Estimation of the Mean Field Bias of Radar Rainfall Estimates

JAMES A. SMITH

Department of Civil Engineering and Operations Research, Princeton University, Princeton, New Jersey

WITOLD F. KRAJEWSKI

Iowa Institute of Hydraulic Research, Department of Civil and Environmental Engineering, The University of Iowa, Iowa City, Iowa

(Manuscript received 31 October 1989, in final form 21 September 1990)

ABSTRACT

In this paper procedures are developed for estimating the mean field bias of radar rainfall estimates. Mean field bias is modeled as a random process that varies not only from storm to storm but also over the course of a storm. State estimates of mean field bias are based on hourly raingage data and hourly accumulations of radar rainfall estimates. The procedures are developed for the precipitation processing systems used with products of the Next Generation Weather Radar (NEXRAD) system. To implement the state estimation procedures, parameters of the bias model must be specified. Likelihood-based procedures are developed for estimating these parameters. A simulation experiment is carried out to assess performance of the parameter estimation procedure. Convergence of parameter estimators is rapid for the cases studied, with data from approximately 25 storms providing parameter estimates of acceptable accuracy. The state estimation procedures are applied to radar and raingage data from the 27 May 1987 storm, which was centered near the NSSL radar in Norman, Oklahoma. The results highlight dependence of the state estimation problem on the parameter estimation problem.

1. Introduction

Ahnert et al. (1983) note that "in spite of efforts to maintain a high level of quantitative accuracy in estimating precipitation from radar data, there are sure to be errors in these estimates. While some of these errors will be localized or perhaps range dependent, others will often produce a uniform multiplicative bias in the radar estimated precipitation." In this paper procedures are developed for estimating the multiplicative bias of radar estimates of rainfall, which we term the mean field bias. Mean field bias is modeled as a random process that varies from storm to storm and, during a storm, on an hourly time scale. The estimation procedures are designed for the precipitation processing systems used for the Next Generation Weather Radar (NEXRAD) system. Discussions of general design features of the precipitation processing systems are given in Hudlow et al. (1985), Ahnert et al. (1983), Shedd et al. (1989), Smith et al. (1989), and Hudlow et al. (1989). In this paper attention is focused on procedures for automated correction of mean field bias using hourly accumulation data from a network of raingages.

Mean field bias represents one component of the error in radar rainfall estimates and is defined in section 2 in the context of a broader model of error structure of radar rainfall estimates. Implicit in the model is the assumption that radar reflectivity data have been subjected to basic quality control procedures to mitigate the effects of ground clutter, anomalous propagation, and isolated nonmeteorological targets (as described, for example, in Hudlow et al. 1983). For stratiform rain, it is also particularly important to account for brightband effects (see, for example, Smith 1986). From the model of section 2, the mean field bias can be interpreted as randomizing the multiplicative parameter in the *Z–R* relationship. It has been recognized for many years that drop-size distributions can vary significantly from storm to storm (see Joss and Waldvogel 1989; Austin 1987; Battan 1973) and during storms (see Cataneo and Stout 1968; Waldvogel 1974; Carbonne and Nelson 1978). One component of the error introduced into rainfall estimates by temporal variation of drop-size distribution can be represented as a time-varying mean field bias. Variation in drop-size distribution is not, however, the only process that can result in a mean field bias. Similar effects can be attributed to processes that affect the power transmission of the radar, for example, wet radome attenuation, and variation in the power output of the transmitter (wet radome effects and the effects of variation in the power output can generally be kept to levels small in comparison with other error sources). The circumstances described previously in which temporal vari-

Corresponding author address: Dr. James A. Smith, Department of Civil Engineering and Operations Research, School of Engineering/Applied Science, Princeton University, Princeton, New Jersey 08544.

ation in radar bias occurs involve the complex interaction of numerous processes. Consequently the mean field bias is modeled as a random process.

Manual processing of radar rainfall estimates in a research setting by a highly trained radar hydrometeorologist with access to automated computer processing tools and extensive ancillary hydrometeorological information will likely provide the highest quality radar rainfall products (as illustrated by Austin 1987). The precipitation processing systems that will reside at the NEXRAD computer and at the National Weather Service meteorological forecast offices are, however, designed to run in a fully automated mode. Furthermore, the external information that is available to the NEXRAD computer consists only of raingage data. These considerations play an important role in the procedures developed in this paper for estimation of the mean field bias.

The procedures developed in this paper follow along the lines of those developed in the work by Ahnert et al. (1984) in using a state estimation formulation of the problem. The sequential estimation procedure of Cain and Smith (1976) is also similar. Other procedures for bias correction are reported in Wilson and Brandes (1979), Collier et al. (1983), and Collier (1986). More generally, optimal estimation procedures for combining gage and radar data are reported in Krajewski (1987), Seo et al. (1989), and Seo and Smith (1990). The bias correction procedure reported in this paper is one component of NEXRAD precipitation processing in which optimal estimation procedures are applied following bias correction.

Our approach to the state estimation problem follows statistical approaches (Harrison and Stevens 1976; Smith and Karr 1985; Karr 1986) rather than engineering control approaches (Gelb 1973). Two consequences are the attention given to parameter estimation (section 4) and exploitation of distributional assumptions in deriving state estimators.

Contents of the sections are as follows. The statistical model relating rainfall rate to equivalent reflectivity factor is presented in section 2. Interpretations of the bias component of the model are presented and distributional results for the bias process are derived. Also section 2 develops the observation equation that relates radar and raingage data to the model bias. In section 3 the state estimation problems of filtering, prediction, and smoothing are presented and updating procedures are developed for solving these problems. The role of each type of state estimation problem in operational use of NEXRAD precipitation estimates is described. In section 4 likelihood-based parameter estimation procedures for the bias model are developed and tested via a Monte Carlo experiment. The state estimation procedures are applied in section 5 to a sample data case from the NSSL radar in Norman, Oklahoma. The summary and conclusions are given in section 6.

2. Model formulation

In this section a statistical model is developed that relates radar measurements of rainfall to the true rainfall field; that is, the rainfall that would be observed at the ground by an error-free space–time sensor. A component of the model is a time-varying random process representing mean field bias. Specification of the model serves two purposes for the present study. It provides a precise interpretation of the mean field bias. It also serves as the basis for development of an observation equation that relates radar and raingage measurements to the mean field bias. The observation equation and model of mean field bias are the principal tools required for procedure development to estimate the mean field bias. Before presenting the model, some notation is introduced.

Precipitation rate at time τ and spatial location x is denoted $\xi_\tau(x)$. The index τ represents time, in hours, since the last period of no rainfall. The precipitation rate process for scan t of hour s averaged spatially over the bin specified by azimuth i and range j is denoted

$$R_{s,t}(i,j) = |D_{ij}|^{-1} \int_{D_{ij}} \xi_\tau(x)dx \quad \tau = (s-1) + t\Delta t \tag{1}$$

where D_{ij} is the land area beneath the radar sample volume with azimuth i and range j, $|D_{ij}|$ is the surface area associated with D_{ij}, and Δt is the time resolution of radar observations (in hours). It is assumed that the region D_{ij} is large enough that wind drift has a minor effect on the rainfall observation process. The number of scans γ during an hour is $1/\Delta t$. For the NEXRAD system the time resolution is approximately 6 min ($\Delta t = 0.1$ h) during precipitation periods so the number of scans during an hour will be 10. As noted in (1) the times τ for which observations are available, relative to the start of the storm, are given by $(s-1) + t\Delta t$, with t ranging from 1 to γ. The equivalent radar reflectivity factor for scan t in hour s at azimuth i and range j is denoted $Z_{s,t}(i,j)$.

The statistical model presented below specifies that rainfall rate can be represented as the product of two terms. The first term is a power function of equivalent reflectivity factor with range-dependent parameters. The second term specifies a multiplicative error model for radar rainfall estimates. The statistical model is expressed as follows:

$$R_{s,t}(i,j) = [a(j)Z_{s,t}(i,j)^{b(j)}][B(s)\epsilon_{s,t}(i,j)]. \tag{2}$$

In the formulation given previously the model can be interpreted as a regression model for log rainfall rate versus log reflectivity factor. The error field ϵ has, for each s, t, i, and j, a lognormal distribution with mean 1 and range-dependent standard deviation. The mean field bias $B(s)$ is a Markov chain with median 1. Its complete distribution is specified in the following. Unlike the mean field bias, the error field ϵ is spatially varying over the radar field and varies from scan to scan. Both error processes are assumed to be mutually independent and to be independent of the reflectivity process. Underlying model assumptions on the error processes is the assumption that rainfall rate and reflectivity factor follow a lognormal distribution. The lognormal assumption for rainfall rate and reflectivity factor has been made by a number of authors (see Kedem and Chiu 1987; Seo and Smith 1990 for discussion and additional references).

An alternative representation of the model is

$$R_{s,t}(i,j) = [A_s(j)Z_{s,t}(i,j)^{b(j)}][\epsilon_{s,t}(i,j)] \quad (3)$$

where

$$A_s(j) = a(j)B(s). \quad (4)$$

This formulation leads to the interpretation of the bias process as producing a randomized Z–R relationship. In this case the randomized multiplicative coefficient is given by the process $\{A_s(j)\}$. The formulation of (3) also leads to an interpretation of the statistical model as a "random coefficient" regression model (see Johansen 1984). This interpretation is useful in development of parameter and state estimation procedures.

Denote the natural logarithm of the mean field bias for hour s by $\beta(s)$; that is,

$$\beta(s) = \ln[B(s)]. \quad (5)$$

The quantity $\beta(s)$ is assumed to be a stationary Markov chain over the integers $[1, \ldots, T]$ satisfying

$$\beta(s) = a_1\beta(s-1) + W(s); \quad W(s) \sim N(0, v) \quad (6)$$

where $0 \leqslant a_1 \leqslant 1$, v is nonnegative, T is the storm duration in hours, and $W(s)$ is a sequence of independent normally distributed random variables with mean 0 and variance v. The log bias process has mean 0. We denote the stationary variance of the log bias process by the parameter a_2; that is,

$$a_2 = \mathrm{var}[\beta(s)]; \quad s = 0, \cdots, T. \quad (7)$$

The correlation function of the log bias process is given by

$$\mathrm{cor}[\beta(s), \beta(s+k)] = a_1{}^k. \quad (8)$$

Two very distinct conceptual models of radar bias have simple representations in terms of model parameters. It follows from the assumption of stationarity that

$$v = a_2(1 - a_1{}^2). \quad (9)$$

Note that if a_1 equals 1, v must equal 0. If the correlation parameter a_1 equals 1, the bias process can vary randomly from storm to storm, but is fixed over the duration of a storm. The log bias, which applies over the duration of a storm, has a normal distribution with mean 0 and variance a_2. Conversely, if the correlation parameter is less than 1, the bias varies not only from storm to storm but also over the course of a storm. A likelihood ratio test is developed in section 4 to distinguish the two cases. It is illustrated in section 5 that bias estimation procedures perform quite differently in the two cases.

To develop procedures for estimating the bias process $B(s)$ it is necessary to specify the relationship between radar and raingage observations and the mean field bias. The number of raingages reporting measurable rainfall for hour s of the storm is denoted $\eta(s)$. The accumulated rainfall measured by gage k during hour s is denoted $G_s(k)$. Gage locations are specified in terms of the radar grid coordinates; the location of the kth gage is denoted $[i(k), j(k)]$. It follows from (2) that

$$B(s) = \frac{\sum_{k=1}^{\eta(s)} \sum_{t=1}^{\gamma} R_{s,t}[i(k), j(k)]}{\sum_{k=1}^{\eta(s)} \sum_{t=1}^{\gamma} a[j(k)]Z_{s,t}[i(k), j(k)]^{b[j(k)]}\epsilon_{s,t}[i(k), j(k)]}. \quad (10)$$

Recall that γ is the number of radar samples in an hour. Based on (10) and the lognormality of $B(s)$, the following approximation is used for the observation equation:

$$Y(s) = \beta(s) + M(s); \quad M(s) \sim N\{(0, \sigma[\eta(s)]^2\} \quad (11)$$

where

$$Y(s) = \ln\left(\frac{\sum_{k=1}^{\eta(s)} G_s(k)}{\sum_{k=1}^{\eta(s)} \sum_{t=1}^{\gamma} a[j(k)]Z_{s,t}[i(k), j(k)]^{b[j(k)]}}\right); \quad (12)$$

$\sigma(\eta)$ is a nonnegative function representing the observation error given that the number of gages with measurable rainfall is η, and $M(s)$ is a sequence of independent normally distributed random variables with mean 0 and variance $\sigma(\eta)^2$. It is assumed that the error function is a power function of the number of gages; that is,

$$\sigma(\eta)^2 = a_3\eta^{a_4}. \quad (13)$$

The power-law form of (13) allows the error model to account for correlation of gage observations and is consistent with results obtained by Rodriguez-Iturbe and Mejia (1974) in a study of sampling errors for rainfall field estimation.

The observation $Y(s)$ is the log ratio of mean gage rainfall to mean radar rainfall at gage locations. The error process $M(s)$ accounts for the approximation of gage observations for the numerator in (10) and for the approximation of radar observations to the random variable in the denominator of (10). Therefore, both discrete approximation and measurement errors are taken into account.

3. State estimation of mean field bias

The formulations of the bias model and observation model in section 2 provide a model system in which procedures can be developed for correcting radar rainfall estimates for mean field bias. The procedures developed in the following will be applied in the NEXRAD precipitation processing algorithms. The resulting precipitation estimates serve multiple purposes. They are used for flash flood forecasting, main stem flood forecasting, routine river flow and stage forecasting, and in preparation of long-term hydrologic forecasts. Automated raingage data are required, along with radar data, for implementation of the bias estimation procedures. For the majority of radar umbrellas in the United States it will be possible to obtain more than 30 gages with automated hourly data. For more than one-third of the radar umbrellas 100 or more automated gages are currently available.

In the flash flood application interest focuses on the most recent rainfall estimates. To correct these estimates for mean field bias the bias must be estimated for the most recent hour given observations prior to and including the most recent hour. This type of state estimation problem is referred to as a filtering problem. For main stem river forecasting, with the longer response times of catchments to precipitation, precipitation estimates prior to the most recent hour are of significant interest. Consequently, in some situations we will want to correct for bias in preceding hours given observations prior to, including, and following a given hour. This state estimation problem is referred to as a smoothing problem. The final type of problem we may face is one in which inadequate gage data are available for the current hour to make a bias computation. In this case we will want to estimate the current bias from observations preceding the current hour. This problem is one of prediction. We define our state estimation problems more formally in the following.

State estimation is distinguished from parameter estimation by virtue of the fact that the objects to be estimated are random variables rather than unknown real-valued parameters. In the problem at hand the mean field bias $\{B(s)\}$ has been modeled as a random process with distributional law specified by (5)–(7). The observations related to the bias process are specified by the observation Eq. (12). State estimators are derived, whenever possible, as the conditional expectation of the process given the observations. Justification of the conditional expectation criterion is given by Karr (1986). We define our state estimation problems in the following.

Let $Y(1), \ldots, Y(T)$ be observations of log radar bias, as defined in (12). The state estimation problem is to compute the conditional expectation of the bias $B(s)$ given observations $Y(1), \ldots, Y(u)$ for u less than or equal to T; that is,

$$\hat{B}(s|u) = E[B(s)|Y(u), \cdots, Y(1)]. \quad (14)$$

If $u \leqslant s$, the problem is one of prediction; if $u = s$, the problem is one of filtering; and if $u > s$, the problem is one of smoothing.

To evaluate accuracy of the state estimators we want to compute the conditional error variance

$$V(s|u) = E\{[\hat{B}(s|u) - B(s)]^2|Y(u), \cdots, Y(1)\}. \quad (15)$$

The bias model is defined in terms of the log bias process $\beta(s)$. State estimators for $B(s)$ will consequently be derived in terms of state estimators for $\beta(s)$. The following argument shows how this is done. From (6) and (11) it can be seen that the random variables $\beta(s), Y(1), \ldots, Y(u)$ have a multivariate normal distribution. It follows from Theorem 2.5.1 in Anderson (1958) that the conditional distribution of $\beta(s)$ given $Y(1), \ldots, Y(u)$ is normal, we will write

$$[\beta(s)|Y(u), \ldots, Y(1)] \sim N[\hat{\beta}(s|u), \textstyle\sum(s|u)] \quad (16)$$

where

$$\hat{\beta}(s|u) = E[\beta(s)|Y(u), \cdots, Y(1)] \quad (17)$$

and

$$\textstyle\sum(s|u) = E\{[\hat{\beta}(s|u) - \beta(s)]^2|Y(u), \cdots, Y(1)\}. \quad (18)$$

It follows that the conditional distribution of $B(s)$ given $Y(1), \ldots, Y(u)$ is lognormal with parameters $\hat{\beta}(s|u)$ and $\sum(s|u)$. Consequently,

$$\hat{B}(s|u) = \exp[\hat{\beta}(s|u) + (1/2)\textstyle\sum(s|u)] \quad (19)$$

and

$$V(s|u) = \hat{B}(s|u)^2\{\exp[\textstyle\sum(s|u)] - 1\}. \quad (20)$$

The remaining problem is to compute state estimates for log bias. We begin with the filtering problem for $\beta(s)$. It will be seen that the smoothing and prediction problems simplify to filtering problems. The conditioning argument u is suppressed for filtering problems, so $\hat{\beta}(s)$ equals $\hat{\beta}(s|s)$ and $\sum(s)$ equals $\sum(s|s)$.

The Kalman filter can be used to recursively compute $\hat{\beta}(s)$ and $\Sigma(s)$. To initiate the procedure, note that

$$\hat{\beta}(0) = E[\beta(0)] = 0 \quad (21)$$

and

$$\Sigma(0) = E\{[\hat{\beta}(0) - \beta(0)]^2\} = a_2. \quad (22)$$

For $s > 1$, the conditional expectations can be computed recursively using the following relations, developed originally by Kalman (1960):

$$\hat{\beta}(s) = a_1\hat{\beta}(s-1) + \{H(s)/[a_3\eta(s)^{a_4} + H(s)]\}e(s) \quad (23)$$

$$\Sigma(s) = H(s)\{1 - H(s)/[a_3\eta(s)^{a_4} + H(s)]\} \quad (24)$$

where

$$e(s) = Y(s) - a_1\hat{\beta}(s-1) \quad (25)$$

is the "innovation" and

$$H(s) = a_1^2 \Sigma(s-1) + a_2(1 - a_1^2). \quad (26)$$

Additional details on the Kalman filter are given in Gelb (1973).

The prediction problem is easily solved by noting that

$$\beta(s+k) = a_1^k\beta(s) + \sum_{j=0}^{k-1} a_1^j W(s+k-j). \quad (27)$$

It follows that

$$\hat{\beta}(s+k|s) = a_1^k\hat{\beta}(s). \quad (28)$$

Similarly,

$$\Sigma(s+k|s) = a_1^{2k}\Sigma(s) + \sum_{j=0}^{k-1} a_1^{2j}a_2(1 - a_1^2). \quad (29)$$

The smoothing problem can be decomposed into two parts: computation of a "backwards" filter and combination with the "forward" filter of (23). The key to solving the smoothing problem is the observation that if the time sequence is reversed, the bias process remains a Markov process. In other words, if we define

$$\tilde{\beta}(s) = \beta(T - s + 1) \quad (30)$$

then $\tilde{\beta}(s)$ is a Markov process with the same distribution as $\beta(s)$. This result follows from the facts that the random variables $\beta(1), \ldots, \beta(T)$ have a multivariate normal distribution with symmetric covariance function (see Weiss 1975 for additional details). To proceed, the following additional notation is required for the backwards filter:

$$\hat{\beta}_b(s) = E[\beta(s)|Y(s), \cdots, Y(T)] \quad (31)$$

$$\Sigma_b(s) = E\{[\hat{\beta}_b(s) - \beta(s)]^2|Y(s), \cdots, Y(T)\}. \quad (32)$$

The backwards filter and its variance are computed using the recursive relations of (23) and (24) with the observations reversed in time.

It follows from normality of the estimators (see Anderson 1958) that for $u > s$,

$$\hat{\beta}(s|u) = c\hat{\beta}(s) + (1 - c)\hat{\beta}_b(s) \quad (33)$$

where

$$c = \frac{\Sigma_b(s)}{\Sigma_b(s) + \Sigma(s)}. \quad (34)$$

We also have

$$\Sigma(s|u) = \frac{\Sigma_b(s)\Sigma(s)}{\Sigma_b(s) + \Sigma(s)}. \quad (35)$$

It can be easily seen that the variance of the smoothed estimator is at least as small as the filter estimator by rewriting (35) as follows:

$$\Sigma(s|u) = \Sigma(s)\left[\frac{1}{1 + \Sigma(s)/\Sigma_b(s)}\right]. \quad (36)$$

In this section it has been implicitly assumed that parameters of the observation equation and bias model are known. Generally, this will not be the case. In order to effectively implement the state estimation procedures we must develop procedures to estimate unknown parameters of the model. We address this problem in the following section.

4. Parameter estimation and hypothesis testing

Paired hourly gage and NEXRAD rainfall estimates will be archived for parameter estimation and quality control applications. In this section parameter estimation and hypothesis testing procedures for the mean field bias model are described. From the paired gage radar observations we can obtain samples of the observation time series $Y(s)$ defined in (12). We will denote the sample log bias for hour s of the ith storm by $Y_i(s)$. This section will develop likelihood-based inference procedures for the bias model using the sample observations $\{Y_i(s); i = 1, \cdots, n; s = 1, \cdots, T_i\}$ where n denotes the total number of archived storms and T_i denotes the number of hours in the ith storm. Seasonality is a prominent feature of our problem. The development given below, however, assumes stationarity not only during a storm, but also from storm to storm. The modifications to account for seasonality are straightforward but notationally cumbersome.

The log-likelihood function of the observation process $\{Y_i(s)\}$ is the natural logarithm of the joint density of $Y_1(1), \ldots, Y_n(T_n)$; that is,

$$L_n(a) = \log\{p[Y_1(1), \cdots, Y_1(T_1), \cdots, Y_n(1), \cdots, Y_n(T_n)|\mathbf{a}]\} \quad (37)$$

where $\mathbf{a}$ is the vector of parameters $\mathbf{a} = (a_1, a_2, a_3, a_4)$. The main result used in parameter estimation is the following representation of the log-likelihood function for the model of (11) and (12) given observations from n storms:

$$L_n(a) = -\frac{1}{2}\sum_{i=1}^{n} \frac{Y_i(1)^2}{a_2 + a_3\eta_i(1)^{a_4}} + \ln[a_2 + a_3\eta_i(1)^{a_4}] - \frac{1}{2}\sum_{i=1}^{n}\sum_{s=2}^{T_i} \frac{[Y_i(s) - a_1\hat{\beta}_i(s-1)]^2}{a_1^2 \Sigma_i(s-1) + a_3\eta_i(s)^{a_4} + a_2(1 - a_1^2)} - \frac{1}{2}\sum_{i=1}^{n}\sum_{s=2}^{T_i} \ln[a_1^2 \Sigma_i(s-1) + a_3\eta_i(s)^{a_4} + a_2(1 - a_1^2)] + C \quad (38)$$

where C is a constant. The proof of this result is given in the Appendix.

Maximum likelihood estimators $\hat{\mathbf{a}} = (\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{a}_4)$ are solutions, if they exist, to the system of equations

$$\nabla L_n(a) = 0 \quad (39)$$

where ∇ is the gradient operator. Maximum likelihood estimators cannot be derived analytically from (39). A gradient search method (Press et al. 1988) was developed to numerically solve the system of likelihood equations.

The parameter estimators obtained from (39) have the attractive large sample properties usually associated with maximum likelihood estimators. Asymptotic normality and consistency (see Bickel and Doksum 1977, for discussion) can be established by procedures similar to those presented in the work by Smith (1987).

A likelihood ratio test can be developed using (40) to distinguish between bias models in which random variation of bias occurs over the course of a storm versus the model in which the only variability in the bias process is storm to storm variability. Recall from section 2 that the latter situation results if the correlation parameter equals 1. Denote the maximum likelihood estimator of the parameter vector $\mathbf{a}$, under the constraint that a_1 equals 1, by $\tilde{\mathbf{a}}$. The log-likelihood ratio statistic

$$\Lambda_n = -2[L_n(\tilde{a}) - L_n(\hat{a})] \quad (40)$$

has a χ^2 distribution with 1 degree of freedom, under the condition that a_1 equals 1 (Bickel and Doksum 1977). The hypothesis that a_1 equals 0 is rejected if $\Lambda_n > \chi^2(1-\alpha)$ where $\chi^2(1-\alpha)$ is the $1-\alpha$ quantile of a χ^2 distribution with one degree of freedom.

Similar tests can be developed to determine whether the bias is nonrandom (test $a_2 = 0$ versus $a_2 > 0$), the number of raingages influences the observation error (test $a_4 = 0$ versus a_4 different from 0), or significant error is present in the bias observations (test $a_3 = 0$ versus $a_3 > 0$). In the case that a_4 equals 0; that is, the number of gages recording measurable rainfall has no influence on the observation error, the system of likelihood equations will not have unique solutions. In this case it is, in effect, impossible to distinguish between observation equation error and variability in the bias process.

In order to examine the small sample properties of the parameter estimators, an extensive Monte Carlo simulation experiment was performed. The true values of the parameters a_1, a_2, a_3, and a_4 were assumed and both the bias process and its observations were generated for a specified number of raingages and storms. The number of hours within a storm was assumed to have a Poisson distribution with a specified mean. The number of raingages reporting nonzero rain was simulated by a Gaussian random variable with a specified mean and standard deviation. This choice was dictated by convenience in investigating the influence of the number of gages on the parameter estimators.

Two different cases were analyzed in detail. The first one corresponds to $a_1 = 0.8$ which means that the bias process varies both from storm to storm and during storms. The values of the other parameters were selected to be $a_2 = 0.1$, $a_3 = 1.0$, and $a_4 = -1.0$. Figure 1 shows the results based on 100 realizations. The points correspond to the mean value of the estimated parameters with the vertical bars marking a range of one standard deviation. A notable feature of the results is that, for the conditions represented by the chosen parameter values, the estimators are quickly converging to the true values as the number of storms increases. For 100 storms (with average duration of 5 h) one should get very good estimates. The situation looks even better for the second scenario (Fig. 2) with $a_1 = 1.0$, $a_2 = 0.1$, $a_3 = 1.0$, and $a_4 = -2.0$. In this case, corresponding to the radar bias varying only between the storms, a dataset of 100 storms provides excellent estimates.

An interesting result is presented in Fig. 3. It shows the effect of variability in the number of raingages reporting nonzero rainfall on the estimation of a_3 and a_4, the two parameters describing the uncertainty in the bias observations. The quality of the estimated values for these parameters does not improve as the average number of gages increases. Since the standard deviation of the number of gages was kept equal to 1

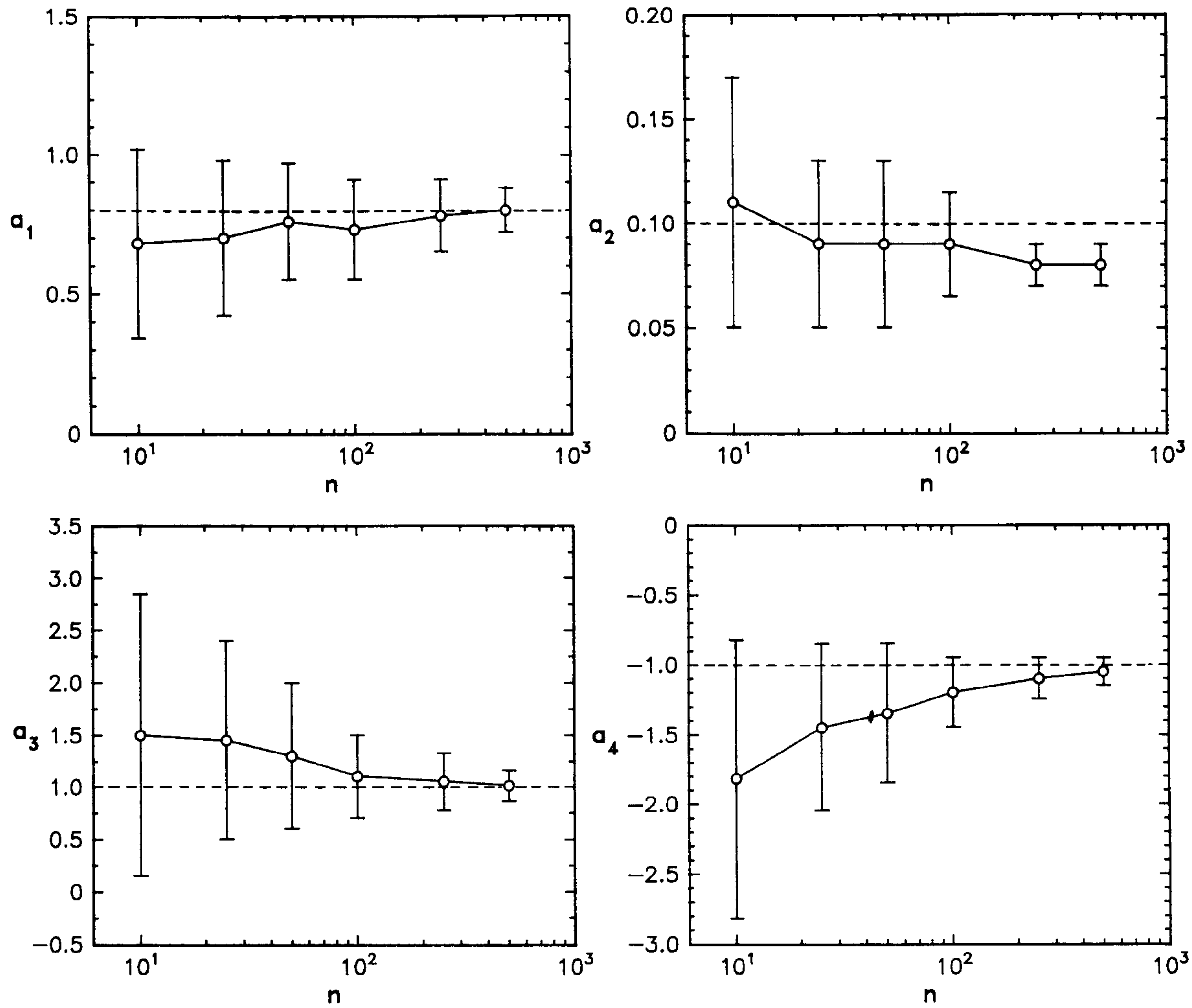

FIG. 1. Results of simulation experiment for $a_1 = 0.8$, $a_2 = 0.1$, $a_3 = 1.0$, and $a_4 = -1.0$. The horizontal scale is in common logarithms of the number of storms. The data points correspond to 10, 25, 50, 100, and 500 storms each with average duration of 5 h. True parameter values are marked with dashed lines. The small circles correspond to the mean value of the estimators from 100 realizations. The vertical bars denote a range of 1 standard deviation. The average number of raingages per hour is 10.

for all the runs here, the results indicate that with a large but essentially fixed number of gages it is difficult to identify the parameters a_3 and a_4. It was noted earlier in this section that if the number of gages has no effect on the bias error; that is, if a_4 equals 0, then the parameters a_2 and a_3 cannot be uniquely distinguished. Figure 3 illustrates a similar feature of the estimation problem. If there is no variability in the number of gages reporting rainfall, the parameters a_3 and a_4 cannot be uniquely distinguished. The practical implications of this result are minimal. Even in the highly unlikely case in which a large relatively constant number of gages is always available, one can perform random deletion to insure adequate variability in the number of gages.

It should be emphasized that the results presented previously, as with any simulation experiment, are tied to the validity of the model. The simulation results imply that if the radar bias model described in section 2 is appropriate for a given setting, one should expect good performance from the parameter estimation procedure. The results of the simulation experiment should be helpful in assessing the value of archive data from the point-of-view of parameter estimation. To evaluate the performance of the bias estimation procedure described in this section and its sensitivity to the model parameters, a separate experiment is underway. Some of these issues are addressed, in a limited way, through application of the state estimation procedure to real data, as described in the following section.

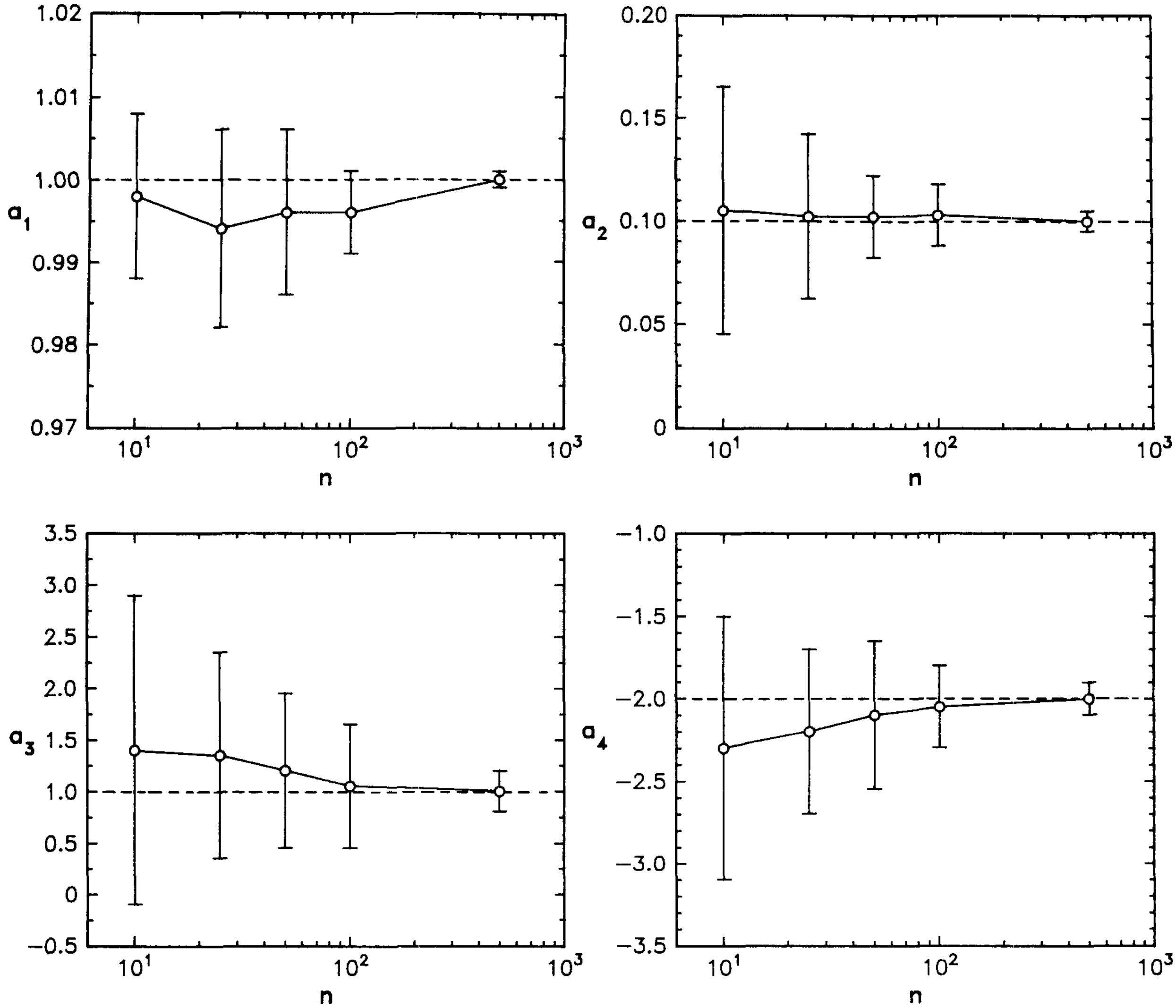

FIG. 2. Results of simulation experiment for $a_1 = 1.0$, $a_2 = 0.1$, $a_3 = 1.0$, and $a_4 = -2.0$. The rest of the experiment is the same as Fig. 1.

5. Application of bias estimation procedures

In this section the state estimation procedures developed in section 3 are applied to radar rainfall estimates derived from the NSSL radar at Norman, Oklahoma, for the storm of 27 May 1987. The storm was responsible for flash floods resulting in loss of life and significant property damage. Meteorological patterns associated with the 27 May storm are typical of flash flood producing storms east of the Rocky Mountains (see Maddox et al. 1979; Chappell 1986). Figure 4 shows rainfall rate estimates derived from the NEXRAD Precipitation Processing System (PPS) for the bin (of size approximately 4 km^2) with highest storm total accumulation. The sequence of peaks in rainfall rate is characteristic of quasi-stationary convective events (Chappell 1986). The Z–R relationship used in the NEXRAD PPS is

$$Z = 300R^{1.4}. \tag{41}$$

The absolute calibration of the NSSL radar is 1 dBZ. Additional details on the NEXRAD PPS are given in Ahnert et al. (1983), Shedd et al. (1989), and Smith et al. (1989).

Hourly raingage observations with positive accumulations and within a range of 150 km of the radar were available for 20 sites over the course of the storm. The maximum hourly gage accumulation was 53 mm (2.1 in); maximum storm total accumulation at a gage was 225 mm (8.9 in). The radar rainfall estimates do not include range correction adjustments (other than the usual $1/r^2$ correction in the radar equation), al-

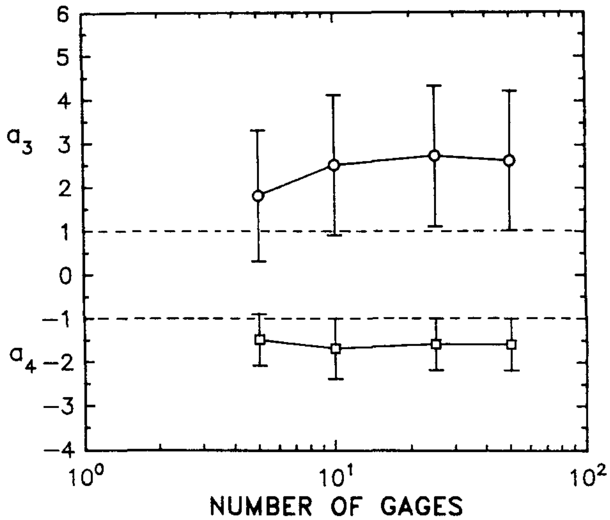

FIG. 3. Effect of variability in the number of raingages on estimates of a_3 and a_4. The horizontal scale shows the average number of raingages reporting rainfall. The standard deviation of the number of gages was held fixed at 1. The number of storms was fixed at 100.

though the NEXRAD PPS does include a parameterized range correction procedure (see Smith et al. 1989). For the present study range effects were minimized by restricting consideration to gages within 150 km of the radar. Over this range, beam elevation can be maintained at a relatively constant elevation (approximately 1.25 km for the center of the beam) using the sectorized hybrid scan strategy of NEXRAD (Shedd et al. 1989).

Table 1 contains summary information by hour on radar and raingage measurements. The sample bias value is the ratio of mean gage rainfall to mean radar accumulation at gage sites [i.e., $\exp\{Y(s)\}$], for $Y(s)$ as defined in (12). The sample bias ranges from a maximum of 2.50 in hour 3 to a minimum of 1.8 in hour 8. Note that the variability of radar observations, as measured by the coefficient of variation (standard deviation divided by the mean), is quite comparable to variability of raingage observations, both in magnitude and temporal pattern. This provides qualitative support for the assumption that multiplicative mean field bias is a significant component of radar error. Note also that the correlation of raingage and radar observations by hour is relatively high, ranging from a minimum of 0.77 to a maximum of 0.96.

Implementation of the state estimation procedure must be carried out in this case without the benefit of parameter estimation. Nominal values of $a_1 = 1.0$, $a_2 = 0.20$, $a_3 = 1.0$, and $a_4 = -1.0$ are used initially for state estimation. Figure 5a shows sample bias values along with bias estimates obtained using the procedures of section 3. The estimated standard deviation of filter and smoothing estimators are shown in Fig. 5b. One clear benefit of the smoothing procedure is that significant improvements over the filter estimates can be obtained for bias values early in a storm, given subsequent observations. The filter estimates can be seri-

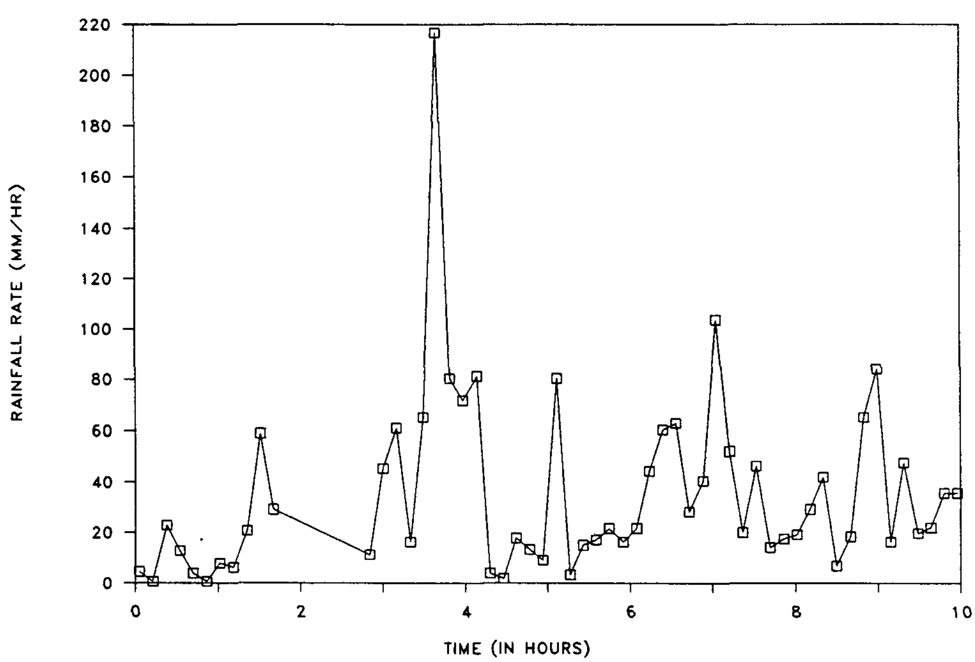

FIG. 4. Estimated rainfall rate from the NSSL radar for the bin with heaviest rainfall during the 27 May 1987 storm in Oklahoma. Precipitation estimates are computed from the NSSL radar data using the NEXRAD Precipitation Processing System.

TABLE 1. Summary statistics for gage radar data.

Hour	Sample bias	Gage mean (mm)	Gage CV	Radar mean (mm)	Radar CV	Correlation
1	1.97	4.43	1.42	2.25	1.35	0.71
2	2.50	4.78	1.73	1.91	1.50	0.86
3	1.71	6.32	1.20	3.69	1.16	0.87
4	1.69	5.88	1.32	3.48	1.46	0.85
5	2.00	6.73	2.02	3.37	2.26	0.96
6	2.56	6.50	1.52	2.54	1.74	0.92
7	2.06	8.93	1.29	4.33	1.42	0.86
8	1.61	6.71	1.24	4.18	1.29	0.70

ously biased for early time periods due to the constraint imposed by the initial conditions of (21) and (22). For the chosen parameters the filter estimates are quite close to the smoothing estimates for hours 2 through 8. Both vary in a small range about the value 2. The estimated standard deviations of the bias estimators illustrate fundamental differences in the filter procedure and smoothing procedure. For the filter estimators, standard deviation of the estimators drops by more than a factor of 2 from hour 1 to hour 8, reflecting the increasing information available to the filter as the storm progresses. For the smoothing procedure the standard deviation varies little over the course of the storm and is slightly smaller than the final standard deviation of the filter estimator.

The bias estimation procedures have qualitatively different behavior if the correlation parameter a_1 is less than 1. Figure 6a shows bias estimates in the case that a_1 equals 0.9 and all other parameters are unchanged. Note that the estimators are more responsive to short-term variation in the sample bias values. As before, the filter estimator is constrained in the first hour by initial conditions. In contrast with the initial case, the filter and smoothing estimators begin to show differences over the course of the storm. Most notably the smoothing estimator is less responsive than the filter to short-term variation in sample bias values. Also in contrast to the initial case the estimated standard deviation (Fig. 6b) for the bias responds to fluctuations in sample values. Note that the estimated sample standard deviation of the smoothing estimator increases substantially from the initial case. A decrease in the correlation parameter translates to a decline in information content for bias estimates.

If the bias value can be accurately characterized, even for a small number of gages, then the bias estimates will be highly compliant with the sample bias values. In Fig. 7a bias estimators are shown for the case in which the observation equation variance parameter a_3 has been dropped from 1.0 to 0.1. Both the smoothing and filter estimators essentially replicate the sample values in this case. Note also in Fig. 7b that if the observation equation variance is quite low, the standard deviation of bias estimators is correspondingly low.

Similar performance of bias estimators is obtained when the variance parameter a_2 is large. In Fig. 8a bias estimates are shown for the case in which a_2 is increased from 0.2 to 1.0 (with a_3 at its initial value of 1). Although the bias estimates are quite similar to those shown in Fig. 7a it is for quite different reasons, as reflected in Fig. 8b, which shows the standard deviation of the estimators. In the present case bias estimates are compliant not because sample bias values are particularly accurate but because the bias equation variance is too large to allow the state estimation procedures to rule out sample fluctuations as insignificant.

The results of this section show that modest changes in parameter values can produce significant changes in performance of the bias estimation procedures for a real-data case. These results underscore the need for efficient parameter estimation procedures and archiving of the necessary data to carry out parameter estimation. The bias estimation procedures developed in this paper are attractive, in comparison to the procedure of Ahnert et al. (1985), in part due to availability of parameter estimation procedures. The bias estimation procedures proposed in this paper are also computationally simpler than the procedure of Ahnert et al. (1985) and have the added capability of operating in the smoothing mode.

6. Summary and conclusions

Mean field bias of radar rainfall estimates is modeled as a random process that varies hourly over the course of a storm. Mean field bias values are independent from storm to storm. Within a storm the sequence of bias values constitutes a Markov process with lognormal distribution. The observation equation that relates radar and raingage data to the random bias is based on the statistical model of equivalent radar reflectivity measurements given in (2). The observation equation specifies that the log bias equals the log of the ratio of mean gage observations to mean radar observations (at gage locations) plus a normally distributed error term. Variance of the error term is specified as a power law function of the number of raingages reporting measurable rainfall.

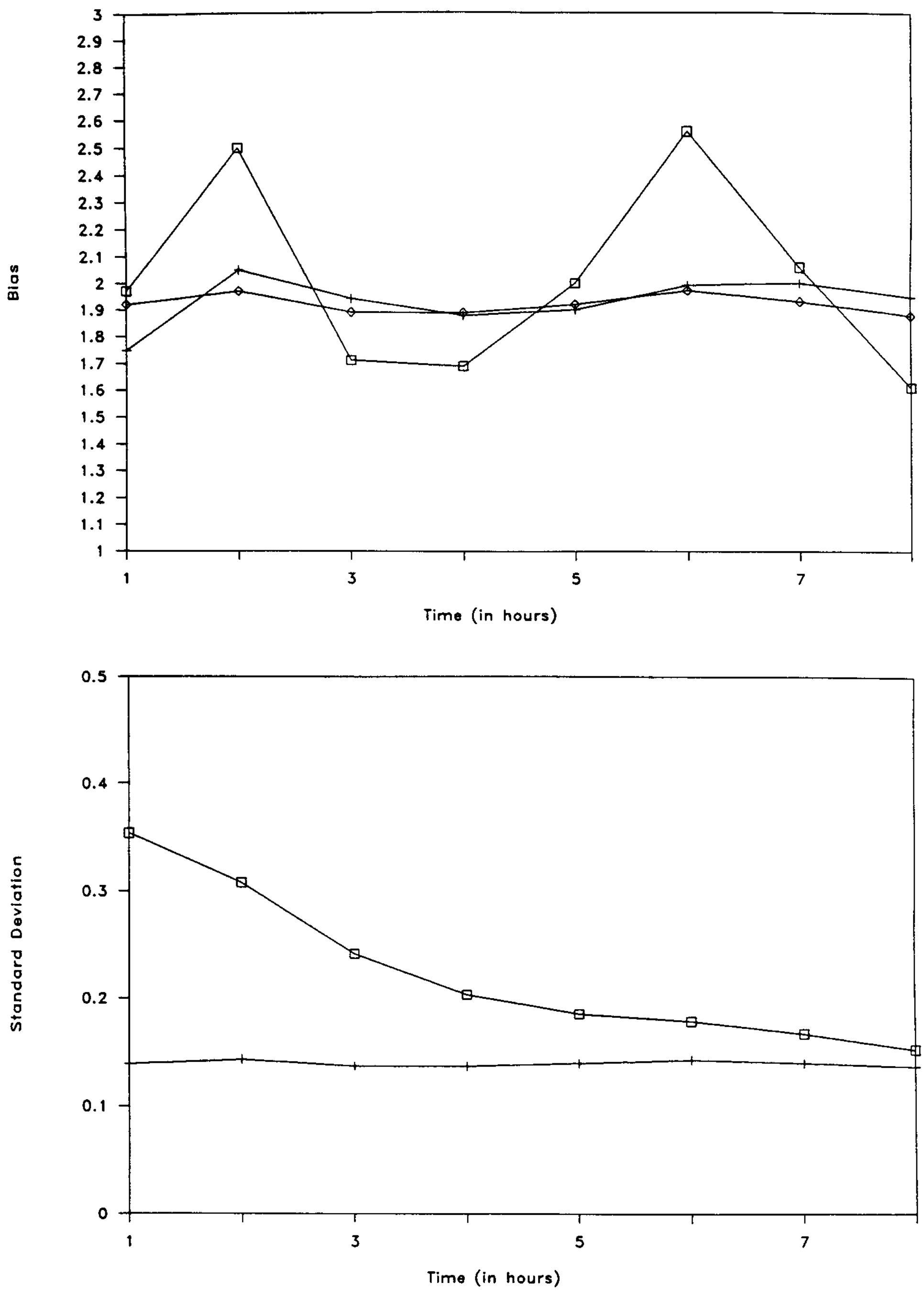

FIG. 5. (a) Bias estimates for the 8 h 27 May data case using parameter values a_1 = 1.0, a_2 = 0.1, a_3 = 1.0, and a_4 = −1.0. Sample bias values are denoted by boxes. Filter estimates are denoted by crosses. Smoothing estimates are denoted by diamonds. (b) Estimated standard deviation of filter estimators (denoted by boxes) and smoothing estimators (denoted by crosses).

Based on the model of mean field bias and the observation equation, state estimators are derived in section 3 for the mean field bias. State estimators are derived for three situations: 1) estimating the bias for a given hour from observations up to and including the hour; this is a filtering problem, 2) estimating the bias for a given hour from observations preceding, during, and following the hour; this is a smoothing problem, and 3) estimating the bias for a given hour from observations preceding the hour; this is a prediction

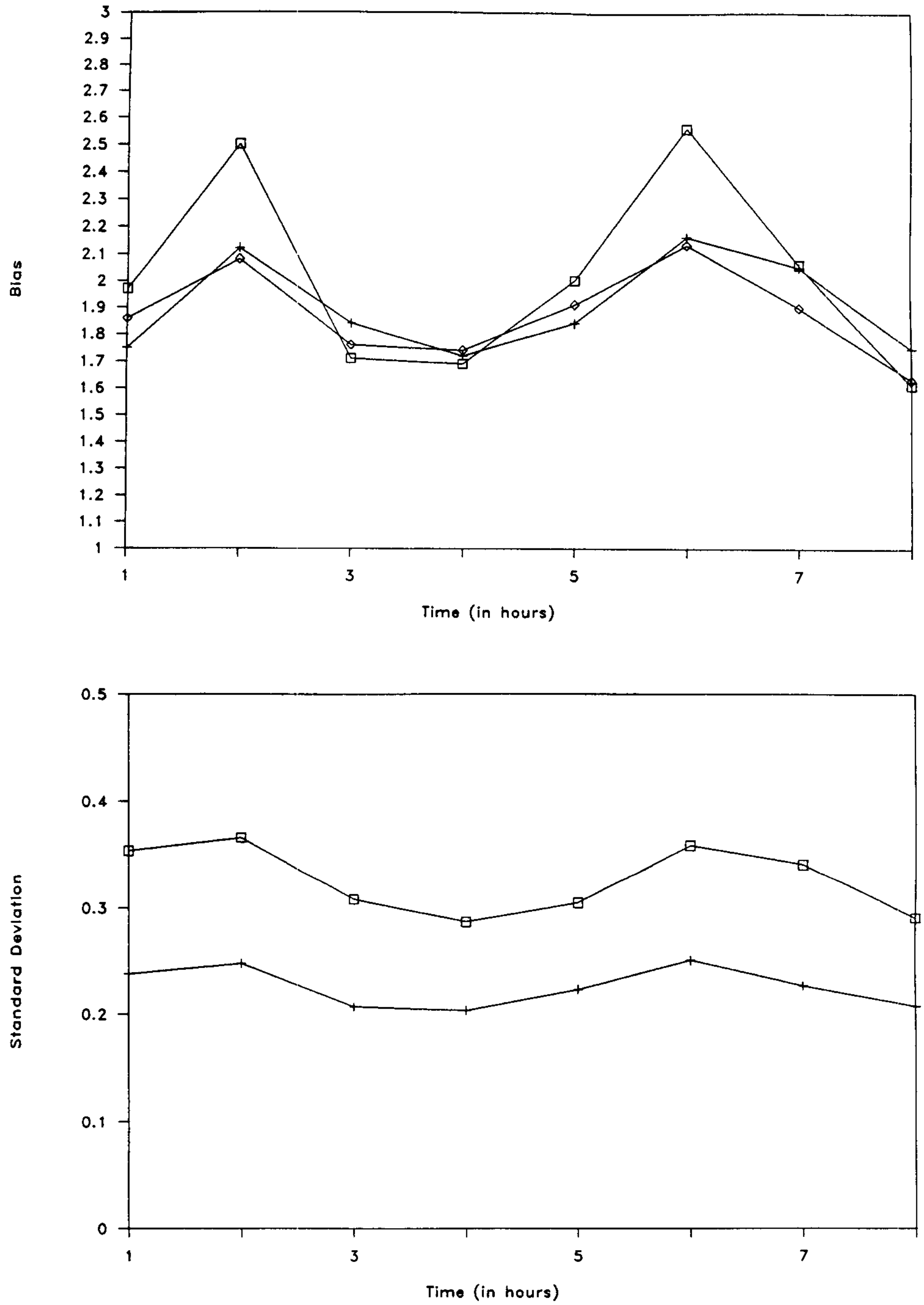

FIG. 6. Same as Fig. 5, but with $a_1 = 0.9$, $a_2 = 0.1$, $a_3 = 1.0$, and $a_4 = -1.0$.

problem. In each case the state estimators are obtained by a sequential procedure which has at its basis the Kalman filter. Each problem has application to the NEXRAD precipitation processing procedures.

Likelihood-based inference procedures are developed for model parameters in section 4. The major result is the representation of the log-likelihood function for the model given in (38). Maximum likelihood estimators are obtained from the likelihood function by a gradient search algorithm. Hypothesis testing procedures are developed to assess structural properties of the bias model and observation equation. In particular a test is developed for distinguishing whether the bias varies over the course of a storm versus an alternative

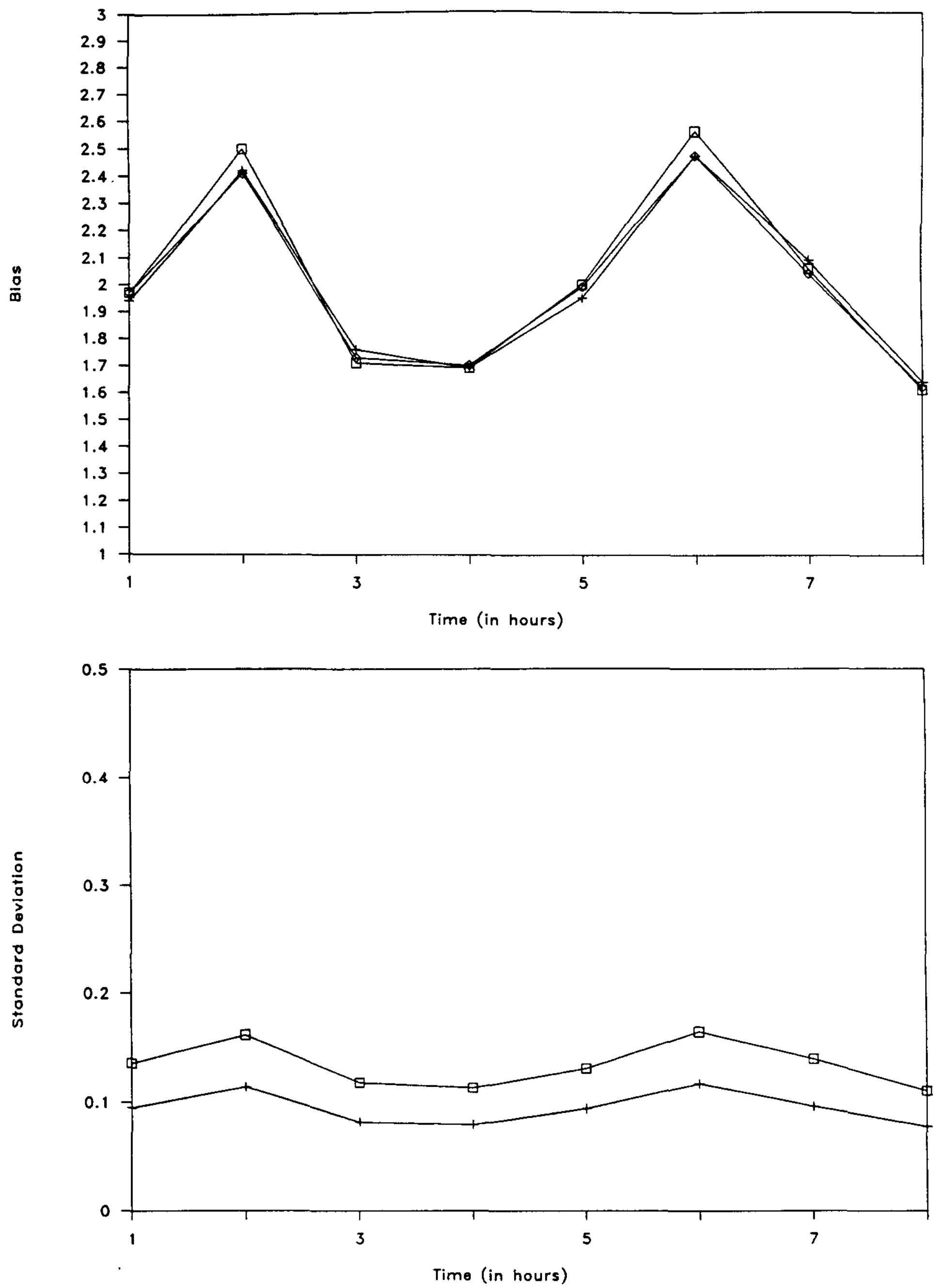

FIG. 7. Same as Fig. 5, but with $a_1 = 0.9$, $a_2 = 0.1$, $a_3 = 0.1$, and $a_4 = -1.0$.

in which the only variability in the bias process is from storm to storm. Simulation results illustrate that convergence of the parameter estimators can be relatively fast, with data from approximately 25 storms providing reasonably accurate parameter estimates.

The state estimation procedures are applied in section 5 to a sample data case from the NSSL radar at Norman, Oklahoma. An attractive feature of the procedure developed in this paper for off-site processing is the flexibility offered by the smoothing procedure. The state estimation results underscore the need for efficient parameter estimation procedures.

In conclusion it should be emphasized that obtaining high-quality radar rainfall products depends on quality

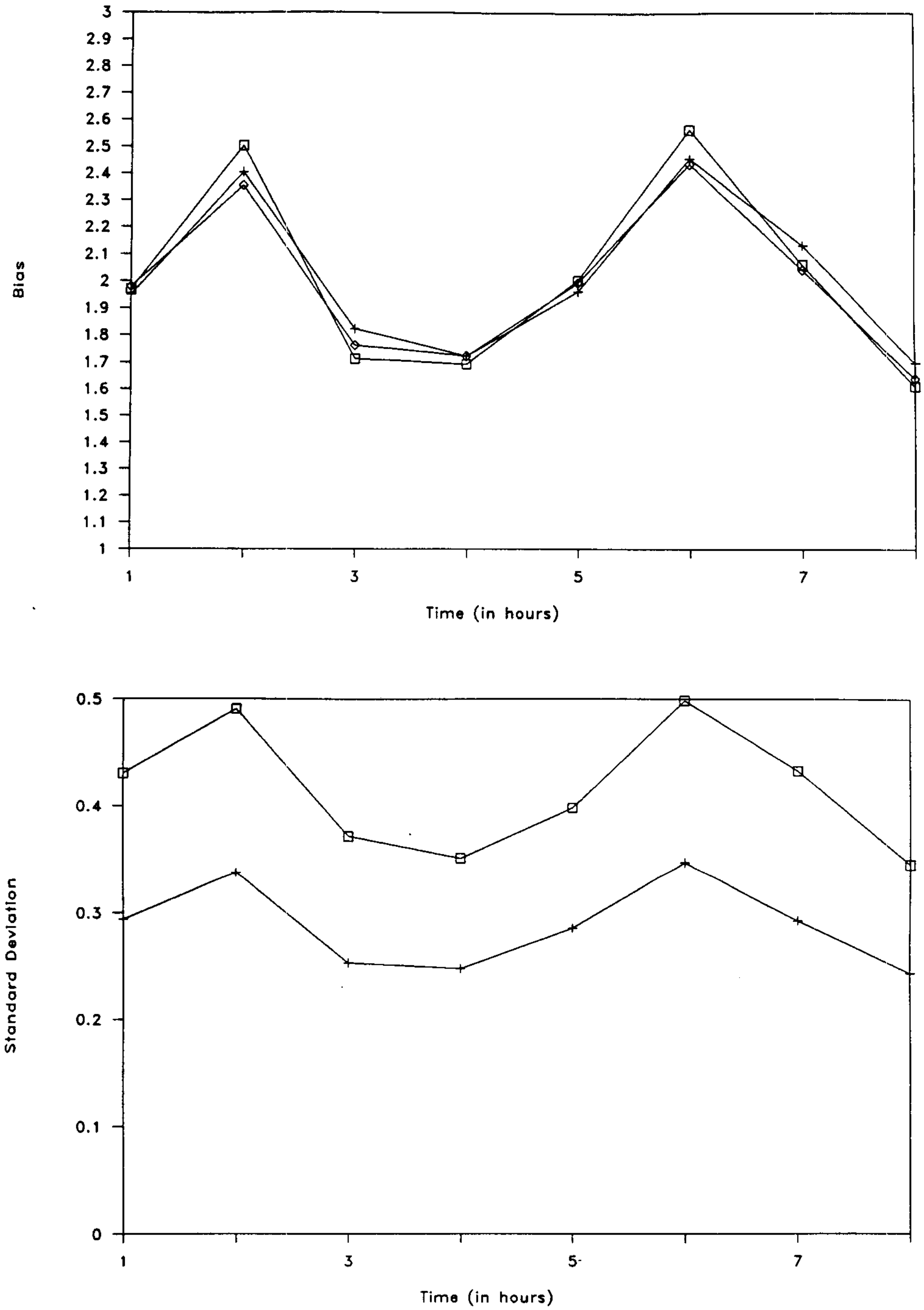

FIG. 8. Same as Fig. 5, but with $a_1 = 0.9$, $a_2 = 1.0$, $a_3 = 1.0$, and $a_4 = -1.0$.

of the radar signal and on development of an integrated processing system with a strong focus on quality control (as discussed by Hudlow et al. 1983). The procedures developed in this paper for bias adjustment are but one piece of an integrated processing system. For the NEXRAD system bias adjustment is a particularly important component because the processing system is fully automated and because raingage data are the only external hydrometeorological data that will be imported to the NEXRAD system.

Acknowledgments. The work of J. Smith was carried out at the Hydrologic Research Laboratory (HRL) of the National Weather Service (NWS). The work of W.

Krajewski was supported by NWS under Cooperative Agreement NA86-AA-H-HY126 between HRL and the Iowa Institute of Hydraulic Research (IIHR). The numerical computations were performed using a minisupercomputer Apollo DN10000 funded by the National Science Foundation (Grant BCS 8906210) and IIHR.

APPENDIX

Derivation of the Likelihood Function

The derivation of the representation for the likelihood function in (38) of section 4 is given below.

Proof:

$$L_n(a) = \ln p[Y_1(1), \cdots, Y_1(T_1); , \cdots, ; Y_n(1), \cdots, Y_n(T_n)] \quad (A1)$$

$$= \ln \prod_{i=1}^{n} p[Y_i(1), \cdots, Y_i(T_i)] \quad (A2)$$

$$= \ln \prod_{i=1}^{n} \left\{ \prod_{s=2}^{T_i} p[Y_i(s) | Y_i(s-1), \cdots, Y_i(1)] \right\} \times p[Y_i(1)] \quad (A3)$$

$$= \sum_{i=1}^{n} \left\{ \sum_{s=2}^{T_i} \ln p[Y_i(s) | Y_i(s-1), \cdots, Y_i(1)] \right\} + \ln[p(Y_i(1)]. \quad (A4)$$

From (10)–(12) we have

$$Y_i(1) = \beta_i(1) + \mathrm{M}_i(1) \quad (A5)$$

implying that

$$Y_i(1) \sim N[0, a_2 + a_3\eta_i(1)^{a_4}]. \quad (A6)$$

It follows from Eqs. (10)–(12) that for s greater than 1 the conditional distribution of $Y_i(s)$ given $Y_i(s-1)$, $\ldots$, $Y_i(1)$ is Gaussian. The moments can be calculated as follows:

$$E[Y_i(s) | Y_i(s-1), \cdots, Y_i(1)]$$

$$= E[\beta_i(s) + M_i(s) | Y_i(s-1), \cdots, Y_i(1)] \quad (A7)$$

$$= \hat{\beta}_i(s). \quad (A8)$$

Similarly,

$$\mathrm{var}[Y_i(s) | Y_i(s-1), \cdots, Y_i(1)]$$

$$= E\langle \{Y_i(s) - E[Y_i(s) | Y_i(s-1), \cdots, Y_i(1)\}^2 | Y_i(s-1), \cdots, Y_i(1)\rangle \quad (A9)$$

$$= E\{[\beta_i(s) + M_i(s) - a_1\hat{\beta}_i(s-1)]^2 | Y_i(s-1), \cdots, Y_i(1)\} \quad (A10)$$

$$= E\{[a_1\beta_i(s-1) + W_i(s) + M_i(s) - a_1\hat{\beta}_i(s-1)]^2 | Y_i(s-1), \cdots, Y_i(1)\} \quad (A11)$$

$$= E\langle \{[a_1\beta_i(s-1) - a_1\hat{\beta}_i(s-1)] + W_i(s) + M_i(s)\}^2 | Y_i(s-1), \cdots, Y_i(1)\rangle \quad (A12)$$

$$= a_1^2 \textstyle\sum_i(s-1) + a_3\eta_i(s)^{a_4} + a_2(1 - a_2^2). \quad (A13)$$

It follows that,

$$[Y_i(s) | Y_i(s-1), \ldots, Y_i(1)] \sim N[a_1\hat{\beta}_i(s-1), a_1^2 \textstyle\sum_i(s-1) + a_3\eta_i(s)^{a_4} + a_2(1 - a_1^2)]. \quad (A14)$$

Equation (38) follows immediately.

REFERENCES

Ahnert, P., M. Hudlow, E. Johnson and D. Greene, 1983: Proposed "on-site" precipitation processing system for NEXRAD. Preprint, *21st Radar Meteorology Conference,* Boston, Amer. Meteor. Soc., 19–23.

——, W. Krajewski and E. Johnson, 1985: Kalman filter estimation of radar-rainfall field bias. Preprint *23d Radar Meteorology Conference,* Boston, Amer. Meteor. Soc.

Anderson, T. W., 1985: *An Introduction to Multivariate Statistical Analysis.* John Wiley and Sons Inc., 373 pp.

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

Battan, L. J., 1973: *Radar Observation of the Atmosphere.* University of Chicago Press,

Bickel, P., and K. Doksum, 1977: *Mathematical Statistics.* Holden Day, Inc., 492 pp.

Cain, D. E., and P. L. Smith, Jr., 1976: Operational adjustment of radar estimated rainfall with gage data: A statistical evaluation. Preprints, *17th Conference on Hydrometeorology,* Boston, Amer. Meteor. Soc., 54–59.

Carbonne, R. E., and L. D. Nelson, 1978: The evolution of raindrop spectra in warm-based convective storms as observed and numerically modeled. *J. Atmos. Sci.,* **35,** 2302–2314.

Cataneo, R., and G. E. Stout, 1968: Raindrop-size distributions in humid continental climates and associated rainfall-rate radar reflectivity relationships. *J. Appl. Meteor.* **7,** 901–907.

Chappell, C. F., 1986: Quasi-stationary convective events. *Mesoscale Meteorology and Forecasting,* P. S. Ray, Ed., Amer. Meteor. Soc., 289–310.

Collier, C. G., 1986: Accuracy of rainfall estimates by radar. Part I: Calibration by telemetering raingages. *J. Hydrol.,* **93,** 207–223.

——, P. R. Larke and B. R. May, 1983: A weather radar correction procedure for real-time estimation of surface rainfall. *Quart. J. Roy Meteor. Soc.,* **109,** 589–608.

Doviak, R. J., and D. S. Zrnić, 1984: *Doppler Radar and Weather Observations.* Academic Press, 458 pp.

Gelb, A., 1973: *Applied Optimal Estimation.* MIT Press, 374.

Harrison P. J., and C. F. Stevens, 1971: A Bayesian approach to short-term forecasting. *Operations Research Quarterly,* **22,** 341–362.

——, and ——, 1976: Bayesian forecasting. *J. Roy. Stat. Soc.,* **38,** 205–247.

Hudlow, M. D., D. Greene, P. Ahnert, W. Krajewski, T. Sivaramakrishnan, M. Dias and E. Johnson, 1983: Proposed off-site precipitation processing system for NEXRAD. Preprint, *21st Radar Meteorology Conference,* Boston, Amer. Meteor. Soc., 394–403.

——, J. A. Smith, M. L. Walton and R. C. Shedd, 1989: NEXRAD—New Era in hydrometeorology in the United States. *Proc. In-*

ternational Symposium on Hydrological Applications of Weather Radar, Salford, England, 12 pp.

Johansen, S., 1984: *Functional Relations, Random Coefficients, and Nonlinear Regression.* Springer Verlag,

Joss, J., and A. Waldvogel, 1989: Precipitation measurement and hydrology. *Radar Meteorology,* D. Atlas, Ed., Boston, Amer. Meteor. Soc., 577–606.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Engineering,* **82D,** 35–45.

Karr, A. F., 1986: *Point Processes and their Statistical Inference.* Marcel Dekker,

Kedem, B., and L. Chiu, 1987: On the lognormality of rainrate. *Proc. Natl. Acad. Sci.,* **84,** 901–905.

Krajewski, W. F., 1987: Cokriging of radar and raingage data. J. Geophys. Res., **92,** 9571–9580.

Maddox, R. A., C. F. Chappell and L. R. Hoxit, 1979: Synoptic and meso-α-scale aspects of flash flood events. *Bull. Amer. Meteor. Soc.,* **60,** 115–123.

Press, W., B. Flannery, S. Teukolsky and W. Vetterling, 1988: *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, 818 pp.

Rodriguez-Iturbe, I., and J. Mejia, 1974: The design of rainfall networks in time and space. *Water Resources Research,* **10,** 713–728.

Seo, D.-J., W. F. Krajewski and D. S. Bowles, 1989: Stochastic interpolation of rainfall data from raingages and radar using cokriging. *Water Resources Research,* **26,** 469–478.

——, and J. A. Smith, 1990: Rainfall estimation using raingages and radar—a Bayesian approach. *J.Stochastic Hydrol. and Hydraulics.*

Shedd, R., J. A. Smith and M. L. Walton, 1989: The sectorized hybrid scan strategy for NEXRAD. *Proc. International Symposium on Hydrological Applications of Weather Radar,* Salford, England, 8 pp.

Smith, C. J., 1986: The reduction of errors caused by bright bands in quantitative rainfall measurements made using radar. *J. Atmos. Oceanic Technol.,* **3,** 129–141.

Smith, J. A., 1987: Statistical modeling of daily rainfall occurrences. *Water Resources Research,* **24,** 885–893.

——, and A. F. Karr, 1985: Statistical inference for point process models of rainfall. *Water Resources Research,* **21,** 73–79.

——, R. Shedd and M. Walton, 1989: Parameter estimation for the NEXRAD hydrology sequence. Preprints, *24th Radar Meteorology Conference,* Amer. Meteor. Soc., Boston, 259–263.

Smith, P. L., and D. E. Cain, 1983: Use of sequential analysis methods in adjusting radar rainfall estimates on the basis of raingage data. Report 83-2, Institute of Atmospheric Sciences, South Dakota School of Mines.

Waldvogel, A. N., 1974: The N_0 jump of raindrop spectra. *J. Atmos. Sci.,* **31,** 1067–1978.

Weiss, G., 1975: Time-reversibility of linear stochastic processes. *J. Appl. Probability,* **12,** 831–836.

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