NASA-CR-195132

N I X I K

IN 47 CR

SEMI-ANNUAL STATUS REPORT ON GRANT NO. NAG 5-386 *205073*

10P

FURTHER STUDIES TO EXTEND AND TEST THE AREA-TIME-INTEGRAL TECHNIQUE APPLIED TO SATELLITE DATA

PERIOD: 1 JULY - 31 DECEMBER 1993

Principal Investigators

Paul L. Smith

Institute of Atmospheric Sciences
South Dakota School of Mines and Technology
501 E. St. Joseph Street
Rapid City, South Dakota 57701-3995

Thomas H. Vonder Haar

ASTeR, Inc.
P. O. Box 466
Fort Collins, Colorado 80522

Prepared for NASA Goddard Space Flight Center
Greenbelt, Maryland 20771

(NASA-CR-195132) FURTHER STUDIES TO EXTEND AND TEST THE AREA-TIME-INTEGRAL TECHNIQUE APPLIED TO SATELLITE DATA Semiannual Status Report, 1 Jul. - 31 Dec. 1993 (South Dakota School of Mines and Technology) 10 p

N94-24705

Unclas

G3/47 0205073

The principal goal of this project is to establish relationships that would allow application of area-time integral (ATI) calculations based upon satellite data to estimate rainfall volumes. The research is being carried out as a collaborative effort between the two participating organizations, with the satellite data analysis to determine values for the ATIs being done primarily by the ASTeR scientists and the associated radar data analysis to determine the "ground-truth" rainfall estimates being done primarily at the South Dakota School of Mines and Technology (SDSM&T). Synthesis of the two separate kinds of data and investigation of the resulting rainfall-versus-ATI relationships is then carried out jointly.

The research has been pursued using two different approaches, which for convenience can be designated as the "fixed-threshold approach" and the "variable-threshold approach". In the former approach, an attempt is made to determine a single temperature threshold in the satellite infrared data that would yield ATI values for identifiable cloud clusters which are most closely related to the corresponding rainfall amounts. In this respect the approach resembles the GOES Precipitation Index (GPI), but we make no assumption of a fixed rainfall rate for each cloudy pixel. Results thus far have indicated that a strong correlation exists between the rain volumes and the satellite ATI values, but the optimum threshold for this relationship seems to differ from one geographic location to another. The difference is probably related to differences in the basic precipitation mechanisms that dominate in the different regions. The average rainfall rate associated with each cloudy pixel is also found to vary across the spectrum of ATI values.

During this reporting period, a journal manuscript summarizing these results was revised in response to review comments and returned to the editors of the Monthly Weather Review. Notice of acceptance of these revisions was received and the paper is now scheduled to appear in the March 1994 issue of MWR.

Work on the second, or "variable-threshold", approach for determining the satellite ATI values was essentially suspended during this period due to exhaustion of project funds.

Most of the ATI work thus far has dealt with cloud clusters from the Lagrangian or "floating-target" point of view. For many purposes, however, the Eulerian or "fixed-target" perspective is more appropriate. For a very large target area encompassing entire cluster life histories, the rain volume-ATI relationship would obviously be the same in either case. The important question for the Eulerian perspective is how small the fixed area can be made while maintaining consistency in that relationship. To investigate that question, a sample of radar data for echo clusters from southeastern Montana was partitioned by dividing the radar surveillance area into successively smaller sectors. If sectors receiving more than 50% of their rainfall from echoes below the threshold (25 dBz) used for the ATI calculation are excluded, the rain volume-ATI relationship remains essentially the same from the overall radar surveillance area of 75,000 km^2 down to 4700 km^2. The cases with weaker echoes could presumably be incorporated by repeating the ATI computation with a lower reflectivity threshold. A paper describing this analysis was accepted for presentation at a Special Session on Hydrometeorology in March 1994; an abstract for that paper appears in Attachment A.

The Graduate Research Assistant working on the project continued his investigation of the Probability Density Functions associated with the radar echo clusters. His effort during the fall semester was supported largely by local funds. Attachment B is a summary of his work during the fall semester.

By the end of this reporting period, no response had been received to the proposal for three additional years of investigation that was submitted to NASA in June. (News of a forthcoming award for the first year was received in January, 1994.)

THE RELATIONSHIP BETWEEN RADAR AREA-TIME INTEGRALS AND AREAL RAINFALL

by

L. Ronald Johnson and Paul L. Smith

For convective elements a strong correlation exists between area-time integrals (ATIs) determined from radar data and rainfall volumes determined using a reference frame centered on the storm through time, i.e. a "floating target" frame. Similar, but somewhat weaker, correlations have been identified using satellite infrared data to determine the ATIs. These correlations can be useful in studies of storm water budgets and regional or global-scale precipitation. However, to make such relationships useful to the hydrology community, they must be established for fixed areas on the ground (e.g. watersheds). This paper reports the results of an ATI analysis using a fixed map grid. The grid could be adjusted to represent a specific watershed or a geo-political boundary. The gridded estimated rainfall amounts would be desirable quantities for hydrologists and forecasters.

The usable size of the grid elements is investigated. If a cloud system plays out its entire lifetime within a single grid element, the ATI and the corresponding rainfall estimate would be the same whether treated in the Lagrangian or Eulerian frame. Thus, the correlation should be equally strong for such cases; the important question is what happens as the grid spacing is reduced. The median "footprint size" for convective radar echo clusters followed in the COHMEX project is about 87.3 km^2 and the quartile range is 37.6 to 183.1 km^2. With grid squares comparable to the upper quartile, the echo footprints frequently cross grid boundaries. The strength of the grid-element rainfall versus ATI correlation in such cases, and the usable range of grid spacings, are examined in the paper.

28 January 1994

MEMORANDUM

FOR: Dr. Paul Smith

FROM: Scott D. Larsen

SUBJECT: Fall 1993 Research Results--A Summary.

This memorandum summarizes the progress of my research so far. The background material is included as preliminary material for the introduction of the thesis and upcoming seminar.

The Search for a Uniform Probability Density Function for Precipitation Processes

Given a set of outcomes described by a set A such that each event A_i is a subset of A, then Pr(A) is the *probability* that the event A_i is an element of A. If the following are also true,

1. $\Pr(A_i) \geq 0$
2. $\Pr(\mathrm{A}) = 1$, and
3. $\Pr(A_1 \cup A_2 \cup A_3 \cup \cdots) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3) + \cdots$

then Pr(A) is called a *probability set function*. If a probability set function can be described in terms of a function such that

$$\Pr(A) = \Pr(X \in A) = \sum_{A} f(X),$$

then $f(X)$ is said to be a *probability density function* with random variable X. The probability density function, or PDF, completely determines the probability of an event occurring within the set of possible outcomes.

In this study the objective is to determine a frequency distribution, given a radar reflectivity value, for the area of coverage of the corresponding precipitation event. Such work with areas and reflectivity values has been extensively done in Area-Time Integral studies. However, much of the work has assumed that a PDF describing the relationship exists without any reference to the function itself. This study attempts to determine if, in fact, such a PDF does exist and if it can be universally applied to a variety of precipitation events.

Attempting to find a PDF, or any function for that matter, ultimately is a problem of fitting a characteristic equation (a "curve") to an observed set of data. While the problem of finding a value if a function is known is fairly straightforward, the reverse problem is not so and usually more complicated.

This study selected radar precipitation data from the 1981 CCOPE project to use as an initial attempt to determine a PDF. This data set was chosen for three reasons:

1. The data set is readily available;
2. The data set has been extensively analyzed in past studies; and
3. The radar data involved full volume scans of the precipitation events.

Precipitation cells were initially chosen to fall into one of three time-duration categories: small, those lasting on the order of 30 minutes; medium, those lasting on the order of 90 minutes; and large, those lasting longer than 180 minutes. Secondly, the data set was divided into six categories based on time of year: May; 1-15 June; 16-30 June; 1-15 July; 16-31 July; and August. Finally, three of each type cells were analyzed from each of the six categories for a total of 54 cells analyzed. Individual cells were selected according to relative isolation from other storms and followed from first to last echo. The area covered by each cell during each scan was analyzed using interactive radar analysis software, which produced the echo area statistics used in this study. Figure 1 shows the distribution of cells within an area-time domain.

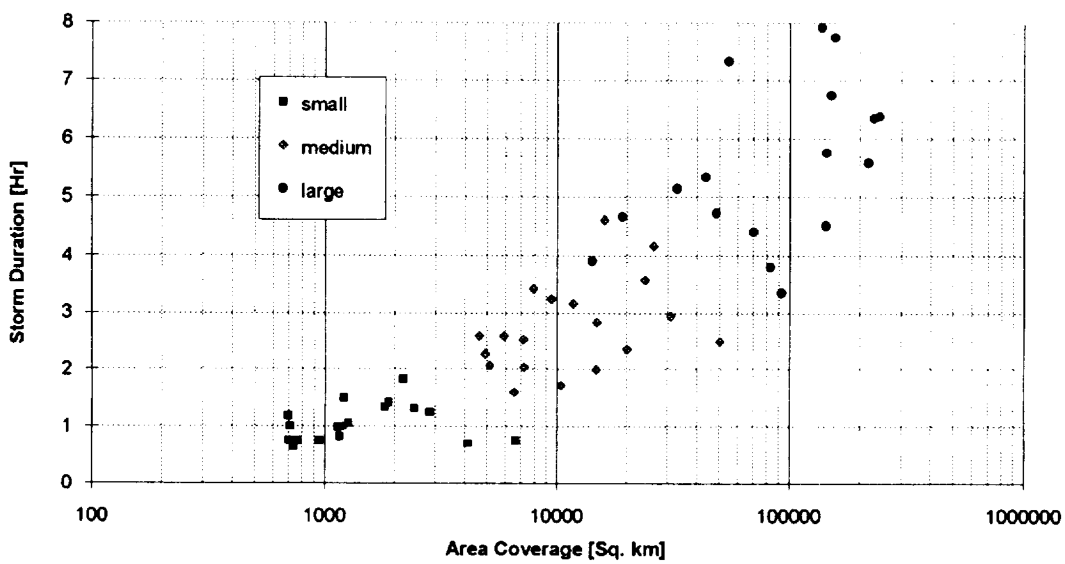

Figure 1. Time as a function of Area Coverage for selected events

As a starting point for determining a PDF, a normal or Gaussian curve fit was attempted. A normal curve is described by the equation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right]$$

Where the symbols μ and σ are the mean and standard deviation of the distribution. This equation generates a curve with the maximum centered at the mean of the distribution and the curve tapering off equally toward either end.

All distributions exhibit a tendency toward a maximum, or a peak. The measurements of central tendency include the mean, median, and mode. The *mean* (also known as the arithmetic mean) is simply the sum of the elements of the distribution divided by the number of elements of the distribution:

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

The *median* of the distribution represents the point where exactly one-half of the distribution lies above *and* below. The *mode* of the distribution represents the most frequently occurring value. In a normal distribution, the mean, median, and mode all coincide at the same location.

Another measure of a distribution is a degree of spread, or deviation from the mean. The deviation is the difference between the mean value and the random variable. The common measurement is the *standard deviation*, which is simply the square root of the *variance*. The variance is calculated as the sum of the square of the deviations divided by the number of elements of the distribution:

$$Var = \frac{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2}{N}$$

Two other measures of the distribution are *skewness* and *kurtosis*. Skewness is a measure of the degree of symmetry about the mean. It is defined as the sum of the cube of the deviations divided by the cube of the standard deviation:

$$Sk = \frac{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)^3}{S.D.^3}$$

Kurtosis is a measure of the peakedness of frequency distributions. It is defined as the sum of the fourth power of the deviations divided by the fourth power of the standard deviation:

$$Ku = \frac{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)^4}{S.D.^4} - 3$$

For a perfectly normal distribution, the measure of kurtosis is 3. Thus, with the "-3" term in the kurtosis equation a perfectly normal distribution would have a kurtosis value of 0.

For the cases studied, values of mean, median, mode, standard deviation, skewness, and kurtosis have been calculated and are displayed in tables 1-3. The theoretical mean and standard deviation, μ and σ, are also calculated and represent the values the distributions would have if perfectly normal.

Cell ID	Min	Max	Mean	Median	Mode	μ	S.D.	σ	Sk	Ku
10_01jul	11	33	16.17	14.38	14.00	22.00	3.71	6.63	1.47	1.89
11_04aug	6	29	11.47	9.52	7.00	17.50	5.12	6.92	1.08	0.52
14_20jul	1	42	17.12	14.31	14.00	21.50	5.80	12.12	1.43	2.31
18_17jun	13	28	15.64	14.20	14.00	20.50	2.83	4.61	1.84	3.61
1_20jul	8	37	16.86	14.71	10.00	22.50	6.37	8.66	0.82	-0.11
28_13jun	11	44	18.08	15.87	13.00	27.50	5.70	9.81	1.39	2.27
2_01aug	12	42	19.45	17.95	18.00	27.00	5.26	8.94	1.33	2.32
2_01jun	9	27	13.01	11.16	10.00	18.00	3.83	5.48	1.35	1.45
2_07jul	1	41	15.14	12.87	8.00	21.00	8.10	11.83	0.70	-0.23
2b_02aug	10	33	15.11	14.08	14.00	21.50	3.42	6.92	1.78	4.85
37_17jun	5	40	15.13	12.76	9.00	22.50	7.17	10.39	1.15	0.85
3_19jun	7	33	12.42	10.47	9.00	17.00	4.58	9.52	2.08	5.16
51_09jun	1	35	9.49	7.29	2.00	18.00	7.23	10.10	1.06	0.67
61_22may	8	36	15.21	12.40	9.00	22.00	6.45	8.37	1.11	0.39
6_01jul	8	43	13.29	10.92	9.00	25.50	5.14	10.39	1.95	4.42
q_21may	6	35	14.55	12.68	8.00	20.50	6.74	8.66	0.84	-0.06
s_20jul	1	17	13.99	13.93	14.00	9.00	2.58	4.90	-4.31	18.63
w_31may	13	35	15.62	14.25	14.00	24.00	3.02	6.63	2.42	7.36

Table 1. Statistics for small time-duration precipitation events. All units are in reflectivity factor (dBz).

Cell ID	Min	Max	Mean	Median	Mode	μ	S.D.	σ	Sk	Ku
11_13jun	3	40	13.55	12.45	12.00	21.50	4.52	10.97	1.11	3.26
12_10jul	10	34	15.48	13.56	13.00	22.00	3.95	7.21	1.76	2.84
16_03aug	2	50	18.13	15.66	9.00	26.00	9.93	14.14	0.68	-0.32
16_21jul	1	52	19.00	16.25	15.00	26.50	6.84	15.00	1.43	2.78
1_07jul	1	41	19.17	17.07	15.00	21.00	6.62	11.83	0.69	-0.03
20_19jun	1	48	13.18	11.02	8.00	24.50	7.54	13.85	0.93	0.81
39_17jun	4	46	16.41	14.47	11.00	25.00	7.60	12.41	0.88	0.46
3_01aug	1	52	23.08	20.58	18.00	26.50	7.54	15.01	0.91	0.27
3_01jun	8	26	12.93	11.20	10.00	17.00	3.52	5.48	1.19	0.62
4_27jun	1	40	13.80	12.23	11.00	20.50	5.16	11.54	0.64	1.13
51_22may	11	42	17.24	15.38	15.00	26.50	5.14	9.23	1.70	3.11
6_02aug	9	35	16.84	16.60	18.00	22.00	3.20	7.79	0.03	0.21
6_03jun	1	30	9.29	7.73	1.00	15.50	6.27	8.65	0.78	-0.04
6_20jul	8	49	16.88	13.80	13.00	28.50	7.02	12.12	1.59	2.21
7_01jul	10	49	16.34	14.16	11.00	29.50	6.13	11.54	1.97	4.87
7_21jul	6	63	16.65	13.66	12.00	34.50	7.91	16.74	1.59	2.46
gg_21may	12	42	17.50	15.49	14.00	27.00	5.04	8.94	2.01	4.33
k_26may	8	41	14.53	13.01	11.00	24.50	5.26	9.81	1.39	2.55
x_18jun	5	46	17.44	15.39	11.00	25.50	7.97	12.12	0.97	0.61

Table 2. Statistics for medium time-duration precipitation events. All units are in reflectivity factor (dBz).

Cell ID	Min	Max	Mean	Median	Mode	μ	S.D.	σ	Sk	Ku
11_12jul	8	58	17.52	14.92	13.00	33.00	6.74	14.72	1.42	2.02
16_17jun	4	46	15.34	13.52	11.00	25.00	5.59	12.41	1.33	2.39
19_19jun	1	58	16.71	15.67	18.00	29.50	7.13	16.74	0.62	0.45
23_11jul	1	57	16.50	14.33	13.00	29.00	9.04	16.45	1.17	1.86
23_13jun	1	56	19.31	18.43	20.00	28.50	6.29	16.16	1.24	3.11
2_06aug	1	45	12.58	10.38	7.00	23.00	8.03	12.98	0.79	0.13
2_20jul	1	55	17.35	15.09	14.00	28.00	6.64	15.87	1.60	3.05
4_01aug	1	53	21.37	18.29	17.00	27.00	8.45	15.30	1.10	0.69
4_10jun	1	39	15.74	15.10	13.00	20.00	6.42	11.25	0.10	-0.53
4_20jul	1	57	18.44	16.29	15.00	29.00	6.57	16.45	1.48	2.69
5_01jul	1	55	19.40	16.48	15.00	28.00	6.92	15.87	1.35	1.61
5_26jun	1	57	16.81	14.93	10.00	29.00	8.96	16.45	0.81	0.57
8_02aug	1	58	17.92	15.54	15.00	29.50	6.64	16.74	1.81	4.18
9_12jun	1	63	18.35	15.75	15.00	32.00	7.56	18.18	1.75	4.06
aa_21may	7	63	18.98	17.83	15.00	35.00	5.67	16.45	1.29	3.22
f1_21may	2	63	20.67	19.52	10.00	32.50	9.21	17.89	0.66	0.14
l_26may	1	55	22.26	22.56	27.00	28.00	8.17	15.87	-0.16	-0.24

Table 3. Statistics for large time-duration precipitation events. All units are in reflectivity factor (dBz).

The only conclusion that can be drawn thus far is that the distributions do not exhibit a normal tendency. Skewness tests indicate most have peaks toward the left side. No goodness-of-fit tests have been applied yet but graphical inspection shows this is the case. The next step in this research involves the following:

1. Apply a goodness-of-fit test to quantify the above observations. This should be fairly straightforward.

2. Determine theoretical distributions for the above distributions based on the Pearson family of curves, specifically types I, III, and V. From the theoretical distributions, some type of goodness-of-fit test needs to be applied so the representative frequency curve can be identified.

3. Within the given parameters, a tolerance value needs to be identified in order to determine which fit is best. In other words, how small do the deviations need to be in order that a fit be accomplished?

I see this thesis becoming an exercise in trying to fit the observed distributions to a few known curves. Since the Pearson curves can be manipulated to become standard distributions (exponential, gamma, etc.), this may be the direction to go. Unfortunately the canned statistics routines I have encountered this far all fall short in being able to specify a distribution. It should not be too hard to code this, however, and once the appropriate parameters for the Pearson curves are identified, implementing that will probably be the next step.

cc: Ron Johnson