HYDROLOGICAL PROCESSES
Hydrol. Process. **18**, 2165–2179 (2004)
Published online 30 June 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.1465

Characterizing the spatial variability of groundwater quality using the entropy theory: I. Synthetic data

Y. Mogheir,[1,2] J. L. M. P. de Lima[3]* and V. P. Singh[4]

[1] *Department of Civil Engineering, Faculty of Science and Technology, Campus 2, University of Coimbra, 3030-290 Coimbra, Portugal*
[2] *Palestinian Water Authority, Building No. 136/61,League of Arab States St. Tel El Hawa, Gaza, Palestine*
[3] *Department of Civil Engineering, Institute of Marine Research, Coimbra Interdisciplinary Centre, Faculty of Science and Technology, Campus 2, University of Coimbra, 3030-290 Coimbra, Portugal*
[4] *Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA 70803-6405, USA*

Abstract:

This paper, the first in a series of two, applies the entropy (or information) theory to describe the spatial variability of synthetic data that can represent spatially correlated groundwater quality data. The application involves calculating information measures such as transinformation, the information transfer index and the correlation coefficient. These measures are calculated using discrete and analytical approaches. The discrete approach uses the contingency table and the analytical approach uses the normal probability density function. The discrete and analytical approaches are found to be in reasonable agreement. The analysis shows that transinformation is useful and comparable with correlation to characterize the spatial variability of the synthetic data set, which is correlated with distance. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS correlation; entropy; information; spatial variability

INTRODUCTION

The entropy (or information) theory, developed by Shannon (1948), recently has been applied in many different fields, such as ecology, biology, data mining and financial time-series analysis (e.g. Darbellay and Wuertz, 2000; Rojdestvenski and Cottam, 2000; Sy, 2001; Ulanowicz, 2001). The entropy theory has also been applied in hydrology and water resources for measuring the information content of random variables and models, evaluating information transfer between hydrological processes, evaluating data acquisition systems, and designing water quality monitoring networks. A comprehensive review of the application of entropy theory in hydrology and water resources is given by Singh (1997).

Design of water quality monitoring networks is still a controversial issue, for there are difficulties in the selection of temporal and spatial sampling frequencies, the variables to be monitored, the sampling duration and the objectives of sampling (Harmancioglu *et al*., 1999). Many studies have applied the entropy theory to assess and optimize the data collection network (e.g. water quality, rainfall, stream flow, elevation data, landscape, etc.). Uslu and Tanriover (1979) analysed the entropy concept for the delineation of optimum sampling intervals in data collection systems, both in space and time. Harmancioglu (1981) investigated the transfer of information between observations of two stream gauging stations. Krastanovic and Singh (1992) used the marginal entropy measure to draw contour maps of the rainfall network in Louisiana and evaluated the network according to the entropy map. Yang and Burn (1994) described an analytical comparison between the correlation and the joint entropy between gauging stations. Lee and Ellis (1997) compared kriging and the maximum entropy estimator for spatial interpolation and their subsequent use in optimizing monitoring

* Correspondence to: J. L. M. P. de Lima, Department of Civil Engineering, Faculty of Science and Technology, Campus 2, University of Coimbra, 3030-290 Coimbra, Portugal. E-mail: plima@dec.uc.pt

Received 21 February 2001
Accepted 1 September 2003

networks. Husain (1989) and Bueso *et al*. (1999) used the entropy theory to illustrate a framework for spatial sampling of a monitoring network. Ozkul *et al*. (2000) presented a method using the entropy theory for assessing water quality monitoring networks. The work was a follow up of earlier work by Harmancioglu and Alpaslan (1992). Mogheir and Singh (2002) used the entropy theory to evaluate and assess a groundwater monitoring network by means of marginal entropy contour maps.

Most of the research cited above used an analytical approach that presumed knowledge of the probability distributions of the random variables under study. The problem of not knowing the probability distributions can, however, be circumvented if a discrete approach is adopted. Furthermore, these investigations give little indication as to how information changes with the distance when the data are spatially correlated or not necessarily correlated. This paper sets out to use a discrete approach for calculating information measures, and to use transinformation (T) and the information transfer index (ITI) to describe the spatial variability of synthetic data that is spatially correlated and fits the normal distribution function. The transinformation model (T model) is a relationship between mutual information measures, specifically T, and the distance between wells. Thus, the objective is to investigate the possibility of developing the T model and compare it with the commonly used correlation model (C model), where both models are obtained by discrete and analytical approaches. Also explored is the capability of the T model for characterizing the spatial variability of synthetic data. The method developed here was also used to analyse data that have low spatial correlation and do not fit the normal distribution function (Mogheir *et al*., 2004).

SYNTHETIC DATA

The data used in this study were obtained using the COVRAN program (Zhou, 1996). The input data for the program were: the mean, the variance, the type of distribution function, the size of generated data, random number generator, the grid size and the type of correlation model. The exponential model was used for the correlation function.

For this hypothetical case, the random field consisted of 10×10 grid points, as shown in Figure 1. The grid size was taken as 100 m in both orthogonal directions. Each point was assumed to represent an observation well in the monitoring network. The generated data can represent any hydrogeological data, such as water level or chemical concentration. Thirty grid points were selected to represent the observation wells for use in the analyses. Table I shows the parameters used for generating the random data, using the COVRAN program (Zhou, 1996).

METHOD

The method developed in this study involves three steps.

1. Computation of information measures by two different approaches, discrete and analytical.

Table I. Input parameters (program COVRAN; Zhou, 1996) used for generating the random data for the hypothetical monitoring network as represented in Figure 1

Input Parameter	
Type of distribution function	Normal
The mean	0
The variance	1
Sizes of generated data	200, 300, 400, 500
Correlation model	Exponential
Correlation length	500
Grid size	100

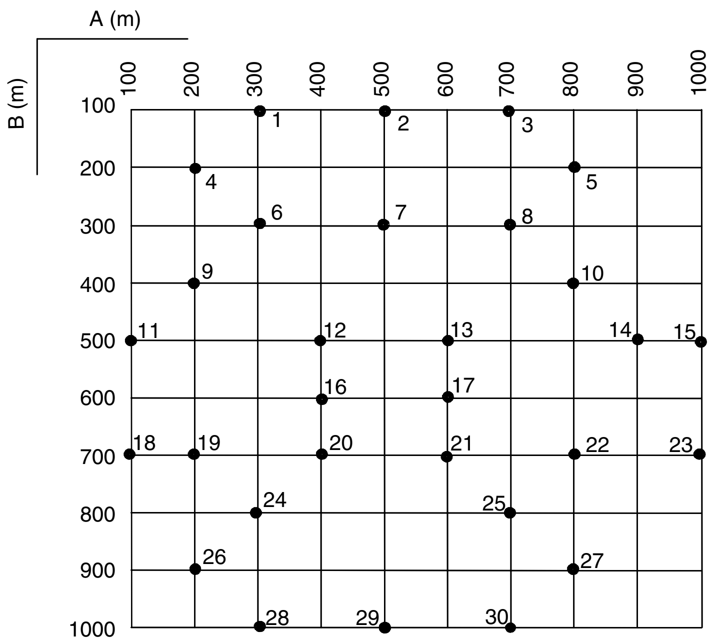

Figure 1. Hypothetical monitoring network using 10 × 10 grid point. The grid length is 100 m and each point (●) represents an observation well (30 wells were selected for use in the analysis)

2. Comparison of the two approaches by means of the transinformation model (T model) and the correlation model (C model).
3. Applying the discrete and analytical approaches for characterizing the spatial variability.

These three steps were carried out using the randomly generated data (synthetic data). The C-model and the information transfer index model (ITI model) were compared to ascertain the applicability of the T-model for characterizing spatial variability.

Discrete approach—correlation model

The correlation coefficient r_{xy} was calculated for each pair of wells (or points) as

$$r_{xy} = \frac{\mathrm{cov}_{xy}}{S_x S_y} \tag{1}$$

where cov_{xy} is the covariance between the random variables x and y, and S_x and S_y are the standard deviation of variables x and y, respectively. The cov_{xy} could be obtained as

$$\mathrm{cov}_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{n - 1} \tag{2}$$

where $\overline{x}$ and $\overline{y}$ and are the means of variable x and y, respectively.

Discrete approach—transinformation model

To calculate the information measures, the joint or conditional probability is needed, and this can be obtained using a contingency table. An example of a two-dimensional contingency table is given in Table II.

Table II. Two-dimensional contingency table (frequency)

x	y					Total
	1	2	3	...	u	
1	f_{11}	f_{12}	f_{13}	...	f_{1u}	$f_{1.}$
2	f_{21}	f_{22}	f_{23}	...	f_{2u}	$f_{2.}$
3	f_{31}	f_{32}	f_{33}	...	f_{3u}	$f_{3.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
v	f_{v1}	f_{v2}	f_{v3}	...	f_{vu}	$f_{v.}$
Total	$f_{.1}$	$f_{.2}$	$f_{.3}$	...	$f_{.u}$	f_x or f_y

To construct a contingency table, let the random variable x have a range of values consisting of v categories (class intervals), whereas the random variable y is assumed to have u categories (class intervals). The cell density or the joint frequency for (i, j) is denoted by f_{ij}, $i = 1, 2, \ldots, v$; $j = 1, 2, \ldots, u$, where the first subscript refers to the row and the second subscript to the column. The marginal frequencies are denoted by $f_{i.}$ and $f_{j.}$ for the row and the column values of the variables, respectively. The construction of two-dimensional contingency tables is illustrated in Appendix A.

The entropy of a random variable is a measure of the information or uncertainty associated with it. The measures of information are: the marginal entropy, conditional entropy, joint entropy and transinformation. For a random variable x, the marginal entropy, $H(x)$, can be defined as the potential information of the variable. For two random variables, x and y, the conditional entropy $H(x/y)$ is a measure of the information content of x that is not contained in the random variable y. The joint entropy $H(x, y)$ is the total information content contained in both x and y. The mutual entropy (information) between x and y, also called transinformation, $T(x, y)$, is interpreted as the reduction in uncertainty in x, due to the knowledge of the random variable y. It also can be defined as the information content of x that is contained in y. These information measures for discrete variables can be expressed as (e.g. Lubbe, 1996; Singh, 1998)

$$H(x) = -\sum_{i=1}^{n} p(x_i) \ln p(x_i) \tag{3}$$

$$H(x, y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \ln p(x_i, y_j) \tag{4}$$

$$H(x/y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \ln p(x_i/y_j) \tag{5}$$

$$T(x, y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \ln \left[\frac{p(x_i, y_j)}{p(x_i)p(y_j)}\right] \tag{6}$$

where x and y are two discrete variables with values x_i, $i = 1, 2, \ldots, n$; y_j, $j = 1, 2, \ldots, m$, defined in the same probability space, each of which has a discrete probability of occurrence $p(x_i)$, $p(x_i, y_j)$ is the joint probability of x_i, y_j and $p(x_i/y_j)$ is the probability of x_i conditional on y_j. Note that $H(x, y) = H(y, x)$. The transinformation $T(x, y)$ also can be expressed as (e.g. Jessop, 1995)

$$T(x, y) = H(x) - H(x/y) \tag{7}$$

$$T(x, y) = H(x) + H(y) - H(x, y) \tag{8}$$

$$T(y, x) = H(y) - H(y/x) \tag{9}$$

$$T(y, x) = H(y) + H(x) - H(y, x) \tag{10}$$

Note that $T(x, y) = 0$ if x and y are independent. Transinformation is an indicator of the capability of information transmission. Although transinformation indicates the dependence of two variables, it is not a good index of dependence because its upper bound varies from site to site (it varies from 0 to marginal entropy H). Therefore, an information transfer index (ITI) is defined by normalizing transinformation, which then indicates the standardized information transferred from one site to another

$$\text{ITI} = \frac{T(x, y)}{H(x, y)} \tag{11}$$

When specifying the unit of information measures, it is important to note that the logarithmic base used to calculate the information theory parameters determines the units of these measurements. If a base 2 is used, then the unit is a 'bit'; for a logarithmic base 10 the unit is decibles, and it is nats (natural units) if the logarithmic base is e (Caselton and Husain, 1980). However, some researchers, such as Harmancioglu and Yevjevich (1987), have used napiers as the unit of the information theory parameters, with base e. Provided that the logarithmic base is used consistently the choice of units is not critical. For convenience, the base e and the unit 'nats' have been used here for computing all the numerical results.

For both transinformation and correlation models, the geometric distance (d) between two wells was calculated as

$$d = \sqrt{(A_1 - A_2)^2 + (B_1 - B_2)^2} \tag{12}$$

where A_1, B_1 are the coordinates of well 1, A_2, B_2 are the coordinates of well 2 and d is the distance between wells 1 and 2 (see Figure 1).

A program, INFOR, was used to compute the correlation coefficient of each pair of wells using Equation (1), the joint frequency and joint probability contingency tables, the marginal entropy using Equation (3), the joint entropy using Equation (4), the transinformation using Equation (6), the ITI using Equation (11) and the distance between pairs of wells using Equation (12).

Smoothing the discrete models

The discrete T values may exhibit a scatter when plotted against the distance between wells. The literature reports several smoothing methods, such as axis transformation (e.g. logarithm transformation), moving average and exponentially weighted moving average (Berthouex and Brown, 1994). In this study, the moving average method is used to smooth the T data using a 100 m distance interval, which is the distance between wells in the hypothetical network used (see Figure 1). For distance 0, the transinformation T_0 was assumed as the average of the marginal entropies of the wells. The moving average method was applied to smooth the lognormal T, ITI and correlation models.

Analytical approach—correlation model

The data used in this study were generated synthetically. They fit the normal probability distribution (with mean 0 and variance 1) and were spatially correlated with distance. In the COVRAN program (Zhou, 1996), the correlation coefficient was calculated using the exponential correlation

$$r(d) = e^{-d/\lambda} \tag{13}$$

where d is the distance between wells (or points) and λ is the correlation length (the inverse of the correlation length is the correlation decay rate). Equation (13) represents the analytical correlation model (ACM).

Analytical approach—transinformation model

Marginal entropy, as a measure of information, was computed analytically using the expression (Lubbe, 1996)

$$H(x) = \ln(S_x) + 1{\cdot}419 \tag{14}$$

where S_x is the standard deviation for the random variable x.

Lubbe (1996) and Kapur and Kesavan (1992) estimated the values of $T(x, y)$, using the correlation coefficient (r_{xy}), as

$$T(x, y) = -0{\cdot}5 \ln(1 - r_{xy}^2) \tag{15}$$

Equations (14) and (15) are applicable only if the mean of the data is 0 and the data fit the normal probability distribution function. Derivation of Equations (14) and (15) is presented in Appendix B.

For the synthetic data, the correlation is represented by Equation (13) and therefore the analytical T model can be computed as

$$T(d) = -0{\cdot}5 \ln\lfloor 1 - (e^{-d/\lambda})^2 \rfloor \tag{16}$$

where r_{xy} is replaced in Equation (15) by $e^{-d/\lambda}$.

Fitting the discrete model with the analytical model

The coefficient of determination was used to quantify the goodness of fit between the analytical and the discrete models. The coefficient of determination (R^2) was computed as

$$R^2 = 1{\cdot}0 - \frac{SS_{\mathrm{reg}}}{SS_{\mathrm{tot}}} \tag{17}$$

where SS_{reg} is the sum of the squares of residuals between the discrete model and the analytical model, and SS_{tot} is the sum of the squares of residuals between the discrete model and the horizontal line through the mean.

COMPARISON OF DISCRETE AND ANALYTICAL APPROACHES

Correlation model (C model)

The discrete C model (DCM) was obtained by calculating the correlation between pairs of wells using Equation (1) and the distance between wells using Equation (12) (see Figure 1 and the description of synthetic data). The moving average method was used to smooth the discrete C model results ($\mathrm{DCM_{MA}}$). The analytical correlation model (ACM) was obtained by using Equation (13) and a distance interval equal to 100 m. Figure 2 shows the discrete C model, the analytical C model and the smoothed discrete C model, given by the moving average method ($\mathrm{DCM_{MA}}$). As expected, the analytical C model and the $\mathrm{DCM_{MA}}$ showed a good fit ($R^2 = 0{\cdot}94$).

Transinformation model (T model)

The discrete T model (DTM) was determined by calculating the discrete T values of pairs of wells using Equation (6), and the distance between wells was computed using Equation (12). The discrete T model was smoothed by the moving average method ($\mathrm{DTM_{MA}}$). Based on the correlation model (Equation 13), the analytical T model was determined using Equation (16) and a distance interval equal to 100 m. The DTM, $\mathrm{DTM_{MA}}$ and analytical T model (ATM) are plotted in Figure 3. This figure shows that the analytical T model fits to the $\mathrm{DTM_{MA}}$ ($R^2 = 0{\cdot}92$). However, the analytical T model does not fit perfectly to the $\mathrm{DTM_{MA}}$ (there is approximately 0·1 nats difference).

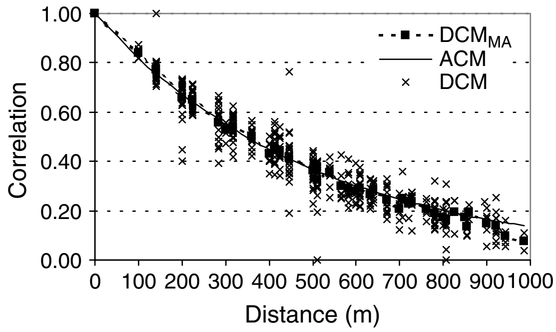

Figure 2. Correlation model applied to the synthetic data using the discrete and the analytical approaches. In the figure: DCM$_{MA}$ = smoothed discrete correlation model by moving average method; ACM = analytical correlation model; DCM = discrete correlation model

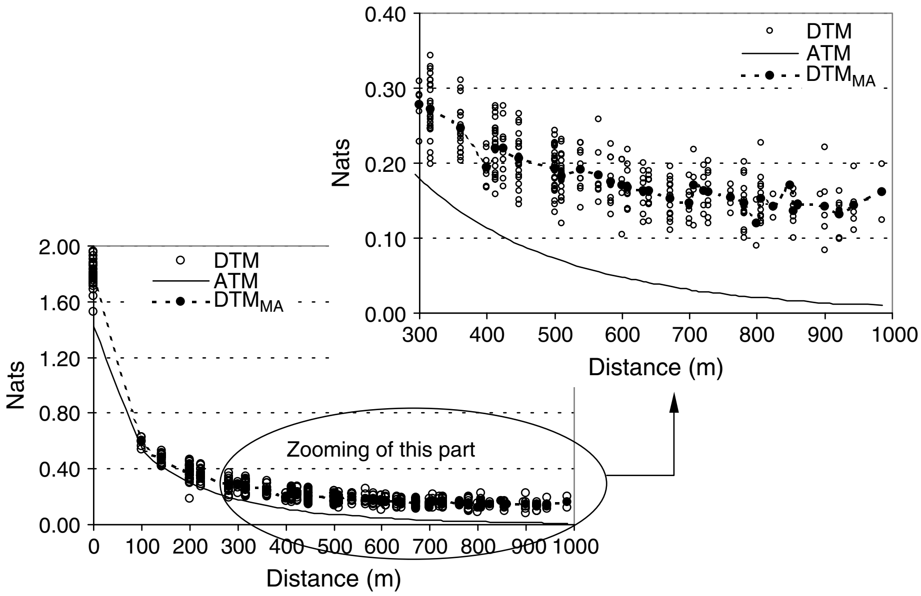

Figure 3. Transinformation model applied to the synthetic data using the discrete and analytical approaches. In the figure: DTM = discrete transinformation model; ATM = analytical transinformation model; DTM$_{MA}$ = smoothed discrete transinformation model by moving average method

Sensitivity analysis of the discrete T model. The discrete T model was obtained taking 200 as the size of generated data and 8 as the number of class intervals. As Figure 3 shows, there is a difference between the analytical T model and DTM$_{MA}$. For that reason a sensitivity analysis was performed for the discrete T model to investigate the factors that influence the difference between these models. The factors included the size of generated data and the number of class intervals.

Size of generated data. Different sizes of generated data were used to construct the discrete T model (200, 300, 400 and 500). The number of class intervals was the same for all the different sizes of generated data (the number of class intervals was 9). As shown in Figure 4, the larger the size of the generated data the less the difference there was between the DTM$_{MA}$ and analytical T model, as expected. This indicates that the discrete T model is sensitive to the size of the data available for analysis, as in the case of actual groundwater data where the data are limited in time or are incomplete.

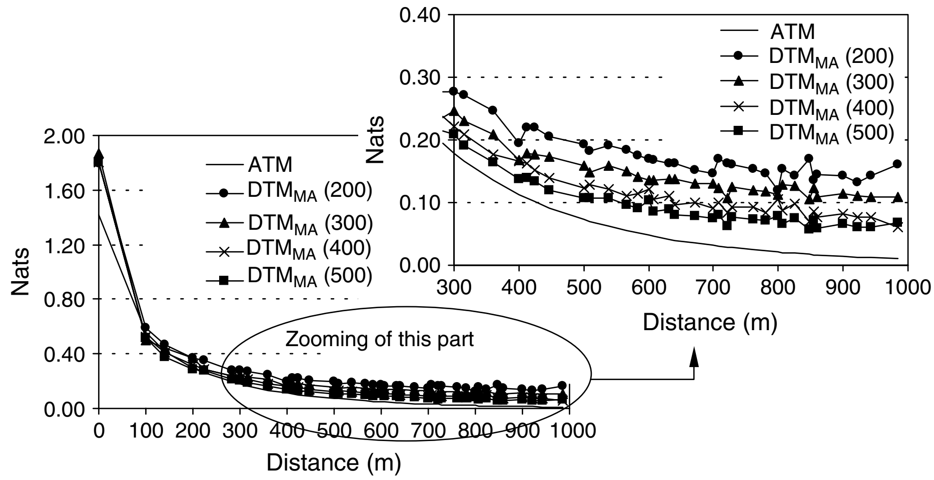

Figure 4. Discrete T-model for different sizes of generated data (200, 300, 400 and 500) compared with the analytical T model. In the figure: ATM = analytical transinformation model; DTM_{MA} = smoothed discrete transinformation model by moving average method

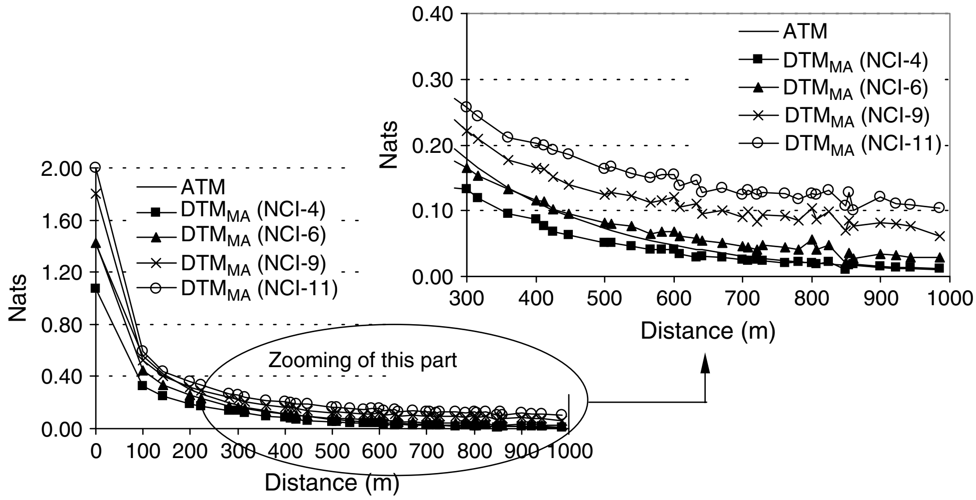

Figure 5. Discrete T-model for different class intervals (4, 6, 9 and 11) compared with the analytical T model. In the figure: ATM = analytical transinformation model; DTM_{MA} = smoothed discrete transinformation model by moving average method; NCI = number of class intervals

Number of class intervals (NCI). In order to evaluate the importance of the class interval, the size of the generated data used was 400 and four class intervals were analysed: 4, 6, 9 and 11. As shown in Figure 5, the discrete T model is influenced by the number of class intervals (NCI); when the class interval decreases the DTM_{MA} comes closer to the analytical T model. For this specific synthetic data example, the difference between the analytical T model and DTM_{MA} is smaller if the NCI is between 6 and 9.

CHARACTERIZATION OF SPATIAL VARIABILITY

In the literature, the C model has been used to characterize the spatial variability (linear dependency) of many types of data in different fields (e.g. Cressie, 1990). It is noted that the synthetic data are correlated by distance, which means that the smaller the distance the higher the correlation, as represented in Figure 2. The T model also can be used to represent the spatial variability of the synthetic data, as shown in Figure 3, which shows that there is a relationship between transinformation and distance. The closely spaced wells have a higher value for T than the ones that are further apart. The T values become essentially constant as the distance increases. That may be because there is still mutual information that can be transferred, even for a long distance. Figures 2 and 3 show that both T and C models can be used to represent the spatial dependency; however, there are some differences between the two models. There is a sharp decrease in the T model, which is not found in the correlation model. Also, the C model appears to have a higher scatter, in this case.

Figure 6 shows the smoothed discrete ITI model by the moving average method ($DITIM_{MA}$), or the normalized smoothed discrete T model. This figure shows that at around 500 m distance (the correlation length) the ITI reaches an essentially constant minimum value. Figure 6 also indicates that the ITI model could further provide a representation of the spatial variability of the synthetic data. This conclusion is also pointed out in Mogheir *et al*. (2004).

CONCLUSIONS

Synthetic data were used to compare the transinformation model (T model) and correlation model (C model). The models were also used to compare discrete and analytical approaches. The results of synthetic data analyses demonstrate that the class interval and the size of the data influence the T model results. It is also found that both C and T models using the discrete approach can be used to characterize the spatial variability by means of exponential curves. Both analytical models fit the discrete models data well (R^2 is quite high).

The analyses indicated that the presumed method can be applied successfully if the data are spatially correlated and fit the normal distribution function with 0 mean. Mogheir *et al*. (2004) used different sets of data from the Gaza Strip (groundwater quality data) to demonstrate the applicability of these procedures to real data. The main concern of these two articles is the characterization of the spatial structure of the groundwater quality variables by means of the transinformation model, which is a preliminary step for using the entropy theory in assessing and redesigning the spatial locations of monitoring wells.

ACKNOWLEDGEMENTS

The authors gratefully thank the Foundation for Science and Technology of the Portuguese Ministry of Science and Technology for sponsoring the fellowship (Reference: SFRH/BD/6089/2001). The fellowship

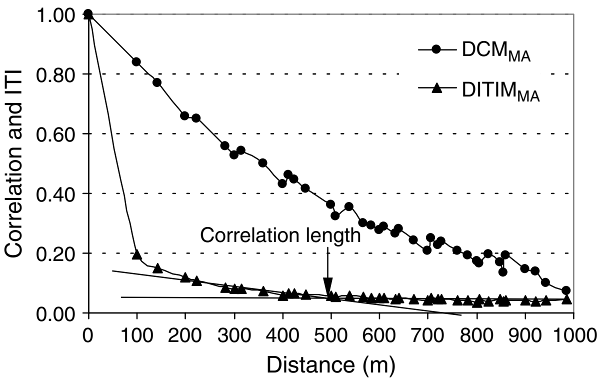

Figure 6. C-model and ITI-model using the discrete approach, for the synthetic data. In the figure: DCM_{MA} = smoothed discrete correlation model by moving average method; $DITIM_{MA}$ = smoothed discrete information transfer index model by moving average method

was provided for the first author's PhD programme entitled 'Quantification of Information for Groundwater Quality Networks'. The programme is being undertaken at the Department of Civil Engineering, Faculty of Science and Technology of the University of Coimbra, Coimbra, Portugal.

APPENDIX A

Two-dimensional contingency table

To illustrate the construction of a two-dimensional contingency table for field data, two wells (H-9 and H-8) were selected from the Gaza Strip groundwater quality monitoring network. The time-series of the chloride data of the two wells are presented in Table AI. For fuller description of the Gaza Strip data see Mogheir *et al*. (2004). The construction involved the following steps.

1. The time-series of a given water quality constituent observed at H-39 and H-8 was drawn in the same figure, as shown for chloride in Figure A1.
2. Each data set was subdivided into class intervals.
3. To fill the first table (frequency table), the rows or the columns were kept constant and the shared data of the other well were counted, as shown in Table AII.
4. The joint probability table was constructed by dividing each count by the total number of the recorded data of one well, as shown in Table AIII.

APPENDIX B

Analytical computation of marginal entropy and transinformation

Marginal entropy (Equation 14). The marginal entropy has been computed analytically by Lubbe (1996) for the normal distribution. A random variable x has a normal or Gaussian distribution if the probability

Table AI. Chloride data for well H-39 and H-8

Date	H-39 (Cl mg/l)	H-8 (Cl mg/l)	Date	H-39 (Cl mg/l)	H-8 (Cl mg/l)	Date	H-39 (Cl mg/l)	H-8 (Cl mg/l)
01-05-1972	644	427	20-03-1980	721	532	27-01-1990	868	700
28-10-1972	679	413	16-09-1980	749	546	26-07-1990	854	707
26-04-1973	721	483	15-03-1981	756	518	22-01-1991	840	770
23-10-1973	805	497	11-09-1981	861	525	21-07-1991	845	784
21-04-1974	693	483	10-03-1982	840	602	17-01-1992	819	770
18-10-1974	805	497	06-09-1982	861	630	15-07-1992	826	784
16-04-1975	693	504	05-03-1983	959	630	11-01-1993	819	805
13-10-1975	679	518	01-09-1983	882	644	10-07-1993	819	805
10-04-1976	721	511	28-02-1984	854	651	06-01-1994	819	784
07-10-1976	805	553	26-08-1984	868	658	05-07-1994	819	784
05-04-1977	658	630	22-02-1985	854	665	01-01-1995	819	777
02-10-1977	756	497	21-08-1985	868	644	30-06-1995	763	791
31-03-1978	735	504	17-02-1986	833	651	27-12-1995	714	777
27-09-1978	756	497	16-08-1986	868	658	24-06-1996	767	829
26-03-1979	735	504	12-02-1987	770	721	21-12-1996	739	921
22-09-1979	728	525	11-08-1987	868	658	19-06-1997	752	822
			07-02-1988	854	721			
			05-08-1988	819	707			
			01-02-1989	840	784			
			31-07-1989	819	707			

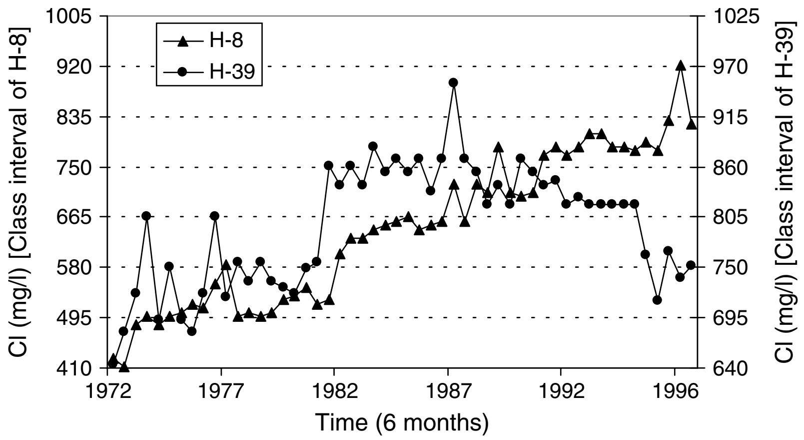

Figure A1. Chloride time-series for a two-well combination (H-8 and H-39). The wells are selected from groundwater quality monitoring network in the middle part of Gaza Strip and used in Mogheir *et al*. (2004)

Table AII. Absolute frequency contingency table for H-8 and H-39 combinations

	410 < Cl < 495	495 < Cl < 580	580 < Cl < 665	665 < Cl < 750	750 < Cl < 835	835 < Cl < 920	Marginal H-39
640 < Cl < 695	3	1	2	2	0	0	8
695 < Cl < 750	2	6	1	1	1	0	11
750 < Cl < 805	1	0	0	4	6	1	12
805 < Cl < 860	0	0	1	4	1	0	6
860 < Cl < 915	0	1	3	10	0	0	14
915 < Cl < 970	0	1	0	0	0	0	1
Marginal H-8	6	9	7	21	8	1	52

Table AIII. Joint probability (contingency) table for H-8 and H-39 combinations

	410 < Cl < 495	495 < Cl < 580	580 < Cl < 665	665 < Cl < 750	750 < Cl < 835	835 < Cl < 920	Marginal H-39
640 < Cl < 695	0·058	0·019	0·038	0·038	0·000	0·000	0·154
695 < Cl < 750	0·038	0·115	0·019	0·019	0·019	0·000	0·212
750 < Cl < 805	0·019	0·000	0·000	0·077	0·115	0·019	0·231
805 < Cl < 860	0·000	0·000	0·019	0·077	0·019	0·000	0·115
860 < Cl < 915	0·000	0·019	0·058	0·192	0·000	0·000	0·269
915 < Cl < 970	0·000	0·019	0·000	0·000	0·000	0·000	0·019
Marginal H-8	0·115	0·173	0·135	0·404	0·154	0·019	1·000

distribution $p(x)$ for $-\infty < x < +\infty$ is given by

$$p(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right] \quad \text{(B1)}$$

where μ_x and σ_x are the normal distribution parameters (mean and variance).

The mean of variable x is defined by

$$\mu_x = \int_{-\infty}^{+\infty} x p(x) \mathrm{d}x \tag{B2}$$

The variance, which is a measure of the variation of the values of x around its mean, is defined by

$$\sigma_x = \int_{-\infty}^{+\infty} (x - \mu_x)^2 p(x) \mathrm{d}x \tag{B3}$$

For the continuous variable x with probability density function $p(x)$, the marginal entropy is equal to

$$H(x) = -\int_{-\infty}^{+\infty} p(x) \ln p(x) \mathrm{d}x \tag{B4}$$

The marginal entropy of x, $H(x)$, is maximum if and only if

$$p(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \tag{B5}$$

Equation (B5) is obtained by substituting $\mu_x = 0$ in Equation (B1). Therefore, the following marginal entropy derivation is valid only for the normalized data where $\mu_x = 0$ and σ_x is the standard deviation of variable x, and is constant.

Substitution of Equation (B5) in Equation (B4) yields

$$H(x) = -\int_{-\infty}^{+\infty} \left[\frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \ln \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right)\right] \mathrm{d}x \tag{B6}$$

$$= -\int_{-\infty}^{+\infty} \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \left[\ln 1 - \ln(\sigma_x \sqrt{2\pi}) - \frac{x^2}{2\sigma_x^2}\right] \mathrm{d}x$$

$$= \int_{-\infty}^{+\infty} \frac{\ln(\sigma_x \sqrt{2\pi})}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \mathrm{d}x + \int_{-\infty}^{+\infty} \frac{x^2}{2\sigma_x^2} \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \mathrm{d}x$$

$$= \ln(\sigma_x \sqrt{2\pi}) \int_{-\infty}^{+\infty} \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \mathrm{d}x + \frac{\ln e}{2\sigma_x^2} \int_{-\infty}^{+\infty} x^2 \left[\frac{1}{\sigma_x \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right)\right] \mathrm{d}x \tag{B7}$$

Substitution of Equation (B3) in Equation (B7) yields

$$H(x) = \ln(\sigma_x \sqrt{2\pi}) + \frac{\ln e}{2\sigma_x^2} \sigma_x^2 \tag{B8}$$

Then, Equation (B8) can be written as

$$H(x) = \ln(\sigma_x \sqrt{2\pi}) + \frac{1}{2} \ln e = \ln \sigma_x + \ln \sqrt{2\pi e} \tag{B9}$$

Therefore, the marginal entropy can be obtained as

$$H(x) = \ln \sigma_x + 1{\cdot}419 \qquad \text{(B10)} \equiv \text{Equation(14)}$$

Transinformation (Equation 15)

Equation (B9) can be written as

$$H(x) = \frac{1}{2}\ln \sigma_x^2 + \frac{1}{2}\ln(2\pi e) \qquad \text{(B11)}$$

For variable y the marginal entropy is

$$H(y) = \frac{1}{2}\ln \sigma_y^2 + \frac{1}{2}\ln(2\pi e) \qquad \text{(B12)}$$

where σ_y is the standard deviation of variable y.

The joint probability densities $p(x, y)$ for variable x and y, where $-\infty < x < +\infty$, $-\infty < y < +\infty$ and considering two-dimensional Gaussian distribution with mean for both x and y equal to 0, could be expressed as (Kapur and Kesavan, 1992)

$$p(x, y) = \frac{1}{2\pi|\mathbf{C}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}[x \quad y]\mathbf{C}^{-1}\begin{bmatrix} x \\ y \end{bmatrix}\right\} \qquad \text{(B13)}$$

where

$$\mathbf{C} = \begin{bmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{yx}\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \qquad \text{(B14)}$$

ρ_{xy} is the correlation coefficient between variable x and y and can be calculated as

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \qquad \text{(B15)}$$

where σ_{xy} is the covariance between the variable x and y.

For the continuous variable x and y with joint probability density $p(x, y)$, the joint entropy is equal to

$$H(x, y) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x, y)\ln p(x, y)\mathrm{d}x\mathrm{d}y \qquad \text{(B16)}$$

By substituting Equations (B13) and (B14) in Equation (B16) we obtain

$$H(x, y) = \ln(2\pi e) + \frac{1}{2}\ln|\mathbf{C}| \qquad \text{(B17)}$$

For computing the transinformation $T(x, y)$ for two random variables, x and y, the following expression can be used (e.g. Jessop, 1995)

$$T(x, y) = H(x) + H(y) - H(x, y) \qquad \text{(B18)}$$

Using Equations (B11), (B12) and (B17) in Equation (B18) we obtain

$$T(x, y) = \frac{1}{2}\ln \sigma_x^2 + \frac{1}{2}\ln(2\pi e) + \frac{1}{2}\ln \sigma_y^2 + \frac{1}{2}\ln(2\pi e) - \ln(2\pi e) - \frac{1}{2}\ln|\mathbf{C}| \qquad \text{(B19)}$$

Then Equation (B19) can be simplified as

$$\begin{aligned} T(x, y) &= \frac{1}{2}\ln(\sigma_x^2\sigma_y^2) - \frac{1}{2}\ln|\mathbf{C}| \\ &= \frac{1}{2}\ln(\sigma_x^2\sigma_y^2) - \frac{1}{2}\ln\left(\sigma_x^2\sigma_y^2\begin{vmatrix} 1 & \rho_{xy} \\ \rho_{yx} & 1 \end{vmatrix}\right) \\ &= \frac{1}{2}\ln(\sigma_x^2\sigma_y^2) - \frac{1}{2}\ln(\sigma_x^2\sigma_y^2) - \frac{1}{2}\ln\begin{vmatrix} 1 & \rho_{xy} \\ \rho_{yx} & 1 \end{vmatrix} \end{aligned} \quad \text{(B20)}$$

Therefore, $T(x, y)$ can be obtained as

$$T(x, y) = -\frac{1}{2}\ln(1 - \rho_{xy}^2) \quad \text{(B21)}$$

Using the sample correlation coefficient r_{xy} in Equation (B21), we obtain

$$T(x, y) = -\frac{1}{2}\ln(1 - r_{xy}^2) \quad \text{(B22)} = \text{Equation(15)}$$

APPENDIX C

List of symbols and abbreviations

Symbols.

$\mathbf{C}$	covariance matrix
$\lvert\mathbf{C}\rvert$	covariance determinant
cov_{xy}	sample covariance between x and y
d	distance between wells (m)
$f_{i.}$	marginal frequency
f_{ij}	joint frequency
$H(x)$	marginal entropy of x (nats)
$H(x, y)$	joint entropy of x and y (nats)
$H(x/y)$	conditional entropy of x given y (nats)
$\mathrm{ITI}(x, y)$	information transfer index between x and y
$p(x)$	probability distribution of x
$p(x_i)$	probability of occurrence
$p(x, y)$	joint probability distribution of x and y
$p(x/y)$	conditional probability distribution of x given y
R^2	coefficient of determination
$r(d)$	correlation as a function of distance (analytical correlation model, ACM)
r_{xy}	sample correlation coefficient between x and y
SS_{reg}	sum of the squares of the residuals between the discrete model and the best-fit curve (analytical model)
SS_{tot}	sum of squares of the residuals between the discrete model and the horizontal line through the mean
S_x and S_y	sample standard deviation of variable x and y respectively
$T(x, y)$	transinformation between x and y (nats)
x and y	discrete variables
$\overline{x}$ and $\overline{y}$	sample mean of variable x and y respectively
μ_x	population mean of variable x
σ_x	population standard deviation of variable x

σ_x^2	population variance of variable x
σ_{xy}	population covariance between x and y
ρ_{xy}	population correlation coefficient between x and y
λ	correlation length (m)

Abbreviations.

ACM	analytical correlation model
ATM	analytical transinformation model
C-model	correlation model
DCM	discrete correlation model
DITIM	discrete ITI model
DTM	discrete transinformation model
ITI-model	information transfer index model
DCM_{MA}	smoothed discrete correlation model by moving average method
$DITIM_{MA}$	smoothed discrete ITI model by moving average method
DTM_{MA}	smoothed discrete transinformation model by moving average method
T-model	transinformation model

REFERENCES

Berthouex PM, Brown, LC. 1994. *Statistics for Environmental Engineers*. CRC Press: Florida; 335 pp.

Bueso MC, Angulo JM, Cruz-Sanjulian J, Carcia-Arostegui JL. 1999. Optimal spatial sampling design in a multivariate framework. *Mathematical Geology* **31**(5): 507–525.

Caselton WF, Husain T. 1980. Hydrologic network: information transmission. *Journal of Water Resources Planning and Management Division ASCE* **106**(WR2): 503–529.

Cressie NC. 1990. *Statistics for Spatial Data*, revised edition. Wiley: New York; 900 pp.

Darbellay GA, Wuertz D. 2000. The entropy as a tool for analysing statistical dependence in financial time series. *Physica A* **287**(3–4): 429–439.

Harmancioglu NB. 1981. Measuring the information content of hydrological processes by the entropy concept. *Centennial of Ataturk's Birth, Journal of the Civil Engineering Faculty of Ege University:* 13–88.

Harmancioglu NB, Alpaslan N. 1992. Water quality monitoring network design. *Water Resources Bulletin* **28**(1): 179–192.

Harmancioglu NB, Fistikoglu O, Ozkul SD, Singh VP, Alpaslan MN. 1999. *Water Quality Monitoring Network Design*. Kluwer Academic Publishers: Boston; 299 pp.

Harmancioglu NB, Yevjevich V. 1987. Transfer of hydrologic information among river points. *Journal of Hydrology* **91**: 103–118.

Husain T. 1989. Hydrologic uncertainty measure and network design. *Water Resources Bulletin* **25**(3): 527–534.

Jessop A. 1995. *Informed Assessments, an Introduction to Information, Entropy and Statistics*. Ellis Horwood: New York; 366 pp–.

Kapur JN, Kesavan HK. 1992. *Entropy Optimisation Principles with Applications*. Academic Press: San Diego; 408 pp.

Krastanovic PF, Singh VP. 1992. Evaluation of rainfall networks using entropy II. *Water Resources Management* **6**: 295–314.

Lee Y, Ellis JH. 1997. On the equivalence of Kriging and maximum entropy estimators. *Mathematical Geology* **29**(1): 131–152.

Lubbe CA. 1996. *Information Theory*. Cambridge University Press: Cambridge; 350 pp–.

Mogheir Y, Singh VP. 2002. Application of information theory to groundwater quality monitoring networks. *Water Resources Management* **16**(1): 37–49.

Mogheir Y, de Lima JLMP, Singh VP. 2004. Characterizing the spatial variability of groundwater quality using the entropy theory: II. Case study from Gaza Strip. *Hydrological Processes* (in press).

Ozkul S, Harmancioglu NB, Singh VP. 2000. Entropy-based assessment of water quality monitoring networks. *Journal of Hydrologic Engineering, American Society of Civil Engineers* **5**(1): 90–100.

Rojdestvenski I, Cottam MG. 2000. Mapping of statistical physics to information theory with application to biological system. *Journal of Theoretical Biology* **202**(1): 43–54.

Shannon CE. 1948. A mathematical theory of communication, II and I. *Bell System Technical Journal* **27**: 379–423.

Singh VP. 1997. The use of entropy in hydrology and water resources. *Hydrological Processes* **11**: 587–626.

Singh VP. 1998. *Entropy-based Parameter Estimation in Hydrology*. Kluwer Academic Publishers: Boston.

Sy BK. 2001. Information–statistical pattern based approach for data mining. *Journal of Statistical Computation and Simulation* **69**(2): 171–201.

Ulanowicz RE. 2001. Information theory in ecology. *Computers and Chemistry* **25**(4): 393–399.

Uslu O, Tanriover A. 1979. Measuring the information content of hydrological process. *Proceedings of the First National Congress on Hydrology*, Istanbul; 437–443.

Yang Y, Burn D. 1994. An entropy approach to data collection network design. *Journal of Hydrology* **157**: 307–324.

Zhou Y. 1996. *Spatial Data Generation Program (COVRAN)*. IHE Delft, The Netherlands.

