ON WHAT CAN BE EXPLAINED BY THE ENTROPY OF A CHANNEL NETWORK

M. FIORENTINO AND P. CLAPS
Dipartimento di Ingegneria e Fisica dell'Ambiente
Università della Basilicata
Potenza, 85100 - Italy

ABSTRACT. In this paper a measure of the entropy associated with a channel network is defined, according to the Shannon informational definition, as the expectation of (-log p_j), where p_j is the ratio of the existing paths at the bifurcation level j to the total number of paths. Here j and p_j are respectively proportional to the arrival time of water to the network outlet and to the number of water parcels arriving from the distance j. Then, the expression for the channel network entropy proposed in this paper is well suited for hydrologic purposes. By analyzing a river network (and related sub-networks) of an Italian basin with surface area of 123 km^2, it is shown that the network entropy is strictly related to basin characteristics such as average elevation, Horton order, and magnitude.

1. Introduction

The theory of entropy developed in thermodynamics and statistical mechanics as well as in communication and information sciences has recently been found to have wide ranging application in hydrology and water resources [for a review see Singh, 1989]. In this paper the attention is focused on the entropy of river networks which is thought to be of a great interest for understanding the basin scale hydrologic response.

Runoff from a drainage basin occurs as a result of expenditure of energy. Quite naturally, the basin tends to spend the least amount of energy to produce that runoff. Producing the runoff hydrograph is the network of the basin.

The links between concepts of energy dissipation and entropy, together with the related impact on the river flow, were first investigated by Leopold and Langbein [1962] who, by invoking a thermodynamic analogy for a river system along with the principle of least work, found some general rules affecting the hydraulic geometry of river channels. The same concepts were revisited by Yang [1971] who introduced the law of least rate of energy expenditure which states that during the evolution toward its equilibrium condition a natural stream chooses its course of flow in such a manner that the rate of potential energy expenditure per unit mass of water along this course is minimum. By considering the principle of maximum entropy equivalent to that of minimum rate of potential energy expenditure, Yang shed more light on some geomorphologic features of river systems.

In both Leopold and Langbein's [1962] and Yang's [1971] papers the river system to which the general laws of entropy and energy are applied were considered to be one-dimensional. In their work, entropy was used as a powerful concept to investigate distributive properties of some energy-related characteristics of the river but no measurable quantity was proposed to evaluate the entropy from data.

V. P. Singh and M. Fiorentino (eds.), Entropy and Energy Dissipation in Water Resources, 139–154.

In this paper an effort is made to seek this quantity in order to investigate the way it is affected by plano-altimetric variations in the river basin configuration.

The opportunity is given by the link existing between the thermodynamic entropy and the *informational entropy* proposed by Shannon [1948], which will be shown in the next section.

The way entropies vary from one stage of the system, e.g. network magnitude, basin order, mean elevation, etc. to another, produces insights into the knowledge of runoff evolution. In fact, the entropy accounts for the flow of information that the system has used to reach the current level of organization. This merits further investigation for, as a guess, one may think of water discharge as the main source of information used by the river system to adjust its configuration.

2. Informational, Primary and Thermodynamic Entropies of the Channel Network

In Thermodynamics, entropy of a system is defined in the difference form as

$$dS' = \frac{dQ}{T} \tag{1}$$

where Q is the thermal energy of the system and T is its absolute temperature. When the system changes from state σ_1 to state σ_2 the change in entropy is given by the curve integral

$$S' = \int_{\sigma_1}^{\sigma_2} \frac{dQ}{T} \tag{2}$$

According to Boltzmann, the entropy S of a system is proportional to the logarithm of the relative probability of its state:

$$S' = k \log P \tag{3}$$

where k is Boltzmann's constant, being the number of states σ_1 , σ_2 , ... (in phase space) each molecule of the system can occupy.

Let us think of a river network as an isolated system consisting of M=2n-1 links (molecules), each of which can occupy one of the states d = [1,2, ... , D], d being the topological distance of the link down node from the outlet. The network diameter D, as is the maximum value of d, is the total number of states. The number of ways to achieve a state of the system in which there are M_d links in state d = 1,2, ..., D is

$$g = \frac{M!}{M_1!\, M_2! \,.......\, M_D!} = \frac{M!}{\prod_{d=1}^{D} M_d!} \tag{4}$$

For large M, using Stirling's approximation $\log M! \sim M \log M - M$, one can get

$$\log g = M \log M - \sum_d M_d \log M_d \tag{5}$$

As g is proportional to the relative probability of the state, the entropy of the state is proportional to log g:

$$S' \sim M \log M - \sum_{d} M_d \log M_d = -M \sum_{d=1}^{D} p_d \log p_d \tag{6}$$

where $p_d = M_d / M$ is the probability of a given link being in the state d. The average entropy per link of the state of the system is given by S' / M.

We define this unit average entropy as the *Informational Entropy* of the river *Network* (IEN), which we refer to as S:

$$S = -\sum_{d=1}^{D} p_d \log p_d \tag{7}$$

The entropy, as given in equation (7), was also proposed by Shannon [1948] in Information theory. It is taken as a measure of the amount of information transmitted per symbol on the average, whereby a symbol is any of the D elements of the string transmitted by the information source and p_d is its probability, given each symbol is transmitted independently. If the logarithms are taken to the base 2 then the information is measured in terms of binary digits (*bits*), and if they are taken to the base *e* then the base is in *nats*. In the following we will use natural logs.

The previous brief discussion provides an idea of the linkage existing between thermodynamical, statistical-mechanical and informational entropies. It suggests, as regards the river network case, that the distribution of links between topological levels is related to the thermodynamical processes involved with the system history. Also, entropy is a measure of the probability of state of the system as it provides an estimate for $\langle -\log p \rangle$ which is a monotonically increasing function of p. This implies that the higher the entropy, the higher the probability of the system of being in that state.

Entropy, as given in equation (7), can be maximized with respect to several kinds of constraints to provide the least biased distribution of p_d's according to the information given by the constraints. If we only consider the obvious constraint

$$\sum_{d=1}^{D} p_d = 1 \tag{8}$$

by maximizing S, we get $p_d = 1/D$, $d = 1, 2, \ldots, D$, which yields

$$S = \log D \tag{9}$$

This quantity is called *Primary Entropy* of the *Network* (PEN) [Kapur, 1990] in order to underline its significance as maximum value in absence of natural constraints.

Indeed, in the real world, river networks never show the same number of links at any topological level, hence equation (9) is never strictly true. However, principally due to the flatness of S in a wide range around its maximum, equation (9) provides a good approximation to S in most cases, such that log D can be assumed as an efficient estimator of the entropy of the river network.

In the following, equation (7) will be used to analyze the relationships which relate, on average, variations of different characteristics of the drainage network.

2.1. ENTROPY AND POTENTIAL ENERGY

In thermodynamics it is possible to associate with the state σ_i of the system the thermal energy E_i, which is in turn related to the probability p_i of state σ_i . Maximizing S, given as in equation (7), under the constraints

$$\sum_{i=1}^{k} p_i E_i = E \tag{10}$$

$$\sum_{i=1}^{k} p_i = 1 \tag{11}$$

of which the (10) is the total energy of the system, yields the Maxwell-Boltzmann distribution, in which

$$p_i = \frac{e^{-E_i/kT}}{\sum_j e^{-E_j/kT}} \tag{12}$$

In the dynamics of a river network the potential energy is the one which plays the most relevant role. This energy is given, in the appropriate scale unit, by the elevation of network nodes above a datum. As it is reasonable that the current state of the river network is the result of its history from the infinite past, it is noteworthy to evaluate the entropy of the system as the most probable state under the constraint provided by the mean node elevation of the network.

Let $<y_d>$ be the mean node elevation of the upstream node of each link being in the topological level d and $<y>_D$ be the mean node-elevation evaluated throughout the network (potential energy of the river network system) of diameter D. Maximizing S, as given in equation (7), under the constraints

$$\sum_{d=1}^{D} p_d <y_d> = <y>_D \tag{13}$$

$$\sum_{d=1}^{D} p_d = 1 \tag{8}$$

gives, in analogy with equation (12),

$$p_d = \frac{e^{-<y_d>/DT'}}{\sum_j e^{-<y_j>/DT'}} \tag{14}$$

in which T ' is assumed to be the (degenerate) temperature of the network. In principle, it can be thought of as a parameter whose definition will arise from the following derivations.

Re-parameterizing equation (14) yields:

$$<y_d> = -\alpha \log (\beta\, p_d) \tag{15}$$

where

$$\alpha = DT' \tag{16}$$

$$\beta = \sum_{d=1}^{D} e^{-<y_d>/\alpha} \tag{17}$$

Taking the mean of equation (15) we get

$$<y>_D = \sum_{d=1}^{D} p_d <y_d> \;=\; -\alpha \sum_{d=1}^{D} p_d \log(\beta\, p_d) \tag{18}$$

and, by inserting equation (7)

$$<y>_D = -\alpha \log (\beta) + \alpha\, S \tag{19}$$

Due to the generality of the hypotheses underlying equation (19), it should hold for any sub-network. The α and β parameters, being dependent upon D and T ', might in principle vary from one sub-network to another. However, for the system as a whole, the distribution of potential energy within the system is constrained by two fundamental quantities D and T ' measured with respect to the entire system. This is reasonable as, in analogy with thermodynamics, T ' can be thought of as proportional to the energy content of the system. The entropy, as defined in equation (7), calculated with probabilities defined in equation (14) is called [after Kapur, 1990] *Thermodynamic Entropy*. Under these assumptions, α and β parameters are easily derived as follows.

If we define $<y>_\delta$ the mean of the elevations of nodes owing to a subbasin with topological diameter δ, at the source of each channel we have $<y>_\delta = 0$. We also have S=0 in sources, hence equation (19) will give $\log \beta = 0$. This means $\beta = 1$ and

$$<y>_\delta = \alpha\, S \tag{20}$$

The α parameter can be eliminated by use of equation (9) to obtain

$$<y>_\delta = \frac{<y>_D}{\log D} S \qquad (21)$$

and

$$<y>_\delta = <y>_D \frac{\log \delta}{\log D} \qquad (22)$$

Finally, In following equations (16), (20) and (21), we get:

$$T' = \frac{<y>_D}{D \log D} \qquad (23)$$

which gives the quantity that acts as the (degenerate) temperature of the network. It is proportional to the energy of the system, as in thermodynamics. The multiplying coefficient is inversely proportional to a monotonically increasing function of D.

3. River Profiles

The longitudinal profile of the channel elevations can be derived by exploiting equations (9) and (19). If one consider that, in a channel network of diameter δ with uniform width function and links elevation fall, the total elevation fall H_δ from the upstream node to the outlet is twice the mean elevation $<y>_\delta$ of the entire network (relative to tho outlet elevation, as we will refer to hereafter), we can introduce a relationship

$$H_\delta = \frac{\delta \; <y>_\delta}{<d>_\delta} \qquad (24)$$

which can be regarded as a good approximation for any general case. In the (24), $<d>_\delta$ is the topological distance of the centroid of the sub-basin width function from its outlet.

Thus, by multiplying both sides of equation (19) by $\delta \, / <d>_\delta$, one get

$$H = -\alpha' \log \beta \; + \alpha' S \qquad (25)$$

in which H is an estimate of the elevation fall from the source to the outlet of the main channel of the network whose entropy is S, and α' is given by multiplying α as in equation (16) for the network diameter divided by $<d>_\delta$.

The α' and β parameters can be evaluated as in the case of equation (19) (see also equation (20) and (21)). Thus, equation (25) becomes

$$H = \alpha' S = \frac{H_D}{\log D} S \qquad (26)$$

Let y_0 be the elevation of the upstream node of the channel and y_δ be the elevation of the downstream node being at a distance δ from the source. Then, by substituting S via equation (9), the elevation profile of the channel of topological length D is given by

$$y_\delta = y_0 - \frac{H_D}{\log D} \log \delta \ ; \qquad \delta \geq 1 \tag{27}$$

4. Entropy, Strahler Order, Magnitude and Fractal Dimension

Let a network have Strahler order Ω. Then a Strahler channel (or stream) of order ω, $2 \leq \omega \leq \Omega$, is defined as the sequence of links of order ω whose last link drains either into a link of order ω' > ω or into the outlet. Let R_B and R_L respectively the bifurcation and the stream length ratio:

$$R_B = N(\omega\text{-}1; \Omega) \,/\, N(\omega; \Omega) \tag{28}$$

$$R_L = L(\omega; \Omega) \,/\, L(\omega-1; \Omega) \tag{29}$$

where N(ω; Ω) is the number of streams of order ω in the basin network of order Ω and L(ω; Ω) is the related average length.

Since scaling in length does not affect R_L , one can think of R_L as the ratio of the number of links of order ω to the number of links of order ω-1. The number of first order streams N(1;Ω) is said to be the magnitude of the network and it will be hereafter referred to as *n*. The total number of links in the network is M=2*n*-1. As the network order increases, the uncertainty about the topological level each link can occupy increases, so that one expects the entropy per link, as defined in equation (7), to increase.

In this section the relationship between the expected value of S and the network order is analyzed. The implications of the derived laws are also discussed. To this end we consider the entropy in the *primary* form S = ln D. For a Hortonian network we have

$$D = \sum_{\omega=1}^{\Omega} R_L^{\,\omega-1} \tag{30}$$

This expression corresponds to a geometric progression of base R_L , which can be expressed in the form

$$D = \frac{1 - R_L^{\,\Omega}}{1 - R_L} \tag{31}$$

which implies

$$\ln D = \ln \left[\frac{R_L^{\ \Omega} - 1}{R_L - 1} \right] \qquad (32)$$

Letting Ω tend to infinity, the member at right hand side of (32) becomes:

$$\lim_{\Omega \to \infty} \left[\ln (R_L^{\ \Omega} - 1) - \ln (R_L - 1) \right] = \Omega \ln (R_L) - \ln (R_L - 1) \qquad (33)$$

This expression provides a direct linear relationship between S and Ω, in which the Strahler order can be thought of as a measure of the network complexity:

$$S = \Omega \ln (R_L) - \ln (R_L - 1) \qquad (34)$$

It is noteworthy to point out that an increase in entropy from order Ω_1 to Ω_2, $(\Omega_2 > \Omega_1)$ accounts for a reduction in availability of energy of the system or, in other words, for the work done by the system to evolve from state Ω_1 to state Ω_2. This is in agreement with equation (26) in which an increase of the entropy of the network corresponds to a reduction of the outlet elevation and hence to a loss of availability of potential energy.

Magnitude (n) is perhaps the most important characteristic of the network as it can serve as a surrogate for the watershed area. The relationship between S and n is derived by equation (34), given

$$n = R_B^{\Omega - 1} \qquad (35)$$

$$\Omega = 1 + \frac{\ln n}{\ln R_B} \qquad (36)$$

and

$$S = \ln n \frac{\ln R_L}{\ln R_B} + \ln \left[\frac{R_L}{R_L - 1} \right] \qquad (37)$$

It is noteworthy to note that equations (9) and (37) yield

$$D = C\, n^{1/F} \qquad (38)$$

where :

$$F = \frac{\ln R_B}{\ln R_L} \qquad C = \frac{R_L}{R_L - 1} \qquad (39)$$

The scaling coefficient F can be considered as the topological expression of the fractal dimension of the network, given

$$\frac{D}{D_1} = \frac{\ln R_B}{\ln R_L} \tag{40}$$

[i.e. Tarboton et al., 1990] in which D and D_1 are the fractal dimensions of the entire network and of the main stream respectively. If $R_B = 4$ and $R_L = 2$ we have $F = 2$ and $C = 2$.

By substituting equation(39) in (37) one obtains

$$S = \ln\left[\frac{R_L}{R_L - 1} n^{1/F}\right] \tag{41}$$

5. Comparison With Real World Data

In order to assess the theoretical assumptions and to check the validity of the interconnections which entropy seems to show with the Strahler order, elevation and topological diameter, a natural channel network has been taken under consideration, namely, the River Arcidiaconata, located in southern Italy, whose characteristics are described in table 1. The topological width function of the channel network (no. of streams vs. topological distance) is showed in figure 1.

table 1. General data of the River Arcidiaconata channel network [from Copertino et al., 1991].

Surface (km^2)	Perimeter (km)	Drainage Density (km^{-1})	1st. Order Streams Freq.
123.9	59.5	2.24	2.05
Mean Topol.Dist.	Magnitude	R_B	R_L
25.08	254	4.12	2.35
2nd Order Streams	3rd Order Streams	4th Order Streams	5th Order Streams
66	17	3	1
Mean Elevation (m a.s.l.)	Maximum Elevation (m a.s.l.)	Minimum Elevation (m a.s.l.)	
538	894	237	

The hypothesis of validity of the primary entropy is tested in figure 2, in which entropies, calculated for each link through equation (7), are compared with the logarithm of the topological distance d . The comparison is reasonably good with respect to other major causes of approximation, yet the deviation between the two quantities is systematically increasing. It is noteworthy that the figure under discussion resembles very much the deviation between the observed and theoretical mainstream profiles computed by Yang [1971] under the hypothesis of uniform drop of Strahler streams.

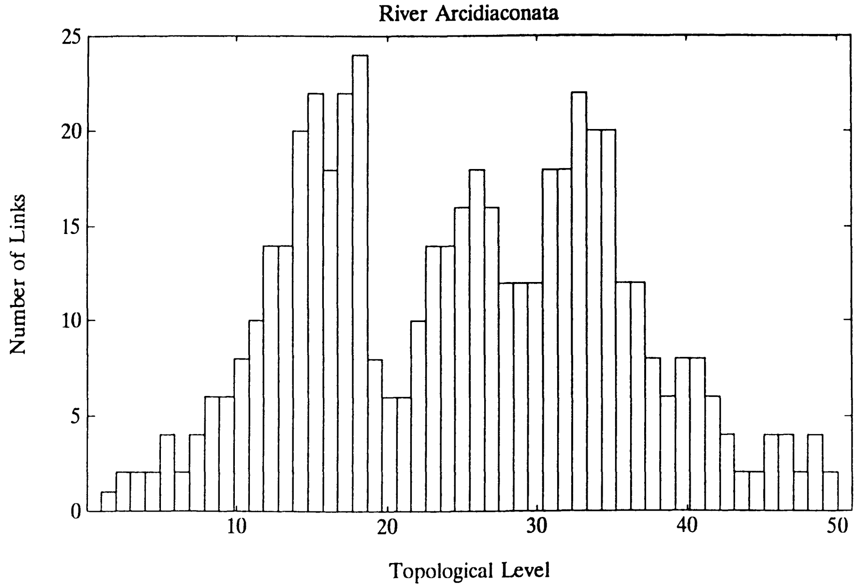

Figure 1. River Arcidiaconata. *Width function* (number of links at same topological distance from the outlet vs. topological distance).

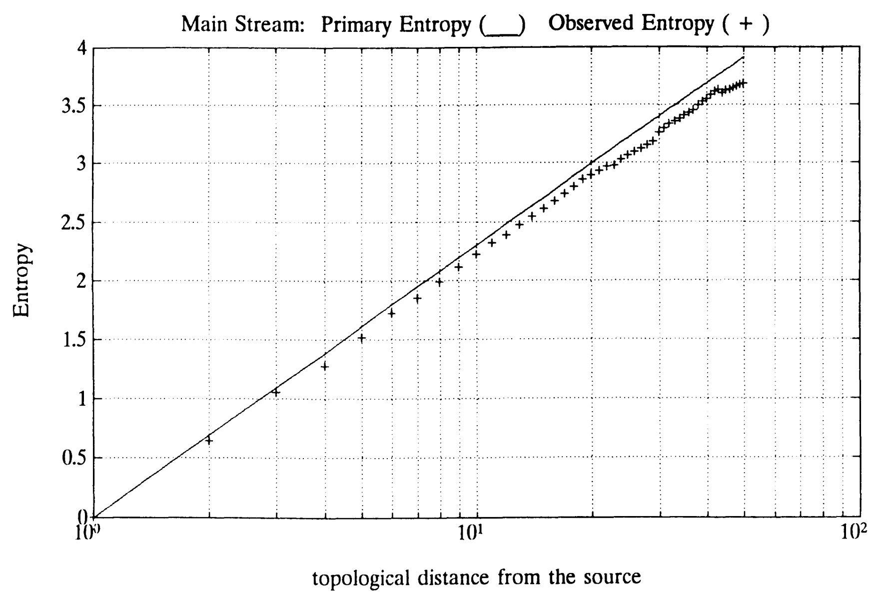

Figure 2. River Arcidiaconata. Deviation of observed entropies from theoretical primary entropy.

The goodness of fit of equation (22) to observed data is shown in figure 3, which refers to the main (longest topologically) channel. It is recognized that the theoretical equation accounts well for the actual mean elevation trend.

The theoretical profile of the main stream, obtained by equation (27) is shown in figure 4, in comparison with the actual one. It is worthwhile remarking the closeness of the two profiles, amongst which the theoretical one is based upon the knowledge of the maximum elevation and the total drop of the stream only. An even better behavior is displayed by the profile computed by mean of the calculated Informational Entropy, as shown in figure 5. A linear regression on the pairs {Informational Entropy, Actual Elevation} can be seen (figure 6) to display the same coefficients of the theoretical relationship (27) .

The last two variables with which entropy has been related are the Strahler order and the network magnitude. In figure 7 the calculated informational entropies averaged with the order (whose values are reported in table 2 compared with theoretical curves (equation 32) referred to as the primary entropy approximation.

table 2. Average entropies of links owing to streams of different Horton-Strahler orders.

Strahler Order	1	2	3	4	5
Avg. Entropy	0	0.979	1.77	3.10	3.68

Figure 8 shows how the theoretical curve $S(n)$ varies for different F and R_L (equation (41)) while in figure 9 actual data of magnitude of each link of the network with $n > 10$ are compared with the regression power law suggested by the equation (38), which is an asymptotically derived one.

6. Conclusions

Theoretical derivations, based upon general entropy laws, provide explanation for well-known empirical observations, are consistent with the results of other theories and produce insights into the understanding of yet unaccounted for river-basin characteristics. In particular, the informational entropy of the network is shown to be related to the thermodynamic entropy and, in turn, it accounts for the distribution of potential energy throughout the network. Empirical measures of the organization level of the network, such as the Strahler order, are shown to be strictly related to entropy.

The primary (unconstrained) entropy is shown to account for most of the informational (observed) entropy. Yet, discrepancy between them tends to increase with the length of the network, thus suggesting that the effect of the available potential energy tends to be more significant as the network grows.

The results of the study suggest the use of the informational entropy of a river network as significant quantity for further understanding the hydrologic behavior of river systems.

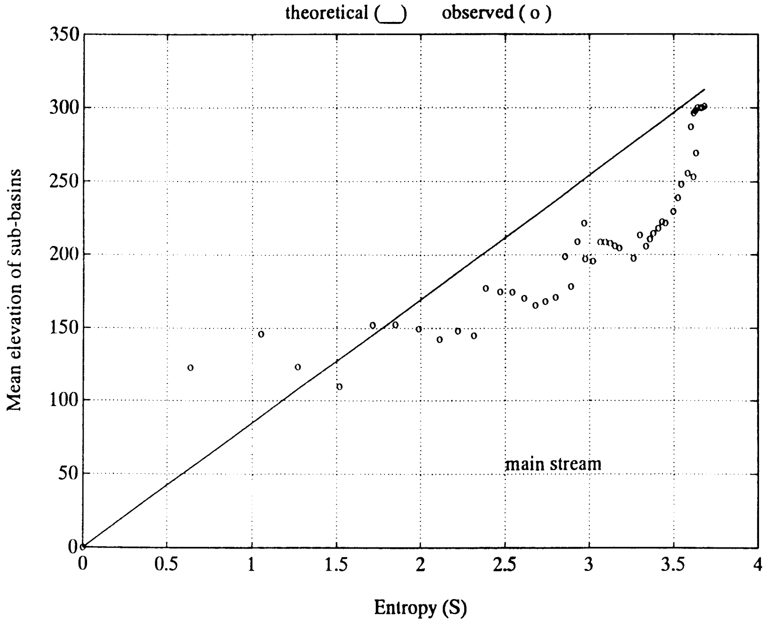

Figure 3. River Arcidiaconata. Observed and theoretical dependence between informational entropy and mean relative elevation of sub-catchment with varying topological diameter (main stream)

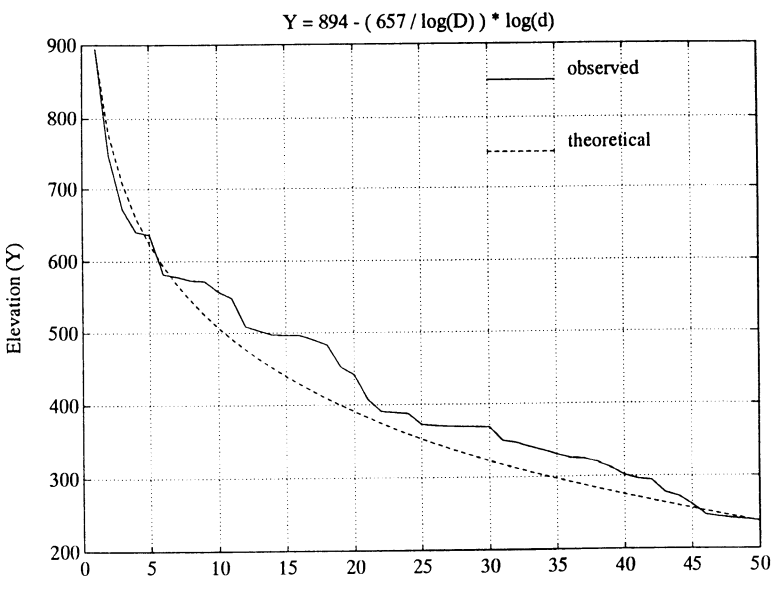

Figure 4. River Arcidiaconata. Actual and theoretical (primary entropy) elevation profile of the main stream.

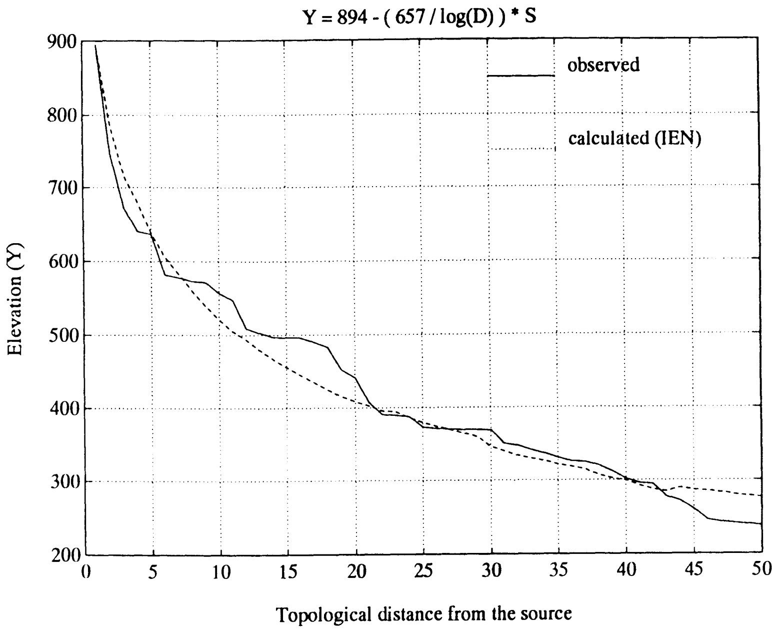

Figure 5. River Arcidiaconata. Actual and calculated (Informational Entropy) elevation profile of the main stream.

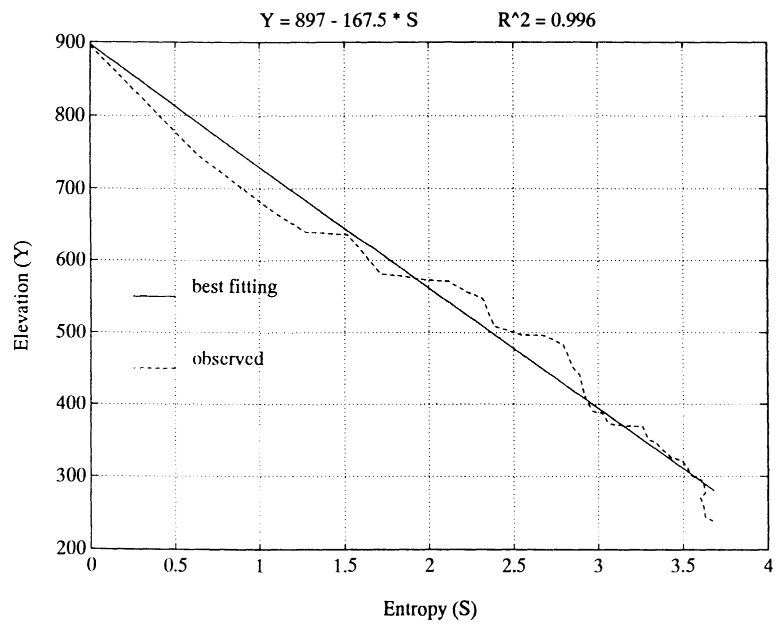

Figure 6. River Arcidiaconata. Linear regression between Elevation and Informational Entropy for the main stream.

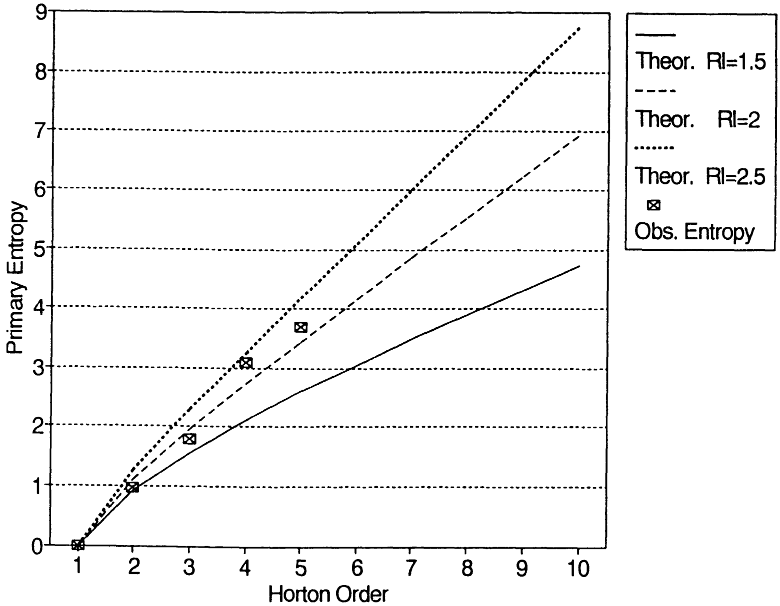

Figure 7. River Arcidiaconata. Theoretical relationships between Primary Entropy and Strahler Order for different values or R_L and calculated IEN.

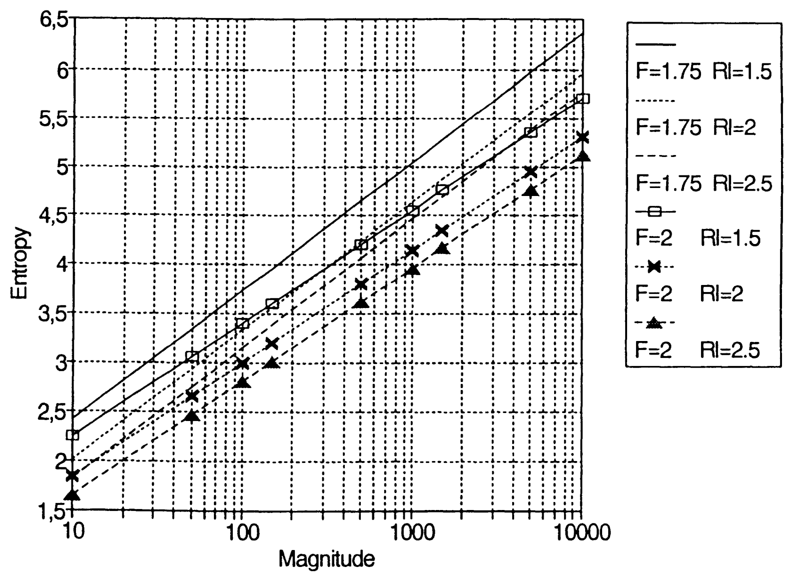

Figure 8. River Arcidiaconata. Theoretical Relationship between Primary Entropy and Magnitude for different values or R_L and F .

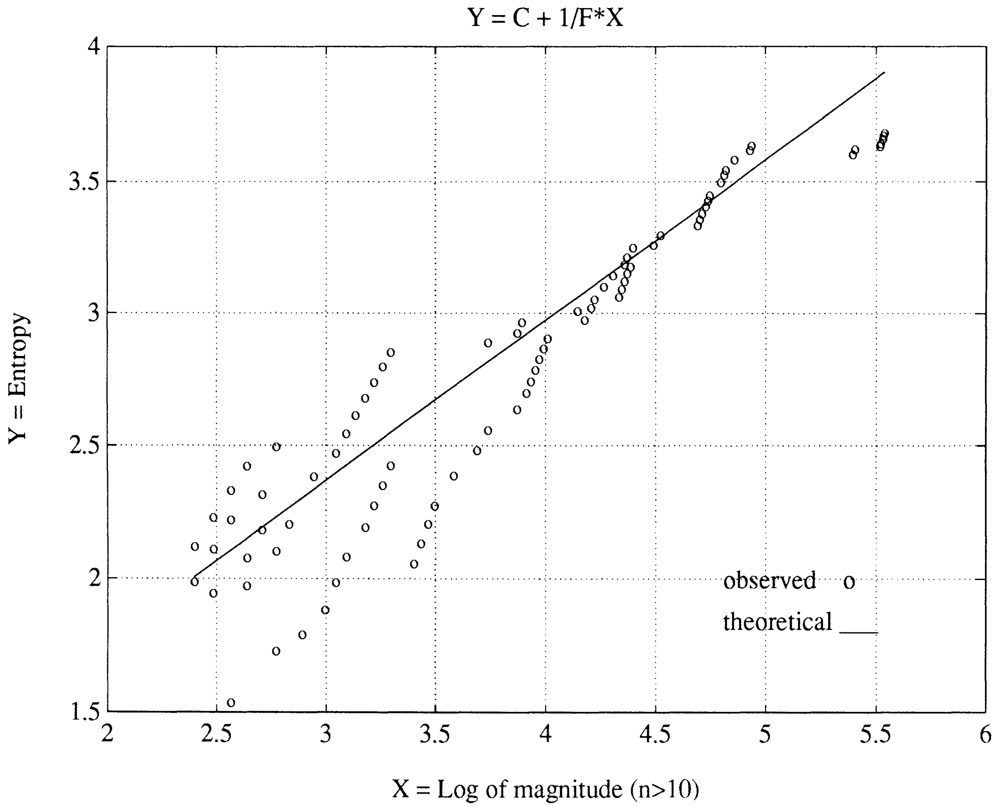

Figure 9. River Arcidiaconata. Theoretical Relationship between Primary Entropy and Magnitude for $n > 10$ (asymptotic law) and calculated average IENs.

Acknowledgements

This work is supported by funds granted by Ministero della Ricerca Scientifica e Tecnologica, Progetto 60% "Criteri di massima entropia in idrologia e idraulica" and Progetto 40% "Fenomeni di trasporto nel ciclo idrologico".

7. References

Copertino, V.A., M. Fiorentino, A.Sole, A.Valanzano, Organizzazione del Sistema Informativo dei Bacini Idrografici Pugliesi (SIBIP), in F. Rossi (Ed.), *Previsione e prevenzione degli eventi idrologici estremi e loro controllo,* Linea 1. Rapporto 1989. GNDCI-CNR, Roma, 1990 (in italian).

Kapur, J.N., *Maximum-Entropy Models in Science and Engineering,* John Wiley, New York, 1990

Leopold, L. B. and W.B. Langbein, The concept of entropy in landscape evolution, *U.S. Geol. Survey Prof. Pap.* 500-A, 1962.

Prigogine, I., *Introduction to Thermodynamics of Irreversible Processes*, John Wiley, New York, 144 p., 1967.

Rinaldo, A., Marani, A. and R.Rigon, Geomorphological dispersion, *Water Resources Research*, **27** (4), 513-525, 1991.

Shannon, C.E., A mathematical theory of communications, *Bell System Technol. Jour.*, vol. 27, 1948.

Singh, V.P., Hydrologic modelling using entropy, *Journal of the Institution of Engineers*, vol 70, Part CV2, pp. 55-60, 1989

Tarboton, D.G., R.l. Bras and I. Rodriguez Iturbe, Comment on " On the fractal dimension of stream networks" by La Barbera, P. and R.Rosso, *Water Resources Research*, **26** (9), 2243-2244, 1990.

Yang, C.T., Potential energy and stream morphology, *Water Resources Research*, **7** (2), 311-322, 1971.

8. Notation

- d topological distance of the link down node from the outlet
- D network diameter (maximum value of d)
- E_i potential energy of the state i
- k Boltzmann's constant
- H_δ total elevation fall from the upstream node to the outlet of a channel of diameter δ
- H total elevation fall from the source to the outlet of the main channel of the network
- $L(\omega; \Omega)$ average length of order ω in the basin network of order Ω
- n Magnitude of the network (number of first order streams)
- $N(\omega; \Omega)$ number of streams of order ω in the basin network of order Ω
- $N(1;\Omega)$ (n) number of first order streams (magnitude of the network)
- P_d probability of a given link being in the state d
- Q thermal energy
- R_B ,R_L bifurcation and stream length ratio
- S ' thermodynamic entropy
- S network entropy
- T temperature
- T ' degenerate temperature of the network
- y_0 elevation of the upstream node of the channel
- y_δ elevation of the downstream node being at a distance δ from the source
- $<y_d>$ mean node elevation of the upstream node of each link being in the topological level d
- $<y>_D$ mean node elevation evaluated throughout the network of diameter D
- $<y>_\delta$ mean of the elevations of nodes owing to a subbasin with topological diameter δ
- α, β coefficients of the law ($<y_d>$, p_d)
- σ system state
- ω Strahler order of a generic channel (or stream)
- Ω Strahler order of the network