Maritime Transport Research 3 (2022) 100056

Contents lists available at ScienceDirect

Maritime Transport Research

journal homepage: www.sciencedirect.com/journal/maritime-transport-research

Marine insurance claims analysis using the Weibull and log-normal models: Compensation for oil spill pollution due to tanker accidents

Ringo Ho Fai Ching [a], Tsz Leung Yip [b,*]

[b] Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[a] STAT.com Limited

ARTICLE INFO

Keywords:
Marine insurance
Oil spill
Maritime safety
Weibull distribution
Log-normal distribution
IOPC

ABSTRACT

In marine insurance, the identification of the distribution of the claim size processes from given observations is of major importance. We evaluate the claim amounts due to accidental oil spill volume from tankers. Previous studies on oil spills were mainly concentrating on the relation between spill amounts and some accident factors. The trend of the spill claims is a major concern to the sustainability of insurance funds, tanker shipping and the whole oil industry. By analysing the real claim data series of IOPC (International Oil Pollution Compensation Funds), we identify the trends and reveal findings which enable policy-makers and insurers for the sufficient financial protection from oil spill damage. This study introduces the Weibull model and the log-normal model to the framework for oil spill claims. The Weibull model and the log-normal model are obtained with Markov chain Monte Carlo (MCMC) simulation method. The statistical properties of oil spill claims are determined in the time series of oil spill claims. Moreover, the relations among claim properties of different settings are examined. The log-normal of 3 groups of gross tonnage is found the best fit according to R-squared statistics and the mean squared errors.

1. Introduction

Oil spills are the most tangible evidence of maritime pollution from shipping activities. They may occur as a result of human or mechanical error during cargo-handling, pumping, refuelling, tank cleaning or ballasting operations, as well as through maritime accidents, e.g., collisions or groundings. International conventions have been developed to provide compensation for victims of oil pollution damage, which regulate civil liabilities resulting from pollution incidents. The overall purpose of the conventions is to establish an international regime on a uniform base of liability for (1) damage done by pollution by escape or discharge of oil, and (2) cost of measures taken reasonably to mitigate such damage.

The first (lower) layer is provided by the International Convention on Civil Liability for Oil Pollution Damage (CLC), which is subject to a limitation determined by the ship's tonnage. The second (upper) layer is provided by the International Convention on the Establishment of an International Fund for Oil Pollution Damage ("the Fund Convention"). According to the Fund Convention, the International Oil Pollution Compensation Funds (IOPC) were established in 1972 and 1992, known as "*1972 Fund*" and "*1992 Fund*" respectively, to provide compensation for oil pollution damage. And IOPC is administrated by the International Tanker Owners

* Corresponding author.
E-mail addresses: chfringo1228@gmail.com (R.H.F. Ching), t.l.yip@polyu.edu.hk (T.L. Yip).

https://doi.org/10.1016/j.martra.2022.100056
Received 29 November 2021; Received in revised form 6 February 2022; Accepted 4 March 2022
Available online 15 March 2022

Pollution Federation (ITOPF). The funds are destined to compensate anyone who has been unable to obtain full and adequate compensation under the CLC.

The data is obtained from the document “Incidents involving the IOPC Funds” which includes information on the claims in two funds, namely “1992 Fund” and “1972 Fund”. According to the explanatory note released by IOPC fund in May 2011, 1992 Civil Liability Convection (CLC 1992) covers pollution damage suffered in territory, territorial sea or exclusive economic zone including preventive measures even when no spill occurs. The Flag State of the tanker and nationality of the shipowner are irrelevant for the scope of application. It establishes exemption of the liability of tanker owners in the following situations:

(1a) the damage resulted from an act of war or a rave natural disaster, or
(1b) the damage was wholly caused by sabotage by a third party, or
(1c) the damage was wholly caused by the negligence of public authorities in maintaining lights or other navigational aids.

In this study, it is assumed that exemption due to reason (1a) and (1b) above is rare and unless otherwise specified, the compensations would be due to reason (1c). In addition, the shipowner is entitled to limit his liability under the 1992 Fund convection as follows:

(2a) for a tanker not exceeding 5000 units of gross tonnage, 4510,000 Special Drawing Rights (SDR);
(2b) for a tanker with a tonnage between 5000 and 140,000 units of tonnage, 4510,000 SDR plus 631 SDR for each additional unit of tonnage; and
(2c) for a ship of 140,000 units of tonnage or over, 89,770,000 SDR.

If it is proved that the pollution damage resulted from the shipowners’ personal act or omission, committed with the intent to cause such damage, or recklessly and with knowledge that such damage would probably result, the shipowner is deprived of the right to limit his liability.

The 1992 Fund pays compensation to those suffering oil pollution damage in a State Party to the 1992 Fund Convention who does not obtain full compensation under the 1992 Civil Liability Convention for one of the following reasons:

(3a) the shipowner is exempted from liability under the 1992 Civil Liability Convention because he can invoke one of the exemptions under this Convention; or
(3b) the shipowner is financially incapable of meeting his obligations under the 1992 Civil Liability Convention in full and his insurance is insufficient to satisfy the claims for compensation for pollution damage; or
(3c) the damage exceeds the shipowner’s liability under the 1992 Civil Liability Convention.

The maximum amount payable by the 1992 fund was 135 million SDR before 1 November 2003, including the amount paid by the shipowner or his insurer. On 1 November 2003, the limited was increased to 203 million SDR and it applies to incidents after this date. An International Oil Pollution Compensation Supplementary Fund was established on 3 March 2005 for providing additional compensation over that available under 1992 Fund Convention. However, no incident has occurred has involved the Supplementary Fund.

The 1969 Civil Liability Convention was established in 1975 and is similar to 1992 Civil Liability Convention except that the 1969 Convention is limited to pollution damage suffered in the territory including the territorial sea) of a State Party to the Convention where threat removal measures, i.e., preventive measures is not accounted. Under the 1969 Civil Liability Convention, the limit of shipowners’ liability is 133 SDR per ton of ship’s tonnage or 14 million SDR whichever is lower. The 1971 Fund was entered ceased to be in force on 24 May 2002 and does not apply to incidents occurring after that date.

The proposed probabilistic models allow not only for the estimation of the amount of claim given a limited set of input variables, but most of them are also available in the Fund records. They can also assist in evaluating the required changes in the Fund in order to reach better management of the Fund. The models can be further used in the process of risk assessment of oil spills for the maritime policy focusing on the assessment of the required fund amounts. In principle, the models can be used in the process of risk-based fund management, linking the probability of accidental oil spills. The Weibull model and log-normal model are parametric models that may be used to extrapolate the financial liabilities in respect of oil pollution. The groups of data can be modelled for each component of risk.

In many studies related to oil spills, the main concerns are the probability of oil spills and the direct consequence of oil spills (e.g. the amount of oil spills). A recent review paper by Chen et al. (2019) discussed the historical oil spills and the current status of global governance. Talley et al. (2012) evaluated bunker spills from non-oil-cargo vessels empirically and found that the amount bunker spills from a non-oil-cargo vessel is approximately 1000 gallons fewer than that from a tank-barge. Yip et al. (2011) reconfirmed the effectiveness of double-hull design empirically such that double hulls can reduce the amounts of accidental oil spills by 20% and 62% in tank barge and tanker ships on average. Zhong et al. (2021) quantified the legal liability risk of oil spills in China.

The financial liability to oil spills is obvious, but the compensation claims of IOPC funds are often neglected Shahriari and Frost (2008). developed a regression model for clean-up costs based on 80 oil spill incidences reported in IOPC Yamada (2009). presented a

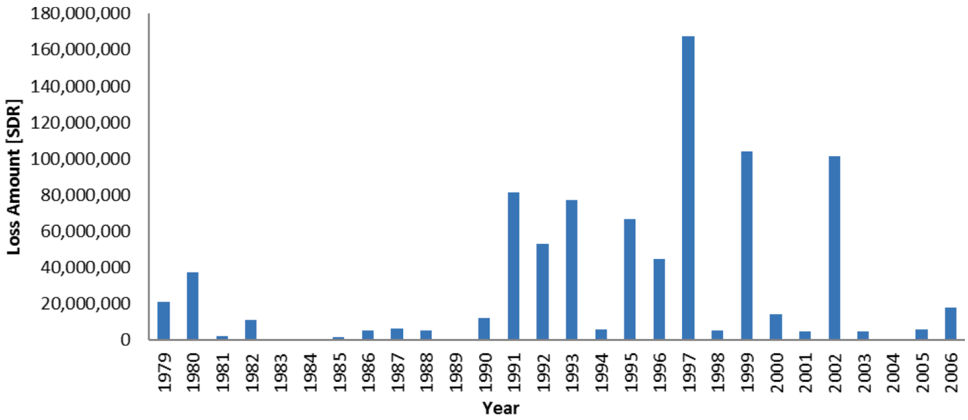

Fig. 1. Loss amounts over years.

regression analysis of historical oil spill data from IOPC. Kontovas et al. (2010) presented a regression analysis on the clean-up cost and total cost of IOPC funds data. Different from previous studies based on linear regression, this study attempts a more robust approach on how financial liability is estimated.

The reminder of the paper is organized as follows: Section 2 reviews the oil spill incidents and highlights the difficulties of modelling. In Section 3, the principle of probabilistic models is inspected and the results are compared Section 4. concludes this study.

2. Data of oil spill accidents

The raw data were gathered from IOPC funds, and the 142 incidents occurred in 1970 to 2008 were analysed. The raw data are not ready for analysis and are manipulated with the currency conversion into SDR (Special Drawing Rights). In estimating the actual total loss amount in SDR, the fund compensation amounts in pounds were converted to SDR versus the year-end exchange rate which is obtained from the “Datastream International (current & historical)”. After manipulation, 89 incidents across 1979 to 2006 are used in this analysis. The data was summarized in Fig. 1.

The liability limit, which varies versus the tonnage as mentioned above, is added to the converted fund compensation amount if there is no further information available. For example, Table 1 shows the summary of the case 31 of 1971 fund:

The clean-up claims RUB 1417,448 was added to the compensation after converting it to SDR. However, the currency rate of Ruble against SDR in year 1987 is not available from Datastream to approximate the total loss amounts. Instead referring to a publication of IMF (1991) the rate of Ruble against US dollars was obtained. The amount in US dollars was converted to SDR accordingly.

There are cases where the total damage did not exceed shipowners’ liability, for example, the case reported in Table 2.

Thus, the liability limit would not be added to the compensation amount.

If the shipowner (or his insurer) has paid for the damage, according to exemption (3b) the amount paid would be added to the fund compensation amount after converted back to SDR versus the rate in the year.

3. Model development

The probabilistic models to be developed are to support decision-making in the presence of uncertainty, by informing the policymakers and fund contributors about estimated claim amount of accidental oil spills. Two types of probabilistic models will be considered, (1) Weibull model and (2) log-normal model, with respect to oil spill amounts. The log-normal model and Weibull model are not popular statistical models. Herein, they are chosen to model extremes, because they can collectively characterize most important aspects of distributional tail behaviour and therefore these two models will make our analysis more general.

Studies have been revealed that the cost of oil spill is related to several factors White (2002). emphasized that and cost of spill has no relationship with the size of tanker because other factors in combination would give rise to a great variation to the cost. Etkin (1999) developed a per-tonnage clean-up cost estimation model based on factors such as location of the spill, type of the oil spilled and total amount spilled, without explanation on how the numbers are obtained. In our data on the estimated total loss amounts, gross tonnage would be accounted as it is known before accidents and thus can be used for predictions. In order to model the great variation, the cost of spills with respect to ships having same gross tonnage and time of sail, the cost amounts would be represented as distributions with parameters varies versus gross tonnages and the time of sail.

The frequencies of the accidents versus years and the gross tonnages (GT) are calculated, part of the result is as shown in Table 3:

Most of the times there are only one data point for each gross tonnage/year pair, it is impossible to obtain distributions according to a gross tonnage in a particular year. Instead, the data would be divided into groups versus GT and years, and the data fall within a

Table 1
A case when the 1971 fund is not responsible for.

Ship	Date of incident [YYYY.MM.DD]	Place of incident	Gross Tonnage [GT]	Compensation (and/or indemnification) paid by the 1971 Fund up to 22.10.2010 [£]	Notes
Antonio Gramsci	1987.02.06	Borgå, Finland	27,706	268,982	Clean-up claims in USRR (RUB 1417,448) not paid by 1971 Fund since USSR not Member of 1971 Fund at time of incident.

group would be assumed to follow a same distribution. The size of the groups would depend on the parameter uncertainty (parameter risk) of the resulting distributions. The grouping would cease when the parameter uncertainty becomes unreasonable, and this would be demonstrated during the development of the model.

Parameter uncertainty has been studied in the field of actuary Kreps (1997). mentioned that parameter uncertainty mostly comes from finite sample size Van Kampen (2003). described the idea that different sets of parameters could have produced the actual data and used simulation method to determine a viable parameter set for log-normal distribution.

In order to estimate boundaries versus gross tonnage and time and describe the **parameters uncertainty Bayesian hierarchical model** would be employed. It allows all unknown parameters to be regarded as random variables and the conditional distribution of the parameters based on known information (posterior distribution) can be evaluated. One method to evaluate Bayesian model is Markov chain Monte Carlo (MCMC) simulation, and OpenBUGS are employed for the MCMC estimation (Spiegelhalter, et al., 2014). The OpenBUGS are run from the R statistical platform.

The following procedure is adopted to determine the probabilities. The estimated total loss amounts exhibit a long tail behaviour, where incidences of small losses occur frequently, and an incidence of extreme large loss amounts exists. This statistical property can be observed through the upward trend of the mean excess plot on the total loss amounts, in which the mean excess is approximated as

$$e(u) = \frac{\sum_{x_i > u}(x_i - u)}{Number\ of\ x_i > u} \tag{1}$$

The plot was obtained through R package "**fExtremes**", and x_i represents the estimated total loss amounts. There are 89 incidences of loss amounts included. This plot confirms the use of long tail distributions like Weibull and Log-normal distribution, which are discussed in the following sections.

3.1. Weibull model of grouped data

First, the Weibull distribution is employed. According to Nelson (2004), one property of Weibull distribution is the linear relationship under a change of variable. We use the default Weibull distribution of OpenBUGS:

$$F(x) = 1 - \exp\left[-\lambda x^k\right] \tag{2}$$

The Weibull plot is:

$$ln\{-\ln[1 - F(x)]\} = k\ln(x) + \ln(\lambda) \tag{3}$$

The scale parameter is defined as:

$$Scale = \left(\frac{1}{\lambda}\right)^k \tag{4}$$

With n data points, $F(x_i)$ can be approximated as $\widehat{F(x_i)} = \frac{i-0.3}{n+0.4}$ for the ith data. As the first step of analysis, a Weibull plot is drawn on all the loss amounts (Fig. 2).

Fig. 2 shows that the Weibull plot, the data agrees with the Weibull distribution with an R-squared statistic of 0.9135. Except for a few incidents of small loss amounts, most of them shows a linear relationship with $ln\{-ln[1 - F(x)]\}$ and so Weibull distribution is a plausible model for the loss amounts:

Table 2
A case when the damage is below the liability.

Ship	Date of incident [YYYY.MM.DD]	Place of incident	Gross tonnage [GT]	Compensation (and/or indemnification) paid by the 1971 Fund up to 22.10.2010 [£]	Notes
AgipAbruzzo	1991.04.10	Livorno, Italy	98,544	635,290	Total damages less than shipowners' liability.

Table 3
The frequencies of accidents occurred versus GT and years.

				Year											
GT	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	2002	2003	2005	2006	Total
19									1						1
31															1
38						1									1
43															1
48					1										1
56															1
57															1
59															1
68							1								1
74															1
82					1										1
86										1					1
89															1
93															1
94															2
…															
107				1											1

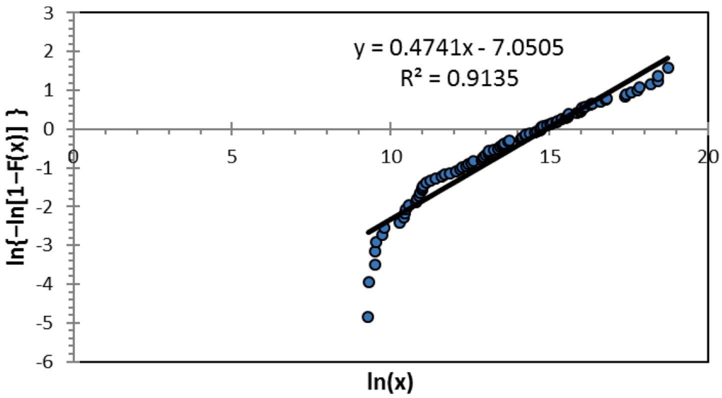

Fig. 2. Weibull plot on all the loss amounts.

$$x_i \sim Weibull(k, \lambda) \tag{5}$$

such that $k \sim Gamma(4,\ 8.5)$ and $\lambda \sim Gamma(4,\ 4\ 611)$.

The hyper-parameters have been set so that the expectation of the Weibull parameters is approximately equal to those obtained from the Weibull plot. The estimation result is shown in Table 4:

A total of two MCMC chains with 100,000 iterations each are performed. For each chain, the first 50,000 iteration results are discarded to ensure convergence, and Gelman-Rubin statistics is checked if it equals to 1 after 50,000 iterations. This practice is kept throughout the rest of the analysis. The mean of the Weibull distribution with the generated parameters is calculated, whereas the mean squared error (MSE) represents the squared difference between the mean and the loss amounts. If the distribution is suitable to describe the data, the squared difference between the data and the mean should be minimized. The MSE is thus a measure of how well the simulated distribution fits the data. The fluctuation on the mean is due to parameter uncertainty. SD stands for the standard deviation of the variable. The columns "*val2.5pc*", "*median*", and "val97.5pc" represent 2.5th, 50th and 97.5th percentile of the variables' posterior distribution respectively. Percentage error or% error is the percentage of the SD (standard deviation) over the mean of the variable, and it is a comparable measure of the parameter uncertainty as it eliminates the effect of the magnitude on the variation.

A plot of the loss amounts versus the ships' gross tonnage is in Fig. 3. From Fig. 3, most loss amounts are associated with smaller

Table 4
Weibull distribution for single group of incidents.

	mean	SD	val2.5pc	median	val97.5pc	% error
K	0.4516	0.02173	0.4111	0.4508	0.4961	4.81
scale	3.51×10^6	8.36×10^5	2.19×10^6	3.40×10^6	5.45×10^6	23.85
mean	8.74×10^6	2.32×10^6	5.30×10^6	8.37×10^6	1.43×10^7	26.57
MSE	5.63×10^{14}	1.01×10^{13}	5.57×10^{14}	5.60×10^{14}	5.82×10^{14}	1.79

Note: SD = standard derivation; MSE = mean squared error; start = 50,000; sample = 100,000.

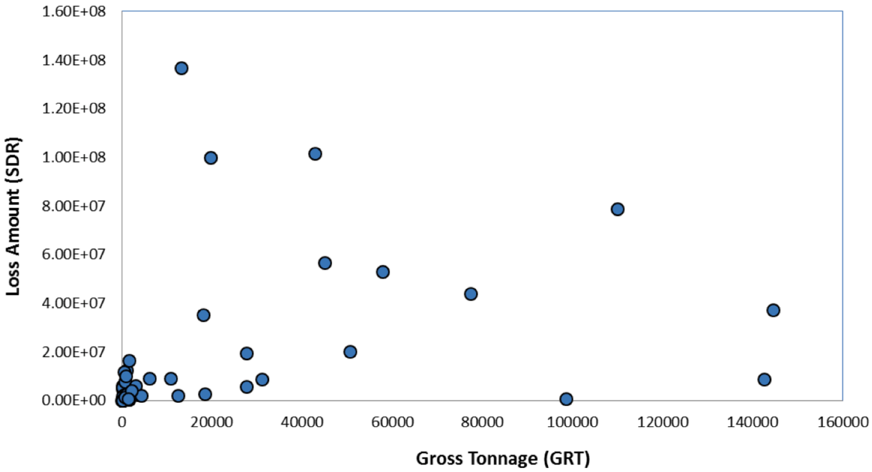

Fig. 3. Loss amounts versus gross tonnage (All Ships).

ships, and the fluctuation of the loss amounts is larger for ships larger than 6000 GT. But this boundary is not clear and ranges from 1000 GT to 12,000 GT.

A closer look at the loss amounts associated with smaller ships is provided by Fig. 4. The fluctuation of loss amounts associated with ships ranges from 400 GT to 1000 GT is larger than those smaller than 400 GT. As an initial investigate the data is grouped into three categories: one for gross tonnage smaller than 400 GT, the other for gross tonnage between 400 GT to 6000 GT. To check if the Weibull distribution is the suitable distribution for the data in these two groups, Weibull Plots are performed on these three data sets, as shown in Fig. 5a–c.

Loss amounts associated with ships larger than 6000 GT have a linear Weibull plot, while those amounts associated with ships smaller than 6000 GT are not that obvious. Most data give R-squared statistic close to 0.9 in Weibull plots. Weibull distribution is worth trying and the two boundaries (g.change1, and g.change2) should become parameters:

$$x_i \sim Weibull(k_g, \lambda_g) \quad (6)$$

where

$$g = \begin{cases} 1 & if\ GT < g.change1 \\ 2 & if\ g.change1 < GT < g.change2 \\ 3 & if\ GT > g.change2 \end{cases} \quad (7)$$

$$k_g = \begin{cases} k_1 \sim Gamma(0.56,\ 1) \\ k_2 \sim Gamma(0.88,\ 1) \\ k_3 \sim Gamma(0.78,\ 1) \end{cases} \quad (8)$$

$$\lambda_g = \begin{cases} \lambda_1 \sim Gamma(4,\ 5\ 600) \\ \lambda_2 \sim Gamma(4,\ 1\ 500\ 000) \\ \lambda_3 \sim Gamma(4,\ 1\ 310\ 000) \end{cases} \quad (9)$$

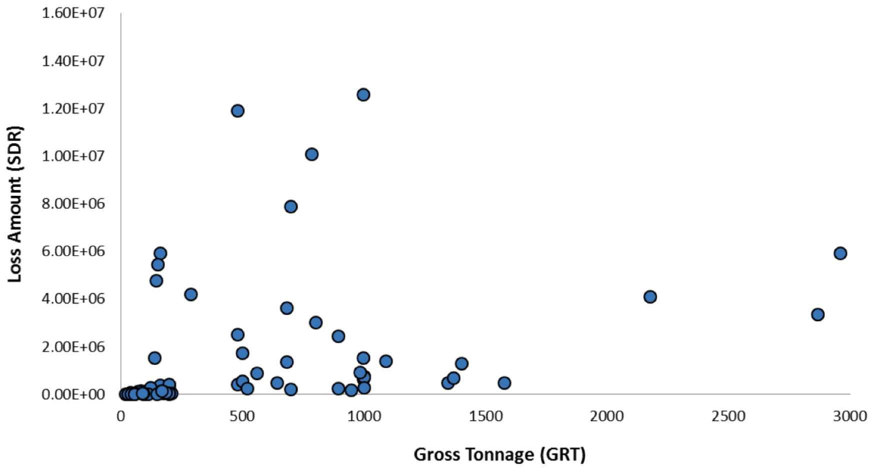

Fig. 4. Loss amounts versus gross tonnage (GT < = 3000).

in which

$$g.change = \begin{cases} \text{g.change1} \sim \text{uniform}(300,\ 900) \\ \text{g.change2} \sim \text{uniform}(900,\ 14\,000) \end{cases} \tag{10}$$

and

$$Scale_g = \left(\frac{1}{\lambda_g}\right)^k \tag{11}$$

The estimation result is in Table 5.

As concluded in previous MCMC simulation, first 50,000 iteration results are discarded, and Gelman-Rubin statistics is checked to ensure convergence. After grouping the MSE of the model is decreased at the cost of the higher parameter uncertainty. A plot of the mean of simulated Weibull distribution obtained before and after the grouping has provided further information.

In Fig. 6, the dotted lines represent the mean and its percentiles of the Weibull distribution estimated before grouping, while the solid lines represent the mean and its percentiles of the Weibull distribution estimated after grouping. The incidents are approximately divided into three GT groups: (1) smaller than 400 GT, (2) between 400 GT and 8510 GT; and (3) larger than 8510 GT. The grouping leads to distinctive estimates of Weibull means, as the interval of the Weibull means after grouping does not contain the estimate made previously except near the boundaries. Thus, the grouping is appropriate.

There are about 20 incidents fall into Group 3. Further dividing Group 3 in two groups would give one single group having fewer than 10 incidents and produce unreliable estimates such that the interval of the Weibull mean is far from that without grouping. A closer study on those incidents is presented in Section 4.

In Fig. 7, there is a larger fluctuation on the 19 incidents of ships bigger than 100 GT than the remaining 16 incidents. We thus perform Weibull plots on those two groups.

The R-squared statistics after grouping is improved compared with the results shown in Fig. 5a. Together with the argument above, these suggest fitting the loss amounts with further grouping through Bayesian model:

Fig. 8.

Eqs. (7), (8), (9), and (10) become, respectively:

$$g = \begin{cases} 1 & if\ GT < g.change1 \\ 2 & if\ g.change1 < GT < g.change2 \\ 3 & if\ g.change2 < GT < g.change3 \\ 4 & if\ GT > g.change3 \end{cases} \tag{12}$$

$$k_g = \begin{cases} k_1 \sim Gamma(1.19,\ 1) \\ k_2 \sim Gamma(0.54,\ 1) \\ k_3 \sim Gamma(0.88,\ 1) \\ k_4 \sim Gamma(0.78,\ 1) \end{cases} \tag{13}$$

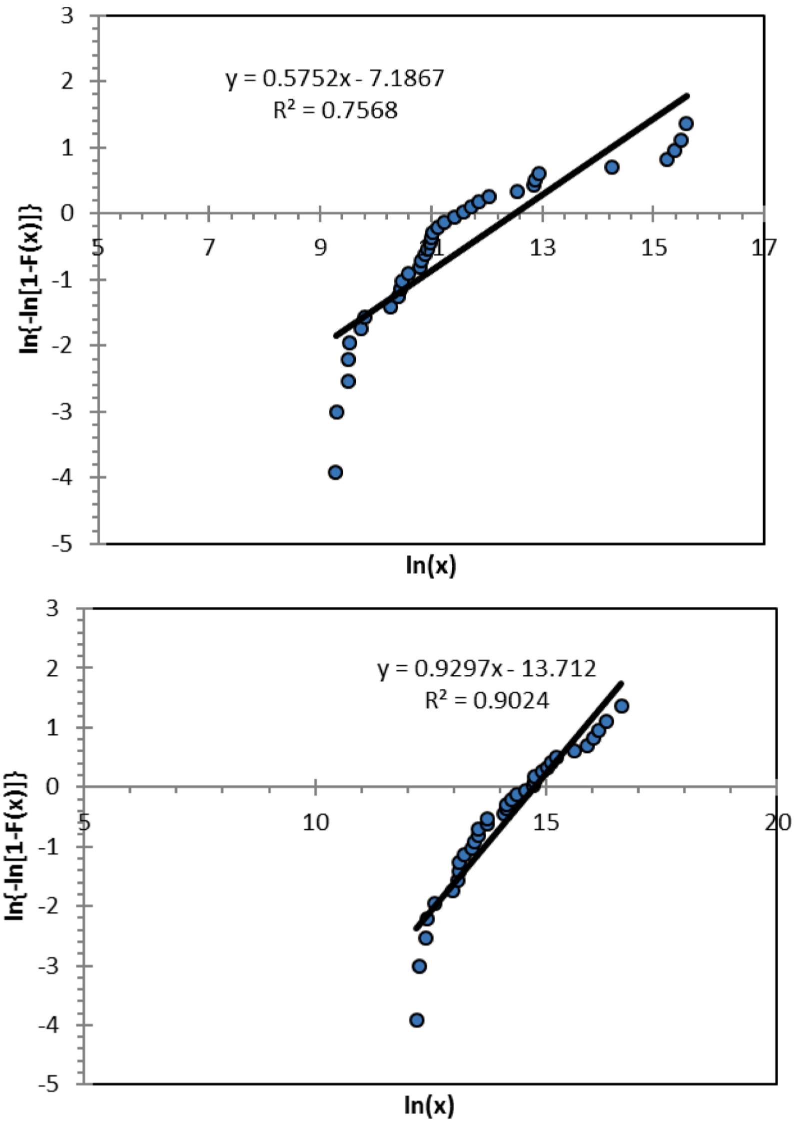

Fig. 5. (a). Weibull plot for incidents associated with ships smaller than 400 GT. (b). Weibull plot for incidents associated with Ships between 400 GT to 6000 GT. (c). Weibull plot for incidents associated with ships LARGER than 6000 GT.

$$\lambda_g = \begin{cases} \lambda_1 \sim Gamma(4,\ 2\ 000\ 000) \\ \lambda_2 \sim Gamma(4,\ 5\ 200) \\ \lambda_3 \sim Gamma(4,\ 1\ 500\ 000) \\ \lambda_4 \sim Gamma(4,\ 13\ 310\ 000) \end{cases} \tag{14}$$

$$g.change = \begin{cases} \text{g.change1} \sim \text{uniform}(50,\ 150) \\ \text{g.change2} \sim \text{uniform}(300,\ 900) \\ \text{g.change3} \sim \text{uniform}(900,\ 14\ 000) \end{cases} \tag{15}$$

The results are reported in Table 6.

Fig. 9 shows that the grouping at boundary about 100 GT produced large parameter uncertainty. The uncertainty included the interval of the mean of Weibull without grouping (or one single group). This further grouping made an indistinguishable estimate thus should be abandoned. The final model of the loss amounts employing Weibull distribution would be Eqs. (6) to (11) and the estimates of the parameters are listed in Table 6.

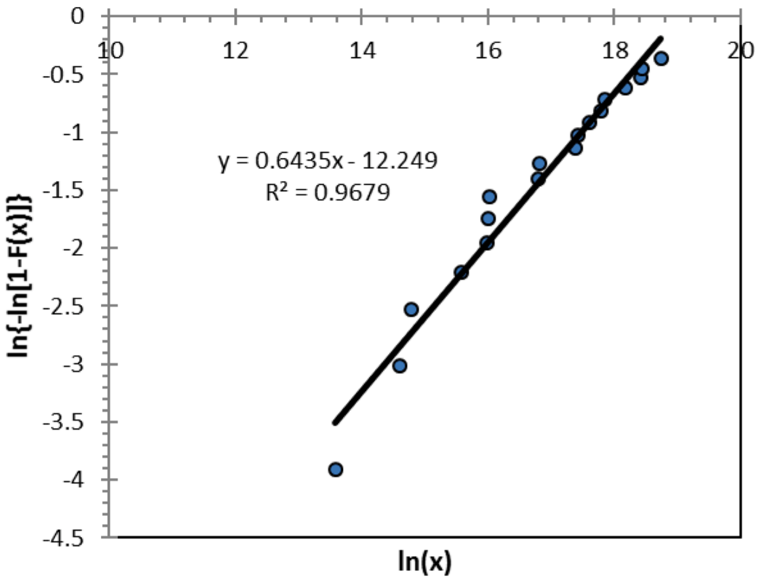

Fig. 5. (*continued*).

Table 5
Weibull distributions on 3 groups of GT.

	mean	SD	val2.5pc	median	val97.5pc	% error
g.change1	392.5	55.97	304.5	391.3	478.6	14.26
g.change2	8518	3008	3586	8457	13,040	35.32
k_1	0.3824	0.01544	0.3517	0.3826	0.4124	4.04
k_2	0.6923	0.01396	0.6644	0.6923	0.7194	2.02
k_3	0.697	0.01513	0.6663	0.6973	0.7257	2.17
$scale_1$	2.02×10^5	8.98×10^4	9.14×10^4	1.83×10^5	4.29×10^5	44.47
$scale_2$	2.59×10^6	6.38×10^5	1.63×10^6	2.50×10^6	4.10×10^6	24.63
$scale_3$	3.69×10^7	1.39×10^7	1.90×10^7	3.41×10^7	7.15×10^7	37.69
MSE	3.99×10^{14}	2.82×10^{14}	2.94×0^{14}	3.42×10^{14}	8.86×10^{14}	70.76

Note: SD = standard derivation; MSE = mean squared error; start = 50,000; sample = 100,000.

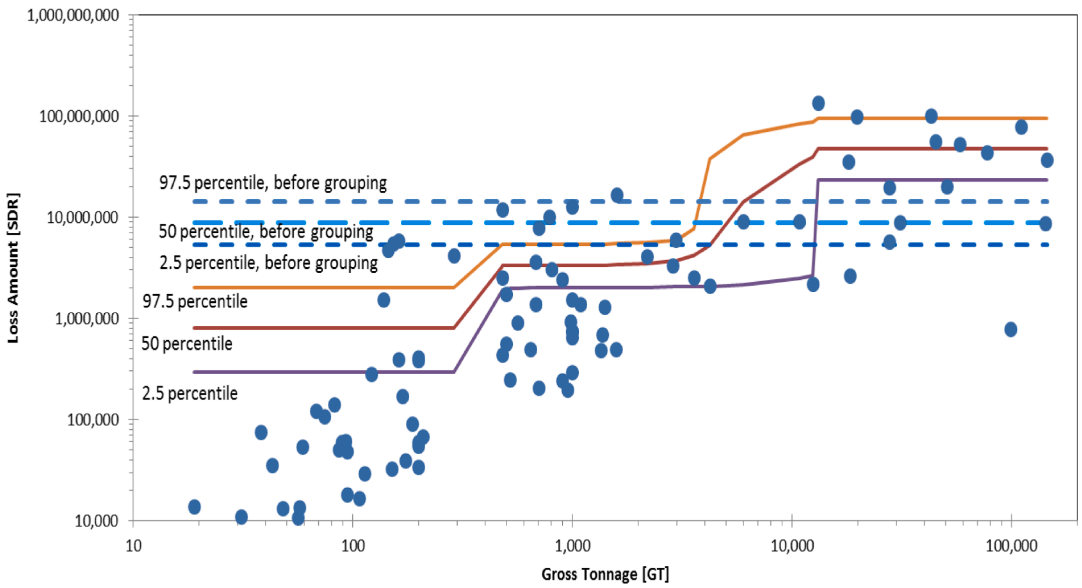

Fig. 6. Estimated mean of Weibull distribution with 3 groupings versus GT.

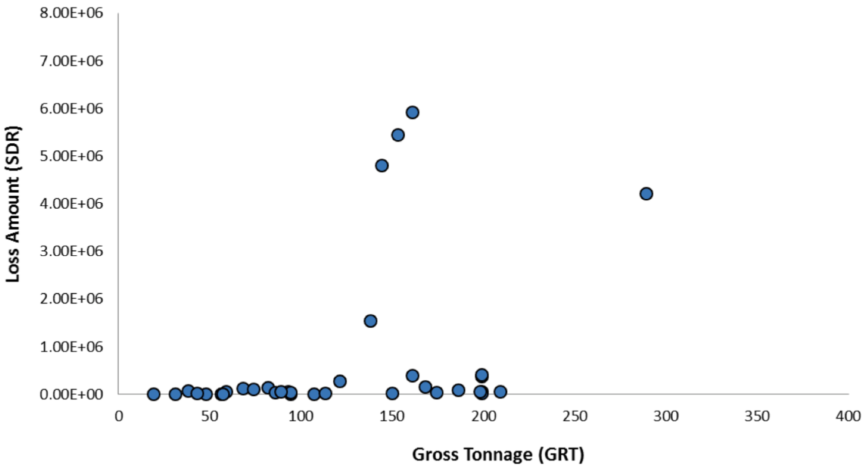

Fig. 7. Loss amounts associated with ships smaller than 400 GT.

3.2. Log-normal model of grouped data

A log-normal distribution is then tested. The log-normal distribution is as follows:

$$f(x) = \sqrt{\frac{\tau}{2\pi}}\frac{1}{x}\exp\left[-\frac{\tau(\ln x - u)^2}{2}\right] \tag{16}$$

It is hypothesised that $y_g = ln(x_g)$ follows the normal distribution if the variable x_g follows log-normal distribution. In Fig. 10, the normal plot of $\ln(x_g)$ are performed on the suitability of log-normal distribution to the loss amounts.

From the high R-squared statistic, a log-normal distribution is suitable for all the data

$$x_g \sim Lognormal(u,\ \tau) \tag{17}$$

$$u \sim Gamma(14,\ 1) \tag{18}$$

$$\tau \sim Uniform(0,\ 1) \tag{19}$$

The estimation result is in Table 7.

In the light of Figs. 3, 4 and 10, and the subsequent analysis normal plots of ln*(x$_i$)* are performed to investigate the suitability of log-normal to the three groups of GT.

Having a large value of R-squared statistic in these three plots of Fig. 11, Bayesian log-normal model is fitted on the three groups of GT:

Fig. 12.

$$x_i \sim Lognormal(u_g, \tau_g) \tag{20}$$

where

$$u_g = \left\{ \begin{matrix} u_1 \sim Gamma(11.5,\ 1) \\ u_2 \sim Gamma(14.2,\ 1) \\ u_3 \sim Gamma(16.8,\ 1) \end{matrix} \right. \tag{21}$$

$$\tau_g = \begin{cases} \tau_1 \sim Uniform(0,\ 1) \\ \tau_2 \sim Uniform(0,\ 1) \\ \tau_3 \sim Uniform(0,\ 1) \end{cases} \tag{22}$$

in which

$$g''.\text{change1}'' \sim \text{uniform}(300,\ 800),\ \text{and g.change2} \ \sim\ \text{uniform}(1\ 000,\ 14\ 000) \tag{23}$$

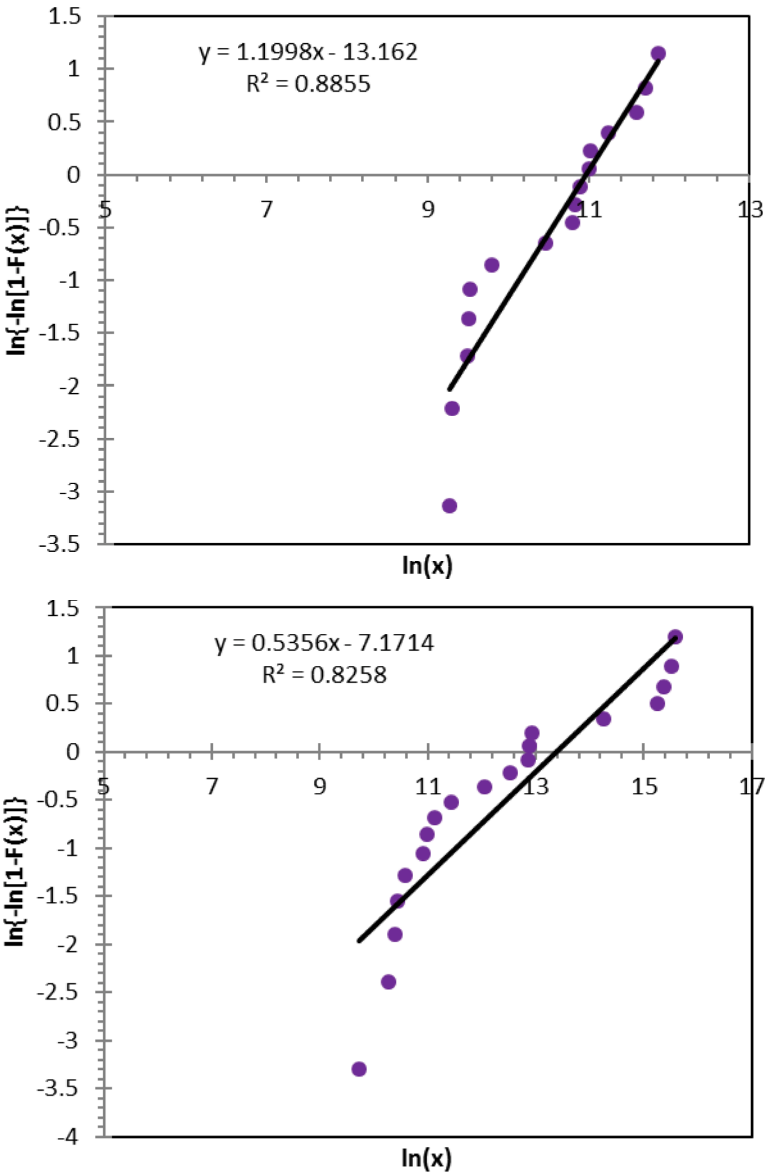

Fig. 8. (a). Weibull plot for Incidents associated with ships smaller than 100 GT. (b). Weibull Plot for Incidents associated with Ships between 100 GT and 400 GT.

Table 6
Weibull distributions on 4 groups of GT.

	mean	SD	val2.5pc	median	val97.5pc	% error
g.change1	118.8	9.368	97.53	118	136.7	7.89
g.change2	397.3	62.51	304.7	394.3	522.6	15.73
g.change3	8516	3006	3611	8452	13,050	35.30
k_1	0.9401	0.02479	0.89	0.9405	0.9872	2.64
k_2	0.3301	0.01836	0.2926	0.3305	0.3647	5.56
k_3	0.692	0.01407	0.664	0.6922	0.7193	2.03
k_4	0.697	0.01521	0.6658	0.6974	0.7254	2.18
$scale_1$	5.41×10^4	1.52×10^4	3.24×10^4	5.15×10^4	9.08×10^4	28.03
$scale_2$	5.67×10^5	6.09×10^5	1.41×10^5	4.16×10^5	1.89×10^6	107.42
$scale_3$	2.59×10^6	6.42×10^5	1.63×10^6	2.50×10^6	4.13×10^6	24.73
$scale_4$	3.70×10^7	1.40×10^7	1.90×10^7	3.40×10^7	7.21×10^7	37.91
MSE	4.19×10^{14}	1.10×10^{15}	2.96×10^{14}	3.46×10^{14}	9.59×10^{14}	262.76

Note: SD = standard derivation; MSE = mean squared error; start = 50,000; sample = 100,000.

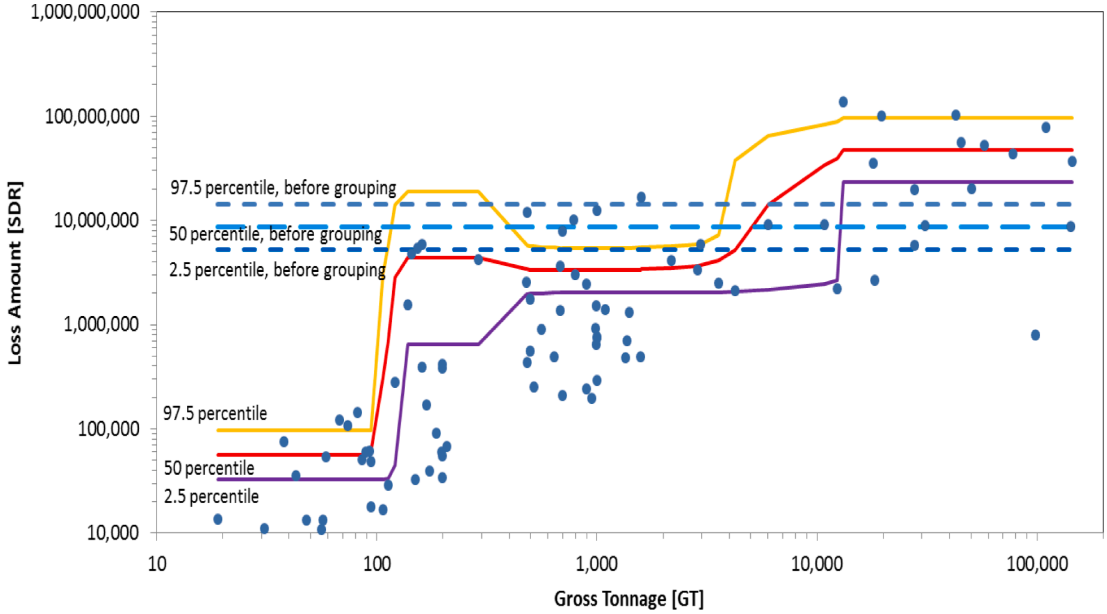

Fig. 9. Estimated mean of Weibull distribution with 4 groups versus GT.

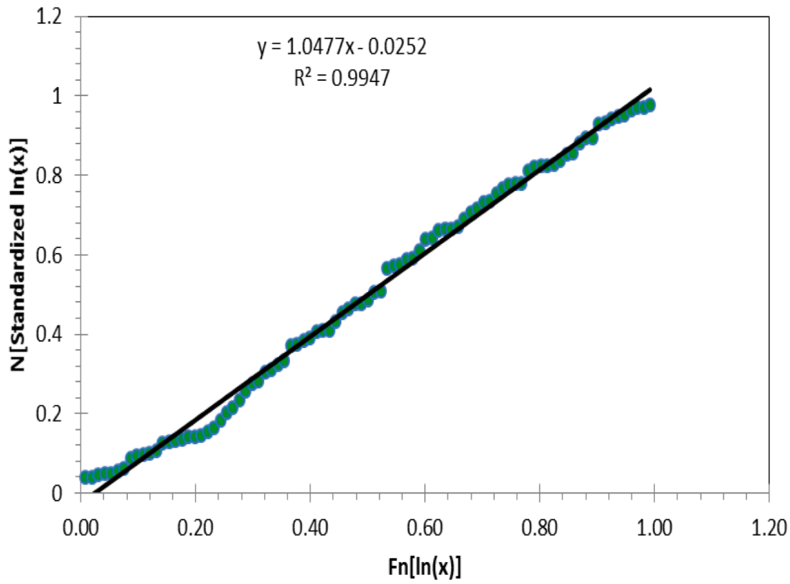

Fig. 10. Normal Plot of $\ln(x_i)$.

Table 7
Log-normal Distribution for one single group of Incidents.

	mean	SD	val2.5pc	median	val97.5pc	% error
u	1.37×10^1	2.65×10^{-1}	1.32×10^1	1.37×10^1	1.42×10^1	1.94
τ	0.1634	0.02434	0.1192	0.1623	0.2147	14.90
mean	2.33×10^7	1.69×10^7	7.70×10^6	1.89×10^7	6.53×10^7	72.40
MSE	1.03×10^{15}	2.29×10^{15}	5.57×10^{14}	6.43×10^{14}	3.66×10^{15}	222.55

Note: SD = standard derivation; MSE = mean squared error; start = 50,000; sample = 100,000.

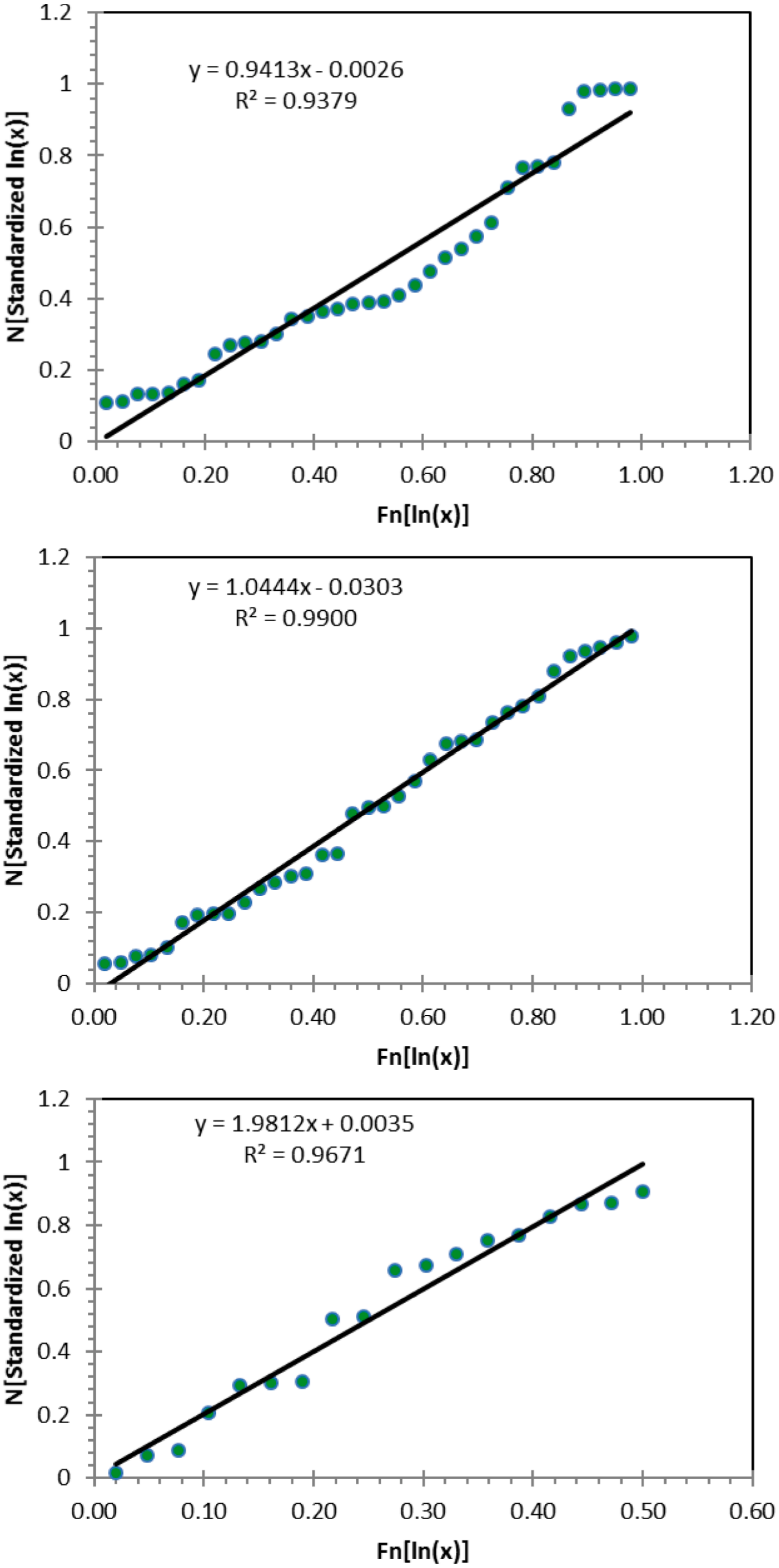

Fig. 11. (a). Normal plot for loss amounts associated with ships smaller than 400 GT. (b). Normal plot for loss amounts associated with Ships between 400 GT and 6000 GT. (c). Normal plot for loss amounts associated with ships larger than 6000 GT.

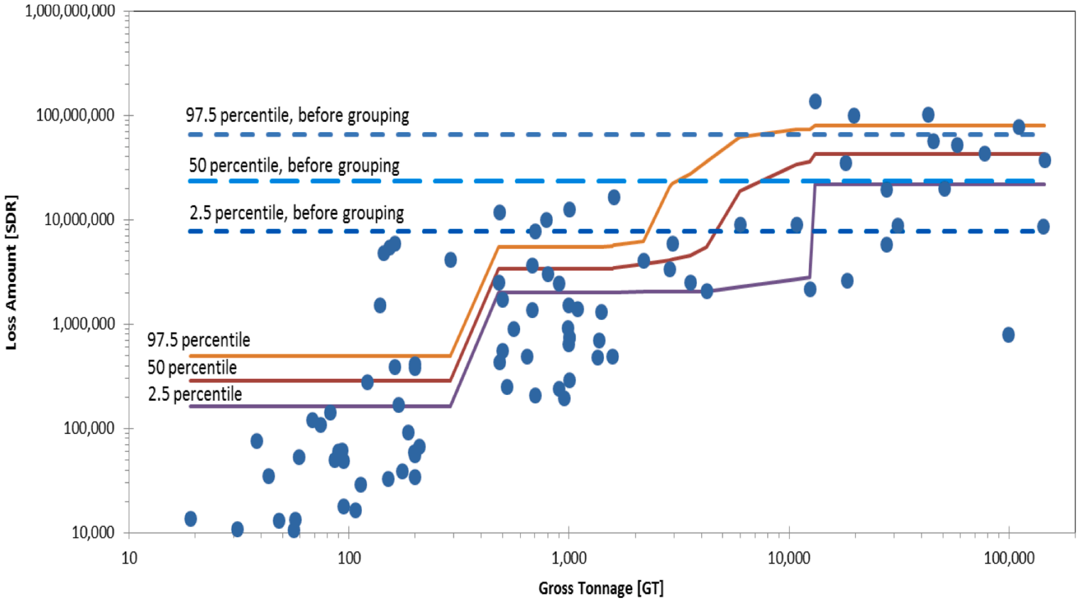

Fig. 12. Estimated mean of log-normal distribution with 3 groups versus GT.

Table 8
Log-normal distribution on 3 groups of GT.

	mean	SD	val2.5pc	median	val97.5pc	% error
g.change1	390.1	52.05	304.5	390.2	475.6	13.34
g.change2	7682	3095	2674	6974	13,040	40.29
u_1	11.52	0.2384	11.05	11.52	11.99	2.07
u_2	14.19	0.2181	13.76	14.2	14.62	1.54
u_3	16.77	0.306	16.17	16.77	17.37	1.82
τ_1	0.5135	0.0791	0.3708	0.5094	0.681	15.40
τ_2	0.6426	0.1037	0.4558	0.6362	0.864	16.14
τ_3	0.6804	0.1081	0.4838	0.6751	0.9085	15.89
MSE	3.68×10^{14}	1.15×10^{14}	2.94×10^{14}	3.41×10^{14}	6.20×10^{14}	31.26

Note: SD = standard derivation; MSE = mean squared error; start = 50,000; sample = 100,000.

The estimation result is reported in Table 8.

Compared with Tables 7 and 8, the parameter uncertainty of the grouped log-normal model is higher, u_1 and π_1 gives% errors larger than u and τ in previous model. The MSE of the grouped log-normal model is smaller than the previous single log-normal model, at the same time the uncertainty of the MSE becomes smaller although the parameter uncertainty is larger. The grouped log-normal model is a better description of the loss amounts than using a single-group log-normal model, and this effect burrs the increase in uncertainty from parameters.

3.3. Comparisons

The MSE of the grouped Weibull model ranges from 2.94×10^{14} to 8.86×10^{14} (see Table 5), and the MSE of the grouped log-normal model ranges from 2.94×10^{14} to 6.20×10^{14} (see Table 8). Both the mean and deviation of MSE of log-normal models are smaller than those of Weibull model, and therefore the grouped log-normal model is a better model with smaller number of groups and associated parameter uncertainty. The grouped loss amounts follow different log-normal distributions in the form of Eqs. (20) to (23). The mean and standard deviation (in parentheses) of the parameters and statistics (SDR) are listed in Tables 9 and 10.

The statistics are evaluated through MCMC simulations using the same practice in this study. Observing that all the means of the statistics are getting larger as moving from a group of smaller ships to a group of larger ships and the 95% intervals of the statistics are non-overlapping, sufficient evidence showing that larger ships have larger probability producing larger loss amounts, the fluctuation of the loss amounts from them are larger, too.

Table 9
The final log-normal model on three groups of GT.

	GT < 390 (52.05)	390.1 < GT < 7682 (52.05) (3095)	GT > 7862 (3095)
u	11.52 (0.2384)	14.19 (0.2181)	16.77 (0.306)
τ	0.5135 (0.0791)	0.6426 (0.1037)	0.6804 (0.1081)
Mean	285,400 (87,150)	3.354×10^6 (900,800)	4.297×10^7 (1.499×10^7)
Variance	6.982×10^{11} (9.71×10^{11})	5.554×10^{13} (5.968×10^{13})	8.376×10^{15} (1.07×10^{16})
Median	103,800 (25,130)	1.494×10^6 (328,600)	2.006×10^7 (6.332×10^6)
Mode	14,780 (5579)	312,200 (102,000)	4.574×10^6 (1.796×10^6)

Remark: () = standard deviation.

Table 10
The 95% interval of the statistics.

	GT < 390 (52.05)	390.1 < GT < 7682 (52.05) (3095)	GT > 7862 (3095)
Mean	[1.620×10^5, 4.970×10^5]	[2.012×10^6, 5.504×10^6]	[2.200×10^7, 7.971×10^7]
Variance	[1.099×10^{11}, 2.805×10^{12}]	[1.099×10^{13}, 1.990×10^{14}]	[1.316×10^{15}, 3.310×10^{16}]
Median	[6.320×10^4, 1.610×10^5]	[9.460×10^5, 2.237×10^6]	[1.050×10^7, 3.513×10^7]
Mode	[5.850×10^3, 2.740×10^4]	[1.410×10^5, 5.370×10^5]	[1.835×10^6, 8.752×10^6]

Remark: () = standard derivation; [] = range.

4. Conclusions

The objective of this paper is to develop a parametric model of claims. A procedure of data grouping has been demonstrated according to R-squared statistic and mean squared errors, and then the model has been specified according to the criterion. The real data set relates to IOPC financial compensation for oil pollution damage due to tanker accidents. The compensation data from IOPC is converted to the actual loss amounts in SDR according to the International Conventions. Bayesian models on the groups according to the gross tonnage and time are developed. Throughout the model building and grouping of the data, the parameter uncertainty is monitored through Bayesian statistics by MCMC algorithm. Unreasonable estimation and the uncertainty on the mean of the estimated distribution are avoided and monitored respectively when the sample size is decreased throughout the grouping according to gross tonnage. From the resulting model, the loss amounts of larger ships are larger in magnitude. The stakeholders are highly sensitive to the upper limits (extremes) of the claim amounts due to tanker oil spills. The limitations of this study include: further studies may consider alternative approaches for modelling the loss amounts. The data grouping can be more complicated, while data grouping of this Bayesian approach retains a natural interpretation of ship size. The predictivity of modelling may change in future, as the oil spills of similar incidence may cause more serious damage over years and on lower risk probabilities for larger oil spills in the long-term.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been partially supported by The Hong Kong Polytechnic University (G-UAFE, P0008760).

References

IMF, 1991. A Study of the Soviet Economy. International Monetary Fund, Washington, D.C.
Kontovas, C.A., Psaraftis, H.N., Ventikos, N.P., 2010. An empirical analysis of IOPCF oil spill cost data. Mar. Pollut. Bull. 60 (9), 1455–1466.
Kreps, R.E., 1997. Parameter uncertainty in (Log)normal distributions. Proc. Casualty Actuar. Soc. 84, 553–580.
Shahriari, M., Frost, A., 2008. Oil spill cleanup cost estimation - developing a mathematical model for marine environment. Process Saf. Environ. Prot. 86 (3), 189–197. https://doi.org/10.1016/j.psep.2007.12.001.
Talley, W.K., Yip, T.L., Jin, D, 2012. Determinants of vessel-accident bunker spills. Transport. Res. Part D 17 (8), 605–609.
van Kampen, C.E., 2003. Estimating the parameter risk of a loss ratio distribution. Casualty Actuarial Society Forum. Spring, pp. 177–213.
White, I.C., 2002. *Factors affecting the cost of oil spills.* International Tanker Owners Pollution federation Limited (ITOPF). International Tanker Owners Pollution Federation Limited (ITOPL, London, U.K.
Yamada, Y., 2009. The cost of oil spills from tankers in relation to weight of spilled oil. Mar. Technol. 46 (4), 219–228.
Yip, T.L., Talley, W.K., Jin, D, 2011. The effectiveness of double hulls in reducing vessel-accident oil spillage. Mar. Pollut. Bull. 62 (11), 2424–2432.
Zhong, H., Lin, Y., Yip, T.L., 2021. A quantitative liability risk assessment of oil spills in oil ports: the case of the Guangdong-Hong Kong-Macao GREATER bay area. Marit. Policy Manag. (online 8 June 2021).