WATER RESOURCES RESEARCH, VOL. 46, W03528, doi:10.1029/2009WR008101, 2010

Information theory–based approach for location of monitoring water level gauges in polders

Leonardo Alfonso,[1] Arnold Lobbrecht,[1,2] and Roland Price[1]

Received 11 April 2009; revised 28 October 2009; accepted 2 November 2009; published 23 March 2010.

[1] Data collection is a critical activity in the management of water systems because it supports informed decision making. Data are collected by means of monitoring networks in which water level gauges are of particular interest because of their implications for flood management. This paper introduces a number of modifications to previously published methods that use information theory to design hydrological monitoring networks in order to make the methods applicable to the design of water level monitors for highly controlled polder systems. The new contributions include the use of a hydrodynamic model for entropy analysis, the introduction of the quantization concept to filter out noisy time series, and the use of total correlation to evaluate the performance of three different pairwise dependence criteria. The resulting approach, water level monitoring design in polders (WMP), is applied to a polder in the Pijnacker region, Netherlands. Results show that relatively few monitors are adequate to collect the information of a polder area in spite of its large number of target water levels. It is found, in addition, that the directional information transfer DIT_{YX} is more effective in finding independent monitors, whereas DIT_{XY} is better for locating sets of monitors with high joint information content. WMP proves to be a suitable and simple method as part of the design of monitoring networks for polder systems.

Citation: Alfonso, L., A. Lobbrecht, and R. Price (2010), Information theory–based approach for location of monitoring water level gauges in polders, *Water Resour. Res.*, *46*, W03528, doi:10.1029/2009WR008101.

1. Background

[2] Polders are developed areas that are below sea level and are drained by canals and pumping stations that discharge excess water either to elevated storage basins or to the sea. In this way, artificial catchments are formed consisting of the combination of a number of polders normally delimited by weirs or pump stations. Consequently, there are a considerable number of different target water levels that need to be maintained rigorously within predefined ranges in relatively small, hydraulically disconnected areas. For this reason, polder areas become very difficult to monitor because it is very expensive to place a water level gauge at every significant location. This research focuses on the design of monitoring networks required to control polder water systems using information theory. In particular, the objective is to find the minimum number of monitors that offer the best description (information content) for controlling a water system. The motivation of this research is in line with the conclusions of *Mishra and Coulibaly* [2009], who stated that most studies on the design of monitoring networks focus on prediction and that research should concentrate on other users and needs.

[3] This paper is organized as follows. First, information theory is introduced, followed by a review of the methods for the evaluation and design of monitoring networks, including some entropy-based approaches. Subsequently, procedures are presented for data preparation, gauge location, and the evaluation of solutions. Then a case study is introduced with an analysis and discussion of the results. A summary of the findings concludes the paper.

[1]Hydroinformatics and Knowledge Management, UNESCO-IHE, Delft, Netherlands.
[2]HydroLogic BV, Amersforot, Netherlands.

0043-1397/10/2009WR008101

2. Information Theory

[4] Information theory, as described by *Shannon* [1948], provides mechanisms for measuring the information content of random variable X, which is defined as a reduction in the uncertainty $H(X)$. This quantity is also known as entropy, information entropy, Shannon entropy, or marginal entropy. Consequently, the terms "entropy" and "uncertainty" will be used interchangeably throughout this paper. The definition of uncertainty indicates how surprising it is, on average, to get a value x from a random variable X that can take the possible values $x_1, x_2, \ldots, x_n$ each with probability $p(x)$:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i). \tag{1}$$

The units of uncertainty are actually given by the base of the logarithm, being "nats" if the base is e and "bits" if it is 2. In this paper the latter base will be used. Another important consideration is that $x \log x \to 0$ as $x \to 0$ and values with zero probability do not change the uncertainty. It is possible to see that entropy is a measure of variability or dispersion but is better than the classical variance when the size of the sample is small [*Hart*, 1971; *Mogheir et al.*, 2006; *Singh*, 1997, 2000]. A more detailed discussion of the advantages of entropy is given by *Wei* [1987].

[5] In some cases, it is necessary to estimate the information content between two random variables X and Y. A frequently used measure is mutual information (or transinformation) I, which quantifies the amount of information of one random variable that is contained in another random variable [*Cover and Thomas*, 1991], and it can be interpreted as the reduction in the uncertainty of X due to knowledge of Y:

$$I(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \quad (2)$$

Given that $p(x, y)$ is the joint distribution between the variables X and Y and $p(y|x)$ is the probability of y given x, other information-related measures for two variables can be defined [*Cover and Thomas*, 1991], such as the joint entropy,

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log p(x_i, y_j), \quad (3)$$

which represents the amount of information contained in both variables, and the conditional entropy,

$$H(Y|X) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log p(y_i|x_j), \quad (4)$$

which represents the amount of information of X not contained in Y.

[6] The use of mutual information has become popular in several fields of science to measure the dependence between variables. Some of its advantages have been reported widely [e.g., *Li*, 1990; *Linfoot*, 1957; *Singh*, 2000; *Steuer et al.*, 2002]. A major restriction on the mutual information is that it is only applicable to two random variables. However, often there is a need to evaluate the dependence among several variables, which involves a difficult assessment of multivariate joint probabilities. A number of pairwise approximations have been proposed for this assessment [*Chow and Liu*, 1968; *Kirshner et al.*, 2004; *Lewis*, 1959]. Another way of looking at multivariate dependence is by assessing the total amount of information that is shared by all variables at the same time, without considering any dependencies among partial combinations of the variables. In this context, the total correlation C,

$$C(X_1, X_2, \ldots, X_N) = \sum_{i=1}^{N} H(X_i) - H(X_1, X_2, \ldots, X_N), \quad (5)$$

can be used [*McGill*, 1954; *Watanabe*, 1960]. The total correlation is always positive since the sum of the entropies of all of the variables will always be greater than the joint entropy of all of them, and C is equal to zero if and only if all the variables being considered are independent. However, C is greater than zero if two of the variables have some dependence, even though the rest of the variables are independent. It is evident that the mutual information is a special case of total correlation when the number of variables is two. The total correlation concept has been used extensively in medicine, neurology, psychology, clustering, feature selection, and genetics. Complete references to the use of total correlation in these fields are given by *Jakulin and Bratko* [2004] and *Fass* [2006]. However, there appear to have been no applications to water resources.

[7] Although equation (5) involves the estimation of multivariate joint entropies (and therefore the estimation of multivariate joint probabilities), it is not really necessary to estimate them. The value of C can be estimated directly by the systematic grouping of bivariate joint entropies. In brief, the grouping property of total correlation states that if new variables are created by the union of two or more original variables in such way that the marginal entropy of a new variable is equal to the joint entropy of the two original variables, then the total correlation of the original variables can be computed by summing the marginal entropies of the new variables [see *Kraskov et al.*, 2005].

3. Monitoring Networks

[8] A vital part of good management practice for a water system is the measurement and collection of data that are needed to describe the hydrological processes of interest in such a system at certain spatial and temporal scales. For this purpose, monitoring networks are designed and sometimes optimized for decision making according to water management objectives [*Loucks and van Beek*, 2005]. An example of a monitoring network is a collection of discrete rain gauges located at convenient points in a catchment to estimate the precipitation over the catchment area. Similarly, a water level gauge network is composed of monitors located in a water system to register water level variations, which, depending on the objectives, can be used for warning, control strategies, or design [see *Mishra and Coulibaly*, 2009].

[9] Although there is extensive literature on different approaches to rain gauge network design [see, e.g., *Bogárdi et al.*, 1985; *Bras and Rodríguez-Iturbe*, 1976; *Moore et al.*, 2000; *Pardo-Igúzquiza*, 1998; *Rodríguez-Iturbe and Mejia*, 1974; *Sansó and Müller*, 1997; *Yeh et al.*, 2006], there are comparatively few papers on the design of water level gauge networks [*Karasev*, 1968; *Moss*, 1976; *Moss and Karlinger*, 1974; *Moss and Tasker*, 1991] and none on water level monitoring in polder systems.

4. Information Theory in Monitoring Network Design

[10] The main idea behind the entropy-based design of a monitoring network is to reduce as much as possible the uncertainty associated with the estimation of values of a certain variable in places where it is not measured directly. The concept of uncertainty has been traditionally linked with the statistical variance, even though *Amorocho and Espildora* [1973] noticed that it was not an objective index of quality when used, for instance, in comparing predictions of a hydrological model and the corresponding data records. A similar observation was made in the field of portfolio management [see *Philippatos and Wilson*, 1972].

[11] A number of authors have applied information theory concepts to the design or evaluation of monitoring networks for general purposes [*Caselton and Zidek*, 1984] and for more specific purposes, such as water quality [*Harmancioglu et al.*, 1999], groundwater quality [*Caselton and Husain*, 1980; *Mogheir et al.*, 2004; *Mogheir and Singh*, 2002; *Mogheir et al.*,

2006], air pollution [*Zidek et al.*, 2000], and rainfall gauging stations [*Krstanovic and Singh*, 1992a; 1992b], among others.

[12] From the point of view of surface water gauging, *Husain* [1989] presented a method for network design based on the information-transmitting capabilities of a hydrologic network in terms of entropy. *Yang and Burn* [1994] criticized this method, pointing out that a continuous distribution function is assumed when calculating the entropy-related measurements. Even though these authors overcame this problem by using a nonparametric estimation of the density distributions, the assumptions of having independent and identically distributed random variables and the assessment of the smoothing factor of the kernel parameter still continue to add vagueness to the process. These authors, nevertheless, presented a normalized version of the mutual information between two gauges, called the directional information transfer index (DIT), to obtain the fraction of information transferred from one site to another as a value between 0 and 1:

$$\mathrm{DIT}_{XY} = \frac{I(X;Y)}{H(X)}. \quad (6)$$

This expression was first introduced in 1970 by C. H. Coombs, R. Dawes, and A. Tversky in the field of mathematical psychology as the coefficient of constraint [*Fass*, 2006]. *Markus et al.* [2003] explained DIT_{XY} as the information received by X from Y. When $H(Y)$ is used in the denominator of equation (6), DIT_{YX} becomes the information sent from X to Y. Markus et al. present a comparison between entropy and the least squares method to evaluate stream gauges, in which the DIT of *Yang and Burn* [1994] was adopted. The authors faced the problem of selecting the bin size for calculating the empirical frequency analysis in order to obtain the probability of a value in a particular interval. They found that, in spite of the differences in entropy values when changing the bin size, the ranking of stations in terms of the difference between the information received and the information sent remained, in general, unchanged. The problem of the bin size was pointed out from the beginning by *Amorocho and Espildora* [1973] and has also been studied in the calculation of the mutual information for discrete variables by *Steuer et al.* [2002]. A related analysis on information transfer in monitoring networks is given by *Ruddell and Kumar* [2009], who use the concept of transfer entropy to estimate the information of the current state of X by looking at the past states of Y that are not present in the past states of X itself.

[13] Two difficulties appear recurrently in the studies that use entropy for monitoring network design. First, there is a problem of establishing the joint probability functions to calculate mutual information. This has been mainly solved by either assuming a Gaussian distribution of the variables or evaluating the transinformation as a function of the correlation coefficient r, as suggested by *Harmancioglu and Yevjevich* [1987]. Second, for the multivariate case, several simplifications are made, for example, by analyzing mutual information of pairs of stations and analyzing the resulting 2-D transinformation matrices [*Filippini et al.*, 1994; *Mogheir and Singh*, 2002] or by assuming a normal distribution to calculate the multivariate joint entropy [*Krstanovic and Singh*, 1992a].

[14] Even though the concept of total correlation for evaluating information among multiple variables is an indirect approach to the evaluation of the multivariate joint probability functions, it is a direct and precise method for evaluating the information dependence among multiple variables. Nevertheless, equation (5) has not been used to date for evaluating or designing stream gauge networks.

5. Methods

[15] An approach called water level monitoring design in polders (WMP) for locating and evaluating water level gauges in a water system composed of a highly controlled canal network is presented here. It consists of five parts: (1) the generation of time series at a very dense set of calculation points using a hydrodynamic model, (2) a quantization method to "clean" the noise from the time series, (3) the use of three different pairwise criteria to evaluate dependence using squared matrices in a similar fashion to that used by *Mogheir and Singh* [2002], (4) a procedure to locate the gauges following a method similar to that of *Krstanovic and Singh* [1992a, 1992b], which aims to find the set of points that together provide the highest information content and are at the same time the most independent of each other, and (5) the evaluation of the multivariate dependence with equation (5), using the grouping property of the total correlation [*Fass*, 2006; *Kraskov et al.*, 2005] in order to establish comparisons between the gauges. The first criterion of the procedure in step 4 is evaluated with equation (1), and the second is evaluated by three different pairwise methods: transinformation (equation (2)) and DIT_{XY} and DIT_{YX} (equation (6)). Although the concept of transfer entropy [*Schreiber*, 2000] and its application in monitoring design [*Ruddell and Kumar*, 2009] promote a new pairwise dependence criterion, the dynamic analysis of the time series is beyond the scope of this paper: a single rainfall event is considered instead of a long time series.

[16] The probabilities required for the estimation of entropy and mutual information are calculated using the well-known histogram-based technique, as described, for example, by *Steuer et al.* [2002]; in this way the choice of a probability distribution to fit the continuous data is avoided. Although there exist a number of nonparametric methods to estimate mutual information [see, e.g., *Moon et al.*, 1995; *Sharma*, 2000], this paper uses bins as an opportunity to take into account water management issues. The subjective determination of the bin size for the histogram construction [*Ruddell and Kumar*, 2009] is addressed here using the quantization method introduced below.

6. Quantization

[17] The quantization concept comes from the theory of communication systems and aims to convert an analog (i.e., continuous) sign into a discrete pulse in order to allow its digital transmission by applying the mathematical floor function (denoted by floor brackets). The conversion of an analog value x to a quantized value x_q, which is rounded to the nearest multiple of a, is performed by

$$x_q = a\left\lfloor\frac{2x+a}{2a}\right\rfloor. \quad (7)$$

The relationship between the bin size b and the parameter a is given by the quotient of the difference between the

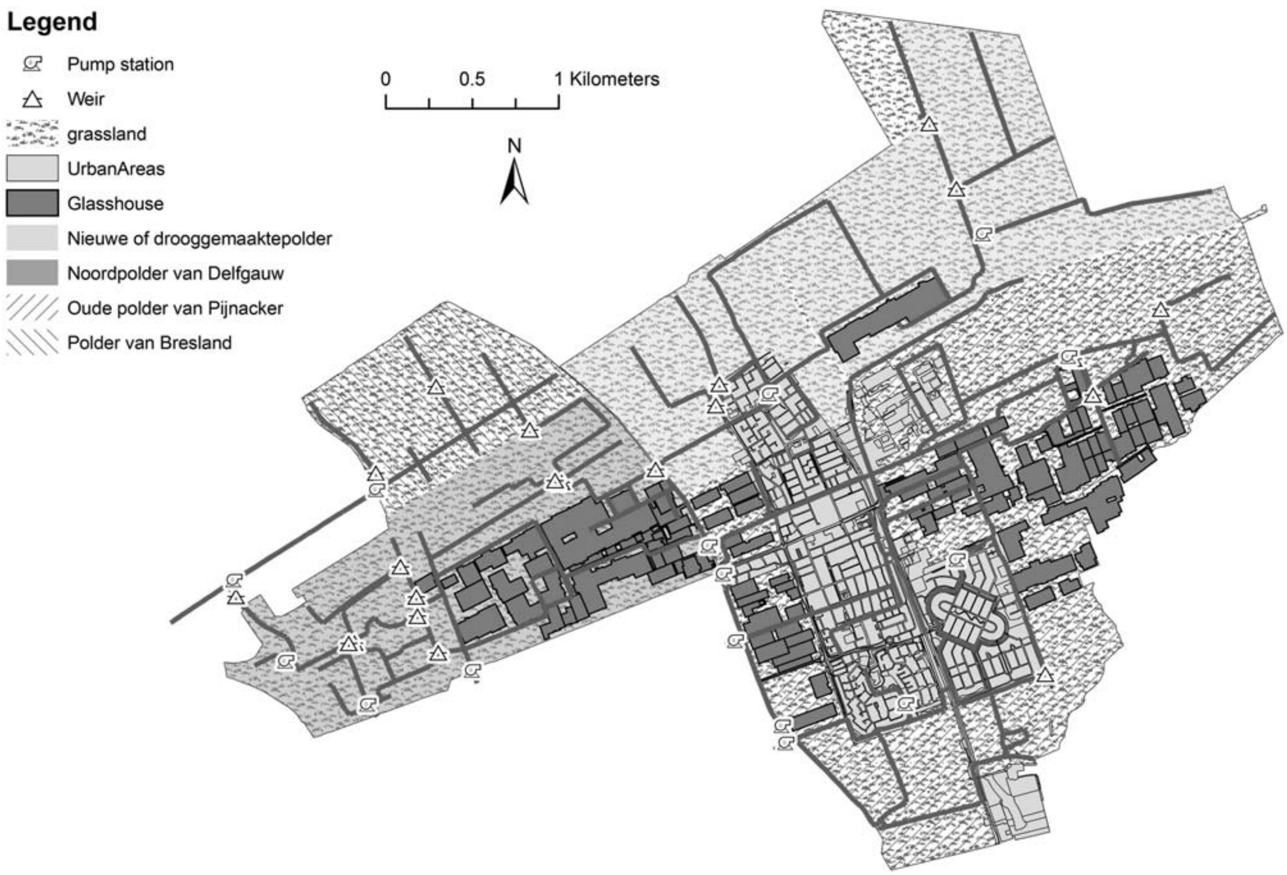

Figure 1. Canal network in Pijnacker region, Delfland, Netherlands. Water is drained from the area through the six pumping stations at its southwest boundary.

maximum and the minimum of the time series and the value *a*.

[18] In the context of this paper, water level time series are transformed through equation (7) into "pulses" of discrete information, which produces a regular discretization of water levels (quantization) at irregular intervals of time. An advantage of the approach is that the quantized water level series are "noise-free" in the sense that high-frequency, low-amplitude water level changes (i.e., dynamic waves) that are generated by neighboring pumping stations are

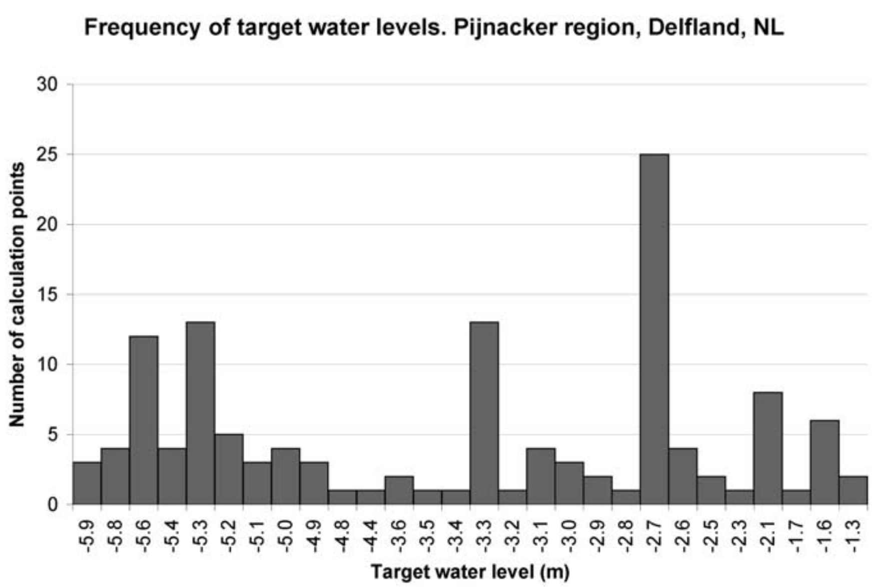

Figure 2. Frequency of target water levels at calculation points, Pijnacker, Delfland. In total, 28 unique target levels out of the 130 existing hydrological units are shown.

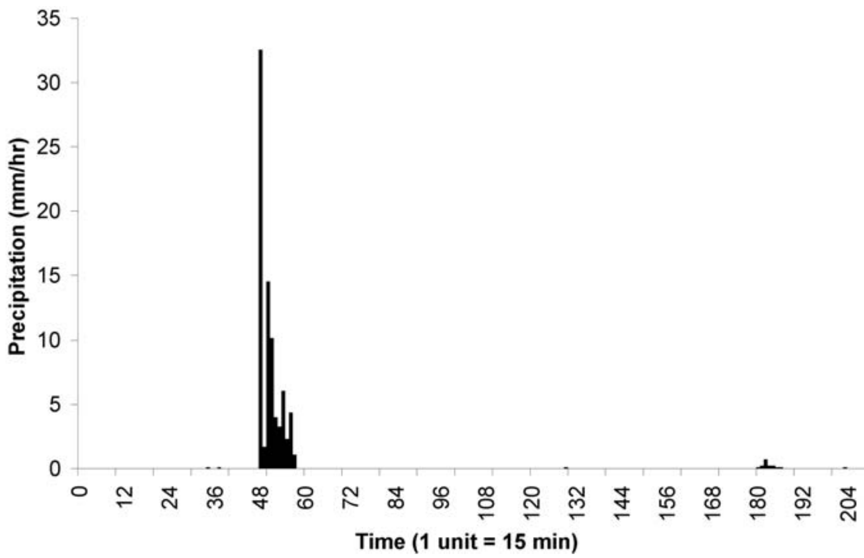

Figure 3. Rainfall event used in the hydrodynamic model.

filtered out. Consequently, the value of a can be seen as the minimum dimensional unit of water level for which the management of the system becomes critical. This is crucial when computing equation (1) since high-frequency water level signals give very high values of entropy but do not necessarily provide information content for water management decision making. It must be noted that even though the quantization alters the water volume at a point, this is not important since only probabilities of occurrence of the values are taken into consideration for the entropy calculation.

7. Location of Gauges

[19] Two different conditions must be considered when locating the gauges: (1) the monitors must be as independent of each other as possible (i.e., have a low pairwise value) and (2) the monitors must provide, individually, the highest information content (i.e., have a high entropy). The procedure is explained as follows.

[20] 1. Read and quantize the water level time series generated by the hydrodynamic model for each of the calculation points s_i (i = 1, 2, …, n, where n is the number of calculation points). Each point has an associated sequence of values X_i.

[21] 2. Calculate the marginal entropy $H(X_i)$ for each s_i with equation (1), from which the values to fulfill condition 1 will be taken.

[22] 3. For each s_i, calculate the mutual information in equation (2) with respect to each of the remaining points and

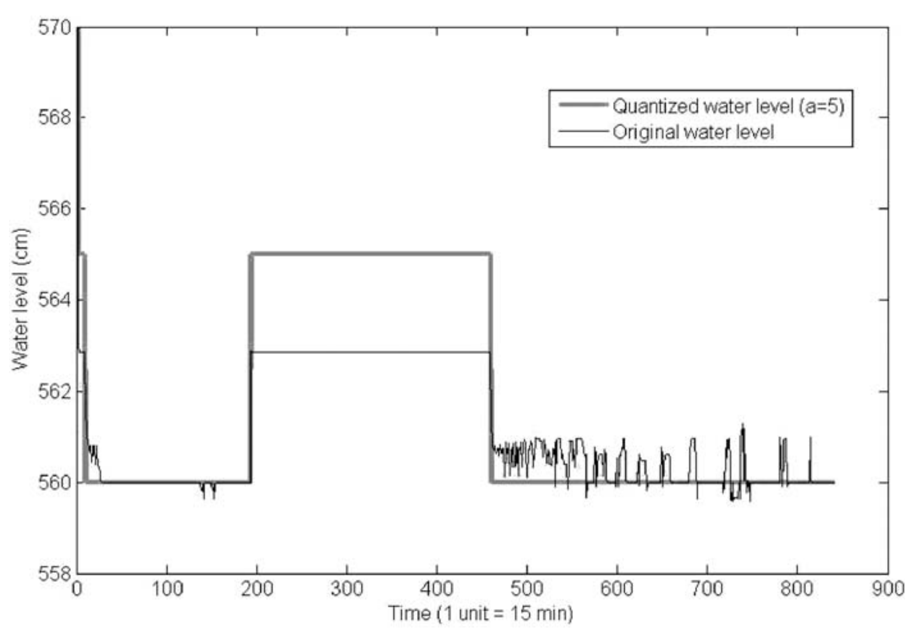

Figure 4. Example of original water level time series and its quantized version at a point located downstream from a pumping station.

Figure 5. Entropy map of the Pijnacker region.

build the symmetric matrix T, in which $I(X_i; X_i)$ is equivalent to $H(X_i)$ [*Cover and Thomas*, 1991],

$$T = \begin{bmatrix} I(X_1;X_1) & I(X_1;X_2) & \dots & I(X_1;X_n) \\ I(X_2;X_1) & I(X_2;X_2) & \dots & I(X_2;X_n) \\ \dots & \dots & \dots & \dots \\ I(X_n;X_1) & I(X_n;X_2) & \dots & I(X_n;X_n) \end{bmatrix}. \quad (8)$$

In this way, the point s_i will have an associated vector of mutual information v_i defined by the ith row (or column) of T. The values to fulfill condition 2 will be taken from this matrix.

[23] 4. The first monitor, m_1, is located at the point that provides the highest information content of the system (i.e., the point with the highest entropy value), so $m_1 = \max(H(X_i))$.

[24] 5. Add m_1 to the matrix of the monitoring points M.

[25] 6. Recover the mutual information vector v_1 of the monitor m_1:

$$v_1 = I(X_i; X_{m_1}), i = \{1, 2, \dots, n\}.$$

[26] 7. The system can then be divided into two sets of points with respect to their dependence on m_1: those that are dependent and those that are independent ($S_{m_1}^{\mathrm{ind}}$). The second monitor, m_2, must be selected from the former in order to fulfill condition 1. The set $S_{m_1}^{\mathrm{ind}}$ is obtained by looking at the elements of v_1 such that $I(X_i; X_{m_1}) < \varepsilon$, with ε being the value of transinformation between X_i and X_{m_1} that is insufficient for the pair to be considered dependent.

[27] 8. To fulfill condition 2, m_2 must have the highest entropy possible of the set $S_{m_1}^{\mathrm{ind}}$ so $m_2 = \max(H(X_i \in S_{m1}^{\mathrm{ind}}))$.

[28] 9. Recover the mutual information vector v_2 of the monitor m_2:

$$v_2 = I(X_i; X_{m_2}), i = \{1, 2, \dots, n\}.$$

[29] 10. The next monitor, m_3, must be selected in a similar way, but now using a modified set of independent points S_{m3}^{ind} given by the common set of independent points in the overlapping transinformation vectors for the previously selected monitors m_1 and m_2. Therefore, $v_3 = v_1 + v_2$.

[30] 11. Set $v_1 = v_3$ and repeat the procedure from step 6 until a maximum number of points is reached or until m_i does not provide a significant information content for the remaining system (i.e., its marginal entropy is too low).

[31] The matrix T in step 3 is replaced by DIT_{XY} and DIT_{YX} when evaluating the DIT-based criteria.

[32] Although the general scheme of the method is based on the studies mentioned in section 6, several changes are

Figure 6. Directional information transfer index (DIT) map for point *A* (bits).

proposed to make it applicable to highly controlled water systems. First, the time series of water level are generated by a hydrodynamic model at a very dense set of computational points, each of which is a potential gauging site. This avoids the use of empirical, historical measurements, which are not available at the required density (and at all significant points in a highly controlled polder system). Second, noisy time series produced by the operation of pumps are filtered by introducing the quantization concept. Third, the independent sets of time series are defined in a new way by using the common set of independent points in the overlapping transinformation vectors for the previously selected monitors. The solutions obtained by each of the pairwise criteria are then evaluated with the concept of total correlation to check the independence among monitors and the joint information provided by the set.

8. Case Study: Pijnacker Region, Netherlands

[33] The case study is located in a low-lying region of Pijnacker, Delfland, Netherlands, which has an area of 18.80 km^2, 15 pump stations, and 21 fixed weirs (Figure 1) that are operated in order to keep the water levels between limits defined by the water management authority of the region. The area is divided into 130 discharge hydrological response units, all contributing to the main storage basin or the big canals used to drain the excess water to the sea, a well-known characteristic of drainage in Netherlands. The region is mainly rural, with some urban developments (5.7 km^2) and glasshouses (2.82 km^2). From the water management point of view, the 130 hydrological units have 28 unique target water levels, which are maintained and controlled by the structures. Figure 2 shows the frequency of these target water levels, where negative values indicate that the region is below sea level.

[34] In order to apply the WMP approach described in section 5, the following actions were taken. The water level time series were generated by a hydrodynamic model built by the Delfland Water Board to make operational decisions regarding the control structures under several scenarios. A dense set of n = 1520 calculation points separated along the canals by a distance of 15 m on average were used with the rainfall event shown in Figure 3. The records were generated in a 10 day simulation with a time step of 15 min (this is the time step considered by the water board as useful for the management of the area and the pump operation). In addition, the parameter a in equation (7) used in the quantization of the time series was determined by looking at the water level variations that are negligible for management. In our case, a = 5 cm is the minimum dimensional unit for which the water management becomes critical. For example, water level variations of less than 5 cm can be due to wind, ship movement, or dynamic waves generated by the operation of pumping stations and are not important for water manage-

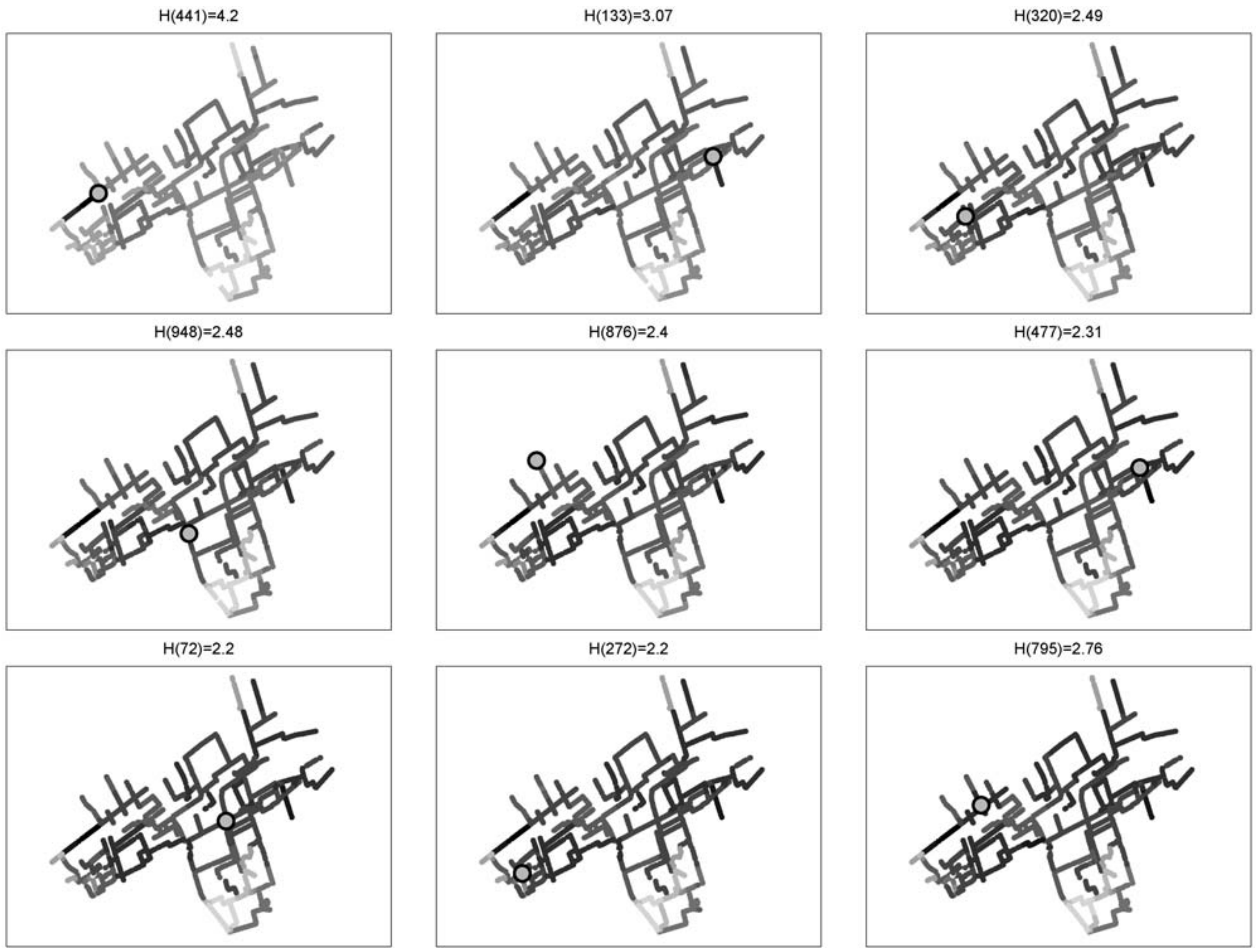

Figure 7. Step-by-step solution for the location of water level monitors using $I(X;Y)$ as the dependence criterion. The entropy of the currently selected point is shown at each step (bits).

ment (i.e., such variations should not require the hydraulic structures to be operated), and so they should be considered as noise in the time series (i.e., these time series should not provide information content, although they would have high entropy values without discretization). As an example, the water level time series at the discharge of one of the pump stations is presented in Figure 4, together with its quantized version used in the entropy calculation. Finally, the value of ε in step 7 was considered to be the mean of the vectors v_i, so that half of the points were considered independent and half dependent. This assumption is valid since only points with low pairwise dependence criterion with respect to a given point are selected.

9. Results

[35] In order to have an initial idea of the behavior of the information content of the Pijnacker region polder system, the maps of marginal entropy and mutual information were drawn. First, the entropy map of the system is shown in Figure 5, where several static information zones can be identified. Few points with null entropy can be found, which corresponds to fixed boundary conditions for water level (i.e., pump stations discharging to big storage bodies with fixed water level, to the southeast of the area). This map provides a first view of the areas (with high entropy) where it is suitable to place the first water level monitoring station, from the information content viewpoint. The subsequent monitors must be as independent as possible of this first monitor. Second, in order to highlight the information dependence in the system, Figure 6 shows the DIT index calculated between an arbitrary point (A) and the rest of the system. In Figure 6, the darker the point is, the greater its dependence on point A. The dependence of A on its neighboring points is evident. However, some other regions seem to have a strong dependence on A in spite of their hydraulic disconnection. It can also be seen that only a few points (the ones with constant water level during the time of simulation) are completely independent (DIT = 0).

[36] In order to look for a set of N gauging stations that can provide the maximum information of water levels in the system, the WMP approach is used, with the dependence criteria $I(X;Y)$, DIT_{XY}, and DIT_{YX} as the basis to create the matrix T in equation (8). In order to facilitate further analyses, n is initially considered to be 9. For the first criterion $I(X;Y)$, Figure 7 is the result, showing the overlapped map $I(X;Y)$ at each step and the value of the marginal entropy of the placed monitor along with the monitor locations. The darker the cal-

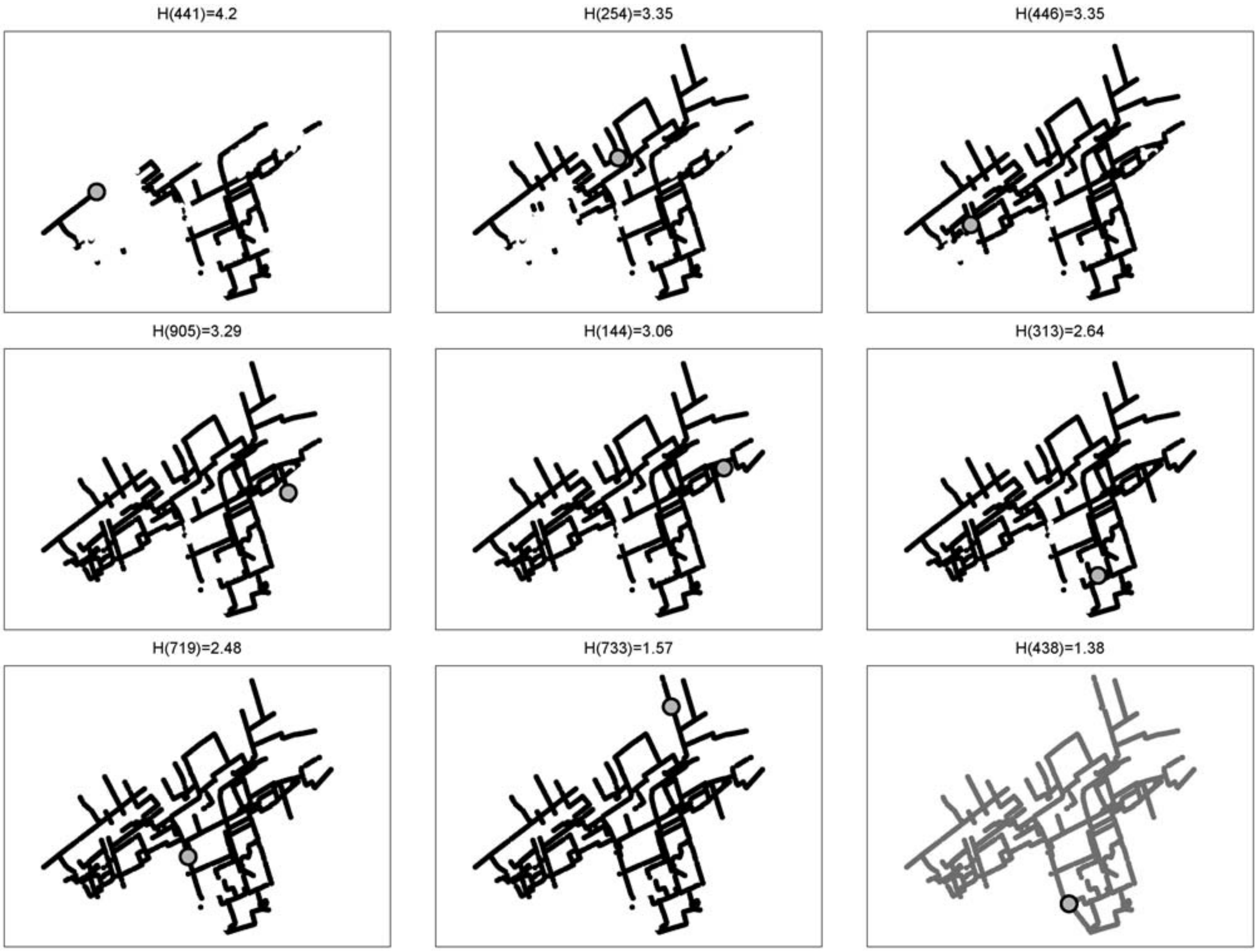

Figure 8. Step-by-step solution for the location of water level monitors using DIT_{XY}. The entropy of the currently selected point is shown at each step (bits).

culation point is, the higher its dependence with respect to the set of previously selected monitors. Similarly, Figure 8 results from applying the second criterion DIT_{XY} in the WMP approach. In this case, a value of 1 was assigned to the dependent points and 0 was assisnged to the independent ones, in such a way that the empty areas indicate where the next monitor must be placed. It can be seen that the ninth monitor does not have independent points associated with it, implying that this ninth monitor is not needed. This is also confirmed by its small value of entropy $H(X_9) = 1.38$ bits, which does not provide much more information to the joint set. Finally, WMP was preformed using the third dependence criterion, DIT_{YX}, yielding the nine monitors presented in Figure 9. A small area to the south with no coverage can be seen, which corresponds to points with very low entropy values (see Figure 5). This implies that any point in the system is completely independent of any point that belongs to this area (Figure 6, for example, confirms this statement for the case of point *A*). This situation explains why the ninth monitor in the previous experiment is not worth being selected.

[37] The location of the sets of monitors obtained by means of the three criteria is shown in detail in Figure 10. The summary of the monitors obtained with the different independence criteria is presented in Table 1, together with their corresponding values of total correlation and joint entropy, the former being calculated from equation (5).

10. Discussion

[38] The solutions obtained with the WMP method are discussed and evaluated in this section. First, the three pairwise criteria give some similar monitor locations in terms of spatial distribution. Besides the monitor at point 441 (selected by the three solutions, since the approach starts with the point with highest entropy), it is noticed that the monitors at points 733, 133, 426, and 438 have been selected by at least two of the solutions. In addition, there are points that are separately identified but are the same from the practical point of view, such as points 319 and 320, 426 and 876, 719 and 88, and 133 and 144.

[39] Note that for each criterion of dependence used, every time a monitor is added to the set, the number of independent points is reduced, and this reduction becomes less evident when new monitors are added. A quantitative way of looking at the reduction of uncertainty in the system when a new monitor is selected is by evaluating the value of joint entropy at every step. Since the calculation points are not completely independent, then $H(X_1, X_2, \ldots, X_N) \neq \sum_{i=1}^{N} H(X_i)$,

Figure 9. Step-by-step solution for the location of water level monitors using DIT_{YX}. The entropy of the currently selected point is shown at each step (bits).

and the concept of total correlation is needed to evaluate the multivariate independence by subtracting the summation of the marginal entropies from the value of total correlation estimated using the grouping property. Figure 11 shows that the three solutions have a similar behavior for both information content and independence. Furthermore, the reduction in uncertainty is strong for the first monitors because more information among them is shared as new monitors come into the solution. This explains why additional points do not reduce the uncertainty in the system. In order to allow further analysis, the following values are computed:

$$C(X_1, X_2, \ldots, X_{n=1520}) = 3510 \text{ bits},$$

$$\sum_{i=1}^{n=1520} H(X_i) = 3519.2 \text{ bits},$$

$$H(X_1, X_2, \ldots, X_{n=1520}) = \sum_{i=1}^{n=1520} H(X_i) - C(X_1, X_2, \ldots, X_{n=1520}) = 9.2 \text{ bits}.$$

It is clear that the total correlation of all points in the system almost equals the sum of their marginal entropies. This means that the amount of information shared among all of the points in the system is practically the same as the total amount of information that each point adds to the system separately. Moreover, the maximum amount of information content that can be extracted from the system is 9.2 bits.

[40] The implication of having C almost as big as the sum of the marginal entropies is that the calculation points are highly dependent on each other. This is like having a Venn diagram with 1520 circles of different size that overlap each other almost completely. If all the circles were independent of each other, then $C = 0$, and they would not overlap. The problem could be seen as selecting the best N circles that "cover" the total "area" of the n circles but that at the same time have little "overlapping area."

[41] The results are summarized in Table 1, in which the value of the total joint entropy and the total correlation are included for each solution. Figure 11 shows that with only the first two monitors it is possible to reach 5.5 bits of information content (60% of the total information content of the system) at a relatively low dependence value (less than 0.5% of the total correlation of the system). This provides a good criterion for evaluating the quality of the results.

[42] In general, the monitors selected are located next to hydraulic structures. This can be explained by the fact that these elements provide high, systematic variations to the water levels in the system during a precipitation event. Although entropy increases considerably by water level variations from pump stations, its informative capability is

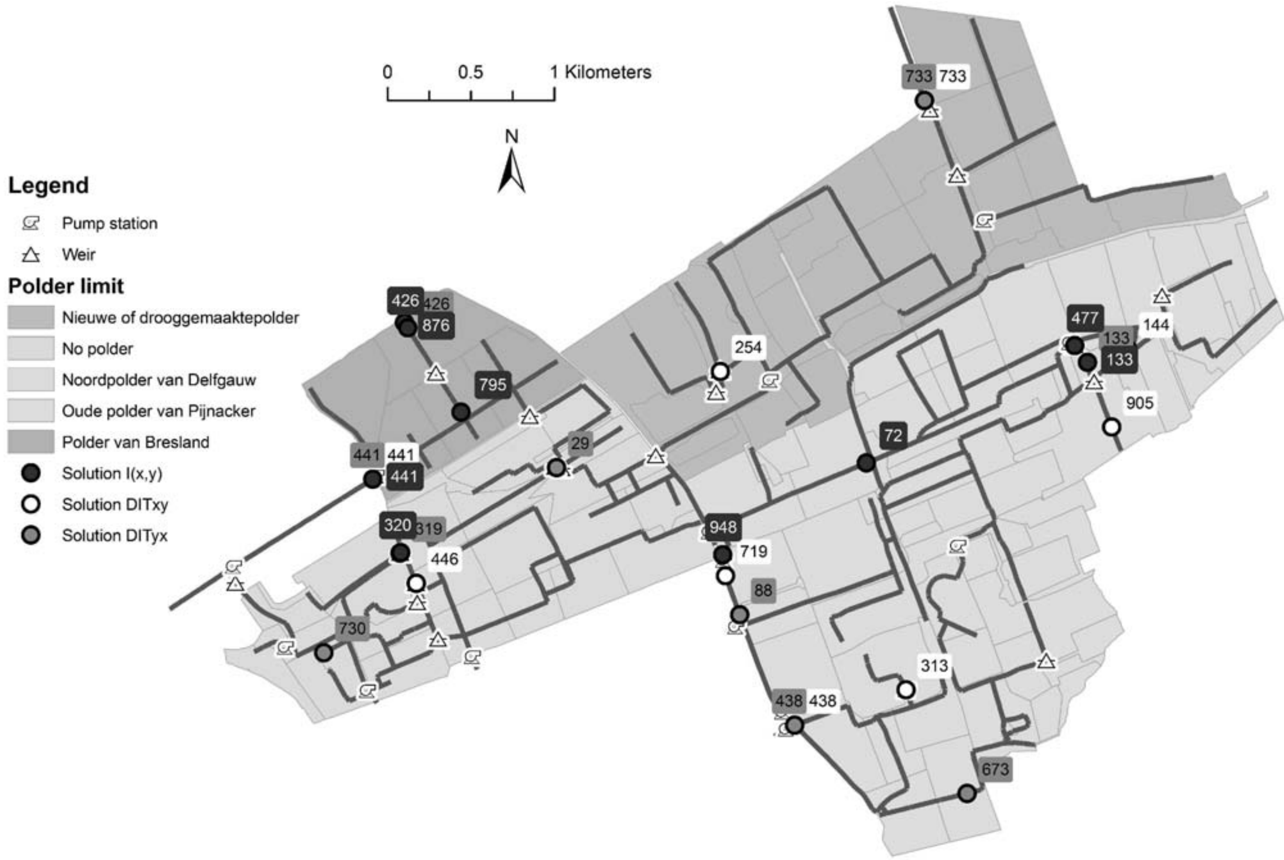

Figure 10. Location of water level monitors obtained by the water level monitoring design in polders approach using $I(X;Y)$, DIT_{XY}, and DIT_{YX} as pairwise dependence criteria.

not important if such variations occur within a small water level range. In this case, quantization appears to be a promising approach to filter out these noisy signals.

[43] The parameter a of equation (7) can be viewed as the minimum dimensional unit of water level for which the management of the system becomes critical. In the case of a typical low-lying polder system in the Netherlands, 5 cm is already decisive for water management in terms of the operation of control structures. Conversely, water level variations smaller than 5 cm (due, say, to wind, ship movement, or dynamic waves generated by the operation of pumping stations) are not important for water management, and so they should be considered as noise in the time series. In other water systems such as rivers, a = 5 cm might be too low.

[44] A sensitivity analysis of a is carried out by comparing monitor locations obtained for integer values of a between 1 and 10 cm and for 15 and 20 cm. The comparison shows that for each value of a between 3 and 8 cm a minimum of 55% of the same locations are shared with the average of the four neighboring solutions that are the closest in number. For values of $a < 3$ cm and $a > 10$ cm, only a few monitor locations are common. Figure 12 shows the average percentage of common monitors when comparing solutions for different values of a. In general, 25% of the locations are common to all solutions, although they are not necessarily selected in the same order. The first monitor location is common to all solutions (which is the one with the highest entropy value, as explained in step 4).

[45] From the hydraulics point of view, it is important to note that the monitors located downstream of a weir cannot give any information about the conditions upstream of the weir when this is working in a modular regime. On the other hand, when the weir regime is drowned, the downstream monitoring point may provide additional information to the system upstream. In general terms, all the weirs break the dependence in information between upstream and downstream points; in a similar way they break the continuity in water levels. However, some weirs working in the modular regime may not cause a discontinuity in information because

Table 1. Summary of Monitors Obtained by Each Dependence Criterion and Corresponding Values of Joint Entropy and Total Correlation

Criterion	Selected Monitor 1	2	3	4	5	6	7	8	9	Joint Entropy $H(X_1,X_2,\ldots,X_9)$	Total Correlation $H(X_1,X_2,\ldots,X_9)$
$I(X;Y)$	441	133	320	948	876	477	72	272	795	7.32	16.79
DIT_{XY}	441	254	446	905	144	313	719	733	438	7.63	17.7
DIT_{YX}	441	133	426	319	88	730	29	733	673	6.82	13.88

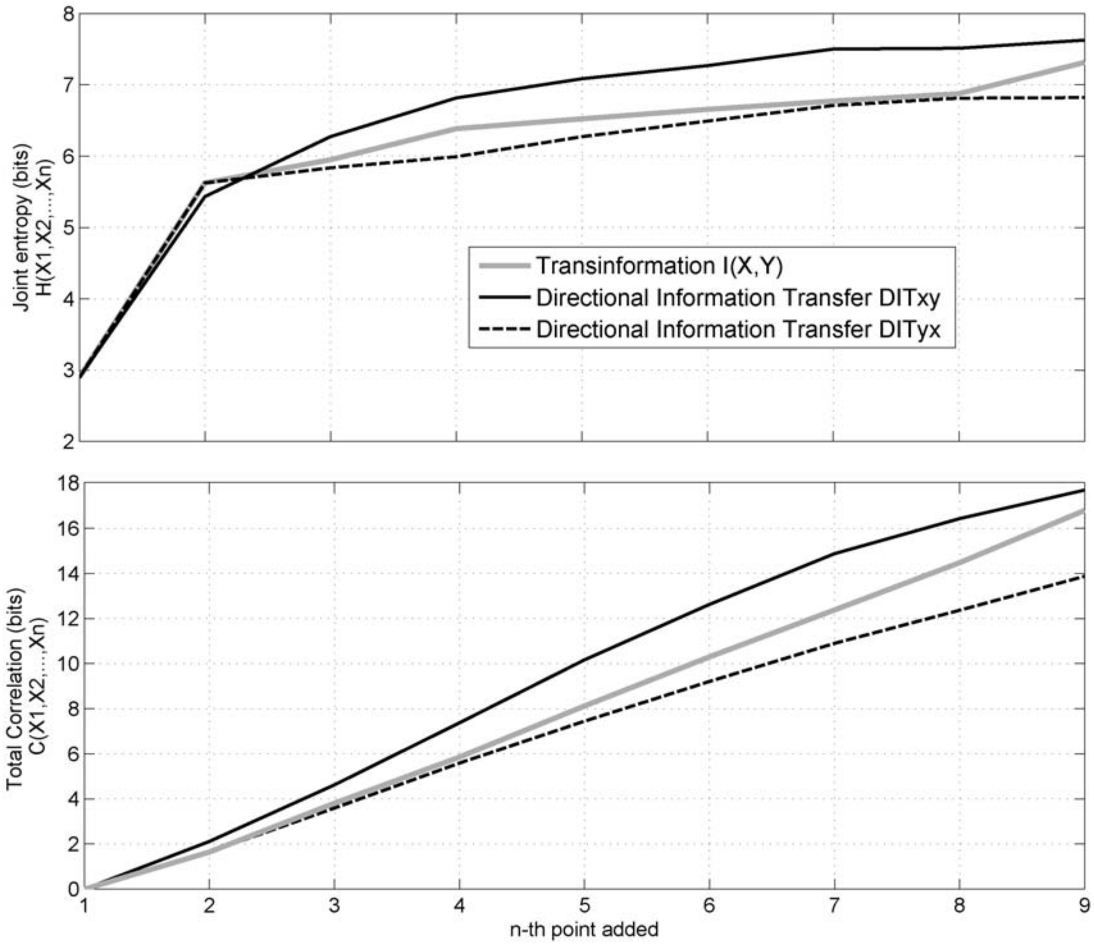

Figure 11. Evolution of the values of joint entropy and total correlation as new monitors are added to the solution set.

the same local hydrological information may be shared upstream and downstream. This situation can be seen in Figures 5 and 6, in which not all the weirs limit the information areas.

[46] The gauge locations will not change significantly when using different time steps in the hydrodynamic model as long as it acceptably replicates the real behavior of the system. Certainly, the quantization of the time series as well as the frequency-based method to estimate the probabilities for the information-related measures would give similar entropy values. In addition, the relative nature of the pairwise criteria makes entropy values unchanged using different time steps.

[47] Even though values of ε equal to 0 will define exactly the set of points that is dependent on a particular point, its applicability might lead to an empty set of independent values because even a small amount of information may be shared between hydraulically disconnected points. In this case the points can be hydrologically related; for example, a precipitation event that affects water levels at two hydraulically unrelated points may induce a correlation between them and thus share some of their information content, and $I(X_i;X_j) > 0$. This is why the independent points were selected considering the mean of the independence criterion as a threshold.

[48] The fact that the monitored signals are dependent is, in practice, a desirable condition since cross-checking is necessary to validate the data from other stations and to detect possible errors. Nevertheless, the theoretical design of monitoring networks should still consider the most independent monitors as priority places for data collection.

11. Summary and Conclusions

[49] The problem of placing water level monitors in a highly controlled polder system is not trivial because of the hydraulic discontinuity caused by hydraulic structures and because the large number of different surface levels where monitoring is required. For the sake of providing a solution an information theory–based method called water level monitoring design in polders (WMP) was presented. Several modifications to previous works that use information theory to evaluate rain gauge networks were introduced in order to make them applicable for the problem addressed, which includes the use of a hydrodynamic model for entropy analysis, the introduction of the quantization concept to filter out noisy time series, and the use of the total correlation to evaluate the performance of three different pairwise dependence criteria.

[50] A number of "information zones" can be identified in both entropy and transinformation maps defined by pump stations and by weirs working in modular regime, which create discontinuities in the information content upstream

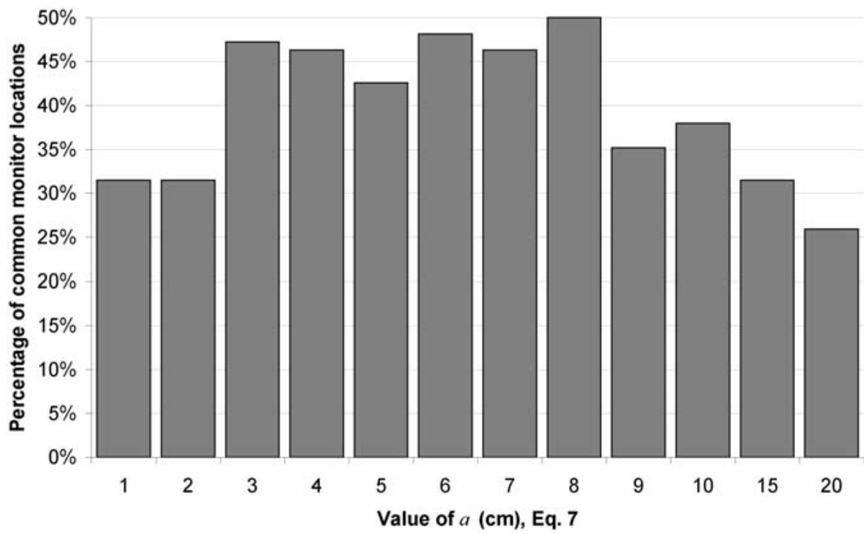

Figure 12. Average percentage of common monitor locations, comparing the solution obtained for each value of a with all other calculated solutions (a = 1, 2, …, 10, 15, 20).

and downstream of the structures. One information zone may include several target water levels.

[51] The solutions obtained with the three pairwise criteria I, DIT_{XY}, and DIT_{YX} have some monitors in common in terms of spatial distribution. However, the solution obtained with the DIT_{XY} criterion provides the highest value of joint entropy for the set of monitors and the highest value of total correlation. A high value of joint entropy reveals a high information content. This is the measure that identifies the preferred solution for the monitors.

[52] In contrast, the solution obtained with the DIT_{YX} criterion provides the lowest information content and the minimum value of total correlation. According to the conditions mentioned in the procedure for gauge location, this implies that DIT_{YX} is more effective at fulfilling condition 1 (independence), whereas DIT_{XY} is better at fulfilling condition 2 (amount of information content). This opens a new possibility for solving the problem of monitor selection out of a dense set of potential monitors, which is to use multiobjective optimization where DIT_{YX} is to be maximized and DIT_{YX} is to be minimized.

[53] Results show that the calculation points may be highly dependent on each other even if some of them are hydraulically disconnected. This dependence is due to the hydrological connection between the points since in relatively small areas the same rainfall events are shared. Although these hydrological dependences make the problem of looking for independent monitors more difficult, the proposed methodology proves to be a suitable method for this purpose.

[54] The values of marginal entropy are sensible for different values of a used in equation (7). Small values lead to high entropy values for locations near pumping stations, whereas larger values tend to filter out small disturbances, causing a decrease in entropy. However, as a affects all the generated time series equally, the selection of the first monitor location (see step 4 in section 7) remains the same for different values of a. In contrast, transinformation and DIT values do not change at all, due to the relative nature of the expressions. The value of a = 5 cm, however, is found to be the minimum dimensional unit for which the water management becomes critical and seems to be good for keeping the informational property, from the management point of view.

[55] Ongoing research is focused on using multiobjective optimization techniques to solve the problem by using the minimization of the total correlation as a first objective and the maximization of the joint entropy as a second objective. The information zones, in this sense, might be further used as a criterion for clustering in order to reduce the search space of such an optimization process.

Notation

a	quantization coefficient.
$C(X,Y)$	total correlation between X and Y, bits.
DIT	directional information transfer.
$I(X;Y)$	transinformation between X and Y, bits.
$H(X)$	entropy of X, bits.
$H(X,Y)$	joint entropy between X and Y.
m	selected monitor.
N	number of gauging stations.
n	number of calculation points in the hydrodynamic model.
(S_m^{ind})	set of points s that is independent of monitor m.
s	calculation point of the hydrodynamic model.
T	dependence matrix built with a given pairwise criterion.
v	dependence vector given by a row or column of matrix T.
x	outcome of X.
x_q	quantized value of x.
X,Y	random variables.

References

Amorocho, J., and B. Espildora (1973), Entropy in the assessment of uncertainty in hydrologic systems and models, *Water Resour. Res.*, *9*, 1511–1522, doi:10.1029/WR009i006p01511.

Bogárdi, I., A. Bárdossy, and L. Duckstein (1985), Multicriterion network design using geostatistics, *Water Resour. Res.*, *21*, 199–208, doi:10.1029/WR021i002p00199.

Bras, R. L., and I. Rodríguez-Iturbe (1976), Network design for the estimation of areal mean of rainfall events, *Water Resour. Res.*, *12*, 1185–1195, doi:10.1029/WR012i006p01185.

Caselton, W. F., and T. Husain (1980), Hydrologic networks: Information transmission, *J. Water Resour. Plann. Manage. Div.*, *106*, 503–520.

Caselton, W. F., and J. V. Zidek (1984), Optimal monitoring network designs, *Stat. Probab. Lett.*, *2*, 223–227, doi:10.1016/0167-7152(84)90020-8.

Chow, C., and C. Liu (1968), Approximating discrete probability distributions with dependence trees, *IEEE Trans. Inf. Theory*, *14*, 462–467, doi:10.1109/TIT.1968.1054142.

Cover, T. M., and J. A. Thomas (1991), *Information Theory*, John Wiley, New York, doi:10.1002/0471200611.

Fass, D. M. (2006), Human sensitivity to mutual information, Ph.D. dissertation, Rutgers–State Univ. of N. J., New Brunswick.

Filippini, F., et al. (1994), Optimization criteria for configuring a network of monitoring stations, *Environ. Software*, *9*, 77–88, doi:10.1016/0266-9838(94)90001-9.

Harmancioglu, N., and V. Yevjevich (1987), Transfer of hydrologic information among river points, *J. Hydrol.*, *91*, 103–118, doi:10.1016/0022-1694(87)90131-4.

Harmancioglu, N., B. O. Fistikoglu, S. D. Ozkul, V. P. Singh, and M. N. Alpaslan (1999), *Water Quality Monitoring Network Design*, Kluwer Acad., Dordrecht, Netherlands.

Hart, P. E. (1971), Entropy and other measures of concentration, *J. R. Stat. Soc., Ser. A*, *134*, 73–85, doi:10.2307/2343975.

Husain, T. (1989), Hydrologic uncertainty measure and network design, *Water Resour. Bull.*, *25*, 527–534.

Jakulin, A., and I. Bratko (2004), Testing the significance of attribute interactions, in *Proceedings of the Twenty-First International Conference on Machine Learning*, *ACM Int. Conf. Proc. Ser.*, vol. 69, p. 52, Assoc. for Comput. Mach., Banff, Alberta, Canada.

Karasev, I. F. (1968), Principles for distribution and prospects for development of a hydrologic network, *Sov. Hydrol.*, *6*, 560–588.

Kirshner, S., P. Smyth, and A. Robertson (2004), Conditional Chow-Liu tree structures for modeling discrete-valued vector time series, in *Uncertainty in Artificial Intelligence: Proceedings of the Twentieth Conference*, pp. 317–324, AUAI Press, Arlington, Va.

Kraskov, A., H. Stogbauer, R. G. Andrzejak, and P. Grassberger (2005), Hierarchical clustering using mutual information, *Europhys. Lett.*, *70*, 278–284, doi:10.1209/epl/i2004-10483-y.

Krstanovic, P. F., and V. P. Singh (1992a), Evaluation of rainfall networks using entropy: I. Theoretical development, *Water Resour. Manage.*, *6*, 279–293, doi:10.1007/BF00872281.

Krstanovic, P. F., and V. P. Singh (1992b), Evaluation of rainfall networks using entropy: II. Applications, *Water Resour. Manage.*, *6*, 295–314, doi:10.1007/BF00872282.

Lewis, P. M. (1959), Approximating probability distributions to reduce storage requirements, *Inf. Control*, *2*, 214–225, doi:10.1016/S0019-9958(59)90207-4.

Li, W. (1990), Mutual information functions versus correlation functions, *J. Stat. Phys.*, *60*, 823–837, doi:10.1007/BF01025996.

Linfoot, E. H. (1957), An informational measure of correlation, *Inf. Control*, *1*, 85–89, doi:10.1016/S0019-9958(57)90116-X.

Loucks, D. P., and E. van Beek (2005), *Water Resources Systems Planning and Management*: *An Introduction to Methods, Models and Applications*, U. N. Educ., Sci. and Cult. Organ., Paris.

Markus, M., H. V. Knapp, and G. D Tasker (2003), Entropy and generalized least square methods in assessment of the regional value of streamgages, *J. Hydrol.*, *283*, 107–121, doi:10.1016/S0022-1694(03)00244-0.

McGill, W. J. (1954), Multivariate information transmission, *Psychometrika*, *19*, 97–116, doi:10.1007/BF02289159.

Mishra, A. K., and P. Coulibaly (2009), Developments in hydrometric network design: A review, *Rev. Geophys.*, *47*, RG2001, doi:10.1029/2007RG000243.

Mogheir, Y., and V. P. Singh (2002), Application of information theory to groundwater quality monitoring networks, *Water Resour. Manage.*, *16*, 37–49, doi:10.1023/A:1015511811686.

Mogheir, Y., J. L. M. P. de Lima, and V. P. Singh (2004), Characterizing the spatial variability of groundwater quality using the entropy theory: I. Synthetic data, *Hydrol. Processes*, *18*, 2165–2179, doi:10.1002/hyp.1465.

Mogheir, Y., V. P. Singh, and J. L. M. P. de Lima (2006), Spatial assessment and redesign of a groundwater quality monitoring network using entropy theory, Gaza Strip, Palestine, *Hydrogeol. J.*, *14*, 700–712, doi:10.1007/s10040-005-0464-3.

Moon, Y.-I., B. Rajagopalan, and U. Lall (1995), Estimation of mutual information using kernel density estimators, *Phys. Rev. E*, *52*, 2318–2321, doi:10.1103/PhysRevE.52.2318.

Moore, R. J., D. A. Jones, D. R. Cox, and V. S. Isham (2000), Design of the HYREX raingauge network, *Hydrol. Earth Syst. Sci.*, *4*, 521–530.

Moss, M. E. (1976), Design of surface water data networks for regional information, *Hydrol. Sci. Bull.*, *21*, 113–127.

Moss, M. E., and M. R. Karlinger (1974), Surface water network design by regression analysis simulation, *Water Resour. Res.*, *10*, 427–433, doi:10.1029/WR010i003p00427.

Moss, M. E., and G. D. Tasker (1991), Intercomparison of hydrological network-design technologies, *Hydrol. Sci. J.*, *36*, 209–221.

Pardo-Igúzquiza, E. (1998), Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing, *J. Hydrol.*, *210*, 206–220, doi:10.1016/S0022-1694(98)00188-7.

Philippatos, G. C., and C. J. Wilson (1972), Entropy, market risk, and the selection of efficient portfolios, *Appl. Econ.*, *4*, 209–220, doi:10.1080/00036847200000017.

Rodríguez-Iturbe, I., and J. M. Mejia (1974), The design of rainfall networks in space and time, *Water Resour. Res.*, *10*, 713–728, doi:10.1029/WR010i004p00713.

Ruddell, B. L., and P. Kumar (2009), Ecohydrologic process networks: 1. Identification, *Water Resour. Res.*, *45*, W03419, doi:10.1029/2008WR007279.

Sansó, B., and P. Müller (1997), *Redesigning a Network of Rainfall Stations*, Inst. of Stat. and Decision Sci., Duke Univ., Durham, N. C.

Schreiber, T. (2000), Measuring information transfer, *Phys. Rev. Lett.*, *85*, 461–464, doi:10.1103/PhysRevLett.85.461.

Shannon, C. E. (1948), A mathematical theory of communication, *Bell Syst. Tech. J.*, *27*, 379–423.

Sharma, A. (2000), Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3—A nonparametric probabilistic forecast model, *J. Hydrol.*, *239*, 249–258, doi:10.1016/S0022-1694(00)00348-6.

Singh, V. P. (1997), The use of entropy in hydrology and water resources, *Hydrol. Processes*, *11*, 587–626, doi:10.1002/(SICI)1099-1085(199705)11:6<587::AID-HYP479>3.0.CO;2-P.

Singh, V. P. (2000), The entropy theory as a tool for modelling and decision-making in environmental and water resources, *Water SA*, *26*, 1–12.

Steuer, R., et al. (2002), The mutual information: Detecting and evaluating dependencies between variables, *Bioinformatics*, *18*, S231–S240.

Watanabe, S. (1960), Information theoretical analysis of multivariate correlation, *IBM J. Res. Dev.*, *4*, 66–82.

Wei, Y. (1987), Variance, entropy and uncertainty measure, in *Proceedings of the Survey Research Methods Section*, pp. 609–610, Am. Stat. Assoc., Alexandria, Va.

Yang, Y., and D. H. Burn (1994), An entropy approach to data collection network design, *J. Hydrol.*, *157*, 307–324, doi:10.1016/0022-1694(94)90111-2.

Yeh, M.-S., Y. P. Lin, and L.-C. Chang (2006), Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms, *Environ. Geol.*, *50*, 101–121, doi:10.1007/s00254-006-0190-8.

Zidek, J. V., W. Sun, and N. D. Le (2000), Designing and integrating composite networks for monitoring multivariate Gaussian pollution fields, *Appl. Stat.*, *49*, 63–79, doi:10.1111/1467-9876.00179.

L. Alfonso, A. Lobbrecht, and R. Price, Hydroinformatics and Knowledge Management, UNESCO-IHE, Westvest 7, NL-2611 AX Delft, Netherlands. (l.alfonsosegura@unesco-ihe.org; a.lobbrecht@unesco-ihe.org; r.price@unesco-ihe.org)