

# Entropy theory-based criterion for hydrometric network evaluation and design: Maximum information minimum redundancy

Chao Li,<sup>1</sup> Vijay P. Singh,<sup>1,2</sup> and Ashok K. Mishra<sup>1</sup>

Received 6 August 2011; revised 14 March 2012; accepted 23 March 2012; published 15 May 2012.

[1] Hydrometric information constitutes the fundamental input for planning, design, operation, and management of water resources systems. How to optimally site monitoring gauges such that they are effective and efficient in gathering the hydrometric information or data has received considerable attention. This paper presents a generic approach for the design (or evaluation) of hydrometric networks. First, an entropy theory-based criterion, named as maximum information minimum redundancy (MIMR), is proposed. The MIMR criterion maximizes the joint entropy of stations within the optimal set, and the transinformation between stations within and outside of the optimal set. Meanwhile, it insures that the optimal set contains minimum duplicated information. An easy-to-implement greedy ranking algorithm is developed to accomplish the MIMR selection. Two case studies are presented to illustrate the applicability of MIMR in hydrometric network evaluation and design. We also compare the MIMR selection with another entropy-based approach. Results illustrate that MIMR is apt at finding stations with high information content, and locating independent stations. The proposed approach is suitable for design (or evaluation) of any type of hydrometric network.

**Citation:** Li, C., V. P. Singh, and A. K. Mishra (2012), Entropy theory-based criterion for hydrometric network evaluation and design: Maximum information minimum redundancy, *Water Resour. Res.*, 48, W05521, doi:10.1029/2011WR011251.

## 1. Introduction

[2] Hydrometric information, which is mainly collected by monitoring networks, constitutes the fundamental input for planning, design, operation, and management of water resources systems. Optimally siting of monitoring gauges such that they are effective and efficient in gathering data (information) has received considerable attention. Although there are myriad concerns in hydrometric network design, this study focuses on the fundamental theme, i.e., selecting an optimum number of stations and their optimum locations. Many approaches have been developed for that purpose. A comprehensive review can be found in the work of *Mishra and Coulibaly* [2009]. Among others, one type of approach is based on entropy theory. The merit of entropy theory is that it directly defines information and quantifies uncertainty [*Harmancioglu and Singh*, 1998; *Mogheir et al.*, 2006]. A significant body of literature employing entropy theory for the design of monitoring networks has been reported, as briefly reviewed in what follows.

[3] *Caselton and Husain* [1980] and *Husain* [1987] used an information maximization principle for identifying optimum locations of rainfall gauges to be retained in a dense network. When an existing network is sparse, *Husain* [1989]

proposed a methodology for expanding it by means of information interpolation. *Krstanovic and Singh* [1992a] developed an entropy-based approach for hydrologic network evaluation. This approach was then used to spatiotemporally evaluate the rainfall network in Louisiana [*Krstanovic and Singh*, 1992b]. *Yang and Burn* [1994] presented a method for data collection design in which a concept of directional information flow was employed. Also in terms of entropy theory, *Mogheir et al.* [2006] evaluated the optimality of the groundwater quality monitoring network in Gaza Strip, Palestine. *Mishra and Coulibaly* [2010] assessed streamflow network in different Canadian watersheds using entropy, joint entropy and transinformation.

[4] Common with the aforementioned studies is that entropy terms were computed using univariate or bivariate formulations. However, joint information retained by multiple stations and their dependence are always required for a more objective evaluation. A few studies did use multivariate distributions but assumed that the data were normally or lognormally distributed as, for example, did *Husain* [1987, 1989], and *Krstanovic and Singh* [1992a, 1992b]. This assumption is debatable, since many natural phenomena are heavy tail distributed, like streamflow and precipitation [*Bernadara et al.*, 2008; *Carreau et al.*, 2009; *Li et al.*, 2012].

[5] Similar to *Krstanovic and Singh* [1992a, 1992b], *Alfonso et al.* [2010a] introduced several adaptations to make the entropy-based method applicable to the design of water level monitors for highly controlled polder system in the Netherlands. They first used multivariate total correlation to assess the performance of three pairwise dependence

<sup>1</sup>Department of Biological and Agricultural Engineering, Texas A&M University, College Station, Texas, USA.

<sup>2</sup>Department of Civil and Environmental Engineering, Texas A&M University, College Station, Texas, USA.

criteria. The selection of optimal monitors was restricted to low dimensional analysis (less than 2).

[6] Later, *Alfonso et al.* [2010b] proposed another criterion by maximizing multivariate joint entropy and minimizing total correlation for optimally siting water level monitors. Yet, this approach failed to account for the information transition ability (transinformation) of a network. It is acknowledged that transferring hydrologic information from points where it is available to those where it is required is one of the purposes of collecting hydrometric information [*Harmancioglu and Yevjevich*, 1987].

[7] Moreover, optimally siting hydrometric monitors is a multiobjective problem. Solving this problem is tricky in practice. *Alfonso et al.* [2010b] exploited a genetic algorithm to approach the multiobjective optimization. The advantage of multiobjective optimization is that it provides different feasible solutions under different scenarios. Nevertheless, selection of the final network is not straightforward. To assist the decision making processes, it is quite admirable to find an easy-to-implement way to solve the multiobjective problem and to provide the end user a unique solution with decent performance.

[8] Our objective therefore is to develop an easy-to-implement approach for the design (or evaluation) of hydrometric networks. To that end, we first propose an entropy theory-based criterion, named as maximum information minimum redundancy (MIMR), satisfying three norms: (1) maximum overall information (joint entropy), (2) maximum information transition ability (transinformation), and (3) minimum redundant information (total correlation). These entropy terms are calculated at multivariate level without any distributional assumption. Thereby interactions among stations can be properly accounted for. Then we present a straightforward greedy selection algorithm to rank the candidate gauges based on MIMR. During selection, the three commensurable norms are additively unified, which circumvents the complexity of multiobjective optimization, while preserving its advantage in achieving different feasible solutions through information-redundancy tradeoff weights.

[9] The paper is organized as follows. Formulating the objectives of the study in this section, basic entropy theory is briefly presented in section 2 for the ease of understanding the MIMR criterion, which, together with a selection algorithm, is discussed in section 3. Through two case studies, section 4 illustrates the applicability of MIMR to the evaluation and design of hydrometric networks. Merits and demerits of MIMR are discussed in section 5. Conclusions are generalized in section 6.

## 2. Entropy Theory

### 2.1. Basic Entropy Terms

[10] Marginal entropy, joint entropy, transinformation, and total correlation constitute basic information measures commonly used in hydrometric network evaluation and design. Let  $[X_1, X_2, \dots, X_n]$  denote a discrete random vector with joint probability mass function  $p(x_1, x_2, \dots, x_n)$  and one-dimensional margins, respectively,  $p_{X_1}(x_1), p_{X_2}(x_2), \dots, p_{X_n}(x_n)$  for  $X_1, X_2, \dots$ , and  $X_n$ . In hydrometric network evaluation and design, one usually wants to know: (1) How much information is retained by a random variable (station)? (2) What is the information conveyed by several var-

iables (stations) together? (3) How much information of a random variable (station) can be inferred from the knowledge of another one? (4) What is the duplicated information among several variables (stations)?

[11] The first question can be answered by marginal entropy. The marginal entropy of a discrete random variable measuring the information retained by it is defined as

$$H(X_1) = -\sum_{x_1} p(x_1) \log_2(p(x_1)), \quad (1)$$

where, for simplicity, subscript of  $p(\cdot)$  is suppressed. The sum notation means summation over all possible outcomes of  $X_1$ . All sum notations in what follows hold the same meaning, if no lower or upper limit is specified.

[12] To answer the second question, joint entropy is defined as a measure of the overall information retained by random variables. The bivariate joint entropy of  $X_1$  and  $X_2$  is defined as

$$H(X_1, X_2) = -\sum_{x_1} \sum_{x_2} p(x_1, x_2) \log_2(p(x_1, x_2)). \quad (2)$$

There is a natural extension for multivariate joint entropy from bivariate if one wants to know the information retained by more than two random variables. The definition of multivariate joint entropy is

$$H(X_1, X_2, \dots, X_n) = -\sum_{x_1} \sum_{x_2} \dots \sum_{x_n} p(x_1, x_2, \dots, x_n) \log_2(p(x_1, x_2, \dots, x_n)). \quad (3)$$

The joint entropy is symmetric with respect to its arguments. Quantitatively, joint entropy is less than or equal to the sum of its one-dimensional marginal entropies. The equality holds if and only if the random variables are stochastically independent.

[13] Concerning the third question, transinformation (also referred to as mutual information) provides the answer. Transinformation of  $X_1$  and  $X_2$ , measuring the information of  $X_1$  (or  $X_2$ ) inferred from that of  $X_2$  (or  $X_1$ ), is computed as

$$T(X_1; X_2) = \sum_{x_1} \sum_{x_2} p(x_1, x_2) \log_2 \frac{p(x_1, x_2)}{p(x_1)p(x_2)}. \quad (4)$$

Transinformation provides a general measure of dependence between random variables. It is superior to the Pearson correlation coefficient, since it captures both linear and nonlinear dependence, whereas the Pearson correlation coefficient is only suitable for linear relationships, or more generally, for spherical and elliptical dependence structures.

[14] To assess the redundancy of a hydrometric network, one is usually interested in the amount of duplicated information among a set of stations, which can be measured by total correlation [*Watanabe*, 1960]. Mathematically, the total correlation is defined as

$$C(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i) - H(X_1, X_2, \dots, X_n). \quad (5)$$

Total correlation is also symmetric, facilitating its computation through grouping property [*Kraskov et al.*, 2005;

Alfonso et al., 2010a, 2010b]. Total correlation  $C$  is non-negative. It assumes 0 if and only if all random variables are independent. Otherwise,  $C$  will be greater than 0. Total correlation answers the fourth question.

## 2.2. Continuous Time Series Discretization

[15] The entropy terms can be computed by discretizing hydrometric time series data collected at stations. Several methods are available for continuous data discretization. Among others, the histogram method and mathematical floor function are frequently used [Caselton and Husain, 1980; Husain, 1987; Mishra and Coulibaly, 2010; Alfonso et al., 2010a, 2010b].

[16] When applying the histogram discretization, it is common to assume a somewhat arbitrary number of bins, like 10, 15, and 20. Yet, this strategy is questionable, since discrete entropy terms are sensitive to the bin size. Optimal bin size estimators can be used, like those in the work of Scott [1979], Freedman and Diaconis [1981], Birgé and Rozenholc [2006], Shimazaki and Shinomoto [2007]. These estimators are derived based on different criteria, nevertheless, none of them are completely convincing in that one has been shown to be better than others. Given a data sample, different estimators may lead to different bin size estimates, which in turn will lead to different entropy values.

[17] The subjective determination of the bin size is addressed to some extent, with the use of a mathematical floor function [Ruddell and Kumar, 2009; Alfonso et al., 2010a, 2010b]. Through a mathematical floor function, a continuous value  $x$  is converted to its nearest lowest integer multiple of a constant  $a$ , i.e.,

$$x_q = a \left\lfloor \frac{2x + a}{2a} \right\rfloor, \quad (6)$$

where  $\lfloor \cdot \rfloor$  represents the conventional mathematical floor function. The advantages of mathematical floor function include: (1) it avoids the choice of a parametric distribution to fit the continuous data; (2) it can incorporate physical considerations in that the resolution of  $a$  should be no less than uncertainties involved in the continuous data, as will be illustrated in section 4.2. However, how to physically determine an appropriate  $a$  is not always explicit in all situations. Then parameter  $a$  might be empirically selected through trial and error, as will be illustrated in section 4.1. Generally, it should be neither too large nor too small. Rules of thumb can be used to guide the selection of parameter  $a$ : (1) it should guarantee that all candidate stations have significant and distinguishable information contents; (2) the spatial and temporal variability of time series of stations should be preserved as much as possible before and after discretization; and (3) the selected stations should be stable as much as possible, when  $a$  fluctuates within some interval centered near its optimal value.

[18] One point worth noting is that after applying the mathematical floor function the marginal entropy is no longer a measure of information of the continuous random variable  $X$ , but the information of  $X$  rounded to its nearest lowest integer multiple of a constant [Papoulis and Pilli, 2001]. Analogous meanings hold for other entropy terms. In the context of hydrometric network design, we do not need to precisely quantify the information retained by stations. A reasonable approximation is sufficient as long as

the relative relationship among stations can be preserved in terms of information content. Indeed, treating a continuous time series as a discrete pulse signal is reasonable, considering the fact that in practice any records are subject to noise, round off errors and errors caused by measurement instruments. It is necessary to filter out such high-frequency low-amplitude fluctuations.

## 2.3. Discrete Variable Merging

[19] Multivariate joint entropy and total correlation can be computed at any dimension with the aid of discrete variable merging. The basic idea for variable merging lies in creating a new variable  $X$  such that the information retained by which is equal to that of the original variables, say  $X_1, X_2, \dots, X_n$ . Consider merging two discrete variables as an example. If it is assumed that  $X_1 = [1, 2, 1, 2, 1, 3, 3]^T$  and  $X_2 = [1, 2, 2, 2, 1, 3, 2]^T$ , then the new variable  $X$  can be obtained by pairwise welding the corresponding digits together [Alfonso et al., 2010b], i.e.,  $X = [11, 22, 12, 22, 11, 33, 32]^T$ . It can be verified that the information amount keeps invariant before and after merging.

[20] The direct welding approach, however, has a deficiency which may cause the so called problem of “out of memory” as the number of variables to be merged increases, especially when the sample size is large. An adapted alternative to the direct welding approach, which avoids the above problem, is generalized in what follows:

[21] Algorithm 1

[22] 1. Create a new sample  $X$  from  $X_1$  and  $X_2$  by the direct welding approach.

[23] 2. Pick out the unique values in  $X$  and rank them in ascending order, resulting in a ranked sample  $X_r$  with a length of  $l$ .

[24] 3. Access the location index of each element of  $X$  in the ranked sample  $X_r$ .

[25] 4. Assign each element of  $X$  a new label as its location index obtained in step 3.

[26] After this, each element in the new merged sample is relabeled by an integer ranging from 1 to  $l$ . Remember that the merging approach is only suitable for discrete variables. The variable merging approach satisfies the law of association and commutation in terms of information content. Taking merging 3 variables as an example, according to the law of association and commutation, the following equalities are satisfied:

$$\begin{aligned} H(\langle X_1, X_2, X_3 \rangle) &= H(\langle\langle X_1, X_2 \rangle, X_3 \rangle) \\ &= H(\langle X_1, \langle X_2, X_3 \rangle \rangle) \\ &= H(\langle\langle X_1, X_3 \rangle, X_2 \rangle), \end{aligned} \quad (7)$$

where  $\langle \cdot \rangle$  denotes the merging operator. Assuming  $X_3 = [1, 1, 2, 2, 1, 3, 3]^T$  and applying algorithm 1 yields

$$\begin{aligned} H(\langle\langle X_1, X_2 \rangle, X_3 \rangle) &= H([1, 4, 2, 3, 1, 6, 5]^T) \\ &= 2.522 \text{ Bits,} \\ H(\langle X_1, \langle X_2, X_3 \rangle \rangle) &= H([1, 2, 4, 3, 1, 6, 5]^T) \\ &= 2.522 \text{ Bits,} \\ H(\langle\langle X_1, X_3 \rangle, X_2 \rangle) &= H([1, 3, 2, 4, 1, 5, 6]^T) \\ &= 2.522 \text{ Bits.} \end{aligned}$$

[27] In terms of variable merging, the multivariate joint entropy,  $H(X_1, X_2, \dots, X_n)$ , can be computed by sequentially applying algorithm 1, i.e.,

$$\begin{aligned} H(X_1, X_2, \dots, X_n) &= H(\langle X_1, X_2 \rangle, X_3, \dots, X_n) \\ &= H(\langle\langle X_1, X_2 \rangle, X_3 \rangle, \dots, X_n) \\ &= H(\langle \dots \langle\langle X_1, X_2 \rangle, X_3 \rangle, \dots, X_{n-1} \rangle, X_n) \\ &\dots \dots \\ &= H(\langle \dots \langle \dots \langle\langle X_1, X_2 \rangle, X_3 \rangle, \dots, X_{n-1} \rangle, X_n \rangle). \end{aligned} \quad (8)$$

[28] Concerning the computation of total correlation, one can proceed in two different ways. One is to use the shortcut formula in equation (5). The other one is to apply the grouping property [Kraskov *et al.*, 2005; Alfonso *et al.*, 2010a, 2010b] together with algorithm 1, as shown in the following:

$$\begin{aligned} C(X_1, X_2, \dots, X_n) &= C(X_1, X_2) + C(\langle X_1, X_2 \rangle, X_3, X_4, \dots, X_n) \\ &= C(X_1, X_2) + C(\langle X_1, X_2 \rangle, X_3) \\ &\quad + C(\langle X_1, X_2, X_3 \rangle, X_4, X_5, \dots, X_n) \\ &\dots \dots \\ &= C(X_1, X_2) + C(\langle X_1, X_2 \rangle, X_3) + C(\langle X_1, X_2, X_3 \rangle, X_4) \\ &\quad + \dots + C(\langle X_1, X_2, \dots, X_{n-1} \rangle, X_n). \end{aligned} \quad (9)$$

Since the total correlation at the bivariate level reduces to transinformation, equation (9) indicates that the  $n$ -dimensional total correlation is factorized as a summation of traditional transinformation values.

[29] With the aid of variable merging, we can define the multivariate transinformation between single and grouped variables, and between two grouped variables. The first type multivariate transinformation,  $T(\langle X_1, X_2, \dots, X_n \rangle; X)$ , measures the information amount of a single variable which can be inferred from that of variables in the group, and vice versa. Transinformation between grouped variables,  $T(\langle X_1, X_2, \dots, X_n \rangle; \langle X'_1, X'_2, \dots, X'_m \rangle)$ , measures the common information shared by the two grouped variables. The above two types of multivariate transinformation can be easily computed by first merging variables in the group and then applying the definition of bivariate transinformation in equation (4).

### 3. MIMR Criterion and Its Implementation

#### 3.1. MIMR Criterion

[30] The idea of MIMR criterion lies in selecting (or ranking) stations from a candidate set, through which, the selected stations can: (1) maximize the overall information, (2) maximize the information transition ability, and (3) minimize the redundant information.

[31] Assume there are  $N$  candidate stations. For each station, there are several years of records of the variable of interest denoted by  $X$ , such as streamflow. Let  $S$  be the set of stations already selected into the optimal network whose elements are denoted as  $X_{S_1}, X_{S_2}, \dots, X_{S_k}$ . Similarly, let  $F$  be the set of candidate stations to be selected whose

elements are denoted as  $X_{F_1}, X_{F_2}, \dots, X_{F_m}$ . Apparently, the sum of  $k$  and  $m$  is equal to  $N$ . The amount of overall information retained by  $S$  can be quantified by multivariate joint entropy

$$H(X_{S_1}, X_{S_2}, \dots, X_{S_k}). \quad (10)$$

The information transition ability of  $S$  can be measured by the sum of transinformation between grouped variables in  $S$  and each station in  $F$ , denoted as  $T_1$  for simplicity,

$$T_1 = \sum_{i=1}^m T(\langle X_{S_1}, X_{S_2}, \dots, X_{S_k} \rangle; X_{F_i}). \quad (11)$$

Also, it can be measured by the transinformation between grouped variables in  $S$  and  $F$ , denoted as  $T_2$ ,

$$T_2 = T(\langle X_{S_1}, X_{S_2}, \dots, X_{S_k} \rangle; \langle X_{F_1}, X_{F_2}, \dots, X_{F_m} \rangle). \quad (12)$$

The redundant information of the optimal set  $S$  can be quantified by the total correlation

$$C(X_{S_1}, X_{S_2}, \dots, X_{S_k}). \quad (13)$$

Then, the MIMR criterion-based objective functions are formulated as

$$\begin{cases} \text{Max.} & H(X_{S_1}, X_{S_2}, \dots, X_{S_k}) \\ \text{Max.} & \sum_{i=1}^m T(\langle X_{S_1}, X_{S_2}, \dots, X_{S_k} \rangle; X_{F_i}) \\ \text{Min.} & C(X_{S_1}, X_{S_2}, \dots, X_{S_k}) \end{cases} \quad (14a)$$

or

$$\begin{cases} \text{Max.} & H(X_{S_1}, X_{S_2}, \dots, X_{S_k}) \\ \text{Max.} & T(\langle X_{S_1}, X_{S_2}, \dots, X_{S_k} \rangle; \langle X_{F_1}, X_{F_2}, \dots, X_{F_m} \rangle). \\ \text{Min.} & C(X_{S_1}, X_{S_2}, \dots, X_{S_k}) \end{cases} \quad (14b)$$

The objectives in equation (14a) [or equation (14b)] indicate that an optimal network should convey effective information as much as possible, while retaining redundant information, if any, as little as possible.

[32] To circumvent the complexity of multiobjective optimization and facilitate end user's decision making, the three objectives can be additively unified considering that they are commensurate. Hence, the multiobjective optimization reduces to single-objective optimization. The integrated objective function corresponding to equation (14a) is

$$\begin{aligned} \text{Max.} \quad & \lambda_1 (H(X_{S_1}, X_{S_2}, \dots, X_{S_k}) + \sum_{i=1}^m T(\langle X_{S_1}, X_{S_2}, \dots, X_{S_k} \rangle; \\ & X_{F_i})) - \lambda_2 C(X_{S_1}, X_{S_2}, \dots, X_{S_k}), \end{aligned} \quad (15a)$$

where  $\lambda_1$  and  $\lambda_2$ , whose summation is 1, respectively, are the information-redundancy tradeoff weights. The purpose of tradeoff weights is to give users the possibility to include

additional knowledge or their preference. Moreover, varying tradeoff weights can achieve different feasible solutions under different scenarios. Therefore, the advantage of multi-objective optimization is retained to some extent. Similarly, the analogous integrated objective function corresponding to equation (14b) is

$$\text{Max. } \lambda_1(H(X_{S_1}, \dots, X_{S_k}) + T(\langle X_{S_1}, \dots, X_{S_k} \rangle; \langle X_{F_1}, \dots, X_{F_m} \rangle)) - \lambda_2 C(X_{S_1}, \dots, X_{S_k}). \quad (15b)$$

### 3.2. Selection Algorithm

[33] Using the MIMR criterion, a selection algorithm for the design of hydrometric networks can be generalized in what follows:

[34] Algorithm 2

[35] 1. Collect data of hydrometric variable of interest at each candidate station;

[36] 2. Discretize the continuous time series data by equation (6);

[37] 3. Initialize the optimal set  $S$  as an empty set and the candidate set  $F$  as the one containing all candidate stations;

[38] 4. Identify the central station as the one with maximum marginal entropy among all candidates;

[39] 5. Update sets  $S$  and  $F$ ;

[40] 6. Select the next optimal station from  $F$  by the MIMR criterion. In this step all stations in  $F$  are scanned sequentially to search the one satisfying equation (15a) [or equation (15b)];

[41] 7. Repeat steps 5 and 6 iteratively until the expected number of stations have been selected.

[42] A pseudo code for algorithm 2 is given in Table 1. Convergence of the selection can be determined by the ratio of joint entropy of the selected stations to that of all the candidates. If the ratio is over a threshold, like 0.90,

the selection stops. If no threshold is provided, then all candidate stations will be ranked in descending order of priority. To reduce the implementation effort, a MATLAB package HydroMIMR was developed.

[43] Other than forward selection, the optimal stations can also be determined in an opposite direction. For backward selection, the criterion should be changed to minimum information reduction and maximum redundancy reduction, which is also based on the principle of MIMR.

## 4. Application

[44] Two case studies are presented in this section to illustrate the applicability of MIMR in evaluating and designing a hydrometric network. The first case shows how to evaluate an existing streamflow gauge network. The second case demonstrates the way to optimally select stations from a dense polder water level monitoring system.

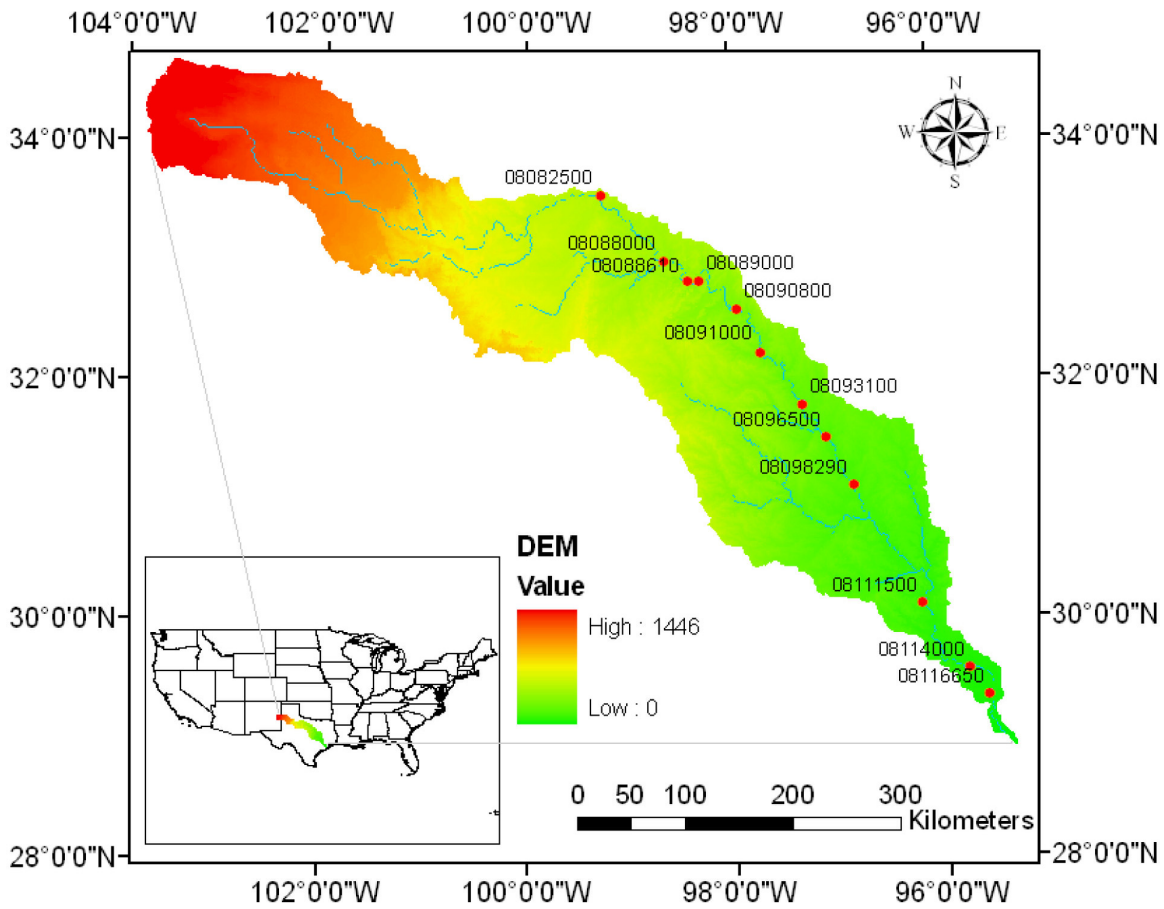
### 4.1. Case 1: Streamflow Network Evaluation

#### 4.1.1. Study Area and Data

[45] This case study aims to rank the importance of streamflow gauges located on the mainstream of the Brazos River in Texas. The Brazos River, with a drainage area of about 118,000 km<sup>2</sup>, extends from eastern New Mexico, and flows for more than 1000 km southeasterly across Texas, to the Gulf of Mexico (Figure 1) [Wurbs *et al.*, 2005]. Monthly scale streamflow observations from 1990 to 2009 of 12 USGS stream gauges were selected for this study, as depicted in Figure 1. Notice that only gauges on the mainstream were chosen. Evaluating gauges separately for each tributary and the mainstream is reasonable, since they belong to different geophysical hydrological units and pooled evaluation would lead to irrational results.

**Table 1.** Pseudo Code for the MIMR-Based Greedy Selection Algorithm

Commands Outline	Comments
1: $F \leftarrow$ candidate set including all candidate stations $S \leftarrow$ empty set	Initialize candidate set $F$ and empty set $S$
2: Discretize the continuous time series.	
3: $Info \leftarrow H(F) \leftarrow$ Equation (8)	Compute the total information ( $Info$ ) of all the potential stations
4: For $i = 1 : N$ $H(X_i) \leftarrow$ Equation (1) End	Compute the marginal entropy of each potential station
5: $X_i \leftarrow \arg \max_{X_i} [H(X_i)]$	Select the first center station
6: $F \leftarrow F - X_i$ $S \leftarrow S + X_i$	Update $F$ and $S$ for the first time
7: For $i = 2 : N$ $m \leftarrow \text{size}(F)$ $n \leftarrow \text{size}(S)$ For $k = 1 : m$ $MIMR_{S+X_{F_k}} \leftarrow$ Equation (14) End $X_{F_k} = \arg \max_{X_{F_k}} (MIMR_{S+X_{F_k}})$ $F \leftarrow F - X_{F_k}$ $S \leftarrow S + X_{F_k}$ End	Sequentially select station from the updated candidate set according to MIMR criterion  Update the candidate set and already selected set successively
8: For $i = 1 : N$ $sInfo \leftarrow H(X_{S_1}, X_{S_2}, \dots, X_{S_i})$ $sInfo$ $pct \leftarrow \frac{sInfo}{Info}$ If $ pct - threshold  < eps$ $S_{final} \leftarrow \{X_{S_1}, X_{S_2}, \dots, X_{S_i}\}$ return End End	Determine the final optimal set $S_{final}$ according to the information fraction of the selected set to the total information. Joint entropy of selected stations is denoted by $sInfo$



**Figure 1.** Location map of the Brazos River basin and streamflow gauges on the mainstream.

Reallocated station identification number (ID) sequentially from upstream to downstream and the corresponding summary statistics of streamflow are given in Table 2. One major finding was the increased variability of streamflow from upstream to downstream as signified by the increased variance. From the perspective of information theory, generally the larger the variance of a station the more information (or uncertainty) it contains. In this sense, the importance of each station also increases from upstream to downstream. Selecting stations simply based on the variance (or standard deviation), however, cannot make sure that the transformation between selected and unselected

stations is high and the redundant information among the selected stations is low.

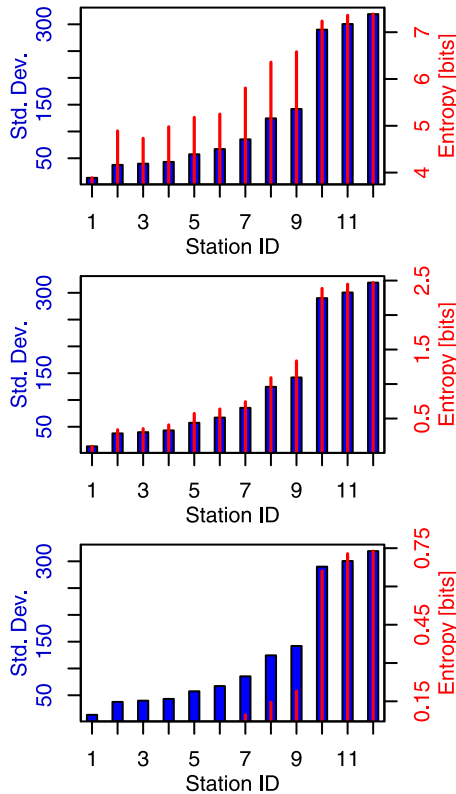
**4.1.2. Results**

[46] Prior to the MIMR ranking, the continuous streamflow time series for each station was discretized using equation (6). A too small value of  $a$  would make the information content retained by every station indistinguishable, leading to inconvenience for further analysis. As illustrated in the top plot of Figure 2, the information content measured by marginal entropy was similar among stations due to a small value of  $a$  ( $1.0 \text{ m}^3 \text{ s}^{-1}$ ). On the other

**Table 2.** USGS Code, Reallocated ID and Summary Statistics of Streamflow for Each Gauge in this Study<sup>a</sup>

USGS Code	ID	Mean	Variance	SD	Maximum	Median	Minimum
08082500	1	7.13	181.65	13.47	99.25	2.47	0.00
08088000	2	18.96	1415.07	37.62	254.48	5.19	0.00
08088610	3	18.92	1588.43	39.86	245.20	6.51	0.77
08089000	4	21.99	1864.03	43.17	256.66	7.78	0.96
08090800	5	28.09	3293.39	57.39	377.18	8.17	0.75
08091000	6	31.68	4502.97	67.10	423.62	9.00	0.35
08093100	7	44.57	7277.09	85.31	652.14	16.69	0.65
08096500	8	71.79	15493.18	124.47	807.02	27.82	0.94
08098290	9	89.79	20146.78	141.94	904.16	35.25	3.04
08111500	10	235.47	84091.05	289.98	1646.06	104.30	10.07
08114000	11	248.88	90336.52	300.56	1727.61	117.87	11.88
08116650	12	258.96	101733.94	318.96	1998.03	135.71	6.58

<sup>a</sup>Streamflow is measured in  $\text{m}^3 \text{ s}^{-1}$ .



**Figure 2.** Marginal entropy after discretization and standard deviation at the original scale of streamflow of each station. The discretization parameters are  $1.0 \text{ m}^3 \text{ s}^{-1}$ ,  $150.0 \text{ m}^3 \text{ s}^{-1}$  and  $1000.0 \text{ m}^3 \text{ s}^{-1}$  for each plot from top to bottom, respectively.

hand, an over large  $a$  would make stations with relatively small values insignificant. As in the bottom plot of Figure 2, the large  $a$  ( $1000 \text{ m}^3 \text{ s}^{-1}$ ) compiled stations with small discharges to be noninformative (e.g., gauges with ID 1, 2, and 3). That is because equation (6) with large  $a$  can be considered as a “high-pass-filter,” by which values below a high threshold will be converted to the same integer. In both of the above cases, the spatial variability of streamflow was distorted after discretization. From the middle plot of Figure 2 corresponding to an appropriate value of  $150 \text{ m}^3 \text{ s}^{-1}$  for  $a$ , which is determined by trial and error, one can see that all stations were informative and distinguishable. The spatial variability of streamflow among stations is also preserved. To confirm the rationale of  $150 \text{ m}^3 \text{ s}^{-1}$  for  $a$ , a sensitivity analysis was carried out by comparing station ranks corresponding to different integers of  $a$  varying from  $130 \text{ m}^3 \text{ s}^{-1}$  to  $170 \text{ m}^3 \text{ s}^{-1}$ . Results showed that station ranks obtained by MIMR were stable. The top 9 selected stations kept invariant no matter how  $a$  changed. These observations together empirically justified the rationale for the value of  $150 \text{ m}^3 \text{ s}^{-1}$  for  $a$ .

[47] Marginal entropy map, multivariate joint entropy and total correlation provide an overall picture of the behavior of information content of all candidate stations. Figure 2 offered an implicit entropy map. The station with minimum marginal entropy was located at the most upstream on the river, whereas the one with maximum entropy was found at the most downstream location. Also we

detected an increasing trend of information content from upstream to downstream, which was consistent with the empirical conclusion achieved based on the streamflow variance. Following this, we suspected that the information amount increased as the controlled drainage area accumulated. The joint entropy of all stations was 4.10 bits, representing the maximum information amount that could be extracted from them. The total correlation was 8.78 bits, signifying the amount of duplicated information among them. The sum of marginal entropies of all stations was 12.88 bits, which was approximately 47% larger than the total correlation, implying the gauging system was of an information-scattered type.

[48] With information weight  $\lambda_1$  adopting a value of 0.8, the MIMR selection was implemented to rank the 12 gauges. Justification for the value 0.8 for  $\lambda_1$  would be detailed in section 4.1.4. Results of MIMR selection are presented in Table 3. We noted the following: (1) as expected, the “central” station, with ID 12, was the one with maximum marginal entropy located at the most downstream area; (2) unlike that derived from marginal entropy or variance, the MIMR based priority rank did not exhibit an increasing pattern from upstream to downstream; (3) the joint entropy of stations in the optimal set did not always increase as more and more stations were added; (4) the top 6 stations explained roughly 80% of the total information retained by the 12 stations; (5) generally, multivariate transinformation  $T_1$  and  $T_2$  first increased and then decreased as selection proceeded; and (6) the total correlation increased as a new station was added to the optimal set.

[49] Observation 2 indicates that siting gauges simply based on marginal entropy cannot obtain an optimal network, but leads to the one with either low transinformation or high duplicated information. Concerning observation 3, theoretically, the joint entropy is nondecreasing as more stations are added. It is invariant if and only if the information of new added stations is thoroughly duplicated among those already selected. The invariant joint entropy over steps 5, 6 and 7 is not surprising, if one looks at the close locations of the selected stations (Figure 1) during these steps (ID 2, 3, and 4). The decreased transinformation in observation 5 is explained by realizing that fewer and fewer stations were left in the candidate set as selection proceeded. Figure 3 (MIMR) shows the spatial locations of the top 6 selected stations. They are nearly uniformly distributed along the stream, which is desirable for obtaining the area mean value. Even though the spatially averaged streamflow is of little practical interest, the MIMR criterion is generic for any hydrometric network evaluation and design, of course suitable for siting gauges across a rainfall field, for which the area mean value is important.

#### 4.1.3. Comparison With WMPs

[50] We compared the performance of MIMR selection with another entropy-based method developed by *Alfonso et al.* [2010a], i.e., water level monitoring design in polders (WMP), whose objective function is generalized as:

$$\begin{cases} \text{Max.} & H(X_1) + \dots + H(X_i) \\ \text{Min.} & \sum_{j=1}^{i-1} T(X_i; X_j) \end{cases}, \quad (16)$$

**Table 3.** Joint Entropy, Transinformation, and Total Correlation of Ranked Streamflow Gauges Obtained by MIMR With  $\lambda_1 = 0.8$  and by WMPs

	Iteration Step											
	1	2	3	4	5	6	7	8	9	10	11	12
	<i>MIMR</i>											
ID	12	6	1	8	2	3	4	7	5	9	10	11
$H$	2.47	2.84	2.87	3.21	3.23	3.23	3.23	3.32	3.33	3.52	3.93	4.10
$T_1$	6.57	7.12	7.76	7.42	7.13	6.78	6.37	5.81	5.24	4.12	2.28	
$T_2$	2.17	2.54	2.59	2.77	2.78	2.77	2.77	2.76	2.58	2.32	2.28	
$C$	0.00	0.27	0.34	1.09	1.40	1.75	2.16	2.82	3.38	4.53	6.50	8.78
	<i>WMP1</i>											
ID	12	9	7	6	5	4	3	2	1			
$H$	2.47	3.07	3.21	3.36	3.37	3.37	3.38	3.38	3.38			
$T_1$	6.57	7.22	6.89	6.81	6.25	5.84	5.51	5.17	5.07			
$T_2$	2.17	2.59	2.66	2.80	2.72	2.69	2.70	2.62	2.58			
$C$	0.00	0.73	1.34	1.82	2.38	2.79	3.13	3.47	3.57			
	<i>WMP2</i>											
ID	12	9	7	6	5	4	3	2	1			
$H$	2.47	3.07	3.21	3.36	3.37	3.37	3.38	3.38	3.38			
$T_1$	6.57	7.22	6.89	6.81	6.25	5.84	5.51	5.17	5.07			
$T_2$	2.17	2.59	2.66	2.80	2.72	2.69	2.70	2.62	2.58			
$C$	0.00	0.73	1.34	1.82	2.38	2.79	3.13	3.47	3.57			
	<i>WMP3</i>											
ID	12	9	6	11	10	8	5	2	1			
$H$	2.47	3.07	3.26	3.57	3.86	4.01	4.02	4.06	4.08			
$T_1$	6.57	7.22	7.26	5.29	3.20	2.31	1.79	1.56	1.46			
$T_2$	2.17	2.59	2.76	2.65	1.41	1.11	0.98	0.98	0.95			
$C$	0.00	0.73	1.20	3.32	5.42	6.36	6.92	7.22	7.32			

where  $X_i$  and  $X_j$  represent stations in sets  $F$  and  $S$ , respectively. Other than transinformation, another two criteria were used by WMP to measure the dependence of stations. One was the directional information transfer index

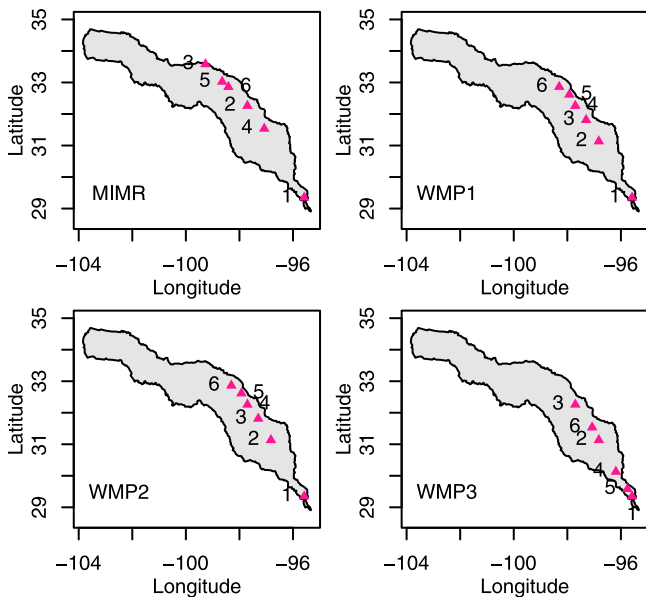
$$DIT_{YX}(X; Y) = \frac{T(X, Y)}{H(Y)}. \tag{17}$$

The other one was also the directional information transfer index but from  $X$  to  $Y$ , i.e.,

$$DIT_{XY}(X; Y) = \frac{T(X, Y)}{H(X)}. \tag{18}$$

For simplicity, WMPs with transinformation,  $DIT_{YX}$  and  $DIT_{XY}$  were denoted by WMP1, WMP2 and WMP3, respectively. The objective functions for WMP2 and WMP3 can be obtained by replacing  $T(X_i; X_j)$  in equation (16) with  $DIT_{YX}(X_i; X_j)$  and  $DIT_{XY}(X_i; X_j)$ , respectively. A brief description of the step by step selection of WMPs is generalized in a way such that their selection concerns could be explicitly embodied. Implementation of WMPs should strictly follow procedures in the Appendix.

[51] A summary of entropy terms associated with WMPs is tabulated in Table 3. First, we compared the performance of MIMR and WMPs in finding stations with high information content. The joint entropy of MIMR-ranked stations was a little bit smaller than that of stations obtained by WMPs. The top 6 stations selected by MIMR explained approximately 80% of the total information and 82% of which was explained by the same number of top stations derived from WMP1 and WMP2. The performance of MIMR was comparable to WMP1 and WMP2 from the point of view of joint entropy. For WMP3, the top 6 stations explained 98% of the total information. The above comparison might seem against MIMR. A network is considered as optimal not only because of its high joint entropy, but it is better to balance its performance in the three respects mentioned in section 3.



**Figure 3.** Spatial distribution of the top 6 streamflow gauges ranked by MIMR with  $\lambda_1 = 0.8$  and by WMPs.



[52] Second, we compared the performance of MIMR and WMPs in locating stations with high information transition ability. By looking into  $T_1$  in Table 3, one can see that MIMR outperformed WMPs, indicating the decent information transition ability of the network obtained from MIMR. From the point of view of  $T_2$ , the performance of MIMR was not always superior to that of WMPs. This is unsurprising since in this study MIMR used function (14a) as an objective in which  $T_2$  was not accounted for. Actually even in this case, MIMR outperformed WMPs after more than 4 stations were selected. The above analysis also suggests a hydrometric network with high joint entropy does not mean that it also has decent information transition ability.

[53] Third, we evaluated the performance of MIMR in searching independent stations. Total correlation of the top 6 stations selected by MIMR occupied less than 20% of the duplicated information among the 12 stations. The percentages for the same number of top stations derived from WMPs were as high as values ranging approximately from 32% to 72%, as inferred from Table 3. Apparently, MIMR is more effective in searching independent stations.

[54] Finally, from the spatial distribution of the top ranked stations by different approaches (Figure 3), an interesting observation emerged. Compared with stations selected by MIMR, stations obtained from WMPs clustered along the stream, explaining why they contained more duplicated information. It is desired to screen the clustering effect of a data collection network. In this sense, MIMR should gain more confidence.

#### 4.1.4. Sensitivity Analysis of $\lambda_1$

[55] Since the primary purpose of data collection is to obtain information, the sensitivity analysis was only carried out for information weight falling between 0.5 and 1.0. Results are summarized in Table 4, which mainly signify the stability of MIMR with respect to information weight. In this case, the stability of MIMR was mainly because of the relatively small number of candidate stations. When an existing network is dense, it is cautioned that paying special attention to the tradeoff weights is required.

[56] As previously stated, the tradeoff weights should reflect the user's knowledge and preference about the hydrometric network under consideration. In practice, however, such kind of prior information or experience may not always be available. Even if it is available, selecting optimal weights is still challenging. One question arises: how to determine a suitable information weight without prior knowledge about the system? Investigating the performance of MIMR with different tradeoff weights can achieve

useful hints. As inferred from Figure 4, a more informative but less independent network can be obtained by increasing the information weight. Maximum information and minimum redundancy are two conflicting objectives. It is better to balance them in order to reach a satisfactory result. The information-redundancy tradeoff weights provide a flexible "handle" to control this balance. As seen from Figure 4, a value of 0.8 for information weight led to a decent balance between information and redundancy, which provided a justification for selecting 0.8 for  $\lambda_1$  in this case study.

[57] The behavior of different optimality measures associated with WMPs is also presented in Figure 4. It was noted that as the information weight increased, MIMR moved toward WMP1 and WMP2. MIMR performed better in maximizing the information transition ability and avoiding redundant stations. Concerning the joint entropy, MIMR performed slightly worse than WMPs. An interesting result of MIMR worth noting is the smaller total correlation (redundancy) than that related to WMPs even when the information weight was 1.0. A value of 1.0 for information weight means that the objective of minimizing redundancy was excluded from the selection. Figure 4 inspires a viable approach to guide the selection of tradeoff weights. Using WMPs as benchmark methods, the information weight is modulated until a satisfactory result is obtained.

## 4.2. Case 2: Polder Water Level Monitoring Network Design

### 4.2.1. Study Area and Data

[58] Water level time series of a highly controlled polder system reported by *Alfonso et al.* [2010a] were used in this case study. The polder system is located in a low-lying region of Pijnacker, Delfland, Netherlands. There were in total 1520 potential monitoring points along the canals separated by a distance of 15 m on average, as depicted in Figure 5. Water level time series for these points were generated from a hydrodynamic model built by the Delfland Water Board. Driven by a given storm event, the model was run for a simulation period of 10 days at a time step of 15 min, resulting in 1520 synthetic water level time series, each with a length of 960. At each computational point, only the first 792 records were used in the present analysis due to data availability.

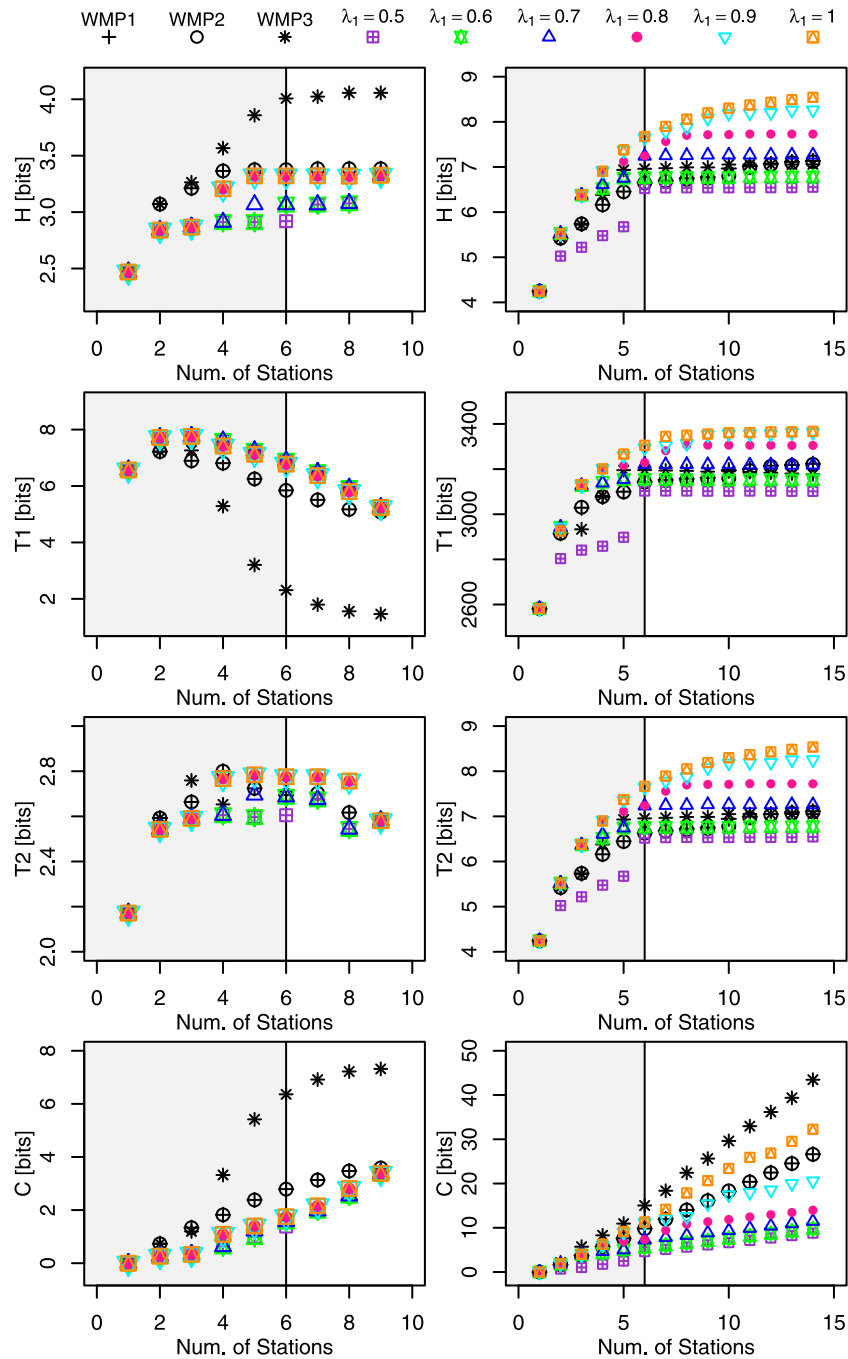
### 4.2.2. Results

[59] The generated continuous water level time series were first discretized by equation (6) with a value of 5 cm for parameter  $a$ . This value was determined based on physical considerations that water level variations smaller than 5 cm were caused by wind, ship movement, or dynamic waves generated by the operation of pumping stations [*Alfonso et al.*, 2010a]. These kinds of variations should be treated as noise. The role of the mathematical floor function was to filter out such noise to produce a noise-free pulse signal. This example illustrates the advantage of mathematical floor function in incorporating physical considerations.

[60] The marginal entropy for each simulation point was superimposed over the location map in Figure 5. The 16 null entropy points were highlighted by black points. These points corresponded to pumping stations discharging to big storage bodies, where fixed boundary conditions for water level should be satisfied. Generally, points with static

**Table 4.** IDs of Selected Stations Step By Step Using the MIMR Criterion With Different Information Weights

$\lambda_1$	Iteration Step											
	1	2	3	4	5	6	7	8	9	10	11	12
0.5	12	6	1	2	3	4	7	5	8	9	10	11
0.6	12	6	1	2	3	7	4	5	8	9	10	11
0.7	12	6	1	2	7	3	4	5	8	9	10	11
0.8	12	6	1	8	2	3	4	7	5	9	10	11
0.9	12	6	1	8	2	3	4	7	5	9	10	11
1.0	12	6	1	8	2	3	4	7	5	9	10	11



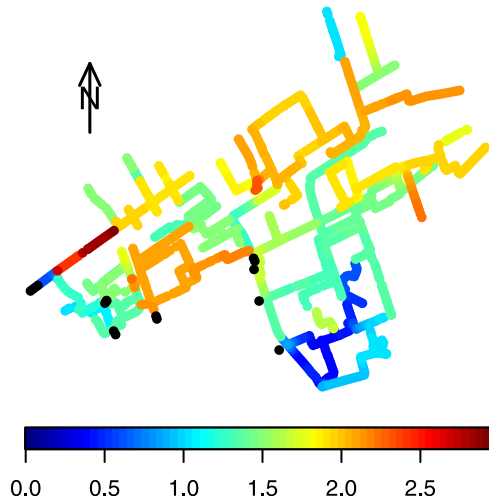
**Figure 4.** Behavior of different entropy terms of stations selected by MIMR with different information-redundancy tradeoff weights and by WMPs.

observations can be determined in advance whether or not to be retained in the final monitoring set based on management considerations. Therefore, they were left out from the following analysis.

[61] Figure 5 provided an overall picture for the information content of the polder system. Besides the null information points, a low information area was identified in the south part of the system. A short canal segment with extremely high information content was found in the northwest part, implying the suitable area for siting the “central” monitor. Excluding the null information points, the joint entropy of the 1504 monitors was 9.09 bits, which represented

the maximum information amount that can be obtained from the system. The total correlation was 3411.49 bits, which quantified the duplicated information of the system. The summation of the marginal entropies was 3420.58 bits, which was almost equal to the amount of duplicated information, suggesting the synthetic network was of an information-clustered type.

[62] A summary of entropy terms for the top 10 monitors obtained by MIMR with a value of 0.8 for the information weight are presented in Table 5. The joint entropy reached a plateau after 9 monitors were selected, which was used as the convergence indicator. The top 9 monitors explained



**Figure 5.** Marginal entropy map of the 1520 calculation points used in this study. Points with null entropy are highlighted with black points.

about 84.9% of the total information, whereas roughly 85% of which was explained after adding one more station. The increment was no more than 0.1%, indicating it was reasonable to choose the top 9 monitors as the final set. Transinformation  $T_1$  was much larger than  $T_2$ . This was expected, since the duplicated information in  $F$  was filtered out in the variable merging process.

**4.2.3. Comparison With WMPs**

[63] Different entropy terms corresponding to MIMR and WMPs were also presented in Table 5. More or less the same results were obtained as in case 1. The better performance of MIMR was more apparent. The joint entropy of selected stations by MIMR was larger than that of stations selected by WMPs. This confirmed the decent performance of MIMR in locating stations with high information content. Both  $T_1$  and  $T_2$  obtained by MIMR

were higher than those obtained by WMPs. Even though equation (14a) was used as the objective in the MIMR selection, implying that  $T_2$  was not accounted for,  $T_2$  obtained by MIMR was still larger than that obtained by WMPs. This is another interesting finding about MIMR. The network selected by MIMR contained notably less redundant information than that contained in the network obtained by WMPs, as indicated by total correlation.

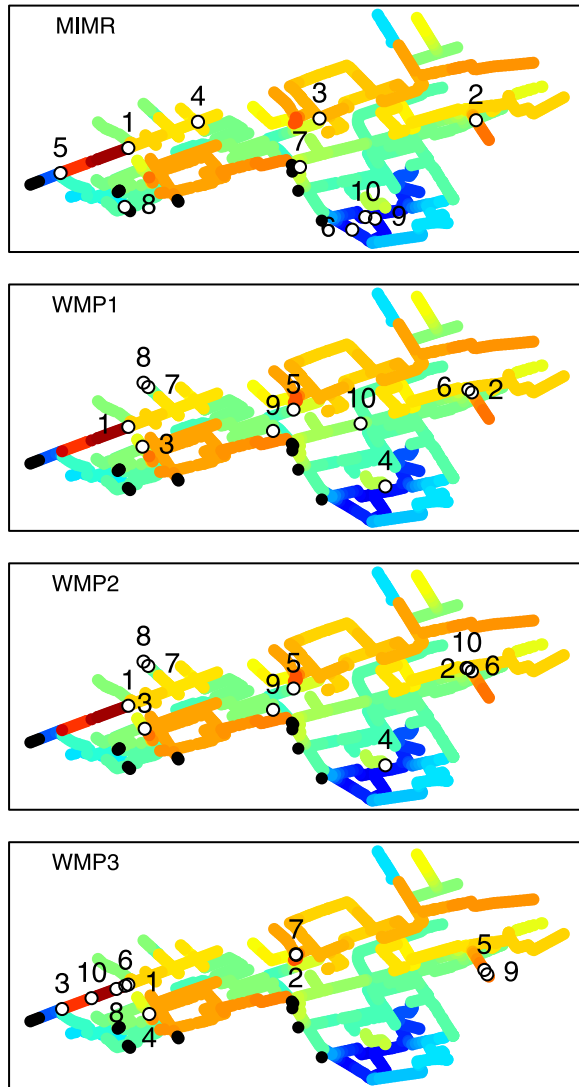
[64] Spatial locations of the selected stations are presented in Figure 6. The top 10 monitors derived by MIMR were widely spaced without clusters across the canal system, confirming the good performance of MIMR in determining a network with minimum redundant information. The spatial distribution of the top 10 stations ranked by WMP1 and WMP2 was almost the same except the one selected at the 10th step. This observation explained why entropy terms associated with WMP1 and WMP2 were similar as given in Table 5. Compared with the widely spaced stations determined by MIMR, station clusters were found in the network determined by WMPs, especially by WMP3. This signified the good performance of MIMR in determining an unbiased monitoring network as long as the information weight was properly selected. A monitoring network with clustered stations is considered as biased in that the underlying spatial variability of the variable of interest cannot be realistically discerned by the data collected from it.

**4.2.4. Sensitivity Analysis of  $\lambda_1$**

[65] The right panel of Figure 4 displays the results of sensitivity analysis, which are similar to those obtained in case 1. As can be seen, the performance of MIMR approached and then surpassed that of WMPs with respect to multivariate joint entropy and transinformation as the information weight increased from 0.5 to 1.0. On the other hand, from the view point of total correlation, MIMR moved toward and then surpassed WMP1 and WMP2 but never reached WMP3. Again it was found that even though the redundancy weight was null, the MIMR criterion still performed better in finding independent stations compared with WMP3.

**Table 5.** Joint Entropy, Transinformation, and Total Correlation of Monitors Obtained by MIMR With  $\lambda_1 = 0.8$  and by WMPs

	Iteration Step									
	1	2	3	4	5	6	7	8	9	10
	<i>MIMR</i>									
$H$	4.25	5.53	6.37	6.92	7.12	7.25	7.57	7.70	7.72	7.73
$T_1$	2581.58	2942.34	3126.60	3201.55	3215.89	3229.46	3281.50	3306.45	3306.40	3306.19
$T_2$	4.24	5.52	6.37	6.91	7.11	7.24	7.56	7.70	7.71	7.71
$C$	0.00	1.76	3.88	6.10	6.89	7.31	9.44	10.83	11.37	11.87
	<i>WMP1</i>									
$H$	4.25	5.42	5.74	6.16	6.46	6.65	6.69	6.75	6.77	6.97
$T_1$	2581.58	2915.38	3028.62	3076.65	3099.92	3143.04	3151.82	3157.58	3161.70	3159.53
$T_2$	4.24	5.42	5.74	6.16	6.45	6.63	6.68	6.73	6.75	6.94
$C$	0.00	1.63	3.70	5.72	7.60	9.83	11.95	14.05	16.17	18.18
	<i>WMP2</i>									
$H$	4.25	5.42	5.74	6.16	6.45	6.65	6.69	6.73	6.75	6.80
$T_1$	2581.58	2915.38	3031.15	3077.77	3098.63	3142.14	3149.10	3152.33	3156.54	3159.98
$T_2$	4.24	5.42	5.74	6.16	6.44	6.62	6.67	6.70	6.73	6.78
$C$	0.00	1.63	3.65	5.67	7.55	9.78	11.90	14.03	16.14	18.34
	<i>WMP3</i>									
$H$	4.25	5.43	5.72	6.38	6.94	6.96	6.97	6.99	6.99	7.05
$T_1$	2581.58	2918.61	2933.07	3081.56	3195.22	3195.47	3193.73	3192.24	3188.90	3190.04
$T_2$	4.24	5.43	5.72	6.37	6.93	6.95	6.97	6.99	6.99	7.05
$C$	0.00	2.16	5.72	8.29	10.94	15.03	18.37	22.43	25.64	29.62



**Figure 6.** Spatial distribution of the top 10 stations ranked by MIMR with  $\lambda_1 = 0.8$  and by WMPs.

[66] Spatial locations of the selected stations by MIMR with different information weights are presented in Figure 7. When the information weight was small, a cluster of stations with small information content was always identified. A small information weight corresponds to a large redundancy weight. With a large redundancy weight, the total correlation has more influence on the objective function. Therefore, stations which would introduce less redundant information were selected. The points with small information content in  $F$  shared a small amount of information with stations in  $S$ . That is why clustered stations were selected from the low information area. It was also noted that as the information weight increased, the selected stations became widely spaced, and then became clustered again (e.g.,  $\lambda_1 = 1.0$ ). This time the clustered stations were found in the high information area. In view of this, it was suggested again, using WMPs as benchmark methods to modulate the information weight until a satisfactory result can be obtained. This provides a practical way for determining suitable information-redundancy tradeoff weights.

### 5. Merits and Demerits of MIMR

[67] It is instructive to compare the criteria of different approaches that have appeared in the literature through which to highlight the merits and demerits of the MIMR criterion.

[68] The idea of the information maximization principle [Caselton and Husain, 1980; Husain, 1987, 1989; Al-Zahrani and Husain, 1998; Yoo et al., 2008] is to select a set of  $p$  stations from a dense network of  $m$  stations such that

$$\text{Max. } \sum_{i=1}^p H(X_i^p) + \sum_{i=1}^p \sum_{j=1}^{m-p} T(X_i^p, X_j^{m-p}), \quad (19)$$

where  $X_i^p$  represents the station in the possible  $p$ -combination of the  $m$  stations and  $X_j^{m-p}$  denotes the station outside of the  $p$ -combination. The combination of these  $p$  stations that satisfies equation (19) will be the optimally retained stations. Redundancy of the network is not accounted for.

[69] The idea of Krstanovic and Singh [1992a, 1992b] is to select the “central” station first then sequentially select others that give the lowest reduction in uncertainty. The objective is generalized as

$$\begin{aligned} \text{Min. } & H(X_1, \dots, X_{i-1}) - H((X_1, \dots, X_{i-1})|X_i) \\ & = \text{Min. } T(X_1, \dots, X_{i-1}; X_i), \end{aligned} \quad (20)$$

where  $X_i$  represents the station selected in the  $i$ th step. Equation (20) focuses on avoiding the redundant information of a network and explicitly neglects its ability in maximizing the joint entropy and the transinformation between selected and unselected stations.

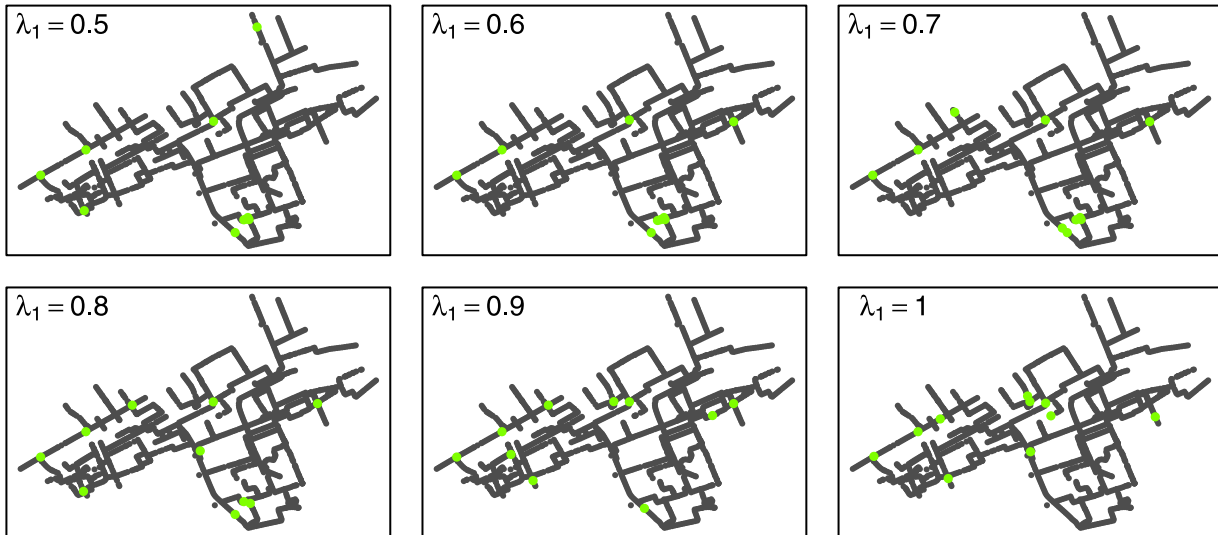
[70] A two-phase approach was described by Yang and Burn [1994]. The first stage uses a hierarchical clustering technique to form groups of hydrometric gauging stations. The second stage involves the selection of a single station from each group for retention in the final network. The single station is selected based on that the maximum fraction of its information can be transferred to other stations in the group. Generally, this approach focuses on the information transition ability and on avoiding redundancy of the network. It omits the ability in providing maximum joint entropy.

[71] Quantitative comparisons between WMPs and MIMR have been detailed. One more point we want to further highlight is that WMPs are implemented through matrix operation. Therefore it is more computationally efficient than the MIMR greedy selection. At each step of MIMR selection, all stations remaining in the candidate set will be scanned to search the optimal one, which would largely increase the computational burden.

[72] Another criterion was described by Alfonso et al. [2010b], which is generalized as

$$\begin{cases} \text{Max. } [H(X_1, \dots, X_M) - M\mu] \\ \text{Min. } [C(X_1, \dots, X_M) + M\mu] \end{cases}, \quad (21)$$

where  $X_i, i = 1, 2, \dots, M$  represent the selected stations in the optimal set and  $\mu$  is a constant with cost units of bits per new station. Equation (21) fails to account for



**Figure 7.** Spatial distribution of selected stations by MIMR with different information-redundancy tradeoff weights.

transformation. The advancement is the consideration for the economic effect. It is easy to incorporate a cost function into MIMR. The important point is to develop cost and benefit functions in information units. Compared with the approach used in the reference by *Alfonso et al.* [2010b] to solve the multiobjective problem, the greedy ranking algorithm for MIMR selection is straightforward and much easier to implement. However, we want to caution that this simplicity is traded by the declined diversity of feasible solutions.

[73] Another limitation of MIMR selection is that it is only suitable for choosing stations from a dense network. How to expand a sparse network is beyond its capacity. Fortunately, this problem can be addressed by simulation. Physically based hydrologic or hydrodynamic models can be used to generate synthetic dense networks, like the polder water level monitors in case 2. Weather generators are accessible to generate synthetic rainfall fields [Wilks, 2009]. High spatial and temporal resolution of radar rainfall products also can be exploited to approach this problem [Volkman et al., 2010]. It is cautioned that the synthetic dense network should preserve the properties of the observed network as much as possible. Otherwise, one should be careful with the potential dangers.

## 6. Concluding Remarks

[74] A generic criterion for the evaluation and design of hydrometric network based on entropy theory is presented. The proposed criterion, named as maximum information minimum redundancy (MIMR), covers the following three optimality measures of a hydrometric network: (1) joint entropy, (2) transinformation, (3) and total correlation. In terms of the MIMR, the selected network can provide the highest information content and avoid dependent stations as much as possible, guaranteeing while the stations within and outside of the optimal set has high common information.

Based on MIMR, a greedy ranking algorithm is proposed to optimally select hydrometric stations from a dense network. The optimum number of stations is determined by the percentages of total information that can be explained by stations in the optimal set.

[75] With the aid of discrete variable merging, all of the entropy terms involved in MIMR can be easily evaluated at high dimension without any distributional assumption. The basic idea of variable merging is to create a new variable whose information content is equal to that retained by the original variables. This technique is only suitable for discrete data. In the context of hydrometric network design (or evaluation), continuous time series data should be discretized first, which is done with the use of a mathematical floor function. Mathematical floor function on one hand avoids the choice of a parametric distribution to fit the continuous data and on the other hand, it can incorporate physical considerations. Three thumb rules suggest how to select a proper discretization parameter suitable for general hydrometric system, even in the case that no clear physical consideration is available.

[76] The MIMR selection is quantitatively compared with another entropy-based approach, i.e., WMPs. Results show that MIMR is better at finding stations with high information content, and better at locating independent stations. The information-redundancy tradeoff weights provide the user a flexible handle to balance the two conflicting objectives: maximum information and minimum redundancy. The tradeoff weights reflect the user's knowledge about the hydrometric network. When such knowledge is unavailable, a promoted approach is to use the computationally efficient WMPs as benchmarks and adjust the information weight until a satisfactory result is obtained. In this study, we only quantitatively compare MIMR with WMPs. Further comparisons with other approaches are merited, especially with *Alfonso et al.* [2010b], to see the difference between monitoring networks determined by greedy selection and by multiobjective optimization.

[77] Even though there are many ways to define the objective for the design of a hydrometric network, MIMR focuses on the fundamental theme, i.e., selecting an optimum number of stations and their optimum locations. Other considerations, like the cost of placing new stations, are not included. Nevertheless, they can be easily incorporated into the objective by introducing an extra penalty function. The crucial point is to find an appropriate way to measure the economic cost in terms of information units. This is a problem pending for future work.

## Appendix A: The Step by Step Implementation Procedure of WMPs'

[78] Following Alfonso *et al.* [2010a], the step by step selection procedure of WMPs is generalized as:

[79] 1. Collect data of hydrometric variable  $X_i$  of interest at each candidate station  $s_i$  ( $i = 1, 2, \dots, n$ , where  $n$  is the number of candidates) and discretize them using equation (6).

[80] 2. Calculate the marginal entropy  $H(X_i)$  for each  $s_i$  with the use of equation (1).

[81] 3. For each  $s_i$ , calculate the transinformation in equation (4) with respect to each of the remaining points and build the symmetric matrix

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

[82] For WMP2 and WMP3, the transinformation is replaced by directional information transfer index in equations (17) and (18), respectively.

[83] 4. Select the first monitor  $m_1$ , which is located at the point with the highest information content measured by marginal entropy, and add the monitor  $m_1$  to the selected set  $M$ .

[84] 5. Recover the transinformation vector  $v_1$  of the monitor  $m_1$ :  $v_1 = V(i, m_1)$ ,  $i = 1, 2, \dots, n$ .

[85] 6. Divide the candidates into two sets of stations with respect to their dependence on  $m_1$ : those that are dependent and those that are independent  $S_{m_1}^{ind}$ . The independent set  $S_{m_1}^{ind}$  is obtained by looking at the elements of  $v_1$  such that  $V(i, m_1)$  is less than a threshold given as mean ( $V(i, m_1)$ ).

[86] 7. From the independent set  $S_{m_1}^{ind}$ , select the second monitor  $m_2$ , which has the highest marginal entropy.

[87] 8. Recover the transinformation vector  $v_2$  of the monitor  $m_2$ :  $v_2 = V(i, m_2)$ ,  $i = 1, 2, \dots, n$ .

[88] 9. Select the next monitor  $m_3$  in a similar way, but now using an updated independent set  $S_{m_3}^{ind}$  given by the common set of independent points in the overlapping transinformation vector for the previously selected monitors  $m_1$  and  $m_2$ . Set  $v_3 = v_1 + v_2$ .

[89] 10. Set  $v_1 = v_3$  and repeat the procedure from step 5 until  $m_i$  does not provide significant information content or until the independent set  $S_{m_i}^{ind}$  is empty.

[90] **Acknowledgments.** The authors would like to express their sincere appreciation to Leonardo Alfonso, at the Hydroinformatics and Knowledge Management, UNESCO-IHE, Delft, Netherlands, who provided part of the data used in this paper. Three anonymous referees raised constructive comments. Their help in improving this paper is gratefully

acknowledged. The work was financially supported in part by the United States Geological Survey (USGS, Project ID: 2009TX334G) and TWRI through the project "Hydrological Drought Characterization for TX under Climate Change, with Implications for Water Resources Planning and Management."

## References

- Alfonso, L., A. Lobbrecht, and P. Price (2010a), Information theory-based approach for location of monitoring water level gauges in polders, *Water Resour. Res.*, *46*, W03528, doi:10.1029/2009WR008101.
- Alfonso, L., A. Lobbrecht, and R. Price (2010b), Optimization of water level monitoring network in polder systems using information theory, *Water Resour. Res.*, *46*, W12553, doi:10.1029/2009WR008953.
- Al-Zahrani, M., and T. Husain (1998), An algorithm for designing a precipitation network in the south-western region of Saudi Arabia, *J. Hydrol.*, *205*, 205–216.
- Bernadara, P., D. Schertzer, E. Sauquet, I. Tchiguirinskaia, and M. Lang (2008), The flood probability distribution tail: How heavy is it?, *Stochastic Environ. Res. Risk Assess.*, *22*, 107–122.
- Birgé, L., and Y. Rozenholc (2006), How many bins should be put in a regular histogram, *ESAIM: Probab. Stat.*, *10*, 24–45, doi:10.1051/ps:2006001.
- Carreau, J., P. Naveau, and E. Sauquet (2009), A statistical rainfall-runoff mixture mode with heavy-tailed components, *Water Resour. Res.*, *45*, W10437, doi:10.1029/2009WR007880.
- Casleton, W. F., and T. Husain (1980), Hydrologic network: Information transmission, *J. Water Resour. Manage. Div. Am. Soc. Civ. Eng.*, *106*(2), 503–520.
- Freedman, D., and P. Diaconis (1981), On the histogram as a density estimator: L2 theory, *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, *57*, 453–476.
- Harmancioglu, N. B., and V. P. Singh (1998), Entropy in environmental and water resources, in *Encyclopaedia of Hydrology and Water Resources*, edited by R. W. Herschy and R. W. Fairbridge, 832 pp., Kluwer, Dordrecht, Netherlands.
- Harmancioglu, N. B., and V. Yevjevich (1987), Transfer of hydrologic information among river points, *J. Hydrol.*, *91*, 103–118.
- Husain, T. (1987), Hydrologic network design formulation, *Can. Water Resour. J.*, *12*(1), 44–59.
- Husain, T. (1989), Hydrologic uncertainty measure and network design, *Water Resour. Bull.*, *25*(3), 527–534.
- 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. E., 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. E., and V. P. Singh (1992b), Evaluation of rainfall networks using entropy: II. application, *Water Resour. Manage.*, *6*, 295–314, doi:10.1007/BF00872282.
- Li, C., V. P. Singh, and A. K. Mishra (2012), Simulation of the entire range of daily precipitation using a hybrid probability distribution, *Water Resour. Res.*, *48*, W03521, doi:10.1029/2011WR011446.
- Mishra, A. K., and P. Coulibaly (2009), Developments in hydrometric network design: A review, *Rev. Geophys.*, *47*, RG2001, doi:10.1029/2007RG000243.
- Mishra, A. K., and P. Coulibaly (2010), Hydrometric network evaluation for Canadian watersheds, *J. Hydrol.*, *380*, 420–437.
- Mogheir, Y., V. P. Singh, and J. L. M. P. de Lima (2006), Spatial assessment and redesign of a groundwater quantity monitoring network using entropy theory, Gaza strip, Palestine, *Hydrogeol. J.*, *14*, 700–712, doi:10.1007/s10040-005-0464-3.
- Papoulis, A., and S. U. Pillai (2001), *Probability, Random Variables and Stochastic Processes*, MacGraw-Hill, New York.
- Ruddell, B. L., and P. Kumar (2009), Ecohydrologic process networks: 1. Identification, *Water Resour. Res.*, *45*, W03419, doi:10.1029/2008WR007279.
- Scott, D. W. (1979), On optimal and data-based histograms, *Biometrika*, *66*(3), 605–610.
- Shimazaki, H., and S. Shinomoto (2007), A method for selecting the bin size of a time histogram, *Neural Comput.*, *19*(6), 1503–1527.
- Volkman, T. H. M., S. W. Lyon, H. V. Gupta, and A. Troch (2010), Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain, *Water Resour. Res.*, *46*, W11554, doi:10.1029/2010WR009145.

- Watanabe, S. (1960), Information theoretical analysis of multivariate correlation, *IBM J. Res. Rev.*, 4, 66–82.
- Wilks, D. S. (2009), A gridded multisite weather generator and synchronization to observed weather data, *Water Resour. Res.*, 45, W10419, doi:10.1029/2009WR007902.
- Wurbs, R. A., R. S. Muttiah, and F. Felden (2005), Incorporation of climate change in water availability modeling, *J. Hydrol. Eng.*, 10(5), 375–385.
- 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.
- Yoo, C., K. Jung, and J. Lee (2008), Evaluation of rain gauge network using entropy theory: Comparison of mixed and continuous distributions function applications, *J. Hydrol. Eng.*, 13(4), 226–235.

---

C. Li, A. K. Mishra, and V. P. Singh, Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX 77843-2117 USA. (lichsunny@tamu.edu; amishra@tamu.edu; vsingh@tamu.edu)