Resenhas IME-USP 1993, Vol. 1, No. 2/3, 217 - 232.

On A Curious Linear Relationship Between Rainfall Averages [1]

B. Kedem and D. A. Short

Abstract: Empirical evidence points to the fact that the area average rain rate and the fraction of the area where rain rate exceeds a given threshold tend to be highly correlated, provided the area is large enough and the threshold is chosen optimally as to increase the correlation. This fact has an important application in rainfall estimation from space using satellite borne instruments. A statistical explanation is provided for the observed linearity, and a method for optimal thresholds is discussed.

Key words: Rain rate, area average, TRMM satellite, mixed distribution, lognormal, space time.

1 Introduction

Meteorologists have been reporting high positive correlations between precipitation amounts and various area statistics for many years. One particular case which came to light only a few years ago in connection with a space mission is the subject of the present paper. Our goal is to describe and explain an observed curious linear relationship between the instantaneous *area average* rain rate (in mm/hr) and the *fraction* of the area where rain rate exceeds a given threshold. Experimental evidence, obtained from quite a few data sets, shows that when the threshold is chosen optimally, the sample correlation between the area average and the fractional area can be as high as 99%. This experimental fact can be explained in more than one way.

The starting point of *our* solution is the answer to the following question: "What is the most characteristic thing about rain ?" "Wet" is not the answer, but *intermittency* is. That is, since it does not rain between rain events, the distribution of rain rate has an atom at 0, and this simple fact together with some assumptions lead to a plausible explanation to the observed linearity. This clearly is not the end of the story, for other approaches, rooted in the theory of random fields, are possible and any such explanation to the intriguing linear relationship by the statistical community is certainly welcome.

To proceed intelligently we must first introduce briefly some scientific facts and terms associated with the *TRMM mission*.

[1] Work supported by NASA S-97607-E, ONR N00014-92-C-0019 and NSF CDR-88-03012.

1.1 The TRMM Mission: Measurement of Rainfall From Space

A series of "Earth Probe" missions is planned for the 1990's by the National Aeronautics and Space Administration (NASA) to advance our understanding of global climate change. Among these is the Tropical Rainfall Measuring Mission (TRMM)–a U.S-Japan joint space project–whose primary goal is the measurement of the annual total volume of tropical rainfall and its variation. The reason is that through tropical rainfall–it accounts for over 2/3 of worldwide precipitation–it is possible to estimate the amount of atmospheric latent heat release and assess its role in driving the circulation of the atmosphere. The projected byproducts are long range weather and climate forecasts connected with the global hydrological cycle. Another important byproduct is an assessment of the relationship between tropical rainfall and the El niño phenomenon. For many more details see Simpson (1988) and Simpson et al. (1988).

So much for science. Now the actual implementation. Putting buckets throughout the tropics is out of the question, for the tropics are covered by massive oceans and jungles and thus to a great extent are practically inaccessible. Adding to the predicament is the fact that our present day existing technology has not been able to provide reliable rain gauges which can be installed over the oceans and sustain heavy storms and very strong waves in the open seas. The NASA solution then is to let a satellite–the TRMM satellite–do the job. However, the satellite instruments–microwave, visible, and infrared radiometers, and a precipitation radar–*do not* measure rainfall but other variables, such as microwave temperature and radar reflectivity, whose relationship to rainfall (really rain rate) is nonlinear and not entirely clear at that. Putting it differently, from a given microwave temperature we may not be able to tell apart light from heavy rain. Another potential problem is that even if all goes well and we use the "correct" $Z-R$ relationship between reflectivity (Z) and rain rate (R), the satellite borne radar has a dynamic range of approximately 80 mm/hr. However, it is well known that rain rate in the range of hundreds of mm/hr is a pretty normal event in the tropics, and this is beyond the radar capability. So what do we do ?

1.2 The Threshold Method

There are various ways to overcome the measurement problem, including the aforementioned observed linear relationship between the area average and the fractional area. The latter only requires the classification of instantaneous rain rate measurements as being *above or below* an optimal threshold–to obtain an estimate of the fractional area–but not the actual measurements themselves, a much simplified problem compared to the requirement of precise point measurements. Inferring the area average rain rate from the fractional area where rain rate exceeds a threshold is called the *threshold method* (Kedem et al. (1990a), Kedem and Pavlopoulos (1991), Braud et al. (1993), Short et al. (1993b)). Krajewski et al. (1992) use *area-threshold method*, and another name–for a slightly modified technique–suggested by Rosenfeld et al. (1990) is *height-area rainfall threshold (HART)*.

Interestingly, the method can be used in reverse as well using the 0 threshold to estimate the storm area from the area average or alternatively the volume of rainfall (Eltahir and Bras (1993)).

Regarding the optimal threshold, it is obtained under some assumptions on the *continuous* part of the distribution of rain rate. Suppose it is lognormal $\Lambda(\mu, \sigma)$. Then the optimal threshold is obtained by maximizing with respect to τ the quantity, $S_\theta(u)$, defined for $u = (\log \tau - \mu)/\sigma$ and $\theta = (\mu, \sigma)$ as

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

If one stretches one's imagination it is not difficult to see that $S_\theta(u(\tau))$ is essentially a *periodogram-like quantity where the threshold τ plays the role of frequency.*

1.3 The GATE Data Set

We shall make reference to the GARP Atlantic Tropical Experiment (GATE) data set comprised of instantaneous rain rate snapshots–taken mostly every 15 minutes–obtained by radar over the eastern Atlantic Ocean in the summer of 1974, some hundreds of kilometers off the coast of west Africa (Hudlow and Patterson (1979), Simpson (1988) p. 37). The size of the area in question is that of a circle 400 km in diameter, and the data are considered close to being "ground truth". There are several phases of GATE of which phases I and II consist of 1716 (18 days) and 1512 (15 days) snapshots, respectively, obtained from radar reflectivity binned into $4 \times 4\ km^2$ pixels. Relative to the size of the area, it is convenient to think of each $4 \times 4\ km^2$ pixel as a point in space. The GATE data have been the source for numerous studies.

2 Empirical Linear Relationships

The meteorological literature offers numerous examples of interesting empirical linear relationships between rainfall and area statistics. Following is a very brief account of several such examples.

A simple example of linear relationship is the empirical fact that the total precipitation is positively correlated with the number of rainy days over a given area. The total volume is obtained by multiplying the average amount of rainfall per day (can be obtained for example ¿from the climatology of the region) times the number of rainy days. This method was used by Supan as early as 1898 to estimate annual rainfall over the Atlantic and Indian Oceans, as a function of latitude, by collecting data from ships at sea. See also Mintz (1981) for related subsequent work.

Another example relates rainfall volume V to area A as follows. Note that

$$V = \int_T \int_A R(a,t)\,da\,dt = \overline{R} \int_T \int_A da\,dt$$

where $R(a,t)$ is the instantaneous rain rate at time t and point a in space, T is the period of observation, and $\overline{R}$ is the space time average rain rate. The area-time integral (ATI) refers to $\int\int dadt$, but it is more convenient to define it as the approximating sum (Doneaud et al. (1984)),

$$ATI \equiv \sum_i A_i \Delta t_i$$

where A_i is the area where rain is detected and Δt_i is the time interval between observations. Doneaud et al. (1984) report high correlations between an estimate of V (obtained ¿from the $Z-R$ relationship $Z = 155R^{1.88}$) and the ATI when the A_i are estimated from radar echos (measured in dBZ) in excess of certain thresholds. For a 25 dZB threshold the correlation is 98%. Here the high correlation is quite subtle since most likely $\overline{R}$ varies from one rain event to the next. Clearly, if $\overline{R}$ were a constant, the high correlation would not come down as a big surprise, its source being two estimates–differing essentially by a constant multiple–of the same quantity, the true rainfall volume. Apparently $\overline{R}$ did not vary much across rain storms during the summers of 1980 and 1981 in western North Dakota, the periods and area over which the data were collected.

Another subtle example of linear relationship is the high correlation between the area covered with radar reflectivity above a threshold and the rainfall amount over that area reported by Hudlow and Scherer (1975), Lovejoy and Austin (1979), as well as by others. Hudlow and Scherer (1975) give an example using a 20 dBZ threshold where the correlation is 95%.

Arkin (1979) found a correlation over 80% and as high as 89% between 6 hr. rainfall accumulations and the corresponding 6 hr. averages of fractional coverage of cloud higher than 10 km using GATE data. Similar results were obtained by Richards and Arkin (1981) and Arkin and Meisner (1987) when the coverage area is defined by clouds colder than certain temperature thresholds. The latter reference also discusses spatial and temporal scale considerations.

Perhaps the most intriguing of all these examples, and the starting point of several exciting investigations by NASA scientists, is the experiment conducted by Chiu (1988) using the entire Phase I and Phase II of the GATE data. Observe that each snapshot from GATE gives rise to an instantaneous area average rain rate and to a percent of the area where the instantaneous rain rate exceeds a given threshold. Thus we have several sets–corresponding to the fixed thresholds–of 1716 pairs from Phase I, and likewise several sets of 1512 pairs from Phase II. For each fixed threshold, Chiu computed the sample correlation between the area average and the fractional area. The results, given in Table 1, reveal the important fact that the correlation is a function of the threshold, and that for a threshold of 5 mm/hr the correlation can reach 99%.

Triggered by the work of Chiu (1988), Atlas et al. (1990) and Rosenfeld et al. (1990) repeated the same experiment using data ¿from GATE III (over 1600 scans) as well as data from other locations–central South Africa (over 2450 scans), Texas (over 1300 scans), and Darwin (Australia, 48 scans). They report similar high correlations, shown in Table 2, well above 90% for thresholds $\tau = 2, 4, 6, 8$ mm/hr. The authors also show that a slight improvement in correlation may be achieved if the snapshots are classified according to storm height.

Krajewski et al. (1992) used the Bell model (Bell (1987)) to simulate artificial space-time rain rate fields. For thresholds $\tau = 0, 5, 10, 20$ mm/hr they obtained correlations 0.9359,0.9879,0.9964,0.9673, respectively.

In what follows we provide an explanation for the high correlation observed between the area average rain rate (from now on "area average") and the fractional area where rain rate exceeds a given threshold (from now on "fractional area") by arguing that under some conditions these quantities are essentially linearly related. We also suggest a method for deriving optimal thresholds by assuming a parametric model for the distribution of rain rate. By assuming lognormal rain rate, conditional on rain, our method yields for GATE-like rain an optimal thresholds which essentially matches that of Chiu (1988).

Table 1. *Sample correlation r between the area average rain rate and the fraction of the area above threshold $\tau = 0, 1, 5, 10, 20$ mm/hr for GATE I, II. Source: Chiu (1988).*

Threshold τ	GATE I r	GATE II r
0	0.883	0.843
1	0.943	0.922
5	0.990	0.985
10	0.975	0.980
20	0.922	0.933

Table 2. *Sample correlation r between the area average rain rate and the fraction of the area above threshold $\tau = 2, 4, 6, 8$ mm/hr for GATE III, central South Africa, Texas, and Darwin (Australia) Source: Rosenfeld et al. (1990).*

Threshold τ	GATE III r	S. Africa r	Texas r	Darwin r
2	0.942	0.914	0.972	—
4	0.965	0.964	0.986	—
6	0.981	0.965	0.990	0.965
8	0.984	0.963	0.991	—

3 Lognormality of Rain Rate

The distribution of rain rate, conditional on rain, plays a fundamental role in rainfall estimation. Empirical evidence shows that the distribution is highly skewed and resembles that of lognormal or gamma, but no acceptable theoretical justification for any distribution exists at present. Nevertheless, there are arguments predicated on time series of rain rate which lend some credibility to the lognormal hypothesis. Kedem and Chiu (1987), modeling rain rate time series by a stochastic regression, provided some *necessary* conditions for lognormality. The conditions were found to hold to a surprising degree by time series from GATE. More recently, Pavlopoulos and Kedem (1992) modeled rain rate in time as a diffusion process with certain intuitive drift and diffusion coefficients and obtained a new parametric family containing the lognormal distribution. Kedem et al. (1990b) found that rain rate ¿from GATE I and II greater than 1 mm/hr (truncated at 1 mm/hr to overcome noisy data close to 0) gave excellent lognormal fits but relatively mediocre gamma fits. This is in line with an earlier work of Houze and Cheng (1977) who found by experimental means that the echo size of GATE tended to be lognormal. On the other hand, Meneghini and Jones (1993), using different rain rate data, had more success with the gamma distribution (compare with LeCam (1961)). Lovejoy and Schertzer (1985) and others argue that the distribution must have a heavy tail altogether. Models for the distribution of space-time rainfall assuming self similarity and random cascading of of rain fields are discussed in detail in Gupta and Waymire (1993).

As far as the present work is concerned, following Kedem et al. (1990b) we shall assume that positive GATE rain rate has a lognormal distribution with both parameters approximately equal to 1.

4 A Space-Time Box

Since what we have is a space-time measurement, namely rain rate, it is convenient to adopt a mathematical model based on a space-time box for reference. Thus we can speak of points in the box and of, say, horizontal slices of the box.

Following Kedem and Pavlopoulos (1991), we think of time as varying on a vertical axis, and of space being horizontal. Suppose rain rate is observed over a given region A and throughout a specific period $[0, T]$, and consider the box $\Omega \equiv A \times [0, T]$. With each each $\omega \in \Omega$ we associate a rain rate value, and let $X(\omega)$ be the random variable which gives the instantaneous value of rain rate associated with ω. Then, X has a *mixed* distribution because it admits the value 0 (no rain) with positive probability, say, $1-p$. That is, $P(X = 0) = 1 - p$. Denote the mixed distribution by the pair (p, f), where f is the probability density with respect to Lebesgue measure of the continuous component of the distribution.

First we refer to the box. If $\varphi(X)$ is an arbitrary integrable function of X, then its expected value is

$$E(X) = \beta_\varphi\{E[\varphi(X)] - \varphi(0)\} \tag{1}$$

where

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

depends only on f but not on p. Here are two examples of (1) and (2) from (Kedem and Short (1989)).

Example 4.1. Let $\phi(X) = X^k$. Then $\phi(0) = 0$ and

$$E(X) = \beta_\phi E[X^k]$$

where

$$\beta_\phi = \frac{E[X|X>0]}{E[X^k|X>0]}$$

depends on f only. □

Example 4.2. The widely used $Z - R$ relationship (Battan (1973), p. 89)

$$Z = cX^b \tag{3}$$

where $c, b > 0$, relates space time point measurements of rain rate X to radar reflectivity Z. This is a nonlinear equation between X and Z, however, the relationship between the *expected values* can be linearized as follows. Define $\phi(X)$ by the conditional expectation,

$$\phi(X) \equiv E(Z|X)$$

This can be done because our space-time box clearly accommodates the *joint* distribution of (X, Z). We have,

$$\begin{aligned} \phi(0) &= E(Z|X=0) = 0 \\ E(\phi(X)) &= E[E(Z|X)] = E(Z) \\ E[\phi(X)|X>0] &= E[E(Z|X)|X>0] = E(Z|X>0) \end{aligned}$$

and so by substitution in (1) and (2) we obtain a precise equation,

$$E(X) = \beta_\phi E(Z) \tag{4}$$

where

$$\beta_\phi = \frac{E(X|X>0)}{E(Z|X>0)} \tag{5}$$

□

Next we refer to a horizontal slice from the box. So, fix t and let $X_t(\mathbf{a})$ be the random variable which gives the value of rain rate at the point $\mathbf{a}$ in space, at time t. As with X, the distribution of X_t is mixed with components (p_t, f_t), and from (2) we get the slope

$$\beta_\varphi(t) = \frac{E[X_t|X_t>0]}{E[\varphi(X_t)|X_t>0] - \varphi(0)} \tag{6}$$

which depends on f_t only but not on p_t.

With reference to the GATE data, there is empirical evidence which suggests that, conditional on rain, the behavior of tropical rain rate is reasonably homogeneous in time and space throughout a short period. For a discussion of this see Kedem and Pavlopoulos (1991) and to some extent the above discussion regarding the ATI. Thus, emboldened by intuition derived from some empirical work we make the following simplifying assumption.

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

It follows that $\beta_\varphi(t) = \beta_\varphi$ does not depend on t.

Observe that no such assumption is made about p_t. Now, the homogeneity of f_t and the variability of p_t give us what we want.

Consider a slice from the box at time t, draw a random sample of rain rate values over the slice, and let $< X_t >$ and $< \varphi(X_t) >$ be the sample averages of X_t and $\varphi(X_t)$ respectively. Clearly, $< X_t >$ and $< \varphi(X_t) >$ are area averages. *We define $< X_t >$ as the area average of rain rate.* By the law of large numbers then, as the sample size increases, $< X_t > \rightarrow E(X_t)$ and $< \varphi(X_t) > \rightarrow E[\varphi(X_t)]$ with probability one. Therefore from (1) and the homogeneity assumption, for a sufficiently large sample we have the approximation for each fixed t,

$$< X_t > \simeq \text{constant} + \beta_\varphi < \varphi(X_t) > \tag{7}$$

This is a linear relationship between two area averages. *Thus, when the homogeneity assumption is satisfied, there is a high correlation between the area average rain rate and the area average of any function thereof.*

Example 4.3 As an example of (7), we considered $< X >$ vs $< X^2 >$ using snapshots from GATE of size 280×280 km^2. Linear regression gives correlation 92% for GATE I and 90% for GATE II. □

We can now specialize (7) to explain the high correlation between the area average and the fractional area illustrated in Tables 1 and 2. To do that fix a threshold τ and define the indicator function,

$$\varphi(x) = \begin{cases} 1 & \text{if } x > \tau \\ 0 & \text{if } x \leq \tau \end{cases} \tag{8}$$

Since $\varphi(0) = 0$, and

$$< \varphi(X_t) > \simeq E[\varphi(X_t)] = P(X_t > \tau) \simeq < I[X_t > \tau] >$$

the general approximation (7) entails, substituting $\beta(\tau)$ for $\beta_\varphi(\tau)$,

$$< X_t > \simeq \beta(\tau) < I[X_t > \tau] > \tag{9}$$

where under homogeneity

$$\beta(\tau) = \frac{E[X|X > 0]}{P(X > \tau|X > 0)}$$

is a constant which depends on f and τ but not on t. Since $< I[X_t > \tau] >\simeq P(X_t > \tau)$, and $P(X_t > \tau)$ is the true fractional area, it seems appropriate to define $< I[X_t > \tau] >$ as the (observed) *fractional area*. And so, (9) provides an explanation to the observed high correlation between the area average and the fractional area for a fixed threshold.

As was emphasized earlier, ours is a particular approach and there must be other explanations. In particular, Braud el al. (1993), using certain approximations, maintain that the high correlation is mostly due to the respective coefficients of variation of both the fractional area and the mean intensity within the delineated area where rain rate exceeds the threshold. Both approaches however are static in the sense that no underlying dynamical structure is assumed.

We close this section with a curious note. According to our formulation, it is the variability as a function of time of the discrete probability of rain p_t which gives rise to the linear relationship (9). If on the other hand the homogeneity assumption also holds for p_t, instead of a scattergram of points aligned along a straight line, we would get a scattergram around a single point.

4.1 A Cautionary Note

Although the regression equation (9) looks very promising, we must take it with a grain of salt since our derivation is based on the homogeneity assumption. Since the behavior of rain is notoriously erratic, it is very possible that in certain situations f_t can change in time due to changes in parameters or a shift to a different distribution altogether. However, if the changes are reasonably small the method is still viable. A limited sensitivity analysis conducted in Kedem et al. (1990a) indicates that the slope is quite insensitive to small changes in the parameters or even a distributional shift, provided the mean rain rate is sufficiently high, as is the case in the tropics. The slope however is very sensitive for low mean rain rate and the method then may not be reliable.

On the positive side, clearly, the threshold method may still work, that is the high correlation may still persist, even if the homogeneity assumption breaks down, for the assumption is only a mathematical convenience.

Furthermore, as indicated in Table 1, the correlation between the area average and the fractional area may not be sufficiently high for poorly chosen thresholds. For the method to be of use, the threshold must be chosen as to maximize the correlation. A method for choosing optimal thresholds is discussed in the next section.

5 Optimal Thresholds

5.1 Optimality Criteria

There are several criteria for optimal thresholds needed for the regression (9). Kedem et al. (1990a) suggest a certain distance, $d(\tau)$, between slopes $\beta(\tau)$–obtained from several different distribution models f–to be minimized as a function of τ, the idea being to render $\beta(\tau)$ resistant to distributional changes as much as possible. Ideally, a choice of τ which equalizes the different $\beta(\tau)$ is clearly optimal as far as shifts between the chosen distributions are concerned. The same idea has been expressed in Short et al. (1993b) using empirically derived distributions. Krajewski et al. (1992) suggest that the use of low thresholds for noisy rain fields. Our approach on the other hand calls for the maximization of the correlation between the area average and the fractional area as a function of τ. A very similar idea is to minimize the residuals sum of squares in the regression of the area average on the fractional area as is done in Short et al. (1993a).

Ideally, the optimal threshold τ gives maximum correlation. However we only follow this route in spirit and use instead a certain approximation suggested in Kedem and Pavlopoulos (1991) and Short et al. (1993a).

5.2 Maximum Likelihood Considerations

5.2.1 An Optimality Criterion

To formulate an optimality criterion, we turn to the estimation of $\beta(\tau)$ in (9) assuming that f is modeled by a parametric density f_θ, and consider the correlation between $< X_t >$, and $\beta_{\hat{\theta}}(\tau) < I[X_t > \tau] >$. It can be shown under some assumptions (Kedem and Pavlopoulos (1991)) that

$$Corr^2[< X_t >, \beta_{\hat{\theta}}(\tau) < I[X_t > \tau] >] \leq \frac{1}{\frac{Var[\beta_{\hat{\theta}}(\tau)]}{\beta_\theta^2(\tau)} + 1} \tag{10}$$

It follows that

$$\frac{Var[\beta_{\hat{\theta}}(\tau)]}{\beta_\theta^2(\tau)}$$

should be minimized with respect to τ.

Let $\hat{\theta}$ be the maximum likelihood estimator of θ. Then under some regularity conditions (see for example Lehmann (1983), p. 429, Billingsley (1986), p. 402), the asymptotic distribution of $\beta_{\hat{\theta}}(\tau)$ is given by

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

In the spirit of the preceding paragraph, our optimal threshold is the one which minimizes,

$$w_\theta(\tau) \equiv \frac{v_\theta(\tau)}{\beta_\theta^2(\tau)}$$

5.2.2 The Lognormal Case

Assume now that f_θ is a lognormal density with parameter $\theta = (\mu, \sigma)$. This is a specific parametric family for which the slope is automatically parametrized by θ,

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

If $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ is the maximum likelihood estimator of θ, then

$$\sqrt{n}(\beta_{\hat{\theta}}(\tau) - \beta_\theta(\tau)) \longrightarrow^{\mathcal{L}} \mathcal{N}(0, v_\theta(\tau)), \ n \rightarrow \infty$$

where,

$$v_\theta(\tau) = (\frac{\partial\beta}{\partial\mu} \ \frac{\partial\beta}{\partial\sigma})\mathbf{I}^{-1}(\theta)\begin{pmatrix} \dfrac{\partial\beta}{\partial\mu} \\ \dfrac{\partial\beta}{\partial\sigma} \end{pmatrix} = \sigma^2[(\frac{\partial\beta}{\partial\mu})^2 + \frac{1}{2}(\frac{\partial\beta}{\partial\sigma})^2] \tag{11}$$

and $\mathbf{I}^{-1}(\theta)$ is the inverse of the information matrix corresponding to the lognormal density f_θ with parameters μ and σ,

$$\mathbf{I}^{-1}(\theta) = \sigma^2\begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$$

Put $u = (\log\tau - \mu)/\sigma$, and define

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

Under lognormality the quantity to be minimized is

$$S_\theta(u(\tau)) = w_\theta(\tau) = \frac{v_\theta(\tau)}{\beta_\theta^2(\tau)} \tag{12}$$

The minimum exists and is unique (Kedem and Pavlopoulos (1991)). To find the optimal threshold under lognormal rain rate, all we have to do is to obtain μ and σ, substitute in (12), and minimize with respect to τ.

Example 5.1. Recall that the lognormal distribution with parameter $\theta = (1, 1)$ was found quite adequate for the GATE data. Plugging $\mu = 1, \sigma = 1$ in (12) and minimizing gives

$$\tau_{optimal} = 5.1 \ mm/hr$$

This is in good agreement with the experimental results presented in Table 1, where 5 mm/hr is optimal.

Table 3 provides optimal thresholds for any given (μ, σ) pair. ¿From the table we can see that, except for relatively high mean rain rate, the optimal threshold is not far removed from the mean rain rate conditional on rain. It is interesting to note that the optimal threshold entries in Table 3 can be approximated by the expression

$$\tau_{optimal} = \exp\{-0.322 - 0.014\sigma + 0.973\sigma^2 + \mu\}$$

Thus for $(\mu,\sigma) = (0.8, 1.3)$, $\tau_{optimal} = 8.2$, and for $(\mu,\sigma) = (1.2, 1.3)$, $\tau_{optimal} =$ 12.24, etc.

Table 5.1. *Optimal threshold level as a function of $\theta = (\mu,\sigma)$. In each pair, the first number is the optimal level $\tau_{optimal}$. The second is the mean of the lognormal distribution $\Lambda(\mu,\sigma)$. Source: Kedem and Pavlopoulos (1991).*

	σ						
μ	0.7	0.8	0.9	1	1.1	1.2	1.3
0.7	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
0.8	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
0.9	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

6 Extensions

The threshold method can be easily extended to the estimation of any area moment. In fact, Short et al. (1993a) point out that the same argument which leads to (9) also gives for a parametric f_θ,

$$< X_t^k > \simeq \beta_\theta(k,\tau) < I[X_t > \tau] > \tag{13}$$

where under homogeneity

$$\beta_\theta(k,\tau) = \frac{E[X^k|X>0]}{P(X>\tau|X>0)}$$

is a constant which depends on f_θ and τ but not on t. When $\hat{\theta}$ is the maximum likelihood estimator of θ and

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

the optimal threshold is the one which minimizes,

$$w_\theta(k,\tau) \equiv \frac{v_{k,\theta}(\tau)}{\beta_\theta^2(k,\tau)}$$

In the lognormal case

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

and

$$\frac{v_\theta(k,\tau)}{\beta_\theta^2(k,\tau)} = \frac{\sigma^2}{[1-\Phi(u)]^2}\{[k(1-\Phi(u)) - \frac{1}{\sigma}\phi(u)]^2 + \frac{1}{2}[k^2\sigma(1-\Phi(u)) - \frac{1}{\sigma}u\phi(u)]^2\}$$

where $\phi(u)$ and $\Phi(u)$ are the density and distribution function of the standard normal distribution, respectively, and $u = \{\log(\tau) - \mu)/\sigma\}$.

As an example, for GATE-like rain with $\mu = 1$ and $\sigma = 1$ the optimal threshold for $k = 2$ is 16.3 mm/hr. Actually, Kedem et al. (1990b) estimate μ and σ as slightly greater than 1. Thus, for example, with $\mu = 1.1$ and $\sigma = 1.05$ the optimal thresholds for the first and second area moments are about 6 and 22 mm/hr, respectively. By another account (Short et al. (1993a)) the parameters are $\mu = 0.685$ and $\sigma = 1.184$, yielding as optimal thresholds for the first and second moments the values 5.5 and 26.3 mm/hr, respectively. As far as the optimal threshold for the first area moment, these figures agree with the experimental result of Chiu (1988).

Using the same ideas, optimal thresholds can be obtained for any rain rate distribution. The gamma and inverse Gaussian cases are discussed in Short et al. (1993a). It is shown there that the optimal threshold does vary with the distribution, but apparently not drastically for practical purposes. The case of mixtures of lognormal has been studied recently in Kayano and Shimizu (1993), and a quadratic relationship between the area *variance* and the fractional area is discussed in Shimizu et al. (1993). Finally, sampling considerations regarding the optimal threshold are discussed in Pavlopoulos (1991) and in Ha (1992).

7 References

Arkin, P.A. (1979), "The relationship between fractional coverage of high cloud and rainfall accumulations during GATE over the B-scale array", *Monthly Weather Review*, 107, pp. 1382-1387.

Arkin, P.A. and B.N. Meisner (1987), "The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982-84. *Monthly Weather Review*, 115, pp. 51-74.

Atlas, D., D. Rosenfeld, and D.A. Short (1990), "The estimation of convective rainfall by area integrals 1. The theoretical and empirical basis," *Journal of Geophysical Research*, 95, No. D3, 2153-2160.

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

Bell, T. L. (1987), "A space-time stochastic model of rainfall for satellite remote-sensing studies," *Journal of Geophysical Research*, 92, pp. 9631-9643.

Billingsley, P. (1886), *Probability and Measure*, 2'nd Edition, New York: John Wiley.

Braud, I., J. D. Creutin, and C. Barancourt (1993), "The relationship between the mean areal rainfall and the fractional area where it rains above a given threshold," *J. Appl. Meteorology.*, 32, pp. 193-202.

Chiu, L.S. (1988), "Rain estimation from satellites: Areal rainfall-rain area relations," 3rd Conf. on Satellite Meteorology and Oceanography, American Meteorological Society, Jan. 31-Feb.4, 1988, Anaheim, CA.

Doneaud, A.A., S.I. Niscov, D.L. Priegnitz and P.L. Smith (1984), "The area-time integral as an indicator for convective rain volumes," *Jour. Climat. Appl. Meteorology*, 23, 555-561.

Eltahir, E.A.B. and R.L. Bras (1993), "Estimation of the fractional coverage of rainfall in climate models," *Journal of Climate*, 6, pp. 639-644.

Gupta, V.K. and E.C. Waymire (1993), "A statistical analysis of mesoscale rainfall as a random cascade," *J. Appl. Meteorology.*, 32, pp. 251-267.

Ha, E. (1992), *Analysis of Spatial Autoregressive Processes and Rain Rate Estimation*, Doctoral Dissertation, Department of Statistics, Texas A&M University, College Station.

Houze, R. A. and C.-P. Cheng (1977), "Radar characteristics of tropical convection observed during GATE," *Mon. Weather Rev.*, 105, 964-980.

Hudlow, M.D. and V. Patterson (1979), *GATE Radar Rainfall Atlas*, NOAA Special Report, NOAA.

Hudlow, M.D. and W.D. Scherer (1975), "Precipitation analysis for BOMEX period III," NOAA Tec. Report EDS 13, (NTIS PB246870/AS).

Kayano, K. and K. Shimizu (1993), "Optimal thresholds for a mixture of log-normal distributions as the continuous part of the mixed distribution." Submitted.

Kedem, B. and L.S. Chiu (1987), "On the lognormality of rain rate," *Proc. National Acad. of Sci., USA*, 84, pp. 901-905.

Kedem, B. and H. Pavlopoulos (1991), "On the threshold method for rainfall estimation: Choosing the optimal threshold level," *Jour. Amer. Stat. Assoc.*, 86, 626-633.

Kedem, B., and D.A. Short (1989), " Linear relationships for rain rate estimation from space," in *Precipitation Measurement*, B. Sevruk, ed., WMO/IAHS/ETH Workshop on Precipitation Measurement, St. Moritz, December 3-7, 1989, pp. 441-445.

Kedem, B., L.S. Chiu and Z. Karni (1990a), "An analysis of the threshold method for measuring area average rainfall," *J. Appl. Meteorology.*, 29, 3-20.

Kedem, B., L.S. Chiu and G.R. North (1990b), "Estimation of mean rain rate: Application to satellite observations," *Jour. Geophysical Research.*, 95, No. D2, 1965-1972.

Krajewski, W.F., M.L. Morrissey, J.A. Smith, and D.T. Rexroth (1992), "The accuracy of the area-threshold method: A model based simulation study," *J. Appl. Meteorology.*, 31, pp. 1396-1406.

LeCam, L. (1961), "A stochastic description of precipitation," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. III, Univ. Calif. Press, Berkeley, CA, 165-186.

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

Lovejoy, S. and G.L. Austin (1979), "The sources of error in rain amount estimating schemes from GOES visible and IR data," *Monthly Weather Review*, 107, pp. 1048-1054.

Lovejoy, S. and D. Schertzer (1985): Generalized scale invariance in the atmosphere and fractal models of rain, *Water Resour. Res.*, 21, 1233-1250.

Meneghini, R. and J.A. Jones (1993): An approach to estimate the areal rain rate distribution from space borne radar by the use of multiple thresholds, *J. Appl. Meteorology*, 32, 386-398.

Mintz, Y. (1981), "A Brief Review of the Present Status of Global Precipitation Estimates," in *Workshop Report on Precipitation Measurements From Space*, D. Atlas and O. Thiele Eds., NASA/GSFC, October, 1981, pp. D1-D4.

Pavlopoulos, H. (1991), *Statistical Inference for Optimal Thresholds*, Doctoral Dissertation, Department of Mathematics, University of Maryland, College Park.

Pavlopoulos, H. and B. Kedem (1992), "Stochastic modeling of rain rate processes: A diffusion model," *Stochastic Models*, 8, 397-420.

Richards, F. and P. Arkin (1981), "On the relationship between satellite-observed cloud cover and precipitation," *Monthly Weather Review*, 109, pp. 1081-1093.

Rosenfeld, D., D. Atlas, and D. Short (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.

Shimizu, K., D.A. Short, and B. Kedem (1993), "Quadratic relationship between the area rain rate variance and the fractional area exceeding a fixed threshold." Submitted.

Short, D.A., K. Shimizu, and B. Kedem (1993a), "Optimal thresholds for the estimation of area rain-rate moments by the threshold method," *J. Appl. Meteor.*, 32, pp. 182-192.

Short, D.A., D.B Wolff, D. Rosenfeld and D. Atlas (1993b), "A study of the threshold method utilizing rain-gauge data," *J. Appl. Meteor.*, 32, pp. 1379-1387.

Simpson, J. (1988), *TRMM, A satellite mission to measure tropical rainfall*, Report of the Science Steering Group, NASA.

Simpson, J, R.F. Adler, and G.R. North, (1988), "A proposed rainfall measuring mission satellite," *Bulletin Of the American Meteorological Society*, 69, 278-295.

Supan, A. (1898), "Die Jährlichen Niederschlagsmengen auf den Meeren, *Petermanns Geographische Mitteilungen*, 44, pp. 179-182.

B. Kedem
Dept. of Mathematics
Institute for Systems Research
University of Maryland
College Park, MD 20742
USA

D. A. Short
Laboratory for Atmospheres
NASA / Goddard FSC
Greenbelt, MD 20771
USA