

# A Hydrologically Useful Station Precipitation Model

## 2. Case Studies

KONSTANTINE P. GEORGAKAKOS

*Hydrologic Research Laboratory, National Weather Service, NOAA, Silver Spring, Maryland*

RAFAEL L. BRAS

*Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge*

A physically based model of precipitation at a point, developed by the authors in an accompanying paper, is used to make rainfall and snow predictions at locations in the United States and Venezuela. The model is utilized in off-line and real-time modes. The latter involves the use of a Kalman filter. Results indicate that the parameterization suggested here is quite robust, comparing favorably with locally calibrated linear regression alternative models. Ideas for future work are given.

### INTRODUCTION

In an accompanying paper, *Georgakakos and Bras* [this issue] developed a rainfall model based on a simplified formulation of cloud physics. The model considers the atmosphere as a reservoir of moisture. The outputs from the reservoir are small hydrometeors blown off the top of the clouds by ascending winds and those hydrometeors large enough to counteract updraft velocities and exit at cloud bottom. Input to the reservoir is the pseudo-adiabatic condensation resulting from moisture-laden air cooling while rising through the atmosphere. The state equation describing the system is given by *Georgakakos and Bras* [this issue] as

$$\frac{dX}{dt} = f(\mathbf{u}, \mathbf{a}_I) - h(\mathbf{u}, \mathbf{a}_0)X \quad (1)$$

where  $X$  is the liquid water content in the cloud system. The nonlinear function  $f(\mathbf{u}, \mathbf{a}_I)$  represents moisture input. The function  $h(\mathbf{u}, \mathbf{a}_0)$  is also nonlinear in variables  $\mathbf{u}$  and  $\mathbf{a}_0$  but acts linearly on the state  $X$ . The vector of meteorological inputs  $\mathbf{u}^T$  is

$$\mathbf{u}^T = [T_0, p_0, T_d] \quad (2)$$

where  $T_0$  is ambient air temperature at ground level,  $p_0$  is the ground-level pressure, and  $T_d$  is the dew-point temperature, also at ground level. There are three main parameters, included in vectors  $\mathbf{a}_I$  and  $\mathbf{a}_0$

$$\mathbf{a}_I^T = [p, v] \quad (3)$$

$$\mathbf{a}_0^T = [p, v, c] \quad (4)$$

where  $p$  is pressure at the top of the cloud,  $v$  is updraft wind velocity, and  $c$  is the inverse of the average hydrometeor diameter, a parameter of the particle distribution function.

The output at cloud base suffers evaporation before reaching the ground as precipitation. *Georgakakos and Bras* [this issue] show that this can be represented by

$$P = \Phi(\mathbf{u}, \mathbf{a}_0)X \quad (5)$$

where  $\Phi(\cdot)$  is a nonlinear function of  $\mathbf{u}$  and  $\mathbf{a}_0$  acting linearly on the state  $X$ .

Equations (1) through (5) form a linear system of determin-

Copyright 1984 by the American Geophysical Union.

Paper number 4W1076.  
0043-1397/84/004W-1076\$05.00

istic equations. It is possible to hypothesize errors in the state dynamics, in the input variables as well as in the precipitation equation. If such errors are taken as additive, the system becomes a classic stochastic differential equation formulation,

$$\frac{dX(t)}{dt} = f(\mathbf{u}, \mathbf{a}_I, t) - h(\mathbf{u}, \mathbf{a}_0, t)X(t) + \Gamma(t)W(t) \quad (6)$$

$$P(k) = \Phi(\mathbf{u}, \mathbf{a}_0, k)X(k) + V(k) \quad (7)$$

where  $k$  is a time index,  $t = k\Delta t$ . The term  $\Gamma(t)W(t)$  is a continuous stochastic error with zero mean and variance density  $Q(t)$  such that

$$Q(t)\delta(t - \tau) = E[\Gamma(t)W(t)\Gamma(\tau)W(\tau)]$$

In the above,  $\delta(t)$  is the Dirac delta function. The error  $V(k)$  in (7) corresponds to a discrete precipitation measurement error with zero mean and variance:

$$E[V(k)V(l)] = \begin{cases} R(k) & k = l \\ 0 & k \neq l \end{cases}$$

The input error component of  $Q(t)$  and  $R(k)$  is additive and is computed by linearization of the model  $f(\cdot)$ ,  $h(\cdot)$ , and  $\Phi(\cdot)$  functions with respect to the input  $\mathbf{u}$ . The details of the formulation are given in *Georgakakos and Bras* [1982b].

Given the stochastic system of (6) and (7), it is possible to use the continuous-discrete Kalman filter [*Gelb*, 1974] in order to combine noisy state dynamics and observations in making predictions of the future. The filter, in this case, takes the form

Prediction

$$\frac{d\hat{X}(\tau|t)}{d\tau} = f(\mathbf{u}, \mathbf{a}_I, \tau) - h(\mathbf{u}, \mathbf{a}_0, \tau)\hat{X}(\tau|t) \quad (8)$$

$$\frac{d\sum(\tau|t)}{d\tau} = h(\mathbf{u}, \mathbf{a}_0, \tau)\sum(\tau|t) + \sum(\tau|t)h(\mathbf{u}, \mathbf{a}_0, \tau) + Q(\tau) \quad (9)$$

Update

$$\hat{X}(k|k) = \hat{X}(k|k-1) + K(k)[P(k) - \Phi(\mathbf{u}, \mathbf{a}_0, k)\hat{X}(k|k-1)] \quad (10)$$

$$\sum(k|k) = [1 - K(k)\Phi(\mathbf{u}, \mathbf{a}_0, k)]\sum(k|k-1) \quad (11)$$

TABLE 1. Storm Data Statistics 1

Storm Group	Average Precipitation Rate, mm/h	Accumulated Precipitation, mm
1	2.59	259
2	1.585	95
3	0.950	62.7
4	2.193	131.6
5	1.831	183.1

$$K(k) = \sum (k|k-1)\Phi(\mathbf{u}, \mathbf{a}_0, k) \cdot [\Phi(\mathbf{u}, \mathbf{a}_0, k)\sum (k|k-1)\Phi(\mathbf{u}, \mathbf{a}_0, k) + R(k)]^{-1} \quad (12)$$

Equations (8) through (12) need initial conditions  $\sum (0|0)$  and  $\hat{X}(0|0)$  to initiate the iterative solution. In the equations,  $\hat{X}(\tau|t)$  stand for the best estimate of the state at time  $\tau$  given information up to time  $t$ . The variance of estimation is similarly given by  $\sum (\tau|t)$ . Predictions can be made by integrating (8), starting from the last updated estimate of the state. Integration of (9) yields the variance of the prediction. In order to make predictions and integrate (8) and (9), values of inputs  $T_0$ ,  $p_0$ , and  $T_d$  are required, they are obtained by observation or prediction. Similarly, the parameters  $p_i$ ,  $v$ , and  $c$  must be determined. Therefore, before going into particular examples to illustrate the prediction capabilities of the model, the paper will address the model parameterization and the nature of its inputs.

#### PRECIPITATION MODEL PARAMETERS

The following have been identified as model physical parameters: the vertically averaged updraft velocity in the unit area column,  $v$ ; the terminal pressure level of the column,  $p_i$ ; and the average water equivalent particle diameter  $1/c$  at the cloud base. It is desirable to express the model equations in terms of parameters that are storm invariant so that robust parameter estimates can be obtained.

#### Updraft Velocity and Terminal Pressure Parameterization

Work by Sulakvelidze [1969] indicates that  $v$  obeys a law of the type

$$v = \varepsilon_1 \cdot (c_p \cdot \Delta T)^{1/2} \quad (13)$$

where

$$\Delta T = T_m - T_s' \quad (14)$$

In the previous equations,  $\varepsilon_1$  is a constant parameter,  $c_p$  is the specific heat of dry air under constant pressure (Joule/(kg · °K)),  $T_m$  is the cloud temperature (°K) at a certain level  $p'$  (mbar) assuming pseudo-adiabatic ascent and  $T_s'$  is the corresponding ambient air temperature (°K). The quantity  $\varepsilon_1$  is analogous to the ratio of kinetic to thermal energy per unit

TABLE 3. Storm Data Statistics 3

Storm Group	Cross-Correlations to Precipitation Rate			
	$T_0$	$p_0$	$T_d$	$T_w$
1	-0.066	0.087	0.170	0.086
2	0.315	-0.101	0.392	0.347
3	0.185	0.204	0.340	0.283
4	0.253	-0.228	0.337	0.303
5	-0.474	0.161	-0.394	-0.431

mass of ascending air at the level  $p'$ . Therefore,  $\varepsilon_1$  is dimensionless.

Due to the difficulty of obtaining radiosonde data in real time,  $T_s'$  is taken as the temperature at level  $p'$  that results from dry-adiabatic ascent. The pressure level  $p'$  is taken where the updraft velocity is equal to the height-averaged  $v$ . Given the assumed vertical distribution of updraft velocity in Figure 2 of Georgakakos and Bras [this issue]:

$$p' = p_s - \frac{1}{4}(p_s - p_i) \quad (15)$$

with  $p_s$  and  $p_i$  defined as the pressures at the cloud base and cloud top, respectively.

The temperatures  $T_s'$  and  $T_m$  obey the constant-adiabat and pseudo-adiabat equations (7) and (16) in Georgakakos and Bras [this issue]

$$T_s' = \frac{T_0}{p_0^{0.286}} \left( \frac{3}{4} p_s + \frac{1}{4} p_i \right)^{0.286} \quad (16)$$

and

$$T_m \left( \frac{p_n}{p'} \right)^{0.286} \cdot \exp \left\{ \frac{L(T_m) \cdot w_s(T_m, p')}{c_p \cdot T_m} \right\} = \theta_e \quad (17)$$

There is then an implicit relationship between  $p_i$  and  $v$ .

Independently, and based on observations of the development of storm clouds, another equation relating  $p_i$  and  $v$  is suggested. It is based on the well-known fact that the stronger the updraft, the more vigorous the storm cloud development and, consequently, the lower  $p_i$  is. However, the value of  $p_i$  also depends on the past history of the storm. Thus as the storm persists for several hours, even for low updraft velocities,  $p_i$  is expected to be relatively low (e.g., in precipitation processes from stratiform clouds). Since no information of the lifetime of the storm system before reaching the drainage basin boundaries is assumed, the new  $p_i$  versus  $v$  relationship is parameterized as follows:

$$\frac{p_i - p_l}{\varepsilon_2 - p_l} = \frac{1}{1 + \varepsilon_3 \cdot v} \quad (18)$$

where  $p_l$  is the lowest value that  $p_i$  can attain, and  $\varepsilon_2$  and  $\varepsilon_3$  are constant parameters.

TABLE 2. Storm Data Statistics 2

Storm Group	Coefficient of Variation					Skewness Coefficient				
	$T_0$	$p_0$	$T_d$	$T_w$	$P_v$	$T_0$	$P_0$	$T_d$	$T_w$	$P_v$
1	0.005	0.008	0.007	0.006	1.066	-0.216	-0.710	-1.106	-0.460	1.691
2	0.002	0.007	0.003	0.002	1.157	0.139	1.125	-0.578	-0.467	1.303
3	0.005	0.009	0.006	0.005	1.266	-0.489	0.746	-0.547	-0.483	2.926
4	0.012	0.009	0.011	0.011	1.418	1.707	-0.614	1.205	1.590	3.061
5	0.012	0.002	0.010	0.010	1.609	-0.141	-0.034	-0.221	-0.198	2.53

TABLE 4. Storm Data Statistics 4: Group 1 Autocorrelations

Variable	Lag, hours					
	1	2	3	4	5	6
$T_0$	0.928	0.840	0.738	0.626	0.513	0.411
$p_0$	0.976	0.945	0.908	0.867	0.823	0.775
$T_d$	0.914	0.814	0.706	0.611	0.525	0.460
$T_w$	0.928	0.837	0.746	0.656	0.554	0.473
$P_v$	0.641	0.441	0.313	0.188	0.066	0.031

Parameter  $p_r$  can be set equal to the pressure value at the troposphere-stratosphere boundary, since very few storms penetrate into the stratosphere. That is about  $p_t = 200$  mbar. Parameter  $\varepsilon_2$  has dimensions of pressure and depends on the history of the storm before it reaches the basin boundaries. Parameter  $\varepsilon_3$  has dimensions of inverse velocity and controls the  $p_t$  versus  $v$  relationship. Note that as  $v$  tends to zero,  $p_t$  tends to  $\varepsilon_2$ , and as  $v$  tends to infinity,  $p_t$  tends to  $p_b$  in agreement with the qualitative arguments described above. Georgakakos and Bras [1982b] present a sensitivity analysis of the implicit set of (13) through (18) with respect to the parameters  $\varepsilon_1$  and  $\varepsilon_3$  and the physical quantities  $v$  and  $p_t$ . Those results can be used for initial parameter estimation when observations of cloud top pressure or height are available.

#### Parameterization of Average Hydrometeor Diameter at Cloud Base

Several processes contribute to determine the parameter  $c$ . Based on past work, Pruppacher and Klett [1978] identify as most important: the condensation, the collision coalescence, and the collisional breakup of the larger particles. They indicate that the stronger the updraft velocity, the larger the number of the larger particles. That implies increasing the average hydrometeor diameter ( $\frac{1}{c}$ , see Georgakakos and Bras [this issue]) as  $v$  increases. In addition, based on theoretical work, they suggest that even a mild updraft (e.g., of  $v$  equal to 0.10 m/s) has a pronounced effect on the particle distribution. For purposes of this work it is assumed that  $c$  is

solely determined by  $v$  from a relationship of the type

$$\frac{1}{c} = \varepsilon_4 \cdot v^k \quad (19)$$

where  $\varepsilon_4$  and  $k$  are constant parameters. The dimensions of  $\varepsilon_4$  are  $s^k m^{(1-k)}$ , while  $k$  is dimensionless.

In summary, the storm invariant model parameters are  $\varepsilon_1$ ,  $\varepsilon_2$ ,  $\varepsilon_3$ ,  $\varepsilon_4$ , and  $k$ . Also needed are the value of the ratio of the upper cloud inverse mean particle diameter to that of inverse mean particle diameter at the bottom of the cloud,  $\gamma$ , and the parameter  $\beta$  of the updraft velocity distribution which is the ratio of the updraft velocity at the top or bottom of the cloud column to the height averaged updraft velocity [see Georgakakos and Bras, this issue].

#### MODEL INPUTS

One of the main arguments for the proposed model is that its inputs are readily available variables that lend themselves to statistical analysis and predictions. To illustrate the above point, data from eight storms in the Boston, Massachusetts, area (Logan airport) and three storms in Tulsa, Oklahoma (International airport) were studied. The storms included convective line storms, a tropical storm, large low-pressure centers of persistent rainfall northeasterlies, snowstorms, and thunderstorms with long dry periods in between. Georgakakos and Bras [1982b] describe the storms in more detail. In order to facilitate analysis the storms were divided into five different groups. Group 1 was a convective storm and a tropical storm (Daisy) in Boston for a total of 100 hours of rainfall. Group 2 was a 60-hour-long cyclonic storm in Boston. Group 3 consists of three winter snowstorms in Boston for a total of 66 hours. Group 4 includes two heavy northeasterlies over Boston for a total of 60 hours of heavy rainfall. Three thunderstorms in Tulsa, Oklahoma, form the 100-hour-long group 5. All groups included significant, sometimes record-setting events. The average precipitation and accumulated precipitation in all groups are shown in Table 1. All records consisted of hourly data of  $T_0$ ,  $p_0$ , and  $T_d$  as well as wet bulb temperature  $T_w$  and precipitation  $P$ . The units of all data are degrees Kelvin for temperatures Pascals ( $\text{kg m}^{-1} \text{s}^{-2}$ ) for pressures,

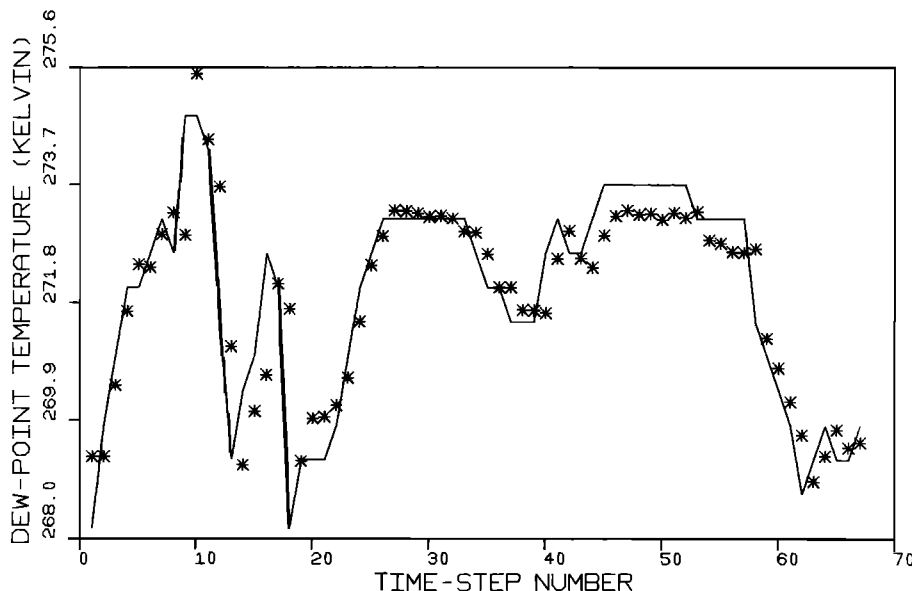


Fig. 1. Dew-point regression hourly predictions (stars) vs. observations (solid line) for storm group 3.

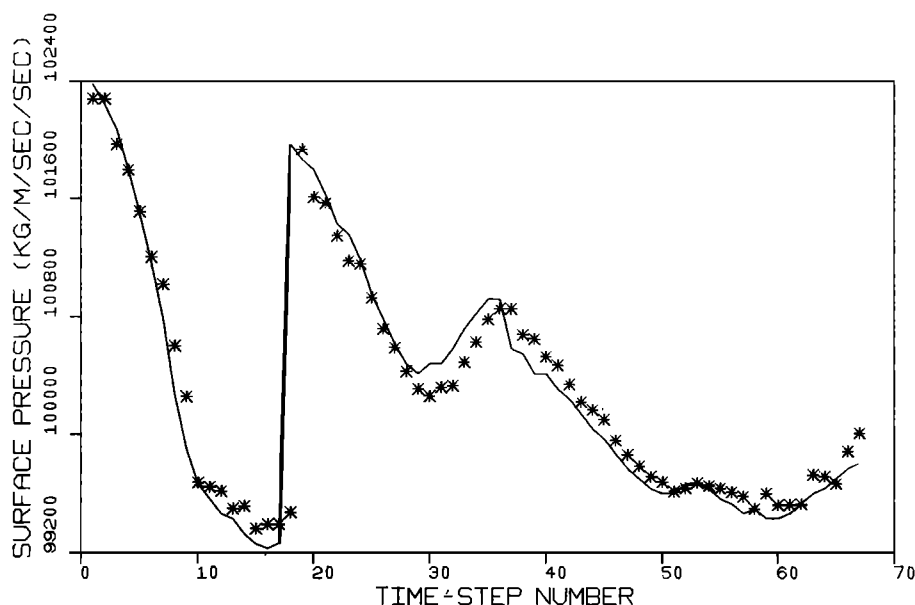


Fig. 2. Surface pressure regression hourly predictions (stars) vs. observations (solid line) for storm group 3.

and mm/hr for precipitation. Table 2 gives the coefficient of variation and skewness of all variables in each group. Table 3 gives the correlation of precipitation with the temperatures and pressure in each group. Table 4 shows the autocorrelation up to lag 6 for the variables in group 1. Group 1 exhibited the most significant autocorrelations of precipitation for all groups. The lag 1 autocorrelation of precipitation for group 2 was 0.408; for group 3, 0.6; for group 4, 0.372; and for group 5, 0.609. Group 1 also had the slowest decay of the precipitation autocorrelation. *Georgakakos and Bras [1982b]* give the complete set of results. The statistical analysis yields the following:

1. The scale of fluctuations of the precipitation variable, as it is expressed by the coefficient of variation (ratio of standard deviation to the mean), is at least two orders of magnitude

greater than those of the temperature and pressure variables for all storm groups.

2. Strong positive skewness is characteristic of the precipitation rate. The skewness coefficient of the other variables were both negative and positive.

3. Low cross correlations of the temperatures and of the pressure to the concurrent hourly precipitation rate were observed, ranging from 0.066 to 0.474 in absolute value. This highlights the difficulty of using the temperatures and pressure as explanatory variables in a linear regression for the precipitation rate prediction.

4. Characteristically high lag-1 (1 hour) auto correlations were obtained for the temperatures and the pressure variables. They ranged from about 0.8 to about 0.98, suggesting that the current value of those variables contains considerable amount

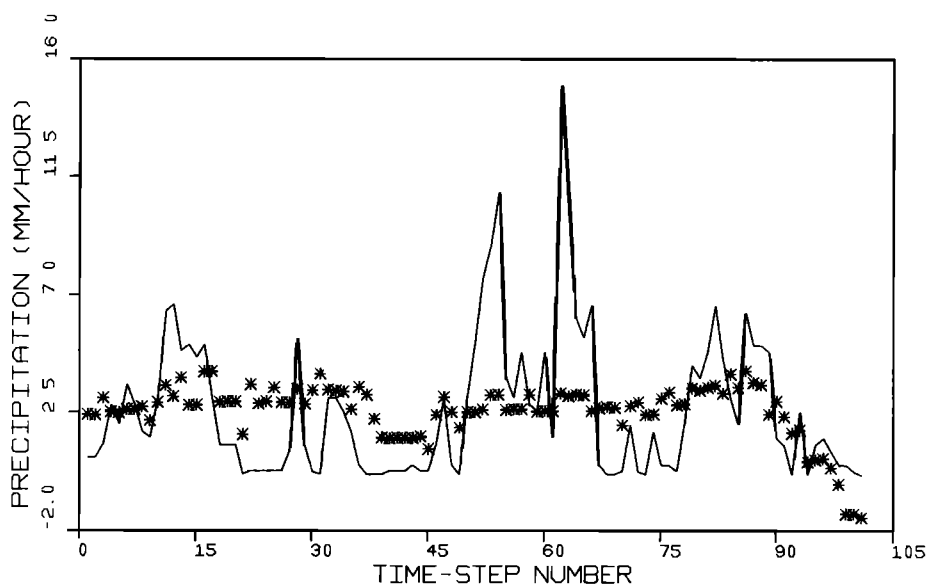


Fig. 3. Precipitation rate hourly predictions (stars) based on linear regression model vs. observations (solid line). Storm group 1. Model does not include lagged precipitation variables.

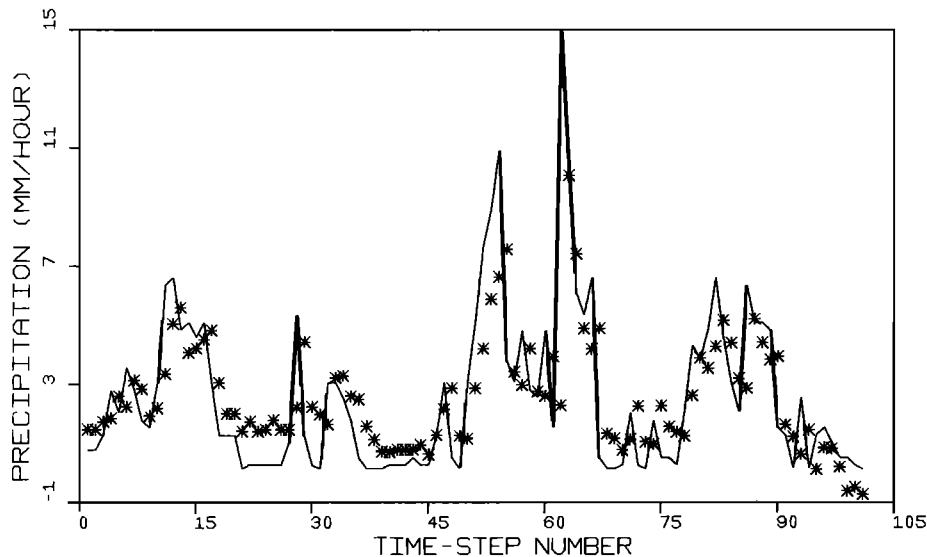


Fig. 4. Precipitation rate hourly predictions (stars) based on the linear regression model vs. observations (solid line). Storm group 1. Model includes lagged precipitation variables.

of information on their value 1 hour later. Therefore, simple linear regression predictors can be used to forecast those variables at least for 1 hour lead time. In some cases, (groups 1, 2, and 5) the correlations are of a high value, even for lags of 6 hours (up to about 0.77). In group 3, the snowstorms group, correlations dropped relatively fast with lag.

5. Autocorrelations of the precipitation rate were lower than those of the other variables for all lags and for all groups. One hour autocorrelation lags ranged from a value of about 0.6 (groups 1, 3, and 5) to a value of about 0.37 (group 4). The auto correlations dropped fast at higher lags and became negative in some cases (groups 2 and 4).

6. A variety of intrinsic statistical characteristics can be observed for the various groups. Therefore, none was considered redundant for the purposes of the precipitation model tests. Most importantly, this indicates that the optimal parameters of linear, purely statistical, predictors of those variables based on past values of the same variables, would be calibration period dependent. This holds true, especially when predictions of the precipitation rate are sought.

7. No significant difference of statistical character was observed between Boston and Tulsa storms. The relatively

strong negative cross correlation of the temperatures to concurrent hourly precipitation rate in Tulsa hold some promise for linear regressions at this site.

The use of simple linear predictors of the temperature and pressure variables is illustrated in the next two figures. Figure 1 displays plots of the observed (solid lines) hourly dew-point temperature together with the corresponding 1-hour lead time predictions (stars) for the snowstorm group 3. This is the worst case in terms of dew-point prediction among the different groups. Analogous plots are shown in Figure 2 for the surface pressure of group 3. Again, shown is the worst case among all the different groups. The regressions had five explanatory variables and a constant. The independent variables were the lag-1 value of the dependent one and the lag-1 value of the other four variables. It must be stressed that each regression was the best least square fit for the particular record. Variation of optimal parameters for different calibration periods is expected. The magnitude of the standard errors indicates large parameter uncertainty for some of the explanatory variables. Nevertheless, the behavior of the models is good, which indicates potential predictive power, as implied by the observed high correlations (Table 4).

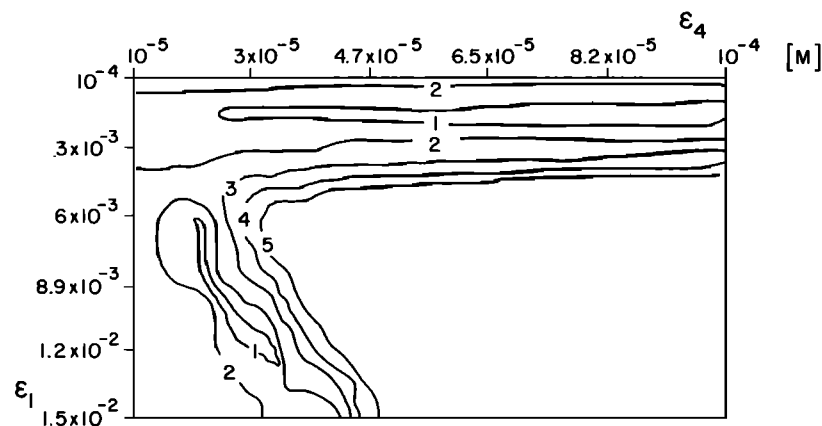


Fig. 5. Absolute proportional average error ( $E_1$ ). Storm group 2. Contour values: 1 = 0.24, 2 = 0.73, 3 = 1.21, 4 = 1.70, 5 = 2.18.

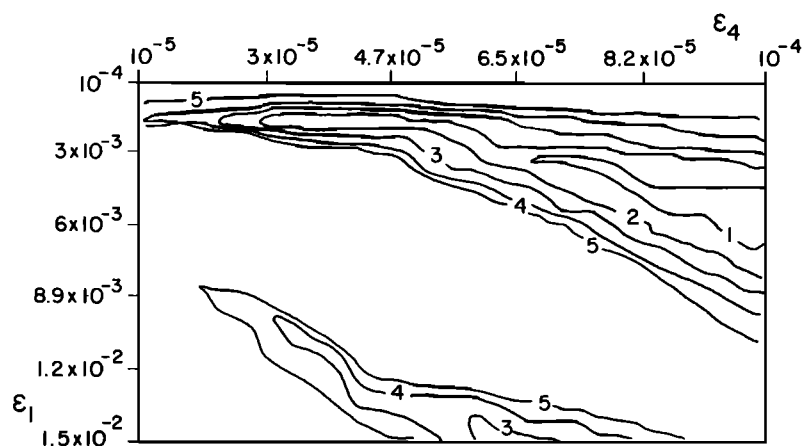


Fig. 6. Proportional standard error ( $E_2$ ). Storm group 2. Contour values: 1 = 0.942, 2 = 0.948, 3 = 0.954, 4 = 0.960, 5 = 0.966.

The potential of a linear regression in predicting precipitation with 1-hour lead is illustrated in Figures 3 and 4, using group 1 as example. Georgakakos and Bras [1982b] give similar results for all other groups. It must be stressed that each group is fitted with the best possible regression equation. Figure 3 is the result of a regression with concurrent time temperatures and pressures as explanatory variables. The regression of Figure 4 also includes the lag-1 precipitation as an explanatory variable. The fit of Figure 3 lacks the large fluctuations about the mean that characterizes the observed values. Due to the unconstrained nature of the regressions, unrealistic negative values of precipitation rate were also predicted. The fit of Figure 4 shows considerable improvement over the previous result, although some prediction delays are still observed. Some negative values were also predicted. The regressions of the various groups exhibited considerable difference in optimal parameters with high standard errors of estimation. This makes the generalization of any one equation impossible.

#### CALIBRATION OF PRECIPITATION MODEL

For purposes of this paper and as a first step toward the testing of the precipitation model developed, uniform vertical velocity profile with height is assumed. Similarly, the distribution of the parameter  $c$  (inverse average level diameter) with height was taken to be uniform. Under those assumptions,

$\beta = 1$  and  $\gamma = 1$  (equation (29) and Figure 2, Georgakakos and Bras [this issue]). No dependency of the parameter  $c$  on the updraft velocity was considered,  $k = 0$  (equation (19)). This implies constant average level diameter for all storms and at all times. The value of parameter  $\varepsilon_3$  (equation (18)) was taken equal to 1 s/m, which is the order of magnitude of the average updraft velocity  $v$ . Parameter  $\varepsilon_2$  (equation (18)) was initially set at the 700-mbar level. Therefore, the cloud top was allowed to vary from the level of 700 mbar to a level of 200 mbar for all cases. For all the model runs to follow, the initial conditions  $X_0$  of the state (mass of condensed liquid water equivalent in cloud storage) was taken equal to 1 kg/m<sup>2</sup>. With those values for  $\beta$ ,  $\gamma$ ,  $\varepsilon_3$ ,  $\varepsilon_2$ , and  $X_0$ , the model was used with input-output data to determine the values of the remaining two parameters:  $\varepsilon_1$  (equation (13)) and  $\varepsilon_4$  (equation (19)). Storm group 2, consisting of only one storm, was selected as the calibration period.

Three model performance criteria were considered. The average error in predicting the precipitation rate, the residual standard deviation, and the cross-correlation coefficient  $E_3$  between the model prediction and the observations. The first indicates the extent to which the model produces the volume of precipitation observed. The second is the standard least squares criterion. The third was included to indicate undesirable lags between observations and predictions. Its value

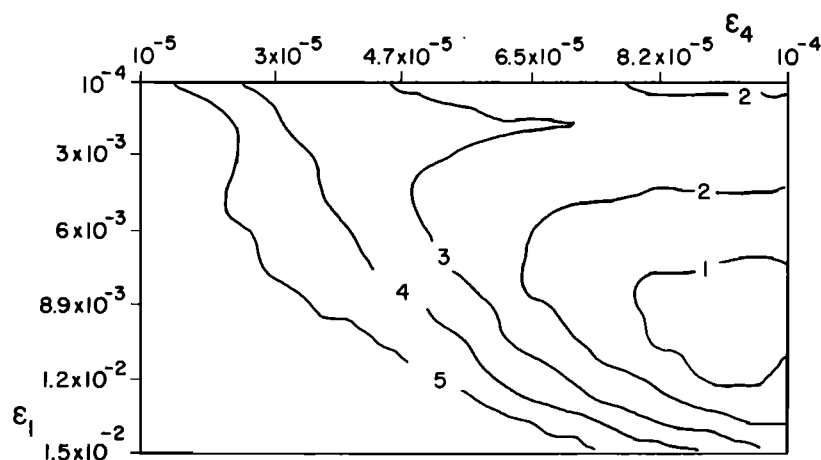


Fig. 7. Cross-correlation coefficient of observations and predictions ( $E_3$ ). Storm group 2. Contour values: 1 = 0.351, 2 = 0.344, 3 = 0.337, 4 = 0.330, 5 = 0.323.

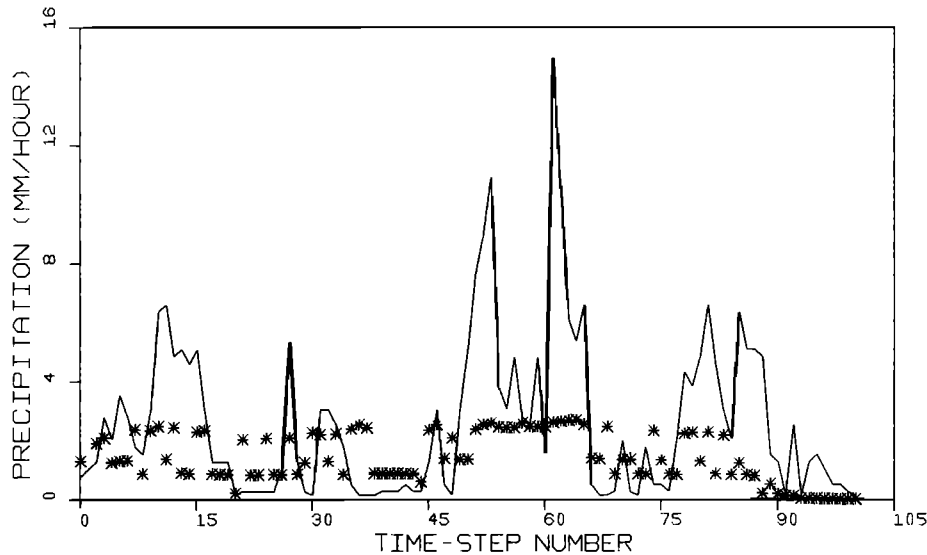


Fig. 8. Deterministic precipitation model hourly predictions (stars) vs. observations (solid line). Storm group 1.

ranges from  $-1$ , for worst performance, to  $1$ , for perfect performance. Normalized quantities for the first and second criteria were used for calibration. Those were the absolute proportional average error  $E_1$  and the proportional standard error  $E_2$ .  $E_1$  is obtained by dividing the residual mean by the observed mean precipitation rate and taking the absolute value of the ratio. Perfect performance leads to a value of  $E_1$  equal to zero.  $E_2$  results by division of the residual standard deviation with the observed precipitation rate standard deviation. Perfect performance yields a value of  $E_2$  equal to zero. Figures 5 to 7 present contours of  $E_1$ ,  $E_2$ , and  $E_3$ , respectively, in the space of the parameters  $\varepsilon_1$  (ordinate) and  $\varepsilon_4$  (abscissa). Parameter  $\varepsilon_1$  ranged from  $10^{-4}$  to  $1.5 \times 10^{-2}$ , while parameter  $\varepsilon_4$  ranged from  $10^{-5}$  m to  $10^{-4}$  m. Note that given  $k = 0$ ,  $\varepsilon_4$  is the average diameter assumed constant.

Examination of Figures 5 to 7 reveals that different parameter sets optimize the different criteria for the selected calibration period. Thus good least squares performance does not guarantee equally good performance in the cross-correlation coefficient  $E_3$ .

Notable also is the fact that changes in the value of one of the parameters result in drastic changes in the gradient of the performance index  $E_1$  with respect to the other parameter. Thus a value of about  $2 \times 10^{-3}$  for  $\varepsilon_1$  gives regions of very mild  $E_1$  gradient with respect to  $\varepsilon_4$ . For  $\varepsilon_1$  near  $9 \times 10^{-3}$ , sharp changes in  $E_1$  occur by changing  $\varepsilon_4$ . Multiple optima occur for  $E_1$  and possibly (for high values of  $\varepsilon_1$ ) for  $E_2$ . Index

$E_3$  was used to decide between these optima. For example the depressions in the lower left part of Figures 5 and 6 were excluded due to their lower value of  $E_3$ .

Even with the exclusion of the lower left part of all figures as a possible optimum, a choice between the parameter sets that optimize the different performance indices has to be made.

$E_1$  was heavily weighted due to the importance of preserving total precipitation volume and due to its sensitivity to parameter  $\varepsilon_1$ . The region of the limited  $\varepsilon_1$ ,  $\varepsilon_4$  space that gives both good performance with respect to  $E_1$  and to  $E_2$  is the one defined by  $\varepsilon_1$  in the range  $1.5 \times 10^{-3}$  to  $2.5 \times 10^{-3}$  and  $\varepsilon_4$  in the range  $3 \times 10^{-5}$  m to  $5.5 \times 10^{-5}$  m. A choice of  $\varepsilon_1 = 2 \times 10^{-3}$  and  $\varepsilon_4 = 4.5 \times 10^{-5}$  m in the prespecified region was arbitrarily made.

The sensitivity to the value of the ill-defined constant  $C_1$  for snow diffusion losses [see Georgakakos and Bras, this issue, equation (52)] was studied for the case of the snowstorm group 3. Two values of  $c_1$ , the one obtained for rain ( $7 \times 10^5$  kg/(m<sup>3</sup>s)) and the one estimated for snow ( $1.4 \times 10^5$  kg/(m<sup>3</sup>s)) were examined. Contours of  $E_1$  and  $E_2$  for  $\varepsilon_1$  in the range  $10^{-4}$  to  $1.3 \times 10^{-2}$  and  $\varepsilon_4$  in the range  $1.2 \times 10^{-5}$  to  $10^{-4}$  were obtained for both values of  $C_1$ . No significant changes are observed among the different plots corresponding to  $E_1$  and  $E_2$  other than a shift to a higher optimal value of  $\varepsilon_4$  when  $C_1$  changes from  $7 \times 10^5$  to  $1.4 \times 10^5$  kg/(m<sup>3</sup>s). Choosing a value of  $c_1 = 1.4 \times 10^5$  for snowstorm, group 3, leads to parameters  $\varepsilon_1$  and  $\varepsilon_4$  similar to the ones already obtained using group 2 (rainfall) data.

TABLE 5. Deterministic Precipitation Model Residual Statistics for Storm Groups 1 to 5

Statistic	Group				
	1	2	3	4	5
Mean, mm/h	1.09	-0.23	0.28	0.82	0.31
Standard deviation, mm/h	2.49	1.701	1.10	2.91	2.68
Lag-1 autocorrelation	0.56	0.46	0.57	0.41	0.47
Lag-2 autocorrelation	0.34	0.22	0.36	0.20	0.31
Lag-3 autocorrelation	0.21	0.12	0.26	0.16	0.13
Regression residual standard deviation	2.64	1.63	1.08	2.94	2.60

TABLE 6. Deterministic Precipitation Model Least Squares Performance Measures for Storm Groups 1 to 5

Description	Group				
	1	2	3	4	5
Efficiency coefficient	0.02	0.10	0.12	0.03	0.15
Determination coefficient	0.19	0.12	0.17	0.13	0.16
Persistence coefficient	-0.39	0.23	-0.12	0.23	-0.10
Extrapolation coefficient	0.44	0.71	0.55	0.70	0.54

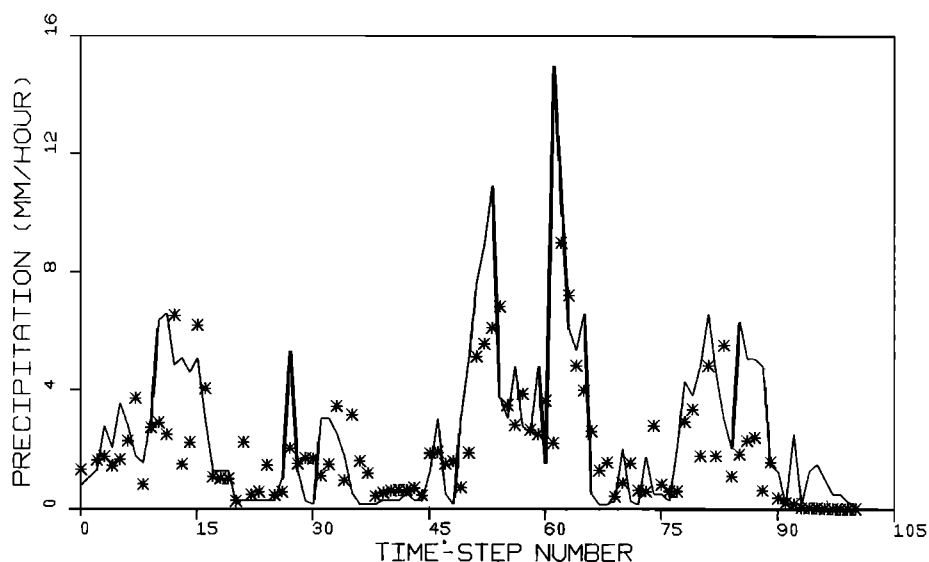


Fig. 9. Stochastic precipitation model hourly predictions (stars) vs. observations (solid line). Storm group 1.

In summary,  $C_1$  was taken as  $1.4 \times 10^5$  for snow group 3 and  $7 \times 10^5$  for all rainfall groups. Parameters  $\varepsilon_1$  and  $\varepsilon_4$  took values of  $2 \times 10^{-3}$  and  $4.5 \times 10^{-5}$ , respectively, for all groups.

#### RESULTS: THE CALIBRATED RAINFALL MODEL PREDICTIONS

The precipitation model developed in *Georgakakos and Bras* [this issue] was used to forecast the hourly precipitation rate for all storm groups, with the parameters defined in the previous section. The input at all times was equal to the current observed value for the meteorological parameters. Performance is judged by the values of the following measures: residual mean, residual standard deviation, lag-1 correlation coefficient of residuals, coefficient of efficiency, coefficient of determination, coefficient of persistence, coefficient of extrapolation, and comparison to the regression runs discussed in a previous section which used the same input and were locally calibrated for all storm groups.

The residual at each time is equal to the difference between

the predicted precipitation rate and the observed one. The coefficients of efficiency, determination, persistence, and extrapolation were introduced in hydrologic forecasting by *Kitanidis and Bras* [1980]. The coefficient of efficiency is a measure of the residual squared error as it compares to the squared error of the quantity to be forecasted. A perfect value is 1. Negative values indicate large residuals relative to the scale of the observations. The coefficient of determination compares the residual squared error to the squared error of the observations after the linear trends have been removed from the residual time series by regression. Comparison of the coefficients of efficiency and determination allows assessment of the possible model systematic errors. The coefficient of persistence compares the model prediction to a simple model that predicts the previous observation. Thus negative values of this measure indicate that in the least squares sense the model is worst than no-model persistence. The coefficient of extrapolation compares the model predictions to the observations of a linear

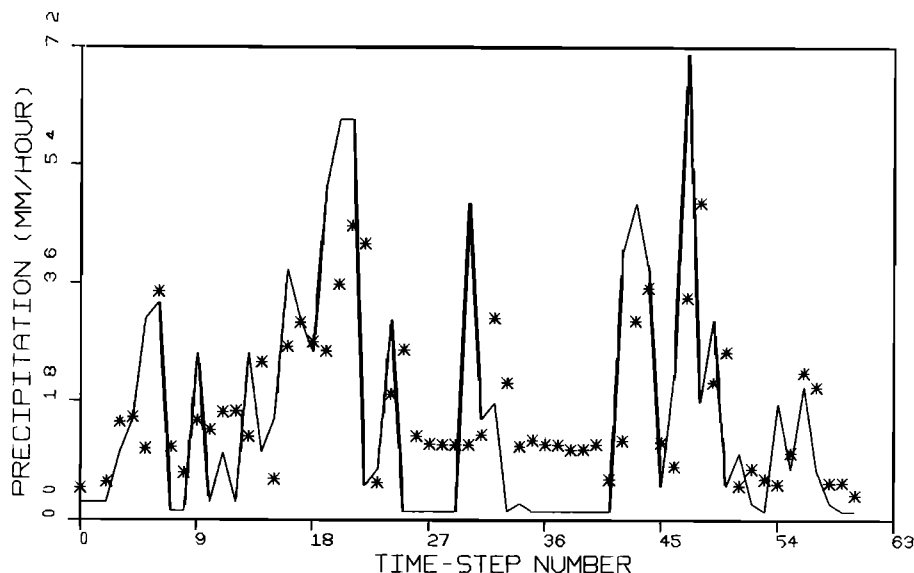


Fig. 10. Stochastic precipitation model hourly predictions (stars) vs. observations (solid line). Storm group 2.



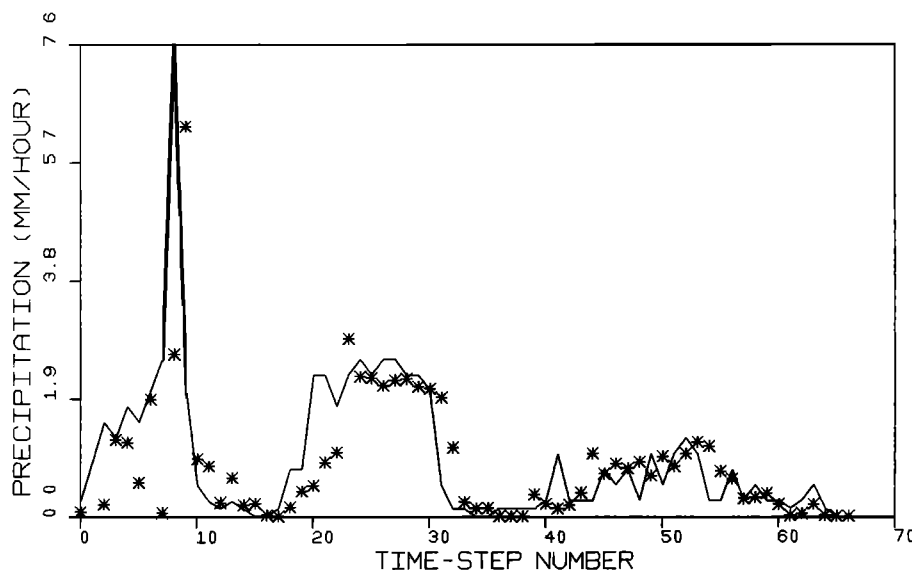


Fig. 11. Stochastic precipitation model hourly predictions (stars) vs. observations (solid line). Storm group 3.

extrapolator using the latest observations (in this work the latest two).

The residual mean multiplied by the number of storm time-steps gives the accumulated volume error in predicting precipitation.

Figure 8 shows the 1-hour lead forecasts, using the rainfall model as a deterministic relationship, of the precipitation rate (mm/hr), in stars together with the corresponding observations in solid lines for storm group 1. It can be compared to Figure 3, a regression fitted exclusively to group 1, which uses  $T_0$ ,  $p_0$ ,  $T_d$ , and  $T_w$  as explanatory variables.

Table 5 gives the residual mean, standard deviations, and three autocorrelation coefficients for each storm group together with the residual standard deviation of the regressions discussed previously. Keep in mind that those regressions were calibrated to each group individually, therefore representing the best possible linear fit. Table 6 gives the least squares performance measures for each storm group.

Examination of Table 5 indicates that the deterministic

model performance, with parameters estimated only with data of group 2, is comparable in the least squares sense, to the locally calibrated regressions. In some cases, groups 1 and 4, the precipitation model had better performance. The value of the residual means of Table 5 implies that in all cases volume preservation was satisfactory. The relatively high residuals correlation values are indicative of possible improvement when the precipitation model is later complemented by a filter.

The performance measures of Table 6 point to the low efficiency of the deterministic model. The difference between the coefficient of determination and the coefficient of efficiency indicate systematic errors in some of the predictions. Some cases had negative persistence coefficients, implying poor performance at a 1-hour lead forecast relative to a simple persistence model. The extrapolation coefficient was large for most storm groups indicating better performance than linear extrapolation.

It should be noted that the precipitation model proposed can predict the beginning and ending of the precipitation

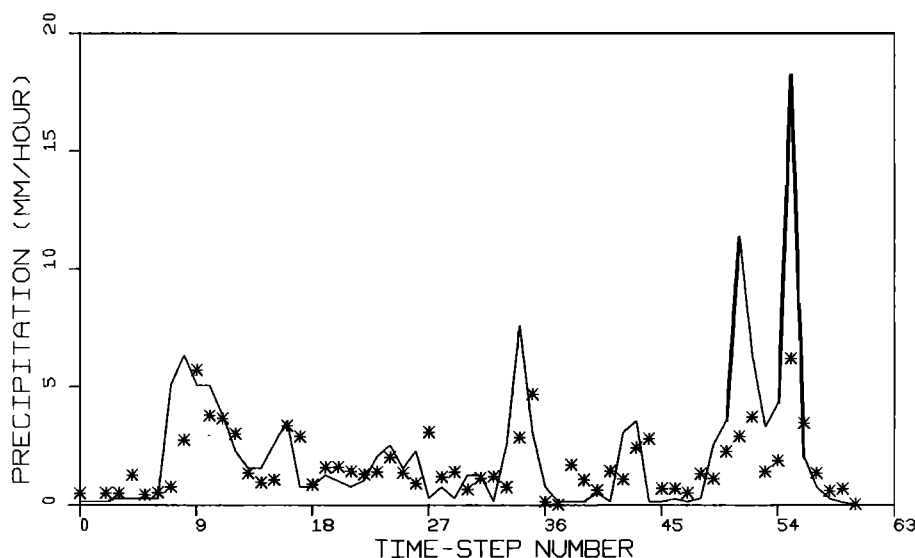


Fig. 12. Stochastic precipitation model hourly predictions (stars) vs. observations (solid line). Storm group 4.

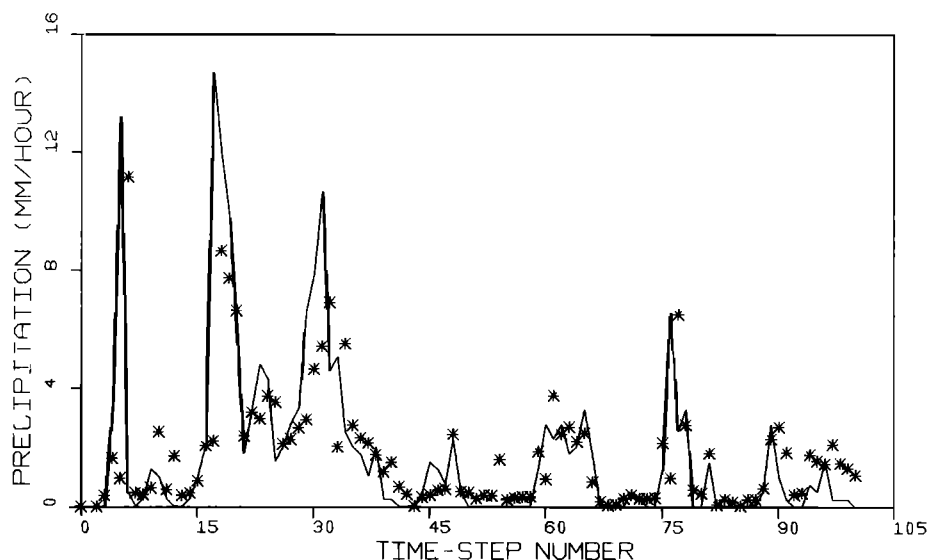


Fig. 13. Stochastic precipitation model hourly predictions (stars) vs. observations (solid line). Storm group 5.

based on the temperature and pressure input. Therefore, when the forecasts of no rain or no snow are taken into consideration for a continuous period of record, performance is considerably improved.

Characteristic to all runs of the deterministic model is its deficiency in predicting excessively high precipitation rates while it does a rather good job in predicting the no rain periods. This suggests the examination of other values of  $k$ ,  $\beta$ , and  $\gamma$  than the ones selected in this calibration. The diffusion losses part of the model seems to respond properly to observed input.

The station precipitation model performance in a stochastic feedback mode was studied next. The continuous-discrete extended Kalman filter in Gelb [1974, pp. 188] was utilized as outlined in the introduction. The input error standard deviation for temperature and dew point were set to  $1^\circ\text{K}$ , while no error was assumed for the pressure input. Zero correlation in the errors of the different input variables was assumed. The observation error standard deviation took the value 1 (mm/hr), while the model error spectral density was set equal to  $0.01 \text{ (kg/m}^2\text{)}^2/\text{s}$ . This spectral density adds about  $0.01 \times 3600 = 36 \text{ (kg/m}^2\text{)}^2$  to the state variance at each step when the order of magnitude of  $X$  is about  $10 \text{ kg/m}^2$ . Such a high value of the model error spectral density was considered necessary in order to avoid filter divergence. This initial state standard deviation was  $0.3 \text{ kg/m}^2$ . No sensitivity analysis was performed for any of the filter parameters.

Performance was judged as in the case of deterministic model forecasts. Now, however, the filter predictions are com-

pared to the set of individually fitted regressions that use the previous precipitation value. Figures 9 to 13 show the model forecasts (dashed line) together with the corresponding hourly precipitation rate (solid line).

Tables 7 and 8 give the performance statistics. Considerable improvement is noted for all storm groups relative to the deterministic model, as expected.

The values of the residual mean have all been reduced such that preservation of precipitation volume is even more successful in this case. Reductions of the standard deviations and the auto correlations of the residuals confirm good filter performance.

Comparison of the filter forecasts to the locally calibrated regression forecasts (see also standard deviations in Table 7) support the robustness and reasonableness of the storm independence of the suggested rainfall model.

Considerable improvement (Table 8) in the efficiency measures is observed, relative to the deterministic case.

All persistence coefficients are now positive and the efficiency has risen an order of magnitude in some cases. Examples of the model performance for longer forecast lead times using observed input are given in Tables 9 and 10. Model forecasts up to a maximum lead time of 6 hours were considered for storm groups 1 and 5. The drop of efficiency is more pronounced in the Boston data (group 1) than in the Tulsa data (group 5). Increase of the persistence coefficient from the low 1-hour lead time value to values of 0.48 and 0.57 supports the forecasting ability of the precipitation model. Table 10 suggests that the stochastic model will have good

TABLE 7. Stochastic Precipitation Model Residual Statistics for Storm Groups 1 to 5

Statistic	Group				
	1	2	3	4	5
Mean	0.58	-0.15	0.104	0.49	0.06
Standard deviation	2.01	1.43	0.96	2.20	2.4
Lag-1 autocorrelation	0.058	0.01	-0.003	0.16	-0.12
Lag-2 autocorrelation	0.009	-0.07	-0.058	0.07	0.09
Lag-3 autocorrelation	0.1	-0.05	0.04	0.10	-0.04
Regression residual standard deviation	2.12	1.51	0.94	2.90	2.30

performance with precipitation data from Tulsa for relatively long forecast lead times.

#### RESULTS: MODEL GENERALIZATION TO CONDITIONS IN VENEZUELA

The generality and robustness of the proposed rainfall model was tested with storms in the tropical mountainous regions of Venezuela.

Initially, meteorologic data from station Macagua in the Caroní River basin (tributary to the Orinoco River) was used with the same parameters discussed previously, which were calibrated with Boston data. The 1-hour predictions obtained with the filtered model and using the latest observations of  $p_0$ ,  $T_0$ , and  $T_d$  are shown in Figures 14 through 17. All performance criteria were positive. Considering that there was absolutely no calibration attempt, the results are surprisingly good. The obvious problems are the delayed response and peak, the underestimated peaks, and the persistent predicted light precipitation early in the storm.

The first two problems can be immediately attributed to known peculiarities of tropical storm behavior. First, clouds may reach elevations well above the 200 mbar level assumed here. Heights of 12 km and pressures of 100–150 mbar are common. Similarly, the lowest cloud top pressure ( $\varepsilon_3$ ) is probably above the 700-mbar level assumed, probably on the order of 400 mbar. Second, updraft velocities are usually large,  $\varepsilon_1$  should be larger. Third, the assumption of uniform mean particle diameter is not adequate. At least the mean particle diameter should be higher. To investigate the previous arguments the updraft velocity was increased by augmenting  $\varepsilon_1$ , to  $0.24 \times 10^{-2}$ . The lowest cloud top pressure was set at 400 mbar. Although the particle size distribution remained uniform throughout the cloud profile, the inverse of the mean particle size,  $\varepsilon_4$ , was increased to  $0.8 \times 10^{-4}$ . Given the larger parameter uncertainty, the standard deviation of the model error was increased by a factor of 5 to 0.5. The new parameters were used in predicting the storm shown in Figure 14. The result is Figure 18. As expected a higher peak and better timing is observed, an indication that calibration can indeed be improved with simple physical arguments.

The problem of the apparent low persistent precipitation of the beginning of the storm was traced to the effects of atmospheric inversions. It was indeed confirmed that they existed in those early morning periods. During these periods the atmosphere is nearly saturated and  $T_0 \approx T_d$ . Equation (1) will incorrectly predict an updraft velocity because it assumes a dry-adiabatic cooling of the ambient air when in fact the temperature profile corresponds to a very stable atmosphere. The saturation conditions with a false updraft velocity leads to precipitation. One way to solve this problem is to change the parameterization of updraft velocity. Possibilities include making it dependent on time of the day (inversions usually occur early in the day in the tropics) or direct observations.

TABLE 8. Stochastic Precipitation Model Least Squares Performance Measures for Storm Groups 1 to 5

Description	Group				
	1	2	3	4	5
Efficiency coefficient	0.36	0.33	0.28	0.38	0.28
Determination coefficient	0.41	0.33	0.35	0.46	0.30
Persistence coefficient	0.10	0.43	0.08	0.50	0.07
Extrapolation coefficient	0.63	0.78	0.63	0.81	0.61

TABLE 9. Least Squares Performance Measures for Storm Group 1

Description	Lead Time, hours					
	1	2	3	4	5	6
Efficiency coefficient	0.36	0.17	0.08	0.04	0.02	0.02
Determination coefficient	0.41	0.36	0.31	0.30	0.29	0.30
Persistence coefficient	0.10	0.26	0.33	0.40	0.47	0.47
Extrapolation coefficient	0.63	0.53	0.48	0.45	0.44	0.43

Maximum lead time: 6 hours.

Another promising alternative is to use simultaneous temperature data from stations at different elevations as a surrogate measure of the temperature profile. Initial investigations indicate that this latter approach may indeed work in the Caroní River basin.

#### CONCLUSIONS AND RECOMMENDATIONS

A set of 11 storms of different type and severity from Boston, Massachusetts, and Tulsa, Oklahoma, was the main hourly data base for the tests of the precipitation model presented in *Georgakakos and Bras* [this issue] and parameterized in this paper. Several performance indices were used to quantify behavior of the developed model. Thus mean, standard deviation and autocorrelation structure of the residual process together with efficiency, determination, persistence, and extrapolation coefficients were utilized. In addition the model predictions were compared to the optimistic predictions of linear regressions of same input calibrated for each storm period separately. In contrast the physical model parameters were manually calibrated with data from a single storm and remained unchanged during all model tests.

Characteristic to all tests of the deterministic precipitation model was inadequate simulation of high precipitation rates and the successful prediction of the no-precipitation periods.

Improved model behavior was observed when the stochastic formulation was used. The filter parameters were based on a priori considerations and remained unchanged for all the stochastic model runs. Again, the model predictions were compared to locally calibrated regression model predictions with rainfall in the last time step as well as meteorologic parameters as explanatory variables. The high extrapolation coefficients indicate superiority of the developed model over linear predictors using the previous two precipitation rates. The low autocorrelations of the residual process suggested near-optimal filter performance.

Good stochastic model performance is also confirmed by the positive persistence coefficients. Positive persistence short-range predictions are rather difficult to attain [e.g., *Kitanidis and Bras*, 1980].

TABLE 10. Least Squares Performance Measures for Storm Group 5

Description	Lead Time, hours					
	1	2	3	4	5	6
Efficiency coefficient	0.28	0.26	0.18	0.34	0.22	0.21
Determination coefficient	0.30	0.37	0.33	0.40	0.30	0.30
Persistence coefficient	0.07	0.41	0.43	0.59	0.54	0.56
Extrapolation coefficient	0.61	0.60	0.55	0.45	0.35	0.35

Maximum lead time 6 hours

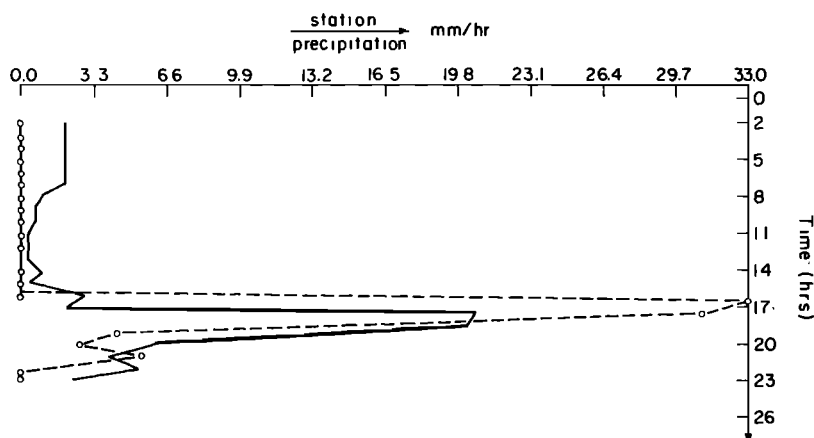


Fig. 14. Stochastic precipitation model hourly predictions (solid line) vs. observations (dashed line). Macagua, Venezuela, storm data.

A valid concern is the relative merit of the model dynamics and filter adjustments. Clearly, the filter is extremely responsive to observations during the update phase. A search for the best model error density should accompany model use. Georgakakos [1984] has, in fact, implemented such procedure. Nevertheless, model dynamics are significant, even in this non-optimal example. The following facts are evidence of this:

1. The model is not calibrated to individual events, yet performs reasonably well.
2. Extended forecasts up to 6 hours lead time (Table 9) do significantly better than a simple persistence model. This is a very difficult and strict test.
3. Even at 1-hour lead predictions the dynamics are significant. For example, in Figure 13 the model predicts a downturn in precipitation at time step 19, even though time step 18 was underestimated in the rising limb of the hydrograph. Blind tracking of data would have adjusted upwards and possibly continue the rising trend in rainfall intensities.

Nevertheless, it has been confirmed in continuing work that

the time constants of the rainfall model are short (on the order of 1 hour). This implies that response is very dominated by inputs ( $T_0$ ,  $T_d$ ,  $p_0$ ).

No significant difference between the Boston and Tulsa hourly forecasts was observed. Nevertheless, when the extended forecasts (up to 6 hours maximum forecast lead time) performance coefficients were examined for storms of the two locations, the Boston forecasts showed a much more pronounced drop in efficiency with lead time. Efficiency was almost constant for all the lead times for Tulsa.

The robustness of the model was confirmed when storms in tropical Venezuela were reasonably predicted without recalibration. It was shown that model parameters would be significantly improved by using simple physical arguments.

Automatic calibration methods of all the model parameters should be used to establish optimal parameter values. The maximum likelihood methodology [Restrepo-Posada and Bras, 1982] and the unbiased sequential estimator in Georgakakos and Bras [1982a] are two possible alternatives. The

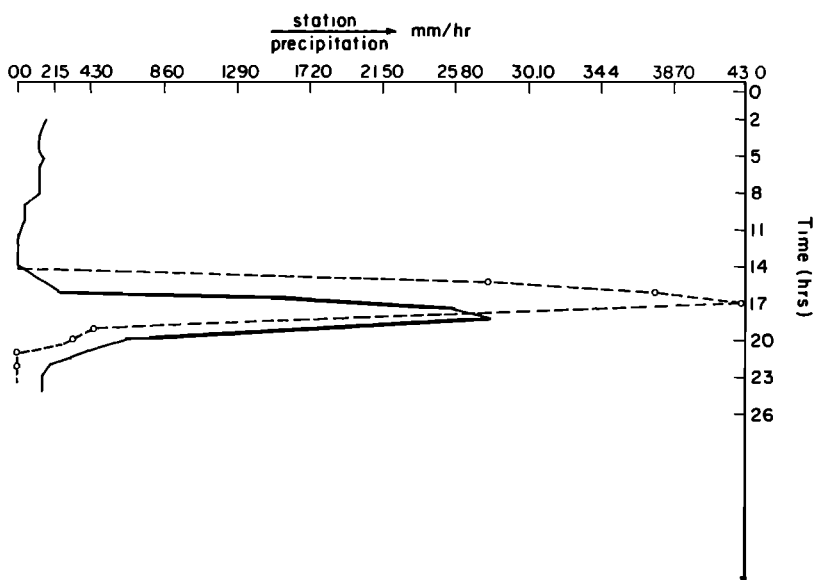


Fig. 15. Stochastic precipitation model hourly predictions (solid line) vs. observations (dashed line). Macagua, Venezuela, storm data.

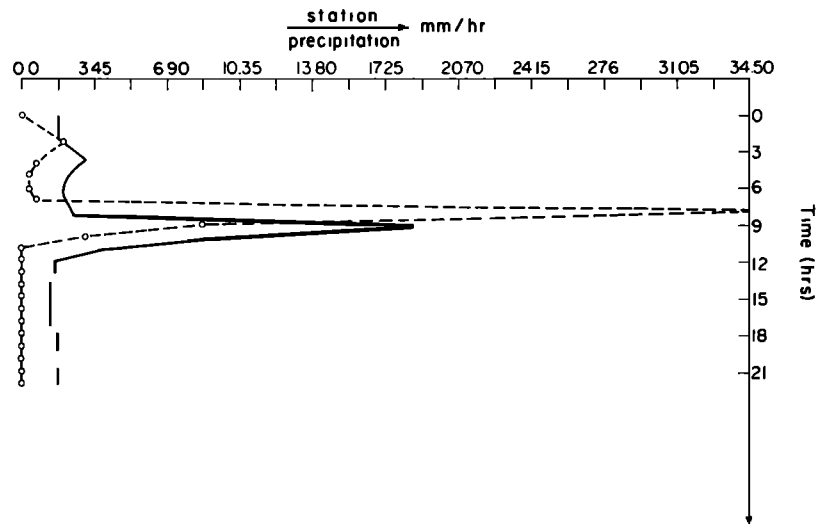


Fig. 16. Stochastic precipitation model hourly predictions (solid line) vs. observations (dashed line). Macagua, Venezuela, storm data.

fact that there is only one state, permits the use of elaborate parameter estimation procedures and many data sets. Caution must be exercised, however, to avoid the local optima indicated by the parameter space maps presented in previous sections. There is no question, see Figure 8, that better calibration will lead to better performance in a deterministic mode, particularly in handling high peaks. Improved calibration will address some of the model dynamics versus filter issues mentioned previously. During this effort, all parameters including  $m$ ,  $\beta$ , and  $\gamma$  should be calibrated.

The calibrated model, complemented by a filter, should be used with several data sets from different locations and for different storms in a sensitivity analysis of the effect of the filter parameters on model performance. The nonstationarity of the model error should be examined in detail. Similarly, the observation error structure needs examination. Related to this, the theoretical framework in Sharon [1980] will be useful. The extent to which the error in the input variables improves model performance should also be evaluated.

Due to the physical interpretation given to the model components, observations of cloud tops, upper-cloud divergence from satellite images, droplet spectra, and updraft velocities can considerably improve the stochastic model performance when used in additional observation equations. Perhaps more important will be observations of the model state (cloud moisture) of the type described in Bunting and Conover [1976].

The low order of the precipitation model permits probabilistic forecasts based on the use of the Bayes law and initial probability distributions of the model state. Formulations in Ho and Lee [1964] can be used. The types of forecasts are particularly useful in decision making.

Slight modifications in the model structure can be done to incorporate the effect of the time variation of some parameters. For example assume a one-parameter Markov model for the time evolution of the average level diameter  $1/c$ . Estimating the additional parameter from storm data permits evaluation of the resultant precipitation for varying initial conditions for  $1/c$ . This can be done in a filter framework with

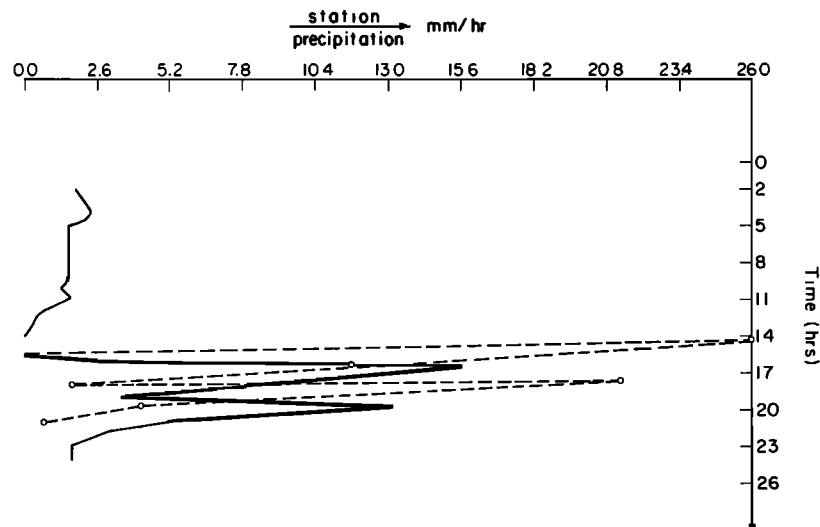


Fig. 17. Stochastic precipitation model hourly predictions (solid line) vs. observations (dashed line). Macagua, Venezuela, storm data.

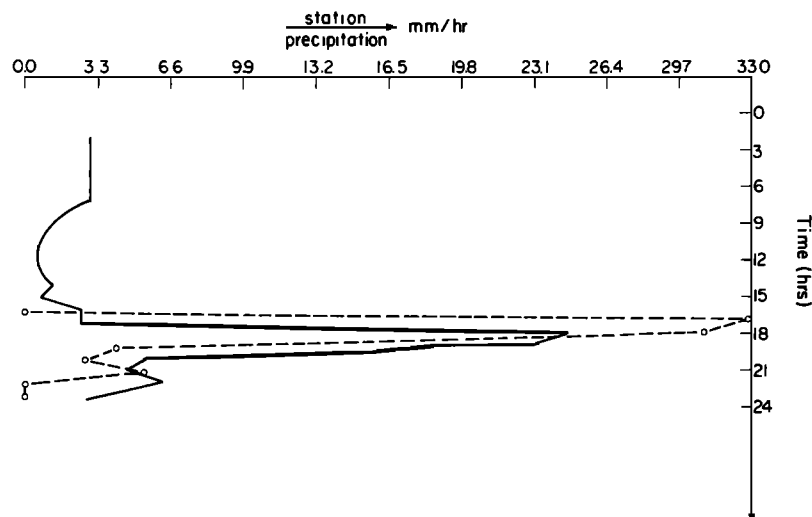


Fig. 18. Stochastic precipitation model hourly predictions (solid line) vs. observations (dashed line). Macagua, Venezuela, storm data. Model parameters recalibrated.

observations of the precipitation rate and possibly of other physical quantities as described above. This type of research can be valuable in assessing the effects of cloud seeding on the precipitation rate in real time.

**Acknowledgments.** This work and publication was sponsored by the U.S. National Weather Service, Hydrologic Research Laboratory, as part of a cooperative research agreement (NA80AA-H-00044) with the Ralph M. Parsons Laboratory, Department of Civil Engineering, Massachusetts Institute of Technology. The use of the model in Venezuela was sponsored by EDELCA, a Venezuelan government utility in charge of the hydroelectric power development of the Coroni River basin. The authors acknowledge the cooperation and work of Michael Hudlow (N.W.S.), Yolanda Fiallo, and Antonio A. Ahogado (both the EDELCA). Portions of this paper were written while Georgakakos was a National Research Council-NOAA Research Associate and Bras was on a sabbatical leave sponsored by the John Simon Guggenheim Foundation and Simon Bolivar University, Caracas, Venezuela. We want to thank Alonso Rhenals and two other anonymous reviewers for their quick response in an unusual situation.

#### REFERENCES

- Bunting, J. T., and J. H. Conover, Estimates from satellites of total ice and water content of clouds, paper presented at International Cloud Physics Conference, Am. Meteorol. Soc., Boulder, Colo., July 26-30, 1976.
- Gelb, A. (Ed.), *Applied Optimal Estimation*, The M.I.T. Press, Cambridge, Mass., 1974.
- Georgakakos, K. P., Model-error adaptive parameter determination of a conceptual rainfall prediction model, paper presented at the 16th Southeastern Symposium on System Theory, Inst. Electron. Electr. Eng., Miss. State Univ., March 25-27, 1984.
- Georgakakos, K. P., and R. L. Bras, Real time, statistically linearized adaptive flood routing, *Water Resour. Res.*, 18(3), 513-524, 1982a.
- Georgakakos, K. P., and R. L. Bras, A precipitation model and its use in real-time river flow forecasting, *Tech. Rep. 286*, Ralph M. Parsons Lab., Hydrol. Water Resour. Syst., Dep. Civil Eng., Mass. Inst. Technol., Cambridge, Mass., 1982b.
- Georgakakos, K. P., and R. L. Bras, A hydrologically useful station precipitation model, 1, Formulation, *Water Resour. Res.*, this issue.
- Ho, Y. C., and R. C. K. Lee, A Bayesian approach to problems in stochastic estimation and control, *IEEE Trans. Automat. Contr.*, AC-9(5), 333-339, 1964.
- Kitanidis, P. K., and R. L. Bras, Real-time forecasting with a conceptual hydrologic model, 2, Applications and results, *Water Resour. Res.*, 16(6), 1034-1044, 1980.
- Pruppacher, H. R., and J. D. Klett, *Microphysics of Clouds and Precipitation*, D. Reidel, Boston, Mass., 1978.
- Restrepo-Posada, P. J., and R. L. Bras, Automatic parameter estimation of a large conceptual rainfall-runoff model: A maximum likelihood approach, *Tech. Rep. 267*, Ralph M. Parsons Lab. Water Resour. Hydrodyn., Dep. Civil Eng., Mass. Inst. Technol., Cambridge, Mass., 1982.
- Sharon, D., The distribution of hydrologically effective rainfall incident on sloping ground, *J. Hydrol.*, 46, 165-188, 1980.
- Sulakvelidze, G. K., *Rainstorms and Hail*, translated from Russian by the Israel Program for Scientific Translations, Jerusalem.
- R. L. Bras, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139.
- K. P. Georgakakos, Hydrologic Research Laboratory, National Weather Service, NOAA, Silver Spring, MD 20910.

(Received October 21, 1983;  
revised July 31, 1984;  
accepted August 1, 1984.)