APPENDIX B

OFFSETS OF POSTERIOR CDFs

TABLE B.1
U.S.-FLAG TANKER SPILLS ASSOCIATED WITH FAULTY OPERATIONAL PROCEDURES AND EQUIPMENT

Volume	Posterior CDF
1.60000	9.95216 E-2
2.80000	0.154429
4	0.193674
5.20000	0.235314
6.40000	0.266934
7.60000	0.291755
8.80000	0.319302
10	0.341062
21.2500	0.400946
32.5000	0.541761
43.7500	0.611060
55	0.688032
66.2500	0.699450
77.5000	0.716021
88.7500	0.737446
100	0.755910
212.500	0.855623
325	0.875157
437.500	0.910731
550	0.934288
662.500	0.943408
775	0.957785
887.500	0.959255
1000	0.965161
2125	0.985260
3250	0.990307
4375	0.993343
5500	0.995030
6625	0.995111
7750	0.996331
8875	0.997336
10000	0.997738
21250	0.999110
32500	0.999425

TABLE B.2
U.S.-FLAG TANKER SPILLS ASSOCIATED WITH HULL RUPTURES

Volume	Posterior CDF
1.60000	0.171480
2.80000	0.284374
4	0.358902
5.20000	0.412327
6.40000	0.455018
7.60000	0.485370
8.80000	0.511917
10	0.534232
21.2500	0.649928
32.5000	0.702953
43.7500	0.735402
55	0.758008
66.2500	0.774982
77.5000	0.788366
88.7500	0.799291
100	0.808444
212.500	0.857544
325	0.879302
437.500	0.892473
550	0.901613
662.500	0.908455
775	0.913841
887.500	0.918235
1000	0.921913
2125	0.941646
3250	0.950403
4375	0.955716
5500	0.959405
6625	0.962173
7750	0.964354
8875	0.966135
10000	0.967627
21250	0.975659
32500	0.979240
43750	0.981419
55000	0.982936
66250	0.984075
77500	0.984975
88750	0.985710
100000	0.986327

TABLE B.3
OFFSHORE PETROLEUM PRODUCTION SPILLS ASSOCIATED WITH WELL BLOWOUTS

Volume	Posterior CDF
10	2.00169 E-2
42	2.58768 E-2
100	3.15226 E-2
420	4.82893 E-2
1000	6.88876 E-2
4200	0.153330
10000	0.243708
42000	0.423133
100000	0.531746
420000	0.698138
1000000	0.785080
4200000	0.893572
10000000	0.931876
42000000	0.960075
100000000	0.970637
4.20000 E+8	0.981629
1.00000 E+9	0.985852
4.20000 E+9	0.991027
1.00000 E+10	0.993525

TABLE B.4
OFFSHORE PETROLEUM PRODUCTION SPILLS NOT ASSOCIATED WITH WELL BLOWOUTS

Volume	Posterior CDF
10	1.46037 E-3
42	1.89041 E-2
100	0.283014
420	0.883106
1000	0.975429
4200	0.998511
10000	0.999973
42000	0.999984
100000	0.999997

TABLE B.5
PIPELINE-ASSOCIATED OIL SPILLS

Volume	Posterior CDF
42	4.29310 E-2
100	0.281036
420	0.768353
1000	0.904968
4200	0.980002
10000	0.992268
42000	0.998353
100000	0.999340
420000	0.999856
1000000	0.999942
4200000	0.999988
10000000	0.999994

APPENDIX C

A SUMMARY OF THE STATUS OF THE BAYESIAN HYPOTHESIS TEST TECHNIQUE

The test developed for this report for the selection of the sampling function family underlying the observed volume distribution is an heuristic extension of the classical Bayesian hypothesis test (Zellner, 1971, page 297). In its classical form, this test assumes the form:

$$P(H_j|\underline{x}) = \frac{P(H_j)\int f(\underline{x}|\underline{\theta}_j,\ H_j)f(\underline{\theta}_j|H_j)\ d\underline{\theta}_j}{f(\underline{x})} \qquad (1)$$

where

$P(H_j|\underline{x})$ is the posterior probability that hypothesis H_j is correct;

$P(H_j)$ is the prior probability on H_j;

$F(\underline{x}|\underline{\theta}_j,\ H_j)$ is the likelihood of sample $\underline{x}$ conditional on hypothesis H_j and parameter values $\underline{\theta}_j$;

$f(\underline{\theta}_j|H_j)$ is the prior distribution on $\underline{\theta}_j$ conditional on H_j; and,

$f(\underline{x})$ is the posterior probability on the sample $\underline{x}$. It is equal to:

$$f(\underline{x}) = \sum_i p(Hi)\int_{\underline{\theta}i} f(\underline{x}|\underline{\theta}i,\ Hi)f(\underline{\theta}i|Hi)d\underline{\theta}i \qquad (2)$$

In the present study, we assumed the $f(\underline{\theta}i|Hi)$ were given by improper priors of the forms $d\theta i$ or $\frac{d\theta i}{i}$ for the infinite and semi-infinite parameter range cases respectively. This leads to nothing more than philosophical problems when the object is to calculate a posterior on $\underline{\theta}i$ given $\underline{x}$ and Hi.

However, in the problem at hand, (1), these improper priors result in a fundamental indeterminancy.

To resolve this difficulty, the hypothesis test used here inserts the posterior distribution on $\underline{\theta}_j$, $f(\underline{\theta}_j|\underline{x}\ H_j)$, for $f(\underline{\theta}_j|H_j)$ in equation (1). Since the posterior pdf on $\underline{\theta}_j$ is a proper density function, the indeterminancy disappears, but now we are using the sample, $\underline{x}$, twice in our equations. Further, this form of the hypothesis test is not derivable from any set of probability axioms that we are familiar with. Thus our substitution is an ad hoc measure, and posterior hypothesis probabilities based on using $f(\underline{\theta}_j|\underline{x},\ H_j)$ in (1) should be regarded as statistics subject to interpretation, rather than as probabilities in the strict sense.

In order to test the inferential power of this statistic, we subjected our ad hoc posterior probabilities to extensive numerical tests. These tests were reported in earlier drafts of Appendix C. We found that it behaved properly for samples drawn from the two parameter Gamma, Lognormal and Inverse Gamma families provided the distributions were not taken from that part of the family's parameter space where the function became assymptotically Normal. This was a qualification of little interest to us because we knew the oil spill data would rarely exhibit Normal volume distributions. In short, we found the statistic to have good inferential powers in identifying underlying distributions like those found in oil spill records. The use of the statistic as a probability, however, remains speculative since it is not theoretically based. Thus posterior cumulative volume distributions composed of weighted sums of the family types, like the blowout CDF, are somewhat speculative. We are unaware of any other options, however, for treating small data sets like that found in the blowout problem. On this basis, we decided to adopt the test for use in this report.

Because of the ad hoc nature of the procedure, we have continued to work on the concept in hopes of finding a more elegant and theoretically justifiable approach. This has led us to several very promising ideas, and we are actively pursuing the problem (MBK and JWD at MIT and RJS at PMEL). Since the method is novel, its presentation must necessarily incorporate both the results underlying this report and our most recent thoughts, which have yet to be proven numerically. We do not, therefore, include a more complete description here, although we hope to have the material published soon.

APPENDIX D

A PRELIMINARY EXTREME VALUE ANALYSIS OF PLATFORM AND PLATFORM SPILLS (1964-1975)

An alternative approach for characterizing oil spill volumes to the Bayesian technique adopted here is the extreme value methodology. This methodology is concerned with the analytical properties of the distribution of the largest random variate observed in a sample of fixed and/or very large size. For example, if y is the largest of n exponentially distributed random variates, x, then y may be shown to have the following probability density function:

$$f_Y(y) = n(1-e^{-y})^{n-1} e^{-y} \tag{D1}$$

where $y = \max(x_1, x_2, \ldots x_n)$

and $f_X(x) = e^{-x}$ (D2)

(Gumbel, 1958, chapter 4)

When n becomes large, the form of $f_Y(y)$ is most strongly influenced by the behavior of the upper tail of the distribution on x since the x's will usually come from this region. There are three assymptotic forms of $f_Y(y)$ that are associated with exponential, geometrical, and maximum value limited tail distributions on x respectively.

Corresponding to the probability density function $f_Y(y)$, shown above, is the cumulative distribution function

$$F_Y(y) = [1-e^{-y}]^n \tag{D3}$$

We may expand this distribution about the modal value of (D1),

$y_m = \ell n\ n$, by multiplying e^{-y} in (D3) by $\frac{e^{y_m}}{n}$ (which equals unity), or

$$F_Y(y) = 1 - \frac{e^{-(y-y_m)}}{n}^{\ n} \text{, and} \tag{D4}$$

then noticing that in the limit, $n \to \infty$, this becomes

$$\lim_{n\to\infty} F_Y(y) = \exp[-e^{-(y-y_m)}] \tag{D5}$$

This is a Type I extreme value distribution and it is the characteristic form for all extreme value distributions based on populations with an exponential tail (e.g., Gamma, Normal, and Lognormal families all have tails of the form $e^{-g(x)}$.) In the same fashion, assymptotic distributions may also be derived for the geometrical and maximum value limited tails.

The important thing to notice in (D3) and (D5) is that the distribution is dependent on n, even in the limit $n \to \infty$, although in this limit all n does is locate the distribution about some modal value. That is, the shape of $F_Y(y)$ assymptotically converges on the form (D5). Even for an n like 50, (D5) is a reasonable approximation to (D3)(e.g., (D3) yields .36714 and (D5) yields .36788 for n = 50 and y = $\ell n 50$.)

Since extreme value distributions deal only with the largest of 'n' samples, they fit naturally into an analysis of low-end censored data (data which disregards observations smaller than a lower cutoff). One need only specify a large enough n to achieve a high degree of certainty that y will be greater than the censoring threshold. As such, the extreme value technique offers some hope of improving the spill volume distributions for platforms and pipelines. As pointed out in the report these spill sources were not well modeled by the Bayesian method. We suggested that this might be due to the known censoring. There is one drawback, however, and that is that the maximum values must come from samples that are of approximately equal size. In earthquake magnitude calculations, this is achieved by using the largest observation in a period of time under the hypothesis that a large and approximately equal number of tremors will occur in that time. For maximum waveheight calculations, the maximum value is again picked out of a fixed time period, usually of order two hours, under the assumption that about 1000 wave crests should be observed in that time.

If we had a year by year record of how many small oil spills occurred but were not recorded, then it would be possible to find periods of time in which similar or equal numbers of oil spills occurred. It would then be a simple matter to find the largest spill in the Event file for that period and proceed. Unfortunately, this information is not available over any significant period. The only alternative is, therefore, to estimate these periods of time from the data at hand, mainly the Platform File construction records, and our knowledge of the spill generation process, which holds platform years to be a useful exposure parameter. On this basis, it is possible to estimate that there were about 18,500 cumulative platform-years in the period January 1964 through December 1975 on U.S. leases in the Gulf of Mexico. If we break this into ten, 1850 platform-year segments, we find the following maximum spill sizes for each of these periods, Table D1. Also shown in this table are the maximum pipeline spills, although we have no way of supporting an assertion that a similar number of pipeline spills occurred in each period.

The simplest way to test this data for its extreme value properties is to plot it on extreme value probability paper (Gumbel, pgs. 177 and 263, 1958). Figures D1 and D2 show this data on such paper. If the data was generated by a process with an exponential tail, then the curve in Figure D1 would be a straight line. Since it isn't we can rule out Type I processes. Figure D2 is slightly better and shows that the process might be modeled with a Type II process, one of which is the Inverse Gamma. Unfortunately the fit is still not very good, mainly due to the absence of extreme values in the range 500 to 5000 barrels. This gap may be real or it might be due to inaccuracies in the reporting of the volume spilled* or it might be due to the particular selection of time intervals on which the data is based. Such speculations,

---

* We might hypothesize, for example, that the larger spill will be more carefully researched and thus more accurately reported while the small spills are more likely to be estimated by guess. Further, it might well be in the estimators interest to report 100 or 200 BBLS versus 1000 or 2000 BBLS.

TABLE D1

Period from	Period to	Max. Platform Spill (BBL)	Max. Pipeline Spill (BBL)
1/64	9/65	5,180 (5)*	<50
10/65	3/67	<50	65 (8)
4/67	6/68	85 (10)	160,639 (9)
7/68	10/69	250 (15)	7,532 (14)
11/69	12/70	53,000 (26)	50 (24)
1/71	3/72	200 (27)	80 [10]
4/72	3/73	9,935 (40)	100 (39)
4/73	12/73	240 (43)	5,000 (42)
1/74	12/74	130 (46)	17,883 (44)
1/75	12/75	100 (54)	414 (55)

*Incident number in ( ) Table D
[ ] Table C

Accidents connected with Federal Oil and Gas Operations on the Outer Continental Shelf. USGS Conservation Division, July 1976.

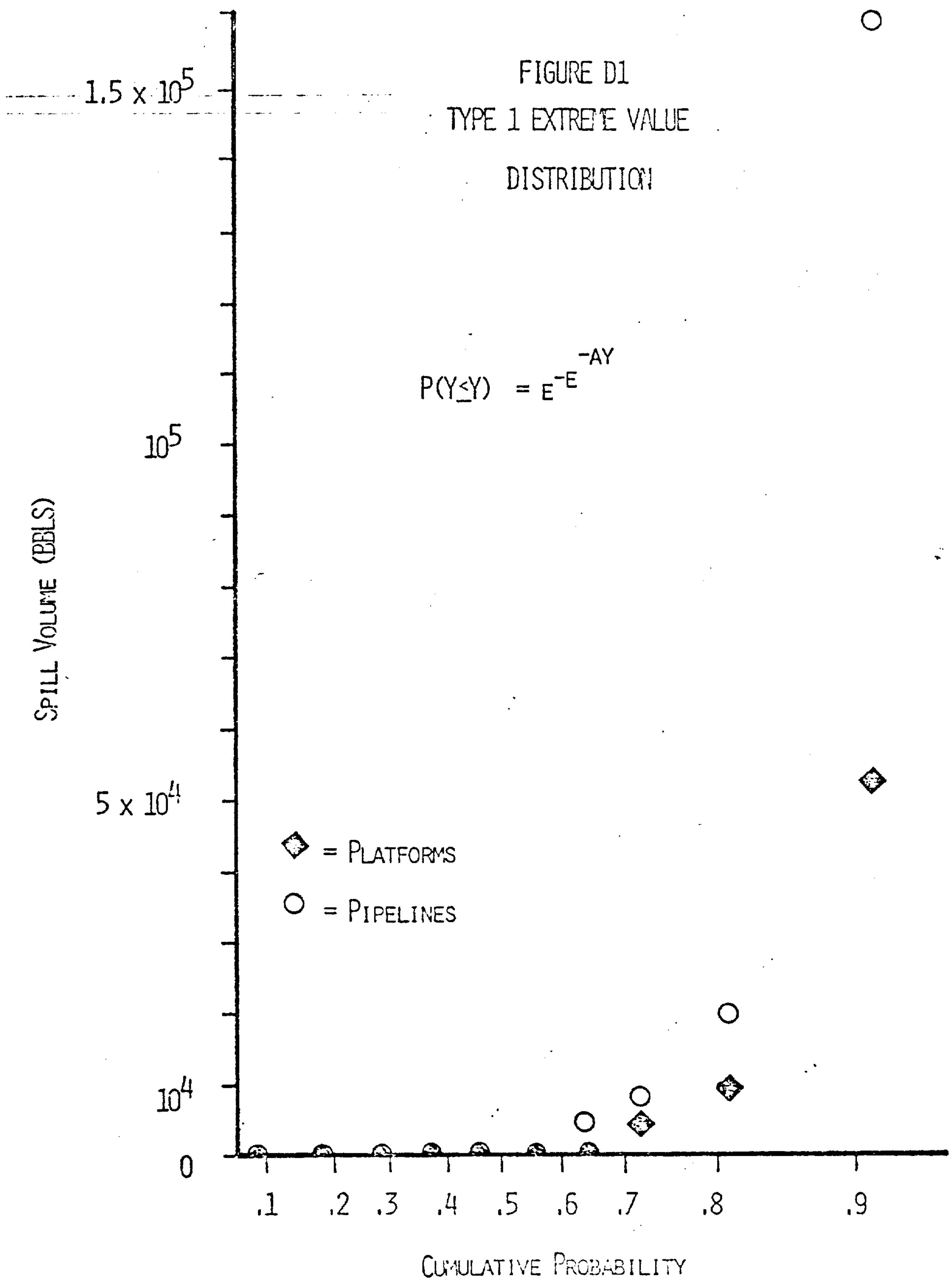
FIGURE D1
TYPE 1 EXTREME VALUE
DISTRIBUTION
P(Y≤Y) = E^-E^-AY
1.5 x 10^5
10^5
5 x 10^4
10^4
0
SPILL VOLUME (BBLS)
= PLATFORMS
= PIPELINES
.1
.2
.3
.4
.5
.6
.7
.8
.9
CUMULATIVE PROBABILITY

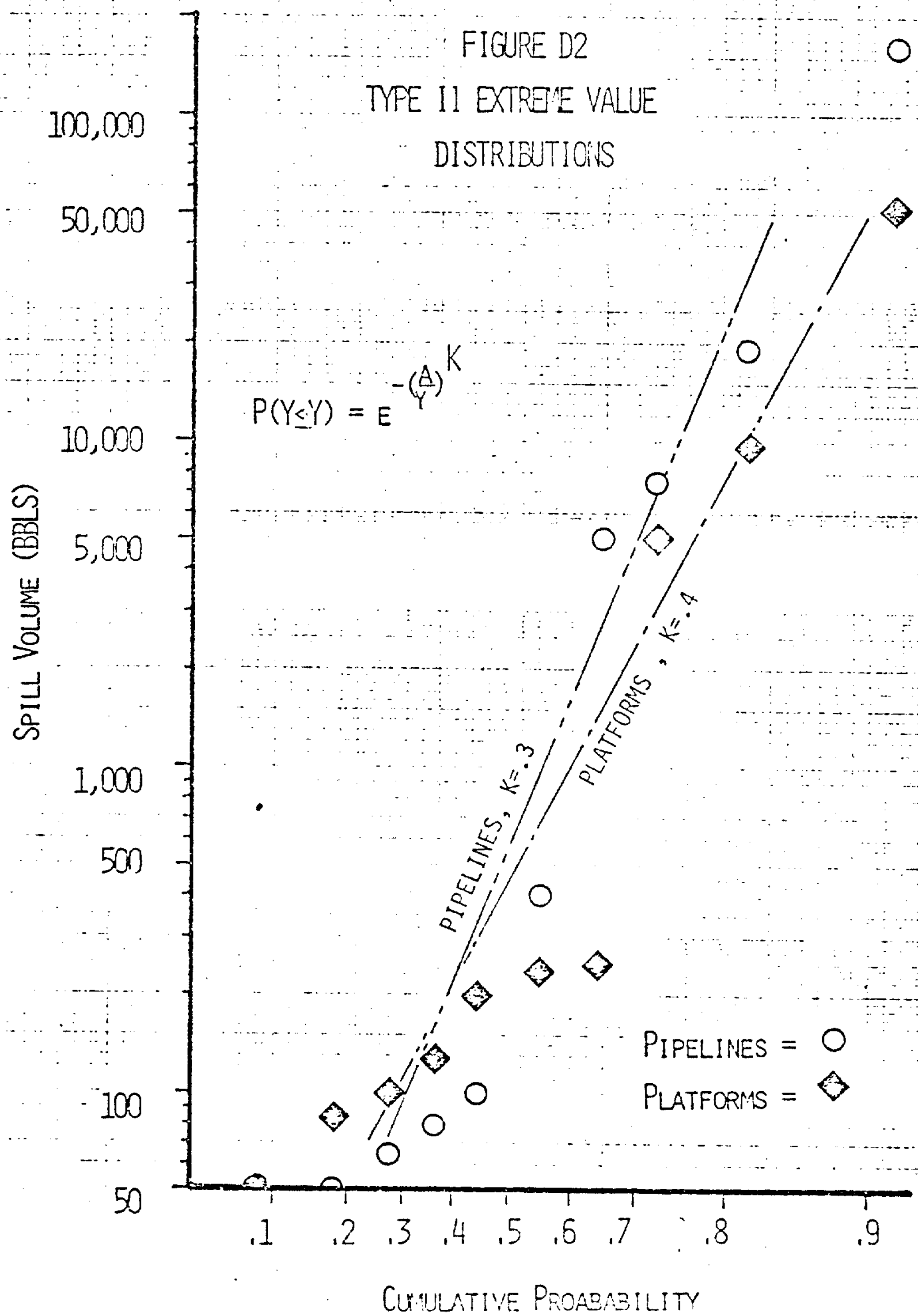

FIGURE D2
TYPE II EXTREME VALUE DISTRIBUTIONS

however, will not resolve the ambiguity, and Figure D2 in its present form would not seem to warrant further analysis.

We conclude, therefore, that an extreme value analysis of the 1964-1975 spill data does not offer much hope of significantly improving the platform and pipeline spill volume cumulative distributions. In our opinion, significant improvement can only come from an analysis of the data that specifically incorporates the censoring in its formulation, or through use of the existing methodology on uncensored data. With regard to the latter point, we should point out that the Bayesian methodology used in the body of the report only requires two or three "sufficient" statistics for each hypothesized distribution, and these statistics can be sequentially modified and do not require the retention of complete data on the incident. For example, if we simply had the number, the sum and the product of all spills less than 1 BBL from platforms, we could use the uncensored Gamma; inclusion of the sum of the reciprocals of the volume spilled in this set of statistics would allow inclusion of the uncensored Inverse Gamma; and so on. Records for these small spills could therefore be kept in tabular form, with a running total for each spill classification of interest.

BIBLIOGRAPHY

Benjamin, J.R., and Cornell, C.A. (1970) Probability, Statistics and Decisions for Civil Engineers. McGraw-Hill Book Co., New York, NY, pp 684.

Gumbel, E.J. (1958) Statistics of Extremes. Columbia University Press, New York, pp 375.

U.S. Geological Survey, Conservation Division. (July 1976) "Accidents Connected with Federal Oil and Gas Operations on the Outer Continental Shelf" USGS, Reston, VA.