On the Threshold Method for Rainfall Estimation: Choosing the Optimal Threshold Level
Author(s): Benjamin Kedem and Harry Pavlopoulos
Source: *Journal of the American Statistical Association*, Vol. 86, No. 415 (Sep., 1991), pp. 626-633
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2290391
Accessed: 07-11-2019 19:56 UTC

On the Threshold Method for Rainfall Estimation: Choosing the Optimal Threshold Level

BENJAMIN KEDEM and HARRY PAVLOPOULOS*

Real data experiments show that, for large areas, the area average rain rate and the fraction of the area covered by rain rate above a fixed threshold are highly correlated. For the right choice of the threshold level the correlation can easily exceed 95% and may even reach 99%. This remarkable fact observed in nature is the basis for the so-called threshold method for measuring rainfall from space via satellite. The method depends critically on the threshold level, showing significant improvement in performance for optimal and nearly optimal thresholds. Under the assumption that the continuous part of the distribution of rain rate is lognormal, two theoretically derived optimal levels agree closely with the best level obtained empirically from a well-known data set of rain rate.

KEY WORDS: GATE; Information matrix; Lognormal; Mixed distribution; Rain rate; Space–time observations.

1. INTRODUCTION

The purpose of this article is to present a statistical solution to a problem of scientific interest in the field of meteorology. The problem is to choose an optimal threshold level needed for the so-called threshold method for measuring rainfall from space via satellite. The threshold method calls for the estimation of the instantaneous area average rain rate from the fraction of the area above which the instantaneous rain rate exceeds a fixed threshold. We suggest two criteria for deriving optimal thresholds. In the first, the optimal level minimizes a certain variance. In the second, the optimal level maximizes a certain correlation. The two schemes provide reasonably close threshold levels across a wide range of rain types.

By *rain rate* we mean *instantaneous* rain intensity measured in mm/hr at points in space. By a *snapshot* we mean the map of rain rate over a given region at a fixed instant. The *area* refers to the area of a given region. From now on we shall refer to the area average rain rate and the fraction of the area covered by rain rate exceeding a given threshold as the *area average* and *fractional area,* respectively. In what follows we assume that rain rate possesses all moments, and thus we exclude "fat-tailed" distributions, conditional on rain, from our discussion. The threshold level is denoted by τ; clearly, $\tau \geq 0$.

It has recently been observed (Chiu 1988; Doneaud, Niscov, Priegnitz, and Smith 1984; Rosenfeld, Atlas, and Short 1990; Short, Wolff, Rosenfeld, and Atlas 1989) that the area average and fractional area are highly correlated. This observation has been confirmed from quite a few data sets from different parts of the world by showing that the sample correlation can easily exceed 95% and in some cases can even exceed 99% (Short et al. 1989). In particular, for the well-known and widely used data set referred to as GATE, to be described later, it has been found empirically by Chiu (1988) that the threshold level $\tau = 5$ leads to a squared correlation of 98% and drops dramatically for other levels. For level $\tau = 0$ the squared correlation drops to below 80%. The drop in squared correlation from a maximum of 98% is shown in Figure 1.

We shall provide, under some assumptions inspired by the GATE data set, a theoretical explanation for the empiricism observed in Chiu (1988). In particular, by means of our optimality criteria, we shall show why threshold levels around $\tau = 5$ mm/hr are optimal for GATE-like rain.

1.1 The Need for the Threshold Method

The National Aeronautics and Space Administration (NASA) is contemplating at present the Tropical Rainfall Measuring Mission (TRMM), a satellite program for measuring rain rate from space over the tropics and subtropics. TRMM will provide monthly and seasonal rainfall estimates averaged over areas of about 10^5 km^2. The idea is to equip a satellite at a low altitude orbit with a precipitation radar, the first ever in space, and passive microwave radiometers operating at several frequencies. Rain rate is to be recovered from radar reflectivity and microwave temperature. The measurements are needed for a better understanding of the variability of the Earth's climate and a description of the global hydrological cycle. An extensive description of TRMM and its scientific goals is given in Simpson 1988 and in Simpson, Adler, and North 1988.

From a statistical point of view, the mission poses great challenge, because rain rate will be measured indirectly through covariates whose precise relationship to rain rate is not entirely clear. For example, microwave temperature is related to rain rate nonlinearly and the relationship itself is empirical (Wilheit 1986). Thus converting microwave temperature to rain rate is quite tricky. Likewise, the relationship between rain rate and radar reflectivity, commonly referred to as the Z–R relationship, is also empirical and is given by the model $Z = aR^b$, where Z stands for radar reflectivity and R denotes rain rate (Battan 1973, p. 89). The parameters a, b are estimated from data by regression methods. Thus, using the Z–R relationship, which gives *point by point* conversion, to obtain the area average rain rate, may lead to gross errors if the "noise" is such that it does

* Benjamin Kedem is Professor and Harry Pavlopoulos is Graduate Student, both in the Department of Mathematics, University of Maryland, College Park, MD 20742. This research was supported in part by the National Aeronautics and Space Administration and the Office of Naval Research. Part of the work was done while Benjamin Kedem was a visitor at the Institute for Physical Science and Technology, University of Maryland, College Park. The authors thank the referees for helpful comments.

Journal of the American Statistical Association
September 1991, Vol. 86, No. 415, Applications and Case Studies

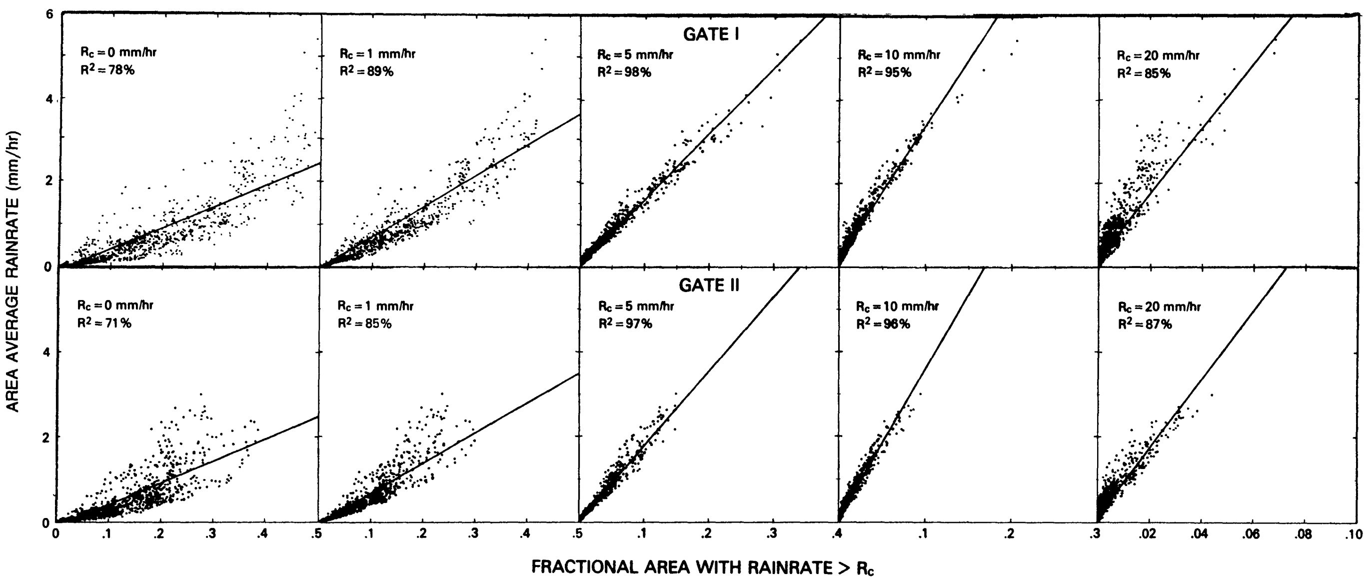

Figure 1. Scattergram and the Corresponding Linear Regression of the Area Average Rain Rate on the Fractional Area. $R_c = \tau$. Source: Chiu (1988).

not average out. Another problem is the inability to tell instrument noise apart from very low rain rates that are close to 0. Adding to these instrument problems the fact that tropical rainfall is highly variable as well as highly capricious (its two most characteristic features), one can have a sense of the difficulties and challenge encountered in measuring rain rate from space.

The threshold method could alleviate some of these problems because all that is required of the instrument is the ability to distinguish correctly between rain rate values that are either above or below a given threshold, not precise point measurements. This, combined with sampling designs that guarantee the validity of the law of large numbers in space (i.e., iid observations), could lead to reasonably accurate estimates of the area mean rain rate. The connection with the law of large numbers will become clearer when we expound the observed linear relationships in Figure 1.

It is quite clear that the threshold level, on whose choice depends the method's reliability, will require periodic revision according to geographical location and month of the year. Intuitively, the threshold level should be linked to the mean rain rate, conditional on rain. As we shall see, this intuition can in fact be justified rigorously under some assumptions.

1.2 The GATE Data Set

GATE stands for Global Atmospheric Research Program, Atlantic Tropical Experiment. It is a field study conducted in 1974 in the eastern Atlantic off the coast of west Africa to study and collect rainfall data. Five ships equipped with precipitation radars that covered an area of about 400 km in diameter collected rainfall data in the form of huge snapshots of radar reflectivity taken every 15 minutes. The ships were also equipped with raingages, needed for calibration, and other relevant atmospheric probes (Simpson 1988, p. 37). The resulting data set is known as the GATE data. Actually, there were several phases in the GATE study and the data that generated Figure 1 come from the first two phases, commonly referred to as GATE–I and GATE–II. GATE–I and GATE–II also stand for the corresponding data sets, consisting of 1,716 (18 days) and 1,512 (15 days) snapshots, respectively. The radar reflectivity data were converted into rain rate data binned into 4×4 km^2 pixels, so what we call a point in space is actually a 4×4 km^2 pixel. For technical details see Arkel and Hudlow (1977) and Patterson, Hudlow, Pytlowany, Richards, and Hoff (1979).

1.2.1 On the Lognormality and Homogeneity of Rain Rate. Experiments with real data point to the curious fact that many convective cloud variables, such as rain rate (conditional on rain), cloud height, duration of a storm, cloud extent, and more, seem to follow the lognormal distribution (Biondini 1976; Lopez 1977). In Kedem and Chiu (1987) the lognormal distribution of rain rate is explained with the help of a stochastic regression model.

The GATE data set has been the source for numerous studies, some of which are summarized in Simpson (1988). What is relevant to the present work is that the lognormal distribution $\Lambda(\mu, \sigma)$ (truncated at 1 mm/hr to overcome the problem of setting apart very low rain rates from instrument noise) provides a close fit to the data from GATE–I and GATE–II, conditional on rain. This was shown in Kedem, Chiu, and North (1990) using different sampling designs in time and space. In both phases of GATE the different time–space designs invariably yielded as estimate for (μ, σ) values close to $(1, 1)$. Thus the lognormal distribution $\Lambda(1, 1)$ is a reasonable model for GATE-like rain rate, conditional on rain. This fact is instrumental in explaining the optimality of the threshold level $\tau = 5$ in Figure 1. We shall now proceed to describe briefly the sampling designs and statistical results of Kedem, Chiu, and North (1990).

A time–space sampling design is characterized by the triple (n, k, l). The first index n refers to sampling frequency in time and is always a multiplier of 15 minutes. The k and

l refer to spatial sampling in east-west and north-south directions, respectively; they are always multipliers of 4 km. For example, the design (1, 10, 10) denotes sampling "instantaneous 4 × 4 pixels" every 15 minutes in time but separated by 40 km in both east-west and north-south directions. Given a specific design, only data values that exceeded 1 mm/hr were used in the estimation of (μ, σ). The results of minimum chi-squared estimation are shown in Table 1. It is seen that, regardless of the design, the values of the estimates $\hat{\mu}$, $\hat{\sigma}$ are remarkably close. This indicates that, conditional on rain, the rain rates are distributed more or less evenly in space and time. This is the basis for our homogeneity assumption (Sec. 2.2).

2. EXPLANATION OF THE LINEAR RELATIONSHIP

2.1 An Outline of Requirements

There are several ways to explain the observed linear relationship between the area average and the fractional area that can lead to a sensible choice of the threshold. Any approach, however, should take into account the following facts.

1. Experimental results point to a high degree of linearity between the area average and the fractional area.
2. The distribution of rain rate is mixed, with an atom at 0.
3. The continuous part of the distribution is markedly skewed.
4. From Figure 1, the data in the upper right portion of the scatterplots are more consistent with heteroscedasticity than with homoscedasticity.
5. Table 1 points to a reasonable degree of homogeneity in time and space of positive rain rates.
6. From Figure 1, the correlation between the area average and the fractional area is a function of the threshold level.
7. Rain rate data are collected in time and space.

Table 1. Minimum χ^2 Estimates From Different Designs

GATE	Design	$\hat{\mu}$	$\hat{\sigma}$
I	(2, 4, 4)	1.137	1.043
	(2, 8, 8)	1.157	1.031
	(4, 4, 4)	1.129	1.048
	(4, 8, 8)	1.140	1.039
	(5, 20, 20)	1.185	1.033
	(6, 6, 6)	1.152	1.028
	(6, 8, 8)	1.169	1.042
	(8, 4, 4)	1.126	1.045
	(8, 6, 6)	1.182	1.030
	(10, 8, 8)	1.162	1.029
	(10, 10, 10)	1.085	1.084
	(20, 10, 10)	1.095	1.035
	(24, 1, 1)	1.159	1.028
	(48, 1, 1)	1.255	1.009
II	(4, 4, 4)	1.065	1.100
	(3, 10, 10)	1.032	1.100
	(5, 3, 3)	1.099	1.077
	(5, 5, 5)	1.056	1.091
	(5, 10, 10)	1.031	1.089
	(8, 8, 8)	1.046	1.053
	(10, 5, 5)	1.043	1.098
	(10, 10, 10)	.960	1.161
	(20, 3, 3)	1.050	1.098
	(20, 5, 5)	.998	1.121
	(20, 10, 10)	.918	1.180
	(30, 5, 5)	.976	1.123
	(30, 10, 10)	.982	1.214
	(48, 1, 1)	1.041	1.028

Source: Kedem, Chiu, and North (1990).

Fitting regression models, including nonlinear and log–log models, as well as modeling by time series models that may include seasonal trends, are all viable possible approaches that can take into account some or even all of these facts. It is not entirely clear, however, how these approaches take into consideration the mixed distribution aspect of rain rate. We believe that the mixed distribution aspect of rain rate is of fundamental importance in rainfall modeling and that it should be addressed clearly, whatever the approach may be. Evidently, the formulation that we adopt in what follows is quite general and does address the foregoing facts, including the mixed distribution aspect of rain rate. In fact, this is our starting point. By this, however, we do not exclude from future consideration appropriate alternative formulations. It would certainly be of interest to see what other approaches produce.

2.2 A Statistical Setup

The ensuing discussion follows Kedem, Chiu, and North (1990) and Kedem and Short (1989). It is helpful to keep in mind that what we are going to describe next amounts to looking at a box and slices of the box. The box refers to points in space and time; a slice refers to points across an instantaneous snapshot for fixed time. It is convenient to think of the points as points on a continuum. We think of space as being horizontal and of time as being vertical. The interplay between the box and its slices enables us to shift from observations in space and time to observations across slices or instantaneous snapshots. Our interpretation bypasses issues of temporal variability and correlation by taking random samples in time and space and from a snapshot by fixing time.

Suppose that rain rate is observed over a given region A and throughout a specific period $[0, T]$. Then, each space–time point $\omega \equiv (\mathbf{a}, t)$, $\mathbf{a} \in A$, $t \in [0, T]$ is endowed with an instantaneous rain rate measurement. The set of all of these endowed ω's constitutes the sample space $\Omega \equiv A \times [0, T]$ of interest. Let $X(\omega)$ be the random variable that gives the value of rain rate associated with ω, so $X : A \times [0, T] \rightarrow [0, \infty)$. Then, X has a *mixed* distribution because it admits the value 0 (no rain) with positive probability, say, $1 - p$. That is, $\Pr(X = 0) = 1 - p$. But otherwise, that is, conditional on rain, X admits values in a continuum and, therefore, has a continuous distribution. Let $G(x)$ denote the distribution function of X. Then $G(x)$ is of a *mixed type* that has the general form

$$G(x) = (1 - p)H(x) + pF(x), \tag{1}$$

where $H(x)$ is a step function defined by

$$\begin{aligned} H(x) &= 0 \quad \text{if } x < 0 \\ &= 1 \quad \text{if } x \geq 0 \end{aligned}$$

and $F(x)$ is a continuous distribution function with positive density $f(x) = F'(x)$ $(x > 0)$. In other words, $F(x)$ is the conditional distribution function of X, given that $X > 0$. We refer to $F(x)$ or, equivalently, $f(x)$ as the *continuous* part of the distribution of X and to $H(x)$ as the *discrete* part. The weight given to each part is determined by p, the probability of rain.

The following elementary argument establishes linear relationships between the expected value of X and the expected value of any integrable function of X. These linear relationships are fundamental for the explanation that we are seeking.

If $\varphi(X)$ is an arbitrary integrable function of X, then its expected value is $E[\varphi(X)] = (1 - p)\varphi(0) + pE[\varphi(X) \mid X > 0]$. In particular, if $\varphi(X) = X$, then $E[X] = pE[X \mid X > 0]$. By equating p from these two equations we obtain the following linear relationship between $E[X]$ and $E[\varphi(X)]$:

$$E(X) = \beta_\varphi\{E[\varphi(X)] - \varphi(0)\}, \tag{2}$$

where the slope β_φ depends only on the continuous part of the distribution of X and $\varphi(0)$ and is given by

$$\beta_\varphi = \frac{E[X \mid X > 0]}{E[\varphi(X) \mid X > 0] - \varphi(0)}. \tag{3}$$

Besides the fact that the slope β_φ is independent of the probability of rain p and depends only on the continuous part of the distribution, characterized by f, we see that, if $\varphi(0) = 0$, then the linear relationship does not contain an intercept term.

Now fix t, and let $X_t(\mathbf{a})$ be the random variable that gives the value of rain rate at the point $\mathbf{a}$ in space, at time t, so that $X_t : A \rightarrow [0, \infty)$. The distribution of X_t is, once again, mixed and is governed by p_t and $f_t(x)$, which play the same role as p and $f(x)$ for X. Clearly, p_t and $f_t(x)$ need not equal p and $f(x)$, respectively. Therefore, the ratio β_φ corresponding to X_t depends on f_t only and, in light of (3), is given by the equation

$$\beta_\varphi(t) = \frac{E[X_t \mid X_t > 0]}{E[\varphi(X_t) \mid X_t > 0] - \varphi(0)}, \tag{4}$$

where now $\beta_\varphi(t)$ depends on t. We shall make the simplifying homogeneity assumption that $\beta_\varphi(t)$ is independent of t. Clearly, the estimation results in Table 1 are consistent with such an assumption and render it more palatable. The precise assumption is as follows.

Homogeneity Assumption. The continuous part of the distribution of X is homogeneous in time and space: $f_t \equiv f$, for all $t \in [0, T]$.

No similar assumption is made on p_t. It follows that $\beta_\varphi(t) = \beta_\varphi$.

The following idealization assumption is instrumental for the understanding of the rest of the article. Each instantaneous pixel is considered a point in time and space (a point in $A \times [0, T]$), and the population of these space–time pixels ("points") is assumed to be infinite. The peculiarity of our space–time structure allows sampling from (p, f) or, by fixing t, from (p_t, f). By the first of these two schemes, we can take a random sample of pixels in space and time and consider the sample values as iid observations from the mixed distribution (p, f). By the second scheme, for every fixed time t, we can take a random sample of pixels from an instantaneous snapshot and consider the sample values as iid observations from the mixed distribution (p_t, f). At a later stage we also assume that those pixels for which positive rain rate has been measured, by any of these two schemes, are independent observations from a lognormal distribution with density f whose parameters are fixed and, in particular, do not depend on t.

Now, let $\langle X_t \rangle$ and $\langle \varphi(X_t) \rangle$ be the sample averages of X_t and $\varphi(X_t)$, respectively, obtained from a random sample over an instantaneous snapshot at time t. $\langle X_t \rangle$ and $\langle \varphi(X_t) \rangle$ are area averages. We define $\langle X_t \rangle$ as the area average of rain rate.

The area averages approximate the corresponding expected values as follows. Fix t, and take a random sample over an instantaneous snapshot at time t. By the law of large numbers, as the sample size increases, $\langle X_t \rangle \rightarrow E(X_t)$ and $\langle \varphi(X_t) \rangle \rightarrow E[\varphi(X_t)]$ with probability 1. Furthermore, under the homogeneity assumption together with (2), for a sufficiently large sample we have the approximation $\langle X_t \rangle \simeq$ constant $+\ \beta_\varphi \langle \varphi(X_t) \rangle$ for each fixed t, where the slope β_φ is given by (3) and is independent of t.

It should be noted that, in principle, once a snapshot has been observed in its entirety over the area, we could compute $E(X_t)$ and $E[\varphi(X_t)]$ exactly. In practice, however, only a fraction of the snapshot is available. For example, the TRMM satellite will not be able to "see" the entire area. In reality it will only give information about some pixels each time it flies over an area. See Simpson (1988). In our idealized construction the observed pixels are treated as a random sample from the snapshot.

A special case is provided by the fractional area. Let $I[X_t > \tau]$ be the indicator function of the event $X_t > \tau$ and observe that $\langle I[X_t > \tau] \rangle$ is an area average. We define $\langle I[X_t > \tau] \rangle$ as the fractional area. We shall return to the fractional area momentarily.

We can now specialize the preceding argument in explaining the linear relationship observed in Figure 1. Fix $\tau > 0$, and define

$$\begin{aligned} \varphi(x) &= 1 \quad \text{if } x > \tau \\ &= 0 \quad \text{if } x \leq \tau. \end{aligned} \tag{5}$$

Then, $\varphi(0) = 0$, $E[\varphi(X_t)] = \Pr(X_t > \tau)$, and from (2) and (3) we easily obtain

$$E(X_t) = \beta_\varphi(\tau)\Pr(X_t > \tau), \tag{6}$$

where, under homogeneity

$$\beta_\varphi(\tau) = \frac{E[X \mid X > 0]}{\Pr(X > \tau \mid X > 0)}$$

depends on f and the *threshold level* τ only, but not on the *time* t. Observe now that $E(X_t)$ and $\Pr(X_t > \tau)$ are approximated by $\langle X_t \rangle$ and $\langle I[X_t > \tau] \rangle$, respectively. Therefore, from (6),

$$\langle X_t \rangle \simeq \beta_\varphi(\tau)\, \langle I[X_t > \tau] \rangle, \tag{7}$$

where $\beta_\varphi(\tau)$ is a constant for each given τ.

The linear relationship (7) provides an explanation for the observed linear relationship portrayed in Figure 1. On the basis of (7) one can expect high correlation between the area average and the fractional area, as is clearly manifested in Figure 1 and many other similar figures obtained by Rosenfeld et al. (1990) and Short et al. (1989), employing data from completely different rain patterns as well as geographical locations.

Now, (7) provides the clue we were seeking all along: an optimal level τ can be chosen to minimize the variance of the maximum likelihood estimate of $\beta_\varphi(\tau)$ under an appropriate choice of a parametric model for f. Another possibility is to choose τ that maximizes the correlation between the area average $\langle X_t \rangle$ and the fractional area $\langle I[X_t > \tau] \rangle$.

This scheme is motivated by the analogous situation in periodogram analysis of a sinusoid plus noise (the case of a single frequency). There, the least squares estimates of the sinusoid amplitudes and hence also the sample correlation between the independent and dependent variables are obtained as functions of the frequency. The frequency is then chosen as the one that maximizes the sample correlation between the independent and dependent variables. This is the same as minimizing the residuals sum of squares.

Remark 1. Because $\varphi(0) = 0$, no intercept term is included in (7). From the best case corresponding to $\tau = 5$ in Figure 1, in both GATE's, we can see that, likewise, the estimated intercepts are small. In fact, from Chiu (1988), they are .05 for GATE–I and .01 for GATE–II.

Remark 2. It is interesting to note a certain consequence of our adopted model. Had we included in our homogeneity assumption also the condition $p_t \equiv p$, then our model suggests that (7) should produce a *point* rather than a line. That is, if $p_t = p$ in addition to $f_t = f$, then by our model the points $(\langle X_t \rangle, \langle I[X_t > \tau] \rangle)$ should cluster about the fixed point $(E[X_t], E[\varphi(X_t)])$. Thus, to accommodate the empirical evidence, adhering to the homogeneity assumption about f implies the inhomogeneity of p_t.

3. CHOOSING AN OPTIMAL THRESHOLD LEVEL

We are now at a position to speak intelligently about an optimal choice of the threshold level τ. For this end we suggest two asymptotic procedures that for GATE-like rain produce similar results for all practical purposes.

Because the case of interest now centers around $\varphi(x)$ in (5), the notation of the slope can be simplified by dropping the reference to φ. It is more convenient now to reparameterize the slope by adding the parameter θ from a continuous parametric distribution. Thus the slope is denoted by the more suggestive notation $\beta_\theta(\tau)$.

We shall first derive the maximum likelihood estimator of $\beta_\theta(\tau)$ by assuming that f is the lognormal density $\Lambda(\mu, \sigma)$, with parameter $\theta = (\mu, \sigma)$:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[\frac{-(\log(x) - \mu)^2}{2\sigma^2}\right], \quad x > 0, \quad (8)$$

where $-\infty < \mu < \infty$, $\sigma > 0$. Of course, many different parametric models are just as plausible, including the gamma and inverse Gaussian distributions; all three are markedly skewed, as is believed of the true distribution of rain rate. But we adhere to the lognormal case because we would like to relate our results to GATE-like rain (see Sec. 1.2.1 again).

Fix τ. Let $Y_1, \ldots, Y_n$, $Y_i > 0$ for all i, be a random sample in time and space from $\Lambda(\mu, \sigma)$, that is, sample rain rate conditional on rain. Note the necessary change in notation here. Denote the maximum likelihood estimator (MLE) of $\beta_\theta(\tau)$ by $\beta_{\hat\theta}(\tau)$, and let $v_\theta(\tau)$ be the variance in the asymptotic distribution of $\sqrt{n}(\beta_{\hat\theta}(\tau) - \beta_\theta(\tau))$. We have two procedures for finding an optimal level. Procedure 1 chooses τ so that $v_\theta(\tau)$ becomes minimum. Procedure 2 chooses τ so that $w_\theta(\tau) \equiv v_\theta(\tau)/\beta_\theta^2(\tau)$ becomes minimum. For sufficiently large samples, it can be argued that the level that minimizes $w_\theta(\tau)$ also maximizes the correlation between the area average and fractional area. This will be shown shortly. Procedure 1 is a simplification of Procedure 2. As we shall see, for GATE-like rain the two procedures yield levels that are quite similar.

3.1 Maximum Likelihood Estimation of the Slope

To obtain the MLE $\beta_{\hat\theta}(\tau)$ of $\beta_\theta(\tau)$ from a random sample of positive rain rates from $\Lambda(\mu, \sigma)$, we first note that the $\Lambda(\mu, \sigma)$ model for f yields

$$\beta_\theta(\tau) = \exp(\mu + \sigma^2/2)/\{1 - \Phi((\log \tau - \mu)/\sigma)\}.$$

Denote the MLE of $\theta = (\mu, \sigma)$ by $\hat\theta = (\hat\mu, \hat\sigma)$. Then,

$$\beta_{\hat\theta}(\tau) = \exp(\hat\mu + \hat\sigma^2/2)/\{1 - \Phi((\log \tau - \hat\mu)/\hat\sigma)\},$$

where

$$\hat\mu = n^{-1} \sum_{i=1}^{n} \log Y_i$$

and

$$\hat\sigma = \left\{ n^{-1} \sum_{i=1}^{n} \log^2 Y_i - \left(n^{-1} \sum_{i=1}^{n} \log Y_i \right)^2 \right\}^{1/2}.$$

It follows from the theory of maximum likelihood estimation (Lehmann 1983, p. 429) that, as $n \to \infty$,

$$\sqrt{n}(\hat\theta - \theta) \xrightarrow{L} N(\mathbf{0}, \mathbf{I}^{-1}(\theta)),$$

where $\mathbf{I}^{-1}(\theta)$ is the inverse of the information matrix associated with $\Lambda(\mu, \sigma)$. Here and in other similar expressions, it is convenient to think of θ, $\hat\theta$, $\mathbf{0}$ as two-dimensional column vectors. Since $\beta_\theta(\tau)$ is differentiable at every point $\theta \in (-\infty, \infty) \times (0, \infty)$, with obvious notation we obtain

$$\frac{\partial\beta}{\partial\mu} = \frac{\exp(\mu + (\sigma^2/2))}{[1 - \Phi(u)]^2} \left[1 - \Phi(u) - \frac{1}{\sigma}\phi(u) \right],$$

$$\frac{\partial\beta}{\partial\sigma} = \frac{\exp(\mu + (\sigma^2/2))}{[1 - \Phi(u)]^2} \left[\sigma(1 - \Phi(u)) - \frac{1}{\sigma}u\phi(u) \right],$$

where we define $u \equiv (\log \tau - \mu)/\sigma$ and $\phi(u)$, $\Phi(u)$ are the density and distribution function of the standard normal distribution, respectively. Therefore, by invoking the δ method [see, for example, Billingsley (1986, p. 402)], we obtain that the asymptotic distribution of $\beta_{\hat\theta}(\tau)$ is given by

$$\sqrt{n}(\beta_{\hat\theta}(\tau) - \beta_\theta(\tau)) \xrightarrow{L} N(\mathbf{0}, v_\theta(\tau)), \qquad n \to \infty,$$

where

$$v_\theta(\tau) = \left(\frac{\partial\beta}{\partial\mu}\frac{\partial\beta}{\partial\sigma}\right)\mathbf{I}^{-1}(\theta)\begin{pmatrix}\frac{\partial\beta}{\partial\mu}\\ \frac{\partial\beta}{\partial\sigma}\end{pmatrix} = \sigma^2\left[\left(\frac{\partial\beta}{\partial\mu}\right)^2 + \frac{1}{2}\left(\frac{\partial\beta}{\partial\sigma}\right)^2\right]. \quad (9)$$

3.2 Optimal Threshold Level for GATE: Procedure 1

By the first procedure, for a fixed θ we obtain an optimal threshold level τ, when the positive rain rate is lognormally distributed, which minimizes the function

$$\begin{aligned} v_\theta(\tau) &= \sigma^2\left[\left(\frac{\partial\beta}{\partial\mu}\right)^2 + \frac{1}{2}\left(\frac{\partial\beta}{\partial\sigma}\right)^2\right] \\ &= \sigma^2\frac{\exp(2\mu+\sigma^2)}{[1-\Phi(u)]^4} \\ &\quad\times\left\{\left[1-\Phi(u)-\frac{1}{\sigma}\phi(u)\right]^2\right. \\ &\quad\left.+\frac{1}{2}\left[\sigma(1-\Phi(u))-\frac{1}{\sigma}u\phi(u)\right]^2\right\}, \end{aligned} \quad (10)$$

where as before $u = (\log\tau - \mu)/\sigma$. By using the fact (Feller 1968, p. 175) that $(1-\Phi(u)) \sim \phi(u)/u$, as $u \to \infty$, it is easy to see that for all $\theta \in (-\infty, \infty) \times (0, \infty)$ the asymptotic behavior of $v_\theta(\tau)$ is dictated by

$$\lim_{\tau\to\infty} v_\theta(\tau) = \infty, \qquad \lim_{\tau\to 0} v_\theta(\tau) = \sigma^2(1+\sigma^2/2)\exp(2\mu+\sigma^2). \quad (11)$$

By (11) the minimum of $v_\theta(\tau)$ exists.

Because we would like to explain the results in Figure 1, we adhere to GATE-like rain and thus, inspired by Table 1, we specialize to the case $\theta = (1, 1)$. So, in accordance with the first procedure, we minimize $v_{(1,1)}(\tau)$ to obtain an optimal level τ_{opt}. This is done as follows. First it is shown that the derivative of $v_{(1,1)}(\tau)$ is always positive for τ greater than some fixed positive value. The minimum then must be located in the interval between 0 and that fixed positive value. The minimum is found by actually computing $v_{(1,1)}(\tau)$ over that interval.

Set $u = \log\tau - 1$, and let $v^*_{(1,1)}(u)$ be given by the right side of (10):

$$v^*_{(1,1)}(u) = \frac{\exp(3)}{[1-\Phi(u)]^4} \times\left\{[1-\Phi(u)-\phi(u)]^2 + \frac{1}{2}[1-\Phi(u)-u\phi(u)]^2\right\}.$$

Then, $dv_{(1,1)}(\tau)/d\tau = \tau^{-1}dv^*_{(1,1)}(u)/du$. From the definition of $v^*_{(1,1)}(u)$ and by using $\phi'(u) = -u\phi(u)$, we obtain $dv^*_{(1,1)}(u)/du = \{\exp(3)\phi(u)/[1-\Phi(u)]^5\}P(u)$, where $P(u)$ is given by the equation

$$P(u) = (u^2+2u+2)(1-\Phi(u))^2 - (u^3+4u+6)\phi(u)(1-\Phi(u)) + (2u^2+4)\phi^2(u).$$

From this and the inequality $(u^2-1)/u^3 < 1-\Phi(u) < \phi(u)/u$, for $u > 0$ (Feller 1968, p. 175), follows the inequality $P(u) > u^{-6}\phi^2(u)Q(u)$, for $u > 1$, where $Q(u)$ is an 8th degree polynomial given by

$$Q(u) = u^8 + u^6 - 4u^5 - 4u^3 - 3u^2 + 2u + 2.$$

But $u^8 - 4u^5 = u^5(u^3-4) > 0$ and $u^6 - 4u^3 - 3u^2 = u^2(u^4 - 4u - 3) > 0$, for $u > 2$. Hence $Q(u) > 0$ for $u > 2$. Therefore, we have shown that $v_{(1,1)}(\tau)$ is a strictly monotone increasing function of τ for $\tau > \exp(3) \doteq 20.0855$. The values of $\log(v_{(1,1)}(\tau))$ are computed in Figure 2 over an interval extending from 0 to a value exceeding $\tau = 21$. The minimum occurs at $\tau_{\text{opt}} \doteq 4.84052$, and this is the value of the desired optimal threshold level obtained by Procedure 1. Evidently, this value agrees well with the best experimental value of $\tau = 5$ obtained by Chiu (1988) and displayed in Figure 1, and it is not too far from the mean of the lognormal distribution $\Lambda(1, 1)$ given by $\exp(3/2) \doteq 4.48169$.

The same analysis was repeated for several values of $\theta = (\mu, \sigma)$ to cover rain types other than GATE. The results are given in Table 2. Each entry in the table gives the pair (τ_{opt}, M) corresponding to a certain $\theta = (\mu, \sigma)$, where τ_{opt} is the determined optimal level at which the minimum occurs and M is the mean of the lognormal distribution with parameter $\theta = (\mu, \sigma)$. From the table we can see that for reasonable rain rates, the mean of the distribution of rain rate, conditional on rain, provides a good approximation for the optimal threshold level.

It is interesting to note that a similar conclusion has been reached by a completely different method in Kedem, Chiu, and Karni (1990). There we minimized a certain distance, as a function of τ, between $(\beta_\theta(\tau) \equiv)\ \beta(\tau)$'s across several distributions all having the same mean and variance. In addition, in that work it was observed that the slope $\beta(\tau)$ shows a surprising degree of stability in the face of reasonable changes in the parameters of f, where f is modeled by the densities of the lognormal, gamma, and inverse Gaussian distributions.

3.3 Optimal Threshold Level for GATE: Procedure 2

Fix t. We wish to find the level τ that maximizes the correlation between the area average $\langle X_t\rangle$ and the fractional

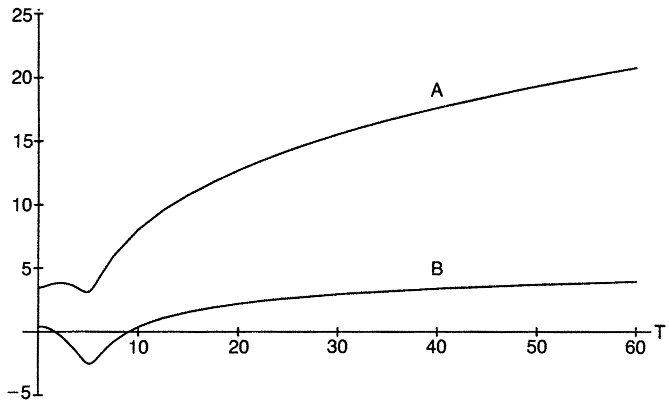

Figure 2. (A) Graph of log $(v(\tau))$. Minimum occurs at $\tau_{opt} \doteq 4.84052$. (B) Graph of $\log(w(\tau))$. Minimum occurs at $\tau^*_{opt} \doteq 5.1309$.

Table 2. Optimal Threshold Level by Procedure 1 as a Function of $\theta = (m, s)$

	s						
m	.7	.8	.9	1	1.1	1.2	1.3
.7	2.12	2.49	2.96	3.58	4.42	5.55	7.11
	2.57	2.77	3.01	3.32	3.68	4.13	4.68
.8	2.34	2.75	3.27	3.96	4.88	6.14	7.86
	2.84	3.06	3.33	3.66	4.07	4.57	5.18
.9	2.59	3.04	3.61	4.37	5.39	6.78	8.69
	3.14	3.38	3.68	4.05	4.50	5.09	5.72
1.0	2.86	3.36	3.99	4.84	5.96	7.49	9.60
	3.47	3.74	4.07	4.48	4.97	5.58	6.32
1.1	2.12	2.49	2.96	3.58	4.42	5.55	7.11
	3.83	4.13	4.50	4.95	5.50	6.17	6.99
1.2	3.50	4.10	4.88	5.91	7.28	9.15	11.73
	4.24	4.57	4.97	5.47	6.07	6.82	7.72
1.3	3.86	4.53	5.39	6.53	8.05	10.12	12.97
	4.68	5.05	5.50	6.04	6.71	7.53	8.54

NOTE: In each pair, the first number is the optimal level τ_{opt} and the second number is the mean of the lognormal distribution $\Lambda(m, s)$.

area $\langle I[X_t > \tau]\rangle$. In the following heuristic discussion we argue that by minimizing $w_\theta(\tau) \equiv v_\theta(\tau)/\beta_\theta^2(\tau)$ with respect to τ, for a fixed θ, we come close to achieving this goal. The argument is not intended as a proof of anything but as motivation for an interesting criterion. With this in mind, we make the following points.

1. Because in practice $\langle X_t\rangle$ is really obtained from $\beta_{\hat\theta}(\tau) \times \langle I[X_t > \tau]\rangle$, we find it convenient to modify the problem slightly and ask for the level τ that maximizes the correlation between $\langle X_t\rangle$ and $\beta_{\hat\theta}(\tau)\langle I[X_t > \tau]\rangle$.
2. Observe that the nature of the problem at hand is such that in practice $\beta_{\hat\theta}(\tau)$ is obtained from an *independent* random sample *prior to* actual prediction. Therefore, $\beta_{\hat\theta}(\tau)$ is independent of $\langle X_t\rangle$ and $\langle I[X_t > \tau]\rangle$.
3. Note that $\beta_{\hat\theta}(\tau)$ is asymptotically unbiased, and assume that the sample is sufficiently large so that $E[\beta_{\hat\theta}(\tau)] \doteq \beta_\theta(\tau)$.

Based on these observations and simplifying assumptions we obtain a useful inequality:

$$\operatorname{corr}^2[\langle X_t\rangle, \beta_{\hat\theta}(\tau)\langle I[X_t > \tau]\rangle]$$

$$\doteq \beta_\theta^2(\tau)\,\frac{\operatorname{var}[\langle I[X_t > \tau]\rangle]}{\operatorname{var}[\beta_{\hat\theta}(\tau)\langle I[X_t > \tau]\rangle]}\operatorname{corr}^2[\langle X_t\rangle, \langle I[X_t > \tau]\rangle]$$

$$\le \beta_\theta^2(\tau)\,\frac{\operatorname{var}[\langle I[X_t > \tau]\rangle]}{\operatorname{var}[\beta_{\hat\theta}(\tau)]E[\langle I[X_t > \tau]\rangle^2] + \beta_\theta^2(\tau)\operatorname{var}[\langle I[X_t > \tau]\rangle]}$$

$$= \frac{1}{\dfrac{E[\langle I[X_t > \tau]\rangle^2]}{\operatorname{var}[\langle I[X_t > \tau]\rangle]}\,\dfrac{\operatorname{var}[\beta_{\hat\theta}(\tau)]}{\beta_\theta^2(\tau)} + 1} \le \frac{1}{\dfrac{\operatorname{var}[\beta_{\hat\theta}(\tau)]}{\beta_\theta^2(\tau)} + 1}. \tag{12}$$

It follows that the quantity $\operatorname{var}[\beta_{\hat\theta}(\tau)]/\beta_\theta^2(\tau)$ must be as small as possible. That is, it must be minimized with respect to τ. This suggests $w_\theta(\tau) \equiv v_\theta(\tau)/\beta_\theta^2(\tau)$ as a criterion for choosing an optimal level for the estimation of the area mean rain rate. Motivated by this idea, we show that, under lognormality $\Lambda(\mu, \sigma)$, $w_\theta(\tau)$ has a minimum, and then we proceed to find its location for the special case $\Lambda(1, 1)$ as was done before to accommodate GATE-like rain.

For $u = (\log \tau - \mu)/\sigma$ define

$$S_\theta(u) \equiv \frac{\sigma^2}{[1-\Phi(u)]^2}\left\{\left[1 - \Phi(u) - \frac{1}{\sigma}\phi(u)\right]^2 + \frac{1}{2}\left[\sigma(1-\Phi(u)) - \frac{u}{\sigma}\phi(u)\right]^2\right\}.$$

Then, $S_\theta(u(\tau)) = w_\theta(\tau) = v_\theta(\tau)/\beta_\theta^2(\tau)$. Again, using the fact that $(1 - \Phi(u)) \sim \phi(u)/u$, as $u \to \infty$, we obtain

$$\lim_{\tau\to\infty} w_\theta(\tau) = \infty, \quad \lim_{\tau\to 0} w_\theta(\tau) = \sigma^2(1 + \sigma^2/2). \tag{13}$$

Therefore, the minimum of $w_\theta(\tau)$ exists. Specializing to the case $\theta = (1, 1)$ and using computation similar to that used before, we can show that

$$dw_{(1,1)}(\tau)/d\tau > \tau^{-1}\{u^{-2}\phi(u)/(1-\Phi(u))\}^3 Q^*(u), \qquad u > 1,$$

where $Q^*(u) = u^6 - 3u^4 - 4u^3 + 3u^2 + 2u - 1$. But $3u^2 + 2u - 1 > 0$ and $u^6 - 3u^4 - 4u^3 > 0$, for $u > 3$. Therefore, $Q^*(u) > 0$, for $u > 3$, and $w_{(1,1)}(\tau)$ is strictly monotone increasing for $\tau > \exp(4) \doteq 54.59815$. The values of $\log(w_{(1,1)}(\tau))$ are computed in Figure 2 on an interval extending from 0 and passing 55. The minimum occurs at $\tau_{\text{opt}}^* = 5.1309$, and this is the optimal threshold by Procedure 2. Again, this is close to the experimental optimum of $\tau = 5$ and not too far from the mean, conditional on rain, $E[X \mid X > 0] = \exp(1.5) = 4.48169$.

Figure 2 also shows some general similarity between the graphs of $v_{(1,1)}(\tau)$ and $w_{(1,1)}(\tau) = v_{(1,1)}(\tau)/\beta_{(1,1)}^2(\tau)$. Both curves show minima in a neighborhood of $\tau = 5$: $\tau_{\text{opt}} < 5 < \tau_{\text{opt}}^*$. Thus we have matched the experimental results.

The same analysis has been repeated for other values of θ. Table 3 gives the corresponding values of the optimal level τ_{opt}^* obtained by Procedure 2, as well as the values of M given in Table 2. In this connection see also the plot of $\log(\beta_\theta^2(\tau))$ in Figure 3 for various values of θ. By Markov inequality, $\log(\beta_\theta^2(\tau))$ is an increasing function of the threshold level τ.

Table 3. Optimal Threshold Level by Procedure 2 as a Function of $\theta = (m, s)$

	s						
m	.7	.8	.9	1	1.1	1.2	1.3
.7	2.33	2.69	3.17	3.80	4.66	5.82	7.42
	2.57	2.77	3.01	3.32	3.68	4.13	4.68
.8	2.57	2.97	3.50	4.20	5.15	6.43	8.20
	2.84	3.06	3.33	3.66	4.07	4.57	5.18
.9	2.84	3.28	3.87	4.64	5.69	7.11	9.07
	3.14	3.38	3.68	4.05	4.50	5.09	5.72
1.0	3.14	3.63	4.27	5.13	6.28	7.86	10.02
	3.47	3.74	4.07	4.48	4.97	5.58	6.32
1.1	3.47	4.01	4.72	5.67	6.95	8.68	11.07
	3.83	4.13	4.50	4.95	5.50	6.17	6.99
1.2	3.84	4.43	5.21	6.27	7.68	9.59	12.24
	4.24	4.57	4.97	5.47	6.07	6.82	7.72
1.3	4.24	4.90	5.77	6.92	8.48	10.60	13.52
	4.68	5.05	5.50	6.04	6.71	7.53	8.54

NOTE: In each pair, the first number is the optimal level τ_{opt}^* and the second number is the mean of the lognormal distribution $\Lambda(m, s)$.

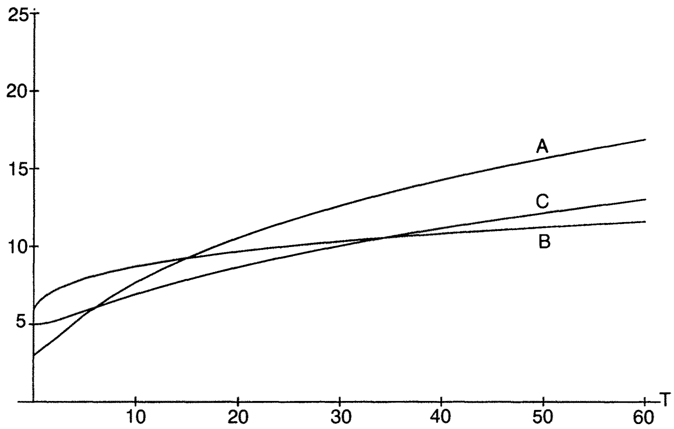

Figure 3. Graph of $\log(\beta_\theta^2(\tau))$. (A) $\theta = (1, 1)$. (B) $\theta = (1, 2)$. (C) $\theta = (2, 1)$.

Evidently, the two optimality criteria produce similar optimal threshold levels. In this regard, it is interesting to note that in the exponential case the optimal threshold level by the two criteria is one and the same and is equal to the mean of the distribution, conditional on rain.

[*Received October 1989. Revised December 1990.*]

REFERENCES

Arkel, R., and Hudlow, M. (1977), *GATE International Meteorological Radar Atlas,* U.S. Dept. of Commerce, NOAA, Washington, DC: U.S. Government Printing Office.

Battan, L. J. (1973), *Radar Observations of the Atmosphere,* Chicago: The University of Chicago Press.

Billingsley, P. (1986), *Probability and Measure* (2nd ed.), New York: John Wiley.

Biondini, R. (1976), "Cloud Motion and Rainfall Statistics," *Journal of Applied Meteorology,* 15, 205–224.

Chiu, L. S. (1988), "Rain Estimation From Satellites: Areal Rainfall-Rain Area Relations," 3rd Conference on Satellite Meteorology and Oceanography, American Meteorological Society, Jan. 31–Feb. 4, Anaheim, California.

Doneaud, A. A., Niscov, S. I., Priegnitz, D. L., and Smith, P. L. (1984), "The Area-Time Integral as an Indicator for Convective Rain Volumes," *Journal of Climate and Applied Meteorology,* 23, 555–561.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications* (Vol. I), New York: John Wiley.

Kedem, B., and Chiu, L. S. (1987), "On the Lognormality of Rain Rate," *Proceedings of the National Academy of Sciences,* 84, 901–905.

Kedem, B., Chiu, L. S., and Karni, Z. (1990), "An Analysis of the Threshold Method for Measuring Area Average Rainfall," *Journal of Applied Meteorology,* 29, 3–20.

Kedem, B., Chiu, L. S., and North, G. R. (1990), "Estimation of Mean Rain Rate: Application to Satellite Observations," *Journal of Geophysical Research,* 95, No. D2, 1965–1972.

Kedem, B., and Short, D. (1989), "Linear Relationships for Rain Rate Estimation From Space," in *Precipitation Measurement,* ed. B. Sevruk, WMO/IAHS/ETH Workshop on Precipitation Measurement, St. Moritz, December 3–7, pp. 441–445.

Lehmann, E. L. (1983), *Theory of Point Estimation,* New York: John Wiley.

Lopez, R. E. (1977), "The Log-Normal Distribution and Cumulus Cloud Populations," *Monthly Weather Review,* 105, 865–872.

Patterson, V. L., Hudlow, M. D., Pytlowany, P. J., Richards, F. P., and Hoff, J. D. (1979), *GATE Radar Rainfall Processing System,* NOAA Technical Memorandum, EDIS 26, Washington, DC.

Rosenfeld, D., Atlas, D., and Short, D. (1990) "The Estimation of Convective Rainfall by Area Integrals 2. The Height-Area Threshold (HART) Method," *Journal of Geophysical Research,* 95, No. D3, 2161–2176.

Short, D. A., Wolff, D. B., Rosenfeld, D., and Atlas, D. (1989), "A Rain-Gage, Radar and Satellite Simulation Study of the Estimation of Convective Rainfall by Area-Time Integrals," 4th Conference on Satellite Meteorology and Oceanography, May 16–19, San Diego, California.

Simpson, J. (1988), "TRMM, a Satellite Mission to Measure Tropical Rainfall," *Report to the Science Steering Group,* National Aeronautics and Space Administration.

Simpson, J., Adler, R. F., and North, G. R. (1988), "A Proposed Rainfall Measuring Mission Satellite," *Bulletin of the American Meteorological Society,* 69, 278–295.

Wilheit, T. T. (1986), "Some Comments on Passive Microwave Measurement of Rain," *Bulletin of the American Meteorological Society,* 67, 1226–1232.