Typology of rainfall fields to improve rainfall estimation in the Sahel by the area threshold method

Abou Amani,[1,2] Thierry Lebel,[3] Jean Rousselle,[1] and Jean Denis Taupin[4]

Abstract. The stratification of rainfall fields to improve specific rainfall models is a subject that has received relatively little attention in the literature. It is shown here that objective classification techniques, based on the intensities and spatial distribution of the rainfall fields, can produce meaningful results in terms of the area threshold method (ATM) model and climatology. Four approaches for rainfall classification, using rain gauge data, are proposed in order to improve the average areal rainfall estimation in the Sahel by the ATM model. Two of them are based on the structural behavior of the rainy area (area where it rains above a given threshold) function against a threshold. Based on this function, a new parameter, called the under profile area (UPA), has been proposed for the classification of rainfall fields. The groups obtained from the method based on this parameter are characterized by different average spatial structures. A significant improvement on the ATM model is observed by considering classification based on the UPA parameter. An average reduction of 34% of the root-mean-square error is observed in a validation term. This improvement is a direct consequence of the fact that the optimal thresholds are different from one group to another, which is an important point when considering the impact of classification on the ATM model.

1. Why a Typology of Rainfall Fields?

For a large majority of the rainfall models presented in the literature, all the available rainfall data are used in the modeling process without any stratification or, at best, with a crude separation based on some elementary characteristics of the rainfall field (its space-averaged value, for instance). Yet it is well known that rainfall data originate from different types of rainstorms, and *Houze* [1981] has shown that a separation between stratiform and convective rainfall is possible for many different precipitation systems. On a synoptic scale, meteorologists have developed the concept of weather type, which has proved efficient in relating the rainfall features on a large scale with some relevant meteorological variables [e.g., *Benichou et al.*, 1988]. However, in many cases it is very difficult, if not impossible, to identify the meteorological nature of a rainfall event in a limited area of study, the more so if only inappropriate meteorological data are available. Furthermore, on smaller scales, which are of interest to hydrologists, relationships between the atmospheric circulation and rainfall characteristics on the ground are much more difficult to find. This is partly due to the lack of proper meteorological data and partly due to the complexity of rainfall distribution in time and space when viewed at a resolution of, say, a few kilometers and a few hours (or less). On such scales, two events triggered by the same meteorological situation may have a different ground signature.

Rainfall stratification is especially useful for satellite-based rainfall estimation algorithms and climatology. For instance, the lower accuracy in satellite estimates during phase 2 of the GARP (Global Atmospheric Research Program) Atmospheric Tropical Experiment (GATE) was attributed by *Woodley et al.* [1980] to the presence of a higher proportion of a different kind of precipitating convective system as compared to phases 1 and 3. In their review of the grid-cell approach for satellite rain estimation over Florida, *Negri and Adler* [1987] felt the necessity to group the data into classes. However, the results of the proposed stratification technique, based on the mean rain rate, were inconsistent, the variability of the mean infrared temperature within each rain rate class being often as great as the differences between rain classes. The authors implied that other classification schemes should be designed. *Desbois et al.* [1988] point out that in the Sahel, a distinction has to be made between local convection and mesoscale convection.

Given the theoretical and practical difficulties of implementing stratification schemes based on a meteorological analysis, it appears appropriate to derive alternate schemes based on the rainfall data set only, since these schemes are the most readily available to users. Since the possible uses of a rainfall typology are numerous, the choice of a proper classification procedure will have to be based on the aims of this classification. Stratification of rainfall events in their stratiform and convective parts is considered to be very important in rainfall estimation by satellite remote sensing. Also, the separation of rainfall events having different statistics (namely mean and variance) is necessary when using methods based on these statistics (satellite rainfall estimation methods). One of such methods widely used is the area threshold method (ATM). The ATM model [*Short et al.*, 1993; *Rosenfeld et al.*, 1990], is derived from the

[1]Département de Génie Civil, Ecole Polytechnique, Montreal, Quebec, Canada.
[2]Now at Groupe Précipitations en Afrique de l'Ouest, Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM), Grenoble, France.
[3]Groupe Précipitations en Afrique de l'Ouest, Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM), Grenoble, France.
[4]Groupe Précipitations en Afrique de l'Ouest, Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM), Niamey, Niger.

Paper number 96WR01100.
[illegible]9.00
ISSN 0043-1397

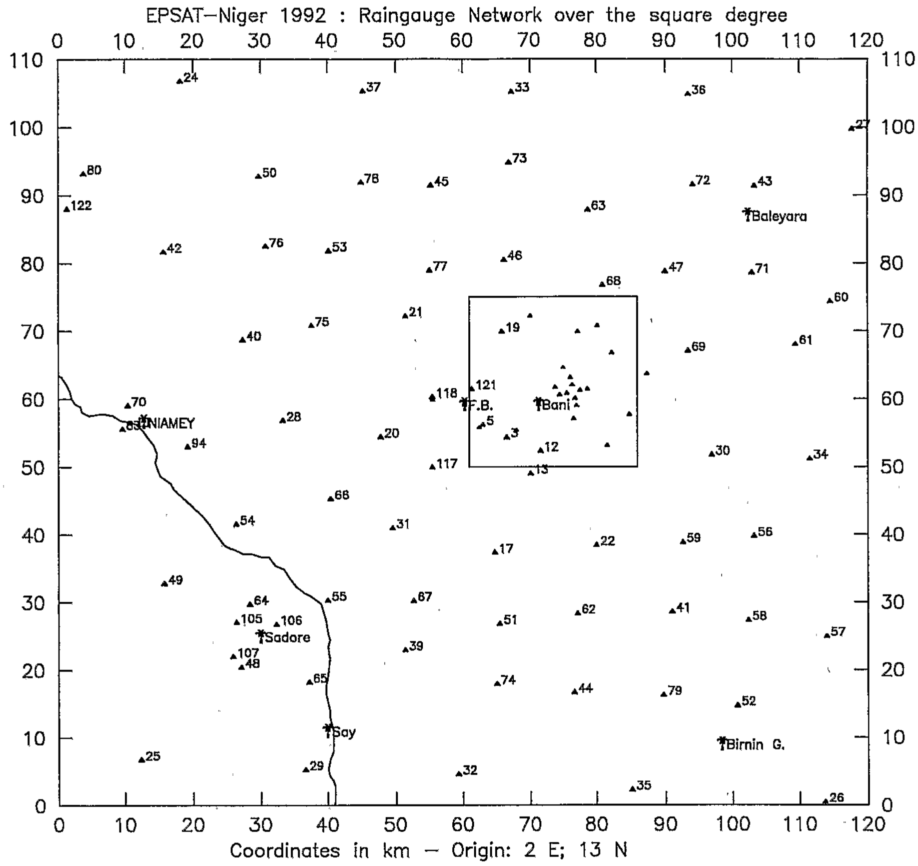


Figure 1. The satellite estimation of precipitation, Niger experiment (EPSAT-NIGER) recording rain gauge network, 1992. The square in the center is the 400-km^2 target area, and the solid line represents the Niger River.

area time integral (ATI) model, first proposed by *Doneaud et al.* [1981] for the estimation of convective rain volume for a given rainfall event. In this study we are interested in classification schemes based on rain gauge data in order to (1) improve the areal rainfall depth estimation by the ATM model and (2) improve our understanding of the Sahelian rainfall climatology.

The topics covered in this paper are presented in the following order: geographical location and data set of the study, review of possible classification criteria, classification methods relevant to rainfall estimation by the threshold method, applications of the classification methods and their contribution for improving Sahelian precipitation estimation by the ATM model, and the conclusion.

2. A Sahelian Case Study

Estimating rainfall over the Sahel (the semiarid region lying south of the Sahara), where the networks of measurement stations are not highly developed, is necessarily more difficult than in a temperate region. This is because of the intermittent nature, in time and space, of the data they provide. It is of vital importance in this setting therefore to combine direct rainfall measurements (ground-based rainfall data) and data obtained through remote sensing. However, the performance of the algorithms designed to integrate these raw data is a function of the homogeneity of the data. In order to test and improve these algorithms, the satellite estimation of precipitation, Niger, experiment (EPSAT-NIGER) [*Lebel et al.*, 1992] was conducted in the Niamey region in which data were obtained through very accurate sampling of rainfall systems by means of a dense network of rainfall gauges and by radar (Figure 1). At this time the area under study, measuring 120 × 110 km, was equipped with about 100 static-memory rain gauges. This network consisted of a basic network of rain gauges installed on a regular grid and spaced approximately 12.5 km apart, at the center of which was a denser network covering an area of 400 km^2.

The various rainfall events observed in the Sahel originate mainly from three types of meteorological systems: isolated convective systems, organized moving convective systems, and squall lines, the latter two being grouped together under the heading "mesoscale convective systems" (MCS). The amounts of precipitation recorded during a rainfall event, whatever its origin, are extremely variable in space, however. The rainfall regime includes a dry season (October–April) and then a rainy season (May–September). The rainfall events observed during the EPSAT-NIGER experiment have been defined as follows: (1) a minimum of 30% of operating rain gauges have recorded more than 2.5 mm of rainfall, and (2) a minimum of 30 min must separate two consecutive events. Respectively, 37, 47, and 49 such events were recorded in 1990, 1991, and 1992 [*Lebel et al.*, 1995], totaling more than 95% of the seasonal rainfall (the remaining rainfall, less than 5%, is produced by isolated showers at the beginning and end of the rainy season).

Another feature of rainfall in this region is that the average annual rainfall is strongly correlated to the number of events during which rainfall accumulation exceeds a given threshold [*Taupin et al.*, 1993]. Recently, *LeBarbé and Lebel* [1996] pointed out that seasonal rainfall depth (week, month, or year) in the Sahel can be easily estimated by the product of climatological mean rainfall depth per event and the number of recorded events. This observation can be improved if we can distinguish between the different contributions in number and total rainfall accumulation of the different types of rainfall events. Also, a robust method stratifying the Sahelian rainfall events can contribute to the improvement of the climatological diagnostics for the rainy season. A major field of interest for classification is to improve areal rainfall estimation models such as the ATM, widely proposed for rainfall estimation from remote-sensing data. Since the Sahelian rainfall is produced by a few different precipitation systems, it would be helpful to classify rainfall fields with the objective of improving the ATM model in the Sahel. An objective meteorological stratification of rainfall events in the Sahel is difficult to implement because of the lack of pertinent meteorological data and the soft gradient of Sahelian pressure fields. Therefore classification based on rain gauge data is very attractive. Since it is based on rainfall characteristics observed at the ground level, the classification can also be useful for other specialists interested in water resources in the Sahel.

3. Stratification Based on Rain Gauge Data: A Review of Possible Criteria

Since no widely accepted stratification method is available, several approaches, based on parameters characterizing the spatial or spatiotemporal features of the rainfall events, will have to be compared. These parameters may be directly connected to the theoretical assumptions of the estimation model(s) of interest, or we can define a priori certain parameters characterizing the basic structure of the rainfall field but with no direct connection with the estimation model(s).

For rainfall field estimation models the classification criteria must be necessarily linked to the spatial or spatiotemporal behavior of the rainfall fields. *Taupin et al.* [1993] and *Lebel et al.* [1995, 1996] have described the high spatial variability of the Sahelian rainfall at the event and annual scales. For hydrological and agronomical studies in the Sahel this variability must be considered. Thus a natural classification criterion for rainfall fields can be the spatial structure of the total rainfall depths of the event. Various attempts carried out by *Amani* [1995] to stratify the Sahelian rainfields according to such statistical parameters as the variogram parameters or the coefficient of variation of the rainfields proved to be not very efficient. We were thus led to consider alternate approaches.

Considering the spatiotemporal behavior of rainfall fields, the procedure proposed by *Kottegoda and Kassim* [1991] for the classification of temporal rainfall structures (punctual hyetographs) was generalized in space-time (areal hyetographs) by *Amani et al.* [1993]. The classification based on these approaches has two main drawbacks: (1) the computation of the rainfall hyetograph type depends on the time step duration of the hyetograph used for the analysis [*Amani et al.*, 1993], and (2) it cannot account for the rainfall distribution in space. An extension of these methods to describe the spatial structure of the rainfall as measured by a rain gauge network is therefore

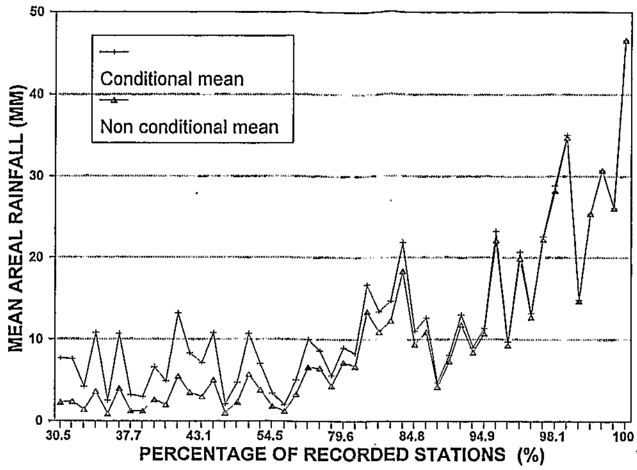


Figure 2. Event rainfall versus proportion of zero rainfall for the events of 1992. Shown are conditional (crosses) and nonconditional (triangles) mean areal rainfall of events over the whole network area.

proposed in the next section as a classification scheme directly related to the ATM method.

Two important characteristics to be considered in the classification schemes are the intermittent nature and intensity of rainfall fields in the Sahel. Even though intermittent rainfall events are generally weak events, we can observe from Figure 2 that moderate events can be nonintermittent events. As we can see from this figure, there is a significant difference between the conditional and the nonconditional mean areal rainfall depths for very intermittent events. A good classification method for Sahelian rainfall must consequently have a separation of events based on their degree of intermittency (proportion of zero-rain area inside the area of study) or their intensity or both.

4. Selection of Methods Relevant to Rainfall Estimation by the ATM Algorithms

Classification methods capable of improving rainfall field estimation by the ATM model must necessarily be based on parameters characterizing the spatial structure of rainfall fields in terms of rainfall accumulation and/or their areal extent. Several classification methods are proposed and compared by *Amani* [1995]. The comparison presented by *Amani* [1995] led us to select three classification methods which are better and more robust than the others. They are method 1, based on the clustering analysis of parameters characterizing the rainfall event intensity; method 2, based on a crossing analysis of the rainy area function (area where it rains above a given threshold) versus threshold (this is an extension in space of the method proposed by *Kottegoda and Kassim* [1991] for the classification of the hyetograph of rainfall events observed at a given rain gauge); and method 3, based on the analysis of the cumulative distribution function of the newly introduced under profile area (UPA) parameter. The UPA is defined as the area under the rainy area function obtained by method 2. Also, the two criteria, the UPA parameter and the parameters used in method 1, are combined to produce a fourth method.

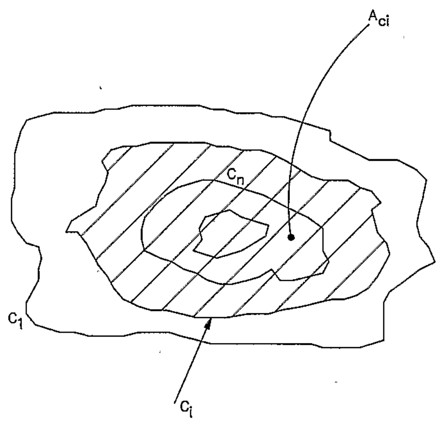


Figure 3. Definition of the rainy area above a threshold (C_i is the ith threshold value, and A_{Ci} is the rainy area where it rains above C_i).

4.1. Method 1

For a rainfall event the maximum total rainfall $C_{M\theta}$ observed over the whole rain gauge network for various periods θ of accumulation are computed:

$$C_{M\theta} = \max\,[R_\theta(i)] \qquad i = 1, n \tag{1}$$

where $R_\theta(i)$ is the maximum rainfall accumulation recorded at rain gauge i for accumulation period θ. Usually, not all the maxima are recorded on the same gauge. This set of maximum values is thus a global indicator of the rainfall pattern over the area of study. Here, five durations θ were chosen: 5, 10, 15, 30, and 60 min. Other meaningful parameters may also be considered, such as total event recorded depth, mean recorded depth, and event duration. The five parameters ($C_{M\theta}$ for θ = 5, 10, 15, 30, and 60 min) retained here proved to be the most significant for application to Sahelian rainfall.

4.2. Method 2

With this method, we propose to use the profile of the rainy area function as a classification parameter. The rainy area is the area over which the rain rate exceeds a given threshold. It is scaled by the total area of the study. Figures 3 and 4 illustrate the rainy-area approach. Let C be a rainfall threshold; then define AC as the percentage of the rainy area where rainfall exceeds this threshold C. AC is estimated from the recorded spatial data using the indicator technique, where an indicator variable is computed as

$$\begin{aligned} I(x, y, t) &= 1 \qquad R(x, y, t) >= C \\ I(x, y, t) &= 0 \qquad \text{otherwise} \end{aligned} \tag{2}$$

The value of AC is then given by

$$AC = E[I(x, y, t)] \tag{3}$$

In practice, AC is calculated for a rainfall field accumulation over a time period θ (θ may be a fixed time step, say 1 hour, or the event duration as considered in this study). The corresponding AC function is denoted by AC_θ and estimated as

$$AC_\theta = \frac{1}{n}\sum_{i=1}^{n} I_\theta(x_i, y_i) \tag{4}$$

where $I_\theta(x_i, y_i)$ is the indicator variable associated with gauge i for the accumulation period θ. AC_θ is a decreasing function of C.

The complementary function $BC = (1 - AC)$ is the area of rainfall below C. In order to compare the BC functions for different events, the threshold axis must be scaled by a scaling parameter C_M

$$C_M = \max\,[R_\theta(i)] \qquad i = 1, n \tag{5}$$

The threshold variate then becomes

$$C^* = C/C_M \tag{6}$$

BC is a function analogous in time to a mass curve. By following the procedure applied by *Kottegoda and Kassim* [1991] to the hyetograph analysis, it is possible to count the number of intersections between the bisector and the observed BC curve, which produces a direct classification (Figure 5). The basic types of the classification are numbered 1a, 1b, 2a, 2b, and so on. The order number is defined as follows: if x is the number of intersections between the mass curve and the bisector excluding the two end points, then the order of the mass curve is given by $(x + 1)$. If the part of the mass curve before the first intersection is under the bisector, then the class of the mass curve is a; otherwise it is b.

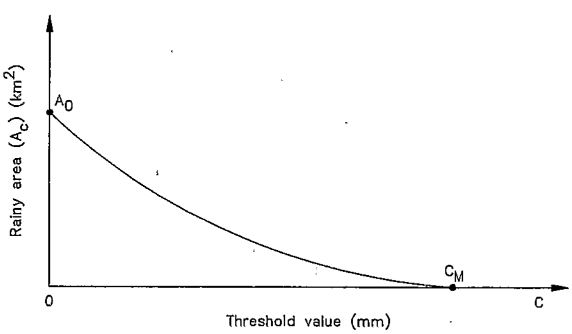


Figure 4. Rainy area (area where it rains above a given threshold) function of the threshold. C_M is the maximum rainfall depth and A_0 is the total area of the network.

4.3. Method 3

This method is based on the BC function defined above. The area under the BC function (UPA) is computed (Figure 6). Since both axes are scaled, UPA takes values between 0 and 1. The value 0 corresponds to a spatial uniform rainfall and the value 1 corresponds to a Dirac-like event. It is demonstrated in section 4.4 that the UPA parameter is related to the coefficient of variation under certain assumptions. It is shown in section 5 that the cumulative distribution of the UPA values obtained from the 133 Sahelian rainfall events leads to a natural classification in three groups. The UPA parameter has some interesting statistical meaning, as will be shown below.

4.4. Links Between the UPA Parameter and CV

First, let us consider the relationship between the rainy area AC and the threshold C. Studying this relationship for each event, it appears that log (AC) is linearly correlated with C. In 1991, only nine events, out of a total of 47, have a coefficient of determination R^2 under 0.90, and 26 events have an R^2 above 0.95. The results for 1990 and 1992 are even better, with, respectively, five events (out of a total of 37) and four events (out of a total of 49) with an R^2 under 0.90, and 20 events (1990) and 37 events (1992) with an R^2 above 0.95. For a given event j, we may thus write

$$AC_j = \exp\,(-C/a_j) \tag{7}$$

It should be noted that the empirical justification of the expression (7) is backed by the finding of *Lebel et al.* [1996] that the cumulative distribution function (cdf) of the event rainfall is exponential. Since there is an obvious, even if not straightforward, relationship between the exceedance probability of a given rainfall event value Z_j and the area a_j corresponding to this value, it is not surprising that the cdf and area distribution function are of the same functional type. Using the scaled threshold C^*, defined in expression (6), (7) may be rewritten:

$$AC_j = \exp\,[-C^*(C_{Mj}/a_j)] \tag{8}$$

where C_{Mj} is the recorded maximum of $R_T(x, y)$ for event j. The scaled spatial profile BC_j is then expressed as

$$BC_j = 1 - \exp\,[-C^*(C_{Mj}/a_j)] \tag{9}$$

BC_j, C_{Mj}, and a_j are all realizations of random variables associated with the random process $R_T(x, y)$. This process is defined by its probability density function, with mean μ_T, standard deviation σ_T, and coefficient of variation CV. From the frequency analysis point of view the maximum rainfall recorded by a network of n independent gauges is also a random variable corresponding to the quantile Q_n. The expected probability of nonexceedance of Q_n, P_n, is given by $(n - a)/(n + b)$, with the a and b parameters taking values between 0 and 1 (the values of a and b required to obtain an unbiased estimate of P_n depend on the probability distribution of $R_\theta(x, y)$; see *Cunnane* [1978] for a review). Q_n may be expressed as

$$Q_n = \mu + \lambda\sigma \tag{10}$$

where λ is a statistic which depends on P_n and on the probability distribution of $R_T(x, y)$. For a given realization (rain event) j, this expression yields

$$C_{Mj} = m_j + l_j s_j \tag{11}$$

where m_j and s_j are the computed mean and standard deviation of the rainfall recorded by the network for the event j; l_j

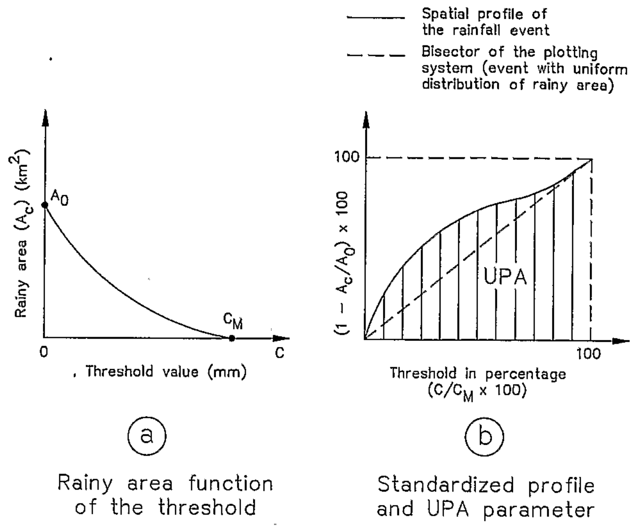


Figure 6. Illustration of the UPA parameter and spatial crossing approach. (a) Rainy area versus threshold and (b) spatial profile of the rainfall event and definition of the under profile area (UPA) parameter which is the integral of the spatial profile.

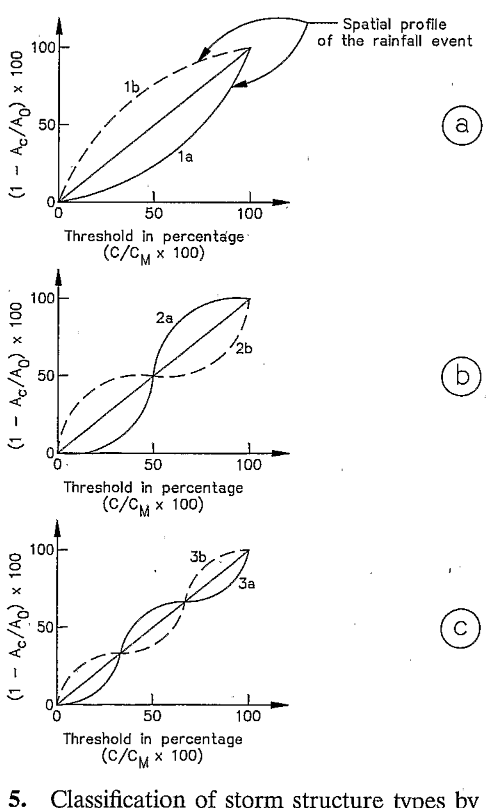


Figure 5. Classification of storm structure types by method 2: (a) types 1b and 1a, (b) types 2b and 2a, and (c) types 3b and 3a. A_0 and A_C are the total network area and the area where it rains above the threshold value C, respectively, and C_M is the maximum rainfall depth recorded through the network. Reprinted from *Journal of Hydrology*, vol. 127, no. 1/4, N. T. Kottegoda and A. H. M. Kassim, Classification of storm profiles using crossing properties, pp. 37–53, 1991, with kind permission of Elsevier Science—NL, Sara Burgerhartstraat 25, 1055 KV Amsterdam, The Netherlands.

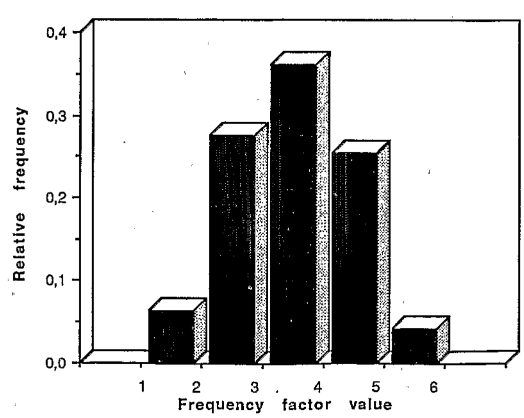


Figure 7. Histogram of the frequency factor (l) values corresponding to the event maximum rainfall depth C_M (3 years combined).

is the frequency factor of the maximum cumulated rainfall and it is linked to the return period of C_M recorded during event j. Substituting (11) into (9) yields

$$BC_j = 1 - \exp\,[-C^*(m_j + l_j s_j)/a_j] \tag{12}$$

For the area under the scaled spatial profile, UPA_j, being the integral of BC for C^* varying from 0 to 1, we get

$$\mathrm{UPA}_j = \int_0^1 1 - \exp\,[-C^*(m_j + l_j s_j)/a_j]\, dC^* \tag{13}$$

Defining

$$K_j = (m_j + l_j s_j)/a_j \tag{14}$$

leads to

$$UPA_j = 1 - [-1/K_j \exp\,(-K_j C^*)]_0^1 \tag{15}$$

or

$$UPA_j = 1 - 1/K_j + \exp\,(-K_j)/K_j \tag{16}$$

Computing the values a_j and m_j for all the available events j shows that m_j may be used as an estimate of a_j, which means that (14) may be written simply as

$$K_j = 1 + l_j CV_j \tag{17}$$

Furthermore, the values of CV_j and l_j show that $l_j CV_j$ is almost always greater than 1 and most often greater than 2. Figure 7 presents the histogram of the frequency factor values l.

Consequently, $\exp\,(-K_j)$ is small compared to 1, and the third term in the summation of (16) may be neglected, so that (16) becomes

$$UPA_j = l_j CV_j/(1 + l_j CV_j) \tag{18}$$

Using a Taylor second-order development, (18) may be written as

Table 1. Parameters of the Linear Regression of UPA Against $1/CV$ and $(1/CV)^2$

Year	R^2	a_0	a_1	a_2	l_1	l_2	n
1990	0.753	1.10	−0.35	0.036	2.8	3.7	37
1991	0.873	1.02	−0.34	0.046	2.8	3.3	47
1992	0.813	1.01	−0.29	0.036	3.5	3.7	49

$UPA = a_0 + a_1/CV + a_2/(CV)^2$; l_1 and l_2 are the estimates of l computed respectively as $l = 1/a_1$ and $l = 1/(2a_2)^{1/2}$ (see expression (19)). R^2 is the coefficient of determination and n is the sample size (number of rainfall events).

$$UPA_j \approx 1 - \frac{1}{l_j CV_j} + \frac{1}{2(l_j CV_j)^2} \tag{19}$$

It is possible to verify the theoretical expression (19) by regressing UPA on $1/CV$ and $(1/CV)^2$. The results are given in Table 1 for all 3 years. It can be seen that depending on the year considered, between 75 and 90% of the variance of UPA is explained by the variance of CV. Thus the UPA distribution is related primarily to that of CV, which is a measure of the global spatial variability of a given rainfield, and also to that of the variable l, which appears to be related to the overall strength of a rainfall event. This explains (at least partly) two findings of our comparison between classification methods: one, not presented here, is that differences were found between the UPA classification and a classification based on the CV only [*Amani*, 1995], the UPA classification performing better; the other (presented below) is that the combination between methods 1 and 2 did not improve the performances obtained with method 3 alone (the information provided by method 1 is essentially related to the strength of a rainfall event).

5. Stratification of Sahelian Event Rain Fields

In order to evaluate the impact of classification on the spatial estimation of rainfall by the ATM model, rainfall data

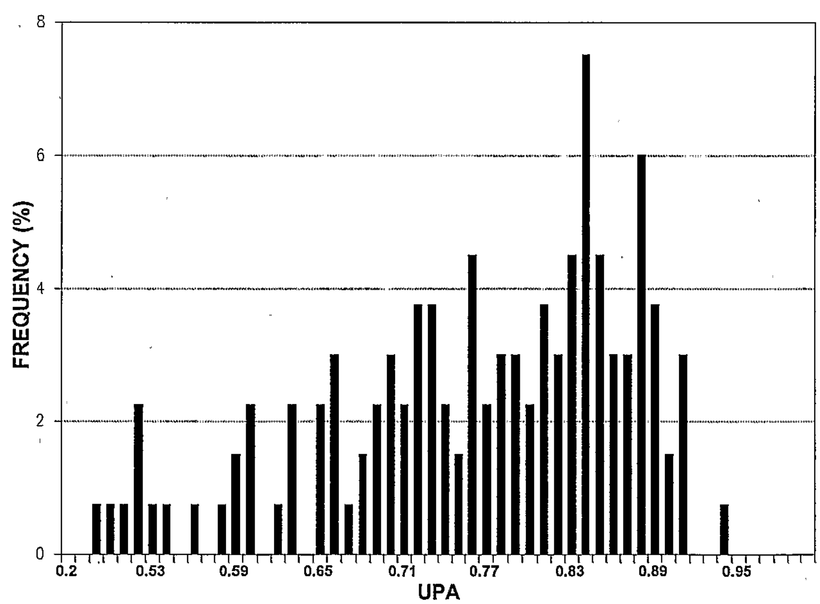


Figure 8. Histogram of the UPA values (3 years combined).

Table 2. Parameters and Group Size for Each Significant Group Obtained From Each of the Four Methods for Seasons 1990, 1991, and 1992

Methods	Group	Group Size 1990	1991	1992	Parameters
1	1	15	15	24	rainfall intensities
	2	19	15	22	
2	1b	32	39	35	rainy area function
	2a, 2b	5	8	12	
3	1	6	10	5	UPA defined from rainy area function
	2	13	21	21	
	3	18	16	23	
4	1	27	23	29	UPA and rainfall intensities
	2	10	20	12	

collected in connection with the EPSAT-NIGER experiment during the years 1990, 1991, and 1992 are used. The four classification schemes presented in section 3; methods 1, 2, 3, and 4 are applied to the stratification of each of the three event rainfall samples of 1990, 1991, and 1992; samples of 1990 and 1991 combined; and the 3 years are combined. The histogram of the UPA values (133 events recorded during the EPSAT-NIGER experiment) is presented in Figure 8. This figure clearly leads to a natural classification of rainfall fields into three distinct groups: group 1, with UPA values less than 0.66; group 2, with UPA values between 0.66 and 0.81; and group 3, with UPA values greater than 0.81. This systematic classification is called the distribution of the UPA parameter (DUPA), or method 3. Thus two methods are systematic (methods 2 and 3), and the other two (methods 1 and 4) are based on a clustering analysis. The commercial software *Statistical Analysis System* (*SAS*) [1982], version 5.18, was used to perform the clustering. A review of clustering techniques may be found in work by *MacQueen* [1967] and *Anderberg* [1973].

Table 2 presents the size of the significant groups for each method. Method 1 leads to two groups having similar sizes. Group 1 is made of the stronger events, while group 2 is made of the less intense events. For method 2, spatial crossing, the grouping of the event is given by the order of its spatial profile function (rainy area against threshold) when the crossing technique is applied. The majority of the events belong to group 1b. The other events, belonging to the other orders (2a or 2b), are characterized by a large spatial extension. The three groups given by method 3 are characterized by different degrees of intermittency. The events of group 1 are large spatial events, while those of group 3 are highly intermittent.

Major points to consider when estimating areal rainfall are the continuity and stationarity of the field. If the field may be assumed to be continuous and stationary in space, several statistical methods, most notably those of the best linear unbiased estimator (BLUE) family, are available. On the other hand, when the intermittency becomes a significant aspect of the field under consideration, it has to be taken into account, as proposed by *Braud et al.* [1993], for instance. In such cases, two parameters play a key role in the areal estimation process: the mean areal rainfall (AR) over the rainy areas, which is a characteristic of the event magnitude, and the percentage of gauges affected by rainfall (PGA), which characterizes the spatial extension of the event. Methods 2 and 3, based on the spatial behavior of the events, tend to classify the rainfall events according to their PGA and their AR, whereas method 1 does so principally according to their AR [*Amani*, 1995].

Three important statistics of the obtained groups can be used to analyze the potential of the classification methods to improve rainfall estimation by the ATM model. They are the mean average areal rainfall, the mean standard deviation, and the mean coefficient of variation. These statistics are presented for the three groups obtained by method 3 (DUPA) in Table 3 as an example. It appears that both the nonconditional and conditional statistics are significantly different from one group to another. The probability of nonzero rainfall may be estimated as the ratio between the nonconditional average (including zero rainfall) and the conditional nonzero average, that is, 97% for group 1, 85% for group 2, and 63% for group 3. Since both CV and the probability of nonzero rainfall are very different from one group to the other, this necessarily has an effect on the ATM model where, usually, a unique conditional distribution function of rain rate is considered.

Also, to point out the possible positive impact of these classification methods on improving the ATM model, the variance of the conditional mean rainfall depth for each group is computed and compared with the variance of the conditional mean rainfall depth (conditional to the threshold considered) of the sample for each year. A reduction in the variance of the conditional average rainfall depth is observed for the majority of groups. Figure 9a illustrates the reduction in the variance of the conditional rain depth against the threshold for method 3 for the 1990 sample. The conditional mean rain depth is also given as a function of the threshold in Figure 9b. Table 3 and Figure 9 display significant differences between the statistics of each group of method 3. Similar results were obtained with the other methods, which leads one to expect an improvement in the ATM model by considering the classification procedure.

6. Comparison of Performance of the ATM Event Rain Depth Estimates With and Without Stratification

The four classification methods are considered now in order to evaluate their impact on the ATM estimates of areal rainfall depth. The ATM model is therefore applied before and after classification, and the resulting errors are compared. These errors are evaluated by considering the optimal ATM model for each group.

6.1. Review of the ATM

The ATM model, initially developed in the meteorological domain for estimating average rainfall based on satellite images and radar, is applied here to the EPSAT-NIGER data to estimate average event rainfall. The ATM model is based on the existence of a strong linear correlation between the average rainfall over an area and the fractional area where the rainfall amounts exceed a given threshold. This relation was first observed empirically by *Doneaud et al.* [1981, 1984]. Thus for a threshold rainfall C the mean rainfall $m(t)$ at time t on the surface under study is given by

$$m(t) = S(c)F(t, c) + b(c) \tag{20}$$

where $F(t, c)$ is the fractional area where the rainfall recorded at time t exceeds threshold C. The parameters $S(c)$ and $b(c)$ are coefficients obtained through linear regression between the

Table 3. Statistics and Contributions of the Different Groups of Method 3 to the Total Rainfall Accumulation in 1990, 1991, and 1992

	Group											
	1				2				3			
Sample	Average Rainfall Depth, mm	Average Standard Deviation, mm	Average Coefficient of Variation	Statistics Defining the Importance of Each Group*	Average Rainfall Depth, mm	Average Standard Deviation, mm	Average Coefficient of Variation	Statistics Defining the Importance of Each Group*	Average Rainfall Depth, mm	Average Standard Deviation, mm	Average Coefficient of Variation	Statistics Defining the Importance of Each Group*
						3 Years						
Nonconditional statistics	22.8	12.8	0.60	···	11.1	10.2	1.15	···	5.3	8.4	1.79	···
Conditional statistics	23.6	12.5	0.56	···	13.1	10.4	0.84	···	8.4	9.6	1.16	···
	···	···	···	21, 15.8, 34.7	···	···	···	55, 41.4, 43.7	···	···	···	57, 42.8, 21.6
						1990						
Nonconditional statistics	22.0	12.8	0.58	···	12.1	12.5	1.12	···	4.8	8.4	1.89	···
Conditional statistics	22.9	12.3	0.54	···	14.2	12.6	0.91	···	8.3	9.8	1.18	···
	···	···	···	6, 16.0, 35.0	···	···	···	13, 35.0, 42.0	···	···	···	18, 49.0, 23.0
						1991						
Nonconditional statistics	21.4	12.7	0.65	···	10.9	10.2	1.01	···	4.9	8.3	1.83	···
Conditional statistics	21.9	12.5	0.61	···	12.3	10.1	0.83	···	8.2	9.6	1.17	···
	···	···	···	10, 21.3, 41.0	···	···	···	21, 44.6, 43.9	···	···	···	16, 34.0, 15.1
						1992						
Nonconditional statistics	26.6	13.0	0.55	···	10.8	8.8	1.30	···	6.1	8.5	1.67	···
Conditional statistics	27.7	12.6	0.49	···	13.1	9.2	0.79	···	8.6	9.5	1.14	···
	···	···	···	5, 10.2, 26.6	···	···	···	21, 42.9, 45.4	···	···	···	23, 46.9, 28.0

*Sample size, percentage represented by the group's events, and percentage of the group's events' contribution to the total rainfall accumulation.

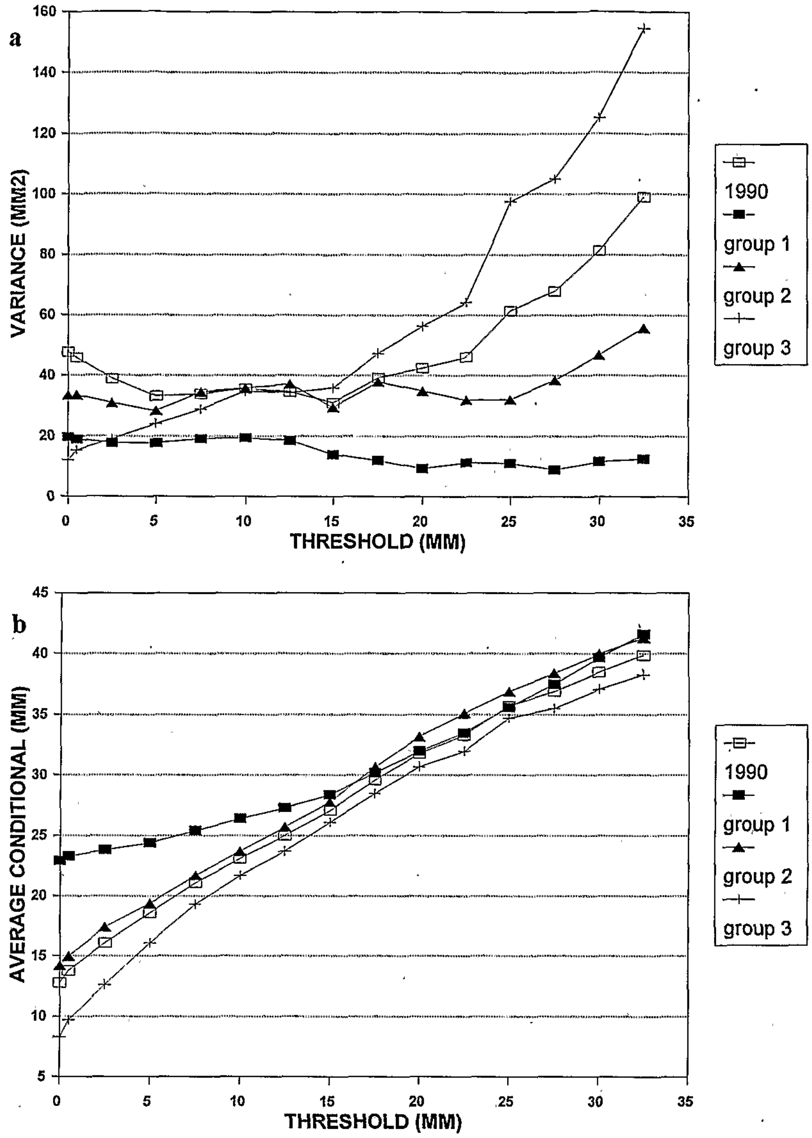


Figure 9. Conditional mean and variance of the areal rainfall depth for the events of 1991 before classification and after classification by the distribution under profile area (DUPA) method (all three groups together) against the threshold considered. (a) Variance and (b) average conditional rain depth.

average rainfall amounts for the various fields and the corresponding fractional rainfall area. These coefficients are a function of the threshold, while the average rainfall is independent of it. The threshold method is characterized by the fact that as the threshold value increases, so does the coefficient of determination R^2 between $m(t)$ and $F(t, c)$ until a maximum value for a certain threshold (optimum threshold) is reached, and then it decreases.

The theoretical framework of the threshold method has been recently revisited by several authors [*Kedem et al.*, 1990; *Atlas et al.*, 1990; *Rosenfeld et al.*, 1990]. *Kedem and Pavlopoulos* [1991] have proposed a statistical method for determining the optimum threshold, and *Braud et al.* [1993] have shown that under certain hypotheses, the theoretical correlation coefficient between $m(t)$ and $F(t, c)$ is a function of the spatial structure of the rainfall depth amounts.

The evaluation of the impact of classification is based on the calculation of a criterion measuring the error between the average rainfall estimated by the model over the whole area under study and the average rainfall recorded by the rain gauge network. The criterion selected here is the root-mean-square error (RMSE). In addition, the improvement brought about by the model is tested in calibration and validation modes. Given a classification method, the RMSE for each group and a global RMSE are computed.

6.2. Relation Between the Rainfall Surface and the Average Accumulated Rainfall for the Event

The method is applied here to estimate average rainfall accumulations during an event over the area under study (13,200 km^2). First, the method is applied without classification of rainfall fields. It is then applied separately for each of the significant groups obtained using the four classification methods. The following nine basic threshold values are considered: 0.5, 2, 5, 7.5, 10, 12.5, 15, 17.5, and 20 mm. For some groups, to reach the optimum threshold, additional threshold values

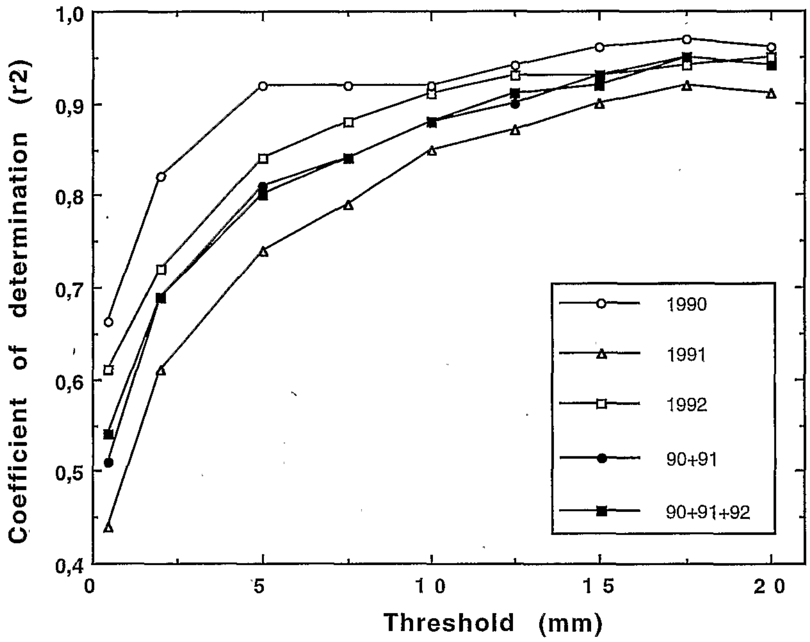


Figure 10. Coefficient of determination (R^2) of the ATM model for various thresholds for each year and combination of years.

are considered. For each of these values and for each rainfall field the fractional rainfall area (an area where the rainfall amounts exceed the fixed threshold value) is calculated using the indicator method. Let C be the value of the threshold and $R(i)$ the total amount of rainfall recorded at station i.

The indicator variable $I(i)$ is defined as

$$\begin{array}{ll} I(i) = 1 & R(i) >\ = C \\ I(i) = 0 & \text{otherwise} \end{array} \tag{21}$$

The fractional rainfall area for rainfall event k is estimated by

$$F(c, k) = \frac{1}{n} \sum_{i=1}^{i=n} I(i) \tag{22}$$

The average rainfall $m(k)$ of field k over the surface under study is calculated by the arithmetic mean, taking into account null values. For each set of rainfall events the estimation model is obtained by carrying out, for a given value of threshold C, a linear regression between the average rainfall values $m(k)$ and the values of the fractional rainfall area $F(c, k)$. Basic models are determined first, without classification, by calculating the correlations for each year separately (1990, 37 values; 1991, 47 values; and 1992, 49 values). More substantial samples are formed subsequently by adding the values for 1990 and 1991 (84 values) and then by including values for all 3 years (133 values). The evolution of the coefficient of determination R^2 as a function of the threshold for each sample is given in Figure 10. From the values of R^2, we see that the optimum threshold for each year is in the neighborhood of 17.5 mm.

6.3. ATM Model and Classification

The ATM is applied to each of the groups obtained by the four classification methods for each of the 3 years, for the years 1990 and 1991 combined, and for the 3 years combined. The most important point in the application of the ATM model for each group is that the optimum threshold and the coefficient of determination R^2 are different from one group to another for all the samples and classification methods considered. As an example, Figure 11 illustrates the variation in the coefficient of determination (regression equation characterizing the ATM model) versus the threshold for the three classification methods (methods 1, 2, and 3) when all 3 years are combined. This result can be expected from the analysis given in section 4, where it was pointed out that the statistics (mean, standard deviation, and coefficient of variation) of the various groups are different. The parameters of the optimal ATM model for methods 1, 2, and 3 are given in Table 4. Figure 12a illustrates the threshold model for the 3-year sample (1990–1992) before classification at the optimal threshold (17.5 mm). For method 3, Figures 12b, 12c, and 12d present also the optimal threshold models for the groups 1, 2, and 3, respectively.

Before stratification, the RMSE value of the optimal ATM model for the 1990–1992 sample was 2.24. After classification, by considering the optimal model for each group, the RMSE values are 1.93, 1.68, and 1.54 for methods 1, 2, and 3, respectively. In terms of percentages this corresponds to RMSE value reductions of 13.8%, 25%, and 31.3%, respectively, for the three methods. Thus classification improves significantly the ATM estimates in the calibration phase, particularly for method 3. However, the effect of the classification has to be evaluated in a validation mode, especially since the model before classification has a smaller number of parameters (two) than the models after classification (2 times the number of groups).

Two cases are considered for validation purposes. In the first case the mean areal rainfall of the 1992 events is reconstituted from the optimal ATM models calibrated by combining the 1990 and 1991 events. A fourth classification scheme combin-

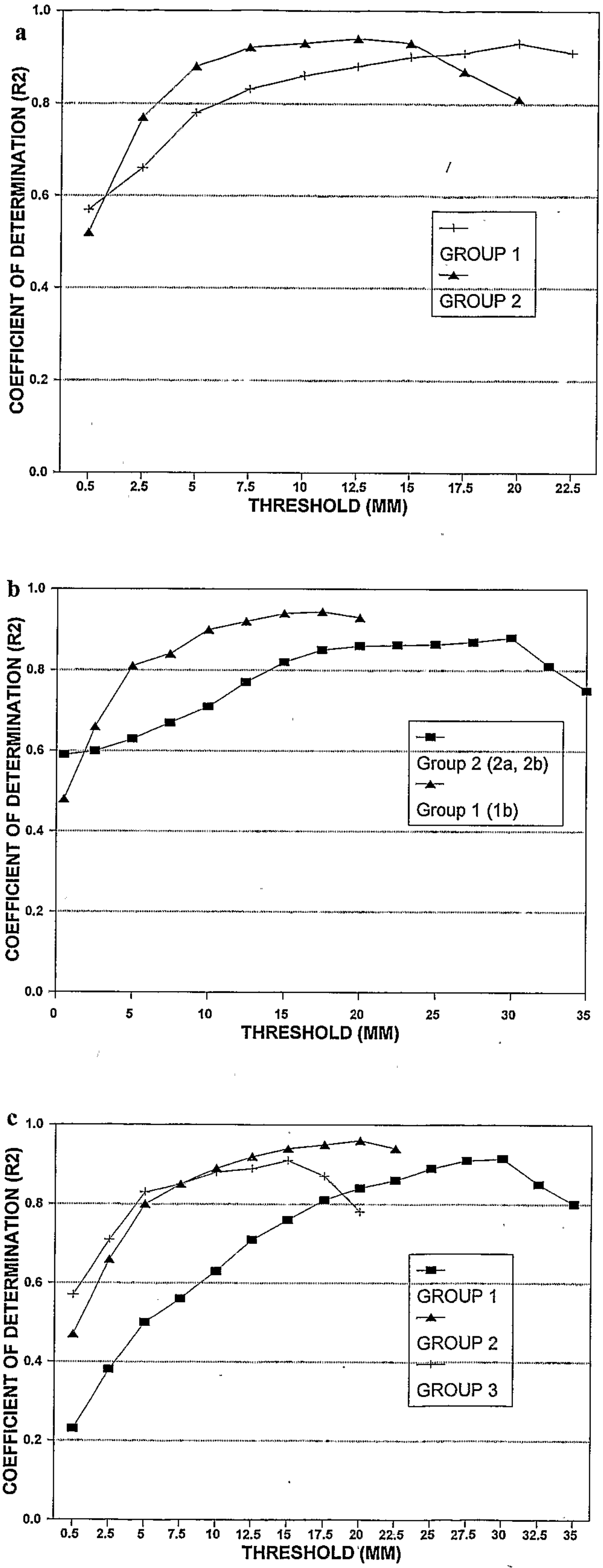


Figure 11. Coefficient of determination (R^2) versus threshold for the ATM models after classification for the different groups. (a) Method 1, (b) method 2, and (c) method 3.

Table 4. Parameters of the Optimal ATM Model for the Different Groups Obtained by Methods 1, 2, and 3 for the Combined Rainfall Events of 1990, 1991, and 1992

Method	R^2	Sample Size N	Optimum Threshold, mm	B, mm	St, mm
1	0.93	56	20	3.5	0.377
	0.94	50	12.5	1.5	0.253
2	0.88	26	30	10.4	0.436
	0.94	107	17.5	1.8	0.393
3	0.91	21	30	9.2	0.454
	0.96	55	20	2.6	0.436
	0.91	57	15	1.3	0.361
Total	0.94	133	17.5	2.1	0.361

R^2, coefficient of determination; St, slope of the regression; B, the ordinate at the origin.

ing methods 1 and 3 is considered here in order to see if improvement can be observed by stratification based on parameters related to the strength and the intermittence of the event rainfall. The parameters of the optimal ATM model are given in Table 5 for the four methods. On the basis of the reduction of the RMSE criterion after classification, the four methods can be ranked as follows: method 3 (23.6%), method 2 (20.8%), method 1 (13.5%), and method 4 (7.6%).

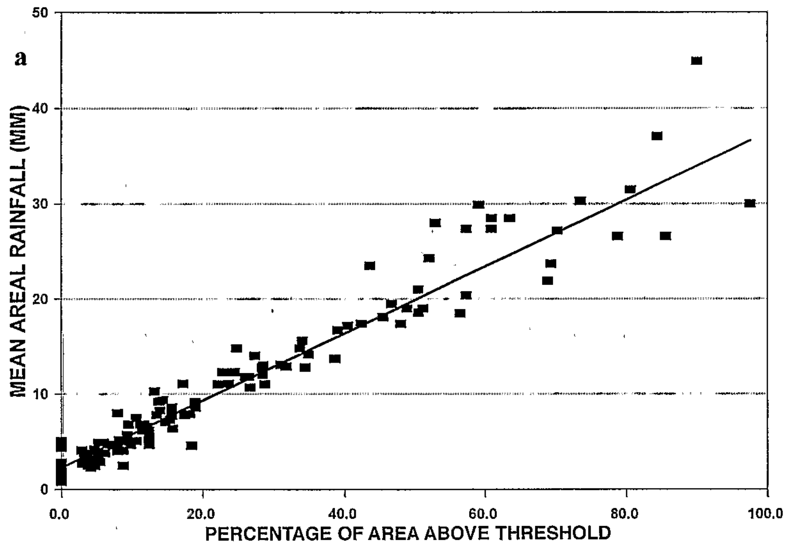
To check for possible bias linked to 1992 being a nonrepresentative year, a second validation procedure was designed. For a given classification method, events are considered in a chronological order inside each group produced by the method. Then, one out of two events is put into a validation sample. This leads us to obtain a calibration and a validation samples of identical size (plus or minus one unit) for each group. The overall calibration and validation samples are the addition of the calibration and validation samples of each group.

The results of this second validation procedure are similar to those obtained with the first one. That is, method 3 produces

Table 5. Calibration Error of the 1992 Mean Areal Rainfall Events Estimated by the ATM Models Before and After Classification

Method	Optimum Threshold, mm	R^2	St, mm	B, mm	RMSEb, mm	RMSEa, mm
3	...	...	...	...	2.50	1.91
	30	0.94	0.430	10.43	5.84	2.99
	20	0.96	0.441	2.83	2.28	2.20
	15	0.88	0.338	1.21	1.07	1.17
2	...	...	...	...	2.50	1.98
	30	0.95	0.432	12.94	4.39	3.24
	17.5	0.95	0.398	1.68	1.24	1.27
1	...	...	...	...	2.67	2.31
	20	0.93	0.417	2.70	3.48	2.92
	15	0.94	0.304	1.90	1.15	1.40
4	...	...	...	...	2.50	2.31
	20	0.88	0.349	5.05	4.00	3.61
	15	0.92	0.308	1.70	1.22	1.27
Total	17.5	0.94	0.335	2.28	2.50	...

R^2, coefficient of determination; St, slope of the regression; B, the ordinate at the origin; RMSEb, the error before classification; RMSEa, the error after classification. For method 1, only the events belonging to the significant groups 1 and 2 are considered; this is why the RMSEb is 2.67 and not 2.50.



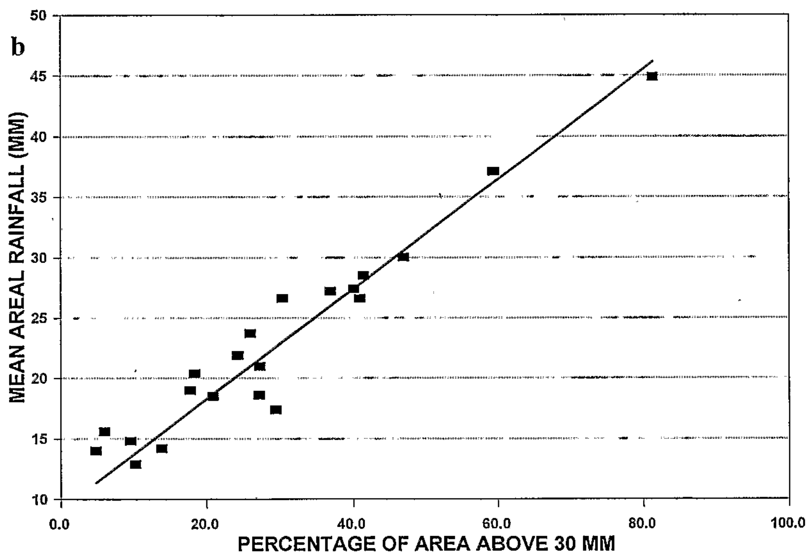


Figure 12. Optimal ATM models for the 3 years combined for the groups of method 3 (solid line represents the regression equation). (a) Total sample, (b) group 1, (c) group 2, and (d) group 3.

the largest improvement of the ATM model performances (Table 6). A reduction of the RMSE criterion after classification of 31.3% is obtained when sample 1 is estimated from sample 2, and a reduction of 36.5% is obtained when sample 2 is estimated from sample 1, that is, an average reduction of 34%. Group 1, composed of major events, contributes greatly to the performance of the method 3. For this group a reduction in RMSE validation values of 52% is obtained.

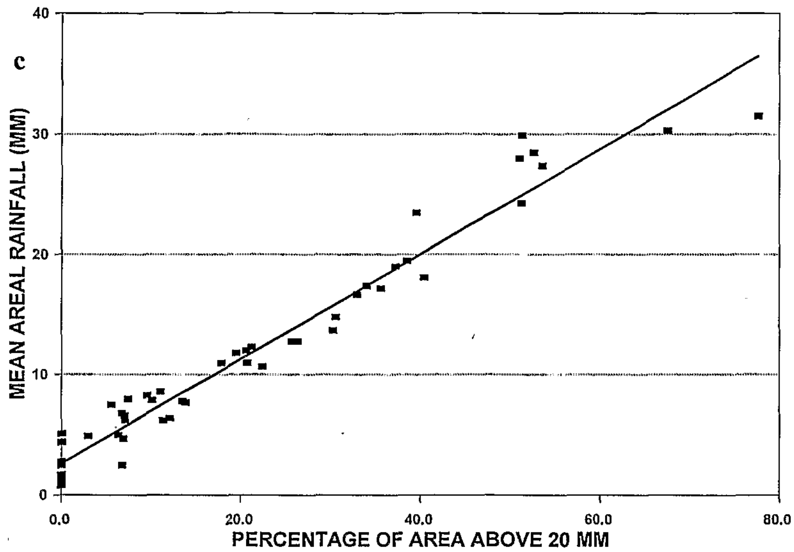
The good results in validation case 1 reveal the robustness of the classification combined with the ATM model, since the event statistics of the 1992 season were somewhat different from those of 1990 and 1991 (see statistics in Table 3). Classification method 4 does not perform as well as method 3. Thus there is no advantage in combining methods 1 and 3. The impact of the classification is very small for medium events and is sometimes negative for small or very intermittent rainfall events.

Analysis of the optimal ATM models and the statistics of the various samples considered (each year, the 1990–1991 combination, the 3 years together, and their corresponding groups for method 3) leads to the following four observations.

1. For each sample before classification the optimal threshold is around 17.5 mm. This means that the 3 years of EPSAT-NIGER data can be considered homogeneous. Thus the threshold method can be used in the Sahel as a climatological approach for areal rainfall estimation over large domains (over 10,000 km^2 in this case). Of course, this is by no way the best model when direct measurement (rain gauges) are available, but it may constitute a reasonable trade-off when only indirect measurements such as satellite infrared temperatures are available.

2. On the other hand, it is observed that for each sample considered, the three groups obtained from the DUPA classification (method 3) have different optimal thresholds and that these optimal thresholds are identical for all the samples. Their values are 30, 20, and 15 mm, respectively, for groups 1, 2, and 3.

3. Further confirmation of the stability and robustness of



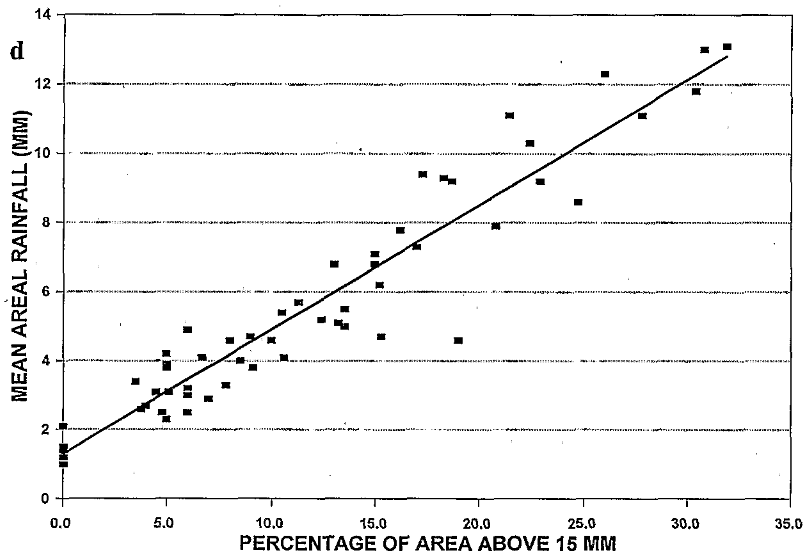


Figure 12. (continued)

classification method 3 is given by the contribution of the different groups in terms of total rainfall accumulation for each sample. Analyzing Table 3 (third row for each sample) reveals that the events in group 1 contribute greatly to the total rainfall relative to their number. It is also seen from this table that the ratio of the contribution of the events in each group relative to the total rainfall (as a percentage) to their number (as a percentage) is relatively constant from sample to sample. These ratios are approximately equal to 2, 1, and 0.5 for groups 1, 2, and 3, respectively.

4. Even though some further work is needed on that point, it is of particular interest to link the three groups of method 3 with the types of precipitation systems listed in section 2, namely, isolated convective systems, organized moving convective systems, and fully developed squall lines. As far as isolated convective systems could be identified to group 3 and fully developed squall lines to group 1, it can be concluded then that 1992 was markedly different from 1990 and 1991 in that the

Table 6. Optimal ATM Model Parameters and the Validation Errors for Validation Case 2 for Method 3

Group	Optimum Threshold, mm	R^2	St, mm	B, mm	RMSEb, mm	RMSEa, mm
			Sample 1			
	17.5	0.94	0.352	2.28	2.11	1.45
1	30	0.91	0.404	11.0	3.91	1.84
2	20	0.96	0.440	2.63	1.80	1.63
3	15	0.90	0.351	1.34	1.18	1.07
			Sample 2			
	17.5	0.94	0.380	1.99	2.55	1.62
1	30	0.93	0.429	10.2	4.77	2.34
2	20	0.96	0.433	2.52	2.39	1.86
3	15	0.93	0.373	1.19	1.22	0.87

R^2, coefficient of determination; St, slope of the regression; B, the ordinate at the origin; RMSEb, the error before classification; RMSEa, the error after classification.

isolated convective systems produced a larger proportion of the seasonal rainfall than usual. More generally, studying the contribution of each group of the UPA stratification to the seasonal rainfall could provide meaningful insights for hydrological and agronomic applications.

7. Conclusion

The impact of classification on the average estimation of rainfall over an area has been evaluated on the ATM model. Four methods of classification have been proposed and tested. These methods are method 1, based on the maximum rainfall accumulation over an array of time steps; method 2, based on the crossing analysis of the rainy area function against a threshold; and method 3, based on a new parameter, called UPA. The fourth method is a combination of methods 1 and 3. It should be noted that the UPA parameter is directly linked to the distribution of the area where it rains above a given threshold and that it is a scaled parameter.

A consequence of the application of these methods is that the nonconditional and conditional statistics (means of average rainfall, standard deviation, and coefficient of variation) for each obtained group differ from one group to another. Also, a reduction in the variance of the average conditional mean rainfall depth is observed for the majority of groups.

These methods were applied to the estimation of mean areal rainfall on the EPSAT-NIGER study area by the ATM model. For every method the optimal threshold of the ATM model is different from one group to another, which points up the importance of classification in improving the performances of the ATM model. At the calibration or validation level, classification methods 2 and 3 are the best in terms of RMSE reduction. The impact of classification is best for groups composed of large events (as in the case of group 1 for method 3). For this group the reduction in validation error are 52% and 48.8%, respectively, for the two validation cases. Natural classification based on the three groups from the UPA distribution analysis (method 3) is the best for the Sahelian rainfall events. A combination of method 1, which characterizes the intensity of the event, and method 3, which characterizes the spatial organization of the event, does not challenge method 3. This means that the spatial organization of the rainfall event is more important than rainfall intensity for estimation with a global model such as the ATM model.

The excellent results of classification applied to the ATM model, combined with the fact that the optimal threshold is different from one group to another, have a major consequence. The hypothesis of one probability distribution function of the conditional rain rate for the sample assumed by the ATM model has to be reconsidered. More specifically, it appears that separating between different types of convective events may prove to be as rewarding as separating between stratiform and convective rainfall.

Finally, it was pointed out that the classification method based on the UPA parameter is climatologically stable and robust. Thus this method can be used to stratify the various events observed during a rainy season in the Sahel based only on rain gauge data and can contribute to the improvement of our understanding of the interannual rainfall fluctuation in this region.

Notation

- AC percentage of area where it rains above a rainfall threshold.
- AC_j value of AC for event j (km^2).
- BC percentage of area where it rains below a rainfall threshold.
- BC_j value of BC for event j (km^2).
- $b(c)$ ordinate at the origin of the linear regression between $m(t)$ and $F(t, c)$ (mm).
- C Rainfall threshold (mm).
- C^* standardized rainfall threshold by C_M.
- C_M maximum rainfall recorded through the rain gauge network (mm).
- C_{Mj} value of C_M for event j (mm).
- $C_{M\theta}$ maximum rainfall recorded through the raingauge network for a particular period θ of accumulation (mm).
- CV coefficient of variation (dimensionless).
- $E[\ ]$ expectation operator.
- $F(C, k)$ fraction area where rain depth exceeds threshold C for event k (dimensionless).
- $F(t, c)$ fraction area where it rains above a threshold at time t (dimensionless).
- $I(i)$ value of $I(x, y, t)$ for gauge i and a given event (dimensionless).
- $I(x, y, t)$ indicator variable at (x, y) for time t (dimensionless).
- l_j frequency factor of the maximum cumulated rainfall for event j (dimensionless).
- m_j mean of the rainfall recorded by the network for event j (mm)..
- $m(t)$ mean areal rainfall at time on the surface of study (mm).
- P_n probability of nonexceedance of Q_n (dimensionless).
- Q_n quantile corresponding to the maximum rain depth recorded (mm).
- R^2 coefficient of determination (dimensionless).
- $R(i)$ rain depth recorded at gauge i for a particular event (mm).
- $R(x, y, t)$ rainfall depth recorded at location (x, y) at time t (mm).
- $S(c)$ slope of the linear regression between $m(t)$ and $F(t, c)$ (mm).
- s_j standard deviation of rain depth recorded by the raingauge network for event j (mm).
- θ period of rainfall accumulation (h).
- μ mean of the random process $R_T(x, y)$ (mm).
- σ standard deviation of the random process $R_T(x, y)$.

Acknowledgments. The suggestions and comments of three anonymous reviewers and also of D. I. Zawadzki from McGill University helped us to improve the presentation and the writing of this manuscript. They are gratefully acknowledged. Special thanks are addressed to the scholarship program of the Gouverment of Canada (Bourse de la Francophonie) for their financial support of the graduate studies of the first author. The implementation of the EPSAT-NIGER experiment involved the TOA and DEC departments of ORSTOM, as well as the Direction de la Méteorologie du Niger and the French Ministry of Cooperation. Their support is gratefully acknowledged. Finally, this research was partially supported by the Natural Sciences and Engineering Research Council of Canada through an individual research grant (OGP0008415).

References

Amani, A., Estimation des pluies au Sahel: Quelques problemes de modélisation analysés à partir des données de l'expérience EPSAT-NIGER, Ph.D. thesis, 308 pp., Dep. Génie Civ., Ecole Polytechnique de Montréal, Montreal, Canada, 1995.

Amani, A., J. Rousselle, T. Lebel, and J. D. Taupin, Hyetogram analysis of Sahelian mesoscale convective systems, in *Precipitation Variability and Climate Change: Proceedings of the International Symposium on Precipitation and Evaporation*, edited by B. Sevruk and M. Lapin, vol. 2, pp. 23–28, Slovak Hydrometeorol. Inst., Bratislava, Slovakia, 1993.

Anderberg, M. R., *Cluster Analysis for Applications*, Academic, San Diego, Calif., 1973.

Atlas, D., D. Rosenfeld, and D. A. Short, The estimation of convective rainfall by area integrals, 1, The theoretical and empirical basis, *J. Geophys. Res.*, *95*(D3), 2153–2160, 1990.

Benichou P., M. Lavielle, and P. Veyre, A study of the systematic behaviour of the French hemispheric NWP model as determined by a non-hierarchical clustering analysis, paper presented at OMM Workshop on Systematic Errors in Models of the Atmosphere, World Meteorol. Org., Toronto, Canada, Sept. 19–23, 1988.

Braud I., J. D. Creutin, and C. Barancourt, The relation between the mean areal rainfall and the fractional area where it rains above a given threshold, *J. Appl. Meteorol.*, *32*, 193–202, 1993.

Cunnane, C., Unbiassed plotting positions—A review, *J. Hydrol.*, *37*(3/4), 205–222, 1978.

Desbois, M., T. Kayiranga, B. Gnamien, S. Guesous, and L. Picon, Characterization of some elements of the Sahelian climate and their interannual variations for July 1983, 1984 and 1985 from the analysis of METEOSAT ISCCP data, *J. Clim.*, *1*(9), 867–904, 1988.

Doneaud, A. A., P. L. Smith, A. S. Dennis, and S. Sengupta, A simple method for estimating convective rain volume over area, *Water Resour. Res.*, *17*(6), 1676–1682, 1981.

Doneaud, A. A., S. I. Niscov, D. L. Priegnitz, and P. L. Smith, The area-time integral as an indicator for estimating convective rain volume, *J. Appl. Meteorol.*, *23*, 555–561, 1984.

Houze, R. A., Jr., Structure of atmospheric precipitation: A global survey, *Radio Sci.*, *16*, 671–689, 1981.

Kedem, B., and H. Pavlopoulos, On the threshold method for rainfall estimation: Choosing the optimal threshold level, *J. Am. Stat. Assoc.*, *86*(415), 626–633, 1991.

Kedem, B., D. A. Chiu, and Z. Karni, An analysis of the threshold method for measuring area-average rainfall, *J. Appl. Meteorol.*, *29*, 3–20, 1990.

Kottegoda, N. T., and A. H. M. Kassim, Classification of storm profiles using crossing properties, *J. Hydrol.*, *127*(1/4), 37–53, 1991.

LeBarbé, L., and T. Lebel, Rainfall climatology of the HAPEX Sahel during the years 1950–1990, *J. Hydrol.*, in press, 1996.

Lebel, T., H. Sauvageot, M. Hoepffner, M. Desbois, B. Guillot, and P. Hubert, Rainfall estimation in the Sahel: The EPSAT-NIGER experiment, *Hydrol. Sci. J.*, *37*(3), 201–215, 1992.

Lebel, T., J. D. Taupin, and M. Gréard, Rainfall monitoring: The EPSAT-NIGER set up and its use for HAPEX-Sahel, in *Hydrologie et Météorologie de Mésoéchelle dans HAPEX-SAHEL: Dispositif de mesures au sol et premiers résultats*, edited by T. Lebel, pp. 31–68, ORSTOM, Bondy, France, 1995.

Lebel, T., J. D. Taupin, and N. d'Amato, Rainfall monitoring during HAPEX-Sahel, 1, General rainfall conditions and climatology, *J. Hydrol.*, in press, 1996.

MacQueen, J., *Some methods for classification and analysis of multivariate observations*, Proceedings. 5th Berkeley Symp. on Math. Sta. and Prob., 1, 281–297, 1967.

Negri, A. J., and R. F. Adler, Infrared and visible satellite rain estimation, I, A grid cell approach, *J. Clim. Appl. Meteorol.*, *26*, 1553–1564, 1987.

Rosenfeld, D., D. Atlas, and D. A. Short, The estimation of convective rainfall by area integrals, 2, The Height-Area Rainfall Threshold (HART) method, *J. Geophys. Res.*, *95*(D3), 2161–2176, 1990.

Short, D. A., D. B. Wolff, D. R. Rosenfeld, and D. Atlas, A study of the threshold method utilizing raingage data, *J. Appl. Meteorol.*, *32*, 1379–1387, 1993.

Statistical Analysis System (SAS), *Statistics User's Guide*, 500 pp., SAS Inst., N. C., 1982.

Taupin, J. D., A. Amani, and T. Lebel, Small-scale spatial variability of the annual rainfall in the Sahel, in *Exchange Processes at the Land Surface for a Range of Space and Time Scales*, edited by H.-J. Bolle et al., *IAHS Publ. 212*, pp. 593–602, 1993.

Woodley, W. L., C. G. Griffith, J. S. Griffin, and S. C. Stromatt, The inference of GATE convective rainfall from SMS-1 imagery, *J. Appl. Meteorol.*, *19*, 388–408, 1980.

A. Amani and T. Lebel, Groupe PRAO, ORSTOM, LTHE, BP 53, 38041 Grenoble, Cedex 9, France. (e-mail: lebel@lthe710.img.fr)

J. Rousselle, Département de Génie Civil, Ecole Polytechnique, C. P. 6079, Succ. Centre-ville, Montreal, Quebec, H3C 3A7 Canada.

J. D. Taupin, Groupe PRAO-Niamey, BP 11416, Niamey, Niger.

(Received July 18, 1995; revised March 25, 1996; accepted April 8, 1996.)