Office of Chief Counsel

SUBJECT: Freedom of Information Act (FOIA) Request (FP-17-007297/FA-17-0003) for document entitled: Research on Technical Applications of Extreme Values; Accession Number: AD0280119; Author(s): Gumbel, Emil J., Columbia University New York School of Engineering And Applied Science; Report Date: 01 Apr 1962; Descriptive Note: Final Report

Mr. John Greenwald

Dear Mr. Greenwald:

This is in response to your Freedom of Information Act request, dated September 29, 2016, which was received by the undersigned on January 4, 2017 for the subject document.

Attached are the documents responsive to your request in their entirety, a total of 31 pages, fully releasable without redactions.

Fees in this matter have been waived in accordance with DOD Directive 5400.7-R.

If you have any questions, you may contact Edith Koleoso at the above address, by telephone at (301) 394-1636, or by e-mail to ARL-FOIA-Request@arl.army.mil.

Sincerely,

Timothy W. Connolly
FOIA Officer
U.S. Army Research Laboratory
NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incur no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.
Final Report  
April 1 1962

RESEARCH ON

TECHNICAL APPLICATIONS OF EXTREME VALUES.

Contract No. DA 30 069 Ord 1061

Statistical Engineering Group
Department of Industrial and Management Engineering
School of Engineering
Columbia University in the City of New York

FOREIGN ANNOUNCEMENT AND DISSEMINATION
OF THIS REPORT BY ASTIA IS LIMITED

Technical Supervisor
Office of Ordnance Research

Project Director, Professor Sebastian B. Littauer
Principal Investigator, Professor Emil J. Gumbel
Research on Contract DA 30-069 ORD 1061 with Columbia University, which began on 1 February 1953, was initiated in response to requests from General Leslie E. Simon and Charles Bicking of the then Office of Ordnance Research to provide methods of analysis of strength of materials. Their aim in encouraging this research was to improve the quality of materials of military materiel and, in particular, the materiel in production under Army Ordnance. The expectation that the methods of extreme values would provide methods beyond those of the statistics of the Normal, Binomial, Poisson and other more familiar distributions for analysing the behavior of materials under various forms of stress has been realized in twenty-two technical reports given in Appendix B. Six of these reports result from the first two years work on this contract from 1 February 1953 to 30 June 1955, which forms an integral part of these researches. Many publications have also appeared and they are given in Appendix C.

The primary applications of the statistical theory of extreme values have been to breaking strength and fatigue of materials. Other applications have been made to hydrology. In order, however, to provide the techniques which were effective in these applications it was necessary to conduct research in general statistical theory as well as in the statistical theory of extreme values. These researches,
then, have four main categories:

I. General Statistical Theory

II. Statistical Theory of Extreme Values

III. Applications to Breaking Strength and Fatigue of Materials

IV. Applications to Hydrology

A roster of technical personnel attached to this contract is given in Appendix A. Other scientific work engaged in by workers on this contract is given in Appendix D. Finally acknowledgement is made of the services of Mrs. Mary Roston who assisted the Principal Scientist and Project Director in carrying out many of the necessary administrative duties.

I. **GENERAL STATISTICAL THEORY**

Some of the phenomena of extreme value theory have significance in general statistical theory. One such useful concept is the "Return Period of Order Statistics", (Technical Report 21, publication 43). The return period \( T(X) \) is a mean number of observations on \( X \) such that for some value \( x \), the expected number among such observations which are equal to or greater than \( x \) is equal to one. If the observations all occur at equidistant times the return period has the dimension of time. Hence the name return period. This is expressed formally as follows: if \( P(X) \)
is the cumulative distribution function of $X$ and $T(X)$ is the return period for $X \geq x$, then

$$T(x) = \left(1 - P(x)\right)^{-1}$$

Although the return period is defined as an expected value verbally, its determination is as a conditional variable. Nevertheless, the return period can be studied under assumptions that the chance variable is both distribution and parameter free and, therefore, the return period of an order statistic is of general statistical interest.

Suppose we consider an order statistic of rank $m$, where the maximum observation in a sample of size $n$ has rank 1. Then one obtains for the expected value of $k$th moment $T_m^k$ of the return period $T_m$ of the order statistic of rank $m$, the expression

$$E(T_m^k) = \frac{n(n-1) \ldots (n-k+1)}{(m-1)(m-2) \ldots (m-k)}$$

This expression is finite only for $k \leq m-1$. An interesting particular result obtains for $n = 2$, where the expected return period is two for the second order statistic but is infinite for the first order statistic. Furthermore, the variance is infinite for either statistic.

Since "rare" events are formally representable as events of low probability they lend themselves to extreme value analysis. This connection has been shown in Technical
Report 14, which was published in a work on measurement theory (publication 29). An interesting consequence of the statistical analysis of rare events is that attempts have been made by researchers in extrasensory perception to validate their findings thereby. Richard von Mises probed this, so to speak, quasi scientific field with his armor of scientific insight and statistical analysis. E. J. Gumbel prepared an article from von Mises' findings which is to be published in his collected works.

Several studies have been devoted to bivariate distributions. As a remnant of nineteenth century opinions many statisticians believe that the so-called bivariate normal distribution may be used as a model for any bivariate distribution. A bivariate distribution, however, is not determined by its two margins. Even if they are normal, the bivariate density function need not be the usual exponential function of a quadratic form in the two variables $X$ and $Y$. The regression curves need not be linear and the curves of equal probability density need not be ellipses. On the contrary, there corresponds to given normal margins an infinity of bivariate distributions. A general form has been developed for effecting the construction of a bivariate distribution for any given margins (publication 18).
A family of bivariate probability functions with specified marginal probability functions \( P_1(X) = 1 - F_1(X) \) and \( P_2(Y) = 1 - F_2(Y) \), with parameter \( a, -1 < a < 1 \), is given by

\[
P(X, Y) = P_1(X) P_2(Y) \left[ 1 + aP_1(X) P_2(Y) \right].
\]

The parameter value \( a = 0 \) corresponds to independence. In the normal case the regression curves are linear functions of the probabilities and the correlation coefficient \( \rho \) is bounded by \(-1/3 < \rho < 1/3\). Tables of the bivariate probability function for given numerical values of the margin have been computed, where \( \rho \) is limited to the domain \(-1/3 < \rho < 1/3\). Generalizations to higher dimensions follow readily. Technical Report 12 and publications 32 and 35 deal with these topics.

Methods have been developed for constructing bivariate distributions from exponential and logistic margins. The resulting distributions are quite different from those previously obtained by statisticians. Technical Reports 15 and 16 and publications 42 and 46 present this material.

II. STATISTICAL THEORY OF EXTREME VALUES

In 1938 the Columbia University Press published E. J. Gumbel's book, *Statistics of Extremes*, which provides a systematic and comprehensive presentation of the statistical theory of extremes up to that date. The book was well.
received by both the profession and reviewers, and a second printing was made in 1959.

The asymptotic distribution of the range was generalized to the $m^{th}$ range in Report 11 and publication 30; the $m^{th}$ range is the difference between the $m^{th}$ value from the top and the $m^{th}$ value from the bottom. In a large sample the asymptotic density function of the $m^{th}$ range can be reduced to the corresponding function of the range proper, i.e. to a Bessel function. The $m^{th}$ mid-range is the sum of the $m^{th}$ extremes. It is shown (in Report 17) that its distribution is a generalization of the logistic distribution. The distributions of the $m^{th}$ range and the $m^{th}$ mid-range converge to normality with increasing rank $m$.

While it is known that the quantiles are asymptotically normally distributed and the two extremes are independent, the joint asymptotic distribution of a quantile and of the extremes was not known. This distribution was shown by Professor Berman (Report 22) to split, with increasing sample size, into the product of the normal distribution of the quantile and the extremal distribution, demonstrating the asymptotic independence of these two statistics.

Some results on the rejection of outliers by means of extreme value theory have been presented in publication Number 36. Extreme value theory seems to be the natural approach to the detection of outliers.
Considerable effort was devoted to study of the asymptotic bivariate and multivariate extremal distribution (Report 19 and publication 41). In the simplest case the bivariate distributions are the products of the corresponding univariate functions (publication 14).

In the bivariate case, there are three special cases and the three extremal distributions to consider.

There are, therefore six-fold infinities of such distributions (publication 14). General formulae which contain undetermined functions of the two marginal extremal probabilities are interesting from a theoretical point of view but are not of much practical use. What is needed are specific cases where the bivariate extremal distributions can be written down expressed by the univariate distributions and a factor depending on the correlation of the two marginal distributions. Two such cases are given in publications 39 and 44. Let

\[(\text{log} \mathfrak{A}(x, y)) = \xi; \quad \text{log} \mathfrak{A}(x) = \xi; \quad \text{log} \mathfrak{A}(y) = \eta\]

be transformations of the bivariate and the two marginal extremal distributions, then the two cases are given by

\[(5) \quad \eta^m = \xi^m + \eta^m \quad ; \quad m \geq 1\]
and
\[ \xi = \psi + \eta + \alpha \left( \frac{1}{\xi} + \frac{1}{\eta} \right)^{-1} \quad 0 \leq \alpha \leq 1 \]

The same formulae hold for bivariate extremal distributions of smallest values. The special cases \( m = 1 \) and \( a = 0 \) stand for independence. The formulae can easily be generalized into more than two dimensions.

III. Applications to Breaking Strength and fatigue of Materials

The study of fatigue and breaking strength of metals is an important field for the application of extreme value theory (publication 4). A popular resumé of the numerous aspects and technical applications in this domain was given in Report 7. There is given, in particular, an analysis of the corrosion problem, i.e. the pit depth of pipe lines, which is of great importance for the transportation of oil: the maximum pit depth increases as a linear function of the length of the line.

Previously only classical methods of statistics and curve fitting were applied to such problems. An important development in the analysis of fatigue was made by Weibull who used the third asymptotic distribution of smallest values, although on a purely empirical basis. The statistical analysis takes the number of cycles at failure under a given stress as a random variable. Although the
Stresses at failure are constant within each experiment it is proper to interpret them also as random variables and to use the same theory for both problems. These notions have been developed in Reports 8 and 9 and publications 8, 9, 10, and 13, where more than thirty sets of data on fatigue failure were analyzed. The most important characteristics are as follows: a) The minimum life $N_{0,s}$, i.e. the largest number of cycles with probability one of survival, which decreases with increasing stress (publication 26). Its inverse $S_{o,N}$ is the stress with probability of survival one. b) The endurance limit $S_0$, i.e. the stress so small that the specimen may survive an infinity of cycles, c) the probability of permanent survival $F(S_0)$ is a function of the stress. None of these values can be observed. In the present theory they take on the role of parameters which can be estimated if a sufficiently large sample of observations is made. The statistical theory is based on the two conditional survivorship functions

\[
\mathcal{L}(N|S) = \exp\left[-\frac{(N - N_{0,s})^2}{V_{S} - N_{0,s}}\right]
\]

\[
\mathcal{L}(S|N) = \exp\left[-\frac{(S - S_{0,N})^2}{V_{S} - S_{0,N}}\right]
\]

where the parameters $N_{0,s}$ and $S_0$ are as explained previously.
The number of cycles $V_g$ and its inverse, the stress $3_y$ are parameters of location while $\alpha_g$ and $\beta_y$ are parameters of scale and shape. Formula 8 can also be used for the probability of permanent survival. The analysis of these parameters and their interdependence is contained in Report 20 and publication 21. The parameters of scale which have no dimensions are linked to the lower limits.

This statistical approach was contrasted to the purely empirical procedures used by Bastenaire and Bennett in publication 12. Several semi-popular articles (publications 5 and 15) have been written on request.

While experiments may be performed so that we approach the minimum life no experiment is possible to approach the endurance limit. Instead of an infinity of cycles, a large number, $10^7$ or possibly $10^8$, is used for its estimation. The existing estimates of the endurance limits are based on an apparent discontinuity of the survivorship functions while the statistical theory must assume continuity. Therefore the estimates given in the literature may be of doubtful value. The problem of the endurance limit is treated in publication 6 and a test program appropriately designed for the estimation of the endurance limit is given in publication 2. Unfortunately no such large scale testing program has ever been undertaken.
A resume of the logical, physical and statistical reasons for preferring the extreme value theory in this analysis is given in publication 47.

IV. Applications to Hydrology

The first large scale application of the theory of extreme values arose from hydrological problems, in particular the analysis of floods and droughts. By their very definition they are the annual largest and smallest values of the daily discharges of a river at a given station, hence extreme values.

At the invitation of the Société Hydrotechnique de France, E. J. Gumbel wrote expository papers explaining his method of analysis, in particular the use of extreme value paper in graphical procedures for the analysis and (publications 7, 19, 31) forecast from drought and flood data. Using the return period scale for forecast, an analysis of records of twenty-five rivers provided fruitful examples of the use of these methods. In the course of the extensive discussion that arose from these papers, the question was raised as to why the first asymptotic distribution seemed to represent adequately the distribution of largest annual discharges (with occasional better representation by the second distribution). A further question was raised, namely,
why in the case of minimum discharges there could be, for some rivers, positive asymptotes and for other rivers zero least value.

At the invitation of the British Journal of the Institution of Water Engineers, an article was written in explanation of both these questions. In response, methods for constructing linear control boundaries which are useful in extrapolation and a graphical method for selecting among the three types of extremal distribution that type most effective for representing a given sequence of data were presented (Report 13, publication 16). The method of discrimination is based on the curvature which is assumed to hold for the whole sequence. In consequence of this analysis, it is shown that the first or second distributions of largest values is adequate for floods. So far, no example has been found where the third distribution (which has an upper limit) is indicated. This result is interesting in view of the fact that the objection of many engineers to the statistical theory of extreme values is made precisely because they feel that should the appropriate distribution have a finite upper bound. On the other hand, as might be expected, the third asymptotic distribution of the smallest values is useful in representing...
droughts. Further interest in the question of the most appropriate statistical representation of floods and droughts by way of communications from various countries was responded to in Report 1 and publication 27.

An important doubt as to the validity of using the present statistical theory of extremes which is based on an assumption of the independence of successive observations of daily discharges has been largely mitigated in work of Professor Berman who showed that the actual absence of independence in the observations did not invalidate the results obtained by the present methods. For further information, a list of recent papers on Extreme Values is added as Appendix E.
APPENDIX A

Personnel Who have served on the Scientific Staff of this C

Professor E. J. Gumbel, Principal Investigator
Professor S. B. Littauer, Project Director

Simeon Berman, now Assistant Professor, Department of Mathematics and Statistics, Columbia University
Cyrus Derman, now Associate Professor, Department of Industrial Engineering, Columbia University
Neil Goldstein, graduate student in mathematical statistics

T. T. Kuo, obtained Ph.D. in Industrial Engineering, Columbia University, 1961, now in Operations Research, National Cash Register Co., Dayton, Ohio
Seiti Sugihara, then a graduate student in mathematical statistics at New York University
Taro Yamane, since Assistant Professor of Statistics, New York University

Phillip G. Carlson, received D.Eng.Sci. in Industrial Engineering, 1962, now on Operations Research staff, Arthur Andersen and Co., New York; co-operated with Professor E. J. Gumbel on researches under that contract without formal appointment to the staff.

Jose Tiago de Oliveira, Professor of the Faculty of Sciences, Lisbon, Portugal.
APPENDIX B - TECHNICAL REPORTS

1. E. J. Gumbel - MINIMUM LIFE IN FATIGUE FAILURES
2. C. Derman, T. T. Kwo and E. J. Gumbel
   STANDARD ERRORS OF ESTIMATE OF PARAMETERS OF
   FATIGUE FAILURE SURVIVORSHIP FUNCTIONS
3. E. J. Gumbel - STATISTICAL ESTIMATION OF THE ENDURANCE LIMIT
4. C. Derman - SOME REMARKS ON THE ENDURANCE LIMIT PROBLEM
5. S. Sugihara - SOME TESTS FOR MINIMUM LIFE OF FATIGUE FAILU
   SURVIVORSHIP FUNCTIONS
6. E. J. Gumbel - STATISTICAL ESTIMATION OF THE ENDURANCE LIMIT
7. E. J. Gumbel - EXTREME VALUES IN TECHNICAL PROBLEMS (Reprin
   from Industrial laboratories, Vol. 7, no. 12 Dec., 1956)
8. E. J. Gumbel - STATISTICAL ANALYSIS OF FATIGUE
9. E. J. Gumbel - THE STATISTICAL ASPECT OF FATIGUE (Reprint f
   Columbia Engineering Quarterly, March 1957, pp. 3-7)
10. E. J. Gumbel + P. G. Carlson
    ON THE ASYMPTOTIC COVARIANCE OF THE SAMPLE M
    AND STANDARD DEVIATION (Reprint from Metron, vol. 13, No. 1, pp. 1-9, Roma, 1956)
11. E. J. Gumbel - THE MTH RANGE
12. E. J. Gumbel - MULTIVARIATE DISTRIBUTIONS WITH GIVEN MARGIN
13. E. J. Gumbel - STATISTICAL THEORY OF FLOODS AND DROUGHTS
    (Reprint from Journal of the Institution of
    Water Engineers, Vol. 12, pp. 157-181, Londo
14. E. J. Gumbel - MEASUREMENTS OF RARE EVENTS
15. E. J. Gumbel - BIVARIATE EXPONENTIAL DISTRIBUTIONS
16. E. J. Gumbel, A. Avishur, A. D. Benham, F. Law, R. W. S.
    Thompson, and D. H. Thompson
    COMMUNICATIONS ON THE STATISTICAL THEORY OF
    FLOODS AND DROUGHTS (Reprint from Journal of
    the Institution of Water Engineers, Vol. 13,
17. E. J. Gumbel - Statistical Theory of Extreme Values (Main Results)
18. E. J. Gumbel - A Bivariate Logistic Distribution
19. E. J. Gumbel - Multivariate Asymptotic Distributions of Extreme Values
20. E. J. Gumbel - On the Parameters of the Distribution of Fatigue Life
21. E. J. Gumbel - The Return Period of Order Statistics
22. S. Berman - The Asymptotic Independence of the Sample Quartile and the Sample Extreme Value
APPENDIX C - PUBLICATIONS


12. Discussion of the contributions of Mr. Bastenaire and Mr. Benn International Conference on Fatigue of Metals, British Institute of Mechanical Engineers, London 1956.


Dr. E. J. Gumbel made a number of trips abroad, disseminating the ideas developed in the researches sponsored by this contract. Descriptions of these activities have been published in articles 17, 25, 34 and 45. He also participated in a number of international scientific congresses, reported on in articles 24 and 40.

A by-product of these researches is a course on the Statistical Theory of Extreme Values and their Technical Applications, given at Columbia University (Catalogue Number: E 6611 x - E 6612 y. Engineering Applications of Extreme Values).

Professor S. B. Littauer (jointly with Professor S. Ehrenfeld) has worked on a text on Statistical Method in Engineering and Science, which is scheduled for publication early in 1963.
APPENDIX E

Publications Concerning Gumbel's Work 1956-1961


2. Howard J. Pincus: Some vector and arithmetic operations on two dimensional orientation variates, with applications to geological data. The Journal of Geology, Vol. 64, No. 6, pp. 533-557, November 1956


37. M. V. Yevdjevich: Hydrology (Serbian) Belgrade, pp. 88


54. H. C. S. Thom: Distributions of extreme winds in the United States, Proc. of the Am. Society of Civil Engineers, Structural Division, pp. 11-24, April 1960


61. Oliveira, J. Tiago: O ensaio X e os ensaios de concordancia, Gazeta de Matematica Nos. 68-69, Septembro/Dezembro, Lisboa 1


64. Phillip G. Carlson: A recurrence formula for the mean range for odd sample sizes, Skandinavisk Aktuartidsskrift 1958, Uppsala, 1959


66. G. N. Alexander: Return period relationships, Journal of Geo-


72. Gringorten, I. E.: Extreme value statistics in meteorology, Air Force Surveys in Geophysics, No. 125, Bedford, Mass, June 1


