Analysis of the Oil Spills from Tanker Ships

Ringo Ching and T. L. Yip

The Hong Kong Polytechnic University
香港理工大學

Department of Logistics & Maritime Studies
物流及航運學系

CY Tung International Centre for Maritime Studies
董浩雲國際海事研究中心

The Data

Included accidents in which International Oil Pollution Compensation (IOPC) Funds were involved, up to October 2009

In this study the spill amounts in tonnes recorded in 1992 Fund and 1971 fund are combined according to cases

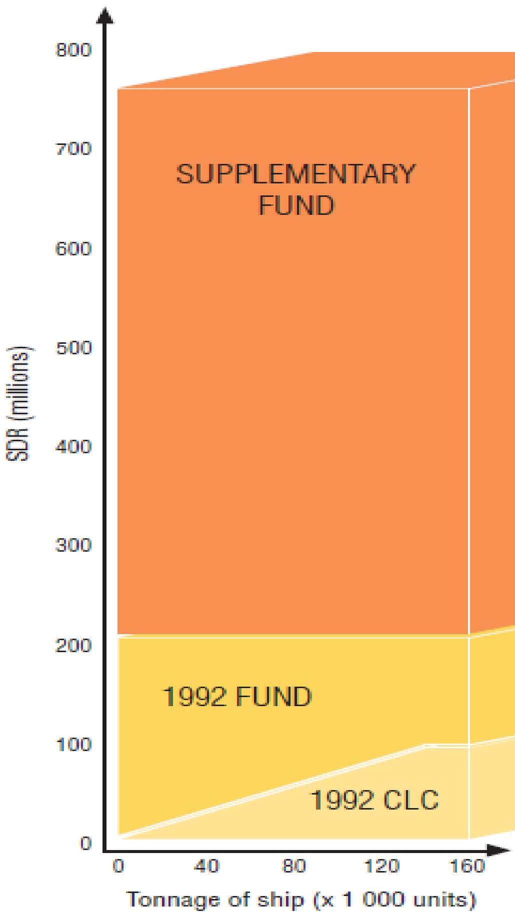
Background

According to annual report of the IOPC fund :

The 1992 Fund pays compensation when:

- the damage exceeds the limit of the shipowner's liability under the 1992 Civil Liability Convention,
- the shipowner is exempt from liability under the 1992 Civil Liability Convention,
- the shipowner is financially incapable of meeting his obligations in full under the 1992 Civil Liability Convention, and the insurance is insufficient to pay valid compensation claims.

Background



Maximum limits laid down in respect of the 1992 Civil Liability and Fund Conventions and the Supplementary Fund Protocol

An overall average does not fully describes the situation, especially for this compensation fund which is responsible for the major spills in excess of the liability limit of ship owners

Background

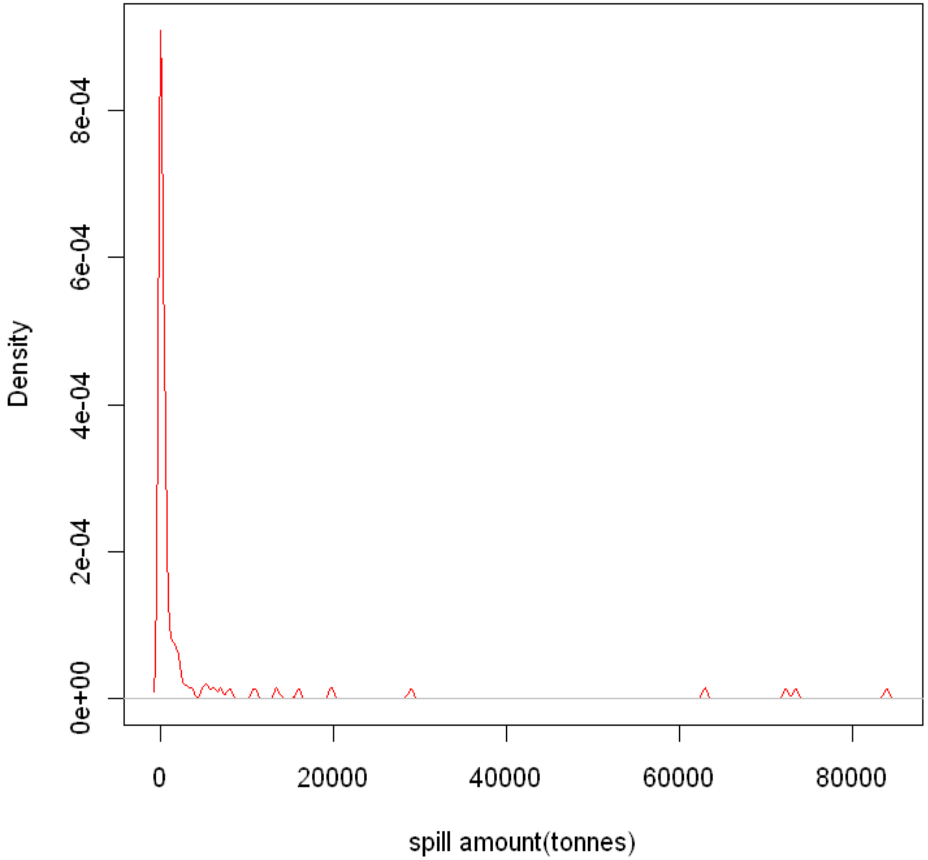
Precise analysis should be done on the larger major spills

If premium is too large, the cost of the business will be increased unreasonably and lower the profit

If the premium was too low, the fund will go bankrupt, the risk sharing mechanism would not work

Accurate estimation would lead to more reasonable premium, making the fund more efficient

Summary Statistics



- Year 1979-2008
- Number of accidents=105, expected value = 4296.99
- Maximum = 84000
- Skewness= 4.43, Kurtosis= 19.46 (normal distribution has skewness = 0; kurtosis = 3)

Most spills are small in amount while some spills are in another extreme

Fitting with a single distribution

Weibull and lognormal distributions are fitted to the spilled amount

Log-Likelihood of fitted lognormal: -785.72

Log-Likelihood of fitted Weibull: -791.17

Observed spill amount: 4282.11 tonnes

Expected spill amount: 11731.70 tonnes

Expected vs observed: 173.9% error

Single distribution is not working well

Possible solution: 2 distributions

Peaks-Over-Threshold Method

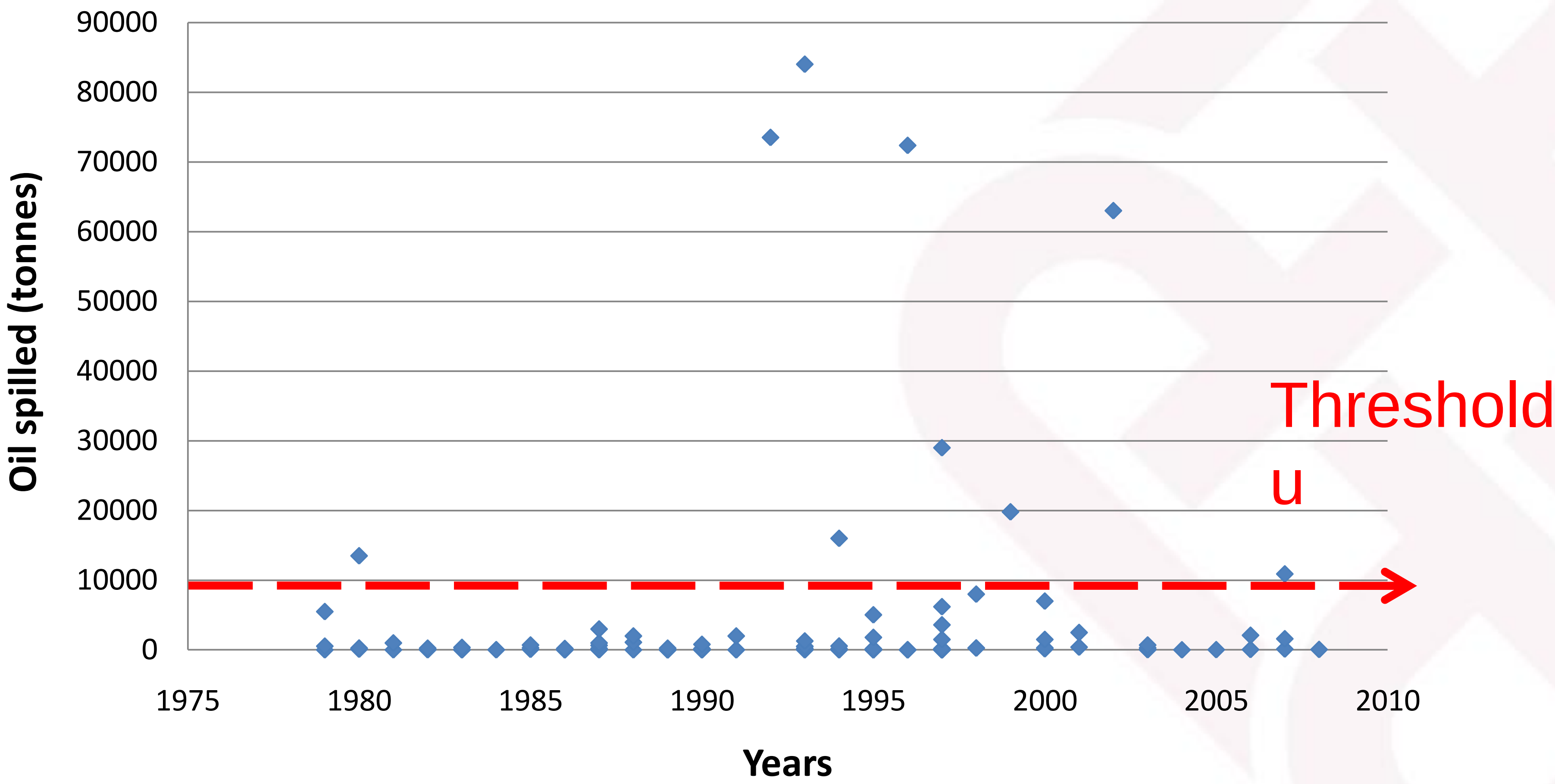
A method used widely in field of hydrology and insurance

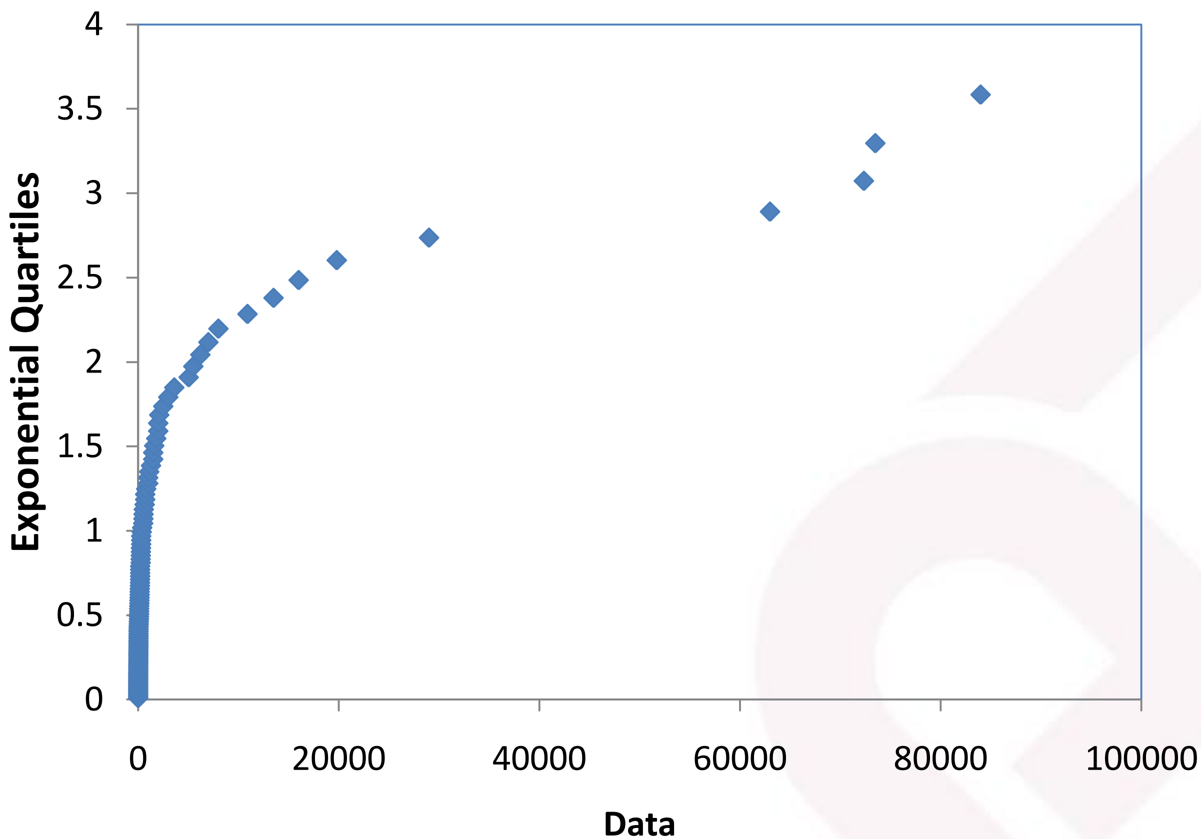
Our random variable X would be the spill amount in tonnes

The approximate distribution F(x) of those X larger than u, would be generalized Pareto distribution (GPD) [1]:

$$G_{\xi,\sigma,u}(x) = \begin{cases} 1 - \left(1 + \frac{\xi\,[x-u]}{\sigma}\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp\left(\frac{-[x-u]}{\sigma}\right) & \xi = 0 \end{cases}$$

Peaks-Over-Threshold Method





Quartile-plot against exponential distribution (GPD, ξ=0)

The plot was obtained by matching observed data to the exponential distribution. Since it is not linear the data cannot be modeled by exponential distribution (GPD with ξ=0)

Peaks-Over-Threshold Method

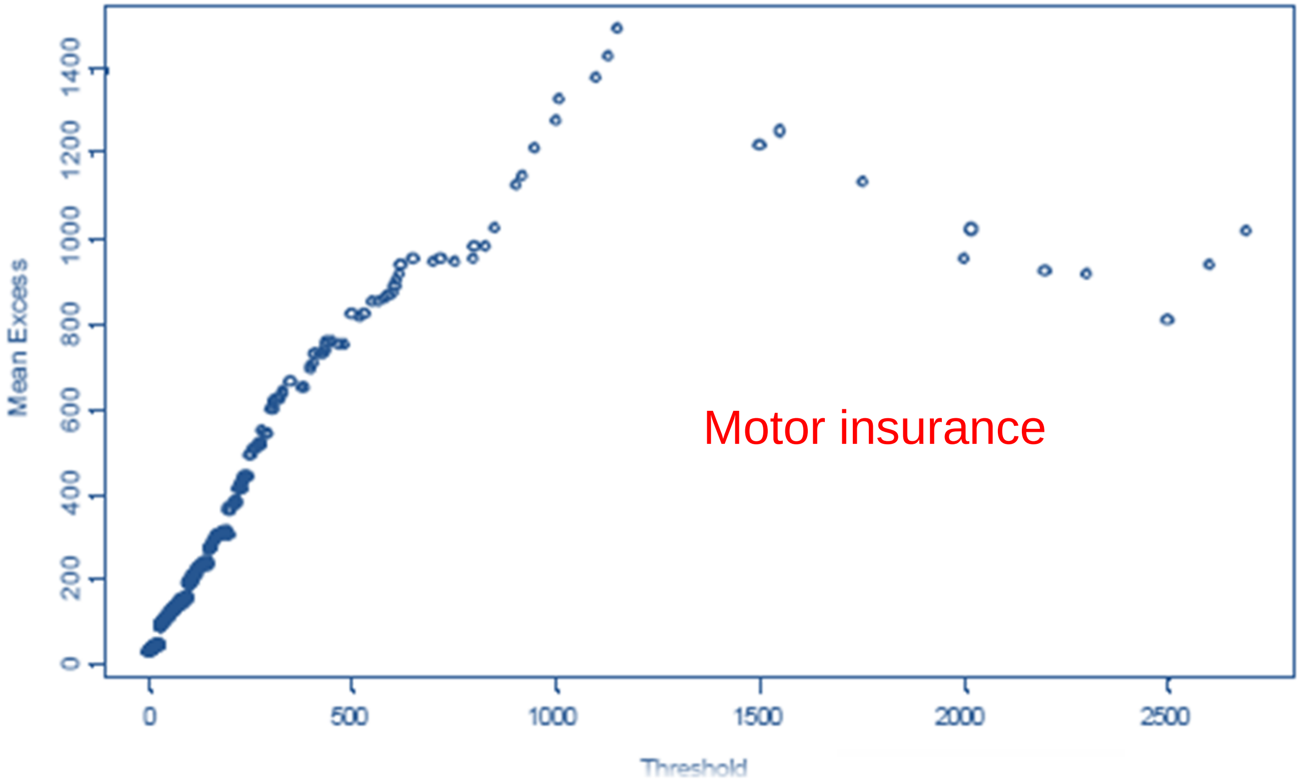
For GPD, if we keep on rising our threshold R larger than the suitable threshold u, the average value of those spills minus R (mean excess) would increase linearly with R with slope

$$\frac{\xi}{1-\xi}$$

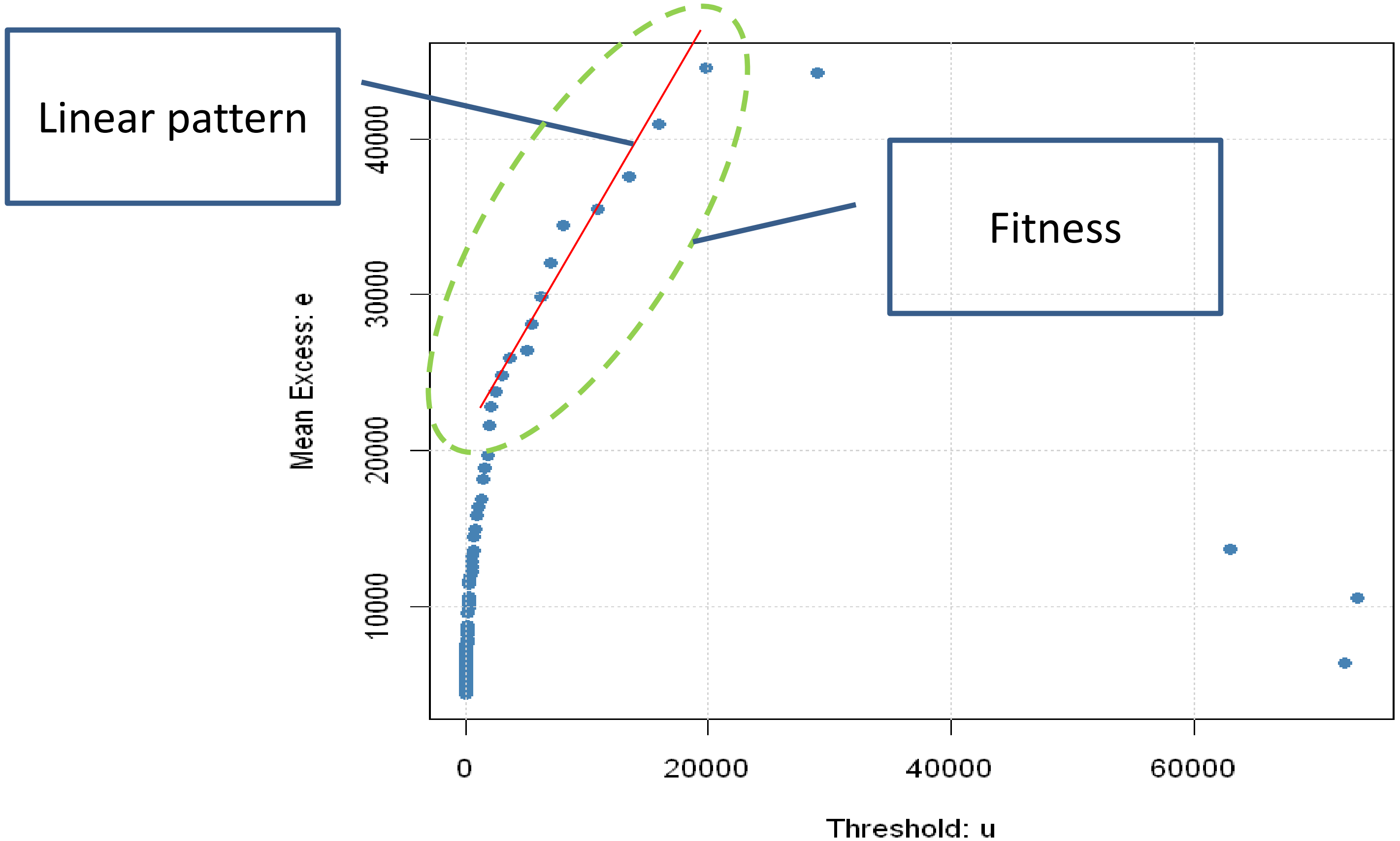
$$E(X - R \mid X > R) = \frac{\sigma + \xi(R - u)}{1 - \xi}, \quad R \geq u$$

Peaks-Over-Threshold Method

An example would be claim data from motor insurance portfolio consists of 172,161 policies, studied by P. Gigante, L. Picech and L. Sigalotti[2]



CY TUNG
International Centre for Maritime Studies
董浩雲國際海事研究中心



Observed mean excess function of spill amount

The linear pattern after reaching the threshold 6000 shows that the data can be modeled by GPD with threshold at around 6000

Peaks-Over-Threshold Method

In other words, the major spill amount can be modeled by GPD by choosing a high enough threshold u

The overall spill amount is represented by 2 distributions, with the GPD responsible for the large spill

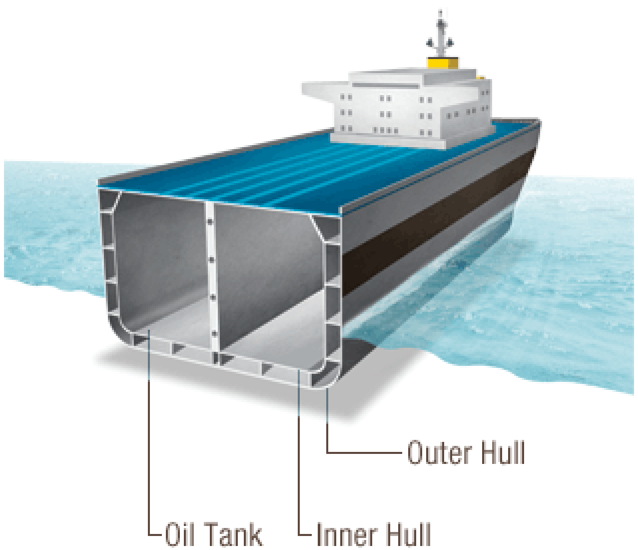
Results

Castillo and Hadi [3] compared the methods for estimating the generalized Pareto distribution. They suggested that for small sample, probability weighted moment method should be employed when there is reason to believe $0 \leq \xi \leq 0.5$

From the linear part of empirical mean excess function, its slope $\frac{\xi}{1-\xi}$ is positive, such that $\xi \geq 0$ and it is approximately 0.12

Results

GPD has finite expectation and variance if and only if ξ is smaller than 0.5. As the amount of spill is limited by the capacity, the expectation and variance of the spill amount should be finite



Results

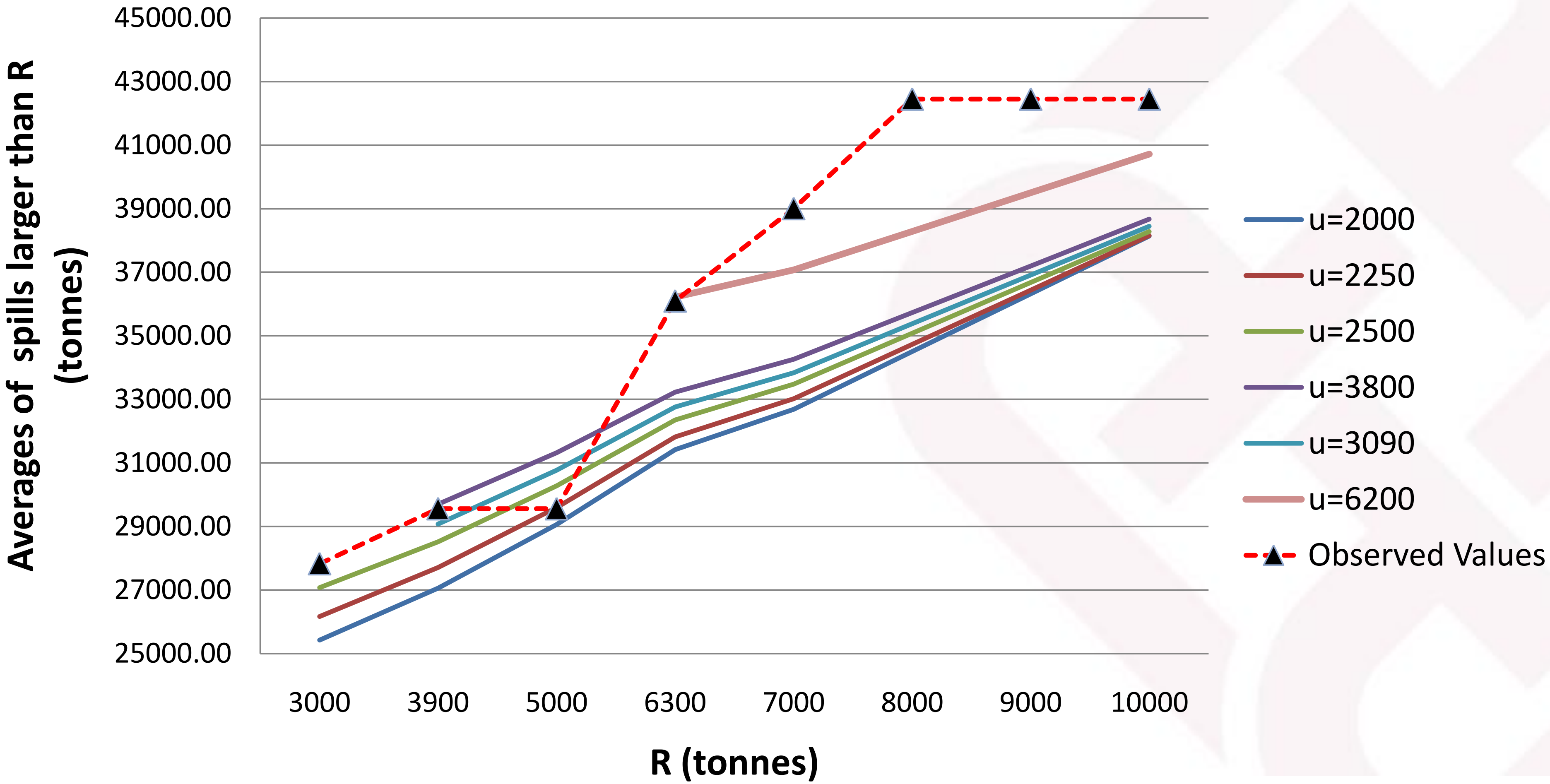
Thresholds u (tonnes)	No. of exceedances	ξ	σ	Average (tonnes) of those spills larger than : 3900	6300	7000	8000	10000
3090	15	0.3489	16110	29076.79	32762.86	33837.96	35373.83	38445.55
3800	14	0.3197	17520	29700.34	33228.19	34257.15	35727.09	38666.97
5700	12	0.2650	20510		34421.09	35373.47	36734.01	39455.10
5900	12	0.2769	20030		34153.35	35121.41	36504.34	39270.21
6100	12	0.2888	19560		33884.03	34868.28	36274.35	39086.50
6200	11	0.1780	24575		36218.54	37070.12	38286.67	40719.76
6500	11	0.1954	23814			36717.77	37960.57	40446.19
6800	11	0.2128	23063			36350.42	37620.70	40161.25
7000	10	0.0751	29603				40087.18	42249.53
7200	10	0.0864	29057				39881.68	42070.86
7500	10	0.1034	28247				39563.69	41794.44
Observed Values				29556.79	36096.36	39006.00	42451.11	42451.11

The Hong Kong Polytechnic University 香港理工大學

Department of Logistics & Maritime Studies 物流及航運學系

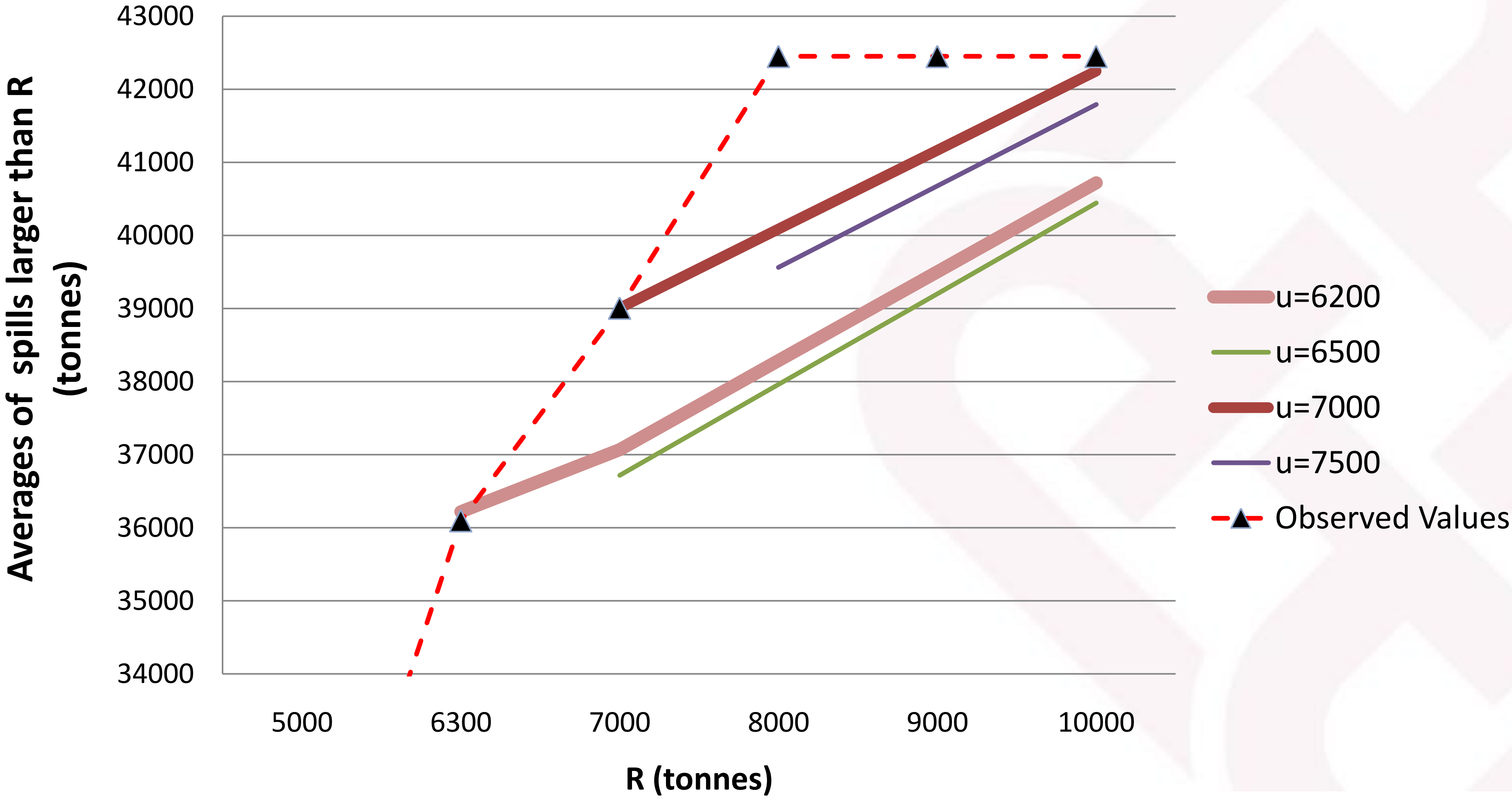
CY TUNG International Centre for Maritime Studies 董浩雲國際海事研究中心

Results

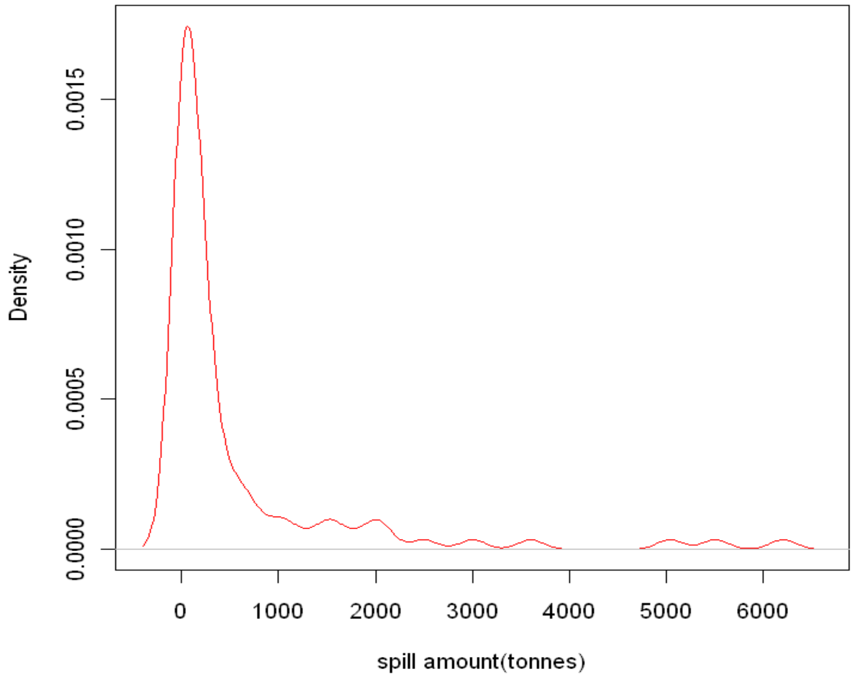


45000.00
43000.00
41000.00
39000.00
37000.00
35000.00
33000.00
31000.00
29000.00
27000.00
25000.00
Averages of spills larger than R (tonnes)
3000
3900
5000
6300
7000
8000
9000
10000
R (tonnes)
u=2000
u=2250
u=2500
u=3800
u=3090
u=6200
Observed Values

Results


43000
42000
41000
40000
39000
38000
37000
36000
35000
34000
Averages of spills larger than R (tonnes)
5000
6300
7000
8000
9000
10000
R (tonnes)
u=6200
u=6500
u=7000
u=7500
Observed Values

Results



Thresholds 6200 and 7000 would be compared.

From the density graph of the spills less than 6300, weibull, gamma and lognormal distributions were fitted to these smaller spill amounts

The Log-Likelihood of the fitted distributions:

Thresholds (tonnes)	Weibull	Gamma	Log-normal
6200	-628.01	-970.4774	-631.2586
7000	-640.7031	-1043.872	-643.5014

The Hong Kong Polytechnic University 香港理工大學

Department of Logistics & Maritime Studies 物流及航運學系

CY Tung International Centre for Maritime Studies 董浩雲國際海事研究中心

Results

Hypothesis tests were conducted on the overall fitness of the mixture distributions

The Kolmogorov-Smirnov (KS) Test is based on the maximum difference between the observed distribution $F_n(x)$ and estimated distribution F(x) [4]:

$$\sup_x |F(x)-F_n(x)|$$

The Anderson-Darling (AD) Test is a modification which puts more weight on the large data:

$$n \int_{-\infty}^{\infty} \frac{[F(x)-F_n(x)]^2}{F(x)[1-F(x)]} dF(x)$$

The Hong Kong Polytechnic University 香港理工大學

Department of Logistics & Maritime Studies 物流及航運學系

CY TUNG International Centre for Maritime Studies 董浩雲國際海事研究中心

Results

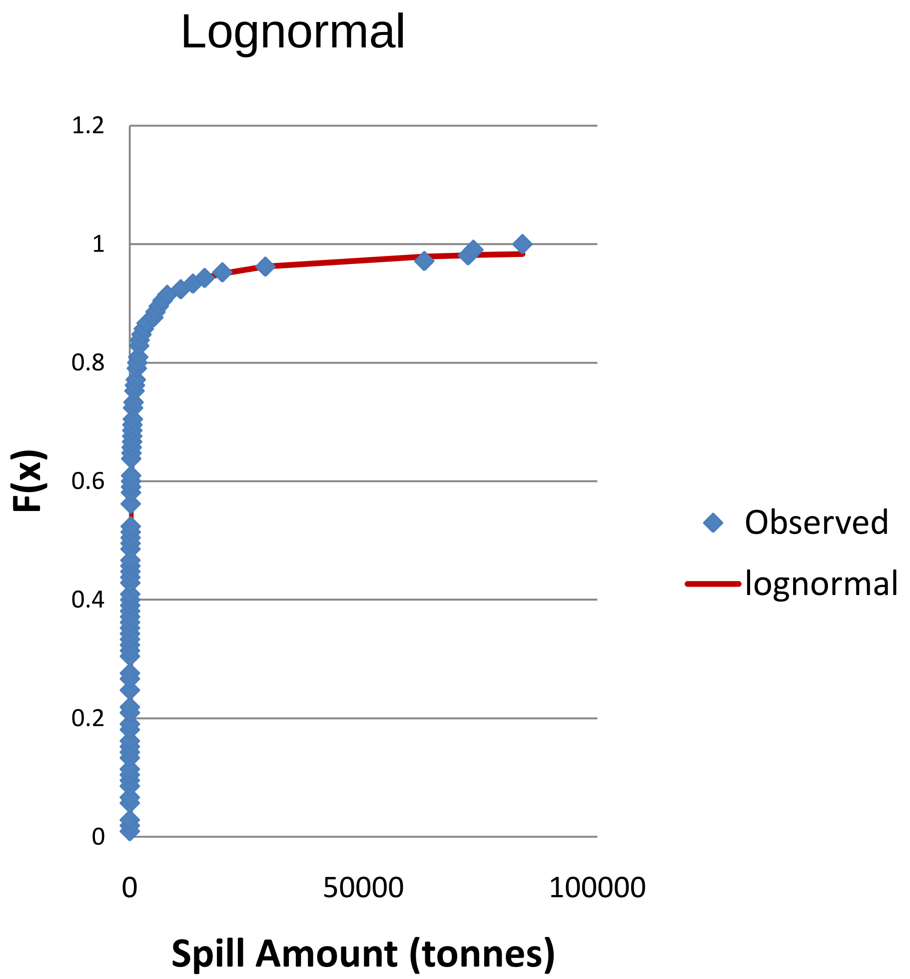
Threshold (tonnes)	Kolmogorov-Smirnov (KS) Test	Anderson-Darling (AD) Test
6200	0.0530	0.2257
7000	0.0565	0.2317
Critical values (5% level of significance)	0.1327	2.492

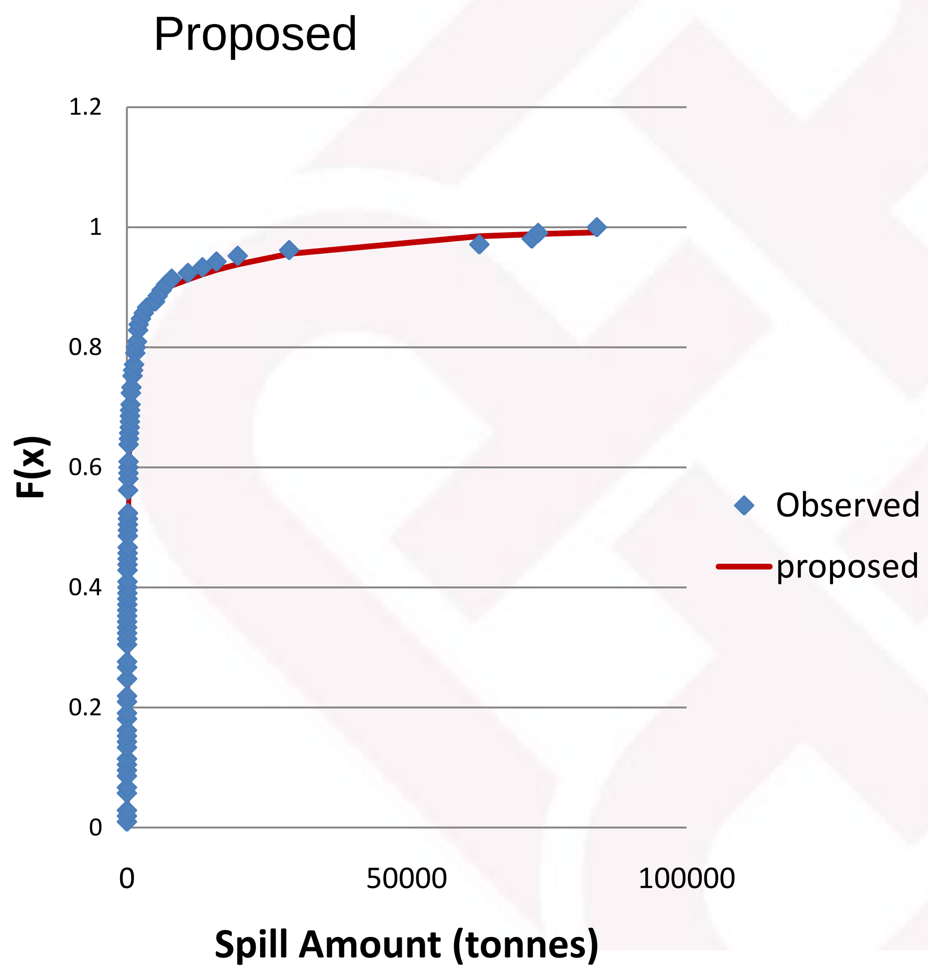
The distribution with threshold 6200 have a slightly better fit to the observed spill amounts

Average spill amount given by this proposed distribution is 4307.08 tonnes with 0.58% percentage error

A log-normal distribution gives an estimate with 173.9% percentage error

Implication



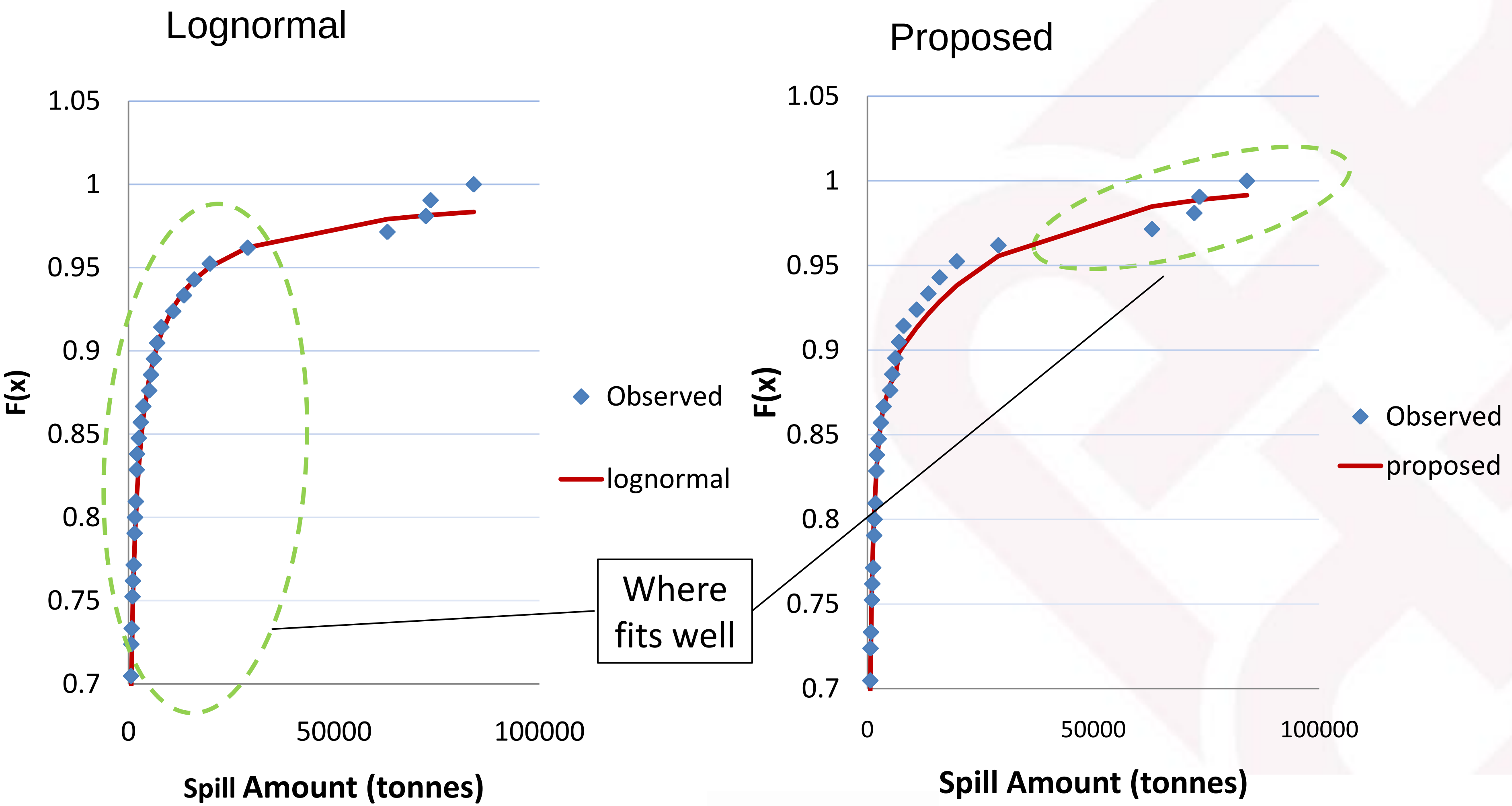


THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Department of LOGISTICS & MARITIME STUDIES
物流及航運學系

CY TUNG International Centre for Maritime Studies
董浩雲國際海事研究中心

Implications



Department of LOGISTICS & MARITIME STUDIES 物流及航運學系

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

CY TUNG International Centre for Maritime Studies 董浩雲國際海事研究中心

Implications

We further compare the performance of a single lognormal and the proposed distribution through hypothesis tests. The test statistics are given below

	Kolmogorov-Smirnov (KS) Test	**Anderson-Darling (AD) Test**
Log normal	0.0433	0.2386
Proposed	0.0530	0.2257
Critical values (5% level of significance)	0.1327	2.492

The proposed distribution performs better when placing more emphasis on the large data

Implications

From the prospective of funds, the estimated average spill amount larger than a level would be put to test

	Average (in tonnes) of those spills larger than :				
	3000	**3900**	**6300**	**8000**	**10000**
Observed	26524.68	28053.00	33671.67	42451.11	42451.11
Log normal	78776.71 (197%)	89943.78 (216%)	116155.39 (245%)	132772.86 (213%)	151009.2 (256%)
GPD (u=6200)			36218.54 (7.56%)	38286.67 (-9.81%)	40719.76 (-4.08%)

Where the percentage errors compared with the observed values are in blankets

Implications

- Through separate treatment of the larger spill amounts with Peak-Over-Threshold method, a more accurate distribution for extreme oil spill data is obtained
- This distribution can be used by funds which are responsible for accidents exceeded the liability limit of ship owners to determine more reasonable premium, making the whole business more efficient

References

- [1] Pickand, J. (1975) “Statistical inference using extreme order statistics” Annals of Statistics, vol 3(1), pp.119-131
- [2]Gigante, P. Picech, L. and Sigalotti, L. (2002) “Rate making and large claims” in XXXIIIrd Astin Colloquium, Match 21-22, 2002, Mexico
- [3] Castillo, E. and Hadi, A.S. (1997) “Fitting the Generalized Pareto Distribution to Data” Journal of American Statistical Association, vol 92(440), pp. 1609-1620
- [4]Lai, L.H. and Wu, P.H. (2008) “Estimating the threshold value and loss distribution: Rice damaged by typhoons in Taiwan” African Journal of Agricultural Research, vol 3(12), pp.818-824

The Overall Distribution

A mixture distribution can be used for the overall spill amount, with the GPD responsible for the larger spill amounts (X>R), the expectation would thus be given as

$$E(X) = P(X \leq R)E(X|X \leq R) + P(X > R)E(X|X > R)$$

E(X|X>R) given by GPD would be

$$E(X|X > R) = \frac{\sigma + \xi(R - u)}{1 - \xi} + R, \quad R \geq u$$

Appendix

Suggested by Castillo and Hadi [2]:

1. If the sample size is large (>500) and it is believed that 0.5> ξ >-0.5, maximum likelihood estimation (MLE) method would be preferred

2. If sample size is not large and it is believed that 0.5 > ξ > 0, probability weighted moment method (PWM)should be used

3. In all other cases, used elemental percentile method (EPM)

4. In all cases, if MLE has convergence problems or if PWM gives nonsensical estimates, then use EPM

Appendix

Probability weighted moment method:

$$\alpha_s = E[\,X\{1 - F(X)\}^s] = \frac{\sigma}{(s+1)(s+1+\xi)}$$

$$\widehat{\alpha_s} = n^{-1}\sum_{j=1}^{n}(1 - p_{j:n})^s x_{j:n}$$

where $x_{j:n}$ is the jth largest sample and $p_{j:n} = \frac{(j - 0.35)}{n}$

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Department of LOGISTICS & MARITIME STUDIES
物流及航運學系

CY TUNG International Centre for Maritime Studies
董浩雲國際海事研究中心