Journal of Environmental Management 132 (2014) 24–31

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Efficient method for optimal placing of water quality monitoring stations for an ungauged basin

Changhyoun Lee, Kyungrock Paik*, Do Guen Yoo, Joong Hoon Kim

School of Civil, Environmental and Architectural Engineering, Korea University, Anam-dong 5 ga, Seongbuk-gu, Seoul 136-713, South Korea

ARTICLE INFO

Article history:
Received 11 December 2012
Received in revised form
9 October 2013
Accepted 14 October 2013
Available online 22 November 2013

Keywords:
Water quality monitoring
Monitoring network
Sampling location
Optimization

ABSTRACT

A core problem in monitoring water quality of a river basin is identifying an optimal positioning of a limited number of water-sampling sites. Various optimality criteria have been suggested for this selection process in earlier studies. However, the search for sets of sampling sites that satisfy such criteria poses a challenging optimization problem, especially for a large basin. Here, we show that for particular types of objective functions, the optimization procedure can be dramatically simplified via an analogy with the formulation of Shannon entropy. On this basis, we propose an efficient algorithm that can easily determine the optimal location of water quality sampling sites in a river network. The proposed algorithm can be used standalone or in conjunction with a heuristic optimization algorithm such as a genetic algorithm. For the latter, the proposed algorithm filters only competitive candidates and makes a contribution to reducing the problem size significantly. The superior performance of the proposed method is demonstrated via its application to actual river networks examined in earlier studies, in which the proposed method determines more optimal solutions in a shorter computation time. The idea presented in this study can also be applied to other problems in which the objective function can be formulated in a similar functional form.

1. Introduction

Regular monitoring of water quality in a river network is crucial for reliable water supply and preservation of a healthy ecosystem. To comprehensively determine water quality status across an entire river network, it would be ideal to have a virtually infinite number of sampling sites that provide spatially continuous data on water quality. However, there are always practical limitations, such as budget constraints, for maintaining sampling sites. Therefore, assessing the overall water quality status of a river network via a minimum number of sampling sites presents an important engineering problem.

In order to design an effective monitoring network, several criteria for the selection of sampling sites have been proposed in previous studies. Park et al. (2006) suggested four individual fitness functions, termed compliance with water quality data, supervision of water use, surveillance of pollution sources, and examination of water quality changes/estimation of pollution loads. The weighed sum of these was proposed as an objective function to determine the optimal location of monitoring sites. Telci et al. (2009) proposed two objective functions of minimizing the average time for contaminant detection and maximizing the reliability of a monitoring system. Some studies have utilized the information theory, in which disorder or uncertainty contained in a signal can be evaluated through entropy quantities such as the marginal entropy, conditional entropy, and transinformation (Shannon, 1948). Ozkul et al. (2000) suggested the criterion of minimum redundancy to determine optimal spatial distribution of sampling sites, and used transinformation as the measure of redundancy. Karamouz et al. (2009a) used not only transinformation but also marginal entropy as measures of information gained from individual sampling sites, and suggested a methodology for expansion or modification of an existing sampling network in terms of maximum total information gained from the sites with minimum redundancy. Karamouz et al. (2009b) studied sampling frequencies as well as locations on the basis of an objective function that minimizes the sum of differences between observed and simulated values of water quality variables.

The criteria described above are used for selecting a specific number of monitoring sites among candidates for which the water quality data have already been observed. In other words, they are not applicable before the installation of monitoring sites, and hence, an alternative method must be employed to design a network of monitoring sites for an ungauged basin or a basin for which there are insufficient field data. In this regard, Sharp (1971)

---

* Corresponding author. Tel.: +82 2 3290 4860; fax: +82 2 928 7656.
E-mail address: paik@korea.ac.kr (K. Paik).

0301-4797/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jenvman.2013.10.012

suggested a topologically optimum water-sampling plan in which the optimal locations of sampling sites can be estimated solely from the topological organization of a river network. Following this early idea of utilizing topological information, Dixon et al. (1999) proposed specific objective functions using variables such as the number of channel reaches η and their length Γ as

$$E = \frac{1}{\eta_0}\sum_{k=1}^{\varepsilon} \eta_k \log_2 \eta_k \tag{1}$$

and

$$F = \frac{1}{\Gamma_0}\sum_{k=1}^{\varepsilon} \Gamma_k \log_2 \eta_k. \tag{2}$$

Here, both E and F indicate the expected amount of work required to determine an exact pollutant source (a channel reach of the original pollution source) when contamination is observed in a river network with ε sampling sites. Therefore, a monitoring network that provides a smaller value for E or F is preferred for the same number of ε. Γ_0 and η_0 are the total length and the number of reaches in the entire river network, respectively, whereas Γ_k and η_k are the total length and the number of reaches, respectively, for so-called the coverage area of a sampling site k. Each sampling site is intended to detect the occurrence of contamination within its coverage area. The coverage area of a sampling site is the upstream area that is exclusively allocated to the sampling site, and differs from the contributing (or drainage) area. If there is another site upstream, the area covered by this upstream site is excluded from the coverage area of the downstream site (Fig. 1).

The formulation of Eq. (1) or (2) determines a reach within a river network at which a sampling site should be placed. That is, these formulations determine macrolocation (Sanders et al., 1983), but do not indicate a more detailed sampling location within a reach. Therefore, we assume that each sampling site is placed at the downstream end of a reach in order that the site can fully cover the given channel reach.

In order to identify optimal sampling sites based on cost functions such as Eqs. (1) and (2), an optimization technique is often used. Dixon et al. (1999) used the simulated annealing method (Kirkpatrick et al., 1983) for the optimization of Eq. (1), and Ouyang et al. (2008) used a genetic algorithm (Holland, 1975; Goldberg, 1989) for the optimization of Eq. (2). Recently, Kao et al. (2012) proposed and applied two linear programming models for several cost functions, including Eqs. (1) and (2). As shown in these earlier studies, the given optimization problem can be approached by any meta-heuristic algorithm. This is a fairly typical application of meta-heuristic optimization methods, in which a search algorithm generates solution candidates, evaluates their performance via the given objective function, and iterates these procedures until the performance criteria are satisfied (mostly global optimal results are difficult to achieve and quasi-optimal results are accepted).

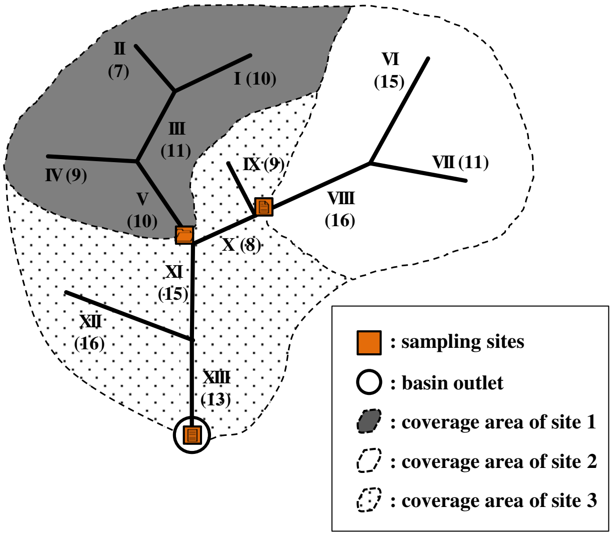

Fig. 1. Hypothetical river network. The Roman numeral next to each reach indicates the reach ID. The Arabic numeral in parentheses is the reach length (no unit is assigned for the length).

Although these typical application practices of meta-heuristic optimization methods can be used for the given problem, herein, we present a notably simple approach to this particular problem. Interestingly, for the given objective functions (Eqs. (1) and (2)), we can dramatically simplify the optimization procedure. The main purpose of this paper is to present this novel idea. Details of the proposed method are given in the next section (Section 2). In Section 3, the proposed method is evaluated using examples covered in earlier studies. Summary and conclusions are given in Section 4.

2. Ideas for efficient problem solving

The key idea underlying the proposed method was derived from the similarity in the functional form between that of marginal entropy (Shannon, 1948) and the given cost functions. Here we show that the search space of the given problem can be substantially reduced by utilizing this similarity. The formulation for the marginal entropy H can be expressed as

$$H = -K\sum_{m=1}^{\beta} p(x_m)\log_2 p(x_m) \tag{3}$$

where $p(x_m)$ is the probability that a variable x_m occurs, β is the number of variables, and K is a constant. The functional similarity between the above formula for marginal entropy and the cost function of Eq. (1) was previously noted by Dixon et al. (1999). We show that both Eq. (1) and Eq. (2) are similar to Eq. (3), and importantly, we explore the potential usefulness of this characteristic.

It is known that H increases as the distribution of $p(x_m)$ approaches the uniform distribution (e.g., Amorocho and Espildora, 1973). Considering the similarity in the functional forms between Eqs. (1) and (3), this fact between H and $p(x_m)$ can be applied to the relationship between E and η_k in Eq. (1). That is, E decreases as the probability distribution of η_k approaches the uniform distribution or in other words, the variance of the group of η_k decreases. This relationship between E and η_k may appear to be the opposite of that between H and $p(x_m)$. However, this is due to the presence of the negative sign in Eq. (3), which is not present in Eq. (1). Therefore, in order to determine an optimal group of monitoring sites that minimizes E, a group of sites that yields minimum variance in η_k should be selected.

Here, we show that the above logic applies not only for minimum-E (Eq. (1)) but also for minimum-F (Eq. (2)). This can be shown based on the relationship that η_k can be expressed as a function of Γ_k in river networks. We demonstrate this functional relationship between η_k and Γ_k using three examples—one natural river network from Sanganmi basin in South Korea, and two theoretical river networks that exhibit ideal self-similar binary tree characteristics (Fig. 2). It is well known that natural river networks exhibit self-similar binary tree topological organization (Horton,