NTNU

Norwegian University of
Science and Technology

# Probabilistic Methods for Estimation of Extreme Ice Loads on Ships

Mathias Borgnes

Marine Technology
Submission date: June 2018
Supervisor: Bernt Johan Leira, IMT

Norwegian University of Science and Technology
Department of Marine Technology

**NTNU**
**Norwegian University of Science and Technology**
*Department of Marine Technology*

Master Thesis, Spring 2018
for
Master Student Mathias Borgnes

# Probabilistic Methods for Estimation of Extreme Ice Loads on Ships

## *Statistiske metoder for estimering av ekstreme islaster på skip*

The Arctic Ocean is characterized by harsh and cold climate, and large areas are covered by ice most of the year. It is accordingly crucial to understand how ice interacts with e.g. ships and other floating structures which are operating in these waters. However, the behaviour of ice is highly complex, and models describing ice-ship interactions and loading on the hull are characterized by great uncertainties.

As part of the search for more accurate knowledge, DNV (now DNV GL) performed measurements on the Norwegian coast guard vessel KV Svalbard in 2006, 2007, 2011 and 2012. The measurements provided a large amount of valuable data, including global motions and strains in the bow region during ice loading. A subset of these data is to be applied as basis for the present study.

The following subjects are to be addressed as part of this work:

- Describe different types of sea ice and their mechanical and physical properties. Also, give a brief review of the ice-breaking process with associated pressures and loads acting on the hull
- A review of the state of the art methods for describing short-term extreme ice loads shall be carried out.
- Implement the methods which are to be used for analyses in MATLAB.
- Investigate the effects of ice thickness and vessel speed on the estimated extreme values.
- Perform a comparative study of state of the art existing methods for describing short-term extreme ice loads, using the measurements from KV Svalbard to reveal their strengths and weaknesses.

The work scope may prove to be larger than initially anticipated. Subject to approval from the supervisor, topics may be deleted from the list above or reduced in extent.

In the thesis the candidate shall present his personal contribution to the resolution of problems within the scope of the thesis work. Theories and conclusions should be based on mathematical derivations and/or logic reasoning identifying the various steps in the deduction.

The candidate should utilise the existing possibilities for obtaining relevant literature.

*NTNU*
*Faculty of Engineering Science and Technology*
*Norwegian University of Science and Technology*
*Department of Marine Technology*

The thesis should be organised in a rational manner to give a clear exposition of results, assessments, and conclusions. The text should be brief and to the point, with a clear language. Telegraphic language should be avoided. The thesis shall contain the following elements: A text defining the scope, preface, list of contents, summary, main body of thesis, conclusions with recommendations for further work, list of symbols and acronyms, references and (optional) appendices. All figures, tables and equations shall be numbered.

The supervisor may require that the candidate, in an early stage of the work, presents a written plan for the completion of the work. The original contribution of the candidate and material taken from other sources shall be clearly defined. Work from other sources shall be properly referenced using an acknowledged referencing system.

The thesis shall be submitted in electronic form:

- Signed by the candidate
- The text defining the scope included
- Drawings and/or computer prints which cannot be bound should be organised in a separate folder.

Supervisor: Professor Bernt J. Leira

Deadline: June 11th 2018

Trondheim, January 15th, 2018

Bernt J. Leira

# Preface

The following report is an individual master's thesis carried out during the spring of 2018 at the Department of Marine Technology, Norwegian University of Science and Technology (NTNU) in Trondheim. This work is the final project of my Master's degree at NTNU within the field of marine structures.

The thesis is partly a continuation of the project thesis conducted in the course "TMR4500 - Marine Structures, Specialization Project", which was within ice engineering. However, my personal interest and fascination for statistics have strongly influenced the subject of the thesis, which is related to short-term extreme value statistics of local ice loads on ships.

Throughout the semester, more time than anticipated has been devoted to implementing various probabilistic methods in MATLAB. However, the implementation has been useful, both in terms of getting a deeper understanding of the theory behind the methods, and in terms of improving my programming skills.

I would like to thank my supervisor Professor Bernt J. Leira for helpful discussions and guidance whenever needed. Further, I wish to thank Professor Kaj Riska for allowing me to attend lectures in the module "Ship Design for Ice" during fall 2017, which provided crucial knowledge about ice and ice engineering. I would also like to thank professors and fellow students who have contributed by answering various questions and for having constructive discussions.

Mathias Borgnes
Trondheim, June 8, 2018

# Abstract

The maritime activity in the Arctic Ocean is ever increasing, as the reduced ice extent reveals new resources and new shipping routes. Ice loading processes due to ship-ice interaction are highly complex and yet not fully understood. Many probabilistic methods for describing the ice loading process and its extreme values are proposed in the literature. However, all of them seem to have flaws or weaknesses, thus no universal method is verified. The purpose of this thesis is to compare different methods for estimation of short-term extreme loads, with a focus on revealing their strengths and weaknesses. Full-scale measurements from KV Svalbard are used for analyses. The applied probabilistic methods are; the classical approach, the asymptotic approach, application of a three-parameter exponential distribution, and the average conditional exceedance rate (ACER) method.

A review of sea ice and ice formation is performed. Mechanical and physical properties of ice are described, i.e., strength, ductility, fracture toughness and Young's modulus. The most critical parameters for the properties were found to be porosity, strain rate, temperature and macrostructure. Moreover, the ice-breaking process is described with associated pressures and loads.

Many of the applied methods are based upon the assumption of stationarity. However, since the ice loading process is highly complex, it is not straightforward which conditions and characteristics that must be stationary to provide stationarity for the measured ice load peaks. Previous works on the same subject present different requirements when stationarity is strived for. Therefore, a brief study on how ice thickness and vessel speed affect the estimated extreme ice load is carried out. The study provides strong indications that the ice thickness has a significant impact on the estimated extreme load, whereas no correlation was found for the vessel speed.

The applied probabilistic methods for estimating extreme loads are described and implemented in MATLAB. Several time series for analyses were identified based on the study related to stationarity. Different categories of load patterns were observed for the data sets. Four of the data sets were selected for comparison of the methods, where the most frequently observed load patterns were represented; one set containing an outlier, one set where two load populations were identified, one set with scattered loads in the upper tail, and a set that exhibited a highly stationary load pattern.

Most of the applied methods provide good fits and estimates of extreme loads if the data sets are selected with some care. Thus, the selection of method should depend on what kind of data that is being analyzed, and what the purpose of the application is. The exponential and Weibull distributions provide good fits to data collected during stationary ice conditions. However, stationary ice conditions are rarely encountered in Arctic waters, which makes these methods unsuitable for live estimation of extreme loads. The three-parameter exponential distribution is more flexible than the regular exponential distribution, by means of the ability to capture a more extensive variety of load patterns. The ACER method provides inconsistent results when applied to time series of short duration (less than 30 min), which made it difficult to compare with the other methods. The asymptotic approach using type I extreme value distribution, i.e., the Gumbel, was found to be the most robust method, in the sense of providing satisfactory fits and estimates of extreme loads for a wide range of load patterns. It was also found to be better than the ACER method for 6 hours non-stationary time series. However, none of the methods were able to capture outliers, which were found for several data sets. It is questionable whether the outliers are actual loads or measuring errors. This should be investigated to be able to develop accurate models for extreme load estimation.

# Sammendrag

Den marine aktiviteten i Arktis øker stadig, da den reduserte isutbredelsen frigjør nye ressurser og nye transportruter. Isbelastninger som følge av skip-is-interaksjon er svært komplekse og ennå ikke forstått fullt og helt. En rekke statistiske metoder for å beskrive isbelastningsprosesser og tilhørende ekstremverdier er foreslått i litteraturen, men alle synes å ha feil eller svakheter, og det eksisterer ingen verifisert universell modell. Formålet med denne oppgaven er å sammenligne ulike metoder for estimering av korttids ekstrembelastninger, med fokus på styrker og svakheter ved de ulike metodene. Fullskala målinger fra KV Svalbard er brukt til analyser. De anvendte metodene er; klassisk tilnærming, asymptotisk tilnærming, anvendelse av en tre-parameter eksponentialfordeling, og "average conditional exceedance rate" (ACER) metoden.

En litteraturstudie av sjøis og isdannelse er gjennomført, og mekaniske og fysiske egenskaper ved is er beskrevet. De viktigste parametrene for isegenskapene ble funnet å være porøsitet, belastningsrate, temperatur og makrostruktur. Videre er isbryting beskrevet, med tilhørende trykkfordelinger og belastninger.

Mange av de anvendte metodene er basert på antagelsen om stasjonæritet. Siden islaster er komplekse kan det være vanskelig å si hvilke betingelser og egenskaper som må være stasjonære for å oppnå en stasjonær isbelastningsprosess. Tidligere arbeider på det samme emnet presenterer ulike krav når stasjonæritet er etterstrebet. Det er derfor utført en kort studie på hvordan istykkelse og skipshastighet påvirker den estimerte ekstremlasten. Studien gir sterke indikasjoner på at istykkelsen har stor innvirkning på den estimerte ekstremlasten, mens det ikke ble funnet noen sammenheng mellom fartøyets hastighet og den beregnede ekstremlasten.

De anvendte statistiske metodene for estimering av ekstremlaster er beskrevet og implementert i MATLAB. Flere tidsserier for analyser ble identifisert basert på studien relatert til stasjonæritet. Ulike kategorier av islastfordelinger ble observert for datasettene. Fire datasett ble brukt for å sammenligne de probabilistiske metodene, der de hyppigst observerte lastfordelingene var representert; ett sett med en "outlier", ett sett hvor to lastpopulasjoner ble identifisert, ett sett med spredte laster i den øvre halen, og ett sett som utviste meget stasjonær lastfordeling.

De fleste av de anvendte metodene gir gode tilpasningskurver og estimater av ekstremlaster dersom datasettene velges med omhu. Valg av metode bør derfor avhenge av hvilken type data som analyseres, og hva formålet med analysen er. Eksponential- og Weibull-fordeling gir gode tilpasninger til datasett som stammer fra stasjonære isforhold. Imidlertid finner man sjelden stasjonære isforhold i arktiske farvann, noe som gjør disse metodene uegnede for live-estimering av ekstremlaster. Eksponentialfordelingen med tre parametre er mer fleksibel enn den ordinære eksponentialfordelingen, i form av evnen til å fange opp et større spekter av underliggende fordelinger. ACER-metoden gir inkonsekvente resultater for tidsserier med kort varighet (mindre enn 30 min), hvilket gjorde det vanskelig å sammenligne metoden med de andre metodene. Den asymptotiske tilnærmingen ved bruk av type I ekstremverdifordeling, det vil si Gumbel-fordeling, viste seg å være den mest robuste metoden, i form av å gi gode estimater av ekstremlaster for et bredt spekter av underliggende fordelinger. Den var også bedre enn ACER-metoden ved analyse av 6 timers ikke-stasjonære tidsserier. Imidlertid var ingen av metodene i stand til å fange opp outliere, som ble funnet i flere datasett. Det kan stilles spørsmål ved om outliere er faktiske laster eller målefeil. Dette bør undersøkes nærmere for å kunne utvikle gode statistiske modeller for ekstreme islaster.

# Contents

# List of Figures



45

# List of Tables

# 1 Background

Over the past years, activities related to exploitation of natural resources, such as oil and gas, minerals and biological resources in the Arctic areas have increased. Decreasing ice extent has revealed the Northeast Passage open for longer periods, which opens up for transportation of goods and commodities in the Arctic areas. The number of transit voyages has increased the last ten years, and the trend is expected to continue.

For ships operating in ice-infested waters, the presence of ice is a challenge, both in a design perspective and an operational perspective, as the ice induces large local loads on the hull. Moreover, the loading process and its extreme values are difficult to describe by probabilistic methods, although several models are proposed in the literature. As a response to these challenges, DNV (now DNV GL) has performed measurements during full-scale trials on the Norwegian coast guard vessel KV Svalbard in 2006, 2007, 2011 and 2012. The measurements are a part of an Ice Load Monitoring (ILM) project. The measurements provide valuable data, including strains in the bow region during ice loading.

Accurate probabilistic models for describing ice loads are valuable, both from a design point of view and for navigation purposes. This thesis provides a review and a comparative study of state of the art existing methods for estimation of short-term extreme ice loads, using the measurements from KV Svalbard to reveal their strengths and weaknesses.

# 2 Ice

## 2.1 Ice Infested Areas

Sea ice is found in the Arctic and Antarctic areas. The ice conditions are somewhat different in the two areas, mainly because Antarctica is a continent surrounded by a sea, whereas the Arctic is a sea surrounded by continents (Riska, 2017a). On average, almost half of the area that is covered with sea ice during winter in the Arctic survives the summer (*Arctic vs. Antarctic*, 2017). Ice that survives the summer will continue to grow the following winter, forming multi-year ice. Multi-year ice has different properties than first-year ice, which will be discussed later on. Ice properties will be discussed in Section 2.3. In the Antarctic, most of the sea ice melts during the Antarctic summer, making the presence of multi-year ice limited to only a few areas. Figure 1 shows the ice extent in the Arctic between 1979 and 2012 measured in March and September. The September trend indicates that Arctic oceans may be completely ice-free during summer in a relatively near future, which opens up for more activity. An important consequence for ice-going vessels is that there will be less multi-year ice, some years maybe not any at all. The decrease in ice extent during winter does not show the same dramatic trend, and the trend indicates that there will be sea ice during winter in the Arctic in the foreseeable future. However, the low decrease rate in ice extent during winter may be explained by more factors than the climate alone. Whereas in summer the ice edge is somewhere at sea, the ice edge in winter is bordered in many places by land (Russian Siberia, Canadian Arctic islands, Alaska, Greenland and Spitsbergen). In truth, there are only three locations where the ice edge is at sea in winter: the Labrador Sea, the Barents/Greenland Sea and the Bering Sea. This is a small fraction of the total ice edge length, and a slightly warmer climate will only affect the location of the ice border in these areas. Thus, the measurements may not fully reflect an increasing temperature.

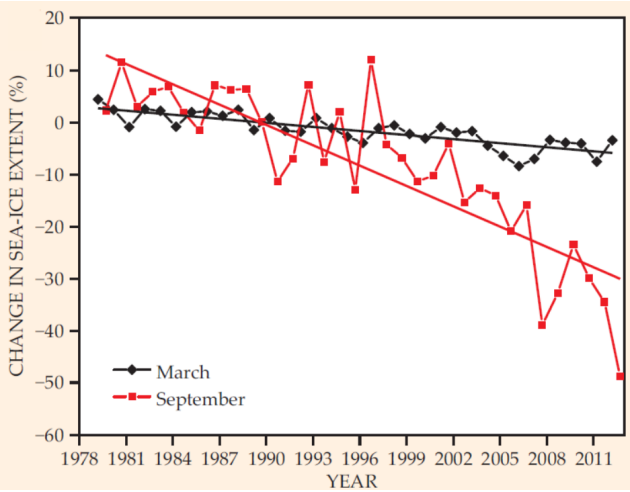



Figure 1: Ice extent in the Arctic (Jeffries et al., 2013).

## 2.2 Formation of Ice and Ice Features

Sea ice forms when seawater reaches the freezing temperature, which depends on the salinity of the water. For salinity of 35 [ppt], which is typical for seawater, the freezing temperature is -1.8 °C (*Salinity and Brine*, 2017). When ice crystals are formed, they are oriented in random directions, and these needle formed crystals start to grow perpendicular to their so-called c-axis. If the sea is calm, these needles develop very rapidly at low temperatures and change to plate crystals (Zubov, 1963). These thin ice sheets, called nilas, will develop in the horizontal plane, resulting in vertical c-axes. With all c-axes pointing in the same direction, the ice becomes anisotropic. As the layers continue to grow, they are pushed on top of each other, see Figure 3 (a). For thin ice, the two layers will alternately be on top of each other, a phenomenon called finger rafting.

As long as the air temperature is colder than the freezing temperature, $T_f$, and the water temperature is sufficiently low, heat will conduct from the water to the air, through the ice, causing the ice to grow thicker. Since the heat must be transferred through the ice, thick ice grows slower than thin ice. Ice covered with snow is also growing slower, as snow is a good insulator. There are many suggested formulas for how to describe the growth rate of the ice thickness. Many of them are variants of Stefan's formula, developed by the Slovenian physicist Josef Stefan in 1891. Stefan's formula says that the ice thickness is proportional to $\sqrt{t \cdot (T_f - T_{air})}$, where $t$ is the time and $T_{air} < T_f$ (Weber, 2009). In the Arctic, first-year level ice is typically 1.5 - 2.0 meters thick at the end of winter. Multi-year ice can be several meters thick.

When seawater freezes, the salt is rejected by the ice. Solutions with very high salinity, called brine, get trapped inside of the ice, making it porous and weak. Over time, brine is drained out by gravity drainage and flushing (Understeiner, 1968), making the salinity in top layers of multi-year ice reduced to less than 10 % compared to newly formed ice (Wettlaufer, 2011).

Icebergs are formed when ice breaks off from glaciers, and its properties are somewhat different compared to sea ice. The glacier ice is created when snow is compressed under its own weight. In the upper layers, the ice crystals are randomly oriented, because the snowflakes have settled randomly. The ice is deforming by gliding along its basal planes, which are perpendicular to the c-axis. As the ice deforms, the c-axes rotate towards the axis of compression, making the crystals no longer randomly oriented (*Ice crystal structure*, 2017). However, glacier ice is often considered as an isotropic material (Sanderson, 1988). Since icebergs are not made out of seawater, there is no brine in the ice, making it less porous and stronger than sea ice formed at sea. Ice-going vessels are not designed for collisions with large icebergs, and impacts are characterized as accidents. While icebergs are easily avoided by ships, they are a bigger issue for offshore installations. However, growlers that are trapped inside the sea ice may be hard to detect, and can cause substantial damages to ships.

Winds and currents induce drifting and stresses in the ice layers (Eriksson et al., 2007). When ice layers are compressed against each other, the ice is crushed and piled up, and ice ridges are formed (see Figure 3 (b)). Ice ridges with keel depths of up to 50 meters have been measured (Johnsten et al., 2009). There is no consensus on how deep such ridges can be, or if there is a limit at all. An asymptotic value of 25 times the ice thickness has been suggested (Riska, 2017a). Although the keel depth of an ice ridge can be very large, the sail height is rarely higher than a few meters, making them difficult to detect. Young ice ridges have a porosity of approximately 30 % due to the void between the crushed ice (Riska, 2017e). As the time goes, the water filling up the voids may freeze,

forming consolidated ridges. Multi-year ice ridges tend to be wider, but less deep than younger ridges, with a typical angle between the keel and the horizontal plane of 25°, as indicated in Figure 3 (e) (Riska, 2017e). Ice ridges usually form perpendicular to the direction of compression, making them more or less parallel to each other. A typical pressure ridge is shown in Figure 2.

Figure 2: Typical pressure ridge in the Arctic Ocean (Alexander & Alexander, n.d.).

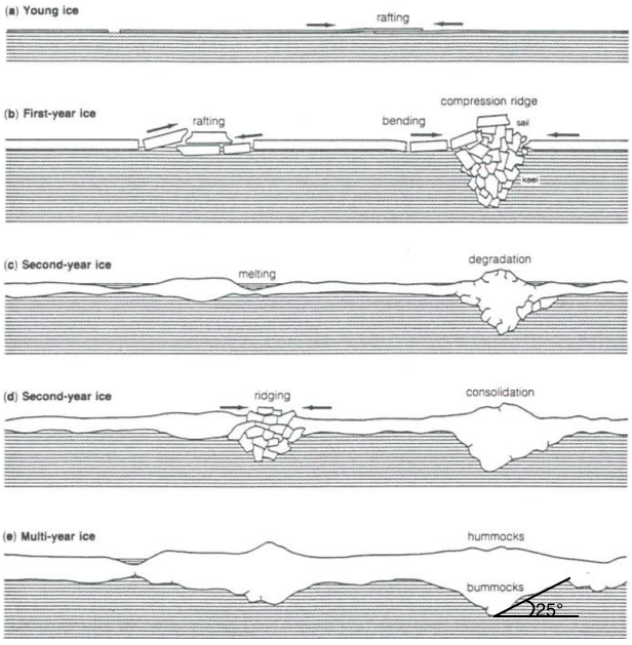



Figure 3: Formation of ice ridges (Riska, 2017b).

## 2.3 Mechanical Properties

The properties of ice come along with great variations, depending on many parameters. Accurate models have turned out to be challenging to develop, even inside the laboratories. Most of the research, data and models are on freshwater ice, which may not always be applicable for saline sea ice due to differences in microstructure (Schapery, 1997). As explained in section 2.2, ice is often anisotropic, thus many mechanical properties are dependent on the direction considered.

It is convenient to divide the parameters affecting the mechanical properties into material type and state type parameters. The most important parameters are discussed below.

### 2.3.1 Porosity and Density

One of many anomalies with water is the fact that it is less dense as a solid than as a liquid. The density of ice varies from less than $600\frac{kg}{m^3}$ to about $930\frac{kg}{m^3}$, depending on, among others, air pockets, brine content, microstructure and temperature (Riska, 2017b). The porosity is the sum of the fractions of brine pockets and air pockets, although the latter is usually much lower than the former (Schulson & Duval, 2009). Porosity is an important material type parameter. The strength and Young's modulus decrease with increased porosity, as seen in Figure 4, and the same applies for reduced density (Langleben & Pounder, 1961). Figure 5 shows that the porosity has a great impact on the fracture toughness for sea ice (Sammonds et al., 1999).

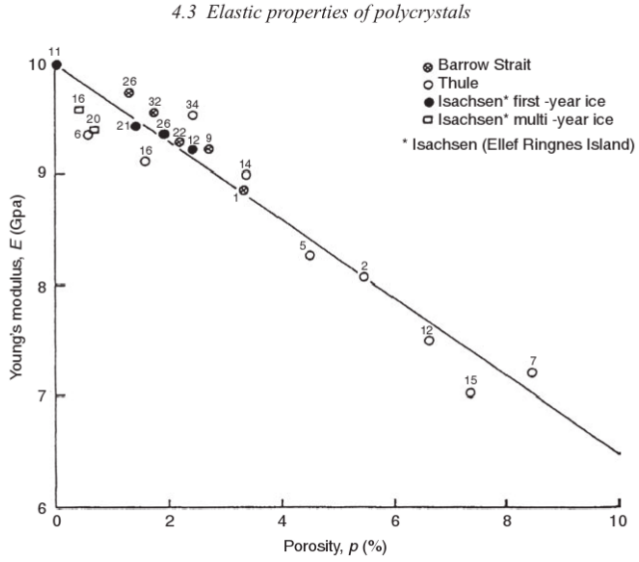



Figure 4: Youngs modulus plotted against porosity for arctic sea ice, based on measurements (Langleben & Pounder, 1961).

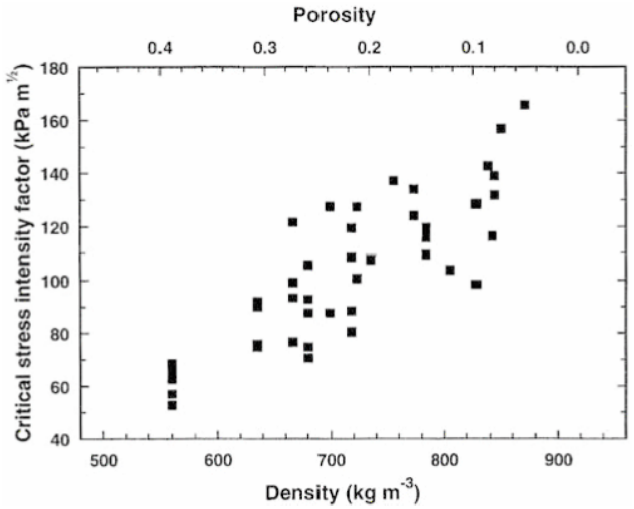



Figure 5: Fracture toughness plotted against porosity based on experiments on fresh water ice (Sammonds et al., 1999).

### 2.3.2 Strain Rate

The strain rate, $\dot{\varepsilon}$, which is defined as the change in strain per second, is an important state type parameter for the mechanical properties. Tensile and compression strength are highly dependent on the strain rate (Jones (1997); Schulson and Duval (2009); and more). For low strain rates, the ice behaves in a more ductile manner than for higher strain rates. This is illustrated in Figure 6 (Schulson & Duval, 2009). As seen in the figure, stress relaxation occurs before failure at low strain rates. As the strain rate increases, a ductile to brittle transition takes place. In Figure 7, compressive strength is plotted against a variation of strain rates for three different data sets (Jones, 1997). Figure 6 indicates that there is a slight reduction in peak stress in the brittle regime. This is partly in compliance with two of the data sets in Figure 7, while the third (+) data set shows that the strength continues to increase for high strain rates, or at least remain steady. Thus, there is some discrepancy about this topic, which is found for other data sets as well. For ice-going vessels, high strain rates will dominate the impacts between ice and the hull, resulting in high peak stresses. One should note that the data shows greater scatter in the brittle range (Jones, 1997).

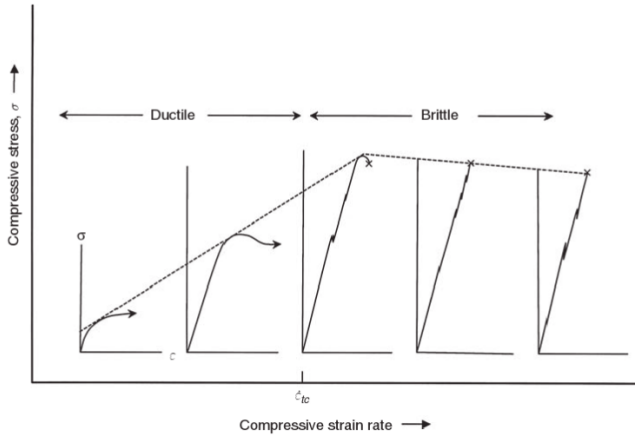



Figure 6: Stress-strain curves for different strain rates (Schulson & Duval, 2009).

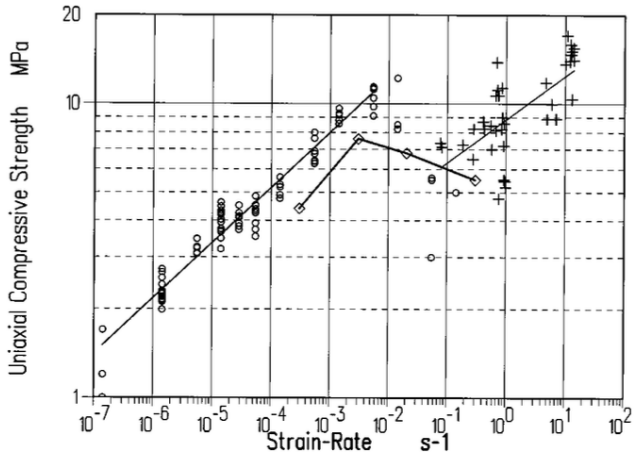



Figure 7: Compressive strength plotted against strain rate for three different data sets; ◯ , ◊ and + (Jones, 1997).

### 2.3.3 Temperature

The temperature $T$ is another important state type parameter for the mechanical properties of sea ice, both during the formation of the ice and afterwards. The surrounding temperature affects the growth rate of the ice, which has an impact on the brine and air content. Fast-growing ice will be more porous, as more brine and air are trapped.

The temperature also has a direct effect on the mechanical properties of the ice. As for most other materials, both brittleness and strength increase for decreasing temperatures. Young's modulus $E$ is also dependent on the temperature. Several models relating $E$ and $T$ have been suggested. Sinha (1989) presented the following relation between Young's modulus and the temperature:

$$E(T) = 9.61 + 1.1 \cdot 10^{-2} \cdot (T_m - T) \tag{1}$$

where $T_m$ is the melting temperature, and $T$ is the temperature of the ice, both taken in Kelvin or Celsius.

For ship-ice interactions, the temperature dependency on the mechanical properties is relatively low compared to the dependency of other parameters, e.g., indention rates. The dependency is also low compared to the large uncertainties related to the properties and material models for sea ice in general. Schulson and Duval (2009) found that the tensile strength of columnar-grained freshwater ice increase by only 10 % when the temperature was reduced from 0 °C to −30 °C. For constant air and water temperature, the temperature varies linearly through the ice sheet, with the temperature $T_m$ closest to the water. Thus, the variation of the average temperature of the ice sheet is significantly lower than for air temperatures.

### 2.3.4 Macrostructure

The macrostructure refers to the grain size and whether the ice has a granular or columnar type of structure. The tensile strength of ice shows great dependency on grain size, which is a material type parameter. As seen from the data in Figure 8, the tensile strength is proportional to $d^{-0.5}$, where $d$ is the grain size (Schulson & Duval, 2009). By comparing the data in the two plots in Figure 8, one can see that the strain rate affects the dependency on the grain size. This illustrates the complexity of ice mechanics, and why it is so difficult to obtain good models. Timco and Frederking (1983) showed that the confined compressive strength is extremely sensitive to the structure of the ice.

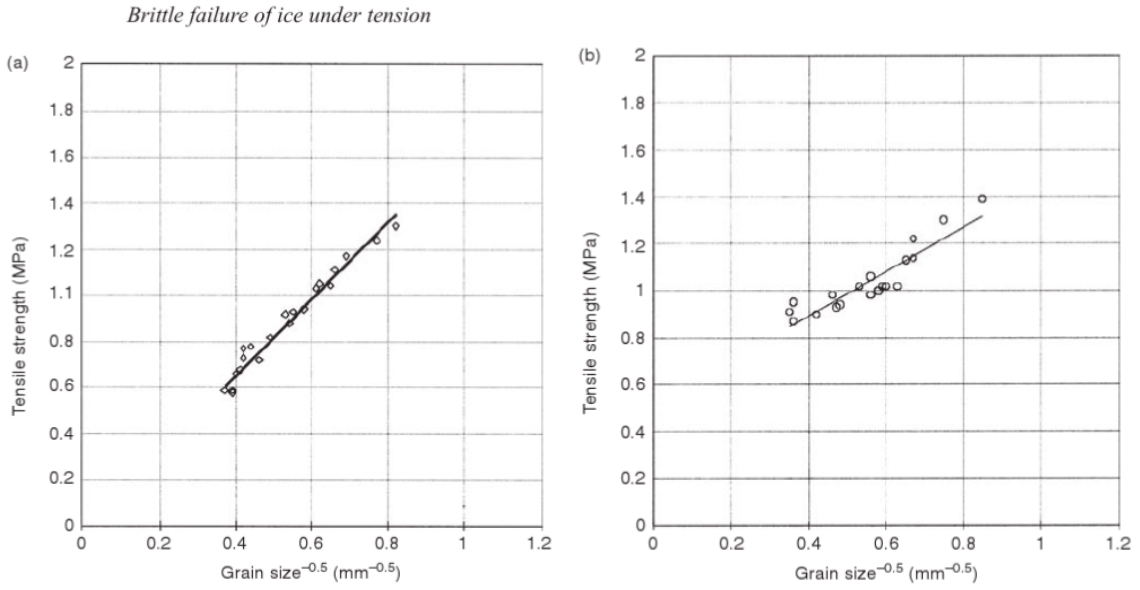



Figure 8: Tensile strength for different grain sizes, at indention rates of $10^{-7}$ $s^{-1}$ (a) and $10^{-3}$ $s^{-1}$ (b) (Schulson & Duval, 2009).

# 3 Ship-Ice Interaction and Ice Loads

Icebreaking vessels are experiencing different types of ice loads, caused by different scenarios. The scenarios can be coarsely divided into three types of ice-ship interactions:

- **Breaking of level ice**. Main pressure loads act in the bow region, and on the outer side aft of the ship when the ship is turning.
- **Collisions with ice features**, such as glacial ice, ice ridges or multi-year ice. These types of ice-ship interactions are characterized by large and concentrated forces, thus denting and minor deformations are expected. Heavy collisions occasionally result in larger deformations and sometimes also critical damages, as seen in Figure 9.
- **Ships in compressive ice**. With ice being pushed into the ship sides, large frictional forces develop, and the ship may get stuck. Compressive ice may also cause damage to the ship sides.

Figure 9: Damage from collision with ice features (Riska, 2017d).

The ice loads on ships are often characterized by high load pressures and non-uniform pressure fields. High velocities govern ice-ship impacts, resulting in high strain rates. High confinement pressures are, in addition to high strain rates, important for the occurrence of high-pressure zones, with local pressures of up to 70-100 MPa (Soares & Garbatov, 2015). As ice is compressed against the hull, pressure builds up. Cracking, spalling and crushing take place, resulting in a constantly changing pressure distribution, both in time and space. For relatively low indention rates, and thus low strain rates, ice will spall off with a regular time interval, giving an oscillating force (O'Rourke et al., 2015). This is not considered as an issue for ships, since they are subjected to higher strain rates, resulting in a more irregular and frequent crushing. Figure 10 shows the response for different indention speeds, increasing from (a) to (d).

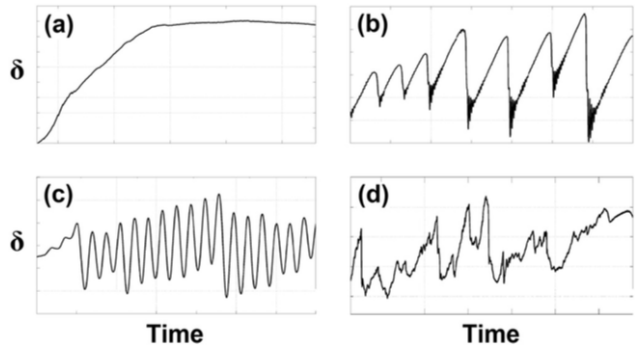
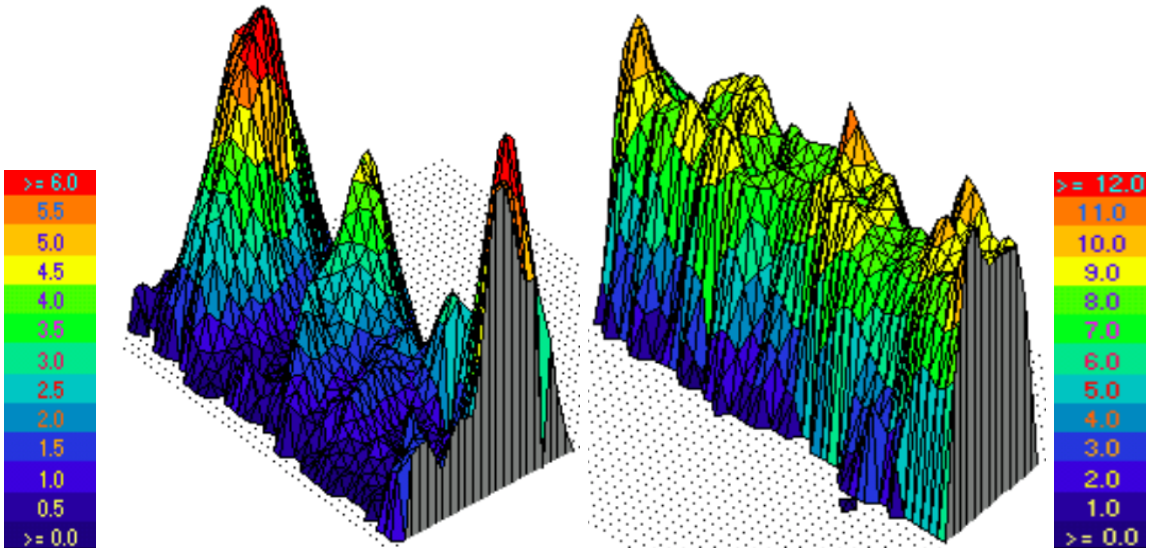



Figure 10: Deformation of structure plotted against time for different indention rates, increasing from (a) to (d) (O'Rourke et al., 2015).

## 3.1 Pressure Distributions

Both small and full-scale tests have shown that the pressure often distributes like a line (although the load varies along the line). Figure 11 shows the spatial pressure distribution during sea ice impact for different indentation rates. One can clearly see the effect of strain rates. High indentation rates, and thus high strain rates, cause brittle failure with a large spatial variation of the pressure. At low strain rates, the ice behaves in a more ductile manner, resulting in a more evenly distributed pressure along the line. Thus, pressure distributions associated with bow impacts and those associated with compressive ice may not behave equally, as the indention rates are different. Higher indention rates often cause higher local pressures, but the total force tends to be smaller than for lower indention rates.




(a) Indentation rate = $502\frac{mm}{s}$. (b) Indentation rate = $0.33\frac{mm}{s}$.

Figure 11: Spatial pressure [MPa] distributions for sea ice impact with different indentation rates (Sodhi, n.d.).

Measurements performed on ice-going vessels, as well as experiments, show a clear trend of decreasing average pressure with increasing contact area. Figure 12 shows a collection of data for contact area and pressure from ship measurements and experiments. Although the scatter is significant, there is a clear trend in the data for contact areas up to about 10 $m^2$. Average contact pressures are commonly expressed as a function of the contact area on the form:

$$P = C_1 \cdot A^{C_2} \tag{2}$$

where $P$ is the average pressure, $C_1$ and $C_2$ are fitted constants, and $A$ is the contact area. For the data sets in Figure 12, $C_1$ and $C_2$ are found to be 8.1 and -0.572 respectively. Equations on the same form are used when the load is considered as a line load instead of a pressure, using the length of the contact line instead of the area and load line intensity instead of pressure. This concept, which was introduced by Riska et al. (I. J. Jordaan, 2001), is suitable for describing loads induced by interaction between level ice and ship hulls. High-pressure zones are found to occur more frequently near the center of the cross-section of ice sheets. For different specimen geometry, high-pressure zones tend to be distributed in certain patterns. Some concepts are seen in Figure 13.

The reason for the pressure-area relationship is widely discussed in the literature. Sanderson (1986) uses the Weibull-effect to explain the relation; the ice is not stronger than the weakest point within the contact area. A large contact area has a higher probability of having a weak point, resulting in a lower expected average pressure.

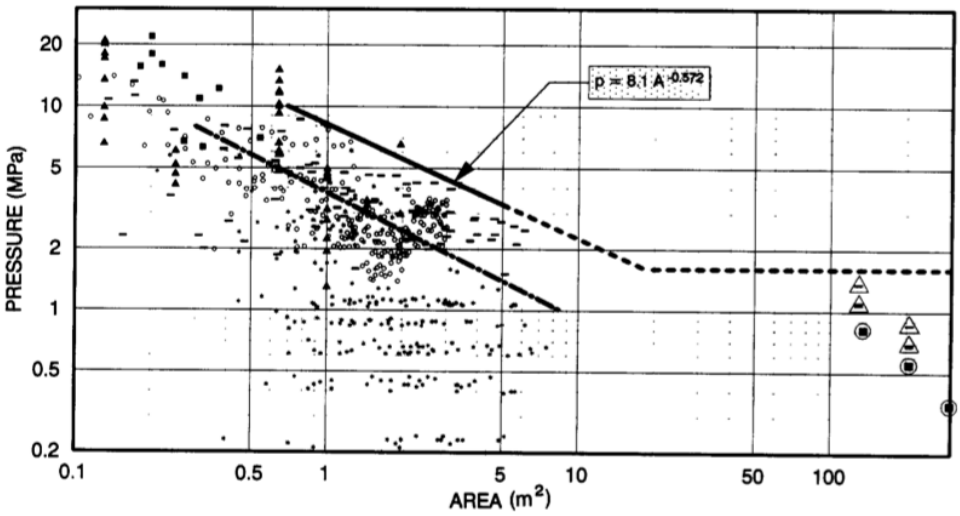



Figure 12: Average pressure plotted against contact area for seven different data sets. $- \cdot -$ is fitted to the mean of all data, solid line is 2STD. (Masterson & Frederking, 1993).

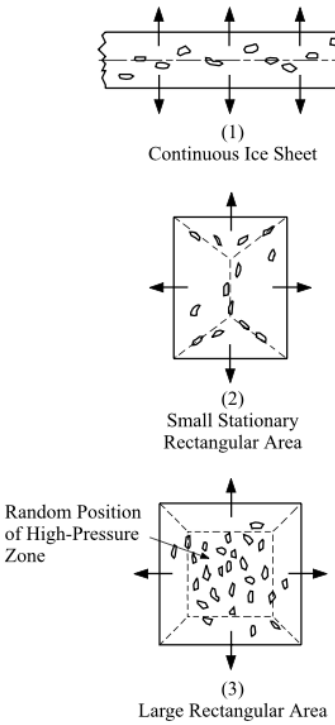



Figure 13: The figures show high-pressure zones for various ice geometries. Dotted lines indicate where high-pressure zones tend to be concentrated. (I. J. Jordaan, 2001).

## 3.2 Breaking of Level Ice

Icebreakers are designed to break the ice by bending rather than crushing, as it requires less energy. Thus, small buttock angles and large frame angles are beneficial. The breaking of level ice is associated with different forces, and the forces on the hull can be divided into normal forces (due to pressure) and frictional forces. An illustration is seen in Figure 14. The forces can be decomposed to a vertical and a horizontal force. The vertical force induces a bending moment in the ice. Crushing and spalling take place in the initial phase of the ice-hull interaction, causing the contact area - and the forces - to increase. The forces will increase until the flexural strength, $\sigma_f$, of the ice sheet is reached, and an ice floe breaks off. When the ice breaks due to bending, it fails in tension. The tensile strength of ice is in the range 0.5-2 MPa, which is only a fraction of the compressive strength (Riska, 2017b).

Ships with vertical (or close to vertical) sides in the bow region will induce small vertical forces, and the ice will fail by crushing instead of bending. Thus, the geometry of the bow is essential for the resistance.

The most important parameters for the bending capacity of the ice are the ice thickness $h_i$ and the flexural strength. A model for the critical vertical force is presented in Equation 3 (Riska, 2017e):

$$F_{V,crit} = C \cdot \sigma_f \cdot h_i^2 \tag{3}$$

where $C$ is a constant, depending on the geometry of the boundary of the ice. By applying Newton's third law, it can be shown that the ice breaking resistance also is proportional to $\sigma_f \cdot h_i^2$.

Figure 14 illustrates the breaking as a 2D problem, in the xz-plane. Seen from above, the cracks are propagating as a curve. This is illustrated in Figure 15a. In Figure 15b one can see both crushing and the breaking pattern. It is important that the channel created by bending failure is wider than the beam of the ship in order to avoid so-called *shoulder crushing*. Shoulder crushing occurs when the unbroken ice sheet is compressed against the shoulder of the ship - the foremost part of the ship with full width. At this part of the ship, the sides are usually vertical, thus no bending is induced. The ice must fail by crushing, which is very energy demanding. The phenomenon is illustrated in Figure 16. The increase in resistance due to shoulder crushing is so large that the design will be considered as unsuccessful if it occurs (Riska, 2017e).

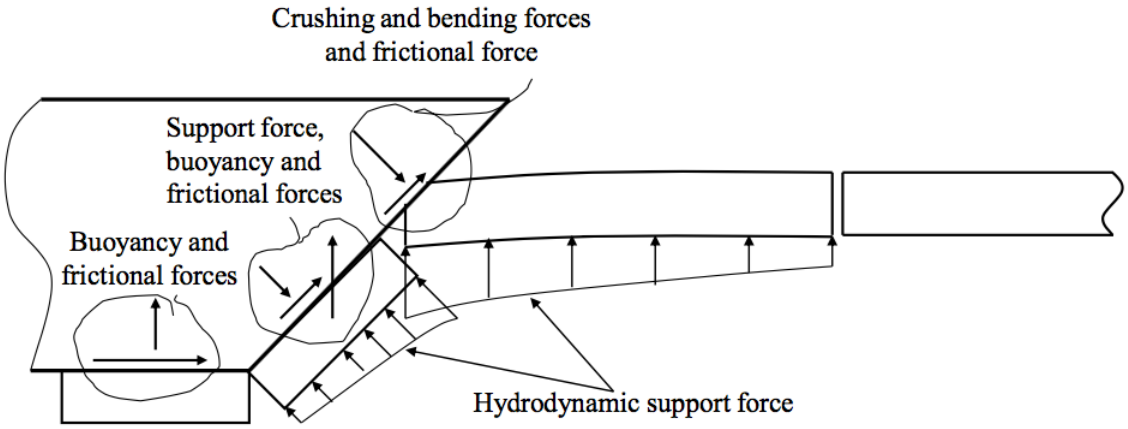
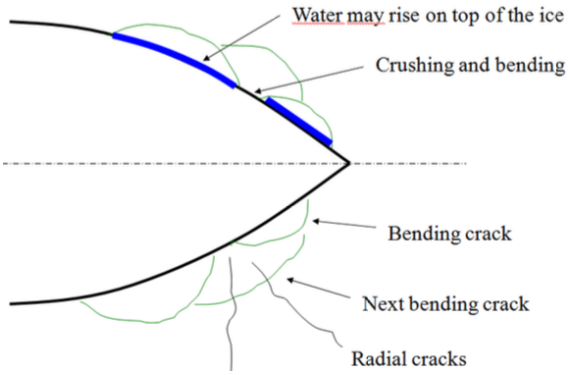



Figure 14: Illustration of forces related to ice-ship interaction (Riska, 2017d).




(a) Illustration of breaking pattern.

(b) Ice crushing and breaking pattern. The picture is from an icebreaker in the Gulf of Bothnia.

Figure 15: Both Figures: (Riska, 2017c).

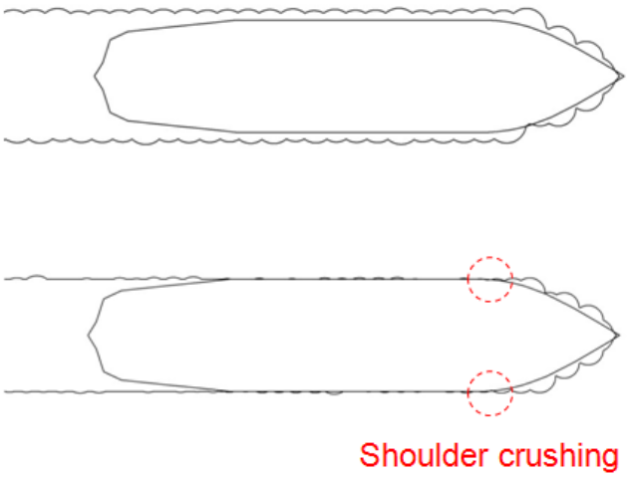



Figure 16: Illustration of a ship without and with shoulder crushing (Riska, 2017c). Note that the created channel have almost perfectly straight lines when shoulder crushing is present.

# 4 Measurements from KV Svalbard

In 2006 DNV (now DNV GL) launched an ice load monitoring program (Chai et al., 2018). KV Svalbard was equipped with sensors and measurement instruments for collecting various data, e.g., speed over ground, direction, position, propulsion power, ice conditions, ice loading and accelerations. Since 2006, several research expeditions have been performed where data have been collected. Measured data from an expedition carried out during March 2007 in the vicinity of Svalbard are used in this thesis. Information and technical data about KV Svalbard are found in Appendix A.

Fiber optic strain sensors were installed on local hull structure. A total of nine locations were investigated, which of eight of them were in the bow region, and one was located amidships on the starboard side. Only the sensors in the bow region are of interest in this thesis, see Figure 17 for arrangement. Figure 18 shows how the strain sensors were mounted. The measured strains are transformed into stresses. The total force acting between two locations can be found as the difference between the shear stress at the two locations. Therefore, a pre-assumed shear stress distribution is integrated over the cross-section in order to get the total load. For details about how the strains measured on KV Svalbard are converted into loads, see Leira and Børsheim (2008). As discussed in Section 3.1, is it convenient to express ice loads as line loads, which is obtained by dividing the total load on the frame spacing.

The data used in this work is processed by Suyuthi et al. (2013a). He identified the ice load peaks and other relevant data from the original, continuous time series. A threshold of 31.25 kN/m (corresponding to 25 kN acting on one frame) was applied in order to identify the ice load peaks, and yet void other noise (Espeland, 2008). Compared to other works, this is a relatively high threshold. As an example, Kujala et al. (2009) used 10 kN/m. However, for extreme value estimations, the lower loads are not important. If fatigue is considered, the results will be more sensitive to the selection of the lower threshold, since most of the fatigue damage is due to the low, but many, load cycles. For further reading about how the data is processed, see Suyuthi et al. (2013a).

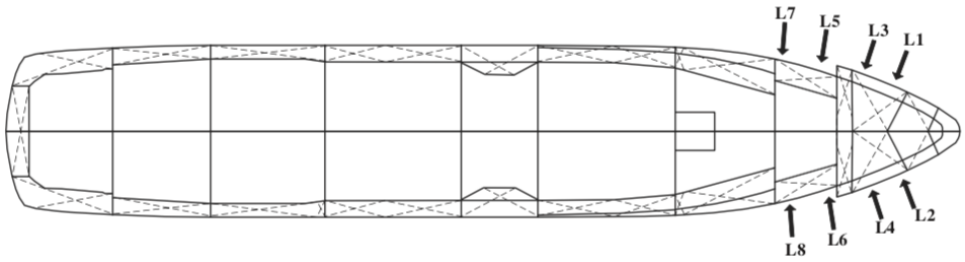



Figure 17: Locations of strain sensors (Suyuthi et al., 2014).

Figure 18: Sensors mounted on hull structure (Leira & Børsheim, 2008).

# 5 Ice Load Peak Statistics

The ice-breaking process, which is described in Section 3.2, is very complex, and the associated pressures and loads vary significantly, even for tests constructed to be carried out under identical conditions. As discusses in Section 2.3 and Section 3, the ice properties come along with great variations, which are affecting the loading to a large extent. Even small and unavoidable variations in ice properties (e.g., porosity, density, temperature, macrostructure, etc.) are sources to great variations in the ice-induced loads. The ice conditions that a vessel encounters in the Arctic Oceans are constantly varying. Winds, currents and weather change the composition of the ice, which makes it non-stationary. Broken ice-fields are common, which can consist of level ice, ice floes, ridges, re-frozen ice, open water and sometimes also multi-year ice. This was experienced when data sets for analyses were selected. From 6 days of constant measurements, it was difficult to find stable conditions, even for periods of just a few minutes. In addition, the ship-ice interaction itself is a source to randomness in the ice-breaking process and the associated loads. The crushing, bending and breaking pattern of the ice are so complex that they for all practical applications can be considered as random by nature (Chai et al., 2018).

With such large variations and degree of randomness, probabilistic models for describing the loads are important, both for tactical navigation purposes (short-term statistics) and for design purposes (long-term statistics). Full-scale trials, such as DNV's ice load monitoring project, provide valuable data for statistical analyses. Other full-scale measurements have been performed, and the collected data have been used for statistical analyses, e.g., Kujala and Vuorio (1986), Lensu and Manninen (2003), Suominen and Kujala (2010). Kujala and Vuorio (1986) used measurements from 1979 to 1985 from the icebreaker Sisu operating in the Baltic waters. They applied a classical approach, and found that the exponential distribution gave the best fit to their data. The classical approach is described in Section 5.3. Suominen and Kujala (2010) also used a classical approach, and found that a Weibull distribution with shape parameter 0.75 was the most suited distribution for the measurements performed on MS Kemira during the winters of 1987 and 1988. The measurements from KV Svalbard have been subject for previous works related to ice load statistics, e.g., Suyuthi et al. (2013a) and Chai et al. (2018). Suyuthi et al. (2013a) applied both a classical method and a non-parametric approach using kernel functions (described in Section 5.6.1). He also came up with a novel approach using a three-parameter exponential distribution, which is an unequally weighted sum of two different exponential distributions. The three-parameter exponential distribution, which is described in Section 5.5, was found to give a better fit than both the one-parameter exponential and the Weibull distribution. Chai et al. (2018) used a quite novel approach known as the average conditional exceedance rate (ACER) method for estimating extreme loads. The ACER method has several advantages compared to the classical methods; it is applicable to non-stationary as well as stationary time series, it accounts for dependencies in the data set, and it puts more weight on the more reliable measurements. The method was also found to perform better than the asymptotic method. The ACER method is described in Section 5.6.2.

## 5.1 Stationarity

In its strictest form, stationarity is achieved if the moments of all orders (i.e., the mean, variance, skewness, etc.) are constant, independent of time. Many of the methods applied in this thesis, e.g.,

the classical approach and the asymptotic approach, are based upon the assumption of stationarity in the data set being analyzed. If the parameters influencing the ice loads behave stationary, one can expect that the ice loads in the same data set will be stationary as well. Suyuthi et al. (2013b) were mainly concerned about stationarity regarding propulsion power when data sets were selected. Suyuthi et al. (2012b) selected time series that showed stationarity for measured propulsion power and vessel speed, and ice thickness was not even mentioned when the selection of time series was discussed. Nevertheless, the combination of stationary propulsion power and vessel speed is most likely achieved in stationary ice conditions. Chai et al. (2018) mentioned stationarity in terms of the measured ice load peaks. However, the variance of the measured peaks is very large, and it is difficult to say whether they are stationary or not. Due to the discrepancies discussed above related to requirements for selecting time series, a study on the effect of ice thickness and vessel speed on the ice loading is carried out in Section 7.1.

For broken ice fields, stationarity as defined above is impossible to have. A less strict requirement can be applied, known as "weak stationarity". For a weak form of stationarity, the $n$ first moments are required to be independent of time, where $n$ is typically a low number, e.g. 1 or 2. In Figure 19, vessel speed and ice thickness are plotted on the left y-axis, and the propulsion power is plotted on the right y-axis for a data set used later on (set 6). In Figure 20, the same data set is plotted together with an averaged ice thickness (blue line) for the same data set. We can see that for $\Delta T = 5$, the averaged ice thickness can be considered as constant for practical applications.

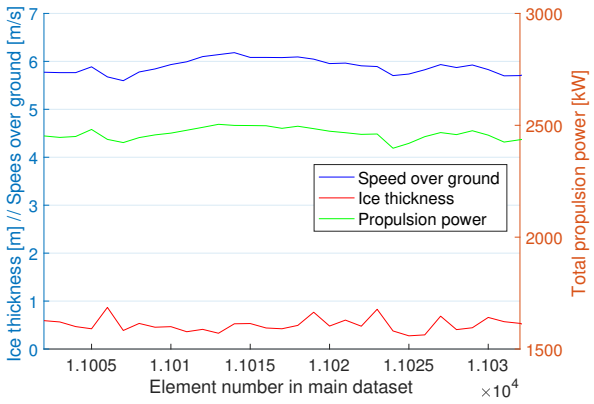



Figure 19: Ice thickness [m], speed over ground [m/s] and total engine power [kW]. x-axis: element nr in main data set.

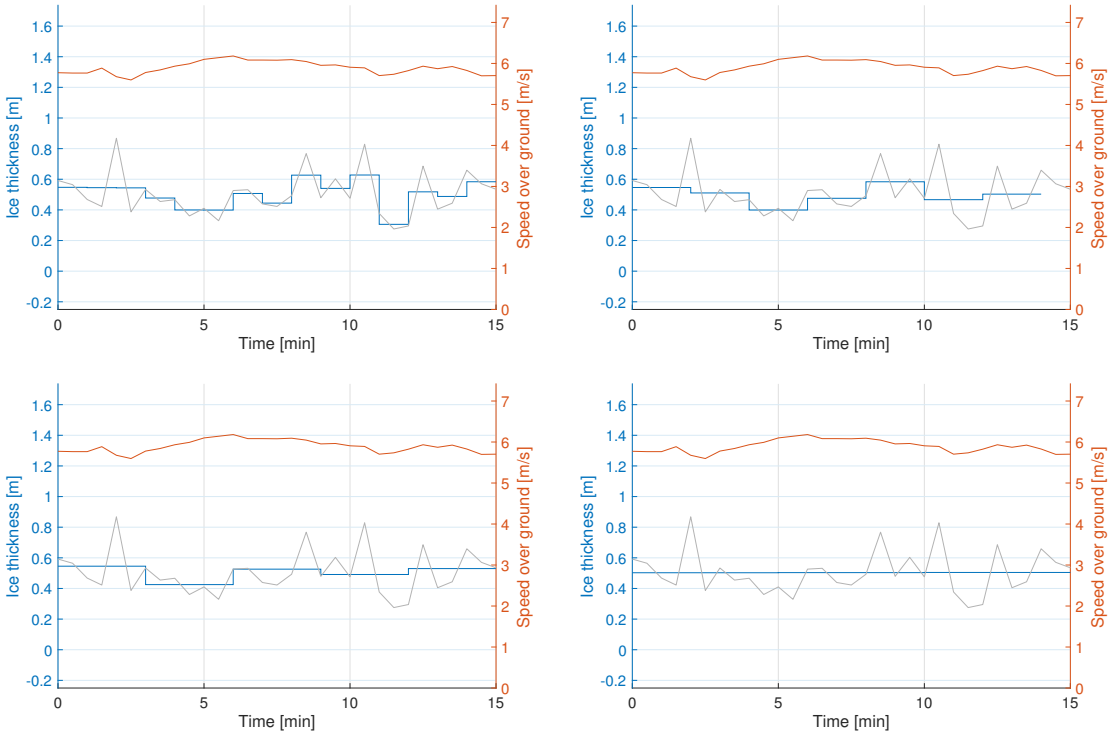



Figure 20: Averaged ice thickness over $\Delta t = 1$ min, 2 min, 3 min and 5 min, respectively.

## 5.2 Selected Data Sets

Possible sets for analyses are extracted from the measurements, and presented in Table 1 below. It was strived for finding sets with a low as possible relative standard deviation (RSD) for the ice thickness and speed over ground, prioritized in that order.

Table 1: Possible sets for statistics.

| Set | Elements in the matrix "data" | Duration | Ice thickness | | | Speed over ground | | |
|---|---|---|---|---|---|---|---|---|
| | | [min] | $\mu$ [m] | $\sigma$ [m] | RSD [%] | $\mu$ [m/s] | $\sigma$ [m/s] | RSD [%] |
| 1 | 7680-7743 | 31.5 | 0.42 | 0.12 | 28.5 | 6.08 | 0.19 | 3.2 |
| 2 | 7770-7830 | 30 | 1.13 | 0.36 | 31.5 | 4.92 | 0.81 | 16.4 |
| 3 | 8167-8203 | 18 | 0.37 | 0.10 | 26.3 | 7.06 | 0.12 | 1.8 |
| 4 | 8311-8340 | 14.5 | 0.65 | 0.16 | 24.2 | 3.24 | 0.46 | 14.3 |
| 5 | 10942-11002 | 30 | 0.82 | 0.24 | 29.0 | 5.06 | 0.54 | 10.7 |
| 6 | 11002-11032 | 15 | 0.51 | 0.14 | 28.4 | 5.90 | 0.16 | 2.6 |
| 7 | 11056-11086 | 15 | 0.99 | 0.20 | 20.7 | 4.87 | 0.54 | 11.1 |
| 8 | 11090-11130 | 20 | 0.98 | 0.28 | 28.5 | 5.33 | 0.49 | 9.1 |
| 9 | 15482-15542 | 30 | 0.51 | 0.15 | 28.5 | 2.68 | 0.61 | 22.7 |
| 10 | 15618-15678 | 30 | 0.72 | 0.23 | 31.6 | 5.58 | 0.47 | 8.7 |
| 12 | 15720-15762 | 21 | 0.46 | 0.09 | 20.3 | 6.93 | 0.09 | 1.4 |
| 13 | 11058-11123 | 32.5 | 1.00 | 0.22 | 21.9 | 5.07 | 0.55 | 10.8 |
| 14 | 11058-11108 | 25 | 1.04 | 0.20 | 19.2 | 4.87 | 0.41 | 8.4 |

## 5.3 Classical approach

Suyuthi et al. (2013b) defines the term "classical approach" as *"any statistical approach which requires the random process to be presented by a distribution model"*. The classical approach is a well established method. Several others have used this approach in an attempt to describe ice-induced loads on ship hulls, e.g., Kujala and Vuorio (1986), Suominen and Kujala (2010), Suyuthi et al. (2013b). Since the method have been applied several times before and is well known, it is useful for comparison purposes and as a reference when other methods are evaluated. Thus, the classical approach is used as a benchmark. The probability density function (PDF) and the cumulative distribution function (CDF) for the applied probability distributions are given in Table 2 and Table 2, respectively.

Table 2: Probability density functions.

| Probability distribution | Probability density function |
|---|---|
| Exponential | $f(y) = \lambda \exp(-\lambda y)$ |
| Log-normal | $f(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right)$ |
| Gumbel | $f(y) = \frac{1}{\beta} \exp\left(-\frac{y-\alpha}{\beta}\right) \exp\left(-\exp\left(-\frac{y-\alpha}{\beta}\right)\right)$ |
| Weibull | $f(y) = \frac{k}{\theta}\left(\frac{y}{\theta}\right)^{k-1} \exp\left(-\left(\frac{y}{\theta}\right)^{k}\right)$ |

Table 3: Cumulative distribution functions.

| Probability distribution | Cumulative distribution function |
|---|---|
| Exponential | $F(y) = 1 - \exp(-\lambda y)$ |
| Log-normal | $F(y) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right)$ |
| Gumbel | $F(y) = \exp\left(-\exp\left(-\frac{y-\alpha}{\beta}\right)\right)$ |
| Weibull | $F(y) = 1 - \exp\left(-\left(\frac{y}{\theta}\right)^k\right)$ |

### 5.3.1 Extreme Value Prediction

The main objective for this thesis is to perform a comparative study of different methods for estimating short-term extreme loads. In order to be able to compare the estimated extreme values with the measured ones, the most probable largest value during the evaluated time series, $\hat{y}_{max}$, is defined as the extreme value. This applies for all the evaluated methods in this thesis. $\hat{y}_{max}$ is the value for which the probability density function for the extreme value reaches its peak. This corresponds to the load that has the cumulative probability $F(y) = 1 - \frac{1}{N}$, where $N$ is the number of measured loads during the analyzed time series, and $F(y)$ is the CDF for the fitted distribution. Thus, $\hat{y}_{max}$ can be found by rearrange the expression for $F$, and applying $F(y) = 1 - \frac{1}{N}$. $y$ is defined as the measured load subtracted the applied threshold: $y = x - 25$. The extreme value for the probability distributions presented in Table 3 is given in Table 4 below:

Table 4: Most probable largest values.

| Probability distribution | Most probable largest value |
|---|---|
| Exponential | $\hat{y}_{max} = -\frac{\ln(1/N)}{\lambda}$ |
| Log-normal | $\hat{y}_{max} = \exp\left(\mu + \sigma \cdot \Phi^{-1}\left(1 - \frac{1}{N}\right)\right)$ |
| Gumbel | $\hat{y}_{max} = \alpha - \beta \cdot \ln\left(-\ln\left(1 - \frac{1}{N}\right)\right)$ |
| Weibull | $\hat{y}_{max} = \theta\left(-\ln\left(\frac{1}{N}\right)\right)^{\frac{1}{k}}$ |

### 5.3.2 Graphical Methods for Finding Underlying Distribution

There exist several methods for identifying the underlying distribution of a random process. Quantile-Quantile plot (Q-Q plot) and percentile-percentile plot (P-P plot) can be used to determine the match between the sample data and a known probability distribution. The P-P plot is better to detect discrepancies in the middle of the distribution than in the tails. Contrary, the Q-Q plot is

more sensitive to discrepancies in the tails, and is therefore more suitable for applications related to extreme value statistics. Consequently, P-P plots are not applied in this thesis.

In a Q-Q plot, equal quantiles from two different data sets are plotted against each other. The two data sets can be two sets of sampled data, or one set of sampled data and a set of quantiles taken from the cumulative distribution function of a known distribution. If the two data sets are taken from the same distribution, the plotted points will lie on a straight line with slope 1, starting at the origin. For a finite sample collected from a real-life experiment, there will always be deviations from the line. This will be the case even if the data are taken directly from the distribution. How well the scatter fit the line indicates how similar the underlying distributions of the data sets are. By comparing Q-Q plots using different distributions, one can find out which of the evaluated distributions that in the best manner describes the data. A disadvantage of this method is that the distribution parameters must be determined in advance. Figure 21 show Q-Q plots for set 14, load sensor L1.

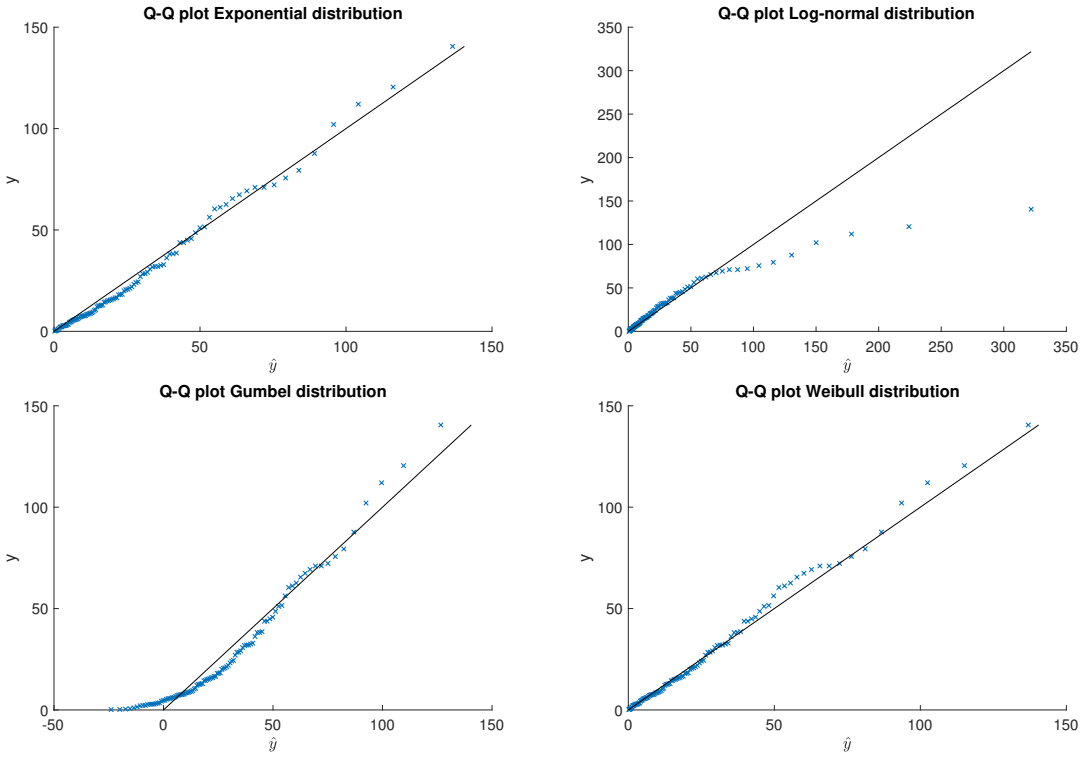



Figure 21: Q-Q plots for set 14 L1.

Another commonly applied graphical method is to plot the data points in a probability paper. Here, the cumulative distribution function $F(y)$ is rearranged such that a function of the cumulative probability $F$ is proportional to a function of the variable $y$; $g(F) \propto h(y)$. In the probability paper, $g(p_i)$ is plotted against $h(y_i)$. $i$ is the index number and $p_i$ is the empirical cumulative probability, given by Equation 4.

$$p_i = \frac{i}{N+1}. \tag{4}$$

Here, $N$ is the number of data points in the data set. Abscissa and ordinate transformations for the applied probability distributions are given in Table 5. After plotting the data in a probability paper, a straight line is fitted to the data. If the data set can be well described by the underlying distribution that the probability paper is based upon (and the number of data points is sufficiently large), all the plotted data should appear close to and randomly scattered around the line. If the data does not fit the line, the data cannot be well described by the distribution evaluated. An advantage of this method is that the distribution parameters do not have to be known in advance. Examples of probability papers are seen in Figure 22.

Table 5: Abscissa and ordinate transformations for probability papers.

| Probability distribution | Abscissa | Ordinate |
|---|---|---|
| Exponential | $y$ | $-\ln(1-F)$ |
| Log-normal | $\ln(y)$ | $\Phi^{-1}(F)$ |
| Gumbel | $y$ | $-\ln(-\ln(F))$ |
| Weibull | $\ln(y)$ | $\ln(-\ln(1-F))$ |

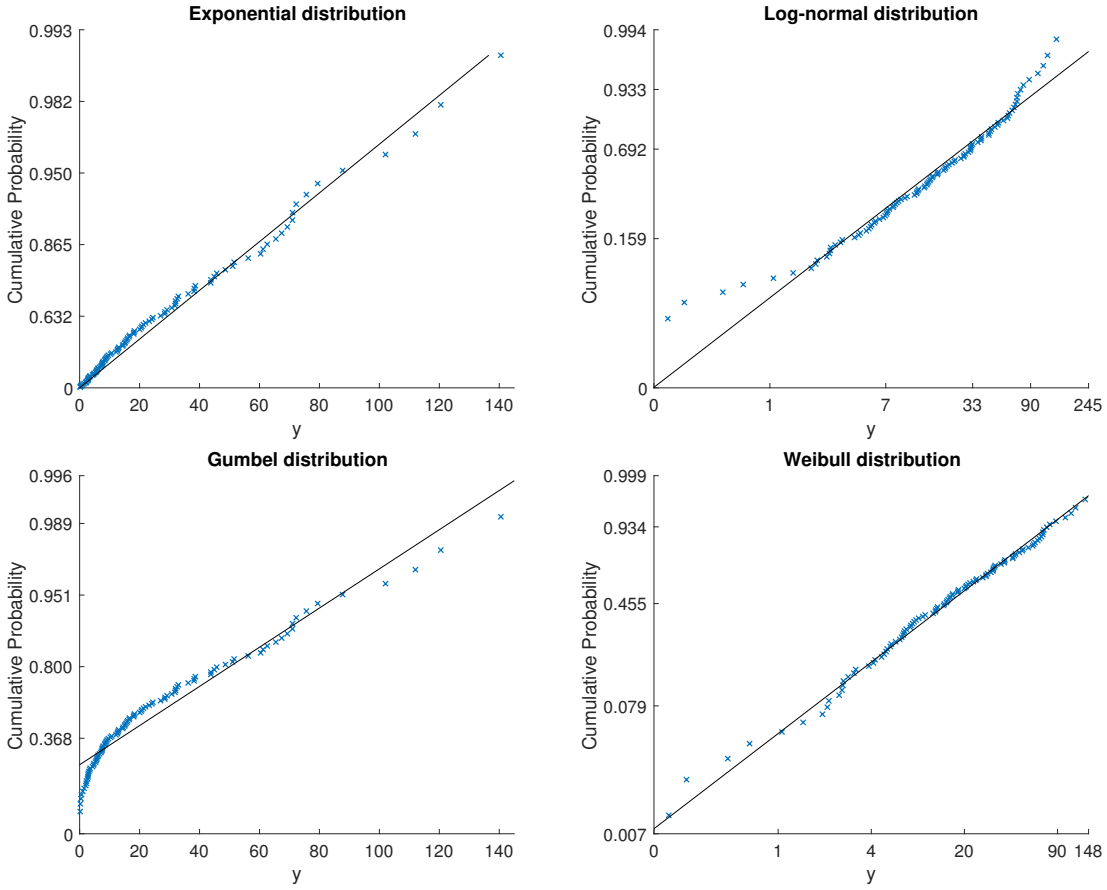



Figure 22: Probability paper for set 14 L1.

### 5.3.3 Distribution Parameters Estimation

When a data set is plotted in a probability paper, a line is fitted using the least square method. The probability distribution parameters can be expressed by the slope $m$, and the interception point $c$ with the ordinate. The expressions for the parameters calculated based on the probability paper are given Table 6.

Table 6: Distribution parameters estimated from probability paper.

| Distribution | Parameters | |
|---|---|---|
| Gumbel | $\hat{\beta} = \frac{1}{m}$ | $\hat{\alpha} = -\hat{\beta} \cdot c$ |
| Log-normal | $\hat{\sigma} = \frac{1}{m}$ | $\hat{\mu} = -\hat{\sigma} \cdot c$ |
| Weibull | $\hat{k} = m$ | $\hat{\theta} = exp\left(\frac{c}{\hat{k}}\right)$ |
| Exponential | $\hat{\lambda} = m$ | - |

Besides, there are two non-graphical methods for estimating the distribution parameters, i.e., the method of moments (MoM), and the maximum likelihood estimate (MLE). When applying MoM, the population moments $\mu_i$ are derived, where $\mu_i$ is the $i$'th moment, which will be functions of the distribution parameters $\theta_1$, $\theta_2$ ... $\theta_n$, where $n$ is the number of parameters. Estimates for the parameters are found by solving the equations $\mu_i = f(\theta_1, \theta_2 \ ... \ \theta_n)$ for $\boldsymbol{\theta}$. For many distributions, the solutions for $\boldsymbol{\theta}$ is easily found and applied.

The maximum likelihood estimate method finds the parameters that are most probable given the selected data, using a so-called likelihood function. For further reading about MLE and MoM, reference is made to Walpole et al. (2012). In this thesis, parameters are estimated based on probability paper and by MLE.

Table 7: Distribution parameters calculated using maximum likelihood estimation.

| Distribution | Parameters | |
|---|---|---|
| Gumbel | $\hat{\beta} = \bar{y} - \frac{\sum_{i=1}^{n} y_i \exp\left(-\frac{y_i}{\hat{\beta}}\right)}{\sum_{i=1}^{n} \exp\left(-\frac{y_i}{\hat{\beta}}\right)}$ | $\hat{\alpha} = -\hat{\beta} \cdot \ln\left[\frac{1}{n}\sum_{i=1}^{n} \exp\left(\frac{-y_i}{\hat{\beta}}\right)\right]$ |
| Log-normal | $\hat{\mu} = \frac{\sum_{i=1}^{n} \ln y_i}{n}$ | $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (\ln y_i - \hat{\mu})^2}{n}$ |
| Weibull | $\hat{\theta} = \left(\frac{1}{n}\sum_{i=1}^{n} y_i^{\hat{k}}\right)^{1/\hat{k}}$ | $\hat{k} = \frac{n}{\frac{1}{\hat{\theta}}\sum_{i=1}^{n} y_i^{\hat{k}} \ln y_i - \sum_{i=1}^{n} \ln y_i}$ |
| Exponential | $\hat{\lambda} = \frac{1}{\bar{y}}$ | - |

### 5.3.4 Confidence Intervals

When data from a time series is fitted to a specific distribution, the distribution parameters can be calculated as described in the previous section. When a finite sample is drawn from a distribution, we know that the underlying distribution is not perfectly recreated. Due to randomness, each sample will have different sample distribution parameters, known as sampling error. Each of the measured data points provide some information about the underlying distribution, thus the bigger

the sample is, the more information we have. The estimated distribution parameters are calculated for the measured data, but we are actually interested in saying something about the underlying distribution. Thus, we want to know something about how accurate the estimates are, i.e., a confidence interval (CI). Bootstrapping using Monte Carlo simulation is used for this purpose (C. Davison & Hinkley, 1997).

The Monte Carlo simulations are performed by first generating a set of random numbers in the range from 0 to 1. Next, the expression for the cumulative probability function for the evaluated distribution is reworked to get an expression for the stochastic variable as a function of the cumulative probability. Lastly, the new data set, also known as the bootstrapped data set, is estimated by substituting the cumulative probability with the randomly generated numbers, and by using the estimated distribution parameters. The generated sets should be of the same size as the original data set (Haver, 2017), and can be treated as an individual data set from the same distribution. The procedure is repeated to get more data sets, typically tens of thousands of times. New, bootstrapped distribution parameters are calculated based on the new data sets. A confidence interval for the estimated parameters for the measured data set can be found based on the bootstrapped ones. The $1 - \alpha$ confidence interval is found by excluding the $\alpha/2$ highest and lowest of the bootstrapped parameters. The results can be plotted in a stem plot, which is illustrated by an example in Figure 23. The bootstrapped parameters are divided into intervals of equal lengths, represented by the stem plot. The black dashed line is the estimated parameter based on the measured data, and the red dashed lines are the bounds of the 95 % confidence interval.

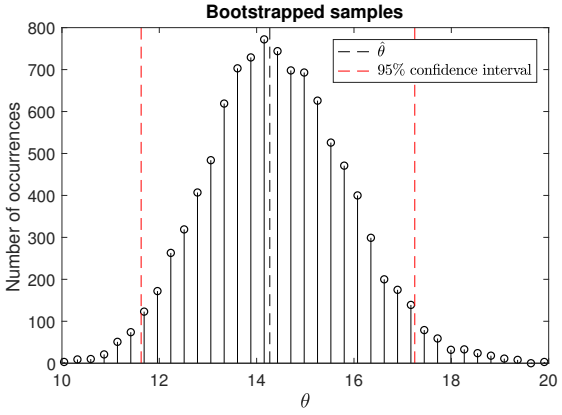



Figure 23: Confidence interval calculated for $\hat{\theta}$ (Weibull distribution) for set 9 L2. 10 000 bootstrapped samples of size N=82.

### 5.3.5 Goodness of Fit

The Kolmogorov-Smirnov test is a goodness of fit test where the largest deviation D between the empirical cumulative distribution function and the fitted CDF for the same load value $y$ is identified. The test is formulated as:

$$D = \max |F_{empirical}(y_i) - F_{fitted}(y_i)| \tag{5}$$

where $F_{empirical}(y_i) = \frac{i}{N+1}$. Here $i$ is the number of events less than or equal to $y$, and $N$ is the number of events. If the test value D is less than a critical value $D_{crit}$, which is dependent on the sample size and the confidence level $(1-\alpha)$, the distribution is not rejected by the test. A table for $D_{crit}$ is given in Appendix B. The method is quick and easy to implement and gives a good visual impression of the fit. The Kolmogorov-Smirnov test is illustrated in Figure 24.

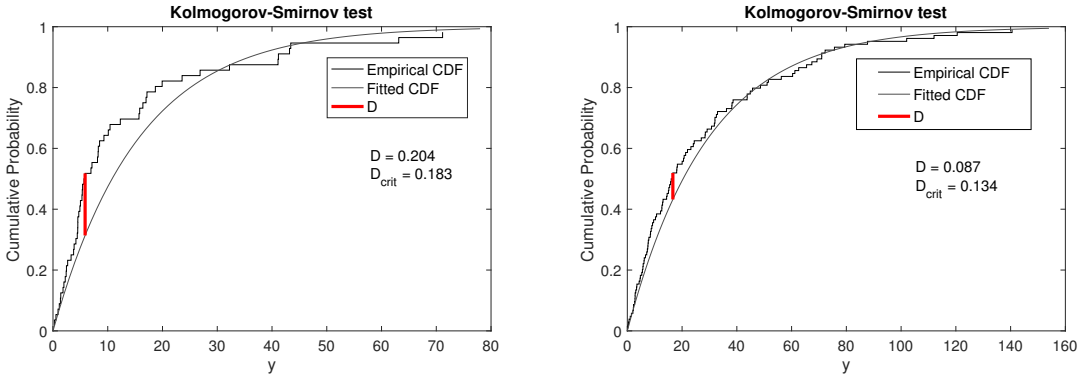



Figure 24: Kolmogorov-Smirnov tests for exponential distribution for set 6 L5 and set 14 L1, respectively. For set 6 L5 the exponential distribution is rejected by the KS-test. $\alpha = 0.05$ for both figures.

Another graphical goodness of fit test is to plot the bootstrapped data points together with the measured data points from the time series, either in linear scale or in a probability paper. All the measured data points should lie inside the range of the bootstrapped ones. If this is not the case, the pre-assumed distribution is unlikely to be the true underlying distribution (Haver, 2017). This test is suitable for checking whether a measured outlier can be explained by the evaluated distribution.

## 5.4 Asymptotic Method

The asymptotic approach is widely used for obtaining extreme values of natural phenomena. The time series can be divided into sub-intervals of constant duration $\Delta t$, where the maximum value for each time interval is identified. The duration of the constant time intervals may vary depending on the nature of the phenomenon investigated, the available data, and the time horizon of the extremes. However, more data points give better estimates of the distribution parameters. Haver (2017) proposes that at least 20-30 points are needed in order to capture the tails and to get acceptable estimates of the distribution parameters.

Gumbel (1958) showed that the maximum value of a random process following an exponential distribution converges to the Gumbel distribution as the sample size increases. As seen in Section 7.3, the exponential distribution is found to provide a good fit for several data sets, thus the maximums within the sub-intervals are expected to be modelled well by the Gumbel distribution. The cumulative Gumbel distribution function is given by the following expression:

$$F_y = exp\left(-exp\left(-\frac{y-\alpha}{\beta}\right)\right) \tag{6}$$

where $\alpha$ and $\beta$ are distribution parameters. The load $y$ corresponding to an exceedance probability of $\lambda$ is found by applying $F = 1 - \lambda$, and rearranging Equation 6:

$$y = \alpha - \beta \ln(-\ln(1-\lambda)) \tag{7}$$

An example is shown in Figure 25, where 1 minute maximums for set 14 L1 are plotted in a Gumbel probability paper. Although 30 data points are relatively few, they seem to fit the straight line quite well. The Kolmogorov-Smirnov test gives $D = 0.106 < D_{crit} = 0.274$, which implies that the asymptotic approach using Gumbel distribution cannot be rejected by this goodness-of-fit test for the evaluated data set. The Kolmogorov-Smirnov test is illustrated in Figure 26.

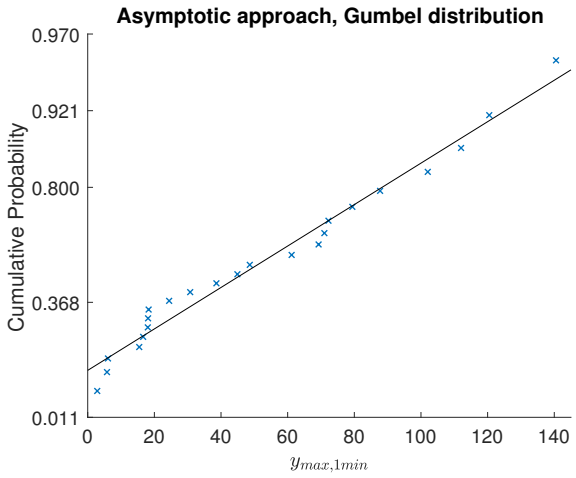



Figure 25: 1 minute maximums plotted in Gumbel probability paper for set 14 L1.

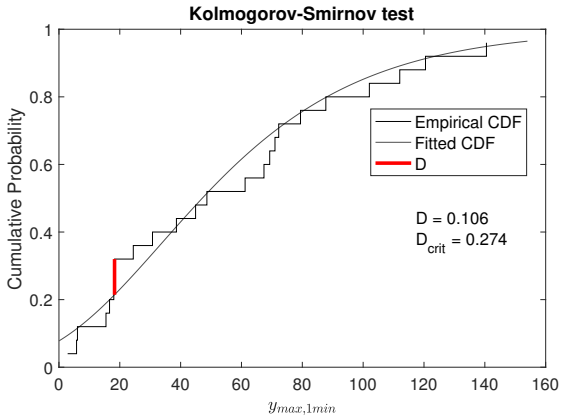



Figure 26: Kolmogorov-Smirnov test performed on the fitted distribution and the data points found in Figure 25. $D = 0.106 < D_{critical} = 0.274 \implies$ not rejected.

## 5.5 Three-Parameter Exponential Distribution

Suyuthi et al. (2014) observed that ice-induced loads on ship hulls often tend to look like a combination of two different populations when they are plotted in an exponential probability paper, with a less steep trend for the upper tail. For such cases, by applying a one-parameter exponential distribution, the overall fit is poor, and extreme values will be underestimated. A possible explanation for the existence of more than one population is the varying ice thickness and presence of ice features. Most of the loads are found in the lower range, which may be a result of, e.g., breaking of level ice. These loads represent *one* population. The loads in the upper range may be a result of, e.g. interaction with ice-ridges and hummocks, representing *another* population. An additional population may occur due to previously open waters that have been refrozen, and thus have a different thickness than the rest of the level ice. It is likely to believe that such varying ice characteristics can lead to several populations acting simultaneously. However, the ice loading process is very complex, and multiple populations may be caused by other factors and conditions as well. In Figure 27 one can clearly see the presence of two populations.

The three-parameter exponential distribution is simply given as a weighted sum of two one-parameter exponential distributions. The probability density function is given as:

$$f(y) = a\lambda_1 \cdot exp\{-\lambda_1 y\} + (1-a)\,\lambda_2 \cdot exp\{-\lambda_2 y\} \tag{8}$$

where $a$ is a weight parameter, restricted by $0 \leq a \leq 1$. $\lambda_1$ and $\lambda_2$ are the parameters of the two one-parameter exponential distributions. The cumulative distribution function can be found by integrating Equation 8, and can be written as:

$$F(y) = a(1 - exp\{-\lambda_1 y\}) + (1-a)\,(1 - exp\{-\lambda_2 y\}) \tag{9}$$

The estimated extreme value, $\hat{y}_{max}$, is found numerically, by requiring $F(\hat{y}_{max}) = \dfrac{1}{N}$.

The fitted three-parameter distribution is seen in Figure 27. For both the upper and the lower distribution the fit is acceptable. It should be noted that the data set used in the figure is selected to demonstrate the benefit of using a three-parameter exponential distribution. For other data sets, the improved fit by using a three-parameter exponential distribution instead of a regular exponential distribution may be smaller. However, the fit is never worse for a three-parameter than for a one-parameter exponential distribution. If a data set is perfectly described by *one* population, the weight constant $a$ will be zero (or one), which corresponds to a one-parameter exponential distribution. Thus, the model will also work for stationary conditions, which is one of the main advantages of the method.

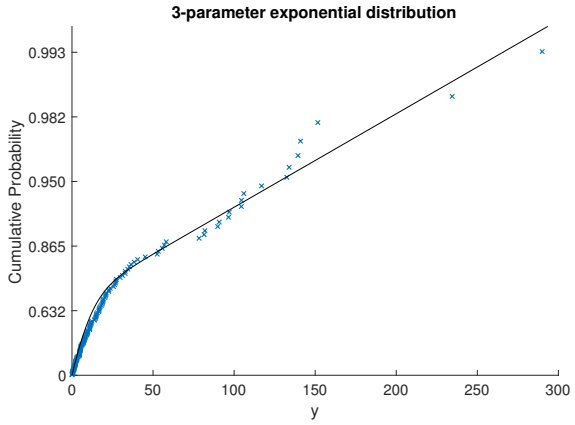



Figure 27: Two populations are observed in the exponential probability paper for set 2 L7.

### 5.5.1 Parameter Estimation

In their work, Suyuthi et al. (2014) investigated several methods for estimating the parameters, i.e., the method of moments, a non-linear least square method, a hybrid approach, and the maximum likelihood estimation approach (MLE). The MLE was found to give the most stable results, by means of always finding a desired and consistent solution. Unlike MLE, the other methods require an initial value, or guess, and a bad initial guess may give invalid or non-converging results. Therefore, Suyuthi et al. (2014) applied the MLE approach in their calculations. Consequently, the MLE approach, which is described below, is used in this thesis as well.

The evaluated data set of size $N$ is sorted and divided into two groups, $n_1$ and $n_2$, where $n_1+n_2 = N$. $n_1$ and $n_2$ represent the two sub-populations. The likelihood function is given as:

$$L(y|\lambda_{y1}, \lambda_{y2}, a) = \frac{N!}{n_1!n_2!} a^{n_1}(1-a)^{n_2}\lambda_{y1}^{n_1}\lambda_{y2}^{n_2} \cdot \exp\left(-\sum_{i=1}^{n_1}\lambda_{y1}y_{1i} - \sum_{i=1}^{n_2}\lambda_{y2}y_{2i}\right) \tag{10}$$

where $n_j$ and $y_{ji}$ are the sample size and the member of the $j$'th sub-population, respectively.

The estimates of the parameters are given by Equation 11 to Equation 13:

$$\hat{a} = \frac{n_1}{N} \tag{11}$$

$$\hat{\lambda}_{y1} = \frac{n_1}{\sum_{i=1}^{n_1} y_{1i}} \tag{12}$$

$$\hat{\lambda}_{y2} = \frac{n_2}{\sum_{i=1}^{n_2} y_{2i}} \tag{13}$$

Since we don't know the optimal size of $n_1$ (and the corresponding size of $n_2$), all possible sets of $n_1$ and $n_2$ must be evaluated, i.e. $n_1 = 0, 1, 2, \cdots, N$, with corresponding $n_2 = N, N-1, N-2, \cdots, 0$.

This gives a total of $N + 1$ sets. The parameters presented in Equation 11 to Equation 13 are calculated for all $(N + 1)$ sets, and the root-mean-square error (RMSE), which is the square root of the average of squared errors, are calculated for each of them. RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{F^*}_i - F^*_{empirical,i})^2}{N}} \tag{14}$$

where $\hat{F^*}_i$ is the ordinate value provided by the three-parameter exponential distribution, and $F^*_{empirical,i}$ is the ordinate value for the measured load. The star (*) indicates the cumulative probability is transformed into a non-linear scale (see Table 5). Since the errors are squared, the method favors solutions with many medium errors rather than some small and some large errors (for equal total error $\sum_{i=1}^{N} |\hat{F^*}_i - F^*_{empirical,i}|$ ). The set of parameters giving the smallest RMSE is identified as the optimal solution.

MATLAB codes related to this section are found in Appendix D.

## 5.6 Non-Parametric Methods

All of the previous methods have in common that they require stationary time series. As seen in Section 5.1, defining stationarity related to ice-induced lads is not straightforward, neither is it possible to obtain in its strictest form. In the Arctic Ocean, winds, currents and weather are constantly changing the composition of the ice, which makes it non-stationary. Broken ice-fields are common, which can consist of level ice, ice floes, ridges, re-frozen ice, open water and sometimes also multi-year ice. The classical approach perform well when the time series are selected with some care, e.g., where stationarity is satisfied to a certain level. However, for some data sets the fit is not satisfactory in the upper tail, which is the important part for estimation of extreme values. Thus, these methods may not be suited for e.g. live estimation of extreme values.

### 5.6.1 Using Kernel Functions

Suyuthi et al. (2012a) proposed a non-parametric probabilistic approach based on an application of the Kernel density estimation. The idea of the method is that each point is represented by a kernel function - or a local PDF -, where the expected value of the kernel function is at the data point. The function can in principle be any pdf, but Suyuthi et al. (2012a) applied a Gaussian kernel function. The total PDF is found by summarizing all the kernel functions. An illustration of the concept is shown in Figure 28. The dashed lines are the Gaussian kernel functions, and the black dots are the data points.

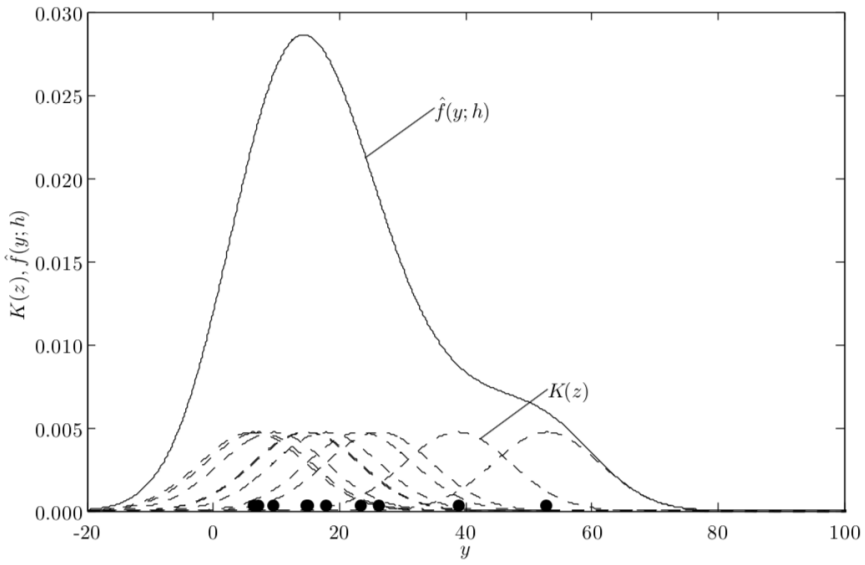



Figure 28: Illustration of the kernel density estimation. The figure is taken from Suyuthi et al. (2012a).

The proposed method was found to give a better fit in the upper tail than the Weibull and one-parameter exponential distribution. The method and its corresponding extreme loads rely solely on

the evaluated data set, making it adequate for "live" estimation of short-term extreme loads onboard the vessel. However, relating the results from this non-parametric method to the prevailing conditions is difficult, and the method is therefore inadequate for long-term statistics and generalization of short-term statistics. For more details about this method, see Suyuthi et al. (2012a).

### 5.6.2 ACER Method

The main problem with the parametric models is that they require stationary processes. As a consequence, the models perform badly for many of the time series, particularly in the tails, which can lead to underestimation of the extreme values. The average conditional exceedance rate method, shortened as the ACER method, is a method that is developed for estimating extreme values from sampled time series. The ACER method was developed and used for the first time by Naess and Gaidai (2009). Their goal was to establish a method for predicting extreme return levels based on sampled time series which takes statistical dependence in the time series into account. Naess and Gaidai (2009) demonstrated the method on measured wind speeds, and the novel method tended to give more accurate results than traditional methods. The ACER method has later been adopted for prediction of extreme value statistics associated with wind speeds (Karpa & Naess, 2013), roll motion for a large container vessel (Gaidai et al., 2016), and ship ice loads (Chai et al., 2018). Compared to the traditional methods, e.g. the peak over threshold and the asymptotic approach, the ACER method have several advantages. Firstly, the method is applicable to both stationary and non-stationary time series. The method also accounts for dependencies in the time series, thus declustering of the data is avoided. There are no further requirements for the data, and all the data are used in the analysis. These properties ensures that one does not have to be very careful when the data set used for analysis is selected, which is the case when the classical approach is applied. Another advantage is that the ACER method puts more weight on the empirical estimates when they are more accurate. This makes the method adequate for real-time series of natural phenomena such as ice-induced loads, which are rarely stationary.

The basic theory and concepts behind the ACER method are presented below. For a more detailed derivation and explanation, see Naess and Gaidai (2009) and Chai et al. (2018). The ACER method provides an extreme value distribution by constructing a set of non-parametric ACER-functions.

The extreme value distribution for a stochastic process X(t), where $M_N = \max\{X_1, X_2, \cdots, X_N\}$ and N is the number of data points, is given as:

$$P(\eta) = P(M_N \leq \eta) = P(X_1 \leq \eta, X_2 \leq \eta, \cdots, X_N \leq \eta) \tag{15}$$

$$P(\eta) = P(X_N \leq \eta | X_{N-1} \leq \eta, \cdots, X_1 \leq \eta) \cdot P(X_{N-1} \leq \eta, \cdots, X_1 \leq \eta) \tag{16}$$

$$= \prod_{j=2}^{N} P(X_j \leq \eta | X_{j-1} \leq \eta, \cdots, X_N \leq \eta) \cdot P(X_1 \leq \eta) \tag{17}$$

$\eta$ [kN] is the actual load value, without an applied threshold. Hence, $\eta$ = y + threshold. If all the data are assumed to be statistically independent, the extreme value distribution is given by the first order approximation:

$$P(\eta) \approx P_1(\eta) = \prod_{j=1}^{N} P(X_j \leq \eta) = \prod_{j=1}^{N} (1 - \alpha_{1j}(\eta)) \tag{18}$$

where $\alpha_{1j}(\eta) = P(X_j > \eta)$. For small values of r, it is known that $1 - r \approx exp(-r)$ with a very small error. Hence, Equation 18 can be rewritten as:

$$P(\eta) \approx P_1(\eta) = \prod_{j=1}^{N} exp(-\alpha_{1j}(\eta)) \approx exp\left(\sum_{j=1}^{N} -\alpha_{1j}(\eta)\right) \tag{19}$$

Statistical dependency can be investigated and accounted for to different levels of extent. The second level approximation $P_2(\eta)$ is based on assuming that the value of each data point is dependent on the previous data point:

$$P(X_j \leq \eta | X_{j-1} \leq \eta, \cdots, X_1 \leq \eta) \approx P(X_j \leq \eta | X_{j-1} \leq \eta), \quad j = 2, \cdots, N \tag{20}$$

From the previous equation it follows that:

$$P(\eta) \approx \prod_{j=2}^{N} P(X_j \leq \eta | X_{j-1} \leq \eta) \cdot P(X_1 \leq \eta) = \prod_{j=2}^{N} (1 - \alpha_{2j}(\eta)) \cdot (1 - \alpha_{11}(\eta))$$

$$\approx \prod_{j=2}^{N} exp(-\alpha_{2j}(\eta)) \cdot exp(-\alpha_{11}(\eta)) \approx P_2(\eta) = exp\left(\sum_{j=1}^{N} -\alpha_{1j}(\eta) - \alpha_{11}\right)$$

For an approximation of order $k$, the final expression in the last equation above can be generalized to:

$$P_k(\eta) = exp\left(-\sum_{j=k}^{N} \alpha_{kj}(\eta) - \sum_{j=1}^{k-1} \alpha_{jj}(\eta)\right), \quad k \geq 2 \tag{21}$$

where $\alpha_{kj}(\eta)$ is the probability of $X_j$ exceeding the value $\eta$, given that the $k-1$ previous data points have not exceeded the same value $\eta$. The expression for $\alpha_{kj}(\eta)$ then becomes:

$$\alpha_{kj}(\eta) = P(X_j > \eta | X_{j-1} \leq \eta, X_{j-2} \leq \eta, \cdots, X_{j-(k-1)} \leq \eta, \quad j = k, \cdots, N$$

For $N \gg k$, which is the case for practical applications, the second term in the exponent in Equation 21 becomes very small compared to the first term, and can therefore be neglected. Thus, the expression can be approximated as:

$$P_k(\eta) \approx exp\left(-\sum_{j=k}^{N} \alpha_{kj}(\eta)\right), \quad k \geq 2 \tag{22}$$

The average exceedance rate (the ACER function) $\varepsilon_k(\eta)$ conditioning on $k$ is introduced:

$$\varepsilon_k(\eta) = \frac{1}{N-k+1} \sum_{j=k}^{N} \alpha_{kj}(\eta), \quad k \geq 2 \tag{23}$$

where $\alpha_{kj}$ can be written as:

$$\alpha_{kj}(\eta) = \frac{E[A_{kj}]}{E[B_{kj}]}, \quad 2 \leq k \leq j \leq N \tag{24}$$

where $E[\cdot]$ is the expected value operator, and $A_kj$ and $B_kj$ are given as:

$$A_{kj} = \begin{cases} 1 & \text{when} \quad X_j > \eta, X_{j-1} \leq \eta, X_{j-2} \leq \eta, \cdots, X_{j-(k-1)} \leq \eta \\ 0 & \text{otherwise} \end{cases} \qquad j = k, \cdots, N; \quad k \geq 2$$

$$B_{kj} = \begin{cases} 1 & \text{when} \quad X_{j-1} \leq \eta, X_{j-2} \leq \eta, \cdots, X_{j-(k-1)} \leq \eta \\ 0 & \text{otherwise} \end{cases} \qquad j = k, \cdots, N; \quad k \geq 2$$

If the data is assumed to be taken from an ergodic process, then the estimate of the ACER function can be expressed as:

$$\bar{\varepsilon}_k(\eta) = \lim_{\eta \to \infty} \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{\sum_{j=k}^{N} B_{kj}(\eta)} \tag{25}$$

When extreme values are predicted, $\eta$ will have a relatively high value. In addition, the sample size $N$ is in practical applications a high number. With that in mind, we can argue that $\sum_{j=k}^{N} B_{kj}(\eta) \approx N$, since

$$\lim_{\eta \to \infty} \sum_{j=k}^{N} B_{kj}(\eta) = N - k + 1 \approx N$$

A modified ACER function, $\tilde{\varepsilon}_k(\eta)$, is introduced such that $\lim_{\eta \to \infty} \tilde{\varepsilon}_k(\eta) / \bar{\varepsilon}_k(\eta) = 1$, where

$$\tilde{\varepsilon}_k(\eta) = \lim_{\eta \to \infty} \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{N-k+1} \tag{26}$$

In this thesis, the ACER method will be used to calculate extreme levels (high values of $\eta$), thus any function that provides correct estimates of the ACER function may be used. Consequently, the empirical ACER function $\hat{\varepsilon}_k(\eta)$ is introduced:

$$\hat{\varepsilon}_k(\eta) = \frac{\sum_{j=k}^{N} A_{kj}(\eta)}{N-k+1} \tag{27}$$

In Figure 29, $\hat{\varepsilon}_k(\eta)$ is plotted for different values of $k$. As one can see, the order of $\hat{\varepsilon}_k(\eta)$ is not important for medium and high values of $\eta$. This indicates that the data in the sample to a short extent show statistical dependency. For small values of $\eta$, $\hat{\varepsilon}_k(\eta)$ is smaller for the same value of $\eta$ for higher values of $k$. Even for a statistically independent random process, this is expected due to randomness in the data points. A 95% confidence interval for the ACER functions can be expressed by Equation 28 (Naess and Gaidai (2009); Gaidai et al. (2016); Chai et al. (2018)).

$$CI_{\pm} = \hat{\varepsilon}_k(\eta)\left(1 \pm \frac{1.96}{\sqrt{(N-k+1)\hat{\varepsilon}_k(\eta)}}\right) \tag{28}$$

The ACER function is used for prediction of extreme values for return periods which extends beyond the data set by extrapolation of the created ACER functions. The extrapolation is based on the assumption of an underlying asymptotic extreme value distribution of the Gumbel type. This proves to be the case for most engineering applications (Gaidai et al., 2016). The ACER functions are in general highly regular in the tail region $\eta \geq \eta_0$, and it is argued by (Naess & Gaidai, 2009) that they follow a function on the form

$$\varepsilon_k(\eta) = q_k \cdot exp\left(-a_k(\eta - b_k)^{c_k}\right), \quad \eta \geq \eta_0 \tag{29}$$

Here $q_k$, $a_k$, $b_k$ and $c_k$ are dependent on the data set and $k$. Equation 29 is asymptotic of the Gumbel type for any order of the ACER function. However, the requirements for an underlying Gumbel extreme value distribution, i.e., stationarity, does not have to be fulfilled.

The constants $q_k$, $a_k$, $b_k$ and $c_k$ are determined by minimizing the mean square error function

$$F(q_k, a_k, b_k, c_k) = \sum_{i=1}^{M} \rho_i |\ln \hat{\varepsilon}_k(\eta_i) - \ln q + a \cdot (\eta_i - b)^c|^2 \tag{30}$$

where $i = 1, ..., M$ are the data points that are empirically estimated by $\hat{\varepsilon}_k(\eta)$. $\rho_i$ is a factor that puts more weight on the more accurate estimates, and it is defined as:

$$\rho_i = (\ln CI^+(\eta_i - \ln CI^-(\eta_i))^{-2} \tag{31}$$

The purpose of $\rho_i$ is to put more weight on data with relatively smaller confidence interval on the log level. However, the exact expression is somewhat arbitrary. For instance, the data points could be more or less equally weighted by increasing or decreasing the exponent, respectively. The effect of using different expressions for $\rho_i$ have not been investigated in this thesis, and the expression is taken from Naess and Gaidai (2009) and Chai et al. (2018).

The task of determining the constants in Equation 29 is an optimization problem. By looking at Equation 30, one can see that with $b$ and $c$ fixed, the problem becomes a weighted linear regression

problem $y_i = -ax_i + ln(q)$ (Karpa & Naess, 2013), where $y_i = ln(\hat{\varepsilon}_k(\eta_i))$ and $x_i = (\eta_i - b)^c$ with the solution

$$a^*(b_k, c_k) = -\frac{Cov(x,y)}{Var(x)} = -\frac{\sum_{i=1}^{M} \rho_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{M} \rho_i (x_i - \bar{x})^2}$$

$$\ln q^*(b,c) = \bar{y} + a^*(b,c) \cdot \bar{x}$$

Here $\bar{x}$ and $\bar{y}$ are the the sum of the weighted $x_i$ and $y_i$, respectively. $Cov(x,y)$ and $Var(x)$ are the sample weighted covariance and variance.

A nonlinear optimization of the Levenberg-Marquardt type is applied to find the optimal values of $b_k^*$ and $c_k^*$ by applying it to the modified mean square function

$$\tilde{F}(b,c) = F(a_k^*(b,c), b, c, q_k^*(b,c)) = Var(y) - \frac{Cov^2(x,y)}{Var(x)}$$

The MATLAB code for the optimization can be found in the script "acer_fit.m" in Appendix D.

Finally, $q_k^*$, $a_k^*$, $b_k^*$ and $c_k^*$ are used in the extrapolation function given in Equation 29, and plotted together with the empirical ACER function. An estimate of a 95% confidence interval for the extrapolation function is obtained by adjusting the empirical confidence bands to the extrapolation function, and perform a similar fitting procedure to them as for the empirical ACER function. Figure 30 demonstrates the concept. The blue stair plot is the empirical ACER function, the solid line is the extrapolation function, while the dashed and dotted lines are the 95% confidence interval boundaries for the empirical ACER function and the extrapolation function, respectively.

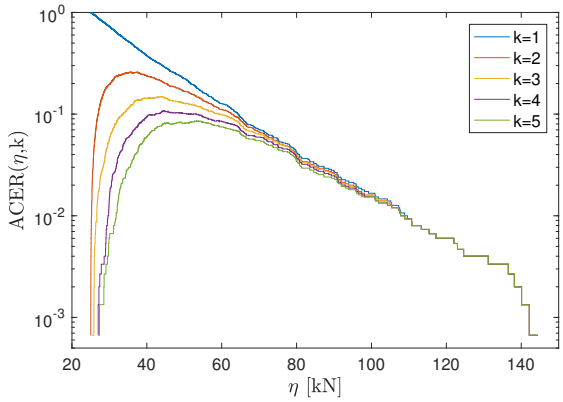



Figure 29: ACER functions k=1, ..., 5. Note that for the upper tail (large $\eta$), the ACER function is independent of $k$.

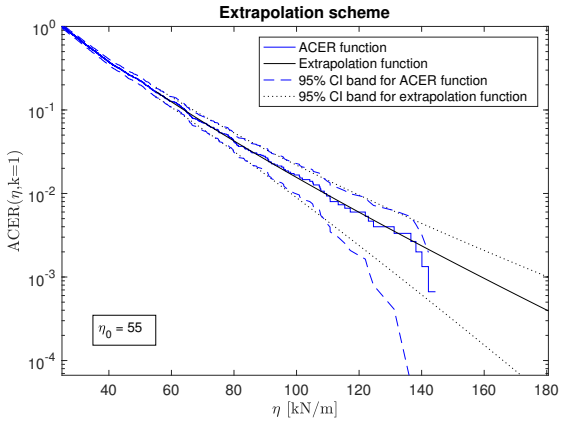



Figure 30: Extrapolation scheme for the ACER function (stairs). The dashed lines indicate the 95% confidence interval for the data points. The dotted lines represent the fitted CI.

# 6 MATLAB Routines

Programming and implementing the different methods using MATLAB has been a significant part of the work in this thesis, both when considering time consumption and challenges during the process. All scripts and codes used in the statistical analyses and plotting of data are developed by the author. For some methods, e.g. for the ACER method (Section 5.6.2), existing programs and scripts could have been used. However, scripts were programmed for these purposes as well, mainly due to two reasons. The primary motivation has been that when a method is transformed from theory to a code, every part of the theory behind the method is covered, and a deeper understanding is required in order to get the code running correctly. Bugs and errors that may occur during the development of the codes are sources for valuable learning and understanding of the theory. Thus, the programming has been an essential part of the learning process, and a tool for getting the theory into the fingertips. Another motivation for doing all the programming is to improve my programming skills in general, which is an important skill to possess as an engineer.

The most important codes will be briefly explained and summarized below, while the scripts themselves are found in Appendix D.

## 6.1 separate.m

### 6.1.1 Inputs and Outputs

Table 8: Inputs and outputs for separate_s.m

| | Name | Type | Comment |
|---|---|---|---|
| Input | peakDB | Matrix | Matrix containing loads and associated data. |
| Output | L$i$ | Matrix | Load matrix for each sensor $i = 1, ..., 9$. |

### 6.1.2 Description

The file peakDB_2007.mat contains the matrix peakDB. The second column in the matrix gives the measured loads. The corresponding time of occurrence, the duration of the load and at which sensor the load is measured is found in column one, three and four, respectively.

separate_s.m creates a matrix for each of the load sensors named L1, L2, ..., L9. Each load incident and its corresponding time of occurrence and load duration is presented in individual rows. The loads and corresponding data are listed chronologically with respect to time of occurrence. The matrices L1, L2, ..., L9 will be used for analyzes of load data, which always are performed for one sensor at a time.

## 6.2 prob_paper.m

### 6.2.1 Inputs and Outputs

Table 9: Inputs and outputs for prob_paper.m.

| | Name | Type | Comment |
|---|---|---|---|
| Input | ILMdata.mat | File | File containing (among others) the matrix "time". |
| | separate_s.m | Script | Described in Section 6.1. |
| | start | Variable | Start point in the file "data"/"time" (from ILMdata.mat). |
| | stop | Variable | Stop point in the file "data"/"time" (from ILMdata.mat). |
| | LX | Variable | The load matrix for the desired load sensor. |
| | distr | Variable | The distribution, given by the user, that is used for fitting the measured data. |
| | threshold | Variable | Lower threshold for the loads. 25 kN is default value, but a higher threshold can be applied if desired. |
| Output | dataset | Vector | Sorted loads from the selected load sensor in the time interval given by "start"/"stop". |
| | p_i | Vector | Empirical cumulative distribution function for "dataset". |
| | figure | Plot | Plot of "dataset" in the desired probability paper with fitted line. |
| | m | Variable | Slope of the fitted line. |
| | c | Variable | Ordinate interception for the fitted line. |
| | R_squared | Variable | Coefficient of determination. |

### 6.2.2 Description

This script fits a distribution to the load data in a probability paper. ILMdata.mat is loaded (containing the matrices "data", "sensors" and "time"), which contains measurements (speed over ground, position, ice thickness, propulsion power, etc.) and the time at which each measurement is collected. These data are given for every 30 seconds. These data are relevant since the selection of time series is based on the ice conditions and operating characteristics. If desired, the user can apply a higher threshold than the default threshold. The user must type in which element numbers in the matrix "data" (or "time") the evaluated time series shall span over. The corresponding time window is recognized in "LX", which is the load matrix for the selected load sensor. The script plots the measured loads in a desired probability paper. The user can choose between five distributions (exponential, log-normal, Gumbel, Weibull and the three-parameter exponential). The built-in function "lsline" is used for fitting a line to the data points using the least square method, and the slope and the ordinate interception are found. For the exponential distribution, the fitted line is forced through the origin. The coefficient of determination, $R^2$, is also calculated.

## 6.3 qq_plot.m

### 6.3.1 Inputs and Outputs

Table 10: Inputs and outputs for qq_plot.m

| | Name | Type | Comment |
|---|---|---|---|
| Input | ILMdata.mat | File | File containing (among others) the matrix "time". |
| | separate_s.m | Script | Described in the Section **??** |
| | start | Variable | Start point in the file "data"/"time" (from ILMdata.mat) |
| | stop | Variable | stop point in the file "data"/"time" (from ILMdata.mat) |
| | LX | Matrix | The load matrix for the desired load sensor |
| | method | Variable | Desired method (maximum likelihood or by fitting to the probability paper) for estimating the parameters of the distribution |
| | parameters.m | Script | Described in Section **??**. |
| Output | Figure | Plot | Quantile-Quantile plot of the measured loads and the desired probability distribution. |

### 6.3.2 Description

qq_plot.m is quite similar to prob_paper.m (see Section 6.2). Instead of plotting the data points in a probability paper, x and corresponding y values for the same quantiles given by the evaluated probability distribution function and the data set, respectively, are plotted.

## 6.4 parameters.m

### 6.4.1 Inputs and Outputs

Table 11: Inputs and outputs for parameters.m

| | Name | Type | Comment |
|---|---|---|---|
| Input | prob_paper.m | Script | Described in Section 6.2, see outputs. |
| | method | Variable | User input: Desired method used for estimating the parameters; based on probability paper or MLE. |
| Output | lambda | Variable | Distribution parameter for exponential distribution. |
| | mu | Variable | Distribution parameter for log-normal distribution. |
| | sigma | Variable | Distribution parameter for log-normal distribution. |
| | alpha | Variable | Distribution parameter for Gumbel distribution. |
| | beta | Variable | Distribution parameter for Gumbel distribution. |
| | k | Variable | Distribution parameter for Weibull distribution. |
| | theta | Variable | Distribution parameter for Weibull distribution. |
| | parameter1 | Variable | First parameter for the evaluated distribution. |
| | parameter2 | Variable | Second parameter for the evaluated distribution. |

### 6.4.2 Description

The purpose of this script is to calculate the distribution parameters. The parameters are calculated either based on the probability paper or by the maximum likelihood estimation method. For the maximum likelihood estimation, some of the parameters are found using built-in functions in MATLAB. The other estimates, both for MLE and estimation based on probability paper, are calculated using the formulas given in Table 6 and Table 7.

## 6.5 asymptotic.m

### 6.5.1 Inputs and Outputs

Table 12: Inputs and outputs for asymptotic.m

| | Name | Type | Comment |
|---|---|---|---|
| Input | ILMdata | File | File containing (among others) the matrix "time". |
| | separate_s.m | Script | Described in Section 6.1. |
| | Dt | Variable | Time interval of which the maximum value is identified. |
| | start | Variable | Start point in the file "data"/"time" (from ILMdata.mat). |
| | stop | Variable | Stop point in the file "data"/"time" (from ILMdata.mat). |
| | LX | Matrix | The load matrix for the desired load sensor. |
| | threshold | Variable | Lower threshold for the loads. 25 kN is default value, but a higher threshold can be applied if desired. |
| Output | figure1 | Plot | Dt Plot of Dt maximas in probability paper, with least square fit. |
| | R_squared | Variable | Coefficient of determination. |
| | D | Variable | Result of the KS-test; maximum difference between empirical and fitted CDF. |
| | figure2 | Plot | Illustration of the Kolmogorov-Smirnov test. |

### 6.5.2 Description

This script covers the asymptotic approach, described in Section 5.4. The user defines the time series, the length of the sub-intervals and the load sensor. Another threshold than 25 kN can be applied if desired. The matrix "peak_int" is created, where the measured loads for each sub-interval are stored in individual rows. The maximum value for each row, or each sub-interval, is identified and stored in the new vector "dataset". The vector is then sorted and plotted in a Gumbel probability paper. A linear line is fitted by the least square method, and the coefficient of determination is calculated. The parameters are then estimated, either based on the probability paper or by the maximum likelihood approach. Finally, the Kolmogorov-Smirnov test is applied, as described in Section 6.11.

## 6.6 three_param_exp.m

### 6.6.1 Inputs and Outputs

Table 13: Inputs and outputs for three_param_exp.m

| | Name | Type | Comment |
|---|---|---|---|
| Input | dataset | Variable | Vector of measured loads being analyzed. |
| Output | a | Variable | Weight parameter. |
| | lambda1 | Variable | Rate parameter of the first distribution. |
| | lambda2 | Variable | Rate parameter of the second distribution. |
| | figure | Plot | The data set and the fitted three-parameter exponential distribution in a exponential probability paper. A least square fit is also plotted for comparison with a regular exponential distribution. |

### 6.6.2 Description

The script estimates the parameters for all possible ways to split the data set, and calculate the Root-Mean-Square Error (RMSE) for each set of parameters. The set providing the smallest RMSE is identified, and they are used when the fitted line is plotted together with the data set in an exponential probability paper. A linear fit is plotted for comparison with a regular one-parameter exponential distribution.

## 6.7 acer.m

### 6.7.1 Inputs and Outputs

Table 14: Inputs and outputs for acer.m.

| | Name | Type | Comment |
|---|---|---|---|
| Input | ILMdata.mat | File | File containing (among others) the matrix "time". |
| | separate_s.m | Script | Described in Section 6.1. |
| | start | Variable | Start point in the file "data"/"time" (from ILMdata.mat). |
| | stop | Variable | Stop point in the file "data"/"time" (from ILMdata.mat). |
| | LX | Variable | The load matrix for the desired load sensor. |
| | threshold | Variable | Lower threshold for the loads. 25 kN is default value, but a higher threshold can be applied if desired. |
| | N_k | Variable | Highest order of the acer function to be included. |
| Output | eta | Vector | Measured loads to be analyzed. |
| | rho | Vector | Weight factor used in the fitting process. |
| | dataset | Vector | Load vector. |
| | ACER | Matrix | Containing the value of the ACER function for each data point for ACER functions up to order N_k. |
| | CIminus | Matrix | Lower bound of the confidence interval for each ACER function up to order N_k. |
| | CIplus | Matrix | Upper bound of the confidence interval for each ACER function up to order N_k. |
| | figure | Plot | All ACER functions up to order N_k plotted in logarithmic scale on the y-axis and linear scale on the x-axis. |

### 6.7.2 Description

The purpose of the script is to calculate and plot the ACER functions. The user defines the start and end of the data set, the load sensor that shall be applied, and the highest order of the ACER functions. Another threshold than 25 kN can be applied if desired. The number of elements smaller or equal to eta, $i$, is calculated, and the ACER function and P (cumulative probability) are calculated for $k = 1$, which is a special case. Then, the same matrices are calculated for $k > 1$, which is the general case. Next, the weight factor, rho, and the bounds of the confidence intervals used for fitting (later) are estimated. Finally, the ACER functions of the orders k = 1, ..., N_k are plotted.

## 6.8 acer_fit.m

### 6.8.1 Inputs and Outputs

Table 15: Inputs and outputs for acer_fit.m.

| | Name | Type | Comment |
|---|---|---|---|
| Input | acer.m | Script | Described in section 6.7. |
| | eta_0 | Variable | The lower load limit from which the ACER function shall be fitted. |
| | target | Variable | The target exceedance probability. |
| | dataset | Vector | Load vector. |
| | rho | Matrix | Weight factors used in the fitting process. |
| | eta | Vector | Load vector. |
| | ACER | Variable | Containing the value of the ACER function for each data point for ACER functions up to order N_k. |
| Output | a | Variable | Optimized parameter used for the extrapolation function. |
| | b | Variable | Optimized parameter used for the extrapolation function. |
| | c | Variable | Optimized parameter used for the extrapolation function. |
| | q | Variable | Optimized parameter used for the extrapolation function. |
| | a_minus | Variable | Optimized parameter used for the lower bound of the confidence interval (CI) for the extrapolation function. |
| | b_minus | Variable | Optimized parameter used for the lower bound of the CI for the extrapolation function. |
| | c_minus | Variable | Optimized parameter used for the lower bound of the CI for the extrapolation function. |
| | q_minus | Variable | Optimized parameter used for the lower bound of the CI for the extrapolation function. |
| | a_plus | Variable | Optimized parameter used for the upper bound of the confidence interval (CI) for the extrapolation function. |
| | b_plus | Variable | Optimized parameter used for the upper bound of the CI for the extrapolation function. |
| | c_plus | Variable | Optimized parameter used for the upper bound of the CI for the extrapolation function. |
| | q_plus | Variable | Optimized parameter used for the upper bound of the CI for the extrapolation function. |
| | figure | Plot | Showing the ACER function plotted against the loads, together with the the confidence interval bands, the extrapolation function and the confidence bands for the extrapolation function. |

### 6.8.2 Description

The purpose of the script is to provide an extrapolation scheme for determining extreme loads by the ACER method. The extrapolation function is a function of the parameters a, b, c and q. Initially, the user must give the value of eta_0, which is the load level where the extrapolation function starts, and the target exceedance probability. The start and end point in the load vector that shall be used for estimating the extrapolation function are found (i and j). The objective function and the optimization characteristics are given, before built-in optimization routines are used for the optimization of a, b, c and q. The same optimization procedure is repeated for the bounds of the confidence interval for the extrapolation function. A plot showing the ACER function and the extrapolation function, with the confidence interval bounds for both functions is generated.

## 6.9 bootstrapping.m

### 6.9.1 Inputs and Outputs

Table 16: Inputs and outputs for bootstrapping.m.

| | Name | Type | Comment |
|---|---|---|---|
| Input | method | Variable | User input: Desired method used for estimating the parameters; based on probability paper or MLE. |
| | parameters.m | Script | Described in Section 6.4 |
| | n_boot | Variable | Number of bootstrapped samples, given by the user. |
| Output | lambda_hat | Variable | Distribution parameter for exponential distribution. |
| | mu_hat | Variable | Distribution parameter for log-normal distribution for the measured data. |
| | sigma_hat | Variable | Distribution parameter for log-normal distribution for the measured data. |
| | alpha_hat | Variable | Distribution parameter for Gumbel distribution for the measured data. |
| | beta_hat | Variable | Distribution parameter for Gumbel distribution for the measured data. |
| | k_hat | Variable | Distribution parameter for Weibull distribution for the measured data. |
| | theta_hat | Variable | Distribution parameter for Weibull distribution for the measured data. |
| | param | Matrix | Estimated distribution parameters for bootstrapped samples. |
| | Figure | Plot | Plot of bootstrapped data points together with measured data points in probability paper. |

### 6.9.2 Description

Initially, the script obtains the distribution parameters from parameters.m for the time series defined in prob_paper.m. The number of bootstrapped samples is defined by the user. A set (of the same size as the evaluated data set) of random numbers between 0 and 1 is generated, representing cumulative probabilities for the bootstrapped sample. The parameters obtained from parameters.m are used to transform the cumulative probabilities into loads. Distribution parameters for this data set are calculated and stored in the matrix "param". The process is repeated "n_boot times, which is defined by the user. The distribution parameters calculated for the measured loads are always used for calculating the bootstrapped data sets. Finally, the bootstrapped data points are plotted in a probability paper together with the measured loads.

## 6.10 confidence_int.m

### 6.10.1 Inputs and Outputs

Table 17: Inputs and outputs for confidence_int.m

| | Name | Type | Comment |
|---|---|---|---|
| Input | conf_int | Variable | User defined confidence interval. |
| | par_inp | Variable | User input; which parameter a confidence interval shall be estimated for. |
| | param | Matrix | Matrix containing estimated distribution parameters for bootstrapped data sets. |
| | n_stem | Variable | User defined number of stems in stem plot of the bootstrapped parameters. |
| Output | CI_min | Variable | Lower bound of the confidence interval. |
| | CI_plus | Variable | Upper bound of the confidence interval. |
| | figure | Plot | Plot showing the distribution of the estimated parameters, the bounds of the confidence interval and the estimated parameter for the measured loads. |

### 6.10.2 Description

The main objective of this script is to estimate a confidence interval for the distribution parameters by using the estimated parameters from the bootstrapped samples. The confidence level "conf_int" is defined by the user of the script. The bounds of the confidence interval, "CI_min" and "CI_plus", are identified. Next, the bootstrapped parameters are sorted and divided into stems, where each stem spans equally large intervals. The stems are plotted together with the confidence interval bounds and the parameter estimated for the measured loads evaluated, $\hat{\mu}$.

## 6.11 kolmogorov_smirnov.m

### 6.11.1 Inputs and Outputs

Table 18: Inputs and outputs for kolmogorov_smirnov.m

| | Name | Type | Comment |
|---|---|---|---|
| Input | method | Variable | User input: Desired method used for estimating the parameters; based on probability paper or MLE. |
| | parameters.m | Script | Described in Section 6.4 |
| Output | D | Variable | Result of the KS-test; maximum difference between empirical and fitted CDF. |
| | figure | Plot | Illustration of the Kolmogorov-Smirnov test. |

### 6.11.2 Description

The script uses the estimated parameters obtained from parameters.m to generate and plot the fitted distribution. The empirical CDF is plotted as a stair plot in the same figure. Finally, the largest deviation between the empirical CDF and the fitted CDF is marked on the figure.

# 7 Numerical Results

In this section, the main results are presented. Relevant plots, e.g. selected probability papers, Q-Q plots, Kolmogorov-Smirnov tests and plots of ACER functions are presented. The figures and plots that are excluded from the main text can be found in Appendix C.

## 7.1 Effects of Different Conditions

As discussed in Section 5.1, which factors that affects the ice loads are not fully understood. In their search for stationary time series, different parameters are given attention by different works (Suyuthi et al. (2013b) Suyuthi et al. (2012b), Chai et al. (2018), Kujala and Vuorio (1986), etc). To be able to say something about how the different parameters, e.g. vessel speed, ice thickness and propulsion power affect the ice loads is valuable, both from a design perspective and for navigation purposes. In this thesis, the effects of ice conditions and vessel speed are investigated by pairwise comparison of different data sets where the ice conditions are similar, whereas the vessel speed is different, and vice versa. Table 1 is used to identify adequate sets, and the selected ones are presented in Table 19 and Table 20. The relation between the most probable largest load during one hour and ice thickness are also investigated for six of the sets in Table 1, and the same is done for speed over ground.

Table 19: Pairwise similar sets with respect to ice conditions.

| Set | Elements in the matrix "data" | Duration | Ice thickness | | | Speed over ground | | |
|---|---|---|---|---|---|---|---|---|
| | | [min] | $\mu$ [m] | $\sigma$ [m] | $\mu_i/\mu_j$ | $\mu$ [m/s] | $\sigma$ [m/s] | $\mu_i/\mu_j$ |
| $i = 6$ | 11002-11032 | 15 | 0.51 | 0.14 | 0.99 | 5.90 | 0.16 | 2.20 |
| $j = 9$ | 15482-15542 | 30 | 0.51 | 0.15 | | 2.68 | 0.61 | |
| $i = 7$ | 11056-11086 | 15 | 0.99 | 0.20 | 1.00 | 4.87 | 0.54 | 0.91 |
| $j = 8$ | 11090-11130 | 20 | 0.98 | 0.28 | | 5.33 | 0.49 | |

Table 20: Pairwise similar sets with respect to speed over ground.

| Set | Elements in the matrix "data" | Duration | Speed over ground | | | Ice thickness | | |
|---|---|---|---|---|---|---|---|---|
| | | [min] | $\mu$ [m] | $\sigma$ [m] | $\mu_i/\mu_j$ | $\mu$ [m/s] | $\sigma$ [m/s] | $\mu_i/\mu_j$ |
| $i = 5$ | 10942-11002 | 30 | 5.06 | 0.54 | 1.00 | 0.82 | 0.24 | 0.82 |
| $j = 13$ | 11058-11123 | 32.5 | 5.07 | 0.55 | | 1.00 | 0.22 | |
| $i = 2$ | 7770-7830 | 30 | 4.92 | 0.81 | 1.01 | 1.13 | 0.36 | 1.14 |
| $j = 7$ | 11056-11086 | 15 | 4.87 | 0.54 | | 0.99 | 0.20 | |

For each pair, $i$ and $j$, of sets with similar conditions, several characteristics and outputs are compared. The data sets are plotted in a Weibull probability paper, a line is fitted by the least square method, and the coefficient of determination is calculated. The Weibull distribution parameters, $k$ and $\lambda$, are estimated by the maximum likelihood approach. The Weibull distribution is chosen because it has, together with the exponential distribution, proved to give the best fit for ice-induced

loads on ships (Kujala and Vuorio (1986), Suominen and Kujala (2010), Suyuthi et al. (2013b)). In addition, the number of measured loads larger than the threshold (per hour) is identified, and the most probable largest load during one hour, $\hat{x}_{max,1h}$, is calculated. Note that $x = y + 25kN$. The results are presented in tables 21 to 24 below. Unfortunately, finding adequate time series for this purpose was difficult. Only two pairs for each of the two categories (similar ice thickness properties and similar speed over ground) were found. When the pairs were selected, it was required that the mean of either the speed over ground or the ice thickness were close to equal, and that their standard deviations, given in Table 1, were relatively small and similar to each other. The time series must also provide enough data points to get reliable results. Since few pairs of similar sets were found, measurements from all the strain sensors, L1, ..., L8, are analyzed. Some of the load incidents may be registered more than one sensor, thus the data sets from different sensors may not be completely independent. On the other hand, they are exposed to the same conditions and give therefore a good indication of the randomness within the data.

**For similar ice thickness:**

Table 21: Ice thickness: $\mu_i = 0.51$, $\mu_j = 0.51$. Speed over ground: $\mu_i = 5.90$, $\mu_j = 2.68$. For more data, see Table 19.

| $i = 6$ | | Sensor | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $j = 9$ | | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 |
| $k$ | $i$ | 0.902 | 1.192 | 1.148 | 0.951 | 0.880 | 0.870 | 1.028 | 0.788 |
| | $j$ | 1.121 | 1.133 | 0.822 | 0.938 | 0.828 | 0.923 | 0.932 | 0.808 |
| $\theta$ | $i$ | 20.11 | 17.18 | 20.25 | 14.96 | 11.71 | 11.30 | 13.98 | 12.40 |
| | $j$ | 17.17 | 14.27 | 13.87 | 20.55 | 14.12 | 20.42 | 15.80 | 12.28 |
| $R^2$ | $i$ | 0.986 | 0.983 | 0.988 | 0.993 | 0.973 | 0.984 | 0.991 | 0.981 |
| | $j$ | 0.987 | 0.974 | 0.983 | 0.988 | 0.987 | 0.989 | 0.992 | 0.989 |
| $Loads/h$ | $i$ | 144 | 154 | 148 | 180 | 220 | 380 | 344 | 156 |
| | $j$ | 48 | 164 | 32 | 180 | 290 | 568 | 304 | 338 |
| $\hat{x}_{max,1h}$ | $i$ | 144.0 | 91.8 | 107.2 | 109.5 | 104.4 | 112.5 | 102.9 | 121.8 |
| | $j$ | 82.4 | 85.1 | 87.9 | 144.1 | 139.8 | 176.2 | 127.5 | 133.7 |

Table 22: Ice thickness: $\mu_i = 0.99$, $\mu_j = 0.98$. Speed over ground: $\mu_i = 4.87$, $\mu_j = 5.33$. For more data, see Table 19.

| $i = 7$ | | Sensor | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $j = 8$ | | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 |
| $k$ | $i$ | 0.914 | 0.893 | 0.828 | 0.829 | 0.972 | 0.942 | 0.779 | 0.803 |
| | $j$ | 0.958 | 0.937 | 0.960 | 0.917 | 0.845 | 0.851 | 1.034 | 0.876 |
| $\theta$ | $i$ | 25.68 | 20.71 | 24.21 | 27.05 | 19.23 | 18.28 | 18.84 | 14.42 |
| | $j$ | 25.31 | 26.07 | 31.31 | 23.09 | 19.3 | 19.78 | 24.47 | 15.35 |
| $R^2$ | $i$ | 0.984 | 0.978 | 0.972 | 0.970 | 0.988 | 0.978 | 0.989 | 0.979 |
| | $j$ | 0.994 | 0.982 | 0.964 | 0.989 | 0.988 | 0.994 | 0.992 | 0.952* |
| $Loads/h$ | $i$ | 248 | 172 | 228 | 204 | 488 | 500 | 244 | 216 |
| | $j$ | 294 | 228 | 270 | 315 | 552 | 618 | 450 | 258 |
| $\hat{x}_{max,1h}$ | $i$ | 191.4 | 154.8 | 211.9 | 227.9 | 150.4 | 152.1 | 206.6 | 142.0 |
| | $j$ | 180.3 | 183.7 | 213.4 | 180.7 | 195.7 | 201.1 | 166.0 | 133.6 |

* Too bad fit to be included in the work

**For similar speed over ground:**

Table 23: Speed over ground: $\mu_i = 5.06$, $\mu_j = 5.07$. Ice thickness: $\mu_i = 0.82$, $\mu_j = 1.00$. For more data, see Table 20.

| $i = 5$ | | Sensor | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $j = 13$ | | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 |
| $k$ | $i$ | 1.004 | 1.055 | 0.909 | 0.903 | 0.833 | 0.886 | 0.854 | 0.800 |
| | $j$ | 0.968 | 0.945 | 0.901 | 0.908 | 0.896 | 0.892 | 0.930 | 0.839 |
| $\theta$ | $i$ | 20.41 | 23.81 | 20.73 | 22.40 | 15.10 | 18.22 | 17.79 | 17.25 |
| | $j$ | 25.51 | 24.94 | 27.86 | 25.74 | 19.17 | 19.65 | 23.27 | 15.16 |
| $R^2$ | $i$ | 0.990 | 0.993 | 0.996 | 0.988 | 0.995 | 0.990 | 0.988 | 0.989 |
| | $j$ | 0.996 | 0.985 | 0.978 | 0.992 | 0.995 | 0.997 | 0.994 | 0.973 |
| $Loads/h$ | $i$ | 234 | 214 | 206 | 244 | 388 | 482 | 382 | 174 |
| | $j$ | 277 | 214 | 258 | 270 | 567 | 585 | 426 | 257 |
| $\hat{x}_{max,1h}$ | $i$ | 135.6 | 142.1 | 155.6 | 172.8 | 153.7 | 167.3 | 168.6 | 159.0 |
| | $j$ | 176.9 | 172.6 | 211.7 | 196.7 | 175.6 | 181.8 | 186.3 | 141.8 |

Table 24: Speed over ground: $\mu_i = 4.92$, $\mu_j = 4.87$. Ice thickness: $\mu_i = 1.13$, $\mu_j = 0.99$. For more data, see Table 20.

| $i = 2$ | | Sensor | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $j = 7$ | | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 |
| $k$ | $i$ | 0.801 | 0.954 | 0.888 | 0.872 | 0.789 | 0.782 | 0.695 | 0.714 |
| | $j$ | 0.914 | 0.893 | 0.828 | 0.829 | 0.972 | 0.942 | 0.779 | 0.803 |
| $\theta$ | $i$ | 20.17 | 20.54 | 20.69 | 32.03 | 24.67 | 22.47 | 20.05 | 12.59 |
| | $j$ | 25.68 | 20.71 | 24.21 | 27.05 | 19.23 | 18.28 | 18.84 | 14.42 |
| $R^2$ | $i$ | 0.976 | 0.985 | 0.947 | 0.985 | 0.978 | 0.995 | 0.982 | 0.931 |
| | $j$ | 0.984 | 0.978 | 0.972 | 0.970 | 0.988 | 0.978 | 0.989 | 0.979 |
| $Loads/h$ | $i$ | 104 | 122 | 86 | 110 | 276 | 372 | 298 | 322 |
| | $j$ | 248 | 172 | 228 | 204 | 488 | 500 | 244 | 216 |
| $\hat{x}_{max,1h}$ | $i$ | 162.3 | 131.4 | 136.2 | 214.0 | 245.1 | 243.3 | 270.3 | 171.9 |
| | $j$ | 191.4 | 154.8 | 211.9 | 227.9 | 150.4 | 152.1 | 206.6 | 142.0 |

As seen from tables 21 to 24, there are large variations in the provided data, and some results are counter-intuitive. This may be explained by that the ice-loading process is highly random by nature, and that not all of the data sets provides equally good fits to the Weibull distribution. Considering the limited amount of data, it is difficult to detect clear patterns in the analyzed data. However, an attempt is made to say something about how the ice thickness and vessel speed influence the most probable largest load during one hour, $\hat{x}_{max,1h}$. In Figure 31, $\frac{\hat{x}_{max,1h,i}}{\hat{x}_{max,1h,j}}$ is plotted against $\frac{\mu_i}{\mu_j}$ for all eight sensors, where $\mu$ is the mean of the ice thickness when data sets with similar speed over ground are analyzed, and vice versa. The plots show great scatter, and for all four pairs of $i$ and $j$, $\frac{\hat{x}_{max,1h,i}}{\hat{x}_{max,1h,j}}$ occurs both above and below 1. However, for the pairs with similar speed over ground and different ice thicknesses (left plot), a trend of higher $\hat{x}_{max,1h}$ is found for thicker ice, as expected. The mean (X) of the values from the first pair (+) has a two-sided 95% confidence interval (CI) equal to [0.796 0.96], whereas the second pair (*) has a 95% CI for the mean of [0.95 1.45]. However, it is not possible to tell how the ice thickness is related to $\hat{x}_{max,1h}$ based on these data, other than that they are likely to be positively correlated. Contrary, no positive trend is found for the pairs with similar ice conditions and different speed over ground. In fact, the means of both pairs are very close to one, pointing towards that $\hat{x}_{max,1h}$ is independent of the vessel speed. For the first pair (*) on the right side figure, $\frac{\mu_i}{\mu_j}$ is close to one, thus it is expected that the mean is close to one, also if $\hat{x}_{max,1h}$ is dependent of the vessel speed. For the other pair (+), $\frac{\mu_i}{\mu_j} = 2.20$, and a mean(X) close to one indicates no correlation between $\hat{x}_{max,1h}$ and the vessel speed. It should be noted that the means (X) are associated with large uncertainties (95 % CI = [0.862 1.14], [0.739 1.24] for the first (*) and second (+) pair, respectively), thus no trend can be concluded.

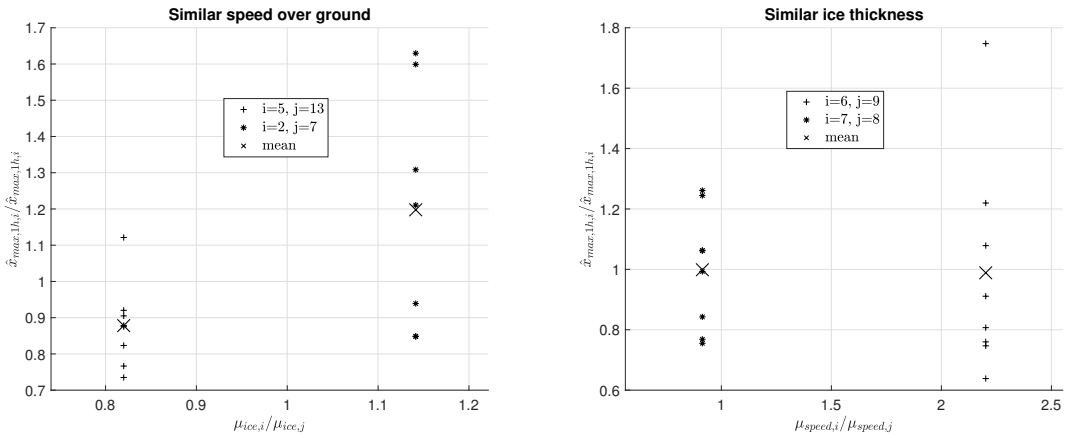
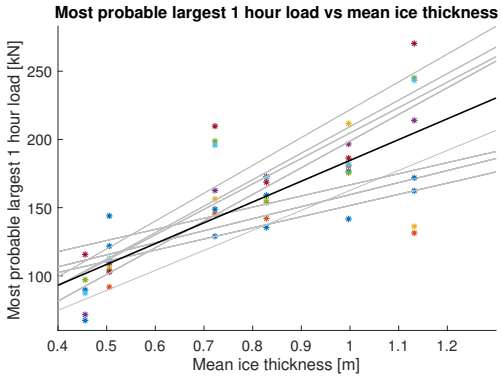
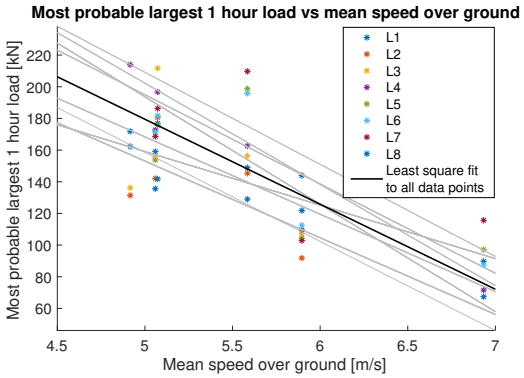



Figure 31: Similar speed over ground and different ice thickness (left), and vice versa (right). Showing data for all eight sensors. 95% confidence intervals for the means (X) are; [0.796 0.96], [0.95 1.45], [0.862 1.14] and [0.739 1.24] from left to right, respectively.

Figure 32 (a) shows $\hat{x}_{max,1h}$ plotted against $\mu_{ice}$ as a scatter plot for all sensors L1 to L8, for the time series in Table 1. Each scatter (*) represent a specific sensor for a specific data set. If a sensor has measured less than 25 loads above the applied threshold during the time series, its most probable largest value is not included in the plot. For each sensor, a least square line is fitted (grey lines). One can see that all the fitted lines show a significant trend of increasing $\hat{x}_{max,1h}$ for increasing ice thickness. It should be noted that other parameters, such as speed over ground and standard deviation of the ice thickness, may vary between the sets and affect $\hat{x}_{max,1h}$. A negative correlation is observed between $\hat{x}_{max,1h}$ and the speed over ground. As a first thought, this is counter-intuitive, but by looking at Figure 33, one can see that there is a significant negative correlation between the speed over ground and the ice thickness. This is probably due to increased total resistance when sailing in thicker ice, and it also explains why it is difficult to find time series with one of the parameters similar and the other different. In order to get an idea about how much of the trend in Figure 32 (b) that is due to the correlation in the data between speed over ground and ice thickness, a similar plot as in Figure 32 (a) is plotted using only the pairs from Table 20 (similar speed). A least square fitted line is plotted for each pair, providing an indication of the trend between $\hat{x}_{max,1h}$ and the ice thickness. The plot can be seen in Figure 34. A brief estimate of the expected correlation between the speed over ground and $\hat{x}_{max,1h}$ (Figure 32 (b)) due to the correlation between the vessel speed and the ice thickness (Figure 33) is obtained by dividing the slopes of the fitted lines from Figure 34 on the slope from Figure 33: $\frac{[142,121]}{-2.26} = [-63, -54]$. This is close to the actual slope in Figure 32 (b), which is -53 kN/(m/s), indicating that the vessel speed has small, or maybe even none, effect on $\hat{x}_{max,1h}$ in the evaluated range of vessel speeds. Few and widely scattered data are used for this brief estimate, thus it should be considered as an indication rather than a scientific estimate. However, the result is supported by the findings in Figure 31, which indicated that $\hat{x}_{max,1h}$ is independent of the vessel speed. Stationarity in terms of ice thickness is therefore preferred when selecting time series in this thesis.




(a) Least square fitted lines (grey) to $\hat{x}_{max,1h}$ estimated from each sensor for all the sets in Table 1 plotted against the mean ice thickness, $\mu_{ice}$, for the time series. The black line shows the least square fit to all of the extreme values.




(b) Least square fitted lines (grey) to $\hat{x}_{max,1h}$ estimated from each sensor for all the sets in Table 1 plotted against the mean speed over ground, $\mu_{speed}$, for the time series. The black line shows the least square fit to all of the extreme values.

Figure 32

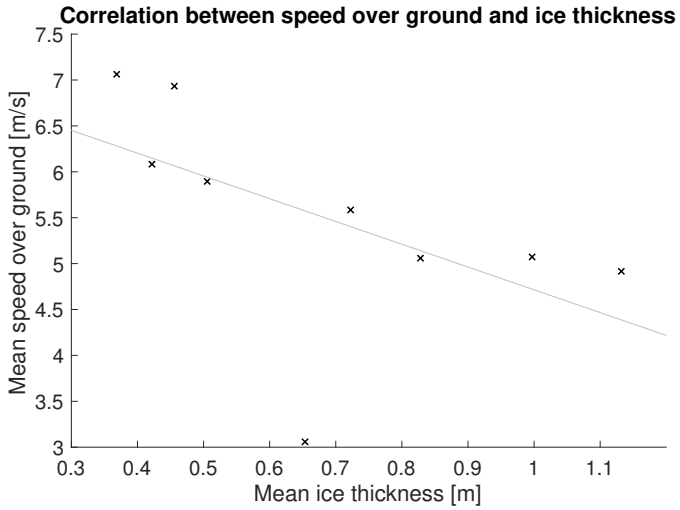



Figure 33: A significant negative correlation between speed over ground and ice thickness. The slope of the fitted line is -2.26 (m/s)/m.

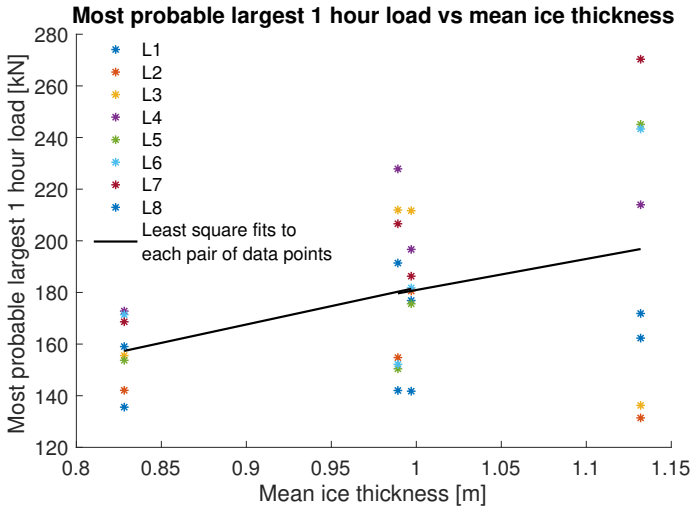



Figure 34: $\hat{x}_{max,1h}$ from two pairs with similar vessel speeds are shown. The first pair has mean ice thicknesses of 0.83 m and 1.00 m, and the second pair has mean ice thicknesses of 0.99 m and 1.13 m. A least square line is fitted for each pair. Their slopes are 142 kN/m and 121 kN/m, respectively.

## 7.2 Selection of Data Sets

For different time series and for different sensors, the measured loads can vary significantly in terms of both magnitude and load pattern. This can be observed by plotting different data sets in probability papers or Q-Q plots. Some of the differences may be due to the internal randomness of the ice-loading process, but in other cases, different underlying distributions govern the ice loads. For some data sets, more than one population seem to exist, while others have outliers. The selection of sets is therefore crucial for how well a method performs. In this study, the goal is to reveal both strengths and weaknesses of the applied methods for describing the load pattern, with a focus on extreme loads. Thus, data sets are selected with the intention of illustrating both strengths and weaknesses. This also means that the selected data sets are not a representative selection for all measurements during the expedition. However, load patterns that were found most frequently when working with the data are represented. A selection of sets used for comparing the different methods is given in Table 25. In addition, other sets may be applied for some of the methods. Relevant plots, e.g. selected probability papers, Q-Q plots, Kolmogorov-Smirnov tests and plots of ACER functions are presented. The figures and plots that are excluded from the main text can be found in Appendix C.

Table 25: Sets for comparison of extreme value estimation methods.

| Set | Sensor | Duration | Ice thickness | | | Number of loads | $y_{max}$ [kN] | Load pattern |
|---|---|---|---|---|---|---|---|---|
| | | [min] | $\mu$ | $\sigma$ [m] | RSD [%] | | | |
| 5 | L5 | 30 | 0.82 | 0.24 | 29.0 | 194 | 253.8 | Outlier |
| 6 | L5 | 15 | 0.51 | 0.14 | 28.4 | 55 | 71.2 | Two populations |
| 9 | L2 | 30 | 0.51 | 0.15 | 28.5 | 82 | 50.5 | Scattered |
| 14 | L1 | 25 | 1.04 | 0.20 | 19.2 | 103 | 140.5 | Stationary |

## 7.3 Classical Approach

Figures 35 to 38 show the sets given in Table 25 plotted in probability papers for exponential, log-normal, Gumbel and Weibull distributions. For the Gumbel distribution, the parameters are fitted based on a least square fitted line. The same method is applied for the exponential distribution, except that the fitted line is forced through the origin, since the exponential distribution by definition starts at the origin. It was experienced that the least square fitted lines gave better fit in the upper tails than when the maximum likelihood estimation (MLE) is applied when the x-axis is given in linear scale, which is the case for the exponential and Gumbel probability papers. MLE is applied for the log-normal distribution and the Weibull distribution. In their probability papers, the x-axis has a logarithmic scale, which can lead to very large scatter in the lower parts (see Figures 37 (b) and (d)). Thus, a least square fitted line will be significantly affected by the lower "outliers", which are solely dependent on the applied threshold on a very accurate level. The MLE is less sensitive to the "outliers" in the lower parts, and is therefore preferred for probability papers with logarithmically scaled x-axis. The selection of parameter estimation methods is made in order to get more accurate results for extreme loads. For other applications, e.g. for fatigue analyses, other methods may be preferred. Tables 26 to 29 present some key statistics related to the probability papers. Confidence intervals (CI) are given if there is a satisfactory visual match between the fitted distribution and the data set.

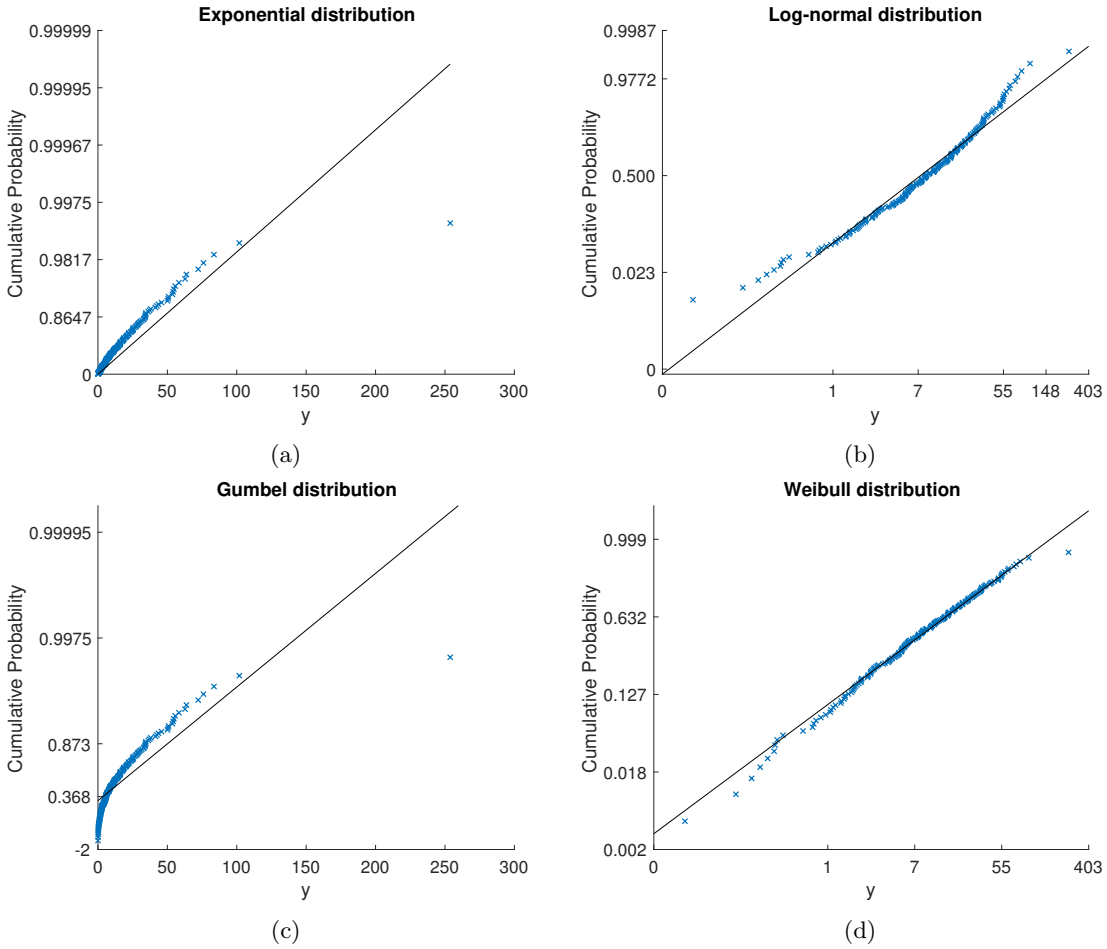



Figure 35: Probability papers for set 5, sensor L5.

Table 26: Set 5, sensor L5.

| Set 5, L5. $y_{max} = 253.8$ kN | Exponential | Log-normal | Gumbel | Weibull |
|---|---|---|---|---|
| Parameter | $\lambda = 0.043$ | $\sigma = 1.42$ | $\beta = 23.26$ | $k = 0.83$ |
| $CI_{\alpha=0.05}$ | | | | $0.75 - 0.93$ |
| Parameter | – | $\mu = 2.05$ | $\alpha = 3.59$ | $\theta = 15.10$ |
| $CI_{\alpha=0.05}$ | | | | $12.58 - 17.90$ |
| $\hat{y}_{max}$ [kN] | 123.5 | 294.8 | 126.1 | 111.0 |
| $R^2$ | 0.687 | 0.964 | 0.721 | 0.993 |
| KS-test, D | 0.189 | 0.065 | 0.307 | 0.030 |
| $D_{\alpha=0.05} = 0.098$ | Rejected | Not rejected | Rejected | Not rejected |

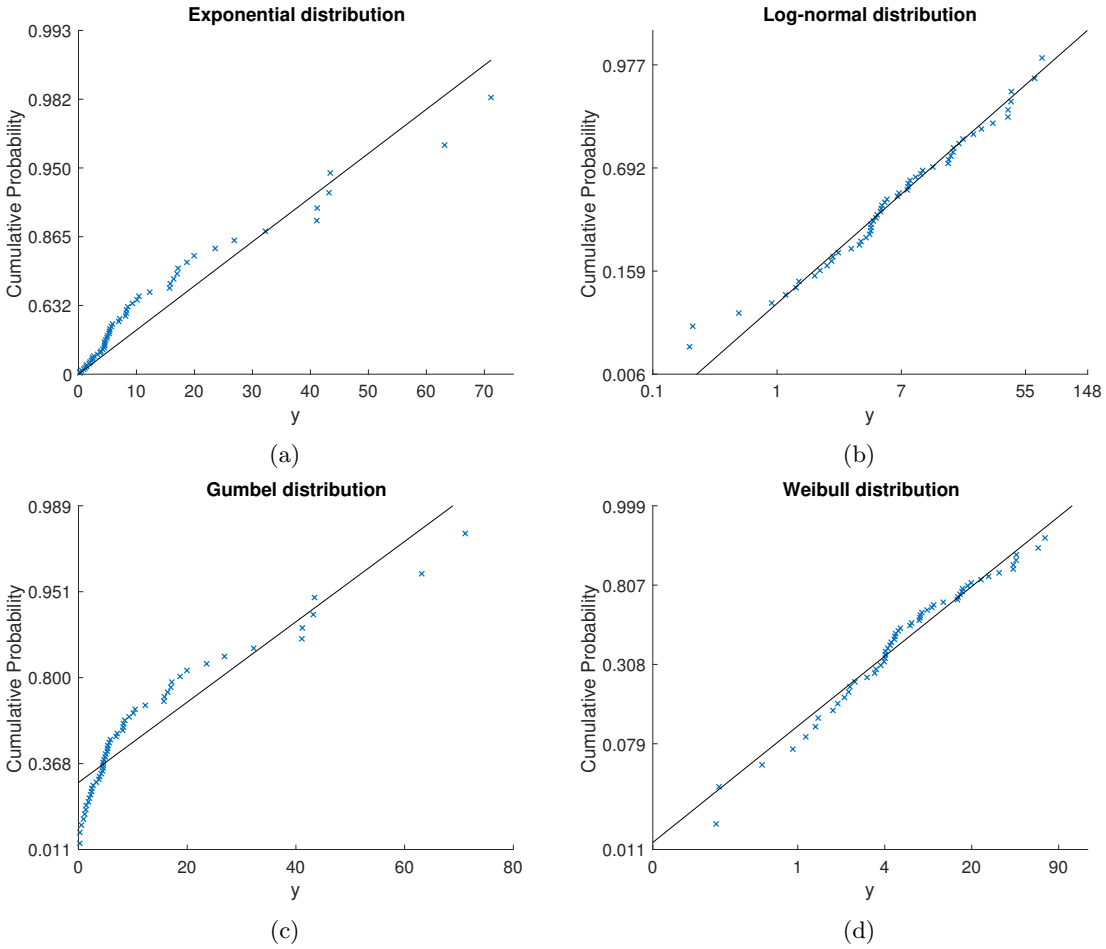



Figure 36: Probability papers for set 6, sensor L5.

Table 27: Set 6, sensor L5

| Set 6, L5. $y_{max} = 71.2$ | Exponential | Log-normal | Gumbel | Weibull |
|---|---|---|---|---|
| Parameter<br>$CI_{\alpha=0.05}$ | $\lambda = 0.064$ | $\sigma = 1.26$ | $\beta = 14.26$ | $k = 0.88$ |
| Parameter<br>$CI_{\alpha=0.05}$ | – | $\mu = 1.85$ | $\alpha = 4.71$ | $\theta = 11.71$ |
| $\hat{y}_{max}$ [kN] | 62.4 | 88.4 | 61.7 | 56.7 |
| $R^2$ | 0.896 | 0.971 | 0.862 | 0.972 |
| KS-test test, D<br>$D_{\alpha=0.05} = 0.183$ | 0.204<br>Rejected | 0.070<br>Not rejected | 0.237<br>Rejected | 0.098<br>Not rejected |

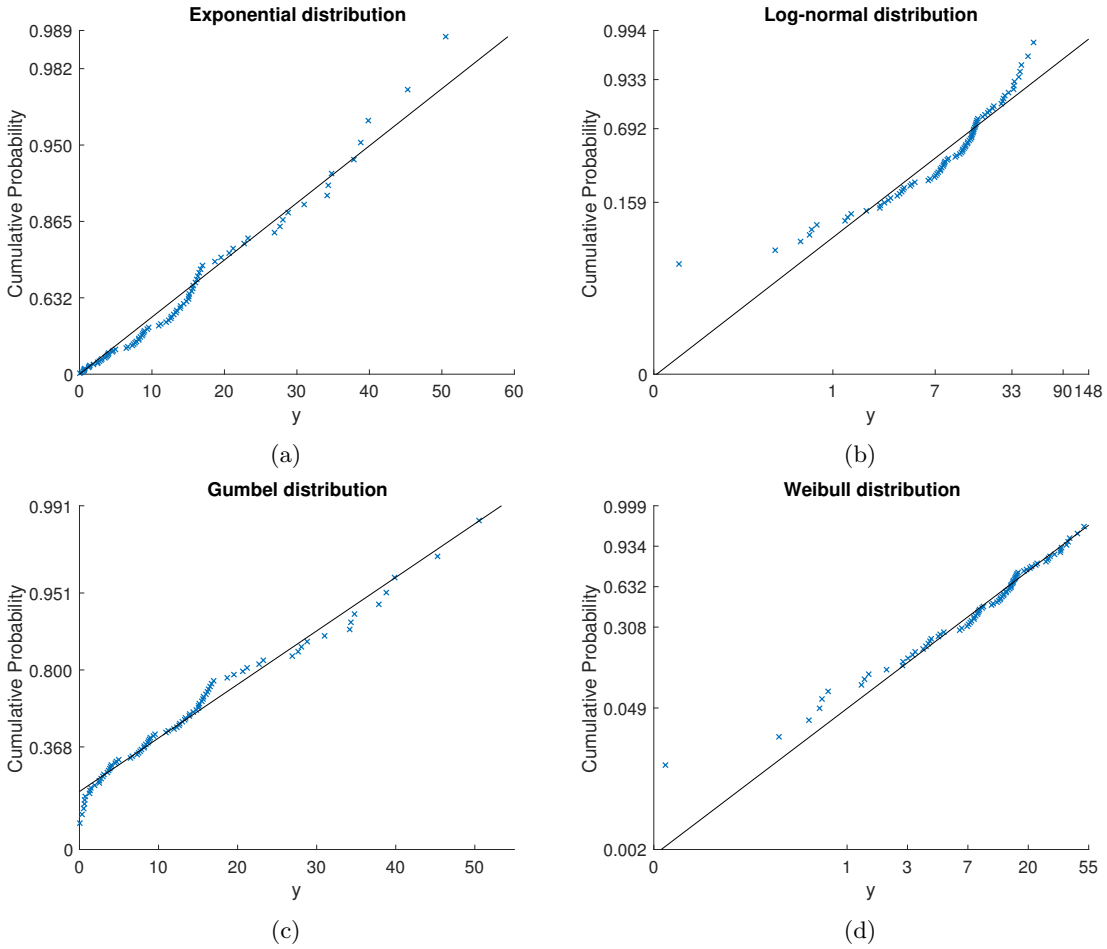



Figure 37: Probability papers for set 9, sensor L2.

Table 28: Set 9, sensor L2

| Set 9, L2. $y_{max} = 50.5$ | Exponential | Log-normal | Gumbel | Weibull |
|---|---|---|---|---|
| Parameter | $\lambda = 0.075$ | $\sigma = 1.24$ | $\beta = 9.58$ | $k = 1.13$ |
| $CI_{\alpha=0.05}$ | $0.055 - 0.090$ | | $8.21 - 12.98$ | $0.97 - 1.33$ |
| Parameter | – | $\mu = 2.12$ | $\alpha = 8.34$ | $\theta = 14.27$ |
| $CI_{\alpha=0.05}$ | | | $5.94 - 10.44$ | $11.63 - 17.25$ |
| $\hat{y}_{max}$ [kN] | 59.0 | 135.4 | 50.5 | 52.8 |
| $R^2$ | 0.976 | 0.873 | 0.978 | 0.938 |
| KS-test, D | 0.107 | 0.143 | 0.093 | 0.065 |
| $D_{\alpha=0.05} = 0.150$ | Not rejected | Not rejected | Not rejected | Not rejected |

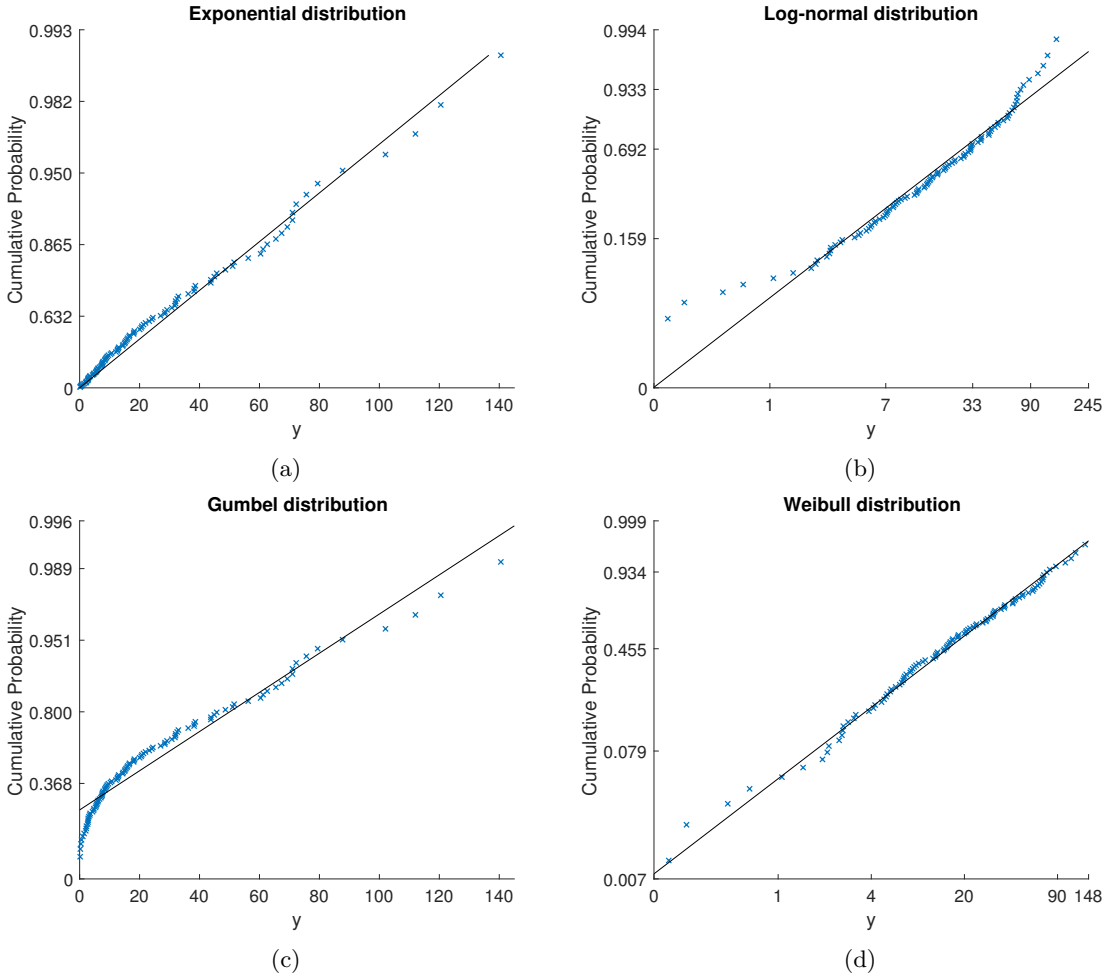



Figure 38: Probability papers for set 14, sensor L1.

Table 29: Set 14, sensor L1

| Set 14, L1. $y_{max} = 140.5$ | Exponential | Log-normal | Gumbel | Weibull |
|---|---|---|---|---|
| Parameter | $\lambda = 0.034$ | $\sigma = 1.33$ | $\beta = 24.37$ | $k = 0.93$ |
| $CI_{\alpha=0.05}$ | $0.026 - 0.040$ | | | $0.81 - 1.10$ |
| Parameter | – | $\mu = 2.65$ | $\alpha = 13.52$ | $\theta = 26.28$ |
| $CI_{\alpha=0.05}$ | | | | $20.79 - 32.40$ |
| $\hat{y}_{max}$ [kN] | 136.2 | 320.3 | 126.4 | 136.7 |
| $R^2$ | 0.987 | 0.950 | 0.943 | 0.993 |
| KS-test, D | 0.087 | 0.061 | 0.168 | 0.042 |
| $D_{\alpha=0.05} = 0.134$ | Not rejected | Not rejected | Rejected | Not rejected |

From Figure 35 one can see that the measured data from set 5, L5, are not well captured by any of the probability papers. Although the fitted Weibull distribution fits most of the data points, the estimated extreme value during the measurement is underestimated by 56%. The log-normal distribution is the only one that overestimates the extreme value, despite the outlier in the measurements. Although the distribution is not rejected by the Kolmogorov-Smirnov test, the data points are not scattered randomly around the fitted line, but tend to follow another curve, which indicates that the underlying distribution is not the log-normal. The Q-Q plot in Figure 39 illustrates the poor fit as well.

For set 6 L5 (Figure 36), the probability paper for the exponential distribution indicates that two populations exist, which is further discussed in Section 7.5. The fitted exponential distribution provides a poor fit, and is rejected by the KS-test. Both the log-normal and Weibull distribution give acceptable fits on their probability papers, but one must remember that their load axes are logarithmic. When the actual loads are read off, one can observe that the points are coarsely scattered in the log-normal probability paper, while the fitted Weibull distribution underestimates the values in the upper tale. This is easier to observe in Q-Q plots (Figure 40), which are more sensitive to deviations in the upper tail. It is also noted that the data points behave close to linearly in the Gumbel probability paper for values higher than approximately 10 kN.

Also for set 9 L2 (Figure 37), the data points exhibit linear behavior in the Gumbel probability paper, except for the lower tail. Despite the poor fit in the lower tail, the fit is good for $y > 5$ kN. The fitted Weibull distribution provides a good fit for the data (except for the lower tail), whereas the fitted exponential distribution somewhat overestimates the extreme value. It is noted that the fitted log-normal distribution, which provides a rather poor fit, is not rejected by the KS-test on a 95% significance level. This is discussed in Section 8.

The exponential distribution is able to capture the measured loads in a satisfactory way for set 14 L1 (Figure 38). Since the exponential distribution is a special case of the Weibull distribution, this also applies for the Weibull distribution. Set 14 was the most stationary of all identified sets, with a relative standard deviation (RSD) for the ice thickness of 19.2 %. This supports previous works (Kujala and Vuorio (1986), Suominen and Kujala (2010), Suyuthi et al. (2013b)), who found that the exponential and/or the Weibull distribution provide best fits to ice-loads when stationary conditions are strived for. Also for this set, it should be noted that the data points can be fitted by a linear line in the Gumbel probability paper above a certain load level.

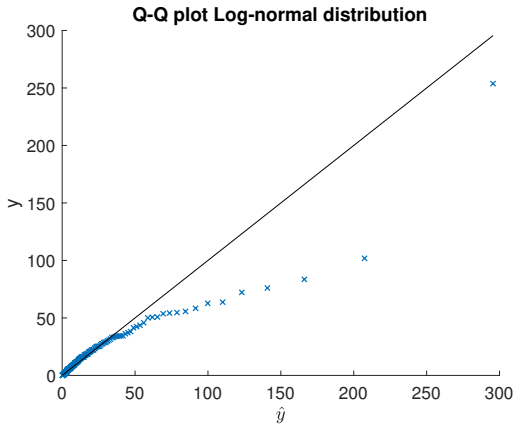
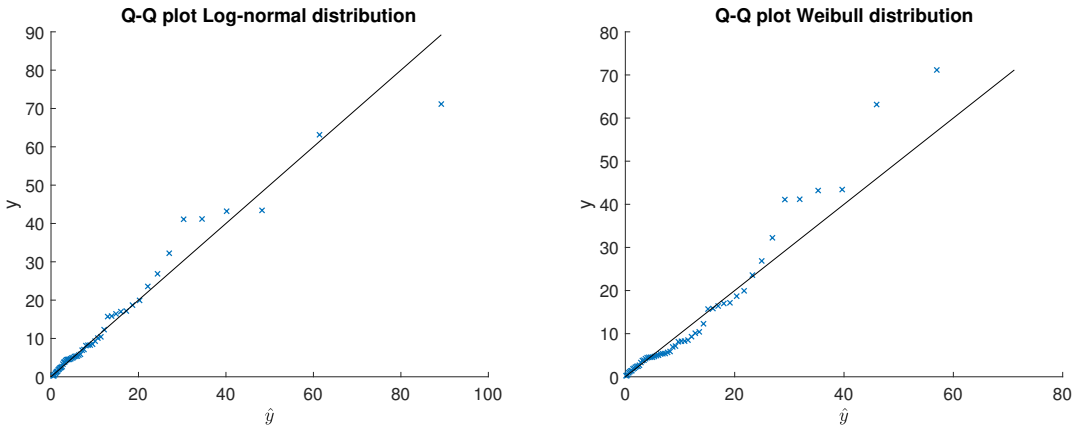



Figure 39: Q-Q plot for set 5, L5, for the log-normal distribution.



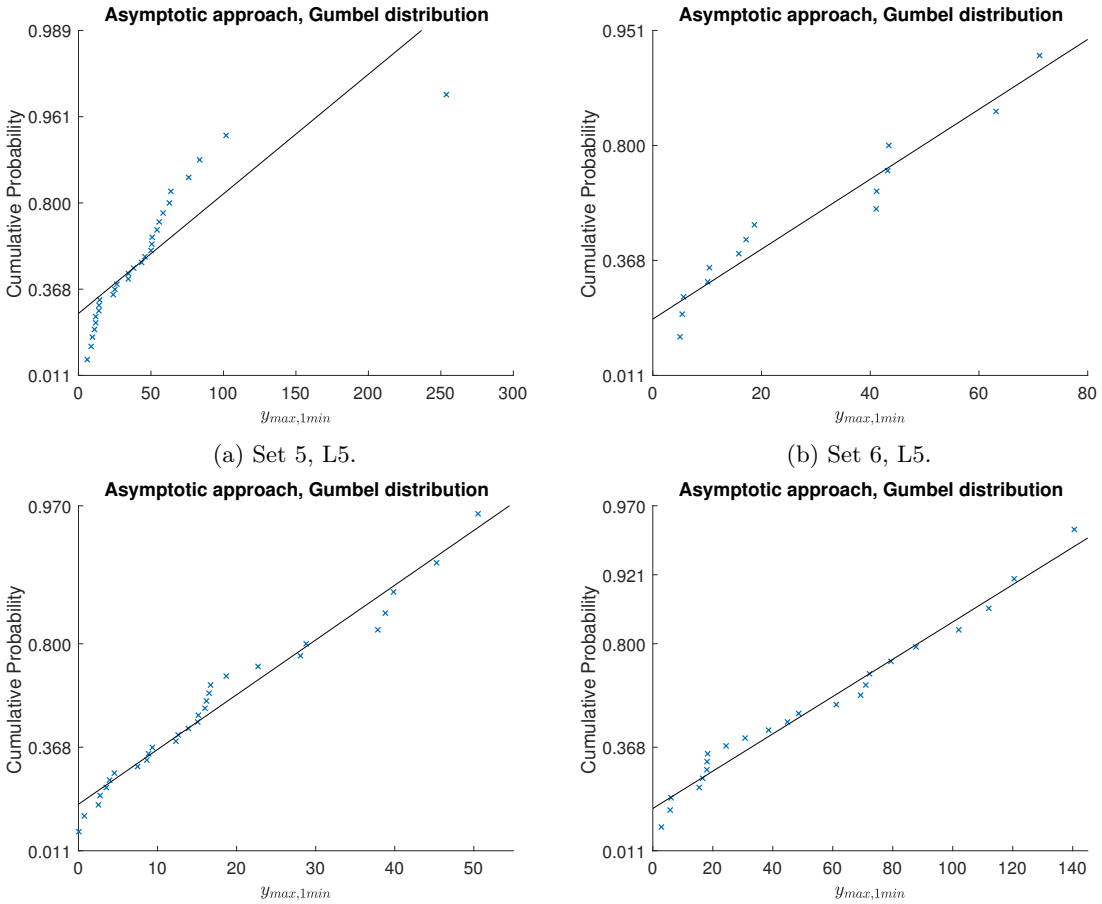

(a) Log-normal distribution.

(b) Weibull distribution.

Figure 40: Q-Q plots for log-normal and Weibull distributions for set 6, sensor L5.

## 7.4 Asymptotic Approach

Figures 41 (a) to (d) show the one minute maximums from the data sets given in Table 25 plotted in Gumbel probability papers. Even though the data sets become quite small (since only one load is included in the plotted data set for each minute), the fitted type I extreme value distributions, i.e., the Gumbel, provide satisfactory fits for three of the sets ((b) to (d)). As for the other methods, the outlier in set 5 L5 is not captured by the asymptotic approach.




(a) Set 5, L5.

(b) Set 6, L5.

(c) Set 9, L2.

(d) Set 14, L1.

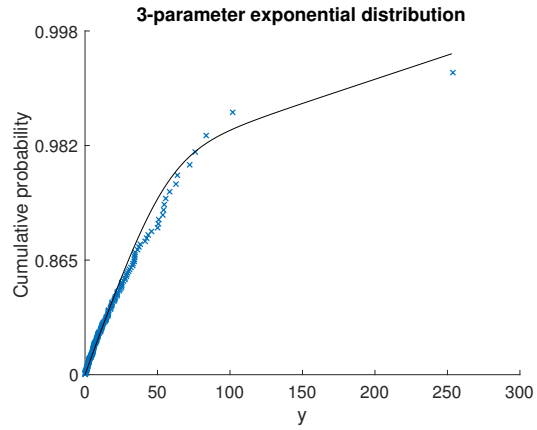
Figure 41: Fitted type I extreme value distribution to 1 min maximas plotted in Gumbel exponential probability papers.

Table 30: Set 14, sensor L1.

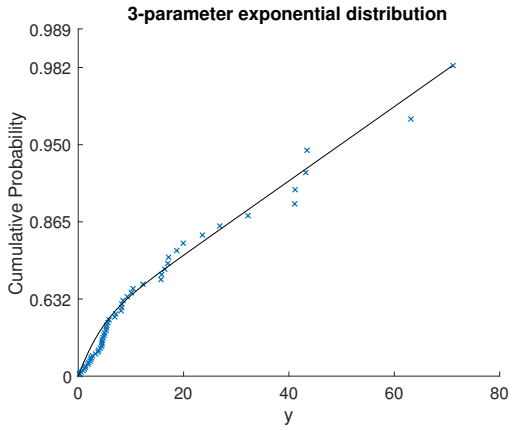
| Asymptotic approach | Set 5, L5 | Set 6, L5 | Set 9, L2 | Set 14, L1 |
|---|---|---|---|---|
| $\beta$ | 48.07 | 21.93 | 12.60 | 36.12 |
| $\alpha$ | 20.32 | 16.78 | 10.41 | 33.87 |
| $y_{max}$ [kN] | 253.8 | 71.2 | 50.5 | 140.5 |
| $\hat{y}_{max}$ [kN] | 181.3 | 73.8 | 52.6 | 147.9 |
| $R^2$ | 0.754 | 0.938 | 0.967 | 0.979 |
| Kolmogorov-Smirnov test, D | 0.227 | 0.133 | 0.122 | 0.106 |
| $D_{\alpha=0.05}$ | 0.246 | 0.354 | 0.246 | 0.274 |
| Rejected / Not rejected | Not rejected | Not rejected | Not rejected | Not rejected |

## 7.5 Three-Parameter Exponential Distribution

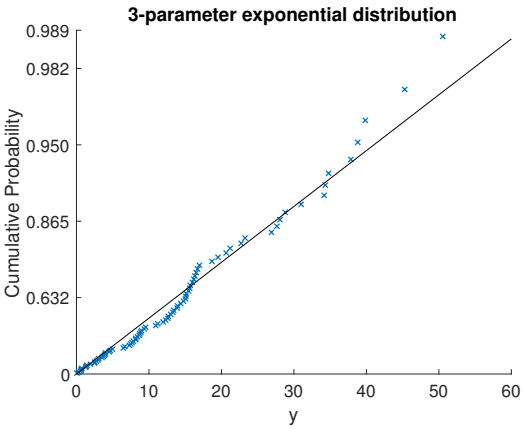
Fitted distributions and the data sets are plotted in exponential probability papers in Figure 42. The fit provided by the three-parameter exponential distribution is better for set 5 L5 than for the other methods. However, from Table 31 it can be seen that the estimated extreme value is 16% on the non-conservative side compared to the measured one, thus also this method fails to capture the outlier. Similar as for the asymptotic approach, and contrary to the classical approach, the three-parameter exponential distribution provides a good fit for set 6 L5. For set 9 L2, the weight parameter $a$ is 0, which means that the fitted distribution, in fact, is a one-parameter exponential distribution. The performance is therefore identical as found for the exponential distribution in Section 7.3, where the extreme value is overestimated by 19%. For set 14 L1, both the overall fit and the estimated extreme value are very good.



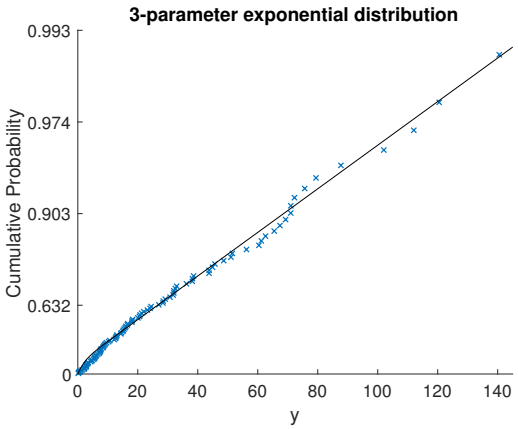

(a) Set 5, L5.




(b) Set 6, L5.




(c) Set 9, L2.




(d) Set 14, L1.

Figure 42: Fitted three-parameter exponential distributions plotted in exponential probability papers.

Table 31: Three-parameter exponential distribution.

| Three-parameter exponential | Set 5, L5 | Set 6, L5 | Set 9, L2 | Set 14, L1 |
|---|---|---|---|---|
| $\lambda_1$ | 0.0722 | 0.3445 | - | 0.5323 |
| $\lambda_2$ | 0.0084 | 0.0481 | 0.0731 | 0.0317 |
| $a$ | 0.969 | 0.455 | 0 | 0.1456 |
| $y_{max}$ [kN] | 253.8 | 71.2 | 50.5 | 140.5 |
| $\hat{y}_{max}$ [kN] | 213.2 | 70.7 | 60.3 | 141.2 |
| $R^2$ | 0.982 | 0.976 | 0.975 | 0.994 |
| Kolmogorov-Smirnov test, D | 0.051 | 0.153 | 0.100 | 0.084 |
| $D_{\alpha=0.05}$ | 0.098 | 0.183 | 0.150 | 0.134 |
| Rejected / Not rejected | Not rejected | Not rejected | Not rejected | Not rejected |

## 7.6 ACER Method

Comparison of ACER functions of different orders can be used as a tool to investigate the dependency of the data. The value of the ACER function is a function of how many of the measured data points that are larger than $\eta$, where the $k-1$ previous measured loads are smaller or equal to $\eta$. Thus, if the ACER functions of different orders are coalescing in the upper tail, it means that the highest loads are spread in time. In general, higher order of $k$ gives lower (or similar) value of the ACER function, since a higher order $k$ is a stricter requirement. ACER functions of order $k$ =1, ...,5 are plotted for set 14 L1 in Figure 43. The plot indicates that the data set is independent, at least for the upper tail. The same applies to the other sets investigated, which justifies the use of the first order ACER function for extreme value estimation.

The first order ACER functions and the extrapolation functions for the sets in Table 25 are plotted in Figure 44. It should be noted that the confidence interval (CI) for the ACER function stops earlier than the ACER function. For the largest loads in many of the data sets, the lower confidence band becomes negative, which does not make sense, nor is it possible to plot on a logarithmic scale. In addition, the weight factor $\rho_i$ used in Equation 30 becomes negative for the same high values of $\eta$, thus they are not used for estimating the extrapolation function. However, these data points are included in the figure for comparison with the extrapolation function. The optimized parameters for the extrapolations function are presented in Table 33. $\eta_{max}$ and $\hat{\eta}_{max}$ are also presented in the table. Note that $\eta_{max}$ and $\hat{\eta}_{max}$ are equivalent to $y_{max}$ and $\hat{y}_{max}$ corrected for the threshold used for the other methods. In addition, $\eta_0$ is given in the table, which is the lower load level from which the extrapolation function starts, selected by the user. In their study, Chai et al. (2018) found that the predicted extreme values were not very sensitive to the selection of $\eta_0$. This was the case for most, but unfortunately not all, of the data sets analyzed in this thesis. A sensitivity study was carried for set 5 L5 and set 9 L2. The results are presented in Table 32.

Table 32: ACER method.

| | | | | | |
|---|---|---|---|---|---|
| Set 5, L5 | $\eta_0$ [kN] | 40 | 45 | 50 | 55 |
| | $\hat{\eta}_{max}$ [kN] | 133.5 | 133.3 | 121.2 | 131.7 |
| Set 14, L1 | $\eta_0$ [kN] | 35 | 40 | 45 | 50 |
| | $\hat{\eta}_{max}$ [kN] | 95.9 | 143.6 | 75.5 | 73.9 |

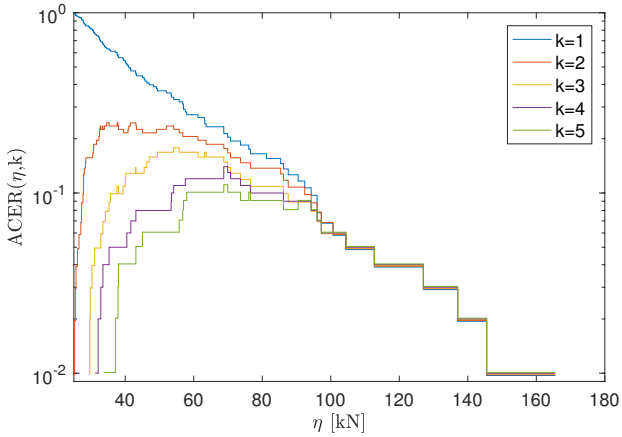
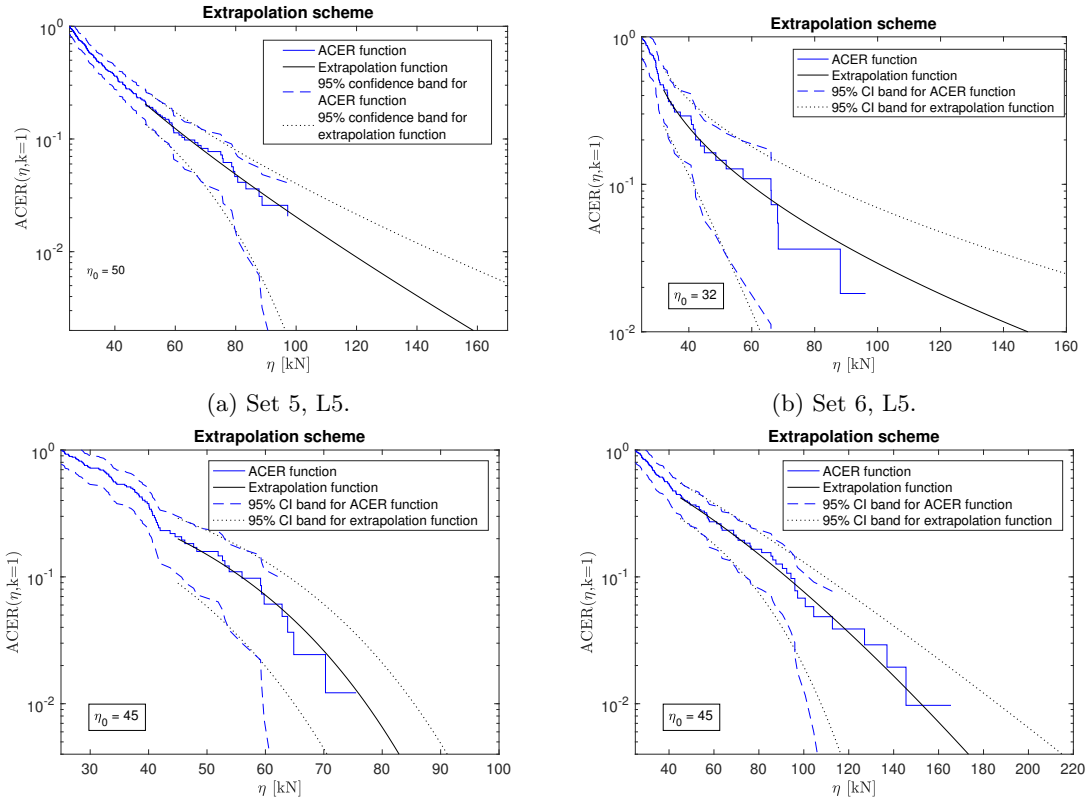



Figure 43: The ACER functions of different orders $k$ plotted for set 14 L1.




(a) Set 5, L5.

(b) Set 6, L5.

(c) Set 9, L2.

(d) Set 14, L1.

Figure 44: ACER functions with fitted extrapolation function for estimation of extreme loads.

Table 33: ACER method.

| ACER method | Set 5, L5 | Set 6, L5 | Set 9, L2 | Set 14, L1 |
|---|---|---|---|---|
| $\eta_0$ [kN] | 50 | 32 | 45 | 45 |
| $a_1$ | 0.123 | 0.377 | 0 | 0.004 |
| $b_1$ | 25.91 | 27.48 | -1425 | -2.45 |
| $c_1$ | 0.802 | 0.522 | 48.40 | 1.415 |
| $q_1$ | 1.00 | 1.00 | 1.00 | 1.00 |
| $\eta_{max}$ [kN] | 278.8 | 96.2 | 75.5 | 165.5 |
| $\hat{\eta}_{max}$ [kN] | 133.3 | 115.9 | 75.5 | 151.8 |

# 8 Discussion

The findings regarding the effects of ice thickness and vessel speed on the estimated extreme values are partly in line with what one can expect. As discussed in Section 3.2, the ice thickness is important for the bending capacity of the ice. However, the critical vertical force that is required to break level ice by bending, $F_{V,crit}$, is proportional to the ice thickness squared (Riska, 2017e), whereas $\hat{x}_{max,1h}$ seems to vary linearly with the ice thickness for the analyzed data. Also, contrary to what is indicated by the analyses above, one could expect that increased hydrodynamic support forces from the water due to greater accelerations will result in a positive correlation between the vessel speed and $\hat{x}_{max,1h}$. On the other hand, higher indention rates are associated with lower total forces (although higher local pressures occur), see Section 3.1. In addition, previous works (Jones (1997), Schulson and Duval (2009), and more) have showed that the indention rate have great impact on both tensile and bending strength, see Section 2.3. However, the ice-loading process is fairly complex, and many effects that are not considered here may have an impact on the relation between extreme loads and ice thickness and vessel speed. In order to investigate these effects in detail and provide good estimates for their correlation with the ice loads and its extreme values, more data is needed. Nevertheless, there are strong indications of that the ice thickness is a much more important parameter for the ice loads than the vessel speed. Stationarity in terms of ice thickness is therefore preferred when selecting time series in this thesis.

As observed in Section 7, the applied methods for estimating extreme loads provide various results for the same data sets. For the most stationary data set, set 14 L1, the exponential and the Weibull distributions are able to fit the data in the whole range of measured loads. This supports previous works (Kujala and Vuorio (1986), Suominen and Kujala (2010), Suyuthi et al. (2013b)), who found that the exponential and/or the Weibull distribution provide best fits to ice loads when stationary conditions are strived for. Also for set 9 L2, the fit is good (for y $>$ 3 kN) for the Weibull distribution. The same partly applies for the exponential distribution (Figure 37), although the highest loads are somewhat overestimated. For the classical approach, other combinations of data sets and distributions are not found to give satisfactory results for both overall fit and in terms of extreme value prediction. Despite this, 11 of 16 of the fitted distributions are not rejected by the Kolmogorov-Smirnov test (KS-test) for $\alpha = 0.05$. As an example, the log-normal distribution is not rejected for set 9 L2 (Figure 37 (b)) although the fit is obviously bad, particularly in the upper tail. The poor fit is confirmed in the Q-Q plot in Figure 45. All the cumulative distribution functions for the distributions applied in the analyses are steep for low values of y, and almost flat in the upper tail. Since the KS-test evaluates the difference in cumulative probability for the same value y, and not the other way around, the test is not sensitive to outliers in the upper tail. This is exemplified by the KS-test for the data set and fitted distribution shown in Figure 35 (d). The measure D (which is very low for this example) is not affected by the extreme value, which is underestimated by 56 %. The KS-test is seen in Figure 46.

Figure 38 (c) illustrates one of the disadvantages of the classical approach for extreme value applications. The data points exhibit close to linear behavior for y $>$ 20. However, the vast amount of data points in the lower load range has a substantial impact on the fitted distribution compared to the few loads in the upper load range. As discussed in Section 7.3, this effect can be reduced to some extent by selecting the method for parameter estimation with care, but the effect will still be an issue. When plotted in Gumbel probability paper, the data points show linear behavior above a certain, although not constant, value y for most of the data sets.

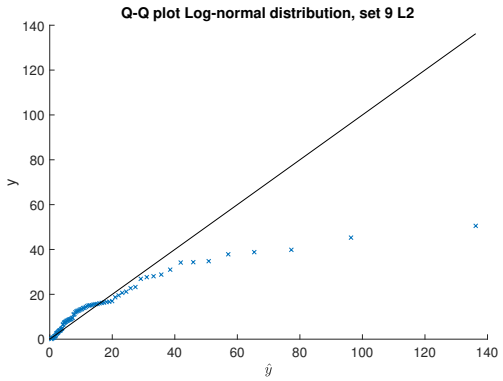



Figure 45: Q-Q plot for set 9 L2 for log-normal distribution.

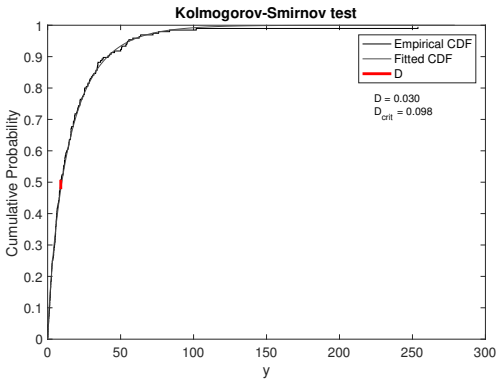
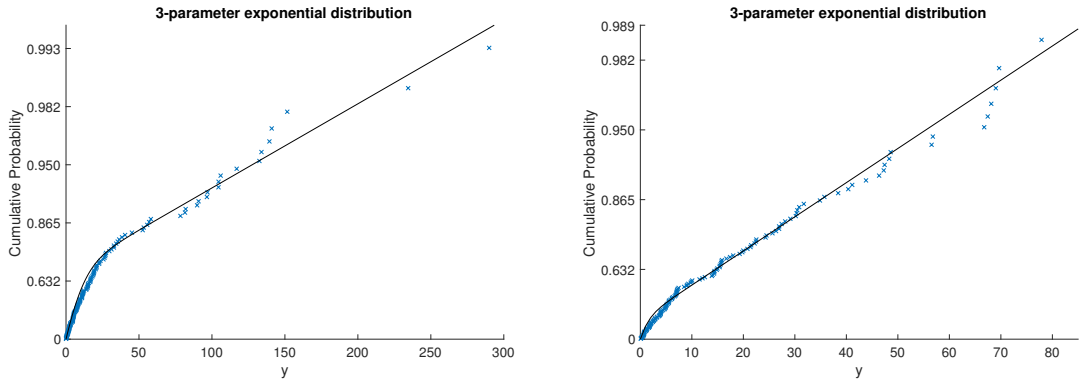



Figure 46: KS-test ($\alpha = 0.05$) for the fitted Weibull distribution for set 5 L5.

This may explain why the asymptotic approach provides good fits and extreme value estimates for the data sets (except for set 5 L5, which in fact is not well fitted by any of the methods). When only the one-minute maximums are included, most of the lower loads are excluded. If the highest loads are spread randomly in time, which yields for a stationary process, most of them are likely to be included in the analysis. However, randomness will ensure that some low loads are included, and some high loads are excluded. Thus, it is not the same as applying a higher threshold, which doesn't give satisfactory results. In this thesis, the asymptotic approach is found to be more robust in the sense of being able to predict extreme values for a wider range of load patterns and underlying distributions compared to the classical approach and the ACER method. This was also the impression when other sets from Table 1 were analyzed.

The three-parameter exponential distribution also avoids the problem with the lower loads having a large effect on the fit in the upper tail. As seen in Figure 42, the method is able to capture a bend in the load pattern when plotted in an exponential probability paper, and the estimated extreme values are close to the measured ones (except for set 5 L5). It should be noted the fitted curve cannot have a positive second derivative, i.e., the slope must be constant or decreasing for increasing values of y. In Figure 42 (c) an "ideal" fit will have a positive second derivative. The consequence is that the weight parameter $a$ becomes zero, meaning that the fitted distribution, in fact, is a one-parameter exponential distribution. However, this ensures that the fitted three-parameter exponential distribution is never less conservative than the fitted one-parameter exponential distribution. From a design point of view, there may be a challenge to generalize the method. Linking three different parameters to physical quantities (e.g. level ice thickness, vessel speed, the presence of ice ridges, ice features, etc.) is difficult. For such a task one needs to know which phenomena that are causing the different populations. In Figure 47 two different types of two-populated data sets are shown, resulting in very different distribution parameters.

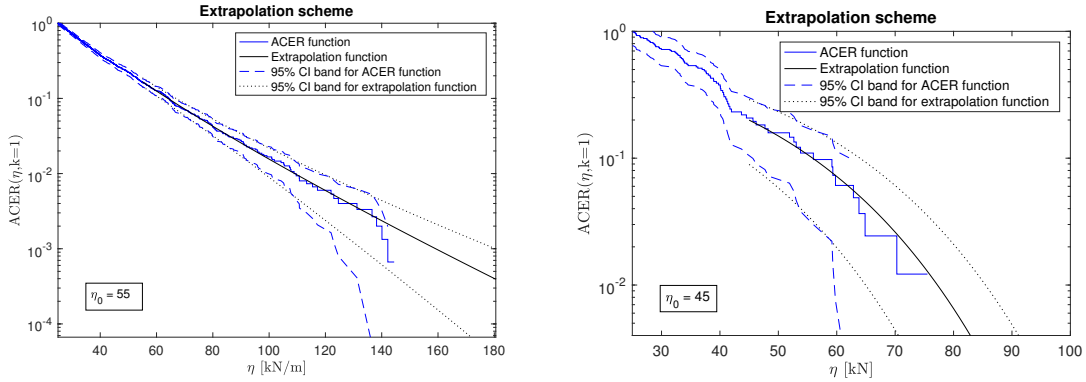



(a) Set 2 L7

(b) Set 9 L5

Figure 47: Fitted three-parameter exponential distribution. $[a, \lambda_1, \lambda_2] = [0.766, 0.049, 0.248]$ and $[0.1456, 0.0145, 0.6846]$ for (a) and (b), respectively.

Various results are obtained by the ACER method, and the results are found to be highly dependent on the tail maker $\eta_0$. This is an issue because $\eta_0$ has to be chosen, and is therefore a somewhat arbitrary value. The findings are in contrast to what was found by Chai et al. (2018): "*... the predicted value is not very sensitive to the tail maker $\eta_0$, provided it is chosen with some care*". The main difference between the analyses performed by Chai et al. (2018) and the analyses in this thesis is the duration of the time series. Chai et al. (2018) applied time series with duration of 6 hours, containing 545 and 1501 data points. The time series used in this thesis are between 15 and 30 minutes, and contain correspondingly fewer data points. Few data points give an irregular ACER function, whereas more data points will result in a smoother ACER function. Figures 48 (a) and (b) show typical ACER functions for small and large data sets. In the fitting process, the weight factor $\rho_i$, given in Equation 31, tends to be strongly decreasing for increasing $\eta$, thus the fitted curve is mainly affected by the data points in a relatively short range after $\eta_0$. With Figure 48 in mind, this explains why Chai et al. (2018) found that the fitted extrapolation function was not very sensitive to $\eta_0$, whereas this thesis came to the opposite conclusion.




(a) Data set used by Chai et al. (2018).

(b) Set 9 L2.

Figure 48: Difference between ACER function for a large (a) and a small (b) data set.

Since the ACER method is found to be inadequate for the data sets applied for the other methods (due to short duration), a brief comparison study is performed between the asymptotic approach and the ACER method for the same time series as applied by Chai et al. (2018). A comparison is performed for all load sensors L1 to L8, and the results are found in Table 34 below. If the sensors L4 and L5 are ignored (due to outliers), the asymptotic approach using $\Delta t = 5$ min gives the best fit for four of six data sets, and the remaining two are fitted best when $\Delta t = 1$ min. The ACER method underestimates the extreme value by more than 9% for four of six data sets. In comparison, the asymptotic approach using $\Delta t = 5$ min underestimates the extreme value by more than 2.5% only once. It should also be noted that the estimated extreme value by the asymptotic approach applying $\Delta t = 1$ min is on the conservative side for only one of the sets. To summarize the brief study; the asymptotic approach applying $\Delta t = 5$ min provides a better fit for these data sets than the ACER method.

Table 34: Comparison between ACER method and the asymptotic approach using type I extreme value distribution (Gumbel). ** indicates that the largest load in the data set was a significant outlier.

| Sensor | | | L1 | L2 | L3 | L4 | L5 | L6 | L7 | L8 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\eta_{max}$ [kN] | | | 141 | 148 | 147 | 210** | 224** | 193 | 144 | 165 |
| $\hat{\eta}_{max}$ [kN] | Asymptotic | $\Delta t = 1min$ | 132 | 131 | 145 | 159 | 165 | 159 | 160 | 118 |
| | | $\Delta t = 5min$ | 140 | 148 | 163 | 189 | 190 | 188 | 170 | 142 |
| | ACER | $\eta_0 = 55$ kN | 128 | 128 | 158 | 145 | 202 | 170 | 168 | 130 |

None of the methods are able to estimate the outliers. This was expected, which is the reason why it is included in the study. Outliers were identified in several data sets, and their presence cannot be ignored. Assumed that the outliers are valid measurements, extreme load estimation is difficult by application of the methods applied in this thesis, at least if not a significantly larger amount of data are analyzed. One of DNV's superior goals for the ILM project was to develop models for live estimation of loads, mainly for navigation purposes. Such a model is impossible to develop if

unpredictable outliers occur from time to time. In Figure 49, a plot of 10 000 bootstrapped samples based on the Weibull distribution fitted to set 5 L5 (estimated by MLE) is shown together with the measured data. The outlier from the data set (the top blue star) is exceeded 36 times by the bootstrapped data points (yellow stars), with the highest bootstrapped load being 56% higher than the outlier. Thus, it is not impossible to have such a high load based on the Weibull distribution. On the other hand, we know that it is not correct to assume that the data points can be explained solely by a Weibull distribution, due to non-stationary conditions. For further work, it should be addressed whether the outliers are actual loads or not. If they appear to be actual loads, it is, particularly from a navigation point of view, valuable to identify which impacts that are causing them.

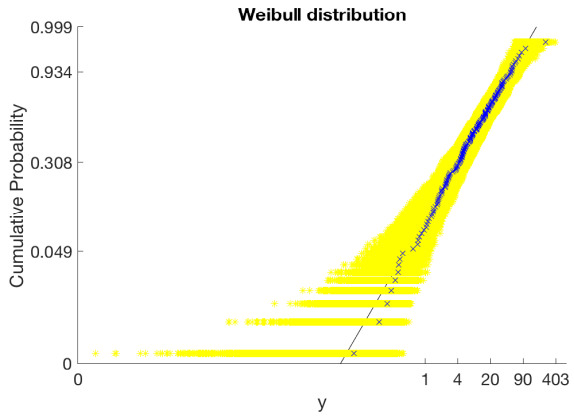



Figure 49: 10 000 samples (yellow stars) bootstrapped to the fitted Weibull distribution. The blue stars represent the data set.

# 9 Conclusion

Few adequate data sets were identified and used for the study on the effects of ice thickness and vessel speed on the estimated extreme values. The results from the study are widely scattered, particularly for the vessel speed. However, there are strong indications of that the ice thickness has a great impact on the estimated extreme value, whereas no correlation is found for the vessel speed.

Most of the applied methods provide good fits and estimates of extreme values if the data sets are selected with some care. Thus, the selection of method should depend on what kind of data that is being analyzed, and what the purpose of the application is. For stationary ice conditions, the exponential and Weibull distributions provide good fits to the data. However, stationary ice conditions are rarely encountered in the Arctic Oceans, which makes these methods unsuitable for live estimation of extreme loads. On the other hand, they may be adequate for design purposes, such as predicting loading due to level ice.

Contrary to the exponential and Weibull distributions, the asymptotic approach and the three-parameter exponential distribution provide good fits to data sets where two populations seem to exist, which was found for several data sets. They also performed well for the stationary data sets, making them more robust than the exponential and Weibull distributions.

The ACER method has several advantages compared to traditional methods, as described in Section 5.6.2. However, the user must select a lower load level, $\eta_0$, from which the extrapolation function is fitted. It was found that the predicted extreme value is very sensitive to $\eta_0$ for small data sets (15-30 min), which made it difficult to compare the ACER method to the other approaches, as stationary conditions are not obtained for long time series. However, the sensitivity to $\eta_0$ decreases for increasing data set sizes. For non-stationary time series of 6 hours, the asymptotic approach provided more accurate and conservative estimates of extreme values than the ACER method.

# 10 Recommendations of Further Work

During the work with this thesis, several tasks have arisen that have not been investigated in detail due to limited data sets or lack of information, time or knowledge. The tasks are presented below.

**Investigate which properties and conditions that affect the ice loading.**

It was discovered that previous works (e.g. Suyuthi et al. (2013b) Suyuthi et al. (2012b), Chai et al. (2018), Kujala and Vuorio (1986)) applied different requirements to identify stationary time series. Some of the requirements are partly in contrast to the theory found in the literature as well. This is briefly discussed in Section 5.1, and a brief study on effects of ice thickness and vessel speed is carried out in Section 7.1. The results indicated that the ice thickness has a great impact on the estimated extreme values, while no correlation was found for the vessel speed. However, very limited appropriate data were found, thus further work on the subject is needed. Other physical properties, conditions or characteristics may also be investigated.

**Relate physical quantities to probabilistic methods.**

If the previous topic is addressed, it is useful to relate physical quantities to the probabilistic methods, i.e., the distribution parameters. This is particularly interesting from a design/class point of view, both related to fatigue and extreme loads.

**Linear fit in the upper tail in Gumbel probability paper.**

J. Jordaan et al. (1993) suggested that in most cases, ice loads exhibit linear behavior in the tail region when plotted in exponential probability paper. The data sets in figures 36 (c), 37 (c) and 38 (c) show similar behaviour in Gumbel probability papers. The linear behavior in the tail may be the explanation for why the asymptotic approach gave good results, but this relation is not investigated in this thesis.

**Investigate outliers.**

Several outliers were found among the measured loads on KV Svalbard. Whether they are actual loads or measuring errors is crucial to know to be able to develop good extreme load estimation models, both for live estimation and for use in design. Thus, it is recommended that a study is carried out to clarify the issue.

# References


Alexander, B., & Alexander, C. (n.d.). Nature Picture Library. Retrieved from http://www.arkive.org/eco-regions/arctic/image-H205 (accessed: 12 September 2017)

*Arctic vs. antarctic.* (2017). Retrieved from https://nsidc.org/cryosphere/seaice/characteristics/difference.html (accessed: 12 September 2017)

C. Davison, A., & Hinkley, D. (1997, 01). Bootstrap methods and their application. , *94*.

Chai, W., et al. (2018). Probabilistic methods for estimation of the extreme value statistics of ship ice loads. *Cold Regions Science and Technology*, *146*, 87-97. doi: 10.1016/j.coldregions.2017.11.012

Eriksson, P., et al. (2007). *Description and operative forecasting of compression in an ice field.* (Research Report No 59). Winter Navigation Research Board. Retrieved from https://www.trafi.fi/filebank/a/1352716465/d61cba0efa01452ec404778d20ade75d/10737-Report_No_59_Ships_in_Compressive_Ice.pdf (accessed: November 11 2017)

Espeland, (2008). *Ice action and response monitoring of ships.* Norwegian University of Science and Technology, Faculty of Engineering Science and Technology, Department of Marine Technology.

Gaidai, O., et al. (2016). Extreme value statistics of large container ship roll. *Journal of Ship Research*, *60*, 92-100. doi: 10.5957/JOSR.60.2.150070

Gumbel, E. J. (1958). *Statistics of extremes.* Columbia University Press.

Haver, S. (2017). *Professor 2.* Said during lecture held spring 2017 in the course "Design of Offshore Structures" at NTNU.

*Ice crystal structure.* (2017). Retrieved from http://www.iceandclimate.nbi.ku.dk/research/flowofice/ice_crystal_structure/ (accessed: 22 October 2017)

Jeffries, M., et al. (2013). The arctic shifts to a new normal. *Physics Today*, *66*, 35. doi: 10.1063/PT.3.2147

Johnsten, M., et al. (2009). Multi-year ice thickness: knowns and unknowns. *Proceedings of the 20th International Conference on Port and Ocean Engineering under Arctic Conditions.* Retrieved from http://nparc.cisti-icist.nrc-cnrc.gc.ca/eng/view/accepted/?id=15b81ab8-4582-4272-9f8c-ff99f9c09cba

Jones, S. J. (1997). High strain-rate compression tests on ice. *The Journal of Physical Chemistry*, *August 1997*. doi: 10.1021/jp963162j

Jordaan, I. J. (2001). Mechanics of icestructure interaction. *Engineering Fracture Mechanics*, *68*, 1923-1960.

Jordaan, J., et al. (1993). Probabilistic analysis of local ice pressures. *Journal of Offshore Mechanics and Arctic Engineering*, *115*, 83-89.

Karpa, O., & Naess, A. (2013). Extreme value statistics of wind speed data by the acer method. *Journal of Wind Engineering and Industrial Aerodynamics*, *112*, 1-10. doi: 10.1016/j.jweia.2012.10.001

Kujala, P., et al. (2009). Statistics of ice loads measured on mt uikku in the baltic. *In Proceedings of the 20th Interna- tional Conference on Port and Ocean Engineering under Arctic Conditions..*

Kujala, P., & Vuorio, J. (1986). *Results and statistical analysis of ice load measurements on board icebreaker sisu in winters 1979 to 1985.* Winter Navigation Research Board.

*Kv svalbard.* (2014). Retrieved from https://forsvaret.no/fakta/utstyr/Sjoe/KV-Svalbard (accessed: September 7 2017)

Langleben, M. P., & Pounder, E. R. (1961). *Elastic parameters of sea ice.* Defense Technical Information Center.

Leira, B. J., & Børsheim, L. (2008). Estimation of ice loads on a ship hull based on strain measurements. *Proceedings of the ASME 27th International Conference on Offshore Mechanics and Arctic Engineering.* (Estoril, Portugal)

Lensu, M., & Manninen, S. (2003). Short term monitoring of ice loads experienced by ships. *17th International Conference on Port and Ocean Engineering under Arctic Conditions (POAC'03).* (Trondheim, Norway)

Masterson, D. M., & Frederking, R. M. W. (1993). Local contact pressures in ship/ice and structure/ice interactions. *Cold Regions Science and Technology*, *January 1993*, 169-184. doi: 10.1016/0165-232X(93)90005-S

Naess, A., & Gaidai, O. (2009). Estimation of extreme values from sampled time series. *Structural Safety*, *31*, 325-334. doi: 10.1016/j.strusafe.2008.06.021

O'Rourke, B. J., et al. (2015). *Experimental investigation of oscillation of loads in ice high-pressure zones, part 1: Single indentor system, page 26.* Cold Regions Science and Technology.

Riska, K. (2017a). Li general. *Ship design for ice operations.* (distributed by email)

Riska, K. (2017b). Lii ice conditions and sea areas. *Ship design for ice operations.* (distributed by email)

Riska, K. (2017c). Lix calculation of ice action. *Ship design for ice operations.* (distributed by email)

Riska, K. (2017d). Lix ice loading. *Ship design for ice operations.* (distributed by email)

Riska, K. (2017e). *Professor 2.* Said during lecture held fall 2017 in the module "Ship Design for Ice" at NTNU.

*Salinity and brine.* (2017). Retrieved from `https://nsidc.org/cryosphere/seaice/characteristics/brine_salinity.html` (accessed: 14 September 2017)

Sammonds, P., et al. (1999). Experimental and theoretical fracture mechanics applied to antarctic ice fracture and surface crevassing. *Journal of Geophysical Research Atmospheres*, *February 1999*, 2973-2986. doi: 10.1029/1998JB900026

Sanderson, T. J. (1986). A pressure-area curve for ice. *Proc 8th IAHR International Symposium on Ice*, 361-384. (Iowa City, Iowa)

Sanderson, T. J. (1988). *Ice mechanics and risks to offshore structures.* Springer Netherlands.

Schapery, R. A. (1997). Linear elastic and viscoelastic deformation behavior of ice. *Journal of cold regions engineering*, *Vol. 11, Issue 4 (December 1997).* doi: 10.1061/(ASCE)0887-381X(1997)11:4(271)

Schulson, E. M., & Duval, P. (2009). *Creep and fracture of ice.* Cambridge University Press.

Sinha, N. K. (1989). Elasticity of natural types of polycrystalline ice. *Cold Regions Science and Technology*, *17*, 127-135. doi: 10.1016/S0165-232X(89)80003-5

Soares, C. G., & Garbatov, Y. (2015). *Proceedings of the 19th international ship and offshore structures congress.* Taylor & Francis Group.

Sodhi, D. S. (n.d.). *Crushing failure during ice-structure interaction.* Retrieved from `http://virtual.vtt.fi/virtual/proj6/arki/julkaisut/strice/Sodhi.pdf` (accessed: 28 October 2017)

Suominen, M., & Kujala, P. (2010). *Analysis of short-term ice load measurements on board ms kemira during the winters 1987 and 1988.* Aalto University, School of Science and Technology, Department of Applied Mechanics.

Suyuthi, A., et al. (2012a). Non-parametric probabilistic approach of ice load peaks on ship

hulls. *Proceedings of the 31st International Conference on Offshore Mechanics and Arctic engineering*. (Rio de Janeiro, Brazil)

Suyuthi, A., et al. (2012b). Short term extreme statistics of local ice loads on ship hulls. *Cold Regions Science and Technology*, *82*, 130-143. doi: 10.1016/j.coldregions.2012.05.017

Suyuthi, A., et al. (2013a). *Prediction of extreme loads and fatigue damage for a ship hull due to ice action*. Norwegian University of Science and Technology, Faculty of Engineering Science and Technology, Department of Marine Technology.

Suyuthi, A., et al. (2013b). Statistics of local ice load peaks on ship hulls. *Structural Safety*, *40*, 1-10. doi: 10.1016/j.strusafe.2012.09.003

Suyuthi, A., et al. (2014). A generalized probabilistic model of ice load peaks on ship hulls in broken-ice fields. *Cold Regions Science and Technology*, *97*, 7-20. doi: 10.1016/j.coldregions .2013.09.012

Timco, G., & Frederking, R. (1983). Confined compressive strength of sea ice. *The Seventh Internationas Conference on Port and Ocean Engineering under Arctic Conditions*. (Helsinki, Finland, 5-9 April 1983)

Understeiner, N. (1968). Natural desalination and equilibrium salinity profile of perennial sea ice. *Journal of Geophysical Research*, *73*, 1251-1257. doi: 10.1029/JB073i004p01251

Walpole, R. E., et al. (2012). *Probability statistics for engineers scientists*. Pearson.

Weber, J. E. (2009). *Oceanography*. EOLSS Publishers.

Wettlaufer, J. S. (2011). Brine fluxes from growing sea ice. *Geophysical Research Letters*, *38*, 1-5. doi: 10.1029/2010GL046288

Zubov, N. N. (1963). *Arctic ice*. San Diego, Calif., U.S. Navy Electronics Laboratory.

# Appendix

## A KV Svalbard

KV Svalbard is a Norwegian icebreaking coastguard vessel. She was built at "Langstein Slip & Båtbyggeri AS", and delivered in 2002. With a displacement of 6375 tonnes (*KV Svalbard*, 2014), she is the largest ship in Norway's military force measured by tonnage. KV Svalbard is the only ice-going vessel in the Norwegian military force, and her main operating areas are in the Arctic part of Norwegian territory; Northern parts of Norway, Barents Sea and the areas around Svalbard. She is designed to operate in 1 meter thick level ice and 4 meter thick ice features. Main dimensions and technical data are given in Table 35.

Table 35: KV Svalbard.

| | |
|---|---|
| Built | 2001 |
| Class (DNV, July 1999) | +1A1, Icebreaker Polar-10 RPS F-A, E0, Heldk-SH, Deice, Firefighter 1 |
| Deplacement | 6375 [t] |
| Length o.a. | 103.7 [m] |
| Length p.p | 88.7 [m] |
| Breadth, moulded | 19.1 [m] |
| Draught | 6.5 [m] |
| Propulsion | 10 000 [kW] |
| Speed | 18 [kn] |
| Crew | 48 + |

## B Critical Values for the Kolmogorov-Smirnov Test

Critical values for the Kolmogorov-Smirnov test, $D_{crit}$, are given for data set size $n$ and confidence level $(1 - \alpha)$ below:

Table 36: Table of $D_{crit}$ for the Kolmogorov-Smirnov test.

| n\$\alpha$ | 0.2 | 0.10 | 0.05 | 0.01 |
|---|---|---|---|---|
| 5 | 0.45 | 0.51 | 0.56 | 0.67 |
| 10 | 0.32 | 0.37 | 0.41 | 0.49 |
| 15 | 0.27 | 0.30 | 0.34 | 0.40 |
| 20 | 0.23 | 0.26 | 0.29 | 0.36 |
| 25 | 0.21 | 0.24 | 0.27 | 0.32 |
| 30 | 0.19 | 0.22 | 0.24 | 0.29 |
| 35 | 0.18 | 0.20 | 0.23 | 0.27 |
| 40 | 0.17 | 0.19 | 0.21 | 0.25 |
| 45 | 0.16 | 0.18 | 0.20 | 0.24 |
| 50 | 0.15 | 0.17 | 0.19 | 0.23 |
| $> 50$ | $1.07/\sqrt{n}$ | $1.22/\sqrt{n}$ | $1.36/\sqrt{n}$ | $1.63/\sqrt{n}$ |

# C Figures

In this appendix, plots and figures that are excluded from the main text are found.

## C.1 Q-Q Plots

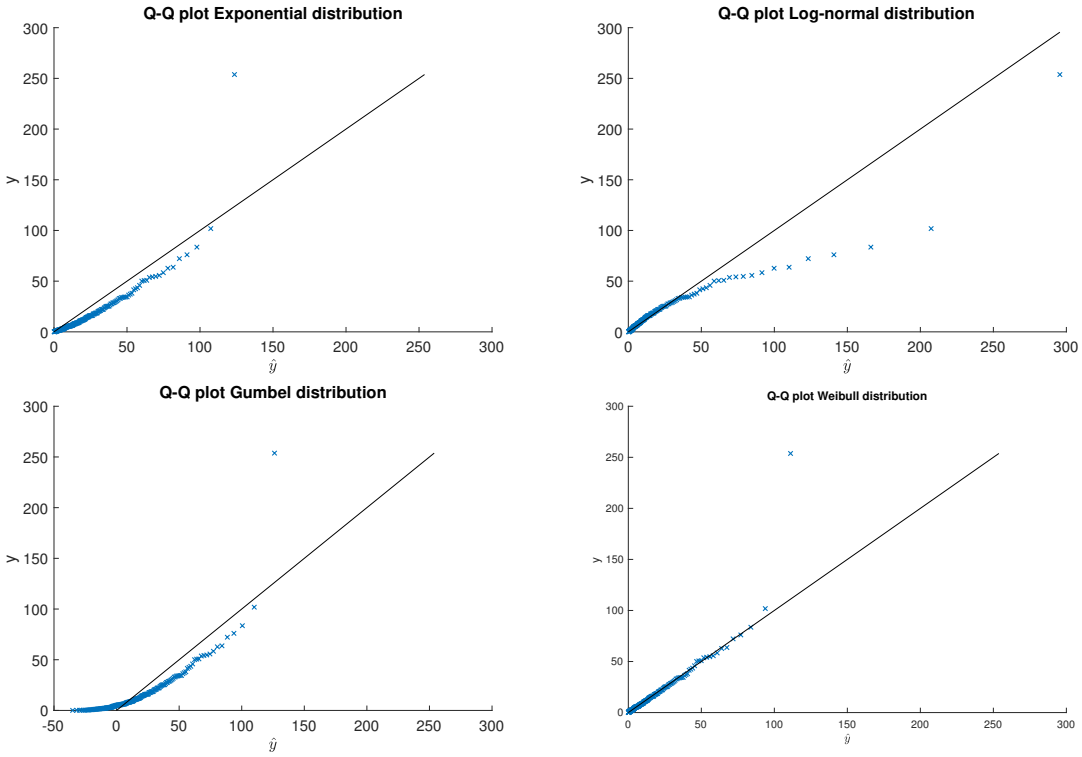



Figure 50: Q-Q plots for set 5 L5.

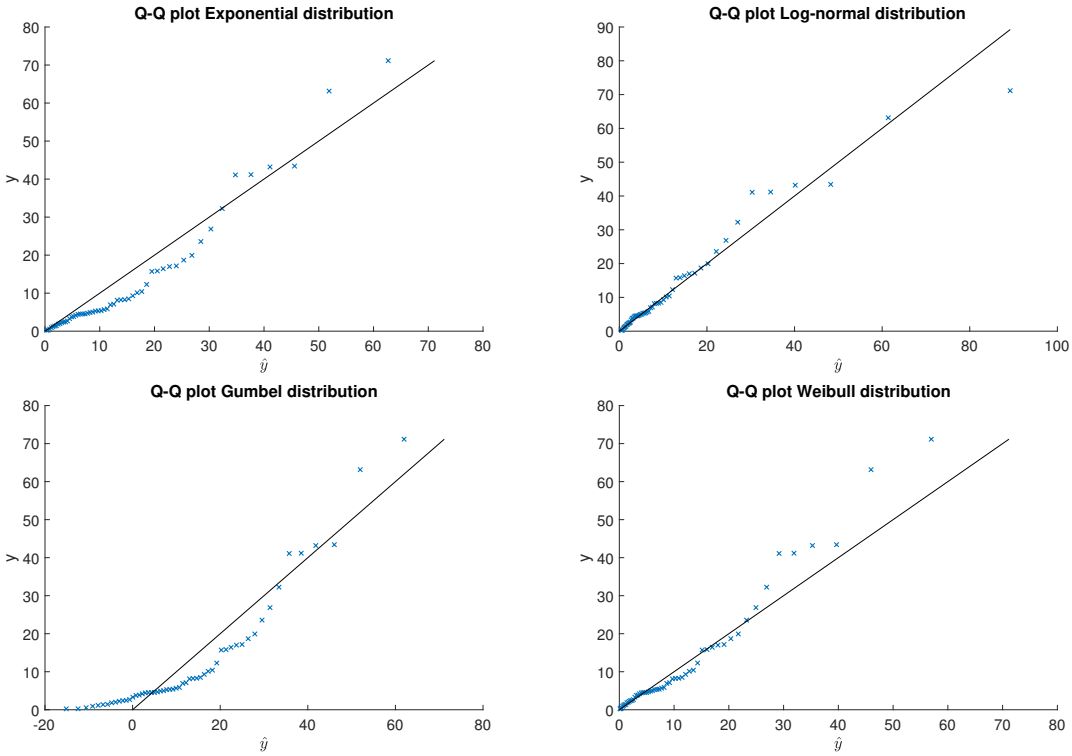



Figure 51: Q-Q plots for set 6 L5.

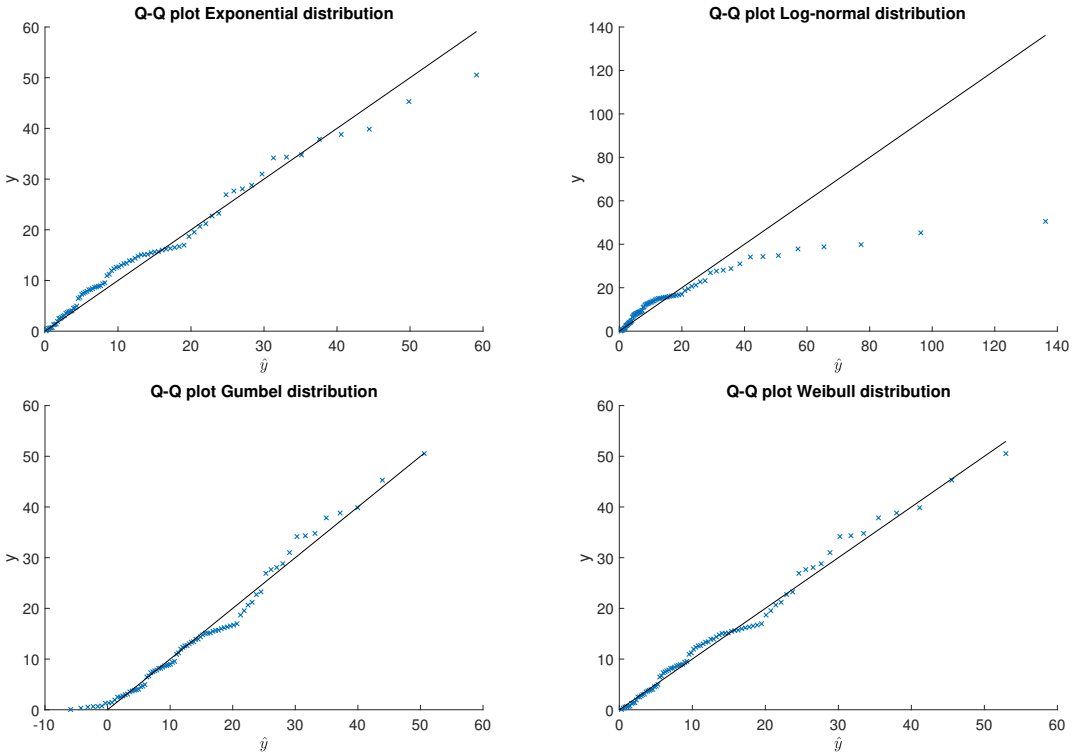



Figure 52: Q-Q plots for set 9 L2.

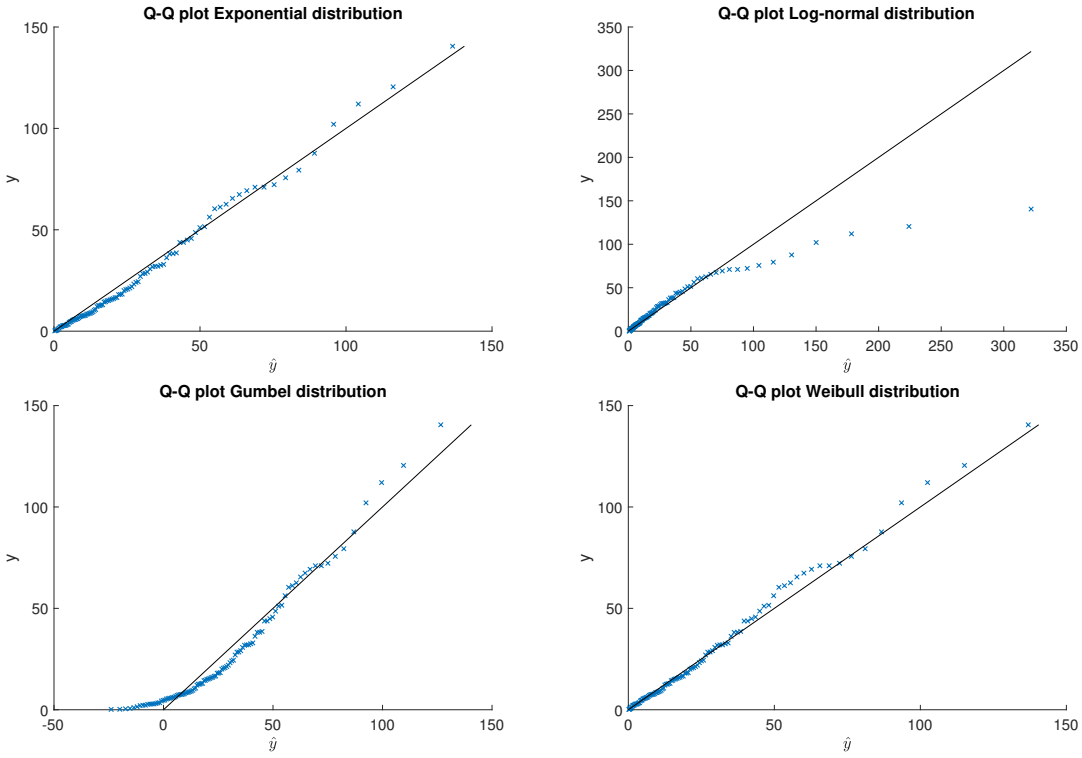



Figure 53: Q-Q plots for set 14 L1.

## C.2 Asymptotic Approach

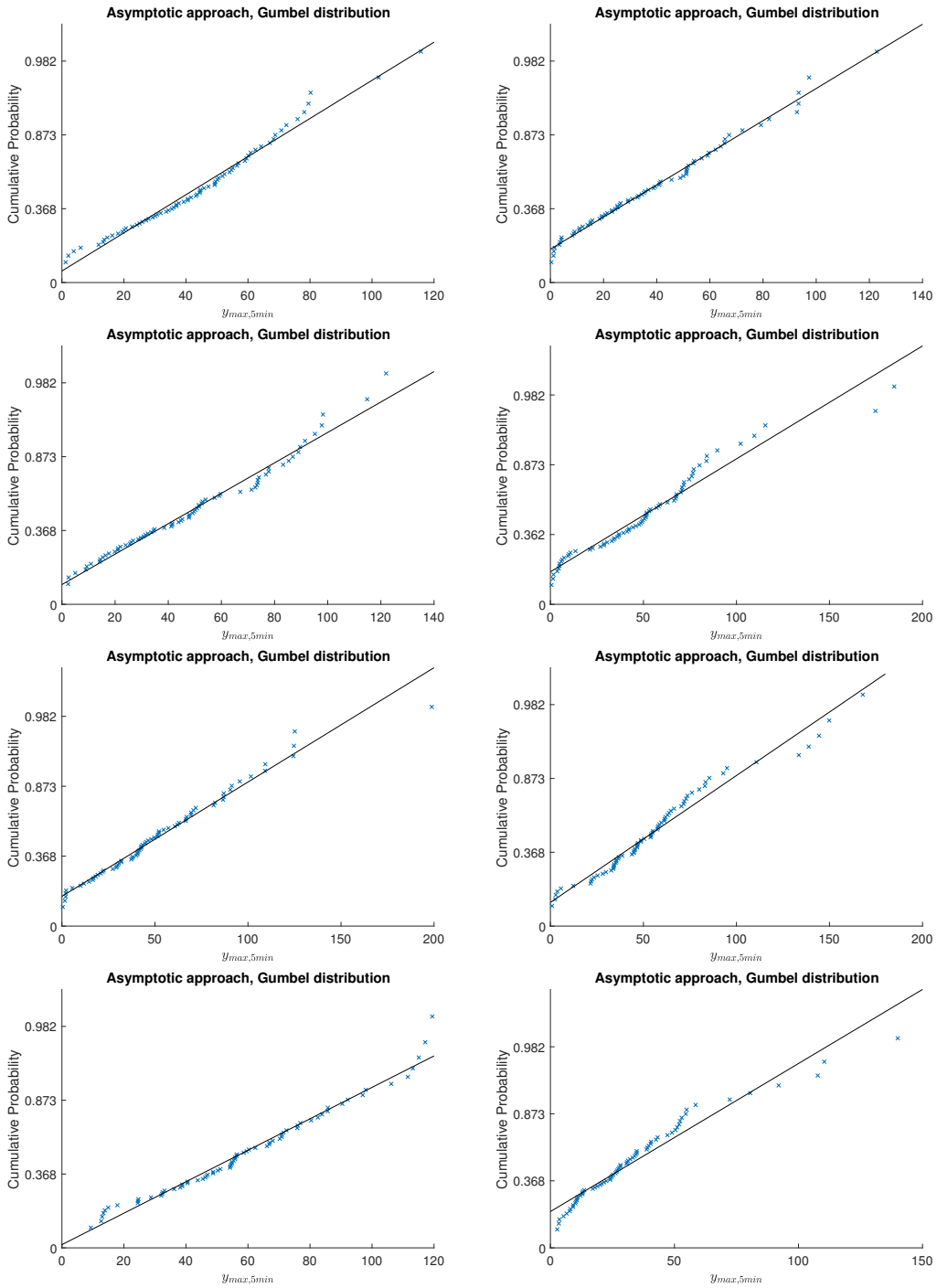



Figure 54: $\Delta t = 5$ min. Sensor L1 to L8 for the 6h time series used by Chai et al. (2018).

## C.3 Kolmogorov-Smirnov Tests

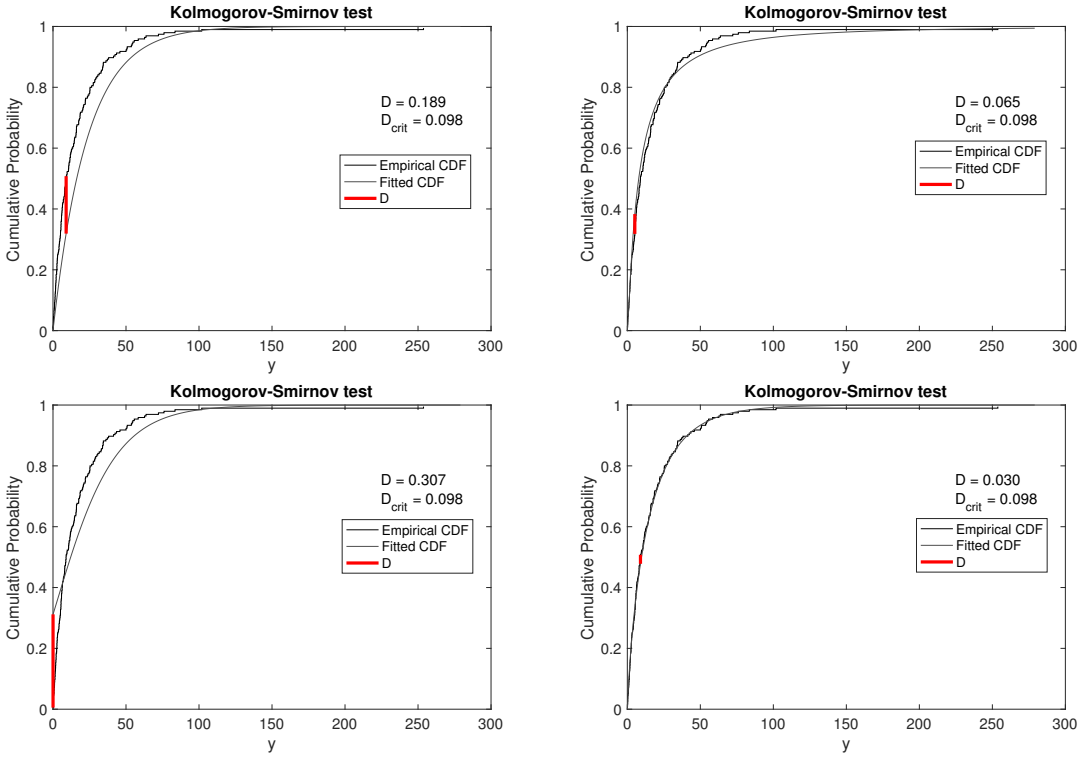



Figure 55: Kolmogorov-Smirnov tests for set 5 L5.

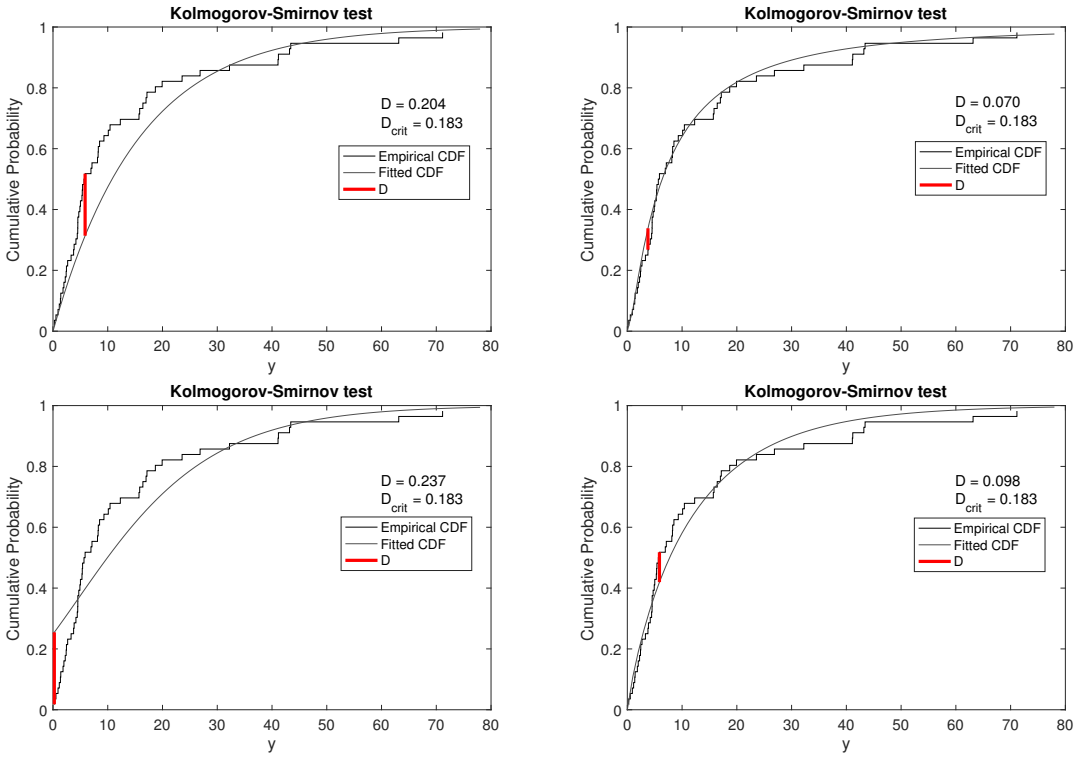



Figure 56: Kolmogorov-Smirnov tests for set 6 L5.

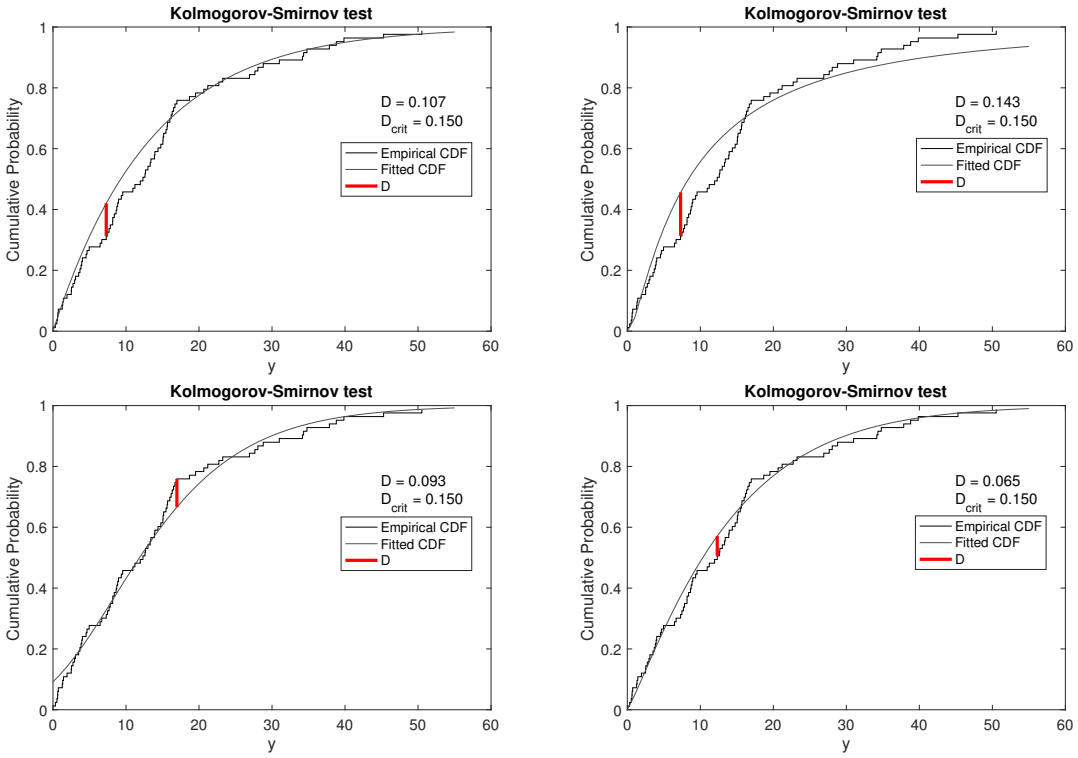



Figure 57: Kolmogorov-Smirnov tests for set 9 L2.

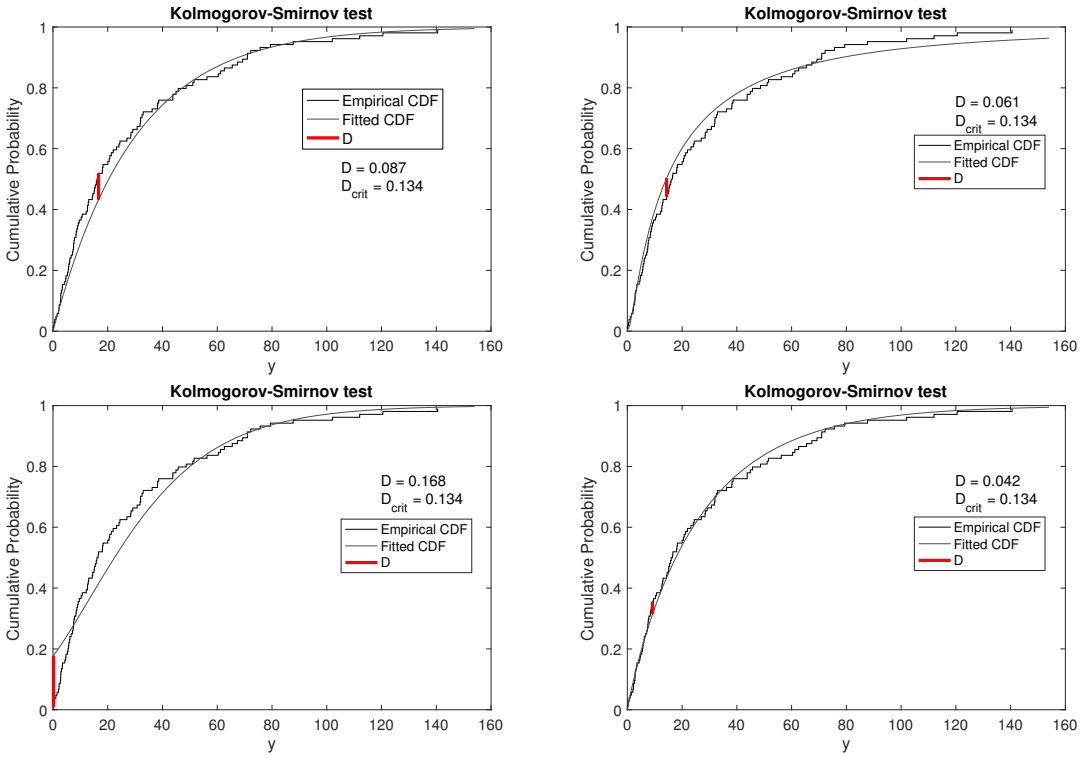
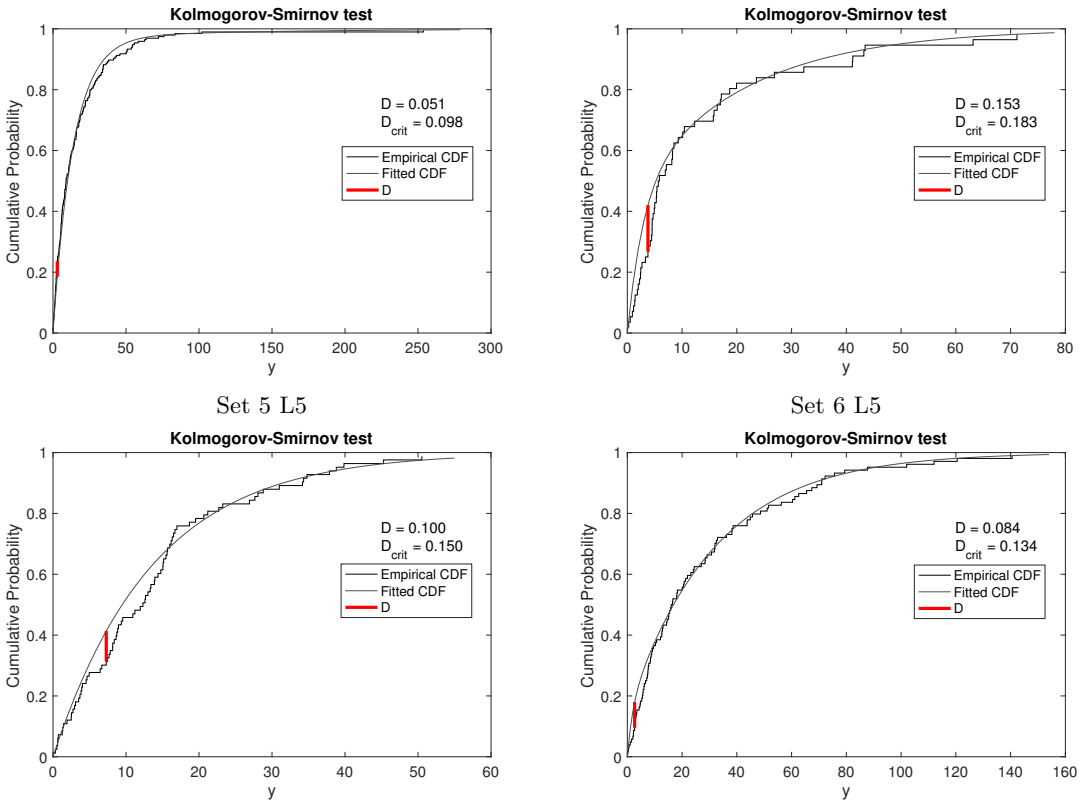



Figure 58: Kolmogorov-Smirnov tests for set 14 L1.

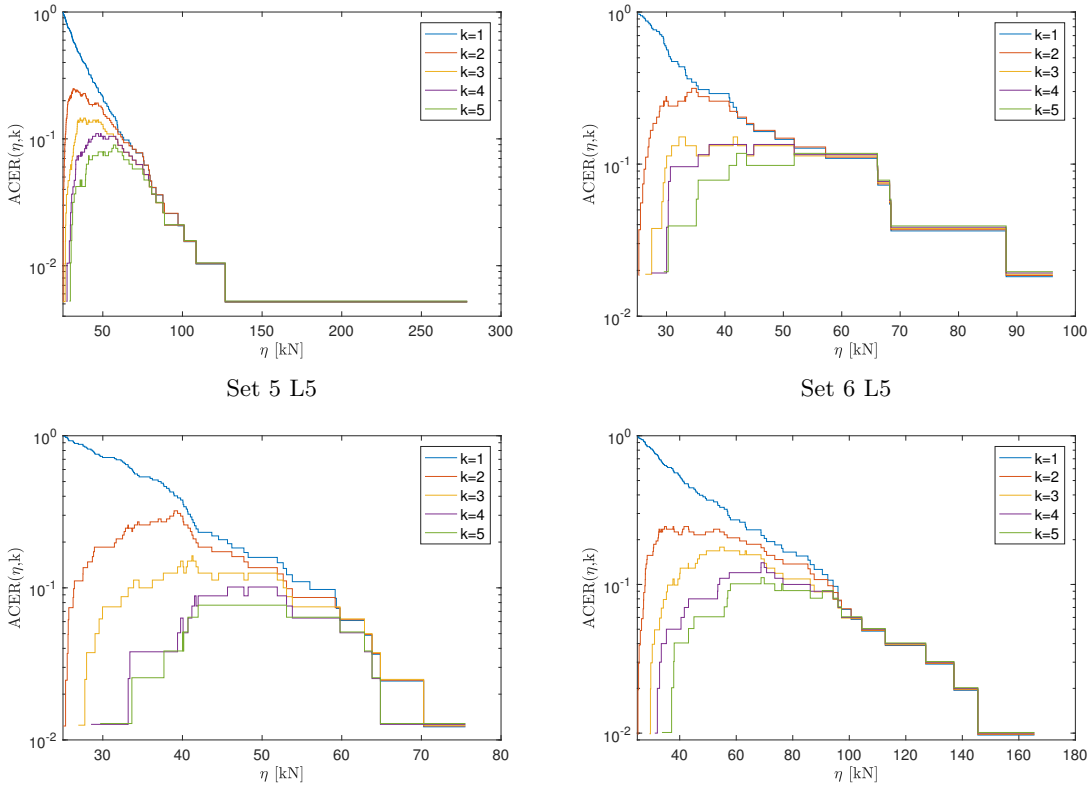



Set 5 L5 Set 6 L5

Set 9 L2 Set 14 L1

Figure 59: Kolmogorov-Smirnov tests for the three-parameter exponential distribution.

### C.4 ACER Functions




Set 5 L5 | Set 6 L5

Set 9 L2 | Set 14 L1

Figure 60: ACER functions.

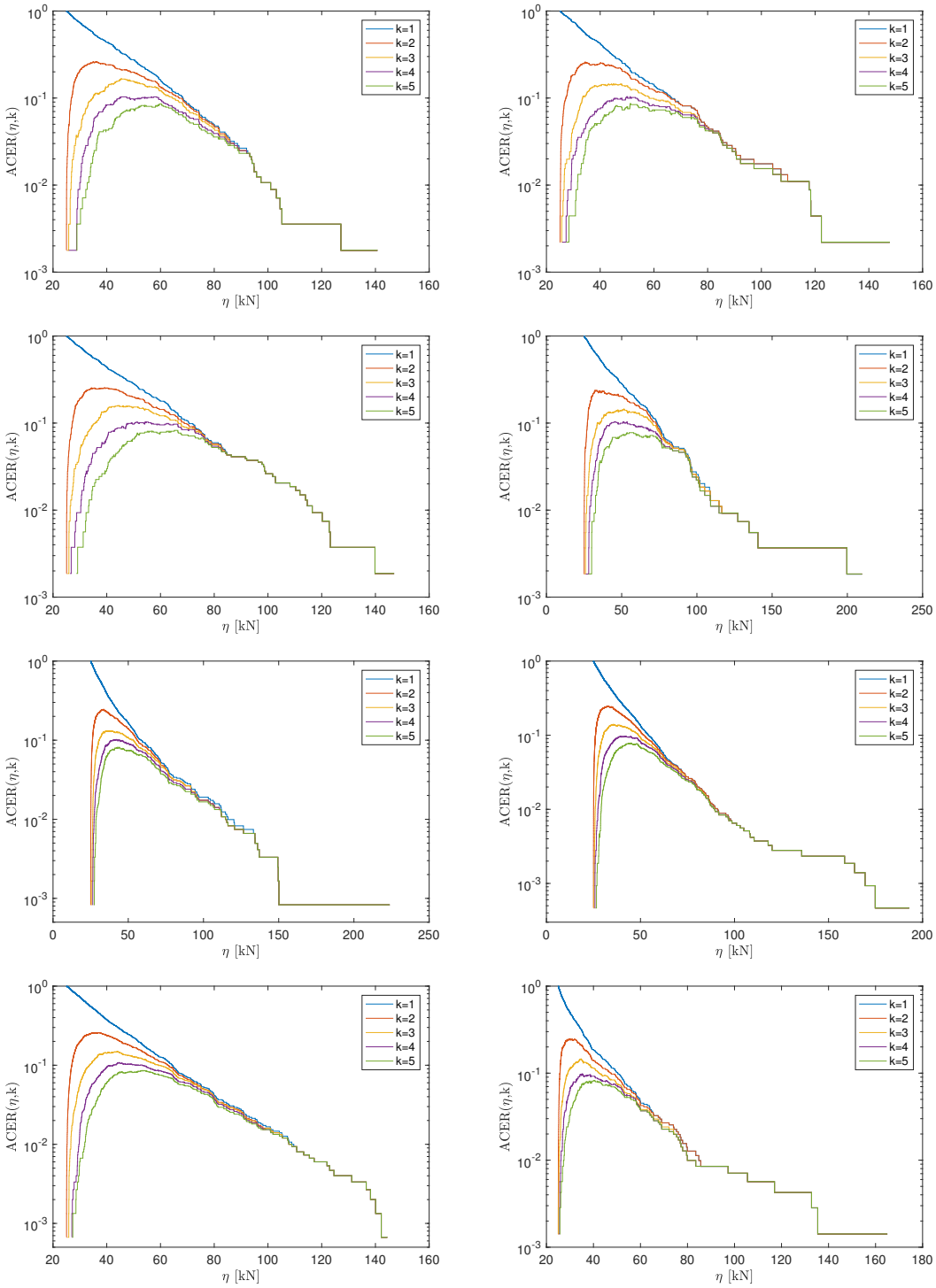



Figure 61: ACER functions for sensor L1 to L8 for the 6h time series used by Chai et al. (2018).

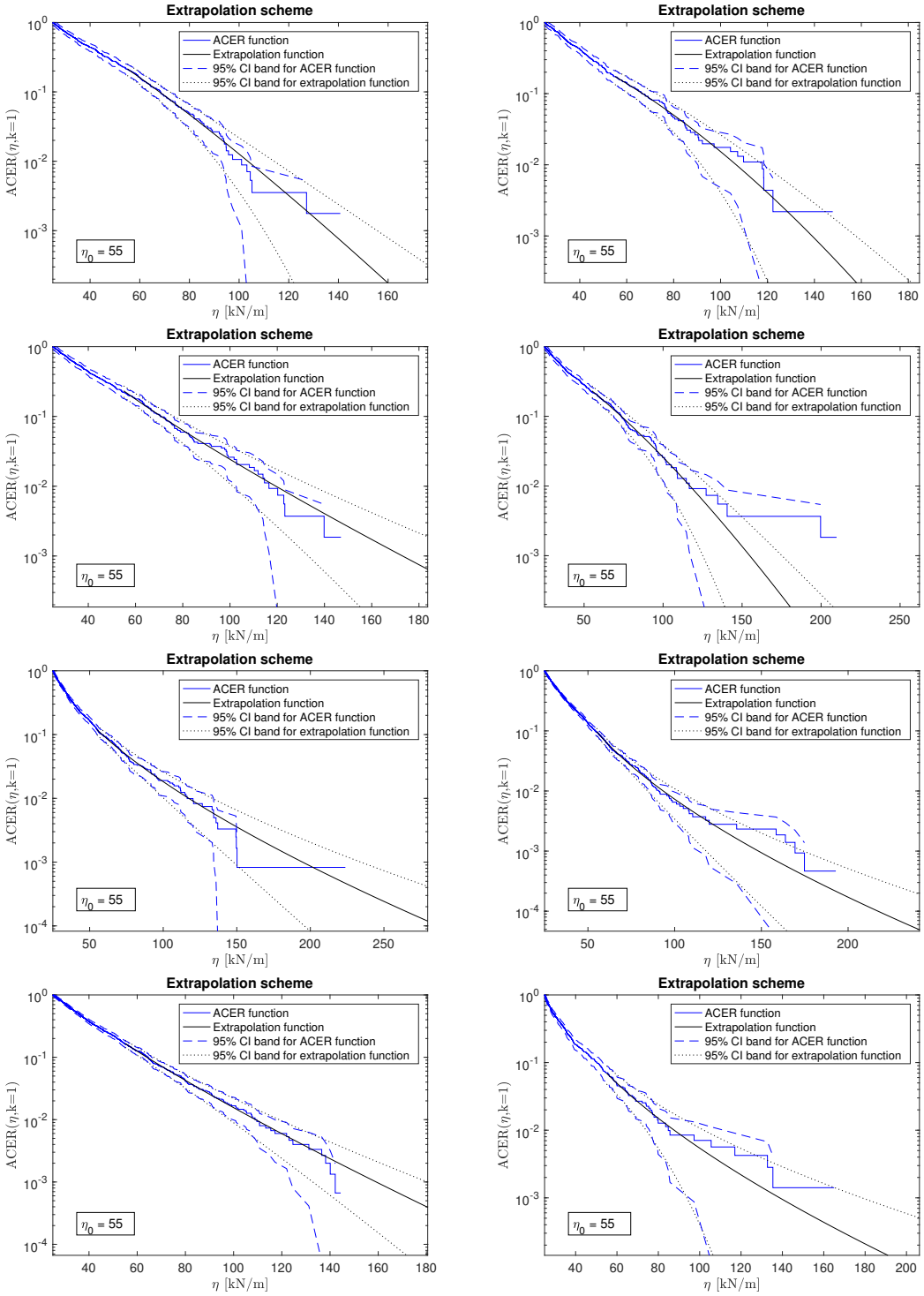



Figure 62: Sensor L1 to L8 for the 6h time series used by Chai et al. (2018).

# D MATLAB Scripts

For a list of inputs/outputs and a description of each script, see Section 6. The most important scripts are included. Scripts used for plotting, checking things, and trivial calculations are not presented.

### separate.m

```
% Start separate.m ------------------------------------------------------

load('peakDB_2007'); %Load vector
n=length(peakDB);

%index variables for each load sensor:
c1=1;
c2=1;
c3=1;
c4=1;
c5=1;
c6=1;
c7=1;
c8=1;
c9=1;

%
for i=1:n
    if peakDB(i,4)==1 %If the sensor is L1
        L1(c1,:)=[peakDB(i,1),peakDB(i,2),peakDB(i,3)];%Add load data to L1
        c1=c1+1;
    elseif peakDB(i,4)==2 %If the sensor is L2
        L2(c2,:)=[peakDB(i,1),peakDB(i,2),peakDB(i,3)];%Add load data to L2
        c2=c2+1;
    elseif peakDB(i,4)==3 %If the sensor is L3
        L3(c3,:)=[peakDB(i,1),peakDB(i,2),peakDB(i,3)];%Add load data to L3
        c3=c3+1;
    elseif peakDB(i,4)==4 %If the sensor is L4
        L4(c4,:)=[peakDB(i,1),peakDB(i,2),peakDB(i,3)];%Add load data to L4
        c4=c4+1;
    elseif peakDB(i,4)==5 %If the sensor is L5
        L5(c5,:)=[peakDB(i,1),peakDB(i,2),peakDB(i,3)];%Add load data to L5
        c5=c5+1;
    elseif peakDB(i,4)==6 %If the sensor is L6
        L6(c6,:)=[peakDB(i,1),peakDB(i,2),peakDB(i,3)];%Add load data to L6
        c6=c6+1;
    elseif peakDB(i,4)==7 %If the sensor is L7
        L7(c7,:)=[peakDB(i,1),peakDB(i,2),peakDB(i,3)];%Add load data to L7
        c7=c7+1;
    elseif peakDB(i,4)==8 %If the sensor is L8
```

```
41          L8(c8,:)=[peakDB(i,1),peakDB(i,2),peakDB(i,3)];%Add load data to L8
42          c8=c8+1;
43      elseif peakDB(i,4)==9 %If the sensor is L9
44          L9(c9,:)=[peakDB(i,1),peakDB(i,2),peakDB(i,3)];%Add load data to L9
45          c9=c9+1;
46      end
47  end
48
49  % End separate.m ----------------------------------------------------------
```

### prob_paper.m

```
% Start prob_paper.m -------------------------------------------------------

if exist('bootstr', 'var')==0 %True if script is not ran by bootstrapping.m
close
clearvars -except method start stop LX
load('ILMdata') %Ice load monitoring data
run('separate');

% User inputs:
start=; %Start row in the matrix "data"/"time"
stop=; %End row in the matrix "data"/"time"
LX=; %Load (strain) sensor

prompt = 'Input: 1 for exponential distribution, 2 for Log-normal,',...
    '3 for Gumbel, 4 for Weibull, 5 for 3-parameter-exponential ';
distr = input(prompt);

% Threshold:
threshold=25; %[kN] >= 25. 25 is default in the data set
if threshold>25
a=1;
LX_copy=LX;
clearvars LX
for i=1:length(LX_copy)
    if LX_copy(i,2)>threshold
        LX(a,:)=LX_copy(i,:);
        a=a+1;
    end
end
end

i=1;
% Find start point "i" for time series in "time":
while LX(i,1)<time(start)
    i=i+1;
end
j=i;
l_LX=length(LX);
% Finds end point "j" for time series in "time":
while LX(j,1)<time(stop) && j<l_LX
    j=j+1;
end

dataset=LX(i:j,2);
end %If the script is not ran by bootstrapping.m
dataset=sort(dataset);

N=length(dataset);
```

```
% Empirical CDF:
p_i=1:N;
p_i=p_i/(N+1);

if exist('bootstr','var')==0 %True if script is not ran by bootstrapping.m
dataset=dataset-threshold; %For fitting in probability papers
end

if distr==1 %Exponential
    %Abscissa and ordinate transformations (x and y):
    x=dataset;
    y=-log(1-p_i);
    hold on
    xlabel('y')
    ylabel('Cumulative Probability')
    title('Exponential distribution')
    scatter(x,y,'x')

elseif distr==2 %Log-normal
    hold on
    xlabel('y')
    ylabel('Cumulative Probability')
    title('Log-normal distribution')
    x=log(dataset);
    y=norminv(p_i); %Normal inverse cumulative distribution function
    scatter(x,y,'x')

elseif distr==3 %Gumbel
    x=dataset;
    y=-log(-log(p_i));
    hold on
    xlabel('y')
    ylabel('Cumulative Probability')
    title('Gumbel distribution')
    scatter(x,y,'x')
    hold off

elseif distr==4 %Weibull
    x=log(dataset);
    y=log(-log(1-p_i));

    hold on
    xlabel('y')
    ylabel('Cumulative Probability')
    title('Weibull distribution')
    scatter(x,y,'x')

elseif distr==5 %Three-parameter exponential
    run('three_param_exp');
    return
```

```
end

% Fitted distribution:
if distr==1 %Exponential
    c=0; %The exponential distribution is by def. starting at the origo
    m=x\y.'; %solve for slope, zero intercept
    yhat=[0 y(end)]; %evaluate over same range from origin
    plot([0 y(end)/m],yhat,'k-') %plot fitted line
elseif distr==2 %Log-normal. Fitted line based on MLE parameters
    mu=sum(log(dataset))/N;
    sigma=sqrt(sum((log(dataset)-mu).^2)/N);
    xx=[logninv(1/length(x),mu,sigma) , logninv((length(x)+1)/(length(x)...
        +2),mu,sigma)];
    xx=log(xx);
    yy=[norminv(1/length(x)) , norminv((length(x)+1)/(length(x)+2))];
    scatter(xx,yy,'.w')%create two points based on the MLE parameters to...
    %create the fitted line
    l=lsline; %plot least square fitted line
    set(l(1),'color',0.0*[1 1 1]) %Color of the fitted line
    m_c = polyfit(xx,yy,1);
    m=m_c(1); %slope
    c=m_c(2); %ordinate interception
elseif distr==3 %Gumbel
    l=lsline; %plot the fitted line
    set(l(1),'color',0.0*[1 1 1]) %Color of the fitted line
    m_c = polyfit(get(l,'xdata'),get(l,'ydata'),1);
    m=m_c(1); %slope
    c=m_c(2); %ordinate interception
elseif distr==4 %Weibull. Fitted line based on MLE parameters
    thetak=wblfit(dataset); %Parameter estimation for Weibull using MLE
    k=thetak(2); %Weibull
    theta=((1/N)*sum(dataset.^k))^(1/k); %Weibull
    xx=[wblinv(1/length(x),theta,k) ,...
        wblinv((length(x)+1)/(length(x)+2),theta,k)];
    xx=log(xx);
    yy=[log(-log(1-(1/length(x)))) ,...
        log(-log(1-((length(x)+1)/(length(x)+2))))];
    scatter(xx,yy,'.w')%create two points based on the MLE parameters...
    %to create the fitted line
    l=lsline; %plot the fitted line
    set(l(1),'color',0.0*[1 1 1]) %Color of the fitted line
    m_c = polyfit(xx,yy,1);
    m=m_c(1); %slope
    c=m_c(2); %ordinate interception
end

if exist('bootstr','var')==0 %True if script is not ran by bootstrapping.m
% Coefficient of Determination (R^2):
f=m*x+c;
y_bar=mean(y);
```

```
SSres=sum((y-f.').^2);
SStot=sum((y-y_bar).^2);
R_squared=1-(SSres/SStot);
end

% End prob_paper.m ----------------------------------------------------------
```

### qq_plot.m

```
% Start qq_plot.m ------------------------------------------------------

clear all
load('ILMdata') %Ice load monitoring data
run('separate');

start=;
stop=;
LX=;

% Threshold:
threshold=25; %[kN] >= 25. 25 is default in the data set
if threshold>25
a=1;
LX_copy=LX;
clearvars LX
for i=1:length(LX_copy)
    if LX_copy(i,2)>threshold
        LX(a,:)=LX_copy(i,:);
        a=a+1;
    end
end
end

i=1;
% Find start point "i" for time series in "time":
while LX(i,1)<time(start)
    i=i+1;
end
j=i;
l_LX=length(LX);
% Finds end point "j" for time series in "time":
while LX(j,1)<time(stop) && j<l_LX
    j=j+1;
end

dataset=LX(i:j,2);
dataset=sort(dataset);
% Empirical CDF:
p_i=1:length(dataset);
p_i=p_i/(length(dataset)+1);

prompt = 'User input: 1 for parameters estimated from probability',...
    'paper, 2 for maximum likelihood estimate (MLE) ';
method = input(prompt);

% Estimate distribution parameters:
run('parameters')
```

```
close all

% Calculate load value for each quantile for fitted distribution
if distr==1 %Exponential
    X=-log(1-p_i)/lambda;
elseif distr==2 %Log-normal
    X=exp(sigma*norminv(p_i) + mu);
elseif distr==3 %Gumbel
    X=alpha - beta*log(-log(p_i));
elseif distr==4 %Weibull
    X=theta*(-log(1-p_i)).^(1/k);
end

% Q-Q plot:
scatter(X,dataset)
hold on
% Plot line with slope 1 through origin (line of perfect match):
plot([0,max(max(dataset,X))],[0,max(max(dataset,X))])
xlabel({'$\hat{y}$'},'Interpreter','latex')
ylabel('y')
if distr==1
    title('Q-Q plot Exponential distribution')
elseif distr==2
    title('Q-Q plot Log-normal distribution')
elseif distr==3
    title('Q-Q plot Gumbel distribution')
elseif distr==4
    title('Q-Q plot Weibull distribution')
end
hold off

% End qq_plot.m ----------------------------------------------------------
```

**parameters.m**

```
% Start parameters.m -------------------------------------------------------

if exist('no_plot','var')==0 %False if bootstrapped samples are analyzed
run('prob_paper')
end
N=length(dataset);
if distr==5 %For Kolmogorov-Smirnov test. Parameters not calculated here
return
end

% Based on probability paper:
if method==1
lambda=m; %Exponential
sigma=1/m; %Log-normal
mu=-sigma*c; %Log-normal
beta=1/m; %Gumbel
alpha=-beta*c; %Gumbel
k=m; %Weibull
theta=exp(-c/k); %Weibull

% Maximum Likelihood Estimate (MLE):
elseif method==2
lambda=1/mean(dataset); %Exponential
mu=sum(log(dataset))/N; %Log-normal
sigma=sqrt(sum((log(dataset)-mu).^2)/N); %Log-normalexp
alphabeta=evfit(-dataset); %Parameter estimation for Gumbel using MLE
alpha=-alphabeta(1); %Gumbel
beta=alphabeta(2); %Gumbel
thetak=wblfit(dataset); %Parameter estimation for Weibull using MLE
k=thetak(2); %Weibull
theta=((1/N)*sum(dataset.^k))^(1/k); %Weibull
%*************************************************************
end

%Relevant for bootstrapped samples:
if distr==1 %Exponential
    parameter1=lambda;
    parameter2=0;
    %Most probable largest value during the evaluated time series:
    x_max=(-log(1/N)/lambda) +25; %Corrected for the threshold
elseif distr==2%Log-normal
    parameter1=sigma;
    parameter2=mu;
    x_max=exp(sigma*norminv(1-1/N) + mu) + 25;
elseif distr==3%Gumbel
    parameter1=beta;
    parameter2=alpha;
    x_max=alpha - beta*log(-log(1-1/N)) + 25;
```

```
49 elseif distr==4%Weibull
50     parameter1=k;
51     parameter2=theta;
52     x_max=theta*(-log(1/N))^(1/k) + 25;
53 end
54
55 % End parameters.m ------------------------------------------------------
```

**asymptotic.m**

```
% Start asymptotic.m ------------------------------------------------------

clear all
load('ILMdata')
run('separate');

% User inputs:
Dt=; %length of applied sub-interval [min]
start=; %in the file "data"/"time"
stop=; %in the file "data"/"time"
LX=; % define strain sensor

% Threshold:
threshold=25; %[kN] >= 25. 25 is default in the data set
if threshold>25
a=1;
LX_copy=LX;
clearvars LX
for i=1:length(LX_copy)
    if LX_copy(i,2)>threshold
        LX(a,:)=LX_copy(i,:);
        a=a+1;
    end
end
end

% Extract and create data set:
peak_int=[];
dataset=[];
start_time=time(start);
stop_time=time(stop);
a1=1; %Evaluated element in LX
a2=1; %Sub-interval nr
a3=1; %Element nr in sub-interval
while start_time>LX(a1,1) %Finds start point in LX
    a1=a1+1;
end

while LX(a1,1)<=stop_time
    int=floor((LX(a1,1)-start_time)/(Dt*60)); %
    if int>a2
        if a3>1
        dataset(a2)=max(peak_int(a2,:)); %Max in subinterval a2 stored
        a2=a2+1; %New sub-interval is created
        a3=1;
        end
    end
    peak_int(a2,a3)=LX(a1,2);
```

```
    a1=a1+1;
    a3=a3+1;
end

dataset(a2)=max(peak_int(a2,:)); %Data set applied in analysis
dataset=sort(dataset);
if exist('bootstr','var')==0 %True if script is not ran by bootstrapping.m
dataset=dataset-threshold; %Subtracted the threshold for fitting purposes
end


% Empirical CDF:
p_i=1:length(dataset);
p_i=p_i/(length(dataset)+1);

% Plot in Gumbel probability paper:
x=dataset;
y=-log(-log(p_i));
hold on
figure(1)
scatter(x,y,'x')
xlabel('$y_{max,1min}$','Interpreter','latex')
ylabel('Cumulative Probability')
title('Asymptotic approach, Gumbel distribution')
hold off

% Least square fit:
l=lsline; %plot the fitted line
set(l(1),'color',0.0*[1 1 1]) %Color of the fitted line
m_c = polyfit(get(l,'xdata'),get(l,'ydata'),1);
m=m_c(1); %slope of fitted line
c=m_c(2); %ordinate interception for fitted line

if exist('bootstr','var')==0 %True if script is not ran by bootstrapping.m
% Coefficient of Determination (R^2):
f=m*x+c; %Fitted line
y_bar=mean(y);
SSres=sum((y-f).^2);
SStot=sum((y-y_bar).^2);
R_squared=1-(SSres/SStot);
end

% Parameter estimation based on probability paper:
beta=1/m;
alpha=-beta*c;

% Parameter estimation by MLE:
% alphabeta=evfit(-dataset);
% alpha=-alphabeta(1);
% beta=alphabeta(2);
```

```
99
100 % Kolmogorov-Smirnov test:
101 X_cdf=0:1:max(dataset)*1.1; %For fitted distribution
102 Y_cdf=exp(-exp(-(X_cdf-alpha)/beta)); %For fitted distribution
103 ks_cdf=exp(-exp(-(dataset-alpha)/beta)); %For fitted distribution
104
105 % Plot empirical and fitted CDF:
106 figure(2)
107 fitted=plot(X_cdf,Y_cdf,'-','color',0.3*[1 1 1]); %Plot fitted CDF
108 hold on
109 empirical=stairs(dataset,p_i,'-k'); %Plot empirical CDF
110
111 %Finding the measure "D":
112 ks=abs(ks_cdf-p_i);
113 D=max(ks);
114
115 i=1;
116 while D>ks(i)
117     i=i+1;
118 end
119 % Plot "D":
120 D_plot=plot([dataset(i),dataset(i)],[ks_cdf(i),p_i(i)],'-r','LineWidth',3);
121
122 title('Kolmogorov-Smirnov test')
123 xlabel('$y_{max,1min}$','Interpreter','latex')
124 ylabel('Cumulative Probability')
125 legend([empirical,fitted,D_plot],{'Empirical CDF','Fitted CDF','D'});
126
127 % Identify extreme loads and other characteristics:
128 N=length(dataset);
129 n_hour=60*N/((stop-start)*0.5)
130 max_data=max(dataset)+25; %Largest measured load in the data set
131 x_max=alpha - beta*log(-log(1-1/N))+25; %Estimated largest for the data set
132
133 % End asymptotic.m ------------------------------------------------------
```

### three_param_exp.m

```
% Start thhree_param_exp.m ------------------------------------------------

% !!!This plot is always ran by another script!!!

RMSE=zeros(N+1,1); %creating vector for Root-Mean-Square Error
a_i=zeros(N+1,1); %Vector of weight parameters a
l1=zeros(N+1,1); %Vector of lambda_1 parameters
l2=zeros(N+1,1); %Vector of lambda_2 parameters
l2(1)=N/(sum(dataset)); %Creating first element. a_i(1)=l1(1)=0

for i=1:N
    a_i(i+1)=i/N;
    l1(i+1)=i/(sum(dataset(1:i)));
    l2(i+1)=(N-i)/(sum(dataset(i:end)));
end

F=zeros(N,N+1); %Fitted cumulative probability for each element in
                %dataset for all possible divisions of the dataset
for i=1:N
    %Each column represent one "set" of parameters. Each row gives F for
    %the same element in "dataset" represented by different parameters.
    F(i,:)=a_i.*(1-exp(-l1*dataset(i)))+(1-a_i).*(1-exp(-l2*dataset(i)));
end

y1=-log(1-F); %Value in probability paper for estimated points
y2=-log(1-p_i); %Value in probability paper for measured points

for i=1:N+1
    RMSE(i)=sqrt(sum((y2'-y1(:,i)).^2)/N); %Root-Mean-Square Error
end

target=min(RMSE); %Lowest value in RMSE
for i=1:N+1
    if RMSE(i)==target %Find where the lowest RMSE is located
        nr=i;
        break
    end
end

% The set of parameters that give the lowest RMSE:
a=a_i(nr);
lambda1=l1(nr);
lambda2=l2(nr);

% Plotting the 3 parameter exp. distr and the dataset in exp. prob. paper:
x=0:1:max(dataset)*1.2;
F_fit=a*(1-exp(-lambda1*x))+(1-a)*(1-exp(-lambda2*x));
yy=-log(1-F_fit);
```

```
scatter(dataset,y2,'x')
hold on
plot(x,yy,'Color',0.0*[1 1 1])
title('3-parameter exponential distribution')
xlabel('y')
ylabel('-ln(1-p)')

% Least square line (1 parameter exponential, for comparison):
m=dataset\y2.'; %solve for slope, zero intercept
yhat=m*[0 dataset(end)]; %evaluate over same range from origin
plot([0 dataset(end)],yhat,'k-') %plot fitted line


% Most probable largest value during the time series:
n_data=length(dataset);
n_hour=60*n_data/((stop-start)*0.5);%loads per hour

%Initial values. Could have been set higher, but the script is fast anyways
F=0;
x=0;

while F<1-(1/n_data)
    x=x+0.1;
    F=a*(1-exp(-lambda1*x))+(1-a)*(1-exp(-lambda2*x));
end

% Coefficient of Determination:
F_fit=a*(1-exp(-lambda1*dataset))+(1-a)*(1-exp(-lambda2*dataset));
yy=-log(1-F_fit);
y_bar=mean(y2);
SSres=sum((y2-yy.').^2);
SStot=sum((y2-y_bar).^2);
R_squared=1-(SSres/SStot);

% Estimated most probable largest value during the analyzed time series:
x_max=x;

% End thhree_param_exp.m --------------------------------------------------
```

**acer.m**

```
% Start acer.m ------------------------------------------------------------

clear all
load('ILMdata')
run('separate');

% User inputs:
start=;
stop=;
LX=;
N_k=; %Highest order of the acer function to be included

% Threshold:
threshold=25; %[kN] >= 25. 25 is default in the data set
if threshold>25
a=1;
LX_copy=LX;
clearvars LX
for i=1:length(LX_copy)
    if LX_copy(i,2)>threshold
        LX(a,:)=LX_copy(i,:);
        a=a+1;
    end
end
end

i=1;
while LX(i,1)<time(start) %finds start point "i" of interval
    i=i+1;
end
j=i;
l_LX=length(LX);
while LX(j,1)<time(stop) && j<l_LX %finds stop point "j" of interval
    j=j+1;
end
dataset=LX(i:j,2);

eta=sort(dataset); %load vector eta
N=length(dataset);
ACER=zeros(N,N_k); %acer function values
P=zeros(N,N_k); %extreme value distribution for large values of eta
rho=zeros(N,N_k); %Weight factor used in optimization of fitted curve
plusminus=zeros(N,N_k); %half of the length of the confidence interval


for k=1:N_k %each order of the ACER function
for r=1:N %Calculating points for element in "dataset"

```

```
if k==1
    q=sort(dataset); %Used for calculating P(etta)

    i=round(N/2); %i: number of elements in "dataset" <= eta
    numerator=0.5;
    step=2;%Given a number >1
    while N*numerator>1 %Continues as long as the step length is > 1
        numerator=numerator/2;
        if q(i)<eta(r)
            i=round(i+N*numerator);
        elseif q(i)>eta(r)
            i=round(i-N*numerator);
        else %q(i)==etta(r)
            break
        end
    end
    if q(i)>eta(r) %Correctin for possible numerical errors
        i=i-1;
    end
    alpha_1j=(N-i)/N;%1-i/(N+1);

    ACER(r,k)=alpha_1j; %=(1/(N-k+1))*N*alpha_1j
    P(r,k)=exp(-N*alpha_1j);


    % Calculate a_kj for each element:
    else %if k is larger than 1
    a_kj=zeros(N,1);
    check=1;
    for j=k:N
        for i=(j-k+1):j-1
            if dataset(i)>eta(r) %Requirement: the previous loads...
            %being smaller than eta
            check=0; %fails
            break
            end
        end
        if dataset(j)<=eta(r) %Requirement: load is larger than eta
            check=0; %fails
        end
        if check==1 %All requirements are fulfilled.
            a_kj(j)=1;
        end
    check=1;
    end

    epsilon_hat=(1/(N-k+1))*sum(a_kj(k:end));
    ACER(r,k)=epsilon_hat;
    P(r,k)=exp(-sum(a_kj(k:end)));
end
```

```
plusminus(r,k)=(ACER(r,k)*1.96)/sqrt((N-k+1)*ACER(r,k));
rho(r,k)=(log(ACER(r,k)+plusminus(r,k))-log(ACER(r,k)-plusminus(r,k)))^-2;
end
end

%Confidence interval (lower and upper bound)
CIminus=ACER-plusminus;
CIplus=ACER+plusminus;

figure1 = figure;
% Create axes
axes1 = axes('Parent',figure1);
hold(axes1,'on');

% Create multiple lines using matrix input to stairs
stairs(eta,ACER,'Parent',axes1);

xlabel({'$\eta$ [kN]'},'Interpreter','latex')
ylabel({'ACER($\eta$,k)'},'Interpreter','latex')
legend({'k=1','k=2','k=3','k=4','k=5'});

box(axes1,'on');
% Set the remaining axes properties
set(axes1,'YMinorTick','on','YScale','log');

% End acer.m ---------------------------------------------------------------
```

### acer_fit.m

```
% Start acer_fit.m ----------------------------------------------------------

run('acer');

eta_0=; % [kN], Load value from where the extrapolation function starts
b0=mean(dataset); %Initial value for optimized constant b
c0=2; %Initial value for optimized constant c
x0=[b0,c0];
k=1; %Order of ACER function used for fitting of the extrapolation function

i=1;
% Find from which element number in dataset a function shall be fitted:
while eta(i)<eta_0
    i=i+1;
end
i=i+1; %Start point in data set for extrapolation function

rho_opt=real(rho(:,k)); %Some imaginary parts due to numerical errors
for j=i:(length(rho_opt)-1)
    if rho_opt(j)<0 %Only positive wheight factors are included
        j=j-1; %End point in load vector for extrapolation function
        break
    end
end

% For the fitting purpose (spanning from i to j):
rho_opt=rho_opt(i:j);
ACER_opt=ACER(i:j,k);
eta_opt=eta(i:j);
ln_acer=log(ACER_opt);

% Objective function:
F = @(x) sqrt(rho_opt).*(ln_acer - sum(rho_opt.*ln_acer)-((sum(rho_opt.*...
    ((eta_opt - x(1)).^x(2).*ln_acer)) - sum(rho_opt.*(eta_opt - x(1))...
    .^x(2))*sum(rho_opt.*ln_acer))/(sum(rho_opt.*((eta_opt-x(1)).^x(2).*...
    (eta_opt - x(1)).^x(2))) - (sum(rho_opt.*(eta_opt - x(1)).^x(2)))...
    ^2))*((eta_opt - x(1)).^x(2) - sum(rho_opt.*(eta_opt - x(1)).^x(2))));

%Optimization characteristics:
ObFun_log_LS_NLINFIT_LM = optimset('Display','off','Algorithm',...
    'levenberg-marquardt','MaxFunEvals', 10000, 'MaxIter', 10000,'TolX',...
    1e-12, 'TolFun', 1e-12);

%Optimization:
x = lsqnonlin(F, x0, [], [], ObFun_log_LS_NLINFIT_LM);
inp = (eta_opt - x(1)).^x(2);
A = (sum(rho_opt.*(inp.*ln_acer)) - sum(rho_opt.*inp)*sum(rho_opt.*...
    ln_acer))/(sum(rho_opt.*(inp.*inp)) - (sum(rho_opt.*inp))^2);
```

```
B = sum(rho_opt.*ln_acer) - A*sum(rho_opt.*inp);
sol(1:4) = [-A x(1) x(2) exp(B)];
a=-A; %Parameter used for extrapolation function
b=x(1); %Parameter used for extrapolation function
c=x(2); %Parameter used for extrapolation function
q=exp(B); %Parameter used for extrapolation function

% Upper band of CI, Same procedure as for the extrapolation function:
ln_CIplus=log(CIplus(i:j,k));
b0plus=mean(CIplus);
c0plus=2;
x0plus=[b,c];
Fplus = @(xplus) sqrt(rho_opt).*(ln_CIplus - sum(rho_opt.*ln_CIplus) - ...
((sum(rho_opt.*((eta_opt - xplus(1)).^xplus(2).*ln_CIplus)) - sum...
(rho_opt.*(eta_opt - xplus(1)).^xplus(2))*sum(rho_opt.*ln_CIplus))/(sum...
(rho_opt.*((eta_opt - xplus(1)).^xplus(2).*(eta_opt - xplus(1)).^...
xplus(2))) - (sum(rho_opt.*(eta_opt - xplus(1)).^xplus(2)))^2))*...
((eta_opt - xplus(1)).^xplus(2) - sum(rho_opt.*(eta_opt - xplus(1)).^...
xplus(2))));

ObFun_log_LS_NLINFIT_LM = optimset('Display','off',...
    'Algorithm', 'levenberg-marquardt', ...
    'MaxFunEvals', 10000, 'MaxIter', 10000,...
    'TolX', 1e-12, 'TolFun', 1e-12);

xplus = lsqnonlin(Fplus, x0plus, [], [], ObFun_log_LS_NLINFIT_LM);
inp = (eta_opt - xplus(1)).^xplus(2);
A = (sum(rho_opt.*(inp.*ln_CIplus)) - sum(rho_opt.*inp)*sum(rho_opt.*...
    ln_CIplus))/(sum(rho_opt.*(inp.*inp)) - (sum(rho_opt.*inp))^2);
B = sum(rho_opt.*ln_CIplus) - A*sum(rho_opt.*inp);
solplus(1:4) = [-A xplus(1) xplus(2) exp(B)];
a_plus=-A;
b_plus=xplus(1);
c_plus=xplus(2);
q_plus=exp(B);

% Lower band of CI, Same procedure as for the extrapolation function:
ln_CIminus=log(CIminus(i:j,k));
b0minus=mean(CIminus);
c0minus=2;
x0minus=[b,c];
Fminus = @(xminus) sqrt(rho_opt).*(ln_CIminus - sum(rho_opt.*ln_CIminus)...
    -((sum(rho_opt.*((eta_opt - xminus(1)).^xminus(2).*ln_CIminus))-sum...
    (rho_opt.*(eta_opt - xminus(1)).^xminus(2))*sum(rho_opt.*...
    ln_CIminus))/(sum(rho_opt.*((eta_opt - xminus(1)).^xminus(2).*...
    (eta_opt-xminus(1)).^xminus(2))) - (sum(rho_opt.*(eta_opt-xminus(1))...
    .^xminus(2)))^2))*((eta_opt - xminus(1)).^xminus(2) - sum(rho_opt.*...
    (eta_opt - xminus(1)).^xminus(2))));

ObFun_log_LS_NLINFIT_LM = optimset('Display','off',...
```

```
    'Algorithm', 'levenberg-marquardt', ...
    'MaxFunEvals', 10000, 'MaxIter', 10000,...
    'TolX', 1e-12, 'TolFun', 1e-12);

xminus = lsqnonlin(Fminus, x0minus, [], [], ObFun_log_LS_NLINFIT_LM);
inp = (eta_opt - xminus(1)).^xminus(2);
A = (sum(rho_opt.*(inp.*ln_CIminus)) - sum(rho_opt.*inp)*sum(rho_opt.*...
    ln_CIminus))/(sum(rho_opt.*(inp.*inp)) - (sum(rho_opt.*inp))^2);
B = sum(rho_opt.*ln_CIminus) - A*sum(rho_opt.*inp);
solminus(1:4) = [-A xminus(1) xminus(2) exp(B)];
a_minus=-A;
b_minus=xminus(1);
c_minus=xminus(2);
q_minus=exp(B);

% Calculate estimated most probable largest value:
N=length(dataset);%length(dataset);
n_hour=60*N/((stop-start)*0.5);%loads per hour
P=((N-1)/N)^N;%Probability of exceeding most probable largest
eta_max_hat=b+((-1/a)*log(-(log(P))/(N*q)))^(1/c);

% For plotting:
x=eta_0:max(eta)*1.1;
y=q.*exp(-a.*(x-b).^c); %Extrapolation function
yplus=q_plus.*exp(-a_plus.*(x-b_plus).^c_plus); %Upper bound for CI
yminus=q_minus.*exp(-a_minus.*(x-b_minus).^c_minus); %Lower bound for CI

figure1 = figure;
% Create axes
axes1 = axes('Parent',figure1);
hold(axes1,'on');
xlim(axes1,[25 , max(eta)*1.25]); %Limits z-axis
ylim(axes1,[min(ACER(1:(end-1),k))*0.1 , 1]);%yplus(1)]); %Limits y-axis
title('Extrapolation scheme')
xlabel({'$\eta$ [kN/m]'},'Interpreter','latex')
ylabel({'ACER($\eta$,k=1)'},'Interpreter','latex')

dataset=sort(dataset);

% Plotting:
% ACER function
acer=stairs(eta(1:end),ACER(1:end,k),'Parent',axes1,'color',[0 0 1]);
% Extraploation function:
fit=semilogy(x,y,'-k');
% Upper confidence band for extrapolation function:
CIfit=semilogy(x,yplus,':k');
%Lower confidence band for extrapolation function:
semilogy(x,yminus,':k');
%Confidence interval bands for ACER function:
CIacer=semilogy(dataset(1:j),CIminus(1:j,k),dataset(1:end-1),...
```

```
149     CIplus(1:end-1,k),'LineStyle','--','Color',[0 0 1]);
150
151 box(axes1,'on');
152 % Set the remaining axes properties:
153 set(axes1,'YMinorTick','on','YScale','log');
154 legend([acer,fit,CIacer(1),CIfit],{'ACER function',...
155     'Extrapolation function','95% CI band for ACER function',...
156     '95% CI band for extrapolation function'});...
157     annotation('textbox',[0.18 0.18 0.1 0.07],'String',...
158     {['\eta_{0} = ',num2str(eta_0)]},'FontSize',12.6,'EdgeColor','k');
159
160 % End acer_fit.m ------------------------------------------------------------
```

### kolmogorov_smirnov.m

```
% Start kolmogorov_smirnov.m ------------------------------------------------

prompt = 'User input: Enter "1" for parameters estimated from',...
    'probability paper, "2" for maximum likelihood estimate (MLE) ';
method = input(prompt);

% Estimate distribution parameters for set defined in prob_paper.m:
run('parameters')
close

% Calculating points in CDF for fitted distribution:
X_cdf=0:1:max(dataset)*1.1;
if distr==1 % Exponential
    Y_cdf=expcdf(X_cdf,1/lambda);
    ks_cdf=expcdf(dataset,1/lambda);
elseif distr==2 %Log-normal
    Y_cdf=logncdf(X_cdf,mu,sigma);
    ks_cdf=logncdf(dataset,mu,sigma);
elseif distr==3 %Gumbel
    Y_cdf=exp(-exp(-(X_cdf-alpha)/beta));
    ks_cdf=exp(-exp(-(dataset-alpha)/beta));
elseif distr==4 %Weibull
    Y_cdf=wblcdf(X_cdf,theta,k);
    ks_cdf=wblcdf(dataset,theta,k);
elseif distr==5
    Y_cdf=a*(1-exp(-lambda1*X_cdf))+(1-a)*(1-exp(-lambda2*X_cdf));
    ks_cdf=a*(1-exp(-lambda1*dataset))+(1-a)*(1-exp(-lambda2*dataset));
end

% Plot empirical and fitted CDF:
empirical=stairs(dataset,p_i,'-k');
hold on
fitted=plot(X_cdf,Y_cdf,'color',0.3*[1 1 1]);

ks=abs(ks_cdf-p_i');

% Finding the measure "D":
D=max(ks);
% Calculate critical value of "D":
D_crit=1.36/sqrt(length(dataset)); %For alpha=0.05, n>50

if D>D_crit
    disp('KS-test: Rejected')
else
    disp('KS-test: Not rejected')
end

i=1;
```

```
while D>ks(i)
    i=i+1;
end

% Plot "D" and add labels, titles, legend, textbox
D_plot=plot([dataset(i),dataset(i)],[ks_cdf(i),p_i(i)],'r-','LineWidth',3);
title('Kolmogorov-Smirnov test')
xlabel('y')
ylabel('Cumulative Probability')
legend([empirical,fitted,D_plot],{'Empirical CDF','Fitted CDF','D'});
annotation('textbox',[0.7 0.65 0.15 0.11],'String',...
    {['D = ',num2str(D,'%.3f')],['D_{crit} = ',num2str(D_crit,'%.3f')]},...
    'FontSize',11,'FitBoxToText','off','EdgeColor','none');

% End kolmogorov_smirnov.m ----------------------------------------------
```

### bootstrapping.m

```
% Start bootstrapping.m -------------------------------------------------

clear all
tic %This script may run for a while!
prompt = 'User input: Enter "1" for parameters estimated from',...
    'probability paper, "2" for maximum likelihood estimate (MLE) ';
method = input(prompt);

% Number of bootstrapped samples:
n_boot=10000;

% Estimate distribution parameters for set defined in prob_paper.m:
run('parameters')

% Matrix for estimated parameters for each bootstrapped sample:
param=zeros(n_boot,2);

% Estimates of parameters from original dataset:
lambda_hat=lambda; %Exponential
sigma_hat=sigma; %Log-normal
mu_hat=mu; %Log-normal
beta_hat=beta; %Gumbel
alpha_hat=alpha; %Gumbel
k_hat=k; %Weibull
theta_hat=theta; %Weibull

% For plotting later: original (measured) dataset stored:
dataset_orig=dataset;

% Bootstrapping:
for i=1:n_boot
% Generate N (sample size) random numbers between 0 and 1 (cum. prob):
u=rand(N,1);

% Generate bootstrapped datasets from distribution with parameters
% calculated from the original data set:
if distr==1 %Exponential:
    dataset=-(log(1-u))/lambda_hat;
elseif distr==2 %Log-normal:
    dataset=exp(sigma_hat*norminv(u) + mu_hat);
elseif distr==3 %Gumbel:
    dataset=alpha_hat - beta_hat*log(-log(u));
elseif distr==4 %Weibull:
    dataset=theta_hat*(-log(1-u)).^(1/k_hat);
end

% Bottstrapped sample (dataset):
dataset=sort(dataset);
```

```
49
50 % Abscissa and ordinate transformations:
51 if distr==1 %Exponential:
52     x=dataset;
53     y=-log(1-p_i);
54 elseif distr==2 %Log-normal:
55     x=log(dataset);
56     y=norminv(p_i); %Normal inverse cumulative distribution function
57 elseif distr==3 %Gumbel:
58     x=dataset;
59     y=-log(-log(p_i));
60 elseif distr==4 %Weibull:
61     x=log(dataset);
62     y=log(-log(1-p_i));
63 end
64
65 % Plot (all) bootstrapped points in same figure:
66 hold on
67 scatter(x,y,'y*')
68
69 %Calculate new parameters based on the bootstrapped data.
70 bootstr=1; %Variable used to maintain the present dataset as "dataset" in
71 no_plot=1; %Used for avoiding new plots of dataset in prob_paper.m
72 %figure(2)
73
74 if method==1 % Parameter estimation based on probability paper:
75     if distr==1 %Exponential distribution
76         c=0; %The exponential distribution is by def. starting at origo
77         m=x\y.'; %solve for slope, intercept=0
78     else
79         m_c = polyfit(x,y',1);
80         m=m_c(1); %slope
81         c=m_c(2); %ordinate interception
82     end
83 end
84
85 % Estimate distribution parameters for bootstrapped samples
86 run('parameters');
87 param(i,1)=parameter1;
88 param(i,2)=parameter2;
89 end %bootstrapping process
90
91 % Plot the orsginal data set in the same figure as the bootstrapped ones:
92 if distr==1
93     %************EXPONENTIAL************
94     x=dataset_orig;
95     y=-log(1-p_i);
96
97 elseif distr==2
98     %************LOGNORMAL************
```

```
    x=log(dataset_orig);
    y=norminv(p_i); %Normal inverse cumulative distribution function

elseif distr==3
    %*************GUMBEL**************
    x=dataset_orig;
    y=-log(-log(p_i));

elseif distr==4
    %*************WEIBULL*************
    x=log(dataset_orig);
    y=log(-log(1-p_i));
end
hold on
scatter(x,y,'bx')

toc
% End bootstrapping.m ---------------------------------------------------
```