NASA CONTRACTOR
REPORT

NASA CR-61370

USE OF THE COMPOUND NEGATIVE BINOMIAL-TRUNCATED POISSON DISTRIBUTION IN THUNDERSTORM PREDICTION

By Michael C. Carter and William O. Williford
University of Georgia
Department of Statistics
Athens, Georgia

December 1971

Prepared for

NASA-GEORGE C. MARSHALL SPACE FLIGHT CENTER
Marshall Space Flight Center, Alabama 35812

~~16090~~
11674

ERRATA

NASA CR-61370
December 1971

N72-14641

The attached material should be substituted in your copy of the above numbered NASA Contractor Report. The cover and title page has been corrected to reflect the proper subject title, author, and University department. Page 4 has been corrected to show a plus sign rather than a minus sign in equation 11.

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
NASA CR-61370		

4. TITLE AND SUBTITLE	5. REPORT DATE
USE OF THE COMPOUND NEGATIVE BINOMIAL-TRUNCATED POISSON DISTRIBUTION IN THUNDERSTORM PREDICTION	December 1971
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR (S)	8. PERFORMING ORGANIZATION REPORT #
Michael C. Carter and William O. Williford	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
University of Georgia Department of Statistics Athens, Georgia	11. CONTRACT OR GRANT NO. NAS8-11175
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT & PERIOD COVERED
NASA Washington, D. C. 20546	CONTRACTOR REPORT
	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES Technical Coordinator: Lee W. Falls, Aerospace Environment Division, Aero-Astrodynamics Laboratory, Marshall Space Flight Center, Alabama.

16. ABSTRACT

A probability model is presented for the distribution of thunderstorms over a small area given that thunderstorm events (1 or more thunderstorms) are occurring over a larger area. The model incorporates the Negative Binomial and Truncated Poisson distributions. Probability tables for Cape Kennedy for spring, summer, and fall months and seasons are presented. The computer program used to compute these probabilities is appended.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
thunderstorm prediction negative binomial distribution truncated poisson distribution statistical analysis probability distributions	UNCLASSIFIED-UNLIMITED WWV E. D. Geissler E. D. GEISSLER Dir, Aero-Astrodynamics Lab, MSFC

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
Unclassified	Unclassified	18	$3.00

MSFC - Form 3292 (May 1969)

USE OF THE COMPOUND NEGATIVE BINOMIAL-TRUNCATED POISSON DISTRIBUTION IN THUNDERSTORM PREDICTION

Michael C. Carter and William O. Williford

University of Georgia

1. Introduction

This article summarizes an investigation into predicting the number of thunderstorms (TH's) over a small area. It has been established by Falls [1] that thunderstorm events (THE's), i.e., one or more TH's, occurring over a large area are reliably predicted by the negative binomial distribution. Williford and Carter [2] have shown that a "modified" negative binomial distribution gives reliable predictions for frequencies of thunderstorms over a "point" irrespective of the frequency of THE occurrences. It is considered likely that the number of TH's within a THE follows the Poisson law with the zero class truncated. The resulting model gives probability predictions over a small area (of undetermined size) as an individual THE is active over a small area.

This research was supported by NASA under contract No. NASA 11175.

2. Probability Model

The probability generating function (PGF) technique will be used to get the joint distribution for THE's and TH's and the marginal distribution of TH's. This technique is given in Bailey [3] or Feller [4].

We make the following assumptions:

a. THE's are distributed as negative binomial, i.e.,

$$\Pr\{X \text{ THE's}\} = \frac{\Gamma(X+r)}{X!\,\Gamma(r)} p^r q^X \quad X = 0,1,2,\ldots, \quad r > 0, \tag{1}$$

with the PGF given by

$$F(t) = p^r(1-qt)^{-r} \tag{2}$$

b. TH's, given a THE has occurred, are distributed as Poisson with the zero class truncated, i.e.,

$$\Pr\{Y \text{ TH's} | x=1\} = \exp(-\mu)\ \mu^Y/Y!(1-\exp(-\mu)) \quad Y = 1,2,\ldots \tag{3}$$

with the PGF given by

$$G(t) = (\exp[\mu(t-1)] - \exp(-\mu))/(1-\exp(-\mu)). \tag{4}$$

Using these results the PGF for TH's is the expression F[G(t)], i.e.,

$$F[G(t)] = p^r\{1-q\,\frac{\exp[\mu(t-1)] - \exp(-\mu)}{1-\exp(-\mu)}\}^{-r} \tag{5}$$

Expanding this expression gives the probability of n TH's as the coefficient of t^n, i.e.,

$$\Pr\{n \text{ TH's}\} = \sum_{i=1}^{n} \frac{\Gamma(i+r)}{i!\Gamma(r)} p^r q^i \frac{\exp(-i\ \mu)(i^n - k(i,n))\mu^n}{n![1 - \exp(-\mu)]^i}, \quad n \geq 1,\ i \leq n \quad (6)$$

and the joint distribution of i THE's and n TH's ($0 < i \leq n$) is given by

$$\Pr\{i \text{ THE's},\ n \text{ TH's}\} = \frac{\Gamma(i+r)}{i!\Gamma(r)} p^i q^r \frac{\exp(-i\ \mu)(i^n - k(i,n))\mu^n}{n![1 - \exp(-\mu)]^i}. \quad (7)$$

The function k(i,n) is a multinomial sum of the form

$$k(i,n) = i^n - \sum_{j_1} \sum_{j_2} \cdots \sum_{j_i} \frac{n!}{j_1! j_2! \cdots j_i!} \quad (8)$$

where $\sum_{m=1}^{i} j_m = n$ and $j_m > 0$ for $m = 1,2, \ldots i$.

Specifically, we have $k(1,n) = 0$, $k(2,n) = 2$, $k(3,n) = 3\{2^n-2\} + 3$ and, in general,

$$k(i,n) = \sum_{j=1}^{i-1} \binom{i}{j}[j^n - k(j,n)], \qquad i > 1,\ n > 0. \tag{9}$$

Using the result that

$$\frac{d\, F[G(t)]}{dt}\bigg|_{t=1} = \text{mean(TH)}$$

and

$$\frac{d^2 F[G(t)]}{dt^2}\bigg|_{t=1} + \frac{d\, F[G(t)]}{dt}\bigg|_{t=1} - \left[\frac{d\, F[G(t)]}{dt}\bigg|_{t=1}\right]^2 = \text{Var(TH)}$$

we have the results

$$\text{mean}(T\hat{H}) = r\, q\, \mu/[(1 - \exp(-\mu))p] \tag{10}$$

and

$$\text{Var(TH)} = \text{mean(TH)}\ \{1 + \mu + \frac{\text{mean(TH)}}{r}\}. \tag{11}$$

3. Estimation

The expressions (5), (6) and (7) are quite involved and difficult to use in estimating p, r, and μ. However, by making one seemingly reasonable assumption a simple estimation procedure is available.

Estimation of negative binomial parameters p and r is straight forward (see Cohen [5]). Assuming the number of TH's, given a THE occurrence, is independent of the number of THE's occurring, the truncated Poisson parameter μ can be estimated by the usual techniques (see Rider [6] and David and Johnson [7]).

Using the data in Appendices A and B and the techniques referenced above estimates for p, r and μ were calculated. Excluding the winter months, these estimates are presented in Table 1.

Table 1

Estimates of the Parameters p, r, and μ.

	$\bar{p}^*$	$\bar{k}^*$	$\bar{\mu}^*$
March	.558	.189	.645
April	.600	.214	.507
May	.567	.460	.707
June	.643	1.354	.624
July	.684	1.893	.766
August	.632	1.391	.686
September	.655	.967	.807
October	.570	.182	.525
Spring			
March, April, May	.557	.271	.650
Summer			
June, July, August	.652	1.523	.760
Fall			
Sept., Oct., Nov.	.571	.302	.570

*These values are obtained from Falls [1].

For evaluation of expressions (6) and (7) a FORTRAN program was written (Appendix C). For the months and seasons in Table 1 and for values of i = 1, ..., 6, n = i, ..., 6, these results are given in Tables 2-12. In these tables the new totals represent the marginal distribution for n TH's and column totals represent the marginal distribution of i THE's. It should be noted that the probabilities in Tables 2-12 will not sum to unity. This is due to the truncation of n and i at 6.

Table 2

Probabilities of i THE's and n TH's for March

	i=0	1	2	3	4	5	6	Total
n=0	.8956							.8956
1	.0	.0533						.0533
2	.0	.0172	.0100					.0271
3	.0	.0037	.0064	.0023				.0124
4	.0	.0006	.0024	.0022	.0006			.0058
5	.0	.0001	.0007	.0012	.0007	.0002		.0028
6	.0	.0	.0001	.0005	.0005	.0002	.0	.0013
Total	.8956	.0748	.0196	.0062	.0018	.0004	.0	

Table 3

Probabilities of i THE's and n TH's for April

	i=0	1	2	3	4	5	6	Total
n=0	.8964							.8964
1	.0	.0589						.0589
2	.0	.0149	.0110					.0259
3	.0	.0025	.0056	.0025				.0106
4	.0	.0003	.0016	.0019	.0006			.0045
5	.0	.0	.0004	.0008	.0006	.0002		.0020
6	.0	.0	.0001	.0002	.0003	.0002	.0	.0009
Total	.8964	.0767	.0186	.0054	.0016	.0004	.0	

Table 4

Probabilities of i THE's and n TH's for May

	i=0	1	2	3	4	5	6	Total
n=0	.7703							.7703
1	.0	.1055						.1055
2	.0	.0373	.0229					.0602
3	.0	.0088	.0162	.0056				.0306
4	.0	.0016	.0067	.005i	.0014			.0156
5	.0	.0002	.0020	.0035	.0020	.0004		.0082
6	.0	.0	.0005	.0015	.0016	.0007	.0001	.0043
Total	.7703	.1534	.0484	.0165	.0050	.0011	.0001	

Table 5

Probabilities for i THE's and n TH's for June

	i=0	1	2	3	4	5	6	Total
n=0	.5499							.5499
1	.0	.1915						.1915
2	.0	.0597	.0579					.1177
3	.0	.0124	.0362	.0167				.0652
4	.0	.0019	.0132	.0156	.0047			.0354
5	.0	.0002	.0035	.0081	.0056	.0013		.0190
6	.0	.0	.0008	.0030	.0039	.0020	.0003	.0101
Total	.5499	.2658	.1115	.0434	.0144	.0033	.0003	

Table 6

Probabilities for i THE's and n TH's for July

	i=0	1	2	3	4	5	6	Total
n=0	.4873							.4873
1	.0	.1940						.1940
2	.0	.0743	.0590					.1333
3	.0	.0190	.0452	.0161				.0803
4	.0	.0036	.0202	.0185	.0041			.0465
5	.0	.0006	.0066	.0118	.0063	.0010		.0264
6	.0	.0001	.0017	.0054	.0053	.0020	.0002	.0147
Total	.4873	.2915	.1328	.0518	.0157	.0030	.0002	

Table 7

Probabilities for i THE's and n TH's for August

	i=0	1	2	3	4	5	6	Total
n=0	.5282							.5282
1	.0	.1882						.1882
2	.0	.0645	.0576					.1221
3	.0	.0148	.0395	.0167				.0710
4	.0	.0025	.0158	.0172	.0047			.0402
5	.0	.0003	.0046	.0098	.0064	.0013		.0225
6	.0	.0	.0011	.0040	.0048	.0022	.0004	.0125
Total	.5282	.2704	.1187	.0477	.0159	.0035	.0004	

Table 8

Probabilities for i THE's and n TH's for September

	i=0	1	2	3	4	5	6	Total
n=0	.6642							.6642
1	.0	.1441						.1441
2	.0	.0581	.0318					.0899
3	.0	.0156	.0257	.0071				.0483
4	.0	.0032	.0121	.0085	.0016			.0253
5	.0	.0005	.0042	.0057	.0025	.0003		.0133
6	.0	.0001	.0012	.0028	.0022	.0007	.0001	.0070
Total	.6642	.2216	.0748	.0241	.0063	.0011	.0001	

Table 9

Probabilities for i THE's and n TH's for October

	i=0	1	2	3	4	5	6	Total
n=0	.9028							.9028
1	.0	.0537						.0537
2	.0	.0141	.0104					.0245
3	.0	.0025	.2254	.0025				.0104
4	.0	.0003	.0017	.0019	.0006			.0046
5	.0	.0	.0004	.0009	.0007	.0002		.0021
6	.0	.0	.0001	.0003	.0004	.0002	.0	.0010
Total	.9028	.0706	.0179	.0055	.0017	.0004	.0	

Table 10

Probabilities for i THE's and n TH's for Spring

	i=0	1	2	3	4	5	6	Total
n=0	.8533							.8533
1	.0	.0727						.0727
2	.0	.0236	.0145					.0382
3	.0	.0051	.0094	.0035				.0180
4	.0	.0008	.0036	.0034	.0009			.0087
5	.0	.0001	.0010	.0018	.0012	.0002		.0043
6	.0	.0	.0002	.0007	.0008	.0004	.0001	.0022
Total	.8533	.1024	.0288	.0094	.0029	.0006	.0001	

Table 11

Probabilities for i THE's and n TH's for Summer

	i=0	1	2	3	4	5	6	Total
n=0	.5213							.5213
1	.0	.1845						.1845
2	.0	.0707	.0541					.1242
3	.0	.0178	.0411	.0148				.0736
4	.0	.0034	.0182	.0168	.0039			.0423
5	.0	.0005	.0059	.0107	.0059	.0010		.0240
6	.0	.0001	.0016	.0049	.0049	.0019	.0003	.0135
Total	.5213	.2763	.1209	.0471	.0146	.0029	.0003	

Table 12

Probabilities for i THE's and n TH's for Fall

	i=0	1	2	3	4	5	6	Total
n=0	.8443							.8443
1	.0	.0812						.0812
3	.0	.0044	.0096	.0041				.0181
4	.0	.0006	.0032	.0035	.0011			.0084
5	.0	.0001	.0008	.0017	.0012	.0003		.0040
6	.0	.0	.0002	.0006	.0018	.0004	.0001	.0030
Total	.8443	.1094	.0305	.0099	.0031	.0007	.0001	

4. Summary

The model is an "in between" result and represents a compromise between Falls [1] who investigated the general phenomenon and Williford and Carter [2] who investigated TH distributions over a point. As an example, Falls [1] gives Pr 1 THE in July = .2915, Table 6 gives Pr 1 TH in July = .1940, (say, in an area including vehicle assembly building and surrounding launch pads) and Williford and Carter [2] give Pr 1 TH over a point = .0723 (say, over vehicle assembly building or one of the launch pads). Then the values

Probability over a large "region"	.2915
Probability over a small "subregion"	.1940
Probability over a "point"	.0723

substantially demonstrate the property asserted above.

Formulas (10) and (11) can be expressed in the forms

$$\text{Mean(TH)} = (\text{mean of Neg. Binomial})(\text{Mean of Truncated Poisson}) \tag{12}$$

$$\begin{aligned}\text{Variance TH} = {} & (\text{Mean of Neg. Binomial})(\text{Variance of Truncated Poisson}) \\ & + (\text{Mean of Truncated Poisson})^2(\text{Variance of Neg. Binomial})\end{aligned} \tag{13}$$

Using these formulas along with Tables 1-12 the results concerning the whole Cape Kennedy region could be restricted to a smaller region, e.g., the vehicle assembly building and surrounding launch pads.

Appendix A*

Frequencies of Thunderstorm Events Containing

X Thunderstorms at Cape Kennedy

X	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1	7	18	50	37	87	182	192	201	122	35	13	12
2	1	4	15	10	25	55	90	60	35	10	2	1
3		1	3	2	5	6	10	10	5	2		
4			1		3	5	4	3	4			
5							2	2	2			

X	Spring	Summer	Fall	Winter	Annual
1	174	575	170	37	956
2	50	205	47	6	308
3	10	26	7	1	44
4	4	12	4		20
5		4	2		6

* This data was furnished by the Terrestrial Environment Branch, Aerospace Environment Division, Aero-Astrodynamics Laboratory, George C. Marshall Space Flight Center, Alabama. The frequencies for classes $x = 2, 3, 4, 5$ are approximate and are based on the information available. The $x = 1$ class is exact and the rest represents a partition of the $x = 2$ or more class.

Appendix B*

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.	Mar.	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
0	335	295	308	299	266	187	177	185	228	311	321	334
1	4	9	20	18	43	77	80	89	54	17	6	3
2	2	4	9	10	25	40	47	30	33	9	3	2
3		2	3	3	3	17	26	24	12	4		2
4			1		3	6	9	10	3			
5					0	2	2	3				
6					1	1						

X	Spring	Summer	Fall
0	873	549	860
1	81	246	77
2	44	117	45
3	9	67	16
4	4	25	3
5	0	7	
6	1	1	

* Reprinted from Falls [1].

Appendix C

The program is written in FORTRAN (Version H) suitable for processing on the IBM 360-67. Inputs required are Negative Binomial parameter estimates P and R, Truncated Poisson parameter estimate PMU and NTH ($\leq$ 10) which is the TH's for which probabilities are computed.

Program Listing

```
      DIMENSION PK(10,10), PROB(10,10), PROBM(10,10), TNBIN(10),
QK(10,10) 1000 READ(5,1)P,RK,PMU,NTH
    1 FORMAT(3F 8.4,I3)
      PI = P**RK
      DO 2 NX = 1,NTH
    2 QK(1,NX) = 0.
      DO 3 NX = 1,NTH
      DO 3 I = 2,NTH
      QK(I,NX) = 0.
      II = I - 1
      FI = I
      FII = FI + 1.
      XI = GAMMA(FII)
      DO 30 J = 1,II
      TJ - J
      FJ - TJ + 1
```

```
   FIJ = FI - TJ + 1.
   XJ = GAMMA(FJ)
   XIJ = GAMMA(FIJ)
   FACT = XI/(XJ*XIJ)
30 QK(I,NX) = QK(I,NX) + FACT*(TJ**NX - QK(J,NX))
 3 PK(I,NX) = QK(I,NX)
   DO 31 NX = 1,NTH
31 PK(1,NX) = 0.
   TM = RK
   XM = GAMMA(TM)
   DO 4 NX = 1,NTH
   PROBM(NX) = 0.
   FNX = NX
   TNX = FNX + 1.
   D1 = GAMMA(TNX)
   DO 5 I = 1,NX
   FI = I
   FII = FI + 1.
   TK = RK + FI
   XI = GAMMA(FII)
   XK = GAMMA(TK)
   FACT = XK/(XM*XI)
   D2 = (1. - EXP(-PMU))**I
   QI = (1. - P)**I
   EX = EXP(-FI*PMU)
   EX1 = (PMU**NX)*(FI**NX - PK(I,NX))
```

```
      PROB(I,NX) = FACT*PI*QI*EX*EX1/(D1*D2)
    5 PROBM(NX) = PROBM(NX) + PROB(I,NX)
    4 CONTINUE
      DO 32 * = 1,NTH
      TNBIN(I) = 0.
      DO 33 J = I,NTH
   33 TNBIN(I) = TNBIN(I) + PROB(I,J)
   32 CONTINUE
      WRITE(6,6)P,RK,PMU
    6 FORMAT(1H1,20X,3H P =,F8.4,10X,5H  RK -,F8.4,10X,6H  PMU -,F8.4)
      WRITE(6,7)
    7 FORMAT(1H0,20X,34H PROBABILITY OF I THE'S AND N TH'S)
      WRITE(6,8) (I,I = 1,10)
    8 FORMAT(1H0,5X,3H I =,10(9X,I2)))
      DO 9 NX = 1,NTH
    9 WRITE(6,10)NX,(PROB(I,NX),I = 1,NX)
   10 FORMAT(1H ,3H N =,12,9X,10(F8.4,3X)
      WRITE(6,11)
   11 FORMAT(1H0,20X,22H PROBABILITY OF N TH'S/)
      DO 12 NX = 1,NTH
   12 WRITE(6,13)NX,PROBM(NX)
   13 FORMAT(1H ,20X,3H N -,I2,5X,10H PROBM(N) =,F8.4)
      WRITE(6,14)
   14 FORMAT(1H0,20X,23H PROBABILITY OF I THE'S/)
      WRITE(6,15)PI
   15 FORMAT(1H ,20X,5H I - 0,5X,10H NGBIN(I) -,F8.4)
```

```
      DO 16 I = 1,NTH
   16 WRITE(6,17)I,TNBIN(I)
   17 FORMAT(1H ,20X,3H I -,I2,5X,10H NGBIN(I) -,F8.4)
      GO TO 1000
      STOP
      END
```

REFERENCES

[1] Falls, Lee W. (1969). "A probability distribution for the number of thunderstorm events at Cape Kennedy, Florida." NASA TM X-53816, Marshall Space Flight Center, Alabama.

[2] Williford, W. O., and Carter, M. C. (1969). "Probability models for the variation in the number of thunderstorm hits per day." NASA Contractor Report No. 61289, Marshall Space Flight Center, Alabama.

[3] Bailey, Norman T. J. (1964). The Elements of Stochastic Processes with Applications to the Natural Sciences. John Wiley and Sons, Inc. New York.

[4] Feller, W. (1962). An Introduction to Probability Theory and Its Applications, Vol. 1 (Second Edition, Seventh Printing). John Wiley and Sons, Inc., New York.

[5] Cohen, A. Clifford, Jr. (1965). "Estimation in the Negative Binomial Distribution". NASA TM X-53372, Marshall Space Flight Center, Alabama.

[6] Rider, Paul R. (1953). Truncated Poisson Distributions, Journal of the American Statistical Association 48, 826-831.

[7] David, F. N., and Johnson, N. L. (1952). The Truncated Poisson. Biometrics 8, 275-286.

MSFC—RSA, Ala