Principal Modes of Variation of Rain-Rate Probability Distributions

THOMAS L. BELL AND R. SUHASINI*

Laboratory for Atmospheres, NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 2 April 1993, in final form 31 December 1993)

ABSTRACT

Radar or satellite observations of an area generate sequences of rain-rate maps. From a gridded map a histogram of rain rates can be obtained representing the relative areas occupied by rain rates of various strengths. The histograms vary with time as precipitating systems in the area evolve and decay and amounts of convective and stratiform rain in the area change. A method of decomposing the histograms into linear combinations of a few empirical distributions with time-dependent coefficients is developed, using principal component analysis as a starting point. When applied to a tropical Atlantic dataset (GATE), two distributions emerge naturally from the analysis, resembling stratiform and convective rain-rate distributions in that they peak at low and high rain rates, respectively. The two "modes" have different timescales and only the high-rain-rate mode has a statistically significant diurnal cycle. The ability of just two modes to describe rain variability over an area can explain why methods of estimating area-averaged rain rate from the area covered by rain rates above a certain threshold are so successful.

1. Introduction

Meteorological radars can be used to generate images of rain-rate fields over areas hundreds of kilometers in diameter. Satellites provide images over even larger domains, though generally less frequently and at lower spatial resolution. For some purposes, which will be discussed later, it can be convenient to ignore the spatial structure of the rain field and to concentrate on the histogram of the rain rates in the field; that is, the field is gridded, a bin size for rain rate is selected, and the number of grid points where rain rate falls in each of the bins is computed.

The rain-rate histogram for an area varies from moment to moment, as storms grow and decay and synoptic conditions evolve. Climatologically, rain-rate distributions are observed to change with the time of day. The changes are more evident for large rain rates than for small ones (e.g., Gibbins 1990), presumably because convective activity tends to be affected by the diurnal cycle of heating more strongly than the longer-lived stratiform precipitation, and convective rain is associated with higher rain rates. In order to characterize the effects of the diurnal changes in heating on rain activity, an economical description of the changes in the distribution of rain rates in terms of just a few parameters was needed.

Another motivation for this work arose from the attempt to understand the success of the area–time integral (ATI) method of estimating area-averaged rainfall. This method, suggested by the results of Chiu (1988) and further developed by Atlas et al. (1990), assumes that the area-averaged rain rate $\bar{R}$ can be estimated from the fraction f_τ of the area where rain rate exceeds a specified threshold τ, using the linear relationship

$$\bar{R} \approx S_\tau f_\tau. \tag{1.1}$$

The threshold τ is chosen to minimize the errors in the estimates. The coefficient S_τ can be shown to depend on the histogram of rain rates in the area in a straightforward way, and its sensitivity to changes in the histogram is of interest in evaluating the robustness of the ATI method to changes in the characteristics of rain with time of day, season, or geographical location. Short et al. (1993a) have shown empirically that the probability distribution of rain tends to vary so that the means and standard deviations of the distributions increase and decrease together, and that S_τ tends to be insensitive to just such variations. Kedem and Pavlopoulos (1991) and Short et al. (1993b) have shown that the threshold that minimizes the variability of S_τ due to sampling fluctuations in the histogram tends to predict the optimal choice for the threshold τ rather well.

A method of describing how the distribution of rain rates in an area changes with time will be investigated here by assuming that the rain in the area is composed

* Universities Space Research Association Resident Associate. Permanent affiliation: Physical Research Laboratory, Navrangpura, India.

Corresponding author address: Dr. Thomas L. Bell, Code 913, Goddard Space Flight Center, Greenbelt, MD 20771.

of several different "types" of rain, each of which has a characteristic distribution of rain rate associated with it. The distribution of rain rates in a field varies with time under this assumption because the relative amounts of the different types vary. This can be true only if the area is sufficiently large that it contains, for example, many convective cells at different stages in their evolution. The areawide rain-rate distribution is assumed to vary more because the number of cells changes than because of the evolution of individual cells, since a change in the distribution due to the evolution of one cell is compensated on average by changes in other cells at different stages in their evolution.

Houze (1981), in a survey of atmospheric precipitation systems, found that over many parts of the globe and in different climatic regimes, a classification of precipitation type as either stratiform or convective could provide a very useful first-order description of precipitation characteristics. Our approach to describing changes in the distribution of rain rates was in part motivated by Houze's (1981) analysis.

The decomposition of probability distributions into linear sums of parametric distributions (e.g., gamma, lognormal) is a well-explored topic in the statistical literature. Sansom and Thomson (1992) describe a particularly interesting application of these methods to the 15-yr-average distribution of pluviograph data from a New Zealand rain gauge. They find that their data are naturally decomposable into two rain types.

A time-dependent linear decomposition of rain-rate distributions into data-adaptive, empirical probability distributions is investigated here. The component probability distributions are not restricted to a particular class of parameterized distributions. This initial approach leads to a numerically difficult problem, but by introducing some approximations an orthogonal basis for the expansion of the time-dependent histograms is obtained that is easily and naturally recast into an approximation of the desired expansion.

The method is described in detail in the next section. In section 3 the method is applied to radar-derived rain maps obtained in the Global Atmospheric Research Program's Atlantic Tropical Experiment (GATE). Section 4 discusses uses of the method and possible avenues for future research, and section 5 contains some conclusions. Mathematical details are given in an appendix.

2. Decomposing time-varying histograms into component probability distributions

a. General approach

Suppose that a time series of gridded images is available, each grid point representing rain rate averaged over its associated grid box. The rain rates $R(\mathbf{x}, t)$ at all grid points $\mathbf{x}$ in an image at time t are histogrammed, and $n(r_i, t)$ is the number of grid boxes with rain rates falling in bin i, with the ith bin delimited by

$$r_i \leqslant R < r_{i+1}. \tag{2.1}$$

The total number of counts in all B bins is

$$N(t) = \sum_{i=1}^{B} n(r_i, t). \tag{2.2}$$

It is hypothesized that a large area contains rain of different types in different regions, and that each type is associated with a different rain-rate distribution. Rain rates for stratiform rain, for example, are typically much lower than rain rates for convective systems. Because the dynamical development of the different types of rain is different, the relative amounts of the types will vary from image to image. This variation will allow us to extract the component distributions associated with each type.

It is thus proposed to describe the time variation of $n(r_i, t)$ by an expression of the form

$$n(r_i, t) \approx \sum_{\alpha=1}^{M} n_\alpha(t) p_\alpha(r_i), \tag{2.3}$$

hoping that only a few "modes" $p_\alpha(r_i)$ are needed to capture the behavior of $n(r_i, t)$ adequately. [The term "component" would have been preferable to "mode" to describe the $p_\alpha(r_i)$, but the term "principal components" is already used for the eigenvectors of a covariance matrix.] Expression (2.3) may be interpreted to mean that out of the $N(t)$ grid points, $n_\alpha(t)$ of them are occupied by rain of type α. Some constraints are imposed on the n_α, p_α by their physical interpretation: it is required that, for each α,

$$n_\alpha(t) \geqslant 0; \tag{2.4a}$$

$$p_\alpha(r_i) \geqslant 0, \quad i = 1, \ldots, B; \tag{2.4b}$$

$$\sum_i p_\alpha(r_i) = 1. \tag{2.4c}$$

If the description of $n(r_i, t)$ in (2.3) were perfect, it would follow that

$$\sum_\alpha n_\alpha(t) = N(t). \tag{2.5}$$

This will not, however, be imposed as a constraint; the degree of agreement with (2.5) will instead be viewed as one measure of the success of the description.

Because of the physical interpretation of the $p_\alpha(r_i)$, they would be expected to be somewhat disjoint. For example, to the extent that the distributions are identifiable as stratiform or convective, a "stratiform" $p_s(r_i)$ would be expected to diminish rapidly at large rain rates relative to a "convective" distribution $p_c(r_i)$. The relative strengths of the two distributions would be reversed at low rain rates. These conditions might be written as

$$\frac{p_s(r_i)}{p_c(r_i)} \quad \text{small for } r_i \text{ large;} \tag{2.6a}$$

$$\frac{p_c(r_i)}{p_s(r_i)} \quad \text{small for } r_i \text{ small.} \tag{2.6b}$$

To simplify the presentation, introduce the vector notation $\mathbf{p}_\alpha \equiv \{p_\alpha(r_i); i = 1, \ldots, B\}$ and $\mathbf{n}(t) \equiv \{n(r_i, t); i = 1, \ldots, B\}$. Our task is to find a set of vectors $\mathbf{p}_\alpha$ that describe the time-dependent histogram data $\mathbf{n}(t)$ with a minimum amount of error. To express this, rewrite (2.3) as

$$\mathbf{n}(t) = \mathbf{n}_E(t) + \epsilon(t), \tag{2.7}$$

where

$$\mathbf{n}_E(t) \equiv \sum_\alpha n_\alpha(t)\mathbf{p}_\alpha \tag{2.8}$$

is the expected histogram count under our assumptions, and $\epsilon(t)$ is the error in the description. A global measure of the error of the description is needed. A simple global measure of error would be just the total squared error $\sum_t \epsilon'(t)\epsilon(t)$, where the prime indicates vector transpose. A better measure of error, suggested by a standard approach to fitting distributions to a histogram of independent samples, will be investigated here instead. Although the assumption that the rain-rate histograms are composed of independent samples is not entirely valid in our case, it suggests a measure of error that has many desirable features that will be discussed later.

If the expected number of counts in bin i is $n_E(r_i, t)$ and the counts are independently distributed, then the error $\epsilon(r_i, t)$ in (2.7) is normally distributed in the limit of a large number of samples, with expected variance $n_E(r_i, t)$; see Cramér (1946), for example. This suggests using as a measure of the goodness of fit the weighted-squares quantity

$$\mathcal{E} = \sum_t \sum_i \frac{\epsilon^2(r_i, t)}{n_E(r_i, t)}, \tag{2.9}$$

with the property that in the limit of a large number of independent samples it is distributed as a chi-squared variable (Cramér 1946). It is standard statistical practice to fit parameterized distributions to sample histograms by minimizing expressions like (2.9). Such fits correspond to maximum-likelihood estimates of the sample distribution. An example of this approach and a discussion of some of the effects of sample correlation on the interpretation of $\mathcal{E}$ is given by Kedem et al. (1990). In general, both the expected mean and variance of $\mathcal{E}$ are inflated by sample correlations over what they would be if the samples were independent.

Given expression (2.9) to minimize, it is in principle a straightforward matter to find the distributions $\mathbf{p}_\alpha$ and corresponding time series $n_\alpha(t)$ that minimize $\mathcal{E}$, subject to the constraints (2.4). In practice, however, because $\mathcal{E}$ is not a quadratic function of the unknowns, standard numerical procedures for least-squares problems are not applicable, and finding the minimum is numerically difficult.

b. An approximation and a new basis

An approximation to the true solution can be obtained by replacing the denominator of (2.9) with its average value,

$$n_E(r_i, t) \approx N(t)p(r_i), \tag{2.10}$$

where $N(t)$ is defined in (2.2) and $p(r_i)$ is the frequency distribution for the entire rain-rate dataset,

$$\mathbf{p} \equiv \frac{\sum_t \mathbf{n}(t)}{\sum_t N(t)}, \tag{2.11}$$

normalized to $\sum_i p(r_i) = 1$. This approximation can be viewed as a first step in an iterative approach to minimizing (2.9).

With this approximation, $n_\alpha(t)$, $\mathbf{p}_\alpha$ must now be found that minimize

$$\mathcal{E}_0 = \sum_t \sum_i \frac{[n(r_i, t) - \sum_\alpha n_\alpha(t)p_\alpha(r_i)]^2}{N(t)p(r_i)}. \tag{2.12}$$

This can be written in vector notation as

$$\mathcal{E}_0 = \sum_t \frac{[\mathbf{n}(t) - \mathbf{n}_E(t)]'\mathbf{W}^2[\mathbf{n}(t) - \mathbf{n}_E(t)]}{N(t)}, \tag{2.13}$$

where the diagonal weighting matrix

$$(\mathbf{W})_{ij} \equiv \frac{\delta_{ij}}{[p(r_i)]^{1/2}} \tag{2.14}$$

has been introduced, δ_{ij} is the Kronecker delta, and $\mathbf{n}_E(t)$ is defined in (2.8).

It is a remarkable fact that the problem of minimizing (2.13) can be converted into a simple matrix eigenvalue–eigenvector problem if one is willing to relax the constraints (2.4). Because the solutions are informative and so easy to obtain and can serve as a first guess for the numerically more difficult problem of minimizing (2.9), their properties will be investigated here.

If the modes $\mathbf{p}_\alpha$ that minimize (2.13) were known in advance, finding the corresponding time series $n_\alpha(t)$ that minimizes (2.13) would be identical to the standard linear least-squares problem. The solutions for $n_\alpha(t)$ are obtained by setting the derivatives of $\mathcal{E}_0$ with respect to each $n_\alpha(t)$ equal to 0. This gives, for each α and t,

$$\sum_\gamma (\mathbf{p}'_\alpha\mathbf{W}^2\mathbf{p}_\gamma)n_\gamma(t) = \mathbf{p}'_\alpha\mathbf{W}^2\mathbf{n}(t). \tag{2.15}$$

This is a linear equation for the $n_\alpha(t)$ that can be solved, albeit in terms of the unknowns $\mathbf{p}_\alpha$. The solutions of

these equations are not, however, guaranteed to be nonnegative: that is, they may not obey constraint (2.4a) at all times t. We will nevertheless proceed using these solutions, substituting them in (2.13) and then attempting to solve for the set of $\mathbf{p}_\alpha$ that minimizes (2.13). The solutions will then be examined to see whether the violations of the constraints (2.4), if any, are acceptable or not.

To this end, it is convenient to expand the $\mathbf{p}_\alpha$ in terms of a new basis set of vectors,

$$\mathbf{p}_\alpha = \sum_\beta c_{\alpha\beta}\mathbf{x}_\beta, \tag{2.16}$$

where the vectors $\mathbf{x}_\beta$ are orthonormal with respect to the weighting $\mathbf{W}^2$;

$$\mathbf{x}'_\beta\mathbf{W}^2\mathbf{x}_\gamma = \delta_{\beta\gamma}, \tag{2.17}$$

so that the coefficients $c_{\alpha\beta}$ are determined by

$$c_{\alpha\beta} = \mathbf{p}'_\alpha\mathbf{W}^2\mathbf{x}_\beta. \tag{2.18}$$

The "change of basis" $\mathbf{p}_\alpha \rightarrow \mathbf{x}_\beta$ diagonalizes (2.15) and enables it to be solved easily:

$$n_\alpha(t) = \sum_\beta m_\beta(t)(\mathbf{c}^{-1})_{\beta\alpha}, \tag{2.19}$$

with $(\mathbf{c})_{\alpha\beta} \equiv c_{\alpha\beta}$, $\mathbf{c}\mathbf{c}^{-1} = 1$, and

$$m_\beta(t) = \mathbf{x}'_\beta\mathbf{W}^2\mathbf{n}(t). \tag{2.20}$$

If this solution for $n_\alpha(t)$ is substituted in (2.13),

$$\mathcal{E}_0 = \sum_t \frac{[\mathbf{n} - \sum_\beta (\mathbf{x}'_\beta\mathbf{W}^2\mathbf{n})\mathbf{x}_\beta]'\mathbf{W}^2[\mathbf{n} - \sum_\gamma (\mathbf{x}'_\gamma\mathbf{W}^2\mathbf{n})\mathbf{x}_\gamma]}{N(t)} \tag{2.21}$$

is obtained, where indication of the dependence of $\mathbf{n}$ on t has been omitted. It is shown in the appendix that the problem of obtaining the unknown vectors $\mathbf{x}_\beta$ that minimize (2.21) reduces to a simple eigenvalue problem: Define the matrix

$$\mathbf{C} \equiv \sum_t \frac{\mathbf{n}(t)\mathbf{n}'(t)}{N(t)} \tag{2.22}$$

and obtain the eigenvectors ψ_β of the symmetric matrix $\mathbf{WCW}$,

$$\mathbf{WCW}\psi_\beta = \lambda_\beta\psi_\beta; \tag{2.23}$$

then a basis set $\mathbf{x}_\beta$ that minimizes (2.21) is just

$$\mathbf{x}_\beta = \mathbf{W}^{-1}\psi_\beta, \tag{2.24}$$

where the subset of M eigenvectors with the largest eigenvalues is chosen.

As mentioned above, however, the constraints (2.4) may possibly not be satisfied in this approximation. The solutions for $n_\alpha(t)$ may be negative sometimes, and a satisfactory set of coefficients $c_{\alpha\beta}$ in (2.16) for combining the vectors $\mathbf{x}_\beta$ to form a set of nonnegative $\mathbf{p}_\alpha$ may not exist. If, however, there exists a set of modes $\mathbf{p}_\alpha$ that describes $\mathbf{n}(t)$ with little error [see (2.3)], the approach above would find them.

The approximation (2.10), although introduced out of necessity, has resulted in the generation of a basis set $\mathbf{x}_\beta$ with a very appealing set of properties. Four of them are given here:

1) As shown in (A.6) in the appendix, when combined with (2.24) above, the first vector of the basis set is identical to the average frequency distribution of rain rate,

$$\mathbf{x}_1 = \mathbf{p}, \tag{2.25}$$

with associated eigenvalue $\lambda_1 = \sum_t N(t)$. This means that the modes $\mathbf{p}_\alpha$ formed from the basis set using (2.16) will always be able to be combined to describe the average histogram (2.11) perfectly. The basis vectors $\mathbf{x}_\beta$, $\beta \geq 2$, describe deviations of the image histogram from the average histogram.

2) The orthogonality of the basis set $\mathbf{x}_\beta$ and the result (2.25) imply

$$\sum_i x_\beta(r_i) = 0, \quad \beta \geq 2, \tag{2.26}$$

as can be shown by setting $\gamma = 1$ in (2.17).

3) The normalization of the $\mathbf{p}_\alpha$, (2.4c), and (2.26) above imply that for all α, in (2.16),

$$c_{\alpha 1} = 1. \tag{2.27}$$

4) Since the basis set $\{\mathbf{x}_\beta; \beta = 1, \ldots, B\}$ is complete, $\mathbf{n}(t)$ can be expanded in terms of the $\mathbf{x}_\beta$,

$$\mathbf{n}(t) = \sum_{\beta=1}^{B} m_\beta(t)\mathbf{x}_\beta, \tag{2.28}$$

with the coefficients m_β defined in (2.20). Using (2.25) and (2.26) and the definition of $N(t)$ in (2.2), it follows immediately that

$$m_1(t) = N(t). \tag{2.29}$$

The first coefficient in the expansion (2.28) is thus just the total number of counts in the histogram.

c. *Obtaining the modes* $\boldsymbol{p}_\alpha$

It has been shown how to obtain a basis set of vectors $\mathbf{x}_\beta$ that it is hoped may be combined according to (2.16) to obtain the modes $\mathbf{p}_\alpha$. The basis set has the property that linear combinations formed from it are able to describe histogram variations with minimal error, as measured by (2.13), with (2.8) replaced by

$$\mathbf{n}_E(t) = \sum_{\beta=1}^{M} m_\beta(t)\mathbf{x}_\beta$$

$$= N(t)\mathbf{p} + \sum_{\beta=2}^{M} m_\beta(t)\mathbf{x}_\beta, \tag{2.30}$$

using (2.25) and (2.29). If an economical description of the variability of rain-rate histograms is sought, expression (2.30) may be sufficient. As discussed above, however, values of $n_E(r_i, t)$ obtained from (2.30) may sometimes be negative.

To obtain the modes $\mathbf{p}_\alpha$, the coefficients in (2.16) must be chosen appropriately. Equation (2.16) may now be written

$$\mathbf{p}_\alpha = \mathbf{x}_1 + \sum_{\beta=2}^{M} c_{\alpha\beta}\mathbf{x}_\beta = \mathbf{p} + \sum_{\beta=2}^{M} c_{\alpha\beta}\mathbf{x}_\beta, \quad (2.31)$$

using (2.27) and (2.25). If just two modes were sufficient to describe the histogram variability, two coefficients would need to be specified, one for each of the modes. Unfortunately, because an approximation has been introduced that linearizes the problem, and because the constraints (2.4a) and (2.4b) have been temporarily relaxed, the coefficients are not uniquely determined. The only guidance available for choosing the coefficients comes from the constraints. Note that the normalization constraint (2.4c) is automatically satisfied by (2.31) because of (2.26).

To proceed, some experience with the behavior of this approach using rain-rate data is needed, and the behavior of the technique using GATE radar data will be explored in the next section. In order not to conclude this section leaving the question of how to choose the coefficients unanswered, we summarize our experience here for the case where we limit ourselves to just two modes ($M = 2$). The first vector $\mathbf{x}_1(r_i)$ is fixed by (2.25). The vector $\mathbf{x}_2(r_i)$ generally changes sign once as r_i increases. As a result, the condition $p_\alpha(r_i) = p(r_i) + c_{\alpha 2}\mathbf{x}_2(r_i) \geq 0$ places bounds on $c_{\alpha 2}$ of the sort

$$-a \leq c_{\alpha 2} \leq b. \quad (2.32)$$

It will be found that if the choices $c_{12} = -a$ and $c_{22} = b$ are used, then two modes $p_\alpha(r_i)$ are obtained that go to zero, respectively, for large and small rain rates r_i, agreeing with the observational experience for stratiform and convective rain summarized in (2.6). The resulting time series $n_\alpha(t)$ determined from (2.19) and (2.20) sometimes go negative, but mostly for cases where the number of counts $N(t)$ is low and the histograms are likely to be "noisy." The portion of time for which the $n_\alpha(t)$ are negative increases if the coefficients are chosen inside the limits (2.32) instead of at the boundaries, and, by that measure, these choices are less desirable. It is our judgment, based on this, that a correct, but numerically difficult, minimization of (2.9) obeying the constraints (2.4) leads to unique modes $\mathbf{p}_\alpha$, similar to what is obtained with our approximation scheme. This has been our experience in the few cases with which we have experimented. The behavior of the method with actual data is explored next.

3. Principal modes of variation of GATE rain

The GATE dataset analyzed here was derived by Hudlow and Patterson (1979) from radar measurements taken in the tropical Atlantic off the west coast of Africa during the summer of 1974. The radar data were converted into rain rates $R(\mathbf{x}, t)$ on a 4-km grid covering a circle 400 km in diameter centered on 8°30'N, 23°30'W, with each gridpoint value representing the instantaneous rain rate averaged over a 4 km × 4 km box. Only the portion of the rain-rate field bounded by a 280-km square centered in the GATE area is used. The square contains 4900 grid points. Rain rates are binned logarithmically into 36 bins, with the bin boundaries in (2.1) given by

$$r_1 = 0.251 \text{ mm h}^{-1},$$
$$r_i = 10^{3(i-1)/40} r_1, \quad i = 1, \ldots, 36,$$
$$r_{36} = 106 \text{ mm h}^{-1},$$
$$r_{37} = \infty. \quad (3.1)$$

The bin sizes were chosen to conform to the digitization of rain rates in the dataset. Note that zero rain rates are excluded from the histogram. Their inclusion will be discussed later.

Phase I of GATE (extending from 28 June to 16 July 1974) contains 1716 rain-rate maps at intervals of approximately 15 min, with occasional gaps. Histograms $n(r_i, t)$ were obtained for each map. The covariance matrix of the histograms was computed weighted as in (2.22), and the eigenvalues obtained as specified in (2.23). The eigenvalues (divided by the number of images $\sum_t 1 = 1716$) are shown in Fig. 1. They decrease rapidly, and the first two account for

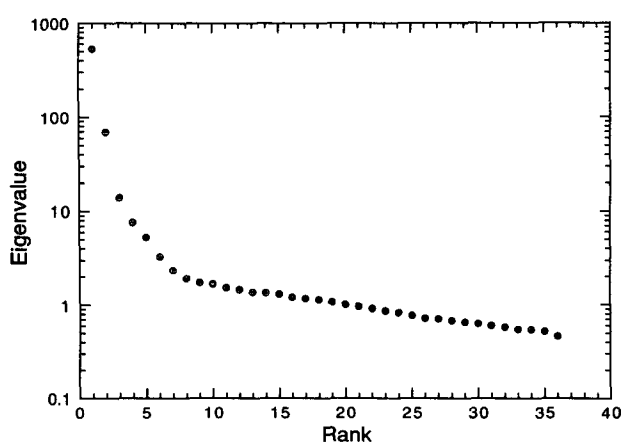

FIG. 1. Eigenvalues of GATE phase I weighted histogram covariance matrix shown in Eq. (2.23), normalized by the number of histograms (1716).

TABLE 1. Histogram variability described by first few eigenvectors for GATE phase I, based on 1716 rain-rate fields (280 km × 280 km at 4-km resolution).

Number M of eigenvectors	Percent trace "explained"	$\frac{\mathcal{E}_0}{\sum_t 1}$
0	0	664
1	80.2	131
2	90.7	61
3	92.8	47

91% of the trace of the matrix being diagonalized, as shown in Table 1. Also shown in Table 1 is the average goodness of fit per histogram as measured by $\mathcal{E}_0/(\sum_t 1)$. To the extent that 1) the quantity $\mathcal{E}$ in (2.9) is distributed as a chi-squared variable, 2) the samples counted by the histograms are independent, and 3) the quantity $\mathcal{E}$ is approximated by $\mathcal{E}_0$ in (2.12), $\mathcal{E}_0$ would be expected to equal the "degrees of freedom" $B \sum_t 1$ less the number of parameters in the fitting procedure (see, e.g., Cramér 1946). It might therefore be expected that $\mathcal{E}_0/(\sum_t 1) \approx B - M = 36 - M$. The numbers in the third column have not dropped to this level even with three eigenvectors, but because of the approximations involved and the spatial correlations, which tend to increase $\mathcal{E}$, these numbers can at best serve as a qualitative measure of the ability of the eigenvectors to describe the variability of the histograms. It is clear, however, that with just two modes much of the variability of the histograms is being captured. Limiting the description to two modes has the additional benefit that the modes are easy to construct and the two-mode description is straightforward to interpret and highly informative.

a. *Principal modes*

The first two orthogonal basis vectors $\mathbf{x}_1 = \mathbf{p}$ and $\mathbf{x}_2$ are shown in Fig. 2. They must be combined according to (2.31) to form the principal modes of variation $\mathbf{p}_\alpha$, subject to the nonnegativity constraint (2.4b). In order to satisfy the constraint, the coefficients $c_{\alpha 2}$ must lie in the range

$$-0.372 \leq c_{\alpha 2} \leq 0.761, \quad \alpha = 1, 2. \qquad (3.2)$$

Several distinct lines of reasoning lead to choosing the two boundary values in (3.2) for the values of $c_{\alpha 2}$:

- This choice produces two nonnegative modes with minimum overlap, as measured by $\mathbf{p}'_\alpha \mathbf{p}_\beta$ $(\alpha \neq \beta)$; that is, they are as nearly "disjoint" as possible.
- This choice yields two modes that conform to the expectation that stratiform rain will have relatively few counts at high rain rates, whereas convective rain will show relatively few low-rain-rate counts [cf. (2.6)]. In fact, the ratio $x_1(r_i)/x_2(r_i)$ tends to level off to a value of 0.372 at high rain rates and −0.761 at low rain rates, indicating that combining them with these coefficients would produce modes that look like stratiform and convective rain distributions.
- When the time series $n_\alpha(t)$ in (2.3) that result from this choice are obtained from (2.19), some values are negative, contrary to the constraint (2.4a). This issue will be discussed more later—but if values of $c_{\alpha 2}$ interior to the bounds in (3.2) are used, the number of occurrences of negative $n_\alpha(t)$ increases. This problem can be minimized by choosing the $c_{\alpha 2}$ at the boundaries.

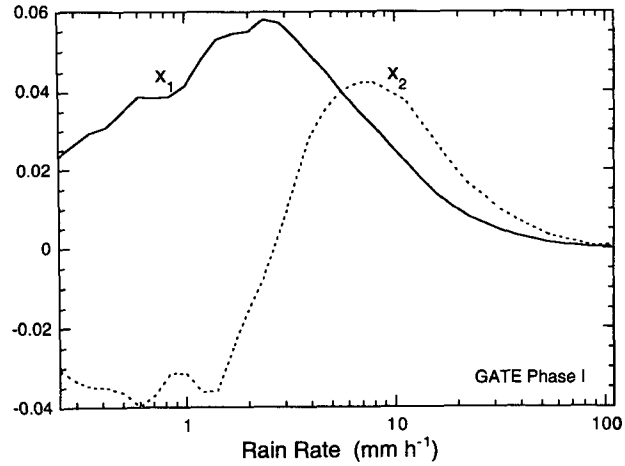

FIG. 2. First two basis vectors from Eq. (2.24) for describing GATE phase I histogram variability.

These arguments suggest constructing the two modes

$$\mathbf{p}_L = \mathbf{x}_1 - 0.372\mathbf{x}_2, \qquad (3.3a)$$

$$\mathbf{p}_H = \mathbf{x}_1 + 0.761\mathbf{x}_2, \qquad (3.3b)$$

which are shown in Fig. 3. The labels L for "low" and H for "high" are used to indicate the rain rates pri-

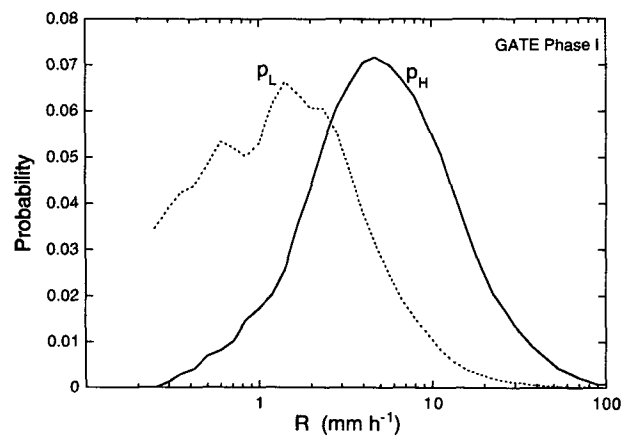

FIG. 3. Principal modes of variability ($M = 2$) of GATE phase I histograms, derived from the basis vectors in Fig. 2. Labels L for "low" and H for "high."

marily associated with each mode. It is tempting to refer to the two rain-rate probability distributions as "stratiform" and "convective," since there is so much observational evidence that rain can often be meaningfully classified as one or the other type, as discussed in the introduction, and these two distributions successfully describe a considerable amount of the variability of the distribution of rain rates over an area and are suggestively dominated by low and high rain rates, respectively. Because there are such strong dynamical connotations to the terms, however, and because these modes are derived from statistical rather than structural considerations, we will limit ourselves to the labels L and H, while noting the strong associations with the characteristics of stratiform and convective rain.

The rain rates represented by $\mathbf{p}_L$ are generally less than 10 mm h^{-1}, as was found for anvil rain in GATE in five cases analyzed by Leary and Houze (1979). The distribution $\mathbf{p}_H$ peaks near 5 mm h^{-1} and accounts for nearly all rain rates above 10 mm h^{-1}. It is fit rather well by a lognormal distribution with parameters $\mu(\ln r) = 1.7$ and $\sigma(\ln r) = 0.985$, when r is in units millimeters per hour. The mean rain rate associated with each mode is

$$\bar{r}_\alpha \equiv \sum_i \bar{r}_i p_\alpha(r_i), \qquad (3.4)$$

where $\bar{r}_i$ is the average rain rate for bin i (recall that r_i is the lower boundary of bin i). Values are shown in Table 2.

b. *Time series $n_\alpha(t)$*

The choices (3.3) for $\mathbf{p}_L$ and $\mathbf{p}_H$ imply corresponding time series $n_L(t)$ and $n_H(t)$ based on (2.19). Figure 4 shows a scatterplot of $n_H(t)$ versus $n_L(t)$ for GATE phase I. A substantial number of values of $n_H(t)$ are negative (700 of the 1716 points). This can be dealt with in several ways:

- The negative values occur because the least-squares fit to the histograms, based on (2.19) and (2.20), was obtained ignoring the nonnegativity constraint (2.4a); the constraint could simply be reimposed by setting all negative values to zero and adjusting the other values to satisfy (2.5). Because the negative values occur mostly when the other mode dominates (i.e., $n_L \gg |n_H|$), this adjustment has a relatively minor impact on the quality of the fits.
- Alternatively, the coefficients $c_{\alpha 2}$ could be permitted to move outside the bounds given in (3.2), so that the modes $p_\alpha(r_i)$ begin to have slightly negative values at some rain rates r_i. With relatively small changes in $c_{\alpha 2}$, the number of occurrences of negative values of $n_H(t)$ decreases dramatically. The negative values of $p_\alpha(r_i)$ that result can be set to zero.
- The best approach would be to take the results so far as a starting point and to return to the constrained, nonlinear problem of minimizing $\mathcal{E}$ in (2.9). We have experimented with this and find that the solution tends toward a point somewhere between what is obtained following the first two approaches above—but the numerical effort required to find the minimum is greater.

The first approach above, adjusting negative values of n_α to zero, is both simple and produces results that are close to what would be obtained from a more exact treatment, based on this limited experience. It will be followed here.

There are five values of $n_L(t)$ with particularly large negative values, all occurring between 1630 and 1745 UTC 7 July (Julian day 188). The highest area-aver-

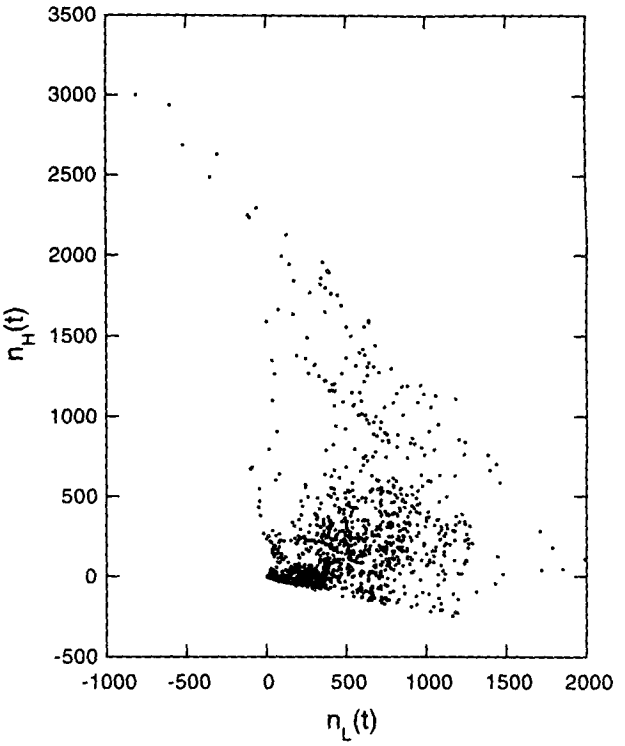

FIG. 4. Scatterplot of coefficients of p_L and p_H in Eq. (2.3), ignoring constraint Eq. (2.4a).

TABLE 2. Characteristics of two principal modes based on GATE phase I data. See text for caveats concerning "stratiform" and "convective" labels. Column 2 shows average rain rate defined in Eq. (3.4). Column 3 shows correlation times of n_L and n_H, column 4 shows relative areas occupied by the two types, and column 5 shows the relative rain volumes attributable to each type. The 95% confidence interval for both relative area and rain volume is estimated to be ±0.19.

Mode	$\bar{r}_\alpha$ (mm h^{-1})	Correlation time (h)	Relative area	Relative rain volume
L ("stratiform")	2.6	11	0.65	0.36
H ("convective")	8.8	6	0.35	0.64

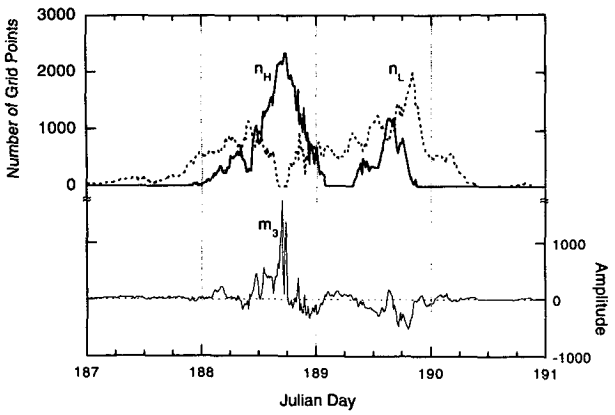

FIG. 5. Evolution in time of $n_L(t)$ and $n_H(t)$ [constrained by Eq. (2.4a)] during portion of GATE phase I. Highest volume of rain fell during day 188. Shown below is the coefficient $m_3(t)$ in Eq. (2.28), a measure of what is not captured by the two-mode expansion.

aged rain rates observed during GATE occurred during these few hours. The histogram for 1645 UTC, for example, when n_L is most negative, peaks at approximately 15 mm h^{-1} and cannot be described well by a mixture of $\mathbf{p}_L$ and $\mathbf{p}_H$ (see Fig. 3). There are clearly some extreme events that cannot be fully captured by just two modes.

A portion of the time series $n_L(t)$ and $n_H(t)$ for Julian days 188–191 is shown in Fig. 5. The constraints $n_\alpha(t) \geqslant 0$ have been imposed on the series, as discussed above. There appears to be some suppression of low rain rates when the strongest convective events occur. This is not an artifact of the fitting process: the dip can also be seen in a plot (not shown) of the area covered by rain rates below 6 mm h^{-1}—there is a sharp drop during the period of most intense rainfall. It is not clear whether this is dynamical in origin or an artifact due to radar attenuation. Note that this is the same event mentioned above that is responsible for the points $n_L \ll 0$ in Fig. 4.

The area of high-rain-rate activity, as represented by $n_H(t)$, develops and dies out much more rapidly than the "stratiform" rain area. The correlation times, determined by the lags when autocorrelations fall to $1/e$, are shown in the third column of Table 2. The correlation time of the area of light rain is nearly twice that of the area of heavy rain.

The relative areas of the two types, determined from

$$A_\alpha = \sum_t n_\alpha(t), \tag{3.5}$$

are shown as fractions $A_\alpha/(A_H + A_L)$ in Table 2. The corresponding rain volumes contributed by the two types, as fractions of the total areawide rainfall in the 280-km square, are shown in the last column of Table 2. They are estimated to be uncertain by ±0.19 (95% confidence interval), based on their correlation times, means, and variances. Cheng and Houze (1979) estimated that stratiform rain contributed about 50% of the rain that fell during GATE phase I. In their analysis, the area averaged over was a circle 520 km in diameter, considerably larger than the area analyzed here, and included more of the intertropical convergence zone (ITCZ); only data from times near 1200 UTC were used, and the definition of convective rain was based on identification of rapidly changing, intense, localized radar echoes. There is a strong diurnal cycle in $n_H(t)$ in phase I. When the relative rain volumes due to types L and H were recomputed for the period 1000–1400 UTC, the rain volume fraction due to low rain rates decreased to 0.28, which suggests that the fraction of total rain attributable to stratiform rain would have been higher in GATE than Cheng and Houze (1979) estimated had their analysis not been limited by practical constraints to a portion of each day.

c. *Other measures of goodness of fit*

The quality of the descriptions of histogram variability can be examined in more detail by defining the "bin-by-bin" quantity

$$\mathcal{E}_0(r_i) \equiv \sum_t \frac{[n(r_i, t) - \sum_\alpha n_\alpha(t)p_\alpha(r_i)]^2}{N(t)p(r_i)}, \tag{3.6}$$

in terms of which $\mathcal{E}_0 = \sum_i \mathcal{E}_0(r_i)$. A graph of $\mathcal{E}_0(r_i)/(\sum_t 1)$ is shown in Fig. 6 for one-, two-, and three-mode expansions. As discussed at the beginning of this section, if the histograms were computed from spatially uncorrelated, independent samples and if $\mathcal{E}_0 \approx \mathcal{E}$, we would expect $\mathcal{E}_0(r_i)/(\sum_t 1) \approx 1$ for satisfactory fits. Histogram counts here are unfortunately not independent, and correlations tend to increase the value of $\mathcal{E}_0$. Nevertheless, with just two modes, values near 1 are

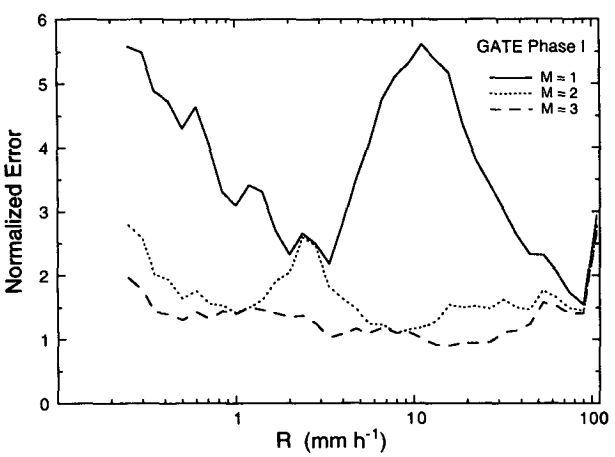

FIG. 6. A measure [Eq. (3.6)] of average error of description of histogram values for each rain-rate bin for one-, two-, and three-mode descriptions. For a satisfactory description, a value of 1 would be expected if the data were uncorrelated and the differences in the denominators of Eqs. (2.9) and (2.12) could be neglected.

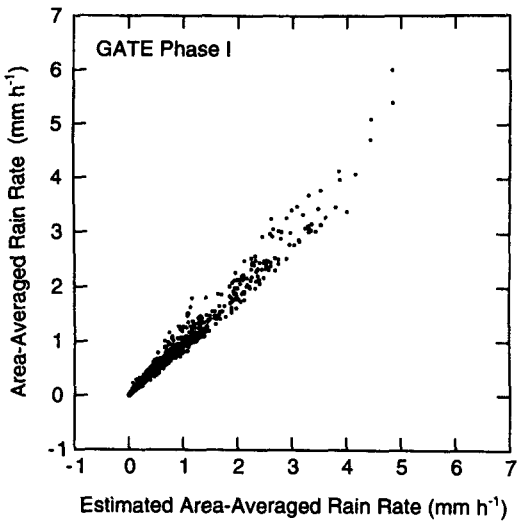

FIG. 7. Plot of radar-derived rain rate and value obtained from two-mode fit to histogram for each of 1716 maps, for 280 km × 280 km area in GATE phase I.

obtained over many rain-rate bins. The "transition" region from $\mathbf{p}_L$ to $\mathbf{p}_H$ at around 2–3 mm h^{-1} is not always captured adequately by the two-mode description, and there may be something to gain in this respect by adding a third mode.

Another measure of the adequacy of the description of the histograms can be obtained by comparing the average rain rate in a rain field with the value predicted from the fit to the rain-rate histogram. The true area-averaged rain rate is given by

$$\bar{R}(t) = \frac{1}{4900} \sum_{\mathbf{x}} R(\mathbf{x}, t), \tag{3.7}$$

and the estimate based on M modes by

$$\hat{\bar{R}}(t) = \sum_{\alpha=1}^{M} n_\alpha(t)\bar{r}_\alpha, \tag{3.8}$$

using (2.3) and (3.4). A plot of $\bar{R}(t)$ versus $\hat{\bar{R}}(t)$ is shown in Fig. 7 for the two-mode fits to the 1716 rain-rate fields of GATE phase I. The two-mode estimates and the actual area averages have a correlation of 0.99. Correlation values for M = 1, 2, and 3 are given in Table 3.

As another indication of how well the histograms are fit at each moment, $m_3(t)$ is plotted beneath n_L and n_H in Fig. 5. This is just the amplitude of the next term in the expansion (2.28) of $n(r_i, t)$. It is generally much smaller than $N(t) = n_H(t) + n_L(t)$ except during the few hours of day 188 discussed above. Of the 1216 fields with $N(t) > 100$ in GATE phase I, 77% have ratios $|m_3(t)/N(t)| < 0.2$.

d. GATE phase II

Data were also analyzed from the second phase of GATE, which extended from 24 July to 15 August 1974 and included 1512 gridded rain-rate maps. The results are close to those for Phase I. The first two modes explain 90.9% of the trace of the matrix diagonalized, and $\mathcal{E}_0/\Sigma_t\ 1 = 46$. Some characteristics of the two modes are given in Table 4. Cheng and Houze (1979) found that the fraction of rain volume attributable to stratiform rain dropped in GATE phase II to 40%, whereas we find an increase due to type L rain. The discrepancy may be due to the larger area and smaller time intervals centered around 1200 UTC that they analyzed, as mentioned earlier. Note that the sampling error is sufficiently large that the change in relative amounts contributed by the two types of rain from phase I to phase II is not statistically significant. Sampling error for the ratios found by Cheng and Houze (1979) are probably larger, since the additional area analyzed is probably not large enough to offset the reduction in sample size from the smaller time period analyzed. The confidence intervals found here for the relative areas and volumes of rain are therefore probably lower limits for the confidence intervals for the estimates of Cheng and Houze (1979).

TABLE 3. Correlation of true area-averaged rain rate [Eq. (3.7)] and estimates [Eq. (3.8)] based on fits to histograms using M modes, for phase I of GATE. Data for M = 2 are shown in Fig. 7.

M	Percent correlation	Percent variance explained
1	90.0	81.0
2	99.0	98.1
3	99.6	99.2

e. Including zero rain rates

Up to this point, the histogrammed data $n(r_i, t)$ have not included the number of grid points with no measurable rain in them. When a bin is added to count zero rain rates and the analysis repeated, it is found that the first three eigenvectors can be recombined into a mode $p_0(r_i)$ with nearly all its mass concentrated in the bin $r_0 = 0$, and into two modes corresponding to $\mathbf{p}_L$ and $\mathbf{p}_H$ obtained above. This is encouraging, since satellite instruments do not always distinguish rain from no-rain areas as clearly as radar, and the application of this method to satellite data will probably involve histograms with nonrainy events included.

TABLE 4. Characteristics of two principal modes based on GATE phase II data, as in Table 2. Uncertainties of relative areas and rain volumes are estimated to be ±0.25.

Mode	$\bar{r}_\alpha$ (mm h^{-1})	Correlation time (h)	Relative area	Relative rain volume
L ("stratiform")	3.4	13	0.74	0.48
H ("convective")	10.5	6	0.26	0.52

f. More than two modes

The need to use three modes when zero-rain-rate counts are included raises the general issue of employing more than two modes to describe rain-rate distributions. It is clear from the discussion of Fig. 6 that variations in the distribution of rain rates in the transition region 2–5 mm h^{-1} could be better described by using three modes instead of two. We find (zero rain rates not counted) that the first three eigenvectors can be combined into three modes, two of which are similar to the modes $\mathbf{p}_L$ and $\mathbf{p}_H$ already described but now peak at lower and higher rain rates, respectively. The third mode peaks at 2.5 mm h^{-1}. The number of instances of negative $n_\alpha(t)$ diminishes dramatically. There is, however, an element of subjectivity in the choices made that is best removed by using the modes obtained this way as a first guess to obtain modes from the constrained nonlinear minimization of (2.9).

4. Discussion

Some possible applications of the method developed here for describing the variation of rain distributions in terms of mixtures of component distributions, or principal modes of variation (PMVs) as they will be called, will be discussed next, as will some unresolved issues.

a. Diurnal cycle of rainfall

The original need for the method arose from a desire to characterize better the diurnal changes in rainfall statistics. A first use of the method was therefore to compare the diurnal variations in $\bar{R}(t)$ [Eq. (3.7)], $n_L(t)$, and $n_H(t)$. Bell and Reid (1993) describe a way to test for the presence of a diurnal cycle in a time series using the amplitude of a diurnal sinusoid fit to the entire time series and the sequence of amplitudes of separate fits of sinusoids to each day of the time series. The ratio of the square of the amplitude of the overall fit to the variance of the daily amplitudes determines the significance of the diurnal cycle. As expected, the diurnal amplitude of $n_H(t)$ is stronger than that of $\bar{R}(t)$, in the sense that the diurnal variation of $n_H(t)$ is significant at the $p = 0.017$ level, as compared to $p = 0.028$ for $\bar{R}(t)$. Spectral analysis bears this out, showing a stronger peak at frequency $(24\text{ h})^{-1}$ for $n_H(t)$ compared with $\bar{R}(t)$. The diurnal variation of $n_H(t)$ is well described by a sinusoid with amplitude ± 123 about a mean of 187 (grid points), peaking at 1550 UTC (1415 LT). Interestingly enough, the diurnal amplitude of $n_L(t)$ is small and does not pass a significance test, though the fit does peak several hours after the diurnal maximum of $n_H(t)$, consistent with the association of $n_H(t)$ and $n_L(t)$ with convective and stratiform activity, respectively. If this association is accepted, then the diurnal cycle in rainfall is mostly due to the convective component. The stratiform component, which contributes one-third to one-half of the total rainfall in GATE phase I (Table 2), tends to obscure the diurnal variability.

b. Stochastic rain model

Output from a space–time stochastic rain model developed to reproduce GATE statistics for satellite sampling studies (Bell et al. 1990) was analyzed using histograms of rain rates from a model field with the same size and resolution as that of the GATE data. The histograms do not decompose cleanly into two distinct types, and the correlation times of the amplitudes $m_\alpha(t)$ in the expansion (2.28) do not decrease with the rank α, in contrast with the GATE results. Since the stochastic model was not constructed to reproduce stratiform and convective types of rain, it is interesting to see how clearly the technique described here reveals this.

c. The ATI method

The decomposition of rain distributions into PMVs was used to investigate the success of the ATI method for estimating area-averaged rain rate using (1.1). The sensitivity of S_τ to varying amounts of stratiform and convective rain (or, more precisely, rain of type L and type H) in the area depends on the threshold τ. With the proper choice, S_τ becomes independent of the relative amounts of the two kinds of rain. An optimal threshold can be selected based on this approach and will be explored elsewhere.

d. Application to rain gauge data

Since rain gauge data are so plentiful, it is interesting to ask how such data might be analyzed with the method described here. The time series from a rain gauge must first somehow be converted into a series of histograms. An obvious way to do this would be to break the time series up into segments, possibly overlapping, and histogram each segment. Contemporaneous data from nearby rain gauges could be combined to increase the number of samples in each histogram. Since the histograms should probably contain at the very least on the order of 100 observations, many years of data for an isolated gauge would probably be necessary. The method would be best suited to analyzing seasonal variability in the distribution of rain rates rather than the hourly changes that could be examined here with radar data.

e. Labeling grid points

One of the results of PMV analysis is that, for each image with histogram $n(r_i, t)$, the number $n_\alpha(t)$ of grid points with rain of type α is found. It would clearly be of interest to be able to identify the type of rain present at each grid point. A step in that direction

would be to assign a probability $q_\alpha(\mathbf{x})$ that a grid point $\mathbf{x}$ is occupied by rain of type α, with $\sum_\alpha q_\alpha = 1$. One plausible way to assign these probabilities would be to assume that they are proportional to the PMVs for each type [i.e., $q_\alpha \propto p_\alpha(r)$, where r is the rain-rate bin into which the gridpoint rain rate falls]. The assignment

$$q_\alpha = \frac{n_\alpha(t)p_\alpha(r)}{\sum_{\beta=1}^{M} n_\beta(t)p_\beta(r)}$$

has the property that $\sum_{\mathbf{x}} q_\alpha(\mathbf{x}) = n_\alpha(t)$ if the PMV fit to the histogram $n(r_i, t)$ is good.

f. Other remarks

Satellite datasets, since they consist of sequences of images, are clearly amenable to analysis using the methods developed here. The histograms need not be confined to data from a single instrument or channel; that is, one set of bins could be assigned to one channel and another set to another channel.

Numerical schemes for obtaining principal modes when their number exceeds two or three will need refinement. Iterative methods using the principal component analysis described here as a starting point seem to be feasible. Minimization of $\mathcal{E}_0$ in (3.6), subject to the constraints (2.4), can be cast as a constrained least-squares problem and is therefore easier to treat than minimizing $\mathcal{E}$ itself; there is a large literature for solving such problems.

A number of other issues remain to be explored. Although the method does not seem to be very sensitive to bin-size choices nor to the number of bins used in our experiments, more experience with the method is needed to suggest objective criteria for making these decisions. The accuracy of the representation of the tails of the PMV distributions needs further exploration. Questions concerning how variable the estimates of PMVs are due to the smallness of the datasets used to derive them (i.e., sampling errors) need to be addressed.

A better choice may be possible for the quantity $\mathcal{E}$ in (2.9) that is minimized to find the PMVs. It was selected with common statistical practice for fitting distributions to data in mind, but the correlations in the data and the fact that histogram counts in some bins sometimes vanish as rain activity shifts from convective to stratiform weaken the arguments for this choice. It nevertheless has several valuable advantages. In its linearized form (2.12) the PMVs formed from linear combinations of the eigenfunctions are always capable of fitting the overall climatological rain-rate distribution exactly. Because of its denominator, errors in the fits contribute to the total error measure in linear proportion to the histogram count rather than proportionally to the square of the count, as would happen if a simpler least-squares criterion for the fits were used. Each histogram thus contributes to the error linearly according to the counts in it, so that a few histograms with a large number of counts do not control the results.

5. Conclusions

Decomposition of time-varying frequency distributions into sums of underlying distributions has a number of potential applications in precipitation research. PMVs generate a rapid and informative description of the types of rain in a space–time volume, useful when trying to characterize large radar- or satellite-derived datasets. They provide an objective means for describing changes in rainfall statistics that can be as interesting as descriptions of total rainfall in an area, as in the case of the diurnal cycle of rainfall. It may have application to the development of algorithms for remote sensing of rain that are based on matching probability distributions, such as have been described recently by Rosenfeld et al. (1993) and Wilheit et al. (1991), and to refinement of methods of estimating rainfall based on the area covered by rain.

Acknowledgments. Helpful remarks by E. Foufoula, D. A. Short, and M. Steiner are gratefully acknowledged.

APPENDIX

Details of Eigenvector Solution to Minimization Problem

A derivation of the vectors $\mathbf{x}_\beta$ that minimize the error measure (2.21) is given here. Expression (2.21) can be rewritten in terms of the matrix $\mathbf{C}$ defined in (2.22) as

$$\mathcal{E}_0 = \mathrm{Tr}(\mathbf{WCW}) - \sum_\beta \mathbf{x}'_\beta \mathbf{W}^2\mathbf{C}\mathbf{W}^2\mathbf{x}_\beta, \quad \text{(A.1)}$$

using the orthonormality property (2.17) assumed for $\mathbf{x}_\beta$. The vectors $\mathbf{x}_\beta$, constrained to be orthonormal, that minimize (A1) can be found by setting the derivative of

$$\mathcal{E}_0 - \sum_\beta \sum_\gamma \Lambda_{\beta\gamma}(\mathbf{x}'_\beta \mathbf{W}^2\mathbf{x}_\gamma) \quad \text{(A.2)}$$

with respect to $\mathbf{x}_\beta$ to zero, where $\Lambda_{\beta\gamma}$ are Lagrange multipliers fixed by the constraints (2.17). This yields the equation

$$\mathbf{W}^2\mathbf{C}\mathbf{W}^2\mathbf{x}_\beta = \sum_\gamma \Lambda_{\beta\gamma}\mathbf{W}^2\mathbf{x}_\gamma. \quad \text{(A.3)}$$

Clearly any subset of the eigenvectors $\psi_\beta = \mathbf{W}\mathbf{x}_\beta$ from (2.23) are solutions of (A.3), with the Lagrange multipliers fixed by the constraints (2.17) to take the values

$$\Lambda_{\beta\gamma} = \lambda_\beta \delta_{\beta\gamma}. \quad \text{(A.4)}$$

In order to minimize (A.1), which can be rewritten as

$$\mathcal{E}_0 = \mathrm{Tr}(\mathbf{WCW}) - \sum_\beta \lambda_\beta, \quad \text{(A.5)}$$

choose the subset of M eigenvectors with the largest eigenvalues, where M is the number of modes $\mathbf{p}_\alpha$ that will be formed from the solutions $\mathbf{x}_\beta$.

The first eigenvector of **WCW** can be shown to be

$$\psi_1(r_i) = [p(r_i)]^{1/2}. \tag{A.6}$$

To show this, write out explicitly the index summations implicit in the eigenvalue equation (2.23):

$$\lambda_1\psi_1(r_i) = \sum_j \frac{1}{[p(r_i)]^{1/2}}\left[\sum_t \frac{1}{N(t)} n(r_i, t)n(r_j, t)\right] \times \frac{1}{[p(r_j)]^{1/2}}\psi_1(r_j),$$

which becomes, with Ansatz (A.6),

$$\lambda_1[p(r_i)]^{1/2} = \frac{1}{[p(r_i)]^{1/2}} \sum_t \frac{1}{N(t)} n(r_i, t) \sum_j n(r_j, t).$$

Using the definition of $N(t)$ in (2.2) and of $p(r_i)$ in (2.11), (A.6) is confirmed, and the value of the eigenvalue is obtained:

$$\lambda_1 = \sum_t N(t). \tag{A.7}$$

REFERENCES

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

Bell, T. L., and N. Reid, 1993: Detecting the diurnal cycle of rainfall using satellite observations. *J. Appl. Meteor.*, **32,** 311–322.

——, A. Abdullah, R. L. Martin, and G. R. North, 1990: Sampling errors for satellite-derived tropical rainfall: Monte Carlo study using a space–time stochastic model. *J. Geophys. Res.*, **95D,** 2195–2205.

Cheng, C. P., and R. A. Houze, Jr., 1979: The distribution of convective and mesoscale precipitation in GATE radar echo patterns. *Mon. Wea. Rev.*, **107,** 1370–1381.

Chiu, L. S., 1988: Estimating areal rainfall from rain area. *Tropical Rainfall Measurements,* J. S. Theon and N. Fugono, Eds., A. Deepak Publishing, 361–367.

Cramér, H., 1946: *Mathematical Methods of Statistics.* Princeton University Press, 575 pp.

Gibbins, C. J., 1990: Rainfall rates, raindrop size distributions and millimetre-wave attenuations. *Proc. URSI Commission F, Open Symp. on Regional Factors in Predicting Radiowave Attenuation Due to Rain,* Rio de Janeiro, Union Radio Scientifique Internationale, 29–34.

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

Hudlow, M. D., and V. L. Patterson, 1979: GATE radar rainfall atlas. NOAA Special Report, 158 pp. [Available from U.S. Government Printing Office, Washington, D.C. 20402.]

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

——, L. S. Chiu, and G. R. North, 1990: Estimation of mean rain rate: Application to satellite observations. *J. Geophys. Res.*, **95,** 1965–1972.

Leary, C. A., and R. A. Houze, Jr., 1979: Melting and evaporation of hydrometeors in precipitation from the anvil clouds of deep tropical convection. *J. Atmos. Sci.*, **36,** 669–679.

Rosenfeld, D., D. B. Wolff, and D. Atlas, 1993: General probability-matched relations between radar reflectivity and rain rate. *J. Appl. Meteor.*, **32,** 50–72.

Sansom, J., and P. J. Thomson, 1992: Rainfall classification using breakpoint pluviograph data. *J. Climate,* **5,** 755–764.

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

——, K. Shimizu, and B. Kedem, 1993b: Optimal thresholds for the estimation of area rain-rate moments by the threshold method. *J. Appl. Meteor.*, **32,** 182–192.

Wilheit, T. T., A. T. C. Chang, and L. S. Chiu, 1991: Retrieval of monthly rainfall indices from microwave radiometric measurements using probability distribution functions. *J. Atmos. Oceanic Technol.*, **8,** 118–136.