



This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Li, Fang; Lu, Liangliang; Suominen, Mikko; Kujala, Pentti Short-term statistics of ice loads on ship bow frames in floe ice fields

Published in: Marine Structures

DOI: 10.1016/j.marstruc.2021.103049

Published: 01/11/2021

Document Version Publisher's PDF, also known as Version of record

Published under the following license: CC BY

Please cite the original version:

Li, F., Lu, L., Suominen, M., & Kujala, P. (2021). Short-term statistics of ice loads on ship bow frames in floe ice fields: Full-scale measurements in the Antarctic ocean. *Marine Structures*, *80*, Article 103049. https://doi.org/10.1016/j.marstruc.2021.103049

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Contents lists available at ScienceDirect

Marine Structures



journal homepage: www.elsevier.com/locate/marstruc

Short-term statistics of ice loads on ship bow frames in floe ice fields: Full-scale measurements in the Antarctic ocean

Fang Li^{*}, Liangliang Lu, Mikko Suominen, Pentti Kujala

Aalto University, School of Engineering, Finland

ARTICLE INFO

Keywords: Ice load statistics Full-scale measurement Floe ice field Truncated distribution Mixture distribution Statistical inference

ABSTRACT

Ship operation and ice loading in floe ice fields have received considerable interest during recent years. There have been several numerical simulators developed by different institutes which can simulate ship navigation through floe ice fields and estimate ship performance and local ice loads. However, public data obtained from full-scale measurement covering comprehensively ship performance and ice loads under various ice thicknesses, concentrations and floe sizes are rare. The 2018/19 Antarctic voyage of the Polar Supply and Research Vessel (PSRV) S.A. Agulhas II gathered considerable data of the ship in floe ice fields under various thicknesses, concentrations, and floe sizes. The aim of this paper is to carry out statistical analysis to seek suitable probability distributions which adequately fit the measured ice load and therefore suitable to be used as parent distributions for long-term estimation. For this aim, three categories of probability distributions, namely standard distributions, truncated distributions and mixture distributions are tested. It is found that truncated distributions can fit the load data better than standard distributions bounded at the threshold. In addition, mixture distributions are shown to have promising features, which fit the data well and are able to separate distribution components. Subsequentially, the well-performed distributions are used as parent distributions to make long-term load estimations. The estimation results demonstrate that long-term estimations are sensitive to the selection of parent distribution, which addresses the importance of finding correct distribution to model short-term ice loads. The data of ten selected cases will be published for the use of other researchers.

1. Introduction

Ships sailing through Arctic and Antarctic oceans frequently encounter floe ice fields. A floe ice field is comprised of large amount of ice floe with various shape and size. The thickness and concentration may also vary locally and between different ice fields. There are increasing number of numerical simulation programs developed by various institutes to simulate ships going through floe ice fields, e. g. Refs. [1–5]. However, public full-scale measurement data regarding ships in floe ice fields are rare. The Polar Supply and Research Vessel (PSRV) S.A. Agulhas II encountered extensive floe ice during its 2018/19 Antarctic voyage, and gathered large amount of data covering wide ranges of thicknesses, concentrations, and floe sizes. This provides valuable resource to investigate ship performance and ice loads related issues in floe ice fields.

Due to the extensive randomness in ice conditions, ice behavior and contact scenarios, ice loads measured on the hull frames always

* Corresponding author. *E-mail address:* fang.li@aalto.fi (F. Li).

https://doi.org/10.1016/j.marstruc.2021.103049

Received 4 March 2021; Received in revised form 4 May 2021; Accepted 13 May 2021

Available online 12 August 2021

^{0951-8339/© 2021} The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

show high randomness. Fig. 1 gives an example of the time history of ice loads measured on a bow frame of S.A. Agulhas II. Typically, the load peaks are of concern. Probabilistic approaches are usually adopted to model the random load peaks. There are two different methodologies in the literature to model ice load peaks. The first works on short-term load peaks, i.e. all individual load peaks extracted from the time history. Standard distributions such as Rayleigh distribution [6], lognormal distribution [6–9], Weibull distribution [6–9] and exponential distribution [7–9] have been deployed to model short-term ice load peaks. Several papers have identified Weibull distribution as the best distribution among the tested standard distributions for short-term load peaks, e.g. Refs. [7–9]. However, most of these papers are not dedicated for ice loads in floe ice fields; instead, considerable level ice navigations are involved in these analyses. Suyuthi et al. [10] pointed out that Weibull distributions which gives better fit to their data for the upper tail. Adequately good fitting to the upper tail is necessary if the short-term distribution is used as the parent distribution for the establishment of long-term distributions. Another method to fit the upper tail well is the Event Maximum Method proposed by Jordaan et al. [11] for ramming analysis and later used by Taylor et al. [12] for continuous measurement, which uses exponential distribution to fit the upper part of the data without attempting to fit the whole dataset. The drawback of such method is that it does not offer an overall description of the whole ice load dataset, which is not suitable if the purpose is fatigue analysis. Also, one needs to define a portion of the data as 'tail part', which is a somewhat subjective judgement.

The second methodology for ice load analysis is to work directly on the maxima of fixed time intervals, e.g. 10-min maxima. Such methodology has been attempted with Gumbel distribution [13,14], average conditional exceedance rate method [15] and Event Maximum Method [16]. One merit of such methodology is that it does not require separate modelling of exposure because it can be easily defined in terms of operational time, while with short-term load peak distributions one needs to estimate the exposure (i.e. number of contacts) separately. Nonetheless, working with fixed time maxima leads to inefficient use of data since all the load peaks except for the maxima are abandoned, which then demands measurement covering rather long duration (e.g. a whole winter). Analysis of short-term load peaks offers more insights into events within small time scale and maximizes the usage of measurement data.

The focus of this paper is on short-term load peaks. Within ship design context, the goal is to find appropriate probability distributions which are able to describe the random nature of ice loading process, and then to use the distributions as parent distributions to calculate the distribution of long-term extremes (e.g. life-time maxima). This paper presents ten selected cases of the ship S.A. Agulhas II in floe ice fields during its 2018/19 Antarctic voyage. The datasets cover extensive information including ship navigation data, machinery data, ice condition data and local ice load measurements. The ice condition covers various ice thicknesses, concentrations, and floe sizes, which gives an overall view of the ship's performance and ice loading in floe ice fields. A preliminary analysis of the whole dataset will be presented in Li et al. [17], which identifies the dependences between ship resistance, ice loads ice conditions, and ship maneuvers. The aim of this paper is to investigate suitable distributions for the ice load data to add further insights into the statistical analysis of short-term ice loads. In addition to standard exponential, Weibull and lognormal distributions, two more distribution categories, namely truncated distributions and mixture distributions are tested. These open new possibilities for the fitting of ice load data and the results are shown to be promising. The paper also investigates how different the estimated extreme load value can be if using different distributions as parent distributions for long-term estimation. The dataset will be published along with his paper, which complements the dataset published along with Suominen et al. [18] where the same ship in level ice condition on the Baltic Sea is in focus.

2. Description of the dataset

The data are collected through the 2018/19 Antarctic voyage of the Polar Supply and Research Vessel (PSRV) S.A. Agulhas II. Fig. 2 illustrates the ship route during the 2018/19 voyage. The ship started from Cape Town, South Africa, on 6 December, heading south to Antarctic along zero meridian; then travelled through Weddell Sea with tasks and finally came back to South Africa on 15 March. The readers are referred to Refs. [19,20] for more information about the 2018/19 voyage.

The ship is instrumented with shear strain gauges at the starboard side on a total of nine frames, including two at the bow, three at the bow shoulder and four at the stern shoulder (see Fig. 3). The forces are converted from the shear strain on the frames according to Suominen et al. [21]. Here the load measurement of the bow region will be analyzed because the number of loads at other locations is too small for statistical analysis. The measured values represent the total ice force on a frame with frame space of 0.4 m. The time



Fig. 1. Time history of measure ice loads in Case 10 on frame #134.5 and Rayleigh separated load peaks.



Fig. 2. Route of the 2018/19 Antarctic voyage.



Fig. 3. Instrumentation of S.A. Agulhas II.

history of measured ice loads is fed into a Rayleigh separator (see Ref. [8]) so that the load peaks can be extracted. An example of the time history of ice loads and the extraction method has been given in Fig. 1. A threshold of 10 kN is adopted since smaller load peaks contain considerable hydrodynamic load. Therefore, it should be made clear that the load data to be analyzed represent *load peaks over* 10 kN.

The ice conditions during the voyage are monitored via two sources. The first is visual observation, which are conducted by dedicated ice observers on the bridge, estimating ice concentration, floe size and thickness approximately every minute and summarizing the results in 10-min interval (see Refs. [22,23] for more information). In addition to that, an ice condition camera (Fig. 3) is installed on the ship to take photos of the ice condition constantly during the voyage. The information published through this paper includes the visually observed ice condition together with the raw photos taken during the voyage.

To investigate ice loads in floe ice fields, ten cases are selected from the whole voyage for further investigation. The cases are selected based on several requirements: i) the cases are expected to cover a wide range of ice condition parameters including thickness, concentration and floe size; ii) the duration of a case should be between 15 and 30 min so that enough loads are gathered for statistical analysis; iii) ice concentration does not vary much during each case; iv) ship speed and heading should not change too much. Table 1 summarize the main ice condition parameters of the selected cases. Note that there is considerable variation in floe diameter within each case. Therefore, the values in Table 1 are the rough estimations of the categories where the majority of ice floe fall in. Fig. 4 gives examples of orthorectified images [24] of representative photos from the ten cases. The scope of each image is about 350 m by 350 m.

3. Methods

This paper investigates suitable probability distributions for the description of measured local ice loads at the bow region. The standard distributions commonly adopted in the literature, including exponential distribution, Weibull distribution and lognormal distribution, are first investigated. In addition, this paper proposes the application of truncated distributions, which offers another possibility to fit ice loads data. Moreover, mixture of two exponential distributions, which has been investigated by Suyuthi et al. [10], is tested. The methodology with mixture distributions is then generalized in this paper to test mixture of multiple exponential distributions.

It is important to note that the loads data employed here (and in many other publications, e.g. Refs. [9,10]) are the load peaks above a certain threshold to exclude loads from waves. An example has been given in Fig. 1. We can denote the obtained load peaks with *X* via the following definition:

X: Ice load peaks over a certain threshold T (e.g. here 10 kN).

It is convenient to denote *Z* as the difference between *X* and the threshold:

Z: Ice load peaks subtracting the threshold, i.e. z = x-T.

We can then define another quantity, Y, as:

Y: All ice load peaks without setting a threshold.

It is difficult, if possible, to extract *Y* from the measurement since the small values are mixed up with loads from waves. Moreover, the measurement accuracy for small loads can be low since the system is set capable of measuring loads with magnitude of several MN, thus the noise interferes. Therefore, *X* is often investigated due to practicality. The distinction is very important firstly because the distribution fitted to one does not necessarily describe the other, and also because it relates to the definition of exposure. For example, to use the distribution fitted with *X*, one should define the exposure as *number of encountered loads over* 10 kN, while with *Y* the definition of exposure is free of threshold.

3.1. Standard exponential, weibull and lognormal distributions bounded at a threshold

The formulations of the probability density function (PDF) and cumulative distribution function (CDF) of exponential, Weibull and lognormal distributions as well as their maximum likelihood estimators (MLE) are summarized in Table 2. Here erf denotes the complementary error function; *T* is the threshold. The quantity *z* is used instead of *x* for CDF and MLE for simplicity. $f_X(x)$ equals $f_Z(z)$ because of the linear transformation between *x* and *z*. The MLEs of exponential and lognormal distribution can be solved directly while those of Weibull distribution need to be approximated numerically by solving the system of nonlinear equations. Usually, exponential distribution bounded at a threshold other than zero is referred to as two-parameter exponential distribution while Weibull distribution as three-parameter Weibull distribution. However, with z = x-T, the PDFs and CDFs can be converted to standard distributions, for which the maximum likelihood estimations are straightforward. Therefore, in this paper we still use the term 'standard distributions' to refer to this distribution category.

| Table 1 | | | |
|------------------------------|-------------------|----------|--------|
| Summary of the ice condition | parameters of the | selected | cases. |

| Case | Date | Start time | Duration (min) | h _{mean} (cm) | h _{max} (cm) | Concentration | Floe diameter (m) |
|------|------------|------------|----------------|------------------------|-----------------------|---------------|-------------------|
| 1 | 7-Jan-2019 | 18:10 | 28 | 73 | 170 | 30% | 0–20 m |
| 2 | 7-Jan-2019 | 16:40 | 20 | 87 | 110 | 30% | 10–30 m |
| 3 | 8-Jan-2019 | 09:48 | 20 | 72 | 90 | 50% | 20–50 m |
| 4 | 8-Jan-2019 | 20:00 | 30 | 130 | 170 | 50% | 0–20 m |
| 5 | 8-Jan-2019 | 21:08 | 16 | 157 | 225 | 50% | 10–30 m |
| 6 | 6-Jan-2019 | 09:11 | 17 | 56 | 90 | 70% | 50–100 m |
| 7 | 8-Jan-2019 | 09:00 | 20 | 106 | 150 | 70% | 20–40 m |
| 8 | 3-Jan-2019 | 22:01 | 24 | 53 | 110 | 90% | 0–30 m |
| 9 | 3-Jan-2019 | 20:01 | 29 | 72 | 150 | 90% | 20–50 m |
| 10 | 3-Jan-2019 | 19:30 | 30 | 80 | 110 | 90% | 20–40 m |



Fig. 4. Orthorectified images of representative photos of the selected cases, from (a) to (j) showing case 1 to 10. The scale in (a) applies to all.

| able 2 |
|--|
| DF, CDF and MLE of exponential, Weibull and lognormal distributions. |

| Distribution | PDF | CDF | MLE |
|--------------|---|--|---|
| Exponential | $ \begin{aligned} f_X(x) &= f_Z(z) = \lambda \exp(-\lambda z) \\ &= \lambda \exp(-\lambda (x-T)) \end{aligned} $ | $F_X(\mathbf{x}) = F_Z(\mathbf{z}) = 1 - \exp(-\lambda(\mathbf{z}))$ | $\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} z_i}$ |
| Weibull | $f_X(x) = f_Z(z) = \frac{\beta}{\alpha} \left(\frac{z}{\alpha} \right)^{\beta-1} \exp\left(- \left(\frac{z}{\alpha} \right)^{\beta} \right) = \frac{\beta}{\alpha} \left(\frac{(x-T)}{\alpha} \right)^{\beta-1} \exp\left(- \frac{z}{\alpha} \right)^{\beta-1} \exp\left(-$ | $F_X(\mathbf{x}) = F_Z(\mathbf{z}) = 1 - \exp\bigg(-$ | $\widehat{\alpha} = \left(\frac{1}{2}\sum_{i=1}^{n} \alpha_{i}\widehat{\beta}\right)^{\frac{1}{\widehat{\beta}}}$ |
| | $\left(\frac{(x-T)}{\alpha}\right)^{\beta}$ | $\left(\frac{z}{a}\right)^{\beta}$ | $\widehat{\beta} = \left(n \sum_{i=1}^{n-1} \widehat{\beta}_i \right)$ |
| Lognormal | $egin{aligned} f_X(x) &= f_Z(z) &= rac{1}{z\sqrt{2\pi\sigma^2}} 	ext{exp}igg(-rac{\left(\ln(z)-\mu ight)^2}{2\sigma^2}igg) &= rac{1}{(x-T)\sqrt{2\pi\sigma^2}} 	ext{exp}igg(-rac{\left(\ln(x-T)-\mu ight)^2}{2\sigma^2}igg) \end{aligned}$ | $F_X(x) = F_Z(z) = rac{1}{2} + rac{1}{2} 	ext{erf}igg(rac{	ext{ln}(z) - \mu}{\sqrt{2\sigma^2}}igg)$ | $\frac{n}{\frac{1}{\widehat{\alpha}}\sum_{i=1}^{n}z_{i}^{\widehat{\beta}}\log z_{i}-\sum_{i=1}^{n}\log z_{i}}$ $\widehat{\mu}=\frac{1}{n}\sum_{i=1}^{n}\ln z_{i}$ $\widehat{\sigma}^{2}=\frac{1}{n}\sum_{i=1}^{n}(\ln z_{i}-\widehat{\mu})^{2}$ |

3.2. Truncated distributions

The idea of using truncated distributions to fit the load peaks above a threshold is natural. If we assume the load peaks arising from ship-ice interaction, i.e. the *Y* defined at the beginning of this section, can be modelled by a distribution bounded by zero at the lower end, e.g. exponential, Weibull or lognormal distributions, the load peaks above a threshold, i.e. the *X* defined at the beginning of this section, will be truncated-distributed. Denoting the PDF and CDF of a standard probability distribution bounded at zero as $f_X(x)$ and $F_X(x)$, the corresponding left-truncated distribution will be:

$$f_X^{trun}(x) = \frac{f_X(x)}{1 - F_X(T)}, \ x > T$$
(1)

F. Li et al.

Following this, the truncated exponential distribution is

$$f_{\chi}^{exprime}(x) = \lambda \exp(-\lambda(x-T)) = \lambda \exp(-\lambda z)$$
⁽²⁾

which is identical to the PDF of exponential distribution bounded at *T* listed in Table 2. Therefore, the truncated exponential distribution is equivalent to the standard exponential distribution bounded at the threshold.

Again, following Eq. (1), the PDF of the left-truncated Weibull distribution can be expressed as:

$$f_X^{wblrun}(x) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(\left(\frac{T}{\alpha}\right)^{\beta} - \left(\frac{x}{\alpha}\right)^{\beta}\right)$$
(3)

where *T* is the threshold which equals 10 kN here. The CDF is then:

$$F_X^{wbltrun}(x) = 1 - \exp\left(\left(\frac{T}{\alpha}\right)^{\beta} - \left(\frac{x}{\alpha}\right)^{\beta}\right)$$
(4)

One can compare the expression with that in Table 2 to notice the difference of the exponents. Truncated Weibull distribution has thicker tails comparing to standard Weibull distribution, which indicates larger extreme loads for the same return period.

The MLE of left-truncated Weibull distribution [25] is

$$\widehat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} (x_i^{\widehat{\beta}} - T^{\widehat{\beta}})\right)^{\frac{1}{\widehat{\beta}}}$$
(5)

$$\frac{1}{\widehat{\beta}} = \frac{\sum_{i=1}^{n} \left(\frac{x_i}{T}\right)^{\widehat{\beta}} \log \frac{x_i}{T}}{\sum_{i=1}^{n} \left(\left(\frac{x_i}{T}\right)^{\widehat{\beta}} - 1\right)} - \frac{1}{n} \sum_{i=1}^{n} \log \frac{x_i}{T}$$
(6)

Similarly to standard Weibull distributions, $\hat{\alpha}$ and $\hat{\beta}$ need to be approximated numerically by solving the system of nonlinear equations. Truncated Weibull distribution does not always have MLE solutions. The following needs to be satisfied by the data for the existence of non-zero solutions of $\hat{\alpha}$ and $\hat{\beta}$:

$$2\left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{x_{i}}{T}\right)^{2} - \frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{x_{i}}{T}\right)^{2} > 0$$
(7)

If Eq. (7) is violated, the only solution to $\hat{\alpha}$ and $\hat{\beta}$ is that they both equals 0. On the contrary, standard Weibull distribution always has MLE solutions regardless of the quality of fitting.

Again, following Eq. (1), the truncated lognormal distribution can be expressed as:

$$f_{X}^{logntrun}(x) = \frac{\frac{1}{x\sqrt{2\pi\sigma^{2}}} \exp\left(-\frac{(\ln(x)-\mu)^{2}}{2\sigma^{2}}\right)}{\frac{1}{2} + \frac{1}{2} erf\left(\frac{\ln(7)-\mu}{\sqrt{2\sigma^{2}}}\right)}$$
(8)



Fig. 5. Numerical experiment of fitting datasets with different thresholds via standard Weibull distribution bounded at the threshold and truncated Weibull distribution, left: histogram of the samples, inserted figure showing values larger than 1000 and vertical axis re-scaled; right: results of fitting as functions of the threshold.

An obvious advantage of using truncated lognormal distribution instead of standard lognormal distribution is that the PDF does not necessarily have a rising part, which potentially leads to improved fitting performance for datasets which have constantly decreasing probability density. The MLE of truncated lognormal distribution cannot be written in closed form. However, the solution to the maximum likelihood problem can be solved numerically, easily with e.g. Matlab *mle* function. Similarly to truncated Weibull distribution, the solution of truncated lognormal distribution with MLE does not always exist.

The major advantage of adopting truncated distributions is that the obtained distribution parameters are independent of the threshold, while it is on the contrary with standard distributions. To demonstrate this, we implement a numerical experiment as an example with 10^4 samples drawn from a standard Weibull distribution bounded at zero:

$$Y \sim Weibull(\alpha = 100, \beta = 0.5) \tag{9}$$

The histogram of the samples is shown in Fig. 5. Next, a threshold *T* is implemented to the dataset to get *X* by deleting the *Y* values lower than *T*. Standard Weibull distribution left-bounded at *T* and truncated Weibull distribution truncated at *T* are then fitted to the remaining data via maximum likelihood estimation. Fig. 5 summarizes the obtained α and β values as functions of the threshold. As shown here, the truncated Weibull distribution maintains the α and β values from the original distribution regardless of the threshold while the fitting with standard Weibull distribution is dependent on the threshold.

3.3. Mixture of distributions

Mixture of exponential distributions, as suggested by Suyuthi et al. [10], may offer good fitting to ice load data, especially to the tail. Suyuthi et al. [10] use the term 'generalized exponential distribution', but this term may as well refer to a different distribution [26]. In this paper it will be referred to as the 'mixture of exponential distributions' [27].

More generally, *mixture of distributions* can be mixture of any distributions, even if the individual distributions are different, e.g. exponential, Weibull and lognormal distributions. An example is the mixture of Gaussian distributions, which is widely used in machine-learning related topics [28]. Mixture of distributions can be universally expressed as

$$f_X^{mix}(x) = \sum_{j=1}^m w_j f_{X_i}(x)$$
(10)

with the constraint $\sum_{j=1}^{m} w_j = 1$, where w_i are the weights; m is the total number of distribution components; and $f_{X_i}(x)$ the individual distribution components.

distribution components.

The rationale of using mixture of distributions is also natural. Ship-ice interaction may come from different processes when a ship goes through a floe ice field. For instance, a load may arise from the force to push away an ice floe, or from the force to break the floe by bending, sometimes with splitting prior to bending. It is natural to assume ice loads arising from different processes are differently distributed. In addition, ice condition parameters in a floe ice field are never as constant as one can set in a simulator. Ship speed may also vary even if the crew try to maintain a constant speed. It is then natural to regard the measured ice load as the result of combining various distributions with different weight.

Using mixture of distributions opens wide possibilities for the modelling of ice loads, since the distribution type as well as number of distributions can be varied flexibly. In this paper, however, we only investigate the mixture of exponential distributions and leave the rest for future research. The mixture of exponential distributions can be expressed as

$$f_X^{mixexp}(x) = \sum_{j=1}^m w_j \lambda_j e^{-\lambda_j (x-T)}$$
(11)

with the constraint $\sum_{j=1}^{m} w_j = 1$. The total number of distribution parameters to be estimated is then 2m-1 due to the constraint of *w*. The

mixture distribution is then the weighted sum of *m* exponential distributions. Suyuthi et al. [10] tested a special case where *m* equals 2, while Eq. (11) gives the generalized version.

Statistical inference with mixture distributions is not as straightforward as those for standard single distributions, due to the fact of being a sum of several components, which leads to no maximum likelihood estimator in closed form. Nonetheless, there is a standard approach to make maximum likelihood estimation for mixture distributions, which is the Expectation-Maximization (EM) algorithm [28]. A derivation of the implementation of EM algorithm for mixture of exponential distributions is given in Appendix I. Here we simply list the procedure to obtain the w_i and λ_i parameters through iteration:

1) Initialize w_i and λ_i , where j = 1, 2, ...m.

2) Evaluate γ_{ij} , where i = 1, 2, ..., n and j = 1, 2, ...m; *n* denotes the number of ice load data in a dataset, according to the following equation

$$\gamma_{ij} = \frac{w_j \lambda_j \exp\left(-\lambda_j (x_i - T)\right)}{\sum_{j=1}^m w_j \lambda_j \exp\left(-\lambda_j (x_i - T)\right)}$$
(12)

3) Update w_i and λ_i by

$$w_{j} = \frac{\sum_{i=1}^{n} \gamma_{ij}}{n}$$
(13)
$$\lambda_{j} = \frac{\sum_{i=1}^{n} \gamma_{ij}}{\sum_{i=1}^{n} (x_{i} - T) \gamma_{ij}}$$
(14)

4) Iterate to Step 2, until w_i and λ_i converge to the desired level.

The above step gives EM algorithm solution for a mixture of any number of exponential distributions. It is natural to ask whether a higher number of mixed exponential distributions leads to better fitting. In this paper, we will investigate the mixture of two, three, five and ten exponential distributions for the fitting of measured ice loads in order to provide an answer to this question.

3.4. Goodness of fit

Although there are various statistical hypothesis testing methods which can test whether to accept a distribution for a dataset, e.g. Kolmogorov-Smirnov test, in this paper these are not adopted due to several reasons. First, the statistical modelling of ice loads in this paper is in the context of long-term extreme value prediction, for which the parent distribution for the measured short-term dataset is sought. Considerable focuses are given to the upper tail region of the probability distributions, where the measurement is scarce. Therefore, results from a statistical testing method may not adequately reveal the goodness of fit to the tail. Second, the aim of the investigation is to identify the most suitable distribution from the ones in focus here. What is needed is an index to compare the fitting performance between distributions. It is then not necessary to know whether a distribution is to be accepted or not according to certain statistical test.

In this paper, to compare the fitting performance between distributions, the resulting distribution and the datasets are plotted in quantile-quantile plot (Q-Q plot) and the R-square values are evaluated and compared. Results based on Q-Q plot gives more focus on the tail region compared to probability-probability plot (P–P plot), therefore suitable for our purpose of comparison. The R-square values are calculated for the whole dataset and for the upper 20% of data (which roughly represents the tail) respectively, in order to offer insights to the fitting performance to the whole dataset as well as to the upper tail.

3.5. Long-term estimation with extreme distributions

For ship design purpose, one needs to estimate the life-time maximum load on the structure. The distribution used to fit measured short-term load peaks is then applied as parent distribution to derive the distribution of life-time maxima. Denote the maximum load among n loads as M, i.e.

$$M = \max(X_1, X_2, \dots, X_n) \tag{15}$$

Here *n* can be the estimated number of ship-ice contact events during a certain period, e.g. 25 years. The CDF and PDF of M can be expressed as

$$F_M(m) = [F_X(m)]^n \tag{16}$$

$$f_M(m) = n[F_X(m)]^{n-1} f_X(m)$$
(17)

where F_X is the CDF of the individual loads (short-term load peaks). F_X is then the parent distribution while F_M is the extreme value distribution.

For distributions within exponential family, e.g. exponential, Weibull and lognormal distributions, when n is large, F_M converges to Type I extreme distribution, i.e. Gumbel distribution [29], which is characterized as

$$F_M(m) = \exp\left(-\exp\left(-\frac{m-\alpha}{\beta}\right)\right)$$
(18)

where α is the characteristic extremes and β the dispersion factor. With standard exponential, Weibull and lognormal distributions, α and β can be expressed in terms of the parent distribution parameters in closed-form. With truncated distributions and mixture of exponential distributions, closed-form solutions may not exist. However, it is convenient to carry out Monte Carlo simulation to sample the extremes and get the Gumbel distribution parameters. Here the following procedure is adopted:

1. Randomly generate n samples from the parent distribution F_X

2. Find the maximum of the *n* samples

3. Repeat above steps and get 10,000 maxima

4. Fit Gumbel distribution to the 10,000 maxima to get α and β

Once α and β are obtained, the PDF and CDF of the extreme distribution are known. One can calculate the characteristic extreme load (which is α) or the load corresponding to a certain probability of exceedance depending on the actual need.

Although all the tested distributions converge to Gumbel distribution, the convergence behaves differently. Referring to Ref. [29], lognormal and truncated lognormal distributions belong to Class E3 [29], which results in flatter Gumbel distribution (larger β) when the number of loads *n* increases. Exponential and mixture of exponential distributions belong to class E2, which results in stable behavior with the shape unchanged (constant β) regardless of *n*. Weibull distribution may belong to class E1 if the shape parameter is less than 1, which results in more peaked Gumbel distribution with larger *n*, or belong to class E2 if the shape parameter equals 1, or E3 if the shape parameter is larger than 1.

4. Results

Since there are in total 20 datasets from the two bow frames of the ten cases, the majority of the results are attached in Appendix II. Only the results with the data of bow frame #134.5 of case 10 are plotted in this section as an example dataset. To enable a direct visual comparison between the fitting of different distributions to the data, the example dataset as well as fitted distributions are plotted on exponential probability paper in Fig. 6, where the abscissa gives the force magnitude while the ordinate gives the CDF in the scale of $-\log(1 - CDF)$. The Q-Q plot of each fitted distribution for the example dataset is then plotted individually in Fig. 7. In addition, the R-square values of the fittings to the entire 20 datasets are summarized in the histograms shown in Fig. 8 distribution-wise.

4.1. Results of distribution fittings

4.1.1. Standard distributions

It is easily seen from Fig. 6 that exponential distribution fits mainly to the lower part of the example dataset, while the larger forces are significantly deviated. The fitting is apparently unsatisfactory judging from a simple visual examination. The Q-Q plot presented in Fig. 7 further proves that exponential distribution is not able to give a proper fit especially to the tail. From Fig. 8 it can be observed that with exponential distribution, the R-square values calculated for the tail are rather low for most of the datasets. It can be concluded that exponential distribution is not suitable to describe the selected datasets.

Comparing to exponential distribution, the Weibull distribution shows clearly better fitting performance to the example dataset as shown in Fig. 6. However, the tail part of the dataset still seems not properly fitted, as the Weibull distribution underestimates the force for the same cumulative probability. This is further indicated by the Q-Q plot in Fig. 7, where the R-square value for the tail part is obviously lower than that for the whole example dataset. The generally better fitting to the whole dataset rather than the tail can be seen in the histograms of Fig. 8, where the R-square values are mostly above 0.9 for all data but rarely exceed 0.9 for tail data.

The lognormal distribution shows similar feature of fitting as the Weibull distribution, that the distribution is better fitted to the lower part of the data but deviates more for the tail part. However, unlike the Weibull distributions which has the R-square values all above 0.5, the fitting performance of the lognormal distributions varies significantly, spreading from negative values to near 1. The performance is not as good as the Weibull distributions for most of the cases as shown in Fig. 8.

Therefore, it can be concluded that Weibull distribution gives the best fitting performance among the three standard distributions. Nonetheless, the fittings to the tail are not very good with all the three distributions.



Fig. 6. Measure data in Case 10 and fitted distributions on exponential probability paper: (a) standard and truncated distributions; (b) mixture of exponential distributions.



Fig. 7. Q-Q plots of the data in Case 10 fitted with different distributions, red marks denoting data in the tail while blue marks the remaining data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.1.2. Truncated distributions

As derived in Section 3.2, the truncated exponential distribution is the same as the standard exponential distribution bounded at the threshold, thus the results of fitting are the same. Fig. 6 shows that the truncated Weibull and lognormal distributions fit the dataset better than the standard Weibull and lognormal distributions. The Q-Q plot in Fig. 7 also proves that. In this case, both truncated distributions give rather good agreement with the data, both for the whole dataset and for the tail parts. The truncated Weibull distributions give slightly higher R-square values, but the difference is rather small. The histograms in Fig. 8 shows that both truncated distributions give R-square values over 0.9 for most of the cases in terms of the whole datasets. The R-square values in terms of the tail part are smaller, but still quite high (mostly over 0.8) in contrast to standard distributions. Again, the truncated Weibull distribution shows better performance than the truncated lognormal distribution, especially for the tail part. It can be concluded here that truncated Weibull and lognormal distributions are more suitable for the available datasets comparing to standard distributions bounded at the threshold.

However, there are five of the 20 datasets which cannot be fitted with truncated Weibull or lognormal distributions with maximum likelihood estimation, including both frames of Case 3, both frames of Case 6 and frame #134 of Case 7. These datasets have the common feature that there seems to be a gap which separates the datasets into two parts. For example, the histogram in Fig. 9 shows the distribution of measured ice loads on frame #134.5 of Case 3. There is a gap between 60 kN and 210 kN where no loads are detected but several loads above this gap detected. The dataset then looks like coming from two different distributions, one resulting in smaller loads while the other generating larger loads. Since the number of ice loads encountered during this case is rather small, this might be caused by the randomness of the sample dataset. It is also likely due to the existence of possible ridges or thick ice in the case, which results in the larger forces with a gap. Similar gaps are also observed in other datasets where truncated distributions fail to give MLE solutions. It should be noted that the truncated distribution parameters can still be estimated through e.g. method of moments. However, these are not adopted in order to maintain the consistency in parameter estimation method.

4.1.3. Mixture of exponential distributions

To illustrate the parameter optimization process described in Section 3.3, an example of the EM algorithm in terms of Case 10 with mixture of two exponential distributions is given here. To start with, the parameters $w_1, w_2, \lambda_1, \lambda_2$ are randomly initialized, shown as the left-most points in Fig. 10. At each step, these parameters are updated according to the procedure in Section 3.3, which results in a new set of distribution parameters. The process continues until the results converge. The same procedure is repeated with different initialization of the parameters, which gives different optimization chains. Finally, the optimized parameters $w_1, w_2, \lambda_1, \lambda_2$ are extracted



Fig. 8. Summary of the R-square values of the 20 datasets, dashed lines at 0.9: (a) R-square in terms of each whole dataset; (b) R-square in terms of the tail of each dataset.

from the chain which leads to the maximum likelihood. Fig. 10 shows that most of the chains lead to rather similar converged results, although the starting points are rather different. The log-likelihoods increase after each iteration and finally converge to a similar level.

Fig. 6(b) shows the fitting to the example dataset with mixture of one (i.e. standard/truncated exponential distribution), two, three, five and ten exponential distributions. It can be clearly observed that the distributions get better fitted to the tail with the increase of *m*



Fig. 9. Histogram of the force on frame #134.5 during case 3.



Fig. 10. Results of EM algorithm with mixture of two exponential distributions with different initial values for the distribution parameters, with Case 10 as the example.

(number of components) up to three, but not much improved when m is further increased to ten. The same conclusion can be drawn from the Q-Q plots in Fig. 7 by calculating the R-square values. For the example dataset, it seems adequate to have three mixed exponential distribution components. From Fig. 8 where the results of all datasets are shown, it can be observed that for most cases the R-square values change very little between fittings with m equaling 3, 5 or 10. For some other cases, e.g. Case 1 frame #134.5, it is sufficient to have two components. Such results indicate that the force data may come from two or three different distributions which might be the result of different ship-ice interaction processes.

Table 3 listed the obtained λ and w values sorted by the weights, with different number of exponential distribution components. It can be seen that when m increases from 1 to 3, there are major changes in λ and the corresponding w. However, when m is further

increased, the major components remains similar, while some minor components with rather small weights come out. The weights of the minor components are so small that it does not lead to much change in the obtained probability distribution, which explains the similar performance when *m* increases from 3 to 10. The same information as given in Table 3 is plotted in Fig. 11, where the minor components with weights below 0.01 are left out. This can be seen as the *spectrum* of the fitted mixture distributions, where *w* are plotted as a function of λ . There seems to be three main components, the first located at around λ equaling 0.035, the second between 0.1 and 0.15 and the third at 0.01. When *m* is less than 3, the components cannot be sufficiently separated and thus mixed up. When *m* is larger than 3, the components are adequately separated and increasing the number of components does not further improve the fitting.

4.1.4. Summary of distribution fittings

The above analysis has shown that Weibull distribution as well as truncated Weibull distribution are the best performed distributions among the categories of standard distribution and truncated distribution. It also shows that mixture of three exponential distribution is able to fit the datasets well while adding more components do not further contribute to the performance. To summarize the findings through the large number of distribution fittings (nine distributions, ten cases, two frames), a comparison is made between standard Weibull distribution, truncated Weibull distribution and mixture of three exponential distributions case-wise and rank their performance in terms of R-square values in Table 4. Here the fittings to loads on frame #134.5 are used since similar conclusions can be drawn in terms of frame #134. Overall, the truncated Weibull distributions show similar performance as mixture of three exponential distributions, while standard Weibull distributions perform poorer compared to the other two.

4.2. Long-term estimations

Following Section 4.1.4, we compare the long-term estimation of extreme ice loads using the obtained standard Weibull distribution, truncated Weibull distribution and mixture of three exponential distributions, in order to check how sensitive the estimation is to the parent distributions. The procedure given in Section 3.5 is followed. The results obtained with dataset of frame #134.5 Case 10 are again used as the example.

Fig. 12 presents an example of the obtained Gumbel distributions with n = 10,000, i.e. the probability distribution of the maximum load among 10,000 ship-ice contacts. The characteristic loads (i.e. the most probable loads) are marked with dashed lines. It can be seen that with different parent distributions, the obtained extreme distributions are rather different. In this case, the extreme distribution with truncated Weibull distribution as parent distribution is much flatter than the remaining distributions, and estimates higher characteristic extreme load. This is because truncated Weibull distributions have thicker tails comparing to the other two distributions. Fig. 13 presents the characteristic loads as functions of the return periods in number of loads with different parent distributions. The same can be observed that the estimated characteristic load is rather sensitive on selected parent distributions. For a comparison, the ice class of S.A. Agulhas II is PC5 while the hull is enforced to DNV ICE 10, which requires capacity about 1500 kN on the instrumented frame. The estimated extremes with Weibull distribution and mixture of three exponential distributions are well below this value while that with truncated Weibull distribution exceeds this when the number of contacts increases to 10^5 . It is worth noting that according to Fig. 7, the truncated Weibull distribution and mixture of three exponential distributions differs significantly.

It should be noted that the obtained Weibull and truncated Weibull distribution shape parameters (β in Table 2 and Eq. (4)) are less than 1, thus belong to class E3 parent distribution (see Section 3.5). The extreme distribution thus gets flatter when *n* increases. However, the flatness of the extreme distribution with mixture of exponential distributions remain unchanged regardless of *n*. This may lead to more significant difference in the estimated load level if a probability of exceedance, e.g. 0.01, is considered for safety reason.

5. Discussions and conclusions

This paper has investigated three different categories of probability distributions, namely standard distributions, truncated

| m = 1 | | m=2 | | m = 3 | | m = 5 | | m = 10 | |
|-------|---|-------|------|-------|------|-------|--------|--------|--------|
| Λ | w | λ | w | λ | w | λ | w | λ | w |
| 0.035 | 1 | 0.091 | 0.64 | 0.035 | 0.49 | 0.033 | 0.45 | 0.032 | 0.44 |
| | | 0.016 | 0.36 | 0.138 | 0.39 | 0.116 | 0.40 | 0.111 | 0.35 |
| | | | | 0.010 | 0.13 | 0.010 | 0.11 | 0.010 | 0.11 |
| | | | | | | 0.416 | 0.04 | 0.129 | 0.03 |
| | | | | | | 0.029 | < 0.01 | 0.137 | 0.03 |
| | | | | | | | | 0.455 | 0.03 |
| | | | | | | | | 0.035 | < 0.01 |
| | | | | | | | | 0.319 | < 0.01 |
| | | | | | | | | 0.092 | < 0.01 |
| | | | | | | | | 0.412 | < 0.01 |

Table 3 List of obtained λ and w with different number of exponential distribution components.



Fig. 11. Distribution spectrum of fitted mixture distributions.

Table 4

Comparison of the fitting performance between standard Weibull (SW), truncated Weibull (TW) and mixture of three exponential distributions (ME).

| Case number | Best | 2nd | 3rd |
|-------------|------|-----|-----|
| 1 | ME | TW | SW |
| 2 | TW | ME | SW |
| 3 | ME | SW | |
| 4 | TW | ME | SW |
| 5 | TW | ME | SW |
| 6 | ME | SW | |
| 7 | ME | TW | SW |
| 8 | TW | ME | SW |
| 9 | ME | TW | SW |
| 10 | TW | ME | SW |



Fig. 12. PDF of the extreme load among 10,000 loads with different parent distributions.



Fig. 13. Characteristic loads as a function of return period (in number of loads) with different parent distributions.

distributions, and mixture of exponential distributions. Here the advantages and shortcomings of these probability distributions will be discussed.

5.1. Standard distributions bounded at the threshold

Standard exponential, Weibull and lognormal distributions bounded at the threshold have the advantage that the maximum likelihood solution always exists. Among these, Weibull distribution has shown its superiority in fitting performance comparing to the others. Despite that, Weibull distribution fails to give satisfactory fitting to the tail part of the majority of the cases. For the aim of long-term extreme value analysis, it is then questionable to use these standard distributions bounded at the threshold as the parent distribution due to the sensitivity on the tail part. Nonetheless, if the aim is to investigate fatigue of structures where more focus is put on the lower part of ice loads, Weibull distribution bounded at the threshold may fulfill the required accuracy.

5.2. Truncated distributions

Truncated distributions have the major advantage that the concept naturally follows common sense, that if all ice load (the defined *Y* in Section 3) on a frame follows a certain distribution, the loads above a threshold should follow the corresponding truncated distribution. Using a threshold is due to the practical reason to filter out the loads from waves. It comes then naturally that the original distribution parameters should be independent of the chosen threshold. Truncated distributions keep these features while standard distributions are not capable of. With truncated Weibull and lognormal distributions, the fitting performance is obviously improved comparing to standard distributions, especially at the tail part. This indicates that truncated distributions may be a promising concept for the modelling of measured ice loads which have been cut off by a certain threshold. The downside of truncated distributions is that it does not always give MLE solutions, especially when there is discontinuity in the datasets. However, in these cases the standard distributions are not able to give good fitting either.

5.3. Mixture distributions

Mixture of exponential distributions show promising feature from another perspective. With mixture distributions, the data are identified as coming from several different distribution components and mixed up with certain weight. This is in line with the knowledge of ship-ice interaction, that a ship may break/clear ice through different means, e.g. bending failure or pushing aside. Different ship-ice interaction processes result in different load distribution and therefore naturally mixed together, yielding mixture distributions. The results of this paper indicate that mixture of distributions can give rather good fitting both in terms of the lower part and tail part. The performance is similar to those with truncated distributions. The fitting performance does not increase all the time with more components. Instead, it stabilizes at a small *m* and rarely increases further. This results in a distribution spectrum which describes the distribution composition of a dataset. It is worth investigating whether such spectrum with several major spikes is common for general ice load datasets. It is also relevant to test the method on longer datasets and see how the number of parameters changes.

Another merit of mixture distribution is that it works as well for dataset which are not pure, e.g. the ice condition and ship speed varying a lot or existence of ridged ice. Mixture distributions potentially can automatically separate these as different components while single distributions always treat them as coming from the same distribution. This broadens the capability of using probability distributions to analyze measured ice loads.

5.4. Estimation of extreme loads based on short-term measurements

The results in Section 4.2 have demonstrated that long-term estimations are very sensitive to the selected parent distributions for short-term load peaks. This tells that the choice of the parent distribution form is of vital importance for long-term distribution, even if the fitting performance on short-term load peaks is similar. This leads to a question: how one can make long-term estimations with confidence based on short-term load analysis? One possibility is to carry out validation with long-term load measurements and compare the estimations with measured extremes, in order to confirm which probability distribution leads to correct estimation. This will be possible with the continuous S.A. Agulhas II measurement in the Antarctic ocean. The number of cases presented in this paper is not large enough to draw a firm conclusion on the choice of parent distributions between truncated Weibull and mixture distributions. It is therefore also important to conduct similar studies on other short-term datasets to further compare the performance with different distributions.

A general issue related to short-term load distribution fitting is the estimation of long-term (e.g. lifetime) exposure, which is needed together with the parent distribution to make long-term estimations. This is not in focus of this paper, and is rarely addressed in the literature. It is important for future work to investigate the exposure so that the parent distributions can actually be used for long-term estimation. Another general issue related to probabilistic modelling of ice loads is that the distribution parameters need to be linked to the prevailing ice conditions, so that long-term estimations can be made given a certain operational area. There have been some examples dealing with this issue, e.g. Refs. [7,13,16], attempting to connect distribution parameters to varying ice thickness or/and ship speed. However, influence of other factors such as floe size, hull form and ship maneuvering has not been thoroughly investigated. Overall, there is still extensive work needed before the methodology of modelling short-term load peaks becomes a feasible tool for long-term load estimation.

5.5. Other future work following the findings

This paper has shown the merit of truncated distributions and exponential distributions. A natural question following this is that: will mixture of truncated distributions perform even better? Mixture of truncated distributions can merge the merits of both truncated distributions and mixture distributions, and yield distributions following common sense. It should be noted that although this paper investigates mixture of exponential distributions, the methodology of mixing distributions and the EM algorithm can be applied to mixture of any type of distributions, and even mixture of different types of distributions, e.g. an exponential distribution and a Weibull distribution. This then opens a much wider range to seek the correct distribution for ice load modelling.

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.

Acknowledgement

This project has received funding from the Lloyd's Register Foundation, a charitable foundation, 563 helping to protect life and property by supporting engineering-related education, public engagement and the application of research www.lrfoundation.org.uk. We also than Andrei Sandru for providing the processed images; and Christof M. van Zijl, James-John Matthee and Annie Bekker from Stellenbosch University, who provide enormous helps on visual observation of ice condition during the SANAP 2018/2019 and Weddell Sea Expedition 2019.

Appendix A. Derivation of EM algorithm for mixture of exponential distributions

EM algorithm is a widely adopted algorithm for mixture distributions. We will not prove its validity here since it has been widely accepted. Here we simply derive Eq. (12-14) following the standard EM procedure.

The starting point is Eq. (11). We will use z which equals x-T instead of x for simplicity. The PDF of z can be written as:

$$f(z) = \sum_{j=1}^{m} w_j \lambda_j e^{-\lambda_j z}$$
(I.1)

The PDF of z can be expressed indirectly via a latent variable, **q**, which is a binary vector (i.e. elements either 0 or 1) of *m* elements, with the following PDF:

$$f(\mathbf{q}) = \sum_{j=1}^{m} w_j^{q_j} \tag{I.2}$$

with the constraint $\sum_{j=1}^{m} q_j = 1$, i.e. only one of the **q** elements equals 1 and the rest equal 0. Equivalently, $f(\mathbf{q}_j = 1) = w_j$.

The PDF of z is expressed via **q** by the basic probability rule:

$$f(z) = \sum_{q} f(\mathbf{q}) f(z|\mathbf{q})$$
(I.3)

where the conditional probability $f(z|\mathbf{q})$ is:

$$f(z|\mathbf{q}) = \prod_{j=1}^{m} \left(\lambda_j \exp(-\lambda_j z)\right)^{q_j}$$
(I.4)

The joint PDF of z and q is

$$f(z,\mathbf{q}) = f(\mathbf{q})f(z|\mathbf{q}) = \prod_{j=1}^{m} \left(w_j \lambda_j \exp\left(-\lambda_j z\right) \right)^{q_j}$$
(I.5)

Therefore, denoting γ_i as the PDF of q_i conditional on z, γ_i can be written as:

$$\gamma_j = f\left(q_j = 1 \mid z\right) = \frac{f\left(z, q_j = 1\right)}{f(z)} = \frac{w_j \lambda_j \exp\left(-\lambda_j z\right)}{\sum_{j=1}^m w_j \lambda_j e^{-\lambda_j z}}$$
(I.6)

For a dataset $\mathbf{z} = (z_1, z_2, ..., z_n)$, Eq. (I.5) and (I.6) can be rewritten as:

$$f(z_i, \mathbf{q}_i) = \prod_{i=1}^{m} \left(w_i \lambda_j \exp\left(-\lambda_j z_i\right) \right)^{q_{ij}}$$
(I.7)

$$\gamma_{ij} = \frac{w_j \lambda_j \exp\left(-\lambda_j z_i\right)}{\sum_{i=1}^m w_j \lambda_j e^{-\lambda_i z_i}}$$
(I.8)

which gives Eq. (13). The vectors q_i forms a matrix, denoted by Q.

With Eq. (I.7), the log-likelihood of the joint PDF of z and Q is

$$f(z, \mathbf{Q} | \mathbf{w}, \boldsymbol{\lambda}) = \log\left(\prod_{i=1}^{n} \prod_{j=1}^{m} \left(w_{j} \lambda_{j} \exp\left(-\lambda_{j} z_{i}\right)\right)^{q_{ij}}\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} q_{ij} \left(\log w_{j} + \log \lambda_{j} - \lambda_{j} x_{i}\right)$$
(I.9)

So the expectation of the log-likelihood with regards to Q is:

$$E[f(\boldsymbol{z}, \boldsymbol{Q} | \boldsymbol{w}, \boldsymbol{\lambda})] = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} \left(\log w_j + \log \lambda_j - \lambda_j x_i \right)$$
(I.10)

The final step is to take the derivative of the expectation with regards to w_i and λ_i and set the derivatives to zero, which gives:

$$w_j = \frac{\sum_{i=1}^n \gamma_{ij}}{n} \tag{I.11}$$

$$\lambda_j = \frac{\sum_{i=1}^n \gamma_{ij}}{\sum_{i=1}^n Z_i \gamma_{ii}}$$
(I.12)

The EM algorithm is realized by iterating with Eq. (I.8), (I.11) and (I.12) to constantly update w and λ . The iteration constantly increases the likelihood function $f(\mathbf{z}|\mathbf{w}, \lambda)$ and converges to a local maximum value. One need to set different initial w and λ and run the

algorithm multiple times to find the global maximum value.

Appendix B. Q-Q plots of all scenarios



Fig. B.1. Q-Q plot of distribution fittings to the data in Scenario 1; top: fr. #134.5, bottom: fr. #134



Fig. B.2. Q-Q plot of distribution fittings to the data in Scenario 2; top: fr. #134.5, bottom: fr. #134



Fig. B.3. Q-Q plot of distribution fittings to the data in Scenario 3; top: fr. #134.5, bottom: fr. #134



Fig. B.4. Q-Q plot of distribution fittings to the data in Scenario 4; top: fr. #134.5, bottom: fr. #134



Fig. B.5. Q-Q plot of distribution fittings to the data in Scenario 5; top: fr. #134.5, bottom: fr. #134



Fig. B.6. Q-Q plot of distribution fittings to the data in Scenario 6; top: fr. #134.5, bottom: fr. #134



Fig. B.7. Q-Q plot of distribution fittings to the data in Scenario 7; top: fr. #134.5, bottom: fr. #134



Fig. B.8. Q-Q plot of distribution fittings to the data in Scenario 8; top: fr. #134.5, bottom: fr. #134



Fig. B.9. Q-Q plot of distribution fittings to the data in Scenario 9; top: fr. #134.5, bottom: fr. #134



Fig. B.10. Q-Q plot of distribution fittings to the data in Scenario 10; top: fr. #134.5, bottom: fr. #134

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marstruc.2021.103049.

References

- [1] Daley C, Alawneh S, Peters D, Colbourne B. GPU-Event-Mechanics evaluation of ice impact load statistics. Arct. Technol. Conf. 2014.
- [2] Lau M, Lawrence KP, Rothenburg L. Discrete element analysis of ice loads on ships and structures. Ships Offshore Struct 2011;6:211-21.
- [3] Huang L, Tuhkuri J, Igrec B, Li M, Stagonas D, Toffoli A, et al. Ship resistance when operating in floating ice floes: a combined CFD&DEM approach. Mar Struct 2020;74:102817.

F. Li et al.

- [4] Yang B, Sun Z, Zhang G, Wang Q, Zong Z, Li Z. Numerical estimation of ship resistance in broken ice and investigation on the effect of floe geometry. Mar Struct 2021;75:102867.
- [5] Lubbad R, Løset S, Lu W, Tsarau A, van den Berg M. Simulator for arctic marine structures (SAMS). In: Proc. ASME 2018 37th int. Conf. Ocean. Offshore arct. Eng. (OMAE2018), madrid, Spain; 2018.
- [6] Kujala P, Vuorio J. On the statistical nature of the ice-induced pressures measured on board IB Sisu. In: Port ocean eng. Under artic cond., narssarssuaq, Greenland; 1985.
- [7] Kotilainen M, Vanhatalo J, Suominen M, Kujala P. Predicting ice-induced load amplitudes on ship bow conditional on ice thickness and ship speed in the Baltic Sea. Cold Reg Sci Technol 2016;135:116–26.
- [8] Suominen M, Kujala P. Variation in short-term ice-induced load amplitudes on a ship's hull and related probability distributions. Cold Reg Sci Technol 2014.
 [9] Suyuthi A, Leira BJ, Riska K. Statistics of local ice load peaks on ship hulls. Struct Saf 2013;40:1–10.
- [0] Stytum A, Leina BJ, Riska K. A generalized probabilistic model of ice load peaks on ship hulls in broken-ice fields. Cold Reg Sci Technol 2014;97:7–20.
- [10] Jordan JJ, Masa K, Agartanza probabilistic index of load ice pressions in Jordan Actic English (193:11583-9)
- [12] Taylor RS, Jordaan IJ, Li C, Sudom D. Local design pressures for structures in ice: analysis of full-scale data. J Offshore Mech Arctic Eng 2010;132:1–7.
- [13] Kujala P. Semi-empirical evaluation of long term ice loads on a ship hull. Mar Struct 1996;9:849-71.
- [14] Kujala P, Jiang Z, Li F, Lu L. Long term prediction of local ice loads on the hull of S. A. Agulhas II. Proc 25th int conf port ocean eng under arct cond delft. Netherl 2019.
- [15] Chai W, Leira BJ, Naess A. Probabilistic methods for estimation of the extreme value statistics of ship ice loads. Cold Reg Sci Technol 2018;146:87–97.
- [16] Shamaei F, Bergström M, Li F, Kujala P, Taylor R. Analysis and development of a probabilistic method for calculating local peak ice pressures on ships. Mar Struct 2020;74.
- [17] Li F, Khawar M, Sandru A, Lu L, Suominen M, Kujala P. Full-scale measurement of ship performance and ice loads in Antarctic ice floe fields. In: Port ocean eng. Under arct. Cond., Moscow, Russia; 2021.
- [18] Suominen M, Li F, Lu L, Kujala P, Bekker A, Lehtiranta J. Effect of maneuvering on ice-induced loading on ship hull: dedicated full-scale tests in the Baltic sea. J Mar Sci Eng 2020;8:1–30.
- [19] Lu L, Kujala P. Full-scale measurements on board S.A. Agulhas II during SANAP 2018/2019 and Weddell Sea expedition 2019. 2019.
- [20] Lu L, Kujala P, Banda OAV, Goerlandt F. Ramming induced ice loads between the ship bow and multiyear thick ice in antarctic weddell sea. In: Dev. Collis. Grounding ships offshore struct. - proc. 8th int. Conf. Collis. Grounding ships offshore struct.; 2020.
- [21] Suominen M, Romanoff J, Remes H, Kujala P. The determination of ice-induced loads on the ship hull from shear strain measurements. Anal Des Mar Struct V 2015:375–83.
- [22] Suominen M, Aalto H, Kulovesi J, Lensu M, Kujala P, Soal K, et al. Ship-based visual Antarctic sea ice thickness measurements. Ann Arbor, Michigan USA: 23nd IAHR Int. Symp. Ice; 2016. p. 1–8.
- [23] Suominen M, Bekker A, Kujala P, Soal K, Lensu M. Visual antarctic sea ice condition observations during austral summers 2012-2016. Proc. Int. Conf. Port Ocean Eng. under Arct. Cond. POAC 2017.
- [24] Sandru A, Hyyti H, Visala A, Kujala P. A complete process for shipborne sea-ice field analysis using machine vision. In: Int. Fed. Autom. Control world congr., berlin, Germany; 2020.
- [25] Wingo DR. The left-truncated Weibull distribution: theory and computation. Stat Pap 1989;30:39-48.
- [26] Gupta RD, Kundu D. Generalized exponential distribution: existing results and some recent developments. J Stat Plann Inference 2007;137:3537-47.
- [27] Jewell BYNP. Mixtures of exponential distributions. Ann Stat 1982;10:479-84.
- [28] Bishop CM. Pattern recognition and machine learning. Springer; 2006.
- [29] Jordaan I. Decisions under uncertainty: probabilistic analysis for engineering decisions, vol. 9780521782; 2005.