Raingauge network optimization in a tropical urban area by coupling cross-validation with the geostatistical technique

Mohd Zaharifudin Muhamad Ali and Faridah Othman

Department of Civil Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur, Malaysia

ABSTRACT
An adequate and reliable raingauge network is essential for observing rainfall data in hydrology and water resource applications. A raingauge network developed for a catchment area is commonly extended periodically to increase data accuracy. Due to financial constraints, the network is reviewed for the optimal number of stations. A new optimization approach is developed in this study by coupling a cross-validation technique with a geostatistical method for raingauge network optimization to prioritize raingauge stations. The spatial interpolation error of the spatial rainfall distribution, measured as the root mean square error ($E_{\text{rms}}$) optimization criterion is applied to a raingauge network in a tropical urban area. The results indicate that this method can successfully optimize the number of rainfall stations in an existing raingauge network, as the stations are prioritized based on their importance in the network.

ARTICLE HISTORY
Received 30 August 2016
Accepted 15 November 2017

EDITOR
D. Koutsoyiannis

ASSOCIATE EDITOR
E. Volpi

KEYWORDS
raingauge network; cross-validation; geostatistical analysis; optimization; root mean square error

Introduction
A raingauge network is a hydrometric network meant to collect rainfall data and facilitate hydrology applications, such as hydrological modelling (Xu et al. 2015), flood forecasting (Kar et al. 2015), and flash flood prediction (Volkman et al. 2010) as well as water resource analysis. Rainfall data from a raingauge network are used to compute spatial rainfall information in terms of areal average, point-based and spatial variability. Accurate rainfall data are very important for hydrology and water resource-related projects at the planning, design and operational levels (Adhikary et al. 2015).

The rainfall data from raingauge networks involve point-based rainfall that is used to compute areal average rainfall for spatial rainfall mapping for instance. The map is produced through an interpolation process using point rainfall values. Accurate rainfall maps are extremely important for any hydrology application. The spatial rainfall interpolation error, measured as root mean square error ($E_{\text{rms}}$) is commonly used as an accuracy indicator. For example, Ali and Othman (2017) employed $E_{\text{rms}}$ as an accuracy indicator to study the best variogram to produce an accurate spatial rainfall map. The spatial interpolation error should be as small as possible, which is achievable with an appropriate number of point rainfall values. The development of a raingauge network is an evolutionary process, beginning with the initial development of a basic network, followed by periodic reviews for upgrading to achieve an optimum network (Vivekanandan and Jagtap 2013). River basin managers around the world adopt this process in raingauge network design and optimization. A raingauge network is reviewed to optimize the appropriate number of point rainfall stations, as studied by Bastin et al. (1984) and Pardo-Iguzquiza (1998). The review process may be based on the procedure suggested by the World Meteorological Organization (WMO), as illustrated in Figure 1.

Earlier research on raingauge network assessment was conducted using classical methods, such as statistical and probabilistic approaches. Nemec and Askew (1986) explained the philosophy of hydrological network design using statistical moments of mean and variance. Şorman and Balkan (1983) applied the same statistical moments to redesign the raingauge network in the Kizilirnmak River basin in Turkey. However, limitations with the statistical ability to explain precise rainfall data have encouraged the application of probability theory.

A probabilistic approach called the entropy method has also been used to design hydrological station networks. This method is also known as the Shannon entropy (Shannon and Weaver 1949) and can be utilized to model system information through transmitting and receiving information as entropy values. The probability distribution logarithm serves to measure the entropy value. According to the literature, this