NASA TECHNICAL
MEMORANDUM

NASA TM X-53867

PROBABILITY DISTRIBUTIONS FOR THUNDERSTORM ACTIVITY AT CAPE KENNEDY, FLORIDA

By Lee W. Falls, William O. Williford, and
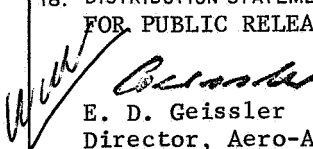   Michael C. Carter
Aero-Astrodynamics Laboratory

May 12, 1970

NASA

George C. Marshall Space Flight Center

Marshall Space Flight Center, Alabama

1. REPORT NO. NASA TM X-53867	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
4. TITLE AND SUBTITLE PROBABILITY DISTRIBUTIONS FOR THUNDERSTORM ACTIVITY AT CAPE KENNEDY, FLORIDA		5. REPORT DATE May 12, 1970
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Lee W. Falls (MSFC), William O. Williford and Michael C. Carter (Univ. of Ga.)		8. PERFORMING ORGANIZATION REPORT #
9. PERFORMING ORGANIZATION NAME AND ADDRESS NASA-George C. Marshall Space Flight Center Marshall Space Flight Center, Alabama 35812		10. WORK UNIT NO.
		11. CONTRACT OR GRANT NO. NAS8-11175
12. SPONSORING AGENCY NAME AND ADDRESS		13. TYPE OF REPORT & PERIOD COVERED Technical Memorandum
		14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

Research partially supported under NASA Contract NAS8-11175 with Aerospace Environment Division, Aero-Astrodynamics Laboratory, Marshall Space Flight Center, Alabama 35812

16. ABSTRACT

Several statistical distribution functions are investigated as prospective models to represent the variation of thunderstorm activity at Cape Kennedy, Florida. Statistical methods are presented using the latest and most comprehensive thunderstorm data available. The conclusion is reached that the negative binomial distribution and a modification of the negative binomial distribution are adequate statistical models to represent "thunderstorm events" and "thunderstorm hits," respectively, at Cape Kennedy, Florida.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
	FOR PUBLIC RELEASE E. D. Geissler Director, Aero-Astrodynamics Laboratory

19. SECURITY CLASSIF. (of this report) UNCLASSIFIED	20. SECURITY CLASSIF. (of this page) UNCLASSIFIED	21. NO. OF PAGES 28	22. PRICE

MSFC - Form 3292 (May 1969)

TECHNICAL MEMORANDUM X-53867

PROBABILITY DISTRIBUTIONS FOR THUNDERSTORM ACTIVITY
AT CAPE KENNEDY, FLORIDA

Lee W. Falls, William O. Williford and Michael C. Carter

SUMMARY

Several statistical distribution functions are investigated as prospective models to represent the variation of thunderstorm activity at Cape Kennedy, Florida. Statistical methods are presented using the latest and most comprehensive thunderstorm data available. The conclusion is reached that the negative binomial distribution and a modification of the negative binomial distribution are adequate statistical models to represent "thunderstorm events" and "thunderstorm hits," respectively, at Cape Kennedy, Florida.

I. INTRODUCTION

Statistical methods of analysis may be divided into two general categories, descriptive and analytical, both of which depend on the basic laws of probability. Descriptive methods reduce large amounts of data to a few meaningful "statistics" such as means and standard deviations. A theoretical statistical model (distribution function) is assumed for the observations, and analytical methods are used to determine how well the empirical data fit this model.

The purpose of this paper is to determine underlying, or basic, theoretical distributions for making probability inferences in regard to two types of thunderstorm activity at Cape Kennedy, Florida. A thunderstorm event will relate to a thunderstorm at Cape Kennedy and immediate surroundings. A thunderstorm hit, denoted by TH, relates to a thunderstorm passing over a point, e.g., a launch site.

Thunderstorms are of primary concern in the design of launch vehicles, in the planning of space missions, and in launch operations at Cape Kennedy because of high winds, lightning hazard, and extreme turbulence associated with this atmospheric phenomenon. The combinations of environmental conditions, including unstable air with a relatively high moisture content, and some type of lifting action present during the summer months make Florida one of the major thunderstorm genesis areas over the entire earth. Two distributions are presented.

The first is the negative binomial distribution to represent the varia-
tion in the number of thunderstorm events per day at Cape Kennedy and
the second is a modified negative binomial distribution to represent
the variation in the number of thunderstorms per day which pass over a
given point, for example, a launch site, at Cape Kennedy.

II. STATISTICAL MODELS

In practical statistics, a discrete probability law is required to
describe events which seem to occur at random; for example, the arrivals
of customers at a service point or the number of accidents and break-
downs in a factory. It is common practice to assume that the frequencies
of such events fit a Poisson distribution. However, the Poisson series
requires the assumption that the probability of the event remains con-
stant. For the Poisson distribution the variance of the distribution
equals its mean. In reality, it is rarely true that the probability of
the event remains constant. Any variation in the probability of the
event, in particular, the tendency for one event to increase the probabil-
ity of another, will increase the variance of the distribution in rela-
tion to the mean -- which means a negative binomial distribution will
better describe the data. A report by the weather observer of a thunder-
storm is proof that the atmosphere is in a state of instability and con-
ditions are present for the formation of further thunderstorm cells;
i.e., the probability of the event is increasing.

Let us consider the first application of the negative binomial
probability distribution by Yule [3]. We will make an analogy between
this application and the distribution of thunderstorms at Cape Kennedy.
Suppose we have a population of people subjected to recurring exposures
to a disease and that during an exposure each member of the population
has an equal probability p of contracting the disease. After x exposures,
the proportions who have contracted the disease 0, 1, 2, ..., times will
be given by

$$q^x, \quad xpq^{x-1}, \quad \frac{x(x-1)}{2!} p^2 q^{x-2}, \quad \ldots \tag{1}$$

where $q = 1 - p$. The terms given by (1) are terms of the binomial series $(q + p)^x$. If k unfavorable exposures to the disease are fatal to the individual, the proportion surviving after x exposures will be given by the first k terms of the binomial $(q + p)^x$. The proportion dying during the xth exposure will be those who contracted the disease (k - 1) times in the first (x - 1) exposures and who contract it again during the xth exposure; i.e., it will be

$$\binom{x - 1}{k - 1} p^{k-1} q^{x-k} \cdot p = \binom{x - 1}{k - 1} p^k q^{x-k} \tag{2}$$

and since deaths do not begin until the kth exposure, the proportion of deaths at the kth, (k + 1)th ... exposure will be

$$p^k \left[1, \ kq, \ k(k + 1) \frac{q^2}{2!}, \ \ldots \right] \tag{3}$$

which are successive terms in the expansion of $p^k(1 - q)^{-k}$, a binomial with a negative index. Thus, the proportions of the original population dying during successive exposures are given by successive terms of the negative binomial distribution with the first deaths occurring at the kth exposure.

Now, the probability of exactly x events (density function) is given by

$$\Pr(x=i) = \binom{x + k - 1}{k - 1} p^k q^x \qquad \text{for} \ \ i \geq 0. \tag{4}$$

Suppose in Yule's [3] classic example we let the people exposed to the disease be analogous to the days in some month, say, June, being exposed to the synoptic condition favorable for the formation of thunderstorms at Cape Kennedy. Now, the number of deaths that result from exposure to the disease will be analogous to the number of thunderstorms that actually develop in June. Now, we have all the days in June subjected to recurrent exposures of synoptic conditions favorable for the formation of thunderstorms. We must assume that each day in June that is exposed to the favorable synoptic conditions has an equal probability p of having a thunderstorm develop. This is a reasonable assumption.

Continuing our analogy, the proportion of thunderstorms that develop at the kth, (k + 1)th ... exposure will be given by (3), successive terms in the expansion of $p^k(1 - q)^{-k}$, a negative binomial whose density function is given by (4). Similar applications of the negative binomial distribution have been performed by Thom [4,5]. Thom considers the Polya distribution (which is a particular case of the negative binomial) in regard to hail frequency series.

Thunderstorms over a point (TH's) require additional assumptions. Only one TH is possible at a given time and a specified time interval must elapse before new activity constitutes a new TH as opposed to a continuation of the previous TH. To incorporate these requirements into the model the negative binomial distribution was modified. The assumptions and modification are presented below.

Assumptions

1. A probability of $\alpha(1 - \alpha)$ is assigned to the possibility of a TH occurrence (nonoccurrence) on any given day.

2. PR (TH occurs in a unit of time | a TH not in progress, $\alpha \neq 0$) = p.

3. T is the number of units of time in the specified time period. The positive integer h is defined by the statement

Pr (TH occurs in a unit of time | a TH in the preceding h - 1 units of time) = 0.

Then the maximum number of occurrences in T units of time is $n \leq [T/h] + 1$, where [T/h] stands for the greatest integer not exceeding T/h.

Under the above assumptions and if x is a random variable denoting the number of TH's per time period T, we have the following model:

$$\Pr(x = 0) = (1 - \alpha) + \alpha q^T \qquad (q = 1 - p)$$

$$\Pr(x = i) = \alpha \left[p^i q^{T-ih} \binom{T - ih + i - 1}{i} \right.$$

$$\left. + p^i q^{T-ih} \sum_{m=1}^{h-1} \binom{T - (i - 1)h + i - m - 2}{i - 1} q^{h-m} \right] \quad \text{for} \quad 0 < i < n$$

(5)

$$\Pr(x = n) = 1 - \Pr(x < n).$$

4

Similar modifications have previously been made on the binomial and Poisson models by Singh [1,2] and Neyman [6]. The parallel in reasoning is best demonstrated by comparing our assumptions with those of Singh [2] in his Poisson birth model. Singh assumes cohabitations are Poisson distributed and proceeds to show that resulting conceptions are modified-Poisson where the value of h was 10 months. Both models have an underlying parent distribution and require special and identical modifications to realize a specific outcome resulting from the activity of the basic population.

III. ESTIMATION

Numerous estimators for the parameters of the negative binomial distribution have been proposed. We have chosen to use the first two-moment method proposed by Cohen [7]. The negative binomial density function given by (4) may be written in terms of the gamma function as

$$Pr(x = i) = \frac{\Gamma(x + k)}{x!\,\Gamma(k)}\, p^k q^x, \quad i \geqq 0, \quad k > 0, \quad 0 \leqq p \leqq 1. \tag{6}$$

The distribution function is given by

$$F(x) = \sum_{x=0}^{n} \frac{\Gamma(x + k)}{x!\,\Gamma(k)}\, p^k q^x, \tag{7}$$

which gives the probability of obtaining a value of x less than or equal to some particular value of x, say x_0.

Now, after some algebraic manipulation of Cohen's estimators, we have for the moment estimators of the parameters k and p

$$k^* = \frac{\bar{x}^2}{s^2 - \bar{x}}, \quad p^* = \frac{k}{k + \bar{x}}, \tag{8}$$

where $\bar{x}$ is the sample mean and s^2 is the sample variance.

The mean M of the negative binomial distribution is given by

$$M = \frac{kq}{p},$$

and the variance V is

$$V = \frac{kq}{p^2} \, .$$

The efficiency of estimating p and k by the method of moments is derived by Fisher [8]. In terms of the parameters used here, the reciprocal of the efficiency is given by

$$\frac{1}{E} = 1 + 2 \left[\frac{1}{3} q \; \frac{2}{(k+2)} + \frac{1}{4} q^2 \; \frac{2 \cdot 3}{(k+2)(k+3)} \right.$$

$$\left. + \frac{1}{5} q^3 \; \frac{2 \cdot 3 \cdot 4}{(k+2)(k+3)(k+4)} + \cdots \right] \, .$$

The modified negative binomial distribution involves two unknown parameters α and p (or q). A statistic is called a minimum Chi-Square (MCS) estimator of α if it is obtained by minimizing, with respect to α, the expression

$$\chi^2 = \sum_{i=0}^{n} \frac{[N_i - NP_i(\alpha,p)]^2}{N_i} \, . \tag{9}$$

See Neyman [9], Kendall and Stuart [10, Vol. II, pp. 91-93] and Singh [1] for a fuller explanation of BAN (best asymptotically normal) estimators. Neyman [9] has shown that the class of MCS estimators are also BAN estimators. These estimators are consistent, asymptotically normal, and asymptotically efficient. Let $P_i(\alpha,p)$ be the probability for i (i = 0, 1, ..., n) TH's per day, and satisfying the regularity conditions in Neyman [9]; and N_i is the observed frequency in the ith class and

$$N = \sum_{i=1}^{n} N_i \, .$$

If $P_i(\alpha,p)$ is linear in α and p, the estimates can easily be found; otherwise, we can linearize them at a properly chosen point $(\bar{\alpha},\bar{p})$ and use the linearized $P_i(\alpha,p)$'s instead of the original $P_i(\alpha,p)$'s to find the estimates. The estimates obtained in this fashion are also BAN, if the point estimates $(\bar{\alpha},\bar{p})$ are consistent. The solutions $(\bar{\alpha},\bar{p})$ to the equations

$$N_0/N = P_0(\alpha,p), \quad N_1/N = P_1(\alpha,p) \tag{10}$$

provides consistent estimates. Letting $P_i'(\alpha,p)$ be the new linearized probabilities, we have

$$P_i'(\alpha,p) = P_i(\bar{\alpha},\bar{p}) + (\alpha - \bar{\alpha}) \left.\frac{\partial P_i(\alpha,p)}{\partial \alpha}\right]_{(\bar{\alpha},\bar{p})} + (p - \bar{p}) \left.\frac{\partial P_i(\alpha,p)}{\alpha P}\right]_{(\bar{\alpha},\bar{p})} \tag{11}$$

as the general equation. In particular, we have

$$P_0'(\alpha,p) = 1 - \alpha(1 - \bar{q}^T) + (p - \bar{p}) \, \bar{\alpha} \, T\bar{q}^{(T-1)} \tag{12}$$

$$P_i'(\alpha,p) = \frac{\alpha}{\bar{\alpha}} P_i(\bar{\alpha},\bar{p}) + (p - \bar{p}) \, \bar{\alpha} \left\{ \bar{p}^{(i-1)}\bar{q}^{(T-ih-1)}[i\bar{q} - (T-ih)\bar{p}]Q(i) \right.$$

$$\left. + \sum_{m=1}^{h-1} \bar{p}^{(i-1)}\bar{q}^{(T-ih+h-m)}[i\bar{q}-(T-ih+h-m)\bar{p}] \, QS(i,m) \right\}, \tag{13}$$

where

$$0 < i < n, \quad Q(i) = \binom{T - ih + i - 1}{i},$$

$$QS(i,m) = \binom{T-(i-1)h + i - m - 2}{i - 1} .$$

$$P_n'(\alpha,p) = 1 - \sum_{i=0}^{n-1} P_i'(\alpha,p). \qquad (14)$$

Replacing $P_i(\alpha,p)$ by $P_i'(\alpha,p)$ in (9), we obtain a modified form of χ^2

$$(\chi^2)' = \sum_{i=0}^{n} \frac{[N_i - P_i'(\alpha,p)]^2}{N_i} \qquad (15)$$

which is minimized to give the MCS estimates α^* and p^*.

IV. DATA SAMPLE

According to standard United States weather observing procedure, a thunderstorm is reported whenever thunder is heard at the station. It is reported along with other atmospheric phenomena on the standard weather observer's form WBAN-10 when thunder is heard and ends 15 minutes after thunder is last heard. Notice that the standard definition of a thunderstorm may include multiple occurrences of thunderstorms. For this reason, we have chosen to use the term "thunderstorm event" as a more appropriate definition for our statistical analysis.

The type of statistical analysis presented is useful primarily for the planning of missions rather than for application to operations. Statistics may be useful up to a few days before a mission. However, at this time the weather forecaster's predictions should be more accurate, and the transition is made from statistical inference to weather forecasting dependent upon the synoptic situation prevailing a few days before the mission.

The data sample used was produced by ESSA, National Weather Records Center, Asheville, North Carolina, under government order number H-76789 for the Terrestrial Environment Branch, Aerospace Environment Division, and is the latest and most comprehensive thunderstorm data available for Cape Kennedy, Florida. The period of record is January 1957 through December 1967.

8

Table 1 summarizes observed frequencies of days that experienced x thunderstorm events for all months, and for the spring, summer, and fall seasons at Cape Kennedy. Table 1a gives the relative frequency of occurrence of days that experienced at least one thunderstorm event at Cape Kennedy for the same reference periods.

Table 1. Frequencies of the Observed Number of Days that Experienced x Thunderstorm Events at Cape Kennedy, Florida for the 11-year Period of Record January 1957 through December 1967.

x	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Spring	Summer	Fall
0	335	295	308	299	266	187	177	185	228	311	321	334	873	549	860
1	4	9	20	18	43	77	80	89	54	17	6	3	81	246	77
2	2	4	9	10	25	40	47	30	33	9	3	2	44	117	45
3		2	3	3	3	17	26	24	12	4		2	9	67	16
4			1		3	6	9	10	3				4	25	3
5					0	2	2	3					0	7	
6					1	1							1	1	
n	341	310	341	330	341	330	341	341	330	341	330	341	1012	1012	1001

Table 1a. Relative Frequency of Days that Experienced at Least One Thunderstorm Event at Cape Kennedy, Florida.

Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Spring	Summer	Fall
.018	.048	.097	.094	.220	.433	.481	.457	.309	.088	.027	.021	.137	.458	.141

Those occurrences which were classified as thunderstorm hits (TH's) from the data sample were of the following types.

1. A thunderstorm was actually reported overhead.

2. A thunderstorm was first reported in a sector and last reported in the opposite sector. This is assuming thunderstorms move in a straight line (over small areas, at least).

9

Only the summer months of June, July and August were selected for TH examination since there were not enough TH data in the sample during the remainder of the year.

The period is 24 hours and T is taken to be 48 units. The value of h is taken as 2 which means that, given a TH occurring, another cannot occur for 30 minutes.

Table 2. Frequencies of the Observed Number of Days that Experience x TH's at Cape Kennedy, Florida for the 11-Year Period of Record January 1957 through December 1967.

x	June	July	August	Summer
0	293	305	300	898
1	27	24	30	81
2	5	6	7	18
3	3	3	2	8
4 or more	2	3	2	7
TOTAL	330	341	341	1012

Table 2a. Relative Frequency of Days that Experienced at Least One TH at Cape Kennedy, Florida.

June	July	August	Summer
.112	.106	.121	.113

V. ANALYSIS

Theoretical summaries of the months and seasons that experience significant thunderstorm events at Cape Kennedy are given in Tables 3 through 13. In all cases, the sample variance was significantly greater

than the sample mean, indicating the negative binomial distribution as the appropriate model.

Notations used in tables 3 through 13 are as follows:

x   = the number of thunderstorm events per day

f_o = observed number of days during the 11-year period of record that experienced x thunderstorm events

r.f.   = the relative frequency of occurrence of x thunderstorm events

F_o = the observed distribution function

f_e = the expected frequencies using the negative binomial distribution

F(x)   = the negative binomial distribution function

$\bar{x}$ = the sample mean

s^2 = the sample variance

k^*, p^* = parameter estimators of the negative binomial distribution

n   = sample size.

Conditional probabilities are also included in the tables.

Consider the month of June (Table 6) as an example. There were 40 days out of 330 days (11 years of Junes) that had exactly two thunderstorm events. This gives a relative frequency (probability) of occurrence of 0.121 of having exactly two thunderstorm events during any day in June. The observed distribution function (F_o) gives a probability of 0.921 of having two or less thunderstorm events during any day in June, or a probability of (1 - 0.921 = 0.079) of having more than two thunderstorm events during any day in June. The negative binomial distribution predicts 36.9 days in June that will experience exactly two thunderstorm events and the probability (F(x)) is 0.928 of having two or less thunderstorm events during any day in June, or a probability of (1 - 0.928 = 0.072) of having more than two thunderstorm events during any day in June. The agreement between theory and observation is very good.

The conditional probabilities are computed from the theoretical frequencies (f_e) by using a double summation technique.

The following tabulation is an example of this technique using the month of June (see Table 6),

x	$\sum$	$\sum\sum$	$\sum\sum\sum$	CONDITIONAL PROBABILITIES i					
				1	2	3	4	5	6
0	181.5								
1	87.7	148.0	244.2	$\frac{244.2}{244.2}=1$					
2	36.9	60.3	96.2	$\frac{96.2}{244.2}=.394$	$\frac{96.2}{96.2}=1$				
3	14.7	23.4	35.9	$\frac{35.9}{244.2}=.147$	$\frac{35.9}{96.2}=.373$	$\frac{35.9}{35.9}=1$			
4	5.7	8.7	12.5	$\frac{12.5}{244.2}=.051$	$\frac{12.5}{96.2}=.130$	$\frac{12.5}{35.9}=.348$	$\frac{12.5}{12.5}=1$		
5	2.2	3.0	3.8	$\frac{3.8}{244.2}=.016$	$\frac{3.8}{96.2}=.040$	$\frac{3.8}{35.9}=.106$	$\frac{3.8}{12.5}=.304$	$\frac{3.8}{3.8}=1$	
6	.8	.8	.8	$\frac{.8}{244.2}=.003$	$\frac{.8}{96.2}=.008$	$\frac{.8}{35.9}=.022$	$\frac{.8}{12.5}=.064$	$\frac{.8}{3.8}=.211$	$\frac{.8}{.8}=1$

Each element in the second summation ($\Sigma\Sigma$) is divided by the appropriate top element in each column as indicated in order to obtain the conditional probabilities; i.e., in each column under conditional probabilities, given i thunderstorm events (i = 1, 2, 3, ...), the probability of having j additional thunderstorm events (j = 0, 1, 2, ...) is given by

$$P(i + j \mid i) = \frac{(i + j)\text{th element}}{(i)\text{th element}} .$$ (16)

(Equation (16) refers to the above tabulation only.) For example, for i = 2; given two thunderstorm events on any day in June, what is the probability of having two additional thunderstorm events (j = 2) on that day in June? From (16),

$$P(4 \mid 2) = \frac{4\text{th element}}{2\text{nd element}} = \frac{12.5}{96.2} = 0.130.$$

12

Also, given four thunderstorm events on any day in June (i = 4), the probability of having one additional thunderstorm event (j = 1) on that same day in June is 0.304.

Table 3. March-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida.

x	f_o	r.f.	F_o	f_e	F(x)	Conditional Probability			
						i			
						1	2	3	4
0	308	.902	.902	305.4	.896				
1	20	.059	.961	25.5	.970	1			
2	9	.026	.987	6.7	.990	.277	1		
3	3	.009	.996	2.2	.996	.078	.281	1	
4	1	.003	1.000	.8	.999	.016	.059	.211	1

$\bar{x} = 0.150 \quad s^2 = 0.268 \quad k^* = 0.189 \quad p^* = 0.558 \quad n = 341$

Table 4. April-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida.

x	f_o	r.f.	F_o	f_e	F(x)	Conditional Probability		
						i		
						1	2	3
0	299	.906	.906	295.9	.897			
1	18	.055	.961	25.3	.973	1		
2	10	.030	.991	6.1	.992	.226	1	
3	3	.009	1.000	1.8	.997	.042	.186	1

$\bar{x} = 0.142 \quad s^2 = 0.237 \quad k^* = 0.214 \quad p^* = 0.600 \quad n = 330$

Table 5. May-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida

x	f_o	r.f.	F_o	f_e	F(x)	Conditional Probability					
						i					
						1	2	3	4	5	6
0	266	.779	.780	262.6	.770						
1	43	.126	.906	52.4	.924	1					
2	25	.073	.979	16.6	.972	.339	1				
3	3	.009	.988	5.9	.989	.120	.354	1			
4	3	.009	.997	2.2	.996	.041	.122	.345	1		
5	0	.000	.997	.9	.998	.013	.037	.106	.306	1	
6	1	.003	1.000	.3	.999	.003	.007	.021	.061	.200	1

$\bar{x} = 0.352$ $s^2 = 0.621$ $k^* = 0.460$ $p^* = 0.567$ $n = 341$

Table 6. June-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida

x	f_o	r.f.	F_o	f_e	F(x)	Conditional Probability					
						i					
						1	2	3	4	5	6
0	187	.567	.567	181.5	.550						
1	77	.233	.800	87.7	.816	1					
2	40	.121	.921	36.9	.928	.394	1				
3	17	.052	.973	14.7	.972	.147	.373	1			
4	6	.018	.991	5.7	.989	.051	.130	.348	1		
5	2	.006	.997	2.2	.996	.016	.040	.106	.304	1	
6	1	.003	1.000	.8	.999	.003	.008	.022	.064	.211	1

$\bar{x} = 0.752$ $s^2 = 1.169$ $k^* = 1.354$ $p^* = 0.643$ $n = 330$

14

Table 7. July-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida.

x	f_o	r.f.	F_o	f_e	F(x)	Conditional Probability i				
						1	2	3	4	5
0	177	.519	.519	166.2	.487					
1	80	.234	.753	99.4	.779	1				
2	47	.138	.891	45.4	.912	.399	1			
3	26	.076	.967	18.6	.967	.143	.357	1		
4	9	.026	.993	7.2	.988	.044	.110	.307	1	
5	2	.006	1.000	2.7	.996	.009	.023	.066	.214	1

$\bar{x} = 0.874$ $s^2 = 1.277$ $k^* = 1.893$ $p^* = 0.684$ $n = 341$

Table 8. August-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida.

x	f_o	r.f.	F_o	f_e	F(x)	Conditional Probability i				
						1	2	3	4	5
0	185	.542	.542	180.2	.528					
1	89	.261	.803	92.2	.799	1				
2	30	.088	.891	40.5	.918	.399	1			
3	24	.070	.961	16.9	.967	.146	.366	1		
4	10	.029	.990	6.8	.987	.046	.116	.316	1	
5	3	.009	1.000	2.7	.995	.010	.026	.070	.221	1

$\bar{x} = 0.809$ $s^2 = 1.280$ $k^* = 1.391$ $p^* = 0.632$ $n = 341$

Table 9. September-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida.

x	f_o	r.f.	F_o	f_e	$F(x)$	Conditional Probability i			
						1	2	3	4
0	228	.691	.691	219.2	.664				
1	54	.164	.855	73.1	.886	1			
2	33	.100	.955	24.8	.961	.316	1		
3	12	.036	.991	8.5	.987	.089	.283	1	
4	3	.009	1.000	2.9	.995	.018	.057	.203	1

$\bar{x} = 0.509 \quad s^2 = 0.777 \quad k^* = 0.967 \quad p^* = 0.655 \quad n = 330$

Table 10. October-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida.

x	f_o	r.f.	F_o	f_e	$F(x)$	Conditional Probability i		
						1	2	3
0	311	.911	.911	307.7	.902			
1	17	.050	.961	24.2	.973	1		
2	9	.026	.987	6.1	.991	.235	1	
3	4	.012	1.000	1.9	.997	.045	.192	1

$\bar{x} = 0.138 \quad s^2 = 0.242 \quad k^* = 0.182 \quad p^* = 0.570 \quad n = 341$

Table 11. Spring (March, April, May)-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida.

x	f_o	r.f.	F_o	f_e	$F(x)$	Conditional Probability i					
						1	2	3	4	5	6
0	873	.863	.863	863.6	.853						
1	81	.080	.943	103.7	.956	1					
2	44	.043	.986	29.2	.985	.312	1				
3	9	.009	.995	9.8	.994	.106	.339	1			
4	4	.004	.999	3.5	.998	.035	.113	.335	1		
5	0	.000	.999	1.3	.999	.011	.034	.101	.303	1	
6	1	.001	1.000	.5	1.000	.002	.007	.022	.066	.217	1

$\bar{x} = 0.215 \quad s^2 = 0.386 \quad k^* = 0.271 \quad p^* = 0.557 \quad n = 1012$

16

Table 12. Summer (June, July, August)-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida.

x	f_o	r.f.	F_o	f_e	F(x)	Conditional Probability i					
						1	2	3	4	5	6
0	549	.542	.542	527.8	.522	1					
1	246	.243	.785	279.6	.798	1					
2	117	.116	.901	122.7	.919	.404	1				
3	67	.066	.967	50.1	.969	.153	.379	1			
4	25	.025	.992	19.7	.988	.054	.133	.352	1		
5	7	.007	.999	7.6	.995	.017	.041	.108	.307	1	
6	1	.001	1.000	2.9	.998	.004	.009	.023	.067	.216	1

$\bar{x} = 0.812 \quad s^2 = 1.245 \quad k^* = 1.523 \quad p^* = 0.652 \quad n = 1012$

Table 13. Fall (September, October, November)-Negative Binomial Distribution for Thunderstorm Events at Cape Kennedy, Florida.

x	f_o	r.f.	F_o	f_e	F(x)	Conditional Probability i			
						1	2	3	4
0	860	.859	.859	845.2	.844	1			
1	77	.077	.936	109.5	.954	1			
2	45	.045	.981	30.6	.984	.286	1		
3	16	.016	.997	10.1	.994	.080	.281	1	
4	3	.003	1.000	3.6	.998	.017	.058	.208	1

$\bar{x} = 0.227 \quad s^2 = 0.397 \quad k^* = 0.302 \quad p^* = 0.571 \quad n = 1001$

17

Using the modified negative binomial distribution as a model for TH activity, we see from Table 14 that the model fits particularly well in the 0, 1, and 2 classes and not so well in the 3 and 4 classes. The notable exception is the August data.

Notations used in table 14 are as follows:

x = the number of TH's per day

f_o = the observed number of days during the 11-year period of record that experienced x TH's

$r.f.$ = the relative frequency of occurrence of x TH's

F_o = the observed distribution function

f_e = the expected frequencies using the modified negative binomial distribution

$F(x)$ = the modified negative binomial distribution function

α^*, p^* = parameter estimators of the modified negative binomial distribution

n = sample size.

Table 14. June, July, August, Summer-Modified Negative Binomial Distribution for TH's at Cape Kennedy, Florida

	JUNE					JULY				
x	f_o	$r.f.$	F_o	f_e	$F(x)$	x	f_o	$r.f.$	F_o	f_e	$F(x)$
0	293	.888	.888	295.18	.894	0	305	.894	.894	306.99	.900
1	27	.082	.970	26.21	.973	1	24	.070	.964	23.05	.968
2	5	.015	.985	6.54	.993	2	6	.018	.982	7.97	.991
3	3	.009	.994	1.02	.996	3	3	.009	.991	1.71	.996
4 or more	2	.006	1.000	1.05	.999	4 or more	3	.009	1.000	1.28	1.000

$\alpha^* = .25467$, $p^* = .01108$, $n = 330$ | $\alpha^* = .19135$, $p^* = .01523$, $n = 341$

18

Table 14 (Continued)

	AUGUST						SUMMER				
x	f_o	r.f.	F_o	f_e	F(x)	x	f_o	r.f.	F_o	f_e	F(x)
0	300	.879	.879	300.55	.881	0	898	.887	.887	902.38	.892
1	30	.088	.967	29.66	.968	1	81	.080	.967	79.00	.970
2	7	.021	.988	8.12	.992	2	18	.018	.985	22.88	.993
3	2	.006	.994	1.38	.996	3	8	.008	.993	4.12	.997
4 or more	2	.006	1.000	1.29	1.000	4 or more	7	.007	1.000	3.61	1.000
$\alpha^*=.26740, \quad p^* = .01214, \quad n = 341$						$\alpha^* = .23461, \quad p^* = .01282, \quad n = 1012$					

VI. CONCLUSIONS

There are many advantages in the use of a theoretical statistical
model for predicting variables such as thunderstorm events or thunder-
storm hits at Cape Kennedy. Once sufficient representative samples have
been collected and analyzed and the validity of the theory is established,
the theoretical model becomes "deterministic" and may be applied uni-
versally to the variable under consideration. Another advantage of theory
over empirical statistics is the use of the acceptable theoretical func-
tion for making probability inferences concerning values of the variable
outside of the range of observation. It is often desired to make pre-
dictions relating to these "never observed" values, and the theoretical
approach permits one to do so. It should be pointed out that no
theoretical function can explain all observations for which it is the
proposed model. Some areas of non-agreement must occur between theory
and observation. These areas should be considered as expected deviations
of the observations from the "fitted" theoretical curve.

The physical properties necessary for the application of the nega-
tive binomial distribution have been shown to be present in our experi-
ment concerning the number of thunderstorm events at Cape Kennedy. In
all the samples considered, the sample variance exceeded the sample
mean, indicating the negative binomial as an appropriate model. Our
comparison with Yule's [3] classic application of the negative binomial

19

substantiates its validity to represent the number of thunderstorm events at Cape Kennedy. Both the negative binomial and Poisson distributions were tried as prospective models for thunderstorm events at Cape Kennedy. The negative binomial gave a "better" fit in all cases than the Poisson distribution.

The modification of the negative binomial distribution presented here has been shown to be an adequate model to represent the variation in the number of thunderstorm hits per day at a launch site. Comparisons were also made using the same modification on the binomial and Poisson distributions as was used here on the negative binomial distribution. The modified negative binomial distribution gave a "better" fit in all cases.

The X^2 "goodness of fit" test was applied to all cases where data were sufficiently large to admit its use. For both the negative binomial and the modified negative binomial distributions, the computed X^2 values were a minimum relative to those of the other distributions mentioned above.

Using statistical methods, it has been demonstrated that the negative binomial and the modified negative binomial distributions are adequate to represent the variation in thunderstorm events and thunderstorm hits at Cape Kennedy, Florida.

REFERENCES

1. Singh, S. N., "Probability Models for the Variation in the Number of Births per Couple," J. Amer. Stat. Assoc., 58, 721-727, 1963.

2. Singh, S. N., "A Chance Mechanism of the Variation in the Number of Births per Couple," J. Amer. Stat. Assoc., 63, 209-213, 1968.

3. Yule, G. U., "On the Distribution of Deaths with Age When the Causes of Death Act Cumulatively, and Similar Frequency Distributions," J.R.S.S., 73, 26-35, 1910.

4. Thom, H. C. S., "The Frequency of Hail Occurrence," Arch. Meteorol., Geophys., Bioklimatol. [B] 8, No. 2, 185-194, 1957.

5. Thom, H. C. S., "A Method for the Evaluation of Hail Suppression," J. Applied Math. and Phys. (ZAMP), 9, 37-64, 1958.

6. Neyman, J., "On Problems of Estimating the Number of Schools of Fish," Univ. Calif. Pub. Stat., 1, 21-36, 1949.

7. Cohen, A. C., "Estimation in the Negative Binomial Distribution," NASA TM X-53372, MSFC, 1965.

8. Fisher, R. A., Contributions to Mathematical Statistics, John Wiley and Sons, Inc., 1950.

9. Neyman, J., "Contribution to the Theory of X^2 Test," Proc. Berk. Symp. Math. Stat. and Prob., 239-273, 1949.

10. Kendall, M. G. and A. Stuart, The Advanced Theory of Statistics, Vol. II, Hafner Publishing Company, 1963.

BIBLIOGRAPHY

1. Cohen, A. C., "A Note on Certain Discrete Mixed Distributions," Biometrics, 22, 566-572, 1966.

2. Neyman, J., "On A New Class of 'Contagious' Distributions, Applicable in Entomology and Bacteriology," Ann. Math. Stat., 10, 35-57, 1939.

3. Kendall, M. G. and A. Stuart, The Advanced Theory of Statistics, Vol. I, Hafner Publishing Company, 1963.

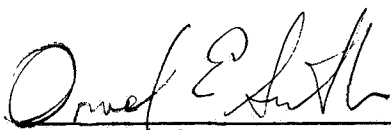PROBABILITY DISTRIBUTIONS FOR THUNDERSTORM ACTIVITY
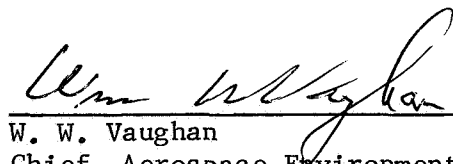AT CAPE KENNEDY, FLORIDA

by

Lee W. Falls, William O. Williford and Michael C. Carter

The information in this report has been reviewed for security classification. Review of any information concerning Department of Defense or Atomic Energy Commission programs has been made by the MSFC Security Classification Officer. This report, in its entirety, has been determined to be unclassified.

This document has also been reviewed and approved for technical accuracy.

O. E. Smith
Chief, Terrestrial Environment Branch

W. W. Vaughan
Chief, Aerospace Environment Division

E. D. Geissler
Director, Aero-Astrodynamics Laboratory