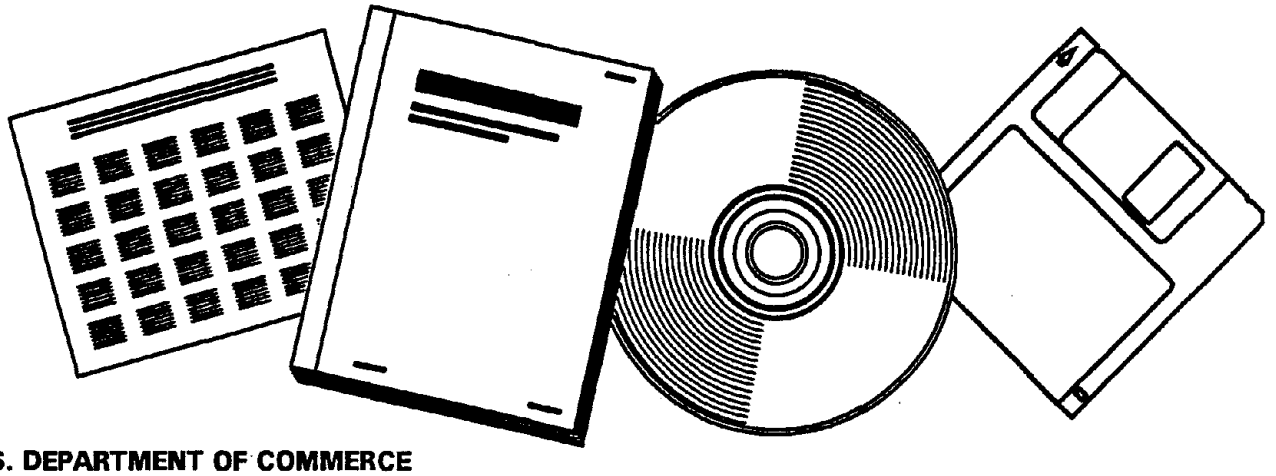
PB95-160594

JOURNAL OF RESEARCH OF THE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY JULY/AUGUST 1994. VOLUME 99, NUMBER 4 SPECIAL ISSUE: EXTREME VALUE THEORY AND APPLICATIONS. PROCEEDINGS OF THE CONFERENCE ON EXTREME VALUE THEORY AND APPLICATIONS VOLUME 2. HELD AT GAITHERSBURG, MARYLAND IN MAY 1993

(U.S.) NATIONAL INST. OF STANDARDS AND TECHNOLOGY
GAITHERSBURG, MD

MAY 93



U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

July–August 1994
Volume 99, Number 4

ISSN 1044-677X

Journal of Research of the

National Institute of Standards and Technology

United States Department of Commerce
Technology Administration
National Institute of Standards and Technology

The National Institute of Standards and Technology was established in 1988 by Congress to "assist industry in the development of technology . . . needed to improve product quality, to modernize manufacturing processes, to ensure product reliability . . . and to facilitate rapid commercialization . . . of products based on new scientific discoveries."

NIST, originally founded as the National Bureau of Standards in 1901, works to strengthen U.S. industry's competitiveness; advance science and engineering; and improve public health, safety, and the environment. One of the agency's basic functions is to develop, maintain, and retain custody of the national standards of measurement, and provide the means and methods for comparing standards used in science, engineering, manufacturing, commerce, industry, and education with the standards adopted or recognized by the Federal Government.

As an agency of the U.S. Commerce Department's Technology Administration, NIST conducts basic and applied research in the physical sciences and engineering and performs related services. The Institute does generic and precompetitive work on new and advanced technologies. NIST's research facilities are located at Gaithersburg, MD 20899, and at Boulder, CO 80303. Major technical operating units and their principal activities are listed below. For more information contact the Public Inquiries Desk, 301-975-3058.

Technology Services
- Manufacturing Technology Centers Program
- Standards Services
- Technology Commercialization
- Measurement Services
- Technology Evaluation and Assessment
- Information Services

Electronics and Electrical Engineering Laboratory
- Microelectronics
- Law Enforcement Standards
- Electricity
- Semiconductor Electronics
- Electromagnetic Fields[1]
- Electromagnetic Technology[1]

Chemical Science and Technology Laboratory
- Biotechnology
- Chemical Engineering[1]
- Chemical Kinetics and Thermodynamics
- Inorganic Analytical Research
- Organic Analytical Research
- Process Measurements
- Surface and Microanalysis Science
- Thermophysics[2]

Physics Laboratory
- Electron and Optical Physics
- Atomic Physics
- Molecular Physics
- Radiometric Physics
- Quantum Metrology
- Ionizing Radiation
- Time and Frequency[1]
- Quantum Physics[1]

Manufacturing Engineering Laboratory
- Precision Engineering
- Automated Production Technology
- Robot Systems
- Factory Automation
- Fabrication Technology

Materials Science and Engineering Laboratory
- Intelligent Processing of Materials
- Ceramics
- Materials Reliability[1]
- Polymers
- Metallurgy
- Reactor Radiation

Building and Fire Research Laboratory
- Structures
- Building Materials
- Building Environment
- Fire Science and Engineering
- Fire Measurement and Research

Computer Systems Laboratory
- Information Systems Engineering
- Systems and Software Technology
- Computer Security
- Systems and Network Architecture
- Advanced Systems

Computing and Applied Mathematics Laboratory
- Applied and Computational Mathematics[2]
- Statistical Engineering[2]
- Scientific Computing Environments[2]
- Computer Services[2]
- Computer Systems and Communications[2]
- Information Systems

---

[1]At Boulder, CO 80303.
[2]Some elements at Boulder, CO 80303.

Journal of Research of the National Institute of Standards and Technology

Volume 99 Number 4 July–August 1994

U.S. Department of Commerce—**Ronald H. Brown,** Secretary
Technology Administration—**Mary L. Good,** Under Secretary for Technology
National Institute of Standards and Technology—**Arati Prabhakar,** Director

The Journal of Research of the National Institute of Standards and Technology features advances in measurement methodology and analyses consistent with the NIST responsibility as the nation's measurement science laboratory. It includes reports on instrumentation for making accurate and precise measurements in fields of physical science and engineering, as well as the mathematical models of phenomena which enable the predictive determination of information in regions where measurements may be absent. Papers on critical data, calibration techniques, quality assurance programs, and well-characterized reference materials reflect NIST programs in these areas. Special issues of the Journal are devoted to invited papers in a particular field of measurement science. Occasional survey articles and conference reports appear on topics related to the Institute's technical and scientific programs.

ISSN 1044-677X **Coden: JRITEF** **Library of Congress Catalog Card No.: 89-656121**

United States Government Printing Office, Washington: 1994

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

Extreme Value Theory and Applications

Proceedings of the Conference on Extreme Value Theory and Applications, Volume II

Gaithersburg, MD
May 1993

Preface

It appears that we live in an age of disasters: the Mississippi and the Missouri rivers flood millions of acres, earthquakes hit Tokyo and California, airplanes crash due to mechanical failure, and powerful windstorms cause increasingly costly damage. While these may seem to be unexpected phenomena to the man in the street, they are actually happening according to well defined rules of science known as extreme value theory. For many phenomena records must be broken in the future, so if a design is based on the worst case of the past then we are not really prepared for the future. Materials will fail due to fatigue: even if the body of an aircraft looks fine to the naked eye, it might suddenly fail if the aircraft has been in operation over an extended period of time. Extreme value theory has by now penetrated the social sciences, the medical profession, economics and even astronomy. We believe this field has come of age. To utilize and stimulate progress in the theory of extremes and promote its application, an international conference was organized in which equal weight was given to theory and practice.

The Proceedings are published in three Volumes. Volume I, published by Kluwer Academic Publishers, contains papers of general interest in extreme value theory and practice. Volume II, this Special Issue of the NIST Journal of Research, contains papers deemed by the Committee to be most directly relevant to NIST's mission. Volume III, NIST Special Publication 866, contains papers selected for their important contribution to a number of specialized topics. All papers have been refereed and we are grateful to the many engineers and scientists from all over the world who served as referees.

The conference was held in May 1993 on the campus of the National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland, with its Statistical Engineering Division (SED) acting as host. It was organized by Temple University, Philadelphia, Pennsylvania, and NIST.

The conference had no external funding, and NIST's support was fundamental to its success. We are particularly grateful to Dr. Lundegard, Chief of SED, whose support was the single most important factor in making the conference happen. The support of NIST's Building and Fire Research Laboratory is also acknowledged with thanks.

The Organizing Committee consisted of Janos Galambos (Chairman), James Lechner, Stefan Leigh (Director of Local Arrangements), James Pickands III, Emil Simiu, and Grace Yang. Stefan's enthusiasm and tireless work was essential for the success of the Conference. The Conference included three special sessions:

The Centennial Session for Emil Gumbel. Churchill Eisenhart introduced the Session. His personal recollections of Gumbel are included in Volume I of the Proceedings. Emil Simiu then spoke on Gumbel's life and work.

The Memorial Session for Josef Tiago de Oliveira. Janos Galambos remembered Tiago, a close friend to many Conference participants, who was on the initial list of invited speakers. M. Ivette Gomes gave a detailed account of his work.

The 80th Birthday Session for B. V. Gnedenko. Janos Galambos summarized the work of Gnedenko as the founder of modern extreme value theory and his contributions to the central limit problem, limit theorems with random sample size, and renewal theory.

The Conference was opened by Dr. Robert Lundegard who emphasized extreme value theory's role in several scientific and engineering fields. It ended with a panel discussion on the future of extreme value theory and its applications. The Panel was chaired by Janos Galambos, and its members were Enrique Castillo, Laurens de Haan, Lucien Le Cam, and Richard L. Smith.

We sincerely thank Julian M. Ives of the NIST Publications Production Program for his invaluable assistance in editing and producing this Volume.

Janos Galambos
James Lechner
Emil Simiu
Editors

Charles Hagwood
Technical Editor

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

Contents

Extreme Value Theory and Applications

Proceedings of the Conference on Extreme Value Theory and Applications, Volume II

Articles

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

News Briefs

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 313 (1994)]

Applications of Extreme Value Theory in Corrosion Engineering

Volume 99 Number 4 July–August 1994

Philip A. Scarf

University of Salford,
Salford M5 4WT, U.K.

and

Patrick J. Laycock

University of Manchester Institute of Science and Technology,
Manchester M60 1QD, U.K.

In the context of corrosion engineering it is often natural to be concerned with extreme events. This is because, firstly, it is these extreme events that often lead to failure and, secondly, it may only be possible to measure the extremes, with much of the underlying measurements by their very nature unobservable. Statistical methods relating to extreme value theory can be used to model and predict the statistical behaviour of extremes such as the largest pit, thinnest wall, maximum penetration or similar assessment of a corrosion phenomenon. These techniques can be applied to the single largest value, or to a given number of the largest values, measured over individual areas or coupons; or to all values exceeding a given threshold. The data can be modeled to account for dependence on environmental conditions, surface area examined, and the duration of exposure or of experimentation. The application of a selection of these techniques is demonstrated on data from industry and from laboratory experiments.

Key words: corrosion; exceedances; extreme values; extreme value distributions; generalized Pareto distribution.

Accepted: March 22, 1994

1. Introduction

Extremes are typically defined in two ways. Either by selecting a suitable threshold and then recording every observation above that threshold; or by sorting the data, according to some *a priori* sampling scheme, so as to select the one, two, or three, etc., largest value(s). The nature by which the extremes are defined and hence measured is then indicative of the techniques appropriate for modeling and prediction. Most of the statistical methods relating to extreme values are based, in the first instance, on the assumption of an underlying large sample of possible measurements, all nominally arising from a single population of such possible measurements. For extreme value theory to be used, it is then only necessary for the actual extremes to be measured. The other possible measurements can be ignored and may even be unobservable with the equipment used to measure the extremes. The nature of the extreme may be that of a maximum value or a minimum value. In this paper we will assume that maximum values are of interest. In applications concerned with minima, negating the variable of interest will transform the problem into one concerned with maxima.

The generalized Pareto distribution (GPD) is the standard family of statistical distributions to be used as a basis for modeling data which arise as exceedances over some threshold. Applications of this approach for the first of the above extreme value definitions is examined in the following section. Methods to ensure the validity of the standard statistical assumptions while accumulating such data are discussed. The generalized extreme value (GEV) distribution can be shown to be the natural

one to use for single extremes. Data can arise as the largest value from each of a set of coupons (individual specimens), or from partitioning an area into equal smaller areas and selecting one maximum from each smaller area. The application of methods considering such single extremes is also considered. The joint generalized extreme value distribution (JGEV) is the appropriate distribution family to use when the r (say) largest values are extracted, instead of just the single largest value. This provides a useful extension to the classical theory in such a way as to match up with the common practice of measuring the few largest pits at any one location undergoing pitting. Using the r extreme order statistics in this way can increase the precision of the estimates in the model and hence improve predictions.

Dependence on time and area can be incorporated for prediction and extrapolation purposes when applying these distributions, and methods for modeling the dependence on environmental conditions, say, through covariates are indicated.

2. Exceedances Above a Threshold

These are data collected on the basis of all values exceeding a specified threshold, taken sufficiently "high" to imply that certain limiting statistical results will hold. The data in Table 1, on pit depths in two stainless steel roofs, were collected with just such a threshold, namely 6 μm, in operation. This threshold qualifies as "high" on the basis that a much lower one, such as 0.06 μm for example, would have produced a very much larger sample of nascent pits. This is consistent with theories of pitting in steel and other metals. See further argument supporting this approach in Ref. [1]. This type of data censoring can arise through built in limits on measurement capabilities or else through deliberate censoring of a given data set, typically a dense time series, so as to isolate the important events. When such data are extracted from a regular grid of values rather than through the engineer visually identifying isolated corrosion phenomena and taking one measurement on each, it may be necessary to edit the values so as to extract only local cluster maxima rather than using all nearby points. This is needed to "decouple" the recorded values and so validate the usual assumption of statistical independence or exchangeability. A careful combination of grid size (to match the scale of the phenomena being studied) and threshold (to select for significant phenomena) may be all that is necessary.

Table 1. Pit depths above 6 μm in stainless steel sheet college roofs (area 500 m^2; samples 10 cm^2; thickness 400 μm)

Roof 1 (50 months)
131 106 35 26 26 25 23 20 20 18 18 18 17 16 16 15 15 15 14 14
14 14 14 14 14 14 14 12 12 12 12 12 10 10 8 8 8 8 8 8 8 8 8
Roof 2 (29 months)
140 106 95 77 72 55 55 53 52 36 33 32 32 30 28 28 26 26 25 24
24 24 22 22 20 18 18 16 16 16 16 14 14 12 12 12 8 8 8

With this form of data set, both the number, n, of observations and their observed values $\{y_i\}$ are necessarily random variables. It can be shown, see for example Ref. [2], that, for sufficiently high thresholds, and for a wide variety of initial distributions, this number, n, of the exceedances, has asymptotically a Poisson distribution (with parameter λ, say) and their sizes, y, have a generalized Pareto distribution:

$$G(y) = 1 - (1 + \xi y/\sigma)^{-1/\xi}, \tag{1}$$

valid for $1 + \xi y/\sigma > 0$, with $\sigma > 0$ and $-\infty < \xi < \infty$. In particular, if these distributional results hold exactly for some particular threshold, u say, then the maximum of this set of values has a generalized extreme value distribution (see next section) exactly, and this will be true for all higher thresholds. A check that the distribution, Eq. (1), holds can be made by graphing the mean excess plot, in which the mean exceedances in the data are plotted against increasing threshold values. This plot should follow a straight line with slope $\xi/(1-\xi)$ and intercept $\sigma/(1-\xi)$; with a horizontal plot corresponding to $\xi = 0$ and a simple exponential distribution for the tail. For extrapolation over larger areas, for extremes derived from random sampling over a large structure, often the quantity of interest is the Nth return level

$$q_N = u - \frac{\sigma}{\xi}[1\text{-}(\lambda N)^{\xi}],$$

where N is either the number of "coupon multiples" as a measure of structure size, or else the number of time intervals into the future. The Nth return level is interpreted as that level which would be exceeded on average once every N units of area (or time).

The data in Fig. 1(a) are 1024 values of "current noise" collected during a study of the electrochemical nature of pitting. This series was "declustered" using a moving window of width 40 to give the isolated maxima in Fig. 1(b). A mean excess plot for the isolated maxima of the current noise data is given in Fig. 1(c). Consideration of this plot suggests that either a large threshold is required or that the exceedances arise from a mixture of the tails of underlying distributions. For an electrochemical interpretation of this latter phenomenon, it can be noted that large narrow current spikes have been described as being typical of intermittent pitting corrosion, while steady broader based but less variable current noise has been associated with general corrosion, see for example Ref. [3]. Intermediate conditions can be associated with persistent pitting, widely recognized as the most threatening scenario for metal structures.

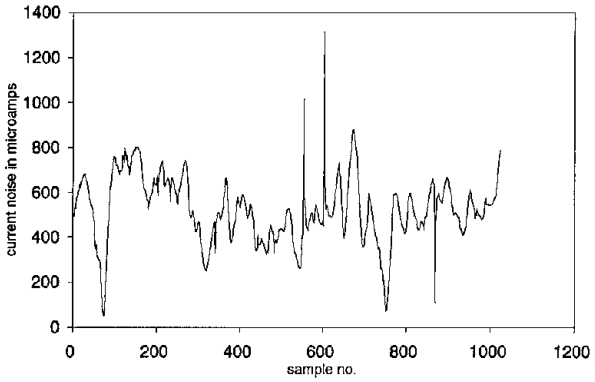


Fig. 1(a). Current noise measurements (sample size = 1024).

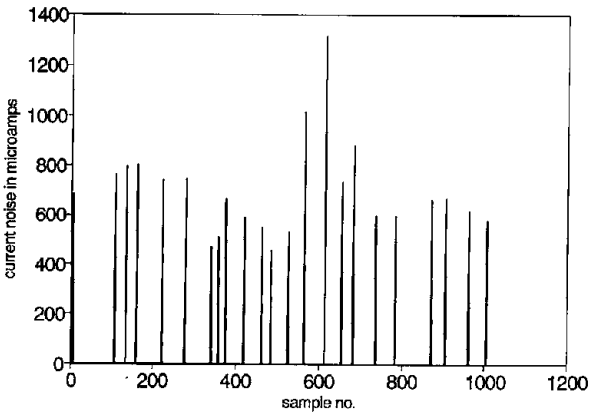


Fig. 1(b). Isolated peaks in current noise measurements.

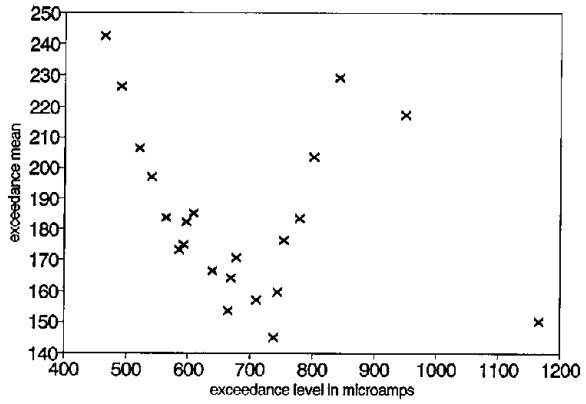


Fig. 1(c). Mean excess plot for current noise measurements.

The main difficulty which can arise with the threshold method is the choice of an appropriate threshold, especially when there is no *a priori* reason for choosing one particular threshold over another. In an experiment to consider the prediction of extreme corrosion rates for carbon steel in a simulated basalt groundwater [4], a number of 200 mm × 200 mm coupons were exposed for varying lengths of time. These coupons, having been first cleaned to remove all corrosion products, were profiled with spot heights taken at the nodes of a 1 mm lattice. This then gave, after making an adjustment for the original coupon surface, a 196 × 196 array of corrosion measurements. False-color histogram-equalization techniques, displayed on computer monitors, were used to validate and inspect the digitized spot heights from these coupons. A mean excess plot for a typical coupon exposed for 26 weeks is shown in Fig. 2(a). Note that this plot was drawn for both the raw exceedances and also for declustered exceedances. The process of declustering essentially amounted to identifying all those "pits" or clusters exceeding a particular threshold and calculating the maximum exceedance for each "pit." The mean excess plot indicates that a range of possible thresholds (300 μm–550 μm) would be appropriate for model fitting. Table 2 gives the results for such model fitting using maximum likelihood for a range of values of threshold. Here λ is the mean exceedance rate per m^2, σ, and ξ are the parameter estimates for the GPD, and q_{25} and q_{250} are those levels that would be exceeded once on average every m^2 and every 10 m^2 respectively. Standard errors are given in brackets. If the q_{25} is considered, we see that its estimated value decreases as the threshold increases, its value being highly sensitive to the value of ξ.

Table 2. Summary of model fitting and prediction using maximum likelihood for the generalized Pareto distribution for a typical 26 week basalt groundwater coupon profile

Threshold	Mean cluster exceedance (μm)	Number of clusters	λ	σ	ξ	q_{25}	q_{250}
300	99	177	4425	98.0	0.01	1158	1406
			(333)	(11)	(0.08)	(260)	(430)
350	92	146	3650	99.0	0.04	1205	1500
			(302)	(11)	(0.09)	(300)	(527)
400	97	96	2400	104.3	−0.08	1004	1102
			(245)	(16)	(0.11)	(214)	(322)
450	83	76	1900	83.4	−0.01	1057	1233
			(218)	(11)	(0.13)	(241)	(405)
500	90	50	1250	102.6	−0.14	963	1037
			(177)	(23)	(0.17)	(213)	(309)
550	87	29	725	108.5	−0.23	918	961
			(135)	(12)	(0.31)	(250)	(339)

For higher thresholds the large negative value of ξ is indicative of a tail distribution which is shorter than exponential so implying lower return values. For lower thresholds the tail appears to be exponential implying relatively higher return values. This effect can be seen further in an exponential probability plot of the exceedances above 300 μm, Fig. 2(b). As the threshold increases more weight is given to the extreme observations, which are themselves smaller than would be expected for an exponential tail. The lack of an objective method for determining the correct threshold therefore leads to difficulties in prediction.

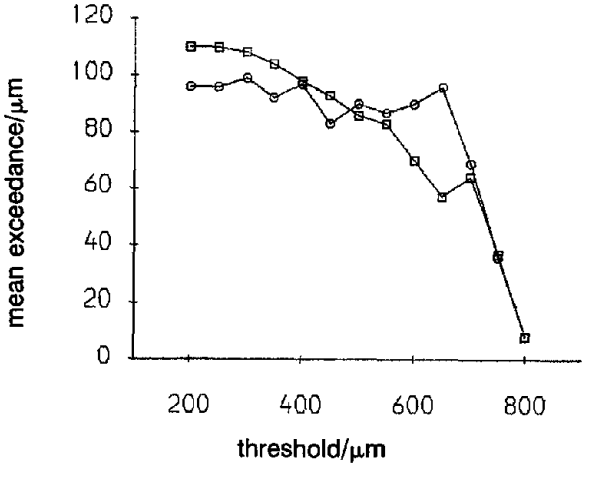


Fig. 2(a). Mean excess plot for typical 26 week basalt groundwater coupon profile: O—mean declustered exceedances; □—mean of all exceedances.

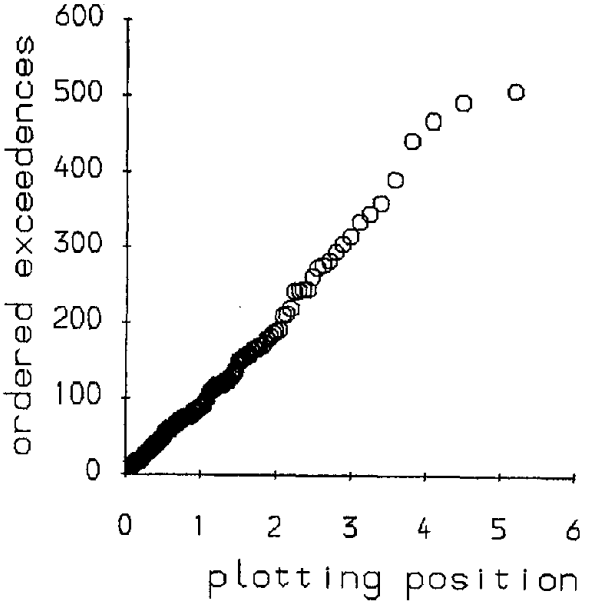


Fig. 2(b). Exponential probability plot of declustered exceedances above 300 μm.

3. Extreme Value Distributions

Data suitable for this type of analysis can arise as the largest value from each of a set of coupons, or from dividing an area into equal smaller areas and selecting one maximum from each smaller area, provided the scale of division and corrosion patterns are compatible in the sense described above for the generalized Pareto distribution. For a sample of independent identically distributed random

variables, $x_1,\ldots,x_n$, the distribution of $x_{\max}$, the data maximum, depends on n. Suppose however that there exist location and scale factors, a_n and b_n say, so that the rescaled variate, $y=a_n+b_nx_{(n)}$, has a distribution which is independent of n. This is the so-called "stability postulate," and leads immediately to the following functional equation (to be solved for F): $F(x)^n=F(a_n+b_nx)$. The solution to this equation is the generalized extreme value (GEV) distribution, which can be written in the following 3-parameter form:

$$F(x)=\exp\{-[1+\xi(x-\mu)/\psi]^{-1/\xi}\},$$

$$\xi x>\xi\mu-\psi=\xi\theta,\ \psi>0. \qquad (2)$$

See for example Ref. [5]. Note also that if the assumption of independence is relaxed, under general conditions the distribution, Eq. (2), is still the appropriate one for maxima. It turns out that almost all standard distributions satisfy the stability postulate asymptotically, although it is only exactly true for the GEV distribution itself. This is exactly analogous to the Central Limit Theorem for averages, which is satisfied asymptotically by almost all standard distributions, but only holds exactly for an initial Normal distribution. As with averages, which are assumed Normal, by the Central Limit Theorem, and then fitted accordingly, so with maxima, it is reasonable to assume a GEV distribution and fit accordingly. Since the dependence of the stability coefficients, a_n, b_n, on n is typically logarithmic, or slower, we can extract maxima from samples which are roughly the same size. In engineering practice this is often almost unverifiable, but nevertheless a plausible assumption, since the bulk of the data, "too small to be seen," may be uncounted, let alone observed. The physical size of components and common conditions may be the only justification.

For extrapolation over larger areas (for extremes derived from random sampling over a large structure) or over longer time periods (for extremes derived from sampling at regular intervals of time), the Nth return level can be defined by solving $F(x)=1-1/N$. Again N is interpreted as in the previous section. Alternatively, after fitting the distribution to the given data, the implied distribution of extreme values from future samples over larger areas and longer lengths of time (with equal base populations) can be deduced and properties such as the mean extreme, etc., inferred from this more fundamental approach. For a full discussion see Ref. [1]. However, the return period method is particularly easy to implement for type I extreme value probability plots. For examples of these plots applied to pit depths in steels exposed to marine environments see Refs. [6,7]. The parameters can also be regressed on covariates as appropriate, to allow for dependence on measured environment variables and/or time, see for example Ref. [8]. A more subtle approach for modeling covariates would use an extreme value regression model of the sort considered in the context of the Weibull distribution [9].

In Ref. [10] each of five circular coupons were exposed to a corrosive medium for each of four different exposure times: 1000 h, 3000 h, 5000 h, and 8000 h. The maximum pit depth was measured in each of six equal sectors on each specimen. Nominally this gave 120 pit depths in all, however, for many coupons, pits overlapped into a number of sectors and so the number of independent maxima was significantly reduced. Figure 3 shows a plot of maximum pit depth against exposure time for resulting data. The plotted mean function and upper bound are based on the fitting of a 4-parameter time dependent GEV distribution for which $\mu_t=\mu t^\beta$, $\psi_t=\psi t^\beta$ and ξ is constant. This model gives

$$\mu_t=0.912(\pm0.063)t^\beta \qquad \psi_t=0.293(\pm0.037)t^\beta$$

$$\beta=0.298(\pm0.051) \qquad \xi=-0.216(\pm0.121).$$

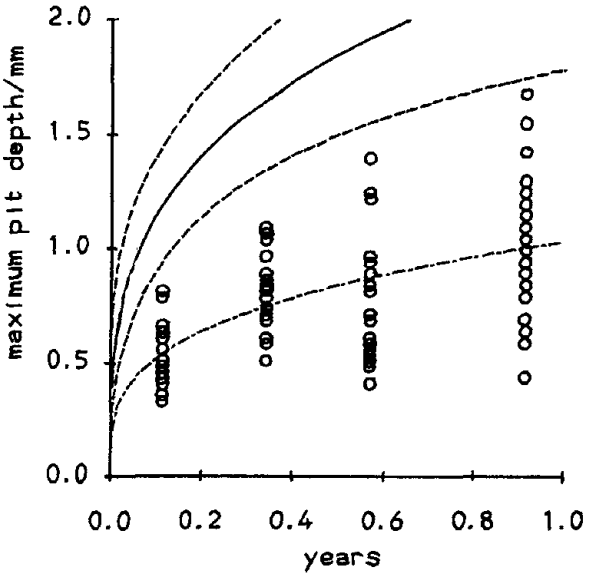


Fig. 3. Maximum pit depths against time for carbon steel in alkaline conditions along with fitted mean function (–·–), upper bound (——) and confidence curves for the upper bound (– – –).

The corresponding mean function is $\eta_t = [\theta + \frac{\psi}{\xi}\Gamma(1-\xi)]t^\beta = \eta t^\beta$, which agrees with the common assumption made in the corrosion literature of a power law growth of the mean maximum pit depth with time [8,11,12]. The implied upper bound is then $\theta_t = \theta t^\beta = (\mu - \psi/\xi)t^\beta$. Such means and bounds can be extrapolated out to larger areas of exposed metal and to longer time periods using the methods described in Ref. [1]. Standard errors on the upper bound were calculated by reparameterizing the problem and constructing a profile likelihood for θ_t as in Ref. [2]. The negative value for the shape parameter ξ has been observed by the authors of this paper consistently for corrosion phenomena of many types and in many environments. This has important consequences for extrapolation since, in corrosion engineering return levels are often very large (e.g., it may only be possible to inspect a small number of one meter sections of a buried pipeline which may be hundreds of kilometers in length), and so for the range of values of ξ encountered by the authors, the maximum will be very close to the upper bound or end point of the distribution. This should be contrasted with the commonly used $\xi = 0$, type I extreme value distribution, [6–8,11] for which there is no upper bound.

4. Extreme Order Statistics

There is a corresponding asymptotic result concerning the joint distribution of the r largest values, $x_{\max} = x_{(1)} \geq \ldots \geq x_{(r)}$, from a sample of independent identically distributed random variables. Data will in general then consist of m sets of such largest values. The joint generalized extreme value distribution (JGEV) has density

$$f(x_1,x_2,\ldots x_r) = \psi^{-r}\exp\{-[1+\frac{\xi}{\psi}(x_r-\mu)]^{-1/\xi}$$

$$-(\frac{1}{\xi}+1)\sum_{j=1}^{r}\log[1+\frac{\xi}{\psi}(x_j-\mu)]\}, \qquad (3)$$

valid for $\xi x_j > \xi\mu - \psi = \xi\theta$, $\psi > 0 (j = 1,\ldots,r)$. See for example Ref. [13]. This is the appropriate distribution to use when the r (say) largest values are extracted from coupons or sampled areas, instead of just the single largest value. This provides a useful extension to the classical theory in such a way as to match up with the common practice of measuring the few largest pits at any one location undergoing pitting. Using all this information rather than just the single largest extreme enables smaller confidence bands to be drawn around predicted values. However care is needed to ensure that r is not taken so large as to invalidate the choice of the asymptotic distribution, Eq. (3).

When $\xi = 0$, this model reduces to the Gumbel form of the JGEV with density

$$f(x_1,x_2,\ldots x_r) = \psi^{-r}\exp\{-\exp[-\frac{1}{\psi}(x_r-\mu)]$$

$$-\sum_{j=1}^{r}\frac{1}{\psi}(x_j-\mu)\}. \qquad (4)$$

A useful diagnostic here is the joint Gumbel plot. When $x_{(1)} \geq \ldots \geq x_{(r)}$ have density, Eq. (4), $E(x_{(i)}) = \mu - \psi\phi(i)$ (all $1 \leq i \leq r$) [14], where $\phi(\cdot)$ is the digamma function. Thus a plot of the order statistics $x_{(i)}$ against $-\phi(i)$ will give a straight line with slope ψ and intercept μ if the Gumbel form of the JGEV distribution is appropriate. Such a plot is shown in Fig. 4 for each of the pitted college roofs data in Table 1. This plot indicates that these extremes arise from perhaps a mixture of two tail distributions. However it was assumed that $\xi = 0$ for both roofs and that for roof 1, the two largest values were to be outliers from the model, Eq. (4). These two values were removed for the purpose of analysis, and the slopes and intercepts resulting

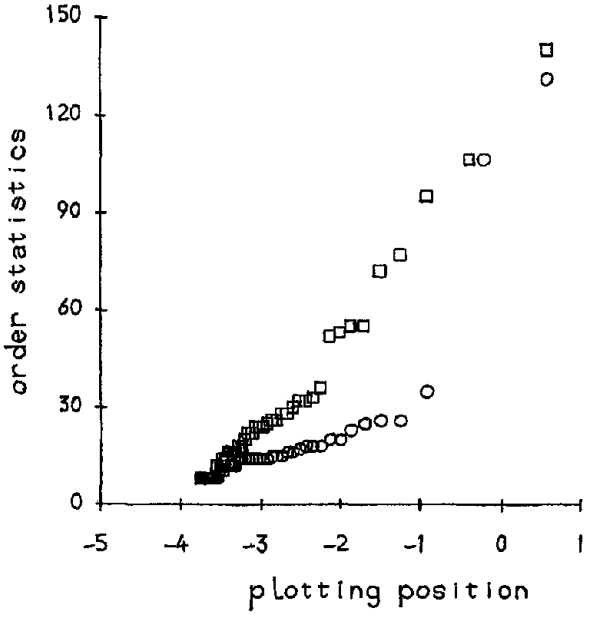


Fig. 4. Joint Gumbel plot for the college roof data: O–roof 1; □–roof 2.

used as starting values for determining the maximum likelihood estimates of the parameters in Eq. (4). The fitted values, with their standard errors, were

$$\mu = 54.2\ (\pm 7.9) \qquad \psi = 12.5\ (\pm 2.1),\ \text{roof 1},$$

$$\mu = 103.2\ (\pm 15.8) \qquad \psi = 26.0\ (\pm 4.2),\ \text{roof 2}.$$

These values are then available for the implied Gumbel distribution of the maximum value, which has mean $\mu + 0.5772\psi$. This gives 61.4 μm for roof 1 and 118.2 μm for roof 2. Extrapolation could now proceed according to the method described in the previous section, noting however that the mean of the maximum for roof 1 is considerably out of line with the observed maximum of 131 μm.

Reference [15] reports on an experiment where 15 low alloy steel specimens were suspended in a deionized warm water bath under free corrosion conditions. Specimens were removed at varying intervals up to 71 days, then after cleaning, pit depths and diameters were measured optically. A 4-parameter JGEV distribution incorporating a power law dependence on time [16] was fitted to these pit-depths, utilizing the two largest pits from each side of the specimens giving parameter values:

$$\mu = 7.041(\pm 0.710)t^{\beta} \qquad \psi = 0.467(\pm 0.066)t^{\beta},$$

$$\beta = 0.609(\pm 0.016) \qquad \xi = -0.513(\pm 0.126).$$

These are the maximum likelihood estimates for their data, for which they were only, at that time, able to report initial probability weighted moment and regression estimates. Figure 5 shows a plot of this data along with the fitted mean function and upper bound, and confidence curves for the upper bound calculated using the profile likelihood method discussed in the previous section.

5. Discussion

A number of statistical techniques relating to extreme value theory have been described and demonstrated on selected sets of corrosion data. Noting that much corrosion data are inherently of an extreme nature, purely statistical considerations along the lines described in this paper may be the only means of determining numerical values for prediction of the maximum pit depth in an area A at time t, for example, along with some estimate of precision or possible error. There is much evidence in the literature that $\xi < 0$ for the GEV distribution in the context of extremes of corrosion phenomena. Return levels are often very large and so, for the range of values of ξ encountered, predicted maxima will often be very close to the implied upper bound or end point of the distribution.

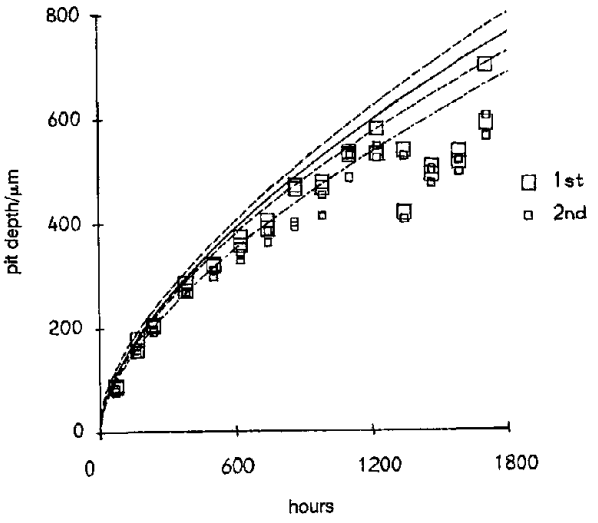


Fig. 5. First and second largest pit depths against time for low alloy steel in deionized warm water, along with fitted mean function (–·–), upper bound (——) and confidence curves for the upper bound (– – –).

It should be noted however, that with all the methods described here, there are pitfalls. When modeling exceedances, for example, it is difficult to choose the threshold objectively, and different thresholds can lead to different predictions. Similar problems exist in the use of the r largest order statistics and also the maximum itself. How many largest order statistics should be used? When recording single maxima, how large should the sampled area be? While some theoretical results are available to answer such questions (e.g., Ref. [17]) these are not very helpful in a practical context.

Acknowledgments

The authors would like to thank the following: Dave Farrell of CAPCIS Ltd., U.K., for the college roof data; Dave Eden of CAPCIS-MARCH Ltd., U.K., for the electrochemical noise data; The National Institute of Standards and Technology (Grant No. 60NANB5D0519) for the basalt groundwater data.

6. References

[1] P. J. Laycock, R. A. Cottis, and P. A. Scarf, Extrapolation of extreme pit depths in space and time, J. Electrochem. Soc. **137**, 64–69 (1990).
[2] A. C. Davison and R. L. Smith, Models for exceedances over high thresholds (with discussion), J. R. Statist. Soc. B **52**, 393–442 (1990).
[3] C. Gabrielli, F. Huet, and M. Keddam, Investigation of metallic corrosion by electrochemical noise techniques, In Electrochemical and Optical Techniques for the Study and Monitoring of Metallic Corrosion, M. G. S. Ferrara and C. A. Melendres, eds., Kluwer, Dordrecht (1991).
[4] R. A. Cottis, P. J. Laycock, and P. A. Scarf, Statistics of Pitting, Quarterly Report, May 1990, NBS/NIST Contract 60NANB5D0519.
[5] M. R. Leadbetter, G. Lindgren, and H. Rootzen, Extremes and Related Properties of Random Sequences and Processes, Springer-Verlag, New York (1983).
[6] F. Blekkenhorst, G. M. Ferrari, C. J. van der Wekken, and F. P. Ijsseling, Development of high strength low alloy steels for marine applications, part 1, Brit. Corros. J. **21**, 163–176 (1986).
[7] F. Blekkenhorst, G. M. Ferrari, C. J. van der Wekken, and F. P. Ijsseling, Development of high strength low alloy steels for marine applications, part 2, Brit. Corros. J. **23**, 165–171 (1988).
[8] G. P. Marsh and K. J. Taylor, An assessment of carbon steel containers for radioactive waste disposal, Corrosion Sci. **28**, 289–320 (1988).
[9] M. J. Crowder, A. C. Kimber, R. L. Smith, and T. L. Sweeting, Statistical Analysis of Reliability Data, Chapman and Hall, London (1991).
[10] B. Daniel, Corrosion Pitting of Carbon Steel in Alkaline Conditions, Ph.D. Thesis, UKAEA, Harwell, U.K (1989).
[11] K. Masamura and I. Matsushima, Statistical characteristics of pit depths observed on inner surface of carbon steel pipes, Zairyo **46**, 47–51 (1987).
[12] M. Romanoff, Underground corrosion, Natl. Bur. Stand. (U.S.) Circular No. 597 (1957).
[13] R. L. Smith, Extreme value theory based on the r largest annual events, J. Hydrol. **86**, 27–43 (1988).
[14] I. Weissman, Estimation of parameters and large quantiles based on the k largest observations, J. Amer. Stat. Assoc. **73**, 812–815 (1979).
[15] D. C. Buxton, R. A. Cottis, and P. A. Scarf, Life prediction in corrosion fatigue, In proceedings of Conference on Life prediction of corrodible structures, NACE, Kanai, November 1991.
[16] P. A. Scarf, R. A. Cottis, and P. J. Laycock, Extrapolation of extreme pit depths in space and time using the r deepest pit depths, J. Electromchem. Soc. **139**, 2621–2627 (1992).
[17] R. L. Smith, Estimating tails of probability distributions, Ann. Statist. **15**, 1174–1207 (1987).

About the authors: *Philip A. Scarf is a lecturer in Statistics at the University of Salford, U.K. Patrick J. Laycock is a senior lecturer in Statistics at the University of Manchester Institute of Science and Technology (UMIST), U.K. Both authors have recently worked on a project entitled "The Statistics of Pitting" funded by NIST.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 321 (1994)]

On the Requirements for a Reasonable Extreme Value Prediction of Maximum Pits on Hot-Water-Supply Copper Tubing

Volume 99 Number 4 July–August 1994

Shigeru Komukai and Komei Kasahara

Tokyo Gas Company, Ltd.,
Fundamental Technology
Research Laboratory,
1-16-25 Shibaura, Minato-ku,
Tokyo 105, Japan

Application of extreme value statistics to the problem of Type-II pits growth prediction on hot-water-supply copper tubing is described. A recommendation is suggested for optimum combinations of the number and the size of unit samples required for reasonable extreme value predictions.

Key words: central hot-water-supply; copper tube; extreme value statistical analysis; maximum pit depth prediction; pitting corrosion.

Accepted: March 22, 1994

1. Introduction

Modern large buildings in the Tokyo metropolitan area usually have a centralized hot-water supply operating 24 h a day. Copper pipe is widely used in such systems because of its relatively high resistance to corrosion, coupled with additional pragmatic merits: it is easy to work with, easy to install, and relatively cheap.

The seriousness of Type-II pitting corrosion, however, has increasingly received high recognition in such hot-water supply systems [1,2,3]. The need to obtain information regarding the degree of pitting corrosion has increased over the last decade because considerable pipe damage may require maintenance, and even replacement, and in that case, proper life prediction is essential to pass reasonable engineering judgements and thereby to perform proper maintenance.

The life prediction of such copper plumbing tubing can first be performed by coupling adequate nondestructive and/or, though less favorable, destructive inspection techniques with reliable statistical analysis. It has been shown that the most promising statistical analysis methods for such a purpose include extreme value statistical analysis [3,4].

Although such tools have become widespread, a general method for evaluating the localized corrosion propensity on existing engineering structures from limited inspection data, and concrete criterion for the number and size of samples required to obtain a reasonable extreme value prediction is still not available.

In the present study, a set of pitting corrosion depth data obtained from 7 year old copper plumbing pipe, one third of which was removed from a centralized hot-water supply system, was examined by extreme value statistics. Emphasis was placed on the effect of the total number and size of maximum pit depths on the accuracy of pit growth prediction. A concept was ultimately proposed to obtain a reasonable prediction of maximum pit depths while minimizing the total sampling area (or, length) for analysis.

2. Test Procedures

2.1 Test Specimens

Copper pipes totaling 8.88 m in length were removed from various parts of the centralized hot-water supply system in an 11-story multi-family dwelling in Tokyo. The system consisted of one stainless steel storage tank (1.8 m^3) and copper plumbing pipe having an overall length of about 28 m. The plumbing material was JIS C 1220T type 25AM (outside diameter 28.58 mm and wall thickness 0.89 mm) copper pipe for building use. The system had been operating for about 2600 d before test piping was removed. In the system, water at a nominal temperature of 60 °C circulated constantly. Average flow (i.e., hot-water consumption) was around 8 m^3/d, the storage tank being supplied automatically with tap water.

The copper piping removed was cut into parts 100 mm long, which were then cut in half to give half-ring specimens. Each half-ring specimen was then completely cleaned ultrasonically in dilute sulfuric acid, followed by marking-off to divide it into 10 virtual half-ring specimens of 10 mm unit length, after which the pit depths were measured by using an optical microscope of 1 μm precision. By coupling two opposite virtual half-ring specimens, a 10 mm long full ring specimen was reassembled, and the area thus surveyed should be representative of the pipe at that particular location. Then, by taking several adjoining full-rings of 10 mm length, for each unit sample sizes of maximum pit depths (hereinafter, s: unit length) in the interval 20 mm up to 200 mm were obtained.

2.2 Extreme Value Statistical Analysis

The extreme value statistical analysis was performed by using a commercial available personal computer software package, EVANS [5].

The basic concept of the present extreme value analysis is briefly reviewed in the following sections.

2.2.1 Extreme Value Probability Plots The first step of the extreme value analysis included the preparation of extreme value probability plots, that is, plots of maximum pit depth data on extreme value probability paper of the cumulative relative frequency ($F(y)$) vs maximum pit depth (x). Maximum pit depths data were arranged in order from largest to smallest and assigned a rank number. The vertical plotting position $F(y)$ for each pit depth value was calculated by the averaged rank method as follows:

$$F(y) = 1 - i/(N+1) \quad (1)$$

where i = rank number,
N = total number.

Several sets of plots were obtained depending on the combinations of N and s. In an exemplifying case of $s = 100$ mm and $N = 7$, a total of 10 sets of plots was subjected to regression analysis.

2.2.2 Regression for the Best Fit Line Regression analysis based on the MVLUE (Minimum Variance Linear Unbiased Estimator) method was made for each data set to determine a straight line of best fit to the plotted extreme values. The equation of a straight line is given as follows:

$$x = \lambda + \alpha y \quad (2)$$

where x = expected maximum pit depth,
y = standardized variable,
λ = location parameter,
α = scale parameter.

The straight line is drawn in Fig. 1 on extreme value probability paper. In the present analysis, emphasis was placed on the optimum combinations of N and s for obtaining a reasonable estimate of the extreme value. Such an analysis becomes feasible through a thorough investigation of all the samples where the depth of the actually detected deepest pit could be regarded as the probable maximum pit depth, that is, the extreme value.

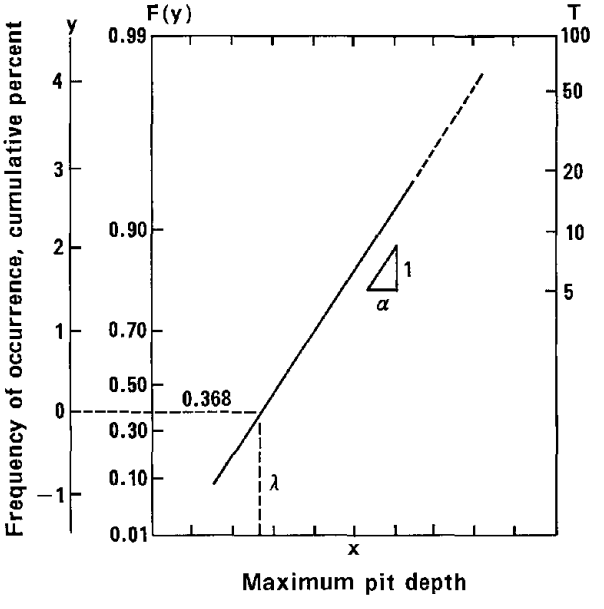


Fig. 1. Extreme value probability paper of double-exponential distribution.

3. Results and Discussion

3.1 Distribution of Pit Depths

Figure 2 is a histogram showing the relationship between the pit depths and the total frequency. In the figure, a total of 970 depth data for all pits found in one 10 mm long full-ring, removed from the 7th floor in the building, was grouped over the pit depth ranges, 0 mm–0.019 mm, 0.020 mm–0.039 mm, etc. As is evident from the figure, the shape of the pit depth distribution is a bell-shaped curve starting at zero, rising to a maximum at around 0.05 mm and thereafter decreasing rapidly with increasing pit depths. The lack of "J"-shaped portion bending to the right in the pit depth range from 0 mm to 0.02 mm, together with the trailing extreme portion of the tail of the curve up to 0.34 mm, indicated that most of the small pits had already had ceased to grow while only a small number of deeper pits continued to grow.

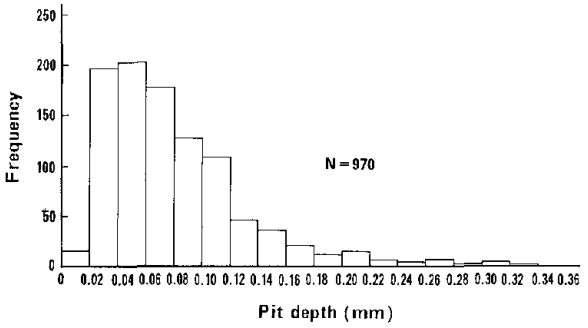


Fig. 2. Pit depths distribution histogram for 10 mm long copper pipe used for 2600 d in a hot-water supply system operated at 60 °C.

In Fig. 3, all the pit depth data represented in Fig. 2 were plotted on a logarithmic-normal distribution diagram. The apparent linearity of the plot indicates that this distribution applies to the logarithmic-normal; and hence the maximum pit depth data obtained in the present pit depth survey were sampled from the parent populations with a logarithmic-normal distribution.

Figure 4 shows the distribution at every floor of the maximum pit depth detected in each 100 mm unit length. Though the maximum pit depths seemed to have a slight tendency to become shallower at upper floors, the distribution of pit depths was regarded as being uniform throughout the building. The actual maximum pit depth value in the present survey was 0.452 mm which was detected at 6th floor.

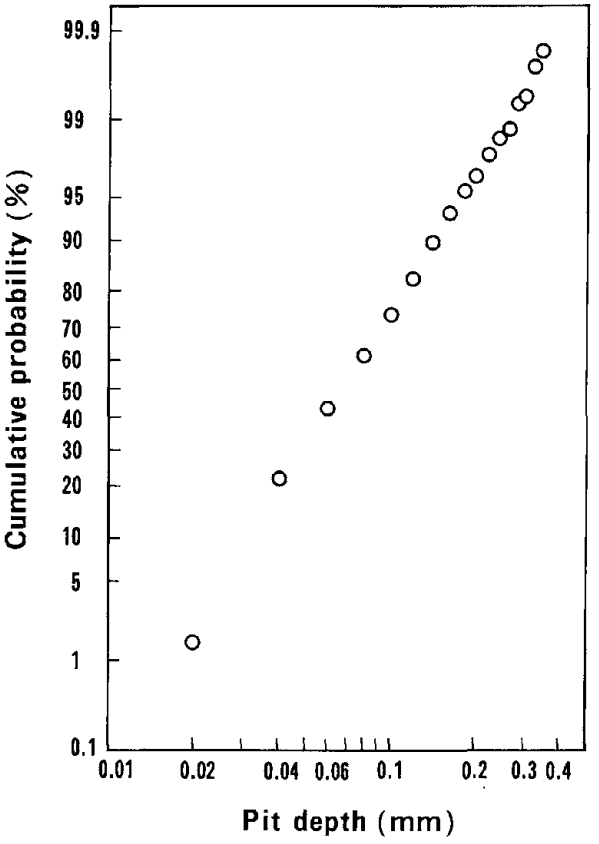


Fig. 3. Plot on a logarithmic-normal distribution diagram of all data presented in Fig. 2.

In Figure 5, the maximum pit depth data for unit lenghs of 20 mm, 100 mm, and 200 mm were evaluated from extreme value analysis on the basis of Gumbel's double-exponential distribution; the linearity of each plot shows that this distribution applies to the maximum pit depth data obtained at the unit lengths between 20 mm and 200 mm.

3.2 Minimum Required N and s

In the practical application of extreme value statistics, the number and the size of unit samples for the pit depths survey are to be decided prior to the destructive (or nondestructive) inspection of the existing structures.

Reliability of the extreme value prediction depends on the following three equations:

$$y = \ln T \tag{3}$$

$$T = S/s \tag{4}$$

$$V(x) = \alpha^2[A(N,n)y^2 + B(N,n)y + C(N,n)] \tag{5}$$

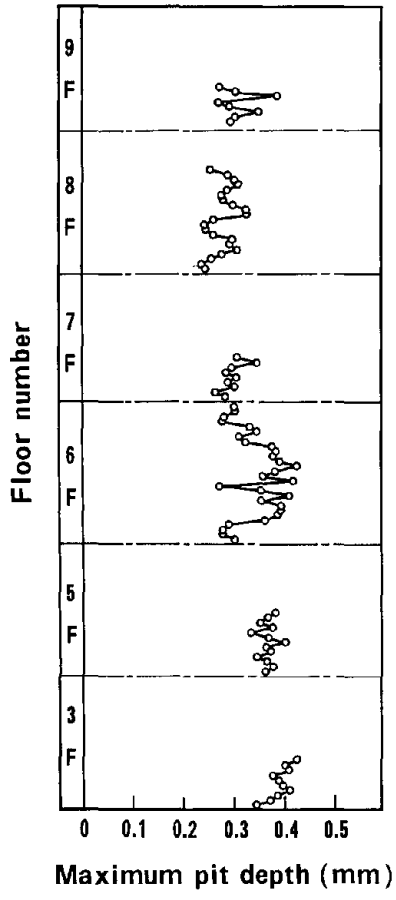


Fig. 4. Distribution at each floor in the building of the maximum pit depth detected in each 100 mm unit length.

where y = standardized variable,
T = return period,
S = total length,
s = unit length,
$V(x)$ = variance of the estimated extreme value,
$A(N,n)$, $B(N,n)$, $C(N,n)$ = MVLUE coefficients,
α = scale parameter.

It is obvious from these equations that for increased reliability, hence for a minimized $V(x)$, it is required to increase s (therefore to decrease T) and/or to increase N. In the interest of economy, however, there must be a natural limitation to the increase of N and s for the extreme value survey.

In theory, the optimum combinations of N and s can be determined from Eqs. (3) to (5) once the distribution parameter (that is, a ratio of the location parameter to the scale parameter, α/λ) is reasonably assumed (that is, empirically or experimentally), and the extent of the standard deviation of the error of extreme value estimates, σ, may be expected to be:

$$\lambda = m\sigma \tag{6}$$

where λ = location parameter,
σ = standard deviation of the estimated extreme value,
m = assumed number (1, 2, 3, etc.)

That is, the decision may be made to reduce σ to 1/m of λ [6].

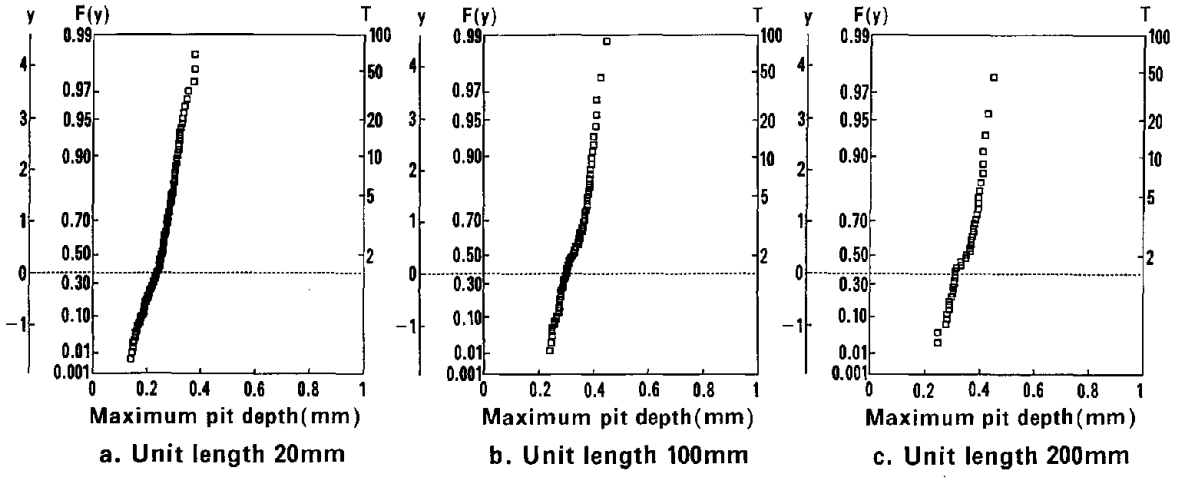


Fig. 5. Double-exponential distribution type of extreme value probability plots for $s = 20$ mm, 100 mm, and 200 mm.

Table 1 shows the result of extreme value analysis to obtain the relationship between unit length and distribution parameter. In the analysis, all pit depth data obtained from the whole lengths were brought into consideration. It can be seen in the table that the distribution parameter, α/λ, decreased with increasing unit length, ultimately approaching to a definite level of 0.15.

Table 1. Location, scale, and distribution parameters determined from all maximum pit depth data obtained in the survey

Unit length (mm)	T	N	x_{max} (mm)	λ (mm)	α (mm)	α/λ
50	177.6	172	0.575	0.28415	0.05621	0.198
100	88.8	86	0.528	0.30446	0.04986	0.16
200	44.4	43	0.506	0.3221	0.04866	0.151

Optimum combinations of N and T (hence, s) required for controlling σ down to $1/m$ of λ were obtained as shown in Fig. 6 corresponding to α/λ of 0.15 and 0.20. It stands to reason, that the number of unit lengths, N, can markedly be reduced by loosening the requirement for the reliability, that is, by decreasing m. N may also be reduced by decreasing α/λ. Though N may be reduced by increasing s (hence decreasing T), it should be noted that the total length required for a pit depths survey can increase on the contrary.

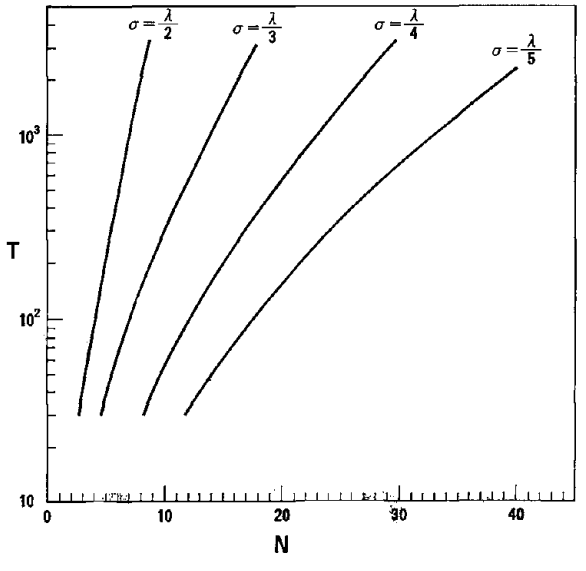


Fig. 6. Optimum combinations of N and T to control σ at levels of 1/2, 1/3, 1/4, or 1/5 of λ under the limitation of $\alpha/\lambda = 0.2$.

3.3 Maximum Pit Depths Prediction Based on the Optimized N and s

Figure 7 shows the results of the maximum pit depths prediction by using the combinations of N and s from Fig. 6. Unit lengths for this evaluation were not sampled at random locations, but were ordered from one end of the lower floors upward. Thus the results were represented by correlating to the locations wherefrom those unit lengths used in the prediction were removed. It can be seen that the scatter of estimated extreme values decreased with increasing s. As would be expected, the scatter was narrower at $m = 5$ as compared with $m = 3$.

To determine the general tendency in the reliability of the present extreme value prediction, Table 2 was developed from data on Fig. 7 by taking an average for each item. The results indicate the following:

1) The assumptions for α/λ and m, that were made prior to the analysis, have conservatively been met.
2) Increased m did not always result in increased reliability which indicated that $m = 3$ might be reasonable in the interest of economy.
3) Except for the cases of $s = 10$ mm, the maximum detected pit depth value of 0.452 mm fell closely between the average of estimates $\pm 1\sigma$.
4) Although the level of s is to be kept as low as possible, because the total sampling length (ie., $N \times s$) may increase with increasing s, at least 2.5% of the whole structure is to be subjected to extreme value evaluation at a return period of 500 or the less.

4. Concluding Remarks

Based on this analysis, the following conclusions can be drawn concerning the optimum conditions for obtaining a reasonable extreme value prediction.

1) The number and the size of unit samples for the extreme values survey may be determined so that the variance of the extreme value estimates is to be minimized under a definite distribution parameter, wherein the standard deviation of the estimates is expected to be 1/3 of the mode of distribution.
2) A return period of 500 with a total sampling length amounting to 2.5% of the entire parts may be desirable for an increased reliability of extreme value prediction.

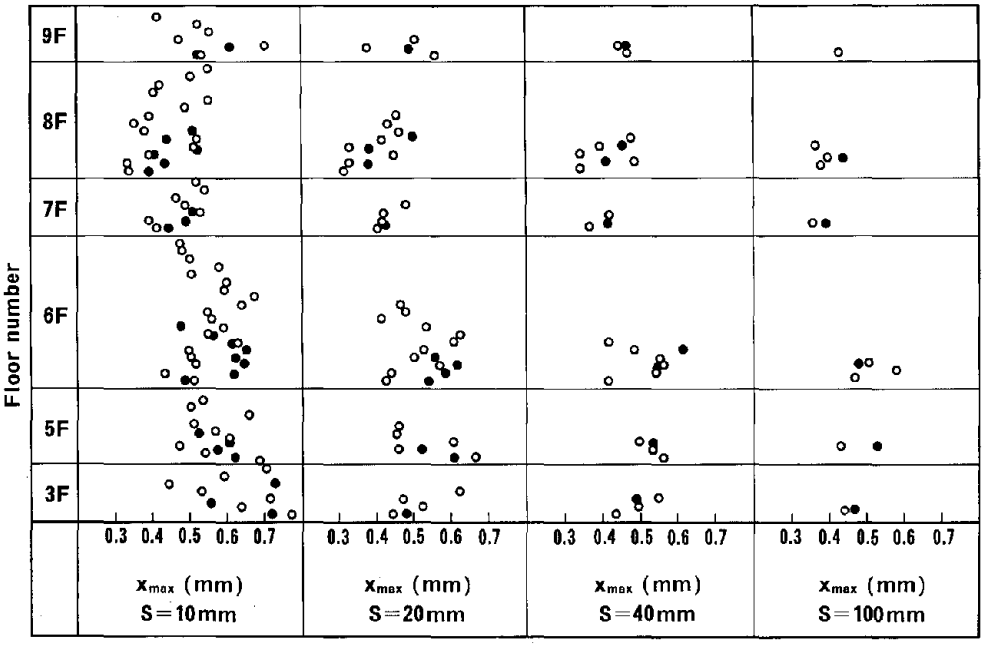


Fig. 7. Distribution at each floor of the estimates of maximum pit depths on the basis of the optimum combinations of N and T under the limitations of $\lambda = 3\sigma$ and 5σ.

Table 2. Summary of the extreme value prediction

	Unit length S (mm)	Return period T	Sample number N	Location parameter λ_{av}	Scale parameter α_{av}	Estimated maximum pit depth $\tilde{x}_{max}$	$\bar{\alpha}_{max}$	$\frac{\lambda_{av}}{\bar{\alpha}_{max}}$
$\lambda = 3\sigma$	10	888	14	0.241	0.043	0.534	0.069	3.50
	20	444	12	0.269	0.034	0.478	0.054	4.98
	40	222	9	0.288	0.035	0.467	0.058	4.98
	100	88.8	7	0.313	0.028	0.428	0.045	6.94
$\lambda = 5\sigma$	10	888	33	0.218	0.046	0.559	0.047	4.63
	20	444	28	0.266	0.040	0.491	0.040	6.67
	40	222	23	0.282	0.039	0.487	0.039	7.25
	100	88.8	18	0.312	0.034	0.463	0.032	9.71

5. References

[1] H. Baba, T. Kodama, and T. Fujii, Boshoku Gijutsu (Corr. Eng.) **34**, 10 (1985).
[2] K. Kasahara and S. Komukai, Repeat **36**, 492 (1987).
[3] K. Kasahara, S. Komukai, and T. Fujiwara, Repeat **37**, 361 (1988).
[4] K. Kasahara and S. Komukai: Proc. 9th European Congress on Corrosion, Vol. 1, FU-037, Utrecht, Holland (1989).
[5] Japan Society of Corrosion Engineering Ed.: EVANS-Software Program Package for Life Prediction by the Extreme Value Analysis (1989) [MARUZEN Press, Tokyo].
[6] Proc. Japan Society of Corrosion Engineering Meeting on Extreme Value Analysis (1986).

About the authors: Shigeru Komukai is the lead researcher of the engineering materials team and Komei Kasahara is general manager of the Fundamental Technology Research Laboratory of Tokyo Gas Co., Ltd.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 327 (1994)]

Application of Extreme Value Statistics to Corrosion

Volume 99 Number 4 July–August 1994

Toshio Shibata

Department of Materials Science and Processing,
Faculty of Engineering,
Osaka University,
2-1 Yamadaoka, Suita 565
JAPAN

Application of extreme value statistics to corrosion is reviewed. It is emphasized that the concept of corrosion probability is important for a quantitative evaluation of corrosion failure and its prediction, especially for localized corrosion. Extreme value statistics is quite useful for assessing the maximum pit depth and/or the minimum time for crack generation. The maximum pit depth depending on the surface area can be evaluated by using the Gumbel distribution with the concept of return period. A standardized procedure is proposed for estimating the maximum pit depth. Examples of corrosion failure analysis using extreme value statistics, which were reported mainly in Japan, are shown. Accumulated experiences suggest that an appropriate classification of data based on the corrosion mechanism is required before applying extreme value analysis.

Key words: corrosion probability; failure time; Gumbel distribution; localized corrosion; maximum pit depth; MVLUE method; return period; Weibull distribution;

Accepted: **March 22, 1994**

1. Introduction

Development of extreme value statistics and its application to various fields, including corrosion, has been described by Gumbel in Ref. [1] and [2]. Evans is one of the pioneers of modern corrosion science, and first established the concept of corrosion probability [3, 4]. Eldredge [5] used extreme value statistics to obtain the maximum value of pit depth on an oil well tube wall as a function of tube surface area. Scott [6] found a logarithmic dependence of the maximum pit depth on surface area, and explained that dependence by referring to Trippet [7]. Aziz [8] and Eledredge [5] discussed almost all important points to be considered for the analysis of corrosion pit data and made use of the return period concept. This concept, originally introduced in the fields of hydrology or meteorology, is now used to obtain a size factor which makes it possible to estimate the maximum pit depth in a large surface area based on the distribution of a small number of pit depth data from the small surface area samples. In Japan, early review papers on corrosion probability and extreme value statistics by Masuko [9] and Shibata [10, 11] contributed to the study of the extreme value statistics as applied to corrosion problems. Ishikawa [12,13,14] and Imagawa [15,16,17] applied extreme value statistics to analyze engineering data. Kase [18, 19] reviewed Lieblein's paper [20], introducing MVLUE (minimum variance linear unbiased estimator) method for estimating the distribution parameters. Lieblein had given the coefficient of MVLUE up to $N=6$. Recently, Tsuge [21] had calculated the coefficients up to $N=45$, and confirmed that the parameters estimated by the MVLUE method are unbiased and efficient and are consistent with values estimated by the method of moments or maximum likelihood when the sample size exceeds more than 20. The committee of Japan Society of Corrosion Engineering (JSCE) proposed a standard procedure [22] to estimate the maximum pit depth from the small

sample size data by using the MVLUE method, and a computer program named EVAN [23] was developed. Recently Laycock et al. [24] reported that a generalized extreme value distribution is more convenient for corrosion depth analysis, because no preliminary assumption on the type of distribution is needed. An introductory book [25] by Kowaka et al. helped to differentiate extreme value analysis among corrosion workers in Japan. General background on extreme value statistics is provided by Ang and Tang [26], and Kinnison [27].

2. Application of the Extreme Value Analysis to Corrosion

In the early 1980s, meetings and symposia [28,29,30,31] were held in Japan for discussing the basic principles of extreme value statistics as well as difficulties and problems in their application to corrosion. In Table 1, several topics for which the Gumbel distribution is applied are listed. Table 2 includes cases analyzed using the Weibull distribution, including the exponential distribution. Before discussing case histories, the standard procedure [22] proposed by the committee is briefly explained; details are available elsewhere [22, 32].

Table 1. Examples of the extreme value analysis for corrosion using the Gumbel distribution

Example	Ref.
1. Life prediction of super heater tubes of the power plant	[49]
2. Application of the extreme value analysis to heating tubes of the boiler	[37]
3. Estimation of the maximum amount of impurity segregation in steel	[42]
4. Failure life estimation of SCC for Ni base alloys	[41]
5. Extreme value analysis of the corrosion depth of the oil tank plate	[35] [36] [50]
6. Life prediction of heat exchanger tubes	[51]
7. Eddy current examination system for heat exchanger tubes with the extreme value analysis	[43]
8. Extreme value analysis of pitting corrosion of heat exchanger tubes	[52]
9. Methods for the parameter estimation of the pit distribution in plants	[16]
10. Ultrasonic method for heat exchanger tubes with the extreme value analysis	[44]
11. Maintenance system for coated heat exchanger tubes	[45]
12. Corrosion of steels in sea water	[53]
13. Analysis of perforation of zinc plating steels by extreme value statistics	[54]
14. Fatigue crack behavior of high strength steel in artificial sea water	[55]

Table 2. Examples of the extreme value analysis for corrosion using the Weibull (exponential) distribution

Example	Ref.
1. Failure life analysis of stress corrosion cracking of stainless steel heat exchanger tubes	[56]
2. Failure life distribution of stainless steels in high temperature and high pressure water	[57]
3. Effect of $CaCl_2$ concentration on SCC life time distributions of stainless steels	[58]
4. Evaluation of SCC failure life of stainless steel in high temperature water	[59]

2.1 The Gumbel Distribution

The procedure is proposed mainly for analyzing pit depth distribution by using the Gumbel distribution and the return period in order to estimate the maximum depth of the larger surface area from which small area specimens are sampled. The Gumbel distribution is expressed as

$$F(x) = \exp(-\exp(-(x-\lambda)/\alpha)), \quad (1)$$

where $F(x)$ is the cumulative probability of pit depth, x, and λ and α are the location and scale parameters. The reduced variate, y,

$$y = (x-\lambda)/\alpha \quad (2)$$

is introduced, and then

$$y = -\ln(-\ln(F(y))) \quad (3)$$

is used for constructing the Gumbel probability paper. Plotting position for the cumulative probability can be calculated simply by

$$F(y) = 1 - i/(1+N), \quad (4)$$

where i is the ith of the ordered value, x, in descending order and N is the total number of sample. Plotting y as a function of x yields a best-fitting straight line; its slope provides $1/\alpha$ and its intercept (at $y=0$) yields λ. Instead of this graphical estimation of the parameters, more reliable estimates of α and λ can be obtained by using the MVLUE (minimum variance unbiased estimator) method, the maximum likelihood and the method of moments. Among them the MVLUE method which is discussed by Lieblein [20] is found to be more efficient and unbiased for small size samples. The MVLUE estimator can be calculated by

$$\lambda = \Sigma a_i\,(N, n)x_i$$

$$\alpha = \Sigma b_i\,(N, n)x_i, \qquad (5)$$

where $a_i\,(N, n)$ and $b_i\,(N, n)$ are weights for each sample depending on the sample size, N, and truncated number, n, which are tabulated in the table given by Tsuge [21, 23] up to $N=45$. The weights, A, B, C, of variance, V,

$$V = \alpha^2\,(A\,(N,n)y^2 + B\,(N,n)y + C\,(N,n)) \qquad (6)$$

are also found in the table given by Tsuge [21, 23]. For the pit depth distribution, the return period, T, is defined as

$$T = S/s, \qquad (7)$$

where S is the surface object (e.g., a tank plate) to be examined and s is the area of the small specimens which are sampled randomly from the objective. The return period, T, is in effect a size factor. The mode, λ, of the pit depth distribution for the small specimen is simply obtained by the MVLUE estimators mentioned above, and the mode for the T times larger surface, x_{max}, is given by

$$x_{max} = \lambda + \alpha \ln\,(T). \qquad (8)$$

The perforation probability, P, of the maximum pit through the wall thickness, d, is given by

$$P = 1 - \exp\,(-\exp(-\,(d - (\lambda + \alpha \ln\,(T)/\alpha)). \qquad (9)$$

Finally, the procedure [22] requires reporting the surface area of the object, S, with the small sample area, s, providing the return period, $T\,(=S/s)$, and the number of samples, N, with data number, n, actually obtained. In addition, the original thickness of the plate, d, and the perforation probability, P, if needed are to be stated. The above procedure does not request to check a goodness of fit of the distribution obtained to the Gumbel distribution, but recommends to examine the fitness by the Kolmogorov-Smirnov or chi-square test if needed.

2.2 The Weibull Distribution

The third type for the smallest value called the Weibull distribution

$$F\,(t) = 1 - \exp\,(-((t-\gamma)/\eta)^m) \qquad (10)$$

can be fitted to the failure life distribution of stress corrosion cracking [33, 34] as shown in Table 2, where γ, η and m are the location, scale and shape parameter, respectively. This third type of asymptotic distribution for the smallest value can be transformed to the first type for the largest value, that is, Eq. (1), by changing $1-F(t)$ to $F(z)$ and by introducing the following reduced variate

$$X = \ln(t-\gamma),\ z = (X-\lambda)/\alpha. \qquad (11)$$

The same MVLUE method used for Eq. (1) can be utilized for parameter estimation, because the following relations exist between the parameters of both distributions;

$$\lambda = \ln(\eta),\ \alpha = 1/m. \qquad (12)$$

The above unified procedure for estimating parameters of the Gumbel and Weibull distribution was coded in the computer program EVAN [23].

3. Examples

Several examples are provided to demonstrate the usefulness of extreme value statistics for analyzing corrosion problems.

3.1 Maximum Pit Depth of Oil Tank Plate

Through the 1960s and 1970s a number of oil tanks were built in Japan. In the late 1970s there occurred frequent oil leakages from tanks due to corrosion failure. Oil refinery or petrochemical industries were located along the seacoast and oil leakage caused serious environmental damages. In 1976, the fire service law was revised to enforce inspection of the thickness of the base and annular plates of oil tanks every time oil was evacuated. On these occasions extreme value analysis was applied and found to be a powerful tool for estimating the maximum pit depth. It is emphasized that data for the base plate and the annular plate should be considered separately because they are characterized by different corrosion damage and mechanisms.

The law requests that plate thickness has to be measured at the corners of every 10 cm square on the whole surface of the plate. This inspection procedure contributed greatly to reducing corrosion leakage, but was time-consuming and costly. The extreme value analysis was then studied intensively in this field [35] [36]. Pit depth distribution sampled from the whole base plate was found to obey the Poisson distribution.

Araki et al. [36] found that the largest value from each small square ($s = 1\ m^2$) being randomly sampled obeys the Gumbel distribution as shown in Fig. 1 [36]. The slope and intersect of the line ($\alpha = 0.694$ and $\lambda = 1.41$) were estimated by the MVLUE estimators of Eq. (5). In this case, the surface area of the base plate, S, was 1535 m^2 and the return period or size factor was calculated to be $T = S/s = 1535$. The maximum depth, x_{max}, was calculated by Eq. (8):

$$x_{max} = 1.41 + 0.694 \times \ln\ (1535) = 6.50\ \text{mm}$$

which is shown also in Fig. 1. These data were obtained for a base plate which was exposed for 12.6 years. Data for both annular and base plate exposed for 7.7 years were plotted in Fig. 2 [36], from which the first leak due to the maximum pit is to be expected after 17.6 years for the base plate and 23.5 years for the annular plate, respectively. The effect of N, s, on the estimates was examined and it was concluded that the MVLUE method is optimal for $N < 20$ and the maximum likelihood method is reliable for $N > 20$.

3.2 Rupture of Heat Exchanger Tubes of the Boiler

Super heater and economizer tubes of boilers are exposed to high temperature gases with salt deposits which cause severe corrosion attack. Corrosion attack is not uniform, but localized at several sites, wall thinning at the localized site resulting in burst. Regular inspection is needed to predict time for replacement of the tube before burst. All tubes have to be examined for predicting exact time with high confidence, but cost of inspection being high, that Fukuda et al. [37] introduced the use of extreme value analysis to supplement the inspection of a small number of tubes. In Fig. 3, the largest values of wall thinning observed for 14 tubes are plotted on Gumbel probability paper. The distribution of wall thinning at every inspection time is seen to obey the Gumbel distribution and the maximum thickness determined by the return period (40 tubes) increases with operation time as shown in Fig. 4. A criterion for a proper replacement time has been proposed, which requires replacement when the wall thickness reaches half of the design thickness, t_{sy}. Risk of burst could be avoided by estimating the depth and noting the proposed criterion.

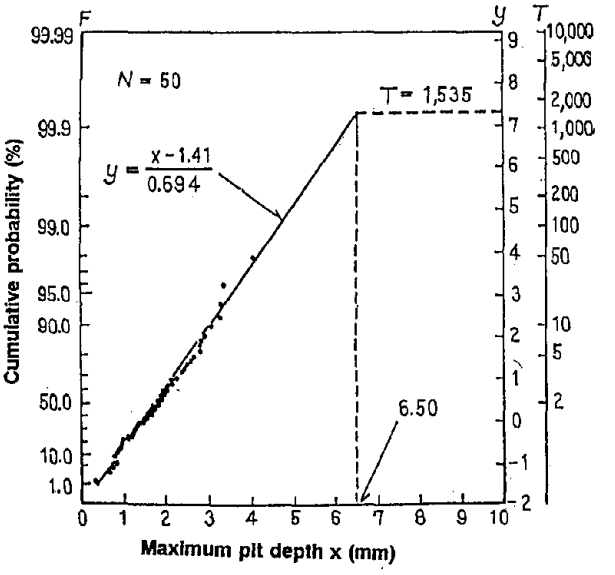


Fig. 1. The Gumbel plot of the maximum pits on the bottom plate of the oil tank.

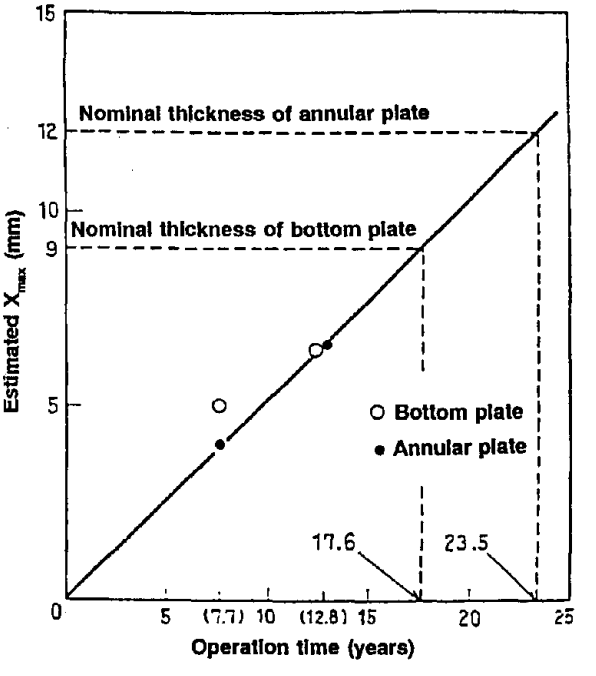


Fig. 2. Estimated depth of the maximum pit on the whole surface as a function of operation years, and prediction of failure life for the bottom and annular plate.

3.3 The Pit Depth Distribution of Steel Piles in Sea Water

Since the 1970s, steel pipes and piles have been used extensively in Japan for harbor construction, because lead time for construction could be reduced compared with using concrete. Recently,

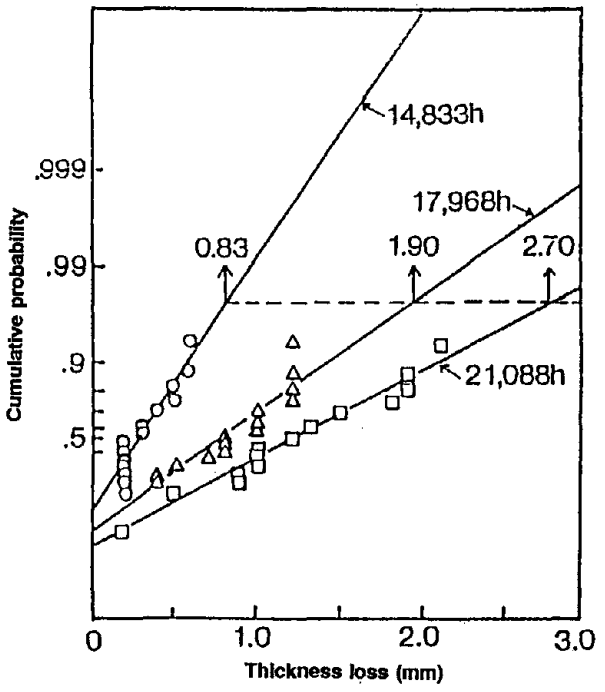


Fig. 3. The Gumbel plots of the maximum thickness loss of boiler tubes used for different operation times.

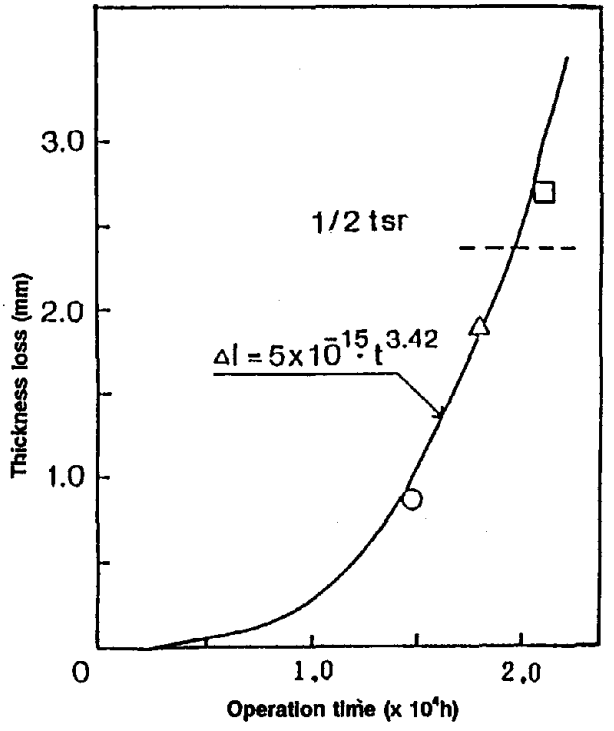


Fig. 4. Thickness loss of boiler tubes as a function of operation time and estimation of rupture time.

corrosion of steel pipes and piles was found to cause the collapse of harbor structures. Then corrosion damage of steel structures exposed in sea water has been inspected and analyzed by using extreme value analysis. Itoh et al. [38] reported that three different types of depth distribution were found for steel piles and plates depending on exposure time and exposure location such as water level and deep sea (Fig. 5). The type A distribution which exhibits a nearly straight line, was found for uniform corrosion loss, its mean value being below 1.0–1.2 mm thickness. The type C distribution obeying the Gumbel distribution was observed for heavily localized specimens. The type B distribution is a mixed type of A and C distributions. The estimated depth using the return period was consistent with observations.

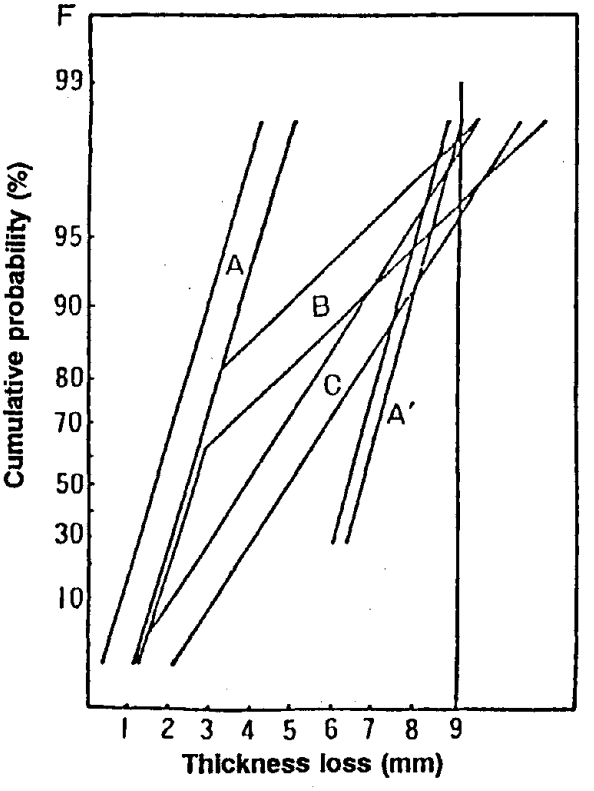


Fig. 5. Various types of the distribution observed for steel piles and pipes exposed in sea water.

3.4 Classification of Data Based on Corrosion Knowledge

In any of the cases mentioned earlier, measured sets of data is fitted by two or three distributions and must be separated from each other before the analysis in order to obtain the maximum value.

Imagawa et al. [15,16,17,39] presented many cases which require classification of data. For example, data for the heat exchanger tubes had to be classified into the inlet and outlet side samples because corrosion form and its degree of damage were different at the two locations owing to exposure to different temperatures. For the oil tank, Imagawa observed that more deep pits were formed on the welding line compared with other parts. He obtained the different estimated value of the pit depth for each classified sample. At the present time, the classification was done on corrosion knowledge and experience, but it is required to establish a procedure based on a common criterion.

3.5 Crack Depth Distribution of Stress Corrosion Cracking

Stress corrosion cracking is one of the most dangerous corrosion failure and shows random occurrence which is a very specific and common feature of materials fracture. The Weibull distribution has been known to be quite useful to analyze the distribution of fracture strength of various materials [40] and also has been found to be applicable for analyzing failure life distribution due to stress corrosion cracking [33, 34].

An interesting application of the Gumbel distribution for analyzing the crack depth distribution has been reported by Tsuge [41]. The laboratory experiment for evaluating the susceptibility of stress corrosion cracking of Type 304 stainless steel was done by using a bent specimen of u-shape. Bending gives stress to the specimen and the environmental condition of high pressure water causes many cracks, which can be revealed by sectioning the specimen after the test as shown in Fig. 6. Distribution of the crack depth plotted in the Gumbel probability paper showed two lines with an inflection

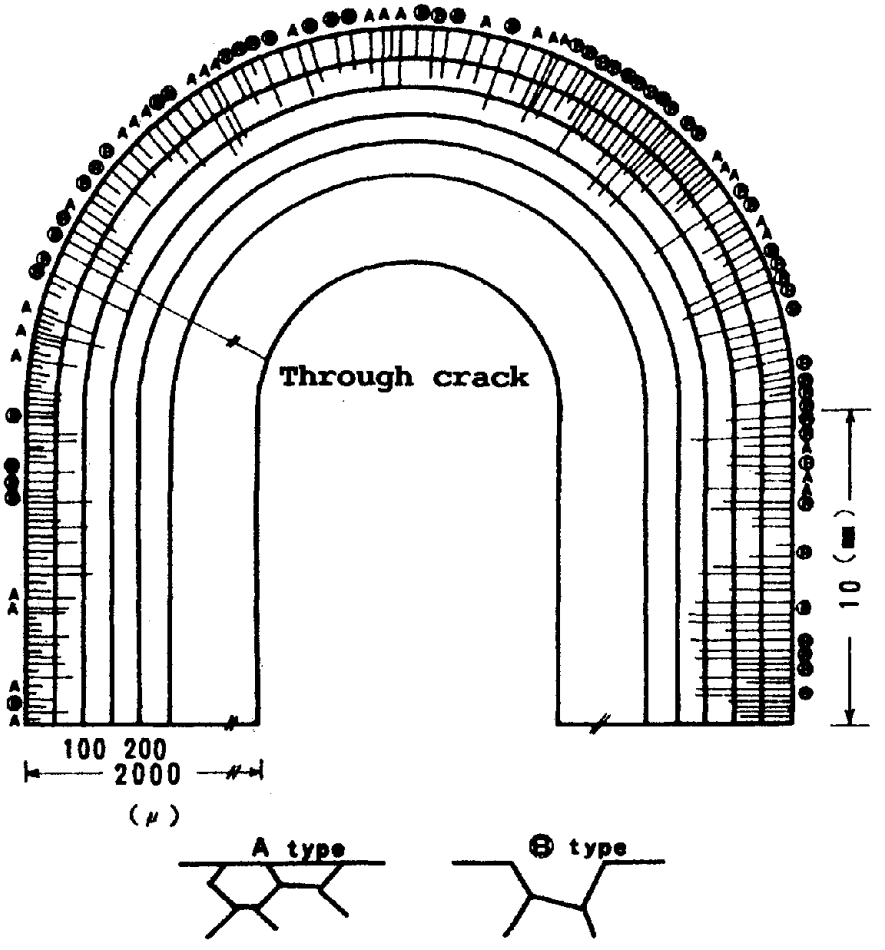


Fig. 6. The distribution of intergranular corrosion attacks and cracks observed for sensitized type 304 stainless steel exposed to the BWR simulated water (DO 8 ppm, 250 °C).

point as can be seen in Fig. 7. This inflection point was found to correspond just to a depth for initiating the intergranular crack. Thus the initiation of the intergranular crack growth could be separated from the initial process of purely chemical intergranular corrosion.

3.6 Estimation of the Maximum Segregation of Impurities in Steel

Continuous casting of steel is one of the innovative technologies achieved by the steel industry. Segregation and its band which are formed during solidification at the center of slab remain after rolling and work as initiation sites for fracture phenomena such as lamellar tear and hydrogen induced cracking (HIC). The maximum amount of segregation was found to be related to the above fracture phenomena, so that extreme value analysis was applied for estimating the maximum amount of segregation in steel plate from small area samples [42], the concentration of impurities being measured by using EPMA. The maximum amount of segregation thus determined can be used to predict the susceptibility to lamellar tear. It should be emphasized that the statistical procedure for the chemical analysis is mainly concerned with the mean and standard deviation which assesses the reliability of the measurement, but not with extreme values. In recent years, highly sensitive analytical methods have been developed, but it is not clear how to correlate the data of small area samples to that of the total or bulk specimen. The ratio of the analytical area to the bulk specimen reaches almost to 10^{-9}, and extreme value statistics is expected to be useful.

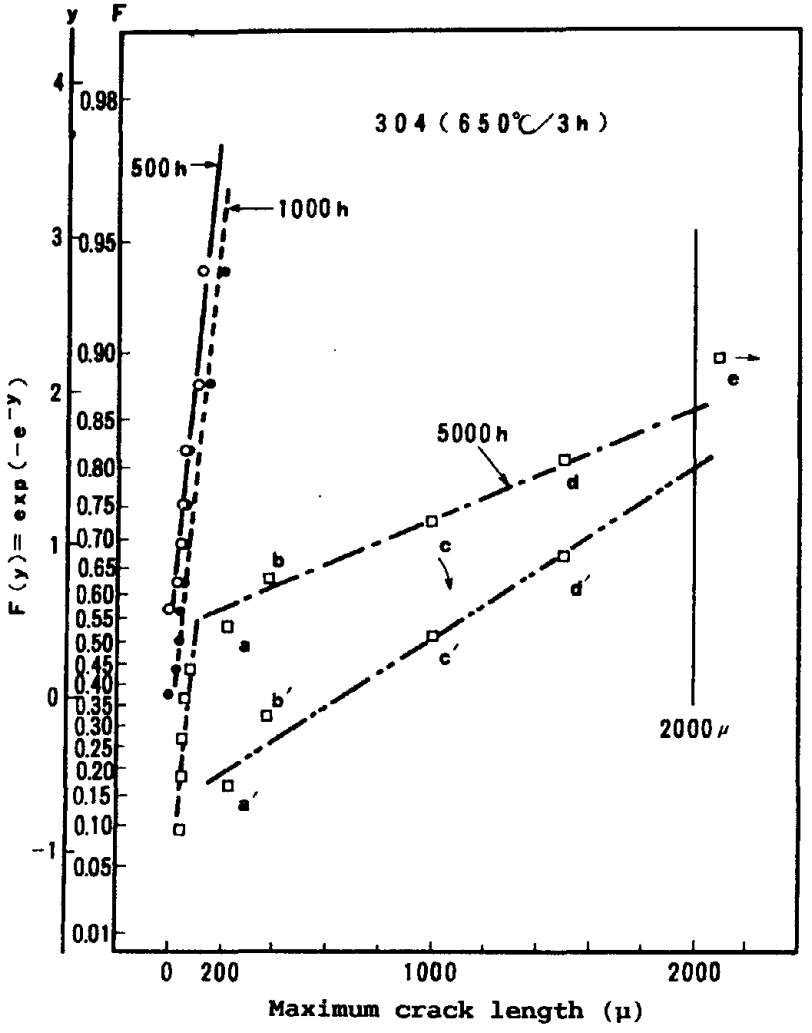


Fig. 7. Distributions changing with exposure time, the initial distribution corresponding to intergranular corrosion and the second to intergranular cracking.

3.7 Non-Destructive Methods With the Extreme Value Analysis

Various types of nondestructive methods are used for inspecting and examining corrosion damage in order to prevent failure. High sensitivity and resolution in time or in space are required for the measurement. In addition, a computer-aided operation becomes popular, because huge amounts of data must be evaluated. For the heat exchanger, a thousand tubes must be checked and the number of data easily exceeds 10^5. An eddy current sensor [43] and ultrasonic sensor probe [44] to steel tubes, and an impedance sensor probe [45] for coating tubes have been developed with the data logger and the extreme value analysis software.

4. Discussion

The size effect on the maximum pit depth is found to be estimated with confidence by introducing the concept of the return period. Theoretical bases of the procedure have been provided by extreme value theory [1]. Our experience shows that the pit depth distribution obeys the normal or exponential distribution, which belong to the exponential distribution family. Thus the maximum values of pit depth extracted from the exponential family distribution may reasonably be expected to obey the Gumbel distribution. Thus the size effect could be rationally predicted by using the concept of return period.

Evans [46] pointed out, however, that some cases as observed by Wormwell et al. [47] does not obey a normal or exponential type distribution, but that the tail of the distribution is limited at a certain depth. Evans emphasized that such a limit is reasonable for the case of anodic reaction control situation and this limited depth gives a rough indication of the greatest pit depth to be expected on a much larger area. Evans, however, did not check another possibility using the type III distribution which has an upper limit. Recently Laycock et al. [24] discussed that usefulness of the generalized extreme value (GEV) distribution:

$$F(x)=\exp\,(-(1-k(x-u)/\alpha)^{1/k}\quad kx<=\alpha+uk, \tag{13}$$

because the distribution subsumes all three types with the sign of a shape parameter, k. When k is zero, negative or positive, the distribution changes to type I, type II, and type III, respectively. They found that the pit distribution on stainless steels in acidified chloride solution fits the GEV distribution with k positive, indicating that the type III for the largest value could be fitted. The type III distribution has a bound or a limit with increasing area, as suggested by Evans.

What sample size, or what size of specimen area should be used are questions from non-specialists in statistics. For this question, we proposed a procedure or criterion for choosing s, N and T based on the variance given by Eq. (6). The surface area, S, of the object is given, and the sampling area, s, is selected so as to include at least one pit. Then $T(=S/s)$ is obtained. Accumulated data of the parameters of α and λ suggest [32] that the ratio of α/λ for localized corrosion is below, or not much larger than, 0.3. Kinnison [48] states that the asymptotic theory predicts a constant ratio of 0.313 for all extreme value distributions. Then it can be assumed that the ratio α/λ, is 0.3. If we wish to control variance within $(\lambda/3)^2$, the following relation can be deduced from Eq. (6)

$$(\lambda/\alpha 3)^2=A(N,n)y^2+B(N,n)y+C. \tag{14}$$

Equation (14) can be solved for y or T as a function of N and α/λ, as plotted in Fig. 8. When the ratio of α/λ can be equated to 0.3 as discussed before, a suitable number of samples can be found for a given return period, T. From this figure, the required size of samples is $N=30$ for $T=1200$, or $N=20$ for $T=274$ and so on. This figure is approximately the same as what was observed empirically.

5. Conclusions

Extreme value statistics has been found to be a powerful tool for estimating the maximum value of localized corrosion depending on the surface area. Accumulation of data and experience, however, reveals that statistics is less important than corrosion experience and knowledge for obtaining a reasonable estimation; measured data must be classified based on the form of corrosion damage and its degree before the analysis. Properly classified data is found to provide a very reasonable value. Nondestructive methods for measuring wall thickness with various types of sensors, combined with extreme value analysis, have been developed in recent years.

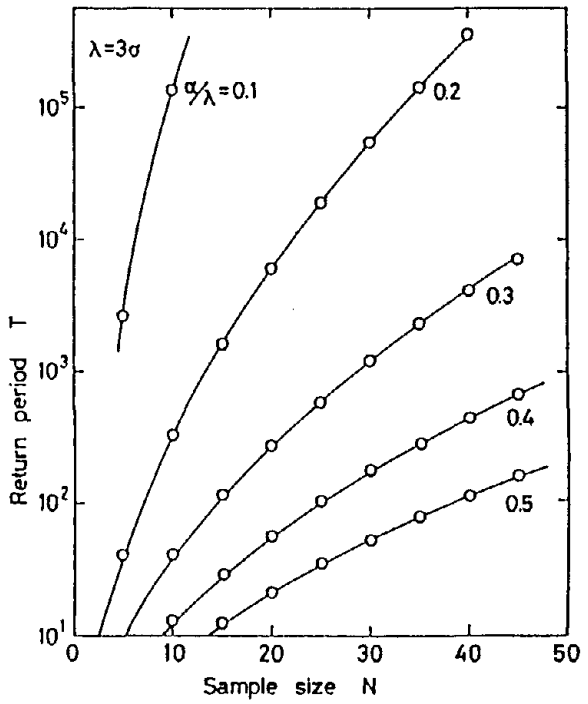


Fig. 8. Calculated curves of the optimum condition to choose T and N at various value of α/λ.

Acknowledgments

The author would like to acknowledge the contribution of the JSCE 60-1 Committee Members, particularly H. Tsuge and M. Akashi for establishing the standardized procedure for estimation of the maximum pit depth.

6. References

[1] E. J. Gumbel, Statistics of Extremes, Columbia University Press, New York and London (1958).
[2] E. J. Gumbel, Statistical Theory of Extreme Values and Some Practical Applications, Natl. Bur. Stand. (U.S.), Appl. Math. Ser. No. 33 (1954).
[3] R. B. Mears and U. R. Evans, The "Probability" of Corrosion, Trans. Faraday Soc. **31**, 527–542 (1935).
[4] U. R. Evans, The Corrosion and Oxidation of Metals: Scientific Principles and Practical Applications, Edward Arnold, London (1960).
[5] G. G. Eldredge, Corrosion **13** (13), 51–60 (1957).
[6] G. N. Scott, Proc. Amer. Petroleum Inst. **15** (IV), 18 (1934).
[7] L. H. C. Trippet, Biometrika **17**, 364–386 (1925).
[8] P. M. Aziz, Corrosion **12** (12), 495–506 (1956).
[9] N. Masuko, Boshoku Gijutsu **21** (8), 347–353 (1972) (in Japanese.)
[10] T. Shibata, Boshoku Gijutsu **27** (1), 23–35 (1978). (in Japanese.)
[11] T. Shibata, Kinzoku Hyomen Gijutsu **31** (6), 336–344 (1980) (in Japanese.)
[12] Y. Ishikawa, Boshoku Gijutsu **28** (5), 278–284 (1979) (in Japanese.)
[13] Y. Ishikawa, T. Ozaki, N. Hosaka, and O. Nishida, J. Soc. Mater. Sci. Jpn. **30** (337), 975–980 (1981) (in Japanese.)
[14] Y. Ishikawa, T. Ozaki, N. Hosaka and O. Nishida, Trans. ISIJ **22** (12), 977–983 (1982).
[15] H. Imagawa and K. Matsuno, Boshoku Gijutsu **31** (6), 406-411 (1982) (in Japanese.)
[16] H. Imagawa and K. Matsuno, J. Soc. Mater. Sci. Jpn. **36** (400), 76–81 (1987) (in Japanese.)
[17] H. Imagawa and K. Matsuno, J. Soc. Mater. Sci. Jpn. **38** (430), 823–828 (1989) (in Japanese.)
[18] S. Kase, Boshoku Gijutsu **31** (7), 481–487 (1982) (in Japanese.)
[19] S. Kase, How to Analyze Reliability Data—Double Exponential Distribution, Ohm Ltd., Tokyo (1983) (in Japanese).
[20] J. Lieblein, A New Method of Analyzing Extreme-Value Data, National Advisory Committee For Aeronautics, Tech. Note 3053, January 1954.
[21] H. Tsuge, J. Soc. Mater. Sci. Jpn. **36** (400), 35–46 (1987). (in Japanese).
[22] JSCE 60-1 Technical Committee, Working Group, Corrosion Eng. **37** (12), 699–705 (1988).
[23] JSCE 60-1 Technical Committee, Working Group, Computer program "EVAN," Maruzen Publ., Tokyo (1989).
[24] P. J. Laycock, R. A. Cottis and P. A. Scarf, J. Electrochem. Soc. **137** (1), 64–69 (1990).
[25] M. Kowaka, H. Tsuge, M. Akashi, K. Masamura and H. Ishimoto, An Introduction to the Life Prediction of Plant Materials—Application of Extreme Value Statistical Methods for Corrosion Analysis, Maruzen Publ., Tokyo (1984) (in Japanese).
[26] A. H-S. Ang and W. H. Tang, Probabilistic Concepts in Engineering Planning and Design, Vol. II—Decision, Risk, and Reliability, John Wiley & Sons, New York (1984) p. 186.
[27] R. R. Kinnison, Applied Extreme Value Statistics, Battelle Press, Macmillan Publishing Co, A Division of Macmillian, Inc., New York (1985).
[28] 41st Corrosion and Protection Symposium, Life Prediction by Extreme Value Statistics, JSCE, January 1982.
[29] 51st Corrosion and Protection Symposium, Life Prediction of Plant Materials, JSCE, September 1983.
[30] 58th Corrosion and Protection Symposium, Life Prediction of Plant Materials—Application of Extreme Value Statistics to Corrosion, JSCE, December 1984.
[31] 73rd Corrosion and Protection Symposium, Life Prediction of Plant Materials—Experience and Practice of the Analysis, JSCE, March 1988.
[32] T. Shibata, ISIJ International **31** (2), 115–121 (1991).
[33] T. Shibata, Localized Corrosion, F. Hine, K. Komai, and K. Yamakawa, eds., Elesvier Applied Science, London and New York (1988) p. 197.
[34] T. Shibata, Proc. Life Prediction of Corrodible Structures, R. N. Parkins, ed., NACE, Houston, to be published.
[35] T. Shibata and K. Okamoto, Boshoku Gijutsu **34** (7), 404-408 (1981) (in Japanese.)
[36] R. Araki, A. Miura, M. Sakai, J. Yokoyama, and S. Yokoya, Atsuryoku Gijutsu **24** (3), 124–132 (1986) (in Japanese.)
[37] Y. Fukuda, H. Kawasaki, and M. Seki, in Ref. [30], p. 24.
[38] S. Itoh, T. Murata, and H. Okada, Poc. Inter. Congr. Metallic Corrosion, Vol. 2, National Research Council of Canada, Ottawa (1984) pp. 330–333.

[39] H. Imagawa, Haikan Gijutsu (Piping Engineering), Special Vol. 91-96 (1987) (in Japanese.)
[40] W. Weibull, J. Applied Mechanics, ASME **18**, 293–297 (1951).
[41] H. Tsuge, in Ref. [29], pp. 16–21 (in Japanese.)
[42] H. Ogawa, in Ref. [30], pp. 31–36.
[43] T. Anzai, H. Yamamoto, K. Wakebe, Bosei Kanri **34** (4), 135–139 (1990) (in Japanese.)
[44] M. Kimura, Haikan Gijutsu, Special Vol. 142–145 (1987) (in Japanese.)
[45] S. Kurusu and T. Sato, Haikan Gijutsu, Special Vol. 156-160 (1987) (in Japanese.)
[46] U. R. Evans, The Corrosion and Oxidation of Metals: Scientific Principles and Practical Applications, Edward Arnold, London (1960) p. 927.
[47] F. Wormwell, G. Butler, and J. G. Beynon, Trans. Inst. Marine Eng. **69**, 109–120 (1957).
[48] R. R. Kinnison, Applied Extreme Value Statistics, Battelle Press (1985) p. 70.
[49] S. Kobayashi, in Ref. [30], pp. 8–13.
[50] K. Matsuno and H. Imagawa, in Ref. [31], pp. 8–15.
[51] H. Ishimoto, in Ref. [29], pp. 56–62.
[52] S. Takasaki, T. Kontani, J. Soc. Mater. Sci. Jpn. **36** (400), 72–75 (1987).
[53] F. Blekkenhorst, G. M. Frerrari, C. J. Van Der Weeken, and F. P. Ijssceling, Br. Corros. J. **21** (3), 163–176 (1986).
[54] H. Sato, K. Shimogori, H. Nishimoto, K. Miki, K. Ikeda, M. Iwai, H. Sakai, and S. Nomura, Testu-to-Hagane **72** (8), 1098–1105 (1986) (in Japanese.)
[55] K. Komai, K. Minoshima, and K. Kim, J. Soc. Mater. Sci. **36** (401), 141–146 (1987) (in Japanese.)
[56] Committee of Materials for Chemical Industry Plants, Society of Chemical Engineering, Japan, SCC of Heat Exchanger Tubes—2nd Report on Survey and Failure Life Analysis, Kagakukogyosha, Tokyo (1984).
[57] W. L. Clarke and G. M. Gordon, Corrosion **29** (1), 1–12 (1973).
[58] T. Shibata, J. Nakata, and S. Fujimoto, Life Prediction of Corrodible Structures, R. N. Parkins, ed., NACE, Houston, to be published.
[59] M. Akashi, Localized Corrosion, F. Hine, K. Komai, and K. Yamakawa, eds., Elesevier Applied Science, London and New York (1988) p. 175.

About the author: *Toshio Shibata is a professor at Osaka University, Osaka, Japan, and chairs the Environmental Materials and Surface Processing Group in the Department of Materials Science and Processing.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 337 (1994)]

Exact Solution to an Interacting Extreme-Value Problem: The Pure-Flaw Model

Volume 99 Number 4 July–August 1994

P. L. Leath

Department of Physics and Astronomy,
Rutgers University,
Piscataway, NJ 08855-0849

and

P. M. Duxbury

Department of Physics and Astronomy,
Michigan State University,
East Lansing, MI 48824-1116

Simple models play a key role in the microstructural analysis of mechanical failure in composites and other materials with complex and often disordered microstructures. Although equal load-sharing-models are amenable to rigorous statistical analysis, problems with local load enhancements near failed regions of the material have so far resisted exact analysis. Here we show for the first time, that some of the simpler of these *local-load-sharing models* can be solved exactly using a sub-stochastic matrix method. For diluted fiber bundles with local load sharing, it is possible to find a compact expression for the characteristic equation of the sub-stochastic matrix, and from it obtain an asymptotic expansion for the largest eigenvalue of the matrix. This in turn gives the asymptotic behavior of the size effect and statistics of the fiber-bundle models. We summarize these results, and show that the important features of the exact result can be obtained from a single scaling analysis we had developed previously. We also compare the statistics of fracture with the appropriate limiting extreme-value survival distribution (a Gumbel distribution), and, as previously indicated by Harlow and Phoenix, note that the Gumbel distribution performs quite poorly in this problem. We comment on the physical origin of this discrepancy.

Key words: extreme-value distributions; scaling analysis; size effect in fracture.

Accepted: March 22, 1994

1. Introduction

It has been known since the pioneering work of Chaplin [1] and well known since the classic work of Griffith [2] that randomly occurring flaws or weak links effectively determine the observed tensile strength of materials. Early on it was realized that the dependence of failure upon the weakest part of a material structure gives rise to non-Gaussian statistics for fracture stress and strain. These developments lead to the classical period of the development of the statistics of extremes by mathematicians such as Dodd [3], Frechet [4], Fisher and Tippett [5], von Mises [6], Gnedenko [7], and Gumbel [8].

Following the work of Duxbury et al. [9–11], there have been many attempts to use random network models to determine the statistics and size dependence of material breakdown [12–16]. These calculations have in many cases elucidated the general behavior and size dependence of breakdown, but few exact results have been produced.

Perhaps the simplest model that shows the statistics of brittle failure has been the pure-flaw, chain-of-bundles model of Harlow and Phoenix [17] which has been studied by Harlow [18] and more recently by Harlow and Phoenix [19]. In this model there is a series or chain of m structurally and

statistically independent bundles of n elements each as shown in Fig. 1, where the vertically applied uniaxial stress is shared by the surviving vertical fibers (bonds). Each element or fiber is independently present with probability p and absent with probability $f=1-p$. The survival probability of the chain of bundles is then the survival probability of a single bundle raised to the power m. The main difficulty in this analysis is calculating the survival probability of a single bundle. The extension of this theory to the survival of two-dimensional networks is straightforward and amounts to the approximation that cracks or flaws only exist and break along the direction transverse to the direction of the applied stress. Following Harlow and Phoenix, we assume the local-load-sharing model for a flaw (crack) of length n to be

$$\sigma_{\text{tip}}=\sigma_0(1+n/2), \tag{1}$$

which is to say that the entire force applied to the cluster is concentrated at the tip (on the fibers adjacent to each end of the flaw or vacancy cluster). Failure of any surviving bond (all of which have the same strength) leads to a rip which causes failure of the entire bundle. Solution of this model requires finding the bond (weakest link) which experiences the largest stress enhancement and that stress which would break this most stressed bond.

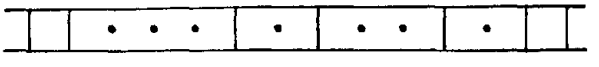


Fig. 1. A one dimensional array of intact bonds (fibers) and flaws (vacancies). The tensile stress σ, is applied vertically.

Duxbury, Leath, and Beale [11] showed how a one-dimensional model such as this could be used as a simple model for fracture or breakdown of a two-dimensional network. If one considers that cracks or flaws only exist and break horizontally, then the two-dimensional model becomes that illustrated in Fig. 2 (i.e., no horizontal bonds break). Then if we impose spiral boundary conditions, where the last site in a row is connected to the first site of the previous row on the other side of the sample, then the $N\times N$ network problem is reduced to a one-dimensional chain of N^2 fibers (or bonds) in parallel like that in Fig. 1.

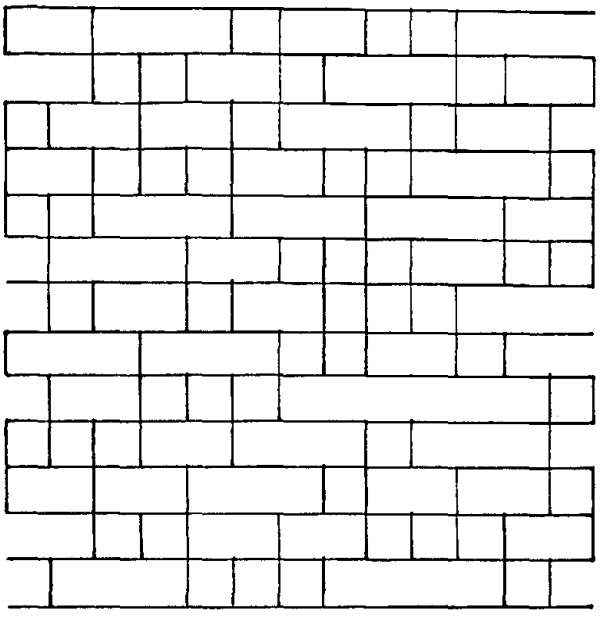


Fig. 2. A two dimensional lattice with only horizontal cracks. Spiral boundary conditions identify each site on the right edge with the site on the left edge of the previous row.

2. Single-Cluster Calculation: The Sub-Stochastic Transition-Matrix Method

In this calculation one assumes that the weakest link is the pure flaw or cluster of vacancies of the largest size that exists in the sample. The survival probability is then closely related to the probability that in a sample of length $L(=N^2)$ that there are *no* clusters of vacant bonds (flaws) of size greater than some prescribed value n. Using a generating-function technique Duxbury, Leath, and Beale [11] calculated exactly the asymptotic form of the probability to be

$$C_L(n)\approx[1-pf^{n+1}]^L \tag{2}$$

in the limit of large L. It is now possible to rederive this result while introducing the sub-stochastic transition-matrix method. Following Harlow [18] we define all possible endings of a fiber bundle of length $L+1$ and the way in which those endings may be generated from a bundle of length L and the probabilities of those endings. Since there are no allowed clusters larger than n, the allowed bundle endings or distinctive endings at a particular site are an occupied site (1) followed by $0\leq r\leq n$ vacancies (0) so that these distinct endings are spanned by the basis vector $\phi_L{}^T=(\phi_{(1)},\phi_{(10)},\phi_{(100)},\ldots,\phi_{(10\ldots0)})_L$ where the last element contains n zeros.

Then the probability increment for going from a cluster of size r to a cluster of size r' on the next site is included in the matrix product

$$M\phi_j = \begin{bmatrix} p & p & p & \dots & p \\ f & 0 & 0 & \dots & 0 \\ 0 & f & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \dots & f & 0 \end{bmatrix} \begin{bmatrix} \phi_{(1)} \\ \phi_{(10)} \\ \vdots \\ \vdots \\ \phi_{(10\dots0)} \end{bmatrix}_j = \begin{bmatrix} \phi_{(1)} \\ \phi_{(10)} \\ \vdots \\ \vdots \\ \phi_{(10\dots0)} \end{bmatrix}_{j+1}, \quad (3a)$$

which is the same as the matrix M operating j times on the probability vector ϕ_0 for the starting site, or

$$M\phi_j = M^{j+1}\phi_0. \quad (3b)$$

The probability $C_L(n)$ that there are only clusters up to size n in the entire bundle (or network) of size L is thus

$$C_L(n) = \sum_{\ell} (\phi_\ell)_L, \quad (4)$$

where the sum is over all the elements of ϕ_L. The simplest and most natural boundary conditions are periodic ones where $C_L(n)$ becomes the trace

$$C_L(n) = tr(M^L) = \sum_k \lambda_k^L, \quad (5)$$

since the 1st and Lth sites must be the same, where λ_k are the eigenvalues of M. We find the eigenvalues of M via its characteristic equation

$$D_n = \det|M/\lambda - I| = \begin{vmatrix} (a-1) & a & \dots & a & a \\ b & -1 & 0 & \dots & 0 \\ 0 & b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 & 0 \\ 0 & \dots & 0 & b & -1 \end{vmatrix} = 0, \quad (6)$$

where $a = p/\lambda$ and $b = f/\lambda = (1-p)/\lambda$. A cofactor expansion of the determinant D_n about its last row, gives immediately the recursion relation

$$D_n = -D_{n-1} + (-1)^n ab^n, \quad (7a)$$

where D_{n-1} is the $n \times n$ determinant for clusters up to size $(n-1)$. With $D_0 = (a-1)$, the solution, upon iteration of Eq. (7a), is the characteristic equation

$$(-1)^n D_n = (a-1) + ab + ab^2 + \cdots + ab^n = 0. \quad (7b)$$

Summing this geometric series we obtain

$$\lambda^{n+2} - \lambda^{n+1} + pf^{n+1} = 0. \quad (8)$$

This equation is the characteristic equation of M times $(\lambda - f)$, so there is an additional spurious root at f (since we are interested in the largest root, this does not affect the analysis). Since M is primitive and non-negative, its largest eigenvalue is real and distinct and it can easily be seen that all the eigenvalues are less than 1 and λ_{max} approaches 1 for large n. Thus we set $\lambda_{max} = 1 - \epsilon$ and expand Eq. (8) to lowest order in ϵ and f^n, which gives us

$$\lambda_{max} \approx 1 - pf^{n+1} + O(f^{2n}). \quad (9)$$

Then, we obtain

$$C_L(n) \cong \lambda_{max}^L \cong (1 - pf^{n+1})^L, \quad (10)$$

which confirms the result Eq. (2) by the sub-stochastic transition-matix method.

In order to find the failure probability as a function of applied stress, we use the load-sharing rule Eq. (1), coupled with the fact that the failure of the bond carrying the largest local stress nucleates catastrophic failure, and thereby use the relation

$$\sigma_b/\sigma = 1 + \frac{n}{2}, \quad (11)$$

where σ_b is the breaking strength of a single fiber. Note that we could have used a variety of other load-sharing rules here, and for example the same expression with n raised to an arbitrary power is also of physical interest. This result combined with Eq. (10) yields the probability $S(\sigma)$ that a fiber bundle will survive at stress σ

$$S(\sigma) = (1 - pf^{2\frac{\sigma_b}{\sigma} - 1})^L. \quad (12)$$

For large n and L, this becomes the modified Gumbel form, introduced previously [10,11] in the analysis of the random fuse network. Although $C_L(n)$ in Eq. (10) becomes a Gumbel distribution in n, the substitution of $n(\sigma)$ from Eq. (11) produces a modified Gumbel distribution that is significantly different from a Gumbel form in σ in the high-reliability tail of the distribution. This modification is discussed further in Sec. 4.

3. Double-Cluster Calculation

Several authors [12,16,19,20] have suggested that the most critical defect is not a single cluster of n vacancies, but rather a double cluster (double co-linear crack) of n vacancies separated by a single occupied site located anywhere within the $n+1$

adjacent sites. Such a double crack is shown in Fig. 3a. This candidate for the most critical crack is appealing because the stress enhancement at the interior occupied site grows as n in network models rather than as $n^{1/2}$ (as for the edges of a single crack in a two-dimensional network) and because the increased entropy of the $(n+1)$ locations of the occupied site makes it more probable. Thus, following Harlow and Phoenix [19], we consider the probability of bundles of length L with repeated double cracks (and single cracks when the occupied site is at either end) not exceeding n vacant sites in any two adjacent cracks separated by a single site. These repeated double cracks are shown in Fig. 3b.

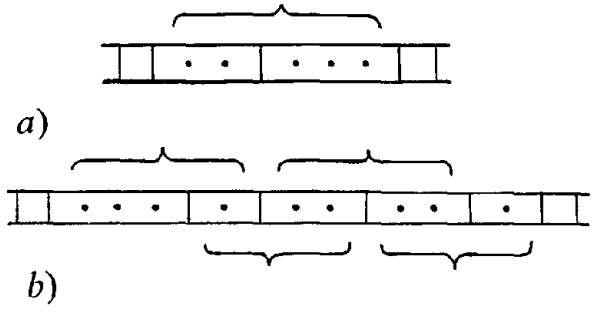


Fig. 3. a) A double cluster of $\leq n$ total vacancies plus one isolated occupied bond. b) Repeated, overlapping double clusters; each pair of clusters as indicated by the brackets contains $\leq n$ total vacancies plus one isolated bond.

Harlow [18] showed that this problem is amenable to analysis by the sub-stochastic transition-matrix method. The load-sharing rule is still given by Eq. (1) as before but now n is the sum of the number of vacant sites immediately on the left and right of any isolated intact bond or fiber. Thus it is necessary to keep track of not only the number of vacancies in the cluster being considered but also those in the previous vacant cluster. There are now $(n+1)(n+2)/2$ distinct endings that must be considered at a site (or bundle ending); these are given by the basis vector $\phi = \phi_{(1)}, \phi_{(10)} \ldots \phi_{(10\ldots0)}; \phi_{(101)}, \phi_{(1010)} \ldots \phi_{(1010\ldots0)}; \phi_{(1001)} \ldots \phi_{(10010\ldots0)}; \ldots; \phi_{(10\ldots01)}$ where there are no more than n total vacancies in any element. With this ordering of states the $n=4$ sub-stochastic matrix for this problem, for example, is given by Eq. (13).

With periodic boundary conditions Eq. (5) still holds and we again analyse the largest eigenvalue of M_n. As a technical point, note that since we are using periodic boundary conditions, we can always start the matrix process at a surviving bond, and so the endings considered above include all possible

$$M_4\phi_L = \begin{bmatrix}
p&0&0&0&0&p&0&0&0&p&0&0&p&0&p\\
f&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&f&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&f&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&\underline{f}&0&0&0&0&0&0&0&0&0&0&0\\
0&p&0&0&0&p&0&0&0&p&0&0&p&0&0\\
0&0&0&0&0&f&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&f&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&\underline{f}&0&0&0&0&0&0&0\\
0&0&p&0&0&0&0&p&0&0&0&p&0&0&0\\
0&0&0&0&0&0&0&0&f&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&\underline{f}&0&0&0&0&0\\
0&0&0&p&0&0&0&0&p&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&\underline{f}&0&0&0&0\\
0&0&0&0&p&0&0&0&0&0&0&0&0&0&0
\end{bmatrix}
\begin{bmatrix}
\phi_{(1)}\\ \phi_{(10)}\\ \phi_{(100)}\\ \phi_{(1000)}\\ \phi_{(10000)}\\ \phi_{(101)}\\ \phi_{(1010)}\\ \phi_{(10100)}\\ \phi_{(101000)}\\ \phi_{(1001)}\\ \phi_{(10010)}\\ \phi_{(100100)}\\ \phi_{(10001)}\\ \phi_{(100010)}\\ \phi_{(100001)}
\end{bmatrix} = \phi_{L+1} \quad (13)$$

survival configurations (we don't have to consider configurations which start with 0's).

A great simplification in the characteristic equation

$$M_n\phi = \lambda\phi, \quad (14)$$

where ϕ are the eigenvectors of M_n, is possible since most of the rows of M_n contain only a single non-zero element f. This gives, for example $f\phi_{(1)} = \lambda\phi_{(10)}$ or $b\phi_{(1)} = (f/\lambda)\phi_{(1)} = \phi_{(10)}$. By such relations, we can eliminate all the rows of M_n except the rows with p's corresponding to the reduced basis vector $\phi = \phi_{(1)}, \phi_{(101)}, \phi_{(1001)}, \ldots, \phi_{(100\ldots01)}$. The resulting $(n+1)$ equations give an $(n+1)\times(n+1)$ matrix M', which satisfies the reduced characteristics equation

$$M'\phi = \begin{bmatrix}
a & a & a & \ldots & \ldots & a\\
ab & ab & \ldots & \ldots & ab & 0\\
ab^2 & ab^2 & \ldots & ab^2 & 0 & \vdots\\
ab^3 & ab^3 & \iddots & 0 & \ldots & \vdots\\
\vdots & \iddots & \iddots & \ldots & \ldots & 0\\
ab^n & 0 & 0 & \ldots & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\phi_1\\ \phi_2\\ \phi_3\\ \vdots\\ \vdots\\ \phi_{n+1}
\end{bmatrix} =
\begin{bmatrix}
\phi_1\\ \phi_2\\ \phi_3\\ \vdots\\ \vdots\\ \phi_{n+1}
\end{bmatrix}, \quad (15)$$

where $a = p/\lambda$, and $b = f/\lambda$. This M' matrix can be considered as a new transition-matrix which adds a *cluster* at a time rather than a bond or fiber at a time. Thus we have the characteristic determinant equation.

$$D_n = \det \begin{vmatrix}
a-1 & a & a & \ldots & \ldots & a\\
ab & ab-1 & \ldots & \ldots & ab & 0\\
ab^2 & ab^2 & ab^2-1 & \iddots & \iddots & \vdots\\
ab^3 & ab^3 & \iddots & \ddots & & \vdots\\
\vdots & \iddots & \iddots & \ldots & -1 & 0\\
ab^n & 0 & 0 & \ldots & 0 & -1
\end{vmatrix} = 0. \quad (16)$$

For small n these determinants can be evaluated directly. For example,

$$D_0 = -1+a, D_1 = 1-a-a^2b \text{ and}$$

$$D_2 = -1+a+ab+a^2b^2-a^3b^3. \quad (17)$$

But by expanding the determinants Eq. (16) by rows and columns, we can show that there is an exact recursion relation

$$bD_n(a,b) = sD_{n-2}(ab,b) - D_{n-4}(ab^2,b), \quad (18)$$

where $D_m(ab^\ell,b)$ is $D_m(a,b)$ with a replaced by (ab^ℓ), and where

$$s = 1+b-a^2b^{n+1}. \quad (19)$$

Note that $s = s(a,b,n) = s(ab,b,n-2) = s(ab^2,b,n-4)$ which is key in the solvability of the recursion relation Eq. (18). After some detailed analysis, we have found (see [21] for details), that this recurrence equation may be solved. The resulting characteristic equation is given by

$$\frac{z_+^{(n-1)/2} - z_-^{(n-1)/2}}{z_+^{(n+1)/2} - z_-^{(n+1)/2}} = s - (1+ab^{(n+1)/2}), \quad (20a)$$

for $n \geq 3$ and odd. While for n even ≥ 4, we find,

$$\frac{z_+^{n/2-1} - z_-^{n/2-1}}{z_+^{n/2} - z_-^{n/2}} = s - \frac{1}{1-ab^{n/2}}. \quad (20b)$$

In Eqs. (20a) and (20b),

$$z_{+,-} = \frac{s \pm \sqrt{s^2-4b}}{2b}. \quad (20c)$$

The key quantity s is given by Eq. (19) above. Equations (20a,b) are the exact expressions for the characteristic equation of the original M in Eqs. (13) and (14).

Again we find that the largest eigenvalue of M is near 1 for large n, so we set $\lambda = 1-\epsilon$ and expand Eqs. (20a,b) and find that in both cases, to lowest order in ϵ and f^n,

$$\epsilon = [(n+2)p^2 - p]f^{n+1} + O(f^{3n/2}). \quad (21)$$

Comparing this double-cluster result to the single-cluster result Eq. (9) we find the expected $(n+2)$ from the possible locations of the single bond in a double cluster of size $(n+1)$. The $(-p)$ is a correction to properly handle the single-cluster cases as well as the double-cluster case, since these are included whenever the isolated bond is located at one end of the double cluster. It is only important for small n.

In order to check and better understand the asymptotic form Eq. (21) of λ_{max} and the importance of the other eigenvalues λ_i we have made several numerical evaluations of the various equations. First, we have numerically found the largest eigenvalue λ_{max} of the original, full, substochastic transition-matrix M as given by Eq. (13). Using the unit vector as a starting vector we repeatedly apply the matrix M to it. Since the largest eigenvalue is unique, this process converges exponentially to the largest eigenvalue. We found in general that convergence occurred to six significant figures with at most 50 matrix products (even for matrices M of dimension $(n+1)(n+2)/2 = 10{,}000$. The sparsity of M with this iterative procedure eliminated matrix-storage problems. The results of this iterative procedure for $C_L'(n)$ versus n are shown in Fig. 4 as solid lines, and the dots give λ_{max}^L with λ_{max} as given by the asymptotic form Eq. (21). Good agreement is seen for all p, with a small deviation in the $p = 0.2$ data. However, a more stringent test is needed in the high-reliability (large n) tail. Thus, in Fig. 5 we plot the quantity

$$\frac{1-\lambda_{max}}{f^{n+1}} = \frac{\epsilon}{f^{n+1}} \approx [(n+2)p^2 - p]. \quad (22)$$

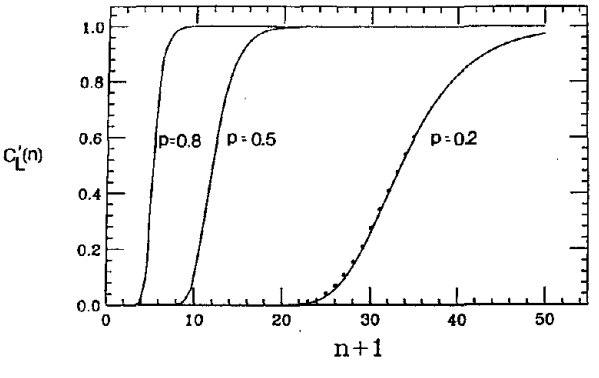


Fig. 4. A plot of the double-cluster probability $C_L'(n)$ vs $(n+1)$, for $p = 0.2$, 0.5, and 0.8. The dotted line is the asymptotic form given by Eq. (24). The solid lines are found from evaluating the largest eigenvalue numerically, and by using $C_L'(n) \approx \lambda_{max}^L$.

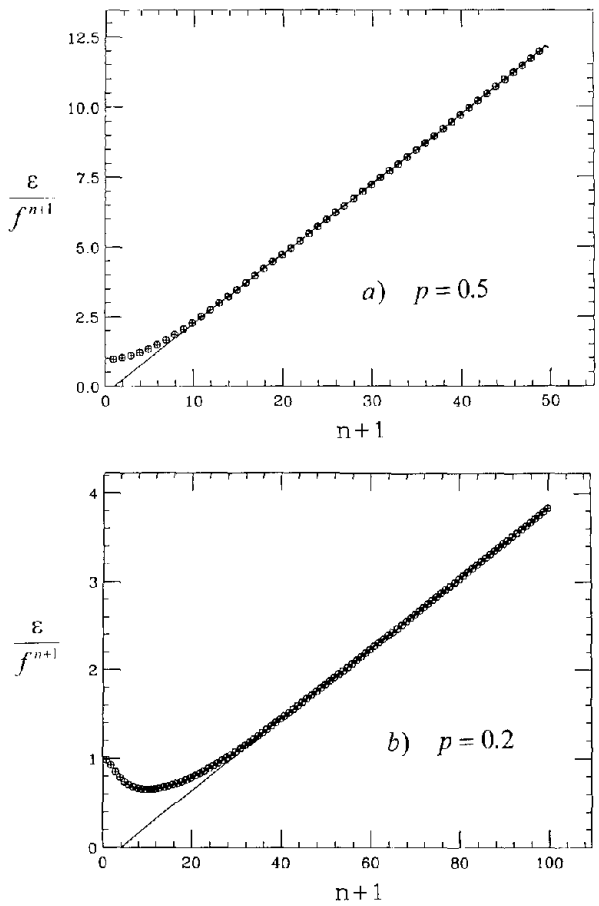


Fig. 5. Plots of ϵ/f^{n+1} vs $n+1$, for a) $p=0.5$ and b) $p=0.2$. The solid lines are the asymptotic form as given by Eq. (22); the circles are the exact values of $(1-\lambda_{\max}^{L})/f^{n+1}$ as obtained by iterating M numerically.

The solid straight line versus $(n+1)$ is the asymptotic result, which is linear in n and this is compared with the iterated numerical values (circular symbols), for $p=0.5$ and 0.2. For large n, in all cases the two calculations agree. But for small values of $p<0.5$ there appears a minimum in $(1-\lambda_{\max})/f^{n+1}$ versus n which corresponds to higher order terms in f^n. In particular the next order term in $\lambda_{\max}^{L}$ is $O(f^{3n/2})$ which would appear as a $O(f^{n/2})$ correction in Fig. 5.

Finally, we test for the accuracy of neglecting all but the maximum eigenvalue $\lambda_{\max}^{L}$. It is possible to directly, numerically evaluate the trace of M^L by iteration since it only requires storage of the matrix and one vector at any time. Using this method, we have evaluated the quantity

$$(1-(tr(M^L))^{1/L})/f^{n+1}, \tag{23}$$

which should converge to $(1-\lambda_{\max})/f^{n+1}$ when the largest eigenvalue is dominant. A numerical test of this convergence is shown in Fig. 6, for $L=1000$, and shows that for large lattice sizes, the most important corrections to $C_L'(n)$ are the higher order contributions to $\lambda_{\max}$, which are of $O(f^{3n/2})$, rather than the neglected smaller eigenvalues of M, which are relatively unimportant here.

Finally, we obtain the asymptotic form

$$C_L'(n)\approx\{1-[(n+2)p^2-p]f^{n+1}+O(f^{3n/2})\}^L. \tag{24}$$

Thus, following the same arguments as at the end of Sec. 2, we find that to leading order the survival probability of the entire network or chain-of-bundles is

$$S_L'(\sigma)=[1-(\frac{2\sigma_b}{\sigma}p^2-p)f^{\frac{2\sigma_b}{\sigma}-1}]^L, \tag{25}$$

which is again of the form of a modified Gumbel distribution with slightly different coefficients from Eq. (12).

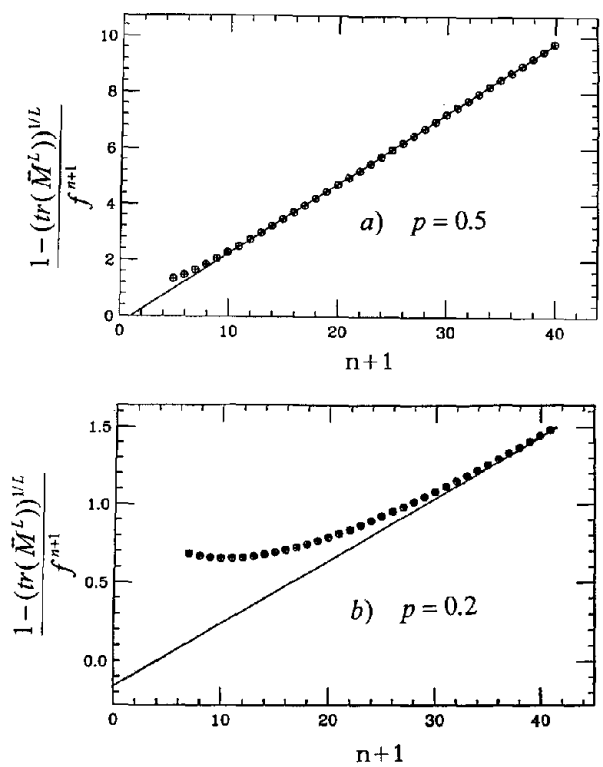


Fig. 6. Plots of the exact value of $\left[1-\left(tr(M^L)\right)^{1/L}\right]/f^{n+1}$ vs $(n+1)$ obtained numerically from M (circles); for $L=1000$, for a) $p=0.5$ and b) $p=0.2$. The solid line is the asymptotic form Eq. (22).

4. Extreme-Statistical Form

For large L, we can easily find the limiting form of $C_L'(n)$ as given by Eqs. (10) and (24) respectively,

$$C_L(n) \underset{L\to\infty}{\sim} \exp[-Lp\exp(-n\log(1/f))], \quad (26)$$

and

$$C_L'(n) \underset{L\to\infty}{\sim} \exp[-L((n+2)p^2-p)\exp(-n\log(1/f))], \quad (27)$$

as the upper limit behavior for large n which is important in the high-surviveability tail of the distribution. This is a Gumbel distribution, as is expected from the exponential [pf^n or npf^n] behavior for the probability of large clusters [Castillo, 1988].

On the other hand, the surviveabilities $S(\sigma)$, and $S'(\sigma)$, in the limit $L\to\infty$, are found to be *modified* Gumbel forms [from Eqs. (12) and (25)],

$$S_L(\sigma) \underset{L\to\infty}{\sim} \exp[-Lp\exp(-(\tfrac{2\sigma_b}{\sigma}-1)\log(1/f))] \quad (28a)$$

and,

$$S_L'(\sigma) \underset{L\to\infty}{\sim} \exp[-L(\tfrac{2\sigma_b}{\sigma}p^2-p)\exp(-(\tfrac{2\sigma_b}{\sigma}-1)\log(1/f))]. \quad (28b)$$

That the dominant behavior of $S_L(\sigma)$, and $S_L'(\sigma)$ is $\exp(-LA\exp(-\frac{B}{\sigma}))$ as σ tends to zero is essential to ensuring that the survival distributions have the proper limiting approach to one when the applied stress approaches zero. Harlow and Phoenix [19] have numerically shown that this high-reliability tail can not be well described by a Gumbel form for σ (such as $\exp(-LA\exp(B(\sigma-\sigma_0)))$). But this failure is obvious since the Gumbel form doesn't approach one until $\sigma\to-\infty$, so at sufficiently small stresses it must be inaccurate. Nevertheless the standard texts on extreme distributions (see, for example [8] or [22]) seem to suggest that the Gumbel distribution is the appropriate one in such cases. The difference clearly is in the form of the normally assumed scaling $\lim_{N\to\infty}[S(\sigma)]^N = S(a_N\sigma+b_N)$, which fits the shift and slope of the limiting function $S(\sigma)$ at its median, but fails near the high-reliability limit $\sigma=0$. It would seem that instead a more general scaling form $\lim_{N\to\infty}[S(n(\sigma))]^N = S(a_N n(\sigma)+b_N)$ must be allowed to also include the proper high-surviveability limit near zero stress. In many practical material-failure problems, this modification of the Gumbel form is essential in order to correctly represent the important high-reliability tail. Note that this is not true of the Weibull distribution, which as well as being a stable limiting extreme-value distribution, does have the physically correct behavior as stress approaches zero. This is one good reason why the Weibull distribution is a very robust form in the analysis of failure problems. We suggest that the family of modified Gumbel distributions of the sort Eq. (28), should be similarly robust, in contrast to the conventional Gumbel distribution which is of limited use in the analysis of the statistics of material failure.

5. General Scaling Behavior

The size dependence and general form of the limiting distribution can usually be found from a back-of-the-envelope scaling calculation which we introduced previously [9,11]. First, for the single-cluster calculation, consider the probability $P_L(n)$ of finding a cluster of size n in a sample of size L. The order of magnitude of this probability is

$$P_L(n) \propto Lp^2f^n/p = Lpf^n, \quad (32)$$

since, for normalization, $\Sigma p^2 f^n = p$, and since there are L different locations in the sample where a cluster could be located. For the maximum cluster size to be expected in a sample of size L we set

$$P_L(n) \sim Lpf^{n_{\max}} = 1, \quad (33)$$

and obtain the size dependence

$$n_{\max} \cong \frac{\log L}{\log(1/f)}, \quad (34)$$

or, from the load-sharing rule Eq. (1), the breakdown stress σ_c

$$\sigma_c^{(L)} \sim \frac{1}{1+n_{\max}/2} \sim \frac{1}{1+\dfrac{\ln L}{2\ln(1/f)}}, \quad (35)$$

scales to zero logarithmically in the thermodynamic limit. A similar argument for double clusters gives

$$P_L'(n) \propto L(n+1)p^2f^n, \quad (36)$$

since there are $(n+1)$ places to put an isolated bond in the n-double-clusters. We then obtain the same limiting form Eq. (35) for $\sigma_c(L)$, although

there are additive (log(log$L$)) corrections in the double-cluster case. The logarithmic scaling law in turn implies that the failure statistics is of the double-exponential form given in Eq. (28). The Weibull and Frechet distributions always give power-law size scaling. These qualitative arguments are very powerful and are confirmed by the rather elaborate, exact calculation described here.

Acknowledgments

One of us (P.L.L.) would like to thank the Department of Physics and Astronomy, Michigan State University, for its hospitality and support during his sabbatical from Rutgers University during the 1993 winter semester when this work was done. We would like to thank S. Leigh Phoenix, and Michael F. Thorpe for helpful discussions during the course of this research. One of us (P.M.D.) acknowledges the support of the DOE under grant #DE-FG02-90ER45418, and the CMSC at MSU.

6. References

[1] W. S. Chaplin, Van Nostrands' Engineering Magazine **23**, 441 (1880); W. S. Chaplin, Proceedings of the Engineers Club, Philadelphia **3**, 15 (1882).
[2] A. A. Griffith, Phil. Trans. Roy. Soc. **221A**, 163 (1920).
[3] E. L. Dodd, Trans. Am. Math. Soc. **25**, 525 (1923).
[4] M. Frechet, Ann. Soc. Math. (Cracow) **6**, 92 (1927).
[5] R. A. Fisher and L. H. C. Tippett, Proc. Camb. Phil. Soc. **24**, 180 (1928).
[6] R. Von Mises, Rev. Math. Union Interbalkanique **1**, 141 (1936).
[7] B. Gnedenko, Ann. Math. **44**, 423 (1943).
[8] E. J. Gumbel, Statistics of Extremes, Columbia University Press, New York (1958).
[9] P. M. Duxbury, P. D. Beale, and P. L. Leath, Phys. Rev. Lett. **57**, 1052 (1986).
[10] P. M. Duxbury and P. L. Leath, J. Phys. A **20**, L411 (1987).
[11] P. M. Duxbury, P. L. Leath, and P. D. Beale, Phys. Rev. B **36**, 367 (1987).
[12] Y. S. Li and P. M. Duxbury, Phys. Rev. B **36**, 5411 (1987).
[13] B. Kahng, G. G. Batrouni, and S. Redner, J. Phys. A **20**, L827 (1987).
[14] B. Kahng, G. G. Batrouni, S. Redner, L. De Arcangelis, and H. J. Herrmann, Phys. Rev. B **37**, 7625 (1988).
[15] P. D. Beale and P. M. Duxbury, Phys. Rev. B **37**, 2785 (1988).
[16] P. D. Beale and D. J. Srolovitz, Rhys. Rev. B **37**, 5500 (1988).
[17] D. G. Harlow and S. L. Phoenix, Int. J. Fracture **17**, 601 (1981).
[18] D. G. Harlow, Proc. Roy. Soc. (Lond.) **A397**, 211 (1985).
[19] D. G. Harlow and S. L. Phoenix, J. Mech. Phys. Solids **39**, 173 (1991).
[20] J. Machta and R. A. Guyer, Phys. Rev. **B36**, 2142 (1987).
[21] P. M. Duxbury and P. L. Leath, Phys. Rev. B **49**, 12676 (1994).
[22] E. Castillo, Extreme Value Theory in Engineering, Academic Press, San Diego (1988).

About the authors: P. L. Leath is a physicist in the Department of Physics and Astronomy at Rutgers University. P. M. Duxbury is a physicist in the Department of Physics and Astronomy at Michigan State University.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 345 (1994)]

Inclusion Rating by Statistics of Extreme Values and Its Application to Fatigue Strength Prediction and Quality Control of Materials

Volume 99 Number 4 July–August 1994

Y. Murakami

Dept. of Mechanical Science and Engineering,
Faculty of Engineering Kyushu University,
6-10-1 Hakozaki, Higashi-ku,
Fukuoka, 812 Japan

The inclusion rating method by statistics of extreme values (IRMSE) using $\sqrt{area}$ of inclusions as the size parameter enables one to discriminate between current super-clean steels. Moreover, IRMSE enables one to predict the size ($\sqrt{area_{max}}$) of maximum inclusions contained in domains larger than the inspection domain. The statistical distribution of $\sqrt{area_{max}}$ can be used for the quality control of materials and for the prediction of a scatter band of fatigue strength. Practical procedures of inclusion rating and prediction of a scatter band of fatigue strength are shown.

Key words: fatigue; high strength steel; nonmetallic inclusion; statistics of extreme values.

Accepted: March 22, 1994

1. Introduction

With the increase in cleanliness of steels, conventional inclusion rating methods are no longer as useful as before, because conventional inclusion rating methods cannot determine the cleanliness of new clean steels. Although the cleanliness of steels has been markedly improved in the last two decades, the fatigue strength of recent clean high strength steels cannot attain the ideal value expected from their high static strength. Nonmetallic inclusions are predominantly the cause of lower fatigue strength even for such clean high strength steels. Thus, in order to predict the fatigue strength behavior and to evaluate quality, we need a new inclusion rating method relevant to recent super-clean steels. The inclusion rating method based on statistics of extreme [1] is most relevant for this purpose. In the following, we call this method Inclusion Rating Method by Statistics of Extreme (IRMSE).

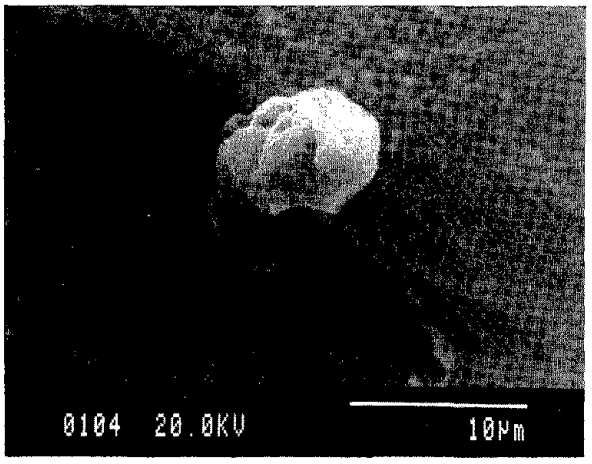
In this study, we shall first show that if we choose an appropriate size parameter for inclusions, the size of inclusions obey the statistics of extreme value theory. The appropriate size parameter is the square root of projected area of the maximum inclusion contained in a standard inspection area or volume, $\sqrt{area_{max}}$. Second, we predict the size of the maximum inclusion which may be contained in a larger area or volume than the standard inspection area and, lastly, we use the size parameter, $\sqrt{area_{max}}$, to predict the scatter band of fatigue strength of hard steels.

The merits of IRMSE, in comparison with conventional methods, are (1) to distinctly discriminate the cleanliness of recent super-clean steels, and (2) to predict the size of larger inclusions contained in a domain larger than the inspection domain. This method is useful for quality control of materials and for improvement of the steel making processes. It also enables one to predict the scatter of the fatigue strength of a large number of mass production products.

2. Nonmetallic Inclusions as a Fatigue Fracture Origin

Figure 1 shows an example of the nonmetallic inclusion which was observed at fatigue origin of a bearing steel under a rotating bending fatigue test. If this inclusion did not exist in this specimen, the fatigue strength of this specimen should have been higher than the applied stress, $\sigma_a = 1078$ MPa. Since the size and location of nonmetallic inclusions scatter randomly, the fatigue strength of high strength steels naturally scatters. Although there has been a firm opinion that the chemical composition and shape of nonmetallic inclusions substantially influences the fatigue limit, Murakami et al. [2–5] have shown the incorrectness of the conventional opinion by their detailed experiments and analyses, and reported distinct experimental evidence that the size of inclusions (defined by $\sqrt{area}$) is the most crucial geometrical parameter. It is empirically known that the intrinsic fatigue strength of steels is determined by the hardness (H_V) of its microstructure. For steels with $H_V < 400$, nonmetallic inclusions contained in current commercial steels are not detrimental and we have the following empirical formula

$$\sigma_w \cong 1.6\, H_V \qquad (1)$$



Vickers hardness $H_V = 745$ kgf/mm^2
Applied stress at surface $\sigma_a = 1078$ MPa
Cycles to failure $N_f = 7.94 \times 10^6$
Square root of projection area of inclusion $\sqrt{area} = 9.8$ μm
Distance from surface $h = 158$ μm
Applied stress at inclusion $\sigma' = 1034$ MPa
Chemical composition of inclusion: Al-Mn-S-O

Fig. 1. A typical example of inclusion observed at the center of fatigue fracture origin [super-clean bearing steel, SUJ2(N)].

where σ_w is the fatigue limit (MPa) and H_V is the Vickers hardness (kgf/mm^2). However, for steels with $H_V > 400$, the effect of inclusions reveals itself and the intrinsic or ideal fatigue limit given by Eq. (1) cannot be attained. The fatigue strength depends on the size ($\sqrt{area}$) and location of the fatal inclusion and H_V of the matrix. Murakami et al.'s [6–9] fatigue limit prediction equations are classified into three categories depending on the location of fatal inclusions (see Fig. 2):

Fatigue limit for a surface inclusion [Fig. 2(a)]

$$\sigma_w = 1.43\ (H_V + 120)/(\sqrt{area})^{1/6} \qquad (2)$$

Fatigue limit for an inclusion in touch with free surface [Fig. 2(b)]

$$\sigma_w = 1.41\ (H_V + 120)/(\sqrt{area})^{1/6}. \qquad (3)$$

Fatigue limit for an internal inclusion [Fig. 2(c)]

$$\sigma_w = 1.56\ (H_V + 120)/(\sqrt{area})^{1/6}, \qquad (4)$$

where the units are σ_w: MPa, $\sqrt{area}$: μm, and H_V: kgf/mm^2.

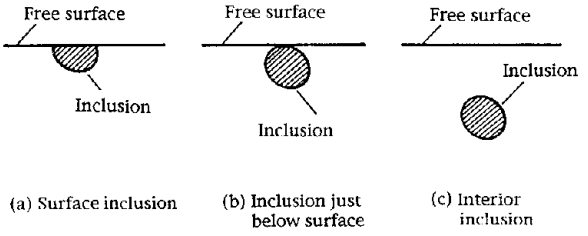


Fig. 2. Various locations of inclusions causing fatigue fracture.

Since for a constant value of area, an inclusion is most detrimental when it exists just in touch with the free surface of a specimen, we can use Eq. (3) in combination with the maximum size $\sqrt{area}_{max}$ obtained by IRMSE to predict the lower bound ($\sigma_{w\ell}$) of scattered fatigue strength of many specimens or machine elements.

3. Inclusion Rating of Various High Strength Steels by Statistics of Extreme

Figure 3 explains the practical procedure to implement the inclusion rating by statistics of extreme values. The details of this method are reported in

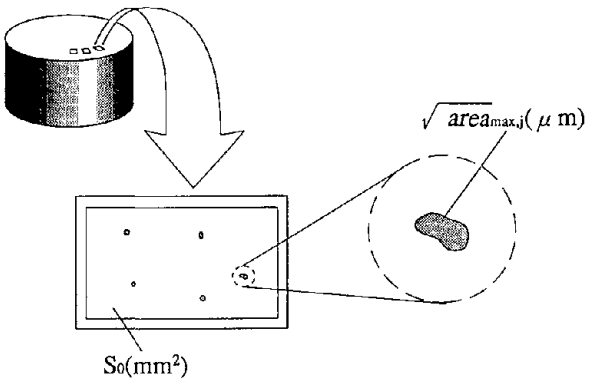


Fig. 3. Practical procedure of the inclusion rating by statistics of extreme values.

Murakami et al.'s papers [3,5,9–11]. The procedure is briefly explained in the following, see [11].

(1) A section perpendicular to the maximum principal stress is cut from the specimen. After polishing with a n°2000 emery paper, the test surface is mirror-finished with buff.

(2) A standard inspection area S_0 (mm^2) is fixed. Generally, it is advisable to take a microscope picture for an area approximately equivalent to S_0. In the area S_0, the inclusion of maximum size is selected. Then, the square root of the projected area $\sqrt{area}_{max}$ of this selected inclusion is calculated. This operation is repeated n times (in n areas S_0) (see Fig. 2).

(3) The values of $\sqrt{area}_{max,j}$ are classified, starting from the smallest, and indexed: (with $j = 1...n$). We then have the following relation:

$$\sqrt{area}_{max,1} \leq \sqrt{area}_{max,2} \leq ... \leq \sqrt{area}_{max,n}.$$

The cumulative distribution function F_j and the reduced variates y_j are then calculated from the equations.

$$F_j = j \times 100/(n+1)$$

$$y_j = -\ln[-\ln(j/(n+1))].$$

(4) The data are then plotted on probability paper. The point j has an abscissa coordinate of $\sqrt{area}_{max,j}$ while the ordinate axis represents either F_j or y_j. An example of the curve is shown in Fig. 4.

Figure 4 shows the inclusion ratings by IRMSE for two kinds of super-clean bearing steels, SUJ2(N) and SUJ2 (H). The total oxygen contained in these steels is 8 ppm for SUJ2(N) and 5 ppm for SUJ2(H). This kind of information

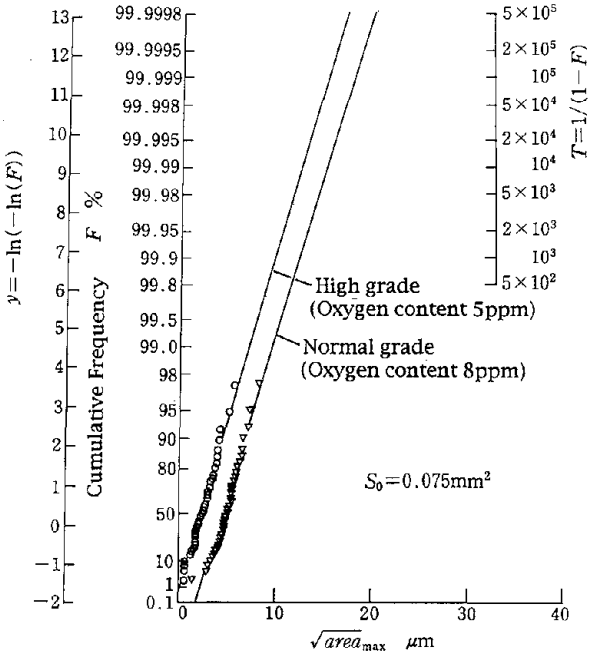


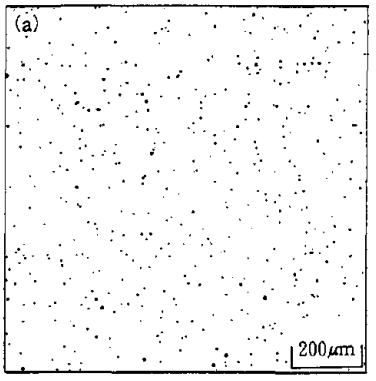
Fig. 4. Cumulative frequency of the extreme values of inclusions [Super-clean bearing steels, SUJ2(N) and SUJ2(H)].

enables one to discriminate quantitatively the difference among the cleanliness levels of the same kind of materials produced by different companies or produced by a company at different periods. Thus, this information will be useful for the quality control of materials and the improvement of the steel making process.

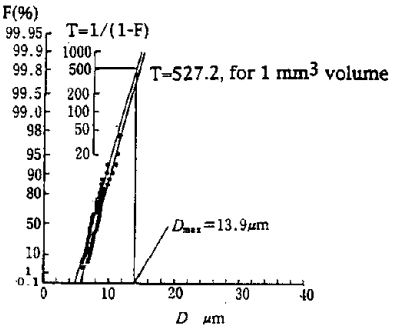
It is not *a priori* evident to what extent the extreme values $\sqrt{area}_{max}$ of inclusions contained in various steels follow extreme statistics value. However, Murakami et al. [3,5–11] have shown many examples of measurements which obey the statistics of extreme value theory. Uemura and Murakami [12] carried out a three-dimensional numerical simulation to find the statistical distribution of the extreme values $\sqrt{area}_{max}$ of inclusions which were distributed in a constant volume with the size (D) distribution of the type,

$$\phi(D) = \frac{1}{m} \exp\left(-\frac{D}{m}\right),$$

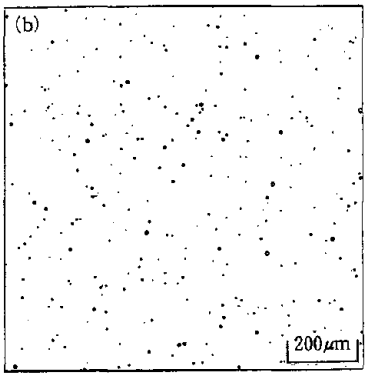
where m is the mean value, and they confirmed the validity of IRMSE (Fig. 5). In addition, they indicated the quantitative difference between two-dimensional and three-dimensional measurements, though the difference virtually vanishes with increasing inspection domains.



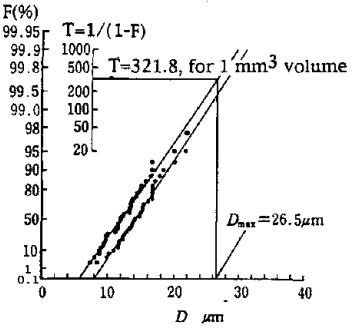
(a-1) Inclusion distribution on inspection section of the material with m=1 μm.



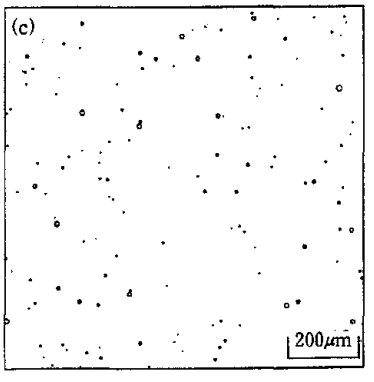
(a-2) Statistical distribution of true(3-D) and apparent(2-D) extreme values of inclusions for material with m=1 μm (S_O=0.25 mm^2, number of inspections = 40).



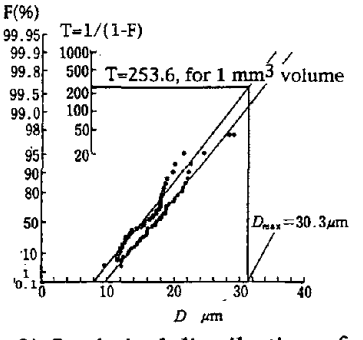
(b-1) Inclusion distribution on inspection section of the material with m=2 μm.



(b-2) Statistical distribution of true(3-D) and apparent(2-D) extreme values of inclusions for material with m=2 μm (S_O=0.25 mm^2, number of inspections = 40).



(c-1) Inclusion distribution on inspection section of the material with m=3 μm.



(c-2) Statistical distribution of true(3-D) and apparent(2-D) extreme values of inclusions for material with m=3 μm (S_O=0.25 mm^2, number of inspections = 40).

Fig. 5. Numerical simulation of the inclusion rating by statistics of extreme values on the materials with the inclusion size distribution of the type $\phi(D)=\frac{1}{m}\exp\left(-\frac{D}{m}\right)$.

4. Application to Prediction of Scatter Band of Fatigue Strength

Figure 6 illustrates the shape and dimension of a tension-compression fatigue specimen [13]. The material used is tool steel, SKH51. The chemical composition is shown in Table 1. Table 2 shows the mechanical properties.

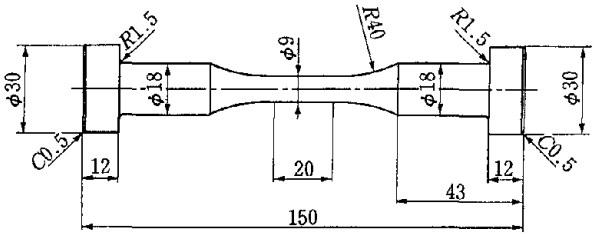


Fig. 6. Shape and dimension of tension-compression fatigue specimen (mm) (Tool steel, SKH51).

Figure 7 shows the extreme value distribution of $\sqrt{area}$ of the inclusions found at the fracture origin of 34 specimens. The data in Fig. 7 are the extreme values obtained by the fatigue test but not by the two-dimensional metallographic method described in Sec. 3. Figure 8 indicates the location of these inclusions on the fracture surface. If the tension-compression fatigue test is not performed correctly, that is, specimens are subject to a bending moment due to a bad alignment or the curving of the specimen axis, nonmetallic inclusions existing near the free surface are likely to appear as the fracture origin on the fracture surface [14]. In such a case, unusually low fatigue strength is likely to be obtained. Since the fatigue fracture origins shown in Fig. 8 are distributed randomly on the section of specimen, these data may be valid for the statistical analysis. However, it should be noted that when

Table 1. Chemical composition in wt% of material (Tool Steel, SKH51)

C	Si	Mn	P	S	Cr	W	Mo
0.81	0.31	0.29	0.018	0.002	3.92	6.10	4.85
V	Co	Cu	Ca	Al	Mg	O	
1.81	0.46	0.07	0.004	0.035	0.0005	0.0018	

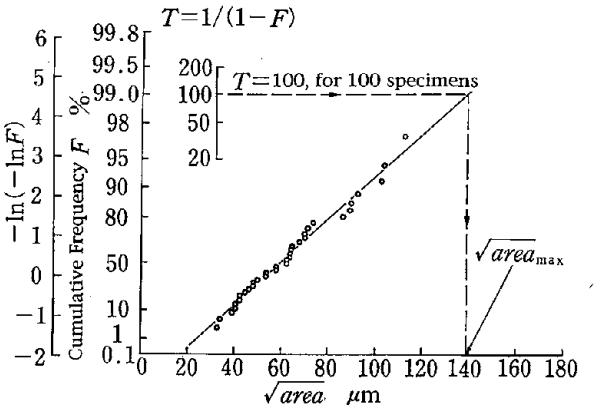


Fig. 7. Statistical distribution of the extreme values, the maximum size of inclusion at the center of fracture origin (Tool steel, SKH51).

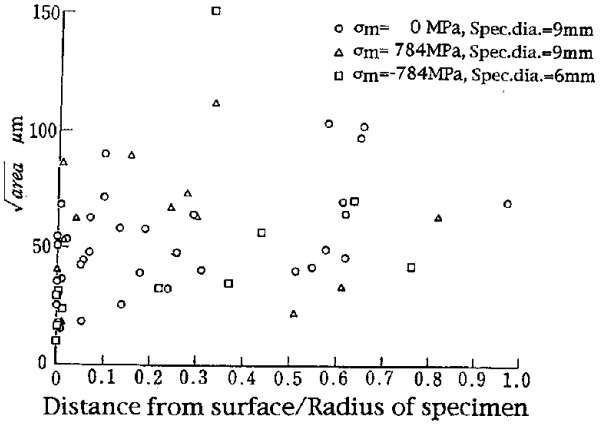


Fig. 8. Relationship between the size ($\sqrt{area}$) and location of inclusions at the center of fracture origin (Tool steel, SKH51).

Table 2. Mechanical properties of quenched and tempered test material (Tool steel, SKH51)

Heat treatment	0.2% Proof stress (MPa)	Tensile strength (MPa)	Elongation (%)	Reduction of area (%)	Vickers hardness H_v (kgf/mm²)
Heat treat. 1	1 820	2 110	2.9	3.7	615
Heat treat. 3	2 270	2 560	2.0	0	654

the surface inclusions became the fracture origins, the data were not plotted on Fig. 7, because such inclusions are more detrimental than an inclusion having the same size and existing internally and accordingly they may be a little smaller than the exact maximum inclusion.

In the case of the data of Fig. 7, the volume of the test part of one specimen (Fig. 6) corresponds to one inspection domain and there are 34 extreme values in Fig. 7. Therefore, Fig. 7 can be used for predicting the expected maximum size of the inclusion which may be contained in more specimens than those used in fatigue tests. For example, an inclusion having $\sqrt{area}_{max} \cong 138$ μm is expected to be contained in 100 specimens ($N=100$). Combining this $\sqrt{area}_{max}$ ($=138$ μm) and Eq. (3), the lower bound (σ_{we}) of fatigue strength of 100 specimens can be predicted.

Figure 9 compares the scatter observed in experiments and the predicted lower bound σ_{we} of the scatter band. The prediction is in good agreement with experiments. The prediction of the lower bound of fatigue strength explained above can be used for the quality control of machine elements which are produced by mass-production and cannot be tested individually.

The data as shown in Fig. 7 offer us reliable information on inclusions expected to be contained in other specimens. However, obtaining the data shown in Fig. 7 requires preparation of many precise specimens and time consuming fatigue tests. To avoid this inconvenience, the author has proposed an alternative two-dimensional method as

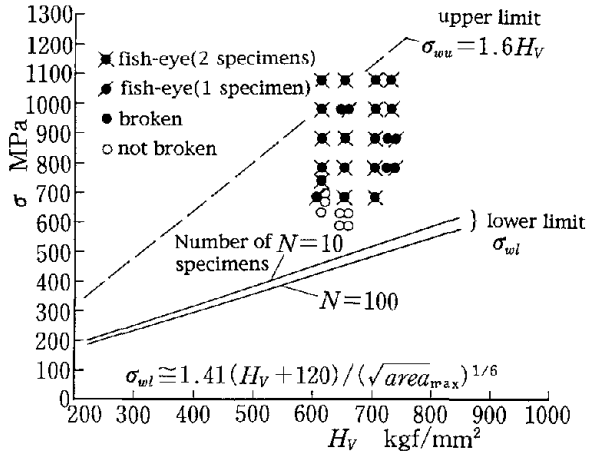


Fig. 9. Comparison between the experimental results and the lower bound of fatigue strength which was predicted on the basis of Eq. (3) and the maximum size of inclusion (Tool steel, SKH51).

explained in Sec. 3. A sufficient number (N) of inspection domains (inspection areas) necessary to predict reliably $\sqrt{area}_{max}$ for more specimens or larger areas should depend on the materials to be inspected and on the inspection area S_0 observed by the image processor combined with an optical microscope. From the author's experience, it is recommended that N be larger than 40 for $S_0=0.031$ mm².

Several Japanese industries have already put the method proposed in this study in practice [15].

5. Conclusions

(1) If we define the size of nonmetallic inclusions contained in commercial steels by the square root of the projected area, $\sqrt{area}$, the maximum values, $\sqrt{area}_{max}$, in a definite inspection domain obey the statistics of extreme value theory.

(2) The inclusion rating method by the statistics of extreme values (IRMSE) based on $\sqrt{area}_{max}$ can be used for a new inclusion rating method. IRMSE enables one to discriminate distinctly between recent super-clean steels, while conventional inclusion rating methods are no longer valid as the method to evaluate the cleanliness of new clean steels.

(3) IRMSE is useful not only for a relative evaluation of materials but also for the prediction of the expected maximum size of inclusions to be contained in a domain larger than the inspection domain. The value of $\sqrt{area}_{max}$ can be used with the fatigue strength prediction equation to predict a scatter band of fatigue strength of high strength steels.

6. References

[1] E. J. Gumbel, Statistics of Extremes, Columbia University Press, New York (1957).
[2] Y. Murakami, S. Kodama, and S. Konuma, Internat. J. Fatigue **11**(5), 291–298 (1989).
[3] Y. Murakami and H. Usuki, Internat. J. Fatigue **11**(5), 299–307 (1989).
[4] Y. Murakami, K. Kawakami, and W. E. Duckworth, Internat. J. Fatigue **13**(6), 489–499 (1991).
[5] Y. Murakami, T. Toriyama, Y. Koyasu, and S. Nishida, J. Iron Steel Inst. Jpn. **79**(6), 60–66 (1993).
[6] Y. Murakami and M. Endo, The $\sqrt{area}$ Parameter Model for Small Defects and Nonmetallic Inclusions in Fatigue Strength: Experimental Evidences and Applications, Theoretical Concepts and Numerical Analysis of Fatigue, Proc. Conf. held 25–27th May 1992, Birmingham Univ., A. F. Blom and C. J. Beevers, eds., EMAS, West Midlands, U.K. (1993) pp. 51–71.

[7] Y. Murakami and T. Toriyama, The $\sqrt{area}$ Parameter Model for Quantitative Evaluation of Effects of Nonmetallic Inclusions on Fatigue Strength, Proc. Fatigue 93, J. P. Bailon and J. I. Dickson, eds., Vol. I (1993) pp. 303–309.
[8] Y. Murakami, Impact of Improved Materials Quality on Properties, Product Performance, and Design, ASME Winter Annual Meeting, MD-Vol. 28, 1991, pp. 89–102.
[9] Y. Murakami, Metal Fatigue: Effects of Small Defects and Nonmetallic Inclusions, Yokendo Ltd., Tokyo (1993).
[10] Y. Murakami, K. Kawakami, and M. Saito, J. of Spring, Japan Society of Spring, Vol. 35 (1990) pp. 1–7.
[11] Y. Murakami, T. Toriyama, and E. M. Coudert, Instructions for a New Method of Inclusion Rating and Correlation with the Fatigue Limit, J. Testing Evaluation **22**(4), 318–326 (1994).
[12] Y. Uemura and Y. Murakami, Trans. Jpn. Soc. Mech. Eng. **56**(521), 162–167 (1990).
[13] Y. Natsume, S. Miyakawa, Y. Uemura, and Y. Murakami, Proceedings of the Fourth International Conference on Fatigue and Fatigue Thresholds, Fatigue 90, Vol. I, 349–354 (1990).
[14] A. Melander, M. Rolfson, A. Nordgren, B. Jansson, H. Hedberg, and T. Lund, Reprint from Swedish Institute for Metals Research, No. IM-2589 (1990) pp. 1–71.
[15] H. Narai, C. Abe, and K. Furumura, Current Advances in Materials and Processes, Report of the Iron and Steel Institute of Japan Meeting, Vol. 4 (1991) pp. 1178–1181.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 353 (1994)]

Critical Levels of Ozone Over the United Kingdom: Mapping Aggregate Exceedances Over Moderate to High Thresholds

Volume 99 Number 4 July–August 1994

R. I. Smith

Institute Terrestrial Ecology, Edinburgh Research Station, Bush Estate, Penicuik, Midlothian EH26 0QB, Scotland

C. W. Anderson

University of Sheffield, School of Mathematics and Statistics, P.O. Box 597, Sheffield, S10 2UN, England

and

D. Fowler

Institute Terrestrial Ecology, Edinburgh Research Station, Bush Estate, Penicuik, Midlothian EH26 0QB, Scotland

The critical level for ozone, above which it has a detectable effect on biological targets, is potentially to be set by the United Nations Economic Commission for Europe at 300 nL·h/L hours per annum over 40 nL/L. It is therefore important to determine the aggregate exceedance over 40 nL/L throughout the United Kingdom. Over most of the UK, ozone concentrations are unknown so we rely on our understanding of the atmospheric processes and on the statistical properties of ozone concentrations to interpolate between monitoring sites. This paper describes the application of statistical models derived for storm severity data to the ozone data for the United Kingdom. Aggregate excess distributions were fitted to data from all rural monitoring sites using a Weibull model with a 40 nL/L threshold. At this threshold the scale parameter has a spatial interpretation, but, with higher thresholds, there were problems with missing data and small scale spatial effects were not detected. The approach appears successful for all except very large aggregate exceedances which deviate from the Weibull predictions.

Key words: aggregate excess distribution; critical level; mapping; ozone.

Accepted: March 22, 1994

1. Introduction

The major public concern with ozone, O_3, in Europe has focused recently on the existence of "ozone holes" in the stratosphere caused by the depletion of ozone as a consequence of chlorofluorocarbon emissions. Ozone is also present in the troposphere and in the planetary boundary layer at concentrations, i.e., volume fractions between 10 and 200 nL/L (i.e., parts per billion, ppb = 10^{-9}). In the second half of the last century European mean concentrations ranged between 10 ppb and 15 ppb [8]. Current mean concentrations are about twice these values and ozone episodes with peak concentrations between 100 ppb and 200 ppb occur, a level known to cause damage to many plant species. Episodes happen if the precursor gases for photochemical ozone production (oxides of nitrogen, NO and NO_2, and volatile organic compounds, VOCs) are present in suitable meteorological conditions for the chemical reactions to occur (ideally hot summer days with clear skies and low wind speeds).

The description of spatial patterns in exposure of vegetation to ozone over Europe has been hampered by the limited availability of monitoring data, the very large spatial variability in ozone concentrations and a poor understanding of the underlying mechanisms regulating the ozone exposure of terrestrial ecosystems. Defining a threshold for phytotoxicity is not simple. However, at 60 ppb of ozone there is little doubt that there is a clear contribution from photochemical production in polluted air and a map of hours over 60 ppb for Europe (Fig. 1) is a guide to some broad trends [4]. In a large area north of the Alps, covering most of Germany and parts of neighbouring countries, 200 hours per year above 60 ppb is common. North and west of this area the annual duration of exposure declines but to the east there is so little information available that mapping is uncertain. The Mediterranean zone of high ozone exposure reflects recent work showing that ozone episodes are common events in this region but the levels have not yet been well quantified.

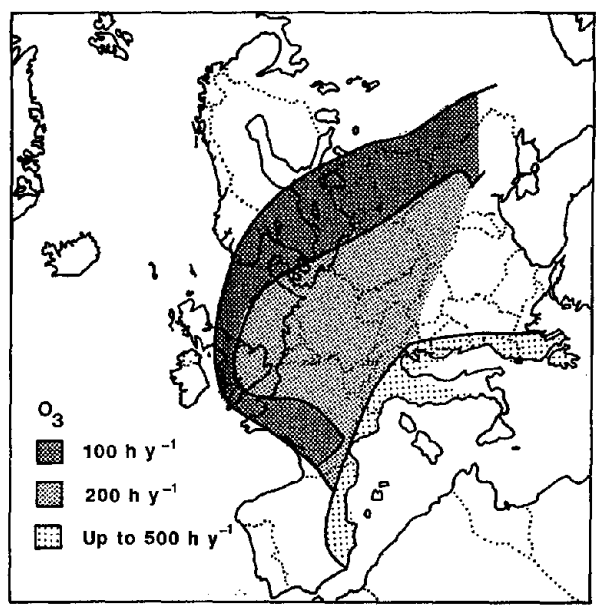


Fig. 1. Hours when ozone exceeds 60 ppb.

Although the meteorological conditions leading to ozone episodes are similar in a general sense at all sites, the climates of northern and southern Europe lead to very different patterns of events. In northern Europe, typical episodes occur when a stationary spring or summer anticyclone provides the conditions for ozone production from the emitted precursor gases to add appreciably to the background concentration of about 30 ppb. Typical production rates give net increases of 10 ppb to 20 ppb per day and a succession of 8 to 10 such days leads to peak concentrations of 150 ppb to 200 ppb. Often in northern Scandinavia, Britain, Ireland and western France the ideal meteorology exists but in the absence of upwind precursors. In Germany and central Europe almost all wind directions provide the precursors and hence the NW-SE gradient in episodes. In southern Europe, the meteorological conditions are more stable and episodes can occur every day for long periods. However the effects of both sea breezes and the development of intense thermal low pressure areas on the air circulation causes very variable patterns of ozone exposure.

Superimposed on this two dimensional surface there is a daily cycle in ozone concentration which is a very important and variable feature. At low altitude inland sites a marked diurnal variation (of the order of 30 ppb) is observed but at high elevation the amplitude of the diurnal cycle gradually reduces to less than 5 ppb at mountain tops. Figure 2 shows data for 1 day at both Great Dun Fell (847 m above sea level) and Wharleycroft (206 m above sea level), two sites which are less than 10 km apart. Hill tops are generally windy sites at which the terrestrial surfaces are well connected to the free troposphere and where the downward supply of ozone to the surface exceeds the rate of deposition. At low level sites the thermal stratification of the atmosphere with the development of a nocturnal inversion restricts the supply of ozone from above during the night and morning. In these conditions both deposition to the surface and the nocturnal atmospheric chemical titration of ozone with nitric oxide causes the surface concentrations of ozone to decline, potentially to negligible levels. At coastal sites the effects of land and sea breezes strongly modify the ozone exposure of the ground.

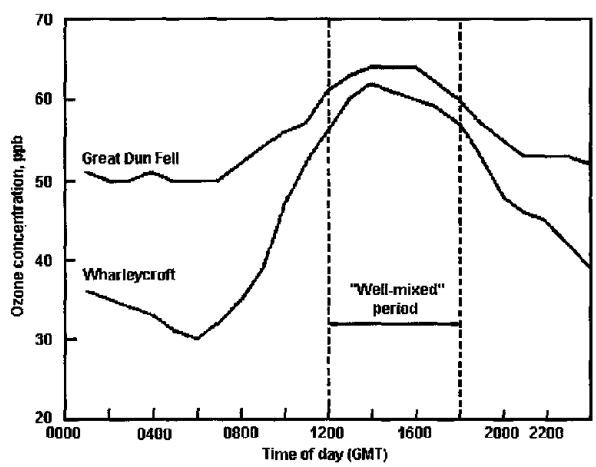


Fig. 2. Typical altitude effect on the diurnal cycles of ozone concentration.

The main concern with rural ground level ozone concentrations is the damage which can be caused to plants and to human and animal health.

For vegetation, some sensitive species show visible or physiological effects following exposure to 40 ppb or 50 ppb [3]. However, the effect of exposure can be modified by the presence of other atmospheric pollutants and, since ozone causes damage to vegetation through stomatal uptake, by nutritional status, light, temperature and humidity. There is genetic variability in the ozone response within species as well as between species and, although considerable attention has focused on crops and forests, little is known about the impact on semi-natural vegetation. Timing of the exposure within the life of the plant can be important as can be the time for recovery between exposures [6]. The United Nations Economic Commission for Europe is considering a tentative proposed critical level of 300 ppb · h above 40 ppb during daylight hours for the growing season of the vegetation. A critical level is defined as one below which ozone has no detectable effect. However, there are a number of outstanding issues which it is hoped to resolve by the end of 1993 and the adopted critical level may well be different. The proposed critical level would probably be exceeded in most of Europe at present.

Concern for human health in the UK at the current levels of ozone is growing but better assessments of population and individual exposure are thought necessary [6]. This aspect may in time be the main argument for emission controls of the major precursor gases.

It is important to differentiate between the dose which a plant or human receives, that is incorporated into the individual's system by some method, and its exposure, that is the level in the atmosphere around the individual. In this paper current methods for determining plant exposure in the UK are described and then the potential application of extreme value theory is explored.

2. Ozone Exposure Maps of the UK

Ozone exposure has recently been mapped for the UK at three concentration thresholds: 40 ppb, 60 ppb and 90 ppb [5]. There were about 17 rural or semi-rural monitoring stations between 1987 and 1991 which recorded hourly mean concentrations in Britain and Ireland (Fig. 3). As the differences between sites which were geographically close was as large as the differences between geographically distant sites, a straight spatial interpolation between sites gave a map similar to that in Fig. 4.

In the summer months, taken as April to September, during the part of the day when the atmospheric boundary layer was well mixed by turbulence, ozone concentrations at neighbouring sites were very similar. For each threshold, an empirical relationship was derived between the hours over the threshold for the whole day and the hours over the threshold for the well mixed period, taken to be 1200 to 1800 GMT. For the threshold at 60 ppb, the relationship was

$$h_{60} = (1.3 + 0.0021\, z)\, t_{60}, \qquad (1)$$

where h_{60} was the total hours over 60 ppb, t_{60} was the hours over 60 ppb between 1200 and 1800 GMT and z was the altitude of the location in meters. This relationship was applied to the spatial interpolation of hours over 60 ppb for 1200 to 1800 GMT to provide a map (Fig. 5) with clear topographical influence. The coastal effect, which can extend for 5 km to 20 km inland depending on meteorological conditions, was ignored; typically coastal ratios were around 2 rather than 1.3. The maps are only for the summer months, April to September, but ozone levels very rarely exceed 40 ppb during the remainder of the year.

This approach emphasizes the spatial variability of ozone exposure within small areas. The relationships for the different thresholds are empirical and must be recalculated for each threshold and they do not provide a general description of high concentration events. Direct estimates of exposure in terms of a dose measurement like ppb.hours are not available although a minimum estimate could be made. If the decision were made to set different windows to match the growing seasons of different vegetation types, the whole procedure could be difficult to implement.

3. Modelling Aggregate Excess

In work on flood levels for the River Thames, Anderson and Dancy [2] modelled the aggregate excess, that is the sum of the exceedances over a threshold, within a cluster using a Weibull distribution. There are similarities between ozone data and flood level data. A Pareto distribution has been shown to predict the peak excesses of ozone concentrations at a rural site using a threshold of 40 ppb [7]. There is a seasonal component in the data, since high concentrations rarely occur over the winter period, but this has not been modelled at present. Also ignored was the probable increase in mean values of ozone concentration over the time

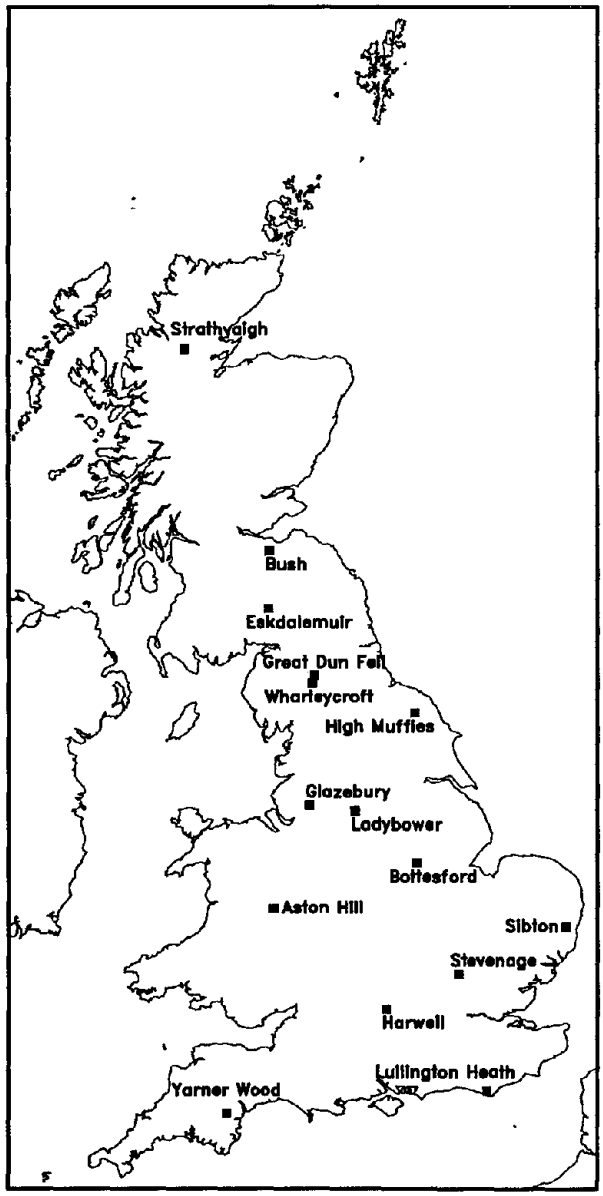


Fig. 3. Locations of 15 monitoring sites on the UK mainland (2 sites, Lough Navar and Mace Head, are on Ireland).

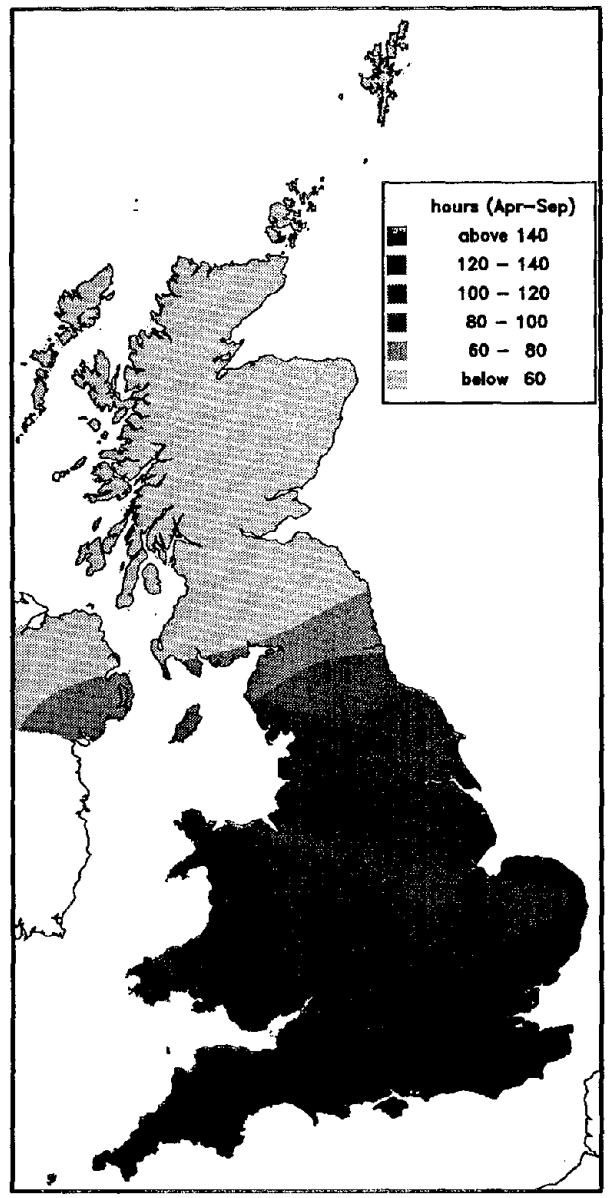


Fig. 4. Interpolated map of the number of hours when ozone exceeds 60 ppb.

period of data collection, as this increase was small compared to both the diurnal fluctuations and the accuracy of the recording methods.

Anderson [1] has looked at ozone data for one site, Stevenage, for a longer time period. There was evidence of nonstationarity in that data set and he shows that there is a need for temperature, or some similar measure, as a covariate. This problem is still under investigation but for the time period considered in this paper, 1986 to 1991, no covariate has been used. Anderson also derived a method of extrapolating to higher thresholds than those used in fitting the models, a very useful tool for determining exposures to plants with different sensitivities to ozone.

The data for the 17 sites have been fitted using a single threshold of 40 nL/L and a Weibull distribution for the aggregate excess. The data were declustered using a minimum time separation of 48 hours. There were about 100 clusters for the sites with relatively complete data sets. Some sites were not operational in the earlier period of collection and one site had only 3 years data.

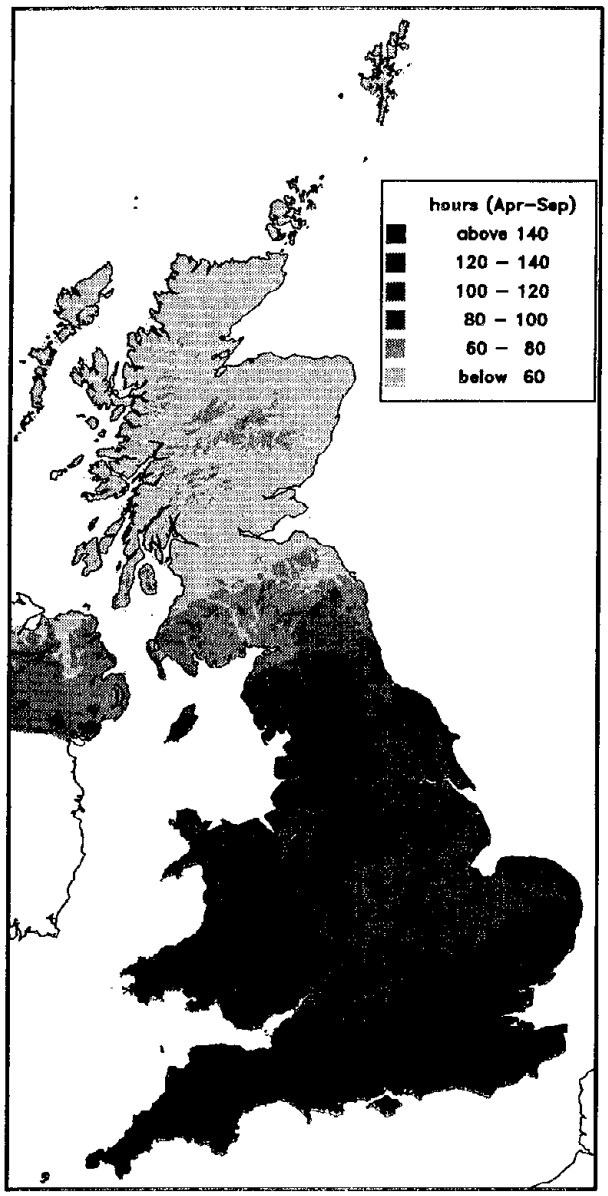


Fig. 5. Altitude adjusted interpolated map of the number of hours when ozone exceeds 60 ppb.

The two parameter Weibull model

$$P(S > s) = \exp(-\alpha \cdot s^{\theta}) \qquad (2)$$

was fitted. The shape parameter, θ, varied between 0.4 and 0.6 for all sites. When θ was constrained to the value 0.5, there were only slight increases in the values of the likelihood function. The spatial variation was therefore explored using only the scale parameter α.

The Q-Q plots showed that, as expected, the fit of the Weibull model varied from site to site. There were some very straight line plots but there were also shapes typically illustrated by the plot for Ladybower (Fig. 6). Most of the data were on a reasonably straight line but the Weibull distribution underpredicted a few data points, usually no more than five, at the higher values.

The shape parameter, α, from the fitted Weibull model was clearly related to a SE-NW trend across the country. To investigate this further, the α values were regressed on other available data.

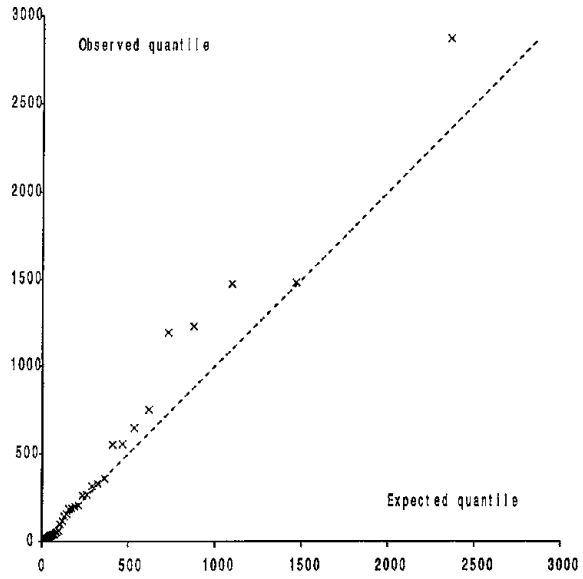


Fig. 6. Q–Q plot for Weibull model (40 ppb threshold, 48 hour separation) fitted to data from Ladybower.

The sites were referenced to a line from Lullington Heath, a site on the south-east coast of England, to Strathvaigh in the north of Scotland using two variables, n_{mod}, the distance north-west along the transect, and e_{mod}, the perpendicular distance from the transect with positive values being to the east. The actual distances were divided by the length of the transect to give manageable numerical values. The relationship between α and n_{mod} was non-linear and the simplest function of n_{mod} which fitted well was n_{mod}^4. A linear function of e_{mod} improved the fit. One site, Bottesford, had a high residual. This site has peculiar local features which can give it the characteristics of an urban site and was removed from the data. The subsequent regression equation

$$\alpha = 0.091 + 0.083\ n_{mod}^4 - 0.023\ e_{mod} \qquad (3)$$

explained about 85% of the variation and produced an acceptable residual pattern. The 2 remote sites, Strathvaigh and Mace Head (on the west coast of Ireland) had the most influence on the fit.

When a threshold of 60 ppb was used, there were problems in fitting the model to the data. Each site had only about four independent clusters per year and fitting a Weibull model was very difficult since the likelihood surface was quite flat. Results were obtained by assuming θ was the same for a 60 ppb threshold as for a 40 ppb threshold. However, it became apparent that there were potentially two difference sources for the higher exceedances and that separation of these sources was critical with the increased threshold level.

4. Discussion

The results of these fits are encouraging although a number of problems have occurred. The Weibull distribution with a threshold of 40 ppb and θ fixed at 0.5 gives an interpretable underlying pattern for the whole country. The SE-NW gradient would be expected. Areas to the east of the chosen transect are more influenced by air masses from continental Europe and would be expected to have more ozone episodes. The lack of detection of an altitude effect at this threshold is not entirely surprising as high altitude sites can have mean ozone concentrations quite close to this threshold. However an altitude effect would be expected at a higher threshold.

At the 60 ppb threshold two main problems occur. The first, and possibly the most important, is lack of data. The declustering algorithm which has been used takes the rather simple approach of removing clusters with missing data. When monitoring stations are running continuously, usually recording several times per minute, there are a whole series of glitches which can occur in the data for reasons wholly unconnected with the concentration values. In particular there may well be a series of instrumentation tests which usually occur during the working day and often at least once per week. Data capture rates of over 90 % on hourly values are regarded as good but not all sites on the network are achieving these rates. Therefore, careful decisions on the treatment of missing data are likely to give more information for analysis.

The second problem is one of determining whether there are two distinct distributions required to model threshold exceedances or whether the Weibull model is the wrong approach. If a meteorological covariate is introduced [1], it is not clear where it should be measured. Rapid ozone production can be occurring 50 km or 100 km downwind in good sunny conditions but the monitoring site may be sitting in quite a different climate. Clearly, further investigation of the peak and close to peak values will be required.

Although large exposure to ozone can be accumulated by a plant at concentrations over 100 ppb, these are relatively rare occurrences in the UK and are often, if not always, associated with very dry conditions. How much of the ozone will enter the plant's system, given that the plant is probably under considerable water stress by the afternoon period, is not clear. Even if the models do not perform very well at the highest exceedances, if they can perform reasonably well for the remainder of the exceedances they could be of considerable benefit when critical levels for vegetation are considered. For human health problems, of course, a different perspective is required.

This approach, when combined with a model of time between clusters, has the potential of producing valuable information for the assessment and mapping of critical levels for vegetation. However, some further progress is required with models for the 60 ppb threshold and with identification of local scale variability in concentration levels.

Acknowledgments

Rognvald Smith and David Fowler gratefully acknowledge the support of the United Kingdom Department of the Environment for this work under contract number PECD 7/12/132.

5. References

[1] C. W. Anderson, The aggregate excess measure of severity of extreme events, J. Res. Natl. Inst. Stand. Technol. **99**, 555 (1994).

[2] C. W. Anderson and G. P. Dancy, The severity of extreme events, Research Report 409/92, Department of Probability and Statistics, University of Sheffield, United Kingdom (1992).

[3] M. R. Ashmore, J. N. B. Bell, and C. L. Reily, A survey of ozone levels in the British Isles using indicator plants, Nature **276**, 813–815 (1978).

[4] D. Fowler, J. N. Cape, R. I. Smith, I. Leith, and J.-W. Erisman, The pollution climate of Europe, in Air pollution and crop responses in Europe (Proceedings of a conference at Tervuren, 23–25 November 1992, organised by the CEC and the Institute for Chemical Research, in press) (1994).

[5] United Kingdom Photochemical Oxidants Review Group, Ozone in the United Kingdom 1993, Third report of the United Kingdom Photochemical Oxidants Review Group, Chapter 8, Effects of ozone upon man, vegetation and materials, Department of the Environment, London (1993).
[6] United Kingdom Photochemical Oxidants Review Group, Ozone in the United Kingdom 1993, Third report of the United Kingdom Photochemical Oxidants Review Group, Chapter 2, Ozone measurements, Department of the Environment, London (1993).
[7] R. I. Smith, D. Fowler, and J. N. Cape, The statistics of phytotoxic air pollutants, J. R. Statist. Soc. A **152**, 183–198 (1989).
[8] A. Volz and D. Kley, The evaluation of the Montsouris series of ozone measurements made in the 19th-century, Nature **332**, 240–242 (1988).

About the authors: *Rognvald I. Smith is a biometrician at the Edinburgh Research Station of the Institute of Terrestrial Ecology. Clive W. Anderson is a lecturer in statistics in the School of Mathematics and Statistics at the University of Sheffield. David Fowler is an environmental physicist and leader of the research section on trace gas fluxes and air pollution at the Edinburgh Research Station of the Institute of Terrestrial Ecology.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 361 (1994)]

Extreme Value Estimation Applied to Aerosol Size Distributions and Related Environmental Problems

Volume 99 Number 4 July–August 1994

Philip K. Hopke
Department of Chemistry,
Clarkson University,
Potsdam, NY 13699-5810

and

Pentti Paatero
Department of Physics,
University of Helsinki,
Helsinki, Finland

This work examines the potential connections between extreme value statistics, problems in aerosol science, and a recent technique of solving ill-posed inversion problems, called EVE (Extreme Value Estimation). EVE estimates functionals of the unknown solution by searching the extreme (maximum and minimum) values of that functional within a set of acceptable solutions. The statistics of occurrence of extreme values in real life were not considered when this method was developed. The results of this technique are more conservative than those of the other methods used to solve the problem of aerosol size distribution estimation like non-linear least squares, expectation-maximization, regularization, etc. The utilization of the customary methods of deconvolution may lead to an underestimation of the possibility of occurrence of extreme values in real life. It is suggested that consideration of extreme value statistics might aid in better defining the limits to be placed on the physically acceptable solutions in the EVE deconvolution. Other problems could also benefit from the application of extreme value statistics including the estimation of the second highest value of measured airborne particle mass in the context of the ambient air quality standard for particulate matter less than 10 μm and the determination of the Maximally Exposed Individual as required under the 1990 revisions to the Clean Air Act.

Key words: aerosol mass concentrations; aerosol size distributions; deconvolution algorithms; maximally exposed individual; particulate matter 10 μm.

Accepted: March 22, 1994

1. Introduction

Although extreme value statistics has been applied to environmental phenomena such as maximum wind speed and wave heights, it has not been applied to air pollution regulations, concentration estimation, or other related problems. Since many of the problems related to the effects of pollution on public health and welfare are dependent on the high end of the distribution of concentrations and/or exposures, there appears to be an opportunity to bring the developments in extreme value statistics to an area that could make good use of such methods. In this paper, three possible applications of extreme value statistics will be presented with the hope of sparking interest in bringing these tools to bear on some difficult but interesting problems.

2. Aerosol Size Distribution Estimation

One common problem in aerosol science is the estimation of the aerosol particles size distribution from measurements of their aerodynamic behavior (penetration or deposition) through a separation device. For small particles (<300 nm), the penetration through a device is governed by the particle's diffusivity while for large particles (>300 nm), inertial impaction is the usual separation mechanism. The response of the device is known either by calculation or measurement using particles of known size. For the unknown aerosol, the penetration is measured through a series of stages that sequentially remove additional particles. From the known characteristics and a limited number of measurements, the size distribution of the aerosol is

estimated. In general there are fewer measurements than parameters to be estimated and there can be collinearity problems in the penetration matrix describing the instrument to complicate the problem further. There are a number of conventional approaches to providing a solution, but since the problem is underdetermined, one cannot insure that they will provide the true solution. It is also difficult to estimate error bounds for these solutions.

2.1 Conventional Methods

The observed sequence of particle concentrations penetrating through each stage of a size segregating device contains information on the size distribution of that aerosol. In general, the number of particles penetrating through a given stage of the system can be expressed by

$$N_i = N_0 \int_0^\infty P(i, \mathrm{d_p})\, f(\mathrm{d_p})\, \mathrm{d}d_\mathrm{p} + \epsilon_i \qquad (1)$$

where N_i is the concentration penetrating through the ith stage, $P(i,d_\mathrm{p})$ is the known particle size penetration characteristics for particles of diameter d_p through stage i, $f(d_\mathrm{p})$ is the size distribution function to be estimated, and ϵ_i is the error in fitting the measurement.

The normal approach to solving this equation is to express it as a series of linear, simultaneous equations relating the particle penetration fraction to discrete values of the size distribution and the stage penetration functions.

$$N_i = \sum_{j=1}^{J} P_{ij} \cdot f_j \qquad i = 1, \ldots, I \qquad (2)$$

where I is the number of stages in the device, J is the number of size interval midpoints in the distribution, P_{ij} is the penetration of the jth particle size, $d_\mathrm{p}(j)$, through the ith stage, and N_i is the number of particles penetrating the ith stage. The f_i values must be nonnegative. However, there is generally no other objective *a priori* information on the nature of the distributions. The size distribution is not normalized so that

$$N_0 = \sum_{j=1}^{J} f_\mathrm{j}\,, \qquad (3)$$

where N_0 is the total airborne concentration that is being partitioned into the various size intervals. Equation 2 can be rewritten in matrix form.

$$N = P \cdot f + E\,. \qquad (4)$$

If I is greater than or equal to J, then the problem is overdetermined and can be solved for a unique solution using methods such as least squares. However, because the size distributions typically cover several orders of magnitude in particle diameter, it is normally necessary to estimate more midpoint values than measurements ($I < J$). There is then no unique solution to the problem.

Because collection by diffusion varies slowly with particle size, the penetration values for adjacent size ranges are often quite similar to one another. The penetration functions for a screen diffusion battery used for separating particles in the 0.5 nm to 500 nm range generally have substantial collinearity and thus, the problem is ill-conditioned as well as underdetermined [1]. Phillips [2] concluded that direct inversion of theses equations rarely produces physically acceptable solutions.

Two techniques for solving the ill-posed set of equations have been developed by Twomey [3] and by Maher and Laird [4]. There is limited theoretical justification for these methods. In practice, however, they have been widely used in the aerosol field with satisfactory results in many cases. Different variations of the Twomey algorithm have been proposed (e.g., [5]).

Other approaches have sought specific solutions within the feasible solution space by incorporating additional constraints on the problem. For example, Wolfenbarger and Seinfeld [6] assume that the distribution is fully smooth from one interval to another. However, it is certainly possible to have aerosol sources that produce particles with a very narrow initial distribution and thus, the overall aerosol size distribution may not be truly smooth. Thus, in all of these solution methods, a solution, but not necessarily the solution will be obtained.

2.2 Extreme Value Estimation

Replogle et al. [7] initially suggested the concept that the primary "solution" is the set of all those points that could produce the observed values. Paatero [8,9] recognized that this approach could be applied to the aerosol inversion problem by considering a one-to-many mapping of the measured N onto f such that there is the set $D(N)$ of possible solutions corresponding to each possible measured N. The set $D(N)$ is defined as the collection of all such solutions f that allow the reproduction of the measured N by Eq. (4) when reasonable values are used for E. Then the true unknown solution f is a member of the set $D(N)$ with a high probability. $D(N)$ is then the set of acceptable solutions.

To initiate the analysis a best fit, f_0, is calculated such that the nonnegative constraints are satisfied. Additional solutions are calculated that are sufficiently close to the best fit estimation that they fall within a criterion for acceptable solutions. For each of the estimated quantities, the largest and smallest values within the set $D(N)$ are taken as the bounds of the confidence interval in which the true solution will fall at some high probability.

The question is then how to define what solutions are acceptable. The likelihood function, $L(N,f)$ is the probability of observing N when f is given. It will be assumed that

$$-\ln(L) = const \cdot \sum_{i=1}^{J} \left| \frac{E_i}{S_i} \right|^2 = const \cdot Q(f) , \quad (5)$$

so that $Q(f)$ is the sum-of-squares for the case in which f and N are substituted into Eq. (1). The optimum solution would then be the one that maximizes L or minimizes Q. The minimum Q value is denoted Q_0 corresponding to f_0. Maintaining the non-negativity constraints, the members of acceptable solution set, D, must be such that

$$\ln[L(f)] \geq \ln[L(f_0)] - const \cdot K \quad (6)$$

or alternatively,

$$Q(f) \leq Q_0 + K , \quad (7)$$

where K is a confidence parameter with a typical value of 3. In this way, the set of acceptable solutions of the original equation that fit sufficiently well are determined. In estimating the effects of exposure to this airborne activity, it may be of interest to estimate a function of the distribution. The dose to cells in the bronchial epithelium could be calculated by

$$g[f(d_p)] = \int_a^b G(d_p) f(d_p) \, dd_p , \quad (8)$$

where $G(d_p)$ is the dose per unit airborne alpha activity in the size range d_p to $d_p + dd_p$ [10]. To examine the original distribution, the cumulative sums are estimated as represented by the following sequence of functionals:

$$F(d) = \sum_{j=1}^{J} \begin{array}{l} \Delta_j = 1 \text{ if } d_p \leq d \\ \Delta_j = 0 \text{ if } d_p > d , \end{array} \quad (9)$$

where the $F(d)$ is the cumulative size distribution for the aerosol. The EVE(P) approach estimates such functionals by determining their confidence intervals.

2.3 Activity-Weighted Size Distributions

Activity-weighted size distribution have been measured in a number of normally occupied houses [11–13] using an automated, semi-continuous graded screen array (ASC-GSA) described by Ramamurthi [14] and Ramamurthi and Hopke [15]. The ASC-GSA measurement system is a diffusion battery that uses a combination of six sampler-detector units operated in parallel. Each sampler-detector unit couples wire screen penetration, filter collection, and activity detection with a solid state detector in a way as to minimize depositional losses. The system samples air simultaneously in all of the units, with a flow of about 15 lpm through the sampler slit between the detector and filter holder section in each unit. The sampled air is drawn through a filter. Complete details of the sampler are provided by Ramamurthi and Hopke [15].

Computer control of sampling, counting, and analysis permits automated, semi-continuous operation of the system with sampling every 1.5 h to 3 h. The activities of each radon progeny are estimated from alpha spectra collected during two counting intervals: the first one during sampling and the second 20 min after end of sampling. The observed concentrations of ^{218}Po, ^{214}Pb, and ^{214}Bi are used to reconstruct the corresponding activity-weighted size distributions using the Expectation-Maximization algorithms [4] in six inferred size intervals in geometric progression within the 0.5 nm–500 nm size range. In addition to the individual size distribution for each decay product, the total airborne activity concentration can be characterized by the Potential Alpha Energy Concentration (PAEC). The *PAEC* can be calculated from the individual progeny concentrations by

$$PAEC(mJm)^{-3} = 5.79 \times 10^{-7} \cdot c_1 + 2.86 \times 10^{-6} \cdot c_2 + 2.10 \times 10^{-6} \cdot c_3 , \quad (10)$$

where c_1, c_2, and c_3 are the activity concentrations of the three radon decay products in Bq m^{-3}.

2.4 Results

Measurements have been made in a number of houses in Northeastern North America. To illustrate the use of the EVE(P) algorithm for deconvoluting the activity size distributions, samples taken in houses in Arnprior, Ontario and Parishville, NY will be presented. In each home, radon and the size

distributions of each of the three decay products and the PAEC were determined at 2 h intervals. The details of the experiments in Arnprior are given by Hopke et al. [11]. In this home, radon concentrations were relatively low (< 100 Bq m^{-3}) and generally in the range of 25 Bq m^{-3} to 45 Bq m^{-3}. The cumulative probability distribution for PAEC is shown in Fig. 1. The outer boundary lines are the EVE(P) results for the 95% and 99% confidence intervals. The solid central line is the EM deconvolution result. Although the specific solution obtained by the EM algorithm should fall within the EVE bounds, it may lie anywhere within the feasible region. The confidence band will not necessarily be symmetrically distributed about the specific solution obtained by any particular algorithm.

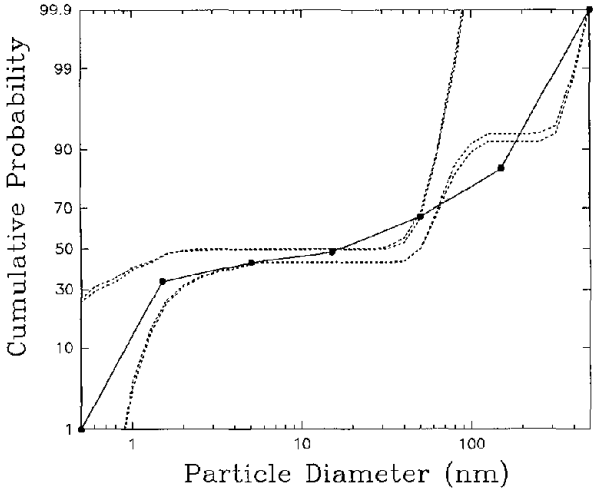


Fig. 1. Cumulative distribution for PAEC for a sample taken in an occupied home in Arnprior, Ontario.

Another analysis was performed on samples from a home in Parishville, NY with much higher radon concentrations (500 Bq m^{-3} to 600 Bq m^{-3}) and thus, the bounds on the feasible region might be smaller [16]. The comparison of the EM size distribution with the EVE(P) distribution for PAEC is shown in Fig. 2. The EM-derived distribution does not appear to fully fit within the EVE(P) bounds. The question is then whether the current EVE(P) approach is the best description of the bounds on the feasible region.

Consideration of extreme value statistics could lead to the following suggestion: it might be possible to define some statistical properties for the extreme members of the set of acceptable solutions, even when there exists no general information about the probability distribution of the solution. Such properties might help in better defining the limits of the set of acceptable solutions. This could help in reducing the confidence intervals of the EVE deconvolution technique without sacrificing the reliability of estimation.

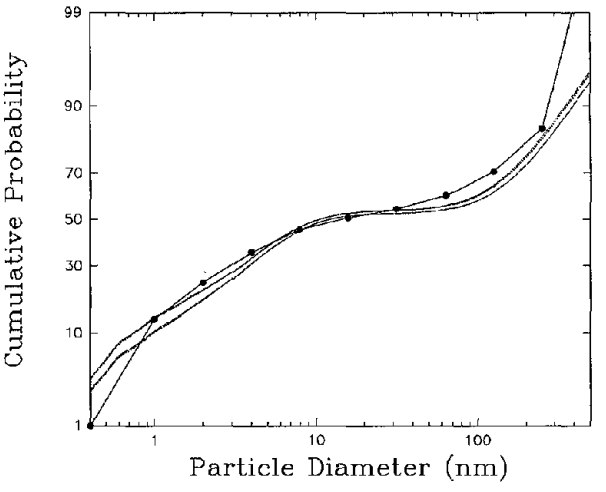


Fig. 2. Cumulative distribution for PAEC for a sample taken in an occupied home in Parishville, NY.

3. Other Applications

3.1 Ambient Air Quality Standard for PM_{10}

In 1987, the U.S. Environmental Protection Agency promulgated a new National Ambient Air Quality Standard (NAAQS) for airborne particulate matter [17] which defined a size-selected portion of the ambient aerosol, particulate matter less than 10 μm or PM_{10}, as important for protection human health and a new way of the determining when the standard had been violated. It is the form of the 24 h standard that involves extreme values. The standard requires that samples taken over 24 h intervals not show more than 1 "expected exceedance" of 150 μg m^{-3} per year averaged over a 3 year period. Particle samples are not usually taken daily because of the manpower requirements needed to manually weigh unexposed filters, change them in the field, and weigh the exposed filters again. A minimum sampling regime would collect samples every 6th day. Thus, over a year approximately 61 samples might be collected. It is assumed that these samples are IID and thus, the number of "expected exceedances" can be estimated as

$$EE_i = OE_i \cdot \frac{n_i}{m_i}, \qquad (11)$$

where for a given year i, EE_i is the number of estimated exceedances, OE_i is the number of observed exceedances, m_i is the number of samples taken, and n_i is the number of days in the year. Thus, if 61 samples are taken in a 365 day year, then 1 observed exceedance becomes 6 expected exceedances. If this observed exceedance is the only one that occurs during a 3 year interval, then the 6 expected exceedances are divided by 3 years to yield an average number of expected exceedances of 2 which is greater than 1 and hence the area is in non-attainment of the standard. In other words, the average number of expected exceedances in any 3 year period is given by

$$E = \frac{1}{3} \cdot \sum_{i=1}^{3} EE_i \,. \tag{12}$$

Davidson and Hopke [18] examined some of the problems that arise as a result of the application of such a standard given incomplete sampling. To illustrate the difficulties, the upper tail of the distribution of airborne mass concentrations will be represented by the following exponential distribution:

$$P(c \le L) = 1 - \frac{1}{365} = 1 - \exp(-y + 2.0) \tag{13}$$

or

$$P(c > L) = \exp\left(2.0 - 7.90\,\frac{c}{L}\right), \tag{14}$$

where c is the mass concentration of airborne particulate matter and L is the maximum concentration allowable under the standard. The probability of an average number of exceedances being greater than 1 will be examined by examining $P(E > 1.05)$.

$$\begin{aligned} P(E \ge 1.05) &= P(\Sigma\, EE_i/3 \ge 1.05) \\ &= P(\Sigma\, EE_i \ge 3.15) \\ &= 1 - P(\Sigma\, EE_i < 3.15) \\ &= 1 - P\,(\Sigma\, OE_i < 3.15 \cdot n/m) \end{aligned} \tag{15}$$

Thus, the probability of nonattainment classification is dependent on the number of measurements per year.

$$\begin{aligned} P(E \ge 1.05) &= 1 - P(\Sigma\, OE_i = 0) \;\; m \le \frac{n}{3.15} \\ &= 1 - P(\Sigma\, OE_i \le 1) \;\; \frac{n}{3.15} < m \le \frac{2n}{3.15} \\ &= 1 - P(\Sigma\, OE_i \le 2) \;\; \frac{2n}{3.15} < m \le \frac{3n}{3.15} \\ &= 1 - P(\Sigma\, OE_i \le 3) \;\; \frac{3n}{3.15} < m \le n \end{aligned} \tag{16}$$

The probabilities of observing 0 to 3 exceedances in any 1 year given the chosen sampling frequency can be estimated using the exponential distribution given in Eq. (17).

$$\begin{aligned} P(\Sigma\, OE_i = 0) &= P_0^3 \\ P(\Sigma\, OE_i = 1) &= 3P_0^2P_0 \\ P(\Sigma\, OE_i = 2) &= 3P_0^2P_2 + 3P_1^2P_0 \\ P(\Sigma\, OE = 3) &= P_1^3 + 3P_0^2P_3 + 6P_0P_1P_2 \end{aligned} \tag{16}$$

A plot of the probability of declaring an area in nonattainment as a function of the number of samples taken per year is shown in Fig. 3. For $c < 1.0L$, classification as nonattainment is a Type I error. For $c > 1.0L$, probability of proper classification represents the power of the approach. The discontinuities occur because of the change in the integer values of the number of expected exceedances that occur at different n/m values. It can be seen that for an area that is exactly in attainment ($c = 1.0L$), there is a probability of up to 60% that it will be misclassified as nonattainment depending on the number of samples taken per year. This form of the standard, therefore, has a high probability of a type I error in order to attain a reasonable power to identify real nonattainment areas.

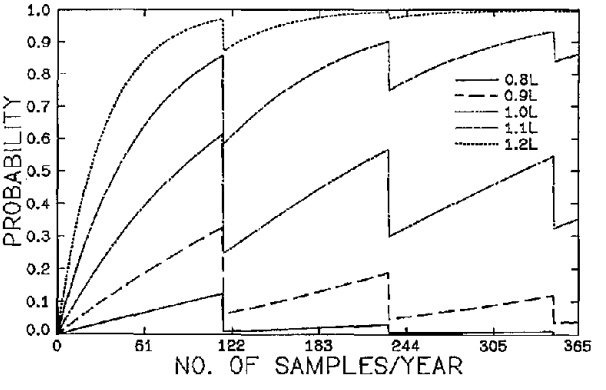


Fig. 3. Probability of classifying an area as being in nonattainment of the 24 h NAAQS for PM_{10} based on an exponential distribution model of the tail.

The goal of this standard is to have the second highest actual value whether measured or not, be at or below the prescribed concentration. Thus, alternative approaches that can more accurately estimate the second highest value in the tail of an extreme valued distribution would potentially provide equal or greater power while lowering the probability of making a misclassification error. Such an estimation process would make the standard more efficient while maintaining or possibly improving its effectiveness.

3.2 Most Exposed Individual

Under the Clean Air Act Amendment of 1990, the Congress has mandated that major emission sources of hazardous air pollutants, defined as materials on a list of 189 substances given in the Act, must install emission control systems. After these systems are in place, the residual risk to the *most exposed individual* must be assessed. If the risk is found to be $>10^{-4}$, the EPA Administrator must decide what additional steps, if any, are to be taken to reduce this risk. Previously the most exposed individual (MEI) has been defined as a person living continuously at the fence line of the facility 200 m from the emission source for 70 years. The idea of a 24 h per day, 70 year lifetime exposure for this individual is obviously an overestimate of the real maximally exposed individual. Recently EPA has revised its guidelines for exposure assessment to support the development of a distribution of exposures that an individual might encounter. However, extreme value statistics is never mentioned in any of the discussions of the use of the upper tail of the distribution to examine exposure and thus risk to the most exposed individual. Since the inaccurate estimation of the residual risk could result in substantial costs for no health benefit if the maximum exposure is overestimated or result in death or adverse health effects if underestimated, the best statistical methodologies should be applied to this important estimation problem. This situation appears ideally suited for extreme value statistics and thus should simultaneously provide interesting statistical problems to solve and value to the society by solving them properly.

4. Conclusions

There appear to be a number of areas in the air pollution field in which rigorous application of extreme value methods could provide useful contributions to solving important environmental problems. The better estimation of the bounds for aerosol size distributions, the determination of attainment or nonattainment of the NAAQS for PM_{10}, and exposure and risk assessments at the high end of the range of possible exposures all could benefit from substantial involvement of extreme value statistical expertise. It is hoped that this report will spark interest in one or more of these problem areas.

Acknowledgments

This work is supported by the U.S. Department of Energy under grant DE FG02 90ER61029 and the National Science Foundation under Grant ATM 9114750.

5. References

[1] D. W. Cooper and L. A. Spielman, Data inversion using non-linear programming with physical constraints: Aerosol size distribution measurements by impactors, Atmospheric Environ. **10**, 723–729 (1976).

[2] D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. ACM **9**, 84–97 (1962).

[3] S. Twomey, Comparison of constrained linear inversion with an iterative nonlinear algorithm applied to the indirect estimation of the particle size distribution, J. Comput. Phys. **18**, 188–200 (1975).

[4] E. F. Maher and N. M. Laird, EM algorithm reconstruction of particle size distribution from diffusion battery data, J. Aerosol Sci. **16**, 557–570 (1985).

[5] E. O. Knutson, Personal Computer Programs for Use in Radon/Thoron Progeny Measurements, U.S. Department of Energy Report EML-517, Environmental Measurements Laboratory, New York (1989) 86 pp.

[6] J. K. Wolfenbarger and J. H. Seinfeld, Inversion of aerosol size distribution data, J. Aerosol Sci. **21**, 227–247 (1990).

[7] J. Replogle, B. D. Holcolmb, and W. R. Burrus, The use of mathematical programming for solving singular and poorly conditioned systems of Equations, J. Math. Anal. Appl. **20**, 310 324 (1967).

[8] P. Paatero, The extreme value estimation deconvolution method with applications in aerosol research, Report Series in Physics No. HU-P-250, University of Helsinki (1990) 42 pp.

[9] P. Paatero, Extreme value estimation, a method for regularizing ill-posed inversion problems, in Ill-Posed Problems in Natural Sciences, A. N. Tikhonov, ed., VSP, Utrecht, the Netherlands (1993) pp. 118–133.

[10] National Research Council (NRC) Dosimetric Extrapolation of BEIR IV Risks to the General Public, National Academy Press, Washington, DC (1991).

[11] P. K. Hopke, N. Montassier, and P. Wasiolek, Evaluation of the effectiveness of several air cleaners for reducing the hazard from indoor radon progeny, Aerosol Sci. Technol. **19**, 268–278 (1993)

[12] C. S. Li and P. K. Hopke, Characterization of radon decay-products in a domestic environment, Indoor Air **4**, 539–561 (1991).

[13] P. T. Wasiolek, P. K. Hopke and A. C. James, Assessment of exposure to radon decay products in realistic living conditions, J. Exposure Analysis Environ. Epidemiology **2**, 309–322 (1992).
[14] M. Ramamurthi, The Detection and Measurement of the Activity Size Distributions ($d_p > 0.5$ nm) Associated with Radon Decay Products in Indoor Air. Ph.D. Thesis, Department of Civil Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 1989.
[15] M. Ramamurthi and P. K. Hopke, An automated, semi-continuous system for measuring indoor radon progeny activity-weighted size distributions, dp: 0.5–500 nm., Aerosol Sci. Technol. **14**, 82–92 (1991).
[16] P. K. Hopke, B. Jensen, and N. Montassier, Evaluation of several air cleaners for reducing indoor radon progeny, J. Aerosol Sci. **25** (2), 395–405 (1994).
[17] Environmental Protection Agency (EPA) Regulations for implementing revised Particulate matter standards, Federal Register **52**, 24, 634 (1987) .
[18] J. Davidson and P. K. Hopke, Implications of incomplete sampling on a statistical form of the ambient air quality standard for particulate matter, Environ. Sci. Technol. **18**, 571–580 (1984).

About the authors: *Philip K. Hopke is the RA Plane Professor of Chemistry at Clarkson University. Pentti Paatero is an Associate Professor of Physics at Helsinki University. Both have interests in the application of statistical methods to the interpretation of air quality data.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 369 (1994)]

A Trivariate Extreme Value Distribution Applied to Flood Frequency Analysis

Volume 99 Number 4 July–August 1994

Carlos A. Escalante-Sandoval

Engineering Graduate Studies Division, Universidad Nacional Autonoma de Mexico, 04510 Mexico, DF, Mexico

and

Jose A. Raynal-Villasenor
Institute of Sciences, Universidad Autonoma de Chihuahua, 31800 Chihuahua, Chih., Mexico

A trivariate extreme value distribution has been derived from the logistic model for the multivariate extreme value distribution. The construction of its corresponding probability distribution and density function is described. In order to obtain the parameters of such a trivariate distribution, a generalized maximum likelihood estimation procedure is described to allow for the cases of samples with different record lengths. Furthermore the reliability of the estimated parameters of the trivariate extreme value distribution is measured through the use of relative information ratios. A region in Northern Mexico with six gauging stations has been selected to apply the trivariate model. Results produced by the proposed model have been compared with those obtained by general extreme value (GEV) distribution functions.

Key words: distribution functions; distribution models; flood frequency analysis; method of maximum likelihood; trivariate extreme value distributions.

Accepted: March 22, 1994

1. Introduction

Flood frequency analysis has been carried out by using univariate distribution functions, the extreme value distributions being an important set of distributions used in this field of study. Generally, parameters of such distributions are estimated from a short record of flows. The variability of these estimates has prompted exploration of joint estimation models which use information from streamflow records of neighboring gauging stations.

In pioneering papers Finkelstein [1], Tiago de Oliveira [2], and Gumbel [3] gave the foundations for the multivariate approach to extreme value distributions. Following this work, several bivariate extreme value models began to appear in the literature. Rueda [4] explored the logistic and mixed models for bivariate extreme value distributions when both marginals are extreme value type I (EVI) distributions. He reported improvements in the estimation of parameters when the bivariate approach is used. Raynal [5] developed and applied three bivariate options from the logistic model of bivariate extreme value distribution for flood frequency analysis. He found that there exists an improvement in the parameter estimation phase, even in the case when both samples have the same record lengths.

Herein, the trivariate approach of multivariate extreme value distribution is presented with a view to its application to flood frequency analysis.

General characteristics, the procedure for estimation, and reliability of parameters of the trivariate extreme value distributions will be described in the following sections. An actual application of the proposed model to six gauging stations in Northern Mexico is presented in the paper.

2. Characteristics of the Trivariate Logistic Model

From the multivariate extension of the logistic model for bivariate extreme value distribution [3], the trivariate approach is:

$$F(x, y, z, \underline{\theta}) = \exp\{-[(\ln F(x))^m + (-\ln F(y))^m + (-\ln F(z))^m]^{1/m}, \quad (1)$$

where m is the association parameter ($m \geq 1$) and $F(s) = F(s, \underline{\theta})$ is the marginal distribution function of s. Equation (1) must satisfy the following inequalities (Tiago de Oliveira [6, 7]):

$$F(x)F(y)F(z) \leq F(x, y, z) \leq \min[F(x), F(y), F(z)] \quad (2)$$

$$[F(x, y,)F(x, z)F(y, z)]^{\frac{1}{4}} \leq F(x, y, z) \leq \frac{[F(x, y)F(x, z)F(y, z)]^{1/2}}{[F(x)F(y)F(z)]^1}. \quad (3)$$

Marginals in Eq. (1) can be either EVI distributions:

$$F(s) = \exp\left(-\exp-\left(\frac{s-u}{\alpha}\right)\right) \quad (4)$$

or GEV distributions:

$$F(s) = \exp-\left(1-\left(\frac{s-u}{\alpha}\right)\beta\right)^{1/\beta}. \quad (5)$$

The combinations have been named (Escalante [8]):

a) Trivariate extreme value distribution type 111 (TEV111) or TriGumbel distribution. All marginals are EVI distributions.
b) Trivariate extreme value distribution type 112 (TEV112) or BiGumbel-GEV distribution.
c) Trivariate extreme value distribution type 122 (TEV122) or BiGEV-EVI distribution.
d) Trivariate extreme value distribution type 222 (TEV222) or TriGumbel distribution. All marginals are GEV distributions.

The particular form of Eq. (1), when the marginals are GEV distributions for the maxima, is (Escalante [8]):

$$F(x, y, z, u_1, \alpha_1, \beta_1, u_2, \alpha_2, \beta_2, u_3, \alpha_3, \beta_3, m_t) = \exp\left[-\left(\left(1-\left(\frac{x-u_1}{\alpha_1}\right)\beta_1\right)^{m_t/\beta_1} + \left(1-\left(\frac{y-u_2}{\alpha_2}\right)\beta_2\right)^{m_t/\beta_2} + \left(1-\left(\frac{z-u_3}{\alpha_3}\right)\beta_3\right)^{m_t/\beta_3}\right)^{1/m_t}\right], \quad (6)$$

where u_i, α_i and β_i, $i = 1, 2, 3$, are the location, scale and shape parameters of the marginal GEV distributions for the maxima. The corresponding probability density function is (Escalante [8]):

$$f(x, y, z, u_1, \alpha_1, \beta_1, u_2, \alpha_2, \beta_2, u_3, \alpha_3, \beta_3, m_t) = \frac{1}{\alpha_1\alpha_2\alpha_3}\left(1-\left(\frac{x-u_1}{\alpha_1}\right)\beta_1\right)^{m_t/\beta_1-1}\left(1-\left(\frac{y-u_2}{\alpha_2}\right)\beta_2\right)^{m_t/\beta_2-1}$$
$$\left(1-\left(\frac{z-u_3}{\alpha_3}\right)\beta_3\right)^{m_t/\beta_3-1}\exp\left[-\left(\left(1-\left(\frac{x-u_1}{\alpha_1}\right)\beta_1\right)^{m_t/\beta_1}+\left(1-\left(\frac{y-u_2}{\alpha_2}\right)\beta_2\right)^{m_t/\beta_2}+\left(1-\left(\frac{z-u_3}{\alpha_3}\right)\beta_3\right)^{m_t/\beta_3}\right)^{1/m_t}\right]$$
$$\left[\left(1-\left(\frac{x-u_1}{\alpha_1}\right)\beta_1\right)^{m_t/\beta_1}+\left(1-\left(\frac{y-u_2}{\alpha_2}\right)\beta_2\right)^{m_t/\beta_2}+\left(1-\left(\frac{z-u_3}{\alpha_3}\right)\beta_3\right)^{m_t/\beta_3}\right]^{1/m_t-3}\left[\left(1-m_t\right)\left(1-2_{m_t}\right)\right.$$
$$+\left(\left(1-\left(\frac{x-u_1}{\alpha_1}\right)\beta_1\right)^{m_t/\beta_1}+\left(1-\left(\frac{y-u_2}{\alpha_2}\right)\beta_2\right)^{m_t/\beta_2}+\left(1-\left(\frac{z-u_3}{\alpha_3}\right)\beta_3\right)^{m_t/\beta_3}\right)^{2/m_t}+\left(3m_t-3\right)$$
$$\left.\left(\left(1-\left(\frac{x-u_1}{\alpha_1}\right)\beta_1\right)^{m_t/\beta_1}+\left(1-\left(\frac{y-u_2}{\alpha_2}\right)\beta_2\right)^{m_t/\beta_2}+\left(1-\left(\frac{z-u_3}{\alpha_3}\right)\beta_3\right)^{m_t/\beta_3}\right)^{1/m_t}\right]. \quad (7)$$

3. Estimation of Parameters

The method of maximum likelihood for estimating the parameters of trivariate extreme value distributions has been chosen due to its characteristics for consistency in large sample estimation and applicability in estimating the parameters of cumbersome density functions.

For the case of trivariate distribution functions, the sample arrangements could allow having either an equal or different record length in any of the samples to be analysed.

In order to consider all possible combinations of data, it is required to have a sufficiently flexible formulation, therefore the following general form of the likelihood function will be used based on the generalization obtained by Anderson [9]:

$$L(x,y,z,\underline{\theta})=\left[\prod_{i=1}^{n_1} f(p_i,\underline{\theta}_1)\right]^{I_1}\left[\prod_{i=1}^{n_2} f(p_i,q_i,\underline{\theta}_2)\right]^{I_2}$$

$$\left[\prod_{i=1}^{n_3} f(x,y,z,\underline{\theta}_3)\right]^{I_3}\left[\prod_{i=1}^{n_4} f(r_i,s_i,\underline{\theta}_4)\right]^{I_4}\left[\prod_{i=1}^{n_5} f(r_i,\underline{\theta}_5)\right]^{I_5}, \tag{8}$$

where:

n_1, n_2 = are respectively the univariate and bivariate record lengths before the common period n_3,

n_4, n_5 = are respectively the bivariate and univariate record lengths after the common period n_3,

p = is the variable with univariate record before the common period,

(p, q) = are the variables with bivariate record before the common period,

(x, y, z) = are the variables with trivariate record during the common period,

(r, s) = are the variables with bivariate record after the common period,

r = is the variable with univariate record after the common period,

I_i = are indicator numbers such that: $I_i=1$ if $n_i>0$ and $I_i=0$ if $n_i=0$.

The logarithmic function will be used instead of the likelihood function. So, Eq. (8) is transformed into:

$$LL(x,y,z,\underline{\theta})=I_1\left[\sum_{i=1}^{n_1}\ln f(p_i,\underline{\theta}_1)\right]$$

$$+I_2\left[\sum_{i=1}^{n_2}\ln f(p_i,q_i,\underline{\theta}_2)\right]+I_3\left[\sum_{i=1}^{n_3}\ln f(x,y,z,\underline{\theta}_3)\right]$$

$$+I_4\left[\sum_{i=1}^{n_4}\ln f(r_i,s_i,\underline{\theta}_4)\right]+I_5\left[\sum_{i=1}^{n_5}\ln f(r_i,\underline{\theta}_5)\right]. \tag{9}$$

The maximum likelihood estimators of parameters for the trivariate extreme value distributions are those values for which Eq. (9) is maximized.

The corresponding logarithmic likelihood function for the trigeneral extreme value (TEV222) distribution function, based on Eq. (9) from [8] is shown in Eq. (10):

$$LL(x,y,z,u_1,\alpha_1,\beta_1,u_2,\alpha_2,\beta_2,u_3,\alpha_3,\beta_3,m_t,m_{b1},m_{b2})=I_1\left\{-n_1\ln\alpha_p\right.$$

$$\left.+\sum_{i=1}^{n_1}\left[-\left(1-\left(\frac{p_i-u_p}{\alpha_p}\right)\beta_p\right)^{1/\beta_p}+\ln\left(1-\left(\frac{p_i-u_p}{\alpha_p}\right)\beta_p\right)^{1/\beta_p-1}\right]\right\}$$

$$+I_2\left\{\sum_{i=1}^{n_2}\left[-\left(\ln\alpha_p+\ln\alpha_q\right)+\ln\left(1-\left(\frac{p_i-u_p}{\alpha_p}\right)\beta_p\right)^{m_{b1}/\beta_p-1}+\ln\left(1-\left(\frac{q_i-u_q}{\alpha_q}\right)\beta_q\right)^{m_{b1}/\beta_q-1}\right.\right.$$

$$+\ln\left(\left(1-\left(\frac{p_i-u_p}{\alpha_p}\right)\beta_p\right)^{m_{b1}/\beta_p}+\left(1-\left(\frac{q_i-u_q}{\alpha_q}\right)\beta_q\right)^{m_{b1}/\beta_q}\right)^{1/m_{b1}-2}$$

$$+\ln\left(\left(m_{b1}-1\right)+\left(\left(1-\left(\frac{p_i-u_p}{\alpha_p}\right)\beta_p\right)^{m_{b1}/\beta_p}+\left(1-\left(\frac{q_i-u_q}{\alpha_q}\right)\beta_q\right)^{m_{b1}/\beta_q}\right)^{1/m_{b1}}\right)$$

$$\left.\left.-\left(\left(1-\left(\frac{p_i-u_p}{\alpha_p}\right)\beta_p\right)^{m_{b1}/\beta_p}+\left(1-\left(\frac{q_i-u_q}{\alpha_q}\right)\beta_q\right)^{m_{b1}/\beta_q}\right)^{1/m_{b1}}\right]\right\}$$

$$+I_3\Bigg\{-n_3\Big(\ln\alpha_1+\ln\alpha_2+\ln\alpha_3\Big)+\sum_{i=1}^{n_3}\Bigg[\ln\Big(1-\Big(\frac{x_i-u_1}{\alpha_1}\Big)\beta_1\Big)^{m_t/\beta_1-1}$$

$$+\ln\Big(1-\Big(\frac{y_i-u_2}{\alpha_2}\Big)\beta_2\Big)^{m_t/\beta_2-1}+\ln\Big(1-\Big(\frac{z_i-u_3}{\alpha_3}\Big)\beta_3\Big)^{m_t/\beta_3-1}$$

$$+\ln\Bigg(\Big(1-\Big(\frac{x_i-u_1}{\alpha_1}\Big)\beta_1\Big)^{m_t/\beta_1}+\Big(1-\Big(\frac{y_i-u_2}{\alpha_2}\Big)\beta_2\Big)^{m_t/\beta_2}+\Big(1-\Big(\frac{z_i-u_3}{\alpha_3}\Big)\beta_3\Big)^{m_t/\beta_3}\Bigg)^{1/m_t-3}$$

$$+\ln\Bigg[\Big(1-m_t\Big)\Big(1-2m_t\Big)+\Bigg(\Big(1-\Big(\frac{x_i-u_1}{\alpha_1}\Big)\beta_1\Big)^{m_t/\beta_1}+\Big(1-\Big(\frac{y_i-u_2}{\alpha_2}\Big)\beta_2\Big)^{m_t/\beta_2}$$

$$+\Big(1-\Big(\frac{z_i-u_3}{\alpha_3}\Big)\beta_3\Big)^{m_t/\beta_3}\Bigg)^{2/m_t}+\Big(3m_t-3\Big)\Bigg(\Big(1-\Big(\frac{x_i-u_1}{\alpha_1}\Big)\beta_1\Big)^{m_t/\beta_1}$$

$$+\Big(1-\Big(\frac{y_i-u_2}{\alpha_2}\Big)\beta_2\Big)^{m_t/\beta_2}+\Big(1-\Big(\frac{z_i-u_3}{\alpha_3}\Big)\beta_3\Big)^{m_t/\beta_3}\Bigg)^{1/m_t}\Bigg]-\Bigg(\Big(1-\Big(\frac{x_i-u_1}{\alpha_1}\Big)\beta_1\Big)^{m_t/\beta_1}$$

$$+\Big(1-\Big(\frac{y_i-u_2}{\alpha_2}\Big)\beta_2\Big)^{m_t/\beta_2}+\Big(1-\Big(\frac{z_i-u_3}{\alpha_3}\Big)\beta_3\Big)^{m_t/\beta_3}\Bigg)^{1/m_t}\Bigg]\Bigg\}$$

$$+I_4\Bigg\{\sum_{i=1}^{n_4}\Bigg[-\Big(\ln\alpha_r+\ln\alpha_s\Big)+\ln\Big(1-\Big(\frac{r_i-u_r}{\alpha_r}\Big)\beta_r\Big)^{m_{b2}/\beta_r-1}+\ln\Big(1-\Big(\frac{s_i-u_s}{\alpha_s}\Big)\beta_s\Big)^{m_{b2}/\beta_s-1}$$

$$+\ln\Bigg(\Big(1-\Big(\frac{r_i-u_r}{\alpha_r}\Big)\beta_r\Big)^{m_{b2}/\beta_r}+\Big(1-\Big(\frac{s_i-u_s}{\alpha_s}\Big)\beta_s\Big)^{m_{b2}/\beta_s}\Bigg)^{1/m_{b2}-2}+\ln\Big(\Big(m_{b2}-1\Big)$$

$$\Bigg(\Big(1-\Big(\frac{r_i-u_r}{\alpha_r}\Big)\beta_r\Big)^{m_{b2}/\beta_r}+\Big(1-\Big(\frac{s_i-u_s}{\alpha_s}\Big)\beta_s\Big)^{m_{b2}/\beta_s}\Bigg)^{1/m_{b2}}\Bigg)-\Bigg(\Big(1-\Big(\frac{r_i-u_r}{\alpha_r}\Big)\beta_r\Big)^{m_{b2}/\beta_r}$$

$$+\Big(1-\Big(\frac{s_i-u_s}{\alpha_s}\Big)\beta_s\Big)^{m_{b2}/\beta_s}\Bigg)^{1/m_{b2}}\Bigg]\Bigg\}+I_5\Bigg\{-n_5\ln\alpha_r$$

$$+\sum_{i=1}^{n_5}\Bigg[-\Big(1-\Big(\frac{r_i-u_r}{\alpha_r}\Big)\beta_r\Big)^{1/\beta_r}+\ln\Big(1-\Big(\frac{r_i-u_r}{\alpha_r}\Big)\beta_r\Big)^{1/\beta_r-1}\Bigg]\Bigg\} \tag{10}$$

where:

m_t trivariate association parameter
m_{b1}, m_{b2} bivariate association parameter before and after the common period, respectively.

$\sum_{i=1}^{n_1}\ln f(p_i,\ \underline{\theta}_1)$ and $\sum_{i=1}^{n_5}\ln f(r_i,\ \underline{\theta}_5)$ take the form:

$$\sum_{i=1}^{n_j}\Bigg[-\ln\ \alpha_t-(1-\Big(\frac{t_i-u_t}{\alpha_t}\Big)\ \beta_t)^{1/\beta_t}+\ln(1-\Big(\frac{t_i-u_t}{\alpha_t}\Big)\ \beta_t)^{1/\beta_t-1}\Bigg]. \tag{11}$$

Similarly, $\sum_{i=1}^{n_2} \ln f(p_i, q_i, \underline{\theta}_2)$ and $\sum_{i=1}^{n_4} \ln f(r_i, s_i, \underline{\theta}_4)$ take the following form (bivariate relationship with both GEV marginals):

$$\sum_{i=1}^{n_1}\Bigg[-\left(\ln\alpha_t+\ln\alpha_w\right)+\ln\left(1-\left(\frac{t_i-u_t}{\alpha_t}\right)\beta_t\right)^{m_b/\beta_t-1}$$

$$+\ln\left(1-\left(\frac{w_i-u_w}{\alpha_w}\right)\beta_w\right)^{m_b/\beta_w-1}$$

$$+\ln\left(\left(1-\left(\frac{t_i-u_t}{\alpha_t}\right)\beta_t\right)^{m_b/\beta_t}\right.$$

$$\left.+\left(1-\left(\frac{w_i-u_w}{\alpha_w}\right)\beta_w\right)^{m_b/\beta_w}\right)^{1/m_b-2}$$

$$+\ln\left(\left(m_b-1\right)+\left(\left(1-\left(\frac{t_i-u_t}{\alpha_t}\right)\beta_t\right)^{m_b/\beta_t}\right.\right.$$

$$\left.\left.+\left(1-\left(\frac{w_i-u_w}{\alpha_w}\right)\beta_w\right)^{m_b/\beta_w}\right)^{1/m_b}\right)$$

$$-\left(\left(1-\left(\frac{t_i-u_t}{\alpha_t}\right)\beta_t\right)^{m_b/\beta_t}\right.$$

$$\left.+\left(1-\left(\frac{w_i-u_w}{\alpha_w}\right)\beta_w\right)^{m_b/\beta_w}\right)^{1/m_b}\Bigg]. \qquad (12)$$

Given the complexity of the mathematical expressions in Eq. (10) and their partial derivatives with respect to the parameters, the constrained multivariable Rosenbrock method, Kuester and Mize [10], was applied to obtain the maximum likelihood estimators for the parameters by the direct maximization of Eq. (10). The required initial values of the parameters to start the optimization of Eq. (10) were provided by the univariate maximum likelihood estimators of the parameters for the case of the location, scale, and shape parameters. The initial values of the association parameters, bivariate and trivariate, were set equal to 2, following the procedure developed by Escalante [8].

4. Reliability of Estimated Parameters

The indicator selected to measure the reliability of estimated parameters when using the trivariate distribution as compared with the univariate counterpart was the asymptotic relative information ratio.

Table 1 shows a sample of relative information ratios obtained by using the following set of parameters:

$$u_1=15.0,\quad \alpha_1=2.0,\quad \beta_1=-0.20$$
$$u_2=12.0,\quad \alpha_2=1.2,\quad \beta_2=-0.15$$
$$u_3=10.0,\quad \alpha_3=1.0,\quad \beta_3=-0.10$$

Table 1. Asymptotic relative information ratios of the parameters of the TEV222 distribution for the maxima $n_3=25$; $m_t=2$; $m_{b2}=2$

Parameter	n_4	0	25	50	75
u_1	0	1.0868	1.3695	1.5055	1.5856
	25	1.4460	1.5942	1.6823	
	50	1.6295	1.7201		
	75	1.7408			
α_1	0	1.0141	1.2274	1.3256	1.3821
	25	1.2712	1.3753	1.4356	
	50	1.3941	1.4555		
	75	1.4662			
β_1	0	1.2405	1.3555	1.4041	1.4312
	25	1.3864	1.4382	1.4671	
	50	1.4514	1.4806		
	75	1.4883			
u_2	0	1.0876	1.3722	1.5094	1.5903
	25	1.0599	1.2263	1.3333	
	50	1.0500	1.1683		
	75	1.0450			
α_2	0	1.0135	1.2151	1.3065	1.3587
	25	0.9835	1.1046	1.1788	
	50	0.9734	1.0604		
	75	0.9684			
β_2	0	1.2442	1.3469	1.3901	1.4140
	25	1.1324	1.2069	1.2507	
	50	1.0934	1.1507		
	75	1.0736			
u_3	0	1.0882	1.0462	1.0302	1.0217
	25	1.0604	1.0417	1.0313	
	50	1.0506	1.0390		
	75	1.0736			
α_3	0	1.0126	0.9553	0.9334	0.9218
	25	0.9822	0.9549	0.9395	
	50	0.9790	0.9542		
	75	0.9669			
β_3	0	1.2437	1.0268	0.9535	0.9166
	25	1.1314	1.0246	0.9704	
	50	1.0923	1.0223		
	75	1.0724			

5. Case Study

A region located in Northern Mexico, with a total of six gauging stations, was selected to apply the proposed methodology to the flood frequency analysis. Tables 2–4 show the results of the application of the trivariate extreme value distributions for the maxima to the data recorded in such gauging stations.

In order to compare the goodness of fit between the univariate and trivariate maximum likelihood estimates of the parameters in stations considered in the case study, the standard error of fit, as defined by Kite [11], was obtained and is displayed in Table 5.

Table 2. Correlation coefficients and relative sample sizes for the triplets of stations for the case study

Triplets of stations	Correlation coefficient	Relative sample sizes n_1	n_2	n_3	n_4	n_5
Acatitan-Sta Cruz-Ixpalino	0.926	9	2	26	0	0
Choix-Huites-Sn Francisco	0.969	0	14	18	7	0

Table 3. Univariate maximum likelihood estimates of the parameters of the GEV distributions defined by the data of the gauging stations of the case study

Station	Location	Scale	Shape
Acatitan	576.21	283.80	−0.62
Choix	236.69	130.15	−0.12
Huites	1564.78	978.87	−0.57
Ixpalino	772.57	473.97	−0.38
Sn Francisco	926.53	532.56	−0.65
Sta Cruz	835.74	440.23	−0.40

Table 4. Trivariate maximum likelihood estimates of the parameters of the TEV222 distribution defined by the data of the gauging stations of the case study

Station	Location	Scale	Shape
Acatitan	568.93	269.44	−0.64
Choix	220.85	128.29	−0.39
Huites	1603.30	1038.53	−0.68
Ixpalino	795.03	490.86	−0.46
Sn Francisco	943.69	540.73	−0.67
Sta Cruz	850.97	467.74	−0.52

Table 5. Standard errors of fit for gauging stations of case study

Station	Standard error of fit Univariate (GEV)	Trivariate (TEV222)
Acatitan	244.40[a]	253.90
Choix	87.70	58.80[a]
Huites	1024.00	831.90[a]
Ixpalino	537.90	393.00[a]
Sn Francisco	350.80[a]	401.50
Sta Cruz	497.20	259.60[a]

[a] Minimum standard error of fit.

6. Conclusions

The logistic model for trivariate general extreme value distribution for the maxima has been proposed. Asymptotic and data base results suggest that the proposed model is a suitable option to be considered when performing flood frequency analysis.

Acknowledgments

The authors express their deepest gratitude to Comision Nacional del Agua, Universidad Nacional Autonoma de Mexico and Universidad Autonoma de Chihuahua for the support given for the publication of this paper. Funding for this study was provided by the Comision Nacional del Agua through the Agreement SARH-CNA-UNAM.

7. References

[1] B. V. Finkelstein, Limiting distributions of extreme terms of variational series of a two dimensional random value, Dokl. Akad. Nauk. SSSR 91, 2 (1953).
[2] J. Tiago de Oliveira, Extremal distributions, Faculdade de Ciencias de Lisboa, 2 serie, A. Mat. VII (1958).
[3] E. J. Gumbel, Distributions del valeurs extremes en plusiers dimensions, Publications de L'Institute de Statistique, Vol. 9, 171–173 (1960).
[4] E. Rueda, Transfer of information for flood related variables, M. Sc. thesis, Civil Engineering Department, Colorado State University (1981) p. 112.
[5] J. A. Raynal, Bivariate extreme value distributions applied to flood frequency analysis, Ph. D. dissertation, Civil Engineering Department, CSU (1985) p. 237.
[6] J. Tiago de Oliveira, Bivariate extremes: extensions, Proc. of the 40th session, Bulletin of the International Statistical Institute, Vol. 46, Book 2, 241–251 (1975).
[7] J. Tiago de Oliveira, Bivariate and multivariate extreme distributions, Statistical Distributions in Scientific Work, G. P. Patil et al., editors, D. Reidel Publishing Co., Vol. 1, (1975) pp. 355–361.

[8] S. C. Escalante, Trivariate extreme value distributions and its applications to flood frequency analysis, Ph. D. dissertation, Engineering Graduate Studies Division, Universidad Nacional Autonoma de Mexico (1991) p. 315 (in Spanish).
[9] T. W. Anderson, Maximum likelihood estimates for a multivariate normal distribution when some observations are missing, Journal of the American Statistical Association, Vol. 52, (1957) pp. 200–203.
[10] J. L. Kuester and J. H. Mize, Optimization with Fortran, McGraw Hill Book Co. (1973) pp. 386–398.
[11] G. W. Kite, Frequency and risk analysis in hydrology, Water Resources Publications (1977) p. 257.

About the authors: *Carlos A. Escalante-Sandoval is an engineer at the Universidad Nacional Autonoma de Mexico and Jose A. Raynal-Villasenor is a mathematical statistician at the Universidad Autonoma de Chihuahua.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 377 (1994)]

Fractal Theory and the Estimation of Extreme Floods

Volume 99 | Number 4 | July–August 1994

Donald L. Turcotte

Department of Geological Sciences,
Cornell University,
Ithaca, NY 14853

Floods and draughts constitute extreme values of great consequence to society. A wide variety of statistical techniques have been applied to the evaluation of the flood hazard. A primary difficulty is the relatively short time span over which historical data are available, and quantitative estimates for paleofloods are generally suspect. It was in the context of floods that Hurst introduced the concept of the rescaled range. This was subsequently extended by Mandelbrot and his colleagues to concepts of fractional Gaussian noises and fractional Brownian walks. These studies introduced the controversial possibility that the extremes of floods and droughts could be fractal. An extensive study of flood gauge records at 1200 stations in the United States indicates a good correlation with fractal statistics. It is convenient to introduce the parameter F which is the ratio of the 10 year flood to the 1-year flood; for fractal statistics F is also the ratio of the 100 year flood to the 10 year flood and the ratio of the 1000 year flood to the 100 year flood. It is found that the parameter F has strong regional variations associated with climate. The acceptance of power-law statistics rather than exponentially based statistics would lead to a far more conservative estimate of future flood hazards.

Key words: Brownian walks; floods; fractals; Gaussian noises; time series.

Accepted: March 22, 1994

1. Introduction

The flow in a river can generally be considered a time series. The extreme values in the time series constitute floods. Floods present a severe natural hazard; in order to assess the hazard and to allocate resources for its mitigation it is necessary to make flood-frequency hazard assessments. The integral of the flow in a river is required for the design of reservoirs and to assess available water supplies during periods of drought.

One estimate of the severity of a flood is the peak discharge at a station V. The magnitude of the peak discharge is affected by a variety of circumstances including: (1) The amount of rainfall produced by the storm or storms in question, (2) the upstream drainage area, (3) the saturation of the soil in the drainage area, (4) the topography, soil type, and vegetation in the drainage area, and (5) whether snow melt is involved. In addition dams, stream channelization, and other man-made modifications can affect the severity of floods.

In order to estimate the severity of future floods, historical records are used to provide flood-frequency estimates. Unfortunately, this record generally covers a relatively short time span and no general basis has been accepted for its extrapolation. Quantitative estimates of peak discharges associated with paleofloods are generally not sufficiently accurate to be of much value. A wide variety of geostatistical distributions have been applied to flood-frequency forecasts, often with quite divergent predictions. Examples of distributions used include power law (fractal), log normal,

gamma, Gumbel, log Gumbel, Hazen, and log Pearson. Many discussions of this work appear in the literature [1–7].

An independent approach to reservoir storage was developed by Hurst [8, 9]. Hurst spent his life studying the flow characteristics of the Nile and introduced the rescaled range (R/S) analysis. He found that the variations of the storage (the range) scaled with the time period considered as a power law. Mandelbrot and Wallis [10–13] introduced the concepts of fractional Gaussian noises and fractional Brownian walks and related these to R/S analysis; all are recognized as fractal distributions. They also introduced the Noah and Joseph effects. The Noah affect is the skewness of the distribution of flows in a river and the Joseph effect is the persistence of the flows. Although the concepts introduced by Hurst and Mandelbrot and Wallis have been considered in a wide variety of applications [14], they have not influenced approaches to flood-frequency forecasting. This point will be a central feature of this paper along with a general discussion of the applicability of fractal statistics.

2. Analysis

In most cases the flow in a river is a continuous function of time, thus it is appropriate to treat the flow as a time series. It is straightforward to study the spectral characteristics of the time series by determining the coefficients of a Fourier expansion. For most river flows there will be a strong annual peak associated with seasonal variations in rainfall. However, it is of interest to examine the longer range trends in the data. If the Fourier coefficients have a power-law dependence on frequency over a significant range of frequencies a fractal dependence is obtained (with some constraints on the power).

If $\dot{V}(t)$ is the volumetric flow in a river as a function of time, the condition that the flow is fractal requires that

$$\Pr\left[\frac{\dot{V}(t+T)-\dot{V}(t)}{T^H} \le y\right]=f(y), \tag{1}$$

where $\dot{V}(t+T)-\dot{V}(t)$ is the difference in flow after a time T, H is known as the Hausdorff measure, and $f(y)$ is a normalized cumulative probability distribution function. When $f(y)$ is the error function and $H=1/2$ this relation defines a Brownian walk. If $0<H<1$ and $f(y)$ is the error function, this relation defines fractional Brownian walks. The fractal dimension of a fractional Brownian walk is related to the Hausdorff measure by [15]

$$D=2-H \tag{2}$$

and with $0<H<1$ we have $1<D<2$.

An extension of the self-similar analysis of rivers as a time series is to treat floods as a discrete fractal set. In order to avoid difficulties with annual variability we hypothesize that the peak annual discharge $\dot{V}_m$ in a time interval T is related to the interval by

$$\dot{V}_m(T)=C_1T^H \tag{3}$$

with T an integer number of years. Self-similar river flows imply a power-law scaling of peak annual discharges and recurrence intervals.

This scale invariant distribution can also be expressed in terms of the ratio F of the peak discharge over a 10 year interval to the peak discharge over a 1 year interval. With self-similarity the parameter F is then also the ratio of the 100 year peak discharge to the 10 year peak discharge. In terms of H and D we have

$$F=10^H=10^{2-D}. \tag{4}$$

The parameter F is a measure of the severity of great floods.

An alternative way of writing Eq. (3) is

$$\dot{N}=C_2\dot{V}^{-\alpha}, \tag{5}$$

where $\dot{N}$ is the number of floods per unit time with flows that exceed $\dot{V}$. This relation will be used to analyse actual flood-frequency data. The quantities $\dot{N}$ in Eq. (5) and T in Eq. (3) are related by

$$\dot{N}=\frac{1}{T}, \tag{6}$$

so that we have

$$H=\frac{1}{\alpha} \tag{7}$$

and from Eq. (2) we have

$$D=2-\frac{1}{\alpha}. \tag{8}$$

Data will be used to obtain α; F, H, and D will then be found from Eqs. (4), (7), and (8).

Before considering actual examples we will also introduce rescaled range (R/S) analysis. Hurst [8, 9]

proposed this empirical approach to the statistics of floods and draughts. The method is illustrated in Fig. 1. Consider a reservoir behind a dam that never overflows or empties, the flow into the reservoir is $\dot{V}(t)$ and the flow out of the reservoir is $\bar{\dot{V}}(T)$ defined by

$$\bar{\dot{V}} = \frac{1}{T}\int_0^T \dot{V}(t)\,dt. \tag{9}$$

The volume of water in the reservoir $V(t)$ is given by

$$V(t) = V(0) + \int_0^T \dot{V}(t')\,dt' - t\,\bar{\dot{V}}(T) \tag{10}$$

and the range is defined by

$$R(T) = V_{max} - V_{min}, \tag{11}$$

where V_{max} is the maximum volume and V_{min} the minimum volume stored during the interval T. The rescaled range is defined as R/S where S is the standard deviation of the flow during the period T

$$S(T) = \left[\frac{1}{T}\int_0^T (\dot{V}(t) - \bar{\dot{V}})^2\right]^{1/2} dt. \tag{12}$$

Hurst et al. [16] found that for many time series the rescaled range satisfies the empirical relation

$$\frac{R}{S} = \left(\frac{T}{2}\right)^{H_1} \tag{13}$$

where H_1 is known as the Hurst exponent. Examples included river discharges, rainfall, varves, temperatures, sunspot numbers, and tree rings. In many cases the value of the Hurst exponent is near 0.7.

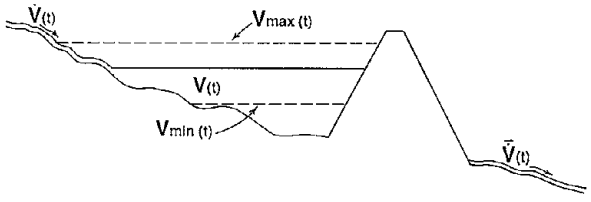


Fig. 1. Illustration of how rescaled range (R/S) analysis is carried out. The flow into a reservoir is $\dot{V}(t)$ and the flow out is $\bar{\dot{V}}(T)$. The maximum volume of water in the reservoir during the period T is $V_{max}(T)$ and the minimum $V_{min}(T)$; the difference is the range $R(T) = V_{max}(T) - V_{min}(T)$.

If a Gaussian white noise sequence of numbers is integrated or summed the result is a Brownian walk. An R/S analysis of the white noise sequence gives a Hurst exponent H_1, thus the Hurst exponent is equal to the Hausdorff measure of the integrated signal, a Brownian walk with $H = 0.5$. Mandelbrot and Wallis [10–13] introduced the concept of fractional Gaussian noises and their integrals, fractional Brownian walks. They showed that the Hurst exponent H_1 of a fractional Gaussian noise is equal to the Hausdorff measure of the corresponding fractional Brownian walk.

If $0.5 < H_1 < 1$ the original time series is said to be persistence; adjacent values are more strongly correlated than if they were random. The higher the value of H_1, the greater the persistence. If $0 < H_1 < 0.5$ the original time series is said to be antipersistent; adjacent values are less correlated than if they were random.

3. Examples

We now turn to the analysis of flood-frequency records. As our first example, the 10 benchmark stations considered by Benson [2] will be studied. Benson [2] applied a variety of geostatistical distributions to the data from these stations, these will be compared with the fractal approach discussed above. The maximum annual floods for two stations are given in Fig. 2. Values for station 1-1805 on the Middle Branch of the Westfield River in Goss Heights, Massachusetts are given in Fig. 2a for the period 1911–1960 [17] and values for station 11-0980 in the Arroyo Seco near Pasadena, California are given in Fig. 2b for the period 1914–1965 [18].

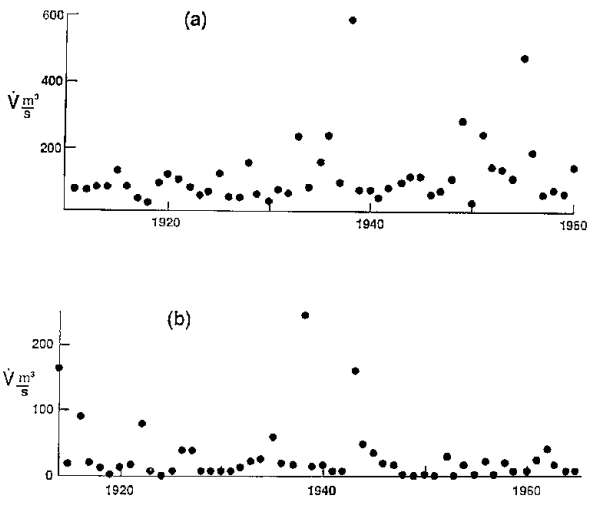


Fig. 2. Maximum annual floods for (a) station 1-1805 on the Middle Branch of the Westfield River, Goss Heights, Massachusetts and (b) station 11-0980 in the Arroyo Seco near Pasadena, California.

In order to assess the applicability of fractal statistics the number of annual floods N with a peak discharge greater than $\dot{V}(m^3/s)$ is divided by the sampling period to give the mean number of floods per year $\dot{N}$ with a peak discharge greater than the specified value. The $\log \dot{N}(\dot{V})$ is then plotted against $\log \dot{V}$. Results for station 1-1805 are given in Fig. 3a, the solid line is the least square fit of Eq. (5) with the data over the range $50 < \dot{V} < 200\ m^3/s$; large floods are omitted from the fit because of their small number. The solid line corresponds to $\alpha = 2.3$; from Eqs. (4), (7), and (8) we have $H = 0.435$, $F = 2.72$, and $D = 1.56$. Results for station 11-0980 are given in Fig. 3b, the solid line is the best fit of Eq. (5) with the data over the range $10 < \dot{V} < 100\ m^3/s$. The solid line corresponds to $\alpha = 1.1$; from Eqs. (4), (7), and (8) we have $H = 0.909$, $F = 8.11$, and $D = 1.09$. In both cases the fit to the scale-invariant (fractal) relation is quite good. The values of H and F in California are considerably larger than in Massachusetts. Large floods are relatively more probable in the arid climate than in the temperate climate.

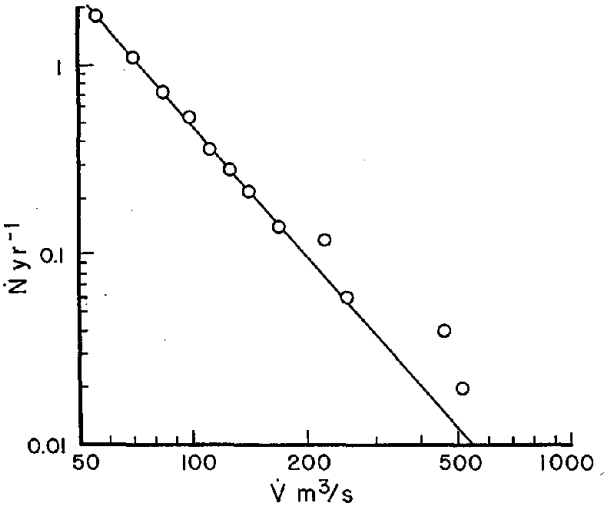


Fig. 3a. Number of floods per year with a peak discharge greater than $\dot{V}$. Station 1-1805 in Goss Heights, Massachusetts during the period 1911-1960.

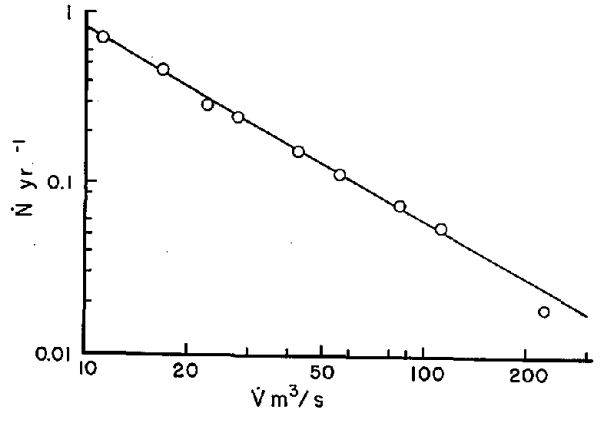


Fig. 3b. Number of floods per year with a peak discharge greater than $\dot{V}$. Station 11-0980 near Pasadena, California during the period 1914–1965.

The values of H, D, and F are given for all ten benchmark stations in Table 1. The correlations with the fractal relation (5) in Fig. 3 are typical of the ten stations. The parameter F is a measure of the relative severity of flooding. The higher the value of F the more likely that severe floods will occur. Our results show that there are clear regional trends in values of F. The values in the southwest including Nevada ($F = 4.13$) and New Mexico ($F = 4.27$) as well as California ($F = 8.11$) are systematically high. The high values can be attributed to the arid conditions and the rare tropical (monsoonal) storm that causes severe flooding. Central Texas ($F = 5.24$) is also high and Georgia ($F = 3.47$) is intermediate. These areas are influenced by hurricanes. The northern tier of states including Massachusetts ($F = 2.72$), Minnesota ($F = 2.95$), Nebraska ($F = 3.47$), and Wyoming ($F = 3.31$) range from low values in the east to intermediate values in the west. Washington ($F = 2.04$) has the lowest value of the stations considered; this low value is consistent with the maritime climate where extremes of climate are rare.

We have also determined the Hurst exponent for the ten benchmark stations. Values of R/S for $T = 5$, 10, 25, and 50 years ($R/S = 1$ for $T = 2$ by definition) are given in Fig. 4a for station 1-1805 (Westfield,

Table 1. Values of the Hausdorff measure H, fractal dimension D, flood intensity factor F, and Hurst exponent H_1 for the 10 benchmark stations

Station	River	(State)	H	D	F	H_1
1-1805	Westfield	(MA)	0.435	1.56	2.72	0.67
2-2185	Oconee	(GA)	0.540	1.46	3.47	0.72
5-3310	Mississippi	(MN)	0.470	1.53	2.95	0.72
6-3440	Little Missouri	(WY)	0.520	1.48	3.31	0.72
6-8005	Elkhorn	(NE)	0.540	1.46	3.47	0.67
7-2165	Mora	(NM)	0.630	1.37	4.27	0.73
8-1500	Llano	(TX)	0.719	1.28	5.24	0.70
10-3275	Humboldt	(NV)	0.616	1.38	4.13	0.66
11-0980	Arroyo Seco	(CA)	0.909	1.09	8.11	0.68
12-1570	Wenatchee	(WA)	0.310	1.69	2.04	0.72

MA) and in Fig. 4b for station 11-0980 (Pasadena, CA). Good correlations are obtained with (13) taking $H_1 = 0.67$ for station 1-1805 and $H_1 = 0.68$ for station 11-0980. Values of H_1 for all ten stations are given in Table 1. The values are nearly constant with a range from 0.66 to 0.73 indicating moderate persistence. It is not surprising that the values of the Hausdorff measures H differ from the values of the Hurst exponent H_1 since the former refers to the statistics of the flood events and the latter to the statistics of the running sum.

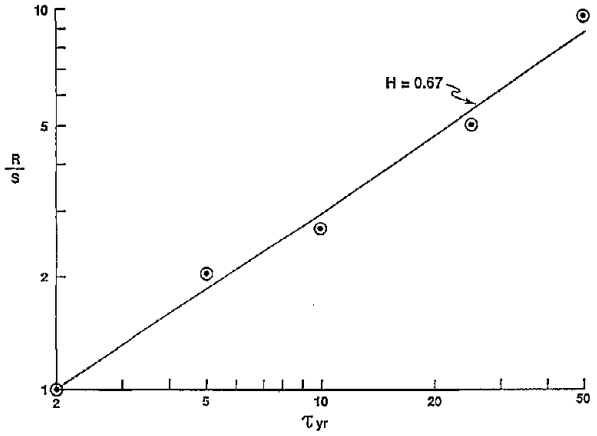


Fig. 4a. The rescaled range (R/S) for several intervals T. Station 1-1805. The correlations are with Eq. (13) and the Hurst exponents H_1 are given.

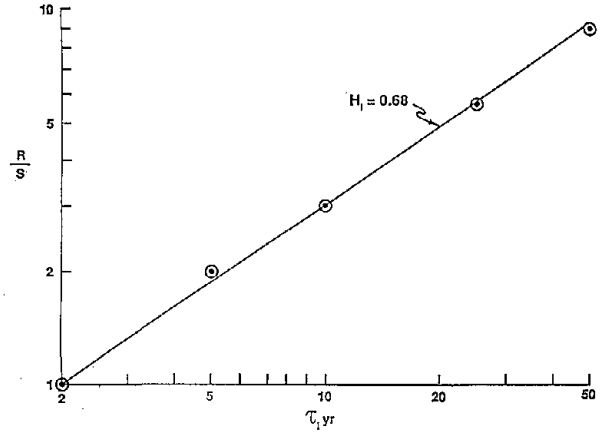


Fig. 4b. The rescaled range (R/S) for several intervals T. Station 11-0980. The correlations are with Eq. (13) and the Hurst exponents H_1 are given.

However, the results indicate that there is considerable variation of α (H, D, and F) but very little variation in H_1. A simple explanation is that the former is sensitive to the Noah effect while the latter is sensitive to the Joseph effect. The relative scaling of floods is sensitive to the skewness of the statistical distribution but is not sensitive to the persistence of flows or floods. An important conclusion is that R/S analysis is not relevant to flood-frequency hazard assessments.

Many statistical distributions have been applied to historical records of floods. Benson [2] has given six statistical correlations for each of his ten benchmark stations. His results for the 2-parameter gamma (Ga), Gumbel (Gu), log Gumbel (LGu), log normal (LN), Hazen (H), and log Pearson type III (LP) are given in Fig. 5a for station 1-1805 and in Fig. 5b for station 11-0980. Also included in each figure is the self-similar (fractal) estimate (F). For large floods the fractal prediction (F) correlates best with the log Gumbel (LGu) while the other statistical techniques predict longer recurrence time for very serious floods. The fractal and log Gumbel are essentially power-law correlations whereas the others are essentially exponential.

While the ten benchmark stations provide a basis for comparing statistical approaches, they hardly made a convincing case that fractal statistics are preferable to alternatives. A principal difficulty is the relatively short time span over which reliable records have been collected. In order to try to overcome this difficulty we have analysed a large number of records and superimposed the results. We have utilized a digitized 40 year data set for 1009 stations unaffected by flood control projects [19]. The distribution of the stations over the United States is given in Fig. 6a. We will separately consider the data from the 18 hydrologic districts, these are illustrated in Fig. 6b.

The largest floods in each of the 40 water years are ordered, the largest annual flood is assigned a period of 40 years, the 2nd largest annual flood a period of 20 years, the 3rd largest annual flood a period of 13.3 years, and so forth. The log of the peak discharge for each flood is plotted against the log of its assigned period and the best straight line, i.e., from Eq. (3), is obtained. Two randomly selected examples are given in Fig. 7.

Results for station 1-860 on the Warner River in Davisville, NH, are given in Fig. 7a. The best fit straight line gives $H = 0.68$; from Eqs. (2), (4), and (7) we have $F = 4.8$, $D = 1.32$ and $\alpha = 1.46$. Results for station 3-2305 on the Big Darby Creek in Darbyville, OH are given in Fig. 7b. The best fit straight line gives $H = 0.386$; from Eqs. (2), (4), and (7) we have $F = 2.43$, $D = 1.61$, and $\alpha = 2.59$.

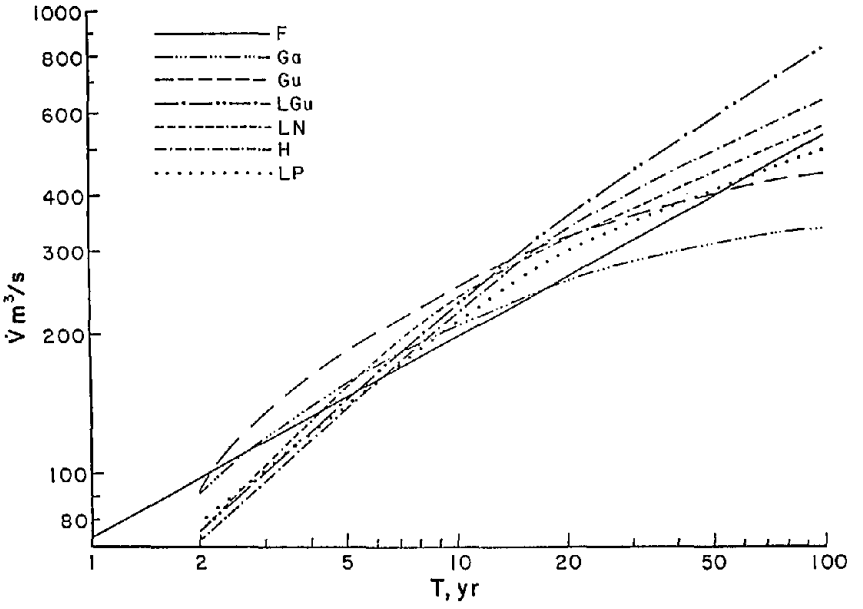


Fig. 5a. Flood frequency predictions for Station 1-1805. The peak discharge $\dot{V}$ is given as a function of recurrence intervals T. The scale-invariant (fractal) prediction, F, is compared with the six statistical predictions given by Benson (1968); 2 parameter gamma (Ga), Gumbel (Gu), log Gumbel (LGu), log normal (LN), Hazen (H), and log Pearson type III (LP).

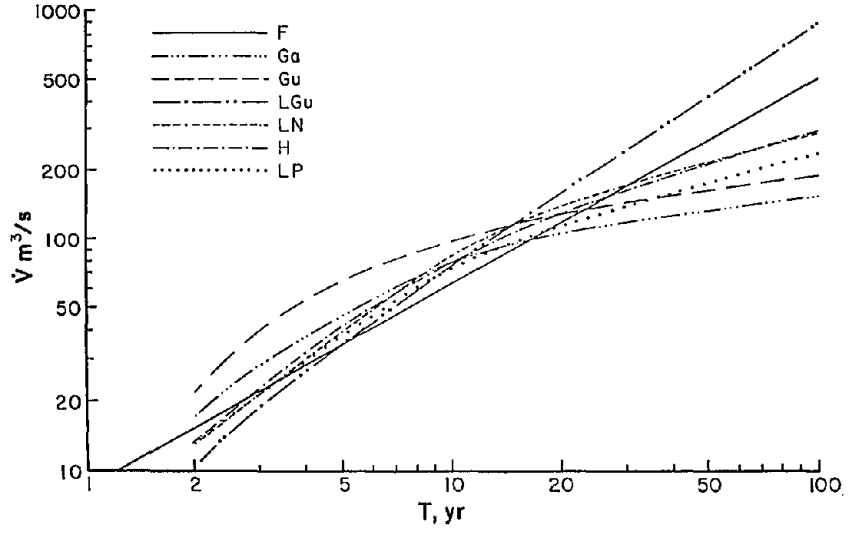


Fig. 5b. Flood frequency predictions for Station 11-0980. The peak discharge $\dot{V}$ is given as a function of recurrence intervals T. The scale-invariant (fractal) prediction, F, is compared with the six statistical predictions given by Benson (1968); 2 parameter gamma (Ga), Gumbel (Gu), log Gumbel (LGu), log normal (LN), Hazen (H), and log Pearson type III (LP).

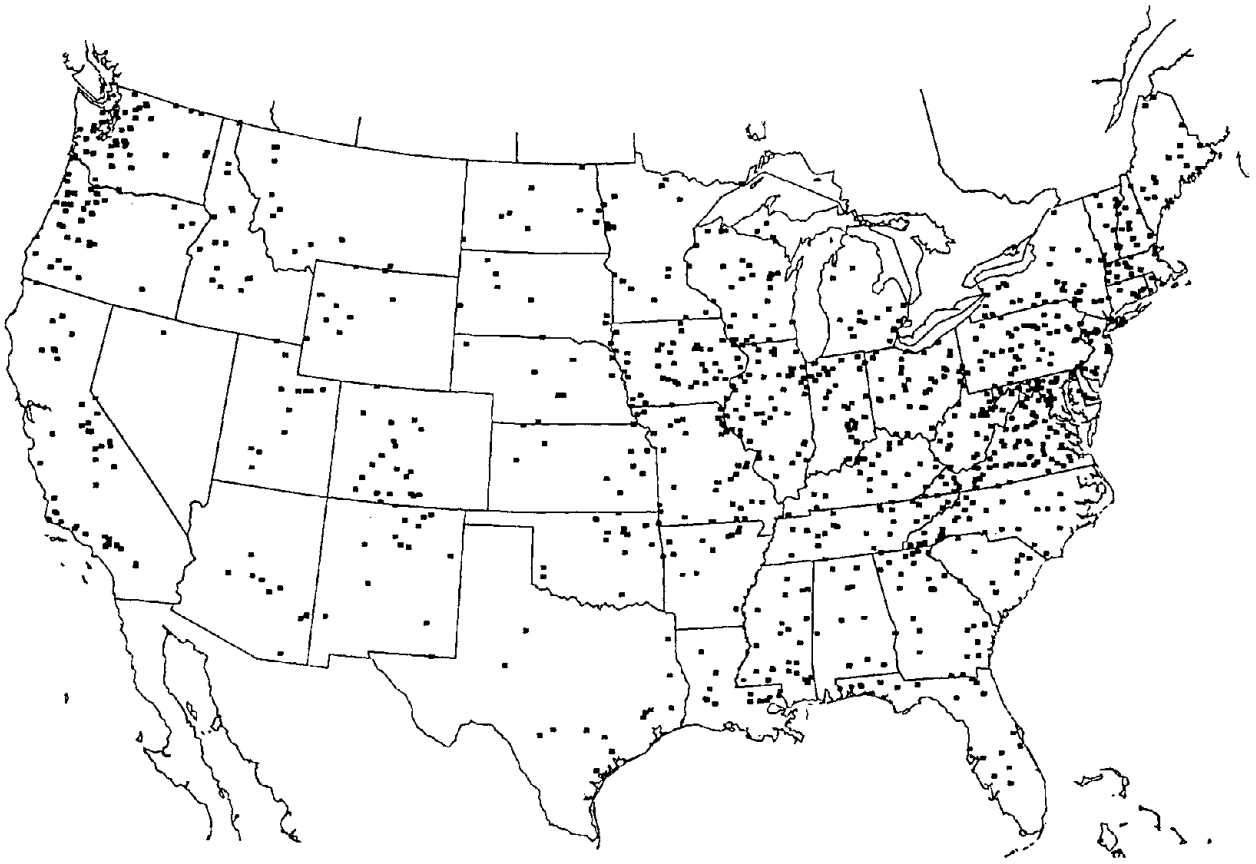


Fig. 6a. Distribution of the 1009 stations that have been analysed.

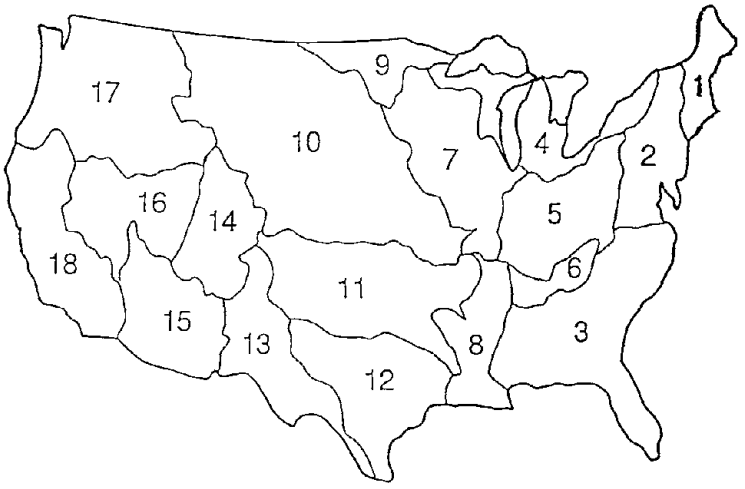


Fig. 6b. Hydrologic regions of the continental United States.

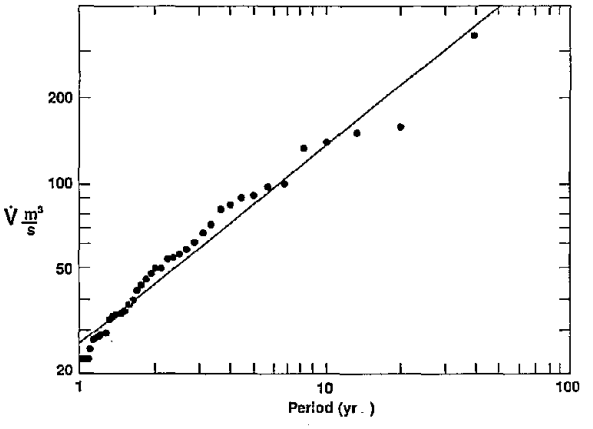


Fig. 7a. The peak daily discharge for the largest annual floods over 40 years as a function of the assigned period: Station 1-0860.

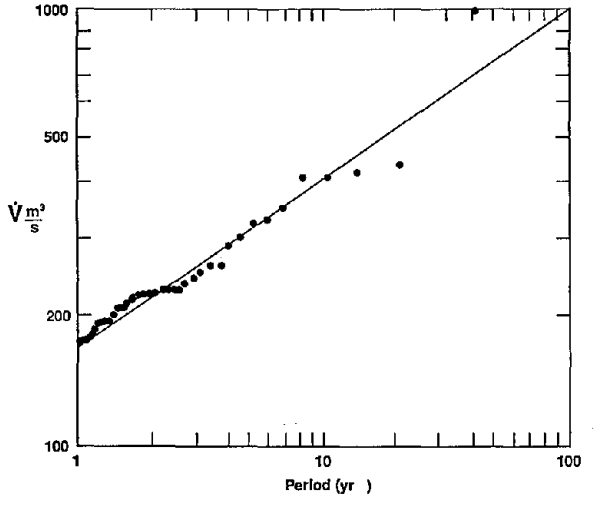


Fig. 7b. The peak daily discharge for the largest annual floods over 40 years as a function of the assigned period: Station 3-2305.

In order to determine the quality of the fit of the data to the fractal relation Eq. (3), the ratio of the measured peak flow to the value predicted by the fractal fit is given for periods of 1, 2, 5, 10, 20, and 40 years in Fig. 8. The 111 stations from hydrologic region 3 are given in Fig. 8a, the 57 stations from region 4 in Fig. 8b, the 10 stations from region 16 in Fig. 8c, and the 100 stations from region 17 in Fig. 8d. If all points were unity the fit would be perfect. The mean deviations from the fractal relation are only a few per cent. The deviations for larger values of the period are greater as would be expected since the individual points are only a few floods. However, the mean values of the 40 year floods are close to the fractal extrapolation. This agreement provides support for the applicability of fractal statistics to the estimation of the flood hazard.

In Fig. 9a the 111 fractal fits for hydrologic region 3 are given, the fits for regions 4, 16, and 17 are given in Figs. 9b, 9c, and 9d. The peak flow at a period of 10 years was normalized by the drainage area upstream of the station. If peak flows were simply proportional to upstream drainage areas in a hydrologic district then all the plots should fall on a single band. In fact, there is more than an order of magnitude variation. This is not surprising but the details of the variations should be helpful in providing a better understanding of the flood hazard.

The regional variations in F are clearly illustrated in Table 2. The highest values of F are generally associated with the arid southwestern states in regions 12, 13, 15, and 18, the mean value of F for these regions is $F = 5.03$. The lowest mean value for F is in region 17, the Pacific Northwest, with $F = 2.08$. In some cases the standard deviations for F in a district are large. For district 18 (primarily California) the mean is 5.34 and the standard deviation is 2.4. In this case much of the deviation can be identified with the presence or absence of snow run off. Those stations with large upstream snow packs have relatively small values for F compared with those stations with little or no upstream snow packs.

4. Conclusions

Historical flood-frequency records have been examined to determine whether fractal (power-law) statistics are applicable. Although it must be recognized that the relatively short duration of historical records restricts the validity of conclusions; nevertheless, quite good agreement is obtained between fractal statistics and observations for 10 benchmark stations and for 1200 other stations in the United States. The basic question in terms of flood hazard assessment is whether extreme floods decay exponentially in time or as a power law. If the power-law behavior is applicable then the likelihood of severe floods is much higher and more conservative designs for dams and land use restrictions are indicated.

For fractal behavior the ratio of the 10 year to the 1 year flood F is also the ratio of the 100 year to the 10 year flood and the ratio of the 1000 year flood to the 100 year flood. We find large regional variations in values of F. In arid regions such as the southwestern United States the values of F are nearly three times the values in more temperate regions such as the northwestern and northeastern corners of the country. Smaller values of F are also found if upstream drainage areas have large snow packs.

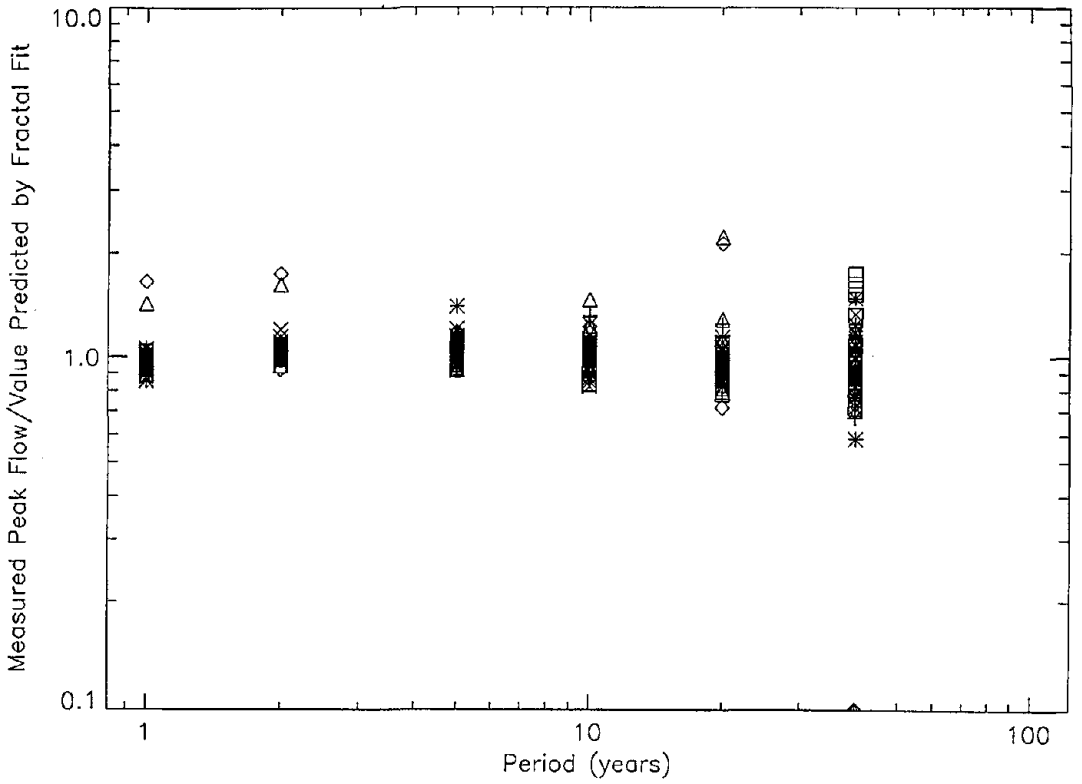


Fig. 8a. Ratio of the observed peak daily discharge to the value predicted by the fractal fit to the data as a function of the assigned period for the 111 stations in region 3.

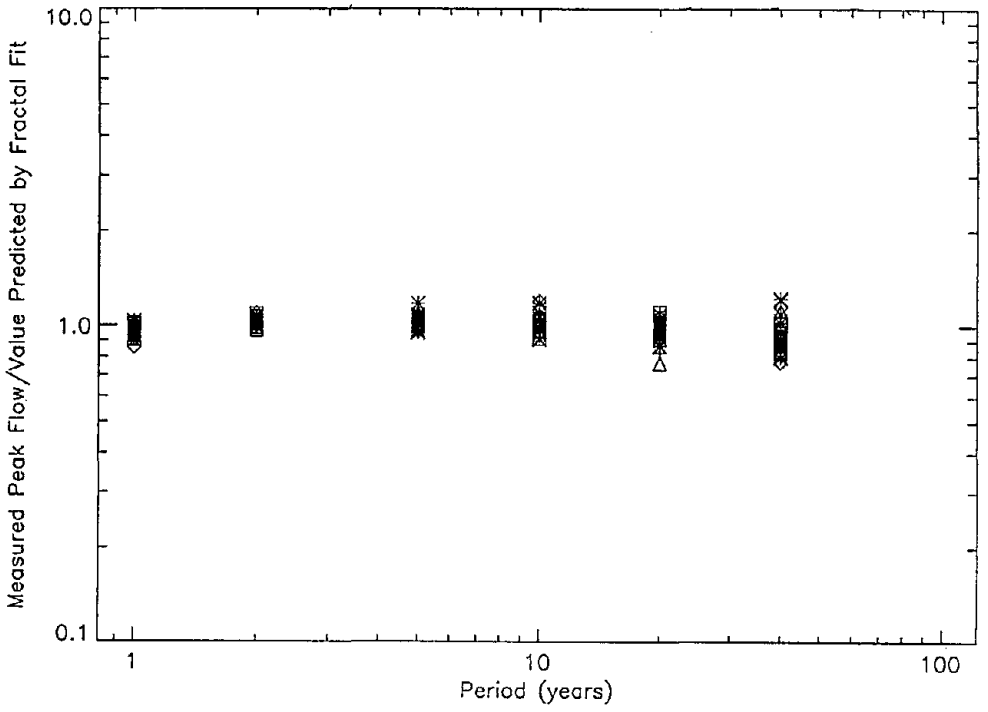


Fig. 8b. Ratio of the observed peak daily discharge to the value predicted by the fractal fit to the data as a function of the assigned period for the 57 stations in region 4.

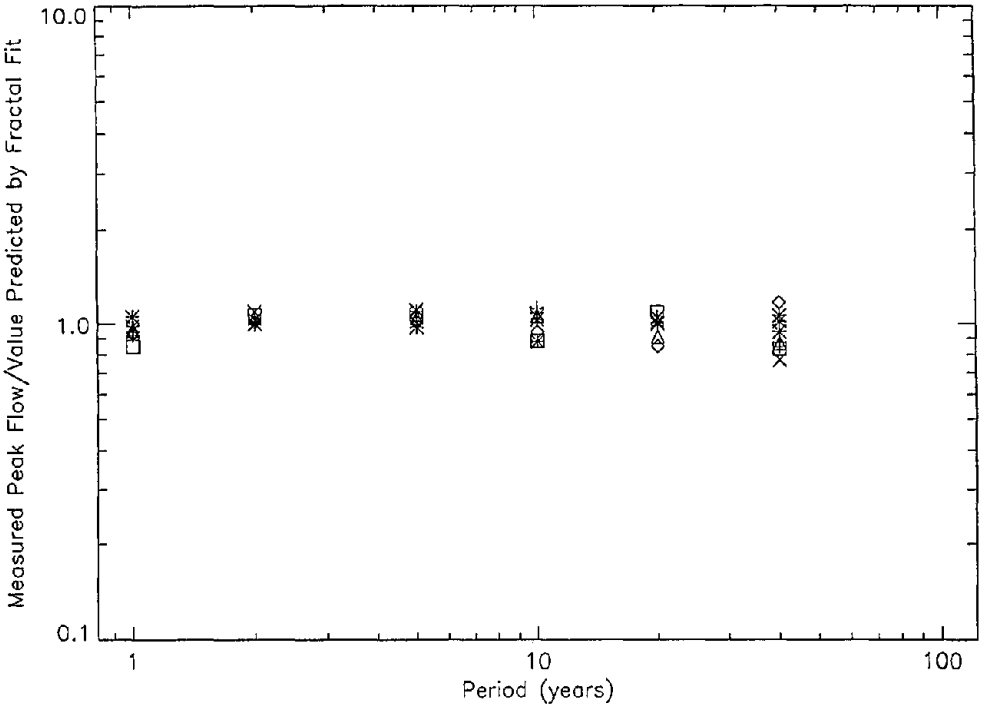


Fig. 8c. Ratio of the observed peak daily discharge to the value predicted by the fractal fit to the data as a function of the assigned period for the 10 stations in region 16.

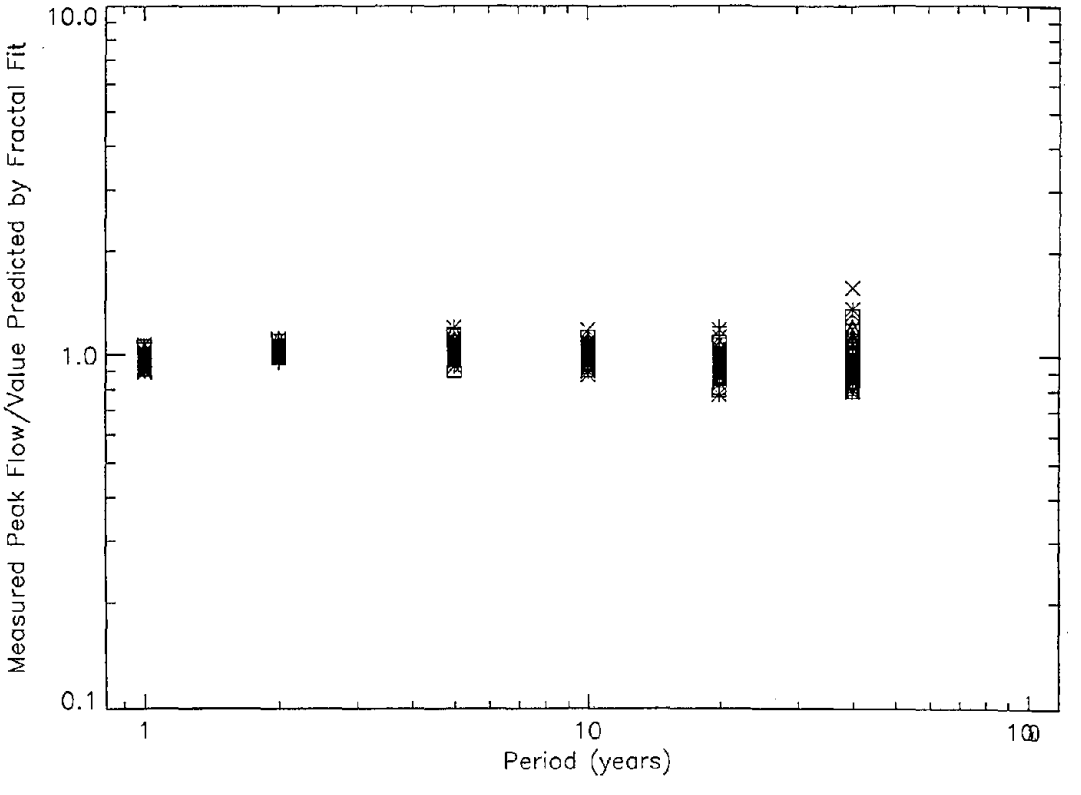


Fig. 8d. Ratio of the observed peak daily discharge to the value predicted by the fractal fit to the data as a function of the assigned period for the 100 stations in region 17.

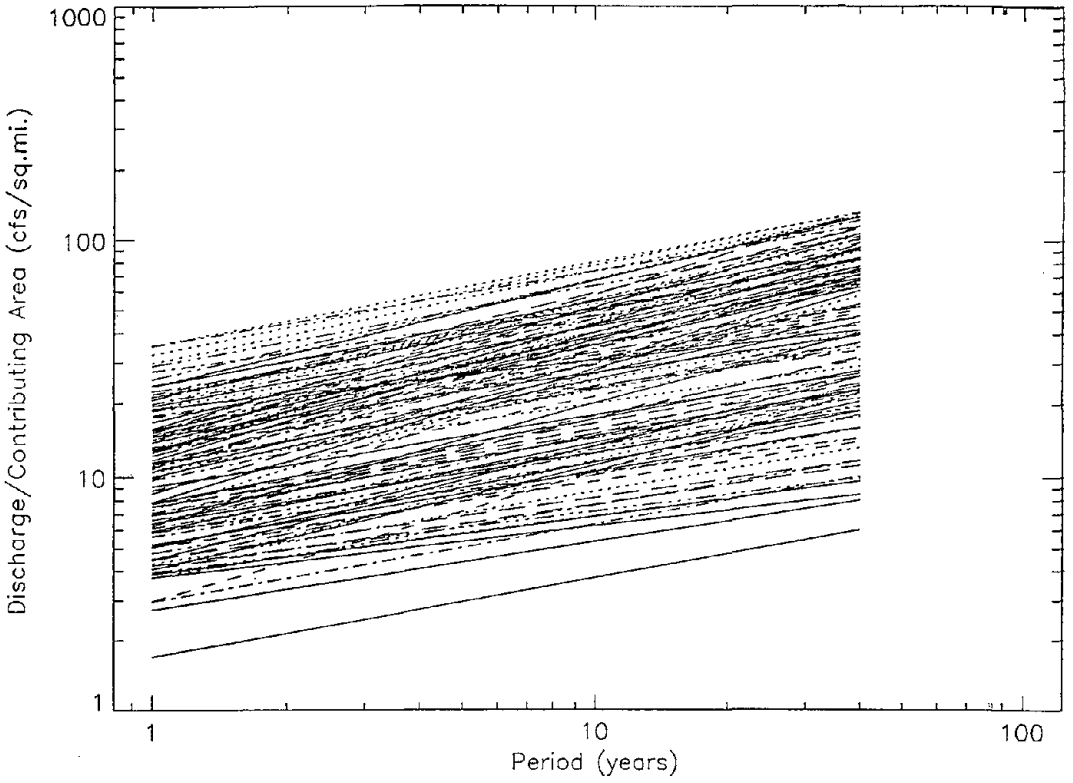


Fig. 9a. Fractal fits of the normalized flood frequency data for the 111 stations in region 3.

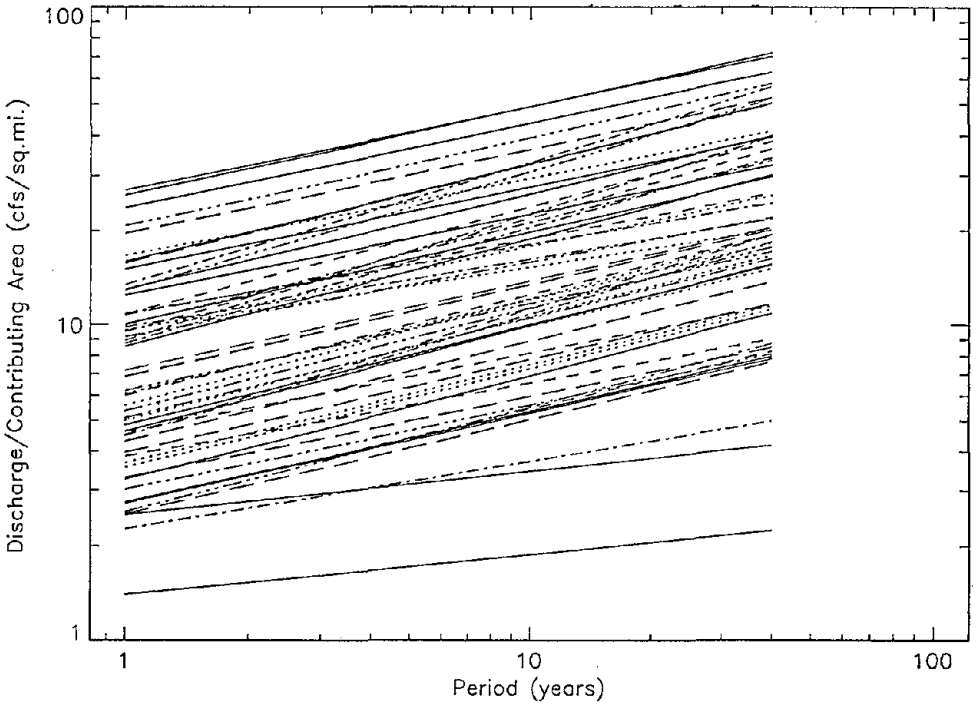


Fig. 9b. Fractal fits of the normalized flood frequency data for the 57 stations in region 4.

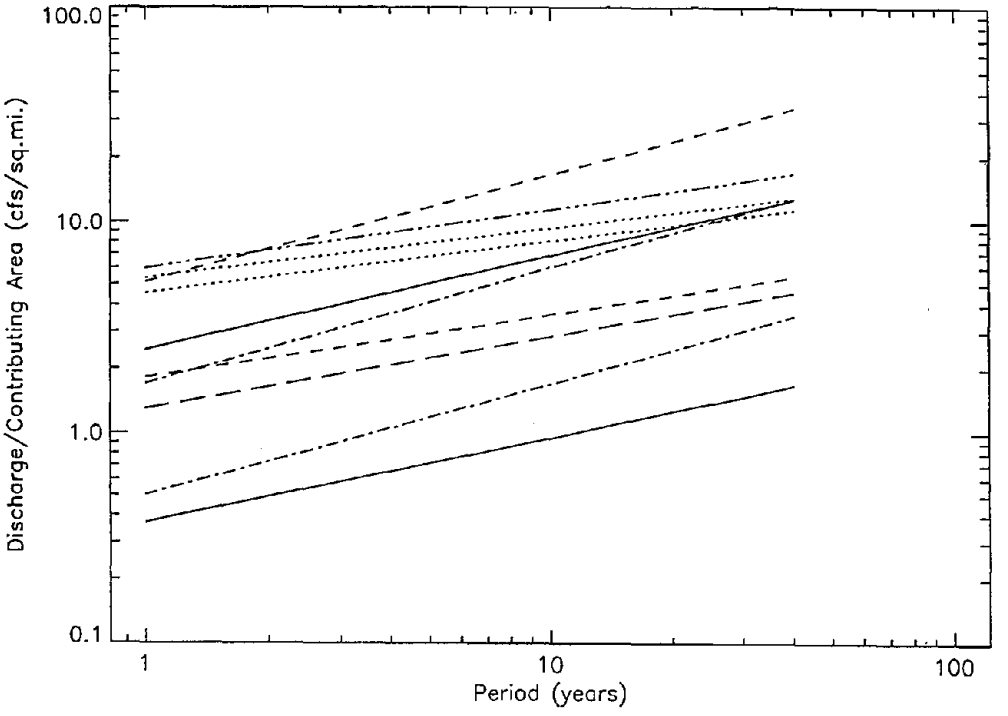


Fig. 9c. Fractal fits of the normalized flood frequency data for the 10 stations in region 16.

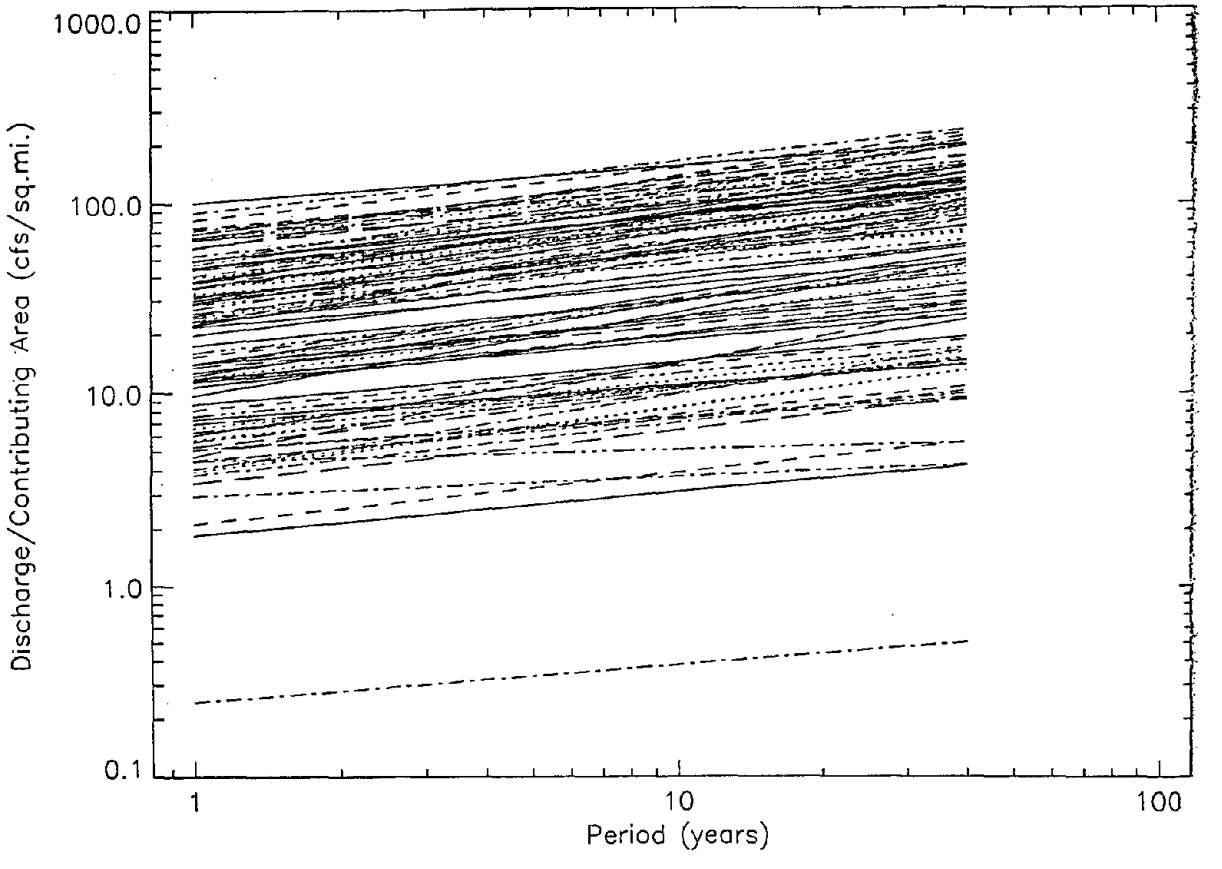


Fig. 9d. Fractal fits of the normalized flood frequency data for the 100 stations in region 17.

Table 2. Average values and standard deviations of the flood intensity factor F for the 18 hydrologic districts

Hydrologic regions	F	Standard deviations	Number of stations
1	2.369	0.377	54
2	2.998	1.313	147
3	2.758	0.617	111
4	2.183	0.289	57
5	2.396	0.509	129
6	2.505	0.324	38
7	2.782	0.738	123
8	3.021	0.979	22
9	4.7	1.586	13
10	3.557	1.677	64
11	3.897	1.801	46
12	4.848	1.559	13
13	4.104	2.121	14
14	2.283	0.51	18
15	6.066	1.08	11
16	2.778	0.752	10
17	2.076	0.357	100
18	5.134	2.4	39

The relevance of R/S analysis to flood frequency forecasting has also been addressed. For the ten bench mark stations we find the Hurst exponent to be $H_1 = 0.7 \pm 0.03$. This value indicates moderate persistence for the floods but also shows that determinations of Hurst exponents are not useful for flood hazard assessments. The Hurst exponent does not correlate with the fractal flood parameter F. In the terms introduced by Mandelbrot and Wallis [10] the Hurst exponent is sensitive to the Joseph effect or persistence of events whereas the fractal flood parameter F is sensitive to the Noah effect or skewness of the statistical distributions of floods.

It certainly remains to be demonstrated that fractal flood frequency statistics are generally valid. However, the success indicated in the results given here raises the interesting question whether the underlying physical processes are inherently fractal. Fractal statistics will be applicable to any scale invariant process. They are also applicable to dynamical systems that exhibit self-organized criticality [20]. One speculative conclusion is that the storms that generate floods are associated with the self-organized critical behavior of the atmosphere.

5. References

[1] A. Hazen, Flood Flows, John Wiley, New York, (1930).
[2] M. A. Benson, Uniform flood-frequency estimating methods for federal agencies, Water Resour. Res. 4 , 891–908 (1968).
[3] H. J. Morel-Seytoux, Forecasting of flows-flood frequency analysis, in Modeling of Rivers, H.W. Shen, ed., John Wiley, New York (1979), pp. 3.1–3.46.
[4] U.S. Water Resources Council, Floods flow frequency, Hydrology Committee, Bull. 17B, Washington, DC (1981).
[5] J. R. M. Hosking and J. R. Wallis, The value of historical data in flood frequency analysis, Water Resour. Res. **22**, 1606–1612 (1986).
[6] M. Jin and J. R. Stedinger, Flood frequency analysis with regional and historical data, Water Resour. Res. **25**, 925–936 (1989).
[7] D. L. Turcotte and L. Greene, A scale-invariant approach to flood-frequeny analysis, Stochastic Hydrol. Hydraul. **7**, 33–40, (1993).
[8] H. E. Hurst, Long-term storage capacity of reservoirs, Am. Soc. Civil Eng. Trans. **116**, 770–799 (1951).
[9] H. E. Hurst, Methods of using long-term storage in reservoirs, Insti. Civil Eng. Proc. Part 1, **5**, 519–590 (1956).
[10] B. B. Mandelbrot and J. R. Wallis, Noah, Joseph, and operational hydrology, Water Resour. Res. **4**, 909–918 (1968).
[11] B. B. Mandelbrot and J. R. Wallis, Computer experiments with fractional Gaussian noises. Parts I, II, III, Water Resour. Res. **5**, 228–267 (1969).
[12] B. B. Mandelbrot and J. R. Wallis, Some long-run properties of geophysical records, Water Resour. Res. **5**, 321–340 (1969).
[13] B. B. Mandelbrot and J. R. Wallis, Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence, Water Resour. Res. **5**, 967–988 (1969).
[14] J. Feder, Fractals, Plenum Press, New York (1988).
[15] D. L. Turcotte, Fractals and Chaos in Geology and Geophysics, Cambridge University Press, Cambridge, (1992).
[16] H. E. Hurst, R. P. Black, and Y. M. Simaika, Long-term Storage, Constable, London (1965).
[17] A. R. Green, Magnitude and Frequency of Floods, Part 1-A, U.S. Geol. Survey, Water Supply Paper 1671 (1964), pp. 212–213.
[18] I. E. Young and R. W. Cruff, Magnitude and Frequency of Floods, Part 11, U.S. Geol. Survey, Water Supply Paper 1685 (1976), pp. 714–715.
[19] J. R. Wallis, D. P. Lettenmaier, and E. F. Wood, A daily hydroclimatological data set for the continental United States, Water Resour. Res. **27**, 1657–1663 (1991).
[20] P. Bak, C. Tank, and K. Wiesenfeld, Self-organized criticality, Phys. Rev. **A38**, 364–374 (1988).

About the author: *Donald L. Turcotte is Maxwell Upson Professor of Engineering and a faculty member of the Department of Geological Sciences at Cornell University.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 391 (1994)]

Short-Record-Based Extreme Wind Simulation

Volume 99 Number 4 July–August 1994

Edmond D. H. Cheng and Arthur N. L. Chiu

Department of Civil Engineering, University of Hawaii at Manoa, Honolulu, Hawaii 96822

In order to utilize limited historical wind records for estimating extreme wind speeds for natural hazards damage mitigation, a Markov chain model for generating long-term annual extreme winds, on the basis of short-term records, is investigated. Basically, this simulation model consists of three components. They are state of wind speeds, wind speed distribution functions, and transition probability matrices.

The basic strategy of our simulation model is to generate the time series of hourly wind speeds in parts: for those winds associated with well-behaved climates and those with extreme winds. Applications of this model to generate long-term extreme winds, on the basis of short records at Houston Intercontinental Airport of Texas, are demonstrated.

Key words: extreme wind speed; Gumbel distribution; Markov chain model; simulation; transition probability matrix.

Accepted: March 22, 1994

1. The Model

Recent efforts have been made to improve the methods of utilizing short-term wind records [1,2,3,4]. Our contribution to this endeavor is the development of stochastic simulation models which generate hourly wind data on a daily cycle basis [5,6].

The three essential elements in the Markov chain model are: state of wind speeds, wind speed distribution functions, and transition probability matrices.

1.1 State of Wind Speeds

In the simulation process for a given wind site, the first step is to divide the entire range of observed wind speeds into a finite number of states. This is performed with reference to the probability histogram derived from the observed wind data for that site. A computer program called WSTAT was developed for performing this task.

1.2 Distribution Functions

The second basic element involves the wind speed distribution functions, viz., the probability density functions (PDF) and the cumulative distribution functions (CDF) of wind speeds in various states. In this paper, four types of PDFs are utilized to fit a wind speed histogram, viz., uniform, linear, exponential and Weibull distribution functions. The exponential and Weibull distributions are exclusively reserved for the last state in which extreme winds are involved.

1.3 Transition Probability Matrices

The transition probability p_{ij} is actually a conditional transition probability of wind speed v_τ going from one state i at hour τ to wind speed $v_{\tau+1}$ of state j at hour $\tau+1$ or

$$p_{ij} = p(v_{\tau+1} = j | v_\tau = i). \quad (1)$$

With m states determined, an $m \times m$ transition probability matrix PM can be determined as

$$PM = [p_{ij}] \text{ for } i,j = 1,2,...,m \quad (2)$$

in which, p_{ij} have the following properties:

$$\sum_{j=1}^{m} p_{ij} = 1, \text{ for } i = 1,2,...,m \quad (3)$$

and

$$p_{ij} > 0, \text{ for all } i \text{ and } j. \quad (4)$$

In this paper, a day is divided into several periods. Similarly, the variation of mean monthly wind speeds is accounted for by grouping consecutive months with similar wind speed trends into a number of seasons for a year. If the number of periods in a day and the number of seasons in a year are R and S, respectively, then there will be $R \times S$ transition probability matrices in the simulation process. A typical transition probability matrix for a given period r and season s can be expressed as

$$PM(s,r) = [p_{ij}^{s,r}] \quad (5)$$

where $s = 1,2,...,S$, and $r = 1,2,...,R$. A computer program called WTPM was developed for calculating the $PM(s,r)$.

2. Simulation Procedures

Some sequential steps of generating hourly wind data points at a given site are briefly described as follows:

1. Divide the historical wind data into subsets so that program WSTAT may be executed.
2. Define R, number of periods in a day, and S, number of seasons in a year, so that program WTPM is activated.
3. Calculate PDFs and CDFs of the historical data.
4. By means of Eq. (5), compute $PM(s,r)$, the transition probability matrix of period r in season s. A total of $R \times S$ transition probability matrices will be obtained.
5. Determine the state of the succeeding hour's wind speed. For any given wind speed in state i of this current hour (with specified period r and season s), the succeeding hour's wind speed state interval "k" can be determined.
6. Determine the value of the succeeding hour's wind speed. With the state k determined in Step 5, the simulated wind speed for the succeeding hour can thus be obtained.

Repeat steps 5 and 6 until a desired period of simulation is attained. A computer program called WSIM was developed for generating the hourly wind data.

3. Markov Property and Stationary Tests

In order to substantiate the major assumptions made earlier, a test must be performed of the Markov property, i.e., the existence of dependency between two adjacent hourly wind speeds. This simulation technique is only applicable to stationary time series; the intended simulation model is a stationary first order Markov chain. Consequently, a test of stationarity of the historical wind speed times series is necessary prior to the acceptance of the simulated results. Anderson and Goodman's method [7] was used in performing these tests in Sec. 4.

4. Application

The simulation model based on the described procedures is applied to wind data collected at the Houston Intercontinental Airport in Texas. In this illustration, three periods in a day (1:00 a.m.–9:00 a.m., 10:00 a.m.–7:00 p.m., and 8:00 p.m.–midnight) and two seasons in a year (November–May and June–October) were considered. The periods of a day were decided from the averaged diurnal wind speeds at the site. Therefore, by using Eq. (5), six transition probability matrices were calculated.

Based on the 20 years (January 1973 to September 1992) of hourly wind records available at Houston Intercontinental Airport, eight simulation runs were made. Each run generated 100 years of hourly wind speeds. Historical record periods for the eight

runs are: the first 5 years (January 1973–December 1977); the second 5 years (January 1978–December 1982); the third 5 years (January 1983–December 1987); the fourth 5 years (January 1988–September 1992); the first 10 years (January 1973–December 1982); the second 10 years (January 1983–September 1992); the first 15 years (January 1973–December 1987); and all 20 years (January 1973– September 1992).

The annual extreme wind speeds of the historical data (1963–1990), and of eight sets, each 100 years long, of simulated hourly data, are plotted on Type I probability paper (Fig. 1). Using the Gumbel line fitted to the historical data as the reference, the performance of the simulation model is summarized in Table 1. As shown in this table, the deviations of the simulated 25-year, 50-year, or 100-year wind speeds from the reference Gumbel line were measured by S_v, Cramer-Rao's standard deviation of the inherent sampling error of the historical records [8,9]. As indicated in Table 1, the differences between the simulated annual extreme wind speeds and the values obtained from the reference Gumbel line of historical data at 25 year or 50 year or 100 year recurrence intervals are all smaller than one S_v. This result is very encouraging.

The plots of cumulative distribution functions for the historical as well as the 100 year generated records at Houston Intercontinental Airport are presented in Fig. 2. As shown in this figure, the curves derived from the generated records closely match those of the historical records, which implies that the characteristics of the historical wind speeds at Houston Intercontinental Airport were adequately represented.

Table 1. Estimated annual extreme wind speeds from historical records and from simulation methods at Houston Intercontinental Airport

Data period	Recurrence interval (in years)	V_a	S_v	V_s	$\frac{V_a - V_s}{S_v}$
1963–1990	25	26.6	1.78		
	50	28.8	2.13		
	100	30.9	2.48		
1/1973–12/1977	25			27.6	−0.56
	50			28.7	0.05
	100			29.9	0.40
2/1978–12/1982	25			27.8	−0.67
	50			29.1	−0.14
	100			30.3	0.24
1/1983–12/1987	25			27.4	−0.45
	50			28.4	0.19
	100			29.4	0.60
1/1988–9/1992	25			27.4	−0.45
	50			28.6	0.09
	100			29.8	0.44
1/1973–12/1982	25			27.4	−0.45
	50			28.6	0.09
	100			29.7	0.48
1/1983–12/1987	25			28.1	−0.84
	50			29.4	−0.28
	100			30.7	0.08
1/1973–12/1987	25			27.6	−0.56
	50			28.8	0.00
	100			30.0	0.36
1/1973–9/1992	25			27.9	−0.73
	50			29.6	−0.38
	100			31.3	−0.16

V_a = extreme wind speed from annual series in m/s and 10 m above ground level.
V_s = simulated extreme wind speed in m/s and 10 m above ground level.
S_v = Cramer-Rao's standard deviation of inherent sampling error of historical records.

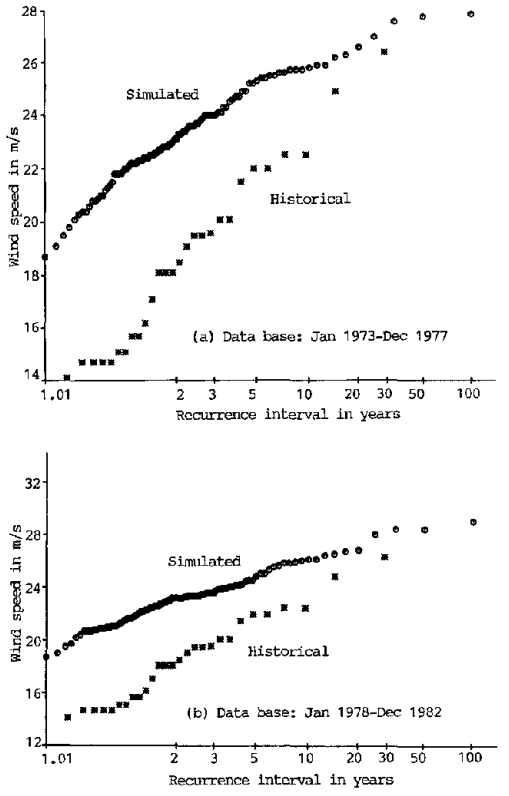


Fig. 1a, b. Type I distributions of historical and simulated annual extreme wind speeds at Houston Intercontinental Airport at 10 m above ground level.

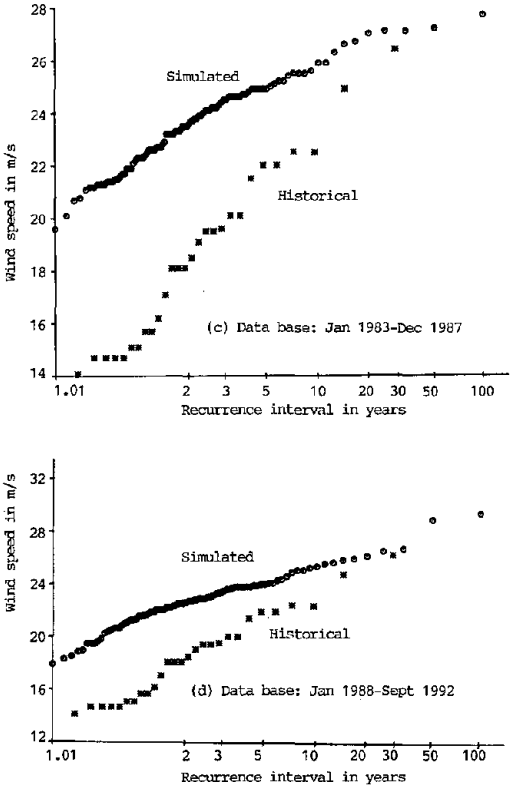


Fig. 1c, d. Type I distributions of historical and simulated annual extreme wind speeds at Houston Intercontinental Airport at 10 m above ground level.

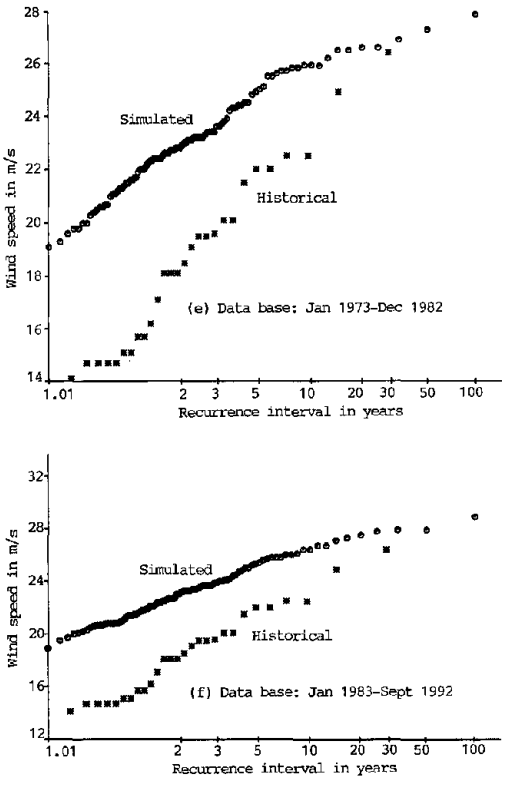


Fig. 1e, f. Type I distributions of historical and simulated annual extreme wind speeds at Houston Intercontinental Airport at 10 m above ground level.

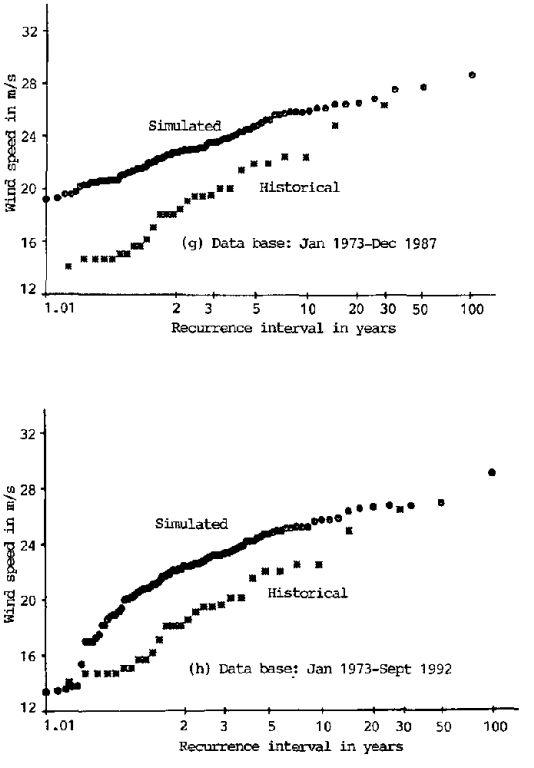
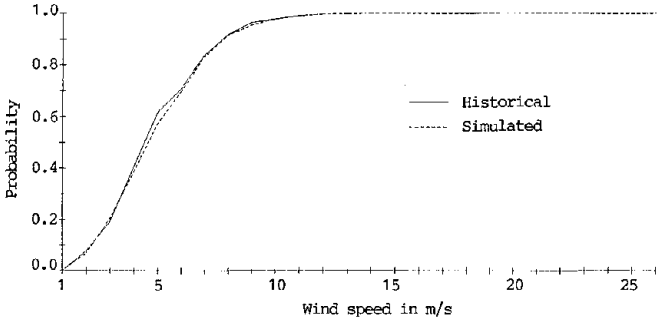
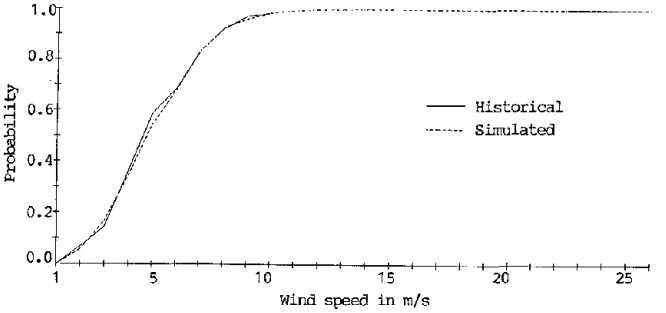


Fig. 1g, h. Type I distributions of historical and simulated annual extreme wind speeds at Houston Intercontinental Airport at 10 m above ground level.

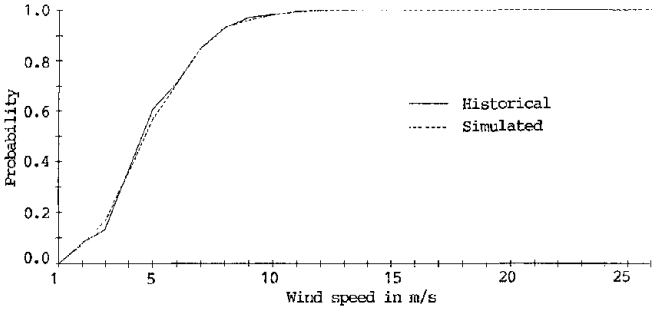


(a) Data base: Jan 1973-Dec 1977

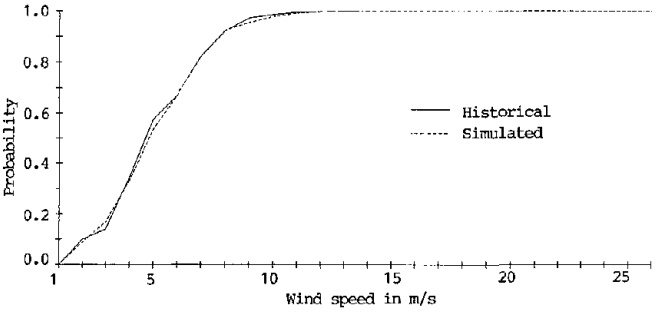


(b) Data base: Jan 1978-Dec 1982

Fig. 2a, b. Cumulative distribution functions of historical and simulated hourly wind speeds at Houston Intercontinental Airport at 10 m above ground level.

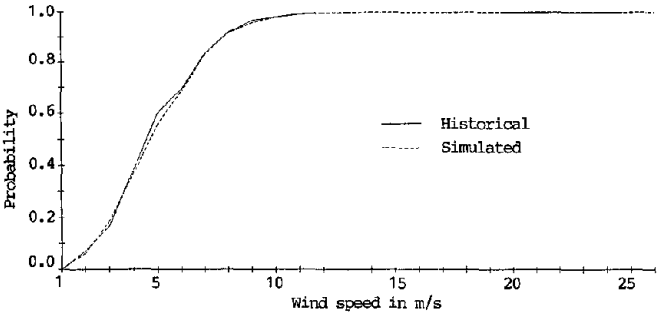


(c) Data base: Jan 1983-Dec 1987

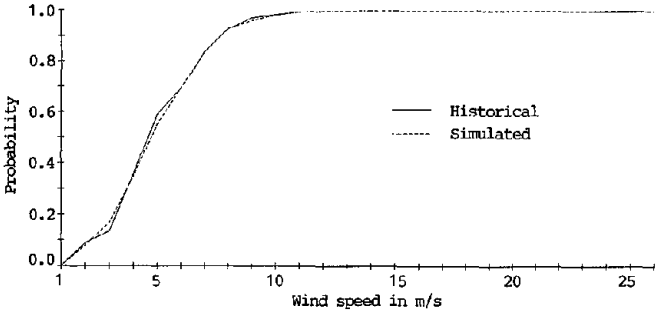


(d) Database: Jan 1988-Sept 1992

Fig. 2c, d. Cumulative distribution functions of historical and simulated hourly wind speeds at Houston Intercontinental Airport at 10 m above ground level.

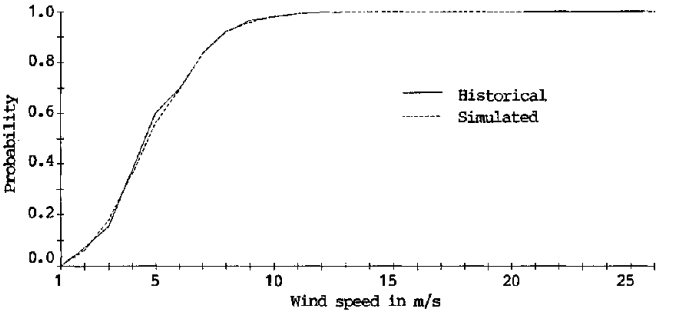


(e) Data base: Jan 1973–Dec 1982

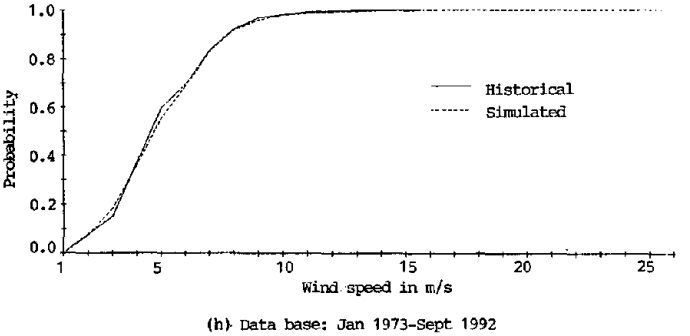


(f) Data base: Jan 1983–Sept 1992

Fig. 2e, f. Cumulative distribution functions of historical and simulated hourly wind speeds at Houston Intercontinental Airport at 10 m above ground level.



(g) Data base: Jan 1973–Dec 1987



(h) Data base: Jan 1973–Sept 1992

Fig. 2g, h. Cumulative distribution functions of historical and simulated hourly wind speeds at Houston Intercontinental Airport at 10 m above ground level.

5. Conclusions

A procedure for predicting extreme wind speeds at a location along the Gulf Coast is demonstrated. The results obtained from the application of this model are very encouraging. Although 20 years of data were available for the particular station in the illustration, computer simulation runs were made on the basis of 20 year, 15 year, 10 year, and 5 year database. It has been shown that it is not necessary to have 20 years of continuous data and that even a 5 year record is adequate for showing good comparison between the simulated results and historical data. Further research effort is being undertaken at the University of Hawaii at Manoa to study the applicability of this method to other stations along the Gulf Coast and other parts of the continental United States.

Acknowledgment

Partial support of this study by the National Science Foundation through Grant No. BCS-9122224 is gratefully acknowledged.

6. References

[1] L. Gomes and B. J. Vickery, On the prediction of extreme wind speeds from the parent distribution, J. Indust. Aerodynam. 2(1), 21–36 (1977).
[2] A. G. Davenport, Wind structure and wind climate, Proceedings of the International Research Seminar on Safety of Structures Under Dynamic Loading, Tapir Publishers, Trondheim, Norway, 209–256 (1978).
[3] E. Simiu, J. J. Filliben, and J. R. Shaver, Short-term records and extreme wind speeds, J. Struct. Div., ASCE 108(ST11), 2571–2577 (1982).
[4] M. Grigoriu, Estimates of design winds from short-records, J. Struct. Div., ASCE 198, 1034–1048 (1982).
[5] E. D. H. Cheng and A. N. L. Chiu, Extreme winds simulated from short-period records, J. Struct. Eng., ASCE 111, 77–94 (1985).
[6] E. D. H. Cheng and A. N. L. Chiu, Extreme winds generated from short records in a tropical cyclone-prone region, J. Wind Eng. Indust. Aerodynam. 28, 69–78 (1988).
[7] T. W. Anderson and L. A. Goodman, Statistical inference about Markov chains, Ann. Math. Statist. 28, 89–110 (1957).
[8] E. J. Gumbel, Statistics of Extremes, Columbia University Press, New York, New York (1958) p. 227.
[9] E. Simiu and R. Scanlan, Wind Effects on Structures, John Wiley & Sons, Inc., New York (1986) p. 87.

About the authors: Edmond D. H. Cheng and Arthur N. L. Chiu are professors of civil engineering, University of Hawaii at Manoa.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 399 (1994)]

Getting the Most From Your Extreme Wind Data: A Step by Step Guide

Volume 99 Number 4 July–August 1994

David Walshaw

University of Newcastle upon Tyne,
Newcastle upon Tyne, U.K.

Models for extremes of environmental processes have been studied extensively in recent years. The particular problems arising when attempting to estimate return levels from sequences of measurements on the appropriate variables have been considered in some detail. In particular, the aspects of seasonal variation and short-range dependence have received a great deal of attention. In this paper we present a case study based on 10 years of hourly wind speed measurements collected at a U.K. site, elucidating the most successful procedure emerging from an extensive study of this data. The basic model (in which an extreme value distribution is fitted to cluster peak excesses over a high threshold) is standard. However the emphasis is on a number of practical problems which will arise when such models are fitted to wind speeds, but which have received little consideration. These include: model selection and assessment of model adequacy when the threshold, and some or all of the parameters, are allowed to vary seasonally; the choice of the best combination of threshold and cluster identification procedure; and the choice of a measure of precision for return level estimates. The aim is to suggest an algorithm which can be generally applied to the problem of gust return level estimation at individual sites.

Key words: extreme value theory; generalized pareto distribution; peaks over threshold; return levels; statistics of extremes; wind speed.

1. Introduction

Threshold models for exceedances have been widely adopted in recent years in the study of extremes of environmental processes. The main advantage of such models over the so-called "classical" extreme value models (in which a limiting distribution is fitted to the largest order statistics selected form fixed time intervals) is their greater flexibility in the manner in which events are identified as "extreme." This generally leads to a larger number of extreme events being available for analysis, and this in turn to more precise estimates for return levels and return periods.

The price paid for the increased efficiency of data exploitation and consequent improvements in estimation precision is, as one would expect, a greater complexity of model. Seasonal variation and short-range correlation, almost always present in environmental time series, can no longer be ignored in the manner of a traditional "annual maxima" analysis (or "Gumbel analysis"). Instead they must be given careful consideration. Models which take account of both of these features have received considerable attention in the literature (e.g., Refs. [2,3,7,8] and the associated discussion).

In this paper we consider a complete study of a sequence of wind speed measurements recorded at a single U.K. site. We address some of the practical complexities that arise when adopting a threshold-based approach to extremes of environmental time series. In particular, the related issues of

1. choosing a threshold large enough for the distribution of excesses to approximate to a limiting form,
2. allowing the threshold and some or all of the parameters to vary seasonally,
3. employing a threshold-based declustering method for identification of storm peaks,

give rise to a situation which requires some careful consideration in terms of the practical application of existing models.

The theoretical arguments supporting the use of threshold models in the manner considered in this paper, already validated in previous studies (e.g., Ref. [8]), suggest that the techniques employed should be applicable at any site at which the natural mechanism underlying the generation of extreme winds is not capable of taking on several distinct forms (e.g., hurricanes and conventional storms). Thus the approach considered here could be viewed as a possible algorithm for the estimation of the extreme wind potential at any site in a temperate climate.

2. Background to the Study

2.1 Wind: The Variable

The behaviour of wind velocity as a continuous variable demonstrates certain characteristics which distinguish it from other environmental variables. In common with other such variables, clear seasonal patterns and short-range dependence are strong features of the wind climate at most locations. However, in comparison with these others, wind velocity is fairly well-behaved in a number of ways. Unlike sea-level (Ref. [9]), wind speed does not naturally break down into distinct components, and unlike rainfall, the wind does not arrive in clearly identifiable episodes. In comparison with many environmental phenomena, wind velocity is not subject to *very* violent departures from the norm. Although a wind velocity of 200 mph may sound rather severe, from a statistical point of view such departures from mean levels are small compared with those occasionally demonstrated by rainfall levels over short periods, flow rates in rivers, and concentrations of certain pollutants. The relative stability of wind velocity is more akin to sea-level behaviour, but wind speed differs from sea-level in being one of the most rapidly varying of all environmental variables. Conditional on the underlying "level" of the wind (characterized by storms and periods of calm), many distinct gusts can be observed in periods as short as several minutes. In a sense therefore, while being rather stable, the wind can provide us with a great deal of information in a relatively short time. This strengthens arguments supporting limiting asymptotic distributions for the most extreme gusts, and potentially allows us to make inferences about long-period return levels from comparatively short runs of measurements.

2.2 Extreme Value Models: Exploiting the Variable

We consider the problem of estimating gust return levels for specified periods of the order of 50 or more years, when data available consist of recorded maximum gusts taken over short intervals (say 1 hour or 1 day), and are collected over a time period which may be short in comparison to the return periods (perhaps less than 10 years). In such situations, a classical approach based on annual maxima is unworkable, due to sparsity of data. Methods which make use of several order statistics from each year (for example the "r largest" approach advocated by Tawn [9] in analysing extreme sea-levels) can produce viable estimates of 50 year gust return levels from as little as 10 years of data [10]. However, such methods must take account of serial correlation, and are vulnerable to the effects of seasonal variation. Seasonal effects could be incorporated into the models, but given the additional complexity this would entail, it is thought preferable to convert to a threshold-based approach. The main advantage over the use of order statistics from fixed time intervals is the additional flexibility in the choice of extreme events for analysis. This arises from allowing the number of such events which occur over a fixed period to vary according to the behaviour of the wind during that time. Serial correlation can be dealt with by identifying clusters of observations above a threshold, which are deemed to be correlated, and discarding all but the largest observation within each cluster. The aim here is to filter out a set of independent "cluster peak excesses" for further analysis (Ref. [7]). Seasonal variability in the behaviour of extremes can be incorporated by allowing the

threshold (above which events are deemed to be extreme), and the distribution of excesses over this threshold, to vary through the year. However the justification for such a model is not immediate and is worth considering in a little more depth.

It is usual in strongly seasonal climates for the occurrence of truly extreme wind speeds to be confined to a certain part of the yearly cycle. In the U.K. for example, it is very unusual for wind damage to occur outside the period October through March. However a model for extreme values which *takes account* of this seasonality will select as extreme events all gusts which are large *given the time of year*. If the probability of important levels being exceeded during certain seasons is negligible, then there is only a point to modelling the extremes observed during these periods if we believe that they can supply additional information about what may happen in the seasons in which genuinely large events *can* occur. For this to be the case, we must assume that there is some homogeneity in the extremal behaviour across the different seasons—that in some sense it is fundamentally the same mechanism which is responsible for the generation of large gusts throughout the year, and it is just some of the associated parameters of this mechanism which change. Fortunately, there are often good reasons for making this assumption. In temperate climates, it is essentially the same alternating passage of anticyclones and depressions which leads to all the storms which occur throughout the year. It is merely the severity of these systems which is seasonally variable. Hence it seems reasonable to assume that the manner in which large events cluster together will be broadly homogeneous throughout the year.

A further, more tentative contention is that the patterns of turbulence caused by the local terrain around a site also remain essentially unchanged throughout the annual cycle. Since it is this turbulence that is the cause of gusting, i.e., very short term fluctuations away from the mean wind speed, and since the systems generating sequences of high or low mean speeds appear to differ from season to season only in their severity, we suggest that the *shape* of the upper tail in the distribution of gusts could well be homogeneous throughout the year (i.e., the distribution of extremes varies seasonally only in terms of location and scale). In terms of fitting extreme value distributions to large gusts, this would be reflected by the shape parameter (denoted here by k) being held constant across all seasons.

Of course homogeneity conditions on both clustering behaviour of large gusts, and the shape of the upper tail in their distribution, must be verified from the data. However, previous studies suggest that such assumptions are often validated, and can then provide an important route to a more efficient exploitation of data. This will be demonstrated in the case study which follows. Working with hourly maximum gusts collected at Sheffield University for the U.K. Meteorological Office over a 10 year period 1975–1984, we identify four steps to the estimation of return levels. Implementing this algorithm, we obtain useful return level estimates for 10, 50, and 1000 years. The level of precision attached to these estimates is greater than any achieved via a whole range of conventional analyses applied to the same data, as well as some more novel models (see Ref. [10]).

3. Step 1—Generating a Stationary Series

3.1 Dealing with Seasonal Variation

Davison and Smith [3] identify two basic approaches for handling seasonal data:

1. the removal of known seasonal components to create a stationary (prewhitened) series;
2. a separate seasons approach, in which a different model is fitted within each of a finite number of seasons.

For wind-speed data there are no clearly defined seasonal components. Also, as Davison and Smith [3] point out, it is important that the seasonal effects identified are those which affect the upper tails, rather than the central portion of the data. We therefore advocate the separate seasons approach (with a different extreme value model being fitted to large gusts from each season) as the more natural choice. However, as stated in Sec. 2, we hope to be able to exploit homogeneities across seasons in the mechanisms underlying generation of extreme gusts. This may involve application of a uniform procedure for identification of clusters of large observations, and/or the fitting of a constant shape parameter across all seasons. Now the assessment of goodness-of-fit of extreme value models generally entails graphical rather than formal methods, due to the intractability of the latter, and the ease of application and interpretation of the former. In particular, the mean excess plot [mean

residual life (MRL), or conditional mean excess (CME)] is advocated for the limiting Generalized Pareto Distribution (GPD) fitted to threshold excesses (see Lechner, Leigh, and Simiu [5,6] for arguments in justification). In order to check our homogeneity assumptions we must be able to assess the adequacy of the model to all the seasons *simultaneously*. For this purpose we suggest the generation of a prewhitened series for the preliminary stages of the analysis *only*, namely the choice of an appropriate seasonally varying threshold, an accompanying method of identifying clusters of observations above this threshold, and the initial assessment of model adequacies.

In this paper, we take our seasonal unit to be 1 month. Experience suggests that by dividing the year into 12 equal-length seasons, we strike a good balance between the two conflicting requirements of a) reflecting reasonably accurately the continuous nature of seasonal changes in climate, and b) retaining a substantial amount of data for analysis within each season. The models we will consider thus consist of a separate GPD fitted to cluster peak excesses within each month, the threshold also being allowed to vary on a monthly basis. We will assume a homogeneous clustering mechanism throughout the year, but retain the option of allowing the shape parameter k to vary from month to month, or constraining it to take a single value across all months. (In other situations where a different length of season is considered appropriate, the arguments laid out below would apply unchanged.)

Under such a separate months model, an appropriate set of prewhitening operations would be provided by separate transformations t_m for each month m (applied to all the observations in month m). In order to know the precise transformations required, we would need to know the parameters in the GPDs fitted to cluster peak excesses within each month. Since we have not yet established how to obtain the cluster peak excesses (CPEs), we cannot know these values. However it is possible to make an educated guess at an appropriate set of monthly transformations, as shown in the following sections.

3.1.1 Homogeneous Shape Parameter k We consider first the situation in which the GPD shape parameter k is assumed constant over all months. It is then easy to show that a set of *linear* transformations $t_m(x)=a_m x+b_m$; $a>0$, $m=1$, ..., 12 can be chosen to render the distribution of CPEs over a *single* threshold homogeneous GPD across all months (see Ref. [10]).

In order to form estimates for the required transformations, we bear in mind that it is the upper tails of the monthly distributions of all recorded maximum gusts (in our case hourly) which will yield the CPEs. Since the clustering mechanism is assumed homogeneous across all months, we suggest that a good approximation to the appropriate transformations will be obtained by making the upper tails of the empirical monthly distributions of all recorded maximum gusts coincide with each other in some sense. Since the required transformations are linear, this can be achieved by transforming two high quantiles (e.g., 0.95 and 0.99) from each month to two distinct arbitrarily specified points, say the corresponding theoretical quantiles of the unit exponential distribution. Explicitly, we would transform empirical monthly quantiles $z_{1,m}$ and $z_{2,m}$ to the corresponding exponential quantiles q_1 and q_2 by solving the simultaneous equations:

$$a_m z_{1,m}+b_m=q_1$$

$$a_m z_{2,m}+b_m=q_2 \qquad (1)$$

for $a_m>0$ and b_m, and for each $m=1$, ..., 12. The precise choice of quantiles is not critical, and is somewhat arbitrary. It is determined by the necessity of moving as far as possible into the upper tails, while still retaining a substantial amount of data between the two quantiles, and above the largest of them (in order to keep sampling error to a minimum).

3.1.2 Variable Shape Parameter k If the shape parameter k is allowed to vary from month to month, the required monthly transformations are no longer linear. However, the arguments leading to approximately the correct transformations being obtained (by causing the upper tails of the empirical monthly distributions of all monthly gusts mutually to coincide) still hold. This time transformations which will lead to monthly cluster peak exceedances being homogeneous GPD over a single threshold are of the form $t_m(x)=a_m\log(x-c_m)+b_m$; $a_m>0$ (easily obtained by considering the transformation which maps one GPD c.d.f. onto another). Estimates can be obtained by transforming *three* high quantiles (e.g., 0.90, 0.95, and 0.99) from each month to distinct arbitrary points. For example the empirical quantiles $z_{1,m}$, $z_{2,m}$, and $z_{3,m}$ from each month m could be transformed to the corresponding theoretical quantiles q_1, q_2, and q_3 of the unit exponential distribution by (numerically) solving the simultaneous equations:

$$a_m \log(z_{1,m} - c_m) + b_m = q_1$$
$$a_m \log(z_{2,m} - c_m) + b_m = q_2 \qquad (2)$$
$$a_m \log(z_{3,m} - c_m) + b_m = q_3$$

for $a_m > 0$, b_m, and c_m, and for each $m = 1, ..., 12$.

3.2 Implementation for the Sheffield Data

For each month, 10 years of hourly maximum gusts constitute approximately 7300 observations. Hence there are about 365 points lying above the 0.95 quantile; 73 above the 0.99 quantile. The sampling error in estimating these quantiles' theoretical values via the empirical equivalents is therefore reasonably small. We initially make the assumption of a homogeneous shape parameter. As we shall see in Sec. 4, this appears to be well-founded. For each month, then, linear transformations which map the two empirical quantiles to their theoretical unit exponential counterparts (2.996 and 4.605), are applied to all hourly maxima. The resulting prewhitened sequence occupies the range [−4.559, 8.596].

4. Step 2—Threshold Selection

4.1 Methodology

Having created an approximately stationary (in the upper tail at least) sequence of hourly maximum gusts, we are in a position to experiment with various choices of threshold and cluster identification procedure.

We propose a constant threshold for the prewhitened series, based on the assumption that the region of the data to be treated as extreme will constitute the same upper quantile for all seasons. Applying the inverses of the prewhitening transformations to this threshold in monthly segments will then provide the seasonally varying threshold for use in the final model.

Exceedances of the threshold will occur in clusters (storms) from which we wish to choose only the peak excesses for modelling. We need to be able to identify these clusters, bearing in mind that some of the observations within a storm may lie below the threshold. Of several possible methods, we opt for a fixed termination time approach, whereby a storm is deemed to have ended when a certain required number of consecutive observations below the threshold are observed. The advantage of this method over some others is that it allows both the duration of storms, and the duration of intervals between them to vary according to the data, reflecting the inherent natural variability of these quantities.

The threshold and the termination interval may be regarded as the two parameters for estimation in this section of the analysis. Formal estimation procedures such as maximum likelihood are inappropriate here: distributional assumptions on CPEs only hold if the threshold is chosen high enough, and we do not wish to impose a specific model structure on the underlying process which generates storms and periods of calm. However, graphical procedures are highly effective in this capacity. In particular the mean excess plot (the mean residual life plot: see Ref. [4]; or conditional mean excess plot: e.g., Ref. [5,6]) performs well. This is produced by simply plotting the mean excess of all model data above threshold u against u for a range of such thresholds. Linearity in the plot corresponds to a good fit of the GPD to excesses of the model data over any threshold above which the linearity holds. In our case the model data will be the selected cluster peak exceedance magnitudes.

Note that the threshold and the termination interval must be chosen in combination, because these two parameters interact in the manner in which they determine the set of cluster peak exceedances actually selected. Basically, provided both are large enough, the set of corresponding CPEs should be iid GPD, because

1. the GPD exhibits a threshold stability property, whereby a good fit above a certain threshold implies a good fit above all higher thresholds, with merely a change in scale parameter, and
2. if the termination interval is long enough for the CPEs to be approximately independent, then this will still hold for increased intervals.

However, subject to this constraint, we wish to make both quantities as *small* as possible, in order to maximize the number of valid CPEs selected for analysis.

In principle, it would be possible to produce a large number of mean excess plots to examine the model adequacy under a whole variety of combinations of threshold and termination interval. In practice however, this would prove a very cumbersome route to making an appropriate choice. Instead we propose a simple modification to the mean excess plot which leads to considerable streamlining of the selection procedure. For a given termination interval z^*, we propose that the

mean excess above threshold u is plotted against u, *with the identification of cluster peak exceedances being carried out separately for each threshold* u. We will call this device a *reclustered excess plot*. The idea here is that linearity in such a plot above a certain threshold $\tilde{u}$ suggests both a good fit of the GPD to CPEs over $\tilde{u}$ selected using termination interval z^*, *and* a robustness of the mean CPE to the threshold at which *declustering* is carried out. Note that if such a robustness were *not* present, it would cast considerable doubt on the validity of the declustering procedure.

By producing individual reclustered excess plots for a range of values of z^* (each one requires surprisingly little computation time), we should be able to identify the smallest such value for which the independence criterion for the CPEs is met to a good approximation. This will be the smallest value yielding a plot which straightens out above a certain level $\tilde{u}$. This value of $\tilde{u}$ is then chosen as the best threshold for the corresponding value of z^*, giving the optimal pairing $(\tilde{u}, z^*)$.

Note that having selected the pair $(\tilde{u}, z^*)$, it is strongly recommended that a conventional mean excess plot is obtained for the CPEs so obtained, the plotting range being $u \geq \tilde{u}$. This is to verify the validity of the choice, and in particular to check that approximate linearity in the reclustered excess plot is not caused by lack of fit of the GPD and non-robustness to the declustering threshold having opposing effects, and thereby cancelling one-another out.

For a more in-depth discussion of reclustered excess plots and their validity, see Ref. [10].

We suggest that we first work with a prewhitened series obtained under the assumption of a homogeneous shape parameter k, since this will provide a very useful improvement in return level estimation precision if it proves to be justified. Only if the reclustered and mean excess plots suggest a poor fit for all trial values of z^* do we recommend relaxing this assumption and working with a prewhitened series created using non-linear transformations.

Note that the effect of a moderate failure in the assumption of homogeneous *clustering* behaviour is not liable to be serious. While this implies that z^* should be allowed to vary seasonally, the above procedure will tend to lead to the selection of the smallest z^* value large enough to work for *all* seasons: any smaller value of z^* will fail in some parts of the annual cycle, and this should show up as a lack of fit of the overall GPD model to CPEs from the prewhitened series.

4.2 Implementation for the Sheffield Data

Figure 1 shows reclustered excess plots produced for termination intervals $z^* = 0$ (all excesses), 6 h, 15 h, 30 h, 60 h, and 120 h. Here we are using the prewhitened series obtained at the end of Sec. 3, based on the homogeneous k assumption. The plots appear to straighten for $z^* = 30$ h (debatable), 60 h, and 120 h, but not for the smaller termination intervals. Conventional mean excess plots (Fig. 2) produced for $z^* = 15$ h, 30 h, 60 h, and 120 h using the corresponding linearity thresholds $\tilde{u} = 2.8$, 2.6, 2.7, and 3.3 (for $z^* = 15$ h we use the inflection point) broadly support the findings, and we conclude that $z^* = 15$ h is too small; $z^* = 30$ h is borderline; and $z^* = 60$ h or $z^* = 120$ h is large enough.

The fact that the fit of a single GPD to this prewhitened series appears good supports the homogeneity assumption on k, and we do not need to abandon this in favour of a model which allows k to vary.

We select the pairs $(\tilde{u} = 2.6,\ z^* = 30)$ and $(\tilde{u} = 2.7,\ z^* = 60)$ as our choices for the next stage of modelling. We retain *two* combinations because of the doubt over the adequacy of the termination interval $z^* = 30$ h, and in order to check on the robustness of final results to the precise choice of CPEs. The 10 years of hourly maximum gusts yield respectively 525 and 352 CPEs under the two pairings. The thresholds 2.6 and 2.7 lie at the 0.923 and 0.935 quantiles in the empirical distribution of transformed hourly maxima.

5. Step 3—Model Verification

5.1 Likelihood Ratio Tests

From any given choice of threshold and termination interval, and the corresponding monthly sets of cluster peak exceedances, we are able to move directly to a separate seasons model for the raw (untransformed) cluster peak exceedances. Under the appropriate model, the excesses of these in month m over a segmented monthly varying threshold (obtained by applying the inverses of the prewhitening transformations to the threshold $\tilde{u}$ identified in Sec. 4) are independent GPD(σ_m, k_m), with distribution functions

$$G_m(y; \sigma_m, k_m) = (1 - k_m y/\sigma_m)^{1/k_m}; \qquad (3)$$

scale parameters $\sigma_m > 0$; shape parameters k_m arbitrary; and G_m defined on $0 < y < \infty$ if $k_m \leq 0$, and $0 < y < \sigma_m/k_m$ if $k_m > 0$. The case $k_m = 0$ is interpreted as the limit $k_m \to 0$, and is the exponential

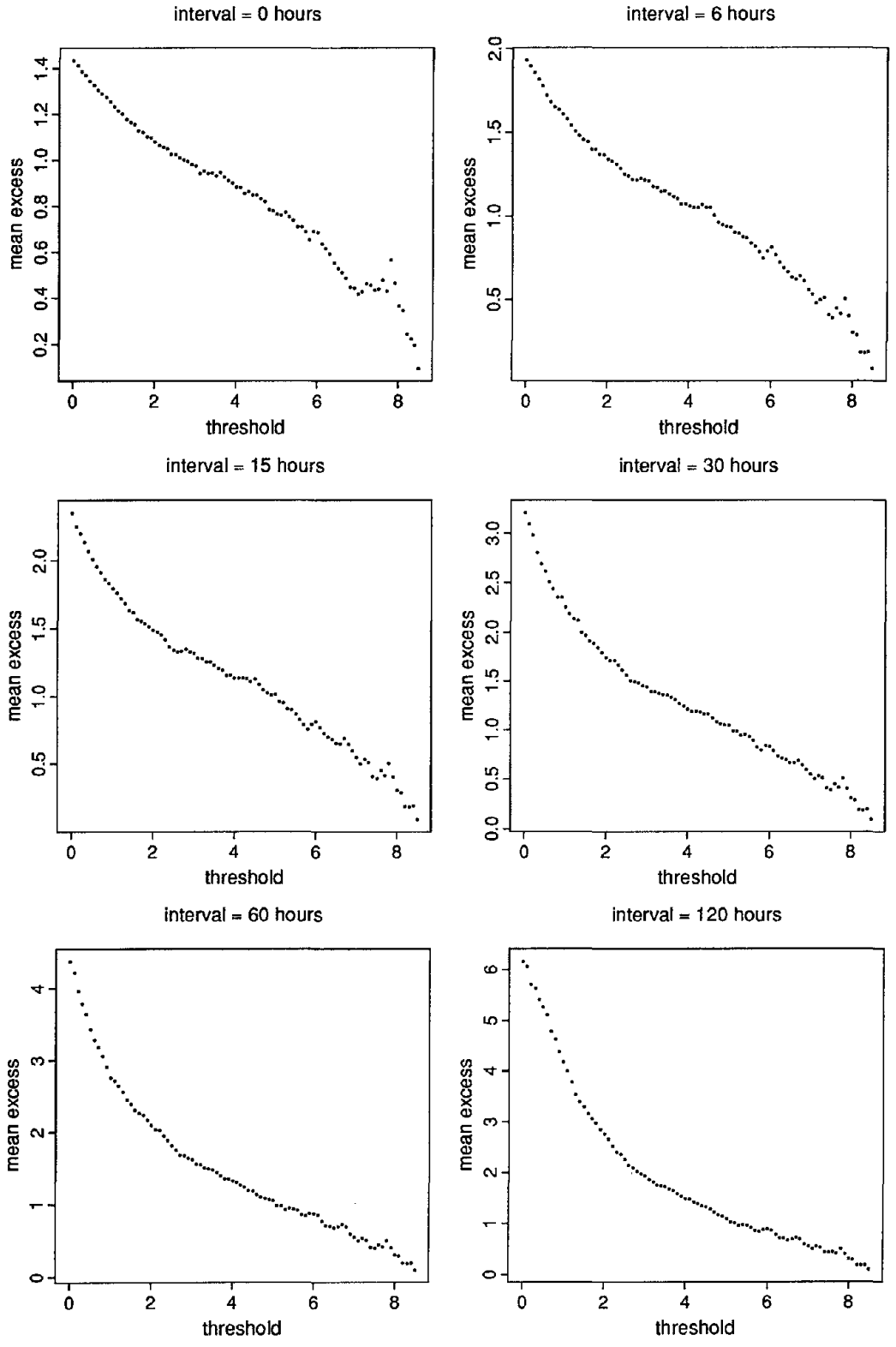


Fig. 1. Reclustered excess plots.

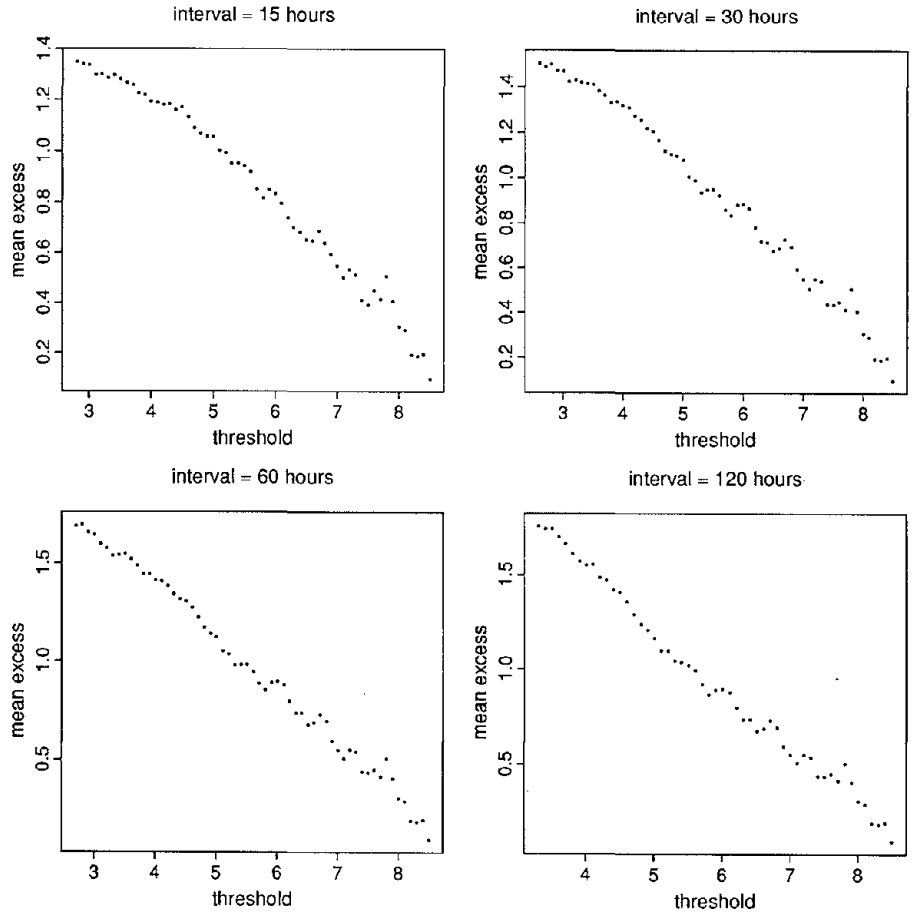


Fig. 2. Mean excess plots.

distribution with mean σ_m. The parameters σ_m and k_m can be estimated via numerical maximum likelihood estimation. (N.B. at this stage of the modelling, the values u_m are treated as fixed constants. Starting values for σ_m and k_m can be provided from the graphical estimates for scale and shape parameters for the prewhitened CPEs obtained using the fact that the fitted line on the mean excess plot should have slope $-k/(1+k)$ and intercept $\sigma/(1+k)$; see Ref. [3]. Applying the inverses of the prewhitening transformations to the GPD(σ,k) will give good preliminary estimates for σ_m and k_m). It is then possible to verify the choice of homogeneous or variable shape parameter k via a likelihood ratio test—twice the decrease in fitted log-likelihood when k is constrained to be homogeneous (over a model in which it can vary from month to month) should be chi-square on 11 degrees of freedom (11 is the change in the number of model parameters) under the null hypothesis of homogeneity. In the surprising event of the test result conflicting with the decision reached in Sec. 4, we recommend the likelihood ratio result as the more reliable, due to its more rigourous justification. In this instance, we would have to be satisfied that the preliminary analysis of Sec. 4 has at least allowed us to get to this stage, while proving to be somewhat misleading!

Notice that once the thresholds and the termination interval have been chosen, a separate seasons model which allows the shape parameter to vary from month to month is in fact equivalent to a model in which each season is treated entirely separately, i.e., no further homogeneities are incorporated. If the extremes occurring in some seasons are not truly large values, then including these seasons in any further analysis will contribute little to return level estimation.

5.2 Graphical Evaluation

The overall fit of the separate months model for the magnitudes of excesses over thresholds can be verified via probability plots or quantile plots (plots of fitted distribution function versus empirical distribution function, or fitted quantile versus empiri-

cal quantile; the plotting points being defined by the cluster peak exceedances). By using the fitted parameter values to transform each monthly set of CPEs to a common margin (say unit exponential), the fit to all seasons can be assessed simultaneously.

5.3 Implementation for the Sheffield Data

Tables 1 and 2 contain thresholds u_m and maximum likelihood estimates for σ_m and k ($=k_m$ for all $m=1, ..., 12$) for the separate months model fitted to the CPEs obtained from $z^*=30$ h and $z^*=60$ h, respectively.

Likelihood ratio tests confirm the validity of the homogeneous k assumption: for the cases $z^*=30$ h and $z^*=60$ h, respectively, 8.23 and 7.56 are compared with a chi-square distribution on 11 degrees of freedom; no evidence that k should vary from month to month.

Table 1. Results when the separate months model is fitted to cluster peak exceedances obtained using $z^*=30$ h

Month (m)	u_m	$\hat{\sigma}_m$	k
1	38.38	16.75 (2.01)	
2	29.68	15.60 (1.99)	
3	34.65	11.37 (1.39)	
4	29.57	10.63 (1.25)	
5	24.85	7.68 (0.79)	
6	25.77	8.75 (0.96)	0.3603 (0.0469)
7	24.26	7.23 (0.79)	
8	23.71	9.22 (1.08)	
9	29.95	12.12 (1.37)	
10	29.52	10.76 (1.26)	
11	34.45	12.34 (1.53)	
12	33.27	16.03 (1.84)	

Table 2. Results when the separate months model is fitted to cluster peak exceedances obtained using $z^*=60$ h

Month (m)	u_m	$\hat{\sigma}_m$	k
1	39.95	23.28 (2.60)	
2	30.94	22.93 (2.60)	
3	35.77	14.09 (1.57)	
4	30.52	13.61 (1.51)	
5	25.65	9.48 (0.95)	
6	26.52	11.52 (1.20)	0.4975 (0.0573)
7	25.07	8.76 (0.90)	
8	24.45	11.97 (1.30)	
9	31.03	16.75 (1.89)	
10	30.69	13.57 (1.44)	
11	35.58	16.32 (1.98)	
12	34.75	19.98 (2.11)	

The overall adequacy of the model in both instances is strongly supported by the probability and quantile plots shown in Fig. 3.

6. Step 4—Return Level Estimation

6.1 Profile Likelihood Confidence Intervals

For given monthly thresholds u_m and GPD parameters σ_m and k_m, $m=1, ..., 12$, the r year return level q_r is obtained as the solution of the equation

$$\sum_{m=1}^{12} \lambda_m[1-k_m(q_r-u_m)/\sigma_m]^{1/k_m}=r^{-1}, \qquad (4)$$

where λ_m is the monthly exceedance rate of threshold u_m. This arises by setting the exceedance rate of level q_r in any given year, given by the LHS in Eq. (4), equal to $1/r$. (Note that if $q_r \leq u_m$ for any m, then the quantity $\lambda_m[1-k_m(q_r-u_m)/\sigma_m]^{1/k_m}$ should be replaced by λ_m; and if for any m $k_m>0$ and $q_r \geq u_m+\sigma_m/k_m$, the replacement should be by zero, because of the range on which the GPD is defined.)

We have not yet considered the monthly exceedance rate parameters λ_m. Assuming a Poisson rate of storm occurrence (following Ref. [7]), the maximum likelihood estimates for these are simply the mean annual numbers of storms occurring in each month. A point estimate for q_r can be obtained by substituting the thresholds u_m, and the parameter estimates for λ_m, σ_m, and k_m into Eq. (4), and solving numerically. Standard errors can be estimated via techniques such as the delta-method, but the construction of symmetrical confidence intervals within a specified number of standard errors either side of the mean is not recommended. Instead, we strongly suggest the use of profile-likelihood. Rather than use the limiting quadratic form of the likelihood surface, profile-likelihood makes use of its actual shape for the data in question. The severe asymmetry of the surface often encountered when it is calculated for return levels suggests that conventional symmetrical confidence intervals are highly misleading.

The details involved in the calculation of profile-likelihood confidence intervals for return levels are not entirely straightforward, and we describe them here. For each of a range of possible values of the r year return level q_r, we maximize the log-likelihood with respect to the model parameters subject to the constraint Eq. (4), which ensures that q_r is in fact the desired return level. Technically this can be

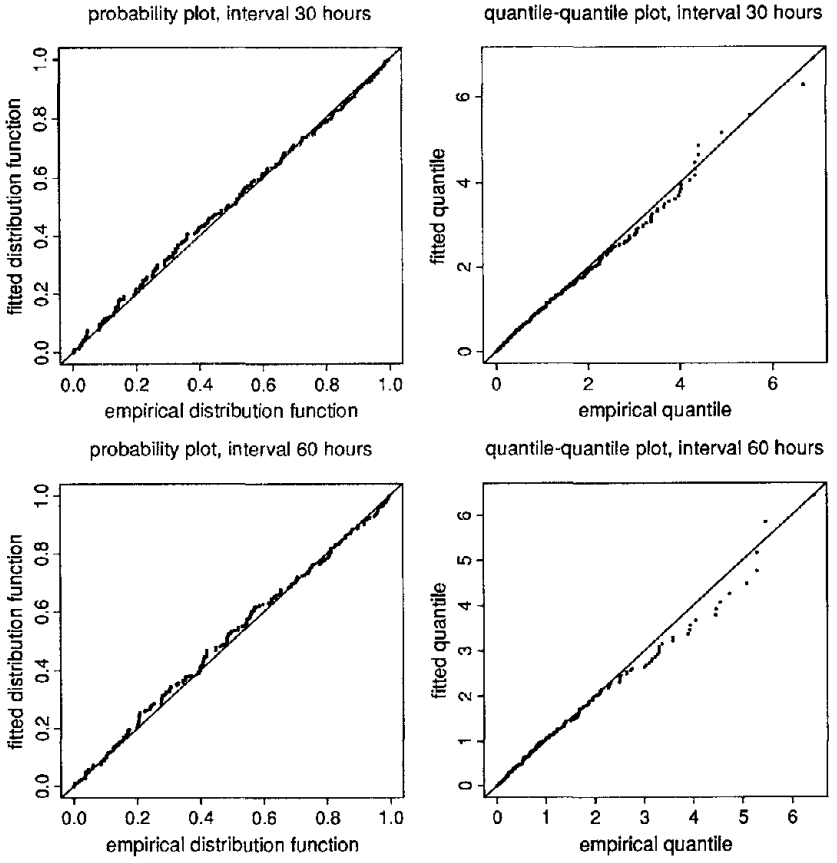


Fig. 3. Probability and quantile-quantile plots.

achieved by making one of the parameters the subject of Eq. (4). Suppose, without loss of generality, that σ_1 is chosen. Then Eq. (4) gives

$$\sigma_1 = k_m(q_r - u_1)/\{1 - [r^{-1} - C)/\lambda_1]^{k_m}\}, \qquad (5)$$

where

$$C = \sum_{m=2}^{12} \lambda_m [1 - k_m(q_r - u_m)/\sigma_m]. \qquad (6)$$

The return level q_r is fixed at the desired level, and the log-likelihood $L \equiv L(q_r)$ maximized with respect to the parameters λ_m, k_m, and σ_2, ..., σ_{12}. At each iteration in the maximization, σ_1 is calculated numerically from Eq. (5), and L is obtained as follows: suppose the CPEs occur over a period of l years, and the number of CPEs in month m in year j is n_{mj}. Let $n_m = \sum_{j=1}^{l} n_{mj}$, and denote the CPEs y_{mi}; $i = 1, ..., n_m$. Then

$$L = \sum_{m=1}^{12} \left[-n_m \log \sigma_m + \left(\frac{1}{k_m} - 1\right) \sum_{i=1}^{n_m} \log\left(1 - \frac{k_m y_{mi}}{\sigma_m}\right) \right]$$

$$- l \sum_{m=1}^{12} \lambda_m + \sum_{m=1}^{12} n_m \log \lambda_m - \sum_{m=1}^{12} \sum_{j=1}^{l} \log(n_{mj}!). \qquad (7)$$

A confidence interval for q_r can then be formed via inversion of a likelihood ratio test, i.e. as the set of values q_0 for which $2[L(\hat{q}_r) - L(q_0)]$ is not significant when compared with a chi-square distribution on one d.f., where $\hat{q}_r$ is the m.l.e. for q_r.

6.2 Implementation for the Sheffield Data

Tables 3 and 4 give point estimates and 95% profile-likelihood confidence intervals for the 10, 50, and 1000 year return levels at the Sheffield site, using the CPE sets obtained via $z^* = 30$ h and $z^* = 60$ h, respectively.

Figure 4 shows the profile-likelihood for q_{50} obtained using $z^* = 30$ h, illustrating the gross asymmetry in the surface. The vertical line is plotted through $\hat{q}_{50} = 82.4$ knots. The horizontal line lies at a level $0.5 \times \chi_1^2(0.95)$ below the maximized log-likelihood, the intersections with the surface thus providing the bounds for the 95% confidence interval.

Table 3. Point estimates and 95% profile-likelihood confidence intervals for some return levels: $z^* = 30$ h

Return period and m.l.e. for return level		
10 ($\hat{q}_{10} = 76.4$)	50 ($\hat{q}_{50} = 82.4$)	1000 ($\hat{q}_{1000} = 88.8$)
95% Profile-likelihood confidence interval		
(72.0, 84.9)	(77.0, 93.9)	(81.8, 103.8)

Table 4. Point estimates and 95% profile-likelihood confidence intervals for some return levels: $z^* = 60$ h

Return period and m.l.e. for return level		
10 ($\hat{q}_{10} = 77.3$)	50 ($\hat{q}_{50} = 82.5$)	1000 ($\hat{q}_{1000} = 85.8$)
95% Profile-likelihood confidence interval		
(73.3, 86.7)	(78.4, 93.1)	(82.3, 97.8)

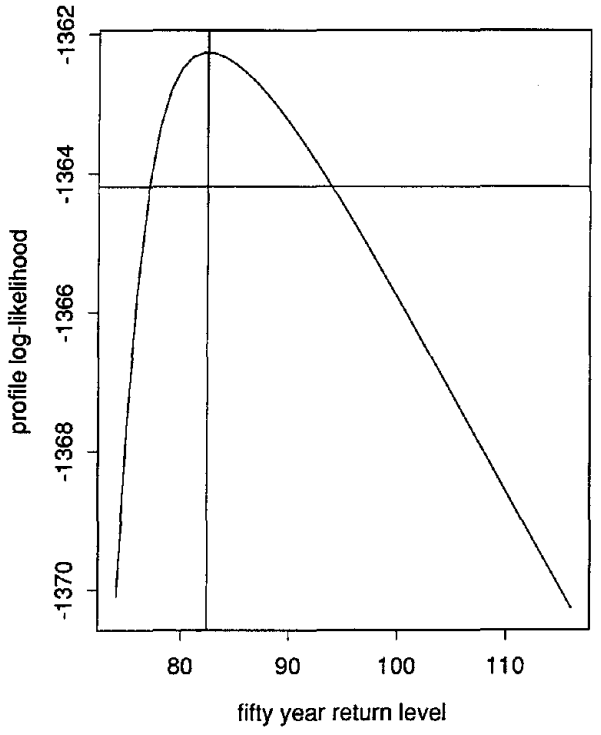


Fig. 4. Profile likelihood for 50 years return level.

7. Discussion

The analysis of the previous four sections seems to be very satisfactory for the data collected at Sheffield. Theoretically motivated models appear to be vindicated by the good fit demonstrated by the plots, and homogeneity arguments pertaining to the wind process in different seasons are supported. The consistency of inferences drawn from the two sets of CPEs obtained using $z^* = 30$ h and $z^* = 60$ h suggests a robustness of results to the informal methods employed in the selection of thresholds and in cluster identification procedures. Finally, while the entire recommended procedure may appear quite complex, once the appropriate software has been set up it can be implemented very quickly and easily, even on a small machine such as a Sun SPARC station.

Despite the success of the algorithm described, which it is expected will be repeated at other sites, it is *very* important to bear in mind a number of cautionary comments. In particular, we must remember that we have relied very heavily on the assumption that there is essentially a single meteorological mechanism which is responsible for the generation of all extreme gusts. It is clear that this is *violated* in climates where several distinct types of storm can generate extreme winds (e.g., both normal temperate zone storms and hurricanes can occur and generate very high velocity winds). At sites at which such climates prevail, considerations different to those presented in this paper apply. For example, it may be that we know that hurricanes *can* occur at a site, but the short run of data available does not include any hurricanes. This highlights a basic limitation in any extreme value analysis—*if we cannot assume that all the physical mechanisms which can generate extremes have been observed in our data, we cannot produce realistic estimates for return levels*. The best we could do under such circumstances is attempt to import knowledge on the unobserved mechanisms from other sites. Any such analysis would, of course, be extremely vulnerable to inter-site differences in behaviour, which could only be assessed theoretically.

In the more favourable situation where instances of all the relevant types of system have been observed, it seems clear that *separate models* should be fitted to the extremes generated by each one. The overall exceedance rate of any particular high level could then be expressed as a sum of components corresponding to each system type, and return levels estimated numerically in a manner similar to that employed in Sec. 6.

Two further aspects of the models considered here are worth brief discussion:

7.1 Piecewise Seasonality

The discontinuous (piecewise) nature of the manner in which all seasonally varying parameters are modelled clearly does not match the *continuous* change inherent in natural processes. However, experimentation with model modifications which allow the parameters to vary continuously [10], suggests that inferences are barely altered in relation to a separate months model for extreme wind gusts. The significant increase in computation time incurred by fitting continuously varying parameters is therefore not thought to be worthwhile.

7.2 Weibull-Type Tails

More interestingly, we note that the shape parameter k fitted to the Sheffield data is very definitely *positive*. A likelihood ratio test overwhelmingly rejects a null hypothesis which constrains k to be zero, in favour of an alternative which allows it to be greater than zero.

Positive k values correspond to a Weibull-type upper tail (with a finite upper endpoint) for the distribution of extremes. Traditional analyses, on the other hand, have been based on the assumption of a Gumbel-type upper tail for extreme wind speeds (with no upper endpoint), following from the notion that there is no natural upper bound to wind velocity anywhere near the orders of magnitude at which wind-speeds are actually observed. However, the findings of this paper concur with those of many other authors. Lechner, Leigh, and Simiu [5], for example, find that a Weibull distribution performs significantly better than a Gumbel distribution for the majority of a sample of 100 stations studied in the United States. These authors point out that convergence to the Gumbel distribution can be *extremely* slow, and that the Weibull distribution, as a penultimate asymptotic approximation, can then often provide a better fit even for sample sizes as large as one billion. In view of this consideration, we contend that the arguments supporting the use of the Gumbel distribution are something of a red herring as far as any practical applications are concerned, and that if the data supports the case for Weibull-type upper tails, then a positive shape parameter should duly be fitted!

7.3 Conclusions

The analysis of the Sheffield data presented in this paper has stood up to a fairly rigourous scrutiny. Further, the assumption of a single meteorological mechanism underlying the generation of extreme gusts is believed to be well-founded in the U.K., and we suggest that even the estimates of 1000 year return levels produced from the 10 years of data can be quoted with some confidence (provided that we remember that the quotation of a 1000 year return level does not incorporate any forecast of a homogeneous climate over the next 1000 years!). It is worth elaborating here on the precise manner in which the *extreme value paradigm* [1] has been applied to our problem of return level estimation. Theoretical (asymptotic) arguments suggest that the GPD should provide a good *approximation* to cluster peak excesses over thresholds, provided the thresholds are *large enough*. Since the approximation *does* appear to be good for all thresholds above a level close to the upper 93rd percentile of the data, we feel justified in assuming that the asymptotic arguments are applicable at these levels. By their very nature, they are then applicable at all higher levels. This enables us to *extrapolate* beyond the upper endpoint of our sample, and hence estimate return levels for periods far longer than those for which data have been recorded. There is obviously a limit to the extent to which this extrapolation is viable, but hopefully this should be self-apparent: provided the method of calculating confidence intervals is not based on unfounded assumptions about the shape of the likelihood surface, any attempt to extrapolate *too* far will simply lead to confidence intervals which are too wide to be of use.

However, this last point leads to a very important cautionary note. Most of the analyses on which current design-level specifications are based make the assumption of Gumbel-type upper tails. The effect of this has almost certainly been to *over-estimate* return levels at most sites. Thus structures have often been designed to be *stronger* than is actually necessary, and the precision of return level estimates has not been of crucial importance. In converting to the more appropriate Weibull-type tails, it becomes *essential* to make adequate allowance for the margins of error associated with return level estimation. To rely, for example, on a procedure such as the delta-method, which does not capture the inherent asymmetry in these error margins, could prove disastrous!

Acknowledgments

The data analysed in this paper were supplied by Peter Smithson at Sheffield University and the U.K. Meteorological Office. I am grateful to both.

8. References

[1] C. W. Anderson, Discussion of paper by S. G. Coles and J. A. Tawn, Appl. Statist. **43**, 1–48 (1994).
[2] S. G. Coles and J. A. Tawn, Statistical methods for multivariate extremes: an application to structural design, Appl. Statist. **43**, 31–33 (1994).
[3] A. C. Davison and R. L. Smith, Models for exceedances over high thresholds, J. Roy. Statist. Soc. B **52**, 393–442 (1990).
[4] W. J. Hall and J. Wellner, Mean residual life, in Statistics and Related Topics, M. Csorgo, D. A. Dawson, J. N. K. Rao, and A. K. Md. E. Saleh, eds., North-Holland, Amsterdam (1981) pp. 169–184.
[5] J. A. Lechner, S. D. Leigh, and E. Simiu, Recent approaches to extreme value estimation with application to wind speeds, part I: the Pickands method, preprint, Statistical Engineering Division, National Institute of Standards and Technology (1993).
[6] J. A. Lechner, S. D. Leigh, and E. Simiu, Recent approaches to extreme value estimation with application to wind speeds, part II: prediction of extreme winds, preprint, Statistical Engineering Division, National Institute of Standards and Technology (1993).
[7] R. L. Smith, Threshold methods for sample extremes, in Statistical Extremes and Applications, J. Tiago de Oliveira, ed., Reidel, Dordrecht (1984) pp. 621–638.
[8] R. L. Smith, Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone, Statist. Sci. **4**, 367–393 (1989).
[9] J. A. Tawn, An extreme value theory model for dependent observations, J. Hydrol. **101**, 227–250 (1988).
[10] D. Walshaw, Statistical Analysis of Extreme Wind Speeds, Ph.D. thesis, University of Sheffield, U.K., 1991.

About the author: *David Walshaw is a statistician in the Department of Speech at the University of Newcastle upon Tyne.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 413 (1994)]

Application of an Empirical Extreme Value Distribution to Load Models

Volume 99 Number 4 July–August 1994

Jun Kanda

University of Tokyo,
Tokyo 113, Japan

An empirical extreme value distribution with lower and upper bounds proposed by the author is applied to represent probability distribution models for maximum load intensities of the earthquake ground motion, the wind speed, and the live load in supermarkets. One of the difficulties in the estimation of the parameters is determining the upper bound value. Nevertheless application of the proposed distribution to the annual maximum earthquake ground motion results in considerable improvements over other models. Possible improvements to the annual maximum wind speed model are discussed. The proposed distribution is also a good candidate for the live load extremes.

Key words: annual maximum; equivalent uniformly distributed load; extreme live load; extreme value distribution; Frechet distribution; Gumbel distribution; maximum bedrock velocity; upper bound value; windspeed.

Accepted: March 22, 1994

1. Introduction

Probabilistic load models are utilized for limit state design procedures and safety assessments of structures. Since lifetime maximum loads have to be applied to these analyses, appropriate probability distributions are needed to represent load intensity models. The Gumbel distribution (Type I extreme value) and the Frechet distribution (Type II extreme value) are often used for such purposes.

When the coefficient of variation (cov) is not large, discrepancies of the upper tails may not be very serious. However for load intensities with fairly large cov such as earthquake ground motions, existing extreme value distributions do not provide good fits to statistical data. Then an empirical extreme value distribution with both upper and lower bounds proposed by the author [1] is a good alternative to improve probabilistic load models.

Statistical data were prepared for the annual maximum earthquake ground motion, the annual maximum wind speed, and live load extremes incrowding situations. The significance of the proposed distribution is discussed in terms of lifetime maximum statistics.

2. Proposed Extreme Value Distribution

Three types of extreme value distributions are commonly used for engineering purposes. Cumulative distribution functions are written as follows [2],

$$F_{\mathrm{I}} = \exp[-\exp\{-a(x-b)\}] \qquad -\infty < x < \infty \qquad (1)$$

$$F_{\mathrm{II}}(x) = \exp\left[-\left(\frac{c}{x-\epsilon}\right)^{\gamma}\right] \qquad \epsilon < x < \infty \qquad (2)$$

$$F_{\mathrm{III}}(x) = \exp\left[-\left(\frac{w-x}{w-v}\right)^{\gamma}\right] \qquad -\infty < x < w \qquad (3)$$

where $a, b, c, \epsilon, \gamma, w, v$ are parameters which characterize the form of distribution. Distributions expressed in Eqs. (1), (2), and (3) are the Gumbel, Frechet and Weibull distributions respectively. In Eq. (1), the random variable x could theoretically vary between $-\infty$ and $+\infty$, while in Eq. (2) the lower bound value, ϵ, and in Eq. (3) the upper bound value, w, exist. When natural phenomena are considered, it seems reasonable that the physical quantity has a positive value with an upper bound limit. On consideration the formula of Eqs. (2) and (3), the following empirical extreme value distribution has been proposed [1]

$$F_J(x) = \exp\left[-\left\{\frac{w-x}{u(x-\epsilon)}\right\}^{\gamma}\right], \quad \epsilon < x < \epsilon \qquad (4)$$

where w and ϵ are upper and lower bound values, respectively, and u and γ are scale and shape parameters, respectively. It can be seen that when x approaches the lower bound, ϵ, Eq. (4) approaches Eq. (2) with $c = (w-\epsilon)/u$ and when x approaches the upper bound, w, Eq. (4) approaches Eq. (3) with $v = w - u(w-\epsilon)$.

In order to demonstrate the form of the proposed distribution, the effect of parameter u with $w = 10$ and $\gamma = 1.0$ for Eq. (4) is shown in Fig. 1 on Gumbel probability paper, with x on the ordinate and reduced variate y on the abscissa. In a similar way the effect of parameter γ with $w = 10$ and $u = \exp(4/\gamma)$ for Eq. (4) is demonstrated in Fig. 2.

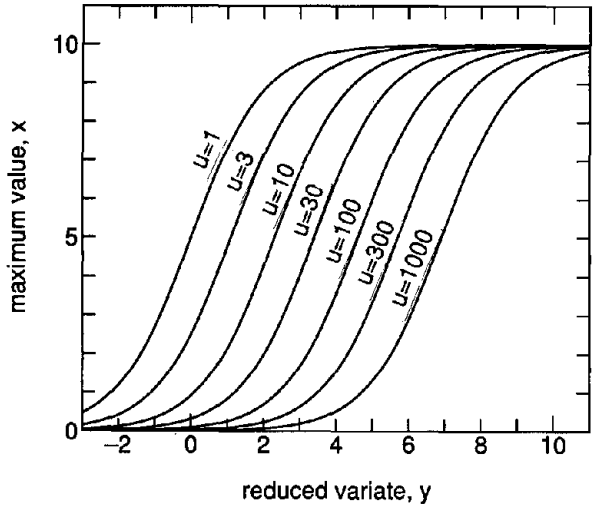


Fig. 1. Effect of parameter u on proposed distribution with $w = 10$, $\gamma = 1.0$. (Gumbel distribution probability paper.)

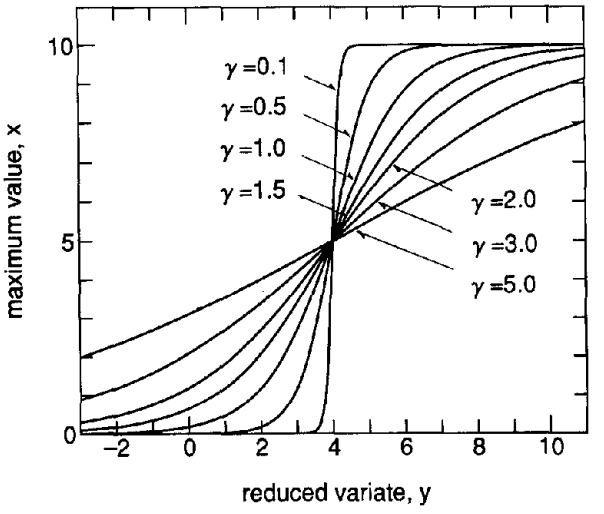


Fig. 2. Effect of parameter γ on proposed distribution with $w = 10$, $u = \exp(4/\gamma)$. (Gumbel distribution probability paper.)

3. Annual Maximum Earthquake Ground Motion Model

The seismic hazard estimation has often been based on earthquake occurrence models assuming a Gutenberg-Richter relationship. An alternative estimation is possible when sufficient number of earthquake records are available to acquire annual maximum earthquake ground motion data at a site. Such an approach is rather common in Japan beginning with Kawasumi's work in 1951 [3]. A recent attempt was made by applying the proposed distribution of Eq. (4) [4]. Some modifications were introduced in this study. Earthquake data for the last 400 years were utilized according to Usami's catalogue [5]. Kanai's attenuation law was chosen as a representative relationship between the bedrock velocity, V, and the magnitude, M, with the hypocentral distance, x, expressed in the following formula [6]

$$V = 10^{0.61M - (1.66 + \frac{3.6}{x}) \log x - (0.631 + \frac{1.83}{x})} \qquad (5)$$

where a focal depth of 30 km is uniformly assumed to calculate x.

The annual maxima of bedrock velocity calculated according to Eq. (5) are plotted on Gumbel probability paper for four sites, i.e., Sendai, Tokyo, Osaka, and Fukuoka in Fig. 3. The 50 largest data from annual maxima in the 400 year period were used, since minor earthquake motions were considered to be missing in historical records or documents and so should be eliminated from the analysis.

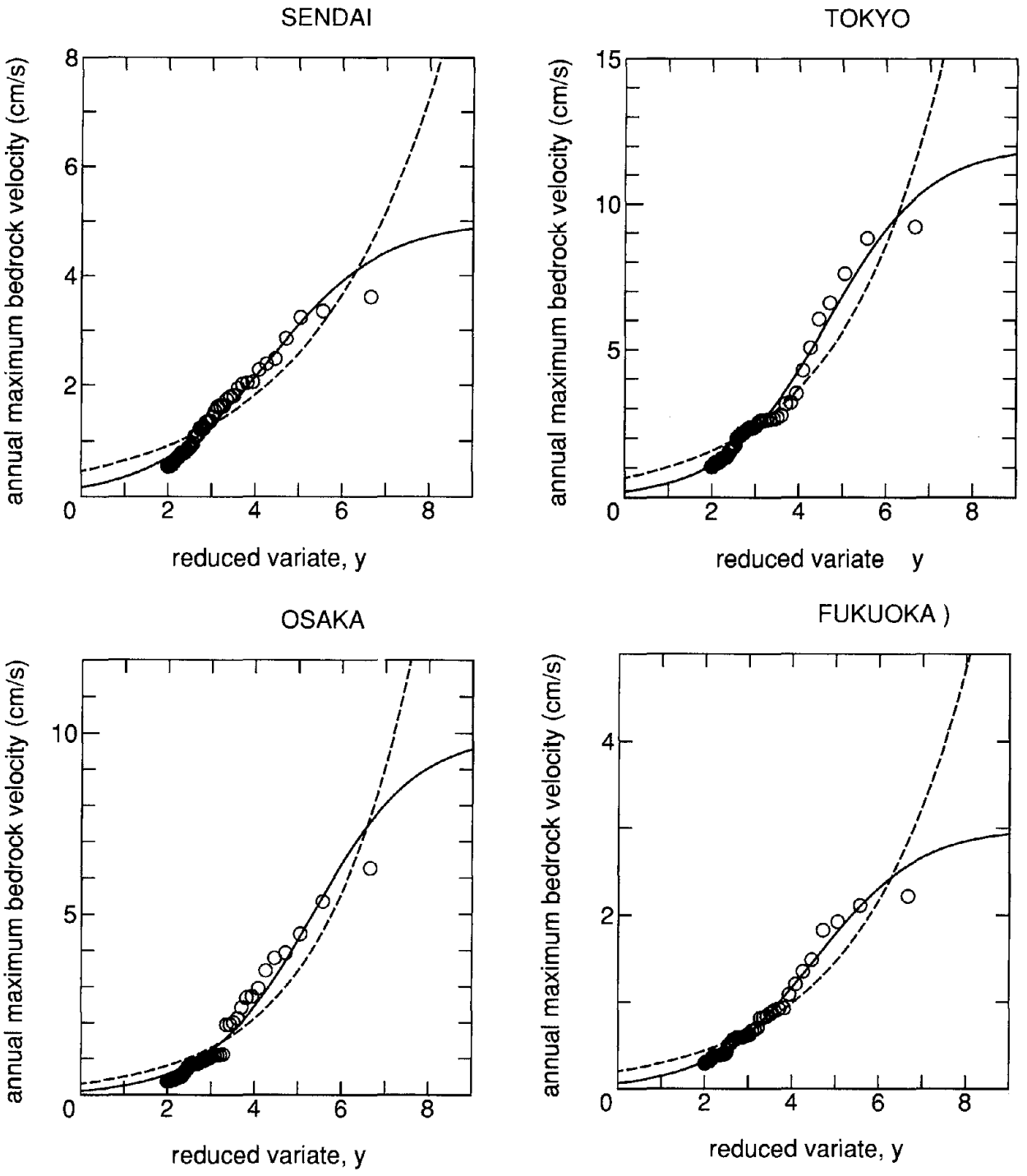


Fig. 3. Extreme value fitting to annual maximum bedrock velocity of earthquake motion in Japan, where the solid and dashed lines indicate the proposed and the Frechet distribution, respectively.

The upper bound value could be assumed from tektonic findings on the fault activity [4], however the upper bound magnitude seems to provide a fairly rough estimate for a particular site as a variation of magnitude by 0.1 causes only 15% change in the estimation of V.

The proposed distribution fitted to plotted data by the least squares method is shown where the value of w in Eq. (4) was chosen by engineering judgment as $w = 5.0$, 12.0, 10.0 and 3.0 for Sendai, Tokyo, Osaka and Fukuoka, respectively. The Frechet distribution fitted by the same method is

also shown in a dashed line for comparison. The existence of saturation tendency indicates that the representation by Eq. (4) is better.

The difference in the results between the proposed distribution and the Frechet distribution can be summarized in Table 1. The error is estimated in terms of the normalized square root of sum of squares error as,

$$E = \sqrt{\frac{1}{n}\Sigma\,(x_i - \hat{x}_i)^2} \Big/ \frac{1}{n}\Sigma\, x_i\,, \qquad (6)$$

where x_i is the ith annual maximum data and $\hat{x}_i$ is the corresponding value estimated from the distribution model, and $n = 50$ for earthquake models.

Although the extimated mean of 50 year maximum values based on the Frechet distribution is similar to that based on the proposed distribution, estimated cov values for the 50 year maxima based on the Frechet distribution, which are the same as those for annual maxima, are considerably greater than those yielded by the proposed distribution. Significant reduction in the error estimate also indicates the appropriateness of the proposed distribution. The cumulative distribution of the 50 year maximum was obtained as the 50th power of the cumulative distribution function of the annual maximum, i.e., it was assumed that the annual maxima are mutually independent. The mean and cov of the 50 year maximum were calculated numerically for the proposed distribution as the closed form relationship between the mean and variance and the parameters in Eq. (4) is not obtainable.

4. Annual Maximum Wind Speed

The Gumbel distribution is often used to represent the annual maximum wind speed distribution. The possibility of improved representation by Eq. (4) is examined for sites where some saturation tendencies are observed, i.e., Aomori, Akita, Nagoya and Kagoshima.

The wind speed data were corrected by taking into account changes of measurement height and the change of the terrain roughness in the period between 1960 and 1970 [7]. Measured data at meteorological agency stations in the period between 1929 and 1991 were utilized. Plotted data and distribution curves fitted to the plots, as was done for the earthquake cases, are shown in Fig. 4, with dashed lines representing the Gumbel distribution. The upper bound $w = 35$ (m/s) was used for Aomori and Akita, while $w = 40$ was used for Nagoya and Kagoshima in Eq. (4). Although the difference between the two types of distributions is not as significant as in the case of earthquakes, error estimates are improved except for Nagoya, where the fitting is rather poor in comparison with other cases as seen in Table 2. The use of a nonzero lower bound value, e.g., $\epsilon = 10$, could improve the fitting for the case of Nagoya. However, this was avoided. Two different major factors, such as the occurrence of typhoons and monsoons, could be the reason for the concave shape of plots on the Gumbel probability paper.

Estimated mean and cov values for the annual maxima and the 50 year maxima are also listed in Table 2. The reduction in estimation of cov of the 50 year maximum for the proposed distribution can be pointed out. When the existence of an upper bound for the extreme value distribution of wind speed is accepted, such a reduction could result in a smaller load factor in the probability-based design.

Table 1. Statistics of maximum earthquake ground motion

Site	Proposed distribution, Eq. (4)								Frechet distribution, Eq. (2)					
					Annual max		50 year max					Annual max		50 year max
	w	u	γ	E	Mean	cov	Mean	cov	c	γ	E	Mean	cov	Mean
Sendai	5.0	28.7	1.30	0.11	0.32	1.56	2.52	0.33	0.455	2.88	0.23	0.63	0.74	2.44
Tokyo	12.0	57.6	1.15	0.125	0.61	1.73	5.46	0.48	0.660	2.34	0.25	1.03	1.30	5.49
Osaka	10.0	90.9	1.18	0.21	0.34	1.99	3.48	0.60	0.303	2.06	0.37	0.52	3.29	3.50
Fukuoka	3.0	45.8	1.19	0.10	0.18	1.58	1.43	0.44	0.199	2.51	0.24	0.30	1.03	1.41

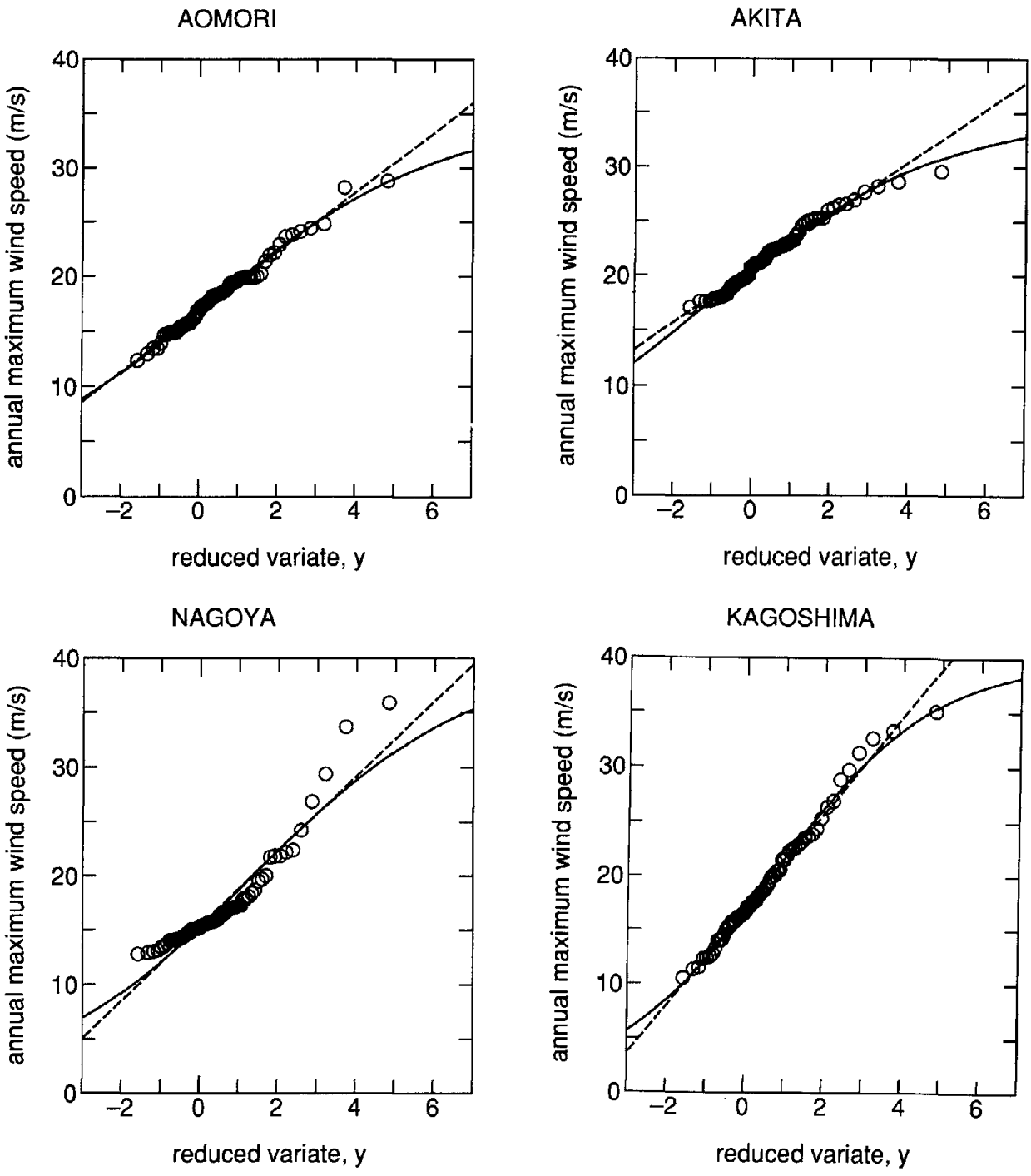


Fig. 4. Extreme value fitting to annual maximum wind speed in Japan, where the solid and dashed lines indicate the proposed and the Gumbel distribution, respectively.

Table 2. Statistics of maximum wind speed

Site	Proposed distribution, Eq. (4)								Gumbel distribution, Eq. (1)						
					Annual max		50 year max					Annual max		50 year max	
	w	u	γ	E	Mean	cov	Mean	cov	a	b	E	Mean	cov	Mean	cov
Amori	35	1.09	3.00	0.02	18.4	0.19	27.9	0.08	0.37	16.8	0.02	18.4	0.19	29.1	0.12
Akita	35	0.70	2.98	0.02	22.0	0.14	30.1	0.05	0.41	20.6	0.03	22.0	0.14	31.7	0.10
Nagoya	40	1.63	2.79	0.08	17.3	0.25	29.8	0.10	0.29	15.4	0.08	17.4	0.25	30.9	0.14
Kagoshima	40	1.39	2.05	0.02	19.4	0.29	34.0	0.08	0.23	16.9	0.03	19.4	0.29	36.8	0.15

5. Simulated Extreme Live Load

Extraordinary live loads for the ultimate limit state design may be estimated based on computer simulations according to a scenario for extraordinary situations. One reported possibility is to model crowd-gathering situations in supermarkets [8]. Room plans with specified shapes and weights of racks with goods and other furnishings were used according to surveyed data, and crowd-gathering situations at one corner of each room were simulated for typical cases such as,

(a) from 0.3 person/m^2 to 5.0 person /m^2

(b) from 1.0 person/m^2 to 10.0 person /m^2

Personnel loads were distributed in the area with no racks or furniture, and 700 N was postulated as the weight of a person.

The equivalent uniformly distributed loads (EUDL) were calculated for slab end bending moments in a shorter span and for girder end bending moments. The detailed procedure is described elsewhere [8].

The plotted data with distribution curves of Eqs. (1) and (4) are shown in Fig. 5 in a similar manner to Figs. 3 and 4. Dashed lines represent the Gumbel distribution of Eq. (1). The upper half of the data were used to obtain parameters of distributions by the least squares method. The upper bound values of 4 kPa and 7 kPa were used for cases (a) and (b), respectively.

Results are summarized in Table 3. Saturation tendency is clear for case (b) where the error is significantly reduced by the proposed distribution. Assumed personnel load intensities in the scenario are somehow arbitrary and a further survey could improve models introduced herein. Nevertheless the usefulness of the proposed extreme value distribution with upper bound can be recognized.

Table 3. Satistics of extreme EUDL for supermarkets

Case	Proposed distribution, Eq. (4)						Gumbel distribution, Eq. (1)		
	w	u	γ	E	Mean	cov	a	b	E
Slab (a)	4000	3.65	1.61	0.04	1674	0.36	0.00182	827	0.07
Slab (b)	7000	13.6	0.80	0.11	2476	0.68	0.00067	193	0.24
Girder (a)	4000	4.22	1.96	0.03	1390	0.36	0.00216	676	0.04
Girder (b)	6000	14.3	0.81	0.10	2075	0.71	0.00076	64.7	0.22

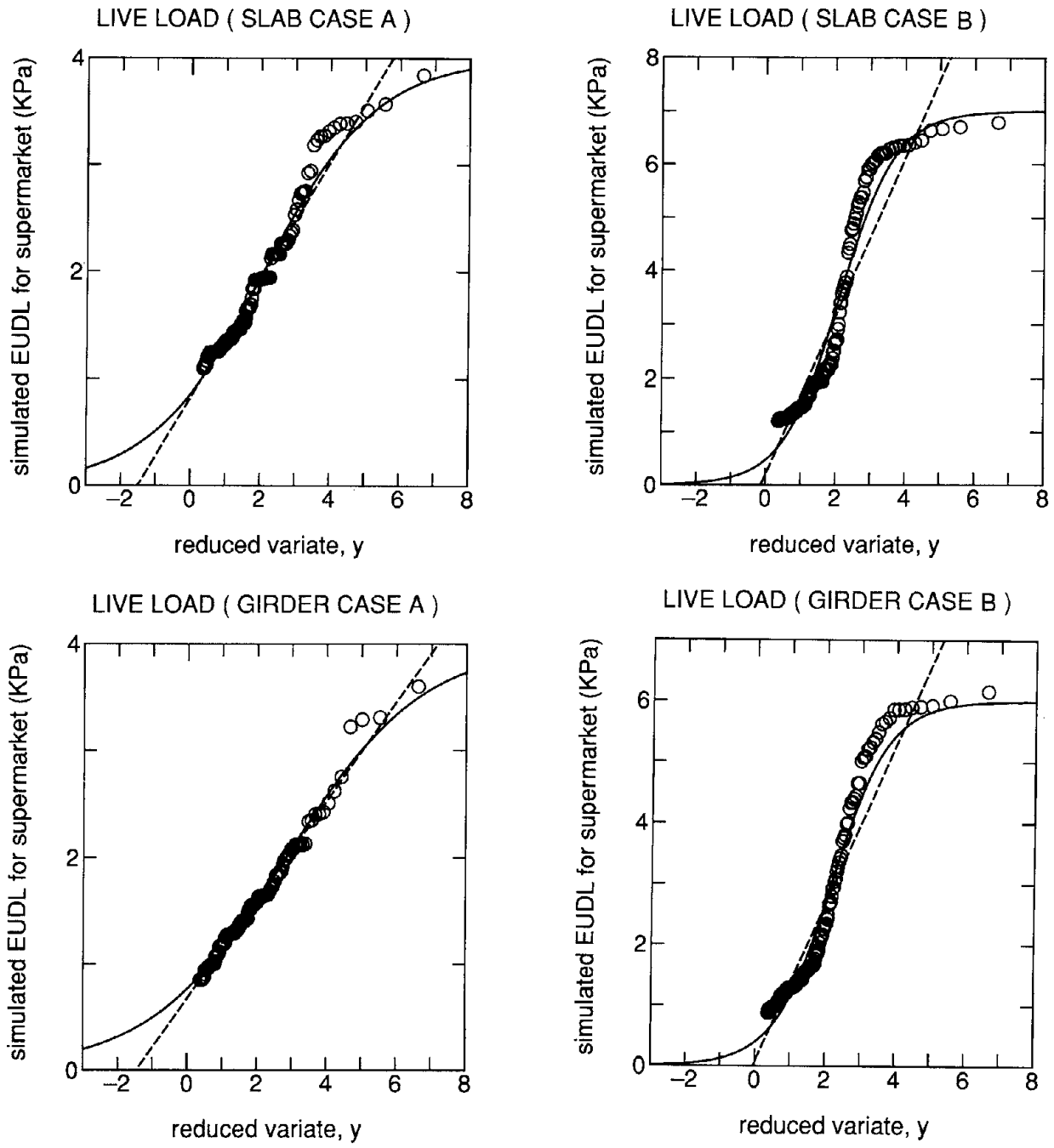


Fig. 5. Extreme value fitting to simulated extraordinary live loads due to crowding concentration in supermarkets, where the solid and dashed lines indicate the proposed and the Gumbel distribution, respectively.

6. Concluding Remarks

An empirical extreme value distribution with both upper and lower bounds was reviewed. The usefulness and improved fit to extreme load intensities available, such as the annual maximum earthquake ground motion, the annual maximum wind speed and the extreme EUDL due to crowding situations in supermarkets, were demonstrated. Use of the simpler commonly used Gumbel or Frechet distributions could cause some significant overestimation in the coefficient of variation for lifetime maximum loads.

Acknowledgments

Appreciation is due to Dr. Kazuo Dan for earthquake data preparation, Mr. Yungbea Chung for wind speed data preparation, and in particular Mr. Kazushige Yamamura for live load data preparation and analyses.

7. References

[1] J. Kanda, A New Extreme Value Distribution with Lower and Upper Limits for Earthquake Motions and Wind Speeds, Theoretical and Applied Mechanics Vol. 31, Tokyo Univ. Press (1981) pp. 351-360.
[2] E. J. Gumbel, Statistics of Extremes, Columbia Univ. Press (1958) pp. 156–164.
[3] H. Kawasumi, Measures of Earthquake Danger and Expectancy of Maximum Intensity throughout Japan as Inferred from the Seismic Activity in Historical Times, Bull. Earthq. Res. Inst., Tokyo, **29**, 469–482 (1951).
[4] J. Kanda and K. Dan, Distribution of Seismic Hazard in Japan based on an Empirical Extreme Value Distribution, Structural Safety **4**, 229–239 (1987).
[5] T. Usami, Revised List of Earthquakes Accompanied by Damages in Japan, Univ. Tokyo Press (1987) (in Japanese).
[6] K. Kanai, K. Hirano, and S. Yoshizawa, Observation of Strong Earthquake Motions in Matsushiro Area, Part 1. (Empirical Formulae of Strong Earthquake Motions), Bull. Earthq. Res. Inst., Tokyo **44**, 1269–1296 (1966).
[7] J. Kanda, T. Saito, and Y. Chung, Return Period Estimation for Maximum Wind Speed caused by Typhoon 9119, Proc. Struct. Engr., A.I.J., **39B**, 1–8 (1993) (in Japanese).
[8] J. Kanda and K. Yamamura, Extraordinary Live Load Model in Retail Premises, Proc. ICOSSAR, San Francisco, (1989) pp. 1799–1806.

About the author: *Jun Kanda is an Associate Professor in the Department of Architecture, Faculty of Engineering, the University of Tokyo, Tokyo 113, Japan.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 421 (1994)]

Seismic Risk Analysis Based on Strain Energy Accumulation in Focal Region

Volume 99 Number 4 July–August 1994

Motoyuki Suzuki and Yoshio Ozaka

Civil Engineering Department,
Tohoku University,
Aoba Aramaki,
Sendai 980, Japan

The object of this paper is to propose a stochastic method for evaluating the magnitude of future earthquakes taking account of nonstationarity in earthquake occurrence. For this purpose, the strain energy accumulation in the focal region was estimated by means of the earthquake data of the past 100 years in Japan. Furthermore, the distributions of maximum ground acceleration were derived by means of the attenuation law. As a result, we found that the distributions of maximum ground acceleration fit the type III extreme value distributions and that the expected values of those distributions depend on the strain energy accumulation significantly. Finally, it is pointed out that the nonstationarity in earthquake occurrence should be taken into consideration in order to evaluate the earthquake load in design.

Key words: earthquake; extreme value distribution; magnitude; maximum ground acceleration; nonstationarity; strain energy accumulation in focal region.

Accepted: March 22, 1994

1. Introduction

Since Japan is located on the subduction zone of a few plates, seismicity is active and many structures have been damaged during large earthquakes. To evaluate the characteristics of earthquake load in design it is important to develop a highly accurate method for estimating the ground motion within the service life of a structure.

Both deterministic and probabilistic methods are available. The former methods estimate the ground motion by means of the dislocation model. Suzuki and Satou [1] have applied this model to a great earthquake expected in the Tokai region. The latter methods evaluate the probability distribution or the expected value of recurrence of the ground motion by considering earthquake occurrence as a probabilistic event. Kawasumi [2] has proposed one such method employing cumulative frequencies. At present, it can be pointed out that probabilistic methods are more suitable for estimating the ground motion than deterministic methods, because a geophysical model and its parameters in earthquake occurrence are not known with certainty [3].

However, in traditional probabilistic methods, it is assumed that the process of earthquake occurrence is temporally stationary, i.e., that the probability of occurrence is invariant in time. Actually, it is rare for another large earthquake to occur in the same region immediately after a large earthquake. Moreover, since the service life of a structure ranges from several decades to about 100 years, it is not reasonable to assume stationarity in earthquake occurrence within the service life of a structure.

The object of this paper is to estimate strain energy accumulation in focal regions at present and to propose a new method for evaluating the ground motion.

2. State of the Art

Some seismic risk analyses in which the non-stationarity in earthquake occurrence is taken into consideration have been suggested. Typical models of seismic risk analyses—the time-predictable model, the slip-predictable model, and the semi-Markov model are reviewed as follows.

2.1 Time-Predictable Model

This model was proposed by Shimazaki et al. [4]. Time history of the stress accumulation and release in a fault is represented schematically in Fig. 1. This is, stress accumulates at a constant rate up to a certain threshold, at which time an earthquake occurs and accumulated stress is released. The size of the earthquake is determined by the level of the released stress. The time when the next earthquake will occur is predictable given the size of the preceding earthquake, but it is difficult to estimate the size of next earthquake. Anagnos et al. [5] described this model by following Markov's renewal process.

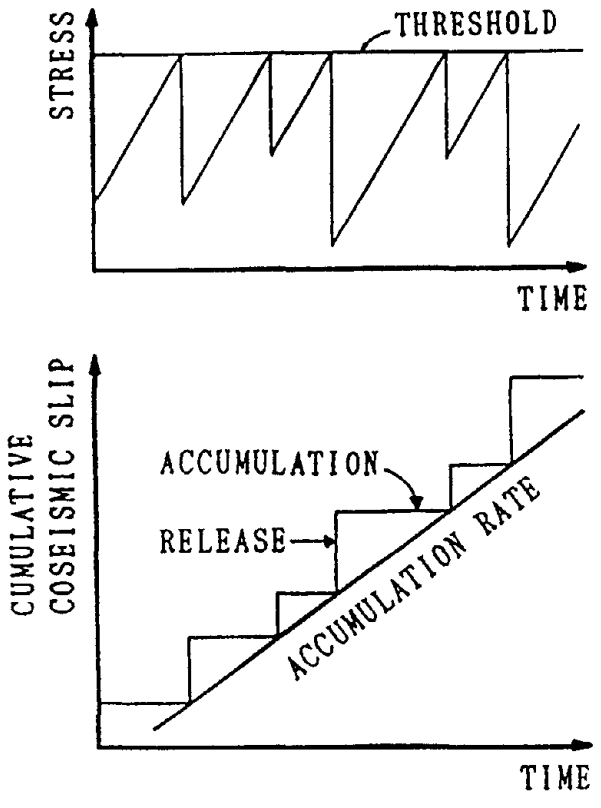


Fig. 1. Time-predictable model.

$$P[Y_{n+1} = J, T_{n+1} - T_n \leq t \mid Y_0, \ldots, Y_n; T_0, \ldots, T_n]$$

$$= P[Y_{n+1} = J, T_{n+1} - T_n \leq t \mid Y_n = I] \quad (1)$$

where

I, J = the state depending on the size of the earthquake

Y_n = the state of the fault after the nth event

T_n = the time of the nth event.

This means that the joint probability from present state to the next state depends only on the present state and is independent of past history. A change of stress release by measuring a coseismic slip in a fault has been proposed, because it is difficult to directly measure the level of stress release.

2.2 Slip-Predictable Model

This model, proposed by Shimazaki et al. [4] predicts the size of an earthquake based on the interval times. Figure 2 shows schematically the stress accumulation and release at a fault. For this model, it is assumed that the stress at the fault drops to zero after each earthquake. The time up to the next event is random, and the longer the interval, the greater the event due to the release of the larger stress. Kiremidjian et al. [6] extend the slip-predictable model to a site hazard model using the attenuation law. However, the occurrences of successive earthquakes are independent according to the above assumption, and the process of foreshock, mainshock, and aftershock at the same fault cannot be rationally explained.

2.3 Semi-Markov Model

This model was proposed by Patwardhan et al. [7]. It is based on the assumption that the size of the earthquake and the interval of time until the next earthquake are influenced by the amount of strain energy released by the previous earthquake. However, a weakness of this model is that subjective assessment is required when classifying the magnitude. That is, the evaluated value is supposed to vary with the classified magnitude because if the magnitude is changed by only 1.0, the released energy varies by about thirty times. Also, the validity

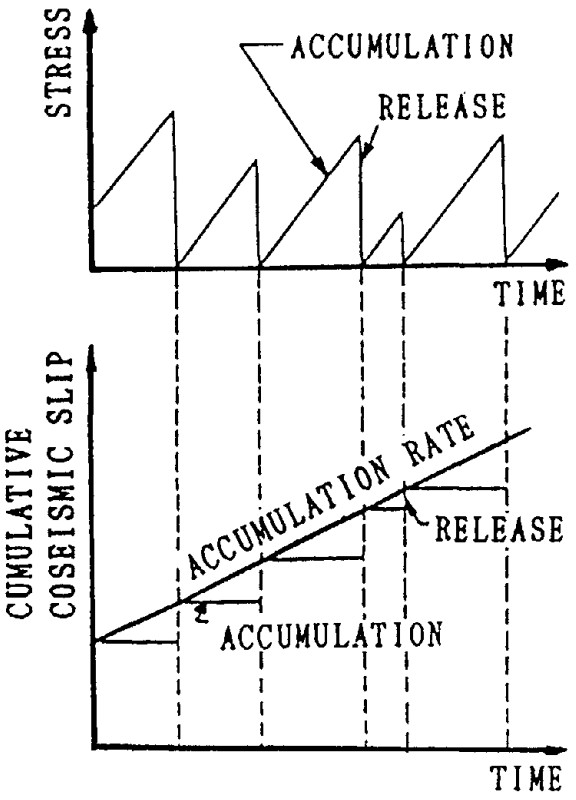


Fig. 2. Slip-predictable model.

of the assumed probability distribution for the time interval is questionable.

As mentioned above, it is necessary to harmonize the stochastic model with the geophysical model of earthquake occurrence for purposes of seismic risk analysis, because the parameters in any model contain some uncertainties. Therefore, we based our research on the theory of plate tectonics [8]. This theory postulates that "the strain energy is accumulated due to the interaction movement of the plates. At the time when the accumulated strain energy reaches a certain extent, an earthquake occurs due to the break of the plates, and the strain energy is released. So, some interval is necessary for the accumulation of strain energy leading to the occurrence of the next event."

In this analysis, it is assumed that the size and the time interval before the occurrence of the next earthquake depend on the strain energy accumulation in the plate at present. A method for forecasting the magnitude of future earthquakes and the distributions of maximum ground acceleration at several main cities in Japan is proposed.

3. Seismic Risk Analysis

In this analysis, the focal region which would influence Japan is restricted to latitudes from 25° N to 50° N and longitudes from 125° E to 150° E. This zone is divided into meshes of 0.5° and relative strain energy accumulation in each mesh at present is estimated. Next, in each mesh, the extremal distributions of magnitude of earthquakes which are expected to occur in the next n years are estimated. Furthermore, the extremal distributions of maximum ground acceleration at main cities are derived by means of the attenuation law.

Presently, it is difficult to estimate the absolute strain energy accumulation, but seismic risk analysis can be performed by estimating the relative strain energy accumulation, according to the following assumptions.

3.1 Earthquake Data

In this analysis, we employ data on earthquakes occurring in or near Japan from 1885 to March 1988 available from the Meteorological Agency [9,10,11,12,13]. However, since the accuracy of methods used in the past to evaluate magnitude is unreliable, the data are corrected by means of the following method [14] proposed by the Ministry of Construction. The method is based on the assumption that "the long-term incline of the curved line of energy accumulation is almost constant and that the incline from 1926 to 1973 shows a value peculiar to Japan." The magnitudes of earthquakes which occurred from 1885 to 1925 are corrected by the following formulas

$$\begin{aligned} &1885\text{–}1895 \;:\; M = M' - 0.5 \\ &1896\text{–}1915 \;:\; M = M' - 0.6 \\ &1916\text{–}1925 \;:\; M = M' - 0.5 \\ &1926\text{–} \qquad : \; M = M' \end{aligned} \tag{2}$$

where M = magnitude before correction

M' = magnitude after correction.

We consider that a deep earthquake (focal depth larger than 100 km) does not have much influence on surface ground motion and that plate thickness is approximately 100 km. Therefore, we limited our investigation to earthquakes with a focal depth of 100 km or less occurring after 1926 when focal depth was added to earthquake data.

3.2 Fault Model

In general, fault movement is not uniform in either time or space. But fault movement must be simplified for the sake of modeling the earthquake occurrence from a technical viewpoint. So in this analysis, it is assumed that a rectangular fault occurs at the time of earthquake occurrence, that its center agrees with the epicenter, that the ratio of its long side to its short side is 2 : 1, and that a section of the fault is at an angle of 45° with the horizontal plane. Moreover, it is assumed that the long side runs parallel to a longitudinal line if the epicenter is located at latitude from 35° N to 41° N, and parallel to a latitudinal line otherwise [15].

Furthermore, in allowance with the concept of the basic fault model by Kanamori [16], it is supposed that strain energy is released uniformly in proportion to some meshed part of the shadow which the rectangular fault casts on the horizontal plane. In practice, however, the areas releasing the strain energy do not always spread around the epicenter, but stretch in only one direction in many cases. Therefore, with regard to data on such faults included in the earthquake fault parameter handbook in Japan [15] and to enable interpretation of the shapes of the faults, it is assumed that the epicenter agrees with the center of the faults. Concerning the relation between section of a fault and magnitude, the proposed equation by Satou [15] is adopted, and the length of the long side of a fault is determined by the following equation:

$$\log L = 0.5M - 1.88, \tag{3}$$

where L = length of the long side of a fault

M = magnitude.

A released amount of strain energy is assumed in allowance with the following equation proposed by Gutenberg and Richter.

$$\log E = 1.5M + 11.8, \tag{4}$$

where E = released amount of strain energy

M = magnitude.

3.3 Cluster Division of Each Mesh

It is assumed that the rate of strain energy accumulation is constant regardless of time. In general, there are areas which are similar with respect to the changing conditions of the plates and release conditions of the strain energy. But it is currently difficult to accurately estimate the accumulation and the release of strain energy. In this analysis, therefore, in order to grasp the relative strain energy accumulation in each mesh, each mesh is classified into the following three clusters based on the distribution of the sum of total released energy in each mesh from 1885 to March 1988. This assumption is based on the thinking that it is more rational to apply ergodicity to the meshes in which the released rate of strain energy is almost equal than to all meshes. The cluster division is determined by considering the relationship between the earthquake magnitude and the amount of earthquake data.

$$\begin{aligned} &1)\ \text{cluster 1} : M_E \leqslant 7.4 \\ &2)\ \text{cluster 2} : 7.4 < M_E \leqslant 7.7 \\ &3)\ \text{cluster 3} : M_E > 7.7 \end{aligned} \tag{5}$$

where M_E = the magnitude into which the annual average released energy in a mesh from 1885 to March 1988 is converted by Eq. (4).

The amount of the annual average released energy for each cluster is averaged, and it is defined as the progress rate of strain energy accumulation. Furthermore, in the case of $M_E < 4.5$, it is regarded as the strain energy released mainly by the inelastic slip and is not dealt with because the released strain energy is small. The result of classifying each mesh is shown in Fig. 3. The meshes not indicated by marks do not belong to any cluster. From this figure, it is recognized that many earthquakes have occurred along the plates.

3.4 Evaluation of the Strain Energy Accumulation in Each Mesh at Present

In order to evaluate the relative strain energy accumulation (E_{ij}) in a mesh (i,j) with latitude i° N and longitude j° E as the center, it is necessary to estimate the strain energy accumulation of the plate at the time of occurrence of the oldest earthquake adopted in this analysis. In general, it is supposed that the recurrence period is peculiar to each focal region, but it is difficult to evaluate them strictly at present. Kanamori [8] reported that the average interval time of a great earthquake with a magnitude on the order of 8.0 is about 100 years on the Pacific side and offing. So in this analysis, it is assumed that all strain energy accumulation is released at least once about every 100 years in each mesh. Based on this assumption, the minimum strain energy accumulation (min E_{ij}) on the

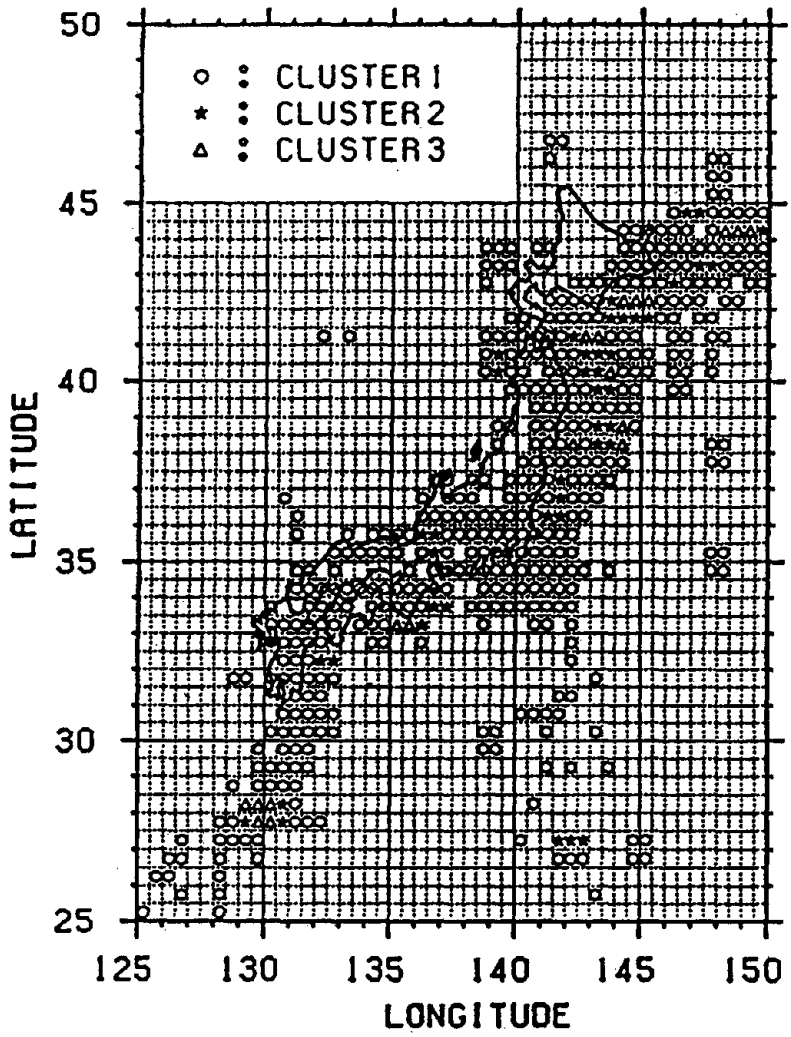


Fig. 3. Cluster division of each mesh.

strain energy-time curve is regarded as being relative strain energy accumulation 0, and the strain energy-time curve is moved in parallel as shown in Fig. 4. The relative strain energy accumulation (E_{ij}) in each mesh at present is estimated by the preceding method. The value of the relative strain energy accumulation in each mesh of each cluster is represented in Figs. 5 to 7. In cluster 3, the relative strain energy accumulation is divided into three classes, i.e., high (more than 300 erg), middle (200–300 erg) and low (less than 200 erg). In cluster 2, the accumulation is divided into high (more than 150 erg), middle (100–150 erg) and low (less than 100 erg) (1 erg = 10^7 joules). As the strain energy accumulations in all meshes of cluster 1 are not high, that cluster is divided into three equal parts.

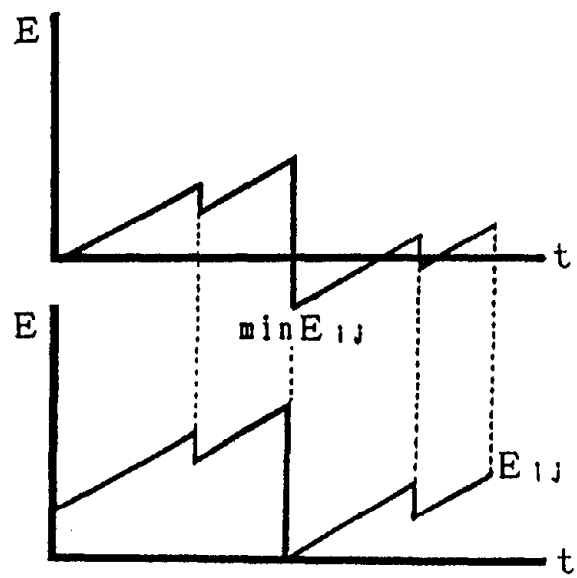


Fig. 4. Parallel movement of the strain energy-time curve.

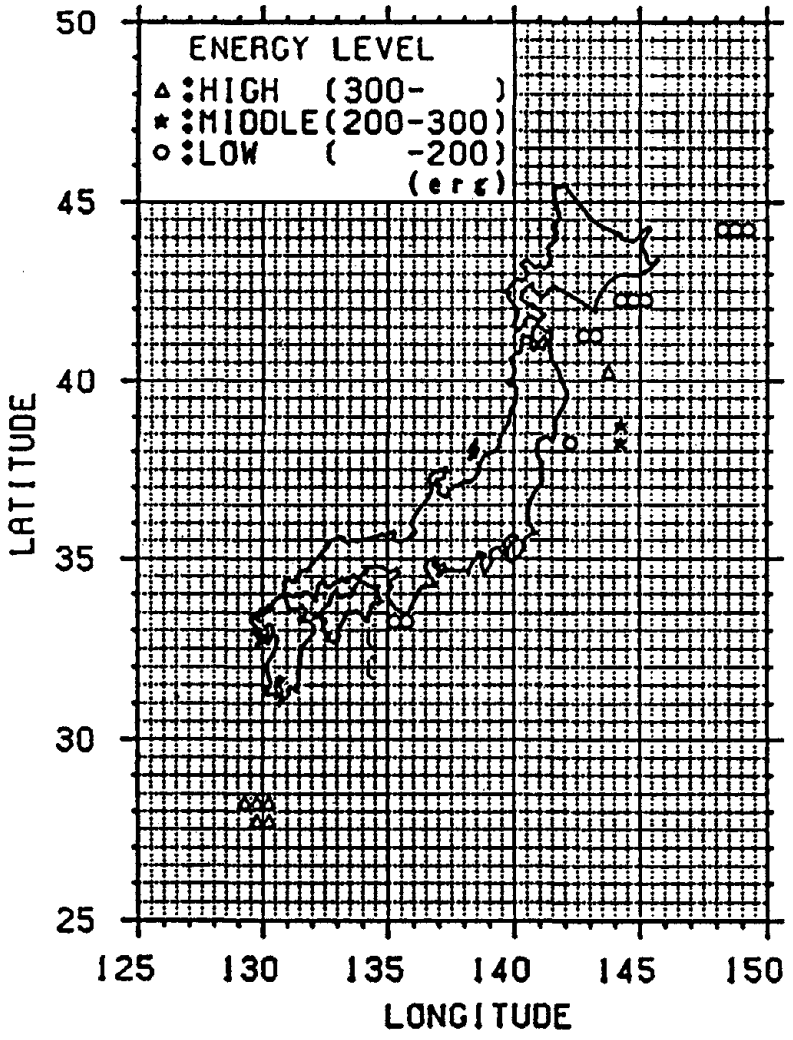


Fig. 5. Strain energy accumulation (Cluster 3).

3.5 Evaluation of Extremal Distribution of Magnitude Considering the Strain Energy Accumulation at Present

In this section, the extremal distributions of magnitude in each mesh are evaluated. The process by which the strain energy is released in allowance with the size of the earthquake, and is again accumulated as time passes, is repeated in each zone. Thus, the strain energy accumulation at present greatly influences the extremal distribution of magnitude of the earthquake expected to occur in the future. If sufficient earthquake data are gathered, it is possible to obtain the extremal distributions of magnitude of each mesh. However, the earthquake data measured by seismographs in Japan are 100 years old at most; the period of observation is not sufficient in light of the recurrence period of great earthquakes. So in this analysis, to evaluate the extremal distributions of magnitude in each mesh, ergodicity is applied to each mesh in the same cluster. Figure 8 shows a flow-chart of the analysis based on this assumption. Figure 9 shows this method schematically. First, the strain energy accumulation of E_{ij} in a mesh (i,j) at present is evaluated, and the strain energy accumulation of $E_{i'j'}$ equal to E_{ij} is determined based on strain energy-time curves in other meshes of the same cluster. Next, this time is defined as $T_{i'j'}$ and the maximum released strain energy ($\max\Delta E_{i'j'}$) for n years from $T_{i'j'}$ is converted into the magnitude by Eq. (4). Some samples from each mesh (i,j) are

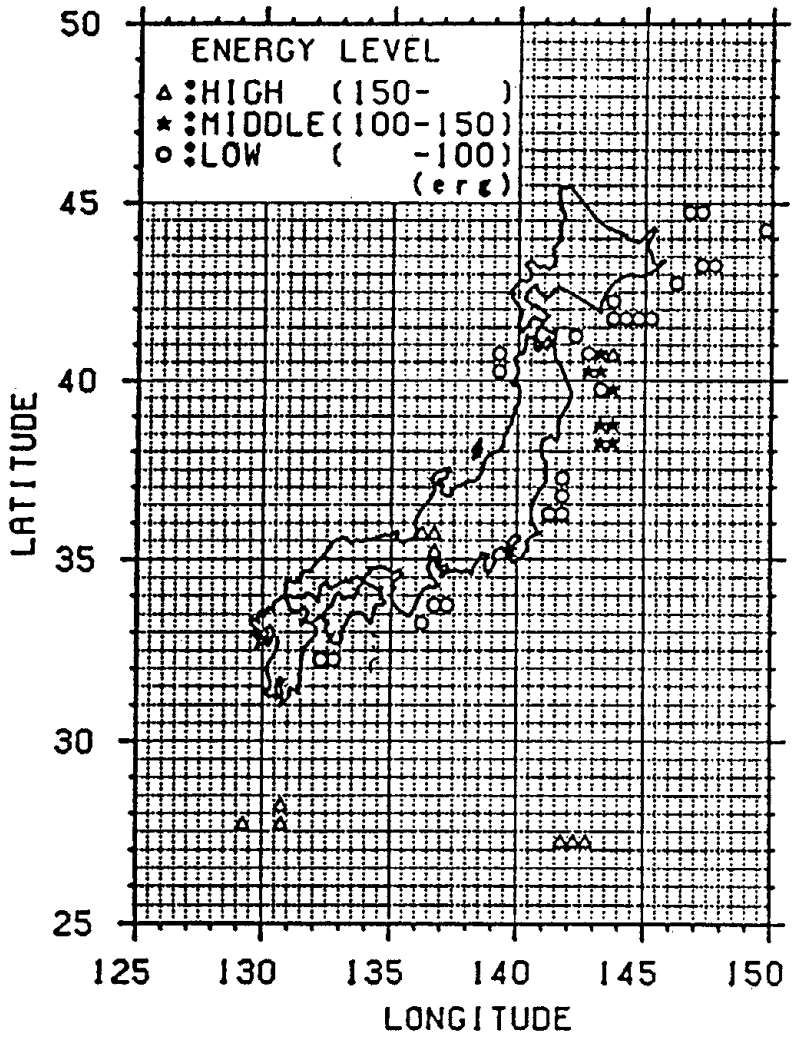


Fig. 6. Strain energy accumulation (Cluster 2).

obtained. To evaluate the form of the distribution, these samples are plotted on the Gumbel probability paper. Assuming that $n = 50$ years, the samples in meshes around Sendai are plotted on the Gumbel probability paper in Figs. 10 to 13. Figure 10 shows the extremal distribution of cluster 3, Figs. 11 and 12 show that of cluster 2 and Fig. 13 shows that of cluster 1. On Gumbel probability paper, the type I extreme value distribution is indicated by a straight line, type II is indicated by a lower convex curve and type III is indicated by an upper convex curve. The upper limit value is decided from maximum sample data rounded off to one decimal and parameters are decided by using the method of least squares.

3.6 Evaluation for Extremal Distribution of Maximum Ground Acceleration

Ten cities in Japan where earthquake observatories are situated are chosen as the points for calculating the maximum ground acceleration. The following attenuation law [17] suited for standard clay is proposed by the Public Works Research Institute of the Ministry of Construction and is adopted in this analysis.

$$Acc_{\max} = 18.4 \times 10^{0.302M} \times \Delta^{-0.8} \qquad (6)$$

where $Acc_{\max}$ = maximum ground acceleration

M = magnitude

Δ = epicentral distance

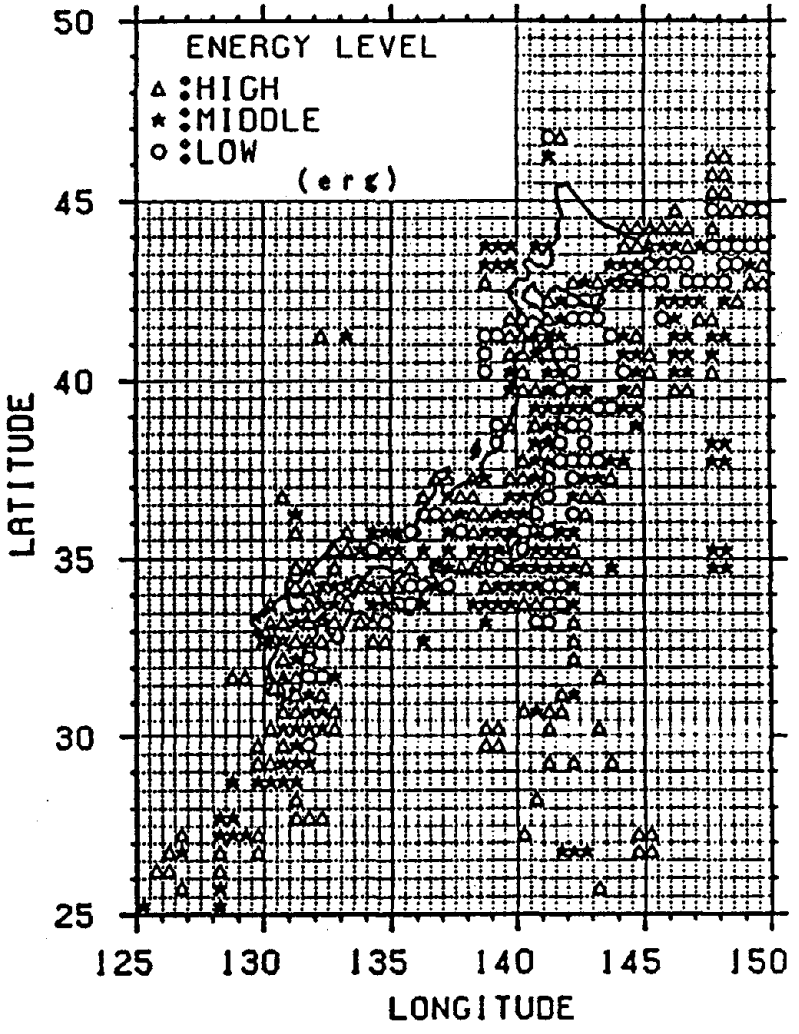


Fig. 7. Strain energy accumulation (Cluster 1).

The extremal distribution of the maximum ground acceleration is estimated as follows.

a) It is assumed that a mesh (i,j) is a hypocenter, and the epicentral distance from the center of the mesh to a city is calculated.

b) Magnitude M_{ij} is obtained from the attenuation law for which the epicentral distance and an acceleration $Acc_{\max}$ are substituted.

$$M_{ij} = g(Acc_{\max}, \Delta_{ij}) \tag{7}$$

c) The value of the distribution function $F_{aij}(a)$ at a maximum ground acceleration $Acc_{\max}$ is evaluated by using the shape parameter, the modal value and the characteristic largest value of the extremal distribution of the magnitude in the mesh (i,j), and M_{ij} is obtained as in b).

d) The previous operation is done for each mesh for maximum ground acceleration. Then using the following equation, the distribution function at city $F_a(a)$ is obtained.

$$F_a(a) = \prod_{i=1}^{n} \prod_{j=1}^{m} F_{aij}(a) \tag{8}$$

e) $F_a(a)$ is obtained by the preceding operation from b) to d) for some $(Acc_{\max})$ accelerations, and the relation between $F_a(a)$ and a is plotted on Gumbel probability paper.

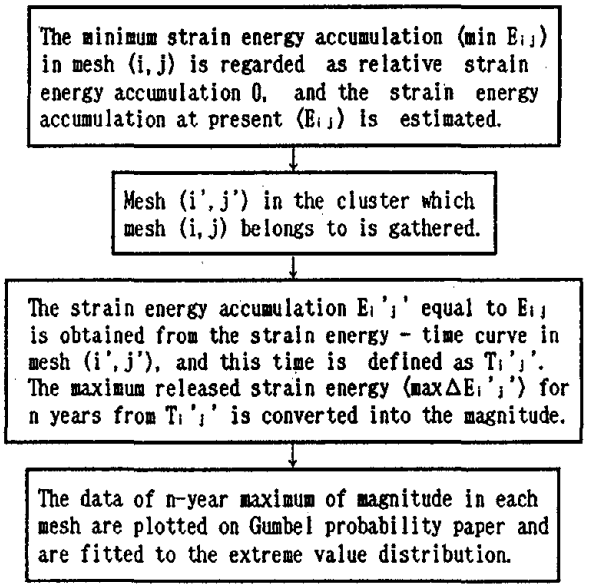


Fig. 8. Flow-chart of the analysis.

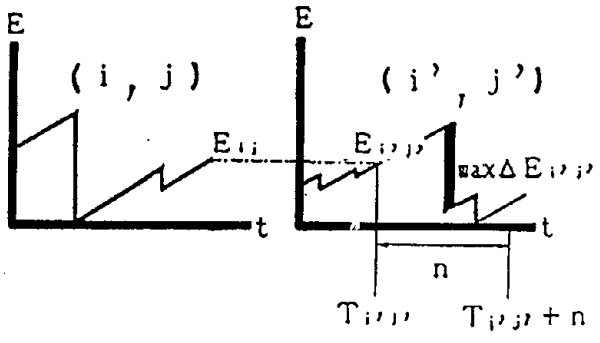


Fig. 9. Schematic representation of the analysis.

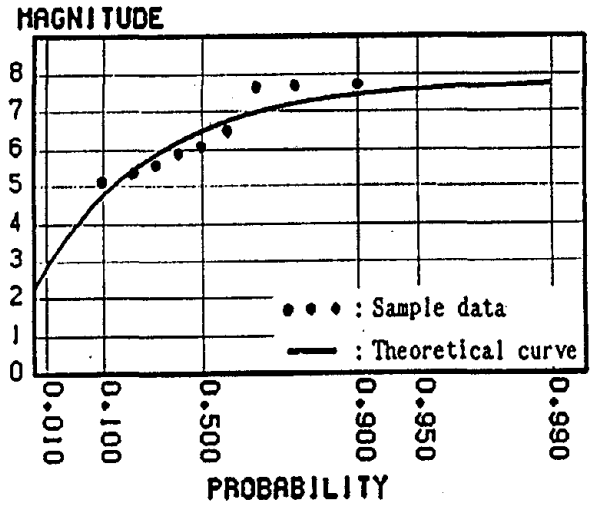


Fig. 10. Distribution of 50 year maximum of magnitude (Cluster 3).

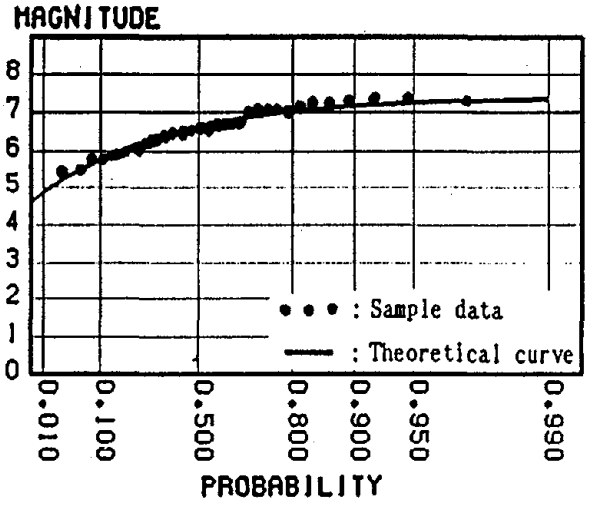


Fig. 11. Distribution of 50 year maximum of magnitude (Cluster 2, Energy level = Middle).

Thus, the extremal distributions of the maximum ground acceleration at main cities are obtained. Figures 14 and 15 show the distribution for 50 year maximum of maximum ground acceleration at Sendai and Tokyo, respectively. Moreover, in order to examine the nonstationarity, the expected values and the coefficients of variation of the distributions for 50, 40, 30, 20, and 10 year maximums of maximum ground acceleration in 1988 are shown in Table 1. Furthermore, the expected values and the coefficients of variation of the distributions for the 50 year maximum of maximum ground acceleration at the different starting points (1968 and 1988) are shown in Table 2.

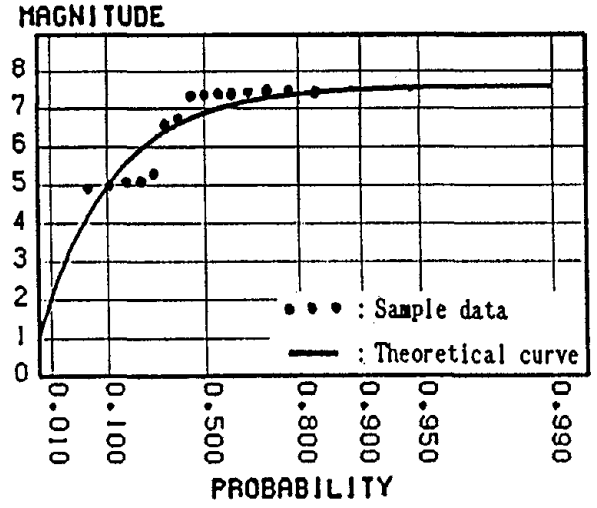


Fig. 12. Distribution of 50 year maximum of magnitude (Cluster 2, Energy level = High).

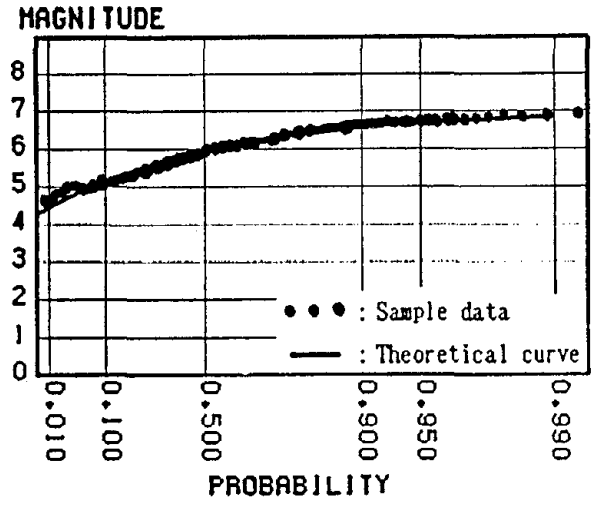


Fig. 13. Distribution of 50 year maximum of magnitude (Cluster 1).

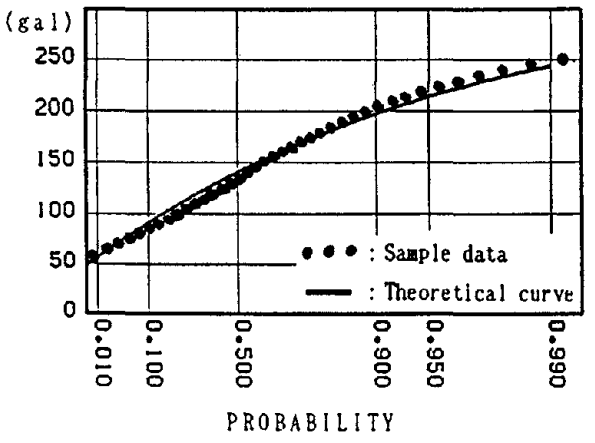


Fig. 14. Distribution of 50 year maximum of maximum ground acceleration at Sendai.

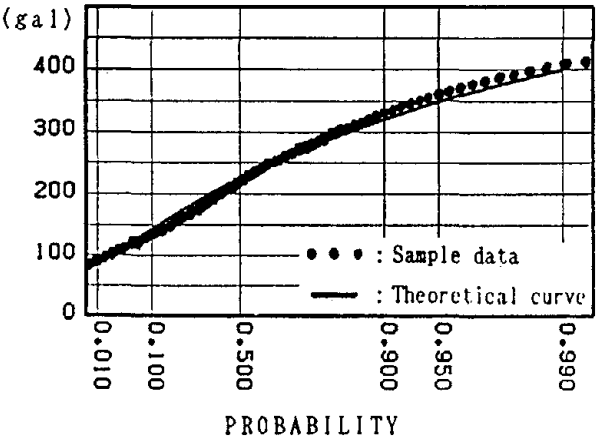


Fig. 15. Distribution of 50 year maximum of maximum ground acceleration at Tokyo.

4. Results and Considerations

4.1 Relative Strain Energy Accumulation

Figures 5 to 7 show the relative strain energy accumulation in each cluster at present. It can be recognized that most meshes in each cluster are distributed near the boundary of the plate of the Pacific side in Kanto, Tohoku, and Hokkaido, and many earthquakes occur in those places. Moreover, it can be assumed that a large earthquake is likely to occur in places in which the strain energy accumulation is high such as in cluster 3 at present.

4.2 Extremal Distribution of the Magnitude in Each Mesh

The distributions for the 50 year maximum of magnitude in the meshes near Sendai are shown in Figs. 10 to 13. Judging from theoretical curve of the type III extreme value distribution, the data accounting for the relative strain energy accumulation at present obviously fit this distribution. The probability of the occurrence of a large earthquake is greater as the strain energy accumulation at present increases. For example, comparing Fig. 11

Table 1. The expected values and the coefficients of variation of the extremal distributions of maximum ground acceleration at main cities (1 gal = 1 cm s^{-2})

Name of City	Statistics	10 year	20 year	30 year	40 year	50 year
Sapporo	Mean value (gal)	41	42	45	46	46
	COV[a](%)	16.2	15.9	15.6	15.4	14.6
Sendai	Mean value (gal)	127	129	139	140	143
	COV (%)	28.7	28.5	28.5	28.3	28.8
Tokyo	Mean value (gal)	189	199	211	224	229
	COV (%)	29.1	29.9	30.0	29.9	30.0
Niigata	Mean value (gal)	52	53	57	58	59
	COV (%)	21.9	22.3	20.7	19.9	19.4
Nagoya	Mean value (gal)	260	260	262	263	263
	COV (%)	5.4	5.4	3.6	3.9	3.9
Kyoto	Mean value (gal)	87	90	91	93	94
	COV (%)	10.6	11.9	12.2	13.1	13.5
Osaka	Mean value (gal)	85	94	94	99	100
	COV (%)	28.9	24.7	27.7	28.3	28.2
Hiroshima	Mean value (gal)	83	96	99	104	108
	COV (%)	23.0	27.2	25.3	26.2	25.2
Takamatsu	Mean value (gal)	107	110	123	124	129
	COV (%)	32.4	30.1	28.4	28.2	27.0
Fukuoka	Mean value (gal)	93	104	117	133	137
	COV (%)	23.4	23.7	22.1	22.4	22.0

[a] COV: Coefficient of Variation.

with Fig. 12, which show the distribution for the 50 year maximum of magnitude in the meshes at cluster 2, the magnitude at a probability exceeding 0.2 is less than 7.5 in Fig. 12 and 7.0 in Fig. 11, respectively, because the strain energy accumulation at present in Fig. 12 is higher than that in Fig. 11.

Therefore, it is thought that the form of the distribution of the 50 year maximum of magnitude remains unchanged, but that the magnitude at the probability of occurrence varies depending on the strain energy accumulation at present.

4.3 Extremal Distribution of Maximum Ground Acceleration at Main Cities

The distributions for the 50 year maximum of maximum ground acceleration at main cities are shown in Figs. 14 and 15. Those distributions fit the type III extreme value distribution as do the distributions for the 50 year maximum of magnitude. The expected values and the coefficients of variation of the distributions for 50, 40, 30, 20, and 10 year maximums of maximum ground acceleration in 1988 are shown in Table 1. Those values reflect the strain energy accumulation at present in the mesh in which the cities are located. Comparing the expected values for $n = 50$ years in Table 1 with the seismic risk map of maximum ground acceleration by Gotou and Kameda [18] in Fig. 16, the expected values yielded by this analysis for Kyoto and Osaka are extremely low. Extensive earthquake data were available for the Kyoto area in which population and culture have been concentrated; the analysis by Gotou and Kameda used historical earthquake data based on estimations from the ancient records.

Table 2. The expected values and the coefficients of variation of the extremal distributions of maximum ground acceleration in 1968 and 1988 (1 gal = 1 cm s^{-2})

Name of City	Statistics	1968	1988
Sapporo	Mean value (gal)	49	46
	COV[a](%)	12.9	14.6
Sendai	Mean value (gal)	151	143
	COV (%)	29.9	28.8
Tokyo	Mean value (gal)	239	229
	COV (%)	30.1	30.0
Niigata	Mean value (gal)	58	59
	COV (%)	18.0	19.4
Nagoya	Mean value (gal)	261	263
	COV (%)	3.8	3.9
Kyoto	Mean value (gal)	96	94
	COV (%)	12.4	13.5
Osaka	Mean value (gal)	139	100
	COV (%)	16.0	28.2
Hiroshima	Mean value (gal)	117	108
	COV (%)	20.3	25.2
Takamatsu	Mean value (gal)	136	129
	COV (%)	21.9	27.0
Fukuoka	Mean value (gal)	144	137
	COV (%)	17.6	22.0

[a] COV: Coefficient of Variation.

4.4 Examination of Nonstationarity in Maximum Ground Acceleration

According to Table 1, the expected values of extremal distribution of maximum ground acceleration at Sapporo and Niigata are almost constant from $n = 10$ years to $n = 50$ years because the seismicities of these cities are not active. However, in other cities, there are large differences in the expected values between $n = 10$ years and $n = 50$ years; in particular difference in Tokyo is 40 gal (1 gal = 1 cm s^{-2}). Moreover, the expected values and the coefficients of variation of the distributions for the 50 year maximum of maximum ground acceleration at the different starting points (1968 and 1988) are shown in Table 2. According to Table 2, the difference of the expected values in Niigata and Nagoya are small, but about 10 gal in Tokyo and Sendai, and 40 gal in Osaka. So, it is recognized that the expected value of maximum ground acceleration varied due to the strain energy accumulation at that time. Therefore, it is necessary to consider the nonstationarity in earthquake occurrence when determining the earthquake load in design.

5. Conclusions

This analysis employs seismic risk analysis in which the focal regions which would have an influence on Japan were restricted. This zone was divided by meshes with 0.5° angles, and relative strain energy accumulation in each mesh was estimated by taking account of the nonstationarity in earthquake occurrence. The distributions for the 50 year maximum of magnitude in each mesh were evaluated. Furthermore, the extremal distributions of maximum ground acceleration at the main cities were derived by means of the attenuation law. From this analysis, the following conclusions can be stated:

(1) A procedure of seismic risk analysis taking account of the relative strain energy accumulation was proposed.

(2) The distributions for the 50 year maximum of magnitude in each mesh fitted the type III extreme value distribution very well.

(3) As the strain energy accumulation at present increases, the value of magnitude at a probability of occurrence becomes greater.

(4) The distributions for the 50 year maximum of maximum ground acceleration at main cities also fitted the type III extreme value distribution.

(5) The expected value of maximum ground acceleration at a city reflected the strain energy accumulation at present in the mesh in which the city is located.

(6) this analysis is capable of forecasting the earthquake load suited to the service life of a structure. That is, it is possible to determine a more rational earthquake load in design by estimating the strain energy accumulation at the time when the structure will be constructed.

(7) This analysis is capable of evaluating the extremal distributions for maximum ground acceleration and those expected values in all parts of Japan, and it seems that these statistics are useful for the criterion of aseismic design.

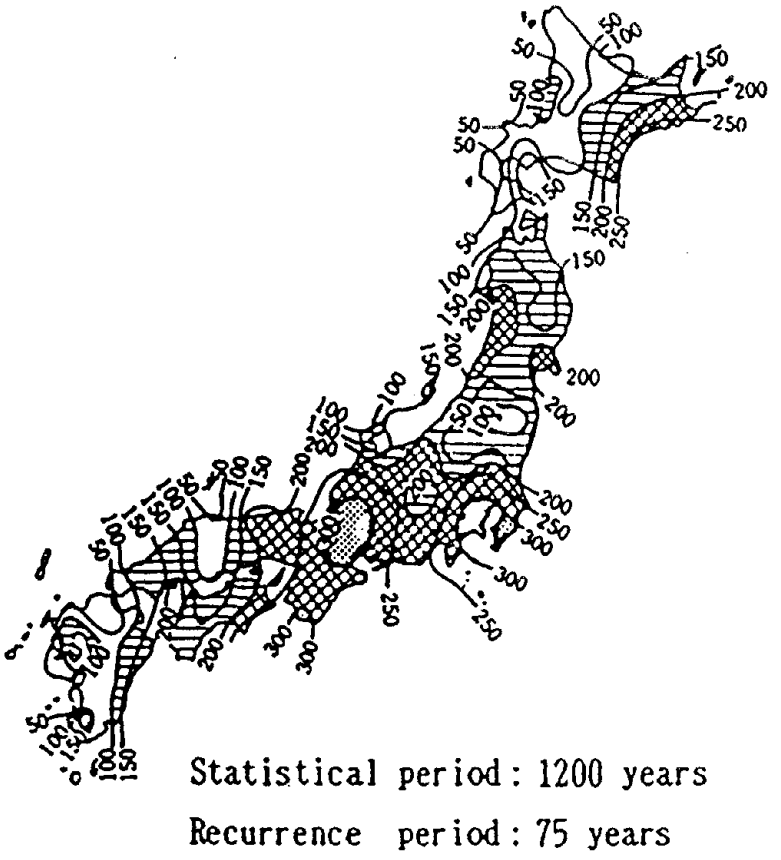


Fig. 16. Seismic risk map by Gotou and Kameda.

As mentioned above, this seismic risk analysis is capable of taking account of the nonstationarity in earthquake occurrence by estimating the strain energy accumulation in each mesh at present. So, with this analysis, it is possible to forecast earthquakes by adopting new earthquake data and to estimate the earthquake load suited to the service life of a structure. However, the data on large earthquakes with recurrence periods of 200 to 300 years are probably insufficient because the earthquake data of the past 100 years in Japan as measured by seismograph are used in this analysis. In this analysis, the seismicity gaps are not treated and the attenuation law is used to cope with standard clay. For obtaining more accurate findings, it is necessary that the attenuation law be suited to each place and condition of clay.

6. References

[1] Y. Suzuki and R. Satou, Evaluation of short-period acceleration and velocity by means of fault model, 8th Symposium on Ground Motion of Architectural Institute of Japan (1980) pp. 37-42 (in Japanese).

[2] H. Kawasumi, Measures of earthquake danger and expectancy of maximum intensity throughout Japan as inferred from the seismic activity in historical times, Bull. Earthq. Res. Inst. **29**, 469-482 (1951).

[3] Y. Matsushima, Seismic response of buildings and probability, J. Archit. Building Sci. **97** (1194), 14-17 (1982) (in Japanese).

[4] K. Shimazaki and T. Nakata, Time predictable recurrence model for large earthquakes, Geophys. Res. Lett. **7**, 279-282 (1980).

[5] T. Anagons and A. S. Kiremidjian, Temporal dependence in earthquake occurrence, 8th World Conference of Earthquake Engineering (1984) pp. 255-262.

[6] A. S. Kiremidjian and S. Suzuki, A stochastic model for site ground motions from temporally dependent earthquakes, Bull. Seism. Soc. Am. **77** (4) 1110-1126 (1987).

[7] A. S. Patwardhan, R. B. Kulkarni, and D. Tocher, A semi-Markov model for characterizing recurrence of great earthquakes, Bull. Seism. Soc. Am. **70** (1) 323-347 (1984).

[8] H. Kanamori, Physics of the Earthquake, Iwanami Kouza Geophysics **8**, Iwanami Shoten (1978) (in Japanese).

[9] Central Meteorological Observatory, The Magnitude Catalogue of Major Earthquakes which Occurred in the Vicinity of Japan (1885-1950) (1952) (in Japanese).

[10] Japan Meteorological Agency, Catalogue of Major Earthquakes which Occurred in and near Japan (1926-1956), The Seism. Bull. of the Japan Meteorological Agency, Supplementary Volume No. 1 (1958) (in Japanese).

[11] Japan Meteorological Agency, Catalogue of Major Earthquakes which Occurred in and near Japan (1957-1962), The Seism. Bull. of the Japan Meteorological Agency, Supplementary Volume No. 2 (1966) (in Japanese).
[12] Japan Meteorological Agency, Catalogue of Major Earthquakes which Occurred in and near Japan (1963-1967), The Seism. Bull. of the Japan Meteorological Agency, Supplementary Volume No. 3 (1968) (in Japanese).
[13] Japan Meteorological Agency, The Seism. Bull. of the Japan Meteorological Agency, January 1952-March 1988 (in Japanese).
[14] Building Research Institute, Ministry of Construction, Earthquake Danger in the Whole Vicinity of Japan Earthquake Danger Map Based on Seismicity and Ground Characteristics, Report of the Building Research Institute, No. 81 (1977) pp. 12-19 (in Japanese).
[15] R. Satou, The Earthquake Fault Parameter Handbook in Japan, Kajima Publishing Company (1989) (in Japanese).
[16] H. Kanamori, A New View of the Earth (by the Physical Society of Japan) (1974) pp. 261-282 (in Japanese).
[17] K. Toki, Earthquake Resistant Analysis of Structures, New System of Civil Engineering, Vol. 11, Gihoudou Co., Ltd. (1981) pp. 33-34 (in Japanese).
[18] Architectural Institute of Japan, Seismic Loading-State of the Art and Future Developments, Maruzen Co., Ltd. (1987) pp. 85-87 (in Japanese).

About the authors: *Motoyuki Suzuki was an associate professor of civil engineering at Tohoku University when this work was done. He is now with the Earthquake Engineering Division of the Ministry of Construction. Yoshio Ozaka is a professor of civil engineering at Tohoku University.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 435 (1994)]

Techniques Used to Determine Extreme Wave Heights from the NESS Data Set

Volume 99 Number 4 July–August 1994

Marc A. Maes

Department of Civil Engineering,
The University of Calgary,
Calgary, Alberta T2N 1N4

and

George Z. Gu

Dallas E&P Engineering (Mobil Research and Development),
Dallas, TX

The objective of the present paper is to explore the application of different extreme value procedures to a selected grid point in the NESS (North European Storm Study) data set. For the purpose of the present benchmark study, each of the participants was invited to submit estimates of a significant wave height with a high return period at a given grid point in the North Sea, together with a detailed description and a rationale of the approach used in deriving extreme wave heights. Submissions from five industry groups were received, compared, and benchmarked against one another. This analysis leads to a number of conclusions regarding the actual use and misuse of extreme value techniques in the fields of engineering and oceanography.

Key words: extreme value analysis; extreme wave heights; hindcast; wave statistics.

Accepted: March 22, 1994

1. Introduction

The NESS project [1] was funded by eleven participants and conducted by a consortium of five research institutes in Europe. The NESS users group built up considerable experience in the field of extremal value (EV) applications. (For a list of acronyms see Appendix A.) Valuable input was received from recognized experts in EV theory. It should therefore be made clear that the objective of the present comparison is not to review/criticize the use of particular extreme value techniques. Rather, the objective is (1) to reflect upon the diversity of the modelling assumptions and the procedures used to determine extreme wave heights, (2) to report on how the different groups set out to deal with difficult issues such as data reduction, statistical and parameter uncertainty, hindcast model uncertainty, and the consideration of measured data, and (3) to seek constructive guidance in this area from the extreme value specialists present at the NIST/Temple University EV Conference.

2. The NESS Data Set

The wave model used in the North European Storm Study (NESS) is an adaptation of the model HYPAS (Hybrid Parametrical Shallow Water wave model by Günther and Rosenthal [2]). The model results used here are from the "fine" grid model, which has a resolution of 30 km and output available every 3 h. Data are available for the periods: 1) 25 continuous 6 month (October–March)

winter periods for the winters of 64/65 through 88/89; 2) three continuous 6 month (April–September) summer periods for the summers of 77 through 79; 3) continuous data for the month of April 85; and 4) 40 discrete storm periods in the April–September summer periods between 1968 and 1988. The significant wave height, *Hs*, from the model is based on the spectral definition of *Hs*, i.e., four times the square root of the spectral variance. It is assumed that *Hs* is representative of a 3 h average sea state in a 30 km by 30 km square around the grid point, and that all storms, which would have any effect on annual extreme values of *Hs*, are included in the data set.

3. Benchmark

The NESS grid point used in the present studies, was selected to be a location in the Forties area of the North Sea with latitude 57.777° N and longitude 0.952° E and a water depth of about 100 m. The NESS participants were invited to provide, as a minimum requirement, their "best" estimate of a 100 year return period significant wave height, *Hs-100*, for this grid point together with a short writeup describing how and why a particular EV procedure was used. Five industry groups submitted contributions; for the purpose of this paper, it was agreed not to identify the contributors; they will be referred to as groups A, B, C, D, and E. All the contributors exceeded the basic requirement of providing a 100 year return wave. Particular emphasis was given to the question of how to account for the uncertainty associated with the hindcast model itself. The contributors' supererogation should not come as a surprise—many of the analysis procedures are coloured by subjective choices and assumptions: it is very much up to individuals to decide what looks good, what techniques are appropriate, how they should be used, and which numbers will finally be acceptable.

The emergence of the NESS database in the North Sea is not the cause of the divergence of extreme value analysis methods. To date extremal analyses have been based on available measured wave data sets—each with their own degree of accuracy and length of record. Only one attempt has been made to use the measured data in a consistent manner in estimation of extremes. The results of that pioneering effort are reported in the U.K. OTH 89 300 Supporting Document [3]. Because those results were intended to provide "indicative" values of extreme environmental criteria, implicit interpretations were made, for example in extreme value extrapolations, to reduce the risk that the results might be underestimates. It was also accepted at the time that a case for other values could be made. For reference, the results at Forties in that document provide an *Hs-100* of 14.3 m.

In the following five Sections, the five benchmark study contributions are summarized. Acronyms are used to denote the several cdfs used by the contributors; to avoid confusion caused by unclear terminology, the distributions corresponding with each acronym are listed in Appendix A. Whenever "storm peaks" are used in a subsequent analysis, the contributors resort to the same peak identification procedure: peaks are identified by determining the maximum wave *Hs* within a moving 18 h window—the average duration of a storm event.

4. Contribution A

Two basic techniques are used. The first one (A1) consists of fitting all 3 hourly data to a (three parameter) Weibull (W3), a Gumbel (G), or an FT3 distribution (the extreme value distribution with an upper bound) using either MOM or LS. The selection between the two distributions is made on the basis of individual judgment or a goodness-of-fit (GOF) criterion.

The second technique (A2) is a peak over threshold (POT) analysis of all peak storm event values exceeding a given level. The threshold data are fitted to either an exponential cdf (EXP) or a two parameter Frechet cdf (F2); MOM or LS are used to estimate the parameters of these conditional distributions; a plotting position $i/n+1$ is used in the case of LS, but it is not clear in which direction errors were considered. Selection is based on best visual fit or GOF. Threshold upcrossings are assumed to be Poisson distributed, with λ estimated as the average number of storms (with peak wave exceeding the threshold) per year. The NESS results for the selected gridpoint are given in Table 1, together with results of the same analysis performed on a set of proprietory measured wave data from the same area, at Forties.

Table 1. Group A results; 100 year return period *Hs* (m)

A1	NESS (3 hourly data)	Measurements ("spot" data)
W3, LS	11.4[a]	13.0*
G, LS	13.2	13.6
FT3, LS	10.3[a]	12.3
W3, MOM	11.9	12.8*
G, MOM	16.3	14.0
A2	50 storm peaks with $Hs > 7.6$ m	52 storm peaks with $Hs > 7.6$ m
FR2, LS	10.7	14.2
EXP, LS	10.9[a]	14.1[a]
F2, MOM	10.4	13.3
EXP, MOM	10.5	13.5[a]
Average of *Hs* with good quality fit	10.8	13.4
Convert "spot" to "3 hourly"	+0.6	−0.9
Account for 1989–1992 data	+1.2	
Offset NESS/measurements	12.6	12.5
Adjusted estimate		
Final estimate	12.6	

[a] Good quality fit.

The 100 year return values of the distributions selected on the basis of a good quality fit are now averaged, and it appears that there is a substantial difference of some 2.6 m between NESS and measured data (Table 1). Three corrections are applied:

(1) The measured data consist of *Hs* estimates taken at a point ("spot data") over a 20 min sample interval and recorded at hourly or 3 hourly intervals; a "new" data base was created by converting them to 3 h averages similar to NESS. Extreme value analyses on the original set and the converted set were compared and it was found that the "3 h average" data consistently gave lower estimates of 100 year *Hs* extremes in comparison to "spot" data. The variation ranges from 0.5 m to 1.2 m, with an average value of 0.9 m, or about 6% (see Table 1).

(2) The NESS database finishes at the end of March 1989. Some storms in the North Sea since that date have been very severe; indeed the most severe storm measured at Forties occurred in December 1990, when an *Hs* of 11.6 m was recorded. The effect of this missing data in the NESS archive was assessed, albeit indirectly, by examining the effect that the equivalent period has on extrapolations of the Forties measured data. From the various analyses performed for the above periods, the effect on 100 year *Hs* estimates of including data recorded in the period April 1989 to May 1992 ranged from increases of 0.2 m to 0.9 m. The average increase across the analyses was 0.6 m, or about 5% of the shorter period estimate.

(3) Both regression and extreme value analyses have been performed on selected overlaps between NESS and measured data. The regression analyses revealed that the mean NESS wave height was some 10% higher than the measured wave height, but when extrapolated to extremes, the 100 year *Hs* estimates from NESS were between 0.5 m and 2.0 m lower than extrapolations from the measured database, with an average difference of 1.2 m. This apparent offset could be due to a wide range of factors, many (if not all) of which are under investigation by the NESS User Group at the time of writing.

All three corrections are captured in Table 1. The conservative view taken in applying the three corrections is indicative of the safety margin associated with the final estimate $Hs\text{-}100 = 12.6$ m; however, no specific uncertainty band is provided.

5. Contribution B

Group B's procedure for wave criteria determination for Forties is as follows:

(1) Extract storm peak data at the reference gridpoint with a threshold of $Hs = 6.0$ m and an 18 h window (298 storm peaks).

(2) Select the Annual Extreme Value (AEV) for each of the 25 years from the 298 peaks. The 25 values range from 6.9 m to 9.9 m. The reasons for using AEV instead of POT method are described to be the following:

- AEV shows consistently better fits than POT (higher correlation coefficients, smaller mean square errors in the case of LS, and larger likelihood functions in the case of MLE);
- The extremes from AEV are not influenced by the threshold, i.e., they are less subjective;
- Extremes based on AEV method tend to be higher than POT (more conservative);
- In the North Sea, due to the high frequency of storms, the highest *Hs* in a year does represent the wave severity for the year in most cases, whereas for POT, when calculating extremes for various thresholds, it is sometimes found that the storm frequency for the best fit is less than 1.0/year, less than that for AEV.

(3) The 25 Annual Extreme Values are fitted to six distributions: G, BM, FR3, FT3, W3, EXP, using two estimation methods: LLS and MLE. All six LLS fits are very good since the correlation coefficients all exceed 0.98. Only three MLE fits are considered acceptable: this was judged on the basis of the relative magnitudes of the likelihood function. The range of the 100 year *Hs* given by the nine good fit cases (6 LLS and 3 MLE) is from 10.6 m to 11.5 m (Table 2). Since Gumbel is theoretically sound for annual extremes and the Gumbel LLS gives an excellent fit, it was decided that Gumbel LLS would be used throughout the analysis. Plotting positions for LLS are $i/n+1$ and squared errors on *Hs* are minimized.

(4) To take consideration of possible bias in NESS, measured, smoothed storm peaks at the Forties location are plotted against the corresponding peak *Hs* from NESS. Only peaks exceeding 6.0 m are considered. A regression analysis yields the best fit linear function

$$Hs, \text{ measured} = 0.63 + 0.9746\, Hs, \text{ NESS} \quad (1)$$

with a standard deviation of 0.93 m.

(5) The 298 NESS peaks are adjusted on the basis of Eq. (1), and the analysis steps 2 and 3 are repeated; this is shown in the second column of *Hs-100* values in Table 2.

(6) Scatter is now considered by adding random errors to the adjusted 298 storm peaks. Random errors are generated using a Monte Carlo simulation assuming a normal distribution with a standard deviation of 1.0 m (1.0 m is selected as a round-off value of the regression model error of 0.93 discussed above). An AEV LLS Gumbel analysis is performed on the *Hs* data containing random errors. After 10,000 simulations the average *Hs-100* is 12.4 m with a sample standard deviation of 0.66 m.

(7) The proposed 100 year return period *Hs* is taken to be 12.4 m.

Table 2. Group B results; 100 year return period *Hs* (m)

	NESS (3 hourly data)	NESS with Linear adjustment
G, LS	11.0	11.4
BM, LS	10.7	11.2
FT2, LS	11.1	11.4
FT3, LS	11.0	11.3
W3, LS	10.8	10.9
F, MLE	10.6	11.1
FT2, MLE	10.8	11.3
W3, MLE	10.6	10.9
Incl. Random Errors (STD = 1.0 m)		12.4
Final estimate	12.4	

6. Contribution C

Group C established the following EV procedure for NESS data. The method accounts for spatial spreading using neighbouring grid points, but this aspect of the procedure will not be described. To start with, the 2 parameter Weibull cdf (W2) is fitted to the cumulative frequency distribution of all the data. In practice, however, a best fit is sought for the top 10% of these data. MML is used to estimate the parameters. POT is suggested as an alternative method for a finer 10×10 km grid, but not for the 30×30 km grid under consideration.

The NESS extremes are now corrected to take into account hindcast model uncertainty by applying all of the following techniques:

C1: add random Gaussian noise at 5%, 8%, and 10% to the W2 cdf (the KESPL method) and record the increase of *Hs-100*.

C2: obtain short return period quantiles, specifically those having exceedance probabilities equal to $12/k$ and $1/k$, where $k = 365 \times 8/2$ is approximately the number of NESS data per year; then multiply these two values with 1.86 and 1.40, respectively, (the RATIO method). The idea of scaling short return period values to 100 year estimates using factors obtained from measurements, originates from the so-called Jenkinson method used by the UK Met Office for deriving extreme wind speeds.

C3: use a linear equation to transform both W2 parameters to "equivalent measured" parameters (the PARAMETER method). The equation derives from an existing regression between hindcast and measured data.

The final step is to interpret the results obtained and to compare them with all available measured data (Table 3). In the case of Forties, the *Hs-100* based on measurements was found to be 13.29 m; this indicates that a 8% noise level is appropriate under C1, and a 1 year ratio method under C2. The 100 year return values for C2 and C3 are averaged, and this value is then averaged with C1. This is considered to be the best "equivalent measure" *Hs-100*. Finally, this result is averaged once more with the direct NESS estimate and a correction factor of 1.03 is applied to take into account that NESS covers only the 6 month winter period in each year. The results in Table 3 should be used with caution: certain values are valid for the averages of 5 gridpoints (including our reference point) covering the Forties area.

7. Contribution D

Group D's Method is a POT method of storm peak values. The threshold is varied in increments of 0.1 m, until a good visual fit is obtained to the following distributions: Gumbel (G), Exponential (EXP), two-parameter Weibull (W2), Pareto (P), lognormal (LN), and generalized gamma (GG); parameters are estimated using LS or MML, except in the last case where a MOM based on Stacy and Mirham [4] is used. The empirical cdf is taken to be $i/n^* + 1$, where n^* is the number of data exceeding the threshold.

In this particular case, little variation in *Hs-100* was detected when the threshold was varied from 7.0 m to about 7.9 m, and reasonable fits were obtained using G, GG, and LN. The final selection of a threshold of 7.8 m was guided by the principle that a POT analysis should ideally be conducted using (approximately) the top 40 data. The best visual fit on a Gumbel plot is obtained by the GG (Table 4). The estimate of *Hs-100* is 10.7 m, which is rounded to 11.0 m.

Measured storm peak data at the Forties are taken into account by multiplying the NESS estimate by 1.07. The value of this multiplication factor is justified on the basis of the following two considerations:

(a) perform a peak-to-peak scatter plot (measured vs NESS): the best fit regression line forced through the origin has a slope equal to 1.07;

(b) a POT analysis (threshold = 7.0 m) is conducted on the measured data and on the corresponding NESS data (i.e., the NESS data occurring simultaneously with the measured data). The 100 year return period on the
former turns out to be 7% greater than the NESS *Hs-100*; the generalized gamma cdf was also used for this purpose.

Table 4 summarizes the intermediate values and shows Group D's best estimate *Hs-100* = 12.0 m.

Table 3. Group C results; 100 year return period *Hs* (m)

1. Directly from NESS (W2, MLE)	11.3
2. Measured data available	13.3[a]
3. C1 using noise at 8%	11.4[a]
4. C2 using 1:100 ratio	12.4
5. C3	12.4
6. Best "Equivalent Measured": 0.5 C1 + 0.25 C2 + 0.25 C3	12.2[a]
7. Final estimate = 1.03 × average of 1 & 6	12.2[a]

[a] These values are averages of 5 grid points located in Forties.

Table 4. Group D results; 100 year return period *Hs* (m)

POT with threshold of 7.8 m ($n^* = 37$)	
1. Gumbel using MML	9.8
2. Gumbel using Ls	10.4
3. LN using MML	9.7
4. LN using LS	9.9
5. P using MML	13.0
6. GG using MOM	10.7[a]
7. NESS best estimate (round off of 6.)	11.0
8. Correction based on Measured Data: (7.)[a] 1.07	12.0

[a] Indicates best fit.

8. Contribution E

POT storm are fitted to one of three distributions that are left-truncated below the threshold x_0: Gumbel truncated (GT), Frechet truncated (FR2T), and Weibull truncated (WT). In the present application, however, the second cdf failed to give a good fit and was discarded. The likelihood expressions involve three parameters: the two basic parameters, together with x_0. In practice, the MML is applied to determine the two basic parameters, given x_0. The associated 95% confidence bands on the corresponding 100 year return values are determined using the 2×2 observed information matrix, given x_0. The candidate distribution is selected on the basis of (1) the correlation coefficient for LS residuals in the *Hs* direction (usually $> 99\%$), (2) the mean error on cumulative probabilities (generally ≤ 0.05), (3) the mean square error on cumulative probabilities (generally ≤ 0.01), and (4) visual assessment.

To determine the threshold x_0, the above procedure is repeated in order to find a range of thresholds over which both the goodness-of-fit statistics, as well as the extrapolated design value are stationary. The selection process is guided by the condition that the annual storm frequency at the site should be between 1.0 (0.5) and 3.0 (4.0). This frequency is proportional to the inverse of the number of peaks exceeding x_0. The half range of the two parameter confidence interval obtained using the MML procedure for the selected x_0, is now added to the *Hs-100* value. The addition of the half-range confidence interval derives from the concern noted in Contribution A that the 3 h average hindcast data overly smooths storm peaks as compared to the 20 min "spot" measured data. In the comparisons to measured data at three North Sea sites which have been performed to date, use of the confidence interval results in an unbiased extrapolation at both the 40 year and the 100 year return period.

As with the previously discussed procedures, the impact of several severe storms which have occurred after the NESS hindcast period was also considered. Measured data from those storms were used to artificially extend the hindcast database. A repeat of the above procedure using the extended database resulted in an increase in the extrapolated 100 year wave height by 0.7 m. The best estimate of *Hs*, shown in Table 5, including the confidence interval and consideration of post-Ness storms, is 12.0 m.

Table 5. Group E Results; 100 year return period *Hs* (m)

			Hs-100 (m)	
Threshold × 0 (m)	Number of points > ×0	Annual storm frequency	Fit to GT	Fit to WT
Stationary range				
7.00	118	4.7	10.8	10.6
7.15	92	3.7	10.6	10.4
7.25	81	3.2	10.7	10.5
7.35	68	2.7	10.6	10.4
7.45	56	2.2	10.5	10.3
7.55	50	2.0	10.6	10.4
7.65	44	1.8	10.6	10.5
7.75	37	1.5	10.6	10.6
7.85	29	1.2	10.7	10.3
Average			10.6	10.4
Best NESS estimate			10.5	
Add 95% CI as described			11.3	
Add 0.7 m to account for post NESS storms			12.0	
Final design value			12.0	

9. Summary

Five NESS participants were asked to provide their best estimate of the 100 year return significant wave height at a given grid point in the North Sea. We cannot help being pleasantly surprised with the astonishing array of approaches used by the participants: all submissions attest to the fact that the contributors have an expert understanding of the NESS statistics and the extreme value methods needed to formulate engineering design criteria. Our second impression is equally compelling: not withstanding the diversity of selected EV methods and the variety of subsequently applied "adjustments/corrections," it is interesting to observe that the recommended *Hs-100* values lie very close to one another.

Table 6 summarizes the final results. The first row lists the *Hs-100* obtained from a direct EV analysis of the NESS data: all values submitted can essentially be rounded off to the same number: 11.0 m. This value may be contrasted with the aforementioned reference approach (R), which was seen to result in a significant wave height of 14.3 m.

Table 6. Summary of Recommended *Hs-100* (m)

	A	B	C	D	E	R
Value based exclusively on NESS data	10.8	11.0	11.3*	11.0	10.5	14.3
Recommended value including all corrections/uncertainties	12.6	12.4	12.2*	12.0	12.0	

10. Evaluation

Each submission contains a fair number of steps that require the use of good judgement and subjective reasoning. Several issues are simply not amenable to quantitative evaluation. For instance, the reason for selecting a particular approach may be that it is a given group's standard way of dealing with extreme value problems, or it may be an approach strongly favored by one or more people, or it may be a series of procedures developed over the years, which enjoys a history of frequent and successful use. At the same time, each group must attempt to derive a result that is theoretically defensible as well as one that will in all likelihood be acceptable to the outside world (management, designers, regulatory agencies, etc.)

Consequently, there are several aspects of the submissions that are difficult to interpret. Keeping these limitations in mind, it seems reasonable to identify the following basic criteria to assess the quality of a particular approach:

(1) How practical and clear is the suggested approach? A convincing procedure must be logical and simple to use.

(2) Is the method theoretically sound and does it lead to accurate results? Is it based on recognized statistical techniques and proven results from extreme value theory?

(3) Can the method be generalized easily to other gridpoints and locations or is it very dependent on a particular data structure? How wide is its range of applicability?

(4) How sensitive is the method to assumptions regarding data, distribution types? Is the method robust? Can confidence intervals easily be constructed? Is parameter/statistical uncertainty taken into account?

(5) Does the method allow for adjustment based on measured data; is the hindcast model uncertainty taken into account?
In fact, two questions should be considered:
(5a) When measured data are available at the site, can a suitable procedure be used to incorporate them in any way? and;
(5b) When no measured data are available at the site, can the intrinsic hindcast model uncertainty be accounted for in a NESS EV analysis?

A detailed evaluation is not attempted, but it is felt that most of the methods used would get a fine score against each of the above criteria, with the exception of criterion (5b). This is due to the lack of a consistent technique to account for the intrinsic hindcast model uncertainty, even in the absence of measured data. Another weakness would be reflected in criterion (4): parameter uncertainty and/or short data uncertainty should be addressed in extrapolating to high return values.

On a theoretical level, we feel somewhat uneasy about the use of "cumulative" data (as opposed to working with storm peaks): the implication on estimating high extreme values is not clear. As far as distribution choices are concerned, three considerations jump to mind. First, we are somewhat surprised that no contribution included the (3 parameter) generalized extreme value cdf in the analysis; this is a particularly flexible distribution and it could virtually be used on its own to model a wide range of tail behaviors. By the same token, no attempt was made to look at an analysis based on seasonal extremes, month-by-month extremes, and the effective use of more than just one high order statistic (for instance through the use of the i dimensional generalized extreme value cdf). Weighted least squares also failed to be selected as a convenient way to correct tail behavior. With regard to the POT analyses, it is somewhat puzzling to see that distributions which would not be expected to yield good fits were included in the analysis (one would expect POT density functions to be monotonically decreasing starting at the threshold).

Further discussion is needed to investigate the quality of the different approaches.

11. Problem Issues

In the course of evaluating the different submissions, it becomes clear that there are a number of grey areas with issues that will need to be dealt with at some point in the future. Guidance needs to be sought from experts in EV analysis and from experienced oceanographers and engineers with regard to these matters. The following list may not be complete, but it does contain a set of both general and particular issues identified in the process of analyzing the five contributions:

11.1 Issues related to EV analysis

1. Selection of storm peaks from 3 hourly data; smoothing/interpolation of peaks; storm duration: can the 18 h window criterion be relaxed?
2. Least Squares Methods: plotting position to be used, particulary in the case of upper tail analysis; in which direction should errors be considered: variable, log (exceedance probability), weights?
3. POT: How many data are needed; How does the threshold need to be selected (almost all of the contributors used different criteria); In deriving EVs, is it preferable to use quantiles simply on the basis of an adjusted exceedance probability or on the basis of a compound Poisson cdf?
4. Develop means to construct confidence intervals associated with some of the more complicated methods. Only contributor E made an attempt to account for parameter uncertainty.
5. When using the "cumulative" (all data) approach, assess the impact of correlation between peaks, particularly when only a small percentage of the top data is used.
6. Evaluate the impact of discontinuous data on determining r year return periods; in the particular case of AEV, what is the impact of using 6 month (winter) extremes?

11.2 Additional Issues

7. Measured Data; the various used/proposed methods require a detailed examination. Clarification and consensus is needed on how to "combine" hindcast and measured data. Some of the approaches reflect a sense of "we know what number we want to get close to, so let's select a method that will get us there"; this arbitrary aspect should be addressed.
8. Spatial Spreading; this issue was not part of the present analysis, but contributor C showed that any method should also be applicable to a series of gridpoints, rather than just one gridpoint.
9. Inclusion of recent storms and/or recently observed high *Hs* values in a NESS extreme value analysis. Only contributor A explicitly addressed this seemingly important issue.

12. Appendix A. Acronyms Used for Distribution Functions and Analysis Methods

A,B,C,D,E: the 5 contributing groups

AEV: Annual Extreme Value (Method)

BM: Borgman cdf

$$F(x) = \exp\left[-\exp\left(-\frac{x^2-a}{b}\right)\right] \quad x > 0;\ b > 0$$

EXP: exponential cdf:

$$F(x) = 1-\exp\left(-\frac{x-a}{b}\right) \quad x > a;\ b > 0$$

EV: Extreme Value

FR2: 2 parameter Frechet (or Fisher-Tippett Type 2, or log extremal) cdf:

$$F(x) = \exp\left[-\left(\frac{x}{b}\right)^{-c}\right] \quad x > 0;\ b, c > 0$$

FR3: 3 parameter Frechet (or Fisher-Tippett Type 2, or log extremal) cdf:

$$F(x) = \exp\left[-\left(\frac{x-a}{b}\right)^{-c}\right] \quad x > 0;\ b, c > 0$$

FR2T: Left-truncated Frechet cdf

FT3: the Fisher-Tippett Type 3 (or, the inverted Weibull) cdf:

$$F(x) = \exp\left[-\left(\frac{a-x}{b}\right)^{c}\right] \quad x < a;\ b, c > 0$$

G: Gumbel cdf

$$F(x) = \exp\left[-\exp\left(-\frac{x-a}{b}\right)\right]$$

$$-\infty < x < +\infty;\ b > 0$$

GG: 3 parameter generalized gamma cdf with probability density function:

$$f(x) = \frac{\lambda\beta}{\Gamma(\alpha)}(\lambda x)^{\alpha\beta-1}\exp\left[-(\lambda x)^{\beta}\right]$$

$$x > 0;\ \alpha, \beta, \lambda > 0$$

GOF: goodness-of-fit

GT: Left-truncated Gumbel cdf

HYPAS: Hybrid Parametrical Shallow Water Wave Model

MOM: method of moments

MML: methods of maximum likelihood

MLE: maximum likelihood estimate

LN: 2 parameter log-normal cdf (Log of the variable has a normal cdf)

LS, LLS: least squares, linear least squares

NESS: North European Storm Study

P: Pareto cdf

$$F(x) = 1 - x^{-\beta}, \quad x > 1;\ \beta > 0$$

POT: peak over threshold

R: the reference approach in the Guidance Notes

W2: the 2 parameter Weibull cdf (or the FT3 for minima)

$$F(x) = 1 - \exp\left[-\left(\frac{x}{b}\right)^{c}\right] \quad x > 0;\ b, c > 0$$

W3: the 3 parameter Weibull cdf (or the FT3 for minima)

$$F(x) = 1 - \exp\left[-\left(\frac{a-x}{b}\right)^{c}\right] \quad x > a;\ b, c > 0$$

W2T: Left-truncated Weibull cdf

Acknowledgments

The authors would like to thank the five contributors for the time spent on preparing their submission, and for their enthusiasastic participation in the study. Authorization from the NESS Users Group (NUG) to publish this paper is appreciated.

13. References

[1] C. K. Grant, J. C. Heideman, D. J. Peters, C. J. Shaw, and D. Szabo, Modelling the North Sea through the North European Storm Study, Presented at OTC 93, Houston TX (1993).

[2] H. Günther and W. Rosenthal, Shallow Water Surf. Wave Model Based on the Texel-Marsen-Arsloe (TMA) Wave Spectrum, Proc. 20th Congress of the Int. Assoc. of Hydraulic Research (IAHR), Moscow/SU (1983).

[3] Department of Energy, Metocean Parameters—Wave Parameters, Supporting Document to: Offshore Installations: Guidance on Design, Construction and Certification—Environmental Considerations, Techword Services, OTH 89 300, London, UK (1989).

[4] E. W. Stacy and G. A. Mirham, Parameter Estimation for a Generalized Gamma Distribution, Ann. Math. Stat. **33**, 1187-1192 (1965).

About the Authors: M. A. Maes, Ph.D., does research in the areas of risk analysis and reliability-based design at the Civil Engineering Department of the University of Calgary in Canada. G. Z. Gu, Ph.D., is a civil engineer with Dallas E&P Engineering at the Mobil Research Lab in Dallas, Texas.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 445 (1994)]

Extreme Value Analysis of Wave Heights

Volume 99 Number 4 July–August 1994

E. Castillo and J. M. Sarabia

University of Cantabria
Avenida de los Castros s/n,
39005,
Santander, Spain

This paper discusses the most common problems associated with the determination of design wave heights. It analyzes two common methods used in fitting wave data and shows some of the stability or inconsistency problems associated with commonly used distributions. Some methods to obtain confidence intervals, detecting of outliers and treatment of missing data are given.

Key words: design; extreme value distributions; wave heights.

Accepted: March 22, 1994

1. Introduction

As in many other fields of engineering, the design of ocean or marine structures is governed by extreme values of wave heights. Several methods have been given in the past for the determination of design values. However, no method is widely accepted by the engineering community.

Traditionally, the analysis of yearly maxima has been considered as a good method for this purpose. However, recently, peak value methods arose as a promising alternative.

The aim of this paper is to compare these two methods and illustrate some of the problems related to their use.

2. Two Standard Methods in the Determination of Wave Height Design Values

In this section we analyze the following two well known procedures for obtaining design wave heights: the *peak value method* and the *yearly maxima method*.

The first one employs the peak wave heights of individual storms and thus composes a set of extreme wave data. The second one uses the yearly maxima.

Several authors have criticized the second method in that it discards large wave heights, when they occur in years with large storms, but includes relatively small wave heights which are maxima of calm years.

2.1 The Peak Value Method

This method consists of the following steps:

1. Fit the peak values of individual storms to a parametric family of distributions

$$F_0(x;\ \lambda_0,\ \delta_0,\ \beta_0)\,, \tag{1}$$

where λ_0, δ_0 and β_0 are the parameters. In some cases these three parameters can be reduced to two or even to a single one. The fitting of the above family can be done either by using all data or only tail data (Peak over threshold (POT) method).

It is worthwhile mentioning that this distribution corresponds to the wave height of a storm, that is, we assume that the cdf of the maximum wave height of storms is Eq. (1).

2. Use the following cdf for the maximum wave height in a period of duration D years:

$$F_0(x;\ \lambda_0,\ \delta_0,\ \beta_0)^{Dk}, \tag{2}$$

where k is the mean number of storms per year, or determine the wave height, x_T, associated with a return period T, that is, solve, for x, the equation

$$F_0(x;\ \lambda_0,\ \delta_0,\ \beta_0)^k = 1-\frac{1}{T}. \tag{3}$$

Note that the cdf in Eq. (2) implies the assumption of independence of storms.

2.2 The Yearly Maxima Method

This method consists of the following steps:

1. Fit the yearly maxima to a parametric family of distributions

$$F_1(x;\ \lambda_1,\ \delta_1,\ \beta_1), \tag{4}$$

where λ_1, δ_1 and β_1 are the new parameters. This is equivalent to assuming that the yearly maxima follow a distribution which belongs to Eq. (4).

2. Use the following formula to extrapolate to the maximum of a period of D years:

$$F_1(x;\ \lambda_1,\ \delta_1,\ \beta_1)^D, \tag{5}$$

or determine the wave height x associated with a return period T, i.e., solve the equation for x:

$$F_1(x;\ \lambda_1,\ \delta_1,\ \beta_1) = 1-\frac{1}{T}. \tag{6}$$

3. Some Problems Related to the Data Analysis

In the analysis of data one has to deal with some problems. Among them we mention the following:

- Selection of the families $F_0(x;\ \lambda_0,\ \delta_0,\ \beta_0)$ or $F_1(x;\ \lambda_1,\ \delta_1,\ \beta_1)$
- Estimation of the parameters of the selected families
- Confidence interval determination
- Outlier detection
- Treatment of incomplete series

3.1 Some Distribution Families Used in the Analysis of Wave Data

The most common used distributions in the analysis of wave heights are the following:

1. The Gumbel family

$$F(x;\ \lambda,\ \delta) = \exp\left[-\exp\left(\frac{\lambda-x}{\delta}\right)\right];$$
$$-\infty < x < \infty \tag{7}$$

2. The maximal Weibull family

$$F(x;\ \lambda,\ \delta,\ \beta) = \exp\left\{-\left(\frac{\lambda-x}{\delta}\right)^{\beta}\right\};$$
$$x \leqslant \lambda \tag{8}$$

3. The maximal generalized extreme value or Jenkinson's family

$$F(x) = \exp\left\{-\left[1+\frac{(x-B)}{kA}\right]^{-k}\right\};$$
$$1+\frac{(x-B)}{kA} \geqslant 0 \tag{9}$$

4. The minimal Weibull family

$$F(x;\ \lambda,\ \delta,\ \beta) = 1-\exp\left\{-\left(\frac{x-\lambda}{\delta}\right)^{\beta}\right\};$$
$$x \geqslant \lambda \tag{10}$$

The Gumbel, maximal Weibull and maximal Jenkinson's families are justified from a theoretical point of view, because they are the limit distributions for maxima (see Galambos [5] or Castillo [2]). It is interesting to note that the Jenkinson's family includes the other two, as particular cases, and the maximal Frechet family (for $k>0$). The Frechet distribution is not justified in this case because wave heights are physically limited, no matter we deal with shallow or deep waters (see Castillo and Sarabia [3] and [4]).

The minimal Weibull distribution, though widely used, is not theoretically justified in the case of maxima. Its only justification is that its range can be made to be consistent with the positive character of wave heights. In addition, we remind the reader that it belongs to the maximal domain of attraction of the Gumbel type, i.e., it is asymtotically equivalent to a Gumbel distribution of the type Eq. (7).

However, due to the fact that this distribution is widely used in the analysis of wave heights, it seems convenient to make here some comments.

Initially we can say that this distribution has the following advantages:

- For $\lambda = 0$, its range is $(0, \infty)$, that is, it does not include negative values of the random variable.
- Assuming that the location parameter, due to physical reasons, is fixed to zero, it depends only on two parameters. This makes the estimation process much simpler.
- Its associated domain of attraction is Gumbel type. Thus, it could be used if this were the actual case.

Its main drawbacks are the following:

- Its range is unbounded on the right. This contradicts the physical reality.
- It does not cover the Weibull domain of attraction that could be the real situation.
- It is an asymptotical minimum law.
- It is not stable with respect to maximum operations. Thus, if the minimal Weibull law is satisfied for yearly maxima the maxima of periods of duration different from one year cannot satisfy this law. This problem can be solved by adding an extra parameter to this family, which leads to the extended minimal Weibull family.

Consequently, the minimal Weibull family could be used if and only if we were sure that the domain of attraction of wave heights is of a Gumbel type.

In order to determine the domain of attraction of a given distribution several methods are available, such as the Pickands' or the curvature methods (see Castillo [2] chapter 6 and Castillo, Galambos and Sarabia [3]).

3.2 Estimation Methods

Several methods have been used to estimate the parameters of the families Eqs. (7) to (10). The most important are:

- The maximum likelihood method
- The method of moments
- The least squares method
- The probability paper method
- The Goda's method
- The percentile method

3.2.1 The Maximum Likelihood Method This method is based on maximizing the likelihood of data with respect to the parameters. The central idea consists of assuming that the sample comes from a population with parent distribution belonging to a parametric family and choosing the parameter values that maximize the probability of ocurrence of the sample data.

This is the best known method in statistics and it is recognized as the most convenient, due to its statistical properties. It leads to the best estimators, which, in addition, are asymptotically normal. This allows asymptotic confidence intervals for the parameters to be easily obtained. Using the δ-method, to be described later, the confidence interval of any regular function of the parameters can be obtained, too. In particular, confidence intervals of percentiles can be obtained in this manner.

In order to estimate an extreme value distribution with the purpose of extrapolation beyond the data range, only high order statistics must be used and the rest must be discarded. Thus, we recommend the method indicated by Castillo [2], in chapter 5.

In the case of the minimal and maximal Weibull families, the estimation process can lead to some problems, either because the likelihood function becomes unbounded ($\beta \leq 1$) or because some nonregularities, for some values of the shape parameter ($1 < \beta < 2$). However, it can be applied to values of the shape parameter larger than or equal to 2 without any problem. Thus, once the estimates are available, it is necessary to check that their values are consistent with the initial hypothesis. Here we give the following recommendations:

- If the shape parameter takes a negative value, this means that the data indicate a Frechet type domain of attraction. This suggests the presence of at least one outlier that gives an erroneous curvature in the right tail.
- If we get a value of $\beta \leq 1$, we can think on the presence of outliers. This value of the shape

parameter indicates that the probability density function is increasing in the tail, which contradicts the physical reality.

- If we get $1<\beta<2$ then the law is far from the Gumbel law (note that Gumbel corresponds to $\beta=\infty$).
- If the value of the λ parameter is less than the maximum of the sample this indicates that there is an outlier.

3.2.2 The Method of Moments This method consists of equating the moments of the sample to the moments of the theoretical distribution. We use as many moments as there are parameters to be estimated and we get the same number of equations from which the parameters can be obtained. The asymptotic properties of the moment estimates are good but worse than those associated with the maximum likelihood estimates.

This method can also be applied to tail estimation, using the moments of the truncated distribution.

3.2.3 The Least Squares Method This method consists of minimizing the sum of squares of the differences between the theoretical and the empirical values. There are many versions of this method. In some cases the random variable scale is used to measure the errors and in other cases the probability or the return period scales are used (see chapter 4 of Castillo [2]).

The main advantage of these methods is that they give an explicit solution and do not depend on convergence of any algorithm, as is the case with the maximum likelihood method.

Nevertheless, these methods are sensitive to the plotting position formulas used in the estimation method.

3.2.4 The Probability Paper Method By probability paper method we understand a visual method, in which the data is drawn on probability paper and a straight line is visually fitted to data.

The main drawback of this method is that it depends on the plotting position formula used in the graphic representation and the subjective criteria for selecting the optimal fit.

3.2.5 The Goda Method Goda [4] fits a minimal Weibull distribution, truncated at the threshold value x_0, to the right tail of data. By right tail are meant the wave heights above a second threshold value $x_1 > > x_0$.

3.2.6 The Percentile Method One way of obtaining quick estimates of the parameters of a distribution is by means of the percentile method. This method consists of equating as many percentiles in the sample and the theoretical distribution as the number of parameters to be estimated.

As an illustrative example we use this method for the estimation of the parameters of a three parameter maximal Weibull family.

The cdf of the maximal Weibull distribution is:

$$G(x)=\exp\left[-\left(\frac{\lambda-x}{\delta}\right)^{\beta}\right]. \tag{11}$$

Thus, the percentile or order p satisfies the equation

$$p=\exp\left[-\left(\frac{\lambda-x_p}{\delta}\right)^{\beta}\right], \tag{12}$$

from which we get

$$x_p=\lambda-\delta(-\log p)^{1/\beta}. \tag{13}$$

Equating the three percentiles of orders p_1, p_2, p_3 of sample and population, we get the following system of equations:

$$x_{pi}=\lambda-\delta(-\log p_i)^{1/\beta};\ i=1,2,3, \tag{14}$$

where p_i can be written, using the Gringorten's formula, as:

$$p_i=\frac{i-0.44}{n+0.12}, \tag{15}$$

where i is the rank of the order statistic associated with p_i.

From Eq. (14) we get

$$\frac{x_{p_2}-x_{p_1}}{x_{p_3}-x_{p_2}}=\frac{(-\log p_2)^{1/\beta}-(-\log p_1)^{1/\beta}}{(-\log p_3)^{1/\beta}-(-\log p_2)^{1/\beta}}, \tag{16}$$

which depends only on the parameter β and thus, it can be easily solved by an iterative method, as the bisection method for example, with a personal computer. Once β is known, the values of λ and δ can be obtained from any two of the equations in Eq. (14). For example:

$$\delta=\frac{x_{p_2}-x_{p_1}}{(-\log p_1)^{1/\beta}-(-\log p_2)^{1/\beta}};$$

$$\lambda=x_{p_i}+\delta(-\log p_1)^{1/\beta} \tag{17}$$

For the estimates to be consistent with the model we must have

$$\lambda > \max(x_1, x_2, \ldots, x_n) \tag{18}$$

where $(x_1, x_2, \ldots, x_n)$ is the sample.

If the percentiles are arbitrarily chosen, this inconsistency can easily appear. Thus, it is good practice to choose as one of the percentiles the maximum of the sample $x_{(n)}$.

In addition, if we are dealing with a tail estimation we must choose the adequate percentiles, that is, percentiles in it.

In order to improve the quality of the estimates we can use three groups of percentiles instead of three percentiles, that is, replace the system Eq. (14) by the system

$$\frac{1}{m_j}\sum_{i=k_j}^{i=k_j+m_j-1} x_{p_i} = \lambda - \frac{\delta}{m_j}\sum_{i=k_j}^{i=k_j+m_j-1} (-\log p_i)^{1/\beta}; \quad j = 1, 2, 3 \tag{19}$$

where m_j, $(j = 1, 2, 3)$ are the numbers of percentiles included in each group. With this, equation Eq. (16) becomes Eq. (20).

$$\frac{\dfrac{1}{m_2}\displaystyle\sum_{i=k_2}^{i=k_2+m_2-1} x_{p_i} - \dfrac{1}{m_1}\sum_{i=k_1}^{i=k_1+m_1-1} x_{p_i}}{\dfrac{1}{m_3}\displaystyle\sum_{i=k_3}^{i=k_3+m_3-1} x_{p_i} - \dfrac{1}{m_2}\sum_{i=k_2}^{i=k_2+m_2-1} x_{p_i}} = \frac{\dfrac{1}{m_2}\displaystyle\sum_{i=k_2}^{i=k_2+m_2-1} (-\log p_i)^{1/\beta} - \dfrac{1}{m_1}\sum_{i=k_1}^{i=k_1+m_1-1} (-\log p_i)^{1/\beta}}{\dfrac{1}{m_3}\displaystyle\sum_{i=k_3}^{i=k_3+m_3-1} (-\log p_i)^{1/\beta} - \dfrac{1}{m_2}\sum_{i=k_2}^{i=k_2+m_2-1} (-\log p_i)^{1/\beta}}. \tag{20}$$

3.2.7 Plotting Position Formulas There is much controversy about the plotting position formulas to be used for representing data on probability paper and the posterior estimation by least squares methods.

The resulting estimates are sensitive to the plotting position formulas being used. This confirms the fact that the least squares method is not optimal. Note that maximum likelihood or moment methods do not depend on plotting positions.

The discussion of the appropriateness of various formulas is intended to avoid or reduce some of the errors involved (in this case authors recommend using formulas leading to unbiased estimators).

However, we mention here that all plotting position formulas are asymptotically equivalent.

3.3 δ-Method

The δ-method (Bishop, Fienberg, and Holland [1]) allows us to obtain confidence intervals of certain regular functions of the parameters, as functions of the parameter estimates, and its variance-covariance matrix.

Let

$$\eta_i = h_i(\lambda_1, \lambda_2, \ldots, \lambda_s); i = 1, 2, \ldots, k \tag{21}$$

be k functions of the set of parameters $\lambda_1, \lambda_2, \ldots, \lambda_s$. Then, according to the δ-method,

$$(\hat{\eta}_1, \hat{\eta}_1, \ldots, \hat{\eta}_k) = (h_1(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_s), h_2(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_s), \ldots, h_k(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_s)) \tag{22}$$

is an estimator of $(\eta_1, \eta_2, \ldots, \eta_k)$ which is asymptotically normal and has mean

$$(h_1(\lambda_1, \lambda_2, \ldots, \lambda_s), h_2(\lambda_1, \lambda_2, \ldots, \lambda_s), \ldots, h_k(\lambda_1, \lambda_2, \ldots, \lambda_s)) \tag{23}$$

and variance-covariance matrix

$$\Sigma^* = \begin{bmatrix} \frac{\partial h_1}{\partial \lambda_1} & \frac{\partial h_2}{\partial \lambda_1} & \cdots & \frac{\partial h_k}{\partial \lambda_1} \\ \frac{\partial h_1}{\partial \lambda_2} & \frac{\partial h_2}{\partial \lambda_2} & \cdots & \frac{\partial h_k}{\partial \lambda_2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial h_1}{\partial \lambda_s} & \frac{\partial h_2}{\partial \lambda_s} & \cdots & \frac{\partial h_k}{\partial \lambda_s} \end{bmatrix}^T$$

$$\Sigma \begin{bmatrix} \frac{\partial h_1}{\partial \lambda_1} & \frac{\partial h_2}{\partial \lambda_1} & \cdots & \frac{\partial h_k}{\partial \lambda_1} \\ \frac{\partial h_1}{\partial \lambda_2} & \frac{\partial h_2}{\partial \lambda_2} & \cdots & \frac{\partial h_k}{\partial \lambda_2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial h_1}{\partial \lambda_s} & \frac{\partial h_2}{\partial \lambda_s} & \cdots & \frac{\partial h_k}{\partial \lambda_s} \end{bmatrix} \tag{24}$$

where Σ is the variance-covariance matrix of $(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_s)$.

3.4 Estimation of Percentiles of the Maximal Weibull Distribution

As a simple example of the δ-method we give the confidence interval of one percentile of the three parameter maximal Weibull distribution. We assume that the parameter β is larger than 2.

3.4.1 Point Estimate The percentile x_p of the maximal Weibull distribution is:

$$x_p = \lambda - \delta(-\log p)^{1/\beta}. \tag{25}$$

Thus, according to the invariance principle, the maximum likelihood estimator of that percentile, is:

$$\hat{x}_p = \hat{\lambda} - \hat{\delta}(-\log p)^{1/\hat{\beta}}. \tag{26}$$

3.4.2 Maximum Likelihood Estimators: Asymptotic Theory We assume here that the sample consists of those observed values above the threshold value t (type II censoring), that is, the probability density function is given by

$$f_t(x) = \frac{f(x; \lambda, \delta, \beta)}{1 - F(t; \lambda, \delta, \beta)} \tag{27}$$

and f and F are the pdf and cdf of the maximal Weibull family.

The maximal Weibull distribution satisfies the necessary regularity conditions for the asymptotic normality if $\beta \geq 2$. Thus, if we have a sufficiently large sample coming from a population with maximal Weibull parent, we can write:

$$\sqrt{n}((\hat{\lambda}, \hat{\delta}, \hat{\beta}) - (\lambda, \delta, \beta)) \xrightarrow{D} N(\underline{0}, \Sigma) \tag{28}$$

where

$$\Sigma = (I + J)^{-1}, \tag{29}$$

where $I = (a_{ij})$ is the information matrix associated with the maximal Weibull family, that is,

$$a_{11} = I_{\lambda\lambda} = \frac{(\beta - 1)^2}{\delta^2} \Gamma(1 - \frac{2}{\beta}) \tag{30}$$

$$a_{12} = I_{\lambda\delta} = -\frac{\beta^2}{\delta^2} \Gamma(2 - \frac{1}{\beta}) \tag{31}$$

$$a_{13} = I_{\lambda\beta} = \frac{-1}{\beta}\Gamma(1 - \frac{1}{\beta}) + \frac{1}{\delta}\Gamma(2 - \frac{1}{\beta}) + \frac{1}{\delta}\Gamma'(2 - \frac{1}{\beta}) \tag{32}$$

$$a_{22} = I_{\delta\delta} = \frac{\beta^2}{\delta^2} \tag{33}$$

$$a_{23} = I_{\delta\beta} = -\frac{1}{\delta}\Gamma'(2) \tag{34}$$

$$a_{33} = I_{\beta\beta} = \frac{1}{\beta^2}(1 + \Gamma''(2)) \tag{35}$$

and J is the matrix of the second order partial derivatives, with respect to the parameters of the model, of the function

$$G(t; \lambda, \delta, \beta) = \log[1 - F(t; \lambda, \delta, \beta)].$$

If we consider now the function:

$$f(\lambda, \delta, \beta) = x_p = \lambda - \delta(-\log p)^{1/\beta}, \tag{36}$$

with partial derivatives:

$$f_1 = \frac{\partial f}{\partial \lambda} = 1$$

$$f_2 = \frac{\partial f}{\partial \delta} = -(-\log p)^{1/\beta} \tag{37}$$

$$f_3 = \frac{\partial f}{\partial \beta} = \frac{\delta}{\beta^2} \log((-\log p))(-\log p)^{1/\beta}$$

then, by the δ-method we have:

$$\sqrt{n}(f(\hat{\lambda},\hat{\delta},\hat{\beta})-f(\lambda,\delta,\beta)) \xrightarrow{D} N(\underline{0},\Sigma^*) \quad (38)$$

where

$$\Sigma^* = (f_1,f_2,f_3)\Sigma(f_1,f_2,f_3)' , \quad (39)$$

from which the confidence interval for the percentile x_p at level α becomes:

$$(\hat{x}_p - z_{1-\alpha/2}\frac{\Sigma^{*1/2}}{\sqrt{n}} ; \hat{x}_p + z_{1-\alpha/2}\frac{\Sigma^{*1/2}}{\sqrt{n}}) . \quad (40)$$

3.5 Outlier Detection

In this section we give a method to detect the presence of outliers in the sample data. The method is based on the fact that if we make the following change of variable:

$$Y = F(X) \quad (41)$$

where $F(x)$ is the cdf of X, the resulting random variable, Y, is uniform $U(0,1)$.

In addition we know that the maximum of a random sample of size n coming from a standard uniform parent has cdf

$$F_{Y_{\max}}(y) = y^n = \text{Prob}[Y_{\max} \leq y] . \quad (42)$$

We shall say that the sample maximum is one outlier if the probability of being exceeded is very small. Thus, the value y_0 can be considered as critical for the maximum value of the sample if

$$\text{Prob}[Y_{\max} > y_0] = F_{Y_{\max}}(y_0) = 1 - y_0^n = \alpha \quad (43)$$

with a very small (0.01, 0.05, etc.). Then, we get

$$y_0 = (1-\alpha)^{1/n} \quad (44)$$

This critical value refers to the random variable Y. Thus, we need to obtain X by means of the inverse of Eq. (41). As one example, for the maximal Weibull distribution we get

$$x_0 = \hat{\lambda} - \hat{\delta}[-\log(1-\alpha)^{1/n}]^{1/\beta} . \quad (45)$$

3.6 Treatment of Incomplete Series

If we know about the existence of r storms in a given series, but we ignore the peak intensities we can perform an estimate based on the known peaks and then make a correction for the unknown peaks. This means estimating the cdf with the known peaks and raise to the power $(n+r)/n$, where n and r are the number of known and unknown peaks, respectively.

4. Critical Analysis

In this section we analyze the previous methods and discuss some of their inconsistencies.

4.1 Inconsistencies due to the Lack of Stability With Respect to Maximum Operations

When several design methods are recognized by the engineering community a certain consistency in the respective results should be expected. We shall see that this is not the case for some of the previous methods.

Let us assume that we try to fit the minimal Weibull family

$$F(x;\lambda_0,\delta_0,\beta_0) = 1-\exp\left\{-\left(\frac{x-\lambda_0}{\delta_0}\right)^{\beta_0}\right\}, \quad (46)$$

where λ_0, δ_0 and β_0 are the parameters. Then, Eq. (2) transforms to

$$F_D(x;\lambda_0,\delta_0,\beta_0) = F(x;\lambda_0,\delta_0,\beta_0)^{Dk} = \left\{1-\exp\left[-\left(\frac{x-\lambda_0}{\delta_0}\right)^{\beta_0}\right]\right\}^{Dk} \quad (47)$$

and the wave height associated with a return period T becomes Eq. (3):

$$x_T = \lambda_0 + \delta_0\left\{-\log\left[1-\left(1-\frac{1}{T}\right)^{1/k}\right]\right\}^{1/\beta_0} \quad (48)$$

Let us assume that now we also use the minimal Weibull family in Eq. (4):

$$F_1(x;\lambda_1,\delta_1,\beta_1) = 1-\exp\left\{-\left(\frac{x-\lambda_1}{\delta_1}\right)^{\beta_1}\right\}, \quad (49)$$

where $\lambda_1, \delta_1, \beta_1$ are the new parameters. Then, Eq. (5) becomes

$$F_D(x; \lambda_1, \delta_1, \beta_1) = F_1(x; \lambda_1, \delta_1, \beta_1)^D = \left\{1-\exp\left[-\left(\frac{x-\lambda_1}{\delta_1}\right)^{\beta_1}\right]\right\}^D \quad (50)$$

and the wave height associated with a return period T, from Eq. (6), is

$$x_T = \lambda_1 + \delta_1 \left\{-\log\left[1-\left(1-\frac{1}{T}\right)\right]\right\}^{1/\beta_1} \quad (51)$$

The minimal Weibull model is inconsistent in the following sense: It is not stable with respect to maximum operations, that is, when the cdf is raised to a given power $s \neq 1$, then, the resulting cdf is not minimal Weibull. Thus, though Eq. (46) is minimal Weibull, Eq. (47) is not minimal Weibull for $Dk \neq 1$. In other words, if we assume a minimal Weibull distribution for the peaks of storms, the yearly maxima cannot be minimal Weibull and vice versa. In fact for equations Eqs. (47) and (48) to be identical to Eqs. (50) and (51), respectively, i.e., for consistency, we must have

$$\lambda_0 = \lambda_1, \delta_0 = \delta_1, \beta_0 = \beta_1, k = 1 \quad (52)$$

which implies $k = 1$, that is a mean number of one storm per year, which is not the case.

However, if, instead of using the minimal Weibull family we use the maximal Gumbel family

$$F_0(x; \lambda_0, \delta_0) = \exp\left[-\exp\left(\frac{\lambda_0 - x}{\delta_0}\right)\right], \quad (53)$$

then, Eqs. (47), (48), (50) and (51) become

$$F_D(x; \lambda_0, \delta_0) = \left\{\exp\left[-\exp\left(\frac{\lambda_0 - x}{\delta_0}\right)\right]\right\}^{Dk} \quad (54)$$

$$x_T = \lambda_0 - \delta_0 \log\left[-\log\left(1-\frac{1}{T}\right)^{1/k}\right] \quad (55)$$

$$F_D(x; \lambda_1, \delta_1) = \left\{\exp\left[-\exp\left(\frac{\lambda_1 - x}{\delta_1}\right)\right]\right\}^D \quad (56)$$

and

$$x_T = \lambda_1 - \delta_1 \log\left[-\log\left(1-\frac{1}{T}\right)\right] \quad (57)$$

and, taking into account that

$$\left\{\exp\left[-\exp\left(\frac{\lambda_1 - x}{\delta_1}\right)\right]\right\}^D = \left\{\exp\left[-\exp\left(\frac{\lambda_1 - x}{\delta_1}\right)\right]\right\}^{Dk\,1/k} = \left\{\exp\left[-\exp\left(\frac{\lambda_1 - \delta_1 \log k - x}{\delta_1}\right)\right]\right\}^{Dk} \quad (58)$$

the coincidence of the pairs Eqs. (54)–(56) and Eqs. (55)–(57) implies

$$\delta_0 = \delta_1 \; y \; \lambda_0 = \lambda_1 - \delta_1 \log k. \quad (59)$$

That is, the coincidence of both is possible for any value of k.

The same conclusion is valid for any of the Weibull Eq. (8) or the Jenkinson's Eq. (9) families.

4.2 Inconsistencies Associated With the Lack of Stability With Respect to Truncation

Goda's method is inconsistent for the following reasons:

1. It gives different estimators for different values of x_0.
2. If the truncated distribution belongs to the minimal Weibull family it cannot belong for a different threshold value. Thus, different designers using different threshold values necessarily arrive to different models.

In the following paragraphs we shall make a detailed analysis of this problem.

With respect to the first inconsistency it is clear that because the method only uses the data above the second threshold value x_1, the resulting estimates should be independent on the first threshold value x_0.

In relation to the second inconsistency, the model should be stable with respect to truncations. With the purpose of clarifying this idea, let us assume that we choose a family of candidate distributions $H(x;y)$, where the second argument y is one parameter, which, without loss of generality, can be assumed to be the threshold value. Then, if the wave height exceeding z has as cdf the function $H(x,z)$, then, the wave height exceeding y should have a cdf given by

$$\frac{H(x;z)-H(y;z)}{1-H(y;z)} = H(x;y), \tag{60}$$

where the right hand term arises from the consistency condition that expresses that the family $H(x;z)$ remains valid for any value of the threshold parameter, which in this case is y.

Equation (60) is a functional equation. Its general solution can easily be obtained by making $z = z_0$, that is,

$$H(x;y) = \frac{G(x)-G(y)}{1-G(y)}, \tag{61}$$

where

$$G(x) = H(x;z_0). \tag{62}$$

For $H(x;y)$ to be a cdf, then $G(x)$ must also be a cdf.

Equation (61) proves that any consistent family $H(x;y)$ must come from another family $G(x)$ by means of a truncation procedure.

The minimal Weibull family, used by Goda, does not satisfy this condition. Thus, it is inconsistent.

With the purpose of having a consistent family in the two previously given senses, one solution would consist of assuming $G(x)$ to be extended minimal Weibull with null location parameter. This would imply that the sample data above the threshold value x_1 should be fitted to the family

$$H(x;y,\delta,\beta) = \frac{[F(x;\delta,\beta)]^n-[F(y;\delta,\beta)]^n}{1-[F(y;\delta,\beta)]^n}, \tag{63}$$

where

$$F(x;\delta,\beta) = 1-\exp\left[-\left(\frac{x}{\delta}\right)^{\beta}\right], \tag{64}$$

Nevertheless, we remind the reader that this solution can be satisfactory only in the case of a parent distribution in the domain of attraction for maxima of a Gumbel type.

Consequently, as a summary, we recommend to fit the sample data above the threshold to one of the following three families:

- If the domain of attraction is Weibull type, fit the right tail to the maximal Weibull family

$$F_0(x;\lambda_0,\delta_0,\beta_0) = \exp\left\{-\left(\frac{\lambda_0-x}{\delta_0}\right)^{\beta_0}\right\} \tag{65}$$

- if the maximal domain of attraction is Gumbel type, fit the right tail to the maximal Gumbel family

$$F_0(x;\lambda_0,\delta_0,) = \exp\left[-\exp\left(\frac{\lambda_0-x}{\delta_0}\right)\right] \tag{66}$$

or to the extended minimal Weibull family

$$F_0(x;\lambda,\delta,\beta) = \left\{1-\exp\left[-\left(\frac{x}{\delta}\right)^{\beta}\right]\right\}^{\eta}; \quad x \geq 0 \tag{67}$$

where β, δ and η are the parameters to be estimated. In the last case we are assuming that the cdf of the maximum wave height in an indeterminate period, to be estimated, is minimal Weibull.

Note that fitting the right tail means fitting a truncated model with basic distribution given by Eqs. (65), (66) or (67).

All these models are consistent in the previously mentioned sense.

4.3 Inconsistencies Associated With the Use of H_s and T_z

It is very common in the Ocean Engineering field to work with the significant wave height, H_s, and period, T_z, as the basic variables for extreme value analysis of waves. However this is not correct because T_z is the mean zero up-crossing period and H_s is defined as the mean of the 1/3 largest waves. These two random variables are convenient to justify normality assumptions in wave spectra, but cannot be accepted if an extreme value analysis of wave height, H, is to be performed. In fact, distributions in different domain of attraction types can lead to the same distribution for H_s and/or T_z, thus, obscuring the tail properties of single waves.

5. Conclusions

From all the above we get the following conclusions:

1. The most convenient families to fit wave height data in the tails are:
   - The maximal Weibull family
   - The maximal Gumbel family

   However, the extended minimal Weibull family can be used too.

   Before fitting the Gumbel or the extended minimal Weibull families, the domain of attraction for maxima must be checked using, for example, the Pickands or the curvature methods. For the estimation of the parameters, the maximum likelihood or the method of moments applied to the truncated distributions is recommended.

   In the case of the maximal Weibull family, the shape parameter β must be larger than unity. If it is not, the data suggests an increasing probability density function in the tail, which contradicts the reality.
2. It is recommended the elimination of outliers by means of the following iterative method:
   (a) Estimate all parameters with all data but the maximum
   (b) Check for the outlier character of the maximum by the previously indicate method
   (c) If it is an outlier, remove the maximum and start the process again; if it is not, repeat the estimation with all the valid data
3. If there are missing data correct the obtained cdf by raising to the power $(n+r)/n$ where n and r are the number of known and unknown data, respectively.
4. Significant wave height H_s and mean up-crossing periods T_z are not adequate variables to analyze the extreme value behaviour of wave heights.

Acknowledgments

The authors are grateful to M. J. Martín Soldevilla, J. Martínez-Aranzábal, and M. García Mañes from the Centro de Estudios de Puertos y Costas, CEPYC-CEDEX for their helpful discussions and to the University of Cantabria and the Dirección General de Investigación Científica y Técnica (DGICYT) (project PB91-0302), for partial support of this work.

6. References

[1] Y. M. M. Bishop, S. E. Fienberg, and P. W. Holland, Discrete Multivariate Analysis: Theory and Practice, The MIT Press, Cambridge, MA (1975).
[2] E. Castillo, Extreme Value Theory in Engineering, Academic Press, New York (1988), p. 389.
[3] E. Castillo, J. Galambos, and J. M. Sarabia, The selection of the domain of attraction of an extreme value distribution from a set of data, Lecture Notes in Statistics 51, Proceedings Oberwolfach (1987) pp. 181-190.
[4] E. Castillo and J. M. Sarabia, Engineering analysis of extreme value data: Selection of models. J. of Waterways, Port, Coastal Ocean Eng., **118** (2),129-146 (1992).
[5] J. Galambos, The Asymptotic Theory of Extreme Order Statistics, Krieger Publishing Company, Malabar, Florida (1987), p. 414.
[6] Y. Goda, On the Methodology of Selecting Design Wave Height, Proceedings 21st. International Conference on Coastal Engineering, Costa del Sol (1988).

About the authors: Enrique Castillo is a Civil Engineer and a Mathematician. He is also a Professor in the Department of Applied Mathematics and Computational Sciences at the University of Cantabria, Spain. José María Sarabia is a Statistician and a Professor of the Department of Economics at the University of Cantabria, Spain.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 455 (1994)]

Dynamic Amplification of Jack-Up Platforms Subjected to Non-Gaussian Wave Loads

Volume 99 Number 4 July–August 1994

Jørgen Juncher Jensen

Department of Ocean Engineering, Technical University of Denmark, Building 101 E, DK-2800 Lyngby, Denmark

Jack-up platforms are sensitive to dynamic amplification in waves because their fundamental period can be as high as 7 s–8 s. Whereas the dynamic motion of the platforms is rather well described by linear theory the excitation depends nonlinearly on the wave height. The stochastic wave loading is thus far from being normally distributed.

In this paper dynamic amplifications obtained by the diffusion theory extended to cover nonnormal excitations are compared with available time simulation results in irregular seaways. Generally, the time simulations seem to yield responses less nonlinear in the wave heights than the responses estimated from the diffusion theory. Explanations for these discrepancies are discussed and mainly attributed to a proper choice of total damping.

Key words: diffusion theory; dynamic amplification; Jack-up platforms; non-Gaussian waveloads; nonlinear wave response.

Accepted: March 22, 1994

1. Introduction

The jack-up drilling rig concept, Fig. 1, has proved to be very convenient in the exploration for oil and gas in offshore areas. Therefore, requests are made for designs able to operate in increasing water depths. Due to their sizes and independent leg configuration the natural periods of their lowest vibration modes become comparable with the dominant wave periods in the design sea states. As an example jack-up rigs with leg lengths of 160 m, hull masses of the order 15,000 t and lowest natural periods around 8 s are currently under construction. For such structures dynamic amplification of the wave load responses is certainly to be expected.

The wave loading on the legs can be estimated using Morison's equation. Usually, the legs are truss-like, with each leg consisting of three (or four) vertical chords connected by horizontal and oblique bracing members. The diameters of the individual members are so small that the wave loads on the legs become drag dominated. Alternative designs for smaller platforms have considered circular cylindrical legs, yielding inertia-dominant waveloads [1]. For larger platforms the circular cylindrical leg design is not feasible as the loadings and thereby the required amount of steel are so much higher than for the truss leg design that it cannot be counterbalanced by lower production costs.

A drag-dominated wave load implies a loading which is nonlinear in the wave height. Furthermore, the integration of the wave load up to the actual position of the wave elevation on a leg and the non-symmetry (Stoke's 5th order wave) of the wave profile magnify these nonlinearities in the base shear, the overturning moment and associated responses.

The structural stiffness of the jack-up rig in the lowest vibration mode is characterized by the leg stiffness, the distance between the legs, the bottom

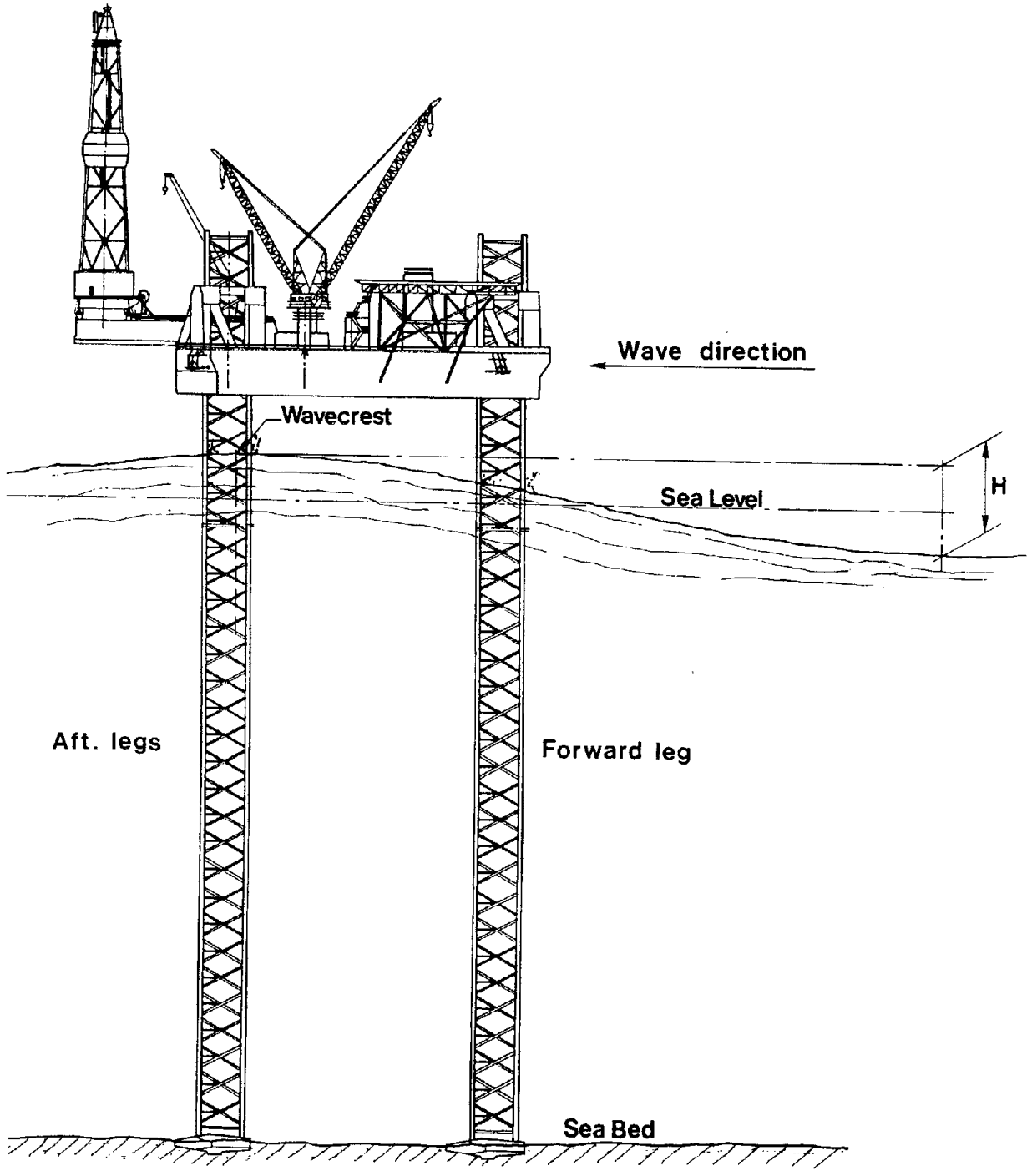


Fig. 1. Jack-up platform with cantilever.

support conditions, the distance L from sea bottom to the platform deck and the leg-jack-house flexibility. The global vibration pattern is normally beam-like with a maximum horizontal deck deflection of the order of 1 %–2 % of L in the design sea states. A linear structural analysis would therefore normally suffice. However, the additional overturning moment in the deflected state due to the high axial leg loads from the deck mass must be included by the so-called $P - \delta$ effect and, in a dynamic analysis, by reducing the leg bending stiffness.

A jack-up rig is a highly stressed structure. Therefore, it is important that an accurate structural evaluation is performed. Such an analysis should be done, not only

when designing a new structure, but also when a jack-up rig is moved to a new location. In order to get uniform and reliable site approval procedures, a large study was initiated by a group of companies involved in jack-up design and operations. A summary of this project is presented in [2]. Generally three different methods are applied:

- Single degree of freedom methods (SDOF).
- Frequency domain methods.
- Time-domain methods.

In the first procedure the quasistatic solution, determined by neglecting the motion of the platform, is amplified by a dynamic amplification factor (DAF) calculated by the classical SDOF formula. This will introduce errors due to the nonlinearities in the wave loading and several approximative procedures have been used [3]–[5], aiming at reducing these errors.

The frequency domain methods rely on a suitable linearization of the wave loads with the wave height. The resulting linear dynamic system is then solved exactly. The non-Gaussian behavior of the extreme values is then simply estimated by multiplying the standard deviation of the linear response with a factor depending on the ratio between the root-mean-square values of the drag and inertia terms in the wave loads.

Due to the assumptions inherent into above-mentioned methods, time simulation procedures are often used. For a specific stationary stochastic sea state random time signals of wave elevation and corresponding wave kinematics are generated, typically by superposition of first order (Airy) wave components. A structural analysis of the jackup-rig including dynamic and nonlinear effects is then carried out using time steps of the order of 0.5 s. The main drawback in this method is that due to excessive computational costs only a limited number of time simulations, each covering a few hours, can be generated. The extrapolation of these results to extreme value predictions for design approval can be difficult. Several applications of time simulation procedures to jack-up rigs have been published [1]–[2], [4]–[7].

In a previous paper by the author [8], an alternative method has been developed. The method is based on exact solution of a linear single-degree-of-freedom system subjected to a non-Gaussian excitation. Like in the SDOF method the present method also needs the nonlinear quasistatic response as input but now the dynamic effects are calculated much more consistently. The present procedure yields all required statistical moments of the response, which makes extreme value predictions very easy. This is done without using any stochastic linearization procedures as required in the frequency domain method. Finally, compared to time simulation procedures, the present method is much faster to apply and does not have the problems with extreme value predictions inherent in time simulation procedures.

The aim of the present paper is to evaluate the proposed procedure [8] by comparing results with those obtained from time simulation procedures. Previous comparisons [8], [9], with results based on a usual SDOF procedure have been very favorable as well as have been comparisons with time simulation results for a large offshore jacket structure [10]. In the next section the present stochastic dynamic procedure is described. Then it is applied to data presented in [5], obtained using a time simulation procedure and the importance of the various approximations and different modelings is discussed.

2. Stochastic Dynamic Analysis

For a linear single-degree-of-freedom system the equation of motion for the response $Y(t)$ can be written

$$\ddot{Y}(t)+2\zeta_0\,\omega_0\dot{Y}(t)+\omega_0^2(t)=\omega_0^2\,Y_0(t),$$

where ω_0 and ζ_0 is the natural frequency and damping ratio, respectively. Time is denoted by t and differentiation with respect to t by $(\cdot)$. The function $Y_0(t)$ is seen to be the quasistatic response obtained neglecting the dynamic behavior of the structure $(\omega_0\to\infty)$.

If $Y_0(t)$ represents a global jack-up response variable like the base shear or overturning moment, then it has been shown, [3], [8]–[10] that it can quite accurately be represented by a polynomial description in terms of the wave height H.

$$Y_0(t)=\sum_{i=0}^{n} A_i\,h(t)^i \tag{2}$$

$$h(t)=\frac{H}{2}\cos\left[2\pi\frac{t}{T}+\epsilon\right], \tag{3}$$

when the jack-up is subjected to a regular long-crested wave. The coefficients A_i will depend on the platform geometry, the water depth, the wave

theory applied and the current profile. No closed-form solution exists and the actual values of A_i must be derived by curve fitting from the numerical results.

The wave period T in the applied regular wave and in equivalent wave elevation h is taken to be uniquely given by the wave height H, using for instance Odland's formula [11]

$$T = 1 + 4.1\, H^{0.4}, \tag{4}$$

with T in seconds and H in meters.

In a stationary stochastic sea state with significant wave heights H_s the individual wave heights H have been found to be Rayleigh distributed with a root-mean-square value close to $H_s/2\sqrt{2}$. Furthermore, the phase lag ϵ in Eq. (3) can be taken to be uniformly distributed. Then the parameter h, Eq. (3), becomes normal distributed with zero mean and standard deviation equal to $H_s/4$. Thereby, the stochastic equivalent of Eq. (2) becomes

$$Y_0(t) = \sum_{i=0}^{n} a_i\, U(t)^i \tag{5}$$

with

$$a_i = A_i\,(H_s/4)^i \tag{6}$$

and where $U(t)$ is a standard Gaussian process with zero mean and unit variance. In most cases a cubic polynomial, $n = 3$, will suffice.

Clearly the curve fitting and the specific values of $T = T(H)$ used to obtain the quasistatic response description, Eq. (2), impose some inaccuracies in the coefficient a_i. Therefore, if time simulation results are available for the stochastic quasistatic response, Y_0, then these results could be used directly to generate proper values of a_i. For example, from the four lowest statistical moments: mean, standard deviation, skewness, and kurtosis, it is straightforward to determine the four coefficients a_i in a cubic description of $Y_0(t)$ [8]. This possibility will be considered in the next section.

The solution of Eq. (1) with the right hand side given by Eq. (5) will be based on the theory of diffusion processes. Therefore the forcing function $\xi(t)$ must be a normal white-noise process with a covariance function satisfying

$$E[\xi(t)\,\xi(t+\tau)] = 2\pi S\,\delta(\tau), \tag{7}$$

where S is the spectral density of ξ and $\delta(\tau)$ is Dirac's delta function. A constant spectral density is a very poor approximation for a wave load process $U(t)$ and the standard procedure to overcome this problem is to pass the white noise process $\xi(t)$ through a filter defined by

$$\ddot{\eta} + 2\,\zeta_g\,\omega_g\,\dot{\eta} + \omega_g^2\,\eta = \xi. \tag{8}$$

Thereby, the spectral shape S_η of η becomes

$$S_\eta(\omega) = \frac{2S}{(\omega^2 - \omega_g^2)^2 + (2\,\zeta_g\,\omega_g\,\omega)^2}\ ;\ \omega \geq 0. \tag{9}$$

Compared to the usual wave spectra of the Pierson-Moskowitz type the spectrum $S_\eta(\omega)$ has the disadvantage that $S_\eta(0) \neq 0$. The spectral shape $S_\varphi(\omega)$

$$S_\varphi(\omega) = S_\eta(\omega)\,(\omega/\omega_g)^2 \tag{10}$$

of the process

$$\varphi(t) = \dot{\eta}(t)/\omega_g \tag{11}$$

much better resembles the Pierson-Moskowitz spectrum. This is illustrated in Fig. 2, where the normalized spectra $S_\eta(\omega)$, $S_\varphi(\omega)$ are compared with the normalized Pierson-Moskowitz spectrum

$$S_{PM}(\omega) = 4\tau_p^{-4}\,\omega^{-5}\,\exp\left(-\left(\tau_p\,\omega\right)^{-4}\right)\ ;\ \omega \geq 0\ . \tag{12}$$

Here

$$\tau_p = \frac{T_p}{2\pi}\left(\frac{4}{5}\right)^{1/4}, \tag{13}$$

where T_p is the spectral peak period for the sea state. The normalizations are such that all three spectra have a unit variance implying that

$$S = \frac{2\,\zeta_g\,\omega_g^3}{\pi} \tag{14}$$

for both S_η *and* S_φ.

Furthermore, the spectral parameters ω_g and ζ_g are chosen such that the peak values and peak frequencies for all three spectra coincide [8], yielding

$$\begin{aligned} S_\eta(\omega):&\ \omega_g = 1.052\,\frac{2\pi}{T_p}\ ;\ \zeta_g = 0.221 \\ S_\varphi(\omega):&\ \omega_g = \frac{2\pi}{T_p}\ \ ;\ \zeta_g = 0.222 \end{aligned} \tag{15}$$

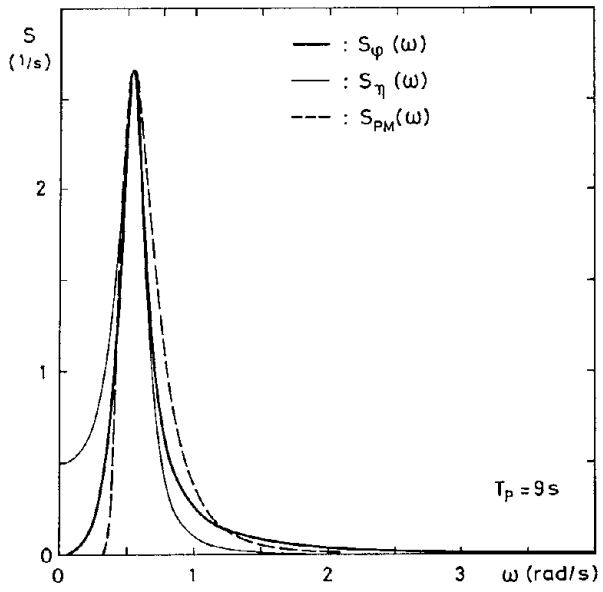


Fig. 2. Comparison between the normalized spectra $S_\eta(\omega)$, $S_\varphi(\omega)$ and $S_{PM}(\omega)$.

In the following normal process $U(t)$ in Eq. (5) will be taken as

$$U(t)=\varphi(t) \tag{16}$$

Eqs. (1), (5), (8), and (11) can be written as Ito differential equations

$$\dot{\boldsymbol{Z}} = \boldsymbol{C}(\boldsymbol{Z}(t))+\boldsymbol{W}(t), \tag{17}$$

where

$$\boldsymbol{Z}=\{Y,\ \dot{Y}/\omega_0,\ \eta,\ \varphi\}^T\equiv\{Z_1,\ Z_2,\ Z_3,\ Z_4\}^T \tag{18}$$

and

$$\boldsymbol{C}(\boldsymbol{Z})=\begin{bmatrix} \omega_0 Z_2 \\ -2\,\zeta_0\,\omega_0 Z_2-\omega_0 Z_1+\omega_0\sum_{i=0}^{n} a_i Z_4^i \\ \omega_g Z_4 \\ -2\,\zeta_g\,\omega_g Z_4-\omega_g Z_3 \end{bmatrix} \tag{19}$$

$$\boldsymbol{W}(t)=\{0,\ 0,\ 0,\ \xi(t)/\omega_g\}^T. \tag{20}$$

Since $\boldsymbol{W}(t)$ satisfies the white noise property

$$E[\boldsymbol{W}(t)\,\boldsymbol{W}(t+\tau)]=\boldsymbol{D}\,\delta(t)\ , \tag{21}$$

it follows from diffusion theory, e.g., [12] that the statistical mean value $E[g(\boldsymbol{Z})]$ of any function g of $\boldsymbol{Z}$ satisfies

$$\sum_i E\left[C_i(Z)\,\frac{\partial g}{\partial Z_i}\right]+$$

$$\frac{1}{2}\sum_i\sum_j D_{ij}E\left[\frac{\partial^2 g}{\partial Z_i\,\partial Z_j}\right]=0 \tag{22}$$

in a stationary sea state. In the present case $i,j=1$, 2, 3, 4 and only the D_{44} component in the 4×4 matrix D is different from zero.

As the vector $\boldsymbol{C}(\boldsymbol{Z})$ is given in polynomial form, it is straightforward to apply the procedure given by Krenk and Gluver [12] to obtain exact values for the statistical moment of Y. Here only the four lowest moments are determined and used to define uniquely a cubic polynomial approximation for $Y(t)$, [8]

$$Y(t)=c_0+c_1\,U(t)+c_2\,U(t)^2+c_3\,U(t)^3\ . \tag{23}$$

Extreme values of $Y(t)$ are finally obtained by replacing $U(t)$ with corresponding extreme values, that is by $\sqrt{2\ln N}$ for the most probable largest peak among N peaks.

3. Numerical Results

For a linear single-degree-of-freedom system subjected to a Gaussian excitation, a dynamic amplification factor can be defined as [11]

$$\text{dynamic amplification factor}=\frac{\left[\int_0^\infty \psi^2(\omega)\,S(\omega)\,\mathrm{d}\omega\right]^{1/2}}{\sigma_s}, \tag{24}$$

where ψ is the classical dynamic amplification factor

$$\psi(\omega)=\frac{\omega_0^2}{\sqrt{(\omega_0^2-\omega^2)^2+(2\zeta_0\,\omega\,\omega_0)^2}} \tag{25}$$

and where $S(\omega)$ is the spectral density of the quasistatic response (the excitation) Y_0. Furthermore, σ_s is the standard deviation of the excitation given by

$$\sigma_s^2=\int_0^\infty S(\omega)\,\mathrm{d}\omega. \tag{26}$$

Examples of dynamic amplification factors determined by Eq. (24) are shown in Fig. 3. Three spectral densities $S(\omega)$ are used, the Pierson-Moskowitz spectrum, Eq. (12), the spectrum $S_\eta(\omega)$, Eq. (9),

and the spectrum $S_\varphi(\omega)$, Eq. (10). The damping ratio is taken to be $\zeta_0 = 0.07$ and it is seen that at resonance the dynamic amplification factor is only about half the value of $\psi(\omega_0) = 1/2\ \zeta_0$.

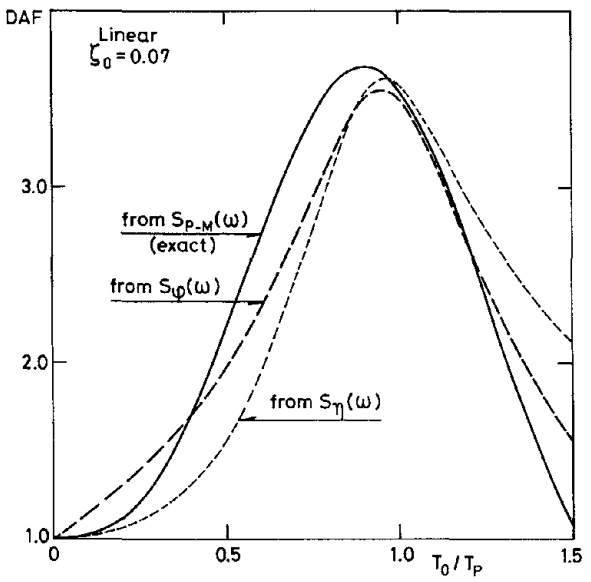


Fig. 3. Dynamic amplification factor (DAF), Eq. (24), for the standard deviation of a linear one degree-of-freedom system, using different excitation spectral densities.

For a linear system subjected to a Gaussian excitation the same dynamic amplification factor will apply to both standard deviations and extreme values. Typical values of the fundamental period $T_0 = 2\pi/\omega_0$ are around 8 s for large jack-up rigs whereas the peak spectral period T_p in the design sea state is about 16 s. From Fig. 3 one could then expect dynamic amplification factors in the vicinity of 2. However, most time simulation results [4]–[7], yield dynamic amplification factors for the extreme values much lower and even sometimes below 1. In the following this difference will be discussed using data for the example jack-up rig considered in [5].

First deterministic, quasistatic results for this jack-up rig were computed using both the Stretched Airy and the Stoke's 5th order wave theory. The results obtained for the overturning moment (OTM) using the Stretched Airy wave theory are shown in Fig. 4. Corrections for P-δ effects have been made. The wave period T is taken in accordance with Eq. (4). The sensitivity of the calculated results to the choice of T is exemplified in Table 1. It is seen that minor variations in T around the value given by Eq. (4) do not change the overturning moment significantly.

Two different curve fitting procedures have been used in Fig. 4 to generate cubic polynomial representations (A, B) of the overturning moment as function of the deterministic wave height H. Similar curves are obtained for the base shear and also when using the Stoke's 5th order wave theory. All these results are expressed in terms of coefficients A_i to be used in Eq. (2).

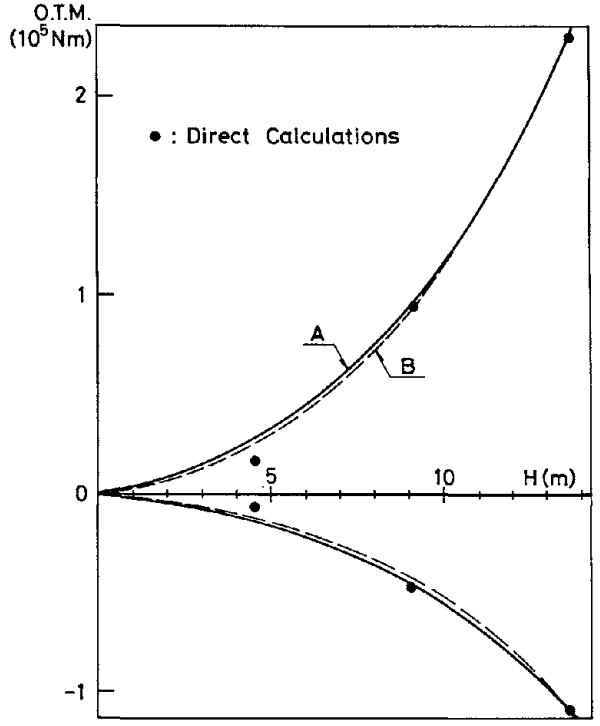


Fig. 4. Overturning moment for the jack-up-rig considered [5], subjected to a regular long-crested stretched Airy wave.

The coefficients A_i are then used in Eq. (6) to obtain values of a_i valid for the stationary stochastic design sea state. From these coefficients the four lowest statistical moments are calculated by the procedure given in [8]. The results are given in Tables 2 and 3 and compared with those presented in [5] from a quasistatic time simulation procedure. Furthermore, these tables contain the dynamic results determined by the present stochastic dynamic procedure and by dynamic time simulations in random seaways [5]. From Table 2 it is clear that the choice of wave theory and curve fitting procedure has only a marginal influence on the quasistatic overturning moment. Therefore Table 3 for the base shear only contains results for one of these choices. Also it appears that the statistical moments calculated from the deterministic results using Eqs. (2)–(6) are remarkably close to those found from time simulations except perhaps for the kurtosis which is somewhat lower in the time simulations.

Table 1. Sensitivity of overturning moment (OTM) to wave period T, $H = 13.7$ m (45 ft), Stoke's 5th order wave

Wave period T (s)	max OTM (10^5 Nm)	min OTM (10^5 Nm)
9.854	3.108	−0.811
11.261	2.635	−0.941
12.690 [Eq. (4)]	2.428	−1.112
14.077	2.457	−1.261

Table 2. Statistical moments and dynamic amplification factors (DAF) for the overturning moment in the design sea state, $H_s = 12.8$ m, $T_p = 15.5$ s

Overturning moment	Stoke's 5th fit A	Stoke's 5th fit B	Stretched Airy fit A	Stretched Airy fit B	Time simulation	Based on (*)
Quasistatic						
Mean/stand. dev.	0.197	0.198	0.192	0.193	0.200 (*)	
Skewness	2.60	2.77	2.50	2.61	2.99 (*)	
Kurtosis	25.7	29.7	24.3	27.0	18.4 (*)	
Dynamic						
Mean/stand. dev.	0.081	0.082	0.079	0.080	0.128	0.080
Skewness	0.17	0.19	0.16	0.18	1.29	0.19
Kurtosis	6.89	7.60	6.67	7.15	8.40	5.30
DAF/stand. dev.	2.43	2.41	2.43	2.41	1.49	2.50
DAF ($N = 1000$)	1.36	1.36	1.36	1.36	1.08	1.39

Table 3. Statistical moments and dynamic amplification factors (DAF) for the base shear in the design sea state, $H_s = 12.8$ m, $T_p = 15.5$ s

Base shear (BS)	Stoke's 5th	Time simulation [5]	Based on (*)
Quasistatic			
Mean/stand. dev.	0.163	0.160 (*)	
Skewness	2.43	2.23 (*)	
Kurtosis	32.2	13.7 (*)	
Dynamic			
Mean/stand. dev.	0.068	0.122	0.065
Skewness	0.17	1.34	0.12
Kurtosis	8.18	9.01	4.71
DAF (stand. dev.)	2.40	1.24	2.46
DAF ($N = 1000$)	1.37	1.05	1.38

The only additional information needed to calculate the dynamic responses by the present stochastic dynamic procedure is the fundamental period T_0 and the total damping ratio ζ_0, see Eq. (1). For the example jack-up rig $T_0 = 8.45$ s whereas the damping ratio is specified to 0.05 [5]. However, the time simulations are carried out in [5] using a single-degree-of-freedom formulation which includes coupled fluid-leg interaction terms. These terms will reduce the dynamic response and therefore act as additional (hydrodynamic) damping. The total damping in these time simulations must thus be greater than 0.05. In the present calculations, Eq. (1), (7)–(23), the damping ratio has been taken to be $\zeta_0 = 0.05$. The consequences of larger actual damping will be discussed later.

It is seen from Tables 2 and 3 that the dynamic amplification factor is nearly the same whether the quasistatic input is taken from Eqs. (2)–(6) or from the quasi static stochastic time simulations performed in [5] (the results marked by (*)). The difference in kurtosis is apparently not important.

For both the base shear and the overturning moment the dynamic amplification factors for the standard deviation turn out to be around 2.4 whereas the dynamic amplification factor for the most probable largest response peak among $N = 1000$ peaks becomes 1.36. As $T_0/T_p = 8.45/15.5 = 0.54$ the dynamic amplification factor for the standard deviation is seen to be in accordance with Fig. 3 taking into account that the damping in the present example is $\zeta_0 = 0.05$. Tables 2 and 3 also show that the dynamic effects tend to reduce the skewness and kurtosis making the response more Gaussian than the quasistatic response. Thereby, the dynamic amplification factors for the extreme values become smaller than for the standard deviation with decreasing values for increasing values of N.

The most severe disagreement between the results from the present stochastic dynamic procedure and the time simulations is clearly in the dynamic amplification factors. They are consistently smaller in the time simulations.

Before looking after possible explanations it should be stressed that the above results only concern one specific jack-up rig. Other results obtained by time simulations have shown larger dynamic amplification factors. For instance the dynamic amplification factor for standard deviation of the overturning moment is found in Ref. [6], Figs. 7 and

11 to be 5.583/2.521 = 2.21 for a comparable jack-up rig and sea state. Also [4] shows dynamic amplification factors around two without specifying precisely the extreme value level.

One source of uncertainty is the damping ratio. In the present stochastic dynamic procedure the total damping has to be specified whereas in the time simulation procedure the hydrodynamic damping is automatically taken into account by the relative velocity terms in Morison's equation. Therefore, it could be interesting to see how much the total damping should be increased before results in accordance with the time simulations are obtained. In Table 4 such results are shown and it is seen that first for a total damping ratio of about 20 % good agreement on dynamic amplification factors is obtained. Such damping is rarely expected in real jack-ups although [4] presents experimental values around 10 % for a model test. A total damping as low as 2.2 % has on the other hand been estimated from full scale measurements [14]. As mentioned previously, the stochastic dynamic time simulation procedure in [5] includes 5 % damping in addition to some damping from fluid-structure interactions (Eq. (26) in [5]). This must result in an effective total damping in the time simulation results greater than 5 % but how much greater it is not possible to say.

A further verification of the present procedure will clearly require more detailed comparisons with time simulation results including estimations of the total damping in the time simulations as function of the severity of the sea state. Until then it seems reasonable to assume that if the total damping is known then the present procedure will yield results with uncertainties mainly related to the assumption of a single degree-of-freedom system. Note that $P-\delta$ effects are included in T_0 through a reduced leg stiffness [11].

To illustrate the potential of the present procedure, Fig. 5 shows the variation with sea state of the non-Gaussian behavior and of the dynamic amplification factor for the most probable peak value among 1000 peaks for the overturning moment. In particular, one should note that even in extreme sea states some dynamic amplification occurs. There are two reasons for this. First the stochastic sea state averages out the classical dynamic amplification factor as shown in Fig. 3. Secondly, the non-Gaussian parts of the quasistatic excitation Y_0 are amplified differently. In Fig. 6 the dynamic amplifications associated with a pure quadratic and a pure cubic excitation are shown. The linear excitation, Fig. 3, has a maximum dynamic amplification factor for $T_0/T_p \simeq 1$, whereas the quadratic excitation yields maxima for $T_0/T_p \simeq 0.5$ and the cubic excitation maxima for $T_0/T_p \simeq 0.33$ and $T_0/T_p \simeq 1$. Thus depending on the relative magnitude of the linear (a_1), quadratic (a_2) and cubic (a_3) terms in the excitation Y_0, Eq. (5), the largest dynamic amplification factor can appear within a range of T_0/T_p values. For the example considered here in Fig. 5, the linear and especially the cubic terms dominate yielding a maximum dynamic amplification factor when the spectral peak period T_p gets close to the

Table 4. Statistical moments and dynamic amplification factors (DAFs) for the overturning moment in the design sea state, ($H_s = 12.8$ m, $T_p = 15.5$ s) as function of total damping ζ_0

ζ_0	Mean μ	Stand. dev. σ	Skewness κ_3	Kurtosis κ_4	DAF	
					Stand dev.	$N = 903$
0.05	20.5	258	0.19	5.3	2.50	1.39
0.10	20.5	193	0.52	7.1	1.87	1.26
0.15	20.5	165	0.90	8.4	1.60	1.19
0.20	20.5	148	1.25	9.4	1.44	1.13
Quasistatic time sim. [5]	20.5	103	2.99	18.4	1.00	1.00
Dynamic time sim. [5]	19.6	153	1.29	8.4	1.49	1.08

fundamental natural period T_0. Note, however, that most jack-up rigs have $T_0/T_p \simeq 0.5$ in the design sea state which is where the dynamic amplification factor from the quadratic term is largest. This is the reason why the dynamic amplification factor in Fig. 5 levels off for H_s around 13.7 m (45 ft).

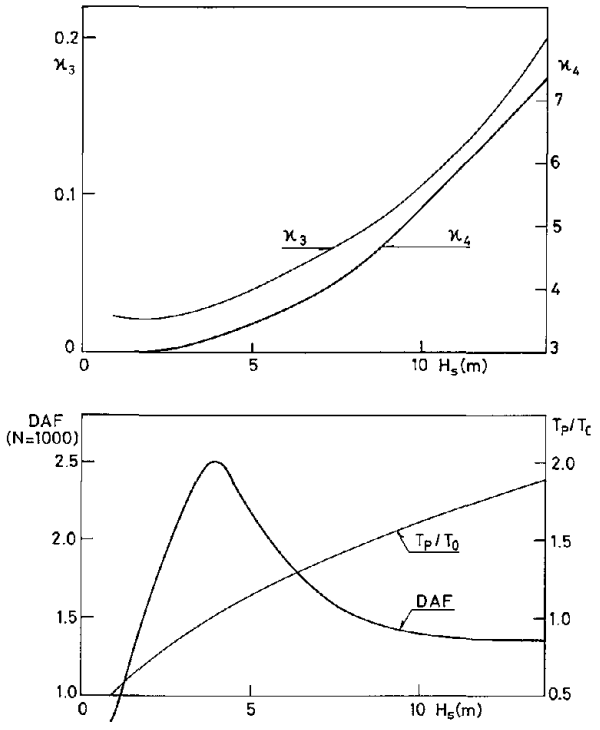


Fig. 5. Skewness κ_3, kurtosis κ_4 and dynamic amplification factors for the dynamic overturning moment as function of the significant wave height.

Finally, Fig. 7 shows for sake of completeness the variations of the skewness and the kurtosis for a dynamic system subjected to a pure quadratic and a pure cubic excitation. Note that a quadratic excitation has $\kappa_3 = 2\sqrt{2}$, $\kappa_4 = 15$ whereas the cubic excitation has $\kappa_3 = 0$, $\kappa_4 = 46.2$. The change towards a Gaussian behavior is clearly significant already for $T_0 > 0.1\ T_p$.

4. Conclusions

A procedure able to predict dynamic global responses of jack-up rigs subjected to wave loads in stationary stochastic seaways has been described.

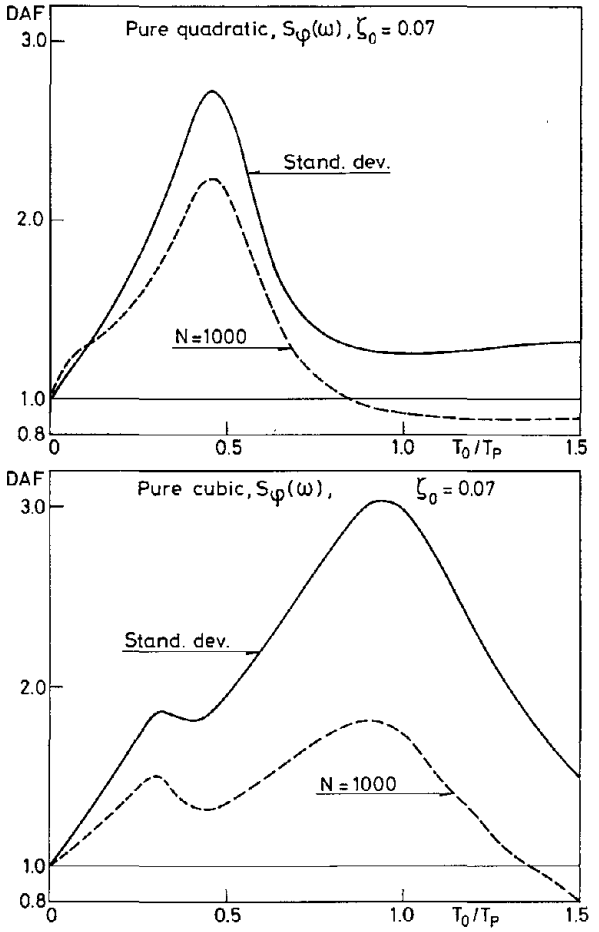


Fig. 6. The dynamic amplification of the standard deviation and the most probable largest peak among 1000 peaks for a pure quadratic and a pure cubic excitation.

The procedure consists of three steps:

(i) determine the wave load response using a suitable nonlinear regular wave theory neglecting the motion of the platform,

(ii) fit a polynomial in the wave height through the calculated response maxima and minima,

(iii) assume a deflection mode in the form of the first horizontal vibration mode and solve the corresponding equation of motion in stationary sea states using the theory of diffusion processes.

For an example jack-up rig it is observed that the dynamic amplification quite significantly changes the statistical behavior of the response toward a Gaussian process. Also, a significant dynamic amplification is found in the extreme sea states

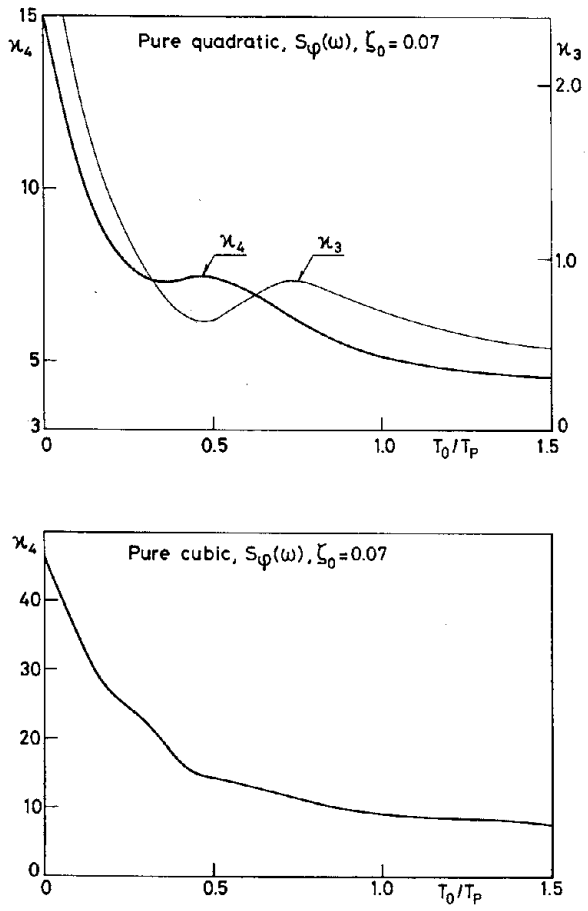


Fig. 7. Skewness κ_3 and kurtosis κ_4 for a dynamic system subjected to a pure quadratic and a pure cubic ($\kappa_3 = 0$) excitation.

where the spectral peak period is about twice the lowest natural period. These results are in agreement with previous findings [1]–[13] using various formulations and jack-up geometries.

Comparisons with time simulations performed for the example jack-up rig considered in [5] have indicated that the main uncertainty in the present procedure relates to a proper choice of total damping. In order to clarify this point estimates of total damping ratios from time simulation results would be extremely helpful. Such dampings would include fluid-structure interaction effects and will vary with time. Suitable average values must then be defined. In this context also doubts expressed on the use of the relative velocity term in Morisons equation should be mentioned [13].

Acknowledgments

The author wants to thank Dr. Y. N. Chen and Dr. Nigel Robinson for many valuable discussions.

5. References

[1] C. J. Mommaass and G. J. Gründlehner, Application of a Dedicated Stochastic Non-Linear Dynamic Time Domain Analysis Program in Design and Assessment of Jack-ups, Proc. Third Int. Conf. on The Jack-Up Drilling Platform, City University, London, Sept. 1991.

[2] M. J. R. Hoyle, Jack-Up Dynamics – A Review of Analysis Techniques, Proc. Third Int. Conf. on The Jack-Up Drilling Platform, City University, London, Sept. 1991.P

[3] J. J. Jensen, A. E. Mansour and P. T. Pedersen, Reliability of Jack-Up Platforms Against Overturning, Marine Structures **4**, 203–229 (1991).

[4] W. T. Bennett and R. V. Patel, Jack-Up Behaviour in Elevated Condition: Model Test and Computer Simulation, Proc. SNAME Spring Meeting/STAR Symposium, New Orleans, April 1989.

[5] Y. N. Chen, Y. K Chen and J. P. Cusack, Extreme Dynamic Response and Fatigue Damage Assessment for Self-Elevating Drilling Units in Deep Water, SNAME Trans., **98**, 143–168 (1990).

[6] H. Kjeøy, N. G. Bøe and T. Hysing, Extreme Response Analysis of Jack-Up Platforms, Proc. Second Int. Conf. on The Jack-Up Drilling Platform, Ocean Eng. Research Center, Dept. of Civil Engineering, London, September 1989.

[7] B. J. Leira and D. Karunakaran, Estimation of Fatigue Damage and Extreme Response for a Jack-Up Platform, Marine Structures **3**, 461–493 (1990).

[8] J. J. Jensen, Dynamic Amplification of Offshore Steel Platform Responses due to Non-Gaussian Wave Loads, Marine Structures **7**, 91–105 (1994).

[9] J. J. Jensen, Wave-Induced Dynamic Amplification of Jack-Up Rigs, Proc. ISMS '91, Shanghai, September (1991).

[10] J. Baatrup and J. J. Jensen, Dynamic Analysis of Offshore Jacket Platforms Subjected to Non-Gaussian Wave Loads, Proc. BRASIL OFFSHORE '91, Rio deJaneiro, 3–6 September 1991, Pentech Press, London (1990).

[11] J. Odland, Response and Strength Analysis of Jack-Up Platforms, Norwegian Maritime Research, No. 4 (1982) pp. 225.

[12] S. Krenk and H. Gluver, An Algorithm for Moments of Response from Non-Normal Excitation of Linear Systems, Stochastic Structural Dynamics: Progress in Theory and Applications, S. T. Ariaratnam et al., eds. Elsevier, Amsterdam (1988).

[13] O. T. Gudmestad and D. Karunakaran, Wave Current Interaction, Volume 26: Environmental Forces on Offshore Structures and their Predictions, Society for Underwater Technology, Kluwer Academic Publishers (1990).

[14] J. N. Brekke, R. B. Campbell, W. C. Lamb and J. D. Murff, Calibration of a Jack-up Structural Analysis Using Field Measurements from a North Sea Jack-up; Proc. Offshore Technology Conference 1990, paper No. OTC 6465, pp. 357–368.

About the author: Jørgen Jensen has a M. Sc. and Ph. D. in Mechanical Engineering from the Technical University of Denmark. He has been at the Department of Ocean Engineering at that university since 1976. Currently he is Head of the Department.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. **99**, 465 (1994)]

Prediction of Extreme Response of Nonlinear Oscillators Subjected to Random Loading Using the Path Integral Solution Technique

Volume 99 Number 4 July–August 1994

Arvid Naess

Faculty of Civil Engineering,
The Norwegian Institute of
Technology,
Rich. Birkelands vei la,
N-7034 Trondheim, Norway

This paper studies the applicability of the path integral solution technique for estimating extreme response of nonlinear dynamic oscillators whose equations of motion can be modelled by the use of Itô stochastic differential equations. The state vector process associated with such a model is generally a diffusion process, and the probability density function of the state vector thus satisfies the Fokker-Planck-Kolmogorov equation. It is shown that the path integral solution technique combined with an appropriate numerical scheme constitutes a powerful method for solving the Fokker-Planck-Kolmogorov equation with natural boundary conditions. With the calculated probability density function of the state vector in hand, one can proceed to calculate the required quantities for estimating extreme response. The proposed method distinguishes itself by remarkably high accuracy and numerical robustness. These features are highlighted by application to example studies of nonlinear oscillators excited by white noise.

Key words: extreme response; nonlinear oscillators; path integral solution; random loading.

Accepted: March 22, 1994

1. Introduction

An important element in the safety assessment of many engineering systems, is the task of estimating the probability of extreme events that may jeopardize the structure in some specified sense. Very often, this problem can be formulated as finding the probability that some time varying random quantity does not exceed a specified capacity level during a given time period. Stated this way, the problem typically reduces to a study of the extreme values of a stochastic process originating as the response of a system subjected to some stochastic loading process.

In this paper the focus will be on the problem of estimating the extreme response of nonlinear dynamic systems subjected to random forcing processes. In recent years the methods of time domain Monte Carlo simulations, see, e.g., Refs. [1-5], have received considerable attention as a tool for estimating response statistics. These methods are versatile and attractive in the sense that nonlinearities can be easily dealt with. The main drawback at present is the large CPU times needed for accurate prediction of extreme responses. Even if this issue seems to become less of an obstacle every year, portending perhaps that such methods may dominate practical estimation of response statistics of nonlinear systems in the not too distant future, it will still be desirable to have available alternative methods of calculating the response statistics, both simplified and more elaborate. Here we shall explore a method based on the theory of Markov diffusion processes. The justification for using this theory is related to the fact that the response of nonlinear dynamic systems to broad band random excitation can very often be accurately de-

scribed by applying the theory of multidimensional Markov processes. By this, the extensive theory of Markov diffusion processes can be brought to bear on these problems. In particular, it can be shown that the probability law of response quantities can be derived by solving a partial differential equation, viz., the Fokker-Planck (-Kolmogorov) (FPK) equation, see Refs. [6,7]. In most cases of practical interest, this equation has to be solved numerically.

In the next section we shall describe a method for solving the FPK equation that is based on a formal solution of the same equation. This solution is obtained by invoking the fact that a Markov diffusion process locally looks like a Brownian motion. By using the Markov property, the global solution can then be constructed by linking the local solutions, which are known explicitly. The obtained solution is generally known as a path integral solution (PIS). The reader is referred to Ref. [7] for a further discussion. One of the first efforts to exploit the PIS method explicitly in developing numerical solution algorithms is described in Ref. [8]. Subsequently, other authors have also used the PIS approach to solve various random vibration problems, cf. Refs. [9-13].

Before embarking on a description of the PIS method, it is expedient to briefly show how the obtained solutions are used in an extreme value analysis. Assuming that the response quantity of interest is a scalar (real) stationary stochastic process, $Z(t)$ say, the PIS method typically provides a numerical estimate of the joint probability density function (PDF) $f_{Z\dot{Z}}(\bullet,\bullet)$ of $Z(t)$ and $\dot{Z}(t)=\mathrm{d}Z(t)/\mathrm{d}t$. It is now assumed that the mean level upcrossing rate $\nu_Z^+(\bullet)$ of $Z(t)$ can be calculated from Rice's formula as follows

$$\nu_Z^+(z)=\int_0^\infty y f_{Z\dot{Z}}(z,y)\mathrm{d}y. \tag{1}$$

Adopting the assumption that upcrossing of high levels are statistically independent events, which leads to Poisson distributed crossings, it follows that an asymptotic approximation of the probability distribution function of the extreme value of the process $Z(t)$ during a time T, denoted by $M(T)$ $(=\sup\{Z(t);0\le t\le T\})$, is given by

$$Prob\{M(T)\le z\}=\exp\{-\nu_Z^+(z)T\} \quad (T\to\infty). \tag{2}$$

The accuracy of Eq. (2) depends to a large extent on the effective bandwidth of the response process $Z(t)$. Decreasing bandwidth leads eventually to a significant clumping effect of large response peaks, invalidating the assumption of statistically independent upcrossing of high levels. Methods that aim at correcting for this effect have been proposed for Gaussian (Refs. [14,15]) and non-Gaussian (Ref. [16]) processes. However, this point will not be pursued any further here. We shall assume that Eq. (2) provides an acceptable approximation. Hence, the central parameter to be determined is the upcrossing frequency $\nu_Z^+(\bullet)$, which is easily calculated once the joint PDF $f_{Z\dot{Z}}(\bullet,\bullet)$ of $Z(t)$ and $\dot{Z}(t)=\mathrm{d}Z(t)/\mathrm{d}t$ has been made available. In the next section it is shown how this PDF can be calculated for the response of a wide range of nonlinear oscillators subjected to white noise or filtered white noise loading.

2. The Path Integral Solution

The path integral solution (PIS) method is suitable for calculating the joint probability density function (PDF) of a vector process $\boldsymbol{X}(t)=[X_1(t),...,X_n(t)]^T$ (T-transposition) satisfying a stochastic differential equation of the following form, cf. Ref. [6],

$$\mathrm{d}\boldsymbol{X}(t)=\boldsymbol{m}[\boldsymbol{X}(t)]\mathrm{d}t+\boldsymbol{Q}[\boldsymbol{X}(t)]\mathrm{d}\boldsymbol{W}(t). \tag{3}$$

Here $\boldsymbol{m}(\bullet)=[m_1(\bullet),...,m_n(\bullet)]^T$, $m_j(\bullet)$ denotes a real function of n real variables. $\boldsymbol{Q}(\bullet)=(q_{ij}(\bullet))$ denotes an $n\times m$-matrix where each $q_{ij}(\bullet)$ is a real function of n real variables. $\boldsymbol{W}(t)=[W_1(t),...,W_m(t)]^T$ where $W_j(t)$, $j=1,...,m$ are standard, real Brownian motion processes, which are mutually independent, see e.g., Refs. [6,7]. That is, $E[W_j(t)]=0$ and

$$E[\mathrm{d}W_i(t)\mathrm{d}W_j(t+\tau)]=\delta_{ij}\delta_{t,t+\tau}\mathrm{d}\tau,\ i,j=1,...,m, \tag{4}$$

where $\delta_{xy}=1$ for $x=y$, $\delta_{xy}=0$ for $x\neq y$. Equation (4) is a short-hand notation for the relation $E[\int\int h(s,t)\mathrm{d}W_i(s)\mathrm{d}W_j(t)]=\delta_{ij}\int h(t,t)\mathrm{d}t$, where $h(\bullet,\bullet)$ is a non-random function.

Equation (3) is interpreted here as an Itô stochastic differential equation (SDE). Since it is often relevant to consider Eq. (3) as being obtained as a limit of equations with band limited noise processes, it may happen that $\boldsymbol{m}(\bullet)$ should contain correction terms to ensure a consistent limiting solution, cf. Ref. [6]. It is assumed here that this consideration has already been made, and that Eq. (3) has the final form to be used subsequently.

It is demonstrated in Ref. [6] that the solution $\boldsymbol{X}(t)$ to Eq. (3) is a Markov vector process. Its transition probability density function (TPD), $p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')$, is defined by the equation

$$Prob\{\boldsymbol{X}(t)\in A\,|\,\boldsymbol{X}(t')=\boldsymbol{x}'\}=\int_A\cdots\int p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')\mathrm{d}\boldsymbol{x}, \tag{5}$$

where $A\subseteq\mathbf{R}^n$ is some event, $\boldsymbol{x},\boldsymbol{x}'\in\mathbf{R}^n$, $\mathrm{d}\boldsymbol{x}=\mathrm{d}x_1...\mathrm{d}x_n$.

Provided that $\boldsymbol{m}(\bullet)$ and $\boldsymbol{Q}(\bullet)$ satisfy certain regularity conditions, see Ref. [6], it can be proved that the TPD $p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')$ $(t\geq t'\geq 0)$ is the solution of a partial differential equation of the form

$$\frac{\partial}{\partial t}p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')=-\sum_{i=1}^{n}\frac{\partial}{\partial x_i}[m_i(\boldsymbol{x})p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')]$$

$$+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2}{\partial x_i\partial x_j}[g_{ij}(\boldsymbol{x})p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')], \tag{6}$$

where $\boldsymbol{G}(\boldsymbol{x})=(g_{ij}(\boldsymbol{x}))=\boldsymbol{Q}(\boldsymbol{x})\boldsymbol{Q}(\boldsymbol{x})^T=(\sum_{k=1}^{m}q_{ik}\,q_{jk})$, and with initial condition $p(\boldsymbol{x},t'\,|\,\boldsymbol{x}',t')=\delta(\boldsymbol{x}-\boldsymbol{x}')$. $\boldsymbol{G}(\bullet)$ will be called the diffusion matrix and Eq. (6) will be referred to as the Fokker-Planck-Kolmogorov (FPK) equation. Since clearly $Prob\{\boldsymbol{X}(t)\in\mathbf{R}^n\,|\,\boldsymbol{X}(t')=\boldsymbol{x}'\}=1$ the TPD satisfies the following normalization condition

$$\int\cdots\int_{\mathbf{R}^n}p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')\mathrm{d}\boldsymbol{x}=1. \tag{7}$$

Let $f(\boldsymbol{x},t)$ denote the PDF of the random vector $\boldsymbol{X}(t)$. If $f(\boldsymbol{x},t')=w(\boldsymbol{x})$ for some initial PDF $w(\boldsymbol{x})$, then it is recognized from Eq. (6) and the relation

$$f(\boldsymbol{x},t)=\int\cdots\int_{\mathbf{R}^n}p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')w(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' \tag{8}$$

that $f(\boldsymbol{x},t)$ itself is a solution of Eq. (6) satisfying the initial condition $f(\boldsymbol{x},t')=w(\boldsymbol{x})$.

In this paper we shall be interested primarily in stationary solutions $f_s(\boldsymbol{x})$ to Eq. (6), that is

$$f_s(\boldsymbol{x})=\lim_{t\to\infty}f(\boldsymbol{x},t)=\lim_{t\to\infty}p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t') \tag{9}$$

provided they exist. Even when both limits exist, it is clear that $\lim f(\boldsymbol{x},t)$ provides the faster convergence when the initial condition $f(\boldsymbol{x},t')\approx f_s(\boldsymbol{x})$. This comment is relevant to the numerical implementation of the PIS method, and will be discussed below.

To obtain the PIS appropriate for the dynamic systems studied in this paper, it is necessary to be more specific on the structure of the matrix function $\boldsymbol{Q}(\bullet)$. In particular, it will be assumed that the first r rows of $\boldsymbol{Q}(\bullet)$ are zero, that is

$$q_{ij}(\bullet)\equiv 0 \text{ for } i=1,\ldots,r;\ j=1,\ldots,m\ (r<n) \tag{10}$$

and that $q_{ij}(\bullet)\not\equiv 0$ for at least one j for every $i=r+1,\ldots,n$. This implies that the diffusion matrix $\boldsymbol{G}(\bullet)$ assumes the form

$$\boldsymbol{G}(\bullet)=\begin{bmatrix}\boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{O} & \tilde{\boldsymbol{G}}(\bullet)\end{bmatrix}. \tag{11}$$

$\boldsymbol{O}$ denote appropriate zero-matrices and $\tilde{\boldsymbol{G}}(\bullet)$ denotes an $(n-r)\times(n-r)$-matrix function with elements $g_{ij}(\bullet)$, $i,j=r+1,\ldots,n$. $\tilde{\boldsymbol{G}}(\bullet)$ will be called the reduced diffusion matrix. Equation (6) can now be rewritten as

$$\frac{\partial}{\partial t}p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')=-\sum_{i=1}^{n}\frac{\partial}{\partial x_i}[m_i(\boldsymbol{x})p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')]$$

$$+\frac{1}{2}\sum_{i=r+1}^{n}\sum_{j=r+1}^{n}\frac{\partial^2}{\partial x_i\partial x_j}[g_{ij}(\boldsymbol{x})p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')]. \tag{12}$$

Proceeding in a manner similar to the derivations in Ref. [7], it can be shown that the TPD for small values of $\tau(=t-t')$ is given by the following expression, which is correct up to terms of order τ^2

$$p(\boldsymbol{x},t+\tau\,|\,\boldsymbol{x}',t)=\{\prod_{i=1}^{r}\delta(x_i-x_i'-m_i(\boldsymbol{x}')\tau)\}$$

$$\bullet(2\pi\tau)^{-\frac{n-r}{2}}|\tilde{\boldsymbol{G}}(\boldsymbol{x}')|^{-\frac{1}{2}}\bullet\exp\{-\frac{1}{2\tau}\sum_{i=r+1}^{n}\sum_{j=r+1}^{n}$$

$$(x_i-x_i'-m_i(\boldsymbol{x}')\tau)\,[\tilde{\boldsymbol{G}}(\boldsymbol{x}')^{-1}]_{i-r,j-r}(x_j-x_j'-m_j(\boldsymbol{x}')\tau)\}, \tag{13}$$

where $|\tilde{\boldsymbol{G}}|$ denotes the determinant of the reduced diffusion matrix $\tilde{\boldsymbol{G}}$, assumed to be positive definite. This implies that $|\tilde{\boldsymbol{G}}|>0$. $[\tilde{\boldsymbol{G}}^{-1}]_{ij}$ denotes the element in position i,j of the inverse matrix of $\tilde{\boldsymbol{G}}$. As shown in Ref. [7], the expression given by Eq. (13) is not unique, but seems to be well suited for our purpose.

Having obtained an explicit expression for the TPD for a short time step, one can now invoke the Markov property. This allows a TPD over a time interval of arbitrary length to be expressed in terms of a product of short-time TPDs. By dividing a given time interval (t',t) into N small time intervals of length $\tau=(t-t')/N$, it is found that $(t_j=t'+j\tau,\ t=t^N,\ \boldsymbol{x}=\boldsymbol{x}^{(N)},\ t'=t_0,\ \boldsymbol{x}'=\boldsymbol{x}^{(0)})$

$$p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')=\int\cdots\int_{\mathbf{R}^{n(N-1)}}\prod_{j=1}^{N}p(\boldsymbol{x}^{(j)},t_j\,|\,\boldsymbol{x}^{(j-1)},t_{j-1})\mathrm{d}\boldsymbol{x}^{(1)}\cdots\mathrm{d}\boldsymbol{x}^{(N-1)}. \tag{14}$$

Similarly, with an initial PDF $f(\boldsymbol{x}',t')=w(\boldsymbol{x}')$, the PDF $f(\boldsymbol{x},t)$ will be given by

$$f(\boldsymbol{x},t)=\int\cdots\int_{\mathbf{R}^{nN}}\prod_{j=1}^{N}p(\boldsymbol{x}^{(j)},t_j\,|\,\boldsymbol{x}^{(j-1)},t_{j-1})w(\boldsymbol{x}^{(0)})\mathrm{d}\boldsymbol{x}^{(0)}\cdots\mathrm{d}\boldsymbol{x}^{(N-1)}. \tag{15}$$

Hence, by combining Eq. (13) with Eqs. (14) or (15), a formal (approximate) solution of the FPK equation can be written. Equations (14) and (15), which are often referred to as PIS, constitute the core of the numerical

solution procedure to be described subsequently. It is realized that a numerical solution according to this method, automatically provides the evolution in time of the (conditional) PDF of the Markov process $X(t)$ from given start conditions in terms of an initial density $f(x',t')=w(x')$, including the degenerate case $f(x',t')=\delta(x'-x_0)$, for some starting point x_0. It is also worth noting how the PIS relates to the physics of the dynamic model, which is expressed through the coefficients $m_j(\bullet)$ and $q_{ij}(\bullet)$, cf. Eq. (3). The evolution in time of the PDF as expressed by the PIS, is seen to be directly determined by these coefficients in an explicit manner. This fact is a very important advantage of the PIS method, and reveals its fundamental physical significance.

3. Numerical Implementation

In the numerical implementation, the PIS is obtained by an iteration process based on the Chapman-Kolmogorov equation expressed as

$$p(x^{(j)},t_j \mid x',t')=\int\cdots\int_{R^n} p(x^{(j)},t_j \mid x^{(j-1)},t_{j-1})p(x^{(j-1)},t_{j-1} \mid x',t')dx^{(j-1)}. \quad (16)$$

The discretization of state space for the numerical solution makes it appropriate to employ an interpolation and smoothing procedure to increase the numerical efficiency. It was found that application of cubic B-splines, as detailed in Ref. [17], offered the desired accuracy and smoothness for the type of problems considered in this paper. This procedure was used as follows. At each time step $t_{j-1}\rightarrow t_j$, $p(x^{(j-1)},t_{j-1} \mid x',t')$ is represented as a cubic B-spline series in the following manner

$$p(x^{(j-1)},t_j \mid x',t')=\sum_{k_1=1}^{M_1}\cdots\sum_{k_n=1}^{M_n}\Gamma^{(j-1)}(k_1,\cdots,k_n)\bigotimes_{i=1}^{n}B_{k_i}(x^{(j-1)}), \quad (17)$$

where M_j=number of grid points for the i'th state variable x_i, $\{\otimes_{i=1}^{n}B_{k_i}(\bullet)\}_{k_i=1}^{M_i}$ is a tensor product basis of cubic B-splines and $\{\Gamma^{(j-1)}(k_1,\cdots,k_n)\}_{k_i=1}^{M_i}$ is the set of interpolation coefficients associated with time t_{j-1}. It is assumed that each set $\{B_{k_i}(\bullet)\}_{k_i=1}^{M_i}$, $i=1,\cdots,n$, is a basis of cubic B-splines associated with the knot sequence determined by the grid points for the i'th variable x_i. The tensor product B-spline is defined by

$$\bigotimes_{i=1}^{n}B_{k_i}(x)=\prod_{i=1}^{n}B_{k_i}(x_i). \quad (18)$$

The representation of $p(x^{(j-1)},t_{j-1} \mid x',t')$ by B-splines makes it possible to retain high numerical accuracy even with a fairly coarse basic grid if $p(x^{(j-1)},t_{j-1} \mid x',t')$ is not too singular. By substituting from Eq. (17) into Eq. (16), Eq. (19) is obtained

$$p(x^{(j)},t_j \mid x',t')=\sum_{k_1=1}^{M_1}\cdots\sum_{k_n=1}^{M_n}\Gamma^{(j-1)}(k_1,\cdots,k_n)\int\cdots\int_{R^n}p(x^{(j)},t_j \mid x^{(j-1)},t_{j-1})\bigotimes_{i=1}^{n}B_{k_i}(x^{(j-1)})dx^{(j-1)}. \quad (19)$$

It is seen from Eq. (13) that since $m_j(\bullet)$ and $g_{ij}(\bullet)$ are not functions of time t, the TPDs cannot depend on absolute time, but only on the time increment. Markov processes whose TPDs have this property, are called homogeneous. It follows that

$$p(x^{(j)},t_j \mid x^{(j-1)},t_{j-1})=p(x^{(j)},\tau \mid x^{(j-1)},0),\ j=1,2,..., \quad (20)$$

which holds for any $t_j-t_{j-1}=\tau\geq 0$.

From Eqs. (19) and (20) it is seen that for a fixed value of the time increment τ, each of the integrals on the right hand side of Eq. (19) need to be calculated only once, and can be stored for repeated use. That is, the following parameters are calculated initially and stored

$$B_{k_1\cdots k_n}^{l_1\cdots l_n}=\int\cdots\int_{R^n}p(x_{(l_1\cdots l_n)}^{(j)},\tau \mid x^{(j-1)},0)\bigotimes_{i=1}^{n}B_{k_i}(x^{(j-1)})dx^{(j-1)}. \quad (21)$$

Here, the index l_i, $i=1,\ldots,n$, refers to grid point number l_i for the state space variable x_i. It may be noted here that due to the properties of the TPD for small time increments τ, the tensor $B_{k_1\cdots k_n}^{l_1\cdots l_n}$ has a strongly banded character with the elements decreasing rapidly away from the main diagonal $k_1=l_1,\ldots,k_n=l_n$. This has important implications for the efficiency of the computer program. Let $p_{l_1\cdots l_n}^{(j)}=p(x_{(l_1\cdots l_n)}^{(j)},t_j \mid x',t')$. Then Eq. (19) can be rewritten as

$$p_{l_1\cdots l_n}^{(j)}=\sum_{k_1=1}^{M_1}\cdots\sum_{k_n=1}^{M_n}\Gamma^{(j-1)}(k_1,\ldots,k_n)B_{k_1\cdots k_n}^{l_1\cdots l_n}. \quad (22)$$

Having calculated the TPD $p(x^{(j)},t_j \mid x',t')$ at the grid points by using Eq. (22), a spline interpolation is again carried out and a new set of interpolation coefficients $\{\Gamma^{(j)}(k_1,\ldots,k_n)\}_{k_i=1}^{M_i}$ are calculated. This provides an updated representation of the TPD for time step j, cf. Eq. (17). For each time step, the normalization condition Eq. (7) is checked. That is, if

$$\int\cdots\int_{R^n}p(x^{(j)},t_j \mid x',t')dx^{(j)}=\sum_{k_1=1}^{M_1}\cdots\sum_{k_n=1}^{M_n}\Gamma^{(j)}(k_1,\ldots,k_n)\prod_{i=1}^{n}\int_{-\infty}^{\infty}B_{k_i}(x)dx=q_j \quad (23)$$

and $q_j \neq 1.0$ within the desired accuracy, then the following replacement is made to restore the correct normalization.

$$\Gamma^{(j)}(k_1,\cdots,k_n)^{\text{new}} \leftarrow q_j^{-1}\,\Gamma^{(j)}(k_1,\cdots,k_n)^{\text{old}} \tag{24}$$

This normalization check and replacement strategy contributes to producing a very stable and accurate numerical procedure.

4. Examples

The accuracy and power of the developed PIS procedure will be illustrated by application to specific case studies taken from two classes of dynamic models. Both models are described by Eq. (3) with n=2 and m=3. This implies a two-dimensional state space vector $\boldsymbol{X}=(X_1,X_2)^T=(Z,\dot{Z})^T$. Further, $\boldsymbol{m}(\cdot)$ and $\boldsymbol{Q}(\cdot)$ are such that $m_1(X_1,X_2)=X_2$ and $q_{1j}(\cdot)=0$ for j=1,2,3. Assuming sufficient restrictions on $\boldsymbol{m}(\cdot)$ and $\boldsymbol{Q}(\cdot)$, cf. Refs. [6,7], $\boldsymbol{X}(t)$ becomes a Markov diffusion process. Invoking Eq. (13), it can be shown that, up to correction terms of order τ^2, the associated TPD assumes the form

$$p(\boldsymbol{x},\tau\,|\,\boldsymbol{x}',0)=\delta(x_1-x_1'-x_2'\,\tau)\cdot\tilde{p}(x_2,\tau\,|\,\boldsymbol{x}',0). \tag{25}$$

$\tilde{p}(x_2,\tau\,|\,\boldsymbol{x}',0)$ is given by the relation

$$\tilde{p}(x_2,\tau\,|\,\boldsymbol{x}',0)=\frac{1}{\sqrt{2\pi\beta(\boldsymbol{x}')\tau}}\exp\left\{-\frac{(x_2-x_2'-m_2(\boldsymbol{x}')\tau)^2}{2\beta(\boldsymbol{x}')\tau}\right\}, \tag{26}$$

where

$$\beta(\boldsymbol{x}')=\sum_{j=1}^{3} q_{2j}(\boldsymbol{x}')^2. \tag{27}$$

By combining Eqs. (25) and (26), and applying the solution technique described in the previous section, the TPD $p(\boldsymbol{x},t\,|\,\boldsymbol{x}',t')$ for large $t-t'$ can be calculated. By this, the time evolution of the system when it starts from rest, for example, can be studied. The stationary PDF is obtained in the limit as $t-t'\to\infty$. For application of the PIS method to other problems involving both two- and three-dimensional state space vectors, the reader may consult Refs. [11–13,18,19].

4.1 Example 1—The Caughey Oscillator

There is a class of dynamic models for which there exist an analytical solution for the stationary joint PDF of $\boldsymbol{X}$. A member of this class may be called a Caughey oscillator, Ref. [20]. The generic equation of motion for this oscillator can be written as

$$\ddot{Z}+\dot{Z}g(E)+h(Z)=\Gamma N(t). \tag{28}$$

$N(t)$ denotes a stationary, zero-mean Gaussian white noise satisfying $E[N(t)\,N(t+\tau)]=\delta(\tau)$, where $\delta(\cdot)$ denotes Dirac's delta function, Γ is a positive constant and $g(E)$ is a function of the total energy $E=E(Z,\dot{Z})$ given as follows

$$E=\frac{1}{2}\dot{Z}^2+V(Z) \tag{29}$$

where

$$V(z)=\int_0^z h(s)\mathrm{d}s. \tag{30}$$

For this example $m_2(z,\dot{z})=-\dot{z}\,g[E(z,\dot{z})]-h(z)$, and we may choose $q_{21}=q_{22}=W_1=W_2=0$. $q_{23}=\Gamma$ and $\mathrm{d}W_3(t)=N(t)\mathrm{d}t$. The stationary, joint PDF, denoted by $p_s(\cdot)$, is then determined by the relation, cf. Refs. [20,21]

$$p_s(z,\dot{z})=C\exp\{-\frac{2}{\Gamma^2}\int_0^{E'} g(s)\mathrm{d}s\}, \tag{31}$$

where $E'=\dot{z}^2/2+V(z)$, and C is a normalization constant to ensure a total probability equal to 1.0.

For the illustration purposes in this paper, we have chosen the following special case of Eq. (28)

$$\ddot{Z}(t)+2\xi\dot{Z}(t)\{1+\varepsilon[\frac{1}{2}\dot{Z}^2(t)+\frac{1}{2}Z^2(t)+\frac{1}{4}\lambda Z^4(t)]^{1/2}\}+Z(t)$$

$$+\lambda Z^3(t)=2\sqrt{\xi}N(t) \tag{32}$$

with parameters ξ, ε, and λ.

The stationary PDF only depends on the parameters ε and λ, and the numerical solution for the following set of parameter values has been calculated (ε,λ)=(0,0) (Gaussian response), (0, 0.2) (Duffing oscillator) and (0.5, 0.1). The calculations were carried out with the same number of grid points on both axes in state space, aviz., 45. Since the resulting PDFs are actually independent of ξ, the value ξ=0.1 was chosen for the Gaussian and Duffing cases, while ξ=0.5 was adopted for the last case. The time increments used were τ=0.0025 s, 0.001 s, and 0.02 s, respectively. The total CPU time on a DEC station 3100[1] was about 5 minutes for each case. In Figs.

[1]Certain commercial equipment, instruments, or materials are identified in this paper to specify adequately the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

1 and 2 are shown the marginal PDFs of the displacement response for the three case studies considered, together with the corresponding analytical solutions. In Fig. 3 are given the corresponding analytical and numerical results for the mean upcrossing rate. It is seen that in all three cases the agreement between the numerical PIS and the analytical solution is very good over the whole range of probability levels given. In fact, the accuracy can be retained down to much lower probability levels ($\simeq 10^{-10}$) at a moderate increase in computer time.

4.2 Example 2—Parametric and External Excitation

In this example, the response statistics of a nonlinear oscillator subjected to both external and parametric random excitation will be illustrated by applying the methodology of the paper to two specific case studies.

The equation of motion of the oscillator is the following

$$\ddot{Z}+2\xi[1+\tilde{N}_1(t)]\dot{Z}+\gamma[Z^2+\frac{Z^2}{\omega_0^2}]\dot{Z}+\omega_0^2[1+\tilde{N}_2(t)]Z=\tilde{N}_3(t). \tag{33}$$

Here ξ, γ, and ω_0 are positive constants, $\tilde{N}_j(t)$, j=1,2,3, are independent Gaussian white noises satisfying

$$E[\tilde{N}_j(t)\tilde{N}_j(t+\tau)]=\Gamma_j^2\,\delta(\tau),\ j=1,2,3, \tag{34}$$

where Γ_j are positive constants. For this example it is found that $m_2(z,\dot{z})=-2\xi\dot{z}-\gamma[z^2+\dot{z}^2/\omega_0^2]-\omega_0^2 z$, $q_{21}(z,\dot{z})=-2\xi\dot{z}\Gamma_1$, $q_{22}(z,\dot{z})=-\omega_0^2 z\Gamma_2$ and $q_{23}(z,\dot{z})=\Gamma_3$.

This model was studied by Dimentberg [22], who showed that when

$$\omega_0^2\,\Gamma_2^2=4\xi^2\,\Gamma_1^2 \tag{35}$$

a closed-form expression for the stationary joint PDF can be obtained. It is given as

$$p_s(z,\dot{z})=C\,\frac{\exp\{-\mu(z^2+\dot{z}^2/\omega_0^2)\}}{(\kappa+z^2+\dot{z}^2/\omega_0^2)^{\delta-\kappa\mu}}, \tag{36}$$

where C is a normalization constant and

$$\kappa=\frac{\Gamma_3^2}{\omega_0^4\Gamma_2^2},\ \delta=\frac{2\xi}{\omega_0^2\Gamma_2^2}+\frac{1}{2},\ \mu=\frac{\gamma}{\omega_0^2\Gamma_2^2} \tag{37}$$

By this, we have the opportunity to test the accuracy of the PIS method for this kind of dynamic model. The results of two particular cases will be presented.

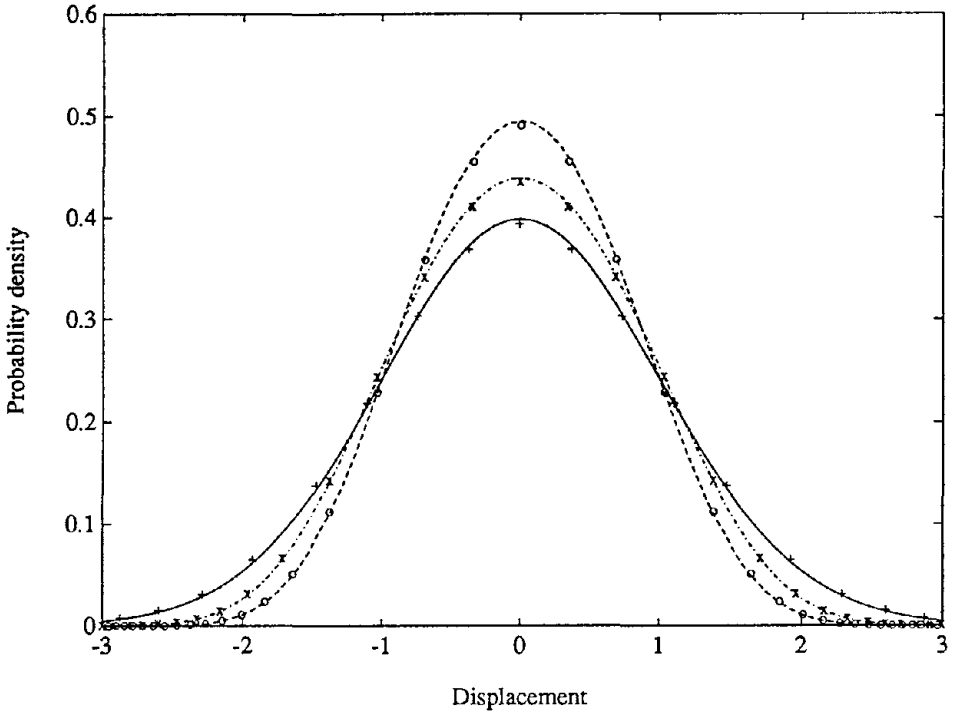


Fig. 1. Probability density function of displacement response for the Caughey oscillator in example 1. Analytical solutions: ———, ε=0, λ=0; – • – •, ε=0, λ=0.2; – – – –, ε=0.5, λ=0.1. Numerical path integral solution: +, ε=0, λ=0; ×, ε=0, λ=0.2; ○, ε=0.5, λ=0.1.

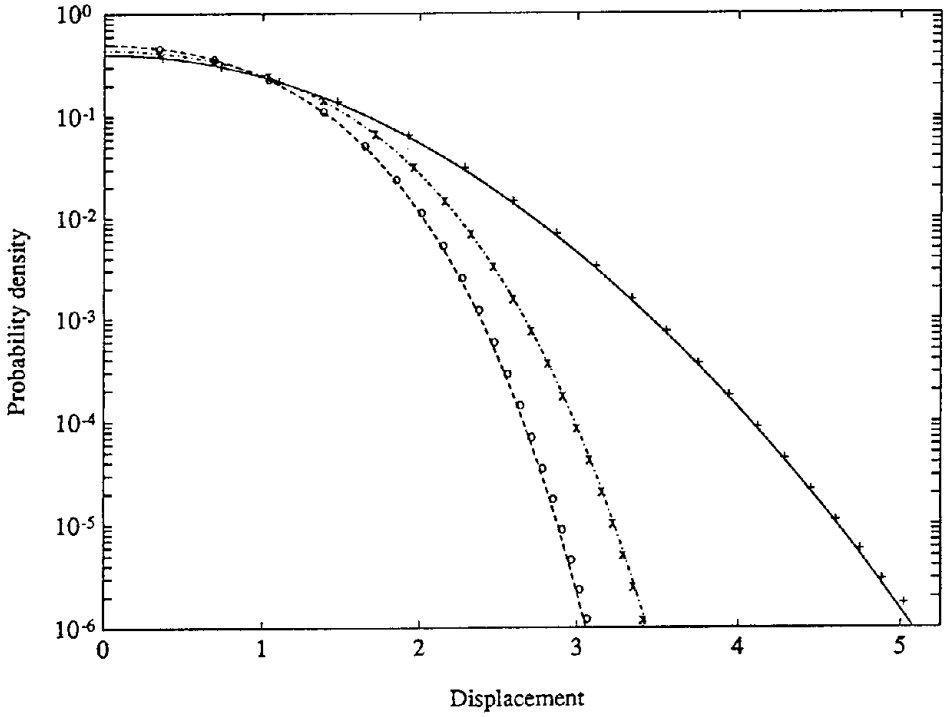


Fig. 2. Logarithmic plot of the probability density function of displacement response for the Caughey oscillator in example 1. Key as in Fig. 1.

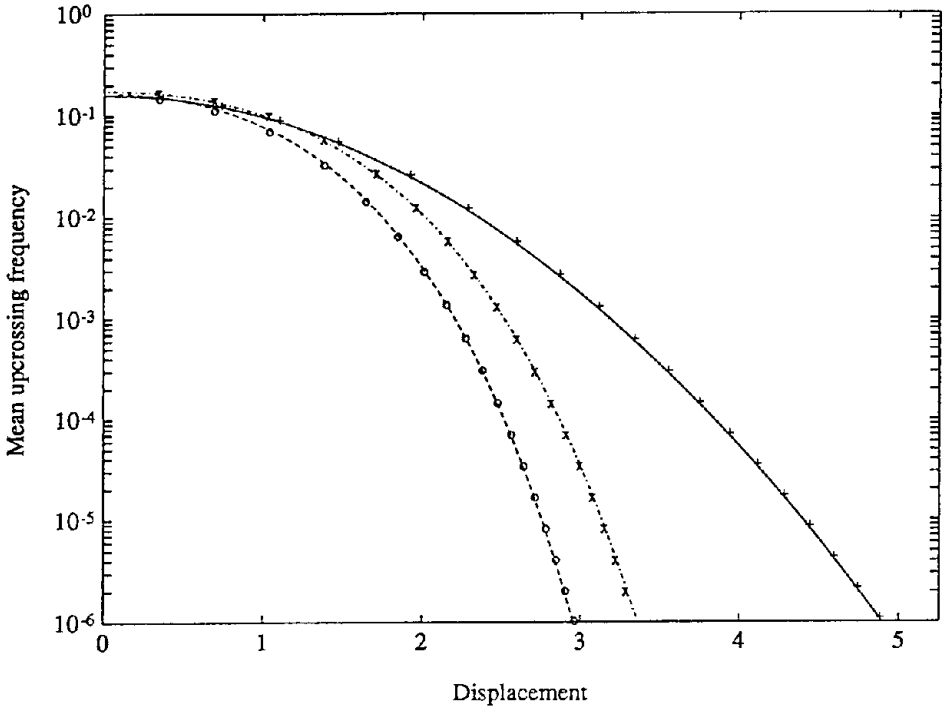


Fig. 3. Mean upcrossing rate of displacement response for the Caughey oscillator in example 1. Key as in Fig. 1.

Case 1: Here the following parameter values were used. ξ=0.1, γ=0.1, ω_0=1.0, Γ_1^2=2.5, Γ_2^2=0.1, Γ_3^2=0.3. For the numerical calculations a grid size of 49×49 points and a time increment τ=0.01 s was used. The total CPU time on a DEC 3100 work station was 3 min for the PIS calculation. The results for the analytical and numerical solutions are given in Figs. 4–6. In Figs. 4 and 5 are shown the marginal PDF of the displacement response and in Fig. 6 is shown the corresponding mean upcrossing rate.

Case 2: In this case the following set of parameters were used. ξ=0.1, γ=0.4, ω_0=1.0, Γ_1^2=5.0, Γ_2^2=0.2, Γ_3^2=0.3. A grid size of 51×51 points together with a time increment τ=0.01 s were chosen. The CPU time was about the same as in the previous case. The same results as for Case 1 are presented in Figs. 4–6.

5. Conclusions

A numerical method for estimating the extreme response of nonlinear oscillators excited by white noise, or filtered white noise, has been described. The example calculations presented show that the method gives very accurate estimates of the required joint PDF. In fact, for every example having analytical solution on which the method has been tested, complete agreement has been found with proper choice of grid size and time increment in the numerical solution procedure. In the present paper, of course, only a few cases can be given. Experience with the method indicates that two-dimensional problems can be solved routinely with high accuracy requiring a few minutes CPU time on a work station (DEC station 3100). The solution of three-dimensional problems requires more care in the sense that computer capacity starts to become an issue of importance. In such cases the CPU time easily runs into hours.

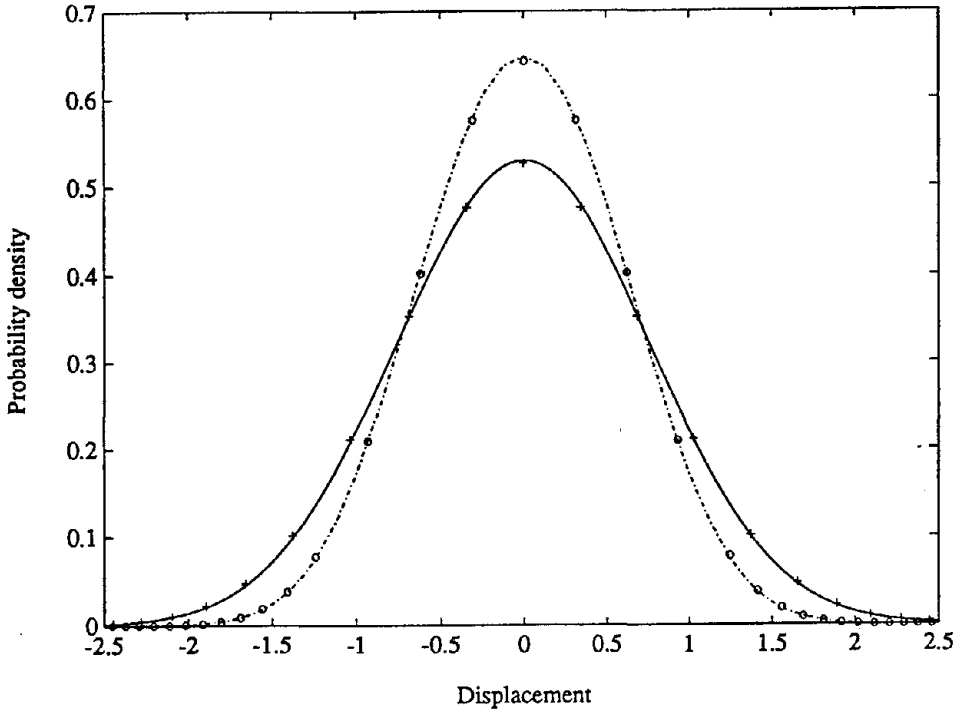


Fig. 4. Probability density function of displacement response for the oscillator in example 2, case 1 and 2. Analytical solutions: ———, case 1; – • – •, case 2. Numerical path integral solution: +, case 1; O, case 2.

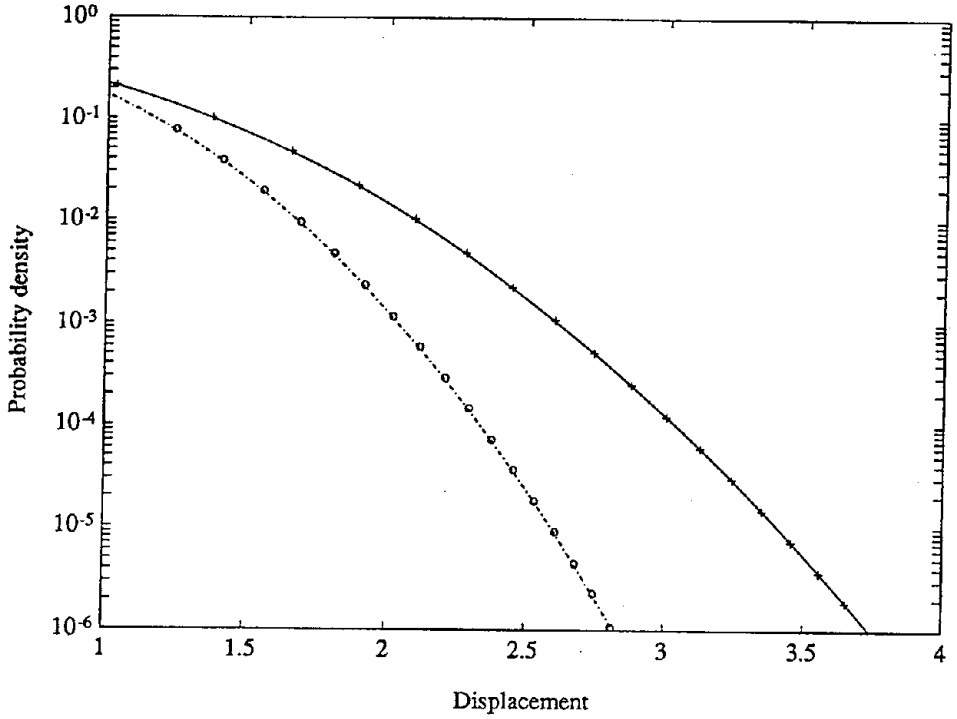


Fig. 5. Logarithmic plot of the probability density function of displacement response for the oscillator in example 2, case 1 and 2. Key as in Fig. 4.

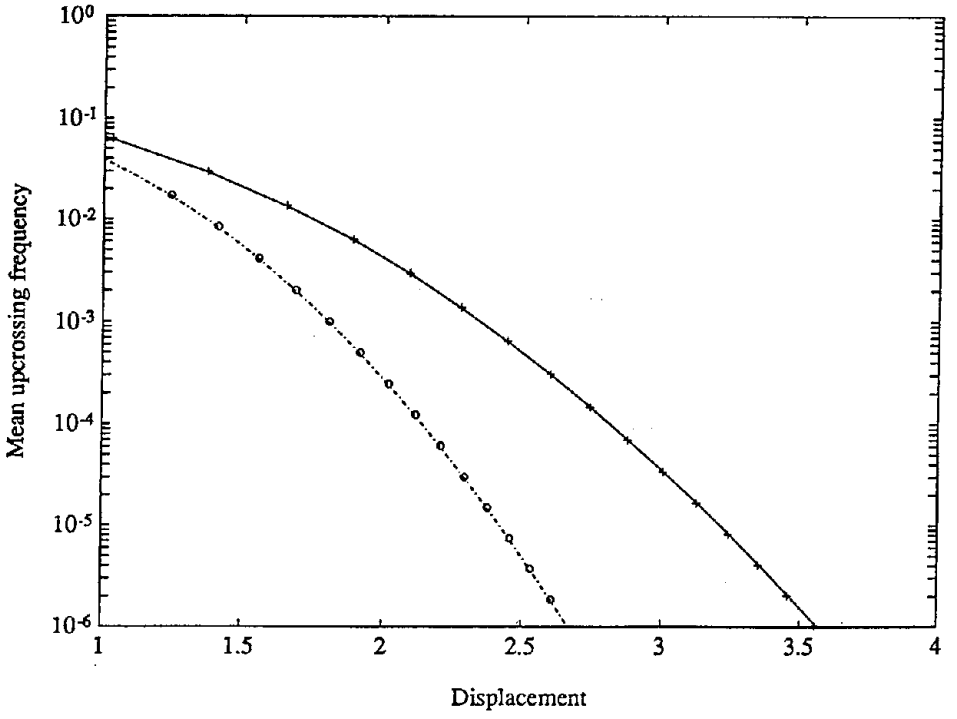


Fig. 6. Mean upcrossing rate of displacement response for the oscillator in example 2, case 1 and 2. Key as in Fig. 4.

6. References

[1] R. Y. Rubinstein, Simulation and the Monte Carlo Method, J. Wiley & Sons, New York (1981).
[2] J. N. Franklin, Numerical simulation of stationary and non-stationary Gaussian random processes, SIAM Review **7** (1), 68–80 (1965).
[3] M. Shinozuka and C.-M. Jan, Digital Simulation of Random Processes and Its Applications, J. Sound Vibration **25** (1), 111–128 (1972).
[4] P. D. Spanos and M. D. Mignolet, ARMA Monte Carlo Simulation in Probabilistic Structural Analysis, Shock Vib. Digest **21**, 3–14 (1989).
[5] F. Yamazaki and M. Shinozuka, Simulation of stochastic fields by statistical preconditioning, J. Struct. Eng., ASCE **116** (2), 268–287 (1990).
[6] E. Wong and B. Hajek, Stochastic Processes in Engineering Systems, Springer-Verlag, New York (1985).
[7] H. Risken, The Fokker-Planck Equation, Second Edition, Springer-Verlag, Berlin (1989).
[8] M. F. Wehner and W. G. Wolfer, Numerical evaluation of path integral solutions to Fokker-planck equations, Phys. Rev. A **27** (5), 2663–2670 (1983).
[9] T. Kapitaniak, Stochastic Response with Bifurcations to Non-Linear Duffing's Oscillator, J. Sound Vibration **102** (3), 440–441 (1985).
[10] A. Naess and J. M. Johnsen, Direct Numerical Simulation of the Response Statistics on Nonlinear Dynamic Systems, Proc. Scandinavian Forum for Stochastic Mechanics, Lund Institute of Technology, Lund, Sweden, August 1990.
[11] A. Naess and J. M. Johnsen, Response Statistics of Nonlinear Dynamic Systems by Path Integration, Proc. IUTAM Symposium on Nonlinear Stochastic Mechanics, Turin, Italy, July 1991.
[12] J. M. Johnsen and A. Naess, Time Variant Wave Drift Damping and its Effect on the Response Statistics of Moored Offshore Structures, Proc. 1st Int. Conference on Offshore and Polar Engineering (ISOPE-91), Edinburgh, August 1991.
[13] A. Naess and J. M. Johnsen, The Path Integral Solution Technique applied to the Random Vibration of Hysteretic Systems, Proc. 1st Int. Conf. on Computational Stochastic Mechanics, Corfu, Greece, September 1991, Computational Mechanics Publications, 1991.
[14] E. H. Vanmarcke, On the distribution of the first-passage time for normal stationary random processes, J. Appl. Mech. **42**, 215–220 (1975).
[15] J. B. Roberts, First passage time for the envelope of a randomly excited linear oscillator, J. Sound Vibration **46**, 1–14 (1976).
[16] A. Naess, Approximate first-passage and extremes of narrow-band Gaussian and non-Gaussian random vibrations, J. Sound Vibration **138**, 365–380 (1990).
[17] C. de Boor, A Practical Guide to Splines, Springer-Verlag, New York (1978).
[18] A. Naess and J. M. Johnsen, Response statistics of nonlinear, compliant offshore structures by the path integral solution method, Probabilistic Eng. Mech. **8**, 91–106 (1993).
[19] A. Naess and B. K. Hegstad, Response statistics of van der Pol oscillators excited by white noise, Nonlinear Dynamics **5**, 287–297 (1994).
[20] T. K. Caughey, Nonlinear theory of random vibrations, in Advances in Applied Mechanics, Academic Press, New York (1971).
[21] Y. K. Lin, Probabilistic Theory of Structural Dynamics, McGraw-Hill, New York (1967).
[22] M. F. Dimentberg, An exact solution to a certain non-linear random vibration problem, Int. J. Non-Linear Mech. **17** (4), 231–236 (1982).

About the author: *Arvid Naess is Professor of Structural Engineering on the Faculty of Civil Engineering of The Norwegian Institute of Technology in Trondheim, Norway.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 475 (1994)]

A Random Field Excursion Model of Salt Induced Concrete Delamination

Volume 99 | Number 4 | July–August 1994

Gordon A. Fenton

Technical University of Nova Scotia,
Halifax, Nova Scotia, B3J 2X4

Salt induced concrete delamination is a problem often encountered in parking garage slabs in northern climates where deicing salts are heavily used. A random field model of the delamination process is investigated and compared against some field data. Delaminated regions of the slab are modeled as excursions of a random field above a prescribed threshold and the growth of these regions with time is obtained by allowing the threshold to fall as a function of time. Simulation based excursion statistics are used to obtain the mean and variability of various aspects of the delamination process using this model.

Key words: concrete degradation; delamination; level excursions; random fields; rehabilitation; reliability.

Accepted: March 22, 1994

1. Introduction

The delamination and spalling of concrete surfaces in parking structures in northern climates is an ongoing and expensive problem. The Transportation Research Board [1] estimates that between 50 and 150 parking structures in the Northeast and Midwest United States will need to be rehabilitated each year for the next 10 years at an average cost of $1 million per structure. Without proper design and/or maintenance, deicing salts brought in by vehicles from the roadways are deposited on the concrete surface along with water. Chloride ions gradually penetrate the concrete and electro-chemical processes lead to corrosion of the reinforcing steel. This results in both a degradation of structural integrity and, since the corrosion products occupy considerably more volume than the original steel, delamination or spalling of the concrete surface and loss of utility.

This paper is aimed at developing a tool to aid in a rational probabilistic approach to the rehabilitation of parking structures. Such an approach allows the optimal allocation of limited resources to this ongoing and rather expensive maintenance issue. A simple stochastic model involving only a few parameters is used herein to represent what is known to be a complex phenomenon. The following factors are suggested by Public Works Canada [2] to have the largest effect on the onset of concrete delamination;

- *chloride ion input:* quantities of deicing salts used,
- *concrete permeability:* influenced in turn by water/cement ratio, intensity and frequency of cracking, surface coatings/sealers and construction practices,
- *ambient temperature, humidity, precipitation*
- *concrete cover depth*
- *pH of local aggregates*
- *conductivity:* wetness of concrete.

It is immediately recognized that virtually all of these factors are highly variable from structure to

structure and within a single structure from point to point. As well, on a practical basis for an existing structure, some of the factors are unknowable except through extensive destructive testing. Clearly a representative model should not require experimental validation on a per structure basis, where the experimentation may be more expensive than the final repairs. This, in fact, is the primary motivation for the use of stochastic models.

In a different approach to the same problem, Attwood et al. [3] develop a limit state function for parking structures based on a critical fraction of floor area delamination, D_{cr},

$$m = D_{cr} - \alpha S(t - t_I), \quad t > t_I, \tag{1}$$

where $m \leq 0$ denotes the "failure" state, S is the annual delamination rate, α is a correction factor accounting for ambient temperature and crack widths, t is time in years, and t_I is the time to initiation of delamination (also in years). Note that the second term is taken to be zero for all $t \leq t_I$. Attwood et al. employ a First-Order Reliability Method (FORM) to estimate reliabilities associated with the delamination of a parking garage structure. In their approach, all of the factors appearing in Eq. (1) are expressed in terms of random variables having assumed distributions and the joint cumulative probabilities are evaluated using FORM.

In this paper a 2-dimensional random field model representing the spatial delamination process over time is investigated. The random field may be loosely interpreted as the out-of-plane stress field at the reinforcement level (which changes randomly from point to point over the area of the slab). Delaminated regions are represented by excursions of the random field above some threshold which can be thought of as a critical concrete resistance to horizontal splitting. In fact the stress field itself will not be considered in that it is unmeasurable in practice. Rather, the field excursions will be used to represent the delaminated regions directly, since these regions are measurable to some extent. A simple random field model is adopted whose primary motivation is to attempt to shed light on the following questions;

1) what is the mean and variability of the total area of delamination as a function of time?

2) what is the average size of individual delaminated regions as a function of time?

3) how many delaminated regions can be expected in a slab?

These are essentially questions regarding the statistics of excursion regions and so some simulation results regarding excursion statistics will be presented in the next section.

Figure 1 illustrates an example of excursions of a random field above some predefined threshold. In the context of this paper, the dark regions can be viewed as areas where concrete delamination has occurred at some fixed time. As time progresses, delaminated regions are expected to grow in size, corresponding to a falling threshold level. A falling threshold is equivalent to a rising mean, in the case of a homogeneous field. Since the excursion statistics in the next section are developed as functions of the threshold, the falling threshold interpretation is used here rather than a rising mean. In either case, a nonhomogeneous field could be employed, if the data so indicated, by considerably extending the simulation based study (in the absence of analytical results). In this preliminary investigation, only a homogeneous field is considered.

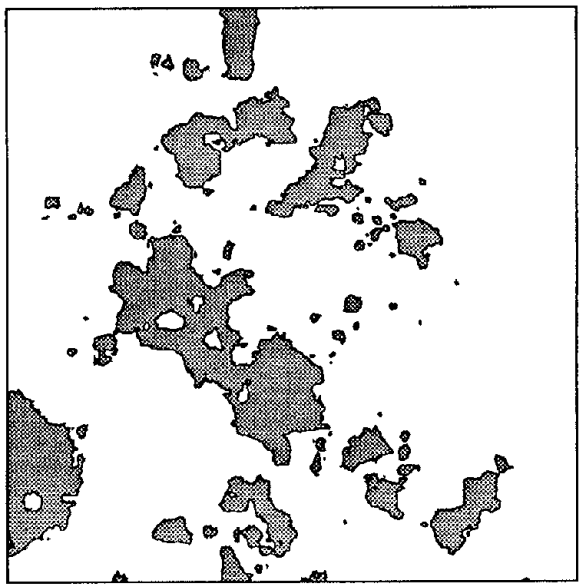


Fig. 1. Excursions of a two-dimensional random field above threshold $b\sigma$.

To represent the delamination process, excursions of an isotropic Gaussian random field having mean zero and unit variance will be used. Although other values of mean and variance are possible, the excursion statistics are dependent purely on the distance between the mean and the threshold. This distance can be expressed in units of σ, the standard deviation, so that there is no advantage in choosing anything other than a mean zero, unit

variance field. The quantity $b\sigma$ will be referred to henceforth as the physical threshold and b alone as the threshold. The choice of an isotropic Gaussian process has been made largely for simplicity, there being little evidence available to clearly justify other types of random functions.

2. Excursion Statistics in Two Dimensions

With respect to excursion statistics, such as the mean number and area of isolated excursions, analytical results developed to date are asymptotic in nature, accurate only at very high thresholds where the excursion process approaches a Poisson point process. Often in engineering problems the interest is in thresholds which are quite a bit lower, such as the delamination process considered herein. In order for the proposed random field model to be useful in this context, excursion statistics should be available. This section summarizes a study in which excursion statistics are obtained through Monte Carlo simulation. Specifically, an ensemble of 2000 realizations of an isotropic zero mean, unit variance Gaussian random field, $Z(x)$, with Markovian covariance function,

$$B(\tau)=\sigma^2\rho(\tau)=\sigma^2\exp\left\{-\frac{2}{\theta}|\tau|\right\}, \tag{2}$$

are produced using the Local Average Subdivision (LAS) method [4, 5], where τ is the lag vector and θ is the scale of fluctuation. The scale of fluctuation is loosely interpreted as the distance over which correlation is significant. Since many of the statistics of interest depend strongly on the scale of fluctuation, 2000 realizations were generated at each of 5 different scales of fluctuation.

Individual realizations are decomposed into excursion regions and "holes," using a space-filling algorithm, over a range of thresholds $b=[-4, 4]$. The mean and variance statistics of the excursions are estimated over the ensemble. It should again be emphasized that b is measured in units of standard deviation, for example $b=2$ implies a threshold at two standard deviations from the mean. For a unit variance field there is no distinction between the value of b and the physical threshold level. However the local averaging performed by the LAS method results in a slight decrease of the variance of the discretized field, as dictated by local averaging theory. In that each field is represented as a discrete lattice of 128×128 "cells," the variance of each cell varies from 0.971 at the smallest scale of fluctuation considered to 0.999 at the largest. The distinction between cell variance and point variance will be ignored in this paper, although the results presented in the various plots to follow are accurate in this respect. For a more rigorous treatment of this issue, see Ref. [6].

Within a given domain $V=[0, L_1]\times[0, L_2]$ of area $A_T=L_1 L_2$, the total excursion area per unit area, D_b, where the process $Z(x)$ exceeds some threshold, can be defined by

$$D_b=\frac{1}{A_T}\int_V I\,(Z(x)-b\sigma)\,dx, \tag{3}$$

where $b\sigma$ is the physical threshold of interest, σ^2 being the variance of the process, and $I(\cdot)$ is the indicator function defined on V (taken to be zero outside the domain V)

$$I(t)=\begin{cases}1 & \text{if } t\geq 0\\ 0 & \text{if } t<0\end{cases}. \tag{4}$$

For a homogeneous process, the expected value of D_b is simply

$$E[D_b]=P[Z\geq b\sigma], \tag{5}$$

which, for a zero-mean Gaussian process yields

$$E[D_b]=1-\Phi(b), \tag{6}$$

where Φ is the standard normal distribution function. The estimate of $E[D_b]$, denoted m_{D_b}, derived using the simulation results is shown in Fig. 2 and is in complete agreement with Eq. (6). Figure 2 (b) shows the estimated standard deviation of D_b denoted s_{D_b}. Note that while $E[D_b]$ is independent of θ, its variance is not. In keeping with the practice of normalizing all results, the scale of fluctuation has been normalized with respect to $L=\sqrt{A_T}$ in these plots. Note also that the horizontal threshold axis values decrease to the right—this is because the threshold axis is associated with time in the next section and time increases to the right as usual.

Figure 3 shows the estimated mean and standard deviation of the number of isolated excursion regions, N_b, denoted m_{N_b} and s_{N_b} respectively, as a function of scale and threshold. Although not readily apparent from the plot, the limiting value of m_{N_b} at the right edge of the plot ($b\to-\infty$) is 1. In other words, a single excursion exists over the entire domain for very low thresholds. At the other extreme ($b\to\infty$), the limiting value is zero as expected. Although the roughness of the estimate s_{N_b}, as seen in Fig. 3(b), is as yet unexplained, it appears unlikely that it arises from statistical uncertainty in the

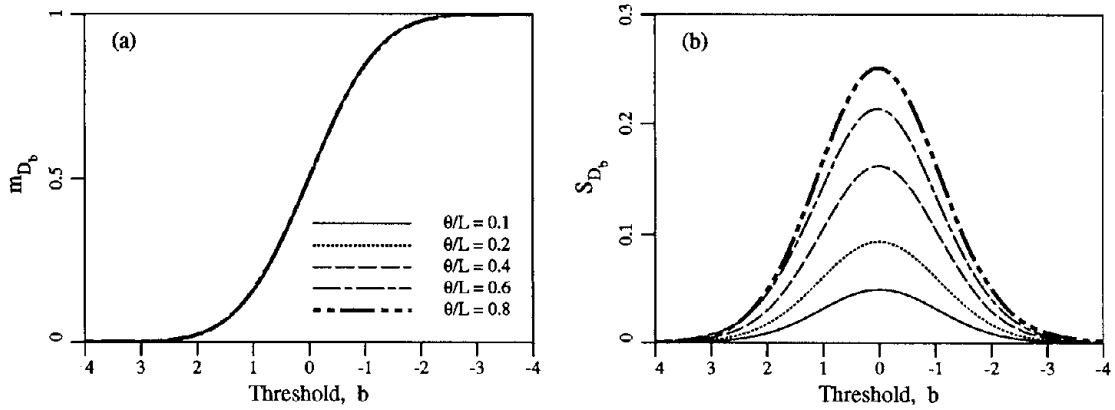


Fig. 2. Estimated statistics of total excursion area per unit area, D_b,: a) mean, b) standard deviation.

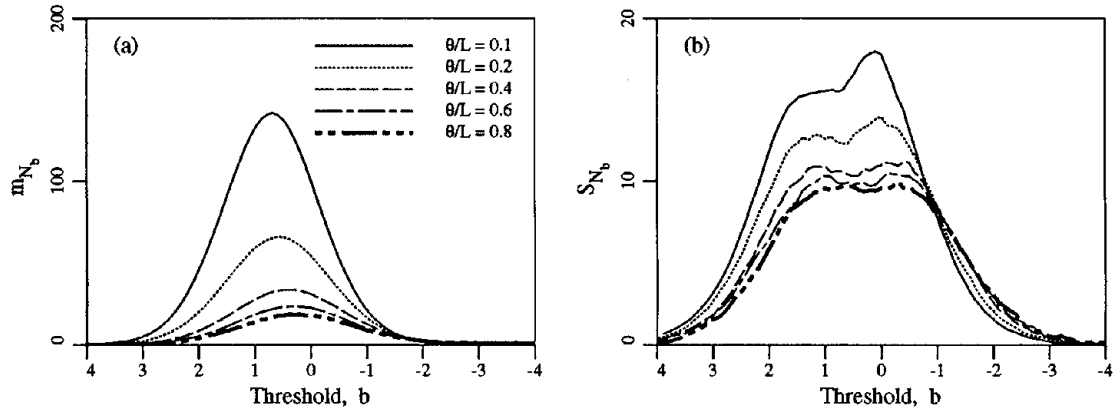


Fig. 3. Estimated statistics of number of isolated excursions, N_b: a) mean, b) standard deviation.

estimation procedure. Based on a sample size of 2000 realizations, a 90 percent confidence bound on the standard deviation of N_b is about $\pm 0.025 s_{N_b}$ which is smaller than the size of most of the "bumps" seen in Fig. 3(b).

Within a given realization, the average area of an isolated excursion per unit domain area, D_e, can be obtained using the number of excursions,

$$D_e = \frac{D_b}{N_b}.$$

Since D_b is the sum of the N_b isolated excursion areas, the expected value of D_e is just

$$E[D_e] = \frac{E[D_b]}{E[N_b]}$$

Figure 4(a) illustrates this result using the estimated mean value of N_b shown in Fig. 3(a). The estimated standard deviation of D_e shown in Fig. 4(b) is derived using the assumption that the sizes of isolated excursions are independent. While this is true from realization to realization, it is clearly not true within a single realization. Thus Fig. 4(b) can only be considered to be a rough indication of the true variability of the area of isolated excursions.

3. Calibrating the Model

There are essentially only two parameters in the random field model considered in this study. These are the scale of fluctuation, θ, and the threshold level, b. Although there is little published experimental evidence to allow a clear statement of what these parameters should be for a given structure, some preliminary estimates are possible. First the relationship between the threshold level and the age of the structure can be obtained by fitting delamination versus age data collected by Trow [7] and presented in Ref. [3] to the curve of Fig. 2(a).

A good fit was obtained using least squares regression ($r^2=0.93$) on the linear relationship $b = 3.17-0.225t$. To obtain this relationship, the total fraction of delaminated area observed by Trow, D_b, was plotted against $b = \Phi^{-1}(1-D_b)$ and a best fit line obtained. Figure 5(a) shows the regression results and Fig. 5(b) shows where the observed delamination results would appear on Fig. 2. To expand the scale, only positive thresholds are shown in Fig. 5(b). The choice of function relating time and threshold is largely arbitrary as long as $b(t)$ is a decreasing function over all times of interest (assuming delaminated areas cannot "heal"). While a quadratic gives a slightly better fit to the raw data, it violates this principle and so cannot be used to extrapolate.

Some additional unpublished data was made available to the author by Public Works Canada based on a survey of a single parking garage structure in Ontario, Canada. Figure 6 illustrates this data. The ± 1 standard deviation curves are obtained using Fig. 2(b) with $\theta/L = 0.1$. Clearly, for such a choice in θ, the variability in D_b predicted by the random field excursion model underestimates the variability in the observations. The additional variance arises because the time-threshold relationship $b(t)$ is itself a function of random coefficients, as implied by the regression analysis. Alternatively and equivalently, the additional variability could be ascribed to the fact that the delamination field is not homogeneous on such a scale. When considered at the scale of a typical bay, some bays show much

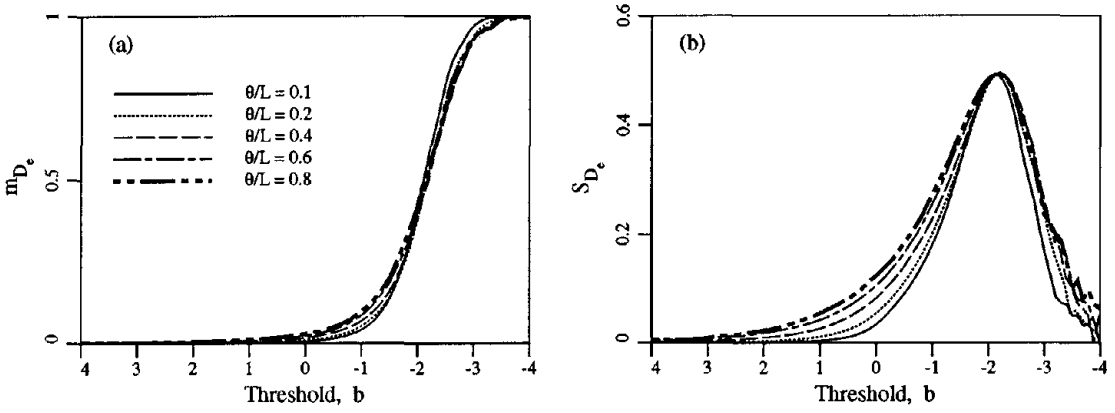


Fig. 4. Estimated statistics of isolated excursion areas per unit area, D_e,: a) mean, b) standard deviation.

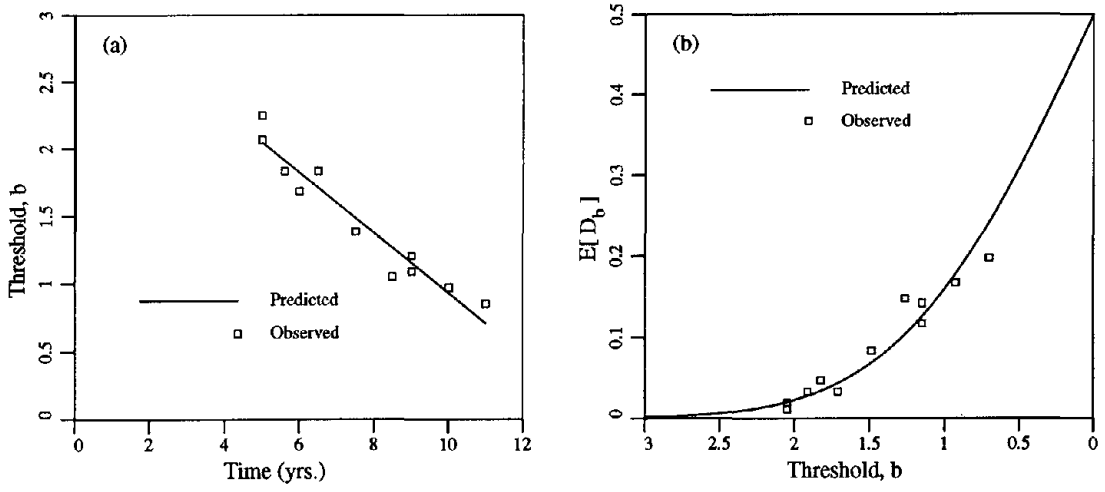


Fig. 5. Regression results using Trow's delamination rate data: a) regression, b) observed versus predicted D_b.

higher delamination rates than others, perhaps corresponding to smaller mean cover depths and/or higher Cl^- inputs. Note also that larger values of θ/L lead to higher standard deviations in D_b so that the variability seen in Fig. 6 could also be explained by larger values of θ. However the choice of θ should not be based on the variability in D_b, rather this observation indicates that perhaps the Gauss-Markov covariance model is not appropriate.

In this study, $b(t)$ will be treated as a deterministic function which corresponds to the choice of a homogeneous random field. It is beyond the scope of this initial investigation to consider estimating the parameters of a nonhomogeneous (or self-similar) field to represent the delamination process generally, although this appears indicated. Since plan views of the delaminations are available from Public Works for individual bays of approximately 5 m × 5 m in extent ($L = 5$), attention will be restricted to data on such a scale over which the field can be considered homogeneous. Choosing one such area from the Public Works data, the measured delamination fraction as a function of time is shown in Fig. 7. In this case the line of best fit was found to be

$$b = 4.13 - 0.172t \tag{7}$$

with $r^2 = 0.86$. Notice that for such a case, the variability in the observations is substantially reduced

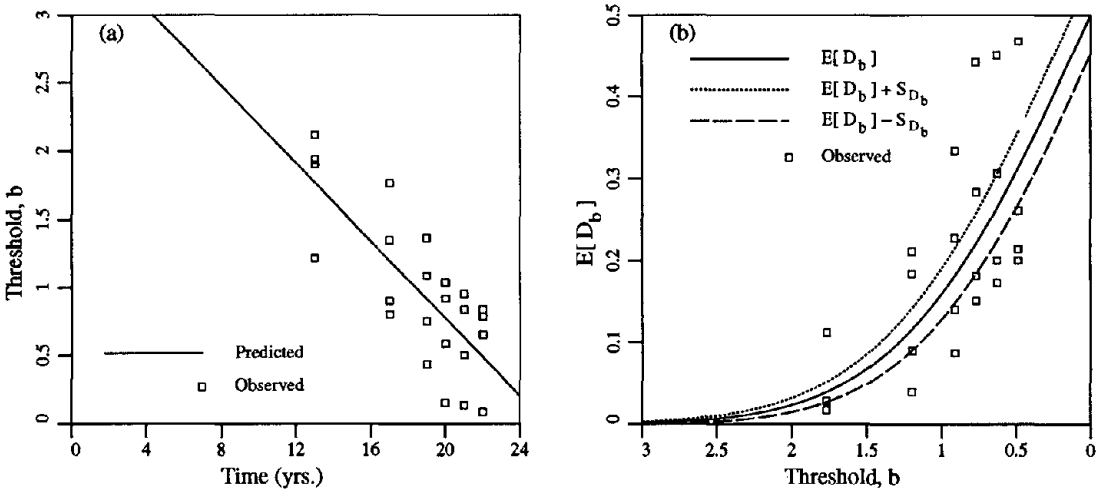


Fig. 6. Delamination rate data provided by Public Works Canada. The fitted line is $b = 3.61 - 0.142t$ with $r^2 = 0.6$.

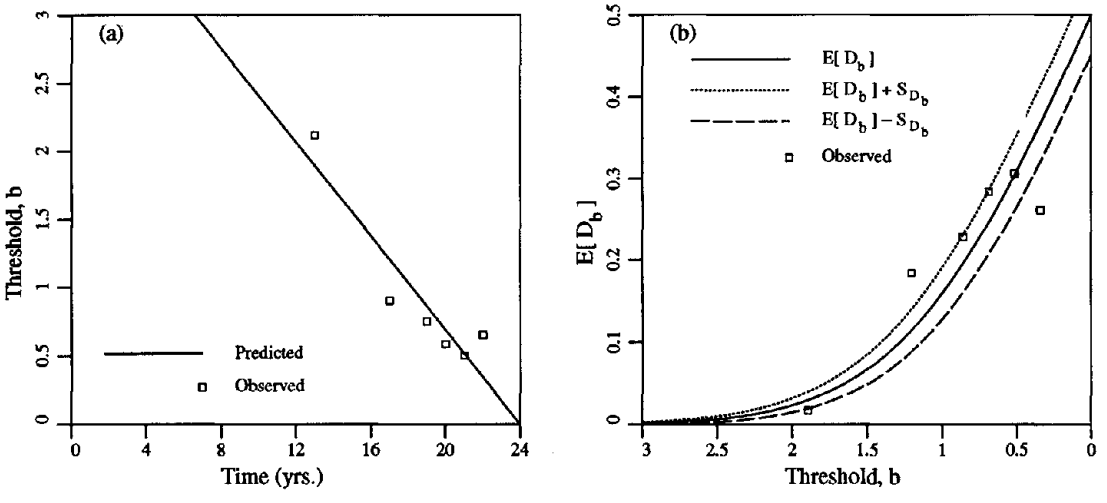


Fig. 7. Delimination rate provided by Public Works Canada over a single slab region.

and that 2/3 of the observations lie on or within the ± 1 standard deviation curves for $\theta/L = 0.1$ (see also Fig. 5). This result appears encouraging although it is recognized that it could be due in part to the reduced number of samples, even though similar results were found for most other bays.

Before considering the estimation of θ, it is worth pointing out a further difference between the random excursion model and the commonly accepted delamination model. Salt induced reinforcement corrosion is generally believed to involve two stages (see Ref. [8]): 1) an initiation phase during which the alkalinity of the concrete surrounding the reinforcement (which renders the steel passive) is reduced by the migrating Cl^- ions, and 2) an active phase during which corrosion takes place. Attwood et al. [3] estimates the initial phase to last 4.7 years. During this phase no corrosion is assumed to take place. Note that such a model can only be applied to points in the slab where the concrete is in contact with the reinforcement.

In contrast with this two stage model, the random field excursion model admits some probability of delamination even at time $t = 0$. Using Eq. (7), one obtains $b = 4.13$ at time $t = 0$, so that the expected total delamination area per unit area is 1.7×10^{-5} or about 17 mm^2/m^2 of slab. At these levels the precise definition of delamination comes into question. If it is strictly interpreted as a loss of bond between the concrete and reinforcement then this result is not unreasonable given the presence of cracks, voids and the initial state of the reinforcement. At time $t = 5$ years, the expected total delamination area is still only about 6 cm^2/m^2 of slab, a level which is probably still largely undetectable at the surface of the slab and presumably would correspond to corrosion in the immediate neighborhood of surface cracks. Although it is believed that cracks do not contribute significantly to the areal delamination process [8], it is not unreasonable to expect that they can be initiators of the corrosion process at discrete points in the slab. If this is the case, then the many year delay before the onset of observable levels of delamination implies that the corrosion growth should be quite slow at first. The results predicted by Fig. 2(a) are in basic agreement with this in that $E[D_b]$ grows very slowly for b decreasing to about 2 [corresponding to $t < 12$ years using Eq. (7)].

Turning now to the estimation of the scale of fluctuation, θ, it becomes apparent that this task is complicated by the type and quality of data available. Ideally, one would take measurements of the corrosion induced stress field over a number of structures, estimate a spatial covariance structure and from this obtain θ. Even if this approach were possible, the nonhomogenieties mentioned above would make it difficult. However, in general the stress field is unmeasurable and what little data is available generally consists of surveys giving the spatial extent of delaminated regions. Figure 8(a) illustrates such a survey while Fig. 8(b) is a realization of the random field excursions using $\theta = 0.5$ m. Once the time-threshold relationship has been established, and for the purposes of this argument Eq. (7) will be used, a possible technique of estimating θ would be to count the average number of excursions and enter Fig. 3(a) at the appropriate threshold to estimate θ/L. Purely on the basis of Fig. 8(a) this yields estimates of $\theta = 2$ to 4 m ($\theta/L = 0.4$ to 0.8). However realizations at this scale yields excursions which are generally far too well connected as shown in Fig. 9. Realizations at such large scales appear like large land masses with many small islands close to shore. Figure 8(a) has "islands" that are more uniform in size and distribution implying a smaller scale of fluctuation. In Fig. 8(b), produced using $\theta = 0.5$, the larger islands are of similar size to the delaminations seen in Fig. 8(a). This along with arguments to follow supports the choice of a smaller scale of fluctuation.

While one cannot expect the plots in Fig. 8 to be identical since they are both independent realizations of a random process, any more than one could expect the pattern of delaminations in another building to be identical, a number of points can be made about the two plots;

1) there is no apparent spatial orientation of the delamination regions in the observations of Fig. 8(a), indicating that the assumption of isotropy is acceptable, at least for this case.

2) about half of the excursions in the random field model (Fig. 8b) are of very small extent. This fraction increases at larger scales. On the other hand, in Fig. 8(a) there are only very few "small' areas appearing in the later survey. It seems reasonable to suspect that additional small delamination regions are in fact occurring in the real slab but that the chain-drag surveying technique is unable to resolve them. Operator bias will almost certainly also be present due to the prior knowledge of existing delamination areas.

3) the random field model is much "rougher" than the observed delamination plots. Again this is likely a problem with the ability of the chain-drag survey to resolve detail.

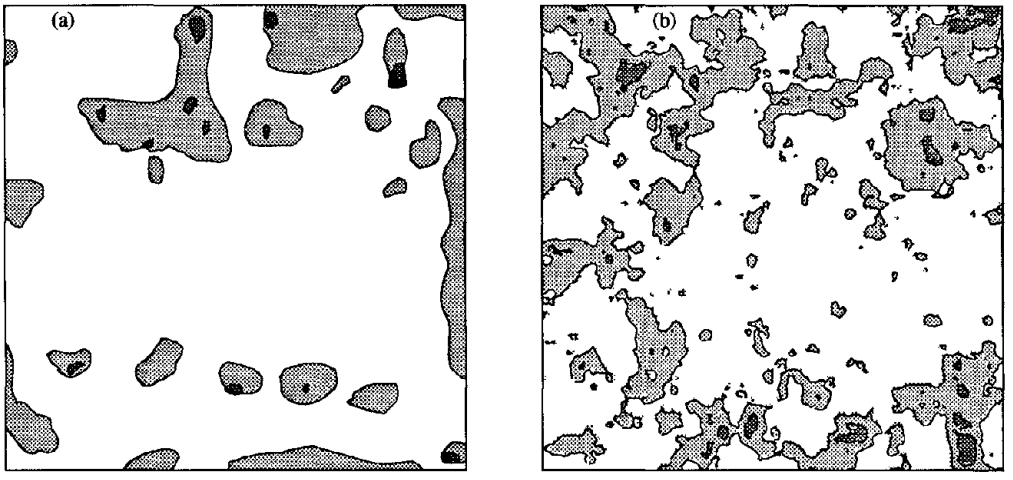


Fig. 8. a) Observed delamination regions at ages $t = 13$ years (dark grey) and $t = 22$ years (light grey) on a 5 m × 5 m portion of a parking garage slab. b) Excursions of a random field above $b = 0.34 (t = 13)$ and $b = 19 (t = 22)$ using $\theta = 0.5$ and Eq. (7).

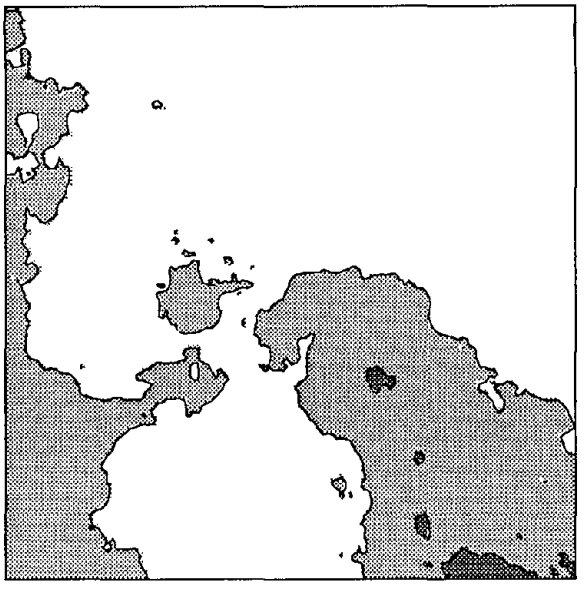


Fig. 9. A realization of excursions at $t = 13$ years (dark grey) and at $t = 22$ years (light grey) using $\theta = 4$ m on a 5 m × 5 m field.

The last two points illustrates the difficulty in estimating θ on the basis of delamination surveys. The chain-drag method depends on setting up reverberations in the delaminated concrete. For delamination details below a certain size, the frequency shifts are undetectable to even the most sensitive human ears, rendering them unnoticeable. Thus many of the delamination details on which an estimation of θ depend are unavailable using current surveying techniques. To some extent the deficiencies in the chain-drag procedure could be accommodated simply by introducing more local averaging in the random field model—effectively smoothing the field. While such a "correction" would not likely result in an improved delamination model, it may allow improved estimates of θ. For the purposes of this investigation, the realization based estimate of $\theta = 0.5$ m is used.

4. Discussion

Conceptually the model proposed herein is quite attractive in that the delamination process is indeed a threshold excursion process in two dimensions. Once the details of the model have been established (type of distribution, time-threshold relationship and scale of fluctuation) and some of the properties of threshold excursions in two dimensions have been determined analytically or via simulation, the model can be used in a reliability context. For example, using Eq. (7), at time $t = 12$ years ($b = 2.07$) Fig. 2 along with Eq. (6) indicates that $m_{D_b} = 0.0194$ and $s_{D_b} = 0.0080$, using $\theta = 0.5$ m. If D_b is assumed to have a Beta distribution, which is properly bounded between 0 and 1, then the probability that D_b is less than 3 % at 4 years is given by

$$P[D_b \leq 0.03] = 0.90 \,. \tag{8}$$

In turn, if a target reliability of 0.9 were chosen against delamination in excess of 3 %, then

inspection of the garage would be recommended at $t = 12$ years for this structure.

Because the model includes the spatial aspects of delamination, it can be used to evaluate testing and mapping procedures. For example, at time $t = 12$ years one could expect about 42 isolated excursions on a 5 m × 5 m slab ($\theta = 0.5$ m) with a standard deviation of about 11. The area of each isolated excursion would average about 0.011 m^2 with a standard deviation of about 0.022 m^2. Surveys which yield results considerably different than these may require verification using other techniques. In addition, the spatial description of delamination could be used in a structural reliability study. For this the simulation approach could be employed to yield a measure of the degree of clustering of the delamination regions (see Ref. [6]).

One recognizes that the excursion model is attempting to predict the cracked state of a concrete slab. In that internal cracks are exceedingly difficult to map, even in controlled laboratory conditions, the model is to some extent intuitive and will likely remain so until improved surveying techniques are developed. Nevertheless the model demonstrates some promising features and can be used as a powerful reliability tool when its parameters are clearly defined in terms of additional data and simulation studies. In particular, the data shown in Fig. 6 indicates that perhaps alternative correlation functions [see Eq. (2)] should be studied – multiple-scale or self-similar type random fields are suggested, showing the small scale behaviour over small regions while reflecting also the large scale, slower variations over larger domains. Also the fact that the model allows delamination to occur at time $t = 0$ (or before) implies that some thought should be given to the assumption of a Gaussian random field and/or the time-threshold relationship. In the interim, however, the choice of a Gaussian field and linear time-threshold relationship leads to results which appear reasonable for any time $t > 0$.

Acknowledgments

The financial support of the Natural Sciences and Engineering Research Council of Canada under Grant OPG0105445 is gratefully acknowledged. The author also extends thanks to Mr. Brian Kyle of Public Works Canada, Ottawa, who kindly provided the results of several delamination surveys for use in this study. Any opinions, findings, and conclusions or recommendations are those of the author and do not necessarily reflect the views of the aforementioned organizations.

5. References

[1] Transportation Research Board, Highway Deicing, Special Report 235, Washington, DC (1991).
[2] Public Works Canada, Development of Reliability-Based Methods for Cost Effective Maintenance of Parking Structures, Technical Report, Ottawa (1989).
[3] D. Attwood, M. A. Nessim, A. Ghoneim, A. Cormeau, and M. S. Cheung, Application of reliability theory to in-service monitoring and maintenance of parking garages, Can. J. Civ. Eng. **18**, 781–788 (1991).
[4] G. A. Fenton, Simulation and Analysis of Random Fields, Ph.D. Thesis, Princeton University, Princeton, New Jersey (1990).
[5] G. A. Fenton and E. H. Vanmarcke, Simulation of Random Fields via Local Average Subdivision, ASCE J. Eng. Mech. **116**(8), 1733–1749 (1990).
[6] G. A. Fenton and E. H. Vanmarcke, Simulation-based excursion statistics, ASCE J. Eng. Mech. **118**(6), 1129–1145 (1992).
[7] Trow Ltd. Consulting Engineers, Parking Structure Deterioration: A Survey and Analysis of its Extent and Influencing Factors, Vol. I Report, Toronto (1984).
[8] A. W. Beeby, Cracking, cover, and corrosion of reinforcement, Concrete International **5**(2), 35–40 (1983).

About the author: Gordon A. Fenton is a civil engineer in the Applied Mathematics Department of the Technical University of Nova Scotia.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 485 (1994)]

Extreme Value Theory Applications to Space Radiation Damage Assessment in Satellite Microelectronics

Volume 99 Number 4 July–August 1994

P. W. Marshall

Naval Research Laboratory,
Washington, DC 20375 and
SFA, Inc.,
Landover, MD 20785

C. J. Dale

SFA, Inc.,
Landover, MD 20785

and

E. A. Burke
Woburn, MA 01801

Calculations of the first and second moments of displacement damage energy distributions from elastic collisions and from nuclear reactions, at proton energies ranging from 10 MeV to 300 MeV, are incorporated into a model describing the probability of damage as a function of the proton fluence and the size of the sensitive micro-volume in Si. Comparisons between the predicted and measured leakage currents in Si imaging arrays illustrate how the Poisson distribution of higher energy nuclear reaction recoils affects the pixel-to-pixel variance in the damage across the array for proton exposures equivalent to mission duration of a few years within the earth's trapped proton belts. Extreme value statistics (EVS) quantify the largest expected damage extremes following a given proton fluence, and an analysis derived from the first-principle damage calculations shows excellent agreement with the measured extremes. EVS is also used to demonstrate the presence of high dark current pixels, or "spikes," which occur from different mechanisms. Different sources of spikes were seen in two different imager designs.

Key words: charge coupled devices; charge injection devices; displacement damage; extreme value statistics; microdosimetry; satellite imagery; satellite microelectronics.

Accepted: March 22, 1994

1. Introduction

Proton-induced displacement damage degrades semiconductor electrical properties by introducing localized energy states within the band-gap which result in increased generation dark current, carrier recombination and charge trapping. On *average*, the permanent proton-induced damage in bulk Si is proportional to the average amount of energy which has been imparted through non-ionizing processes following elastic and inelastic scattering of Si atoms [1-3]. However, on micro-volume scales appropriate for microelectronics, average damage is a poor indicator of damage effects because of differences in the number of incident particles and fluctuations in energy deposition which are an unavoidable consequence of collision kinematics.

Characterization of displacement damage in Si micro-volumes has particular importance for satellite imaging array applications. Device radiation hardening solutions have largely solved problems associated with ionization effects. However, particle irradiation seriously degrades charge transfer efficiency through carrier trapping and increases dark current by carrier generation. Permanent dark current increases from single particle interactions have been reported in sensor arrays following proton and neutron irradiation [4,5]. Pixel-to-pixel

variations in dark current increases following multiple interactions within each pixel have also been shown to depend on the incident particle and energy [3,6].

Orbital proton energy spectra, whether from the earth's trapped radiation belts or solar flares, typically peak at very low (and more damaging) energies and decrease exponentially with increasing proton energy. Typical spacecraft structural shielding effectively attenuates lower energy protons resulting in spectra extending from a few MeV to several hundred MeV with average energies over 20 MeV. Proton linear accelerators and cyclotrons are therefore well suited for monoenergetic characterizations of damage verses proton energy which can then be incorporated in damage predictions for a given environment and shielding configuration.

For the proton energy range of 10 MeV to 300 MeV, this work explains the average damage and pixel-to-pixel damage fluctuations in terms of calculated parameters reflecting the energy dependence of the proton-silicon interactions. The analysis predicts the damage distribution within a given array as illustrated for the particular case of a charge injection device (CID) depletion volume and the cross-sections and Si recoil energies applicable to 12 MeV, 22 MeV, and 63 MeV proton induced damage. This enables a direct comparison between the predicted damage distributions and the observed dark current histograms reported in [3] across a range of energies important for orbital environments.

Dark current extremes, which may follow from damage extremes, are a particularly serious concern for a variety of satellite imager applications. These "hot pixels" of "spikes" interfere with the instruments ability to resolve small, dim objects such as low magnitude stars which might be used for a star tracker guidance system. Also, spikes in a image can place overhead on data compression algorithms and burden telemetry channels. Extreme value statistics are well suited for characterizing the frequency and magnitude of these spikes. These tools are applied to proton damaged CID imagers to illustrate this approach, and we show that for one particular CID design, the spikes can be accurately predicted based on the calculated probabilities and kinematics of proton-initiated nuclear reactions.

2. Recoil Spectra Parameters

For proton energies of practical interest in satellite orbits, the damage is caused by recoiling atoms from collisions with Si atoms. As depicted in Fig. 1 [7], elastic scattering by the Coulombic field of the nucleus dominates for protons below 10 MeV, though at higher energies, nuclear elastic scattering also becomes important. By 60 MeV, about half of the displacement damage is due to nuclear inelastic reactions which dominate above 100 MeV. Elastic cross-sections are relatively high with recoil energies typically less than 1 keV as opposed to infrequent nuclear reactions emitting very damaging MeV-range recoils. In this work, the first and second moments of the recoil spectra are calculated separately for each type of interaction.

The average damage energy from all *elastic* recoils is obtained by numerically integrating the product of the differential cross section weighted by the corresponding recoil damage energy, over all scattering angles. Damage energy is defined here as the portion of energy lost by a recoil through mechanisms other than ionization as calculated by Lindhard et al. [8]. Note that this represents an important adjustment to the total energy imparted by the reaction atoms which must be assessed for evaluating either the nonionizing or the ionizing energy imparted. The second moment calculations proceed in the same manner, except the recoil damage energies now appear to the second power. The variance follows as the second moment minus

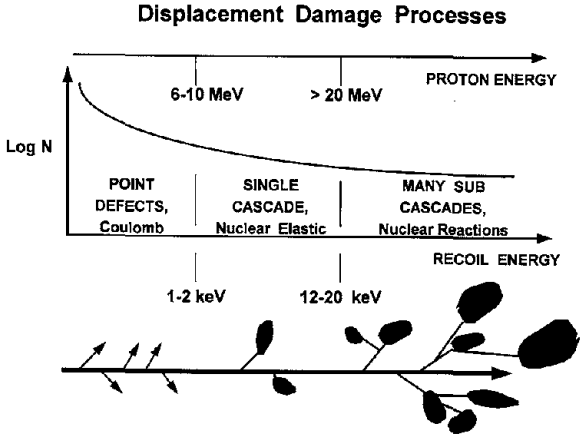


Fig. 1. Frequent coulombic scattering from protons of a few MeV initiate low energy recoil atoms resulting in isolated defect sites. More energetic protons can impart more energy to recoil atoms via nuclear elastic and inelastic reactions resulting in less frequent but more complex damage structures.

the first moment squared, as is customary. Figure 2 plots the mean and variance of non-ionizing energy for proton energies from 10 MeV to 300 MeV, and Table 1 lists the values of experimental interest, along with the total elastic cross-sections and recoil energies.

The *inelastic* reaction cross-sections are estimated according to the empirical formula of Letaw et al. [9]. Calculations of primary recoil energies consider both the initial intranuclear cascade and subsequent evaporation of nucleons. The momentum imparted during the evaporation phase is estimated using a Brownian motion model. Next, the average and variance of the damage energy are calculated as in the elastic case, and the results are summarized in Table 1. Further details and comparisons with data are discussed in [1,3].

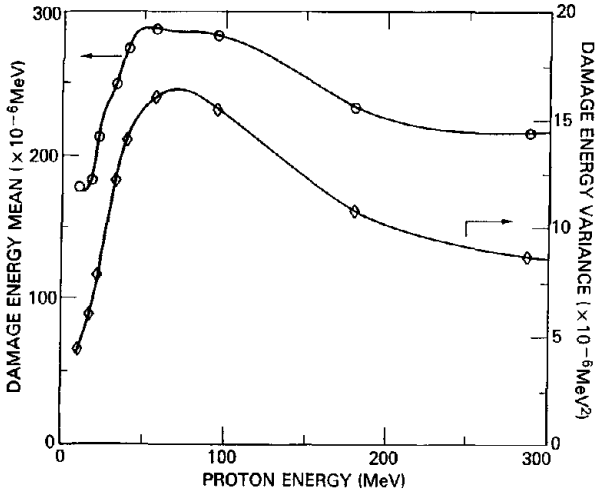


Fig. 2. The mean and variance of the total elastic damage energy are plotted versus proton energy along with a best-fit curve. The moments were calculated based on elastic differential cross-section data [1 and references therein] indicated by circles and triangles.

3. Damage Calculations

For a given proton energy, the mean and variance describing the probability density function (pdf) for damage from single interactions, as listed in Table 1, allow independent evaluation of the damage expected from the elastic and inelastic recoil categories. For the *elastic* category the mean for the pdf describing damage at a given proton fluence is the product of the number of interactions and the mean of the pdf for single interactions. The number of interactions is the product of the average cross-section, the incident particle fluence, and the number of Si atoms in the sensitive volume.

The elastic scattering component of the variance associated with the fluence dependent pdf is estimated as the product of the number of interactions and the single interaction pdf variance shown in Fig. 2. This is possible because Poisson fluctuations in the number of elastic recoils per pixel do not contribute significantly to the final result. In the regime where N, the average number of interactions per volume element, is greater than 20, the N-fold convolution of the single interaction pdf with itself leads to a Gaussian elastic damage distribution with mean and variance as described above.

For sensitive volumes and fluences of interest here, the average number of *inelastic* recoils ranges typically from a fraction to a few, and a discrete Poisson distribution determines the probability of a given number of inelastic recoils. The pdf governing the inelastic damage energy for a pixel with N_i

Table 1. Proton recoil spectrum parameters

Proton energy (MeV)	Cross section (BARNS)	Mean recoil energy (MeV)	Mean damage energy (MeV)	Variance of damage energy $(MeV)^2$
		Elastic reactions		
12	1548	3.40×10^{-4}	1.76×10^{-4}	4.77×10^{-6}
22	857	4.68×10^{-4}	2.13×10^{-4}	7.71×10^{-6}
63	318	7.77×10^{-4}	2.87×10^{-4}	1.62×10^{-5}
		Inelastic reactions		
12	0.670	0.267	0.0765	2.05×10^{-3}
22	0.723	0.569	0.111	$2{,}71\times10^{-3}$
63	0.523	1.44	0.152	3.11×10^{-3}

inelastic recoils reflects the N_i-fold convolution of the pdf for single inelastic damage. For purposes of this analysis, the form of the single event pdf for inelastic recoil products is approximated as a two parameter gamma distribution with mean and variance as indicated in Table 1.

Since the elastic and inelastic processes are independent random variables, the *combined* damage for pixels in which both occur follows as the convolution of the pdfs describing each of the two categories. Figure 3 illustrates this simulation for the specific case of the imaging array used in this study in which damage from a fluence of 4.0×10^{10} 12 MeV protons/cm^2 occurs, and each pixel's sensitive volume is 1300 μm^3. The Gaussian distribution, shown in Fig. 3a as the $N_i = 0$ case, describes damage corresponding to an average of 4,000 events per pixel. Figure 3a curves for $N_i = 1$ through 10 inelastic recoils per pixel reflect increases in both the means and variances as the shape tends toward Gaussian. Figure 3b shows the pdf for total combined damage as the superposition of the pdfs in Fig. 3a, after weighting by their associated Poisson probabilities according to the average of 1.8 inelastic recoils per pixel. This average is arrived at by considering the number of silicon atoms present in the 1300 μm^3 volume, and the composite cross-section for nuclear inelastic reactions for 12 MeV protons as shown in Table 1.

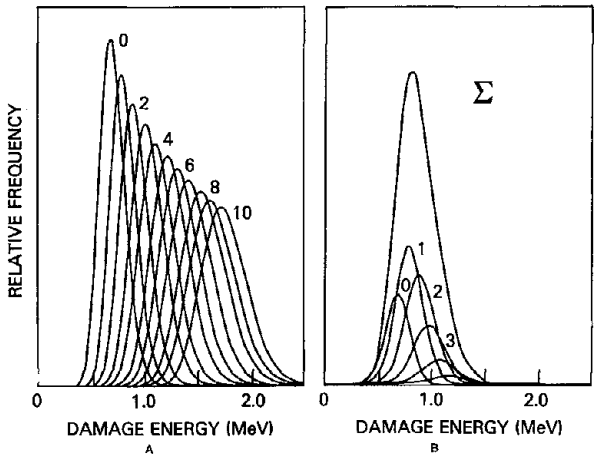


Fig. 3 (a). The Gaussian distribution with no inelastic recoils describes elastic damage and convolved distributions show combined damage from elastic and 1 through 10 inelastic recoils. (b) Weighting according to the Poisson probabilities precedes the superposition to determine combined damage probabilities. The simulation applies to the Si CID sensitive volume of 1300 μm^3 and 4.0×10^{10} 12 MeV protons/cm^2.

Early in a space mission or in a relatively benign orbit, the fluences may be 1-2 orders of magnitude lower, at about 10^8 cm^{-2}. In this low fluence regime, the very low probability of inelastic recoils suggests that two would probably not be observed in the same volume element. The number of elastic recoils per volume would be correspondingly low resulting in very large relative changes within the pixels where nuclear reactions occur. The product of the low probability of an inelastic event with the large number of pixels determines the pixel population for which damage exceeds the average by factors of up to 1,000.

4. Predicting Damage Extremes

In addition to being a necessary tool for assessing radiation-induced fixed pattern noise, the probability density function describing damage throughout the array can be used to predict the number of elements sustaining exceedingly large damage increases after a specified exposure. In [6] it was shown how extreme value statistical analysis can describe the measured distribution of pixels with the largest damage increases following 12 MeV and 63 MeV proton damage to the Si CID. For a broad range of proton energies and fluence levels, the largest extremes were shown to obey a Type 1 extreme value distribution. Next it will be shown that the particular Type 1 distribution describing proton-induced damage extremes can be predicted from the calculated pdf described above.

Figure 4 shows an expanded view of the tail region in Fig. 3b which identifies the contributions to the pdf from the 11 populations containing 0 through 10 inelastic recoil events per pixel. The damage energy distribution has a mean of 0.85 MeV, and the skewed high energy tail extends to about 1.8 MeV. Individual distributions are identified according to the number of inelastic recoils. Figure 4 illustrates how several of the component distributions contribute to the probability of exceeding large damage energies. Based on a total pixel population of 61,504, the inset presents the number of pixels expected above the specified damage level, E_d. This number is the total population multiplied by p_d, the probability of exceeding damage energy E_d within the whole array. This probability is calculated as the summed pdf integrated from E_d to infinity.

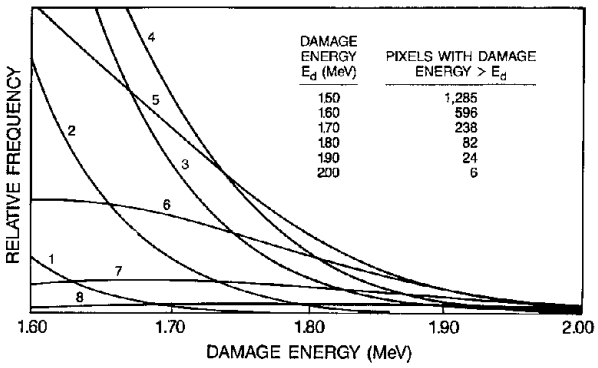


Fig. 4. For the simulation depicted in Fig. 3, volumes containing from 1 to 10 inelastic recoils contribuite to the population of pixels with the most damage. The inset shows the number expected above a given damage energy for a 61,504 pixel array.

Two steps are necessary to compare these results on the basis of the cumulative Type 1 extreme value distribution. As discussed in [6,10-12], extreme value analysis can be applied to data to evaluate the probability of exceeding a certain value within any population size by evaluating a set of largest values extracted from subsets of a given population. In the next section we will treat the case where the 61,504 pixel population has been subdivided into 248 groups of 248 pixels each. Using p_d as defined above, the probability, p_0, of having no pixels exceeding E_d within the group of 248 pixels can be evaluated using the discrete binomial distribution as:

$$p_0 = (1-p_d)^{248}. \quad (1)$$

The associated standard variate specific to the extreme value cumulative probability plot is given by:

$$S(p_0) = -\ln[-\ln(p_0)]. \quad (2)$$

Thus E_d, or a proportional quantity such as dark current, can be plotted against the corresponding standard variate to predict the Type 1 extreme distribution specific to the pdf from which it is generated. Detailed discussions of extreme value analysis are discussed in the references [11,12], and applications to this study will be illustrated in the following section.

5. Comparison with Dark Current Data

Calculations described in the previous section are compared here to measured dark current increase distributions specific to proton-induced damage in a General Electric 256 pixel × 256 pixel Si CID. Devices are fabricated in an n-type Si epitaxial layer doped with 5×10^{14} P atoms/cm^2. A field isolation oxide confines the collection area to about 17 mm × 17 mm, but for purposes of dark current studies only the 1300 μm^3 depletion volume leads to carrier generation.

All dark current data reported here were acquired at 18.0 °C and correspond to a 248 × 248 subset of the array. After each proton exposure and measurement the dark current increase for each element was calculated by a pixel-by-pixel subtraction of the pre-irradiation value. This correlation removes imager spatial noise not resulting from radiation. Temporal read-out noise accounts for less than 5% of the dark current spreads reported here. More detailed descriptions of the imaging array and the dark current measurement are provided in [6].

Proton irradiations with energies of 12 MeV, 22 MeV, and 63 MeV were performed at the University of California at Davis cyclotron facility. The beam line and dosimetry have been described previously [13]. Irradiations were conducted at a nominal dose rate of 1 kRad(Si)/s with all leads grounded. Dark current measurements were initiated about 15 minutes post irradiation and repeated after 1 day and again after about 1 week. No significant annealing was observed over this period.

In Fig. 5, comparisons are made between dark current data histograms and calculated damage energy distributions in the CID pixels. The calculation approach described above has been exercised for three 12 MeV proton fluences corresponding to averages of (1.8, 4.5, and 9.0) inelastic recoils/pixel. Based on the population of 61,504 pixels and Poisson statistics, the maximum numbers of inelastic recoils expected in any single pixel are 10, 16, and 24, respectively. For comparing the calculations to dark current data, a conversion factor relates the average dark current and the mean damage energy. For the three fluences, the average conversion factor of 2.2 nA/cm^2 per MeV of damage energy varies by up to 10%, which reflects the experiment's dosimetry uncertainty. The calculated damage curves in Fig. 5, based on the first and second moments of the non-ionizing energy imparted by the recoil spectrum, describe the dark current distribution to a remarkable degree of accuracy.

Comparisons for 22 MeV and 63 MeV proton damage show similar agreement. The coefficient of variation, defined as the ratio of the standard deviation to the mean damage (or dark current), is a dimensionless figure-of-merit. At 12 MeV, 22 MeV, and 63 MeV, the experimental and calcu-

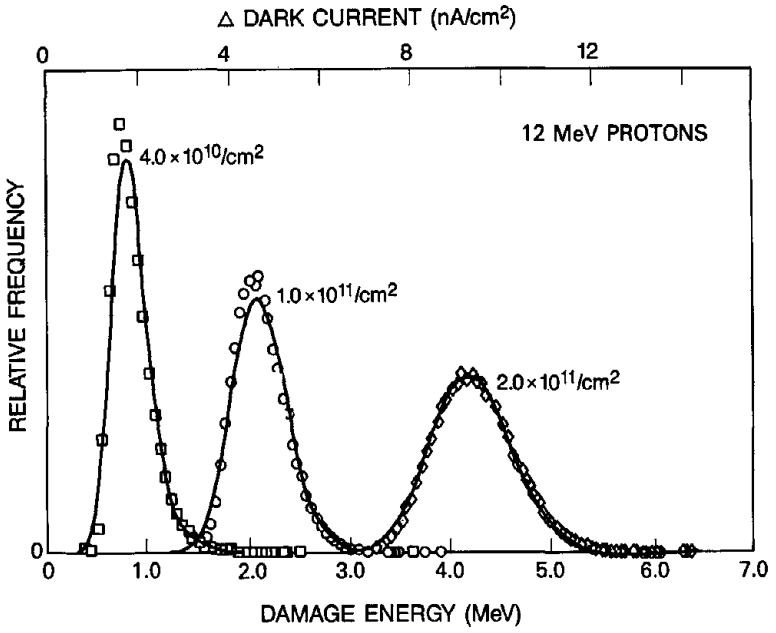


Fig. 5. Calculated damage energy distributions show excellent agreement with measured dark current histograms from a Si CID damaged by 12 MeV protons. Calculations are based on averages of (1.8, 4.5, and 9.0) inelastic recoils per pixel, and the damage distribution shapes reflect the associated discrete Poisson distributions.

lated results agreed within 2%, 12%, and 15% respectively [14]. Also at 63 MeV, with 45% of the damage caused by inelastic recoils, the means of the two distributions are normalized by a factor of 2.0 nA/cm² per MeV of damage. This does not differ significantly from the conversion factors determined for 12 MeV, thus demonstrating that the average damage is proportional to the energy lost through non-ionizing processes, and that the expected damage from both the elastic and inelastic categories is present.

The somewhat better agreement between calculated and measured damage distributions at the lower proton energy of 12 MeV could be influenced by characteristics associated with high energy recoils. At proton energies of 12 MeV and 63 MeV, the contribution to the total damage from inelastic reaction recoils increases from roughly 15% to 45%. Also, as this fraction increases, the average inelastic recoil energy (and range) also increases, and at higher proton energies the higher energy recoil ranges approach the smallest dimension of the sensitive volume (about 2 μm). These issues would be even more important for smaller sensitive volumes (i.e., CTE loss in a CCDs buried channel).

6. Largest Dark Current Extremes

Here the measured largest dark current increases are compared to the calculated damage maxima for the specific cases of the three 12 MeV proton fluences of Fig. 5. For each proton energy and fluence level, the dark current extreme populations are generated by subdividing the 61,504 pixel population into 248 groups of 248 pixels each. The largest value from each group comprises the population of extremes. Figure 6 depicts how the extreme distribution is derived for the case of the lowest fluence level shown in Fig. 5 (note this example also corresponds to the calculations for Figs. 3b and 4). After ranking and estimating the probability according to the [rank/($n+1$)] for n samples as in [6], the standard variate follows from Eq. (2), and the measured dark current extremes can be compared with the Type 1 extreme value distribution using a Type 1 cumulative probability chart.

Likewise, damage maxima calculated as described in section C can be compared to the same Type 1 extreme probability distribution using Eqs. (1 and 2) and the normalization constant of 2.2 nA/cm² per MeV of damage energy. Figure 7

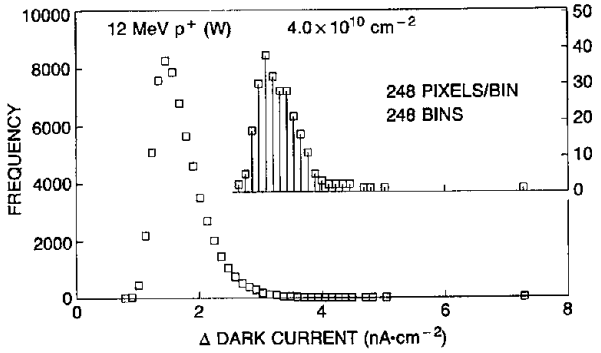


Fig. 6. Measured dark current histogram for 61,504 pixels following exposure to 4.0×10^{10} 12 MeV protons/cm^2. The 248 extremes are from groups of 248 pixels.

compares measured dark current extremes, for the three 12 MeV proton fluences treated in Fig. 5, to predicted damage maxima according to the Type 1 distribution. The linear character of the data and calculation show that they obey a Type 1 distribution, and the close agreement at each fluence demonstrates the robustness of the analysis. The return period abscissa at the top of Fig. 7 identifies the largest expected dark current increase for a given number of array subsets. For example, at the fluence of 2.0×10^{11}/cm^2 the return period value of 10 corresponds to about 13 nA/cm^2 indicating the largest expected increase within a set of 10 groups or 2,480 pixels. Good agreement also exists between the measured and predicted extremes from 63 MeV protons.

The ability of the calculation to predict the largest measured dark current changes offers insight into the mechanisms responsible for proton-induced damage extremes. The linear response on the Type 1 plot indicates that a single mechanism is probably responsible for largest values while the slope reflects the variance. As pointed out in Fig. 4, the largest damage regions in this fluence regime follow from the probabilistic treatment of pixel populations damaged by several inelastic reaction recoils.

When the probability of an inelastic recoil per pixel is much less than one, as is the case in many natural space environments, the analysis can determine the total number of pixels expected with damage above a given level. In this regime, where the background radiation-induced damage can be quite low, largest damage regions can be several hundred times the average. Some of the array subsets would have largest changes dominated by single inelastic recoil damage and others by the largest of the less damaging elastic recoils. In this case, agreement with the Type 1 cumulative chart could be expected only with sufficiently larger bin sizes so that each bin would include at least 1 pixel with damage from an inelastic reaction.

A qualitative comparison of such a situation follows from our evaluation of the proton response of an alternate CID imager design. The important aspects of this "narrow row" design were previously discussed in [6], with the key difference resulting in spurious high electric field profiles near the row

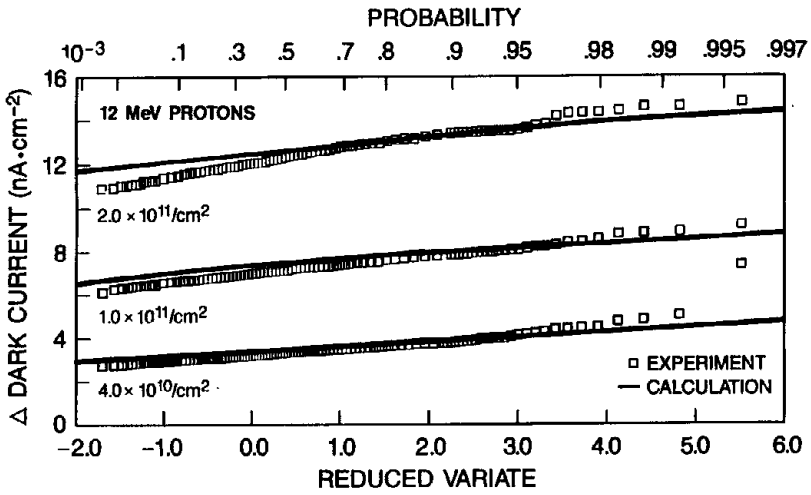


Fig. 7. Cumulative probability distributions demonstrate excellent agreement between calculated damage extremes and the measured dark current extremes based on a 248 pixel by 248 group extreme value analysis. Though not shown here, similar agreement is obtained for damage from 63 MeV protons.

electrodes. The comparison of the dark current and extreme distributions for this device type, shown in Fig. 8, can be made with the previously discussed design at the same proton exposure level, as in Fig. 6. Note that the average dark current is doubled, but more importantly, the character of the extreme distribution is markedly different. The consequence of this is more evident in the probability chart of Fig. 9. Clearly the narrow row design results in an extreme distribution which is not Type 1 when analyzed as before. Rearrangement of the array to 31 bins of 1984 pixels offered a better match with the Type 1 distribution. Even so, the extremes for this case cannot be understood based on first principles analysis of damage mechanisms as before. We concluded that in this case, the largest extremes were not caused by conventional charge generation, and extreme value statistics played a critical role toward quantifying the likelihood and magnitudes of this other mechanism. In [4] we discuss supplemental measurements and analysis which have lead us to conclude that the high field regions were causing localized lowering of the band-gap resulting in field enhanced emission and tunneling currents. Thus the statistics of extremes are applied to evaluate design variations and to assure that optimum imager performance can be assured. We also concluded that acceptable designs should have extreme characteristics as depicted in Fig. 6 which are limited only by unavoidable consequences of particle-semiconductor physics.

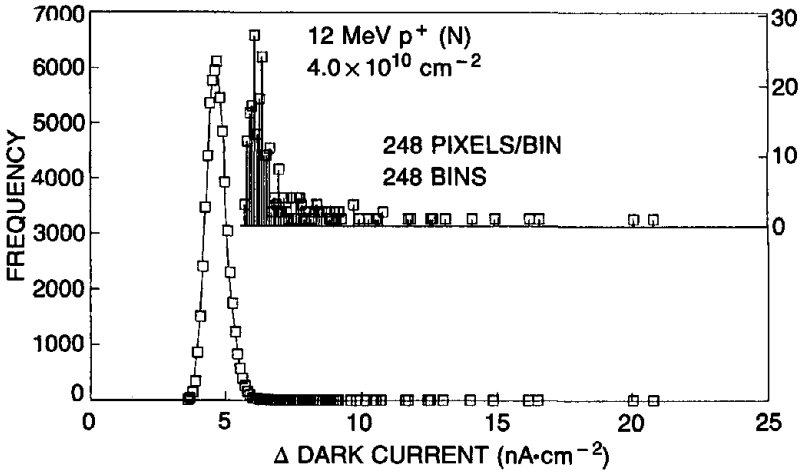


Fig. 8. The high electric field CID design yields a different dark current and extreme response as compared to the same conditions shown in Fig. 6. High electric fields are thought enhance the leakage currents when associated with damage.

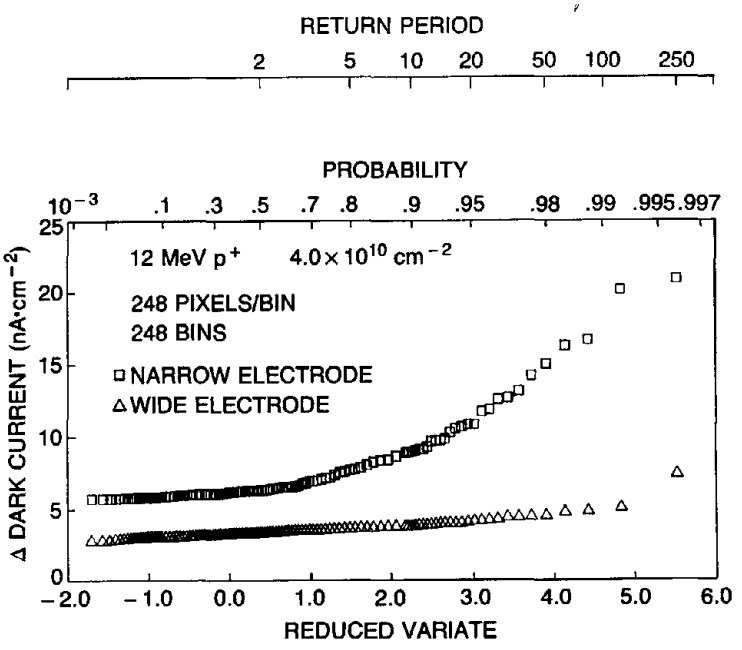


Fig. 9. The probability chart comparing the responses shown in Figs. 6 and 8 suggests the role of field enhanced mechanism in causing the largest leakage extremes.

7. Conclusions

This paper presents an analytic approach for determining the pixel-to-pixel distribution of particle-induced displacement damage in micro-volumes representing sensitive volumes in sensor arrays. The calculation is based on interaction cross-sections as well as parameters describing the damage imparted by the spectrum of particle-initiated recoils. It predicts the dark current distribution and largest dark current changes in a Si CID following incremental damage with 12 MeV, 22 MeV, and 63 MeV protons. These proton energies span a regime important to the natural space environment; lower energy protons for which the damage is dominated by elastic scattering and higher energies where nuclear reactions become increasingly important. The analysis illustrates how high energy recoils from nuclear reactions influence the pixel-to-pixel variance in proton-induced damage and cause the largest damage occurrences. To understand the important exception, we rely on extreme value statistics to identify and quantify the role of electric field enhanced emission as a mechanism for excessive leakage currents.

The calculation is general in the sense that once the parameters describing the recoil spectrum are determined, the particle-induced damage distribution can be calculated as a function of particle type, particle fluence, sensitive volume, and material. The significance of these results is that once the factor relating the mean dark current to the damage energy is known from a single measurement on a particular array, the radiation response in a specified environment can be predicted. In addition to providing a means for assessing the radiation response of a given imager, this analysis has flexibility enabling the design-phase evaluation of the radiation response of different pixel geometry and materials in a variety of environments.

Extreme value statistics play a critically important role in understanding leakage current spikes and in assuring reliable satellite performance. In ongoing related research we continue to rely on this valuable tool for assessing damage and single particle ionization extremes in infrared imaging arrays and in optoelectronic detector materials for high data rate spacecraft data links, each of which must perform to exacting standards to assure reliable performance of extremely valuable space assets.

8. References

[1] E. A. Burke, Energy Dependence of Proton-Induced Displacement Damage in Silicon, IEEE Trans. Nucl. Sci. **33** (6), 1276 (1986).
[2] G. P. Summers, E. A. Burke, C. J. Dale, E. A. Wolicki, P. W. Marshall, and M. A. Gehlhausen, Correlation of Particle-Induced Displacement Damage in Silicon, IEEE Trans. Nucl. Sci. **34** (6), 1134 (1987).
[3] C. J. Dale, P. W. Marshall, E. A. Burke, G. P. Summers, and G. E. Bender, The Generation Lifetime Damage Factor and its Variance in Silicon, IEEE Trans. Nucl. Sci. **36** (6), 1872 (1989).
[4] J. R. Srour, R. A. Hartmann, and K. S. Kitazaki, Permanent Damage Produced by Single Proton Interactions in Silicon Devices, IEEE Trans. Nucl. Sci. **33** (6), 1597 (1986).
[5] J. R. Srour and R. A. Hartmann, Effects of Single Neutron Interactions in Silicon Integrated Circuits, IEEE Trans. Nucl. Sci. **32** (6), 4195 (1985).
[6] P. W. Marshall, C. J. Dale, E. A. Burke, G. P. Summers, and G. E. Bender, Displacement Damage Extremes in Silicon Depletion Regions, IEEE Trans. Nucl. Sci. **36** (6), 1831 (1989).
[7] Susan Wood, N. Doyle, J. A. Spitznagel, W. J. Choyke, R. M. Moore, J. N. McGruer and R. B. Irwin, Simulation of Radiation Damage in Solids, IEEE Trans. Nucl. Sci. **28** (6), 4107 (1981).
[8] J. Lindhard, V. Nielsen, M. Scharff and P. V. Thomsen, Integral Equations Governing Radiation Effects (Notes on Atomic Collisions, III), Mat. Fys. Medd. Dan. Vid. Selsk. **33** (10), 1-42 (1963).
[9] J. R. Letaw, R. Silberberg and C. H. Tsao, Proton-Nucleus Total Inelastic Cross Sections: An Empirical Formula For $E > 10$ MeV, Astr. J. Suppl. **51**, 271-276 (1983).
[10] P. W. Marshall, C. J. Dale, and E. A. Burke, Proton-Induced Displacement Damage Distributions and Extremes in Silicon Micro-Volumes, IEEE Trans. Nucl. Sci. **37** (6), 1776 (1990).
[11] E. J. Gumbel, Statistics of Extremes, Colombia University Press, New York, (1958).
[12] E. Castillio, Extreme Value Theory in Engineering, Academic Press, San Diego (1988).
[13] K. M. Murray, W. J. Stapor, and C. Castenada, Calibrated Charged Particle Radiation System with Precision Dosimetric Measurement and Control, Nucl. Inst. Meth. **A281**, 616 (1989).
[14] C. J. Dale, P. W. Marshall, and E. A. Burke, Proton-Induced Spatial Dark Current Fluctuations in Focal Plane Arrays, IEEE Trans. Nucl. Sci. **37** (6), 1784 (1990).

About the authors: Paul W. Marshall is employed by SFA, Inc. providing technical support to the Naval Research Laboratory Radiation Effects Branch for investigations of space radiation effects in satellite micro-electronic and photonic systems. Cheryl J. Dale is employed by the Condensed Matter and Radiation Sciences Division at the Naval Research Laboratory, where her research includes radiation effects in CCDs, and in technologies critical for photonics applications on board satellites. Edward A. Burke is a radiation effects consultant and has been employed at Spire Corporation, Mission Research Inc., and also headed the Radiation Effects Section at the Rome Air Development Center.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 495 (1994)]

Performance Comparison Between a Statistical Model, a Deterministic Model, and an Artificial Neural Network Model for Predicting Damage From Pitting Corrosion

Volume 99 Number 4 July–August 1994

M. Urquidi-Macdonald and D. D. Macdonald

Pennsylvania State University, University Park, PA 16802

Various attempts have been made to develop models for predicting the development of damage in metals and alloys due to pitting corrosion. These models may be divided into two classes: the empirical approach which employs extreme value statistics, and the deterministic approach based on perceived mechanisms for nucleation and growth of damage. More recently, Artificial Neural Networks (ANNs), a nondeterministic type of model, has been developed to describe the progression of damage due to pitting corrosion. We compare the three approaches above – statistical, deterministic, and neural networks. Our goal is to illustrate the advantages and disadvantages of each approach, in order that the most reliable methods may be employed in future algorithms for predicting pitting damage functions for engineering structures. To illustrate the difficulty that we face in predicting cumulative pitting damage, we selected a set of data that was collected in the laboratory. We compare and contrast the three approaches by reference to this data set.

Key words: artificial neural networks; deterministic; mathematical modeling; pitting corrosion; statistics.

Accepted: March 22, 1994

1. Introduction

On the basis of laboratory studies [1], and through the analysis of field data collected over the past decade by Battelle Columbus Laboratory [2], several factors have been identified as contributing to the development of pitting damage in gas fired heat exchangers in domestic and industrial service:

(i) The type of alloy used for fabricating the heat exchanger
(ii) Chloride concentration in the flue gas condensate
(iii) Temperature
(iv) Exposure time
(v) Ambient versus indoor air
(vi) pH
(vii) Electrochemical potential

Unfortunately, few of these factors are simply related to the damage functions or to one another. Accordingly, it is seldom possible to establish a simple empirical equation for predicting pitting damage as a function of these variables. The case cited above is not atypical, and it illustrates the difficulties faced by those who seek to develop predictive models for assessing corrosion damage. Indeed, the data base established by Battelle is probably one of the best that currently exists for the development of pitting damage in an industrial system. A full interpretation of the Battelle data in terms of statistical, deterministic, and artificial neural network models is published elsewhere [3].

In the present paper, we use a more restricted database to illustrate how various classes of models are used to analyze the damage caused by pitting corrosion. These models include a statistical approach based on the Weibull distribution function, a deterministic model based on a physicochemical

mechanism, and an Artificial Neural Network (ANN) that assumes neither a mathematical model nor a physical model, but which seeks to establish relationships between the dependent and independent variables by examining the patterns contained within the data set.

2. Experimental Data

We used laboratory data to illustrate the time and potential dependencies of pitting damage. To do this, we chose a laboratory data set for which the following independent variables were identified: 1) concentration of minor alloy elements in weight percentages, 2) difference in oxidation state between host metal and minor alloy elements, 3) applied potential, and 4) time of observation. Independent variables 1) and 2) are related to the type of alloy; and independent variable 4) together with solution composition (which was maintained constant) is determined by the electrochemical potential. Temperature, solution composition, and pH were maintained constant. The dependent variable was the total number of pits.

We then used this set to illustrate the prediction of cumulative damage for pitting corrosion using three different models: statistical, deterministic, and artificial neural networks. The data were measured by English and Macdonald at SRI International [1].

Several alloys of nickel were fabricated. Each of the alloys tested was arc-melted from powders under an Argon gas blanket in a sealed container. Binary nickel alloys containing Al, Ta, and Mo in nominal concentrations of 0.1 %, 0.5 %, 1 %, 3 %, 5 %, and 8 % by weight were cast as 100 g buttons and were sectioned in an acrylic plastic before polishing. The alloying elements were selected on the basis of their oxidation states relative to nickel (oxidation state = 2). The excess oxidation states range from 1 for Al to 4 for Mo.

The polished specimens were placed in a cell. The electrode potential was swept in the positive direction at 1 mV/s from an initial potential of 0.0 V. This results in a distribution in breakdown potentials. Alternatively, the potential was stepped from 0.0 V to 0.325 V, 0.375 V, 0.4 V, and 0.45 V. This resulted in a distribution in induction (or observation) times for the nucleation of pits.

In both types of experiments the pit nucleation and growth events were photographed at 65 × magnification at regular intervals. The number of pits were counted on the pictures taken at different times and conditions.

The pitting data were measured several times on a similar sample to explore reproducibility. The reproducibility in pure nickel appeared satisfactory (about 10% difference between runs), but the reproducibility from alloy composition to alloy composition was different. Reproducibility was better at high potentials perhaps because the total number of pits developed was higher. Reproducibility appeared to be better at high minor alloy contents and high oxidation states (about 20 %), than at low minor alloy contents and low oxidation states (about 50 %). Regardless of the poor reproducibility in some of the samples, a general trend was observed: a) The cumulative number of pits diminishes with 1) minor alloy element content, and 2) with increasing difference in oxidation state between the base alloy and the minor alloy element; and b) The cumulative number of pits grows with increasing applied potential and observation time.

Cumulative pitting damage is an irreversible, dynamic, time decay, environmentally related process. The literature is abundant in pitting corrosion data, but there is a lack of good quality data because of the difficulty of measuring pitting corrosion when controlling all the environmental parameters.

All model building is concerned with an attempt to increase our knowledge of complex physical reality. The parameters plus the validity of the model must be determined from the data. The philosophy behind the type of model is different. The information obtained from a purely probabilistic model (statistic and stochastic models) is about finding embodied in the data trends that can be used in future predictions. The information obtained from a deterministic model is about the physical meaning of the phenomena itself. The information obtained with a ANN model is about the dependency and importance of input /output relationships. In any case, the model capabilities need to be tested.

We can start the process of solving our problem by listing facts, listing observations, and listing existing laws relating variables and outputs. Then we have to ask ourselves which will be the best model to describe the problem, and what do we expect from the model. Later we need to identify the model or models to use; specify the constraints, choose the coordinates, and apply the laws dictated by the model. Important questions related to the choice of a correct model are: Is the process static or dynamic? Is the process stationary or not?; Are the available data distributed or not? What do the data mean? What is the data variance? What are the correlations?. In any case, the fitted model you

use for analyzing your data is the nearest representation of the true situation you have available.

3. Statistical Approach and Results

Stochastic processes are dynamic, and good examples are fatigue, wear, and crack or pit growth. There are two main types of stochastic processes: stationary and nonstationary.

It is well known from experimental data that cumulative pitting damage is a nonstationary phenomenon. It is well known that nonstationary models and their estimation are notoriously difficult problems to handle except for special cases. Discrete state continuous Markoff processes are good examples of models that describe nonstationary stochastic phenomena. However, there is no literature on problem solving using nonstationary finite Markoff chains [4]. On the other hand, for the last data set [1] (measured at the laboratory), the cumulative damage versus time was measured, but the pit depth versus number of pits was not. Therefore, it is impossible to derive a dynamic model for pitting damage using that data set. The only option available is to try to fit a static model (i.e., our hypothesis is that the numbers of pits versus pit depth does not change with time). We choose a 2 parameter Weibull distribution; for which we assume that the independent parameters are potential, and the oxidation state and concentration of the minor alloying elements. The dependent variables are the cumulative number of pits and the induction time. The Weibull distribution function is

$$F(x) = (1 - \exp(-(x/\beta)^{\alpha}))$$

where α and β are fitting parameters and x is the dependent variable of interest.

We normalized the data set to 80 % of its maximum value, allowing 20 % of the pits to nucleate if the time would have been extended to infinity. For each potential, oxidation state, and percentile of minor alloy element, we performed a nonlinear fit to estimate the Weibull fitting parameters.

The choice of a Weibull distribution is arbitrary; we chose a Weibull distribution instead of some other probabilistic distribution because of the flexibility that this distribution offers in fitting different shapes obtained when plotting cumulative damage versus dependent variables.

The nonlinear fits were acceptable (sum of square errors between fit and data < 20 % for α or β). We used those data sets for which smooth changes of α and β were calculated as a function of potential. That reduced the data base to about 50 % of the total available (the total data base had 1400 data lines containing number of pits at different observation times, applied potential, oxidation states, and percentile of minor elements). We plotted the α and β values as a function of potential. Figures 1a and 1b show the results. The beta parameter of the Weibull distribution appears to not change with applied potential at high concentration of minor alloy elements (5 %), but it changes drastically with applied potential at low concentrations of minor alloy element (3 %, 1 %). We then fit polynomials describing α and β as functions of applied potential, oxidation state, and percentile of minor alloy elements.

The Weibull distribution with α and β as parameters was used to generate the cumulative damage function. Figures 2a, b, and c show the predictions obtained with this statistical model. When we compared the predictions obtained with this model and the measured data, we observed that both trends are similar. However, it would be very risky to use the model to make predictions for other oxidations states, percentile of minor alloys elements, or applied potentials outside the range for which the Weibull-parameters were calculated.

It is well known that the Weibull distribution is a sufficiently flexible function that practically any set of data can be fitted by it. However, the problem we faced is that we do not know *a priori* the correct relationships between α and β and the independent variables.

Predictions with the same model for oxidation states greater than 3–2 gave cumulative probability of zero at any time and are not shown. The designation "3–2" referes to the oxidation state of the alloying element (Al = 3) and the host metal (Ni = 2).

4. Deterministic Model

A completely successful model must account for all of the phenomenological correlations that exist between pitting susceptibility and pit velocity, and various environmental and electrochemical factors, such as temperature, pH, [Cl], potential, time, and alloy composition. The Point Defect Model (PDM) [5, 6] accounts for the effects of electrochemical potential, alloy composition, and environmental conditions on the nucleation of pits.

The deterministic model is based on the PDM and the Solute Vacancy Interaction Model (SVIM) [7–10]. The PDM proposes that passivity breakdown occurs because of an enhanced flux of cation vacancies from the film/solution to the metal/film

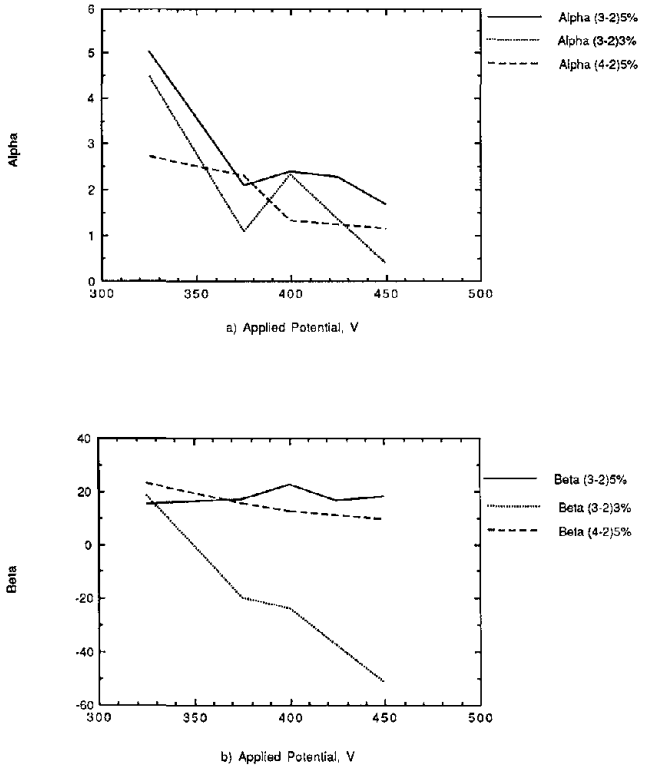


Fig. 1. (a) Alpha parameter of the Weibull distribution versus applied potential; at several percentiles of minor alloying elements, and oxidation states. (b) Beta parameter of the Weibull distribution versus applied potential; at several percentiles of minor alloying elements, and oxidation states.

interface. If the excess of vacancies arriving at the interface between the metal and the film can not be absorbed into the metal or be annihilated by some appropriate mechanism at high enough rate, they accumulate to form a vacancy condensate at the metal film/interface, which then grows to a critical size. The PDM is used to calculate the breakdown potential and induction time. The effect of the minor alloying elements in the oxide film on the breakdown parameters is modeled using the SVIM. The SVIM is based on the hypothesis that highly oxidized solutes in the passive film electrostatically complex with the mobile cation vacancies.

The PDM and SVIM results in distributed values of the breakdown potential and induction time, and complexing between the immobile alloying element in the film (the "solute") and the mobile vacancies diminishes the flux of vacancies across the film. This leads to an increase in the breakdown potential and the induction time for film breakdown. The higher the net oxidation state (minor alloy element oxidation-host ion oxidation) and/or the higher the percentile of minor alloying elements in the film, the greater the effect on reducing the flux of vacancies and hence in increasing the pitting potential and the induction time. Once the pits nucleate, they grow at different rates. To calculate the pit growth rate we used 1) a simplistic steady state model suggested by Alkire [11]; and 2) a nonstationary model developed by us [12–13]. The stationary model is expected to be adequate for only short times.

The overall model (combination of the PDM, SVIM, and pit growth) requires the defining of a number of parameters, as shown in Table 1.

Figure 3a shows the cumulative probability of the number of pits (normalized to 1) as a function of pit depth and observational time of 50 s, for an applied voltage of 0.325 V and for several concentrations of the minor alloying element with oxidation state of 3–2 (example aluminum in nickel). Figures 3b and 3c show similar plots for oxidation states of 4–2 (e.g., titanium–nickel) and 6–2 (e.g., molybdenum–nickel), respectively, for the same conditions. It is interesting to note the great effect of minor alloying elements with high oxidation states. The model predicts that the cumulative probability of the number

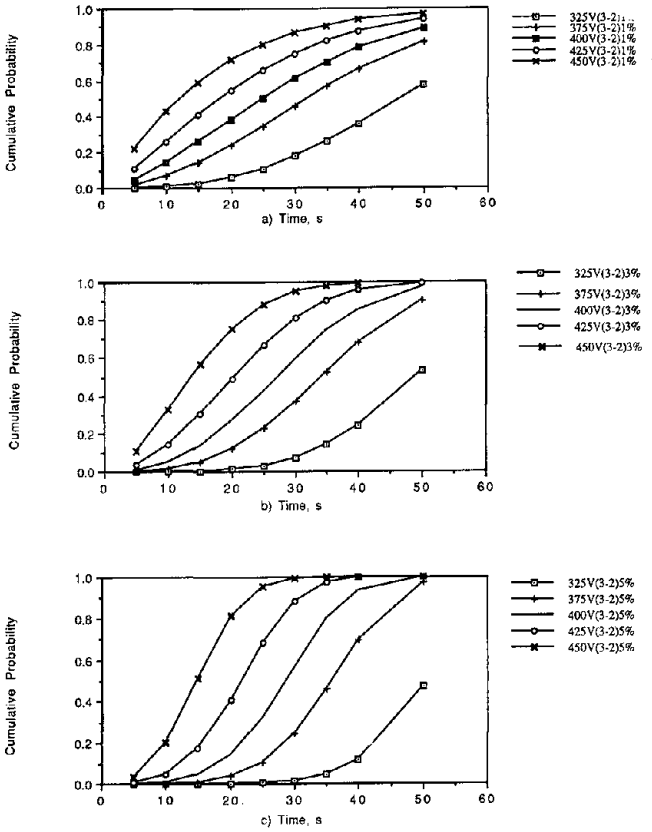


Fig. 2. Predicted cumulative probability, obtained using the statistical model, versus time of observation, at several applied potentials. (a) Oxidation state 3–2; and 1 % of minor alloying element segregated in the film. (b) Oxidation state 3–2; and 3 % of minor alloying element segregated in the film. (c) Oxidation state 3–2; and 5 % of minor alloying element segregated in the film.

Table 1. Input data used in the calculation of the deterministic/probabilistic model

Parameters	Value	Units
Stoichiometry	2	
Avogrado constant	6.023 E + 23	mol^{-1}
Mol vol. of oxide cation	30	cm^3/mol
Gibbs energy change[a]	−40,000	J/mol
Gibbs energy change[a]	−10,000	J/mol
Mean diffusion coefficient	5 E − 20	cm^2/s
Standard deviation	0.75 Dmean	cm^2/mol
Chloride activity	0.573/2	
Electrical field across film	1.1 E + 6	V/cm
Alpha	0.65	
Beta	−0.01	V/pH unit
Critical area vacancy size[a]	1 E + 16	$No./cm^2$
Critical vacancy flux[a]	15.87 E + 12	$No./cm^{2 \cdot s}$
Temperature	298.15	K
Applied potential	−0.55	V SHE
Molar gas constant	8.314	$J\,K^{-1}\,mol^{-1}$
Electrical potential film/sol	−0.5	V SHE

[a] Variables that were used to adjust one datum point to scale the results properly.

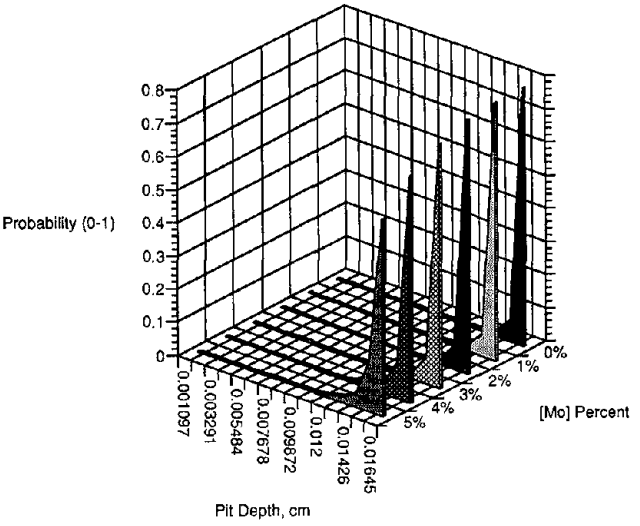


Fig. 3a. Cumulative probability, calculated using the deterministic model, versus pit depth for several concentrations of minor alloying element segregated in the film for oxidation state (3–2).

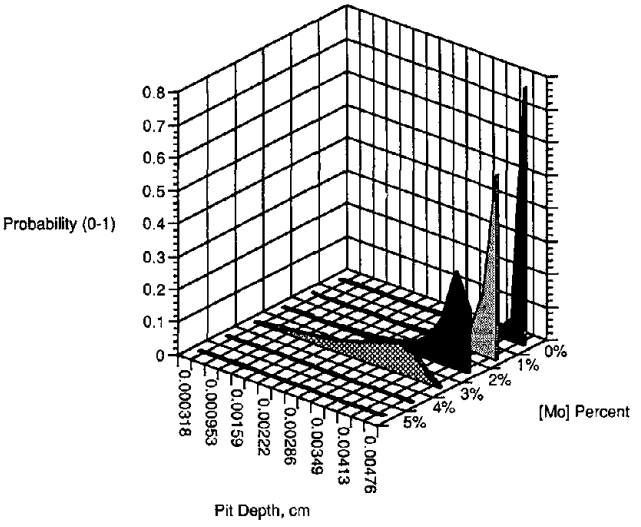


Fig. 3b. Cumulative probability, calculated using the deterministic model, versus pit depth for several concentrations of minor alloying element segregated in the film for oxidation state (4–2).

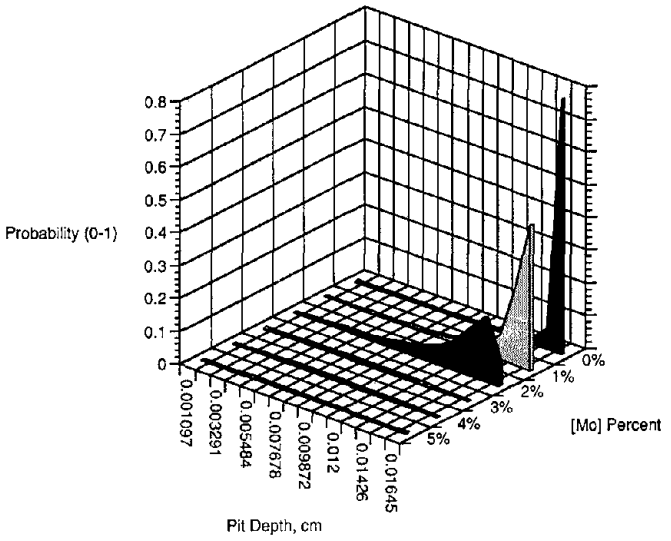


Fig. 3c. Cumulative probability, calculated using the deterministic model, versus pit depth for several concentrations of minor alloying element segregated in the film for oxidation state (6–2).

of pits at all pit depths is higher at lower minor alloying element oxidation state and at lower concentration of the minor alloying element. Because the model does not assume a total number of breakdown sites only a normalized probability is obtained.

Figure 4 shows the beneficial effect of adding minor alloying elements with high oxidation states.

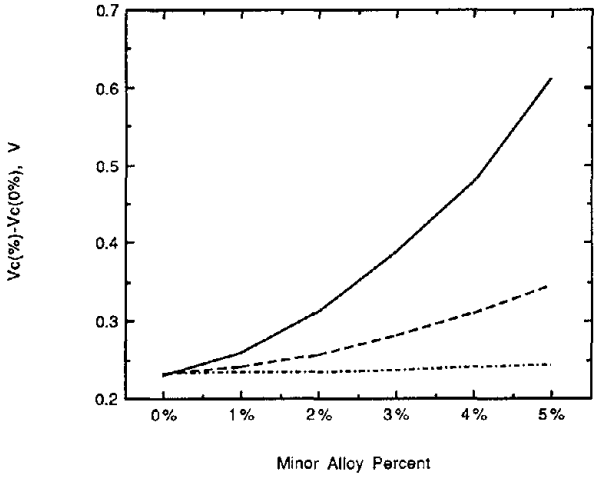


Fig. 4. Difference between calculated breakdown potential of nickel containing 0 %–5 % of minor alloying elements with oxidation states of (___3–2: Ni–Al), (- - - 4–2: Ti–Ni), and (- . - . - . 6–2: Mo–Ni) and calculated breakdown potential of pure nickel (containing 0 % of minor alloying elements).

The breakdown potential is shifted in the positive direction, indicating that higher potentials are necessary to achieve the same damage.

5. Artificial Neural Network Model and Results

Probably the most efficient method, when data are available, of establishing relationships between inputs and results is to use artificial intelligence techniques. Accordingly, we describe here an Artificial Neural Network (ANN) for predicting pitting damage functions for condensing heat exchangers. When the net is trained with reliable data and knowledge, we are able to accurately predict damage outside the ranges of the input variables.

An ANN is a highly interconnected system inspired by the brain and formed by simulated "neurons" represented by a transfer function, and "weights" associated to the connections of the "neurons." The back propagation training algorithm allows experimental acquisition of input/output mapping knowledge within multilayer networks. Because we have experimental data on the cumulative numbers of pits versus time of observation, as a function of oxidation state, minor alloying element, and applied potential, we decided to use an ANN backward propagation technique with supervised learning. During training of the ANN, the cumulative numbers of pits were used as "output" and the applied potential, oxidation state,

minor alloying element concentration, and time of observation as "inputs." We explored several topologies to obtain the best compromise between learning and computing time for an ANN with 2 hidden layers.

The maximum training time was set to 12 hours on a Macintosh II microcomputer with a threshold of 10 % of the normalized input values (input-ouput) [2].

The ANN had the following features:

(i) Heteroassociative memory, for which the patterns on recall from the memory are purposely different from the input pattern, because the inputs and outputs are different and belong to different classes of information.

(ii) Delta rule type of learning, where the neuron weights are modified to reduce the difference between the desired output and the actual output of the processed element. The weights are changed in proportion to the error calculated. This rule also limits the learning, if the error at the output of the network is lower than a given threshold. The learning rates of those layers close to the output are set lower than the learning rates of the other layers.

(iii) A momentum term, which is used to smooth out the changes.

(iv) A sigmoid transfer function, which is a monotonically continuous mapping function.

The ANN predictions are in good agreement with the measured data. Figure 5 shows that correlation. Considering the difficulty of obtaining high quality data, we consider that the correlation factor is satisfactory.

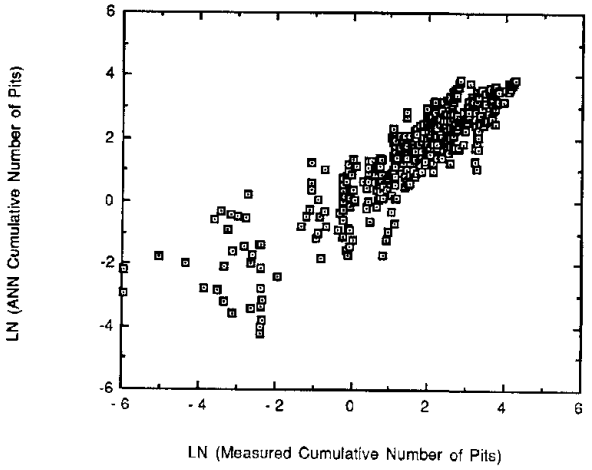


Fig. 5. Natural logarithm of the predicted ANN total number of pits versus natural logarithm of laboratory measured total number of pits. The measurements included several: applied potentials, observational times, oxidation states, and percent of minor alloying elements.

After the ANN was trained, it was used to make predictions of the number of pits at different applied potentials, observation times, oxidation states, and percentages of minor alloying elements in the film. The total number of pits predicted by the ANN decreased with increasing percentage of minor alloying elements in the film, and with increasing oxidation state of the alloying element (Figs. 6a and 6b). Behavior similar to that predicted by ANN was observed experimentally.

The ANN, once trained, can be used to explore the importance of the relationship between "output" and "inputs." We found that the results were strongly dependent on observation time (t^3, t), have a medium dependency on applied potential ($V^{1/2}$), and show weak dependencies on oxidation state ($Z^{1/3}$) and concentration of minor alloying element in the film ($[\%]^{1/4}$).

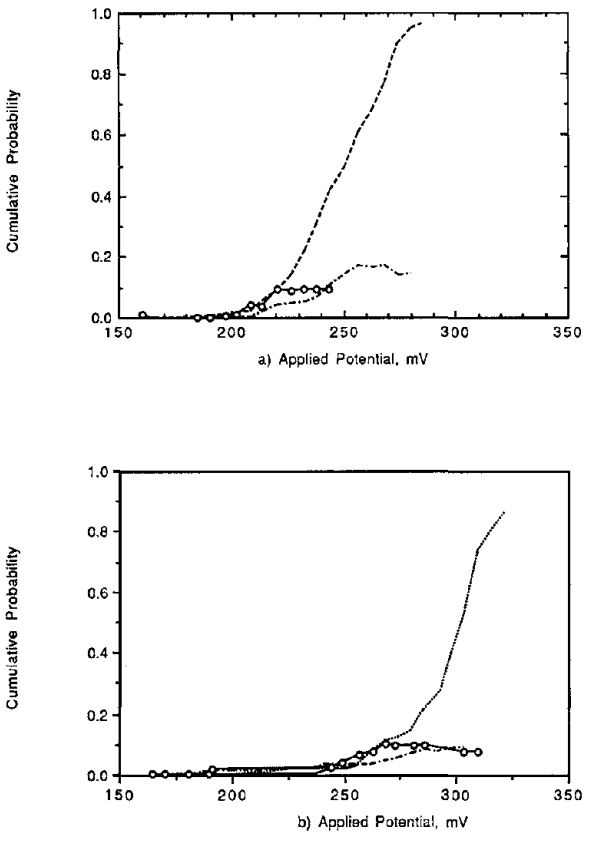


Fig. 6. ANN prediction of cumulative number of pits versus applied potential, at 50 s time of observation; and several percentiles of minor alloying elements. (6a) Oxidation state (4-2). (6b) Oxidation state (6-2).

6. Discussion

Figures 7a to 7d show the best predictions obtained with the three models compared with the laboratory data.

The deterministic model predicts that the cumulative probability at low applied potentials is not only described by a flat curve but that the curve is displaced to higher times. This prediction coincides with the experimental observations. The predictions with the deterministic model at high potentials indicated that the plateau corresponding to higher times is reached sooner than that measured. The deterministic model is the only model (compared to the other two models) that brings together an understanding of the problem as well a predictive tool. Another advantage that the deterministic model has over any nondeterministic model is that to fit the model, only an experimental datum point is necessary to calibrate the model to the data. The deterministic model was developed to predict damage and cumulative number of pits simultaneously. This last capability makes it very attractive to the user.

The results obtained with the probabilistic model are in general agreement with the experimental observations. As with the deterministic model, the plateau in cumulative damage is reached sooner than the measured one. However, the curves are flatter at lower potentials than at high potentials, but they are not displaced to higher times. The probabilistic approach needs a large data base, and the predictive capabilities are limited to the ranges of variables confined in the data base.

The ANN model describes the cumulative number of pits very close to the experimental measurements. The plateaus on cumulative damage correspond very well to the plateau obtained experimentally. The ANN predictions at low number of pits are inaccurate, but they are very close to the experimental observation at higher cumulative numbers of pits.

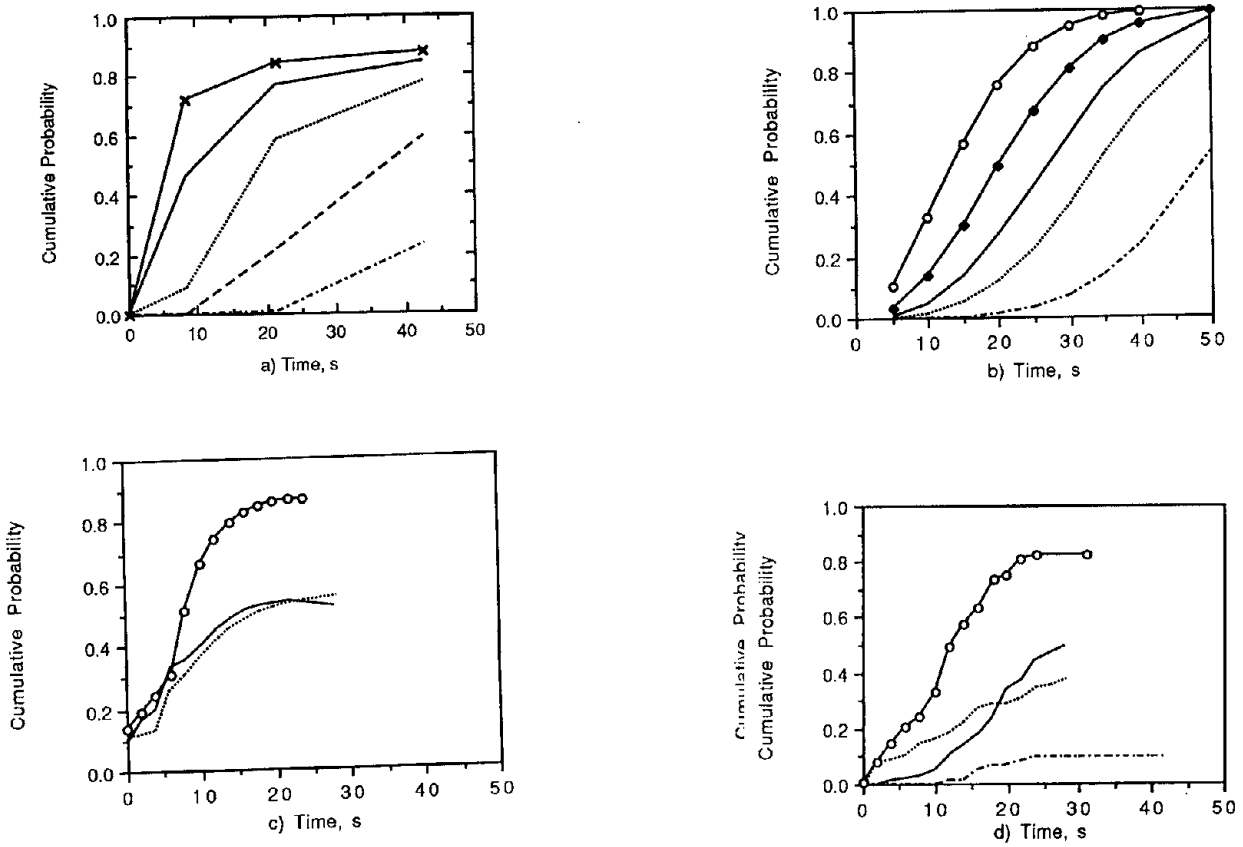


Fig. 7. Cumulative number of pits versus time of observation at several applied potentials; Oxidation state (3-2); and 3% of minor alloying element. (7a) Predictions using the deterministic model. (7b) Predictions using Weibull Distribution model with 2 fitting parameters. (7c) Predictions using the ANN. (7d) Laboratory measurements.

We conclude that little is learned about the phenomena when nondeterministic models are used; however, they can represent invaluable predictive tools. The statistical model is in a sense more demanding that an ANN model. It requires a robust and large data base. We found that an ANN can learn from "noisy" data, and that the range of prediction can be extended outside the range for which it was trained, if trained correctly [13]. Damage functions can also be calculated using the deterministic model based on the PDM and SVIM. The deterministic model does not need to have an extensive data base that includes pit depth distributions. On the other hand, the nondeterministic models need a large data base, and they are unable to make predictions of partial damage if the pit depth versus number of pits is not confined to the data base. Cumulative damage can be interpolated and extrapolated to other voltages and times, and to other applied potentials, for any of the three models. In general the results obtained with the three models were found to be in reasonable agreement with experimental data [12].

We do not intend to emphasize here the importance of deterministic models over nondeterministic models, but we have to keep in mind, when picking a model, to choose the one that best represents the observations and that is reasonably easy to solve. Clearly, the reliability of the extrapolation, in particular, depends critically on the quality of the data and on the veracity of the model.

Acknowledgments

The authors gratefully acknowledge the support of this work by the Department of Energy/Basic Energy Sciences.

7. References

[1] SRI Final Report, June 1989-July 1991, Material Research Laboratory, SRI International, Menlo Park, CA.
[2] R. Razgaitis, J. H. Payer, S. G. Talbert, B. Hindin, E. L. White, D. W. Locklin, R. A. Cudnik, and G. H. Stikford, Condensing Heat Exchanger for Residential/Commercial Furnaces Boilers, Phase IV, Battelle Report to DOE/BNL, BNL Report No. 5194
[3] M. Urquidi-Macdonald and D. D. Macdonald, Corrosion, to be submitted for publication (1994).
[4] J. L. Bogdanoff and F. Kozin, Probabilistic Models of Cumulative Damage, John Wiley & Sons (1930).
[5] C. Y. Chao, L. F. Lin, and D. D. Macdonald, J. Electrochem. Soc. **128**, 1187 (1981).
[6] L. F. Lin, C. Y. Chao, and D. D. Macdonald, J. Electrochem. Soc. **128**, 1194 (1981).
[7] D. D. Macdonald and M. Urquidi-Macdonald, Electrochem. Acta **31**, 1079 (1986).
[8] M. Urquidi-Macdonald and D. D. Macdonald, J. Electrochem. Soc. **132** (3), 555 (1986).
[9] M. Urquidi-Macdonald and D. D. Macdonald, J. Electrochem. Soc. **132**, 555 (1985).
[10] D. D. Macdonald, M. Urquidi-Macdonald, and S. Lenhart, (1986).
[11] T. R. Beck and R. C. Alkire, J. Electrochemical Soc. **126**, 1662 (1979).
[12] D. D. Macdonald and M. Urquidi-Macdonald, Corrosion **46**, 380 (1990).
[13] M. Urquidi-Macdonald, unpublished data, Engineering Science and Mechanics, Pennsylvania State University (1993).

About the authors: M. Urquidi-Macdonald is an associated professor at the department of Engineering Science and Mechanics of Penn State. Her primary interests are in mathematical modeling of damage and life prediction. D. D. Macdonald is the Director of Penn State's Center for Advanced Materials and is a Professor of Materials Science and Engineering. One of his principal research interests is the development of deterministic models for predicting damage due to localized corrosion.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 505 (1994)]

Identification of Failure Origin Through Testing and the Weibull Risk-of-Rupture

Volume 99 Number 4 July–August 1994

John A. Tesk, Martin Y. M. Chiang, and Spurgeon M. Keeny, III

National Institute of Standards and Technology,
Gaithersburg, MD 20899

Jun Tang

Shanghai Pharmaceutical Industry Design Institute,
Shanghai, China

and

Yuuji Sato

Hiroshima University School of Dentistry,
Hiroshima, Japan

The stress distribution in bond layers of two different thicknesses (50 μm and 200 μm) was calculated by finite element analysis for pairs of rectangular cross section metal bars bonded to each other and subjected to four point bending. These stresses were used to aid in identification of the failure origin by use of the Weibull risk-of-rupture (RR) function. By use of the stress distributions, the characteristic strength from 50 μm bond test specimens could be correlated with that for 200 μm bond test specimens when the failure was assumed to have an interfacial origin. The finite element meshes were refined twice and the ratios of characteristic strengths were recalculated and remained virtually unchanged by each of the mesh refinements. Hence, the identification of the interface as the failure origin remained consistent. Further, the use of stresses extrapolated to zero mesh size also produced the same ratios. Therefore, the RR calculations do not appear to be sensitive to the mesh sizes used for the stress calculations when the meshes are comparable or when changed in a comparable manner. The results show this method can be consistent and a useful adjunct for identification of failure origins.

Key words: failure analysis; failure in bending; failure origin; failure stress; failure stress and size effect; finite element analysis; finite element stress; origin of failure; Weibull analysis; Weibull hazard function; Weibull risk-of-rupture function.

Accepted: March 22, 1994

1. Introduction

In previous work [1,2] bending tests were conducted on adhesively bonded specimens of a dental alloy. The purpose was to determine:

a) how much the bond thickness influenced the test results;
b) whether the failure origins appeared to be the same for the two different bond thicknesses employed;
c) failure origin through analysis using the risk-of-rupture (RR) function.

In this paper, the finite element method was used to arrive at the stress distributions used for the RR analyses that employ the Weibull risk-of-rupture function[1] [3]. An ancillary purpose, therefore, was to ascertain how sensitive the analyses were to the fineness of the finite-element mesh and, hence, whether the method can be applied with confidence to the analyses conducted for a, b, and c above.

[1] Today, a risk-of-rupture function, defined by Weibull, would be recognized as a hazard function.

2. Materials and Methods

The failure of brittle materials is typically catastrophic and in many instances the failure stresses obtained from a set of test specimens follow a Weibull distribution. For a homogeneous isotropic material subjected to a uniform tensile stress, σ, the probability of failure, $P(\sigma)$, is given by:

$$P(\sigma)=1-\exp-[\delta(\sigma/\sigma_{0,u})^m], \tag{1}$$

where $\sigma_{0,u}$ is a characteristic strength for a specimen of unit dimension; m is a shape factor (Weibull modulus); and δ is a size factor (the ratio of the failure originating dimension to a unit dimension of the same kind) and represents the volume, V, area, A, or other dimension in which reside the flaws from which the failure originates [3]. Eq. (1) is often written as:

$$P(\sigma)=1-\exp-[(\sigma/\sigma_0)^m], \tag{1a}$$

where the size of the specimen, δ, is subsumed into σ_0. This form of the equation is commonly used when analyzing test data and the effects of specimen size are ignored. It is clear from Eq. (1) that for specimens of two sizes, δ_1 and δ_2, with the same failure origins and presenting the same distribution of failure stresses ($m_1=m_2$), there will be two different values of σ_0 for Eq. (1a), with the larger size, call it δ_1, leading to a value, $\sigma_{0,1}$, that is less than $\sigma_{0,2}$.

For such specimen sets, the relation between the characteristic strengths calculated by Eq. (1a) is [4]:

$$\sigma_{0,2}=\sigma_{0,1}[\delta_1/\delta_2]^{1/m}. \tag{2}$$

For a nonuniform tensile stress field, a more general form of Eq. (1) is necessary:

$$P(\sigma)=1-\exp-[\textstyle\int_\delta(\sigma/\sigma_0)^m d\delta], \tag{3}$$

where the region of integration over δ is the region critical to failure (rupture) and it can be in one, two, or three dimensions. Then the relationship between the values of σ_0 [Eq. (1a)] as determined from experiments on sets of specimens having either one or the other of the bond thicknesses, is given by the ratios of the exponents of Eq. (3), i.e.,

$$\frac{(\sigma/\sigma_{0,1})^m}{(\sigma/\sigma_{0,2})^m}=\frac{\int_{\delta 1}(\sigma_1/\sigma_{0,u})^m d\delta_1}{\int_{\delta 2}(\sigma_2/\sigma_{0,u})^m d\delta_2}. \tag{4}$$

Canceling terms in $\sigma_{0,u}$ on the right side and σ on the left side leads to

$$\frac{\sigma_{0,2}}{\sigma_{0,1}}=\left[\frac{\int_{\delta 1}(\sigma_1)^m d\delta_1}{\int_{\delta 2}(\sigma_2)^m d\delta_2}\right]^{1/m} \tag{4a}$$

When $\sigma_i(\delta)$ is not known as an explicit function, the relationship between $\sigma_{0,1}$ and $\sigma_{0,2}$ may, in principle, be approximated by [3]):

$$\sigma_{0,2}=\sigma_{0,1}[(\textstyle\sum_j \sigma_{1,j}^m \Delta\delta_{1,j})/(\sum_i \sigma_{2,i}^m \Delta\delta_{2,i})]^{1/m}, \tag{5}$$

and the validity of the approximation must be checked by computation. Here the summations are over all the elements considered to be involved with the failure (interface, volume, etc.) and the stresses can be evaluated by the finite-element method of analysis.

Note: The stress field in the bond region is typically three dimensional; the analyses of this paper utilize unidirectional tensile stresses because alterations to the principal stresses were found to be minor and may be ignored. We also note that for a variety of reasons (plasticity, change in composition, properties, or flaw populations) this analysis method may not apply for very thin bond layers approaching micrometers or less.

Each assumed failure origin for a specimen has its own specific δ with its associated stresses. When the ratios of volumes, surface areas, interface areas, edge lengths etc. (any dimensions containing the flaws from which failures may originate) are properly chosen to be different for experimental tests, only one set of δ_i's, σ_i's, δ_j's, and σ_j's should produce coincidence between the experimentally determined ratio of σ_0's and the ratios of either the integrals shown by Eq. (4a) or the summations as shown in Eq. (5).

As with any analysis employing the finite element (FE) method for determination of the stresses, a critical question arises as to the FE-

mesh sensitivity of Eq. (4). If sensitive, then the method would not, in actuality, be useful for the correlation of results from differently sized specimens.

To provide insight into the ability of this approach to identify sources of failure, rectangular bond specimens as shown in Fig. 1 were prepared for testing in four-point bending, with either 50 μm or 200 μm bond-thicknesses. Rectangular bars of the bulk bonding material were also tested in three-point bending. The details of specimen preparation were given in a presentation by Keeny et al. [1]. The number of specimens and the results for each test series are shown in Table 1.

A three-dimensional, finite-element elasticity model[2] was used for evaluation of the stress distribution throughout the volume of the bond region.

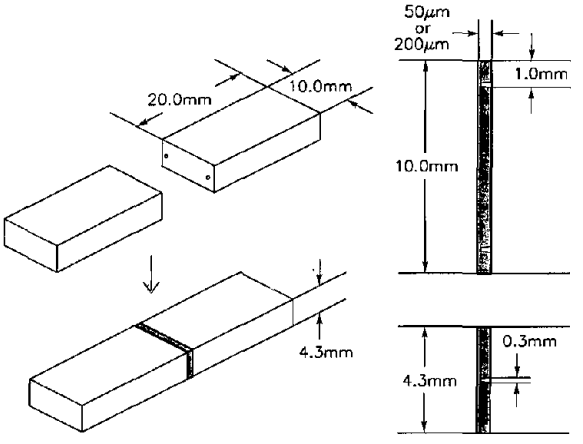


Fig. 1. Specimen used for bond testing. The shaded area represents the bonding material between rectangular beams of alloy that were bonded together. Two small projections were used to control the width of the bond at either 50 μm or 200 μm.

The original bond model (Fig. 2) consisted of 2,197 elements in one quadrant of the specimen which had three planes of symmetry. Subsequent refinements of interface and surface elements led to elements 1/2 and 1/3 the original size. The validity of the model was checked by comparison of the finite-element results for a homogeneous beam with the analytical solution. Examples of how the bending tensile stresses change as a function of the thickness of the bond layer are shown in Figs. 3, 4, and 5, for which $E_b \cong E_a/50$ where: E_b is Young's modulus for the bonding material; and E_a is Young's modulus for the alloy.

If the failure stresses are referenced to the stresses along the surface, the ratio of the operative (effective) dimensions, δ, from which the failures originate are given by Eq. (6):

$$\frac{\delta_{\text{eff},1}}{\delta_{\text{eff},2}} = \left(\frac{\sigma_{R,2}}{\sigma_{R,1}}\right)^m \frac{\sum_j \sigma_{1,j}^m \Delta\delta_{1,j}}{\sum_i \sigma_{2,i}^m \Delta\delta_{2,i}}, \qquad (6)$$

where σ_R denotes the reference stress.

For these calculations the bending stress at the surface was used as the reference stress for calculations of bending strength and $\sigma_{R,2} = \sigma_{R,1}$. Then, from Eq. (2) and Eq. (6)

$$\sigma_{0,2}/\sigma_{0,1} = (\delta_{\text{eff},1}/\delta_{\text{eff},2})^{1/m} = \left[\frac{\sum_j \sigma_{1,j}^m \Delta\delta_{1,j}}{\sum_i \sigma_{2,i}^m \Delta\delta_{2,i}}\right]^{1/m}, \qquad (7)$$

which is equivalent to Eq. (5).

Table 1. Uncorrected Weibull parameters and mean strengths

Test	Gap (μm)	N^a	σ_0 (MPa) {range 95%}[b]	m {range 95%}
4-Pt Bend	50	25	110 {106–115}	11.3 {8.2–14.0}
"	200	25	107 {103–111}	11.5 {8.3–14.2}
3-Pt Bend	Bars of bulk bond material	54	85 {81–89}	6.8 {5.5–7.93}

[a] N = the number of specimens.
[b] { } = the associated confidence bounds on σ_0 and m as determined from the data.

[2] Developed at NIST.

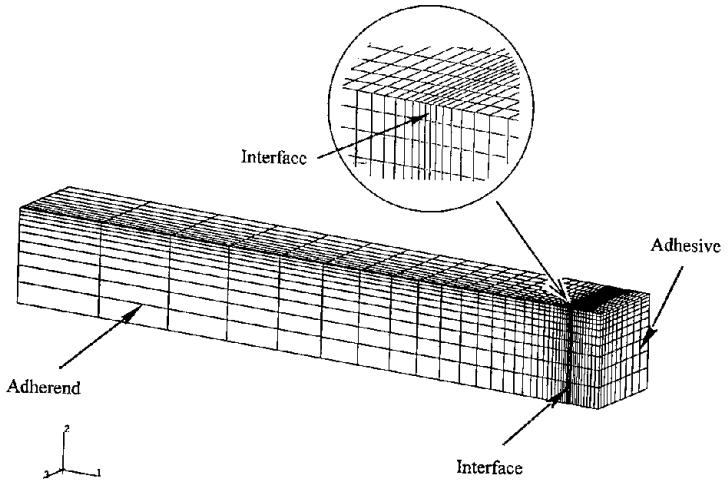


Fig. 2. A 1/8 section of a three-dimensional model for finite element calculations of stress. The specimens (Fig. 1) had three planes of symmetry.

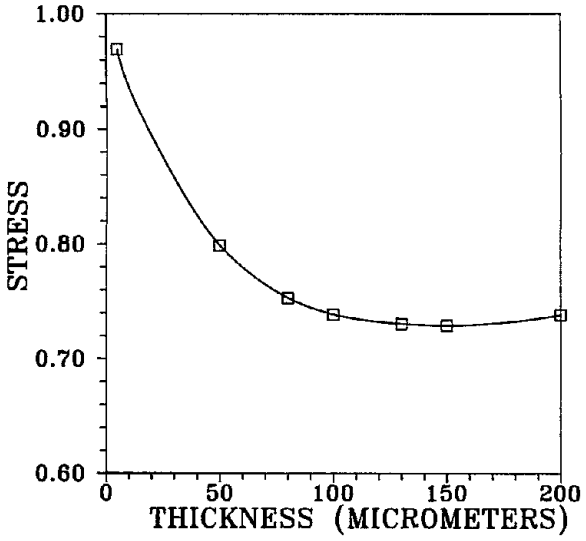


Fig. 3. Result of finite element calculation of the near-surface bending stress at the bond midplane that bisects the bend specimens into symmetrical halves.

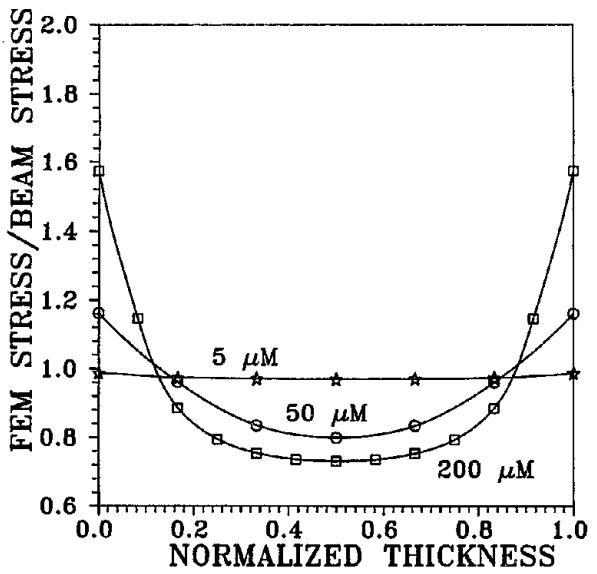


Fig. 4. Surface bending stress across the surface of the bond material, from one interface junction to the opposite one, as calculated for three thicknesses. The 5 μm thickness is presented to illustrate the trend toward beam stress calculations as the thickness approaches zero. The deviations from beam theory calculations are appreciable for thick bond specimens, showing the need to use the more robust finite element method for the failure analysis employing the RR function.

3. Results

By the use of the right-hand side of Eq. (7) and the finite-element-derived stress distributions, characteristic strength ratios were calculated for four potential regions (Table 2) where the failure of the bond could originate, i.e., volume, surface, interface, and interface-line-junction failure origins. These were then compared with the results obtained from the left hand side of Eq. (7). For these calculations, a value of $m = 11.5$ was used [in Table 1, m was obtained from Newton-Raphson iteration for fitting experimental data to Eq. (1)].

The ratio of the experimental characteristic strengths has a 95% confidence range of 0.955 to 1.11. When this ratio is compared with ratios calculated from the finite element analyses, the interface (Table 2) is identified as the origin of failures,

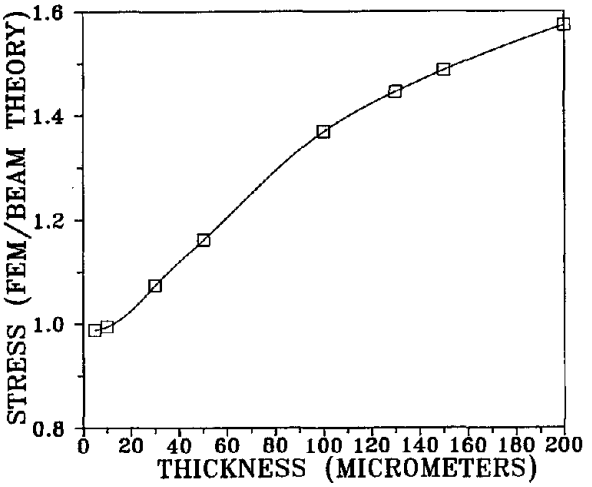


Fig. 5. Finite element calculated stress for the adherend-adherent interface surface junction line, illustrating the dramatic effect of bond thickness on interface stresses.

Table 2. Ratios of characteristic strengths calculated from risks of rupture: (strength, 200 μm)/(strength, 50 μm). Four-Point-bending

Assumed failure origin	Coarse mesh	Refined mesh 1	Refined mesh 2
Volume	1.12	1.12	1.12
Interface	.995	1.01	1.06
Surface	1.26	1.36	1.32
Interface-surface junction line	1.24	1.25	1.41
Interface-surface junction line (extrapolated)[a]	1.38	1.37	1.37

[a] Obtained from extrapolations of finite element derived stresses at centroids to the interface between alloy and bond layer.

with all other failure origins excluded. The row with the next closest match of strength ratios is that for volume failures which, with a ratio of 1.12, lies just outside the 95% confidence range, so this argument, by itself, is somewhat unconvincing. However, volume failures are ruled out because the m value of 11.5 from the bond tests differs, at the 90% confidence level, from the value of $m = 6.8$ which was obtained from the bulk specimen test data. The bulk specimens can fail only by volume or surface failures. The strength ratio calculations for bond specimens rule out surface failures. The m value differences then are used to rule out volume failures.

Hence, the most reasonable explanation is that the bond specimens fail by interfacial failures. This is consistent with features of the failed specimens which always presented regions showing interfacial debonding.

There is some possibility that the strength of the bond itself would depend on the bond thickness due to a change in material response (i.e., formation of plasticity). Such effects obviously cannot be dealt with by the linear elastic analysis presented and within the confines of this analysis, interfacially initiated failure is concluded.

4. Summary

An analysis by Weibull RR for bonded specimens of two different sizes tested in bending has shown:

1) Correlations between characteristic strengths, σ_0's, were possible through the use of finite-element derived stresses in the RR analysis.
2) The correlations were not sensitive to the particular mesh chosen.
3) For the interface, surface, and interface-surface junction line, the stress ratio calculations employing the element centroid stresses are not significantly changed by use of stresses from extrapolations to the interface. The largest difference is for the interface-surface junction line and these are shown in Table 2.
4) Because the absolute magnitude of each RR calculation changes, mesh of the same size and configuration must be used for each set of comparisons.
5) The origins of failures can be determined by suitable testing and analysis of different size bond specimens and bulk specimens of the bonding material. This involves the use of a combined approach, analysis of the σ_0's and m values.
6) The determination of failure origin by this approach can be useful for focusing attention on the proper parameters if improvements in system strength or performance are sought.

Acknowledgments

The authors wish to acknowledge the helpful discussions with Professor Sam Saunders, Washington State University, and Dr. James Lechner, National Institute of Standards and Technology. This work was partially supported by Interagency Agreement 2Y01 DE 3001 with the National Institute of Dental Research, Bethesda, MD 20892.

5. References

[1] S. Keeny, Y. Sato, and J. A. Tesk, Bond Strength of Resin-Bonded Systems in Tension and Bending, J. Dent. Res., 69, Special Issue, Abst. # 796 (1990) p. 208.
[2] J. A. Tesk, M. Y. M. Chiang, J. Tang, and S. Keeny, Stress in an "Adhesive" Bond Layer, J. Dent. Res., 69, Special Issue, Abst. # 1998 (1990) p. 358.
[3] W. Weibull, A Statistical Theory of the Strength of Materials, Ingeniorsvelenskapsakademiens Handlinger NR 151 (1939) pp. 5–45.
[4] D. G. S. Davies, The Statistical Approach to Engineering Design in Normally Stresses Brittle Materials, J. Am. Ceram. Soc. **47** (6), 268–274 (1973).

About the authors: *John A. Tesk is a General Physical Scientist and leader of the Dental and Medical Materials Group in the Polymers Division. Martin Chiang is a Mechanical Engineer in the Polymers Division. Spurgeon Keeny, III is a Biomedical Engineer, formerly with the Polymers Division. Jun Tang and Yuuji Sato are Guest Scientists with the Polymers Division. The National Institute of Standards and Technology is an agency of the Technology Administration, U.S. Department of Commerce.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 511 (1994)]

Conical Extremes of a Multivariate Sample

Volume 99 Number 4 July–August 1994

Alexander V. Gnedin

University of Göttingen
Göttingen, Germany

We introduce multivariate extremes in the direction of a given cone. Convergence results for the number of the kth extremes are obtained for sampling from a distribution having asymptotically independent radial and spherical components and regularly varying tail of the radial component.

Key words: conical extremes; multivariate extremes; regular variation.

Accepted: March 22, 1994

1. Introduction

Let $\mathcal{H}_n = \{X_1, \ldots, X_n\}$ be a point set of independent identically distributed d-dimensional random vectors sampled from the probability measure μ, and K be a punctured at the origin *cone* in $\mathbf{R}^d$, $d > 1$. We define *the kth layer* as

$$\mathcal{L}^{(k)}(\mathcal{H}_n) = \{X_i : \#(K_{X_i} \cap \mathcal{H}_n) = k-1\} \quad k = 1,2, \ldots,$$

where $K_x = x+K$ is the translated cone with vertex in $x \in \mathbf{R}^d$. Intuitively, the kth layer is the set of the kth extremes of $\mathcal{H}_n$ in the direction K. The prime examples we have in mind are (1) *the Pareto-optimal points* corresponding to the first layer in the direction of the positive orthant, and (2) *the total maximum*, which may be considered as the first layer in the direction of the cone, complement to the negative orthant. We are interested here in the distributions of random variables

$$V_n^{(k)} = \# \mathcal{L}^{(k)}(\mathcal{H}_n),$$

counting the number of points in the kth layer. These distributions depend essentially on both K and μ.

From a more general viewpoint, the first layer can be regarded as the set of *maximal elements* [4] with respect to the binary relation $\mathcal{R}$ in $\mathbf{R}^d$ defined as $x\mathcal{R}y \Leftrightarrow x-y \in K$. Alternatively, any scale and translation invariant binary relation generates a cone by setting $K = \{x \in \mathbf{R}^d: x\mathcal{R}0\}$ and the maximal elements are conical extremes.

Two above cases of the counting problem have been considered in the literature under the assumption that μ is either a product of one-dimensional marginal measures or a multivariate normal distribution [2,10,11,12]. It is well known, for example, that if μ is a product measure in $\mathbf{R}^d$ then the average number of Pareto points is of the order of $(\log n)^{d-1}$, while the probability that the multiple maximum exists is n^{1-d}.

In this paper we focus on a class of distributions μ already studied in connection with multivariate extreme-value theory [8] and statistics of convex hulls [1,5,6,9]. These distributions are characterized by regular variation of the tail of the radial component and asymptotical independence of radial and angular components. We show that typically the $V_n^{(k)}$'s converge in distribution and the expectations have finite limits as $n \to \infty$. In the special case of slow variation we calculate explicitly the limiting distributions.

2. Preliminaries

We define a *cone* as a punctured at the origin, scale-invariant Borel set in $\mathbf{R}^d$, i.e., $0 \notin K$, $tK = K\ \forall t > 0$. Each cone is uniquely determined by its spherical base $S_+ = K \cap S$, where S denotes the unit sphere. We associate with K also the spherical set S_- obtained by reflection about the origin, $S_0 = S \setminus (S_+ \cup S_-)$ and $S_\pm = S_+ \cap S_-$. The cone with spherical base $C \subset S$ will be denoted cone (C).

Set $B_r = \{x \in \mathbf{R}^d : \|x\| \le r\}$, $B_r^c = \mathbf{R}^d \setminus B_r$ and $A_{r,C} = \text{conc}\ (C) \cap B_r^c$.

We fix in what follows a cone K and a multidimensional probability distribution μ satisfying the following conditions:

(i) There exists $\alpha \ge 0$ and a probability measure ρ on S such that

$$\lim_{t\to\infty} \frac{\mu(B_{tr}^c)}{\mu(B_t^c)} = r^{-\alpha} \quad r > 0, \tag{1}$$

(ii) For all ρ-continuous $C \subset S$

$$\lim_{t\to\infty} \frac{\mu(A_{t,C})}{\mu(B_t^c)} = \rho(C), \tag{2}$$

(iii) ρ (int S_+) > 0, and

(iv) μ has no atom at the origin.

Consider an *iid* sample from μ, $\mathcal{H}_n = \{X_1, \ldots, X_n\}$, represented in the polar form as the product of radial and spherical components: $X_i = R_i Z_i$, where $R_i = \|X_i\|$, $Z_i = X_i/\|X_i\|$. The above conditions on μ have a natural probabilistic interpretation. Condition (i) means that the distribution function of the radial component,

$$F(r) \stackrel{\text{def}}{=} \mu\,(B_r),$$

has a regularly varying tail. Condition (ii) is translated as

$$\lim_{r\to\infty} \mathbf{P}\{Z_1 \in \cdot\, | R_1 > r\} = \rho\,(\cdot)$$

and is to be interpreted as the asymptotic independence of radial and spherical components, where the limiting distribution ρ does not disappear in the interior of S_+ (condition (iii)). The last condition is not essential and assumed for technical reasons.

Given a Borel set $B \subset \mathbf{R}^d$, we represent the number of the kth layer points in B as the sum of random indicators

$$\#\mathcal{L}^{(k)}(\mathcal{H}_n \cap B) = \sum_{i=1}^{n} 1_{\{X_i \in \mathcal{L}^{(k)}(\mathcal{H}_n) \cap B\}},$$

and using the *iid* assumption write for the expectations

$$\mathbf{E}\#\mathcal{L}^{(k)}(\mathcal{H}_n \cap B) = n\mathbf{P}\{X_1 \in \mathcal{L}^{(k)}(\mathcal{H}_n) \cap B\} =$$

$$n\binom{n-1}{k-1}\mathbf{P}\{X_1 \in B; X_2, \ldots, X_k \in K_{X_1}; X_{k+1}, \ldots, X_n \notin K_{X_1}\} =$$

$$n\binom{n-1}{k-1}\int_B (\mu(K_x))^{k-1}(1-\mu(K_x))^{n-k}\,\mathrm{d}\mu(x)\ . \tag{3}$$

The following lemmas will be used to estimate these integrals.

Lemma 1. *There exists* $\tau > 0$ *such that* $\mu(K_x) > \tau(1-\mu(B_{\|x\|}))$ *for all* $x \in \mathbf{R}^d$.

Proof. Consider first the case where there exists a linear isomorphism which maps K onto the positive orthant. Let y be the inverse image of the vector $(1, \ldots, 1)$ under this isomorphism. By convexity, $K_y \subset K_x$ for all $x \in B_1$.

Condition (iii) allows one to select a compact ρ-continuous set $C \subset S_+$ with $\rho\,(C) > 0$. It is easy to see that $y \in \text{int}\,K$, the sets K_{sy}, $s>0$, are increasing as $s \downarrow 0$ and $\cup_{s>0} K_{sy} = \text{int}\,K$. It follows that $C \subset K_{sy}$ for sufficiently small s. Furthermore, for small s we have also $A_{1,C} \subset K_{sy}$. Indeed, the points of $A_{1,C}$ are representable as tx, with $t > 1$, $x \in C$, thus, by convexity, $x \in K_{sy}$ implies $tx \in K_{tsy} \subset K_{sy}$. Homogeneity implies $A_{t/s,C} \subset K_{ty}$. It follows now from Eqs. (1) and (2) that

$$\frac{\mu(K_{ry})}{\mu(B_r^c)} > \frac{\mu(A_{r/s,C})}{\mu(B_r^c)} = \frac{\mu(A_{r/s,C})}{\mu(B_{r/s}^c)}\frac{\mu(B_{r/s}^s)}{\mu(B_r^c)}$$

$$\to s^\alpha \rho(C),\ r\to\infty\ . \tag{4}$$

From $K_x \supset K_{\|x\|y}$ we derive for sufficiently large r_0 and $\|x\| > r_0$ that

$$\frac{\mu(K_x)}{\mu(B_{\|x\|}^c)} \ge \frac{\mu(K_{\|x\|y})}{\mu(B_{\|x\|}^c)} > \frac{1}{2}s^\alpha\rho(C)\ .$$

For $x \in B_{r_0}$ we have $K_x \supset K_{r_0 y}$, therefore Eq. (4) along with the inclusion $K_{tr_0 y} \supset K_{r_0 y}$, $t > 1$, implies

$$\frac{\mu(K_x)}{\mu(B_{\|x\|}^c)} \ge \frac{\mu(K_{r_0 y})}{\mu(B_{\|x\|}^c)} > \mu(K_{r_0 y}) > 0\ .$$

The assertion follows in this case by setting $\tau = \min(\mu(K_{r_0 y}), \frac{1}{2}s^\alpha\rho(C))$.

For arbitrary K one can find a smaller cone $K' \subset K$, which is linearly isomorphic to the positive orthant and still has the interior of its spherical base of positive ρ-measure. This is possible since the spherical d-simplexes build a measure-generating class on S. It remains to note that $\mu(K_x) \ge \mu(K'_x)$ for any translation, whence the estimate holds in general. □

Lemma 2. *If* $\mathbf{E}\, V_n^{(1)}$ *has a limiting value* $v \in [0,\infty)$ *then all* $\mathbf{E}V_n^{(k)}$, $k = 2,3,\ldots$ *converge to this limit as* $n \to \infty$.

Proof. Let $m(t)$, $t \in [0,1]$, be the distribution function of the image measure induced by the mapping $x \mapsto \mu(K_x)$. Changing variables transform Eq. (3) to the one-dimensional Lebesgue-Stiltjes integral

$$\mathbf{E}V_n^{(k)} = n\binom{n-1}{k-1}\int_0^1 t^{k-1}(1-t)^{n-k}\mathrm{d}m(t).$$

A slight modification of the standard Tauberian theorem as found in [14] assures that the limiting value of this integral for $k = 1$ exists iff $m(t)$ is left-differentiable at $t = 1$, in which case the limit and the derivative have the same value. Applying this theorem in the reverse direction one can easily see that all the $\mathbf{E}V_n^{(k)}$'s must have the same limit. □

Lemma 3. *Assume* γ_n *is an increasing sequence such that* $\lim_{n\to\infty} n(1-F(\gamma_n)) \to \gamma$, $\gamma > 0$, *then*

$$\limsup_{n\to\infty} \mathbf{E}V_n^{(k)} \leq \tau^{-1}, \tag{5}$$

$$\limsup_{n\to\infty} \mathbf{E}\,\#(\mathscr{L}^{(1)}(\mathscr{H}_n) \cap B_{\gamma n}) \leq \mathrm{e}^{-\gamma\tau}\tau^{-1}, \tag{6}$$

with τ *determined by Lemma* 1.

Proof. Set

$$f(t) = \int_{F(r)\leq t} \mathrm{d}F(r),\ t \in [0,1].$$

Regular variation of F at infinity implies readily that for all sufficiently distant discontinuity points the ratio (jump-size)/(distribution tail) is close to zero. It follows that $(1-f(t))/(1-t) \to 1$ as $t \uparrow 1$ (for continuous F this is obvious since $f(t) = t$). Lemma 1, a change of variables and the Tauberian theorem yield

$$\mathbf{E}V_n^{(1)} = n\int_{\mathbf{R}^d} (1-\mu(K_x))^{n-1}\mathrm{d}\mu(x) \leq$$

$$n\int_{\mathbf{R}^d} (1-\tau(1-\mu(B_{\|x\|}))^{n-1}\mathrm{d}\mu(x) =$$

$$\int_{\mathbf{R}} (1-\tau(1-F(r))^{n-1}\mathrm{d}F(r) = \int_0^1 (1-\tau(1-t))^{n-1}\mathrm{d}f(r) \to \tau^{-1}.$$

Similarly,

$$\limsup_{n\to\infty} \mathbf{E}\#(\mathscr{L}^{(1)}(\mathscr{H}_n)\cap B_{\gamma_n}) \leq n\int_0^{\gamma_n} (1-\tau(1-F(r))^{n-1}$$

$$\mathrm{d}F(r) =$$

$$n\int_0^{F(\gamma_n)} (1-\tau(1-t))^{n-1}\mathrm{d}f(t) \sim n\int_0^{\gamma} (1-\tau(1-t))^{n-1}\,\mathrm{d}t$$

$$\to \mathrm{e}^{-\gamma\tau}\tau^{-1},$$

where the equivalence can be justified by partial integration. □

3. Pareto-Tails: $\alpha > 0$

In this section we study the limiting behaviour of $V_n^{(k)}$ under the assumption that the regular variation index α in Eq. (1) is positive. Our plan is to translate Eqs. (1) and (2) into the convergence, of a suitably normalized sample, to a Poisson process[6,15] and then apply a continuity argument to prove also the convergence of the $V_n^{(k)}$'s.

Compactify $\mathbf{R}^d$ by adjoining the infinite point ∞ and then puncture in the origin. The resulting topological space, say $\hat{\mathbf{R}}^d$, is isomorphic to $\mathbf{R}^d$ and canonically embedded into its compactification, bounded from the origin Borel sets $B \subset \hat{\mathbf{R}}^d$ being relatively compact. We endow the space $M(\hat{\mathbf{R}}^d)$ of Radon measures with the vague topology: $m_n \xrightarrow{v} m$ iff $m_n(B) \to m(B)$ for all relative compacts.

There exists a sequence of positive constants $a_n \to \infty$ such that the measures $\nu_n(\cdot) \stackrel{\mathrm{def}}{=} n\mu(a_n\,\cdot)$ converge vaguely to the measure ν determined by

$$\nu(A_{r,C}) = r^{-\alpha}\,\rho\,(C),\ \ \nu(\{\infty\}) = 0. \tag{7}$$

The limiting measure is in $M(\hat{\mathbf{R}}^d)$, being infinite on balls centered at the origin as well as on the sets cone (C) with $\rho(C) > 0$. In particular, condition (iii) implies $\nu(\mathrm{int}\,K) = \infty$. Clearly, ν is a product measure in polar coordinates and has no atoms.

Let ξ be a Poisson point process in $\hat{\mathbf{R}}^d$ with intensity measure ν, and ξ_n be the random element of $M(\hat{\mathbf{R}}^d)$ associated with the scaled sample $a_n^{-1}\mathscr{H}_n$. Obviously, the operation of taking a layer commutes with rescaling: $\mathscr{L}^{(k)}(a\mathscr{H}_n) = a\mathscr{L}^{(k)}(\mathscr{H}_n)$, $a > 0$, therefore the number of points in each layer remains invariant under scale transformations. One can expect in this situation that $V_n^{(k)}$ converges in some sense to an analogous functional of the Poisson process.

Define the kth layer of the Poisson sample as

$$\mathscr{L}^{(k)}(\xi) = \{x \in \hat{\mathbf{R}}^d : \xi\,(\{x\}) = 1,\ \xi\,(K_x) = k-1\},$$

and denote $V^{(k)} = \#\mathscr{L}^{(k)}(\xi)$ the number of points in the kth layer.

Using the polar representation, $x = rz$, and homogeneity we can write $\nu(K_x) = r^{-\alpha}\phi(z)$, where $\phi(z) = \nu(K_z)$ is a function of the spherical argument. Using Palm probabilities and integrating along radial rays we represent the expectations as

$$\mathbf{E}V^{(k)} = \int_{\hat{\mathbf{R}}^d} e^{-\nu(K_x)} \frac{(\nu(K_x))^{k-1}}{(k-1)!} d\nu(x) =$$

$$\int_S \int_{\mathbf{R}_+} e^{-r^{-\alpha}\phi(z)} \frac{(r^{-\alpha}\phi(z))^{k-1}}{(k-1)!} d\rho(z) d(-r^{-\alpha}) =$$

$$\int_S \frac{d\rho(z)}{\phi(z)}. \tag{8}$$

The resulting integral does not depend on k, as it is suggested by Lemma 2. The integration area can be reduced to $S \backslash \text{int } S_-$ since ϕ is infinite on int S_-.

The following lemma is found in [5].

Lemma 4. *Let E be a locally compact, Hausdorff and separable space; $h_0, h_1, \ldots$ be a uniformly bounded sequence of real measurable functions commonly supported by a relatively compact set; and $m_0, m_1, \ldots$ be a sequence of Radon measures on E such that $m_n \xrightarrow{v} m_0$. The set $D = \{x \in E : \exists\{x_n\}, x_n \to x, h_n(x_n) \not\to h(x)\}$ is measurable and if $m_0(D) = 0$ then $\int h_n dm_n \to \int h_0 dm_0$.*

Now we are ready to prove a convergence result.

Theorem 1. *Assume* (i)–(iv), $\alpha > 0$, *and*

$$\nu(-\partial K) = 0, \tag{9}$$

$$\nu \times \nu\{(x,y) \in \hat{\mathbf{R}}^d \times \hat{\mathbf{R}}^d : (x-y) \in \partial K\} = 0. \tag{10}$$

Then for all $k = 1,2,\ldots$

$$(V_n^{(1)}, \ldots, V_n^{(k)}) \xrightarrow{d} (V^{(1)}, \ldots V^{(k)}), \tag{11}$$

$$\mathbf{E}V_n^{(k)} \to \mathbf{E}V^{(k)}, \; n \to \infty. \tag{12}$$

Proof. By Skorohod's theorem we can find random point measures $\hat{\xi}_n, \xi \in M(\hat{\mathbf{R}}^d)$ satisfying $\hat{\xi}_n \overset{d}{=} \xi_n$, $\hat{\xi} \overset{d}{=} \xi$ and $\hat{\xi}_n \xrightarrow{v} \hat{\xi}$ a.s. Thus to prove the convergence in distribution (10) it suffices to consider the case $\xi_n \xrightarrow{v} \xi$ a.s. In what follows we fix a typical realization of ξ and assume n sufficiently large.

Since $\nu(\text{int } K) = \infty$, ξ lays in the cone interior infinitely many points. Select k of them, say $x_1, \ldots, x_k$. Pick r sufficiently small to satisfy $B_r \subset \cap_{j=1}^k -K_{x_j}$ as well as $\xi(\partial B_r) = 0$ and also $B_r \cap \{x_1, \ldots, x_k\} = \emptyset$. The complement B_r^c is relatively compact hence the processes ξ_n and ξ have there a finite number of points, say $y_{n,1}, \ldots, y_{n,p}$ and $y_1, \ldots, y_p$, respectively. These points may be labeled so that $y_{n,i} \to y_i$, as it follows from the vague convergence. By the construction, any translated cone K_x with $x \in B_r$ contains at least k points of ξ, thus $B_r \cap \mathcal{L}^{(j)}(\xi) = \emptyset$ and also $B_r \cap \mathcal{L}^{(j)}(\xi_n) = \emptyset$ for $j = 1, \ldots, k$.

For $y_i \in -\text{int } K$, the cone K_{y_i} contains a vicinity of the origin, where ξ has infinitely many points. Therefore $y_i \notin \cap_{j=1}^k \mathcal{L}^{(k)}(\xi)$, $y_{n,i} \notin \cap_{j=1}^k \mathcal{L}^{(k)}(\xi_n)$.

The condition shown in Eq. (9) assures that no one of $y_1, \ldots, y_p$ lies on $-\partial K$, almost surely.

For $y_i \notin -\text{cl } K$, the shifted cone K_{y_i} is bounded from the origin. Therefore there exists an open vicinity of $\text{cl}(\cup_{i=1}^p K_{y_i})$ which is still relatively compact, and hence contains at most a finite number of points in addition to $y_1, \ldots, y_p$. By Eq. (10), $\xi(\partial K_{y_j}) = 0$ a.s. Again the poinwise convergence implies $\xi(K_{y_{n,i}}) = \xi(K_{y_i})$, whence $(V_n^{(1)}, \ldots, V_n^{(k)}) = (V^{(1)}, \ldots, V^{(k)})$ and thus (11).

Now turn to the convergence in mean. It is enough to prove Eq. (12) for the first layer, $k = 1$. It is easy to see that

$$\mathbf{E}V_n^{(1)} \sim \int \exp(-\nu_n(K_x)) d\nu_n(x) = \int_{B_r} (\ldots) d\nu_n +$$

$$\int_{B_r^c} (\ldots) d\nu_n \quad r > 0$$

Take a point $x \notin -\text{cl } K$ and a sequence $x_n \to x$ and consider the indicator functions of the sets K_x and K_{x_n} as the h's in Lemma 4. The divergence set D is ∂K_x, whence by (9) and the lemma $\nu_n(K_{x_n}) \to \nu(K_x)$.

For $x \in -\text{int } K$, $x_n \to x$ we have $\nu_n(K_{x_n})$ to $\nu(K_x) = \infty$ since K_{x_n} contains some fixed vicinity of the origin,for all sufficiently large n. Therefore in this case also $\nu_n(K_{x_n}) \to \nu(K_x) = \infty$.

To make further use of Lemma 4, consider this time the functions $h_0(x) = \exp(-\nu(K_x))$, $h_n(x) = \exp(-\nu(K_{x_n}))$. For the discontinuity set we have $D \subset -\partial K \cup \{x : \nu(\partial K_x) > 0\}$. The assumptions in Eqs. (9) and (10) imply $\nu(D) = 0$, hence for any r

$$\int_{B_r^c} \exp(-\nu_n(K_x)) d\nu_n(x) \to \int_{B_r^c} \exp(-\nu(K_x)) d\nu(x).$$

Now apply Lemma 3 to derive the estimate

$$\limsup_{n\to\infty} \int_{B_r} \exp(-\nu_n(K_x)) d\nu_n(x) \leq \tau^{-1} \exp(-\nu(B_r^c)\tau).$$

The right–hand side here tends to zero as $r \to 0$.

Putting this all together and comparing with Eq. (8) we conclude

$$\limsup_{n\to\infty} \mathbf{E}V_n^{(1)} \leq \mathbf{E}V^{(1)}.$$

The reverse inequality,

$$\liminf_{n\to\infty} \mathbf{E}V_n^{(1)} \geq \mathbf{E}V^{(1)},$$

follows from the convergence in distribution □

Remark. The continuity conditions of Eqs. (9) and (10) are actually some properties of the spherical measure ρ. The first one trivially translates as $\rho(-\partial K) = 0$, but we have not been able to find a re–formulation for the second. Sufficient conditions for Eq. (10) are: ρ is non–singular, and ∂S_+ lies in a $(n-2)$–dimensional set; or K is convex, ∂K has no two–dimensional facets and $\rho(\partial S_+) = 0$.

Example. Here is a remarkable case where the expectations are explicitly computed. Consider the two–dimensional Cauchy distribution specified by the density $d\mu(x) = (2\pi)^{-1}(1+\|x\|)^{-3/2}dx$, $x \in \mathbf{R}^2$. The radial tail is regularly varying with $\alpha = 1$ and the circular measure is uniform, i.e., $d\nu(rz) = (2\pi)^{-1}r^{-2}dr\,dz$, $r > 0$, $z \in [0,2\pi)$.

Assume first that K is the positive quadrant. The kth layer are those X_i's which are exceeded by exactly $k-1$ points of $\mathcal{H}_n$ in both components. Integrating yields

$$\phi(z) = \nu(K_z) = \frac{\cos z + \sin z - 1}{2\pi \sin z \cos z} \quad z \in (-\pi/2, \pi)$$

and $\phi(z) = \infty$ otherwise. Computing the integral in Eq. (8) we obtain

$$\lim_{n\to\infty} \mathbf{E}V_n^{(k)} = 1 + \frac{3\pi}{4}.$$

For $k = 1$ we have the limiting mean of the number of Pareto points.

Now suppose K is the complement to the negative quadrant. The kth layer consists of those X_i's which exceed all except some $k-1$ sample points in both components. We get

$$\phi(z) = \frac{\cos z + \sin z + 1}{2\pi \sin z \cos z} \cdot \quad z \in (0,\pi/2)$$

and $\phi(z) = \infty$ otherwise. Computing the integral Eq. (8) in this case, we obtain

$$\lim_{n\to\infty} \mathbf{E}V_n^{(k)} = 1 - \frac{\pi}{4}.$$

The first layer is either empty or just one point, maximizing both components, thus this mean value coincides with the limiting probability of the total maximum.

The limiting distribution and higher moments of the $V_n^{(k)}$'s can be, in principle, expressed in terms of some integrals similar to Eq. (8). These expressions do not seem tractable by analytical methods because of the complicated integration domains.

4. Slowly Varying Tails: $\alpha = 0$

The case of slowly varying radial tail, with $\alpha = 0$ in Eq. (1), is of special interest. The above Poisson approximation method does not work, since the sample cannot be rescaled to provide a non-degenerate limit. To get around, we extend here a method already exploited in [1], where the number of convex hull extremes of a sample under slighly stronger assumptions on the distribution has been studied.

We assume for technical reasons that F is continuous though, in fact, slow variation is all that is needed.

Let $X_n^{[1]}, \ldots, X_n^{[n]}$ be the elements of $\mathcal{H}_n$ arranged in the norm-decreasing order, i.e., $\|X_n^{[1]}\| > \ldots > \|X_n^{[n]}\|$. Set $R_n^{[i]} = R_j$ and $Z_n^{[i]} = Z_j$, iff $X_n^{[i]} = X_j$; $i, j = 1, \ldots, n$. One can recognize in the $R_n^{[i]}$'s the radial order statistics. The associated spherical variables, $Z_n^{[i]}$'s will be called *concomitants*. Note that the continuity hypothesis make the definitions correct since the radial components are different with probability one.

Maller and Resnick [13] proved that slow variation is equivalent to

$$\frac{R_n^{[i+1]}}{R_n^{[i]}} \xrightarrow{\mathbf{P}} 0 \quad i = 1,2,\ldots. \tag{13}$$

Our convergence results effectively exploit this fact combined with the asymptotic independence of the concomitants shown next.

Let $Z^{[1]}, Z^{[2]}, \ldots$ be iid S-valued random variables with distribution ρ.

Lemma 5. *Assume that $F(r) = \mu(B_r)$ is continuous and Eq. (2) holds. Then*

$$(Z_n^{[1]}, \ldots, Z_n^{[k]}) \xrightarrow{d} (Z^{[1]}, \ldots, Z^{[k]}) \quad k = 1,2,\ldots$$

Proof. For ρ-continuous $C \subset S$ write (2) as

$$\lim_{\gamma\to\infty} \frac{1-F_C(r)}{1-F(r)} = \rho(C), \tag{14}$$

where $F_C(r) = \mu(\text{cone}(C) \cap B_r)$. Select arbitrary $k \in \mathbf{N}$ and ρ-continuous spherical sets $C_1, \ldots, C_k$.

We have

$$\mathbf{P}\{Z_n^{[1]} \in C_1, \ldots, Z_n^{[k]} \in C_k\} = \frac{n!}{(n-k)!}$$

$$\mathbf{P}\{Z_n^{[1]} \in C_1, X_n^{[1]} = X_1, \ldots, Z_n^{[k]} \in C_k, X_n^{[k]} =$$

$$X_k\} = \frac{n!}{(n-k)!}\mathbf{P}\{Z_n^{[1]} \in C_1, R_n^{[1]} = R_1, \ldots,$$

$$Z_n^{[k]} \in C_k, R_n^{[k]} = R_k\} = \frac{n!}{(n-k)!}\mathbf{P}\{Z_n^{[1]} \in C_1, \ldots,$$

$$Z_n^{[k]} \in C_k, R_1 > \ldots > R_k; R_k > R_i \text{ for } i = k+1, \ldots, n\} =$$

$$\frac{n!}{(n-k)!} \int_{\substack{r_1 > \ldots > r_k \\ z_1 \in C_1, \ldots, z_k \in C_k}} (F(r_k))^{n-k} \mathrm{d}\mu(r_1 z_1) \ldots \mathrm{d}\mu(r_k z_k) =$$

$$\frac{n!}{(n-k)!} \int_{r_1 > \ldots > r_k} (F(r_k))^{n-k} \mathrm{d}F_{C_1}(r_1) \ldots \mathrm{d}F_{C_k}(r_k) =$$

$$n\binom{n-1}{k-1} \int_{r_1, \ldots r_{k-1} > r_k} (F(r_k))^{n-k} \mathrm{d}F_{C_1}(r_1) \ldots \mathrm{d}F_{C_k}(r_k) =$$

$$n\binom{n-1}{k-1} \int_0^\infty (F(r_k))^{n-k} \mathrm{d}F_{C_k}(r_k) \int_{r_k}^\infty \ldots \int_{r_k}^\infty \mathrm{d}F_{C_1}(r_1)$$

$$\ldots \mathrm{d}F_{C_{k-1}}(r_{k-1}) = n\binom{n-1}{k-1} \int_0^\infty (F(r_k))^{n-k}$$

$$(1-F_{C_1}(r_k)) \ldots (1-F_{C_{k-1}}(r_k))\, \mathrm{d}F_{C_k}(r_k) =$$

$$n\binom{n-1}{k-1} \left(\int_0^r (\ldots) + \int^\infty (\ldots) \right) \sim n\binom{n-1}{k-1} \int_0^\infty (\ldots) =$$

(for large r uniformly in n)

$$n\binom{n-1}{k-1} \rho(C_1) \ldots \rho(C_{k-1}) \int_r^\infty (1-F(r_k))^{k-1} (F(r_k))^{n-k}$$
$$\mathrm{d}F_{C_k}(r_k) + \epsilon \to \rho(C_1) \ldots \rho(C_k) + \epsilon,$$

as $n \to \infty$, where we have used Eq. (14) and applied an argument similar to that in Lemma 3. Asymptotically, the probability is factorized, whence the statement □

To prove the convergence we combine in what follows the above lemma and Eq. (13). The idea is that the points with top layer ranks have also small ranks in the radial components. On the other hand, conical extremality of the points with small layer ranks is determined by their, almost independent, spherical components.

Introduce the random variables

$$T_n^{(0)} = 0;\ T_n^{(k)} = \min\{i : \cup_{m=1}^k \mathscr{L}^{(m)}(\mathscr{H}_n) \cap \{X_n^{[i+1]}, \ldots, X_n^{[n]}\} = \emptyset\},\ k = 1, \ldots, n,$$

which count the X_i's in the norm-decreasing order until the first k layers having been filled. Clearly, $V_n^{(k)} \leq T_n^{(k)} \leq T_n^{(k+1)}$. Denote S^∞ the product of infinitely many spheres, and set

$$\hat{T}^{(0)}(\mathbf{z}) = 0;\ \hat{T}^{(k)}(\mathbf{z}) = \min\{j : (S_+ \cap \{z_1, \ldots, z_j\}) = k\},$$
$$\hat{V}^{(k)}(\mathbf{z}) = \#\{j : \hat{T}^{(k-1)}(\mathbf{z}) < j \leq \hat{T}^{(k)}(\mathbf{z}),\ z_j \in S \backslash S_-\},$$

where $\mathbf{z} = (z_1, z_2, \ldots) \in S^\infty$ and $\inf \emptyset = \infty$. For $i \geq j$ the set

$$\{\mathbf{z} \in S^\infty : \hat{V}^{(k)}(\mathbf{z}) = j, \hat{T}^{(k)}(\mathbf{z}) = i\}$$

is a finite-dimensional cylinder in S^∞.

Denote $\mathbf{Z} = (Z^{[1]}, Z^{[2]}, \ldots)$ the sequence with *iid* components distributed according to ρ, $T^{(k)} = \hat{T}^{(k)}(\mathbf{Z})$, $V^{(k)} = \hat{V}^{(k)}(\mathbf{Z})$. It follows from the definition and condition (iii) of Sec. 2 that $T^{(k)}$, $k = 1,2, \ldots$ is a strictly increasing sequence of finite stopping times with respect to $Z^{[1]}, Z^{[2]}, \ldots$

Theorem 2. *Assume* (i)-(iv), $\alpha = 0$, *and* $\rho(\partial S_+ \cup \partial S_-) = 0$. *Then for any* $k = 1,2, \ldots$

$$(V_n^{(1)}, T_n^{(1)}, \ldots, V_n^{(k)}, T_n^{(k)}) \xrightarrow{d} (V^{(1)}, T^{(1)}, \ldots, V^{(k)}, T^{(k)}).$$

Proof. Fix integers $v_1, \ldots, v_k; t_1, \ldots, t_k = t$ satisfying $0 < t_1 < \ldots < t_k$ and $0 \leq v_i \leq t_i - t_{i-1}$ for $i = 1, \ldots, k$. We need to prove that

$$\lim_{n\to\infty} \mathbf{P}\{V_n^{(i)} = v_i, T_n^{(i)} = t_i;\ \ i \leq k\} =$$
$$\mathbf{P}\{V^{(i)} = v_i, T^{(i)} = t_i; i \leq k\}.$$

We endow S^t with the product measure ρ^t and the Euclidean metric. Define

$$D = \partial S_+ \cup \partial S_-, \mathscr{D} = \{(z_1, \ldots, z_t) \in S^t : \{z_1, \ldots, z_t\} \cap D = \emptyset\},$$

$$\mathscr{A} = \{(z_1, \ldots, z_t) \in S^t : \hat{V}^{(i)}(z_1, \ldots, z_t) = v_i, \hat{T}^{(i)}(z_1, \ldots, z_t) = t_i;\ i \leq k\}.$$

The definition of $\mathscr{A}$ is correct due to the cylindrical property. It is easy to see that $\mathscr{D}$ is compact, $\mathscr{A} \backslash \mathscr{D}$ is open and, by the assumption, $\rho^t(\mathscr{D}) = 0$. It follows,

$\rho'(\mathcal{A}) = \rho'(\mathcal{A}\backslash\mathcal{D})$. For any δ there exists $\mathcal{A}_\delta$ with the properties:

$$\mathcal{A}_\delta \subset \mathcal{A}\backslash\mathcal{D},\ \rho(\mathcal{A})-\rho(\mathcal{A}_\delta)<\delta,\ \mathrm{dist}(\mathcal{A}_\delta,\mathcal{D})>0,$$
$$\rho^t(\partial\mathcal{A}_\delta)=0. \tag{16}$$

To prove this, take $\mathcal{C}_\theta = \{a \in \mathcal{A}\backslash\mathcal{D} : \mathrm{dist}\,(a,\partial(\mathcal{A}\backslash\mathcal{D})) > \theta\}$, then $\mathcal{C}_\theta$ is an open set, increasing to $\mathcal{A}\backslash\mathcal{D}$ as $\theta \downarrow 0$. We have $\rho'((\mathcal{A}\backslash\mathcal{D})\backslash\mathcal{C}_\theta) < \delta$ for sufficiently small θ. On the other hand, $\partial\mathcal{C}_\theta \subset \{a \in S' : \mathrm{dist}\,(a,\partial(\mathcal{A}\backslash\mathcal{D})) = \theta\}$, these sets being disjoint for different θ. Hence the set of the values of θ with $\rho'(\partial\mathcal{C}_\theta) > 0$ is at most countable. Select an appropriate θ and set $\mathcal{A}_\delta = \mathcal{C}_\theta$.

We derive from Eq. (16) with the help of some topological considerations that for sufficiently small ϵ

$$\bigcup_{i \le t} (z_i + B_\epsilon) \cap \mathrm{cone}\,(D) = \emptyset \quad (z_1, \ldots, z_t) \in \mathcal{A}_\delta. \tag{17}$$

Assume now that the compound event

$$(Z_n^{[1]}, \ldots, Z_n^{[t]}) \in \mathcal{A}_\delta;\ \epsilon R_n^{[i]} > R_n^{[i+1]} \quad i = 1, \ldots, k \tag{18}$$

occurs. We show next that in this case

$$T_n^{(i)} = t_i,\ V_n^{(i)} = v_i \quad i = 1, \ldots, k. \tag{19}$$

Let Q be an element of the finite algebra of spherical sets generated by S_+ and S_-. The following equivalence holds:

$$X_n^{[i]} - X_n^{[j]} \in \mathrm{cone}(Q) \Leftrightarrow Z_n^{[i]} \in Q \text{ for } 1 \le i \le t, i < j \le n. \tag{20}$$

Indeed, note first that $\partial Q \subset$ D. By (17), $Z_n^{[i]} \in Q$ implies $Z_n^{[i]} + B_\epsilon \subset \mathrm{cone}\,(Q)$. From Eq. (18) we have also $X_n^{[i]} + B_{R\backslash\epsilon} \subset \mathrm{cone}\,(Q)$. But $-X_n^{[i]} \in B_{R\backslash\epsilon}$ thus $X_n^{[i]} - X_n^{[j]} \in \mathrm{cone}\,(Q)$. Use Q^c instead of Q to prove the reverse implication.

The definition of $\mathcal{A}$ and Eq. (18) yield $Z_n^{[t_i]} \in S_+$, $i = 1, \ldots, k$. Setting $Q = S_+$ in Eq. (20) we have $X_n^{[t_i]} - X_n^{[j]} \in K$, $t_i < j \le n$. Therefore,

$$\{X_n^{[t_i+1]}, \ldots, X_n^{[n]}\} \cap \Big(\bigcup_{m=1}^{k} \mathcal{L}^{(m)}(\mathcal{H}_n)\Big) = \emptyset.$$

Let $t_{i-1} < j < t_i$ and $Z_n^{[j]} \in S_-$. Setting $Q = S_-$ in Eq. (20), we have $\{X_n^{[j+1]}, \ldots, X_n^{[n]}\} \subset (X_n^{[j]} + K)$. Setting $Q = S_+$ we have further $\{X_n^{[t_1]}, X_n^{[t_2]}, \ldots X_n^{[t_{i-1}]}\} \in (X_n^{[j]} + K)$. That is, $X_n^{[j]} \notin \cup_{m \le k} \mathcal{L}^{(m)}(\mathcal{H}_n)$.

Let $t_{i-1} < j < t_i$ and $Z_n^{[j]} \in S_0$. Substituting $Q = S_0$ into Eq. (20) we get $X_n^{[j]} - X_n^{[p]} \in \mathrm{cone}\,(S_0)$, $j+1 \le p \le n$, together with $S_0 = -S_0$ and $S_0 \cap S_+ = \emptyset$ this yields $\{X_n^{[j+1]}, \ldots, X_n^{[n]}\} \cap (K + X_n^{[j]}) = \emptyset$. For $Q = S\backslash S_+$ and $p \in \{1, \ldots, j-1\}\backslash\{t_1, t_2, \ldots, t_{i-1}\}$ we have $X_n^{[p]} \notin (K + X_n^{[n]})$. Similarly, for $p \in \{t_1, t_2, \ldots, t_{i-1}\}$ we have

$$\{X_n^{[t_1]}, X_n^{[t_2]}, \ldots, X_n^{[t_{i-1}]}\} \subset X_n^{[j]} + K.$$

Thus in this case $X_n^{[j]} \in \mathcal{L}^{(i-1)}(\mathcal{H}_n)$.

In the same manner, $Z_n^{[t_i]} \in S_+\backslash S_-$ implies $X_n^{[t_i]} \in \mathcal{L}^{(i-1)}(\mathcal{H}_n)$.

Summarizing, if Eq. (18) holds then $\mathcal{L}^{(i)}(\mathcal{H}_n) = \{X_n^{[t_i]} : t_{i-1} < j \le t_i,\ Z_n^{[j]} \in S_+\backslash S_-\}$, whence Eq. (19).

Now from Eq. (13) and Lemma 5 (recall that $t = t_k$)

$$\mathbf{P}\{(Z_n^{[1]}, \ldots, Z_n^{[t_k]}) \in \mathcal{A}_\delta,\ \epsilon R_n^{[i]} > R_n^{[i+1]};\ i = 1, \ldots, t_k\} \to \rho'^k(\mathcal{A}_\delta).$$

Recalling the definitions of $\mathcal{A}$ and $\mathcal{A}_\delta$ we get

$$\liminf_{n\to\infty} \mathbf{P}\{V_n^{(i)} = v_i, T_n^{(i)} = t_i;\quad i \le k\} >$$
$$\mathbf{P}\{V^{(i)} = v_i, T^{(i)} = t_i;\quad i \le k\} - \delta.$$

Take $\delta = \delta(v_1, \ldots, v_k; t_1, \ldots, t_k)$ with $\Sigma\delta(v_1, \ldots, v_k; t_1, \ldots, t_k) = \beta$ and choose a diagonal subsequence of the values of n to get the convergence of the probabilities in the left-hand side. Recalling that probabilities sum to one, we derive Eq. (15) by setting $\beta \to 0$ □

Convergence in mean does not require additional restrictions, as shown next.

Theorem 3. *Under the assumptions of Theorem 2*

$$\mathbf{E}V_n^{(k)} \to \mathbf{E}V^{(k)},\ n \to \infty.$$

Proof. It is sufficient to consider only the case $k = 1$. Denote by $I_n^{[i]}$ and $I^{[i]}$ the indicator functionsof the events $\{X_n^{[i]} \in \mathcal{L}^{(1)}(\mathcal{H}_n)\}$ and $\{Z^{[i]} \in S\backslash S_-,\ i \le T^{[1]}\}$, respectively. Clearly,

$$V_n{}^{(1)} = I_n^{[1]} + \ldots I_n^{[n]},\ V^{(1)} = I^{[1]} + I^{[2]} + \ldots$$

By an argument similar to that used in Theorem 2 we show that

$$(I_n^{[1]}, \ldots, I_n^{[m]}) \xrightarrow{d} (I^{[1]}, \ldots, I^{[m]}) \quad m = 1,2, \ldots \tag{21}$$

Choose γ_n satisfying $np_n \to \lambda$, where $\lambda > 0$, $p_n = 1-\mu(B_{\gamma_n})$. The random variable $N = \#(\{X_1, \ldots, X_n\} \cap B_{\gamma_n}^c)$ has binomial distribution with parameters (n, p_n). By Eq. (6),

$$\mathbf{E}(I_n{}^{[N+1]} + \ldots + I_n^{[n]}) = \mathbf{E}(\mathcal{L}^{(1)}(\mathcal{H}_n) \cap B_{\gamma_n}) \le e^{-\lambda\tau}\tau^{-1}. \tag{22}$$

Fix m and write the expectation as

$$\mathbf{E}V_n^{(1)} = \mathbf{E}(I_n^{[1]} + \dots I_n^{[n]}) = \mathbf{E}(I_n^{[1]} + \dots I_n^{[m]}) +$$

$$\mathbf{E}(I_n^{[m+1]} + \dots I_n^{[N]}) 1_{\{N \geq m\}} + \mathbf{E}(I_n^{[N+1]} + \dots I_n^{[n]}) 1_{\{N \geq m\}} +$$

$$\mathbf{E}(I_n^{[m+1]} + \dots I_n^{[n]}) 1_{\{N < m\}}.$$

The first term converges by Eq. (21):

$\mathbf{E}(I_n^{[1]} + \dots I_n^{[m]}) \to \mathbf{E}(I^{[1]} + \dots I^{[m]})$. The first and the third terms are estimated by Eq. (22) as

$$\mathbf{E}(I_n^{[N+1]} + \dots I_n^{[n]}) 1_{\{N \geq m\}} + \mathbf{E}(I_n^{[m+1]} + \dots I_n^{[n]}) 1_{\{N < m\}}$$

$$\leq \mathbf{E}(I_n^{[N+1]} + \dots + I_n^{[n]}) \leq e^{-\lambda\tau}\tau^{-1}.$$

Since N is binomially distributed, we have for the second term

$$\mathbf{E}(I_n^{[m+1]} + \dots I_n^{[N]}) 1_{\{N \geq m\}} < \mathbf{E}(N 1_{\{N \geq m\}}) \to$$

$$\lambda \mathbf{P}\{N_\lambda \geq m-1\}, \quad n \to \infty$$

where N_λ is a Poisson random variable with parameter λ. Selecting λ and then m sufficiently large, we prove $\limsup_{n\to\infty} \mathbf{E}V_n^{(k)} \leq \mathbf{E}V^{(k)}$.

The inverse inequality involving lim inf follows from the convergence in distribution □

It is not hard to find the limiting distributions of the $V_n^{(k)}$'s. Note first that

$$(V^{(1)}, T^{(1)}), (V^{(2)}, T^{(2)} - T^{(1)}), \dots$$

are all *iid*, therefore it is sufficient to consider only the first pair. Clearly $T^{(1)}$ is geometrically distributed with parameter $\rho(S_+)$. The probability law of $V^{(1)}$ is found from the following scheme: throw down the *iid* points $Z^{[1]}, Z^{[2]}, \dots$ in S according to the probability law ρ until the first point falls into S_+, then count all the points falling into $S \backslash S_-$. To make this precise denote

$$p_0 = \rho(S_0),\ p_\pm = \rho(S_\pm),\ p_- = \rho(S_-),\ p_+ = \rho(S_+)$$

(thus $p_0 + p_- + p_+ - p_\pm = 1$). The joint distribution of $V^{(1)}$ and $T^{(1)}$ is this:

$$\mathbf{P}\{V^{(1)} = i, T^{(1)} = j\} = \mathbf{P}\{\{Z^{[1]}, \dots, Z^{[j-1]}\} \cap S_+ = \emptyset,$$

$$Z^{[j]} \in S_+, \#(\{Z^{[1]}, \dots, Z^{[j]}\} \cap (S \backslash S_-)) = i\} =$$

$$\mathbf{P}\{\{Z^{[1]}, \dots, Z^{[j-1]}\} \cap S_+ = \emptyset,$$

$$Z^{[j]} \in S_+ \backslash S_-,\ \#(\{Z^{[1]}, \dots, Z^{[j-1]}\} \cap (S \backslash S_-)) = i-1\} +$$

$$\mathbf{P}\{\{Z^{[1]}, \dots, Z^{[j-1]}\} \cap S_+ = \emptyset,\ Z^{[j]} \in S_\pm,$$

$$\#(\{Z^{[1]}, \dots, Z^{[j-1]}\} \cap (S \backslash S_-)) = i\} =$$

$$\mathbf{P}\{\#(\{Z^{[1]}, \dots, Z^{[j-1]}\} \cap S_0) = i-1,$$

$$\#(\{Z^{[1]}, \dots, Z^{[j-1]}\} \cap (S_- \backslash S_+)) = j-i\} \times$$

$$\mathbf{P}\{Z^{[j]} \in S_+ \backslash S_-\} + \mathbf{P}\{\#(\{Z^{[1]}, \dots, Z^{[j-1]}\} \cap S_0) = i,$$

$$\#(\{Z^{[1]}, \dots, Z^{[j-1]}\} \cap (S_- \backslash S_+)) = j-i-1\} \mathbf{P}\{Z^{[j]} \in S_\pm\} =$$

$$\binom{j-1}{i-1} p_0^{i-1}(p_- - p_\pm)^{j-i}(p_+ - p_\pm) + \binom{j-1}{i} p_0^i (p_- - p_\pm)^{j-i-1} p_\pm,$$

where $i \geq 0, j \geq 1$ and $j \geq i$. Summing over j we arrive at the limiting distribution of points in the kth layer:

$$\mathbf{P}\{V^{(k)} = 0\} = \frac{p_\pm}{1 - p_- + p_\pm},$$

$$\mathbf{P}\{V^{(k)} = i\} = \frac{p_0^{i-1}(1-p_-)p_+}{(1-p_- + p_\pm)^{i+1}} \quad i = 1,2,\dots$$

If the cone satisfies $K \cap -K = \emptyset$ (or, more generally $\rho(S_+ \cap S_-) = 0$) then $p_\pm = 0$ and $V^{(k)}$ is geometrically distributed. A little additional work is needed to find the expectation:

$$\mathbf{E}V^{(k)} = \frac{1-p_-}{p_+}. \tag{23}$$

Example. Assume that the radial tail is slowly varying and ρ is the uniform spherical measure.

For $K = \mathbf{R}_+^d$ we have $p_+ = p_- = 2^{-d}$, and Eq. (23) yield $\mathbf{E}V_n^{(k)} \to 2^d - 1$. In particular, the mean number of Pareto points in two dimensions converges to 3.

Taking the complement to the negative orthant, we have $p_+ = p_- = 1 - 2^{-d}$ and $\mathbf{E}V_n^{(k)} \to (2^d - 1)^{-1}$. In two dimensions, the probability that the sample has the double maximum tends to 1/3.

Appearing of inverse numbers in the above example is a general phenomenon.We write further $V_n^{(k)}(K)$ to emphasize the dependence on the cone.

Theorem 4. Under the assumptions of Theorem 2

$$\lim_{n\to\infty} \mathbf{E}V_n^{(k)}(K)\, \mathbf{E}V_n^{(k)}(-K^c) = 1,$$

provided one of the numbers p_+ or $1-p_-$ is positive.

Proof. This follows from Theorem 3, Eq. (23), and the formulas $p_-(-K^c) = 1 - p_+(K)$, $p_+(-K^c) = 1 - p_-(K)$ □

Remark. Given a binary relation, say $\mathcal{R}$, on a sampling space, and a random sample $\mathcal{H}_n$, there are two natural ways to define the "kth extremes" *of* $\mathcal{H}_n$: (1) sample elements X_i which are in $\mathcal{R}$ with all other sample elements with the exception of some $k-1$ points; or (2) the elements X_i such that there are exactly $k-1$ sample points which are in the relation with X_i. In the theory of partially ordered sets extremes ($k = 1$) of the first type are called the greatest points, of the second type—maximal [4]. This is best illustrated by the natural partial order of $\mathbf{R}^d$: total maximum is the greatest point, while Pareto set consists of maximal points. If the binary relation $\mathcal{R}$ is generated by a cone K, as mentioned in Introduction, then the K-extremes are maximal points, while the $-K^c$-extremes are the greatest points w.r.t. $\mathcal{R}$. Baryshnikov [3] has proved that the asymptotic upper bound for the product of expectations of the numbers of the extremes of both types is at most 1, for any fixed $\mathcal{R}$ and k. Theorem 4 shows that this bound is sharp.

Remark. Normal multivariate distributions can be viewed as the case of fast decreasing radial tails, $\alpha = \infty$. The mean number of conical extremes demonstrates typically the following behavior: for any k, $\mathbf{E}V_n^k$ infinitely grows if K is contained in a half-space, and tends to zero if K contains a half-space [10,11].

Acknowledgment

This work was supported by Deutsche Forschungsgemeinschaft.

5. References

[1] D. J. Aldous, B. Fristedt, P. S. Griffin, and W. E. Pruitt, The number of extreme points in the convex hull of a random sample, J. Appl.Prob. **28**, 287-304 (1991).
[2] O. Barndorff-Nielsen and M. Sobel, On the distribution of the number of admissible points in a vector random sample, Theory Prob. Appl. **11**, 283-305 (1966).
[3] Yu. M. Baryshnikov, Counting optimal elements in a sample, University of Osnabrück, Preprint (1991).
[4] G. Birkhoff, Lattice Theory, Amer. Math. Soc. Colloq. Publ. 25, Amer. Math. Soc., Providence, RI (1967).
[5] H. Brozius, Convergence in mean of some characteristics of the convex hull, Adv. Appl. Prob. **21**, 526-542 (1989).
[6] H. Brozius and L. De Haan, On limiting laws for the convex hull of a sample, J. Appl. Prob. **24**, 852-862 (1989).
[7] R. A. Davis, E. Mulrow and S. I. Resnick, The convex hull of a random sample in $\mathbf{R}^2$. Commun. Statist. Stoch. Models, **3**, 1-27 (1987).
[8] L. De Haan and S. I. Resnick, Limit theory for multivariate sample extremes, Z. Wahrsch. Gebiete **40**, 317-333 (1977).
[9] W. F. Eddy and J. D. Gale, The convex hull of a spherically symmetric sample, Adv. Appl. Prob. **13**, 751-763, 1981.
[10] A. V. Gnedin, Multiple maxima of a normal sample. Preprint, University of Göttingen (1993).
[11] V. M. Ivanin, Asymptotic estimate of the number of elements in the Pareto set, Kibernetika **11**, 108-113 (1975).
[12] A. I. Kuksa and N. Z. Shor, The method of evaluating the number of conditionally optimal trajectories of discrete separable dynamic programming, Kibernetika **8**, 910-912 (1972).
[13] R. A. Maller and S. I. Resnick, Limiting behaviour of sums and the term of maximum modulus, Proc. London Math. Soc. **49**, 385-422 (1984).
[14] A. G. Postnikov, Introduction to the Analytical NumberTheory, Nauka, Moscow (1971).
[15] S. I. Resnick, Point processes, regular variation and weak convergence. Adv. Appl. Prob. **18**, 66-138 (1986).

About the author: *Alexander V. Gnedin is a mathematician at the Institut für Mathematische Stochastik, Universität Göttingen, Lotzestrasse 13, 37083 Göttingen, Germany.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 521 (1994)]

Bayesian Forecasting of Extreme Values in an Exchangeable Sequence

Volume 99 | Number 4 | July–August 1994

Bruce M. Hill

Department of Statistics,
University of Michigan,
Ann Arbor, MI 48109

This article develops new theory and methodology for the forecasting of extreme and/or record values in an exchangeable sequence of random variables. The Hill tail index estimator for long-tailed distributions is modified so as to be appropriate for prediction of future variables. Some basic issues regarding the use of finite, versus infinite idealized models, are discussed. It is shown that the standard idealized long-tailed model with tail index $\alpha \leq 2$ can lead to unrealistic predictions if the observable data is assumed to be unbounded. However, if the model is instead viewed as valid only for some appropriate finite domain, then it is compatible with, and leads to sharper versions of, sensible methods for prediction. In particular, the prediction of the next record value is then at most a few multiples of the current record. It is argued that there is no more reason to eschew posterior expectations for forecasting in the context of long-tailed distributions than to do so in any other context, such as in the many applications where expectations are routinely used for scientific inference and decision-making. Computer simulations are used to demonstrate the effectiveness of the methodology, and its use in forecasting is illustrated.

Key words: Bayesian forecasting; exchangeability; long-tailed distributions; record values; tail-index estimator.

Accepted: March 22, 1994

1. Introduction

Consider a sequence $X_1,...,X_n$ of positive random variables that is exchangeable. We say that X_{n+1} is a (new) record value if $X_{n+1} > X_i$, for $i = 1,...,n$. See [2] for some related discussion of record values in the iid case. The problem that we address concerns forecasting of the next observation, X_{n+1}, given that it is a record value, conditional upon the data $X_i = x_i$, for $i = 1,...,n$. In other words, given that X_{n+1} sets a new record, how large will it be?

In the Bayesian approach, with squared error loss, the forecast of X_{n+1}, conditional upon the data $X_1,...,X_n$, and upon $X_{n+1} > max[X_1,...,X_n]$, is simply the posterior expectation of X_{n+1} conditional upon the same information. Note that if a sequence is exchangeable, then the future variables are also conditionally exchangeable, given the realization of the first n variables. Hence each of the next N observations has in fact the same posterior predictive distribution. The posterior expectation for X_{n+j}, conditional upon X_{n+j} being larger than each of the first n observations, is then the same for each $j \geq 1$. It may be noted that there are two quite different questions that arise concerning the forecasting of future record values. The first concerns the forecasting of when the next record value will occur, while the second concerns the forecasting of the

magnitude of the next record value. In this article we only consider the second question.[1]

Although we focus attention here only on the prediction of the magnitude of X_{n+1} given that it sets a new record, there is a relatively straight-forward extension of these results to the evaluation of the posterior expectation of X_{n+j}, given that it sets a new record. To obtain the prediction of the next record value, conditional upon the data $x_1,...,x_n$, and upon X_{n+j} being the next new record value, we must evaluate the posterior expectation of X_{n+j}, conditional upon the collection of inequalities that define the event that X_{n+j} is the next record value. This can be done by a generalization of the procedure for forecasting X_{n+1}, conditional upon its being the next record value. For example, the posterior expectation of X_{n+2}, conditional upon its setting a new record, can be obtained by conditioning upon the event that X_{n+1} sets a new record, and then making the same type of evaluation as above for X_{n+1}, given that it is a record value; or alternatively, by conditioning upon the event that $X_{n+1}<max[X_1,...,X_n]$, and then evaluating the posterior expectation of X_{n+2}, given that it is larger than $max[X_1,...,X_n]$. Since in the Bayesian framework with a specified *a priori* distribution, the posterior probability that X_{n+1} sets a new record is known, there is no difficulty in principle in extending the analysis for the posterior expectation of X_{n+1}, given that it sets a new record, to the forecasting of the magnitude of future record values. Explicit algorithms for doing so will appear in a later paper.

Although the present paper deals only with the evaluation of the posterior expectation of X_{n+1}, given that it sets a record, we shall nonetheless sometimes speak of forecasting the magnitude of future record values, since this can be achieved by the same basic methods. Similarly, one can obtain the posterior expectation of the maximum over some finite horizon, say the maximum of $X_{n+1},...,X_{n+N}$, given that this maximum exceeds our current record value. This is a problem of considerable practical importance both in economic forecasting of interest rates, and in engineering design, where for example, one desires to build a structure capable or withstanding severe winds or earthquake tremors over a certain period of time. To the best of my knowledge such forecasting has never been attempted before in the sense of providing a procedure that could be recommended for serious consideration in real-world problems.

If we assume a conventional statistical model with some unknown parameter θ, then in principle these are straight-forward Bayesian problems, since one can integrate out unknown parameters with respect to their posterior distribution, to obtain the predictive distribution for a new observation; and then condition also upon such a new observation being a record value, in order to answer the question. For example, one could obtain the posterior expectation and variance for X_{n+1}, given that it is a record value. However in typical real-world problems involving forecasting of such extreme values, the model is always uncertain and often unreliable. This is especially so in the tails of the distribution, where there is little, if any, past data to rely upon. Thus to obtain reliable forecasts requires serious attention to model uncertainly. See Hill [7] for discussion of the selection of models from a Bayesian viewpoint, Poirier [8] for a Bayesian analysis of some theoretical models in economics, and Singpurwalla and Meinhold [9] for Bayesian robustification theory in a closely related area.

In this paper we attempt to deal with the problem by using the formulation for inference about the tails of the distribution initiated in [1]. See [10] for an exposition, and Csörgö et al. [11] for related asymptotic theory. This approach utilizes only the upper order statistics of the past data for inference about the upper tail, since it is only such order statistics that fall in the upper tail where the form of the distribution is assumed known. Seriously to utilize the information in the other order statistics requires knowledge concerning the global form of the distribution, and such knowledge is often unavailable. Suppose that given the parameter α, the upper tail of a distribution F on the positive real line is of algebraic form, with tail index α. We assume that

$$1-F(t)=P(X>t|\alpha)\approx C\times t^{-\alpha},$$

for $C>0,\alpha>0$, and t in some interval (A,k) that is considered relevant for prediction of future observations. It is supposed that a random sample $X_i=x_i$, for $i=1,...,n$, from the distribution is available, and based upon this data we wish to forecast the next observation X_{n+1}. Such prediction in the Bayesian context amounts to putting forth a posterior distribution for X_{n+1}, that is obtained by

[1] For those unfamiliar with exchangeability, it may be remarked that exchangeable sequences are strictly stationary processes, and can be strongly dependent. An interesting and important class of exchangeable processes consists of the Markov-Pólya processes, discussed in [3,4,5,6], which play a major role in the theory of stochastic chaos.

integrating out unknown parameters such as α, with respect to their posterior distribution, and then making appropriate forecasts by minimizing posterior expected loss with respect to some loss function. In this article we consider only squared error loss, but our methods can be used in connection with any loss function believed appropriate. See Aitchision and Dunsmore [12] and Maret [13] for the Bayesian theory and methodology of such predictive distributions.

Often a simple summary of the posterior predictive distribution, such as the posterior expectation and variance of X_{n+1}, suffices for many practical purposes. In typical applications A will be the largest order statistic of the past data. k can sometimes be $+\infty$, but for reasons discussed below will often instead be some modest multiple of A. We might be interested, for example, in forecasting the next observation, X_{n+1}, conditional upon its being between $x^{(1)}$ and $5\times x^{(1)}$, where $x_{(1)}$ is the largest order statistic of the past data. Forecasting of such a record value is an especially difficult part of the overall forecasting problem, since by assumption there is no past data of this magnitude. Yet in forecasting extreme values, it is necessary to consider precisely the situation in which the observation is more extreme than anything yet seen. For example, in designing a structure to resist high winds, one must make allowance for forces more extreme than have yet been experienced. It would be foolish to imagine that such forces have already been observed at their maximum.

The best that one can do in such circumstances is to use what relevant theory exists, making sure that such theory is compatible with the data that has been seen. In this article we shall rely on the theory of long-tailed distributions, in which the tail is known to be of algebraic form at least in some interval. Many data sets are known to be of this form. Examples include income distributions, city size distributions, distributions of genera by species, insurance claim sizes, word frequency distributions, stock market fluctuations, and many others. See Zipf [14] for graphical presentation of a great variety of data in support of his theory for long-tailed distributions. Several theoretical models have been proposed for such data. These include the probability models of Yule [15], Hill [16,17,18], Hill and Woodroofe [19,20], and Hill, Lane and Sudderth [3,4]. See Johnson and Kotz [21] for discussion of the model of Hill [22,17], which was the starting point for the later models. As pointed out by Chatterjee and Yilmaz [23], some of these models are related to stochastic models for chaos.

We are particularly interested in the case where α is not large, so we are dealing with a truly long-tailed distribution. For any $\alpha>0$ the distribution of X_{n+1} is proper, even when $k=\infty$. However, for fixed known $\alpha\leq 1$ the expectation of X_{n+1} is infinite if there is no finite upper bound for the data, and the variance of X_{n+1} is infinite if $\alpha\leq 2$. Also, if $\alpha\geq 1$ is unknown, which is ordinarily the case, the posterior distribution for α must give sufficiently small weight to values of α near 1, in order for the posterior expectation of X_{n+1} to be finite. This gives rise to an important practical issue for Bayesians, since the predictions are then very sensitive to the precise form of the *a priori* distribution for α near 1, and the results are not robust. Similarly, if $\alpha\geq 2$ is unknown, the posterior distribution for α must give sufficiently small weight to values of α near 2, in order for the posterior variance of X_{n+1} to be finite.[2]

In view of such nonrobustness, it is necessary to proceed more carefully than in most problems of statistical inference and prediction. Our method is to take explicit account of the boundedness of the observations. In many real world applications of extreme value theory, where one deals with maximal temperatures, wind velocities, rain fall, etc., the data are generally considered to be bounded. For example, a wind velocity even double the highest ever previously experienced, must be regarded as extremely improbable. Even if such could occur, it might be regarded as indicating a basic change in climate such as would invalidate all standard assumptions, and so require modification of existing theory. This suggests that a realistic analysis of the problem should incorporate a finite upper bound, say K, for the data.[3] Such a bound might be taken a

[2] Some may think that because of such issues one should be considering inference about percentiles, such as the median, rather than the expectation. However, means are often of particular interest and importance in real-world problems, and of course are appropriate for squared error loss. If there were no technical difficulties at infinity with the expectation, would anyone argue against its use for prediction?

[3] Instead of requiring that the mass be exactly 0 beyond a certain known bound K, one can alternatively require that the mass beyond this bound be so negligible as to be of no interest. In the subjective Bayesian approach it would be remarkable for anyone to have a probability of 0, to infinitely many decimal points, for a logically possible event. However, whether or not 0 is taken literally, in effect one ordinarily ignores values of the observation larger than the bound. For the purposes of this article we treat such negligible mass as though it were 0. An alternative and nearly equivalent way to deal with the problem is to consider only conditional inference, given that the observations are no larger than the bound. A general theory and methodology for such conditional inference is proposed in [24].

good deal larger than is ordinarily believed reasonable. A 10-fold increase above a previous record value that was based upon substantial data would often be too large, but is worthy of consideration. If such an upper bound is incorporated in the analysis, then as shown below, even if $\alpha \leq 1$ there is no problem with infinite moments. We will typically assume some known finite upper bound K, perhaps much too large, but we will not necessarily assume that $\alpha \geq 1$, and will let the data speak for themselves in this regard. Since the density in the tail is proportional to $t^{-\alpha-1}$, we see that $\alpha = 0$ corresponds (in the tail) to a uniform distribution for the logarithm of the observation. Such a distribution is often used by Bayesians to represent diffuse *a priori* knowledge about a positive quantity such as a variance.

Our precise model is as follows. We assume that there exists a known constant K such that $0 \leq X \leq K$, so that K is a known upper bound for the data. In applications, ordinarily $K < \infty$, but for completeness we shall also discuss the case $K = \infty$, which is sometimes appropriate and is mathematically convenient when $\alpha > 2 + \epsilon > 2$, in which case no problems arise due to infinite first or second moments. We do not assume in applications that one can necessarily determine a smallest such K, but merely that one can pick some bound. We also assume that there exist constants k and A with $K > k > A > 0$, such that the tail is algebraic, to an adequate approximation, for $A \leq t \leq k$, with 0 mass beyond K. Let $x_{(1)} > \cdots > x_{(n)}$ be the descending order statistics of the past data. Ordinarily we take A to be the largest order statistic of the past data, $A = x_{(1)}$. The quantity k is the key variable in our analysis. It represents the point up to which the algebraic assumption is assumed to be valid. k is not a parameter in the usual sense, but is more in the nature of a decision variable, since in applications the tail will not be *exactly* algebraic in any interval, but it will nevertheless be reasonable to act as if it were approximately of this form for some intervals. The selection of k in part acts as a means to specify the portion of the distribution that we are particularly interested in. Even if $X > k$ we may not be interested in forecasting X for such extreme values, since the occurrence of such would force us to reconsider our modelling assumptions, as in [7,24,25].

We are in effect assuming a model in which the algebraic behavior holds, given α, to a satisfactory approximation for $A \leq X \leq k$, and that eventually there is 0 (or negligible) mass beyond some known $K > k$. We assume that the same k is appropriate for all values of α being given positive weight. Between k and K there must be a transition from the algebraic tail behavior up to k and the negligible mass beyond K. In this transition zone the tail of the distribution may not even be approximately algebraic, and if algebraic, may have a different tail index. The mass between k and K need not be entirely negligible, but we assume there is no data-based or other information concerning the form of the mass distribution in this interval, apart from the fact that the total mass in the interval is smaller then $C \times k^{-\alpha}$, as is required by the model. If k is large enough, then $C \times k^{-\alpha}$, although not entirely negligible, may be sufficiently small so that the mass between k and $K < \infty$ has only a slight effect upon the posterior moments for X_{n+1}. We shall assume that this is the case, so that the tail distribution is of algebraic form from A to k, while beyond k, although not 0 or entirely negligible, the mass is of no practical importance for the assessment of the posterior moments of X_{n+1}.

Typically, the posterior expectation of $C \times x_{(1)}^{-\alpha}$ will be of order of magnitude $1/(n+1)$ based on a previous sample of size n. Compare the maximum-likelihood estimator $\hat{C}_1$ of [1, p. 1168]. This also corresponds to the fiducial analysis of Fisher [26, p. 210], and to the Bayesian non-parametric procedure A_n of Hill [22,27,35]. Thus before observing $X_1,\ldots,X_n$, because of the exchangeability there is an unconditional probability of $1/(n+1)$ that X_{n+1} will be the maximum, which suggests that even conditionally this will often be of the right order of magnitude. As shown in [5], there is an explicit parametric model, called a splitting process, for which this evaluation holds exactly, and such an evaluation is coherent in the sense of de Finetti [28,29].

The constant K plays virtually no direct role in the following analysis, but is important because of the delicate issues that arise when $\alpha \leq 2$. In this case if there were no finite upper bound K and the algebraic tail were assumed valid everywhere beyond A, then the posterior predictive variance of the next observation would be infinite; and the predictive expectation would also be infinite unless the *a priori* distribution for α gave sufficiently small weight to values near 1. There is no known reason that α must be larger than 2, or even larger than 1, and the data may in fact clearly suggest that it is smaller than 1. But an infinite predictive expectation would not correspond to any real world problem that I know of concerning extreme data, and I doubt that one could seriously recommend such predictions. For example, they would lead to

terribly poor performance if predictions were made and assessed according to some proper scoring rule or loss function. This change in viewpoint to reflect the boundedness of the data gives rise to some surprising consequences with regard to prediction.

The key choice concerns not K but k, since even if there were a known finite upper bound K for the data, it might still not be appropriate to assume the algebraic form all the way up to K, but only that in the domain of practical importance the tail is of this form, say up to k, which is equal to some appropriate upper percentile of the distribution. This is in essence a modelling assumption, just as when we assume that the normal model for data is sufficiently closely satisfied to be useful in the analysis of that data. Modelling assumptions are rarely exactly true, but they are sometimes indispensable in order to proceed, and often give useful results. See [7,25,27]. The form of analysis that we recommend is a conditional analysis, given a specification of k. For example, with $A = x_{(1)}$, we consider predictive inference about the next observation given that it lies between $x_{(1)}$ and some $k > x_{(1)}$. If $L = k/x_{(1)}$, then we find that it typically makes a great difference whether L is of order 5 or order 100, both with respect to the posterior predictive mean and the posterior predictive variance for the next observation. Based upon the mathematical and computer analysis in the next sections, we recommend that the forecaster make a choice of L, usually with $L \leq 10$ and sometimes even with $L = 2$. To illustrate, when L is chosen to be 3, the adequacy of our modelling assumption depends on whether it is or is not the case that the algebraic form holds between $x_{(1)}$ and $3 \times x_{(1)}$, with the mass beyond $3 \times x_{(1)}$ no longer even approximately of the algebraic form with the same α as between $x_{(1)}$ and $3 \times x_{(1)}$, and also with the mass beyond $3 \times x_{(1)}$ sufficiently small so that for practical purposes it can be ignored. In principle the optimal choice of k is the largest value for which the algebraic assumption holds exactly (or in a suitable sense, approximately); while beyond that k the tail is no longer of that same form, and also is of little practical importance in the evaluation of the first two posterior predictive moments. It would be difficult if not impossible in typical real-world problems to find such an optimal k, and so we recommend that several values of k be chosen, yielding different values for the posterior predictive moments, and then by means of judgment and data-analytic methods that a choice be made to yield a forecast. See for example Sec. 5 of [1] for a closely related type of data-analysis. Such analyses must be made on a computer, rather than purely mathematically, and can be quite demanding computationally.

We emphasize that it does not seem possible to avoid such considerations as to the choice of k, since in even the best of cases, where the tail of the distribution is known to be of the algebraic form in the domain of interest, the only alternative to such an analysis is to simply ignore the boundedness of the data, and take $k = \infty$. But then our prediction of the next record value can become infinite, which is absurd in most real-world problems. Hence the algebraic tailed model with $1 \leq \alpha \leq 2$ is not compatible with unbounded data unless the *a priori* distribution is chosen to give suitably small weight to values of α close to 1. There may be little or no evidence for choosing the *a priori* distribution in this way, and it does not seem appropriate to do so merely to avoid the issue, just as it does not seem appropriate to replace the expectation by the median merely to avoid the issue. At any rate, this article shows that effective predictions can be made with any prior distribution for α, including cases where $\alpha \leq 1$, provided that one can justify some finite upper bound K for the observations.

Our underlying motivation is that given the unreliability of assessments of the far upper tail of a distribution, for predictive purposes it may be appropriate to ignore this far upper tail, i.e., the part beyond k, or equivalently, to condition upon X falling in some finite interval, say $(x_{(1)}, L \times x_{(1)})$, for which the algebraic assumption is believed to be valid, and beyond which there is no assumption that is believed trustworthy. It is implicit in this analysis that there is little mass beyond k, and that in ignoring the case $X \geq k$ for some appropriately chosen k, one loses little, while gaining the power of a statistical analysis based upon the extreme value model with some $\alpha > 0$. In the case of a known finite upper bound K, in effect we perform conditional inference, given that the observation is not too large, and then examine sensitivity to the choice of k. The same is true if the random variable is unbounded and $K = \infty$, since again beyond a certain percentile one would have no empirical basis for any assumption in the far upper tail. Whatever extreme value theory exists for tails of distributions could not be expected to hold literally in the far upper tail of the distribution, where no data has been observed. Nevertheless, one may have to make some forecasts, and it would appear reasonable to assume that the algebraic assumption holds for at least some distance beyond $x_{(1)}$. If this, or some other assumed model does not hold beyond $x_{(1)}$ then plainly no serious theory-based

forecasting is possible. But if through data analysis, as in [1,26], it has been discovered that the algebraic assumption is acceptable for say the upper $r+1$ order statistics of the past data, then it would be reasonable to anticipate that this will also be true for some distance beyond $x_{(1)}$. A Bayesian theory of data analysis is put forth in [25] which indicates how the classical Bayesian approach must be modified to deal with issues that arise from such data analysis.

Finally, real world data sets of interest in regard to the forecasting of extreme values are not necessarily of the long-tailed algebraic form that we have discussed. In this case we recommend that a transformation be first applied to the data in order to make the upper tail of the long-tailed form. For example, if the tail is of Weibull form, then the transformation to $\exp X^s$ yields an algebraic tail, as discussed in [1,10]. When the form of the tail is unknown, data-analytic methods can be used to determine an appropriate transformation. In this way, having learned how to forecast extreme tails for the long-tailed distributions as a type of standard case, we can also apply our methods to distributions not of this form in the upper tail, and then take the inverse transformation to forecast the extreme values in the original units in which the data were measured. Such methods are quite common in statistics, for example in transforming data in order to obtain approximate normality, using normal methods for analysis of the data, and then transforming back to the original units. In the Bayesian scenario it is even possible to provide a strong justification for these methods, since conditional upon the data, one can quite freely transform the parameters, and obtain the posterior distribution for the new parameters by the usual calculus of transformations.

2. Predictive Moments for Known α

Our object is to evaluate, as meaningfully and robustly as possible, the posterior moments

$$E(X_{n+1}^i | A \leq X_{n+1} \leq k),$$

for specified A and k, and $i = 1,2$. The primary application will be in the case where there has been a previous sample, $X_1,\ldots,X_n$. Let D denote the data $X_1 = x_1,\ldots,X_n = x_n$. Given this data, we wish to forecast the next observation X_{n+1}. It is notationally convenient to refer to X_{n+1} as X from now on. Since A will usually be held fixed, we suppress it in the notation. To evaluate the posterior predictive expectation of X we first condition on α, to obtain

$$f(k,\alpha) = E(X | A \leq X \leq k, \alpha),$$

and then we take the expectation of this quantity with respect to the posterior distribution of α to obtain the predictive expectation of primary interest.

Based upon our assumption that the tail is algebraic between A and k, we obtain

$$f(k,\alpha) = \frac{\int_A^k x^{-\alpha}\,dx}{\int_A^k x^{-\alpha-1}\,dx}.$$

For $L = \frac{k}{A}$, this yields:

$$f(k,\alpha) = \begin{cases} A \times \frac{\alpha}{\alpha-1} \times \frac{1-L^{1-\alpha}}{1-L^{-\alpha}} & \text{if } \alpha \neq 0,1 \\ A \times \ln(L) \times \frac{L}{L-1} & \text{if } \alpha = 1 \\ \frac{A \times (L-1)}{\ln(L)} & \text{if } \alpha = 0. \end{cases} \quad (1)$$

For $\alpha \neq 0$, 1, we can also write:

$$f(k,\alpha) = A \times \frac{\alpha}{\alpha-1} \times L \times \frac{L^{\alpha-1}-1}{L^{\alpha}-1}. \quad (2)$$

A similar equation is available for $f^{(2)}(k,\alpha) \equiv E(X^2 | A \leq X \leq k, \alpha)$. We obtain:

$$f^{(2)}(k,\alpha) = \begin{cases} A^2 \times \frac{\alpha}{\alpha-2} \times \frac{1-L^{2-\alpha}}{1-L^{-\alpha}} & \text{f } \alpha \neq 0,2 \\ 2 \times A^2 \times \ln(L) \times [\frac{L^2}{L^2-1}] & \text{if } \alpha = 2 \\ A^2 \times \frac{L^2-1}{2 \times \ln(L)} & \text{if } \alpha = 0. \end{cases} \quad (3)$$

The posterior predictive variance for a future record value X, given α, is therefore

$$V(k,\alpha) \equiv f^{(2)}(k,\alpha) - [f(k,\alpha)]^2. \quad (4)$$

It follows from (2) that for $\alpha > 1$, as $L \to \infty$ we have

$$f(k,\alpha) \sim A \times \frac{\alpha}{\alpha-1}. \quad (5)$$

When $\alpha > 1$, the right-hand side of (5) decreases from ∞ for $\alpha = 1$ to the value $2 \times A$ when $\alpha = 2$, with the value $3 \times A$ when $\alpha = 1.5$. Provided that α is bounded away from 1 this expectation remains bounded.

For $\alpha \leq 2$, the posterior predictive variance goes to ∞ as $L \to \infty$. If we define $\epsilon = 2 - \alpha > 0$ then for large L

$$f^{(2)}(k,\alpha) \approx A^2 \times \alpha \times \frac{L^{\epsilon}-1}{\epsilon}. \quad (6)$$

For each $L>1$, and for $\epsilon>0$, the function $\phi(\epsilon)=\frac{L^{\epsilon}-1}{\epsilon}$ is monotonically increasing in ϵ. For $0<\epsilon\leq 2$ it has a maximum value of $\frac{[L^2-1]}{2}$ when $\epsilon=2$, and an infimum of $\ln(L)$ as $\epsilon\to 0$. For large L, as $\epsilon\to 0$ we see from Eqs. (3), (5), and (6), that

$$V(k,\alpha)\approx A^2\times(\alpha\ln(L)-[\frac{\alpha}{\alpha-1}]^2). \quad (7)$$

From Eq. (3), it follows that for $\alpha>2$ the posterior predictive variance remains bounded, and as $L\to\infty$ tends to the limiting value

$$A^2\times\frac{\alpha}{(\alpha-2)(\alpha-1)^2}. \quad (8)$$

Now consider the forecasting of the maximum of N future observations. Define

$$M=max[X_{n+1},\ldots,X_{n+N}],$$

and let $\pi^*(\alpha,C)$ be the posterior distribution for α,C, based upon the data D. The likelihood function $L_2(\alpha,\beta)$ of [1], when converted from lower tail to upper tail inference, can be used to obtain this posterior distribution. For $t>A$, we have

$$P(M>t|D)=\int_0^{\infty}\int_0^{\infty}[1-(1-C\times t^{-\alpha})^N]$$

$$\pi^*(\alpha,C)\,d\alpha\,dC. \quad (9)$$

When $N=1$ this gives the posterior predictive distribution for a single new observation considered earlier, except that here we have not yet conditioned upon $X\geq A$. Just as before, one can consider the posterior moments of M, given that $M\geq A$. When N is not small it is very probable that $M\geq x_{(1)}$, so that a new record will be set. Thus for large N the predictive distribution of M will be approximately the same as the predictive distribution of M, given $M\geq x_{(1)}$.

In Table 1 we present for several values of α the predictive moments as obtained by numerical integration. The predictive mean is denoted by $E^*(X)$ and the predictive standard deviation by $SD^*(X)$. The column labelled DIST gives the posterior predictive probability that X is larger than 2, 3, and 5 times A. Values of α go from .10 to 1.90, and values for L go from 1.25 to 10^6. It can be checked that the above asymptotic formulas hold quite closely for fixed α.

We see from Table 1 that the posterior expectation of X, given that $X>A$, is only a few multiples of A, even when α is as small as .10, provided that $L\leq 10$. In an important class of application A is taken to be $x_{(1)}$, so that the real action takes place with regard to a few multiples of the largest observation yet observed. When $L\leq 2$ we see that the value of α between .10 and 1.90 has very little effect on the posterior predictive first and second moments. On the other hand, when L is very large the value of α has a huge effect. For example, the posterior expectation drops from $37{,}297\times A$ when $L=10^6$ to $2.11\times A$, as α changes from .10 to 1.90. The choice of L can make a huge difference when $\alpha\leq 1$. However, in many applications of extreme value theory, it could safely be assumed that $L\leq 10$, in which case L has only a minor effect even when $\alpha\leq 1$. The choice of L has a greater effect with regard to the predictive variance, but again if $L\leq 10$ there is substantial robustness.[4] Thus the first conclusion that we draw is that in a real-world problem, where there has been substantial data, such as with regard to wind velocities, temperatures, etc., and where one does not take seriously the possibility of the next record value being an enormous multiple of the current maximum, the precise choice of α and L has a limited effect upon the forecast. This is precisely what we are aiming for, namely an approach in which one can seriously input *a priori* knowledge regarding α and L in such a way as to see clearly the real but limited effect of such choices.

Table 1 refers to the case of known α. In practice α will ordinarily be unknown. The Bayesian approach is to employ some *a priori* distribution π for α, obtain the posterior distribution for α given D, and then obtain the posterior expectation of X, given that $A\leq X\leq k$. For a specified k, this posterior expectation can be written as

$$f(k)=E[E(X|D,A\leq X\leq k,\alpha)]=E[f(k,\alpha)], \quad (10)$$

where the last expectation is taken with respect to the posterior distribution of α. Similarly, the posterior second moment for X is obtained by evaluating

$$f^{(2)}(k)=E[E(X^2|D,A\leq X\leq k,\alpha)]=E[f^{(2)}(k,\alpha)]. \quad (11)$$

We employ the theory of [1] to obtain a likelihood function for the parameter α based upon the upper order statistics of the past data. We first condition upon the upper $r+1$ order statistics of the data lying in the region where the tail is of

[4] See [30] for a general formulation of the robustness problem in Bayesian statistics.

Table 1. Fixed ALPHA

ALPHA	PRED		DIST			BOUND
α	$E^*(X)$	$SD^*(X)$	2	3	5	L
.10	1.12	.07	.93	.85	.79	1.25
.10	1.23	.14	.93	.85	.79	1.50
.10	1.44	.29	.93	.85	.79	2
.10	1.80	.57	.93	.85	.79	3
.10	2.43	1.12	.93	.85	.79	5
.10	3.75	2.45	.93	.85	.79	10
.10	18.70	23.46	.93	.85	.79	100
.10	734.88	1715.25	.93	.85	.79	10^4
.10	37297.27	1.28×10^5	.93	.85	.79	10^6
.50	1.12	.07	.71	.45	.32	1.25
.50	1.22	.14	.71	.45	.32	1.50
.50	1.41	.28	.71	.45	.32	2
.50	1.73	.56	.71	.45	.32	3
.50	2.24	1.07	.71	.45	.32	5
.50	3.16	2.22	.71	.45	.32	10
.50	10.00	16.43	.71	.45	.32	100
.50	100.05	571.77	.71	.45	.32	10^4
.50	1001.62	18257.56	.71	.45	.32	10^6
.90	1.12	.07	.54	.23	.13	1.25
.90	1.22	.14	.54	.23	.13	1.50
.90	1.39	.28	.54	.23	.13	2
.90	1.66	.54	.54	.23	.13	3
.90	2.05	1.00	.54	.23	.13	5
.90	2.67	1.93	.54	.23	.13	10
.90	5.35	10.12	.54	.23	.13	100
.90	13.62	142.79	.54	.23	.13	10^4
.90	26.86	1806.66	.54	.23	.13	10^6
1.10	1.12	.07	.47	.17	.08	1.25
1.10	1.21	.14	.47	.17	.08	1.50
1.10	1.38	.28	.47	.17	.08	2
1.10	1.63	.53	.47	.17	.08	3
1.10	1.97	.96	.47	.17	.08	5
1.10	2.46	1.78	.47	.17	.08	10
1.10	4.09	7.73	.47	.17	.08	100
1.10	6.62	69.46	.47	.17	.08	10^4
1.10	8.24	554.67	.47	.17	.08	10^6
1.50	1.11	.07	.35	.09	.03	1.25
1.50	1.21	.14	.35	.09	.03	1.50
1.50	1.36	.27	.35	.09	.03	2
1.50	1.57	.50	.35	.09	.03	3
1.50	1.82	.87	.35	.09	.03	5
1.50	2.12	1.49	.35	.09	.03	10
1.50	2.70	4.44	.35	.09	.03	100
1.50	2.97	16.98	.35	.09	.03	10^4
1.50	3.00	54.72	.35	.09	.03	10^6
1.90	1.11	.07	.27	.05	.01	1.25
1.90	1.20	.14	.27	.05	.01	1.50
1.90	1.34	.27	.27	.05	.01	2
1.90	1.51	.48	.27	.05	.01	3
1.90	1.69	.78	.27	.05	.01	5
1.90	1.87	1.22	.27	.05	.01	10
1.90	2.08	2.61	.27	.05	.01	100
1.90	2.11	4.93	.27	.05	.01	10^4
1.90	2.11	7.23	.27	.05	.01	10^6

algebraic form, i.e., larger than D of [1], and then condition upon the values of the ratios of upper order statistics $v_i = x^{(i)}/x^{(i+1)}$, for $i = 1,...,r$. As shown in [1], if we are indeed in the upper tail of the distribution where the algebraic form holds, then conditional upon α, the quantities $e_i = i \times \ln v_i$ are independent with a common exponential distribution having parameter α. A sufficient statistic for α, conditional upon the v_i and r, is then

$$t = t(r) = \sum_{i=1}^{r} e_i. \tag{12}$$

The (conditional) likelihood function based upon r and t is then

$$L(\alpha) \propto \alpha^r \times \exp[-\alpha t], \tag{13}$$

for $\alpha > 0$. In conjunction with some *a priori* distribution for α this likelihood function can be used to obtain the posterior distribution for α. If k is large and $\alpha > 1$, we see from (5) that

$$E(X|D, A \leq X \leq k, \alpha) \approx A \times \frac{\alpha}{\alpha - 1}. \tag{14}$$

In general, the predictive moments of X can only be obtained by numerical integration. In Sec. 4 we examine the sensitivity of such quantities to the data, choice of L, and choice of *a priori* distribution for α. The case $k = \infty$, however, has a closed form analytic solution for a Gamma *a priori* distribution of α, and this contributes some insight into the behavior or the posterior moments of X.

3. $k = \infty$

In this section we examine the special case in which the distribution is known to be algebraic everywhere beyond A. In this case, in order for posterior moments to be finite, we will have to assume that α is sufficiently large. It follows from Eq. (1) that the posterior expectation of X_{n+1}, given that it is in the upper tail and α, is finite if and only if $\alpha > 1$. In the Bayesian analysis, with an *a priori* distribution for α, the unconditional posterior expectation of X is finite if and only if the *a priori* distribution sufficiently downweights values of α near 1.

We can gain some insight by supposing that $\alpha > 1$ has the prior distribution

$$\pi(\alpha) = c \times (\alpha - 1)^{\delta - 1} \exp[-\beta(\alpha - 1)],$$

for $\delta, \beta > 0$, where $c = \Gamma(\delta)/\beta^{\delta}$ is a proportionality constant. In other words, we give $\alpha - 1 > 0$ a Gamma *a priori* distribution. If $\delta > 1$ we obtain from Eq. (1) that the posterior expectation of X/A, given $X \geq A$, is

$$E\left(\frac{\alpha}{\alpha - 1} | D\right) = \frac{\int_0^{\infty} (1+s)^{r+1} \times s^{\delta - 2} \times \exp[-(t+\beta)s]\,ds}{\int_0^{\infty} (1+s)^{r} \times s^{\delta - 1} \times \exp[-(t+\beta)s]\,ds}. \tag{15}$$

This expectation is finite provided that $\delta > 1$.

For positive integral values of r we can expand the powers of $1+s$ using the binomial theorem, and this allows us to make explicit evaluations. To illustrate, if $r = 1$ as in the forecasting of city sizes in Tables 6 and 7, we have

$$E\left(\frac{\alpha}{\alpha - 1} | D\right) = \frac{1 + 2(\delta - 1)/(t+\beta) + \delta(\delta - 1)/(t+\beta)^2}{1 + \delta/(t+\beta)} \times \frac{t+\beta}{\delta - 1}. \tag{16}$$

This reveals the manner in which the expectation blows up as $\delta \to 1$. When $\delta = 2$, the right-hand side can be written as

$$1 + \frac{(t+\beta)(t+\beta+1)}{t+\beta+2}.$$

For $t + \beta = 1$, we obtain the value 1.67. This is comparable with the values in Tables 2, 3, and 4, when $r = t = 1$, and $L \leq 5$. For $r = 1$ and $\delta = 2$, $f(k)$ is approximately $(1 + t + \beta) \times A$, provided that $t + \beta$ is sufficiently large. Similarly, other integral values of r yield closed form expressions, which provide some insight as to the behavior or the posterior expectation of X.

From Eqs. (3) and (11), the posterior predictive second moment for X, given that $X \geq A$, is

$$f^{(2)}(k) = A^2 \times E\left[\frac{\alpha}{\alpha - 2} | D, X \geq A\right]. \tag{17}$$

If $\alpha > 2$ and the *a priori* distribution for $\alpha - 2$ is of the Gamma form, with parameters δ, β, the posterior predictive variance for X will be finite, provided that $\delta > 1$. Closed form expressions can be obtained when r is a positive integer, just as with the corresponding predictive first moment.

Table 2. Uniform prior, LB = 1.001, UB = 1.999, prior mean = 1.50, SD = .29

DATA		POST		PRED		DIST			BOUND
r	t	$E^*(\alpha)$	$SD^*(\alpha)$	$E^*(X)$	$SD^*(X)$	2	5	10	L
1	1	1.47	.29	1.11	.07	.37	.10	.04	1.25
1	1	1.47	.29	1.21	.14	.37	.10	.04	1.50
1	1	1.47	.29	1.36	.37	.27	.10	.04	2
1	1	1.47	.29	1.58	.51	.27	.10	.04	3
1	1	1.47	.29	1.84	.88	.37	.10	.04	5
1	1	1.47	.29	2.16	1.54	.37	.10	.04	10
1	1	1.47	.29	2.96	5.37	.37	.10	.04	10^2
1	1	1.47	.29	3.82	38.95	.37	.10	.04	10^4
1	1	1.47	.29	4.30	305.19	.37	.10	.04	10^6
3	2	1.50	.28	1.11	.07	.36	.10	.04	1.25
3	2	1.50	.28	1.21	.14	.36	.10	.04	1.50
3	2	1.50	.28	1.36	.27	.36	.10	.04	2
3	2	1.50	.28	1.57	.51	.36	.10	.04	3
3	2	1.50	.28	1.83	.88	.36	.10	.04	5
3	2	1.50	.28	2.14	1.52	.36	.10	.04	10
3	2	1.50	.28	2.88	5.16	.36	.10	.04	10^2
3	2	1.50	.28	3.63	35.99	.36	.10	.04	10^4
3	2	1.50	.28	4.03	276.83	.36	.10	.04	10^6
2	3	1.37	.27	1.11	.07	.39	.12	.05	1.25
2	3	1.37	.27	1.21	.14	.39	.12	.05	1.50
2	3	1.37	.27	1.37	.28	.39	.12	.05	2
2	3	1.37	.27	1.59	.51	.39	.12	.05	3
2	3	1.37	.27	1.87	.90	.39	.12	.05	5
2	3	1.37	.27	2.24	1.61	.39	.36	.36	10
2	2	1.37	.27	3.22	5.99	.39	.36	.36	10^2
2	3	1.37	.27	4.43	46.85	.39	.36	.36	10^4
2	3	1.37	.27	5.16	377.37	.39	.36	.36	10^6
5	1	1.67	.25	1.11	.07	.32	.07	.03	1.25
5	1	1.67	.25	1.21	.14	.32	.07	.03	1.50
5	1	1.67	.25	1.35	.27	.32	.07	.03	2
5	1	1.67	.25	1.55	.49	.32	.07	.03	3
5	1	1.67	.25	1.77	.84	.32	.07	.03	5
5	1	1.67	.25	2.02	1.40	.32	.07	.03	10
5	1	1.67	.25	2.49	4.05	.32	.07	.03	10^2
5	1	1.67	.25	2.80	21.99	.32	.07	.03	10^4
5	1	1.67	.25	2.93	152.19	.32	.07	.03	10^5
1	5	1.22	.20	1.11	.07	.43	.15	.07	1.25
1	5	1.22	.20	1.21	.14	.43	.15	.07	1.50
1	5	1.22	.20	1.37	.28	.43	.15	.07	2
1	5	1.22	.20	1.61	.52	.43	.15	.07	3
1	5	1.22	.20	1.93	.93	.43	.15	.07	5
1	5	1.22	.20	2.36	1.71	.43	.15	.07	10
1	5	1.22	.20	3.68	6.97	.43	.15	.07	10^2
1	5	1.22	.20	5.61	60.75	.43	.15	.07	10^4
1	5	1.22	.20	6.94	512.81	.43	.15	.07	10^6
30	20	1.52	.23	1.11	.07	.35	.09	.03	1.25
30	20	1.52	.23	1.21	.14	.35	.09	.03	1.50
30	20	1.52	.23	1.36	.27	.35	.09	.03	2
30	20	1.52	.23	1.57	.50	.35	.09	.03	3
30	20	1.52	.23	1.82	.87	.35	.09	.03	5
30	20	1.52	.23	2.12	1.50	.35	.09	.03	10
30	20	1.52	.23	2.78	4.81	.35	.09	.03	100
30	20	1.52	.23	3.29	28.02	.35	.09	.03	10^4
30	20	1.52	.23	3.48	185.91	.35	.09	.03	10^6

Table 2. Uniform prior, LB = 1.001, UB = 1.999, prior mean = 1.50, SD = .29—Continued

DATA		POST		PRED		DIST			BOUND
r	t	$E^*(\alpha)$	$SD^*(\alpha)$	$E^*(X)$	$SD^*(X)$	2	5	10	L
20	30	1.08	.07	1.12	.07	.47	.18	.08	1.25
20	30	1.08	.07	1.22	.14	.47	.18	.08	1.50
20	30	1.08	.07	1.38	.28	.47	.18	.08	2
20	30	1.08	.07	1.63	.53	.47	.18	.08	3
20	30	1.08	.07	1.98	.96	.47	.18	.08	5
20	30	1.08	.07	2.48	1.80	.47	.18	.08	10
20	30	1.08	.07	4.21	8.01	.47	.18	.08	100
20	30	1.08	.07	7.27	78.63	.47	.18	.08	10^4
20	30	1.08	.07	9.71	706.23	.47	.18	.08	10^6
300	200	1.50	.09	1.11	.07	.35	.09	.03	1.25
300	200	1.50	.09	1.21	.14	.35	.09	.03	1.50
300	200	1.50	.09	1.36	.27	.35	.09	.03	2
300	200	1.50	.09	1.57	.50	.35	.09	.03	3
300	200	1.50	.09	1.82	.87	.35	.09	.03	5
300	200	1.50	.09	2.12	1.49	.35	.09	.03	10
300	200	1.50	.09	2.71	4.48	.35	.09	.03	10^2
300	200	1.50	.09	3.00	18.24	.35	.09	.03	10^4
300	200	.50	.09	3.04	67.25	.35	.09	.03	10^6
200	300	1.01	.01	1.12	.07	.50	.20	.10	1.25
200	300	1.01	.01	1.22	.14	.50	.20	.10	1.50
200	300	1.01	.01	1.39	.28	.50	.20	.10	2
200	300	1.01	.01	1.65	.53	.50	.20	.10	3
200	300	1.01	.01	2.01	.97	.50	.20	.10	5
200	300	1.01	.01	2.55	1.85	.50	.20	.10	10
200	300	1.01	.01	4.59	8.73	.50	.20	.10	10^2
200	300	1.01	.01	8.88	95.94	.50	.20	.10	10^4
200	300	1.01	.01	13.01	941.80	.50	.20	.10	10^6

4. $k < \infty$

One of our purposes in this article is to show that prediction can be very sensitive to the *a priori* information introduced regarding L, and that it is essential to incorporate strong *a priori* information as to the magnitude of this quantity in order to obtain realistic forecasts. No closed form results are available apart from those of the last section. We consider now various *a priori* distributions for α. In the previous analysis it was not possible to give α a uniform distribution, since this would require $\beta = 0$ and $\delta = 1$, in which case with infinite k the expectation is infinite. However, with a finite upper bound for X, we obtain a finite expectation for any $\alpha \geq 0$, and in fact even for negative α, although this case is of little interest.

Table 2 displays results for the case of a uniform *a priori* distribution for α, using a finite grid of possible values for α between LB = 1.001 and UB = 1.999, several values of r and t, and several choices of L. The prior expectation and standard deviation for α and 1.50 and .29, respectively. Table 3 gives such results for a uniform *a priori* distribution, using a finite grid of values between LB = .001 and UB = 1.999, in which case the prior expectation and standard deviation for α are 1.00 and .58, respectively. In these tables the column labelled "POST" gives the posterior expectation and standard deviation for α, the column labelled "PRED" gives the posterior predictive expectation and standard deviation for the next observation X, and the column labelled "DIST" gives the posterior probability that X is larger than 2, 5, and 10 times A, respectively.

So far we have only considered very strong *a priori* knowledge, such as in Table 1 where α is known, and very weak *a priori* knowledge, such as the uniform distributions of Tables 2 and 3. In applications it is important also to be able to input an *a priori* distribution for α in which some values are singled out as being given substantially more weight than others. A useful family of *a priori* distributions for α for this purpose is the three-

Table 3. Uniform prior, LB = 0.001, UB = 1.999, prior mean = 1.00, SD = .58

DATA		POST		PRED		DIST			BOUND
r	t	$E^*(\alpha)$	$SD^*(\alpha)$	$E^*(X)$	$SD^*(X)$	2	5	10	L
1	1	1.09	.51	1.12	.07	.50	.24	.15	1.25
1	1	1.09	.51	1.21	.14	.50	.24	.15	1.50
1	1	1.09	.51	1.38	.28	.50	.24	.15	2
1	1	1.09	.51	1.64	.53	.50	.24	.15	3
1	1	1.09	.51	1.99	.98	.50	.24	.15	5
1	1	1.09	.51	2.55	1.90	.50	.24	.15	10
1	1	1.09	.51	5.71	11.70	.50	.24	.15	100
1	1	1.09	.51	59.13	474.46	.50	.24	.15	10^4
1	1	1.00	.51	1599.34	26459.07	.50	.24	.15	10^6
3	2	1.31	.42	1.11	.07	.42	.15	.08	1.25
3	2	1.31	.42	1.21	.14	.42	.15	.08	1.50
3	2	1.31	.42	1.37	.28	.42	.15	.08	2
3	2	1.31	.42	1.60	.52	.42	.15	.08	3
3	2	1.31	.42	1.90	.93	.42	.15	.08	5
3	2	1.31	.42	2.32	1.70	.42	.15	.08	10
3	2	1.31	.42	3.98	8.23	.42	.15	.08	100
3	2	1.31	.42	15.97	209.46	.42	.15	.08	10^4
3	2	1.31	.42	187.75	8561.51	.42	.15	.08	10^6
2	3	.90	.44	1.12	.07	.56	.29	.19	1.25
2	3	.90	.44	1.22	.14	.56	.29	.19	1.50
2	3	.90	.44	1.39	.28	.56	.29	.19	2
2	3	.90	.44	1.67	.54	.56	.29	.19	3
2	3	.90	.44	2.07	1.01	.56	.29	.19	5
2	3	.90	.44	2.72	2.01	.56	.29	.19	10
2	3	.90	.44	6.76	13.07	.56	.29	.19	100
2	3	.90	.44	70.77	511.32	.56	.29	.19	10^4
2	3	.90	.44	1619.79	26176.42	.56	.29	.19	10^6
5	1	1.64	.29	1.11	.07	.33	.08	.03	1.25
5	1	1.64	.29	1.21	.14	.33	.08	.03	1.50
5	1	1.64	.29	1.35	.27	.33	.08	.03	2
5	1	1.64	.29	1.55	.50	.33	.08	.03	3
5	1	1.64	.29	1.78	.85	.33	.08	.03	5
5	1	1.64	.29	2.04	1.43	.33	.08	.03	10
5	1	1.64	.29	2.62	4.57	.33	.08	.03	100
5	1	1.64	.29	3.57	50.32	.33	.08	.03	10^4
5	1	1.64	.29	7.89	1247.04	.33	.08	.03	10^6
1	5	.40	.28	1.12	.07	.77	.57	.47	1.25
1	5	.40	.28	1.23	.14	.77	.57	.47	1.50
1	5	.40	.28	1.42	.29	.77	.57	.47	2
1	5	.40	.28	1.75	.56	.77	.57	.47	3
1	5	.40	.28	2.29	1.09	.77	.57	.47	5
1	5	.40	.28	3.33	2.31	.77	.57	.47	10
1	5	.40	.28	12.74	19.46	.77	.57	.47	100
1	5	.40	.28	309.24	1125.56	.77	.57	.47	10^4
1	5	.40	.28	11841.58	72682.37	.77	.57	.47	10^6
30	20	1.51	.24	1.11	.07	.36	.09	.04	1.25
30	20	1.51	.24	1.21	.14	.36	.09	.04	1.50
30	20	1.51	.24	1.36	.27	.36	.09	.04	2
30	20	1.51	.24	1.57	.50	.36	.09	.04	3
30	20	1.51	.24	1.82	.87	.36	.09	.04	5
30	20	1.51	.24	2.13	1.50	.36	.09	.04	10
30	20	1.51	.24	2.81	4.92	.36	.09	.04	100
30	20	1.51	.24	3.42	32.25	.36	.09	.04	10^4
30	20	1.51	.24	3.80	284.87	.36	.09	.04	10^6

Table 3. Uniform prior, LB = 0.001, UB = 1.999, prior mean = 1.00, SD = .58—Continued

DATA		POST		PRED		DIST			BOUND
r	t	$E^*(\alpha)$	$SD^*(\alpha)$	$E^*(X)$	$SD^*(X)$	2	5	10	L
20	30	.70	.15	1.12	.07	.62	.33	.21	1.25
20	30	.70	.15	1.22	.14	.62	.33	.21	1.50
20	30	.70	.15	1.40	.28	.62	.33	.21	2
20	30	.70	.15	1.70	.55	.62	.33	.21	3
20	30	.70	.15	2.14	1.03	.62	.33	.21	5
20	30	.70	.15	2.91	2.09	.62	.33	.21	10
20	30	.70	.15	7.49	13.50	.62	.33	.21	100
20	30	.70	.15	46.87	366.86	.62	.33	.21	10^4
20	30	.70	.15	357.77	10632.53	.62	.33	.21	10^6
300	200	1.50	.09	1.11	.07	.35	.09	.03	1.25
300	200	1.50	.09	1.21	.14	.35	.09	.03	1.50
300	200	1.50	.09	1.36	.27	.35	.09	.03	2
300	200	1.50	.09	1.57	.50	.35	.09	.03	3
300	200	1.50	.09	1.82	.87	.35	.09	.03	5
300	200	1.50	.09	2.12	1.49	.35	.09	.03	10
300	200	1.50	.09	2.71	4.48	.35	.09	.03	100
300	200	1.50	.09	3.00	18.24	.35	.09	.03	10^4
300	200	1.50	.09	3.04	67.25	.35	.09	.03	10^6
200	300	.67	.05	1.12	.07	.63	.34	.22	1.25
200	300	.67	.05	1.22	.14	.63	.34	.22	1.50
200	300	.67	.05	1.40	.28	.63	.34	.22	2
200	300	.67	.05	1.70	.55	.63	.34	.22	3
200	300	.67	.05	2.16	1.04	.63	.34	.22	5
200	300	.67	.05	2.94	2.10	.63	.34	.22	10
200	300	.67	.05	7.62	13.58	.63	.34	.22	100
200	300	.67	.05	41.81	330.65	.63	.34	.22	10^4
200	300	.67	.05	212.71	7448.90	.63	.34	.22	10^6

parameter log-normal family. Suppose that $\ln(\alpha-\gamma)\sim N(\mu,\sigma^2)$. This is the three-parameter log-normal distribution with threshold parameter γ, and is a very convenient and interesting family with which to make inference about α. See Aitchison and Brown [31], and Hill [32] for some properties of this distribution. The integrations in this case again have to be done by numerical analysis. In Table 4 we present results for the case $\gamma=1$, with α taking values between LB = 1.001 and UB = 10. The prior mean and standard deviation for α are 1.50 and .61, respectively.

5. Discussion of Tables

If $\alpha>2$ then for fixed known α there is no problem with infinite first and second moments. This is also the case when α is unknown, except that the *a priori* distribution for α must give sufficiently small weight to values near 2 in order that the second moment be finite. However, the case $\alpha>2$, although of some interest, does not deal with truly long-tailed distributions. For $\alpha>1$, and using a Gamma prior distribution for $\alpha-1$ with $\delta>1$, as $k\to\infty$ the posterior moments of X converge to the limiting results discussed in Sec. 3, such as in Eq. (16). We observe, however, that the convergence is quite slow. For values of k in the practical range, say $L\leq 10$, the results are not very sensitive to the precise value of L, but are quite different from the limiting results, because the convergence is so slow. For example, the theoretical value for the multiplier of A when $r=0,t=1,\delta=2,\beta=1$, is 3. Using UB = 10, when $L=10^{12}$ the calculated value for this multiplier is 2.86, and it is still only 2.98 when $L=10^{80}$. For $L\leq 10^6$, however, the multiplier is less than 2.16, and for values $L\leq 10$, it is at most 2. Thus even in this case, where the posterior expectation exists for $k=\infty$, it can still be important to use a realistic value for L. Although this case can be described as a genuine long-tailed distribution, in order for the posterior expectation of X to be finite when $k=\infty$, it is necessary to take $\delta>1$, and so the *a priori* expectation for α must be larger than $1+1/\beta$.

Table 4. Log-normal prior, LB = 1.001, UB = 10, $\gamma = 1$, $\mu = -1.19$, $\sigma = 1$, prior mean = 1.50, SD = .61

DATA		POST		PRED		DIST			BOUND
r	t	$E^*(\alpha)$	$SD^*(\alpha)$	$E^*(X)$	$SD^*(X)$	2	5	10	L
1	1	1.39	.38	1.11	.07	.39	.12	.05	1.25
1	1	1.39	.38	1.21	.14	.39	.12	.05	1.50
1	1	1.39	.38	1.37	.28	.39	.12	.05	2
1	1	1.39	.38	1.59	.51	.39	.12	.05	3
1	1	1.39	.38	1.87	.90	.39	.12	.05	5
1	1	1.39	.38	2.24	1.62	.39	.12	.05	10
1	1	1.39	.38	3.26	6.09	.39	.12	.05	100
1	1	1.39	.38	4.48	46.61	.39	.12	.05	10^4
1	1	1.39	.38	5.15	353.95	.39	.12	.05	10^6
3	2	1.41	.38	1.11	.07	.39	.12	.05	1.25
3	2	1.41	.38	1.21	.14	.39	.12	.05	1.50
3	2	1.41	.38	1.36	.27	.39	.12	.05	2
3	2	1.41	.38	1.59	.51	.39	.12	.05	3
3	2	1.41	.38	1.86	.90	.39	.12	.05	5
3	2	1.41	.38	2.22	1.60	.39	.12	.05	10
3	2	1.41	.38	3.18	5.92	.39	.12	.05	100
3	2	1.41	.38	4.30	44.35	.39	.12	.05	10^4
3	2	1.41	.38	4.90	332.81	.39	.12	.05	10^6
2	3	1.27	.23	1.11	.07	.42	.14	.06	1.25
2	3	1.27	.23	1.21	.14	.42	.14	.06	1.5
2	3	1.27	.23	1.37	.28	.42	.14	.06	2
2	3	1.27	.23	1.61	.52	.42	.14	.06	3
2	3	1.27	.23	1.91	.92	.42	.14	.06	5
2	3	1.27	.23	2.31	1.67	.42	.14	.06	10
2	3	1.27	.23	3.50	6.57	.42	.14	.06	100
2	3	1.27	.23	5.02	52.43	.42	.14	.06	10^4
2	3	1.27	.23	5.02	52.43	.42	.14	.06	10^6
5	1	2.34	1.17	1.11	.07	.25	.06	.02	1.25
5	1	2.34	1.17	1.20	.14	.25	.06	.02	1.50
5	1	2.34	1.17	1.32	.26	.25	.06	.02	2
5	1	2.34	1.17	1.48	.47	.25	.06	.02	3
5	1	2.34	1.17	1.65	.77	.25	.06	.02	5
5	1	2.34	1.17	1.83	1.26	.25	.06	.02	10
5	1	2.34	1.17	2.22	3.82	.25	.06	.02	100
5	1	2.34	1.17	2.58	24.18	.25	.06	.02	10^4
5	1	2.34	1.17	2.74	170.55	.25	.06	.02	10^6
1	5	1.18	.14	1.11	.07	.44	.15	.07	1.25
1	5	1.18	.14	1.21	.14	.44	.15	.07	1.50
1	5	1.18	.14	1.38	.28	.44	.15	.07	2
1	5	1.18	.14	1.62	.52	.44	.15	.07	3
1	5	1.18	.14	1.94	.94	.44	.15	.07	5
1	5	1.18	.14	2.38	1.73	.44	.15	.07	10
1	5	1.18	.14	3.77	7.11	.44	.15	.07	100
1	5	1.18	.14	5.73	60.45	.44	.15	.07	10^4
1	5	1.18	.14	6.95	483.26	.44	.15	.07	10^6
30	20	1.40	.22	1.11	.07	.38	.11	.04	1.25
30	20	1.40	.22	1.21	.14	.38	.11	.04	1.50
30	20	1.40	.22	1.36	.27	.38	.11	.04	2
30	20	1.40	.22	1.58	.51	.38	.11	.04	3
30	20	1.40	.22	1.86	.90	.38	.11	.04	5
30	20	1.40	.22	2.21	1.58	.38	.11	.04	10
30	20	1.40	.22	3.09	5.58	.38	.11	.04	100
30	20	1.40	.22	3.92	36.97	.38	.11	.04	10^4
30	20	1.40	.22	4.28	252.19	.38	.11	.04	10^6

Table 4. Log-normal prior, LB = 1.001, UB = 10, $\gamma = 1$, $\mu = -1.19$, $\sigma = 1$, prior mean = 1.50, SD = .61 – Continued

DATA		POST		PRED		DIST			BOUND
r	t	$E^*(\alpha)$	$SD^*(\alpha)$	$E^*(X)$	$SD^*(X)$	2	5	10	L
20	30	1.10	.07	1.11	.07	.47	.17	.08	1.25
20	30	1.10	.07	1.21	.14	.47	.17	.08	1.50
20	30	1.10	.07	1.38	.28	.47	.17	.08	2
20	30	1.10	.07	1.63	.53	.47	.17	.08	3
20	30	1.10	.07	1.97	.96	.47	.17	.08	5
20	30	1.10	.07	2.46	1.78	.47	.17	.08	10
20	30	1.10	.07	4.09	7.76	.47	.17	.08	100
20	30	1.10	.07	6.75	71.88	.47	.17	.08	10^4
20	30	1.10	.07	8.61	607.81	.47	.17	.08	10^6

A case of substantial practical importance is that in which the *a priori* information about α is weak, apart from the knowledge that $1 < \alpha \leq 2$. There is substantial empirical data on incomes, stock-market prices, city sizes, the distribution of biological genera and species, and many other variables, for which $\alpha \leq 2$. See Yule [15] and Zipf [14]. However, there is no known theoretical reason for taking the *a priori* distribution of α to be of the Gamma form, or for taking $\delta > 1$. In the case of weak *a priori* information, the likelihood function is approximately proportional to the posterior density for α. See the stable estimation argument of Savage [33] and Edwards, Lindman and Savage [34]. For either classical statisticians, to whom the *a priori* distribution is non-existent or "unknown," or to Bayesians who prefer to use some form of "uninformative" prior distribution, the results of Table 2 should be quite reassuring. It is possible, despite the delicacy at ∞ to obtain robust answers. It may be noted in this table that typically the posterior predictive expectation of X_{n+1}, given that it is between $x_{(1)}$ and $10 \times x_{(1)}$, is some modest multiple of the largest observation, at most $3 \times x_{(1)}$; and it is at most $5 \times x_{(1)}$ when $L \leq 100$. This is as it should be. One does not, for example, anticipate wind strengths that are some enormous factor times the largest yet experienced, even given that we set a new record wind strength. By comparing Table 1 for $\alpha = 1.50$ known, with Table 2 for the case $r = 3, t = 2$, we see that there is little sensitivity in either the predictive moments or the predictive probabilities. For example, when $L = 5$, Table 1 gives predictive moments of 1.82 and .87, and predictive probabilities of .35, .09, and .03; while Table 2 gives predictive moments of 1.83 and .88, and predictive probabilities of .36, .10, and .04. The greatest discrepancies occur for very large values of L, such as 10^6, which are inappropriate for most real-world applications.

Another case of substantial interest is that in which α is uniform from 0 to 2, so that even more extreme long-tailed behavior is possible. Again results are not very sensitive to the choice of *a priori* distribution, provided that L is not too large. For example, Table 3 with $r = 3, t = 2, L = 5$, gives the predictive moments as 1.90 and .93, and the predictive probabilities as .42, .15, and .08. Although there is a real change from the results of Tables 1 and 2, it is of limited extent, and is in the direction of making the predictive distribution longer-tailed, as was to be expected. If anything, one might be surprised that allowing α to get close to 0, as with this *a priori* distribution, did not move the predictive distribution much further to the right.

The final case of great interest is where some definite *a priori* information is input, as we do here with the log-normal distribution. Table 4, for the case $\gamma = 1, r = 3, t = 2, L = 5$, gives 1.86 and .90 as predictive moments, and .39, .12, and .05, as predictive probabilities. These results are close to those of Table 2, in which α has the same *a priori* expectation as in Table 4.

The reader may compare these various tables for other values of the parameters, to examine the effect of long-tailed sample data, greater sample sizes, cases where the *a priori* information is less concordant with the data, and the effect of L. For example, in Table 3 with $r = 2, t = 3$, so that $\hat{\alpha} = .67$, and $L = 5$, the predictive moments are 2.07 and 1.01, while the predictive probabilities are .56, .29, and .19. Again, provided that a realistic upper bound for L is chosen, such as 10, the changes from previous values are real but of limited magnitude, and in the direction to be anticipated.

Armed with this information, let us now examine real-world data on city sizes. Table 5 gives the sizes of the 30 largest cities in the United States in 1940 and 1988. They are first presented in descending

Table 5. City size $\times 10^{-3}$ data

1940		1988	
Ordered	Permuted	Ordered	Permuted
7455	1931	7353	987
3397	859	3353	727
1931	302.3	2978	532
1623	305	1698	1647
1504	3397	1647	522
878	368	1070	599
859	816	1036	2978
816	587	987	465
771	399	941	1036
672	1623	924	502
663	456	751	434
635	387	738	511
587	771	732	1070
576	635	727	7353
495	492	645	481
492	301.2	635	732
456	495	617	941
430	663	599	3353
399	306	578	578
387	878	570	570
385	7455	532	1698
368	325	521	617
325	322	511	924
322	319	502	738
319	576	492	645
306	302.2	481	751
305	672	465	492
302.3	430	439	635
302.2	1504	434	427
301.2	385	427	439

order, and then in a randomly chosen permutation. The data for 1940 was previously analysed in [1] to illustrate use of the tail-index method. The upper tail of such city size data is generally regarded as being modelled by Zipf's law, with some tail-index α. Tables 6 and 7 give the running forecasts, and their standard deviations, for the next observation, based upon the permutation. We imagine, in other words, that a random sample has been taken from the population, and that we successively forecast the magnitude of each upcoming record value. In this way we simulate the actual forecasting of future record values based upon a random sample from a population. It is well known that sampling (with or without replacement) from a finite population generates an exchangeable sequence. Because our forecast of the magnitude of the next record value depends only upon the upper order statistics of the past data, and not directly upon how many past values have been observed, we put forth the same expectation for the magnitude of the next record value, until we observe a new record value.

The record values (with the first value taken as a record value by default) for Table 6 occurred at times 1, 5, 21, and had the values 1931, 3397, 7455, respectively. Table 6 gives the 1940 forecasts for $L=3,5,10$, where each forecast is based upon all the past data up to the time of the forecast, and uses only the current upper two order statistics of the data, so $r=1$. The column labelled $\hat{\alpha}$ gives the current maximum-likelihood estimate of α based upon the two upper order statistics, so $t=\frac{1}{\alpha}$. The first row of Table 6 would be read as follows. Based upon the two largest order statistics (1931, 859) at time 2 in the 1940 permuted sequence, the estimate of α is 1.235. This data (with $r=1$ and $t=.810$) is used to obtain the posterior distribution for α, for a uniform *a priori* distribution on the interval from 0 to 2. Forecasts and standard deviations are then presented for $L=3,5,10$. For example the $L=3$ forecast of the next record value

Table 6. Forecast of 1940 city sizes $\times 10^{-3}$

City size	$\hat{\alpha}$	Forecast			Forecast SD		
		3	5	10	3	5	10
3397	1.235	3146	3810	4831	1023	1869	3596
7455	1.770	5500	6621	8295	1787	3244	6166
(?)	1.272	12137	14694	18608	3944	7209	13844

Table 7. Forecast of 1988 city sizes $\times 10^{-3}$

City size	$\hat{\alpha}$	Forecast			Forecast SD		
		3	5	10	3	5	10
1647	3.271	1588	1899	2351	516	929	1743
2978	1.953	2663	3202	4001	865	1568	2973
7353	1.688	4824	5810	7290	1569	2847	5420
(?)	1.106	12007	14581	18574	3904	7154	13824

is 3146 with a standard deviation of 1023, this forecast being made using only the previous records of 1931 and 859. The realized value turned out to be 3397. Note that most of the actual values are well within 1 standard deviation of the forecast. The row '?' forecasts a next record value, based upon all the past data, as though the population were not complete, and is given only for illustrative purposes. Table 7 repeats the analysis for the 1988 city size data. The record values occurred at trials 1, 4, 7, 14, and had the values 987, 1647, 2978, 7353, respectively.

This type of forecasting problem, based upon a random sample from a fixed population, is used to illustrate the procedure in connection with an exchangeable sequence of observations. As shown by de Finetti, and discussed in [35], one can always represent real-world exchangeable sequences in terms of limits arising in sampling from a finite population. The exchangeable case is the simplest scenario in which our methods can be usefully applied. More generally, one must deal with evolutionary processes, as for example when successive records are set over time. For example, if we consider the successive Olympic High Jump records, since 1880, we must keep in mind that we are not sampling from a fixed population, and that changes in technique and general level of physical fitness over time, may have a substantial effect. Similarly, in considering the next record value of some stock market index, such as the Dow Jones, there may be time trends that must be taken into account. However, even in such examples as these, local exchangeability over sufficiently short time periods may be a reasonable assumption, and appropriate modification of the basic forecasting procedure proposed in this article can be developed.

6. Conclusions

We believe that the above studies indicate that it is possible to make effective inference and predictions about record values. Our methodology can be used both with uniform *a priori* distributions, such as represented in Tables 2 and 3, and with more informative *a priori* distributions such as in Table 4. The case that is perhaps of greatest interest for applications is that of the three-parameter log-normal distribution with threshold taken to be 1 or 0, as may seem appropriate. Uniform *a priori* distributions can, for practical purposes, be represented as special cases of such log-normal distributions. We believe that it is important to study sensitivity of results to choice of *a priori* distribution, as recommended in [36,30]. The choice of r and of L can be implemented by Bayesian data-analytic techniques, such as described in [1,25]. Here in our forecast of city sizes we took $r = 1$, but substantial improvements could result from a Bayesian decision-theoretic choice of r.

There are some basic issues concerning the use of finite models, versus infinite idealized models, that are especially pertinent in connection with the problem of prediction for long-tailed distributions. If one took the conventional idealized model literally in our example, then the analysis of Secs. 1 and 2 demonstrates that there are some logical difficulties, if one also views the observations as unbounded. For in the case of greatest interest, where it is known that $1 \leq \alpha \leq 2$, the posterior first moment may be infinite, even though it is plainly unreasonable to make a prediction of more than a few multiples of the largest observation yet seen. The issue is resolved here by treating the algebraic model for the tail as only an approximation, valid in some finite domain. In this case the algebraic tail is compatible with both the data, and with putting forth sensible predictions for squared error loss. See [24] for discussion of the finite/infinite question in connection with Steinian shrinkage estimators.

The issue regarding infinite predictive moments thus turns out to be largely irrelevant for forecasting, provided that one is comfortable with using some reasonable upper bound for the observable variables. Careless use of infinite models, ignoring the fact that realistic finite upper bounds are usually available, might instead have led one to the conclusion that theory-based forecasting is impossible in the case $\alpha \leq 2$. Since all statistical analyses must eventually be done on a computer with finite memory, such infinite models are at best only useful guides, and their careless use can lead to numerous apparent paradoxes, which have no real-world importance. The primary conclusion of this article is that provided that a finite upper bound for the observations can be supplied, as is ordinarily the case, it is possible to make effective predictions of future record values. The forecasts that we have obtained, employing such finite upper bounds, are by no means perfect, but they do at least put one in the right ballpark, with predictions that are at most a few multiples of the previous record value. I am not aware of other methods available at present that do so.

Forecasting is always difficult, and perhaps even more so for the case of record values in the case of long-tailed distributions. Nonetheless, often such forecasts are important in the decision-making process, and must somehow or other be put forth. We have suggested a Bayesian methodology which can make systematic use both of *a priori* information and of the current data. When used with care, we believe these methods can be of value in a variety of areas.

Acknowledgment

This work was supported by the National Science Foundation under grant DMS-9201056.

7. References

[1] B. M. Hill, A simple general approach to inference about the tail of a distribution, Ann. Stat. **3**, 1163–1174 (1975).
[2] W. Feller, An Introduction to Probability Theory and Its Applications, Volume 2, John Wiley and Sons, Inc., New York (1971) p. 15.
[3] B. M. Hill, D. Lane, and W. Sudderth, A strong law for some generalized urn processes, Ann. Prob. **8**, 214–226 (1980).
[4] B. M. Hill, D. Lane, and W. Sudderth, Exchangeable urn processes, Ann. Prob. **15**, 1586–1592 (1987).
[5] B. M. Hill, Parametric models for A_n: splitting processes and mixtures, J. R. Statist. Soc. B **55**, 423–433 (1993).
[6] B. M. Hill, Bayesian forecasting of economic time series, to appear in Econometric Theory, 1994.
[7] B. M. Hill, Some subjective Bayesian considerations in the selection of models (with discussion), Econometric Revs. **4** (2), 191–288 (1985–86).
[8] D. J. Poirier, A Bayesian view of nominal money and real output through a new classical macroeconomic window (with discussion), J. Bus. Econ. Stat. **7**, 125–161 (1991).
[9] N. D. Singpurwalla and R. J. Meinhold, Robustification of Kalman Filter Models, J. Amer. Statist. Assoc. **84**, 479–486 (1989).
[10] B. M. Hill, Tail Probabilities, in Encyclopedia of Statistical Sciences Vol. 9, S. Kotz and N. Johnson, eds., John Wiley and Sons, Inc., New York (1988).
[11] S. Csörgö, P. Deheuvels, and D. M. Mason, Kernel estimates of the tail index of a distribution, Ann. Stat. **13**, 1050–1077 (1985).
[12] J. Aitchison and I. R. Dunsmore, Statistical Prediction Analysis, Cambridge University Press, Cambridge (1975).
[13] X. Maret, Estimating the Structure of Stochastic Dynamic Linear Systems: A Bayesian Approach and Economic Applications, Doctoral dissertation, The University of Michigan (1988).
[14] G. K. Zipf, Human Behavior and the Principle of Least Effort, Addison-Wesley Publishing Co. (1949).
[15] G. U. Yule, A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, F. R. S., Phil. Trans. B **213**, 21–87 (1924).
[16] B. M. Hill, Zipf's Law and prior distributions for the composition of a population, J. Amer. Statist. Assoc. **65**, 1220–1232 (1970).
[17] B. M. Hill, The rank frequency form of Zipf's Law, J. Amer. Statist. Assoc. **69**, 1017–1026 (1974).
[18] B. M. Hill, A theoretical derivation of the Zipf (Pareto) Law, in Studies on Zipf's Laws, H. Guiter, ed., Sprachwissenschaftliches Inst., Ruhr-Universitat Bochun (1981).
[19] B. M. Hill and M. Woodroofe, Stronger forms of Zipf's law, J. Amer. Statist. Assoc. **70**, 212–219 (1975).
[20] M. Woodroofe and B. M. Hill, On Zipf's Law, J. Applied Prob. **12**, 425–434 (1975).
[21] N. L. Johnson and S. Kotz, Urn Models and Their Application, John Wiley and Sons, Inc., New York (1977) p. 350.
[22] B. M. Hill, Posterior distribution of percentiles: Bayes' theorem for sampling from a population, J. Amer. Statist. Assoc. **63**, 677–691 (1968).
[23] S. Chatterjee and M. R. Yilmaz, Chaos, fractals and statistics, Statistical Sci. **7**, 49–68 (1992).
[24] B. M. Hill, On Steinian shrinkage estimators: The finite/infinite problem and formalism in probability and statistics, to appear.
[25] B. M. Hill, A theory of Bayesian data analysis, in Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard, S. Geisser, J. S. Hodges, S. J. Press, and A. Zellner, eds., North-Holland (1990) pp. 49–73.
[26] R. A. Fisher, Conclusions Fiduciare, Annales de l'Institut Henri Poincaré **10**, 191–213 (1948).
[27] B. M. Hill, De Finetti's theorem, induction, and A_n or Bayesian non-parametric predictive inference (with discussion), in Bayesian Statistics 3, J. Bernardo, M. DeGroot, D. V. Lindley, and A. F. M. Smith, eds., Oxford University Press (1988) pp. 211–241.
[28] B. de Finetti, La prévision: ses lois logiques, ses sources subjectives, Annales de l'Institut Henri Poincaré **7**, 1–68 (1937).
[29] B. de Finetti, Theory of Probability, Vol. 1, John Wiley and Sons, Inc., London (1974).
[30] B. M. Hill, Statistics, Robustness, in Encyclopedia of Physical Science and Technology Vol 15, Academic Press, Inc. (1992).
[31] J. Aitchison and J. A. C. Brown, The Lognormal Distribution, Cambridge University Press, Cambridge (1957).
[32] B. M. Hill, The three-parameter log-normal distribution and Bayesian analysis of a point-source epidemic, J. Amer. Statist. Assoc. **58**, 72–84 (1962).
[33] L. J. Savage, The Foundations of Statistical Inference, A Discussion, Methuen and Co. Ltd. (1962).
[34] W. Edwards, H. Lindman, and L. J. Savage, Bayesian statistical inference for psychological research, Psychological Review **70**, 193–242 (1963). Reprinted in Robustness of Bayesian Analysis, J. Kadane, ed., North-Holland: Amsterdam, 1–62 (with discussion) (1984).
[35] B. M. Hill, Bayesian nonparametric prediction and statistical inference (with discussion), in Bayesian Analysis in Statistics and Econometrics, P. K. Goel and N. S. Iyengar, eds., Lecture Notes in Statistics Series, Springer-Verlag (1992) pp. 43–94.
[36] B. M. Hill, Robust analysis of the random model and weighted least squares regression, in Evaluation of Econometric Models, J. Kmenta and J. Ramsey, eds., Academic Press (1980) pp. 197–217.

About the author: Bruce M. Hill is a statistician at the University of Michigan. His interest is Bayesian statistics.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 539 (1994)]

On the Convergence of the Number of Exceedances of Nonstationary Normal Sequences

Volume 99 Number 4 July–August 1994

J. Hüsler

University of Bern,
Bern, Switzerland

and

M. Kratz

University Paris VI,
Paris, France

It is known that the number of exceedances of normal sequences is asymptotically a Poisson random variable, under certain restrictions. We analyze the rate of convergence to the Poisson limit and extend the result known in the stationary case to nonstationary normal sequences by using the Stein-Chen method. In addition, we consider the cases of exceedances of a constant level as well as of a particular nonconstant level.

Key words: exceedances; nonstationary; normal sequences; rate of convergence; Stein-Chen method.

Accepted: March 22, 1994

1. Introduction and Result

The extreme value theory of Gaussian sequences has interested many authors, for instance Refs. [1,4,7,8], dealing with the limit distribution of the suitably normalized extreme value.

Let $\{X_i, i \geq 1\}$ be a standardized normal sequence with correlations $E(X_iX_j) = r_{ij}$, $i,j \geq 1$, and $\Phi(\cdot)$ the distribution function of X_i.

Let

$$N_n = \sum_{i=1}^{n} 1(X_i > u_{ni})$$

denote the number of exceedances of a boundary given by a triangular array $\{u_{ni}, i \leq n, n \geq 1\}$. Then it was also found that N_n converges in distribution to a random variable having a Poisson distribution $\mathbb{P}(\lambda_n)$, if the mean number of exceedances $\lambda_n = \Sigma_{i \leq n}(1 - \Phi(u_{ni}))$ remains bounded (cf. Ref. [4]).

For practical use of the asymptotic theory, it is rather important to know the rate of convergence or at least some upper bound for this rate.

For the stationary case, results on the rate of convergence have been obtained for instance by Refs. [2,3,9,10]. The aim of this paper is to give an upper bound for the total variation distance d_{tv} between N_n and $\mathbb{P}(\lambda_n)$, in the nonstationary case, extending the results of these mentioned papers.

Suppose that for some sequence ρ_n: $|r_{ij}| \leq \rho_{|i-j|}$ for $i \neq j$, and that the two conditions

$$\rho_n < 1 \text{ for all } n \geq 1 \tag{1}$$

$$\rho_k \leq A/\log k, k \geq 2, \text{ for some constant } A \tag{2}$$

are satisfied. Define ρ as $\rho = \max(0, r_{ij}, i \neq j) < 1$.

In addition, we assume that the boundary values tend uniformly to ∞:

$$u_{n,\min} = \min_{1 \leq i \leq n} u_{ni} \to \infty \text{ as } n \to \infty. \tag{3}$$

The exceedances of a constant boundary $u_{ni} = u_n$, $1 \leq i \leq n$, are considered first, where only the tools given in Ref. [3] are used. We show in the second result that the method of Ref. [3] can be used also for nonconstant boundaries $\{u_{ni}\}$. But these boundaries are restricted such that the condition

$$\limsup_{n\to\infty} n(1-\Phi(u_{n,\min})) < C < \infty \tag{4}$$

holds. If we want to extend the results to a more general class of boundaries such that only

$$\limsup_{n\to\infty} \lambda_n < C < \infty \tag{5}$$

holds, we need to combine the method developed by Ref. [4] with that of Ref. [3] to get satisfactory results (see Ref. [6]).

Our first result for boundaries which are constant for fixed n, shows that the given upper bound of the rate of convergence depends mainly on the largest positive correlation value ρ.

Theorem 1: *Let $\{X_i, i \geq 1\}$ be a standardized nonstationary normal sequence with correlations $\{r_{ij}, i,j \geq 1\}$. Suppose that $|r_{ij}| \leq \rho_{|i-j|}$ for $i \neq j$, such that Eqs. (1) and (2) hold. Let the boundary values $\{u_{ni} = u_n, 1 \leq i \leq n\}$ and λ_n be real values with $\lambda_n = n(1-\Phi(u_n))$. Suppose that $\lambda_n \leq C < \infty$, for some constant C. Then as $n \to \infty$*

$$d_{tv}(N_n, \mathbb{P}(\lambda_n)) = O\Big(n^{-\frac{1-\rho}{1+\rho}} \cdot (\log n)^{-\frac{\rho}{1+\rho}} + \frac{\log n}{n^2}\sum_{k=1}^{n-1}(n-k)\rho_k\Big). \tag{6}$$

This extends the result of Ref. [3] showing that for a constant boundary their upper bound of the rate of convergence in the stationary case holds also in the nonstationary case.

Theorem 2: *Let $\{X_i, i \geq 1\}$ be a standardized nonstationary normal sequence with correlations $\{r_{ij}, i,j \geq 1\}$ as in Theorem 1 satisfying Eqs. (1) and (2). Suppose that the boundary values $\{u_{ni}, 1 \leq i \leq n, n \geq 1\}$ are such that Eq. (4) holds. Then Eq. (6) holds.*

The first term of Eq. (6) dominates the rate of convergence in cases with $\Sigma_{k\geq 1}\rho_k < \infty$ and $\rho > 0$.

Then the rate of convergence depends only on the lowest value $u_{n,\min}$ of the particular boundary u_{ni} and also on the largest positive correlation ρ. It extends naturally the results of the stationary case with boundary values which are constant for fixed n. This rate is only good if $u_{n,\min}$ is not a uniquely low value, which is supposed for reasonable boundaries. For the case $u_{n,\min}$ is uniquely low, the rate can be improved.

2. Proof

The proof of Theorem 1 is an adaption of that used by Ref. [3] in the stationary case. We use the following lemma which is a straightforward extended version of Lemma 3.4 of Ref. [3].

Lemma 1: *Suppose that*

$$(X_i, X_j) \overset{d}{=} N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & r_{ij}\\ r_{ij} & 1\end{pmatrix}\right).$$

Define $Z_i = 1(X_i > u_{ni})$ for some boundary values u_{ni} where

$$1 - \Phi(u_{ni}) \leq C/n$$

for some finite constant C.
Then for some constant K depending on C only and for all $n \geq 2$:

i) If $0 \leq r_{ij} < 1$, then

$$0 \leq cov(Z_i, Z_j) \leq K \cdot \frac{1}{\sqrt{1-r_{ij}}} n^{-2+\frac{2r_{ij}}{1+r_{ij}}} (\log n)^{-\frac{r_{ij}}{1+r_{ij}}}$$

$$\leq K n^{-2}\left(\frac{n^2}{\log n}\right)^{\frac{\rho_{|i-j|}}{1+\rho_{|i-j|}}}$$

ii) If $0 \leq r_{ij} \leq 1$, then

$$0 \leq cov(Z_i, Z_j) \leq K \frac{r_{ij}\log n}{n^2} e^{2r_{ij}\log n}$$

$$\leq K \frac{\rho_{|i-j|}\log n}{n^2} e^{2\rho_{|i-j|}\log n}$$

iii) If $-1 < r_{ij} \leq 0$, then

$$0 \geq cov(Z_i, Z_j) \geq -K \frac{|r_{ij}|\log n}{n^2} \geq -K \frac{\rho_{|i-j|}\log n}{n^2}$$

iv) If $-1 \leq r_{ij} \leq 0$, then

$$0 \geq cov(Z_i, Z_j) \geq -K \frac{1}{n^2}$$

We need for Theorem 2 an extension of Lemma 1.

Lemma 2: *Suppose that* (X_i,X_j) *and* Z_i *are as in Lemma 1. For any* i,j, *define* $u_{nij}=\min(u_{ni},u_{nj})$ *and* $v_{nij}=\max(u_{ni},u_{nj})$.

Then for some constant K depending only on C and for all $n \geq 2$:

i) If $0 \leq r_{ij} \leq 1$, *then*

$$0 \leq cov(Z_i,Z_j) \leq K \frac{1}{\sqrt{1-r_{ij}}} \left(\frac{\varphi(u_{nij})}{u_{nij}}\right)^{\frac{2}{1+r_{ij}}} \cdot u_{nij}^{-\frac{2r_{ij}}{1+r_{ij}}}$$

with $K \geq (2\pi)^{-\frac{r_{ij}}{1+r_{ij}}} \cdot (1+r_{ij})^{3/2}$.

ii) If $0 \leq r_{ij} \leq 1$, *then*

$$0 \leq cov(Z_i,Z_j) \leq K r_{ij}\, \varphi(u_{nij}) \left(\varphi(v_{nij})\right)^{\frac{1-r_{ij}}{1+r_{ij}}}$$

iii) If $-1 \leq r_{ij} \leq 0$, *then*

$$0 \geq cov(Z_i,Z_j) \geq -(1-\Phi(u_{nij}))(1-\Phi(v_{nij}))$$

$$\geq -(1-\Phi(u_{nij}))^2$$

iv) If $-1 \leq r_{ij} \leq 0$, *then*

$$0 \leq |cov(Z_i,Z_j)| \leq K \min(1, |r_{ij}| u_{nij} v_{nij} + r_{ij}^2 v_{nij}^2)$$

$$(1-\Phi(u_{nij})) \cdot (1-\Phi(v_{nij}))$$

(The proof of this lemma is given in [6]).

Theorem 1 follows also by Theorem 2. Therefore, we prove now Theorem 2 by using the method of Ref. [3].

By Theorem 3.1 of Ref. [3], we have

$$d_{tv}(N_n, \mathbb{P}(\lambda_n)) \leq \frac{1-e^{-\lambda_n}}{\lambda_n} \left(\frac{\lambda_n^2}{n} + \sum_{i \neq j} |cov(Z_i,Z_j)|\right). \quad (7)$$

If $\rho=0$, i.e. if $r_{ij} \leq 0$ for $i \neq j$, then using Lemma 2(iv) for the second term of Eq. (7), we get the second term of Eq. (6) which dominates the first one in this case, if $\rho_k > 0$ for some k. Obviously, if $\rho_k = 0$ for all k, then the result holds, since the second term in Eq. (7) is 0.

Thus suppose from now on that $\rho > 0$.

Because of Eq. (2) the sum

$$S_n = \sum_{1 \leq i < j \leq n} |cov(Z_i,Z_j)| = \sum_{1 \leq i < j \leq n} c_{ij}$$

is split up into three parts, by using $\delta > 0$ such that

$$3\delta < \frac{\rho}{1+\rho}$$

There are only finitely many k's with $\rho_k > \delta$. This will be treated first as Case (i). For indices k with $\rho_k \leq \delta$, we distinguish Case (ii) with $k < n^{\delta}$ and Case (iii) with $k \geq n^{\delta}$.

Case (i): Each term c_{ij} of the sum S_n is bounded above by

$$K n^{-2/(1+\rho)} (\log n)^{-\rho/(1+\rho)}$$

if $r_{ij} \geq 0$ by Lemma 2(i)
or bounded by

$$K n^{-2}$$

if $r_{ij} < 0$ by Lemma 2(iii).
Since there are finitely many k's with $\rho_k > \delta$, the number of terms c_{ij} with $|i-j|=k$ and $\rho_k > \delta$ is of the order $O(n)$. Hence the sum on these terms is bounded by

$$K n^{-(1-\rho)/(1+\rho)} (\log n)^{-\rho/(1+\rho)} + K n^{-1}.$$

Case (ii): There are at most n^{δ} terms ρ_k such that $0 \leq \rho_k \leq \delta$ and $k < n^{\delta}$. Hence there are at most $O(n^{1+\delta})$ terms c_{ij} with such a $k = |i-j|$. But each such term c_{ij} of S_n is bounded by

$$K n^{-2/(1+\delta)} (\log n)^{-\delta/(1+\delta)} \leq K n^{-2+2\delta}$$

if $r_{ij} \geq 0$ by Lemma 2(i)
or bounded by

$$K n^{-2}$$

if $r_{ij} < 0$ by Lemma 2(iv).
Then the sum of these terms c_{ij} is bounded by $O(n^{-1+3\delta})$.

Case (iii): Finally we consider the terms such that $\rho_k \leq \delta$, with $k \geq n^{\delta}$. Note that we have

$$0 \leq \rho_k \leq \frac{A}{\log k} \leq \frac{A}{\delta \log n}. \quad (8)$$

Each such term c_{ij} of S_n gives a contribution

$$K\,\rho_k\left(\frac{\sqrt{\log n}}{n}\right)^{1+\frac{1-\rho_k}{1+\rho_k}} \leq K\,\rho_k \frac{\log n}{n^2}$$

if $r_{ij} \geq 0$ by Lemma 2(ii) and by using Eq. (8)

$$K\,\rho_k\left(\frac{\log n}{n^2}\right)$$

if $r_{ij} < 0$ by Lemma 2(iv).
Taking now the sum on all terms (i,j) with $|i-j| \geq n^{\delta}$, we get the second term of Eq. (6).

Finally, adding up all these upper bounds of Cases (i), (ii), and (iii), the result Eq. (6) of Theorem 2 follows.

3. References

[1] S. M. Berman, Limit theorems for the maximum term in stationary sequences, Ann. Math. Statist. **33**, 502–516 (1964).
[2] P. Hall, The rate of convergence of normal extremes, J. Appl. Probab. **16**, 433–439 (1979).
[3] L. Holst and S. Janson, Poisson approximation using the Stein-Chen method and coupling: number of exceedances of Gaussian random variables, Ann. Probab. **18**, 713–723 (1990).
[4] J. Hüsler, Asymptotic approximations of crossing probabilities of random sequences, Z. Wahrscheinlichkeitstheorie verw. Geb. **63**, 257–270 (1983).
[5] J. Hüsler, Extreme values of non-stationary random sequences, J. Appl. Probab. **23**, 937–950 (1986).
[6] J. Hüsler and M. F. Kratz, Rate of convergence of the number of exceedances of nonstationary normal sequences. Preprint (1993).
[7] M. R. Leadbetter, G. Lindgren, and H. Rootzén, Extremes and related properties of random sequences and processes, Springer Series in Statistics, Springer, New York (1983).
[8] V. I. Piterbarg, Asymptotic expansions for the probability of large excursions of Gaussian processes, Soviet Math. Dokl. **19**, 1279–1283 (1978).
[9] H. Rootzén, The rate of convergence of extremes of stationary random sequences, Adv. Appl. Probab. **15**, 54–80 (1983).
[10] R. Smith, Extreme value theory for dependent sequences via the Stein-Chen method of Poisson approximation, Stoch. Process. Appl. **30**, 317–327 (1988).

About the authors: *Jürg Hüsler is Professor at the Department of Mathematical Statistics of the University of Bern. Marie Kratz is now Maître de Conférences at the Department of Mathematics and Computer Science of the University René Descartes at Paris.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 543 (1994)]

On the Multivariate Extremal Index

Volume 99 Number 4 July–August 1994

S. Nandagopalan

Colorado State University,
Fort Collins, CO 80523

The exceedance point process approach of Hsing et al. is extended to multivariate stationary sequences and some weak convergence results are obtained. It is well known that under general mixing assumptions, high level exceedances typically have a limiting Compound Poisson structure where multiple events are caused by the clustering of exceedances. In this paper we explore (a) the precise effect of such clustering on the limit, and (b) the relationship between point process convergence and the limiting behavior of maxima. Following this, the notion of multivariate extremal index is introduced which is shown to have properties analogous to its univariate counterpart. Two examples of bivariate moving average sequences are presented for which the extremal index is calculated in some special cases.

Key words: dependence function; exceedance; extremal index; multivariate; point process; stationary.

Accepted: March 22, 1994

1. Introduction

Extreme value theory for multivariate iid sequences has been studied for quite some time now but attention to the dependent case has been relatively recent. For univariate sequences it is known that local dependence causes extreme values to occur in clusters, which in turn results in a stochastically smaller distribution for the maximum than if the observations were independent. We begin with a brief review of these results, which we shall later extend to the multivariate case.

Let $\{\xi_n\}$ be a univariate stationary sequence. Write $M_n=\max\{\xi_1,\ldots,\xi_n\}$ and for $\tau>0$, let $\{u_n(\tau)\}$ denote a sequence satisfying $\lim_{n\to\infty} nP\{\xi_1>u_n(\tau)\}=\tau$. Under quite general mixing assumptions there exist constants $0\le\theta'\le\theta''\le 1$ such that

$$\limsup_{n\to\infty} P\{M_n\le u_n(\tau)\}=e^{-\theta''\tau} \text{ and}$$

$$\liminf_{n\to\infty} P\{M_n\le u_n(\tau)\}=e^{-\theta'\tau}$$

for all τ. (See Ref. [1], although the idea actually dates back to Refs. [2–4].) Thus if $P\{M_n\le u_n(\tau_0)\}$ converges for some τ_0, then $\theta'=\theta''(=\theta$, say) and hence $\lim_{n\to\infty} P\{M_n\le u_n(\tau)\}=e^{-\theta\tau}$ for all $\tau>0$. The common value θ is then called the *extremal index* of $\{\xi_n\}$. We shall assume θ to be positive whenever it exists, since the case $\theta=0$ corresponds to a degenerate limiting distribution for M_n. Note that $\theta=1$ for iid sequences. Let $\{\hat{\xi}_n\}$ be an iid sequence with $\hat{\xi}_1=^d \xi_1$, called the *associated iid sequence*, and write $\hat{M}_n=\max\{\hat{\xi}_1,\ldots,\hat{\xi}_n\}$. If $\{\xi_n\}$ has extremal index θ and $\lim_{n\to\infty} P\{M_n\le v_n(t)\}=H(t)$ for a suitable family of normalizing constants $\{v_n(t)\}$, then it follows (upon identifying $e^{-\theta\tau}$ with $H(t)$) that $\lim_{n\to\infty} P\{\hat{M}_n\le v_n(t)\}=\hat{H}(t)$ where

$$H(t)=\hat{H}(t)^{\theta}. \tag{1}$$

The extremal index is thus a measure of the effect of

dependence on the limiting distribution of M_n. The stochastically smaller limiting distribution of M_n is in fact a direct result of the clustering of extremes, as explained below. See Ref. [5] for details.

For fixed $\tau>0$ let the *exceedance point process* $N_n=N_n^{(\tau)}$ be defined by

$$N_n(B)=\sum_{i=1}^{n} I_{\{\xi_i>u_n(\tau)\}}I_{\{i/n\in B\}}, \quad B\subset[0,1],$$

where I_A denotes the indicator function of the event A. Then for a broad class of weakly dependent sequences, the limit in distribution of N_n, if it exists, is a Compound Poisson process with intensity $\theta\tau$ and multiplicity distribution π on $\{1,2,...\}$. The Poisson events may in fact be regarded as the positions of "exceedance clusters" while the multiplicities correspond to cluster sizes. More explicitly, one may divide the n observations into k_n blocks of roughly equal size and regard exceedances within each block as forming a single "cluster", so that the cluster sizes are given by $N_n(J_i)$, $i=1,...,k_n$, where $J_i(=J_{n,i})=(\frac{i-1}{k_n}, \frac{i}{k_n}]$. For a suitable choice of k_n depending on the mixing rate of $\{\xi_n\}$, one then has

$$\lim_{n\to\infty} P\{N_n(J_1)=j \mid N_n(J_1)>0\}=\pi(j), \quad j\geq 1,$$

and

$$\lim_{n\to\infty} P^{k_n}\{N_n(J_1)=0\}=\lim_{n\to\infty} P\{N_n[0,1]=0\}$$

$$=\lim_{n\to\infty} P\{M_n\leq u_n(\tau)\}=e^{-\theta\tau},$$

so that in particular, $\lim_{n\to\infty} k_n P\{N_n(J_1)>0\}=\theta\tau$. Hence

$$\lim_{n\to\infty} E\,N_n[0,1]=\lim_{n\to\infty} k_n E\,N_n(J_1)$$

$$=\lim_{n\to\infty} k_n E(N_n(J_1) \mid N_n(J_1)>0)P\{N_n(J_1)>0\}$$

$$=\theta\tau \lim_{n\to\infty} E(N_n(J_1) \mid N_n(J_1)>0),$$

while on the other hand, $\lim_{n\to\infty} E\,N_n[0,1]=\lim_{n\to\infty} nP\{\xi_1>u_n(\tau)\}=\tau$. The cluster size distribution and the extremal index are therefore related by

$$\lim_{n\to\infty} E(N_n(J_1) \mid N_n(J_1)>0)=1/\theta. \tag{2}$$

Now let $\{\xi_n=(\xi_{n1},...,\xi_{nd}), n\in\mathbb{Z}\}$ be a multivariate stationary sequence where $d\geq 1$ is a fixed integer, and write $M_n=(M_{n1},...,M_{nd})$ where $M_{nj}=\max\{\xi_{1j},...,\xi_{nj}\}$, $j=1,...,d$. The study of multivariate extremes began in the early 1950s, focusing mainly on the limiting behavior of M_n under a linear normalization, when the observations are iid. The resulting class of limiting distributions was characterized in Ref. [6] and domains of attraction criteria were given in Ref. [7]. See also Ref. [8], Chapter 5, for an account of the literature surrounding this theory. For stationary sequences satisfying a general mixing assumption, it is known (see Refs. [9, 10], and Theorem 1.1 below) that the class of limiting distributions of M_n is the same as for iid sequences. In this paper we explore the precise effect of dependence on the limiting distribution by extending the univariate theory described above to the multivariate case. Essentially, this involves studying the inter-relationship between the two dependence structures present, one due to dependence over time and the other due to the dependence between the various components of the multivariate observations. The ideas become most transparent when presented in terms of so-called dependence functions [8]. Here we adopt the slightly modified definition found in Ref. [9]. A distribution function D on $[0,1]^d$ is called a *dependence function* if $D_j(D_j(u))=D_j(u)$, $u\in[0,1]$, $j=1,...,d$, where the subscript j signifies the jth marginal. The dependence function of a distribution F on $\mathbb{R}^d$ is defined by $D_F(\boldsymbol{u})=P\{F_1(X_1)\leq u_1,...,F_d(X_d)\leq u_d\}$, $\boldsymbol{u}=(u_1,...,u_d)\in[0,1]^d$, where $(X_1,...,X_d)$ is a random vector with distribution F. More generally, any dependence function satisfying $F(x)=D(F_1(x_1),...,F_d(x_d))$ could be defined to be a dependence function of F, although the present choice is a natural one.

Write $T=(0,1)^d\setminus\{\mathbf{1}\}$ where $\mathbf{1}=(1,...,1)\in\mathbb{R}^d$, and for $\mathbf{t}=(t_1,...,t_d)\in T$, let $v_n(\mathbf{t})=(v_{n1}(t_1),...,v_{nd}(t_d))$ where $v_{nj}(t_j)$ satisfies $\lim_{n\to\infty} nP\{\xi_{1j}>v_{nj}(t_j)\}=-\log t_j$. Let H_n denote the distribution function of M_n (i.e., $H_n(x)=P\{M_n\leq x\}$), with marginals H_{nj}, $j=1,...,d$. Then (see Refs. [8, 11]),

$$H_n(v_n(\mathbf{t}))\to^w H(\mathbf{t}) \tag{3}$$

if and only if

$$H_{nj}(v_{nj}(t_j))\to^w H_j(t_j),\ j=1,...,d, \text{ and } D_{Hn}(\mathbf{t})\to^w D_H(\mathbf{t}).$$

The limiting behavior of M_n can therefore be separated into two parts, one pertaining to the convergence of the marginals (a univariate problem) and the other to the convergence of the dependence functions. Here we focus attention exclusively on the latter. It should be noted that the choice of normalising constants does not affect the dependence function of the limit distribution H, but only alters the marginals (see Ref. [9], Lemma 3.2). Since our main interest is in the dependence function, the present choice of normalising constants is appropriate in view of the fact that it results in Uniform$[0,1]^d$ marginals for the limit distribution when $\{\xi_n\}$ is iid, so that in particular $D_H=H$. According to Theorem 3.3 of Ref. [9], the class of all possible limits H in Eq. (3) (for iid $\{\xi_n\}$) is precisely the class of *extreme* dependence functions, that is those that satisfy

$$D^n(\mathbf{t})=D(t_1^n,\ldots,t_d^n) \tag{4}$$

for each $n>1$ and $\mathbf{t}=(t_1,\ldots,t_d)\in[0,1]^d$. Theorem 1.1 below shows that the same is true also if $\{\xi_n\}$ is a stationary sequence satisfying the following mixing condition.

For $\mathbf{t}\in T$, let

$$\mathcal{B}_k^l(v_n(\mathbf{t}))=\sigma\{(\xi_{ij}>v_{nj}(t_j)) : k\le i\le l,\ j=1,\ldots,d\}$$

and for $1\le l\le n-1$, define

$$\alpha_{n,l}=\sup\{\,|P(A\cap B)-P(A)P(B)| : A\in\mathcal{B}_1^k(v_n(\mathbf{t})),$$

$$B\in\mathcal{B}_{k+l}^n(v_n(\mathbf{t})),\ 1\le k<k+l\le n\}.$$

The mixing condition $\Delta(v_n(\mathbf{t}))$ is then said to hold if $\alpha_{n,l_n}\to 0$ for some sequence $\{l_n\}$ satisfying $l_n/n\to 0$. This is the multivariate version of the mixing condition used in Ref. [5] and is slightly stronger than the $D(u_n)$ condition in Ref. [9]. Henceforth $\{\xi_n\}$ will be assumed to satisfy $\Delta(v_n(\mathbf{t}))$, for some or all $\mathbf{t}$, as required.

THEOREM 1.1. *Let $\{\xi_n\}$ satisfy $\Delta(v_n(\mathbf{t}))$ for all $\mathbf{t}\in T$ and suppose that $P\{M_n\le v_n(\mathbf{t})\}\to^w H(\mathbf{t})$, non-degenerate. Then D_H is an extreme dependence function and hence, in particular, $H(\mathbf{t}^c)=H^c(\mathbf{t})$ for each $\mathbf{t}\in[0,1]^d$ and $c>0$ (where $\mathbf{t}^c=(t_1^c,\ldots,t_d^c)$).*

PROOF: The first part is an immediate consequence of Theorem 4.2 of Ref. [9] while the second part follows from the definition of extreme dependence functions upon noting that (by the univariate theory described above), the marginals of H are of the form $H_j(t_j)=t_j^{\theta_j}$ where θ_j is the extremal index of $\{\xi_{nj}\}$, the jth-component sequence of $\{\xi_n\}$.

In the next section we apply the exceedance point process approach to multivariate extremes and obtain some weak convergence results. The multivariate extremal index is then defined (in Sec. 3), based on the multivariate analogue of Eq. (1). It is seen to be a function of only $d-1$ variables and its properties naturally extend those of the univariate extremal index. Finally in Sec. 4 we consider two examples of bivariate moving average sequences for which the computation of the extremal index is demonstrated.

2. Exceedance Point Processes

Fix $t\in T$ and let $\delta_{ij}=I_{\{\xi_{ij}>v_{nj}(t_j)\}}$, $i=1,\ldots,n$, $j=1,\ldots d$, and put $\delta_i=(\delta_{i1},\ldots,\delta_{id})$. The *multivariate exceedance point process* $N_n=N_n^{(\mathbf{t})}$ is then defined by

$$N_n(B)=\sum_{i=1}^{n} I_{\{i/n\in B\}}\delta_i,\quad B\in[0,1]. \tag{5}$$

Assume that $\{\xi_n\}$ satisfies $\Delta(v_n(\mathbf{t}))$. If also $N_n\to^d N_0$, then it may be shown (as in the univariate case) that the limit N_0 is a point process on $[0,1]$ which is of Compound Poisson type. More precisely, the Laplace Transform of N_0 is given by

$$-\log E\exp\{-\sum_{j=1}^{d}\int_{[0,1]} f_j\,\mathrm{d}N_{0j}\}=\nu\int_{x\in[0,1]}\int_{\mathbf{y}\in\mathbb{Z}_+^{d\wedge}}$$

$$(1-\exp\{-\sum_{j=1}^{d} y_j f_j(x)\})\mathrm{d}\pi(\mathbf{y})\mathrm{d}x. \tag{6}$$

Here N_{0j} denotes the jth-component of N_0, ν is a positive constant, π is a probability distribution on $\mathbb{Z}_+^{d\wedge}=\{0,1,2,\ldots\}^{d\wedge}\setminus\{\mathbf{0}\}$ and f_j's are non-negative functions on $[0,1]$.

Let $\{k_n\}$ be any sequence of positive integers satisfying

$$k_n\to\infty,\ k_n l_n/n\to 0,\ \text{and}\ k_n\alpha_{n,l_n}\to 0,\quad \text{as } n\to\infty. \tag{7}$$

Set $r_n=[n/k_n]$ (the largest integer not exceeding n/k_n) and put $J_{n1}=[0,r_n/n]$. Define the probability distribution π_n on $\mathbb{Z}_+^{d\wedge}$ by

$$\pi_n(\mathbf{y})=P\{N_n(J_{n1})=\mathbf{y}\,|\,N_n(J_{n1})\ne\mathbf{0}\},\ \mathbf{y}\in\mathbb{Z}_+^{d\wedge}.$$

The following theorem which gives a useful characterization of the convergence of N_n is an immediate consequence of the results in Sec. 5 of Ref. [12].

THEOREM 2.1. *$N_n\to^d N_0$ if and only if $\pi_n\to^w\pi$ and $P\{M_n\le v_n(\mathbf{t})\}\to e^{-\nu}$, and in that case the Laplace Transform of N_0 is given by Eq. (6).*

Next we consider the iid case in some detail and obtain an interesting connection with Theorem 5.3.1 of Ref. [8].

PROPOSITION 2.2. *Let $\{\xi_n\}$ be iid and for fixed $\mathbf{t}\in T$ let N_n be defined by Eq. (5). If $N_n\to^d N_0$ then the multiplicity distribution π in Eq. (6) is supported on the set $S=\{0,1\}^{d\wedge}\setminus\{\mathbf{0}\}$.*

PROOF: Observe that $\Delta(v_n(\mathbf{t}))$ is trivially satisfied since $\alpha_{n,1}=0$ so that we may take $l_n\equiv 1$ and $k_n=n$. Then $\pi_n(\mathbf{y})=P\{\delta_1=\mathbf{y}\,|\,\xi_1\not\le v_n(\mathbf{t})\}$, $\mathbf{y}\in\mathbb{Z}_+^{d\wedge}$, which is clearly supported on S. The result is now immediate since S is a closed set and $\pi_n\to^w\pi$ by Theorem 2.1. □

Making the dependence on $\mathbf{t}$ explicit, we now write $N_n=N_n^{(\mathbf{t})}$, $N_0=N_0^{(\mathbf{t})}$, $\nu=\nu^{(\mathbf{t})}$ and $\pi=\pi^{(\mathbf{t})}$. In addition we shall

require the following notation from Ref. [8]. For $1\le k\le d$, let $\mathbf{j}(k)=(j_1,\ldots,j_k)$ denote a vector with integer-valued components $1\le j_1<j_2<\cdots<j_k\le d$, and for $\mathbf{x}=(x_1,\ldots,x_d)\in\mathrm{IR}^d$ write $\mathbf{x}_{\mathbf{j}(k)}=(x_{j_1},\ldots,x_{j_k})$. Define the "survival function"

$$G(\mathbf{x})=P\{\xi_{11}>x_1,\ldots,\xi_{1d}>x_d\}$$

and write $G_{\mathbf{j}(k)}(\mathbf{x})=P\{\xi_{1j_1}>x_{j_1},\ldots,\xi_{1j_k}>x_{j_k}\}$. For each $\mathbf{j}(k)$, let $\mathbf{y}_{\mathbf{j}(k)}$ denote the element in $S=\{0,1\}^d\setminus\{\mathbf{0}\}$ whose jth component equals 1 if and only if $j=j_i$ for some $i=1,\ldots,k$. (This defines a natural 1-1 correspondence between S and the $\mathbf{j}(k)$'s.)

THEOREM 2.3. *Let $\{\xi_n\}$ be iid. Then $N_0^{(t)}\to^d N_0^{(t)}$ for some fixed $\mathbf{t}\in T$ if and only if*

$$\lim_{n\to\infty} nG_{\mathbf{j}(k)}(v_n(\mathbf{t}))=h_{\mathbf{j}(k)}(\mathbf{t})<\infty \tag{8}$$

for each $\mathbf{j}(k), 1\le k\le d$. In that case $N_0^{(\mathbf{t})}$ has Laplace Transform given by Eq. (6) with

$$\nu^{(\mathbf{t})}=\sum_{k=1}^{d}(-1)^{k+1}\sum_{1\le j<\cdots<j_k\le d} h_{\mathbf{j}(k)}(\mathbf{t}) \tag{9}$$

and with $\pi^{(\mathbf{t})}$ determined by the relations

$$h_{\mathbf{j}(k)}(\mathbf{t})=\nu^{(\mathbf{t})}\sum_{\mathbf{y}\ge\mathbf{y}_{\mathbf{j}(k)}}\pi^{(\mathbf{t})}(\mathbf{y}). \tag{10}$$

PROOF: Write $S_k(\mathbf{t})=\Sigma_{1\le j_1<\cdots<j_k\le d}G_{\mathbf{j}(k)}(\mathbf{t})$, so that $P\{\xi_1\not\le v_n(\mathbf{t})\}=\Sigma_{k=1}^{d}(-1)^{k+1}S_k(v_n(\mathbf{t}))$. If Eq. (8) holds for each k, then

$$\lim_{n\to\infty} nP\{\xi_1\not\le v_n(\mathbf{t})\}=\sum_{k=1}^{d}(-1)^{k+1}\sum_{1\le j_1<\cdots<j_k\le d} h_{\mathbf{j}(k)}(\mathbf{t})=\nu^{(\mathbf{t})}, \tag{11}$$

and hence $\lim_{n\to\infty}P\{M_n\le v_n(\mathbf{t})\}=e^{-\nu^{(\mathbf{t})}}$.

Next observe that for each $\mathbf{j}(k)$, $1\le k\le d$,

$$G_{\mathbf{j}(k)}(v_n(\mathbf{t}))=\sum_{\mathbf{y}\ge\mathbf{y}_{\mathbf{j}(k)}}P\{\delta_1=\mathbf{y}\}=P\{\xi_1\not\le v_n(\mathbf{t})\}\sum_{\mathbf{y}\ge\mathbf{y}_{\mathbf{j}(k)}}\pi_n(\mathbf{y}), \tag{12}$$

where $\pi_n(\mathbf{y})=P\{\delta_1=\mathbf{y}\mid\xi_1\not\le v_n(\mathbf{t})\}$. Moreover, this relationship is invertible in the sense that each of the probabilities $\pi_n(\mathbf{y})$, $\mathbf{y}\in S$, can be expressed as a linear combination of the $G_{\mathbf{j}(k)}(v_n(\mathbf{t}))$'s. Therefore by Eq. (8), $\lim_{n\to\infty}\pi_n(\mathbf{y})=\pi^{(\mathbf{t})}(\mathbf{y})$ (say) exists and satisfies Eq. (10). Hence by Theorem 2.1 $N_n^{(\mathbf{t})}\to^d N_0^{(\mathbf{t})}$ where $N_0^{(\mathbf{t})}$ has the specified parameters. Conversely if $N_n^{(\mathbf{t})}\to^d N_0^{(\mathbf{t})}$ then π_n and $P\{M_n\le v_n(\mathbf{t})\}$ converge (by Theorem 2.1 again), and hence Eq. (8) follows by virtue of Eqs. (11) and (12). □

COROLLARY 2.4. *Let $\{\xi_n\}$ be iid. Then $N_n^{(\mathbf{t})}\to^d N_0^{(\mathbf{t})}$ for each $\mathbf{t}\in T$ if and only if $P\{M_n\le v_n(\mathbf{t})\}\to^w H(\mathbf{t})$. Moreover H and $\{\nu^{(\mathbf{t})},\pi^{(\mathbf{t})}\}_{\mathbf{t}\in T}$ determine each other.*

PROOF: (Sketch) The first part follows from Theorem 2.3 above and Theorem 5.3.1 of Ref. [8] which states that $P\{M_n\le v_n(\mathbf{t})\}\to^w H(\mathbf{t})$ if and only if Eq. (8) holds for each $\mathbf{t}\in T$. Note that $H(\mathbf{t})=e^{-\nu^{(\mathbf{t})}}$ so that H and the $\nu^{(\mathbf{t})}$'s can be obtained from each other. Also the $\pi^{(\mathbf{t})}$'s can be obtained from the $\nu^{(\mathbf{t})}$'s by first inverting Eq. (9) to get the $h_{\mathbf{j}(k)}(\mathbf{t})$'s and then inverting Eq. (10). (The inversion of Eq. (9) is carried out inductively using the fact that the weak convergence of $H_n(v_n(\mathbf{t}))$ implies that of all lower dimensional marginals.) □

Analogous results for the dependent case take on a slightly different form. Let $\{\xi_n\}$ be a stationary sequence satisfying $\Delta(v_n(\mathbf{t}))$ for each $\mathbf{t}\in T$. As before let $r_n=[n/k_n]$ where $\{k_n\}$ is any sequence satisfying Eq. (7), and define

$$G_{r_n\mathbf{j}(k)}(v_n(\mathbf{t}))=P\{M_{r_nj_1}>v_{n,j_1}(t_{j_1}),\ldots,M_{r_nj_k}>v_{n,j_k}(t_{j_k})\}.$$

THEOREM 2.5. *Let $\{\xi_n\}$ be a stationary sequence satisfying $\Delta(v_n(\mathbf{t}))$ for each $\mathbf{t}\in T$. Then $P\{M_n\le v_n(\mathbf{t})\}\to^w H(\mathbf{t})$ if and only if*

$$\lim_{n\to\infty} k_nG_{r_n\mathbf{j}(k)}(v_n(\mathbf{t}))=h_{\mathbf{j}(k)}(\mathbf{t})<\infty$$

for each $\mathbf{j}(k)$, $1\le k\le d$ and $\mathbf{t}\in T$, and in that case

$$H(\mathbf{t})=\exp\{\sum_{k=1}^{d}(-1)^k\sum_{1\le j_1<\cdots<j_k\le d} h_{\mathbf{j}(k)}(\mathbf{t})\}.$$

PROOF: Observe that the mixing condition $\Delta(v_n(\mathbf{t}))$ implies that $\{\xi_{n,\mathbf{j}(k)}\}$ satisfies $\Delta(v_{n,\mathbf{j}(k)}(\mathbf{t}))$ for each $\mathbf{j}(k)$ (with obvious notation). Hence it may be shown as in the univariate case (see Lemma 2.1 of Ref. [1]) that

$$P\{M_{n,\mathbf{j}(k)}\le v_{n,\mathbf{j}(k)}(\mathbf{t})\}-P^{k_n}\{M_{r_n\mathbf{j}(k)}\le v_{n,\mathbf{j}(k)}(\mathbf{t})\}\to 0, \tag{13}$$

for each $\mathbf{j}(k)$. The result may therefore be proved in exactly the same way as Theorem 5.3.1 of Ref. [8]. □

REMARK: Under the hypothesis of Theorem 2.5, if $N_n^{(\mathbf{t})}\to^d N_0^{(\mathbf{t})}$, $\mathbf{t}\in T$, with parameters $\nu^{(\mathbf{t})}$ and $\pi^{(\mathbf{t})}$ then $P\{M_n\le v_n(\mathbf{t})\}\to^w H(\mathbf{t})=e^{-\nu^{(\mathbf{t})}}$, as in the iid case. However it is not possible in general to recover the $\pi^{(\mathbf{t})}$'s from H since the clustering of exceedances may cause the support of $\pi^{(\mathbf{t})}$ to extend beyond S. References [9, 10] give sufficient conditions (analogous to the $D'(u_n)$ condition of Ref. [13]) under which clustering does not occur, so that Corollary 2.4 can be extended to stationary sequences satisfying this condition.

A distribution function F on IR^d is said to be *independent* if $F(\mathbf{x})=\Pi_{j=1}^d F_j(x_j)$, $\mathbf{x}\in\mathrm{IR}^d$. If $\{\xi_n\}$ is iid and $P\{M_n\le v_n(\mathbf{t})\}\to^w H(\mathbf{t})$, then it follows from Corollary 5.3.1 of Ref. [8] that H is independent if and only if the marginals of H are pairwise independent. The analogous result for the dependent case is stated below. The proof (which is omitted) is essentially the same as for the iid case, but uses Theorem 2.5 instead of Theorem 5.3.1 of Ref. [8].

COROLLARY 2.6. *Let $\{\xi_n\}$ be a stationary sequence satisfying $\Delta(v_n(\mathbf{t}))$ for each $\mathbf{t}\in T$ and suppose that $P\{M_n\le v_n(\mathbf{t})\}\to^w H(\mathbf{t})$. Then H is independent if and only if $k_nP\{M_{r_{nj}}>v_{nj}(t_j),\ M_{r_{nl}}>v_{nl}(t_l)\}\to 0$ for each $1\le j<l\le d$, $\mathbf{t}\in T$, i.e., if and only if $k_nG_{r_{nj(k)}}(v_n(\mathbf{t}))\to 0$ for each $\mathbf{j}(2)$ and each $\mathbf{t}\in T$.*

It is shown in Ref. [14] that H is independent if $H(\mathbf{t})=\Pi_{j=1}^d H_j(t_j)$ for some $\mathbf{t}\in(0,1)^d$. Although the result in [14] only stated for iid sequences under a linear normalization, the proof essentially rests on the defining property of extreme dependence functions, namely Eq. (1). Consequently the result extends to the present more general situation allowing dependence and non-linear normalizations. Corollary 2.6 can therefore be improved as follows.

COROLLARY 2.7. *Let $\{\xi_n\}$ be as in Corollary 2.6 and suppose that $P\{M_n\le v_n(\mathbf{t})\}\to^w H(\mathbf{t})$. Then the following are equivalent:*

(i) H is independent,

(ii) $H(\mathbf{t})=\Pi_{j=1}^d H_j(t_j)$ for some $\mathbf{t}\in(0,1)^d$,

(iii) $k_nG_{r_{nj(k)}}(v_n(\mathbf{t}))\to 0$ for each $\mathbf{j}(2)$, for some $\mathbf{t}\in(0,1)^d$.

It should be noted that Refs. [9, 10] give some interesting sufficient conditions for H to be independent when $\{\xi_n\}$ is a stationary sequence. A natural question to ask in the present context is whether H is independent whenever $\hat{H}$ is. Proposition 3.4 gives a necessary and sufficient condition for this in terms of the extremal index, but the answer in general is negative and a counter-example can be found in [10]. It seems more plausible that the converse may be true, i.e., that $\hat{H}$ is independent whenever H is. In fact however, this too is not the case, as shown by an interesting counter-example in [15].

We conclude this section by stating a result which extends Theorem 5.1 of [5] and is proved similarly.

THEOREM 2.8. *Let $\{\xi_n\}$ be a stationary sequence satisfying $\Delta(v_n(\mathbf{t}))$ for each $\mathbf{t}\in T$ and suppose that $N_n^{(\mathbf{t})}\to^d N_0^{(\mathbf{t})}$ for some $\mathbf{t}\in T$. Then $N_n^{(\mathbf{t}^c)}\to^d N_0^{(\mathbf{t}^c)}$ for each $c>0$ and furthermore, $\nu^{(\mathbf{t}^c)}=c\nu^{(\mathbf{t})}$ and $\pi^{(\mathbf{t}^c)}=\pi^{(\mathbf{t})}$ (where $\mathbf{t}^c=(t_1^c,\ldots,t_d^c)$).*

3. The Multivariate Extremal Index

Let $\{\xi_n\}$ be a stationary sequence and $\{\hat{\xi}_n\}$ the associated iid sequence. Suppose that $P\{M_n\le v_n(\mathbf{t})\}\to^w H(\mathbf{t})$ and $P\{\hat{M}_n\le v_n(\mathbf{t})\}\to^w \hat{H}(\mathbf{t})$. The *multivariate extremal index* of $\{\xi_n\}$ is then defined by the relation $H(\mathbf{t})=\hat{H}^{\theta(\mathbf{t})}(\mathbf{t})$ (see Eq. (1)), or more explicitly

$$\theta(\mathbf{t})=\log H(\mathbf{t})/\log \hat{H}(\mathbf{t}),\quad \mathbf{t}\in T.$$

Observe that $\theta(\mathbf{t})$ is well defined since $\hat{H}$ has Uniform[0,1] marginals and hence, $0<\hat{H}(\mathbf{t})<1$ on T. The following results describe some basic properties of the multivariate extremal index.

PROPOSITION 3.1. *Assume that $\{\xi_n\}$ satisfies $\Delta(v_n(\mathbf{t}))$ for each $\mathbf{t}\in T$ and has extremal index $\theta(\mathbf{t})$. Then*

(i) $\theta(\mathbf{t})=\theta(\mathbf{t}^c)$ for each $\mathbf{t}\in T$ and $c>0$, and

(ii) for each $j=1,\ldots,d$, $\{\xi_{nj}\}$ has extremal index $\theta_j=\theta(\mathbf{t})$ where $\mathbf{t}\in T$ has all coordinates equal to 1 except the jth.

(Note that by (i), $\theta(\mathbf{t})$ is constant along the contours $L_t=\{\mathbf{t}^c: c>0\}$, $\mathbf{t}\in T$, and hence θ_j in (ii) is well-defined.)

PROOF: Recall that (by Theorem 1.1) $H(\mathbf{t}^c)=H^c(\mathbf{t})$ and $\hat{H}(\mathbf{t}^c)=\hat{H}^c(\mathbf{t})$ so that (i) follows from the definition of the extremal index. Next, for $\mathbf{t}\in T$ with all coordinates but the jth equal to 1, $P\{M_n\le v_n(\mathbf{t})\}=P\{M_{nj}\le v_{nj}(t_j)\}$ and hence

$$\lim_{n\to\infty} P\{M_{nj}\le v_{nj}(t_j)\}=\lim_{n\to\infty} P\{M_n\le v_n(\mathbf{t})\}=H(\mathbf{t})=H_j(t_j).$$

Therefore by Theorem 2.2 of Ref. [1], $\{\xi_{nj}\}$ has extremal index θ_j (say) so that $H_j(t_j)=t_j^{\theta_j}$. Now $H(\mathbf{t})=\hat{H}^{\theta(\mathbf{t})}(\mathbf{t})$ by definition of the extremal index, and for the present choice of $\mathbf{t}$ this is the same as $H_j(t_j)=t_j^{\theta(\mathbf{t})}$, whence it follows that $\theta(\mathbf{t})=\theta_j$ for all such $\mathbf{t}$. □

For $\mathbf{t}\in T$, let $\tilde{N}_n^{(\mathbf{t})}$ denote the one-dimensional point process obtained from $N_n^{(\mathbf{t})}$ via the map $\mathbf{y}\to I_{\{\mathbf{y}\neq 0\}}$ from $\{0,1\}^d$ to $\{0,1\}$, i.e., $\tilde{N}_n^{(\mathbf{t})}(B)=\Sigma_{i=1}^n I_{\{i/n\in B\}}I_{\{\delta_i\neq 0\}}$, $B\in\mathcal{B}$. Thus $\tilde{N}_n^{(\mathbf{t})}$ has unit mass at i/n if and only if $\xi_i\not\le v_n(\mathbf{t})$. Assume that $\{\xi_n\}$ satisfies $\Delta(v_n(\mathbf{t}))$ and with J_{n1} as in Sec. 2, let

$$\tilde{\pi}_n^{(\mathbf{t})}(y)=P\{\tilde{N}_n^{(\mathbf{t})}(J_{n1})=y \mid \tilde{N}_n^{(\mathbf{t})}(J_{n1})>0\},\quad y\ge 1.$$

PROPOSITION 3.2. *Assume that $\{\xi_n\}$ satisfies $\Delta(v_n(\mathbf{t}))$ for each $\mathbf{t}\in T$ and has extremal index $\theta(\mathbf{t})$. Then $\theta(\mathbf{t})=(\lim_{n\to\infty}\Sigma_{y\ge 1} y\tilde{\pi}_n^{(\mathbf{t})}(y))^{-1}$.*

PROOF: Observe that

$$\sum_{y\geq 1} y\tilde{\pi}_n^{(\mathbf{t})}(y)=E(\tilde{N}_n^{(\mathbf{t})}(J_{n1})\,|\,\tilde{N}_n^{(\mathbf{t})}(J_{n1})>0)=\frac{r_nP\{\xi_1\not\leq v_n(\mathbf{t})\}}{P\{\tilde{N}_n^{(\mathbf{t})}(J_{n1})>0\}}=$$

$$\frac{k_nr_n}{n}\,\frac{nP\{\xi_1\not\leq v_n(\mathbf{t})\}}{k_nP\{M_{r_n}\not\leq v_n(\mathbf{t})\}}.$$

Now $\lim_{n\to\infty} P\{\tilde{M}_n\leq v_n(\mathbf{t})\}=\hat{H}(\mathbf{t})$ and $\lim_{n\to\infty} P\{M_n\leq v_n(\mathbf{t})\}=H(\mathbf{t})$ (by assumption), so that $\lim_{n\to\infty} nP\{\xi_1\not\leq v_n(\mathbf{t})\}=-\log \hat{H}(\mathbf{t})$ and (by Eq. (13)) $\lim_{n\to\infty} k_nP\{M_{r_n}\not\leq v_n(\mathbf{t})\}=-\log H(\mathbf{t})$. Therefore $\lim_{n\to\infty} \Sigma_{y\geq 1} y\tilde{\pi}_n^{(\mathbf{t})}(y)=\log \hat{H}(\mathbf{t})/\log H(\mathbf{t})=1/\theta(\mathbf{t})$, as required. □

REMARK: Proposition 3.2 is simply the multivariate version of Eq. (1) and shows how the extremal index is related to the clustering of "exceedances." Indeed, according to the present viewpoint, an exceedance occurs at time i if $\xi_i\not\leq v_n(\mathbf{t})$, i.e., if $\xi_{ij}>v_{nj}(t_j)$ for at least one j. Thus Propositions 3.1 and 3.2 show that while the degree of clustering may depend on $\mathbf{t}$, it is constant on each $L_{\mathbf{t}}$. Note also the connection to Theorem 2.8.

The next result gives the relation between the dependence functions of H and $\hat{H}$, which is seen to involve the extremal index in an intricate manner.

PROPOSITION 3.3. *If $\{\xi_n\}$ has extremal index $\theta(\mathbf{t})$, $\mathbf{t}\in T$, then*

$$D_H(t_1^{\theta_1},\ldots,t_d^{\theta_d})=D_{\hat{H}}^{\theta(\mathbf{t})}(\mathbf{t}),\quad \mathbf{t}\in T, \tag{14}$$

where θ_j is the extremal index of $\{\xi_{nj}\}$, $j=1,\ldots,d$.

PROOF: By definition of the dependence function, $D_H(\mathbf{t})=P\{H_1(X_1)\leq t_1,\ldots,H_d(X_d)\leq t_d\}$ where $(X_1,\ldots,X_d)$ is a random vector with distribution H. Therefore, since $H_j(t_j)=t_j^{\theta_j}$,

$$D_H(t_1^{\theta_1},\ldots,t_d^{\theta_d})=P\{X_1\leq t_1,\ldots,X_d\leq t_d\}=H(\mathbf{t})=\hat{H}^{\theta(\mathbf{t})}(\mathbf{t})$$

$$=D_{\hat{H}}^{\theta(\mathbf{t})}(\mathbf{t}),$$

or required. □

REMARKS

1.) Note that $\mathbf{s}=\mathbf{t}^c$ (for some $c>0$) if and only if $\log s_j/\log s_d=\log t_j/\log t_d=a_j$ (say), $j=1,\ldots,d-1$. Therefore we may write $L_{\mathbf{t}}=L_{\mathbf{a}}$ where $\mathbf{a}=(a_1,\ldots,a_{d-1})$, and hence by the remark following Proposition 3.2, $\theta(\mathbf{t})=\theta(\mathbf{a})$, i.e., the extremal index is a function of $d-1$ variables only.

2.) By Proposition 3.3, $D_H(t_1^{\theta_1},\ldots,t_d^{\theta_d})=D_{\hat{H}}^{\theta(\mathbf{t})}(\mathbf{t})=D_{\hat{H}}(\mathbf{t}^{\theta(\mathbf{t})})$. Also, if $\mathbf{t}\in L_{\mathbf{a}}$ then $(t_1^{\theta_1},\ldots,t_d^{\theta_d})\in L_{\mathbf{a}^*}$ where $\mathbf{a}^*=(a_1\theta_1/\theta_d,\ldots,a_{d-1}\theta_{d-1}/\theta_d)$. Thus D_H is obtained by translating the values of $D_{\hat{H}}(=\hat{H})$ on $L_{\mathbf{a}}$ onto $L_{\mathbf{a}^*}$.

3.) While the above results illustrate some of the basic properties of the multivariate extremal index, they are far from complete. For instance; it would be useful to identify the set of all "admissible" $\theta(\cdot)$ for a given $\hat{H}$, that is the set of all $\theta(\cdot)$ such that $D_H(\cdot)$ defined by Eq. (14) is a probability distribution on $[0,1]^d$. It would also be of interest to study the properties of $\theta(\cdot)$ when one or both of $\hat{H}$ and H are independent. In this context we have the following simple result.

PROPOSITION 3.4. *If $\hat{H}$ is independent, then H is independent if and only if*

$$\theta(\mathbf{t})=\sum_{j=1}^{d}\theta_j\log t_j/\sum_{j=1}^{d}\log t_j,\quad \textit{for some } \mathbf{t}\in(0,1)^d.$$

In particular, if both $\hat{H}$ and H are independent then $\theta(\mathbf{t})$ is a convex combination of the θ_j's.

PROOF: If $\hat{H}$ is independent, then $H(\mathbf{t})=\hat{H}(\mathbf{t})^{\theta(\mathbf{t})}=(\Pi_{j=1}^d t_j)^{\theta(\mathbf{t})}$. The conclusion follows immediately from Corollary 2.7 (iii) upon taking logarithms and noting that if H is independent, then $H(\mathbf{t})=\Pi_{j=1}^d t_j^{\theta_j}$.

The extremal index can be given the following more general formulation. Let $\hat{\mu}$ and μ be the probability measures on $(0,1)^d$ corresponding to $\hat{H}$ and H, respectively. Thus for instance,

$$\mu(A)=\lim_{n\to\infty} P\{M_n\in v_n(A)\}$$

where $v_n(A)=\{v_n(\mathbf{s}) : \mathbf{s}\in A\}$, $A\subset(0,1)^d$. We now define $\tilde{\theta}(A)$ via the relationship $\mu(A)=\hat{\mu}^{\tilde{\theta}(A)}(A)$, or more directly $\tilde{\theta}(A)=\log\mu(A)/\log\hat{\mu}(A)$, for subsets $A\subset(0,1)^d$ such that $\hat{\mu}(A)>0$ and $\mu(A)>0$.

Note that $\theta(\mathbf{t})=\tilde{\theta}((0,t_1)\times\cdots\times(0,t_d))$ for $\mathbf{t}\in T$. Thus if $\{\theta(\mathbf{t}) : \mathbf{t}\in T\}$ is known along with either of H or $\hat{H}$, then it is possible at least in theory to obtain $\{\tilde{\theta}(A) : A\subset(0,1)^d\}$. In practice, however, it may not be possible to obtain $\tilde{\theta}(A)$ in a tractable form, but frequently one is only interested in certain special sets, typically rectangles of the form $\Pi_{j=1}^d(a_j,b_j)$, and for such sets the computation is easy.

The definition of M_n as the vector of componentwise maxima actually corresponds to regarding ξ_i as an extreme observation if $\xi_{ij}>v_{nj}(t_j)$ for some j. More generally, one may define ξ_i to be an extreme value if $\xi_i\in v_n(A)$ for some $A\subset(0,1)^d$, in which case $\tilde{\theta}(A)$ has an interpretation as a measure of the clustering of such extremes. Note that the original definition of extremes corresponds to letting $A=((0,t_1)\times\cdots\times(0,t_d))^c$. Alternately one may consider taking $A=(t_1,1)\times\cdots\times(t_d,1)$ which corresponds to defining ξ_i as an extreme observa-

tion if $\xi_{ij}>v_{nj}(t_j)$ for all j. Yet another choice is $A=\{\mathbf{t}:\Sigma t_j^2>c\}$.

4. Examples

We conclude with two examples, both involving bivariate stationary sequences.

EXAMPLE 4.1 Let $\{\eta_n\}$ be an iid sequence, and put $\xi_{n1}=\eta_n$, and $\xi_{n2}=\max\{\eta_{n-1},\eta_n\}$. Let F denote the distribution of $\xi_n=(\xi_{n1},\xi_{n2})$ with marginals F_1 and F_2. Then $F_2(x)=P\{\xi_{n2}\le x\}=P\{\eta_{n-1}\le x,\ \eta_n\le x\}=F_1^2(x)$ and

$$F(x_1,x_2)=P\{\xi_{n1}\le x_1,\xi_{n2}\le x_2\}=\begin{cases}F_1^2(x_2), & \text{if } x_1\ge x_2\\ F_1(x_1)F_1(x_2), & \text{if } x_1<x_2.\end{cases}$$

If $v_{nj}(t_j)$ satisfies $F_j^n(v_{nj}(t_j))\to t_j$, $j=1,2$, then $\lim_{n\to\infty}F_1^n(v_{n2}(t_2))=t_2^{1/2}$ so that $v_{n2}(t_2)=v_{n1}(t_2^{1/2})$. Moreover $v_{n1}(t_1)\ge v_{n2}(t_2)$ if and only if $t_1\ge t_2^{1/2}$, and so

$$H_n(v_n(t))=P\{M_n\le v_n(t)\}\to^w H(t)=\begin{cases}t_2^{1/2}, & \text{if } t_1\ge t_2^{1/2}\\ t_1, & \text{if } t_1<t_2^{1/2}.\end{cases}$$

The marginals of H are therefore $H_1(t_1)=t_1$ and $H_2(t_2)=t_2^{1/2}$ so that $\theta_1=1$ and $\theta_2=1/2$, and the dependence function of H is $D_H(t)=H(t_1,t_2^2)=t_1\wedge t_2$. For the associated iid sequence $\{\hat{\xi}_n\}$ on the other hand, it is easily verified that

$$\hat{H}_n(v_n(t))=P\{\hat{M}_n\le v_n(t)\}\to\hat{H}(t)=\begin{cases}t_2, & \text{if } t_1\ge t_2^{1/2}\\ t_1t_2^{1/2}, & \text{if } t_1<t_2^{1/2}.\end{cases}$$

from which it follows that

$$\theta(t)=\begin{cases}\frac{1}{2}, & \text{if } t_1\ge t_2^{1/2}\\ \frac{\log t_1}{\log t_1t_2^{1/2}}, & \text{if } t_1<t_2^{1/2}.\end{cases}$$

We next consider a moving average sequence studied in Ref. [16].

EXAMPLE 4.2. Let $\{\mathbf{Z}_k=(Z_{k1},Z_{k2})'\}$, $-\infty<k<\infty$, be a sequence of iid random vectors in IR^2. We assume the existence of a sequence of positive constants $a_n\to\infty$, and a measure ν on IR^2 which is finite on sets of the form $\{\mathbf{x}:\|\mathbf{x}\|>r\}$, $r>0$ (where $\|\cdot\|$ denotes the Euclidean norm in IR^2), such that $nP\{a_n^{-1}\mathbf{Z}_0\in\cdot\}\to^v\nu(\cdot)$. (Here '$\to^v$' denotes vague convergence of measures on IR^2 with respect to the metric $d(\mathbf{x}_1,\mathbf{x}_2)=|r_1^{-1}-r_2^{-1}|\vee|\theta_1-\theta_2|$, where for $i=1,2$, r_i and θ_i denote the polar coordinates of $\mathbf{x}_i$, and $a\vee b=\max\{a,b\}$.) The measure ν is necessarily of the form $\nu(\{\mathbf{x}:\|\mathbf{x}\|>r,\theta(\mathbf{x})\in A\})=r^{-\alpha}S(A)$ for $r>0$ and $A\subset[0,2\pi)$, where $S(\cdot)$ is a probability measure on $[0,2\pi)$ and $\alpha>0$. Hence in particular [17],

$$\nu(cA)=c^{-\alpha}\nu(A) \tag{15}$$

for all $c>0$ and all sets A with $\nu(A)<\infty$.

Define the bivariate moving average process $\mathbf{X}_n=\Sigma_{j=0}^{\infty}C_j\mathbf{Z}_{n-j}$, where $\{C_j=[c_{j,kl}]_{k,l=1}^2\}_{j\ge0}$ is a sequence of real 2×2 matrices satisfying $\Sigma_{j=0}^{\infty}|c_{j,kl}|^{\delta}<\infty$, $k,l=1,2$, for some $\delta\in(0,\alpha)$, $\delta\le1$. For $\mathbf{x}=(x_1,x_2)'\in\mathrm{IR}^2$, write $A_{\mathbf{x},j}=\{\mathbf{z}:C_j\mathbf{z}\in((-\infty,x_1)\times(-\infty,x_2))^c\}$, where A^c denotes the complement of a set $A\subset\mathrm{IR}^2$. Then [16],

$$\lim_{n\to\infty}P\{a_n^{-1}\hat{M}_n\le\mathbf{x}\}=\exp\{-\hat{\gamma}(\mathbf{x})\},\text{ and}$$

$$\lim_{n\to\infty}P\{a_n^{-1}M_n\le\mathbf{x}\}=\exp\{-\gamma(\mathbf{x})\},\ \mathbf{x}\in\mathrm{IR}^2,$$

where $\hat{\gamma}(\mathbf{x})=\Sigma_{j=0}^{\infty}\nu(A_{\mathbf{x},j})$ and $\gamma(\mathbf{x})=\nu(\cup_{j=0}^{\infty}A_{\mathbf{x},j})$. The extremal index is therefore $\theta(\mathbf{x})=\gamma(\mathbf{x})/\hat{\gamma}(\mathbf{x})$, $\mathbf{x}\in\mathrm{IR}^2$. It follows from the definition of $A_{\mathbf{x},j}$ and Eq. (15) that this is in fact a function of x_1/x_2. Note that the extremal index defined above differs from that in Sec. 3 in that it is defined on IR^2 rather than $[0,1]^2$. However the two definitions are equivalent as may be seen by means of a suitable transformation from IR^2 to $[0,1]^2$.

The actual calculation of $\theta(\mathbf{x})$ may be quite difficult in general, but possible to carry out under appropriate simplifying assumptions.

Case (i). If $C_j=c_jC$ where $C=[c_{kl}]_{k,l=1}^2$ and the c_j's are non-negative constants, then $A_{\mathbf{x},j}=c_j^{-1}B(\mathbf{x})$ and $\cup_{j=0}^{\infty}A_{\mathbf{x},j}=cB(\mathbf{x})$, where $B(\mathbf{x})=\{\mathbf{z}:C\mathbf{z}\in((-\infty,x_1)\times(-\infty,x_2))^c\}$ and $c=\max\{c_j:j\ge0\}$. Therefore by Eq. (15), $\nu(A_{\mathbf{x},j})=c_j^{\alpha}\nu(B(\mathbf{x}))$ and $\nu(\cup_{j=0}^{\infty}A_{\mathbf{x},j})=c^{\alpha}\nu(B(\mathbf{x}))$ so that $\theta(\mathbf{x})=c^{\alpha}/\Sigma_{j=0}^{\infty}c_j^{\alpha}$.

Case (ii). If the C_j's are diagonal, i.e., $C_j=\mathrm{diag}[c_{j1},c_{j2}]$ with $c_{ji}\ge0$, $i=1,2$, then $A_{\mathbf{x},j}=\{\mathbf{z}:c_{j1}z_1>x_1\text{ or }c_{j2}z_2>x_2\}$ and $\cup_{j=0}^{\infty}A_{\mathbf{x},j}=((-\infty,x_1/c_1)\times(-\infty,x_2/c_2))^c$ where $c_i=\max\{c_{ji}:j\ge0\}$, $i=1,2$. In particular, taking $x_2=\infty$ and using Eq. (15) as in Case (i), we have $\nu(A_{\mathbf{x},j})=c_{j1}^{\alpha}\nu(\{\mathbf{z}:z_1>x_1\})$ and $\nu(\cup_{j=0}^{\infty}A_{\mathbf{x},j})=c_1^{\alpha}\nu(\{\mathbf{z}:z_1>x_1\})$, so that the extremal index of $\{X_{n1}\}$ is $\theta_1=c_1^{\alpha}/\Sigma_{j=0}^{\infty}c_{j,1}^{\alpha}$.

Case (iii). Let D denote the support of ν. If $D\subset\{\mathbf{z}:z_1=0\text{ or }z_2=0\}$ (which is the case if the coordinates of $\mathbf{Z}_0$ are independent), then we may write

$$\nu((-\infty,x_1)\times(-\infty,x_2))^c=a_1x_1^{-\alpha}+a_2x_2^{-\alpha},\quad x_1,x_2\ge0, \tag{16}$$

for suitable constants $a_1\ge0$ and $a_2\ge0$. Once again, assuming the $c_{j,kl}$'s to be non-negative and writing $c_{kl}=\max\{c_{j,kl}:j\ge0\}$ for $k,l=1,2$, we have (writing $a\wedge b=\min\{a,b\}$)

$$A_{\mathbf{x},j}\cap D=((-\infty,x_1/c_{j,11}\wedge x_2/c_{j,21})\times(-\infty,x_1/c_{j,12}\wedge x_2/c_{j,22}))^c\cap D$$

and

$$\cup_{j=0}^{\infty} A_{x,j} \cap D = ((-\infty, x_1/c_{11} \wedge x_2/c_{21}) \times (-\infty, x_1/c_{12} \wedge x_2/c_{22}))^c \cap D,$$

so that using Eq. (16)

$$\theta(\mathbf{x}) = \frac{a_1(x_1/c_{11} \wedge x_2/c_{21})^{-\alpha} + a_2(x_1/c_{12} \wedge x_2/c_{22})^{-\alpha}}{a_1 \sum_{j=0}^{\infty}(x_1/c_{j,11} \wedge x_2/c_{j,21})^{-\alpha} + a_2 \sum_{j=0}^{\infty}(x_1/c_{j,12} \wedge x_2/c_{j,22})^{-\alpha}}.$$

Thus putting $x_2=\infty$, we have $\theta_1=(a_1c_{11}^{\alpha}+a_2c_{12}^{\alpha})/(a_1\sum_{j=0}^{\infty}c_{j,11}^{\alpha}+a_2\sum_{j=0}^{\infty}c_{j,12}^{\alpha})$ and similarly, $\theta_2=(a_1c_{21}^{\alpha}+a_2c_{22}^{\alpha})/(a_1\sum_{j=0}^{\infty}c_{j,21}^{\alpha}+a_2\sum_{j=0}^{\infty}c_{j,22}^{\alpha})$.

If also $c_{j,12}=c_{j,21}=0$ for each j, (that is if the C_j's are diagonal), then

$$\theta(\mathbf{x}) = \frac{a_1 x_1^{-\alpha} c_{11}^{\alpha} + a_2 x_2^{-\alpha} c_{22}^{\alpha}}{a_1 x_1^{-\alpha} \sum_{j=0}^{\infty} c_{j,11}^{\alpha} + a_2 x_2^{-\alpha} \sum_{j=0}^{\infty} c_{j,22}^{\alpha}},$$

and in particular, $\theta_1=c_{11}^{\alpha}/\sum_{j=0}^{\infty}c_{j,11}^{\alpha}$ and $\theta_2=c_{22}^{\alpha}/\sum_{j=0}^{\infty}c_{j,22}^{\alpha}$. Note that in this case the limiting distributions of M_n and $\hat{M}_n$ are both independent, and hence (in accordance with Proposition 3.4) $\theta(\mathbf{x})$ is a convex combination of θ_1 and θ_2.

The non-negativeness of the C_j's assumed above is not crucial and may be relaxed, although at the cost of more involved calculations.

Acknowledgments

It is a pleasure to thank Professor M. R. Leadbetter for his encouragement and guidance during the course of this research. Part of this work was carried out during a brief visit to the University of Bern (supported by the Swiss National Science Foundation), and I am very grateful to Professor J. Hüsler for several useful discussions. Research supported by the Air Force Office of Scientific Research Contract No. F49620 85C 0144.

5. References

[1] M. R. Leadbetter, Extremes and local dependence in a stationary sequence, Z. Wahr-sch. verw. Gebiete. **65**, 291–306 (1983).
[2] G. F. Newell, Asymptotic extremes for m-dependent random variables, Ann. Math. Statist. **35**, 1322–1325 (1964).
[3] R. M. Loynes, Extreme values in uniformly mixing stochastic processes, Ann. Math. Statist. **36**, 993–999 (1965).
[4] G. L. O'Brien, The maximum term of uniformly mixing stationary processes, Z. Wahr-sch. verw. Gebiete. **30**, 57–63 (1974).
[5] T. Hsing, J. Hüsler, and M. R. Leadbetter, On the exceedance point process for a stationary sequence, Probab. Theory Rel. Fields. **78**, 97–112 (1988).
[6] L. de Haan and S. Resnick, Limit theory for multivariate sample extremes, Z. Wahrsch. verw. Gebiete. **40**, 317–337 (1977).
[7] A. W. Marshall and I. Olkin, Domains of attraction of multivariate extreme value distributions, Ann. Probab. **11**, 168–177 (1983).
[8] J. Galambos, The asymptotic theory of extreme order statistics, Wiley, New York (1987).
[9] T. Hsing, Extreme value theory for multivariate stationary sequences, J. Multivariate Anal. **29**, 274–291 (1989).
[10] J. Hüsler, Multivariate extreme values in stationary random sequences, Stochastic Proc. Appl. **35**, 99–108 (1990).
[11] P. Deheuvels, Caractèrisation compléte des lois extrêmes mulitvariées et de la convergence des types extrêmes, Publ. Instit. Statist. Univ. Paris. **23**, 1–36 (1978).
[12] S. Nandagopalan, M. R. Leadbetter, and J. Hüsler, Limit theorems for multi-dimensional random measures, 92-14, Department of Statistics, Colorado State University, November 1992.
[13] M. R. Leadbetter, (1974). On extreme values in stationary sequences, Z. Wahrsch. verw. Gebiete. **28**, 298–303 (1974).
[14] R. Takahashi, Some properties of multivariate extreme value distributions and multivariate tail equivalence, Ann. Inst. Statist. Math. **39**, 637–647 (1987).
[15] N. Catkan, University of Bern, Private communication, (1993).
[16] R. A. Davis, J. Marengo, and S. Resnick, Extremal properties of a class of multivariate moving averages, L1-4, Proceedings of the 45th Session of the I.S.I., Amsterdam, (1985) pp. 1–14.
[17] S. Resnick, Extreme values, regular variation, and point processes, Springer-Verlag, New York (1987).

About the author: *S. Nandagopalan is an Assistant Professor of Statistics at Colorado State University.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 551 (1994)]

Domains of Attraction of Multivariate Extreme Value Distributions

Volume 99 | Number 4 | July–August 1994

Rinya Takahashi

Kobe University of Mercantile Marine,
Kobe, Hyogo 658, Japan

Some necessary and sufficient conditions for domains of attraction of multivariate extreme value distributions are shown by using dependence functions. The joint asymptotic distribution of multivariate extreme statistics is also shown.

Key words: asymptotic joint distribution; dependence function; multivariate extreme statistics.

1. Introduction

Multivariate extreme value distributions have been studied by many authors, and their contributions are summarized by Galambos [1] and Resnick [2]. The purpose of this paper is to obtain some necessary and sufficient conditions for domains of attraction of the multivariate extreme value distributions. The joint asymptotic distribution of multivariate extreme statistics is also obtained. To study multivariate extreme value distributions and their domains of attraction, Sibuya [3] introduces the notion of a dependence function which is also used by Galambos [1]. A dependence function or copula is a useful notion to construct a family of joint distributions.

In this paper, basic arithmetical operations are always meant componentwise (see Galambos [1], Chapt. 5).

Let $(X_{1j},X_{2j},...,X_{kj})$, $j=1,2,...,n$, be a sample of size n, of a k-dimensional random vector with a distribution function $F(\boldsymbol{x})$. The i-dimensional distribution function of the components $X_{j_1},X_{j_2},...,X_{j_i}$ will be denoted $F_{j_1j_2...j_i}(x_{j_1},x_{j_2},...,x_{j_i})=F_{J(i)}(\boldsymbol{x}_{J(i)})$. We shall also use the notation $\bar{F}_{j_1...j_i}(x_{j_1},...,x_{j_i})=\bar{F}_{J(i)}(\boldsymbol{x}_{J(i)})=P(X_{j_1}>x_{j_1},...,X_{j_i}>x_{j_i})$. For $k=1$ and $p\in(0,1)$, let $F^{-1}(p)=\inf\{x:F(x)\geq p\}$.

Let $\mathbf{Z}_n=(Z_{1n},...,Z_{kn})$, where $Z_{in}=\max\{X_{i1},...,X_{in}\}$, $i=1,2,...,k$, and let us call $\mathbf{Z}_n$ a multivariate extreme statistic.

If there exist $\boldsymbol{a}_n>\mathbf{0},\boldsymbol{b}_n\in R^k,n=1,2,...$ ($\boldsymbol{a}_n>\mathbf{0}$ means $a_{in}>0,i=1,...,k$) such that $(\mathbf{Z}_n-\boldsymbol{b}_n)/\boldsymbol{a}_n$ converges in distribution to a random vector $\boldsymbol{U}$ with a nondegenerate distribution H (i.e., all univariate marginals of H are nondegenerate), then F is said to be in the domain of attraction of H, $F\in D(H)$ by symbol, and H is said to be a multivariate extreme value distribution. The convergence in distribution is equivalent to the condition

$$\lim_{n\to\infty} F^n(\boldsymbol{a}_n\boldsymbol{x}+\boldsymbol{b}_n)=H(\boldsymbol{x}) \tag{1}$$

for all $\boldsymbol{x}$, because multivariate extreme value distributions are continuous.

We shall need the following lemma to prove a proposition in Sec. 2.

Lemma 1.1 *Equation (1) is equivalent to*

$$\lim_{n\to\infty} n[1-F(\boldsymbol{a}_n\boldsymbol{x}+\boldsymbol{b}_n)]=-\log H(\boldsymbol{x})$$

for all $\boldsymbol{x}$ *such that* $0<H(\boldsymbol{x})<1$. (See Marshall and Olkin [4].)

2. Domains of Attraction

For any k-dimensional distribution F,

$$D_F(\mathbf{y})=F(F_1^{-1}(y_1),...,F_k^{-1}(yk)),\ \mathbf{y}=(y_1,...,y_k)\in(0,1)^k$$

is called the dependence function of F. In this section, we derive necessary and sufficient conditions for domains of attraction in terms of the dependence function.

Proposition 2.1 *Let F be a k-dimensional distribution and let H be a multivariate extreme value distribution with univariate marginals H_i, $i=1,...,k$. Then the following statements are equivalent:*

1) $F\in D(H)$.

2) $F_i\in D(H_i)$, $i=1,...,k$, and

$$\lim_{s\to\infty} s[1-D_F(\mathbf{y}^{1/s})]=-\log D_H(\mathbf{y}) \text{ for all } \mathbf{y}\in(0,1)^k.$$

3) $F_i\in D(H_i)$, $i=1,...,k$, and

$$\lim_{x\uparrow 1} \frac{1-D_F(\mathbf{y}^{1-x})}{1-x}=-\log D_H(\mathbf{y}) \text{ for all } \mathbf{y}\in(0,1)^k.$$

4) $F_i\in D(H_i)$, $i=1,...,k$, and

$$\lim_{t\downarrow 0} \frac{1-D_F(\mathbf{y}^t)}{1-D_H(\mathbf{y}^t)}=1 \text{ for all } \mathbf{y}\in(0,1)^k.$$

Proof. The proof is straightforward from Lemma 1.1, Theorem 5.2.3 and Lemma 5.4.1 of Galambos [1]. □

Proposition 2.2 *Let F be a k-dimensional distribution and let H be a multivariate extreme value distribution with univariate marginals H_i, $i=1,...,k$.*

(A) $F\in D(H)$ if and only if $F_i\in D(H_i)$, $i=1,...,k$, and the functions

$$d_{J(i)}(\mathbf{y}_{J(i)})=\lim_{n\to\infty} n\bar{D}_{F_{J(i)}}((\mathbf{y}_{J(i)})^{1/n})$$

for each fixed vector $J(i)(i>1)$ and for all $\mathbf{y}\in(0,1)^k$ are finite, and the function

$$D_H(\mathbf{y};r)=y_1\cdots y_k \exp\{\sum_{i=2}^{k}(-1)^i \sum_{1\le j_1<\cdots<j_i\le k} d_{J(i)}(\mathbf{y}_{J(i)})\}$$

is a dependence function of H.

(B) The following inequalities hold.

$$D_H(\mathbf{y};2r+1)\le D_H(\mathbf{y})\le D_H(\mathbf{y};2r),$$

for a nonnegative integer r, where

$$D_H(\mathbf{y};r)=y_1\cdots y_k \exp\{\sum_{i=2}^{r}(-1)^i \sum_{1\le j_1<\cdots<j_i\le k} d_{J(i)}(\mathbf{y}_{J(i)})\}$$

and $D_H(\mathbf{y};r)$ is a dependence function of a multivariate extreme value distribution.

Proof. It is easily seen that for all $s>0$,

$$sd_{J(i)}((\mathbf{y}_{J(i)})^{1/s})=d_{J(i)}(\mathbf{y}_{J(i)}).$$

From Theorems 5.3.1 and 5.2.4 of Galambos [1], we have the result. □

Example 2.1 (See Examples 5.2.2 and 5.2.3 of Galambos [1].) For a Mardia's distribution

$$F(x_1,x_2)=1-e^{-x_1}-e^{-x_2}+(e^{x_1}+e^{x_2}-1)^{-1},$$

$$D_F(y_1,y_2)=y_1+y_2-1+\left[\frac{1}{1-y_1}+\frac{1}{1-y_2}-1\right]^{-1},$$

and

$$n\bar{D}_F(y_1^{1/n},y_2^{1/n})=n\left[\frac{1}{1-y_1^{1/n}}+\frac{1}{1-y_2^{1/n}}-1\right]^{-1}$$

$$\to-\frac{(\log y_1)(\log y_2)}{\log y_1+\log y_2},\ \text{as } n\to\infty.$$

Thus, by Proposition 2.2 we have $F\in D(H)$, where

$$D_H(y_1,y_2)=y_1\cdot y_2 \exp\left[-\frac{(\log y_1)(\log y_2)}{\log y_1+\log y_2}\right],$$

$$H(x_1,x_2)=\Lambda(x_1)\Lambda(x_2)\exp\{1/(e^{x_1}+e^{x_2})\},$$

and $\Lambda(x)=\exp(-e^{-x})$.

Proposition 2.3 *Let F and G be k-dimensional distributions and let H be a multivariate extreme value distribution.*

1) If $F,G\in D(H)$, then

$$\lim_{t\downarrow 0} \frac{1-D_F(\mathbf{y}^t)}{1-D_G(\mathbf{y}^t)}=1 \text{ for all } \mathbf{y}\in(0,1)^k.$$

2) If $F\in D(H)$, $G_i\in D(H_i)$, $i=1,...,k$, and

$$\lim_{\mathbf{y}\uparrow\mathbf{1}} \frac{1-D_F(\mathbf{y})}{1-D_G(\mathbf{y})}=1, \text{ where } \mathbf{1}=(1,...,1),$$

then $G\in D(H)$.

3. Marginally Independent or Perfect Dependent Multivariate Extreme Value Distributions

Let H be a multivariate extreme value distribution with univariate marginals H_i, $i=1,...,k$. Let $H_*(\boldsymbol{x})=H_1(x_1)\cdots H_k(x_k)$ and $H^*(\boldsymbol{x})=\min\{H_i(x_i), i=1,...,k\}$, then it holds

$$H_*(\boldsymbol{x})\le H(\boldsymbol{x})\le H^*(\boldsymbol{x})$$

for all $\boldsymbol{x}\in R^k$. Both bounds, H_* and H^*, are multivariate extreme value distributions. Characterizations of these distributions are obtained by Takahashi [5].

In the bivariate case Sibuya [3] obtains necessary and sufficient conditions for $F\in D(H_*)$ and $F\in D(H^*)$. In this section we generalize his results.

Proposition 3.1 *Let F be a k-dimensional distribution and let H_i be a univariate extreme value distribution, $i=1,...,k$. Then the following statements are equivalent:*

1) $F\in D(H_*)$.

2) $F_i\in D(H_i)$, $i=1,...,k$, and there exists $\boldsymbol{y}\in(0,1)^k$ such that

$$\lim_{n\to\infty}(D_F(\boldsymbol{y}^{1/n}))^n=y_1y_2\cdots y_k.$$

3) $F_i\in D(H_i)$, $i=1,...,k$, and

$$\lim_{y\uparrow 1}\frac{1-D_F(y\mathbf{1})}{1-y}=k.$$

4) $F_i\in D(H_i)$, $i=1,...,k$, and

$$\lim_{y\uparrow 1}\frac{1-D_F(y\mathbf{1})}{1-y^k}=1.$$

Proof. The proof is straightforward from Theorems 2.2 and 4.1 and Corollary 2.4 of Takahashi [6]. □

Remark. If $k=2$, we have the same result as Proposition 3.1 by Corollary 2.2 of Takahashi [6].

Example 3.1 (See Example 5.2.3 of Galambos [1].) For the Morgenstern distribution

$$F(x_1,x_2)=1-e^{-x_1}-e^{-x_2}+e^{-x_1-x_2}[1+\alpha(1-e^{-x_1})(1-e^{-x_2})],$$

$$D_F(y_1,y_2)=y_1\cdot y_2[1+\alpha(1-y_1)(1-y_2)]$$

where $-1\le\alpha\le 1$, and

$$\lim_{y\uparrow 1}\frac{1-D_F(y,y)}{1-y^2}=1.$$

By Proposition 3.1 4) we have $F\in D(H_*)$, where $H_*(\cdot,\cdot)=\Lambda(\cdot)\Lambda(\cdot)$.

Proposition 3.2 *Let F be a k-dimensional distribution and let H_i be a univariate extreme value distribution, $i=1,...,k$. Then the following statements are equivalent:*

1) $F\in D(H^*)$.

2) $F_i\in D(H_i)$, $i=1,...,k$, and there exists $y\in(0,1)$ such that

$$\lim_{n\to\infty}(D_F(y^{1/n}\mathbf{1}))^n=y.$$

3) $F_i\in D(H_i)$, $i=1,...,k$, and

$$\lim_{y\uparrow 1}\frac{1-D_F(y\mathbf{1})}{1-y}=1.$$

Proof. The proof is straightforward from Theorem 3.1 and Corollary 3.1 of Takahashi [6]. □

4. Joint Asymptotic Distribution of the Multivariate Extreme Statistics

In this section, we show the joint asymptotic distribution of several multivariate extreme statistics along the arguments in Sec. 2.3 of Leadbetter et al. [7]. For simplicity we shall consider the bivariate case.

Let $(X_1,Y_1),...,(X_n,Y_n)$ be a sequence of independent random vectors with common distribution F. The order statistics of the components will be denoted by

$$X_{1:n}\le X_{2:n}\le\cdots\le X_{n:n};\ \text{and}\ Y_{1:n}\le Y_{2:n}\le\cdots\le Y_{n:n}.$$

For $i=0,1,...,r-1$, define

$$\boldsymbol{Z}_{n-i}=(X_{n-i:n},Y_{n-i:n})$$

and let us call $\boldsymbol{Z}_{n-i}$ an $(i+1)$-th multivariate extreme statistic.

Proposition 4.1 *Suppose that*

$$P\{(\boldsymbol{Z}_n-\boldsymbol{b}_n)/\boldsymbol{a}_n\le\boldsymbol{x}\}\xrightarrow{w}H(\boldsymbol{x})$$

for some nondegenerate distribution H. Then, for $\boldsymbol{x}_1=(x_1,y_1)>\boldsymbol{x}_2=(x_2,y_2)$,

$$P\{(\boldsymbol{Z}_n-\boldsymbol{b}_n)/\boldsymbol{a}_n\le\boldsymbol{x}_1,\ (\boldsymbol{Z}_{n-1}-\boldsymbol{b}_n)/\boldsymbol{a}_n\le\boldsymbol{x}_2\}\xrightarrow{w}H_J(\boldsymbol{x}_1,\boldsymbol{x}_2)$$

where

$$\begin{aligned}H_J(\boldsymbol{x}_1,\boldsymbol{x}_2)=H(\boldsymbol{x}_2)\{&1+\log H(\boldsymbol{x}_1)-\log H(\boldsymbol{x}_2)\\&+[\log H_1(x_1)-\log H_1(x_2)+(h(x_1,y_2)-h(\boldsymbol{x}_2))]\\&\times[\log H_2(y_1)-\log H_2(y_2)+(h(x_2,y_1)-h(\boldsymbol{x}_2))]\}\end{aligned}$$

and $h(x)=\lim_{n\to\infty}\overline{F}(a_n x+b_n)$.

Proof. Define

$$S_0^n=\#\{j \mid X_j>a_{1n}x_1+b_{1n} \text{ or } Y_j>a_{2n}y_1+b_{2n}, j=1,2,\ldots,n.\},$$
$$S_1^n=\#\{j \mid a_{1n}x_2+b_{1n}<X_j\le a_{1n}x_1+b_{1n} \text{ and } Y_j\le a_{2n}y_2+b_{2n}, j=1,2,\ldots,n.\},$$
$$S_2^n=\#\{j \mid X_j\le a_{1n}x_2+b_{1n} \text{ and } a_{2n}y_2+b_{2n}<Y_j\le a_{2n}y_1+b_{2n}, j=1,2,\ldots,n.\},$$
$$S_{12}^n=\#\{j \mid a_n x_2+b_n<(X_j,Y_j)\le a_n x_1+b_n, j=1,2,\ldots,n.\}$$

then, we have

$$P\{(Z_n-b_n)/a_n\le x_1, (Z_{n-1}-b_n)/a_n\le x_2\}$$
$$=P\{S_0^n=0, S_{12}^n=0, S_1^n\le 1, S_2^n\le 1\}$$
$$+P\{S_0^n=0, S_{12}^n=1, S_1^n=0, S_2^n=0\}.$$

On the other hand, by using Theorem 5.3.1 of Galambos [1], we can evaluate the asymptotic probabilities of the evens

$$\{S_0^n=i, S_{12}^n=j, S_1^n=k, S_2^n=m\}$$

for $i,j,k,m=0,1$. Thus we have the result. □

Corollary 4.1 *Suppose that*

$$P\{(Z_n-b_n)/a_n\le x\}\xrightarrow{w}H(x)$$

for some nondegenerate distribution H. Then, for fixed $r\ge 1$ *and* $x_1>\cdots>x_r$

$$\left|P\{(Z_n-b_n)/a_n\le x_1,\ldots,(Z_{n-r+1}-b_n)/a_n\le x_r\}\right.$$
$$\left.-P\{(Z_n^*-\beta_n)/\alpha_n\le x_1,\ldots,(Z_{n-r+1}^*-\beta_n)/\alpha_n\le x_r\}\right|\to 0,$$
$$\text{as } n\to\infty,$$

where Z_{n-i}^* *is the* $(i+1)$*-th multivariate extreme statistic from the distribution H,* $i=0,\ldots, r-1$, *and* $H^n(\alpha_n x+\beta_n)=H(x)$, $n=1,2,\ldots$

Example 4.1 Let F be the bivariate normal distribution with the correlation coefficient less than one. Then

$$\left|P\{(Z_n-b_n)/a_n\le x_1,\ldots,(Z_{n-r+1}-b_n)/a_n\le x_r\}\right.$$
$$\left.-P\{(Z_n^*-(\log n)\mathbf{1}\le x_1,\ldots,(Z_{n-r+1}^*-(\log n)\mathbf{1}\le x_r\}\right|\to 0,$$
$$\text{as } n\to\infty,$$

where Z_{n-i}^* is the $(i+1)$-th multivariate extreme statistic from the bivariate exponential distribution whose marginals are equal to the standard exponential distribution and they are independent. For the univariate case, it is a well known result (see Weissman [8], Theorem 3).

Acknowledgments

The author thanks Prof. M. Sibuya and Prof. J. Galambos for their advice.

5. References

[1] J. Galambos, The Asymptotic Theory of Extreme Order Statistics. 2nd ed., Krieger, Malabar, FL (1987).
[2] S. I. Resnick, Extreme Values, Regular Variation and Point Processes. Springer-Verlag, Berlin (1987).
[3] M. Sibuya, Bivariate extreme statistics. Ann. Inst. Statist. Math. **11**, 195–210 (1960).
[4] A. W. Marshall and I. Olkin, Domains of attraction of multivariate extreme value distributions. Ann. Probab. **11**, 168–177 (1983).
[5] R. Takahashi, Characterizations of a multivariate extreme value distribution. Adv. Appl. Probab. **20**, 235–236 (1988).
[6] R. Takahashi, Asymptotic independence and perfect dependence of vector components of multivariate extreme statistics. Statist. Probab. Letters **19**, 19–26 (1994).
[7] M. R. Leadbetter, G. Lindgren, and H. Rootzén, Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, New York (1983).
[8] I. Weissman, Estimation of parameters and large quantiles based on the k largest observations. J. Amer. Statist. Assoc. **73**, 812–815 (1978).

About the author: *Rinya Takahashi is Associate Professor, Department of Information Systems Engineering, Kobe University of Mercantile Marine, Fukaeminami 5-1-1, Higashinada, Kobe, 658, Japan.*

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 555 (1994)]

The Aggregate Excess Measure of Severity of Extreme Events

Volume 99 Number 4 July–August 1994

Clive W. Anderson

Department of Probability and Statistics,
University of Sheffield,
Sheffield, UK

It is suggested here that in many environmental and other contexts the severity of an extreme event might usefully be represented by the sum of the excesses of a measured variable over a high threshold. The general form of the limiting distributions of such sums for a wide class of models has been derived by Anderson and Dancy, and has suggested methods for the statistical analysis of data concerning extreme severity. This work is reviewed here, and some extensions to the distributional theory are presented. An application of the methods to atmospheric ozone levels, which calls for the extension of the approach to take account of covariate information is reported.

Key words: aggregate excess; extreme event; flood; generalized Pareto distribution; ozone; point process; severity; storm; threshold model; Weibull distribution.

Accepted: March 22, 1994

1. Introduction

The severity of a storm or a flood is often a function not only of the peak value of whichever environmental variable is concerned, but also of other aspects of the extreme event, such as its duration and temporal shape. An extended run of days with temperatures just below freezing, for example, can be more disruptive to everyday human activity and to animal and plant life than a single day with a much sharper frost. Similarly, sustained moderately high water levels in a river or the sea can lead to greater flooding than a more extreme level lasting for only a short time. To attempt to analyze such examples in a way which captures the notion of severity implicit in them demands an extension of traditional statistical methods for extremes, which have tended to concentrate largely on the modelling of maxima or storm peaks. In Ref. [1] it was suggested that for an important class of applications a simple way to quantify the idea of severity is in terms of the sum of the excesses of the environmental variable over a high threshold during the extreme event. In the case of a flood, for example, this sum or aggregate excess is a discrete approximation to the total volume of water overtopping the threshold, and in the case of temperatures the analogous quantity defined for low values, the aggregate deficit, is a measure of exposure or cumulative damage. In the earlier paper some distribution theory was developed for aggregate excesses, and an application to flood data was discussed. Here I review that work and present some extensions of its distributional results, and discuss a new application to ozone concentrations.

2. Preliminaries

The techniques to be described are related to threshold methods for extremes [2], and the distributional results are formulated in terms of the Mori-Hsing point process representation [3, 4] for the structure of high values of a stationary sequence. We briefly recall ideas from these two areas.

Suppose $\{X_j\}$ denotes a sequence of observations, and let u be a high threshold. Times j at which $X_j > u$ are referred to as *exceedances* of u by $\{X_j\}$, and the sizes of overshoots $X_j - u$ at exceedances are called *excesses* over the level u. In environmental applications exceedances are often found to occur in clusters corresponding to physical storms. Threshold methods are based on the modelling of the peak excess within each cluster by a generalized Pareto distribution, with distribution function of the form

$$G(x:\xi\sigma) = 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi}, \qquad (1)$$

where $\sigma > 0$ is a scale parameter, $\xi (-\infty < \xi < \infty)$ is a shape parameter, and the range of x is such that $\xi x/\sigma > -1$.

Let N denote the number of exceedances within a cluster, and suppose that $\zeta_1 \geq \ldots \geq \zeta_N$ are the corresponding excesses. Then the suggestion above is that the aggregate excess within a cluster

$$S = \sum_1^N \zeta_j$$

is for some purposes a reasonable measure of the severity of a storm event. For statistical modelling we are interested in the distribution of S, particularly for high thresholds u. Since $S \geq \zeta_1$, we expect S to have (in the limit as u increases) a tail no lighter than that of the limiting generalized Pareto distribution of ζ_1. The distribution of S is also expected to reflect the cluster size and the pattern of dependence between individual excesses ζ_j.

Suppose now that $M_n = \max_{1 \leq i \leq n} X_i$. It is known that for many $\{X_i\}$ sequences M_n may be normalized to converge in distribution to some nondegenerate limit. Suppose in fact that there is a continuous and strictly decreasing function $u_n(\tau)$ such that, for each $\tau > 0$,

$$\lim_{n\to\infty} P(M_n \leq u_n(\tau)) = e^{-\tau}. \qquad (2)$$

Let u_n^{-1} denote the inverse function of u_n. Consider now the two-dimensional point process with points $(j/n, u_n^{-1}(X_j))$. In Ref. [4], which generalizes Ref. [3], it is proved under a weak long-range mixing condition Δ that if this point process converges as $n \to \infty$ then its limit has points of the form $(S_i, T_i Y_{ij})$, $i \geq 1$, $1 \leq j \leq K_i$, where (S_i, T_i), $i \geq 1$ are the points of a unit Poisson process in $\Re^{+2}$, and for each i, $\{Y_{ij} : j = 1, \ldots, K_i\}$ with $Y_{i1} \equiv 1$, is a point process on $[1, \infty)$ with a random number K_i of points. Moreover the processes $\{Y_{ij} : j = 1, \ldots, K_i\}$ for each i are independent of each other and of the $\{(S_i, T_i)\}$ process, and are identically distributed.

A natural interpretation of this convergence result is that large values of the $\{X_i\}$ sequence occur in clusters, located in time at the points of a simple Poisson process, and that values within a cluster (from the peak downwards) are given, after transformation, by $T_i, T_i Y_{i2}, \ldots$ respectively (reading upwards). Note that, since the transformation is decreasing, a cluster *peak* corresponds to the *lower* endpoint of a vertical string of points in the limiting point process.

In what follows it will be convenient to suppose that the point process associated with each cluster contains infinitely many points Y_{ij} arranged in increasing order of size

$$1 \equiv Y_{i1} \leq Y_{i2} \leq \ldots$$

but that infinite values of the Y_{ij} are allowed after the first point, so that K_i, the number of points in a cluster, is just the index of the last finite Y_{ij}. By this means stochastic properties of K_i are subsumed notationally in those of $\{Y_{ij}\}$.

We are interested in particular in clusters of exceedances by $\{X_j\}$ of a high threshold u. Let $v = u_n^{-1}(u)$. Then $X_j > u$ is equivalent to $u_n^{-1}(X_j) < v$, and so, in the limit, clusters of exceedances of u correspond exactly to those clusters in the point process for which $T_i < v$. Given that we are dealing with such a cluster (as we assume from now on) it follows from the unit Poisson nature of $\{(S_i, T_i)\}$ that T_i is uniformly distributed over $(0, v)$.

For many $\{X_j\}$ the transformation u_n is related in a simple way to the marginal distribution function, F say, of X_j. Suppose in fact that $\{X_j\}$, still satisfying condition Δ, has a positive extremal index θ. Then ([5], Theorem 3.7.2)

$$\lim_{n\to\infty} P(M_n \leq u_n(\tau)) = \lim_{n\to\infty} F^{n\theta}(u_n(\tau)). \qquad (3)$$

Hence, if the tail function $1 - F$ of F is denoted by $\mathcal{F}$, it follows from Eq. (2) that

$$n\theta \mathcal{F}(u_n(\tau)) \sim -n\theta \log F(u_n(\tau)) \sim \tau,$$

for large n. We may therefore define u_n by

$$u_n(\tau) = \mathcal{F}^{-1}(\tau/n\theta). \qquad (4)$$

In particular therefore

$$n\theta\,\mathcal{F}(u)=v,$$

and so the excesses within a cluster, in decreasing order of size, are in the limit (dropping the cluster index i, no longer relevant)

$$\begin{aligned}\zeta_j &= u_n(TY_j)-u\\ &= \mathcal{F}^{-1}(TY_j\,\mathcal{F}(u)/v)-u\\ &= \mathcal{F}^{-1}(T'Y_j\,\mathcal{F}(v))-u,\\ j &= 1, 2, \ldots, N \end{aligned} \tag{5}$$

where $T'=T/v$ is uniformly distributed over $(0,1)$. The aggregate excess for the cluster is

$$S=\sum_1^N \zeta_j,$$

where N, the number of exceedances in the cluster, is

$$N=\max\{j: T'Y_j<1\},$$

and T' is independent of the Y_j process.

3. Asymptotic Distributions of Aggregate Excess

In this section we outline various asymptotic distributional properties of aggregate excesses which follow from the preceding discussion. The asymptotic distribution of aggregate excess S itself turns out to depend on the Y process partly through random sums

$$R_j=\sum_{i=1}^{j} Z_{ij}$$

where the Z_{ij} are defined in terms of $\{Y_i\}$ by

$$Z_{ij}=\begin{cases}\left(\frac{Y_{j+1}}{Y_i}\right)^{\xi}-1 & \text{for } \xi>0\\ \log\left(\frac{Y_{j+1}}{Y_i}\right) & \text{for } \xi=0\\ 1-\left(\frac{Y_{j+1}}{Y_i}\right)^{\xi} & \text{for } \xi<0.\end{cases} \tag{6}$$

3.1 Limit Distributions of S

Suppose that the stationary sequence $\{X_j\}$ satisfies Hsing's mixing condition Δ and has positive extremal index, and that the marginal distribution F of the X_j is such that the limiting distribution of peak excesses within a cluster is generalized Pareto with shape parameter ξ. Suppose too that the corresponding point process $\{(j/n, u_n^{-1}(X_j))\}$ converges to a limiting process with the structure described in Sec. 2. Then, as the threshold level u tends to the upper end point, x_+ say, of the support of X,

$$\lim_{u\to x_+} P\left(\frac{S}{\gamma_\xi(u)}>s\right)=$$

$$\begin{cases} E\left[\left(\frac{\sum_1^{j_s} Y_i^{-\xi}}{s+j_s}\right)^{1/\xi}\right] & \text{for } \xi>0\\ E\left[\exp\left(-\frac{s+\sum_1^{j_s}\log Y_i}{j_s}\right)\right] & \text{for } \xi=0\\ E\left[\left(\frac{\sum_1^{j_s} Y_i^{-\xi}}{j_s-s}\right)_+^{1/\xi}\right] & \text{for } \xi<0\end{cases} \tag{7}$$

where

$$\gamma_\xi(u)=\begin{cases} u & \text{for } \xi>0\\ l(1/\mathcal{F}(u)) & \text{for } \xi=0\\ x_+-u & \text{for } \xi<0\end{cases} \tag{8}$$

for a suitable slowly varying function l, and

$$j_s=\min\{j: Y_{j+1}=\infty \text{ or } R_j\geq s\}. \tag{9}$$

Expectations in Eq. (7) are taken with respect to the point process $\{Y_j\}$.

This result is a consolidation and re-statement of the main limit forms found in Ref. [1]. The proof—essentially a weak convergence argument based on the Mori-Hsing process—exploits regular and slow variation properties of $\mathcal{F}$ implied by the assumption that cluster peak excesses are, in the limit, generalized Pareto distributed. For example, when $\xi=0$,

$\mathcal{F}$ belongs to the domain of attraction of the Gumbel extreme value distribution, so that, as $x \to \infty$,

$$\mathcal{F}^{-1}(e^{-w-x}) - \mathcal{F}^{-1}(e^{-x}) \sim wl(e^{x})$$

for each $w > 0$, for some slowly-varying function l (see, for example, Ref. [6], Sec. 8.13). Thus

$$\zeta_j \sim \mathcal{F}^{-1}(T'Y_j\mathcal{F}(u)) - \mathcal{F}^{-1}(\mathcal{F}(u)) \sim$$

$$(-\log(T'Y_j))l(1/\mathcal{F}(u)),$$

as $u \to x_+$, which establishes the connection between the limiting behaviour of S and the Y-process.

We note that Eq. (7) reveals in reasonably explicit form the dependence of the distribution of S on the number and pattern of excesses within a cluster.

3.2 Joint Limit Distributions

The techniques used to obtain these results may be extended to give limiting distributions for other quantities. As an example (motivated by a question from a reservoir engineer about peak water level and total overtopping discharge at a dam wall) the joint distribution of peak and aggregate excesses is as follows.

Under the same assumptions as in Sec. 3.1, and with the same notation:

$$\lim_{u \to x_+} P\left(\frac{S}{\gamma_\xi(u)} > s, \frac{\zeta_1}{\gamma_\xi(u)} > z\right) =$$

$$\begin{cases} E\left[\min\left\{\left(\frac{\sum_1^{j_s} Y_i^{-\xi}}{s + j_s}\right)^{1/\xi}, (1+z)^{-1/\xi}\right\}\right] & \text{for } \xi > 0 \\ E\left[\min\left\{\exp\left(-\frac{s + \sum_1^{j_s} \log Y_i}{j_s}\right), e^{-z}\right\}\right] & \text{for } \xi = 0 \\ E\left[\min\left\{\left(\frac{\sum_1^{j_s} Y_i^{-\xi}}{j_s - s}\right)_+^{1/\xi}, (1-z)_+^{-1/\xi}\right\}\right] & \text{for } \xi < 0 \end{cases} \tag{10}$$

Similar joint limiting distributions may also be found for ζ_1 and $S - \zeta_1$. Like Eq. (10) they are singular. Methods of statistical analysis based on them have yet to be explored.

3.3 More Explicit Forms for $P(S/\gamma_\xi(u) > s)$

When specific models are assumed for the X process the limiting distributions Eq. (7) take on more explicit forms. Several examples were studied in Ref. [1]. Writing

$$\lim_{u \to x_+} P\left(\frac{S}{\gamma_\xi(u)} > s\right) = \begin{cases} (1 + \text{sign}(\xi)V(s,\xi))^{-1/\xi} & \text{for } \xi \neq 0 \\ \exp(-V(s,0)) & \text{for } \xi = 0 \end{cases} \tag{11}$$

it was found that $V(s,\xi)$ had the same general form in all cases considered: that of a concave increasing function of s dominated by s when $\xi \geq 0$, and by min $\{1, s\}$ when $\xi < 0$. See Fig. 1.

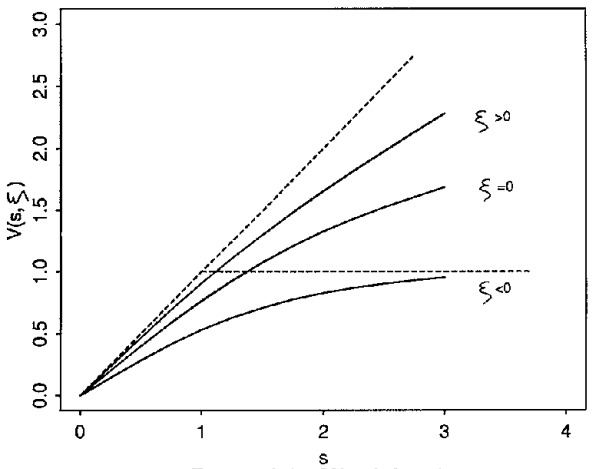


Fig. 1. Forms of the $V(s, \xi)$ function.

The findings and examples above motivate an attempt to fit aggregate excess data by a distribution with tail function of this general form. Two such attempts are described in Sec. 4.

3.4 Higher Thresholds

As often in extreme value Statistics, an aim in many applications will be extrapolation to longer time periods or higher levels than seen in data. In particular, for aggregate excesses, extrapolations to higher *thresholds* will often be of interest. For example, in flood applications knowledge of the aggregate excess above a higher threshold might be vital in estimating the reduction in the size of floods that would result from improved river or sea defences. The following presents a simple relationship on which extrapolation of aggregate excesses could be based.

Suppose that S_u and $S_{u'}$ denote aggregate excesses above levels u and u', respectively, with $u < u'$, in a cluster in which level u is exceeded (so that S_u, but not necessarily $S_{u'}$, is greater than zero). In a slightly more refined notation than used earlier, the limiting forms in Sec. 3.1 are limits, $\mathcal{H}(s)$ say, of $P(S_u/\gamma_\xi(u) > s | S_u > 0)$ as $u \to x_+$. We are now interested in $P(S_{u'}/\gamma_\xi(u) > s | S_u > 0)$. But

$$P\left(\frac{S_{u'}}{\gamma_\xi(u)} > s | S_u > 0\right) =$$

$$P\left(\frac{S_{u'}}{\gamma_\xi(u')} > s\frac{\gamma_\xi(u)}{\gamma_\xi(u')} | S_{u'} > 0\right)$$

$$P(S_{u'} > 0 | S_u > 0) \approx$$

$$\mathcal{H}\left(s\frac{\gamma_\xi(u)}{\gamma_\xi(u')}\right) P(\zeta_1 > u'), \tag{12}$$

for high u, where ζ_1 is the peak excess in the cluster. Thus the distribution of aggregate excesses with respect to the higher threshold u' has a point probability at 0 corresponding to the event $P(\zeta_1 \le u')$ that no exceedance of u' occurred, together with a form over the strictly positive half-line which is the the same as that of the original distribution of aggregate excesses except for an increased scale parameter. Estimation of this distribution may therefore be based, through Eq. (12) on estimation of $\mathcal{H}$ from data on aggregate excesses of u, and of $P(\zeta_1 > u')$ from data on peak excesses of u fitted to the Generalized Pareto distribution Eq. (1). Relationship Eq. (12) should also be useful as a means of checking the fit of specific models for $\mathcal{H}$, though this aspect has yet to be investigated.

4. Applications

4.1 Floods on the River Thames

In Ref. [1] an application of some of the limiting results above to data on levels of the River Thames is described. The aim was largely exploratory: to see whether there is support in an important data set for a model of the general kind suggested in Secs. 3.1 and 3.3, and, if there is, to seek an appropriate parametric form for the model. The results were surprisingly positive: confirmation was found for the general form of distribution predicted by asymptotic arguments, and in particular a simple Weibull distribution with

$$P(S > s) = \exp(-\alpha s)^\phi), \tag{13}$$

for some parameters $\alpha > 0$ and ϕ was found to give an acceptable fit to data on S.

4.2 Ozone Concentrations

An analysis of a further set of data, which calls for the extension of the models above to take account of covariate information, is now reported.

The data consist of hourly mean ozone concentrations at a suburban site in Stevenage, about 25 miles north of London, over the years 1978--1989. High levels of ozone are known to cause direct damage to vegetation (see, for example, Ref. [7]). One tentative suggestion is that a plant or tree suffers damage in proportion to cumulative exposure to ozone at concentrations above some threshold. The threshold is not known, and indeed is likely to be different for different plants, but a figure in the range 40 ppb–90 ppb might be plausible. Though this theory is at present no more than a working hypothesis, it prompts an interest in the occurrence of high values of aggregate excesses of ozone concentrations above moderately high thresholds. The analysis summarized below is a preliminary investigation into the possibility of using the aggregate excess models of Sec. 3 to describe such high doses. A more complete account of the biological background, and of the application of the method to spatial variation of exposure over the UK, is given in Ref. [8].

For the theory of Secs. 2 and 3 to be applicable it is desirable that we work with independent clusters of high values. The hourly data were therefore subjected to a preliminary declustering procedure, which selected episodes when concentrations above a specified 'declustering threshold' were experienced, and ensured that such episodes were separated far enough in time to give some plausibility to the independence assumption. Figure 2 shows a time plot of the resulting aggregate excesses above a threshold of 60 ppb, obtained with a declustering threshold of 50 ppb and with a time separation between clusters of at least 48 hours—these values being chosen as typical of those of possible scientific interest. An immediate observation from the plot is that the assumption of stationarity is suspect: the middle years 1982–1986 contain some values higher than seen earlier or later. (There are known diurnal

patterns in ozone concentrations too, but they are of too short a duration to affect the present analysis.) In view of the apparent nonstationarity a simple model of the kind found useful in the earlier analysis would not on its own be expected to be particularly successful here: and indeed the Weibull model Eq. (13) fitted to aggregate excesses above 60 ppb appears to underestimate the sizes of the highest aggregates.

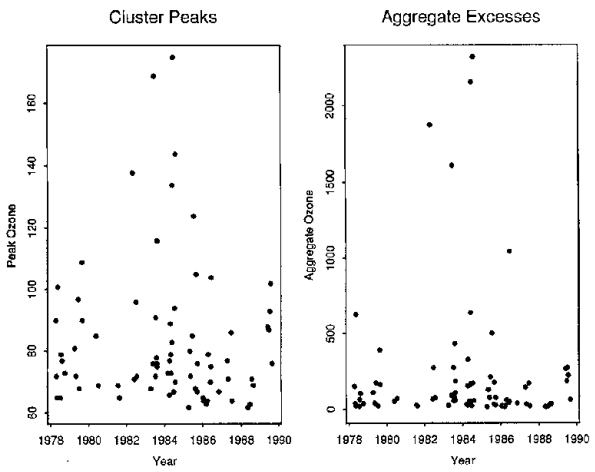


Fig. 2. Hourly mean ozone concentrations over 60 ppb: Stevenage 1978–1989.

The processes leading to the formation of ozone in the atmosphere are photo-chemical – driven by strong sunlight. It is possible therefore that unusual weather conditions in the early to mid 1980s may have had some bearing on the possible inhomogeneity. Unfortunately sunlight was not recorded at the Stevenage monitoring site, nor was temperature, which is a crude surrogate for it. Temperature data were not readily obtainable either from nearby meteorological stations, but were to hand for Sheffield, 140 miles north. Figure 3, showing aggregate excesses over 60 ppb against monthly averages of daily maximum Sheffield temperatures, illustrates that in spite of the geographical separation there is nevertheless some connection. It appears that the summers over the relevant years contained some quite warm spells, presumably experienced in Sheffield as well as Stevenage. Accordingly Weibull models which incorporate temperature t as a covariate were fitted. Two forms were used:

$$P(S>s)=\exp-(s/\delta(t))^{\phi}, \qquad (14)$$

in which the scale parameter δ depends on t in the form $\delta(t)=\delta e^{\beta t}$; and secondly a model suggested by the evidence from Fig. 3 that not all occurrences of high temperatures t at the time of an ozone cluster are necessarily associated with a high aggregate ozone dose. This suggests a model in which ozone clusters are assumed to be of two types, the first showing temperature dependence of the kind above, and the second showing no dependence on temperature. Thus

$$P(S>s)=\begin{cases}\exp-(s/\delta(t))^{\phi} & \text{for type 1 clusters}\\ \exp-(s/\delta')^{\phi'} & \text{for type 2 clusters}\end{cases} \qquad (15)$$

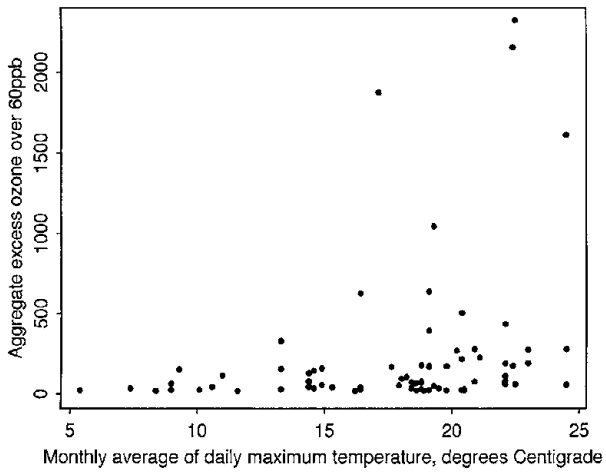


Fig. 3. Aggregate excess ozone over 60 ppb vs temperature.

(Since sunlight/high temperature is at best only one of the preconditions known to be necessary for the formation of ozone, there is some general scientific justification for a model of this form.) In fitting, clusters with aggregate excesses above a specified level were taken to be of type 1. A likelihood ratio test shows that model Eq. (15) represents a very worthwhile improvement over Eq. (14) even after allowing for the inclusion of two extra parameters ($W=27.26$, $p<10^{-4}$, cut-off level for type 1 = 500). Q-Q type plots for the two covariate models are shown in Figs. 4 and 5 respectively. (These are constructed as follows: under model Eq. (14) $S/(\delta e^{\beta t})$ reduces to a standard Weibull variable with unit scale parameter and shape parameter ϕ: $P(S/(\delta e^{\beta t})>s)=\exp(-s^{\phi})$. Thus a plot of the ordered values of $S/(\delta e^{\beta t})$ from a sample of size n against $[-\log(i/n+1)]^{1/\phi}$ should yield an approximate line of unit slope. Figure 4 is a plot of this kind, and Fig. 5 is constructed similarly from model Eq. (15).)

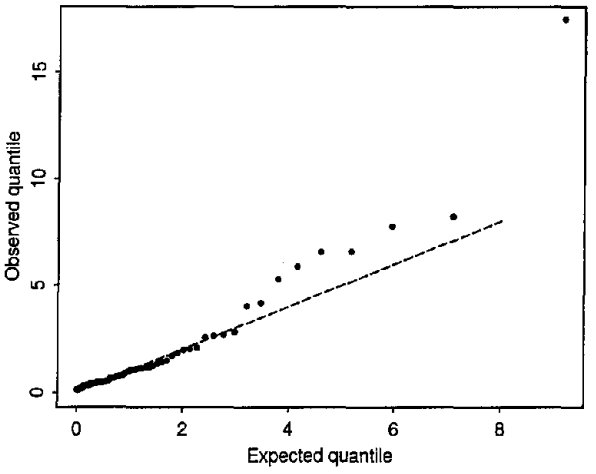


Fig. 4. Q-Q plot for aggregate excess ozone: simple covariate model Eq. (14).

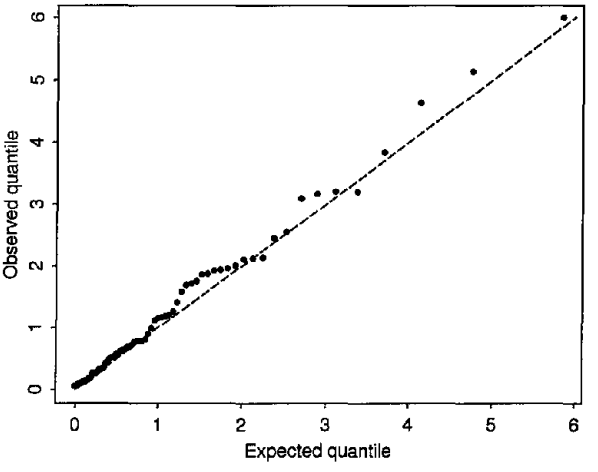


Fig. 5. Q-Q plot for aggregate excess ozone: two-type covariate model Eq. (15).

Both plots appear to show a quite good fit to the Weibull model after allowing for dependence on temperature, model Eq. (15) doing a little better than model Eq. (14). Further refinements of the models allowing temperature-dependence also of the shape parameter ϕ gave no worthwhile improvement in fit as judged by a likelihood test.

Though this is only a preliminary analysis (which we hope to complete with better temperature data), the results so far are encouraging. They appear to show again that models of the form suggested in Sec. 3.3, and in particular a Weibull model—after allowance in this case for nonstationarity—can represent aggregate excess data reasonably well. If this is confirmed, then for example these models will be useful in estimating return levels of future high doses of ozone above 60 ppb or, following the results of Sec. 3.4, above higher thresholds.

Acknowledgment

The numerical work reported here used equipment funded by SERC under the Complex Stochastic Systems Initiative.

5. References

[1] C. W. Anderson and G. P. Dancy, The severity of extreme events, Research Report 92/593 Department of Probability and Statistics, University of Sheffield (1992) 24 pp.

[2] A. C. Davison and R. L. Smith, Models for exceedances over high thresholds (with discussion), J. Roy. Statist. Soc. B **52**, 393–442 (1990).

[3] T. Mori, Limit distributions of two-dimensional point processes generated by strong-mixing sequences, Yokohama Math. J. **25**, 155–168 (1977).

[4] T. Hsing, On the characterization of certain point processes, Stoch. Processes Appl. **26**, 297–316 (1987).

[5] M. R. Leadbetter, G. Lindgren, and H. Rootzén, Extremes and Related Properties of Random Sequences and Processes, Springer, New York (1983).

[6] N. H. Bingham, C. M. Goldie, and J. L. Teugels, Regular Variation, Cambridge University Press, Cambridge (1987).

[7] UK Photochemical Oxidants Review Group, Ozone in the United Kingdom. Department of the Environment, London (1987).

[8] R. I. Smith, C. W. Anderson, and D. Fowler, Critical levels of ozone over the United Kingdom: mapping aggregate exceedances over moderate to high levels, J. Res. Natl. Inst. Stand. Technol. **99**, 353 (1994).

About the author: Clive W. Anderson is a statistician in the Probability and Statistics Section of the School of Mathematics and Statistics, University of Sheffield, UK.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

[J. Res. Natl. Inst. Stand. Technol. 99, 563 (1994)]

The Measurement of Averages and Extremes of Environmental Variables

Volume 99 Number 4 July–August 1994

C. W. Anderson
University of Sheffield,
Sheffield, UK

and

K. F. Turkman

University of Lisbon,
Lisbon, Portugal

The relationship between the statistics of environmental measurements averaged over different time scales is related to extreme levels of the variables. Results on the asymptotic joint distributions of extreme averages over different time periods are treated.

Key words: averages; extremes; extremal index; Poisson random measures.

Accepted: March 22, 1994

1. Introduction

Measurements of many environmental variables such as wind speeds, rainfall and the concentrations of atmospheric and aquatic pollutants are generally duration-specific: the actual quantities measured are averages over a specific time interval rather than instantaneous values. Thus concentrations of ozone are typically measured as parts per 100 million averaged over an hour, and wind speeds are routinely recorded as hourly or daily mean speeds. In practice however, the scientists wishing to understand the environmental processes which lie behind such measurements, and the regulatory body which monitors pollution, often wish to deal with characteristics measured over some other time interval: peak concentrations over a day, for example, or high wind gust values, which in practice correspond to peak 3–5 s averages. There is therefore a need to understand the relationship between the statistics of environmental measurements averaged over different time scales, and to relate these statistics to extreme levels of the variables. In this paper, results on the asymptotic joint distributions of extreme averages over different time periods will be treated. These results will make it possible, for example, to link long historical data series containing information about extremes of daily rainfall (sometimes extending back to the early years of the century) to the shorter series of extreme hourly rainfall which have been recorded only in the past 20 or 30 years. Thus important historical information could properly be taken account of in the estimation of floods, something recognised as higly desirable by hydrologists. Another area of application of the results is in the study of the dispersal of airborne pollutants. Here, it is known (see, for example Fakrell and Robins [9]) that instruments used to measure the concentration of pollutants dispersing in a turbulent flow cannot resolve the finest scales present in such flows. Measurements of concentration are therefore invariably obtained only as averages of the characteristics of primary physical interest, and so a statistical theory which links extremes of averages over different ranges would be of great value

to scientists working in the area (Mole [10]). One related area is the study of joint distributions of averages and maxima of random sequences. Interest in such distributions is motivated by analysis of extreme winds. For purposes of building design or public safety, it is often important to estimate the speed of the most extreme wind likely to occur at a particular location over a period of years and to do this it is natural to apply the methods of extreme value theory to data on maximum gusts. The precision of estimates obtained may be low due to the limited amount of relevant data—often no more than 10 or 20 years, so it is desirable to try to improve the precision by introducing into the estimation procedure other information relevant to extreme winds. It is natural, thefore, to ask whether the data on gusts could be augmented by that on hourly means, as gusts and means are evidently related. One source of guidance here may be provided by the limit properties of the joint distribution of means and the maxima.

In Sec. 2, we give a summary of the results that exist on the joint limiting forms of sums and maxima of stationary sequences and in Sec. 3 we give some results on the asymptotic joint distributions of extreme averages over different time-periods of sequences which have moving average representations. Possible solutions for the general stationary case are also indicated.

2. Extremes and Averages

Let $\{X_i\}$ be a stationary sequence of random variables with marginal distribution function $P(X_i \le x) = F(x)$ and let

$$S_n = \sum_{i=1}^{n} X_i,\ M_n = \max_{1 \le i \le n} X_n\ ,\ n = 1, 2, \ldots$$

We study here the joint limiting distributions of

$$(\tilde{S}_n, \tilde{M}_n) = \left(\frac{S_n - nb_n}{a_n}, \frac{M_n - d_n}{c_n} \right) \tag{1}$$

as $n \to \infty$ for suitable constants $a_n > 0, c_n > 0, b_n$ and d_n.

Case 1: Light Tailed Case

Assume that $Var(X_i) < \infty$ and $F \in D(\Lambda)$ or $F \in D(\Phi_\alpha)$, $\alpha > 2$ or $F \in D(\Psi_\alpha)$, $\alpha > 0$, where $\Lambda(.)$, $\Phi_\alpha(.)$, $\Psi_\alpha(.)$ are respectively the Gumbel, Frechet and Weibull distributions.

Then for the associated iid sequence (Chow and Teugels [5])

$$(\tilde{S}_n, \tilde{M}_n) \to^d (N, \Lambda),$$

or

$$(\tilde{S}_n, \tilde{M}_n) \to^d (N, \Phi_\alpha),\ \alpha > 2,$$

$$(\tilde{S}_n, \tilde{M}_n) \to^d (N, \Psi_\alpha),\ \alpha > 0,$$

where the limit components are independent.

Can dependence amongst X_i modify this limiting independence? As the following theorem (Anderson and Turkman [1]) shows, under quite weak conditions, dependence does not affect the limiting distribution.

Theorem 2.1

Assume that $\{X_i\}$ is strong mixing and has positive extremal index and for some a_n, c_n and d_n,

$$\tilde{S}_n = \frac{S_n}{a_n} \to^d N(0,1),$$

$$\tilde{M}_n = \frac{(M_n - d_n)}{c_n} \to^d G,$$

where $G = \Lambda$ or $G = \Phi_\alpha$, for some $\alpha > 2$ or $G = \Psi_\alpha$, for $\alpha > 0$. Assume further that $\{X_i\}$ satisfies the condition

$$\lim_{k \to \infty} k \limsup_{n \to \infty} D'(a_n, u_n) = 0,$$

where

$$D'(a_n, u_n) = \sum_{j=1}^{n/k} = E\left[\left| \exp\left(ita_n^{-1} \sum_{\substack{l=1 \\ l \ne j}} X_l \right) \right. \right.$$

$$\left. \left. - 1 \right| \left[\chi(X_j > u_n) \right] \right. \tag{2}$$

Then $\tilde{S}_n$ and $\tilde{M}_n$ are asymptotically independent.

Local dependence condition $D'(a_n, u_n)$ is quite weak and satisfied, for example, by m-dependent sequences and by Gaussian sequences with summable covariances.

Case 2: Heavy Tailed Case

Assume that

$$1 - F(x) = px^{-\alpha} L(x),$$

$$F(-x) = qx^{-\alpha} L(x)$$

where $L(x)$ is a slowly varying function and $0<\alpha<2$. Then for the associated iid sequence (Chow and Teugels [5])

$$(\tilde{S}_n, \tilde{M}_n) \to^d (U, V),$$

where

$$E(e^{itU}\chi(V\le v)) = W_\alpha(t,p)\exp\int_v^\infty e^{itkw}\,dw^{-\alpha}.$$

Here $\chi(A)$ denotes the indicator function of the event A, $W_\alpha(t,p)$ is the characteristic function of a stable law of index α and parameter p, and k is a constant depending on α and p. Note that for the heavy tailed case, U and V are dependent. Can the type of local dependence of the X-sequence make a difference to the Chow-Teugels limit? It can! If large values are cancelled by large negative values, then sums and maxima can be asymptotically independent. We show this by constructing an example:

Let $\{Y_i\}$ be a stationary sequence with

$$1-F_Y(y) = 1-y^{-\epsilon},$$

$\epsilon<1$, $y\ge 1$ and for $\nu>0$ such that $\epsilon+\nu<1$, let

$$X_i = \begin{cases} Y_i & \text{with probability } Y_{i-1}^{-\nu} \\ -Y_{i-1} & \text{with probability } 1-Y_{i-1}^{-\nu} \end{cases}$$

Then it can be shown that $\{X_i\}$ is stationary, 1-dependent and $1-F_X(x) = \epsilon(\epsilon+\nu)^{-1}x^{-\epsilon}$, $F_X(-x) = x^{-\epsilon}$. Hence the limit distribution of $(\tilde{S}_n, \tilde{M}_n)$ for the associated iid sequence is the Chow-Teugels limit with $\alpha=\epsilon$ and $p=\epsilon(2\epsilon+\nu)^{-1}$. The components of this limit are dependent. However, $(\tilde{S}_n, \tilde{M}_n)$ of the dependent $\{X_i\}$ process can be shown to be asymptotically independent due the cancellation of large positive values by large negative values values, thus showing that local dependence may make a difference on the limit distribution. However, if we rule out this type of cancellation, then the limit distribution is not affected by the dependence in $\{X_i\}$, as the following theorem demonstrates (Anderson and Turkman [4]). One possible local condition which rules out this type of cancellation is Davis' [6] $D'(a_n)$ condition, which we assume in the theorem. This restrictive technical condition also rules out clustering of large and small values above and below certain thresholds. Types of processes which satisfy this condition (and others which we need in the theorem) can be seen in Davis [6].

Let $\{E_i\}$ be an iid unit mean exponential sequence, $\Gamma_j = \sum_{i=1}^{j} E_i$ and $\{\delta_i\}$ iid taking values $+1$ and -1 with probabilities p and q respectively. Then

Theorem 2.2

(i) If $0<\alpha<1$, $p>0$ and conditions $D(a_n)$, $D'(a_n)$ of Davis [6] hold then

$$(\tilde{S}_n, \tilde{M}_n) \to^d \left(\sum_{j=1}^{\infty}\delta_j\,\Gamma_j^{-1/\alpha},\ \Gamma_D^{-1/\alpha}\right),$$

where $D=\min\{j:\delta_j=1\}$.

(ii) If $1\le\alpha<2$, $p>0$ and conditions $D(a_n)$, $D'(a_n)$ and $D''(a_n)$ of Davis [6] hold then

$$(\tilde{S}_n, \tilde{M}_n) \to^d$$

$$\left(\sum_{j=1}^{\infty}\{\delta_j\,\Gamma_j^{-1/\alpha} - (p-q)\,E(\Gamma_D^{-1/\alpha(0,1)})\},\ \Gamma_D^{-1/\alpha}\right).$$

(iii) Under the conditions of (i) and (ii),

$$E[e^{it\tilde{S}_n}\chi(\tilde{M}_n\le x)] \to^d W_\alpha(t,p)\,\Phi_\alpha(x)e^{-h(t,x)},$$

where

$$h(t,x) = \int_x^\infty [e^{it(C\delta)^{1/\alpha}y} - 1]\,d(-y^{-\alpha}),$$

and

$$\delta = p\frac{2-\alpha}{\alpha},$$

and C is a positive constant. Note that the value of the limit does not depend on the dependence structure of $\{X_i\}$.

These results seem to be discouraging for statistical applications. For example, for sequences with finite variance, the independence of (U, V) does not offer a basis for the use of average wind speeds in inferences about gusts, contrary to the evidence shown in data. This may be due to:

(i) Time intervals are not long enough in practice for asymptotic results to give adequate approximations,

(ii) The correlation structure of the data is not well represented by our mixing and local dependence conditions,

(iii) Residual seasonality remains in the data. Based on these possible deviations, statistical models establishing connection between means and extreme events are suggested in Anderson and Turkman [3].

3. Extremes of Averages Over Different Time Scales

The specific problem to be adressed in this section is as follows: Let X_t represent the instantaneous value of the environmental variable at time t, and denote by $X_{T,t}$ a moving average of $\{X_t\}$ over the range T:

$$X_{T,t} = \frac{1}{T} \sum_{i=0}^{T-1} X_{t-i}. \tag{3}$$

Then we are interested in the paired series $\{X_{T,t}, X_{S,t}\}$ for different fixed S and T and in particular, in the joint distributional properties of extremes of the pair $\{X_{T,t}, X_{S,t}\}$.

We will give results only for sequences with the heavy tailed distributions. The light tailed case is more complicated, since in this case large values of the moving averages may occur due to the contribution of several relatively large values of the sequence in contrast to the heavy tailed case when large values of the moving averages are dominated by the largest value of the sequence. The techniques to be used to study these questions will be developments of those used by Davis and Resnick [8].

Suppose that $\{X_i\}_{i=1}^{\infty}$ are iid random variables with

$$P(X_i > x) \sim x^{-\alpha} L(x),$$

where $L(x)$ is slowly varying as $x \to \infty$, that is $X \in D(\Phi_\alpha)$. Take constants a_n such that

$$nP(X_i > a_n x) \to x^{-\alpha},$$

Consider a point process P_n which puts points at $(\frac{k}{n}, a_n^{-1} X_k)$, $k = 1, 2, \ldots$. Hence P_n is the random point measure on sets in $R^+ \times R$

$$\sum_{k=1}^{\infty} \epsilon_{(\frac{k}{n}, a_n^{-1} X_k)}(\cdot),$$

where

$$\epsilon_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if not.} \end{cases}$$

It is known that P_n converges weakly as $n \to \infty$ to a point process P, a Poisson process with mean measure μ on $R^+ \times (R^+ - \{0\})$. (Davis and Resnick call this Poisson measure, the Poisson random measure $PRM(\mu)$ and consider a more general case which involve the left tail of the distribution as much as the right tail. Here due to the special simple form of the moving averages, we restrict ourselves to the space $(0, \infty)$.)

Here

$$d\mu = dt \times \alpha x^{-\alpha-1} \epsilon_x(0, \infty) dx.$$

Hence

$$\sum_{k=1}^{\infty} \epsilon_{(\frac{k}{n}, a_n^{-1} X_k)}(\cdot) \to \sum_{k=1}^{\infty} \epsilon_{(t_k, j_k)}(\cdot),$$

where $\{(t_k, j_k)\}_{k=1}^{\infty}$ are the points of P.

Davis and Resnick [8] show that, correspondingly, for $\{X_{T,i}\}$ with same normalization

$$\sum_{k=1}^{\infty} \epsilon_{(\frac{k}{n}, a_n^{-1} X_{T,k})}(\cdot) \to \sum_{k=1}^{\infty} \sum_{i=1}^{T} \epsilon_{(t_k, j_k c_i)}(\cdot)$$

on $R^+ \times (R^+ - \{0\})$, where

$$c_i = \begin{cases} \frac{1}{T} & i = 1, 2, \ldots, T. \\ 0 & \text{otherwise} \end{cases}$$

and (t_k, j_k) are as above.

Results for the point process generated simultaneously by $\{X_{T,i}\}$ and $\{X_{S,i}\}$ processes can be obtained with a straightforward generalization:

$$\sum_{k=1}^{\infty} \epsilon_{(\frac{k}{n}, a_n^{-1} X_{T,k}, a_n^{-1} X_{S,k})}(\cdot) \to \sum_{k=1}^{\infty} \sum_{i=1}^{T} \epsilon_{(t_k, j_k c_i, j_k \tilde{c}_i)}(\cdot) \tag{4}$$

on $R^+ \times (R^+ - \{0\})^2$, where

$$\tilde{c}_i = \begin{cases} \frac{1}{S} & i = 1, 2, \ldots, S. \\ 0 & \text{otherwise} \end{cases}$$

The joint limit distribution of $M_{T,n}$ and $M_{S,n}$ now can easily be calculated since

$$P(M_{T,n} \le a_n x,\ M_{S,n} \le a_n y)$$

$$= P\Big(\sum_{k=1}^{\infty} \epsilon_{(\frac{k}{n}, a_n^{-1} X_{T,k}, a_n^{-1} X_{S,k})}$$

$$((0,1] \times ((-\infty, x] \times (-\infty, y])^c) = 0 \Big)$$

$$\to P\Big(\sum_{k=1}^{\infty} \sum_{i=1}^{T} \epsilon_{(t_k j_k c_{i,j_k} \tilde{c}_i)} ((0,1] \times ((-\infty, x] \times$$

$$(-\infty, y])^c) = 0 \Big)$$

$$= P\Big(\sum_{k=1}^{\infty} \epsilon_{(t_k, T^{-1} j_k, S^{-1} j_k)} ((0,1] \times ((-\infty, x] \times$$

$$(-\infty, y])^c) = 0 \Big)$$

$$= P\Big(\sum_{j=1}^{\infty} \epsilon_{(t_k j_k)} ((0,1] \times (Tx \wedge Ts, \infty)) = 0 \Big)$$

$$= \exp[-\mu\ ((0,1] \times (Tx \wedge Sy, \infty))]$$

$$= \exp\Big[-\int_{Tx \wedge Sy}^{\infty} q(x)dx\Big]$$

$$= \begin{cases} \exp\ [-p(Tx \wedge Sy)^{-\alpha}] & \text{if } Tx \wedge Sy > 0 \\ 0 & \text{if not} \end{cases} \qquad (5)$$

Special Cases

[1] $S = T$. Then

$$P(a_n^{-1} M_{T,n} \le x,\ a_n^{-1} M_{S,n} \le y) = P(a_n^{-1} M_{T,n} \le x \wedge y),$$

so from theorem 3.1 of Davis and Resnick [8], we should have

$$P(a_n^{-1} M_{T,n} \le x,\ a_n^{-1} M_{S,n} \le y) = \to e^{-p(T(x \wedge y))^{-\alpha}}.$$

Hence the result Eq. (5) is consistent with the existing results on this special case.

[2] $S = 1 < T$, so $M_{S,n} = \max_{i \le n} X_i$, and the result Eq. (5) says that

$$P(a_n^{-1} M_{T,n} \le x, a_n^{-1} M_{S,n} \le y) \to e^{-p\ (Tx \wedge y)^{-\alpha}}.$$

Note that the above distribution is the joint limit distribution of the maximum of the $\{X_i\}$ process and the maximum of a moving average of it.

The above set up is a very simple case. The immediate question is: what if the $\{X_i\}$ are themselves a dependent sequence?

We can get a partial answer by taking $\{X_i\}$ to be itself a moving average:

$$X_i = \sum_{j=1}^{\infty} a_j Z_{i-j},$$

say, where Z_i satisfy the same conditions as the X_i and $a_j \ge 0$ (for simplicity)

Thus

$$X_{T,i} = \{X\} * \left\{\frac{1}{T}\right\},$$

and

$$X = \{Z\} * \{a\}.$$

Hence

$$X_{T,i} = \{Z\} * \{a\} * \left\{\frac{1}{T}\right\} = \{Z\} * \{d\},$$

say, where

$$d = \{a\} * \left\{\frac{1}{T}\right\}.$$

Note that for any $l = 1, 2, \ldots,$

$$d_l = \frac{1}{T} \sum_{i=1}^{T \wedge l - 1} a_{l-i},\ i = 1, 2, \ldots$$

As before

$$P(a_n^{-1} M_{T,n} \le x,\ a_n^{-1} M_{S,n} \le y)$$

$$\to P\Big(\sum_{k=1}^{\infty} \sum_{i=1}^{\infty} \epsilon_{(t_k, j_k d_{i,j_k} \tilde{d}_i)} ((0,1] \times$$

$$((-\infty, x] \times (-\infty, y])^c) = 0 \Big), \qquad (6)$$

where

$$\tilde{d}_i = \frac{1}{S} \sum_{i=1}^{S \wedge l-1} a_{l-i}.$$

Note that

$$\{\sum_{i=1}^{\infty} \epsilon_{(t_k, j_k d_k, j_k \tilde{d}_l)}((0,1] \times ((-\infty, x] \times (-\infty, y])^c) = 0\} \equiv$$

$$\{j_k(d_i, \tilde{d}_i) \in (-\infty, x] \times (-\infty, y]\},$$

for every i for which one of the components is nonzero. This is the case iff $j_k d^+ \leq x$ and $j_k \tilde{d}^+ \leq y$, where $d^+ = \max d_i$ and $\tilde{d}^+ = \max \tilde{d}_i$. Hence the limit of Eq. (6) is given by

$$P\left(\sum_{k=1}^{\infty} \epsilon_{j_k}\left(\frac{x}{d^+} \wedge \frac{y}{\tilde{d}^+}, \infty\right) = 0\right)$$

$$= e^{-\int \frac{x}{d^+} \wedge \frac{y}{\tilde{d}^+} q(x)dx}$$

$$= \begin{cases} e^{-p\left(\frac{x}{d^+} \wedge \frac{y}{\tilde{d}^+}\right)^{-\alpha}} & \text{if } \left(\frac{x}{d^+} \wedge \frac{y}{\tilde{d}^+}\right) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Hence we see that this kind of dependence in the underlying process does not change the form of the limit distribution. It would be interesting to obtain similar results for general stationary sequences. This could be done by using characterization of the limit point processes for the sequence of point processes with points $(t/n, a_n^{-1} X_t, t = 1, 2, \ldots, n)$ given by Davis and Hsing [7] when $\{X_t\}$ is a stationary sequence with regularly varying tails. In their paper, Davis and Hsing show that when $\{X_t\}$ satisfies a proper mixing condition then

$$N_n = \sum_{t=1}^{n} \epsilon_{\left(\frac{t}{n}, a_n^{-1} X_t\right)}(.)$$

converges to a point process N of the form

$$N = \sum_{t=1}^{\infty} \sum_{j=1}^{\infty} \epsilon_{(S_t, P_t Q_{tj})},$$

where

$$\sum_{t=1}^{\infty} \sum_{j=1}^{\infty} \epsilon_{(S_t, P_t)}$$

is PRM(ν) with

$$d\nu = dt \times \gamma \alpha x^{-\alpha-1} dx, \; x > 0,$$

(This is slightly weaker form of Davis-Hsing limit, since we consider the convergence only on $R^+ - \{0\}$ not involving the left tail.) They also show that under the proper mixing condition, the convergence of N_n to N is equivalent to:

For $r_n \to \infty$, $r_n/n \to 0$ as $n \to \infty$, and $k_n = [n/r_n]$,

$$\lim_{n \to \infty} k_n P\left(\max_{1 \leq i \leq r_n} X_i > a_n x\right) = \gamma x^{-\alpha},$$

and

$$P\left(\sum_{j=1}^{r_n} \epsilon_{X_j / \max_{1 \leq j \leq r_n} X_j} \in . \Big/ \max_{1 \leq j \leq r_n} X_j > a_n x\right)$$

$$\to Q(.)$$

Here $Q(.)$ is the distribution of the iid point processes

$$\left\{\sum_{j=1}^{\infty} \epsilon_{Q_{tj}}(.)\right\}_{t=1}^{\infty}.$$

From this basic result it may be possible to obtain the limiting form of

$$N_n' = \sum_{t=1}^{n} \epsilon_{\left(\frac{t}{n}, a_n^{-1} X_{T,t}, a_n^{-1} X_{S,t}\right)}(.),$$

where

$$X_{T,t} = \frac{1}{T} \sum_{j=1}^{T} X_{t-j},$$

and

$$X_{S,t} = \frac{1}{S} \sum_{j=1}^{S} X_{t-j}.$$

One would expect that if the above sequence of point processes converges, then the limit point process should be of the form

$$N_n' \rightarrow \sum_{t=1}^{\infty} \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \epsilon_{(S_t,\, c_i P_t Q_{i\,j},\, c_i P_t Q_{ij})},$$

where

$$c_i = \begin{cases} \frac{1}{T} & i = 1,2 \ldots, T \\ 0 & \text{otherwise} \end{cases}$$

$$c_i' = \begin{cases} \frac{1}{S} & i = 1,2 \ldots, S \\ 0 & \text{otherwise} \end{cases}$$

Straightforward adjustment of Resnick's [11] arguments which involve consecutive application of the continuous mapping theorem is not possible, since these arguments use the convergence of

$$nP(a_n^{-1} X_1 > \delta,\, a_n^{-1} X_2 > \delta) \rightarrow 0, \delta > 0$$

which is clearly not satisfied by most sequences with strong local dependence.

4. References

[1] C. W. Anderson and K. F. Turkman, The joint limiting distributions of sums and maxima in stationary sequences, J. Appl. Probab. **28**, 33–44 (1991).
[2] C. W. Anderson and K. F. Turkman, Limit distributions for sums and maxima, J. Appl. Probab. **28**, 715–716 (1991).
[3] C. W. Anderson and K. F. Turkman, Limiting joint distributions of sums and maxima in a statistical context, Theoriya Veroyatnostei ee Prim. **37**, 352–354 (1992).
[4] C. W. Anderson and K. F. Turkman, Sums and maxima of stationary sequences with heavy tailed distributions, submitted for publication (1992).
[5] T. L. Chow. and J. L.Teugels, The sums and the maximum of i.i.d. random variables, In Proc. Second Prague Symp. Asymptotic Statistics, 81–92 (1978).
[6] R. A. Davis, Stable limits for partial sums of dependent random variables, Ann. Probab. **11**, 262–269 (1983).
[7] R. A. Davis and T. Hsing, Point process and partial sums convergence for weakly dependent random variables with infinite variance, to be published (1993).
[8] R. A. Davis and S. Resnick, Limit theory for moving averages of random variables with regulary varying tail probabilities, Ann. Probab. **13**, 179–195 (1985).
[9] J. E. Fakrell and A. G. Robins, Concentration fluctuations and fluxes in plumes from point sources in a turbulent boundary layer, L. Fluid Mech. **117**, 1–26 (1982).
[10] N. Mole, A model of instrument smoothing and tresholding in measurements of turbulent dispersion, Atmos. Environ. **24**, 313–323 (1990).
[11] S. I. Resnick, Extreme values, Regular Variation, and Point Processes, Springer-Verlag, New York (1987) pp. 231–237.

About the authors: C. W. Anderson is a lecturer in the Department of Probability and Statistics of the University of Sheffield,UK and K. F. Turkman is a professor in the Department of Statistics and Operations Research of the University of Lisbon, Portugal.

Volume 99, Number 4, July–August 1994
Journal of Research of the National Institute of Standards and Technology

News Briefs

General Developments

Inquiries about News Briefs, where no contact person is identified, should be referred to the Managing Editor, Journal of Research, National Institute of Standards and Technology, Building 416, Room 119, Gaithersburg, MD 20899-0001; telephone: 301/975-3572.

"ELECTRONIC EYE" IMPROVES ACCURACY IN LIGHTING

A new device developed at NIST is giving the lighting industry its most accurate measure of brightness ever. This "electronic eye" is twice as accurate as the lighting standard it replaces, explains a NIST physicist. NIST scientists are using the electronic eye to maintain the candela, the international base unit for measuring light. The electronic eye—a photometer with a green filter (to measure brightness), a silicon photodiode (to direct current flow as determined by illumination) and an electronic circuit for signal processing—has an aperture that works much the same way as the human iris. This new candela standard will help the lighting industry meet new light bulb labeling requirements for brightness, energy efficiency, and color rendering. It also will help ensure proper illumination for vehicle control displays. With the electronic eye, NIST has shifted from a light source to a light detector as a primary standard. Although manufacturers can still purchase standard bulbs from NIST, they now can send their own detectors to NIST for on-site calibration. For more information on calibrations, contact the NIST Radiometric Physics Division, A221 Physics Building, Gaithersburg, MD 20899-0001, (301) 975-3216.

NAVY REVIEW POINTS WAY TO IMPROVED CMM PRACTICES

Coordinate measuring machines are state-of-the-art inspection tools for measuring the three-dimensional geometries of manufactured parts and assemblies of parts. A NIST survey of how CMMs are used in U.S. Navy manufacturing facilities identifies 11 high-priority issues that are key to improving CMM applications and quality assurance practices in the production of high-precision parts for weapons systems. The review of 22 Navy in-house and contractor-run production facilities found a significant portion of CMMs to be underutilized or used improperly—a problem also reported in studies of other industries. An estimated 20 000 machines are found in factories and laboratories worldwide. Areas warranting the Navy's attention include: integration of CMM inspection into design and manufacturing; calibrations and interim tests of CMM performance; effects of temperature, humidity and other environmental influences on CMM measurements; measurement methods for assuring the dimensional accuracy of large parts and assemblies; increased measurement throughput and flexibility; and improved understanding of measurement uncertainty and error budgets. The review by NIST researchers was commissioned by the Navy's Manufacturing Technology Program. Single copies of U.S. Navy Coordinate Measuring Machines: A Study of Needs (NISTIR 5379) are available from David Stieren, B113 Metrology Building, NIST, Gaithersburg, MD 20899-0001; (301) 975-3197; fax: (301) 869-0822, e-mail: dstieren@enh.nist.gov (via Internet).

STANDARD TO FOCUS ON TURNING CENTER PERFORMANCE

For prospective buyers of turning centers, determining how one machine tool stacks up against another can be as futile as comparing apples and oranges. That's because there are no agreed-upon methods for comparing the performance—especially the accuracy and repeatability—of different turning centers or, for that matter, periodically assessing the capabilities of machines already on the

shop floor. A solution is in the offing. In an effort partially supported by NIST, a committee of the American Society of Mechanical Engineers (ASME) has turned its attention to developing performance evaluation standards for turning centers. The emerging standard will build on a set of prescribed procedures for testing machining-center performance developed by the same committee and issued as an ASME standard in 1993. That standard, adopted by a growing number of machine tool users, incorporates many practical tests developed or validated at NIST. NIST researchers are now working with industrial collaborators to modify existing testing methods and devices or develop new ones tailored specifically to turning centers. For more information, contact Denver Lovett (who chairs the ASME Technical Committee on Machining and Turning Centers), Room 142 Shops Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-3503, e-mail: lov@micf.nist.gov (via Internet).

ACCREDITATION FOR CALIBRATION LABS ANNOUNCED

The National Voluntary Laboratory Accreditation Program at NIST now will accept applications from calibration labs seeking accreditation to perform calibration services based on compatibility with international standards. The program will help manufacturers, exporters, testing labs and others to gain acceptance of U.S. calibration and test results between countries to avoid barriers to trade. NVLAP programs are operated in conformance with ISO/IEC Guide 58: 1993—Calibration and Testing Laboratory Accreditation System-General Requirements for Operation and Recognition. Accreditation is available to commercial labs; manufacturer's in-house labs; university labs; and federal, state and local government facilities. Foreign-based labs also may be accredited if they meet the same requirements as domestic labs and pay any additional fees required for travel expenses. To obtain an application package or for further information, contact James L. Cigler, program manager, NVLAP, A162 TRF Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-4016, fax: (301) 926-2884.

1994 NVLAP PROGRAM DIRECTORY AVAILABLE

Manufacturers and governments are requiring laboratory accreditation as a key element in the acceptance of products in domestic and international markets. Recent revisions to procedures for the National Voluntary Laboratory Accreditation Program, administered by NIST, have made them fully compatible with the international standards (ISO/IEC Guides 58 and 25) used by other accreditation systems. The National Voluntary Laboratory Accreditation Program 1994 Directory (NIST Special Publication 810) lists more than 700 domestic and foreign laboratories that are accredited by NVLAP as of January 1994. The labs are listed alphabetically, by field of testing and by state. For a copy of SP 810 (1994 edition), send a self-addressed mailing label to: NVLAP, A162 TRF Building, NIST, Gaithersburg, MD 20899, (301) 975-4016, fax: (301) 926-2884.

EXAMINERS NEEDED FOR 1995 BALDRIGE AWARD

NIST is seeking applicants to serve on the 1995 board of examiners for the Malcolm Baldrige National Quality Award. Applicants for the board must be experts in quality management and capable of evaluating large and small manufacturing and service businesses. In addition, quality experts from the health care and education communities are needed to participate in a pilot program. NIST currently is working to adapt the Baldrige criteria and award program for these communities. Those selected for the board must take part in a three-day preparation course based on the award criteria and examination process. In addition, examiners are expected to spend 10 days or more reviewing applications, preparing feedback reports to applicants and, in some cases, participating in site visits. The Board of Examiners currently has 260 quality experts. Applications to serve on the 1995 board will be available in September from the Malcolm Baldrige National Quality Award Office, A537 Administration Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-2036, fax: (301) 948-3716, e-mail: oqp@micf.nist.gov (via Internet). Examiner applications are due Nov. 1, 1994.

ATP PROJECT DEVELOPS WORLD'S BRIGHTEST GREEN LED

A private company, working with the Department of Physics at North Carolina State University (NCSU), has announced the development of a new, extremely bright, LED (light-emitting diode) with peak wavelengths in the pure green region of the spectrum. In work supported by the NIST Advanced Technology Program, the private company

combined its technology for producing high-quality crystals of pure zinc-selenide with novel fabrication techniques developed at NCSU to build LEDs that are estimated to be more than 50 times brighter than commercial green LEDs. Moreover, the company's LED, which is brightest around 510 nm, is significantly more "green" than existing commercial diodes (at about 555 nm), which have a pronounced yellowish cast. The ATP project is aimed at developing improved technologies for producing blue and green light microsources—among the hardest colors to achieve, but important for building full-color displays and in optical data systems. Researchers at the private company and NCSU now are attacking the problem of device degradation. With useful lifetimes of several hundred to 1000 h, the current LEDs are significant improvements on the state of the art but still degrade too quickly for commercial use. The company also hopes to modify the technology to produce true green/blue laser diodes.

WANT TO TRAP CESIUM ATOMS? USE MICROWAVES

Physicists at the National Institute of Standards and Technology and Harvard University have demonstrated for the first time that a cloud of atoms can be held at a temperature near absolute zero with microwaves. This experiment, reported in the May 16, 1994 Physical Review Letters, brings scientists one step closer to observing an exotic phase transition known as Bose-Einstein (or just Bose) condensation. Bose condensation is related to other macroscopic quantum phenomena such as superfluidity and superconductivity. In theory, certain atoms will condense into an unusual gas as the temperature approaches absolute zero. In such an ideal Bose condensed gas, a large number of atoms would have essentially no velocity. The microwave trap holds a small cloud of cesium atoms at a few microkelvins above absolute zero (-273 °C). In the experiments at NIST, scientists first used intersecting laser beams to slow and cool the cesium atoms in a vacuum chamber. Then the lasers were shut off, and microwave radiation plus a magnetic field held the atoms in place. Trapping atoms with microwaves offers physicists an edge in their quest to observe Bose condensation. The newly demonstrated trap has the advantage that unlike previous traps, it can, in principle, hold atoms in their lowest energy spin-state. This would increase the stability of the gas, making Bose condensation easier to obtain. As a next step, the researchers plan to test the microwave trap on hydrogen atoms at Harvard.

UNITED STATES HARMONIZES STANDARDS WITH UKRAINE

To enhance trade between the United States and Ukraine, NIST and the State Committee of Ukraine for Standardization, Metrology, and Certification (known as DERJSTANDART) have signed a memorandum of understanding on scientific and technical cooperation to remove nontariff trade barriers between the two countries. The MOU was completed on May 28, 1994, at the first informal meeting of the U.S./Ukraine Standards Working Group in Kiev. It recognizes the growing importance of the harmonization of standards and conformity assessment measures to improve international trade. For information, contact the Office of Standards Services, A603 Administration Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-4000, fax: (301) 963-2871.

NETWORK FOR NEW YORK MANUFACTURERS EXPANDED

U.S. Commerce Secretary Ronald H. Brown announced a cooperative agreement with the state of New York on June 17, 1994, that officially launched the New York Manufacturing Extension Partnership. Headquartered at the Rensselaer Technology Park in Troy, NY, the New York MEP is the result of four recent awards from the Technology Reinvestment Project (TRP), the federal government's program to provide funds for dual-use (military and civilian) technology development, deployment and utilization. The TRP awards to New York State were used to expand the extension services provided by the former Northeast Manufacturing Technology Center, one of seven original MTCs established by NIST. Specifically, the expansion includes the operation of four regional New York MEP centers in New York City, Fishkill, Endicott and Amherst. The New York MEP, operated by the New York State Science and Technology Foundation, is affiliated with NIST's Manufacturing Extension Partnership. It is one of 28 extension programs funded through the TRP, bringing the current number of centers in the national MEP to 35. For more information on the New York MEP, contact Jeanne Selmer, New York

MEP, 385 Jordan Rd., Troy, NY 12180-8347, (518) 283-1010, e-mail: jeanne_selmer@mailgate.nemtc.itn.org (via Internet).

PAPER DETAILS CALIBRATION SYSTEM FOR POWER METERS

Using tunable laser diodes, NIST has developed a new method to calibrate optical power meters, the most common type of test equipment used in the optical fiber industry. This system is in addition to the existing one which NIST has used for years to calibrate optical power meters at fixed wavelengths using both collimated beam and fiber/connector configurations. To minimize measurement errors associated with the source wavelength and detector spectral responsitivity, NIST has installed tunable laser sources in the measurement system. With this system, optical power meters can be calibrated both at a particular source wavelength or over the range of wavelengths of the tunable laser diode. For a paper explaining the new system in detail, contact Sarabeth Moynihan, Div. 104, NIST, Boulder, CO 80303-3328, (303) 497-7765, e-mail: moynihan@bldrdoc.gov (via Internet).

NIST REPORTS ON METRIC, THE FEDS AND INDUSTRY

Most U.S. firms that do business abroad are predominantly metric because of global sourcing of parts, service, components, and production. A new report from NIST, A Metric for Success (NISTIR 5425), states, "The metric system cannot be avoided in international trade and commerce. It will be interesting to see how long the United States can hold out against the worldwide use of the International System of Units (known as SI), the modern metric system. What is even more curious is why would the world's leading industrial nation want to resist using a world standard?" The report discusses the worldwide metric momentum, inefficiencies and benefits, and emphasizes that "metric equals standardization." Also discussed is the role of the Commerce Department and NIST in the national transition to the metric system. A chronology of metric in U.S. history is included for reference. Copies of NISTIR 5425 are available from the Metric Program, A146 TRF Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-3690, fax: (301) 948-1416.

ADVANCES IN FIBER OPTIC SENSORS FEATURED IN PAPER

Fiber optic sensors often have been used for measuring magnetic fields and electric current. These polarimetric sensors incorporate the Faraday effect whereby the plane of polarization of a light beam is rotated by a magnetic field. A recent paper highlights NIST research on the use of diamagnetic glasses and iron garnets as sensor materials. One part describes how the Faraday effect in glass (silica) can be used to make inexpensive current sensors from optical fiber. The sensor is formed by looping a single-mode optical fiber around a conductor. Current sensors using this fiber are approaching commercial availability for use in the electric power industry. The paper also describes NIST research using iron garnets, a class of ferromagnetic materials, for high-sensitivity magnetic field and electric current sensors. NIST has developed two current sensors employing iron garnets, and it expects additional materials research will result in iron garnet compositions specifically tailored for magneto-optic sensing applications. For a copy of technical paper 29-94, contact Sarabeth Moynihan, Div. 104, NIST, Boulder, CO 80303-3328, (303) 497-7765, e-mail: moynihan@bldrdoc.gov (via Internet).

BALDRIGE AWARD PROGRAM TRAINS PILOT HEALTHCARE AND EDUCATION EVALUATORS

The Malcolm Baldrige National Quality Award Program has just completed the training of its initial 44 pilot evaluators for education and healthcare. These pilot evaluators were trained using case study material developed for the 1994 Baldrige Award Board of Examiners and using case item material developed specifically for healthcare and educational institutions. Pilot evaluators will use the lessons they learned during training to assess healthcare and educational institution case studies. These case studies, together with evaluation notes and a sample feedback report (such as a Baldrige applicant would receive) will be made available to the public in the fall of 1994. It is hoped that the case study material will both educate healthcare and education institutions on use of the Baldrige Criteria in quality management assessment of their institutions, and also provide a learning opportunity by disseminating some current leading practices through the content of the case studies.

In 1995, full-scale pilot efforts are planned in healthcare and education. Organizations will be invited to submit applications; the applications will receive full review, with feedback to the applicants on their strengths and areas for improvement. Since this will be a learning activity for both the Baldrige Program and applicant organizations, no awards will be presented in the 1995 cycle. Subsequent to the pilot effort, a decision will be made on next steps for implementation of new award categories in healthcare and education.

NVLAP PROCEDURES

The Federal Register, Vol. 59, No. 84 of Tuesday, May 3, 1994, announced the final rule making changes to regulations pertaining to the operation of the National Voluntary Laboratory Accreditation Program (NVLAP). The NVLAP procedures are redesignated as Part 285 of Title 15 of the Code of Federal Regulations.

These procedures have been expanded to include accreditation of calibration laboratories; updated to achieve compatibility with conformance assurance and assessment concepts; modified to ensure consistency with relevant International Organization for Standardization documents; and revised to facilitate and promote acceptance of calibration and test results between countries to avoid barriers to trade.

The revised procedures will foster cooperation among laboratories and with other bodies, and establish the basis for bilateral and multilateral agreements.

NEW DIGITAL BRIDGE TO PROVIDE NIST CUSTOMERS IMPROVED IMPEDANCE SERVICES

NIST scientists have designed, developed, and fabricated a digital impedance bridge that will permit NIST to offer improved impedance services in response to customers' needs when it is put on line later this year. Commercial impedance bridges are used to make critical, practical measurements of component parameters; electronic circuits providing advanced performance, together with reduced size of components, make these measurements more difficult and more critical. Standards that will be calibrated by the new digital impedance bridge will in turn be used to calibrate commercial bridges and their ancillary equipment. Until the new bridge is in service, NIST is forced to rely on a venerable Maxwell-Wien bridge as the basis for its low-frequency impedance services. The new bridge measures two- or three-terminal impedances with precisions of 200 to 2000×10^{-6} depending on the measurement frequency in the range 20 kHz to 100 kHz, the magnitude of the impedance, and the nature of the transfer standard. The automated character of the bridge makes it highly flexible, and many more data points can be taken than were practical before.

The result is that customers will be able to get a comprehensive analysis of the performance of the transfer standards they submit, including information on drift and on frequency dependence. Other advantages of the digital bridge are that it offers more flexible scaling through its sampling system and it can be calibrated with "unlike" standards, i.e., calibration of a capacitive transfer standard at a given frequency does not require use of a standard capacitor but could use a standard resistor at that frequency. NIST scientists also have developed a special probe intended for calibrating inductors. The probe consists of a number of programmable ac resistors, each of which can be switch-selected so as to be within a few centimeters of the terminals of a standard inductor to minimize the effect of parasitic impedances. The inductor is compared to the appropriate probe resistor by equalizing the currents through both, using the bridge and its control software. The voltage across the inductor and the probe resistor is then digitized and the data are processed using a four-parameter sine fit to determine the phase and amplitude of each signal.

NIST HELPS MANUFACTURER INVESTIGATE TACKY SPOTS FOUND ON PARKING AREAS OF HARD DISKS

In a recent collaboration with a major U.S. manufacturer of computer disk drives, a NIST scientist applied a method known as lateral force microscopy to help the company investigate the presence of micrometer-sized sticky or tacky features on the surface of a hard disk. The possibility of these features occurring on any part of the surface of a disk where the read/write heads can be positioned interferes with the design goal of reducing head flying height, an essential requirement as the areal bit storage density increases in advanced magnetic media. Flying heights already are well below 100 nm. The features are micrometers in extent and 20 nm to 30 nm high, comparable to the

disk peak surface roughness of about 20 nm and therefore difficult to image by such common methods as optical or electron-beam microscopy. To the naked eye, the features appear as a dull ring coinciding with the area where the heads are parked.

The scientist's work shows that morphology and friction can be correlated using lateral force microscopy to identify sources of increased friction. The lateral force microscope is a derivative of the atomic force microscope, which can achieve atomic resolution. In the atomic force microscope, a small cantilever having an integral tip is scanned over the surface of a specimen, and deflections up and down are plotted to provide a topographic map. Servoing arrangements maintain the force between tip and specimen surface nearly constant. In lateral force microscopy, the tip is scanned over the specimen perpendicular to the long axis of the cantilever. When a variation of friction between the tip and specimen surface is encountered, the cantilever twists slightly. It is this twisting that is plotted as a function of tip position. Collaboration is continuing with the disk drive manufacturer regarding the source and chemical composition of the sticky features.

OPERATION OF SNS JUNCTION FOR VOLTAGE STANDARD DEMONSTRATED AT 38 K

NIST scientists recently carried out an experiment that serves as a proof-of-principle demonstration for the operation of superconductor-normal-superconductor junctions for voltage standards at elevated temperatures, near 40 K. This temperature can be reached with closed-cycle refrigeration systems and does not require liquid helium as a cryogen. Although potentially offering uncertainties of parts in 10^{10}, Josephson-Junction voltage array standards have found limited acceptance in industry even where their capabilities would be advantageous because they require liquid helium for cooling to near 4 K. As an original inventor and developer of the array standards, NIST is investigating approaches that would increase their utility.

As part of this work, the division team fabricated a single yttrium/barium/copper oxide step-edge junction with a gold barrier and illuminated it with 62 GHz electromagnetic radiation. When they measured the current-voltage characteristic, they found the amplitude of the first quantized voltage step to be 1 mA at 38 K, a value large enough to show that the junction could operate at elevated temperatures without being limited by thermal noise. The results of this experiment were not consistent with predictions of the commonly used resistively shunted junction model, and a NIST scientist formulated a new long-junction model that provided a better understanding of the results. The theoretical understanding coupled with the empirical results offer optimism that the proposed rapidly programmable array can be implemented. The next experiment will be to demonstrate operation of a junction over a ground plane; following that, methods for reliably fabricating literally tens of thousands of nearly uniform junctions will need to be developed (not necessarily at NIST) to result in a practical device. The work is described in a submission to Applied Physics Letters.

CRADA ESTABLISHED WITH PRIVATE COMPANY TO EVALUATE NEW CALIBRATION TECHNIQUE AND THERMAL TRANSFER INSTRUMENT

NIST scientists are working with a private company to evaluate the high-voltage (100 V to 1000 V) performance of a new automated thermal transfer instrument, Model TRS-104, and to verify the innovative step-up techniques recommended by the manufacturer for characterization at those voltages. A cooperative research and development agreement (CRADA) has been established for these purposes.

The project will involve the exchange of well-characterized thermal converters to be used by the private company to study the performance of the new instrument and the impact of the new step-up technique. Although the philosophy underlying the step-up technique is well established, the company's staff believes that its new method will permit customers to minimize the number of calibration points required to calibrate the instrument. The CRADA will enable the company to determine if such an approach is feasible. In the course of this project, NIST staff will re-examine the techniques used to calibrate NIST working standards used for ac-dc difference measurements at high voltages; reassess the performance of the NIST high-voltage range resistors; and perform extended measurements on the company instruments. This effort is expected to yield improvements in NIST services as well as a means of verification for the new instrumentation.

This is the fourth CRADA undertaken by NIST in the past 5 years to evaluate the performance of state-of-the-art instrumentation in this critical measurement area. Such agreements familiarize NIST staff with the characteristics of instruments and standards soon to be on the market and help industry market superior products.

NII CHALLENGES ADDRESSED IN WORKSHOP

NIST recently co-sponsored a Workshop on Advanced Digital Video in the National Information Infrastructure (NII). The wide range of applications that are proposed for NII—including marketing, manufacturing, medicine, education, and entertainment—impose varying and complex demands on the NII, including capabilities and flexibilities that are not available in existing computer or television systems, and are exciting intense interest in how the NII develops. In response, the organizers of the workshop provided a forum for defining a vision of the role of digital video within the NII, identifying the architectural, scaling, and performance issues in realizing this vision, and recommending the research, experiments, and steps to be taken to resolve these issues. Experts in information services, broadcasting, computing, consumer electronics, and government policy broke up into groups to focus on architectural considerations, modular decomposition, interoperability, display performance, image capture and display requirements, and digital delivery service.

A full report of the workshop is being prepared for publication. A feature of the workshop was a demonstration of the proponent high-definition television systems that were incorporated into the Grand Alliance proposal to the Federal Communications Commission for adoption as a broadcast standard in the United States.

PROJECT FOR IMPROVING PISTON TURNING MACHINES COMPLETED WITH SUCCESS

A collaborative project to improve piston turning machines for the automotive industry has been conducted for the past 13 months at NIST with industrial participants from automobile manufacturers, a machine tool builder, and the National Center for Manufacturing Sciences. NIST researchers characterized one of the new-generation piston turning machines loaned from the machine builder, modeled its thermal behavior, and generated software to predict the behavior during the actual cutting operation.

At the end of the development phase, NIST researchers carried out several days of testing on the shop floor, cutting aluminum piston billets over 6 h periods. In an attempt to simulate actual production conditions, certain environmental conditions were varied and production stoppages corresponding to plant coffee breaks and lunch hours were incorporated into test periods. Results showed up to five times improvement in the accuracies of the billets machined with NIST thermal compensation technology as compared with those machined without it.

NEW, POWERFUL THEORY FOR ELECTRON-IMPACT IONIZATION CROSS SECTIONS

In collaboration with the University of Nebraska-Lincoln, a NIST scientist has developed a new, powerful theory that predicts electron-impact ionization cross sections, needed for modeling and understanding the interaction of electron beams with gases, liquids, and solids. This new theory is expected to provide sorely needed ionization cross sections for applications in plasma processing, radiation damage monitoring, fusion device design, and in basic atomic and molecular physics.

The theory, using readily available atomic and molecular data, produces cross sections that agree with known experimental results within 5% to 10% from threshold to several keV in incident energy. The theory has been verified successfully for light atoms, ions, neon, and hydrogen and water molecules. The theory is being applied to angular distributions of ejected electrons and proton-impact ionization cross sections. Preliminary results on angular distributions are very promising too. When utmost accuracy is not required, ionization cross sections based on a simpler version of this new theory can be calculated using a PC program such as LOTUS-123.

LABORATORY SPECTRA OF A STRATOSPHERIC CHLORINE RESERVOIR MOLECULE OBTAINED

The chlorine nitrate ($ClONO_2$) molecule serves an important role in the stratospheric chlorine cycle. It is formed in darkness by the reaction of chlorine

monoxide with nitrogen dioxide, and the chlorine removed by this reaction is thus prevented from reaction with the stratospheric ozone. During daylight hours, however, it is photolyzed back to its original components. It is sometimes referred to as a chlorine reservoir. The concentration of stratospheric chlorine nitrate is monitored using infrared spectrometers usually carried on balloon or aircraft flights.

Previous laboratory studies of chlorine nitrate were unable to resolve the rotational fine structure of the infrared bands, which is obscured by several very low frequency vibrational states. In order to model the temperature dependence (and, therefore, shape) of the observed stratospheric chlorine nitrate infrared bands, it is necessary to know the energy levels and transition probabilities of each band used for concentration monitoring.

In collaboration with a private company, NIST scientists have obtained completely resolved spectra of $ClNO_3$ using a newly constructed pulsed nozzle molecular beam apparatus coupled with a diode laser spectrometer. The resulting cooling obtained as the molecular beam passes into a vacuum greatly simplifies the spectrum by removing the lines originating in vibrational states. The $ClNO_3$ band centered at 1292.5 cm^{-1}, one of the IR bands used for remote sensing, has been completely resolved at a beam temperature of 25 K and with spectral resolution of 0.001 cm^{-1} or better. With this molecular information, models are being developed to predict the spectrum of this band at stratospheric temperatures in order to improve the monitoring accuracy.

NIST COMMISSIONS MEDICAL AND INDUSTRIAL RADIATION FACILITY

NIST has installed an electron accelerator as the heart of a new user facility for the medical and industrial radiation communities. The Medical & Industrial Radiation Facility (MIRF) is based on an rf-powered, traveling-wave electron linac (linear accelerator) donated by the Radiation Therapy Center of Yale University-New Haven Hospital. The accelerator provides electron energies from 7 MeV to 32 MeV at an average beam current of up to 0.1 mA.

Medical linacs are used for treating 500 000 cancer patients annually in 1300 treatment facilities in the United States. The medical dosimetry applications of the MIRF relate to development and testing of instruments and dosimetry systems for use in the clinical facilities. The MIRF also offers unique opportunities for industrial research and a number of companies and consortia have expressed interest. Other accelerator applications currently under investigation include electron-beam treatment of waste water (University of Maryland), curing of polymer composites (Oak Ridge National Laboratory), radiation effects on electronics (University of Maryland), and use of the accelerator for production of channeling radiation and coherent bremsstrahlung (Catholic University and George Washington University).

IMPROVED TIME SCALE RELIABILITY

NIST scientists have substantially enhanced the reliability of the NIST time scale through development of completely redundant clock measurement systems. Two identical measurement systems, controlled by independent PCs, observe the same physical clocks and both run independent copies of the AT1 time scale. Both drive independent microsteppers to provide redundant physical realizations of UTC (NIST) in real time. One system is used as the official output, but the other can be switched into service should there be any failure of the primary system. The dual systems also are being used to study the performance of the AT1 algorithm in general and especially its robustness in the presence of measurement noise.

This work provides the basis for operational changes in the division's long-range plan. First, these systems will facilitate replacing all of the large computers with arrays of relatively cheap PCs. In addition to reducing maintenance costs, such a network (combined with GPS receivers) could facilitate widespread sharing of data and network-based dissemination of UTC (NIST) at very high accuracy. The development of dual measurement systems also provides the basis for moving the time scale when building and remodeling begins on the site. The longer term objective is to separate (at geographically different locations on the site) two groups of clocks with their independent measurement systems. These two clock systems would be interconnected, but the physical separation would guard against time-scale disruption by a catastrophe such as a major fire.

MECHANISM OF MATERIAL REMOVAL IN MACHINING OF CERAMICS

Grinding with diamond wheels is the most prevalent method of machining of advanced ceramics. In this machining method, material is removed by the individual diamond particles producing a series of parallel grooves on the machined surface. Existing models are based on the assumption that material removal occurs by the formation and propagation of cracks below the surface of each groove. These models have been developed for amorphous materials and, therefore, are not suitable for polycrystalline ceramics.

Recent research results at NIST obtained as part of the Ceramic Machining Consortium have shown that material removal in polycrystalline ceramics, e.g., aluminum oxide, occurs by three processes: (1) microfracture and chipping within grains; (2) intergranular fracture and grain dislodgement; and (3) removal of large segments by formation and propagation of subsurface cracks. The influence of each process depends on the load applied to the diamond particle and the microstructure of the ceramic. The researchers have found that the present models describing the material removal process as propagation of subsurface cracks are adequate for machining of fine-grained ceramics under large loads, but, under normal grinding conditions, the other microfracture processes provide a more accurate description of the removal process. These results provide new insights into the basic mechanisms of material removal and the specific role of ceramic microstructure, as well as the design of ceramics with improved machinability, thereby reducing the cost of machining.

MOLECULAR DYNAMICS OF ALTERNATIVE REFRIGERANTS

The importance of the development of alternative refrigerants to CFCs (chlorofluorocarbons) is highlighted by the looming Jan. 1, 1996 ban on CFC productions. For compressors in home refrigerators and auto air-conditioning systems, the chosen replacement is the hydrofluorocarbon (HFC) 134a (F_3C-CFH_2). Although an enormous effort has been given to the development of the HFCs, there is still much to be learned of the properties, chemistry, and interactions of these molecules.

A recent collaborative research effort has been initiated between scientists at NIST, and a private company, to investigate by neutron-scattering techniques the bonding state and molecular dynamics of HFC 134a and its isomer HFC 134 (HF_2C-CF_2H), encaged in the cavities of zeolite molecular sieves. Several neutron-scattering techniques are being combined with complimentary methods, such as infrared, Raman, and NMR, to obtain a better understanding of the guest-host interactions, bonding geometry and reorientational dynamics of the adsorbed HFCs 134 and 134a. The insight gained from this collaborative research effort is directed toward the development of improved methods for the separation and storage of the molecular isomers 134a and 134, both of which are formed during industrial production of 134a.

SECRETARY OF COMMERCE APPROVES DIGITAL SIGNATURE STANDARD AS FEDERAL INFORMATION PROCESSING STANDARD (FIPS)

The Federal Register of May 19 announced that the Secretary of Commerce has approved the Digital Signature Standard (DSS) as FIPS 186. To be effective in 6 months, the DSS provides the capability to generate digital signatures that cannot be forged. This capability is needed by federal government agencies to carry out their responsibilities for electronic exchanges and to improve government operations through the use of information technology.

NIST AND THE U.S. NUCLEAR REGULATORY AGENCY (NRC) COLLABORATE ON NUCLEAR SAFETY

NIST Special Publication 500-216, Proceedings of the Digital Systems Reliability and Nuclear Safety Workshop, presents the results of a workshop held September 13–14, 1993, in Rockville, MD. The NRC, in cooperation with NIST, conducted the workshop to provide a forum for the exchange of information among experts within the nuclear industry and experts from other industries, regulators, and academia. Topics covered a broad range of safety issues such as areas where current software engineering practices may be inadequate for safety-critical systems, methods for reducing risk in such systems, and research directions in the use of digital systems in nuclear power plants.

NEW PUBLICATION PRESENTS PROCEEDINGS OF TEXT RETRIEVAL CONFERENCE

NIST Special Publication 500-215, The Second Text REtrieval Conference (TREC-2), constitutes

the proceedings of this symposium held at NIST August 31–September 2, 1993. Co-sponsored by NIST and the Advanced Research Projects Agency, the conference was the second in an ongoing series of workshops to evaluate new technologies in text retrieval. The event was attended by 150 people involved in the 31 participating groups composed of industry, academia, and government researchers.

The goal of the conference was to bring together research groups to discuss their retrieval results using a large test collection. Attendees reported on a wide variety of retrieval techniques, including methods using automatic thesauri, sophisticated term weighting, natural language techniques, relevance feedback, and advanced pattern matching. As results had been analyzed with a common evaluation package, groups compared the effectiveness of different techniques and discussed how differences between the systems affected performance.

DISTRIBUTED SUPERCOMPUTING SOFTWARE SUBJECT OF NEW REPORT

NISTIR 5381, Distributed Supercomputing Software: Experiences with the Parallel Virtual Machine—PVM, focuses on defining the profile requirements culminating from NIST's assessment of the Parallel Virtual Machine (PVM). Developed by researchers at the Oak Ridge National Laboratory and Emory University, PVM is a distributed system consisting of a portable suite of software designed for use by parallel and supercomputing application engineers. NIST researchers are studying PVM to assist them in defining the system service requirements needed to support parallel programming and supercomputing activities in the general-purpose distributed setting.

FIRST CUSTOMER USES NEW CALIBRATION SERVICE

A step-gage calibration done for a private company inaugurated a pilot measurement program offered by NIST at the Y-12 facility at the Department of Energy Oak Ridge Centers for Manufacturing Technology. The calibration was performed on Y-12's M-60 large-volume, high-accuracy coordinate measuring machine. The new service leverages NIST's measurement expertise and Y-12's specialized equipment, enabling the two organizations to respond to a previously unmet measurement need voiced by automobile, aircraft and heavy-equipment manufacturers. Under the arrangement, NIST metrologists supervise calibrations of end standards and step gauges up to 1.35 m long, and issue a calibration report. Manufacturers use these standards to verify the accuracy of their own measurement machines. Previously, U.S. companies were forced to obtain such services from foreign measurement laboratories. In terms of accuracy, cost and turnaround time, the new domestic calibration service tops the performance of the best foreign supplier, which has been benchmarked by NIST. For more information, contact David Stieren, B113 Metrology Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-3197, e-mail: dstieren@enh.nist.gov (via Internet).

NEW THERMOCOUPLE CAN "TAKE THE HEAT"

A NIST researcher has invented a class of materials for thin-film thermocouples that allows measurement of temperature changes up to 1200 °C in microsecond increments. Made with molybdenum silicide or titanium silicide, the NIST thermocouples can measure temperature changes in hostile environments such as inside diesel or jet engines. Widely used fine-wire thermocouples have much slower response rates (tenths of seconds). More advanced thin-film platinum/rhodium thermocouples have microsecond response rates but deteriorate at 700 °C to 900 °C. The patented NIST materials are made with a thin, outer layer of heat-resistant silicon dioxide that protects an electrically conducting molybdenum or titanium silicide layer and a pure silicon layer underneath. Temperature changes cause proportional changes in voltage in the silicide layer. For more information, contact Kenneth Kreider, A303 Physics Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-2619, e-mail: kkreid@enh.nist.gov (via Internet).

MOTIONLESS REFRIGERATOR LIQUEFIES NATURAL GAS

A new cooperative research and development agreement between NIST and a private company aims to apply refrigeration technology developed by NIST and Los Alamos National Laboratory to the task of liquefying natural gas. NIST and LANL scientists invented the thermoacoustically driven orifice pulse tube refrigerator (or TADOPTR), which has no moving parts, contains tubes of helium gas and is capable of producing a temperature of 112 K. The helium gas is repeatedly compressed and expanded with sound waves rather than with a mechanical compressor as in most conventional cooling systems. The company has obtained the development license as well as the exclusive license to the patents through LANL.

The NIST and LANL scientists will assist the company in upscaling the TADOPTR into two versions that will be manufactured, tested and marketed by the company. The first will liquefy 1900 L of natural gas per day, while the second will increase that output to 38 000 L/d. The cost of TADOPTR liquefaction plants will be very economical, with liquefaction taking place on-site and eliminating transportation costs. This means TADOPTRs may help in areas removed from natural gas pipelines and in the production of natural gas to fuel the "clean-car" vehicles of the future. For more information, contact Ray Radebaugh at Div. 836.02, NIST, Boulder, CO 80303-3323, (303) 497-3710.

CONSORTIUM SEEKS MORE PREDICTABLE PAINTS

Private companies and the Federal Highway Administration have joined NIST in a cooperative research and development consortium to help get new, highly predictable paint products more quickly to market. Because of health and environmental concerns, the chemical makeup and manufacturing processes for making paints have changed tremendously over the past decade. Unlike older paints, new formulas do not have a well-established history of performance. Also, reliable methods of predicting performance have not kept pace with the changes. As a result, potential problems with painted products could cost manufacturers millions of dollars to repair. The consortium's goal is to find a better way to predict the service life of paint. It is expected to last 3 years. For further information contact Jonathan Martin, B348 Building Research Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-6717.

NIST, CHILE TO COLLABORATE ON ANALYTICAL METHODS

Chilean and American scientists are collaborating to improve techniques for analyzing environmental samples and advanced materials. A new international agreement between NIST and Chile's Commission of Nuclear Energy will expand the analytical capabilities of both institutes. An early goal of the joint program is to develop and apply two neutron beam analytical methods. Scientists at NIST will help the Chileans develop instruments for neutron depth profiling and prompt gamma neutron activation analysis. These techniques offer advantages over other methods in that they do not destroy samples. Chilean scientists will provide NIST with advanced materials samples, such as lithium-based superconductors and ceramics, for analysis and comparison of results. Following the nuclear methods work, the collaboration could be extended to other analytical methods, such as electrochemistry and mass spectrometry.

VIRTUAL REALITY TESTBED UNDER WAY

Virtual reality allows a user to interact with a simulated environment as though it were real. Among its many applications is the ability to realistically practice jobs that are too dangerous, too expensive or impossible to carry out for real. Private companies are using virtual reality techniques to help factory workers fabricate complex wiring assemblies, to test visibility from construction equipment before it is built or to study improved methods of automobile assembly. To enhance the usability and further the development of virtual reality technology, NIST has established an Open Virtual Reality Testbed. The testbed, which includes prototype virtual reality systems, was set up to encourage the development of standard interfaces so that component virtual reality systems from different vendors are interoperable. A description and sample demonstration of testbed activities is available online through the World Wide Web at http://nemo.ncsl.nist.gov/~sressler/OVRThome.html. For more information, contact Sandy Ressler, B266 Technology Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-3549, e-mail: sressler@oops.ncsl.nist.gov (via Internet).

"FARSIGHTED" DETECTOR SEES MORE INFRARED

Night vision goggles that detect infrared radiation are a common feature in spy thrillers. But seeing in the far infrared (longer wavelength infrared radiation) has proved more difficult. Now NIST researchers have developed a detector sensitive enough to do the job. The detector consists of an antenna and strip of superconducting material built into an integrated circuit. The circuit is cooled to 90 K, the material's superconducting transition temperature. Even tiny amounts of far infrared radiation collected by the antenna heat the superconductor near its transition point causing relatively large changes in electrical resistance. The device is twice as sensitive and responds 1000 times faster than other nitrogen-cooled thermal detectors. Arrays of the new detectors could be useful for atmospheric studies. For more information, contact Joseph Rice at B208 Physics Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-2133.

JOINT OPTOELECTRONICS AGREEMENT NOW IN PLACE

The United States and Japan have agreed on the last critical step before realization of a joint program to further the design and development of the advanced computing technologies integrating optical and electronic components. The agreement was reached at a recent meeting between two Japanese organizations, the Ministry of International Trade and Industry and the Real World Computing Partnership, and five U.S. agencies. It establishes the process for funding a "broker" through which the Joint Optoelectronics Project will be carried out in the United States. The broker is a service that links designers of advanced computer systems dependent on optoelectronic devices and modules with suppliers of such components in research and development divisions of companies in both countries. Each country will have its own broker, but suppliers in both nations will be available to Japanese and American designers. According to the agreement, NIST will oversee the selection of the broker in the United States. The two brokers will cooperate to bring together designers of innovative, advanced computer systems and fabricators of optoelectronic components, allowing the designers to evaluate their ideas by manufacturing experimental prototypes.

SOFTWARE "BUILDS" PROCESS CONTROL SYSTEMS

New NIST-developed software shows promise as a tool for rapid prototyping of manufacturing control systems. Developed as part of a NIST-led consortium, the software was used successfully to design and operate the control system for a laboratory-scale process for making metal powders. The Expert Control System Shell, or ECSS, features a rule-based expert system to assist designers and a graphical user interface resembling a control panel. Designers can choose from among on-screen versions of push buttons, toggle switches, slide controls and other types of actuators. User-interface options for displaying process data include dials, charts and graphs. The control system can be operated manually or under the full or partial command of the expert system. If the designer chooses, the overall control system can be partitioned into a series of concurrently running modules that synchronize their activities by swapping electronic messages. Data also are logged for post-process analyses as well as for simulations of process changes. A NIST computer scientist who is the ECSS designer suggests that with further refinement and field testing, the software could be used to build control system architectures for other types of dynamic manufacturing processes as well as be adapted to other computer platforms. For more information on ECSS, contact Steve Osella at A127 Metrology Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-4263, e-mail: oscll a@cme.nist.gov (via Internet).

DEMO PUTS INTEROPERABILITY TO THE TEST

A group of computer networking equipment makers have shown that the Integrated Services Digital Network could be used to link their products and provide communication between diverse applications. The multicompany hookup used the Point-to-Point Protocol running over an ISDN basic rate interface B-channel to demonstrate interoperable, local- and wide-area-network connectivity at a recent meeting of the North American ISDN Users' Forum, or NIUF, held at NIST. PPP is a set of protocols developed by the Internet Engineering Task Force that allows dissimilar LAN connection equipment to negotiate quickly which features and protocols will be supported by both ends of a network connection. As part of the demonstration, vendors and end users accessed Internet, read their electronic mail and sent files back home. Interoperability, a key component of the National Information Infrastructure, will enable rapid expansion of telecommuting, remote Internet access and connection with branch offices. Seven vendors from the United States, Canada and Europe participated in the demonstration. For technical information, contact Jeff Fritz of the NIUF's Enterprise Network Data Interconnectivity Family on (304) 293-2060 or jfritz@wvnvm.wvnet.edu (via Internet).

CRADA PARTNERS SEEK BETTER-BEHAVED VAV SYSTEMS

An air-delivery system known as variable air volume, or VAV, can save energy dollars by delivering just the right amount of air to areas of a building that need it. But outdated and complex control strategies for VAV systems are creating problems such as control instability, poor air quality, and inadequate humidity control and ventilation supply. Now, NIST and a private company are teaming under a cooperative research and development agreement to help improve the control and operation of VAV systems. Using a special NIST

laboratory that contains a VAV air-handling system, researchers from NIST and the company will evaluate currently used control methods, then develop and test alternative control techniques and strategies. Also, various methods for detecting and diagnosing faults on-line will be explored. The CRADA project is expected to end in April 1996. For more information, contact George Kelly, B114 Building Research Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-5870, e-mail: gekelly@enh.nist.gov (via Internet), or John Seem, Johnson Controls Inc., 507 E. Michigan St., Milwaukee, Wis. 53201, (414) 274-4677.

LUNAR REFLECTOR WORKS THROUGH SILVER ANNIVERSARY

One of the space program's longest-running experiments—and one with a NIST connection—celebrated its 25th anniversary in July by continuing to return data. During their pioneering moon landing on July 20, 1969, the Apollo 11 astronauts set up a laser reflector to make precise measurements of the distance between the Earth and moon. The still-operational experimental station reflects a powerful laser pulse aimed at it from telescopes on Earth. By measuring how long the pulse takes to return to Earth (the round trip takes about 2.5 s), scientists have defined the Earth-moon distance to within 2.5 cm. The reflector was designed primarily by a NIST scientist at the Joint Institute for Laboratory Astrophysics, operated cooperatively by NIST and the University of Colorado. It consists of a briefcase-sized aluminum panel studded with 100 corner reflectors (the corners of precision-ground glass cubes that have been cut off at 45 degree angles), each about 3.8 cm across. When a ray of light enters the cut-off surface, it is internally reflected from the three sides of the corner, exits the cut-off surface parallel to its entry path and then returns to its source. The same principle is used in bicycle reflectors. The Apollo 14 and 15 missions delivered two other reflectors, including one with 300 cube corners. All three reflectors are targeted almost nightly by scientists at observatories in Texas and France.

NIST TO COOPERATE WITH ARGENTINA AND ECUADOR

In order to provide a mechanism for scientific and technical cooperation in chemistry, physics, and engineering measurement sciences, NIST recently signed a Memoranda of Understanding with the National Institute for Industrial Technology of the Secretary of Industries, Argentine Republic; the Secretariat of Science and Technology, Argentine Republic; and the Ecuadorian Institute of Standardization, Republic of Ecuador. These memoranda will provide a framework for the exchange of scientific and technical knowledge, services, and the augmentation of scientific and technical capabilities. The agreements will be effective immediately for a period of 5 years.

NIST DATA CLARIFIES MODEL FOR TIME-DEPENDENT DIELECTRIC BREAKDOWN

A NIST scientist has carried out work that appears to resolve the choice of competing models for predicting voltage breakdown in integrated circuits. For devices to be reliable, silicon dioxide insulator layers on integrated circuit chips must resist voltage breakdown through years of normal use, and device designers need to be able to estimate this performance. The reliability of oxide layers can be predicted by subjecting the devices to high-voltage stress at elevated temperatures and extrapolating device failure rates to voltages and temperatures at which the devices would normally be used. Calculations of expected device failure times from the time-dependent dielectric breakdown (TDDB) data depend on the model chosen for performing the extrapolation. Until recently, two models contended for calculating t_{50} from the TDDB data, where t_{50} is the time at which half of the devices would have failed. One model assumed that t_{50} was depended on reciprocal electric field; the other model assumed a linear dependence on electric field.

The scientist realized that if he could heat his specimen devices to temperatures higher than those normally used in this type of stress testing, he could reduce the electric field to values much nearer those actually encountered by devices in normal use. At the same time, by using lower fields, he could avoid failures from breakdown mechanisms that occur at high fields but that are not likely to apply in normal device operation. Accordingly, the NIST scientist used a specially designed wafer chuck to stress his specimens at temperatures as high as 698 K (425 °C). The resulting low-field data revealed a linear dependence of t_{50} on electric field. The field dependence of data taken at higher electric fields can not be clearly distinguished; either the linear or reciprocal field dependence can be used to describe the data. The scientist presented these results in a paper on TDDB at the recent International Reliability Physics Symposium.

UNCERTAINTIES IDENTIFIED FOR RADAR CROSS-SECTION MEASUREMENTS

In a study of major government facilities, NIST scientists have identified significant sources of measurement uncertainties encountered in both static and dynamic radar cross-section ranges. Radar cross section is a parameter that defines how easy it is to detect the presence of an object illuminated by a radar beam for a given orientation of the object with respect to the beam. Radar cross-section measurements are used to determine the stealthy performance of aircraft and missiles and other military vehicles; they are now becoming of interest for civilian applications, including air traffic control, highway traffic control, and shipping operations. The division team developed methods for estimating component uncertainties and for combining these into an overall uncertainty budget.

The goal was to provide a reasonable and uniform methodology for evaluating radar cross-section measurements that can be used for both indoor and outdoor test ranges to produce compatible estimates of uncertainty. A benefit of the process is that uncertainty estimation identifies sources of error on which attention needs to be concentrated to improve measurement quality. The team studied and compared alternative methods of estimating uncertainties. The overall uncertainty budget for a specific measurement was developed as a hierarchy of uncertainty budget tables, where the lower-level tables provide the details that are summarized in the upper-level tables. A spread sheet computer program was developed to facilitate the construction of such a hierarchy of uncertainty tables. A thorough analysis of measurement uncertainty is a first step in the development and implementation of a program for certifying radar cross-section ranges. Additional details can be found in a report Proposed Uncertainty Analysis for RCS Measurements (NISTIR 5019).

METHOD FOR ASSESSING ACCURACY OF ON-WAFER MICROWAVE MEASUREMENTS BENEFITS INDUSTRY IN NIST VISITS

NIST researchers in collaboration with the NIST/Industrial MMIC Measurement Consortium, have developed a method for experimentally determining the accuracy of on-wafer measurements when performed using commercially available instrumentation in industrial environments. On-wafer measurements play a critical role in the design of monolithic microwave integrated circuits (MMICs). On-wafer measurements also have reduced test costs and thereby significantly reduced the total cost of MMICs making them cost effective in commercial applications such as personnel communications, wireless networks, and intelligent highway vehicles. Heretofore, the industry lacked adequate methods of evaluating the accuracy of on-wafer measurements.

Beginning in July 1993, NIST researchers began testing the verification method at industrial sites to determine its effectiveness in quantifying measurement errors. Over the following 9 months, they tested the method at several locations comparing actual measurement errors to bounds on those errors predicted by the NIST-developed verification method. The results showed that the technique effectively revealed errors in the industrial measurements. These errors were sometimes as large as 30%, large enough to frustrate the design process and lead to increased cost. The magnitudes of the errors were so large and the experiments so convincing that several participants have been prompted to adopt entirely new calibration procedures. The results of the verification method were also useful in diagnosing the causes of the calibration errors. This ability to pinpoint the source of calibration failure turned out to be an extremely important aid to engineers who otherwise would have had no guidance in how to correct such problems. The NIST team has begun to explore ways of establishing the method as an industry standard.

MONOLITHIC SINGLE-FREQUENCY SOLID-STATE WAVEGUIDE LASER DEMONSTRATED

Working in collaboration with a researcher at a University and with a CRADA partner, NIST scientists have demonstrated a monolithic single-frequency solid-state waveguide laser. This development will have important ramifications for metrology since the rare-earth doped glass waveguide laser is extremely stable with respect to thermally induced wavelength drift and optical feedback that plagues typical semiconductor lasers. Furthermore, the waveguide laser, because of its centimeter-scale size, is more thermally and more mechanically stable than a similarly doped optical fiber laser. These attributes show the potential for applications of the new device as both stable reference standards and stable sources.

The waveguide laser is forced to operate in a single longitudinal mode by virtue of a distributed-Bragg reflector grating that was photolithographically formed on the surface of the waveguide. The

laser was fabricated in a special neodymium-doped phosphate glass that was developed in collaboration with a private company. The device is distinctive since it demonstrates single-frequency laser action is possible in a range of approximately 20 nm around the fluorescence peak at 1052 nm. The waveguide geometry is compatible with optical fibers. This aspect of the device could be important since many new local-area network fiber optic communication systems are envisioned to operate in this wavelength band. Work is also under way to demonstrate similar results in erbium-doped glass operating around the traditional telecommunication wavelength of 1544 nm. Initial results indicate that it now should be possible to fabricate arrays of lasers that operate at identical or displaced wavelengths.

NIST SOFTWARE SIMPLIFYING RESISTIVITY DETERMINATION IN DEMAND

A NIST scientist has developed and published a collection of computer programs for calculating resistivity profiles of semiconductor wafers from measurements of their spreading resistance or four-probe resistance. There has been considerable interest in the NIST package, called RESPAC; several U.S. semiconductor companies have requested it. Resistivity is the most important property for specifying semiconductor material and for evaluating its suitability for use in integrated circuits, power devices, or other applications. The NIST scientist's set of 10 FORTRAN77 programs incorporate simplified routines for the necessary mathematical operations and are suitable for use on a personal computer or workstation. NIST SP 400-91, Semiconductor Measurement Technology: A Collection of Computer Programs for Two-Probe (Spreading) Resistance and Four-Probe Resistance, documents the software and supplies background material to enable the reader to use it in an optimal manner.

FIRST DIRECT MEASUREMENTS DEMONSTRATE LOW-NOISE PERFORMANCE POTENTIAL OF HIGH-TEMPERATURE JOSEPHSON JUNCTIONS

Two NIST scientists have made the first direct measurements of the microwave-frequency noise of Josephson junctions made from high-critical-temperature superconductor material. Using yttrium/barium/copper oxide step-edge junctions developed in previous work and an ultralow-noise 1 GHz measurement system, they have measured (for the lowest-noise device) a maximum available noise temperature of $32\ K \pm 2\ K$ at a physical temperature of 4.3 K. The peak noise temperature is in good agreement with published simulations of the ideal resistively shunted junction model. The results imply that such junctions are capable of noise performance in millimeter-wave and terahertz detection and mixing applications that approach the fundamental quantum mechanical limits.

WORKSHOP ON TESTING STRATEGIES TRANSFERS NIST METHODOLOGY TO INDUSTRY FOR ANALOG AND MIXED-SIGNAL PRODUCTS

NIST scientists hosted and presented the third workshop on Testing Strategies for Analog and Mixed-Signal Products at NIST in June to transfer NIST-developed methodology to industry. Efficient testing is an essential ingredient for competitive manufacturing of electronic products. In some cases, the testing cost can approach and even exceed the original manufacturing cost. For the past several years, NIST has conducted a program to develop widely applicable analytical tools that can lead to reduction in the number of test-point measurements and at the same time an improved prediction of the performance of the item under test. Participants were introduced to these tools and a comprehensive framework for developing and implementing efficient tests for analog and mixed-signal devices and instruments. For many of these products it is not physically or economically feasible to perform exhaustive testing. Therefore, test engineers must formulate abbreviated strategies that are economical to execute but still yield accurate measures of overall performance.

The attendees represented diverse application interests, ranging from down-hole instrumentation (such as used in oil exploration) to the testing of pacemakers, to the testing of two-dimensional, charge-coupled device arrays. The agenda consisted of tutorial material on matrix algebra, lectures on the theory of a mathematical procedure known as QR factorization and test point selection, and the development of accurate device error models using physical, a priori, and empirical approaches. Participants were given hands-on training in the use of a commercial software product to implement the matrix operations needed to develop a given testing strategy. Specific examples were presented on how to apply the NIST-developed methods to the problem of opti-

mizing the testing of data converters (analog-to-digital and digital-to-analog) and other mixed-signal devices, as well as the testing of instruments such as multirange ac-dc transfer standards.

ATP AND PED SPONSORS WORKSHOP ON ELECTRON BEAM MODELING

Accurate electron beam modeling is primary to the development of standards for both microanalysis and linewidth metrology. A NIST workshop on "Electron Beam Interaction Modeling for Metrology and Microanalysis in the Scanning Electron Microscope," the first of its kind, was held during the SCANNING 94/SouthEast Electron Microscopy Society (SEEMS) 94 Meeting May 17–20 in Charleston, SC. The NIST workshop was truly international with attendees coming from Japan, Europe, and Russia.

The workshop included four segments: modeling and its relationship to microanalysis, theoretical aspects of electron beam interaction modeling, modeling and its relationship to metrology, and a discussion and functional workshop session. During that session, a modeling round robin was agreed upon based on two defined standard structures, one for microanalysis and the other for dimensional metrology. During both afternoon laboratory sessions, participants demonstrated their own modeling programs. The output from this recent workshop will be archived in a special issue of SCANNING to be published in late 1994 or 1995. It is planned that the results of the round robin and any further work will be reported at a similar workshop in 1995.

NIST HOSTS WORKSHOP ON ADVANCED MACHINE TOOL STRUCTURES RESEARCH

NIST held a Workshop on Advanced Machine Tool Structures Research on April 29, 1994 to discuss issues related to the development of a new class of machine tools with parallel kinematics. Machine tools in which actuators work in parallel to produce tool or part motions—similar to the operation of Stewart platform flight simulators—promise many advantages over conventional machine tools. These include an expected order-of-magnitude increase in stiffness and accuracy, large acceleration capability, lower part production cost, and a full six axes of motion. An Advanced Machine Tool Structures Testbed (AMTST) is currently being established at NIST to identify and extend the limits of capability of these new machines. A primary objective of the workshop was to explore possibilities for using this testbed for joint research projects with industry. Approximately 40 attendees from industry, government, and academia participated in the workshop.

A NIST scientist presented the research activities planned for the NIST AMTST. The primary research objectives are to:

- adapt current performance evaluation procedures to this new class of machine tools;
- investigate high-accuracy metrology systems for parallel machine tools;
- examine the use of microactuation to enable extremely precise motions; and
- advance the state of the art in open architecture machine tool controllers as applied to parallel machines.

The characteristics of two experimental prototype parallel machines currently being developed by private industry were presented. The precision machining of titanium jet engine rotors as an application that could benefit from the stiffness and multiaxis machining capability of a parallel machine was discussed. Five additional speakers described research work and programs related to parallel machines.

A great deal of interest was expressed by the workshop attendees in participating in the research to be carried out in the AMTST. In the afternoon, there was discussion of the possible modes of interaction and participation. These ranged from supplying example test parts to be machined to more formal arrangements such as sending guest researchers and setting up cooperative research and development agreements. Attendees were encouraged to submit ideas and proposals for cooperative research in the AMTST to NIST.

BODY DIMENSIONS FOR APPAREL

Anthropometric data and apparel sizing is an important component of apparel quality. Apparel can not be top quality unless it fits consumers satisfactorily. In the United States, current sizing standards rely on body-measurement data that were gathered by the U.S. Department of Agriculture during the late 1930s. Changes are needed to accurately represent today's U.S. population. A NIST scientist has published a report entitled "Body Dimensions for Apparel." The report represents a preliminary set of body dimensions that are necessary in the manufacturing and fitting of apparel. It is the result of a comparison of five

body-measurement reports, including documents on national and international apparel sizing standards. The author expects that the information in the report will contribute to future body-measurement surveys as well as the development of new or improved sizing standards. In addition, access to anthropometric databases is expected to be an integral component of emerging technology for apparel design engineering.

In recent years, NIST has been developing the apparel product data exchange standard (APDES). The Defense Logistics Agency (DLA) is sponsoring the APDES project to extend the emerging international Standards for the Exchange of Product Model Data (STEP) to include apparel product data. This work is part of a larger DLA program to improve apparel manufacturing technology. These extensions will lay the groundwork for computer integration of the apparel product life cycle, and it will enable clothing manufacturers to reap the benefits of standardized product data representation. The report of body dimensions will serve as input for developing APDES.

HIGH-FIDELITY SENSOR

In past years, NIST has developed a high-fidelity sensor to measure dynamic material displacement in solids. Theoretical modeling work has provided a firm foundation for the interpretation of the sensor output from various excitation sources on plate structures. The sensor includes a conical-shaped piezoelectric ceramic for point pick-up of a surface displacement and is known as the "NBS conical transducer." The ceramic provides an effective coupling to metals, which possess a comparable acoustic impedance equal to the product of material density and sound speed. A calibration facility for acoustic emission (AE) technology in the kilohertz to megahertz frequency range has been implemented based on this research. Industry's application of this technology includes the monitoring of pressure vessel and aircraft structures for incipient material failure.

Recently, a parallel effort has shown promise for the development of a high-fidelity AE sensor that is applicable to polymeric materials, which have an acoustic impedance very different from that for metals. This particular sensor is designed with a polyvinyl diflouride piezoelectric film, attached to a plastic cone tip, to provide the displacement-voltage transduction. A potential application is for the general assessment of composite material quality and specifically for the bond integrity in multilayered structures.

COLLABORATION WITH HUNGARY'S NATIONAL OFFICE OF MEASURES

Scientists from Hungary's National Office of Measures (OMH) visited NIST recently to continue their collaborative research in pH and electrolytic conductivity under the auspices of the U.S./Hungary Science and Technology Program.

NIST and Hungary scientists conducted intercomparisons of the pH buffer standards that serve as the respective national standards for this important and ubiquitous analytical measurement. Under particular evaluation during this visit was sodium borate or borax, which is the buffer material for pH measurements around pH 9. These pH 9 measurements are important for the treatment of industrial boiler water. This material exhibits some unexplained liquid junction potential effects that may be related to impurities in the material.

Additional studies were initiated on standards for low-level electrolytic conductivity measurements critical to power plants, the pharmaceutical industry, and other users of high-purity water. Two different approaches have been adopted by NIST and OMH; NIST choosing potassium chloride or benzoic acid in an alcohol-water mixture, and OMH choosing dilute solutions of boric acid. These studies involved the comparison of data for the accuracy, reliability, stability, and compatibility of the two approaches. The goals of this program are international comparability and global harmonization of the standards for the measurements of pH and conductivity.

NEUTRON INTERFEROMETRY FACILITY OPERATIONAL

One of the results from modern quantum physics is that the elementary constituents of matter behave not only as particles but also as waves. This wave-particle duality means that elementary particles can exhibit interference effects analogous to those observed with light. Interference with light has been understood for centuries and is now widely used as a technological tool. Interference with "matter waves" is an emerging tool with great promise in both fundamental and applied science.

Its development and application is the thrust of the NIST Neutron Interferometry Facility, which recently became operational at the NIST Cold Neutron Research Facility. In this neutron interferometer, a beam of low-energy neutrons is coherently split and recombined by near-perfect crystals of silicon, just as a light beam can be split and recombined by partially reflecting mirrors. The interferometry facility employs a number of novel techniques to obtain the high degree of noise reduction (acoustic, vibrational, and thermal) required for these sensitive measurements. Particularly noteworthy is a stabilization system that maintains the position of a 41 metric ton, vibrationally isolated mass within ± 1 μm and ± 1 s of arc for operating periods of weeks to months. Interferograms taken at the new facility show phase contrast of up to 70% and phase stability of better than one degree per day. These characteristics, which represent considerable improvements on previous installations, will allow a variety of new investigations to be pursued.

The ionizing radiation research program at the NIST Neutron Interferometry Facility will include fundamental metrology studies and a variety of materials science studies. One very promising avenue of research is the application of neutron phase topography for the investigation of hydrogen and hydrogenous impurities in metals and other media. The NIST Neutron Interferometry Facility will be opened as a national user facility in 1995; proposals for time allocation are already being considered.

HIGH MAGNETIZATION ADVANCED MAGNETIC NANOCOMPOSITES

A cooperative research and development agreement (CRADA) between NIST and a private company has resulted in the improved engineering of magnetic nanocomposites comprised of a high concentration of 5 nm diameter γ-Fe_2O_3 clusters finely dispersed in a polymer resin. These materials, prepared by the ionic exchange of Fe ions with protons inside the polymer, can be designed to be superparamagnetic at room temperature and thereby possess a high magnetization without remanence. Since the CRADA was signed a year ago, this collaboration has enabled the private company to optimize the processing of these nanocomposites, resulting in an increase in the saturation magnetization of the ferrofluid version of the superparamagnetic material by a factor of 2. This ferrofluid version now possesses magnetization values approaching 17 $A \cdot m^2/kg$, i.e., 15% larger than that possessed by common magnetic fluids. These materials are particularly attractive as advanced toner materials, new generation magnetic refrigerants, high-density magnetic recording media, and high-sensitivity indicators in medical diagnostics.

DEVELOPMENT OF NEW STANDARDS FOR THE CONTINUOUS STEEL STRIP INDUSTRY

Electrogalvanized steel for automobile body panels has become one of the steel industry's major products. Because of this, a great deal of effort has been focused on improving the quality of the coated sheets and reducing manufacturing costs. One of the primary ways the industry can accomplish this is through on-line monitoring of processing and coating parameters. In response to steel industry requests, NIST is developing new zinc on steel Standard Reference Materials for calibrating on-line coating thickness gauges.

Because plating processes vary from line-to-line, the basic goal of this development effort is to provide industry with a uniform, well-characterized zinc coating as a baseline for their thickness measurements. NIST researchers have been working closely with coated strip steel, automotive, and instrumentation manu- facturers to ensure a standard that will meet their needs for assessing manufacturing as well as performance capability.

Other applications for the zinc-coating thickness standard are calibration of laboratory x-ray fluorescence instruments and acceptance testing by end users of the electrogalvanized product.

IMPROVED ACCURACY IN QUANTITATIVE PHASE ANALYSIS

The accurate measurement of phase abundance by powder diffraction methods is important to a broad range of technologies. It has particular application to analyzing the polymorphs of silicon nitride and zirconia, two ceramic materials being developed for structural applications in automotive and aircraft engines. The Rietveld method, wherein the entire diffraction pattern is used for a refinement of crystallographic, microstructural, and instrument parameters, provides the most accurate and precise method of obtaining results from powder diffraction data.

Therefore, the Rietveld analysis of x-ray and neutron powder diffraction is being investigated at NIST, in collaboration with the Los Alamos National Laboratory, for use in certification of Standard Reference Materials (SRMs) for quantitative analysis.

The accuracy of a quantitative analysis can be considered in terms of an ability to match a calculated result from properly characterized specimens. Rigorous evaluations of the models used in the Rietveld code were possible by working with well-characterized samples prepared from SRMs and SRM-candidate materials. Several improved models were incorporated in the Rietveld code, the most notable of which concerned the contribution of thermal diffuse scattering to the background. Using the Rietveld method, the researchers have verified that SRM 676, an alumina powder, is phase pure to within 0.2% with 95% confidence. They also have certified SRM 656, for analysis of two polymorphs of silicon nitride, not only with respect to the crystalline phase composition but also to the amorphous content.

NONDESTRUCTIVE EVALUATION OF NATURAL GAS PIPELINES USING GAS-COUPLED ULTRASONICS

Collaborating with Southwest Research Institute (SWRI), NIST researchers have shown that it is feasible to use gas as a couplant for ultrasonic inspection of natural gas pipelines. In the past, the use of gas couplants was restricted to the transmission and pitch-catch geometries. The NIST/SWRI work demonstrates that such restrictions can be overcome so that the same transducer can be used to send and receive the probing ultrasonic signals at MHz frequencies. In addition to static tests using a specially designed apparatus at NIST, measurements were carried out in a flowing gas in a pipeline test facility at SWRI. The experimental data indicate that it may be feasible to adapt the basic NIST/SWRI method for in-service inspection of natural-gas pipelines for thickness gaging and flaw detection.

ADDITION TO NIST PROFICIENCY SAMPLE FACILITY DEDICATED

An addition to the Proficiency Sample Facility of the NIST Construction Materials Reference Laboratories (CMRL) was dedicated on May 18, 1994. CMRL is jointly sponsored by NIST, the American Society for Testing and Materials, and the American Association of State Highway and Transportation Officials. The addition was constructed to keep up with growing demand for proficiency samples of construction materials such as concrete, cement, soil, and asphalt. The CMRL supplies over 8000 samples annually to construction materials testing laboratories to help evaluate testing equipment and procedures. Construction of this facility at NIST is an excellent example of cooperation between the federal and state governments and private industry in promoting the quality of construction.

NIST DEVELOPS LARGE BUILDING INPUT FILES FOR MULTIZONE INDOOR AIR QUALITY MODEL

NIST has developed input files for the multizone airflow and indoor air quality model CONTAM88. These input files describe four large buildings: a 12-story multifamily residential building, a five-story mechanically ventilated office building with an atrium, a seven-story mechanically ventilated office building with an underground parking garage, and a one-story school building. The physical characteristics of each building and its idealization as a multizone airflow system are described. These input files enable a user to employ CONTAM88 (and the most recent version of the program CONTAM93) to study airflow and contaminant dispersal in large buildings without developing building idealizations and inputting them into the program. Results of selected computer simulations are presented to demonstrate the effects of wind speed, indoor-outdoor temperature difference, and the percentage of outdoor air intake in the supply air on building air change rates and interzonal airflows in these four buildings. The report describing these input files (NISTIR 5440, CONTAM88, Building Input Files for Multi-Zone Air Flow and Contaminant Dispersal Modeling) also contains an appendix with a database of building component air leakage values.

NIST CO-SPONSORS WORKSHOP ON STANDARDS DEVELOPMENT AND THE NATIONAL INFORMATION INFRASTRUCTURE (NII)

On June 15–16, 1994 NIST, the Science, Technology and Public Policy Program at Harvard University, and the Technology Policy Working Group of the Information Infrastructure Task Force co-sponsored an invitational workshop on standards development and the evolving NII. NIST Director Arati Prabhakar and Lewis M. Branscomb, director of the Science, Technology and Public Policy Center for Science and International Affairs at Harvard, opened the workshop discussions. Prabhakar said that the Information Infrastructure Task Force is working in partnership with the private sector to set public goals for the NII and that the issue of standards is pervasive to every application area. Branscomb focused on the importance of standards

as components of the government's technology policy and the need to define the roles of government, industry, and other groups in standards development.

Workshop presenters and panelists discussed new practices and new institutions in standards development, conceptualizing the standards process, the evolution of standards institutions, the role of the government, requirements for interoperability, and the impact of intellectual property rights on standards development. James Burrows, CSL director, participated in a panel on standards policy and the NII. Papers presented at the workshop and position papers contributed will be published and will also be available in electronic form.

SPECIFICATIONS OF AN ELECTRONIC RESEARCH NOTEBOOK FOR THE NIST SCIENTIFIC STAFF ISSUED

NISTIR 5395, Preliminary Functional Specifications of a Prototype Electronic Research Notebook for NIST, presents a study on the feasibility of using electronic research notebooks (ERNs) at NIST. The study team interviewed the NIST technical and scientific staff to survey current notekeeping practices and identify specific needs for the ERN. A set of basic and enhanced ERN features resulted from the survey; current technologies and products were then assessed to see how requirements could be met. The publication concludes with a proposed system configuration where functional specifications for a basic ERN are defined.

Standard Reference Data

PROTEIN DATABASE NOW INCLUDES NASA EXPERIMENTS

A newly expanded database on crystal growth conditions of biological macromolecules is now available to help the pharmaceutical and food industries improve medicines, vaccines, food products and other industrial processes. The NIST/NASA/CARB Biological Macromolecule Crystallization Database has been expanded to include the NASA Protein Crystal Growth Archive and data from international microgravity experiments. The database was developed by a NIST research chemist and associate director of the Center for Advanced Research in Biotechnology. Version 3.0 of the Biological Macromolecule Crystallization Database includes data on more than 2000 crystal structures of 1500 biological macromolecules. Users can search for data by 20 different parameters, such as macromolecule name, crystal density, crystallization method and year reported. The expanded database is available for $415. To order the NIST/NASA/CARB Biological Macromolecule Crystallization Database, contact the NIST Standard Reference Data Program, A323 Physics Building, Gaithersburg, MD 20899-0001, (301) 975-2208, fax: (301) 926-0416, e-mail: srdata@enh.nist.gov (via Internet).

Volume 99, Number 4, July–August 1993
Journal of Research of the National Institute of Standards and Technology

Calendar

October 4–7, 1994
NORTH AMERICAN ISDN USERS' FORUM (NIUF)

Location: National Institute of Standards and Technology
Gaithersburg, MD

Purpose: To develop user-defined applications, implementation agreements for existing standards, and tests needed for a transparent, ubiquitous, and user-driven Integrated Services Digital Network (ISDN).
Topics: National Information Infrastructure, applications analysis and multimedia, ISDN wiring and powering issues, and other related ISDN topics.
Format: Tutorials, users' and implementors' workshops, and working group meetings.
Audience: ISDN users, implementors, and service providers.
Sponsor: NIST.
Contact: Sara Caswell, B364 Materials Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-4853.

October 5–7, 1994
FEDERAL WIRELESS USER'S FORUM

Location: Gaithersburg Hilton
Gaithersburg, MD

Purpose: To educate users about wireless telecommunication opportunities, expose industry to government needs, explore the similarity of government needs to those of non-government users, and develop application scenarios and solutions.
Topics: Wireless issues, including land mobile radio, the intelligent vehicle highway system, and wireless applications.
Format: General and breakout sessions.
Audience: Current or potential federal, state, and local government wireless telecommunication users; nongovernment users; and the wireless telecommunications industry.
Sponsors: Federal Wireless Users' Forum, National Communications System, and NIST.
Contact: Tish Antonishek, Secretariat, A216 Technology Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-2922.

October 11–14, 1994
17th NATIONAL COMPUTER SECURITY CONFERENCE

Location: Baltimore Convention Center
Baltimore, MD

Purpose: To provide a forum for addressing traditional security concerns, as well as security issues associated with the emerging National Information Infrastructure (NII).
Topics: Tracks on research and development, integration and applications, and management and administration, in addition to a tutorial track. Topics include directions of IT security, network security, viruses, risk management, contingency planning, and privacy.
Format: Five tracks, which provide a combination of peer-reviewed papers and panel sessions, and opening and closing plenary sessions on present subjects and issues of interest and importance to the community.
Audience: A large, diverse national and international audience—approximately 1500 to 2000 are expected from industry, government, and academia.
Sponsors: NIST and National Computer Security Center.
Contact: Dennis Gilbert, A216 Technology Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-3872.

October 17–20, 1994
ANNUAL CONFERENCE ON FIRE RESEARCH

Location: Gaithersburg Hilton
Gaithersburg, MD

Purpose: To report and discuss advances in fire science, with the intent of stimulating new products that are more fire-safe and new ways to capture that value in the ways products are tested and approved for use.
Topics: The conference will focus on the phenomenology of fire, including chemistry and physics of materials combustion, soot and toxic gas formation, fire signatures and their detection, fire-induced flows, and flame spread and extinction.
Format: Parallel sessions of oral presentations and poster sessions.
Audience: All parties with an interest in advances in fire safety, including researchers on the fundamentals of fire behavior, fire safety practitioners, and manufacturers of products potentially impacted by fire safety regulations.
Sponsors: Building and Fire Research Laboratory, NIST.
Contact: Sheilda Smith, B250 Polymer Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-6864.

October 18–21, 1994
47th ANNUAL GASEOUS ELECTRONICS CONFERENCE

Location: National Institute of
Standards and Technology
Gaithersburg, MD

Purpose: To report and discuss basic phenomena and processes in ionized gases, and relevant theory and experiments concerning basic atomic collision processes.
Topics: Fundamental processes in electrical breakdown of gases and maintenance of discharges of all types, in-gas lasers, ion sources, plasma chemistry and innovative plasma applications, plasma treatment of waste and pollutants, and diamond film deposition.
Format: Oral presentations of invited, contributed talks, and poster presentations.
Audience: Scientists and engineers from academic and industrial laboratories involved in specified and related topics.
Sponsors: Division of Atomic, Molecular, and Optical Physics, American Physical Society, and NIST.
Contact: Jean Gallagher, A323 Physics Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-2204.

November 7–9, 1994
NATIONAL EDUCATORS' WORKSHOP

Location: National Institute of
Standards and Technology
Gaithersburg, MD

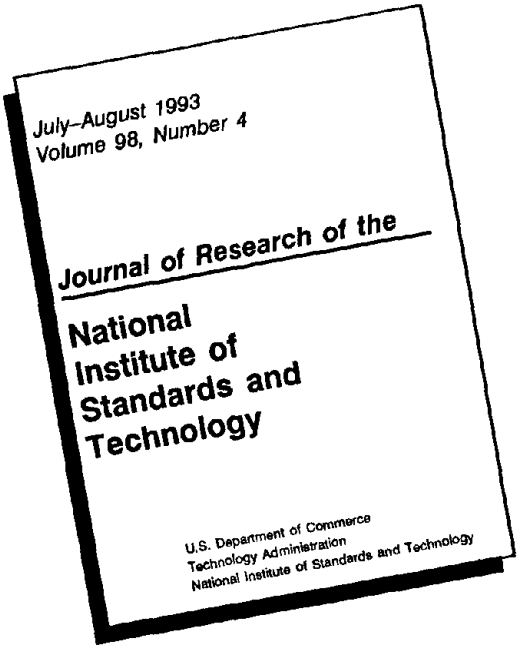
Purpose: This eighth annual workshop is aimed at improving the teaching of materials science, engineering, and technology.
Topics: Experiments and demonstrations of materials science and engineering will be presented.
Format: Lectures and plenary sessions will be held in addition to mini-workshops for small groups in various laboratories throughout NIST.
Audience: College and university-level teachers.
Sponsors: NIST, NASA, Norfolk State University, and DOE.
Contact: Anna Fraker, B254 Materials Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-6009.

November 22, 1994
WORKSHOP ON TISSUE ENGINEERING: FROM BASIC SCIENCE TO PRODUCTS

Location: National Institute of
Standards and Technology
Gaithersburg, MD

Purpose: To provide a forum for the review of research and development advances in tissue engineering and their application to biomedical products.
Topics: Basic science and engineering aspects related to the potential for manufacture of products. Topics will include scaffolding, encapsulation, cell technologies, and biomedical applications.
Format: General session and panel discussions.
Audience: Scientists, engineers, and clinicians with an interest in tissue engineering research and development.
Sponsors: NIST, NIH, FDA, and NSF.
Contact: Stan Abramowitz, A409 Administration Building, NIST, Gaithersburg, MD 20899-0001, (301) 975-2587.

AT $27.00 A YEAR, CAN YOU AFFORD NOT TO KNOW WHAT'S GOING ON AT THE NATION'S MEASUREMENT SCIENCE LABORATORY?



The *Journal of Research* Brings You Up-to-Date Scientific Articles and Information on:

- Measurement Science and Technology
- Calibration Services
- Standard Reference Materials
- Cooperative Research Opportunities and Grants
- Conference Reports

AND MUCH MORE!

It's All At Your Fingertips In the ***Journal of Research of the*** **NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY**

SUBSCRIBE TODAY!

Journal of Research of the
National Institute of Standards and Technology

Superintendent of Documents **Subscription** Order Form

☐ **YES,** send me ___ subscriptions to the **JOURNAL OF RESEARCH OF THE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY** at $27 per subscription (6 times a year) so I can stay up to date on the latest developments in measurement science and technology.

Order Processing Code
6596

1. Please Type or Print

_____
(Company or personal name)

_____
(Additional address/attention line)

_____
(Street address)

_____
(City, State, ZIP Code)

( )
(Daytime phone including area code)

May we make your name/address available to other mailers? YES ☐ NO ☐

2. The total cost of my order is $_____. All prices include domestic postage and handling. International customers please add 25 percent.

3. Please Choose Method of Payment:

☐ Check payable to the Superintendent of Documents

☐ GPO Deposit Account ☐☐☐☐☐☐☐–☐

☐ VISA, or MasterCard
☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

_____ *Thank you for your order!*
(Credit Card Expiration Date)

_____
(Purchase Order No.)

_____
(Signature) (10/93)

4. MAIL TO: New Orders, Superintendent of Documents, P.O. Box 371954, Pittsburgh, PA 15250–7954.

HELP COMMEMORATE THE "PIONEER" OF QUALITY

DR. W. EDWARDS DEMING

The U.S. Bureau of the Census and the International Statistical Institute (ISI) are sponsoring an annual lecture to commemorate and advance the work of Dr. W. Edwards Deming. Dr. Deming was exceptional in the duration and impact of his accomplishments, and his commitment to the goal of continuous quality improvement. He is perhaps best known for his ability to release the power of quality leadership, develop and use statistical analysis methods, and insist on practical operating improvements.

The purpose of this joint Census/ISI lecture series is to provide an opportunity to reflect on the wide variety of Dr. Deming's accomplishments, including but not limited to their implications for government and industry. The lecture will be a featured event of the Census Bureau's Annual Research Conference, and will be available in an accompanying paper published by each sponsoring organization. Research conferences have been held each year since 1985 and attract an international audience of 700–800 persons. The next research conference will be held in March 1995 at a Washington, DC area location.

Persons from a variety of professional backgrounds and experiences are invited to submit abstracts for this lecture. The sponsoring organizations will select the lecturer based on a competitive review of the abstracts. The full published lecture paper will be distributed internationally. To be considered for the 1995 lecture, you should submit a 3–4 page abstract by October 4, 1994, to:

Maxine Anderson-Brown
Office of the Director, Room 2270-3
Bureau of the Census
Washington, DC 20233

For additional information about the lecture, please contact her at the above address or at 301-763-1150 or FAX 301-763-4887.

International Workshop on Semiconductor Characterization: Present Status and Future Needs

January 30—February 2, 1995
Gaithersburg, Maryland, U.S.A.

The International Workshop on Semiconductor Characterization: Present Status and Future Needs will be held Monday, January 30 through February 2, 1995, at NIST in Gaithersburg, Maryland. The Workshop provides a forum to present and discuss critical issues, problems and limits, evolving requirements and analysis needs, future directions, and key measurement principles, capabilities, applications, and limitations. It will be comprised of formal invited presentation sessions, poster sessions for contributed papers, and panel sessions. Invited sessions are planned on:

- Si Process Development and Manufacturing — The Drivers
- Analytical Technology and Metrology Requirements for Beyond 0.35 μm Technology
- Process and Characterization Issues
- Above-Si Processing
- Critical Analytical Methods
- Si and Compounds: In-Situ; Real-Time Diagnostics, Analysis, and Control
- Frontiers in Compound Semiconductors

Contributed poster papers are solicited on new breakthroughs and major improvements in measurement techniques for silicon or compound semiconductors. Authors are requested to submit a one-page abstract by September 23, 1994 to the Conference Contact below. An author's kit will be mailed following paper acceptance.

Sponsors: The Advanced Research Projects Agency, SEMATECH, National Institute of Standards and Technology, Army Research Office, U.S. Department of Energy, National Science Foundation, and SEMI.

Conference Chair: David G. Seiler, NIST

For information, contact:
Jane Walters, NIST
B344 Technology Bldg.
Gaithersburg, MD 20899-0001
Phone: 301/975-2050
Fax: 301/948-2081
e-mail: walters@sed.eeel.nist.gov

APPLIED DIAMOND CONFERENCE 1995

3rd International Conference on the
Applications of Diamond Films and Related Materials
August 21—24, 1995
National Institute of Standards and Technology
Gaithersburg, Maryland 20899, USA

Sponsor: National Institute of Standards and Technology

Cosponsors: American Physical Society, Materials Research Society

Cooperating Societies: American Carbon Society, American Ceramic Society, ASM International®, IEEE—Electron Devices Society, Japan New Diamond Forum, SPIE

OBJECTIVES: Major mechanical, thermal, optical, electronic, medical, and chemical applications of diamond, diamond-like carbon, cubic boron nitride, C-N compounds, and related wide-bandgap superhard materials may soon be realized due to the rapid progress being made in the requisite processing technologies. The objective of this biennial international conference is to allow manufacturers and end-users of diamond and related technologies to interact with scientific researchers for the purpose of identifying technical barriers that are hindering the large scale commercial applications of these materials and research strategies for overcoming these barriers. Topics to be covered include but are not limited to:

- acoustics
- active electronics
- bearings
- biomedical implants
- cutting tools
- deposition optimization
- economic analyses
- electronic packaging
- fabrication processes
- heat management
- low temperature growth
- optoelectronics
- oriented growth
- performance evaluation
- polishing
- process monitoring
- protective coatings
- scale-up
- sensors
- wear resistant surfaces
- windows, lenses, domes

CALL FOR PAPERS: Papers are solicited both on practical applications and on the basic science needed for removing impediments to applications of CVD diamond, diamond-like carbon, cubic boron nitride, C-N compounds, and related wide-bandgap superhard materials. A number of contributed papers will be chosen for oral presentation and the remainder will be presented at poster sessions. Each contributing author is expected to submit the following material:

— An abstract up to one page long, due **January 30, 1995**. Please provide 5 copies.

— A manuscript, four pages long, in camera—ready form, for publication in the proceeding, due April 15, 1995. Instructions for manuscript preparation will be sent after receipt of the abstract. Manuscripts of all accepted papers will be published as submitted in the Conference Proceedings which will be distributed at the conference.

Authors will be notified whether abstracts have been accepted by February 21, 1995. Notification will be sent to the author listed first on the abstract unless alternate instructions are provided in a letter accompanying the abstract. It would be helpful to have available the telephone number, FAX number and/or e-mail address of the contact author. You may include this in your letter.

TECHNOLOGY DEMONSTRATIONS: Demonstrations of products, processing equipment, and prototype specimens that show progress in the development of applications are encouraged, although sales promotions are not permitted at NIST. For further information, please contact the conference chairman.

ADDRESS ALL CORRESPONDENCE TO: Albert Feldman, Chairman ADC'95, National Institute of Standards and Technology, A329 Materials Building, Gaithersburg, MD 20899, USA.
FAX: 301/990-8729; email: feldman@micf.nist.gov

NIST Technical Publications

Periodical

Journal of Research of the National Institute of Standards and Technology—Reports NIST research and development in those disciplines of the physical and engineering sciences in which the Institute is active. These include physics, chemistry, engineering, mathematics, and computer sciences. Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Institute's technical and scientific programs. Issued six times a year.

Nonperiodicals

Monographs—Major contributions to the technical literature on various subjects related to the Institute's scientific and technical activities.

Handbooks—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications—Include proceedings of conferences sponsored by NIST, NIST annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

Applied Mathematics Series—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

National Standard Reference Data Series—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NIST under the authority of the National Standard Data Act (Public Law 90-396). NOTE: The Journal of Physical and Chemical Reference Data (JPCRD) is published bimonthly for NIST by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements are available from ACS, 1155 Sixteenth St., NW, Washington, DC 20056.

Building Science Series—Disseminates technical information developed at the Institute on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

Technical Notes—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NIST under the sponsorship of other government agencies.

Voluntary Product Standards—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NIST administers this program in support of the efforts of private-sector standardizing organizations.

Consumer Information Series—Practical information, based on NIST research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

Order the **above** *NIST publications from: Superintendent of Documents, Government Printing Office, Washington, DC 20402.*

Order the **following** *NIST publications—FIPS and NISTIRs—from the National Technical Information Service, Springfield, VA 22161.*

Federal Information Processing Standards Publications (FIPS PUB)—Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NIST pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

NIST Interagency Reports (NISTIR)—A special series of interim or final reports on work performed by NIST for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Service, Springfield, VA 22161, in paper copy or microfiche form.

Reproduced by NTIS

National Technical Information Service
Springfield, VA 22161

This report was printed specifically for your order from nearly 3 million titles available in our collection.

For economy and efficiency, NTIS does not maintain stock of its vast collection of technical reports. Rather, most documents are printed for each order. Documents that are not in electronic format are reproduced from master archival copies and are the best possible reproductions available. If you have any questions concerning this document or any order you have placed with NTIS, please call our Customer Service Department at (703) 487-4660.

About NTIS

NTIS collects scientific, technical, engineering, and business related information — then organizes, maintains, and disseminates that information in a variety of formats — from microfiche to online services. The NTIS collection of nearly 3 million titles includes reports describing research conducted or sponsored by federal agencies and their contractors; statistical and business information; U.S. military publications; audiovisual products; computer software and electronic databases developed by federal agencies; training tools; and technical reports prepared by research organizations worldwide. Approximately 100,000 *new* titles are added and indexed into the NTIS collection annually.

For more information about NTIS products and services, call NTIS at (703) 487-4650 and request the free *NTIS Catalog of Products and Services*, PR-827LPG, or visit the NTIS Web site **http://www.ntis.gov**.

NTIS
Your indispensable resource for government-sponsored information—U.S. and worldwide

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Technical Information Service
Springfield, VA 22161 (703) 487-4650