AA Technical Memorandum EDIS 29

GAMMA DISTRIBUTION SHAPE PARAMETER BIAS

Washington, D.C.
August 1980

U.S. DEPARTMENT OF COMMERCE / National Oceanic and Atmospheric Administration / Environmental Data and Information Service

TECHNICAL MEMORANDUMS

... and Information Service Series

... and ... on Service (EDIS) is responsible for storing, retrieving, and pub- ... by NOAA and for developing systems to process and present NOAA data in the most ... statistical form. These data, which relate to the solid earth, ocean, atmosphere, ... space, are the basic input in scientific and engineering studies having broad ... ure, commerce, defense, and industry.

... OAA Technical Memorandums facilitates rapid distribution of studies and reports ... y in nature and which may be published formally elsewhere at a later date. Publi- ... a former series of Weather Bureau Technical Notes (TN); publications 3 to 17 are in a former series of ESSA Technical Memorandums designated as Environmental Data Service Technical Memorandums (EDSTM). Beginning with 18, publications are now part of the NOAA Technical Memorandum EDIS series and are designated as either EDS or EDIS.

Publications listed below are available from the National Technical Information Service, U.S. Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, VA 22161. Price: Varies for paper copy; $3.50 microfiche. Order by accession number shown in parentheses at end of each entry.

Weather Bureau Technical Notes

TN 33 / EDS 1 — Part 1: Weather and Disease. H. E. Landsberg. Part 2: Solar Radiation and Skin Cancer Deaths. G. D. Brinckerhoff. February 1966. (PB-169-813)

TN 43 / EDS 2 — A Bibliography of Meso- and Micro-Environmental Instrumentation. John F. Griffiths and M. Joan Griffiths, July 1966. (PB-173-088)

ESSA Technical Memorandums

EDSTM 3 — Climatological Aspects of Air Pollution in West Virginia. Robert O. Weedfall, January 1967. (PB-176-905)

EDSTM 4 — Estimates of Winds at Constant Heights Generated From Winds Observed at Constant Pressure Surfaces. H. L. Crutcher, H. B. Harshbarger, R. L. Durham, F. T. Quinlan, J. D. Matthews, and W. B. Tschiffely, February 1967. (PB-174-499)

EDSTM 5 — Synoptic Aspects of Mortality: A Case Study. Dennis M. Driscoll, February 1967. (PB-174-500)

EDSTM 6 — A Technique for Determination of a Climatology of High Altitude Balloon Trajectories. Frank T. Quinlan and Lee R. Hoxit. October 1968. (PB-180-200)

EDSTM 7 — Minimum Temperature Patterns Over New York State on Two Winter Nights. A. Boyd Pack, February 1969. (PB-183-398)

EDSTM 8 — Coastal Weather and Marine Data Summary for Gulf of Alaska, Cape Spencer Westward to Kodiak Island. Harold W. Searby, May 1969. (PB-184-784)

EDSTM 9 — A Bibliography of Weather and Architecture. John F. Griffiths and M. Joan Griffiths, April 1969. (PB-184-969)

EDSTM 10 — A Note on Climatology of Thailand and Southeast Asia. H. L. Crutcher, S. Bintasant Nash, and D. K. Kropp, May 1969. (PB-187-569)

EDSTM 11 — Meteorological Drought in West Virginia. Robert O. Weedfall, September 1969. (PB-187-474)

EDSTM 12 — Snowfall, Snowfall Frequencies, and Snow Cover Data for New England. R. E. Lautzenheiser, December 1969. (PB-194-221)

EDSTM 13 — Applications of Air Mass Classifications to West Virginia Climatology. Robert O. Weedfall, January 1970. (PB-191-966)

EDSTM 14 — Georgia Tropical Cyclones and Their Effect on the State. Horace S. Carter, January 1970. (PB-190-768)

(Continued on inside back cover)

NOAA Technical Memorandum EDIS 29

GAMMA DISTRIBUTION SHAPE PARAMETER BIAS

Harold L. Crutcher
Formerly, Science Advisor
National Climatic Center
Asheville, N.C.

Raymond L. Joiner
Formerly, Digital Systems Advisor
National Climatic Center
Asheville, N.C.

Washington, D.C.
August 1980

UNITED STATES
DEPARTMENT OF COMMERCE
Philip M. Klutznick, Secretary

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION
Richard A. Frank, Administrator

Environmental Data and
Information Service
Thomas D. Potter, Director

ACKNOWLEDGMENTS

Acknowledgment is made to Mr. Bob Ford and his group for preparation of the figures, and to Mrs. Margaret Larabee and Mrs. Martha Lakenan for typing.

The senior author appreciates the many courtesies extended in discussions with Dr. Leon Harter and subsequent correspondence involving exchange of reprints.

The use of the McGill University "Super-Duper" Random Number Generator of Marsaglia et al. and the algorithms of Phillips and Beightler based on Johnck's rejection technique is acknowledged. The last is taken from the Journal of Statistical Computer Simulation and is used with permission of Gordon and Breach Science Publishers Ltd.

Appreciation is expressed to the staff at EDIS's Environmental Science Information Center, in particular to Mr. Patrick McHugh for the care given to editing this paper.

CONTENTS

FIGURES

maximum estimates of the shape parameter and the known shape parameter. For example, the number 2 on the ordinate represents 100 percent bias; i.e., the computed estimate must be divided by 2 to obtain the shape parameter to be used to compute quantities or probabilities. The abscissa is the shape parameter. The curves represent sample size.

Figure 3b Bias in the moment estimators of shape parameters in a gamma (Pearson Type III) distribution. The bias is a function of both the shape parameter and sample size. The ordinate values are the ratios of the average compiled maximum estimates of the shape parameter and the known shape parameter. For example, the number 2 on the ordinate represents 100 percent bias; i.e., the computed estimate must be divided by 2 to obtain the shape parameter to be used to compute quantities or probabilities. The abscissa is the sample size. The curves represent the shape parameter.

Figure 4a Gamma distribution density curve for 61 years of 24-hour flood discharge data (1,000's cu ft sec^{-1}) at Chattanooga, TN 1874-1934.

Figure 4b Gamma distribution observed frequencies versus expected frequencies for equal probability class intervals of 61 years of 24 hour flood discharge data (1,000's cu ft sec^{-1}) at Chattanooga, TN 1874-1934.

Figure 4c Gamma distribution cumulative frequency diagram for equal probability class intervals for 61 years of 24-hour flood discharge data (1,000's cu ft sec^{-1}) at Chattanooga, TN 1874-1934.

Figure 4d Gamma distribution line of best bit with confidence bands and a data overplot for 61 years of 24-hour flood discharge data (1,000's cu ft sec^{-1}) at Chattanooga, TN 1874-1934.

TABLES

TABLES

Symbols used in this report

e	exponential, also exp
cu	cubic
f	function
ft	foot (feet)
g	function, skew factor
g_1	skew factor estimate
g_2	kurtosis factor estimate
i	index
j	index
k	real number, k-statistic or cumulant
k´	unbiased cumulant
ln	natural logarithm
log	common logarithm
m	mean
mse	mean square error
n	index, number in sample
p	probability, precipitation
r	index, moment
s	sample standard deviation
s^2	sample variance
sec	second
t	variable
dt	derivative of t
u	random number, $0 < u \leq 1$
x	variable, used as a subscript
x´	transformed variable
y	variable, uniform random number, also used as a subscript
z	variable
A	transformed variable, shape parameter transformation ratio of arithmetic mean to geometric mean; random number
B	random number
D	deviate
E	expected value

F	function
G	function, jackknife
$G(\hat{\gamma})$	jackknife estimator
K-S	Kolmogorov-Smirnov
M	moment
ME	moment estimator
ML	maximum likelihood
MLE	maximum likelihood estimator
Q	negative natural logarithm of a random number
X	negative product of logarithms
α	alpha, origin
β	beta, scaling parameter
γ	gamma, shape parameter
$\hat{\gamma}$	Maximum Likelihood Estimator (MLE)
$\hat{\gamma}'$	debiased MLE
$\overset{*}{\gamma}$	Thom estimator
$\overset{*}{\gamma}'$	debiased Thom estimator
$\tilde{\gamma}$	Tilde estimator
μ	moment
μ_i	indexed moment
ν	degrees of freedom
π	product
$\int$	integral
Σ	summation
τ	quantile

$d\tau$	derivative of τ
Γ	gamma
$\Gamma(\ )$	gamma function
χ^2	chi-square
$<$	less than
$\leq$	equal to or less than
$=$	equal
$>$	greater than
$\geq$	equal to or greater than
$\overline{\ \ }$	overbar, indicates averaging process
$\simeq$	approximately
$\tilde{}$ ($\sim$)	as overscript, tilde estimator
∞	infinity, indeterminant
*	footnote reference
©	copyright symbol
$-$	minus
$+$	plus
/	solidus implying devision
[]	truncated integer
!	factorial
Λ	maximum likelihood indicator

ABSTRACT

The generalized gamma distribution has three parameters, the origin, the scale, and the shape. As a model, it serves to approximate univariate distributions in many disciplines. To be an acceptably accurate model, the model parameters estimates should contain no bias or should contain as little a bias as possible. To be a more useful model, the model should show the confidence in the fit it provides to a data set.

Rarely is the bias removed in actual practice. Only recently have procedures to remove bias been developed. The two most widely used estimators, the moments and the maximum likelihood estimators, exhibit a bias which is a function of the sample size and of the shape parameters. The Thom estimator, which is an approximation to the maximum likelihood estimator, should and does exhibit the same bias. The origin is assumed to be fixed. Only two parameters are thus involved, the shape and the scale.

A recent paper by the authors treats the subject of this paper for shape parameters equal to or greater than one. This paper re-examines the region from a shape parameter of 5 down to 1 and extends the removal of bias procedures down to shape parameters of two-hundredths, 0.02, with a sample size 2 or greater.

GAMMA DISTRIBUTION SHAPE PARAMETER BIAS

Harold L. Crutcher and Raymond L. Joiner

I. INTRODUCTION

The gamma distribution function model is an extremely useful one. It serves many disciplines. Among the many applications are those often made to life tests, other quality control and assurance problems, reliability, maintenance, and precipitation. Wherever a distribution is bounded at one end and unlimited at the other, the gamma distribution is one candidate among many that might be used to describe that distribution.

The general gamma distribution with origin parameter $\alpha(-\infty < \alpha < +\infty)$, scale parameter $\beta(\beta > 0)$, and shape parameter $\gamma(\gamma > 0)$ has the probability density function shown in

$$f(y:\alpha,\beta,\gamma) = \beta^{-\gamma}(\Gamma(\gamma))^{-1}(y-\alpha)^{\gamma-1}e^{-(y-\alpha)/\beta};\ y > \alpha;\ -\infty < \alpha, y < +\infty$$

and $f(y:\alpha,\beta,\gamma) = 0,\ y \leq \alpha.$ (1)

The distribution function given by

$$F(y;\alpha,\beta,\gamma) = \int_{\alpha}^{y} f(t;\alpha,\beta,\gamma)dt \qquad (2)$$

is for all $y > \alpha$.

With the origin α obtained, the following expression is pertinent:

$$x = y - \alpha. \qquad (3)$$

Then (1) becomes

$$f(x;0,\beta,\gamma) = \beta^{-\gamma}(\Gamma(\gamma))^{-1}\ x^{\gamma-1}e^{-x/\beta},\quad 0 < x < \infty$$

and $f(x;0,\beta,\gamma) = 0, \quad x \leq 0.$ (4)

Thom (1968 and in his earlier papers) utilizes this form. As shown by Thom (1958) and by Wilk et al. (1962), if the variate x assumes a transform by division of the scale parameter β, the distribution function develops as

$$F(x';0,1,\gamma) = (\Gamma(\gamma))^{-1} \int_{0}^{x'} \tau^{\gamma-1}\ e^{-\tau}\ d\tau, \qquad x' > 0$$

and $F(x';0,1,\gamma) = 0, \quad x' \leq 0,$ (5)

that is a standard form with $\alpha = 0$ and $\beta = 1$ and is positive when $x > 0$.

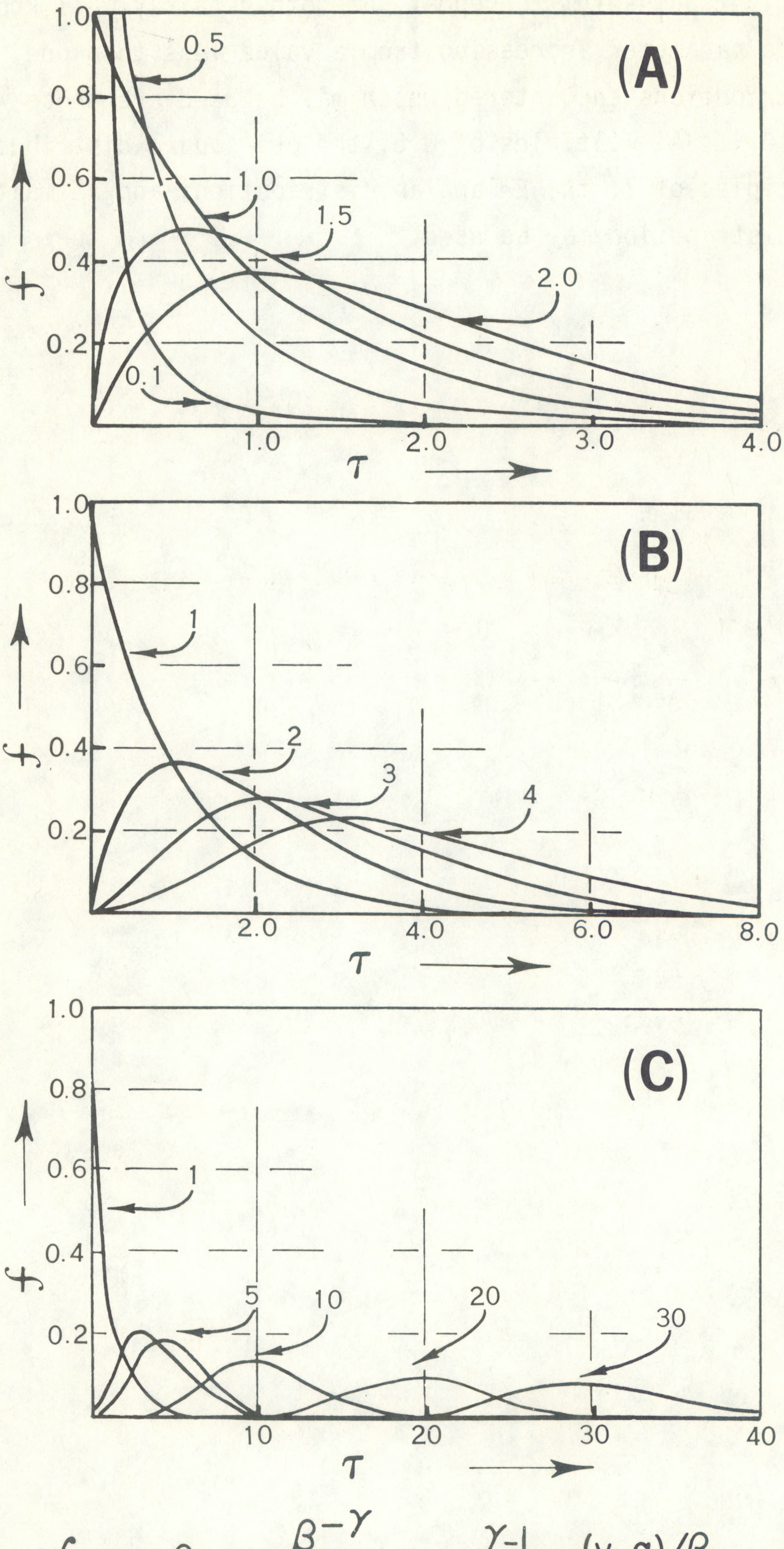
Figure 1 represents equation 5, the standard case, where the shape of the curve depends only on the parameter γ, the shape parameter.

There are three parameters of the gamma distribution to be estimated. If the origin is zero, is known, or arbitrarily assigned, the problem further resolves down to the two parameters, the scale and the shape. Presumably, the bias in the estimators will have been removed so that only the imprecision of the estimates need be considered. This imprecision is, of course, the basis for the need for visual and numerical assessment of just how good the estimated cumulative distribution is.

The mean and the shape parameter are independent but this is not the case with the scale and shape parameters. As Mann et al. (1974) point out, no general method has been developed to obtain confidence intervals for the scale parameter.

Fisher (1922) first develops the maximum likelihood (ML) equation for the solution for the incomplete gamma distribution known commonly as the gamma distribution. It is incomplete in the sense that the integral limits of the function do not range from $-\infty$ to $+\infty$ but from some finite point such as α to $+k$ where k is some real number. If the origin parameter α is zero, this distribution is a special case of the Pearson Type III distribution. The solution of the ML equation as developed by Fisher is difficult. Therefore, Thom (1947) expanding on work by Norlund (1924) develops approximate solutions. Chapman (1956), Greenwood and Durand (1960), Gupta (1960), Harter and Moore (1965), and Wilk et al. (1962) also provide methods to estimate the gamma distribution parameters. There are other methods used to estimate the distribution parameters but only the moments (ME) and the Thom estimators will be studied in any detail in this paper. Others briefly mentioned are the jackknife and tilde estimators of Anderson and Roy (1975).

With respect to moment estimators, Fisher (1922) showed that the moment method of estimating may be inefficient. Kendall and Stuart (1963) indicate that the efficiency may be as low as 22 percent. Thom (1958) gave values of the efficiency which went as low as 50 percent for the scale parameter β and 40 percent for the shape parameter γ.



$$f(y;\alpha,\beta,\gamma) = \frac{\beta^{-\gamma}}{\Gamma(\gamma)}(y-\alpha)^{\gamma-1}\, e^{-(y-\alpha)/\beta}$$

FIGURE 1. SELECTED GAMMA DISTRIBUTION FUNCTION CURVES.

Again, it might be useful to repeat some other fairly well known items. With the shape parameter increasing from a value less than one, some of the specific distributions encountered which may be used are those which are multiples of 0.5. At multiples of 0.5, the chi-square distribution may be used; at multiples of 1, the Erlangian distribution; and at multiples of 2, the Poisson distribution may be used.

II. RANDOM NUMBER GENERATORS

For integer values of the shape parameter, a form, the usual formula used is

$$y = \Sigma_{i=1}^{\gamma} (-\beta \ln y_i) \tag{6}$$

where y_i denotes a random number from the uniform distribution with a range (0, 1). The usual procedure is to call up such a generator from some statistical package within the computer. It has been the experience of the senior author that most of these are deficient in some respect or another. All so-called random number generators are actually pseudo-random number generators operating over certain sub-spaces. Many have truncation errors. However, they generally provide usable and acceptable output.

Bowman and Beauchamp (1975) discuss "Pitfalls With Some Gamma Variate Simulation Routines." This is an excellent paper and should be reviewed by anyone working with the gamma distribution and its applications. Sowey (1972) provides a chronological and classified bibliography on random number generation and testing.

The McGill University "Super-Duper" Random Number Generator of Marsaglia et al. (1972) for integer values of the shape parameter was used by Crutcher and Joiner (1978). The output is, from our prejudiced view, somewhat better. Even with this, there appears to be an insufficient number of large random variates for the fractional shape parameter values.

Phillips and Beightler (1972) discuss these random number generators and propose one to provide gamma distributed variates for shape parameters greater than one-tenth (0.1). The algorithm they propose, based in part on Johnck's (1964) rejection techniques, and the one which is used in this paper where the input is obtained from the "Super-Duper" random number generator follows.

"Define i) γ is a non-integer shape parameter

ii) $\gamma_1 = [\gamma]$ is the truncated integer of γ

iii) u_i is the ith random number $1 \geq u_i \geq 0$.

"1. Let $X = -\ln \prod_{i=1}^{\gamma_1} u_i$ (7)

"2. Let $A = \gamma - \gamma_1$; $B = 1 - A$ (8)

a) Set $j = 1$

b) Generate a random number, u_j, and set $y_1 = (u_j)^{1/A}$ (9)

c) Generate a random number, u_{j+1}, and set $y_2 = (u_{j+1})^{1/B}$ (10)

d) If $y_1 + y_2 \leq 1$, go to (f) (11)

e) Set $j = j + 2$; go to (b) (12)

f) Let $z = y_1/(y_1+y_2)$ so that z is a beta variable with parameters A and B. (13)

"3. Generate a random number, u_n, and let $Q = -\ln (u_n)$ (14)

"4. The needed deviate is $D = (x + zQ)\beta$" (15)

Once a random number has been obtained, it is used in sets of numbers where the moments and maximum likelihood estimators are obtained. Reference may be made to Crutcher and Joiner(1978). A considerable number of attempts to obtain a random number may be rejected. The procedure has been extended in this paper to obtain random numbers from distributions with shape parameters as small as 0.02. The procedure simply was not tried for yet smaller shape parameter distributions.

III. DEVELOPMENT FOR MLE AND ME BIAS CORRECTION

A. Development

The objectives of this work are to (1) substantiate the work of previous investigators, (2) obtain more information on the variation of parameter estimates, and (3) develop procedures to remove the bias in estimation of the parameters of the gamma distribution for shape parameter values 0.02 or greater with special interest in the region $0.02 \leq \gamma \leq 1.0$.

With the exception of the random number generator previously discussed, the procedures, format, and output are the same as given by Crutcher and Joiner (1978). The user is referred to that paper for more detail. The appendix contains the computer program and appropriate algorithms where available. It is stressed here that the moments estimates computed are based on n.

For all moments above the first, this induces a bias. This bias is a part of the bias discussed in previous papers. This will be discussed in more detail in the next section.

Tables 1 and 2 provide a relatively smooth output of information on the bias in gamma distribution shape parameter estimates. These are for the bias in the maximum likelihood and moments estimates. First, initial tables were prepared from the computer output. Then graphs of shape parameters versus the bias by sample size and sample size versus the bias by shape parameter were prepared. The bias is represented by the ratio of the computed estimate to the known shape parameter. Two successive smoothings amongst the tables and respective graphs were made. The tables and figures shown here represent the smoothed field features of the relationships of known shape parameter, computed shape parameter estimate, and sample size. The largest anomaly in the bias removed in the smoothing process was five percent for sample size 4.

γ ↓ / SAMPLE SIZE →	100	50	30	20	10	8	7	5	4	3	2
5.0	1.030	1.055	1.105	1.155	1.365	1.565	1.710	2.320	3.595	12.570	10.000
4.0	1.035	1.060	1.110	1.160	1.368	1.570	1.710	2.330	3.585		
3.0	1.036	1.065	1.110	1.165	1.370	1.565	1.695	2.335	3.570	13.117	
2.0	1.039	1.070	1.110	1.170	1.370	1.540	1.680	2.315	3.530	13.270	
1.0	1.042	1.060	1.110	1.160	1.360	1.500	1.640	2.250	3.420	11.400	
0.9	1.045	1.060	1.096	1.150	1.350	1.505	1.630	2.230	3.400	11.000	
0.8	1.048	1.076	1.095	1.145	1.335	1.487	1.625	2.200	3.380	10.750	
0.7	1.051	1.072	1.096	1.150	1.320	1.482	1.618	2.160	3.350	10.400	
0.6	1.055	1.080	1.100	1.155	1.310	1.460	1.600	2.100	3.330	10.000	
0.5	1.070	1.090	1.105	1.160	1.315	1.440	1.580	2.080	3.250	9.500	
0.4	1.085	1.105	1.120	1.175	1.330	1.430	1.540	1.970	3.050	8.900	
0.3	1.125	1.140	1.160	1.210	1.360	1.440	1.540	1.910	2.885	8.200	
0.2	1.185	1.205	1.225	1.260	1.415	1.480	1.580	1.870	2.590	7.200	
0.1	1.385	1.395	1.415	1.440	1.570	1.642	1.720	1.983	2.575	5.500	
0.05	1.715	1.731					2.050	2.262	2.653	4.428	
0.04								2.457	2.816	4.000	
0.03				(NOT COMPUTED)				2.738	3.148	5.160	
0.02								3.334	5.777	3.697	

TABLE 1 BIAS IN THE MAXIMUM LIKELIHOOD ESTIMATES OF SHAPE PARAMETERS IN A GAMMA (PEARSON TYPE III) DISTRIBUTION. 1.000 REPRESENTS NO BIAS. 1.030 REPRESENTS THREE PERCENT BIAS. 2.000 REPRESENTS 100 PERCENT BIAS.

γ \ SAMPLE SIZE →	100	50	30	20	10	8	7	5	4	3	2
5.0	1.035	1.066	1.112	1.193	1.430	1.631	1.790	2.398	3.639	> 10	>10.000
4.0	1.038	1.075	1.117	1.210	1.432	1.640	1.796	2.427			
3.0	1.041	1.086	1.126	1.224	1.467	1.649	1.798	2.450	3.660		
2.0	1.044	1.094	1.136	1.236	1.488	1.669	1.824	2.491	3.680		
1.0	1.050	1.107	1.160	1.255	1.589	1.772	1.908	2.625	3.730		
0.9	1.053	1.110	1.166	1.270	1.600	1.775	1.927	2.650	3.740		
0.8	1.056	1.119	1.176	1.285	1.610	1.826	1.976	2.670	3.760		
0.7	1.060	1.122	1.186	1.305	1.633	1.866	2.001	2.700	3.780		
0.6	1.064	1.135	1.203	1.335	1.660	1.876	2.006	2.725	3.808		
0.5	1.078	1.157	1.228	1.375	1.741	1.925	2.034	2.750	3.900		
0.4	1.099	1.180	1.259	1.410	1.811	2.022	2.200	2.903	4.000		
0.3	1.112	1.210	1.310	1.480	1.928	2.164	2.370	3.072	4.200		
0.2	1.150	1.271	1.413	1.609	2.144	2.453	2.681	3.508	4.581		
0.10	1.239	1.425	1.646	1.925	2.781	3.230	3.600	4.703	6.225		
0.05	1.403	1.702					5.280	7.241	9.319		
0.04								8.569	11.058		
0.03			(NOT COMPUTED)					10.677	13.837		
0.02								14.922	19.481		
0.01											

TABLE 2 BIAS IN MOMENTS ESTIMATES OF SHAPE PARAMETERS IN A GAMMA (PEARSON TYPE III) DISTRIBUTION. 1.000 REPRESENTS NO BIAS. 1.035 REPRESENTS 3.5 PERCENT BIAS. 2.000 REPRESENTS 100 PERCENT.

Figure 2a, pertaining to maximum likelihood estimators, is a graph of (i) the known shape parameter plotted against (ii) the computed shape parameter estimates divided by the known shape parameter for various sample size. Figure 2b also pertains to maximum likelihood estimators and is a graph of sample size plotted against the computed shape parameter estimate divided by the known shape parameter for various shape parameters. Thus, the family of lines are marked in Figures 2a and 2b respectively by the sample sizes and by known shape parameter. The division by the known shape parameter is a normalizing feature. Figures 3a and 3b present similar information as Figures 2a and 2b do, but the estimators are the moment estimators.

In the previous paper (Crutcher and Joiner, 1978) there was some indication that bias in the maximum likelihood and moments estimators is a function of the shape parameter in addition to being a function of the sample size. This was one of the reasons why this work on bias for small samples and for small shape parameters is continued. This report represents the progress to this point.

The algorithms given by Crutcher and Joiner (1978) still serve reasonably well where sample sizes are 10 or greater and the shape parameter is one or greater.

Table 1 and Figures 2a and 2b quite clearly show that the region for shape parameters below one do not follow the algorithms developed for shape parameters greater than 1, Crutcher and Joiner (1978). Some apparent anomalies exist in this region. The expected correction, given more sampling, is indicated by the dashed lines. Remember that the solid lines are smoothed lines derived from initial tabular data.

The slight swell of the curves for shape parameters 1 to 5 may be the result of insufficient sampling. As it exists in all samples made in the moments as well as the maximum likelihood estimators, it is retained.

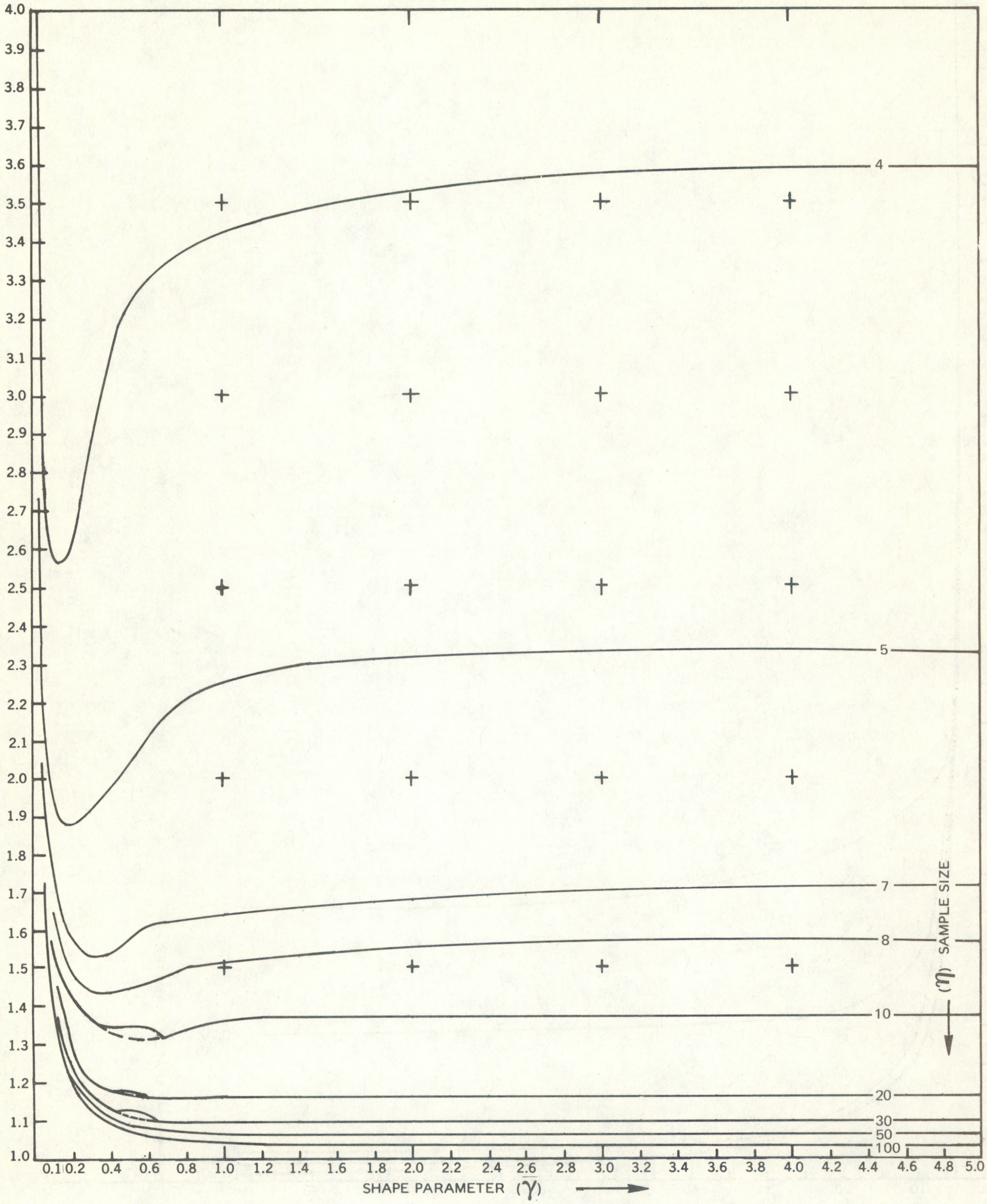


Figure 2a Bias in the maximum likelihood estimators (Thom's approximations) of shape parameters in a gamma (Pearson Type III) distribution. The bias is a function of both the shape parameter and the sample size. The ordinate values are the ratios of the average compiled maximum likelihood estimates of the shape parameter and the known shape parameter. For example, the number 2 on the ordinate represents 100 percent bias; i.e. the computed estimate must be divided by 2 to obtain the shape parameter to be used to compute quantiles or probabilities.

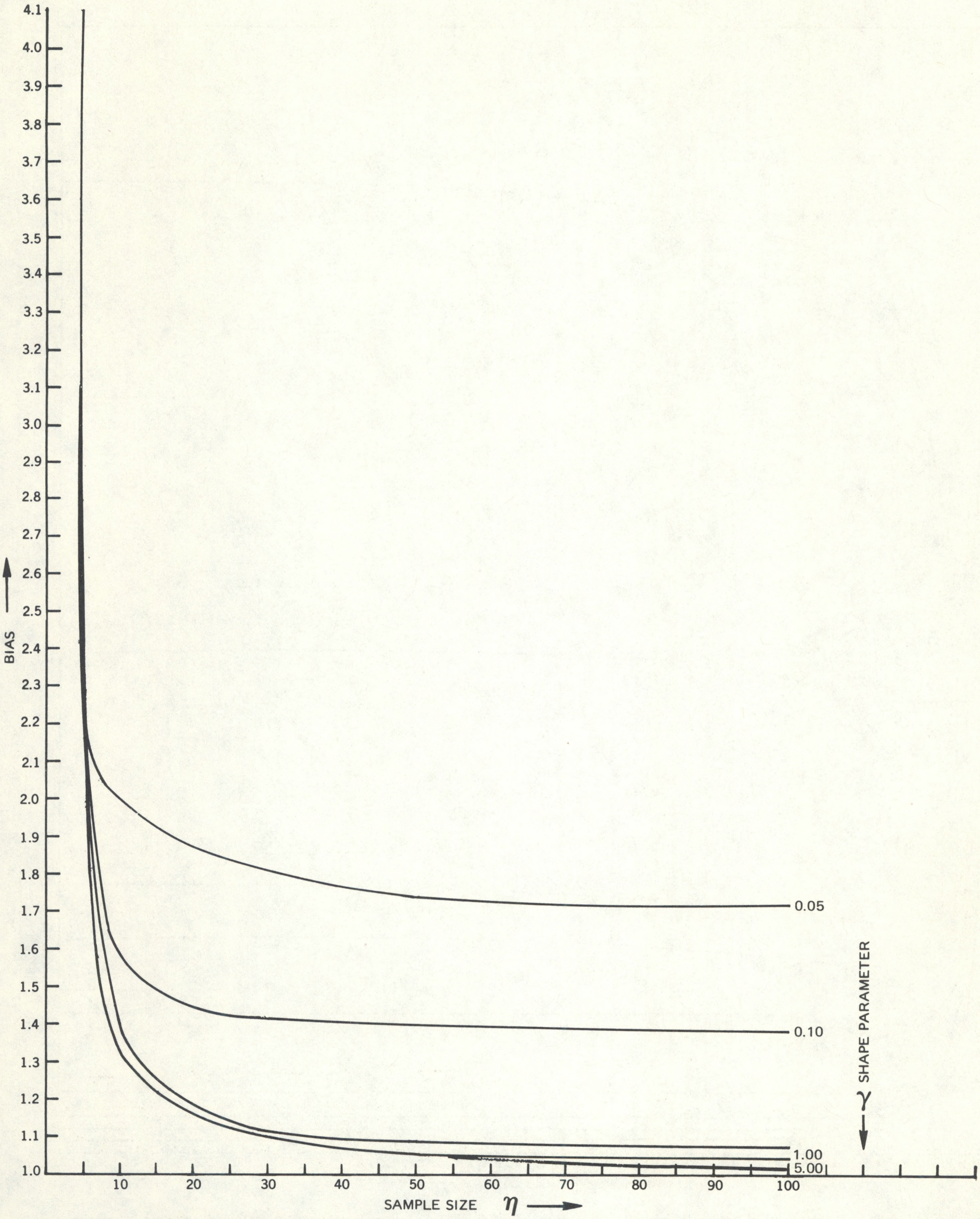


Figure 2b Bias in the maximum likelihood estimations (Thom's approximation) of shape parameters in a gamma (Pearson Type III) distribution. The bias is a function of both the shape parameter and the sample size. The ordinate values are the ratios. The bias is a function of both the shape parameter and the sample size. The ordinate values are the ratio of the average compiled maximum likelihood estimates of the shape parameter and the known shape parameter. For example, the number 2 on the ordinate represents 100 per cent bias, i.e. the compiled estimates must be divided by 2 to obtain the shape parameter to be used to compute quantiles or probabilities.

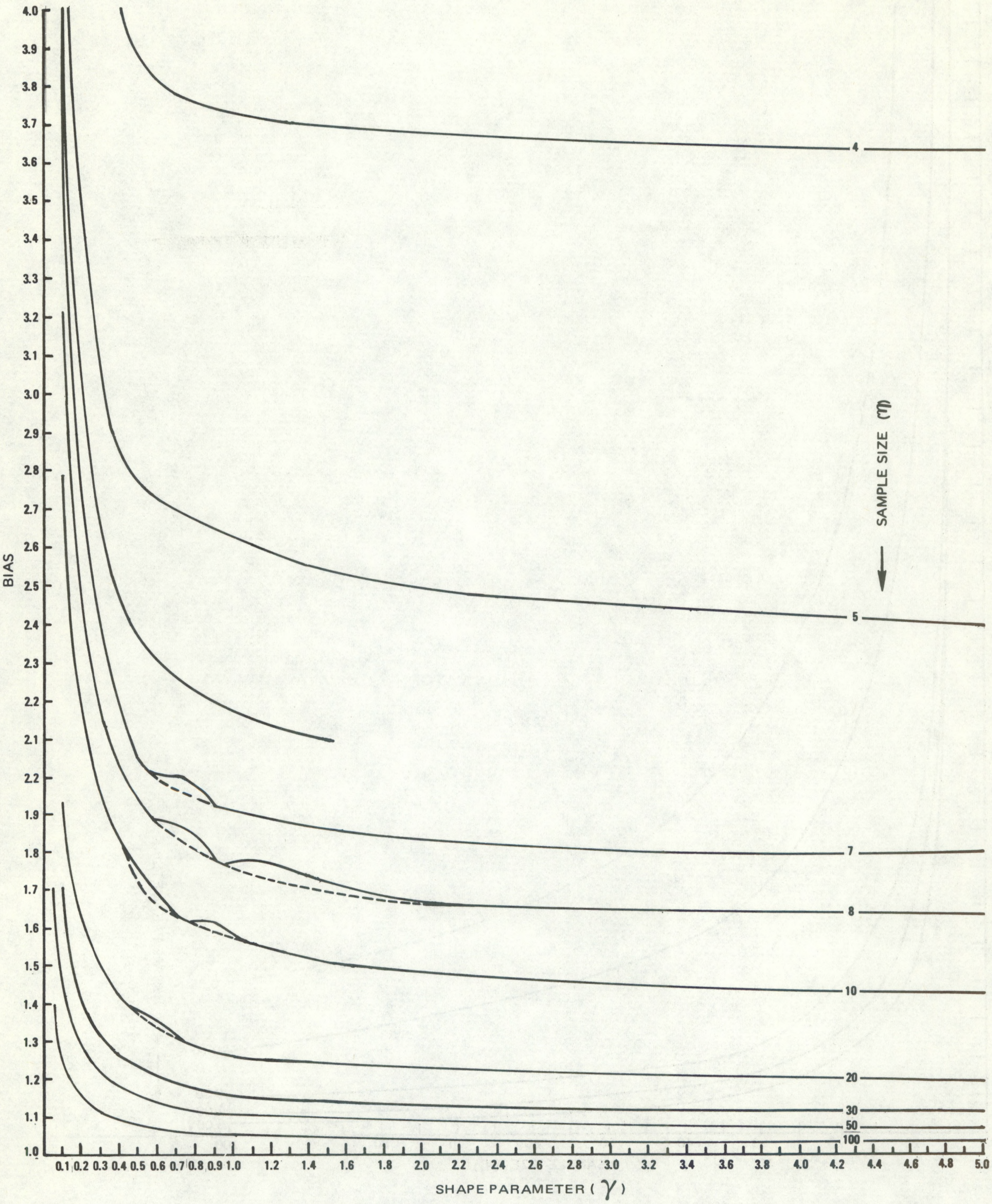


Figure 3a Bias in the moments estimates of shape parameters in a gamma (Pearson Type 111) distrbution. The bias is a function of both the shape parameter and the sample size. The ordinate values as the ratio of the average computed moments estimates of the shape parameter and the known shape parameter. For example the number 2 on the ordinates represents 100 percent bias; i.e. the compiled estimates must be divided by 2 to obtain the shape parameter to be used to compute quantiles or probabilities.

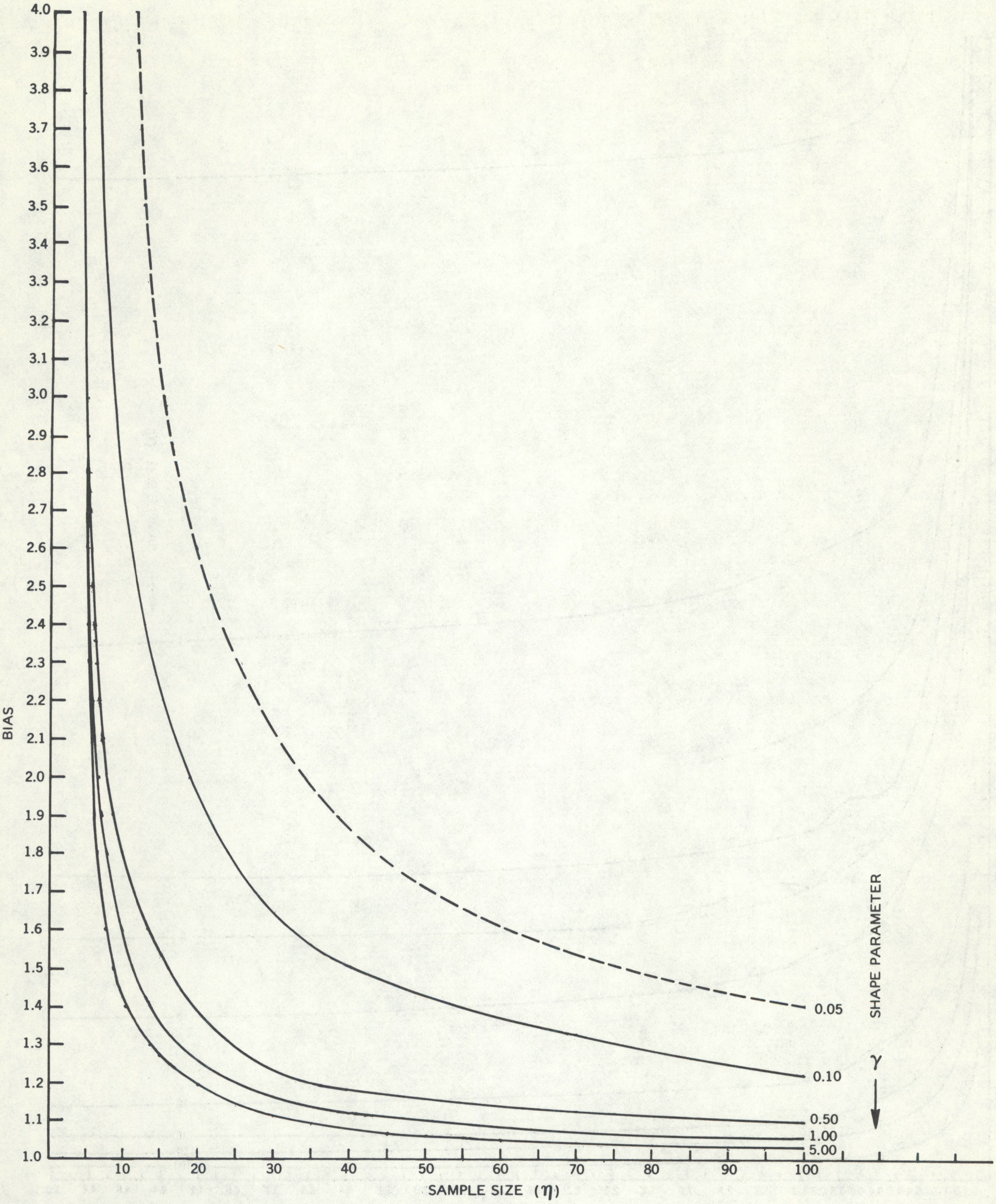


Figure 3b Bias in the moment estimates of shape parameters in a gamma (Pearson Type III) distribution. The bias is a function of both the shape parameter and the sample size. The ordinate values are the ratio of the average computed moments estimates of the shape parameter and the known shape parameter. For example, the number 2 on the ordinate represents 100 percent bias, i.e. the computed estimate must be divided by 2 to obtain the shape parameter to be used to compute quantiles or probabilities.

The sharp dip in the curves appears to be real for decreasing shape parameters until a minimum is reached. This minimum is a function of both sample size and shape parameter. The minimum occurs at a lower shape parameter value with decreasing sample size. The increase of bias is rapid once the minimum has been reached. For sample size 2 the bias is well over 10,000 percent and rather unstable due to limited sampling. Limited here means economic and time requirements do not permit the 1000 or more samples that would have to be made in this case. The minimum is most evident in the data for sample size 3 at a shape parameter of about 0.04 shown in table 1. A graph of this is not made. Neither is a graph for sample size 2 shown, as insufficient sampling was done as indicated above. Difficulties in this region were expected.

Figure 2b clearly shows the bias as a function of shape parameter less than 1. The curves for the various shape parameters cross and the lines become confused at sample size below 10. The figure shows that the algorithms (Crutcher and Joiner, 1978) are valid for shape parameters greater than 1.

Because of time limitations, no functional algorithm is suggested for bias correction at this time. Given a sample size, $(\exp(1/(n^{(n-3)/2}))$ and computed estimate, a bias correction as a divisor may be obtained by look-up of appropriate table 1 and table 2 data which can be contained in a program. Linear interpolation may be used for the shape data while the exponential expression, $(\exp(1/n^{(n-3)/2}))$ may be computed for a sample size between the sample sizes contained in the tables, figures, and/or look-up tables.

B. Errors in Probability Levels and/or in Quantiles

The usual computation of shape parameters in a gamma distribution provides biased estimates. (The usual debiasing of moments estimators is acceptable for large samples but is certainly not acceptable for small samples or for small shape parameters. Small shape parameters imply large skew. The use of these biased estimates produces errors in the probability levels and in quantiles. Table 3 provides an insight into the extent of these errors. These are given simply as functions of tabular data given by many authors. Two tabular presentations are Roy, Gnanadesikan, and Srivastava (1971) and Crutcher, McKay, and Fulbright (1978). Errors are discussed briefly by Bridges and Haan (1972).

A proposed table of shape parameters and sample numbers with biases and algorithms to do interpolation has not yet been prepared. Table 3 is prepared with data selected from presently available tables (Roy, Gnanadesikan, and Srivastava, 1971, and Crutcher, McKay, and Fulbright, 1978). As a result, the sample size and biases may not be the same from example to example.

Table 3 tabular data are prepared by the following steps.

(1) Take a particular gamma shape parameter, say $\gamma = 1$

(2a) Take a particular maximum likelihood estimate sample size, say, $n = 10$, and

(3a) Obtain from table 1 the corresponding bias of 1.36 (or 1.36 percent);

(2b) Or take a particular bias, say, 1.33 obtained by use of $(n + 2)/(n - 1)$ as in the top half of Table 3.

(3b) Obtain from table 1 the approximate corresponding sample size, say, $n = 10$.

(4) Divide the gamma shape parameter estimate by the bias. This will give a new shape parameter unbiased.

(5) For a set of probability levels with biased quantiles, select the 0.10, 0.20, 0.50, 0.70, 0.90, and 0.99 probability level quantiles, for this table. Other levels could have been selected.

(6) For a set of probability levels with unbiased quantiles, select quantiles for the same probability levels as in (5) above.

(7) Multiply each of the unbiased quantiles by the bias value 1.33. This is necessary as the tabular values used are for a scale parameter of one. The debiasing of the shape parameter by division necessarily must increase the scale parameter by multiplication by the same factor.

(8) The difference ((5) - (7)) divided by (7) will give the error in terms of the quantiles. Multiplication by 100 will provide

the percentage error. The data provided in table 3 are approximate. These values were obtained using tabular data (Roy, Gnanadesikan, and Srivastava, 1971, and Crutcher, McKay, and Fulbright, 1978) for the nearest available shape parameter. Future work can provide appropriately computed data through the use of algorithms given by Crutcher, McKay, and Fulbright (1978).

Inferences from examination of the information contained in table 3 indicate that the percent errors for sample sizes less than 100 are near zero only at some narrow probability interval between 0.90 and 0.99 for shape parameters 0.05, 0.10, between 0.70 and 0.90 for shape parameter 0.50. For shape parameter 1, minimum percent error occurs between the 0.70 and 0.90 probability levels. For shape parameters 2 and 3, minimum percent error occurs between the 0.60 and 0.70 probability levels. For shape parameter 10 the minimum error occurs between the 0.50 and 0.60 probability levels. These intervals are narrow in each case for a plus or minus 5 percent error. Outside these intervals the percent error increases rapidly.

Essentially, table 3 indicates that only with large sample sizes, i.e., greater than 100, and large shape parameters greater than 10, can the usual computation of probabilities or quantiles (amounts) be used with any degree of confidence. The inference drawn is that unbiased estimates of the shape parameter should be used at all times.

Table 3.---A few selected approximate errors in quantile induced by gamma distribution shape parameter maximum likelihood estimate bias as a function of probability level. To change to percent error multiply by 100. A plus indicates a positive error or overstatement. A negative indicates a negative error or understatement.

γ	1.		2.			3.		10.	
N	10.	50.	4.	8.	13.	10.	50.	10.	50.
Bias	1.333	1.050	2.000	1.429	1.250	1.333	1.050	1.333	1.050
$\hat{\gamma}^*$	1.267	0.788	4.000	2.000	2.000	3.000	3.031	13.094	7.875
$\hat{\gamma}^{*\prime}$	0.951	0.750	2.000	1.400	1.600	2.250	2.857	9.821	7.500
Prob.									
0.1	--	0.098	≥1.523	0.491	0.263	0.243	0.034	0.092	0.014
0.2	0.503	0.108	≥0.848	0.304	0.169	0.157	0.023	0.061	0.010
0.5	0.144	0.021	≥0.211	0.083	0.047	0.041	0.006	0.011	0.002
0.6	0.082	0.012	≥0.104	0.039	0.022	0.015	0.002	-0.001	0.000
0.7	0.030	0.004	≥0.013	-0.001	-0.002	-0.009	-0.002	-0.014	-0.002
0.9	-0.068	-0.005	≥-0.155	-0.052	-0.086	-0.062	-0.011	-0.045	-0.007
0.99	-0.137	-0.024	≥-0.279	-0.096	-0.150	-0.112	-0.019	-0.078	-0.073

γ	γ=0.05		0.10		0.50	
N	N=5.	100.	5.	30.	8.	30.
Bias	2.262	1.715	1.983	1.418	2.080	1.105
$\hat{\gamma}^*$	0.045	0.051	0.099	0.099	0.728	0.497
$\hat{\gamma}^{*\prime}$	≈0.02	≈0.03	≈0.05	≈0.07	≈0.35	≈0.45
Prob.						
0.1	0.408E 19	0.127E 12	0.524E 10	0.139E 05	0.448E 01	0.539E 00
0.2	0.381E 13	0.129E 10	0.509E 08	0.714E 04	0.206E 01	0.326E 00
0.3	0.449E 09	0.613E 04	0.534E 03	0.139E 02	0.467E 00	0.990E-01
0.6	0.141E 07	0.539E 03	0.349E 03	0.546E 01	0.278E 00	0.600E-01
0.7	0.182E 05	0.686E 02	0.178E 02	0.237E 01	0.140E 00	0.289E-01
0.9	0.934E 01	0.155E 00	0.735E 00	0.242E 00	-0.730E-01	-0.266E-01
0.99	0.992E 00	-0.180E 00	-0.263E 00	-0.149E 00	-0.187E 00	-0.603E-01

C. ME Involving Skew and Kurtosis Parameters

1. General

One of the peculiarities of scientific research is that sometimes it is inhibited by an extremely parochial and limited discipline. There may be little or no crossover or cross reference to even such closely related disciplines of climatology (meteorology) and hydrology. This parochial view may be displayed here in this paper also for it does not present the ideas of other environmental groups, geophysical groups, astronomical groups, and the multitude of other groups. However, the stochastic, statistical, and mathematical treatments may be the same though the languages, the idiomatic developments, and the aliases of many parameters confuse and obfuscate the issue and the understanding.

All models should at least satisfy the physical restraints and constraints of the physical world of which they are used to portray certain features of interest. Descriptive and not causal features are generally the best that can be modeled. The researcher is at a real disadvantage to determine the genesis of the information or its genetic control. This paper does not attempt to determine information genesis.

Perhaps this paper will do a bit to show the relationship being used in climatology and hydrology to reach a consensus for occurrences of phenomena.

Pearson (1899, 1916) developed his well-known families of distributions based on moments considerations. Elderton (1953), Elderton and Johnson (1969) discuss these. The discussion in this section concerns one of these families, namely, the "Pearson Type III Distribution" known also as the "Gamma Distribution." Of the many variants of the Pearson Type III which are extremely useful, only the generalized and two parameters and the log gamma are discussed here. Elderton's work was originally published in 1906.

The gamma distribution and its application to hydrology are discussed in this section. Either original data measurements or transformation of the original data may be used. The transformation more specifically for use here is the logarithmic transformation. That is, logarithms of the original data are fitted by the gamma distribution. The model is then known as the log gamma or the log Pearson.

Forerunners to this study have been EDS-11, EDS-24, EDIS-30, and NOAA Technical Memorandum EDIS-28. The last is in the final stages of completion. It contains the electronic computer programs for EDIS-30 and is being adapted to a new computer system at the National Climatic Center, the Univac 1100. In those studies only the first two moments were used. The three parameters were the origin, scale, and shape parameters estimated by means of the first two moments. The present paper extends the treatment to all four moments, measures of the mean, variance, skew, and kurtosis.

The "Guidelines for Determining Flood Flow Frequency" (1977) contains discussions from the practical as well as theoretical points of view. This is Bulletin No. 17A of the Hydrology Committee of the United States Water Resources Council. The recommended model is the Log Pearson (Log Pearson Type III or the Log Gamma).

Moments computations performed in EDS-11, EDS-24, EDIS-30, and NOAA Technical Memorandum EDIS-28 involved the use of n. This was done for computational reasons where future comparisons or combinations would be made. All moments beyond the first are and were known to be biased. However, the estimated moments are biased beyond the above bias. Comparison of these biases will be made in this paper. Within the limits of data generally used in hydrology, the usual bias reduction will be sufficient for large samples. Here, as the model is examined from a more general point of view, the biases will be examined from the view of their dependence on the moments themselves and on the sample size. These comparisons will be made through the use of the shape parameter, the skew, and the kurtosis. This will involve the very close and natural relation among the various statistical parameters often used. Some of these are the moments themselves, the skew, the kurtosis, the scale, the shape, and the coefficient of variations and its reciprocal.

It is well-known that the distribution of the squares of a normally distributed variate is Rayleighian or Maxwellian. This produces a positively skewed distribution. Thus, if the square roots of the variates of skewed distributions are obtained, the distribution will tend to approach a normal distribution shape though it is zero bound. The logarithms of the variates of such a positively skewed distribution will also tend to approach the normal distribution shape. The resulting distribution will not be zero bound. From this point of view it is a better candidate to transform positively skewed distribution data to the more symmetrical model, the normal distribution. However, some distributions which are normal or negatively skewed will, upon logarithmic transformation, become negatively skewed or even more negatively skewed. Therefore, the logarithmic transformation should not be used blindly by the investigator.

If the data are negatively skewed to begin with, various other transformation may be used such as data reciprocals or distances to some arbitrary origin somewhat greater than the maximum. The net effect is to work with a mirror image transformation, do the necessary calculations and then transform back to the original variate.

A historical review of the Pearson Type III gamma model contribution to hydrology may be in order. A review of the many papers in the field would provide the same and more information. One of the peculiarities of research in any one field is that it becomes extremely parochial. A language is developed. Research builds along one line. Often it does not even cross over into other fields of research where the same model, variants of the model, and even other acceptable models are being developed. The present paper may be guilty of the same narrow limited parochial view. However, the discussion may tend to shed some light on present developments.

2. Applications to Flood Data

Horton (1914), an early worker in hydrology, in a discussion of a paper by Fuller (1914) stated that he had used logarithmic paper in hydrology prior to 1908. Hazen (1914), and in later work, used the logarithmic scales in lieu of arithmetic scales. He discusses these developments in a discussion of a paper by Hall (1921). His work was empirical. He developed a table of logarithmic skew factors. The factors used were the coefficient of variation and the coefficient of skew.

Apparently Hall (1921) was the first to propose the use of the Pearson Type III frequency curves in the analysis of flood flow data. Foster (1924, 1936) developed these ideas further. As is usual, the empirical methods and the theoretical methods have competed with each other, have complemented and have supplemented each other. With more and better data, with more extensive experience and sampling, a feel for their usefulness is developed. Gumbel's (1941a,b) models, Lieblein (1954), Mann et al. (1974) have undergone and are still being extensively tested. By 1962 (Beard, 1967) the use of the Pearson Type III (Gamma) distribution to fit logarithmically transformed flood data had become acceptable. By 1967 the U. S. Water Resources Council proposed that this method of analysis be made standard for Federal agency use. This proposal was made again in the updating of the 1967 issue in June 1977, Guidelines for Determining Flood Flow Frequency.

K. Pearson (1899, 1916) in early papers discusses the skew variation. However, Foster prepared the first published tables of the Pearson Type III skew course factors. Harter (1964, 1969) provides basic tables based on the chi-square distribution in the first reference and in the second reference he provides more extensive and more accurate tables than those previously available such as those published by Foster (1930). The work by Foster was exceptionally good in view of the materials and developments then available to him.

It is advisable here to iterate that the chi-square distribution for integral degrees of freedom, ν, is a Pearson Type III (gamma) distribution with a shape parameter equal to $\nu/2$. It has a mean $\mu = \gamma = \nu/2$, variance equal to 2ν, and a skew factor $g_1 = (8\nu^{-1})^{\frac{1}{2}}$. These relationships will be discussed in more detail later.

3. K-Statistics

As shown by Keeping (1962), the k-statistics (cumulants) may be expressed in terms of the population moments about the mean. These are biased. In unbiased form

$$k_1' = \mu_1' = \mu_1 = \mu \qquad (16)$$

$$k_2' = (n/(n-1))\mu_2 \qquad (17)$$

$$k_3' = n^2/(n-1)(n-2)\mu^3 \qquad (18)$$

$$k_4' = (n^2/(n-1)(n-2)(n-3)\ ((n+1)\mu_4 - 3(n-1)\mu_2^2)) \qquad (19)$$

Now

$$k_1 = \mu \quad \text{always unbiased} \qquad (20)$$

$$k_2 = \sigma^2 \quad \text{now unbiased} \qquad (21)$$

$$g_1 = n^{\frac{1}{2}}(n-1)^{\frac{1}{2}}(n-2)^{-1}\mu_3/\mu_2{}^{3/2} \quad \text{unbiased} \qquad (22)$$

$$g_2 = (((n+1)(n-1)/(n-2)(n-3))\mu_4/\mu_2{}^2)-3(n-1)^2/(n-2)(n-3) \quad \text{unbiased} \qquad (23)$$

The word unbiased here implies only that usual sampling size bias is removed.

With increasing n, and as $n \to \infty$

$$g_1 = \mu_3/(\mu_2)^{3/2} = 2\gamma^{-1/2} \qquad (24)$$

$$g_2 = (\mu_4/(\mu_2)^2)-3 = 6\gamma^{-1} \qquad (25)$$

the forms previously shown by Wallis et al. (1974) in their equation (26) which show the first kurtosis factor to include a plus 3. The plus 3 and the minus 3 balance leaving $g_2 = 6\gamma^{-1}$.

Comparison is now made between the debiasing factor for the skew shown above $n^{1/2}((n-1)^{1/2}(n-2)^{-1})$ and the debiasing factor given by Crutcher and Joiner (1978) $(n+3)/(n-1)\ \exp\ (1/(n^{(n-3)/2}))$ for the shape parameter. These are related through the expression $g_1 = 2\gamma^{-1/2}$. For sample size 10 the debiasing to two decimals rounded would be $(1.19)^2 \simeq 1.42$ versus 1.44. If the value 1.19 is multiplied by $(1 + 8.5/n)$, as suggested by Hazen (1930), the value $(2.20)^2$ is obtained which appears to be far too great. The last debiasing algorithm is not included here or in the USWRC Paper #17, 1977.

Table 4.--Comparison of bias factors for gamma distribution moments estimators (1) Wallis et al. (1974), and (2) Crutcher and Joiner (1978): (1) ;$(n)(n-1)(n-2)^{-2}$ and (2) ;$(n+3)(n-1)^{-1} \exp(n^{-(n-3)/2})$

n	1	2	3	4	5	7	10	15	30	50	100	1000
(1);	0	∞	6.00	3.00	2.22	1.68	1.41	1.24	1.11	1.06	1.04	1.003
(2);	∞	20.56	8.15	3.84	2.44	1.67	1.44	1.29	1.14	1.08	1.04	1.004
(2)/(1);	∞	0	1.36	1.28	1.10	0.99	1.03	1.04	1.02	1.01	1.01	1.001

Table 4 provides some comparison for the theoretical bias in the third moment with the empirical bias given in Crutcher and Joiner (1978). Here the shape parameter is greater than 2 which implies that the skew factor $g_1 < 2^{1/2}$ (Wallis et al., 1974). These differences of the quadratic and exponential fits reflect in the local maximum and minimum at samples sizes 7 and 15. The expression used by Crutcher and Joiner (1978) is simply an estimated model of best fit to sampling. It is realized that this in itself may be in error but if so, the error is not believed to be large. Another speculative point is that the random number generators may be deficient. Still another is that the biases are a function of the shape parameter itself. This last appears to be so for small shape parameters and small sample size. Figures 2 and 3 for small shape parameters and sample sizes sharply point up this inference.

4. Comparison Sample

Harter (1969) provides tables for both positive and negative skew in the Pearson Type III distribution. These may be used with either data in their original form or in a transform. The transform used is the logarithmic in consonance with the recommended practices by the National Water Research Council (USWRC) publications (1967 and 1977).

The data used are the peak 24-hour flood discharges (in thousands of cubic feet per second) on the Tennessee River at Chattanooga, Tennessee, during each of the 61 years 1874-1934. These are duplicated and are arranged in order from lowest to highest.

85.9	142	161	180	188	195	217	248	259	271
108	143	162	181	189	201	220	252	261	275
123	143	163	182	189	202	221	252	266	283
125	146	164	183	190	210	229	252	266	285
133	146	167	183	195	210	229	253	267	310
134	157	178	188	195	212	246	254	269	349
									361

Various authors have analyzed these same data. For comparison purposes, though these are not used directly, the estimates prepared by Harter (1969) are used. Harter's estimates for the original data x and the transformed data $y = \log_{10} x$ are $m_x = 207.359$, $s_x = 57.340$, $g_{1.x} = 0.37096$, $m_y = 2.29972$, $s_y = 0.12481$, $g_{1.y} = -0.40747$. Please note the comment of Harter that the skew of the transformed data is greater than that of the original data. It is also opposite in sign. The usual expectation is the reduction of the magnitude of the skew for this is the reason for the use of the transformation. This does not happen here.

Previously, Harter indicated that $g_{1.x} = (8/\nu)^{\frac{1}{2}}$. But $\nu = 2\gamma$ so that $g_{1.x} = (8/2\gamma)^{\frac{1}{2}} = 2/\gamma^{\frac{1}{2}}$. This relationship has also been shown. With $g_{1.x}$ equal to 0.37096 and $\frac{1}{2}\chi^2 = 2\gamma$ or $\chi^2 = 4\gamma$,

$$g_{1.x} = (8/4\gamma)^{\frac{1}{2}} = (2/\gamma)^{\frac{1}{2}} \tag{26}$$

$\gamma = 2/(g_{1.x})^2 = 2/(0.37096)^2 = 14.5337$. By use of tables in Roy, Gnanadesikan, and Srivastava (1971), with interpolation, the quantile for the 0.99 probability level is 24.8369. The skew factor (0.37096) has been obtained with the debiasing factor of $(n/(n+1-r)!)$ where r for the skew is 3 and the factor reduced by $(n/(n-2))$. When this quantile is multiplied by the scale parameter

$\beta = m_x/\gamma$ or 207.359/14.5337, the expected 100-year flood is obtained. This is 24.8369 times 14.2675 or 354.4 to one decimal place. The units are thousands of cubic feet per minute. This value of 354.4 is compared with the 356.1 for x and the 356.7 for y as defined by Harter, where y is equal to ln x.

The closeness of these numbers to each other suggests that, at least for this data set, the use of the transformed values or untransformed values produced by Harter's procedures as recommended by USWRC (1977) provides essentially the same answers. These also compare favorably with the use of the same estimates with available probability and quantile values. However, the procedures do not consider the use of debiased values in consideration of shape and sample size as discussed in this paper. These are now discussed.

The computer routines of EDS-24 and NOAA Technical Memorandum EDIS-28 may be used. The latter has the options to use either the moments estimators (ME) or the maximum likelihood estimators (MLE) in usual form or in debiased form. Also, there is the argument as to whether (n) or (n-1) should be used in the estimation of the variance. Thus, there are several other estimates that may be made for the 100-year flood from the same set of Chattanooga data. The programs used here do not consider the possible existence of negatively skewed sets of data. This happens in this case for the logarithmic transformed set of data. There are ways to handle this situation but such are not a part of this discussion. Harter (1969) implies or the present authors infer from Harter's work and their own work, that little is accomplished in transformation if the resultant skew is larger in magnitude than that which also resides in the data set and in particular when it is of opposite sign. The results for four runs of the Chattanooga using the electronic computer program of NOAA Technical Memorandum EDIS-28 are given in Table 5 here for the 100-year flood.

Table 5. Five 100-year estimates of 24-hour flood discharges (1000's cu. ft. sec^{-1}) for Chattanooga, TN (1874-1934) based on 61 years of data; moments (ME) and maximum likelihood estimates (MLE); biased (b) and debiased values (db). The asterisk (*) indicates that the factor (n-1) rather than (n) was used to compute the variance estimate.

	Mean	Origin	Scale	Shape	X^2	Prob. X^2	100-year flood (1000's cu ft sec^{-1})
ME(b)*	207.359	0	15.856	13.077	6.377	.504	363.498
ME(db)*	207.359	0	16.899	12.271	6.705	.540	369.276
ML(b)	207.359	0	16.018	12.945	6.377	.540	364.404
MLE(db)	207.359	0	16.782	12.356	6.705	.540	368.629
ME(b)	207.359	0	15.596	13.295	4.738	.308	362.026

The X^2 values and the Prob X^2 values are estimates of the fit of the gamma distribution to the data set. The null hypothesis that there is no difference in the data distribution and the gamma model is not rejected.

Figures 4a, b, c and d show the computer plot output for the next to the last set of estimates of Table 5, the debiased maximum likelihood estimates. The mean, scale and shape parameter estimates for the Chattanooga flood discharge data are 207.359, 16.782 and 12.356, respectively. The shape is dimensionless while the dimensions of the mean and the scale are thousands of cubic feet per second. Figure 4a shows the density distribution curve. Figure 4b shows the frequency differences between the expected and the actual frequencies by equal probability intervals except the last. Figure 4c shows the cumulative frequencies derived from data of Figure 4b. The data of Figure 4b and 4c, respectively, provide the bases of the chi-square and the Kolmogorov-Smirnov (K-S) tests.

Figure 4d is perhaps more important than the others, or perhaps easier to use, in the visual assessment of the fit of the model to the data. Against the line of best fit, the heavy 45^{o} line, the data and four confidence bands are plotted. Visual assessment by the authors infers the fit to be good which concurs with the previous assessment using the chi-square and Kolmogorov-Smirnov test data.

There are now eight estimates for the 100-year flood discharge at Chattanooga, TN based on the same set of data. These are, to one decimal, and to whole numbers, respectively:

(1)	354.4	354	Uses skew to estimate the shape parameter and tables by Roy, Gnanadesikan, and Srivastava
(2)	356.1	356	Uses skew correction tables by Harter and untransformed data
(3)	356.7	357	Uses skew correction tables and logarithmic transformed data
(4)	362.0	362	Uses biased moment estimator (ME) based on (n) rather than (n-1)
(5)	363.5*	364	Uses biased moment estimator (ME)
(6)	364.4	364	Uses biased maximum likelihood estimator (MLE)
(7)	368.6	369	Uses debiased MLE
(8)	369.3*	369	Uses debiased ME

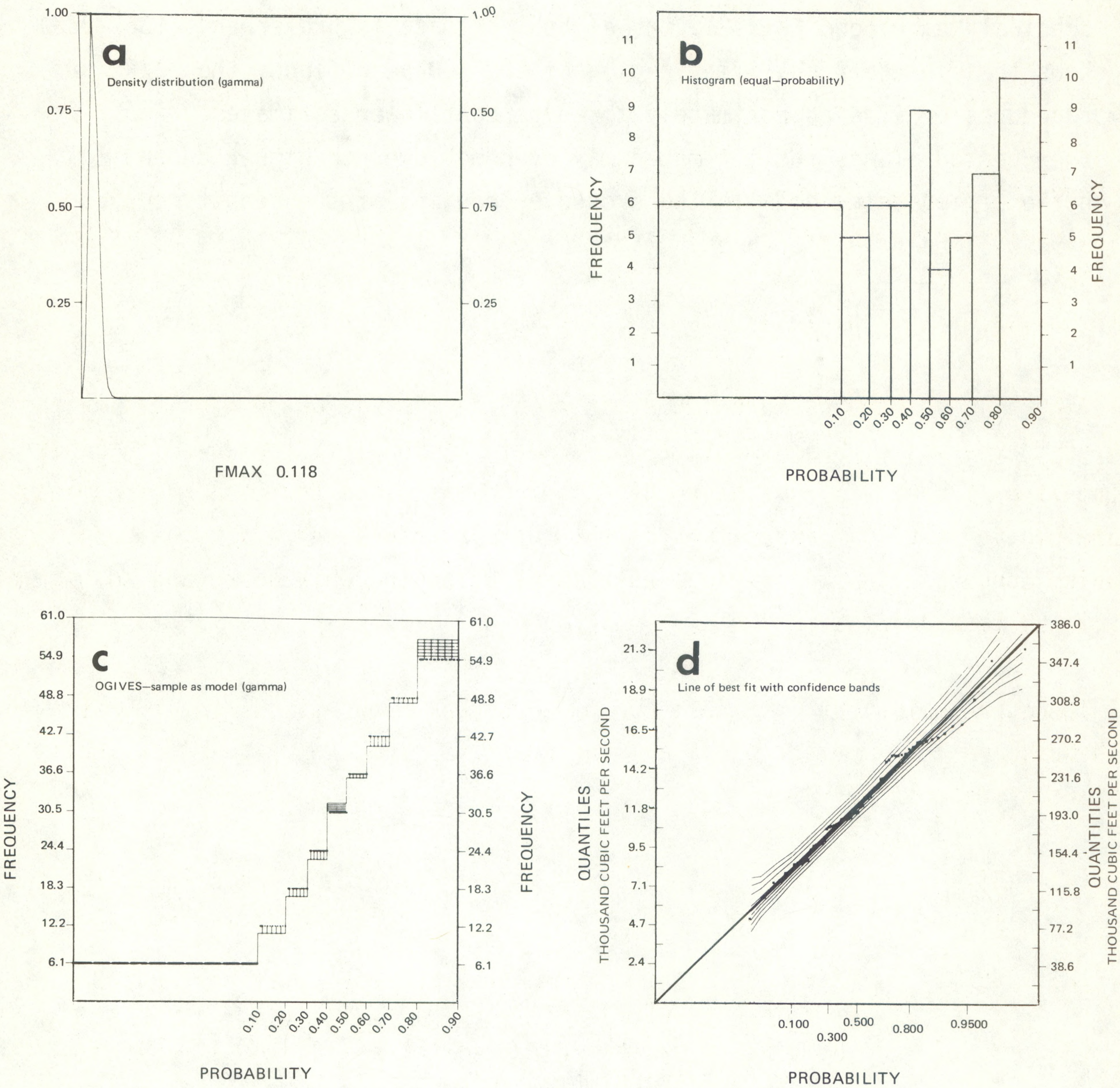


Figure 4. 24–HOUR FLOOD DISCHARGE AT CHATTANOOGA, TENNESSEE 1874–1934
(THOUSAND CUBIC FEET PER SECOND)

The asterisk (*) indicates that the factor (n-1) rather than (n) was used to compute the variance estimate.

Based on the work and arguments presented in this paper and in NOAA Technical Memorandum EDIS-28, the eighth estimate is preferred. (367-369) shows that in the probability region of the 100-year flood, the USWRC procedures provide approximately three percent lower estimates. For other period floods the estimates may even be lower or higher depending on the period selected. In the crossover regions of the density curves there will be no differences.

D. Summary Discussion of MLE and ME Bias Reduction

Prior work by several investigators indicates the need to use unbiased estimates of the shape parameter in the gamma distribution. Some indication for shape parameters below 0.5 is shown by Anderson and Roy (1975).

This paper extends the work of Crutcher and Joiner (1978) to shape parameter values as low as 0.02 and for sample sizes down to and including 2. Acceptable percent errors in the probability levels occur only in very narrow ranges of the quantiles. When no debiasing is made, similarly, and not unexpectedly, acceptable quantile percent errors exist only in very narrow probability intervals.

The usual argument to produce unbiased estimates where the moments are used is reflected in a multiplier, $(n/(n+1-r)!)$ where r is the moment under consideration and $0 > r < \infty$. The zeroth moment is excluded. This procedure does not provide adequate debiasing.

This paper clearly demonstrates that bias in ME's and MLE's of the shape parameter is a function of sample size as well as the shape parameter itself. As the skew and the coefficient of variation are directly related to the shape parameter bias in estimates of these inherently exist also.

The debiasing algorithms for the moments estimators in this paper are to be used where only "n" has been used and the argument $(n/(n+1-r)!)$ has not been utilized. Where the argument $(n/(n+1-r)!)$ has been used in the computation of the skew the bias for skew is a function of $(n^{\frac{1}{2}}(n-1)^{\frac{1}{2}}(n-2)^{-\frac{1}{2}})$. If the sample is large, the bias in the skew is negligible.

Where the shape parameter is less than 2 and the sample size is small, less than 100, the bias should be removed. The ME bias generally is larger than the MLE bias for the shape parameter. For the skew parameter only the bias for the ME has been made. No Monte Carlo study has been made here for the skew though as it is the inverse square of the shape parameter, equivalent bias is expected to hold.

It is suggested that if ME's or MLE's are used, the procedures given in this paper will work for any data set where the causal or genetic drive is known to be a gamma drive. This holds where the resultant data may be described by the gamma distribution (Pearson Type III). These procedures also will give comparable answers in hydrologic flood problems to the log gamma (Log Pearson) techniques recommended by the National Water Resources Council (1977). For small samples the answers should be better.

The differences between the output for the procedures presented in this paper and the log Pearson procedures which utilize the skewed normal might, speculatively, be attributed to the arguments presented in the discussion of Figure 1. The two procedures will produce different results whether or not the bias is removed. Practically, the results may not be significantly different though they will be different at some probability level.

This paper extends the work of the authors to the examination of the inherent bias in the estimation of shape parameters of a value equal to or less than 1. Tables are given which provide sufficient point values among which interpolation can be done as required. Algorithms are not produced here.

The thrust of this paper is that if the gamma distribution is to be used at all, unbiased estimates of the parameters, especially the shape parameter, should be used.

Items for Future Work.

1. In EDIS-30, algorithms were developed for the debiasing of estimates of the gamma parameters for shape parameter equal to or greater than one and the placement of confidence bands on the line of best fit. These algorithms were incorporated respectively into electronic computer programs described in NOAA Technical Memorandum EDIS-28. This paper provides the schematic for refinements of the above algorithm and extends the debiasing factors down to a shape parameter of 0.02 and a sample size of 2. By table look-up, ME's and MLE's can be debiased.

There remains the work to incorporate the results of this study into a computer program.

2. In EDIS-30, Tilde ($\sim$) and Jackknife $G(\hat{\gamma})$ estimators (Anderson and Roy (1975)) were discussed briefly and tables from Anderson and Roy (1975) were shown. These estimators drastically reduce the bias but the values shown indicate that some improvement yet is possible.

There remains work to be done to look at the use of the Tilde $\sim$ and Jackknife, $G(\hat{\gamma})$ estimates and to attempt to remove the remaining bias in these estimates.

3. Work needs to be done on the estimation of the origin parameter. This paper and its predecessors discuss some of the difficulties involved. In the near future, it is hoped some work will be done on the estimation of the origin parameter.

4. There still remains the considerable work to be done to establish the three-way relationship of the mean, scale and shape parameters. The current studies have considered only the shape parameter, essentially holding the mean fixed. This then determines the scale parameter estimate. Estimation where all three parameters are inter-dependent (or perhaps intra-dependent) is most difficult.

REFERENCES

Anderson, C. W., and Roy, W.D., 1975: Improved maximum likelihood estimators for the gamma distribution. Communications in Statistics, Vol. 4 (5), 437-448 pp.

Beard, Leo R., 1974: Flood Flow Frequency Techniques. A report for Office of Water Research and Technology and Water Resources Council, Oct. 1974. Center for Research in Water Resources, University of Texas at Austin. Technical CRWR-119 Report.

Bowman, K. O., and Beauchamp, J. J., 1975: Pitfalls with some gamma variate simulation routines. Journal of Statistical Simulation, Vol. 4, 141-154 pp.

Bridges, T. C., and Haan, C. T., 1972: Reliability of precipitation probabilities estimated from the gamma distribution. Monthly Weather Review, Vol. 100 (8), 607-611 pp.

Chapman, Douglas G., 1956: Estimating the parameters of a truncated gamma distribution. Annals of Mathematical Statistics, Vol. 27, 498-506 pp.

Crutcher, Harold L., McKay, Grady F., and Fulbright, Danny C., 1977: A note on a gamma distribution computer program and computer produced graphs. NOAA Technical Report EDS 24, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Washington, DC, 55 pp. plus 3 appendices.

Crutcher, Harold L., Fulbright, Danny C, McKay, Grady F., 1980: A note on a gamma distribution computer program. NOAA Technical Memorandum EDIS-28, National Oceanic and Atmospheric Administration, U. S. Department of Commerce, Washington, DC.

Crutcher, Harold L., Joiner, Raymond L, 1978: Gamma Distribution Bias and Confidence Limits, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Washington, DC, NOAA Technical Report EDIS-30, 1-46 pp plus appendices.

Elderton, William Palin, 1953: Frequency Curves and Correlation, Harren Press, D.C. 4ed 272 pp.

Elderton, William Palin and Johnson, Norman Lloyd, 1969: Systems of Frequency Curves, Cambridge University Press, London, England.

Fisher, Ronald A., 1922: On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London, Series A, Vol. 222, England, 309-368 pp.

Foster, H. Allen, 1924: Theoretical frequency curves and their application to engineering problems. Transaction ASCE Vol. 87 142-123 pp, 1924. Discussion pp 174-203.

Foster, H. Allen, 1934: Duration curves. Transaction ASCE Vol. 99, pp 1213-1267.

Fuller, W. E., 1914: Flood flows. Transactions ASCE, Vol, 77, 564-617 pp.

Greenwood, J. Arthur, and Durand, David, 1960: Aids for fitting the gamma distribution by maximum likelihood. Technometrics, Vol. 2 (1), 55-65.pp

Gumbel, E. J., 1941a: The return period of flood flows. Annals of Mathematical Statistics, Vol. 12, 163-190 pp.

Gumbel, E. J., 1941b: Probability interpretation of the observed return period of floods. Transactions of the American Geophysical Union, Vol. 22, 836-849 pp. Discussion 849-880 pp.

Hall, L. S., 1921: The probable variations in yearly run-off as determined from a study of California streams. Transactions AMSE, Vol. 84, 191-213 pp.

Harter, H. Leon, 1964: New tables of the incomplete gamma function ratio of percentage points of the chi-square and beta distribution. Aerospace Research Laboratories, Office of Aerospace Research, U.S. Air Force, Wright-Patterson Air Force Base, OH

Harter, H. Leon, and Moore, Albert H., 1965: Maximum likelihood estimation of parameters of gamma and Weibull populations from complete and from censored samples. Technometrics, Vol. 7, 4, 639-643 pp.

Harter, H. Leon, 1969: A new table of percentage points for the Pearson type III distribution. Technometrics, Vol. 11, No. 1, Richmond, VA. Feb. 1969, 177-187 pp.

Hazen, A., 1914: The storage to be provided in impounding reservoirs for municipal water supply, Transactions ASCE, Vol. 77, 1549 pp.

Hazen, A., 1930: Flood flows, John Wiley and Sons, New York, 1930: 199 pp.

Horton, R. E., 1914: Discussion of paper, flood flows. Transactions ASCE Vol. 77, 663-670 pp.

Johnck, M.D., 1964: Erzung von betaverteilter and gammaverteilter zufallszahlen, Metrika, Vol. E, 2.

Keeping, E.S., 1962: Introduction to Statistical Inference. D van Nostrand Company, Inc., New York, 451 pp.

Kendall, M. G., and Stuart, A., 1963: The Advanced Theory of Statistics, Vol II. Charles Griffin and Company, London, 521 pp.

Lieblein, Julius, 1954: A new method of analyzing extreme-value data. National Advisory Committee for Aeronautics, Technical Note 30-53.

Mann, Nancy R., Schafer, R. E., and Singpurwalla, Nozer D., 1974: Methods for Statistical Analysis of Reliability and Life Data. John Wiley & Sons, New York, NY, 564 pp.

Marsaglia, G. K., Anathanarayanan, N., and Paul, N., 1972: Random number generator package, "Super-Duper" uniform, normal and random number generator. McGill University, School of Computer Science, Computing Center, Montreal, Canada.

Norlund, Niels Erik, 1924: Vorlesungen uber differenzen-rechnung. Die Grundlehren der Mathematischen Wissenshaften. In Einzeldarstellungen Mit Besonder Beruchsichtigung Der Anwendungssgebiete with 54 text figures, Published by Julius Springer, 1921. pp 99-100

Pearson, Karl 1894: Contributions to the mathematical theory of evolution Philosophical Transactions of the Royal Society of London. Part 1, Series A, Vol. 185, England pp 71-110.

Pearson, Karl, 1916: Mathematical contributions to the theory of evolution XIX. Second supplement to a memoir on skew variation. Philosophical Transactions of the Royal Society of London, Series A, Vol. 216, England 429-457 pp.

Phillips, Don T., and Beightler, Charles S., 1972: Procedures for generating gamma variates with non-integer parameter sets. Journal - Statistical Computing Simulation Vol 1, pp 197-208. Gordon and Breach Science Publishers Ltd, Great Britain.

Roy, S. N., Gnanadesikan, R., Srivastava, J. N., 1971: Analysis and Design of Certain Quantitative Multiresponse Experiments. Pergamon Press, New York, NY 304 pp.

Sowey, E. R., 1972: A chronological and classified bibliography on random number generation and testing. International Statistical Review, Vol. 40 (3), 355-371 pp.

Thom, Herbert C. S., 1947: A note on the gamma distribution. Statistical Laboratory, Iowa State College, Ames, IA, 14 pp (unpublished manuscript).

Thom, Herbert C. S., 1958: A note on the gamma distribution. Monthly Weather Review, Vol. 86 (4), 117-122 pp.

Thom, Herbert C. S., 1968: Direct and inverse tables of the gamma distribution. ESSA Technical Report EDS-2, Environmental Science Services Administration, U.S. Department of Commerce, Silver Spring, MD, 30 pp.

United States Water Resources Council, 1977: Guidelines for determining flood flow frequency. Bulletin #174 of the Hydrology Committee First issued as Bulletin #15, 1967. United States Water Resources Council, 2120 L Street, NW, Washington, DC 20037.

Wallis, J. R., N.C. Matalas, and J. R. Slack, 1974: Just a moment. Water Resources Research, Vol 10,

AUTHOR INDEX

(Continued from inside front cover)

EDSTM 15 Improved Estimates of Winds at Standard Heights Generated From Winds Recorded at Standard Pressure Levels. Harold L. Crutcher, Russell F. Lee, and H. B. Harshbarger, March 1970. (PB-192-904)

EDSTM 16 Georgia Tornadoes. Horace S. Carter, July 1970. (PB-194-209)

EDSTM 17 Ground Rainfall Data for the 1968 Florida Cloud Seeding Experiment. B. G. Holzman and Marcella Thom, August 1970.

NOAA Technical Memorandums

EDS 18 The Effect of Atmospheric Aerosol on Climate With Special Reference to Surface Temperature. J. Murray Mitchell, Jr., November 1970. (COM-71-00341)

EDS 19 Terrain and Climate. Robert L. Schloemer, April 1971.

EDS 20 Drought Bibliography. Wayne C. Palmer and Lyle M. Denny, June 1971. (COM-71-00937)

EDS 21 Bibliography of the Urban Modification of the Atmospheric and Hydrologic Environment. John F. Griffiths and M. Joan Griffiths, February 1974. (COM-74-10962/AS)

EDS 22 Probability of Sequences of Wet and Dry Days for Tennessee. John M. Safley, Jr., Henry A. Fribourg, John V. Vaiksnoras, and Rodney H. Strand, September 1974. (COM-75-10650)

EDS 23 Probability of Low and High Temperatures in Tennessee. Henry A. Fribourg and Dewayne L. Ingram, March 1978. (PB-280-741)

EDS 24 Heating and Cooling Degree Days for Tennessee. Dewayne L. Ingram and Henry A. Fribourg, March 1978. (PB-280-740)

EDIS 25 Cost-Benefit Analysis of Selected Environmental Data Service Programs. Robert C. Anderson and Norman F. Meade, January 1979, 61 p. (PB-297-189)

EDIS 26 GATE Radar Rainfall Processing System. Vernon L. Patterson, Michael D. Hudlow, Peter J. Pytlowany, Frank P. Richards, and John D. Hoff, May 1979, 41 p. (PB-297-530)

EDIS 27 GATE Radar CAPPI Derivation for Three-Dimensional Reflectivity Analyses. Peter J. Pytolowany, F. P. Richards, and M. D. Hudlow, July 1979, 36 p. (PB-301-263)

EDIS 28 A Note on a GAMMA* Distribution Computer Program. Harold L. Crutcher, Danny C. Fulbright, and Grady F. McKay, March 1980.

NOAA SCIENTIFIC AND TECHNICAL PUBLICATIONS

The National Oceanic and Atmospheric Administration was established as part of the Department of Commerce on October 3, 1970. The mission responsibilities of NOAA are to assess the socioeconomic impact of natural and technological changes in the environment and to monitor and predict the state of the solid Earth, the oceans and their living resources, the atmosphere, and the space environment of the Earth.

The major components of NOAA regularly produce various types of scientific and technical information in the following kinds of publications:

PROFESSIONAL PAPERS — Important definitive research results, major techniques, and special investigations.

CONTRACT AND GRANT REPORTS — Reports prepared by contractors or grantees under NOAA sponsorship.

ATLAS — Presentation of analyzed data generally in the form of maps showing distribution of rainfall, chemical and physical conditions of oceans and atmosphere, distribution of fishes and marine mammals, ionospheric conditions, etc.

TECHNICAL SERVICE PUBLICATIONS — Reports containing data, observations, instructions, etc. A partial listing includes data serials; prediction and outlook periodicals; technical manuals, training papers, planning reports, and information serials; and miscellaneous technical publications.

TECHNICAL REPORTS — Journal quality with extensive details, mathematical developments, or data listings.

TECHNICAL MEMORANDUMS — Reports of preliminary, partial, or negative research or technology results, interim instructions, and the like.

Information on availability of NOAA publications can be obtained from:

ENVIRONMENTAL SCIENCE INFORMATION CENTER (D822)
ENVIRONMENTAL DATA AND INFORMATION SERVICE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
U.S. DEPARTMENT OF COMMERCE

6009 Executive Boulevard
Rockville, MD 20852

NOAA--S/T 80-181