Spatial interpolation of rain-field dynamic time-space evolution based on radar rainfall data

Peng Liu and Yeou-Koung Tung

ABSTRACT

Accurate and reliable measurement and prediction of the spatial and temporal distribution of rain field over a wide range of scales are important topics in hydrologic investigations. In this study, a geostatistical approach was adopted. To estimate the rainfall intensity over a study domain with the sample values and the spatial structure from the radar data, the cumulative distribution functions (CDFs) at all unsampled locations were estimated. Indicator kriging (IK) was used to estimate the exceedance probabilities for different preselected threshold levels, and a procedure was implemented for interpolating CDF values between the thresholds that were derived from the IK. Different probability distribution functions of the CDF were tested and their influences on the performance were also investigated. The performance measures and visual comparison between the observed rain field and the IK-based estimation suggested that the proposed method can provide good results of the estimation of indicator variables and is capable of producing a realistic image.

Key words | exceedance probabilities, indicator kriging, spatial interpolation

Peng Liu (corresponding author)
State Key Laboratory of Hydrology-Water
  Resources and Hydraulic Engineering,
Nanjing Hydraulic Research Institute,
223 Guangzhou Road, Nanjing 210029,
China
E-mail: *pliu@nhri.cn*

Yeou-Koung Tung
Department of Civil and Environmental
  Engineering,
The Hong Kong University of Science and
  Technology,
Clear Water Bay, Kowloon,
Hong Kong

INTRODUCTION

As rainfall constitutes the main source of freshwater for our planet and all living things on earth, accurate and reliable measurement and prediction of its spatial and temporal distribution over a wide range of scales is an important subject for hydrologic investigations. The estimation of the rainfall amount over the spatial domain from the data taken at measurement stations is a crucial stage in many hydrological applications. The problem of estimating spatial distribution of rainfall may be described as follows. Obtain an estimate of the unknown rainfall amount, $r(x_0)$, at an arbitrary location, x_0, within the estimation domain, $\mathcal{A}$, using nearby measurements in and around $\mathcal{A}$, $r(x_1)$, $r(x_2)$, $\ldots$, $r(x_n)$, at locations, x_1, x_2, $\ldots$, x_n, respectively. The value of $r(x_0)$ can be estimated as a weighted average of $r(x_1)$, $r(x_2)$, $\ldots$, $r(x_n)$ (Seo 1998).

To construct new data points within the range of a discrete set of known sample data points, interpolation is needed. Many schemes for spatial interpolation of rainfall have been developed, ranging from simple techniques, such as Thiessen polygons (Thiessen 1911), which simply assign the record of the nearest raingauge to the unsampled location, or inverse-distance weighting (IDW) schemes (Cressman 1959; Shepard 1968; Barnes 1973), to more complex approaches, such as kriging (Matheron 1965). In the Thiessen polygons method, only one raingauge is contributed to the estimated value of each unsampled location and the information on rainfall gradients is lost (Dirks *et al.* 1998). To overcome this shortcoming in estimating the spatial variation of rainfall within the considered domain, numerous studies have been proposed and tested. A simple scheme is based on IDW function with different forms of the weighting functions. The IDW scheme is a relatively simple, deterministic interpolation method, and it requires the determination of the weighting functions prior

to the interpolation. Furthermore, the arbitrary nature of choosing the weighting functions may significantly affect the accuracy of the interpolated field. More studies have been made on geostatistical tools, like kriging, to interpolate rainfall field. Kriging avoids the need for preselecting parameters by calculating the semi-variogram for each pair of observations in the network as a measure of the degree of spatial dependence between the points. Kriging is increasingly preferred in the study of spatial interpolation of precipitation because it allows one to capture the spatial correlation between neighboring observations to estimate and predict attribute values at unsampled locations (Goovaerts 2000). Especially in a region with intense and strongly varying rainfall events, a kriging procedure generally provides a much better estimation than any of the more commonly used methods (Creutin & Obled 1982). Similar conclusions were drawn in several other studies (Shaw & Lynn 1972; Tabios & Salas 1985; Phillips et al. 1992; Abtew et al. 1993). In kriging, the semi-variogram is used to determine the weighting function.

Kriging is a local estimation technique of the best linear unbiased estimator (BLUE) for the unknown values of spatial and temporal variables. It is based on the use of a random field, and a number of assumptions such as stationarity and spatial ergodicity, so as to reduce the needed information to a so-called variogram that can be estimated from the available measurements (Loquin & Dubois 2010). Based on the variogram, optimal weights are assigned to known values in order to calculate unknown ones. The variogram changes with the distance and the weights depend on the known sample distribution. As an interpolation tool, it has some very useful properties (Grimes 2008). It takes account of spatial structure inherent in the data and provides unbiased, optimal (in a least-squares sense) measured values. In addition, it can be adapted to account for secondary information (e.g., topography) and provides the degree of uncertainty for each interpolated value. Ordinary kriging (OK) is the most commonly used algorithm in the kriging family. Another best-known technique is simple kriging (SK), which is based on the assumption that the expected value of the regionalized variable is a known constant. There are many variants of the kriging method, such as co-kriging which takes into account the correlation of the main variable (e.g., rainfall) and secondary variables

(e.g., topography), kriging with external drift (KED) which assumes that the mean of a considered variable is linearly related to a secondary variable and indicator kriging (IK) which carries out the kriging procedure on a binary pattern and provides the probability of occurrence of a considered variable being greater than a given threshold at the estimated locations. The new method, besides allowing spatial information to be incorporated into the modeling and estimation process, also allows temporal information to be incorporated. This technique employs temporal data as secondary co-kriged variables.

Quite a number of works have been carried on to interpolate rainfall data using geostatistical kriging methods (Lebel & Laborde 1988; Barancourt et al. 1992). Goovaerts (2000) compared three multivariate geostatistical algorithms to incorporate a digital elevation model into the spatial prediction of rainfall with the straightforward linear regression techniques, such as the inverse distance or Thiessen polygon. The results confirmed that in a region with low-density raingauge network, geostatistical interpolation outperformed the other univariate methods. Kyriakidis et al. (2001) presented a geostatistical framework that integrated atmospheric and terrain characteristics into the spatial interpolation of seasonal average daily precipitation. The results indicated that geostatistical framework could provide more accurate representations of the spatial distribution of rainfall than those found in the traditional analyses. However, the degree of the improvement depended on the density of the rain gauge stations and the spatial variability of rain field. Pardo-Igúzquiza & Dowd (2005) described a two-step procedure of IK and OK, made a comparison with the empirical kriging and discussed the pros and cons of each method. Kriging with external drift (KED) and indicator kriging with external drift (IKED) were used for the spatial interpolation of hourly rainfall from raingauges using additional information from the radar by Haberlandt (2007). The impact of the semi-variogram estimation (isotropic or anisotropic behavior) on the interpolation performance was analyzed, and it was concluded that the automatic fitting procedure with isotropic variograms offered the best results in his study.

A merging method combining a mean precipitation field interpolated from raingauge observations, along with information about the spatial variability from radar data, was

developed by Ehret et al. (2008). This method was proved to offer a reliable basis for a flood forecasting system. Goudenhoofdt & Delobbe (2009) tested several radar-gauge merging methods with various degrees of complexity and found that geostatistical merging methods outperformed the others. Verworn & Haberlandt (2011) applied the multivariate kriging method with external drift (MKED) using additional information from topography, rainfall data from the denser daily networks and weather radar data and concluded that the weather radar was proved to be the most valuable additional information for convective summer events, while daily rainfall was sufficient for winter events. Tobin et al. (2011) compared different interpolation methods, including IDW, OK and KED, and showed that KED outperformed other methods, which was consistent with the observation by Velasco-Forero et al. (2009). Wagner et al. (2012) analyzed seven different rainfall interpolation schemes including Thiessen polygons and geostatistical approaches with regard to their suitability to produce spatial rainfall estimates in a monsoon-dominated region with scarce rainfall measurements. They concluded that the precipitation interpolation approaches using appropriate covariates performed the best. Sideris et al. (2014b) expanded the well-tested technique (KED) by introducing and coupling temporal data with spatial data through co-kriging and this proposed technique achieved to serve as a real-time operational tool in Switzerland. Lebrenz & Bárdossy (2019) applied quantile kriging (QK) to the variable of observed monthly precipitation from raingauges in South Africa. The cross-validations proved that QK performed the best among OK and external drift kriging (EDK). Huang et al. (2019) proposed a new spatial precipitation interpolation method based on the information diffusion principle that has been proven to effectively reduce data or information uncertainties from point ground observations.

However, kriging is limited in that it depends on the existence of a reliable representation of the spatial correlation structure and is mathematically more complicated and computationally more intensive. The usefulness of kriging depends on whether the distribution of data agrees with certain statistical assumptions (e.g., Gaussian distribution) and a representative semi-variogram can be defined. In modeling spatial variability of rainfall field, these assumptions are not always satisfied. A number of technical studies are carried out to fix this problem and are well documented in Goovaerts (1997). IK is often qualified as a 'non-parametric' approach because it does not rely on a prespecified spatial distribution model for the attribute under study. It is, therefore, applicable for any data sets, which explains the popularity of the method. Liu & Tung (2015) established a geostatistical model based on indicator variograms fed with radar-derived rainfall data that could effectively evaluate the dynamic evolution of the spatial-temporal variation of rain field in Hong Kong. Apart from this non-Gaussian distribution issue, still several questions need to be further addressed for an improvement of interpolation, like the influence of the semi-variogram estimation on the interpolation performance of short time step rainfall in Hong Kong is relatively unknown.

The main objective of this paper is to contribute to evaluate the performance of IK schemes using different sample variograms and different probability distribution functions for fitting rainfall intensity. The paper is organized as follows. First, a brief description of the study area and the data preparation is presented. Then, the methodology to estimate the rainfall intensity by different sample variograms and different probability distribution functions is introduced. This is followed by discussion of the results and finally the conclusions that can be drawn from this study are presented.

STUDY AREA AND DATA PREPARATION

Study area

Hong Kong, with a population of about 7 million people, is situated on China's south coast and is enclosed by the Pearl River Delta and the South China Sea. Heavy rain in summer in Hong Kong is typically associated with the southwest monsoon and tropical cyclones. The threat of flooding is greatest during storm surges generated by the passage of typhoons due to high wind. Many parts of the territory are densely populated urban areas surrounded by or situated within catchments that are typically small with relatively steep slopes. These urban areas and catchments have a relatively short time-of-concentration and are particularly susceptible to the threat of flash floods and extensive

flooding rendering significant economic losses and possible fatalities. Apart from urban flash flooding, another disaster caused by persistent heavy rain in Hong Kong is landslides because of the hilly terrain.

Data preparation

There are numbers of factors influencing the evolution of the spatial structure of rainfall during a rainstorm event, including the weather conditions, the topography of the study area, type of precipitation, and the spatial and temporal scale under consideration. The spatial structure of rainfall could be quite dynamic in space and time. To capture and describe the dynamic time-space evolution of the rainfall, Liu & Tung (2015) proposed a geostatistical approach based on indicator variograms of rain fields and found that the variability of the spatial structure of rain field was rationally retrieved by the main features of the variograms. Therefore, effective and reliable estimation of the variograms at each time step during the rainstorm event is of great importance for the geostatistical interpolation. The variogram estimation has to be made for each time step.

Traditionally, rainfall information has been collected in the form of raingauge records and rainfall is observed at a particular point on the Earth's surface. While the temporal resolution of the rainfall records from the raingauge stations is acceptable, the density of raingauge network is sometimes too sparse to reliably estimate the spatial structure. The idea of using radar reflectivity measurements is appealing by providing the high space-time resolution of observations. In recent years, increasing use has been made of radar-based rainfall estimates and radar images. Due to the superiority of the high spatial and temporal resolution, radar rainfall data have been used more frequently as an input for hydrological modeling.

However, careful system design and sophisticated data processing are required for radar rainfall estimation because of sampling error and measurement-algorithm induced error. The relative influence for different sources of error as well as relations between point raingauge rainfall records and radar measurements are described at length elsewhere (Huffman 1997; Ciach & Krajewski 1999; Harrison *et al.* 2000; Germann & Joss 2001; Sokol 2003). Because of the existence of frequently occurring large space-time variable bias, radar data that have not been corrected are insufficient for the rainfall estimation or prediction (Cole & Moore 2008; Verworn & Haberlandt 2011). Quite a number of studies which compare different algorithms for deriving estimates of precipitation from radar rainfall data have been carried out in conjunction with measurements at raingauges to map precipitation (Wagner *et al.* 2012).

The radar data used in this study is collected by the Hong Kong Observatory (HKO) from a rainstorm nowcasting system SWIRLS (Short-range Warning of Intense Rainstorms in Localized Systems). The radar reflectivity with a high spatial (1 km^2) and temporal (6 min) resolution which is detected at 1 km height is correlated every 5 min with the rainfall recorded by the raingauges underneath to gain the optimal parameters a and b in the $Z - R$ relationship ($Z = aR^b$). A total of 132 raingauge stations, 46 from the HKO and 86 from the Geotechnical Engineering Office (GEO), over the Territory of Hong Kong were used (Figure 1).

To choose the thresholds, the quantiles of the rainfall intensity values are obtained for each rainstorm event (see Figure 2). For the purpose of illustration, rainstorms occurred on 18 May 2007 and 19 April 2008 are used herein.

On 18 May 2007, a trough of low pressure developed over inland Guangdong and moved across the south China coast. A squall line swept across Hong Kong from northwest to southeast that evening, bringing heavy showers and severe gusts to the territory. A peak gust over 100 km/h was recorded at Waglan Island. With the trough of low pressure lingering along the south China coast, there were heavy showers and thunderstorms between 19 and 22 May.

Under the combined effect of Typhoon Neoguri and the northeast monsoon, local winds started to pick up on 18 April 2008. Local winds became generally strong on the afternoon of 19 April 2008 when Neoguri was about 200 km to the west-southwest of Hong Kong. When Neoguri approached the coast of western Guangdong, the warm southerly winds associated with Neoguri met the relatively cooler northeast monsoon and formed
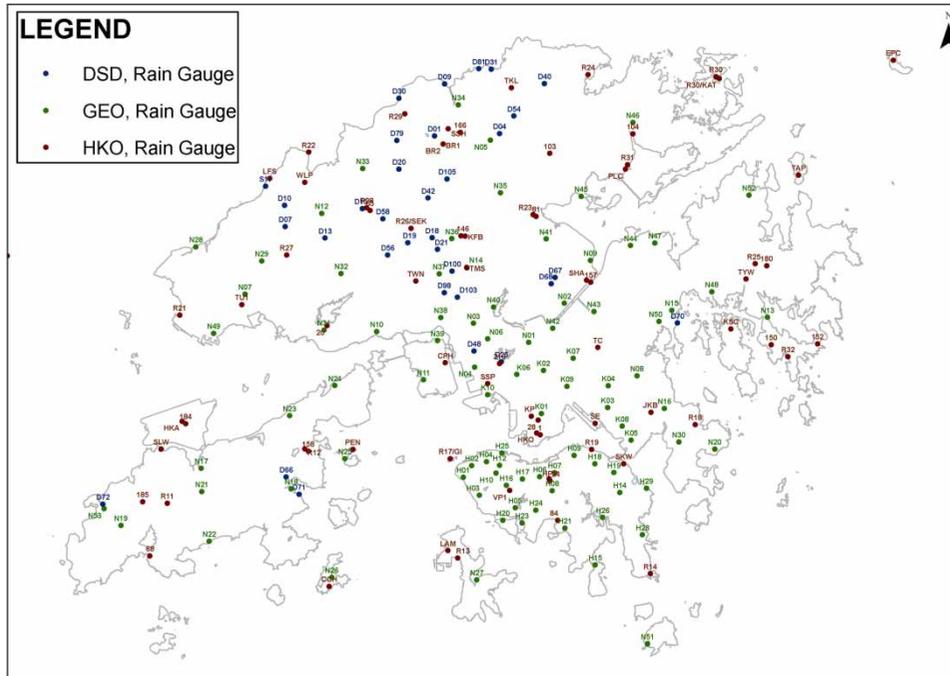
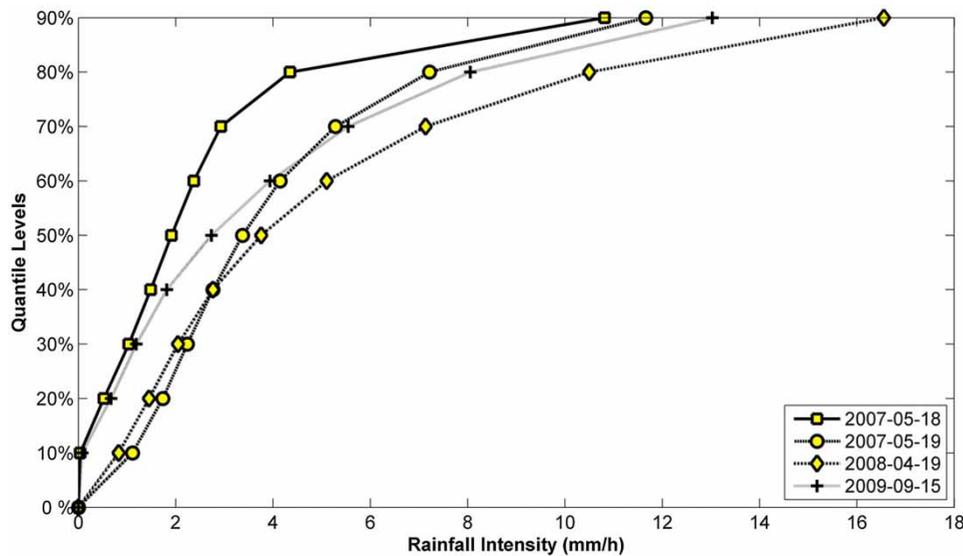**Figure 1** │ Location of raingauge stations in Hong Kong.

Figure 2 │ Quantiles of the distribution of the rain rate values for the four studied rainstorm events.

a warm front with severe convective activities over the coastal waters of Guangdong. The warm front moved from south to north across the coast of Guangdong and brought heavy rain to Hong Kong on that day. The total daily rainfall recorded at the HKO on that day was 233.4 mm, the highest daily rainfall amount recorded in April since records began. After making landfall and moving further inland, Neoguri weakened rapidly. Locally, rain also eased off rapidly with just a few showers on 20 April 2008.

METHODOLOGY

IK scheme

Theoretically, the variogram estimation needs to be carried out for each time step. Since anisotropy is considered in this study, the indicator variograms of rain field at a given time instant are found to fit an exponential model for 24 different directions (Liu & Tung 2015). However, due to the sample error, the parameters of the variogram at a given time instant may vary erratically even in adjacent directions. This sampling variability may induce erratic results in the kriging estimations. Figures 3 and 4 show the examples of estimated ranges of indicator variogram for different directions for each selected rainstorm event. Hence, to capture the essential feature of anisotropic behavior of spatial correlation while avoiding the erratic remove its sampling fluctuations, which can be expressed as the path of a point:

$$\begin{cases} X(\omega) = X_{center} + A\cos\omega\cos\varphi - B\sin\omega\sin\varphi \\ Y(\omega) = Y_{center} + A\cos\omega\sin\varphi + B\sin\omega\cos\varphi \end{cases} \tag{1}$$

where the angle parameter ω varies from 0 to 2π. In Equation (1), (X_{center}, Y_{center}) is the center of the ellipse which is set to

Figure 4 │ Example of estimated ranges of indicator variogram at the threshold of 10 mm/h for different directions at 20:00 pm during the 2008-04-18 rainstorm event.

be (0, 0) during the fitting procedure; A and B are, respectively, the semi-major and semi-minor axes of the ellipse and φ is the angle between the x-axis and the major axis of the ellipse.

Then, the value of the semi-variogram parameter (e.g., range a for direction ω) can be derived from the corresponding $X^*(\omega)$ and $Y^*(\omega)$ as follows:

$$a^*(\omega) = 2 \times \sqrt{X^*(\omega)^2 + Y^*(\omega)^2} \tag{2}$$

To determine the optimal parameters of the elliptic model (A, B, φ), the average square of distance (ASD) between the original estimated range for each direction and the corresponding elliptic-fitted value is minimized:

$$\text{ASD}(A, B, \varphi) = \frac{\sum_{i=1}^{24} [a^*(\theta_i, A, B, \varphi) - a(\theta_i)]^2}{24} \tag{3}$$

where $a(\theta_i)$, $i = 1, \ldots, 24$ is the original estimated range for the considered 24 directions and $a^*(\theta_i, A, B, \varphi)$, $i = 1, \ldots, 24$ is the elliptic-fitted value of the range.
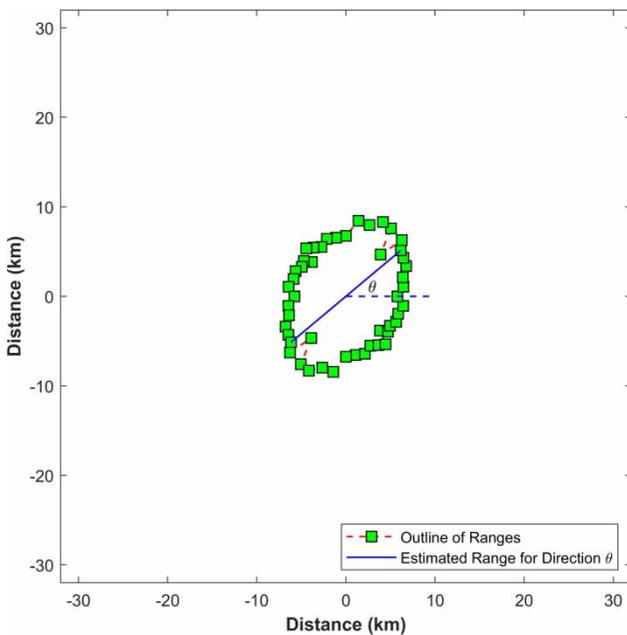

Figure 3 │ Example of estimated ranges of indicator variogram at the threshold of 5 mm/h for different directions at 16:30 pm during the 2007-05-18 rainstorm event.
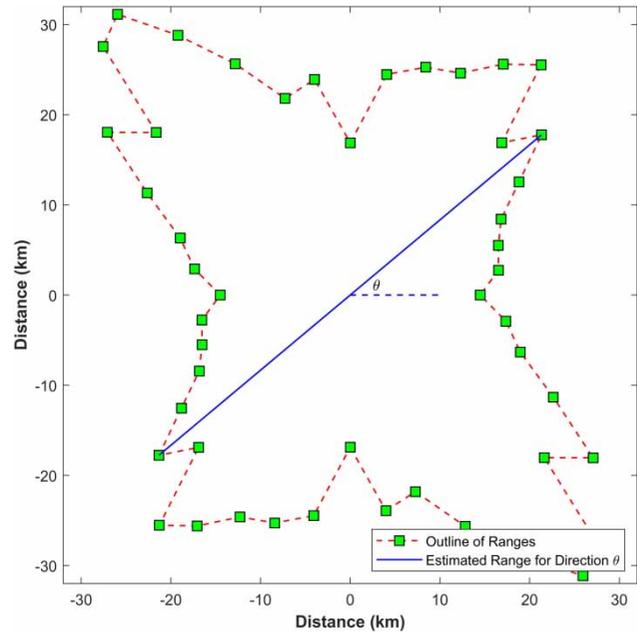
By IK, the actual rain process is transformed into a binary process to distinguish whether the rainfall intensity $R(x)$ at location x at time t is above a special threshold k or not. IK does not aim at estimating the indicator transform $i(x, k)$ set to 1 or 0 but provides an estimate of the cumulative distribution function (CDF) value conditional on threshold level k. A linear estimate of the conditional probability that $R(x_0)$ greater than a given threshold k from the sample data $R(x_\alpha)$, $\alpha = 1, \ldots, n$, which is also the mean of $i(x_0, k)$, is given by the following equation:

$$
\begin{aligned}
E[i^*_{\mathrm{IK}}(x_0, k)] &= 1 \times \Pr\{i(x_0, k) = 1|(n)\} + 0 \times \Pr\{i(x_0, k) = 0|(n)\} \\
&= \Pr\{R(x_0) > k|(n)\} \\
&= \sum_{\alpha=1}^{n} \lambda^{\mathrm{IK}}_\alpha(x_0, k) i(x_\alpha, k)
\end{aligned} \tag{4}
$$

Denoting the covariance of the indicator variables for the threshold level k at location x_τ and x_ν by $C_I(k, x_\tau - x_\nu)$, the ordinary IK weights $\lambda^{\mathrm{IK}}_\alpha(x_0, k)$, $\alpha = 1, \ldots, n$ are obtained by solving the following OK system as follows:

$$
\begin{cases}
\displaystyle\sum_{\beta=1}^{n} \lambda^{\mathrm{IK}}_\beta(x_0, k) C_I(k, x_\alpha - x_\beta) = C_I(k, x_\alpha - x_0), & \alpha = 1, \ldots, n \\[2mm]
\displaystyle\sum_{\beta=1}^{n} \lambda_\beta(x_0, k) = 1
\end{cases} \tag{5}
$$

where $C_I(k, x_\alpha - x_\beta)$, $\alpha = 1, \ldots, n$; $\beta = 1, \ldots, n$ is the $n \times n$ matrix of data-to-data covariance; $C_I(k, x_\alpha - x_0)$, $\alpha = 1, \ldots, n$ is the $n \times 1$ vector of covariance between the data and the target and $\lambda^{\mathrm{IK}}_\beta(x_0, k)$, $\beta = 1, \ldots, n$ is the $n \times 1$ vector of unknown weights to be solved.

Since 1983 when IK was introduced by Journel (1983), it has grown to become a widely applied estimation technique. The method offers a non-parametric, distribution-free solution to the spatial interpolation problem of a variable at an unsampled location. IK has a number of advantages: (1) it does not require the prior knowledge about the spatial distribution of the original sample data; (2) it can handle moderate mixing of diverse sample populations and is resistant to outliers; (3) it takes into account the structure of indicators at each considered threshold; (4) it provides full information of the CDF at the selected thresholds and (5) it provides an estimation variance.

Despite its wide use, IK suffers from several limitations and impediments (Emery & Ortiz 2013), some practical issues need to be addressed which include issues associated with kriging neighborhood and the order relations problem and their correction.

1. Selection of kriging neighborhood

The kriging theory is always derived as if all sample data points were used in the estimation; it is the so-called global neighborhood case. The minimum mean squared error is achieved when all the sample data are included. If a smaller neighborhood is chosen, the estimation can be considered as a constrained optimization with zero weights assigned on the discarded points. However, a global neighborhood may result in a kriging matrix that is too large to be numerically inverted. Another concern is that the geostatistical model may only be valid over short distances. The experimental variogram is generally accurate for small distance lag, and the uncertainties grow rapidly as the separation distance gets large.

The other alternative is called 'moving neighborhood', and the point selection is restricted to a subset of the sample data which is changing with the estimated point. In this study, only the sample data that lie within the preselected distance δ_θ from the point of estimation are counted as conditioning in kriging, and this distance δ_θ is a function of direction θ because of the anisotropy (Figure 5). For the exponential variogram model, Gilgen (2006) suggested not to interpolate for distances larger than the range.

2. Problem of the order relation

The IK algorithm itself does not necessarily ensure that the resulting probability estimates satisfy non-decreasing (or non-increasing) relationships between the threshold level k and non-exceedance probability (or exceedance probability). Let

$$
\Pr\{R(x) > k|(n)\} = \Pr\{i(x, k) = 1|(n)\} = \bar{F}\{x, k|(n)\} \tag{6}
$$

be the exceedance probability of a continuous variable $R(x) > k$ and the corresponding indicator variable $i(x, k) = 1$. Then, for two threshold levels $k' > k$, the legitimate
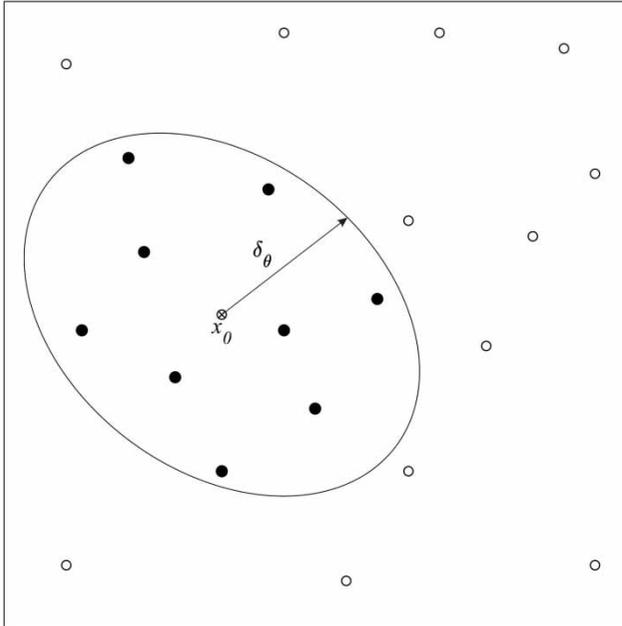
Figure 5 │ Illustration of kriging with moving neighborhood with the solid circle indicating the conditioning sample data used to estimate $R(x_0)$ and the empty circle indicating the discarded points.

exceedance probability relationship should satisfy

$$\bar{F}\{R(x), k'|(n)\} \leq \bar{F}\{x, k|(n)\} \tag{7}$$

However, the violation of such order relation could happen during the estimation.

There are two sources that could cause the violation of order relation: negative IK weights and insufficient number of data in some classes. Practice has shown that the majority of order relation problems are due to the lack of sufficient data, more precisely, to cases when IK is attempted at a cutoff k_s which is the lower bound of a class $[k_s, k_{s+1})$ that contains no data. In such a case, the indicator data set is the same for both cut-off levels k_l and k_{s+1}, and yet, the corresponding indicator variogram models are likely to be different. Therefore, the resulting exceedance probability values will likely be different with a good chance of violating the non-increasing relationship.

A correction of the IK-returned exceedance probability value is necessary. Correcting for order relation problems is quite delicate because of the ordering of the cumulative indicators. The following correction algorithm considers

the average of an upward and downward correction (Deutsch & Journel 1998): The upward correction starts with the lowest cut-off level k_1. If the IK-returned exceedance probability value $\bar{F}\{x, k_1|(n)\}$ is not within $[0, 1]$, reset it to the closest bound. Proceed to the next cut-off level k_2. If the IK-returned exceedance probability value $\bar{F}\{x, k_2|(n)\}$ is not within $[0, \bar{F}\{x, k_1|(n)\}]$, reset it to the closest bound. A loop through all remaining cut-off levels k_s, $s = 3, \ldots, S$ with S being the total number of cut-off levels. The downward correction starts with the largest cut-off level k_S. If the IK-returned exceedance probability value $\bar{F}\{x, k_S|(n)\}$ is not within $[0, 1]$, reset it to the closest bound. Proceed to the next cut-off level k_{S-1}. If the IK-returned exceedance probability value $\bar{F}\{x, k_{S-1}|(n)\}$ is not within $[\bar{F}\{x, k_S|(n)\}, 1]$, reset it to the closest bound. A loop downward through all remaining cut-off levels k_s, $s = S - 2, \ldots, 1$. Average the two sets of corrected exceedance probability values.

Since the exceedance probabilities for different threshold levels are available by IK, it is possible to estimate the mean rainfall intensity at all unsampled locations within the studied domain. The CDF value for the rainfall intensity at each unsampled location can be obtained from the exceedance probability by the following equation:

$$F(x, k_s) = 1 - \bar{F}(x, k_s), \ s = 1, 2, \ldots, S \tag{8}$$

where $\bar{F}(x, k_s), \ s = 1, 2, \ldots, S$ is the exceedance probability at cutoff k_s for location x gained by IK.

Various estimates for the unknown value $R(x)$ can be derived from the conditional CDF, e.g., expectation (E-type), median (M-type) or a maximum probability type (P-type) (Bierkens & Burrough 1993). In this study, the E-type estimate $R_E^*(x)$ is used to estimate the performance IK and is defined as follows:

$$R_E^*(x) = \int_0^\infty R(x) dF^*(x) \tag{9}$$

where $F^*(x)$ is the estimated CDF for the rainfall intensity at location x. The standard deviation $sd^*(x)$ of rainfall intensity can also be calculated to investigate the corresponding

uncertainties:

$$sd^{*2}(x) = \int_0^\infty [R(x) - R_E^*(x)]^2 dF^*(x) \tag{10}$$

To evaluate this integral, the procedures for interpolation between the IK-derived conditional CDF values are critical. Knowing a conditional CDF at several cut-off values, the integral can be estimated by assuming the distribution of the rainfall intensity at the considered location. In this study, several commonly used parametric probability distribution functions are used to fit the IK estimated CDF at different cut-off rainfall intensities, as well as a discrete approximation.

1. Discrete approximation

Discrete approximation (DA) is actually assuming linearity between two adjacent IK estimated CDF values. The mean value of DA-based rainfall intensity at location x can be estimated by the following equation:

$$R_E^*(x) = \sum_{s=1}^{S+1} R'(x) \times [F^*(x, k_s) - F^*(x, k_{s-1})] \tag{11}$$

where R' is the mean value of the class $(k_{s-1}, k_s]$ as obtained from the conditional CDF model.

The standard deviation can be estimated by the following equation:

$$sd^{*2}(x) = \sum_{s=1}^{S+1} [R'(x) - R_E^*(x)]^2 \times [F^*(x, k_s) - F^*(x, k_{s-1})] \tag{12}$$

where $k_s, \ s = 1, \ldots, S$ is the S cutoffs and $k_0 = 0$, $k_{S+1} = R_{max}$ are the minimum and maximum R-values.

2. Log-normal distribution

When a log-normal distribution (LN) is assumed for the random rainfall intensity at an unsampled location, its CDF corresponding to each cutoff $k_s, s = 1, 2, \ldots, S$, can be

expressed as follows:

$$F(x, k_s) = \Pr\{R(x) \le k_s\} = \Pr\{\ln R(x) \le \ln k_s\}$$
$$= \Pr\left\{\frac{\ln R(x) - \mu_{\ln R(x)}}{\sigma_{\ln R(x)}} \le \frac{\ln k_s - \mu_{\ln R(x)}}{\sigma_{\ln R(x)}}\right\} = \Phi\left(\frac{\ln k_s - \mu_{\ln R(x)}}{\sigma_{\ln R(x)}}\right) \tag{13}$$

where $R(x)$ is the value of rainfall intensity at location x; $\mu_{\ln R(x)}$ is the mean of $\ln R(x)$; $\sigma_{\ln R(x)}$ is the standard deviation of $\ln R(x)$ and $\Phi()$ is the standard normal CDF. $\mu_{\ln R(x)}$ and $\sigma_{\ln R(x)}$ can be estimated by linear regression of

$$\ln k_s = \mu_{\ln R(x)} + \sigma_{\ln R(x)} z_s \tag{14}$$

where $z_s = \Phi^{-1}[F(x, k_s)]$.

The mean value of LN-based rainfall intensity at location x can be estimated by the following equation:

$$R_E^*(x) = \mu_{R(x)} = \exp\left\{\mu_{\ln R(x)} + \frac{1}{2}\sigma_{\ln R(x)}^2\right\} \tag{15}$$

whereas the standard deviation of $R(x)$ can be estimated by the following equation:

$$sd^{*2}(x) = \sigma_{R(x)}^2 = \exp\{\sigma_{\ln R(x)}^2\} - 1 \tag{16}$$

3. Generalized extreme value distribution

When generalized extreme value (GEV) distribution is assumed for random rainfall intensity at each unsampled location, the CDF corresponding to each cutoff $k_s, s = 1, 2, \ldots, S$, can be expressed as follows:

$$F(x, k_s) = \Pr\{R(x) \le k_s\}$$
$$= \left\{\begin{array}{ll} \exp\left\{-\left[1 - \frac{\kappa_x(k_s - \xi_x)}{\alpha_x}\right]^{\frac{1}{\kappa_x}}\right\}, & \text{for } \kappa_x \ne 0 \\ \exp\left\{-\exp\left[-\frac{k_s - \xi_x}{\alpha_x}\right]\right\}, & \text{for } \kappa_x = 0 \end{array}\right. \tag{17}$$

where $\xi_x, \ \alpha_x, \ \kappa_x$ are, respectively, location, scale and shape parameters. Furthermore, $\xi_x + \frac{\alpha_x}{\kappa_x} \le x < \infty$ if $\kappa_x < 0$; $-\infty \le x < \infty$ if $\kappa_x = 0$ and $-\infty < x \le \xi_x + \frac{\alpha_x}{\kappa_x}$ if $\kappa_x > 0$.

The shape parameter κ_x governs the tail behavior of the distribution and the sub-family defined by $\kappa_x = 0$ corresponds to the Gumbel (or extreme value type I, EV1) distribution.

The inverse CDF is:

$$k_s(F) = \begin{cases} \xi_x + \dfrac{\alpha_x}{\kappa_x}\{1 - [-\ln F(x, k_s)]^{\kappa_x}\}, & \text{for } \kappa_x \neq 0 \\ \xi_x - \alpha_x \ln\{-\ln F(x, k_s)\}, & \text{for } \kappa_x = 0 \end{cases} \quad (18)$$

The mean value of GEV-based rainfall intensity at location x can be estimated as follows:

$$R_E^*(x) = \mu_{R(x)}$$
$$= \begin{cases} \xi_x + \dfrac{\alpha_x[1 - \Gamma(1 + \kappa_x)]}{\kappa_x}, & \text{for } \kappa_x \neq 0, \kappa_x > -1 \\ \xi_x + 0.5772\alpha_x, & \text{for } \kappa_x = 0 \end{cases} \quad (19)$$

and the standard deviation of $R(x)$ can be estimated by the following equation:

$$sd^{*2}(x) = \sigma_{R(x)}^2$$
$$= \begin{cases} \dfrac{\alpha_x[\Gamma(1 + 2\kappa_x) - \Gamma^2(1 + \kappa_x)]}{\kappa_x^2}, & \text{for } \kappa_x \neq 0, \kappa_x > -0.5 \\ \dfrac{\alpha_x \pi^2}{6}, & \text{for } \kappa_x = 0 \\ Not\ exists, & \text{for } \kappa_x \leq -0.5 \end{cases} \quad (20)$$

4. Gumbel

When Gumbel distribution is assumed for random rainfall intensity at each unsampled location, the CDF corresponding to each cutoff $k_s, s = 1, 2, \ldots, S$, can be expressed as follows:

$$F(x, k_s) = \Pr\{R(x) \leq k_s\} = \exp\left\{-\exp\left[-\dfrac{k_s - \mu_x}{\beta_x}\right]\right\} \quad (21)$$

where μ_x and β_x are, respectively, location and scale parameters.

The inverse CDF of Gumbel is:

$$k_s(F) = \dfrac{1}{\beta_x}\exp\left\{-\dfrac{k_s - \mu_x}{\beta_x} - \exp\left[-\dfrac{k_s - \mu_x}{\beta_x}\right]\right\} \quad (22)$$

The mean value of Gumbel-based rainfall intensity at location x can be estimated as follows:

$$R_E^*(x) = \mu_{R(x)} = \mu_x + \gamma_x \beta_x \quad (23)$$

where γ_x is Euler's constant, and the standard deviation of $R(x)$ can be estimated by the following equation:

$$sd^{*2}(x) = \sigma_{R(x)}^2 = \beta_x^2 \zeta(2) \quad (24)$$

where $\zeta(\cdot)$ is the Riemann zeta function.

5. Generalized Pareto distribution

When generalized Pareto (GPA) distribution is assumed for the random rainfall intensity at each unsampled location, the CDF corresponding to each cutoff $k_s, s = 1, 2, \ldots, S$, can be expressed as follows:

$$F(x, k_s) = \Pr\{R(x) \leq k_s\}$$
$$= \begin{cases} 1 - \left[1 - \dfrac{\kappa_x(k_s - \xi_x)}{\alpha_x}\right]^{\frac{1}{\kappa_x}}, & \text{for } \kappa_x \neq 0 \\ 1 - \exp\left\{-\dfrac{k_s - \xi_x}{\alpha_x}\right\}, & \text{for } \kappa_x = 0 \end{cases} \quad (25)$$

where $\xi_x, \alpha_x, \kappa_x$ are, respectively, location, scale and shape parameters. $\xi_x \leq x < \infty$ if $\kappa_x \leq 0$ and $\xi_x \leq x < \xi_x + \dfrac{\alpha_x}{\kappa_x}$ if $\kappa_x > 0$. When the shape parameter $\kappa_x = 0$, the GPA distribution is reduced to the Pareto (PA) distribution.

The inverse CDF of GPA is given as follows:

$$k_s(F) = \begin{cases} \xi_x + \dfrac{\alpha_x}{\kappa_x}[1 - (1 - F(x, k_s))^{\kappa_x}], & \text{for } \kappa_x \neq 0 \\ \xi_x - \alpha_x \ln\{1 - F(x, k_s)\}, & \text{for } \kappa_x = 0 \end{cases} \quad (26)$$

The mean value of GPA-based rainfall intensity at location x can be estimated as follows:

$$R_E^*(x) = \mu_{R(x)} = \xi_x + \frac{\alpha_x}{1 + \kappa_x} \tag{27}$$

and the standard deviation of $R(x)$ can be estimated by the following equation:

$$\text{sd}^{*2}(x) = \sigma_{R(x)}^2 = \frac{\alpha_x^2}{(1 + \kappa_x)^2 + (1 + 2\kappa_x)} \tag{28}$$

6. Generalized logistic distribution

When the generalized logistic (GLO) distribution is assumed for the random rainfall intensity at each unsampled location, for each cutoff k_s, $s = 1, 2, \ldots, S$, can be expressed as follows:

$$F(x, k_s) = \Pr\{R(x) \le k_s\} = \frac{1}{1 + \exp(-y)} \tag{29}$$

where

$$y = \begin{cases} -\dfrac{1}{\kappa_x} \ln\left\{1 - \dfrac{\alpha_x(k_s - \xi_x)}{\alpha_x}\right\}, & \text{for } \kappa_x \neq 0 \\[2mm] \dfrac{k_s - \xi_x}{\alpha_x}, & \text{for } \kappa_x = 0 \end{cases}$$

where $\xi_x, \alpha_x, \kappa_x$ are, respectively, location, scale and shape parameters. $\xi_x + \dfrac{\alpha_x}{\kappa_x} \le x < \infty$ if $\kappa_x < 0$; $-\infty \le x < \infty$ if $\kappa_x = 0$ and $\xi_x < x \le \xi_x + \dfrac{\alpha_x}{\kappa_x}$ if $\kappa_x > 0$. The GLO distribution is reduced to logistic (LO) distribution when the shape parameter κ_x is zero.

The inverse CDF of GLO is given as follows:

$$k_s(F) = \begin{cases} \xi_x + \dfrac{\alpha_x}{\kappa_x}\left[1 - \left(\dfrac{1 - F(x, k_s)}{F(x, k_s)}\right)^{\kappa_x}\right], & \text{for } \kappa_x \neq 0 \\[3mm] \xi_x - \alpha_x \ln\left\{\dfrac{1 - F(x, k_s)}{F(x, k_s)}\right\}, & \text{for } \kappa_x = 0 \end{cases} \tag{30}$$

The mean value of GLO-based rainfall intensity at location x can be estimated as follows:

$$R_E^*(x) = \mu_{R(x)} = \begin{cases} \xi_x + \dfrac{\alpha_x(1 - g_1)}{\kappa_x}, & \text{for } \kappa_x \neq 0 \\[2mm] \xi_x, & \text{for } \kappa_x = 0 \end{cases} \tag{31}$$

And the standard deviation can be estimated by the following equation:

$$\text{sd}^{*2}(x) = \sigma_{R(x)}^2 = \begin{cases} \dfrac{\alpha_x^2(g_2 - g_1^2)}{\kappa_x^2}, & \text{for } \kappa_x \neq 0 \\[2mm] \dfrac{\alpha_x^2 \pi^2}{3}, & \text{for } \kappa_x = 0 \end{cases} \tag{32}$$

where $g_r = \Gamma(1 - r\kappa_x)\Gamma(1 + r\kappa_x)$.

Statistical indices

Various estimates for the unknown value $R(x)$ can be derived from the conditional CDF, e.g., expectation (E-type), median (M-type) or a maximum probability type (P-type) (Bierkens & Burrough 1993). In this study, the E-type estimate $R_E^*(x)$ is used to estimate the performance IK and is defined as follows:

$$R_E^*(x) = \int_0^\infty R(x) dF^*(x) \tag{33}$$

where $F^*(x)$ is the estimated CDF for the rainfall intensity at location x. The standard deviation $\text{sd}^*(x)$ of rainfall intensity can also be calculated to investigate the corresponding uncertainties:

$$\text{sd}^{*2}(x) = \int_0^\infty [R(x) - R_E^*(x)]^2 dF^*(x) \tag{34}$$

To evaluate this integral, the procedures for interpolation between the IK-derived conditional CDF values are critical. Knowing a conditional CDF at several threshold values, the integral can be estimated by assuming the distribution of the rainfall intensity at the considered location. In this study, several commonly used parametric probability distribution functions, including LN distribution, GEV distribution, GPA distribution, GLO distribution, are used to fit the IK estimated CDF at different threshold rainfall intensities, as well as DA.

There are many factors affecting the performance of spatial interpolation methods, including sampling density, sample spatial distribution, sample clustering, surface type, data variance, data normality, quality of secondary

information, stratification and temporal scale. Among them, data variation is a dominant factor (Li & Heap 2011). Large data variation indicates a high heterogeneity in the data. Since the original rainfall data have transformed into binary variables to distinguish whether the rainfall intensity is higher than the considered threshold in this study, variation in data itself is not a concern that affects the performance of spatial interpolation methods. Herein, the interpolation performance, as well as the impact of the semi-variogram estimation on the interpolation performance, is evaluated. IK results, along with different distributional assumptions of the random rainfall intensity at unsampled locations, were used to estimate the mean rainfall intensity value. The estimated values of rainfall intensity within the domain of interest were compared with the observed values at every sample point using the following performance criteria.

BIAS measures the averaged differences between the estimated rainfall values with the reference to the observed values as follows:

$$\text{BIAS} = \frac{1}{n} \sum_{i=1}^{n} [R_E^*(x_i) - R(x_i)] \tag{35}$$

where $R_E^*(x_i)$ is the E-type estimate at location x_i; $R(x_i)$ the observed value at location x_i. Since the negative and positive estimates counteract each other, the resultant BIAS tends to be lower than the other performance criteria considered below. The positive-valued BIAS indicates that the estimated rainfall value on average is higher than that of the observed. It should not be used as an indicator of prediction accuracy.

The mean absolute error (MAE) summarizes the averaged absolute differences between the observed and estimated rainfall values:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |R_E^*(x_i) - R(x_i)| \tag{36}$$

The root mean squared error (RMSE) is a commonly used performance indicator of overall model prediction accuracy as compared to the observed values. It is more sensitive to outliers than the MAE as large errors exert heavier influence on the value of RMSE:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [R_E^*(x_i) - R(x_i)]^2} \tag{37}$$

The accuracy of estimated indicator variables (AIEV) is given as follows:

$$\text{AEIV} = \text{mean} \left\{ \frac{\sum_{i=1}^{n} 1 - abs[I^*(x_i, k) - I(x_i, k)]}{n} \right\} \tag{38}$$

where $I^*(x, k)$ is the indicator transformation of the estimated rainfall intensity for threshold k at location x; $I(x, k)$ is the indicator transformation of the observed rainfall intensity for threshold k at location x.

RESULTS AND DISCUSSION

To estimate the probability distribution as well as the mean rainfall intensity at an unsampled location within the studied domain, several thresholds need to be selected. At each threshold, the data are indicator-transformed and their resulting indicator variograms are used to describe the changes in spatial variability along with the CDF. Hence, threshold values need to be selected to provide the maximum possible resolution defining CDF. More thresholds, in theory, provide a better definition of the CDF, but the more intensive calculation is required.

In this study, six threshold values of rainfall intensity were chosen, based on a numerical experiment indicating that increasing the number of thresholds brings only a minor additional reduction in estimation accuracy. The threshold value k was set to be 1, 2, 3, 5, 10 and 20 mm/h, respectively, for modeling the prior distribution.

Isotropic variogram

One important problem for geostatistical interpolation of rain field is the effective and reliable estimation of the variogram for each time step. Variograms were inferred from radar rainfall data because of the higher resolution in space compared to the recording raingauge stations. To investigate the influence of variogram estimation on interpolation performance,

isotropic experimental semi-variograms are estimated for every 6-min time step under the considered six threshold levels during each selected rainstorm event. The automatically estimated parameters (e.g., range and sill) of the isotropic variogram for an exponential model for the two selected rainstorm events are shown in Figures 6 and 7.

There are some features that can be noticed from the estimated parameters of the isotropic variograms for the selected rainstorm events.

Most of the sills of indicator variograms at all considered cut-off levels and time steps are ranging from 0 to 0.4.

For the 2007-05-18 rainstorm event, it is clear that the values of the range and sill have a negative correlation with the rainfall intensity thresholds. The larger the threshold, the smaller the range and sill are. Situations are different for the 2008-04-19 rainstorm event from the 80th 6-min to the 200th 6-min. This correlation offers information about the number of the rainstorm centers. Confirmed from the rainfall intensity images of the 2008-04-19 rainstorm event, there is indeed more than one rainstorm center during those time periods, which makes the values of the variogram parameters for lower cut-off levels (e.g., 1 and 2 mm/h) go down, even to the smallest. 'Rainstorm centers'

here are referring to the areas where the rainfall intensity is greater than 10 mm/h. The instantaneous rain field of the 2007-05-18 rainstorm event at the 33rd 6-min shown in Figure 8 clearly indicates that the only one rainstorm center is located in the northwest. For the 160th 6-min of the 2008-04-19 rainstorm event (see Figure 9), the rainstorm centers scatter at several different locations. This result consists of the information obtained from the correlation between the mean length and range of the indicator variogram (Liu & Tung 2015). Because of the existence of multiple rainstorm centers, the correlation coefficients of the mean length and range for the 2008-04-19 rainstorm event are slightly lower than those for the other two considered rainstorm events.

Anisotropic variogram with elliptic-fitted parameters

The spatial structure of rain field at each threshold level reveals strong anisotropy. Hence, the indicator variograms are fitted with the exponential variogram model for 24 directions at each time step. The variogram parameters (e.g., range and sill) are expected to show its directional association and tend to be similar if the
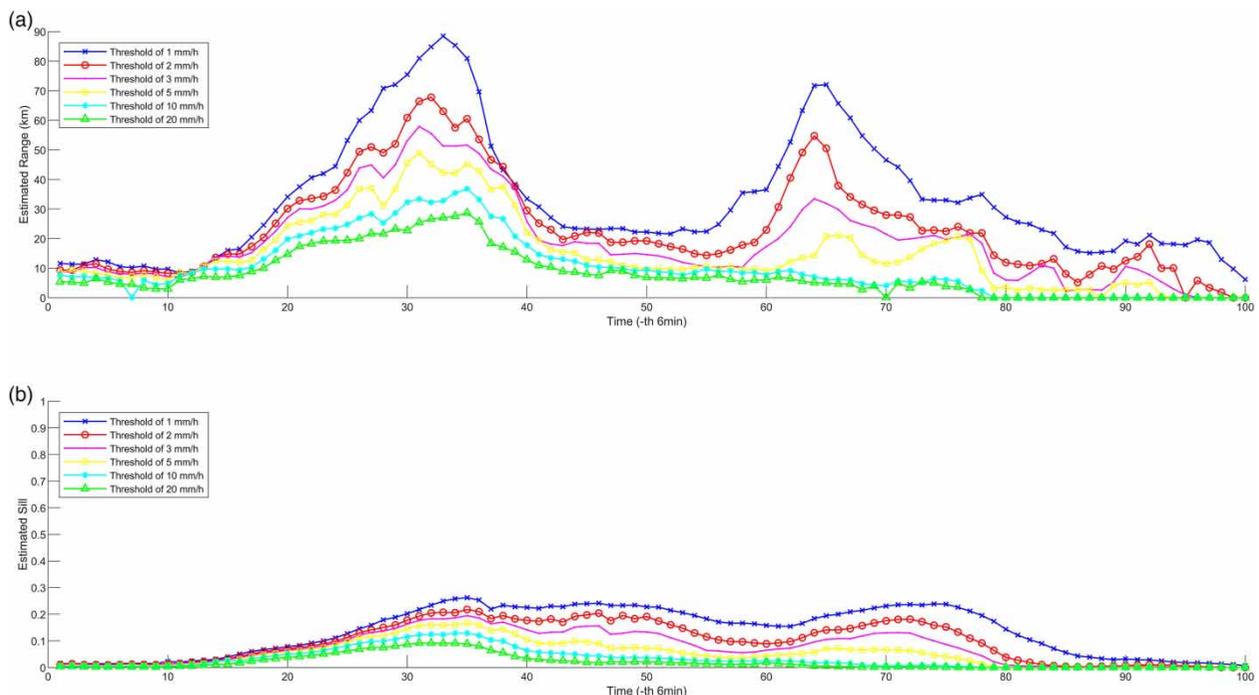

Figure 6 | Time variation of estimated variogram parameters ((a) range and (b) sill) of the exponential model under the isotropic condition for the 2007-05-18 rainstorm event.
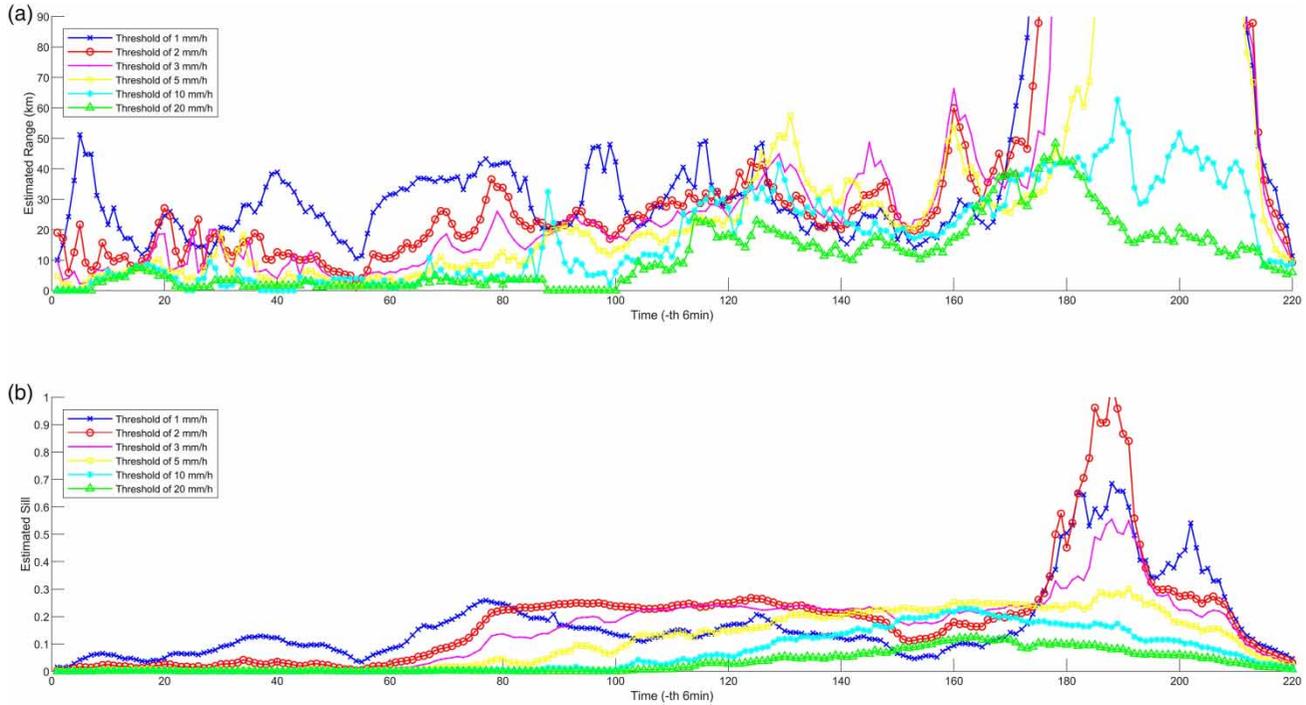
Figure 7 │ Time variation of estimated variogram parameters ((a) range and (b) sill) of the exponential model under the isotropic condition for the 2008-04-19 rainstorm event.

directions are close. However, the presence of sampling variation could result in significant variability of variogram parameters with the direction even the two considered directions only differ by about 10°. Herein, the parameters of anisotropic variograms are fitted with

an elliptic model. Figures 10 and 11 show the examples of sample estimated and elliptic-fitted ranges of anisotropy variogram for the selected rainstorm events. The procedure of fitting the variogram parameters with the elliptic model removes the fluctuation of the original
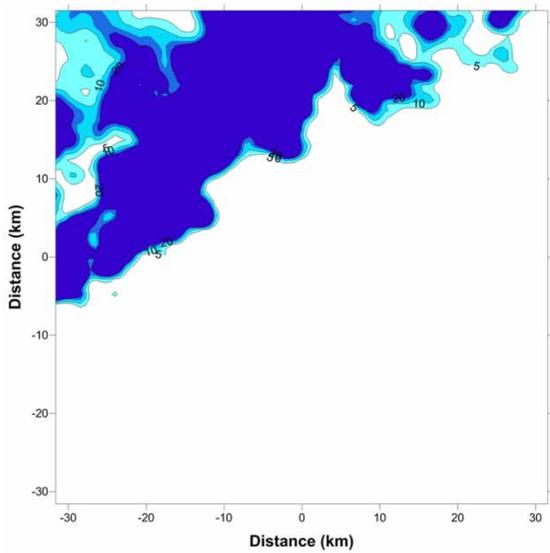

Figure 8 │ Instantaneous rainfall intensity contour (in mm/h) over the study area during the 2007-05-18 rainstorm event at 33rd 6-min.
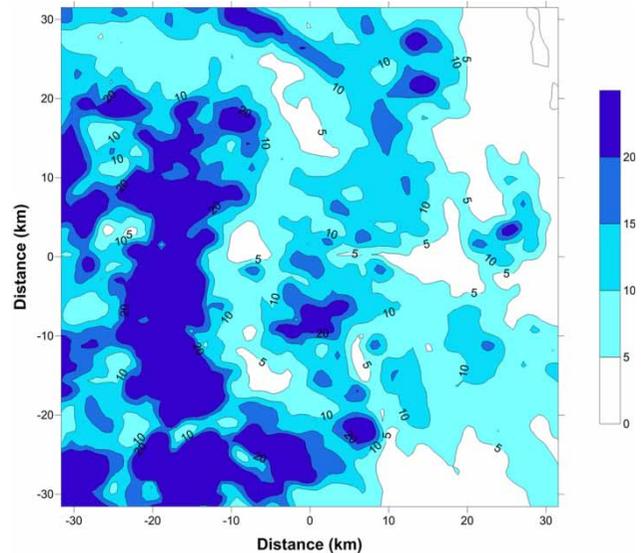

Figure 9 │ Instantaneous rainfall intensity contour (in mm/h) over the study area during the 2008-04-19 rainstorm event at 160th 6-min.

parameters that are directly estimated from the theoretical variogram model for 24 directions. In addition, the parametric representation of variogram parameters by the elliptic model reduces the number of the parameters of the model from 48 (both ranges and sills for all 24 considered directions at one time) to 6 (the biggest range and sill, the smallest range and sill and the angle of the biggest range and sill).

IK is used to analyze the probability that the rainfall intensity is greater than the considered threshold for the unsampled location and the non-decreasing CDF values at the predetermined threshold levels are available. Eight different interpolation schemes are tested to fit the estimated CDF from IK by the least-square method and to estimate the mean rainfall intensity within the study area (see examples in Figure 12).

Discussion

The results of the performance of IK and mean rainfall intensity estimation (compared with the 132 observed rain-gauge estimations) are shown in Tables 1 and 2 under different assumptions about the semi-variogram model and different probability distributions of the CDF.
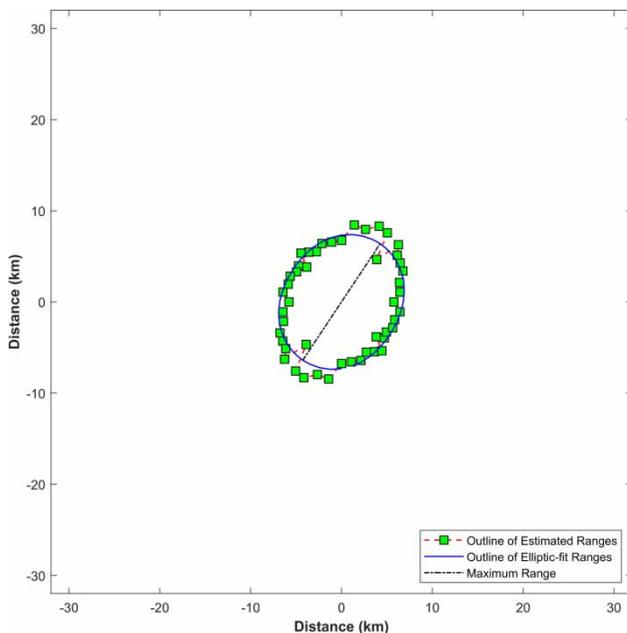
Figure 10 │ Sample and elliptic-fitted ranges of anisotropy variogram for the threshold of 5 mm/h at 18:30 pm during the 2007-05-18 rainstorm event.
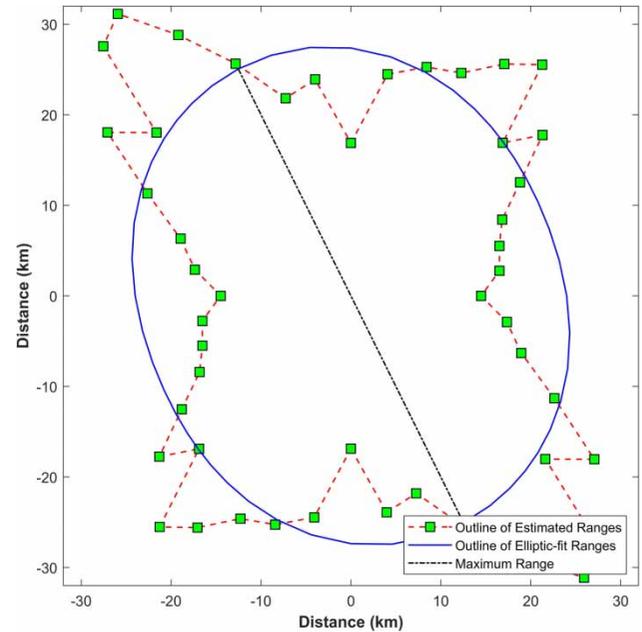
Figure 11 │ Sample and elliptic-fitted ranges of anisotropy variogram for the threshold of 10 mm/h at 20:00 pm during the 2008-04-19 rainstorm event.

Comparing the results based on different variogram models, it is clear that IK based on the variogram under isotropic assumption shows the worst performance for most times. This result signifies the necessity of capturing the anisotropy feature of rain field although it is more computationally intensive and complex to fit the variogram model for 24 different directions. The other finding is that the IK based on the anisotropic variogram with strictly sampled estimated parameters and that of elliptic-fitted parameters show similar performances and the former one leads to slightly better results. Without sacrificing too much in the accuracy, the sampling fluctuation of the variogram parameters for 24 directions can be removed by fitting with an elliptic model and the number of the parameters is reduced from 48 to 6.

It can be seen that the estimation on the value of the rainfall intensity is not always satisfactory. However, performance varies from one rainstorm event to another. The worst performance was found during the 2008-04-19 rainstorm event, the values of BIAS, MAE and RMSE are around -4.1, 9.0 and 21.8 mm/h, respectively. IK shows reasonable performance for the 2007-05-18 rainstorm event (BIAS: -1.3 mm/h, MAE: 2.3 mm/h and RMSE: 6.3 mm/h). This accuracy of estimation can be improved
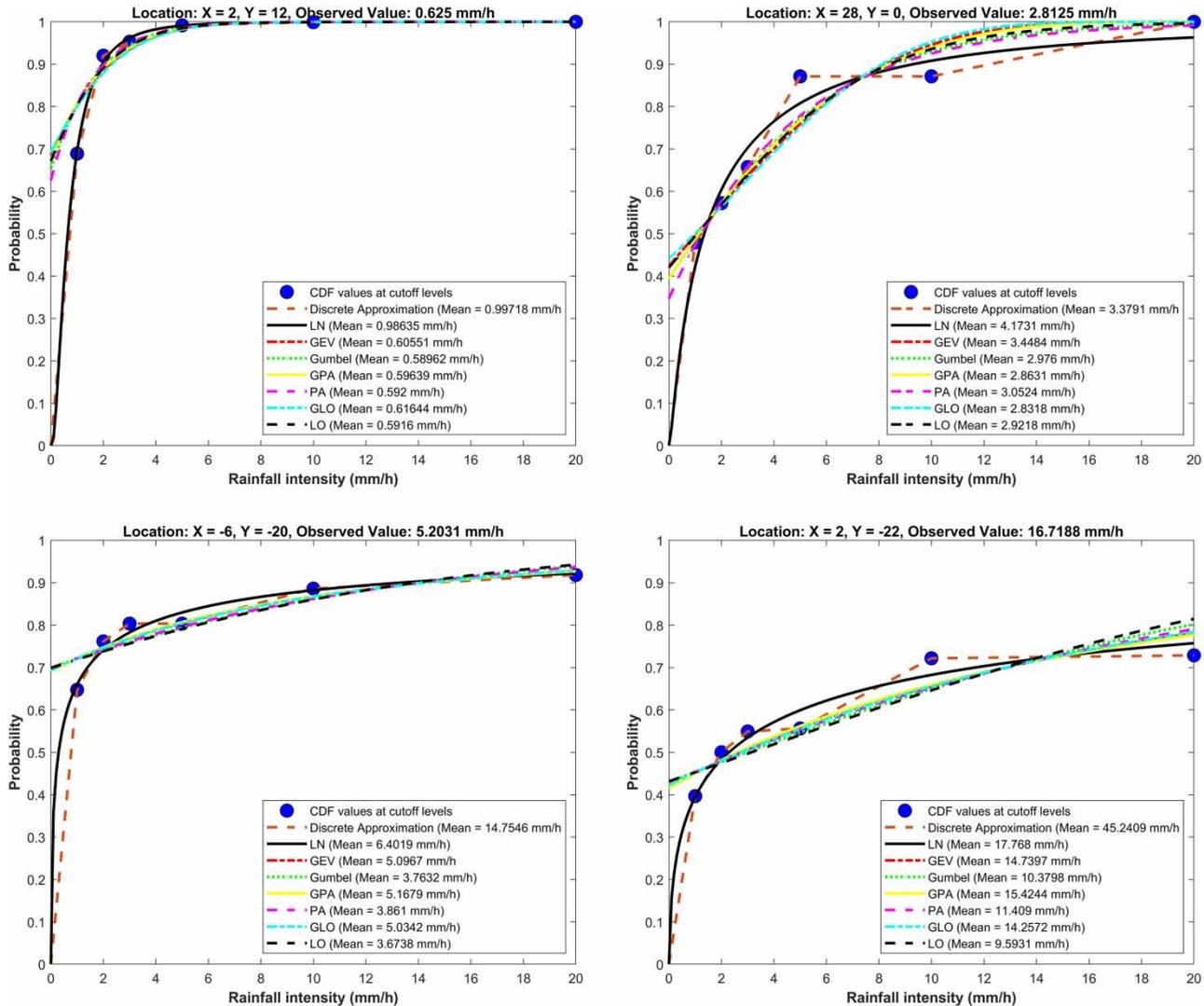
Figure 12 │ Examples of curve fitting for CDF at four different locations for the 2007-05-18 rainstorm event (18:30 pm).

by increasing the number of the thresholds, but at the same time, the indicator variograms associated with the new thresholds need to be calculated which is very time-consuming. The other performance measure, AEIV, shows that, although the estimated mean rainfall intensity value sometimes may not be that precise, most of the times the estimation is located in the right range of the threshold levels. It can be concluded that IK can usually adequately provide the probability that an unknown rainfall intensity at an unsampled location is greater than the given threshold value (or the value of the indicator variable), and this information is the basis for sequential indicator simulation and

critical decisions in agriculture, hydrology, hydraulic structural design and other areas.

For further analysis, the BIAS, MAE and RMSE are calculated for each threshold range of the rainfall intensity. Most contribution of the error comes from the large range with rainfall intensity greater than 10 mm/h.

In general, the three-parameter distribution models usually show slightly better performance than the two-parameter distributions. Fitting the CDF with a GPA distribution outperformed the other six distribution models considered according to the values of BIAS, MAE, RMSE and AEIV.

Table 1 │ Example of performance assessment of IK based on sample isotropic and anisotropic variograms and elliptic-fitted anisotropic variogram at 18:30 pm during the 2007-05-18 rainstorm event

Distribution	IK based on sample isotropic variogram	IK based on sample anisotropic variogram	IK based on elliptic-fitted anisotropic variogram	Distribution	IK based on sample isotropic variogram	IK based on sample anisotropic variogram	IK based on elliptic-fitted anisotropic variogram
	BIAS				**RMSE**		
DA	1.992	−0.529	0.526	DA	15.682	6.439	8.951
LN	−0.462	−0.932	−0.526	LN	6.910	6.291	6.556
GEV	−1.282	−1.907	−1.567	GEV	9.017	6.319	6.371
Gumbel	−0.293	−1.387	−1.132	Gumbel	9.561	6.707	6.958
GPA	−0.336	−1.333	−0.928	GPA	8.847	6.296	6.403
PA	−0.198	−1.356	−1.059	PA	8.791	6.336	6.505
GLO	−1.053	−1.528	−1.199	GLO	9.037	6.325	6.345
LO	−0.343	−1.392	−1.174	LO	9.934	6.44	6.655
	MAE				**AEIV**		
DA	4.926	2.577	3.468	DA	0.847	0.828	0.830
LN	2.707	2.443	2.632	LN	0.847	0.844	0.840
GEV	3.141	2.297	2.374	GEV	0.847	0.852	0.850
Gumbel	3.334	2.431	2.64	Gumbel	0.818	0.837	0.825
GPA	3.250	2.298	2.413	GPA	0.848	0.852	0.855
PA	3.089	2.313	2.478	PA	0.843	0.851	0.849
GLO	3.063	2.297	2.348	GLO	0.848	0.851	0.850
LO	3.307	2.363	2.542	LO	0.821	0.845	0.834

Figures 13 and 14 compare the observed rainfall intensity with that of the estimated value under the assumption of GPA distribution for the CDF. From the standard deviation map, it is clear to see that large uncertainty in estimating unknown rainfall intensity at unsampled location is associated with the large rainfall intensity. The AIEV is high when the rainfall intensity exceeds 5 mm/h. From the results of BIAS, it can be seen that overestimation often occurs when estimating the lower range rainfall intensity, while it tends to underestimate in the range of higher rainfall intensity.
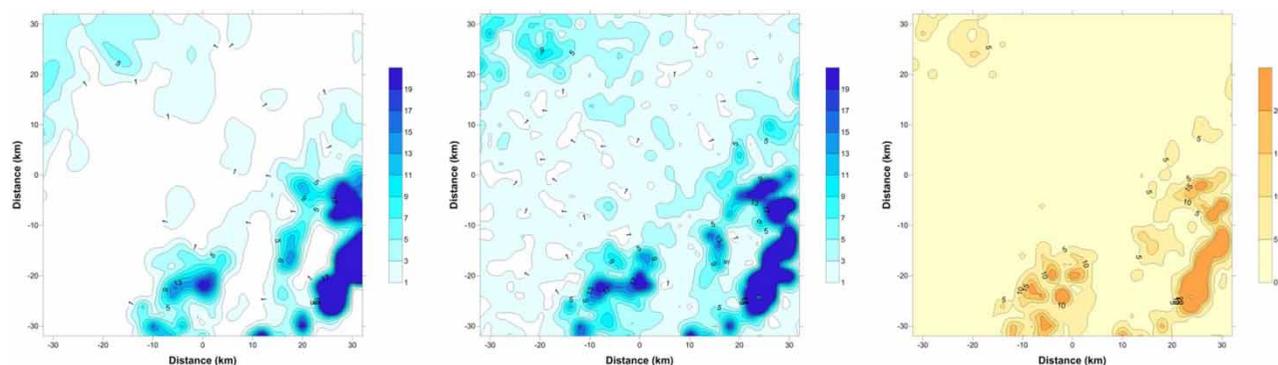
CONCLUSIONS

IK has the capability of estimating the probability that an unknown quantity at an unsampled location is greater than a given threshold value by capturing the proportion of neighboring data valued above the considered threshold.

In this study, IK is used to estimate the exceedance probabilities for different rainfall intensity threshold levels at all unsampled locations within the studied domain. Six rainfall intensity thresholds are chosen based on a numerical experiment, indicating that increasing the number of thresholds brings only a minor additional improvement in estimation accuracy. A procedure for the interpolation between the IK-derived conditional CDF values is implemented.

The performance accuracy of spatial interpolation results as well as influences of different variogram models (isotropic and anisotropic) and different fitted schemes for the estimated conditional CDF are studied. Under the isotropy assumption, values of the exponential variogram model parameters (e.g., range and sill) have a negative correlation with the rainfall intensity threshold except for the time periods when multiple rainstorm centers exist. When different time resolutions (6-min, 12-min, 30-min and 60-min) are considered, the time series of isotropic variogram parameters for the four-time resolutions share a similar

Table 2 │ Example of performance assessment of IK based on sample isotropic and anisotropic variograms and elliptic-fitted anisotropic variogram at 20:00 pm during the 2008-04-19 rainstorm event

Distribution	IK based on sample isotropic variogram	IK based on sample anisotropic variogram	IK based on elliptic-fitted anisotropic variogram	Distribution	IK based on sample isotropic variogram	IK based on sample anisotropic variogram	IK based on elliptic-fitted anisotropic variogram
	BIAS				**RMSE**		
DA	28.089	16.345	18.327	DA	48.423	25.197	26.179
LN	1.247	−2.311	−0.373	LN	22.139	20.396	19.622
GEV	− 0.878	−3.358	−1.604	GEV	20.916	20.783	19.546
Gumbel	0.657	−5.690	−4.380	Gumbel	20.973	22.321	21.116
GPA	1.359	−2.810	−1.077	GPA	21.359	20.440	19.173
PA	2.293	−4.925	−3.366	PA	21.626	21.639	20.226
GLO	− 0.909	−3.593	−1.922	GLO	20.048	21.035	19.740
LO	− 1.822	−6.177	−5.050	LO	21.664	22.787	21.753
	MAE				**AEIV**		
DA	29.971	19.937	21.594	DA	0.584	0.856	0.595
LN	10.282	9.712	10.140	LN	0.699	0.878	0.722
GEV	9.910	9.338	9.644	GEV	0.708	0.841	0.721
Gumbel	10.563	7.777	10.031	Gumbel	0.703	0.916	0.709
GPA	9.773	9.773	9.393	GPA	0.717	0.925	0.721
PA	10.057	9.768	9.421	PA	0.713	0.902	0.716
GLO	10.041	8.486	9.703	GLO	0.710	0.855	0.718
LO	10.931	7.728	10.452	LO	0.698	0.916	0.702

Figure 13 │ Example of observed rainfall intensity (left), estimated mean rainfall intensity (middle) and standard deviation of the estimation (right) in mm/h under the assumption of GPA distribution at 18:30 pm during the 2007-05-18 rainstorm event.

trend. IK based on these isotropic variograms shows the worst performance for most times. While IK based on the anisotropic variogram with strictly sampled estimated parameters and that of elliptic-fitted parameters show similar performances. This confirms the necessity of capturing the anisotropy feature of rain field and the procedure of fitting anisotropic variogram parameters with an elliptic model.

Fitting the conditional CDF values at the selected threshold levels with a GPA model shows slightly
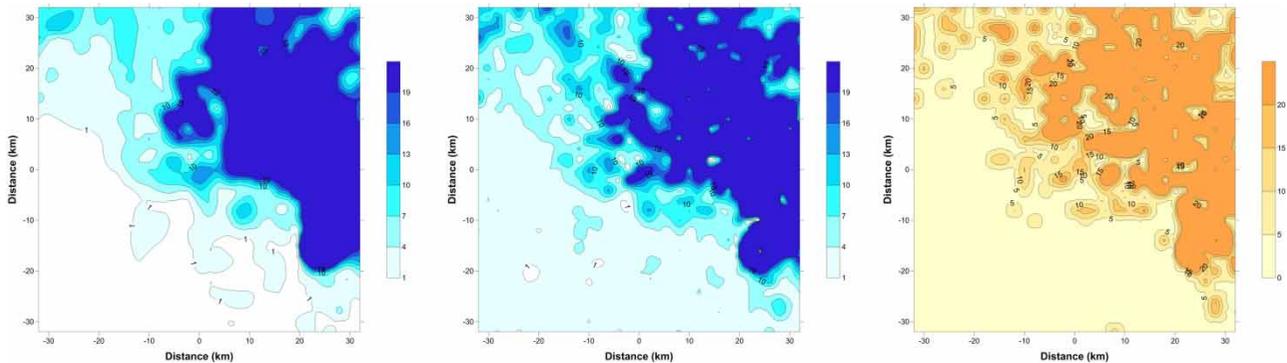
Figure 14 │ Example of observed rainfall intensity (left), estimated mean rainfall intensity (middle) and standard deviation of the estimation (right) in mm/h under the assumption of GPA distribution at 20:00 pm during the 2008-04-19 rainstorm event.

better performance when estimating the mean rainfall intensity. The corresponding AEIV is the highest among the eight considered distributions for most considered rainstorm events. The proposed IK method tends to overestimate the rainfall intensity values in a lower range and underestimate them in a higher range. And the performance varies from one rainstorm event to another. The standard deviation maps indicate that there is high uncertainty associated with the large rainfall intensity values. However, it can provide a good estimation of an indicator variable for rainfall intensity based on the predetermined thresholds, which can serve a good basis for sequential indicator simulation and critical decisions in agriculture, hydrology, hydraulic structural design and other areas.

ACKNOWLEDGEMENTS

REFERENCES

Abtew, W., Obeysekera, J. & Shih, G. 1993 Spatial analysis for monthly rainfall in south Florida. *JAWRA Journal of the American Water Resources Association* **29** (2), 179–188.

Barancourt, C., Creutin, J. & Rivoirard, J. 1992 A method for delineating and estimating rainfall fields. *Water Resources Research* **28** (4), 1133–1144.

Barnes, S. L. 1973 Mesoscale Objective Map Analysis Using Weighted Time Series Observations. NOAA Tech Memo, NSSL, pp. 62–68.

Bierkens, M. & Burrough, P. 1993 The indicator approach to categorical soil data. *Journal of Soil Science* **44** (2), 361–368.

Ciach, G. J. & Krajewski, W. F. 1999 On the estimation of radar rainfall error variance. *Advances in Water Resources* **22** (6), 585–595.

Cole, S. J. & Moore, R. J. 2008 Hydrological modelling using raingauge- and radar-based estimators of areal rainfall. *Journal of Hydrology* **358** (3), 159–181.

Cressman, G. P. 1959 An operational objective analysis system. *Monthly Weather Review* **87** (10), 367–374.

Creutin, J. & Obled, C. 1982 Objective analyses and mapping techniques for rainfall fields: an objective comparison. *Water Resources Research* **18** (2), 413–431.

Deutsch, C. & Journel, A. 1998 *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York.

Dirks, K., Hay, J., Stow, C. & Harris, D. 1998 High-resolution studies of rainfall on Norfolk Island: part II: interpolation of rainfall data. *Journal of Hydrology* **208** (3), 187–193.

Ehret, U., Götzinger, J., Bárdossy, A. & Pegram, G. G. 2008 Radar-based flood forecasting in small catchments, exemplified by the Goldersbach catchment, Germany. *International Journal of River Basin Management* **6** (4), 323–329.

Emery, X. & Ortiz, J. N. M. 2013 Shortcomings of multiple indicator kriging for assessing local distributions. *Applied Earth Science* **113**, 249–259.

Germann, U. & Joss, J. 2001 Variograms of radar reflectivity to describe the spatial continuity of Alpine precipitation. *Journal of Applied Meteorology* **40** (6), 1042–1059.

Gilgen, H. 2006 *Univariate Time Series in Geosciences: Theory and Examples*. Springer, Berlin.

Goovaerts, P. 1997 *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.

Goovaerts, P. 2000 Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* **228** (1), 113–129.

Goudenhoofdt, E. & Delobbe, L. 2009 Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrology and Earth System Sciences* **13** (2), 195–203.

Grimes, D. I. 2008 Hydrological Modelling and the Water Cycle. Springer, Irvine, CA, pp. 117–143. doi:10.1007/978-3-540-77843-1_6.

Haberlandt, U. 2007 Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology* **332** (1), 144–157.

Harrison, D., Driscoll, S. & Kitchen, M. 2000 Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorological Applications* **7** (2), 135–144.

Huang, H., Liang, Z., Li, B. & Wang, D. 2019 A new spatial precipitation interpolation method based on the information diffusion principle. *Stochastic Environmental Research and Risk Assessment* **33**, 765–777.

Huffman, G. J. 1997 Estimates of root-mean-square random error for finite samples of estimated precipitation. *Journal of Applied Meteorology* **36** (9), 1191–1201.

Journel, A. G. 1983 Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology* **15** (3), 445–468.

Kyriakidis, P. C., Kim, J. & Miller, N. L. 2001 Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics. *Journal of Applied Meteorology* **40** (11), 1855–1877.

Lebel, T. & Laborde, J. 1988 A geostatistical approach for areal rainfall statistics assessment. *Stochastic Hydrology and Hydraulics* **2** (4), 245–261.

Lebrenz, H. & Bárdossy, A. 2019 Geostatistical interpolation by quantile kriging. *Hydrology and Earth System Sciences* **3** (23), 1633–1648.

Li, J. & Heap, A. D. 2011 A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecological Informatics* **6** (3), 228–241.

Liu, P. & Tung, Y. 2015 Characterization of dynamic evolution of the spatio-temporal variation of rain-field in Hong Kong. *Hydrology Research* **47**, nh2015129.

Loquin, K. & Dubois, D. 2010 Kriging with Ill-known variogram and data. In: *International Conference on Scalable Uncertainty Management*. Toulouse, France, September 27–29, 2010.

Matheron, G. 1965 *Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature*. Université de Paris, Paris.

Pardo-Igúzquiza, E. & Dowd, P. A. 2005 Multiple indicator co-kriging with application to optimal sampling for environmental monitoring. *Computers & Geosciences* **31** (1), 1–13.

Phillips, D. L., Dolph, J. & Marks, D. 1992 A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology* **58** (1), 119–141.

Seo, D.-J. 1998 Real-time estimation of rainfall fields using rain gage data under fractional coverage conditions. *Journal of Hydrology* **208** (1), 25–36.

Shaw, E. M. & Lynn, P. 1972 Areal rainfall evaluation using two surface fitting techniques. *Hydrological Sciences Journal* **17** (4), 419–433.

Shepard, D. 1968 *A Two-Dimensional Interpolation Function for Irregularly-Spaced Data*. ACM, pp. 517–524.

Sideris, I. V., Gabella, M., Erdin, R. & Germann, U. 2014b Real-time radar–rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland. *Quarterly Journal of the Royal Meteorological Society* **140**, 1097–1111.

Sokol, Z. 2003 The use of radar and gauge measurements to estimate areal precipitation for several Czech river basins. *Studia Geophysica et Geodaetica* **47** (3), 587–604.

Tabios, G. Q. & Salas, J. D. 1985 A comparative analysis of techniques for spatial interpolation of precipitation. *JAWRA Journal of the American Water Resources Association* **21** (3), 365–380.

Thiessen, A. H. 1911 Precipitation averages for large areas. *Monthly Weather Review* **39** (7), 1082–1089.

Tobin, C., Nicotina, L., Parlange, M. B., Berne, A. & Rinaldo, A. 2011 Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region. *Journal of Hydrology* **401** (1), 77–89.

Velasco-Forero, C. A., Sempere-Torres, D., Cassiraga, E. F. & Jaime Gómez-Hernández, J. 2009 A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. *Advances in Water Resources* **32** (7), 986–1002.

Verworn, A. & Haberlandt, U. 2011 Spatial interpolation of hourly rainfall–effect of additional information, variogram inference and storm properties. *Hydrology and Earth System Sciences* **15** (2), 569–584.

Wagner, P. D., Fiener, P., Wilken, F., Kumar, S. & Schneider, K. 2012 Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. *Journal of Hydrology* **464–465**, 388–400.