int_{g}^{\infty} h(g_*)\,\mathrm{d}g_* \tag{62c}$$

The failure density function $f(g)$ determined from empirical data represents randomness, referred to as an R-distribution when data are limited. The parameters of $f(g)$ are uncertain and this uncertainty is represented with an uncertainty or U-distribution. The fewer the data available, the wider the U-distribution. If no data are available, fragility may be evaluated based on judgment. Corresponding distributions are called state-of-the knowledge distributions.

The probability of failure given by Equation (61c) can be presented in four ways: (1) M-format if the probability of failure is presented as a mixed distribution; (2) R-format if an expectation of all uncertainty parameters is estimated with a U-distribution and the probability of failure with an R-distribution; (3) U-format, if an expectation of all random parameters is estimated with an R-distribution and the probability of failure with a U-distribution; (4) T-format, if an expectation of all random and uncertainty parameters is estimated and the probability of failure is the total expectation.

The entropy theory can be used to quantify safety and fragility. In the absence of data, which is not uncommon in hydrology, state-of-knowledge distributions based on judgement are used. A good choice is the log-normal distribution or use of the distribution having a fragility curve confined between $g_{\min}$ and $g_{\max}$. POME offers an objective basis to select a correct state-of-knowledge distribution relevant to the problem in question. Thus, the state-of-knowledge distribution should be the one yielding the maximum entropy, given the knowledge. Since the Shannon entropy is invariant, any Edgeworth–Kapteyn (Korn and Korn, 1968) distribution given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\left[\frac{y(x)-\mu}{\sigma}\right]^2\right)\left|\frac{\mathrm{d}y}{\mathrm{d}x}\right| \tag{63}$$

maximizes Equation (16a) as well as the normal distribution with mean μ and standard deviation σ. The analytical form of the function $y(x)$ depends upon the additional information to be incorporated, such as the range, regularity, analyticity, single-valuedness, etc.: (1) $y(x)$ and its inverse $x(y)$ are single valued; (2) the range of $y(x)$ is from $-\infty$ to ∞; (3) $y(x)$ is analytic in its domain except at the end points; (4) $f(x)$ vanishes at the end-points of $y(x)$.

As an example, consider a random variable x, $x \geqslant 0$. A general expression for $y(x)$ is

$$y(x) = \ln x + \sum_{i=1}^{n} a_i x^i - \sum_{j=1}^{m} \frac{b_j}{x^j} \tag{64a}$$

where a_i and b_j are positive constants and the coefficient of $\ln x$ can be made equal to unity without loss of generality by changing the parameter σ. The above $y(x)$ has $n+m$ parameters. Thus, we can set $n+m$ moments and maximize entropy.

If the random variable x is confined to some finite interval a and b, $a \leqslant x \leqslant b$, then $y(x)$ can be expressed as

$$y(x) = \ln(x - a) - \beta \ln(b - x) - \sum_{i=1}^{n} \frac{A_i}{(x - a)^i} + \sum_{j=1}^{m} \frac{B_j}{(b - x)^j} \tag{64b}$$

where β, A_i and B_j are parameters and are positive.

To summarize, for given mean and standard deviation of x, $-\infty < x < \infty$, the state-of-knowledge distribution is the log-normal distribution. If $a \leqslant x \leqslant b$, and μ and σ are given, then the state-of-knowledge distribution is the modified log-normal distribution:

$$f(x) = \frac{b - a}{\sqrt{2\pi}\sigma(x - a)(x - b)} \exp\left(-\frac{1}{2}\left[\frac{\ln(x - a) - \ln(b - x) - \mu}{\sigma} \right]^2 \right) \tag{65a}$$

The next step is to construct fragility curves based on the assumed risk format. In hydrological modelling, one may employ the probability-of-frequency format. Two distributions are therefore needed: the R-distribution $f(g; \alpha_1, \alpha_2, \ldots, \alpha_m)$ and the U-distribution $\phi(\alpha_1, \alpha_2, \ldots, \alpha_m)$. Here, g is a failure indicator; $\alpha_1, \alpha_2, \ldots, \alpha_m$ are parameters of the distribution f; and ϕ reflects the uncertainty of these parameters owing to insufficient data, model limitation and other reasons. The R-distribution is frequently taken to be the normal distribution with mean μ and standard deviation σ:

$$f(g; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{g - \mu}{\sigma} \right)^2 \right] \tag{65b}$$

Then, following equation (64), the fragility curve $F(g; \mu, \sigma)$ can be expressed as

$$F(g; \mu, \sigma) = \Phi\left(\frac{g - \mu}{\sigma} \right) \tag{65c}$$

where Φ is a cumulative standardized normal distribution.

For n measurements, let m and s be the mean and unbiased standard deviation. Then U-distribution and confidence intervals for parameters μ and σ can be established using student's t distribution of parameter t and chi-square distribution of parameter χ^2 with $n - 1$ degrees of freedom where

$$t = \frac{\mu - m}{s} \tag{66a}$$

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2} \tag{66b}$$

If the failure indicator g has a log-normal R-distribution with median G and logarithmic standard deviation S_R. It is assumed that G has some uncertainty and this uncertainty can be described with a log-normal U-distribution having median G and standard deviation S_U. The R- and U-distributions can be expressed as

$$f_R(g; G, S_R) = \frac{1}{\sqrt{2\pi}S_R G} \exp\left[-\frac{1}{2}\left(\frac{\ln g - \ln G}{S_R} \right)^2 \right] \tag{67a}$$

$$f_U(G; G, S_U) = \frac{1}{\sqrt{2\pi}S_U G} \exp\left[-\frac{1}{2}\left(\frac{\ln g - \ln G}{S_U} \right)^2 \right] \tag{67b}$$

The fragility curve can be expressed analytically as

$$F(g; G, S_R) = \Phi\left(\frac{\ln g - \ln G}{S_R}\right) \tag{67c}$$

The composite distribution of failure indicator g is given by the product of the conditional probability $f_R(g; G, S_R)$ for g, given G and the probability $F_U(G; G, S_U)$ of the parameter G:

$$f_c(g; G; S_R, S_U) = f_R(g; G, S_R) \cdot f_U(G; G, S_U) \tag{67d}$$

Thus, a family of fragility curves can be constructed with three parameters G, S_R and S_U. The parameters are estimated from data. One way to estimate them is using a safety factor method based on the relationship between the safety factor F and the failure indicator g:

$$F = \frac{g}{g_s} \tag{68a}$$

where g_s is a design value of the failure parameter g for the event.

The safety factor model expresses F as a product of several random factors F_i:

$$F = \prod_{i=1}^{k} F_i \tag{68b}$$

For a dam failure such factors could be strength factor, drain factor, etc. Each factor has a mean and logarithm standard deviation, which presumably can be determined and combined as

$$F = \prod_{i=1}^{k} F_i \tag{69a}$$

$$G = g_s \cdot F \tag{69b}$$

$$S_R^2 = \sum_{i=1}^{k} [S_R^{(i)}]^2 \tag{69c}$$

$$S_U^2 = \sum_{i=1}^{k} [S_U^{(i)}]^2 \tag{69d}$$

RELIABILITY

Reliability of a system can be defined as the probability that the system will perform the 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 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 the 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.

Reliability function

Let the failure time (or time to failure) of a system under specified conditions be denoted by T and the pdf of T by $f(t)$. The probability of failure as a function of time can be expressed as

$$P(T \leqslant t) = F(t) = \int_0^t f(s)\, \mathrm{d}s, \quad t \geqslant 0 \tag{70}$$

where $F(t)$ is the probability that the system will fail by time t, or the cumulative distribution function (cdf), and can be designated as the failure cdf or unreliability function. If the reliability is defined as the probability of success, that is the system will perform its intended function for at least a period of time t, then

$$R(t) = P(T \geqslant t) = \int_t^\infty f(s)\, \mathrm{d}s = 1 - F(t) \tag{71}$$

where $R(t)$ is the reliability function, which can be computed directly from the knowledge of the failure time distribution.

Mean time to failure

The mean time to failure (MTTF) is the expected time during which the system will perform successfully

$$E(T) = \int_0^\infty t f(t)\, \mathrm{d}t \tag{72}$$

where $E(T)$ is also called the expected life. Another way to determine the MTTF is

$$E(T) = \int_0^D R(t)\, \mathrm{d}t \tag{73}$$

if

$$\lim_{t \to \infty} tR(t) = 0.$$

If the system is resurrected through repair and maintenance then $E(T)$ is known as the mean (operating) time between failures (MTBF). It may also be appropriate to define the mean time to failure of a system that has survived to time t:

$$M(t) = \frac{1}{R(t)} \int_0^\infty \tau f(t + \tau)\, \mathrm{d}\tau \tag{74}$$

which is the mean residual life. Thus,

$$M(0) = E(T) \tag{75}$$

Failure and hazard rates

Consider a time interval $[t_1, t_2]$. Then the rate at which failures occur in that time interval is the failure rate. It 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, t_1. The interval failure rate is

$$F(t_1, t_2) = \left[\frac{R(t_1) - R(t_2)}{R(t_1)}\right]\left[\frac{1}{t_2 - t_1}\right] \tag{76}$$

where the first bracketed term expresses the conditional probability of failure during $[t_1, t_2]$, given survival to time t_1, and the second factor is a dimensional characteristic used to express the conditional probability on a per unit time basis.

The hazard rate (or hazard rate function or simply hazard function) is defined by the limit of failure as the length of the interval $[t_1, t_2]$ approaches zero. This implies the instantaneous failure rate. The hazard rate $h(t)$ is expressed as

$$h(t) = \lim_{\Delta t \to 0} \frac{R(t) - R(t + \Delta t)}{\Delta t\, R(t)} = \frac{1}{R(t)}\left[- \frac{\mathrm{d}R(t)}{\mathrm{d}t} \right] = - \frac{\mathrm{d}\ \ln\ R(t)}{\mathrm{d}t} = \frac{f(t)}{R(t)} \tag{77}$$

Since $-\mathrm{d}R(t)/\mathrm{d}t = f(t)$, the failure time pdf, $f(t)\,\mathrm{d}t$, represents the probability that a system that has survived to time t will fail in the small interval of time t to $t + \mathrm{d}t$. Thus, $h(t)$ is the rate of change of the conditional probability of failure, given survival to time t. Also to be noted is that $f(t)$ is the rate of change of the ordinary (unconditional) probability of failure. The hazard rate indicates the change in failure rate over the lifetime of a population of systems. If $h(t)$ is increasing (decreasing) in $t \geqslant 0$, then $f(t)$ is said to be increasing failure rate (IFR) [decreasing failure rate (DFR)] distribution.

A typical hazard rate follows a bath-tub shape, with three distinct failure regions: (1) initial failure region, characterized by a decreasing failure rate; (2) chance or random failure region, characterized by a near constant failure rate; and (3) wear out failure region, characterized by an increasing failure rate.

An important mathematical relationship between hazard rate and the reliability function can be expressed as

$$R(t) = \exp\left[- \int_0^t h(x)\ \mathrm{d}x \right] \tag{78}$$

From Equations (77) and (78), it follows that

$$f(t) = h(t) \exp\left[- \int_0^t h(x)\ \mathrm{d}x \right] \tag{79}$$

Thus, $f(t)$, $R(t)$ and $h(t)$ are all related and a knowledge of one determines the other two. These functions can be derived with employment of POME in a usual manner. The importance of entropy in reliability engineering lies in its ability to incorporate one's prior knowledge in the estimated reliability. A comprehensive discussion on the use of POME in reliability estimation is given by Tribus (1963).

APPLICATION IN HYDROLOGY AND WATER RESOURCES

Derivation of distributions

POME has been employed to derive a variety of distributions some of which have found wide applications in hydrology and water resources. Many of these distributions have been summarized by Singh and Fiorentino (1992). Kapur (1982) derived POME-based distributions for continuous random variates over a finite interval. Li *et al* (1987) and Singh (1992) summarized derivations of a wide range of continuous distributions commonly employed in environmental and water sciences. Griffith (1972) presented a computer solution of deriving the discrete maximum entropy distributions. Basu and Templeman (1984) described an algorithm generating the maximum entropy distributions for probabilistic data. Verdugo Lazo and Rathie (1978) presented a table of differential entropies for various continuous probability distributions.

When there are no constraints, or x_i ranges from 1 to n, then POME yields a uniform distribution, with all $p_i = (1/n)$. As more constraints are introduced, the distribution becomes more peaked and possibly skewed. In this way, the entropy H reduces from a maximum for the uniform distribution, to zero when the situation is deterministic. The increase of entropy by adding the fuzzy entropy, as in Equation (28), tends to make the revised distribution $\bar{p}_i$ more platykurtic than p.

Levine (1980) investigated the inversion procedure providing the most conservative (probable) distribution from partial data. Jaynes (1984) expounded on the view of treating inversion problems as inference problems and the resulting decision theory.

Montroll and Shlesinger (1983) employed POME to derive distributions with long tails, and showed that such distributions with inverse power tails would be a consequence of specifying the constraint as the average value of a complicated logarithmic function. The distributions derived were Levy and log-normal.

Downson and Wragg (1973) derived POME-based distributions having prescribed first and second moments, and derived the conditions under which such distributions would exist. Solona and Arteaga (1987) examined the feasibility of POME for second-order uncertainties over finite and infinite domains. Collins and Wragg (1977) discussed a numerical procedure for construction of probability distributions as histograms.

Kapur (1983a,b) applied POME to derive the logistic distribution. Efron (1986) used POMCE in the context of double exponential families and their use in generalized linear regression. Sogawa *et al.* (1986) employed POME to derive a multivariate distribution and investigate its applicability to annual rainfall and annual maximum daily rainfall. Lind and Solona (1988), Solona and Lind (1989) and Lind and Hong (1991) employed POMCE in conjunction with fractile constraints to derive several probability distributions. They applied their method to model compound extreme hydrological phenomena, such as extreme sea levels under storm conditions. Lind (1989) presented a procedure for estimation of distributions of continuous variables on the basis of sample data in two steps: (1) selection of a reference distribution from a restricted model, and (2) selection of a posterior distribution using the sample rule. Thus, the procedure involves both POME and POMCE. Lind and Chen (1987) and Solona and Lind (1989) investigated the monotonic property of distributions based on POME and POMCE with fractile constraints. The principle of consistency is satisfied by the posterior distribution and holds for system reliabilities in the multivariate case. Singh and Cruise (1990) employed POME with fractile constraints in reservoir operation and management.

Singh and Guo (1995a) applied POME to derive the two-parameter Pareto distribution, Singh and Guo (1997b) the two-parameter generalized Pareto distribution and Singh and Guo (1995b) the three-parameter generalized Pareto distribution. The two-parameter log-logistic distribution was derived by Singh and Guo (1995c) and the three-parameter log-logistic distribution by Singh *et al.* (1993).

Parameter estimation

POME has been successfully applied to estimate the parameters of a number of probability distributions and Singh and Fiorentino (1992) summarized some of these studies. Singh and Guo (1995a, b, c) estimated parameters of the Pareto distributions (two-parameter, generalized two-parameter and generalized three-parameter); Singh and Guo (1995d) parameters of the two-parameter log-logistic distribution; and Singh *et al.* (1993) parameters of the three-parameter log-logistic distribution. In all of these studies the POME method was found to be either superior to or comparable to the methods of moments, maximum likelihood estimation and some others. Kam and Brown (1983) discussed a procedure using both symmetric and monotonically increasing fuzzy entropies to update parameters of a truncated normal distribution in light of subjective information.

Wragg and Downson (1970) discussed a computational procedure for estimating a probability distribution over $[0, \infty)$ when several low-order moments are known. They showed that the POME-based method broke down if $\mu_2 > 2\mu_1^2$ where μ_1 and μ_2 are the first-order and second-order moments about the origin. Agmon *et al.* (1979) developed an algorithm for determining a POME-based distribution subject to specified constraints. It incorporates two preliminary stages. The first stage verifies that the constraints are linearly independent and the second checks if a feasible solution exists.

Liu (1972) discussed the method of Lagrange multipliers for exploitation of the entropy principle. Mead and Papanicolaou (1984) derived necessary and sufficient conditions for the existence of a maximum entropy solution subject to specified moments. Landau (1987) attempted to establish links between the entropy concept, trigonometric moment problem and orthogonal decomposition. Because the entropy of a

distribution measures the number of ways in which it can be realized, the POME-based distribution occurs by far the most frequently of the available distributions, and is thereby viewed as the least informative choice. The least informative process consistent with the data is that for which the prediction ahead is poorest.

Jaworski (1987) examined the contribution to the total entropy arising from the extra information corresponding to higher order moments. From a purely thermodynamic point of view, this extra information is non-essential and can be neglected in the POME method. He concluded that the maximum entropy inference has a certain stability property with respect to information corresponding to higher order moments.

Larimore (1983) proposed an asymptotic likelihood principle which is an intuitively compelling justification for use of entropy as a natural measure of approximation to the actual distribution by a predictive distribution. This offers a rational justification for use of entropy for evaluating parameter estimation as well as model order and structure determination procedures.

Akaike (1973) showed that the classical maximum likelihood principle can be founded on the information theoretical considerations and that its applicability widened enormously. Dutta (1966) established an equivalence between the maximum entropy estimation and the maximum likelihood estimation with an exponential distribution.

Entropy spectral analysis for flow forecasting

The maximum entropy spectral analysis (MESA) was introduced by Burg (1968, 1972, 1975). It has several advantages over conventional spectral analysis methods. It has short and smooth spectra with high degree resolutions (Fougere *et al.*, 1976). 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. Ulrych and Clayton (1976) reviewed principles of MESA and the closely related problem of autoregressive time series modelling. Shore (1979) presented a comprehensive discussion of minimum cross-entropy spectral analysis.

The relationship between spectrum $W(f)$ with frequency f of a stationary process $x(t)$ and entropy $H(f)$ can be expressed as

$$H(f) = \frac{1}{2}\ln(2w) + \frac{1}{4w}\int_{-w}^{+w} \ln[W(f)]\,\mathrm{d}f \tag{80}$$

where w is the frequency band. Equation (80) is maximized subject to the constraint equations given as autocorrelations until log m

$$\rho(n) = \int_{-w}^{+w} W(f)\exp(i2\pi fn\Delta t)\,\mathrm{d}f, \quad -m \leqslant n \leqslant +m \tag{81}$$

where Δt is the sampling time interval and $i = (-1)^{1/2}$. Maximization of Equation (80) is equivalent to maximizing

$$H(f) = \int_{-w}^{+w} \ln[W(f)]\,\mathrm{d}f \tag{82}$$

which is known as the Burg entropy. The spectrum $W(f)$ can be expressed in terms of the Fourier series as

$$W(f) = \frac{1}{2w}\sum_{n=-\infty}^{\infty} \rho(n)\exp[-i2\pi nf\Delta t] \tag{83}$$

Substitution of Equation (81) in Equation (83) and maximization lead to MESA. 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 (1987, 1991a, b) employed MESA for long-term stream flow forecasting. Krstanovic and Singh (1993a, b) extended the MESA method to develop a real-time flood forecasting model. Padmanabhan and Rao (1986, 1988) applied MESA to analyse rainfall and river flow time series. Rao *et al.* (1980) compared a number of spectral analysis methods with MESA and found MESA to be superior. Eilbert and Christensen (1983) analysed annual hydrological forecasts for central California and found that dry years might be more predictable than wet years. Dalezios and Tyraskis (1989) employed MESA to analyse multiple precipitation time series.

Basin geomorphology

Culling (1987, 1988) presented a comprehensive discussion on use of entropy in characterization of landscape. Fiorentino *et al.* (1993) undertook an entropy-based morphological analysis of river basin networks. Using the connection between entropy and potential energy, the mean basin elevation was found to be linearly related to the basin entropy:

$$\bar{y} = -a \ln b + aH \tag{84}$$

where $\bar{y}$ is mean basin elevation (potential energy), H is basin entropy and a and b are parameters.

A relation between the mean elevation of a drainage subnetwork y_d and its diameter d was derived:

$$\bar{y}_d = \bar{y} \frac{\ln d}{\ln D} \tag{85}$$

where D is the basin diameter. The parameter a was expressed as

$$a = \frac{\bar{y}}{\ln D} \tag{86}$$

In a similar manner, the relation between the fall in elevation from the source to the outlet of the main channel, Y, and the entropy of its drainage basin H was found to be linear and so also was the case between the elevation of a node and the logarithm of its distance from the source.

$$Y = -a_0 \ln b + a_0 H \tag{87}$$

and

$$y_d = y_0 - Y \frac{\ln d}{\ln D} \tag{88}$$

where y_0 is the elevation of the source of the channel, y_d is the elevation of the downstream node at a distance d from the source and a_0 is related to a as $a_0 = ad/\bar{d}_d$, where $\bar{d}_d$ is the topological distance of the centroid of the subbasin width function.

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

$$H = \ln(R_L^W - 1) - \ln(R_L - 1) \tag{89}$$

where R_L is the stream length ratio and W is the basin order. If $R_L > 1$, then

$$H = W \ln R_L - \ln(R_L - 1) \tag{90}$$

This relation can be characterized 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.

$$H = \ln n \frac{\ln R_L}{\ln R_B} + \ln\left[\frac{R_L}{R_L - 1}\right] \tag{91}$$

where R_B is the bifurcation ratio and n is the magnitude of the basin $n = R_B^{W-1}$. 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:

$$H = \ln\left(\frac{R_L}{R_L - 1} n^{1/F}\right) \tag{92}$$

where F is an exponent = 1·75, and

$$F = \frac{\ln R_B}{\ln R_L} \tag{93}$$

$$D = Cn^{1/f}, \quad C = \frac{R_L}{R_L - 1} \tag{94}$$

These relationships were verified on three drainage basins in southern Italy and the results were found to be in excellent agreement.

Design of hydrological networks

Krstanovic and Singh (1992a, b) developed an entropy-based approach for space and time evaluation of rainfall networks in Louisiana. The evaluation was made for five sampling intervals: daily, two-day, weekly, monthly and yearly, and for two separate seasons. Space and time dependencies between rain gauges were examined by autocovariance and cross-covariance matrices. Multivariate distributions, associated with different dependencies, were derived using POME. For each distribution marginal entropy, joint entropy and transinformation were obtained. 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, and the lines of equal information were constructed to assist with the decision on expanding the existing network or deleting the unnecessary gauges.

Yang and Burn (1994) employed the entropy concept to develop a methodology for data collection network design. The methodology employed a measure of information flow, called directional information transfer index (DIT), between gauging stations in the network. This measure was based on the entropy of gauging stations and pairs of gauging stations. The entropy was calculated with multivariate probability distributions obtained using non-parametric methods. The DIT was defined as

$$DIT = \frac{T}{H} = \frac{(H - H_{lost})}{H} = 1 - \frac{H_{lost}}{H} \tag{95}$$

where H is entropy (information content) of station X, T is transinformation (or mutual information) between two stations X and Y and H_{lost} is the information lost (amount of information transmitted). The value of DIT varies from 0, where no information is transmitted and the stations are independent, to 1, where no information is lost and the stations are fully dependent. Note DIT is not symmetrical, for $DIT(X, Y) = T/H(X)$ for station X is, in general, not equal to $DIT(Y, X) = T/H(Y)$ for station Y. $DIT(X, Y)$ describes the fractional information inferred by station X about Y, whereas $DIT(Y, X)$ the fractional information inferred by station Y about 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 the other side.

The concept of DIT can be applied to regionalization of the network. If both $DIT(X, Y)$ and $DIT(Y, X)$ are high then the two stations X and Y are strongly dependent and should, therefore, be in the same group. If neither DIT is high, they are independent and should be in separate groups. If only DIT [say $DIT(X, Y)$] is high, then the station Y can join station X if Y does not belong to any other group, otherwise it remains in its own group. The base station X does not join station Y's group under any circumstance, because, with eventual elimination of X, information at that site would be lost.

DIT is both a measure of the information transmission capability and an indicator of the dependency of a station pair. With application of DIT, Yang and Burn (1994) developed an informational measure of relationship S-DIT between a pair of stations, called information connection:

$$S - DIT_i = \sum_{j=1;\, j\neq i}^{m} DIT_{ij} \tag{96}$$

where DIT_{ij} is the information inferred by station i about station j and m is the number of stations in the group. The station in each group with the highest value for $S - DIT$, in comparison with other members of the group, should be retained in the group.

Krstanovic and Singh (1992c) investigated transfer of information in a 116-year rainfall series at San Jose, California, on a monthly basis. This series included series of unusually dry periods two to eight years long, and excessive wet periods one to three years long. The consecutive dry and wet sequences appeared random and unrelated. The entropy concept was used to define the relationship of drought and flood sequences to the average rainfall, the relationship between selected drought sequences, the relationship between flood sequences, the drought and flood patterns and the worth of information transfer in the rainfall record.

Reliability of water distribution systems

Awumah *et al*. (1990) employed entropy-based measures for evaluation of reliability and redundancy of water distribution networks. They noted that these measures accurately reflect changes in the network reliability. The network redundancy (or reliability) measure was derived to be

$$S = \sum_{j=1}^{N} \left[\frac{Q_j}{Q_0} S_j \right] - \sum_{j=1}^{N} \left[\frac{Q_j}{Q_0} \ln \frac{Q_j}{Q_0} \right] \tag{97}$$

where S is network redundancy, Q_0 is sum of flows in all links in the network, S_j is total flow into node j and N is number of nodes. S_j is expressed as

$$S_j = -\sum_{i \in \bar{U}_j} \frac{q_{ij}}{Q_j} \ln \frac{q_{ij}}{Q_j} \tag{98}$$

in which q_{ij} is flow in link from node i to node j and $\bar{U}_j$ is set of nodes on the upstream ends of links incident on node j. The right side of Equation (97) is a sum of two terms. The first term expresses an algebraic sum of weighted entropy measure at each of the constituent nodes, and the second term redundancy among the N nodes.

However, the redundancy of a network also depends on the ability of network to respond to the failure of one of its links. To take account of this feature Awumah *et al*. (1990) developed the concept of transmissivity of redundancy, wherein upstream nodes transfer their redundancy to nodes immediately downstream (one link downstream) of them in direct proportion to the ratio of their flow entering the downstream node from the upstream node to the total flow entering the downstream node. Thus, alternative paths need to be enumerated. The value of the number of equivalent paths was obtained as

$$a_j = \sum_{i \in \bar{U}_j} ND_{ij} \left[1 - \left(\sum_{k=1}^{MD_{ij}} (d_k - 1) \Big/ \sum_{k=1}^{MD_{ij}} d_k \right) \right] \tag{99}$$

where a_j is the effective number of independent paths from the source to demand node j; ND_{ij} is the number of paths, independent or dependent, through the link from node i to node j; MD_{ij} is the number of links in

ND_{ij} paths; and d_k is the degree of link k, i.e. number of paths in which link k is a member. If a_{ij} denotes the number of equivalent independent paths through the link from node i to demand node j then

$$a_{ij} = ND_{ij}\left[1 - \left(\sum_{k=1}^{MD_{ij}}(d_k - 1)\Big/\sum_{k=1}^{MD_{ij}} d_k\right)\right] \tag{100}$$

Thus, Equation (98) changes to

$$S_j = -\sum_{i\in\bar{U}_j}\left[\left(\frac{q_{ij}}{Q_j}\ln\frac{q_{ij}}{Q_j}\right)\right] + \sum_{i\in\bar{U}_j}\left[\frac{q_{ij}}{Q_j}\right]\ln a_{ij} \tag{101}$$

The first term represents the redundancy measure for the node under the assumption that each incident link constitutes exactly one path from this source to this node. The second term depends upon the true number of alternative paths and represents a correction factor to reduce the number of alternative paths if some of the paths are dependent.

Awumah *et al.* (1990) applied the above entropy-based method to treat reliability/redundancy related to pipe failures. Pumps and storage also contribute to reliability and redundancy but were not included in their method. They applied the method to a range of network layouts and showed that the entropy-based redundancy measure is a good indicator of the relative performance implications of different levels of redundancy, and is capable of developing reliable network layout and component designs without having to use the large numbers of load patterns or intensive iterative approaches normally required.

Subsurface hydrology

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 confined flow the total probability was the constraint used. The cumulative probability distribution (cdf) of piezometric head h in a confined aquifer was found to be

$$F(h) = \frac{h - h_0}{h_u - h_0} \tag{102}$$

where h_0 is the piezometric head at the lower end of the aquifer ($x = 0$) and h_u is the piezometric head at the upper end of the reservoir ($x = L$) in which L is the length of the aquifer. $F(h)$ increases linearly from $h = h_0$ to $h = h_u - h_0$.

$$f(h) = \frac{1}{h_u - h_0} \tag{103}$$

which is a uniform distribution. The information content of this system is

$$H = \ln(h_u - h_0) \tag{104}$$

ln a similar manner, $f(h)$, $F(h)$ and H were derived for unconfined flow.

Bos (1990) employed POME and Bayes' theorem for aquifer parameter identification. He considered steady-state conditions for predicting flow pattern in an aquifer on a regional scale and hence the only parameter of interest was transmissivity. From a few measurements of transmissivity (T) based on pumping tests and of piezometric head, Bos (1990) was able to derive the pdf of T.

In probabilistic modelling of contaminant transport, Woodbury and Ulrych (1993) used the principle of minimum relative entropy (POMRE). In groundwater engineering, it is often true that few measurements of, say, hydraulic conductivity are available, and there is a large degree of uncertainty in the measured values of fundamental flow parameters. On the other hand, assigning probability distributions to these parameters is extremely useful. Based on limited information a prior pdf of a set of model parameters was determined. For the advective–dispersive equation (ADE) the parameters were porosity, hydraulic conductivity, hydraulic

gradient (assumed steady and parallel to the x axis), macroscopic field-scale dispersivity and the pore-scale diffusion coefficient. Based on upper and lower bounds and expected values of the parameters, Woodbury and Ulrych (1993) assigned prior probabilities using POMRE. This pdf was then used in Monte Carlo simulations to provide expected values in the field such as concentration, spatial moments and confidence limits.

Application in hydraulics

Chiu and Murray (1992) applied POME to determine the probability distribution of velocity in a non-uniform open channel flow. The analysis of the pdf velocity permitted determination of the flow discharge, and momentum and energy coefficients. The parameters of velocity distribution were obtained from the values of maximum and mean velocity, or from a set of such typical data as hydraulic radius, roughness and channel or energy slope. The derived velocity distribution was

$$\frac{u}{u_m} = \frac{1}{m}\ln\left[1 + (e^M - 1)\frac{y - y_0}{y_m - y_0}\right] \tag{105a}$$

where u is velocity that monotonically increases with y, a dimensionless variable, within the range $(y_0, y_m) \simeq (0, 1)$; M is an entropy parameter; u_m is maximum velocity in a channel cross-section. A value of y is assigned to each isovel along which the velocity has a constant value; y_m is the maximum value of y at which $u = u_m$; y_0 is the minimum value of y that occurs along the channel bed (boundary), also an isovel along which $u = 0$. Chiu (1989) derived a relation between M, u_m and the cross-sectional mean velocity:

$$\frac{\bar{u}}{u_m} = e^M(e^M - 1)^{-1} - \frac{1}{M} \tag{105b}$$

Equation (105a) is equivalent to

$$\frac{y - y_0}{y_m - y_0} = \int_0^u f(u)\, du \tag{106}$$

in which $f(u)$ is the pdf of u:

$$f(u) = \exp(a_0 + a_1 u) \tag{107}$$

where a_0 and a_1 are Lagrange multipliers. The entropy-based velocity distribution fits experimental data very well and is of great practical value in hydraulic modelling. It permits an efficient use of limited hydraulic data routinely collected.

Chiu and Abdin Said (1995) suggest that under a wide range of discharge and of water depth, the equilibrium state of a channel section tends to approach a constant value of the entropy parameter. This parameter of a section can be determined from the relation between the mean and maximum velocities. The use of the constant value of the entropy parameter predetermined for a channel section greatly simplifies the discharge estimation.

Yang (1994) showed that the fundamental theories in hydrodynamics and hydraulics can be derived from variational approaches based on maximization of entropy, minimization of energy or minimization of energy dissipation rate. These approaches are consistent with vectorial approaches used in classical hydraulics, provided that the concept of entropy and energy are properly defined and correctly used.

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 activities 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 results

in 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 a 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, based on an ecological rule) is related to the degree of pollution. In general, the number of species decreases as the degree of pollution increases. Thus, DI can be calculated from the Shannon entropy as

$$DI = -\sum_{i=1}^{n} p_i \log p_i \tag{108}$$

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 employed to evaluate water quality of a water body in time and space, as well as to compare water quality of different water bodies (Tai and Goda, 1980, 1985; Goda *et al.*, 1981).

The purpose of water treatment is to decrease the entropy that was increased by environmental pollution. In a waste water treatment system, when clean water (treated water) is mixed with waste water (sludge), entropy changes as a result of this mixing. The feed water is separated into effluent (permeate) and sludge (concentrate) by a separation system such as reverse osmosis. The entropy per mole of ideal solution or dilute real solution can be calculated as

$$H = -R\Sigma x_i \ln x_i + \Sigma x_i h_i \tag{109}$$

where H is entropy per mole of the solution, x_i is mole fraction of component (substance) i, h_i is standard entropy of component i and R is the gas constant. The first term in Equation (109) refers to the entropy of mixing per mole of ideal or dilute real solution, and the second term standard entropy per mole of the solution. Equation (109) is similar to fuzzy entropy or algorithmic entropy.

Consider the polluted water as the dilute real solution. In the treatment system, there are $(1 - A)$ moles of waste water mixed with A moles of clean water. The entropy change resulting from this mixing can be expressed as

$$\Delta H = AH_2 + (1 - A)H_3 - H_1 \tag{110}$$

where A is the ratio of effluent flow to raw water flow, H_1 is the entropy per mole of raw water:

$$H_1 = -R\Sigma x_i \ln x_i + \Sigma x_i h_i \tag{111}$$

H_2 is the entropy per mole of effluent:

$$H_2 = -R\Sigma x_i^1 \ln x_i^1 + \Sigma x_i^1 h_i \tag{112}$$

and H_3 is the entropy per mole of sludge (or concentrate):

$$H_3 = -R\Sigma x_i^2 \ln x_i^2 + \Sigma x_i^2 h_i \tag{113}$$

where x_i, x_i^1 and x_i^2, are, respectively, mole fractions of component i of the raw water, the effluent and the concentrate.

The decreasing rate of entropy of feed water in the reverse osmosis system is obtained as

$$\frac{\Delta H}{\Delta t} = \Delta H \cdot \frac{Q_r}{18} \tag{114}$$

where ΔH is obtained from Equation (110), Q_r is feed water flow rate and one mole of feed water has been assumed to be 18 cm^3. The interval entropy production is calculated using thermodynamic considerations. 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. Goda and associates (Tai and Goda, 1980, 1985; Goda *et al*., 1981) employed this rationale for evaluating the efficiency of the reverse osmosis processes, as well as of other waste water treatment processes.

Design of water quality networks

Harmancioglu *et al*. (1992a,b,c) presented a comprehensive review of design of water quality networks and discussed the potential of the entropy concept in this content. Harmancioglu *et al*. (1994) reviewed the advantages as well as the limitations of the entropy method as applied to the design of water quality monitoring networks. Application of the entropy method 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.

Kusumulyono and Goulter (1994, 1995) applied POME to develop a methodology for predicting water quality values at discontinued sites. The methodology provides unbiased predictions of water quality levels at upstream tributaries and on the mainstream of a river, given observed changes in the distribution of the same water quality parameter at a downstream location. The method also has potential for identifying the location of causes of observed changes in water quality at a downstream location. The methodology employs the principle of minimum discrimination information (MDI) due to Kullback and Leibler (1951). Assume that the water quality at a downstream location, a, is a function of the water quality at two upstream locations, b and c:

$$x_a = f(x_b, x_c) \tag{115}$$

The water quality level at the downstream location a has now changed appreciably owing to changes in water quality levels at either of locations b and c, or both. With use of POME or MDI, the distribution of water quality levels at the upstream locations is predicted. To that end,

$$\max\ H = -\sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij}\ \ln\left[\frac{p_{ij}}{(q_{ij}/m)}\right] \tag{116}$$

subject to

$$\sum_{i=1}^{n} p_{ij} = \frac{1}{m}, \quad j = 1, 2, \ldots, m \tag{117}$$

$$\sum_{i=1}^{n} p_{ij}x_{ij} \Big/ \sum_{i=1}^{n} p_{ij} = \mu_j, \quad j = 1, 2, \ldots, m \tag{118}$$

$$\sum_{i=1}^{n} p_{ij}x_{ij}^2 \Big/ \sum_{i=1}^{n} p_{ij} = \mu_j^2 + \sigma_j^2 = 1, 2, \ldots, m \tag{119}$$

$$\mu = f(\mu, \mu_2, \ldots, \mu_m) \tag{120}$$

$$0 \leqslant\ p_{ij} \leqslant 1 \tag{121}$$

$$0 \leqslant\ q_{ij} \leqslant 1 \tag{122}$$

$$u_j \geqslant\ 0 \tag{123}$$

$$\sigma_j \geqslant\ 0 \tag{124}$$

where x_{ij} is possible water level i at station j, q_{ij} is prior probability of event x_{ij}, μ_j is mean of the water quality level at station j from the prior distribution, σ_j is standard deviation of water quality level at station j from the prior distribution, μ is observed (changed) mean of the water quality level at the downstream location, p_{ij}

is probability of event x_{ij} to be assigned, given the mean of water quality level downstream μ, m is number of upstream stations and n is number of intervals (discrete water quality values) at each station.

Given an observed change in water quality levels at a downstream location, the above formulation predicts the probabilities of each possible water quality level at each of the upstream stations. These probabilities are then used to develop new unbiased estimates of mean values of the water quality at the upstream stations. The q_{ij} values can be the probabilities that existed prior to the observed change in water quality downstream, or those probabilities that can be associated with new known conditions.

Kusumulyono and Goulter (1995) also discussed computational aspects of their entropy-based methodology, including the level of discretization used in converting the continuous probability distribution of water quality values to discrete levels required for the entropy function, and the choice of the interval of time for which to assign the value of water quality (period of time averaging) through the entropy. They found that, depending on the length of the record, the choice of the time interval for which the water quality values were assigned (period for time averaging) exercised a significant effect on the accuracy of predicted water quality values.

Optimization

The Shannon entropy and the Jaynes maximum entropy have been utilized in constrained optimization by Templeman and associates (Templeman and Xingsi, 1987, 1989; Xingsi and Templeman, 1988; Simoes and Templemen, 1989; Templeman, 1989) and others (Kapur, 1972; Elfving, 1980; Censor, 1982; Erlander, 1981). To discuss an entropy application to optimization we consider, following Templeman and Xings (1989), the problem of finding a local solution X^* of problem A:

$$\underset{X}{\text{minimize: }} f(X) \tag{125}$$

subject to:

$$g_j(X) \leqslant 0, \quad j = 1, 2, \ldots, m \tag{126}$$

where $F(X)$ and $g_j(X)$, $j = 1, 2, \ldots, m$, represent real-valued smooth functions of a vector $X = x_i$, $i = 1, 2, \ldots, n$. The problem A can be solved through the use of the surrogate problem B:

$$\underset{X}{\text{minimize: }} f(X) \tag{127}$$

subject to:

$$\sum_{j=1}^{m} a_j g_j(X) \leqslant 0 \tag{128}$$

The single surrogate constraint of Equation (128) is defined as a positive linear combination of the original constraints of Equation (126). Thus, a single equation replaces m constraints in constraint surrogation. $A = a_i, j = 1, 2, \ldots, m$, is an m-vector of non-negative weights, named surrogate multipliers, which may be normalized, without loss of generality, such that

$$\sum_{j=1}^{m} a_j = 1, \quad a_j \geqslant 0, \quad j = 1, \ldots, m \tag{129}$$

If the values of a_j are chosen then the solution of problems A and B will be equivalent (Glover, 1968; Gould, 1969; Greenberg and Pierskalle, 1970).

The Lagrangean of problem B is

$$L(X, b, A) = f(X) + b\sum_{j=1}^{m} a_j g_j(X) \tag{130}$$

in which b is the Lagrange multiplier for constraint Equation (128). An essential condition for the equivalence of problems A and B is that L must satisfy the Lagrangean saddle-point condition:

$$L(X, b^*, A^*) \geqslant L(X^*, b^*, A^*) \geqslant L(X^*, b, A) \tag{131}$$

This condition suggests a two-phase approach for solution of problem A via B. In the first phase, an initial set of surrogate multipliers $A^{[0]}$ is chosen and the problem B is solved by minimization over X corresponding to the left-hand inequality in Equation (131) to yield $X^{[0]}$. The second phase involves updating the multipliers to $A^{[1]}$ using some maximization scheme subject to the right-hand inequality in Equation (131), and then again solving the problem B to yield $X^{[1]}$. This process is repeated until the sequence $(A^{[0]}, X^{[0]})$, $(A^{[1]}, X^{[1]}), \ldots$, converges to a solution of problem B and hence also of problem A at (A^*, X^*).

The updating of surrogate multipliers A is accomplished with entropy maximization. Note that if the m components of A are interpreted as discrete probabilities, then Equation (129) represents the axiomatic normality and non-negativity conditions of such probabilities. The constraint of Equation (128) can be shown to be an equality constraint and hence has the form of an expected value constraint. Thus, the least-biased estimates for $A^{[k]}$ at the kth iteration can be found by using POME as

$$\underset{A^{[k]}}{\text{maximize: }} H = -K\sum_{j=1}^{m} a_j^{[k]} \ln a_j^{[k]} \tag{132}$$

subject to:

$$\sum_{j=1}^{m} a_j^{[k]} = 1 \tag{133}$$

$$\sum_{j=1}^{m} a_j^{[k]} g_i(X^{[k-1]}) = \varepsilon \tag{134}$$

where H is the Shannon entropy, K is a positive constant and ε is an error term owing to the use of $g^{[k-1]}$ in place of $g^{[k]}$, and should approach zero as the iterations proceed. The values of $A^{[k]}$ from solution of Equations (132)–(134) are given by

$$a_j^k = \exp[cg_j(X^{k-1})/K] / \sum_{j=1}^{m} \exp[cg_j(X^{k-1})/K], \quad j = 1, 2, \ldots, m \tag{135}$$

where c is the Lagrange multiplier for Equation (135). c/K may be considered as a control parameter. Equation (135) represents the entropy-based updating formula for the surrogate multipliers A in the optimization methodology.

There is an alternative way to include entropy in the optimization algorithm. This is done by directly augmenting $f(X)$ in problem B with a multiplier entropy term and then treating the resulting problem as a minimax optimization over both sets of variables X and A. The problem B then becomes

$$\underset{X}{\text{minimize}}\ \underset{A}{\text{maximize: }} f(X) - \frac{1}{d}\sum_{j=1}^{m} a_j \ln a_j \tag{136a}$$

subject to:

$$\sum_{j=1}^{m} a_j = 1 \tag{136b}$$

$$\sum_{j=1}^{m} a_j g_j(X) = 0 \tag{136c}$$

where d is a positive constant. The Lagrangean of this problem is

$$L = f(X) + b\sum_{j=1}^{m} a_j g_j(X) + r\left[\sum_{j=1}^{m} a_j - 1\right] - \frac{1}{d}\sum_{j=1}^{m} a_j \ln a_j \tag{137}$$

which is seen to be an extension of Equation (130) owing to addition of multiplier normality and entropy expressions. Here r is the Lagrange multiplier for constraint Equation (128). Stationarity of L with respect to a_j and r yields

$$a_j = \exp[dbg_j(X)]/\sum_{j=1}^{m} \exp[dbg_j(X)], \quad j = 1, 2, \ldots, m \tag{138}$$

which is similar to Equation (135), with c/K replaced by db. Substitution of Equation (138) in Equation (137) yields

$$L^* = f(X) + \left(\frac{1}{d}\right) \ln \sum_{j=1}^{m} \exp[dbg_j(X)] \tag{139}$$

Minimization of L^* over X and with bd taking an increasing positive sequence of values, yields the solution of problem A. Thus, the entropy augmentation approach combines the two-phase approach into a single phase. The entropy-based optimization seems promising but its application in hydrology is yet to be made.

IMPLICATIONS FOR DEVELOPING COUNTRIES

One of the main problems plaguing hydrology in developing countries is the lack of data or lack of sufficient data. Frequently, either the data are missing or are incomplete, or 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, judgement and thumb or empirical rules have played a significant role in hydrological 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 maximizes 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 survey of entropy concepts and entropy-based modelling in hydrology and water resources presented here reveals that 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 whether the available information is adequate or not, and if not then additional information should be sought. In this way it brings a model and its modeller closer. As an objective measure of information or uncertainty, the entropy theory allows us to communicate with nature as illustrated by its application to design of data acquisition systems, design of environmental and hydrological networks, and assessment of the reliability of these systems or networks. In a similar vein, it helps us to 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 the use of entropy.

REFERENCES

Aczel, J. 1984. 'Measuring information beyond communication theory — Why some generalized measures may be useful, others not', *Aequat. Math.*, **27**, 1–19.

Aczel, J. and Daroczy, Z. 1975. *On Measures of Informations and their Characterizations*. Academic Press, New York.

Agmon, N., Alhassid, Y., and Levine, R. D. 1979. 'An algorithm for finding the distribution of maximum entropy', *J. Comput. Phys.*, **30**, 250–258.

Akaike, H. 1972. 'Information theory and an extension of the maximum likelihood principle', *Proceedings, 2nd International Symposium on Information Theory, Supplement to Problems of Control and Information Theory*, Budapest, Hungary. pp. 267–281.

Akaike, H. 1973. 'Information theory and an extension of the maximum likelihood principle', in Petrov, B. N. and Csaki, F. (Eds), *Proceedings, 2nd International Symposium on Information Theory*. Publishing House of the Hungarian Academy of Sciences, Budapest.

Akaike, H. 1974. 'A new look at the statistical model identification', *IEEE Trans. Auto. Control*, **AC-19**(6), 716–723.

Akaike, H. 1985. 'Prediction and entropy', in Atkinson, A. C. and Fienberg, S. E. (Eds), *A Celebration of Statistics*. Springer Verlag, Heidelberg.

Alpaslan, N., Harmancioglu, N. B., and Singh, V. P. 1992. 'The role of entropy concept in design and evaluation of water quality mentoring networks', in Singh, V. P. and Fiorentino, M. (Eds), *Entropy and Energy Dissipation in Water Resources*. Kluwer Academic Publishers, Dordrecht. pp. 261–282.

Awumah, K., Goulter, I., and Bhatt, S. K. 1990. 'Assessment of reliability in water distribution networks using entropy-based measures', *Stoch. Hydrol. Hydraul.*, **4**, 309–320.

Awumah, K., Goulter, I., and Bhatt, S. K. 1991. 'Entropy-based redundancy measures in water-distribution networks', *J. Hydraul. Engng*, **117**(51), 595–614.

Barbe, D. E., Cruise, J. F., and Singh, V. P. 1994. 'Derivation of a distribution for the piezometric head in groundwater flow using stochastic and statistical methods', in Hipel, K. W. (Ed.), *Hydrology and Environmental Engineering*, Vol. 2. Kluwer Academic Publishers, Dordrecht. pp. 151–161.

Basu, P. C. and Templeman, A. B. 1984. 'An efficient algorithm to generate maximum entropy distributions, *Int. J. Num. Meth. Engng*, **20**, 1039–1055.

Bos, R. 1990. 'Aquifer parameter identification by the maximum entropy and Bayes' theorem', Unpublished Report, Technion-Israel Institute of Technology, Haifa, Israel.

Brown, C. B. 1980. 'Entropy constructed probabilities', *J. Engng Mech. Div. ASCE*, **106** (EM4), 633–640.

Burg, J. P. 1968. 'A new analysis technique for time series data', Paper presented at the *NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics*, Enschede, The Netherlands.

Burg, J. P. 1972. 'The relationship between maximum entropy spectral and maximum likelihood spectra', *Geophysics*, **37**(2), 375–376.

Burg, J. P. 1975. 'Maximum entropy spectral analysis', Ph.D. Thesis, Stanford University, Palo Alto, California, 123 pp.

Campbell, L. L. 1970. 'Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities', *Ann. Math. Statist.,* **41**(3), 1011–1015.

Censor, Y. 1982. 'Entropy optimization via entropy projections', in Drenick, R. F. and Kozin, F. (Eds), *System Modeling and Optimization*. Springer-Verlag, New York. pp. 450–454.

Chiu, C. L. 1989. 'Velocity distribution in open channel flow', *J. Hydraul. Engng.,* **115**, 576–594.

Chiu, C. L. and Abidin Said, C. A. 1995. 'Maximum and mean velocities and entropy in open-channel flow', *J. Hydraul. Engng,* **121**(1), 26–35.

Chiu, C. L. and Murray, D. W. 1992. 'Variation of velocity distribution along nonuniform open-channel flow', *J. Hydraul. Engng*, **118**(1), 989–1001.

Collins, R. and Wragg, A. 1977. 'Maximum entropy histograms', *J. Phys. A*, **10**(9), 1441–1464.

Culling, W. E. H. 1987. 'Ergodicity, entropy and dimension in the soil covered landscape', *Trans. Japanese Geophys. Union*, **8**(3), 157–174.

Culling, W. E. H. 1988. 'Dimension and entropy in the soil covered landscape', *Earth Surf. Process. Landf.*, **13**, 619–648.

Dalezios, N. R. and Tyraskis, P. A. 1989. 'Maximum entropy for regional precipitation analysis and forecasting', *J. Hydrol.*, **109**, 25–42.

DeLuca, A. and Termini, S. 1972. 'A definition of a nonprobibalistic entropy in the setting of fuzzy sets theory', *Information Control*, **20**, 301–312.

Downson, D. C. and Wragg, A. 1973. 'Maximum entropy distribution having prescribed first and second moments, *IEEE Trans. Inform. Theory*, **IT-19**, 689–693.

Dutta, M. 1966. 'On maximum (information-theoretic) entropy estimation', *Sankhaya*, **28**(A4), 319–328.

Efron, B. 1986. 'Double exponential families and their use in generalized linear regression', *J. Am. Stat. Assoc.*, **81**(395), 709–721.

Eilbert, R. F. and Christensen, R. A. 1983. 'Performance of the entropy minimax hydrological forecasts for California, water years 1948–1977, *J. Climate Appl. Meteorol.*, **22**, 1654–1657.

Elfving, T. 1980. 'On some methods for entropy maximization and matrix scaling', *Linear Algebra Appl.*, **34**, 321–339.
Erlander, S. 1981. 'Entropy in linear programs', *Mathem. Progr.*, **21**, 137–151.
Evans, R. A. 1969. 'The principle of maximum information', *IEEE Trans. Rel.*, **R-18**(3), 87–90.
Evans, R. B. 1978. 'A new approach to deciding upon constraints in the maximum entropy formalism', in Levine, R. D. and Tribus, M. (Eds), *The Maximum Entropy Formalism*. MIT Press, Cambridge. pp. 169–206.
Fama, E. F. 1965. 'Tomorrow on the New York stock exchange', *J. Business*, **38**, 285–299.
Fiorentino, M., Claps, P., and Singh, V. P. 1993. 'An entropy-based morphological analysis of river basin networks', *Wat. Resour. Res.,* **29**(4), 1215–1224.
Forte, B. 1969. 'Measures of information: the general axiomatic theory', *R.I.R.O., 3me Annee*, **R/2**, 63–90.
Forte, B. 1972. 'Information and probability: collectors and compositivity', *Symposia Mathematica*, **IX**, 121–129.
Forte, B. 1984. 'Entropies with and without probabilities: application to questionnaires', *Inform. Process. Manage.*, **20**(1), 397–405.
Forte, B. and Cictleo, H. 1979. 'Measures of uncertainty with the local property of order one and random variables', *J. Combin. Inform. System Sci.*, **4**(3), 179–204.
Fougere, P. F., Zawalick, E. J., and Radoski, H. R. 1976. 'Spontaneous life splitting in maximum entropy power spectrum analysis', *Phys. Earth Planet. Interiors*, **12**, 201–207.
Glover, F. 1968. 'Surrogate constraints', *Oper. Res.*, **16**, 741–749.
Goda, T., Tai, S., and Yamane, A. N. 1981. 'Evaluation of thermodynamical efficiency of reverse osmosis process using entropy', *Wat. Res.*, **15**, 1305–1311.
Goodman, J. 1985. 'Structural fragility and principle of maximum entropy', *Struct. Safety*, **3**, 37–46.
Gould, F. J. 1969. 'Extensions of Lagrange multipliers in nonlinear programming', *SIAM J. Appl. Math.*, **13**(6), 1280–1297.
Greenberg, H. J. and Pierskalla, W. P. 1970. 'Surrogate mathematical programming', *Oper. Res.*, **18**, 924–939.
Griffith, D. S. 1972. 'Computer solution of the discrete maximum entropy problem', *Technometrics*, **14**(4), 891–897.
Gull, S. F. 1991. 'Some misconceptions about entropy', in Buck, B. and Macauley, V. A. (Eds), *Maximum Entropy in Action*. Oxford University Press, Oxford.
Gull, S. F. and Skilling, J. 1984. 'The maximum entropy method in image processing', *IEEE Proc.*, **131F**, 646–659.
Harmancioglu, N. B., Alpaslan, N., and Singh, V. P. 1992a. 'Application of the entropy concept in design of water quality monitoring networks', in Singh, V. P. and Fiorentino, M. (Eds), *Entropy and Energy Dissipation in Water Resources*. Kluwer Academic Publishers, Dordrecht. pp. 283–302.
Harmancioglu, N. B., Alpaslan, N., and Singh, V. P. 1992b. 'Design of water quality monitoring networks', in Chowdhury, R. N. (Ed.), *Geomechanics and Water Engineering in Environmental Management*. A. A. Balkema, Rotterdam. pp. 267–297.
Harmancioglu, N. B., Singh, V. P., and Alpaslan, N. 1992c. 'Versatile uses of the entropy concept in water resources', in Singh, V. P. and Fiorentino. M. (Eds), *Entropy and Energy Dissipation in Water Resources*. Kluwer Academic Publishers, Dordrecht. pp. 91–117.
Harmancioglu, N. B., Alpaslan, N., and Singh, V. P. 1994. 'Assessment of the entropy principle as applied to water quality monitoring network design', in Hipel, K. W., McLeod, A. I., Panu, U. S., and Singh, V. P. (Eds), *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, Vol. 3. Kluwer Academic Publishers, Dordrecht. pp. 135–148.
Hart, P. E. 1971. 'Entropy and other measures of concentration', *J. R. Stat. Soc.*, **134**(1), 73–85.
Horibe, Y. 1985. 'Entropy and correlation', *IEEE Trans. Systems, Man, Cybernetics*, **SMC-15**(5), 641–642.
Horowitz, A. R. and Horowitz, I. 1976. 'The real and illusory virtues of entropy-based measures for business and economic analysis', *Decision Sci.*, **7**, 121–136.
Jaworski, W. 1987. 'Higher-order moments and the maximum entropy interference: the thermodynamic limit approach', *J. Phys. A*, **20**, 915–926.
Jaynes, E. T. 1957a. 'Information theory and statistical mechanics, I', *Phys. Rev.*, **106**, 620–630.
Jaynes, E. T. 1957b. 'Information and statistical mechanics, II', *Phys. Rev.*, **108**, 171–190.
Jaynes, E. T. 1965. 'Gibbs versus Boltzmann entropies', *Am. J. Phys.*, **33**, 391–398.
Jaynes, E. T. 1978. 'When do we stand on maximum entropy', in Levine, R. D. and Tribus, M. (Eds), *The Maximum Entropy Formalism*. MIT Press, Cambridge. pp. 15–118.
Jaynes, E. T. 1982. 'On the rationale of maximum entropy methods', *Proc. IEEE*, **70**, 939–952.
Jaynes, E. T. 1984. 'Prior information and ambiguity in inverse problems', *SIAM-AMS Proc.*, **14**, 151–166.
Kam, T. Y. and Brown, C. B. 1983. 'Updating parameters with fuzzy entropy', *J. Engng Mech., ASCE*, **109**(6), 1334–1343.
Kaplan, S. and Garrick, B. J. 1981. 'On the quantitative definition of risk', *Risk Analysis*, **1**(1), 11–27.
Kapur, J. N. 1972. 'Measures of uncertainty: mathematical programming and physics', *J. Indian Soc. Agric. Stat.*, **24**, 47–66.
Kapur, J. N. 1982. 'Maximum-entropy probability distribution for a continuous random variate over a finite interval', *J. Math. Phys. Sci.*, **16**(1), 97–109.
Kapur, J. N. 1983a. 'Derivation of logistic law of population growth from maximum-entropy principle', *Natl Acad. Sci. Lett.*, **6**(12), 429–431.
Kapur, J. N. 1983b. 'Twenty-five years of maximum entropy principle', *J. Math. Phys. Sci.*, **17**(2), 103–156.
Kapur, J. N. and Kesavan, H. K. 1992. *Entropy Optimization Principles with Applications*. Academic Press, Inc., New York. 408 pp.
Kesavan, H. K. and Kapur, J. N. 1989. 'The generalized maximum entropy principle', *IEEE Trans. Systems, Man, Cybernetics*, **19**(5), 1042–1052.
Keyes, J. M. 1921. *A Treatise on Probability*. Macmillan and Co., London.
Korn, G. A. and Korn, J. M. 1968. *Mathematical Handbook for Scientists and Engineers*. McGraw-Hill Book Company, New York.
Krstanovic, P. F. and Singh, V. P. 1987. 'A multivariate stochastic flood analysis using entropy', in Singh, V. P. (Ed.), *Hydrol. Frequency Modeling*. D. Reidel, Dordrecht. pp. 515–539.
Krstanovic, P. F. and Singh, V. P. 1991a. 'A univariate model for long-term streamflow forecasting: 1. Development', *Stoch. Hydrol. Hydraul.*, **5**, 173–188.

Krstanovic, P. F. and Singh, V. P. 1991b. 'A univariate model for long-term streamflow forecasting: 2. Application', *Stoch. Hydrol. Hydraul.*, **5**, 189–205.
Krstanovic, P. F. and Singh, V. P. 1992a. 'Evaluation of rainfall networks using entropy: 1. Theoretical development', *Wat. Resour. Manage.* **6**, 279–293.
Krstanovic, P. F. and Singh, V. P. 1992b. 'Design of rainfall networks using entropy: II. Application', *Wat Resour. Manage.*, **6**, 295–314.
Krstanovic, P. F. and Singh, V. P. 1992c. 'Transfer of information in monthly rainfall series of San Jose, California', in Singh, V. P. and Fiorentino, M. (Eds), *Entropy and Energy Dissipation in Water Resources*. Kluwer Academic Publishers, Dordrecht. pp. 155–173.
Krstanovic, P. F. and Singh, V. P. 1993a. 'A real-time flood forecasting model based on maximum entropy spectral analysis: 1. Development', *Wat. Resour. Manage.*, **7**, 109–129.
Krstanovic, P. F. and Singh, V. P. 1993b. 'A real-time flood forecasting model based on maximum entropy spectral analysis: 2. Application', *Wat. Resour. Manage.*, **7**, 131–151.
Kullback, S. and Leibler, R. A. 1951. 'On information and sufficiency', *Ann. Math. Stat.,* **22**, 79–86.
Kusumulyono, A. and Goulter, I. 1994. 'Entropy principles in the prediction of water quality values at discontinued monitoring stations', *Stoch. Hydrol. Hydraul.*, **8**, 301–317.
Kusumulyono, A. and Goulter, I. 1995. 'Computational aspects in use of entropy theory in predicting water quality levels at discontinued stations', *Stoch. Hydrol. Hydraul.*, **9**, 215–237.
Kvalseth, T. O. 1987. 'Entropy and correlation: some comments', *IEEE Trans. Systems, Man, Cybernetics*, **SMC-17**(3), 517–519.
Landau, H. J. 1987. 'Maximum entropy and the moment problem', *Bull. Am. Math. Soc.*, **16**(1), 47–77.
Larimore, W. E. 1983. 'Predictive inference, sufficiency, entropy and an asymptotic likelihood principle', *Biometrika*, **70**(1), 175–181.
Lavenda, B. H. 1978. *Thermodynamics of Irreversible Processes*. John Wiley & Sons, New York.
Lebowitz, J. L. 1993. 'Boltzmann's entropy and time's arrow', *Physics Today*, **46**(9), 33–38.
Levine R. D. 1980. 'An information theoretical approach to inversion problems', *J. Phys. A*, **13**, 91–108.
Levine, R. D. and Tribus, M. (Eds) 1978. *The Maximum Entropy Formalism*. MIT Press, Cambridge. 498 pp.
Lewis, G. N. and Randall, M. 1961. *Thermodynamics*, 2nd edition, revised by Pitzer, K. S. and Brewer, L. McGraw-Hill, New York.
Li, Y., Singh, V. P., and Cong, S. 1987. 'Entropy and probability distributions', in Singh, V. P. (Ed.), *Hydrologic Frequency Modeling*. D. Reidel Publishing Company, Dordrecht. pp. 367–382.
Lind, N. H. 1989. 'Information theory and estimation of random variables', *Paper No. 15*. Institute for Risk Research, University of Waterloo, Waterloo, Canada. 17 pp.
Lind, N. H. and Chen, X. 1987. 'Consistent distribution parameter estimation for reliability analysis', *Struct. Safety*, **4**, 141–149.
Lind, N. H. and Hong, H. P. 1991. 'Entropy estimation of hydrological extremes', *Stoch. Hydrol. Hydraul.*, **5**, 77–87.
Lind, N. H. and Solona, V. 1988. 'Cross-entropy estimation of random variables with fractile constraints', *Paper No. 11*. Institute for Risk Research, University of Waterloo, Waterloo, Canada. 26 pp.
Linfoot, E. H. 1957. 'An informational measure of correlation', *Inform. Control*, **1**, 85–89.
Liu, I. S. 1972. 'Method of Lagrange multipliers for exploitation of the entropy principle', *Arch. Rational Mech. Anal.*, **46**(2), 131–148.
Marschak, J. 1971. 'Economics of information systems', *Frontiers of Quantitative Economics*. North Holland Publishing Company, Amsterdam. pp. 32–107.
Mead, L. R. and Papanicolaou, N. 1984. 'Maximum entropy in the problem of moments', *J. Math. Phys.*, **25**(8), 2404–2417.
Montroll, E. W. and Shlesinger, M. F. 1983. 'Maximum entropy formalism, fractals, scaling phenomena, and $1/f$ noise: a taile of tails', *J. Stat. Phys.*, **32**(2), 209–230.
Mukherjee, D. and Ratnaparkhi, M. V. 1986. 'On the functional relationship between entropy and variance with related applications', *Commun. Stat. Theory Meth.*, **15**(1), 291–311.
Murphy, D. C. and Hasenjaeger, J. T. 1973. 'Entropy as a measure of decentralization', *Proc. Am. Inst. Decision Sci.*, Boston, Massachusetts. pp. 67–69.
Padmanabhan, G. and Rao, A. R. 1986. 'Maximum entropy spectra of some rainfall and river flow time series from southern and central India', *Theor. Appl. Climatol.*, **37**, 63–73.
Padmanabhan, G. and Rao, A. R. 1988. 'Maximum entropy spectral analysis of hydrologic data', *Wat. Resour. Res.,* **24**(9), 1591–1533.
Philippatos, G. C. and Wilson, C. J. 1972. 'Entropy market risk and the selection of efficient portfolios', *Appl. Econ.*, **4**, 209–220.
Pierce, J. G. 1978. 'A new look at the relation between information theory and search theory', in Levine, R. D. and Tribus, M. (Eds), *The Maximum Entropy Formalism*. MIT Press, Cambridge. pp. 339–402.
Prigogin, I. 1967. *Introdution to Thermodynamics of Irreversible Processes*, 3rd ed. John Wiley & Sons, New York.
Rajagopal, A. K., Teitler, S., and Singh, V. P. 1987. 'Some new perspectives on maximum entropy techniques in water resources research', in Singh, V. P. (Ed.), *Hydrologic Frequency Modeling*. D. Reidel Publishing Co., Dordrecht. pp. 247–366.
Rajski, C. 1961. 'A metric system space of discrete probability distributions', *Inform. Control*, **4**, 371–377.
Rao, A. R., Padmanabhan, G., and Kashyap, R. L. 1980. 'Comparison of recently developed methods of spectral analysis', *Proceedings, Third International Symposium on Stochastic Hydraulics*, Tokyo, Japan. pp. 165–175.
Rosenkrantz, R. D. (Ed.) 1983. *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*. D. Reidel Publishing Company, Boston. 435 pp.
Shannon, C. E. 1948a. 'A mathematical theory of communications, I and II', *Bell System Tech. J.*, **27**, 379–443.
Sahnnon, C. E. 1948b. 'A mathematical theory of communications, III–V', *Bell System Tech. J.* **27**, 623–656.
Shannon, C. E. and Weaver, W. 1949. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, Illinois.
Shore, J. E. 1979. 'Minimum cross-entropy special analysis', *NRL Memorandum Report 3921*. Naval Research Laboratory, Washington D.C.
Shore, J. E. and Johnson, R. W. 1980. 'Axiomatic derivation of the principle of minimum cross-entropy', *IEEE Trans. Inform. Theory*, **IT-26**, 26–37.

Simoes, L. M. C. and Templeman, A. B. 1989. 'Entropy-based synthesis of pretensioned cable net structures', *Engng Optimization*, **15**, 121–140.
Singh, V. P. 1989. 'Hydrologic modelling using entropy', *J. Inst. Eng. Civil Engng Div.*, **70**, 55–60.
Singh, V. P. 1992. 'Entropy-based probability distributions for modeling of environmental and biological systems', in Iyengar, S. S. (Ed.), *Structuring Biological Systems: A Computer Modeling Approach*. CRC Press, Boca Raton. pp. 167–208.
Singh, V. P. and Cruide, J. F. 1990. 'Stochastic streamflow modeling for reservoir planning and management', *IAHS Publ.*, **194**, 11–28.
Singh, V. P. and Fiorentino, M. 1992. 'A historical perspective of entropy applications in water resources', in Singh, V. P. and Fiorentino, M. (Eds), *Entropy and Energy Dissipation in Water Resources*. Kluwer Academic Publishers, Dordrecht. pp. 155–173.
Singh, V. P. and Guo, H. 1995a. 'Parameters estimation for 2-parameter Pareto distribution by POME', *Wat. Resour. Manage.*, **9**, 811–893.
Singh, V. P. and Guo, H. 1995b. 'Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME)', *Hydrol. Sci. J.*, **4**, 165–181.
Singh, V. P. and Guo, H. 1995c. 'Parameter estimation for 2-parameter log-logistic distribution (LLD2) by maximum entropy', *Civil Engng Systems*, **12**, 343–357.
Singh, V. P. and Guo, H. 1997. 'Parameter estimation for 2-parameter generalized Pateto distribution by (POME)', *Stoch. Hydrol. Hydraul.*, in press.
Singh, V. P. and Rajagopal, A. K. 1987. 'Some recent advances in application of the principle of maximum entropy (POME) in hydrology', *IAHS Publ.*, **164**, 353–364.
Singh, V. P., Guo, H., and Yu, F. X. 1993. 'Parameter estimation for 3-parameter log-logistic distribution (LLD3) by POME', *Stoch. Hydrol. Hydraul.*, **7**, 163–177.
Smith, C. R. and Grandy, Jr., W. T. (Eds) 1985. *Maximum Entropy and Bayesian Methods in Inverse Problems*. D. Reidel Publishing Company, Boston.
Snickers, F. and Weibull, J. W. 1977. 'A minimum information principle: theory and practice', *Reg. Sci. Urban Econ.*, **7**, 137–168.
Sogawa, N., Araki, M., and Imai, T. 1986. 'Studies on multivariate conditional maximum entropy distribution and its characteristics', *J. Hydrosci. Hydraul. Engng*, **4**(1), 79–97.
Solona, V. and Arteaga, A. 1987. 'Second order uncertainty modeling over finite and infinite domains using entropy', in Lind, N. C. (Ed.), *Reliability and Risk Analysis in Civil Engineering*. University of Waterloo, Waterloo, Canada.
Solona, V. and Lind, N. C. 1989. 'A monotonic property of distributions based on entropy with constraints', in Skilling, J. (Ed.), *Maximum Entropy and Bayesian Methods*. Kluwer Academic Publishers, Dordrecht. pp. 481–490.
Tai, S. and Goda, T. 1980. 'Water quality assessment using the theory of entropy', in Stiff, M. (Ed.), *River Pollution Control*. Ellis Horwood Limited, Chichester. pp. 319–330.
Tai, S. and Goda, T. 1985. 'Entropy analysis of water and wastewater treatment processes', *Int. J. Environ. Stud.*, **25**, 13–21.
Templeman, A. B. 1989. 'Entropy-based minimax applications in shape-optimal design', in Eschenauer, H. A. and Thierauf, G. (Eds), *Discretization Methods and Structural Optimization-Procedures and Applications*. Springer-Verlag, New York. pp. 335–342.
Templeman, A. B. and Xingsi, L. 1987. 'A maximum entropy approach to constrained nonlinear programming', *Engng Optimization*, **12**, 191–205.
Templeman, A. B. and Xigsi, L. 1989. 'Maximum entropy and constrained optimization', in Skilling, J. (Ed.), *Maximum Entropy and Bayesian Methods*. Kluwer Academic Publishers, Dordrecht. pp. 447–454.
Theil, H. and Leehder, C. T. 1965. 'Tomorrow on the Amsterdam stock exchange', *J. Business*, **38**, 277–284.
Tikochinsky, Y., Tishby, N. Z., and Levine R. D. 1984a. 'Consistent inference of probabilities for reproducible experiments', *Phys. Rev. Lett.*, **52**(16), 1357–1360.
Tikochinsky, Y., Tishby, N. Z., and Levine R. D. 1984b. 'Alternative approach to maximum entropy inference', *Phys. Rev. A*, **30**(5), 2638–2644.
Tribus, M. 1961a. 'Information theory as the basis for thermostatistics and thermodynamics', *J. Appl. Mech.*, **28**, 1–8.
Tribus, M. 1961b. *Thermostatics and Thermodynamics*. D. van Nostrand, New York.
Tribus, M. 1963. 'The use of the maximum entropy in the estimation of reliability', *Recent Developments in Information and Decision Processes*. The Macmillan Company, New York.
Tribus, M. 1969. *Rational Description, Decision and Designs*. Pergamon Press, New York.
Tribus, M., Evans, R., and Crellin, C. 1964. 'The use of entropy in hypothesis testing', *Proceedings, 10th National Symposium on Reliability and Control*. pp. 579–590.
Turkman, K. F. 1985. 'The choice of extremal models by Akaike's information criterion', *J. Hydrol.*, **82**, 307–315.
Ulrych, T. and Clayton, R. W. 1976. 'Time series modeling and maximum entropy', *Phys. Earth Planet. Interiors*, **12**, 188–199.
van Campenhout, J. M. and Cover, T. M. 1981. 'Maximum entropy and conditional probability', *IEEE Trans. Inform. Theory*, **IT-27**, 483–489.
Verdugo Lazo, A. C. G. and Rathie, P. N. 1978. 'On the entropy of continuous probability distributions', *IEEE Trans. Inform. Theory*, **IT-24**(1), 120–122.
Wehri, A. 1978. 'General properties of entropy', *Rev. Modern Phys.*, **50**(2), 221–260.
Whitaker, E. T. and Robinson, G. 1944. *The Calculus of Observation*, 4th edn. Blackie, Glasgow. p. 341
Woodbury, A. D. and Ulrych, T. J. 1993. 'Minimum relative entropy: forward probabilistic modeling', *Wat. Resour. Res.*, **29**(8), 2847–2860.
Wragg, A. and Downson, D. C. 1970. 'Filtering continuous probability density functions over $[0, \infty)$ using information theory ideas, *IEEE Trans. Inform. Theory*, **IT-16**, 226–230.
Xingsi, L. and Templeman, A. B. 1988. 'Entropy-based optimum sizing of trusses', *Civil Engng. Systems*, **5**, 121–128.
Yang, C. T. 1994. 'Variation theories in hydrodynamics and hydraulics', *J. Hydraul. Engng*, **120**(6), 737–756.
Yang, Y. and Burn, D. H. 1994. 'An entropy approach to data collection network design', *J. Hydrol.* **157**, 307–324.
Zurek, W. H. 1989. 'Algorithmic randomness and entropy', *Phys. Rev. A*, **40**(8), 4731–4751.

