Environmental Research 161 (2018) 61–75

Contents lists available at ScienceDirect

Environmental Research

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

A kriging and entropy-based approach to raingauge network design

Pengcheng Xu[a], Dong Wang[a,*], Vijay P. Singh[b], Yuankun Wang[a], Jichun Wu[a], Lachun Wang[c], Xinqing Zou[c], Jiufu Liu[d], Ying Zou[d], Ruimin He[e]

[a] Key Laboratory of Surficial Geochemistry, Ministry of Education, Department of Hydrosciences, School of Earth Sciences and Engineering, State Key Laboratory of Pollution Control and Resource Reuse, Nanjing University, Nanjing, PR China
[b] Department of Biological and Agricultural Engineering, Zachry Department of Civil Engineering, Texas A & M University, College Station, TX 77843, USA
[c] School of Geographic and Oceanographic Science, Nanjing University, Nanjing, PR China
[d] Nanjing Hydraulic Research Institute, Nanjing, PR China
[e] State Key Laboratory of Hydrology, Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, PR China

ARTICLE INFO

Keywords:
Network density
Kriging method
Entropy
NI-KSE based criterion
Optimal network design

ABSTRACT

Hydrological data, such as precipitation, is fundamental for planning, designing, developing, and managing water resource projects as well as for hydrologic research. An optimal raingauge network leads to more accurate estimates of mean or point precipitation at any site over the watershed. Some studies in the past have suggested increasing gauge network density for reducing the estimation error. However, more stations mean more cost of installation and monitoring. This study proposes an approach on the basis of kriging and entropy theory to determine an optimal network design in the city of Shanghai, China. Unlike the past studies using kriging interpolation and entropy theory for network design, the approach developed in the current study not only used the kriging method as an interpolator to determine rainfall data at ungauged locations but also incorporated the minimum kriging standard error (KSE) and maximum net information (NI) content. The approach would thus lead to an optimal network and would enable the reduction of kriging standard error of precipitation estimates throughout the watershed and achieve an optimum rainfall information. This study also proposed an NI-KSE-based criterion which is dependent on a single-objective optimization. To evaluate the final optimal gauge network, areal average rainfall was estimated and its accuracy was compared with that obtained with the existing rain gauge network.

1. Introduction

Hydrological stations provide basic data on precipitation, water quality, air quality, soil moisture, infiltration, streamflow, and groundwater. An optimally designed network should provide needed information for use in a variety of fields, such as agriculture, hydrometeorology, water resources, environmental management, ecology, and hydrogeology. Other uses of data include conflict resolution, climate change impact assessment, impact of land use-land cover change, risk assessment, and calibration and verification of mathematical models. The data should be accurate and spatially representative. For example, planning hydraulic facilities for flood prevention economically requires rainfall data to be obtained from an optimal rain gauge network, which would minimize hydrological and economic risks. The best raingauge network should indicate the spatial and temporal characteristics of the network structure precisely and is able for better rainfall estimates with an appropriate number of stations in the watershed (Mishra and Coulibaly, 2009). Thus, the goal of designing a raingauge network is to identify and select the best station combination with adaptive number and location (Basalirwa et al., 1993), so that it can economically provide optimal information content (Pardo-Iguzquiza, 1998).

The widely used approaches for network assessment and design can be summarized as: (1) statistically based, (2) spatial interpolation, (3) information theory-based, and (4) hybrid. For over half a century, statistical regression techniques have been applied widely to locate gauging stations (Moss and Tasker, 1991). The most representative statistical method is the generalized least squares (GLS) method which can lead to effective regional networks.

Geostatistical interpolation methods have been employed in the past three decades using either point-based or areally-based values, taking covariance into account. The technique of spatial interpolation allows to predict data at any site over the watershed (Morrissey et al., 1995; Camera et al., 2014; Bechler et al., 2015; Kilibarda et al., 2014; Manz et al., 2016). The kriging method, a commonly used spatial interpolation method, can perform nonlinearly due to different spatial

* Corresponding author.
E-mail address: wangdong@nju.edu.cn (D. Wang).

http://dx.doi.org/10.1016/j.envres.2017.10.038
Received 9 August 2017; Received in revised form 13 September 2017; Accepted 10 October 2017

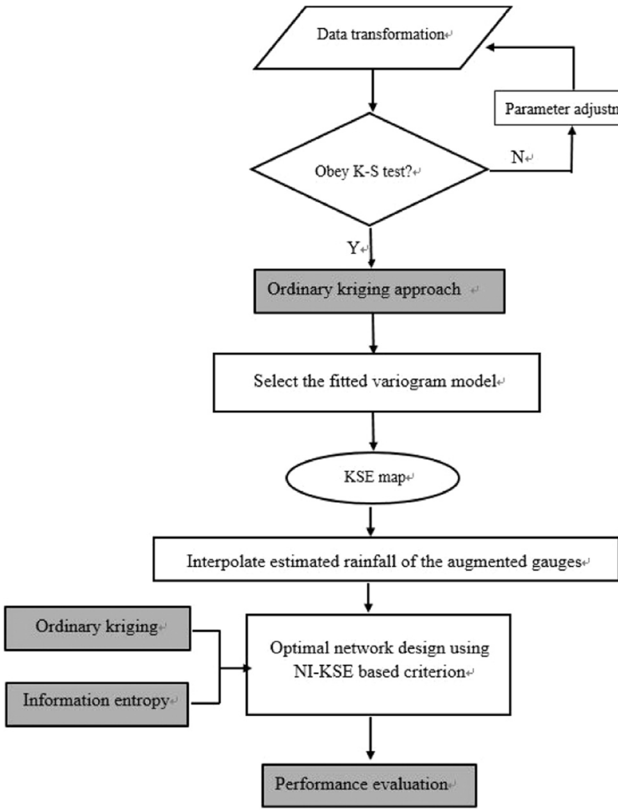

Fig. 1. Schematic overview of the gauge network design process.

coordinates (Biau et al., 2010; Wackernagel, 2003). It is usually composed of two parts: spatial variation analysis using variogram and calculation of distance weights for interpolation. A relatively adaptive network should a systematically search for the best station configuration by enhancing the estimation accuracy. Specific information on kriging-based method is included in (Webster and Oliver, 2007; Goovaerts, 2000; Fernández-Casal et al., 2003; Li and Shao, 2010; Adhikary et al., 2015; Liang et al., 2014; Zhang et al., 2014; Wang and Lin, 2015; Peng et al., 2016; Rousis et al., 2017). Sivakumar and Woldemeskel (2014) contend that the network related approach is useful to verify the relationship between hydrological data and that the approach has good potential for interpolation, extrapolation and prediction.

The entropy theory has been widely used in the hydrometeorology and water resources literature since the 1970s (Singh, 1997). The fundamental basis in gauge network design using the information entropy is that the gaging stations should have as little overlapping information (or mutual information) as possible, which means that all the gauges should be made as independent from each other as possible. A frequently used measure is mutual information, which quantifies the amount of information of one random variable that is contained in another random variable (Cover and Thomas, 1991). Krstanovic and Singh (1992a, 1992b) evaluated the rainfall network in Louisiana using joint entropy and transinformation for which multivariate distributions were determined using the principle of maximum entropy (POME). Mogheir et al. (2006) made an objective assessment of a ground water quality monitoring network using the entropy concept in the Gaza Strip, Palestine. Mishra and Coulibaly (2010) investigated the streamflow network design over different Canadian catchments by three entropy terms (i.e. marginal entropy, joint entropy and transinformation). Alfonso et al. (2010) showed that the entropy theory was capable of achieving the optimum design of water level monitors across a well-conceived and supervised polder system in the Netherlands. They found the total correlation in multivariate cases as a core to evaluate the performance of mutual dependence regulation. Li et al. (2012) developed a maximum information minimum redundancy criterion (MIMR) for hydrometric assessment and planning. The above studies shows that mutual information has been usually used for quantifying the common information among all the stations of the system.

A hybrid method attempts to combine the advantages of single methods or use e the output of one method as the initial condition for the other method, thus making the network more reasonable. Pardo-Iguzquiza (1998) presented an approach to build an optimal station configuration by achieving an accurate estimation of areal mean value of precipitation. This approach aims to satisfy two objective functions: accurate estimation of areal mean and economic data collection. In order to reduce the kriging variance, the accuracy of mean rainfall estimation is realized in the form of simulated alleviation. The process of simulated alleviation is an algorithm of mathematical optimization that helps achieve the best strategy for network configuration by decreasing the error and considering another target function such as economic limitation. It may be noted that the proposed method also combines the accuracy of rainfall estimation (i.e. the kriging standard error reduction) and the optimal information content for the gauge network. Some studies integrated the entropy theory with other approaches, such as the kriging method (Chen et al., 2008) and multidimensional variable analysis (Shaghaghian and Abedini, 2013). Yeh et al. (2011) established a model combining interpolation with entropy for rain gauge network design. For optimizing the gauge network with greater transmitted information at the minimum number of rain gauges, a probability distribution function for measuring the information intensity is often assumed. However, in this study, the kriging method was solely applied to predict, by spatial interpolation, the precipitation data at ungauged locations where augmented gauges would be fixed, while the information entropy was employed to calculate information intensity corresponding to each station. These two methods were actually used for different objectives. The entropy approach was only used to prioritize stations (Adhikary et al., 2015). That is, the optimum station configuration cannot be realized by reducing the kriging variance and maximizing the information content at the same time.

The objective of this study therefore was to develop a kriging and entropy-based approach for designing an optimal rain gauge network which can reduce the kriging variance for either areal or point rain estimates across the watershed (Pardo-Iguzquiza, 1998) and at the same time achieve the optimum rainfall information by optimally redesigning (discontinuing) some overlapping stations as well as optimizing the installation of augmented gauges. The major contribution of this study is that unlike the study of Yeh et al. (2011), the developed methodology not only uses the kriging method as an interpolator to determine rainfall data at ungauged stations but also determines the best network configuration considering the minimum kriging standard error (KSE) and the maximum net information (NI). Inspired by the optimum regulation, named maximum information minimum redundancy (MIMR) proposed by Li et al. (2012), this study proposes an NI-KSE based criterion which is also dependent on a single-objective optimization.

The paper is organized in 4 sections. In Section 2, kriging interpolation and entropy are introduced. The framework of the proposed approach is also included in this section. Section 3 presents the methodology for determining optimal gauge network for the city of Shanghai, China. Results are summarized in Section 4 and finally the conclusions are stated.

2. Methodology

The methodology developed for rain gauge network design was based on kriging and entropy and was composed of five main steps or phases (see Fig. 1): (1) Select the best theoretical variogram model for each scenario of datasets based on the existing gauge network; (2) decide the number and locations of additional gauges according to the