Stochastic Hydrology and Hydraulics 9, 215-237 © Springer-Verlag 1995

Computational aspects in use of entropy theory in predicting water quality levels at discontinued stations

A. Kusmulyono and **I. Goulter**
University of Central Queensland, Rockhampton, Queensland 4702, Australia

Abstract: The computational aspects of using a new, entropy-based, theory to predict water quality values at discontinued water quality monitoring stations are discussed. The main computational issues addressed are the level of discretization used in converting the continuous probability distribution of water quality values to the discrete levels required for the entropy function, and the choice of the interval of time for which to assign the value of the water quality (period of time averaging) through the entropy function. Unlike most cases of entropy applications involving discretization of continuous functions the results of using entropy theory to predict water quality values at discontinued monitoring stations in this application appear to be insensitive to the choice of the level of discretization even down to the very coarse level discretization associated with only eight intervals. However, depending on the length of record available the choice of the time interval for which the water quality values are assigned (period for time averaging) appear to have a significant impact on the accuracy of the results.

Key words: Computation, discretization, entropy, networks, time averaging, water quality.

1 Introduction

The entropy concept as a measure of information or a measure of uncertainty was first introduced by Shannon (1948). Shannon's concept of a measure of uncertainty is based on the principle that the greater the uncertainty about the outcomes, the more uniform should be the probabilities assigned to the outcomes. Expressed in entropy specific terms, the greater the uncertainty about the outcome of a process, the greater the value of the entropy of that process. In a later study by Jaynes (1957), the concept of entropy was applied, through the maximization of the entropy value, to the assignment of probabilities to events.

Knowledge about the distribution of values to which the probabilities were to be assigned was introduced by the incorporation of constraints into the entropy maximization formulation. This concept is known as the Principle of Maximum Entropy (POME) and is based on the theory that the probability distribution with the greatest entropy value is the probability distribution that can be realized in the greatest number of ways. In other words, it is the most likely probability distribution.

It has been shown by Singh et al. (1986) that the POME can be used to develop a range of types of probability distribution by specifying different constraints specific to the probability distribution. The probability distributions so developed using the entropy method are always ensured of having the most unbiased probability distribution consistent with the information specified within the constraints.

The advantages of using the principle to develop the probability distribution is that, if there is any additional information about the events, that information can always be introduced in the formulation through additional constraints. Redundancy of the constraints is also not an issue, since the technique will eliminate the redundant information expressed in the constraints through its algorithmic processes. The only problem that can arise occurs in the case of conflicting information, wherein the technique cannot solve the algorithm and therefore, would not give any results (Jaynes, 1968)

The POME has also been applied in many different disciplines of research. Examples of the use of the POME in different fields of study have been identified by Kapur (1983) as statistical inferences, non-linear spectral analysis, transportation models, population density models, brand switching in marketing, finance insurance and marketing, image reconstruction, pattern recognition, operations research and engineering, some biological medical and technological problems, and non-parametric density estimation.

The use of the entropy principle in water resources engineering has developed relatively recently. An early application of the entropy principle in hydrologic frequency analysis was reported by Sonuga (1972). In that study the entropy principle was used to develop a minimally biased probability distribution appropriate for hydrologic frequency analysis where only small amounts of data are available. In subsequent study, Sonuga (1976) applied the entropy principle to the rainfall-runoff process.

The entropy concept was also used by Harmancioglu (1984) to determine the optimal sampling intervals in water quality monitoring networks. Another application of the entropy principle was in modelling the velocity distribution in open channel flows (Chiu, 1987, 1988, 1989, 1991; Chiu and Chiou, 1986).

Other uses of the entropy principle in water resources have been reported; evaluation of information transfer between hydrologic processes (Harmancioglu and Yevjevich, 1987), assessment of the recharge system for a river basin (Harmancioglu and Baran, 1989), redundancy measures in water distribution network design (Awumah et al., 1990,1991; Awumah and Goulter, 1992), and water quality monitoring network design (Harmancioglu and Alpaslan, 1992).

One of the major practical difficulties encountered in the application of entropy theory is the need to approximate the continuous probability distribution function by a discrete function so it can be analyzed by the basic entropy function of $\Sigma X_i \ln X_i$. This discretization was shown by Harmancioglu et al. (1985) to be critical to the value of the entropy provided by the analysis and also have the potential to change the decision arising from the entropy based analysis. In this study the impact of the level of discretization of the continuous function on the result obtained from use of the discrete entropy function, in the form of MDI, as employed by Kusmulyono and Goulter (1994) to predict water quality values at discontinued water quality monitoring stations is examined. The changes in the MDI predicted values of water quality at upstream locations arising from variation in the length of time used to identify changes in water at a downstream locations was also examined.

The data used to examine the effects of each of these two computational aspects of the use of the entropy theory to predict water quality values were taken from the Fitzroy River basin in Central Queensland as shown in Figure 1. The data used in the study are summarized in Table 1.

2 Definition of problems in the use of the entropy principle to predict water quality values at discontinued monitoring stations

One of the aspects in the use of entropy to predict water quality values at unmonitored locations (in this case discontinued stations) noted earlier is the need to approximate the 'true' continuous probability distribution functions by a discrete approximate function. In the study of Kusmulyono and Goulter (1994), the more general form of the POME, which is known as Kullback-Leibler's principle of minimum discrimination information (the MDI principle), was applied to develop the most unbiased probability distribution of water quality values at discontinued sampling stations on the tributaries and main channel upstream of an existing continuing water quality monitoring station. The prediction itself was undertaken knowing the historical probability distribution at each discontinued station, observed water quality values at the downstream continuing sampling station, and an expression conveying information related to the historical relationship between the water quality value at the downstream station and the water quality values at the upstream tributary and main channel stations. The new unbiased probability distributions derived from this approach for each upstream discontinued locations were then used to estimate the mean water quality at those same upstream locations.

The mathematical formulation employed to predict the water quality values at these upstream stations is:

$$\text{Max } H = -\sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij} \ln\,[p_{ij}/(q_{ij}/m)] \tag{1}$$

Subject to:

$$\sum_{i=1}^{n} p_{ij} = 1/m \qquad (j = 1, 2, ..., m) \tag{2}$$

$$\sum_{i=1}^{n} p_{ij}x_{ij}/\sum_{i=1}^{n} p_{ij} = \mu_j \qquad (j = 1, 2, ..., m) \tag{3}$$

$$\sum_{i=1}^{n} p_{ij}x_{ij}^2/\sum_{i=1}^{n} p_{ij} = \mu_j^2 + \sigma_j^2 \qquad (j = 1, 2, ..., m) \tag{4}$$

$$\mu = f(\mu_1, \mu_2, ..., \mu_m) \tag{5}$$

$$0 \leq p_{ij} \leq 1 \qquad \text{for all i, j} \tag{6}$$

$$0 \leq q_{ij} \leq 1 \qquad \text{for all i, j} \tag{7}$$

Table 1. Water quality data at MacKenzie River and Isaac River sub-basin.

a. STATION 130401 (at Isaac River)

CALIBRATION

Year	Conductivity @ 25 C (mS/m)	Dissolved Ions (mg/l)	Dissolved Solids (mg/l)	Hardness (mg/l)
71	270.0	196.1	-	78.0
72	351.3	265.9	-	102.8
73	237.8	173.8	140.8	73.5
74	471.0	301.0	249.5	136.5
75	350.0	211.4	176.0	89.0
76	550.0	359.4	303.0	157.7
77	311.0	204.3	174.2	86.4
78	457.5	304.9	263.0	129.0
79	660.0	411.0	343.0	181.0
80	429.0	269.6	231.8	113.0
Mean	408.8	269.7	235.2	114.7
Standard Deviation	131.0	76.4	68.1	35.9

VERIFICATION

Year	Conductivity @ 25 C (mS/m)	Dissolved Ions (mg/l)	Dissolved Solids (mg/l)	Hardness (mg/l)
81	425.0	200.8	218.5	106.5
82	316.7	200.7	173.3	83.3
83	-	-	-	-
84	332.5	219.3	185.0	83.5
85	283.3	176.7	156.7	72.0
86	276.7	182.6	150.0	75.7
Mean				
(81-85)	339.4	199.4	183.4	86.3
(82-86)	302.3	194.8	166.3	78.6
(81,82,84)	358.1	206.9	192.3	91.1
(82,84,85)	310.8	198.9	171.7	79.6
(84,85,86)	297.5	192.9	163.9	77.1
(81,82)	370.8	200.7	195.9	94.9
(82,84)	324.6	210.0	179.2	83.4
(84,85)	307.9	198.0	170.8	77.8
(85,86)	280.0	179.7	153.3	73.8

Table 1 (continued)

b. STATION 130106 (at Mackenzie River)

CALIBRATION

Year	Conductivity @ 25 C (mS/m)	Dissolved Ions (mg/l)	Dissolved Solids (mg/l)	Hardness (mg/l)
71	230.0	199.3	-	91.0
72	204.8	167.1	-	78.6
73	146.7	135.8	100.3	61.7
74	168.6	136.1	104.2	61.7
75	192.2	151.0	117.3	65.7
76	410.0	347.2	257.0	162.0
77	240.6	198.6	151.4	85.2
78	160.0	127.5	100.5	55.2
79	340.0	257.6	199.0	101.0
80	230.3	165.9	135.0	69.3
Mean	232.3	188.6	145.6	83.1
Standard Deviation	83.2	68.2	56.1	31.3

VERIFICATION

Year	Conductivity @ 25 C (mS/m)	Dissolved Ions (mg/l)	Dissolved Solids (mg/l)	Hardness (mg/l)
81	199.3	151.0	114.8	58.8
82	211.3	163.3	122.5	72.5
83	-	-	-	-
84	206.7	161.3	123.3	62.0
85	242.5	166.2	140.0	65.5
86	185.0	150.2	110.0	61.0
Mean				
(81-85)	214.9	160.4	125.2	64.7
(82-86)	211.4	160.3	124.0	65.3
(81,82,84)	205.8	158.5	120.2	64.4
(82,84,85)	220.1	163.6	128.6	66.7
(84,85,86)	211.4	159.2	124.4	62.8
(81,82)	205.3	157.2	118.7	65.7
(82,84)	209.0	162.3	122.9	67.3
(84,85)	224.6	163.7	131.7	63.8
(85,86)	213.8	158.2	125.0	63.3

Table 1 (continued)

c. STATION 130105 (at Mackenzie River)

CALIBRATION

Year	Conductivity @ 25 C (mS/m)	Dissolved Ions (mg/l)	Dissolved Solids (mg/l)	Hardness (mg/l)
71	-	-	-	-
72	290.0	199.7	148.5	80.5
73	215.0	159.6	135.8	61.8
74	350.0	233.5	194.0	97.0
75	235.0	158.4	133.7	63.7
76	516.7	336.2	274.0	150.3
77	350.0	248.6	204.0	102.3
78	367.5	258.2	207.5	110.5
79	-	-	-	-
80	411.7	220.0	220.0	107.7
Mean	343.2	231.2	189.7	96.7
Standard Deviation	97.6	58.7	48.3	28.8

VERIFICATION

Year	Conductivity @ 25 C (mS/m)	Dissolved Ions (mg/l)	Dissolved Solids (mg/l)	Hardness (mg/l)
81	248.0	176.7	139.0	71.0
82	260.0	167.0	150.0	63.0
83	-	-	-	-
84	226.3	161.0	130.0	64.0
85	235.0	161.0	140.0	61.5
86	255.0	187.9	146.7	80.7
Mean				
(81-85)	242.3	166.4	139.8	64.9
(82-86)	244.1	169.2	141.7	67.3
(81,82,84)	244.8	168.2	139.7	66.0
(82,84,85)	240.4	163.0	140.0	62.8
(84,85,86)	238.8	170.0	138.9	68.7
(81,82)	254.0	171.9	144.5	67.0
(82,84)	243.1	164.0	140.0	63.5
(84,85)	230.6	161.0	135.0	62.8
(85,86)	245.0	174.4	143.3	71.1

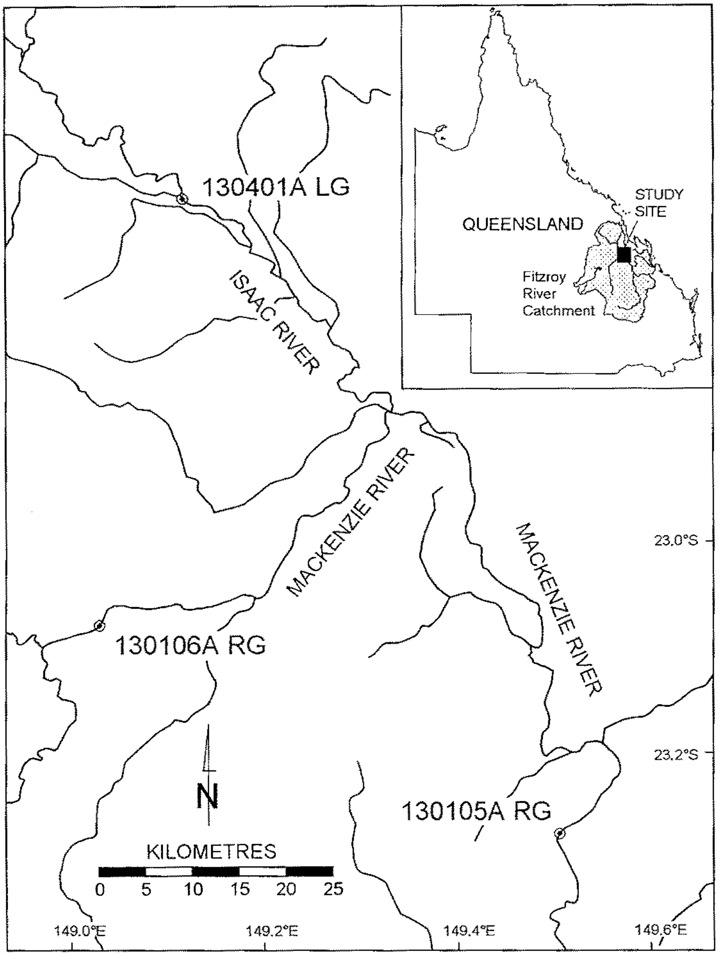


Figure 1. Location of the water quality monitoring stations.

$$\mu_j \geq 0 \qquad \text{for all j} \qquad (8)$$

$$\sigma_j \geq 0 \qquad \text{for all j} \qquad (9)$$

where:

x_{ij} = possible water quality level i at station j
q_{ij} = prior probability of event x_{ij}
μ_j = mean of the water quality level at station j
σ_j = standard deviation of water quality level at station j based on the prior distribution
μ = observed (changed) mean of the water quality level at the downstream location

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

On the basis of a study using one set of data both to define the original (prior) distributions and relationship between upstream and downstream water quality values, i.e., to calibrate the model, and a different set of data from the same stations to validate the model, Kusmulyono and Goulter (1994) report that this formulation gives very good predictions of the water quality at unmonitored sites.

One of the most common problems that may arise in the discrete approximation of a continuous function such as that employed for p_{ij} in the above formulation is the size of the discretization interval. Different levels of discretization may give different results for the same problem. Harmancioglu et al. (1985), have shown this problem to be a significant issue when using entropy principles to calculate the transfer of information between variables.

As well as the concern that the different degrees of discretization may give different results, and therefore give rise to erroneous decisions, computational requirements related to the level of discretization must also be addressed. It is generally true that the greater the number of intervals, i.e., the finer the discretization, the more precise the result. However, finer discretization also requires longer computation time which is not an insignificant issue in the non-linear formulation represented by Equations (1)-(9). Therefore, the choice of discretization has to be based on a trade-off between accuracy and computational effort.

In this model, the relationship between the data at the downstream and at the upstream station(s) is expressed in Equation (5). To develop this function, the values from each of the stations involved are required. The particular issue in this case is, what values should be used in the development of this function and in the corresponding entropy formulation.

More specifically, should either instantaneous values or values averaged over some period or interval of time be used? It is not generally possible to use instantaneous values because either 1) the values at the different stations may not be measured at the same time or 2) if the values are measured at the same time, the values may not be properly related because of the lag time (travel time of pollutants etc.) between the two locations if one monitoring station is located downstream of the other. If values averaged over a given time period are to be used, the question is one of what time interval to use. The particular model proposed by Kusmulyono and Goulter (1994) upon which the analysis described in this paper is based was used to predict changes in the value of water quality in the long term, i.e., on a time span of years. In doing so the model uses mean values of the water quality parameters over a specified period of years, i.e., an annual 'time-averaging' approach. The question therefore becomes one of which time averaging period, i.e., number of years over which the water quality values are to be averaged, to use. It should be noted that the length of data record available controls to some extent the range of time averaging periods which can be contemplated.

3 Implications of the choice of level of discretization

In order to clarify the effects of the choice of the number of intervals (level of discretization) on the prediction of the long term mean of water quality values at discontinued sampling stations, the MDI formulation of Equations (1)- (9) was solved for a six different levels of discretization, namely 8, 16, 24, 32, 40 and 48 intervals spread over four standard deviations either side of the mean. The results derived from the MDI formulation for this range of values are summarized in Table 2 and Figure 2. (The duration of the interval over which the water quality values were averaged was 4 years for this case. The discussion on the implications of varying the duration of time averaging is given in the following section).

It is clear from the results shown in Table 2, that variation in the level of discretization has very little effect on the value of the entropy H; doubling of the number of intervals from 24 to 48 results in a maximum change of less than 1% in the value of H. Similarly, the variation in the mean values of the water quality values assigned by the entropy function with the different levels of discretization is extremely small with the only changes occurring in the second decimal place. These observations indicate, that unlike the applications of the entropy function reported by Harmancioglu et al. (1985; 1992), the model of Kusmulyono and Goulter is quite insensitive to the level of discretization and therefore very robust in its assignment of water quality values. Hence the level of discretization does not appear to be a significant problem.

There is, however, a great difference in the processing time for different levels of discretization. The model was run using an IBM-PC Compatible with Intel 80486 processor, 33 MHz, using the non-linear package program GRG2 (Lasdon and Warren, 1986) to solve the model. The processing time required for the various levels of discretization are also shown in Table 2. This processing time increases rapidly as the discretization interval becomes smaller. Since the difference in results for the different levels of discretization is very small, the appropriate level of discretization can be quite coarse. In this case the level of discretization associated with having only eight intervals would appear to be quite satisfactory.

4 Implication of the choice of the length of time averaging

The method proposed by Kusmulyono and Goulter (1993) is intended primarily to predict the mean value of water quality over a selected period of time. The data used to examine the impact of varying the time period on the predictions provided by the model are shown in Table 1. The data from 1971 up to 1980 were used for the calibration step and the data from 1981 on were used in the validation step. In order to observe the effect of the selection of the period of time over which the functions expressed in Equation (5) are developed, the entropy formulation was solved for a range of different time intervals in the development of that function. Due to the limited data (only eight years of data are available for all stations, except for Dissolved Solids which has only seven years of data in the same year available for the calibration step, and only 5 years of data for all stations in the same year available for the validation step) only a limited number of variations in the averaging time period were able to be analyzed. In this study, 5, 4, 3 and 2 years duration averaging periods were selected to develop Equation (5). From 8 data points available for calibration, only 4 data points can be used to calculate the regression coefficients for the 5 year

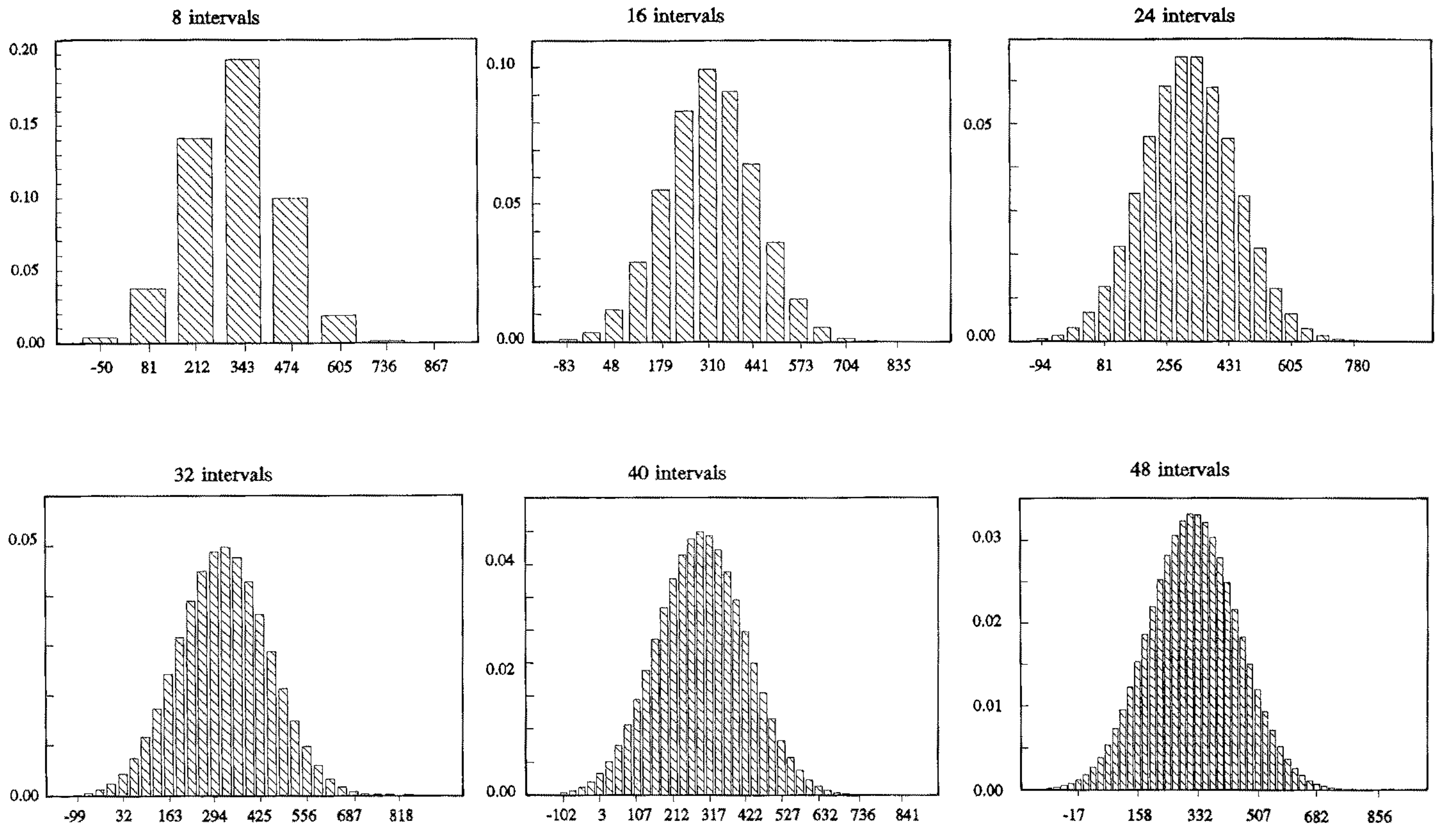


Figure 2a. Assigned probabilities at station 130401 conductivity.

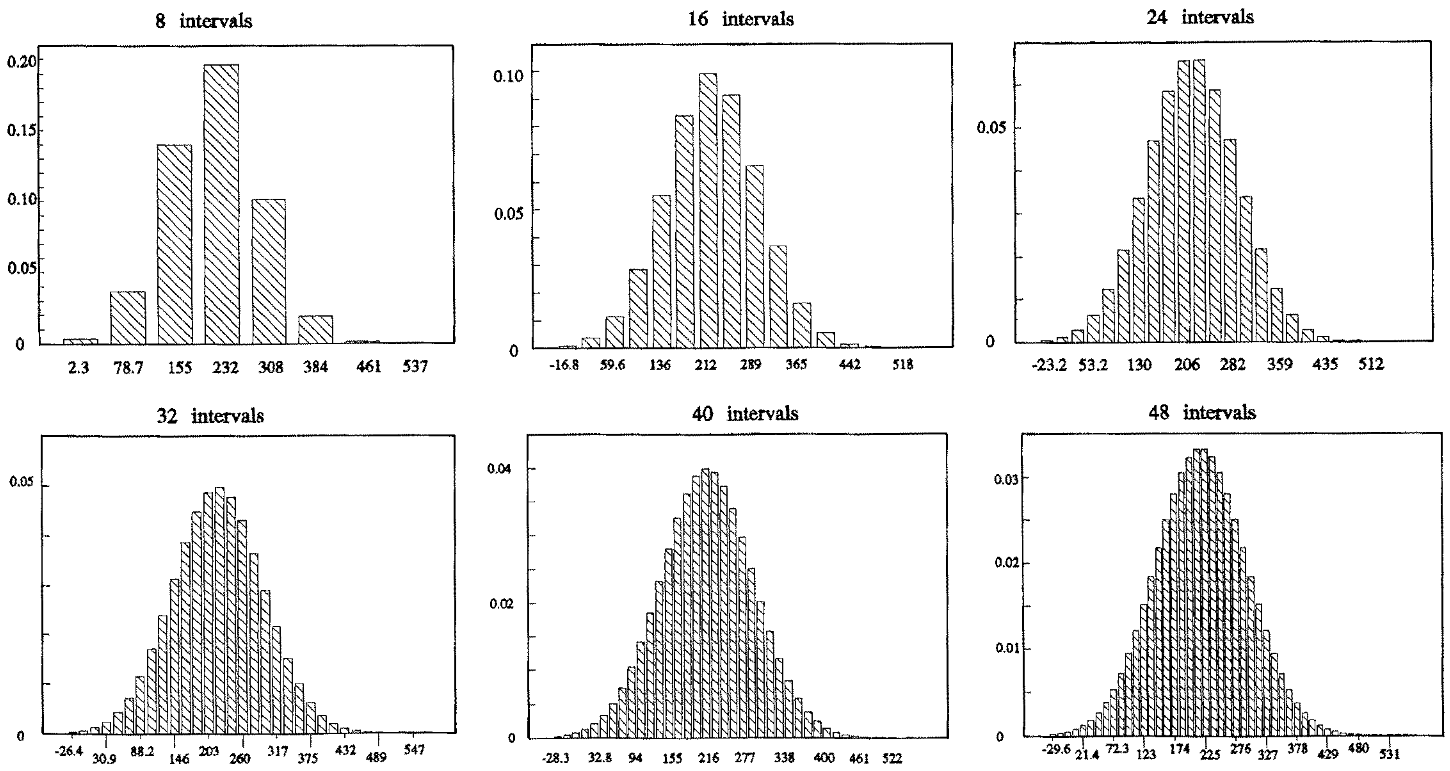


Figure 2b. Assigned probabilities at station 130401 dissolved ions.

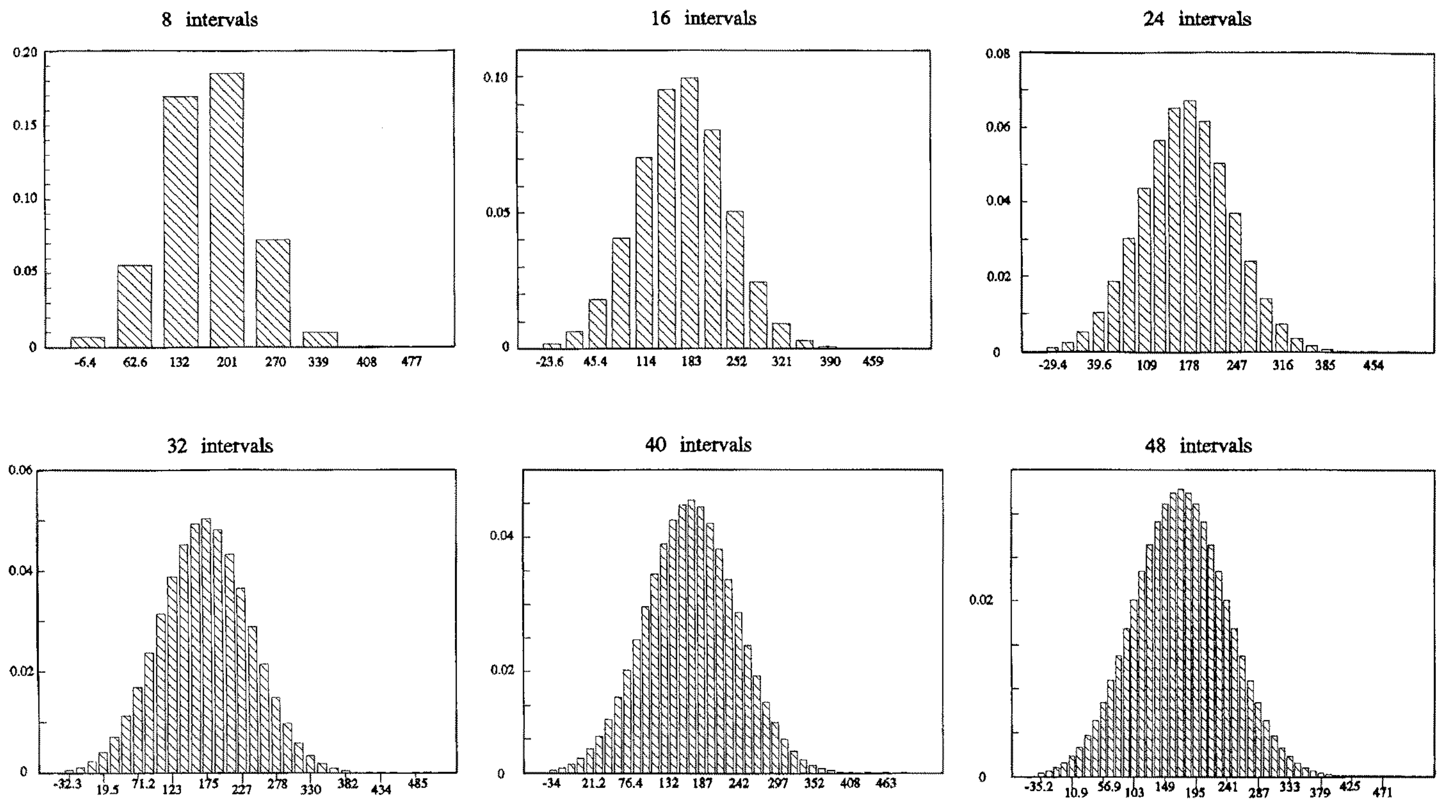


Figure 2c. Assigned probabilities at station 130401 dissolved solids.

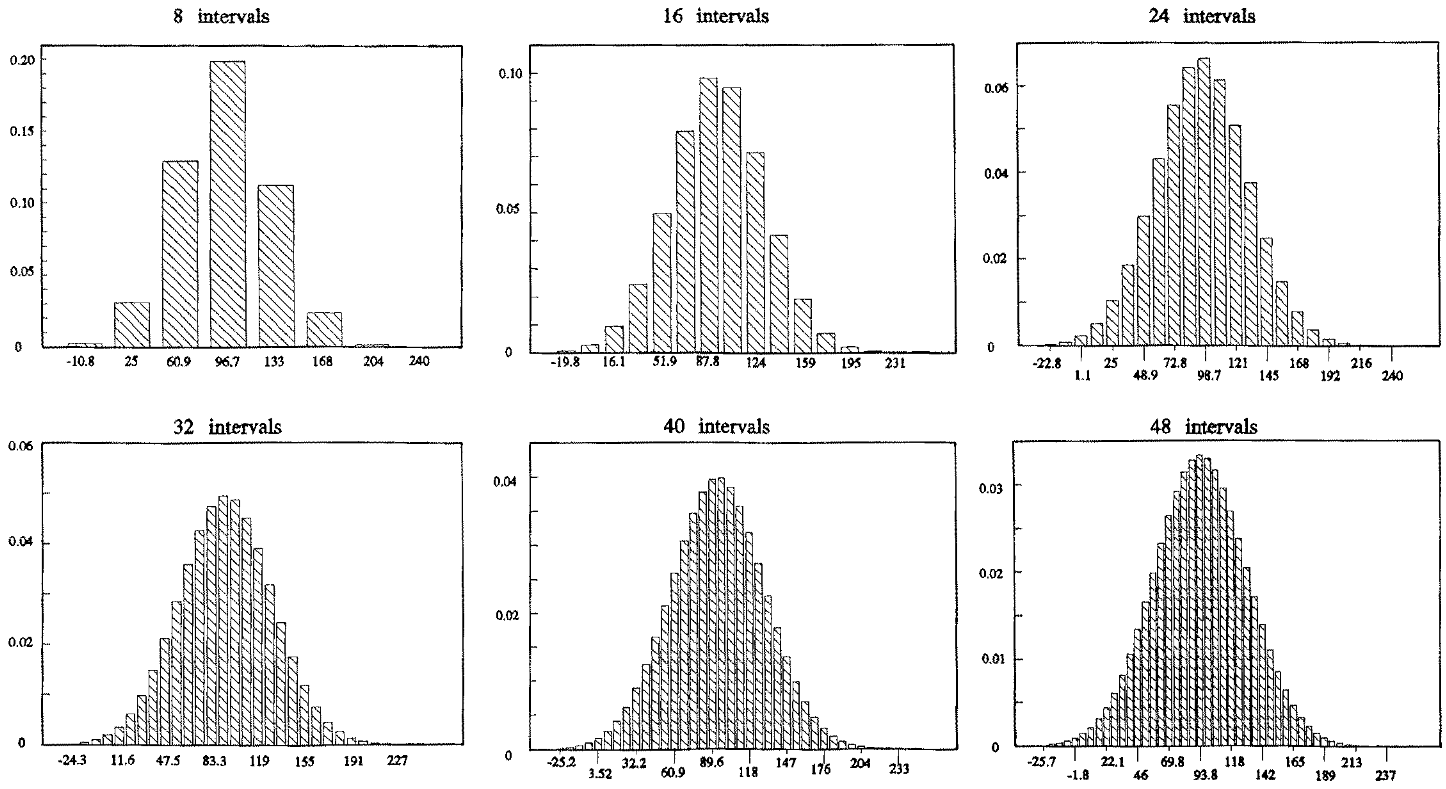


Figure 2d. Assigned probabilities at station 130401 hardness.

Table 2. Results of assignment of mean water quality values with varying levels of discretization.

			Number of Intervals					
Water Quality	Mean Value at Station 130401	Mean Value at Station 130106	8		16		24	
Parameter			Entropy Value (H)	Time to Reach Solution	Entropy Value (H)	Time to Reach Solution	Entropy Value (H)	Time to Reach Solution
Conductivity	318.6	213.4	-0.1314	7.74"	-0.1315	69.2"	-0.1315	422.9"
Dissolved Ions	215.9	166.4	-0.1506	6.42"	-0.1509	104.4"	-0.1509	494.4"
Dissolved Solids	171.0	123.5	-0.2554	5.39"	-0.2556	39.8"	-0.2558	248.0"
Hardness	92.3	74.2	-0.1181	5.58"	-0.1182	96.5"	-0.1183	457.8"

			Number of Intervals					
Water Quality	Mean Value at Station 130401	Mean Value at Station 130106	32		40		48	
Parameter			Entropy Value (H)	Time to Reach Solution	Entropy Value (H)	Time to Reach Solution	Entropy Value (H)	Time to Reach Solution
Conductivity	318.6	213.4	-0.1318	551.9"	-0.1320	963.3"	-0.1330	5233.0"
Dissolved Ions	215.9	166.4	-0.1508	204.2"	-0.1512	921.1"	-0.1512	3665.8"
Dissolved Solids	171.0	123.5	-0.2561	461.1"	-0.2565	2192.9"	-0.2565	1812.0"
Hardness	92.3	74.2	-0.1185	1996.7"	-0.1185	1525.8"	-0.1185	3726.8"

moving average; 5 data points for the 4 year moving average; 6 data points for the 3 year moving average and 7 data points for the 2 year moving average.

Development of Equation (5) from the formulation was conducted by regression analysis, with an assumption that the water quality values at the downstream location are a mixed value of the two upstream water quality values. This value can be calculated as the summation of the value of water quality from each upstream location, weighted by the discharge of each tributary in order to recognize the proportional contribution of each tributary to the water quality at the downstream station.

The regression coefficients for each time averaging period are shown in Table 3. The graphical presentation of the variation of this regression line can be seen in Figure 3a for Conductivity, Figure 3b for Dissolved Ions, Figure 3c for Dissolved Solids and Figure 3d for Hardness. It can be seen in Table 3, that in most cases, the 4 year moving average gives the highest coefficient of correlation between the water quality value at the downstream station and the summation of the water quality values, weighted by their average flows, from the two upstream stations.

It was initially expected that the best entropy prediction would occur for those time intervals with the highest correlation coefficients, in this case, the 4 year interval. The results of the prediction step, as shown in Table 4, show that it is not always true. For example, for a four year averaging interval the correlation coefficient (see Table 3) is significantly larger than those for the other three time intervals considered for all water quality parameters. However, it can be seen that the predictions using the four year time interval are no better and in some cases worse than using the other time periods. This situation arises to a large extent from the variation in the number of points used to develop the regression equation, [Equation (5) in the formulation] caused by change in the duration of the interval used to average the water quality values for a fixed length of record; longer averaging periods result in fewer points for the expression. For example, in the case of a 5 year moving average only 4 data points are available. The smaller numbers of data points may give good correlations but, in fact, the resulting expression may not be an appropriate representation of reality. In the case of smaller intervals over which to average the values, the number of points to calculate the regression coefficients will increase, the resulting expression is likely to be a better reflection of reality, but the short term variation of the data may be more dominant with a correspondingly smaller correlation coefficient.

The duration of the averaging interval may therefore be decided on the basis of the time span upon which it is desired to calculate average water quality values rather than on a formal statistical evaluation of what time interval is optimal in terms of accuracy of prediction. The same 'time-averaging' period must of course be used in the calibration and prediction steps. It should also be recognized that use of time averaged values for water quality damps out the effects of short term transient fluctuation in water quality, which although important at their time of occurrence, do not relate directly to the background water quality conditions and to the long term and possibly permanent variation in those background conditions.

5 Summary and conclusion

Two computational aspects, namely, level of discretization, and duration of the period of time over which to average the water quality values to be predicted, involved in using entropy theory to predict the value of water quality at discontinued monitoring stations were examined.

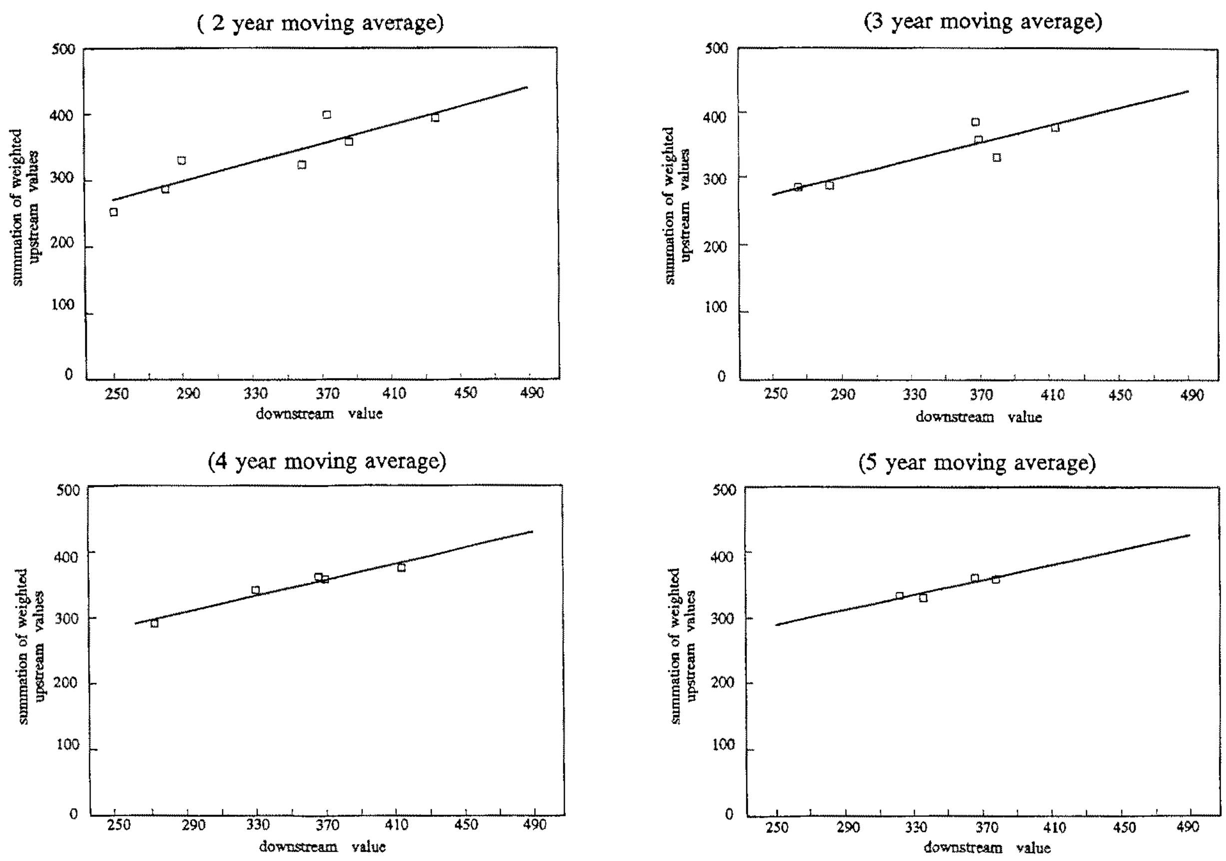
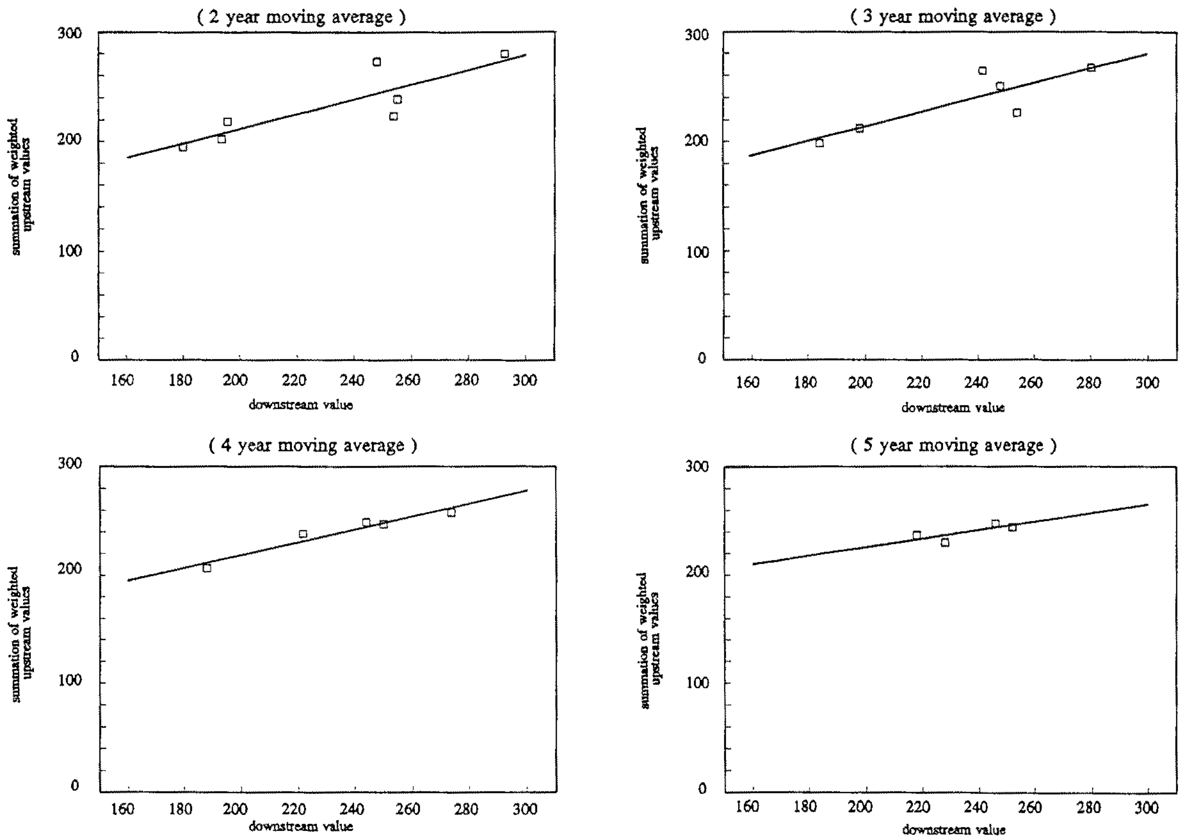


Figure 3a. Plot of regression lines for different periods of time over which the conductivity is averaged.


(2 year moving average)

(3 year moving average)

(4 year moving average)

(5 year moving average)

Figure 3b. Plot of regression lines for different periods of time over which the dissolved ion is averaged.

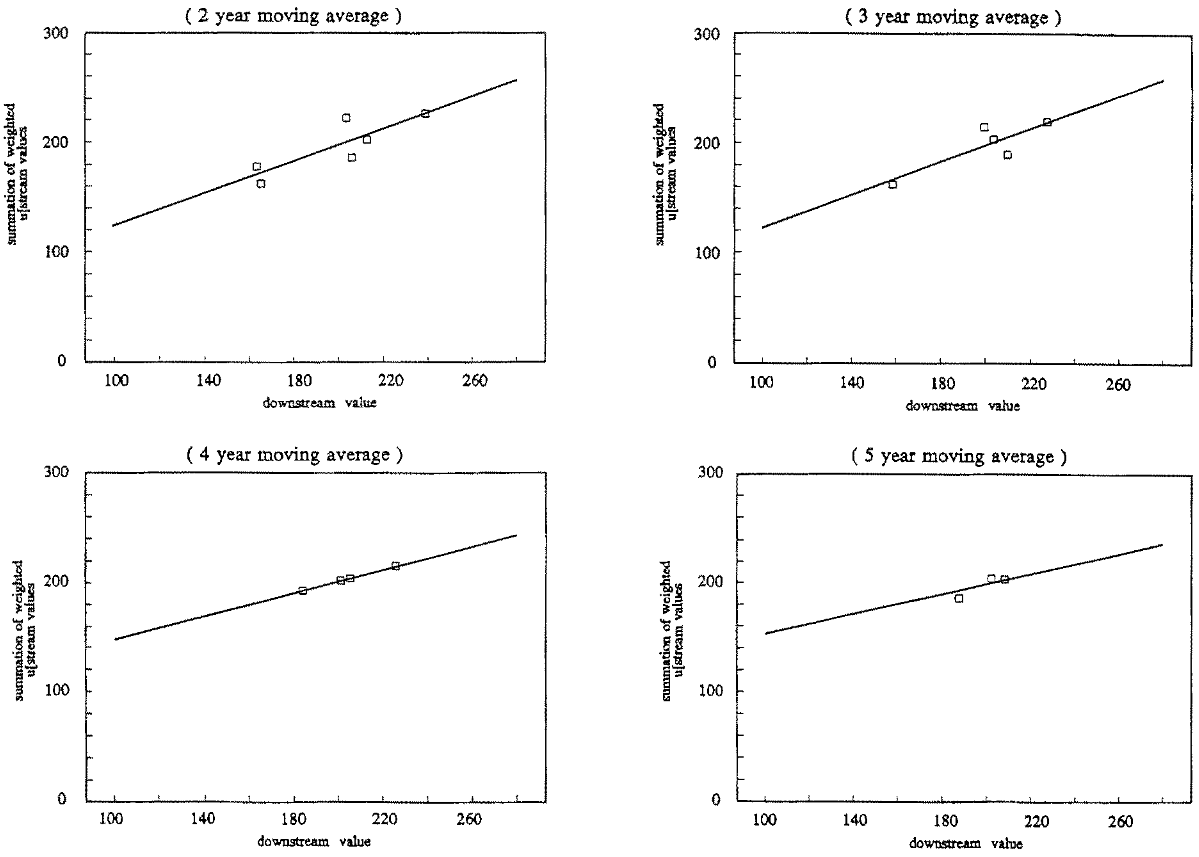


Figure 3c. Plot of regression lines for different periods of time over which the dissolved solid is averaged.

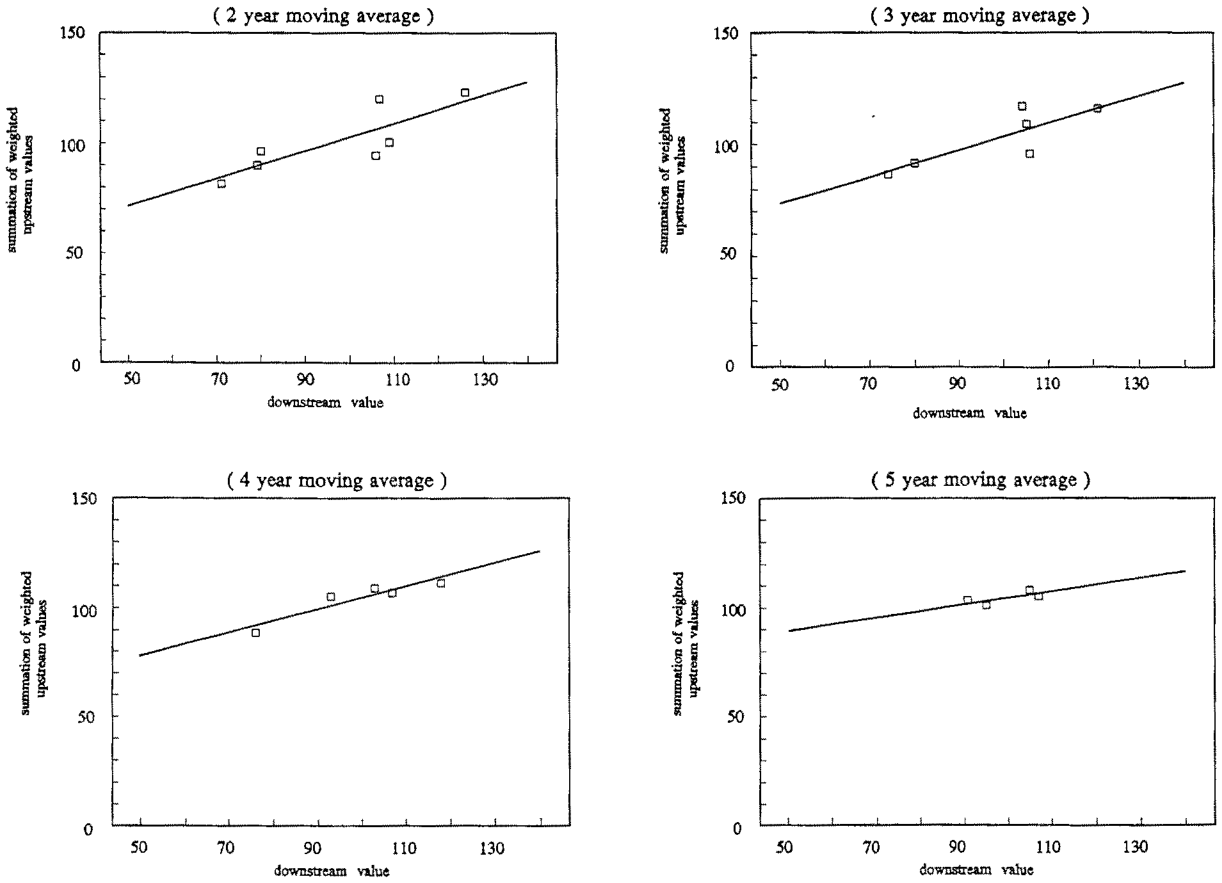


Figure 3d. Plot of regression lines for different periods of time over which the hardness is averaged.

Table 3. Regression coefficients for various periods of time

Summation of Weighted Upstream Values = $B_0 + B_{1*}$ (Downstream Value)

Conductivity

	2 year moving average	3 year moving average	4 year moving average	5 year moving average
B_0	95.96	112.82	137.72	147.89
B_1	0.6990	0.6580	0.5980	0.5689
correlation coefficient	0.8767	0.8828	0.9688	0.9322

Dissolved Ions

	2 year moving average	3 year moving average	4 year moving average	5 year moving average
B_0	77.875	81.546	100.58	147.23
B_1	0.6693	0.6602	0.5914	0.3936
correlation coefficient	0.8526	0.8619	0.9730	0.8175

Dissolved Solids

	2 year moving average	3 year moving average	4 year moving average	5 year moving average
B_0	50.586	46.682	50.978	105.98
B_1	0.7355	0.7577	0.7423	0.4668
correlation coefficient	0.8607	0.8818	0.9832	0.8112

Hardness

	2 year moving average	3 year moving average	4 year moving average	5 year moving average
B_0	40.273	44.021	51.357	74.369
B_1	0.6247	0.5966	0.5351	0.3034
correlation coefficient	0.8347	0.8238	0.9390	0.7853

Table 4. Results of assignment of mean water quality values for various time intervals

STATION 130401

Water Quality	2 Yearly				3 Yearly			
Parameter	Data	Assigned	% Error	Average	Data	Assigned	% Error	Average
Conductivity	370.8	306.1	-17.46		358.1	306.7	-14.34	
	324.6	295.7	-8.90	10.12%	310.8	302.7	-2.62	6.07%
	307.9	283.7	-7.86		297.5	301.2	1.24	
	280.0	297.5	6.25					
Dissolved Ions	200.7	208.4	3.82		206.9	208.0	0.52	
	210.0	201.8	-3.90	6.39%	198.9	203.7	2.42	3.85%
	198.0	199.3	0.67		192.9	209.5	8.62	
	179.7	210.5	17.17					
Dissolved Solids	195.9	173.7	-11.34		192.3	168.1	-12.57	
	179.2	169.4	-5.45	8.23%	171.7	168.4	-1.90	5.54%
	170.8	164.7	-3.59		163.9	167.4	2.14	
	153.3	172.6	12.57					

Table 4 (continued)

Water Quality	2 Yearly				3 Yearly			
Parameter	Data	Assigned	% Error	Average	Data	Assigned	% Error	Average
Hardness	94.9	87.3	-8.02		91.1	88.9	-2.43	
	83.4	84.5	1.30	9.98%	79.6	86.5	8.65	9.68%
	77.8	84.0	8.04		77.1	90.9	17.97	
	73.8	90.5	22.57					

STATION 130106

Water Quality	2 Yearly				3 Yearly			
Parameter	Data	Assigned	% Error	Average	Data	Assigned	% Error	Average
Conductivity	205.3	210.8	2.68		205.8	210.9	2.50	
	209.0	208.6	-0.17	3.34%	220.1	210.1	-4.56	2.60%
	224.6	206.1	-8.23		211.4	209.8	-0.75	
	213.8	208.9	-2.27					
Dissolved Ions	157.2	163.2	3.85		158.5	163.0	2.81	
	162.3	160.4	-1.19	2.85%	163.6	161.2	-1.47	2.34%
	163.7	159.4	-2.65		159.2	163.6	2.75	
	158.2	164.1	3.75					
Dissolved Solids	118.7	124.5	4.92		120.2	122.5	1.89	
	122.9	123.0	0.07	3.39%	128.6	122.6	-4.67	2.76%
	131.7	121.3	-7.87		124.4	122.3	-1.72	
	125.0	124.1	-0.72					
Hardness	65.7	72.2	9.95		64.4	72.9	13.12	
	67.3	71.1	5.72	10.77%	66.7	71.9	7.85	12.75%
	63.8	70.9	11.22		62.8	73.7	17.29	
	63.3	73.5	16.21					

STATION 130401

Water Quality	4 Yearly				5 Yearly			
Parameter	Data	Assigned	% Error	Average	Data	Assigned	% Error	Average
Conductivity	339.4	318.6	-6.12	6.00%	326.8	324.9	-0.58	0. 58%
	302.3	320.1	5.89					
Dissolved Ions	199.4	215.9	8.30	10.10%	196.0	235.2	20.00	20.00%
	194.8	218.0	11.90					
Dissolved Solids	183.4	171.0	-6.75	5.34%	176.7	193.0	9.22	9.22%
	166.3	172.8	3.94					
Hardness	86.3	92.3	6.91	13.17%	84.2	103.5	22.92	22.92%
	78.6	93.9	19.93					

STATION 130106

Water Quality	4 Yearly				5 Yearly			
Parameter	Data	Assigned	% Error	Average	Data	Assigned	% Error	Average
Conductivity	214.9	213.4	-0.72	0.91%	209.0	214.7	2.73	2.7 3%
	211.4	213.7	1.11					
Dissolved Ions	160.4	166.4	3.71	4.02%	158.4	174.4	10.10	10.10%
	160.3	167.2	4.34					
Dissolved Solids	125.2	123.5	-1.33	0.76%	122.1	131.1	7.37	7.37%
	124.0	124.2	0.19					
Hardness	64.7	74.2	14.67	14.37%	64.0	78.7	22.97	22.97%
	65.3	74.9	14.79					

Unlike other applications of entropy theory, the level of discretization does not appear to effect the results significantly. The choice of the discretization or number of intervals can therefore be determined in part by computation time, which increases rapidly with increasing number of intervals.

The period of time for which the water quality values are assigned by the model appears to have a significant impact on the accuracy of the prediction. This time interval is used both in the prediction and calibration steps of the model and must be the same for both steps for a given application of the model. The time interval associated with the highest correlation coefficient in the regression used for calibration does not necessarily correspond to the time interval associated with the most accurate prediction in the entropy formulation derived solution. This difference is believed to be due in part to the amount of data that is available in the calibration phase for a fixed period of record; the longer the intervals over which the water quality values are averaged for a given period of record the smaller the number of data points upon which to base the model. Care must therefore be taken in specifying the time intervals over which water quality values will be averaged in the model.

References

Awumah, K.; Goulter, I.; Bhatt, S. 1990: "Assessment of Reliability in Water Distribution Networks Using Entropy Based Measures", Stochastic Hydrology and Hydraulics, 4, 309-320

Awumah, K.; Goulter, I.; Bhatt, S. 1991: "Entropy Based Redundancy Measures in Water Distribution Network Design", Journal of Hydraulic Engineering, ASCE 117(10), 595-614

Awumah, K; Goulter, I. 1992: "Maximising Entropy Defined Redundancy of Water Distribution Networks", Engineering Optimisation 20(11), 57-80

Chiu, C. 1987: "Entropy and Probability Concepts in Hydraulics", Journal of Hydraulic Engineering, ASCE 113(5), 583-600

Chiu, C. 1988: "Entropy and 2-D Velocity Distribution in Open Channels", Journal of Hydraulic Engineering, ASCE 114(7), 738-756

Chiu, C. 1989: "Velocity Distribution in Open Channels Flow", Journal of Hydraulic Engineering, ASCE 115(5), 576-594

Chiu, C. 1991: "Application of Entropy Concept in Open Channel Flow Study", Journal of Hydraulic Engineering, ASCE 117(5), 615-628

Chiu, C.; Chiou, J.D. 1986: "Structures of 3-D Flow in Rectangular Open Channels" Journal of Hydraulic Engineering, ASCE 112(11), 1050-1068

Harmancioglu, N.B. 1984: "Entropy Concept as used in Determination of Optimum Sampling Intervals", Proceedings of Hydrosoft '84, International Conference on Hydraulic Engineering Software, Portoroz, Yugoslavia, 6-99

Harmancioglu, N.B; Alpaslan, N. 1992: "Water Quality Monitoring Network Design A Problem of Multi-Objective Decision Making", Water Resources Bulletin, A.W.R.A., 28(1), 179-192

Harmancioglu, N.B; Baran, T. 1989: "Effects of Recharge System on Hydrologic Information Transfer Along Rivers", IAHS, Proc. of the Third Scientific Assembly of the International Association of Hydrologic Science - New Directions for Surface Water Modelling, IAHS Publ. 181, 223-233

Harmancioglu, N.B.; Yevjevich, V.; Obeysekera, J.T.B. 1985: "Measures of Information Transfer Between Variables", Proceedings of Fourth International Hydrology Symposium, July 15-17, 1985, 481-499

Harmancioglu, N.B.; Yevjevich, V. 1987: "Transfer of Hydrologic information Among River Points", Journal of Hydrology 91, 103-118

Jaynes, E.T. 1957: "Information Theory and Statistical Mechanics", Physical Review 106, 620-630

Jaynes, E.T. 1968: "Prior Probabilities" IEEE Transactions on System Science and Cybernetics 4(3), 227-241

Jaynes, E.T. 1983: "Papers on Probability Statistics and Statistical Physics", D. Reidel Publishing Company, Dordrecht, Holland, 434 p

Kapur, J.N., 1983, "Twenty five years of maximum entropy principle", Journal of Mathematical Physical Science, 17 (2), pp 103-156.

Kapur, J.N. 1989: "Maximum Entropy Models in Science and Engineering", Wiley Eastern Ltd., New Delhi, India, 635 p

Kusmulyono, A.; Goulter, I. 1994: "Entropy Principles in the Prediction of Water Quality Values at Discontinued Monitoring Stations", Stochastic Hydrology and Hydraulics (This issue)

Lasdon, L.S.; Waren, A.D. 1986: "GRG2 User's Guide Department of General Business Administration", The University Texas at Austin, Austin, Texas, 60 p

Shannon, C.E. 1948: "A Mathematical Theory of Communication", Bell System Technical Journal 27(3), 379-423, 623-659

Sonuga, J.O. 1972: "Principle of Maximum Entropy in Hydrologic Frequency Analysis", Journal of Hydrology 17, 177-191

Sonuga, J.O. 1976 "Entropy Principle Applied to the Rainfall-Runoff Process", Journal of Hydrology 30, 81-94