Journal of Hydrology 208 (1998) 25–36

Real-time estimation of rainfall fields using rain gage data under fractional coverage conditions

D.-J. Seo[1]

Hydrologic Research Laboratory, National Weather Service, 1325 East-West Highway, Silver Spring, MD 20910, USA

Received 11 October 1996; accepted 25 March 1998

Abstract

Two new attempts at optimal estimation of rainfall fields using hourly rain gage data under fractional coverage conditions are reported; (1) double optimal estimation using analogues of indicator and simple kriging to estimate probability of rainfall and rainfall amount given raining, respectively, and (2) single optimal estimation using an analogue of simple kriging to directly estimate rainfall amount. They are evaluated via cross validation using hourly rain gage data from the operational network under the Tulsa, Oklahoma, WSR-88D (Weather Surveillance Radar-1988 Doppler version) umbrella.

Keywords: Rainfall; Estimation; Fractional coverage; Rain gage

1. Introduction

Over a basin, rainfall coverage is generally fractional, and hence procedures for optimal estimation of spatial distribution of rainfall must account for not only within-storm variability of rainfall but also variability due to fractional coverage of rainfall. The existing optimal rainfall estimation procedures (see, e.g. Creutin and Obled, 1982; Tabios and Salas, 1985), with the notable exception of Barancourt et al. (1992), do not explicitly consider rainfall variability due to fractional coverage, and therefore their estimates are less than optimal (in the sense that error variance is minimized) under fractional coverage conditions.

Barancourt et al. (1992) describes a procedure that explicitly considers variability due to fractional coverage of rainfall. Though the procedure was shown to perform significantly better than what was referred to as "global kriging", it has significant drawbacks in that (1) independence between within-storm variability and variability due to fractional coverage has to be assumed, and (2) estimation variance cannot be produced. Also, because the derivation of the estimation procedure lacks mathematical rigor, it is difficult to ascertain exactly what conditional estimates are sought and how the independence assumption impacts the estimation.

The first of the two estimation procedures described in this work is structurally similar to Barancourt et al. (1992) in that optimal linear estimation is performed twice to estimate probability of rainfall and rainfall amount given raining. However, owing to a set of mathematical identities involving conditional expectations and variances of occurrence of rainfall and rainfall amount, the procedure does not require the independence assumption, and provides estimation

[1] Tel.: 001 301 713 0640; fax: 001 301 713 0963; e-mail: dongjun.seo@noaa.gov

0022-1694/98/$19.00
PII S0022-1694(98)00140-1

variance as well. The second procedure involves optimal linear estimation based on the correlation function that accounts for both within-storm variability of rainfall and variability due to fractional coverage of rainfall. The work reported here represents a part of the continuing effort in the National Weather Service (NWS) to provide River Forecast Centers (RFCs) and Weather Forecast Offices (WFOs) with more accurate real-time precipitation analysis procedures in support of operational hydrologic forecasting.

The organization of this paper is as follows. Section 2 describes the estimation problem and the solution approach. Sections 3 and 4 describe the first estimation procedure. Section 5 describes the second estimation procedure. Section 6 describes parameter estimation. Section 7 describes the validation experiment and presents the results. Section 8 provides summary and conclusions.

2. Description of the problem and the proposed approach

The problem of estimating spatial distribution of rainfall may be described as follows. Obtain an estimate of the unknown rainfall amount, z_0, at an arbitrary location, u_0, within the estimation domain, A, using nearby rain gage measurements in A, $z_1, \ldots, z_n$, at locations, $u_1, \ldots, u_n$, respectively. Assuming, for the sake of argument, isotropy in the correlation structure of rainfall, we may define the estimation domain, A, to be the circle centered at the point of estimation, u_0, whose radius equals the spatial correlation scale of rainfall. The goal is then to obtain an estimate that approximates as closely as possible the conditional expectation, $E[Z_0|Z_1 = z_1, \ldots, Z_n = z_n]$, where $E[\cdot]$, Z_j, and z_j denote the expectation operator, the random variable representing rainfall amount at u_j, and the observed rainfall amount at u_j, respectively. For notational brevity, we define $z \equiv \{Z_1 = z_1, \ldots, Z_n = z_n\}$. Throughout this paper, random variables are denoted by uppercase letters whereas their experimental values are denoted by corresponding lowercase letters.

If the joint probability distribution of $Z_0, Z_1, \ldots, Z_n$ is multivariate normal, the optimal linear estimate (Schweppe, 1973, p. 96), or equivalently the simple-kriging estimate (Journel and Huijbregts, 1978, p. 561), is identical to the conditional expectation $E[Z_0|z]$, and hence is the best statistical estimate obtainable in every reasonable sense of the word (Schweppe, 1973, p. 97). For rainfall, however, the optimal linear estimate is not as good as the conditional expectation because its probability distribution is skewed in general.

In full coverage situations, estimates that are comparable to conditional expectation may be obtained by using nonlinear estimation procedures such as disjunctive kriging (Matheron, 1975; Journel and Huijbregts, 1978; Azimi-Zonooz et al., 1989) or indicator (co-) kriging (Journel, 1983, Seo, 1996), but at the expense of about an order-of-magnitude (or larger) increase in computational requirements. In fractional coverage situations, the goodness of optimal linear estimates deteriorates because of probability mass at zero, which adds skewness to the probability distribution of rainfall. Also, in fractional coverage situations, nonlinear estimation procedures based on transformation of continuous variables, such as disjunctive kriging, are not directly applicable because of the discrete probability mass at zero. Indicator (co-) kriging, on the other hand, is capable of handling fractional coverage situations (Seo, 1996) but only at the expense of a significant increase in computational requirements, and hence is not considered operationally viable at this time. In light of these observations, the objective of this work is to develop estimation procedures that explicitly account for rainfall variability due to fractional coverage, but are computationally significantly less demanding than indicator (co-) kriging.

The first step in the first estimation procedure, referred to herein as the Double Optimal Estimation (DOE), is to rewrite the conditional expectation, $E[Z_0|z]$, as follows:

$$E[Z_0|z] \equiv \int_0^\infty z_0 f_{Z0|Z1,\ldots,Zn}(z_0|z_1,\ldots,z_n)\,dz_0 \quad (1a)$$

$$= \int_{0+}^\infty z_0 f_{Z0|Z1,\ldots,Zn}(z_0|z_1,\ldots,z_n)\,dz_0 \quad (1b)$$

$$= \int_{0+}^\infty z_0 f_{Z0|Z1,\ldots,Zn}(z_0|z_1,\ldots,z_n)/Pr[Z_0 > 0|z]\,dz_0 \cdot Pr[Z_0 > 0|z] \quad (1c)$$

$$= E[Z_0|z, z_0 > 0]\cdot Pr[Z_0 > 0|z] \quad (1d)$$

In the above, $f_{Z0|Z1,\ldots,Zn}(z_0|z_1,\ldots,z_n)$ denotes the conditional probability density function of Z_0 given z, and $Pr[\cdot]$ denotes the probability that the event bracketed will occur. Similarly, the conditional variance $Var[Z_0|z]$ may be written as:

$$Var[Z_0|z] \equiv E[Z_0^2|z] - E^2[Z_0|z] \tag{2a}$$

$$= E[Z_0^2|z, z_0 > 0] \cdot Pr[Z_0 > 0|z] - E^2[Z_0|z, z_0 > 0] \cdot Pr^2[Z_0 > 0|z] \tag{2b}$$

$$= Var[Z_0|z, z_0 > 0] \cdot Pr[Z_0 > 0|z] + E^2[Z_0|z, z_0 > 0] \cdot Pr[Z_0 > 0|z] \cdot \{1 - Pr[Z_0 > 0|z]\} \tag{2c}$$

The next step in DOE is then to estimate $E[Z_0|z, z_0 > 0]$, $Var[Z_0|z, z_0 > 0]$, and $Pr[Z_0 > 0|z]$, from which estimates of $E[Z_0|z]$ and $Var[Z_0|z]$ may be obtained via Eq. (1a)–(1d), Eq. (2a)–(2c), respectively. Motivation for such an approach stems from the preceding discussion on skewness: because conditioning on $\{z_0 > 0\}$ reduces skewness in the probability distribution of Z_0, the optimal linear estimate of $E[Z_0|z, z_0 > 0]$ is expected to be a better approximation of $E[Z_0|z, z_0 > 0]$ than the direct optimal linear estimate of $E[Z_0|z]$ is of $E[Z_0|z]$. Hence, if $Pr[Z_0 > 0|z]$ can be estimated accurately, we may expect that the indirect estimate of $E[Z_0|z]$ based on Eq. (1a)–(1d) be more accurate than the direct estimate of $E[Z_0|z]$. In the following two sections, we describe how $Pr[Z_0 > 0|z]$, $E[Z_0|z, z_0 > 0]$, and $Var[Z_0|z, z_0 > 0]$ in Eq. (1a)–(1d), Eq. (2a)–(2c), may be estimated.

3. Estimation of $Pr[Z_0 > 0|z]$

The conditional probability $Pr[Z_0 > 0|z]$ is estimated from the indicator kriging estimator (Journel, 1983; Solow, 1986; Seo, 1996) via the following conditional-probability approximation of the conditional expectation of an indicator random variable:

$$Pr[Z_0 > 0|z] \approx E[I_0|I_1 = i_1, I_2 = i_2, \ldots, I_n = i_n] \tag{3}$$

In the above, the indicator random variable, I_j, is defined according to:

$$i_j = \begin{cases} 1 \text{ if } z_j > 0 \\ 0 \text{ if } z_j = 0 \end{cases} \tag{4}$$

where, following our notational convention, i_j denotes the experimental value that the random variable I_j takes on. For notational brevity, we define $\boldsymbol{i} \equiv \{I_1 = i_1, I_2 = i_2, \ldots, I_n = i_n\}$. By the definition in Eq. (4), we have $E[I_j] = Pr[Z_j > 0]$ and $Var[I_j] \equiv E[I_j^2] - E^2[I_j] = Pr[Z_j > 0](1 - Pr[Z_j > 0])$.

As in indicator kriging, the right-hand side of Eq. (3) is estimated by the following simple kriging estimator:

$$E^*[I_0|\boldsymbol{i}] = E[I_0] + \sum_{j=1}^{n} \lambda_j (i_j - E[I_j]) \tag{5}$$

where the asterisk signifies that the estimate is only an approximation of the conditional expectation. The optimal weights, λ_j, are obtained by minimizing $E[\{E^*[I_0|\boldsymbol{i}] - I_0\}^2]$:

$$(\lambda_1, \ldots, \lambda_n) = P_0 P^{-1} \tag{6}$$

where P_0 and P are the $(1) \times (n)$ and $(n) \times (n)$ indicator covariance vector and matrix whose jth and ijth entries are given by $Cov[I_0, I_j]$ and $Cov[I_i, I_j]$, respectively. The estimation variance is given by:

$$Var^*[I_0|\boldsymbol{i}] = Var[I_0] - P_0 P^{-1} P_0^T \tag{7}$$

From the definition of the indicator random variable in Eq. (4), $Cov[I_i, I_j]$ is written in terms of uni- and bivariate probabilities at and between u_i and u_j, respectively:

$$Cov[I_i, I_j]$$

$$\equiv E[I_i I_j] - E[I_i]E[I_j] \tag{8a}$$

$$= Pr[Z_i > 0, Z_j > 0] - Pr[Z_i > 0]Pr[Z_j > 0] \tag{8b}$$

In practice, in order to ensure positive definiteness of the covariance matrix P (and hence to guarantee valid solutions of the optimal weights) in Eq. (6), it is necessary to assume that occurrence of rainfall is wide-sense second-order homogeneous (Karlin and Taylor, 1975, p. 445). Then, $Cov[I_i, I_j]$ may be expressed as:

$$Cov[I_i, I_j] = \sigma_I^2 \cdot \rho_I(|u_i - u_j|) \tag{9}$$

In the above, σ_I^2 denotes the locally (i.e. within A) homogeneous indicator variance, and $\rho_I(|u_i - u_j|)$ denotes the indicator correlation coefficient, where $|u_i - u_j|$ denotes the Euclidean distance between u_i and u_j. The locally homogeneous indicator variance,

σ_I^2, can be rewritten as:

$$\sigma_I^2$$

$$= E[I^2] - E^2[I] \tag{10a}$$

$$= E[\|A\|^{-1} \int_A I^2(u)\, du] - E^2[\|A\|^{-1} \int_A I(u)\, du] \tag{10b}$$

$$= m_I(1 - m_I) \tag{10c}$$

where m_I denotes the expected fraction of rain area in A. Under the assumption of local homogeneity, we also have $E[I_0] = E[I_j] = m_I$ in Eq. (5). The homogeneity assumption implies that occurrence of rainfall is equally likely everywhere in A, which is reasonable in the absence of pronounced local orographic or lake effects.

Because of lack of rain gage data, real-time estimation of time-varying $\rho_I(\cdot)$ is extremely difficult (if not impossible). For this reason, we used climatological estimates of $\rho_I(\cdot)$ in this work (see also Section 6). Being a measure of spatial association of rainy or non-rainy patches, the indicator correlation function, $\rho_I(\cdot)$, does not in general undergo abrupt changes. Hence, if stratified according to the stage of storm development (e.g. developing, mature, and dissipating), storm type (e.g. convective and stratiform), and seasonality (e.g. warm and cold), climatological estimates should provide a good approximation to time-varying $\rho_I(\cdot)$. Advantages of using climatological variograms in rainfall estimation have long been recognized by a number of researchers (see, e.g. Bastin et al., 1984; Lebel and Bastin, 1985; Lebel et al., 1987; Seo, 1996). On the other hand, the fraction of rain area in A can undergo drastic changes over a very short period of time, depending on how much the areal extent of the storm happens to be overlapping with the fixed field of view, A. For this reason, the expected fraction of rain area in A, m_I, is estimated in this work every hour from real-time gage data (see also Section 6).

4. Estimation of $E[Z_0|z, z_0 > 0]$

The conditional expectation $E[Z_0|z, z_0 > 0]$ is estimated using the following linear estimator:

$$E^*[Z_0|z, z_0 > 0] = E[Z_0|z_0 > 0] + \sum_{i=1}^{n} \Gamma_i(z_i - E[Z_i|z_0 > 0]) \tag{11}$$

The optimal weights, Γ_i, in Eq. (11) are obtained by minimizing the conditional error variance, $E[(Z_0 - E^*[Z_0|z, z_0 > 0])^2|z_0 > 0]$:

$$(\Gamma_1, \Gamma_2, \ldots, \Gamma_n) = Q_0 Q^{-1} \tag{12}$$

where Q_0 and Q are the $(1)\times(n)$ and $(n)\times(n)$ covariance vector and matrix whose jth and ijth entries are given by the conditional covariances, $Cov[Z_0, Z_j|z_0 > 0]$ and $Cov[Z_i, Z_j|z_0 > 0]$, respectively. The estimation variance is given by:

$$Var^*[Z_0|z, z_0 > 0] = Var[Z_0|z_0 > 0] - Q_0 Q^{-1} Q_0^T \tag{13}$$

Under the assumption that both the rainfall amount, Z, and the occurrence of rainfall, I, are locally second-order homogeneous, the expected amount of positive rainfall in A, m_R, may be written as:

$$m_R$$

$$= E[Z|z > 0] \tag{14a}$$

$$= E[Z]/Pr[Z > 0] \tag{14b}$$

$$= E[Z]/E[I] \tag{14c}$$

$$= E[\|A\|^{-1} \int_A Z(u)\, du]/E[\|A\|^{-1} \int_A I(u)\, du] \tag{14d}$$

$$= m/m_I \tag{14e}$$

where Eq. (14b) follows from Eq. (1d), and m denotes the expected spatially-averaged rainfall over A. The assumption that rainfall amount is locally homogeneous is reasonable in the absence of pronounced local orographic or lake effects. The locally homogeneous variance of positive rainfall, σ_R^2, may be written as:

$$\sigma_R^2$$

$$= Var[Z|z > 0] \tag{15a}$$

$$= Var[Z]/Pr[Z > 0] - E^2[Z|z > 0](1 - Pr[Z > 0]) \tag{15b}$$

$$= \sigma^2/m_I - m_R^2(1 - m_I) \tag{15c}$$

In the above, Eq. (15b) follows from Eq. (2c), and σ^2

denotes the locally homogeneous (unconditional) variance of rainfall in A:

$$\sigma^2 = E[Z^2] - E^2[Z] \quad (16a)$$

$$= E[\|A\|^{-1} \int_A Z^2(u)\,du] - E^2[\|A\|^{-1} \int_A Z(u)\,du] \quad (16b)$$

Because both m_R and σ_R^2 are specified largely by the storm dynamics, they can be very fast-varying in time. For this reason, m_R and σ_R^2 are estimated in this work every hour from real-time rain gage data (see also Section 6). The next two subsections describe how the conditional expectation $E[Z_i|z_0>0]$ in Eq. (11) and the conditional covariances, $Cov[Z_0, Z_i|z_0>0]$ and $Cov[Z_i, Z_j|z_0>0]$, in Eq. (12) may be specified.

4.1. Specification of $Cov[Z_0, Z_i|z_0>0]$

The conditional covariance $Cov[Z_0, Z_i|z_0>0]$ may be obtained from the following set of operations:

$$Cov[Z_0, Z_i|z_0>0] \equiv E[Z_0Z_i|z_0>0] - E[Z_0|z_0>0]\cdot E[Z_i|z_0>0] \quad (17a)$$

$$= E[Z_0Z_i|z_i>0, z_0>0]\cdot Pr[Z_i>0|z_0>0] + E[Z_0Z_i|z_i=0, z_0>0]\cdot Pr[Z_i=0|z_0>0] - E[Z_0|z_0>0]\cdot(E[Z_i|z_0>0, z_i>0]\cdot Pr[Z_i>0|z_0>0] + E[Z_i|z_0>0, z_i=0]\cdot Pr[Z_i=0|z_0>0]) \quad (17b)$$

$$= (E[Z_0Z_i|z_i>0, z_0>0] - E[Z_0|z_0>0]\cdot E[Z_i|z_i>0])\cdot Pr[Z_i>0|z_0>0] \quad (17c)$$

$$= Cov[Z_0, Z_i|z_0>0, z_i>0]\cdot Pr[Z_i>0|z_0>0] \quad (17d)$$

Under the assumption that positive rainfall amount is wide-sense second-order homogeneous (Karlin and Taylor, 1975, p. 445), we may write $Cov[Z_0, Z_i|z_0>0, z_i>0]$ in Eq. (17d) as:

$$Cov[Z_0, Z_i|z_0>0, z_i>0] = \sigma_R^2\cdot\rho_R(|u_i-u_j|) \quad (18)$$

In the above, $\rho_R(|u_i-u_j|)$ denotes the correlation coefficient of positive rainfall, which is estimated climatologically in this work (see also Section 6). The conditional probability $Pr[Z_i>0|z_0>0]$ in Eq. (17d) is given by Eq. (8a), (8b) and Eq. (9) as:

$$Pr[Z_i>0|z_0>0] = (1-m_I)\cdot\rho_I(|u_i-u_0|) + m_I \quad (19)$$

We then have:

$$Cov[Z_0, Z_i|z_0>0] = \sigma_R^2\cdot\rho_R(|u_i-u_0|)\cdot((1-m_I)\cdot\rho_I(|u_i-u_0|) + m_I) \quad (20)$$

4.2. Specification of $Cov[Z_i, Z_j|z_0>0]$

By definition, we have:

$$Cov[Z_i, Z_j|z_0>0] \equiv E[Z_iZ_j|z_0>0] - E[Z_i|z_0>0]\cdot E[Z_j|z_0>0] \quad (21)$$

In the above, $E[Z_iZ_j|z_0>0]$ can be rewritten as:

$$E[Z_iZ_j|z_0>0] = E[Z_iZ_j|z_0>0, z_i>0, z_j>0]\cdot Pr[Z_i>0, Z_j>0|z_0>0] + E[Z_iZ_j|z_0>0, z_i=0, z_j>0]\cdot Pr[Z_i=0, Z_j>0|z_0>0] + E[Z_iZ_j|z_0>0, z_i>0, z_j=0]\cdot Pr[Z_i>0, Z_j=0|z_0>0] + E[Z_iZ_j|z_0>0, z_i=0, z_j=0]\cdot Pr[Z_i=0, Z_j=0|z_0>0] \quad (22a)$$

$$= E[Z_iZ_j|z_0>0, z_i>0, z_j>0]\cdot Pr[Z_i>0, Z_j>0|z_0>0] \quad (22b)$$

$$= E[Z_iZ_j|z_i>0, z_j>0]\cdot Pr[Z_i>0, Z_j>0|z_0>0] \quad (22c)$$

In arriving at Eq. (22c) from Eq. (22b), we have used the assumption that positive rainfall is locally second-order homogeneous, and therefore conditioning on $\{z_0>0\}$ has no effect on $E[Z_iZ_j|z_i>0, z_j>0]$.

The conditional expectation $E[Z_i|z_0>0]$ in Eq. (21) can be rewritten as:

$$E[Z_i|z_0>0] = E[Z_i|z_i>0, z_0>0]\cdot Pr[Z_i>0|z_0>0]$$

$$+E[Z_i|z_i=0, z_0>0]\cdot Pr[Z_i=0|z_0>0] \tag{23a}$$

$$=E[Z_i|z_i>0]\cdot Pr[Z_i>0|z_0>0] \tag{23b}$$

We then have:

$$Cov[Z_i, Z_j|z_0>0]$$

$$=E[Z_iZ_j|z_i>0, z_j>0]\cdot Pr[Z_i>0, Z_j>0|z_0>0]$$

$$-E[Z_i|z_i>0]\cdot Pr[Z_i>0|z_0>0]\cdot E[Z_j|z_j>0]$$

$$\cdot Pr[Z_j>0|z_0>0] \tag{24}$$

In the above, $E[Z_iZ_j|z_i>0, z_j>0]$ and $Pr[Z_i>0|z_0>0]$ are given by Eqs. (18) and (19), respectively. Specification of $Pr[Z_i>0, Z_j>0|z_0>0]$ in Eq. (24) requires estimation of trivariate cumulative distribution among Z_i, Z_j, and Z_0 as functions of $|u_i-u_j|$, $|u_i-u_0|$, and $|u_j-u_0|$ at the cutoff of zero rainfall. Empirical estimation of this distribution using, for example, radar rainfall data, is left as a future endeavor. In this work, $Pr[Z_i>0, Z_j>0|z_0>0]$ is approximated by the indicator kriging estimator as described below. First, using the identity $Pr[Z_j>0]=Pr[Z_0>0]$ under the assumption of local homogeneity, we write:

$$Pr[Z_i>0, Z_j>0|z_0>0]=Pr[Z_0>0|z_i>0, z_j>0]$$

$$\cdot Pr[Z_i>0|z_j>0] \tag{25}$$

where $Pr[Z_i>0|z_j>0]$ is given by Eq. (19). To specify $Pr[Z_0>0|z_i>0, z_j>0]$ in Eq. (25), we use, as in the estimation of $Pr[Z_0>0|z]$ in Eq. (3), the conditional-probability approximation of the conditional expectation of an indicator random variable, $Pr[Z_0>0|z_i>0, z_j>0] \approx E[I_0|I_i=1, I_j=1]$. Then, from Eq. (5), we have:

$$Pr^*[Z_0>0|z_i>0, z_j>0]=m_I+(\omega_i+\omega_j)(1-m_I) \tag{26}$$

where

$$\omega_i=\{\rho_I(|u_0-u_i|)-\rho_I(|u_i-u_j|)\rho_I(|u_0-u_j|)\}/$$
$$\{1-\rho_I^2(|u_i-u_j|)\} \tag{27}$$

$$\omega_j=\{\rho_I(|u_0-u_j|)-\rho_I(|u_i-u_j|)\rho_I(|u_0-u_i|)\}/$$
$$\{1-\rho_I^2(|u_i-u_j|)\} \tag{28}$$

Reasonableness of Eq. (26) can be easily checked by evaluating its dependence on the geometry among u_0, u_i, and u_j as specified by $|u_0-u_i|$, $|u_i-u_j|$, and $|u_0-u_j|$. It is also easily verified that $Cov[Z_i, Z_j|Z_0>0]$ thus obtained via Eq. (24) is reduced, as it should, to Eq. (20) when $j = 0$.

5. Direct estimation of $E[Z_0|z]$

A simpler and computationally less expensive alternative to DOE described above is to use the simple kriging estimator to directly estimate rainfall amount, based on the correlation function that accounts for both within-storm variability and variability due to fractional coverage. This approach, referred to herein as the Single Optimal Estimation (SOE), is described below. As will be seen, SOE is identical to the so-called best linear unbiased estimator (BLUE) if the rainfall coverage is full. Hence, the only distinction being made by the term SOE is that it specifically refers to BLUE under fractional coverage conditions. The estimator used in SOE is the same as that in simple kriging:

$$E^*[Z_0|z]=E[Z_0]+\sum_{i=1}^{n}\Lambda_i(z_i-E[Z_i]) \tag{29}$$

Under the assumption that rainfall amount is locally homogeneous, the unconditional expectations, $E[Z_0]$ and $E[Z_i]$, in Eq. (29) are given by m in Eq. (14a)–(14e). The optimal weights, Λ_is, are obtained by minimizing the error variance, $E[(Z_0-E^*[Z_0|z])^2]$:

$$(\Lambda_1, \Lambda_2, \ldots, \Lambda_n)=C_0C^{-1} \tag{30}$$

where C_0 and C are the $(1)\times(n)$ and $(n)\times(n)$ covariance vector and matrix whose jth and ijth entries are given by $Cov[Z_0, Z_j]$ and $Cov[Z_i, Z_j]$, respectively. The estimation variance is given by:

$$Var^*[Z_0|z]=Var[Z_0]-Q_0Q^{-1}Q_0^T \tag{31}$$

Under the assumption that rainfall amount is locally second-order homogeneous, the unconditional variance $Var[Z_0]$ in Eq. (31) is given by σ^2 in Eq. (16a) and Eq. (16b).

The first step in specifying $Cov[Z_i, Z_j]$ is to write:

$$Cov[Z_i, Z_j]$$

$$\equiv E[Z_iZ_j]-E[Z_i]E[Z_j] \tag{32a}$$

$$= E[Z_i Z_j | z_i > 0, z_j > 0] \cdot Pr[Z_i > 0, Z_j > 0] - E[Z_i | z_i > 0] \cdot E[Z_j | z_j > 0] \cdot Pr[Z_i > 0] \cdot Pr[Z_j > 0] \quad (32b)$$

Under the assumption that both positive rainfall amount and occurrence of rainfall are wide-sense second-order homogeneous (Karlin and Taylor, 1975, p. 445), we use Eq. (18), Eq. (8a), Eq. (8b), Eq. (9), Eq. (14a)–(14e) to specify the terms in Eq. (32b), which yields:

$$Cov[Z_i, Z_j] = \sigma_R^2 \cdot m_I(1 - m_I) \cdot \rho_R(|u_i - u_j|) \cdot \rho_I(|u_i - u_j|) + m_R^2 \cdot m_I(1 - m_I) \cdot \rho_I(|u_i - u_j|) + \sigma_R^2 \cdot m_I^2 \cdot \rho_R(|u_i - u_j|) \quad (33)$$

Eq. (33) has been empirically verified in the study of climatological variability of mean areal rainfall through fractional coverage (Seo and Smith, 1996). Note in Eq. (33) that, if the rainfall coverage is full (i.e. $m_I = 1$), the covariance is specified exclusively by the inner variability term (the last term), and that, if there is no inner variability (i.e. $\sigma_R = 0$), the covariance is specified exclusively by the intermittency term (i.e. the second term). It is easy to ensure that Eq. (33) constitute a valid covariance model: for example, if both $\rho_R(|\cdot|)$ and $\rho_I(|\cdot|)$ are of the exponential or the gaussian model (Journel and Huijbregts, 1978, pp. 164–165), $\rho_R(|\cdot|) \cdot \rho_I(|\cdot|)$ is also of the same model, and hence Eq. (33) is a valid covariance (Journel and Huijbregts, 1978, p. 172).

6. Parameter estimation

Both DOE and SOE require specification of the following parameters; (1) m_R, the expected amount of positive rainfall in A, (2) σ_R^2, the variance of positive rainfall in A, (3) m_I, the expected fraction of rain area in A, (4) $\rho_R(|h|)$, the spatial correlation function of positive rainfall, where $|h|$ denotes the separation distance, and (5) $\rho_I(|h|)$, the spatial correlation function of occurrence of rainfall.

Reliable estimation of time-varying m_I, m_R, and σ_R^2 based on rain gage data only from the current hour is very difficult due to lack of data points. Ideally, recursive estimation procedures that make use of rain gage data not only from the current hour but also from the previous hours will have to be used. We only note here that such procedures are under development, and will be reported in the near future. In this work, m_I, m_R, and σ_R^2 are estimated every hour using rain gage data only from the current hour: the estimators used are $m_I = n_p/n$, $m_R = (1/n_p)\sum z_{pj}$, and $\sigma_p^2 = \sum z_{pj}^2/(n_p - 1) - (\sum z_{pj})^2/\{n_p(n_p - 1)\}$, $j = 1, \ldots, n_p$, where n_p is the number of positive gage data in A, n is the number of all gage data in A, and z_{pj} is the jth positive gage datum in A. Implicit in the above estimators is the assumption that rain gage data are spatially independent of one another, which, according to sensitivity analyses, is reasonable for purposes of estimating m_I and m_R.

Climatological estimation of $\rho_I(|h|)$ and $\rho_R(|h|)$ based solely on hourly rain gage data is extremely difficult (if not impossible) due to sparsity of gage networks. In this work, $\rho_I(|h|)$ and $\rho_R(|h|)$ are estimated climatologically from hourly radar rainfall data from the Weather Surveillance Radar-1988 Doppler version (WSR-88D) at Tulsa, Oklahoma. The period covered is late August 1993 through early May 1994. The reader is kindly referred to Seo and Smith (1996) for indicator and conditional (on occurrence of rainfall) correlograms at adjacent WSR-88D sites (i.e. Frederick, Oklahoma, and Little Rock, Arkansas): those at Tulsa are very similar, and hence are not shown here. They suggest that both indicator and conditional correlation functions are well-modeled by the isotropic exponential model with nugget effect; for example, for $\rho_I(|h|)$:

$$\rho_I(|h|) = \begin{cases} 1 & \text{if } h = 0 \\ \rho_{I0} \exp(-|h|/L_I) & \text{if } h > 0 \end{cases} \quad (34)$$

where ρ_{I0} is the lag-0 indicator correlation coefficient and L_I is the scale parameter.

Use of radar rainfall data in estimation of correlation functions of gage rainfall is not completely without base, for example if radar rainfall is a linear sum of gage rainfall and a white-noise error in space, the conditional correlation function of gage rainfall is identical to that of radar rainfall (Creutin et al., 1988; Seo and Smith, 1991). Because of various sources of systematic errors in radar observation of rainfall (Wilson and Brandes, 1979; Austin, 1987; Smith et al., 1996; Seo et al., 1998) and differences in sampling scale between radar and rain gages, however, the above assumption is not true in general (Krajewski et al., 1996). Representativeness of the correlation functions of hourly gage rainfall as estimated from hourly radar rainfall data is currently under investigation, and will be reported in the near future.

7. Validation

To evaluate DOE and SOE, cross validation was performed using hourly rain gage data, which involved the following steps: (1) at each hour, withhold rain gage data one at a time, (2) at each time, estimate rainfall at the withheld gage location using surrounding, non-withheld rain gage measurements available for that hour, and (3) for each withheld gage location, compare the estimates against the revealed gage measurement. The hourly gage data used are from the operational network under the WSR-88D radar umbrella at Tulsa, Oklahoma (see Fig. 1: the radar umbrella of radius of 230 km circumscribes the square area), covering April, June, July, and August of 1994. Because the estimation procedures do not require parameter-tuning of any kind, no calibration data sets were necessary. Because SOE and DOE are being developed as potential alternatives to the existing precipitation analysis procedures, validation was focused largely on how SOE and DOE compare with the reciprocal distance-squared method, abbreviated as RDS (also known as the inverse distance-squared method, Chow et al., 1988; National Weather Service, 1993), which is one of the long-standing gage precipitation analysis procedures for operational hydrologic forecasting in NWS.

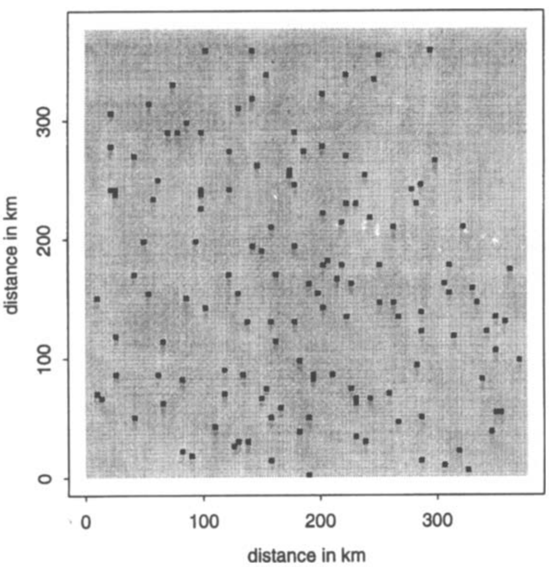

Fig. 1. Hourly rain gage network under the Tulsa, Oklahoma, WSR-88D umbrella.

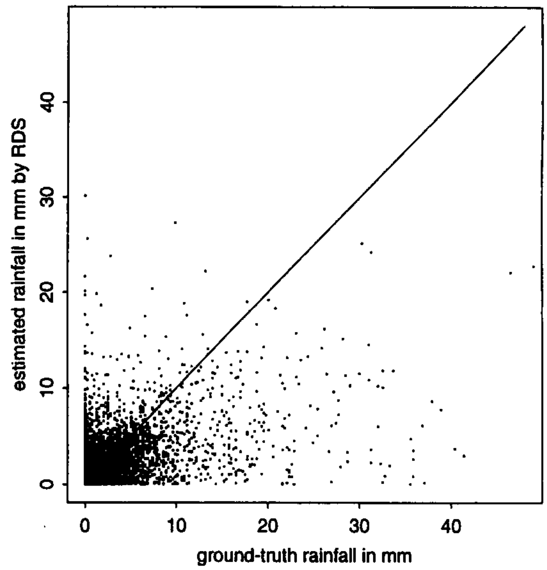

Fig. 2. Scatter-plot between observed and estimated rainfall from RDS.

The number of nearest rain gage data used is 15 for all estimation procedures. Estimates less than 0.25 mm are truncated to zero in accordance with the fact that most rain gages in the network have the minimum detectable rainfall amount of one-hundredth of an inch. Figs 2 and 3 show scatter-plots of estimated versus observed hourly rainfall from the cross validation for RDS and DOE, respectively. The scatter-plot for SOE (not shown) is closer to that of DOE than that of RDS. The tighter the scatter is around the straight line, the better the estimates are. Out of 19 204 data points, 13 506 are associated with the observed rainfall amount of zero. It is seen that the DOE estimates form a tighter scatter around the straight line than the RDS estimates. The type of scatters seen in the figures are not at all unusual for cross validation of hourly point rainfall (even though the study area has one of the densest operational rain

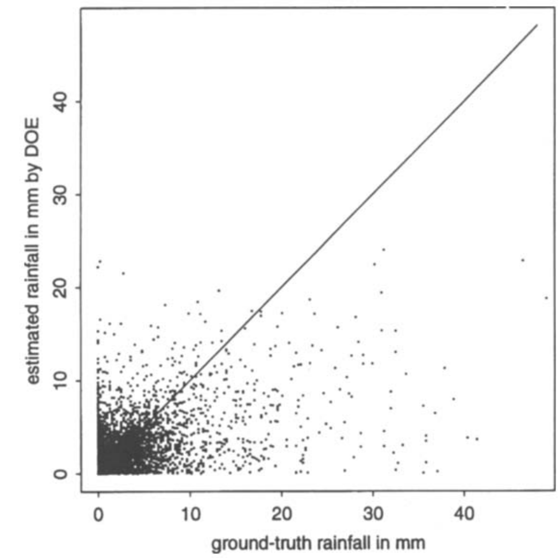

Fig. 3. Scatter-plot between observed and estimated rainfall from DOE.

gage networks in the US), and serves to illustrate extremely large variabilities of rainfall at that space–time scale. Also contributing to the large scatter is the fact that the validation data come largely from the convective season.

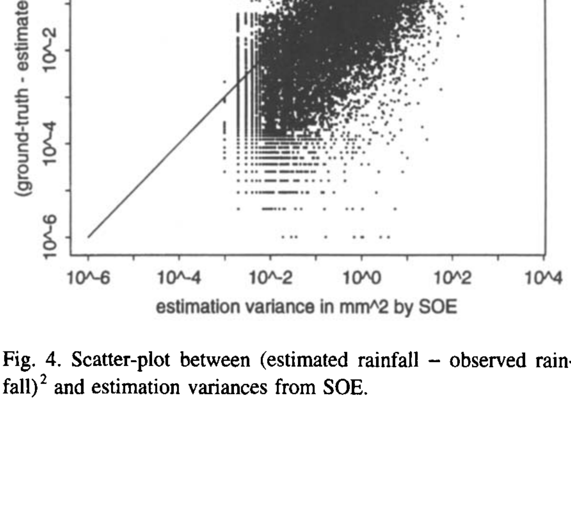

Fig. 4. Scatter-plot between (estimated rainfall – observed rainfall)^2 and estimation variances from SOE.

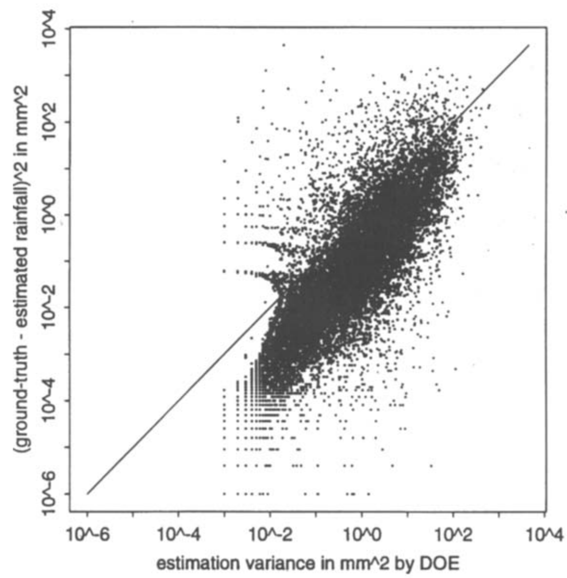

Fig. 5. Scatter-plot between (estimated rainfall – observed rainfall)^2 and estimation variances from DOE.

Figs 4 and 5 show scatter-plots of (estimated rainfall – observed rainfall)^2 versus estimation variance from SOE and DOE, respectively. The tighter the scatter is around the straight line, the better the variance estimates are. It is seen that the DOE estimates form a tighter scatter. The tendency to overestimate estimation variance by both SOE and DOE is due primarily to sampling errors associated with estimating conditional variance, σ_R^2, and expected fraction of rain area, m_I, from a small number of rain gage data in A: similar results have been observed in Seo et al. (1990) and Seo and Smith (1991). Other contributing factors include nonhomogeneities in occurrence of rainfall and positive rainfall in A, unrepresentative correlation functions due to lack of stratification according to rainfall climatology and storm dynamics, and to lack of representativeness in the correlation functions of gage rainfall estimated from radar rainfall data. The biases seen in Figs. 4 and 5 could have been reduced if we had obtained estimation variances by cross-validating against the actual error variance in real time. Such an approach, however, requires that the gage network be of high density, and hence is not considered operationally viable in many parts of the country. The traces of parabolic curves appearing in Fig. 5 over the straight line are due to the fact that $E[Z_0|z]$ and $Var[Z_0|z]$ are

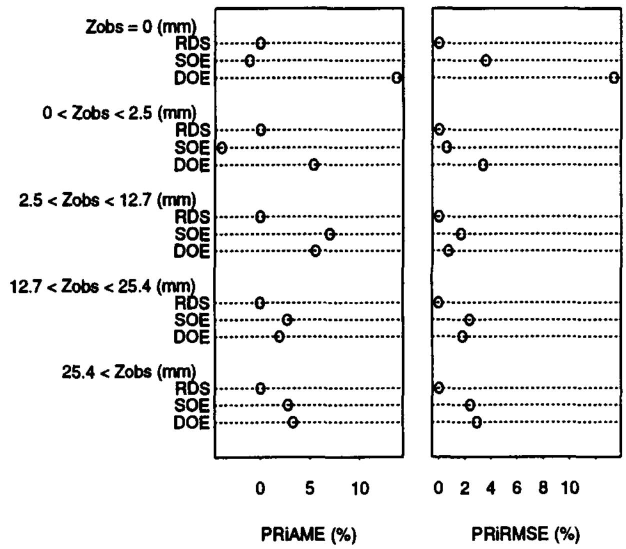

Fig. 6. Percent reduction in mean error and root mean square error over various ranges of amount of observed rainfall.

functionally related to each other by the common terms, $E[Z_0|z, z_0 > 0]$ and $Pr[Z_0 > 0|z]$ (see Eq. (1d) and Eq. (2c)).

To help quantitatively assess the performance of SOE and DOE relative to that of RDS, we introduce two measures; percent reduction in absolute mean error and percent reduction in root mean square error, abbreviated as PRiAME and PRiRMSE, respectively. For example, the percent reduction in absolute mean error by SOE over RDS estimates, or PRiAME-(SOE), is defined as:

$$\mathrm{PriAME(SOE)} = 100 \times \{\mathrm{AME(RDS)} - \mathrm{AME(SOE)}\}/\mathrm{AME(RDS)} \quad (35)$$

where AME(RDS) and AME(SOE) denote the absolute mean errors (i.e. the absolute values of mean error) of RDS and SOE estimates, respectively. Analogously, the percent reduction in root mean square error by DOE over RDS estimates, or PRiRMSE(DOE), is defined as:

$$\mathrm{PriRMSE(DOE)} = 100 \times \{\mathrm{RMSE(RDS)} - \mathrm{RMSE(DOE)}\}/\mathrm{RMSE(RDS)} \quad (36)$$

where RMSE(RDS) and RMSE(DOE) are the root mean square errors of RDS and DOE estimates, respectively. By definition, we have PRiAME(RDS) = PRiRMSE(RDS) = 0 (%), and perfect estimates will yield PRiAME(·) = PRiRMSE(·) = 100 (%).

Fig. 6 shows, in the form of dotcharts, PRiAME(·) and PRiRMSE(·) over various ranges of observed rainfall amount. They indicate that: (1) DOE estimates are more accurate and less biased than RDS estimates over all ranges of rainfall amount, (2) the improvement of DOE over RDS is most significant over areas of no to light rainfall, and (3) SOE estimates are in general more accurate than RDS estimates, but more biased over ranges of no to small rainfall amount.

Because forecasters interpret estimated rainfall fields primarily through visualization, it is important

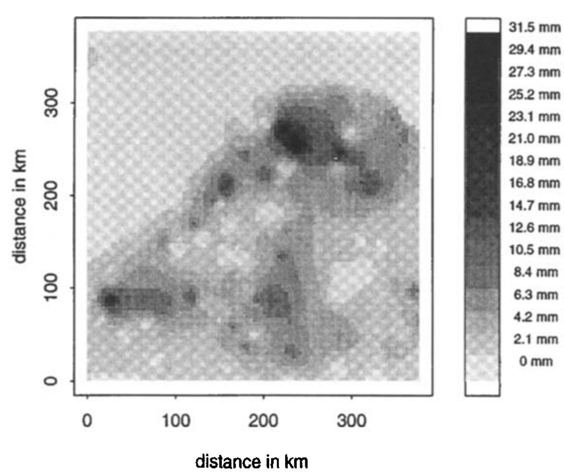

Fig. 7. An example of RDS-estimated hourly rainfall field.

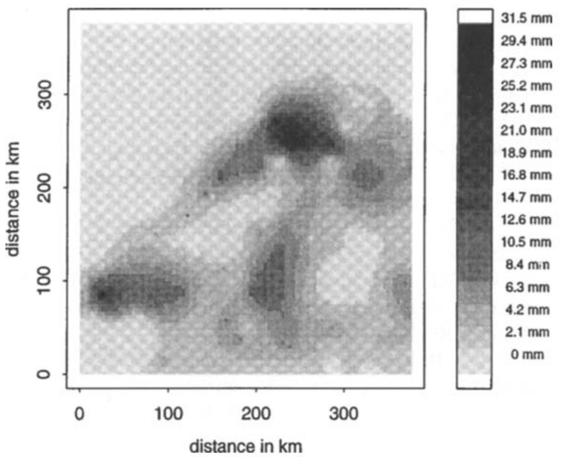

Fig. 8. An example of DOE-estimated hourly rainfall field.

to examine visual differences among them. Figs. 7 and 8 show examples of RDS- and DOE-estimated rainfall fields, respectively. They are obtained by performing estimation on the HRAP grid (National Weather Service, 1993). The grid size is approximately 4×4 km^2 in the study area. The radar umbrella circumscribes the square plotting area in the figures. The SOE-estimated field looks similar to the DOE-estimated, and hence is not shown. The corresponding radar rainfall field, which is also on the HRAP grid and free of visually recognizable errors associated with radar observation of rainfall, such as ground clutter, ground returns from anomalous propagation (AP), bright band enhancement, etc., is shown in Fig. 9. It is seen that Fig. 8 has a more ''structure'' to both

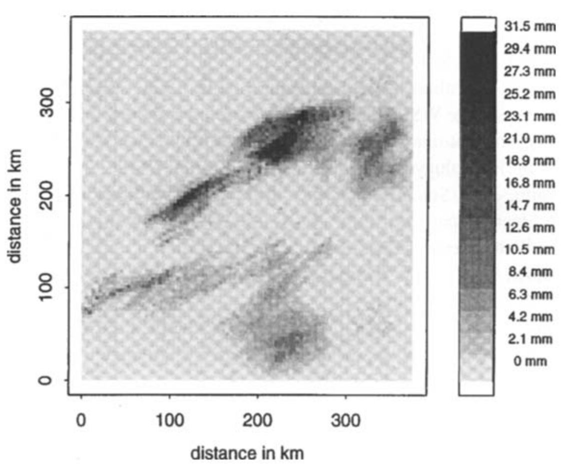

Fig. 9. Hourly radar rainfall field corresponding to Figs. 7 and 8.

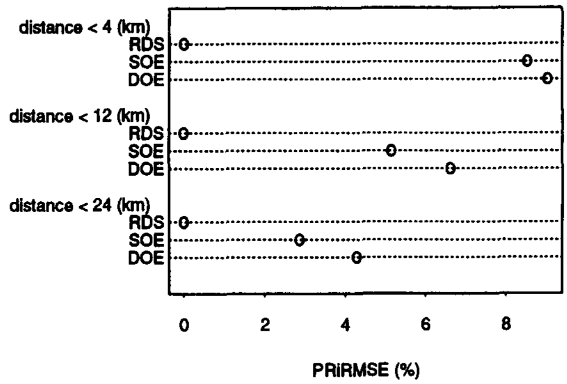

Fig. 10. Percent reduction in root mean square error over various distances to the nearest gage from the point of estimation.

no-rain areas and rainfall over rain areas than Fig. 7, and, in reference to the radar rainfall field, looks to be the more realistic of the two. In this visual comparison, it is extremely important to keep in mind that what is meant by an ''estimated'' rainfall field in this work refers to a conditional ''expectation'' field, as opposed to a conditional ''simulation'' field (see, e.g. Deutsch and Journel, 1992): the distinction is that, whereas the latter represents a plausible realization of what could have occurred, the former represents the ensemble average of many such realizations (hence the circular rainfall gradients surrounding rain gage locations in Figs. 7 and 8).

In Fig. 7, the unrealistically pronounced patterns of circular gradients in ''expected'' rainfall around the gage locations suggest that RDS estimates may be prone to larger errors than SOE or DOE estimates near gage locations. Fig. 10, which shows PRiRMSE(·) as stratified according to the distance between the point of validation and the location of the nearest gage, confirms that SOE and DOE estimates are significantly more accurate in the immediate vicinity of rain gage locations than RDS estimates.

8. Summary and conclusions

Two gage-only rainfall analysis procedures that explicitly account for both within-storm variability of rainfall and variability due to fractional coverage are described; (1) double optimal estimation (DOE) involving analogues of indicator and simple kriging to estimate probability of rainfall and rainfall amount given raining, respectively, and (2) single optimal estimation (SOE) involving an analogue of simple kriging. To compare the estimation procedures with the reciprocal distance-squared method (RDS), cross validation was performed using hourly rain gage data from the operational network under the Tulsa, Oklahoma, WSR-88D umbrella.

Results indicate that: (1) DOE estimates are more accurate and less biased than RDS estimates over all ranges of rainfall amount, (2) the improvement of DOE over RDS estimates is most significant over areas of no to light rainfall, (3) SOE estimates are in general more accurate than RDS estimates, but more biased over ranges of no to small rainfall amount, and (4) SOE and DOE produce much more realistic-looking ''expected'' rainfall fields than RDS.

Acknowledgements

This work was supported by the Advanced Weather Information Processing System (AWIPS) Program and the NEXRAD Product Improvement (NPI) Program, both of the National Weather Service. This support is gratefully acknowledged.

References

Austin, P.M., 1987. Relation between measured radar reflectivity and surface rainfall, Mon. Wea. Rev., 115, 1053–1070.

Azimi-Zonooz, A., Krajewski, W.F., Bowles, D.S., Seo, D.-J., 1989. Spatial rainfall estimation by linear and nonlinear cokriging of radar rainfall and rain gage data, Stoch. Hydrol. Hydraul., 3, 51–67.

Barancourt, C., Creutin, J.D., Rivoirard, J., 1992. A method for delineating and estimating rainfall fields, Water Resour. Res., 28 (4), 1133–1144.

Bastin, G.B., Lorent, C., Duque, M., 1984. Optimal estimation of the average rainfall and optimal selection of raingage locations, Water Resour. Res., 20 (4), 463–470.

Chow, V.T., Maidment, D.R., Mays, L.W., 1988. Applied Hydrology. McGraw-Hill, New York.

Creutin, J.D., Obled, C., 1982. Objective analysis and mapping techniques for rainfall fields: an objective comparison, Water Resour. Res., 18 (2), 413–431.

Creutin, J.D., Delrieu, G., Lebel, T., 1988. Rain measurement by raingage–radar combination: a geostatistical approach, J. Atmos. Oceanic Technol., 5 (1), 102–115.

Deutsch, C.V., Journel, A.G., 1992. GSLIB Geostatistical Software Library and User's Guide. Oxford University Press, New York.

Journel, A.G., Huijbregts, Ch.J., 1978. Mining Geostatistics. Academic Press, London.

Journel, A.G., 1983. Nonparametric estimation of spatial distributions, Math. Geol., 15 (3), 445–468.

Karlin, S., Taylor, H.M., 1975. A First Course in Stochastic Processes. Academic Press, New York.

Krajewski, W.F., Anagnostou, E.N., Ciach, G.J., 1996. Effects of the radar observation process on inferred rainfall statistics. J. Geophy. Res. 101 (D21), 493–502.

Lebel, T.G., Bastin, G., 1985. Variogram identification by the mean-square interpolation error method with application to hydrology fields, J. Hydrol., 77, 31–56.

Lebel, T.G., Bastin, G., Obled, C., Creutin, J.D., 1987. On the accuracy of areal rainfall estimation: A case study. Water Resour. Res. 23 (11), 2123–2134.

Matheron, G., 1975. A simple substitute for conditional expectation: disjunctive kriging. In: Guarascio, M., David, M., Huijbregts, C., (Eds.), NATO ASI Geostat 1975. Reidel, Dordrecht, The Netherlands.

National Weather Service, 1993. National Weather Service River Forecast System (NWSRFS) User's Manual. Office of Hydrology, National Weather Service, Silver Spring, MD.

Schweppe, F.C., 1973. Uncertain Dynamic Systems. Prentice-Hall, Englewood Cliffs, NJ.

Seo, D.-J., 1996. Nonlinear estimation of spatial distribution of rainfall—an indicator cokriging approach, Stoch. Hydrol. Hydraul., 10 (2), 127–150.

Seo, D.-J., Smith, J.A., 1991. Rainfall estimation using raingages and radar—a Bayesian approach: 2. An application, Stoch. Hydrol. Hydraul., 5, 31–44.

Seo, D.-J., Smith, J.A., 1996. Characterization of the climatological variability of mean areal rainfall through fractional coverage, Water Resour. Res., 32 (7), 2087–2095.

Seo, D.-J., Krajewski, W.F., Azimi-Zonooz, A., Bowles, D.S., 1990. Stochastic interpolation of rainfall data from rain gages and radar using cokriging. 2. Results, Water Resour. Res., 26 (5), 915–924.

Seo, D.-J., Fulton, R.A., Breidenbach, J.P., 1998. Rainfall estimation in the WSR-88D era for operational hydrologic forecasting in the National Weather Service. Preprints. Special Symposium on Hydrology. Am. Meteor. Soc., Boston, MA.

Smith, J.A., Seo, D.-J., Baeck, M.L., Hudlow, M.D., 1996. An intercomparison study of NEXRAD precipitation estimates, Water Resour. Res., 32 (7), 2035–2045.

Solow, A.R., 1986. Mapping by simple indicator kriging, Math. Geol., 18 (3), 335–352.

Tabios III, G.Q., Salas, T.D., 1985. A comparative analysis of techniques for spatial interpolation of precipitation, Water Resour. Bull., 21 (3), 365–380.

Wilson, J.W., Brandes, E.A., 1979. Radar measurement of rainfall, Bull. Amer. Meteor. Soc., 60, 1048–1058.