ARTICLE IN PRESS

Marine Pollution Bulletin xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Pollution Bulletin

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

Semivariance analysis and transinformation entropy for optimal redesigning of nutrients monitoring network in San Francisco bay

Amir Boroumand[a],*, Taher Rajaee[a], Fariborz Masoumi[b]

[a] Civil Engineering Dept., Faculty of Engineering, University of Qom, Qom, Iran
[b] Civil Engineering Dept., Faculty of Engineering, University of Mohaghegh Ardabili, Ardabil, Iran

ARTICLE INFO

Keywords:
Optimum distance
Nutrients
San Francisco Bay
Semivariance
Transinformation entropy
Water quality monitoring networks

ABSTRACT

This paper introduces a Semivariance-Transinformation (S-T) based method for designing an optimum bay water nutrients monitoring network in San Francisco bay (S.F. bay), USA. Phosphorus and nitrogen are the most important nutrients that lead to eutrophic condition. The monthly phosphate and nitrate + nitrite data recorded during September 2006 to August 2015 was obtained over 14 active stations located at S.F. bay and was used in the research. Semivariance and discrete transinformation entropy have been applied to calculate the optimum range of the monitoring distance. The study indicated the ranges of 28 to 82 and 37 to 50 km for the phosphate and nitrate + nitrite respectively. Useful information can be obtained from the monitoring network, if the monitoring distance is included in the mentioned intervals. The findings of the research introduce a new approach in the field of water quality monitoring networks design.

1. Introduction

Knowing qualitative and quantitative problems in water resources monitoring systems is one of the most important steps in water resources system management and pollution reduction plans (Boroumand and Rajaee, 2016). In this regard, water resources must be monitored continuously. Difference between required and provided data is one of the most important problems in the monitoring networks. Locations of Sampling stations, time frequencies, qualitative variables specifications, sampling duration and the monitoring objective are the parameters that affect the monitoring results. So monitoring systems should be revised and modified to obtain useful information and prevent cost loss due to high monitoring expenses. In spite of the abilities and investments in the field of water quality monitoring networks, more researches are needed. More studies were done in the field of water resource monitoring systems design but just few cases have been described as below. These studies illustrated that the entropy theory can quantify information content and measure the monitoring networks information.

Harmancioglu et al. (1992) developed a model based on entropy theory to design time frequency and location of sampling stations. The results of their research showed that using this theory is very applicable in qualitative monitoring networks design. Ozkul et al. (2000) followed this study and presented a new method to evaluate and design water quality monitoring networks considering time frequency and sampling locations simultaneously.

Information entropy was used by Krstnovice and Singh (1992) to design rainfall network and also by Goulter and Kusmulyono (1993) to consider of water quality monitoring stations in Australia. These studies on data collection systems showed ability of entropy concept to design optimal monitoring networks. Mogheir and Singh (2002) showed that the distance of groundwater quality monitoring stations is related to transinformation and marginal entropy. Salark and Sorman (2006) optimized and evaluated river stream monitoring network using continuous entropy theory. Masoumi and Kerachian (2008) proposed an entropy-based approach to assess the location of salinity monitoring stations in the Tehran Aquifer. The authors used transinformation entropy to find the optimal distance among stations and showed the applicability and the efficiency of the Entropy in assessing the groundwater monitoring systems. Zhang et al. (2011) and Ridolfi et al. (2011) applied information entropy for rainfall network assessment and obtained the satisfied results in this field. Ridolfi et al. (2012) located monitoring sensors of Dee River basin and then eliminated low efficient stations using entropy theory. Mondal and Singh (2012a, 2012b) used a ratio of transinformation to marginal entropy of rainfall as a measure for assessing natural recharge in unconfined aquifers from Southern India. In another study, they applied entropy theory to evaluate and assess the groundwater monitoring network in Kodaganar River basin of Southern India. They showed that existing monitoring network in Kodaganar River basin leads to excess data and 15 stations can be covered the monitoring network (2012). Fahle et al. (2012) showed

---

* Corresponding author.
E-mail address: a.broumand@stu.qom.ac.ir (A. Boroumand).

http://dx.doi.org/10.1016/j.marpolbul.2017.10.057
Received 26 February 2017; Received in revised form 1 October 2017; Accepted 21 October 2017

Please cite this article as: Boroumand, A., Marine Pollution Bulletin (2017), http://dx.doi.org/10.1016/j.marpolbul.2017.10.057

high potential of entropy theory in evaluating wetlands monitoring systems especially in monitoring and analyzing of surface and groundwater level, which are strongly correlated in wetlands. Lee (2013) used entropy theory in conjunction with genetic algorithm to determine optimal water quality monitoring points in sewer systems. Genetic algorithm has been applied to select the points that maximize the total information among the collected data at multiple locations. Su and You (2014) proposed a spatial information estimation model for the analysis of precipitation gauge networks, to improve previous methods based on information theory. They employed a two-dimensional transinformation–distance relationship in conjunction with multivariate information approximation to estimate transinformation to ungauged locations from existing stations, while taking into consideration the influence of multiple stations and anisotropy. Xu et al. (2015) used an entropy theory based on multi-criteria method to resample the rain gauge networks. Boroumand and Rajaee (2016) applied transinformation entropy in discrete mode for salinity monitoring network design in S.F. bay. They trended two lines to the transinformation-distance data and reported the intersection point of them as optimal distance between monitoring stations for salinity qualitative index.

In this study a Semivariance-Transinformation (S-T) based method has been developed to optimal redesigning of phosphate and nitrate + nitrite monitoring network in bays. When the semivariance is plotted versus distance, semivariogram is produced. Semivariogram is the spatial distribution function which contains the structural information on the variable. Transinformation index measures the redundant or mutual information between two localities. So, both semivariance and transinformation contain the structural information and can be applied to bays monitoring network spatial assessment. Nutrients have a significant impact on bays water quality. An excess amount of nutrients in bay water may lead to low levels of dissolved oxygen and negatively alter various plant life and organisms. Nutrient additions cause a progression of eutrophic symptoms that most often begin with observations of high concentrations of chlorophyll *a* and/or macro algal blooms (Bricker et al., 2008). The need for evaluating the eutrophication status of estuarine and coastal systems, in order to support policy definition, has led to the development of different methods which use symptoms-based multiparameter assessment (Bricker et al., 2003). Well-known example is the United States National Estuarine Eutrophication Assessment (NEEA) (Bricker et al., 1999). Chlorophyll-*a* is the most important indicator that describes the bay's trophic status. Satula et al. (2017) submitted Chlorophyll-a threshold values as a basis to assess eutrophication status of S.F. bay and to inform nutrient management actions. Chlorophyll is the green pigment in plants leaves that allows them to create energy light through photosynthesis. By measuring chlorophyll, the amount of photosynthesizing plants is indirectly measured. In a bay water sample, these plants would be algae or phytoplankton. Chlorophyll is a measure of all green pigments whether they are alive or dead. Chlorophyll-*a* is a measure of the portion of the pigment that is still alive. Sunlight, temperature, nutrients, and wind all affect both algae numbers and Chlorophyll-a concentration (Rajaee and Boroumand, 2015). Nutrients' load of phosphate and nitrate plays a key role in outbreak and growth of algal blooms. In other words, phosphate and nitrate are the indices that control the level of Chlorophyll-a in bays. Management of bays upstream catchment land use, implementation of total maximum daily load plan and or nutrients' load allocation plans (increasing or decreasing the consumption of the fertilizers and chemical pesticides in upstream farms) may control eutrophication at the bays (Tu, 1996). So, with respect to importance of nutrients in bays, the methodology of assessing and optimizing bay water nutrients monitoring network which take into account the values of the semivariance and transinformation was applied in S.F. bay, USA. Based on the author's findings, the S-T methodology is a new approach in the field of monitoring networks assessment and design. The aim of this research is to prevent monitoring systems redundant information and lost data, so using S-T approach may lead to desired results.

2. Study areas and data

2.1. Study area

S.F. bay is a part of the more complex S.F. bay estuary system, which includes San Pablo Bay and Suisun Bay, the Carquinez Strait, the tidal marshes surrounding these waters, and river tributaries. S.F. bay estuary, which consists of 480 mile^2, 12 islands, and two trillion gallons of salt water, can be thought of as two separate areas: the northern, which passes south and westward from the delta through Suisun and San Pablo Bays, and the southern (also called the South Bay) which extends south eastward toward San Jose. These two areas join in the Central Bay near the Golden Gate Bridge and flow out to the Pacific Ocean. The entire bay is relatively shallow, with narrow, deep channels near the Golden Gate.

2.2. Data

In recent decades, human activities have considerably increased the nutrients delivery to many estuarine and coastal areas. Eutrophic conditions, which include low dissolved oxygen concentrations, declining sea grasses and harmful algal blooms, may impact the uses of estuarine and coastal resources by reducing the success of commercial and sport fisheries, fouling swimming beaches, and causing other problems due to the decay of excess amounts of algae (National Research Council, 2000; Duda, 1982). Eutrophic condition has been monitored at the bays, in terms of both temporal and spatial variation. A long-term nutrients monitoring program is running in the S.F bay. There are several reports in this field (e.g. Cloern and Schraga, 2016). The measurements are done monthly at 37 fixed sampling locations spaced 3–6 km apart (Fig. 1). Fourteen stations have been illustrated in red color on the figure. They are measured more completely than the others. So in this research, the data is obtained from these stations. Data is available from 1969 but the measurements before 2006 were not complete. So, this study used phosphate and nitrate + nitrite records consisted of data from September 2006 to August 2015 were obtained from the United State Geological Survey (USGS) website (http://sfbay.wr.usgs.gov/access/wqdata). The sampling depth is 2 m below water surface.

3. Method

3.1. Entropy theory

The entropy of a random variable is a measure of the information or uncertainty associated with it (Mogheir et al., 2004). The entropy theory is a method of quantifying information and even controlling existing data sufficiency. Chaos and turbulence in a system can be measured using entropy theory either. Turbulence existing in a data set means many disordered variations, which can generate repeated data that is so costly and redundant. Information fluctuation and not to following a specific rule will lead to reduction in our knowledge from system and make uncertainties. Entropy theory provides a quantitative measure of the uncertainty or the information content of a random variable (Shannon, 1948). The entropy indices are: the marginal entropy, conditional entropy, joint entropy and transinformation. The indices are defined as follows for the discrete random variables $X = \{x_i\}, i = 1, 2, \ldots, n$ and $Y = \{y_j\}, j = 1, 2, \ldots, m$ (e.g. Mogheir and Singh, 2002 and Vivekanandan, 2014).

3.1.1. Marginal entropy

Marginal entropy for the discrete random variable X is defined as:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \ln(P(x_i)) \tag{1}$$

where, $P(x_i)$ is the probability of i_{th} random variable $X = \{x_i\}$, and n is the number of observations. The total of probability values in the scope