Article:

Dai, Q., Bray, M., Zhuo, L. orcid.org/0000-0002-5719-5342 et al. (2 more authors) (2017) A scheme for rain gauge network design based on remotely sensed rainfall measurements. Journal of Hydrometeorology, 18 (2). pp. 363-379. ISSN 1525-755X

https://doi.org/10.1175/jhm-d-16-0136.1

A Scheme for Rain Gauge Network Design Based on Remotely Sensed Rainfall Measurements

QIANG DAI

Key Laboratory of Virtual Geographic Environment of Ministry of Education, Nanjing Normal University, Nanjing, China, and Water and Environmental Management Research Centre, Department of Civil Engineering, University of Bristol, Bristol, United Kingdom, and Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China

MICHAELA BRAY

Hydro-Environmental Research Center, Cardiff University, Cardiff, United Kingdom

LU ZHUO

Water and Environmental Management Research Centre, Department of Civil Engineering, University of Bristol, Bristol, United Kingdom

TANVIR ISLAM

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

DAWEI HAN

Water and Environmental Management Research Centre, Department of Civil Engineering, University of Bristol, Bristol, United Kingdom

(Manuscript received 16 June 2016, in final form 12 September 2016)

ABSTRACT

A remarkable decline in the number of rain gauges is being faced in many areas of the world, as a compromise to the expensive cost of operating and maintaining rain gauges. The question of how to effectively deploy new or remove current rain gauges in order to create optimal rainfall information is becoming more and more important. On the other hand, larger-scaled, remotely sensed rainfall measurements, although poorer quality compared with traditional rain gauge rainfall measurements, provide an insight into the local storm characteristics, which are sought by traditional methods for designing a rain gauge network. Based on these facts, this study proposes a new methodology for rain gauge network design using remotely sensed rainfall datasets that aims to explore how many gauges are essential and where they should be placed. Principal component analysis (PCA) is used to analyze the redundancy of the radar grid network and to determine the number of rain gauges while the potential locations are determined by cluster analysis (CA) selection. The proposed methodology has been performed on 373 different storm events measured by a weather radar grid network and compared against an existing dense rain gauge network in southwestern England. Because of the simple structure, the proposed scheme could be easily implemented in other study areas. This study provides a new insight into rain gauge network design that is also a preliminary attempt to use remotely sensed data to solve the traditional rain gauge problems.

1. Introduction

Rain is a major component of the water cycle on Earth, which is a significant issue in many scientific fields such as ecosystems, agriculture, and water environment. Methods for measuring rainfall need to take into account its mutability in order to minimize uncertainty and the errors found within the recorded data. The use of rain gauges is one of the oldest and most common methods employed in the world for measuring rainfall. Nowadays, rain gauge rainfall is a vital source of information used for the calibration of remotely sensed rainfall and verification of numerical weather model

Corresponding author e-mail: Qiang Dai, q.dai@njnu.edu.cn; dqgis@hotmail.com

DOI: 10.1175/JHM-D-16-0136.1

rainfall products. However, hydrologists and meteorologists are facing a dilemma of rain gauge management. Additionally, rain gauges are extremely desired in many regions to improve the quality of local rainfall measurement, especially in ungauged catchments, while on the other hand, the expensive cost of operating and maintaining rain gauges leads to shutting down of available rain gauges. A decline of roughly 50% of the number of rain gauges has occurred in the period of 1989–2006 in Europe, South America, and Africa (Lorenz and Kunstmann 2012; Walsh 2012), and there has been an approximately 50% decrease in the number of valid daily reports in the period 2000–07 for APHRODITE (Overeem et al. 2013; Yatagai et al. 2012). In an effort to address this dilemma, an efficient and feasible scheme for rain gauge network design is needed to capture maximum rainfall information with a minimum number of gauges.

The traditional rain gauge network design is classified by two types: haphazard manner and quantitative method. The haphazard manner is generally based on numerous technical guidelines or considerations. At present, there is still no standard procedure in place in most parts of the world for rain gauge network design. The reason for this is because of the complexity of the problem faced by both hydrologists and meteorologists. Rain gauges were distributed according to the population, in order to be close to observers. This led to areas of high rainfall having relatively few gauges. Design requirements consist of determining the number of gauges and their locations, given the frequency in time of sampling that minimizes the uncertainty of rainfall estimation. Other factors also need consideration, that is, the nature of the catchment, its topographic influences, its drainage patterns, the accessibility and suitability of proposed locations, and the cost of installing and maintaining the gauges. Moreover, the purpose of the network and the regional climate should be taken into consideration. The high variability of rainfall over time and space is a significant issue and is difficult to address; to typify rainfall patterns of high variability and intermittency, a much denser network would be needed (Barancourt et al. 1992; Rodríguez-Iturbe and Mejía 1974).

Rain gauge network design based on quantitative analysis attracts more attention. There are many methods employed, such as spatial correlation, variogram analysis, and entropy theory (Al-Zahrani and Husain 1998; Bastin et al. 1984; Bogárdi et al. 1985; Bradley et al. 2002; Bras and Rodríguez-Iturbe 1985; Krstanovic and Singh 1992; Mishra and Coulibaly 2009; Moore et al. 2000; Pardo-Igúzquiza 1998; Tsintikidis et al. 2011; Volkmann et al. 2010; Yang and Burn 1994). Statistical techniques such as variance reduction algorithm, state-space stochastic models, and generalized least squares are also adopted in numerous studies (Bradley et al. 2002; Morrissey et al. 1995; Moss and Tasker 1991; Shih 1982; Stedinger and Tasker 1985; Tasker and Moss 1979). The more sophisticated techniques of rain gauge network design can provide some insight into the location of rain gauges as well as the density of the network. These methods are generally based on analyzing the available limited rain gauge information or duplicating the learned knowledge from a mature network to a pioneering area. The nature of these methods makes them difficult to implement and often requires subjective parameter adjustments. Another shortcoming of most of the aforementioned methods is that they require exhausting all possible candidate networks to explore the optimum one.

Compared with the traditional methods that collect and excavate the limited rainfall-related information (such as rainfall spatial correlation and rainfall patterns) from inside or outside a study area, adoption of remotely sensed rainfall measurements for rain gauge network design is obviously a more direct and efficient way. Currently, remotely sensed rainfall estimates are available in most parts of the world, as some satellites have global rainfall observing ability, such as the Tropical Rainfall Measuring Mission (TRMM) and the Global Precipitation Measurement (GPM). Weather radars are also extensively installed around the world, especially in developed regions. Because of the poorer quality, these rainfall products cannot replace rain gauge measurements at present, and rain gauges will still be the first choice in most hydrological applications in the near future. However, these relatively inaccurate but larger-scaled rainfall measurements are an ideal dataset to provide insight into the local storm characteristics, such as rainfall patterns and local topographic influences on rainfall, which are the key components that the traditional rain gauge network-designed methods eagerly expect to explore. Through analyzing the long-term remotely sensed rainfall dataset, we can reveal these characteristics and consequently investigate the most important locations for deploying rain gauges in the study area. For this reason, this study presents a new approach to rain gauge network design based on a combination of principal component analysis (PCA) and variable selection criteria using a weather radar rainfall dataset to provide both the optimum rain gauge density and rain gauge location. This scheme offers a new insight into rain gauge network design and could be easily adopted and implemented in other study areas.

This paper is organized as follows. After the introduction, section 2 illustrates the study area and

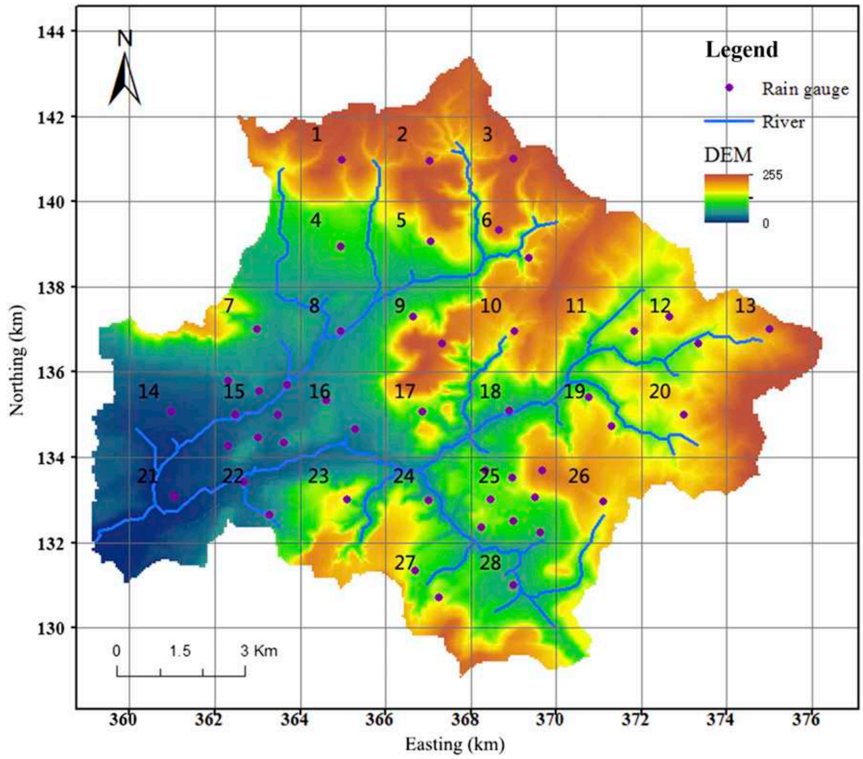

FIG. 1. The Brue catchment terrain map. The purple dots represent the rain gauges and the grid represents the real radar grids. The labeled number is the index of the corresponding grid.

datasets used. Section 3 describes the algorithms of determining the redundancy of networks and optimum locations of gauges. Section 4 presents the outcomes of the rain gauge network design and evaluations of the proposed method. Section 5 elaborates on three key issues associated with this scheme, and section 6 summarizes the key findings and the future work.

2. Study area and data source

The Brue catchment in Somerset, southwestern England (51.08°N, 2.58°W), covering an area of 135 km^2 to its river gauging station at Lovington, is chosen as a case study for the investigations carried out in this work. It is mainly pastureland with some areas of woodland in the higher eastern half of the domain. The choice of catchment is based on the availability of quality data; furthermore, the characteristics of the Brue catchment are considered to be representative of rural U.K. catchments used for rainfall–runoff modeling. With 6 years of continuous data provided by a dense 49-rain-gauge network (see Fig. 1), the Brue catchment provides an ideal study area for the analysis of rain gauge network design. The designed scheme can be carefully evaluated with the large number of available rain gauges. Operationally, most catchments are only serviced by two rain gauges at best. The cost of maintaining a denser network is too prohibitive for it to be feasible on a larger scale. Clearly, the rich dataset provided by a dense rain gauge network is rare and contains information on rainfall events that would normally be missed by catchments with just one or two rain gauges.

Radar and rain gauge datasets were maintained by the National Rivers Authority as part of the Hydrological Radar Experiment (HYREX). The dense rain gauge network, radar, and a variety of related meteorological data are available through the British Atmospheric Data Centre (BADC). The radar datasets are from the Wardon Hill radar, located at a range around 40 km from the center of the catchment. The radar completes one cycle through four different scan elevations every 5 min, and the rainfall intensity is recorded on two Cartesian grids: 5 and 2 km. The rain gauges installed on the Brue catchment are typical of those used by the U.K. Environment Agency: a Casella tipping-bucket gauge mounted vertically on a concrete paving slab. The bucket size was 0.2 mm and the gauge aperture was 400 cm^2. The tip time was recorded up to a time resolution of 10 s. The first

TABLE 1. Ten typical rainfall events measured over the Brue catchment. The averaged rainfall represents the accumulated radar rainfall for an event and areal-averaged rainfall in the catchment.

Event ID	Start date	End date	Averaged rainfall (mm)
1	0500 UTC 5 Oct 1993	2200 UTC 5 Oct 1993	126.39
2	0200 UTC 30 Dec 1993	1800 UTC 30 Dec 1993	153.45
3	1400 UTC 8 Nov 1994	2200 UTC 9 Nov 1994	334.51
4	0800 UTC 27 Jan 1995	0800 UTC 28 Jan 1995	230.02
5	1100 UTC 16 Jul 1995	0300 UTC 17 Jul 1995	150.68
6	1400 UTC 26 Nov 1995	1000 UTC 27 Nov 1995	115.41
7	2000 UTC 11 Feb 1996	1400 UTC 12 Feb 1996	167.29
8	0400 UTC 24 Oct 1998	0000 UTC 25 Oct 1998	230.53
9	1200 UTC 18 Dec 1999	0100 UTC 19 Dec 1999	260.52
10	0500 UTC 18 Apr 2000	0400 UTC 19 Apr 2000	180.65

valid day was considered to be the day after which the gauge recorded its first value (Wood et al. 2000). The dense rain gauge network was designed so that all the rain gauges would lie entirely in the catchment and that there would be at least one rain gauge in the center of each 2-km radar grid square.

Rain gauge network design is carried out for 373 rainfall events from September 1993 to April 2000. Except for some events where either radar or gauge data are missing, these events cover almost all significant storms during the period. To display the outcomes, 10 demonstrated typical events are chosen from them. The event identifier (ID), durations, and averaged rainfall over the catchment are listed in Table 1. The events cover a wide range of scenarios; some of the rainfall is consistent over several hours (events 3 and 4), and some events are short lived but intense (events 2 and 9). Averaged rainfall represents the areal-averaged rainfall over a catchment with accumulated rainfall for an event. It is calculated using radar rainfall measurements. With possible systematic errors, the radar rainfall estimates tend to be larger than gauge rainfall values in the Brue catchment. Discussion of radar rainfall adjustment and uncertainty analysis is outside of the scope of this work. Interested readers can refer to our previous papers for more information (Dai and Han 2014; Dai et al. 2014).

3. Methodology

a. PCA application for radar grid network redundancy analysis

PCA is used first to examine the redundancy existing in the dense radar grid network and second to isolate the grids that provide the most significant contribution to the principal components. The selected radar grids from the existing radar grid network are considered to be ideal places for deploying rain gauges, which are named optimum grids (OGs). The center of each OG indicates one potential location for a rain gauge. The appropriate number and location of OGs are dependent upon the amount of original variance the network should retain. PCA is a technique used in multivariate analysis where it is suspected that a number of variables are interrelated. It is primarily used for compressing a dataset while at the same time minimizing the loss of information. This is achieved by generating a new set of variables called principal components (PCs). In this study, redundancy analysis of existing radar grid network based on PCA is used to determine the number of OGs.

Given a dataset of n variables (i.e., the number of existing radar grids in the study area) with p observations (i.e., the number of hours, days, months, etc. observed), the $n \times n$ covariance matrix $\mathbf{C}$ of the dataset is first calculated. For PCA to work properly, the original dataset is normalized by subtracting the mean from each of the data dimensions. Since the eigenvectors of the covariance matrix $\mathbf{C}$ are orthogonal, the n eigenvectors can be used as a basis from which the principle components are built, which is shown as follows:

$$\text{eigenvector} = \left(\text{eig}_1 \quad \text{eig}_2 \quad \text{eig}_3 \quad \cdots \quad \text{eig}_n\right). \tag{1}$$

Thus, the original data can be represented in terms of the n eigenvectors via a linear transformation from the original dataset $\mathbf{X}$ to a new dataset $\mathbf{Z}$, where the variance of each of the components is its corresponding eigenvalue:

$$\mathbf{Z}_i = X_1\text{eig}_{i,1} + X_2\text{eig}_{i,2} + \cdots + X_n\text{eig}_{i,n}, \quad i = 1, \ldots, n, \tag{2}$$

where $\mathbf{X}_i$ and $\mathbf{Z}_i$ are the vectors of original and new datasets, respectively. These new variables do not contain any redundant information since each PC is a linear combination of the original variables and all PCs are orthogonal to each other (Jolliffe 1986).

Since the eigenvalue of each component is also its variance, the eigenvector with the highest eigenvalue is the PC of the dataset; this property is used to determine the PCs that carry a set percentage of the variance found in the data. Assessment of the network redundancy is achieved by deciding on a threshold of desired variance explained and then noting how many PCs are required to obtain this threshold. If the threshold is q, then we wish to maintain q (%) of variance found in the data, and correspondingly, we accept a loss of $(1 - q)$ (%) of the original information contained by original radar grids. Thus, the number of principal components (say, k) required to give us q (%) of explained variance is the number of optimum grids. If k is much less than n, then we can conclude that the network is heavily redundant and many of the radar grids can be removed without significant loss of information. On the contrary, if k is quite close to n, then there is little redundancy in the network. The choice of the variance threshold is based on two conditions. Basically, the required number k obviously grows with the increase of variance threshold. The growth rate will change when the threshold reaches a critical value, which could be regarded as a threshold value. In addition, some subjective factors are also worth considering. For example, the budget-saving-prone or information-remaining-prone designers may have different claims of threshold value under the same condition. A detailed discussion of threshold selection is given in section 4c.

b. *Selection criteria for determining optimum rain gauge locations*

The PCs are just linear data combinations of all original variables (radar grids rainfalls), so it is necessary to interpret the components in terms of the original variables to select the OGs. In effect, we can try to choose a subset of the original variables that approximate the information retained in k PCs. In this study, cluster analysis (CA) is used to allocate the original variables to a subset of clusters derived from the average linkage method. One variable is retained from each cluster and is chosen as the representative of that cluster. The advantage of CA is that there is no prior knowledge about which elements belong to which clusters. With CA, we can select the OGs and consequently determine the possible locations for rain gauges.

Formal definition of a cluster, group, or class is difficult and is often down to the judgment of the user. Cormack (1971) talks of internal cohesion and external isolation in defining clusters. Although there is no standard definition of a cluster, it is generally felt that it must have something to do with the recognition of relative distance between members, so certain properties are attributed to clusters, such as density, variance, shape, and separation, which are formed by assessing the similarity and dissimilarity of each pair of members to be clustered. The proximity of each member to the other can be measured in many ways depending on the type of variables under investigation. As we already have concluded, in the number of clusters (k as mentioned above) based on PCA, k-means clustering is used to partition original n variables into k clusters, in which each variable belongs to the cluster with the nearest mean, which uses an iterative algorithm to minimize the sum of distances from each object to its cluster centroid, over all clusters. Cluster analysis for selecting OGs works with five main steps: 1) initialize the centroid of the clusters; 2) attribute the closest cluster to each radar grid; 3) set the centroid position of each cluster to the mean of all radar grids belonging to that cluster; 4) repeat steps 2 and 3 until it converges (each centroid stays in a stable location or calculated time reaches the maximum iteration) to form a given number of clusters; and 5) the OG selected to represent each cluster is chosen in two different ways, the radar grid giving the maximum event-averaged rainfall in a given cluster (CA Max) and the radar grid giving the median of the event-averaged rainfall (CA Med). Cluster analysis can measure distances that are Euclidean (can be measured with a ruler) or distances based on similarity. The Euclidean distance is the most straightforward and generally acceptable way of computing distances between objects in a multidimensional space. Moreover, the rainfall connection among different radar grids relies on their separated distances. So we adopt the Euclidean distance measure in the proposed scheme. Thus, each cluster has one OG, and the given number of OGs makes up a new compact but high-information engaged radar grid network, which brings out a new rain gauge network. A flowchart of the proposed method is shown in Fig. 2.

c. *Validation methods*

Performance evaluation of rain gauge network design is a challenge because no standard assessment criteria exist to indicate what kind of network is the most appropriate one for the given study area. In essence, the designed network should be an effective network, which means the network should contain maximum information at the cost of a minimal number of gauges. This principle can be interpreted by two major rules for this study. First, the selected small number of OGs should maintain the dominating rainfall information of the original radar grid network. Second, further increase in the number of OGs should not significantly increase the amount of rainfall information.

The second rule can be achieved through use of an information-component curve, which will be discussed in section 4c. For the first rule, two indicators (the Pearson correlation coefficient and Nash–Sutcliffe coefficient) are introduced herein. The Pearson correlation coefficient r can estimate the systematic deviation

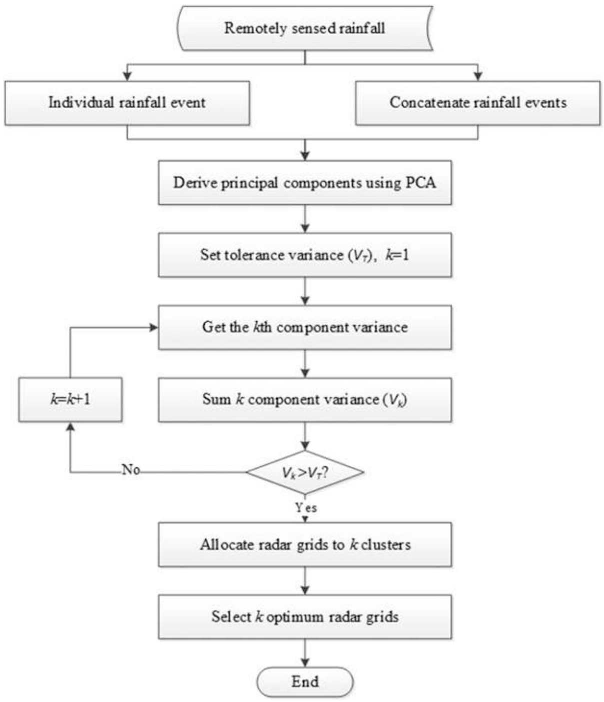

FIG. 2. Flowchart of the proposed method.

between OG rainfall R_m and the original radar grid rainfall R_o, which is written as

$$r_{R_o,R_m} = \frac{E(R_mR_o) - E(R_m)E(R_o)}{\sqrt{[E(R_m^2) - E(R_m)^2] \times [E(R_o^2) - E(R_o)^2]}}, \tag{3}$$

where E calculates the mean value of the corresponding vector. The Nash–Sutcliffe coefficient (Nash and Sutcliffe 1970) is generally used to assess how well hydrological models predict events. The rainfalls of the designed network and existing network are regarded as modeled and observed values, respectively.

The Nash–Sutcliffe coefficient (NS) is calculated as follows:

$$\mathrm{NS} = 1 - \frac{\sum(R_o^t - R_m^t)^2}{\sum[R_o^t - E(R_o)]^2}, \tag{4}$$

where the superscript t refers to the time step of the storm and NS $\in [1, -\infty)$. The closer NS is to 1, the more accurate the designed scheme is.

In addition, because radar measurements have nonnegligible uncertainty, we speculate that this uncertainty could be propagated to the design processing and contaminate the designed network. Therefore, in the second stage of validation, the less accurate radar dataset was compared to the rain gauges dataset after implementing the proposed scheme. An existing dense and high-redundancy rain gauge network (with 49 rain gauges allocated within an area of 135 km^2) was used as a reference network. The proposed scheme was carried out using the rain gauge network and radar grid network, and the corresponding gauge- and radar-designed networks were produced. By comparing the differences of gauge numbers and locations between the gauge- and radar-designed networks, we can investigate the possible errors of the radar-designed network caused by radar rainfall uncertainty.

The mean number error E_N and the mean location error E_L are defined to quantitatively describe the designed errors. The mean number error represents the averaged deviation of sizes between gauge- and radar-designed networks, which is derived as

$$E_N = \frac{1}{\mathrm{VN}} \sum_{v=1}^{\mathrm{VN}} |N_o^v - N_m^v|, \tag{5}$$

where N_o^v and N_m^v refer to the number of gauges in the gauge- and radar-designed networks for the given variance threshold v. To avoid the contingency of performance evaluation due to using a certain variance threshold, dozens of threshold values are used to design rain gauge networks. Variable VN is the number of involved variance thresholds.

The mean location error is used to illustrate the disparity in space between two designed networks. The selected radar grids and selected rain gauges are first paired and the distances between each pair are accumulated. The combined scheme with the least accumulated distance is adopted and the distance is defined as E_L. If the number of radar grids is smaller than that of rain gauges, one radar grid may correspond to two rain gauges and vice versa. As the difference of numbers between them is quite small, the error that may be introduced is negligible. The radar-designed network is explicitly optimal if its gauge locations are exactly the same as those of the gauge-designed network. In such cases, E_L is equal to zero; E_L is calculated for each variance threshold.

4. Model computations and results

a. Radar grid network redundancy

With 28 radar grids located within an area of 135 km^2, a redundancy of rainfall should exist in the radar grid network. To show this, the correlation matrices of 28 radar grids are calculated and drawn in Fig. 3 for events 1 and 2. The Pearson correlation coefficients are computed for every two radar grids. Almost all correlation coefficients in both events are larger than 0.5. For event 2, the coefficients are generally even larger than 0.8. The rainfall patterns of the last two radar grids are relatively

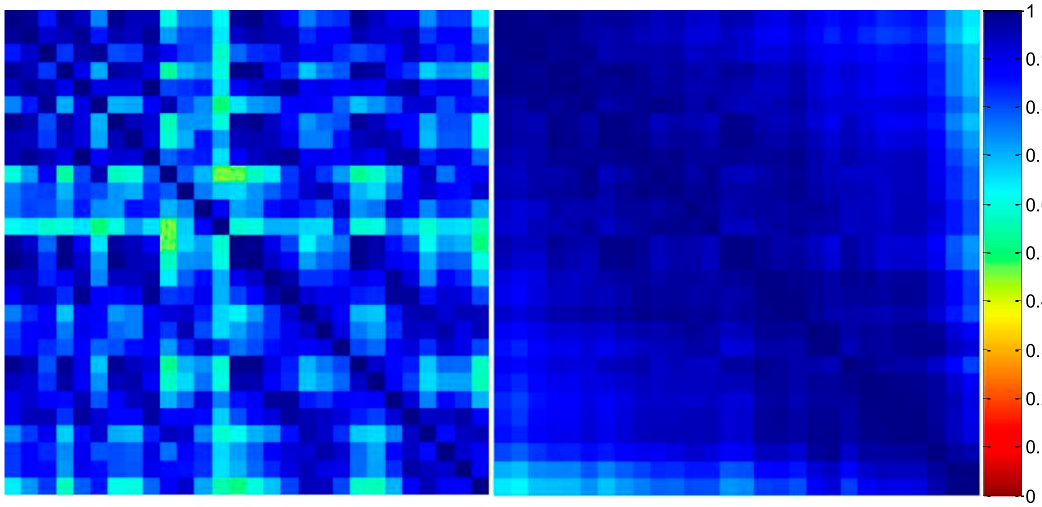

FIG. 3. Correlation matrices for events (left) 1 and (right) 2, with each pixel representing a correlation coefficient between the grid indexed from left to right and the grid indexed from up to down.

exceptional compared with other grids in event 2. This is explicable as the last two radar grids are located in the southern boundary of the catchment and a bit far away from others, so it is not surprising that the correlations between them and other grids are relatively weak in a certain event. Despite this, two rainfall measurements (one for the southern region and the other for the remaining region) may be enough to represent the rainfall diversity in this event. In summary, the radar grid network could be heavily redundant and the analysis of its redundancy can reveal the optimum number of key locations for deploying rain gauges.

PCA applied to the dense radar grid network provides a measurement of the network redundancy for an accepted loss of total information. The analysis is conducted on each storm and highlights the different event characteristics, which are shown in Fig. 4. For most of the events, the first principal component carries close to 90% of the total variance, with the second component bringing this to over 95% of the total variance. This

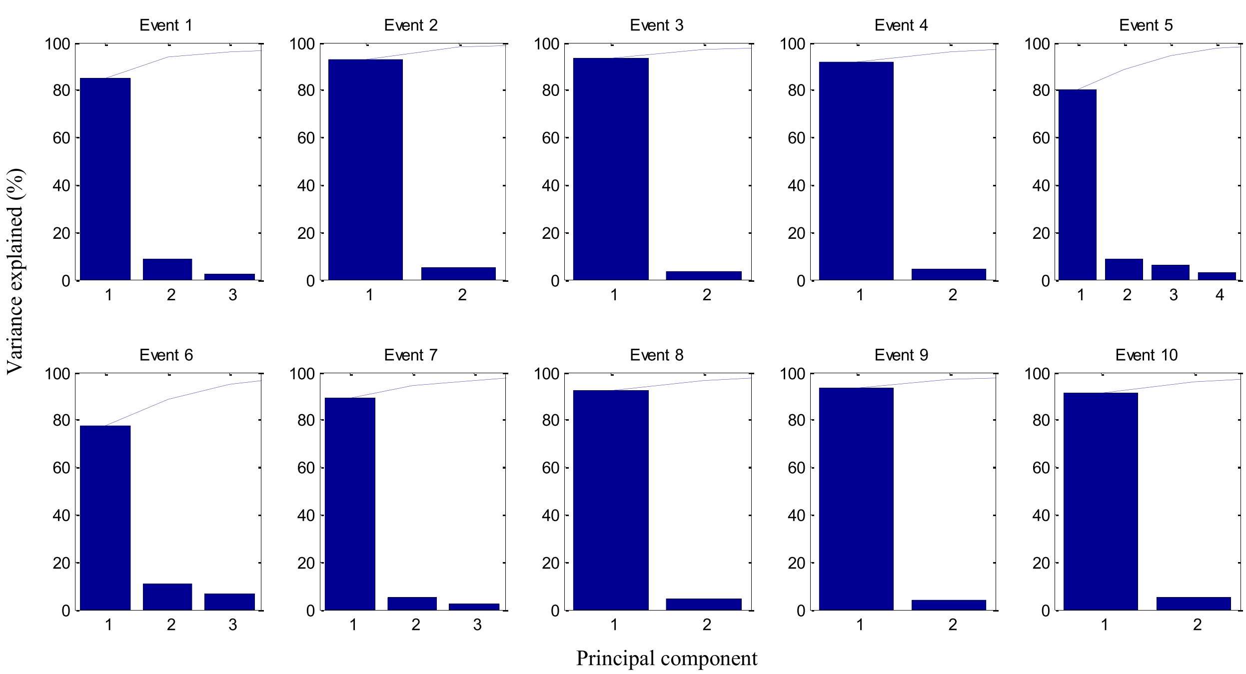

FIG. 4. Variances explained by the PCs for the 10 typical events.

TABLE 2. The number of components to reach the variance threshold for individual events.

Variance	1	2	3	4	5	6	7	8	9	10
75%	1	1	1	1	1	1	1	1	1	1
80%	1	1	1	1	1	2	1	1	1	1
85%	1	1	1	1	2	2	1	1	1	1
90%	2	1	1	1	3	3	2	1	1	1
92.5%	2	1	1	2	3	3	2	2	1	2
95%	3	2	2	2	4	3	3	2	2	2
97.5%	4	2	3	3	5	4	4	3	3	3
99%	7	4	4	5	5	5	6	4	4	5

again indicates a very high level of redundancy in the network, so much so that just one component contains 90% of the total information. However, events 5 and 6 have a noticeably different component weighting. In the case of event 5, the first component can account for around 80% of the total information; in order to reach 90%, both the first and second components are required. The situation is worse for event 6, where the first component carries just around 75% of the information; in order to retain 90% of the information, three components are required. This shows that there is less redundancy in the network for events 5 and 6, indicating that the rainfall amount measured in the radar grids is more varied. This suggests that these two events are less uniform than the other eight events, and so the network will require more radar grids. However, considering 28 radar grids in total, many radar grids are still unessential for these two events. It is worth remarking that the analysis of radar grid network redundancy is served for rain gauge network design. The argument that most radar grids in the current network are unnecessary does not mean we encourage one to apply only part of radar dataset in other applications. In reality, one of the most important advantages of weather radar is that it can take millions of measurements from a single platform and consequently reveal spatial variation of rainfall.

To better show the relationship between principal component numbers and variance explained, thresholds of desired variance explained are set to 75%, 80%, 85%, 90%, 92.5%, 95%, 97.5%, and 99%. The required numbers of components are summarized in Table 2. Leaving aside events 5 and 6, it can be seen that two components are sufficient (although not always necessary) to retain 90% of the information. In addition, to be sure to get at least 95% of the information, three components should be used (although two components is sufficient for six of the events).

b. *Location of rain gauges*

After establishing the level of redundancy in the radar grid network, it is necessary to determine which grids to select so that the maximum level of information will be retained and unnecessary repeated measurements would be removed. Since the components do not represent physical radar grids, cluster analyses (CA Max and CA Med) are used.

The locations of OGs for each event derived from PCA and CA Max are shown for retaining at least 90% of the total variance, as explained in Fig. 5. Each event has a different combination of radar grids and suggest that the optimum locations of a small number of radar grids depends on the rainfall event itself. The selected optimum radar grids tend to be located at the boundary

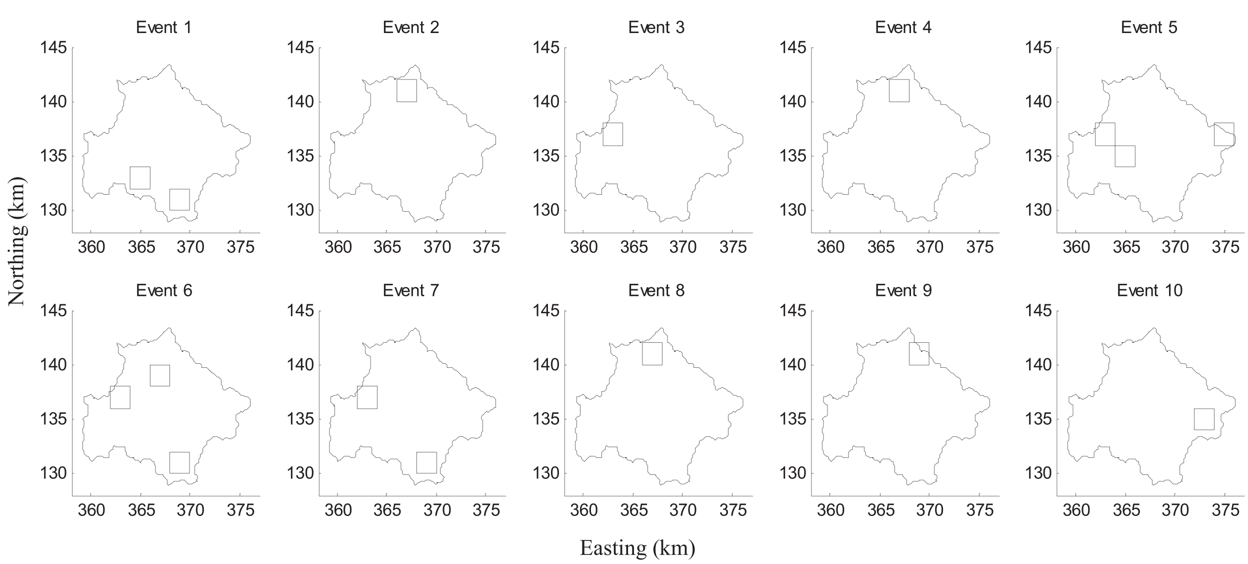

FIG. 5. Locations of the OGs derived from the CA Max method for the 10 typical events (total variance explained ≥90%).

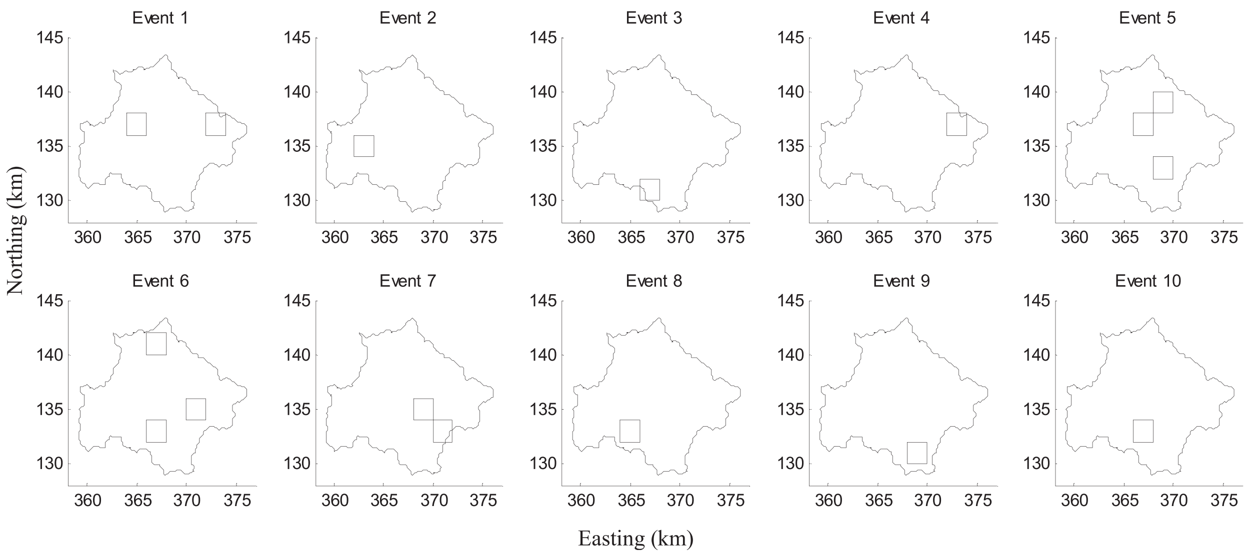

FIG. 6. Locations of the OGs derived from the CA Med method for the 10 typical events (total variance explained ≥90%).

of the catchment, especially for events that claim only one or two grids. For example, in event 1, two OGs are located at the southern boundary of the catchment, while the OG is chosen from the northern boundary in event 2. The corresponding derived OGs using the CA Med method are shown in Fig. 6. Inspection of Figs. 5 and 6 shows that the different variable selection methods produce different OG locations. In some cases, the OG locations differ just slightly (e.g., event 10), and in other cases the locations differ substantially (e.g., events 8 and 9).

It is not practical or cost effective to choose the most suitable location based on one type of rainfall event. Figures 7 and 8 show the envelope of all event-based locations for each method of variable selection. The total number of radar grids needed to satisfy each event is nine for the CA Max case, which is at least 3 times the requirement of each individual case but still less than half of the original number of 28 radar grids. Figure 7 shows the distribution of the required nine gauges is predominately at the boundary region of the catchment. The total number of radar grids increases to 14 for the CA Med case (see Fig. 8). Figure 8 shows that the distribution of the required grids is more evenly spread in contrast to the CA Max selection criterion. These envelopes give an interesting insight into the preferential areas for rain gauge location; however, analysis of individual events demonstrates that just three radar grids are needed to provide a catchment average rainfall close to the 28 radar grids catchment average (see section 4d); therefore the envelope of all radar grid locations does not lead to an efficient network.

c. *Comparison of designed networks by radar and gauge datasets*

The selection procedure of OGs is carried out for each event, while each event gives different outcomes. In an effort to produce an efficient and reliable network for all events, PCA is repeated on the concatenated set of rainfall data, by concatenating the 373 events that cover a period of 6 years. The numbers of components to reach the given variance thresholds (75%, 80%, 85%, 90%, 92.5%, 95%, 97.5%, and 99%) for concatenated events are shown in Table 3. The first principal component carries at least 85% of the total information held in the dense radar grid network. A further two components are required to carry 90% of the total information. It can

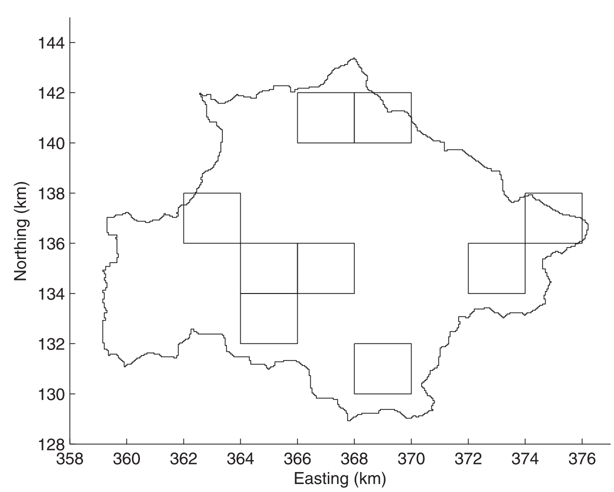

FIG. 7. Locations of the OGs satisfying all typical events. Derived from PCA and CA Max method (total variance explained ≥90%).

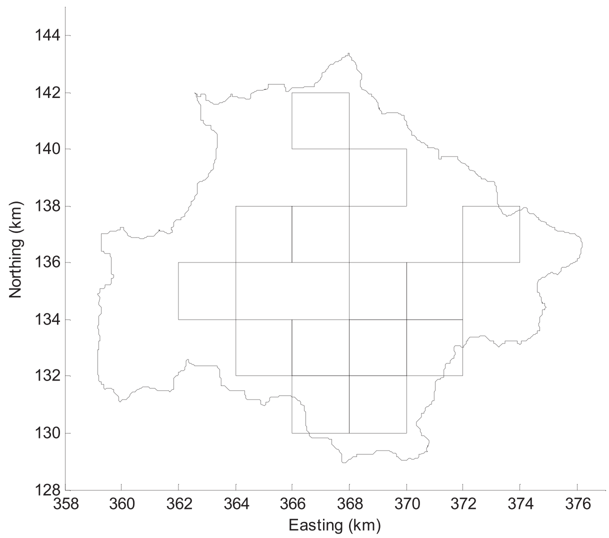

FIG. 8. Locations of the OGs satisfying all typical events. Derived from PCA and CA Med method (total variance explained ≥90%).

TABLE 3. The number of components to reach the variance threshold for concatenated events.

Variance	Component
75%	1
80%	1
85%	1
90%	2
92.5%	3
95%	3
97.5%	6
99%	12

be seen that the 28-radar-grid network still has a high level of redundancy. However, the number of components required to retain 99% of the total information increased from 4 to 7 for most individual events to at least 12 for the set of concatenated events.

To evaluate the performance of rain gauge network design by radar data, a dense rain gauge network is also used as a comparison. As mentioned above, the mean number error and the mean location error are two major indicators for the evaluation. The relationships between variances explained and principal components for concatenated events using radar and rain gauge data, respectively, are shown in Fig. 9. As the variances explained refer to the information the given number of components can achieve, the relationship is also called the information-component curve. It is clear that the disparity between radar and gauge is quite small. For the variance less than 98%, the differences are no more than one component. Moreover, the mean number error calculated using Eq. (1) is only 0.59. These facts prove the radar data have similar performance as rain gauge data in determining the optimum number of gauges. For both radar and rain gauge cases, the required number grows gradually with the increment of the threshold value when the variance threshold is small. From Fig. 9, we can observe that the inflection point appears at around 95% variance threshold. The required components merely climb from one at 84% to three at 95% for both radar and gauge cases. Nevertheless, the growth rate is remarkable when the variance is larger than 95%. In other words, if we expect to maintain more than 95% information, only quite limited additional information can be gained while spending more on numerous additional components. This is obviously not cost effective compared to the easy harvest of information when variance is less than 95%. In summary, analysis of the information-component curve is a promising method to investigate the efficiency of principal components and guide the determination of variance threshold.

As in the individual event, cluster analysis is used to derive the best locations for the rain gauges using concatenated events that correspond to different levels of variance explained given by the principal components (85%, 90%, 95%, and 99%). Figure 10 displays the locations of the selected radar grids and rain gauges by CA Max criterion using radar and gauge datasets, respectively. The dots represent gauges of the gauge-designed network while boxes illustrate the grids of the radar-designed network. It can be seen from the figure that although selected radar grids (OGs) cannot capture all rain gauge points, the distributions of them are quite similar. Take 90% variance, for example: two

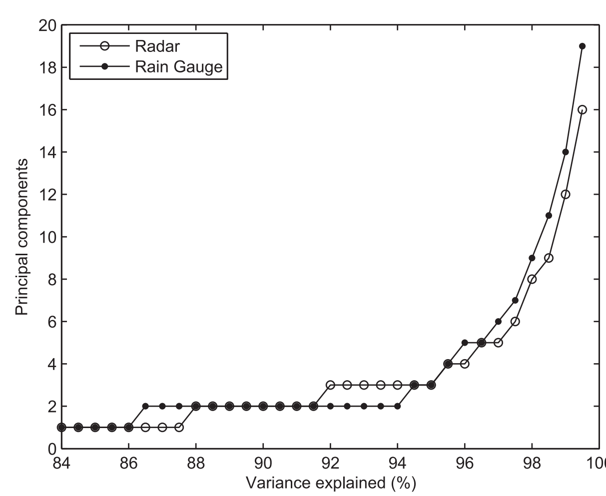

FIG. 9. Relationships between the variance explained and PCs for concatenated events derived by radar and rain gauge datasets.

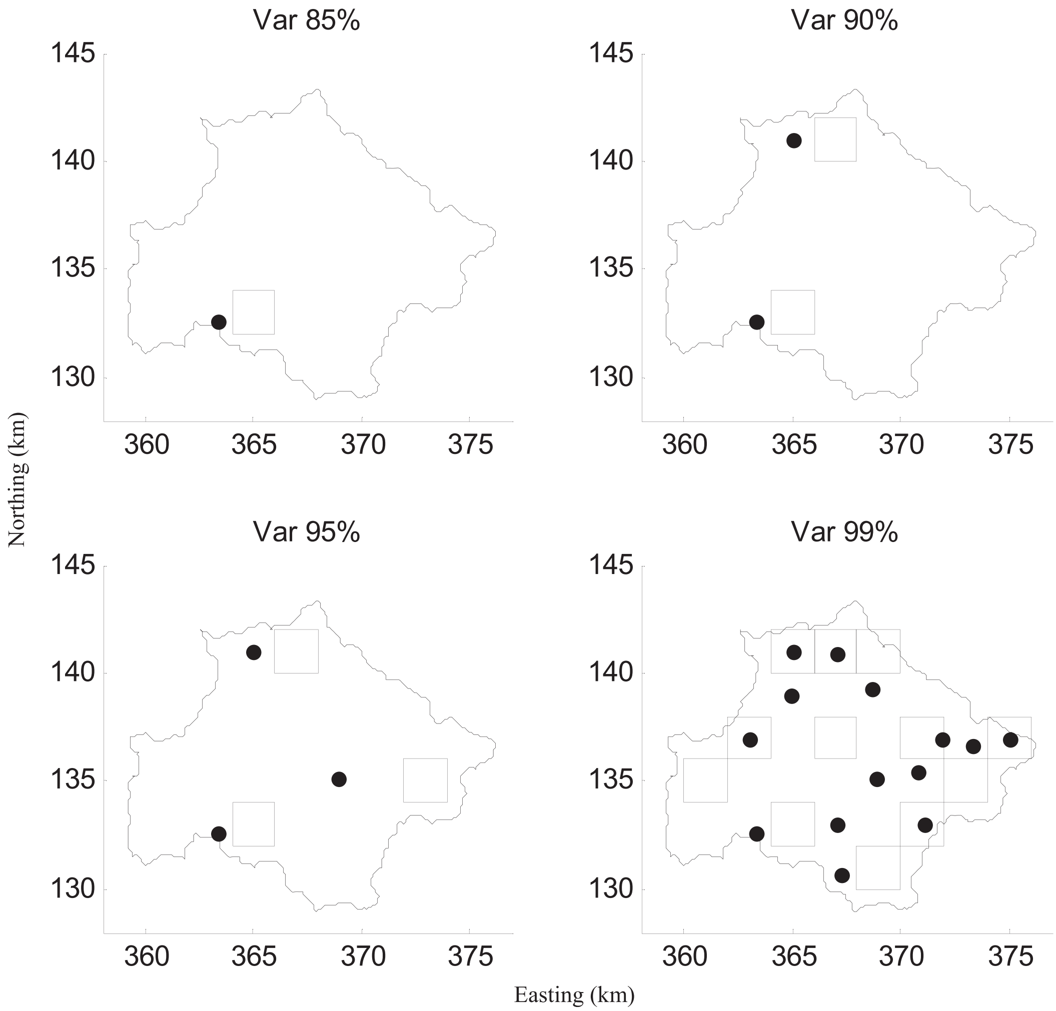

FIG. 10. Comparison of radar grids and rain gauges from radar- and gauge-designed networks with variance thresholds of 85%, 90%, 95%, and 99% using the CA Max method. "Var" refers to the corresponding variance threshold.

rain gauges just locate in the neighboring grids of selected OGs. The same conclusion can also be drawn from Fig. 11, which shows the CA Med case. One can observe that a rain gauge is exactly captured by an OG in 90% variance case. As many gauges and grids are required to retain 99% variance, the corresponding relationship between radar- and gauge-designed networks is not clear in Figs. 10 and 11. To quantitatively describe the differences of locations between two networks, the mean location errors of the radar-designed network are calculated, which is shown in Table 4. The center of the OG is used to measure distance toward the rain gauge, so mean location error still exists even when the OG captures the corresponding rain gauge. In Table 4, it is observed that the mean location errors are less than 3 km in all cases. For 90% variance using CA Med, the errors are as low as 1.02 km. The averaged values of 1.85 and 1.99 km for CA Max and CA Med, respectively, also indicate the strong agreement between radar- and gauge-designed networks. Thus, it can be said that the rain gauge network design using radar dataset can represent gauge dataset; in other words, the effect of radar rainfall uncertainty on designing rain gauge network using the proposed scheme is inconsiderable.

d. *Rain gauge network design evaluation*

As mentioned above, it is important to evaluate whether the radar-designed network can maintain the dominated information of the original radar grid network. Figures 12 and 13 show the scatterplots of each event comparing the 28 radar grids' catchment average with the catchment average produced by each of the variable selection method. Overall, both methods (CA

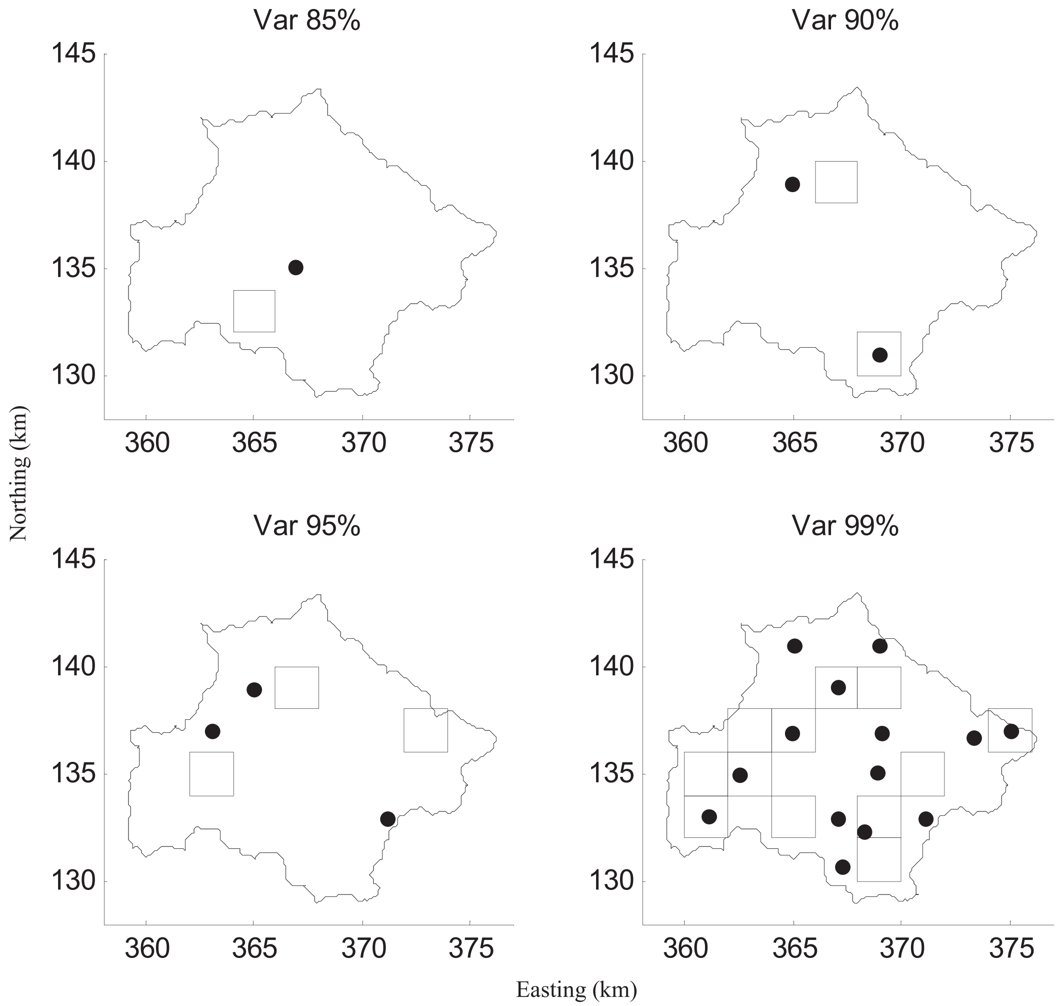

FIG. 11. Comparison of radar grids and rain gauges from radar- and gauge-designed networks with variance thresholds of 85%, 90%, 95%, and 99% using the CA Med method.

Max and CA Med) produce good estimates for average catchment rainfall on an event basis. This is confirmed in Table 5, which gives the corresponding Pearson's correlation coefficient for each method and event. It is found that all correlation coefficients are larger than 0.90 and even reach 0.99 for event 6 using CA Max.

The ability to produce dependable catchment average for each individual event was again determined by the Nash–Sutcliffe coefficient, and these results are given in Table 6. The values are generally larger than 0.90 for the CA Med case and fairly stable for different events. In terms of the CA Max case, we cannot only observe events with poor coefficients (e.g., 0.42 of event 4 and 0.45 of event 10), but also some high-performance events (e.g., 0.97 of event 6). A measure of the range of coefficients given by the interquartile range can provide a clue to the reliability of each method for producing a rain gauge network suitable for all tested events (see Fig. 14). The greater range occurs with CA Max, which suggests that for this case study this method produces the less reliable network.

Scatterplots of concatenated events comparing the 28 radar grids' catchment average are also shown in Fig. 15, together with Pearson's correlation coefficients listed in Table 7. Correlation coefficients as high as 0.92 and 0.93 indicate a fine agreement between the radar-designed network and the original radar grid network. In terms of Nash–Sutcliffe coefficient, CA Max is much smaller than

TABLE 4. Mean location errors (km) for different variance thresholds derived by the two selection methods.

Method	85%	90%	95%	99%	Averaged
CA Max	1.74	1.87	2.62	1.17	1.85
CA Med	2.80	1.02	2.83	1.29	1.99

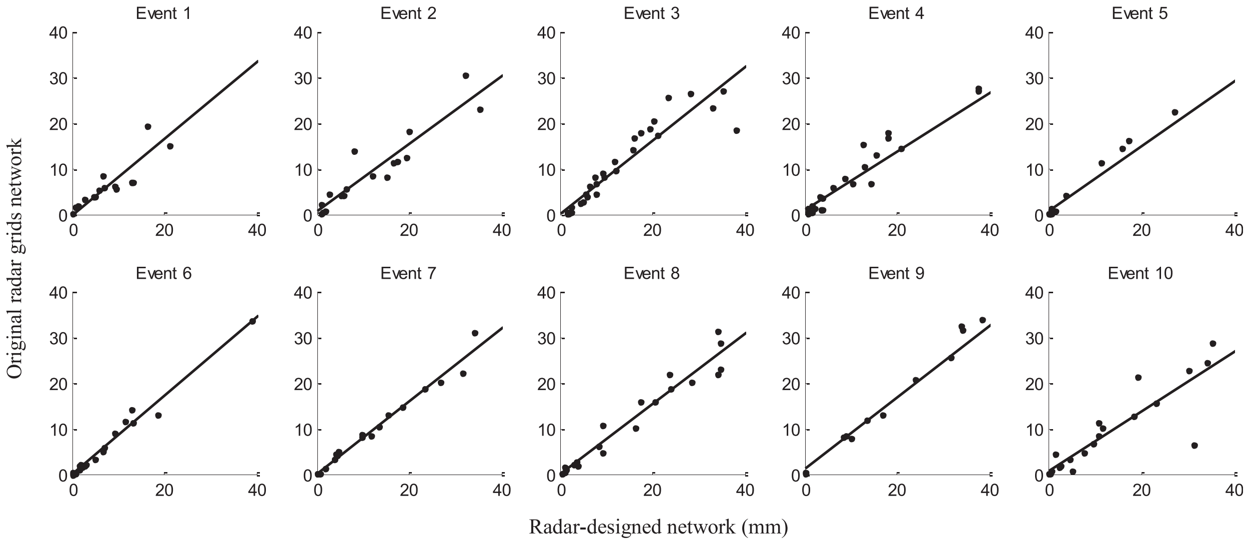

FIG. 12. Rainfall correlations between the original 28 radar grids and the radar-designed network using the CA Max method.

CA Med (0.69 vs 0.84). For this reason, the rain gauge network designed by the CA Med method can maintain more information than CA Max does in the Brue catchment with the same number of OGs.

5. Discussion

A simple, efficient, and quantified method is proposed in this study, and a series of evaluations proves the good performance of the method. However, there are still some key concerns that readers should be aware of. All methods consistently reduce the dense radar grid network from 28 grids down to five or fewer. The main difficulty is in determining the optimal positioning of the rain gauges and the stability of the designed network. The initial study, which analyzed the events individually, produced significant differences in location for each event and method to the point that it was not possible to identify any general pattern. The envelope of all chosen rain gauges produced a network capable of detecting the variability of the rainfall field for all events; however, this led to networks of between 9 and 14 radar grids, substantially more than needed for any one of the events. To tackle this problem, the 373 events covering a period of 6 years were considered as a whole. The proposed scheme was implemented using such concatenated events, and the required number of radar

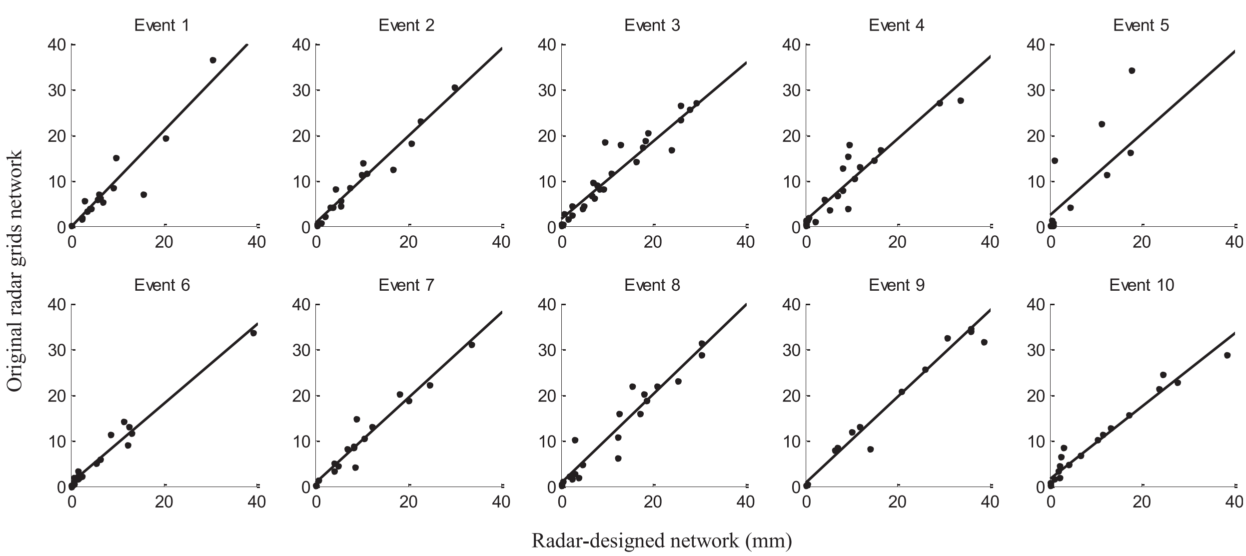

FIG. 13. Rainfall correlations between the original 28 radar grids and the radar-designed network using the CA Med method.

TABLE 5. Correlation coefficients between the 28-radar-grid network and the radar-designed network for individual events.

Method	1	2	3	4	5	6	7	8	9	10
CA Max	0.97	0.93	0.95	0.91	0.98	0.99	0.98	0.97	0.98	0.90
CA Med	0.94	0.98	0.97	0.96	0.98	0.98	0.97	0.96	0.98	0.98

grids dropped to three. The locations of the selected OGs given by each selection criterion are more consistent. The analysis of Nash–Sutcliffe and correlation coefficients between the radar-designed and original network also supports the rationality of this scheme.

In addition, concerns may be raised regarding whether the remotely sensed rainfall (radar rainfall in this study) uncertainty will influence and contaminate the outcomes of the rain gauge network design. It is true that there are numerous yet-to-be-tackled problems associated with weather radar, such as ground clutter, anomalous propagation, signal attenuation, beam blockage, and vertical variability of the reflectivity (Cluckie et al. 2000; Villarini and Krajewski 2010). Many groups have made significant efforts to adjust or describe radar rainfall errors (Borga et al. 2002; Ciach et al. 2007; Collier 1986; Dai et al. 2013, 2015; Kirstetter et al. 2010; Rico-Ramirez and Cluckie 2007; Villarini et al. 2008). In fact, the overall bias of weather radar will not significantly influence this study. For example, if rainfall values of all radar pixels is multiplied by a given ratio, the designed network using the proposed scheme will not change. However, the random error of radar will to some extent affect the final results. To evaluate the possible errors of the radar-designed network caused by radar rainfall uncertainty, this study investigated two networks using radar and rain gauge datasets, respectively. The small differences of gauge numbers and locations between the gauge- and radar-designed networks prove the effectiveness of the proposed scheme. In addition, to reduce the effect of radar rainfall uncertainty on the proposed scheme, we adopted a relatively trustworthy radar dataset herein, which is from the Hydrology Radar Experiment conducted by the Natural Environment Research Council (NERC) Special Topic Programme. Typical errors for radar data have been identified and reduced in the initial processing (Bringi et al. 2011; Moore et al. 2000). Moreover, we used long-term radar rainfall records instead of individual events for the rain gauge network design, which could, to some extent, remove the outliers and produce a more stable outcome. In fact, we propose that a stricter way to solve the uncertainty problem is to integrate the radar rainfall uncertainty model with the proposed scheme. This approach may be necessary in other regions, such as hilly areas where weather radar suffers more problems. The radar rainfall uncertainty model refers to a mathematical approach that elaborately formulates all uncertainties associated with radar rainfall (Dai et al. 2014; Krajewski et al. 1991). An ensemble generation of a large number of probable ''true rainfall'' is currently a popular type of radar rainfall uncertainty model (AghaKouchak et al. 2010; Dai et al. 2014; Germann et al. 2009). For example, we can generate 100 rainfall values that satisfy the error distribution and other restricted conditions of radar rainfall and input them into the proposed scheme to produce 100 possible rain gauge networks. Thus, the designed outcome can be expressed in a probabilistic form instead of a determinate network. A decision-making scheme under uncertainty could be introduced to choose the optimum rain gauge network.

This study makes the assumption that only the center of the radar grid can be used as the potential location of a rain gauge. The spatial resolution of radar data used herein is 2 km. Weather radars with higher spatial resolutions such as 1 km or hundreds of meters are becoming increasingly popular all over the world (Emmanuel et al. 2012; Sandford 2015; Smith et al. 2012; Thorndahl et al. 2014; Wright et al. 2013). Considering the natural spatial continuity of rainfall, there should be a distance error tolerance in rain gauge network design. The so-called distance error refers to the distance differences between the practical designed rain gauge network and the ideal rain gauge network that perfectly satisfies the requirements of maximum rainfall information with a minimum number of gauges. For example, in this study, where there are no dramatic changes of land surface terrain, a distance error of hundreds of meters is considered to be acceptable. However, the resolution of satellite rainfall is relatively lower, although it increases rapidly. For example, the

TABLE 6. Nash–Sutcliffe coefficients between the 28-radar-grid network and the radar-designed network for individual events.

Method	1	2	3	4	5	6	7	8	9	10
CA Max	0.86	0.67	0.80	0.42	0.86	0.97	0.73	0.75	0.82	0.45
CA Med	0.89	0.95	0.91	0.92	0.94	0.95	0.95	0.92	0.95	0.90

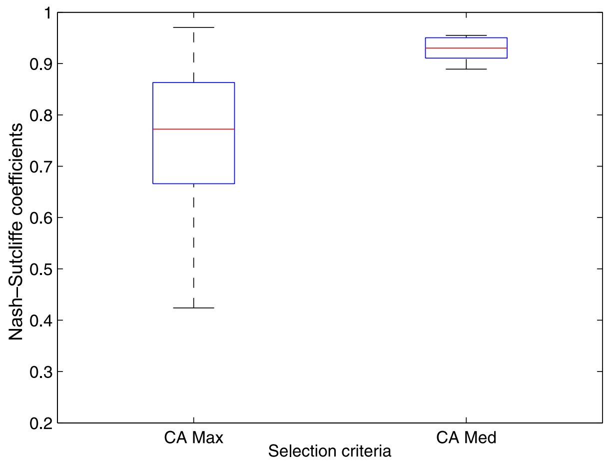

FIG. 14. The range of Nash–Sutcliffe coefficients for the two selection criteria applied on 10 typical events.

spatial resolutions of postprocessed rainfall products from TRMM and GPM are just 5 and 4 km, respectively (Matsui et al. 2013; Simpson et al. 1988). In such situations, scale differences between point rain gauges and areal satellite grids may be worth notice. Such point-to-area error has been studied by numerous hydrologists and meteorologists in remotely sensed rainfall fields, and abundant methods of describing or reducing this error have been proposed (Bringi et al. 2011; Ciach and Krajewski 1999; Habib et al. 2004). Integration of these methods and the proposed scheme offers a promising solution to rain gauge network design using satellite rainfall measurements.

6. Conclusions

The results of this study show that PCA combined with simple selection criteria is an effective tool for rain gauge network design and for the given case study. Moreover, this new methodology can be used in ungauged catchments, as it only requires rainfall data that could be provided by weather radar, satellite, or other remote sensors. The very nature of PCA is to identify how much information in a dataset is useful; this property has been successfully exploited to provide the number of rain gauges needed for a chosen level of retained information. The principal components derived from the PCA method do not represent physical rain gauges; therefore, criteria selection methods are required to identify the best rain gauge locations. Cluster analysis methods presented in this study are both simple and effective in determining rain gauge locations. For individual events, the best rain gauge locations vary significantly for different events, and this means that it is impossible to have an optimum network for individual events unless all the best rain gauge locations of all the events are installed. However, such a network would be impractical and expensive to implement. Therefore, a compromise must be made based on the concatenated events to derive an overall optimized rain gauge network. For the presented case study, two selection criteria (CA Max and CA Med) are used to determine the optimum locations of rain gauge network. Both methods can ensure the network designed by radar dataset has similar characteristics as that by the gauge dataset. It is found that CA Max tends to pick the radar grids located at the boundary of catchment, while the selected radar grids from CA Med are distributed more evenly over the catchment. Moreover, CA Med produces a higher-performance network from the view of Nash–Sutcliffe coefficient.

Except for cluster analysis, there is also a range of other selection criteria can be used to choose a subset of the original variables that approximate the retained principal components. For example, loading combination criteria (LC), variable deselection (B2), and variable retention (B4) methods (Al-Kandari and Jolliffe 2001, 2005) can also be introduced into the proposed scheme. LC, B2, and

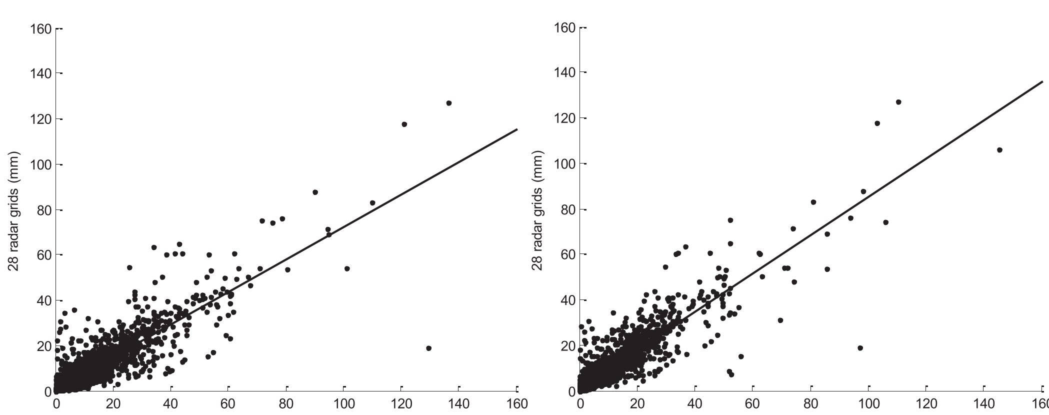

FIG. 15. Rainfall correlations between the original 28 radar grids and the reduced radar grids of the radar-designed network using the (left) CA Max and (right) CA Med methods for concatenated events.

TABLE 7. Correlation coefficients and Nash–Sutcliffe coefficients between the 28-radar-grid network and the radar-designed network for concatenated events.

Method	Correlation coefficient	Nash–Sutcliffe coefficient
CC Max	0.92	0.69
CC Med	0.93	0.84

B4 selection criteria choose variables according to a given association with the original p variables and the loadings of either the first few components (for variable retention) of or the last $(p - q)$ components (for variable deselection). A detailed description of these methods is given by Al-Kandari and Jolliffe (2001). The cluster analysis can be easily replaced by these methods in the proposed scheme. There is no intrinsic disparity among these selection criteria, but they may have different performances in different study areas. One of the major advantages of this study is that the local storm characteristics have already been contained in radar measurements and can be derived through the analysis of long-term radar data. So, we believe this study can be easily and effectively extended to other study areas. Since this is the first time that PCA has been used in rain gauge network design, we hope more study areas with diverse climate and geographical conditions could be explored by the research community to further verify and improve the proposed scheme.

As stated above, hydrologists and meteorologists are facing a decrease in the number of the available rain gauges (Overeem et al. 2013), so the proposed scheme can be used for both rain gauge network design in an ungauged catchment and to help reduce the number of current gauges in a reasonable form. Since the emergence of remotely sensed rainfall measurements, rain gauges continually assist remote sensors in offering more trustworthy rainfall measurements. This study is a preliminary attempt at using remote sensor datasets to solve the traditional rain gauge problems. Based on remotely sensed rainfall information, other problems such as wind effect on rain gauges may also be studied with a new insight, and the original complicated problems may be solved by a simple scheme just as what has been discussed in this study.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant 41501429), the Priority Academic Program Development of Jiangsu Higher Education Institutions program (Grant 164320H116), and the University Natural Science Project of Jiangsu Province (Grant 16KJA170001). The authors acknowledge the British Atmospheric Data Centre for providing the data. This work was done as a private venture and not in the author's (TI) capacity as an employee of the Jet Propulsion Laboratory, California Institute of Technology.

REFERENCES

AghaKouchak, A., A. Bárdossy, and E. Habib, 2010: Copula-based uncertainty modelling: Application to multisensor precipitation estimates. *Hydrol. Processes*, **24**, 2111–2124.

Al-Kandari, N. M., and I. T. Jolliffe, 2001: Variable selection and interpretation of covariance principal components. *Commun. Stat. Simul. Comput.*, **30**, 339–354, doi:10.1081/SAC-100002371.

——, and ——, 2005: Variable selection and interpretation in correlation principal components. *Environmetrics*, **16**, 659–672, doi:10.1002/env.728.

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, doi:10.1016/S0022-1694(97)00153-4.

Barancourt, C., J. Creutin, and J. Rivoirard, 1992: A method for delineating and estimating rainfall fields. *Water Resour. Res.*, **28**, 1133–1144, doi:10.1029/91WR02896.

Bastin, G., B. Lorent, C. Duqué, and M. Gevers, 1984: Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. *Water Resour. Res.*, **20**, 463–470, doi:10.1029/WR020i004p00463.

Bogárdi, I., A. Bárdossy, and L. Duckstein, 1985: Multicriterion network design using geostatistics. *Water Resour. Res.*, **21**, 199–208, doi:10.1029/WR021i002p00199.

Borga, M., F. Tonelli, R. J. Moore, and H. Andrieu, 2002: Long-term assessment of bias adjustment in radar rainfall estimation. *Water Resour. Res.*, **38**, 1226, doi:10.1029/2001WR000555.

Bradley, A. A., C. Peters-Lidard, B. R. Nelson, J. A. Smith, and C. B. Young, 2002: Raingage network design using NEXRAD precipitation estimates. *J. Amer. Water Resour. Assoc.*, **38**, 1393–1407, doi:10.1111/j.1752-1688.2002.tb04354.x.

Bras, R. L., and I. Rodríguez-Iturbe, 1985: *Random Functions and Hydrology*. Addison-Wesley, 704 pp.

Bringi, V., M. Rico-Ramirez, and M. Thurai, 2011: Rainfall estimation with an operational polarimetric C-band radar in the United Kingdom: Comparison with a gauge network and error analysis. *J. Hydrometeor.*, **12**, 935–954, doi:10.1175/JHM-D-10-05013.1.

Ciach, G. J., and W. F. Krajewski, 1999: On the estimation of radar rainfall error variance. *Adv. Water Resour.*, **22**, 585–595, doi:10.1016/S0309-1708(98)00043-8.

——, ——, and G. Villarini, 2007: Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data. *J. Hydrometeor.*, **8**, 1325–1347, doi:10.1175/2007JHM814.1.

Cluckie, I., R. Griffith, A. Lane, and K. Tilford, 2000: Radar hydrometeorology using a vertically pointing radar. *Hydrol. Earth Syst. Sci.*, **4**, 565–580, doi:10.5194/hess-4-565-2000.

Collier, C., 1986: Accuracy of rainfall estimates by radar, part I: Calibration by telemetering raingauges. *J. Hydrol.*, **83**, 207–223, doi:10.1016/0022-1694(86)90152-6.

Cormack, R. M., 1971: A review of classification. *J. Roy. Stat. Soc.*, **134A**, 321–367, doi:10.2307/2344237.

Dai, Q., and D. Han, 2014: Exploration of discrepancy between radar and gauge rainfall estimates driven by wind fields. *Water Resour. Res.*, **50**, 8571–8588, doi:10.1002/2014WR015794.

——, ——, M. A. Rico-Ramirez, and T. Islam, 2013: The impact of raindrop drift in a three-dimensional wind field on a radar–gauge rainfall comparison. *Int. J. Remote Sens.*, **34**, 7739–7760, doi:10.1080/01431161.2013.826838.

——, ——, ——, and P. K. Srivastava, 2014: Multivariate distributed ensemble generator: A new scheme for ensemble radar precipitation estimation over temperate maritime climate. *J. Hydrol.*, **511**, 17–27, doi:10.1016/j.jhydrol.2014.01.016.

——, M. A. Rico-Ramirez, D. Han, T. Islam, and S. Liguori, 2015: Probabilistic radar rainfall nowcasts using empirical and theoretical uncertainty models. *Hydrol. Processes*, **29**, 66–79, doi:10.1002/hyp.10133.

Emmanuel, I., H. Andrieu, and P. Tabary, 2012: Evaluation of the new French operational weather radar product for the field of urban hydrology. *Atmos. Res.*, **103**, 20–32, doi:10.1016/j.atmosres.2011.06.018.

Germann, U., M. Berenguer Ferrer, D. Sempere Torres, and M. Zappa, 2009: REAL—Ensemble radar precipitation estimation for hydrology in a mountainous region. *Quart. J. Roy. Meteor. Soc.*, **135**, 445–456, doi:10.1002/qj.375.

Habib, E., G. J. Ciach, and W. F. Krajewski, 2004: A method for filtering out raingauge representativeness errors from the verification distributions of radar and raingauge rainfall. *Adv. Water Resour.*, **27**, 967–980, doi:10.1016/j.advwatres.2004.08.003.

Jolliffe, I. T., 1986: *Principal Component Analysis*. Springer, 271 pp.

Kirstetter, P.-E., G. Delrieu, B. Boudevillain, and C. Obled, 2010: Toward an error model for radar quantitative precipitation estimation in the Cévennes–Vivarais region, France. *J. Hydrol.*, **394**, 28–41, doi:10.1016/j.jhydrol.2010.01.009.

Krajewski, W. F., V. Lakshmi, K. P. Georgakakos, and S. C. Jain, 1991: A Monte Carlo study of rainfall sampling effect on a distributed catchment model. *Water Resour. Res.*, **27**, 119–128, doi:10.1029/90WR01977.

Krstanovic, P. F., and V. P. Singh, 1992: Evaluation of rainfall networks using entropy: I. Theoretical development. *Water Resour. Manage.*, **6**, 279–293, doi:10.1007/BF00872281.

Lorenz, C., and H. Kunstmann, 2012: The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *J. Hydrometeor.*, **13**, 1397–1420, doi:10.1175/JHM-D-11-088.1.

Matsui, T., and Coauthors, 2013: GPM satellite simulator over ground validation sites. *Bull. Amer. Meteor. Soc.*, **94**, 1653–1660, doi:10.1175/BAMS-D-12-00160.1.

Mishra, A. K., and P. Coulibaly, 2009: Developments in hydrometric network design: A review. *Rev. Geophys.*, **47**, RG2001, doi:10.1029/2007RG000243.

Moore, R. J., D. A. Jones, D. R. Cox, and V. S. Isham, 2000: Design of the HYREX raingauge network. *Hydrol. Earth Syst. Sci.*, **4**, 521–530, doi:10.5194/hess-4-521-2000.

Morrissey, M. L., J. A. Maliekal, J. S. Greene, and J. Wang, 1995: The uncertainty of simple spatial averages using rain gauge networks. *Water Resour. Res.*, **31**, 2011–2017, doi:10.1029/95WR01232.

Moss, M. E., and G. D. Tasker, 1991: An intercomparison of hydrological network-design technologies. *Hydrol. Sci. J.*, **36**, 209–221, doi:10.1080/02626669109492504.

Nash, J., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.*, **10**, 282–290, doi:10.1016/0022-1694(70)90255-6.

Overeem, A., H. Leijnse, and R. Uijlenhoet, 2013: Country-wide rainfall maps from cellular communication networks. *Proc. Natl. Acad. Sci. USA*, **110**, 2741–2745, doi:10.1073/pnas.1217961110.

Pardo-Igúzquiza, E., 1998: Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *J. Hydrol.*, **210**, 206–220, doi:10.1016/S0022-1694(98)00188-7.

Rico-Ramirez, M., and I. Cluckie, 2007: Bright-band detection from radar vertical reflectivity profiles. *Int. J. Remote Sens.*, **28**, 4013–4025, doi:10.1080/01431160601047797.

Rodríguez-Iturbe, I., and J. M. Mejía, 1974: The design of rainfall networks in time and space. *Water Resour. Res.*, **10**, 713–728, doi:10.1029/WR010i004p00713.

Sandford, C., 2015: Correcting for wind drift in high resolution radar rainfall products: A feasibility study. *J. Hydrol.*, **531**, 284–295, doi:10.1016/j.jhydrol.2015.03.023.

Shih, S. F., 1982: Rainfall variation analysis and optimization of gaging systems. *Water Resour. Res.*, **18**, 1269–1277, doi:10.1029/WR018i004p01269.

Simpson, J., R. F. Adler, and G. R. North, 1988: A proposed Tropical Rainfall Measuring Mission (TRMM) satellite. *Bull. Amer. Meteor. Soc.*, **69**, 278–295, doi:10.1175/1520-0477(1988)069<0278:APTRMM>2.0.CO;2.

Smith, J. A., M. L. Baeck, G. Villarini, C. Welty, A. J. Miller, and W. F. Krajewski, 2012: Analyses of a long-term, high-resolution radar rainfall data set for the Baltimore metropolitan region. *Water Resour. Res.*, **48**, W04504, doi:10.1029/2011WR010641.

Stedinger, J. R., and G. D. Tasker, 1985: Regional hydrologic analysis: 1. Ordinary, weighted, and generalized least squares compared. *Water Resour. Res.*, **21**, 1421–1432, doi:10.1029/WR021i009p01421.

Tasker, G. D., and M. E. Moss, 1979: Analysis of Arizona Flood Data Network for regional information. *Water Resour. Res.*, **15**, 1791–1796, doi:10.1029/WR015i006p01791.

Thorndahl, S., J. E. Nielsen, and M. R. Rasmussen, 2014: Bias adjustment and advection interpolation of long-term high resolution radar rainfall series. *J. Hydrol.*, **508**, 214–226, doi:10.1016/j.jhydrol.2013.10.056.

Tsintikidis, D., K. P. Georgakakos, J. A. Sperfslage, D. E. Smith, and T. M. Carpenter, 2011: Precipitation uncertainty and raingauge network design within Folsom Lake watershed. *J. Hydrol. Eng.*, **7**, 175–184, doi:10.1061/(ASCE)1084-0699(2002)7:2(175).

Villarini, G., and W. F. Krajewski, 2010: Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. *Surv. Geophys.*, **31**, 107–129, doi:10.1007/s10712-009-9079-x.

——, P. V. Mandapaka, W. F. Krajewski, and R. J. Moore, 2008: Rainfall and sampling uncertainties: A rain gauge perspective. *J. Geophys. Res.*, **113**, D11102, doi:10.1029/2007JD009214.

Volkmann, T. H., S. W. Lyon, H. V. Gupta, and P. 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.

Walsh, D., 2012: The tricky business of counting rain. *New York Times*, 2 July. [Available online at http://green.blogs.nytimes.com/2012/07/02/do-not-publish-the-tricky-business-of-counting-rain/#more-143581.]

Wood, S., D. Jones, and R. Moore, 2000: Accuracy of rainfall measurement for scales of hydrological interest. *Hydrol. Earth Syst. Sci.*, **4**, 531–543, doi:10.5194/hess-4-531-2000.

Wright, D. B., J. A. Smith, G. Villarini, and M. L. Baeck, 2013: Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition. *J. Hydrol.*, **488**, 150–165, doi:10.1016/j.jhydrol.2013.03.003.

Yang, Y., and D. Burn, 1994: An entropy approach to data collection network design. *J. Hydrol.*, **157**, 307–324, doi:10.1016/0022-1694(94)90111-2.

Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh, 2012: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Amer. Meteor. Soc.*, **93**, 1401–1415, doi:10.1175/BAMS-D-11-00122.1.