Multivariate Space–Time Analysis of PRE-STORM Precipitation

ILYA POLYAK, GERALD R. NORTH, AND JUAN B. VALDES*

Climate System Research Program, College of Geosciences and Maritime Studies, Texas A&M University, College Station, Texas

(Manuscript received 21 January 1993, in final form 31 December 1993)

ABSTRACT

This paper presents the methodologies and results of the multivariate modeling and two-dimensional spectral and correlation analysis of PRE-STORM rainfall gauge data. Estimated parameters of the models for the specific spatial averages clearly indicate the eastward and southeastward wave propagation of rainfall fluctuations. A relationship between the coefficients of the diffusion equation and the parameters of the stochastic model of rainfall fluctuations is derived that leads directly to the exclusive use of rainfall data to estimate advection speed (about 12 m s^{-1}) as well as other coefficients of the diffusion equation of the corresponding fields.

The statistical methodology developed here can be used for confirmation of physical models by comparison of the corresponding second-moment statistics of the observed and simulated data, for generating multiple samples of any size, for solving the inverse problem of the hydrodynamic equations, and for application in some other areas of meteorological and climatological data analysis and modeling.

1. Introduction

Precipitation has assumed an increasingly prominent role in several current areas of research ranging from the mesoscale to the global scale. At the global scale, several satellite missions are currently in the fabrication phase [e.g., the Tropical Rainfall Measuring Mission (TRMM) as described by Simpson et al. (1988)]. The confirmation of algorithms for such effects as beam filling (e.g., Chiu et al. 1990; Short and North 1990), ground truthing, and sampling error analysis (e.g., Kedem et al. 1990; Shin and North 1988; Bell et al. 1990; North and Nakamoto 1989) requires some prior understanding of the statistical structure of rain fields. Most of the previous analyses have been based upon GATE [GARP (Global Atmospheric Research Program) Atlantic Tropical Experiment] data, a tropical oceanic rain-rate archive based upon gridded radar data (Bell 1987; Crane 1990). There is a need to study other datasets to ascertain whether there is a difference between the conclusions drawn from GATE-like analyses and those based upon data from other areas, such as subtropical or midlatitude land areas. One recent experiment provides data that can be used for such an analysis—the PRE-STORM (Preliminary Regional Experiment for STORM-Central) data from the Kansas–Oklahoma region. One study based upon PRE-STORM data has already been published (Graves et al. 1993). In the present study, we examine these data from the viewpoint of multivariate time series and investigate such features as the horizontal propagation of second-moment rain patterns. The horizontal motion of the precipitation systems is, of course, well known to meteorologists, but its significance has not yet been understood in the context of analyzing the error budget for the satellite observation problem nor as it regards reflection in estimates of the second moments.

The approach developed here makes simple assumptions, taking the fields to be governed by the multivariate autoregressive process of the first order. Ease of analytical manipulation, various estimation procedures, and physical interpretation of the numerical results follow from these assumptions. The intent of this paper is, first, to develop a strategy for spatial averaging of rainfall observations and, second, to explore the implications of these results for the estimation of the parameters of stochastic and physical models.

2. Data

For a dataset, the time series of rainfall recorded in the PRE-STORM field experiment were used. This experiment took place from May through June of 1985. Aggregated 5-min rainfall observations were obtained by the system of 42 Portable Automated Mesonet (PAM) stations, most of which were situated in Kansas, and 42 Stationary Automated Mesonet (SAM) stations situated in Oklahoma. PRE-STORM rain observations were described in detail by Meitin and Cunning (1985)

* Additional affiliation: Department of Civil Engineering, Texas A&M University, College Station, Texas.

Corresponding author address: Dr. Ilya Polyak, Climate System Research Program, College of Geosciences and Maritime Studies, Texas A&M University, College Station, TX 77843-3150.

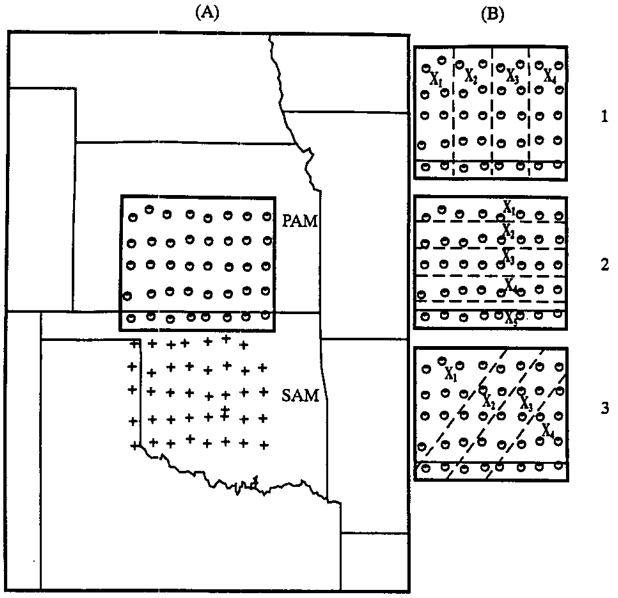

FIG. 1. (a) PRE-STORM mesonet sites over the central part of the United States. The circles indicate the PAM locations, and the crosses indicate the SAM sites. Panel (b) shows the specific regions for each spatial-averaging scheme for the PAM data.

and analyzed by Graves et al. (1993). Consideration in the present study is limited to rainfall data collected at 40 PAM stations as shown in Fig. 1. Missing and spurious data are replaced by zero. Each of the time series has a very special property: more than 99% of their terms are equal to zero.

For the preliminary analysis, some statistics were evaluated separately for May and June data. To compute the second moments, the methodology developed by Kagan (1979) for analysis of time series with missing data was used. The estimates for the point gauges of different stations vary significantly. Table 1 shows that the rains in June occurred, on average, almost two times more often than in May, and their mean duration was 1.5 times as long. A shortage of data in May is displayed by its higher variability, and it is expected that the accuracy of different June statistics must be higher than May statistics. As the values at the bottom of Table 1 show, spatial averaging (over all points) of rainfall observations leads to an increase of the first two autocorrelations.

TABLE 1. Mean characteristics of PRE-STORM rainfall data.

Characteristic	May	June
Mean number of onsets of rain	35	63
Mean duration of rain (h)	0.21	0.31
Mean interval between onsets of two consecutive rainfalls (h)	38	27
Mean relative frequency of observations with rain	0.004	0.008
Mean first autocorrelation coefficient (for 5-min lag)	0.52	0.66
Mean second autocorrelation (for 10-min lag)	0.43	0.44
First autocorrelation coefficient (for 5-min lag) of spatially averaged time series	0.69	0.81
Second autocorrelation coefficient (for 10-min lag) of spatially averaged time series	0.48	0.61

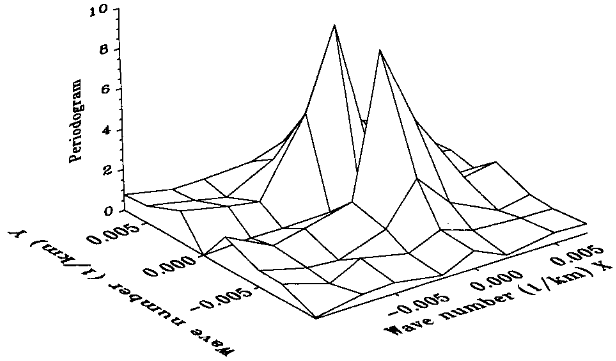

FIG. 2. Two-dimensional periodogram of standardized monthly sums of May rainfall field of PAM stations (X—north–south direction; Y—west–east direction).

3. Spectral and correlation analysis of monthly sums of rainfall fields

Sums of rainfall data were found for each point for May and June separately. The resulting two fields of monthly sums are samples of the two-dimensional random field. The advantage of such consideration is that there are no gaps in the monthly data and the standard statistical technique (Box et al. 1976; Christakos 1992) can be applied to their spectral analysis.

Figures 2 and 3 present the periodograms of the fields. The values have a random character, but a noticeable concentration of a significant power along the

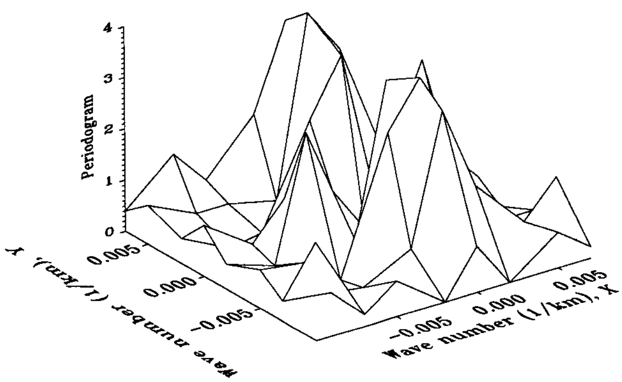

FIG. 3. Two-dimensional periodogram of standardized monthly sums of June rainfall field of PAM stations (X—north–south direction; Y—west–east direction).

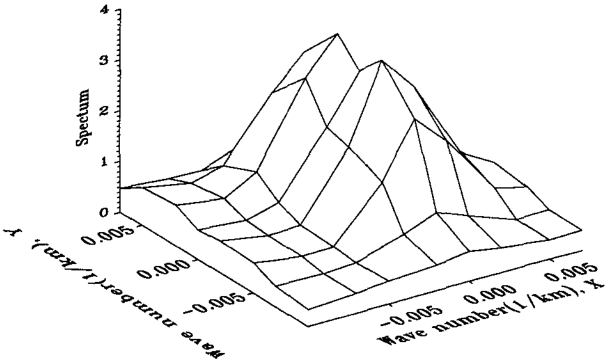

FIG. 4. Estimates of two-dimensional spectrum (of standardized monthly sums of May rainfall field of PAM stations), computed by smoothing the periodogram shown in Fig. 2 (X—north–south direction; Y—west–east direction).

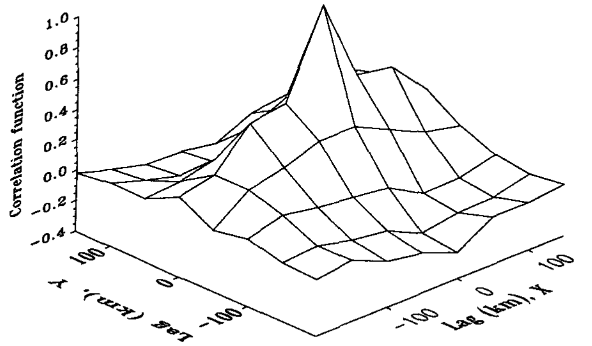

FIG. 6. Estimates of two-dimensional correlation function (of standardized monthly sums of May rainfall field of PAM stations), obtained by Fourier transform of the spectrum field shown in Fig. 4 (X—north–south direction; Y—west–east direction).

east–west axis can be seen. The estimation of the spectra was done by applying a digital filter, which is a two-dimensional generalization of a Tukey spectral window (Polyak 1975, 1979). Two-dimensional spectra (Figs. 4 and 5) show the main features of the wavenumber composition of these fields, but their resolutions are too low to reveal the details of the spectral structure. Only the general picture is clear: powerful westerly spatial propagation waves dominate all other waves. Meridional propagations are significantly smaller. Of course, the prevalence of eastward transport of rain clouds is expected, but its domination of two-dimensional spectra is not a trivial result.

Two-dimensional correlation functions (Figs. 6 and 7) show that for distances of about 50 km the space correlation coefficient can reach values of about 0.35–0.45. In spite of limitations in spatial resolution, the spatial dependence of monthly rainfall fluctuations is clear. Thus, the well-known synoptic eastward transport of rain clouds is reflected in the spectral representation of the second moments of monthly sums of rainfall. The above analysis reveals that the rain data employed are representative samples of a multivariate random process.

4. Algorithm for multivariate modeling

Fitting linear stochastic processes to the multiple time series (Kashyap and Rao 1976; Polyak 1989; Smith et al. 1985) leads to the creation of a stochastic model (SM) that allows us to describe the fluctuations of observed processes in terms of a system of linear differential (or, more exactly, difference) equations. That, in turn, provides the opportunity to make short-range forecasts, to analyze the interactions and feedbacks of the involved processes, to compare statistical and physical modeling, and to solve some other problems connected with the design and development of optimal filters. The most common stochastic models—autoregressive ones—are, in fact, approximations of stationary time series. The order of an autoregressive model corresponds to the order of the corresponding differential equations.

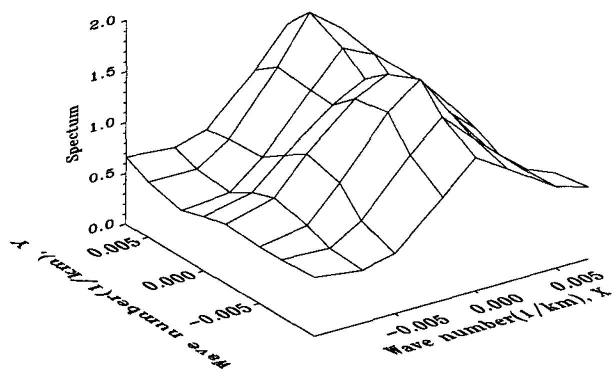

FIG. 5. Estimates of two-dimensional spectrum (of standardized monthly sums of June rainfall field of PAM stations), computed by smoothing the periodogram shown in Fig. 3 (X—north–south direction; Y—west–east direction).

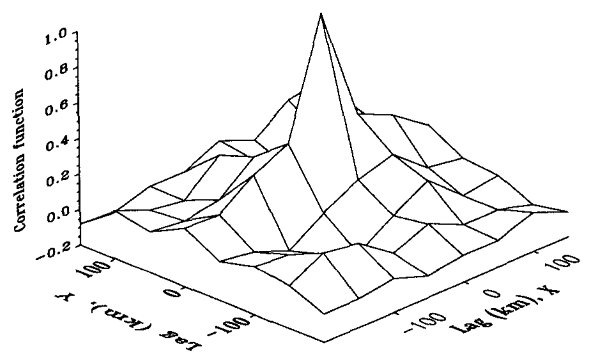

FIG. 7. Estimates of two-dimensional correlation function (of standardized monthly sums of June rainfall fields of PAM stations), obtained by Fourier transform of the spectrum field shown in Fig. 5 (X—north–south direction; Y—west–east direction).

Let us consider briefly an algorithm for building first-order multivariate autoregressive models. We will look at a q-variate random process

$$U_t = \{u_{it}\}_{i=1}^{q}, \quad (1)$$

where u_{it} ($i = 1, 2, \cdots, q$) are discrete stationary univariate random processes with zero expected values. All auto $v_{ii}(\tau)$ and cross $v_{ij}(\tau)$, $i\#j$ covariance functions of processes x_{it} are denoted

$$\mathbf{V}(\tau) = \{v_{ij}(\tau)\}_{i,j=1}^{q}. \quad (2)$$

The matrices $\mathbf{V}(\tau)$ ($\tau = 0, 1, 2, \cdots$) represent the matrix covariance function of a random process U_t.

The random process U_t is an autoregressive process of order p if it can be represented in the form

$$U_{t+1} = \mathbf{A}_1 U_t + \mathbf{A}_2 U_{t-1} + \cdots + \mathbf{A}_p U_{t-(p-1)} + N_{t+1}, \quad (3)$$

where $\mathbf{A}_i$ ($i = 1, 2, \cdots, p$) are the matrices of parameters, each one $q \times q$ in size, and N_{t+1} is a q-variate white noise, determined by the covariance matrix

$$\mathbf{E} = \{e_{ij}^2\}_{i,j=1}^{q}. \quad (4)$$

Identification of the order p and estimation of the elements of the matrices $\mathbf{A}_i$ and $\mathbf{E}$ from the observed q time series u_{it} represents the main practical problem of fitting the multivariate autoregressive model. In the case of the first-order model, the equation for estimation of the matrix of parameters $\mathbf{A}$ is

$$\mathbf{V}^{\mathrm{T}}(1) = \mathbf{A}\mathbf{V}(0). \quad (5)$$

If

$$\tilde{U}_{t+1} = \mathbf{A}U_t \quad (6)$$

is the forecast for one step ahead, then the covariance matrix of forecast errors is represented as

$$\mathbf{E} = \mathbf{V}(0) - \mathbf{A}\mathbf{V}(0)\mathbf{A}^{\mathrm{T}}. \quad (7)$$

If the model is built for the standardized variables x_i, then each variance e_{ii}^2 situated on the main diagonal of the matrix $\mathbf{E}$ cannot be greater than 1.

5. Multivariate stochastic rainfall models

Interpretations of the parameters of the multivariate stochastic models of spatially and temporally averaged data can provide the basis for a statistical description of a transport of rainfall fluctuations. To solve this problem, one must take into account a high level of white noise associated with the fluctuations measured at separate point gauges and attempt to reduce it. Signal detection may be enhanced through spatial averaging and time aggregation of observational data with consecutive fitting of multivariate autoregressive models to the resulting time series. The goal of this procedure is not, of course, forecasting in the sense of predicting the onset or the end of rainfall. What interests us is the diagnostic analysis of the matrices of estimated parameters of such models in terms of the interactions and feedbacks of observed processes. Such a stochastic model of rainfall fluctuations will be sensitive, in the general case, to choices made as to the regions of spatial averaging and the intervals of temporal aggregation.

Three different spatial-averaging strategies are illustrated in Fig. 1b. The first strategy—averaging in four longitude bands—is intended for the detection of west–east interdependence. The second—averaging in five latitude bands—is intended to explore the extent of meridional propagations of rainfall perturbations across latitude zones. The third strategy is useful for the analysis of southeastward transport. Five different temporal aggregation intervals were employed for comparison: 5, 15, 30, 45, and 60 min. Fifteen models were constructed from the combinations of spatial averaging and temporal aggregation assumptions. To reduce the seasonal contribution to rainfall perturbation, these models were built separately for May and June data. The most interesting results are presented here.

Each particular model is identified as model (number, month, aggregation), where "number" corresponds to one of the three strategies in Fig. 1b, "month" is May (m) or June (j), and "aggregation" is the aggregation interval. The time step between the two terms U_{t+1} and U_t determines the corresponding aggregation time. Matrices of the estimated parameters of all the models are given in Tables 2, 3, and 4. Each row of the tables with parameter estimates is associated with an autoregressive equation.

Consider, for example, model (1, m, 5), which is fitted to the indicated data, as shown in Table 2. The values listed in the columns e are the standard errors of the forecasts of the standardized observations for one step ahead. The closer the error is to 1, the worse the forecast. The autoregressive equations corresponding to model (1, m, 5) are shown below:

$$u_{1t+1} = 0.63u_{1t} + 0.09u_{2t} + 0.04u_{3t} + 0.00u_{4t}$$

$$u_{2t+1} = 0.23u_{1t} + 0.50u_{2t} + 0.08u_{3t} + 0.00u_{4t}$$

$$u_{3t+1} = -0.08u_{1t} + 0.12u_{2t} + 0.56u_{3t} + 0.05u_{4t}$$

$$u_{4t+1} = 0.00u_{1t} + 0.00u_{2t} + 0.00u_{3t} + 0.06u_{4t}.$$

We would expect to find a statistical dependence of the fluctuations of rainfall to the east upon the past rainfall fluctuations to the west. In the matrices of parameters for the various models, this statistical dependence should manifest in the closeness to zero of the parameter estimates above the main diagonals of the matrices. For May data—when aggregation intervals are small (5 and 15 min)—this effect is clear but not strong. For temporal aggregations greater than 15 min, however, either parameter values above main diagonals are zeros or the number of statistically significant estimates of above-main diagonals is less than the number below. For the models of June data the results are more

TABLE 2. Estimates (exceeding the 99% significance level) of parameters of four-variate rainfall autoregressive models for spatial-averaging scheme 1 (see Fig. 1b).

May					June				
Parameters				e	Parameters				e
(1, *m*, 5)					(1, *j*, 5)				
0.63	0.09	0.04	—	0.76	0.72	0.09	—	—	0.67
0.23	0.50	0.08	—	0.80	—	0.77	—	0.11	0.62
−0.08	0.12	0.56	0.05	0.79	−0.05	0.05	0.75	—	0.64
—	—	—	0.60	0.80	—	—	0.07	0.67	0.73
(1, *m*, 15)					(1, *j*, 15)				
0.49	—	0.13	−0.06	0.86	0.65	0.08	—	—	0.74
—	0.64	−0.23	—	0.78	0.09	0.60	—	—	0.78
—	−0.06	0.50	0.14	0.85	—	0.14	0.66	—	0.71
—	—	—	0.58	0.81	0.10	−0.14	0.21	0.56	0.77
(1, *m*, 30)					(1, *j*, 30)				
0.37	0.12	—	—	0.90	0.66	—	—	—	0.73
0.23	0.41	−0.14	—	0.86	0.24	0.49	—	—	0.81
—	0.27	0.42	—	0.83	—	0.21	0.63	—	0.71
—	—	0.18	0.68	0.68	0.09	−0.15	0.28	0.58	0.71
(1, *m*, 45)					(1, *j*, 45)				
0.21	0.18	—	−0.15	0.94	0.59	—	—	—	0.79
0.26	0.38	—	—	0.86	0.32	0.57	—	—	0.71
—	0.33	0.47	—	0.79	—	0.35	0.60	—	0.64
0.10	−0.15	0.35	0.64	0.65	—	—	0.40	0:52	0.68
(1, *m*, 60)					(1, *j*, 60)				
0.31	0.12	—	—	0.93	0.50	—	—	—	0.85
0.41	0.33	−0.15	—	0.80	0.26	0.58	−0.26	0.13	0.75
—	0.43	0.59	−0.18	0.58	0.14	0.34	0.54	—	0.68
0.22	−0.17	0.45	0.41	0.74	—	—	0.42	0.43	0.68

dramatic, especially for the 30- and 45-min aggregation intervals, when all the estimates above the main diagonals are zeros.

We now consider the five-variate models (Table 3) for zonal spatial averaging (see Fig. 1b, scheme 2). In this case, a large number of small and zero estimates above main diagonals would be evidence that the rainfall fluctuations from the north influence the rainfall fluctuations of regions to the south. The estimates in Table 3 do not suggest such an influence. On the whole, it seems that the numbers of zeros above and below main diagonals are approximately the same.

Models of the third type—based on diagonal band spatial averages (see Fig. 1b, scheme 3)—provide a clear quantitative picture of the strong dependence of rainfall fluctuations of any region upon the southeastward advection of rain. All (except one) of the parameters above the main diagonals of June models and most such parameters of May models are zeros. Rainfall fluctuations of each region with aggregation intervals larger than 15 min are partially determined by their own past (parameters on the main diagonals are the largest) and by the history of fluctuations of the closest region to the northwest (the values of the parameters of the lower second diagonals are not zeros). So, on average, southeastward transport of rain clouds in May and June of 1985 finds its reflection in rainfall statistics of the second moments when appropriate spatial averaging and temporal aggregation are used to detect the signal.

Consider the matrix autocorrelation function for the model (3, j, 30) of data spatially averaged over ten-gauge regions, as shown in Fig. 1b (scheme 3), and aggregated over 30-min time intervals (six sequential points). The estimates for the first two lags are represented in Table 5. Statistically significant estimates of space–time correlations reach values of 0.33 and 0.42, for example, and allow us to build a stochastic model.

The constructed models describe the behavior of rainfall fluctuations with different aggregations. They show that in May and especially in June, when it rained more often, such fluctuations were most strongly dependent upon advection from the northwest. The computed matrices of the parameters of multivariate autoregressive models for longitudinal and diagonal band averaging have a clear and stable triangular form with zero elements above main diagonals. A methodology of appropriate spatial averaging and the approximation of multiple time series by multivariate autoregressive models has the potential to detect the structure of second-moment rainfall fluctuations in spite of the significant noise associated with observations.

6. Autoregressive models and diffusion process

Motivation for the description of the rainfall fluctuations by the diffusion equation (as well as its connection with autoregressive models) was presented by North and colleagues (see, e.g., Cahalan et al. 1982; North and Nakamoto 1989). The relationship between stochastic models and a diffusive description of the rainfall field is explored below.

The stochastic models obtained above describe standardized *anomalies* of the observed processes. Spatial averaging of data reduced dimensionality and brought the rainfall field to the coordinate system (x, t). This case will be considered, but our reasoning will hold for the field of any dimension.

The diffusion equation is

$$\frac{\partial u}{\partial t} = s\frac{\partial^2 u}{\partial x^2} - v\frac{\partial u}{\partial x} - qu + f(x, t). \tag{8}$$

However, a multivariate model is the system of difference equations corresponding to a system of ordinary differential equations. In order to exploit this relation-

TABLE 3. Estimates (exceeding the 99% significance level) of parameters of five-variate rainfall autoregressive models for spatial-averaging scheme 2 (see Fig. 1b).

May						June					
Parameters					e	Parameters					e
		(2, m, 5)						(2, j, 5)			
0.64	−0.06	0.04	—	—	0.77	0.60	0.11	0.07	−0.09	−0.07	0.74
—	0.48	—	−0.08	0.12	0.83	—	0.72	0.07	—	—	0.66
—	0.09	0.55	0.04	—	0.82	—	0.08	0.71	—	—	0.68
0.04	−0.10	−0.04	0.54	0.08	0.82	−0.11	—	0.11	0.63	—	0.74
—	0.16	0.04	−0.07	0.55	0.76	—	—	0.08	—	0.68	0.71
		(2, m, 15)						(2, j, 15)			
0.39	—	—	0.16	—	0.90	0.37	0.19	0.09	—	—	0.86
0.15	0.51	—	—	—	0.82	0.16	0.45	0.17	—	—	0.77
—	0.32	0.28	—	−0.10	0.91	—	0.08	0.53	—	—	0.81
0.06	0.15	—	0.50	−0.16	0.87	−0.14	—	0.26	0.39	0.06	0.83
—	0.14	0.11	0.09	0.46	0.80	−0.09	0.08	—	—	0.44	0.89
		(2, m, 30)						(2, j, 30)			
0.34	—	0.10	0.28	—	0.86	0.20	0.27	0.18	—	0.11	0.86
0.30	0.34	—	—	0.11	0.82	0.36	0.27	0.36	—	—	0.65
—	0.67	0.17	0.14	−0.25	0.80	—	0.18	0.33	0.18	—	0.83
—	0.13	0.16	0.38	−0.12	0.90	—	—	0.17	0.39	—	0.86
0.17	0.24	0.12	0.15	0.19	0.85	—	—	—	—	0.32	0.94
		(2, m, 45)						(2, j, 45)			
0.27	—	0.12	0.21	0.15	0.88	0.46	0.15	—	—	0.15	0.79
0.44	0.41	—	—	—	0.80	0.33	0.34	0.25	—	0.14	0.65
—	0.98	0.15	—	−0.59	0.75	—	—	0.47	0.24	—	0.72
—	−0.21	0.35	0.20	0.14	0.90	−0.13	0.15	0.20	0.36	—	0.85
0.34	—	—	0.16	0.40	0.83	—	—	—	—	0.43	0.88
		(2, m, 60)						(2, j, 60)			
0.23	—	0.13	—	0.21	0.88	0.38	—	0.20	—	0.16	0.80
0.45	0.48	−0.14	0.12	−0.13	0.78	0.37	0.15	0.42	−0.15	0.12	0.67
0.23	0.47	0.23	—	−0.14	0.78	—	—	0.34	0.30	—	0.76
—	0.35	0.32	0.49	−0.47	0.78	—	—	0.24	0.45	—	0.79
0.35	—	—	0.15	0.27	0.85	−0.16	—	—	0.21	0.39	0.88

ship, we will need to approximate the diffusion equation by a system of ordinary differential equations. The straight lines method (Berezin and Jidkov 1960) achieves such an approximation as follows:

if

$$x = x_k = kh, (k = 0, 1, 2, \cdots); \quad (9)$$

then the replacement of the derivatives $\partial^2 u/\partial x^2$ and $\partial u/\partial x$ by the corresponding differences gives

$$\frac{du_k(t)}{dt} = \frac{s}{h^2}[u_{k+1}(t) - 2u_k(t) + u_{k-1}(t)]$$

$$- \frac{v}{2h}[u_{k+1}(t) - u_{k-1}(t)] - qu_k(t) + f_k(t). \quad (10)$$

This system of ordinary differential equations approximates the diffusion equation and corresponds to the autoregressive equations of the multivariate model. Different forms of the difference representation for the derivatives $\partial^2 u/\partial x^2$ and $\partial u/\partial x$ lead to the different coefficients of the differential equations. Thus, the procedure of estimation of the model parameters gives the optimal solution for the inverse problem of the differential equations.

The relationship between the parameters of the autoregressive model and the coefficients of the diffusion equation can be established as follows: let us replace the derivative du/dt in (10) by the simplest difference (l is the time step or the aggregation time). We have

$$u_{k(t+1)} = au_{(k-1)t} + bu_{kt} + cu_{(k+1)t} + lf, \quad (11)$$

where values a, b, c, l, h, q, s, and v are interconnected by the relationships

TABLE 4. Estimates (exceeding the 99% significance level) of parameters of four-variate rainfall autoregressive models for spatial-averaging scheme 3 (see Fig. 1b).

May					June				
Parameters				*e*	Parameters				*e*
(3, *m*, 5)					(3, *j*, 5)				
0.68	0.05	—	−0.05	0.73	0.77	—	−0.03	—	0.63
—	0.58	—	—	0.81	0.06	0.80	—	—	0.59
—	—	0.65	—	0.76	—	—	0.77	—	0.63
−0.12	−0.05	0.04	0.63	0.76	—	−0.06	—	0.69	0.71
(3, *m*, 15)					(3, *j*, 15)				
0.57	—	−0.11	—	0.81	0.61	—	—	—	0.78
—	0.45	0.06	—	0.88	0.16	0.64	—	—	0.71
—	0.07	0.49	0.06	0.86	−0.08	0.12	0.62	—	0.75
−0.14	—	0.14	0.52	0.81	—	−0.08	0.08	0.56	0.81
(3, *m*, 30)					(3, *j*, 30)				
0.46	—	—	—	0.88	0.55	—	—	—	0.85
0.22	0.43	0.12	—	0.85	0.27	0.59	—	—	0.69
0.23	0.19	0.44	—	0.82	—	0.25	0.49	—	0.81
—	—	0.22	0.59	0.73	0.12	−0.11	0.17	0.52	0.80
(3, *m*, 45)					(3, *j*, 45)				
0.25	—	—	−0.12	0.96	0.51	—	—	—	0.84
0.41	0.55	—	—	0.71	0.35	0.52	—	—	0.69
0.30	0.27	0.43	—	0.78	—	0.38	0.55	—	0.67
—	—	0.24	0.62	0.69	—	−0.14	0.34	0.52	0.76
(3, *m*, 60)					(3, *j*, 60)				
0.28	—	—	—	0.96	0.32	—	—	—	0.93
0.55	0.35	—	—	0.74	0.44	0.43	—	—	0.71
0.30	0.29	0.47	—	0.75	—	0.44	0.42	—	0.74
0.20	—	0.40	0.34	0.79	—	—	0.43	0.49	0.71

$$s = \frac{h^2}{2l}(a + c), \qquad (12)$$

$$v = \frac{h}{l}(a - c), \qquad (13)$$

$$q = \frac{1}{l}(1 - a - b - c). \qquad (14)$$

With sets of the estimates (a, b, c) of the parameters of the autoregressive equations, we can compute the coefficients of the diffusion equation—taking into account that these estimates present appropriate samples for consequential estimation of the coefficients of the diffusion equation. This makes sense when there is a real diffusive process in a given direction. This is the case in the present rainfall analysis for temporal aggregations of a half-hour or more and for eastward and southeastward horizontal atmospheric motions.

We assume that $h \approx 100$ km (the approximate distance between the centers of the regions of averaging) and $c = 0$. The value of b was taken as the mean of the four estimates on the main diagonal of the corresponding matrix of the parameters of the autoregressive model (Tables 2 and 4) for the regions averaged under schemes 1 and 3 (Fig. 1b). The value of a is the mean of the three estimates on the diagonal below the main diagonal of the same model. The results of such computations are shown in Table 6. The advection speed v varies between 34 and 48 km h^{-1} with a mean equal to 44 km h^{-1} (about 12 m s^{-1}). This result is a reasonable approximation of the mean speed of the synoptic fronts of the considered region of the United States. Also note that the value of the diffusion coefficient $k = s^{1/2}$ varies from 41 to 49 with mean equal to 47. It seems to us that among many considered statistics the estimates of v and k are the most stable characteristics of rainfall.

If the mean values in Table 6 are accepted as the estimates of the coefficients of the diffusion equation, we get the following:

TABLE 5. Matrix autocorrelation function for first two lags (0 and 30 min) of the model (3, *j*, 30).

	U_1	U_2	U_3	U_4
$\tau = 0$				
U_1	1.00	0.29	−0.11	0.20
U_2	0.29	1.00	0.17	−0.07
U_3	−0.11	0.17	1.00	0.08
U_4	0.20	−0.07	0.08	1.00
$\tau = 1$ (30 min)				
U_1	0.52	0.42	0.00	0.18
U_2	0.11	0.68	0.33	−0.09
U_3	−0.08	0.08	0.54	0.18
U_4	0.07	−0.04	0.01	0.56

TABLE 6. Estimates of the coefficients of the diffusion equation for different months, regions, and aggregation intervals *l* (h).

l	*a*	*b*	$s^{1/2}$	*v*	*q*
Region 1, May					
0.50	0.23	0.47	48	46	0.60
0.75	0.31	0.42	45	41	0.36
1.00	0.43	0.41	46	43	0.16
Region 1, June					
0.50	0.24	0.59	49	48	0.34
0.75	0.36	0.57	49	48	0.09
1.00	0.34	0.51	41	34	0.15
Region 3, May					
0.50	0.21	0.48	46	42	0.62
0.75	0.31	0.46	45	41	0.31
1.00	0.41	0.36	45	41	0.23
Region 3, June					
0.50	0.23	0.54	48	46	0.46
0.75	0.36	0.54	49	48	0.13
1.00	0.44	0.42	47	44	0.14
Mean	0.32	0.48	47	44	0.30

$$\frac{\partial u}{\partial t} \approx 47^2 \frac{\partial^2 u}{\partial x^2} - 44 \frac{\partial u}{\partial x} - 0.30u + f(x, t). \quad (15)$$

Therefore the multivariate stochastic model of the rainfall fluctuations is seen to be a system of ordinary differential equations that happens also to approximate the diffusion equation. The mean values of the coefficients of the diffusion equation can be determined from appropriate averaging and aggregation of systematic rainfall data. This methodology may help to identify the appropriate priorities for satellite monitoring programs in that it reveals possible uses of different kinds of atmospheric measurements and their empirical models.

7. Comments

The stochastic methodology applied in section 6 is similar to the methodology developed by Hasselmann (1988) and applied in a specific linear case by von Storch et al. (1988). However, discontinuity of the rainfall time series (and, as a result, irregular errors in the estimates of the second moments) and limited data volumes (and their high variability) made it inadvisable to complicate the analysis by consideration of empirical orthogonal functions. Attempts to optimally average discontinuous observations can lead to physically uninterpreted results.

The approximation of the partial differential equation that generates the fluctuation field by multivariate SM is a powerful methodology with a wide range of meteorological applications, including diagnostic analysis, signal detection, and GCM confirmation. The model's core (matrix of parameters) carries a physical meaning; it determines a statistical law (of fluctuations in behavior) that summarizes the logically necessary outcome of complex interactions and feedbacks. This law is mathematically and physically derived and empirically tested. Reliable estimation of each parameter a_{ij} allows us to determine the contribution of fluctuations u_{jt} to the variability of $u_{i(t+1)}$.

From the analysis of sets of time series of the different processes characterizing local or global rainfall, it is possible to estimate their contribution to one another's variability. Moreover, matrices of estimated parameters of the multivariate SM contain all information about the second moments in a concise and physically interpreted form. The comparison of several such models allows us to make statistical inferences about the similarity and differences among observed processes in a straightforward way.

The above results have additional uses: for example, qualitative comparison with analogous estimates from the TRMM or other kinds of rain observations, or generation of multiple samples of any size. They can likewise be used for the confirmation of simulations of the physical models. Our approach can help to solve the inverse problem of the hydrodynamic equations that has important implications for the identification of the number of variates (and their types) and the order of stochastic models, for the choice of appropriate regions for the spatial averaging of time series, for the comparison of different statistical approximations, and for the analysis of physical regularities and patterns. The multivariate SM accumulates both physical and statistical qualities in accordance with the requirements of the system identification theory.

8. Conclusions

The methodologies and results presented show that the formalization of the statistical description of rainfall fluctuations can be achieved by applying a standard technique of multivariate modeling and multidimensional spectral and correlation analysis to data averaged spatially and temporally. Two-dimensional spectra and correlation functions as well as the structure of the matrices of the parameters of multivariate autoregressive models paint a clear and stable picture of the eastward and southeastward wave propagation of rainfall fluctuations. A relationship between coefficients of the diffusion equation and parameters of the stochastic model was found that has led, for example, to the estimation of the advection speed (about 12 m s^{-1}) of the corresponding fields using only rainfall data.

Acknowledgments. Research for this paper was made possible by NASA Grants NAGW 2427 and 6896. We wish to thank Larry Goldberg for helpful discussions and editing the final version of the paper.

REFERENCES

Bell, T. L., 1987: A space–time stochastic model of rainfall for satellite remote-sensing studies. *J. Geophys. Res. Atmos.*, **92**(D8), 9631–9640.

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

Berezin, I. S., and N. P. Jidkov, 1960: *Numerical Methods.* Vol. 2. Physical and Mathematical Publishing House, 620 pp. (in Russian).

Box, G. E., and G. M. Jenkins, 1976: *Time Series Analysis.* Holden-Day, 575 pp.

Cahalan, R. F., D. A. Short, and G. R. North, 1982: Cloud fluctuation statistics. *Mon. Wea. Rev.*, **110,** 26–43.

Chiu, L. S., G. R. North, A. Short, and A. McConnel, 1990: Rain estimation from satellites: Effect of finite field of view. *J. Geophys. Res.*, **95,** 2177–2185.

Christakos, G., 1992: *Random Field Models in Earth Sciences.* Academic Press, 474 pp.

Crane, R. K., 1990: Space–time structure of rain rate fields. *J. Geophys. Res.*, **95,** 2011–2020.

Graves, C. E., J. B. Valdes, S. S. P. Shen, and G. R. North, 1993: Evaluation of sampling errors of precipitation from space-borne and ground sensors. *J. Appl. Meteor.*, **32,** 374–385.

Hasselmann, K., 1988: PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.*, **93,** 11 015–11 021.

Kagan, R. L., 1979: *Averaging of the Meteorological Fields.* Hydrometeoizdat, 214 pp. (in Russian).

Kashyap, R. L., and A. Rao, 1976: *Dynamic Stochastic Models from Empirical Data.* Academic Press, 344 pp.

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

Meitin, J., and J. Cunning, 1985: The Oklahoma–Kansas Preliminary Regional Experiment for STORM-Central. Volume I: Daily operations summary. NOAA Tech. Memo. ERL ESG-20, 313 pp.

North, G. R., and S. Nakamoto, 1989: Formalism for comparing rain estimation designs. *J. Atmos. Oceanic Technol.,* **6,** 985–992.

Polyak, I. I., 1975: *Numerical Methods for Data Analysis.* Hydrometeoizdat, 211 pp. (in Russian).

——, 1979: *Methods for Random Processes and Fields Analysis.* Hydrometeoizdat, 254 pp. (in Russian).

——, 1989: *Multivariate Stochastic Models of Climate.* Hydrometeoizdat, 184 pp. (in Russian).

Shin, K. S., and G. R. North, 1988: Sampling error study for rainfall estimates by satellite using a stochastic model. *J. Appl. Meteor.,* **11,** 1218–1231.

Short, D. A., and G. R. North, 1990: The beam filling error in the Nimbus 5 electronically scanning microwave radiometer observation of Global Atlantic Tropical Experiment rainfall. *J. Geophys. Res.,* **95,** 2187–2193.

Simpson, J., R. Adler, and G. R. North, 1988: A proposed Tropical Rainfall Measuring Mission (TRMM) satellite. *Bull. Amer. Meteor. Soc.,* **69,** 278–295.

Smith, J. A., and A. F. Karr, 1985: Parameter estimation for a model of space time rainfall. *Water Resour. Res.,* **21**(8), 1251–1258.

Storch, H. von, T. Bruns, I. Fisher-Bruns, and K. Hasselmann, 1988: Principal oscillation patterns analysis of the 30- to 60-day oscillation in general climate model equatorial troposphere. *J. Geophys. Res.,* **93,** 11 022–11 036.